

面向遥感图像目标感知的群目标检测框架

张鸿伟¹, 金磊¹, 邹学超², 方宇强³, 尹璐⁴, 赵健⁵, 兴军亮⁶

1. 北京邮电大学 电子工程学院, 北京 100876;
2. 青海大学 计算机技术与应用学院, 西宁 810016;
3. 航天工程大学 研究生院, 北京 101416;
4. 北京遥感信息研究所, 北京 100192;
5. 北方电子设备研究所, 北京 100071;
6. 清华大学 计算机学院, 北京 100084

摘要: 光学遥感是航天侦察和地质勘测中的常用技术, 拍摄得到的可见光图像能够提供非常丰富的信息, 在目标监视、态势预判等方面都具有重要应用。近年来以轮船、飞机等物体检测为代表的光学遥感图像目标感知取得了显著进展, 但对于目标尺度变化大, 目标数量多而小的遥感图像目标感知场景中还存在巨大技术挑战, 也就是在当前的光学遥感图像目标感知场景存在很多目标小并且多目标集中的情况, 容易导致误检和漏检。为了解决现有遥感图像目标检测算法不同目标独立检测的内在低效性, 本文提出了一种新的检测框架, 即群目标检测, 以期通过检测群目标的状态信息来缓解单一目标感知信息不足、可靠性差等问题, 进而得到更为可靠的多目标检测结果。本文首先对群目标的概念进行定义, 然后基于该定义提出了一种群目标自动化标注方案, 在公开数据集上对原有标签进行分析, 无需任何手动标注, 就能得到含有群目标标注的注释信息。基于群目标自动化标注, 本文提出了群目标检测算法, 即在检测群目标的同时, 利用群目标的空间约束提升单一目标检测结果。实验证明, 与近年来的遥感图像检测算法相比, 本文提出的群目标检测在最热门的大型遥感目标检测数据集DOTA上验证时, 性能最佳。

关键词: 遥感图像, 目标检测, 边界框, 群目标, 自动化标注, DOTA, 目标感知, 多目标

中图分类号: P2

引用格式: 张鸿伟, 金磊, 邹学超, 方宇强, 尹璐, 赵健, 兴军亮. 2024. 面向遥感图像目标感知的群目标检测框架. 遥感学报, 28(7): 1802-1811

Zhang H W, Jin L, Zou X C, Fang Y Q, Yin L, Zhao J and Xing J L. 2024. A group object detection framework for remote sensing image object perception. National Remote Sensing Bulletin, 28(7): 1802-1811 [DOI: 10.11834/jrs.20233263]

1 引言

遥感技术是获取大规模地理信息的重要手段, 随着遥感图像获取技术的不断改进和发展, 遥感图像被广泛应用于土地利用、城市规划、交通监测、灾害防治、生态保护 (Chen等, 2006; Lenhart等, 2008; Liu和Wu, 2016; Lopez和Frohn, 2017; 李强等, 2022) 等众多民用领域。在遥感图像中, 地物目标是最关键的信息之一, 包括建筑物、道路、水体、森林等。然而, 由于遥感图像具有分

辨率高、数据量大、复杂多样等特点, 传统目标检测方法已经无法满足对大规模遥感图像数据的自动化处理需求, 因此需要开发高效、准确、自动化的遥感图像目标检测技术。

遥感目标检测主要目的是从遥感图像中自动检测出目标物。目前, 遥感目标检测已经取得了显著的研究成果, 如基于深度学习的目标检测算法和多源遥感数据融合技术等, 但仍然存在一些问题, 如遥感图像中目标物种类繁多、光照和天气等因素的干扰以及遥感图像分辨率低等。因此,

收稿日期: 2023-06-30; 预印本: 2023-12-03

基金项目: 卫星信息智能处理与应用技术实验室基金 (编号: 2022-ZZKY-JJ-06-01)

第一作者简介: 张鸿伟, 研究方向为计算机视觉。E-mail: zhw@bupt.edu.cn

通信作者简介: 兴军亮, 研究方向为视觉感知与博弈决策。E-mail: jlking@tsinghua.edu.cn

如何提高遥感目标检测的准确率和效率，是当前遥感目标检测研究亟待解决的问题。

现有遥感图像目标检测方法主要从以下3个角度来提升性能：(1) 增强特征鲁棒性，能够提高检测的性能和稳定性，使得算法在不同的场景和条件下都能够准确地识别目标，现有手段主要有多尺度特征融合，多模态特征融合及注意力机制。代表性算法有 ICN (Azimi 等, 2019), CAD-Net (Zhang 等, 2019), DRN (Pan 等, 2020) 和改进的 CenterNet (田壮壮 等, 2023)。ICN 提出了一个带多个不同尺度卷积核的级联网络来提取鲁棒的多尺度特征，以提升检测精度。CAD-Net 利用注意力调制特征以及全局和局部上下文来解决从遥感图像中检测目标的挑战。DRN 提出了一个由特征选择模块和动态优化头组成的动态优化网络，它能够自动调制感受野并以对象感知的方式动态地改进检测效果。(2) 改善损失传导机制，从这方面可以提高网络对目标的检测准确度和鲁棒性。其中，多尺度损失，联合损失，梯度裁剪应用广泛，代表性算法有 RetinaNet (Lin 等, 2017), DAL (Ming 等, 2021) 和 R3Det (Yang 等, 2021)。DAL 提出了一种动态锚点学习方法，利用新定义的匹配度算法来综合评估锚点的定位潜力，并进行更有效的标签分配。R3Det 提出了一个端到端的精细化单级旋转检测器，通过使用从粗到细粒度的渐进回归方法来快速准确地检测目标。(3) 从语义级提升检测性能，能够利用更高级别的语义信息，包括目标物体的形状、大小、方向和背景等因素来提高遥感目标检测的准确性和鲁棒性。代表性算法有 ROI Transformer (Ding 等, 2019), CenterMap (Wang 等, 2021) 和 Mask-OBB (Wang 等, 2021)。CenterMap 将边界框回归转换为概率图预测问题，从而在很大程度上消除了目标定义和背景像素上的模糊性。Mask-OBB 将边界框回归转换为像素级分类问题，这可以在很大程度上消除歧义。

上述方法虽然取得了较好的检测精度，但仍存在一些问题。增强特征鲁棒性可以普遍提高检测的性能和稳定性，适应更多的场景，而针对目标小而密集的场景表现得不够出色。改善损失传导机制可以提高网络对目标的检测准确度和鲁棒性，而无法提高小目标的召回率。同时，现有的语义关系联合上下文背景信息，忽略了周围前景

信息可能带来的影响，因此需要更高级别的语义信息来提高检测的准确性。本文通过对现有算法分析发现造成上述问题的根本原因是单目标检测因其受限的感受野和内在的学习机制始终无法充分地利用上下文信息。因此，本文创新性地构建了群目标检测框架，首先提出群目标的概念定义，并采用搜索式自动化标注设计了群目标自动化标注方案，最后提出了一个群目标检测算法，旨在充分挖掘多个相同或者不同类型目标之间的上下文关联信息，对同属一个群体的目标进行关联，从而更进一步地提升检测精度。

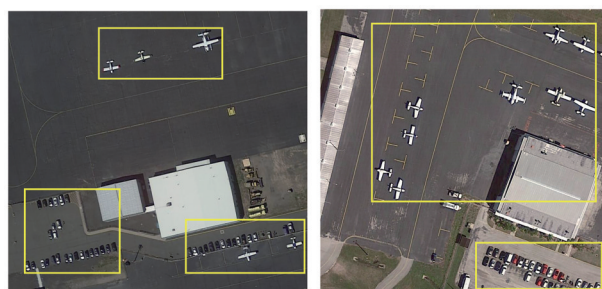
2 群目标检测框架

2.1 群目标的概念定义

传统的目标检测算法是独立地对每个目标进行逐一定位，忽略了多个相同或不同类型目标之间的上下文情景信息，空间位置信息，而群目标检测不是独立地对每个目标进行逐一定位，能够充分挖掘多个相同或者不同类型目标之间的上下文情境信息，对不同位置的目标进行关联分析。本文提出通过检测群目标，将单目标在空间上加以约束，从而提升检测效果。

一般来讲，群目标可以定义为：在表现形状、空间位置等方面具有相符性的多个目标联合体被称为一组群目标。群目标既可以包含不同类型、但相互关联的多个目标，也可以包含同一类型多个实例联合在一起的多个目标，可以根据不同的目标特性、关联关系及应用场景进行定义。在具体实践中，群目标检测需要考虑目标之间的关联性和群体结构，即同时检测出包含多个目标实例的群目标整体，并识别它们之间的群体关系或群体属性。

针对本文关注的遥感图像目标感知应用场景，基于以下两个方面给出群目标的定义：(1) 空间接近：当不同目标间的距离非常接近时，这些目标无论是否属于同一类别都可以被视为一个群体。如图 1 (a) 所示，右下角的小车和飞机排列紧密，两类物体共同组成群目标。(2) 类别相同：当目标之间属于相同的类别或者具有相似的特征时，这些目标即使不符合空间接近也可以被视为一个群体。如图 1 (b) 所示，所有的小车组成群目标，所有的飞机组成群目标。



(a)空间接近 (b)类别相同
(a) Spatial approach (b) Same category

图1 群目标两种定义方式

Fig. 1 Two definition of group object

上述两种定义方式对群目标的侧重点有所不同，第1种定义方式侧重于空间上的距离关系，而第2种定义方式侧重于目标间的语义关联。对比两种定义方式，空间接近更符合实际情况，因此本文将基于空间接近的方式进行标注。

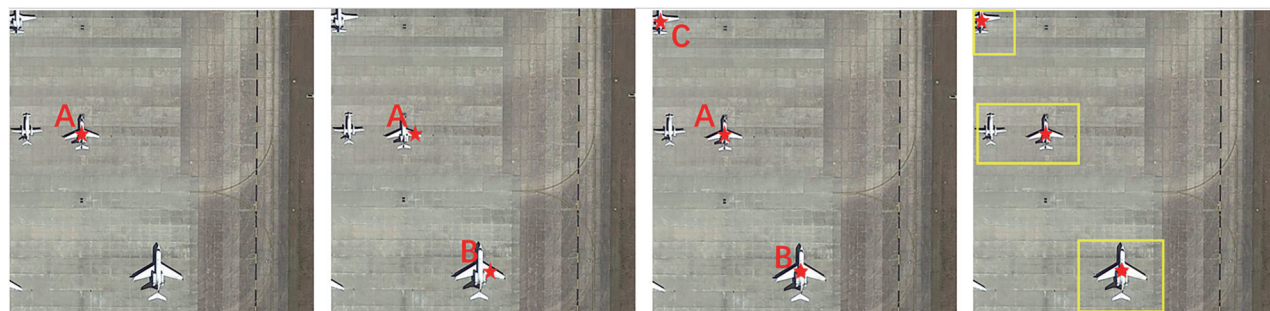
2.2 群目标自动化标注方案

2.2.1 聚类式自动化标注

基于上述群目标的第一种定义，即空间接近，本文在已有目标检测实例标注的基础上设计了群目标自动化标注方案，以DOTA-v1.0数据集为例，标注结果如图3所示。本文基于Kmeans聚类算法(易茹, 2020)在已有数据标签上进行标注，Kmeans算法根据给定的分类数量随机选取一定簇的数量进行聚类，而在遥感数据集中，目标大小

及数量不是固定的，因此需要根据空间距离分布和数量大小动态地调整分簇数量。具体来说，计算每张图片最远两个目标间的距离 dis ，按照以下逻辑计算分类数量 $group_num$ 。具体计算过程如下：当 $dis < dis1$ 时， $group_num=1$ ；如果 $dis > dis2$ ，若 $num < nums1$ ， $group_num=3$ ，若 $num > nums2$ ， $group_num=5$ ，其余情况 $group_num=4$ ；如果 $dis > dis1$ 且 $< dis2$ ，若 $num > nums1$ ， $group_num=2$ ，若 $num > nums2$ ， $group_num=4$ ，其余情况 $group_num=3$ 。

其中， $dis1$ 、 $dis2$ 代表距离阈值， $nums1$ 、 $nums2$ 代表数量阈值，得到分类数量后，为了避免K-means算法随机选取分簇中心造成分类结果重叠分布，相互覆盖，本文在分簇中心点选择上做出调整，K-means方法一次性随机选取 n 个点作为初始分簇中心，不同的是，本文首先选取一个中心点，然后计算其和所有其他点的相对距离，选取最远的点作为下一个中心点，然后再次遍历所有其他点到已有中心点的距离平均值，选取平均距离最大的点作为下一个中心点。如图2所示，首先随机选取一点A(图2(a))，计算其他目标与之距离，得到距离最远的B(图2(b))，然后计算剩余目标与已有中心点的距离，得到平均距离最大的C点(图2(c))，由此得到3个聚类中心进行聚类，最终群目标标注如图2(d)所示。针对不同数据集，需要根据其分布动态调整自适应分簇阈值区间。基于此标注方法，得到的标注结果如图3所示。



(a)随机标记 (b)距离最远标记 (c)平均距离最远标记 (d)标注结果
(a)Random labelling (b)Farthest distance labelling (c) Farthest average distance labelling (d) Labelling result

图2 群目标自动化标注流程

Fig. 2 Automatic labelling process of group object

2.2.2 搜索式自动化标注

由于聚类方法的初始点选取是随机的，聚类结果存在一定的不确定性，因此本文提出了第2种

聚类方法，搜索式聚类。受图论思想启发，结合空间接近的定义，本文制定了搜索式群目标标注的规则，即对于任一目标A，如果在以其中心点为

圆心, 半径为 r 的范围内, 存在另一个目标B, 那么A和B共同归属于一个群目标。对于一张图片而言, 遍历所有的目标, 计算每一个目标和哪些目标的距离在规定范围内, 记录所有的符合条件的目标, 然后根据这些目标生成矩形框, 标注为集群, 结果如图4所示。相比于聚类式方法, 搜索式方法更加简单直观, 具有更好的确定性, 只要给定搜索距离, 分类结果就是确定的, 而聚类方法初始一个点的选择是随机的。因此, 本文后续的实验选用了第二种搜索式方法进行标注, 并在消融实验证明了该方法的优越性, 同时对搜索距离进行了定性分析。



□ 群目标 □ 单目标飞机

图3 聚类式标注示例

Fig.3 Examples of K-means labelling



□ 群目标 □ 单目标飞机

图4 搜索式标注示例

Fig.4 Examples of search labeling

2.3 群目标检测算法

2.3.1 概述

与传统单一目标检测不同, 群目标检测在大范围上预测一群目标的位置, 为了利用群目标带来的空间位置约束, 使单目标检测向群目标靠拢, 本文提出了对称区域卷积神经网络 (Dual-RCNN), 利用群中心约束关联群目标和单目标的位置信息。算法整体流程如图5所示, 给定一个输入的图像, 通过骨干特征提取网络获得多尺度特征, 然后将特征输入到上下两个并行检测分支, 即群目标检测分支和单目标检测分支。每个分支与传统检测算法 Faster RCNN (Girshick, 2015; Ren 等, 2017) 类似。每个分支除了预测群目标和单目标之外, 本方法还设计了群中心约束, 对上下2个分支的提议检测框结果进行匹配, 利用中心点偏移量在空间位置上加以约束, 以优化单目标以及群目标分支检测结果。

2.3.2 单目标分支

该分支由 RPN, ROIpooling 以及 RCNNPredictor 构成。RPN 由 AnchorGenerator 及 RPNHead 构成, RPNHead 由两个 3×3 的卷积分支构成, 分别预测由 AnchorGenerator 产生的锚框的类别 (前景或背景), 以及该锚框的修正参数 (中心点偏移量, 长宽偏移量), 最后 RPNHead 和 AnchorGenerator 会得到含有目标的提议框, 由于目标大小不同, 得到的提议框大小是变化的。通过 ROIpooling 将维度不一致的提议框转化成统一大小的特征向量, 然后做展平处理, 之后通过 RCNNPredictor 进行预测, 该模块主要由几个全连接层构成, 得到最终的检测类别和检测框位置。

2.3.3 群目标分支

群目标分支和单目标分支组成类似, 同样由 RPN, ROIpooling 以及 RCNNPredictor 构成。不同的地方在于群目标分支对应的地面实测为群目标标注。

2.3.4 群中心约束

为了充分利用群目标提供的空间信息, 本文在算法中引入了群中心约束, 从群目标检测分支得到群目标预测提议框, 从单目标检测分支得到

单目标预测提议框，获得这两个结果后，对两者进行匹配，为每个单目标匹配对应的群目标，即该单目标归属于哪个群。具体匹配规则为，针对

每一个单目标，计算所有群目标中心点和该单目标中心点的绝对距离，距离最近的群目标为该单目标的唯一匹配（图6）。

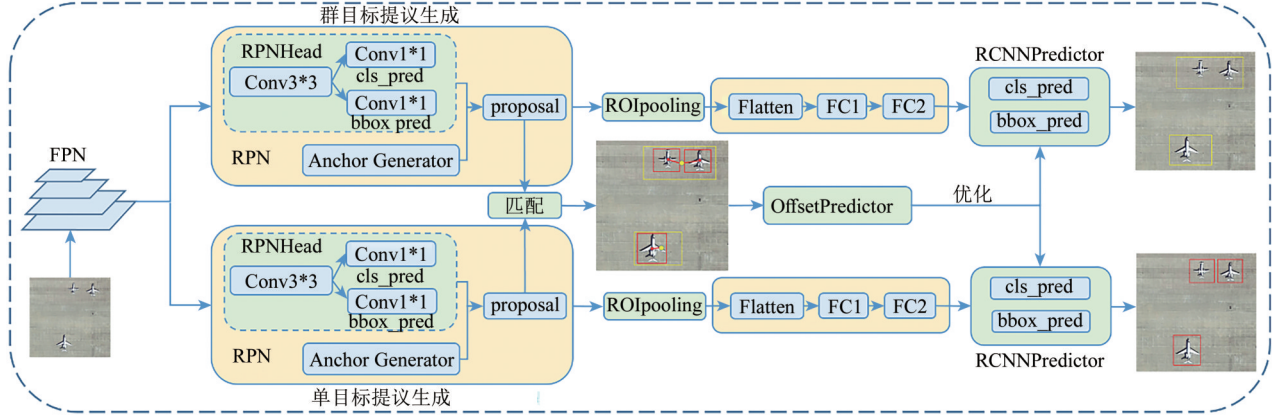


图5 群目标检测算法框架图

Fig. 5 Framework of group object detection

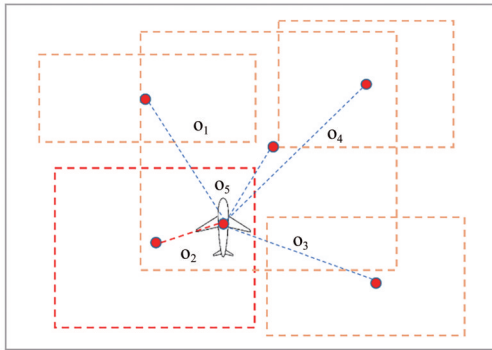


图6 单目标匹配群目标示意图

Fig. 6 Single object matches group object

如图6所示，对于一个单目标来说，在 o_1 , o_2 , o_3 , o_4 , o_5 这5个距离中， o_2 最小，所以该单目标就与 o_2 对应的群目标匹配。匹配过程中，每个单目标匹配一个群目标，而每个群目标会匹配多个单目标。匹配完成后，计算所有单一目标与其所匹配的群目标之间的中心点相对偏移量，让所有单目标预测该偏移（对应框架图中OffsetPredictor）。并计算匹配的单目标和群目标中心点的绝对距离，然后除以其所匹配的群目标框的对角距离，得到相对偏移量。按照同样的方法计算真实的相对偏移量，相对偏移损失计算公式如下：

$$L_{\text{offset}} = \frac{1}{N} \sum_{k=1}^N |\hat{o}_{pk} - o_k| \quad (1)$$

式中， \hat{o}_{pk} 代表预测的偏移值， o_k 代表真实的偏移值， N 代表预测的单目标提议数量。

模型总的损失函数为

$$L_{\text{total}} = L_{\text{RPN}} + L_{\text{RCNN}} + L_{\text{offset}} \quad (2)$$

式中， L_{RPN} 代表RPN部分的提议框分类和回归损失， L_{RCNN} 代表RCNNPredictor部分的分类和回归损失。 L_{RPN} 计算公式如下：

$$L_{\text{RPN}} = \sum_i L_{\text{cls}}(p_i, p_i^*) + \sum_i p_i^* L_{\text{reg}}(t_i, t_i^*) \quad (3)$$

式中， L_{cls} 是Softmax loss，用于分类锚框属于前景还是背景； L_{reg} 为Smooth L1 loss，用于修正锚框，前面乘一个 p_i^* 表示只回归有目标的框； p_i 表示网络预测出来第 i 个锚框是目标的概率， p_i^* 表示对应真实值。若第 i 个锚框与某个真实目标的交并比 >0.7 ，则 $p_i^*=1$ ；若 $\text{IOU}<0.3$ ，则 $p_i^*=0$ ； t_i 表示参数化坐标向量，表示预测框和锚框之间的偏移； t_i^* 表示对应真实偏移。 L_{RCNN} 的表示方法和 L_{RPN} 相同，不同的在于回归的类别数量不同。

2.3.5 训练

给定一个输入大小为1024像素×1024像素×3通道的图片，经过骨干特征提取网络后，获得5个多尺度特征，经过RPN网络，对所有提议框筛选得到分数排在前2000的提议框。得到2000个提议之后，在训练阶段，从2000个提议里随机选取256个正样本，256个负样本用于后续的训练，然后通过ROI融合所有尺度，得到（批大小，512，7，7）的特征，最后通过RCNNPredictor得到输出结果，包括512个样本的分类结果（512，类别数）以及每个样本的检测框位置参数（512，4）。

2.3.5 推理

在推理过程，移除偏移分支，直接得到单目

标和群目标提议，结合预测的回归参数计算最终的bbox坐标，移除背景信息和低概率小尺寸目标，然后进行NMS处理，返回最终预测结果。

3 测试实验

为了验证本方法的有效性，本文在目前最热门的DOTA-v1.0 (Xia等, 2018; Ding等, 2022)数据集上进行了广泛实验。

3.1 数据准备

DOTA-v1.0数据集包含了2806张来自不同传感器和平台的航空影像，每张影像大小都在4000像素×4000像素内，分为飞机(PL)、舰船(SH)、储罐(ST)、棒球场(BD)、网球场(TC)、篮球场(BC)、田径场(GTF)、海港(HA)、桥梁(BR)、大型车辆(LV)、小型车辆(SV)、直升机(HC)、环岛(RA)、足球场(SBF)、游泳池(SP) 15个类别，总计188282个目标，每个目标都标记为一个带方向的边界框。同时，对每张图片添加了额外的群目标标注，每张图像被分成1—5个集群，每个集群以矩形框的形式进行标注。本文使用训练集和验证集进行训练。通过将测试结

果发送到DOTA的在线评估服务器来获得最终的检测精度。

3.2 训练参数配置

本方法使用4张英伟达3090显卡来进行训练，训练中单卡批大小为1，采用同步批归一化，测试时也是在1张3090显卡进行，实验在mmdetection开源框架上进行。在训练过程中统一使用水平、垂直及对角翻转的图片增强。本方法采用SGD优化器优化整个网络，动量为0.9，权重衰减为0.0001。将原始图像裁剪成1024像素×1024像素，裁剪的步长设置为824，即两个相邻部分之间的像素重叠为200。对于多尺度训练和测试，首先在0.5、1.0和1.53个尺度上调整原始图像的大小。训练12个epoch，初始学习率设置为0.005，并在epoch为8和11时降为原有的十分之一。

3.3 实验结果及分析

本文将提出的群目标检测框架和其他12种主流的旋转物体检测框架进行了对比，结果如表1所示。

表1 DOTA数据集检测精度对比

Table 1 Comparison of detection accuracy on DOTA dataset

方法	骨干	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
DRN	H-104	88.9	80.2	43.5	63.4	73.5	70.7	84.9	90.1	83.9	84.1	50.1	58.4	67.6	68.6	52.5	70.7
R3Det	R-101	88.8	83.1	50.9	67.3	76.2	80.3	86.7	90.7	84.6	83.2	61.9	61.3	66.9	70.6	53.9	73.7
RSDet	R-101	89.8	82.9	48.6	65.2	69.5	70.1	70.2	90.5	85.6	83.4	62.5	63.9	65.6	67.2	68.0	72.2
DAL	R-50	88.6	76.5	45.0	66.8	67.0	76.7	79.7	90.8	79.5	78.4	57.7	62.2	69.0	73.1	60.1	71.4
S ² ANet	R-50	89.1	82.8	48.3	71.1	78.1	78.3	87.2	90.8	84.9	85.6	60.3	62.6	65.2	69.1	57.9	74.1
SCRDet	R-50	89.9	80.6	52.0	68.3	68.3	60.3	72.4	90.8	87.9	86.8	65.0	66.6	66.2	68.2	65.2	72.6
Faster-RCNN	R-101	88.4	73.6	44.8	59.0	73.2	71.4	77.1	90.8	78.9	83.9	48.5	62.9	62.1	64.9	56.1	69.0
CAD-Net	R-50	87.8	82.4	49.4	73.5	71.1	63.5	76.6	90.9	79.2	73.3	48.4	60.9	62.0	67.0	62.2	69.9
FAOD	R-101	90.2	79.5	45.4	76.4	73.1	68.2	79.5	90.8	83.4	84.6	53.4	65.4	74.1	69.6	64.8	73.2
CenterMap	R-50	88.8	81.2	53.1	60.6	78.6	66.5	78.1	88.8	77.8	83.6	49.3	66.1	72.1	72.3	58.7	71.7
RetinaNet	R-101	88.6	77.6	41.8	58.1	74.5	71.6	79.1	90.2	82.1	74.3	54.7	60.6	62.5	69.6	60.6	68.4
Mask OBB	R-50	89.6	85.0	51.8	72.9	75.2	73.2	85.5	90.3	82.0	85.0	55.7	68.3	71.6	69.8	66.3	74.8
Dual-RCNN	R-50	89.9	85.4	54.8	78.6	80.3	84.7	88.3	90.9	88.3	85.3	66.6	65.7	75.7	72.6	65.7	78.2

注：加粗数据的显示代表该列的最大值。

可见在DOTA-v1.0数据集上，该方法超越了所有参与对比的其他方法，获得了78.2%的mAP。本文算法Dual-RCNN的可视化结果如图7所示。可见该方法在困难的情境下发现了原来未检测到的目标，提高了召回率。图8为该算法在有无群中心约束情况下的检测结果对比，可见在小目标密

集的情况下，该方法检测到了更多小目标，更好地关注到了小目标周围区域的其他小目标，加强了目标与目标之间的位置信息关联。同时，本文对比了一些算法在群目标检测上的精度，结果见图9。可见在群目标和单目标的相互优化下，本算法在群目标检测上表现地更加出色。本文通过对

15 种类目标尺度分析, 统计了所有标注种类的平均长宽, 得到像素统计结果如下: 飞机 107、舰船 33、储罐 43、棒球场 135、网球场 104、篮球场 117、田径场 231、海港 103、桥梁 41、大型车辆 39、小型车辆 24、直升机 65、环岛 145、足球场 259、游泳池 45。本文认定长宽小于 50 的类别为小目

标, 因此在统计后将桥梁, 小车, 大车, 轮船, 存储罐, 泳池归类为小目标, 其余为大目标。由此得到的大小目标检测的 mAP 如表 2 所示, 结果表明, 本文提出的群目标检测在面对小目标时有着更加突出的表现, 通过群目标检测很好的提高了小目标检测准确率和召回率。



图7 Dual-RCNN 在DOTA数据集的可视化检测结果(■群目标 □单目标)

Fig.7 Visualization of Dual-RCNN on DOTA dataset. Group prediction with blue rectangle, single prediction with other color

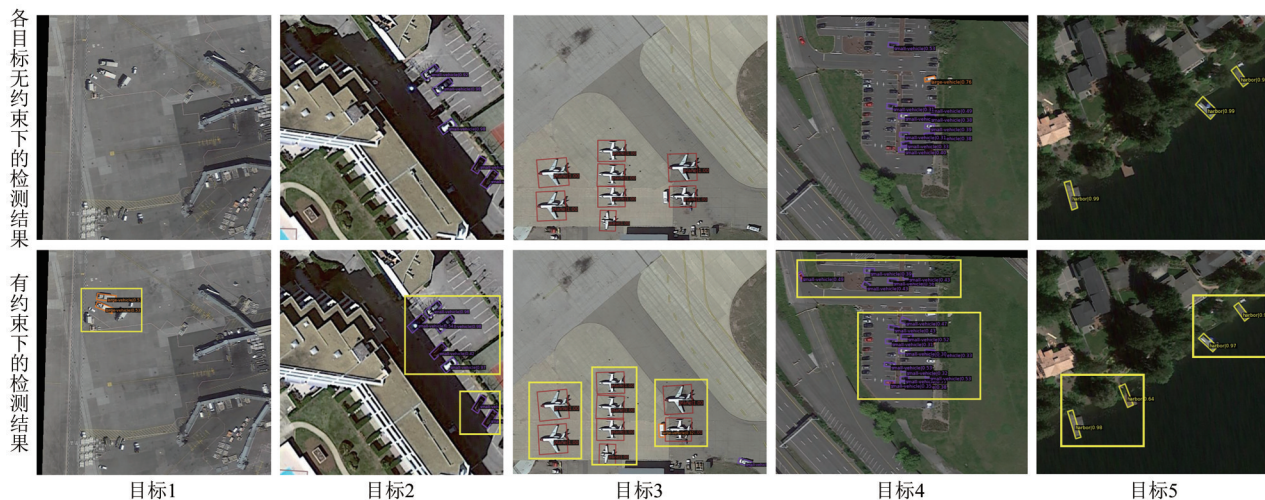


图8 群约束对检测结果的影响对比

Fig. 8 Visualization comparison of constraint

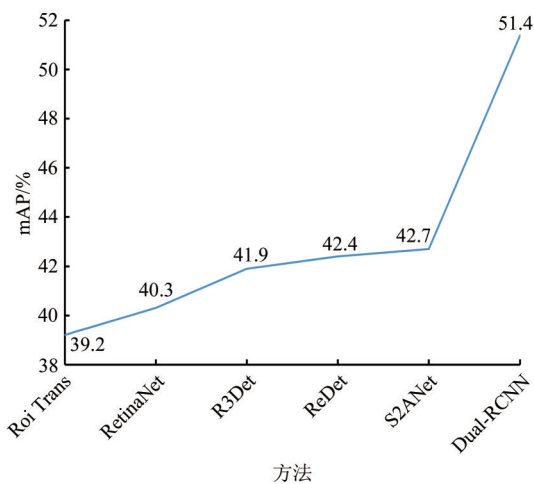


图9 不同方法的群目标检测精度对比

Fig. 9 Comparison of object detection results of different methods on group object detection

表2 不同尺度目标的检测精度对比

Table 2 Comparison of object detection results of different scales

方法	小目标 mAP	大目标 mAP	/%
DRN	70.88	70.57	
R3Det	74.65	73.06	
RSDet	68.16	74.88	
DAL	69.98	72.34	
SANet	74.43	73.87	
SCRDet	68.00	75.66	
ICN	65.41	69.88	
CAD-Net	66.81	71.95	
FAOD	70.06	75.28	
CenterMap	72.03	71.47	
FR-Est	73.06	74.95	
Mask OBB	73.41	75.72	
Dual-RCNN (本文)	77.67	78.56	

3.4 消融实验

3.4.1 群中心约束分析

验证群中心约束对检测结果的影响如表3所示。可见在有群中心约束监督的情况下，检测精度提升了1.7个百分点。这表明，群中心约束在空间范围上对单目标进行约束，提高了目标检测精度。

表3 群中心约束对mAP的影响

Table 3 Effect of group center constraint on mAP

群中心约束	mAP
有	76.51
无	78.21

3.4.2 标注方式分析

本文对提出的两种标注方法进行了实验比较。针对聚类式标注，group_num的取值首先与分类的最大集群数量有关，基于此我们在考虑分类的最大集群数量时进行了实验验证，固定其他参数不变，调整最大分类数量，得到实验结果如图10所示，由此可见，选择最大分类数量为5是最合适的。然后，统计得出在DOTA数据集中，平均每张图片包含67个目标，为了保证每张图片群的个数不唯一，第一个阈值nums1不宜过大，限定其在平均数量的1/4范围内，nums1可选值为5、10和15。阈值nums2不超过平均数量的1/2，nums2可选值为20、25、30和35。我们对这些取值进行组合实验，并得出了如表4的实验结果。因此，我们选择了nums1=15，nums2=25。最后实验结果如表5所示，其中，L代表图片对角线长，r代表搜索距离。聚类式标注方法在各种情况下的最好结果是77.42%，搜索式标注方法最好结果是78.21%，显著优于聚类式标注。搜索式具有更强的规律特点，仅取决于搜索距离，效果更好。

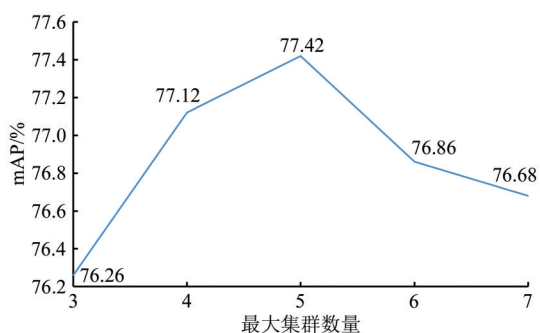


图10 群目标最大集群数量与Map相关性

Fig. 10 Correlation between group object quantity and Map

表4 不同nums参数的对比

Table 4 Results of different nums

nums1	nums2	mAP/%	nums1	nums2	mAP/%
5	20	77.34	5	25	77.37
5	30	76.64	5	35	76.95
10	20	77.19	10	25	77.39
10	30	76.56	10	35	77.14
15	20	76.87	15	25	77.42
15	30	76.97	15	35	76.90

表5 不同标注方式及距离参数对结果的影响

Table 5 Effect of different annotation methods and distance parameters on results

标注方式	dis1	dis2	r/像素	mAP/%
聚类式	L/4	L/2	—	76.89
	L/4	3L/4	—	76.97
	L/3	2L/3	—	77.42
	L/3	3L/4	—	77.35
搜索式	—	—	50	77.11
	—	—	75	78.21
	—	—	100	77.98
	—	—	125	78.03
	—	—	150	77.30

注：黑体表示最优结果。

3.4.3 群目标标注搜索距离分析

对于搜索式标注方法，确定搜索距离是一个关键的选择，本文对不同的搜索距离进行了实验对比，结果如表5所示。可见搜索距离偏大和偏小都会阻碍群目标检测性能，因为当搜索距离偏小时，群目标框将会趋近于单目标框，搜索距离偏大时，一个标注将可能涵盖过多单目标，使模型的针对性下降。针对DOTA数据集图片尺度为1024像素×1024像素，实验确定搜索距离为75是最有效的。

4 结论

本研究提出了一个面向遥感图像感知的群目标检测框架，旨在解决当前遥感图像感知中目标特征信息不足、准确率差的问题，针对小目标集群检测效果出色。主要结论如下：(1) 提出群目标检测理念，并对群目标概念进行定义；(2) 基于群目标定义，提出群目标自动化标注方法；(3) 提出群目标检测算法Dual-RCNN，利用中心约束引导检测框回归。实验结果表明，相比于其他遥感目标检测算法，Dual-RCNN能够召回更多小目标，

有效提升小目标检测精度。本文的研究使目标检测不再局限于单个目标, 将集群信息利用起来, 以缓解遥感目标分辨率低, 准确率差的问题。然而, 不同的定义方法及标注方式还是会对群目标检测产生一定的影响, 在后续的研究中, 将会对这些变量作进一步探索, 以找到最佳方案, 进一步地提高群目标检测能力。

参考文献(References)

- Azimi S M, Vig E, Bahmanyar R, Körner M and Reinartz P. 2019. Towards multi-class object detection in unconstrained remote sensing imagery//Proceedings of the 14th Asian Conference on Computer Vision. Perth: Springer: 150-165 [DOI: 10.1007/978-3-030-20893-6_10]
- Chen X L, Zhao H M, Li P X and Yin Z Y. 2006. Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes. *Remote Sensing of Environment*, 104(2): 133-146 [DOI: 10.1016/j.rse.2005.11.016]
- Ding J, Xue N, Long Y, Xia G S and Lu Q K. 2019. Learning RoI transformer for oriented object detection in aerial images//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 2844-2853 [DOI: 10.1109/CVPR.2019.00296]
- Ding J, Xue N, Xia G S, Bai X, Yang W, Yang M Y, Belongie S, Luo J B, Datcu M, Pelillo M and Zhang L P. 2022. Object detection in aerial images: a large-scale benchmark and challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11): 7778-7796 [DOI: 10.1109/TPAMI.2021.3117983]
- Girshick R. 2015. Fast R-CNN//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE: 1440-1448 [DOI: 10.1109/ICCV.2015.169]
- Lenhart D, Hinz S, Leitloff J and Stilla U. 2008. Automatic traffic monitoring based on aerial image sequences. *Pattern Recognition and Image Analysis*, 18(3): 400-405 [DOI: 10.1134/S1054661808030061]
- Li Q, Geng D, Zhang J F and Gong L X. 2022. Application and trend analysis of remote sensing technology for earthquake emergency investigation. *National Remote Sensing Bulletin*, 26(10): 1920-1934 (李强, 耿丹, 张景发, 龚丽霞. 2022. 面向地震应急调查的遥感应应用现状及趋势分析. *遥感学报*, 26(10): 1920-1934 [DOI: 10.11834/jrs.20210078])
- Lin T Y, Goyal P, Girshick R, He K M and Dollár P. 2017. Focal loss for dense object detection//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE: 2999-3007 [DOI: 10.1109/ICCV.2017.324]
- Liu Y and Wu L Z. 2016. Geological disaster recognition on optical remote sensing images using deep learning. *Procedia Computer Science*, 91: 566-575 [DOI: 10.1016/j.procs.2016.07.144]
- Lopez R and Frohn R. 2017. *Remote Sensing for Landscape Ecology: Monitoring, Modeling, and Assessment of Ecosystems*. 2nd ed. Boca Raton: CRC Press [DOI: 10.1201/9781315152714]
- Ming Q, Zhou Z Q, Miao L J, Zhang H W and Li L H. 2021. Dynamic anchor learning for arbitrary-oriented object detection//Proceedings of the 35th AAAI Conference on Artificial Intelligence. [s.l.]: AAAI: 2355-2363 [DOI: 10.1609/aaai.v35i3.16336]
- Pan X J, Ren Y Q, Sheng K K, Dong W M, Yuan H L, Guo X W, Ma C Y and Xu C S. 2020. Dynamic refinement network for oriented and densely packed object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE: 11204-11213 [DOI: 10.1109/CVPR42600.2020.01122]
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Tian Z Z, Zhang H W, Wang K, Liu S Q, Zou Q J, Zhao Z and Chen Y B. 2023. Application of an improved CenterNet in remote sensing images object detection. *National Remote Sensing Bulletin*, 27(12): 2706-2715 (田壮壮, 张恒伟, 王坤, 刘盛启, 邹前进, 赵镇, 陈育斌. 2023. 改进 CenterNet 在遥感图像目标检测中的应用. *遥感学报*, 27(12): 2706-2715 [DOI: 10.11834/jrs.20231638])
- Wang J W, Yang W, Li H C, Zhang H J and Xia G S. 2021. Learning center probability map for detecting objects in aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(5): 4307-4323 [DOI: 10.1109/TGRS.2020.3010051]
- Xia G S, Bai X, Ding J, Zhu Z, Belongie S, Luo J B, Datcu M, Pelillo M and Zhang L P. 2018. DOTA: a large-scale dataset for object detection in aerial images//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 3974-3983 [DOI: 10.1109/CVPR.2018.00418]
- Yang X, Yan J C, Feng Z M and He T. 2021. R3Det: refined single-stage detector with feature refinement for rotating object//Proceedings of the AAAI Conference on Artificial Intelligence. [s.l.]: AAAI: 3163-3171 [DOI: 10.1609/aaai.v35i4.16426]
- Yi R. 2020. The research on digital media recommendation method based on K -means clustering algorithm. *Journal of Changchun Institute of Technology (Natural Science Edition)*, 21(4): 99-102 (易茹. 2020. 基于 K 均值聚类算法的数字媒体推荐方法研究. *长春工程学院学报(自然科学版)*, 21(4): 99-102 [DOI: 10.3969/j.issn.1009-8984.2020.04.020])
- Zhang G J, Lu S J and Zhang W. 2019. CAD-Net: a context-aware detection network for objects in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12): 10015-10024 [DOI: 10.1109/TGRS.2019.2930982]

A group object detection framework for remote sensing image object perception

ZHANG Hongwei¹, JIN Lei¹, ZOU Xuechao², FANG Yuqiang³, YIN Lu⁴, ZHAO Jian⁵, XING Junliang⁶

1. School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China;

2. School of Computer Technology and Applications, Qinghai University, Qinghai 810016, China;

3. Graduate School, University of Aerospace Engineering, Beijing 101416, China;

4. Beijing Institute of Remote Sensing Information, Beijing 100192, China;

5. North Institute of Electronic Equipment, Beijing 100071, China;

6. School of Computing, Tsinghua University, Beijing 100084, China

Abstract: Optical remote sensing is a widely used technology in aerospace reconnaissance and geological exploration. Visible light images captured by this technology provide a wealth of information and have important applications in intelligence gathering, object monitoring, and situational forecasting. Considerable progress in remote sensing image object perception has been achieved, particularly in ship and airplane detection. However, technical challenges, including with large object-scale variations and numerous small objects, in remote sensing image object perception remain. Existing work has mainly focused on improving boundary box representations, and single-object detection models fail to fully exploit spatial correlation information from surrounding or similar objects. To address the inherent inefficiency of existing remote sensing image object detection algorithms that detect different objects independently, this paper proposes a novel detection framework called group object detection. By detecting the state information of a group object, our framework alleviates problems, such as insufficient perception information and poor reliability of single-object perception, generating reliable multi-object detection results. This paper introduces a concept of group objects and proposes an automated annotation scheme for group objects. By analyzing existing labels on a public dataset, the proposed scheme obtains annotated information with group object labels without manual annotation. Based on the automated annotation of group targets, a group target detection algorithm is presented, which enhances single-object detection results by utilizing the spatial constraints of group objects. Experimental results on the DOTA dataset, a widely-used remote sensing object detection benchmark, demonstrate that the proposed group target detection algorithm outperforms state-of-the-art methods.

Key words: remote sensing image, image object, group object detection, automatic annotation, DOTA, object sensing, multi object

Supported by Foundation of the Key Laboratory of Satellite Information Intelligent Processing and Application Research (No. 2022-ZZKY-JJ-06-01)