

基于对比学习的遥感图像小样本目标检测算法

李宗霖, 符颖

(成都信息工程大学 计算机学院, 四川 成都 610225)

摘要: 针对当前遥感图像目标检测算法依赖于大规模标注数据集才能发挥良好性能、检测新类时准确率低的问题,在经典的检测模型 Faster-RCNN 基础上提出一种新的遥感图像小样本目标检测模型。首先,通过引入元训练方法有效缓解了不平衡数据集上的过拟合现象;其次,在超空间中利用对比学习辅助模型更好地编码新类目标和基类目标,从而更准确地进行目标分类,在提升对新类样本检测效果的同时,减缓由于加入新类所产生的灾难性遗忘。此外,提出一种适用于小样本目标检测场景下的非极大抑制算法,有效解决了同一目标的重复检测问题。实验结果表明,在大规模基准数据集 DIOR 上,该方法相比次优网络 MM-RCNN,新类检测的准确率提升了 3.5%,基类检测的准确率提升了 8.7%,所有类别检测的准确率提升了 6.4%。同时,在 NWPU VHR-10 数据集上进一步证明了该方法的有效性。

关键词: 目标检测;对比学习;非极大抑制;小样本;遥感图像

DOI: 10.11907/rjtk.241162

中图分类号: TP391.41

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2025)006-0151-09



Few-Shot Object Detection Algorithm in Remote Sensing Images Based on Contrastive Learning

LI Zonglin, FU Ying

(College of Computer, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: Aiming at the problem that the current remote sensing image object detection algorithm relies on large-scale labeled datasets to perform well, and the accuracy rate is low when detecting novel classes, a new remote sensing image few-shot object detection model is proposed based on the classic detection model Faster-RCNN. First, the model effectively alleviates overfitting on balanced sub-datasets by introducing meta-training methods. Second, it utilizes contrastive learning in the hyperspace to assist the model in better encoding novel class objects and base class objects, thereby achieving more accurate object classification. This approach improves the detection performance on novel class objects while mitigating the catastrophic forgetting caused by the introduction of novel classes. In addition, a non-maximal suppression algorithm suitable for few-shot object detection scenario is proposed to effectively solve the repeated detection of the same object. Experimental results on the large-scale benchmark dataset DIOR show that, compared with the suboptimal network MM-RCNN, the proposed method improves the accuracy of new class by 3.5%, base class by 8.7%, and all classes by 6.4%. At the same time, the effectiveness of the method is further proved by NWPU VHR-10 dataset.

Key Words: object detection; contrastive learning; non-maximal suppression; few-shot; remote sensing image

0 引言

随着卫星技术的发展,遥感图像进入了海量数据时

代,依靠计算机辅助解译遥感图像能够高效地进行地表资源管理、城市规划、灾害风险评估、环境监测和生态研究。在遥感图像解译中,目标检测是一项不可或缺的任务^[1]。然而,即使是最先进的目标检测器,也需要基于大量标注

收稿日期:2024-02-24

扫描二维码阅读全文:



基金项目:四川省科技厅重大科技专项(2019ZDZX0005)

作者简介: 李宗霖(1999-),男,成都信息工程大学计算机学院硕士研究生,研究方向为计算机视觉、图像处理;符颖(1988-),女,博士,成都信息工程大学计算机学院副教授、硕士生导师,研究方向为信号与图像处理、非参贝叶斯方法在图像复原中的应用。本文通讯作者:符颖。

样本进行模型训练,才能达到令人满意的效果。尽管目前已发布了一系列公开的遥感数据集,但与自然场景数据集相比,图像数量和目标类别仍然较少。在标注样本不充足的情况下,一方面会使深度学习模型出现过拟合问题,另一方面在检测不存在的标注类别时需要耗费大量成本对新类图像进行标注。

为了解决这些问题,小样本目标检测成为一种备受关注的解决方案^[2-3]。其核心思想是只利用少量标注的新类图像即够实现对新类目标的检测,因此可带来巨大的便利性,且具有较强的实用性。



Fig. 1 Different classes of objects in remote sensing images

图1 遥感图像中不同类别的目标

此外,小样本深度学习模型在检测新类时,普遍会出现定位不准确、置信分数低的情况,导致现有的非极大抑制(Non-Maximum Suppression, NMS)算法不能很好地解决同一目标重复检测的问题。本文提出一种适用于小样本目标检测场景的NMS优化算法,以减少同一目标的重复检测框。

1 相关工作

小样本学习的目的是学习到可以迁移的知识,从而推广到新类别,只需要少量新类标注图像即可进行目标检测。近年来,在自然场景中的小样本目标检测算法取得了很好的效果,大致可分为基于数据增强的方法、基于迁移学习或微调的方法和基于元学习的方法。Li等^[5]通过数据增强技术来扩张小样本图像数量,以改善模型容易过拟合的问题。Wang等^[6]以Faster-RCNN为框架,提出一种基于两阶段微调的迁移学习方法。该方法第一阶段在大量基类上训练模型,优化所有参数;第二阶段冻结骨干网络,仅优化检测头的参数,从而进行新类检测。Kang等^[3]利用大量基类数据训练模型,并在单阶段检测架构中使用元特征学习器和重加权模块,以快速适应新类。特征提取器使用具有足够样本的基类训练数据,提取可泛化的元特征,重加权模块将新类的支持样本转换为一个全局向量对新类进行预测。

相较于自然图像,遥感图像的小样本目标检测工作更少。Li等^[7]提出一种基于元学习的方法,在YOLOv3架构上构建小样本目标检测模型,并开发了一个多尺度目标检

当前遥感图像面临两大挑战,即同类目标的尺度差异大和不同类目标的类间相似性小^[4]。在小样本场景下,由于训练样本不足难以充分学习不同类的特征,使该问题变得更加严重。在遥感图像中,不同类别的目标在视觉上通常难以区分,如图1所示。针对该问题,本文使用自监督对比学习来寻找不同类的类间差异性和相同类的类间相似性,从而提升遥感图像目标的分类性能。在小样本目标检测模型中添加需要检测的新类,对比学习模块能够在超空间中有效重新分隔所有类别(基类+新类),从而更大程度上避免灾难性遗忘的发生。

测框架。该模型主要包含3个模块:①通过元特征提取器从输入图像中提取元特征图;②通过特征重加权模块从支持图像中学习特定于类的重加权向量(Reweighting Vectors),并使用其重新校准元特征图;③通过边界框预测模块对重加权特征图进行目标检测。

Cheng等^[8]在Meta RCNN的基础上提出带有原型引导的RPN(Region Proposal Network),为高质量的候选提议生成提供类别信息。该方法主要由3部分组成,将支持图像转换为类别感知原型的原型学习网络(Prototype Learning Network, PLN),用于更好地生成区域建议的原型引导区域建议网络(P-G RPN),以及扩展基于Faster RCNN的检测头以进一步提高性能。该工作首先在大规模数据集DIOR上对新类和基类进行了合理的分割,为后续研究提供了一个较好的基线模型,但没有考虑加入新类时产生的灾难性遗忘问题,导致基类样本准确率下降。Zhao等^[9]在TFA(Two-stage Fine-tuning Approach)^[10]的基础上提出一种基于微调的方法(PAM-det),在路径聚合模块的辅助下学习多尺度特征,以自底向上地缩短信息传播路径,并利用底层特征的语义信息进行定位。在该工作的启发下,在设计网络时使用了特征融合机制,以改善同一类别尺度差异大导致的错检和漏检问题。Li等^[11]提出一种基于元记忆的方法(Meta-Memory-Based RCNN, MM-RCNN),采用记忆模块存储训练阶段学习到的各个类别的知识,通过基于记忆的外部注意力机制(Memory-Based External Attention, MEA),同时聚合各个类别的信息,并使用特征增强模块进一步提高性能。该框架和文献[7]同时证明了元学习在遥感领域小样本目标检测中的有效性,并提供一些额外机

制提升检测的准确率,但其没有考虑遥感图像中类与类之间的关系,很容易造成分类错误。本文通过引入对比学习分支,在减小同类目标编码差异的同时,最大化不同类目标编码的差异,有效提升了对新类的检测效果。

2 算法设计

本文方法的网络模型结构如图 2 所示。训练过程可分为两阶段:第一阶段在丰富的基类训练数据集上进行常规的监督学习,输入图像通过 Faster RCNN 网络训练一个

有效的特征提取器;第二阶段使用第一阶段训练好的特征提取器,并连接对比分支进行监督对比学习。在此阶段,使用提供的新类小样本图像和少量基类图像构成新的平衡子数据集,将该子数据集分为查询集和支持集,通过网络获取目标的 RoI 高维特征,并通过降维得到可对比的类编码向量,使用对比损失函数计算对比损失。同时,采用元训练策略在额外的对比损失帮助下反向传播优化网络参数,使其具有更好的泛化性能和更高的目标分类准确率。将在测试阶段需要检测的图像作为查询集,无须使用对比分支,可直接进行检测。

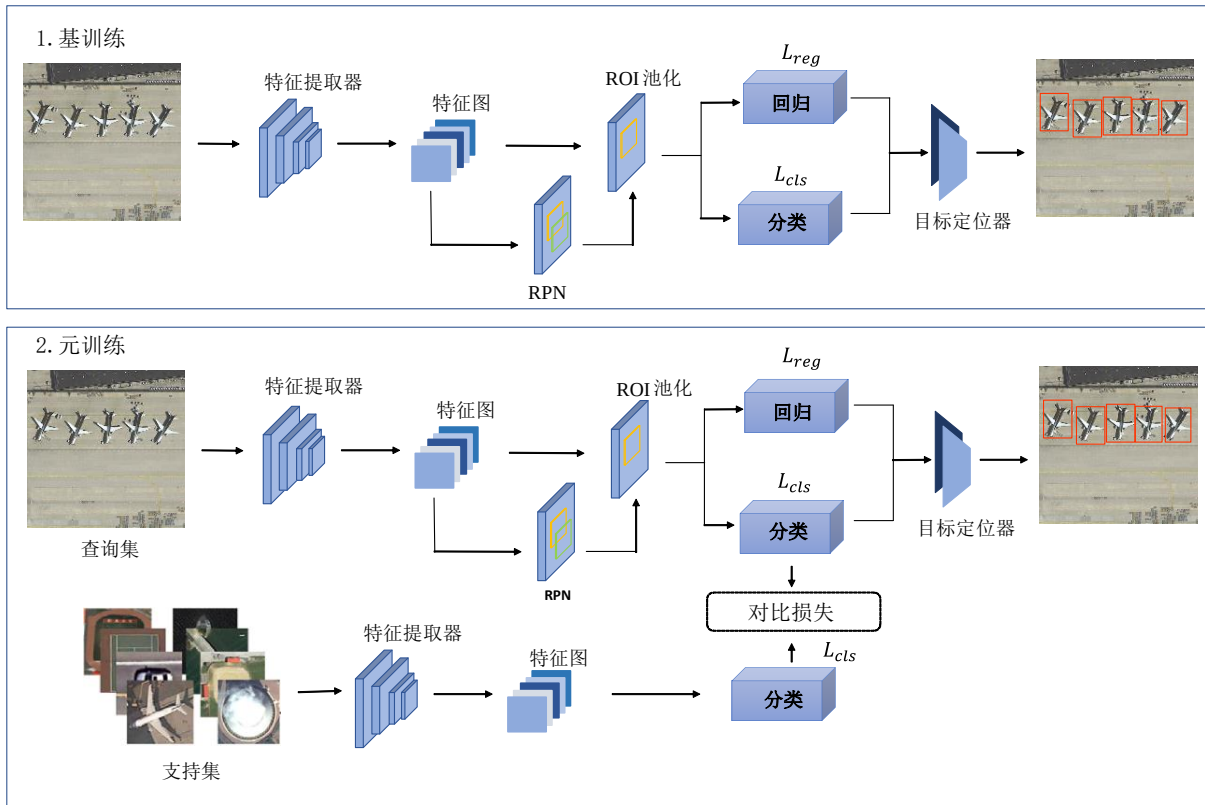


Fig. 2 Model structure of the proposed method

图2 本文方法模型结构

2.1 对比学习模块

对比学习尝试学习相似实例的共同特征,并区分不同实例之间的差异,将样本分为正样本和负样本。在缩小正样本之间距离的同时,扩大负样本之间的距离,从而更好地对不同类别的样本进行分类。

在经典的两阶段目标检测器 Faster R-CNN 中,将经过主干网络提取的特征作为 RPN 的输入,从而生成区域建议。RoI 模块将这些区域建议编码为 ROI 特征,并对目标框信息进行分类和回归。对比学习分支不需要关注具体实例中繁琐的细节,只需关注抽象语义层面的特征空间。该方式对于遥感图像这种复杂场景更加友好,如图 3 所示。使用多层感知机(Multi-Layer Perceptron, MLP)将 ROI 分类特征编码为 128 维的特征向量,以适配对比学习任务。

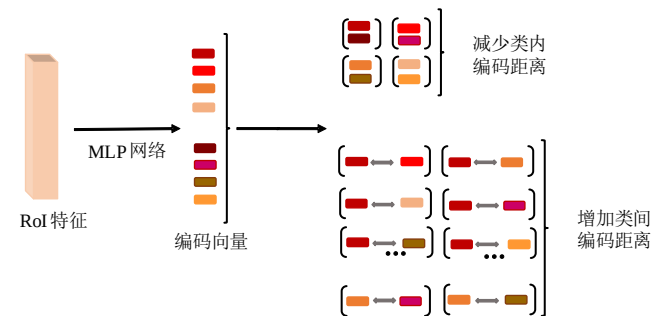


Fig. 3 The branch of supervised contrastive learning

图3 监督对比学习分支

在监督对比学习中,分别从属于同一类与不同类的图像中提取正对和负对。给定一组 N 个随机样本 $\{x_k, y_k\}_{k=1}^N$, x_k 为第 k 张图像, y_k 为第 k 张图像的标签,对其进行两次数据增强得到 $2N$ 个样本 $\{\tilde{x}_i, \tilde{y}_i\}_{i=1}^{2N}$ 。其

中, $\tilde{x}_{2k}$ 和 $\tilde{x}_{2k-1}$ 是 x_k 的两个随机增强样本, 并且 $\tilde{y}_{2k} = \tilde{y}_{2k-1} = y_{2k}$ 。设 $i \in I := \{1, \dots, 2N\}$ 为任意增广样本的索引。根据文献[12]、[13], 定义有监督对比类编码损失 L_{CCE} 如公式(1)所示。

$$L_{CCE} = \frac{-1}{|I|} \sum_{i \in I} \frac{1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(\tilde{z}_i \cdot \tilde{z}_p / \tau)}{\sum_{a \in A_i} \exp(\tilde{z}_i \cdot \tilde{z}_a / \tau)} \quad (1)$$

其中, $P(i) := \{p \in A(i) : \tilde{y}_p = \tilde{y}_i\}$ 为集合 I 中不同于 i 的所有正样本集合, $|I|$ 为 I 的基数数量, $|P(i)|$ 为 P 的基数数量, $\cdot$ 为内积运算符, $\tau \in R^+$ 为正标量的温度参数。对数函数中分子表示锚点 i 与 p 之间的距离, 分母表示负样本距离之和。监督对比分支的优点是考虑了属于同一类的不同图像之间的特征相关性。与自监督场景相比, 监督学习通常具有更均衡的正负样本分布。

2.2 元训练框架

传统的小样本目标检测方法在检测新类时通常选取少量基类标注图像和新类图像构成新的平衡子集, 但在小规模数据集上直接训练非常容易造成过拟合现象。为了充分利用大量基类数据, 将训练分为两个连续的阶段: 第一阶段在基础网络结构上利用基类数据, 获得一个强健的特征提取器; 第二阶段为元训练阶段, 将训练分为多个训练子任务, 每个子任务的数据集分为支持集和查询集, 分别送入第一阶段的特征提取器提取特征, 并联合对比损失和目标定位分类损失共同优化网络参数。

在第一阶段的训练过程中, 采用与当前普遍的目标检测相似的方法。使用 ResNet 101 构建 Faster RCNN 模型, 并在 ImageNet 上进行预训练。在一个包含丰富训练样本的基类数据集 (不包含新类样本) 上通过监督学习获取基类的图像特征和上下文信息。其损失函数可以定义为:

$$L_{FR} = L_{RPN} + L_{RCNN} \quad (2)$$

损失函数 L_{RPN} 的定义如公式(3)所示。

$$L_{RPN} = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (3)$$

其中, p_i 表示第 i 个 anchor 预测为真实标签的概率, p_i^* 为正样本时值为 1, 负样本时值为 0; t_i 表示预测第 i 个 anchor 的边界框回归参数, t_i^* 表示第 i 个 anchor 对应的真实框; N_{cls} 表示一个 mini-batch 中的所有样本数量, N_{reg} 表示 anchor 位置的个数。RPN 分类结果只有两个类别: 背

景和前景, 并使用二分类交叉熵损失。边界框回归损失使用 *smoth L1* 损失。

损失函数 L_{RCNN} 的定义如公式(4)所示。

$$L_{RCNN} = L_{cls}(p, u) + \lambda[u \geq 1] L_{loc}(t^u, v) \quad (4)$$

其中, p 是分类器预测的 *softmax* 概率分布, $p = (p_0, \dots, p_k)$, u 对应目标真实类别的标签, t^u 为边界框回归器预测的回归参数 $(t_x^u, t_y^u, t_w^u, t_h^u)$, v 对应真实目标的边界框回归参数 (v_x, v_y, v_w, v_h) 。RCNN 损失函数包括两部分: 分类损失和边界框回归损失。分类损失使用 *softmax* 多分类交叉熵损失, 边界框回归损失使用 *smoth L1* 损失。

在第二阶段的元训练过程中有多个元任务, 一个元任务中包含了不同的类别组合。元训练开始之前对模型添加新类分类参数并随机进行初始化。元任务会从所有类别中选取一个类标签集。根据标签集, 采样一个支持集和查询集的样本用于训练, 通过随机水平翻转、随机裁剪和颜色抖动等图像数据增强方法进一步防止过拟合。查询集图像和支持集图像经过特征提取后, 通过全局平均池化得到 3×256 维的编码, 使用一个 MLP 网络降维为一个 128 维的特征向量。支持集向量为锚向量, 查询集向量根据标签构成锚向量的正负样本, 以方便进行对比学习, 反向优化分类参数。第二阶段的损失函数定义见公式(4), L_{CCE} 定义见公式(1)。

$$L_{MFR} = L_{RPN} + L_{RCNN} + \lambda L_{CCE} \quad (5)$$

测试阶段不同于元训练阶段, 无需引入支持集分支。

2.3 小样本 NMS 算法

基于深度学习的目标检测模型依赖于大规模数据集才能获得良好的检测准确率, 当训练样本不足时, 性能往往会出现大幅下降, 具体可体现为定位准确率不高和预测置信分数偏低。这一问题导致传统应用于大规模基准数据集的 NMS 算法并不能很好地适用于遥感图像小样本目标检测模型。

如图 4 所示, 同一目标出现了大量重复检测。在小样本目标检测模型中, 这是一个普遍现象。因此, 解决同一目标的重复检测问题具有重要的实际意义。

当前的文献研究普遍将 AP50 作为默认指标, 在该指标算法下, 优化 NMS 算法无法带来明显的性能提升, 反而可能导致性能略微下降, 因而成为一个明显却被忽视的问题。



Fig. 4 Repeated detection of the same object

图 4 同一目标重复检测

Soft-NMS对传统的NMS算法进行简单的改动,通过降低置信度分数来保留可能存在的正确检测,但仍不能解决当前问题。本文在Soft-NMS基础上提出一种新的能够有效减少小样本设置下重复检测现象的NMS算法。具体算法如下:

```
1. Input:  $B = \{b_1, \dots, b_N\}$ ,  $S = \{s_1, \dots, s_N\}$ ,  $N_i$ ,  $N_T$ 
2. begin
3.  $D \leftarrow \{\}$ 
4. while  $B \neq \text{empty}$  do
5.  $m \leftarrow \operatorname{argmax} S$ 
6.  $M \leftarrow b_m$ 
7.  $D \leftarrow D \cup M$ ;  $B \leftarrow B - M$ 
8. for  $b_i$  in  $B$  do
9. // if  $\operatorname{iou}(M, b_i) \geq N_i$  then
10. //  $B \leftarrow B - b_i$ ;  $S \leftarrow S - s_i$ 
11. //end
12.  $s_i \leftarrow s_i f(\operatorname{iou}(M, b_i))$ 
13. if  $s_i \leq s_t$  or  $\operatorname{iou}(\min\_box(M, b_i), \operatorname{iou}(M, b_i)) \geq N_T$  then
14.  $B \leftarrow B - b_i$ ;  $S \leftarrow S - s_i$ 
15. end
16. end
17. end
18. return  $D$ ,  $S$ 
```

其中, B 表示初始的检测框, S 表示对于检测框的置信分数, N_i 、 S_i 表示经典NMS算法设置的阈值, N_T 表示本文NMS算法的阈值(通常取0.5)。第9~11行为经典的NMS算法流程,本文通过第12~14行进行改进。第12行的 $f()$ 表示Soft-NMS函数,最后返回经算法过滤后的检测框和置信分数。

3 实验结果及分析

3.1 数据集与实验环境设置

为了验证本文提出算法的有效性,在DIOR数据集^[14]和NWPU VHR-10数据集^[15]上进行实验。DIOR数据集是RSIs中公开的大型目标检测数据集,与其他数据集相比,该数据集的图像和对象实例具有更大的类内变化性与类别之间的相似性,且标注数量远远超过其他数据集。因此,DIOR是最权威的用于遥感图像目标检测算法训练和评估的数据集。

DIOR数据集包含20个类别,为了使实验数据更具说服力,遵循Cheng等^[8]的小样本目标检测数据集划分设置,采用4种不同的随机新类/基类拆分方法,在DIOR数据集上进行类别的交叉验证。每种拆分方法都包括5个新类和15个基类(剩余的其他类别),详见表1。

NWPU VHR-10数据集包括来自Google Earth和部分来自Vaihingen数据集的高分辨率遥感图像,该数据集的图像大小为400×400像素,是一个广泛使用的遥感目标检

Table 1 Comparison table of categories contained in different segmentation groups of DIOR dataset

表1 DIOR数据集不同分割组所含类别对照表

分割序号	新类					基类
1	棒球场	篮球场	桥梁	烟囱	船只	其他
2	飞机	机场	高速收费站	港口	田径场	其他
3	大坝	高尔夫球场	储罐	网球场	车辆	其他
4	高速服务区	立交桥	体育场	火车站	风车	其他

测基准数据集^[16]。在实验中,遵循Li等^[11]的小样本目标检测数据集划分设置,将飞机、棒球场和羽毛球场设置为新类,其他类别设置为基类。

实验采用的硬件环境与软件环境如表2所示。

Table 2 Hardware environment and software environment of the experiment

表2 实验硬件环境与软件环境

环境	设备	型号
硬件环境	CPU	AMD Ryzen 5 3600X 6-Core Processor 4.4GHz
	GPU	NVIDIA GeForce RTX 3090 24GB
	操作系统	Linux Ubuntu 20.04
软件环境	编程语言	Python 3.8
	深度学习框架	Pytorch 1.8.0
	CUDA	11.1
	CuDNN	8.0.5

实验采用ResNet101作为骨干网络,使用随机梯度下降(Stochastic Gradient Descent, SGD)方法作为模型优化器,设置初始学习率为0.001,动量(momentum)为0.9,权重衰减(weight_decay)为0.000 1。利用L2范数进行梯度裁剪(grad_clip),梯度的最大范数设置为20。训练过程中基于迭代次数进行训练循环,最大迭代次数为36 000次。学习率采用分段常数衰减策略,即在训练过程的不同阶段使用不同的学习率。具体而言,采用阶梯式(step)调整策略,同时包括一个线性预热(warmup)阶段。其中,预热迭代次数为100,预热比例为0.001。在6 000次迭代后,学习率会进行阶梯式调整。

3.2 实验评价标准

本文采用两种评估方法,一种是当交并比(Intersection over Union, IoU)不低于0.5时的平均准确率(mean Average Precision, mAP)。mAP是目标检测领域常用的评价指标,用于衡量算法对于不同目标类别的检测精度,又称查准率。另一种是在一定假阳性(False Positive, FP)条件下计算出召回率(Recall)的平均值,又称查全率。结果均为小数,0.5代表准确率为50%。

3.3 实验结果分析

为验证本文方法的有效性,将本文方法与近年来一些优秀的前沿遥感图像小样本目标检测方法进行比较。在DIOR数据集上与文献[11]给出的不同网络在测试集上的基准分数进行比较,包括经典的两阶段目标检测方法Faster RCNN^[17]、高效的无锚检测算法YOLOX^[18]、基于微调的

方法 RepMet^[19]、被广泛使用的基于迁移学习的方法 TFA^[20]、FSCE^[3]、基于元学习的方法 Meta YOLO^[3]、Meta RCNN^[2]、拥有高质量提议网络的 P-CNN^[8]和基于元记忆的方法 MM-RCNN^[11]。比较结果如表 3、表 4 所示。其中, Split 表示不同的分割方法,选取不同的类作为新类和基类,具体类别如表 1 所示。分别选取 $shot = \{3, 5, 10, 20\}$ 进行实验,表示新类标注的样本图像分别包含 $\{3, 5, 10, 20\}$ 张。

Table 3 Comparison of detection results of different methods on DI-OR dataset (Split1 and Split2)

表 3 不同方法在 DIOR 数据集上检测结果比较 (Split1 与 Split2)							
Shot	方法	Split 1			Split 2		
		新类	基类	所有类别	新类	基类	所有类别
3	Faster RCNN	0	0.459	0.344	0	0.457	0.343
	YOLOX	0	0.465	0.349	0	0.451	0.338
	RepMet	0.113	0.478	0.387	0.052	0.496	0.385
	TFA	0.116	0.477	0.387	0	0.488	0.366
	FSCE	0.123	0.489	0.398	0.096	0.478	0.383
	Meta YOLO	0.075	<u>0.494</u>	0.389	0.048	0.485	0.376
	Meta RCNN	0.140	0.463	0.382	0.105	0.482	0.388
	P-CNN	0.180	0.470	0.398	<u>0.145</u>	0.489	0.403
	MM-RCNN	0.198	0.472	0.404	0.156	0.487	0.404
	本文方法	<u>0.182</u>	0.615	0.497	0.144	0.612	0.495
5	Faster RCNN	0	0.441	0.331	0	0.463	0.347
	YOLOX	0	0.456	0.342	0	0.488	0.366
	RepMet	0.113	<u>0.494</u>	0.399	0.056	<u>0.511</u>	0.397
	TFA	0.162	0.488	0.407	0.036	0.495	0.380
	FSCE	0.178	0.496	0.417	0.086	0.498	0.395
	Meta YOLO	0.121	0.497	0.403	0.070	0.468	0.369
	Meta RCNN	0.168	0.482	0.404	0.125	0.502	0.408
	P-CNN	<u>0.228</u>	0.484	0.420	0.149	0.491	0.406
	MM-RCNN	0.239	0.488	<u>0.426</u>	<u>0.155</u>	0.499	<u>0.413</u>
	本文方法	0.216	0.589	0.496	0.156	0.598	0.488
10	Faster RCNN	0	0.442	0.332	0.018	0.432	0.329
	YOLOX	0	0.481	0.361	0	0.501	0.376
	RepMet	0.121	0.493	0.400	0.059	0.512	0.399
	TFA	0.229	0.492	0.426	0.052	0.485	0.377
	FSCE	0.239	0.504	0.438	0.084	0.506	0.401
	Meta YOLO	0.181	0.495	0.417	0.090	0.464	0.371
	Meta RCNN	0.220	0.494	0.426	0.153	0.503	0.416
	P-CNN	0.276	<u>0.509</u>	<u>0.451</u>	0.189	<u>0.525</u>	<u>0.441</u>
	MM-RCNN	<u>0.288</u>	0.501	0.448	0.201	0.521	<u>0.441</u>
	本文方法	0.316	0.594	0.525	<u>0.195</u>	0.598	0.497
20	Faster RCNN	0.071	0.458	0.361	0.018	0.463	0.352
	YOLOX	0.081	0.471	0.374	0.035	0.478	0.367
	RepMet	0.122	0.486	0.395	0.068	0.504	0.395
	TFA	0.237	0.491	0.428	0.078	0.485	0.383
	FSCE	0.245	0.509	0.443	0.089	0.493	0.392
	Meta YOLO	0.220	0.500	0.430	0.141	0.435	0.362
	Meta RCNN	0.243	0.493	0.431	0.205	0.523	0.444
	P-CNN	0.296	0.522	0.466	0.228	0.516	0.444
	MM-RCNN	<u>0.308</u>	<u>0.529</u>	<u>0.474</u>	<u>0.239</u>	<u>0.525</u>	<u>0.454</u>
	本文方法	0.333	0.598	0.532	0.251	0.596	0.497

注:加粗表示最优值,加下划线表示次优值。

由表 3、表 4 可知,传统的目标检测器 Faster R-CNN 和 YOLOX 在小样本条件设置下容易过拟合。基于微调的小样本检测器的性能优于 Faster RCNN 和 YOLOX,表明微调

Table 4 Comparison of detection results of different methods on DI-OR dataset (Split3 and Split4)

表 4 不同方法在 DIOR 数据集上检测结果对比 (Split3 与 Split4)							
Shot	方法	Split 3			Split 4		
		新类	基类	所有类别	新类	基类	所有类别
3	Faster RCNN	0	0.510	0.383	0	0.484	0.363
	YOLOX	0	0.521	0.391	0	0.503	0.377
	RepMet	0.055	0.487	0.379	0.109	0.474	0.383
	TFA	0.111	0.519	0.417	0.094	0.490	0.391
	FSCE	0.125	<u>0.526</u>	<u>0.426</u>	0.105	0.512	0.410
	Meta YOLO	0.078	0.455	0.361	0.037	0.482	0.371
	Meta RCNN	0.122	0.495	0.402	0.116	0.504	0.407
	P-CNN	0.165	0.495	0.413	<u>0.152</u>	0.498	0.412
	MM-RCNN	0.167	0.504	0.420	0.164	<u>0.517</u>	<u>0.429</u>
	本文方法	<u>0.166</u>	0.634	0.517	0.149	0.561	0.458
5	Faster RCNN	0	0.487	0.365	0	<u>0.518</u>	0.389
	YOLOX	0	0.513	0.385	0	0.503	0.377
	RepMet	0.057	0.496	0.386	0.116	0.492	0.398
	TFA	0.138	0.524	0.428	0.096	0.496	0.396
	FSCE	0.149	<u>0.544</u>	<u>0.445</u>	0.111	0.508	0.409
	Meta YOLO	0.137	0.479	0.394	0.068	0.485	0.381
	Meta RCNN	0.154	0.500	0.414	0.136	0.512	0.418
	P-CNN	0.188	0.499	0.421	0.175	0.499	0.418
	MM-RCNN	<u>0.197</u>	0.495	0.421	<u>0.187</u>	0.507	<u>0.427</u>
	本文方法	0.227	0.651	0.545	0.189	0.567	0.473
10	Faster RCNN	0	0.491	0.368	0	0.508	0.381
	YOLOX	0	0.512	0.384	0	0.510	0.383
	RepMet	0.074	0.523	0.411	0.128	0.505	0.411
	TFA	0.146	<u>0.528</u>	0.433	0.114	0.485	0.392
	FSCE	0.171	0.517	0.431	0.136	0.497	0.407
	Meta YOLO	0.138	0.445	0.368	0.072	0.457	0.361
	Meta RCNN	0.203	0.522	0.442	0.170	<u>0.523</u>	0.435
	P-CNN	0.234	0.521	0.449	0.189	0.517	0.435
	MM-RCNN	0.250	0.524	<u>0.456</u>	<u>0.203</u>	0.521	<u>0.442</u>
	本文方法	<u>0.243</u>	0.624	0.529	0.220	0.546	0.465
20	Faster RCNN	0.112	0.485	0.392	0.024	0.452	0.345
	YOLOX	0.157	0.509	0.421	0.081	0.511	0.404
	RepMet	0.078	0.524	0.413	0.167	<u>0.535</u>	0.443
	TFA	0.189	0.524	0.440	0.128	0.502	0.401
	FSCE	0.204	<u>0.541</u>	0.457	0.142	0.515	0.422
	Meta YOLO	0.185	0.435	0.373	0.122	0.444	0.364
	Meta RCNN	0.263	0.537	0.469	0.220	0.534	0.456
	P-CNN	0.288	0.531	0.470	0.257	0.523	0.457
	MM-RCNN	<u>0.305</u>	0.536	<u>0.478</u>	<u>0.271</u>	0.531	<u>0.466</u>
	本文方法	0.316	0.621	0.545	0.280	0.539	0.474

注:加粗表示最优值,加下划线表示次优值。

策略是有效的。与基于微调的方法相比,基于元学习的检测器取得了更好的结果,证明了元学习策略能够快速适应新类。

本文方法在大部分新类情况下都取得了更好的结果,在基类检测任务中准确率提升明显,最高可达 15%。在所有类别检测任务中的准确率比次优的 MM-RCNN 平均高 6.4%,体现了该方法的有效性。从实验结果来看,对比分支不仅能够提高新类的检测准确率,而且能保证添加新类后基类的准确率,具有极强的抗灾难性遗忘的能力。

为进一步验证本文方法的有效性,在 NWPU VHR-10

数据集上与不同网络在测试集上的基准分数进行比较。除 DIOR 数据集中对比的方法外,还额外添加了基于微调的方法 PAMSDet^[21]和基于元学习的方法 FSODM^[8]。检测结果如表 5 所示。

Table 5 Comparison of detection results of different algorithms on NWPU VHR-10 dataset (AP50)

表 5 不同算法在 NWPU VHR-10 数据集上检测结果比较 (AP50)					
Shot	方法	飞机	篮球场	羽毛球场	平均
3	Faster RCNN	0	0	0	0
	YOLOX	0.03	0	0	0.01
	TFA	0.12	0.61	0.13	0.29
	FSCE	0.13	0.58	0.11	0.27
	PAMSDet	<u>0.21</u>	0.76	0.16	0.37
	Meta YOLO	0.13	0.12	0.11	0.12
	FSODM	0.15	0.57	<u>0.25</u>	0.32
	Meta RCNN	0.18	0.61	0.20	0.33
	MM-RCNN	0.23	0.81	0.24	<u>0.43</u>
	本文方法	0.15	<u>0.73</u>	0.44	0.44
	Faster RCNN	0.11	0.24	0.06	0.14
	YOLOX	0.15	0.21	0.09	0.15
5	TFA	0.51	0.78	0.19	0.49
	FSCE	0.48	0.76	0.16	0.47
	PAMSDet	0.55	<u>0.88</u>	0.20	0.55
	Meta YOLO	0.24	0.39	0.11	0.24
	FSODM	<u>0.58</u>	0.84	0.16	0.53
	Meta RCNN	0.50	0.83	0.17	0.50
	MM-RCNN	0.57	0.89	<u>0.21</u>	<u>0.56</u>
	本文方法	0.59	0.71	0.51	0.60
	Faster RCNN	0.34	0.58	0.21	0.38
	YOLOX	0.36	0.61	0.25	0.41
	TFA	0.60	0.85	0.49	0.65
	FSCE	0.57	0.82	0.46	0.62
10	PAMSDet	0.61	0.88	0.50	0.66
	Meta YOLO	0.20	0.74	0.26	0.40
	FSODM	0.60	0.88	0.48	0.65
	Meta RCNN	0.50	0.83	0.45	0.59
	MM-RCNN	<u>0.63</u>	0.90	<u>0.51</u>	<u>0.68</u>
	本文方法	0.68	<u>0.89</u>	0.60	0.72

注:加粗表示最优值,加下划线表示次优值。

由表 5 可见,传统的目标检测器 Faster RCNN 和 YOLOX 仍比所有的小样本检测器性能差。MM-RCNN 在多个项目中取得了最优值,证明了元训练方法在小样本设置下的有效性。在平均 mAP 指标上,本文方法显著优于所有对比方法,在 $shot = \{3, 5, 10\}$ 下相比次优值分别提升了 1%、4%、4%。

3.4 抗灾难性遗忘

灾难性遗忘是在小样本目标检测中,模型在学习新类别时对已学习过的类别检测性能下降的现象。灾难性遗忘是迁移学习和小样本学习中一个重要挑战,在实际应用中,人们通常希望模型能够不断地适应新任务,而不是丧失已经学到的知识。为了应对该问题,研究者们提出了一些方法,如领域自适应、元学习等,以减轻或避免灾难性遗忘,同时更好地适应新任务。因此,考虑模型的抗灾难性遗忘性能也是一项重要指标。

图 5 表示的是在 Split1 下 $shot = 10$ 时模型在添加新类前后检测的准确率,指标为 AP50。左侧柱状图表示添加新类前的检测准确率,右侧为添加后的准确率,共包含 15 个基类和 5 个新类。由图 5 可见,基类检测准确率下降最多的为储罐,下降了 26.3%;下降最少的为网球场,下降了 0%。平均准确率下降了 7%,证明了基于对比学习和元训练的方法具有不错的抗灾难性遗忘效果。

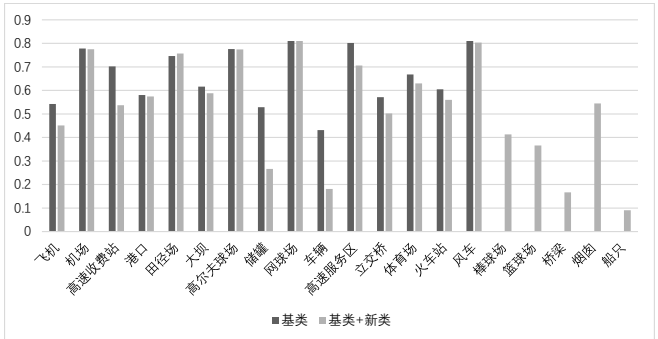


Fig. 5 Comparison of accuracy before and after adding the novel class
图 5 添加新类前后准确率对比

3.5 消融实验

为进一步分析本文算法的有效性,在 DIOR 数据集上进行两项消融实验。实验采用 Split1, $shot = 10$ 的设置,报告了 AP50 下的 mAP 的值和 $FP = 1$ 时的 mAR 值。

消融实验首先探究采用的对比分支的有效性,实验结果如表 6 所示。

Table 6 Ablation experiment results 1 on DIOR dataset
表 6 在 DIOR 数据集上消融实验结果 1

对比分支	mAP		mAR	
	新类	基类	新类	基类
√	0.14	0.51	0.19	0.55
	0.32	0.59	0.39	0.65

注:“√”表示在本文方法中采用该模块。

由表 6 可知,加入对比分支模块时,新类与基类的 mAP 值分别提升了 18% 和 8%,mAR 值分别提升了 20% 和 10%,有效提高了新类的准确率,验证了对比分支的可行性与有效性。

在第二阶段的元训练过程中需要冻结骨干网络、Neck 与 RoI 部分参数,冻结该模块是否能够帮助提升准确率需要消融实验来证明。在 Split1 分割设置下,新类样本数量分别选取 5 和 10。实验结果如表 7 所示。

由表 7 可知,在新类任务中最好的结果比最差的结果提升了 6%,在基类任务中最好的结果比最差的结果提升了 1%,说明在标注数据不足的情况下,并不能很好地驱动模型更新所有参数,冻结部分参数有助于提升新类的检测准确率。如果是大规模数据集,冻结参数则没有太大意义。

3.6 检测结果示例

图 6 为本文方法与 P-CNN 在 DIOR 数据集上的检测结果对比。每组的第一列为真实标签,第二列为本文方法检

Table 7 Ablation experiment results 2 on DIOR dataset

表 7 在 DIOR 数据集上消融实验结果 2

Shot	骨干网络	Neck	RoI	AP		AR	
				新类	基类	新类	基类
10		√	√	0.27	0.58	0.32	0.64
	√		√	0.28	0.58	0.34	0.63
	√	√		0.27	0.57	0.33	0.65
	√	√	√	0.32	0.59	0.39	0.65
		√	√	0.28	0.56	0.34	0.65
20	√		√	0.31	0.58	0.36	0.62
	√	√		0.29	0.58	0.34	0.64
	√	√	√	0.34	0.60	0.42	0.67

注：“√”表示在本文方法中采用该模块。

测结果,第三列为 P-CNN 检测结果,可以看出本文方法有更高的检测准确率和置信分数,且分类错误更少。

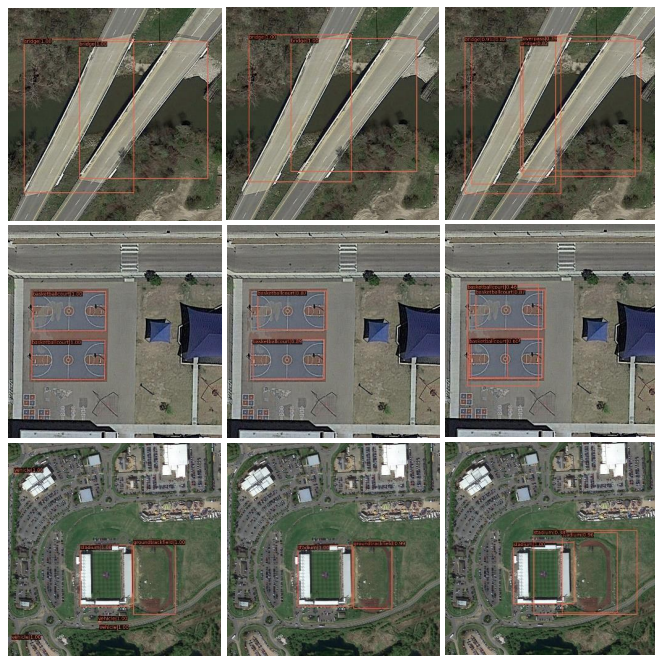


Fig. 6 Comparison of detection results on DIOR dataset

图 6 DIOR 数据集上检测结果对比

图 7 为本文使用的 NMS 算法与经典 NMS 算法在 DIOR 数据集上 $shot = 10$ 条件下的检测结果对比。第一列为真实标签,第二列为本文 NMS 算法检测结果,第三列为经典 NMS 算法检测结果,可以看出本文的 NMS 算法可有效减少对同一目标的重复检测。

图 8 为本文方法在 NWPU VHR-10 数据集上的检测结果,可以看出本文方法能够有效地检测出新类和基类目标。

4 结语

由于遥感图像具有内容丰富复杂、相同类别尺度差异大、不同类别相似性高等特点,导致遥感图像小样本目标检测成为一个巨大的挑战。本文提出一种基于对比学习的小样本遥感图像目标检测方法,通过对比学习区分相似

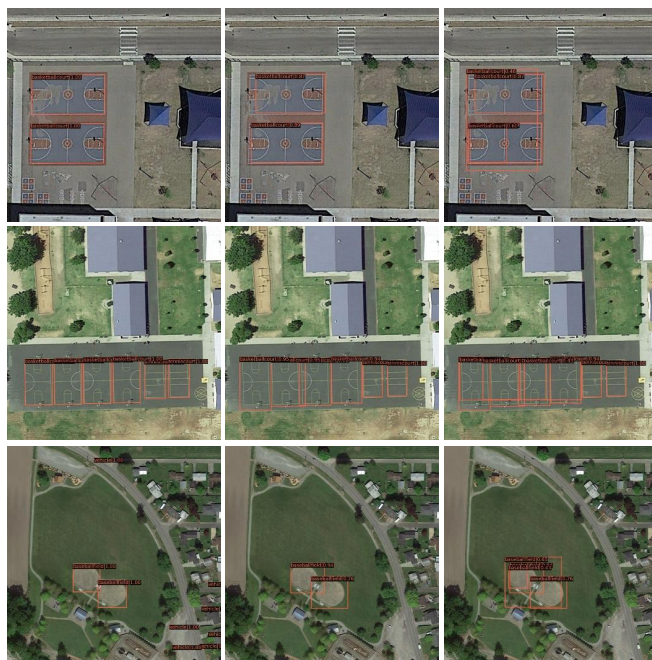


Fig. 7 Detection results comparison of the NMS algorithm in this paper and the classical NMS algorithm

图 7 本文 NMS 算法与经典 NMS 算法检测结果对比



Fig. 8 Detection result of the proposed method on NWPU dataset

图 8 本文方法在 NWPU 数据集上检测结果

的类别,以减少模型分类时的误判,提升检测的准确率,并提出了适合小样本的 NMS 算法以减少同一目标多余的检测框。在 DIOR 和 NWPU VHR-10 数据集上的实验结果表明了该方法的优异检测性能和泛化能力。未来工作可探索多模态数据在小样本场景下的协同学习机制,以进一步提升对复杂地物目标的判别能力。

参考文献:

- [1] LI K, WAN G, CHENG G, et al. Object detection in optical remote sens-

- ing images: a survey and a new benchmark [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 296–307.
- [2] YAN X P, CHEN Z L, XU A N, et al. Meta R-CNN: towards general solver for instance-level low-shot learning [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9577–9586.
- [3] KANG B Y, LIU Z, WANG X, et al. Few-shot object detection via feature reweighting [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 8420–8429.
- [4] XIE X X, CHENG G, YAO Y Q, et al. Dynamic feature fusion for object detection in remote sensing images [J]. Chinese Journal of Computers, 2022, 45(4): 735–747.
- 谢星星,程塔,姚艳清,等. 动态特征融合的遥感图像目标检测[J]. 计算机学报, 2022, 45(4): 735–747.
- [5] LI A X, LI Z G. Transformation invariant few-shot object detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 3093–3101.
- [6] WANG X, HUANG T E, DARRELL T, et al. Frustratingly simple few-shot object detection [DB/OL]. <https://arxiv.org/abs/2003.06957>, 2020.
- [7] LI X, DENG J Y, FANG Y. Few-shot object detection on remote sensing images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1–14.
- [8] CHENG G, YAN B, SHI P, et al. Prototype-CNN for few-shot object detection in remote sensing images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1–10.
- [9] ZHAO Z, TANG P, ZHAO L, et al. Few-shot object detection of remote sensing images via two-stage fine-tuning [J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1–5.
- [10] WANG X, HUANG T E, DARRELL T, et al. Frustratingly simple few-shot object detection [DB/OL]. <https://arxiv.org/abs/2003.06957>.
- [11] LI J, TIAN Y, XU Y, et al. MM-RCNN: toward few-shot object detection in remote sensing images with meta memory [J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1–14.
- [12] KHOSLA P, TETERWAK P, WANG C, et al. Supervised contrastive learning [J]. Advances in Neural Information Processing Systems, 2020, 33: 18661–18673.
- [13] SUN B, LI B H, CAI S C, et al. FSCE: few-shot object detection via contrastive proposal encoding [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 7352–7362.
- [14] LI K, WAN G, CHENG G, et al. Object detection in optical remote sensing images: a survey and a new benchmark [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 296–307.
- [15] CHENG G, HAN J W, ZHOU P C, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 98: 119–132.
- [16] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//Proceedings of the European Conference on Computer Vision, 2016: 21–37.
- [17] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137–1149.
- [18] GE Z, LIU S, WANG F, et al. YOLOX: exceeding YOLO series in 2021 [DB/OL]. <https://arxiv.org/abs/2107.08430>.
- [19] KARLINSKY L, SCHWARTZ E, SHTOK J, et al. Representative-based metric learning for classification and few-shot object detection [P]. America, 10832096, 2020–11–10.
- [20] WANG X, HUANG T E, DARRELL T, et al. Frustratingly simple few-shot object detection [DB/OL]. <https://arxiv.org/abs/2003.06957>.
- [21] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132–7141.

(责任编辑:黄 健)