多模态引导下基于不确定挖掘的开放集目标检测

韩嘉雯，陈 莹†

(江南大学 轻工过程先进控制教育部重点实验室， 江苏 无锡 214122)

摘 要: 开放集目标检测中, 现有对比学习方法虽能增强类间分离与类内紧凑性, 但由于未充分建模未知目标分布, 当已知与未知目标特征相似时易产生误判. 针对已知与未知目标在视觉和语义表示层面的特征混淆, 以及模型在判别边界附近过度自信的问题, 提出一种多模态引导下的样本不确定性挖掘框架. 该方法首先设计区域生成模块, 提高候选区域对未知目标的保留能力; 其次构建区域-文本匹配模块, 利用区域-文本对齐损失与视觉特征对比损失增强已知类表征的判别性; 进一步提出基于双重不确定性的伪未知样本挖掘机制, 结合归因梯度与视觉定位质量校准筛选高质量伪未知样本, 建立自适应的已知–未知判别边界. 实验结果表明, 在 VOC-COCO-60 数据集上, 所提方法将未知类平均精度提升 165.14%, 验证了其有效性与优越性.

关键词: 开放集；目标检测；多模态；不确定性；区域文本匹配；归因梯度

中图分类号: T18 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2025.1014

引用格式: 韩嘉雯, 陈莹. 多模态引导下基于不确定挖掘的开放集目标检测 [J]. 控制与决策.

Uncertainty mining-based open-set object detection with multimodal guidance

HAN Jia-wen，*CHEN Ying*†

(Key Laboratory of Advanced Process Control for Light Industry of Ministry of Education， Jiangnan University， Wuxi 214122， China)

Abstract: In open-set object detection, existing contrastive learning methods can improve inter-class separability and intra-class compactness; however, due to their insufficient modeling of the distribution of unknown objects, misclassification is prone to occur when known and unknown objects exhibit similar features. To address feature confusion between known and unknown objects in both visual and semantic representation spaces, as well as the model's overconfidence near decision boundaries, this paper proposes a multimodal-guided sample uncertainty mining framework. Specifically, a region generation module is first designed to enhance the retention of candidate regions containing potential unknown objects. A region-text matching module is then constructed, where a region-text alignment loss and a visual feature contrastive loss are jointly employed to improve the discriminability of known-class representations. Furthermore, a pseudo-unknown sample mining mechanism based on dual uncertainty is introduced, which integrates attribution gradients with visual localization quality calibration to identify high-quality pseudo-unknown samples and establish an adaptive decision boundary between known and unknown classes. Experimental results on the VOC-COCO-60 benchmark demonstrate that the proposed method improves the average precision of unknown classes by 165.14%, validating its effectiveness and superiority.

Keywords: open-set；object detection；multimodal；uncertainty；region-text matching；attribution gradient

0 引 言

开放集目标检测 (Open-Set Object Detection, OSOD) 作为计算机视觉领域的重要研究方向, 其核心任务是构建一个能够在测试环境中准确识别训练时未见的类别, 并将其正确分类为未知类别的检测器. 这一任务面临着两大关键挑战: 首先是如何有效降低高置信度未知类别的误判率, 即避免将未知类别错误地归类为已知类别; 其次是如何在维持已知类别分类性能的同时, 显著提升检测器对未知类别的检测能力. 这些挑战的解决对于推动目标检测系

收稿日期: 2025-09-26；**录用日期:** 2026-03-31.
基金项目: 国家自然科学基金项目 (62173160).
责任编委: 魏秀琨.
†通信作者. E-mail: chenying@jiangnan.edu.con

统在现实场景中的实际应用具有重要意义.

当前的主流方法主要采用对比学习或聚类等方法[1] 来增强类间分离性和类内紧凑性. 这类方法通过在特征空间中拉大不同类别之间的距离, 同时缩小同类样本之间的距离, 从而提升模型的判别能力. 然而, 当已知类别与特征相近的未知对象共存时, 现有方法往往表现出明显的局限性: 由于未考虑到未知类在特征空间中的分布, 靠近已知类的未知类通常会被识别为高置信度的已知类.

该问题在现有基于对比聚类的开放集目标检测(OSOD) 研究中屡见不鲜[2-4], 其根本原因在于已知类别与未知类别在视觉和语义上下文层面存在相似性, 导致模型难以在特征空间中有效区分二者. 本文将该现象统称为“特征混淆问题”. 如图 1 所示, 第一列为原始图像, 第二列为 OpenDet 的可视化结果, 第三列为本文方法的可视化结果. 在 VOC 数据集上训练的开放集检测器, 通常将未知类别 (如斑马) 误分类为视觉或语义相近的已知类别 (如马、牛或羊). 从特征层级分析, 该混淆现象主要体现在低层视觉特征与高层语义特征两个层面. 首先, 在低层视觉特征层面, 斑马与马在几何结构和轮廓形态上具有较高相似性, 尽管其纹理不同, 但检测网络在早期阶段更多关注边缘和形状特征, 导致二者在低层特征空间中接近. 其次, 在高层语义特征层面, 斑马、牛、羊等目标常出现在相似的场景 (如草原), 因此模型在融合区域语义和上下文信息后, 进一步缩小了它们在高层特征空间中的分布差距. 上述分析表明, 单纯依赖视觉模态的目标检测模型在不同特征层级上都可能出现判别失效. 因此, 仅依靠视觉特征难以在特征空间中建立清晰且稳健的已知–未知判别边界, 特别是当未知类别与已知类别在视觉和语义上具有较高相似度时, 传统的对比聚类方法难以有效缓解开放集场景下的误分类问题.

原图 OpenDet Ours

图1 特征混淆示例图

为解决上述特征混淆问题及模型在判别边界区域易产生过度自信预测的问题, 本文提出了一种基于多模态引导的样本不确定性挖掘框架. 该框架通过引入文本信息, 与视觉特征协同建模. 其中, 文本模态为模型提供稳定的类别级语义结构约束, 有助于缓解跨域或分布偏移条件下的高层语义特征漂移问题; 然而, 由于区域–文本对齐主要针对全局语义一致性建模, 对不同视觉实例之间的细粒度差异刻画有限, 因此仍需引入视觉特征层面的判别约束, 二者共同优化已知类别表示. 具体而言, 本文首先设计了区域生成模块, 通过改进目标性得分生成高覆盖率的候选框; 接着, 本文引入区域–文本匹配模块, 从视觉与文本双模态的角度对候选框特征进行联合对齐, 进而构建更加紧凑、稳定的已知类别边界, 减少与未知类别之间的特征混淆.

需要注意的是, 在本文的开放集目标检测设定中, 未知类别在训练与推理阶段均不提供任何语义描述或文本提示. 因此, 在缺乏真实未知样本及其语义先验的情况下, 模型需要从已知样本中挖掘具有高不确定性的候选区域, 将其作为伪未知目标参与训练, 以构建更可靠的已知–未知判别边界. 为此, 本文进一步提出了基于双重不确定性的伪未知样本挖

掘机制. 该模块在区域–文本匹配分数的引导下, 结合归因梯度思想, 设计了一种新的不确定性度量方法: 通过对中间特征层求导获得视觉特征梯度, 并据此计算特征不确定性分数, 同时借助候选框的视觉定位质量进行校准优化. 通过视觉特征梯度信息与定位质量约束的联合建模, 模型能够更准确地筛选出高质量的伪未知样本, 进而构建更加清晰且自适应的已知–未知判别阈值, 降低未知目标被误分类的风险.

本文贡献如下:

1) 为解决 OSOD 中特征混淆导致的误分类问题, 设计新的检测框架, 利用区域-文本对齐模块实现语义相近的未知类和已知类的有效分离.

2) 为挖掘伪未知样本, 提出检测未知优化模块, 通过归因梯度思想解释不确定性, 筛选伪未知类, 并利用视觉定位质量对其进行校准, 从而提升伪未知样本的质量.

3) 在 VOC-COCO-60 数据集下, 相比目前的 SOTA, 该方法将未知类的平均精度提升了 165.14%.

1 相关工作

1.1 开放集目标检测

开放集目标检测旨在在开放环境下同时实现对已知目标的高效检测以及对未知目标的识别与分类.

尽管开放集识别研究成果颇丰, 但开放集目标检测仍面临未知目标易被误判等关键问题. Dhamija 等[5] 率先揭示 OSOD 中未知目标常被错分为已知类别的问题. Miller 等[6] 借助 Dropout Sampling 等技术建模认知不确定性, 提升了检测精度, 却因依赖不确定性阈值, 在复杂场景中易产生误判.

为提升检测性能, 一些学者基于对比学习探索未知类别表示学习, 如 Han 等[3] 提出 OpenDet, 通过分离潜在空间中的高密度和低密度区域识别未知物体; PUDet[7] 则利用类反点概念将已知目标推理到开放空间. 另一些方法直接采用距离阈值或不确定性阈值区分已知与未知, 这类方法容易造成二者误判. 此外, Josep 等[2]、Zheng 等[8]、Wu 等[9] 和 Zhao 等[10] 在训练过程中采用基于目标性得分的区域提议网络 (Region Proposal Network, RPN) 筛选伪未知目标, 但这类方法可能将背景误判为未知目标, 从而混淆前景与背景. 以上方法仍存在不足, 为进一步提高未知目标检测性能, 本文从定位和分类两个角度改进开放集目标检测.

1.2 开放词汇目标检测

开放词汇目标检测通常借助多模态学习与视觉语言预训练模型 (如 CLIP[11]) 提升类别泛化能力. 该类方法通常引入文本模态与预训练视觉语言模型, 利用预训练编码器获取视觉特征与文本嵌入, 并通过对比学习损失对齐双模态表示, 从而将文本语义信息融入视觉特征空间, 实现对未见类别的检测能力. CLIP 通过大规模图像–文本对联合训练, 展现了强大的多模态表示能力.

基于上述思想, Zareian 等人提出 OVR-CNN 框架[12], 明确了开放词汇检测的任务定义与实验设置; Zhou 等[13] 通过引入 ImageNet-21K 类别扩充训练数据, 显著提升了 Detic 的开放词汇检测性能. 然而, CLIP 难以直接应用于区域级目标检测任务. 为此, Zhong 等 [14] 提出 RegionCLIP, Li 等 [15] 提出 GLIP, Gao 等[16] 提出 Grad-OVD, 通过不同方式增强区域级多模态特征对齐与跨类别泛化能力.

尽管上述方法在开放词汇检测任务中取得了显著进展, 但其推理阶段通常依赖显式文本标签或类别描述完成目标识别, 因此并未处于完全无先验的开放环境. 在缺乏未知类别语义描述时, 这类方法难以对未知目标形成独立判别, 其检测行为更多表现为对已知类别集合的扩展识别.

相比之下, 开放集目标检测关注在训练阶段与推理阶段均不提供未知类别语义先验的条件下, 显式区分已知与未知目标. 基于此, 本文从语义表示与判别置信两个互补维度出发, 探索在无未知类别语义先验条件下对未知目标的有效建模方式, 从而提升目标检测模型在真实开放环境中的适用性.

1.3 不确定性估计

随着深度学习在现实场景中的广泛应用, 模型可靠性问题受到越来越多关注, 而评估神经网络不确定性对于获得可靠预测和良好校准的置信度至关重要[17]. 部分研究[18-20] 通过最大对数概率分数识别和处理未知环境中高不确定性样本. 部分研究通过模型预测熵评估样本不确定性[21], 例如 Chan 等[22] 以像素级 softmax 熵为基础, 通过最大化分布外样本的 softmax 熵提升不确定性表征能力. 还有一些研究[23] 利用能量函数构建不确定性估计模型. 在分布外检测领域, 新兴方法则通过神经网络可解释性技术, 即归因梯度方法[24] 实现不确定性量化. 本文提出基于归因梯度的不确定性计算方法, 相比其他不确定性估计方法, 对于多模态引导下的视觉特征具有更强的可解释性.

2 方法介绍

现有的大部分开放集目标检测研究都基于对比

学习或聚类等方法来增大类间距离、减少类内距离, 但由于未考虑未知类在特征空间中的分布, 导致当未知类特征与已知类特征分布相近时, 容易造成特征混淆.

针对上述问题, 本文构建了如图 2 所示的网络框架. 输入图像首先经过预训练的图像编码器提取特征图, 并送入区域池化层; 同时, 图像经预训练的区域生成模块产生大量候选框, 并计算对应的目标性得分, 从中筛选出高质量候选区域后送入区域池化层, 获得对应的区域特征. 第i个候选框的特征进一步输入注意力层, 生成最终视觉特征$F(x_i)$. 对于已知类别, 其文本标签通过预训练的文本编码器生成文本嵌入T_j, 并与视觉特征$F(x_i)$计算区域文本匹配分数$s_{ij}(s_{ij}=F(x_i)\cdot T_j)$, 以上构成本文的基础框架. 同时, 由于开放集设定下仅提供已知类文本信息, 无法直接得到未知类和背景类的匹配分数, 因此额外引入两个全连接层分别输出未知类得分和背景类得分.

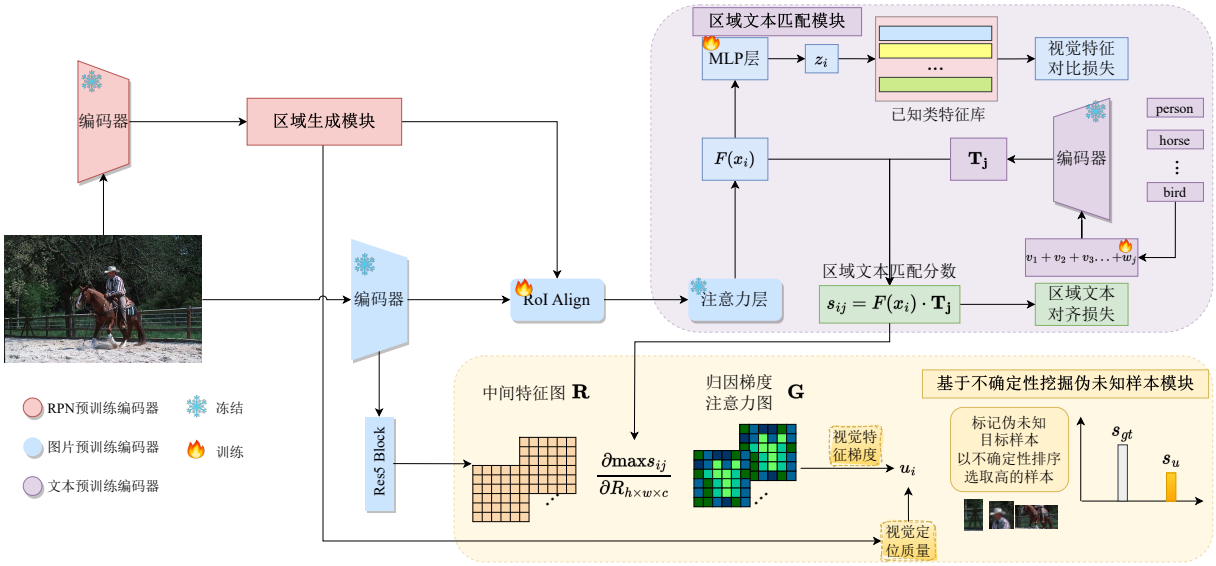


图2 网络框架

在此基础上, 本文为提升开放集目标检测性能设计了三个关键模块. 首先, 区域生成模块通过改进目标性得分, 减弱候选框筛选过程对已知类的偏向. 其次, 区域-文本匹配模块利用区域-文本对齐损失优化匹配分数, 使区域视觉特征与文本嵌入在特征空间中逐渐对齐; 同时, 将视觉特征经 MLP 降维后结合已知类样本特征库, 引入视觉特征对比损失, 以进一步聚拢已知类特征并构造更加清晰的已知边界, 从而缓解特征混淆. 最后, 针对未知信息缺失的问题, 本文引入伪未知样本挖掘模块: 通过对区域文本最大匹配分数关于中间特征图$\mathbf{R}$求导, 获得特征目标层的梯度注意力图$\mathbf{G}$, 据此计算不确定性分数, 并结合视觉定位质量对其进行校准, 从而筛选出更具未知特征的伪未知样本, 进一步构建已知与未知之间的判别边界.

2.1 区域生成模块

目前, 大部分 OSOD 方法采用传统 RPN 模块产生大量候选框. 然而, 由于 RPN 采用基于二元分类的目标性预测机制, 模型在训练过程中容易对已知类别产生偏置, 更倾向于保留已知类目标而忽视未知类别目标, 从而使未知类获得较低的目标性得分, 限制了 RPN 在开放集目标检测场景中的有效性, 特别是在[4,25] 中讨论的 OSOD 设置下这一问题尤为明显. 需要指出的是, 该问题并非源于 RPN 定位能力不足, 而是由于目标性分支隐式学习到了对已知类别的偏置, 使得具有较低语义置信但定位质量较高的未知目标在候选生成阶段容易被过早过滤.

为解决上述问题, 本文在 RPN 网络中并行引入一个独立的卷积分支, 用于计算候选区域的中心度得分, 其网络架构如图 3 所示, 并据此改进目标性得分. 中心度得分作为一种与类别语义无关的定位质量先验, 被用于调节候选框筛选过程, 结合目标的位置信息和形状特征优化候选框的几何特性, 从而缓解未知目标在早期阶段因目标性得分偏低而被过度抑制的问题, 提升其进入后续判别与不确定性建模阶段的概率.

$$S_{oa}=\sqrt{S_{cen}\times S_{obj}}. \tag{1}$$

其中, S_{oa}为改进后的目标性得分, S_{cen}为中心度得分, S_{obj}为原先的目标性得分. 需要强调的是, 中心度并不直接参与类别判别, 而是通过影响候选区域的

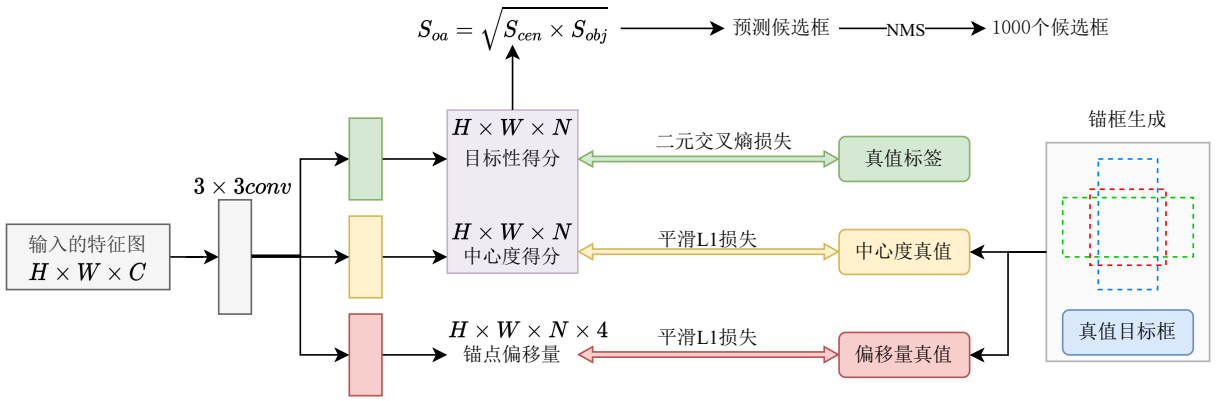


图3 区域生成模块

保留分布, 在区域生成层面缓解模型对已知类别的隐式偏向, 从而增强模型对未知目标的检测能力.

2.2 区域-文本匹配模块

在以往的开放集目标检测研究中, 大多数方法仅依赖视觉信息训练分类器, 而忽略了文本信息的重要性, 这种单一模态的依赖容易导致特征混淆, 尤其是在视觉特征和语义特征相近的类别之间, 模型难以有效区分已知类别与未知类别. 为解决这一问题, 本文引入基于提示学习 (Prompt Learning)[26] 的图像-文本对齐训练方法, 主要参考 CoOp 框架[27]. 相较于开放词汇目标检测引入大量类别文本, 本文为遵守开放环境的设定, 仅引入已知类文本信息, 使模型在检测时仍能够识别非已知类样本并将其标记为未知; 而开放词汇目标检测在推理时只能检测给定文本对应的类别, 这并不符合开放环境的设定.

与传统固定提示模板 (如“a photo of a …”) 不同, 本文将提示模板中的上下文单词替换为可学习参数, 从而能够动态优化提示内容以适应特定任务需求. 具体地, 本文将第j类的 prompt 参数定义为:

$$\mathbf{t}_j=\{v_1,v_2,\cdots,v_L,w_j\}. \tag{2}$$

其中, $v_1,v_2,\cdots,v_L$表示相同维度的可学习向量, w_j为第j类对应标签的文本嵌入.

文本编码器对$\mathbf{t}_j$进行特征编码, 输出文本特征T_j, 并与第i个候选框x_i的视觉特征形成区域-文本训练对$(F(x_i),T_j)$, 据此定义区域文本对齐损失为:

$$L_S=-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K+2}y_{ij}\log\frac{\exp(F(x_i)\cdot T_j/\tau)}{\sum_{c=1}^{K}\exp(F(x_i)\cdot T_j/\tau)}. \tag{3}$$

式中, y_{ij}是指示符, 若第i个候选框属于第j类, 则$y_{ij}=1$; 反之, $y_{ij}=0$. $F(x_i)\cdot T_j$为二者的余弦相似度, τ为温度系数.

尽管语义对齐能够通过文本和图像的交互形成语义簇, 但其主要依赖文本模态与视觉模态之间的直接对齐, 难以充分刻画不同视觉实例之间的潜在关系. 研究表明, 充分利用视觉表征之间的内在关联可以显著提升下游任务性能[28]. 因此, 本文进一步引入视觉表征增强的语义对比学习方法, 以优化模型表示能力. 具体而言, 本文通过 MLP 层将视觉特征映射到潜在空间, 生成 128 维潜在嵌入z_i. 这一过程不仅保留了视觉特征的丰富信息, 还将其压缩到低维且更具判别性的空间中. 本文采用 Han 等人[3] 提出的样本特征库实现入栈出栈操作, 并定义如下视觉特征对比损失:

$$L_V=\frac{1}{N}\sum_{i=1}^{N}L_V(z_i)$$

$$L_V(z_i)=-\frac{\nu}{|M_{c_i}|}\sum_{z_j\in M_{c_i}}\log\frac{\exp(z_i\cdot z_j/\varepsilon)}{\sum_{z_k\in\mathbf{M}\setminus M_{c_i}}\exp(z_i\cdot z_k/\varepsilon)}. \tag{4}$$

式中, N为批次大小, ν和ε为超参数, c_i为第i个候选框的类标签, M_{c_i}为类别c_i的样本特征库, $\mathbf{M}$为总样本特征库, $\mathbf{M}\setminus M_{c_i}$为去除$M_{c_i}$后的样本特征库. 样本特征库中包含该类别相似度最低的样本特征, 据此拉近同类样本边缘视觉表征的距离, 聚拢已知类特征, 并推远不同类别样本的视觉表征距离.

从功能上看, 区域–文本对齐损失主要在语义层面对视觉特征施加类别级约束, 有助于在跨场景或分布偏移条件下稳定已知类别的语义结构; 视觉特征对比损失则直接作用于视觉表征空间, 通过建模不同实例之间的相对关系, 弥补仅依赖跨模态对齐时对实例级视觉差异刻画不足的问题. 二者在语义约束与视觉判别两个层面形成互补, 共同优化特征空间中的类别分布, 从而增强类内特征紧凑性、提升类间可分性, 并为未知类别预留更加清晰的分布区域.

2.3 伪未知样本挖掘模块

在开放集场景中, 由于训练阶段缺乏未知类别

数据, 如何在仅利用已知类别样本的条件下有效筛选潜在未知目标, 成为开放集目标检测中的关键问题. 现有方法通常基于不确定性估计策略 (如证据不确定性[7]) 挖掘伪未知样本. 然而, 传统不确定性计算方法在本文设置下存在一定局限性: 一方面, 本文的类别得分由区域–文本匹配分数与两个全连接层输出共同构成, 其分布形式与证据不确定性假设下的得分分布存在明显差异; 另一方面, 由于采用多预训练编码器架构, 骨干网络在大规模已知类别数据上充分训练后, 模型注意力机制容易偏向已知类别特征, 从而削弱对潜在未知目标的响应能力.

Chen 等的研究[29]表明, 对于分布内 (In-Distribution, ID) 样本, 网络注意力通常集中于与最终分类结果高度相关的判别特征, 使得大量零贡献特征对应的梯度占据主导; 而对于分布外 (Out-of-Distribution, OOD) 样本, 由于模型缺乏稳定的注意力分布, 非零贡献特征的梯度响应更为显著. 因此, 对特征梯度进行聚合分析能够在 ID 与 OOD 样本之间形成有效区分, 从而提升分布外检测性能.

受此启发, 本文提出一种基于归因梯度[24]的不确定性分析方法, 通过量化输入特征对最终预测分数的敏感度, 对模型判别行为进行可解释建模. 具体而言, 本文针对文本引导下的视觉特征进行可信度估计, 利用特征梯度刻画候选区域的不确定性水平.

需要说明的是, 本文中的伪未知样本并非真实未知类别数据, 而是指在仅包含已知类别标注的训练数据中, 模型在判别过程中表现出较高不确定性的候选区域. 这类样本通常位于已知类别的判别边界附近, 在特征空间中难以被模型稳定归类, 可作为约束已知–未知判别边界的重要信号.

在训练阶段, 伪未知样本以在线方式动态挖掘, 其流程包括: 基于第二阶段模型对候选区域进行预测, 结合区域–文本匹配得分与中间特征表示; 随后通过归因梯度与定位质量校准评估候选区域的不确定性; 最后从前景与背景候选框中选取不确定性值最高的样本, 作为伪未知目标参与后续训练.

如图 2 所示, 将中间特征层$\mathbf{R}$作为特定输入层, 对于第i个候选框x_i, 其视觉特征$F(x_i)$输出的得分记为s_{ij}. 为获得网络注意力主要集中的特征梯度, 本文选择最大输出分数对中间特征层求导:

$$G^i_{h\times w\times c}=\frac{\partial \max_{j\in K+2} s_{ij}}{\partial R_{h\times w\times c}}. \tag{5}$$

其中, h,w,c分别代表中间特征层的高度、宽度和通道数, $\max_{j\in K+2} s_{ij}$表示第i个候选框最终输出的所有类别得分中的最大值, K为已知类别数, 2 表示 1 类未知类和 1 类背景类.

基于特征梯度$G^i_{h\times w\times c}$, 可计算不确定性分数u_i. 为进一步明确不确定性分数, 本文将每个通道中的非零梯度绝对值进行聚合, 并对通道求和得到更稳定的不确定性度量, 定义如下:

$$u_i=\frac{1}{C}\sum_{c\in C}\Big(\sum_{h\in H}\sum_{w\in W} mask_{h\times w\times c}|G^i_{h\times w\times c}|\Big). \tag{6}$$

式中, $mask_{h\times w\times c}$为指示函数, 当$G^i_{h\times w\times c}\neq 0$时, $mask_{h\times w\times c}=1$; 反之为 0.

为进一步校准候选框区域不确定性, 本文从视觉定位质量角度进行分析. 对于未知目标而言, 候选框与真实标签框的交并比 (IOU) 是衡量其定位质量的重要指标. 一般来说, 当候选框与真实标签框的 IOU 越大时, 说明候选框包含已知目标的可能性越高, 其未知不确定性应越低; 反之, 当 IOU 越小时, 候选框与已知真实标签框差异较大, 不确定性也相应更高. 另一方面, 改进 RPN 模块为每个候选框生成目标性得分, 该得分反映候选框内包含目标的可能性. 因此, 当候选框 IOU 较小但目标性得分较高时, 其包含未知目标的可能性也更高. 由此, 候选框不确定性与 IOU 成反比, 与目标性得分成正比, 故设计视觉校准不确定性损失如下:

$$\begin{aligned}&\omega=(1-I_{\hat{a}_i,a_{gt}})\cdot S_{oa}\\&L_\omega=\frac{1}{N}\sum_{i=1}^{N}-\omega\cdot\log(1-u_i)-(1-\omega)\log(u_i).\end{aligned} \tag{7}$$

式中, $I_{\hat{a}_i,a_{gt}}$为候选框与真实标签框的 IOU, u_i为第i个候选框目标的不确定性值.

在仅含已知类信息的数据集中, 当候选框x_i具有较高的u_i时, 说明其包含未知类目标的可能性越高. 因此, 本文从前景和背景候选框中分别选取不确定性值最高的k个候选框作为伪未知目标, 据此构建已知与未知的判别边界, 其损失函数定义为:

$$L_U=-\frac{1}{2k}\sum_{i=1}^{2k}\log\frac{\exp(s_u)}{\sum_{j=1}^{K+2}\exp(s_{ij})-\exp(s_{gt})}. \tag{8}$$

式中, s_{gt}为第i个候选框的真值输出分数, s_u为未知概率, 实验中设定$s_u=(1-s_{gt})\cdot s_{gt}$.

2.4 总体损失

为保持 RPN 的无类别分类效果, 本文将其单独训练作为第一阶段, 旨在优化候选框生成并避免引入类别信息. 其损失函数设计如下:

$$L_{stage1}=L_{obj}+L_{cen}+L_{reg}. \tag{9}$$

其中, L_{obj}为目标性得分的二分类损失, L_{cen}为中心度分数计算的 L1 损失, L_{reg}为定位回归损失.

第二阶段的主要目标是训练视觉特征与文本嵌入在特征空间中匹配, 其损失函数定义为:

$$L_{stage2} = L_{reg} + L_S + L_V. \tag{10}$$

第三阶段的主要目标是挑选伪未知目标并优化未知类得分, 其损失函数定义为:

$$L_{stage3} = L_{reg} + \lambda_t(L_U + L_\omega) + L_S. \tag{11}$$

式中, $\lambda_t = \exp(\log(\lambda)\cdot(1-t/T)) \in [\lambda, 1]$. 随着当前迭代轮次$t$逐渐增加并达到最大轮次$T$时, λ_t增大至 1.

这种设计使得模型在第一阶段以学习无类别的区域生成模块以主要任务, 第二阶段以学习区域文本匹配为主要任务, 建立清晰的已知边界, 随后在第三阶段逐渐建立起清晰的已知类和未知类边界.

3　实 验

3.1　实验设置和细节

(1) 数据集设置

本文采用 PASCAL VOC[30] 与 MS COCO[31] 构建 OSOD 基准测试数据集 (参考 Han 等人[3] 的工作). 其中, VOC 的 trainval 集合用于闭集训练; COCO 中选取 20 个 VOC 类别与 60 个非 VOC 类别, 用于评估方法在不同开放集条件下的表现. 定义两种实验设置: VOC-COCO{T1, T2}. T1 设置通过逐步增加开放集类别构建三个联合数据集, 均包含$n = 5000$张 VOC 测试图像, 及分别含$\{20, 40, 60\}$个非 VOC 类别的$\{n, 2n, 3n\}$张 COCO 图像; T2 设置通过提升荒野值 (Wilderness Ratio, WR) 构建四个联合数据集, 均包含n张 VOC 测试图像, 及$\{0.5n, n, 4n\}$张与 VOC 类别无交集的 COCO 图像.

(2) 训练参数设置

实验在 CUDA 12.2 环境下使用 NVIDIA 2080 Ti GPU 训练, 因网络参数多经预训练初始化, 故采用较小学习率并分三阶段进行: 第一阶段 (区域生成模块学习) 训练改进后 RPN 网络, 批次大小=4, 学习率=0.02; 第二阶段 (区域文本匹配模块训练) 学习视觉-文本特征匹配, 优化已知类特征空间为微调铺垫, 批次大小=1, 学习率=0.0002, 最大迭代次数=70000次; 第三阶段 (已知-未知边界微调) 建立明确判别边界, 批次大小=1, 学习率=0.0002, 最大迭代次数=70000次, 采用多步长学习率调整, 第 55000/65000 轮次分别降至原值 1/10, 加速收敛并提升泛化能力.

其他超参数设置: 区域-文本匹配模块温度系数$\tau = 0.5$, 视觉对比损失权重ν与对比学习温度ε均为 1, 整体损失中不确定性损失权重$\lambda = 10^{-4}$, 区域分数融合比例$\beta = 0.5$.

测试阶段为提升未知类检测性能, 最终检测分数计算公式如下:

$$p_{ij} = softmax(s_{ij})^{\beta} \cdot sigmoid(S^{i}_{oa})^{1-\beta}. \tag{12}$$

式中, s_{ij}为模型输出得分, S_{oa}为区域生成模块得分, β为区域分数融合比例.

3.2　OSOD 评价指标

1) 理想的开放集目标检测器应当在不受新增数据干扰的情况下, 保持对已知类别的识别能力. 为此, 本文提出通过未知影响指数 (Wilderness Impact, WI) 评估模型性能——在固定召回率条件下, 量化未知类别样本对已知目标检测精度的影响, 以此衡量检测器的开放集鲁棒性.

$$\mathrm{WI} = \left(\frac{\mathrm{P}K}{\mathrm{P}K \cup U} - 1\right) \times 100. \tag{13}$$

式中, P_K为闭集下的准确率, $P_{K\cup U}$为开集下的准确率.

2) 绝对开集误差 (Absolute Open-Set Error, AOSE) 为未知类被误判为已知类的候选框数量.

3) AP_U为未知平均精度, 用于评估模型对未知类别的检测性能.

4) mAP_K为已知类的平均精度, 用于评估模型对已知类的检测性能.

3.3　实验分析

如表 1 和表 2 所示, 本文提出的方法在多个指

表1　VOC 和 VOC-COCO-T1 的实验结果

Method	骨干网络	VOC	VOC-COCO-20				VOC-COCO-40				VOC-COCO-60			
		$\mathrm{mAP}_{\mathcal{K}\uparrow}$	$\mathrm{WI}_\downarrow$	$\mathrm{AOSE}_\downarrow$	$\mathrm{mAP}_{\mathcal{K}\uparrow}$	$\mathrm{AP}_{\mathcal{U}\uparrow}$	$\mathrm{WI}_\downarrow$	$\mathrm{AOSE}_\downarrow$	$\mathrm{mAP}_{\mathcal{K}\uparrow}$	$\mathrm{AP}_{\mathcal{U}\uparrow}$	$\mathrm{WI}_\downarrow$	$\mathrm{AOSE}_\downarrow$	$\mathrm{mAP}_{\mathcal{K}\uparrow}$	$\mathrm{AP}_{\mathcal{U}\uparrow}$
FR-CNN[34](2016)	R+F	80.10	18.39	15118	58.45	0.00	22.74	23391	55.26	0.00	18.49	25472	55.83	0.00
PROSER[35](2021)	R+F	79.68	19.16	13035	57.91	10.92	24.15	19831	54.66	7.62	19.64	21322	55.20	3.25
ORE[2](2021)	R+F	79.80	18.18	12811	58.25	2.60	22.40	19752	55.30	1.70	18.35	21415	55.47	0.53
DS[36](2018)	R+F	80.04	16.98	12868	58.35	5.13	20.86	19775	55.31	3.39	17.22	21921	55.77	1.25
OpenDet[3] (2022)	R+F	80.02	14.95	11286	58.75	14.93	18.23	16800	55.83	10.58	14.24	18250	56.37	4.36
AKCR[37](2024)	R	78.06	**9.55**	8267	58.52	18.45	**11.89**	14057	56.10	12.56	**10.96**	19153	56.47	5.10
Ours	R	**80.20**	11.96	**8174**	**59.36**	**22.48**	14.70	**11862**	**57.06**	**18.93**	11.74	**13122**	**57.28**	**11.56**

表2 VOC-COCO-T2 的实验结果

Method	骨干网络	VOC-COCO-0.5n				VOC-COCO-n				VOC-COCO-4n			
		$WI_{\downarrow}$	$AOSE_{\downarrow}$	$mAP_{\mathcal{K}\uparrow}$	$AP_{\mathcal{U}\uparrow}$	$WI_{\downarrow}$	$AOSE_{\downarrow}$	$mAP_{\mathcal{K}\uparrow}$	$AP_{\mathcal{U}\uparrow}$	$WI_{\downarrow}$	$AOSE_{\downarrow}$	$mAP_{\mathcal{K}\uparrow}$	$AP_{\mathcal{U}\uparrow}$
FR-CNN[34](2016)	R+F	9.25	6015	77.97	0.00	16.14	12409	74.52	0.00	32.89	48612	63.92	0.00
PROSER[35](2021)	R+F	9.32	5105	77.35	7.48	16.65	10601	73.55	8.88	34.60	41569	63.09	11.15
ORE[2](2021)	R+F	8.39	4945	77.84	1.75	15.36	10568	74.34	1.81	32.40	40865	64.59	2.14
DS[36](2018)	R+F	8.30	4862	77.78	2.89	15.43	10136	73.67	4.11	31.79	39388	63.12	5.64
OpenDet[3] (2022)	R+F	6.44	3944	**78.61**	9.05	11.47	8366	**75.16**	12.58	26.69	32419	**65.55**	16.76
AKCR[37](2024)	R	—	—	—	—	9.51	6875	73.47	14.13	22.31	27362	64.28	17.77
Ours	R	**4.67**	**2660**	76.51	**13.25**	**9.22**	**5598**	73.14	**19.49**	**22.81**	**22077**	62.91	**21.41**

标上显著提升了性能, 达到了当前最优水平. 表中加粗数据为最优结果, 下划线数据为次优结果, 指标中↓表示值越低越好, ↑表示值越高越好. 表中, R 为 ResNet50[32], F 为金字塔网络 (Feature Pyramid Networks, FPN)[33].

(1) VOC 和 VOC-COCO-T1 实验分析

从表 1 可见, 尽管本文方法是为开集检测设计, 但在 VOC 闭集数据集测试中, 其性能优于当前最优的二阶段目标检测器 Faster R-CNN(80.20 vs. 80.10). 对于开放集检测, 本文方法在逐步增加未知类的 VOC-COCO-20、VOC-COCO-40 和 VOC-COCO-60 数据集上, 未知类检测性能显著提升. AP_U 分别增长了 50.57%、78.92%、165.14%, 验证了方法在开放集环境下的鲁棒性, 即使未知类别数量增加, AP_U 依然保持稳定, 未如其他方法急剧下降. 与此同时, 以 VOC-COCO-20 为例, AOSE 下降了 27.57%, 表明误分类的未知实例减少. WI 下降了 17.56%, 进一步验证了方法的鲁棒性. 在提升未知类检测性能的同时, 检测已知类别的性能也未下降, mAP_K 提升了 1.03%-2.2%. 相比之下, OpenDet 在闭集和开放集指标上均落后于本文方法, 充分证明了本文方法的优越性.

与同样使用文本信息的 AKCR[37] 方法相比, 本文方法在闭集检测的性能 (80.20 vs. 78.06) 以及其他指标上均表现更好. 尽管如此, 本文方法的 WI 值稍低, 实际上是因为 AKCR 方法在已知类上的准确率低, 导致式 (13) 计算出的 WI 值降低.

(2) VOC-COCO-T2 实验分析

在表 2 中, 随着未知类目标图片数量的增加, 本文方法相较于 OpenDet, 在开放集性能上, AP_U 增长了 27.75%-54.93%, AOSE 下降了 31.90%-33.09%, 误分类的未知实例数量进一步减少. WI 下降了 14.54%-27.48%, 显示出在未知类检测上的优越性. 然而, mAP_K 下降了 2.60%-4.00%, 随着未知图片数量的增加, mAP_K 的下降幅度加大. 主要原因有二: 一是由于预训练图片编码器的骨干网络缺乏金字塔网络, 导致在长尾目标和困难小样本检测上存在局限性; 二是改进后的区域生成模块框选了更多包含未知类的候选框, 从而减少了已知类候选框的比例, 导致已知类的平均精度下降.

3.4 消融实验

为验证本文方法在开放集目标检测中的有效性, 所有消融实验均在相同网络结构、实验设置及训练阶段下进行, 性能比较基于 VOC-COCO-20 数据集.

(1) 微调阶段总体分析

区域-文本匹配模块对应的区域文本对齐损失 L_S 为网络基础分类损失, 本节通过第三阶段微调的消融实验验证各个损失函数的必要性, 结果如表 3 所示. 第二阶段模型通过 L_{stage2} 优化已知类特征空间, 实现类间分离与类内聚拢, 第三阶段仅用 L_S 优化时, 模型缺乏未知类检测能力.

表3 各个损失在第三阶段的消融实验结果

L_S	L_V	L_U	L_ω	$WI_{\downarrow}$	$AOSE_{\downarrow}$	$mAP_{K\uparrow}$	$AP_{U\uparrow}$
✓				9.55	17556	58.77	0.00
✓	✓			**5.74**	56339	58.46	0.00
✓			✓	8.99	17254	**59.44**	9.09
✓		✓		12.12	8339	59.20	22.16
✓	✓	✓	✓	12.39	8485	59.15	21.81
✓		✓	✓	11.96	**8174**	59.36	**22.48**

仅用 L_S 和 L_V 优化时, AOSE 显著增加、mAP_K 下降, 表明模型对已知类别过拟合, 削弱未知类区分能力; 加入 L_ω 后, 模型减少绝对开放集误差并提升未知类检测性能, mAP_K 和 AP_U 显著高于仅用 L_S 的模型 (如表 3 第 3 行所示).

仅用 L_U 和 L_S 优化时, 模型性能优于仅用 L_S, 但低于最优模型 (如表 3 第 4 行所示), 核心原因是未通过 L_ω 校准不确定性, 导致已知-未知类边界混淆, 印证 L_ω 的必要性. 将 L_V 加入最优模型后, 性能反而下降 (如表 3 第 5 行所示), 因第二阶段已实现类内聚拢, 第三阶段过度强化 L_V 会压缩未知类空间, 且与不确定性模块所需分散性冲突, 导致开放集性能

降低.

结合表 3 结果, 各组件必要性验证如下: 第三阶段通过L_S和L_U可获得基础已知/未知分类效果, 加入L_ω进一步提升性能, 而L_V在此阶段不利于模型优化, 充分验证了本文方法的有效性.

(2) 不确定性量化方法对比实验

为验证本文提出的归因梯度不确定性量化方法优于传统证据不确定性方法, 本节对比不同不确定性计算方式对实验结果的影响, 分析基于证据、概率、归因梯度的三种不确定性方法在开放集目标检测中的性能 (数据如表 4 所示).

表4　不确定性量化方法的性能对比结果

方法	$WI_{\downarrow}$	$AOSE_{\downarrow}$	$mAP_{K\uparrow}$	$AP_{U\uparrow}$
基于证据的不确定性估计	9.27	13980	58.01	21.03
基于概率的不确定性估计	**9.13**	13948	58.23	22.06
基于归因梯度的不确定性估计	11.96	**8174**	**59.36**	**22.48**

结果表明, 基于归因梯度的不确定性方法表现最优: AOSE 误差减少约 41.5%(从 13980 降至 8174), 有效降低未知类样本误分类率; 与证据不确定性方法相比, AP_U提升约 6.9%(从 21.03 提升至 22.48), mAP_K提升约 2.3%(从 58.01 提升至 59.36), 印证该方法在已知类与未知类检测性能上均优于其他方法. 此外, 基于概率的不确定性方法虽 WI 值较低, 但其核心原因是 P_K(已知类预测准确率) 偏低, 并非方法本身的不确定性量化能力更优.

(3) 区域生成模块有效性验证实验

为验证本文区域生成模块的有效性, 本节在相同实验设置下对比测试原始 RPN 模块[34] 与本文模块, 结果如表 5 所示.

表5　RPN 改进前后的对比实验结果

方法	$WI_{\downarrow}$	$AOSE_{\downarrow}$	$mAP_{K\uparrow}$	$AP_{U\uparrow}$
原始的RPN模块	13.80	9090	56.09	20.28
区域生成模块	**11.96**	**8174**	**59.36**	**22.48**

实验结果表明, 原始 RPN 模块的未知样本筛选性能远低于本文区域生成模块: AOSE 增加 916, 未知样本筛选误判率显著升高; mAP_K 降低 5.51%、AP_U 下降 9.79%, 已知类与未知类检测性能均大幅下滑, 印证其在开放环境下的局限性.

相比原始 RPN 模块, 区域生成模块在减少未知样本误判的同时, 保持了已知类检测性能的稳定性, 且未知类检测性能有所提升, 表明其设计更适配开放环境下的目标检测任务.

(4) 区域-文本匹配模块比较实验

为验证区域-文本匹配模块的有效性, 本文在第二阶段实验中将该模块替换为普通全连接层网络并进行对比. 需注意的是, 此阶段网络尚不具备区分已知类别与未知类别的能力. 实验结果如表 6 所示: 与普通全连接层网络相比, 使用区域-文本匹配模块后, AOSE 显著下降 (从 66942 降至 25377), mAP_K 明显提升 (从 53.28% 提升至 56.86%), 表明该模块能有效缓解已知与未知特征混淆, 提升已知类别的区分能力, 同时更好地保留潜在未知区域.

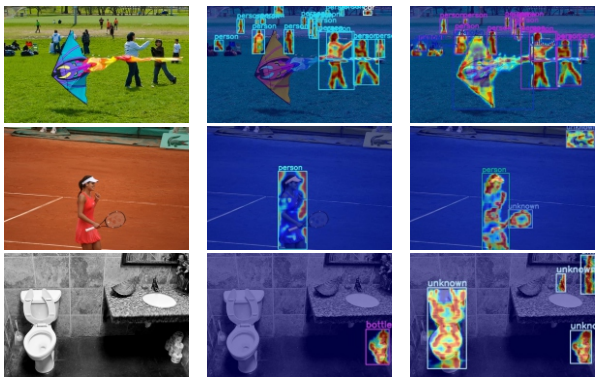
表6　区域文本匹配模块对比实验

方法	$WI_{\downarrow}$	$AOSE_{\downarrow}$	$mAP_{K\uparrow}$	$AP_{U\uparrow}$
全连接层网络	**6.14**	66942	53.28	0.00
区域-文本匹配模块	6.2	**25377**	**56.86**	0.00

3.5　可视化

(1) 特征可视化注意力热图

为进一步验证本文各模块的有效性, 将模型输出特征进行可视化, 结果如图 4 所示, (a) 为原图, (b) 为本文方法第二阶段后模型输出特征的注意力可视化结果, (c) 为第三阶段后模型输出特征的注意力图可视化结果. 可见, 本文方法在第二阶段仅能关注已知类, 这也是部分开放词汇目标检测的共性局限性, 而经第三阶段微调后, 模型能明显将注意力聚焦于未知类, (c) 图中不仅关注已知类, 同时对未知类给予了相当的注意力. 图 4 的可视化结果充分印证了本文伪未知样本挖掘方法的优越性.

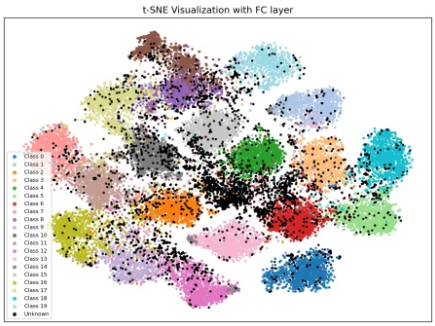


(a) 原图　(b) 第二阶段　(c) 第三阶段

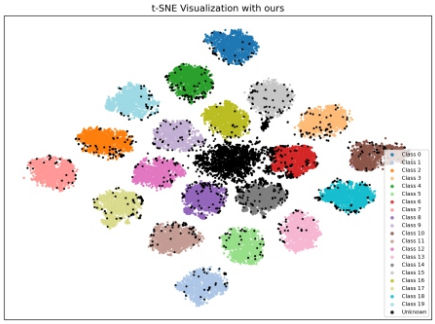
图4　特征可视化注意力热图

(2) 区域文本匹配模块特征可视化

如图 5 所示, 本文对第二阶段实验中基于全连接层 (softmax 损失) 分类与区域-文本匹配模块 (引入文本模态、替换全连接层) 的特征分布进行可视化对比, 特征为 ROI head 最后一层. 图中彩色点表示已知类别, 黑色点表示未知类别. 结果显示, 基于全连接层网络的特征中, 同一已知类别未完全聚拢、分



(a) 全连接层网络



(b) 区域文本匹配模块

图5 特征分布 tsne 可视化图

布分散, 且未知类别 (黑色点) 分布同样散乱; 相比之下, 区域-文本匹配模块的特征可视化结果中, 每个已知类别的特征几乎完全聚拢为单簇, 未知类别虽仍呈分散状态, 但其分布范围显著小于全连接层网络的情况. 这表明该模块能更有效地增强已知类别的语义一致性, 同时在一定程度上约束未知类别分布.

(3) 检测结果可视化

本节可视化检测结果如图 6 所示, 图中红色标注框表示未知类, 白色标注框表示已知类. 本文方法主要解决开放集目标检测中的语义混淆与视觉混淆问题, 尤其能有效降低视觉特征和语义特征上与未知类相似的已知类被误分类的概率, 同时能准确框定每个未知类, 后两行结果显示, 本文方法成功检测到其他方法未识别的未知目标并精准分类, 减少了其他方法在未知类定位中存在的漏检或定位不准确问题. 如图 6 第一行所示, 本文方法准确定位并检测出未知类“熊”, 而 OpenDet 不仅错误定位其位置, 还误识别为“狗”, 如图 6 第 4 行所示, 本文方法准确检测到未知类公园长椅, 其他方法均未检测该目标.

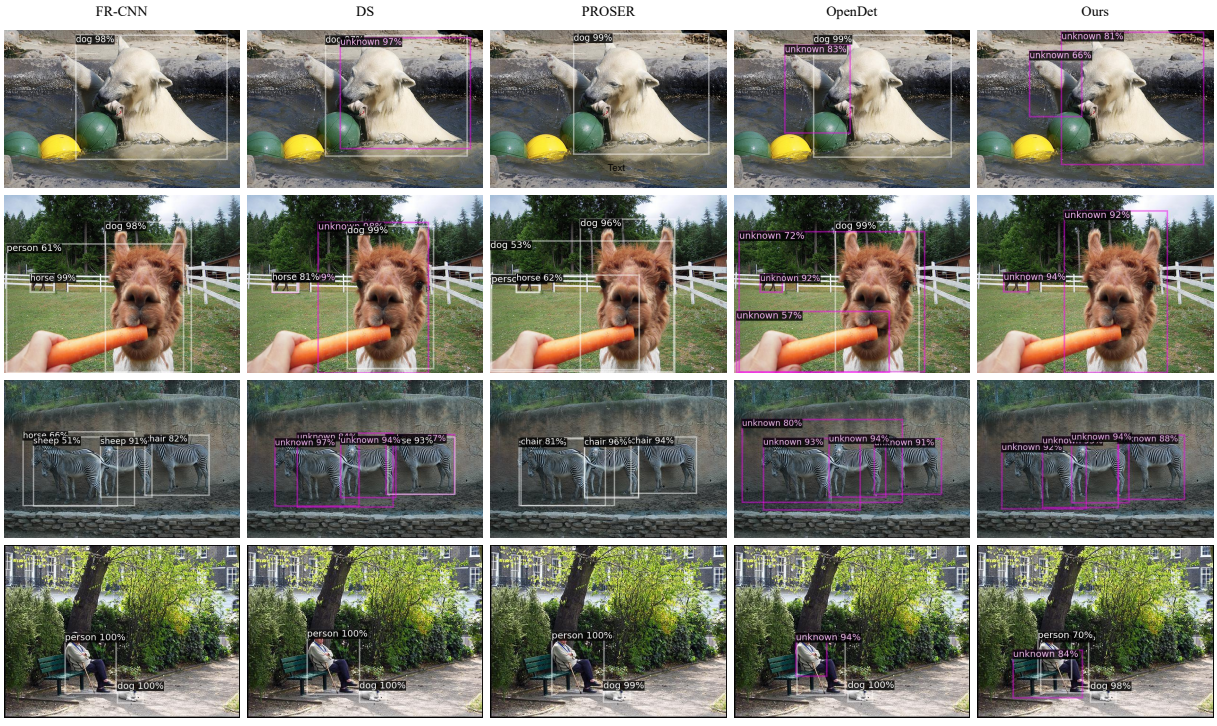


图6 检测结果可视化

4 结 论

本文围绕开放集目标检测任务, 提出了一种基于区域–文本匹配的检测框架, 旨在缓解已知类别与未知类别在特征表示空间中的混淆问题. 针对开放集场景下未知目标难以建模的挑战, 本文设计了区域–文本匹配模块与基于不确定性的伪未知样本挖掘模块, 两者相互协作, 共同提升模型对未知目标的检测能力.

具体而言, 为缓解视觉特征与语义特征相近所导致的类别误分类问题, 本文通过引入文本模态, 构建区域–文本匹配机制, 从多模态角度增强已知类别的结构化表达, 并在特征空间中拉开已知与未知目

标的分布间隔. 进一步地, 针对训练阶段缺乏真实未知样本的问题, 本文从模型判别行为出发, 联合建模基于归因梯度的特征不确定性与基于视觉定位质量的不确定性, 挖掘高质量伪未知样本, 从而构建更加清晰且自适应的已知–未知判别边界, 有效提升了未知目标检测的准确性与鲁棒性.

尽管本文方法在开放集检测场景下取得了显著性能提升, 但仍存在一定局限性. 首先, 区域–文本匹配模块依赖预训练视觉语言模型, 其固有的语义偏置可能在特定场景下影响特征分布结构; 其次, 基于归因梯度的不确定性建模在训练阶段引入了额外的反向传播计算, 导致训练开销有所增加; 此外, 在跨域场景或分布偏移较大的环境中, 模型性能仍存在一定下降空间. 未来工作将重点探索更加鲁棒的跨域特征对齐与不确定性建模策略, 并进一步将该方法扩展至更大规模类别集合及更复杂的多模态感知任务中.

参考文献 (References)

[1] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]. Proceedings of the 37th International Conference on Machine Learning. Virtual, 2020: 1597-1607.

[2] Joseph K J, Khan S, Khan F S, et al. Towards open world object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, 2021: 5830-5840.

[3] Han J M, Ren Y Q, Ding J, et al. Expanding low-density latent regions for open-set object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, 2022: 9591-9600.

[4] Zhou Z X, Yang Y F, Wang Y, et al. Open-set object detection using classification-free object proposal and instance-level contrastive learning[J]. IEEE Robotics and Automation Letters, 2023, 8(3): 1691-1698.

[5] Dhamija A R, Gunther M, Ventura J, et al. The overlooked elephant of object detection: Open set[C]. IEEE Winter Conference on Applications of Computer Vision. Snowmass Village, 2020: 1010-1019.

[6] Miller D, Nicholson L, Dayoub F, et al. Dropout sampling for robust object detection in open-set conditions[C]. IEEE International Conference on Robotics and Automation. Brisbane, 2018: 3243-3249.

[7] Han J W, Chen Y. Pseudo-unknown uncertainty learning for open set object detection[J]. Knowledge-Based Systems, 2024, 303: 112414.

[8] Zheng J Y, Li W H, Hong J, et al. Towards open-set object detection and discovery[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans, 2022: 3961-3970.

[9] Wu Z H, Lu Y, Chen X Y, et al. UC-OWOD: Unknown-classified open world object detection[C]. Computer Vision — ECCV 2022. Tel Aviv, 2022: 193-210.

[10] Zhao X W, Ma Y Q, Wang D R, et al. Revisiting open world object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(5): 3496-3509.

[11] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]. Proceedings of the 38th International Conference onMachine Learning. Virtual, 2021: 8748-8763.

[12] Zareian A, Rosa K D, Hu D H, et al. Open-vocabulary object detection using captions[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, 2021: 14393-14402.

[13] Zhou X Y, Girdhar R, Joulin A, et al. Detecting twenty-thousand classes using image-level supervision[C]. Computer Vision — ECCV 2022. Tel Aviv, 2022: 350-368.

[14] Zhong Y W, Yang J W, Zhang P C, et al. RegionCLIP: Region-based language-image pretraining[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, 2022: 16793–16803.

[15] Li L H, Zhang P C, Zhang H T, et al. Grounded language-image pre-training[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, 2022: 10965-10975.

[16] Gao M F, Xing C, Niebles J C, et al. Open vocabulary object detection with pseudo bounding-box labels[C]. EuropeanConference on Computer Vision. Tel Aviv, 2022: 266-282.

[17] 王杰, 周志杰, 胡昌华, 等. 不确定性信息表示及推理[J]. 控制与决策, 2023, 38(10): 2749-2763.
(Wang J, Zhou Z J, Hu C H, et al. Expression and inference of uncertain information[J]. Control and Decision, 2023, 38(10): 2749-2763.)

[18] Hendrycks D, Basart S, Mazeika M, et al. Scaling out-of-distribution detection for real-world settings[C]. International Conference on Machine Learning. Baltimore, 2022: 8759-8773.

[19] Sun Y, Guo C, Li Y. ReAct: Out-of-distribution detection with rectified activations[C]. Advances in Neural Information Processing Systems 34. Virtual, 2021: 144-157.

[20] Vaze S, Han K, Vedaldi A, et al. Open-set recognition: A good closed-set classifier is all you need[C]. Proceedings of the 10th International Conference on Learning Representations. Virtual, 2022: 1647.

[21] Holub A, Perona P, Burl M C. Entropy-based active learning for object recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition Workshops. Anchorage, 2008: 1-8.

[22] Chan R, Rottmann M, Gottschalk H. Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation[C].

IEEE/CVF International Conference on Computer Vision. Montreal, 2021: 5108-5117.

[23] Liu W, Wang X, Owens J D, et al. Energy-based out-of-distribution eetection[C]. Advances in Neural Information Processing Systems 33. Vancouver, 2020: 21464-21475.

[24] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]. IEEE International Conference on Computer Vision. Venice, 2017: 618–626.

[25] Kim D, Lin T Y, Angelova A, et al. Learning open-world object proposals without learning to classify[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 5453-5460.

[26] Liu P F, Yuan W Z, Fu J L, et al. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing[J]. ACM Computing Surveys, 2023, 55(9): 1-35.

[27] Zhou K Y, Yang J K, Loy C C, et al. Learning to prompt for vision-language models[J]. International Journal of Computer Vision, 2022, 130(9): 2337-2348.

[28] Sun Y Y, Li Y X. OpenCon: Open-world contrastive learning[J/OL]. 2023, arXiv: 2208.02764.7.

[29] Chen J G, Li J J, Qu X Y, et al. GAIA: Delving into gradient-based attribution abnormality for out-of-distribution detection[C]. Advances in Neural Information Processing Systems 36. New Orleans, 2023: 1-10.

[30] Everingham M, van Gool L, Williams C K I, et al. The pascal visual object classes challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[31] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common objects in context[C]. European Conference on Computer Vision. Zurich, 2014: 740-755.

[32] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 770-778.

[33] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 936-944.

[34] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[35] Zhou D W, Ye H J, Zhan D C. Learning placeholders for open-set recognition[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, 2021: 4401-4410.

[36] Miller D, Nicholson L, Dayoub F, et al. Dropout sampling for robust object detection in open-set conditions[C]. 2018 IEEE International Conference on Robotics and Automation. Brisbane, 2018: 3243-3249.

[37] Sarkar H, Chudasama V, Onoe N, et al. Open-set object detection by aligning known class representations[C]. IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, 2024: 218-227.

作者简介

韩嘉雯 (2000−), 女, 硕士生, 主要研究方向为开放环境下的目标检测, E-mail: 6221905004@stu.jiangnan.edu.cn;

陈莹 (1976−), 女, 教授, 博士, 博士生导师, 主要研究方向为计算机视觉、模式识别、多媒体信息融合、深度模型压缩, E-mail: chenying@jiangnan.edu.cn.