

面向无人机场景的深度学习模型压缩与加速设计方法

张佳奇¹, 郭 宣¹, 李巍然², 王胜科¹

(1. 中国海洋大学信息科学与技术学院, 山东 青岛 266100; 2. 青岛第九中学, 山东 青岛, 266000)

摘要: 随着无人机应用日益广泛, 无人机应用场景下的目标检测技术已在诸多领域有着重要的应用价值和迫切的应用需求。在无人机边缘设备实时检测的场景中, 由于无人机高分辨率图像存在大量的微弱目标实例, 将直接降低图像分辨率, 从而造成微弱目标丢失, 因此输入网络时保持高分辨率信息至关重要。为此, 将高分辨率图像裁成多个图像块, 将其作为网络输入, 以保持网络精度的同时使其能在边缘智能设备上运行。同时, 为了加快模型推理速度, 使用分组卷积与通道混洗策略对检测算法的主干网络进行轻量化设计, 并使用通道注意力机制提升网络精度。实验表明, 所提方法相较于YOLOv5, 在无人机数据集VisDrone与自建数据集OUC-UAV-DET上的精度分别提升2%、4%, 在英伟达硬件(Xavier)上推理速度减少1 ms。在网络剪枝层面, 结合具体数据集对检测模型进行通道剪枝, 能在维持算法精度的前提下将推理速度减少2 ms。此外, 对于无人机单类别检测任务, 针对其目标实例尺寸相对固定的特点优化输出部分, 能减少30%的模型参数量, 使推理速度最多减少2 ms。

关键词: 无人机; 高分辨率图像; 目标检测; 轻量化网络; 网络剪枝; 模型部署

DOI: 10.11907/rjtk.231712

开放科学(资源服务)标识码(OSID):

中图分类号: TP183

文献标识码: A

文章编号: 1672-7800(2023)012-0056-07



Compression and Acceleration Design Methods of Deep Network Models for UAV Scenes

ZHANG Jiaqi¹, GUO Xuan¹, LI Weiran², WANG Shengke¹

(1. School of Information Science and Technology, Ocean University of China, Qingdao 266100, China;

2. Qingdao No. 9 Middle School, Qingdao 266000, China)

Abstract: With the increasingly widespread application of drones, target detection technology in drone application scenarios has important application value and urgent application needs in many fields. In the scene of real-time detection of drone edge devices, due to the presence of a large number of weak target instances in the high-resolution images of drones, the image resolution will be directly reduced, resulting in the loss of weak targets. Therefore, it is crucial to maintain high-resolution information when inputting into the network. To achieve this, high-resolution images are cut into multiple image blocks and used as network inputs to maintain network accuracy while enabling them to run on edge intelligent devices. At the same time, in order to accelerate the model inference speed, group convolution and channel shuffling strategies are used to lightweight design the backbone network of the detection algorithm, and channel attention mechanism is used to improve network accuracy. The experiment shows that compared to YOLOv5, the proposed method improves accuracy by 2% and 4% on the unmanned aerial vehicle dataset VisDrone and the self built dataset OUC-UAV-DET, respectively, and reduces inference speed by 1 ms on the Nvidia hardware (Xavier). At the network pruning level, combining specific datasets for channel pruning of detection models can reduce inference speed by 2 ms while maintaining algorithm accuracy. In addition, for the single category detection task of unmanned aerial vehicles, optimizing the output

收稿日期: 2023-07-04

基金项目: 国家重点研发计划项目(2018AAA0100400); 山东省自然科学基金项目(ZR2020MF131, ZR2021ZD19); 青岛市科技计划项目(21-1-4-ny-19-nsh); 中国海洋大学社团培训项目(202265007); HY项目(LZY2022033004)

作者简介: 张佳奇(1999-), 男, 中国海洋大学信息科学与技术学院硕士研究生, 研究方向为深度学习、计算机视觉; 郭宣(1997-), 男, 中国海洋大学信息科学与技术学院硕士研究生, 研究方向为深度学习、计算机视觉; 李巍然(2007-), 女, 青岛第九中学学生; 王胜科(1978-), 男, 博士, 中国海洋大学信息科学与技术学院副教授, 研究方向为深度学习、计算机视觉。本文通讯作者: 张佳奇。

part based on the relatively fixed size of the target instance can reduce the number of model parameters by 30%, resulting in a maximum reduction of 2 ms in inference speed.

Key Words: unmanned aerial vehicle; high resolution images; object detection; lightweight network; network pruning; model deployment

0 引言

近年来,随着无人机技术快速发展,无人机低空遥感成为了卫星遥感数据的重要补充,被广泛应用于农业、安防、交通和战场等领域。相较于卫星遥感图像,无人机图像独特的优势主要体现在便携性好、成本较低及能实现多角度拍摄等方面。

目前,无人机目标检测任务通常通过离线处理无人机拍摄的图像或将图像传输至服务器进行处理。然而,第一种方式并未充分利用无人机实时检测分析的优势;第二种方式则依赖于网络,既需要考虑数据传输过程中的隐私保护问题,还需要承担较高成本。此外,由于无人机图像具有高分辨率特点,在实际应用场景中对自然场景的检测算法效果不佳,无法直接应用于工业应用中。因此,改进现有目标检测算法,在无人机上搭载的边缘智能设备,实现在线分析高分辨率图像这一功能是目前一大挑战,也具有重要的实际应用价值。

本文主要对无人机高分辨率图像下的目标检测算法与模型压缩方法进行研究,结合具体硬件平台进行模型部署实验。主要工作如下:①通过分组卷积和通道混洗策略对目标检测算法的主干网络进行轻量化设计,结合通道注意力机制提升网络精度;②在网络剪枝层面,结合具体数据集对目标检测模型进行通道剪枝;③针对无人机单类别检测任务中目标实例尺寸相对固定的特点,对输出部分进行优化,最多可减少 30% 的参数量。

通过实验与在实际硬件上的运行情况可知,本文方法相较于 YOLOv5 算法,在无人机数据集 VisDrone 和自建数据集 OUC-UAV-DET 上的精度分别提升 2%、4%,在英伟达硬件(Xavier)上推理速度约减 1 ms。通过通道剪枝方法,能在保持算法精度的同时减少 2 ms 的推理速度;对无人机单类别检测任务中输出部分进行相关优化后,推理时间最多能减少 2 ms。

1 相关工作

通过对目标检测算法的发展历史进行研究后,将其划分为传统算法阶段和深度学习阶段。2012年前,传统目标检测阶段主要采用手工提取特征的方法。通过滑窗遍历技术,从整张图像中提取有价值的特征,并利用支持向量机算法^[1]或 AdaBoost 分类方法^[2]分类提取特征。此外,通过对整个图像进行特征提取,能将目标检测看作为回归问题,从而实现快速预测。这两种方法可看作为两个阶段目标检测算法和单阶段目标检测算法的雏形。

早期,将卷积神经网络用到目标检测领域的是 OverFeat^[3],在此之前卷积神经网络已在图像分类任务中取得了一定的成果。随后,目标检测主要分为以 R-CNN 系列^[4-6]为代表的两阶段目标检测算法和以 YOLO 系列为代表的单阶段目标检测算法。其中,R-CNN 系列算法通过划分区域来检测对象,并非同时检测整张图像。这种方式具有两个特点:算法需要多次计算才能提取所有对象;在算法执行过程中不同系统接替工作。因此,R-CNN 系列算法属于两阶段算法。

单阶段目标检测算法(You Only Look Once, YOLO)的诞生是目标检测领域的一次突破,即仅用单个神经网络预测图像的类别概率和边界框坐标^[7]。由于算法自身只有一个神经网络可实现端到端优化,这种端到端的统一检测设计使得 YOLO 架构在推理过程中训练更快,实现实时训练,也保证了接近两阶段目标检测算法的高平均精度。目前,最具代表性的 YOLO 框架 YOLOv5 已可满足一般使用场景,但在无人机高分辨率图像的场景下,边缘设备上的推理速度无法满足实际应用需求,其网络结构仍有压缩和优化空间。

2016 年来,由于实际应用需要,轻量化的卷积神经网络研究开始逐渐浮现,为视觉深度学习模型在边缘智能设备上的应用提供了可能性。其中,轻量级研究较为经典的工作包括基于深度可分离卷积的 MobileNet 系列^[8-10]。在 ShuffleNetv2^[11]中,从内存访问代价和显卡并行性方向分析了如何设计卷积神经网络,才能进一步减少运行时间,并提出轻量化网络设计的 4 条准则及 CSPNet^[12]、VoVNet^[13]、RepVGG^[14]、GhostNet^[15]等轻量化网络结构。此外,模型压缩方式还包含神经网络剪枝、知识蒸馏等从大模型中产生小模型的方法。

综上,本文使用分组卷积与通道混洗策略对检测算法 YOLOv5 的主干网络进行轻量化设计,既使用通道注意力机制提升网络精度,又在网络剪枝层面结合具体数据集对检测模型进行通道剪枝。此外,对于无人机单类别检测任务,针对其目标实例尺寸相对固定的特点优化了模型的输出部分。

2 轻量化目标检测算法

2.1 基于统计的语义相似度算法

本文对 YOLOv5 的主干网络进行改进,主要基于分组卷积通道混洗策略和深度可分离卷积。图 1 展示了网络的整体结构,该网络去除了 Fcous 结构,使用 S1、S2 模块替代 YOLOv5 中的 C3 模块,在颈部采用更简单的直接相加方式进行特征融合。

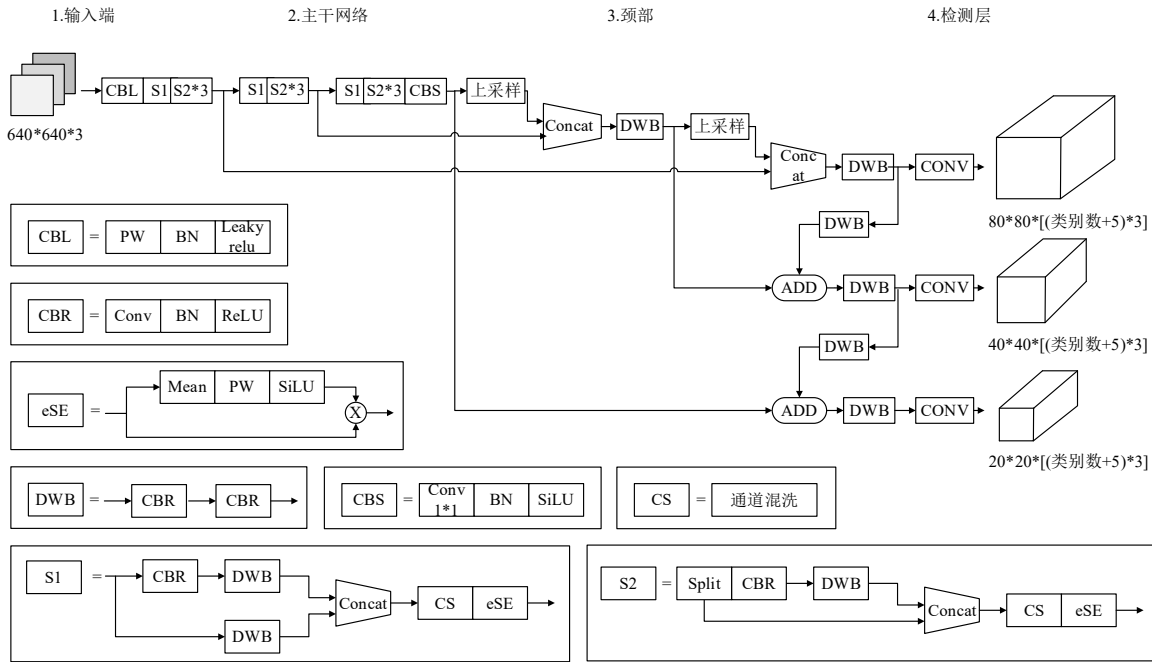


Fig. 1 Overall structure of the network

图1 网络整体结构

YOLOv5在输入层后引入一种名为Focus的结构,如图2所示。该结构对特征图进行间隔取取得到4个图像,并将它们连接起来以使通道数量增加4倍;然后通过卷积操作生成新的特征图,以便在下采样后保留所有空间信息。实际上,在浅层网络中该结构的效果并不显著,主要从减少计算量和参数量角度所设计。

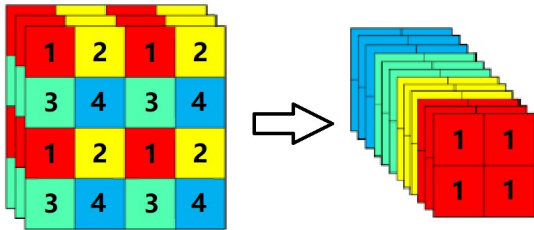


Fig. 2 Focus module

图2 Focus模块

为此,本文使用3x3的卷积网络替代Focus结构,尽管Focus结构的设计初衷是为了减少计算量和参数量,但实验证明使用卷积替代Focus可获得更好的效果。特别是,对于不含GPU和NPU加速的芯片,频繁的切片操作会增加缓存占用、加重计算负担,导致在部署时,许多推理框架无法高效适配Focus的高性能算子。

本文使用3x3卷积网络在改造目标检测网络时,充分考虑了网络轻量化设计的4个准则^[11]:①使用相同的通道尺寸可最小化内存访问量;②避免过度使用分组卷积;③避免网络过于碎片化;④不能忽略元素间的操作(例如shortcut, Add)。

YOLOv5的主干网络主要由C3层组成,该层是CSP结构改进后的版本,在与CSP近乎相似的损耗上能取得更好

的结果。然而,YOLOv5因多次使用高通道数通道的C3层,将导致中间层通道之间的阶跃差距较大。根据准则①可知,高通道数的C3层对CPU性能并不友好。为了解决这个问题,本文将整个特征图按通道分为 c' 、 $c - c'$ 两个部分,主要有以下优点:将整个特征图分为两组,这种分组方式与分组卷积不同,不会增加卷积时的组数,符合准则②;经过特征图分组后,将 c' 组通过短连接通道,而 $c - c'$ 组经过一系列卷积层后,使输入输出通道数保持一致,符合准则①。

在特征融合部分,本文依然采用特征金字塔网络和路径聚合结构。为了满足准则①中的最优化内存访问和使用,在路径聚合结构采用相同的通道数进行特征图融合,使用特征图直接相加的方式实现特征图融合。

2.2 高效的通道注意力模块

注意力机制在深度神经网络中可被看作是一个对象选择过程。SENet^[16]开创了通道注意力的先河,核心是一个用来收集全局信息的压缩和激励模块。

首先,全局空间信息通过全局平均池化层汇集到压缩模块中,激励模块使用全连接层和非线性激活函数获得注意力向量;然后,输入特征的每个通道通过乘以注意力向量进行加权。此外,SE模块能起到强调重要通道并抑制噪音的作用,缺点是在压缩模块中,全局平均池化无法很好的提取复杂的全局信息。同时,在激励模块中,两个全连接层增加了模型复杂性。

对于本文而言,通道注意力机制的重要作用是降低模型的复杂性。为了在被分割后的混洗通道中恢复被打乱的通道信息,促进通道间信息的交流,提升模型检测性能。本文受SENet启发,在S1与S2模块中添加通道注意力eSE

模块,如图3所示。同时,为了提升SE模块的执行效率,本文将SENet中两个全连接层替换为一个点卷积,以降低模型的复杂性,提升执行效率。由特征图的可视化(见图4,彩图扫OSID可见,下同)可见,注意力机制对模型存在积极影响。

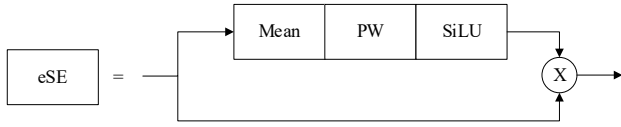


Fig. 3 eSE module

图3 eSE 模块



(a) Add eSE attention mechanism

(a) 添加eSE注意力机制



(b) No eSE attention mechanism

(b) 没有eSE注意力机制

Fig. 4 Visualization results of feature maps

图4 特征图可视化结果

3 目标检测算法剪枝

3.1 基于批量归一化层的通道剪枝

剪枝技术可分为结构化剪枝和非结构化剪枝。其中,非结构化剪枝技术可有效平衡参数数量和模型性能,但由于网络拓扑结构的变化,需要特殊硬件设备支持其正确运行。相比之下,结构化剪枝技术可显著减少网络中滤波器组和特征通道数量,使模型高效运行,无需复杂的算法设计和特殊的硬件。为此,本文采用批量归一化层系数评估每个通道网络参数的重要性,以实现网络的结构化剪枝。

批量归一化可将每个层的神经元输出标准化,以维持激活值的分布稳定,从而加快模型训练速度,有效抑制过拟合现象发生。批量归一化的主要优点之一是有助于稳定深度神经网络训练。深度神经网络对权重的初始化非常敏感,不良的初始权重会导致训练难以收敛,而批量归一化可减少神经网络对权重初始化的敏感性,使训练过程更稳定。此外,通过批量归一化还可有效改善神经网络的性能。在训练过程中,由于输入分布的变化,网络可能产生不准确的预测结果,而批量归一化可有效减少内部协变量偏移,从而提升神经网络预测的准确性。

批量归一化层的计算公式如式(1)所示,每个通道的输出值 z_{out} 与系数 γ 呈正相关,在 γ 接近于0时激活值非常小。本文假设去掉 γ 趋于0的通道,对于模型的影响不大,

并且实际实验结果也符合预想。

$$\hat{z} = \frac{z_{in} - \mu_{\beta}}{\sqrt{\sigma_{\beta}^2 + \varepsilon}}; z_{out} = \hat{z}\gamma + \beta \quad (1)$$

由于不同数据集批量归一化层系数分布的规律不同,可通过增加约束方法使其符合要求。在OUC-UAV-DET数据集上训练发现,批量归一化层的系数呈正态分布,但其期望并不在0附近,无法直接根据 γ 进行剪枝。因此,本文在损失函数中添加L1正则约束,尝试将批量归一化层的系数分布期望调整到0附近。

$$L = \sum_{(x,y)} l(f(x, W), y) + \lambda \sum_{\gamma \in \Gamma} g(\gamma) \quad (2)$$

其中:第一项是正常训练的损失函数;第二项是正则化约束; $g(s) = |s|$; λ 为正则系数。

在反向传播时,损失函数的导数由式(3)所示。

$$L' = \sum l' + \lambda \sum g'(\gamma) = \sum l' + \lambda \sum |\gamma|' = \sum l' + \lambda \sum \text{sign}(\gamma) \quad (3)$$

在添加正则项后,随着训练进行,批量归一化层参数正态分布的期望趋向于0,此时可进行剪枝操作。在设定 γ 阈值时,应不能大于任何通道批量归一化层的最大 γ ,通过该阈值可有效剪除低于阈值的通道。通常情况下,剪枝结束后还需要重新训练模型以恢复模型精度。

3.2 针对单类目标检测的输出层剪枝

YOLOv5结构在网络的3个或4个不同阶段输出检测结果。其中,3个输出层对应的输出由上至下分别通过P3、P4和P5进行表示,第4个输出层根据具体情况可能会由P2或P6进行表示。当输入图像大小为640×640时,经过8倍下采样的P3检测层特征图大小为80×80,特征图中每个像素点可检测原始大小在8×8像素以上的目标。以此类推,P4经过16倍下采样负责检测16×16以上的目标,P5经过32倍下采样负责检测16×16以上的目标。图5为对原始的YOLOv5输出层剪枝后的输出结构。

在VisDrone数据集、OUC-UAV-DET数据集中,某些类别的样本尺寸相对固定。本文认为在检测尺寸相对固定的类别时,对多个尺度的特征进行融合并不能提升模型性能。为此,本文针对YOLO的输出层进行剪枝,表1为YOLOv5s模型通过裁剪掉输出层对参数数量和计算量的影响。

4 实验结果与分析

4.1 数据集

本文使用的数据集包括VisDrone数据集和中国海洋大学无人机目标检测(OUC-UAV-DET)数据集。

VisDrone航拍小目标数据于2018年提出,旨在为无人机应用场景下的物体检测提供一个相对通用和全面的大规模数据集。该数据集包含10 209张静态图像(6 471张用于训练,548张用于验证,1 610张用于测试及1 580张未

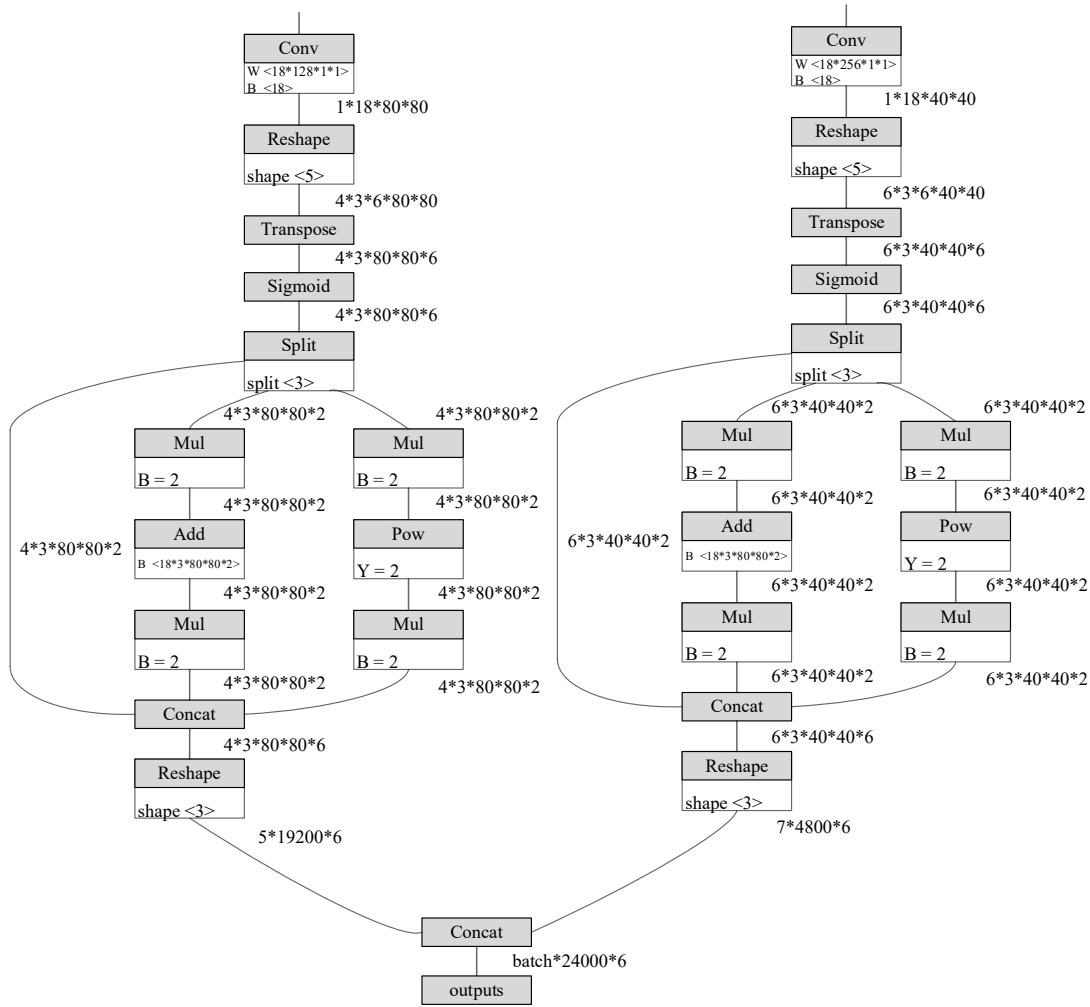


Fig. 5 YOLOv5 output layer pruning

图5 YOLOv5 输出层剪枝

Table 1 Comparison of YOLO output layer parameters and computation

表1 YOLO 输出层参数及计算量比较

输出层	层数	参数量	GFLOPs
P3+P4+P5	157	7 012 822	15.8
P3+P4	140	5 232 324	14.3
P3	123	4 784 434	12.9

标注的图像),主要关注人类和车辆,并定义了直立行人、非站立状态的人、小汽车、货车、公共汽车、卡车、面包车、自行车、遮阳篷三轮车和三轮车共10个感兴趣的对象类别。

中国海洋大学无人机目标检测(OUC-UAV-DET)数据集由中国海洋大学视觉实验室采集和标注。该数据集主要在校园、海滩和街道等人车密集的场所拍摄,采用大疆无人机作为图像采集设备。目前,该数据集包括2 071张高分辨率的静态图像,包括人、自行车、汽车、面包车、货车、自行车、公共汽车、摩托车和船共9类目标。

4.2 实验环境

本文训练的GPU为2080Ti(显存12 GB),部署时使用

英伟达的Jetson AGX Xavier平台,算力为30 Tops,最大功耗为30 W。实验的操作系统为Ubuntu18.4,安装了包括TensorRT 8.5.2、cuDNN 8.6.0、CUDA 11.4.19、OpenCV 4.5.4等软件库的JetPack5.1软件包。

4.3 实验评价指标

在目标检测中,平均值(Average Precision, AP)、均值平均精度(mean Average Percision, mAP)是两个重要的衡量指标^[17,18],mAP为AP值除以类别数,由式(1)所示。mAP@50指交并比阈值为0.5的mAP,mAP@50:95指交并比阈值分别在0.5与0.95之间每隔0.05的mAP的平均值,AP为准确率—召回率曲线下的面积。

$$mAP = \frac{AP}{\text{类别数}} \tag{4}$$

PR曲线由改变IOU阈值产生的多组准确率与召回率,在平面坐标系下对应的点连接而成。目前,主流计算PR曲线下面积(即AP)的方式有VOC数据集和COCO数据集的计算方式。本文使用COCO方式进行AP计算,实验时延指推理时间与非极大值抑制耗时的总和。

4.4 结果分析

表2为本文方法与原始YOLOv5的比较结果。由此可

知, 本文方法相较于 YOLOv5n 及其他模型精度更高, 推理速度更快, 更适用于无人机实时检测场景, 体现了本文设计网络的优越性。表 3 的消融实验证明, 添加高效的注意力模块 (eSE) 和使用 ADD 进行特征融合, 能有效提升模型的 mAP, 降低时延。表 4 为使用批量归一化层的通道剪枝方法后, YOLOv5s 在 OUC-UAV-DET 数据集上的测试结果。由此可知, 当阈值为 0.65 时模型精度大幅下跌, 且速度提升已不明显, 因此本文选用 0.6 为剪枝阈值。

Table 2 Model evaluation on the VisDrone dataset

表 2 VisDrone 数据集模型评估

模型名称	mAP@50/%	mAP@50:95/%	时延/ms
YOLOv6n ^[19]	39.7	20.3	6.0
YOLOv7-tiny ^[20]	41.3	22.9	5.8
YOLOv5n	42.3	21.3	2.8
Ours	42.9	23.5	2.1

Table 3 Ablation experiments

表 3 消融实验

eSE	sE	PAN	mAP/%	时延/ms
		Concat	20.7	2.1
		ADD	20.4	2.0
√		Concat	23.6	2.5
	√	ADD	23.4	2.2
√		ADD	23.5	2.1

Table 4 Channel pruning results on the OUC-UAV-DET dataset

表 4 OUC-UAV-DET 数据集通道剪枝结果

剪枝阈值	mAP@50/%	mAP@50:95/%	时延/ms
0	72.5	47.4	12
0.50	70.4	45.5	5.3
0.60	69.3	43.9	4.5
0.65	56.2	30.9	4.3

由表 5 可知, 单类目标的 YOLO 输出层剪枝在 VisDrone 数据集和 OUC-UAV-DET 数据集上的可行性, 在只保留 P3 层的情况下模型精度并未下降, 只在 P3 层输出时的检测效果甚至更好, 证实了在检测尺寸相对固定的类别时, 对多个尺度的特征进行融合并不能提升模型性能。

4.5 边缘设备部署结果分析

模型转换后, 模型精度略微下降, 可能是由于单精度 (fp32) 浮点数转半精度 (fp16) 浮点数所带来的损失。表 6 为各模型在边缘设备 Xavier 上的性能, 结果与表 2 类似。由此可知, 本文改进的轻量化目标检测模型在 VisDrone 与 OUC-UAV-DET 数据集上的表现优于 YOLOv5、YOLOv6 及 YOLOv7 的轻量版本。表 7 展示了推理一张高分辨率的 4 000×2 250 图像, 从端到端所消耗的时间及帧率。由此可知, 本文方法理论上可达到 6.1 FPS 的预测速度。

Table 5 Comparison of vehicle detection using different output layers

表 5 车辆检测使用不同输出层比较

数据集	输出层	mAP@50/%	mAP@50:95/%	时延/ms
VisDrone	P3+P4+P5	87.0	62.7	7.2
VisDrone	P3+P4	87.2	63.1	6.7
VisDrone	P3	87.6	63.0	5.6
OUC-UAV-DET	P3+P4+P5	95.6	70.8	7.3
OUC-UAV-DET	P3+P4	95.7	71.3	6.7
OUC-UAV-DET	P3	95.4	71.6	5.6

Table 6 Model performance on Jetson platform

表 6 Jetson 平台模型性能

数据集	模型	mAP@50/%	mAP@50:95/%	时延/ms
OUC-UAV-DET	YOLOv5s	66.3	43.9	9.1
OUC-UAV-DET	YOLOv5n	61.7	36.3	7.3
OUC-UAV-DET	Ours	65.2	41.4	5.7
VisDrone	YOLOv5n	41.9	21.1	6.2
VisDrone	YOLOv6n	39.1	19.7	6.5
VisDrone	YOLOv7-tiny	41.6	22.6	8.4
VisDrone	Ours	42.2	23.1	5.7

Table 7 Inference speed of the model on Jetson platform (high-resolution image cropping)

表 7 Jetson 平台模型推理速度 (高分辨率图像切图)

模型	时延/ms	FPS
YOLOv7-tiny	238	4.2
YOLOv6n	186	5.4
YOLOv5n	176	5.7
Ours	163	6.1

5 结语

本文从轻量化网络设计与模型剪枝方向出发, 对 VisDrone 数据集与 OUC-UAV-DET 全类别和特定类别进行优化, 从模型与算法层面提升了无人机目标检测算法的性能。通过轻量化改造 YOLOv5 的主干网络, 引入了一种更高效的通道注意力机制提升网络精度。实验证明, 该方法在公开无人机数据集 VisDrone 和自建数据集 OUC-UAV-DET 上, 能在保持较高精度的基础上具备更快的推理速度。

同时, 本文针对无人机单类别检测任务中目标实例尺寸相对固定的特点, 采用网络剪枝方法减少网络输出层, 结合 OUC-UAV-DET 数据集中批量归一化层系数服从正态分布的规律, 在损失函数中添加正则项, 逐渐将批量归一化层系数分布的期望值趋于 0。然后, 根据批量归一化层的系数绝对值大小, 评估网络中各通道的重要性, 并根据阈值对不重要通道进行裁剪。实验证明, 两种剪枝方式

在相关数据集上取得了良好的成效。

此外,本文在英伟达 Jetson 平台上对模型进行部署和测试,结果表明在边缘智能设备上,本文设计的轻量化目标检测模型和剪枝方法均取得了一定成效。基于此次研究,能使人们深入理解和掌握无人机目标检测算法和应用,为无人机目标检测应用提供重要的技术支持。未来,将进一步扩充高分辨率航拍图像的数据量,通过更多实际场景数据提升模型的泛化性,以应对在不同场景的实际检测过程中可能遇到的问题。

参考文献:

- [1] PLATT J. Sequential minimal optimization: a fast algorithm for training support vector machines [EB/OL]. <https://www.microsoft.com/en-us/research/publication/sequential-minimal-optimization-a-fast-algorithm-for-training-support-vector-machines/>.
- [2] FREUND Y, SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. *Journal of Computer and System Sciences*, 1997, 55(1):119-139.
- [3] SERMANET P, EIGEN D, ZHANG X, et al. Overfeat: integrated recognition, localization and detection using convolutional networks [DB/OL]. <https://arxiv.org/abs/1312.6229>.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2014: 580-587.
- [5] GIRSHICK R. Fast R-CNN [C]// *Proceedings of the IEEE International Conference on Computer Vision*, 2015: 1440-1448.
- [6] 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, 36(6): 1137-1149.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [DB/OL]. <https://arxiv.org/abs/1506.02640>.
- [8] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications [DB/OL]. <https://arxiv.org/abs/1704.04861>.
- [9] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: inverted residuals and linear bottlenecks [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 4510-4520.
- [10] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetv3 [C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 1314-1324.
- [11] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: practical guidelines for efficient CNN architecture design [C]// *Proceedings of the European Conference on Computer Vision*, 2018: 116-131.
- [12] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNET: a new backbone that can enhance learning capability of CNN [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020: 390-391.
- [13] LEE Y, HWANG J W, LEE S, et al. An energy and GPU-computation efficient backbone network for real-time object detection [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019: 752-760.
- [14] DING X, ZHANG X, MA N, et al. RepVGG: making VGG-style convnets great again [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 13733-13742.
- [15] HAN K, WANG Y, TIAN Q, et al. GhostNet: more features from cheap operations [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 1580-1589.
- [16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7132-7141.
- [17] ZHU X H, QIAN L P, BO W. Overview of research on image data enhancement technology [J]. *Software Guide*, 2021, 20(5): 230-236.
朱晓慧,钱丽萍,傅伟. 图像数据增强技术研究综述 [J]. *软件导刊*, 2021, 20(5): 230-236.
- [18] CHEN Y J, CHEN C C, LYU G J, et al. Sentiment analysis model for tourism evaluation integrating attention mechanism and Bi-LSTM [J]. *Software Guide*, 2022, 21(2): 43-47.
陈玉娇,陈崇成,吕贵杰,等. 融合注意力机制和Bi-LSTM的旅游评价情感分析模型 [J]. *软件导刊*, 2022, 21(2): 43-47.
- [19] LI C, LI L, JIANG H, et al. Yolov6: a single-stage object detection framework for industrial applications [DB/OL]. <https://arxiv.org/abs/2209.02976>.
- [20] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Yolov7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]// *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 7464-7475.

(责任编辑:刘嘉文)