

DOI:10.19651/j.cnki.emt.2211140

融合非对称卷积的遥感图像目标检测算法*

齐梦林^{1,2} 陈炳才^{1,2} 张繁盛² 潘旭^{1,2} 彭相澍¹

(1. 新疆师范大学计算机科学技术学院 乌鲁木齐 830054; 2. 大连理工大学计算机科学与技术学院 大连 116024)

摘要: 遥感图像中的目标具有背景复杂、方向多变等特点。利用传统方法进行遥感图像目标检测过程复杂且费时,存在精度低,漏检率高等问题。针对以上问题,提出一种改进的YOLOv5-AC算法,该算法以YOLOv5s模型为基础,首先在原有的Backbone中构建非对称卷积结构,增强模型对翻转和旋转目标的鲁棒性;其次在主干网络的C3模块中引入坐标注意力机制提升特征提取能力,并使用Acon自适应激活函数激活;最后使用CIOU作为定位损失函数以提升模型定位精度。改进后的YOLOv5-AC模型在NWPU VHR-10和RSOD数据集上进行实验,平均精确度均值分别达到了94.0%和94.5%,分别比原版YOLOv5s提升了1.8%和2.3%,有效提高了遥感图像目标检测精确度。

关键词: 遥感图像;目标检测;YOLOv5;非对称卷积;注意力机制

中图分类号: TP399 **文献标识码:** A **国家标准学科分类代码:** D520.2

Object detection algorithm of remote sensing image based on asymmetric convolution

Qi Menglin^{1,2} Chen Bingcai^{1,2} Zhang Fansheng² Pan Xu^{1,2} Peng Xiangshu¹

(1. School of Computer Science and Technology, Xinjiang Normal University, Urumqi 830054, China;

2. College of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China)

Abstract: The object of remote sensing image has the characteristics of complex background and changeable direction. The process of object detection in remote sensing image using traditional methods is complex and time-consuming, with low accuracy and high rate of missed detection. To solve the above problems, we propose an improved YOLOv5-AC algorithm. This algorithm bases on the YOLOv5s model. First, an asymmetric convolution structure is built in the original Backbone to enhance the robustness of the model to flipped and rotated targets; Secondly, coordinate attention mechanism is introduced into C3 module of backbone network to improve feature extraction capability, and Acon (Activate Or Not) adaptive activation function is used for activation; Finally, we use CIOU as the location loss function to improve the positioning accuracy of the model. The improved YOLOv5-AC model was tested on NWPU VHR-10 and RSOD datasets, and the average accuracy reached 94.0% and 94.5%, respectively, 1.8% and 2.3% higher than the original YOLOv5s, which effectively improved the object detection accuracy of remote sensing images.

Keywords: remote sensing image; object detection; YOLOv5; asymmetric convolution net; attention mechanism

0 引言

近年来,我国航天事业不断发展,遥感图像越来越多的应用于军事、农业、环境检测等领域。遥感图像主要分为航空相片和卫星相片两种,遥感图像记录了地面物体的位置和类别信息,遥感图像目标检测可以获取特定目标的具体位置和类别,在诸多领域中有重要应用^[1-3]。随着人工智能

与计算机视觉技术的发展,基于深度学习的目标检测算法层出不穷,其中按阶段又分为两大类:一类是以Ren等^[4]在2016年提出的Faster-RCNN(faster regions with convolutional neural network)为代表的两阶段目标检测算法;另一类是以Liu等^[5]在2016年提出的SSD(single shot multibox detector),和Redmon等^[6]在2016年提出的YOLO(you only look once)为代表的单阶段目标检测算

收稿日期:2022-08-22

* 基金项目:国家自然科学基金(61961040,61771089)、新疆维吾尔自治区“天山青年计划”(2018Q024)、新疆自治区区域协同创新专项(科技援疆计划)(2020E0247,2019E0214)资助

法,单阶段目标检测算法由于不需要生成候选区域往往检测速度远快于两阶段目标检测算法,在实际生产使用中也以单阶段目标检测算法居多。

遥感图像目标检测的难点在于小目标物体和任意方向物体的检测,小目标物体像素低,难以提取有效信息,而任意方向的物体由于大小和方向不一,难以找出合适的锚框,容易造成漏检误检等问题。近年来许多学者都针对遥感图像目标检测提出了改进方案,Liu 等^[7]提出多区域特征融合的遥感图像目标检测算法,通过生成任意方向的候选区域,在多区域融合模块将候选区域特征与全局特征进行融合,以此来提升任意方向遥感舰船检测的精度。Chen 等^[8]提出 Plou Loss 利用角度和 IoU 进行更精确的目标定位,有效提升了多方向遥感目标的检测精度。张艳等^[9]提出结合 FPN (feature pyramid networks)^[10] 和 HRNet (high-resolution net)^[11] 的改进 SSD 算法,通过特征图融合以提升小物体的检测精度,但模型计算量大训练缓慢。李澄非等^[12]在 YOLOv4 中引入注意力模块,并对 PANet 进行改进,提出了提升小目标检测能力的 YOLOv4 算法,王恒涛等^[13]对 YOLOv5 的检测层进行调整,提出了轻量化的 YOLOv5 航拍图像检测算法。李壮飞等^[14]通过在 YOLOv5 网络中引入注意力机制来提升遥感图像小目标检测精度,闫钧华等^[15]提出融合多层次特征的 CC-YOLO 算法,对空间金字塔进行跨层通道特征融合,提升了遥感小

目标的检测精度。王明阳等^[16]提出基于关键点的遥感图像旋转目标检测算法,通过在网络结构中加入旋转因子提供角度信息,增强了网络对旋转目标的检测效果。Lim 等^[17]提出结合上下文特征和注意力机制的目标检测算法,提高了模型检测能力。

本文提出一种结合注意力机制和改进卷积结构的 YOLOv5s 改进模型 (YOLOv5-AC)。首先在 YOLOv5s 的 C3 模块中添加坐标注意力机制^[18]提升模型特征提取能力,降低误检率;针对遥感图像里的目标大小和方向具有不定性,在 C3 模块和 Conv 模块中构建非对称卷积结构,增强模型对图像翻转和旋转的特征提取能力;为了提高模型拟合非线性函数的能力,使用 Acon^[19]激活函数代替 Conv 模块中 SiLU 激活函数,由转换因子决定是否激活神经元。经实验证明,该算法可以有效提升遥感图像目标检测精度。

1 相关工作与改进

YOLOv5 (you only look once version 5) 算法由 Ultralytics LLC 公司于 2020 年 5 月提出,分为 4 个版本分别是 YOLOv5m, YOLOv5l, YOLOv5x 和 YOLOv5s。其中 YOLOv5s 是模型最小的版本,也是计算量最少的版本,对性能要求低便于部署。YOLOv5 网络结构由 Backbone 和 Head 两部分组成,其原始结构框架如图 1 所示。

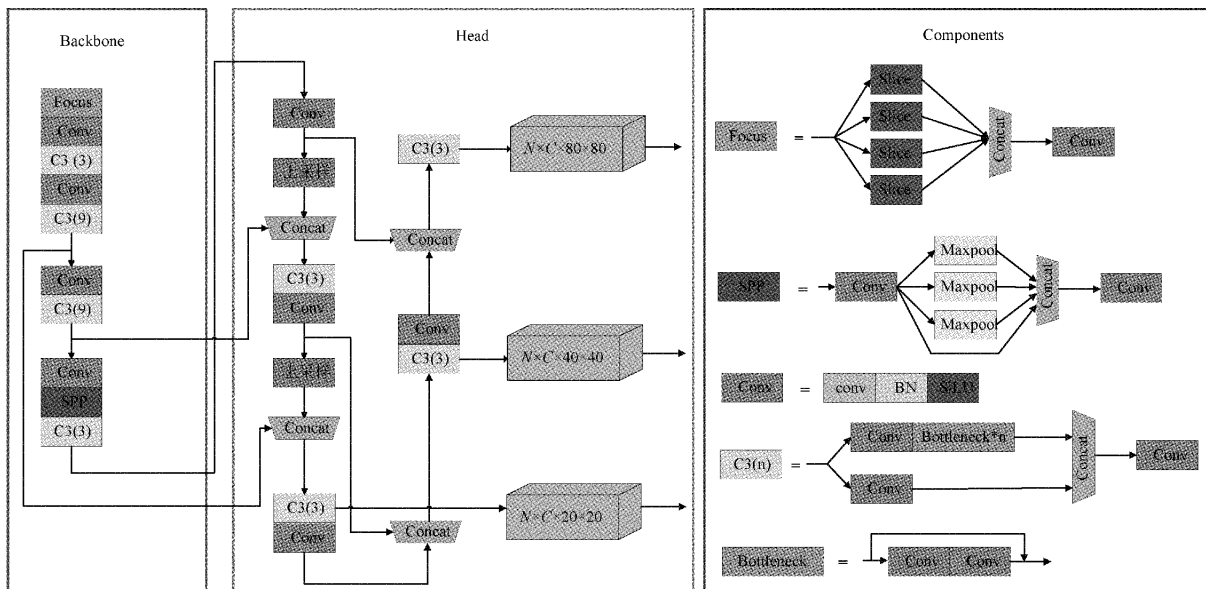


图 1 YOLOv5s 网络结构

Backbone 是 YOLOv5s 的基准网络,由 Focus、Conv、C3 和空间金字塔池化 (spatial pyramid pooling, SPP)^[20] 模块构成。其中 Focus 模块先对输入图片进行切片操作,得到 4 张图片,然后对图片进行拼接将通道数扩充为原来的 4 倍,最后将得到的新图片通过卷积操作,最终得到 $32 \times 320 \times 320$ 的特征图。Conv 模块是 YOLOv5 的卷积操作

单元,由卷积层 BN 层和 SiLU 激活函数组成。C3 模块由多个 Bottleneck 模块组成,Bottleneck 是一种残差结构由两个 Conv 模块和上一层特征组成,在深度网络中可以保留浅层网络的特征。SPP 模块输入特征先经过 Conv 模块卷积然后经过卷积核大小分别为 [5, 9, 13] 的最大池化操作,将结果经过 Concat 拼接,然后再经过 Conv 操作输出结果。

Head部分由FPN(feature pyramid network)和PANet(path aggregation network)^[21]组合而成,通过将深层的特征信息上采样与浅层的特征信息传递融合,更好的保持浅层特征与深层特征的优势,实现多尺度目标检测,提升模型特征提取能力。因为目标检测识别物体的大小不一,为了对大目标与小目标均能实现良好的检测效果,Head部分设置了大小分别为20×20、40×40和80×80三个不同尺度的特征图进行多尺度检测。

2 改进的YOLOv5算法

2.1 坐标注意力机制

注意力机制是当前计算机视觉领域的研究热点之一,被广泛应用于各种目标检测任务。注意力机制可以使模型更关注局部关键信息,通过增加权重的方式抑制无用信息突出有效信息,从而提升模型准确率。遥感图像背景复杂,在识别目标时,容易造成漏检、误检,为了有效提高识别精确度,本文在YOLOv5的C3模块中引入了CA(coordinate attention)注意力机制。CA注意力机制于2021年由Hou等^[18]提出,CA注意力机制通过对特征图的宽和高进行平均池化操作,并将位置信息嵌入到通道注意力中,有效提升了模型的特征提取能力,CA注意力机制原理如图2所示。

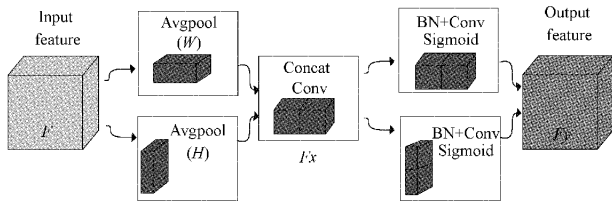


图2 Coordinate Attention原理

如图2所示为CA注意力机制的实现原理,该过程分为4步:1)首先先将原始输入特征图F的高和宽两个维度经过通道注意力的全局平均池化处理;2)将得到的特征图在高和宽维度上进行拼接与卷积操作得到中间特征图Fx;3)对中间特征图Fx进行规范化、非线性处理、对高宽两个维度进行卷积以及进行Sigmoid激活函数激活处理;4)最后将特征图Fx与原特征相乘得到经过CA注意力加权的特征图Fy。CA注意力计算如式(1)、(2)所示。

$$F_x = \text{Conv}(\text{Cat}(\text{AvgPool}(W)), (\text{AvgPool}(H))) \quad (1)$$

$$F_y = \sigma(\text{Conv}(\text{BN}(W))) \times \sigma(\text{Conv}(\text{BN}(H))) \times F \quad (2)$$

其中, F_x 表示中间特征图, W 表示当前输入特征的宽, H 表示当前输入特征的高, AvgPool表示全局平均值池化, Cat表示Concat拼接操作, Conv表示卷积操作。BN表示一系列规范化操作, σ 表示Sigmoid激活操作, F_y 为最终的输出特征。

2.2 卷积结构的改进

在遥感图像中,目标的大小不一,方向也具有任意性,

这给目标检测带来了很大的困难。为了解决这个问题,本文在Backbone阶段通过构建非对称卷积结构,以此来提升具有方向性的目标的检测能力。卷积核的大小决定了模型每次提取特征的感受野的范围,传统卷积神经网络中卷积核通常是长宽3×3、5×5等大小相等的正方形,在2019年的ICCV上Ding等^[22]提出了一种全新的卷积结构ACNet(asymmetric convolution Net),作者使用3个尺寸大小分别为3×3、1×3和3×1的卷积核代替3×3卷积,实验结果表明,将3个卷积核卷积后的特征融合后,特征提取效果优于只使用一个3×3的卷积核,对于具有方向性的目标,检测精度提升明显,ACNet结构如图3所示。

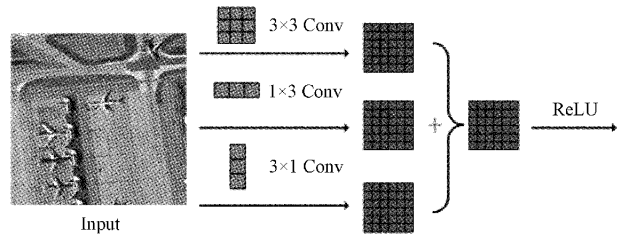


图3 ACNet结构

改进的卷积结构原理如图4所示,图4中(a)、(c)为原始图片,图4(b)、(d)为翻转旋转后的图片,可以看出当只有3×3卷积核时在特征图的相同区域卷积后会得到不同的特征,而1×3的卷积则会保留原始信息,3×1的卷积同理。实验结果表明,将3个卷积核卷积后的特征进行融合,特征提取效果优于只使用一个3×3的卷积核,对于翻转和旋转的物体,检测精度明显提升。

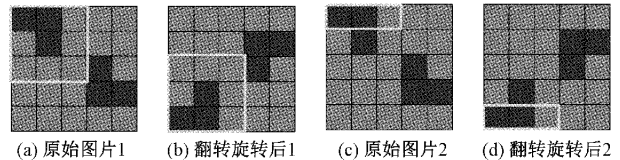


图4 非对称卷积原理

2.3 YOLOv5-AC网络结构

针对遥感图像背景复杂,目标方向多变,容易出现漏检和识别率低等问题,本文提出YOLOv5-AC模型。首先,为提升目标定位精度,降低漏检率,在主干特征提取网络的C3模块中添加CA注意力机制并命名为CAC3,改进后的CAC3模块结构如图5所示。

其次,为提高模型对方向多变目标的定位能力,受Ding等人启发本文在原始的Backbone中构建非对称卷积结构,将原有的卷积模块中增加一个1×n的横向卷积核和一个n×1的纵向卷积核。使用非对称卷积结构卷积后的融合特征代替原始特征,以增强对具有方向性目标的特征提取能力。改进后卷积的特征提取过程如图6所示,首先利用融合卷积核对输入特征图进行卷积,该卷积核是由3种尺度卷积核构成的非对称卷积,可以在提升模型特征提取能力的同时增强翻转和旋转鲁棒性。然后将得到的特

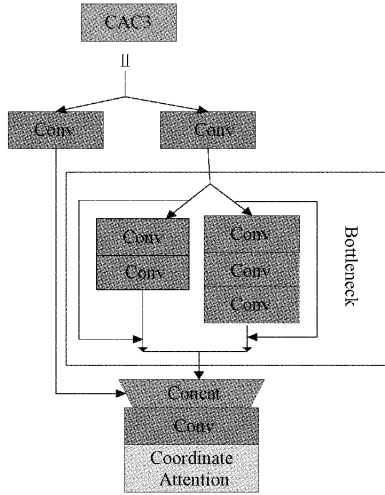


图 5 改进后的 CAC3 模块

征相加融合后进行 BN(BatchNorm2d)归一化处理,并使用 ACON 自适应激活函数进行激活,最终得的输出特征图。

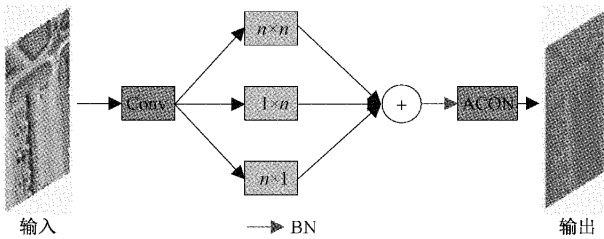


图 6 特征提取过程

YOLOv5-AC 模型推理阶段的结构如图 7 所示,该阶段使用经过 3 个卷积核融合后的卷积核参数来初始化当前网络,这个过程不增加任何计算量,只是网络参数使用了特征提取能力更的参数,即融合后的卷积核参数。

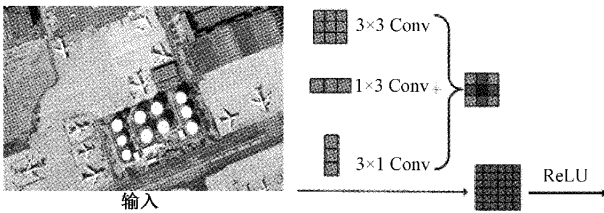


图 7 推理阶段结构

特征融合如式(3)所示。

$$F(x) = BN(Conv(x) + crop(h_c(x) + v_c(x))) \quad (3)$$

其中, x 表示输入特征, $F(x)$ 表示输出特征, BN 表示批标准化(batch normalization), $Conv$ 表示 $n \times n$ 卷积, $crop$ 表示特征图裁剪,目的是使裁剪后经过横向卷积和纵向卷积的特征图与 $n \times n$ 卷积后的特征图尺寸匹配, h_c 表示水平方向卷积操作, v_c 表示垂直方向卷积操作。

本文将改进的 CAC3 模块放在了 Backbone 和 Head 的不同位置进行比较,实验结果表明将其放在 Backbone 中效果最好,最终将改进的 CAC3 模块放在了 YOLOv5-

AC 模型 Backbone 中 C3 的位置,YOLOv5-AC 模型去除 Head 层的整体结构如表 1 所示。

表 1 YOLOv5-AC 网络结构

序号	参数	模块	arguments
0	6 112	Focus	[3, 32, 3]
1	31 424	Conv	[32, 64, 3, 2]
2	83 672	CAC3	[64, 64]
3	124 288	Conv	[64, 128, 3, 2]
4	981 192	CAC3	[128, 128, 3]
5	494 336	Conv	[128, 256, 3, 2]
6	3 879 240	CAC3	[256, 256, 3]
7	1 971 712	Conv	[256, 512, 3, 2]
8	4 595 968	SPP	[512, 512, [5, 9, 13]]
9	3 706 928	CAC3	[512, 512, False]
10	920 320	Conv	[512, 256, 1, 1]
11	0	Upsample	[None, 2, 'nearest']
12	0	Concat	[1]
13	1 745 152	C3	[512, 256, 1, False]
14	230 784	Conv	[256, 128, 1, 1]
15	0	Upsample	[None, 2, 'nearest']
16	0	Concat	[1]
17	438 400	C3	[256, 128, 1, False]
18	247 168	Conv	[128, 128, 3, 2]
19	0	Concat	[1]
20	1 286 400	C3	[256, 256, 1, False]
21	985 856	Conv	[256, 256, 3, 2]
22	0	Concat	[1]

3 实验结果与分析

3.1 实验环境与数据集

本文实验环境操作系统为 Ubuntu 16.04,CUDA 版本 10.1,CPU 型号为 Intel Xeon E5 V3 2600,内存大小 32 G, GPU 型号为 NVIDIA RTX2080TI,显存大小 11 G,深度学习框架为 Pytorch1.8.0 版本。

本文选取公开的遥感数据集 RSOD^[23-24] 和 NWPU VHR-10^[25] 进行实验和测试。RSOD 数据集是武汉大学于 2015 年制作并发布的,包含飞机图片 446 张,飞机目标 4 993 架,操场图片 189 张,操场目标 191 个,立交桥图片 176 张,立交桥目标 180 座,油罐图片 165 张,油罐目标 1 586 个。NWPU VHR-10 数据集由西北工业大学于 2014 年发布,共计 800 张图片,其中 650 张包含飞机,舰船,油罐,棒球场,网球场,篮球场,田径场,港口,桥梁,车辆十类目标,其余 250 张为负样本。实验在上述数据集上对比,进一步分析算法,并提升算法检测的性能。

3.2 损失函数

YOLOv5-AC 的损失函数包含 3 部分,分别是分类损

失,回归损失和置信度损失其中分类损失和置信度损失使用的是 BCEWithLogitsLoss 损失函数,BCEWithLogitsLoss 是在 BCELoss 的基础上融合了 sigmoid 操作,先将样本标签经过 sigmoid 函数归一化到(0,1)的范围,再使用 BCELoss 计算损失。BCELoss 的计算公式如式(4)所示。

$$BCELoss(x_i, y_i) = -w_i [y_i \log x_i + (1 - y_i) \log(1 - x_i)] \quad (4)$$

在式(4)中, $-w_i$ 是超参数, y_i 表示第 i 个样本的标签, x_i 表示第 i 个样本的预测值。

回归损失经过实验最终在 GIoU Loss^[26], DIoU Loss 和 CIoU Loss^[27] 中选择使用 CIoU Loss, CIoU Loss 计算公式如下:

$$CIoULoss = 1 - \left(IoU - \frac{D_p^2}{D_c^2} - \alpha\beta \right) \quad (5)$$

$$\text{其中, } \alpha = \frac{\beta}{(1 - IoU) + \beta}, \beta = \frac{4}{\pi^2} \left(\arctan \frac{w^{gr}}{h^{gr}} - \arctan \frac{w}{h} \right)^2,$$

D_p 表示预测框中心点与真实框中心点之间的欧式距离, D_c 表示预测框与真实框最小外接矩形对角线的距离。 $\frac{w^{gr}}{h^{gr}}$

和 $\frac{w}{h}$ 分别表示真实框的宽高比和预测框的宽高比。CIoU 是对 GIoU 的改进, CIoU 在 GIoU 的基础上考虑了检测框尺度的损失,增加了宽高比信息,使得预测框检测效果更好。

3.3 评价指标

本文采用平均准确率均值(mean average precision, MAP)作为评价指标。在评估过程中,根据准确率(precision)和召回率(recall)绘制一条 P-R 曲线,计算该曲线与坐标轴围成的面积就是 AP(average precision),准确率和召回率可以根据 TP、TN、FP、FN 计算。Precision、Recall、AP 如式(6)~(8)所示。

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$AP = \int_0^1 p(x) dx \quad (8)$$

其中, TP 为预测正确的正例, TN 为预测正确的反例, FP 为预测错误的正例, FN 为预测错误的反例, P 为准确率, r 为召回率。mAP 计算公式如式(9)所示。

$$mAP = \frac{\sum_{i=1}^n AP_i}{n} \quad (9)$$

其中, n 表示类别个数, i 表示第 i 类。

3.4 消融实验

为验证激活函数,非对称卷积结构和 CA 注意力机制对模型准确率的影响,本文分别做了两组消融实验,使用 YOLOv5-AC 模型分别使用 ACON 激活函数和 SiLU 激活

函数在 NWPU 和 RSOD 数据集上得到的结果如表 2 所示。

表 2 不同激活函数下的准确率

数据集	激活函数	mAP
RSOD	SiLU	0.940
RSOD	ACON	0.945
NWPU	SiLU	0.939
NWPU	ACON	0.940

从表 2 结果可以看出,使用 ACON 激活函数比使用 SiLU 激活函数得到的 mAP 值更高。为验证注意力机制和非对称卷积结构的作用,分别使用 4 种结构对数据集进行测试,实验结果如表 3 所示。

表 3 注意力机制和非对称卷积结构下的准确率

模型	船	汽车	立交桥	飞机
YOLOv5	0.852	0.875	0.849	0.985
YOLOv5-A	0.859	0.896	0.846	0.997
YOLOv5-C	0.850	0.877	0.858	0.987
YOLOv5-AC	0.875	0.926	0.859	0.996

表 3 中展示了准确率变化较大的几种类别,其中 YOLOv5-A 表示只使用非对称卷积结构, YOLOv5-C 表示只添加 CA 注意力机制。由表 3 结果可知,非对称卷积结构对船,汽车和飞机等具有方向性的目标提升效果明显,分别提升了 0.7%, 2.1% 和 1.2%,而注意力机制则对图片数量较少的立交桥提升较大,提升了 0.9% 的准确率。

3.5 实验结果分析

本文模型分别在 RSOD 数据集和 NWPU VHR-10 数据集上进行了训练和测试,由于两个数据集中目标种类不同,因此没有将它们混合测试。本文模型使用 One Cycle^[28] 余弦退火学习率调整策略,周期性的对学习率放大缩小,可以使模型跳出局部最优解。模型在验证集上的损失变化如图 8 所示。

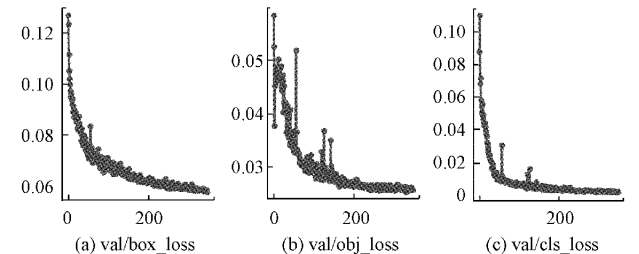


图 8 损失变化

其中, box_loss 是定位损失, obj_loss 是置信度损失, cls_loss 是分类损失,可以看出在学习率突然增大时,会使模型损失也突然增大,然后随着学习率减小,损失缓慢下降,通过周期性的放大缩小学习率,以此达到跳出局部最优解的目的,使模型效果更加收敛。为验证算法在遥感图

像目标检测时的可靠性,本文分别在 NWPU VHR-10 数据集和 RSOD 数据集上与其他算法进行了对比实验,RSOD 数据集和 NWPU VHR-10 数据集的标签分布如图 9 和 10 所示,其中类别数量分布图横坐标表示数据集中的类别,纵坐标表示具体类别的数量,类别尺度分布图横坐标表示目标归一化后的宽度,纵坐标表示目标归一化之后的高度,越靠近原点的点所代表目标尺度越小。

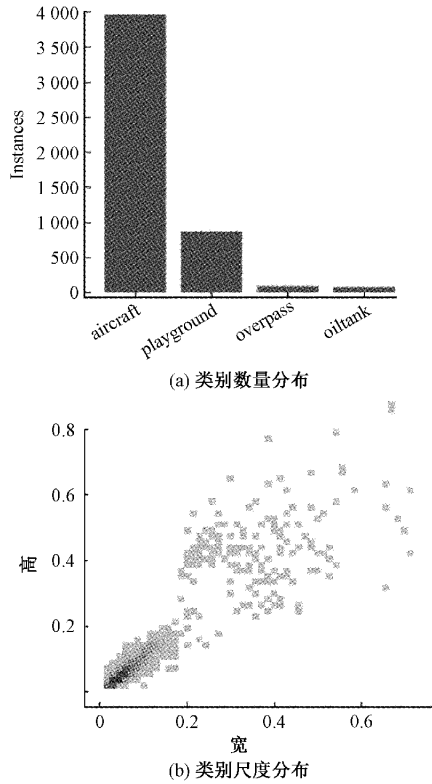


图 9 RSOD 数据集标签分布

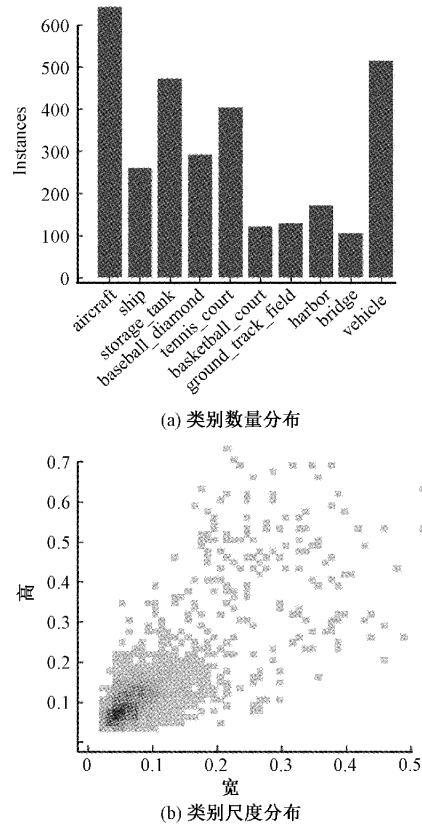


图 10 NWPU VHR-10 数据集标签分布

由图 9 和 10 可知,从尺度分布图可以看出 NWPU VHR-10 数据集的尺度更为集中,从类别数量分布图可以看出桥梁在两个数据集中样本数量都很少,所以两个数据集对桥梁的检测精度都不高,各算法在 NWPU VHR-10 数据集和 RSOD 数据集上测试结果如表 4 和 5 所示。

表 4 各算法在 NWPU VHR-10 数据集上的测试结果

目标	SSD	Faster-RCNN	YOLOv3	YOLOv3-spp	YOLOv4	YOLOv5	YOLOX	Ours
飞机	0.957	0.828	0.996	0.946	0.965	0.985	0.977	0.996
船	0.836	0.776	0.794	0.816	0.863	0.852	0.854	0.875
油罐	0.709	0.624	0.886	0.896	0.932	0.937	0.933	0.933
棒球场	0.994	0.963	0.993	0.982	0.992	0.989	0.992	0.995
网球场	0.877	0.829	0.880	0.913	0.926	0.875	0.939	0.943
篮球场	0.920	0.688	0.945	0.943	0.950	0.944	0.949	0.946
田径场	0.986	0.984	0.982	0.985	0.995	0.995	0.995	0.995
海港	0.946	0.825	0.906	0.925	0.919	0.923	0.924	0.928
桥	0.970	0.788	0.836	0.832	0.856	0.849	0.837	0.859
车辆	0.745	0.738	0.901	0.892	0.910	0.875	0.891	0.926
mAP	0.894	0.804	0.911	0.913	0.931	0.922	0.929	0.940

由表 4、5 可以看出,改进后的 YOLOv5 对几乎所有类别目标都有提升,尤其是对于飞机,车辆和船这种具有方向性的小目标,在 NWPU VHR-10 数据集上提升分别达到了 1.1%,

5.1%和 2.3%。在 NWPU VHR-10 数据集和数据集 RSOD 上各类别目标的检测效果如图 11、12 所示,可以看出改进后的 YOLOv5-AC 算法对各类别目标都有不错的检测效果。

表5 各算法在RSOD数据集上的测试结果

目标	SSD	Faster-RCNN	YOLOv3	YOLOv3-spp	YOLOv4	YOLOv5	YOLOX	本文
飞机	0.799	0.769	0.926	0.952	0.955	0.941	0.945	0.956
油罐	0.934	0.969	0.975	0.967	0.972	0.951	0.957	0.969
立交桥	0.830	0.890	0.832	0.818	0.842	0.805	0.850	0.856
操场	1.000	1.000	1.000	0.990	0.995	0.991	1.000	1.000
mAP	0.890	0.907	0.933	0.932	0.941	0.922	0.938	0.945

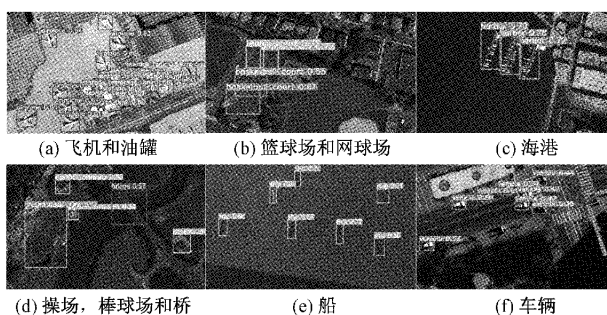


图11 NWPU VHR-10数据集上各类别目标的检测结果



图12 RSOD数据集上各类别目标的检测结果

4 结 论

针对遥感图像目标检测算法中精度低误检率高的问题,本文提出了一种基于YOLOv5的改进算法YOLOv5-AC,通过在改进的C3模块中引入坐标注意力机制并改进卷积层,并在主干网络中构建非对称卷积结构,将经过3种尺寸卷积核卷积后的特征进行相加融合并使用ACON激活函数进行激活,有效提升了模型的特征提取能力和对翻转旋转目标的鲁棒性。最终在NWPU VHR-10和RSOD数据集上mAP达到了94.0%和94.5%,相比于其它模型有明显提升,下一步将考虑使用轻量级特征提取网络作为主干网络,在保证精度的前提下尽可能提升检测速度。

参考文献

- [1] WU J, CAO C Q, ZHOU Y D, et al. Multiple ship tracking in remote sensing images using deep learning[J]. Remote Sensing, 2021, 13(18): 3601.
- [2] WANG P, LIU J Z, ZHANG Y B, et al. A novel cargo ship detection and directional discrimination method for remote sensing image based on lightweight network [J]. Journal of Marine Science and Engineering, 2021, 9(9): 932.

- [3] LIU J P. RETRACTED ARTICLE: Mountain soil characteristics and agricultural economic growth based on high-resolution remote sensing images[J]. Arabian Journal of Geosciences, 2021, 14(15): 1-15.
- [4] 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, 2016, 39(6): 1137-1149.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
- [6] 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.
- [7] LIU Q W, XIANG X Q, YANG Z, et al. Arbitrary direction ship detection in remote-sensing images based on multitask learning and multiregion feature fusion [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(2): 1553-1564.
- [8] CHEN Z M, CHEN K A, LIN W Y, et al. Piou loss: Towards accurate oriented object detection in complex environments[C]. European Conference on Computer Vision. Springer, Cham, 2020: 195-211.
- [9] 张艳, 杜会娟, 孙叶美, 等. 基于改进SSD算法的遥感图像目标检测[J]. 计算机工程, 2021, 47(9): 8.
- [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, 2017: 2117-2125.
- [11] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5693-5703.
- [12] 李澄非, 蔡嘉伦, 邱世汉, 等. 基于改进YOLOv4的铝材表面缺陷检测方法[J/OL]. 中国测试: 1-7[2022-10-

- 21]. <http://kns.cnki.net/kcms/detail/51.1714.TB.20221007.1346.002.html>.
- [13] 王恒涛,张上,陈想,等.轻量化无人机航拍目标检测算法[J/OL].电子测量技术:1-8[2022-08-21].<http://kns.cnki.net/kcms/detail/11.2175.TN.20220817.1209.002.html>.
- [14] 李壮飞,杨风暴,郝岳强.一种基于残差网络优化的航拍小目标检测算法[J].国外电子测量技术,2022,41(8):27-33,DOI:10.19652/j.cnki.femt.2203998.
- [15] 闫钧华,张琨,施天俊,等.融合多层次特征的遥感图像地面弱小目标检测[J].仪器仪表学报,2022,43(3):221-229,DOI:10.19650/j.cnki.cjsi.J2108699.
- [16] 王明阳,王江涛,刘琛.基于关键点的遥感图像旋转目标检测[J].电子测量与仪器学报,2021,35(6):102-108,DOI:10.13382/j.jemi.B2003525.
- [17] LIM J S, ASTRID M, YOON H J, et al. Small object detection using context and attention[C]. 2021 International Conference on Artificial Intelligence in Information and Communication(ICAIC), 2021: 181-186.
- [18] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [J]. ArXiv Preprint, 2021, ArXiv:2103.02907.
- [19] MA N N, ZHANG X Y, LIU M, et al. Activate or Not: Learning customized activation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 8032-8042.
- [20] HE K M, ZHANG X Y, REN S Q, 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.
- [21] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [22] DING X H, GUO Y C, DING G, et al. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1911-1920.
- [23] LONG Y, GONG Y P, XIAO Z F, et al. Accurate object localization in remote sensing images based on convolutional neural networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(5): 2486-2498.
- [24] XIAO Z F, LIU Q, TANG G F, et al. Elliptic Fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images [J]. International Journal of Remote Sensing, 2015, 36(2): 618-644.
- [25] CHENG G, HAN J W, ZHOU P C, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 98:119-132.
- [26] REZATOFIHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658-666.
- [27] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [28] SMITH L N. A disciplined approach to neural network hyper-parameters: Part 1—learning rate, batch size, momentum, and weight decay[J]. ArXiv Preprint,2018, ArXiv:1803.09820.

作者简介

齐梦林,硕士研究生,主要研究方向为计算机视觉,深度学习,目标检测等。

E-mail:821219924@qq.com

陈炳才(通信作者),博士,教授,博士生导师,主要研究方向为计算机视觉,物联网等。

E-mail:china@dlut.edu.cn