

基于单阶段全卷积检测器的遥感图像形状自适应椭圆标签分配方法<sup>①</sup>禹鑫焱<sup>②</sup> 卢江平 林 密 周利波 欧林林<sup>③</sup>

(浙江工业大学信息工程学院 杭州 310023)

**摘 要** 基于无锚框的检测方法在目标检测领域中发展迅速。然而在遥感图像中,目标存在角度任意、密集排列以及形状差异大等难点,使得遥感图像的检测仍是一项挑战。为此,本文提出了基于单阶段全卷积检测器(FCOS)改进的无锚框检测方法。首先,为了挖掘更多潜在的高质量锚点,提出基于椭圆方程的形状自适应特征点采样方法。然后,为进一步降低边界低质量样本点的影响,提出椭圆中心度量方法,相较原有的中心度量方法提供更合理的权重。此外,针对分类与回归的不一致问题,提出交并比(IoU)联合指导策略,将椭圆中心度量与 IoU 得分相结合作为质量分数监督分类分支,进一步提升检测精度。在 DOTA 1.0 数据集上的平均精度达到了 79.17%,优于现有多数无锚框检测算法。

**关键词** 遥感图像;深度学习;目标检测;标签分配

通用目标检测一般通过水平矩形框检测并定位图像中的目标,近年来取得了长足的进展,比如 faster-RCNN(region with CNN)系列算法<sup>[1,2]</sup>、YOLO(you only look once)系列算法<sup>[3,4]</sup>、单阶段全卷积检测器(fully convolutional one-stage,FCOS)<sup>[5]</sup>、YOLOX<sup>[6]</sup>等,在智能交通、行人检测等众多领域都有着广泛的应用。对于遥感图像,其目标存在方向任意、排列密集、背景复杂、目标尺度变化大等问题,采用水平检测框往往会引入过多的背景信息,且密集分布的检测框极有可能在非极大值抑制(non-maximum suppression,NMS)过程中被剔除,从而造成漏检。因此,遥感图像目标检测一般利用包围物体的最小外接矩形框,即旋转框检测。现有的遥感图像目标检测方法大多是从通用目标检测中添加目标的旋转角度而来的。

标签分配是指卷积神经网络(convolutional neural network,CNN)在训练阶段将标签分配给多个候选位置,以提高鲁棒性。基于 CNN 的训练本质上是一个学习从多个候选位置预测到真实对象的一对多

映射的过程。不同的标签分配策略在训练阶段将生成不同的候选样本,这会直接影响检测性能。因此,对于目标检测任务来说,构建一个符合对象形状和方向的候选样本空间尤为重要。

为了尽可能覆盖整张图片,找到合适的正样本,大多数目标检测算法都采用基于锚框的方法,多尺度、多比例、多角度地在特征图上铺设锚框,虽然取得了不错的效果,但大量的锚框极大地增加了计算量,且遥感图像的背景复杂,导致正负样本的极不平衡。此外,基于锚框策略的参数设置比较依赖数据集自身特性,不具有普适性。以 FCOS<sup>[5]</sup>、CornerNet<sup>[7]</sup>等为代表的无锚框方法,去除了预设锚框,通过锚点直接预测目标的边界框,从而省去了复杂计算,使训练过程变得相对简单。CornerNet 将目标框角点位置处设为正样本,其余位置设为负样本。FCOS、Foveabox<sup>[8]</sup>将目标框的中心区域位置设为正样本。CenterNet<sup>[9]</sup>和 BBAVectors<sup>[10]</sup>利用关键点估计得到物体中心。但上述方法在构建正负样本空间时都没有充分考虑目标的形状和方向。这主要是由

① 浙江省自然科学基金(LY21F030018,LQ22F030021)和国家自然科学基金(62203392)资助项目。

② 男,1979年生,博士,副教授;研究方向:机器人控制与规划;E-mail:yuxinyinet@163.com。

③ 通讯作者,E-mail:linlinou@zjut.edu.cn。

(收稿日期:2023-02-06)

于对于水平框检测而言,物体的实际方向与矩形框的形状并不匹配,从而无法完全挖掘潜在的优质锚点。

为此,本文提出了基于 FCOS 的形状自适应椭圆标签分配方法。首先,提出了一种基于椭圆方程的特征点采样方法,选择目标框内接椭圆内部位置作为候选正样本空间;然后,根据目标的形状(长宽比)自适应地选择中心区域;最后,针对提出的椭圆样本空间,提出了椭圆中心度量,相比于 FCOS 的中心度量计算方法,其分布更加合理,更能准确度量样本点的质量。为了解决分类得分与质量评估的不一致问题,提出了交并比(intersection over union, IoU)联合指导策略,动态计算训练过程中的 IoU,并与椭圆中心度量相结合以监督分类分支,进一步提升了检测精度。本文所提方法在多个公开数据集 DOTA v1.0<sup>[11]</sup>、DIOR-R<sup>[12]</sup>、HRSC 2016<sup>[13]</sup>和 UCAS\_AOD<sup>[14]</sup>上均取得了较好结果,优于目前绝大多数无锚框检测方法。

## 1 相关工作

### 1.1 旋转目标检测器

近年来,得益于卷积神经网络的不断发展以及大规模遥感数据集的开放,遥感图像旋转目标检测算法取得了快速的发展。现有的方法大多基于成熟的通用目标检测算法,主要可以分为基于锚框的方法和无锚框方法。为了克服遥感图像目标角度的任意性,一些工作如文献[15]和[16]通过增设不同角度的锚框,以找到最合适的锚框,但这极大增加了计算量。因此,文献[17]提出了感兴趣区域(region of interest, RoI)转化器,利用一个可学习模块将水平 RoI 转化为旋转 RoI,从而避免了大量锚框。文献[18]将分割掩码预测与边界框预测相结合,以增强网络的语义信息。文献[19]提出一种滑动水平框顶点的方式来表示旋转框,通过预测一个四点多边形避免预测目标角度。文献[20]提出了一种单次对齐网络(single-shot alignment network, S2A-Net),通过深度特征对齐缓解定位与分类的不一致问题。文献[21]提出了高斯建模的方法,将旋转矩

形框转化为二维高斯分布,通过计算 2 个高斯分布之间的相似度来近似 2 个旋转框的相似性,避免了旋转框的边界问题。此外,文献[22,23]提出了角度分类的思想,如环形平滑标签(circular smooth label, CSL)<sup>[22]</sup>、密集编码标签(densely coded label, DCL)<sup>[23]</sup>,将角度划分为特定个数类别,通过分类的方法预测目标角度。

### 1.2 标签分配策略

无锚框检测器和基于锚框的检测器的本质区别在于如何定义训练样本<sup>[24]</sup>,正负样本的选择将很大程度上决定检测器的性能。常用的标签分配方法以真实框与锚框之间的 IoU 作为分配原则,但 IoU 对于细长的目标并不友好,细微的角度偏差就有可能造成巨大的 IoU 变化。最近,许多改进标签分配的工作被提出。文献[25]提出了一种学习匹配方式,可以使目标在最大似然原则下灵活选择锚点。文献[26]在此基础上,进一步让模型学习定位正负样本位置并自动分配标签。文献[27]利用高斯混合模型拟合锚框的概率分布,并得到锚框的得分从而确定模型的正负样本。文献[28]指出基于 IoU 的标签分配策略导致了次优的定位能力评估,因而引入新的匹配度测量锚框的定位潜力,以实现寻找高质量的锚框。文献[29]将标签分配任务视为最优传输距离问题,通过在全局上最小化运输成本以找到最优的分配方案。文献[30]提出了一种双重加权的标签分配方式,为每个锚点动态分配单独的正负权重,从而在密集场景下细致地区分正负样本。文献[31]关注于小目标检测,提出一种基于高斯感受野的标签分配策略。通过计算高斯感受野与真实框的相似度代替 IoU。但这些方法都没有充分考虑对象位置、形象和方向上的特征。文献[32]为匹配遥感图像中物体长宽比的变化,提出了一种自适应长宽比的锚框,该锚框可以自适应学习每个类别的合适长宽比。文献[33]根据目标长宽比为目标设计最佳 IoU 阈值。

## 2 方法

### 2.1 网络结构

如图 1 所示,本文以 FCOS 框架为基线,主要针

对遥感图像的特点设计标签分配策略。首先,形状自适应采样策略能为目标匹配更广泛、更优质的样本,同时椭圆中心度量将这些样本赋予不同的距离

权重。然后,引入IoU联合指导训练,提供更加合理的定位评估,避免分类结果与回归结果的不一致。

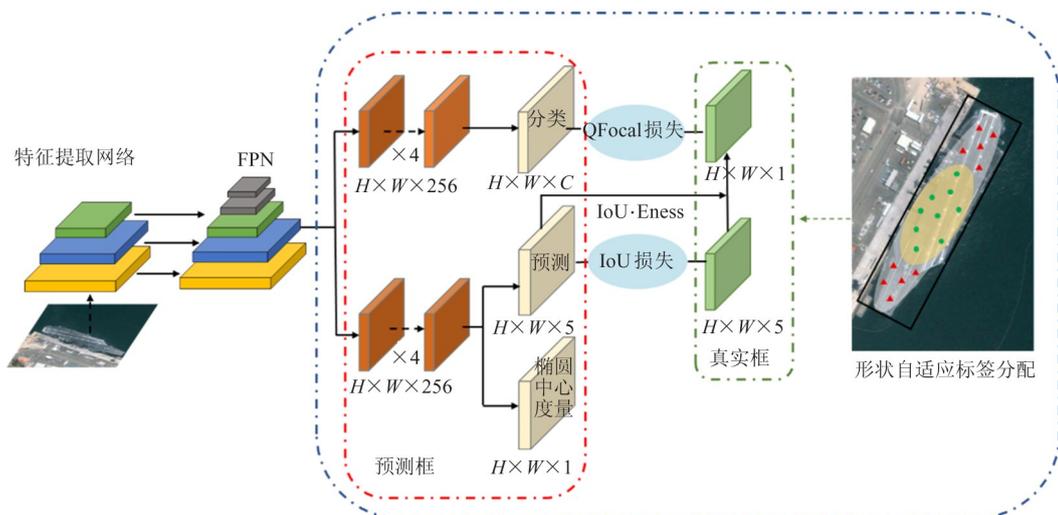


图1 所提方法框架示意图

## 2.2 基于椭圆方程的形状自适应特征点采样方法

FCOS使用中心采样策略来选取靠近物体中心区域的点作为正样本点,如图2(a)所示,从而减少低质量样本的影响以提升检测精度。首先,通过特征点与目标边界框4条边的距离来筛选出位于目标边界框内部的特征点,划分规则可以表示为

$$\begin{cases} \text{正样本} & m_{i-1} < \max(l, t, r, b) \leq m_i \text{ 且} \\ & \min(l, t, r, b) > 0 \\ \text{负样本} & \text{其他} \end{cases} \quad (1)$$

其中,  $i \in \{1, 2, 3, 4, 5\}$  代表5层特征图,  $\{m_j\}_{j=0}^5$  为自定义的范围, 设置为  $\{0, 64, 128, 256, 512, \infty\}$ ,  $(l, t, r, b)$  分别表示特征点到边界框4条边的距离。然后,手动设置一个  $3 \times 3$  的水平区域,用同样的方法进一步筛选出中心区域的样本点。最后,通过取交集的方式得到最终的正样本点。但这仅适用于通用目标检测,对于遥感图像目标,这种中心采样的方式会减少大长宽比目标的采样点数量。为了避免这些影响,本文提出了基于椭圆方程的自适应特征采样方法,如图2(b)所示。首先,提出了基于椭圆方程的特征点定位方法,对于一个旋转目标矩形框  $(x, y, w, h, \theta)$ , 以中心点  $(x, y)$  为圆心,以  $\theta$  方向为  $X$  轴正方向建立椭圆方程:

$$\frac{x^2}{\left(\frac{w}{2}\right)^2} + \frac{y^2}{\left(\frac{h}{2}\right)^2} = 1 \quad (2)$$

然后,只需要获得其在椭圆坐标系下的相对坐标即可判断其位置关系。

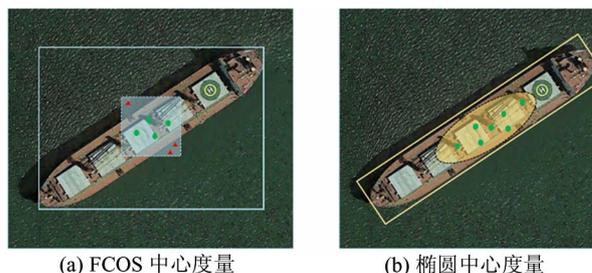


图2 中心采样策略

如图3所示,对于任意一个特征点  $(m, n)$ , 其相对于椭圆圆心的偏差可以表示为  $(\Delta x, \Delta y)$ 。其中,  $\Delta x = m - x, \Delta y = n - y$ 。那么,该特征点的相对坐标可以表示为

$$x' = \Delta x \cos \theta + \Delta y \sin \theta \quad (3)$$

$$y' = \Delta y \cos \theta - \Delta x \sin \theta$$

结合式(1)和(2),可以得到:

$$\alpha = \frac{x'^2}{\left(\frac{w}{2}\right)^2} + \frac{y'^2}{\left(\frac{h}{2}\right)^2} \quad (4)$$

因此,通过  $\alpha$  的值即可判断特征点与目标边界框的位置关系:

$$\begin{cases} \alpha > 1 & \text{在外部} \\ \alpha \leq 1 & \text{在内部} \end{cases} \quad (5)$$

为了进一步减少低质量样本的影响,同时保留足够的候选特征点,本文提出了形状自适应椭圆中心采样方法。在目标中心用一个随目标长宽比自适应变化的椭圆区域代替 FCOS 中的固定中心区域,其椭圆方程可以表示为

$$\frac{x^2}{kr^2} + \frac{y^2}{r^2} = 1 \quad (6)$$

其中,  $r$  为超参数,代表椭圆短半轴,在本文中设置为 2;  $k = w / h$ , 代表目标的长宽比。因此,当真实目标的长宽比越大,中心椭圆也会相应地扩大采样范围,以提供更多潜在优质样本锚点

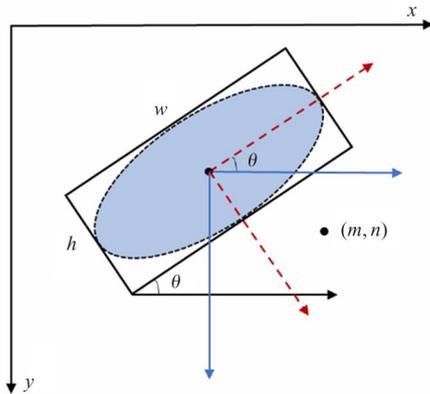


图 3 基于椭圆方程的特征点采样

### 2.3 椭圆中心度量

FCOS 将模型预测的分类置信度与中心度相结合,作为 NMS 过程中的最终排序依据。其中心度的计算方法如下:

$$centerness = \sqrt{\frac{\min(l,r) \times \min(l,r)}{\max(l,r) \times \max(l,r)}} \quad (7)$$

中心度衡量的是样本点与目标中心的位置关系,样本点越接近于中心,则中心度越接近 1;反之,当样本点越靠近目标边缘,则中心度逐渐衰减为 0。当一个样本点越接近目标的中心区域,那么该样本点作为正样本点的可能性就越大,从这一点来说,中心度方法使得模型更加关注目标中心区域的样本点,从而提高了 FCOS 的精度。

本文对中心度量的方法进行了可视化,如图 4(a)

所示,用颜色的深浅代表其靠近中心的程度。可以发现,其分布在中心区域呈近似菱形分布,越接近边缘,其形状逐渐膨胀为一个矩形。为此,本文提出了椭圆中心度量的方法,其公式可以表达为

$$E_{\text{centerness}} = \sqrt{1 - \frac{(r-l)^2}{(r+l)^2} - \frac{(b-t)^2}{(b+t)^2}} \quad (8)$$

其中,  $l, t, r$  和  $b$  分别表示特征点到边界框 4 条边的距离。同样,本文对其作了可视化分析,如图 4(b) 所示,其分布成椭圆形状,当特征点的位置不断靠近边缘时,其值向外均匀变小,与所提出的采样空间完全一致,因此能准确评估候选区域中样本点的质量。

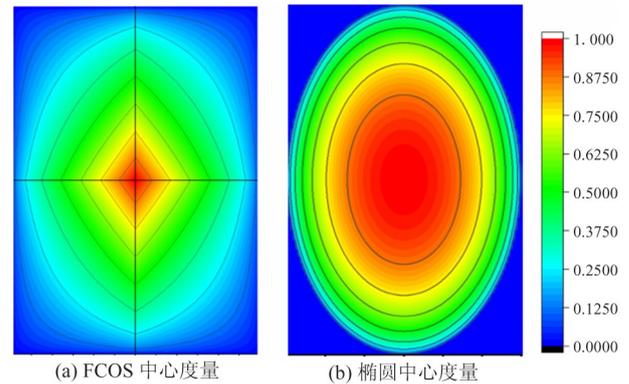


图 4 中心度量策略

### 2.4 IoU 联合指导标签分配

回归分支和分类分支是独立训练的。但在推理过程中,分类分支的输出分数用于对回归分支预测的框进行排序的依据。这导致了训练和推理之间的不一致。此外,固定中心度的评分不能适应物体的各种形状和姿态,不能始终提供对定位质量的合理估计。以往的工作<sup>[34-35]</sup>添加了单独的分支来预测 NMS 的 IoU 得分。

在本文的工作中,通过动态计算训练期间预测的边界框和真实框的 IoU,并耦合到分类分支中。具体来说,在每次训练迭代过程中,动态计算预测框与真实框的 IoU 得分,并与椭圆中心度量相乘得到最终分数,最终分数由以下公式得到。

$$\text{label} = \text{IoU} \times E_{\text{centerness}} \quad (9)$$

在训练初期,椭圆中心度量有助于稳定训练,而动态更新的 IoU 得分则有助于挑选更高质量的锚点,从而重塑分数分布,提高 NMS 的排名。此外,该策略可以使 2 个分支相互作用,以保持训练和推理

的一致性。

## 2.5 损失函数

本文定义损失函数如下:

$$L(\{p_{x,y}\}, \{t_{x,y}\}) = \frac{1}{N_{\text{pos}}} \sum_{x,y} L_{\text{cls}}(p_{x,y}, c_{x,y}^*) + \frac{\lambda}{N_{\text{pos}}} \sum_{x,y} I_{|c_{x,y}^* > 0|} L_{\text{reg}}(t_{x,y}, t_{x,y}^*) \quad (10)$$

其中,  $p_{x,y}$  代表预测类别概率,  $t_{x,y}$  代表预测坐标;  $c_{x,y}^*$  代表该位置处真实类别,  $t_{x,y}^*$  代表真实坐标;  $L_{\text{cls}}$  是质量集点损失函数<sup>[36]</sup>;  $L_{\text{reg}}$  代表 IoU 损失;  $N_{\text{pos}}$  代表正样本的数量;  $\lambda = 1$  用于平衡分类与回归损失;  $I_{|c_{x,y}^* > 0|}$  代表指示函数, 当  $c_{x,y}^* > 0$  时为 1, 否则为 0。

## 3 实验

### 3.1 数据集

为验证本文所提方法的有效性, 本文主要在 DOTA 1.0、DIOR-R、HRSC 2016 以及 UCAS-AOD 数据集上进行实验。

DOTA 1.0 是目前最常用的旋转目标检测数据集, 包含 2 806 个图像和 188 282 个实例, 共 15 个类别, 分别为飞机 (PL)、棒球场 (BD)、桥梁 (BR)、田径场 (GFT)、小型车辆 (SV)、大型车辆 (LV)、船舶 (SH)、网球场 (TC)、篮球场 (BC)、储油罐 (ST)、足球场 (SBF)、环形交叉路口 (RA)、港口 (HA)、游泳池 (SP) 和直升机 (HC)。图片大小在  $800 \times 800$  到  $4\,000 \times 4\,000$  像素之间。由于尺寸过大, 将每张图像切割为  $1\,024 \times 1\,024$  像素子图像, 并重叠 200 像素值。在多尺度训练情况下, 其图像参数设置为  $\{0.5, 1.0, 1.5\}$ 。

DIOR-R 是 DIOR<sup>[12]</sup> 数据集的旋转标注形式, 与其共享相同的图片。共有 23 463 张图像和 192 518 个实例, 涵盖了 20 个常用类别, 包括飞机 (APL)、机场 (APO)、棒球场 (BF)、篮球场 (BC)、桥梁 (BR)、烟囱 (CH)、高速服务区 (ESA)、高速收费站 (ETS)、大坝 (DAM)、高尔夫球场 (GF)、田径场 (GTF)、港口 (HA)、立交桥 (OP)、船舶 (SH)、体育场 (STA)、储油罐 (STO)、网球场 (TC)、火车站 (TS)、车辆 (VE) 和风车 (WM)。所有图像的尺寸为  $800 \times 800$

像素。

HRSC 2016 是一个船舶数据集, 包含 1 061 个图像和 2 976 个实例。图像大小从  $300 \times 300$  到  $1\,500 \times 900$  像素。训练集包括 617 张图像, 测试集包括 444 张图像。在训练和验证过程中, 所有图像调整为  $800 \times 800$  像素。

UCAS-AOD 是一个汽车和飞机数据集, 共包含 1 510 张图片共 14 596 个实例, 其中训练集共 1 057 张图片, 测试集共 453 张图片。

### 3.2 评价指标

在旋转目标检测任务中, 通常采用 IoU 阈值为 0.5 的平均精度均值 (mean average precision, mAP) 来评价检测精度, 它遵循 PASVAL VOC<sup>[37]</sup> 标准。mAP<sub>50:95</sub> 代表 IoU 阈值以 0.05 为步长在 0.50 至 0.95 之间的均值, 因此精度要求更高。

### 3.3 实验设置

本文在开源框架 MMrotate<sup>[38]</sup> 上进行所有的实验, 其中集成了许多先进的旋转检测器。所有的实验都在一张 GeForce RTX 2080 Ti 上进行训练。除非另有说明, 实验由 ResNet50 初始化。本文使用随机梯度下降 (stochastic gradient descent, SGD) 优化器来训练模型, 动量衰减和权重衰减分别为 0.900 0 和 0.000 1, 批处理大小为 2, 初始学习率为 0.000 250。对于 DOTA 数据集, 总共训练了 12 个迭代 (epoch), 在第 8 个 epoch 和第 11 个 epoch 时学习率分别降低了 10 倍; 如果使用数据增强或多尺度训练, 学习率和迭代次数将翻倍。对于 DIOR-R 数据集, 总共训练了 24 个 epoch, 在第 16 个 epoch 和第 22 个 epoch 时学习率分别降低了 10 倍。对于 UCAS\_AOD 数据集, 总共训练了 12 个 epoch, 在第 8 个 epoch 和第 11 个 epoch 学习率分别降低了 10 倍。对于 HRSC 2016, 总共训练了 72 个 epoch, 学习率为 0.000 125, 在第 48 个 epoch 和第 66 个 epoch 时学习率降低了 10 倍。其他 FCOS 相关参数均与 MMrotate 上的 rotated\_FCOS 配置保持一致。

### 3.4 消融实验

为了验证本文所提出的标签分配策略的有效性和泛化性, 本文在 2 个大规模遥感图像数据集 DOTA 1.0 和 DIOR-R 数据集上对所提方法进行了

单尺度的消融实验(为了方便评估,在 DOTA 数据集的消融实验中,仅将训练集进行训练,将验证集作为测试集)。

以 MMrotate 上集成的 rotated\_FCOS 为基线,依次加入基于椭圆方程的形状自适应采样策略、椭圆中心度量和联合 IoU 策略,结果如表 1 所示。

表 1 在 DOTA 和 DIOR-R 上的消融实验 (%)

	DOTA		DIOR-R
	mAP	mAP <sub>50:95</sub>	mAP
基线	65.53	35.65	62.53
+ 形状自适应采样	67.65	37.20	63.46
+ 椭圆中心度量	68.15	37.46	64.43
+ 联合 IoU 指导	68.61	38.68	65.21

对于 DOTA 数据集,基线的 mAP 为 65.53%,加入形状自适应采样后上涨了 2.12%;添加椭圆中心度量方法后,精度又提升了 0.5%;进一步加入联合 IoU 指导策略,mAP 进一步提升了 0.46%。mAP<sub>50:95</sub> 从基线的 35.65% 提升到 38.68%。对于 DIOR-R 数据集,mAP 从基线的 62.53% 提升到 65.21%。图 5 展示了所提方法与 FCOS 的检测对比图,可以看出所提方法对目标回归的结果更加准确且拥有更高的置信度。这表明了本文提出的标签分配策略能够为模型挖掘出更高质量的锚点,并为选出的锚点分配更准确合理的权重,联合IoU指导策略也能尽可能

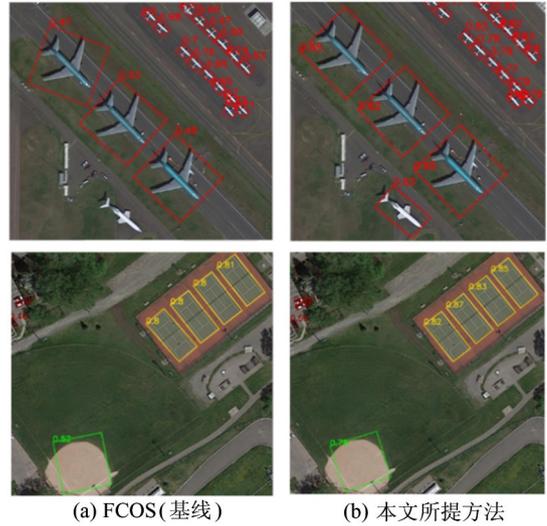


图 5 检测效果对比图

避免分类与回归的不一致,进而提升检测精度。

### 3.5 实验结果

为进一步证明所提方法的有效性,本文在更多遥感数据集上进行了广泛实验,并与其他先进的旋转目标检测方法进行了比较。

(1)DOTA1.0 上的结果。对于旋转目标任务中,大多数方法都使用 DOTA 遥感数据集进行性能比较和分析。消融实验已经充分表明了所提方法的有效性和通用性。为了评估所提方法的最终性能,选择 Train 和 Val 的图像作为训练集,Test 中的图像为测试集,最终结果由官方服务器评估后给出,并与目前先进的旋转目标检测器进行比较,结果如表 2

表 2 不同方法在 DOTA 1.0 数据集上的检测精度 (%)

方法	主干	MS	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
RRPN <sup>[15]</sup>	R101	✓	88.52	71.20	31.65	59.30	51.85	56.19	57.25	90.81	72.84	67.38	56.69	52.84	53.08	51.94	53.58	61.01
ICN <sup>[39]</sup>	R101	✓	81.36	74.30	47.70	70.32	64.89	67.82	69.98	90.76	79.06	75.20	53.64	62.90	67.02	64.17	50.23	68.16
RoI Trans <sup>[17]</sup>	R101	✓	88.64	78.52	43.44	75.92	68.81	73.68	83.59	90.74	77.27	81.46	58.39	53.54	62.83	58.93	47.67	69.56
DAL <sup>[28]</sup>	R101	✓	88.61	79.69	46.27	70.37	65.89	76.10	78.53	90.84	79.98	78.41	58.71	62.02	69.23	71.32	60.65	71.78
RSDet <sup>[40]</sup>	R101	✓	89.90	82.90	48.60	65.20	69.50	70.10	70.20	90.50	85.60	83.40	62.50	63.90	65.60	67.20	68.00	72.20
SCRDet <sup>[41]</sup>	R101	✓	89.98	80.65	52.09	68.36	68.36	60.32	72.41	90.85	87.94	86.86	65.02	66.68	66.25	68.24	65.21	72.61
CFC Net <sup>[42]</sup>	R50	✓	89.08	80.41	52.41	70.02	76.28	78.11	87.21	90.89	84.47	85.64	60.51	61.52	67.82	68.02	50.09	73.50
CV <sup>[19]</sup>	R101		89.64	85.00	52.26	77.34	73.01	73.14	86.82	90.74	79.02	86.81	59.55	70.91	72.94	70.86	57.32	75.02
BBAVector <sup>[10]</sup>	R101	✓	88.63	54.06	52.13	69.56	78.26	80.40	88.06	90.87	87.23	86.39	56.11	65.63	67.10	72.08	63.96	75.36
CSL <sup>[22]</sup>	R152	✓	90.25	85.53	54.64	75.31	70.44	73.51	77.62	90.84	86.15	86.69	69.60	68.04	73.83	71.10	68.93	76.17
R3Det <sup>[43]</sup>	R152	✓	89.80	83.77	48.44	66.77	78.76	83.27	87.84	90.82	85.38	85.51	65.57	62.68	67.53	78.56	72.67	76.47
SCRDet ++ <sup>[44]</sup>	R101	✓	90.05	84.39	55.44	73.99	77.54	71.11	86.05	90.67	87.32	87.08	69.62	68.90	73.74	71.29	65.08	76.81
DCL <sup>[23]</sup>	R152	✓	89.26	83.60	53.54	72.76	79.04	82.56	87.31	90.67	86.59	86.98	67.49	66.88	73.29	70.56	69.99	77.37
S2ANet <sup>[20]</sup>	R101	✓	89.28	84.11	56.95	79.21	80.18	82.93	89.21	90.86	84.66	87.61	71.66	68.23	78.58	78.20	65.55	79.15
FCOS(base)	R50		88.74	75.77	50.07	62.10	80.20	79.41	87.34	90.87	81.52	84.83	53.75	67.81	66.73	69.63	44.08	72.19
本文算法	R50		89.46	76.67	54.80	75.90	80.69	79.63	87.64	90.90	85.77	85.16	63.44	67.54	69.76	72.06	49.61	75.27
本文算法	R50	✓	89.24	84.50	53.76	79.61	82.56	84.58	88.04	90.81	85.13	86.15	67.18	69.53	76.54	80.22	69.63	79.17

所示。可以看出,在 MMrotate 上 FCOS 的基线指标为 72.19%,在同等条件下,所提出的标签分配方法在此基础上提升了 3.08%,达到了 75.27%,且对于该数据集下的每一类目标都有不同程度的提升,这表明所提的标签分配策略确实能为模型选择更高质量

的锚点。此外,在多尺度训练和测试的加持下,所提方法的 mAP 最终达到了 79.17%,在现有无锚框的检测方法中取得了非常有竞争力的结果。其在 DOTA 数据集上的检测结果如图 6 所示。



图 6 本文方法在 DOTA 1.0 数据集上的验证可视化效果图

(2)其他数据集上的结果。本文进一步在 DIOR-R、HRSC2016、UCAS-AOD 等公开的旋转目标遥感数据集上进行全面验证,并与其他方法进行对比。如表 3 所示,所提方法在更大规模的 DIOR-R 数据集上相比于基线提升了 3.08%,达到了 65.21%,优于绝大部分现有检测器模型。对于具有较大长宽比

的船舰数据集 HRSC2016 以及长宽比较小的汽车飞机数据集 UCAS-AOD,所提方法在 mAP 上相比于基线分别提升了 1.23% 和 0.77%,分别达到了 90.24%、90.17%,超越了绝大部分现有的方法,充分表明了所提方法对于任意长宽比的目标都具有不错的提升效果。其对比结果分别如表 4 和 5 所示。值得注

表 3 不同方法在 DIOR-R 数据集上的检测精度 (%)

方法	主干	APL	APO	BF	BC	BR	CH	DAM	ETS	ESA	GF	GTF	HA	OP	SH	STA	STO	TC	TS	VE	WM	mAP
Faster-RCNN-O	R-50	62.79	26.8	71.72	80.91	34.2	72.57	18.95	66.45	65.75	66.63	79.24	34.95	48.97	81.14	64.34	71.21	81.44	47.31	50.46	65.21	59.54
RetinaNet-O <sup>[45]</sup>	R-50	61.49	28.52	73.57	81.17	23.98	72.54	19.94	72.39	58.2	69.25	79.54	32.14	44.87	77.71	67.57	61.09	81.46	47.33	38.01	60.24	57.55
Gliding Vetex <sup>[19]</sup>	R-50	63.53	28.87	74.96	81.33	33.88	74.31	19.58	70.72	64.7	72.3	78.68	37.22	49.64	80.22	69.26	61.13	81.49	44.76	47.71	65.04	60.06
RoI Trans <sup>[17]</sup>	R-50	65.34	37.88	71.78	87.53	40.68	72.6	26.86	78.21	68.09	68.96	82.74	47.71	55.61	81.21	78.23	70.26	81.61	54.86	43.27	65.52	63.87
AOPG <sup>[46]</sup>	R-50	62.39	37.79	71.62	87.63	40.90	72.47	31.08	65.42	77.99	73.20	81.94	42.32	54.45	81.17	72.69	71.31	81.49	60.04	52.38	69.99	64.41
本文算法	R-50	68.52	40.57	71.52	89.52	39.41	77.06	26.46	70.27	80.01	81.18	75.60	37.59	55.33	89.06	74.01	76.40	87.95	46.72	50.89	66.14	65.21

表 4 不同方法在 HRSC2016 数据集上的检测精度

方法	主干	mAP/%
RoI-Trans <sup>[17]</sup>	ResNet-101	86.20
Gliding Vertex <sup>[19]</sup>	ResNet-101	88.20
BBAVectors <sup>[10]</sup>	ResNet-101	88.60
R <sup>3</sup> Det <sup>[43]</sup>	ResNet-101	89.26
R <sup>3</sup> det-DCL <sup>[23]</sup>	ResNet-101	89.46
FPN-CSL <sup>[22]</sup>	ResNet-101	89.62
DAL <sup>[28]</sup>	ResNet-101	89.77
R <sup>3</sup> Det-GWD <sup>[21]</sup>	ResNet-101	89.85
S <sup>2</sup> ANet <sup>[20]</sup>	ResNet-101	90.17
FCOS(base)	ResNet-50	89.01
本文算法	ResNet-50	90.24

表 5 不同方法在 UCAS\_AOD 数据集上的检测精度 (%)

方法	小汽车	机场	mAP
YOLOv3	74.63	89.52	82.08
RetinaNet <sup>[45]</sup>	84.64	90.51	87.57
FR-O	86.87	89.86	88.36
RoI Trans <sup>[17]</sup>	87.99	89.90	88.95
FCOS(base)	88.18	90.62	89.40
本文算法	89.78	90.56	90.17

意的是,由于 HRSC2016 和 UCAS-AOD 数据集规模较小,整体精度较高,因此提升的空间相对较小。综上所述,本文在多个数据集上进行了广泛而深入的实验,验证了所提出的标签分配策略的有效性。

## 4 结论

本文提出了一种新的标签分配策略,该策略主要由基于椭圆方程的形状自适应采样策略、椭圆中心度量以及联合 IoU 指导 3 部分组成。针对遥感图像中目标任意方向以及长宽比变化大的问题,所提出的采样策略根据目标框的长宽比自适应地调整采样区域,以提供更加高质量的锚点。椭圆中心度量策略根据所提出的采样区域重新分配锚点权重,并动态联合 IoU,使得新的权重更加合理。在 DOTA 等大量数据集上取得了卓越的结果并验证了所提方法的有效性和优越性。但是,该方法对于遥感图像中的一些小目标的检测效果仍不理想,未来工作将

探索遥感图像中的小目标检测。

## 参考文献

- [ 1 ] GIRSHICK R. FastR-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015:1440-1448.
- [ 2 ] 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.
- [ 3 ] 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, USA: IEEE, 2016: 779-788.
- [ 4 ] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017:7263-7271.
- [ 5 ] TIAN Z, SHEN C, CHEN H, et al. FCOS: a simple and strong anchor-free object detector[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020,44: 1922-1933.
- [ 6 ] GE Z, LIU S, WANG F, et al. Yolox:exceeding yolo series in 2021[EB/OL]. (2021-07-18)[2022-12-11]. <https://arxiv.org/pdf/2107.08430.pdf>.
- [ 7 ] LAW H, DENG J. Cornernet:detecting objects as paired keypoints[C]//Proceedings of the European Conference on Computer Vision. Munich, Germany: Springer, 2018: 734-750.
- [ 8 ] KONG T, SUN F, LIU H, et al. Foveabox:beyond anchor-based object detection[J]. IEEE Transactions on Image Processing, 2020,29:7389-7398.
- [ 9 ] DUAN K, BAI S, XIE L, et al. Centernet:keypoint triplets for object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Seoul, Korea: IEEE, 2019:6569-6578.
- [ 10 ] YI J, WU P, LIU B, et al. Oriented object detection in aerial images with box boundary-aware vectors[C]//Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Virtual: IEEE, 2021:2150-2159.
- [ 11 ] XIA G S, BAI X, DING J, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 3974-3983.
- [ 12 ] LI K, WAN G, CHENG G, et al. Object detection in optical remote sensing images:a survey and a new bench-

- mark[J]. *ISPRS Journal of Photo Grammetry and Remote Sensing*, 2020,159:296-307.
- [13] LIU Z, YUAN L, WENG L, et al. A high resolution optical satellite image dataset for ship recognition and some new baselines[C]//International Conference on Pattern Recognition Applications and Methods. Porto, Portuguese: SciTePress, 2017,2:324-331.
- [14] ZHU H, CHEN X, DAI W, et al. Orientation robust object detection in aerial images using deep convolutional neural network[C]//2015 IEEE International Conference on Image Processing. Quebec, Canada: IEEE, 2015: 3735-3739.
- [15] NABATI R, QI H. RRPN: radar region proposal network for object detection in autonomous vehicles[C]//2019 IEEE International Conference on Image Processing. Taipei, China: IEEE, 2019:3093-3097.
- [16] JIANG Y, ZHU X, WANG X, et al. R2CNN: rotational region CNN for orientation robust scene text detection [EB/OL]. (2017-06-29) [2022-12-11]. <https://arxiv.org/ftp/arxiv/papers/1706/1706.09579>.
- [17] DING J, XUE N, LONG Y, et al. Learning RoI transformer for oriented object detection in aerial images[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2019: 2849-2858.
- [18] WANG J, DING J, GUO H, et al. Mask OBB: a semantic attention-based mask oriented bounding box representation for multi-category object detection in aerial images [J]. *Remote Sensing*, 2019,11(24):2930.
- [19] XU Y, FU M, WANG Q, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020,43(4):1452-1459.
- [20] HAN J, DING J, LI J, et al. Align deep features for oriented object detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021,60:1-11.
- [21] YANG X, YAN J, MING Q, et al. Rethinking rotated object detection with Gaussian Wasserstein distance loss [C]//International Conference on Machine Learning. Virtual: ACM, 2021:11830-11841.
- [22] YANG X, YAN J. Arbitrary-oriented object detection with circular smooth label[C]//European Conference on Computer Vision. Virtual: Springer, 2020:677-694.
- [23] YANG X, HOU L, ZHOU Y, et al. Dense label encoding for boundary discontinuity free rotation detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Virtual: IEEE, 2021: 15819-15829.
- [24] 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 Conference on Computer Vision and Pattern Recognition. Virtual: IEEE, 2020:9759-9768.
- [25] ZHANG X, WAN F, LIU C, et al. FreeAnchor: learning to match anchors for visual object detection[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2019:147-155.
- [26] ZHU B, WANG J, JIANG Z, et al. AutoAssign: differentiable label assignment for dense object detection[EB/OL]. (2020-07-07) [2022-12-11]. <https://arxiv.org/pdf/2007.03496>.
- [27] KIM K, LEE H S. Probabilistic anchor assignment with IoU prediction for object detection[C]//European Conference on Computer Vision. Virtual: Springer, 2020: 355-371.
- [28] MING Q, ZHOU Z, MIAO L, et al. Dynamic anchor learning for arbitrary-oriented object detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Virtual: AAAI Press, 2021:2355-2363.
- [29] GE Z, LIU S, LI Z, et al. OTA: optimal transport assignment for object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Virtual: IEEE, 2021:303-312.
- [30] LI S, HE C, LI R, et al. A dual weighting label assignment scheme for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022:9387-9396.
- [31] XU C, WANG J, YANG W, et al. RFLA: Gaussian receptive field based label assignment for tiny object detection[C]//European Conference on Computer Vision. Tel Aviv, Israel: Springer, 2022:526-543.
- [32] HOU J B, ZHU X, YIN X C. Self-adaptive aspect ratio anchor for oriented object detection in remote sensing images[J]. *Remote Sensing*, 2021,13(7):1318.
- [33] HOU L, LU K, XUE J, et al. Shape-adaptive selection and measurement for oriented object detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Virtual: AAAI Press, 2022:923-932.
- [34] JIANG B, LUO R, MAO J, et al. Acquisition of localization confidence for accurate object detection [C]//Proceedings of the European Conference on Computer Vision. Munich, Germany: Springer, 2018:784-799.
- [35] WU S, LI X, WANG X. IoU-aware single-stage object detector for accurate localization [J]. *Image and Vision Computing*, 2020,97:103911.

- [36] LI X, WANG W, WU L, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection[J]. *Advances in Neural Information Processing Systems*, 2020,33:21002-21012.
- [37] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (voc) challenge[J]. *International Journal of Computer Vision*, 2010, 88: 303-338.
- [38] ZHOU Y, YANG X, ZHANG G, et al. Mmrotate: a rotated object detection benchmark using pytorch[C]// *Proceedings of the 30th ACM International Conference on Multimedia*. Lisbon, Portuguese: ACM, 2022: 7331-7334.
- [39] AZIMI S M, VIG E, BAHMANYAR R, et al. Towards multi-class object detection in unconstrained remote sensing imagery[C]// *Asian Conference on Computer Vision*. Perth, Australia: Springer, 2018:150-165.
- [40] QIAN W, YANG X, PENG S, et al. Learning modulated loss for rotated object detection[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. Virtual: AAAI Press, 2021,35(3):2458-2466.
- [41] YANG X, YANG J, YAN J, et al. Scrdet: towards more robust detection for small, cluttered and rotated objects [C]// *Proceedings of the IEEE International Conference on Computer Vision*. Seoul, Korea: IEEE, 2019:8232-8241.
- [42] MING Q, MIAO L, ZHOU Z, et al. CFC-Net: a critical feature capturing network for arbitrary-oriented object detection in remote-sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021,60:1-14.
- [43] YANG X, YAN J, FENG Z, et al. R3det: refined single-stage detector with feature refinement for rotating object [C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. Virtual: AAAI Press, 2021:3163-3171.
- [44] YANG X, YAN J, LIAO W, ET al. Scrdet++: detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022,45(2):2384-2399.
- [45] 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, Italy: IEEE, 2017:2980-2988.
- [46] CHENG G, WANG J, LI K, et al. Anchor-free oriented proposal generator for object detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022,60:1-11.

## Shape-adaptive ellipse label assignment for remote sensing image based on FCOS

YU Xinyi, LU Jiangping, LIN Mi, ZHOU Libo, OU Linlin

(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023)

### Abstract

Anchor-free object detection algorithms have experienced rapid development in object detection in recent years. However, in remote sensing images, the objects with arbitrary angles, dense distribution, and large shape differences make the detection still a challenge. Therefore, an anchor-free method based on improved fully convolutional one-stage (FCOS) is proposed. Firstly, to mine more potential high-quality anchor points, a shape-adaptive feature point sampling method based on the ellipse equation is proposed. To further reduce the negative influence of low-quality anchor points, the ellipse centerness is proposed. It can provide more accurate and reasonable weights than the traditional centerness. In addition, to address the inconsistency between classification and regression, a joint intersection over union (IoU) guidance strategy is proposed. The proposed ellipse centerness and IoU score are combined as quality scores to guide the training of the classification branch and to make the results of regression more accurate. The mean average precision on the DOTA 1.0 dataset reaches 79.17%, which is better than most existing anchor-free detection methods.

**Key words:** remote sensing image, deep learning, object detection, label assignment