

基于 DTA-FSAF 的无人机小目标检测研究

赵侃,汪慧兰,郭娇娇,王桂丽

(安徽师范大学 物理与电子信息学院,安徽 芜湖 241002)

摘要:随着无人机的应用越来越广泛,基于无人机下的交通场景目标检测的需求也越来越高。但是现有算法在无人机视角下检测精度不高,鲁棒性也不够稳定。为了解决交通场景下无人机视角的车辆和行人的目标检测问题,该文提出 DTA-FSAF 的目标检测网络。首先,将可变卷积融入骨干网络 ResNet-50 中提高 FSAF 网络 (Feature Selective Anchor-Free) 的特征学习能力,采用 PAFPN (Path Aggregation Feature Pyramid Network) 实现多尺度融合,从而提高小目标的检测精度与网络的拟合能力;其次,使用任务对齐检测头减小网络的分类与定位任务在检测小目标时出现的分类与定位任务的错位,从而进一步提高网络的鲁棒性;最后,通过调整 IoU 损失提高网络整体的检测效果。通过在无人机数据集 VisDrone 上进行实验和分析比较可知,相比于其他网络,在不同的交通场景下,DTA-FSAF 网络在满足实时性需求的同时检测精度达到了 41.3%。相比于 FSAF 网络提升了 19.6%。通过实验证明改进算法能有效地在各种复杂交通场景下完成对行人和车辆的目标检测。

关键词:目标检测;小目标检测;Feature Selective Anchor-Free;无人机;标签分配

中图分类号:TP391.41;V19

文献标识码:A

文章编号:1673-629X(2024)04-0101-08

doi:10.20165/j.cnki.ISSN1673-629X.2024.0016

Research on Small Object Detection of UAV Based on DTA-FSAF

ZHAO Kan, WANG Hui-lan, GUO Jiao-jiao, WANG Gui-li

(School of Physics and Electronic Information, Anhui Normal University, Wuhu 241002, China)

Abstract: With the increasing application of UAV, the demand for object detection in traffic scenes based on UAV is also increasing. However, existing algorithms have low detection accuracy and insufficient robustness from the perspective of UAV. In order to effectively solve the object detection problem of vehicles and pedestrians from the perspective of UAV in traffic scenes, we propose the DTA-FSAF network for object detection. Firstly, deformable convolution is integrated into the backbone network ResNet-50 to improve the feature learning ability of the FSAF (Feature Selective Anchor-Free) network, and PAFPN (Path Aggregation Feature Pyramid Network) is used for multi-scale fusion to improve the detection accuracy of small object and the fitting ability of the network. Secondly, task alignment detection heads are used to reduce the misalignment of classification and positioning tasks in detecting small object, thus further improving the robustness of the network. Finally, the IoU loss is adjusted to improve the overall detection performance of the network. Through experiments and analysis on the drone dataset VisDrone, it is known that compared with other networks, the DTA-FSAF network can achieve a detection accuracy of 41.3% in different traffic scenes while meeting real-time requirements. This is a 19.6% improvement over the FSAF network. The experimental results demonstrate that the improved algorithm can effectively complete the object detection of pedestrians and vehicles in various complex traffic scenes.

Key words: object detection; tiny object detection; Feature Selective Anchor-Free; UAV; label assignment

0 引言

近年来,随着无人机制造成本的下降,无人机在生活中的使用变得愈发普遍。无人机在交通监控、物流配送等领域的应用越来越广泛,但高空拍摄带来了多种挑战,如目标大小、形状、姿态、遮挡等变化,以及普

通尺寸目标在图像中变得更小,使得无人机小目标检测变得更加困难。因此,在一些复杂场景中实现高效、准确、鲁棒的小目标检测是一个极具挑战性的问题,对目标检测算法提出了更高的要求^[1-3]。基于深度学习的目标检测框架主要由以 Faster-RCNN^[4], Cascade R-

收稿日期:2023-07-01

修回日期:2023-11-02

基金项目:安徽省自然科学基金(1708085QF133);安徽师范大学创新基金项目(2018XJJ100);安徽省智能机器人信息融合与控制工程实验室资助(IFCIR2020004)

作者简介:赵侃(1998-),男,硕士生,通讯作者,研究方向为深度学习、计算机视觉、目标检测;汪慧兰(1978-),女,副教授,研究方向为深度学习、行人检测。

CNN^[5] 为代表的二阶段检测网络和以 YOLO^[6], FCOS^[7] 为代表的单阶段检测网络为主。二阶段网络的精确度较高,但是速度较慢,很难达到实时性的检测要求。单阶段网络检测速度快,精确度虽不及二阶段网络,但有着端到端的检测优势。单阶段网络又分为 Anchor-Base 网络和 Anchor-Free 网络,其不同主要在于是否采用预设锚框回归。当前基于深度学习的目标检测算法中,针对大型目标的检测方法趋于成熟,但对于无人机视角下的小目标检测仍然存在巨大的进步空间。Zhan 等人^[8]通过在 YOLOV5 中添加一个小目标检测层来提高检测精确度,但是过深的网络层数使得检测速度明显下降。Song 等人^[9]通过融合不同深度的特征层构建特征金字塔,构建基于多尺度深度特征学习网络的目标检测器,虽然强化了语义信息和特征信息的融合,但降低了检测速度。Zhu 等人^[10]将 Vision Transformer 结构的检测头应用到 YOLOV5 来提高小目标检测效果,虽然效果得到了显著提高,但是其模型参数过大,使得网络的推理速度较慢,同时,由于 VisDrone 数据集中部分类别的样本较少,其针对部分小样本的目标检测效果并不好。

针对小目标检测,不仅要提高网络对目标的特征提取融合能力,同时针对小样本分配也不能忽视。采用不同的样本分配策略能直接影响网络的学习和拟合的效果。在小目标检测中,由于检测时产生的负样本远远多于正样本,基于 Anchor-Base 结构的 RetinaNet 网络使用 Focal Loss 来提高小目标检测能力^[11],但采用预设锚框来预测目标大小和形状导致效果不佳。以 FCOS (Fully Convolutional One-Stage) 网络为主的 Anchor-Free 结构网络可以有效避免这个问题,Anchor-Free 结构网络不需要预设一些先验框,直接预测目标的位置和大小,可以有效避免锚框尺寸不当导致的误检和漏检问题。但是在检测小目标时,因为过多的负样本依然会导致正负样本不平衡,从而使检测效果不佳。为了进一步降低 Anchor-Free 网络的漏检率和误检率,不同的 Anchor-Free 网络采用了不同的样本分配方法,ATSS (Adaptive Training Sample Selection) 网络使用中心度分支增强锚点预测的分类分数,并为相应锚点的定位损失分配更大的权重^[12]。FoveaBox 网

络则将预定义中心区域内的锚视为正样本^[13]。这些启发式设计取得了很好的效果,但分类和定位任务的独立以及大量与任务无关的样本分配使得检测效果仍然不佳。TOOD 则开创性地使用任务损失对齐将分类与定位损失进行损失对齐来平衡正负样本不平衡的问题^[14]。

针对上述问题,该文采用基于 Anchor-Free 结构的 FSAF 网络 (Feature Selective Anchor-Free Network)。FSAF 网络通过自适应特征选择为每一个样本选择合适的特征提取层,在训练过程中,根据 IoU 重叠匹配的方法,将不同尺寸的物体依据其与 FPN 每一层的适配程度,分配到不同分辨率的层上进行学习^[15]。这样能充分适配各层感受野和空间信息到最符合样本大小的特征提取层中,提高不同尺寸目标检测的准确性。为了提高小目标的检测精度,提升网络鲁棒性,该文通过改进 FSAF 网络的骨干网络和 FPN 结构来提高网络针对小目标的特征学习能力。其次,采用任务对齐样本分配策略来解决正负样本之间巨大的不平衡问题。

1 FSAF 网络结构

FSAF 网络的主干网络采用 ResNet-50,连接部分采用了 FPN 结构,检测头则分别采用了 Focal loss 和 IoU loss^[16]来计算分类损失和定位损失。

FSAF 模块的主要思路是在每个特征图中自适应地选择最有用的特征点,并将它们组合成最终的检测结果。这样做的好处是可以避免 Anchor-Based 方法中存在的大量超出目标范围的框,并且可以减少模型的复杂度,提高检测速度和精度,如图 1 所示。具体选择过程如下:每个特征层计算分类损失 (classification loss) 和回归损失 (box regression loss);给定一个实例 I ,定义在特征金字塔上 l 层的输出 P_l 上的分类损失和回归损失为 $L_{FL}^l(l)$ 和 $L_{IoU}^l(l)$,通过对有效框区域 b_e^l 的 Focal loss 和 IoU loss 分别进行平均而获得:

$$L_{FL}^l(l) = \frac{1}{N(b_e^l)} \sum_{(i,j \in b_e^l)} FL(l, i, j) \quad (1)$$

$$L_{IoU}^l(l) = \frac{1}{N(b_e^l)} \sum_{(i,j \in b_e^l)} IoU(l, i, j) \quad (2)$$

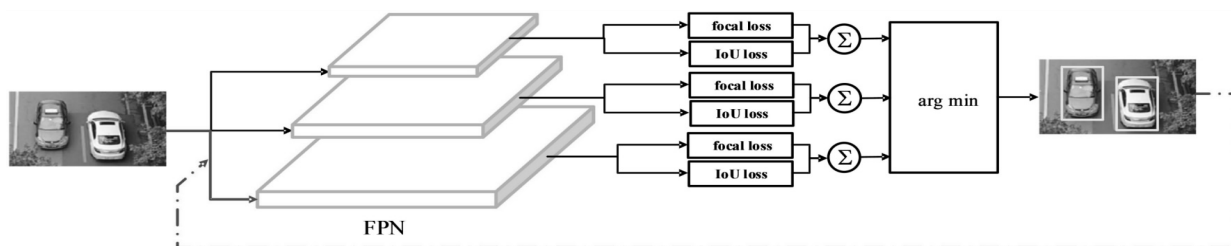


图 1 FSAF 网络基本结构

在所有特征层中选择损失最小的作为梯度进行反向传播。

$$l^* = \arg \min_l L_{\text{FL}}^l(l) + L_{\text{IoU}}^l(l) \quad (3)$$

选择特征点的过程是通过计算每个特征点与目标之间的相似度来实现的,相似度高的特征点会被优先选择。这个相似度计算是基于卷积操作实现的,可以在模型中进行端到端训练。

2 DTA-FSAF 网络

为了提高 FSAF 网络对特征的学习能力,该文从两个方面对 FSAF 网络模型进行改进。首先,将可变形卷积(Deformable Convolution v2, DCNv2)应用在 ResNet-50 网络中以提高骨干网络中针对小目标的特征提取能力。然后,采用 PAFPN 将提取的特征进行多尺度融合,以实现多尺度目标的检测,提高小目标的检测精度,同时还能提高网络的拟合能力,弥补 Anchor-Free 网络特征学习能力不足的缺陷。

2.1 可变形卷积

DCNv2 是一种基于卷积神经网络的卷积操作,可以动态调整卷积核的形状来适应更加复杂的物体形状

和空间变化,提高模型在目标检测任务中的准确性。相比传统卷积操作,DCNv2 能更好地学习目标的轮廓特征,有效帮助 Anchor-Free 网络提高模型的建模能力^[17]。

普通的卷积如图 2(a) 所示,以 3×3 卷积为例,对于每个输出 $y(p_0)$, $p_0 = (0, 0)$,都要从输入特征图 x 上面采样 9 个位置,这 9 个位置都在中心位置 $x(p_0)$ 向四周扩散, $(-1, -1)$ 代表 $x(p_0)$ 的左上角, $(1, 1)$ 代表 $x(p_0)$ 的右下角。采样区域 $R = \{(-1, -1), (-1, 0), \dots, (1, 1)\}$ 。因此,普通卷积的输出为:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \bullet x(p_0 + p_n) \quad (4)$$

其中, p_n 是采样网格中的第 n 个点, $w(p_n)$ 为对应的卷积核权重系数。DCNv2 如图 2(b) 所示,由可变形卷积核、偏移量学习模块和卷积计算模块构成。可变形卷积在普通卷积的基础上添加偏移量学习模块 Δp_n ,通过对输入数据进行卷积运算,得到一组偏移量 $x(\Delta p_n)$ 。这些偏移量可以调整卷积核的形状,适应不同形状的特征。

$$y(p_0) = \sum_{p_n \in R} w(p_n) \bullet (p_0 + p_n + \Delta p_n) \quad (5)$$

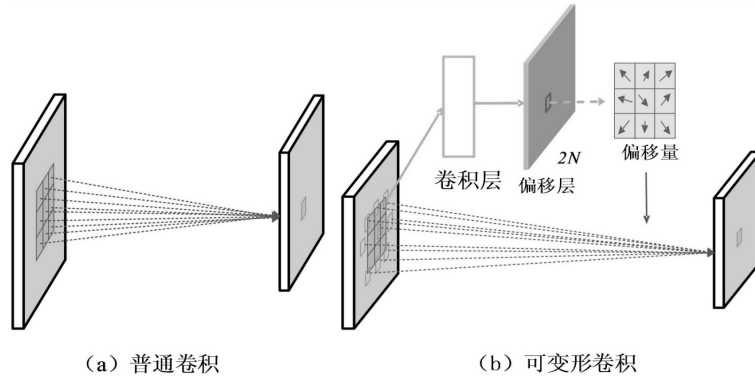


图2 卷积对比

对于输入的特征图,假设原来的卷积操作是 3×3 的,那么为了学习偏移量(offset),定义另外一个 3×3 的卷积层,输出其实就是只用常规卷积输出特征图的大小,但是通道数等于 2×3^2 (分别表示 x 与 y 方向的偏移,3为卷积核的尺寸)。

常规 ResNet-50 的残差模块如图 3(a) 所示,其瓶颈模块包括三个卷积层,分别是 1×1 卷积层、 3×3 卷积层和 1×1 卷积层^[18]。其中, 3×3 卷积层作为该模块中的特征提取层,将 DCNv2 添加入 ResNet-50 层中,如图 3(b) 所示,即瓶颈模块中的 3×3 卷积层。

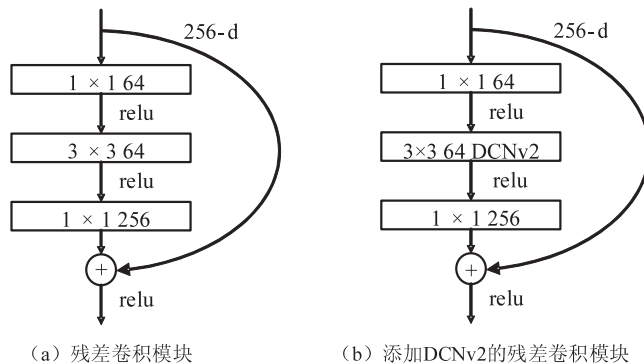


图3 残差模块对比

使用 DCNv2 可以对卷积核进行微小的位置和形状调整,更好地感知小目标的位置和形状,捕捉小目标的特征,提高模型对背景噪声、光照变化等因素的鲁棒性,从而提高对小目标特征的学习能力和检测率。因此,DCNv2 在特征提取骨干网络 ResNet-50 上的应用可以提高模型对小目标的检测能力,进而提高目标检测的准确性和鲁棒性。

2.2 改进特征金字塔网络

PAFPN 的结构由两个部分组成: PAN 和 FPN。FPN 模块用于提取不同分辨率的特征图,而 PAN 模块则用于将这些特征图进行多尺度融合。PAFPN 将高层的特征信息先通过上采样的方式进行传递融合,再通过下采样融合方式得到预测的特征图,最终输出由 3 个特征层组成的结果,能将顶层的特征信息与底层的语义信息充分融合^[19]。

PA 模块计算公式为:

$$F_{PA}(x) = \sum_{i=1}^m \alpha_i F_i(x) \quad (6)$$

其中, F_{PA} 是 PA 模块输出的特征图, $F_i(x)$ 是输入的不同分辨率的特征图, m 是特征图的数量, α_i 是特征图的权重。其中 PA 模块的权重计算公式为:

$$\alpha_i = \frac{e^{w_i}}{\sum_{j=1}^m e^{w_j}} \quad (7)$$

其中, w_i 是根据特征图的大小和分辨率计算出的权重系数。

最终得出 PAFPN 结构的公式:

$$F_{PAFPN}(x) = F_{PA}(\sum F_{P_i}(x)) \quad (8)$$

其中, $F_{P_i}(x)$ 是 FPN 模块输出的第 i 层特征图, $F_{PAFPN}(x)$ 是 PAFPN 模块输出的特征图。

PAFPN 结构将顶层的特征信息与底层的语义信息充分融合,增强网络的特征学习能力和特征表达能力。其相对于 FPN 增加了两个高分辨率的特征层同时添加跨层连接,有效地避免了重复采样和信息丢失。

2.3 任务对齐检测头

任务对齐检测头(Task-Aligned Head, TAH)将检测头分解为检测头分支和任务对齐分支,前者预测目标的位置和类别,后者根据任务要求对检测结果进行损失对齐处理,这样能增强分类与定位之间的相互作用,弥补 Anchor-Free 网络容易出现假阳性的缺点,提升网络整体的鲁棒性。

常规网络检测头如图 4 所示。分类任务和定位任务之间互相独立,不利于网络对目标的学习。

TAH 的结构如图 5 所示。这种设计不仅方便了任务的交互,而且为这两个任务提供了多层次的特征和多尺度的有效感受野^[14]。

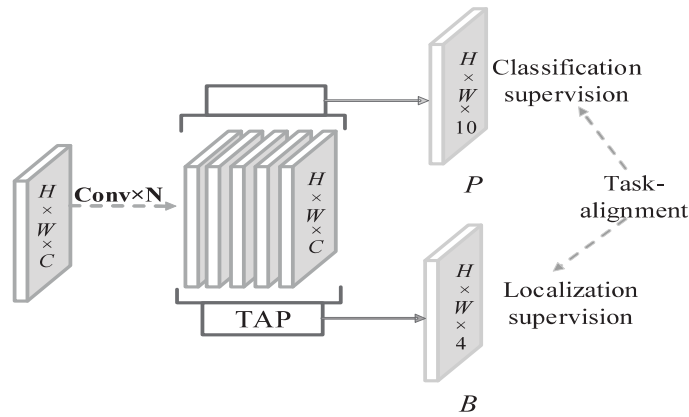


图 4 常规检测头

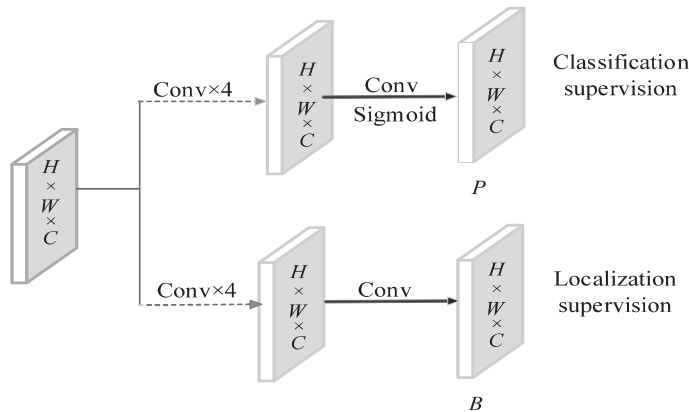


图 5 TAH 检测头

形式上,FPN 特征 $X_k^{\text{fpn}} \in R^{H \times W \times C}$, 其中 H , W 和 C 分别表示高度、宽度和通道数。特征提取器使用 N 个带有激活函数的连续卷积层来计算任务交互特征:

$$X_k^{\text{inter}} = \begin{cases} \delta(\text{conv}_k(X_k^{\text{fpn}})), k = 1 \\ \delta(\text{conv}_k(X_{(k-1)}^{\text{inter}})), k > 1 \end{cases} \quad (9)$$

其中, $\forall k \in \{1, 2, \dots, N\}$, conv_k 和 δ 分别指第 k 个卷积层和 ReLU 函数。因此,利用 head 的单个分支从 FPN 特征中提取丰富的多尺度特征。然后,将计算出的任务交互特征送入两个 TAP 进行分类定位。

2.4 DTA-FSAF 网络

DTA-FSAF(Deformable and Task Aligned-FSAF) 网络如图 6 所示,用于复杂场景下的无人机目标检测。首先,针对 FSAF 的骨干网络进行修改,使用 DCNv2 替换 ResNet-50 中残差模块的 3×3 卷积层,并将 FPN 替换为 PAFPN,在增强网络对目标的特征描述能力的同时考虑了特征提取时将语义和特征信息进行融合,更好地适应不同大小的目标。其次,使用任务对齐思

想的 TAH 作为分类和定位检测头,使网络在训练中得到逐步统一分类和定位预测框,提高了网络的准确度和鲁棒性,同时减小了误检率和假阳性的概率。

在对目标预测框质量的判别标准上,原始 IoU loss 仅计算真实框和预测框的交并比,没有考虑目标框之间的其他因素。小目标的检测具有复杂性和密集性,该文进一步改进 IoU loss,提高网络在复杂场景下对小目标的检测效果。GIoU 是对 IoU 的改进,GIoU (Generalized IoU) 不仅考虑了交集和并集的面积,还考虑了两个框之间的距离,计算时在 IoU 的基础上将预测框和真实框的中心点距离加入分母中,用来衡量位置偏差。这样,GIoU 就能够更加全面地评估预测框和真实框之间的匹配程度^[20]。CIoU 对 GIoU 进一步做出了改进,CIoU(Complete IoU)将目标框的中心点距离和宽高比差异添加到 IoU 计算中,更全面地衡量两个目标框之间的相似度^[21]。

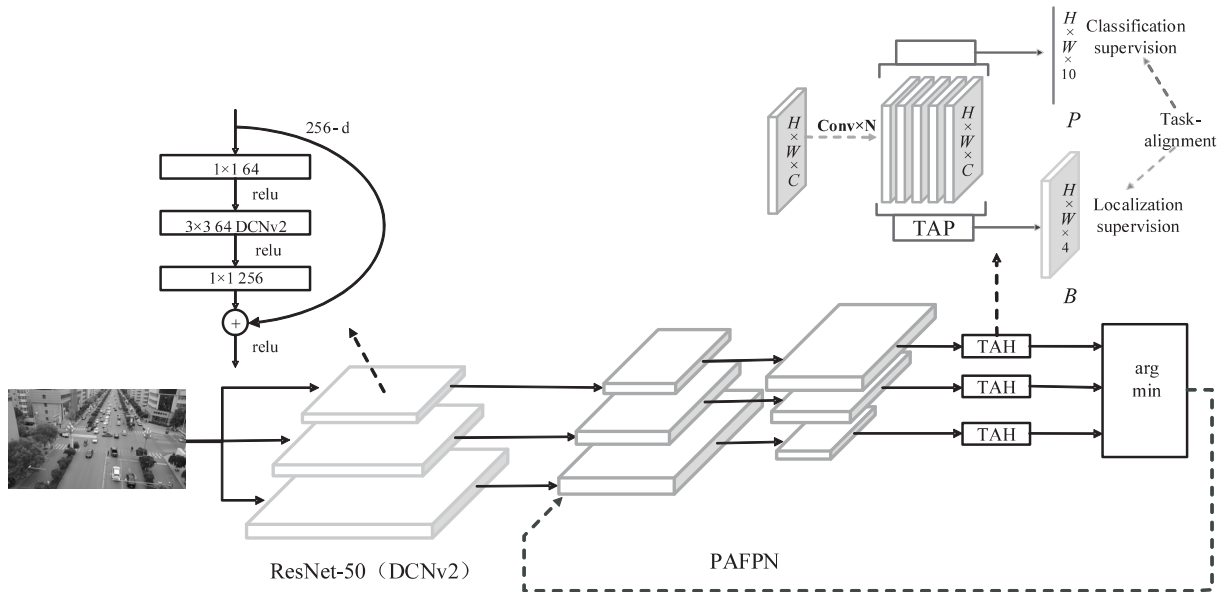


图6 DTA-FSAF 网络结构

该文对比了上述三种 IoU 检测方法,如表 1,最终采用精确度更高,评估更全面的 CIoU 作为网络的定位损失函数。

表1 不同 IoU 对网络的影响

Model	AP ₅₀	AP _S
DTA-FSAF+IoU	40.6	13.8
DTA-FSAF+GoU	40.9	14.1
DTA-FSAF+CoU	41.3	14.5

3 实验结果与分析

实验采用由天津大学机器学习团队开发的

VisDrone 数据集,该数据集是一个公开的、大规模的、高质量的无人机数据集,由 7 019 张静态图像组成,包含由各种无人机摄像头拍摄,在不同场景、不同天气和照明条件下收集的各种行人和车辆序列。数据集划分为由 6 471 张图片组成的训练集和 548 张图片组成的验证集,总共分为十个类别:行人、人、汽车、面包车、公共汽车、卡车、摩托车、自行车、雨棚三轮车和三轮车。其中行人和汽车样本数量最多,分别有 79 937 和 144 866 个样本,三轮车和雨棚三轮车的样本数量最少,仅有 4 812 和 3 246 个样本^[22]。

实验配置:实验室使用的系统为 Ubuntu 20.4,实验环境为 PyTorch 深度学习框架和两张 NVIDIA 3080TITAN 显卡。

3.1 AP 指标

AP 指标指的是平均精度 (Average Precision)。它是衡量目标检测网络性能的一种指标。AP 值越高,表示检测器检测出的物体越准确。AP 值的计算方法是首先在不同的置信度阈值下计算出精确率 (Precision) 和召回率 (Recall),再计算精确率和召回率所围成曲线的面积得出 AP 值。其中 AP_{50} 指标是指当交并比 IoU (Intersection over Union) 大于等于 0.5 时的 AP 值。这个指标通常用于评估目标检测算法在检测目标时的精度。

3.2 对比实验和消融实验

为验证 DTA-FSAF 网络的有效性,通过消融实验对 DTA-FSAF 网络中添加的每个模块进行研究,从而确定不同模块带来的不同影响。通过对比实验,同时对 RetinaNet、FCOS 等经典网络以及 YOLOX^[23]、FoveaBox、TPH-YOLO、MSF-YOLO、YOLOv8 等先进网络在 VisDrone 数据集上的表现。

表 2 DTA-FSAF 网络消融实验

Model	AP/%	FPS/s
FSAF Base line	17.6	28.5
+TAH	20.4(+2.8)	27.7
+PAFPN	22.6(+2.2)	26.6
+DCNv2	23.9(+1.3)	25.5
+CIoU	24.2(+0.3)	25.1

从表 2 可以看出,FSAF 网络在添加了 TAH 之后 AP 得到了显著提升,充分证明了在网络学习能力不变的情况下,合适的样本分配策略使网络能在复杂的交通场景依然有较高的检测精度。在此基础上,将 ResNet-50 中的普通卷积层替换为可变形卷积层,替换 FPN 为 PAFPN,强化网络的特征学习能力和多尺度融合能力,高效地将特征图的语义信息和特征信息进

行融合,提高网络的特征表达能力。取得了 23.9 的 AP。最后改进 IoU loss,采用 CIoU 作为评价指标,能更加全面地评估复杂场景下的检测效果,对小目标检测效果的评估也更加精确,最终取得了 24.2 的 AP,同时 FPS 达到 25.1 帧/秒,参数量仅有 15 M 左右。

表 3 不同网络对小样本类别检测效果

模型	遮阳棚三轮车/%	三轮车/%	参数量/M
MSF-YOLO ^[10]	9.8	13.6	16
TPH-YOLO ^[11]	11.8	15.1	22
YOLOv8	11.1	17.5	14
FSAF	9.3	12.8	13
DTA-FSAF	14.1	19.3	15

由表 3 可以看出,MSF-YOLO 和 TPH-YOLO 在样本数量最少的雨棚三轮车和三轮车这两个类别上的检测效果不佳,尤其是在样本最少的雨棚三轮车上,MSF-YOLO 的精度只有 10.8,而 TPH-YOLO 的精度虽然有 11.8,但过大的参数量减少了网络的推理速度,很难达到实时性的要求。YOLOv8 虽然使用了混合 IoU,但仍然没有很好解决小样本检测效果较低的问题。其检测雨棚三轮车和三轮车的精度也仅有 11.1 和 17.5。DTA-FSAF 网络在采用了任务对齐检测头 (TAP) 后,针对小样本的检测效果大大提高,在样本最少的两类目标中均有较高的精度,分别达到了 14.1 和 19.3。使得网络对小样本类别检测的效果较 MSF-YOLO 在小样本检测上提升了 28%。并且比 YOLOv8 提升了 22%。

表 4 为 DTA-FSAF 网络与不同单阶段网络在 VisDrone2021 验证集上的检测结果。该表列出了不同模型的指标,包括使用的骨干网络、平均精度 (AP) 和不同 IoU 阈值下的检测精度 (AP_{50} 、 AP_{75}) 以及不同尺度范围内的精度 (AP_s 、 AP_m 、 AP_L)。

表 4 VisDrone 验证集下各网络对比 %

Method	Backbone	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_L
RetinaNet ^[11]	ResNet-50	17.8	31.9	17.6	8.8	28.5	36.1
YOLOv5	CSPDarknet-53	19.1	33.9	18.7	\	\	\
FCOS ^[7]	ResNet-50	17.6	31.4	17.4	8.6	28.1	35.7
CenterNet ^[24]	ResNet-50	20.4	34.3	21.7	12.6	30.6	38.1
YOLOX ^[23]	CSPDarknet-53	21.6	35.7	22.4	12.8	31.9	38.8
FoveaBox ^[13]	ResNet-50	18.2	32.0	18.2	8.9	29.7	36.3
ATSS ^[12]	ResNet-50	20.7	36.1	20.9	10.9	33.3	41.2
FSAF ^[15]	ResNet-50	17.6	33.2	16.6	9.9	27.0	34.1
DTA-FSAF	ResNet-50+DCNv2	24.2	41.3	24.6	14.5	36.2	45.1

由表4数据看出,在检测无人机视角下的行人和车辆时,基于 Anchor-Base 的 RetinaNet 网络采用了 Focal loss,在检测普通大小的目标时效果较好, AP_M 达到了 28.5,但是由于其采用预设框检测,在检测小目标时精度不足,仅有 8.8。FCOS 和 CenterNet 都是基于 Anchor-Free 网络,但是 FCOS 落在真实框中心一定范围之内即为正样本的样本分配方案使得其在检测小目标和复杂背景目标时效果较差, AP_S 仅有 8.6, AP_M 为 28.1,对比其他网络均较低。CenterNet 采用了高斯分布的方法,将中心点作为高斯分布的中心,利用高斯核算法将关键点分布在特征图上,取得了不错的效果^[24]。 AP 达到了 20.4,在不同尺寸的目标检测上均取得较好的精度。FoveaBox 和 ATSS 均为在样本分配算法上做出改进的 Anchor-Free 网络,但 ATSS 能自适应根据对象的统计特征自动选择正负样本,在检测小目标和复杂背景时具有良好的鲁棒性, AP_S 达到了 10.9, AP_M 达到 33.3, AP_L 达到 41.2。证明了合适的样本分配方法能有效提高小目标的检测精度和复杂场景下网络的鲁棒性。

DTA-FSAF 网络从两个方面针对小目标和复杂场景检测进行改进。首先提高了网络对特征的学习能力,其次改进了样本分配策略,让 DTA-FSAF 网络能更高效地学习小目标的特征,也能在复杂场景下更稳定的检测。DTA-FSAF 对比 FSAF, AP 提升了 27%,达到了 24.2。在小目标检测上提升了 31%, AP_S 达到了 14.5。在复杂场景下的目标检测精度 AP_M 达到 36.2, AP_L 达到 45.1。充分证明了 DTA-FSAF 网络具有更高的精确度和良好的稳定性。

3.3 实验结果可视化展示

为了直观展示提出的算法在复杂场景下对小型目标的检测能力,在模糊,密集,暗光,过曝这四个有代表性的复杂场景下展示训练结果,如图7所示。从效果图可以看出,改进后的 DTA-FSAF 网络在密集和遮挡环境下依然有着不错的精确度,并且在模糊,过曝等极端环境下比其他网络有着更少的误检。因此,DTA-FSAF 网络在拥有较好的检测性能的同时,鲁棒性也得到了提高,让网络能充分适应各种复杂场景。

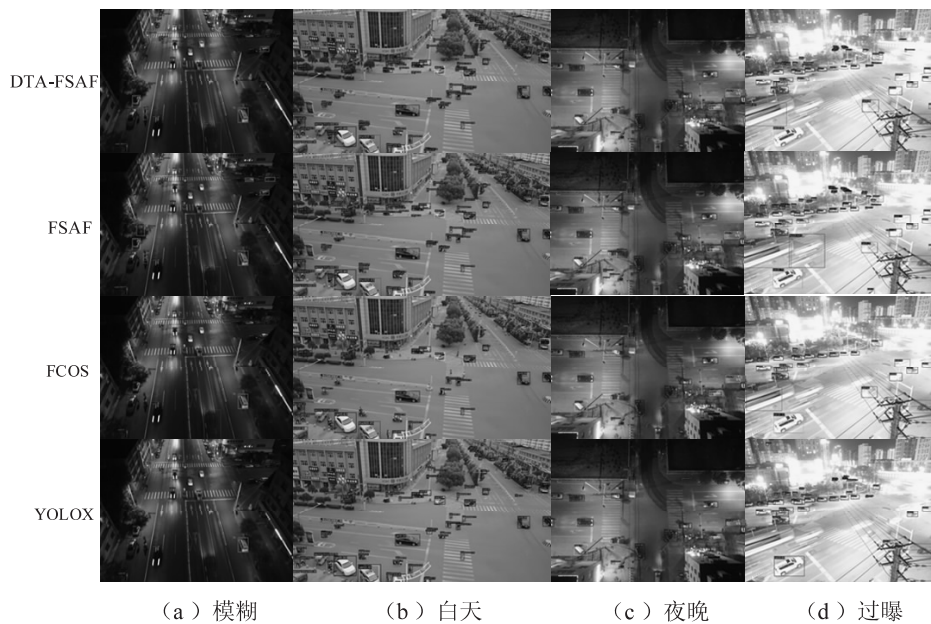


图7 各网络在不同场景下的对比

4 结束语

该文提出的 DTA-FSAF 网络,以 FSAF 网络为基础,针对小目标和复杂场景下的目标检测做出改进。首先使用可变形卷积和 PAFPN,提高网络的特征学习能力、特征融合能力和特征表达能力。其次,使用任务对齐检测头对齐分类和定位损失,防止检测时出现空间错位。最后,使用 CIoU 作为评估标准,更全面、更准确地对检测效果做出评估。

最终,DTA-FSAF 不仅在检测精度上对比 YOLOX 等网络有提升,同时在小样本检测上对比其

他先进网络也有一定的提升。弥补了小目标检测中针对小样本检测精度不佳的问题。提高了网络在无人机视角下小目标检测的鲁棒性和精确度,为工业应用提供了良好的参考。但无人机体积较小,针对小目标检测的运算量较大,未来将进一步研究在维持高检测精度的同时减少网络的参数量以及提高网络的推理速度。

参考文献:

- [1] 李科岑,王晓强,林 浩,等.深度学习中的单阶段小目标检测方法综述[J]. 计算机科学与探索,2022,16(1):41-

- 58.
- [2] 冷佳旭,莫梦竟成,周应华,等. 无人机视角下的目标检测研究进展[J]. 中国图象图形学报, 2023, 28(9): 2563–2586.
- [3] KISANTAL M, WOJNA Z, MURAWSKI J, et al. Augmentation for small object detection[J]. arXiv:1902.07296, 2019.
- [4] REN S Q, HE K M, GIRSHICK R, et al. Faster RCNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149.
- [5] CAI Z, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 6154–6162.
- [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. Las Vegas: IEEE, 2016: 779–788.
- [7] TIAN Z, SHEN C, CHEN H, et al. Fcos: fully convolutional one-stage object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. Kunming: IEEE, 2019: 9627–9636.
- [8] ZHAN W, SUN C, WANG M, et al. An improved Yolov5 real-time detection method for small objects captured by UAV[J]. Soft Computing, 2022, 26(1): 361–373.
- [9] SONG Z, ZHANG Y, LIU Y, et al. MSFYOLO: feature fusion-based detection for small objects[J]. IEEE Latin America Transactions, 2022, 20(5): 823–830.
- [10] ZHU X, LYU S, 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. Montreal: IEEE, 2021: 2778–2788.
- [11] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. Venice: IEEE, 2017: 2980–2988.
- [12] ZHANG S, CHI C, YAO Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle: IEEE, 2020: 9759–9768.
- [13] KONG T, SUN F, LIU H, et al. Foveabox: beyond anchor-based object detection[J]. IEEE Transactions on Image Processing, 2020, 29: 7389–7398.
- [14] FENG C, ZHONG Y, GAO Y, et al. Toood: task-aligned one-stage object detection[C]//2021 IEEE/CVF international conference on computer vision (ICCV). Montreal: IEEE, 2021: 3490–3499.
- [15] ZHU C, HE Y, SAVVIDES M. Feature selective anchor-free module for single-shot object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Long Beach: IEEE, 2019: 840–849.
- [16] YU J, JIANG Y, WANG Z, et al. Unitbox: an advanced object detection network[C]//Proceedings of the 24th ACM international conference on multimedia. Ottawa: ACM, 2016: 516–520.
- [17] ZHU X, HU H, LIN S, et al. Deformable convnets v2: more deformable, better results[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Long Beach: IEEE, 2019: 9308–9316.
- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016: 770–778.
- [19] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 8759–8768.
- [20] REZATOFIGHI 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. Long Beach: IEEE, 2019: 658–666.
- [21] 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. New York: AAAI, 2020: 12993–13000.
- [22] LIU Z, HE Z, WANG L, et al. VisDrone-CC2021: the vision meets drone crowd counting challenge results[C]//Proceedings of the IEEE/CVF international conference on computer vision. Montreal: IEEE, 2021: 2830–2838.
- [23] GE Z, LIU S, WANG F, et al. Yolox: exceeding yolo series in 2021[J]. arXiv:2107.08430, 2021.
- [24] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as points[J]. arXiv:1904.07850, 2019.