

基于DSM-YOLO v5的无人机航拍图像目标检测

陈卫彪^{1,2}, 贾小军^{1,2}, 朱响斌¹, 冉二飞^{2,3}, 谢昊^{2,3}

1. 浙江师范大学 计算机科学与技术学院, 浙江 金华 321004

2. 嘉兴学院 信息科学与工程学院, 浙江 嘉兴 314001

3. 浙江理工大学 计算机科学与技术学院(人工智能学院), 杭州 310018

摘要:针对传统无人机航拍图像目标检测算法存在漏检率高、检测成功率低、模型体积大等问题,提出一种新的基于深度可分离多头网络结构的目标检测方法DSM-YOLO v5(depthwise separable multiplex YOLO v5)。通过在YOLO v5网络结构上增加一个尺寸为160×160的小目标检测头,并将其与高层网络进行残差连接,以提升小目标检测能力,同时在Conv模块中引入深度可分离卷积算法,将其中的普通卷积替换为深度可分离卷积,能有效减少网络的参数量,降低模型体积。实验结果表明,基于DSM-YOLO v5网络结构的目标检测的mAP@0.5为36.8%,较YOLO v5s提高3.6个百分点,参数量和模型体积较YOLO v5s整体下降21.1%和19.7%,能够有效地应用于无人机航拍图像目标检测任务。

关键词:DSM-YOLO v5; 目标检测; 深度可分离; YOLO v5; 无人机图像

文献标志码:A **中图分类号:**TP391.4 **doi:**10.3778/j.issn.1002-8331.2302-0324

Target Detection for UAV Image Based on DSM-YOLO v5

CHEN Weibiao^{1,2}, JIA Xiaojun^{1,2}, ZHU Xiangbin¹, RAN Erfei^{2,3}, XIE Hao^{2,3}

1. School of Computer Science and Technology, Zhejiang Normal University, Jinhua, Zhejiang 321004, China

2. College of Information Science and Engineering, Jiaxing University, Jiaxing, Zhejiang 314001, China

3. School of Computer Science and Technology (School of Artificial Intelligence), Zhejiang Sci-Tech University, Hangzhou 310018, China

Abstract: Aiming at the problems of high missed detection rate, low detection success rate and large model size in traditional UAV (unmanned aerial vehicle) aerial image target detection algorithms, a new target detection method based on depthwise separable multiplex networks structure, DSM-YOLO v5 (depthwise separable multiplex YOLO v5) is proposed. By adding a small target detection head with a size of 160×160 to the YOLO v5 network structure and performing residual connections with the high-level network to improve the detection ability of small objects. At the same time, the depthwise separable convolution algorithm is introduced in the Conv module, replacing ordinary convolution with depthwise separable convolution, which can effectively reduce the number of network parameters and reduce the model size. The experimental results show that the mAP@0.5 of target detection based on DSM-YOLO v5 network structure is 36.8%, which is 3.6 percentage points higher than that of YOLO v5s. The parameter volume and model volume are reduced by 21.1% and 19.7% respectively compared with YOLO v5s, which can be effectively applied to UAV aerial image target detection tasks.

Key words: DSM-YOLO v5; target detection; depthwise separable; YOLO v5; unmanned aerial vehicle(UAV) image

随着无人机(unmanned aerial vehicle, UAV)技术的发展和制造成本的降低,无人机的应用愈发地深入各行各业。目标检测任务是无人机应用最重要的组成部分,广泛应用于抢险救灾、测绘航测、应急救援等领域。

目前目标检测算法主要分为两类,分别是以背景减法、光流法和帧差法为代表的传统算法和以深度学习为代表的新型算法。在航拍图像的目标检测任务中, Yalcin等^[1]提出了利用光流法检测运动目标的特征,该

基金项目:浙江省公益技术应用研究计划(LGG20F010010)。

作者简介:陈卫彪(1997—),男,硕士研究生,研究方向为图像处理、计算机视觉;贾小军(1974—),通信作者,男,博士,教授,硕士生导师,研究方向为图像处理与模式识别、计算机视觉等, E-mail: xjjia@zjxu.edu.cn。

收稿日期:2023-02-23 **修回日期:**2023-06-19 **文章编号:**1002-8331(2023)18-0226-08

方法对运动目标检测效果明显,但检测精度以及对静止目标检测效果差。Mei等^[2]提出模板匹配检测目标的方法。将视频中含有运动图像的每帧单独提取,然后用提前准备好的车辆等物体的模板进行匹配检测。该方法对车辆等形状明显的物体检测效果较好,但存在计算量过大、难以部署的问题。

以深度学习为基础的目标检测算法又分为两类,分别是单阶段式(one-stage)和两阶段式(two-stage)。YOLO(you only look once)^[3]是一种典型的单阶段式目标检测算法,它直接将目标边界定位问题转化为回归问题,图像被缩小到统一尺寸,并以网格进行均等划分。模型仅需要处理图像一次便能得到边界框坐标和类概率。两阶段式的检测算法先生成一系列的候选边界框作为样本,而后由卷积神经网络对样本进行分类,例如R-CNN^[4]、Fast R-CNN^[5]、Faster R-CNN^[6]、R-FCN^[7]等。张瑞倩等^[8]根据多尺度空洞卷积感受野大的特点,提出一种新的网络结构,提升了无人机航拍图像目标检测任务的查准率和查全率,但未能对网络做轻量化处理。陈旭等^[9]通过增加特征融合模块、残差空洞卷积等方式,有效地提高了网络对无人机航拍图像的检测能力,改造之后的模型具有遮挡目标识别率高、虚警率低的优点,但网络所需要的算力大幅增长。韩玉洁等^[10]以YOLO v4为基础,通过改变激活函数、损失函数和以自对抗训练的方式增强数据集,有效地提高了网络在无人机航拍图像目标检测任务中的检测精度,但依旧存在小目标检测能力差、网络模型体积大的问题。刘鑫等^[11]以CenterNet为基础,通过增加自适应基础模块、全局注意力模块和高质量解码模块,在模型复杂度不变的前提下,提升了网络的检测精度,但网络所需参数量大、算力要求高,难于部署在无人机平台中。王浩雪等^[12]针对无人机航拍图像小目标多、易被遮挡的特点,提出了S-YOLO v4网络,通过增加SE(squeeze and excitation)注意力模块、小目标检测层和改进损失函数的方式,增强了网络的鲁棒性,提升了网络的检测精度,但带来了网络体积和总参数量变大的问题。王胜科等^[13]以CenterNet为基础,通过增加可变形双重注意力机制、改变极大值抑制方式和引入LegoNet卷积单元的手段,有效提高了网络的检测精度。但在模型的推理速度和模型体积轻量化改进上还有待于提升。Zhu等^[14]通过添加Transformer Encoder模块和注意力机制,提高了无人机目标检测任务的检测精度,刷新了VisDrone2021数据集的检测纪录,但算法所需的参数、网络层数以及算力大幅度提高。Liu等^[15]通过修改YOLO v3的网络结构,增加ResNet模块,在高层网络结构添加额外的卷积结构以丰富空间信息,提高了小目标物体的检测能力。Lin等^[16]通过使用多尺度融合的特征聚合框架和高精度的上采样方法,改善了在高分辨率图像中的采样精度不足问

题。Zhang等^[17]通过密集裁剪和增加局部注意力模块的方式,改善了行人等目标由于占图像面积小、检测率低的问题,有效提高了对于无人机视角下的行人检测精度。Luo等^[18]在YOLO v5l的基础上,对主干网络中的特征提取模块进行替换,并通过在Focus模块中增加注意力机制、优化锚框尺寸等方法,提高了网络的精度,但其网络参数量高达2 700万,难以部署至无人机等边缘设备中。Zhang等^[19]通过在YOLO v5中增加CA(coordinate attention)注意力机制和CFEM(context feature enhancement module)模块,提高了网络对小目标的检测能力,但增加的模块也带来了参数量和算力的提升。Yang等^[20]在YOLO v5的基础上,通过优化锚框尺寸、增加注意力机制、添加小目标检测层的方式提高了检测精度,但未考虑网络的参数量增加和算力需求变大等问题。李壮飞等^[21]基于YOLO v5网络,在其中引入SENet注意力机制、增加小目标检测层、改进损失函数,提高了对于小车辆目标的查准率,但同时也带来了计算量增加的问题。苏凯第等^[22]通过增加卷积层数来提高算法深度,采用多次循环神经网络提高训练速度,有效地提高了YOLO v5的识别速度和精度,但也带来了更多的参数和更高的算力要求,难以适配无人机等边缘设备。

目前主流的目标检测算法采取基于卷积神经网络的深度学习方式,其特征提取和表达能力强,由于这些算法多用于自然视角下的目标检测场景,在对无人机航拍图像数据中的车辆、行人等小目标进行检测时,难以取得让人满意的效果。与普通场景的目标检测相比,无人机航拍图像存在以下几个特点:(1)检测目标像素占比少;(2)无人机硬件限制,难以提供足够的算力和空间;(3)航拍图像复杂,外界干扰因素大。

针对以上问题,本文以YOLO v5算法为基础,提出一种基于深度可分离的多头网络结构,将其命名为DSM-YOLO v5(depthwise separable multiplex YOLO v5)算法,通过增加检测头的数量提升检测精度,增加深度可分离结构降低网络的算力要求。与当前主流无人机航拍图像目标检测算法不同的是,在聚焦提高网络检测能力的同时降低了网络的参数量。

1 DSM-YOLO v5网络结构

1.1 YOLO v5网络结构

YOLO v5作为目前目标检测领域中的主流算法,分为v5s、v5m、v5l、v5x四个版本,可以根据不同的应用领域动态调整网络结构和深度。基于无人机平台的算力和内存空间有限的特点,本研究选择参数量最小的YOLO v5s^[23]进行对比实验。YOLO v5s主要由Backbone、Head和Output层三部分组成,相较于YOLO v4,Backbone中添加了focus层,能够对特征图进行切片,提高特征提取能力。Head层保持和YOLO v4同样的网

络结构,其主要功能有特征融合、目标检测,采取自顶向下的网络结构,将高层特征与底层特征融合,以此加强网络的特征融合能力。通过对不同尺度的特征图进行检测,最大限度地提高目标检测的精度,Output层则根据预设好的三个锚框对输出特征图进行预测,使用非极大值抑制去除低可信度的数据,最终输出检测的结果。

然而,在应对无人机航拍图像目标检测任务时,YOLO v5s算法所需要的体积和算力仍然超出目前无人机平台所能提供的硬件算力,难以有效部署在无人机等移动设备上实时检测。为了解决这一问题,对YOLO v5s算法进行轻量化改进具有重要的实用价值。

1.2 深度可分离算法

深度可分离算法由谷歌团队提出,其核心思想是将一个普通卷积替换成深度卷积和逐点卷积^[24]。深度卷积先用Channel为1的滤波器对每个输入通道进行单独卷积,在此过程之中,Padding及Stride均为1,保证输出特征图的尺寸和通道保持不变。在逐点卷积的过程中,对深度卷积输出的特征图使用大小为 1×1 的滤波器进行卷积,最终输出特征图,特征图的尺寸和通道数都和普通卷积输出的结果一致。深度可分离模块结构如图1所示。

假设普通卷积使用 M 个大小为 D_K 的卷积核,输入特征图的通道数为 N ,大小为 D_F ,则普通卷积的计算量 F_{conv} ^[25]为:

$$F_{conv} = D_F \times D_F \times N \times M \times D_K \times D_K \quad (1)$$

在使用深度可分离模块替换普通卷积的前提下,计算分为两个步骤,分别是深度卷积和逐点卷积,其计算量 F_{depth} ^[25]为:

$$F_{depth} = D_K \times D_K \times N \times D_F \times D_F + N \times M \times D_F \times D_F \quad (2)$$

用公式(1)除以公式(2),最终得到的结果便是:

$$\frac{F_{conv}}{F_{depth}} = \frac{1}{M} + \frac{1}{D_K^2} \quad (3)$$

由于在MobileNet网络中,卷积核的大小 D_K 一般为3,由公式(3)可以推出,深度可分离模块的计算量为普通卷积模块计算量的 $1/9 \sim 1/8$ ^[25]。

1.3 改进的YOLO v5网络结构

在无人机航拍图像目标检测中,由于检测目标所占图像的比例较小,传统YOLO v5s的Backbone层难以有效提取目标物的特征,导致最终的查准率、查全率低,并且在Head层生成特征图的过程中,未与网络高层进行残差连接,致使特征图中丢失部分检测目标细节。为了提高YOLO v5s模型的小目标感知能力,在原有三个检测头的基础上加深网络结构,额外添加一个新的检测头,该检测头在高层网络中,输出 160×160 大尺寸特征图,由于YOLO v5s网络整体下采样倍数较大,而无人机航拍图像中小目标物体面积占比极小,经过高倍数下采样之后,小目标物体的特征已和背景融为一体,难以提取。相较于传统YOLO v5s的三个小尺寸 80×80 、 40×40 、 20×20 检测头,大尺寸检测头能够从网络高层提取小目标特征,有效加强对于小目标物体的感知能力。在Head层中,为了提高网络对于小目标物体特征的提取能力,将Head层与Backbone的高层进行残差连接。

为了降低网络的参数,将原有Conv模块中的普通卷积替换为深度可分离卷积,组成新的深度可分离卷积模块Conv_ds,加入深度可分离卷积模块Conv_ds后,DSM-YOLO v5相较于YOLO v5s,额外加深了网络层数,使得DSM-YOLO v5具有更好的特征表达能力。虽然网络层数增加了,但是由于深度可分离模块的特点,整个网络的体积及参数量却降低了。通过残差连接的方式,加深了特征融合的能力,能更有效地提取小目标物体的信息。改进的整个网络结构如图2(a)所示,图2(b)为其中的子模板结构。

2 实验结果与分析

2.1 实验数据

实验采用VisDrone2019^[26]数据集。VisDrone数据集由天津大学机器学习与数据挖掘实验室的AISKEYE团队发布,数据集取自中国14个不同的城市,环境多种多样,城市与乡村、各种天气和照明条件。图像检测目

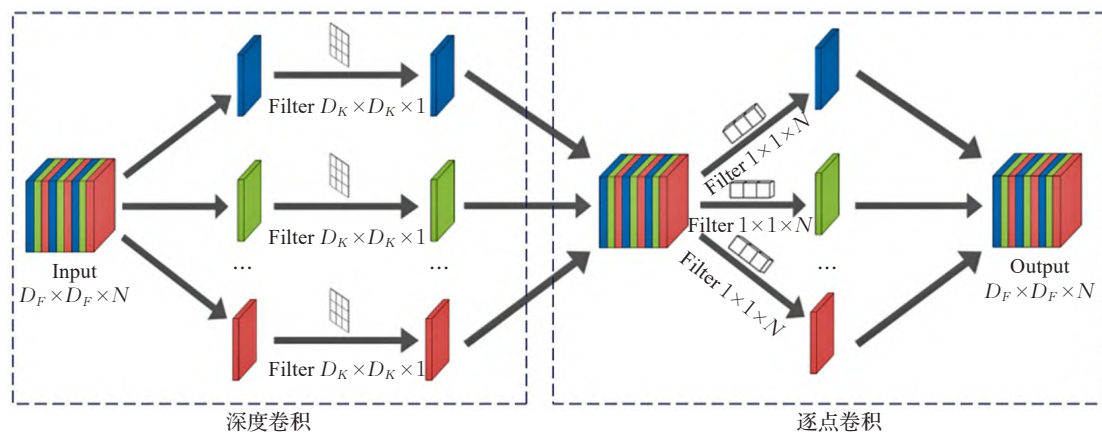
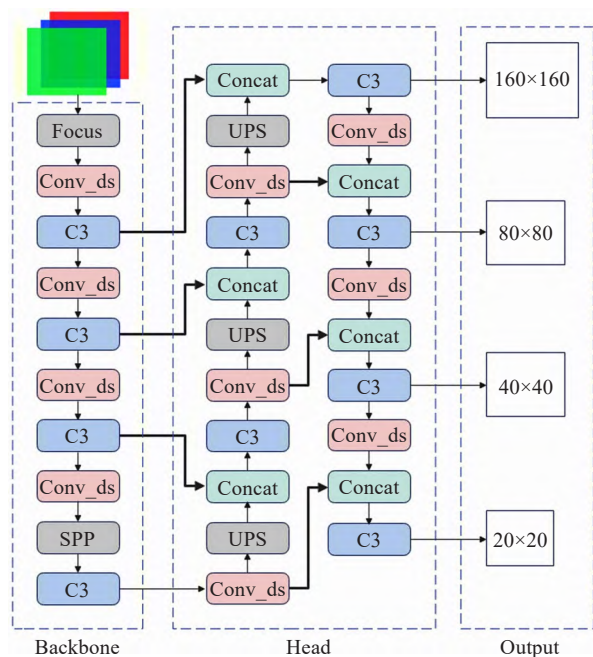
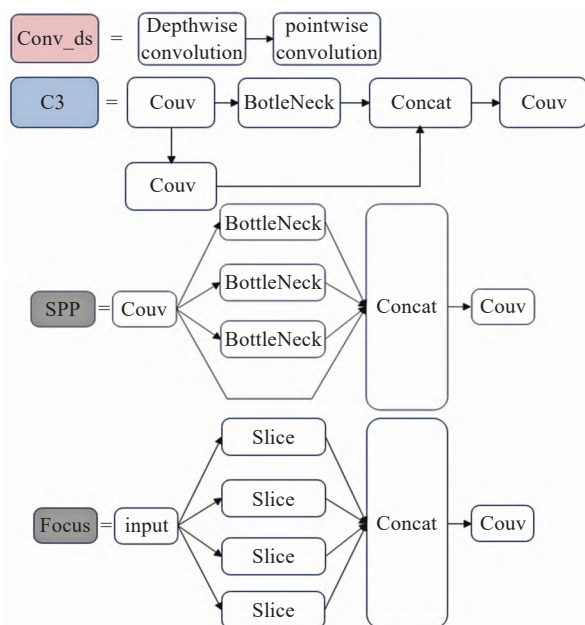


图1 深度可分离模块结构

Fig.1 Depthwise separable model block structure



(a)DSM-YOLO v5网络结构

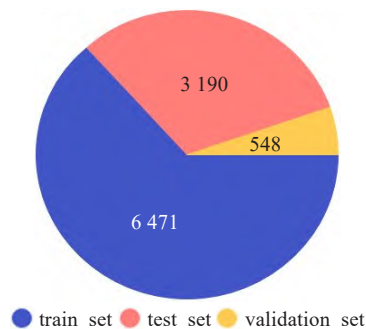


(b)DSM-YOLO v5子模块结构

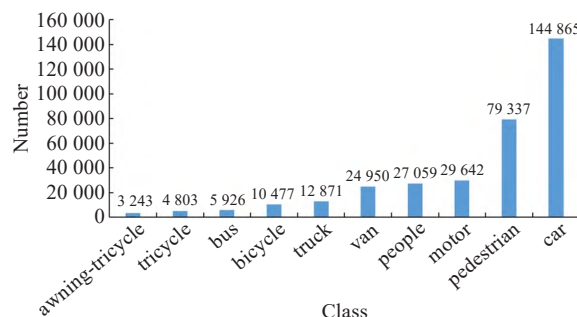
图2 DSM-YOLO v5网络结构

Fig.2 DSM-YOLO v5 network structure

标密度广,稀疏与密集均有,是目前中国无人机航拍数据量最大、覆盖范围最广、环境最复杂的数据集之一。如图3所示,VisDrone2019数据集共包含10类航拍检测目标,由6471张训练集、548张验证集、3190张测试集图片组成。其中,训练集平均每张图片含有53个标签,测试集平均有71个标签,并且每个类别均有不同程度的遮挡。数据集类别由pedestrian、people、bicycle、car、van、truck、tricycle、awning-tricycle、bus、motor组成。由于使用的网络结构为深度可分离模块,所以在预训练权重上使用了基于imagenet数据集训练的mobilenetv1权重文件。



(a)数据集分类



(b)检测目标类别

图3 VisDrone2019数据集

Fig.3 VisDrone2019 dataset

为了方便训练,将VisDrone的文件结构修改为PASCAL VOC结构,并在训练开始前,对所有数据使用Mosaic技术进行数据增强^[27]。随机选取四张图片,对其进行翻转、缩放、色域调整等操作,再将其组合成一张新的图片,以此增强数据样本的多样性,并且能够在硬件资源有限的情况下,提高网络的训练能力。

2.2 训练环境

实验平台采用12核Intel® Xeon® Platinum 8255C CPU @ 2.50 GHz,内存为45 GB,操作系统为Ubuntu 20.04。以Python 3.8实现网络模型,代码编辑器采用PyCharm Community Edition版本,使用Pytorch框架。显卡为RTX2080Ti,cuda版本为11.3。

2.3 评价指标

为更有效直观地展示DSM-YOLO v5的改进效果,以平均精度值(mean average precision, mAP)、查全率(recall)、查准率(precision)作为网络结构在无人机航拍图像目标检测任务中学习能力的的评价依据。以网络模型的体积大小和所需要的参数量作为网络对无人机航拍图像目标检测任务适应能力的指标。模型需要的体积和占用的参数越小,则代表算法更加匹配无人机航拍图像目标检测任务。

2.4 实验结果

整个训练过程迭代300个epoch。为保证模型的稳定,在前3个epoch采用warmup预热学习策略,初始学习率设置为0.01,后续则采用余弦退火算法控制学习率衰减,余弦退火算法超参数lrf为0.2,最低学习率为0.002。经过300个epoch的训练之后,得到的损失曲线

如图4所示,其中Box_loss代表模型预测边界框和真实边界框之间的差异大小,Class_loss为分类损失,用于判断模型是否能准确识别出图像中的对象,并将其分类到正确的类别中。Object_loss为置信度损失,用于监督grid中是否存在物体,计算网络的置信度。从图4中可以看出,当epoch为300轮时,DSM-YOLO v5网络的各项loss值已经不再下降,网络已收敛,趋于稳定。

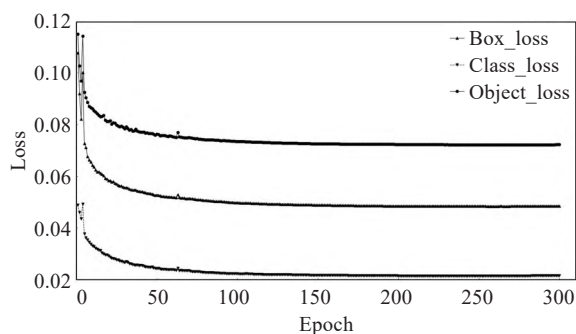
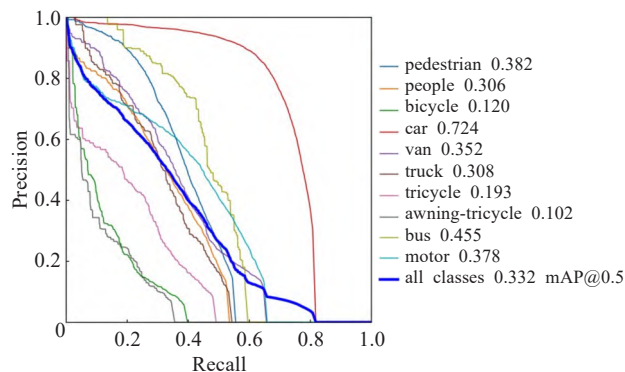


图4 DSM-YOLO v5 损失曲线

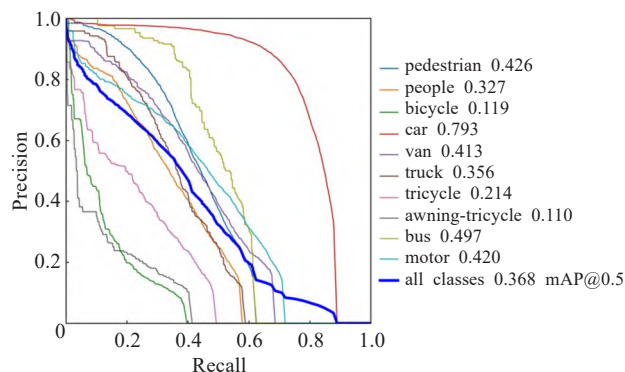
Fig.4 DSM-YOLO v5 loss curve

利用训练后的网络结构进行目标检测,使用的数据为测试集的样本数据。如图5所示,其中图5(a)为YOLO v5s的实验结果,图5(b)为DSM-YOLO v5的实验结果。可以看出DSM-YOLO v5在除bicycle分类外,其他各个分类的精度都优于YOLO v5s,pedestrian、people、car、van、truck、tricycle、awning-tricycle、bus、motor九个类别分别提高0.044、0.021、0.069、0.061、0.048、0.021、0.008、0.042、0.042。在IoU(intersection over union)为0.5时,DSM-YOLO v5的mAP值为36.8%,高于YOLO v5s。

为了对网络结构性能进行验证,选取DetNet59^[28]、CornerNet^[28]、Fast R-CNN^[29]、CenterNet^[30]、MixedYOLOv3-LITE^[31]、YOLO v5s算法进行目标检测对比实验,实验结果如表1所示。在mAP@0.5上,DSM-YOLO v5取得了明显的提升,相较于DetNet59提高21.5个百分点、CornerNet提高19.4个百分点、Fast R-CNN提高15.1个百分点、CenterNet提高10.6%、MixedYOLOv3-LITE提高8.3个百分点、YOLO v5s提高了3.6个百分点。在pedestrian、people、car、van、truck、tricycle、bus以及motor八个类别上达到了最大的AP,分别为42.6%、32.7%、



(a)YOLO v5s



(b)DSM-YOLO v5

图5 Precision-Recall 曲线图

Fig.5 Precision-Recall curve

79.3%、41.3%、35.6%、21.4%、49.7%以及42%。实验结果表明,改进的算法能够较明显地提高网络对于小目标的检测能力,在无人机航拍图像目标检测任务中表现较好。

相较于普通的目标检测任务,无人机航拍图像对网络结构的特征提取能力要求更高,对目标的精确识别难度更大,传统的目标检测算法难以满足要求。DSM-YOLO v5通过加深网络结构,增加浅层和深层网络残差连接,以及增加小目标检测头的方式,有效地提高了网络对于无人机航拍图像目标检测任务的识别能力,对于复杂场景、高密度场景的目标,经过高倍率下采样之后,其特征已经无法识别,密集目标和背景混杂在一起,和深层连接的检测头难以识别。新增的160×160检测头直接和浅层网络相连接,能够在进行高倍率下采样之

表1 不同算法在VisDrone2019数据集上的目标检测结果

Table 1 Target detection results of different algorithms on VisDrone2019 dataset

Networks	AP/%										mAP@0.5/%
	pedestrian	people	bicycle	car	van	truck	tricycle	awning-tricycle	bus	motor	
DetNet59 ^[28]	15.3	4.1	3.1	36.1	17.3	20.9	13.5	10.5	26.0	10.9	15.3
CornerNet ^[28]	20.4	6.6	4.6	40.9	20.2	20.5	14.0	9.3	24.4	12.1	17.4
Fast R-CNN ^[29]	21.4	15.6	6.7	51.7	29.5	19.0	13.1	7.7	31.4	20.7	21.7
CenterNet ^[30]	22.6	20.6	14.6	59.7	24.0	21.3	20.1	17.4	37.9	23.7	26.2
MixedYOLOv3-LITE ^[31]	34.5	23.4	7.9	70.8	31.3	21.9	15.3	6.2	40.9	32.7	28.5
YOLO v5s	38.2	30.6	12.0	72.4	35.2	30.8	19.3	10.2	45.5	37.8	33.2
DSM-YOLO v5	42.6	32.7	11.9	79.3	41.3	35.6	21.4	11.0	49.7	42.0	36.8

表2 YOLO v5与DSM-YOLO v5各项指标对比
Table 2 Comparison of indicators between YOLO v5 and DSM-YOLO v5

Networks	Parameters/10 ⁶	Model Size/MB	Recall/%	Precision/%	mAP@0.5/%
YOLO v5s	7.10	13.70	36.00	47.60	33.20
YOLO v5l	46.68	89.40	39.00	57.00	39.00
DSM-YOLO v5	5.60	11.00	37.00	51.10	36.80

前获得复杂场景、高密度场景的目标的特征信息,能有效识别目标。由于无人机硬件本身的限制,小型无人机平台对于算法的参数数量、模型体积比较敏感,DSM-YOLO v5与YOLO v5中的两种模型s型和l型的性能比较结果如表2所示,DSM-YOLO v5通过使用深度可分离的思想,将传统YOLO v5s的Conv普通卷积模块替换为深度可分离模块Conv_ds,有效地减小了网络的参数数量和模型体积,YOLO v5l虽然在检测精度上小幅优于DSM-YOLO v5,但其所需的参数数量和模型的体积分别为DSM-YOLO v5的8.4倍和8.1倍。DSM-YOLO v5在保证精度的前提下,大大降低了网络结构对硬件的要求,更有利于算法在无人机等小型设备上使用。

为了验证DSM-YOLO v5网络的有效性,选择额外的DroneVehicle数据集和RSOD数据集进行测试。DroneVehicle数据集为无人机航拍下的红外汽车图像数据集,其中训练集共有17 957张图片,测试集8 980张图片,验证集共有1 467张图片,共含有car目标428 086个、truck目标25 960个、bus目标16 590个、van目标12 708个和feright car目标17 173个。RSOD为遥感图像数据集,其中训练集818张图片,测试集99张图片共含有aircraft目标4 993个、oiltank目标1 586个、overpass目标180个和playground目标191个。DSM-YOLO v5的检测结果如表3所示,对于DroneVehicle数据集,AP值最高为car类98.1%,Recall最高为car类95.5%,Precision最高为car类94.9%,mAP@0.5为81.3%。在RSOD数据集中,AP值最高为aircraft类98.8%,Recall最高为overpass类98.9%,Precision最高为oiltank类98.7%,mAP@0.5为90.6%。实验数据表明DSM-YOLO v5的检测效果较好,在不同的数据集中能达到令人满意的检测效果。与VisDrone2019相比,DroneVehicle数据集属于红外图像,RSOD数据集属于遥感图像,整体色调、明暗等画面风格统一,减小了目标检测的难度,DroneVehicle和RSOD

数据集中的目标轮廓明显,不存在重叠遮挡、目标俯视图不会随视角大幅度变化。

表3 DSM-YOLO v5在不同数据集上的检测结果
Table 3 Detection results of DSM-YOLO v5 on different dataset

Dataset	Class	AP/%	Recall/%	Precision/%	mAP@0.5/%
DroneVehicle	car	98.1	95.5	94.9	81.3
	truck	74.2	70.1	77.4	
	bus	96.5	92.8	94.0	
	van	65.5	62.1	73.1	
	feright car	72.1	62.4	78.3	
RSOD	aircraft	98.8	90.9	98.1	90.6
	oiltank	98.1	91.5	98.7	
	overpass	92.5	98.9	72.7	
	playground	73.1	82.1	66.7	

为了验证DSM-YOLO v5中所提出的大尺寸检测头和MobileNet轻量化改进的实际效果,对DSM-YOLOv5网络进行消融实验,实验结果如表4所示。其中第一行到第四行分别屏蔽不同尺寸的检测头。屏蔽160×160检测头时,mAP@0.5为33.2%,屏蔽80×80检测头时,mAP@0.5为34.8%,屏蔽40×40检测头时,mAP@0.5为36.3%,屏蔽20×20检测头时,mAP@0.5为38.1%,从数据可以看出,160×160的检测头对精度的提升最大,能有效提高网络的检测能力。第五行和第六行的对比可以看出,MobileNet轻量化模块的加入导致检测精度小幅度下降,Precision值从53.7%下降为51.1%,Recall值从39.7%下降为37.0%,mAP@0.5从38.4%下降为36.8%,下降值并不明显,但其网络结构所需参数数量却大大降低。考虑到无人机平台的算力低和空间小特性,算法的轻量化处理存在必要性。

2.5 目标检测

为了更加直观地展示DSM-YOLO v5在无人机航拍图像上的目标检测能力,选择了部分实验结果,如图6

表4 消融实验
Table 4 Ablation experiment

160×160	80×80	40×40	20×20	MobileNet	Precision/%	Recall/%	mAP@0.5/%
×	√	√	√	×	44.4	37.8	33.2
√	×	√	√	×	51.5	36.4	34.8
√	√	×	√	×	49.7	38.8	36.3
√	√	√	×	×	52.1	40.8	38.1
√	√	√	√	×	53.7	39.7	38.4
√	√	√	√	√	51.1	37.0	36.8

所示。图6上面一行的三张图像是使用YOLO v5s网络结构的检测结果,下面一行的三张图像则是使用DSM-YOLO v5网络结构检测的结果。显然,对于远处小而密集的物体,DSM-YOLO v5在检测时能够准确地识别,而YOLOv5s的检测效果逊色于DSM-YOLO v5。综合来看,DSM-YOLO v5其检测效果优于YOLO v5s算法。

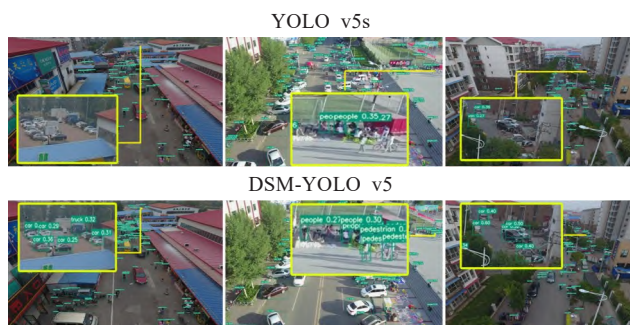


图6 目标检测结果

Fig.6 Target detection results

为了验证DSM-YOLO v5算法的鲁棒性,本文选取了不同环境和光线下的场景作为检测对象。检测结果如图7所示。通过图7第一行的三张图片,可以看到在光线较弱的情况下,DSM-YOLO v5仍然能够有效地检测出大多数有效目标。这表明DSM-YOLO v5算法具有一定的光线适应性,能够适应多种光照条件下的目标检测任务。此外,DSM-YOLO v5算法还能够在各种室内和室外环境中正常运行,这进一步证明了其鲁棒性。第二行的三张图片显示了DSM-YOLO v5在目标密集的情况下的检测效果。这是一个具有挑战性的场景,因为目标过于密集,算法很容易将它们误判为一个单独的目标,或者错过其中的一些目标。然而,DSM-YOLO v5能够有效地检测出每一个目标,并且准确地识别它们的位置和类别,表现出较强的鲁棒性和准确性。



图7 DSM-YOLO v5目标检测结果

Fig.7 DSM-YOLO v5 target detection results

3 结语

根据无人机航拍图像的特点,提出了一种能够在复杂环境、密集场景中进行目标检测的网络结构DSM-YOLO v5。通过加深特征融合部分的网络深度、增加额外的小目标检测头的方式,能够有效地提高无人机航拍图像的检测效率,提高查准率、查全率等各项指标。

将普通的卷积模块替换为深度可分离模块,能明显减少网络所需要的参数,以及模型所占用的空间,使其更有效地应用在无人机硬件设备中。根据在VisDrone2019数据集、DroneVehicle数据集和RSOD数据集上的实验结果,DSM-YOLO v5在无人机航拍图像目标检测任务之中,各项指标都优于目前主流的目标检测网络结构。虽然DSM-YOLO v5算法在无人机航拍图像的目标检测任务当中表现较好,但仍需要进行改进,如网络层数较多,目标检测的精度和速度还有待提升,接下来将继续在网络结构、体积、参数量、检测速度等方面进行更深入的研究。

参考文献:

- [1] YALCIN H, HEBERT M, COLLINS R, et al. A flow-based approach to vehicle detection and background mosaicking in airborne video[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005: 1202.
- [2] MEI X, LING H B. Robust visual tracking and vehicle classification via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(11): 2259-2272.
- [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] 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, 2014: 580-587.
- [5] GIRSHICK R. Fast R-CNN[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440-1448.
- [6] REN S, HE K, GIRSHICK R, et al. Faster RCNN: towards real-time object detection with region proposal networks[C]//Proceedings of Advances in Neural Information Processing Systems, 2015: 91-99.
- [7] DAI J, LI Y, HE K, et al. R-FCN: object detection via region-based fully convolutional networks[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016: 379-387.
- [8] 张瑞倩, 邵振峰, Aleksei Portnov, 等. 多尺度空洞卷积的无人机影像目标检测方法[J]. 武汉大学学报(信息科学版), 2020, 45(6): 895-903.
- [9] ZHANG R Q, SHAO Z F, ALEKSEI P, et al. Multi-scale dilated convolutional neural network for object detection in UAV images[J]. Geomatics and Information Science of Wuhan University, 2020, 45(6): 895-903.
- [9] 陈旭, 彭冬亮, 谷雨. 基于改进YOLOv5s无人机图像实时目标检测[J]. 光电工程, 2022, 49(3): 69-81.

- CHEN X, PENG D L, GU Y. Real-time object detection for UAV images based on improved YOLOv5s[J]. Opto-Electronic Engineering, 2022, 49(3): 69-81.
- [10] 韩玉洁, 曹杰, 刘琨, 等. 基于改进YOLO的无人机对地多目标检测[J]. 电子测量技术, 2020, 43(21): 19-24.
HAN Y J, CAO J, LIU K, et al. UAV ground multi-target detection based on improved YOLO[J]. Electronic Measurement Technology, 2020, 43(21): 19-24.
- [11] 刘鑫, 黄进, 杨涛, 等. 改进CenterNet的无人机小目标捕获检测方法[J]. 计算机工程与应用, 2022, 58(14): 96-104.
LIU X, HUANG J, YANG T, et al. Improved small object detection for UAV acquisition based on CenterNet[J]. Computer Engineering and Applications, 2022, 58(14): 96-104.
- [12] 王浩雪, 曹杰, 邱诚, 等. 基于改进YOLOv4的航拍图像多目标检测方法[J]. 电光与控制, 2022, 29(5): 23-27.
WANG H X, CAO J, QIU C, et al. Multi-target detection method of aerial images based on improved YOLOv4 algorithm[J]. Electronics Optics & Control, 2022, 29(5): 23-27.
- [13] 王胜科, 任鹏飞, 吕昕, 等. 基于中心点和双重注意力机制的无人机高分辨率图像小目标检测算法[J]. 应用科学学报, 2021, 39(4): 650-659.
WANG S K, REN P F, LÜ X, et al. Small target detection algorithm of UAV high resolution image based on center point and dual attention mechanism[J]. Journal of Applied Sciences, 2021, 39(4): 650-659.
- [14] ZHU X, LYU S C, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 2778-2788.
- [15] LIU M J, WANG X H, ZHOU A J, et al. UAV-YOLO: small object detection on unmanned aerial vehicle perspective[J]. Sensors, 2020, 20(8): 2238.
- [16] LIN H, ZHOU J K, GAN Y F, et al. Novel up-scale feature aggregation for object detection in aerial images[J]. Neurocomputing, 2020, 411: 364-374.
- [17] ZHANG X Q, FENG Y, ZHANG S, et al. Finding non-rigid tiny person with densely cropped and local attention object detector networks in low-altitude aerial images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 4371-4385.
- [18] LUO X D, WU Y Q, WANG F Y. Target detection method of UAV aerial imagery based on improved YOLOv5[J]. Remote Sensing, 2022, 14(19): 5063.
- [19] ZHANG T Y, ZHAO Z Q, TIAN D W, et al. Improved YOLOv5 network with attention and context for small object detection[C]//18th International Conference on Intelligent Computing Methodologies, Xi'an, China, August 7-11, 2022. Cham: Springer International Publishing, 2022: 341-352.
- [20] YANG R J, LI W F, SHANG X N, et al. KPE-YOLOv5: an improved small target detection algorithm based on YOLOv5[J]. Electronics, 2023, 12(4): 817.
- [21] 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.
LI Z F, YANG F B, HAO Y Q. Small target detection algorithm for aerial photography based on residual network optimization[J]. Foreign Electronic Measurement Technology, 2022, 41(8): 27-33.
- [22] 苏凯第, 赵巧娥. 基于YOLOv5算法的无人机电力巡检快速图像识别[J]. 电力科学与工程, 2022, 38(4): 43-48.
SU K D, ZHAO Q E. Fast image recognition of UAV power inspection based on YOLOv5 algorithm[J]. Electric Power Science and Engineering, 2022, 38(4): 43-48.
- [23] 许德刚, 王露, 李凡. 深度学习的典型目标检测算法研究综述[J]. 计算机工程与应用, 2021, 57(8): 10-25.
XU D G, WANG L, LI F. Review of typical object detection algorithms for deep learning[J]. Computer Engineering and Applications, 2021, 57(8): 10-25.
- [24] HOWARD A, ZHU M L, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[J]. arXiv: 1704.04861, 2017.
- [25] ZHOU Y, LIU Y Y, HAN G J, et al. Face recognition based on the improved MobileNet[C]//2019 IEEE Symposium Series on Computational Intelligence (SSCI), 2019: 2776-2781.
- [26] DU D W, ZHU P F, WEN L Y, et al. VisDrone-DET2019: the vision meets drone object detection in image challenge results[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019: 213-226.
- [27] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection[J]. arXiv: 2004.10934, 2020.
- [28] 刘展威, 陈慈发, 董方敏. 基于YOLOv5s的航拍小目标检测改进算法研究[J/OL]. 无线电工程: 1-10[2023-02-23]. <https://kns.cnki.net/kcms/detail/13.1097.TN.20230411.1645.026.html>.
LIU Z W, CHEN C F, DONG F M. Improved aerial small object detection algorithm based on YOLOv5s[J/OL]. Radio Engineering: 1-10[2023-02-23]. <https://kns.cnki.net/kcms/detail/13.1097.TN.20230411.1645.026.html>.
- [29] YU W P, YANG T J N, CHEN C. Towards resolving the challenge of long-tail distribution in UAV images for object detection[C]//2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021: 3258-3267.
- [30] ZHOU X, WANG D, KRHENBUHL P. Objects as points[J]. arXiv: 1904.07850, 2019.
- [31] ZHAO H P, ZHOU Y, ZHANG L, et al. Mixed YOLOv3-LITE: a lightweight real-time object detection method[J]. Sensors, 2020, 20(7): 1861-1878.