计算机研究与发展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202440054
61(9): 2128−2141, 2024

开放世界物体识别与检测系统:现状、挑战与展望

聂　晖　王瑞平　陈熙霖
(中国科学院计算技术研究所　北京　100190)
(中国科学院大学　北京　100049)
(hui.nie@vipl.ict.ac.cn)

Open World Object Recognition and Detection Systems: Landscapes, Challenges and Prospects

Nie Hui, Wang Ruiping, and Chen Xilin
(*Institute of Computing Technology*, *Chinese Academy of Sciences*, *Beijing* 100190)
(*University of Chinese Academy of Sciences*, *Beijing* 100049)

Abstract　We explore the transition from closed environments to open world environments and its impact on visual perception (focusing on object recognition and detection) and the field of deep learning. In open world environments, software systems need to adapt to constantly changing conditions and demands, presenting new challenges for deep learning methods. In particular, open world visual perception requires systems to understand and process environments and objects not seen during the training phase, which exceeds the capabilities of traditional closed systems. We first discuss the dynamic and adaptive system requirements brought about by technological advances, highlighting the advantages of open systems over closed systems. Then we delve into the definition of the open world and existing work, covering five dimensions of openness: open set learning, zero-shot learning, few-shot learning, long-tail learning, and incremental learning. In terms of open world recognition, we analyze the core challenges of each dimension and provide quantified evaluation metrics for each task dataset. For open world object detection, we discuss additional challenges compared with recognition, such as occlusion, scale, posture, symbiotic relationships, background interference, etc., and emphasize the importance of simulation environments in constructing open world object detection datasets. Finally, we underscore the new perspectives and opportunities that the concept of the open world brings to deep learning, acting as a catalyst for technological advancement and deeper understanding of the realistic environment challenges, offering a reference for future research.

Key words　open world; visual perception; object detection; object recognition; evaluation metrics; simulation environment

摘　要　探究了从封闭环境到开放世界环境的转变及其对视觉感知(集中于物体识别和检测)与深度学习领域的影响.在开放世界环境中,系统软件需适应不断变化的环境和需求,这为深度学习方法带来新挑战.特别是,开放世界视觉感知要求系统理解和处理训练阶段未见的环境和物体,这超出了传统封闭系统的能力.首先讨论了技术进步带来的动态、自适应系统需求,突出了开放系统相较封闭系统的优势.接着,深入探讨了开放世界的定义和现有工作,涵盖开集学习、零样本学习、小样本学习、长尾学习、增量学习

收稿日期:2024-01-29;**修回日期**:2024-06-06
基金项目:科技创新2030—"新一代人工智能"重大项目(2021ZD0111901);国家自然科学基金项目(U21B2025, U19B2036)
This work was supported by the National Key Research and Development Program of China (2021ZD0111901) and the National Natural Science Foundation of China (U21B2025, U19B2036).
通信作者:王瑞平(wangruiping@ict.ac.cn)

等 5 个开放维度. 在开放世界物体识别方面，分析了每个维度的核心挑战，并为每个任务数据集提供了量化的评价指标. 对于开放世界物体检测，讨论了检测相比识别的新增挑战，如遮挡、尺度、姿态、共生关系、背景干扰等，并强调了仿真环境在构建开放世界物体检测数据集中的重要性. 最后，强调开放世界概念为深度学习带来的新视角和机遇，是推动技术进步和深入理解世界的机会，为未来研究提供参考.

关键词 开放世界；视觉感知；物体检测；物体识别；评价指标；仿真环境

中图法分类号 TP391

随着信息技术的飞速发展，开放世界物体识别与检测系统已成为现代社会不可或缺的一部分，广泛应用于教育、工业、医疗等众多领域. 这些系统不仅需要处理复杂的视觉数据，还要适应不断变化的环境和实时的动态场景. 伴随着深度学习、计算机视觉、边缘计算等新兴技术的突破，以及 GPU 加速卡、高速网络设备等新硬件的发展，开放世界物体识别与检测系统的设计和实现面临着前所未有的新挑战和机遇.

近年来，深度学习特别是在自动驾驶[1]、监控安全[2]、医疗影像分析[3]和对话系统[4]等领域取得了显著的进步. 在自动驾驶领域，深度学习已被成功应用于车辆环境感知和决策系统，提高了自动驾驶车辆的安全性和可靠性. 在监控安全方面，基于深度学习的方法已能有效识别和追踪监控视频中的行人和物体，极大地提高了公共安全的监控效率. 而在医疗影像分析领域，研究人员能够更准确地诊断疾病，如使用深度学习进行皮肤癌的早期检测和分类. 随着模型架构的不断创新，例如 Transformer[5] 和 GPT[6] 系列模型的出现，深度学习在处理语义理解和自然语言处理方面也取得了巨大的成功. 这些进展为开放世界物体识别与检测系统提供了强大的技术支持和灵感源泉.

面对越来越复杂的应用需求和深度学习方法的快速发展，研究人员和从业者开始探索如何更好地将深度学习方法应用于开放世界物体识别与检测系统. 这不仅涉及到对基本的物体识别和定位，还包括对复杂场景的深度理解、动态变化的适应能力的挑战，如图 1 所示.

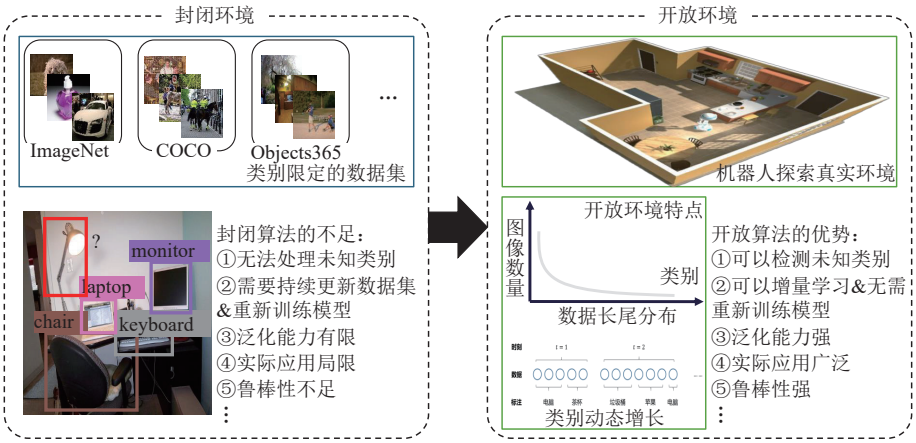


Fig. 1 Comparison between closed and open environments

图 1 封闭环境和开放环境的对比

本文中，封闭环境是指由用户预先定义、限定条件或者背景下的操作环境，其中的变量、条件和可能遇到的情景都是已知且可控的. 在封闭环境中，系统或模型面对的任务、输入数据的种类和范围都是事先设定好的，不会出现系统未曾训练或预备过的新情况和数据类型. 因此，封闭环境通常用于特定的、受限的应用场景，如工厂自动化、特定场景下的物体识别等. 开放环境则相反，指的是没有事先定义所有可能遇到的条件和情景的操作环境. 在开放环境中，系统或模型可能遇到未知的、新颖的或者在训练过程中未曾出现过的情况和数据类型. 这要求系统具有更高的适应性和泛化能力，能够处理和识别新的情景和物体. 开放环境更贴近于现实世界的复杂多变，如街道行驶的自动驾驶汽车、实时监控和响应的安全系统等. 本文将探讨在开放世界物体识别与检测系统中应用深度学习的现状、面临的主要挑战以

及未来的发展趋势，旨在为相关领域的研究和实践提供参考和启示.

1 从封闭走向开放

在信息时代的浪潮中，物体识别与检测系统经历了从封闭到开放的重要转变. 传统上，物体识别与检测系统依赖于封闭的、标注完备的数据集如 ImageNet[7]、COCO[8] 以及 Objects365[9] 等进行训练，这些数据集为系统提供了丰富而准确的标签信息. 然而，这种封闭环境的数据集存在明显的局限性，包括但不限于类别的封闭性、场景的静态性和环境的理想化，这使得模型难以应对现实世界的多样性和复杂性.

随着技术的发展，开放世界物体识别与检测系统[10-11] 应运而生，它要求模型能够识别和理解在训练阶段未出现过的环境和物体. 这种系统面对的挑战包括类别的动态增加、场景的实时变化以及环境的不确定性. 在开放世界条件下，物体识别与检测系统必须具备更高的适应性和鲁棒性，例如能够在复杂的办公或家居场景中，即使遇到未知物体或遭受视角、光照变化的影响，也能准确地完成物体的识别与检测任务.

2 国内外相关研究

开放世界物体识别是一个综合任务，包含开集物体识别、零样本物体识别、小样本物体识别、长尾物体识别和增量物体识别等多个单一开放维度的子任务. 开集识别可分为 3 类：决策改进类[12-15]，通过更新神经网络输出层，优化决策过程，区分已知与未知类别，减少误分类；表示优化类[16-20]，通过改进网络特征表示，结合监督与无监督技术提取判别性特征，提升对未知类别的识别；数据生成类[21-24]，使用 GANs 生成或增广数据，模拟未知类，训练模型以识别新类别. 零样本识别方法分为 2 类：非生成式方法和生成式方法. 非生成式方法[25-27] 将视觉特征投影至语义空间或者公共空间做判别，生成式方法[28-29] 利用语义信息生成未知类视觉特征来训练. 小样本识别可分为 3 类：度量学习式法[30-31]，通过大样本类别的数据学习样本间相似性，再应用于小样本类别的分类；样本生成法[32-33]，在大样本类别上学习增广技术后用于小样本类别的数据增广，解决样本稀缺；元学习法[34-35]，通过大样本类别的训练数据学习优化策略和初始化，以快速适应小样本任务. 在长尾识别任务中，常通过重采样技术[36] 或修改损失函数[37] 来增加模型对少数类别的关注. 增量识别方法主要分为下面 3 个类别，以应对在学习新任务时发生的“灾难性遗忘”问题：结构型策略[38-39] 通过设计新型的网络架构或集成附加网络模块，旨在有效减缓遗忘现象；正则化策略[40-41] 在训练过程中引入特定的“防遗忘”约束条件，以降低遗忘的可能性，这些约束根据施加的位置不同，可进一步细分为权重正则化和激活正则化；回顾型策略[42-43] 通过保存并定期复习旧任务的关键信息来抑制遗忘，这一信息可通过保留代表性旧类样本或利用能够表征旧数据分布的生成模型来实现. 开放世界识别方法分类总览如图 2 所示.

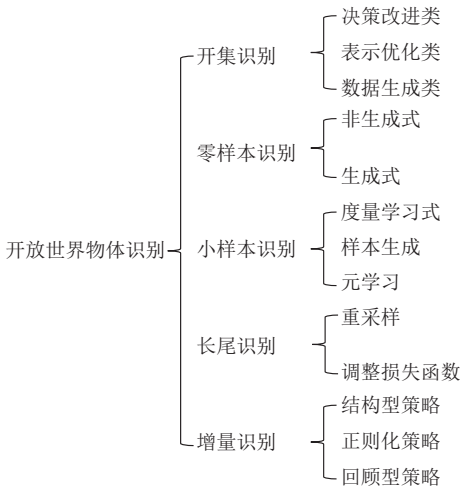


Fig. 2 Classification of open world recognition methods

图 2 开放世界识别方法分类

开放世界物体检测同样是一个综合任务，包含开集物体检测、零样本物体检测、小样本物体检测、长尾物体检测和增量物体检测等多个单一开放维度的子任务. 开集物体检测方法[44] 可以分为非生成式和生成式 2 种类型. 早期研究采用的非生成式方法[45] 通过把已知类别暂时视为未知类别，来训练未知类别的分类器，或者通过比较已知类别的预测概率值与一个预定的阈值来判断一个实例是否属于已知类别. 而最新的研究使用的生成式方法[46] 则通过创造未知类别的样本来进行不确定性的评估. 零样本物体检测方法同样也分为非生成式和生成式 2 种类型，非生成式[47] 主要通过将视觉特征投影至语义空间做判别得到，但是在这种范式下未知类不参与训练，会导致最终模型的预测偏向已知类，最新的研究通常采用生成式的方法[48]，通过未知类的类别语义合成对应的视觉特征参与训练，取得了良好的效果. 小样本物体检测方法主要分为元学习[49] 和微调[50] 两种类

型，元学习的方式训练成本小，微调的方式实现简单. 长尾物体检测中同样采用重采样[51]或者调整损失函数[52]的方式，使模型更关注尾部类别. 增量物体检测主要解决的是灾难性遗忘问题，希望模型学习新类的同时防止旧类遗忘，目前的工作一般采用特征蒸馏[53]或者样例回放[11]的方式防止遗忘. 开放世界检测方法分类总览如图 3 所示.

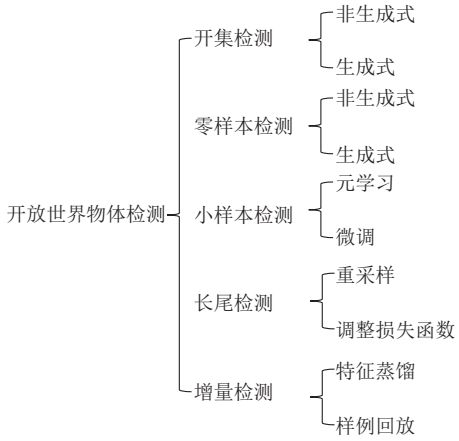


Fig. 3 Classification of open world detection methods

图 3 开放世界检测方法分类

总的来说，所谓的开放，超越传统封闭集的概念，是对真实世界动态变化性的一种模拟. 根据模拟的角度不同，目前主要有 5 种常见的任务设定：开集学习、零样本学习、小样本学习、长尾学习、增量学习，本文称之为开放性维度. 开集学习关注识别训练中未见过的类别，反映了现实世界中不断出现新物体的情况. 零样本学习强调在没有直接样本支持的情况下识别新类别，适应现实世界中未知物体的出现. 小样本学习涉及从极少量样本中快速学习新类别的能力，对于常见的现实情景中仅有少量数据的新物体类别至关重要. 长尾学习应对现实世界中常见类别和罕见类别的不平衡分布，能够处理稀有物体的识别. 增量学习强调模型在学习新知识时保持对旧知识的记忆，适应环境的持续变化. 目前各个维度的开放性任务大多都是孤立研究的，近年来有一种趋势，方法研究从单一维度转向复合维度，但是目前复合维度最多只考虑了 2 个，对于更为复杂的综合了更多开放性维度的任务设定则缺乏考虑. 在本文中，提出囊括上述 5 个主要开放性维度的广义开放世界物体识别与检测任务，超越了以往工作中的开放性任务设定，缩小了与真实世界的差距.

现有的开放世界学习定义，虽然着重强调了模型具备开集学习和增量学习的能力，但这种定义实际上仍然较为狭窄，没有完全捕捉到开放世界环境的复杂性和多样性. 在更为全面和深入的考虑中，开放世界物体识别与检测的概念应当包含更广泛的学习场景和挑战. 例如，零样本学习和小样本学习在开放世界任务中扮演着关键角色. 在这些场景中，模型需要能够识别在训练过程中未见过或仅见过少量样本的物体类别. 此外，长尾学习也是开放世界任务的一个重要方面，因为现实世界中的物体类别分布通常是长尾形态，意味着大量的稀有物体类别和少数的常见类别共存. 上述各个任务是孤立研究的，但它们都强调了真实世界开放性的一部分，因此本文定义一个更加全面的任务广义开放世界学习来囊括以上所有任务.

此外，从是否使用预训练模型这个角度可以将开放世界的相关研究工作分成 2 类：一类是传统的开放世界的设定[10-11]，从头开始训练，不使用预训练模型；另一类是使用了预训练模型如 CLIP[54] 和 SAM[55] 的工作，它们在训练过程中使用了未知类别的数据，不符合严格的开放世界的设定，但是取得了很好的效果. 这 2 种设定都有各自的研究价值，第一种符合传统的开放世界的设定，是一种更加纯粹的研究范式，第二种实际应用效果出色，本文主要集中于第一种.

数据集方面，对于开放世界物体识别领域，不同开放性维度任务所采用的数据集不同. 开集识别通常采用 MNIST, CIFAR10 等数据集，MNIST 包含 10 个手写数字类别，每类约 7 000 张图片，CIFAR10 包含 10 类，每类 6 000 张图片. 零样本识别通常采用 AWA[56], CUB[57] 等数据集，AWA 包含了 50 个动物类别共 30 000 张图片，每个类别都附带了描述其属性的信息，CUB 数据集专注于鸟类的细粒度识别，包含 200 个鸟类类别共 11 000 张图片. 小样本识别通常采用 miniImageNet[58] 和 TieredImageNet[59] 等数据集，miniImageNet 是从大规模图像分类数据集 ImageNet 中抽取的一个子集，包含 100 个类别，每个类别有 600 张图片，与 miniImageNet 相比，TieredImageNet 在类别上的分布更为广泛和平衡. 长尾识别通常采用 iNaturalist[60] 等数据集，iNaturalist 数据集是一个真实世界的生物物种识别数据集，具有明显的长尾分布特征. 增量识别通常采用 CIFAR100 和 ImageNet 等数据集，两者分别包含 100 类和 1 000 类的自然图片. 常用的物体检测数据集包含 PASCAL VOC[61], MSCOCO, Objects365 等. PASCAL VOC 最早由牛津大学于 2005 年发布，至 2012 年每年都会发布一个新的版本. 目前

常用的有2个版本：VOC07和VOC12，前者包含约5 000张的训练图片和约5 000张的测试图片，后者包含约11 000张的训练图片和约11 000张的测试图片，总共包含20类物体. MSCOCO数据集是目前检测任务最常用的数据集，总共包含80类物体，超过120 000张图片. 开放世界物体检测及其各单一开放性维度子任务主要是对这些常用检测数据集进行类别划分和数量限制使之符合不同任务的要求，从而得到各自的数据集. 比如零样本物体检测中对MSCOCO进行类别划分，选择其中65类作为已知类，剩下15类作为未知类；小样本物体检测同样也基于MSCOCO进行类别划分，选择和MSCOCO中与PASCAL VOC重合的20类作为小样本类，剩下的作为基类. AP(average precision)是物体检测领域常用的评价指标之一，用于衡量模型在检测任务上的性能. AP是在不同置信度阈值下计算得到的检测精度(precision)和召回率(recall)曲线下面积的平均值，它旨在评估模型在检测精度和召回率之间的平衡能力. 目前，一般采用COCO AP，即计算交并比(IoU)从0.5到0.95(以0.05为步长)所有阈值下AP的平均值，能够提供关于模型性能的全面评估. 另外，AP50指标也较为常用，AP50指的是在IoU阈值为0.5时的平均精度. 也就是说，当预测的边界框与真实边界框的IoU大于等于0.5时，这个预测被认为是正确的. 对于开放世界相关的任务，一般会评测已知类和未知类的AP. 开放世界相关方法在数据集上(如MSCOCO)的表现不如全监督检测方法的表现，且由于采用不同数据和划分，所以无法统一比较不同单一开放性维度方法的性能，一般来说长尾和增量方法的性能会更高一些，而零样本和开集方法的性能更低.

现有的广义开放世界子任务数据集对常用的物体识别与检测数据集(如COCO)进行简单的类别划分，这种方法无法全面评估模型在开放世界条件下的性能. 这种单一的数据集构造方式不仅限制了对方法缺陷的诊断，也可能导致模型对特定训练场景产生偏见，使其在遇到新场景时性能下降. 由于数据集的收集、处理和标注成本较高(尤其是检测数据集)，通过网络爬取和人工标注的方式构建具有不同分布和划分的数据集变得不现实. 不同于之前的做法，本文提出一种新方法充分利用仿真平台(如AI2-THOR[62])低成本获取大量标注好的数据，此外本文解耦了广义开放世界的5个核心难度指标，并通过调整指标数值采样由AI2-THOR产生的元数据生成任务特定的数据.

3 开放世界物体识别：挑战、实践与展望

在过去的十几年中，传统物体识别技术取得了显著的进展. 然而，这些技术通常局限于封闭的数据集，即所有类别在训练阶段均被视为已知且充足. 这种假设与现实世界的情况相去甚远. 现实世界是一个开放的环境，其中类别数量不断变化，新的未知类别不断出现. 开放世界的挑战包含类别变化、噪声学习、领域差异、算法效率等. 在这4个挑战中：1)类别变化是指在开放世界环境中系统可能遇到新的或未知的类别，这些类别在训练过程中未被考虑. 对于这种情况，模型需要有能力识别并适当处理这些新类别. 这个挑战涉及到如何让系统能够有效地适应或扩展到新类别，而不需要从头开始训练. 2)噪声学习[63-64]是指在实际应用中，训练数据可能包含噪声，例如错误的标签、低质量的输入数据或者不相关的信息. 噪声学习需要设计算法去识别、处理或抵抗这些噪声，以便不会对模型的学习过程和最终性能产生负面影响. 3)领域差异[65-66]是指训练环境(源域)与实际部署环境(目标域)之间的差异. 这些差异可能是由于数据分布的变化、不同的数据采集过程或环境条件造成的. 处理领域差异在于如何使模型能够适应新领域，或者如何将源域学到的知识迁移到与之不同的目标域. 4)算法效率[67-68]涉及到如何设计能够快速、准确、资源高效地处理任务的算法. 在开放世界环境中，算法可能需要在有限的计算资源下处理大量的数据并做出及时的决策. 其中类别变化的挑战最为关键，因为开放世界的核心特征之一就是环境的动态性，尤其是类别的不断变化和扩展. 这直接触及到开放世界物体识别与检测系统设计的根本目的——在不断变化的环境中保持有效性和适应性. 相比其他挑战，如算法效率、计算资源限制等，类别变化更深刻地体现了开放世界环境的本质特征. 因此，如图4所示，为了更贴近真实世界的复杂性和开放性，开集学习[69]、零样本学习[70-71]、小样本学习[72-73]、长尾学习[74]以及增量学习[75]多个开放性维度的研究应运而生. 这些维度分别关注不同的挑战：开集学习针对未知类别的识别；零样本学习和小样本学习聚焦于在极少量或无样本的情况下学习新类别；长尾学习应对类别分布的不平衡；增量学习旨在模型学习新知识的同时保持对旧知识的记忆以适应环境的持续变化. 虽然这些维度各自取得了一定的进展，但它们

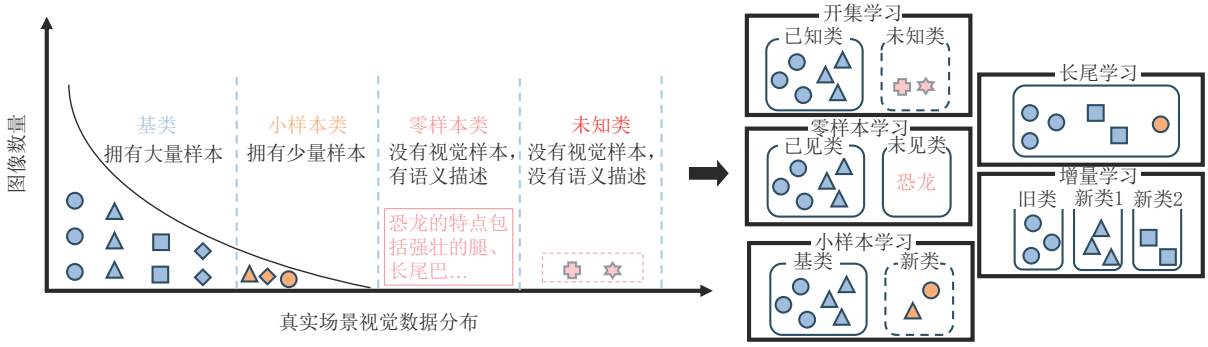


Fig. 4 Simulating the openness of the real scenes from different perspectives

图 4 从不同角度模拟真实世界的开放性

多数局限于单一维度的探索(比如只考虑开集学习设定)或者只考虑 2 种复合维度(比如开放世界，包含增量学习和开集学习 2 个开放性维度)，缺少一种综合考虑更多开放性维度的全面视角. 针对该问题，本文从更宏观的角度提出包含 5 个开放性维度的更加综合的设定.

首先思考为什么这些开放性维度相互割裂缺乏统一视角？究其原因，本质上是对于开放世界的评测基准存在的不足导致的研究方法相对独立，即现有评测基准大多只关注单一开放性维度的评测，大多数方法追求在各自开放性维度对应评测基准上的极值性能，尚未广泛考虑从一个更统一的视角综合评估多个开放性维度的问题. 此外，相关的数据集通常是基于对常用物体识别数据集的单一划分，只提供一种难度级别的测试，这可能无法有效区分不同方法的性能. 这就像设计试卷来考察考生一样，如果试卷难度单一，无论是太难还是太简单，都无法有效区分考生的实际水平，也无法指导考生分析自身的不足.

随着信息技术的快速发展和人工智能、大数据等新兴技术的不断进步，本文面临着将这些技术应用于开放世界物体识别与检测领域的挑战与机遇. 如图 5 所示，为解决当前开放世界物体识别任务评测基准中存在的问题，本文引入了一种新的广义开

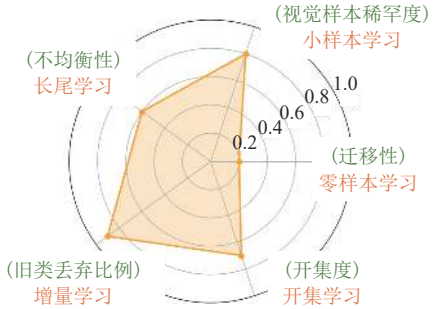


Fig. 5 Openness dimensions and their core difficulty metrics

图 5 开放性维度及其核心难度指标

放世界评估范式. 该范式首先将挑战细分为 5 个关键的开放性维度，每个维度都代表了开放世界场景中的一个核心要素. 在此基础上，本文提出了一个综合框架，旨在全面涵盖所有开放性维度.

通过精心设计的核心难度指标，本文可以为每个任务生成具有不同难度的数据. 这些指标不仅可以用于准确评价各个任务数据集的难度，而且还可以用于根据需求生成自定义难度的数据集. 本文通过解耦开放世界设定下各子任务的相关维度指标，使得数据集的生成既可控制又灵活，满足不同研究和实践的需求. 这种方法的引入，不仅是对现有评测方法的一大改进，而且为开放世界物体识别与检测领域的研究带来了新的视角和可能性. 这一全新的评估范式，对于深入理解和有效应对开放世界物体识别与检测中的复杂性和多样性具有重要的意义.

具体每个指标的定义如下：

1) 开集度. 是开集学习任务的核心难度指标，表示模型在测试环境中潜在的未知程度. 本文将其定义为未知类别数量占所有类别(包含已知类别和未知类别)数量的比例：

$$M_{\text{osd}} = \frac{|C_{\text{uk}}|}{|C_{\text{uk}}|+|C_{\text{k}}|}, \tag{1}$$

其中，$|C_{\text{k}}|$是已知类别数量，$|C_{\text{uk}}|$是未知类别的数量.

2) 迁移性. 表示知识(如属性、模式、特征等)在零样本学习中可以从已见类迁移到未见类的程度，是核心难度指标. 本文定义迁移性为

$$M_{\text{tran}} = \frac{1}{|C_{\text{u}}|}\sum_{a\in C_{\text{u}}} Max_{b\in C_{\text{s}}} Sim(sem_a, sem_b), \tag{2}$$

其中，C_{s}和C_{u}分别表示已见类和未见类类别，sem表示一个类别的语义向量(例如词向量或者属性向量)，$Sim(\cdot,\cdot)$表示相似度计算的方式，这里本文具体使用余弦相似度来计算.

3)视觉样本稀罕度. 是小样本学习任务的核心

难度指标. 本文将训练数据中类别实例数量N_{f}低于N_{fmax}的类别作为小样本类(现有工作中一般认为实例数量超过 30 的不是小样本类，即$N_{\mathrm{fmax}}=30$). 视觉样本稀罕度定义为

$$M_{\mathrm{svs}}=\frac{N_{\mathrm{fmax}}-N_{\mathrm{f}}}{N_{\mathrm{fmax}}}, \tag{3}$$

该指标数值越大，意味着小样本的平均样本数量越低，相应地，任务难度也越大. 这里，一般情况下可以直接使用现有工作的设定[72-73]，即设定N_{fmax}=30. 如果实际应用场景数据比较独特，可以参考 *K*-means 聚类中 *K* 值选择的手肘法确定N_{fmax}的值. 以新类的性能作为N_{fmax}选择的一个度量，当新类平均样本数量低于N_{fmax}，增加新类样本数量，新类性能将大幅度提升，当新类平均样本数量高于N_{fmax}，增加新类样本数量，新类性能提升幅度变缓. 新类性能和新类平均样本数量的关系呈现出一个倒 L 形的曲线形状，而 L 形曲线的拐点所对应的新类平均样本数量即为N_{fmax}.

4) 不均衡性. 是长尾学习任务的核心难度指标. 具体而言，评价一个数据集类别实例数量的分布长尾效应是否足够明显，本文可以用熵来衡量. 通过统计每个类别实例数量的占比，得到每个类别实例的频率f_i，由此可以计算出熵H_f

$$H_f=-\sum_{i=1}^{n}f_i\log(f_i). \tag{4}$$

为了统一表示，本文通过一个映射函数将H_f映射为值域为 [0, 1] 的值以表示不均衡性的大小，计算方式为

$$M_{\mathrm{imb}}=\frac{2\mathrm{e}^{-H_f}}{1+\mathrm{e}^{-H_f}}, \tag{5}$$

该数值越大，表示长尾效应越明显，即大部分的数据类别都集中在头部类别上.

5) 旧类丢弃比例. 增量学习有 2 种简单且极端的解决方式：如果把旧数据全部丢掉，每次只用新数据学习，那么虽然学习十分高效，但容易出现旧知识的灾难性遗忘问题；如果把旧数据全部保留，每次用新旧数据联合学习，那么虽然遗忘问题得到了解决，但是这种做法的时空代价高. 因此，本文将增量过程中旧数据的丢弃比例作为衡量增量任务难度的核心指标，指标数值越大意味着任务难度也就越大，其定义为

$$M_{\mathrm{ocdr}}=\frac{1-\mathrm{e}^{-n_{\mathrm{task}}}}{2+2\mathrm{e}^{-n_{\mathrm{task}}}}\frac{1}{n_{\mathrm{old}}}\sum_{i=1}^{n_{\mathrm{old}}}\frac{N^i_{\mathrm{discard}}}{N^i_{\mathrm{total}}}, \tag{6}$$

其中，n_{old}表示旧类类别数量，N^i_{discard}表示增量学习阶段旧类丢弃的样本数量，N^i_{total}表示旧类的样本总数，n_{task}表示任务数量.

广义开放世界评估方法既可用于评估现有数据集的难度，又可以用于指导生成自定义难度的数据. 当广义开放世界评估方法用于评估一个给定数据集的难度时，我们获得数据集的类别划分以及实例数量等信息后，可以直接代入式(1)~(6)计算得到 5 个维度的难度指标. 比如某个给定的开集识别数据集的类别划分为已知类 60 个、未知类 20 个，我们可以直接计算得到其开集度为 0.25. 当广义开放世界评估方法用于生成给定难度数据集时，以生成开集识别数据集为例，我们需要先确定好类别总数(包含已知类和未知类)，然后根据对应维度的式(1)计算即可得到如已知类和未知类的具体划分比例等信息.

为了改善当前开放性维度数据单一的问题，本文通过组合并改变 5 个开放维度指标的数值，实现任务特定难度可控的数据生成. 既可只控制单一维度生成某个单一任务的数据，又可以同时控制多个维度生成复合任务的数据. 本文可以通过从元数据集中采样的方式实现，这里的元数据集可以是任意现有的物体识别数据集(如 ImageNet 数据集)或者它们的组合.

采样阶段中本文使用 5 个开放维度指标约束采样算法，使其最终采样的数据符合指标定义的数值. 采样阶段中，不同任务生成的过程实际上略有差异. 开集学习、小样本学习和增量学习生成的数据只涉及类别划分，如生成开集学习数据只要根据开集程度数值挑选未知类别的数量达到要求即可. 对于零样本学习数据集的生成，实际上无法准确得到任意迁移数值的类别划分，因为给定一个类别集合和零样本类别的数量后，能够得到的迁移性取值是离散的. 因此只能挑选一个近似给定迁移性数值的类别划分，需要先确定未见类别的数量，然后枚举所有已见-未见类别的划分，从而得到不同的迁移难度，最后挑选与给定数值最接近的划分即可. 对于长尾数据的生成，采用指数分布近似模拟长尾分布.

在探讨开放世界物体识别与检测系统时，评价指标的定义至关重要. 目前本文采用的 5 个指标：开集度、迁移性、视觉样本稀罕度、不均衡性和旧类丢弃比例.这些指标的公式化定义虽然为量化任务难度提供了有效途径，但它们仅仅代表了诸多可能性中的一种实现方式. 这些指标以一种量化的方式捕捉了每个维度的核心难点，但它们的设计和实现还有很大的探索空间.

未来的研究可以在这些初步框架的基础上继续

发展和完善，设计出更为全面和深入的评价指标. 未来的评价指标需要能够更精准地反映开放世界物体识别与检测系统面临的多维度挑战. 例如，它们可能需要综合考虑各种因素，如数据的多样性、类别的动态性、环境的复杂性和模型的泛化能力. 此外，这些评价指标应当能够适应不同的应用场景和需求，从而推动开放世界物体识别与检测系统的部署和应用.

综上所述，尽管当前的评价指标已经提供了对开放世界物体识别与检测系统评估的初步方法，但随着技术的进步和应用需求的发展，本文预见到在评价指标设计方面将会有更广阔的探索空间和发展潜力. 未来的研究不仅将继续完善现有的指标，还将探索新的指标，以更全面地评价和指导开放世界物体识别与检测系统的研究和实践. 在探索开放世界的物体识别的挑战时，我们了解到类别变化是核心难点之一. 为了适应这种不断变化的分类环境，我们必须开发出能够持续进化和适应新类别的系统. 然而，类别的变化仅仅是开放世界挑战的冰山一角. 第4节将继续本节的讨论，并将关注点扩展到如何在实际应用中精确地定位和识别这些不断变化的类别. 我们将详细探讨开放世界环境中物体检测所面临的独特问题，尤其是在复杂场景中对物体进行精确定位的挑战.

4 开放世界物体检测：挑战、实践与展望

继第3节对开放世界物体识别的挑战进行了初步分析之后，本节将进一步探讨开放世界物体检测所面临的挑战. 物体检测不仅要识别物体的类别，而且要在视觉场景中准确地定位物体的位置. 这项任务在开放世界的背景下变得更加复杂.

在开放世界物体检测的研究中，面对的不仅是物体的识别，还包括在复杂的、不断变化的环境中对物体进行精确定位. 与物体识别任务相比，开放世界物体检测要处理的挑战更为多样和复杂. 在开放环境中，物体可能会因为遮挡而部分或完全不可见，尺度变化使得相同物体在不同距离下的表现差异巨大，姿态的多样性要求检测系统能识别同一物体的各种展现形式，而共生关系和背景干扰则进一步增加了识别的难度. 例如，一个物体可能与环境中的其他物体紧密相连，或者在复杂的背景前几乎隐匿，这些因素都极大地提高了检测的难度.

现有的开放性物体检测方法通常依赖于对已有数据集的划分来实现，比如零样本物体检测中通常将ImageNet或COCO数据集中的类别划分为已见类和未见类. 然而，这些数据集通常只能提供有限的场景变化，且其数据分布往往固定并偏向于特定的几个类别，这使得模型难以适应现实世界中类别数量不一、不断变化和新类别不断出现的开放性环境. 从现有数据源（例如COCO）生成结合多维度和不同难度级别的高质量数据集. 这种方法可能受限于现有的类别分布和实例数量，通常缺乏稀有对象和场景，且可能存在未标注对象的问题，这将阻碍物体检测器发现新类别，如图6所示. 此外，搜集大量新数据并对其进行标注的成本极高.

Fig. 6 There are many unannotated objects in COCO dataset

图6 在COCO数据中存在许多未标注物体

为了克服这些限制，开放世界物体检测的研究需要采用更为灵活和综合的方法. 这可能包括开发能够适应新类别出现的增量学习算法，设计能够处理大量类别和实例不均匀分布的长尾学习方法，以及探索能够从少量或零样本中学习新类别的模型. 此外，还需要创建新的数据集，这些数据集能够更真实地反映开放世界中的多样性和复杂性，包括各种遮挡情况、尺度变化、不同姿态、复杂的共生关系以及丰富多变的背景.

开放世界物体检测领域的发展，不仅要求技术上的创新，而且需要对现有研究方法和数据资源重新思考. 本文跳出传统数据集的框架，采用更加综合的方法来面对开放世界环境下的挑战，从而推动这一领域向前发展.

本文探索利用仿真环境来构建开放世界物体检测数据集的独特优势. 仿真环境提供了一个可控且灵活的平台，使研究人员能够创造出接近现实世界的多变场景，这在传统数据集中是难以实现的，如图7所示，通过操控仿真环境可以控制光照、纹理、位姿等变化. 以下是使用仿真环境构建数据集的6个关键优势：

1）多样性. 仿真环境可以渲染出多种背景、光照条件和天气状况下的场景，增强数据集的多样性，从

Fig. 7 Controlling lighting, texture, and pose variations in simulation environments

图 7 在仿真环境中控制光照、纹理和位姿变化

而提高模型的泛化能力.

2)可定制性. 研究人员可以根据需要定制场景的具体参数, 如物体的大小、颜色、纹理等, 以适应特定的测试或训练需求.

3)复杂场景的生成. 仿真技术能够生成包含复杂交互和物体关系的场景.

4)标注成本的减少. 在仿真环境中生成的数据通常可以自动获取精确的标注信息, 如边界框、分割掩码和物体类别, 从而减少了人工标注的成本和时间, 如图 8 所示.

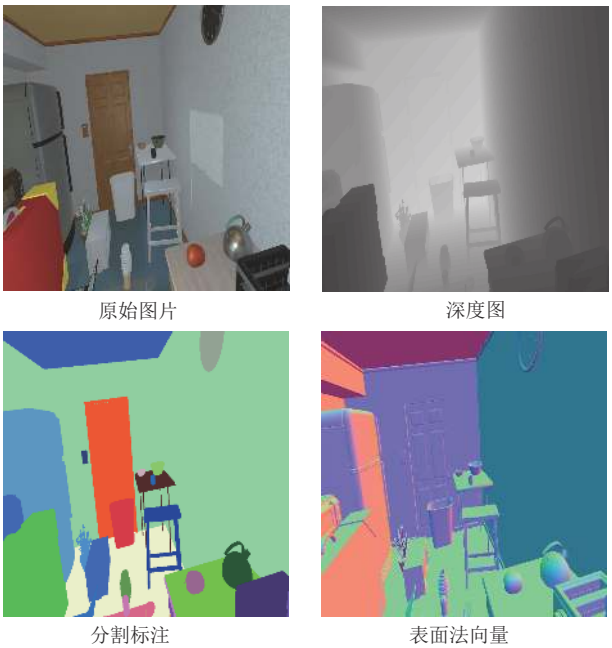


Fig. 8 Diverse annotations provided by the simulation environment

图 8 仿真环境自带的多样化标注

5)未知类别的引入. 仿真环境允许研究人员设计并引入未在现实世界中出现过的虚构类别, 为开放集和零样本学习提供了理想的测试平台.

6)遮挡和姿态变化的模拟. 仿真环境可以特意设计场景以模拟不同程度的遮挡和各种姿态的变化, 为物体检测算法提供更为严峻的测试条件.

利用这些优势, 仿真环境可以成为开放世界物体检测研究的强大工具. 它不仅能够支持传统物体检测任务, 而且能够帮助研究人员探索新的方法和定义新的开放性任务, 收集和构建自定义的开放世界物体检测数据集用以促进方法研究, 从而解决开放世界环境中未知的和不断演变的挑战. 通过这种方式, 仿真环境极大地拓展了数据集的边界, 推动了开放世界物体识别与检测系统的进步. 未来的研究可以在仿真环境中不断探索新的算法和模型, 从而不断推进开放世界物体检测技术的发展.

本文基于 AI2-THOR 环境实现了一个自动化的数据集收集平台, AI2-THOR 是一个基于 Unity 实现的开源仿真环境, 提供由专业 3D 艺术家手工设计的场景, 并且能够渲染得到图像的深度、语义分割、实例分割等标注信息, 可以用于具身智能、机器人以及计算机视觉领域的研究. 但是 AI2-THOR 的物体实例有限, 只有 3 578 个独立物体模型, 而且场景中 Agent 的视角和高度固定, 想要收集大规模的数据存在一定困难, 无法满足本文的需求.

因此, 本文根据需求补充了一些新的功能: 将现有的三维物体模型库(如 ShapeNet[76])的模型导入 AI2-THOR 中, 来扩充 AI2-THOR 的物体模型资源(共扩充了 73 类物体, 包含 77 689 个实例), 使其能够正确识别与初始化新加入的模型; 为了缩小合成数据与真实数据之间的领域差异, 通过随机化场景中物体的位置、姿势、方向、材料以及相机的位置与角度来实现域随机化, 使收集的合成数据更加多样化; 根据从通用检测数据集(如 COCO)中统计的物体共生频率, 设计将物体模型加入到场景的数量与类别, 进一步缩小合成数据与真实数据分布差异. 在完善了上述功能后, 本文实现了基于 AI2-THOR 的自动化的数据集收集平台, 在进行数据收集时, 将会遍历 AI2-THOR 的所有场景, 并对场景进行物体扩充与随机化, 然后通过随机位置与角度的相机进行渲染, 得到数据及其实例分割标注, 之后根据需求对数据集进行处理, 最终得到所需要的数据集.

为了方便实验, 减少不必要的渲染时间, 本文首先获取了一个元数据集, 通过调整不同的开放世界指标对该元数据集进行采样控制, 从而得到后续实验中不同的任务所需的特定数据集. 本文构建了一

个包含 58 470 张图像的元数据集，其中每张图像都有对应的像素级实例分割标注和目标检测标注，涵盖了 96 个类别，并通过对物体位姿、材质、相机位置的随机化以及光照条件的改变实现域随机化、缩小合成数据与真实数据的差异. 数据集一共有 504 574 个目标检测实例标注，平均每张图像拥有 8.63 个目标检测标注.

接下来将介绍利用仿真环境生成的数据去评测开放性检测相关的方法. 本文选择了 2 种当前工作涉及较多的典型单一开放维度任务：零样本检测和长尾物体检测. 训练集和测试集的比例为 4∶1.

在表 1 中，比较了 3 种零样本物体检测方法 DPIF[77]、RRFS[78] 和 ZSDSCR[79] 在不同程度的迁移性 M_{tran}(0.30, 0.46, 0.57)下的表现. 这些方法的主要性能指标包括已见类的平均精度($AP50_s$)、未见类的平均精度($AP50_u$). 对于现有方法，较小的迁移性提供了较大的区分能力，未见类的 AP 最多相差 2 个百分点. 在研究中，零样本检测的主要目标是在未见类别上获得较好的性能，因为这直接反映了模型对于新颖类别的识别能力. 尽管对已见类别的表现在某种程度上也是重要的，但它并不是我们评估模型性能的主要标准. 在表 1 中，虽然 DPIF 模型在中等程度迁移性时的整体表现最优，但我们发现在未见类别上的表现并不总是与此相符. 这表明即使在已见类别上取得了相对较好的结果，模型在未见类别上的表现仍然是不确定的，这与我们的研究重点相契合. 我们的结论侧重于分析模型在未见类别上的表现，而不是仅仅基于总体性能.

Table 1 Experimental Results of Zero-shot Object Detection with Varying Transferability Metrics

表 1 变化迁移性指标的零样本物体检测实验结果

方法	M_{tran}	$AP50_s$	$AP50_u$
DPIF	0.30	51.5	2.7
	0.46	57.2	3.4
	0.57	52.6	4.5
RRFS	0.30	53.5	1.6
	0.46	44.8	1.9
	0.57	49.9	3.9
ZSDSCR	0.30	53.6	0.7
	0.46	44.8	1.2
	0.57	50.0	3.5

在表 2 中，比较了 2 种长尾物体检测方法 EQLV2[80] 和 Seesaw[81] 在不同程度的不均衡性 M_{imb}(0.1, 0.5, 0.9)下的表现. 对于长尾检测，极端平衡(M_{imb}=0.1)的分布或极端不平衡(M_{imb}=0.9)的分布均不利于有效区分现有方法，意味着过高或过低的难度级别缺乏明显的区分性. M_{imb}=0.5 的设置更适合区分当前的方法. 当然，更精确的数值需要进一步的实验来验证.

Table 2 Experimental Results of Long-tailed Object Detection with Varying Imbalance Metrics

表 2 变化不均衡性指标的长尾物体检测实验结果

方法	M_{imb}	AP	AP_{50}
EQLV2	0.1	50.3	64.0
	0.5	30.8	39.5
	0.9	17.2	21.7
Seesaw	0.1	51.2	64.9
	0.5	32.9	42.0
	0.9	17.6	22.1

表 1 和表 2 的实验结果表明，迁移性、不均衡性等开放性指标的变化可能在某些情况下对模型的性能造成一定影响. 根据所分析的特定数据集和任务(仿真数据上的长尾、零样本物体检测任务)，我们观察到当数据分布呈现更为明显的长尾特性时，或者已见类和未见类的迁移性更小时，模型的性能往往有所下降. 然而，我们也认识到不均衡性、迁移性等开放性指标对性能的具体影响可能因任务、数据集的不同而存在变化. 因此，上述结论需要在更广泛的实验中进一步验证，并考虑到在不同任务和数据条件下可能出现的多变性.

此外，本文进一步在真实的物体检测数据集上做了零样本物体检测实验，原始数据集为 PASCAL VOC 数据，实验中将 20 个类别划分成 16 个已见类和 4 个未见类，训练集只包含已见类数据，剔除了包含未见类的图片，实验结果如表 3 所示，实验结论与在仿真数据上的一致，即：迁移性可能在某些情况下对未见类的性能产生有利的影响. 根据所分析的特定数据集和真实数据上的零样本物体检测任务，观察到当数据分布呈现更为明显的迁移性时，未见类的性能会有所提升.

本文对该项研究的未来发展和改进进行了深入的思考. 仿真环境作为一种强大的工具，在开放世界物体检测的研究和应用中发挥着重要作用. 然而，当前的仿真环境仍存在一些局限性，需要进一步的改进和发展. 目前大多数仿真环境主要集中在室内场景的模拟. 未来，考虑到开放世界环境的多样性，本文需要扩展仿真环境，包括更加丰富和复杂的室外场景. 比如自然环境(如森林、沙漠)、城市景观(如

Table 3 Experimental Results of Zero-shot Object Detection with Varying Transferability Metrics

表 3 变化迁移性指标的零样本物体检测实验结果

方法	M_{tran}	$AP50_s$	$AP50_u$
DPIF	0.25	44.4	0.8
	0.45	37.6	21.8
	0.67	32.3	38.3
RRFS	0.25	70.3	0.3
	0.45	62.6	7.7
	0.67	59.4	19.4
ZSDSCR	0.25	70.3	0.3
	0.45	62.6	6.6
	0.67	59.4	16.9

街道、广场)和特殊环境(如工业区、灾难现场).这样的扩展将为物体检测算法提供更加全面和现实的测试环境.当前的仿真环境主要依赖于传统的3D渲染技术,这在一定程度上限制了环境和物体的多样性和真实感.未来,可以考虑将仿真方法与最新的生成技术相结合,例如扩散模型.这种方法可以利用深度学习模型生成更加逼真、多样化的图像和场景,提高仿真环境的质量和效果.综上所述,虽然当前的仿真环境已经为开放世界物体检测提供了宝贵的支持,但未来的发展方向将是更加广阔和多元化.通过不断地技术创新和改进,仿真环境将成为开放世界物体检测研究的一个更加强大和有效的工具.

5 总结和展望

本文深入探讨了开放世界物体识别与检测问题,并指出了对于一个综合性评测框架的迫切需求.这样的框架能够有效地应对这一动态变化领域所提出的多样化挑战.目前,尽管开集学习、零样本学习、小样本学习、长尾学习和增量学习各自在其领域取得了良好进展,但其各自的评估环境仍相对割裂且不够全面.这种局限性,很大程度上是由于常规数据集的划分所限.本文所提出的方法通过解耦并构建跨5个开放维度的核心难度指标,创新性地定义了广义开放世界的多样化任务,为深入思考这一错综复杂和广阔的研究领域提供了新的视角.

利用AI2-THOR仿真平台,本文成功生成了一个多样化的开放世界物体检测数据集,降低了与传统数据收集方法所需的高昂成本和资源需求.本文通过模拟各种环境条件和物体变化,丰富了数据集,使其更加贴近真实世界的不可预测性和多样性.这为开放世界物体识别与检测模型的评估提供了一个更准确、更为全面的基准数据集.

本文通过实验验证了所构建的仿真数据集和所提出的度量标准的有效性和实用性.后续工作将集中于提升仿真环境的真实感和多样性,以及开发更精细和更全面的评价指标.

开放世界物体识别与检测系统目前在类别的开放性(开集、零样本、小样本、长尾、增量)方面已经逐渐走向统一,由只关注单一维度逐步走向关注复合维度,本文进一步考虑涵盖5个开放性维度的设定.开放世界系统的未来研究将围绕提升系统的解释性、多模态学习能力、安全性和效率等方面展开:1)随着开放世界系统在关键领域的应用增多,如自动驾驶、医疗诊断等,其决策过程的透明度和可解释性变得越来越重要.未来的研究需要着力于提升模型的解释能力,使非专业用户也能理解模型的决策逻辑,从而提高人们对这些系统的信任度.2)开放世界系统将面临来自不同源的、形式多样的数据.因此,未来的研究方向之一是如何有效地整合视觉、语音、文本等多种类型的数据,实现跨领域的学习和知识迁移.这不仅可以提高模型的泛化能力,还可以拓宽其应用范围.3)随着开放世界系统在社会生活中的应用日益广泛,如何保护用户数据的安全和隐私成为一个重要问题.未来的研究需要探索新的算法和技术,以确保在数据收集、处理和存储过程中用户的隐私得到有效保护,同时也要保证系统本身免受恶意攻击.4)对于在资源受限的设备上运行的开放世界系统,如智能手机和边缘计算设备,未来的研究将重点关注开发低能耗、高效能的算法.这不仅包括提升算法的计算效率,还包括优化模型的大小,使其能在不牺牲性能的前提下,在资源有限的设备上顺畅运行.通过解决这些关键问题,我们可以推动开放世界系统在更广泛的应用场景中发挥更大的作用.

作者贡献声明: 聂晖提出算法实现方案,开展实验并完成论文撰写;王瑞平提出论文整体研究思路,指导算法与实验方案设计,并修改论文;陈熙霖提出指导意见并修改论文.

参 考 文 献

[1] Hu Yihan, Yang Jiazhi, Chen Li, et al. Planning-oriented autonomous driving[C]//Proc of the IEEE/CVF Conf on Computer Vision and

Pattern Recognition. Piscataway, NJ: IEEE, 2023: 17853−17862

[2] Zhou Kaiyang, Yang Yongxin, Cavallaro A, et al. Omni-scale feature learning for person re-identification[C]//Pro of the IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 3702−3712

[3] Hu Qixin, Chen Yixiong, Xiao Junfei, et al. Label-free liver tumor segmentation[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 7422−7432

[4] Liu Haotian, Li Chunyuan, Li Yuheng, et al. Improved baselines with visual instruction tuning[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2024: 26296−26306

[5] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2017: 5998−6008

[6] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[C]//Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2020: 1877−1901

[7] Deng Jia, Dong Wei, Socher R, et al. ImageNet: A large-scale hierarchical image database[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2009: 248−255

[8] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common objects in context[C]//Proc of the European Conf on Computer Vision. Cham: Springer, 2014: 740−755

[9] Shao Shuai, Li Zeming, Zhang Tianyuan, et al. Objects365: A large-scale, high-quality dataset for object detection[C]//Proc of the IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 8430−8439

[10] Bendale A, Boult T. Towards open world recognition[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 1893−1902

[11] Joseph K J, Khan S, Khan F, et al. Towards open world object detection[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 5830−5840

[12] Bendale A, Boult T E. Towards open set deep networks[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 1563−1572

[13] Shu Lei, Xu Hu, Liu Bing. Doc: Deep open classification of text documents[J]. arXiv preprint, arXiv: 1709.08716, 2017

[14] Yang Hongming, Zhang Xuyao, Yin Fei, et al. Convolutional prototype network for open set recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(5): 2358−2370

[15] Zhang Hongjie, Li Ang, Guo Jie, et al. Hybrid models for open set recognition[C]//Proc of the European Conf on Computer Vision. Cham: Springer, 2020: 102−117

[16] Geng Chuanxing, Chen Songcan. Collective decision for open set recognition[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(1): 192−204

[17] Yoshihashi R, Shao Wen, Kawakami R, et al. Classification-reconstruction learning for open-set recognition[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 4016−4025

[18] Hassen M, Chan P K. Learning a neural-network-based representation for open set recognition[J]. arXiv preprint, arXiv: 1802.04365, 2018

[19] Sun Xin, Yang Zhenning, Zhang Chi, et al. Conditional Gaussian distribution learning for open set recognition[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 13480−13489

[20] Perera P, Morariu V I, Jain R, et al. Generative-discriminative feature representations for open-set recognition[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 11814−11823

[21] Ge Zongyuan, Demyanov S, Chen Zetao, et al. Generative openmax for multi-class open set classification[J]. arXiv preprint, arXiv: 1707.07418, 2017

[22] Neal L, Olson M, Fern X, et al. Open set learning with counterfactual images[C]//Proc of the European Conf on Computer Vision. Cham: Springer, 2018: 613−628

[23] Yu Yang, Qu Weiyang, Li Nan, et al.[J]. arXiv preprint, arXiv: 1705.08722, 2017

[24] Ditria L, Meyer B J, Drummond T. Opengan: Open set generative adversarial networks[C]//Proc of the Asian Conf on Computer Vision. Cham: Springer, 2020: 474−492

[25] Yu F X, Cao Liangliang, Feris R S, et al. Designing category-level attributes for discriminative visual recognition[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2013: 771−778

[26] Peng Peixi, Tian Yonghong, Xiang Tao, et al. Joint learning of semantic and latent attributes[C]//Proc of the European Conf on Computer Vision. Cham: Springer, 2016: 336−353

[27] Song Jie, Shen Chengchao, Lei Jie, et al. Selective zero-shot classification with augmented attributes[C]//Proc of the European Conf on Computer Vision. Cham: Springer, 2018: 468−483

[28] Xian Yongqin, Lorenz T, Schiele B, et al. Feature generating networks for zero-shot learning[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 5542−5551

[29] Xian Yongqin, Sharma S, Schiele B, et al. F-VAEGAN-D2: A feature generating framework for any-shot learning[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 10275−10284

[30] Gidaris S, Komodakis N. Generating classification weights with GNN denoising autoencoders for few-shot learning[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 21−30

[31] Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning[C]//Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2017: 4077−4087

[32] Afrasiyabi A, Lalonde J F, Gagné C. Associative alignment for few-shot image classification[C]//Proc of the European Conf on Computer Vision. Cham: Springer, 2020: 18−35

[33] Guan Jiechao, Lu Zhiwu, Xiang Tao, et al. Zero and few shot learning with semantic feature synthesis and competitive learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(7): 2510−2523

[34] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast

adaptation of deep networks[C]//Int Conf on Machine Learning. New York: PMLR, 2017: 1126−1135

[35] Ravi S, Larochelle H. Optimization as a model for few-shot learning[C]// Proc of Int Conf on Learning Representations. 2016 [2024-05-12]. https://openreview.net/pdf?id=rJY0-Kcll

[36] Kang Bingyi, Xie Saining, Rohrbach M, et al. Decoupling representation and classifier for long-tailed recognition[J]. arXiv preprint, arXiv: 1910.09217, 2019

[37] Tan Jingru, Wang Changbao, Li Buyu, et al. Equalization loss for long-tailed object recognition[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 11662−11671

[38] Kemker R, Kanan C. FearNet: Brain-inspired model for incremental learning[J]. arXiv preprint, arXiv: 1711.10563, 2017

[39] Rusu A A, Rabinowitz N C, Desjardins G, et al. Progressive neural networks[J]. arXiv preprint, arXiv: 1606.04671, 2016

[40] Kirkpatrick J, Pascanu R, Rabinowitz N, et al. Overcoming catastrophic forgetting in neural networks[J]. Proceedings of the National Academy of Sciences, 2017, 114(13): 3521−3526

[41] Zenke F, Poole B, Ganguli S. Continual learning through synaptic intelligence[C]//Proc of Int Conf on Machine Learning. New York: PMLR, 2017: 3987−3995

[42] Rebuffi S A, Kolesnikov A, Lampert C H. iCaRL: Incremental classifier and representation learning[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 5533−5542

[43] Lopez-Paz D, Ranzato M. Gradient episodic memory for continual learning[C]//Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2017: 6467−6476

[44] Dhamija A, Gunther M, Ventura J, et al. The overlooked elephant of object detection: Open set[C]//Proc of the IEEE/CVF Winter Conf on Applications of Computer Vision. Piscataway, NJ: IEEE, 2020: 1021−1030

[45] Miller D, Nicholson L, Dayoub F, et al. Dropout sampling for robust object detection in open-set conditions[C]// Proc of Int Conf on Robotics and Automation. Piscataway, NJ: IEEE, 2018: 3243−3249

[46] Du Xuefeng, Wang Zhaoning, Cai Mu, et al. VOS: Learning what you don't know by virtual outlier synthesis[J]. arXiv preprint, arXiv: 2202.01197, 2022

[47] Rahman S, Khan S, Porikli F. Zero-shot object detection: Learning to simultaneously recognize and localize novel concepts[C]//Proc of the Asian Conf on Computer Vision. Cham: Springer, 2018: 547−563

[48] Hayat N, Hayat M, Rahman S, et al. Synthesizing the unseen for zero-shot object detection[C]//Proc of the Asian Conf on Computer Vision. Cham: Springer, 2020: 155−170

[49] Yan Xiaopeng, Chen Ziliang, Xu Anni, et al. Meta R-CNN: Towards general solver for instance-level low-shot learning[C]//Proc of the IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 9577−9586

[50] Wang Xin, Huang T E, Darrell T, et al. Frustratingly simple few-shot object detection[J]. arXiv preprint, arXiv: 2003.06957, 2020

[51] Feng Chengjian, Zhong Yujie, Huang Weilin. Exploring classification equilibrium in long-tailed object detection[C]//Proc of the IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2021: 3417−3426

[52] Li Yu, Wang Tao, Kang Bingyi, et al. Overcoming classifier imbalance for long-tail object detection with balanced group softmax[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 10991−11000

[53] Joseph K J, Rajasegaran J, Khan S, et al. Incremental object detection via meta-learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(12): 9209−9216

[54] Radford A, Kim W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//Proc of the Int Conf on Machine Learning. New York: PMLR, 2021: 8748−8763

[55] Kirillov A, Mintun E, Ravi N, et al. Segment anything[J]. arXiv preprint, arXiv: 2304.02643, 2023

[56] Lampert C H, Nickisch H, Harmeling S. Learning to detect unseen object classes by between-class attribute transfer[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2009: 951−958

[57] Wah C, Branson S, Welinder P, et al. The Caltech-UCSD Birds−200−2011 dataset[R]. Pasadena, CA: California Institute of Technology, 2011

[58] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning[C]//Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2016: 3630−3638

[59] Ren Mengye, Triantafillou E, Ravi S, et al. Meta-learning for semi-supervised few-shot classification[J]. arXiv preprint, arXiv: 1803.00676, 2018

[60] Horn V G, Aodha O M, Song Yang, et al. The inaturalist species classification and detection dataset[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 8769−8778

[61] Everingham M, Van G L, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88: 303−338

[62] Kolve E, Mottaghi R, Gordon D, et al. AI2-THOR: An interactive 3D environment for visual AI[J]. arXiv preprint, arXiv: 1712.05474, 2017

[63] Wu Zhifan, Wei Tong, Jiang Jianwen, et al. NGC: A unified framework for learning with open-world noisy data[C]//Proc of the IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2021: 62−71

[64] Wang Yisen, Liu Weiyang, Ma Xingjun, et al. Iterative learning with open-set noisy labels[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 8688−8696

[65] Tachet des Combes R, Zhao Han, Wang Yuxiang, et al. Domain adaptation with conditional distribution matching and generalized label shift[C]//Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2020, 33: 19276−19289

[66] Zhang Yuchen, Liu Tianle, Long Mingsheng, et al. Bridging theory and algorithm for domain adaptation[C]//Proc of the Int Conf on

Machine Learning. New York: PMLR, 2019: 7404−7413

[67] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 779−788

[68] Qin Zheng, Li Zeming, Zhang Zhaoning, et al. ThunderNet: Towards real-time generic object detection on mobile devices[C]//Proc of the IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 6718−6727

[69] Gao Fei, Yang Liu, Li Hui. A survey on open set recognition[J]. Journal of Nanjing University (Natural Sciences), 2022, 58(1): 115−134 (in Chinese)
(高菲, 杨柳, 李晖. 开放集识别研究综述[J]. 南京大学学报(自然科学版), 2022, 58(1): 115−134)

[70] Lampert C H, Nickisch H, Harmeling S, et al. Learning to detect unseen object classes by between-class attribute transfer[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2009: 951−958

[71] Bansal A, Sikka K, Sharma G, et al. Zero-shot object detection[C]//Proc of the European Conf on Computer Vision. Cham: Springer, 2018: 384−400

[72] Sung F, Yang Yongxin, Zhang Li, et al. Learning to compare: Relation network for few-shot learning[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 1199−1208

[73] Kang Bingyi, Liu Zhuang, Wang Xin, et al. Few-shot object detection via feature reweighting[C]//Proc of the IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 8420−8429

[74] Liu Ziwei, Miao Zhongqi, Zhan Xiaohang, et al. Large-scale long-tailed recognition in an open world[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 2537−2546

[75] Shmelkov K, Schmid C, Alahari K. Incremental learning of object detectors without catastrophic forgetting[C]//Proc of the IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 3400−3409

[76] Chang A X, Funkhouser T, Guibas L, et al. ShapeNet: An information-rich 3D model repository[J]. arXiv preprint, arXiv: 1512.03012, 2015

[77] Li Yanan, Li Pengyang, Cui Han, et al. Inference fusion with associative semantics for unseen object detection[C]//Proc of the AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2021: 1993−2001

[78] Huang Peiliang, Han Junwei, Cheng De, et al. Robust region feature synthesizer for zero-shot object detection[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 7622−7631

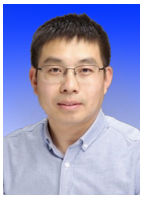
[79] Sarma S, Kumar S, Sur A. Resolving semantic confusions for improved zero-shot detection[C]//Proc of the British Machine Vision Conf. Durham: BMVA, 2023: 347−361

[80] Tan Jingru, Lu Xin, Zhang Gang, et al. Equalization loss v2: A new gradient balance approach for long-tailed object detection[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 1685−1694

[81] Wang Jiaqi, Zhang Wenwei, Zang Yuhang, et al. Seesaw loss for long-tailed instance segmentation[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 9695−9704

Nie Hui, born in 1996. PhD candidate. His main research interests include open world object detection and computer vision.

聂　晖, 1996 年生. 博士研究生. 主要研究方向为开放世界物体检测、计算机视觉.



Wang Ruiping, born in 1981. PhD, professor, PhD supervisor. His main research interests include computer vision, pattern recognition, and machine learning.

王瑞平, 1981 年生. 博士, 教授, 博士生导师. 主要研究方向为计算机视觉、模式识别、机器学习.

Chen Xilin, born in 1965. PhD, professor, PhD supervisor. His main research interests include computer vision, pattern recognition, image processing, and multimodal interfaces.

陈熙霖, 1965 年生. 博士, 教授, 博士生导师. 主要研究方向为计算机视觉、模式识别、图像处理、多模式人机接口.