E-mail: csa@iscas.ac.cn http://www.c-s-a.org.cn Tel: +86-10-62661041

基于全局上下文注意力特征融合金字塔网络的 遥感目标检测^①



¹(南京信息工程大学人工智能学院,南京210044) ²(南京信息工程大学计算机学院,南京210044) ³(南京理工大学计算机科学与工程学院,南京210094) 通信作者:孙文赟, E-mail: wenyunsun@nuist.edu.cn



关键词:遥感图像;定向目标检测;注意力特征融合;特征金字塔网络

引用格式: 孙文赟,车嘉航,金忠.基于全局上下文注意力特征融合金字塔网络的遥感目标检测.计算机系统应用,2024,33(9):114-122. http://www.cs-a.org.cn/1003-3254/9631.html

Remote Sensing Object Detection Based on Global Context Attentional Feature Fusion Pyramid Network

SUN Wen-Yun¹, CHE Jia-Hang², JIN Zhong³

¹(School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing 210044, China) ²(School of Computer Science, Nanjing University of Information Science and Technology, Nanjing 210044, China) ³(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: Remote sensing object detection usually faces challenges such as large variations in image scale, small and densely arranged targets, and high aspect ratios, which make it difficult to achieve high-precision oriented object detection. This study proposes a global context attentional feature fusion pyramid network. First, a triple attentional feature fusion module is designed, which can better fuse features with semantic and scale inconsistencies. Then, an intralayer conditioning method is introduced to improve the module and a global context enhancement network is proposed, which refines deep features containing high-level semantic information to improve the characterization ability. On this basis, a global context attentional feature fusion pyramid network is designed with the idea of global centralized regulation to modulate shallow multi-scale features by using attention-modulated features. Experiments have been conducted on multiple public data sets, and results show that the high-precision evaluation indicators of the proposed network are better than those of the current advanced models.

Key words: remote sensing image; oriented object detection; attentional feature fusion; feature pyramid network



① 收稿时间: 2024-03-13; 修改时间: 2024-04-10; 采用时间: 2024-05-06; csa 在线出版时间: 2024-07-26 CNKI 网络首发时间: 2024-07-29

¹¹⁴ 系统建设 System Construction

1 引言

在遥感图像目标检测等特殊领域中,物体通常密 集排列并且呈现出任意方向的分布.针对这种情况,在 遥感图像中使用通用的水平目标检测方法会导致一系 列问题,包括目标边界框的大量重叠以及目标框包围 背景范围过大等^[1],如图 1(a)所示.相比之下,采用旋 转矩形表示的定向边界框提供了一种更为紧密和准确 地描述遥感图像中具有任意方向目标的有效替代方案. 首先,它能够准确反映目标的真实宽高比.其次,定向 边界框有助于有效地将目标与背景分离,特别是对于 密集排列的目标更具优势.因此,针对遥感图像的目标 检测任务常被称为定向目标检测.



图 1 遥感图像目标检测示意图

在目标检测任务中,精确评估算法性能的指标是 至关重要的. 平均精度 (AP) 被广泛用于量化检测模型 在不同交并比 (IoU) 阈值下的准确性. 其中, AP50 和 AP75 分别代表的是当预测框与真实目标框的交并比 达到 50% 和 75% 时的检测精度, 是主要的评价指标. 目前定向目标检测的算法性能主要是由 AP50 来衡量^[2], 意味着预测框和真实值之间的倾斜交并比 (SkewIoU) 大于 0.5 就被确定为正样本. 然而航空遥感图像的某些 类别物体往往具有宽高比过大的特征,如桥梁、大巴 车和舰船等. AP50 这一评价指标无法精准衡量具有过 大宽高比例的目标的预测角度准确性. 这是因为不同 宽高比的目标框在 SkewIoU 方面呈现两种明显趋势: 当宽高比小于 1.5 时, SkewIoU 对于角度偏差不敏感, 几乎始终大于 0.5; 而当宽高比大于 1.5 时, SkewIoU 随 着角度偏差的增大快速衰减,但仍保持较大范围的角 度偏差容忍限度^[3]. 然而在定向目标检测中, 角度作为 一个独特的参数在实现高精度检测方面发挥着至关重 要的作用.因此,AP50不适合定向目标检测任务,尤其 是高精度检测. 在这一背景下, 更严格的度量 (AP75)

成为关键的性能评价标准. AP75 要求检测算法在 IoU 为 75% 时取得良好的性能,这正是在处理大宽高比目标时尤为重要的一个阈值. 通过这一指标,衡量定向目标检测算法是否更关注角度的预测,并促进算法专注于更有意义的高精度定向目标检测.

此外, 航空遥感图像通常具有广阔的视场, 大多数 目标呈微小且密集排列的状态, 如图 1(b) 所示. 同时由 于航空遥感图像背景的复杂性, 目标周围经常存在各 种干扰因素, 例如建筑物、道路、植被等, 导致目标的 边缘特征和纹理不够明显. 这给目标的密集预测精确 定位带来了巨大挑战, 增加了检测的难度.

针对上述挑战,本文设计了一种全局上下文注意 力特征融合网络 (global contexts attentional feature fusion pyramid network, GCAFFPN), 包含一个三重注意力特 征融合模块 (triple attentional feature fusion, TAFF), 更 有效地融合带有不同语义和尺度特性的特征,从而提 高模型的整体性能.此外,通过层内调节方法的引入,本 文改进并提出全局上下文信息增强网络 (global contexts enhance network, GCEN), 通过对深层特征进行细化, 动态增强密集排列小目标的边缘纹理及.在此基础上, 本文借鉴全局集中调节的思想构建了 GCAFFPN, 通过 注意力调制特征,自上而下地调节浅层多尺度特征,提 高模型对不同尺度下目标的全局感知能力.本文对提 出的方法在两个流行的遥感图像旋转目标检测数据集 基准 (DOTA-v1.0^[1]和 HRSC2016^[4]) 上进行了全面评 估,在不同数据集上均表现出卓越的高精度定向目标 检测性能.总的来说,本文的主要贡献包括以下内容.

 针对遥感图像背景复杂,目标尺寸微小、宽高 比过大的问题,通过引入层内特征调节方法改进并设 计了一个全局上下文信息增强网络(GCEN)对主干网 络提取的最深层特征进行细化调节.此外,本文提出了 一种三重注意力特征融合模块(TAFF)来融合从主干 网络中提取的多尺度特征图以及经过全局上下文信息 增强的深层特征.

 本文在提出的三重注意力特征融合模块和全局 上下文信息增强网络的基础上,改进并设计了一种全 局上下文注意力特征融合金字塔网络 (GCAFFPN),进 一步提高模型对不同尺度间信息融合的能力,提升网 络高精度检测的性能.

● 在目前主流的遥感图像目标检测开源数据集上的广泛实验: 在 DOTA-v1.0 和 HRSC2016 数据集上展

示了本文提出的方法的有效性. GCAFFPN 在定向目标 检测高精度评价指标上实现了较为先进的性能.

2 相关工作

近年来,由于遥感技术应用的深入研究,定向目标 检测作为目标检测应用的扩展受到了广泛关注[2,5-10]. 具体来说, RoI Transformer^[2]提出了一个有监督的旋转 RoI 学习器, 它从 RPN 网络生成的水平 RoI 中学习旋 转的 Rol, 这种方法大大提高了有向物体的检测精度. Yang 等人^[5]在通用目标检测算法 Faster R-CNN 的框 架上构建了一种旋转目标检测方法 SCRDet, 这一方法 针对小目标提出了一种特征融合结构,并引入了多维 注意网络来减少背景噪声.为了进一步解决遥感目标 检测的挑战, Yang 等人在文献[6,7]的工作中使用密集 特征金字塔 (DFPN) 进行多尺度特征融合以及旋转回 归的结合用于遥感图像中船舶的定向检测. Zhang 等 人^[8]提出的 CAD-Net 设计了一个上下文感知网络来学 习全局和局部上下文,并利用注意调制特征和空间和 尺度感知的注意模块来引导网络在适当的特征尺度上 关注更多信息区域,同时抑制无关信息.Xie 等人^[9]提 出的 Oriented R-CNN 设计一个精简的定向区域建议 网络 (Oriented-RPN) 来直接生成高质量的定向候选区 域,大大提高了两阶段检测器的效率.最近,ReDet^[10]使 用旋转等变主干来编码旋转等变特征,并设计了一种 新颖的旋转不变 RiRol Align 模块来从旋转等变特征 中自适应地提取旋转不变特征.

在遥感图像目标检测领域,研究定向目标检测的 方法旨在提升算法对尺度变化的适应性和鲁棒性,这 些方法着眼于解决遥感图像中目标尺度多样性和复杂 背景带来的挑战,为提高检测性能提供有效手段,以满 足实际应用对高效、准确检测的需求.目前的大多数 方法主要仅关注 AP50 评价指标的结果,并在 AP50 指 标上实现了较高性能^[2,11-13].但在面对大巴车、桥梁、 舰船和码头等目标时,继续使用 AP50 指标无法精确衡 量预测角度的准确性.现有的大多数方法往往无法很 好地应对遥感图像中小目标密集排列以及大宽高比目 标等具有挑战性的场景,导致高精度检测性能存在明 显的缺陷.同一模型的实验结果 (AP50) 相较于精度要 求更高的 AP75 评价指标存在较大的差距.

最近,针对高精度定向目标检测这一需求,一些方

法[3,14]通过引入特定的损失函数和优化策略,致力于解 决传统回归损失在处理大宽高比目标时容易导致的问 题. 具体而言, Yang 等人^[14]提出的 KLD 损失将旋转后 的边界框变换成光滑的高斯分布,计算高斯分布之间的 KL 散度作为回归损失, 使其在处理大宽高比目标时能 够更加准确和稳健. 在之后的工作中, Zeng 等人^[3]提出 应用 Transformer 的高精度定向检测器 ARS-DETR, 设计了一种新的角度分类方法,称为 AR-CSL, 采用基 于宽高比的动态权重系数来计算角度损失以更好地适 应目标的尺度和长宽比变化,从而提高检测模型对于 大宽高比目标的角度定位能力. Xu 等人[15]针对边界框 旋转角度不连续问题提出了基于极坐标分解的角度修 正模块 ACM, 旨在以更合理的方式平滑角度标签, 将 模型输出的平滑值转换为锯齿状的角度值,从而消除 了因断点问题导致的角度预测的不稳定性. 通过这一 策略,模型能够更好地利用类 IoU 损失来进行边界框 回归参数的优化,在高精度定向目标检测方面表现出 具有竞争力的性能.

上述工作都提倡使用 AP75 作为高精度定向目标 检测更可靠的指标, 避免 AP50 对角度偏差灵敏度的不 足, 并为高精度定向目标检测方法的发展提供了有效 的途径. 此外, 前面提到的 ReDet^[10]在高精度定向方面 也有着显著提升, 显示了旋转等变主干网络 ReResNet50 对定向目标高精度定位的卓越能力. 因此本文的方法 以 ReDet 为基线模型, 致力于改进不同尺度间信息融 合的能力, 实现高精度定向目标检测.

3 模型概述

1.8

本文提出的整体框架如图 2 所示, 全局上下文注 意力特征融合金字塔网络 (GCAFFPN) 主要由以下部 分组成: 输入图像、主干网络 ReResNet50 用于提取 P2-P5 多尺度特征、改进的全局上下文信息增强网络 (GCEN)、所提出的三重注意力特征融合 (TAFF) 和检 测头部网络 (由分类损失和回归损失共同组成) 用于目 标检测. 如图 2 所示, 本文的 GCAFFPN 以旋转等变检 测器 ReDet^[10]为基线模型, 其建立在经典的两阶段检 测网络 Cascade-RCNN^[16]和 FPN^[17]的结构上. GCAFFPN 的设计使其能够轻松地与现有的检测网络集成, 无需 进行任何适应或额外的监督信息. 更多细节将在后续 小节中展开讨论.



图 2 全局上下文特征融合金字塔网络架构图

3.1 三重注意力特征融合

注意力机制是近年来卷积网络中常用的特征调制 方法,用于提升模型自适应的能力.经典的通道注意力 SE-Net^[18]只利用了全局通道注意,偏向于全局特征范 围的上下文信息.而多尺度通道注意力模块 MS-CAM^[19] 不仅实现了全局尺度注意,还自适应地聚合了局部通 道上下文,这有助于网络关注背景噪音较少的对象,更 有利于微小目标的检测. MS-CAM 通过尺度不同的两 个分支来提取全局特征*G*(*X*)与局部特征*L*(*X*),如图 3(a) 所示.



这部分操作可以表示为:

$$L(\mathbf{X}) = BN(PWConv_2(\sigma(BN(PWConv_1(\mathbf{X})))))$$
(1)

$$F(X) = G(X) \oplus L(X) = GAP(L(X)) \oplus L(X)$$
(2)

其中, PWConv₁将输入特征X的通道数变为原通道数的1/r, r为通道缩放比例, BN(·)表示批归一化层, σ (·)表示 ReLU 激活函数, 通过 PWConv₂ 点卷积将特征 X的通道数恢复到原输入通道数, 获得局部特征L(X). 全局特征的计算公式G(X)需对输入特征X先进行一次 全局平均池化 GAP (global average pooling), 因此局部 特征L(X)与G(X)尺寸信息不同, \oplus 表示两者相加得到 输出F(X)需要采取广播操作, $F(X) \in \mathbb{R}^{C \times H \times W}$ 与原输入 尺寸相同.

在此之后,本文设计了一个能够分配三重特征层 权重的分配模块 (weight allocation module, WAM),如 图 3(b) 所示,包含 stack 堆叠、Softmax 计算和 split 分 割等操作,计算共同输入的三重特征各自的注意力权 重信息.本文在引入多尺度通道注意力 (MS-CAM) 的 基础上,设计了一种三重注意力特征融合模块,见图 3(c). *X*,*Y*,*Z*这 3 个输入分别通过 MS-CAM (只保留未经通 道维度上权重分配的部分),而后将三路输出同时输入 到权重分配模块 (WAM)中,先后经过 stack 在通道维 度上堆叠、Softmax 函数、最后再按原通道数进行分 割,通过训练,让网络在*X*,*Y*,*Z*之间进行自适应地选择 关注,确定 3 个特征层计算注意力权重分配后的结果. 计算之后的权重值用来对三重输入特征*X*,*Y*,*Z*做注意 力操作后得到输出*X'*,*Y'*,*Z'*计算公式如下:

$$X' = X \otimes WAM(F(X)) \tag{3}$$

Y,**Z**的计算过程与式(3)相同,输入的**X**,**Y**,**Z**分别表示深层高级语义特征、当前尺度特征层和下一级感受野更大的特征层.其中,*F*(**X**)来自式(2),WAM表示权重分配的操作,⊗表示两个特征图逐元素乘积.

最终借鉴迭代注意力特征融合 iAFF^[19]的方法,将 第1轮得到的注意力特征*X'*,*Y'*,*Z'*再次以同样的方式 进行迭代得到*X''*,*Y''*,*Z''*,最后对*X''*,*Y''*,*Z''*进行融合. 这一过程可以表示为:

$$\mathbf{X}^{\prime\prime} = \mathbf{X}^{\prime} \otimes WAM(F(\mathbf{X}^{\prime})) \tag{4}$$

$$O = X'' + Y'' + Z''$$
(5)

同理, Y",Z"的计算过程也与式(4)相同,输出融 合特征O∈ℝ^{C×H×W}. 三重注意力特征融合 TAFF 在注 意力模块内聚合多尺度的上下文信息,使初始输入特 征也有尽可能完整的感知,弥补不同尺度间的特征不 一致性,实现三重注意力特征融合.

3.2 全局上下文信息增强网络

现有的特征金字塔网络^[17,20]常使用骨干提取最深 层的 4 层或 5 层特征作为自上而下特征融合的输入, 主要集中在层间特征交互上, 然而却忽略了层内特征 的规范对小目标密集预测任务的有效性^[21,22]. 在本文的 工作中, 改进了一个全局上下文信息增强网络 (GCEN), 其结构如图 4 所示. 首先引入层内特征调节方法: 显式 视觉中心^[23], 同时借鉴 InceptionV3 的理论, 将原来的 输入颈块替换为一个 *Inception* 结构.



图 4 全局上下文信息增强网络结构图

如图 4 所示, GCEN 主要由两个并行的模块组成, 分别是一个轻量化 MLP 和一个可学习的视觉中心, 不 仅可以提取全局远程依赖关系, 还尽可能地保留输入 图像的局部角点区域信息. GCEN 的输出是两个模块 的结果特征图沿着通道维度上连接在一起的特征图. 上述过程可以表述为:

 $\boldsymbol{X} = conv^{1 \times 1}(Cat(MLP(\boldsymbol{X}_{in}), LVC(\boldsymbol{X}_{in})))$ (6)

其中, X为 GCEN 的输出. *MLP*(X_{in})和*LVC*(X_{in})分别 表示两个并行模块所使用的轻量级 *MLP* 和可学习视 觉中心机制的输出特征. *Cat*(·)表示沿通道维度的特征

118 系统建设 System Construction

图拼接. conv^{1×1}(·)为1×1卷积操作,将通道数调整到 256. X_{in}为输入特征平滑的结果,由式(7)求得:

$$X_{\rm in} = \sigma(BN(Inception(P_4))) \tag{7}$$

其中, *Inception*(·)表示 *Inception* 结构, *BN*(·)表示批归 一化层, σ(·)表示 ReLU 激活函数. 为了便于表示,本文 省略了式 (6) 中的 *MLP* 和 *LVC*^[23].

在本文提出的架构中使用 GCEN 对最顶层特征 层 P4 进行全局上下文信息增强,并将经过 GCEN 调节 后的更精细的顶层空间特征称为全局上下文信息增强 特征,这对于遥感图像的小目标宽高比过大、密集排 列的预测任务非常重要.

3.3 全局上下文注意力特征融合金字塔网络

基于第 3.1 节和第 3.2 节提出的 TAFF 和 GCEN, 本文借鉴全局集中调节的思想进一步以自上而下的方式 设计一个全局上下文注意力特征融合金字塔 GCAFFPN. 如图 2 中所示, 本文的 GCEN 首先在金字塔最顶层上 实现, 然后使用所提出的 TAFF 自适应地融合来自主 干网络的多尺度特征以及最顶层 GCEN 得到的上下文 增强特征,利用高级语义信息同时调节所有浅层特征 (即 P3-P1). 实现过程中, 在每个相应的低级特征上, 将 深层调节过的特征上采样到与低级特征相同的空间尺 度,然后沿着通道维度拼接,并将拼接后的特征通过 1×1卷积下采样到 256 通道大小. 最后将这些多尺度特 征聚合成一个解耦的头部检测网络进行分类和回归任 务. 最终设计的 GCAFFPN, 能够在自顶向下路径中显 式地增加特征金字塔每一层多尺度表达的能力,避免 直接融合不同尺度的特征所带来大量的冗余信息和冲 突信息,从而提升模型处理遥感图像背景复杂且目标 尺度变化大等关键问题的能力.

4 实验

本节介绍实验,包括数据集和评估指标、实现细节、两个公共遥感目标检测数据集上的实验结果以及 所提出的 GCAFFPN 的消融研究.

4.1 数据集和实现细节

DOTA-v1.0^[1]是遥感图像定向目标检测应用中最 大的数据集之一. 该数据集涵盖了各种图像尺寸、地 面样本距离 (GSDs)、传感器类型等多方面的差异, 捕 捉到的物体在尺度、形状和方向上呈现丰富的变化. 它由来自不同传感器和平台的 2806 个大型航拍图像 组成,其尺寸范围从大约 800×800 到 4000×4000 不等, 包含了 188282 个实例.本文在训练和验证过程中使用 了训练集和验证集,并在测试中使用了测试集.为了处 理大尺寸图像,本文将图像划分为 1024×1024 的子图 像,重叠 200 像素.

HRSC2016^[4]是另一个具有挑战性的任意方向船 舶检测的遥感图像数据集,包含1061张图片,大小从 300×300到1500×900不等.训练集(436张图像)和验 证集(181张图像)被用于模型训练,其余444张图像 则用于测试.所有图像尺寸都被调整为800,同时锁定 宽高比.在训练过程中使用随机水平翻转.

本文的实验设置与基线模型 ReDet 在训练和测试 期间的设置基本一致.使用一台 NVIDIA RTX 3090 GPU 进行训练和测试,本文中的所有模型均采用基于 PyTorch 框架的 MMRotate^[24]实现,并使用 SGD 优化 器进行训练.初始学习率设定为 0.0025,在每个衰减步 骤中学习率除以 10,动量衰减和权重衰减分别为 0.9 和 0.0001.每个批量包含 2 张图像,默认使用"3x"训练计 划.并且只以单尺度训练和测试,因此本文模型只与基 线方法的未经数据增强的结果进行比较.

采用上述方法,本文在两个具有不同特性和难度的遥感图像数据集,即 DOTA-v1.0 和 HRSC2016 上进行了详尽的实验评估.在主张高精度定向目标检测的方法^[3,14]之后,本文继续采用 AP75 作为主要指标,AP50 为辅助指标,旨在全面验证所提出方法在应对遥感图像复杂场景时的鲁棒性以及高精度检测的性能表现.

4.2 消融实验

为了研究全局上下文信息增强网络 (GCEN) 和三 重注意力特征融合 (TAFF) 在高精度定向目标检测任 务中的有效性,本文在表 1 中进行了详细的消融研究, 使用带有旋转等变金字塔 (ReFPN) 的 ReDet^[10]作为基 线.为了公平比较,本文设置相同的超参数进行消融 实验.

表 1	不同组件在 DOTA-v1.0 上的消融研究 (%)				
GCEN	TAFF	AP50	AP75		
	_	75.78	50.40		
\checkmark	—	75.82 (+0.04)	51.60 (+1.20)		
\checkmark	\checkmark	76.06 (+0.24)	52.25 (+0.65)		

从表 1 可知, GCEN 和 TAFF 都能够显著提升高 精度定向检测的性能,在 AP75 指标上, GCEN 提高了

1.20%, 这表明 GCEN 对网络深层特征提取全局上下文 信息的有效性, 从而大幅提升模型的检测性能. 而加入 的 TAFF 模块进一步向浅层特征中融合来自 GCEN 的 高级语义信息, 以达到充分利用深层高级语义改善高 精度定向检测效果的目的. 这一实验的 AP75 在提升 0.65% 的同时, AP50 的结果也有着一定提升. 具体而 言, 在高精度定向目标检测任务中, 当同时使用 GCEN 和 TAFF 模块时, 均较不添加或单独使用时提升更为 明显, 这表明全局上下文信息增强和多尺度特征融合 机制的结合使用对于大宽高比目标的深层语义信息激 活能起到相互促进的效果, 对于解决遥感图像中高精 度目标检测的挑战具有明显的贡献.

4.3 对比实验

本文将提出的 GCAFFPN 方法与 8 种应用 DOTA 数据集的定向目标检测方法和 5 种应用 HRSC2016 数 据集的目标检测方法进行了比较. 表 2 和表 3 分别报 告了 DOTA-v1.0 和 HRSC2016 数据集的详细对比结 果. 其中 R50 表示 ResNet50^[25], Swin-T 表示 Swin-Transformer^[26], ReR50 表示 ReResNet50^[10].

模型	Backbone	AP50	AP75			
Rotated RetinaNet ^[27]	R50-FPN	69.23	40.96			
R3Det ^[11]	R50-FPN 73.74		41.69			
RoI Transformer ^[2]	R50-FPN 74.61		46.54			
S2A-Net ^[12]	R50-FPN	74.12	43.93			
Oriented R-CNN ^[9]	R50-FPN	75.87	47.73			
ARS-DETR ^[3]	R50-FPN	74.16	49.41			
ARS-DETR ^[3]	Swin-T	75.47	51.77			
ReDet ^[10]	ReR50-ReFPN	75.78	50.40			
GCAFFPN	ReR50-GCAFFPN	76.06	52.25			

表 2 不同方法在 DOTA-v1.0 上的比较 (%)

表 3 不同方法在 HRSC2016 上的比较 (%)

模型	Backbone	mAP (2007)	mAP (2012)
RoI Transformer ^[2]	R101	86.20	_
R3Det ^[11]	R101	89.26	96.01
S2A-Net ^[12]	R101	90.17	95.01
Oriented R-CNN ^[9]	R50	90.40	96.50
ReDet ^[10]	ReR50	90.46	97.63
GCAFFPN	ReR50	90.52	97.87

• DOTA-v1.0 的结果. 如表 2, 本文在 DOTA-v1.0 数据集上进行了实验, 将提出的 GCAFFPN 与遥感目标检测领域其他最先进的方法进行了实验对比. 所有结果均通过单尺度训练和测试得到. 在 AP50 指标上,

120 系统建设 System Construction

本文方法的准确率达到了 76.06%, 与当前最先进的方 法相比具有很强的竞争力, 这表明 GCAFFPN 在检测 大多数定向目标时具有较高的准确率. 然而, 更值得关 注的是 GCAFFPN 在 AP75 指标上的表现. 在这一更为 严格的高精度定向评价标准下, 本文提出的方法取得 了 52.25% 的准确率, 对基线模型 ReDet (双阶段旋转 等变检测器) 有较大的提升, 提高了 1.85%. 这一结果 表明 GCAFFPN 在预测角度时的偏差更小, 尤其在处 理大宽高比目标的情况下表现出色. 与其他先进方法 相比, GCAFFPN 在更严格的高精度评价指标下取得了 更好的性能, 这进一步证实了其在高精度定向目标检 测中的有效性.

• HRSC2016 的结果. 如表 3, 列出了不同方法在 PASCAL VOC 2007 和 VOC 2012 指标下的评估结果. 本文提出的方法在 mAP (2007) 和 mAP (2012) 两个指 标上分别将性能提高了 0.06%、0.24%, 都展现出了令 人鼓舞的性能表现. 值得注意的是, 本文方法能够在两 个不同的评价标准下都取得最佳精度, 这进一步验证 了其在定向目标检测任务中的有效性和可靠性. 这些 结果反映了本文方法在处理复杂背景遥感图像中的任 意方向目标时的卓越性能, 尤其是在小目标密集排列 和大宽高比目标高精度定向检测的情况下.

4.4 可视化分析

除了使用量化评价指标对本文模型的高精度定向 检测效果进行分析外,图 5 列出了基线模型与本文方 法检测结果的可视化对比.因篇幅限制,仅列举出 DOTA-v1.0数据集中 5 种类别的检测效果.从图 5(a) 和 (b)中可以观察到,基线模型在针对大宽高比目标 (如桥梁、海港)的高精度检测时存在角度偏差的缺陷, 而本文提出的 GCAFFPN 方法在处理大宽高比目标时 相对基线模型的角度偏差问题有一定的减轻.此外, 图 5(c)-(e)展示了基线模型在检测密集排列的小目标 (如舰船、直升机、泳池)时均存在一定程度的漏检和 误检情况,相比之下,本文方法相对于基线模型在漏检 和误检方面有所改善,检测精度得到了显著提升.

总体而言,本文的实验结果证明了集成 GCEN 和 TAFF 的 GCAFFPN 对高精度检测性能的提升的有效 性,在不同数据集上均表现出较高的准确率,具有较强 的竞争力.为解决遥感图像高精度定向目标检测问题 提供了可行的解决方案.



图 5 本文模型检测结果可视化对比

5 总结

针对航空遥感图像中的复杂背景、微小目标、宽 高比过大等问题,本文提出了一种全局上下文注意力 特征融合金字塔网络(GCAFFPN).通过全局上下文信 息增强网络(GCEN)对最深层特征进行层内调节,丰 富目标与背景之间的高级语义信息和空间特征.并设 计了一种三重注意力特征融合模块(TAFF),实现了多 重输入特征自适应地融合,有效解决了输入特征之间 语义和尺度的不一致问题.通过全局上下文信息增强 和注意力特征融合,自上而下地对多尺度特征进行全 局集中调节,提高了模型对不同尺度信息的整合能力, 进一步优化模型的高精度定向目标检测的表现.与其 他先进的定向目标检测方法相比,本文提出的方法在 主流公开数据集中更严格的高精度评价指标上表现出 色.这表明本文的方法在应对遥感图像中具有挑战性 的场景时取得了显著的性能提升.

参考文南

- Xia GS, Bai X, Ding J, et al. DOTA: A large-scale dataset for object detection in aerial images. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 3974–3983. [doi: 10.1109/cvpr.2018.00418]
- 2 Ding J, Xue N, Long Y, *et al.* Learning RoI Transformer for oriented object detection in aerial images. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 2844–2853. [doi: 10. 1109/CVPR.2019.00296]
- 3 Zeng Y, Chen YS, Yang X, et al. ARS-DETR: Aspect ratio-

sensitive detection Transformer for aerial oriented object detection. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1–15. [doi: 10.1109/TGRS.2024.3364713]

- 4 Liu ZK, Yuan L, Weng LB, et al. A high resolution optical satellite image dataset for ship recognition and some new baselines. Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods. Porto: SciTePress,2017.324–331.[doi:10.5220/0006120603240331]
- 5 Yang X, Yang JR, Yan JC, et al. SCRDet: Towards more robust detection for small, cluttered and rotated objects. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 8231–8240. [doi: 10.1109/ICCV.2019.00832]
- 6 Yang X, Sun H, Sun X, *et al.* Position detection and direction prediction for arbitrary-oriented ships via multitask rotation region convolutional neural network. IEEE Access, 2018, 6: 50839–50849. [doi: 10.1109/ACCESS.2018.2869884]
- 7 Yang X, Sun H, Fu K, *et al.* Automatic ship detection in remote sensing images from Google earth of complex scenes based on multiscale rotation dense feature pyramid networks. Remote Sensing, 2018, 10(1): 132. [doi: 10.3390/rs10010132]
- 8 Zhang GJ, Lu SJ, Zhang W. CAD-Net: A context-aware detection network for objects in remote sensing imagery. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(12): 10015–10024. [doi: 10.1109/TGRS.2019. 2930982]
- 9 Xie XX, Cheng G, Wang JB, *et al.* Oriented R-CNN for object detection. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 3500–3509. [doi: 10.1109/ICCV48922.2021. 00350]
- 10 Han JM, Ding J, Xue N, et al. ReDet: A rotation-equivariant

detector for aerial object detection. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 2785–2794. [doi: 10. 1109/CVPR46437.2021.00281]

- 11 Yang X, Yan JC, Feng ZM, *et al.* R3Det: Refined singlestage detector with feature refinement for rotating object. Proceedings of the 35th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2021. 3163–3171. [doi: 10. 1609/aaai.v35i4.16426]
- 12 Han JM, Ding J, Li J, *et al.* Align deep features for oriented object detection. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5602511. [doi: 10.1109/TGRS. 2021.3062048]
- 13 Yang X, Yan JC. On the arbitrary-oriented object detection: Classification based approaches revisited. International Journal of Computer Vision, 2022, 130(5): 1340–1365. [doi: 10.1007/s11263-022-01593-w]
- 14 Yang X, Yang X, Yang J, et al. Learning high-precision bounding box for rotated object detection via Kullback-Leibler divergence. Proceedings of the 2021 International Conference on Neural Information Processing Systems. NeurIPS, 2021. 18381–18394.
- 15 Xu H, Liu XY, Xu HN, *et al.* Rethinking boundary discontinuity problem for oriented object detection. arXiv: 2305.10061, 2023.
- 16 Cai ZW, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 6154–6162. [doi: 10.1109/CVPR. 2018.00644]
- 17 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 936–944. [doi: 10.1109/CVPR.2017. 106]
- 18 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141. [doi: 10.1109/CVPR.2018.00745]
- 19 Dai YM, Gieseke F, Oehmcke S, et al. Attentional feature

fusion. Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021. 3559–3568. [doi: 10.1109/WACV48630.2021.00360]

- 20 Tan MX, Pang RM, Le QV. EfficientDet: Scalable and efficient object detection. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 10778–10787. [doi: 10. 1109/CVPR42600.2020.01079]
- 21 Cao Y, Xu JR, Lin S, *et al.* GCNet: Non-local networks meet squeeze-excitation networks and beyond. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop. Seoul: IEEE, 2019. 1971–1980. [doi: 10. 1109/ICCVW.2019.00246]
- 22 Zhang D, Zhang HW, Tang JH, *et al.* Feature pyramid Transformer. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 323–339. [doi: 10.1007/978-3-030-58604-1 20]
- 23 Quan Y, Zhang D, Zhang LY, *et al.* Centralized feature pyramid for object detection. IEEE Transactions on Image Processing, 2023, 32: 4341–4354. [doi: 10.1109/TIP.2023. 3297408]
- 24 Zhou Y, Yang X, Zhang GF, *et al.* MMRotate: A rotated object detection benchmark using PyTorch. Proceedings of the 30th ACM International Conference on Multimedia. Lisboa: ACM, 2022. 7331–7334. [doi: 10.1145/3503161. 3548541]
- 25 He KM, Gkioxari G, Dollár P, et al. Mask R-CNN. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2980–2988. [doi: 10. 1109/ICCV.2017.322]
- 26 Liu Z, Lin YT, Cao Y, et al. Swin Transformer: Hierarchical vision Transformer using shifted windows. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 9992–10002. [doi: 10.1109/ ICCV48922.2021.00986]
- 27 Lin TY, Goyal P, Girshick R, *et al.* Focal loss for dense object detection. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2999–3007. [doi: 10.1109/icev.2017.324]

(校对责编:张重毅)