



DOI:10.12404/j.issn.1671-1815.2403508

引用格式:徐阳,苏树智,朱彦敏,等.基于形状感知与类平衡优化的开放世界目标检测[J].科学技术与工程,2025,25(11):4647-4655.

Xu Yang, Su Shuzhi, Zhu Yanmin, et al. Open world object detection based on shape perception and class balance optimization[J]. Science Technology and Engineering, 2025, 25(11): 4647-4655.

基于形状感知与类平衡优化的开放世界目标检测

徐阳¹, 苏树智¹, 朱彦敏^{2*}, 王超¹

(1. 安徽理工大学计算机科学与工程学院, 淮南 232001; 2. 安徽理工大学机械工程学院, 淮南 232001)

摘要 针对开放世界目标检测中未知类目标预测性能不佳的问题,提出了一种基于形状感知与类平衡优化的开放世界目标检测方法。未知类指在训练阶段未标注的类别,由于缺乏标签的指导,未知类目标的检测是一个具有挑战性的任务。构建了一种未知类增强探测器,作为未知类检测分支,在训练阶段只利用已知类标签进行监督,让探测器学习已知类目标特征的相似性,进而推广到未知类目标。为了提高探测器对未知类的敏感度,利用区域生成网络(region proposal network, RPN)模块区分前景和背景的特性,使用特定筛选方式,从RPN输出中选择“具有未知类潜力”的结果作为伪标签参与探测器训练过程。由于缺乏置信度得分,传统非极大值抑制(non-maximum suppression, NMS)方法难以应用于未知目标的后处理,因此设计了一种冗余未知类目标框抑制器,该抑制器由基于中心点的分组策略和基于形状感知冗余度得分矩阵构成。其中基于中心点的分组策略包含三种根据未知类中心点的分组方法,用于确定抑制范围。接着根据组内每一个预测框的冗余度得分构建冗余度得分矩阵,从而抑制高冗余预测结果。在开放世界目标检测数据集上的实验结果表明基于形状感知与类平衡优化的开放世界目标检测方法在保证未知类召回率的同时,具有较高的未知类预测精度。基于形状感知与类平衡优化的开放世界目标检测方法能有效应对开放世界的难题,避免产生大量的无用预测结果。

关键词 开放世界目标检测; 未知类目标检测; 基于中心点的分组策略; 形状感知; 冗余度得分

中图分类号 TP391;

文献标志码 A

Open World Object Detection Based on Shape Perception and Class Balance Optimization

XU Yang¹, SU Shu-zhi¹, ZHU Yan-min^{2*}, WANG Chao¹

(1. School of Computer Science and Engineering, Anhui University of Science & Technology, Huainan 232001, China;

2. School of Mechanical Engineering, Anhui University of Science & Technology, Huainan 232001, China)

[Abstract] An open world object detection method based on shape perception and class balance optimization was proposed to address the issue of poor prediction performance of unknown class objects in open world object detection. Unknown classes referred to classes that were not labeled during the training phase. Due to the lack of guidance from labels, detecting unknown class objects was a challenging task. An unknown class enhanced detector has been constructed as an unknown class detection branch. During training, this detector was supervised using only known class labels, allowing it to learn the similarities in features of known class objects and generalize to unknown class objects. To improve the detector's sensitivity to unknown classes, the region proposal network (RPN) module's ability to distinguish between foreground and background was utilized. A specific filtering method was employed to select results with "unknown class potential" from the RPN output, which were then used as pseudo labels in the training process. Due to the absence of confidence scores, traditional non-maximum suppression (NMS) methods were difficult to apply for post-processing unknown objects. Therefore, a redundant unknown object suppression mechanism was designed, consisting of a center point-based grouping strategy and a redundancy score matrix based on shape perception. The center point-based grouping strategy included three methods based on the unknown class center points to determine the suppression range. Subsequently, a redundancy score matrix was constructed based on the redundancy scores of each prediction box within the group to suppress highly redundant predictions. Experimental results on open world object detection datasets demonstrated that the open world object detection based on shape perception and class balance optimization maintained high recall rates for unknown classes while achieving high prediction accuracy. This method effectively addressed the challenges of open world scenarios and avoided generating a large number of useless predictions.

收稿日期: 2024-05-12 修订日期: 2024-08-01

基金项目: 国家自然科学基金(52374155); 安徽省高等学校自然科学研究项目(2022AH040113)

第一作者: 徐阳(2000—), 男, 汉族, 江苏淮安人, 硕士研究生。研究方向: 目标检测。E-mail: yangxu_my@163.com。

* 通信作者: 朱彦敏(1988—), 女, 汉族, 山东泰安人, 博士, 讲师。研究方向: 图像处理与模式识别。E-mail: zyanmin1988@163.com。

投稿网址: www.stae.com.cn

[**Keywords**] open world object detection; unknown class object detection; center point-based grouping strategy; shape perception; redundancy score

目标检测^[1]是一项基础性的计算机视觉任务,它需定位和识别图像中的感兴趣目标,在传统的目标检测任务中这些感兴趣目标是面向封闭世界的,它们的种类以及数量都是固定的,因此,传统目标检测难以应对开放环境下的需求。例如,在自动驾驶领域,传统目标检测只能检测出学习过的目标,而忽略那些未被学习过的未知目标带来的危险。开放世界目标检测由 Joseph 等^[2]首次提出,其主要任务是在没有类标签监督的情况下检测出潜在的“未知目标”,在现实世界中,随着学习的类别越来越多,未知类目标开始转变为已知类目标。这个过程类似于人类认知新事物的过程,随着学习的深入,越来越少的事物被标记为未知。由于面向目标转变为开放的世界,开放世界目标检测比传统目标检测有更加广泛的应用前景以及更有价值的现实意义。

目前的开放世界目标检测方法对于未知类目标的检测性能有很大的提升空间,根本原因在于模型训练过程中没有标签的指导,模型很难准确检测出未知类目标。在检测过程中很容易出现忽略潜在的未知类目标的情况,同时一些非目标类会被误当作未知类目标干扰实验结果。对于未知类目标预测结果,无法像传统目标检测一样使用非极大值抑制等方法去除同一未知类目标上冗余预测结果。这些问题导致开放世界目标检测的检测性能较差,各项指标都处于较低水准。

1 相关工作

1.1 开放集识别

开集识别^[3]的目的是处理分类或检测任务中遇到的未知样本。开集识别假设通过训练集获得的知识是不完整的,即在实际应用中可能会遇到新的未知类。关于这个问题,相关学者已经探索了使用自监督学习^[4]和基于重构的无监督学习^[5]方法进行开集识别。

1.2 基于 Faster R-CNN 的开放世界目标检测

OWOD(open world object detection)任务是由 Joseph 等^[2]提出的,由于其潜在的现实影响而引起了广泛关注。在他们的工作中,Joseph 等^[2]介绍了 ORE(open world object detector),它利用具有特征空间对比聚类的 Faster R-CNN(region-based convolutional network)模型^[6]、基于 RPN 的未知检测器^[7]和用于 OWOD 对象的基于能量的未知标识符(ener-

gy based unknown identifier, EBUI)^[8]。Yu 等^[9]试图通过将特征聚类的数量设置为类的数量来最小化嵌入特征空间中已知和未知类的重叠分布,并减少已知和未知对象之间的混淆,从而扩展 ORE。同时,Zhang 等^[10]试图通过引入第二个基于位置的对象检测头来扩展 ORE,并报告了未知对象回忆的好处,刺激了对象在 OWOD 中的效用。UC-OWOD(unknown-classified open world object detection)^[11]还对未知对象进行分类,以在测量未知类时获得比 ORE 更好的结果。Zhao 等^[12]通过选择性搜索纠正了自动标注的建议,并通过类特定驱逐函数校准了过度自信的激活边界。然而,自动标记步骤生成许多伪未知样本,这些样本实际上并不代表未知样本,这限制了它们将知识从已知转移到未知的能力。Liang 等^[13]提出了一种广义对象置信度来更好地区分未知类对象,并对未知类实现了良好的召回率。然而,它们忽略了未知类的预测准确度,导致未知类的不平衡预测。

1.3 基于 Transformer 的开放世界目标检测

Gupta 等^[14]将可变形(detection transformer, DETR)模型^[15]应用于开放世界对象,并提出了(open world detection transformer, OW-DETR)。最近,基于 Transformer^[16]的方法在开放世界对象中显示出了巨大的潜力。OW-DETR 使用伪标记方案来监督未知对象检测,其中选择具有高骨干激活的不匹配对象建议作为未知对象。在 OWOD 方面的工作促使人们使用基于变压器的模型和集成对象来实现强大的 OWOD 性能。

2 基于形状感知与类平衡优化的开放世界目标检测方法

2.1 问题描述

在开放世界目标检测任务定义中,假设有一个已知类集合 $C^k = \{1, 2, \dots, N\}$ 和一个未知类集合 $C^u = \{N + 1, N + 2, N + 3, \dots\}$,模型在训练的过程中已知类集合 C^k 中类别的目标具有类标签参与训练,在测试阶段模型不仅需要检测出已知类集合 C^k 中类别的目标,还需要能够检测出未知类集合 C^u 中类别的目标。

2.2 网络结构

基于形状感知与类平衡优化的开放世界目标检测方法的网络结构如图 1 所示,其中包含主干网络、RPN、RoI Pooling、未知类增强探测器、目标分类头、

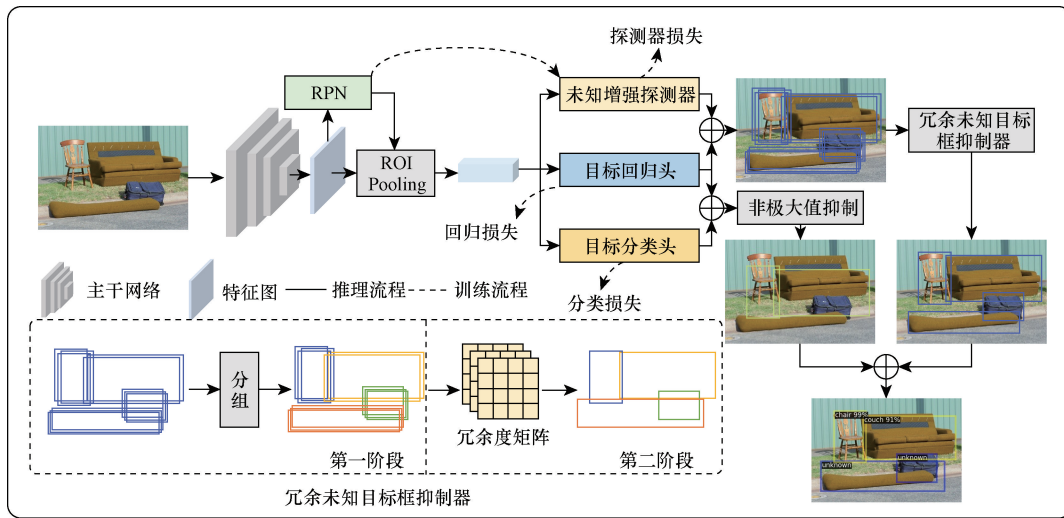


图1 网络结构

Fig. 1 Network structure

目标回归头、冗余未知目标框抑制器 (redundant unknown object box suppressor, RUOBS) 七个部分。基于形状感知与类平衡优化的开放世界目标检测方法首先通过主干网络提取特征,然后利用 RPN 模块获取潜在的目标区域,接着 RoI Pooling 部分将 RPN 输出结果进行池化,得到大小尺寸一致的特征图。然后通过目标分类头和目标回归头配合得到已知类目标预测结果,接着通过未知类增强探测器和目标回归头的配合进一步过滤掉非目标类结果,保留未知类检测结果,得到未知类目标的预测结果。最后通过 RUOBS 处理未知类目标的冗余预测框,得到未知类最终预测结果。

2.3 未知类增强探测器

由于在训练过程中缺乏标签的指导,未知类目标难以被有效检测出。为此,构建了一个新的分支,专门用于未知类目标检测。与已知类目标检测不同,用于未知类检测的分支不依赖未知类标签。它仅从不同类别目标特征的角度出发,通过学习不同类别目标特征的相似性来区分区域中是否包含潜在目标,从而实现检测未知类目标的效果。如图2所示,在特征空间中,尽管未知类目标与已知类目标存在差异,但当已知类数量达到一定程度时,未知类目标在特征空间中会分布在已知类目标周围。而非目标类由于缺乏完整的语义信息,其特征与已知类和潜在未知类的特征距离相对较大。利用这一特性,未知类目标检测分支类似于一个过滤器,它可以过滤出所有包含完整语义信息的目标,并将它们作为未知类预测结果。未知类增强探测器,在功能上类似于一个二分类的过滤器。与已知类目标的预测不同的是,未知类增强探测器不再将物体划分到各种各样的类别当中,所有候选区域中

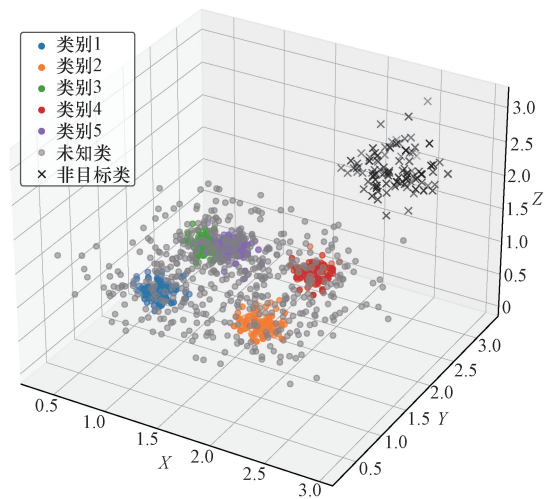


图2 特征空间示意图

Fig. 2 Schematic diagram of feature space

的所有类别中的物体都会被归为两个宽泛的类别,这里称为开放世界类与非开放世界类。

2.3.1 伪标签筛选策略

为了更好地增强未知类探测器的检测性能,利用了 RPN 模块能够区分前景和背景的特性,筛选部分 RPN 输出结果作为伪标签参与训练。这些伪标签将被视作开放世界类的扩充,使未知类探测器能够学习到更多类别目标的特征,从而检测出更多潜在的未知目标。伪标签筛选过程如图3所示。①模块计算真实标签 (ground truth, GT) 与经过 RPN 模块生成的候选区域的 (intersection over union, IoU); ②筛选出 IoU^[17] 小于阈值 k_1 的候选区域结果; ③计算保留下来的候选区域中的框的面积,并筛选出面积大于阈值 k_2 的候选区域结果作为潜在目标; ④对潜在目标使用非极大值抑制方法筛选出最优潜在目标作为伪标签。加入伪标签协调训练后,未知类

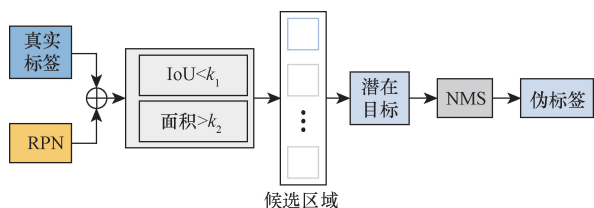


图3 伪标签产生过程

Fig. 3 Pseudo label generation process

探测器能够源源不断地学习到各种类别目标的特性,从而更好地检测出潜在的未知目标。

2.3.2 损失函数

未知类增强探测器的训练过程如图4所示,伪标签样本产生后与经过GT采样后的正样本合并产生最终正样本,其余样本均作为负样本参与计算损失。未知类增强探测器引用 Focal Loss^[18]作为损失函数,表达式为

$$L_{de} = -\alpha_t(1-p_t)^\gamma \lg p_t \quad (1)$$

式(1)中: p_t 为模型的预测结果; α_t 为模型的权重系数; γ 为超参数。 p_t 和 α_t 表达式为

$$p_t = \begin{cases} p, & y = 1 \\ 1 - p, & y \neq 1 \end{cases} \quad (2)$$

$$\alpha_t = \begin{cases} \alpha, & y = 1 \\ 1 - \alpha, & y \neq 1 \end{cases} \quad (3)$$

式中: y 为类标签; $p \in [0,1]$ 为模型对标签 $y = 1$ 的预测概率; $\alpha \in [0,1]$ 为加权系数。

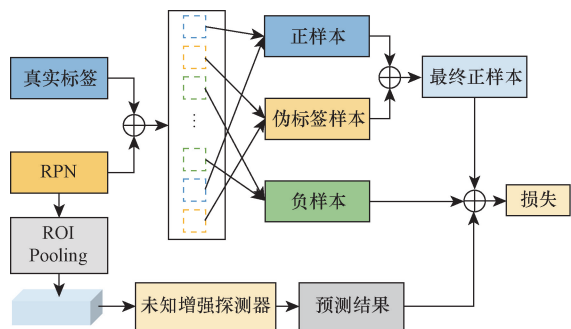


图4 未知类增强探测器训练过程

Fig. 4 Training process of unknown class enhanced detector

2.4 基于中心点的分组策略

由于置信度得分^[19]不适用于未知目标,同时未知目标的数量是不确定的,这就导致了传统的(non-maximum suppression, NMS)方法无法确定最佳预测框。因此,提出了一种基于预测框中心点位置的分组策略,以确定冗余未知类目标框抑制的范围。采用基于中心点的分组策略是为了将相对位置接近的预测框划分到同一组中,方便后续计算每一组预测框的冗余度得分。最理想的分类状态是将每一个未知类目标上预测框分为一组,然而难点在于预测过程中未知类的数量未知,因此传统的聚类方法

如K-means系列的分类方法无法获得类别数量,在分类过程中容易出现如图5展示的问题。面对三个预测框,在不知类别数目的情况下,难以确定将其分为一组、两组或者三组。而在实际预测中,期望将其分到同一组。

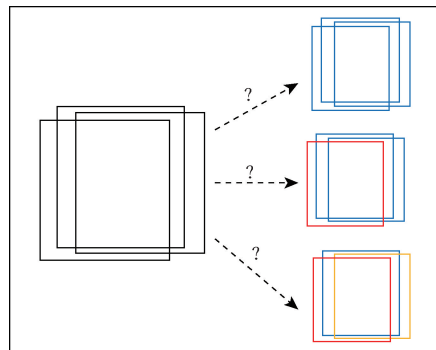


图5 传统分类方法

Fig. 5 Traditional classification methods

2.4.1 基于中心点邻接图的分类

针对类别数量不确定导致难以分组的问题,设计了一种基于中心点密度的分类方法,利用所有未知类预测框的每一个中心点与其他所有点相连,构成邻接图。接着使用深度优先搜索的方法遍历这个邻接图并将与节点欧式距离小于阈值 k 的点分到同一组,最终获得分类结果。这种方法的优点在于不需要事先知道类别的数量,但是容易造成偏差传递,遇到密集点位容易产生分类错误的问题,分类结果如图6所示。

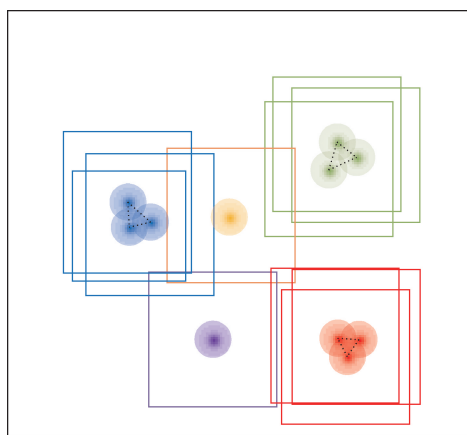


图6 基于中心点邻接图的分类

Fig. 6 Classification based on center point adjacency graph

2.4.2 基于中心点所在网格分类

针对2.4.1节中的问题,设计了一种基于中心点所在网格分类的方法,如图7所示,输入图片根据长宽比被划分成 $N \times M$ 个网格区域,根据预测框的中心点坐标计算每个预测框中心点所在的区域,并将中心点位于同一区域的预测框分组。这种方法的

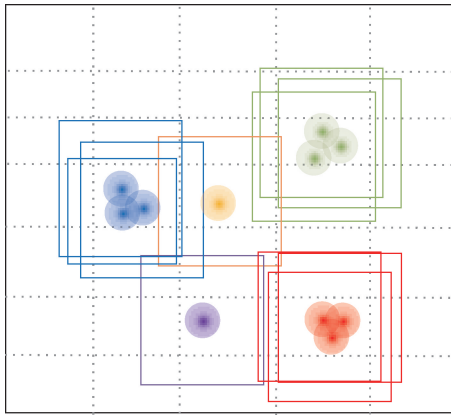


图7 基于中心点所在网格分类

Fig. 7 Classification based on the grid where the center point is located

优点在于密集点位的包容性较强,不会产生偏差传递,但由于网格分割是完全按照图片比例,因此在网格边缘的点容易被错误地分到别的组。

2.4.3 基于中心点自适应层次分类

层次聚类的合并算法通过计算两类数据点间的距离,对最为接近的两类数据点进行组合,并反复迭代这一过程,直到将所有数据点合成一类,并生成聚类谱系图。然而此方法依旧依赖类别数量信息作为参考,因此提出了一种自适应层次分类方法。此方法在层次聚类迭代的过程中通过判断类内距离与阈值的关系,不断优化类别数量,最终确定分类的结果。基于中心点自适应层次分类的工作流程如下。①将类别数目设定为1;②进行层次聚类;③计算每一类的类内距离;④判断类内距离与阈值关系,若所有的类内距离都小于阈值则分类结束,否则类别数目加1;重复②~④直到分类结束。此处类间距离采用平均欧式距离,首先计算每一个点到同一类里所有的点的欧式距离,接着用欧式距离之和除以总边数。图8所示为一个自适应层次分类的示例,每一次分类之后构建每一类的类内距离矩阵,直到所有的类内距离都小于阈值。基于

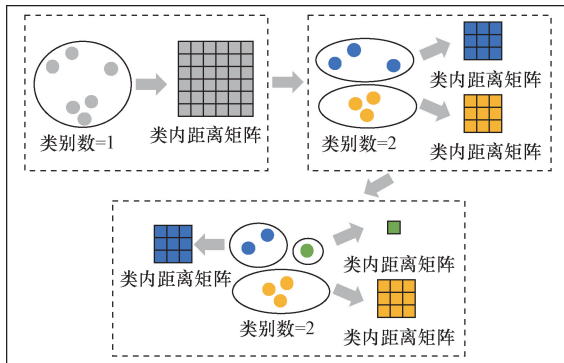


图8 自适应层次分类

Fig. 8 Adaptive hierarchical classification

中心点自适应层次分类方法,可以有效避免类别数目缺失问题,同时能应对密集点位以及边缘点位错分的问题。

2.5 基于形状感知的冗余度得分

对于同一分组的未知类目标预测框,设计了一种基于形状感知的冗余度得分,根据预测框在IoU以及在形状方面的相似程度得到冗余度得分(redundancy score, RDS),RDS可表示为

$$RDS(a, b) = \lambda IoU(a, b) + \rho \text{sim}(a, b) \quad (4)$$

$$IoU(a, b) = \frac{a \cap b}{a \cup b} \quad (5)$$

$$\text{sim}(a, b) = 1 - \sqrt{[AD(a, b)]^2 + [ARD(a, b)]^2} \quad (6)$$

$$AD(a, b) = \frac{A_a - A_b}{\max(A_a, A_b)} \quad (7)$$

$$ARD(a, b) = |H_a / (H_b - W_a) / W_b| \quad (8)$$

式中: a 和 b 为待计算RDS的目标框; $a \cap b$ 为目标框的交集面积; $a \cup b$ 为目标框的并集面积; λ 和 ρ 为用于平衡的超参数; A 为预测框的面积; H 为预测框的高度; W 为预测框的宽度。

基于预测框中心点位置的分组策略旨在尽可能将集中的未知类预测框进行分组。同时,基于形状感知的冗余度得分矩阵则反映了同一组内每个预测框与其他所有预测框之间的冗余程度。针对冗余度得分较高的预测框,模型会优先保留索引靠前的预测框。这一设计的目的是在相对集中的预测框中过滤掉冗余度较高的预测框。这种方法可以帮助模型更好地处理冗余预测框,从而提高未知类预测精度方面表现。

3 实验

3.1 数据集

实验选用的数据集为UnSniffer方法中提出的开放世界数据集,在训练过程中,仅使用Pascal VOC数据集^[20],Pascal VOC仅包含20个类,而在测试阶段,将分别从已知类识别和未知类识别来评估模型。

如表1所示,测试集包含三个部分,其中VOC测试数据集由4 952个图像组成,其中仅包含所学习的20个类的标记信息,而COCO-OOD由504个图

表1 测试数据集

Table 1 Test dataset

数据集	图片	已知	未知
VOC-Test	4 952	5.09	0
COCO-OOD	504	0	3.28
COCO-Mixed	897	2.96	2.82

注:表1展示了不同测试集的图片数量以及每张图片对应的已知类和未知类实例的平均数量。

像构成,其中仅包括 1 655 个未知类的对象的标记信息。COCO-Mixed 数据集包含已知和未知类,由 897 张图像组成,其中包含 2 533 个未知对象和 2 658 个已知对象的注释信息。MS-COCO 数据集^[21]中与 Pascal VOC 对象类重叠的原始标签被用作已知对象的标签,所有标签都由 MS-COCO 中的原始标签和基于 MS-COCO 定义的扩展标签组成。评估模型在检测未知和已知对象方面的可靠性更具挑战性,因为 COCO-Mixed 中的每个图像都包含更多具有复杂类和集中位置的对象实例。

3.2 评估指标

开放世界目标检测不仅需要检测已知对象,还需要检测未知对象,因此评估指标也分为已知对象和未知对象两个方面。为了评估已知物体的检测性能,使用平均准确度(mean average precision, mAP),mAP 的计算公式为

$$P = \frac{TP}{TP + FP} \quad (9)$$

$$R = \frac{TP}{TP + FN} \quad (10)$$

$$AP = \int_0^1 P(R) dR \quad (11)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (12)$$

式中:TP 为已知类中的阳性样本;FP 为已知类的假阳性情况;FN 为已知类内的假阴性情况; P 为模型识别的阳性样本与真阳性样本的比率; R 为模型正确识别的样本与阳性样本的比例; AP 为 P 和 R 曲线下的区域。对于未知物体检测性能的评估,使用未知平均准确度(AP_u)、未知召回率(R_u)、未知 F_1 分数(F_{1u})和未知准确度(P_u)计算公式为

$$P_u = \frac{TP_u}{TP_u + FP_u} \quad (13)$$

$$R_u = \frac{TP_u}{TP_u + FN_u} \quad (14)$$

$$F_{1u} = \frac{2P_u R_u}{P_u + R_u} \quad (15)$$

$$AP_u = \int_0^1 P_u(R_u) dR_u \quad (16)$$

式中: TP_u 为未知类中预测的阳性样本; FP_u 为未知类中的假阳性样本;而 FN_u 则为未知类的假阴性样本; R_u 为正确识别未知类中阳性样本的比率; P_u 为未知类中正性样本与真阳性样本的比例; F_{1u} 为 P_u 和 R_u 的调和平均值。

3.3 对比实验

表 2 为在 VOC-test 数据集上进行的,基于形状

感知与类平衡优化的开放世界目标检测方法虽然并不是最佳算法,但是与最优算法差距仅有 2.5%,VOC-test 数据集中只包含已知类,除了 OW-DETR 外而对于已知类目标的预测都是依靠 Faster R-CNN 模型,因此在已知类预测的平均准确度上,各个算法表现差异不大。

表 3 中结果表明,基于形状感知与类平衡优化的开放世界目标检测方法在 AP_u 、 F_{1u} 和 P_u 指标上表现优异,尤其是在 P_u 方面,比第二名高了 13%。

表 4 中的结果表明,基于形状感知与类平衡优化的开放世界目标检测方法在复杂度较高的 COCO-Mix 数据集上也有不错的表现,虽然相对于 ORE 方法,在未知类目标召回率 R_c 表现相对一般,但是由于 ORE 产生大量的无用预测框导致其未知类预测准确度(P_u)方面表现较差。

表 2 VOC 数据集上实验结果

方法	mAP
Faster R-CNN	0.487
OW-DETR	0.420
VOS ^[21]	0.491
ORE	0.243
UnSniffer	0.462
本文方法	0.466

表 3 COCO-OOD 数据集上实验结果

方法	AP_u	F_{1u}	P_u	R_c
Faster R-CNN	—	—	—	—
OW-DETR	0.033	0.056	0.030	0.380
VOS	0.234	0.296	0.277	0.319
ORE	0.214	0.255	0.153	0.782
UnSniffer	0.396	0.475	0.487	0.464
本文方法_v1	0.367	0.485	0.602	0.407
本文方法_v2	0.392	0.498	0.619	0.417
本文方法_v3	0.402	0.507	0.617	0.430

注:v1 版本是使用基于中心点邻接图的分类方法,v1 版本是使用基于中心点所在网格分类方法,v3 版本使用的是基于中心点自适应层次分类方法。

表 4 COCO-Mix 数据集上实验结果

方法	mAP	AP_u	F_{1u}	P_u	R_c
Faster R-CNN	—	—	—	—	—
OW-DETR	0.414	0.007	0.025	0.014	0.161
VOS	0.364	0.117	0.164	0.191	0.144
ORE	0.213	0.140	0.175	0.103	0.592
UnSniffer	0.362	0.121	0.252	0.201	0.338
本文方法_v1	0.375	0.102	0.234	0.227	0.242
本文方法_v2	0.375	0.098	0.228	0.225	0.233
本文方法_v3	0.375	0.138	0.252	0.243	0.263

3.4 消融实验

为了验证 RUOBS 这个模块的三个不同版本的有效性,设计了针对该模块的消融实验。实验结果如表 5 和表 6 所示。

由表 5 和表 6 可以看出不同的方法加入 RUOBS 后性能均有不同程度的提升。在 COCO-OOD 数据集上,VOS 加入 RUOBS 模块后提升最为明显,未知目标 P_u 最好的效果提升了 23.5%, F_{1u} 提升了 8%。ORE 加入 RUOBS 模块后 P_u 提升 17.5%了, F_{1u} 提升了 19.5%,UnSniffer 加入 RUOBS 模块后 P_u 提升了 0.5%。在 COCO Mix 数据集上,VOS 在加入 RUOBS

模块后 P_u 最好的效果提升了 7.5%, F_{1u} 提升了 1%。ORE 在加入 RUOBS 模块后 P_u 最好的效果提升了 7.4%, F_{1u} 提升了 9.4%。UnSniffer 加入 RUOBS 模块后 P_u 最好的效果提升了 2.6%, F_{1u} 提升了 1.8%。由此可以得出,RUOBS 模块可以有效抑制未知目标的冗余预测框,从而提升未知目标的预测准确度。

3.5 可视化分析

图 9 展示了各种算法在开放世界环境下的检测结果,可知,本文算法在不同环境下都有较好的检测性能,尽管对于一些小型未知类目标存在一定的漏检,但是未知类目标错检结果相对是最少的。

表 5 COCO-OOD 数据集上消融实验结果

Table 5 Results of ablation experiments on the COCO-OOD dataset

RUOBS_v1	RUOBS_v2	RUOBS_v3	方法	P_u	R_c	F_{1u}
×	×	×	VOS	0.277	0.319	0.296
√	×	×		0.507	0.301	0.377
×	√	×		0.476	0.302	0.369
×	×	√		0.512	0.298	0.376
×	×	×	ORE	0.153	0.782	0.255
√	×	×		0.317	0.721	0.440
×	√	×		0.320	0.723	0.443
×	×	√		0.328	0.717	0.450
×	×	×	UnSniffer	0.487	0.468	0.477
√	×	×		0.492	0.455	0.472
×	√	×		0.488	0.466	0.476
×	×	√		0.492	0.457	0.473
×	×	×	本文方法	0.582	0.437	0.499
√	×	×		0.607	0.433	0.505
×	√	×		0.612	0.428	0.503
×	×	√		0.617	0.430	0.507

注:粗体表示最优值,其中 v1 版本是使用基于中心点邻接图的分类方法,v2 版本是使用基于中心点所在网格分类方法,v3 版本使用的是基于中心点自适应层次分类方法。√表示引入该模块的结果,×表示没有引入改模块的结果。

表 6 COCO-Mix 数据集上消融实验结果

Table 6 Results of ablation experiments on the COCO Mix dataset

RUOBS_v1	RUOBS_v2	RUOBS_v3	方法	P_u	R_c	F_{1u}
×	×	×	VOS	0.191	0.144	0.164
√	×	×		0.257	0.130	0.172
×	√	×		0.266	0.129	0.174
×	×	√		0.265	0.129	0.173
×	×	×	ORE	0.103	0.592	0.175
√	×	×		0.152	0.580	0.240
×	√	×		0.177	0.562	0.269
×	×	√		0.165	0.578	0.257
×	×	×	UnSniffer	0.201	0.338	0.252
√	×	×		0.217	0.332	0.262
×	√	×		0.212	0.336	0.259
×	×	√		0.227	0.336	0.270
×	×	×	本文方法	0.205	0.275	0.234
√	×	×		0.237	0.270	0.252
×	√	×		0.225	0.267	0.244
×	×	√		0.243	0.263	0.252

注:粗体表示最优值,其中 v1 版本是使用基于中心点邻接图的分类方法,v2 版本是使用基于中心点所在网格分类方法,v3 版本使用的是基于中心点自适应层次分类方法。√表示引入该模块的结果,×表示没有引入改模块的结果。

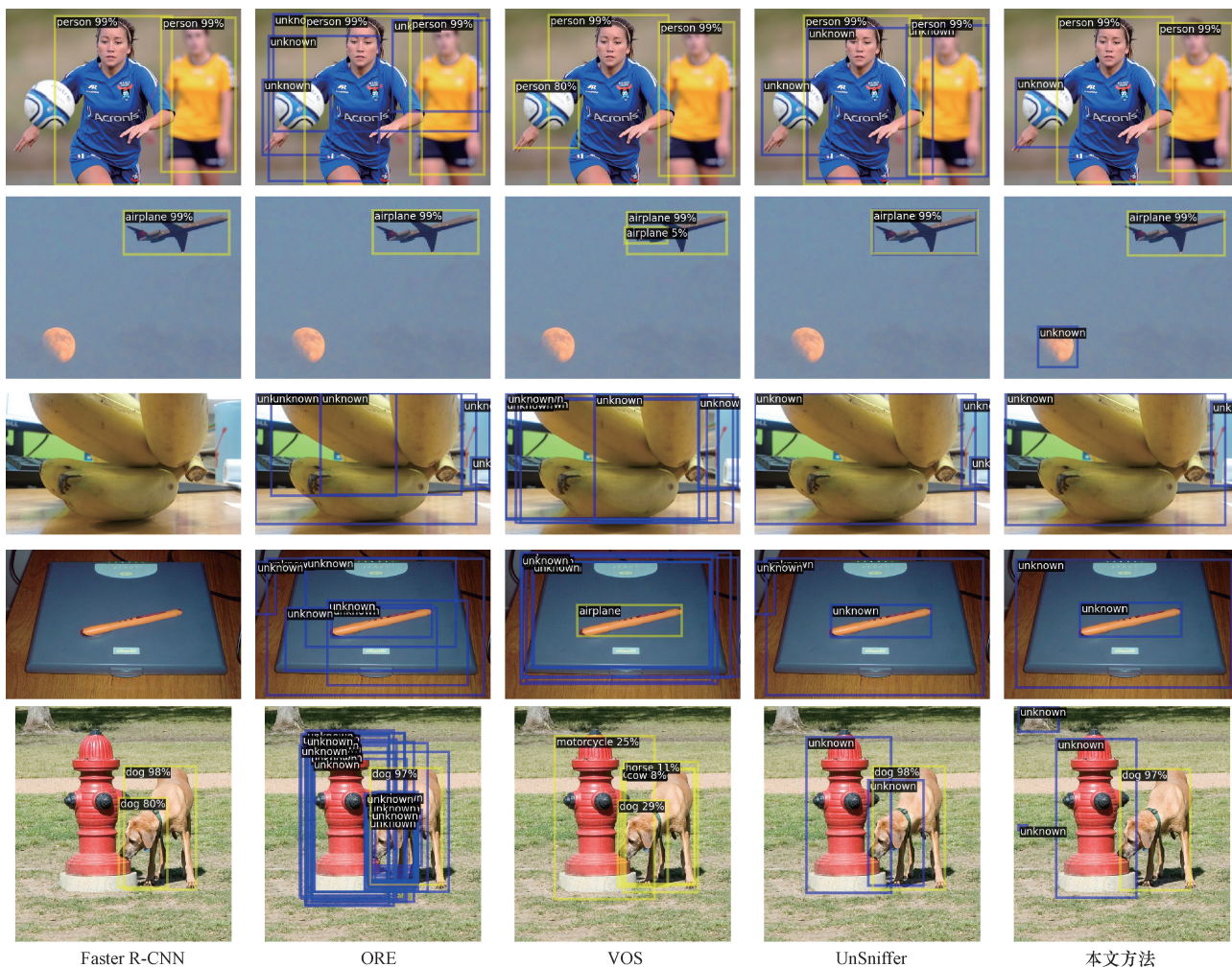


图9 推理结果可视化

Fig.9 Visualization of inference results

ORE 和 VOS 虽然在未知类目标的召回率方面表现出优势,但是检测结果中包含了大量无效预测框,这将极大影响未知目标的预测精度。Faster R-CNN 由于只能预测已知类别的目标因此图中第三行和第四行的预测结果为空。

4 结论

对开放世界目标检测任务,提出了一种形状感知与类平衡优化的开放世界目标检测方法,通过结合未知类增强探测器,基于预测框中心点位置的分组策略以及基于形状感知的冗余度得分矩阵等新颖方法,有效地应对了开放世界目标检测任务中冗余未知类预测框和未知类目标检测的挑战。通过实验验证,在开放世界目标检测数据集上的大量实验结果显示,这种基于形状感知与类平衡优化的开放世界目标检测方法在各个方面均有优异的表现,特别是在未知类目标的预测准确度方面表现出了远超其他方法的性能。

参 考 文 献

- [1] 吴晨曦, 应保胜, 许小伟, 等. 基于改进单步多框目标检测的道路小目标检测算法[J]. 科学技术与工程, 2023, 23(5): 2051-2058.
Wu Chenxi, Ying Baosheng, Xu Xiaowei, et al. Road small target detection algorithm based on improved SSD[J]. Science Technology and Engineering, 2023, 23(5): 2051-2058.
- [2] Joseph K J, Khan S, Khan F S, et al. Towards open world object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New York: IEEE, 2021: 5830-5840.
- [3] Zheng J, Li W, Hong J, et al. Towards open-set object detection and discovery[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2022: 3961-3970.
- [4] Jaiswal A, Babu A R, Zadeh M Z, et al. A survey on contrastive self-supervised learning[J]. Technologies, 2020, 9(1): 8-15.
- [5] Schmarje L, Santarossa M, Schröder S M, et al. A survey on semi-, self-and unsupervised learning for image classification[J]. IEEE Access, 2021, 9: 82146-82168.
- [6] 郑涵, 田猛, 赵延峰, 等. 基于改进 Faster R-CNN 的手部姿势

- 估计方法[J]. 科学技术与工程, 2023, 23(3): 1160-1167.
- Zheng Han, Tian Meng, Zhao Yanfeng, et al. Research on hand pose estimation based on improved Faster R-CNN method[J]. Science Technology and Engineering, 2023, 23(3): 1160-1167.
- [7] Gavrilescu R, Zet C, Fosalău C, et al. Faster R-CNN: an approach to real-time object detection[C]//2018 International Conference and Exposition on Electrical and Power Engineering (EPE). New York: IEEE, 2018: 165-168.
- [8] Zhang Y, Dai L, Wong E W M. Optimal BS deployment and user association for 5G millimeter wave communication networks[J]. IEEE Transactions on Wireless Communications, 2020, 20(5): 2776-2791.
- [9] Yu J, Ma L, Li Z, et al. Open-world object detection *via* discriminative class prototype learning[C]//2022 IEEE International Conference on Image Processing (ICIP). New York: IEEE, 2022: 626-630.
- [10] Wu Z, Lu Y, Chen X, et al. UC-OWOD: unknown-classified open world object detection[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 193-210.
- [11] Zhang Y, Zhang X Y, Shi H. OW-TAL: learning unknown human activities for open-world temporal action localization[J]. Pattern Recognition, 2023, 133: 109027.
- [12] Zhao X, Ma Y, Wang D, et al. Revisiting open world object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(5): 1201-1215.
- [13] Liang W, Xue F, Liu Y, et al. Unknown sniffer for object detection: don't turn a blind eye to unknown objects[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2023: 3230-3239.
- [14] Gupta A, Narayan S, Joseph K J, et al. Owdetr: open-world detection transformer[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2022: 9235-9244.
- [15] Zhu X, Su W, Lu L, et al. Deformable detr: deformable transformers for end-to-end object detection[J]. arXiv preprint arXiv: 2010.04159, 2020.
- [16] Wenkel S, Alhazmi K, Liiv T, et al. Confidence score: the forgotten dimension of object detection performance evaluation[J]. Sensors, 2021, 21(13): 4350.
- [17] Zhou D, Fang J, Song X, et al. IoU loss for 2d/3d object detection[C]//2019 international conference on 3D vision (3DV). New York: IEEE, 2019: 85-94.
- [18] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2017: 2980-2988.
- [19] Wenkel S, Alhazmi K, Liiv T, et al. Confidence score: the forgotten dimension of object detection performance evaluation[J]. Sensors, 2021, 21(13): 4350.
- [20] Tong K, Wu Y. Rethinking PASCAL-VOC and MS-COCO dataset for small object detection[J]. Journal of Visual Communication and Image Representation, 2023, 93: 103830.
- [21] Bideaux M, Phe A, Chaouch M, et al. 3D-COCO: extension of MS-COCO dataset for image detection and 3D reconstruction modules[J]. arXiv preprint arXiv: 2404.05641, 2024.