

引用格式:LIU Xin, HUANG Jin, YANG Yingwei, *et al.* Improved remote sensing rotating object detection based on CenterNet[J]. Remote Sensing Technology and Application, 2023, 38(5): 1081-1091. [刘鑫, 黄进, 杨瑛玮, 等. 基于CenterNet的改进遥感旋转目标检测[J]. 遥感技术与应用, 2023, 38(5): 1081-1091.]
DOI:10.11873/j.issn.1004-0323.2023.5.1081

基于CenterNet的改进遥感旋转目标检测

刘鑫¹, 黄进¹, 杨瑛玮¹, 李剑波²

(1. 西南交通大学 电气工程学院, 四川 成都 611756;

2. 西南交通大学 计算机与人工智能学院, 四川 成都 611756)

摘要: 遥感影像由于目标角度各异且普遍密集、小目标占比高、背景复杂等特点, 检测精度低。针对水平框算法不再适用于遥感旋转目标, 以及主流五参数法存在角度回归的周期性边缘互换性问题, 提出VR-CenterNet, 采用向量表示法来进行旋转目标的检测与损失设计, 规避角度回归的根本性问题, 优化细长目标的偏移高敏问题; 针对浅层特征融合的高冗余问题, 引入自适应通道激活过滤杂质信息, 为强化关键点信息, 在主干输出部分引入改进后的全局上下文自适应层激活注意力块。首先在HRSC2016与UCAS-AOU数据集上进行不同算法的性能比较; 再在两数据集上进行方法消融实验, 以验证各改进方法的有效性。实验结果表明: 在HRSC2016与UCAS-AOU数据集上分别取得了88.48%与90.35%的精度。改进算法能够提升遥感旋转目标的检测精度, 为遥感旋转目标的准确检测提供了另外一种解题思路。

关键词: 遥感影像; 目标检测; 自适应激活; 注意力机制; Anchor-free

中图分类号: TP75 **文献标志码:** A **文章编号:** 1004-0323(2023)05-1081-11

1 引言

目标检测即定位和识别, 以卷积神经网络为典型实现。基于卷积网络的目标检测算法根据有无预设锚框可分为两类, 有锚框算法如RCNN家族^[1-3]、YOLO系列^[4-6]、ResNet系列^[7-8]等, 无锚框系列以CornerNet^[9]、ExtremeNet^[10]、CenterNet^[11]等为代表。

俯视角度的遥感影像目标检测, 场景相对于一般的角度来说更为复杂多变, 不同类别间的尺度变化极大, 且目标普遍排列密集角度多变。虽然水平目标检测技术已趋于完善, 但对于角度各异且普遍密集的遥感目标检测来说, 其由于IoU-NMS (Intersection over Union, Non-Maximum Sup-pression)所带来的局限尤为突出。具体可见图1(a), 当

检测两个紧靠的倾斜物体, IoU-NMS总会滤掉一个, 水平框检测适用性大大降低, 而若采用与目标平行的框去检测(图1(b)), 理想情况下两检测框交并比为0。为了解决这种倾斜物体常规检测框不适用问题, 派生出了许多经典的旋转检测算法, 例如应用在遥感影像检测中的R3Det^[12]、SCRDet^[13]、RoI Trans^[14]、R2CNN^[15]、BL2^[16]、FPN-ROI^[17]等算法。

在旋转检测算法中, 以五参数 (x, y, w, h, θ) 算法为主流, 与水平检测的回归方式相似, 其角度也通过回归的方式进行预测。虽然现今也取得了相当不错的结果^[12-15], 但这种回归方式存在一些根本性的缺陷。由于角度具有周期性与边缘互换性, 这会导致边界回归的不连续。为了解决这个问题, 往

收稿日期: 2020-02-16; 修订日期: 2023-04-07

基金项目: 国家自然科学基金项目(61733015), 高铁联合基金(U1934204), 四川省重点研发计划(2020YFQ0057), 四川省自然资源科研项目(KYL202106-0099)。

作者简介: 刘鑫(1997-), 女, 四川成都人, 硕士研究生, 主要从事深度学习在图像处理方面的应用研究。E-mail: 1252054823@qq.com

通讯作者: 黄进(1977-), 男, 四川成都人, 副教授, 主要从事深度学习在图像处理方面的应用研究。E-mail: 396341096@qq.com

往需要以更为复杂的形式进行边界处的回归,例如使用 Smooth-IoU 损失^[18]来解决边界损失跳变的问题,而这毫无疑问会增加模型复杂度,同时边界处的预测难度也会陡增。而这对于需要高精度旋转契合检测的物体来说是致命的,一般来说,细长目标对角度的偏移极为敏感,具体可见图 1(c)。

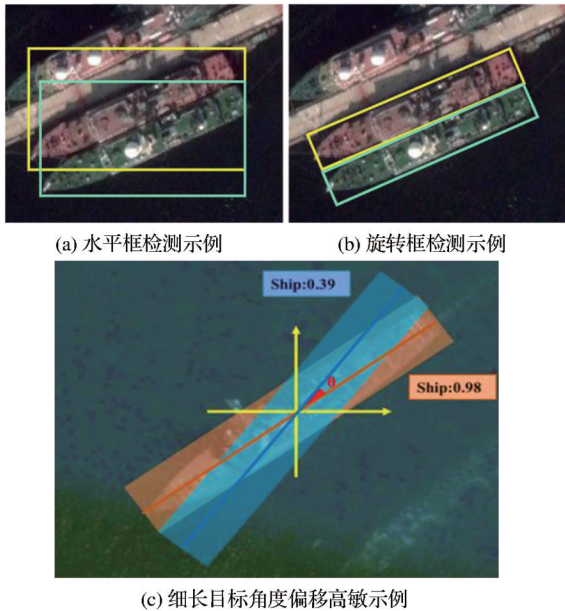


图 1 遥感目标检测示例图

Fig.1 The diagram of remote sensing object detection

为了规避角度回归的本质性问题, Yang 等^[19]设计了环形平滑标签,将角度由回归问题转化为分类问题,解决了角度边界互换问题,并采用密集编码方法(Densely Coded Labels, DCL)^[20]有效缩短编码长度,极大减轻了角度分类头的压力,验证了角度分类的可行性。

也有研究者将角度由回归问题转化为几何回归问题。Lu 等^[21]采用最小水平包围框与最小包围框同心圆切分进行旋转框的表示,该方法过分依赖最小包围框的准确度,而且特殊情况过多,导致框方向与大小选定的损失容易出现较高偏差。He 等^[22]采用旋转内切椭圆进行旋转框的预测,有效的解决了基于角度进行旋转回归的角度周期性问题,但在方向边界处理损失方面还存在欠缺,且没有考虑到正方形框预测时的方向不确定性。宋文龙等^[23]提出目标框对角向量与其相邻向量的投影长度来进行目标表示与预测,随规避了角度问题,但是在对细长物体的检测损失处理上仍存在欠缺。

遥感图像中小目标占比极高,提升小目标检测的精度对遥感目标检测而言极为重要。普遍的提

升小目标检测精度的方法是进行多层特征融合,例如早期为实现不同尺度目标检测的图像金字塔,后期升级的将细节较多的浅层特征与语义丰富的顶层特征融合的特征金字塔 FPN^[24]。Yang 等^[25]将稠密连接与 FPN 结合,通过横向连接与密集连接来获取更高分辨率特征,以提升小目标检测效果; Wang 等^[26]在 FPN 横向连接部分使用改进 Inception 模块以加强浅层细节特征的传输。特征融合虽然一定程度上解决了小目标检测困难的问题,但是卷积网络带来的图像特征的高冗余问题,在这个过程中往往是被忽略的。为了减少杂质冗余的影响, Fu 等^[27]在进行不同层特征融合时,采用一个权重平衡因子来平衡特征的融合,但该因子取定的人为先验性太强,拓展应用时鲁棒性较差。Han 等^[28]指出卷积网络输出的特征图的不同通道中存在很多高度相似的冗余特征,某些通道层其实并不需要在整个数据流过程中进行激活。与此同时,越复杂深层的特征融合会带来显著的内存压力,故而优化浅层的特征融合就显得尤为重要。

有锚框算法需要进行锚框设定,在应对类别极度不平衡且尺度差异极大的问题上,需要极其复杂的计算才能得到预设锚框,与无锚框算法相比短板明显,故而本文采用无锚框算法中的经典算法 CenterNet 作为基础框架来进行算法改进,提出 VR-CenterNet 旋转检测算法,在融合的单层特征图上做全尺度遥感旋转目标的预测和回归。

本文首先针对角度回归本质性问题,采用向量表示法重新设计损失来进行旋转框的预测回归,并将细长目标的偏移敏感性考虑到损失设计中,优化细长目标的回归准确度。其次针对浅层特征融合带来的高冗余问题进行优化,引入自适应激活模块与改进全局上下文层激活注意力模块。在不影响模型复杂度的前提下,有效提升遥感影像的目标检测精度。

2 基于 CenterNet 的改进遥感旋转目标检测算法

2.1 CenterNet 基础算法原理

CenterNet 是一种端到端的像素级目标检测网络,将目标视作关键点的回归问题,以中心偏移预测,尺度预测与类别预测来进行目标的检测。其特征提取结构有 DLA (Deep Layer Aggregation) 和 Hourglass。其中 DLA 小巧而高效,作为本文算法

改进的基础模型结构,其特征提取结构如图2所示,是一种记忆型特征提取结构,后端特征与前端特征高度链接,实现高效的特征记忆矫正性,可削弱特征重炼而导致的部分特征丢失问题。

2.2 改进旋转目标检测算法

本文所用网络的总体结构如图3所示,将DLA最后一层特征提取层(dla5提取层)减枝后作为主干框架,简称DLAs,底层模块使用自适应通道激活SA来进行搭建;遥感影像中小目标密集,故而以下采样4的步长输出融合特征图;将得到的特则图使用GC-SAL进行二次过滤,强化关键点的表达。

经过解码操作后,最后将分类头与回归头解耦,回归头拆分为旋转框的偏移部分、旋转框的尺寸部分与旋转框的朝向预测部分。本节接下来的篇幅将对网络各部分做详细说明。

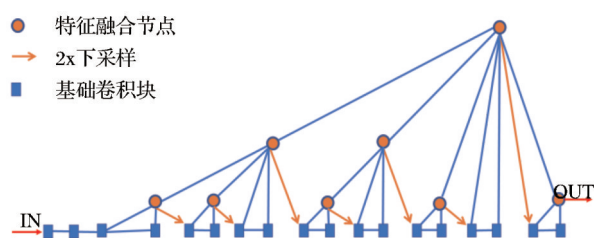


图2 DLA结构图

Fig.2 The structure diagram of DLA

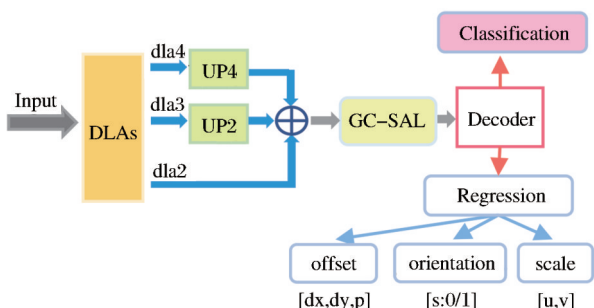


图3 网络总体结构图

Fig.3 The overall structure of the network

2.2.1 SA 激活

激活函数在神经网络的多方位应用发展中扮演着极为重要的角色,非线性激活函数的引入,让计算机在处理现实世界复杂多变的任务中变得游刃有余。最为常用的激活函数有Sigmoid、ReLU、LeakyReLU、Tanh、Mish等。在对特征的激活操作上,普遍都是对任意输出特征的全部通道层以相同的激活函数进行统一操作。但是特征层中某些特征通道对于网络学习来说并不重要,而某些通道起着至关重要的作用,若网络能自适应的按通道对特

征进行激活,具体来说:对不重要的甚至是干扰的通道进行灭活操作,而对关键特征通道进行强化激活,这不仅能提升网络特征提取的效率也能大大增进网络对关键信息学习的能力。ACON^[29]激活采用3个可学习变量(p_1, p_2, β),自适应的控制网络的各通道激活方式。

文献[29]指出激活函数多是由最大值输出函数Maxout进行二元线性变化得来。对Maxout函数做平滑近似,可得式(1)。

$$S_{\beta}(x_1, x_2, \dots, x_n) = \frac{\sum_{i=1}^n x_i e^{\beta x_i}}{\sum_{i=1}^n e^{\beta x_i}} \quad (1)$$

令 $\max(\eta_a(x), \eta_b(x))$ 作为激活得通用形式,其中 $\eta_a(x), \eta_b(x)$ 表示任意线性函数,对式(1)进行二元线性变换,得到式(2)。

$$\begin{aligned} S_{\beta}(\eta_a(x), \eta_b(x)) &= \eta_a(x) \cdot \frac{e^{\beta \eta_a(x)}}{e^{\beta \eta_a(x)} + e^{\beta \eta_b(x)}} + \eta_b(x) \cdot \frac{e^{\beta \eta_b(x)}}{e^{\beta \eta_a(x)} + e^{\beta \eta_b(x)}} \\ &= \eta_a(x) \cdot \frac{1}{1 + e^{-\beta(\eta_a(x) - \eta_b(x))}} + \eta_b(x) \cdot \frac{1}{1 + e^{\beta(\eta_a(x) - \eta_b(x))}} \end{aligned} \quad (2)$$

令 $\sigma(x) = \frac{1}{1 + e^{-x}}$, $\eta_a(x) = p_1 x$, $\eta_b(x) = p_2 x$, 得出式(3)。

$$\begin{aligned} S_{\beta}(\eta_a(x), \eta_b(x)) &= (p_1 - p_2) x \cdot \sigma(\beta(p_1 - p_2)x) + p_2 x \end{aligned} \quad (3)$$

式(3)即ACON激活的自适应激活的函数表达式,其中 p_1, p_2 是网络需要学习的激活权重,来控制激活的类型, β 为通道激活强度范围控制量,其作用如图4所示,可根据特征进行选择性的增强。

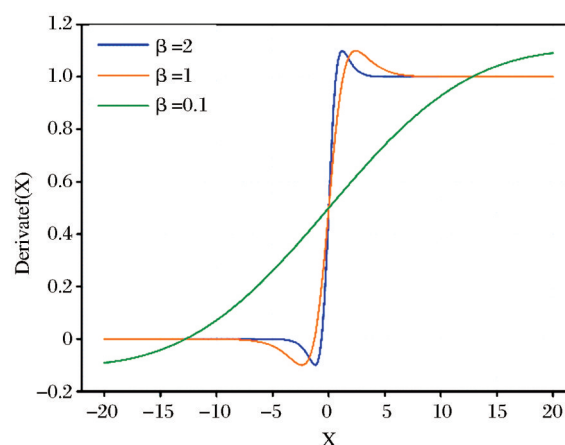


图4 β 对激活的影响

Fig.4 The impact of β on activation

在ACON激活基础上,研究提出SA激活(Self-adaptive Activation),以全新的方式(式(4))生成通道激活强度范围控制量 β ,增进SA激活的通

道适配性。

$$\beta = a \cdot \text{Sigmoid}(CVBN_2(CVBN_1(INS_{sum(2,3)}(x)))) \quad (4)$$

其中: x 表示输入特征图, $INS_{sum(2,3)}$ 表示实例归一化后对维度(2,3)即HM维求和压缩至1x1维, CVBN表示卷积+批量标准化操作, 下标数字表示顺序, Sigmoid将输出限制到[0,1]之间, a 是范围调节因子, 实验中取10, 可以将输出映射到[0, a], β 即生成的通道激活强度范围控制量。

使用SA激活所构建新基础模块SAB, 结构如图5所示, 新的基础模块在进行特征提取和特征融合时, 可自动对各通道进行增强或抑制, 有效提升网络在浅层的语义鉴别提取能力, 为后续预测提供强有力的信息支撑。

2.2.2 GC-SAL注意力模块

遥感影像中, 背景复杂多变, 为了更好地提取到关键目标信息, 滤除无用信息, 提升检测效率, 融入注意力机制。GC-Block^[30]是在两经典注意力模块SE与NL的基础上提出的全局上下文注意力模块, 但是在通道冗余剥离处理上, 还是有待改进。本文提出GC-SAL(Global Context with Self-adaptive Layer Activation Attention Block), 对GC-Block进行改进。在注意力提炼部分, 将原来的层归一化与ReLU激活, 替换为可进行有效冗余剥离的改进自适应激活模块SAL, 并进行16倍的通道升维操作。相较于底层模块中的SA, SAL中 β 的生成公式做了相应修改, 见式(5)。

$$\beta = a \cdot \text{Sigmoid}(CVBN_2(CVBN_1(LN(x)))) \quad (5)$$

式(5)中LN表示层归一化, 其余各参数含义与式(4)相同, 不再赘述。

GC-SAL运算操作如式(6):

$$O = X_i + W_{v2} \text{SAL} \left(W_{v1} \sum_{j=1}^{N_p} \frac{e^{W_k X_j}}{\sum_{m=1}^{N_p} e^{W_k X_m}} X_j \right) \quad (6)$$

其中: O 是运算输出, $X = \{X_i\}_{i=1}^{N_p}$ 是输入特征, 对于图像数据来说 N_p 表示单层通道上的像素点数量 $H \times W$, W 是线性变换矩阵, 网络中采用 1×1 卷积实现, SAL即前文提出的自适应通道激活SA与层归一化结合的新模块。

2.2.3 向量表示法

为了解决常用五参数旋转框检测中角度的周期性与边缘置换性问题, 文献[23]提出向量表示法, 将旋转框表示为以框中任意一条对角向量, 与该向量相邻的任意一个顶点到该向量同向投影长度的表示方法, 以2分类的方式确定框的朝向, 解决

了角度本质性的问题。以图6(a)为例, 框的中心点由C表示, 以该点为零点, 建立直角坐标系, 从C出发选定任意一个顶点D连接, 作为基准向量 CD , 再选定任意与点D相邻的一个框顶点E, 约束 CD 与 CE 的夹角小于等于90度, 即此时 CE 在 CD 上的投影 CP 与 CD 同向。由图6(a)易知, 当 CD 的坐标分量同号时, CD 处于一、三象限, 异号时处于二、四象限, 使用简单的二分类即可确定框的朝向。

向量表示法使用6个参数(x_c, y_c, u, v, s, ρ), 即可表示任意方向的矩形框, 其中(x_c, y_c)即框的中心坐标, (u, v)是 CD 在以中心点为原点建立的直角坐标系上的分量模长, s 确定 CD 的朝向, $\rho = |CP|/|CD|$ 确定矩形框的具体形状。

当 CD 落点在坐标轴附近时, 如图6(b)所示, 朝向的两种选取也均可接受, 此时将 s 的损失应该减小。

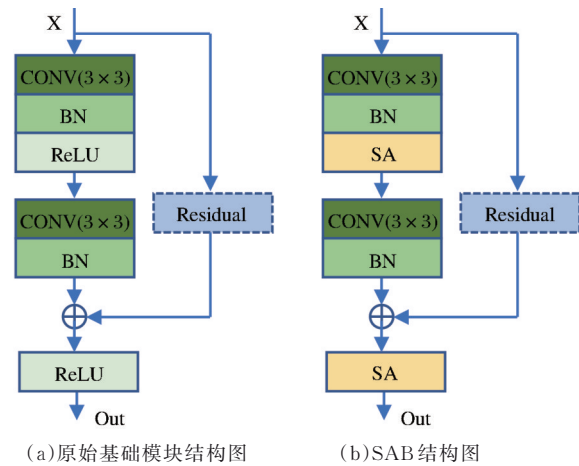


图5 基础模块对比图

Fig.5 The comparison diagram of basic modules

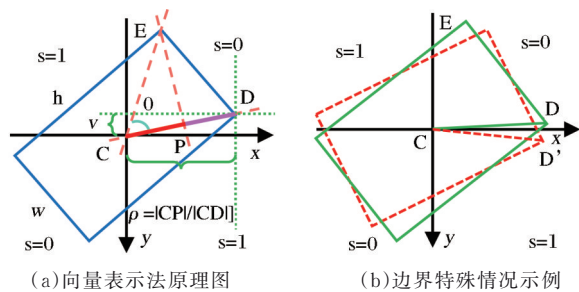


图6 向量表示法

Fig.6 The vector representation

观察到当 CD 在坐标轴附近时, 两分量模长的差距会极大, 两者差距可由式(7)表示。

$$\text{disp}(u, v) = \left(\frac{|u| - |v|}{|u| + |v|} \right)^2 \quad (7)$$

差距越大该式越接近于1, 差距越小, 此式越接

近于0。为朝向损失引入权重衰减控制因子 η 。

$$\eta = \frac{1}{1 + e^{L \cdot (disp(u,v) \cdot g)}} \quad (8)$$

式(8)中 L 为放缩率,设为100, g 为差距控制量,实验中设为0.7,均可根据实际情况予以调整。

特殊地,当框为正方形时, ρ 恒为0,此时框的朝向并不重要,将 CD 的落点限制在 $s=0$ 区域与 x 轴上,即可解决正方形框的检测问题。

2.2.4 损失函数

网络总损失函数为4个部分的加权和,分别是类别损失、偏移损失、尺度损失与朝向损失。

类别损失 L_{cls} 采用Focal-loss^[31],使网络能自动聚焦于难例的区分,具体计算如式(9)所示。

$$L_{cls} = -\frac{\alpha}{N} \sum_c \begin{cases} (1 - Y^p)^\gamma \log(Y^p), & \text{if } Y^g = 1 \\ (1 - Y^g)^\beta (Y^p)^\gamma \log(1 - Y^p), & \text{other} \end{cases} \quad (9)$$

其中: $\alpha=0.25, \beta=4, \gamma=2$, N 是真值框数量, c 表示类别 c , Y^p 表示预测结果概率值, Y^g 表示真值。

偏移损失 L_{offset} 包含两个部分,一是目标框中心以下采样倍数在原图上进行网格划分的所对应网格的偏移损失,另一部分为框具体形状的决定量 ρ 的损失,两部分输出均在 $[0, 1)$ 之间。偏移损失采用均方误差,见式(10)。

$$L_{offset} = \frac{1}{N} \sum_{i=1}^N (\sqrt{f(xy)} + (\rho_i^p - \rho_i^g)^2) \quad (10)$$

$$f(xy) = (dx_i^p - dx_i^g)^2 + (dy_i^p - dy_i^g)^2$$

其中: N 与前式相同, (dx_i^p, dy_i^p) 与 (dx_i^g, dy_i^g) 分别表示第 i 个预测框与真值框的中心相对偏移量, ρ 即框具体形状的决定量。

尺度预测损失 L_{uv} 也采用均方误差,由于细长目标对于角度偏移极为敏感,由图6易知,当 ρ 越趋近于1时,则代表目标越细长,故而引入细长目标基准向量偏移损失控制量 $0.5e^\rho$,用于控制基准向量的尺度预测,从而较好地解决细长目标的偏移高敏性。

$$L_{uv} = \frac{e^\rho}{2N} \sum_{i=1}^N \sqrt{(u_i^p - u_i^g)^2 + (v_i^p - v_i^g)^2} \quad (11)$$

其中: (u_i^p, v_i^p) 和 (u_i^g, v_i^g) 分别表示第 i 个预测框与真值框的基准向量的分量模长。

朝向损失 L_s 采用二分类交叉损失。

$$L_s = \frac{\eta}{N} \sum_{i=1}^N (s_i^g \log s_i^p + (1 - s_i^g) \log (1 - s_i^p)) \quad (12)$$

式(12)中 s_i^g 和 s_i^p 分别表示第 i 个真值框和预测框的朝向, η 为朝向坐标轴附近的权重衰减控制因子。

总体损失为4个部分的加权和,如式(13)。

$$L_{total} = \lambda_1 L_{cls} + \lambda_2 L_{offset} + \lambda_3 L_{uv} + \lambda_4 L_s \quad (13)$$

实验中 $\lambda_1=1, \lambda_2=1.3, \lambda_3=0.4, \lambda_4=0.7$ 。

3 实验与分析

实验基于python 3.6进行, Linux操作系统CU-DA 10.2, PyTorch 1.10, 硬件配置为Tesla T4@16Gx2, 深度学习框架采用基于Pytorch的高度集成化框架Detectron 2.0.6。

3.1 实验数据

采用遥感目标检测领域最为常用的两个数据集HRSC2016^[32]和UCAS-AOD^[33]用于模型训练与有效性验证。

UCAS-AOD数据集是由中国科学院大学在Google Earth上截取的全球部分地区的遥感影像制成的数据集,分为两个类别,汽车与飞机,共计1510张样本图,其中飞机样本图有1000张,汽车样本图有510张,每个类别的实例样本总数目几近一致。

HRSC2016数据集同样来自Google Earth,由西北工业大学标注制成,共计1070张图片样本。根据任务的不同,目标类别也有所差异,分别是针对单类船只的目标检测、三类别舰船检测与细分的26类别细分舰船检测。本文采取与文献[14-16]一致的单类任务进行网络效率评定。

3.2 实验方法

训练时,使用官方DLA34预训练权重初始化新建网络,修改部分采用Xavier初始化策略,网络输入大小 640×640 。数据增强采用原图随机放缩与固定大小随机位置截取的在线数据增强操作,为了使模型能更全面的学习到旋转信息,又引入了水平翻转、上下翻转、以及随机旋转 90° 的一些数据增强操作,随机的对比度、光照、饱和度也是为了应对不同的环境光亮带来的影响的一些惯用操作。训练轮次设定为40k,批处理大小设为32,权重衰减因子为0.0001,初始学习率0.001,采用余弦退火学习率调整策略。

实验时,基于CenterNet与向量表示法,分别融入SA激活,SA激活+原始GC注意力块,SA激活+改进GC-SAL注意力块。最后一种融入方式即本文所提VR-CenterNet算法。

3.3 评价指标

采用可反应全局的性能指标平均准确度(mean Average Precision, mAP)作为网络的评估指标,在各式各样的机器学习评估中最为常用。在遥感旋

转目标检测中,常规的 IoU 已不再适用, PolyIoU 可计算任意多边形之间的交并比,常用于遥感旋转检测的正负例判别。mAP 的计算涉及到两个常用准度评估量,即精确度(Precision)与召回率(Recall)。其中精确率指在所有检测为真的样本中实际为真的比例,召回率指所有检测为真且实际为真占真值样本集的比例。式(14)为单类的 mAP 计算式。

$$AP = \int_0^1 P(r) dr \quad (14)$$

其中: r 代表召回率, $P(r)$ 代表 r 对应处的精确度。

对所有类别的 AP 求平均即可得到 mAP 值,式(15)中 C 代表总类别数。

$$mAP = \frac{\sum_{i=1}^C AP_i}{C} \quad (15)$$

3.4 结果分析

3.4.1 对比实验

为了验证本文所提算法 VR-CenterNet 的有效性,表 1 与表 2 将其与一些经典的算法模型分别在 HRSC2016 与 UCAS-AOD 数据测试集上关于主干类型、是否无锚框、网络输入尺寸与平均准确度等方面进行了比较,算法根据是否是端到端一步到位的网络,分为两阶段算法(Two-stage methods)与单阶段算法(One-stage methods)。表中 mAP@50 表示在多边形交并比大于等于 0.5 时测出来的平均准确度,其最优结果以粗体显示。

由表 1 可知,在 HRSC2016 数据测试集上, VR-CenterNet 以更为轻巧的模型获得了 88.48% 精度,其主干 DLAs 表示在 DLA34 结构基础上移除最后一个特征提取层之后的剪枝主干。

由表 2 可知,在 UCAS-AOD 数据集上, VR-CenterNet 获得了 90.35% 的平均准确度,为表中最优,但在两个类别上的单独准确度却有较为明显的差异,由表 2 可知 VR-CenterNet 对于飞机类别的识别精度最高,为 94.39%,而对汽车的识别精度较差,仅有 86.31%,后文可视化结果部分会对该差别予以说明。

表 1 不同模型在 HRSC2016 数据测试集上精度评估结果
Table 1 Accuracy evaluation results of different models on HRSC2016 test data set

Models	Backbone	Anchor-Free	mAP@50 (%)	InputSize
两阶段算法				
BL2[16]	ResNet101	×	69.60	half of original size
R2CNN[15]	VGG16	×	73.07	800 x 800
RoI Tran[14]	ResNet101	×	86.20	512 x 800
单阶段算法				
IENet[35]	ResNet101	✓	75.01	1024 x 1024
OriCenterNet[36]	ResNet101	✓	78.15	800 x 800
CSL[19]	ResNet101	×	89.62	800 x 800
SAR[21]	ResNet101	✓	88.45	896 x 896
RIE[22]	HRGANet-W48	×	91.27	800 x 800
VR-CenterNet	DLAs	✓	88.48	640 x 640

表 2 不同模型在 UCAS-AOD 数据集上精度评估结果

Table 2 Accuracy evaluation results of different models on the UCAS-AOD data set

Models	Backbone	AnchorFree	Car(%)	Plane(%)	mAP@50(%)	InputSize
两阶段算法						
Faster RCNN[3]	ResNet50	×	86.87	89.86	88.36	800 x 800
RoI Tran[14]	ResNet50	×	88.02	90.02	89.02	800 x 800
RIEDet-Q[34]	ResNet50	×	88.50	89.96	89.23	800 x 800
单阶段算法						
R-Yolov3[6]	Darknet53	×	74.63	89.52	82.08	800 x 800
R-RetinaNet[31]	ResNet50	×	84.64	90.51	87.57	800 x 800
VR-CenterNet	DLAs	✓	86.31	94.39	90.35	640 x 640

3.4.2 消融实验

为有效验证本文针对向量表示法存在的细长目标弱敏性所提出的细长目标权重损失控制项的有效性,本文在包含大量细长船舶目标的 HRSC2016 数据集上进行消融实验验证,结果如表 3 所示,其中 $Ws = 0.5e^0$,即细长目标权重损失控制

量, mAP50、与 mAP75 分别表示多边形交并比大于等于 0.5 与 0.75 时的检测精度。

由表 3 可知,在无任何优化的条件(baseline)下,高 IoU 阈值精度(mAP75)仅能达到 35.35%;而在仅引入 Ws 后,低 IoU 阈值精度提升 2.47 个百分点,高 IoU 阈值精度能提升 9.78 个百分点。

表 3 细长目标权重损失控制量在 HRSC2016 数据集的消融实验结果

Table 3 Ablation experiment results of slender target weighting control volume on HRSC2016 data set

Model	mAP@50/%	mAP@75/%
baseline	82.39	35.34
+ W_s	84.86 (+2.47)	45.12 (+9.78)

更具体地,将检测结果可视化。由图 7 可知,在未引入 W_s 前(baseline),原始向量表示法在进行细长目标旋转框回归时,预测框与目标的贴合度较差;而在引入细长目标损失控制量后($+W_s$),检测效果明显改善。以上结果证明了本文针对向量表示法所提细长目标损失权重控制量的有效性。

为进一步验证各冗余优化模块的有效性,在 UCAS-AOD 数据集上进行模块消融实验,使用 mAP 与参数量(Parameters)作为评价指标,以原始 DLAs 作为模型主干基线。为了公平起见,实验中

所有实验数据与参数设置均严格一致。使用不同模块的模型评估结果如表 4 所示。

表 4 VR-CenterNet 算法各冗余优化模块在 UCAS-AOD 数据集上的消融实验结果

Table 4 Ablation experiment results of each redundant optimization module of VR-CenterNet on UCAS-AOD data set

SA	GC	GC-SAL	mAP@50 /%	Parameters
			85.34	30.8 M
✓			89.21	31.1 M
✓	✓		89.04	31.1 M
✓		✓	90.35	32.2 M

由表 4 可知,SA 的加入提升 3.87% 的精度,在 SA 的基础上加入原始 GC 精度反而不如单独引入 SA,但 SA 与 GC-SAL 结合却能在参数少量增加的情况下再次提升 1.14% 的准确度,证明了本文所提算法的有效性。

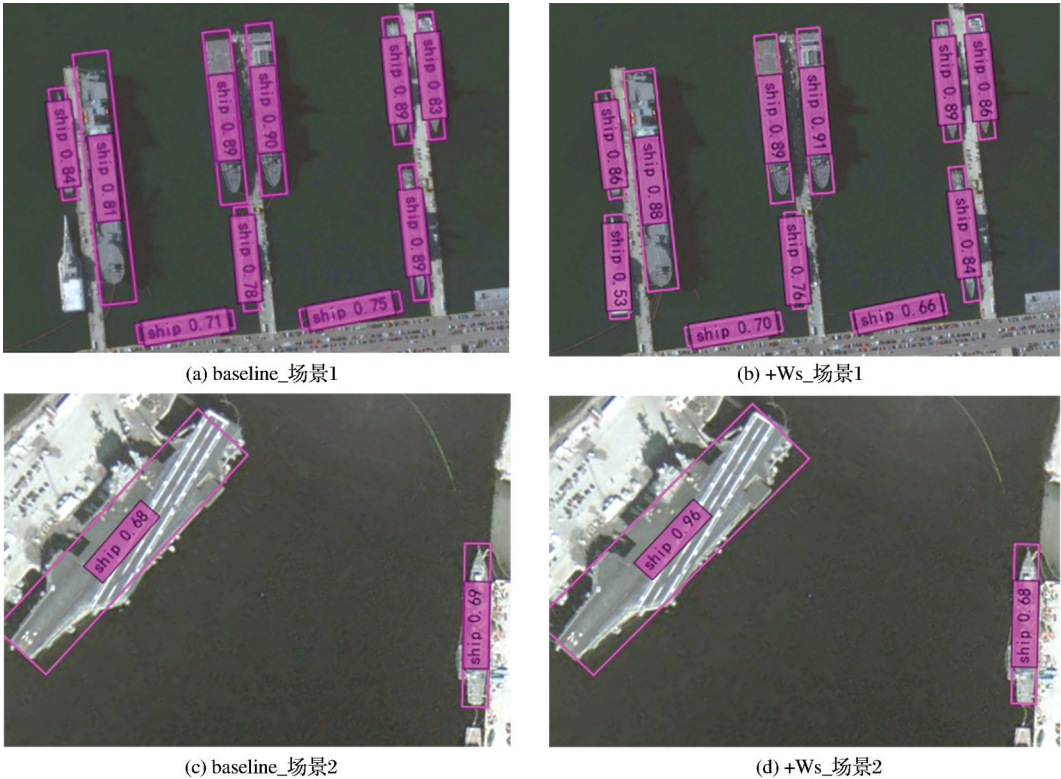


图 7 细长目标权重控制量的消融实验检测对比图

Fig.7 Ablation experimental detection comparison chart of slender target weighting control volume

为进一步地说明各模块的有效性,将不同模块组合的消融实验结果进行可视化。以图 8 为例,对采用不同模块组的网络主干输出的特征图以图像形式展出,如图 9 所示,其中 a 到 d 的序号分别对应

表 4 中从上往下的顺序。

图 9 中可以明显看到原始模块网络输出的特征图(a)中存在大量的杂质信息,且关键信息的亮度是最弱的;在融入了 SA 激活后(b)不但滤掉了多数

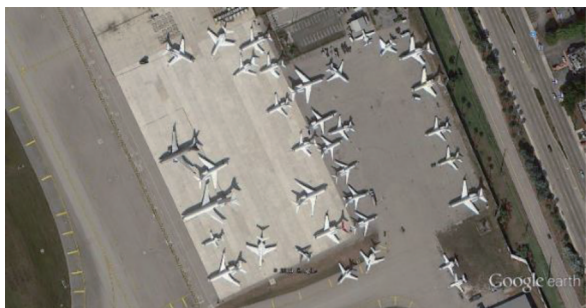


图 8 消融实验样例

Fig.8 Example of ablation experiments

的杂质信息,关键信息的强度也进行了增强,许多之前未被检测出的目标也被捕获;在 SA 激活的基础上引入原始 GC,可以看到特征图(c)中背景中的杂质信息由于通道 4 倍压缩精炼再全局放大后被反向增强,相比之下弱化了关键信息的表达,影响最

终目标的预测,故而检测的精度反而些许下降;而相较于原始 GC,改进后的 GC-SAL 采用了 16 倍通道膨胀,再结合 SA 激活进行杂质过滤,最后再压缩回原通道数量,可以最大程度上过滤掉背景中的干扰信息,使得最终的预测可以充分的利用特征图中的有效信息,故而消融实验结果表现最优。

3.4.3 检测效果展示

本文提出的 VR-CenterNet 在两个数据集进行具体检测效果的示例展示,如图 10 所示,其中,左图为原始标签,右图展示检测结果。

针对之前 UCAS-AOD 数据集上进行精度数据展示时汽车类别识别精度低的情况,根据图 10 中间行左图可知,UCAS-AOD 数据集中车辆的标注较为粗糙,且包含较多漏标的情况,故而导致该类在测试集上精度稍低。

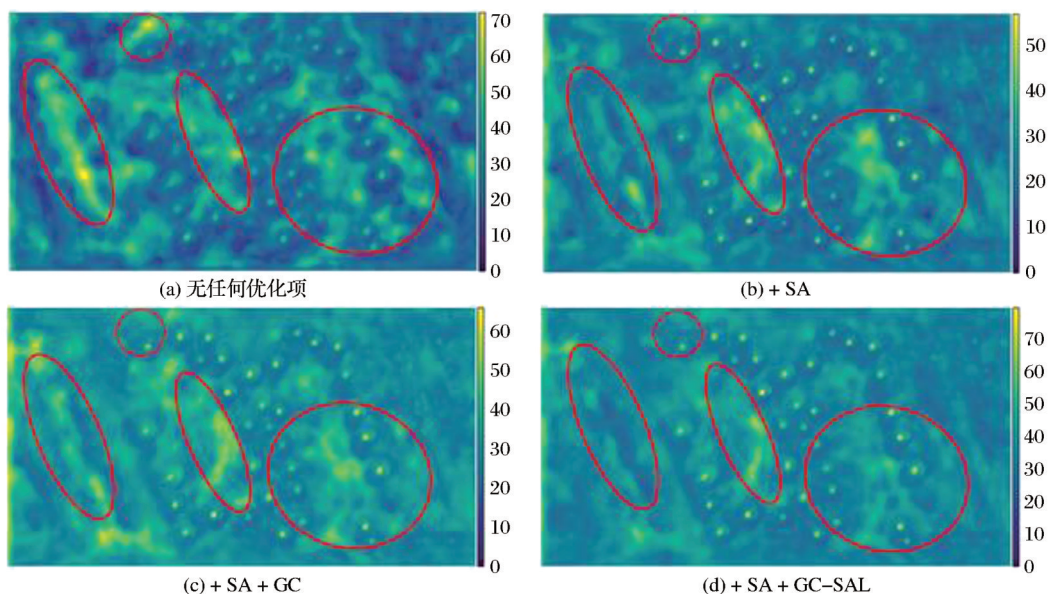


图 9 冗余优化模块的消融实验主干输出特征图

Fig.9 The output features of backbone in ablation experiment for redundant optimization modules

4 讨论

基于 CenterNet 框架规避了五参数法中角度的本质性问题,实现了遥感影像旋转目标的高精度检测,具有一定参考意义。为了深入理解实验细节,为相关研究提供参考建议,特对研究方法与不足之处进行探讨。

从回归方式来看,选用向量表示法作为旋转框的表示方法,不仅有效解决了水平目标检测算法在遥感目标检测中不再适用的问题,也有效规避了遥感旋转目标检测中主流五参数法的角度周期性与边缘互换性问题,这与文献[23]的研究结果一致,

但其没有考虑到细长目标的回归偏移敏感性问题。针对该问题,在损失设计时引入一个的偏移损失控制量,可有效提升细长目标的回归准确度。在对比实验中的结果表明,结合其他特征提取优化方法,VR-CenterNet 分别在 HRCS2016 与 UCAS-AOD 数据集上取得了 88.48% 与 90.35% 的检测精度。

从特征融合冗余优化来看,在 UCAS-AOD 数据集上的消融实验结果表明,选用 SA 激活对各阶段特征进行杂质过滤后再进行特征融合,优于以 ReLU 激活为底层模块的原始 DLAs-CenterNet 算法,仅引入了 0.3 M 的参数量,获得了 3.87% 的精度

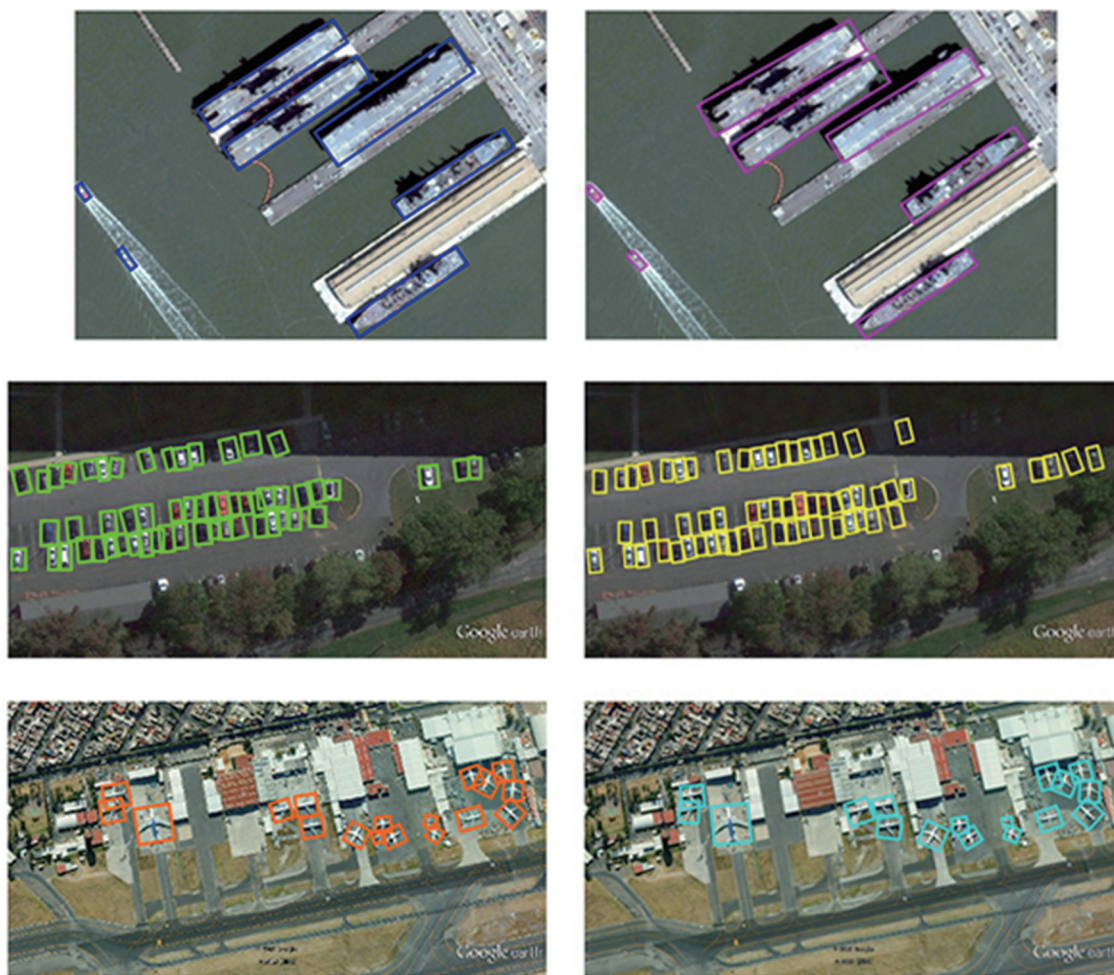


图10 HRSC2016(上)与UCAS-AOD(中、下)数据集的检测结果图

Fig.10 The detection results on HRSC2016(Top) and UCAS-AOD(Middle,Bottom) data set

提升,这与Ma等^[29]的研究相似,不同的是针对遥感影像背景复杂、目标密集等特点,使用了适配性更强的通道激活产生方式和扩充的激活增强范围,从而达成杂质信息的高效过滤。与众多注意力模块^[30]的引入目标一致,对过滤后的特征进行关键信息强化,选用SA激活结合改进GC-SAL注意力模块的消融实验结果优于SA激活与原始GC注意力模块结合算法,既验证了改进注意力模块的有效性,也验证了SA激活针对不同问题的可塑性。

针对上述实验,仍存在不足之处:①实验仅在单层融合特征图上进行全尺度目标预测,若同一图像中尺度差异极大,可能会存在大目标高关注性。在后续研究中,可引入空洞卷积增大单层特征图的感受野,以削弱目标大小的影响。②此外,实验中未进行单注意力模块细化验证,可以在原始算法基础上进行轮换实验,验证模块有效性。③实验中未进行原始水平检测算法的对比实验,在后续研究中应对该部分予以补充。

5 结 语

针对水平检测算法不再适用于遥感旋转目标,以及主流遥感旋转检测五参数法的周期性与边缘互换性问题,以目标检测中的经典无锚框算法CenterNet为基础,采用向量表示法进行新的回归头与损失函数设计,有效规避了角度的本质性问题,同时将细长目标的偏移敏感度考虑到损失设计中,引入一个偏移损失控制量,实现细长目标的高精度回归。针对浅层特征融合带来的杂质高冗余问题,提出SA激活与GC-SAL注意力模块,分别融入网络的基础结构与主干输出部分,抑制杂质信息,增强关键信息,提升小目标的检测精度。在公开遥感检测数据集HRSC2016和UCAS-AOD中,改进算法分别取得了88.48%与90.35%的平均准确度,相较于其他算法以更轻巧的模型结构获得了较好的评估结果。改进算法VR-CenterNet为普遍密集且角度各异的遥感目标的精确检测,提供了一个新思路。

参考文献 (References):

- [1] GIRSHICK R, DONAHUE J, DARRELL T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation [C] // Proceedings of the IEEE conference on computer vision and pattern recognition, 2014: 580-587.
- [2] GIRSHICK R. Fast R-CNN [C] // Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [3] REN S, HE K, GIRSHICK R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks [S]. Advances in Neural Information Processing Systems, 2015, 28.
- [4] 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. 2016: 779-788.
- [5] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C] // Proceedings of the IEEE Conference On Computer Vision and Pattern Recognition. 2017: 7263-7271.
- [6] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. arXiv preprint arXiv:1804.02767, 2018.
- [7] 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. 2016: 770-778.
- [8] ZHANG H, WU C, ZHANG Z, *et al.* Resnest: Split-attention networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 2736-2746.
- [9] LAW H, DENG J. Cornernet: Detecting objects as paired keypoints [C] // Proceedings of the European Conference on Computer Vision (ECCV). 2018: 734-750.
- [10] ZHOU X, ZHUO J, KRAHENBUHL P. Bottom-up object detection by grouping extreme and center points [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 850-859.
- [11] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as points [S]. arXiv preprint arXiv:1904.07850, 2019.
- [12] 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. 2021, 35(4): 3163-3171.
- [13] YANG X, YANG J, YAN J, *et al.* Srdet: Towards more robust detection for small, cluttered and rotated objects [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 8232-8241.
- [14] DING J, XUE N, LONG Y, *et al.* Learning roi transformer for oriented object detection in aerial images [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2849-2858.
- [15] IANG Y, ZHU X, WANG X, *et al.* R2CNN: Rotational region CNN for orientation robust scene text detection [S]. arXiv preprint arXiv:1706.09579, 2017.
- [16] LIU Z, HU J, WENG L, *et al.* Rotated region based CNN for ship detection [C] // 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017: 900-904.
- [17] AZIMI S M, VIG E, BAHMANYAR R, *et al.* Towards multi-class object detection in unconstrained remote sensing imagery [C] // Computer Vision-ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2-6, 2018, Revised Selected Papers, Part III. Cham: Springer International Publishing, 2019: 150-165.
- [18] YANG X, YAN J, LIAO W, *et al.* Srdet++: 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.
- [19] YANG X, YAN J. On the arbitrary-oriented object detection: Classification based approaches revisited [J]. International Journal of Computer Vision, 2022, 130(5): 1340-1365.
- [20] YANG X, HOU L, ZHOU Y, *et al.* Dense label encoding for boundary discontinuity free rotation detection [C] // Proceedings of the IEEE/CVF Conference On Computer Vision and pattern recognition. 2021: 15819-15829.
- [21] LU J, LI T, MA J, *et al.* SAR: Single-stage anchor-free rotating object detection [J]. IEEE Access, 2020, 8: 205902-205912.
- [22] HE X, MA S, HE L, *et al.* Learning rotated inscribed ellipse for oriented object detection in remote sensing images [J]. Remote Sensing, 2021, 13(18): 3622.
- [23] SONG WENGLONG, TANG RUI, YANG KUN, *et al.* An annotation method of arbitrary-oriented rectangle b-box and analysis of its application in remote sensing object detection [J]. Journal of China Institute Water Resources and Hydro-power Research, 2021, 19(1): 165172. [宋文龙, 唐锐, 杨昆, 等. 一种倾斜矩形范围框标注方式及遥感目标检测应用分析 [J]. 中国水利水电科学研究院学报, 2021, 19(1): 165-172.]
- [24] 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. 2017: 2117-2125.
- [25] YANG X, SUN H, SUN X, *et al.* Position detection and direction prediction for arbitrary-oriented ships via multitask rotation region convolutional neural network [J]. IEEE Access, 2018, 6: 50839-50849.
- [26] 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.
- [27] FU Y, WU F, ZHAO J. Context-aware and depthwise-based detection on orbit for remote sensing image [C] // 2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018: 1725-1730.
- [28] HAN K, WANG Y, TIAN Q, *et al.* Ghostnet: More features from cheap operations [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition

- tion. 2020: 1580–1589.
- [29] MA N, ZHANG X, LIU M, *et al.* Activate or not: Learning customized activation [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 8032–8042.
- [30] CAO Y, XU J, LIN S, *et al.* Gcnet: Non-local networks meet squeeze-excitation networks and beyond [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019: 0–0.
- [31] 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. 2017: 2980–2988.
- [32] CHEN Z, CHEN K, LIN W, *et al.* Piou loss: Towards accurate oriented object detection in complex environments [C] // Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16. Springer International Publishing, 2020: 195–211.
- [33] 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 (ICIP). IEEE, 2015: 3735–3739.
- [34] MING Q, MIAO L, ZHOU Z, *et al.* Optimization for arbitrary-oriented object detection via representation invariance loss [J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 1–5.
- [35] LIN Y, FENG P, GUAN J, *et al.* IENet: Interacting embedding one stage anchor free detector for orientation aerial object detection [S]. arXiv preprint arXiv:1912.00969, 2019.
- [36] XIAO Z, QIAN L, SHAO W, *et al.* Axis learning for oriented objects detection in aerial images [J]. Remote Sensing, 2020, 12(6): 908.

Improved Remote Sensing Rotating Object Detection based on CenterNet

LIU Xin¹, HUANG Jin¹, YANG Yingwei¹, LI Jianbo²

(1.School of Electrical Engineering, Southwest Jiaotong University, Chengdu, Sichuan 611756, China;

2.School of Computer and Artificial Intelligence, Southwest Jiaotong University, Chengdu, Sichuan 611756, China)

Abstract: Remote sensing images have low detection accuracy due to the characteristics of different object angles, generally arranged densely, high proportion of small objects and complex background. In view of the inapplicability of the horizontal detection algorithm for remote sensing rotating object detection, and the periodicity and edge interchangeability of angle in the mainstream five-parameter method, a VR-CenterNet is proposed, which used the vector representation to detect the rotating box and design the loss function to avoid the problem of angle regression, and to optimize the high displacement sensitive problem of slender objects. For the high redundancy problem of shallow feature fusion, self-adaptive channel activation is introduced to automatically filter impurity information. In order to strengthen the key point information, an improved global contextual self-adaptive layer activation attention block is introduced in the output of backbone. First, the performance of different algorithms is compared on HRSC2016 and UCAS-AOD data sets. Then, the module ablation experiment is conducted on the two data sets to verify the effectiveness of each improved method. Experimental results show that: 88.48% and 90.35% accuracy are obtained on HRSC2016 and UCAS-AOD data sets respectively. The improved algorithm can improve the detection accuracy of remote sensing rotating objects, and provide another problem-solving idea for the accurate detection of remote sensing rotating objects.

Key words: Remote sensing image; Object detection; Adaptive activation; Attention mechanism; Anchor-free