## Abstract

## 【√】Introduction

近年来，学术界对于开放世界目标检测任务的兴趣逐渐增加，这是因为传统目标检测方法存在一些局限性。传统目标检测要求模型在测试时所遇到的目标类别都必须是在训练阶段见过的，这导致了繁琐且昂贵的数据标注工作。然而，现实世界中的情况往往更加复杂，新的目标类别可能会不断涌现，传统方法无法灵活适应。因此，开放世界目标检测成为一个备受学术关注的课题。研究人员致力于开发能够应对未见过目标类别的检测任务的方法，从而在更加开放和动态的环境下实现准确的目标检测。这一领域的研究不仅有助于解决现实世界中的实际问题，还推动了目标检测技术的创新与进步。

开放世界目标检测算法最初由【ore】提出，旨在解决一项重要任务：将在检测过程中遇到的未曾见过的类别判定为“未知类”，并且在保持对已知类别对象的检测能力的同时，逐步学习如何检测新的类别，以应对不断变化的现实环境，从而实现更加全面和可靠的目标检测能力。目前，许多先进的研究已经为开放世界目标检测领域做出了重要贡献。这些研究可以总结为两种主要趋势：一是在训练过程中不使用未知类伪标签，另一种是在训练过程中使用未知类伪标签。在前一种方法中，研究者通常需要仔细地配置模型，以确定将预测判定为未知类的阈值。有时，甚至需要设置多个阈值，这要求研究人员具备高水平的先验知识。这种方法虽然有效，但在一些情况下可能难以找到合适的阈值，且其准确性受到预设参数的影响。另一种方法则是在训练过程中使用未知类伪标签。这种方法通常基于有限的已知类标注数据，例如根据候选框的 objectness 得分等选择性地为某些候选框分配未知类伪标签。然而，这样的方法可能导致模型对于已知类产生偏见，同时也可能混淆模型对已知类和未知类的精确区分，从而影响了整体性能。

本文的主要目标在于改进目前基于伪标签的方法对已知类数据时可能存在的偏见问题，基于Featurized Query R-CNN，提出了一种简单而又有效的伪标签标注策略。具体而言，在训练过程中，对于那些与真实标注框（gt）没有匹配的候选建议框进行标注时，采用了一种独特的策略。首先，从这些未匹配的候选建议框中排除了最有可能包含已知类对象的一部分，因为这些候选框可能导致已知类偏见的问题。接着，它进一步排除了最有可能是背景的另一些候选建议框，以确保不会将背景误标注为未知类。在这两轮排除之后，剩余的候选建议框中，随机选择了k个候选框，并为它们打上“未知类”的标签。这一方法的核心思想是通过排除潜在的已知类和背景候选框，从而更加准确地标注未知类。这种选择性随机的标注方式有助于缓解模型对于已知类的偏见，同时确保了未知类的有效学习。通过随机选择一定数量的候选建议框并为其打上“未知类”的标签，模型可以在学习过程中逐步适应新的类别，实现对开放世界目标检测任务的更好应对能力。这种方法在解决基于伪标签方法的问题上提供了一个架构简单、新颖且有前景的解决方案。其次，在基于查询的目标检测中，虽然其端到端训练的优势明显，但是采用的一对一匹配方式存在监督信息不足的问题。实际上，大量研究已经确认了这一问题的存在。鉴于此，本文决定采用OTA方法，以一对多的标签分配方式来替代传统的匹配方式。OTA方法的核心思想在于，通过为每个目标真实标注框分配至少一个查询，为模型提供更加丰富的监督信息。这种一对多的标签分配策略能够更好地反映目标与查询之间的复杂联系，从而在模型训练中提供更有针对性的引导。通过这种方式，模型可以更准确地学习目标的特征和位置，从而提高检测性能。最后，正如在【Enhancing Your Trained DETRs with Box Refinement】一文中所指出的，基于查询的检测模型的定位能力使模型整体性能的瓶颈问题。为应对这一问题，本文选择将传统的定位相关损失函数如 l1 范数损失和基于 giou 的损失改为采用 mpDIoU 损失。通过引入mpDIoU损失函数，以提升模型在检测框定位方面的精确性。

## Related work

### 【√】Open world object detection

尽管传统的目标检测方法，如以Faster R-CNN为代表的两阶段目标检测算法、YOLO系列的单阶段目标检测算法，以及DETR系列的端到端目标检测算法，在封闭集目标检测数据集上表现出色，然而，它们往往难以应对多变的开放世界环境。在这种环境下，检测任务可能涉及到训练阶段从未涉足过的类别，这进一步加大了问题的复杂性。为了应对这一挑战，约瑟夫首次提出了开放世界目标检测的任务设定，并在此背景下基于Faster R-CNN框架提出了ORE算法。该算法通过在特征空间中进行聚类，选择背景proposals中对象性得分最高的作为未知类别的伪标签，同时采用基于能量的分类方法，使得模型具备检测未知类别对象的能力。随着时间的推移，许多杰出的研究逐渐深入探讨了这一领域。显然，对于这一任务而言，最为重要的挑战之一在于模型对未知类对象的识别能力。在此过程中，伪标签的应用日益引起关注，并在学术界受到广泛关注【一波文献】。{一波文献的展开} 其中，许多工作如【ORE, SA, UC-OWOD, OW-DETR, RandBox】都是基于 proposal box 特征是前景对象的可能性得分为基础，从而获得未知类伪标签。通常，这些方法将具有高 objectness 得分，但却不与任何 gt 对象匹配的 proposal box 作为潜在的未知类候选框；另外，还有一些工作例如【RE-OWOD, Open World DETR, CAT】，则通过融合额外的候选框生成方法（例如selective search算法）来协助选择潜在的未知类对象候选框；在【...“Brain”】中，通过预训练的大规模视觉语言模型来生成伪标签，而【Hyp-OW】则通过以 proposal box 特征到超类中心的距离为依据进行伪标签的标注。这些精心设计的伪标签标注方法，均在实践中取得了显著的效果。当然，还有一些方法没有采用伪标签策略，例如【OCPL】运用原型学习，致力于在特征空间中学习紧凑的特征表示。其核心思想在于，通过最小化特征空间中已知类和未知类分布的重叠程度，从而构建更具判别性的特征表示；【2B-OCD】通过训练一个以对象为中心的校准器，在预测时将背景中得分大于某一阈值的预测框预测为未知类；【OW-RCNN】方法则采用基于 IoU 的 objectness 得分和高斯混合模型，通过精心设置多个阈值，进行未知类感知分类，并最小化开放集误差；【PROB】方法则利用类别无关的高斯分布（均值和方差为可学习参数）对对象特征进行建模，并将某个 query 的分类结果表述为该query是一个前景对象的概率乘以该前景对象是某个类别的概率；【Annealing...】方法则将目标检测视为对象级特征交织的过程，在学习过程中使用 Label-Transfer Learning 范式，以解耦已知类和未知类特征，从而实现对已知类和未知类的区分能力。本文关注于基于伪标签的开放世界目标检测算法，旨在设计一种简单而有效的伪标签标注策略。

### 【√】Open vocabulary object detection

随着大规模预训练视觉语言模型的盛行，另一个开放环境目标检测设定为开放词汇目标检测（Open Vocabulary Object Detection，简称OVOD），以解决传统目标检测中词汇表限制的问题，使模型能够根据语言识别和检测未在训练数据中出现的目标类别。该领域工作借助预训练好的大规模视觉语言模型在特征空间中已经对齐了视觉信息和语言信息的优势，通常使用相应的文本嵌入特征作为预测分类结果的依据。该领域工作主要可以划分为train with extra data，knowledge distillation，region text align，prompting modeling这四个主流方向。OVR-CNN是OVOD领域的先驱性工作，使用的是train with extra data方法，其提出了一种两阶段的训练策略。首先，在第一阶段，该方法利用大量额外的图像-标题对来学习开放的视觉语义空间。随后，在第二阶段学习检测器时，通过利用第一阶段学到的开放视觉语义空间，使模型能够泛化到新类别上；GradOVD则通过从多个图像标注数据集中自动生成感兴趣物体的边界框标注，以扩展训练数据并提高模型性能。Detic则采用了联合训练的方法，将ImageNet21K分类图像数据集和目标检测数据集结合，共同用于模型的训练。UniDetector通过使用多各包含不同标签空间的数据集进行预训练，对齐图像和文本的特征空间。使用knowledge distillation的典型方法有【ViLD，OADP】，【ViLD】首次引入多模态模型，例如CLIP，通过对象级别的视觉-视觉知识蒸馏，来促进OVOD性能。【OADP】则认为仅从大型预训练模型中蒸馏对象级别信息忽略了不同对象之间的交互关系。为此，OADP提出了全局、块和对象三级金字塔蒸馏结构，以便在蒸馏过程中保留对象之间的上下文关系。VLDet、GLIP和RegionCLIP等方法则专注于对齐视觉区域和相应语言嵌入特征的对齐。其中 VLDeT 将对象-语言对齐表述为一组图像区域特征和一组单词嵌入之间的集合匹配问题；GLIP和RegionCLIP则通过进行region-word 级别的大规模预训练，从而提高模型 OVOD 的性能。DETR系列目标检测算法则开发了基于提示学习的 OVOD 算法。OV-DETR提出通过衡量条件输入与预测结果的匹配性来学习OVOD检测器；Prompt-OVD则通过引入RoI-based masked attention和RoI pruning技术，利用CLIP的零样本分类能力来优化检测性能，实现了更加高效的OVOD。CORA提出了Region prompting技术缓解 CLIP 全局到区域的分布差异，以及Anchor pre-matching技术通过类别感知匹配机制有助于学习通用的目标定位能力。OVOD检测时需要预定义感兴趣的检测对象，并且需要依赖于视觉语言预训练模型对视觉空间和语言空间的对齐能力，OVOD和OWOD是开放环境下相交的两个研究领域。

### 【√】Class-Agnostic Object Detection

开放世界类不可知对象检测的目标是基于有限数量的已知类训练数据集，培养一个能够在图像中捕捉所有前景物体的目标检测器，而无需进行类别区分。【2021，WACV】指出在实际应用的某些场景中，确定物体的存在和准确位置往往比识别特定类别更加重要，提出了类不可知的对象检测作为一个新兴问题，通过对抗性训练的方式，使模型能够定位所有目标，而无需考虑它们的具体类别。【OLN】的研究则认为，目前的目标检测使用的候选框提取方法中的二元分类器往往过于关注训练中的已知类类别，从而容易出现过拟合。因此，该研究选择采用仅考虑区域位置和形状与真实目标框的重叠程度（例如，中心度）来估计每个区域的目标相关性。在【2022，SIBGRAPI】的研究中，"Pixel-level class-agnostic object detection using texture quantization"采用卷积网络和纹理灰度量化来解决类别无关的目标检测任务。【LDET】提出了一种全新的数据增强方法，通过将标注的前景区域剪切并粘贴到合成的背景画布上，从而创造出“完全标记”的训练图像，期望以此达到检测所有前景对象的目的。【GOOD】的研究认为，传统基于RGB的检测模型主要依赖于外观相似性来检测新的物体，这容易导致对纹理和具有明显区分度的部分的过拟合。因此，该研究提出通过整合几何线索（如深度和法线信息），来挖掘潜在的新类别对象。

### Debiasing by Randomness

## Method

介绍owod设置

### Preliminaries

介绍 Featurized query R-CNN

### Our method（RandUnknown）

结合网络架构图，总体介绍本文方法。

#### Pseudo-label by random selection

说明随机选择的方法。

#### One-to-many label assignment via OTA

说明 OTA 用于标签分配的方法。

## Experiment

### Datasets

### Evaluation Metrics

### Implementation Details

### Main Results

### Ablation Experiments

### Additional Qualitative Results

## Conclusion

## References