视图一致性网络下的弱监督遥感影像旋转目标检测

方婷婷[1,2]，刘斌[1,2]，陈春晖[3,4]，厉香蕴[3,4]

1. 中国科学技术大学 网络空间安全学院, 合肥 230026;
2. 中国科学院电磁空间信息重点实验室, 合肥 230026;
3. 安徽省基础测绘信息中心, 合肥 230031;
4. 自然资源部江淮耕地资源保护与生态修复重点实验室, 合肥 230031

摘 要：为了减轻旋转目标检测的标注负担、提高遥感旋转目标检测数据集的可用性和丰富性，本文提出了一种仅使用图像级别标注的旋转目标检测模型。该模型的核心在于利用3个一致性约束：图像级别标注的一致性、旋转检测框位置的一致性和聚类中心分布的一致性，来设计目标检测网络。首先，本文引入了遥感场景下的弱监督旋转目标检测范式，通过逐阶段优化预测结果，利用图像级别标注的一致性来训练模型。其次，为了提高旋转角度的预测准确性，模型通过利用不同旋转视图下的旋转检测框的空间位置一致性约束来调整角度预测。最后，为了保持不同旋转视图下聚类中心节点的分布一致性，设计了匈牙利损失来度量聚类中心的分布一致性。通过这些一致性约束，模型可以更好地利用不同旋转视图下的旋转等变性，提高检测性能。为了验证该模型的有效性，本文在两个主流的遥感目标检测数据集DIOR和DOTA-v1.0上进行了实验。实验结果显示，模型在DIOR数据集上取得37.7% mAP，相较于水平弱监督最先进模型提升了9.4% mAP；在DOTA数据集上取得28.1% mAP，在多个检测类别上优于难度更小的、基于水平框标注的旋转弱监督的最先进模型。这表明该模型在解决标注困难和提高遥感目标检测性能方面具有良好的潜力。未来，研究者可以进一步拓展这种基于图像级别标注的旋转目标检测模型，结合更先进的技术和不同数据源，探索其在遥感、地理空间分析和其他相关领域的更广泛应用。

关键词：遥感，旋转目标检测，目标检测，弱监督，深度学习，旋转一致性，图像级标注，匈牙利损失，DOTA数据集

中图分类号：TP753/P2

引用格式：**方婷婷,刘斌,陈春晖,厉香蕴.2024.视图一致性网络下的弱监督遥感影像旋转目标检测.遥感学报,28(12):3213-3230**

Fang T T, Liu B, Chen C H and Li X Y. 2024. View-consistency network for weakly supervised oriented object detection in remote sensing images. National Remote Sensing Bulletin, 28(12): 3213-3230[DOI: 10.11834/jrs.20243138]

1 引 言

近年来，遥感影像的目标检测加快了土地动态监测和卫片执法的数字化进程。由于遥感影像拍摄的俯瞰视角，待检测目标不止局限于垂直或水平位置，而是可能以各种角度排列。此时传统目标检测下用于定位物体的水平检测框已无法准确描述目标位置（张磊 等，2022），针对遥感影像的旋转目标检测便成为视觉领域一项重要的研究任务。

随着卷积神经网络的提出，很多基于深度学习的旋转目标检测工作涌现（Han等，2021；Yang等，2021a；Xie等，2021；Li等，2022a）。无一例外，这些模型的训练都基于大量的标注数据（袁一钦 等，2023），以该领域主流的两个基准数据集DOTA-v1.0（用于航空图像中目标检测的大规模数据集）（Xia等，2018）和DIOR（用于光学遥感图像目标检测的大规模基准数据集）（Cheng等，2022）为例，前者包含2806张图像及188282个实例，后者包含23463张图像及192472个实例，实例标注形式为包围待检测目标旋转框的四角点八参数坐标。大规模的实例数量和复杂的标签格式耗费大量人力物力，延长开源标注数据集的制作周期，不利于深度学习模型的发展和性能提升。

收稿日期：2023-04-26；**预印本：**2023-12-08

基金项目：安徽省自然资源科技项目(编号:2021-K-14)

第一作者简介：方婷婷,研究方向为遥感目标检测。E-mail:fountain@mail.ustc.edu.cn

通信作者简介：刘斌,研究方向为多媒体信息处理和人工智能安全。E-mail:flowice@ustc.edu.cn

为了降低目标检测的标注成本，很多工作进行了弱监督设置下的探索。

首先对全监督和弱监督目标检测WSOD（Weakly Supervised Deep Object Detection）任务进行比较说明。传统的全监督目标检测需要标注每个目标的边界框和类别标签来训练目标检测模型；而弱监督目标检测则采用更简单的标注信息，最典型的是利用图像级别标签（即仅标注整个图像中是否存在某类目标）完成训练。本文也正是针对基于图像级别标签的弱监督设置进行研究。

现有的图像级别标签的弱监督目标检测方法大多基于弱监督深度卷积网络WSDDN（Weakly Supervised Deep Detection Network），将任务建模为多示例学习问题MIL（Multiple Instance Learning）（Bilen和Vedaldi，2016）。其中，OICR（Online Instance Classifier Refinement）模型首次为弱监督目标检测任务设计多阶段细化分支（Tang等，2017）；PCL（Proposal Cluster Learning）引入候选框聚类算法旨在解决图像级标注下模型只学习物体显著部分的问题（Tang等，2020）；最新的一些模型则通过融入上下文信息或添加额外的分割网络提升检测性能（Wei等，2018；Yan等，2019；Li等，2019；Huang等，2020）。还有些工作致力于遥感场景下的WSOD任务，如使用动态课程学习渐进式生成高质量初始目标样本（Yao等，2021）、针对遥感影像中同类别实例数量多且易漏检问题分别提出了多实例图学习机制和旋转不变性检测网络等（Wang等，2022；Feng等，2020a）。但无论是普通场景还是遥感场景，上述模型均为弱监督下的水平目标检测，没有使用图像级别标注完成旋转目标检测的先例。

考虑到仅使用图像级别标注的任务难度，有些工作放宽了弱监督的限制，以水平检测框为标注预测旋转检测框：如有工作提出了使用水平检测框标注进行旋转影像分割的模型，对上述分割掩码寻找最小边界矩形即可预测旋转检测框（Tian等，2021；Li等，2022b）；或通过设计弱监督和自监督模块学习两个视图的一致性，从而实现物体角度的预测（Yang等，2022）。尽管缓解了部分标注负担，但水平框的标注依然繁琐。上述工作均未实现真正意义上基于图像级别标注的弱监督旋转目标检测。

本文希望提出一种新颖的旋转目标检测模型，只利用少量标注信息就可以有效地实现遥感影像中旋转目标的检测。传统全监督方法依赖于完整的旋转框注释，这些注释由人工标注，制作周期长且具有主观性，在不同的场景下可能存在标注标准不一致的问题。若只利用图像级别标注，所提出的模型有助于减少注释工作量，加速注释过程，减轻不同情形下的标注混淆，可增强遥感目标检测在各种场景和领域中的可扩展性和适用性。以此为出发点，本文旨在研究仅使用图像级别标注的弱监督旋转目标检测方法。针对图像级别标注中缺乏旋转角度信息的问题，以不同旋转视图下图像级别标签、旋转检测框位置和聚类中心分布的3个一致性约束为核心设计神经网络结构，首次提出了遥感场景下基于图像级别标注的弱监督旋转目标检测模型。首先，根据弱监督任务中典型的水平目标检测模型设计旋转目标检测范式，通过发掘图像级别的隐式约束对细化分支的预测结果进行渐进式优化调整；其次，基于不同旋转视图下旋转检测框形状及角度的一致性，实现对预测角度的约束；最后，由于不同旋转视图间旋转检测框的各聚类中心节点分布类似，引入匈牙利匹配损失衡量其一致性并指导模型的优化方向。模型在DIOR和DOTA-v1.0数据集上进行了实验。结果显示：在DIOR数据集的相同实例上，模型性能相比于弱监督水平目标检测最先进SOTA（State Of The Art）方法的性能有显著提升，部分类别指标优于基于水平检测框标注的旋转目标检测方法和全监督目标检测方法；在DOTA-v1.0数据集上模型部分类别指标优于基于水平检测框标注的旋转目标检测方法。

2 弱监督旋转目标检测模型

2.1 模型概述

首先对弱监督设置进行定义：假设待检测目标有C个类别，已知图像级别标注$\boldsymbol{y}=[y_1, \cdots, y_c, \cdots, y_C | y_i \in \{0, 1\}]$，其中$y_c$指示当前图像是否存在类别为$c$的目标。模型仅利用图像级别标注预测图像中包围待检测物体的旋转检测框位置及类别概率。

图1展示了用于弱监督旋转目标检测的视图一致性网络框架。该模型包含一个基本的旋转目标检测器和3个视图一致性增强模块。旋转目标检测

器基于弱监督水平目标检测中的经典网络PCL（Tang等，2020），为适配旋转检测框，在其上添加旋转感兴趣区域RRoI（Rotated Region of Interest）学习器以生成角度预测，实现基本的检测功能。视图一致性增强模块则针对图像级别标签缺乏旋转角度信息的问题，将输入图像拓展到多个视图，将不同视图之间图像级别标签、旋转检测框位置和聚类中心分布的3个一致性作为约束条件指导模型优化方向，提高检测性能。下面介绍具体的训练和推理流程。

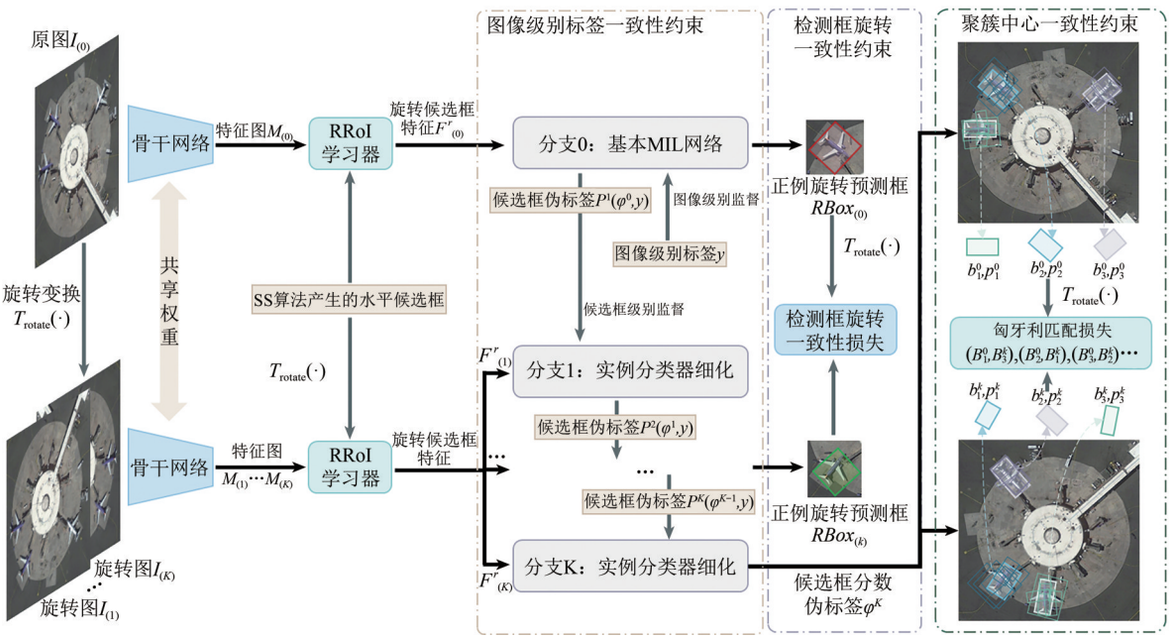


图1 用于弱监督旋转目标检测的视图一致性网络框架

Fig. 1 Overview of View-Consistency Network for weakly supervised oriented object detection

和大多数弱监督目标检测模型类似，首先使用传统目标检测的候选区算法（Proposal Method）为图像生成水平的候选区域（Proposal Region）。然后对图像进行随机的旋转变换将原图拓展至多个旋转视图，同时对水平候选区域也进行相同的旋转变换。将所有视图输入至相同权重的骨干网络获得多个特征图，特征图和其对应变换后的候选区域框共同作为RRoI学习器模块的输入，从而获得带有旋转角度的候选框特征向量。

由于旋转变换不会改变目标类别，不同视图应具有相同的图像级标签和候选框级类别概率。因此不同视图下旋转候选框特征向量产生的类别概率在PCL网络的在线实例分类细化结构中可以互为监督，即在统一的图像级标签下，任一视图产生的候选框级类别伪标签可作为另一视图类别分数预测的监督信号。该模块仅约束预测类别，并未显式地约束旋转检测框的预测角度，为此引入检测框旋转一致性和聚类中心一致性约束模块。

由上文知，基本的旋转目标检测器为每个候选框预测了类别概率，根据此概率可设定阈值划分正例和负例检测框。取出各视图对应的正例旋转框，由对应检测框旋转变换后形状及位置的一致性设计损失函数。同时，以细化分支最后一阶段得到的类别概率为基准，根据概率对候选框进行聚类和图构建。由于不同视图下的候选框聚类中心节点分布应当一致，设计匈牙利匹配损失衡量不同视图下聚类中心分布的一致性。这两个模块的设计强迫RRoI学习器在不同旋转视角下给出具有旋转等变性（Rotation Equivalence）的角度预测，从而捕捉到目标的旋转特征。

推理过程只使用基本旋转检测器，并将输入预处理中的旋转变换替换为恒等变换。最终输出为RRoI学习器生成的旋转候选框及其对应的所有细化分支预测的实例类别得分的平均值。

2.2 基本的旋转目标检测

2.2.1 候选区域生成

本模型基于R-CNN系列算法，需要先生成候

选框，再对候选区域进行分类和回归（Girshick等，2015；Ren等，2015）。由于缺乏检测框级别标注，模型无法直接通过锚定框来生成候选框，因此使用选择性搜索SS（Selective Search）（Uijlings等，2013）算法生成可能的候选区域。该方法首先对图像进行区域分割，再根据区域间的颜色、纹理、尺度和填充相似度进行区域合并，快速生成可能包含目标的水平感兴趣区域HRoI（Horizontal Region of Interest）。

2.2.2 RRoI学习器

RRoI学习器结构如图2所示。输入图像$\boldsymbol{I}$，骨干网络提取其特征图$\boldsymbol{M}\in \mathrm{R}^{h\times w\times l}$，其中$h$、$w$、$l$分别为特征图长度、宽度及通道维度。SS算法为图像生成HRoI集合$\{\boldsymbol{H}_1,\cdots,\boldsymbol{H}_N\}$，其中$N$为HRoI的个数，$\boldsymbol{H}_n$可表示为$(x,y,w,h)$，即中心点二维坐标、宽度及长度形式。使用RoI Align（He等，2017）在$\boldsymbol{M}$上提取$\boldsymbol{H}_n$对应位置的候选区域特征向量$\boldsymbol{F}_n\in \mathrm{R}^{d_1\times d_1\times l}$，其中$d_1\times d_1$为RoI Align划分子区域（bins）的个数。

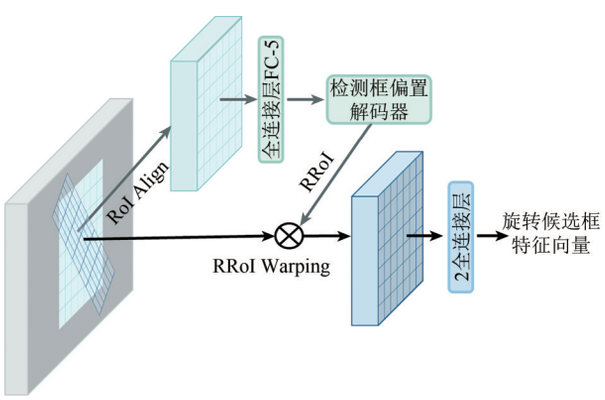


图2 RRoI学习器

Fig. 2 RRoI learner

全连接层$\boldsymbol{\Psi}_{FC-5}$以$\boldsymbol{F}_n$为输入，预测对应RRoI的偏移量和旋转角度。形式化地，$\boldsymbol{\Psi}_{FC-5}$对于特征$\boldsymbol{F}_n$的输出向量$(t_x,t_y,t_w,t_h,t_\theta)$具有如下回归目标：

$$\begin{cases} t_x=\dfrac{1}{w_r}\left((x-x_r)\cos\theta_r+(y-y_r)\sin\theta_r\right) \\ t_y=\dfrac{1}{h_r}\left((y-y_r)\cos\theta_r+(x-x_r)\sin\theta_r\right) \\ t_w=\log\dfrac{w}{w_r},t_h=\log\dfrac{h}{h_r} \\ t_\theta=\dfrac{1}{2\pi}\left(\theta_r \bmod 2\pi\right) \end{cases} \tag{1}$$

式中，$(x_r,y_r,w_r,h_r,\theta_r)$表示基于$\boldsymbol{H}_n$生成的旋转候选框$\boldsymbol{H}_n^{\text{rotate}}$中心点二维坐标、宽度、长度及旋转角度。检测框偏置解码器以$\boldsymbol{H}_n=(x,y,w,h)$和全连接层输出向量$(t_x,t_y,t_w,t_h,t_\theta)$为输入，根据式（1）解码出$\boldsymbol{H}_n^{\text{rotate}}$的5个参数，获得RRoI集合$\{\boldsymbol{H}_1^{\text{rotate}},\cdots,\boldsymbol{H}_n^{\text{rotate}}\}$。

使用RRoI Warping（Ding等，2019）在$\boldsymbol{M}$上提取$\boldsymbol{H}_n^{\text{rotate}}$对应位置的候选区域特征向量$\boldsymbol{F}_n^{\text{rotate}}\in \mathrm{R}^{d_2\times d_2\times l}$，与RoI Align类似，其中$d_2\times d_2$为RRoI Warping划分子区域的个数。$\boldsymbol{F}_n^{\text{rotate}}$经过两个线性-ReLu层进一步处理得$\boldsymbol{F}_n{}^r$，被映射至$d_3$维度。生成最终的旋转候选框特征向量$\boldsymbol{F}^r\in \mathrm{R}^{d_3\times N}$，作为下一模块的输入。

2.2.3 基于PCL的在线实例细化

本模块遵循经典弱监督水平目标检测的网络结构，通过构造多级实例分类器细化分支，不断优化实例级别的标签准确性，旨在解决弱监督目标检测器只捕捉物体显著部分的问题。本节只讨论单视图下的PCL在线实例细化过程，多视图设置将在2.3节详细说明。

首先将RRoI学习器提取的旋转候选框特征$\boldsymbol{F}^r\in \mathrm{R}^{d_3\times N}$输入基本的MIL网络，其结构如图3所示。$\boldsymbol{F}^r$作为两个数据流各经过全连接层$\boldsymbol{\Psi}_{\text{cls}}$和$\boldsymbol{\Psi}_{\text{det}}$分别获得两个矩阵$\boldsymbol{X}^c\in \mathrm{R}^{C\times N}$和$\boldsymbol{X}^d\in \mathrm{R}^{C\times N}$。第一个数据流中$\boldsymbol{X}^c$表示每个区域的类别预测得分，对$\boldsymbol{X}^c$进行类别维度的Softmax操作得到$[\sigma_{\text{cls}}(\boldsymbol{X}^c)]_{ij}=e^{x_{ij}^c}/\sum_{k=1}^{C}e^{x_{kj}^c}$，该Softmax用于比较每个候选区域的类别得分，预测了哪一个类别和当前区域相关联。第二个数据流对$\boldsymbol{X}^d$进行候选框维度的Softmax操作得到$[\sigma_{\text{det}}(\boldsymbol{X}^d)]_{ij}=e^{x_{ij}^d}/\sum_{k=1}^{N}e^{x_{ik}^d}$，该Softmax用于比较每个类别下不同候选区域的得分，预测了哪一个区域更有可能包含当前类别的特征信息。对上述两评价矩阵计算Hadamard乘积得到每个候选框的最终得分$\boldsymbol{\varphi}^0\in \mathrm{R}^{C\times N}$，在候选框维度求和即可得到图像级别预测分数$\boldsymbol{\varphi}(c)=\sum_{n=1}^{N}\varphi_{cn}^0$。最后根据图像级别标签$\boldsymbol{y}=[y_1,\cdots,y_C]$计算多分类交叉熵损失，进行图像级别的分类约束：

$$L_{\text{MIL}}=-\sum_{c=1}^{C}y_c\log\boldsymbol{\varphi}(c)+(1-y_c)\log(1-\log\boldsymbol{\varphi}(c)) \tag{2}$$

然而基本的MIL网络只能捕捉物体最具有鉴

别性的部分，PCL引入在线实例细化机制，进行多阶段的实例分类器细化逐步发现完整物体。如图4所示，第k阶段的实例分类器细化以$k-1$阶段候选区域得分产生的实例级别伪标签$P^k(\boldsymbol{\varphi}^{k-1}, \boldsymbol{y}) = \{\boldsymbol{y}_1^k, \cdots, \boldsymbol{y}_n^k, \cdots, \boldsymbol{y}_N^k\}$作为监督信号，生成当前阶段的候选区域得分$\boldsymbol{\varphi}^k$。其中，$P(\cdot)$为实例级别的伪标签生成函数，通过对候选框类别分数进行聚类筛选出预测概率较高的检测框及与其重叠面积较大的相邻区域，共同赋予正例标签作为下一阶段的实例级监督信号；$\boldsymbol{y}_n^k = [y_{1n}^k, \cdots, y_{(C+1)n}^k]^{\mathrm{T}}$则是$k$阶段第$n$个候选框实例的类别标签。注意不同于基本MIL分支，细化分支的全连接层将候选框特征投射到包括背景类的$C+1$个类别维度。每个细化阶段均产生一个候选框级别的分类损失，对其进行加权求和便可用于各阶段的细化网络参数优化：

$$L_{\mathrm{PCL}} = \sum_{k=1}^{K} L^k(\boldsymbol{F}_{(k)}^r, W_k, P^k(\boldsymbol{\varphi}^{k-1}, \boldsymbol{y})) = \sum_{k=1}^{K}(-\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C+1}\lambda_n^k y_{cn}^k \log \varphi_{cn}^k) \tag{3}$$

式中，λ_n^k是k阶段第n个候选框实例所在的聚类中心置信度，作为自适应损失权重减轻早期细化阶段的噪声影响。

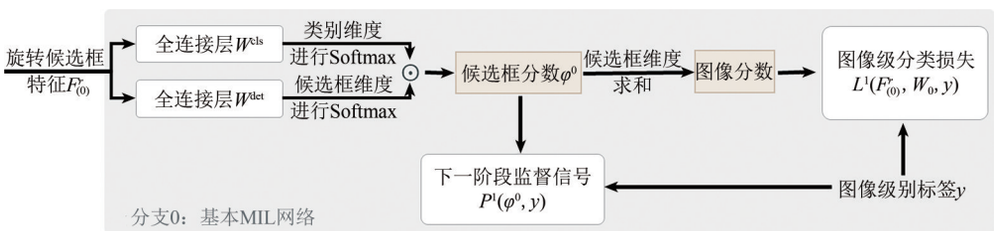


图3 基本MIL网络

Fig. 3 Basic Multiple Instance Learning network

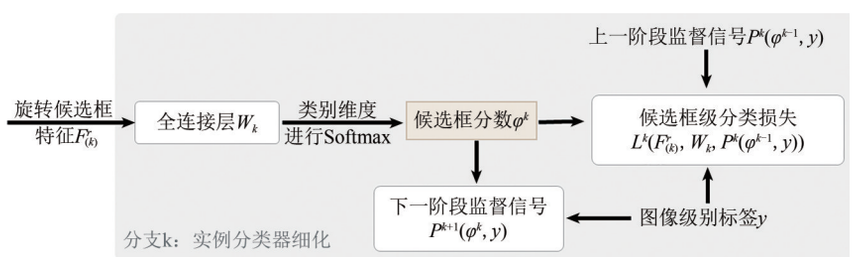


图4 实例分类器细化分支

Fig. 4 Basic Multiple Instance Learning network

2.3 视图一致性约束

图像级类别标签缺乏物体旋转角度信息，而基本的旋转目标检测器只利用图像级标注，这不利于遥感场景中丰富的旋转角度预测。因此将输入图像拓展至不同的旋转视图，挖掘不同视图下基于旋转等变的一致性，构造对角度预测的约束条件。

给定图像$\boldsymbol{I}_{(0)}$和其HRoI集合$\boldsymbol{H}_{(0)}$，对$\boldsymbol{I}_{(0)}$进行K次随机旋转变换得到旋转视图集合$\{\boldsymbol{I}_{(k)} = T_{\mathrm{rotate}}^{(k)}(\boldsymbol{I}_{(0)})\}_{k=1}^K$，对$\boldsymbol{H}_{(0)}$进行相同的旋转变换得到各视图对应的HRoI集合$\{\boldsymbol{H}_{(k)} = T_{\mathrm{rotate}}^{(k)}(\boldsymbol{H}_{(0)})\}_{k=1}^K$，其中旋转角$\Delta\theta \in \{0, \pi, \pi/2, 3\pi/2\}$。将原图和旋转视图输入相同的骨干网络得到特征图集合$\boldsymbol{M} = \{\boldsymbol{M}_{(k)}\}_{k=0}^K$，以$(\boldsymbol{M}_{(k)}, \boldsymbol{H}_{(k)})$为RRoI学习器输入，生成相应的旋转候选框特征$\boldsymbol{F}_{(k)}^r$，用于后续的3个一致性约束模块。

2.3.1 图像级别标签一致性约束

由于旋转变换不会改变物体类别，不同旋转视图关于类别标签存在以下两个一致性：（1）不同视图共享图像级别标签$\boldsymbol{y} = [y_1, \cdots, y_C]$，因此由共享的图像级别标签在MIL分支中为某一视图生成的实例级伪标签可以用来监督其他视图；（2）不同视图对应的第i个旋转候选框特征向量$\{\boldsymbol{F}_{i(k)}^r\}_{k=1}^K$生成的类别得分应当一致，因此由某一视图在实例分类器细化阶段生成的实例级别伪标签可以用来监督其他视图。

基于上述两个性质，将$K+1$个视图的候选框特征分配至PCL网络的不同分支以实现视图间的相互监督：作为初始阶段的基本MIL网络以原图生成的旋转特征$\boldsymbol{F}^r_{(0)}$为输入，在图像级分类标签$\boldsymbol{y}$的约束下生成实例级伪标签；而优化阶段的K个实例分类器细化分支，分别以K个变换视图的旋转特征$\{\boldsymbol{F}^r_{(k)}\}^K_{k=1}$为输入，依次生成伪标签作为下一阶段监督信号。

2.3.2 检测框旋转一致性约束

已知原图$\boldsymbol{I}_{(0)}$到某一旋转视图$\boldsymbol{I}_{(k)}$的变换矩阵为$\boldsymbol{R}_{\Delta\theta}$，则RRoI学习器为原图预测的旋转候选框$\mathbf{RBox}\boldsymbol{x}_{(0)}(x_{(0)},\ y_{(0)},\ w_{(0)},\ h_{(0)},\ \theta_{(0)})$可经过相同变换得到$\mathbf{RBox}\boldsymbol{x}^*_{(0)}(x^*_{(0)},\ y^*_{(0)},\ w^*_{(0)},\ h^*_{(0)},\ \theta^*_{(0)})$，转换关系如下：

$$\begin{cases}\boldsymbol{R}_{\Delta\theta}=\begin{pmatrix}\cos\Delta\theta & -\sin\Delta\theta\\ \sin\Delta\theta & \cos\Delta\theta\end{pmatrix}\\ (x^*_{(0)},y^*_{(0)})=(x_{(0)}-x_c,y_{(0)}-y_c)\boldsymbol{R}^{\mathrm{T}}_{\Delta\theta}+(x_c,y_c)\\ (w^*_{(0)},h^*_{(0)})=(w_{(0)},h_{(0)})\\ \theta^*_{(0)}=\theta_{(0)}+\Delta\theta\end{cases}\tag{4}$$

式中，$(x_c,\ y_c)$代表旋转中心，即输入图像的中心坐标。此时$\mathbf{RBox}\boldsymbol{x}^*_{(0)}$和第$k$个旋转视图对应的候选框$\mathbf{RBox}\boldsymbol{x}_{(k)}(x_{(k)},\ y_{(k)},\ w_{(k)},\ h_{(k)},\ \theta_{(k)})$具有空间位置和旋转角度一致性，据此分别构造中心点损失和角度损失。

为了减少噪声干扰，参与每个阶段损失计算的只包括预测概率较高的实例，即由PCL算法分配的正例候选框。给定正例候选框的索引集合$\boldsymbol{Ind}^+$，计算第k阶段的旋转一致性损失：

$$L^k_{\mathrm{RC}}=\frac{1}{|\boldsymbol{Ind}^+|}\sum_{i\in\boldsymbol{Ind}^+}\lambda^k_i L_{\mathrm{reg}}(\mathbf{RBox}\boldsymbol{x}^*_{i(k)},\mathbf{RBox}\boldsymbol{x}_{i(0)})\tag{5}$$

式中，λ^k_i表示k阶段的第i个正例候选框所在的聚类中心置信度。

中心点损失L_{center}和角度损失$L_{\mathrm{wh\theta}}$共同构成回归损失L_{reg}：

$$L_{\mathrm{reg}}=\alpha_1 L_{\mathrm{center}}+\alpha_2 L_{\mathrm{wh\theta}}\tag{6}$$

$$L_{\mathrm{center}}=L_1\left((x^*_{(0)},y^*_{(0)}),(x_{(k)},y_{(k)})\right)\tag{7}$$

$$L_{\mathrm{wh\theta}}=\min\begin{Bmatrix}L_{\mathrm{IoU}}(\boldsymbol{HB}^*_{(0)},\boldsymbol{HB}^1_k)+|\sin(\theta^*_{(0)}-\theta_{(k)})|,\\ L_{\mathrm{IoU}}(\boldsymbol{HB}^*_{(0)},\boldsymbol{HB}^2_k)+|\cos(\theta^*_{(0)}-\theta_{(k)})|\end{Bmatrix}\tag{8}$$

式中，L_1和L_{IoU}分别表示L1和交并比IoU（Intersection of Union）损失，α_1和α_2表示中心点和角度损失的权重。$L_{\mathrm{wh\theta}}$遵循H2RBox模型（Yang等，2022）的设计，考虑了由角度周期性和边的可交换性（Yang等，2021b）导致的损失不连续问题，满足：

$$\begin{aligned}\boldsymbol{HB}^*_{(0)}&=(-w^*_{(0)},-h^*_{(0)},w^*_{(0)},h^*_{(0)})\\ \boldsymbol{HB}^1_{(k)}&=(-w_{(k)},-h_{(k)},w_{(k)},h_{(k)})\\ \boldsymbol{HB}^2_{(k)}&=(-h_{(k)},-w_{(k)},h_{(k)},w_{(k)})\end{aligned}\tag{9}$$

K个细化阶段的输入分别对应K个旋转视图，将每一阶段产生的正例候选框和相应的原图候选框计算旋转一致性损失，最终损失为各阶段损失之和：

$$L_{\mathrm{RC}}=\sum_{k=1}^{K}L^k_{\mathrm{RC}}\tag{10}$$

2.3.3 聚类中心一致性约束

基于候选框聚类（Proposal Cluster）的PCL算法主要分为3个步骤：（1）根据候选框的类别得分$\{\boldsymbol{\varphi}^k_c=[\varphi^k_{c1},\ \cdots,\ \varphi^k_{cN}]\}^C_{c=1}$生成聚类中心，每个中心节点都代表一个检测目标；（2）根据重叠程度将剩余的每个候选框与一个聚类中心节点相关联或为其分配背景类标签。（3）根据上述分配结果生成实例级伪标签作为下一阶段监督信号。

首先对步骤（1）中聚类中心的生成算法进行回顾。假设图像中存在类别为c的目标，在第k次细化阶段中，使用K-Means算法基于候选框的c类别得分$\boldsymbol{\varphi}^k_c\in\mathrm{R}^N$进行聚类，动态挑选出置信度排名靠前的候选框集合$\boldsymbol{H}^{\mathrm{rotate}}_{c(k)}=\{h^k_{c_1},\ \cdots,\ h^k_{c_N}\}$，为$\boldsymbol{H}^{\mathrm{rotate}}_{c(k)}$建立一个无向无权图$\boldsymbol{G}^k_c=(\boldsymbol{V}^k_c,\ \boldsymbol{E}^k_c)$。顶点集合$\boldsymbol{V}^k_c$与$\boldsymbol{H}^{\mathrm{rotate}}_{c(k)}$中候选框元素对应，边集合则满足$\boldsymbol{E}^k_c=\{e_{ij}|e_{ij}=\boldsymbol{1}_{\{\mathrm{IoU}(h^k_i,\ h^k_j)\geqslant\mathrm{Thresh}\}}\}$，其中$\boldsymbol{1}(\cdot)$为指示函数，Thresh为相似度阈值，即两个顶点的空间相似度决定了它们是否相连。最后以贪心的方式为c类别生成一些聚类中心：将图中当前连接数最多的顶点添加至聚类中心集合，再从图中移除该顶点及其关联边，依次迭代。

本模块利用了不同视图下候选框聚类中心分布的一致性。如图1所示，利用最后一个细化分支产生的实例级别伪标签$\boldsymbol{\varphi}^K$，分别在原图的旋转候选框集合$\boldsymbol{H}^{\mathrm{rotate}}_{(0)}$和第$k$个变换视图的旋转候选框集合$\boldsymbol{H}^{\mathrm{rotate}}_{(k)}$上生成聚类中心$\boldsymbol{B}^0=\{(\boldsymbol{b}^0_1,\ \boldsymbol{p}^0_1),\ \cdots,\ (\boldsymbol{b}^0_{M_0},\ \boldsymbol{p}^0_{M_0})\}$和$\boldsymbol{B}^k=\{(\boldsymbol{b}^k_1,\ \boldsymbol{p}^k_1),\ \cdots,\ (\boldsymbol{b}^k_{M_k},\ \boldsymbol{p}^k_{M_k})\}$，其中$\boldsymbol{b}_i\in\mathrm{R}^5$为旋转框五元组，$\boldsymbol{p}_i\in\mathrm{R}^C$指示类别概率。由于RRoI学习器在不同视图下预测的旋转候选框实际分布

不可能完全一致，因此根据空间相似度构建的聚类中心也存在差异，即 $Sim < T_{\text{rotate}}^{(k)}(\boldsymbol{B}^0)$，$\boldsymbol{B}^k >\neq 1$。为此引入聚类中心一致性损失 L_{CC} 衡量这种差异。

L_{CC} 基于DETR类目标检测模型中的匈牙利匹配损失（Carion等，2020）。首先使用∅（表示无目标）元素将 $\boldsymbol{B}^{*0} = T_{\text{rotate}}^{(k)}(\boldsymbol{B}^0)$ 与 $\boldsymbol{B}^k$ 集合填充到相同大小 $M = \max\{M_0, M_k\}$。构建匹配代价 L_{match} 需要同时考虑类别相似度 L_{cls} 和检测框空间位置相似度 L_{box}，使用匈牙利算法为两个集合寻找代价最小的二分匹配 $\hat{\sigma}$，具体定义如下：

$$\hat{\sigma} = \arg\min \sum_i^M L_{\text{match}}(\boldsymbol{B}_i^{*0}, \boldsymbol{B}_{\sigma(i)}^k) \tag{11}$$

$$L_{\text{match}} = L_{\text{cls}}(\boldsymbol{p}_i^{*0}, \boldsymbol{p}_{\sigma(i)}^k) + L_{\text{box}}(\boldsymbol{b}_i^{*0}, \boldsymbol{b}_{\sigma(i)}^k) \tag{12}$$

$$L_{\text{cls}} = -1_{\{(B_i^{*0} \neq \varnothing) \cap (B_{\sigma(i)}^k \neq \varnothing)\}} \sum_c^C \boldsymbol{p}_{\sigma(i)}^k(c) \log \boldsymbol{p}_i^{*0}(c) \tag{13}$$

$$L_{\text{box}} = 1_{\{(B_i^{*0} \neq \varnothing) \cap (B_{\sigma(i)}^k \neq \varnothing)\}} \left(1 - \frac{1}{1 + f(D_{kl}(\boldsymbol{b}_i^{*0}, \boldsymbol{b}_{\sigma(i)}^k))}\right) \tag{14}$$

L_{cls} 为两集合中检测框类别概率的交叉熵损失。L_{box} 基于KL散度损失（Kullback-Leibler Divergence）（Yang等，2021c），其中 $D_{kl}(\cdot)$ 用于计算被转化为二维高斯分布的旋转检测框的KL距离，$f(\cdot)$ 对距离进行非线性变换使损失更加平滑。

最终的聚类一致性损失即为最优匹配下所有匹配对的代价均值：

$$L_{\text{Hungarian}}^k(\boldsymbol{B}^{*0}, \boldsymbol{B}^k) = \frac{1}{M} \sum_i^M L_{\text{match}}(\boldsymbol{B}_i^{*0}, \boldsymbol{B}_{\hat{\sigma}(i)}^k) \tag{15}$$

K 个旋转视图各与原图计算一次聚类一致性损失，最终损失为各视图损失之和：

$$L_{\text{Hungarian}} = \sum_{k=1}^K L_{\text{Hungarian}}^k \tag{16}$$

2.4 最终损失

模型的最终损失为图像级类别损失、PCL细化阶段产生的实例级损失和两个一致性约束损失加权求和。

$$L = L_{\text{MIL}} + \omega_1 L_{\text{PCL}} + \omega_2 L_{RC} + \omega_3 L_{\text{Hungarian}} \tag{17}$$

3 实验结果与分析

3.1 实验数据

3.1.1 DOTA-v1.0数据集

DOTA-v1.0是用于航空图像中目标检测的大规模数据集，包含2806张尺寸在800 × 800到4k × 4k范围的高分辨率遥感图像。该数据集共包含188282个实例，每个实例的标注形式包括水平边界框和旋转边界框两类，可分别用于水平检测和旋转检测两种目标检测任务，且两种任务包含的实例一一对应。待检测目标被分为15个类别：飞机（PL），棒球场（BD），桥梁（BR），地面轨道场（GTF），小型车辆（SV），大型车辆（LV）、船（SH）、网球场（TC）、篮球场（BC）、储油罐（ST）、足球场（SBF）、环岛（RA）、海港（HA）、游泳池（SP）和直升机（HC）。遵循DOTA数据集提供的划分设置，训练集、验证集和测试集的比例约为1/2，1/6和1/3。

3.1.2 DIOR数据集

DIOR数据集包含水平标注（Li等，2020）和旋转标注（Feng等，2020b）两个版本，包含23463张尺寸为800 × 800的遥感图像，共190288个实例。和DOTA类似，两个版本的数据集可分别用于水平检测和旋转检测两种目标检测任务，且两类任务包含的实例一一对应。待检测目标被分为20个类别：飞机（PL）、机场（AP）、棒球场（BF）、篮球场（BC）、桥梁（BR）、烟囱（CM）、大坝（DA）、高速公路服务区（ES）、高速公路收费站（ET）、高尔夫球场（GF）、地面跑道场（GTF）、港口（HB）、立交桥（OP）、船舶（SH）、体育场（SD），储罐（ST），网球场（TC），火车站（TS），车辆（VH），风车（WM）和直升机（HC）。遵循DIOR数据集提供的划分设置，训练集、验证集和测试集的比例约为1/4，1/4和1/2。

3.2 实验设置

SS算法为一张图像预生成的水平候选框个数被限制在4000以内，由旋转变换拓展得到的视图个数 $K = 2$。训练过程中，每次随机选择 $N = 1000$ 个水平候选框参与模型训练。RRoI学习器中RRoI Align和RRoI Pooling的输出维度 $d_1 = d_2 = 7$。检测框旋转一致性约束模块的损失计算中，中心点损失和角度损失的权重为 $\alpha_1 = 1$，$\alpha_2 = 0.4$。聚类中心一致性损失约束模块中，生成边集合使用的空间相似度阈值Thresh = 0.6，非线性函数为 $f(x) = \log(x + 1)$ 形式。最终损失中各部分损失权重取 $w_1 = 1$，$w_2 = w_3 = 0.1$。

模型骨干网络使用ResNet-50，采用随机梯度下降SGD（Stochastic Gradient Descent）优化器，初始学习率、动量和权重衰减分别为0.01、0.9和0.0005。使用4张TITAN RTX GPU以16的批大小训练12个周期。

3.3 实验结果与分析

3.3.1 旋转目标检测和水平目标检测的任务难度对比

对于水平目标检测任务，目标通常具有固定的方向和位置，因此常采用基于锚框（Anchor）和候选框（Proposal）的单阶段或双阶段方法，模型仅需聚焦于每个候选框的分类和边界回归任务。

而对于旋转目标检测任务，由于目标在图像中方向和角度的不确定性，除了致力于原始的分类和回归任务，模型还要设计对旋转角度的预测。同时，旋转目标检测不仅面临常规目标检测中存在的问题，还需关注由旋转角引入的其他问题，例如如何为卷积神经网络中引入旋转不变性、如何优化旋转边界框的表示以及如何消除旋转目标间感受野的混叠影响等。

再以DOTA-v1.0数据集的两个检测任务为例，任务2（水平目标检测）的当前最佳mAP（mean Average Precision）性能为83.1%，高于任务1（旋转目标检测）的最佳mAP性能82.4%，也说明旋转目标检测具有更高的任务难度。综上所述，旋转目标检测相较于水平目标检测具有更复杂的网络设计，在实际问题中也更具挑战性。

3.3.2 基于图像级标注和基于水平框级标注的弱监督任务难度对比

基于水平框标注的弱监督目标检测任务已经具备了目标检测中最核心的监督信号——每个目标的位置信息，这是全监督目标检测模型生成候选框的基础。在弱监督任务设置中保留这样重要的监督信息，有利于模型准确预测目标位置。

基于图像级标注的弱监督目标检测任务使用的监督信号更为抽象和模糊，因为图像级别标签只指示了某个类别目标的存在与否，无法提供具体的目标位置和方向信息。在这种情况下，只能使用特殊的网络结构或策略来训练模型，如生成伪标签进行自我监督等，但受限于监督信息的匮乏，模型的自我监督能力仍是有限的。

综上所述，基于水平框标注的弱监督任务相较于基于图像级标注的弱监督任务，监督信号更为丰富和具体，因此任务难度更小。

3.3.3 实验的同类模型设置

由于本文首次提出了基于图像级别标签的弱监督旋转检测模型，缺乏先例模型，因此选择限制条件更为宽松、难度更小的弱监督检测模型作为同类模型进行比较，具体包括以下两类。

（1）由3.3.1节可知，旋转目标检测的难度高于水平目标检测。因此对于同时标注旋转边界框和水平边界框且实例一一对应的数据集（由3.1节可知，DOTA和DIOR数据集均满足上述要求），若提出的弱监督旋转目标检测模型在旋转任务上的性能高于已有的弱监督水平目标检测SOTA模型在水平任务上的性能，则说明了方法的有效性。

（2）由3.3.2节可知，基于图像级别标注的弱监督任务比基于水平框级别标注的弱监督任务难度更高。因此在相同的数据集上，分别以图像级别和水平框级别的监督信号作为网络训练输入，若前类模型（本文模型）的部分类别准确度得分高于或者与后类模型水平相当，则说明了方法的有效性。

3.3.4 各数据集的实验结果

（1）DIOR的实验结果。表1将模型与任务难度更低的水平弱监督方法进行了性能比较。结果显示：模型在DIOR测试集上取得37.7% mAP，相较于水平弱监督SOTA方法RINet（Feng等，2022）的28.3% mAP，提升了9.4%；同时在飞机、棒球场、烟囱等12个类别上都取得了性能最优。图5展示了表1对比方法在相同图像上与本文模型的可视化对比结果：OICR和PCL作为水平弱监督模型的经典范式，无法很好地处理遥感场景下特有的密集物体问题，存在漏检现象；RINet利用目标的旋转等变性进行设计，漏检现象大大缓解，但仍有部分目标未被检出。

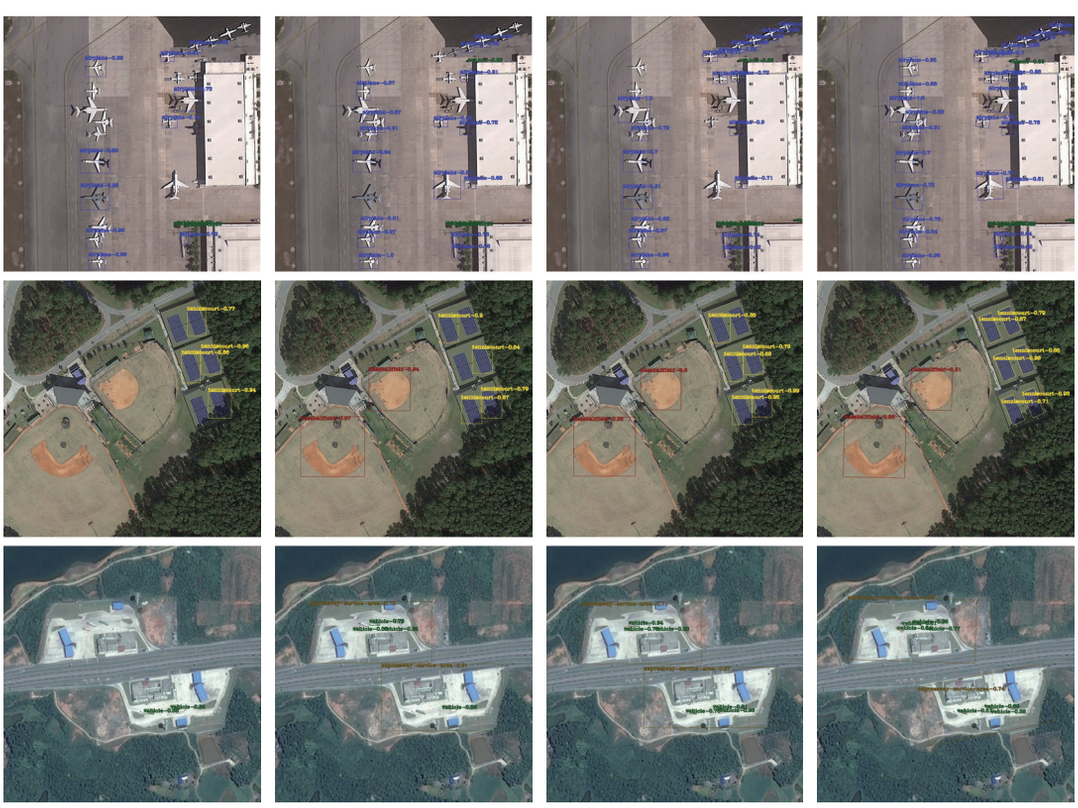
表1 本文模型与基于图像级别标注的水平弱监督方法在DIOR上的AP性能比较

Table 1 AP results of our model and horizontal detection methods based on image labels on DIOR

/%

方法		类别																			
		PL	AP	BF	BC	BR	CM	DA	ES	ET	GF	GTF	HB	OP	SH	SD	ST	TC	TS	VH	WM	mAP
基于图像级别标注的水平弱监督方法	OICR	8.7	28.3	44.1	18.2	1.3	20.2	0.1	0.7	29.9	13.8	57.4	10.7	11.1	9.1	59.3	7.1	0.7	0.1	9.1	0.4	16.5
	PCL	21.5	35.2	59.8	23.5	3.0	43.7	0.1	0.9	1.5	2.9	56.4	16.8	11.1	9.1	57.6	9.1	2.5	0.1	4.6	4.6	18.2
	DCL	20.9	22.7	54.2	11.5	6.0	61.0	0.1	1.1	31.0	30.9	56.5	5.1	2.7	9.1	63.7	9.1	10.4	0.0	7.3	0.8	20.2
	PCIR	30.4	36.1	54.2	26.6	9.1	58.6	0.2	9.7	36.2	32.6	58.8	8.6	21.6	12.1	64.3	9.1	13.6	0.3	9.1	**7.5**	24.9
	TCANet	25.1	30.8	62.9	**40.0**	4.1	67.8	**8.1**	23.8	29.9	22.3	53.9	24.8	11.1	9.1	46.4	13.7	31.0	1.5	9.1	1.0	25.8
	RINet	26.2	**57.4**	62.7	25.1	**9.9**	69.2	1.4	13.3	**36.2**	51.4	53.9	**28.6**	4.8	9.1	52.7	15.8	20.6	**12.9**	9.1	4.7	28.3
基于图像级别标注的水平弱监督方法基于图像级别标注的旋转弱监督方法	视图一致性网络（本文）	**79.4**	25.1	**78.4**	26.6	8.9	**81.5**	1.2	**34.8**	10.9	**56.3**	**58.9**	24.5	**23.6**	**12.2**	**80.8**	**65.4**	**62.4**	9.8	**12.3**	0.5	**37.7**

注：加粗表示当前类别下AP的最高精度。



（a）OICR （b）PCL （c）RINet （d）本文模型

（a）OICR （b）PCL （c）RINet （d）Our model

图5 本文模型与基于图像级别标注的水平弱监督方法在DIOR上的可视化结果比较

Fig. 5 Visualization results of our model and horizontal detection methods based on image labels on DIOR

表2将模型与基于水平框标注的旋转弱监督和全监督方法进行了性能比较。由于图像级别标注和检测框级别标注的信息差距过大，模型性能无法超越基于水平框标注的SOTA性能59.0% mAP（需要注意的是，由于BoxInst-Rbox（Tian等，2021）和H2RBox（Yang等，2022）未在论文中给出各类别具体指标，这里以及后文DOTA-v1.0的mAP结果都是根据论文的开源代码及权重测试得到，与原文指标有些许出入，但总体水平相当）。但飞机、棒球场、烟囱和体育场这4个类别的性能均超过基于水平框标注的弱监督和全监督模型。说明即使在监督信息差距如此巨大的情况下（图像级

别标注只给出了图像中存在哪些类别，而水平检测框标注还包含了目标数量、位置及旋转物体外接矩形的检测关键信息），模型仍然在某些类别目标的检测上占有优势，验证了模型的有效性。图6和图7分别展示了表2对比方法在相同图像上与本文模型的可视化对比结果：图6是全监督模型Fast RCNN和Faster RCNN的预测结果，全监督设置下几乎不存在漏检现象，但检测框的准确度一般；图7是基于水平检测框标注训练的弱监督旋转检测模型，其中BoxInst-Rbox基于物体掩码寻找最小外接矩形，其旋转方向与真实标注偏差较大。

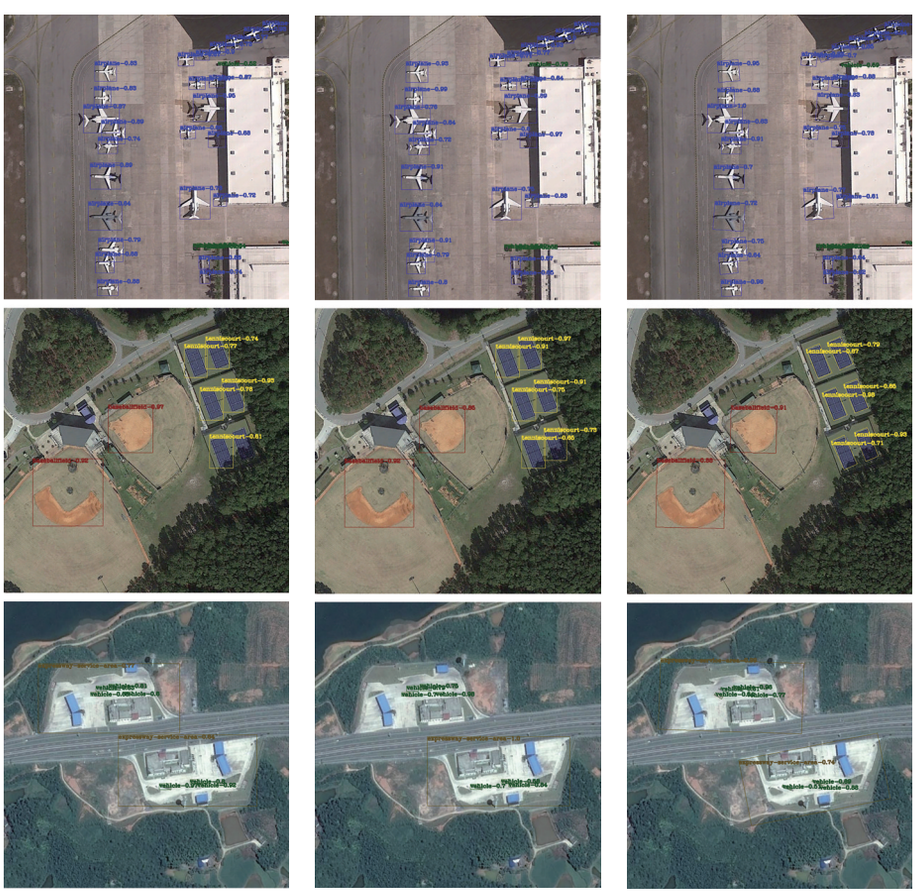
表2 本文模型与基于水平框标注的旋转弱监督方法在DIOR上的AP性能比较

Table 2 AP results of our model and oriented detection methods based on horizontal annotations on DIOR

/%

方法		类别																			
		PL	AP	BF	BC	BR	CM	DA	ES	ET	GF	GTF	HB	OP	SH	SD	ST	TC	TS	VH	WM	mAP
全监督方法	Fast RCNN	44.2	**66.8**	67.0	60.5	15.6	72.3	**52.0**	65.9	44.8	72.1	62.9	**46.2**	38.0	32.1	71.0	35.0	58.3	37.9	19.2	38.1	50.0
	Faster RCNN	50.3	62.6	66.0	80.9	**28.8**	68.2	47.3	58.5	48.1	60.4	67.0	43.9	**46.9**	58.5	52.4	42.4	79.5	48.0	34.8	**65.4**	55.5
基于水平框标注的旋转弱监督方法	BoxInst-RBox	67.5	29.4	74.6	**85.6**	27.7	75.2	30.5	**69.5**	**60.9**	**75.9**	76.8	40.4	45.9	79.5	70.5	**69.7**	81.2	**53.8**	38.0	26.9	**59.0**
	H2RBox	67.9	12.9	76.4	84.0	21.1	72.2	21.8	64.4	58.5	71.9	**80.0**	37.1	45.1	**79.6**	70.1	69.5	**81.5**	48.6	**40.2**	60.3	58.2
基于图像级别标注的旋转弱监督方法	视图一致性网络（本文）	**79.4**	25.1	**78.4**	26.6	8.9	**81.5**	1.2	34.8	10.9	56.3	58.9	24.5	23.6	12.2	**80.8**	65.4	62.4	9.8	12.3	0.5	37.7

注：加粗表示当前类别下AP的最高精度。

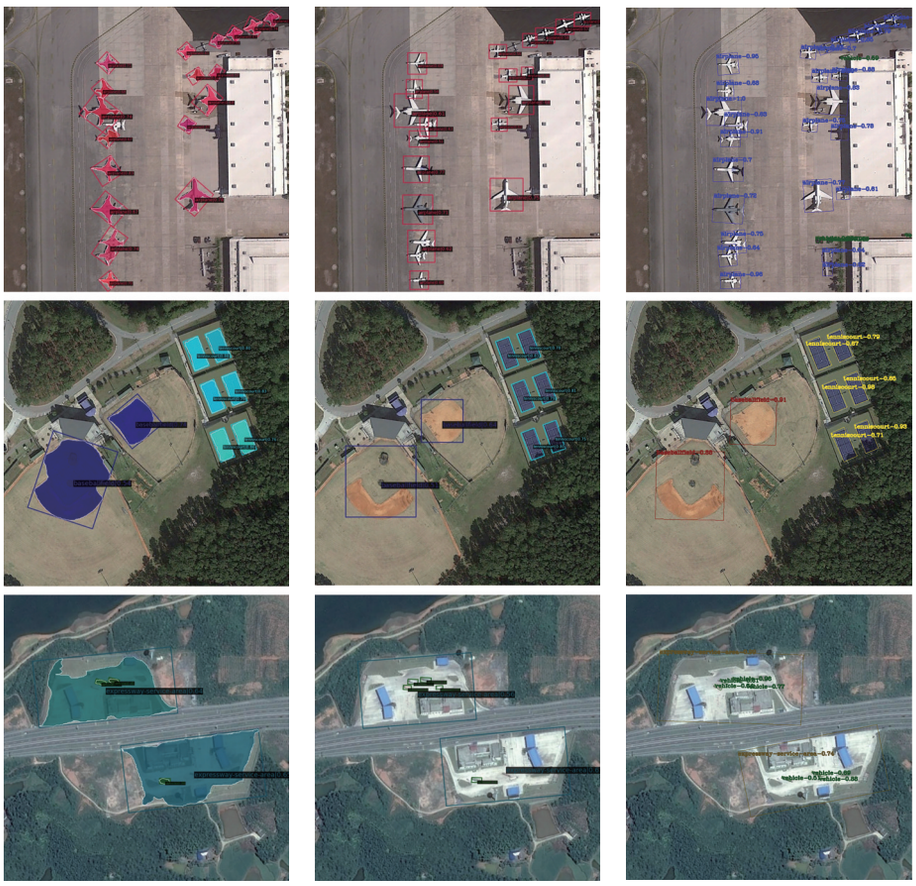


（a）Fast RCNN （b）Faster RCNN （c）本文模型

（a）Fast RCNN （b）Faster RCNN （c）Our model

图6 本文模型与水平全监督方法在DIOR上的可视化结果比较

Fig. 6 Visualization results of our model and fully supervised horizontal detection methods on DIOR



（a）BoxInst-Rbox （b）H2RBox （c）本文模型

（a）BoxInst-Rbox （b）H2RBox （c）Our model

图7 本文模型与基于水平框标注的旋转弱监督方法在DIOR上的可视化结果比较

Fig. 7 Visualization results of our model and horizontal detection methods based on horizontal annotations on DIOR

（2）DOTA-v1.0的实验结果。由于DOTAv-1.0数据集的小物体排列密集且数量众多，这对于缺乏目标数量及位置先验知识的图像级标注弱监督模型并不友好。因此模型在DOTAv-1.0测试集上取得28.1% mAP，相较于DIOR性能有所下降。表3将模型与基于水平框标注的旋转弱监督方法进行了性能比较，虽然仅在游泳池类上优于SOTA模型H2RBox，但棒球场、地面轨道场、网球场和足球场4个类别的性能均高于任务难度更小的BoxInst-Rbox。

表3 本文模型与基于水平框标注的旋转弱监督方法在DOTAv-1.0上的AP性能比较

Table 3 AP results of our model and oriented detection methods based on horizontal annotations on DOTAv-1.0

/%

方法		类别															
		PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
基于水平框标注的旋转弱监督方法	BoxInst-RBox	65.5	30.1	25.1	18.5	50.3	55.2	55.4	66.1	58.5	80.6	32.4	44.4	52.2	66.5	31.4	48.8
	H2RBox	**88.5**	**73.2**	**40.8**	**58.0**	**77.7**	**64.9**	**78.0**	**90.9**	**81.9**	**85.0**	**54.2**	**65.1**	**52.7**	64.4	**36.6**	**67.45**
基于图像级别标注的旋转弱监督方法	视图一致性网络（本文）	56.3	36.7	10.0	25.6	33.9	49.3	15.1	69.2	24.3	73.0	42.3	19.9	12.1	**69.3**	25.5	28.1

注：加粗表示当前类别下AP的最高精度。

图8对模型在DIOR和DOTA-v1.0上的检测结果进行了可视化。由图8可知，模型基本克服了图像级别弱监督任务中常见的漏检问题，成功捕捉到图中相同类别下的多个目标。

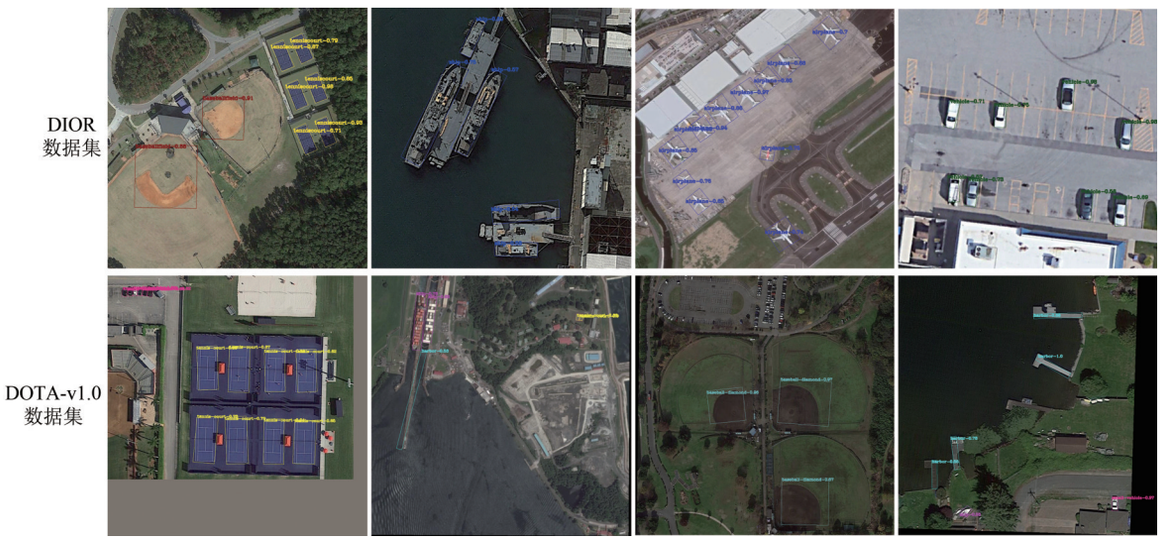


图8 在DIOR和DOTA-v1.0数据集上的检测结果

Fig. 8 Detection results on DIOR and DOTA-v1.0 datasets

（3）预测不佳的实例（bad case）分析。为了更深入地分析模型的不足之处，收集了模型在DIOR数据集上预测不佳的检测实例（即bad case），并将其汇总如图9所示。在图9中绿色细线旋转框表示真实目标位置，而其他颜色的粗线旋转框则表示模型的预测结果。

（a）训练样本少
（a）Lack of sample

（b）前后景难以区分
（b）Similar foreground and background

（c）高长宽比目标
（c）Object with high aspect ratio

（d）目标具有完整语义
（d）Object with complete semantics

图9 DIOR数据集上的badcase

Fig. 9 Badcase on DIOR dataset

观察上述实例后，总结出bad case大致可分为以下4种类型：

（1）待检测类别占总体样本类别比例较少，导致模型对于该类样本的欠拟合。表4列出了DIOR数据集中20个类别实例在训练验证集（Trainval set）的样本占比，图9（a）展示了模型在某些类别下出现的漏检情况。例如，图中漏检类别——大坝DA（dam）和火车站TS（trainstation）的样本占比分别为0.0075和0.0074。这表明模型对于此类训练样本实例较少、且类别比例不均衡的情况存在缺陷。

（2）待检测目标的颜色纹理与背景非常相似，

导致模型难以区分前景和背景。图9（b）展示了模型在这种情况下的漏检问题：上图中风车的白色外观几乎与背景融为一体，因此，两个风车WM（windmill）实例只成功检测到一个；下图中的机场AP（airport）由于其纹理与背景极为相似，也存在漏检情况。

表4 DIOR数据集中20个类别实例在训练验证集(trainval set)的样本占比

Table 4 The sample proportion of the twenty category instances in the DIOR data set in the training verification set

	类别									
	PL	AP	BF	BC	BR	CM	DA	ES	ET	GF
样本占比	0.0277	0.0097	0.0350	0.0158	0.0201	0.0095	0.0075	0.0090	0.0159	0.0075
	类别									
	GTF	HB	OP	SH	SD	ST	TC	TS	VH	WM
样本占比	0.0171	0.0347	0.0195	0.4018	0.0087	0.0447	0.0720	0.0074	0.2016	0.0347

（3）待检测目标具有高长宽比，对旋转角度偏移非常敏感。图9（c）展示了由角度偏移引起的影响：图9中高长宽比的桥梁BR（bridge）和船只SH（ship）由于模型预测角度的轻微偏转，远离了真实的目标位置，导致检测精度急剧下降。为解决这种情况，模型在学习高长宽比物体的旋转角度时应更为慎重。

（4）待检测目标是一个包含完整语义信息的场景，例如港口HB（Harbor）和机场AP（Airport）。图9（d）展示了这种类型的目标漏检情况：上图中的机场范围较大，涵盖多种地物类型，因此模型难以从复杂的地物类型中提取出机场的语义信息；下图中模型虽然可以检测出语义信息明确且具体的船只SH（ship），但却无法捕捉到更大范围语义的港口目标。

通过对这些不同类型的bad case进行分析，我们可以更好地了解模型在旋转检测方面存在的局限性，为改进模型性能提供有益的指导。

3.3.5 消融实验

（1）一致性约束模块。在DIOR数据集对3个一致性约束模块进行消融实验，性能结果如表5所示，可视化结果如图10所示。初始的基线模型只在单一视图下进行基本的弱监督检测，性能较低且虚报率较高。引入图像类别的一致性约束后，模型虚报减少且mAP提升了7.1%，但此时模型为了迎合不同视角下输出的一致性，倾向于为所有检测框预测相似的旋转角，如图10（b）所示。因此接下来引入可以显式约束旋转角度的旋转一致性和聚类中心一致性模块：如图10（c）和（d）所示，加入旋转一致性模块后，模型预测的旋转角度更加丰富；而加入聚类中心一致性模块后，模型预测的旋转角度更加准确。

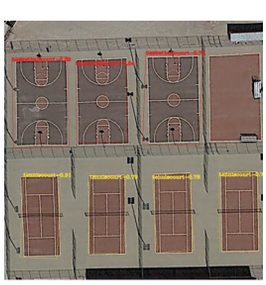
表5 3个一致性模块在DIOR上的消融实验

Table 5 Ablation study of three consistency modules

基线模型	图像类别一致性	检测框旋转一致性	聚类中心一致性	mAP/%
				12.1
	√			19.2
√	√	√		30.2
	√		√	29.8
	√	√	√	**37.7**

注：加粗表示当前模块设置下mAP的最高精度。



(a) 基线模型
(a) Baseline

(b) 为(a)增加类别一致性
(b) Class-consistency+(a)

(c) 为(b)增加旋转一致性
(c) Rotation-consistency+(b)

(d) 为(c)增加聚类一致性
(d) Cluster-consistency+(c)

图10 消融实验可视化结果

Fig. 10 Visualization of ablation study

(2) 损失函数。在DIOR数据集上对各一致性约束模块的损失权重进行消融，即对式(17)中w_i的取值进行实验。为了确保各模块对深度学习网络权重更新的贡献度，通常使用超参将损失函数的各部分调整到相同量级。实验结果如表6所示，首先选取合适的w_i值使得$\omega_1 L_{\mathrm{PCL}}$ ： $\omega_2 L_{\mathrm{RC}}$ ： $\omega_3 L_{\mathrm{Hungarian}} \approx 1$ ： 1 ： 1，即将3种一致性约束损失调整至相同比重。再对哪种一致性约束占梯度更新的主导损失进行消融，分别使用2∶1和5∶1两种设置完成实验。分析可知，选取w_i使得$\omega_1 L_{\mathrm{PCL}}$：$\omega_2 L_{\mathrm{RC}}$：$\omega_3 L_{\mathrm{Hungarian}} \approx 1$ ： 2 ： 1时，模型精度最高，即旋转一致性模块占梯度主导时，模型训练效果达到最佳。但总体上看，相同量级下，损失函数超参的选取对模型最终精度的影响不大。

表6 各一致性约束损失所占比例对模型精度的影响

Table 6 Ablation study of loss weights for three consistency modules

$w_1 L_{\mathrm{PCL}}$	$w_2 L_{\mathrm{RC}}$	$w_3 L_{\mathrm{Hungarian}}$	mAP/%
1	1	1	37.5
2	1	1	37.6
1	**2**	**1**	**37.7**
1	1	2	37.5
5	1	1	37.1
1	5	1	37.4
1	1	5	37.2

注：加粗表示当前一致性约束损失比例设置下mAP的最高精度。

3.3.6 模型复杂度分析

本文提出的旋转弱监督模型只需要图像级别标签即可训练，减轻标注负担的同时也提升了训练效率：与主流的全监督旋转检测模型和基于水平框的弱监督旋转检测模型相比，相同的批大小（batch size）设置下，本文模型占用内存更小，且推理阶段单张图片的推理速度更快。具体推理时间及内存占用量见表7。

表 7 不同模型的内存消耗及推理时间

Table 7 Used memory and time consumption of different models

方法		图片尺寸	批大小	GPU	推理时间/s	内存/GB
全监督旋转目标检测模型	RoI Transformer	1024×1024	2	TiTanXp	19.3	7.56
	Gliding Vertex				16.4	8.45
	ReDet				14.1	7.89
	Oriented RCNN				16.1	8.37
基于水平框标注的弱监督旋转目标检测模型	H2RBox	1024×1024	2	TiTanXp	28.5	7.02
	视图一致性网络（本文）	1024×1024	2	TiTanXp	**30.6**	**6.79**

注：加粗表示当前模型的推理速度最快，内存消耗最低。

4 结 论

本文针对旋转检测框标注复杂且耗时的问题，首次提出了遥感场景中仅使用图像级别标注的弱监督旋转目标检测模型。模型设计了一种视图一致性网络，通过构造不同视图下图像类别、检测框旋转变换和聚类中心分布的3个一致性约束，指导模型优化方向。经过在主流遥感目标检测数据集上的实验验证，该方法显示出了可行性和优越性。

然而，我们也必须认识到模型本身存在一些局限性。首先，当前的模型精度与基于旋转检测框标注的全监督模型仍有差距，图像级别标注和检测框级别标注之间的先验知识差距仍无法完全弥合。虽然我们的模型在弱监督条件下表现出了良好的性能，但在特定数据上它无法达到全监督模型的准确程度。

其次，从检测不佳的预测实例的分析中，我们发现模型针对遥感影像特有的检测难点仍然存在局限性。例如，遥感影像俯瞰视角下地物纹理的复杂度，以及大尺度背景下语义信息的丰富性，都为遥感旋转目标检测带来了困难。这些特殊情况需要更深入的研究和创新来进一步提高模型的性能。

未来，基于图像级别的弱监督遥感检测模型将有广泛的应用场景。例如，可以应用于从大规模弱标注卫星图像数据中进行目标检测，以实现自动化的遥感影像分析。此外，该模型在城市规划、环境监测、农业智能等领域也有着潜在的应用价值。

为进一步提升模型性能和推动相关领域的发展，未来的研究可以着重探索以下方向：（1）优化视图一致性网络，提升模型对特殊遥感场景的适应性；（2）进一步发掘模型克服遥感场景中特有检测难点的能力，以更好地应对遥感旋转目标检测的挑战；（3）结合更多先进的深度学习技术和遥感数据处理方法，进一步提高模型的鲁棒性和泛化能力；（4）将模型拓展至更广泛的遥感应用领域，如海洋监测、资源调查、应急响应等，以实现遥感技术的全面应用和价值发挥。

综上所述，本文提出的基于图像级别标注的弱监督旋转目标检测模型为遥感影像处理领域带来了重要的进展。虽然模型存在一定的局限性，但其应用潜力和未来发展方向仍然非常广阔。相信随着进一步的研究和探索，该模型将为遥感目标检测和相关领域带来更多有价值的应用和成果。

参考文献（References）

Bilen H and Vedaldi A. 2016. Weakly supervised deep detection networks//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE: 2846-2854 [DOI: 10.1109/CVPR.2016.311]

Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S. 2020. End-to-end object detection with transformers//16th European Conference on Computer Vision. Glasgow: Springer: 213-229 [DOI: 10.1007/978-3-030-58452-8_13]

Cheng G, Wang J B, Li K, Xie X X, Lang C B, Yao Y Q and Han J W.

2022. Anchor-free oriented proposal generator for object detection. IEEE Transactions on Geoscience and Remote Sensing, 60: 5625411 [DOI: 10.1109/TGRS.2022.3183022]

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 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 2844-2853 [DOI: 10.1109/CVPR.2019.00296]

Feng X X, Han J W, Yao X W and Cheng G. 2020a. Progressive contextual instance refinement for weakly supervised object detection in remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 58(11): 8002-8012 [DOI: 10.1109/TGRS.2020.2985989]

Feng X X, Han J W, Yao X W and Cheng G. 2020b. TCANet: triple context-aware network for weakly supervised object detection in remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 59(8): 6946-6955 [DOI: 10.1109/TGRS.2020.3030990]

Feng X X, Yao X W, Cheng G and Han J W. 2022. Weakly supervised rotation-invariant aerial object detection network//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE: 14126-14135 [DOI: 10.1109/CVPR52688.2022.01375]

Girshick R. 2015. Fast R-CNN//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago: IEEE: 1440-1448 [DOI: 10.1109/ICCV.2015.169]

Han J M, Ding J, Xue N and Xia G S. 2021. ReDet: a rotation-equivariant detector for aerial object detection//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 2785-2794 [DOI: 10.1109/CVPR46437.2021.00281]

He K M, Gkioxari G, Dollár P and Girshick R. 2017. Mask R-CNN//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.322]

Huang Z Y, Zou Y, Bhagavatula V and Huang D. 2020. Comprehensive attention self-distillation for weakly-supervised object detection//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc.: 1409

Li K, Wan G, Cheng G, Meng L Q and Han J W. 2020. Object detection in optical remote sensing images: a survey and a new benchmark. ISPRS Journal of Photogrammetry and Remote Sensing, 159: 296-307 [DOI: 10.1016/j.isprsjprs.2019.11.023]

Li W T, Chen Y J, Hu K X and Zhu J K. 2022a. Oriented RepPoints for aerial object detection//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE: 1819-1828 [DOI: 10.1109/CVPR52688.2022.00187]

Li W T, Liu W Y, Zhu J K, Cui M M, Hua X S and Zhang L. 2022b. Box-supervised instance segmentation with level set evolution//17th European Conference on Computer Vision. Tel Aviv: Springer: 1-18 [DOI: 10.1007/978-3-031-19818-2_1]

Li X Y, Kan M N, Shan S G and Chen X L. 2019. Weakly supervised object detection with segmentation collaboration//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE: 9734-9743 [DOI: 10.1109/ICCV.2019.00983]

Ren S Q, He K M, Girshick R and Sun J. 2015. Faster R-CNN: towards real-time object detection with region proposal networks//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press: 91-99

Tang P, Wang X G, Bai S, Shen W, Bai X, Liu W Y and Yuille A. 2020. PCL: proposal cluster learning for weakly supervised object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(1): 176-191 [DOI: 10.1109/TPAMI.2018.2876304]

Tang P, Wang X G, Bai X and Liu W Y. 2017. Multiple instance detection network with online instance classifier refinement//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 3059-3067 [DOI: 10.1109/CVPR.2017.326]

Tian Z, Shen C H, Wang X L and Chen H. 2021. BoxInst: high-performance instance segmentation with box annotations//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 5439-5448 [DOI: 10.1109/CVPR46437.2021.00540]

Uijlings J R R, van de Sande K E A, Gevers T and Smeulders A W M. 2013. Selective search for object recognition. International Journal of Computer Vision, 104(2): 154-171 [DOI: 10.1007/s11263-013-0620-5]

Wang B L, Zhao Y Q and Li X L. 2022. Multiple instance graph learning for weakly supervised remote sensing object detection. IEEE Transactions on Geoscience and Remote Sensing, 60: 5613112 [DOI: 10.1109/TGRS.2021.3123231]

Wei Y C, Shen Z Q, Cheng B W, Shi H H, Xiong J J, Feng J S and Huang T. 2018. TS2C: tight box mining with surrounding segmentation context for weakly supervised object detection//Proceedings of the 15th European Conference on Computer Vision. Munich: Springer: 454-470 [DOI: 10.1007/978-3-030-01252-6_27]

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 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 3974-3983 [DOI: 10.1109/CVPR.2018.00418]

Xie X X, Cheng G, Wang J B, Yao X W and Han J W. 2021. Oriented R-CNN for object detection//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE: 3500-3509 [DOI: 10.1109/ICCV48922.2021.00350]

Yan G, Liu B X, Guo N, Ye X C, Wan F, You H H and Fan D R. 2019. C-MIDN: coupled multiple instance detection network with segmentation guidance for weakly supervised object detection//Pro-

ceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE: 9833-9842 [DOI: 10.1109/ICCV. 2019. 00993]

Yang X, Yan J C, Feng Z M and He T. 2021a. R3Det: refined single-stage detector with feature refinement for rotating object//Proceedings of the 35th AAAI Conference on Artificial Intelligence. [s.l.]: AAAI: 3163-3171 [DOI: 10.1609/aaai.v35i4.16426]

Yang X, Hou L P, Zhou Y, Wang W T and Yan J C. 2021b. Dense label encoding for boundary discontinuity free rotation detection//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 15814-15824 [DOI: 10. 1109/CVPR46437.2021.01556]

Yang X, Yang X J, Yang J R, Ming Q, Wang W T, Tian Q and Yan J C. 2021c. Learning high-precision bounding box for rotated object detection via Kullback-Leibler divergence//Proceedings of the 35th International Conference on Neural Information Processing Systems. [s.l.]: Curran Associates Inc.: 1405.

Yang X, Zhang G F, Li W T, Zhou Y, Wang X H and Yan J C. 2022. H2RBox: horizontal box annotation is all you need for oriented object detection//The Eleventh International Conference on Learning Representations. Kigali: ICLR

Yao X W, Feng X X, Han J W, Cheng G and Guo L. 2021. Automatic weakly supervised object detection from high spatial resolution remote sensing images via dynamic curriculum learning. IEEE Transactions on Geoscience and Remote Sensing, 59(1): 675-685 [DOI: 10.1109/TGRS.2020.2991407]

Yuan Y Q, Li L, Yao X W, Li L J, Feng X X, Cheng G and Han J W. 2023. A comprehensive review of optical remote-sensing image object detection datasets. National Remote Sensing Bulletin, 27(12): 2671-2687 (袁一钦, 李浪, 姚西文, 李玲君, 冯晓绪, 程塬, 韩军伟. 2023. 光学遥感图像目标检测数据集综述. 遥感学报, 27(12): 2671-2687) [DOI: 10.11834/jrs.20233457]

Zhang L, Zhang Y S, Yu Y, Ma Y Z and Jiang H G. 2022. Survey on object detection in tilting box for remote sensing images. National Remote Sensing Bulletin, 26(9): 1723-1743 (张磊, 张永生, 于英, 马永政, 姜怀刚. 2022. 遥感图像倾斜边界框目标检测研究进展与展望. 遥感学报, 26(9): 1723-1743) [DOI: 10.11834/jrs. 20210247]

View-consistency network for weakly supervised oriented object detection in remote sensing images

FANG Tingting[1,2], LIU Bin[1,2], CHEN Chunhui[3,4], LI Xiangyun[3,4]

1. School of Cyber Science and Technology, University of Science and Technology of China, Hefei 230026, China;
2. Key Laboratory of Electromagnetic Space Information, Chinese Academy of Science, Hefei 230026, China;
3. Anhui Provincial Geomatic Center, Hefei 230031, China;
4. Key Laboratory of JiangHuai Arable Land Resources Protection and Eco-restoration, Hefei 230031, China

Abstract: The objective of this study is to propose a novel oriented object detection model that can effectively detect objects in remote sensing images, while alleviating the challenges of labor-intensive and time-consuming annotation processes. Specifically, the aim is to develop an efficient and effective approach that only requires image-level annotations, which can improve the availability and diversity of remote sensing rotation object detection datasets. The proposed model is designed to overcome the limitations of traditional annotation methods that rely on bounding box annotations, which can be subjective, inconsistent, and time-consuming to create. By leveraging image-level annotations, the proposed model can greatly reduce the annotation effort, accelerate the annotation process, and enhance the scalability and applicability of remote sensing object detection in various scenarios and domains.

The proposed method introduces a novel weakly supervised oriented object detection paradigm for remote sensing scenes. The model is trained in a progressive manner, starting with coarse image-level annotations and gradually refining the detection results. This approach allows the model to learn from limited annotations and adapt to the complexities of remote sensing data, which often exhibit large scale, diverse appearance, and significant rotation variations. The model incorporates consistency constraints of image-level annotation, the oriented bounding box position and the cluster center distribution in different rotation views, to enhance the accuracy and robustness of the detection performance. The oriented bounding box position consistency ensures that the predicted rotation angles of the bounding boxes are consistent across different views of the same object, while the distribution consistency of clustering centers is measured using the Hungarian loss, which ensures that the predicted object centroids are consistent across different views. These consistency constraints are designed to improve the accuracy of the model in handling rotation variations, which is a common challenge in remote sensing object detection.

The proposed model is evaluated on two mainstream remote sensing object detection datasets, DIOR and DOTA-v1.0. The experimental results demonstrate that the performance of the proposed model is significantly improved compared to state-of-the-art weakly supervised remote sensing object detection models. Results show that our performance is significantly improved compared to the state-of-

the-art object detectors with less strict weakly-supervised settings in remote sensing images, which also highlight the effectiveness and potential of the proposed image-level annotation-based oriented object detection model in addressing the challenges of remote sensing object detection. Furthermore, the model's ability to generalize well to different datasets showcases its robustness and versatility.

In conclusion, the proposed image-level annotation-based oriented object detection model is an innovative approach that addresses the challenges of labor-intensive and time-consuming annotation processes in remote sensing object detection. By leveraging image-level annotations and incorporating consistency constraints, the proposed model achieves improved detection performance while reducing the annotation effort and improving annotation efficiency. The experimental results on DIOR and DOTA-v1.0 datasets demonstrate the superior performance of the proposed model compared to state-of-the-art models, highlighting its potential for practical applications in remote sensing object detection and related fields. Future research can further explore the potential of image-level annotation-based oriented object detection models by incorporating more advanced techniques, exploring different data sources, and investigating real-world applications in remote sensing, geospatial analysis, and other related fields.

Key words: remote sensing, oriented object detection, object detection, weakly supervised, deep learning, rotation consistency, image-level label, hungarian loss, DOTA dataset

Supported by Natural Resources Science and Technology Project of Anhui Province (No. 2021-K-14)