

# 基于迅颈模块与单解耦头的航拍车辆检测方法

高宏伟, 王雨桐

(沈阳理工大学 自动化与电气工程学院, 沈阳 110159)

**摘要:** 无人机航拍图像的目标检测需要大量计算资源, 导致其在移动端上的应用受到限制, 为此提出一种轻量化目标检测网络 Faster S。首先, 在颈部网络设计了由部分卷积和组卷积组成的轻薄迅颈模块, 用以有效聚合不同尺度的信息; 其次, 在检测头部分设计了一种高效、轻量化的单解耦头输出结构 (SDHEAD), 用以进一步分离目标检测中分类和定位两个子任务的特征编码。在航拍车辆数据集 CARPK 上的实验结果表明: 与特征金字塔网络 (FPN) 结构相比, 采用迅颈模块提高了模型的检测精度, 减少了模型的参数量; 相比于其他检测头, SDHEAD 具有更强的特征辨别能力; Faster S 模型检测的平均精度 ( $AP_{0.5}$ ) 达到了 70.9%, 较 Yolo-FastestV2 提高了 12.7%, 较 Fastestdet 提高了 3.5%, 推理时间仅为 29.56 ms, 较 Nanodet-m 降低了 76%, 模型参数量仅为 0.335 M。

**关键词:** 目标检测; 航拍图像; 轻量化; 单解耦头; 迅颈模块

**中图分类号:** TP391.41 **文献标志码:** A **DOI:** 10.3969/j.issn.1003-1251.2024.02.001

## Aerial Vehicle Detection Method Based on Fast Neck Module and Single Decoupled Head

GAO Hongwei, WANG Yutong

(Shenyang Ligong University, Shenyang 110159, China)

**Abstract:** Object detection in drone aerial images require intensive computational resources, which limits the application on mobile terminals. Therefore, a lightweight target detection network Faster S is proposed. Firstly, a light and thin neck module composed of partial convolution and group convolution is designed in the neck network to effectively aggregate information of different scales. Secondly, an efficient and lightweight single decoupled head output structure (SDHEAD) is designed in the detection head part, which can further separate the feature encoding of the two subtasks of classification and localization in object detection. The results of experimental on the aerial vehicle dataset CARPK show that compared with the feature pyramid network (FPN) structure, the fast neck module improves the detection accuracy of the model and reduces the parameters of the model. Compared with other detection heads, SDHEAD has stronger feature discrimination ability. The average accuracy ( $AP_{0.5}$ ) of Faster S model detection reaches 70.9%, which is 12.7% higher than Yolo-FastestV2, 3.5% higher than Fastestdet, and the inference time is only 29.56 ms, which is 76% lower than Nanodet-m. The parameter size is only 0.335 M.

**Key words:** object detection; aerial image; lightweight; single decoupled head; fast neck module

无人机是一种高效的大规模环境传感与监测设备<sup>[1]</sup>,广泛应用于军事、民用等多个领域<sup>[2]</sup>。航拍图像的目标检测是无人机应用研究的热点内容,目标检测也是计算机视觉领域中最基础且最具挑战性的任务之一<sup>[3]</sup>。目标检测算法中,两阶段模型通常精度较高<sup>[4]</sup>,但效率较低,难以在计算资源有限的无人机平台上实现。近年来,基于YOLO的单阶段模型被广泛应用于嵌入式系统<sup>[5]</sup>,但基于锚框的YOLO模型,在资源极度有限的无人机设备上难以满足实时性需求。此外,YOLO需要通过数据集获取锚框的尺寸,这种方式使得模型在其他数据集上的泛化能力较弱。

针对上述问题,许多研究人员致力于开发高效的检测器架构,提出了Nanodet、YOLOX-Nano、YOLOX-Tiny和PP-PicoDet等网络模型<sup>[6-7]</sup>。尽管这些检测器促进了轻量化目标检测研究的进展,但应用于边缘设备上的无人机检测中仍存在两个实际问题:一是GPU设备上多检测头导致网络参数和计算量(GFLOPs)显著增加,特别是随着多分支结构的进一步扩展,GFLOPs增加更为明显<sup>[8]</sup>;二是现有的轻量化模型没有考虑到航拍时角度变化会导致图像中目标的分布稀疏及正负样本不平衡<sup>[9]</sup>。

为解决以上问题,本文提出一种对无人机边缘GPU友好的使用无锚策略且为单输出解耦头

的轻量级深度学习架构Faster S。首先,提出一种称为迅颈模块的轻薄结构,用以替代特征金字塔网络(FPN)<sup>[10]</sup>和路径聚合网络(PANet)<sup>[11]</sup>的颈部层,迅颈模块由简单而高效的运算器部分卷积(PCConv)和分组卷积构成,可大大减少网络参数,更好地满足轻量化网络的需求。其次,提出一种轻量级单输出解耦头(single decoupled head, SD-HEAD),该解耦头根据目标检测中分类和定位两个子任务的特点,设计了两个分支,即用于分类的语义特征提取器和用于定位的边界特征提取器,两分支利用不同语义上下文的特征,以适应不同的目标尺度和复杂度。SDHEAD通过两个分支的特征融合,结合不同层级的感受野,实现单一输出头检测不同尺度物体的功能。最后,为解决航拍图像正负样本不平衡问题,采用动态标签分配策略(adaptive training sample selection, ATSS)优化训练过程<sup>[12]</sup>,获得全局优化结果。

## 1 网络结构与原理

本文提出的Faster S网络整体结构如图1所示,图中各变量含义详见下文中各部分说明。Faster S的主体部分由主干网络、颈部网络和单输出解耦检测头组成。

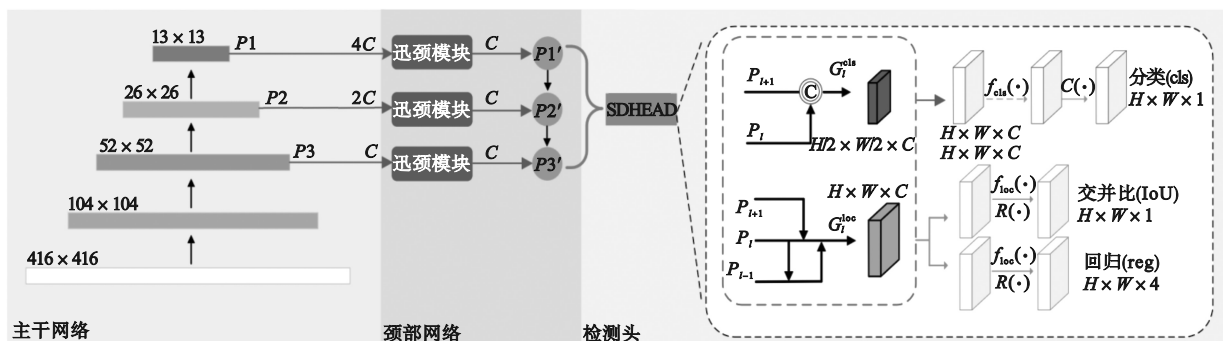


图1 Faster S结构

Fig. 1 The structure of Faster S

### 1.1 轻量化主干网络

本文主干网络采用高效而强壮的深度神经网络 ShuffleNet。相比于其他轻量化主干网络,如 MobileNet<sup>[13]</sup>、GhostNet<sup>[14]</sup>等, ShuffleNet<sup>[15]</sup>使用通道洗牌技术,将输入特征图分成较小的组,并将这些组重新组合,从而减少模型参数数量。这种分组卷积的方法不仅可以有效减少计算量,还可以避免过拟合和降低内存占用。此外, ShuffleNet还具有较好的鲁棒性,其采用深度可分离卷积和残差块等先进技术,在小数据集上的表现比其他

轻量化主干网络更好,可在资源受限的环境下高效完成目标检测任务。

### 1.2 轻量化颈部网络

本文在Faster S的颈部层设计了一种轻量化迅颈网络,计算公式如下。

$$P1' = F(P1) \quad (1)$$

$$P2' = \text{Concat}(F(P2), US_{2 \times}(P1')) \quad (2)$$

$$P3' = \text{Concat}(F(P3), US_{2 \times}(P2')) \quad (3)$$

式中: $P1$ 为骨干网络的最高层; $P1'$ 为迅颈网络的最高层; $P2$ 为骨干网络的第四层; $P2'$ 为迅颈网络

的第二层; $P_3$  为骨干网络的第三层; $P_3'$  为迅颈网络的第一层; $US_{2 \times}(\cdot)$  表示 2 倍下采样; $F(\cdot)$  表示迅颈模块; $\text{Concat}(\cdot)$  表示通道拼接操作。

$P_1$  经过迅颈模块得到  $P_1'$ ,  $P_2$  经过迅颈模块并与  $P_1'$  通道拼接后得到  $P_2'$ ,  $P_3$  经过迅颈模块并与  $P_2'$  通道拼接后得到  $P_3'$ 。迅颈网络利用迅颈模块降低了深层网络的通道数,但并不降低深层网络的特征提取能力。

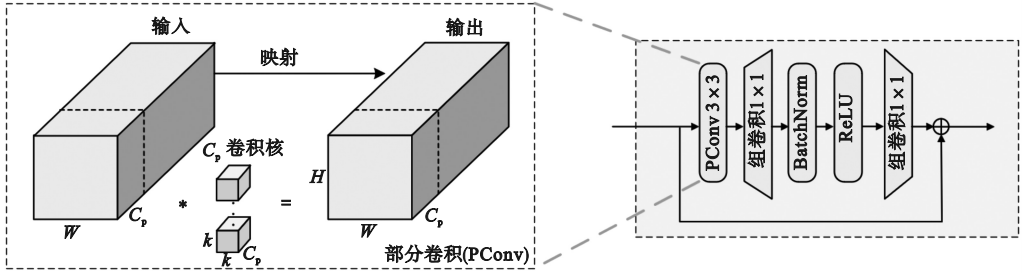


图 2 迅颈模块结构

Fig. 2 The structure of fast neck module

PConv 结构计算公式如下。

$$Y = \text{Concat}(\text{Conv}(X_{C_p}), X_{C-C_p}) \quad (4)$$

式中: $Y$  为输出特征; $C$  为输入特征的通道数; $C_p$  为进行 PConv 的通道数; $X_{C_p}$  为进行 PConv 的输入特征; $X_{C-C_p}$  为进行映射的输入特征; $\text{Conv}(\cdot)$  表示卷积操作。

PConv 只对部分输入通道进行常规卷积,而其余通道保持不变,这种方式可以简单有效地利用特征图中不同通道的信息。PConv 的计算量 (FLOPs) 为  $H \times W \times k^2 \times C_p^2$ , 其中  $H$  和  $W$  分别为特征图的高和宽,  $k$  为卷积核大小。

当 PConv 的通道数为输入特征通道数的 1/4 时, PConv 的 FLOPs 仅为普通卷积的 1/16。此外, PConv 具有较小的内存访问量, 仅为常规卷积的 1/4。

### 1.3 检测头 SDHEAD

本文通过分析目标检测任务中分类和定位两个主要子任务的差异, 并融合不同层间感受野特征, 优化检测头的设计, 提出一种适用于轻量级网络的单输出解耦头 SDHEAD。

#### 1.3.1 检测头框架

本文采用解耦头的传统目标检测器, 将分类和定位两个任务分开处理, 以充分利用不同类型的信息, 提高目标检测的准确性。

给定一个颈部层  $l$  的真实对象, 具有边界框  $B$  和类别标签  $C$ , 采用解耦头的目标检测器在特征图  $P_l$  上的损失  $L$  包括分类损失  $L_{\text{cls}}$  和定位损失

迅颈模块结构如图 2 所示。该模块由一个  $3 \times 3$  的 PConv 和两个组卷积组成, 这些结构一起形成倒置的残差块, 中间层的组卷积具有扩展通道数量的功能, 并通过残差连接以重用输入特征。此外, 在第一个  $1 \times 1$  组卷积运算后, 加入归一化层 (BatchNorm) 和激活层 (ReLU), 以保持特征多样性并实现更低的延迟。

$L_{\text{loc}}$ , 即

$$L = L_{\text{cls}}(F_c(P_l), C) + L_{\text{loc}}(F_r(P_l), B) \quad (5)$$

式中: $F_c(\cdot) = \{f_{\text{cls}}(\cdot), S(\cdot)\}$ , 表示分类分支; $F_r(\cdot) = \{f_{\text{loc}}(\cdot), R(\cdot)\}$ , 表示定位分支; $f_{\text{cls}}(\cdot)$  和  $f_{\text{loc}}(\cdot)$  分别表示用于提取分类和定位特征的投影函数; $S(\cdot)$  和  $R(\cdot)$  分别表示分类分支和定位分支中的最后一层模块, 用于将特征解码为分类分数和边界框位置。

传统解耦头的  $f_{\text{cls}}(\cdot)$  和  $f_{\text{loc}}(\cdot)$  共享相同的结构, 但该设计不能充分获得分类和定位所需要的不同语义信息和边界信息。为此, SDHEAD 在源头上解耦两个任务的特征编码, 使用特定任务的输入特征为两个分支提供输入 (图 1 中  $G_l^{\text{cls}}$  和  $G_l^{\text{loc}}$ ), 故式 (5) 可以写为

$$L = L_{\text{cls}}(F_c(G_l^{\text{cls}}), C) + L_{\text{loc}}(F_r(G_l^{\text{loc}}), B) \quad (6)$$

对于分类分支, 本文生成空间更粗糙但语义更丰富的特征图; 对于定位分支, 本文为其提供了包含更加详细的纹理和边界信息的特征图。

如图 1 检测头部分所示, SDHEAD 接收三个颈部层的特征图, 对这些特征图进行融合以获取不同的感受野信息, 并将生成的解耦特征图用于分类和定位任务。在分类阶段, 采用两层特征投影函数, 在定位阶段仅采用一层特征投影函数, 在减少参数的同时保证模型的高精度。SDHEAD 包括用于分类的语义特征提取器和用于定位的边界特征提取器两部分。

### 1.3.2 语义特征提取器

语义特征提取器结构如图3所示。语义特征提取器利用来自  $P_l$  和  $P_{l+1}$  两个颈部层的特征图生成语义丰富的特征图进行分类。

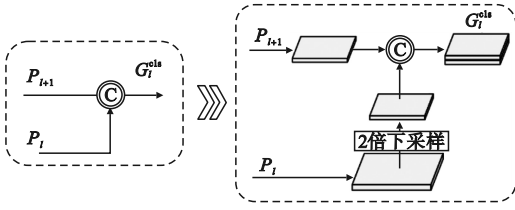


图3 语义特征提取器结构

Fig.3 The structure of semantic feature extractor

首先对  $P_l$  进行2倍下采样,下采样采用平均池化操作,卷积核大小为  $3 \times 3$ ,步距为2。将其与  $P_{l+1}$  聚合,生成最终的  $G_l^{cls}$ ,公式为

$$G_l^{cls} = \text{Concat}(\text{Avg}(P_l), P_{l+1}) \quad (7)$$

式中  $\text{Avg}(\cdot)$  表示下采样层。

输入  $G_l^{cls}$  传递给  $F_c(\cdot)$ ,用以预测分类分数。 $f_{cls}(\cdot)$  使用卷积核大小为  $3 \times 3$  的卷积操作,并应用组卷积操作降低参数量; $S(\cdot)$  使用卷积核大小为  $3 \times 3$  的卷积操作进行最终的分类预测。这种深而狭窄的  $f_{cls}(\cdot)$  设计可以获得更多的语义信息,且只需很少的额外计算成本,有利于提高分类准确率。

通过上述操作,语义特征提取器融合了  $P_l$  层和  $P_{l+1}$  层间的感受野,不仅可利用显著特征的稀疏性,还可受益于更高金字塔级别的上下文语义,有助于更有效地推断对象类别,特别是没有纹理的对象和被严重遮挡的对象。

### 1.3.3 边界特征提取器

边界特征提取器结构如图4所示。

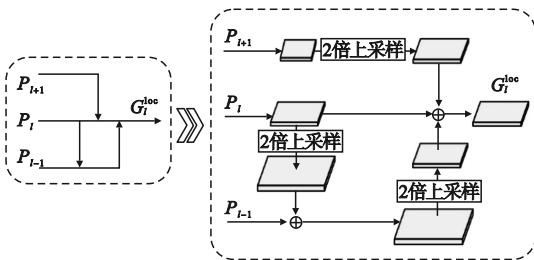


图4 边界特征提取器结构

Fig.4 The structure of boundary feature extractor

边界特征提取器利用来自颈部层的特征图  $P_{l-1}$ 、 $P_l$  和  $P_{l+1}$ ,在融合感受野的同时, $P_{l-1}$  提供了更多的细节和边缘特征, $P_{l+1}$  提供了更全面的物体视角。为提高计算效率,本文对  $P_{l-1}$ 、 $P_l$  和  $P_{l+1}$  进行简单融合。 $P_l$  首先进行2倍上采样,然后与  $P_{l-1}$  相加,之后进行下采样恢复到  $P_l$  的分辨率,下

采样采用平均池化操作,卷积核大小为  $3 \times 3$ ,步距为2。该设计有效保留了  $P_{l-1}$  中的详细信息,且几乎没有额外的计算开销。最后对  $P_{l+1}$  进行上采样并相加,生成最终的  $G_l^{loc}$ 。公式表示为

$$G_l^{loc} = P_l + \mu(P_{l+1}) + \text{Avg}[\mu(P_l) + P_{l-1}] \quad (8)$$

式中  $\mu(\cdot)$  表示上采样。

输入  $G_l^{loc}$  传递给  $F_r(\cdot)$ ,用以预测回归框。 $f_{loc}(\cdot)$  和  $R(\cdot)$  使用卷积核大小为  $3 \times 3$  的卷积操作进行最终的回归预测。

## 2 实验结果和分析

### 2.1 数据集

本文使用 CARPK 数据集,包含由无人机(大疆,型号为 PHANTOM 3 PROFESSIONAL)拍摄的来自4个不同停车场的89 777辆汽车图像。无人机拍摄高度约为40 m,图像大小为  $1\,280 \times 720$  像素。

### 2.2 实验细节与评价指标

本文实验平台硬件配置为:AMD Ryzen 55600 处理器,16 GB 运行内存,NVIDIA RTX3070 GPU。系统环境配置为:Ubuntu22.04,PyTorch 1.11,CUDA11.3。

实验中采用  $416 \times 416$  像素大小的三通道图像作为网络的输入,初始学习率为0.001,使用SGD优化器,动量为0.949,权重衰减为0.0005,训练批次大小为32,训练的迭代次数为500次。使用负对数似然函数和 Smooth L1 作为损失函数。推理时间包括前向推理时间和后处理时间。由于轻量级网络的数据增强不适合采用复杂的方法,如Moscia和Mixup等,故本文采用简单的数据增强方法,如随机平移和随意缩放等。

本文通过对比实验和消融实验对Faster S算法的性能进行分析,采用的评价指标包括平均精度均值(mAP)、FLOPs、参数量和推理时间。mAP是所有检测类别平均精度(AP)的均值,其数值越大,说明模型准确度越高,本研究中数据集仅有车辆一个类别,故mAP即为AP。本文使用的评价指标包括  $\text{AP}_{0.5:0.9}$ 、 $\text{AP}_{0.5}$  和  $\text{AP}_{0.75}$ ,其中  $\text{AP}_{0.5:0.9}$  表示在不同交并比(IoU)阈值(0.5:0.95,步长0.05)上的平均AP, $\text{AP}_{0.5}$  和  $\text{AP}_{0.75}$  分别表示IoU设置为0.5和0.75时所得到的AP值。FLOPs代表模型的复杂程度,即计算量。参数量是衡量模型复杂度和容量的一个重要指标。推理时间代表模型在推理单张图片时所用的时间,其数值越小,说明推理速度越快。

## 2.3 对比实验

### 2.3.1 不同网络对比

为证明 Faster S 的优越性能,本文在 CARPK

数据集上同时使用其他较为先进的轻量级网络(包括 Fastestdet、Yolo-FastestV2 和 Nanodet-m)进行实验。实验结果比较如表 1 所示。

表 1 不同网络在 CARPK 数据集上的实验结果比较

Table 1 Comparison of experimental results of different algorithms on the CARPK dataset

网络	参数量/M	AP <sub>0.5:0.95</sub> /%	AP <sub>0.5</sub> /%	AP <sub>0.75</sub> /%	FLOPs/s <sup>-1</sup>	推理时间/ms
Fastestdet	0.236	33.3	68.5	28.4	177	27.48
Yolo-FastestV2	0.238	32.3	62.9	27.8	155	29.39
Nanodet-m	0.936	50.1	78.7	59.3	720	124.00
<b>Faster S</b>	<b>0.335</b>	<b>36.7</b>	<b>70.9</b>	<b>34.4</b>	<b>225</b>	<b>29.56</b>

由表 1 可以看出:Faster S 相较于 Fastestdet, AP<sub>0.5</sub>提高了 3.5%,参数量仅增加 0.099 M,推理时间增加极少;Faster S 相较于 Yolo-FastestV2, AP<sub>0.5</sub>提高了 12.7%,参数量仅上升了 0.097 M,推理时间几乎不变;Faster S 相较于 Nanodet-m,尽管检测精度有所降低,但参数量和 FLOPs 约为前者的 1/3,推理时间则降低了 76%。可见,Faster S 具有较佳的综合性能。

### 2.3.2 不同主干网络对比

为探究不同主干网络对 Faster S 检测性能的影响,本文对目前流行的轻量级主干网络(包括 Ghostnet、Mobilenetv3-small 和 Shufflenetv2)在 CARPK 数据集上进行实验。在不影响网络检测精度的前提下,为减少主干网络的重量,本文使用三种主干网络的 0.5 倍大小的模型结构。实验结果如表 2 所示。

表 2 不同主干网络对检测性能的影响

Table 2 The impact of different backbone networks on the algorithm

主干网络	参数量/M	AP <sub>0.5:0.95</sub> /%	AP <sub>0.5</sub> /%	AP <sub>0.75</sub> /%	FLOPs/s <sup>-1</sup>	推理时间/ms
Mobilenetv3-small × 0.5	0.763	24.7	59.2	16.1	419.2	40.30
Ghostnet × 0.5	0.780	24.8	54.1	20.4	228.5	38.00
<b>Shufflenetv2 × 0.5</b>	<b>0.335</b>	<b>36.7</b>	<b>70.9</b>	<b>34.4</b>	<b>225.0</b>	<b>29.56</b>

由表 2 可以看出:本文使用 Shufflenetv2 × 0.5 为主干网络时,Faster S 参数量仅为 0.335 M,远少于 Mobilenetv3-small × 0.5 和 Ghostnet × 0.5 为主干网络时的参数量;采用 Shufflenetv2 × 0.5 的模型 AP<sub>0.5</sub>较 Ghostnet × 0.5 提高了 31.1%,较 Mobilenetv3-small × 0.5 提高了 19.8%;采用 Shufflenetv2 × 0.5 的模型计算量 FLOPs 较 Mobilenetv3-small × 0.5 降低了 194.2 s<sup>-1</sup>,且其推理时间也较其他主干网络更低。

尽管 Ghostnet 和 Mobilenetv3 被设计为轻量化模型,但其最轻量化版本在特征提取能力上不

及 Shufflenetv2,因而精度较低。此外,前两者的输出通道数量远超过 Shufflenetv2 网络,需在颈部进行大量的卷积操作以调整通道数量,导致参数量大幅增加。

## 2.4 消融实验

为探究 Faster S 中不同结构(迅颈网络、SD-HEAD)对参数量、准确率、FLOPs 和推理时间的影响,本文以 Shufflenetv2 × 0.5 主干网络、FPN 颈部网络、单独检测头(YOLOX 解耦头)为基础网络进行消融实验研究,消融实验结果如表 3 所示。

表 3 CARPK 数据集上的消融实验结果

Table 3 Comparison of ablation experimental results on the CARPK dataset

基础网络	迅颈网络	SDHEAD	参数量/M	AP <sub>0.5:0.95</sub> /%	AP <sub>0.5</sub> /%	FLOPs/MB	推理时间/ms
✓			0.843	30.1	65.6	672.83	29.90
	✓		0.532	35.2	69.6	357.26	29.60
		✓	0.569	31.9	68.2	487.75	28.80
	✓	✓	0.335	36.7	70.9	225.00	29.56

在基础网络中主干网络和检测头不变的情况下,将FPN替换为迅颈网络,实验结果如表3中第二行所示。与基础网络比较,在参数量降低0.311 M的同时, $AP_{0.5:0.9}$ 和 $AP_{0.5}$ 分别提高了16.9%和6.1%,推理时间降低了0.3 ms,验证了新的颈部网络的有效性。迅颈网络不仅减小了网络规模,而且保持了出色的检测精度。这是因为自顶向下结构可以保证深层语义信息传递到浅层特征图,浅层特征图可为无人机拍摄的小对象提供语义支持。

保持基础网络中主干网络和颈部网络不变的情况下,将检测头替换为SDHEAD,实验结果如表3中第三行所示。相较于单独的YOLOX解耦头,使用SDHEAD的模型参数量降低了0.274 M, $AP_{0.5:0.95}$ 提高了6.0%, $AP_{0.5}$ 提高了4.0%,推理时间降低了1.1 ms。该结构不仅可以提高检测精度,而且可以提高推断速度,充分展现了SDHEAD的高效性能。

保持基础网络中主干网络不变,将FPN替换为迅颈网络、检测头替换为SDHEAD,实验结果如表3中第四行所示。可见,Faster S参数量仅有0.335 M, $AP_{0.5:0.95}$ 高达36.7%,推理时间仅为29.56 ms。相比于其他几种网络结构,Faster S参数量更低、精度更高、推理速度更快,总体性能更好。

### 3 结论

本文提出了一种新型的轻量级目标检测网络Faster S,用于无人机的实时目标检测。首先,设计了一种轻薄的迅颈模块,引入迅颈模块的颈部层在保持网络精度的前提下,有效减少了网络的参数量。其次,设计了一种单解耦头输出结构SDHEAD,通过语义特征提取器生成空间粗糙但语义更丰富的特征图用于分类,通过边界特征提取器生成包含更多边缘信息的高分辨率特征图,以更好地回归物体边界。在CARPK数据集上的大量实验表明,与目前流行的轻量级模型相比,Faster S在准确性和延迟之间实现了更好的平衡。

### 参考文献(References):

[1] LI K, NI W, TOVAR E, et al. Joint flight cruise control and data collection in UAV-aided Internet of Things; an onboard deep reinforcement learning approach [J]. IEEE Internet of Things Journal, 2021, 8(12): 9787-9799.

[2] 岳基隆, 张庆杰, 朱华勇. 微小型四旋翼无人机研究进展及

关键技术浅析[J]. 电光与控制, 2010, 17(10): 46-52.

YUE J L, ZHANG Q J, ZHU H Y. Research progress and key technologies of micro quad-rotor UAVs [J]. Electronics Optics & Control, 2010, 17(10): 46-52. (in Chinese)

[3] 程旭, 宋晨, 史金钢, 等. 基于深度学习的通用目标检测研究综述[J]. 电子学报, 2021, 49(7): 1428-1438.

CHENG X, SONG C, SHI J G, et al. A survey of generic object detection methods based on deep learning [J]. Acta Electronica Sinica, 2021, 49(7): 1428-1438. (in Chinese)

[4] 张荣, 李伟平, 莫同. 深度学习研究综述[J]. 信息与控制, 2018, 47(4): 385-397, 410.

ZHANG R, LI W P, MO T. Review of deep learning [J]. Information and Control, 2018, 47(4): 385-397, 410. (in Chinese)

[5] DIWAN T, ANIRUDH G, TEMBHURNE J V. Object detection using YOLO: challenges, architectural successors, datasets and applications [J]. Multimedia Tools and Applications, 2023, 82(6): 9243-9275.

[6] ZHONG L T, TAN H, LIAO J R. YOLOX-Nano: intelligent and efficient dish recognition system [C]//2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP). Xi'an, China; IEEE, 2022: 1391-1395.

[7] YU G H, CHANG Q Y, LV W Y, et al. PP-PicoDet: a better real-time object detector on mobile devices [EB/OL]. arXiv: 2111.00902 (2021-11-01) [2023-04-10]. <https://arxiv.org/abs/2111.00902>.

[8] 杨小冈, 高凡, 卢瑞涛, 等. 基于改进YOLOv5的轻量化航空目标检测方法[J]. 信息与控制, 2022, 51(3): 361-368.

YANG X G, GAO F, LU R T, et al. Lightweight aerial object detection method based on improved YOLOv5 [J]. Information and Control, 2022, 51(3): 361-368. (in Chinese)

[9] 卢壮. 基于YOLO网络架构的无人机视角下的小目标检测算法研究与应用[D]. 南昌: 南昌大学, 2022.

[10] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA; IEEE, 2017: 936-944.

[11] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA; IEEE, 2018: 8759-8768.

[12] ZHANG S F, CHI C, YAO Y Q, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA; IEEE, 2020: 9756-9765.

[13] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. arXiv: 1704.04861 (2017-04-17) [2023-04-10]. <https://arxiv.org/abs/1704.04861>.

[14] 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). Seattle, WA, USA; IEEE, 2020: 1577-1586.

[15] 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. Salt Lake City, UT, USA; IEEE, 2018: 6848-6856.

(责任编辑: 宋颖韬)