

一种 Anchor Free 的无人机检测方法

耿文^{1,2}, 孙涵^{1,2}

(1. 南京航空航天大学 计算机科学与技术学院/人工智能学院, 江苏 南京 211106;
2. 南京航空航天大学 模式分析与机器智能工信部重点实验室, 江苏 南京 211106)

摘要:目前几乎所有最先进的目标检测方法都依赖于预先定义的 anchor,但是由于 anchor 的存在,网络会增加与 anchor 相关的计算开销,而且在现实世界中,目标尺度多种多样,比如地对空拍摄的无人机目标,预先定义的 anchor 不可能穷举所有的无人机尺度。因此,该文抛弃预先定义的 anchor,采用 anchor free 的方法来检测无人机。该 anchor free 方法以目标中心区域的像素为训练样本来逐像素预测边界框的偏移量。针对地对空拍摄的无人机目标大部分都是小尺度的情况,采用高分辨率网络 HRNet 作为主干网络来提取小目标细粒度的特征,从而提高小目标检测的精度。相对于基于 anchor 的目标检测方法,该网络框架简单、灵活,并且可以自适应地预测无人机目标的边界框。在自己设计的无人机数据集上,该方法获得了更高的召回率和精确率。

关键词:目标检测; anchor free; 无人机检测; 小目标检测; 逐像素预测; HRNet

中图分类号: TP391.41

文献标识码: A

文章编号: 1673-629X(2021)01-0054-07

doi: 10.3969/j.issn.1673-629X.2021.01.010

An Anchor Free Method for Detecting Drones

GENG Wen^{1,2}, SUN Han^{1,2}

(1. School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;

2. MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: At present, almost all the most advanced object detection methods rely on a pre-defined anchor. However, due to the existence of anchor, the network will increase computation and memory footprint, and object scales are varied in the real world, anchor cannot exhaustive all drone scales. Therefore, we abandon a pre-defined anchor and adopt anchor free method to detect drones, which uses the pixels in the center area of the object as training samples to predict the offset of the bounding box. At the same time, in order to solve the problem that most drones are small, a high-resolution network HRNet is used as the backbone network to extract fine-grained features of small objects, thereby improving the accuracy of small object detection. Compared with the anchor-based object detectors, this network framework is simple, flexible, and can adaptively predict the bounding box of the drone. The proposed anchor free method achieves higher precision and recall on the drone dataset designed by ourselves.

Key words: object detection; anchor free; drone detection; small object detection; per-pixel detection; HRNet

0 引言

目前,无人机已经被广泛应用于商业和娱乐等领域,它在带来极大便利的同时,也在威胁着个人隐私,甚至会威胁到民航的安全,因此对于无人机的检测变得越来越重要。

目前的无人机检测技术都是建立在通用目标检测算法基础之上的。起初,通用目标检测算法采用滑动

窗口的方式来预测目标的位置。随着深度学习的崛起,这种自动从数据中学习特征表示的方式被广泛应用到目标检测中。R-CNN^[1]和 Fast R-CNN^[2]通过区域建议生成来提取感兴趣目标区域。之后,基于 anchor 的 RPN^[3]很快取代了区域建议生成这种方式。由此,anchor 被广泛应用于各种目标检测应用中。目前主流的目标检测算法,比如 Faster R-CNN^[3]、SSD^[4]

收稿日期: 2019-12-23

修回日期: 2020-04-24

基金项目: 国防科技创新特区项目(保密)

作者简介: 耿文(1992-),女,硕士,CCF会员(89147G),研究方向为图像处理和计算机视觉;孙涵,博士,副教授,研究方向为数字图像处理、计算机视觉、模式识别。

和 YOLOv2^[5] 都依赖于一组预先定义的 anchor。

尽管这些基于 anchor 的目标检测算法取得了优异的成绩,但是仍然存在一些问题:

(1) 像文献[3,6]中的目标检测算法的性能与 anchor 的尺寸、长宽比和数量有着直接的关系。如果预先定义的 anchor 参数不合适,这会极大地影响目标检测算法的性能。

(2) anchor 的尺寸和长宽比是固定的,因此目标检测算法在处理形状变换比较大的目标时依旧会遇到困难,特别是对于小目标。预先定义的 anchor 还限制了目标检测算法的泛化能力^[7]。

(3) 基于 anchor 的目标检测算法一般会采用密集 anchor 的方式来获得比较高的召回率。比如特征金字塔网络(FPN^[8])有超过 18 万个 anchor,在训练过程中,大部分的 anchor 都被标记为负样本,这会加重正负样本的不平衡性。

(4) 在训练时,需要计算所有 anchor 与真实框的 IoU 得分,过多的 anchor 会显著增加计算量和内存占用。

为了解决以上问题,最直接的办法就是抛弃这些预先定义的 anchor,采用 anchor free 的方法来检测目标。一阶段目标检测算法 YOLOv1^[9] 是最早的 anchor free 检测方法,它是在每个网格上直接预测两个候选边界框的参数。目前,越来越多的 anchor free 检测方法被提出。CornerNet^[10] 预测的是两组角点(左上角和右下角)。ExtremeNet^[11] 是通过预测极值点和中心点来检测目标的。FSAF^[12] 添加了两个卷积层,这两个卷积层分别负责 anchor free 方法的目标分类和边界框回归。FCOS^[13] 为了剔除远离目标中心的低质量检测框,提出了中心度分支。

1 相关工作

1.1 目标检测

起初,R-CNN^[1] 和 Fast R-CNN^[2] 通过区域建议生成方法来提取感兴趣目标区域。接着,基于 anchor 的 Region Proposal Network^[3] (RPN) 很快取代了区域建议生成方法。由此,anchor 被广泛应用于各种目标检测框架中。

目前基于 anchor 的检测器主要分为两类:两阶段和一阶段。两阶段的目标检测方法主要有 Faster R-CNN^[3]、SPP-Net^[14] 和 HyperNet^[15]。这些方法首先生成一系列的目标候选边界框,然后使用卷积神经网络去分类这些目标。一阶段的目标检测方法主要有 SSD^[4] 和 YOLOv2^[5]。与两阶段目标检测方法相比较,一阶段目标检测方法直接将目标定位问题转化为目标回归问题。

为了改善 anchor 的能力,目前也有许多成功的尝试。MetaAnchor^[7] 可以从任意预先定义的 anchor 中动态生成 anchor 函数。Guided-Anchoring^[16] 联合预测目标中心位置以及不同位置上可能存在的目标尺度和长宽比,以此来动态地检测目标。

解决 anchor 产生的问题最直接的办法其实是抛弃这些预先定义的 anchor,采用 anchor free 的方法。一阶段目标检测器 YOLOv1^[9] 是最早的 anchor free 检测方法。后来,anchor free 的目标检测算法逐渐分为两类:基于关键点的和基于密集采样的。基于关键点的方法主要有 CornerNet^[10] 和 ExtremeNet^[11]。基于密集采样的方法主要有 FSAF^[12] 和 FCOS^[13]。

1.2 小目标检测

最初对于小目标的检测大都使用传统方法。文献[17]利用目标的高频特性来检测小目标;文献[18]是基于卡尔曼滤波的小目标检测与追踪;文献[19]中使用 Difference of Gaussian (DoG) 算法检测小目标,然后使用卡尔曼滤波与粒子滤波结合的方式来追踪小目标。

随着深度学习的大热,基于深度学习的小目标检测算法得到了广泛应用。2017年,陈江昀基于卷积神经网络 CNN 和超像素算法提出了一种新型的面向微小目标的检测方法^[20]。2018年,郭之先提出了一种改进的 Faster R-CNN 检测算法^[21] 用来检测交通场景中的小目标。

1.3 无人机检测

关于无人机检测的算法可以分为传统的方法和基于深度学习的方法。传统方法中,文献[22]利用卡尔曼模型来动态地检测和跟踪无人机。文献[23]从无人机目标的移动性、空间性、振动性这三个方面建立了无人机检测系统。

基于深度学习的方法中,文献[24]结合 CNN 和运动补偿来检测视频中的飞行目标。文献[25]提出了一种基于卷积神经网络(CNNs)的学习网络,实现了无人机目标的检测。文献[26]提出了两种基于 SSD 的无人机检测方法,一种方法利用 ResNet 网络进行特征提取来检测无人机,另一种方法是直接将 SSD 检测到的无人机目标图像送入到 AlexNet 网络中进行微调。文献[27]针对固定相机和运动相机,设计了两种不同的方法来检测视频中的小型无人机。

2 模型和方法

受 FSAF 和 FCOS 的启发,该文提出了一种 anchor free 的无人机检测方法。

首先,针对地对空拍摄的无人机目标大部分都是小尺度的情况,主干网络要选取适合提取小目标特征

的网络,该文选取 HRNet^[28] 作为主干网络。HRNet 网络可以得到目标丰富的高分辨率特征表示,这对于网络的分类和回归是有益的。对应的,neck 网络由 FPN 变为了 HRNet^[28] 作者提出的 HRFPN,使得金字塔特征图在下采样的过程中保留更多的细粒度信息。

接着,采用逐像素的方式来预测感兴趣区域,采用 soft-NMS 方法对边界框进行排名。最后,对于置信度

损失和回归损失,分别采用 Focal Loss 损失和 Smooth L1 损失。Focal Loss 损失关注困难样本和平衡正负样本数量。

2.1 网络模型

如图 1 所示,该网络模型主要包括三个部分:主干网络、HRFPN 和 head。

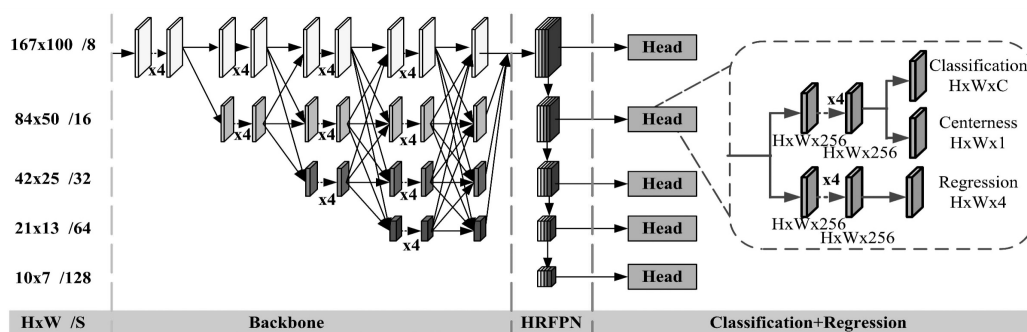


图 1 网络结构

对于主干网络,放弃了常用的 ResNet 和 ResNeXt,采用高分辨率网络 HRNet。HRNet 由四个阶段的子网络组成,第一阶段子网络输出高分辨率特征图,然后每新增一个阶段,该阶段子网络的输出特征图分辨率会降低 2 倍。为了得到目标丰富的高分辨率表示特征,该网络将多分辨率子网并行连接,然后进行多次的多尺度融合,这样使得每一个高分辨率到低分辨率的表示特征都从其他并行表示中反复接收信息。因此,预测的特征图会更加具有细粒度的特征,在空间上的表示也更加精确。HRFPN 代替传统的 FPN,用来构建一个多层次的特征金字塔图。首先低分辨率的特征表示进行双线性上采样,然后将这些特征表示堆叠起来,最后通过多层次的平均池化进行下采样获得多层金字塔特征图。head 模块包括分类和回归分支。该模块主要是在感兴趣区域采用逐像素预测的方法来进行目标分类和回归。

2.2 正负样本划分

一般 anchor free 的目标检测方法需要对特征图从左到右从上到下进行扫描,然后每个像素点都预测一个分类结果和回归边界框的偏移量。对这些像素点划分正负样本的时候,一般的方法是如果像素点落在目标区域内,该像素点就是正样本,否则为负样本,但是考虑到目标真实边界框边缘的像素点都远离目标中心,如果将这些像素点作为正样本,对网络模型的训练会造成困难。尤其对于小目标来讲,它的边缘点与背景像素更为接近,这无疑增加了小目标检测的困难度。综合考虑,将目标真实边界框中的边缘附件的像素点作为负样本,将落在目标中心附件的像素点作为正样本,如图 2 所示。

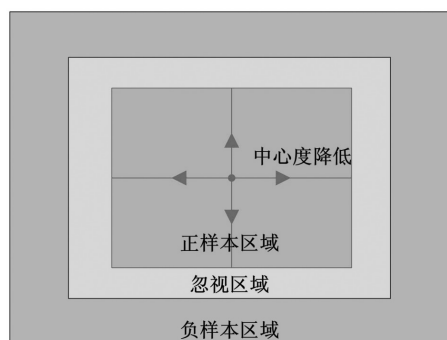


图 2 目标区域的正负样本划分

该文借鉴了 Kong Tao、Sun Fuchun 和 Liu Huaping 等人提出的“Object Fovea”思想,通过参数 δ_1 来调节目标的宽和高,将真实边界框的区域向目标中心收缩一些,收缩后的边界框内的像素点作为正样本;通过参数 δ_2 再次调节边界框的宽和高,将边界框向外扩展一些,扩展后的边界框边缘的像素点作为负样本,处于两个边界框之间内的像素点作为忽略样本,不参与网络分类和回归。忽略样本区域的存在增加了正负样本之间的判别度,从而提高了网络的性能。该文使用 (x_1, y_1, x_2, y_2) 表示目标的真实边界框,也就是目标左上角和右下角的位置。目标边界框的缩放公式如下所示:

$$\begin{cases} c_x = x_1 + 0.5 * (x_2 - x_1) \\ c_y = y_1 + 0.5 * (y_2 - y_1) \\ x'_1 = c_x - 0.5 * (x_2 - x_1) * \delta_1 \\ y'_1 = c_y - 0.5 * (y_2 - y_1) * \delta_1 \\ x'_2 = c_x + 0.5 * (x_2 - x_1) * \delta_1 \\ y'_2 = c_y + 0.5 * (y_2 - y_1) * \delta_1 \end{cases} \quad (1)$$

(c_x, c_y) 是目标真实边界框的中心位置, $(x'_1, y'_1,$

x'_2, y'_2) 是缩放后的正样本区域的左上角和右下角坐标。将公式中的缩放因子 δ_1 替换为扩张因子 δ_2 就可以获得负样本区域的左上角和右下角坐标。

由于该文采用了逐像素预测目标的方法,虽然能很好地提升网络召回率,但是在逐像素预测的过程中会产生许多低质量的预测边界框。虽然用目标的中心区域作为正样本区域已经丢弃了一些远离目标中心的低质量边界框,但是,正样本区域的边缘区域依旧还是会存在一些低质量的边界框。基于此,该文借鉴了 FCOS 的 centerness 概念,提出利用中心度来衰减远离目标中心的训练像素点的置信度。中心度越大,表明像素点越靠近目标中心,反之,像素点离目标中心越远。这样中心度就可以抑制检测到的低质量的边界框,并且中心度策略不会引入任何超参数。中心度计算公式如下:

$$\text{中心度} = \sqrt{\frac{\exp(\min(t_{x_1}, t_{x_2}))}{\exp(\max(t_{x_1}, t_{x_2}))} * \frac{\exp(\min(t_{y_1}, t_{y_2}))}{\exp(\max(t_{y_1}, t_{y_2}))}} \quad (2)$$

其中, $(t_{x_1}, t_{y_1}, t_{x_2}, t_{y_2})$ 是由式(3)计算出来的边界框回归的偏移量。网络训练时,中心度采用二元交叉熵损失(binary cross entropy)来约束中心度取值的范围。中心度取值的范围为(0,1)。中心度策略就是在每层预测特征图上添加一个网络分支,该网络分支与分类并行,这就相当于给网络模型添加了一个中心度损失。中心度损失会保证预测的边界框尽可能地靠近目标中心。在训练的过程中,约束中心度的值,使其接近于0。这样在测试网络模型的过程中,soft-NMS 就可以更好地滤除远离目标中心的低质量边界框,从而提高网络模型检测性能。

2.3 边界框回归

对目标区域进行正负样本划分后,采用多尺度策略,利用 HRFPN 网络生成的多层金字塔特征图来检测不同尺度的目标。每个特征图要预测的目标都有一个正样本区域(感兴趣区域),对每个正样本区域进行逐像素预测。如图3所示,预测由两个分支网络负责,一个分支网络负责分类,另一个分支网络负责边界框回归。

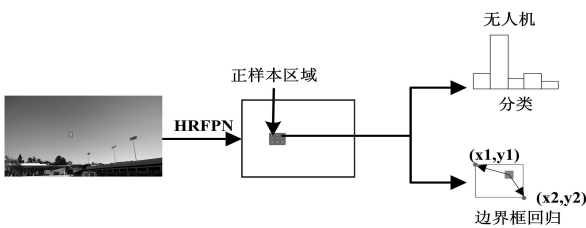


图3 目标分类和回归示意

负责边界框回归的分支网络,将感兴趣区域内的像素坐标 (x, y) 映射到原始图上的坐标,然后计算映

射后的坐标与其所属目标样本的真实边界框的偏移量。偏移量计算公式如下:

$$\begin{cases} t_{x_1} = \log \frac{2^l(x + 0.5) - x_1}{s} \\ t_{y_1} = \log \frac{2^l(y + 0.5) - y_1}{s} \\ t_{x_2} = \log \frac{x_2 - 2^l(x + 0.5)}{s} \\ t_{y_2} = \log \frac{y_2 - 2^l(y + 0.5)}{s} \end{cases} \quad (3)$$

其中, 2^l 是第 l 层金字塔特征图的 stride 值, s 应该选择与目标边框面积最接近的尺度。通过调整 s , 可以控制每个特征图上覆盖目标的尺度范围。 $(t_{x_1}, t_{y_1}, t_{x_2}, t_{y_2})$ 的直观含义可以参考图4, 这里 $(t_{x_1}, t_{y_1}, t_{x_2}, t_{y_2})$ 不再是简单地衡量与边界框四周的距离值, 而是衡量与边界框四周的偏移程度, 使用 log 函数来计算这种偏移程度。

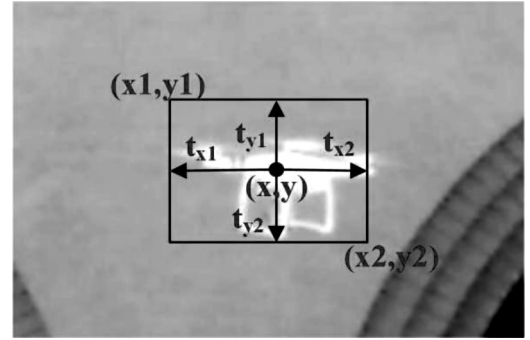


图4 偏移量直观含义

2.4 损失函数

训练损失包括置信损失和定位损失, 损失定义如下:

$$L(p, c, l, g) = \frac{1}{N_{\text{pos}} \sum_{x,y}} L_{\text{cls}}(p, c) + \frac{\lambda}{N_{\text{pos}} \sum_{x,y}} \zeta(c > 0) L_{\text{loc}}(l, g) \quad (4)$$

置信损失 L_{cls} 采用的是 Focal loss, 定位损失 L_{loc} 采用的是 Smooth L1 损失。 N_{pos} 是正样本的数量, $c = \{0, 1\}$ 是来匹配预测的框与类别 p 的真实边界框, 如果 $c = 1$, 预测框是正样本, 否则为负样本。 λ 在实验中为1用来平衡定位损失 L_{loc} 的权重。 l 代表预测框的参数, g 是标记框的参数。定位损失展开如下:

$$L_{\text{loc}} = \frac{\lambda}{N_{\text{pos}} \sum_{x,y}} \zeta(c > 0) L_{\text{smoothL1}}(l, g) \quad (5)$$

文中的 anchor free 的方法采用的是逐像素预测边界框, 将像素的位置 (x, y) 作为训练样本, 直接对每个位置的目标边界框进行归一化处理。采用计算像素位置 (x, y) 与真实框 (x_1, y_1, x_2, y_2) (左上角与右上角位置) 的偏移量 $(t_{x_1}, t_{y_1}, t_{x_2}, t_{y_2})$ 的方法。偏移量计算如式(3)。

3 实验结果和分析

3.1 无人机数据集

无人机数据采集与标注是进行实验的基础。如图 5 所示,这些无人机图像的采集主要有三个方面,一个是自己采集的被监控摄像头捕捉到的图像,另外两个是从文献[29-30]收集到的图像。总共 9 名学生花了近两个月的时间来标记和重新标记这些无人机图像,从而创建了一个包括 32 237 张图像的无人机数据集。从图 5 中可以看出,无人机(矩形标记框包围的目标)的尺寸太小,人眼根本无法发现。因此,无人机检测是一项具有挑战性的任务。笔者设计的无人机数据集共有 32 237 张图像,其中 26 237 张图像作为训练集,1 000 张图像作为验证集,5 000 张图像作为测试集。



(a) 自己采集的图像



(b) 文献[29]中的图像



(c) 文献[30]中的图像

图 5 无人机数据集

3.2 实验过程

使用无人机数据集的训练集来训练网络。在 Py-

torch 框架下,使用一块 1080 的 NVIDIA GPU 进行实验。实验中,使用训练好的 FCOS 网络模型作为预训练模型来训练该论文提出的网络,然后在训练集上微调网络模型。为了定量和定性地分析比较网络,所有网络的基本参数设置如下:迭代次数为 20 epochs,动量为 0.9,权重衰减为 0.000 1。网络训练的初始学习率为 10^{-3} 。当网络迭代到 5 epochs、10 epochs 和 15 epochs 时,学习率变为 10^{-4} 、 10^{-5} 和 10^{-6} 。

3.3 实验结果

3.3.1 消融实验

文中方法是基于 FCOS 网络的,然后对基础网络 FCOS 采取了以下几种策略:(1)基础网络使用中心区域思想回归边界框;(2)主干网络由传统的 ResNet 和 ResNeXt 变为 HRNet;(3)neck 网络也由传统的 FPN 变为 HRFPN;(4)测试网络模型时,由 NMS 策略变为 soft-NMS 策略。为了充分说明这几种策略对网络是有益的,该文进行了消融实验,实验结果如表 1 和表 2 所示。

可以发现采用 soft-NMS 策略后,AP 提升了大约 1.8 个精度,AP₅₀ 提升了大约 0.7 个精度,AP₇₅ 提升了 3.6 个精度。对于 AR 的各个指标,除了 AR_{max=1} 没有提升,其他均提升了大约 5 个精度。这个结果充分证明了检测算法采用 soft-NMS 策略的有效性。

对于 HRNet 网络,分别与 FPN 和 HRFPN 网络组合使用,发现使用 FPN 后,网络性能反而下降了,而使用 HRFPN 后,网络性能指均有所提高。AP 提升了大约 2.5 个精度,AP₅₀ 提升了 0.6 个精度,AP₇₅ 提升了 0.6 个精度,AR_{max=1} 提升了 4.8 个精度,AR_{max=10} 提升了 1.1 个精度,AR_{max=100} 也提升了 0.8 个精度。这些充分证明了 HRNet 和 HRFPN 组合使用的有效性。

HRNet 和 HRFPN 组合使用后,使用中心区域思想回归边界框,整体检测性能有所提升,其中 5 个检测指标有提升,只有 AP₅₀ 没有提升,略微下降了一些。AP 和 AR 这 6 个指标中,AP 提升了大约 0.8 个精度,AP₅₀ 稍微下降了 0.4 个精度,AP₇₅ 提升了 6.2 个精度,AR_{max=1} 提升了 4.8 个精度,AR_{max=10} 提升了 1.1 个精度,AR_{max=100} 也提升了 0.8 个精度。这些充分证明,使用中心区域思想回归边界框是非常有效的。

表 1 关于 AP 的消融实验

FCOS	Soft-NMS	HRNet	FPN	HRFPN	中心区域	AP	AP ₅₀	AP ₇₅
√			√			54.0	96.1	54.3
√	√		√			55.8	96.8	57.9
√	√	√	√			48.2	93.6	43.6
√	√	√		√		58.3	97.0	58.5
√	√	√		√	√	59.1	96.6	64.7

表 2 关于 AR 的消融实验

FCOS	Soft-NMS	HRNet	FPN	HRFPN	中心区域	$AR_{\max=1}$	$AR_{\max=10}$	$AR_{\max=100}$
✓			✓			55.8	60.0	60.1
✓	✓		✓			55.8	65.1	65.4
✓	✓	✓	✓			54.4	58.8	61.7
✓	✓	✓		✓		60.6	66.2	66.2
✓	✓	✓		✓	✓	62.5	67.7	67.8

表 3 目标检测算法在无人机数据集上的 AP 结果

方法	backbone	neck	AP	AP_{50}	AP_{75}
Faster RCNN	ResNet	FPN	37.1	78.0	29.4
Cascade R-CNN	ResNet	FPN	57.5	95.5	63.6
SSD	VGG16	-	53.5	95.5	54.7
FSAF	RetinaNet-r101-FPN	FPN	45.1	93.9	35.3
FCOS	ResNet-r101-FPN	FPN	55.8	96.8	57.9
FCOS	ResNeXt-x101-FPN	FPN	47.4	93.8	41.2
Ours	HRNet	HRFPN	59.1	96.6	64.7

表 4 目标检测算法在无人机数据集上的 AR 结果

方法	backbone	neck	$AR_{\max=1}$	$AR_{\max=10}$	$AR_{\max=100}$
Faster RCNN	ResNet	FPN	43.1	45.6	45.8
Cascade R-CNN	ResNet	FPN	60.5	64.4	64.4
SSD	VGG16	-	58.9	64.0	65.7
FSAF	RetinaNet-r101-FPN	FPN	61.1	58.2	60.2
FCOS	ResNet-r101-FPN	FPN	55.8	65.1	65.4
FCOS	ResNeXt-x101-FPN	FPN	53.3	58.9	61.9
Ours	HRNet	HRFPN	62.5	67.7	67.8

3.3.2 实验对比

为了充分说明文中方法的优越性,对基于 anchor 的经典目标检测算法以及 anchor free 的目标检测算法也在无人机数据集上进行了实验,实验结果如表 3 和表 4 所示。

基于 anchor 的目标检测方法,对经典的两阶段目标检测算法 Faster R-CNN、Cascade R-CNN 和一阶段目标检测算法 SSD 进行了复现。对于 anchor free 的目标检测算法,对 FSAF 和 FCOS 进行了复现。测试时,这些目标检测算法均采用了 soft-NMS 策略。从表 3 和表 4 中可以发现,文中方法的表现是最优异的,仅仅是 AP_{50} 的结果略低于基于 ResNet-r101-FPN 的 FCOS 算法。对于 AP 、 AP_{75} 、 $AR_{\max=1}$ 、 $AR_{\max=10}$ 和 $AR_{\max=100}$,文中方法相比于对应的第二名的算法分别提高了 1.6、1.1、1.4、1.6 和 2.1 个精度,这些结果充分证明了文中方法的优越性。

4 结束语

提出了一种 anchor free 的无人机检测方法,该方

法完全避免了与 anchor 相关的计算和超参数。针对无人机大部分都是小目标的情况,提出使用 HRNet 网络作为主干网络来获得具有丰富细粒度信息的特征。相应的,HRFPN 代替传统的 FPN 来生成多层金字塔特征图。相比较于其他基于 anchor 和 anchor free 的目标检测方法,该方法对于无人机检测获得了更好的表现,具体表现为:相比较于基于 anchor 的目标检测器,该方法更加具有鲁棒性,可以自适应地预测目标边界框,相比较于其他的 anchor free 的目标检测器,该方法获得了更好的精确率和召回率。另外,该方法也可以用来检测其他小目标,而不仅仅是无人机。

参考文献:

- [1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE conference on computer vision and pattern recognition. Columbus, OH: IEEE, 2014: 580-587.
- [2] GIRSHICK R. Fast R-CNN[C]//2015 IEEE conference on computer vision and pattern recognition. Santiago: IEEE,

- 2015;1440–1448.
- [3] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149.
- [4] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//European conference on computer vision. Amsterdam: Springer, 2016: 21–37.
- [5] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//30th IEEE conference on computer vision and pattern recognition. Honolulu, HI: IEEE, 2017: 6517–6525.
- [6] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 99: 2999–3007.
- [7] YANG T, ZHANG X Y, LI Z M, et al. MetaAnchor: learning to detect objects with customized anchors [C]//Advances in neural information processing systems 31. Sweden: Curran Associates Inc, 2018: 320–330.
- [8] 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. Honolulu, HI: IEEE, 2017: 936–944.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE conference on computer vision and pattern recognition. Las Vegas, NV: IEEE, 2016: 779–788.
- [10] LAW H, DENG J. CornerNet: detecting objects as paired keypoints [J]. International Journal of Computer Vision, 2020, 128: 642–656.
- [11] ZHOU X, ZHUO J, KRÄHENBÜHL, et al. Bottom-up object detection by grouping extreme and center points [C]//2019 IEEE conference on computer vision and pattern recognition. Long Beach, CA: IEEE, 2019: 850–859.
- [12] ZHU C, HE Y, SAVVIDES M. Feature selective anchor-free module for single-shot object detection [C]//2019 IEEE conference on computer vision and pattern recognition. Seoul, South Korea: IEEE, 2019: 840–849.
- [13] TIAN Z, SHEN C, CHEN H, et al. FCOS: fully convolutional one-stage object detection [C]//2019 IEEE international conference on computer vision and pattern recognition. Seoul, South Korea: IEEE, 2019: 9627–9636.
- [14] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916.
- [15] KONG T, YAO A B, CHEN Y R, et al. HyperNet: towards accurate region proposal generation and joint object detection [C]//2016 IEEE conference on computer vision and pattern recognition. Las Vegas, NV: IEEE, 2016: 845–853.
- [16] WANG J, CHEN K, YANG S, et al. Region proposal by guided anchoring [C]//2019 IEEE international conference on computer vision and pattern recognition. Seoul, South Korea: IEEE, 2019: 2965–2947.
- [17] 侯洁, 辛云宏. 基于高通滤波和图像增强的红外小目标检测方法 [J]. 红外技术, 2013, 35(5): 279–284.
- [18] 高建伟, 李磊, 姚睿, 等. 基于卡尔曼滤波的弱小目标实时检测与跟踪 [J]. 计算机工程, 2012, 38(2): 4–7.
- [19] 孙继刚. 序列图像红外小目标检测与跟踪算法研究 [D]. 长春: 中国科学院研究生院 (长春光学精密机械与物理研究所), 2014.
- [20] 陈江昀. 一种基于深度学习的新型小目标检测方法 [J]. 计算机应用与软件, 2017, 34(10): 227–231.
- [21] 郭之先. 基于深度卷积神经网络的小目标检测 [D]. 南昌: 南昌航空大学, 2018.
- [22] WU Y W, SUI Y, WANG G H. Vision-based real-time aerial object localization and tracking for UAV sensing system [J]. IEEE Access, 2017, 5: 23969–23978.
- [23] 周伟. 基于 CSI 的无人机检测与 3D 定位研究 [D]. 大连: 大连理工大学, 2018.
- [24] ROZANTSEV A, LEPETIT V, FUA P. Detecting flying objects using a single moving camera [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(5): 879–892.
- [25] 蒋兆军, 成孝刚, 彭雅琴, 等. 基于深度学习的无人机识别算法研究 [J]. 电子技术应用, 2017, 43(7): 84–87.
- [26] 李秋珍, 熊饶饶, 王汝鹏, 等. 基于 SSD 算法的实时无人机识别方法研究 [J]. 舰船电子工程, 2019, 39(5): 30–35.
- [27] 王子哲. 基于视频的无人机目标检测算法研究 [D]. 郑州: 郑州大学, 2019.
- [28] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation [C]//2019 IEEE conference on computer vision and pattern recognition. Long Beach, CA, USA: IEEE, 2019: 5686–5696.
- [29] CHEN Y R, AGGARWAL P, CHOI J, et al. A deep learning approach to drone monitoring [C]//2017 Asia-Pacific signal and information processing association annual summit and conference. Kuala Lumpur: IEEE, 2017: 686–691.
- [30] AKER C, KALKAN S. Using deep networks for drone detection [C]//2017 14th IEEE international conference on advanced video and signal based surveillance. Lecce: IEEE, 2017: 1–6.