

基于改进 YOLOv5 的无人机航拍车辆检测算法

相雨涛, 李 博, 万 涛

(中北大学 仪器科学与动态测试教育部重点实验室, 太原 030051)

摘要: 针对目前无人机航拍车辆检测中存在背景环境复杂、目标较小和计算复杂会造成漏检误检等问题, 提出了一种改进 YOLOv5 的无人机车辆检测算法 YOLOv5-R; 采用轻量化模块 GhostNetV2 对主干网络进行更改, 作为模型的主干特征提取网络, 在缓解网络冗余的同时提高模型的检测速度; 在主干网络中引入坐标注意力 CA 模块, 增强模型对目标车辆检测的特征表达能力, 从而提升模型在复杂背景下的检测精度; 颈部引入加权双向特征金字塔网络 BiFPN, 增强模型的多尺度特征表达和融合能力, 提升对小目标的检测精度; 最后将原始的头部分别替换为动态检测头 DyHead, 通过大小、任务和空间感知的三者统一, 进一步提高模型检测性能; 实验结果表明, 与原有的算法相比, 改进 YOLOv5 算法的准确率和平均精确度分别提高了 6.5% 和 5.1%, 且算法检测速度达到 99.7 FPS, 满足检测实时性的要求, 与其他主流模型相比, 该模型在公开数据集上有更好的检测效果, 验证了其可行性和有效性。

关键词: 车辆检测; YOLOv5; 轻量化; 坐标注意力; 检测头

UAV Photography Detection Algorithm Based on Improved YOLOv5

XIANG Yutao, LI Bo, WAN Tao

(Key Laboratory of Instrumental Science and Dynamic Testing, Ministry of Education,
North University of China, Taiyuan 030051, China)

Abstract: To address the issues of complex background environments, small targets, and complex calculations, it leads to missed and false detection in current unmanned aerial vehicles (UAVs), a novel UAV detection algorithm YOLOv5-R is proposed. The lightweight module GhostNetV2 is used to modify the backbone network, as the backbone feature extraction network of the model, it alleviates network redundancy while improving the detection speed of the model; The coordinate attention (CA) module is introduced into the backbone network to enhance the feature representation ability of the model on the target vehicle detection, thereby improving the detection accuracy of the model in complex backgrounds; The weighted bidirectional feature pyramid network (BiFPN) is introduced into the neck enhances the model's multi-scale feature representation and fusion capabilities, improving the detection accuracy of small targets; Finally, the original head is replaced with the dynamic detection head (DyHead), which further improves the detection performance of the model by unifying the size, task, and spatial perception. Experimental results show that compared with the original algorithm, the improved YOLOv5 algorithm improves the accuracy and average accuracy by 6.5% and 5.1%, respectively, and the detection speed of the algorithm reaches up to 99.7 FPS, meeting the requirements of real-time detection. Compared with other mainstream models, the model has better detection performance on public datasets, verifying its feasibility and effectiveness.

Keywords: vehicle detection; YOLOv5; lightweight; coordinate attention; detection head

0 引言

当今世界经济与科技的高速发展, 极大地改善了人们的出行条件, 车辆数量逐渐增多, 车辆检测作为交通

智能管理、车辆信息管理与自动驾驶的一项关键技术, 一直以来受到国内外研究者的广泛关注, 由于车辆外观属性和检测过程中状态的不同, 车辆检测一直是目标检测领域中最基础、最艰巨的任务。随着无人机相关技术

收稿日期:2024-03-20; 修回日期:2024-04-09。

基金项目:国家自然科学基金(61471325);国家自然科学基金青年科学基金(52006114)。

作者简介:相雨涛(1999-),男,硕士研究生。

李 博(1972-),男,博士,副教授。

引用格式:相雨涛,李 博,万 涛. 基于改进 YOLOv5 的无人机航拍车辆检测算法[J]. 计算机测量与控制, 2025, 33(4): 48-56, 66.

的进步,无人机由于其出色的运载能力和灵活的机动性,使其成为了多种目标检测的载体。无人机航拍车辆具有高空优势,利用无人机技术监测和识别地面车辆,对智能交通、城市规划提供更高的管理效率,具有很高的应用价值^[1-3]。

无人机检测车辆算法的核心在于从无人机捕获的图像或视频中识别并定位车辆^[4]。但与标准图像相比,航拍车辆图像呈现出更复杂的背景,物体尺度变化更大,目标较小,并且目标一般比较密集,计算冗余,在网络中传输容易发生特征信息丢失,为无人机航拍车辆的检测带来了挑战^[5]。

如今的目标检测通常分为两类:经典检测和基于深度学习的检测。传统方法包括SIFT、Harr、HOG等,在图像上设置不同比例的滑动窗口,然后提取特征,然后进行分类和定位。文献[6]提出改进HOG特征的支持向量机工程车辆识别方法。文献[7]提出融合SIFT和级联分类器的特种车辆自动检测识别。但传统的车辆检测算法大多依靠人工进行特征提取,难以提取抽象特征,且在背景复杂的情况下,漏检误检现象较多。因此,传统的目标检测算法往往不适合直接应用于无人机航拍车辆检测。与传统检测方法相比,基于深度学习的检测方法实现了端到端的检测和识别,自动提取高层特征,特征表征能力和鲁棒性,显著提高了检测精度和速度。现有的基于深度学习的目标检测方法主要分为两类:一类是基于候选框的两阶段目标检测算法,代表有RCNN^[8]、Fast RCNN^[9]、Faster RCNN^[10]等。这类算法首先提取目标区域,然后对这些区域进行卷积分类识别,以准确性高而闻名,但参数量较大检测速度较慢。文献[11]提出基于Faster R-CNN的轻量化遥感图像军用飞机检测模型,使用残差拆分注意力网络来捕获特征,利用可变形卷积学习特征,用实验对比精简骨干网络,引入Soft NMS算法,实现了较好的综合性能;文献[12]提出基于改进RPN的Faster-RCNN网络SAR图像车辆目标检测方法,基于VGG-16网络和卷积核提取特征,对处理后的特征进行信息融合。另一类是单阶段端到端目标检测算法,以SSD^[13]、YOLO^[14]算法为代表。这些算法直接使用卷积网络进行回归,以更快的检测速度而闻名,但可能会牺牲一些准确性。文献[15]提出一种基于目标空间分布特征的无人机航拍车辆实时检测技术,SSD为基础,引入深度可分离卷积和抗混叠低通滤波器对网络结构进行优化,成对抗CGAN思想构建密度估计网络生成器,车辆密度估计网络结合,决策层融合检测结果,由此实现对航拍视角道路交通场景下车辆目标的精确高效检测。

与两阶段网络相比,YOLO系列方法直接对输入图像应用算法并输出类别和相应的定位,实现了最先进

的检测性能,已广泛应用于各种场合,如行人检测、车辆检测和遥感检测。在复杂的无人机航拍场景中,如何实现对车辆的精准、实时检测和识别是当前亟待解决的问题。文献[16]提出一种基于改进的YOLOv3多目标小尺度车辆检测算法,采用EfficientNet作为特征网络,并采用4个特征尺度融合,提升小尺度车辆的检测效率。文献[17]提出了LDS-YOLO检测模型,针对无人机图像中小目标特征难以提取的问题,该模型以Yolov5为基础,构造了一个新的特征提取模块,在SPP模块中引入SoftPool结构,并采用深度可分离卷积替代传统卷积。文献[18]对YOLOv5s算法引入多尺度分割注意力,以应对小目标背景复杂及特征提取困难等问题,设计了一种自适应加权特征融合,优化各层输出权重,在复杂背景下提高了无人机检测的性能。文献[19]在YOLOv5网络的颈部增加上采样,形成用于收集小目标特征的特征图和新的特征提取层,增强了对小目标的检测效果。文献[20]提出一种改进YOLOv5的轻量化网络,在颈部网络中引入C3Ghost和Ghost模块;引入卷积块注意模块(CBAM, convolutional block attention module),在这些基础上对损失函数CIoU进行改进,提高了算法的检测精度。文献[21]对YOLOv5算法进行改进,设计了在无人机捕获场景中的小目标检测提出了一种特征增强块,将其集成到空间金字塔池,创建包含多尺度信息功能,加权融合特征,提升了对小目标特征的提取能力。

以上研究虽然在速度与精度上都有所提升,但上述作者的研究都局限于对目标较小、背景复杂,计算冗余和精度低等问题无法多者兼顾,并且大部分模型没有一个特别突出的点。当前在无人机航拍车辆检测领域依然面临着难以提取复杂环境下的特征信息、小目标细节在操作过程当中会丢失、检测算法的不同尺度差异导致漏检误检等问题。结合无人机图像特点和YOLO系列单阶段目标检测算法的实时性和准确性,YOLOv5作为一种高效的目标检测算法,适合对无人机目标车辆检测做出针对性改进。考虑到无人机设备有限的硬件性能,本文采用权值与参数量最小的YOLOv5n作为算法模型,因此,本文基于YOLOv5n对模型进行优化,提出了一种基于YOLOv5n的无人机实时车辆检测算法,本文的主要工作如下:

1) 采用轻量化模块GhostNetV2对主干网络进行更改,替换原有的C3模块,在保证原有精度的同时建立轻量化目标检测网络,解决网络参数量大、计算复杂、检测速度慢等问题。

2) 在特征提取模块引入坐标注意力(CA, coordinate attention)增强精准定位能力,减少无关背景特征对车辆检测的干扰,提高模型在复杂环境的准确率。

标检测的核心部分, 经过骨干网络的特征在颈部网络压缩和融合后进行分类和预测。头部通常包括多个输出层, 对应不同分辨率的特征图。高分辨率的特征图可以更好地检测小目标, 而低分辨率的特征图则更适合检测大目标。

2 算法改进

针对现有 YOLO 的无人机航拍车辆检测算法存在计算量大、易受背景环境干扰、小目标检测能力差会造成漏检误检等问题, 对 YOLOV5n 算法进行具体改进, 如图 2 所示。首先使用轻量化网络 GhostNetV2 替换主干网络的 C3 模块, 其次在主干网络中加入 CA 注意力机制, 采用 BiFPN 进行跨层加权特征信息融合, 最后将 YOLOv5n 的头部更换为 DyHead 模块进行预测输出。轻量化网络以更少的计算资源输出更有效的特征信息, 提升网络检测速度, 缓解网络参数量大、计算复杂的问题; CA 注意力机制减少无关背景特征对车辆检测的干扰, 提高模型在复杂环境下的准确率; BiFPN 能快速的进行多尺度特征融合, 充分捕获小目标信息, 提高了对小目标车辆的检测能力; Dyhead 检测头通过结合 3 种注意力机制使得模型在各种场景下的性能进一步提升。

2.1 GhostNetV2 轻量化网络

随着时代的发展, 对于无人机航拍检测车辆来说,

移动设备已成为主要载体。然而, 大规模深度学习网络模型参数多, 计算要求高, 难以移植到移动终端。轻量级网络模型的出现很好地解决了这个问题, 与大规模深度网络模型相比, 轻量级网络模型的参数和计算量非常少, 模型性能不逊色于大规模深度网络模型, 能够适应移动终端、嵌入式设备等硬件资源有限的场景, 针对原有 YOLOv5 网络参数量大、计算复杂、检测速度慢等问题, 将 GhostNetV2^[26] 骨干网模块集成到原有网络骨干网中, 降低网络参数和计算量, 从而降低硬件要求轻量级 GhostNetV2 由两个 Ghost 模块和一个 DFC (DFC, decoupled fully connected) 模块组成。具有更多通道的特征图的生成由第一个 Ghost 模块承担。随后, 特征图的通道维度被第二个 Ghost 模块有效压缩, 从而减少了冗余。此外, 该文还采用解耦全连接 DFC 注意力机制来捕获水平和垂直方向的长距离像素依赖关系。通过这些组合组件的协同操作, 网络提取有意义特征的能力得到增强。模块如图 3 所示。

Ghost 模块将普通卷积分解成两个步骤, 首先是传统卷积计算, 然后进行一系列简单的线性变换, 生成原有特征图的 Ghost 特征图, 最后将其拼接到一起生成最终的特征图, 以此消除特征图冗余, 当步长为 1 时, Ghost 模型和 DFC 保持分支并行输入, 以进一步将特征传输到下一个 Ghost 模块进行输出。第一个 Ghost 和 DFC 并行模块主要捕获不同空间位置像素的远程依赖

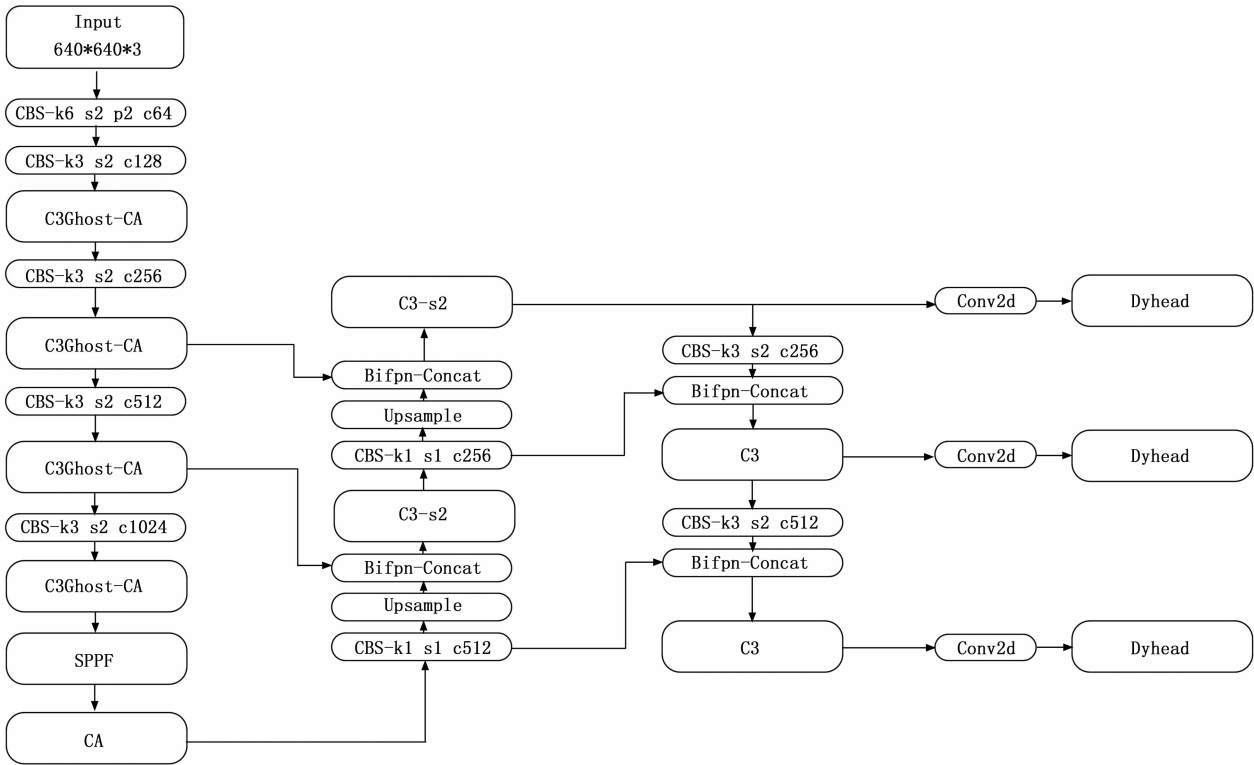


图 2 YOLOv5-R 网络结构

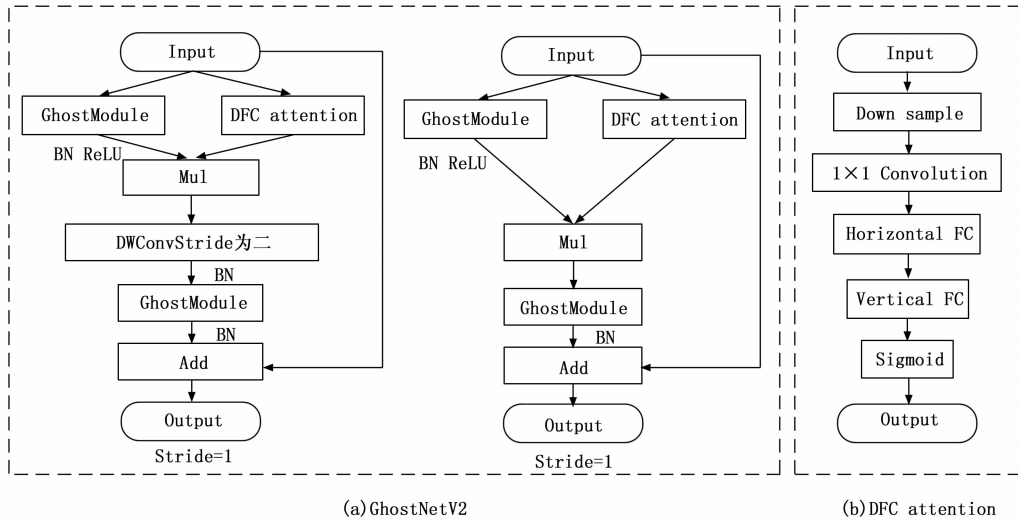


图 3 GhostNetV2 结构

性, 并增加输入特征图通道的数量。第二个 Ghost 模块用于减少输入特征图中的通道数, 并匹配原始输入网络中的通道数。步长 1 和步长 2 的区别在于使用了 DW-Conv, DWConv 主要用于压缩前后两个 Ghost 模块的宽度和高度, 并改变输入要素图层的形状, 如图 4 所示。相比于普通卷积, Ghost 模块的压缩比为:

$$\frac{h' \times w' \times n \times k \times k \times c}{h' \times w' \times m \times k \times k \times c + h' \times w' \times \frac{n}{s} \times d \times d \times (s-1)} \approx \frac{s \times c}{s + c - 1} \approx s \quad (1)$$

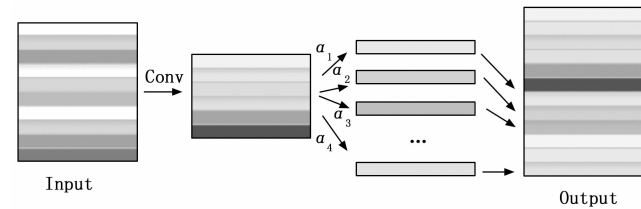


图 4 Ghost 模块示意图

Ghost 模块替换普通卷积后, 网络模型中卷积层参数数量和计算量降低至原来的 $1/s$, 在主干网络中使用 Ghost 模块替换普通卷积能大大减少模型的复杂度。

由于模型空间信息捕获能力较弱, 采用动态特征校正 DFC 注意力机制来捕获模型的远端空间信息。该机制将全连接层分解为水平层和垂直全连接层, 以聚合二维特征图中的像素。在 DFC 的注意力机制下, Ghost 模块可以聚合本地信息, 建立远程依赖。由于 Ghost 模块和 DFC 注意力机制之间的直接并行性会降低 Ghost 模块的效率, 因此在水平和垂直方向上对 DFC 模块中的输入特征图进行下采样可以减小特征的大小, 并允许它们对较小的特征进行操作。

2.2 CA 注意力机制

注意力机制则起源于人类对信息选择, 一般来说, 注意机制的目的是为确定一个特定地点的聚焦位置或者显著特征的表达而设计的。注意力可以使资源在卷积运算过程中得到有效利用, 从而为神经网络强调更有价值的信息。对于无人机视角下检测车辆来说, 通常情况下具有背景环境复杂、所检测车辆目标过小的问题。为了增加特征图的信息量、提高信息准确度、增强网络在复杂背景下的检测能力, 因此在改进 YOLOv5 主干网络提取出来的特征图上增加 CA 注意力机制^[27], 如图 5 所示。

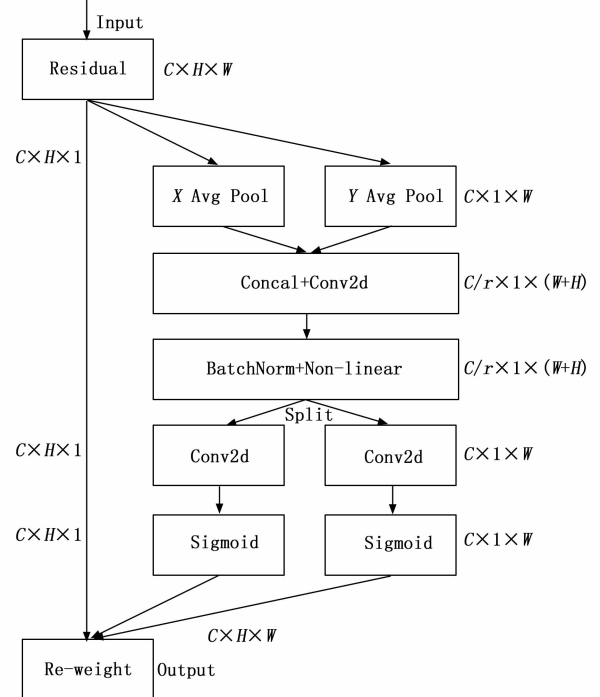


图 5 坐标注意力模块结构图

本文CA注意力机制与其他注意力机制相比,坐标注意力机制作为轻量级混合注意力机制,既关注了通道信息,又兼顾了空间信息,解决了以往注意力机制只能关注通道信息、无法产生长期依赖关系的问题,提高了神经网络的性能。CA注意力机制在通道和空间两个方面都有所关注,利用Concat融合了高度与宽度的空间信息,并采用卷积、Split等技术进一步实现对空间信息的提取。CA模型采用两步来获取信道关系以及长时依赖关系,即坐标信息的嵌入和坐标注意力的产生。首先,基于全局平均池化方法,以获取精确的位置信息;避免将空间信息压缩到信息通道中。CA注意力通常将通道注意力分解为2个一维特征,分别用 $H \times 1$ 和 $1 \times W$ 两个池化核分别对输入特征 X 水平和垂直方向的每个通道进行池化,生成尺寸为 $C \times H \times 1$ 和 $C \times 1 \times W$ 的特征图,即:

$$Z_c^h(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (2)$$

$$Z_c^h(h) = \frac{1}{W} \sum_{0 \leq j < W} x_c(h, i) \quad (3)$$

接着将生成的特征图在第三维度中两个带有一定方位信息的特征图进行Concat操作。然后使用共享的 1×1 卷积变换函数 F_1 进行变换,得到其空间信息特征图,即:

$$f = \delta\{F_1[(z^h, z^w)]\} \quad (4)$$

将 f 在空间维度上拆分为 f^h 与 f^w ,尺寸分别为: $C/r \times 1 \times H$ 和 $C/r \times 1 \times W$,经 1×1 卷积和Sigmoid激活函数得到两个具有长程依赖的注意力权重图,即:

$$g^h = \sigma[F_h(f^h)] \quad (5)$$

$$g^w = \sigma[F_w(f^w)] \quad (6)$$

最终将特征图与特征权值相结合,从而加强了特征表征能力,得到CA的表达式:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (7)$$

2.3 加权双向特征金字塔 BiFPN

无人机拍摄视角高,对于检测微小的目标,它的像素数量很少,检测性能会下降。FPN通过简单的自上而下和水平特征融合方法,从原始特征图进一步重构特征金字塔,但其传统的自上而下结构受到单向信息流限制的影响。PANet基于FPN进行改进,不仅自上而下进行特征融合,在此基础上,还加入了一条自底向上的特征融合通道,该通道既包含了上层的语义信息,也包含了下层的位置信息,可以提取更多的低级特征与高级特征。但在这些特征融合方法中,由于输入特性的分辨率不同,平等地对待每个输入特征图,忽视了不同分辨率输入特征对特征融合贡献的不平等性,对最终结果可能产生更大影响。因此,在颈部网络引入了一种简单、高效的BiFPN^[28]网络结构,以提

高检测精度,如图6所示。

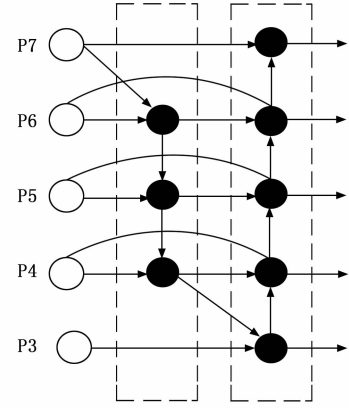


图6 BiFPN结构

BiFPN可以理解为具有权重的双向特征融合金字塔网络。不仅可以实现双向特征信息流,还可以增加交叉连接。首先,它引入了双向特征传播,包括自上而下和自下而上的特征流,促进了更彻底、更细致的跨层次信息传递和融合,丰富了模型在不同层次上捕捉不同特征的能力。其次,通过特征调整 and 选择操作对合并结果进行优化;该过程确保了关键特征的表示,并动态选择最有价值的特征,有助于提高目标检测的准确性和有效性。最后,通过精心设计的特征融合和调优操作,BiFPN增强了多尺度特征的融合,丰富了特征的语义信息,使多尺度特征的融合更加灵活,提高了对小目标的检测精度,而且成功地降低了计算需求。

2.4 DyHead模块

现实世界的图像中存在多个物体,每个物体都具有不同的比例和大小,这加剧了这种困境。此外,当从不同的角度观察时,这些物体可能会表现出明显不同的形状和位置。为了克服这一挑战,物体检测的头部片段必须具有一定的空间感知能力,本研究使用的无人机检测车辆数据集呈现出复杂的背景,包括不同场景和不同尺度的小目标车辆。

因此,检测算法必须具备全面的感知能力。此外,从模型颈部特征金字塔中提取的特征元素图中的目标具有不同的空间位置和形状,需要检测算法有效地捕获空间信息。为了应对这些挑战,我们引入了DyHead^[29]检测模块,该模块同时提供大小、任务和空间感知,每个注意力集中在一个特定的方向上,其维度为水平 $\times$ 空间 $\times$ 通道。尺度感知注意模块将不同尺度的特征根据其语义重要性进行融合,空间感知注意模块首先使用可变形卷积学习稀疏性,然后在同一空间位置聚合跨层特征,任务感知注意模块动态打开或关闭特征通道,选择不同的任务,它使输入数据具有适应性。增强了检测能力,

同时努力优化计算效率。

这种分解有助于处理不同层次特征与目标尺度之间的关系,改善不同层次特征的表示学习,在检测层上的三维特征张量为 $F \in R^{L \times S \times C}$,该注意力函数计算公式如下所示:

$$W(F) = \pi_C \{ \pi_S [\pi_L (F) \cdot F] \cdot F \} \quad (8)$$

$\pi_L(\cdot)$ 、 $\pi_S(\cdot)$ 、 $\pi_C(\cdot)$ 是 3 种不同的注意力模块,并分别应用在 L 、 S 、 C 上的 3 个不同维度上拓宽感受野,图 7 展示了 3 种注意力模块。这 3 个模块应用于检测头部,可以多次叠加使用,在本模型的设计过程中,经过实验对比,使用一个 dyhead 模块让检测头具备更好的性能。

3 实验及结果分析

3.1 数据集

本文使用的数据集是在 Roboflow 平台上发布的公开汽车数据集。该数据集将被检测车辆分为小型货车、野营车、汽车、摩托车、小型卡车、大型卡车、拖拉机、大型货车与其它车等 9 个类别,共有 2 007 张图像。按照 7:2:1 的比例将车辆图片划分为训练集、验证集和测试集。其中 1 407 张图像作为训练集,400 张图像作为验证集,200 张图像作为测试集。按照 7:2:1 的比例将车辆图片划分为训练集、验证集和测试集。其中每张图片包括了多个训练对象,且排列较为密集,更符合无人机航拍视角下的密集小目标车辆检测情况。

3.2 实验设置及评价标准

本模型所有实验环境都是在基于 ubuntu20.04 系统下进行的,硬件配置为:CPU: 14vCPU, Intel (R) Xeon (R) Gold 6330 CPU, GPU: RTX 2080Ti, 24 GB GPU, 编译环境为 Python 3.8.10 + Pytorch 1.10 + Cuda11.3。在实验时,将初始学习率设置为 0.01,批量大小设置为 32,权重的衰减率设置为 0.000 5,热身轮数设置为 3,色相的增强系数设置为 0.015,饱和度的增强系数为 0.7,亮度的增强系数为 0.4,训练周期设置为 200。

为了证明改进模型的优越性,并与原模型进行比较,本研究主要从准确率 (P , Precision)、召回率 (R , Recall)、平均精密度 (mAP , mean average precision)、每秒帧数 (FPS , frames per second)、参数个数、文件权重大小等方面对改进模型进行评价。 P 、 R 、 mAP 的计算公式如下所示。

准确率,准确率主要检查预测结果是否正确,即样本的所有正例中找对的正例所占百分比,计算公式如下所示:

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

模型的召回率,检查预测结果是否全面,即样本所有本应该被找对的正例中通过型计算后所找到的正例占比:

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

其中: TP 表示检测结果中被成功预测为正确的车辆数量, FP 代表结果中被错误判断为正确车辆的数量, FN 表示样本中所有正确车辆目标中未被检测到的数量。

mAP 是所有检测类别的精度的平均值,是对 m 个类别的准确率 (P) 与召回率 (R) 综合考虑,较全面的反映网络的性能计算公式如下:

$$mAP = \frac{1}{m} \sum_{i=1}^m \int_0^1 P(R) dR \quad (11)$$

每秒帧数反映了模型的推理速度, FPS 越大,检测速度越快。Giga 浮点运算每秒 (GFLOPs) 与权重文件 (MB) 是度量模型复杂度的指标。GFLOPs 与权重文件 (MB) 越小,表示模型所需要的计算能力就越小。

3.3 串联不同数量 DyHead 模块的性能对比

DyHead 模块串联的数量决定着模型检测的性能。通过控制不同数量 DyHead 块的堆叠,我们评估了其对模型性能和计算成本的影响。由表 1 看出,经过实验对比并不是串联的 Dyhead 模块越多性能就会越好,反而随着串联模块的增加,模型的性能却逐渐降低。因此,本文只串联 1 块 DyHead 模块进行预测输出。

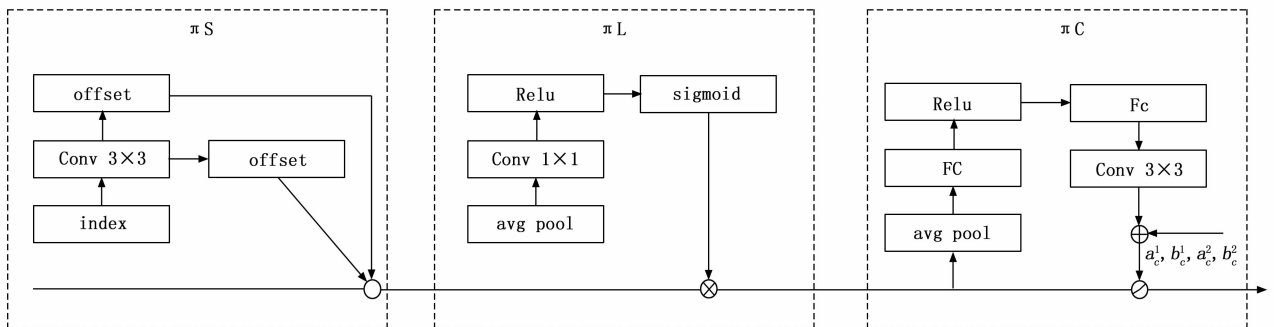


图 7 DyHead 模块结构图

表 1 不同数量 DyHead 模块性能对比

DyHead 数量	FPS	权重文件/MB	P/%	mAP50/%
1	107.5	6.3	72.9	48.6
2	85.3	8.0	73.4	47.8
3	70.2	8.6	71.9	47.3
4	61.6	9.3	66.8	47.2

3.4 不同注意力机制对比

为了进一步检验 CA 模块在模型中的作用，本实验将原始模型与不同的注意力机制结合进行对比，实验结果如表 2 所示。相较于其他两种注意力机制 SE (Squeeze and Excitation)、CBAM，CA 作为轻量型的注意力机制，在模型的参数量和计算量几乎没有增长的情况下，使模型的性能提升最高。

表 2 不同注意力机制对比

注意力机制	FPS	权重文件/MB	P/%	mAP50/%
CBAM	112.3	6.2	68.6	41.3
SE	125.0	5.4	67.2	39.2
CA	125.2	5.6	78.8	48.0

3.5 不同特征融合结构对比

针对不同特征融合结构产生不同的结果。为了验证本文使用不同融合方式的性能优异性，使用不同结构进行实验，从表 3 中可以看出，BiFPN 相比于其他 3 种方式自适应空间特征融合 (ASFF, adaptively spatial feature fusion)，渐近特征金字塔网络 (AFPN, asymptotic feature pyramid network) 有着更高的检测精度，且模型所需的计算能力满足要求。

表 3 颈部网络不同特征融合结构对比

融合结构	FPS	权重文件/MB	P/%	mAP50/%
AFPN	130.3	5.6	71.7	43.9
ASFF	135.6	6.4	68.9	41.2
BiFPN	128.7	5.9	76.2	46.6

3.6 消融实验

为了更好地检验各个模块和各个模块间的结合对原有模型性能的提高效果。本文设计了消融实验。在原有模型的基础上，逐条增加改进措施，并进行了实验。

实验结果如表 4 所示。其中 ✓ 表示有加入该模块，× 则表示没有加入该模块。通过表 4 的实验可以发现，每一步的 YOLOv5 优化都会对 YOLOv5 模型产生影响。其中，轻量化的 GhostNetV2 设计使模型的 mAP@0.5 提高了 3.4%，也使模型 FPS 提高了 16%。轻量化后精度也没有下降，模型大小也减小了 0.9 MB。使用 CA 注意力机制的网络在几乎不减小 FPS 和模型权重文件基本不变的情况下，mAP50 提高了 3.8%，这说

明 CA 注意力机制在计算复杂度和识别精度上取得了较好的平衡。改进加权特征融合网络 BiFPN 后，模型的 mAP50 提高了 2.4%，计算量略有增加，对于该模型来说是可以接受的。DyHead 模块由于 3 种注意力的统一，在能保证实时检测的情况下，mAP50 相较于原模型提升了 4.4%，且模型大小和计算速度都没有太大影响。组合使用多个模块 (GhostNetV2、CA、Bifpn、DyHead) 可以进一步提升准确率，但会影响 FPS 和模型大小。最终，将不同的模块组合在一起，结合各个模块的优势，该组合模型具有较高的准确率、平均准确率且模型大小减小，当 FPS 达到 99.7.6 的时候，仍然可以满足实时检测的需要。

表 4 消融实验

GhostNetV2	CA	Bifpn	DyHead	FPS	权重文件/MB	P/%	mAP50/%
×	×	×	×	138.3	5.3	72.4	44.2
✓	×	×	×	161.7	4.4	78.5	47.6
×	✓	×	×	125.2	5.6	78.8	48.0
×	×	✓	×	128.7	5.9	76.2	46.6
×	×	×	✓	107.5	6.3	72.9	48.6
✓	✓	×	×	130.6	4.6	78.1	48.1
✓	✓	✓	×	126.3	4.7	71.7	48.3
✓	✓	✓	✓	99.7	5.2	78.9	49.3

3.7 实际效果验证

原始的 YOLOv5n 算法在处理不同大小、复杂背景以及较小的图像时没有达到很好的效果。在经过改进后，针对上述情况，对 YOLOv5-R 算法的性能进行了研究。选择有针对性的照片做测试，测试结果如图 8 所示。



图 8 YOLOV5-R 检测图

3.8 不同算法之间的对比实验

为了检验改进网络对无人机航拍检测车辆的有效性和先进性，并从对无人机检测任务实时性的需求出发，

将该算法与目前的几种主流的算法 Efficientnet、RT-DETR、YOLOv3n、YOLOv4n、YOLOv5n、YOLOv6n、YOLOv7n、YOLOv8n 进行对比实验。结果如表 5 所示。与 Efficientnet 和 RT-DETR 算法相比较, YOLOv5-R 准确率分别提升了 12.6% 和 8.7%, $mAP50$ 提升了 12.2% 和 9.1%。且每秒帧数 FPS 更高, 模型权重文件更小, 实验结果表明相较于这两种的优秀模型 YOLOv5-R 具有更好的检测性能。而与 YOLOv3n、YOLOv5n、YOLOv6n、YOLOv7n 和 YOLOv8n 等其他轻量化 YOLO 系列算法比较, YOLOv5-R 算法在仍然满足实时性检测和模型权重文件最小的情况下, 准确率分别提升了 12.5%、6.5%、10.4%、5.1% 和 6.7%, $mAP50$ 分别提升了 9.4%、5.1%、10.8%、5.2% 和 4.1%, 涨幅明显。实验结果表明, 与其它单阶段目标检测算法比较, YOLOv5-R 算法以较低的模型体积实现了极高的检测精度, 且仍然满足实时性的检测需求, 具有高效性与优异性。

表 5 不同算法对比

模型	FPS	权重文件/MB	$P/\%$	$mAP50/\%$
Efficientnet-b0	88.2	8.9	66.3	37.1
RT-DETR-L	49.2	42.6	70.2	40.2
YOLOv3n	122.2	7.9	67.4	39.9
YOLOv5n	138.3	5.3	72.4	44.2
YOLOv6n	133.6	8.3	68.5	38.5
YOLOv7n	120.8	7.6	73.8	44.1
YOLOv8n	116.2	6.6	72.2	45.2
YOLOv5-R	99.7	5.2	78.9	49.3

4 结束语

针对在无人机检测车辆容易出现误检、漏检等问题, 本文提出了一种改进 YOLOv5n 的无人机航拍车辆检测算法, 通过引入 GhostNetV2 简化主干网络, 以更少的计算资源输出更有效的特征信息, 提升网络检测速度, 缓解网络冗余的问题。在主干网络中引入坐标注意力 CA 模块, 对骨干网络的输出特性进行优化, 降低特征信息的丢失, 提升模型在复杂背景下的检测精度。在颈部网络中将特征金字塔结构替换为 BiFPN 模块, 增强网络的多尺度特征表达和融合能力, 充分捕获小目标信息, 提高了对小目标车辆的检测能力。在头部网络, 使用集成了尺度、空间和任务 3 种注意力函数的 Dy-Head 检测头进一步提升了模型的性能。通过实验表明, 改进后的算法与现有模型对比, 计算精度高于原算法, 对于在复杂环境下和小目标车辆检测都有更好的检测效果, 且满足实时检测要求。考虑到目前发展趋势, 未来的研究重点应放在轻量级设计, 在各项性能不降低的情况下, 使计算速度和精度之间平衡, 对模型进行压缩,

以满足部署无人机等嵌入式设备的需求, 实现轻量化模型搭载无人机实时对车辆的检测任务。

参考文献:

[1] 茅智慧, 朱佳利, 吴鑫, 等. 基于 YOLO 的自动驾驶目标检测研究综述 [J]. 计算机工程与应用, 2022, 58 (15): 68-77.

[2] 高港耀. 基于深度学习的无人机视角小目标检测算法研究 [D]. 成都: 电子科技大学, 2023.

[3] 曹家乐, 李亚利, 孙汉卿, 等. 基于深度学习的视觉目标检测技术综述 [J]. 中国图象图形学报, 2022, 27 (6): 1697-1722.

[4] LIU C, YANG D, TANG L, et al. A lightweight object detector based on spatial-coordinate self-attention for UAV aerial images [J]. Remote Sensing, 2022, 15 (1): 83.

[5] LIU W, WANG M, ZHANG S, et al. Research on vehicle target detection technology based on UAV aerial images [C] //2022 IEEE International Conference on Mechatronics and Automation (ICMA), IEEE, 2022: 412-416.

[6] 吴逸, 吴静静. 改进 HOG 特征的支持向量机工程车辆识别方法 [J]. 计算机与数字工程, 2023, 51 (10): 2405-2412.

[7] 唐海涛, 吴果林, 范广义, 等. 融合 SIFT 和级联分类器的特种车辆自动检测识别 [J]. 计算机技术与发展, 2023, 33 (9): 182-189.

[8] 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.

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

[10] REN S, HE K, GIRSHICK R, et al. Faster r-CNN: Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015: 91-99.

[11] 党玉龙, 叶成绪. 基于 Faster R-CNN 的轻量化遥感图像军用飞机检测模型 [J]. 激光杂志, 2024, 45 (7): 111-117.

[12] 曹磊, 王强, 史润佳, 等. 基于改进 RPN 的 Faster-RCNN 网络 SAR 图像车辆目标检测方法 [J]. 东南大学学报 (自然科学版), 2021, 51 (1): 87-91.

[13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] //Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, Part I 14. Springer International Publishing, 2016: 21-37.

(下转第 66 页)