

基于向量叉乘标签分配的遥感图像目标检测算法^①

禹鑫燚^② 林 密 卢江平 欧林林^③

(浙江工业大学信息工程学院 杭州 310023)

摘 要 近年来遥感图像目标检测受到了广泛的关注,主流的遥感图像目标检测器通过预设锚框与真实框之间的交并比(IoU)进行正负样本的划分。为了解决基于 IoU 的标签分配方法在遥感图像小而密集目标中存在复检和漏检的问题,本文提出了一种基于向量叉乘标签分配的遥感图像目标检测算法 YOLOXR。首先,提出了一种标签粗分配策略,通过向量叉乘的方法判断特征图的像素点是否在旋转目标内或者目标中心点附近的旋转正方形框内,从而确定其是否为候选正样本。其次,为了降低边缘低质量候选正样本对标签分配的影响,提出了旋转中心度量方法,通过向量叉乘判断像素点距离中心点的远近程度进而赋予不同的权重。最后,基于最优传输的方法(simOTA)选取真实框和样本点的最优匹配对,使得总体代价最小,进而为旋转目标分配合适的标签。此外,为了解决旋转 IoU 损失不可导以及 Smooth L1 损失难以权衡旋转框各个参数的问题,通过计算真实框和预测框二维高斯分布的 Kullback-Leibler 散度(KLD)来替代 IoU。在公开的遥感图像目标检测数据 DOTA、HRSC 2016 和 UCAS-AOD 上的大量实验表明,所提方法优于目前绝大多数旋转目标检测算法。

关键词 遥感图像;目标检测;标签分配;向量叉乘

目标检测是机器视觉中最具挑战性的问题之一。近年来随着深度卷积神经网络的快速发展,许多优秀的水平目标检测算法相继被提出,包括快速区域卷积神经网络(faster region-based convolution neural network, Faster-RCNN)^[1]、RetinaNet^[2]、全卷积单阶段目标检测(fully convolutional one-stage object detection, FCOS)^[3]和 YOLOX(you only look once X)^[4]等,这些算法在人脸检测、行人计数和自动驾驶等方面取得了较好的结果。与自然场景下的目标不同,遥感图像中的目标往往排列密集,尺度、长宽比以及方向变化比较大,因而水平目标检测算法难以在遥感图像中实现准确的检测^[5]。

近年来,许多优秀的遥感图像目标检测算法相继被提出,它们通过引入额外的角度来实现遥感图

像中任意方向目标的检测,即旋转目标检测。这些检测器首先预设大量的先验框(也称为锚点^[6])来实现与真实框的空间对齐,然后通过预设框与真实目标的交并比(intersection-over-union, IoU)来选择正负样本,最后根据损失来实现正样本的边界回归^[7]。其中,基于预设框和真实框 IoU 选择正负样本的过程被称为基于 IoU 的标签分配策略。由于遥感图像中目标在尺度、长宽比和方向上的变化很大,因此这种基于 IoU 的标签分配策略需要预设大量的锚框来更好地匹配真实目标,这极大地增加了训练过程的计算量,同时加剧了正负样本的不平衡。此外,在密集的目标检测中存在分类和回归不一致的问题,导致了目标的复检或者漏检,进而降低了模型的检测性能。

① 浙江省自然科学基金白马湖实验室区域创新发展联合基金(LBMHD24F030002)和国家自然科学基金(62373329)资助项目。

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

③ 通信作者,E-mail: linlinou@zjut.edu.cn。

(收稿日期:2022-02-21)

为了解决上述问题,本文提出了一种基于改进 YOLOX 的无锚框旋转目标检测算法。首先,在 YOLOX 边界回归的基础上通过引入角度信息使其变成旋转目标检测算法。然后,提出了一种基于向量叉乘的无锚框标签粗分配策略,通过判断特征图的像素点是否位于旋转目标内或者中心点附近的旋转正方形框内,进而确定其是否为候选正样本。此外,通过增加旋转中心度量的方法,使得正样本的分布更加趋近于中心点附近。最后,将基于向量叉乘的标签粗分配策略和中心点度量的方法添加到基于最优传输分配(sim-optimal transmission assignment, simOTA)的代价函数中,通过最优传输得到最终的正负样本。本文的主要贡献如下。

(1) 提出了一种简单且快速的单阶段无锚框遥感图像目标检测算法 YOLOXR,基于向量叉乘的标签粗分配策略,可以为训练样本分配更加合适的候选正样本。

(2) 为了降低旋转目标边缘低质量样本的权重,基于旋转框向量叉乘的比值,设计了旋转中心度量方法,使得正样本更加趋近于中心点附近。

(3) 提出的方法在 DOTA v1.0、HRSC 2016 和 UCAS-AOD 数据集上分别实现了 78.21%、97.11% 和 98.00% 的检测精度,超越了目前大多数旋转目标检测算法,并且保持了较高的检测速度。

1 网络结构设计

目前的旋转目标检测主要是通过生成旋转的锚框或者增加特征精练的两阶段方法来解决,但是这些方法增加了旋转框回归过程的推理时间和计算量。在本文的工作中,基于改进 YOLOX 设计了单阶段无锚框旋转目标检测器,与基于锚框的两阶段方法相比,该方法更简单、更快,而且取得了不错的检测性能。

1.1 YOLOX 网络结构

YOLOX^[4]是目前最先进的单阶段无锚框目标检测器之一,主要由 CSPDarkNet 53、FPN + PAN 和 Decoupled Head 这 3 个部分构成。

YOLOX 的主干检测网络延续了 YOLO v4^[8]和

YOLO v5 的主干网络,由跨阶段局部网络(cross stage partial network, CSP^[9])和 DarkNet 53^[10]构成。为了解决网络优化过程中的梯度信息重复导致推理计算过高的问题,本文采用了 CSP 模块,将网络分成基础层和映射层 2 部分,然后通过跨阶段层次结构将它们合并,在保证准确率的同时可以有效降低模型的计算量。CSPDarkNet 53 在 DarkNet 53 的基础上加入了 CSP 模块,在提高卷积神经网络学习能力的同时,有效地降低了计算量,提升了模型的推理速度。

YOLOX 的颈部网络由特征金字塔网络(feature pyramid network, FPN)^[11]和路径聚合网络(path aggregation network, PAN)这 2 部分组成。FPN 也叫特征金字塔,是针对图像目标中多尺度的特点提出的,其将高层特征通过上采样和低层特征融合得到进行预测的特征图,为每一层特征图赋予了更强的捕获语义信息的能力。然而 FPN 的自底向上的融合方式略微简单,只是将高层的语义信息传递到低层,而低层的纹理信息并没有传递到高层。为此,YOLOX 在 FPN 层的后面还添加了一个自底向上的特征金字塔 PAN。FPN 把深层的语义特征传到浅层,从而增强多个尺度上的语义表达,而 PAN 则把浅层的定位信息传导到深层,增强多个尺度上的定位能力,通过融合自底向上和自顶向下 2 条路径的方式增强了骨干网络的表征能力。

YOLOX 的头部网络采用了解耦头 Decoupled Head,将检测和分类分开输出,由于其具有更好的表达能力,所以可以加速网络的收敛并且提高检测精度。此外,YOLOX 首先通过寻找目标边界框内的像素点以及目标中心点附近的像素点作为正样本的候选样本,然后将标签分配问题转化为最优传输问题,通过求解真实框和预测框的最优匹配对,进而为目标匹配合适的标签。

1.2 YOLOXR 网络结构设计

由于 YOLOX 是水平框的目标检测算法,而遥感图像的目标检测是旋转框目标检测,因此本文在保留 YOLOX 的基础网络以及特征聚合网络情况下,主要对 YOLOX 的头部网络进行修改,使其变成旋转框的目标检测器 YOLOXR,其网络结构图如

图 1 所示。

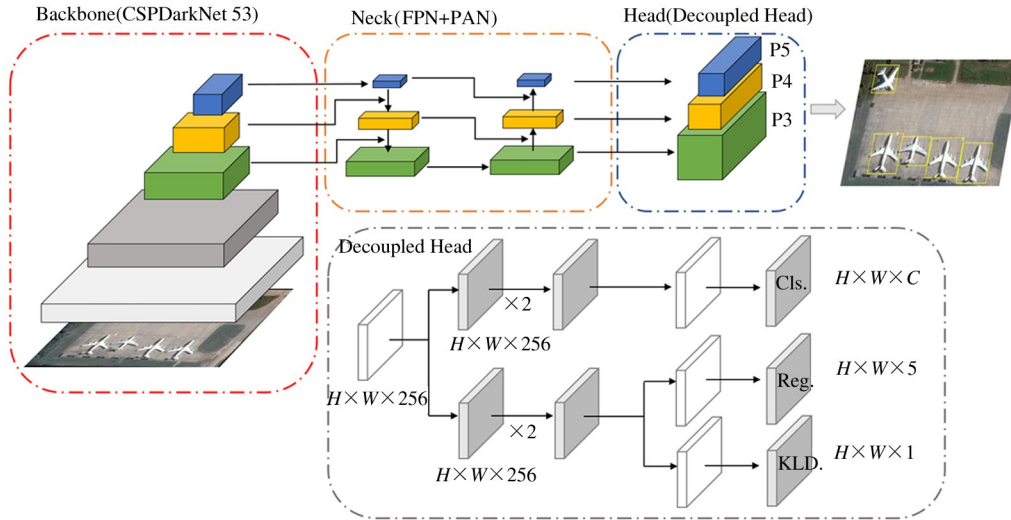


图 1 YOLOXR 网络结构图

首先,YOLOX 目标检测是针对水平框设计的目标检测器,用 (x, y, w, h) 来表示水平框,而本文设计的 YOLOXR 是针对旋转框设计的目标检测,相比于 YOLOX 多了一个角度信息,因而采用 (x, y, w, h, θ) 来表示旋转框。其次,YOLOX 的标签分配策略是针对水平框设计的,用在旋转目标检测中显然不合适,为此,YOLOXR 设计了新的标签分配策略。通过判断特征图的像素点是否位于目标旋转框内或是否位于目标中心点附近的旋转正方形框内,进而确定其是否为候选正样本。此外,通过向量叉乘判断像素点距离中心点的远近程度进而赋予不同的权重,使得标签分配更加趋向于中心点附近。同样地,YOLOX 采用了 simOTA 方法来确定最终的正负样本。最后,修改了 YOLOX 的损失函数。由于 IoU 损失函数在水平框检测中具有较强的适用性,但在旋转目标检测中 IoU 损失函数不可导,而 Smooth L1^[12] 损失函数又不能准确地表示旋转框的各个参数之间的联系,因此,YOLOXR 使用了 Kullback-Leibler 散度 (Kullback-Leibler divergence, KLD)^[13] 损失函数来代替原本的 IoU 损失函数,从而有效地解决了上述问题。

2 基于向量叉乘的标签分配策略

目前大多数旋转目标检测器都是基于锚框的,

通过预设锚框与真实框之间的 IoU 进行正负样本的划分,IoU 大于预设阈值的样本则作为正样本,否则作为负样本。这种基于锚框的正负样本分配方法需要预设一定数量的锚框,在旋转目标检测中包括尺寸、长宽比和角度不同的锚框,这极大地增加了训练过程中的计算量和时间。此外,基于 IoU 标签分配策略在密集的目标检测中存在分类和回归不一致的问题,进而导致了目标的复检或者漏检。

为了解决上述问题,本文基于 YOLOX 设计了旋转目标检测器 YOLOXR,并提出了基于旋转框的标签分配方法。首先通过向量叉乘的方法来判断像素点是否位于真实旋转框或中心点附近的旋转正方形框内,初步筛选出正负样本;然后根据向量叉乘的最小值和最大值之间的比值来确定特征图中像素点距离中心的远近程度,进而赋予不同的权重;最后采用 simOTA 方法来确定最终的正负样本。

2.1 基于向量叉乘的标签粗分配策略

对于任意给定的一个旋转真实框 (x, y, w, h, θ) ,其坐标可以转化为八参数表示为 $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$,按照顺时针的顺序对其排序,如图 2 所示。通过向量叉乘判断特征图的像素点是否位于旋转真实框或者目标中心点附近的旋转正方形框内,初步筛选出满足该条件的像素点作为候选正样本,其余的作为负样本。图 2 表示的是遥感图

像中的一艘船,通过向量叉乘判断特征图的像素点是否位于这艘船所在的真实框以及它的中心点附近的正方形框内,如果在则将这个像素点标注出来(如图2所标注的像素点),作为候选正样本,其余的则作为负样本处理。

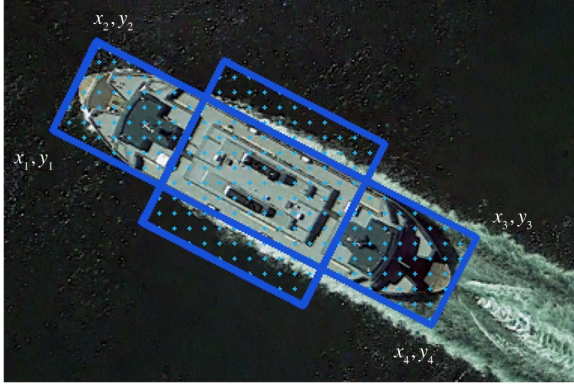


图2 标签粗分配策略示意图

YOLOX 检测器定义了 $\{P_3, P_4, P_5\}$ 3 个级别的特征图,它们的步长 $\{s_3, s_4, s_5\}$ 分别为 8、16 和 32。选取任意一层特征图上的任意一个像素点 P ,其像素坐标值可以表示为 (x_p, y_p) 。如图3所示,选取出真实框的任意一个顶点 A ,它与 P 组成的向量为 $\overrightarrow{AP}$,另外,它与顺时针方向的下个顶点 B 组成向量为 $\overrightarrow{AB}$, $\overrightarrow{AB} \times \overrightarrow{AP}$ 的值可以用式(1)来表示。

$$\overrightarrow{AB} \times \overrightarrow{AP} = (x_2 - x_1, y_2 - y_1) \times (x_p - x_1, y_p - y_1) \quad (1)$$

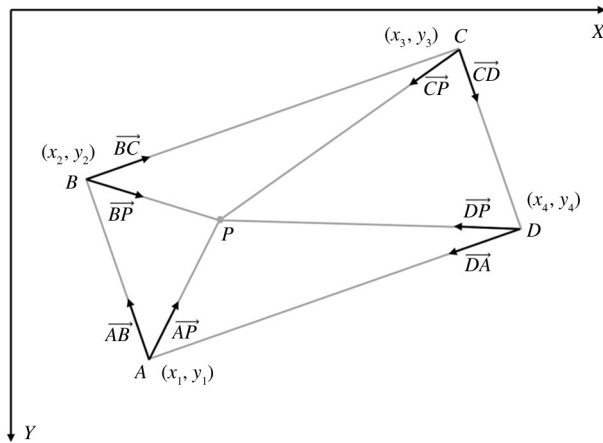


图3 像素点和旋转框的位置关系

同样地,选取出真实框顶点 A 的对角顶点 C ,它与 P 组成的向量为 $\overrightarrow{CP}$,此外,顶点 C 顺时针方向的

下个顶点为 D ,它们所组成的向量为 $\overrightarrow{CD}$, $\overrightarrow{CD} \times \overrightarrow{CP}$ 的值如式(2)所示。

$$\overrightarrow{CD} \times \overrightarrow{CP} = (x_4 - x_3, y_4 - y_3) \times (x_p - x_3, y_p - y_3) \quad (2)$$

联立式(1)和式(2),即可得到 $(\overrightarrow{AB} \times \overrightarrow{AP}) \cdot (\overrightarrow{CD} \times \overrightarrow{CP})$ 的值,该值的正负决定了点 P 是否位于直线 AB 和直线 CD 之间。当 $(\overrightarrow{AB} \times \overrightarrow{AP}) \cdot (\overrightarrow{CD} \times \overrightarrow{CP}) > 0$ 时,点 P 位于直线 AB 和直线 CD 之间;当 $(\overrightarrow{AB} \times \overrightarrow{AP}) \cdot (\overrightarrow{CD} \times \overrightarrow{CP}) = 0$ 时,点 P 位于直线 AB 或者直线 CD 之上;当 $(\overrightarrow{AB} \times \overrightarrow{AP}) \cdot (\overrightarrow{CD} \times \overrightarrow{CP}) < 0$ 时,点 P 位于直线 AB 和直线 CD 之外。因此,点 P 和直线 AB 以及直线 CD 的关系可以用式(3)来表示。

$$\begin{cases} (\overrightarrow{AB} \times \overrightarrow{AP}) \cdot (\overrightarrow{CD} \times \overrightarrow{CP}) > 0 & P \in \{AB, CD\} \\ (\overrightarrow{AB} \times \overrightarrow{AP}) \cdot (\overrightarrow{CD} \times \overrightarrow{CP}) = 0 & P \in AB \vee P \in CD \\ (\overrightarrow{AB} \times \overrightarrow{AP}) \cdot (\overrightarrow{CD} \times \overrightarrow{CP}) < 0 & P \in \text{其他情况} \end{cases} \quad (3)$$

由于旋转目标 $ABCD$ 是一个矩形框, AB 与 CD 是 2 条相互平行的直线,而 BC 与 DA 也是相互平行的 2 条直线,同理可得,点 P 和直线 BC 和直线 DA 的关系为

$$\begin{cases} (\overrightarrow{BC} \times \overrightarrow{BP}) \cdot (\overrightarrow{DA} \times \overrightarrow{DP}) > 0 & P \in \{BC, DA\} \\ (\overrightarrow{BC} \times \overrightarrow{BP}) \cdot (\overrightarrow{DA} \times \overrightarrow{DP}) = 0 & P \in BC \vee P \in DA \\ (\overrightarrow{BC} \times \overrightarrow{BP}) \cdot (\overrightarrow{DA} \times \overrightarrow{DP}) < 0 & P \in \text{其他情况} \end{cases} \quad (4)$$

通过式(3)和式(4),可以确定点 P 是否位于 AB 和 CD 之间以及 BC 和 DA 之间,进而可以确定点 P 是否在旋转真实框 $ABCD$ 内部。对于特征图上的任意一个坐标点 P ,如果其满足:

$$\begin{cases} (\overrightarrow{AB} \times \overrightarrow{AP}) \cdot (\overrightarrow{CD} \times \overrightarrow{CP}) > 0 \\ \text{并且} \\ (\overrightarrow{BC} \times \overrightarrow{BP}) \cdot (\overrightarrow{DA} \times \overrightarrow{DP}) > 0 \end{cases} \quad (5)$$

则点 P 位于旋转框 $ABCD$ 的内部,它是一个正样本的候选样本;否则,点 P 位于旋转框的边上或者外部,需要进一步判断其是否位于目标中心点附近的旋转正方形框内。

已知任意一个旋转目标可以表示为 (x, y, w, h, θ) ,则它的中心点坐标为 (x, y) ,其旋转角度为 θ 。根据这个旋转角度,向该角度的特征图上下左右

扩张 3 个像素点,得到一个正方形的旋转框 $EFGH$, 该旋转框的 4 个顶点坐标 $(x_5, y_5, x_6, y_6, x_7, y_7, x_8, y_8)$ 分别为

$$\begin{cases} x_5 = (-cendis) \cdot \cos(\theta) - (-cendis) \cdot \sin(\theta) + x \\ y_5 = (-cendis) \cdot \sin(\theta) + (-cendis) \cdot \cos(\theta) + y \\ x_6 = cendis \cdot \cos(\theta) - (-cendis) \cdot \sin(\theta) + x \\ y_6 = cendis \cdot \sin(\theta) + (-cendis) \cdot \cos(\theta) + y \\ x_7 = cendis \cdot \cos(\theta) - cendis \cdot \sin(\theta) + x \\ y_7 = cendis \cdot \sin(\theta) + cendis \cdot \cos(\theta) + y \\ x_8 = (-cendis) \cdot \cos(\theta) - cendis \cdot \sin(\theta) + x \\ y_8 = (-cendis) \cdot \sin(\theta) + cendis \cdot \cos(\theta) + y \end{cases} \quad (6)$$

其中, $cendis$ 表示旋转正方形框距离中心点的距离, 由于如图 1 所示的 $\{P_3, P_4, P_5\}$ 3 个特征图步长分别为 8、16 和 32, 因此它在 3 个特征图中表示的距离分别为 24、48 和 96。

在确定旋转正方形框 4 个顶点的坐标之后, 可以通过向量叉乘的方法来判断像素点 $P(x_p, y_p)$ 是否存在于旋转正方形框内。因此, 对于特征图上的任意一个坐标点 P , 如果其满足:

$$\begin{cases} (\vec{EF} \times \vec{EP}) \cdot (\vec{GH} \times \vec{GP}) > 0 \\ \text{并且} \\ (\vec{FG} \times \vec{FP}) \cdot (\vec{HE} \times \vec{HP}) > 0 \end{cases} \quad (7)$$

则点 P 位于旋转框 $EFGH$ 的内部, 它是一个正样本的候选样本, 否则, 点 P 作为负样本处理。

综上所述, 对于位于特征图上的任意一个像素点, 只要其满足式(5)或式(7), 那么该像素点就是正样本的候选点, 否则该像素点作为负样本处理。基于向量叉乘的标签粗分配策略如算法 1 所示。

2.2 基于向量叉乘的旋转中心度量方法

基于向量叉乘的标签粗分配方法能够判断特征图的像素点是否位于真实框或者真实框中心点附近的旋转正方形框内, 但是不能判断这个像素点距离中心点的远近程度, 进而造成远离中心点的边缘低质量样本参与正样本的回归, 影响到最终的检测精度。为此, 本文提出了一种简单而有效的策略来抑制这些检测到的低质量边界框, 而无需引入任何超参数。本文在代价函数之中添加了与分类损失和位

置回归损失并行的旋转中心度量损失, 使得标签分配更加趋近于中心点附近的正样本, 进而降低边缘低质量样本对标签分配造成的影响。

算法 1 基于向量叉乘的标签粗分配策略

输入:

G 是真实框的集合

P 是头部网络特征图中像素点的集合

输出:

S_p 是候选正样本的集合

S_N 是负样本的集合

```

1: for 每个真实框  $g \in G$  do
2:   计算真实框  $\square ABCD$  的 4 个顶点坐标  $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$ ;
3:   根据式(6), 计算真实框中心点附近的旋转正方形框  $\square EFGH$  的 4 个顶点坐标  $(x_5, y_5, x_6, y_6, x_7, y_7, x_8, y_8)$ ;
4:   for 每个像素点  $p \in P$  do
5:     根据式(5), 计算像素点  $p$  和真实框  $\square ABCD$  向量叉乘的值  $V_A$ ;
6:     根据式(7), 计算像素点  $p$  和旋转正方形框  $\square EFGH$  向量叉乘的值  $V_E$ ;
7:     if  $V_A > 0 \vee V_E > 0$  then
8:        $S_p = S_p \cup p$ ;
9:     else
10:       $S_N = S_N \cup p$ ;
11:     end if
12:   end for
13: end for
14: return  $S_p, S_N$ 

```

由 2.1 节可以得知, 对于任意一个像素点 P , 它与 4 个顶点 $ABCD$ 组成向量之间叉乘的值分别为 $\vec{AB} \times \vec{AP}, \vec{CD} \times \vec{CP}, \vec{BC} \times \vec{BP}$ 和 $\vec{DA} \times \vec{DP}$, 为了方便描述, 使用 $\vec{a}, \vec{c}, \vec{b}$ 和 $\vec{d}$ 来分别表示它们。由于向量叉乘的模表示的是 2 个向量所围成的平行四边形的面积, 当点 P 越靠近真实框的中心点时, 向量 $\vec{a}, \vec{c}, \vec{b}$ 和 $\vec{d}$ 表示的平行四边形面积越接近, 也就是它们的模相近; 而当点 P 越远离中心点时, 向量 $\vec{a}, \vec{c}, \vec{b}$ 和 $\vec{d}$ 中靠近 P 点的平行四边形面积比较小, 而远离 P 点的平行四边形面积比较大, 进而造成它们的模相差较大。因此, YOLOXR 的旋转中心度量 (center-ness) 公式为

$$center_ness = \frac{\min(\vec{a}, \vec{b}, \vec{c}, \vec{d})}{\max(\vec{a}, \vec{b}, \vec{c}, \vec{d})} \quad (8)$$

由于图像坐标系相比于通常的笛卡尔坐标系是关于 X 轴对称的,因此,在图像坐标系下顺时针排序的真实框,换算成常用的笛卡尔坐标系下是逆时针,这也导致了任何位于图像真实框内的点,它们的向量 $\vec{a}$ 、 $\vec{c}$ 、 $\vec{b}$ 和 $\vec{d}$ 均为正数并且位于同一个方向。因此,在式(8)中,直接采用它们的最大最小值来计算该点距离中心点的远近程度,而无需对它们进行取模操作。图4给出了基于向量叉乘的中心度量方法的取值情况,当距离中心点较近时,中心值越高,距离中心点越远时,中心值越低。

在得到中心值之后,相应的中心度量损失($center_ness$ loss)为 $center_ness_loss = 1 - center_ness$,它表示的是像素点 P 距离中心点的远近程度。像素点 P 距离中心点越远,损失越大,它在代价矩阵的值也越大,因此它成为正样本的可能性也就越小。

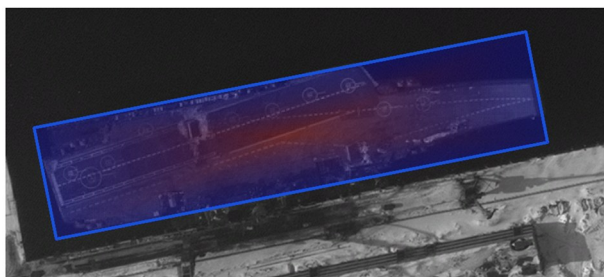


图4 基于向量叉乘的旋转中心度量方法

2.3 基于最优传输的标签分配

目前大多数的正负样本分配是根据预选框和真实框的 IoU 来进行分配的,最优传输分配(optimal transmission assignment, OTA)^[14]认为目标在不同的大小、形状和遮挡条件下的正负样本的分配应该不一样,同时应该考虑上下文信息,因此将标签分配问题当成一个线性规划中的最优传输问题来处理,其核心思想是建立一个代价矩阵。假设有 M 个真实框和 N 个样本(特征图的像素点),那么代价矩阵的大小就为 $M \times N$,矩阵中的每个元素就是从不同真实框到不同样本点的代价(用损失来表示)。最优传输的目的是去选取真实框和样本点的匹配对,使得总体代价最小。

由于 OTA 使用 Sinkhorn-Knopp 迭代求解最优匹配对的过程,因此在训练过程中比较耗时,而 sim-

OTA^[4]直接用动态 top-k 的方式选取匹配对,在保证精度的情况下极大地降低了训练的成本。simOTA 首先获取当前真实框 KLD 值最大的 15 个结果,把这 15 个 KLD 进行求和并向下取整,得到当前真实框的正样本数量 D_p ;然后针对真实框挑选代价值最低的 D_p 个候选框;最后过滤掉共用的样本点,对于不同的真实框均为正样本的样本点,选取代价值较小的点作为正样本,其余的作为负样本。需要注意的是代价包括分类损失、回归损失以及 2.2 节所提到旋转中心度量损失。simOTA 能够自动地分析每个真实框需要正样本的数量,并且能够自动决定每个真实框从哪个特征图来检测。

3 损失函数构建

常见的水平框检测通常是通过 4 个偏移量的形式来对框的大小和位置进行预测的,借鉴于此,目前大多数的旋转目标检测在水平框的基础上加上了角度参数的回归,如常见的 Smooth L1^[12]损失。但是,这 5 个参数是独立优化的,这使得损失(或检测精度)对任何参数的欠拟合都很敏感。此外,不同类型的物体对这 5 个参数的灵敏度也不同。例如,角度参数对于检测大长宽比的物体是非常重要的,而中心点参数对小目标非常重要。这需要在训练过程中给一个特定目标样本选择一组适当的权重,这显然难以实现。

IoU 损失是水平框目标检测中最常用的损失函数,它有效地结合水平框的大小和位置信息,并且能反映出预测检测框与真实检测框的检测效果,更重要的是它具有尺度不变性。然而,在旋转目标检测中,旋转的 IoU 是不可导的,因此不能直接使用旋转的 IoU 来代替水平框的 IoU。为此,基于 KLD^[12]的思想,将旋转框转化为二维高斯分布,然后计算真实框和预测框二维高斯分布之间的 Kullback-Leibler 散度(KLD)作为回归损失,以此来代替 IoU 损失,有效解决旋转 IoU 损失不可导的问题,进而提升旋转目标的检测精度。

4 实验结果及分析

4.1 数据集

为了验证本文提出的 YOLOXR 的有效性,本文

在公开的遥感数据集 DOTA、HRSC 2016 和 UCAS-AOD 数据集上进行了实验。

DOTA^[15]是目前公开的最全面的用于目标检测的光学遥感图像数据集。DOTA v1.0 数据集包含来自不同卫星和传感器的 2 806 张图像,图像的大小在 800×800 和 $4\,000 \times 4\,000$ 的大小之间。该数据集标注了 15 个类别的 188 282 个实例,每个实例都由一个四边形边界框标注,顶点按顺时针顺序排列。官方使用 1/2 的图像作为训练集,1/6 作为验证集,1/3 作为测试集,测试集的标签没有公开,需要上传至服务器。

HRSC 2016^[16]数据集所有图像均来自 6 个著名的港口,包含海上船只和近海船只 2 种场景的图像。图像尺寸范围从 300×300 到 $1\,500 \times 900$,训练、验证和测试集分别包含 436 个图像(包括 1 207 个样本),181 个图像(包括 541 个样本)和 444 个图像(包括 1 228 个样本)。

USCA-AOD^[17]是中国科学院大学模式识别与智能系统开发实验室发布的用于车辆和飞机检测的遥感数据集。该数据集包含 1 510 张航空图像,分辨率大小为 $659 \times 1\,280$,共 14 596 个实例。本文随机选择了 1 110 张图像用于训练,400 张图像用于测试。

4.2 实验设置

本文所有的实验都是基于 YOLOX 来设计的,以 CSPDarkNet 53 作为特征提取骨干网络,并利用在 COCO^[18]数据集预训练的模型参数对网络进行初始化。本文在 DOTA、HRSC 2016 和 UCAS-AOD 数据集上分别训练了 150、150 和 200 个训练轮数,采用随机梯度下降法(stochastic gradient descent, SGD)进行训练,动量和权重衰减分别设置为 0.9 和 5×10^{-4} ,采用余弦退火学习率,大小为初始学习率 $lr \times \text{批量大小} / 64$,初始学习率为 0.01。由于在 DOTA 数据集中,不同类别的实例分布极其不均衡,因此将类别的损失函数修改为焦点损失 Focal loss^[2],进而缓解类别不平衡带来的影响。值得注意的是,在没有特殊说明的情况下,其他与程序相关的超参数设置和 YOLOX 一致。

4.3 消融实验

为了验证本文所提出的标签分配策略的有效

性,本文在基准模型 YOLOX 基础上加入了标签粗分配策略以及旋转中心度量方法。所有的消融实验均以 YOLOX-s 作为骨干网络,并在 DOTA v1.0、HRSC 2016 和 UCAS-AOD 数据集上训练并测试。为了保证实验的公平和准确性,除了标签分配策略,其他实验参数设置严格保持一致。实验结果如表 1 所示,在加入标签分配粗策略之后,本文的方法相比于原始的方法在 DOTA v1.0、HRSC 2016 和 UCAS-AOD 数据集上检测精度分别提高了 2.48%、0.35% 及 0.31%。这说明本文设计的标签粗分配策略(判断样本点是否在旋转目标内或目标中心点附近的旋转正方形框内)有利于正负样本的选择,进而提高遥感图像目标检测的精度。由于 DOTA 数据集包含的目标类别较多,分辨率变化比较大,基准模型的检测精度比较低,因此提升效果也比较明显。而 HRSC 2016 和 UCAS-AOD 数据集类别少,分辨率变化不大,检测精度已经达到 95.00% 以上,因此提升的效果并不明显。此外,本文还在此基础上加上了中心度量的方法,在 DOTA v1.0、HRSC 2016 和 UCAS-AOD 数据集上检测精度进一步提高了 1.73%、0.22% 和 0.27%。这证明了中心度量方法有助于减少低质量的边缘样本点,从而进一步提高检测的精度。

表 1 本文方法在不同遥感数据集上的消融实验结果(%)

方法	YOLOX	YOLOX + 标签粗分配	YOLOX + 标签粗分配 + 中心度量
模型	YOLOX-s	YOLOX-s	YOLOX-s
DOTA v1.0	71.20	73.68	75.41
HRSC 2016	95.63	95.98	96.20
UCAS-AOD	97.17	97.48	97.75

4.4 实验结果

4.4.1 在 DOTA 数据集上的实验结果

DOTA 是目前公开的最全面的用于目标检测的遥感图像数据集,其中不仅包含了车辆、船只等小目标,还包括了桥梁等大长宽比的目标,这有利于检验旋转目标检测算法的性能。为了评估本文方法的性能,实验中的方法均选择相同的数据作为训练集,实验结果来源于 DOTA 的评估服务器所测试的结果。在 DOTA v1.0 数据集上的实验结果如表 2 所示,其中 R-101 表示 ResNet-101, H-104 表示 Hourglass-

104,其他以此类推。从表 2 的检测结果可以看出, YOLOXR 在 DOTA v1.0 上取得了 78.21% 检测精度, 优于目前绝大多数旋转目标检测器。其中小型车辆、大型车辆和船只这些类别取得了较高的检测精度, 这说明 YOLOXR 的标签分配策略可以为小目标分配更加细致的标签, 从而避免了复检和漏检, 有

利于密集排列的小目标检测。而在精度较高的飞机、网球场、篮球场以及游泳池这些场景下, 本文的方法同样取得了和其他优秀算法相当的结果。这些检测结果充分说明了 YOLOX 可以为遥感目标分配更加合适的标签, 进而提升目标检测的精度, 其在 DOTA v1.0 上的检测结果如图 5 所示。

表 2 不同方法在 DOTA v1.0 数据集上的检测精度 (%)

方法	骨干网络 /模型	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
两阶段模型:																	
RRPN	R-101	88.52	71.20	31.66	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	R-101	81.36	74.30	47.70	70.32	64.89	67.82	69.98	90.76	79.06	78.20	53.64	62.90	67.02	64.17	50.23	68.16
RoI Trans.	R-101	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
SCRDet	R-101	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
Gliding Vert.	R-101	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
MASKOBB	RX-101	89.56	85.95	54.21	72.90	76.52	74.16	85.63	89.85	83.81	86.48	54.89	69.64	73.94	69.09	63.32	75.33
CSL	R-152	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
SCRDet ++	R-101	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
一阶段模型:																	
DRN	H-104	88.91	80.22	43.52	63.35	73.48	70.69	84.94	90.14	83.85	84.11	50.12	58.41	67.62	68.60	52.50	70.70
DAL	R-101	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	R-101	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
CFC-Net	R-50	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
RIDet-O	R-101	88.94	78.45	46.87	72.63	77.63	80.68	88.18	90.55	81.33	83.61	64.85	63.72	73.09	73.13	56.87	74.70
BBAVector	R-101	88.63	84.06	52.13	69.56	78.26	80.40	88.06	90.87	87.23	86.39	56.11	65.62	67.10	72.08	63.96	75.36
R ³ Det	R-152	89.80	83.77	48.11	66.77	78.76	83.27	87.84	90.82	85.38	85.51	65.67	62.68	67.53	78.56	72.62	76.47
PolarDet	R-101	89.65	87.07	48.14	70.97	78.53	80.34	87.45	90.76	85.63	86.87	61.64	70.32	71.92	73.09	67.15	76.64
YOLOXR	YOLOX-s	85.39	84.04	51.98	64.68	77.44	83.26	88.34	90.83	80.26	87.18	60.43	72.42	67.58	71.67	65.58	75.41
YOLOXR	YOLOX-x	89.97	85.04	54.73	68.76	84.51	85.78	90.83	90.68	86.85	87.06	65.20	67.81	72.67	75.04	68.21	78.21



图 5 YOLOXR 在 DOTA v1.0 数据集上的检测效果图

4.4.2 在 HRSC 2016 数据集上的实验结果

HRSC 2016 数据集的实验结果在 PASCAL VOC 2007 和 VOC 2012 这 2 种指标下进行评估,以便公平地比较不同方法的性能。如表 3 所示,YOLOXR 的小模型 YOLOX-s 在 VOC 2007 和 VOC 2012 指标下分别实现了 89.68% 和 96.20% 的精度,优于 R³Det 等旋转目标检测检测器。更重要的是,在该模型下,所提的方法在 RTX 2080 Ti GPU 上实现了 49 FPS 的超快检测,是目前该数据集下最快的旋转目标检测器。与此同时,YOLOXR 的大模型 YOLOX-x 在 VOC 2007 和 VOC 2012 指标下分别实现了90.25% 和 97.11% 的精度,和目前最先进的旋转目标检测器性能相当,这充分说明了所提方法在面对大横纵比的目标时仍然具有优势。通过上面的结果可以看出,YOLOXR 可以在 HRSC 2016 数据集中实现准确且快速的旋转目标检测。图 6 展示了 YOLOXR 在 HRSC 2016 上的检测效果图。

表 3 不同方法在 HRSC2016 数据上的检测精度和速度比较

方法	骨干网络 /模型	$mAP(07)$ /%	$mAP(12)$ /%	FPS
R ² CNN	ResNet101	73.07	79.73	2.0
RRPN	ResNet101	79.08	85.64	3.5
ROI-Transformer	ResNet101	86.20	—	6.0
RSDet	ResNet50	86.50	—	—
Gliding Vertex	ResNet101	88.20	—	10.0
R ³ Det	ResNet101	89.26	96.01	12.0
R ³ Det-DCL	ResNet101	89.46	96.41	—
CSL	ResNet50	89.62	96.10	—
CFC-Net	ResNet101	89.70	—	28.0
DAL	ResNet101	89.77	—	—
S ² A-Net	ResNet101	90.17	95.01	12.7
AOPG	ResNet101	90.34	96.22	—
YOLOXR(本文)	YOLOX-s	89.68	96.20	49.0
YOLOXR(本文)	YOLOX-x	90.25	97.11	14.0

4.4.3 在 UCAS-AOD 数据集上的实验结果

为了进一步检验 YOLOXR 对密集小目标的检测性能,在车辆和飞机检测的小目标数据集 UCAS-AOD 上进行了实验,实验结果如表 4 所示。结果显示,本文提出的方法在 UCAS-AOD 数据集上取得了 98.00% 的检测精度,优于R³Det^[19]和PolarDet^[20]等



图 6 YOLOXR 在 HRSC 2016 数据集上的检测效果图

表 4 不同方法在 UCAS-AOD 数据集上的检测精度比较(%)

方法	飞机	汽车	mAP
YOLOv2	96.60	79.20	87.90
R-DFPN	95.60	82.50	89.20
DRBox	94.90	85.00	89.95
S ² ARN	97.60	92.20	94.90
RetinaNet-H	97.34	93.60	95.47
R ³ Det	98.20	94.14	96.17
RSDet	98.04	94.97	96.50
SCRDet ++	98.93	94.97	96.95
PolarDet	99.08	94.96	97.02
YOLOXR (YOLOX-s)	99.24	96.26	97.75
YOLOXR (YOLOX-x)	99.28	96.72	98.00

优秀的旋转目标检测器。在车辆的检测上,YOLOXR 相比与最优秀的检测算法更是取得了 1.75% 的提升,这充分说明了 YOLOXR 在小而密集目标检测中的优异性能。由于本文提出的标签分配方法在面对小而密集目标时,总是能分配到最适合的标签,而不会出现目标复检和漏检的情况,因而实现了精准的密集小目标检测。YOLOXR 在 UCAS-AOD 数据集上检测的可视化结果如图 7 所示。



图 7 YOLOXR 在 UCAS-AOD 数据集上的检测效果图

5 结 论

本文提出了一种基于向量叉乘的遥感图像目标检测算法 YOLOXR。根据遥感图像目标方向任意、小而密集等特点,提出了基于向量叉乘的标签分配策略,通过判断像素点是否在旋转目标内或者目标中心点附近的旋转正方形框内,进而确定其是否为候选正样本。此外,为了降低边缘低质量样本的权重,引入了旋转中心度量的概念,通过向量叉乘最小值与最大值的比值来确定像素点距离中心的远近,进而赋予不同的权重。之后,通过最优传输(simOTA)来寻找真实框和样本点之间最优的匹配对,进而确定最终的正负样本。为了解决 Smooth L1 损失不精确以及旋转 IoU 损失不可导的问题,引用了 KLD 损失。实验结果表明,YOLOXR 在 DOTA、HR-SC 2016 和 UCAS-AOD 这 3 个公开的遥感数据集上取得了卓越的检测精度和速度。但是,对 DOTA 数据集中个别类别的检测仍然欠佳,如桥和直升机。因此,在下一步的工作中,根据目标的形状大小等进一步细化标签分配,为大长宽以及数量较少的实例目标分配更加合适的标签,从而提升检测效果。

参考文献

- [1] 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, 2016,39(6):1137-1149.
- [2] 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.
- [3] TIAN Z, SHEN C, CHEN H, et al. FCOS: fully convolutional one-stage object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Seoul, Korea: IEEE, 2019:9627-9636.
- [4] GE Z, LIU S, WANG F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. (2021-08-06)[2022-04-30]. <https://arxiv.org/pdf/2107.08430.pdf>.
- [5] 朱煜,方观寿,郑兵兵,等. 基于旋转框精细定位的遥感目标检测方法研究[J]. 自动化学报, 2023,49(2):415-424.
- [6] 伏轩仪,张釜景,梁文科,等. 锚点机制在目标检测领域的发展综述[J]. 计算机科学与探索, 2022,16(4):791-805.
- [7] 聂光涛,黄华. 光学遥感图像目标检测算法综述[J]. 自动化学报, 2021,47(8):1749-1768.
- [8] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2022-04-30]. <https://arxiv.org/pdf/2004.10934.pdf>.
- [9] 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 Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020:390-391.
- [10] REDMON J, FARHADI A. YOLOV3: an incremental improvement[EB/OL]. (2018-04-08)[2022-04-30]. <https://arxiv.org/pdf/1804.02767.pdf>.
- [11] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017:2117-2125.
- [12] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015:1440-1448.
- [13] YANG X, YANG X, YANG J, et al. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence[EB/OL]. (2022-04-18)[2022-04-30]. <https://arxiv.org/pdf/2106.01883.pdf>.
- [14] 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. Nashville, USA: IEEE, 2021:303-312.
- [15] 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, USA: IEEE, 2018:3974-3983.
- [16] LIU Z, YUAN L, WENG L, et al. A high resolution optical satellite image dataset for ship recognition and some new baselines[C]//Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods. Porto, Portugal: ICPRAM, 2017,2:324-331.

- [17] ZHU H, CHEN X, DAI W, et al. Orientation robust object detection in aerial images using deep convolutional neural network[C]//Proceedings of the 2015 IEEE International Conference on Image Processing. Quebec, Canada: IEEE, 2015:3735-3739.
- [18] LIN T Y, MAIRE M, BELONGIES, et al. Microsoft coco: common objects in context[C]//Proceedings of the European Conference on Computer Vision. Zurich, Switzerland: ECCV, 2014:740-755.
- [19] 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. Vancouver, Canada: AAAI Press, 2021: 3163-3171.
- [20] ZHAO P, QU Z, BU Y, et al. PolarDet: a fast, more precise detector for rotated target in aerial images[J]. International Journal of Remote Sensing, 2021, 42 (15): 5831-5861.

Remote sensing object detection algorithm based on vector cross product label assignment

YU Xinyi, LIN Mi, LU Jiangping, OU Linlin

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

Abstract

Object detection in aerial images has received extensive attention in recent years, and the mainstream remote sensing image object detectors divide positive and negative samples by the intersection-over-union (IoU) between the preset anchor box and ground-truth box. In order to solve the problem of duplicate detection and missed detection in remote sensing images with small and dense objects in the label assignment method based on IoU, a remote sensing image object detection method YOLOXR based on vector cross product label assignment is proposed. Firstly, a rough label assignment strategy is proposed, which uses the vector cross product method to determine whether a pixel is in the oriented object or the rotating square box near the center of the object, so as to determine whether it is a candidate positive sample. Secondly, a rotation center measurement approach is provided to limit the influence of low-quality candidate positive samples on label assignment by judging the distance between the pixel point and the center point using vector cross product and then assigning different weights. Finally, optimal transmission assignment (simOTA) is used to select the optimal matching pairs of ground-truth boxes and the sample points, which minimizes the overall cost, and then assigns the appropriate label to the rotating object. In addition, IoU is replaced by computing the Kullback-Leibler divergence (KLD) of the two-dimensional Gaussian distribution of the ground-truth box and the predicted box to overcome the problem that the rotation IoU loss is not differentiable and the Smooth L1 loss is difficult to be used to weigh the parameters of the oriented bounding boxes. Extensive experiments on public remote sensing image object detection datasets DOTA, HRSC 2016 and UCAS-AOD show that the proposed method outperforms most of the current oriented object detection algorithms.

Key words: remote sensing image, object detection, label assignment, vector cross product