

基于双分类头的遥感图像精细化目标检测方法

张锋¹, 滕书华^{1,2}, 韩幸³, 王应谦⁴, 王雪莹⁴

1. 湖南第一师范学院 电子信息学院, 长沙 410205;

2. 南宁学院 人工智能与软件学院, 南宁 530299;

3. 湖南师范大学 医学院, 长沙 410081;

4. 国防科技大学 电子科学学院, 长沙 410073

摘要: 高分辨率遥感图像的可获得性大幅提升, 使得遥感图像目标精细化检测成为了遥感以及计算机视觉领域重要的研究方向。针对遥感图像目标精细化检测中存在的相似数据利用不充分、错误标签影响模型精度和相似类别难以区分的问题, 本文提出了一种基于双分类头的遥感图像精细化目标检测方法。首先, 针对遥感图像精细化目标检测中无法有效利用相似数据的问题, 提出一种双分类检测头, 不同的分类头分别对不同数据集训练, 让类别定义不同的相似数据共同参与训练, 进而有效利用相似数据, 显著提升了模型精度。其次, 针对训练标签噪声问题, 设计一种基于预测的错误标签过滤方法, 减小错误标签对模型训练的影响。最后, 针对精细化目标检测中类内差异大、类间差异小的问题, 定义一种 Margin 交叉熵损失, 通过增大分类边界提升了模型精度。在精细化遥感目标检测竞赛数据集和 FAIR1M 数据集上的实验表明, 本文提出的方法显著提高了遥感影像目标精细化检测的精度和鲁棒性。代码已开源在 <https://github.com/zf020114/DCH>。

关键词: 遥感, 深度学习, 精细化目标检测, 有向目标检测, 双分类头, 错误标签过滤, Margin 交叉熵损失

中图分类号: TP751/P2

引用格式: 张锋, 滕书华, 韩幸, 王应谦, 王雪莹. 2026. 基于双分类头的遥感图像精细化目标检测方法. 遥感学报, 30(1): 170-182

Zhang F, Teng S H, Han X, Wang Y Q and Wang X Y. 2026. A fine-grained object detection method for remote sensing images based on dual classification head. National Remote Sensing Bulletin, 30(1): 170-182 [DOI: 10.11834/jrs.20254243]

1 引言

随着遥感技术的快速发展以及成像平台等硬件设备性能的提升, 卫星影像分辨率已达亚米级。在轨卫星数量飞速增长, 为遥感影像目标检测识别提供了海量的数据资源 (李德仁等, 2017)。高分遥感影像是指通过高分辨率卫星和飞机等平台获取的高分辨率的地球表面影像。这些影像具有清晰度高、色彩还原度高和几何精度高的特点, 通过高分遥感图像可以获取地球表面的详细信息, 为各种地理信息系统的应用提供基础数据支持。高分辨率遥感影像的可获得性大幅提升, 使得遥

感影像目标的精细化检测识别成为遥感技术以及计算机视觉领域的重要研究方向 (姚艳清等, 2021; 罗会兰和陈鸿坤, 2020; 肖进胜等, 2022)。高分遥感图像具有目标类型多、目标数量大、目标方向多变和成像视角单一等特点。此外, 不同平台、不同光照、不同天气条件、不同大气参数等都会对遥感图像获取产生影响。这些因素使得高分遥感图像目标检测与自然图像目标检测相比面临着更多的难点和更大的挑战。

遥感影像精细化目标检测主要体现在两个方面: 一是目标表示方式的精细化。遥感图像中的目标具有任意的方向, 通用目标检测中的水平矩

收稿日期: 2024-06-14; 预印本: 2025-09-18

基金项目: 国家自然科学基金 (编号: 62207030); 湖南省自然科学基金 (编号: 2023JJ30185); 湖南省教育厅科学研究重点项目 (编号: 22A0640)

第一作者简介: 张锋, 研究方向为计算机视觉、目标检测、实例分割、目标跟踪。E-mail: 0201zhangfeng@163.com

通信作者简介: 滕书华, 研究方向为计算机视觉、数据挖掘。E-mail: tengshuhua1979@sohu.com

相关代码链接: <https://github.com/zf020114/DCH>

形框难以紧凑、精准地表示被检测物体的位置和范围, 因此遥感图像的目标精细化检测需要设计基于旋转矩形框的检测模型, 使用紧凑的包围目标的旋转矩形框来表示被检测目标, 提升对目标的精确定位能力 (张磊 等, 2022)。二是目标分类的精细化。精细化检测识别任务需要并识别出常规目标中的子类, 类间差异更小。虽然目前遥感图像目标检测领域取得了长足的发展, 在一些常见的遥感图像数据集如 DOTA (Xia 等, 2018)、DIOR (Li 等, 2020) 和 FAIR1M (Sun 等, 2022) 上均取得了较好的检测结果, 但是在遥感图像精细化目标检测还存在以下问题:

(1) 相似数据利用问题。高精度的模型需要大规模的数据集, 但大规模高精度的人工标注往往需要巨大的成本, 因此很难继续扩大训练数据集规模。同时, 通常存在大量的相似数据集, 这些数据规模大于根据特定任务需求收集和标注的数据集。对于基于深度学习的模型, 在较小的数据规模下要精细区分数量较多的类别将会面临训练数据不足的问题, 并且当前的精细化检测方法无法有效利用相似数据集, 这将极大影响模型的精度。目前很少有工作研究如何构建网络结构使得网络不仅能使用该任务的训练集, 还能使用其他相似的数据进行训练。

(2) 错误标签问题。大规模精细化的人工标注不可避免会出现一些错误标注, 这些错误标注将影响模型的精度, 而目前的遥感图像目标检测领域对于如何妥善处理错误标签、减小错误标签对模型训练的影响研究不足。

(3) 精细分类问题。遥感图像目标精细化检测不同于普通的图像检测识别任务, 在遥感影像中, 各型号飞机、舰船的不同子类之间具有较高的相似性 (如轮廓、结构等), 导致飞机和舰船的精细化检测具备极高难度。同时, 高分辨率光学遥感影像在提供目标丰富的纹理、颜色、形状等特征信息的同时也引入了大量的冗余信息, 导致了影像背景极为复杂。

为此, 本研究为解决遥感图像目标精细化检测中存在的相似数据利用不充分、错误标签影响模型精度和相似类别难以区分的问题, 提出一种基于双分类头的遥感图像精细化目标检测方法。使用双分类检测头, 不同的分类头分别对不同数据集训练, 让类别定义不同的相似数据共同参与

训练, 进而有效利用相似数据; 设计一种基于预测的错误标签过滤方法, 减小错误标签对模型训练的影响; 定义一种 Margin 交叉熵损失, 通过增大分类边界提升了模型精度。本文对提出模型进行了实验验证, 以期对遥感图像精细化目标检测提供一种有效利用相似数据集的方法, 提升遥感图像精细化目标检测的精度。

2 研究方法

遥感图像目标精细化检测主要表现在两个方面: 一是表示方式的精细化, 使用旋转矩形框表示被检测目标, 但旋转矩形框的常用参数表示方式在边界条件下会产生突变, 造成训练不稳定; 二是检测类别的精细化, 使得模型能够检测更加精细的目标类别。但是要区分精细的类别, 需要提取更加具有鉴别力的特征。同时, 更加精细的类别对标注人员提出了更高要求, 难免会出现类别标注错误的现象。

2.1 表示方式的精细化

现有的遥感图像有向目标检测方法大多基于通用目标检测框架做相应改进 (张省 等, 2022)。目前多数的遥感图像精细化目标检测工作专注于表示方式的精细化, 这类方法使用旋转矩形框表示目标 (张磊 等, 2022)。RRPN (Rotation Region Proposal Networks) (Ma 等, 2018), 该方法首次将旋转框引入检测网络, 通过生成大量旋转锚框检测有向目标, 提出了一种旋转区域推荐网络, 该网络在每个位置使用 6 种角度、3 种尺度和 3 种长宽比, 生成 54 个旋转锚框。该方法还使用旋转区域池化层 (RRol pooling) 提取区域的特征, 但是由于生成锚框过多导致推理速度较慢。R²CNN (Jiang 等, 2018) 在 Faster RCNN (Ren 等, 2015) 的基础上, 使用多尺度推荐区域池化 (RoI Pooling) 提取感兴趣区域 (ROI) 特征, 同时预测目标的水平和旋转包围框。RoI Transformer (Ding 等, 2018) 使用一个轻量级转换模块将 RPN 生成的水平锚框转换成有向锚框并使用 RRol Warping 模块提取方向不变的特征以增强特征表达能力。Oriented RCN (Xie 等, 2021), 使用四边形的外接矩形以及与外接矩形顶边和右边的偏移量来表示推荐区域 (Proposal), 然后使用有向区域推荐网络 (Oriented RPN) 直接生成有向推荐区域, 最后通

过旋转区域池化 (Rotated RoIAlign) 提取特征, 并通过有向目标检测头执行分类和回归。EIRNet (Han 等, 2022) 设计了具有两个融合方向的密集特征融合网络 (DFF-Net), 以最大限度地利用多层信息, 减少信息冗余, 实现高分遥感图像上舰船的精细化检测与识别。

无锚框的有向目标检测方法能够避免锚框带来的一系列问题。BBAVectors (Yi 等, 2021) 首先检测目标的中心点, 然后通过回归包围框的边缘感知向量 (BBAVectors) 来获取旋转矩形。PolarDet (Zhao 等, 2021) 在 CenterNet (Zhou 等, 2019) 的基础上通过回归中心点、中心点的偏置、4 个顶点的极角和极径、一个较短边和 4 个极比来表示有向目标。Oriented RepPoints (Li 等, 2022) 利用自适应点表示有向目标, 能够捕捉任意方向目标的几何信息。DARDet (Zhang 等, 2021) 通过迭代回归的方式在特征图的每个位置先后预测旋转框的 5 个参数以及偏置来获得旋转检测框。Oriented-DETR 使用点和轴表示遥感目标, 这种表示方式将位置和旋转分离, 解决了传统基于边界框的方法中常见的损失不连续问题 (Zhao 等, 2024)。

2.2 检测类别的精细化

目前也有部分工作通过提取更加精细的特征提升遥感图像精细化目标检测精度。由于遥感目标具有方向任意性, 但多数深度神经网络具有方向敏感性, 相同目标在不同方向下提取的特征不同, 因此提取目标的旋转不变性特征或更具鉴别力的特征可以提升模型的检测精度。 S^2 Anet (Han 等, 2021b) 首先使用特征对齐模块 (FAM) 来提取与旋转锚框对齐的特征, 然后使用定向检测模块 (ODM) 同时产生方向不变和方向敏感的特征, 以缓解有向目标检测中分类和定位不一致的问题。ReDet (Han 等, 2021a) 使用具有方向不变性的骨干网络来提取方向不变特征, 使用方向不变推荐区域对齐模块 (RiRoI Align) 自适应地提取方向不变特征, 提高了有向目标检测精度。ARC 提出了一种自适应旋转卷积 (Adaptive Rotated Convolution) 模块, 该模块通过自适应旋转卷积核来提取不同图像中不同方向的目标特征 (Pu 等, 2023), Large Selective Kernel Network (LSKNet) (Li 等, 2023) 设计了新的骨干网络, 首先通过使

用大小为 5×5 的普通卷积与大小为 7×7 的膨胀卷积相结合来达到大小为 23×23 的大核卷积的效果, 从而使得特征提取网络具有大感受野。然后通过使用普通 2 维卷积与 Sigmoid 函数相结合设计了一种模块选择网络, 使得网络可以针对不同的检测目标动态选取不同感受野大小的卷积, 从而提升骨干网络的特征分辨能力, 获得更好的检测结果。STD 提出了一种称为空间变换解耦的新方法, 为使用视觉变换器 (Vision transformer) 进行遥感目标检测提供了一种简单而有效的解决方案 (Yu 等, 2024)。由于待检测的类别更加精细, 在标注遥感目标过程中难免会引入噪声。为减轻错误标签对模型的影响, Generalized Cross Entropy Loss (Zhang 等, 2018) 和 Symmetric Cross Entropy Loss (Wang 等, 2019) 能够在噪声环境下保持较好的统计一致性, 从而增强模型鲁棒性。Co-teaching (Han 等, 2018) 和 Co-teaching+ (Yu 等, 2019) 使用两个或多个网络模型相互学习, 通过筛选出预测结果不一致的样本进行训练, 从而避免模型对噪声数据的过拟合。Hu 等 (2022) 提出了一种矫正最大化训练的策略, 并设计一种概率微分滤波器来过滤和修正错误标签, 提升了在有噪标签下模型的精度。

3 研究方法

针对上述遥感图像精细化目标检测中存在的 key 问题, 本文提出了一种基于双分类头的高分遥感图像精细化目标检测方法。首先, 提出了一种全新双分类头, 让相似的数据集全程参与模型训练; 其次, 提出了一种基于预测的错误标签过滤模块, 在训练过程中根据预测置信度过滤错误标签; 最后定义了 Margin 交叉熵损失函数以增大精细分类中的分类边界。下面首先给出本文方法的基础框架, 在此基础上对上述 3 个方法展开介绍。

3.1 基础框架

由于双阶段目标检测算法在小目标检测方面的性能通常好于单阶段的目标算法, 结合高分辨率遥感图像中存在大量小目标的特点, 本文选择双阶段旋转目标检测算法 Oriented R-CNN (Xie 等, 2021) 为基础框架。图 1 为 Oriented R-CNN 总体框架图。该方法通过在水平包围框的基础上预测有向候选框的顶点相对于水平框顶边和右边中

点的偏移量来获得目标的4个角点，具有检测精度高，推理速度快的特点。此外周围背景环境可以提供的目标形状、方向和其他特征的重要信息，

物体的成功识别往往还依赖于其背景，因此本研究采用LSKNet（Li等，2023）作为骨干网络，以便更有效的检测遥感目标。

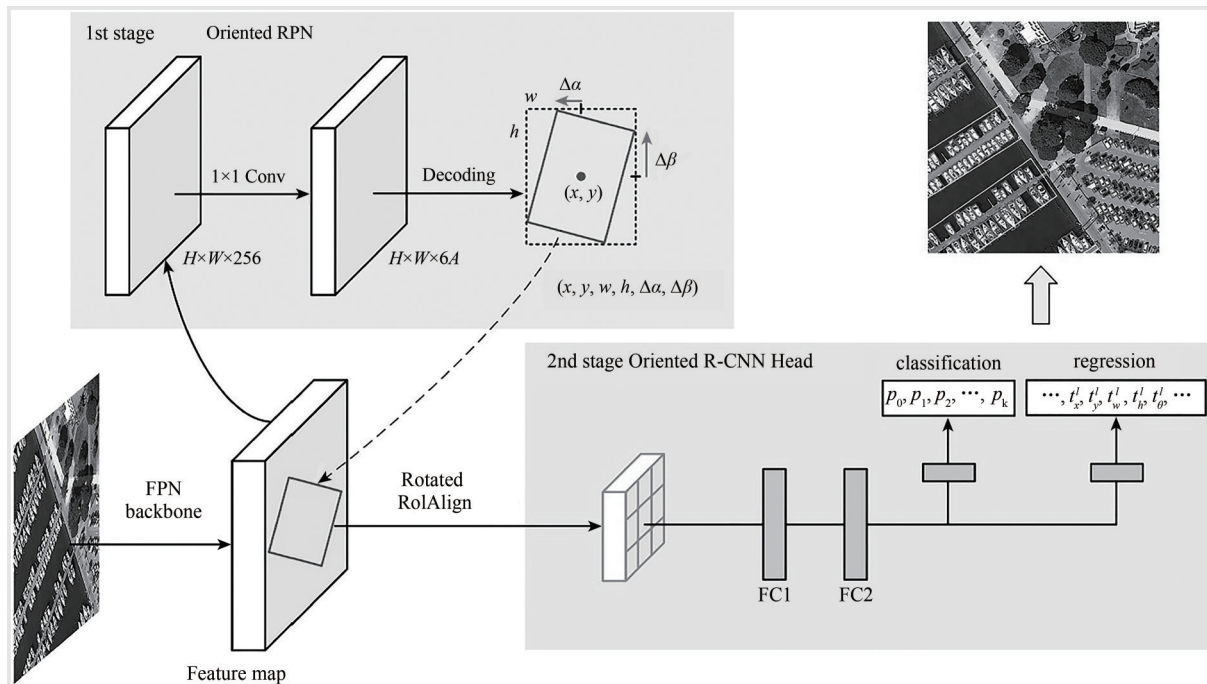


图1 Oriented R-CNN的总体框架图

Fig. 1 Overall framework of oriented R-CNN

3.2 双分类检测头

在多数遥感精细化目标检测任务中均存在一些公开数据集，这些数据集与训练数据集相似，其影像场景相似，包含的目标相似，只是目标的类别标签定义不同，类别标注的精细化程度不同。精细化目标检测竞赛数据集包含了两个相似数据集（训练集和参考数据集），其示例如图2所示。目前主流的目标检测方法对相似数据的使用策略是先用相似数据集进行预训练，然后使用该预训练模型初始化模型参数。这种方式只是将相似数据集训练的模型作为第二次训练的初始化参数，参考数据集并没有真正参与训练，因而这种方法对数据的利用效率较低，没有充分发挥相似数据的价值。对于相似的数据集由于其对目标类别的定义不同，同样的目标在以上两个数据集可能属于不同类别，如果强行合并两个数据集，那么就会出现同样目标属于不同类别的情况，会在网络训练过程中产生误导。因此虽然他们的场景相似，目标相似，但是仍然无法合并训练。而双分类头能够很好的解决这一问题，假设检测网络分别单

独在类别定义相似的两个数据集上训练，则这两个检测模型分类分支中区分前景与背景以及区分不同类别所需要的分类特征也应该相似。如果能设计合理的结构让模型能够同时对两个数据集进行训练，那么不同的数据集中的数据会将分类特征往相同的方向优化，则相似数据集就可以提升模型的精度。基于上述假设，本文提出了一种全新的双分类头，不同分类头针对不同数据集展开训练，使得类别定义不同的相似数据全程参与模型训练。图3给出了现有的普通检测头和本文提出的双分类检测头示意图。

双分类头在分类分支的最后一层特征后面额外添加一个全连接层（class2）以执行对参考数据集的分类任务，添加的全连接层与原始的分类全连接层共享分类特征。在训练阶段，通过类别编号过滤类别预测和真值，将属于原始训练集的类别预测和相应的真值单独计算一个分类损失，将属于参考数据集获得的类别预测与相应的真值计算另一个分类损失，两个分类损失之和为最终的分

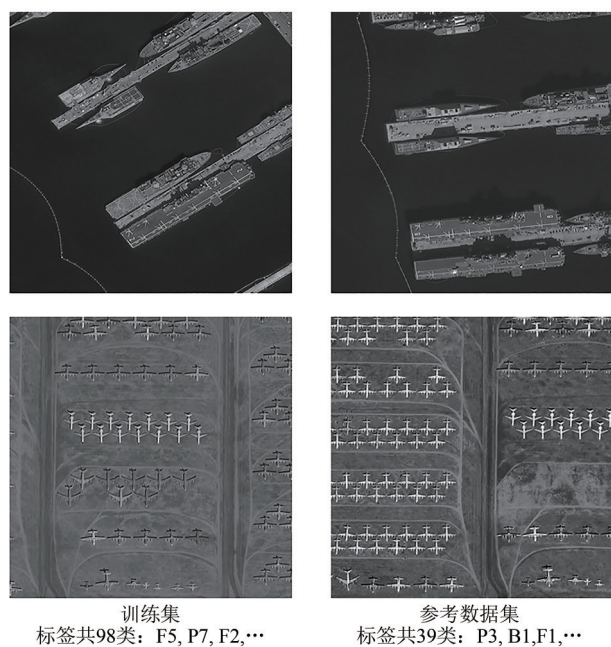


图2 精细化目标检测竞赛数据集样例图

Fig. 2 Sample image of the fine-grained object detection competition dataset

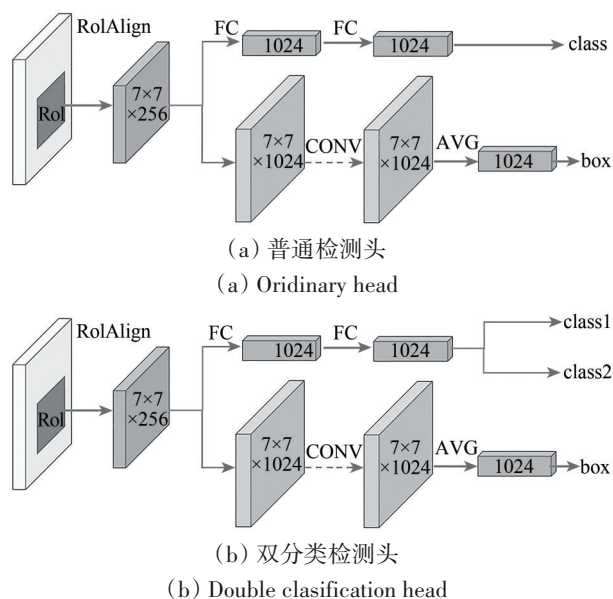


图3 双分类检测头和普通检测头比较示意图

Fig. 3 The comparisons the dual classification detection head proposed in this paper with the ordinary detection head

在推理阶段, 直接忽略额外添加的全链接层, 只输出原始分类头的类别预测。这种方式能够充分利用相似数据的训练模型, 对于原检测网络而言, 参考数据参与了除最后一层全连接层之外的所有层的训练, 相似数据获得了充分的利用, 有效缓解了数据不足的问题, 也将提高模型的精度。该方法可以用于任何相似数据集的共同训练, 而且对于目标分类、目标检测、实例分割、语义分

割等任务同样适用, 有望成为相似数据利用的通用方法。由于该方法在推理阶段直接忽略了额外添加的分类头, 因此对于模型的推理速度没有任何影响。

3.3 错误标签过滤

大规模高精度的人工标注经常会出现一些标注错误。检测网络为了拟合错误标签, 会将网络参数朝相反的方向优化, 这将极大影响网络的整体性能, 因此需要减小错误标签对网络训练的影响。通常情况下, 经过训练的网络能够正确预测目标的类别且获得较高的置信度。本文以此为依据设计了一种基于预测的错误标签过滤模块。该模块在网络的训练阶段计算分类损失前使用, 主要为过滤实际标注遥感图像过程中出现的少量的类别混淆的标注。训练中先对推理获得的类别分数通过额外的 SoftMax 操作来获得该目标的类别和置信度, 如果预测类别与真值标签不同且置信度高于某一阈值, 则认为该标签是错误标签, 本文将过滤错误标签后的相应预测值和真值用于计算分类损失。

标签过滤过程举例如下: 如图4所示, 图4中真值为第0类, 预测为第1类, 且置信度为0.9 (假设给定的阈值为0.88), 0.9高于给定过滤阈值, 则认为该标签为错误标签。通过设置不同的过滤阈值来控制过滤目标的比例, 该方法可以过滤大部分错误标签, 进而降低错误标签对模型训练的影响。同时该方法将不会影响网络在初始阶段的训练, 因为在该阶段, 网络的分类分支从随机初始化开始训练, 此时网络远未收敛, 经过 SoftMax 操作后, 预测的类别置信度通常低于过滤模块设置的置信度阈值。

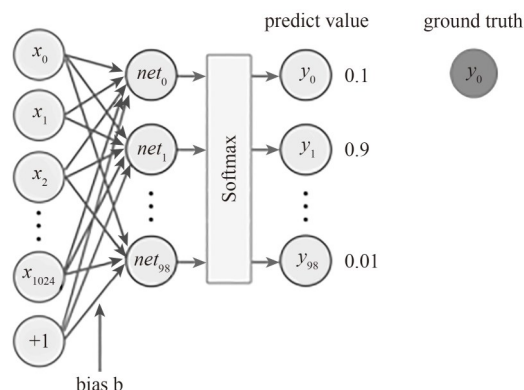


图4 错误标签过滤示意图

Fig.4 Diagram of error label filtering

该方法具有较强的通用性，能够用于目标分类、目标检测以及实例分割等任务中的错误标签的过滤，从而提高模型精度。由于该方法只是在网络预测后额外添加一个SoftMax操作，因此对模型训练速度几乎没有影响，并且完全不影响模型推理速度。

3.4 Margin 交叉熵损失

在通用目标检测任务中，算法更加关注回归分支的损失函数设计，而忽视了分类分支损失函数。由于缺少精细目标检测需求的牵引，从Faster RCNN (Ren 等, 2015) 检测网络到最近的DINO (Zhang 等, 2022) 检测网络，几乎都使用交叉熵损失作为分类分支的损失函数。然而，精细目标检测中目标类别多、类间差异小，类内差异大，传统的交叉熵损失已不能满足精细检测的需求。借鉴SVM思想，通过增大不同类别的分类边界来提高检测精度。参考人脸识别中ArcFace (Deng 等, 2019) 损失函数加上角度Margin，以加法的方式惩罚深度特征与其相应权重之间的角度，从而同时增强了类内紧度和类间差异。据此，本文在训练中通过人为将正样本的网络预测值减去常数项后再计算交叉熵损失，进而提出了Margin交叉熵损失函数，普通的交叉熵损失计算公式如下所示：

$$L = \frac{1}{N} \sum_i L_i = \frac{1}{N} \sum_i -\log \left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}} \right) = -\sum_{i=1}^m \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}} \quad (1)$$

式中， $x_i \in \mathbf{R}^d$ ，表示属于第 y_i 类的深度特征， W 和 b 为最后一层全连接层的权重和偏置， m 和 n 分别表示类别数量和一个训练批次中该类别目标的数量。Margin交叉熵损失函数的计算公式为

$$L = -\frac{1}{N} \sum_{i=1}^m \log \frac{e^Y}{\sum_{j=1}^n e^{W_j^T x_i + b_j}} \quad (2)$$

$$Y = \begin{cases} W_{y_i}^T x_i + b_{y_i} & \text{if } x_i \notin \text{POS} \\ W_{y_i}^T x_i + b_{y_i} - \Delta & \text{if } x_i \in \text{POS} \end{cases} \quad (3)$$

式中，POS表示该样本为正样本， Δ 为常数（最佳 Δ 通过验证集的结果来选择）。Margin交叉熵损失增大分类边界示意图如图5所示。绿色和橘黄色分

别代表两种分类，虚线为分类决策边界，通过加入Margin（即 Δ ）后，类别的分类边界被人为拉大，这将有利于提升对子类的鉴别能力。这种方式增大了不同类别之间的分类边界，使得模型具有更好的精细分类能力。由于该方法只改进分类分支，因此可以用于其他需要精细分类的任务，包括水平框检测和实例分割等任务，具有较强通用性。同时该方法只是改进了训练阶段的损失函数，因此也完全不影响模型推理速度。

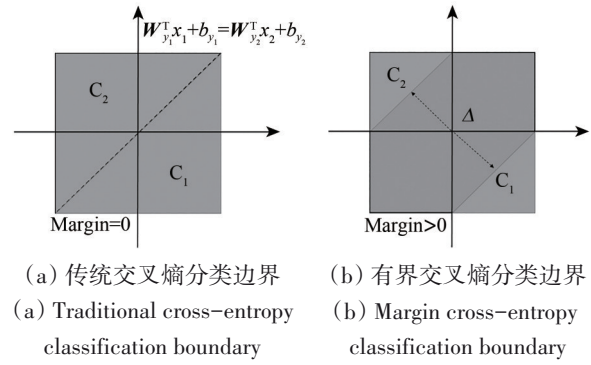


图5 传统交叉熵与Margin交叉熵分类边界示意图
Fig. 5 Schematic diagram of classification boundary between traditional cross-entropy and Margin cross-entropy

4 实验结果与分析

4.1 实验数据集介绍

精细化目标检测竞赛数据集为2023全国大数据与计算智能挑战赛中基于亚米级影像的精细化目标检测赛道发布的数据集。数据来源为四维高景卫星可见光载荷融合图像（空间分辨率0.5 m）。该数据集包含训练集、测试集及参考数据三部分。其中训练集、测试集包含98类飞机和舰船目标，分别有8000张和3000张影像；参考数据集包含39类粗粒度目标，有11000张影像，所有图像的尺寸均为1024×1024，目标位置由4个角点的8个坐标值组成，其影像与目标示例如图2所示。

FAIR1M (Sun 等, 2022) 是遥感图像精细目标检测和识别的大规模数据集。该数据集有37个精细目标类别。例如将飞机细分到具体型号波音737、波音747、波音777、波音787、空客220、空客321、空客330、空客350、C919和ARJ21，将船舶细分为液货船、干货船、渔船、邮轮、拖船和工程船等。该数据集包含有100多万个精细化

标注、多角度分布的目标, 包含 15000 张遥感影像, 每幅影像的大小在 1000×1000 到 10000×10000 像素范围内, 空间分辨率为 $0.3\text{--}0.8\text{ m}$, 实验中本文使用该数据集的 2.0 版本, 分别使用训练集和验证集来训练模型和验证算法的精度, 并使用 DOTA 数据集作为参考数据集。

DOTA (Xia 等, 2018) 是大型航拍图像目标检测的数据集, 包含飞机, 轮船, 储罐, 棒球场, 网球场, 篮球场, 地面跑道, 港口, 桥梁, 大型车辆, 小型车辆, 直升机, 环形交叉路口, 足球场和篮球场共 15 种常见目标类别。数据集包含 2806 幅卫星或者航拍影像, 每张影像的大小在 800×800 到 4000×4000 像素范围内, 包含 18.8 万个不同尺度、方向和形状的物体, 本文将 DOTA 作为 FAIR1M 的参考数据。

4.2 实验评价指标及参数设置

本次实验利用平均精度 (mAP) 在 IoU 阈值为 0.5 的情况下衡量检测性能。当目标的边界框与真值的交并比大于阈值 IoU 且预测的类标签正确时, 认为检测结果真阳性 (TP)。

实验环境包括: Intel i7 13700K CPU, NVIDIA GeForce RTX3090Ti GPU. Ubuntu20.04 系统, 本文代码基于 MMRotate 工具箱完成。

所有实验中, 均使用随机梯度下降 (SGD) 优化器, 学习率 (Learning rate) 设置为 0.001, 动量 (momentum) 为 0.9, 权重衰减 (weight decay) 为 0.0001, 批次大小 (batch size) 设置为 8, 其他设置与 MMRotate 工具箱中 Oriented R-CNN (Xie 等, 2021) 网络的默认设置保持一致。在消融实验中, 为了加快实验进度, 尽快验证网络模型的效果, 本文使用较小 LSKNet-t 作为特征提取骨干网络, 模型共训练 12 轮 (epoch), 特征融合网络使用特征金字塔 (FPN), 仅使用简单的水平翻转作为数据扩充方法。在 FAIR1M (Sun 等, 2022) 数据集上使用 LSKNet-s 作为骨干网络。在性能对比实验中, 为了提高模型精度与更好的与其他方法对比, 本文使用使用 LSKNet-s 作为骨干网络, 训练设置与其他方法的性能对比实验相同, 即额外添加图像旋转的数据扩充方式, 并训练 24 轮。在推理过程中, 使用预设的 0.02 置信度阈值来过滤掉背景, 并使用阈值为 0.1 的旋转非极大值抑制 (R-NMS) 滤除多余的检测框。推理速度是在批次大小为 1 的

情况下获得的。

4.3 消融实验

4.3.1 双分类检测头实验

双分类头能够很好的解决相似数据集利用的问题, 因此能够提高模型精度。如表 1 所示, 在基线模型的基础上引入双分类检测头, 由于该设计能够充分利用参考数据集参与训练, 相当于增加了数据集的规模, 同时由于参考数据集不存在人为的类别混淆, 因此更有利于形成稳定的分类特征。所以, 网络在加入双分类头后, 在竞赛数据集上模型精度从初始的 58.4% 增加到 72.4%, 精度大幅提升 14%。

表 1 不同模型和方法在精细化目标检测竞赛数据集平均精度对比

Table 1 Comparison of average accuracy of different models and methods on the fine-grained object detection competition dataset

		/%
模型和方法	参训数据集	mAP(50)
基线模型	训练集	58.4
基线模型+双分类头	训练集+参考集	72.4
基线模型+双分类头+错误标签过滤	训练集+参考集	73.7
基线模型+双分类头+错误标签过滤+ Margin 交叉熵损失	训练集+参考集	74.7

由于 FAIR1M 数据集规模较大, 有 37 个精细化的目标类别, 包含 100 多万个精细化标注的遥感目标; 而作为参考数据集的 DOTA 数据集数据规模较小, 只有 15 类 18 万多个标注的目标。以往的方法无法将两个数据集共同训练以提升模型精度。使用本文提出的双分类头后, 能够将 DOTA 数据集作为参考数据集加入训练过程。由于数据规模和标注精细度上的差别, 理论上即使 DOTA 数据集能够加入训练, 其对于模型总体精度的提升将十分有限。如表 2 所示, 在 FAIR1M 数据集上, 在加入双分类头后, 由于能够充分利用 DOTA 数据集的信息, 本模型仍然将检测精度提升 0.4%。

4.3.2 错误标签过滤模块实验

错误标签过滤模块通过减少错误标签对训练的影响来提升模型精度。如表 1 所示, 加入错误标签过滤模块后, 在竞赛数据集上模型精度从 72.4% 继续增加到 73.7%, 增加了 1.3%。由于 FAIR1M 数

据集中并未进行人为的标签混淆，错误标签的问题不明显，因此未在该数据集上使用错误标签过滤模块。

表2 不同模型和方法在FAIR1M数据集上的平均精度对比
Table 2 Comparison of average accuracy of different models and methods on the FAIR1M dataset

模型方法	参训数据集	mAP(50) /%
基线模型	FAIR1M	40.6
基线模型+双分类头	FAIR1M + DOTA	41.0
基线模型+双分类头+ Margin 交叉熵损失	FAIR1M + DOTA	41.3

标签过滤模块根据过滤阈值的不同，被过滤的错误标签的比例也有所差别。在不同的过滤阈值下模型在竞赛数据集上的精度如表3所示，从表中可以看出不同的过滤阈值对模型的性能有较大影响，当过滤阈值取0.9时在竞赛数据集上获得了最佳精度。随着过滤阈值逐渐降低，满足过滤条件的标签逐渐增加，过滤掉的标签逐渐增加，更多的错误标签被过滤，此时模型的精度逐渐升高。当继续降低过滤阈值时，此时有一些正确的标签也被过滤，因此模型精度逐渐降低。

表3 在精细化检测竞赛数据集上过滤阈值 α 对检测网络精度的影响
Table 3 The influence of filtering threshold α on the accuracy of the model on the fine-grained object detection competition dataset

过滤阈值	模型精度 mAP(50)/%
0.86	73.2
0.87	73.8
0.88	72.9
0.90	74.7
0.95	74.1
0.98	74.1

4.3.3 Margin 交叉熵损失实验

Margin 交叉熵损失，强行让网络在训练时拉大分类决策边界，增加了模型检测精度。如表1所示，网络在使用 Margin 交叉熵损失后，在竞赛数据集上模型精度从73.7%继续增加到74.7%，增加了1%。如表2所示，在FAIR1M数据集上模型精度从41.0%继续增加到41.3%，增加了0.3%。

基线模型加上双分类头结构和标签过滤模块

后在不同的边界阈值下检测网络在竞赛数据集上精度如表4所示，从中可以看出损失函数的边界阈值 Δ 对模型精度有较大影响，当 Δ 比较小时，分类边界被拉开的距离不大，此时对模型的精度提升有限。随着 Δ 的增加，分类边界逐渐扩大，此时模型的精度逐渐提升，在精细化目标检测竞赛数据集上，当 Δ 为0.12时可以获得最佳精度。继续增加 Δ 时，模型持续拉大分类边界，此时检测精度出现下降。不同的边界阈值下检测网络的在FAIR1M数据集上精度如表5所示，加入Margin交叉熵损失后，检测精度提升了0.3%，当 Δ 为0.10时可以获得最佳精度。

表4 在精细化检测竞赛数据集上边界阈值 Δ 对检测网络精度的影响
Table 4 The influence of margin threshold Δ on the accuracy of the model on the fine-grained object detection competition dataset

边界阈值 Δ	模型精度 mAP(50)/%
0.11	73.6
0.12	74.7
0.13	74.4
0.15	73.8

表5 在FAIR1M数据集上边界阈值 Δ 对检测网络精度的影响
Table 5 The influence of Margin threshold Δ on the accuracy of the model on the FAIR1M dataset

边界阈值 Δ	模型精度 mAP(50)/%
0.09	41.1
0.10	41.3
0.11	41.1
0.12	40.7

4.4 性能对比实验

在精细化目标检测竞赛以及FAIR1M数据集上进行性能对比实验。

4.4.1 在精细化目标检测竞赛数据集上的实验结果

表6为本文方法在竞赛数据集上的精度，可见本文提出的方法检测精度显著提升，将检测精度从58.4%，提升到74.7%，共提升16.3%。本研究使用更大的骨干网络后只是略微减小了推理速度，从每秒29.9帧降低为每秒25.7帧，而使用1312×1312像素推理精度提升到76.5%，速度降为每秒

18.1 帧，模型仍然具有较好的实时性。本文方法使用较小的骨干网络和较少的训练技巧获得了 2023 全国大数据与计算智能挑战赛“基于亚米级影像的精细化目标检测”赛道的第一名，超越其他所有团队。这些团队尝试使用很多检测方法，

使用较大的骨干网络等和诸多的训练技巧。由于复赛测试集的真值标注暂未公开，因此目前无法获得其他方法在该数据集上的精度，但是通过竞赛可以看出，本文方法拥有最高的精度和较快的推理速度。

表 6 不同方法在精细化目标检测竞赛数据集上的实验结果

Table6 Experimental results of different methods on the fine-grained object detection competition dataset				
模型方法	骨干网络	图片尺寸/像素	mAP(50)/%	推理速度(FPS)
基线模型	LSKNet-t	1024	58.4	36.3
基线模型+双分类头	LSKNet-t	1024	72.4	36.3
基线模型+双分类头+错误标签过滤	LSKNet-t	1024	73.7	36.3
基线模型+双分类头+错误标签过滤+ Margin 交叉熵损失	LSKNet-t	1024	74.7	36.3
基线模型+双分类头+错误标签过滤+ Margin 交叉熵损失	LSKNet-s	1024	75.7	36.2
基线模型+双分类头+错误标签过滤+ Margin 交叉熵损失	LSKNet-s	1312	76.5	18.1

本文方法在竞赛数据集上的可视化检测结果如图 6 所示，图中检测框的不同颜色代表不同的类别。可见本文提出的方法能稳定的提升模型精度，

在竞赛数据集上取得最佳结果，这充分说明本文方法的有效性。无论是对微小的飞机还是密集分布的舰船，本文方法都能实现精确检测。

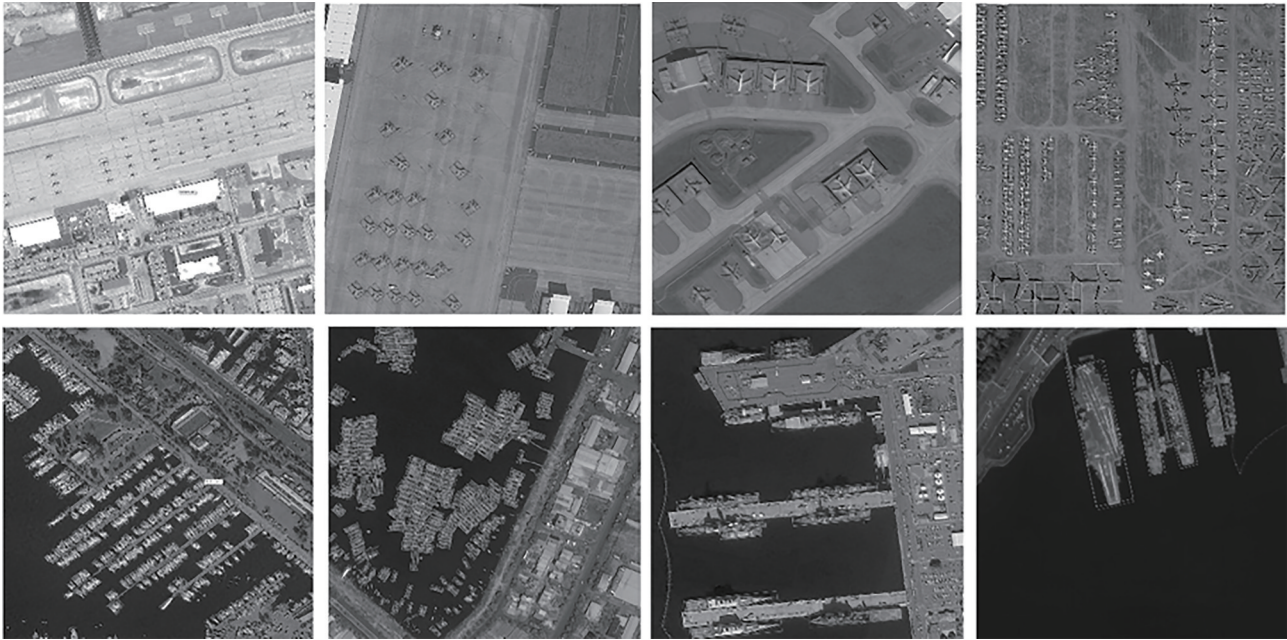


图 6 本文方法在精细化目标检测竞赛数据集上的可视化检测结果
Fig.6 Sample object detection results of our method on the fine-grained object detection competition dataset

4.4.2 在 FAIR1M 数据集上的实验结果

表 7 为本文方法在 FAIR1M 数据集上与其他旋转目标检测器方法的比较结果。可见其他方法的精度和速度为使用 MMRotate 平台中基于 DOTA 数据集的默认设置、在 FAIR1M 训练集上训练并在验证集上测试获得的结果。因显存占用大，STD

(Yu 等，2024) 只能使用尺寸为 800×800 像素的图像训练。由表 7 可见本文方法同时获得了 41.3% 的次高精度和每秒 29.1 帧的最高推理速度。由于对比的其他算法只能以在 DOTA 数据集上的预训练模型初始化训练参数，而本文算法能直接使用 DOTA 数据参与训练，因此本文算法有效提升模型精度。

表7 不同方法在FAIR1M数据集上的实验结果
Table 7 Experimental results of different methods on the FAIR1M dataset

模型方法	骨干网络	图片尺寸/像素	mAP(50)/%	推理速度(FPS)
S ² Anet(Han 等,2021b)	ResNet50	1024	29.8	21.1
R ³ Det(Yang 等,2019)	ResNet50	1024	16.2	23.6
ROI Transformer(Ding 等,2018)	ResNet50	1024	33.2	23.6
Gliding Vertex (Xu 等,2021)	ResNet50	1024	30.4	26.7
ReDet(Han 等,2021a)	Re-ResNet50	1024	34.8	8.8
Oriented RCNN(Xie 等,2021)	ResNet50	1024	36.1	25.8
STD(Yu 等,2024)	LSKNet-s	800	41.7	2.2
LSKNet(Li 等,2023)	LSKNet-s	1024	40.6	29.1
本文方法	LSKNet-s	1024	41.3	29.1

在FAIR1M数据集上的可视化检测结果如图7所示。图中检测框的不同颜色代表不同的目标类别，可见无论数据集中的各种目标还是对各类道

路设施和各类场地，本文方法都能实现精确检测，取得较好的检测结果，这充分说明本文方法的有效性。

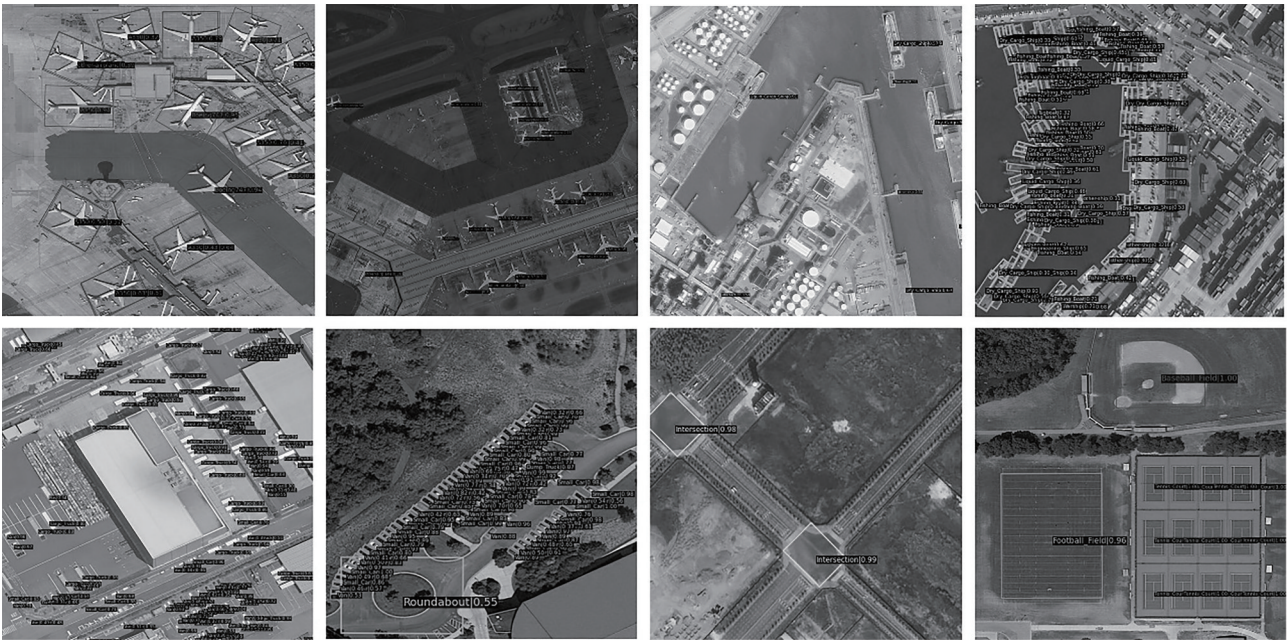


图7 本文方法在竞赛数据集上的检测结果
Fig. 7 Detection results of this method on the competition dataset

5 结 论

针对遥感图像精细化目标检测存在的问题，本文提出了双分类头结构，并对本文提出方法与已有模型和方法进行对比和验证。主要结论如下：

- (1) 本文提出的双分类检测头，通过不同的分支分别训练不同的数据集，解决了遥感图像精细化目标检测中无法有效利用相似数据的问题；
- (2) 基于预测的错误标签过滤方法，能减小

- 了错误标签对模型训练的影响；
 - (3) Margin 交叉熵损失，通过增大分类边界提高了模型精度；
 - (4) 通过对精细化目标检测竞赛数据集、FAIR1M数据集和DOTA数据集的实验，证明本文提出的方法显著提高了模型的检测精度和鲁棒性。
- 同时本文提出的双分类头也具有一定的局限性，对于数据规模相对较小的训练集，在加入相似数据集共同训练后有较大的性能提升。但对于

数据规模本身较大的训练集,即使加入相似的数据共同训练,性能的提升并不明显(在FAIR1M数据集只提升0.4%)。此外本文提出的错误标签过滤模块仅能针对标签类别混淆的情况,无法过滤位置错误的标注。在未来的研究工作中,我们将继续研究更多的相似数据的复用方法和过滤位置错误标注的方法,进一步提升遥感图像精细化目标检测的精度。

志 谢 本文算法是作者为参与全国大数据与计算智能挑战赛“基于亚米级影像的精细化目标检测”赛道而提出的,实验数据均为赛事组织方——大数据与决策(国家级)实验室提供,在此表示衷心的感谢!

参考文献(References)

- Deng J K, Guo J, Xue N N and Zafeiriou S. 2019. ArcFace: additive angular margin loss for deep face recognition//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE: 4685-4694 [DOI: 10.1109/CVPR.2019.00482]
- Ding J, Xue N, Long Y, Xia G S and Lu Q K. 2018. Learning RoI transformer for oriented object detection in aerial images//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE: 2844-2853 [DOI: 10.1109/CVPR.2019.00296]
- Han B, Yao Q M, Yu X R, Niu G, Xu M, Hu W H, Tsang I W and Sugiyama M. 2018. Co-teaching: robust training of deep neural networks with extremely noisy labels//Proceedings of the Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: Curran Associates Inc.: 8536-8546 [DOI: 10.5555/3327757.3327944]
- Han J M, Ding J, Xue N and Xia G S. 2021a. ReDet: a rotation-equivariant detector for aerial object detection//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 2785-2794 [DOI: 10.1109/CVPR46437.2021.00281]
- Han J M, Ding J, Li J and Xia G S. 2021b. Align deep features for oriented object detection. IEEE Transactions on Geoscience and Remote Sensing, 60: 5602511 [DOI: 10.1109/TGRS.2021.3062048]
- Han Y Q, Yang X Y, Pu T and Peng Z M. 2022. Fine-grained recognition for oriented ship against complex scenes in optical remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 60: 5612318 [DOI: 10.1109/TGRS.2021.3123666]
- Hu Z B, Gao K, Zhang X D, Wang J W, Wang H and Han J W. 2022. Probability differential-based class label noise purification for object detection in aerial images. IEEE Geoscience and Remote Sensing Letters, 19: 6509705 [DOI: 10.1109/LGRS.2022.3172983]
- Jiang Y Y, Zhu X Y, Wang X B, Yang S L, Li W, Wang H, Fu P and Luo Z B. 2018. R2CNN: rotational region CNN for arbitrarily-oriented scene text detection//Proceedings of the 24th International Conference on Pattern Recognition. Beijing: IEEE: 3610-3615 [DOI: 10.1109/ICPR.2018.8545598]
- Li D R, Wang M, Shen X and Dong Z P. 2017. From earth observation satellite to earth observation brain. Geomatics and Information Science of Wuhan University, 42(2): 143-149 (李德仁, 王密, 沈欣, 董志鹏. 2017. 从对地观测卫星到对地观测脑. 武汉大学学报(信息科学版), 42(2): 143-149) [DOI: 10.13203/j.whugis.20160526]
- 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. 2022. Oriented RepPoints for aerial object detection//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE: 1819-1828 [DOI: 10.1109/CVPR52688.2022.00187]
- Li Y X, Hou Q B, Zheng Z H, Chen M M, Yang J and Li X. 2023. Large selective kernel network for remote sensing object detection//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE: 16748-16759 [DOI: 10.1109/ICCV51070.2023.01540]
- Luo H L and Chen H K. 2020. Survey of object detection based on deep learning. Acta Electronica Sinica, 48(6): 1230-1239 (罗会兰, 陈鸿坤. 2020. 基于深度学习的目标检测研究综述. 电子学报, 48(6): 1230-1239) [DOI: 10.3969/j.issn.0372-2112.2020.06.026]
- Ma J Q, Shao W Y, Ye H, Wang L, Wang H, Zheng Y B and Xue X Y. 2018. Arbitrary-oriented scene text detection via rotation proposals. IEEE Transactions on Multimedia, 20(11): 3111-3122 [DOI: 10.1109/TMM.2018.2818020]
- Pu Y F, Wang Y R, Xia Z F, Han Y Z, Wang Y L, Gan W H, Wang Z D, Song S J and Huang G. 2023. Adaptive rotated convolution for rotated object detection//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE: 6566-6577 [DOI: 10.1109/ICCV51070.2023.00606]
- 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 29th International Conference on Neural Information Processing Systems. Montreal: MIT Press: 91-99 [DOI: 10.5555/2969239.2969250]
- Sun X, Wang P J, Yan Z Y, Xu F, Wang R P, Diao W H, Chen J, Li J H, Feng Y C, Xu T, Weinmann M, Hinz S, Wang C and Fu K. 2022. FAIR1m: a benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. ISPRS Journal of Pho-

- togrammetry and Remote Sensing, 184: 116-130 [DOI: 10.1016/j.isprsjprs.2021.12.004]
- Wang Y S, Ma X J, Chen Z Y, Luo Y, Yi J F and Bailey J. 2019. Symmetric cross entropy for robust learning with noisy labels//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE: 322-330 [DOI: 10.1109/ICCV.2019.00041]
- 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 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 3974-3983 [DOI: 10.1109/CVPR.2018.00418]
- Xiao J S, Zhang S H, Chen Y H, Wang Y F and Yang L H. 2022. Remote sensing image object detection based on bidirectional feature fusion and feature selection. *Acta Electronica Sinica*, 50(2): 267-272 (肖进胜, 张舒豪, 陈云华, 王元方, 杨力衡). 2022. 双向特征融合与特征选择的遥感影像目标检测. *电子学报*, 50(2): 267-272 [DOI: 10.12263/DZXB.20210354]
- Xie X X, Cheng G, Wang J B, Yao X W and Han J W. 2021. Oriented R-CNN for object detection//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE: 3500-3509 [DOI: 10.1109/ICCV48922.2021.00350]
- Xu Y C, Fu M T, Wang Q M, Wang Y K, Chen K, Xia G S and Bai X. 2021. Gliding vertex on the horizontal bounding box for multi-oriented object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(4): 1452-1459 [DOI: 10.1109/TPAMI.2020.2974745]
- Yang X, Yan J C, Feng Z M and He T. 2019. R³Det: refined single-stage detector with feature refinement for rotating object//Proceedings of the 39th AAAI Conference on Artificial Intelligence. Washington: AAAI Press: 3163-3171 [DOI: 10.1609/aaai.v35i4.16426]
- Yao Y Q, Cheng G, Xie X X and Han J W. 2021. Optical remote sensing image object detection based on multi-resolution feature fusion. *National Remote Sensing Bulletin*, 25(5): 1124-1137 (姚艳清, 程塔, 谢星星, 韩军伟). 2021. 多分辨率特征融合的光学遥感图像目标检测. *遥感学报*, 25(5): 1124-1137 [DOI: 10.11834/jrs.20210505]
- Yi J R, Wu P X, Liu B, Huang Q Y, Qu H and Metaxas D. 2021. Oriented object detection in aerial images with box boundary-aware vectors//Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE: 2149-2158 [DOI: 10.1109/WACV48630.2021.00220]
- Yu H T, Tian Y J, Ye Q X and Liu Y F. 2024. Spatial transform decoupling for oriented object detection//Proceedings of the 39th AAAI Conference on Artificial Intelligence. Vancouver: AAAI Press: 6782-6790 [DOI: 10.1609/aaai.v38i7.28502]
- Yu X R, Han B, Yao J C, Niu G, Tsang I and Sugiyama M. 2019. How does disagreement help generalization against label corruption, Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR: 7164-7173 10.48550/arXiv.1901.04215
- Zhang F, Wang X Y, Zhou S L and Wang Y Q. 2021. DARDet: a dense anchor-free rotated object detector in aerial images. *IEEE Geoscience and Remote Sensing Letters*, 19: 8024305 [DOI: 10.1109/LGRS.2021.3122924]
- Zhang H, Li F, Liu S L, Zhang L, Su H, Zhu J, Ni L M and Shum H Y. 2023. DINO: DETR with improved denoising anchor boxes for end-to-end object detection Proceedings of the International Conference on Learning Representations(ICLR) Kigali City [DOI: 10.48550/arXiv.2203.03605] 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]
- Zhang S, Li S S, Wei G F, Zhang X N and Gao J W. 2022. Refined multi-scale feature-oriented object detection of the remote sensing images. *National Remote Sensing Bulletin*, 26(12): 2616-2628 (张省, 李山山, 魏国芳, 张新耐, 高建威). 2022. 面向精细化多尺度特征的遥感图像目标检测. *遥感学报*, 26(12): 2616-2628 [DOI: 10.11834/jrs.20221801]
- Zhang Z L and Sabuncu M R. 2018. Generalized cross entropy loss for training deep neural networks with noisy labels//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: Curran Associates Inc.: 8792-8802 [DOI: 10.5555/3327546.3327555]
- Zhao P B, Qu Z S, Bu Y J, Tan W M and Guan Q Y. 2021. PolarDet: a fast, more precise detector for rotated target in aerial images. *International Journal of Remote Sensing*, 42(15): 5831-5861 [DOI: 10.1080/01431161.2021.1931535]
- Zhao Z Y, Xue Q L, He Y H, Bai Y F, Wei X and Gong Y H. 2024. Projecting points to axes: oriented object detection via point-axis representation//Proceedings of the 18th European Conference on Computer Vision - ECCV 2024. Milan: Springer: 161-179 [DOI: 10.1007/978-3-031-73390-1_10]
- Zhou X Y, Wang D Q and Krähenbühl P. 2019. Objects as points. *arXiv preprint arXiv: 1904.07850* [doi: 10.48550/arXiv.1904.07850]

A fine-grained object detection method for remote sensing images based on dual classification head

ZHANG Feng¹, TENG Shuhua^{1,2}, HAN Xing³, WANG Yingqian⁴, WANG Xueying⁴

1. School of Electronic Information, Hunan First Normal University, Changsha 410205, China;

2. College of Artificial Intelligence and Software, Nanning University, Nanning 530299, China;

3. Hunan Normal University of Medicine, Changsha 410081, China;

4. College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China

Abstract: With the availability of remote sensing images improved, fine-grained object detection in remote sensing images has become an important research topic in the field of remote sensing and computer vision, and a large number of remote sensing image object detection datasets have been published. However, the current fine-grained object detection method cannot effectively use these publicized datasets, although they have similar image scenes and similar categories with the fine-grained object detection datasets, which will greatly affect the accuracy of the detectors. To solve the problem of insufficient utilization of similar data, incorrect labels affecting model accuracy, and difficulty in distinguishing similar categories, a fine-grained object detection method based on a dual classification head is proposed in this paper. Firstly, to solve the problem of similar data not being able to be used effectively in fine-grained object detection in remote sensing images, a dual classified detection head is proposed in this paper. This head uses different classified branches to train different datasets and allow similar data with varying definitions of category to participate in the training, which can effectively use similar data and significantly improve the accuracy of the model. Secondly, to solve the problem of noise labels in training data in fine-grained object detection, an error label filtering module based on prediction is proposed to reduce the impact of error labels. The category and confidence score of the object were obtained by an additional SoftMax operation in training. If the predicted category is different from the ground truth label and the confidence score is higher than a certain threshold, the label is considered to be incorrect. Finally, to solve the problem of large intra-class distance and small inter-class distance in fine-grained object detection in high-resolution remote sensing images, a margin cross-entropy loss function is designed to increase the classification boundary. The detection accuracy was improved by increasing the classification boundary of different categories. The Margin cross-entropy loss function calculates the loss by artificially subtracting the constant term from the network-predicted value of the positive sample and the detection accuracy was improved by increasing the classification boundary of different categories. Experiments on remote sensing fine-grained object detection datasets and FAIR1M datasets show that the proposed method improves the accuracy and robustness of fine-grained object detectors significantly in remote sensing images. The code is open source at <https://github.com/zf020114/DCH>.

Key words: remote sensing, deep learning, fine-grained object detection, rotated object detection, dual classification heads, error label filtering module, Margin cross-entropy loss

Supported by National Natural Science Foundation of China (No. 62207030); Hunan Provincial Department of Education Key Scientific Research Project (No. 22A0640)