

# 基于因果提示蒸馏的开放世界目标检测

赵佳琦<sup>1,2</sup>, 王平安<sup>1,2</sup>, 周 勇<sup>1,2\*</sup>, 杜文亮<sup>1,2</sup>, 姚 睿<sup>1,2</sup>, 刘 兵<sup>1,2</sup>

(1. 中国矿业大学计算机科学与技术学院, 江苏徐州 221116; 2. 矿山数字化教育部工程研究中心, 江苏徐州 221116)

**摘 要:** 开放世界目标检测旨在在动态环境中同时识别已知与未知类别,并在收到未知类别的标签后逐步实现对新增类别的识别能力。然而,现有方法因缺乏未知类别的语义表征能力,已知与未知类别间的指导信息相互耦合,导致检测性能受限。为此,本文提出一种基于因果提示蒸馏的开放世界目标检测方法。该方法创新性地将视觉-语言模型与因果推理结合,以解决开放场景中的类别间存在的语义偏差问题。具体而言,本文通过构建结构因果模型,从因果视角揭示了已知类别与未知类别间的语义干扰路径;接着提出了因果提示学习,通过生成未知类别的语义向量,显式引入开放场景的语义先验以增强模型对未知目标的感知能力;最后针对知识传递中的语义偏差问题提出因果蒸馏机制,利用双重蒸馏损失解耦教师模型对已知类别与未知类别的指导信息。实验结果表明,该方法在多个数据集上取得了良好效果,已知类别的平均检测精度(mAP)提升了1.3%,未知类别的召回率(U-Recall)提升了6.5%,这些结果验证了本文方法的有效性。

**关键词:** 提示学习;知识蒸馏;因果干预;开放世界;目标检测;计算机视觉

**基金项目:** 国家自然科学基金(No.62272461, No.62172417, No.62276266, No.62277046);中国矿业大学自主创新与社会服务双一流项目(No.2022ZZCX06);江苏省六大人才高峰项目(No.2015-DZXX-010, No.2018-XYDXX-044)

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 0372-2112(2025)06-2079-11

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20250211

## Open World Object Detection Based on Causal Prompt Distillation

ZHAO Jia-qi<sup>1,2</sup>, WANG Ping-an<sup>1,2</sup>, ZHOU Yong<sup>1,2\*</sup>, DU Wen-liang<sup>1,2</sup>, YAO Rui<sup>1,2</sup>, LIU Bing<sup>1,2</sup>

(1. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China;

2. Mine Digitization Engineering Research Center of the Ministry of Education, Xuzhou, Jiangsu 221116, China)

**Abstract:** Open world object detection aims to simultaneously identify both known and unknown categories in dynamic environments, while enabling incremental learning of new categories. However, due to the lack of semantic representation ability of unknown categories, the guidance information between known and unknown categories is mutually coupled, resulting in limited detection performance. To solve this problem, this paper proposes an open world object detection based on causal prompt distillation, which innovatively combines visual-language model with causal inference to solve the problem of semantic bias between categories in open scenes. Specifically, by constructing a structural causal model, this paper reveals the semantic interference path between known and unknown categories from the perspective of causality. Then, causal prompt learning is proposed, which explicitly introduces the semantic prior of the open scene by generating semantic vectors of unknown categories to enhance the model's perception of unknown objects. Finally, in order to solve the problem of semantic bias in knowledge transfer, a causal distillation mechanism is proposed, and the guidance information of the known and unknown categories is decoupled by the double distillation loss decoupling teacher model. Experimental results demonstrate that this method has achieved good effects on multiple datasets, with an improvement in mean average precision (mAP) for known categories by 1.3% and a rise in recall rate (U-Recall) for unknown categories by 6.5%. These results validate the effectiveness and robustness of the proposed approach.

**Key words:** prompt learning; knowledge distillation; causal intervention; open world; object detection; computer vision

**Foundation Item(s):** National Natural Science Foundation of China (No.62272461, No.62172417, No.62276266, No.62277046); Double First-Class Project of China University of Mining and Technology for Independent Innovation and Social Service (No.2022ZZCX06); Six Talent Peaks Project in Jiangsu Province (No.2015-DZXX-010, No.2018-XYDXX-044)

## 1 引言

随着计算机视觉<sup>[1]</sup>技术的快速发展,目标检测<sup>[2]</sup>作为场景理解的核心任务,已在自动驾驶<sup>[3]</sup>、行人重识别<sup>[4]</sup>等动态环境中得到广泛应用.然而,传统目标检测方法依赖预定义封闭类别标签的局限性日益凸显,难以适应开放世界中未知类别的持续涌现.开放世界目标检测<sup>[5]</sup>(Open World Object Detection, OWOD)旨在解决这一挑战,要求模型既能识别已知类别目标,又能主动发现未知类别目标.模型在发现未知类别目标后会赋予未知类别标签,并在获取其标签后,通过增量学习<sup>[6]</sup>的方式逐步掌握这些新类别,持续扩展模型的认知边界.开放世界目标检测的核心难点在于如何在没有未知类别标注的情况下,实现已知与未知类别的精准区分,并避免增量学习中的灾难性遗忘.值得注意的是,开放集识别(Open Set Recognition, OSR)与开放世界目标检测虽同属开放环境下的学习范式,但存在显著差异.OSR的核心目标是在推理阶段识别未知类别,但无需对未知类别进行增量学习;而OWOD不仅要求检测未知目标,还需在获取其标注后通过增量学习持续扩展模型的认知能力.

现有的开放世界目标检测方法大多通过复杂的筛选策略进行推理,依赖先验知识或启发式方法来辨识未知类别.然而,这些方法在面对不断变化的环境时,往往难以保持高效的语义信息利用,并且在实际应用中容易受到类别间相似性的干扰.此外,现有的增量学习方法虽然在一定程度上解决了新类别的学习问题,但由于无法有效地整合已知类别的知识和未知类别的推理,这些方法的泛化能力仍然较为有限.本文引入视觉-语言模型<sup>[7]</sup>,为开放世界目标检测提供了新的思路.

典型的视觉-语言模型CLIP<sup>[8]</sup>(Contrastive Language-Image Pre-training)通过在大规模图像-文本数据集上进行对比学习,能够将图像和文本特征映射到同一特征空间,从而为模型在开放世界场景中的语义理解提供了巨大潜力.然而,受到Li等人<sup>[9]</sup>在少样本目标检测<sup>[10]</sup>中应用CLIP的启发,本文通过实验验证了直接将CLIP引入开放世界目标检测任务进行知识蒸馏<sup>[11]</sup>并未显著提升模型性能.一方面,在蒸馏过程中,CLIP作为指导模型在对未知类别的监督能力上存在不足,可能导致学生模型在未知类别上的泛化能力受到限制.另一方面,CLIP未能有效地区分已知类别与未知类别的指导信息,这限制了模型性能的进一步提升.

针对上述问题,本文提出了一种基于因果提示蒸馏的开放世界目标检测,结合因果干预<sup>[12]</sup>思想来增强模型对未知类别的感知能力.开放世界目标检测任务中,训练数据往往缺乏未知类别的标注信息,而现有方法大多依赖类别间的相关性进行学习,这导致了偏差

的引入.例如,已知类别的语义信息可能会干扰未知类别的检测,导致模型无法准确区分两者.因果干预的核心目标是通过揭示变量间的因果关系,从根本上解决由相关性引起的模型偏差问题.本文通过引入因果机制显式优化提示信息的作用,分离已知类别与未知类别的指导信息,从而提升模型对未知类别的感知能力和整体泛化性能.综上所述,本文的主要贡献包括以下:

(1)引入结构因果模型分析开放世界目标检测任务中的潜在语义偏差来源,并通过因果干预提供了解决这一问题的理论支持.

(2)提出了一种结合因果干预的提示学习,利用视觉-语言模型CLIP的文本编码器生成未知类别的语义向量,通过显式干预增强模型对未知类别的感知能力.

(3)提出了一种因果蒸馏机制,针对知识传递中的语义偏差,设计了双重蒸馏损失函数来分离已知类别与未知类别的指导信息,优化教师模型与学生模型之间的信息传递.

(4)多个公开的实验数据集表明,本文提出的方法总体优于对比方法,有效验证了本文方法的有效性.

## 2 相关工作

### 2.1 视觉-语言模型

视觉-语言模型<sup>[13]</sup>近年来在计算机视觉领域获得了广泛关注.通过在海量图像-文本对上进行对比学习,模型学习到统一的跨模态嵌入空间,实现了图像和文本的语义特征能够进行精确对齐.CLIP<sup>[8]</sup>作为代表性工作,通过在约4亿对图像-文本数据上训练,展现了强大的语义理解能力和泛化性能.在下游任务中,CLIP能够通过设计文本提示直接生成类别的语义嵌入,并与图像特征进行匹配,从而在零样本分类<sup>[14]</sup>、少样本学习<sup>[15]</sup>、行人重识别<sup>[16]</sup>等任务中取得了领先性能.视觉-语言模型在众多领域的成功应用为本文研究提供了启发,促使本文思考利用其强大的语义理解和泛化能力来缓解开放世界目标检测中语义表征不足的问题.

### 2.2 开放集识别

在计算机视觉领域,开放集识别(Open Set Recognition, OSR)是为应对真实世界场景需求而兴起的重要研究方向.OSR要求模型在推理时不仅能准确分类训练阶段见过的类别(即已知类别),还能有效识别出从未见过的新类别(即未知类别或实例),且通常无需对未知类别进行进一步处理.Scheirer等人<sup>[17]</sup>率先将此问题形式化,提出了开放集风险的概念,并通过约束已知类别在特征空间中的边界来减小该风险,但未能对未知类别区域施加有效约束.Feng等人<sup>[18]</sup>提出了一种多专家多元化注意力融合模块,旨在以判别性方式学习

多样的特征表征. 通过自适应融合各专家的分类得分, 并结合精心设计的评分函数来识别未知样本, 从而提升了模型对未知类别的辨识能力. Yang 等人<sup>[19]</sup>从因果推理视角出发, 提出了一种新颖的方法. 该方法引入协变量因果解耦模块, 旨在减轻不同类别间虚假相关性对判别特征泛化能力的负面影响, 从而使模型能更准确地识别未知类别, 提升了模型性能.

### 2.3 开放世界目标检测

尽管开放集识别与开放世界目标检测密切相关, 但 OWO 面临着更独特的挑战. 开放世界目标检测不仅要求模型识别训练时未见的实例并将其标记为“未知”, 还要求在接收到部分未知实例的人工标注后, 能够通过增量学习<sup>[6]</sup>的方式, 逐步将这些新类别纳入其知识体系, 实现模型的持续学习与迭代更新, 以适应动态变化的现实世界. Joseph 等人<sup>[20]</sup>首次提出了开放世界目标检测的概念, 并基于对比聚类和基于能量的未知识别方法强调模型在检测已知类别的同时能够识别并区分未知类别. 在此基础上, Ma 等人<sup>[21]</sup>提出基于标签迁移学习的策略, 通过退火优化增强了模型对未知类别的感知能力. 然而, 已知类别对未知类别的干扰问题仍未解决, 于是 Wang 等人<sup>[22]</sup>提出通过生成随机建议框的方法, 减少了已知类别特征对未知类别的干扰. 为了进一步提高未知类别的识别性能, Sun 等人<sup>[23]</sup>通过正交化策略减少了类别间的混淆, 以此提高了未知类别的识别性能. 尽管这些工作在提升 OWO 模型区分已知与未知类别的能力方面取得了进展, 但较少探索如何有效利用视觉-语言模型所蕴含的丰富先验语义知识. 本文旨在填补这一空白, 通过借助视觉-语言模型引入未知类别的语义先验, 并设计机制优化已知与未知类别指导信息的分离, 以期进一步提升模型对已知和未知目标的识别性能, 为 OWO 领域提供新的解决思路.

### 2.4 因果干预

因果干预近年来在机器学习和计算机视觉领域引起了广泛关注, 其核心目标是通过揭示变量间的因果关系, 从根本上解决由相关性引起的模型偏差问题. Judea 等人<sup>[24]</sup>提出的结构因果模型以因果图形式描述变量间的依赖关系, 并通过干预操作阻断混淆路径. 在计算机视觉领域, 因果干预被广泛应用于消除伪相关性和偏差. 邵志文等人<sup>[25]</sup>将因果干预应用于面部动作单元识别, 通过调整经验风险缓解标签不平衡引起的偏差. 前述提及的 Wang 等人<sup>[22]</sup>利用结构因果模型分析了候选区域生成过程中已知特征对未知类别的干扰, 并借助扩散模型来生成随机候选框以减少此干扰. 上述方法的成功应用给予了本文利用因果学习的理论解决开放世界目标检测问题的新思路, 通过借助因果干预思想, 从语义角度分析类别间的伪相关性, 并通过

显示干预来减轻类别间的干扰.

### 2.5 提示学习

提示学习<sup>[26]</sup>是一种新兴且有效的方法, 尤其在大型预训练模型的应用中表现突出. 该方法源于自然语言处理领域<sup>[27]</sup>, 其核心思想是通过设计合适的提示, 下游特定任务形式重新调整为接近模型预训练时的任务形式, 从而充分利用预训练模型的能力. 提示通常以自然语言的形式嵌入到输入中, 引导模型根据上下文生成与任务相关的输出. 例如, CoOp<sup>[28]</sup> (Context Optimization) 提出学习连续的向量来表示提示中的上下文词, 实现了提示工程的自动化, 并提升了模型在下游图像识别任务中的少样本或零样本性能. 王楠井等人<sup>[29]</sup>提出了一种基于图像内容理解的判别性类别提示学习方法, 通过结合图像的内容特征和自注意力机制, 提升了文本提示在图像分类中的判别性. 受上述方法的启发, 提示学习可以帮助开放世界目标检测模型挖掘出已知类别或未知类别的语义信息, 并结合图像特征进行目标类别的定位与识别.

### 2.6 知识蒸馏

知识蒸馏最早由 Hinton 等人<sup>[30]</sup>的工作提出, 近年来受到了广泛关注, 其核心思想是通过将大型复杂模型(教师模型)的知识传递给较小的模型(学生模型), 使得学生模型能够在保持较小规模的同时, 获得与教师模型相似甚至更好的性能. 鉴于 CLIP 模型通过海量图文数据预训练获得了强大的语义理解能力, 本文选择其作为教师模型. 本文期望通过知识蒸馏, 将 CLIP 的语义知识有效传递给 OWO 学生模型, 以增强其对已知及未知类别识别能力. 在传统目标检测领域中, Yang 等人<sup>[31]</sup>提出了焦点知识蒸馏和全局知识蒸馏, 目的是分离前背景, 赋予不同的权重, 并利用教师的空间与通道注意力作为权重, 共同指导学生模型进行学习, 计算重点蒸馏损失. Li 等人<sup>[9]</sup>在小样本目标检测中针对目标样本的标签是否属于正样本, 将知识蒸馏目标分解为四个特定的项, 使学生模型在知识蒸馏过程中学习到正确的相关性. 上述方法的成功实践给予开放世界目标检测一定的启发, 本文结合因果干预思想, 利用知识蒸馏对已知类别和未知类别进行显式的分离蒸馏, 并利用双重蒸馏损失解耦教师模型对已知类别与未知类别的指导信息.

## 3 本文方法

### 3.1 问题定义

开放世界目标检测包含  $T$  个增量任务. 在第  $t$  个任务中 ( $t \in \{1, 2, \dots, T\}$ ), 已知类别和未知类别分别表示  $K_t$  和  $U_t$ , 其中  $K_t \cap U_t = \emptyset$  且  $U_t \in U$ . 在训练阶段, 类别标签  $y^{K_t}$  和边界框标注 ( $b = [x, y, w, h]$ ) 仅分配给已知类别

的实例. 然而, 在推理阶段, 已知类别  $K$  和未知类别  $U$  都存在, 并且应由开放世界目标检测模型识别. 已知实例应被准确分类到各自的类别中, 而未知实例应被识别为“unknown”类别. 在第  $t+1$  个任务中, 感兴趣的未知类别  $\bar{U}_t \in U$  被标记, 然后添加到已知类别中  $K_{t+1} = K_t \cup \bar{U}_t$ . 当前的未知类别集为  $U_{t+1} = U \setminus \bar{U}_t$ . 这个过程一直持续到  $U = \emptyset$ . 此外, 开放世界目标检测模型需要逐步识别之前的已知类别和当前的已知类别, 同时将剩余的未知类别识别为“unknown”类别. 值得注意的是, 每个任务仅存储少量之前已知类别的实例, 以进行增量学习, 从而节省计算和存储资源.

本文提出的因果提示蒸馏方法旨在解决开放世界目标检测任务中未知类别标注缺失所导致的监督偏差问题, 总体框架如图 1 所示. 首先借助因果推断方法, 构建结构因果模型分析任务中的潜在偏差来源. 图 1①是开放世界目标检测的基础模型. 图 1②呈现的是因果提示学习, 通过生成未知类别的语义向量, 并借助因果干预将其有效地融入模型, 以显式增强对未知目标的语义感知. 图 1③所示的是因果蒸馏机制, 通过设计双重蒸馏损失函数来解耦教师模型中对已知与未知类别的指导信息, 从而缓解知识传递中的语义偏差, 提升学生模型的泛化性.

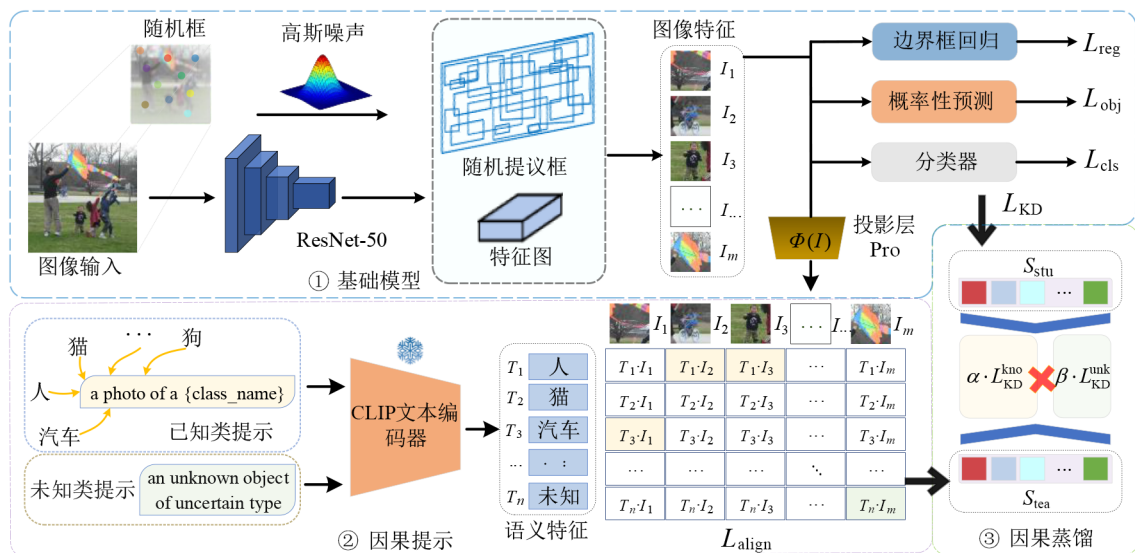


图 1 网络结构示意图

### 3.2 结构因果模型

在开放世界目标检测任务中, 训练数据缺乏未知类别的标注信息, 现有方法依赖类别间的相关性, 这容易引入类别间的语义偏差. 为揭示这种偏差来源并设计有效的解决方案, 本文借鉴因果推断方法, 构建结构因果模型分析开放世界目标检测的因果关系.

图 2 展示了本文所构建的结构因果模型. 其中,  $X$  表示输入图像特征;  $K$  表示模型从训练数据中学习到的已知类别的语义特征, 该特征用于指导模型对已知类别的识别;  $P$  包含已知类别的提示信息 and 未知类别的潜在信息;  $Y$  表示模型对输入图像的最终检测结果, 包括类别标签和位置信息; 路径  $X \rightarrow K$  表示输入图像的图像特征通过已知类别的语义知识  $K$  影响模型的检测能力, 路径  $K \rightarrow Y$  表示已知类别的语义知识  $K$  决定了模型的检测结果, 路径  $P \rightarrow K$  表示提示信息对已知类别语义表达的影响, 而  $P \rightarrow Y$  表示提示信息面对前景类或背景类时的直接影响, 特别是在未知类别的检测过程中, 模型没有未知类的先验知识而影响模型判断.

在开放世界目标检测任务中, 提示信息  $P$  的存在既

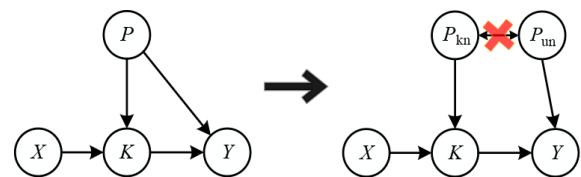


图 2 结构因果示意图

可能有助于模型学习, 也可能引入干扰. 已知类别的提示信息通过  $P \rightarrow K$  提供语义指导信息, 从而增强模型对已知类别的学习能力. 然而, 未知类别的潜在语义信息可能通过两条路径影响检测结果. 一方面,  $P \rightarrow K$  会影响语义学习  $K$ , 引发已知类别和未知类别之间的混淆; 另一方面,  $P \rightarrow Y$  会通过前景和背景的分直接影检测结果  $Y$ , 导致未知类别被误判为已知类别或背景. 这些偏差路径表明, 传统的相关性学习方法无法有效区开放世界目标检测的已知类别和未知类别. 因此, 通过因果干预显式控制提示信息  $P$  的作用可消除其负向影响. 基于上述分析, 本文提出了对提示信息  $P$  进行显式干预的策略, 从而优化提示信息在模型中的作用. 具

何地,本文使用式(1)表示 $P$ 对未知类别的干预作用,式(2)表示 $P$ 对已知类别的干预作用.

$$P(Y|\text{do}(P)) = P(Y|P_{\text{un}}, X) \quad (1)$$

$$P(Y|\text{do}(P)) = P(Y|K) \cdot P(K|P_{\text{kn}}, X) \quad (2)$$

其中,  $\text{do}(P)$  表示对提示信息  $P$  的干预,  $P(Y|P_{\text{un}}, X)$  表示未知类预测受提示信息  $P_{\text{un}}$  和输入特征  $X$  的影响,  $P(K|P_{\text{kn}}, X)$  表示已知类别信息  $P_{\text{kn}}$  和输入特征  $X$  一起决定语义学习  $K$  的优化,  $P(Y|K)$  表示语义知识  $K$  对最终检测结果的决定作用.

### 3.3 因果提示学习

为进一步提升模型对未知类别的检测性能,并具体落实上一节提出的因果干预思想,本文提出了一种因果提示学习方法. 该方法的核心在于主动引入并有效利用未知类别的语义信息,旨在增强模型区分已知与未知类别的能力,并为这两类目标提供解耦的、有针对性的语义指导.

本文采用视觉-语言模型 CLIP 的文本编码器,将所有已知类别的名称映射到高维语义空间中. 具体而言,每个类别  $c_i$  被设计为一个标准文本提示,例如“a photo of a {class\_name}”,并通过 CLIP 文本编码器生成对应类别的语义向量  $t_i$ ,如式(3)所示:

$$\begin{cases} T = \{t_1, t_2, \dots, t_n\} \\ t_i = \text{Encoder}_{\text{CLIP}}(c_i) \end{cases}, i = 1, 2, \dots, n \quad (3)$$

其中,生成的语义向量  $t_i$  为  $d$  维. 这些语义向量捕捉了类别之间的语义关系,为后续图像-文本特征对齐提供了先验信息. 为了增强对未知类别上的建模能力,本文设计了一种特殊的文本提示“an unknown object of uncertain type”,并同样通过 CLIP 的文本编码器生成未知类别的语义向量  $t_u$ . 该设计旨在对结构因果图中的路径  $P \rightarrow Y$  进行干预,显式引入未知类别的语义信息,使得模型不仅依赖已知类别的语义信息,也能为未知类别提供指导. 在下一个任务中,模型确认某一类别为未知类别后,使用对应的已知类别提示“a photo of a {class\_name}”进行语义上的特征提取,确保  $P \rightarrow Y$  的过程只进行一种未知类提示的干预,避免已知类别和未知类别之间产生关联.

在提示学习过程中,输入图像首先通过骨干网络提取特征图,并基于随机提议框从中选取类别无关的边界框,得到区域特征集合,  $\varphi(I) = \{r_1, r_2, \dots, r_m\}$ , 每个特征的维度为  $c$ , 随后,这些区域特征通过线性映射层  $\text{Pro}$  投影到与文本特征相同的  $d$  维语义空间中,生成投影后特征  $F = \{f_1, f_2, \dots, f_m\}$ , 具体如式(4)所示:

$$F = \text{Pro}(\varphi(I)) \quad (4)$$

其中,映射层  $\text{Pro}$  通过全连接层实现,输入维度为  $c$ , 输

出维度为  $d$  (与 CLIP 文本编码器输出维度一致), 从而完成图像与文本特征的对齐. 尽管 CLIP 的图像编码器在多模态特征对齐方面表现出色,但其结构复杂且计算量大. 为追求效率,本文采用一个轻量级的线性投影层来替代完整的 CLIP 图像编码器,仅通过单层线性变换实现高效的特征空间映射,这有助于降低计算和内存开销,同时保留了关键的语义对齐能力. 为了度量图像区域特征  $f_j$  与类别语义特征  $t_i$  的匹配程度,本文计算两者之间的余弦相似度,如式(5)所示:

$$S_{i,j} = \cos(f_j, t_i) = \frac{f_j \cdot t_i}{|f_j| \cdot |t_i|} \quad (5)$$

其中,  $S_{i,j}$  表示图像区域  $f_j$  与类别  $c_i$  的相似度得分,该得分在后续因果蒸馏过程中用于辅助分类与优化. 为进一步优化图像区域与类别语义特征之间的对齐,通过引入对齐损失  $L_{\text{align}}$ , 约束图像区域特征与类别语义向量之间的特征偏差,对齐损失计算方式如式(6)所示:

$$L_{\text{align}} = \sum_{i=1}^n \sum_{j=1}^m \left( 1 - \frac{f_j \cdot t_i}{|f_j| \cdot |t_i|} \right) \quad (6)$$

其中,  $m$  表示图像候选框数量,  $n$  表示类别语义数量. 该损失函数鼓励模型最大化图像区域特征与正确类别语义特征之间的相似度,从而加强图像与类别之间的对齐.

通过因果提示学习,引入未知类别的语义向量  $t_u$ , 显式干预结构因果模型中的路径  $P \rightarrow Y$ , 增强模型对未知目标的感知能力,并通过多模态特征对齐优化,确保图像候选区域特征能够与已知类别及未知类别的语义向量进行联合匹配,实现已知与未知类别在统一语义空间下的判别.

### 3.4 因果蒸馏机制

在因果提示的基础上,为了进一步缓解开放世界目标检测任务中由已知与未知类别监督信号耦合而导致的语义偏差问题,本文提出了一种因果蒸馏机制. 该机制的核心在于通过因果干预的思想指导知识蒸馏的过程,为此设计了双重蒸馏损失,用于显式地解耦教师模型中针对已知类别和未知类别的指导信息.

在开放世界目标检测中,本文设定教师模型为大规模预训练的 CLIP,其强大的语义理解能力能够为学生模型(即基准目标检测器)提供丰富的知识指导. 通过将教师模型的语义知识传递给学生模型,学生模型能够更好地理解未知类别的语义特征,提高对未知类别的识别能力. 知识蒸馏的核心是通过计算教师模型和学生模型输出概率分布之间的 KL (Kullback-Leibler) 散度,量化两者的差异并将其最小化. KL 散度的损失函数定义如式(7)所示:

$$L_{\text{KD}} = \text{KL}(P^{\text{tea}} \| P^{\text{stu}}) = \sum_{i=1}^{N_C} P_i^{\text{tea}} \log \left( \frac{P_i^{\text{tea}}}{P_i^{\text{stu}}} \right) \quad (7)$$

其中,  $\text{KL}(\cdot\|\cdot)$  表示 KL 散度,  $P^{\text{tea}}$  和  $P^{\text{stu}}$  分别表示教师模型和学生模型的预测概率分布,  $N^C$  表示当前任务的类别总数,  $P_i^{\text{tea}}$  和  $P_i^{\text{stu}}$  分别为类别  $i$  的概率预测值. 本文的目标是通过最小化式(7), 提升学生模型对已知类别和未知类别的学习能力.

然而, 正如本文在后续实验章节所验证的, 若将这种传统蒸馏机制直接应用于开放世界目标检测, 可能会导致知识传递中的语义偏差, 即教师模型的指导信息未能有效地区分已知类别与未知类别. 本文提出的因果蒸馏机制旨在通过受因果启发的解耦, 来分离已知类别与未知类别之间的指导信息, 从而优化蒸馏过程中对已知类别和未知类别的指导信息传递, 使得教师模型和学生模型之间可以传递正确的知识, 减少已知类别对未知类别的干扰, 提高模型的学习效率与泛化能力.

为了区分已知类别和未知类别的指导信息, 本文引入了一种双重 KL 散度蒸馏损失. 首先, 分别为教师模型和学生模型的输出  $S_{\text{tea}}$  和  $S_{\text{stu}}$  引入温度缩放  $T$ , 用于平滑二者的概率分布, 如式(8)和式(9)所示:

$$p_{\text{tea}} = \text{softmax}(S_{\text{tea}}/T) \quad (8)$$

$$p_{\text{stu}} = \text{softmax}(S_{\text{stu}}/T) \quad (9)$$

其中,  $T$  为温度参数, 用于放大小概率类别的贡献, 从而增强蒸馏信号在类别区分上的表现. 接下来, 本文定义了针对已知类别和未知类别的独立蒸馏损失, 如式(10)和式(11)所示:

$$L_{\text{KD}}^{\text{kno}} = \frac{1}{I_{\text{kno}}} = \sum_{i \in I_{\text{kno}}} \text{KL}(p_{\text{tea}}^i \| p_{\text{stu}}^i) \quad (10)$$

$$L_{\text{KD}}^{\text{unk}} = \frac{1}{I_{\text{unk}}} = \sum_{i \in I_{\text{unk}}} \text{KL}(p_{\text{tea}}^i \| p_{\text{stu}}^i) \quad (11)$$

其中,  $I_{\text{kn}}$  和  $I_{\text{un}}$  分别表示已知类别和未知类别的样本索引集合. 最终, 本文通过加权组合已知类别与未知类别的损失项, 定义的总蒸馏损失函数如式(12)所示:

$$L_{\text{KD}} = \alpha \cdot L_{\text{KD}}^{\text{kno}} + \beta \cdot L_{\text{KD}}^{\text{unk}} \quad (12)$$

其中,  $\alpha$  和  $\beta$  为权重系数, 分别控制已知类别与未知类别损失的贡献度.

通过对教师模型的输出进行干预, 显式优化了已知类别与未知类别之间的指导信息分离, 使得蒸馏过程中对未知类别的指导信息得到强化, 同时避免了已知类别的干扰. 通过这种方式, 因果蒸馏机制不仅提升了模型在已知类别上的精度, 也增强了其在未知类别检测中的鲁棒性, 进一步优化了模型的泛化能力.

## 4 实验结果与分析

### 4.1 数据集及实验环境

为评估本文提出方法的性能, 本文在 2 个公开的开放世界目标检测数据集 MS-COCO<sup>[32]</sup> 和 Pascal VOC<sup>[33]</sup>

上进行了实验. 对于 M-OWOD 任务, 将 MS-COCO 数据集的 80 个类别划分为 4 个任务, 每个任务的训练集为 MS-COCO 和 Pascal VOC 中的训练集, 验证集也是两个数据集的测试集, 数据划分表如表 1 所示. 对于 S-OWOD 任务, 为避免跨任务和跨数据集的泄漏问题, 仅使用 MS-COCO 的训练集和测试集, 每个任务类别划分一致. 对于增量学习任务, 将 Pascal VOC 的数据集分为 3 类增量学习设置, 分别是 10+10 设置、15+5 设置和 19+1 设置. 每个阶段逐步引入新的类别, 增量学习的训练集和测试集使用 Pascal VOC 数据集的训练集和测试集.

表 1 数据集划分表

单位: 幅

	Task 1	Task 2	Task 3	Task 4
训练图像	16 551	45 520	39 402	40 260
测试图像	4 952	1 914	1 642	1 738
训练实例	47 223	113 741	114 452	138 996
测试实例	14 976	4 966	4 826	6 093

本文所有的实验基于 Pytorch, 代码依赖于 Detectron2, 使用 4 块显存为 11 GB 的 RTX2080Ti 显卡; 实验模型在 OrthoDet<sup>[20]</sup> 的基础上进行改进, 随机提议框引用 RandBox<sup>[19]</sup> 中的高斯扩散方法, 得到类别无关的边界框, 骨干网络为 ResNet-50; 优化器使用 AdamW<sup>[34]</sup>, 学习率为  $2.5 \times 10^{-5}$ , 权重衰减为  $1 \times 10^{-4}$ , 批量大小为 12.

### 4.2 评价指标

为全面评估模型的性能, 本文采用两类核心指标 mAP 和 U-Recall. 对于已知类别, 使用全类平均精度 (mAP) 作为度量指标, 用于评估已知类别的检测性能; 对于未知类别, 使用未知类别召回率 (U-Recall) 作为主要指标, 用于衡量模型对未知类别的检索能力. 为了进一步评估开放世界目标检测中增量任务中的灾难性遗忘, mAP 分为先前已知 (Previously Known, PK) 和当前已知 (Current known, CK) 两类, PK 表示在当前任务之前已经是已知类的 mAP, CK 表示在当前任务中检测到的新的已知类别的 mAP. 此外, 在增量学习任务中, 本文也进行了相关的实验验证, 并使用 mAP 衡量模型在不同增量设置下对所有类别的综合检测性能.

### 4.3 实验对比

本文方法在 M-OWOD、S-OWOD 和增量学习任务上均表现出色. 表 2 展示了本文方法与现有主流方法 (包括 OST<sup>[35]</sup>、OW-DETR<sup>[36]</sup>、UC-OWOD<sup>[37]</sup>、PROB<sup>[38]</sup>、CAT<sup>[39]</sup> 等) 的对比结果.

从表 2 中可以看出, 对于已知类别 (mAP), 本文方法在 Task1、Task3 和 Task4 任务中取得了最优结果, 分别为 62.3%、42.0% 和 39.2%, 在 Task2 上接近最先进的 OrthoDet (46.4% 对比 47.0%). 对于未知类别 (U-Recall), 本文方法在前 3 个任务中的 U-Recall 分别为 25.1%、

29.3%和35.6%,均优于其他方法,体现了显著的性能提升.值得注意的是,此时80个类别已经被模型完全观察到,不存在未知类别,于是在Task4中没有比较U-Recall.

表3展示了S-OWOD任务的对比结果.对于已知类别(mAP),本文方法在所有任务上均优于其他模型(Task1~Task4分别为72.8%、52.4%、49.3%和48.6%).对于未知类别(U-Recall),本文方法在Task1和Task3中取

得了显著的性能提升,分别达到38.0%和34.9%.

表4中对比了本文方法与ILOD<sup>[40]</sup>、Faster ILOD<sup>[41]</sup>、Meta-ILOD<sup>[42]</sup>等方法的增量学习结果.3种增量设置(10+10设置、15+5设置和19+1设置)下,本文方法均取得了最优性能.其中,10+10设置的mAP达到75.4%,相比于OrthoDet的72.3%有显著提升;15+5设置的mAP达到77.1%,优于现有方法的最高值74.7%;19+1设置的mAP达到77.3%,继续保持领先.

表2 基于M-OWOD任务的实验结果

单位:%

Task IDs(→)	Task 1		Task 2				Task 3			Task 4			
Method	U-Recall(↑)	mAP(↑)	U-Recall(↑)	mAP(↑)			U-Recall(↑)	mAP(↑)			mAP(↑)		
		CK		PK	CK	Both		PK	CK	Both	PK	CK	Both
ORE	4.9	56.0	2.9	52.7	26.0	39.4	3.9	38.2	12.7	29.7	29.6	12.4	25.3
OST	—	56.2	—	53.4	26.5	39.9	—	38.0	12.8	29.6	30.1	13.3	25.9
OW-DETR	7.5	59.2	6.2	53.6	33.5	42.9	5.7	38.3	15.8	30.8	31.4	17.1	27.8
UC-OWOD	—	50.7	—	33.1	30.5	31.8	—	28.8	16.3	24.6	25.6	15.9	23.2
ALLOW	13.6	59.3	10.0	53.2	34.0	45.6	14.3	42.6	26.7	38.0	33.5	21.8	30.6
PROB	19.4	59.5	17.4	55.7	32.2	44.0	19.6	43.0	22.2	36.0	35.7	18.9	31.5
CAT	23.7	60.0	19.1	55.5	32.7	44.1	24.4	42.8	18.7	34.8	34.4	16.6	29.9
RandBox	10.6	61.8	6.3	—	—	45.3	7.8	—	—	39.4	—	—	35.4
OrthoDet	24.6	61.3	26.3	55.5	38.5	47.0	29.1	46.7	30.6	41.3	42.4	24.3	37.9
Ours	25.1	62.3	29.9	55.8	37.0	46.4	35.6	47.0	32.1	42.0	43.4	26.4	39.2

表3 基于S-OWOD任务的实验结果

单位:%

Task IDs(→)	Task 1		Task 2				Task 3			Task 4			
Method	U-Recall(↑)	mAP(↑)	U-Recall(↑)	mAP(↑)			U-Recall(↑)	mAP(↑)			mAP(↑)		
		CK		PK	CK	Both		PK	CK	Both	PK	CK	Both
ORE	1.5	61.4	3.9	56.5	26.1	40.6	3.6	38.7	23.7	33.7	33.6	26.3	31.8
OW-DETR	5.7	71.5	6.2	62.8	27.5	43.8	6.9	45.2	24.9	38.5	38.2	28.1	33.1
PROB	17.6	73.4	22.3	66.3	36.0	50.4	24.8	47.8	30.4	42.0	42.6	31.7	39.9
CAT	24.0	74.2	23.0	67.6	35.5	50.7	24.6	51.2	32.6	45.0	45.4	35.1	42.8
OrthoDet	24.6	71.6	27.9	64.0	39.9	51.3	31.9	52.1	42.2	48.8	48.7	38.8	46.2
Ours	38.0	72.8	25.9	64.4	41.5	52.4	34.9	52.5	43.0	49.3	50.5	42.3	48.5

表4 基于增量学习任务的实验结果

单位:%

Method	ILOD	Faster ILOD	ORE	Meta-ILOD	ROSETTA	OW-DETR	PROB	CAT	OrthoDet	Ours
10+10设置 mAP	63.2	62.1	64.5	66.3	66.8	65.7	66.5	67.7	72.3	75.4
15+5设置 mAP	65.8	67.9	68.5	67.8	69.2	69.4	70.0	72.2	74.7	77.1
19+1设置 mAP	68.2	68.5	68.8	70.2	69.6	70.0	72.6	73.8	75.6	77.3

#### 4.4 消融实验

为了验证本文提出模型不同模块的有效性,本文在M-OWOD任务1上进行了消融实验,具体分析了3种影响因素.普通蒸馏仅使用传统的蒸馏方法,不包括任何因果干预;因果提示为引入因果干预中的提示学习;因果蒸馏为在因果提示的基础上,进一步通过因果干

预优化蒸馏过程.

实验结果如表5所示,相比于基准模型(Baseline),普通蒸馏(式(7)所示)带来了小幅度的性能提升(mAP从61.0%提升到61.1%,U-Recall从22.2%提升到22.3%).在加入因果提示后,mAP值提升至62.1%,U-Recall提升至23.2%,表明因果提示增强了模型对未知

类别的检测能力. 最终,在因果提示的基础上加入因果蒸馏机制(式(12)),mAP进一步提升至62.3%,U-Recall提升至25.1%,表明因果蒸馏机制对未知类别有更好的区分度.

表5 消融实验结果 单位:%

Baseline	普通蒸馏	因果提示	因果蒸馏	mAP	U-Recall
√	—	—	—	61.0	22.2
√	√	—	—	61.1	22.3
√	√	√	—	62.1	23.2
√	—	√	√	62.3	25.1

为了进一步验证超参数设置的有效性,本文在Task1任务上对蒸馏过程中引入的超参数 $\alpha$ 和 $\beta$ 进行了实验分析(见表6).实验结果表明,当 $\alpha=3$ 且 $\beta=6$ 时,模型在mAP和U-Recall上取得最优性能.在 $\alpha=3$ 且 $\beta=6$ 的基础上,本文在Task3进一步对蒸馏时温度 $T$ 的选择进行了消融实验(见表7).实验结果表明,当 $T=10$ 时,模型在mAP和U-Recall上能取得更好的效果.可适当增加 $\beta$ 的值以加强对未知类别蒸馏的权重;若更看重已知类别的精度,则可以适当提高 $\alpha$ 的值.温度参数 $T$ 可根据类别间相似性调整,差异较大时增大 $T$ 以增强小类贡献.

表6  $\alpha$ 和 $\beta$ 取值分析

	mAP/%	U-Recall/%
$\alpha=1, \beta=2$	22.5	61.6
$\alpha=2, \beta=4$	23.1	61.9
$\alpha=3, \beta=6$	25.1	62.3
$\alpha=4, \beta=8$	22.1	62.1

表7 温度 $T$ 取值分析

$T$	mAP/%	U-Recall/%
2	41.5	29.1
5	41.3	32.8
10	42.0	35.6
20	41.8	34.9

#### 4.5 可视化

为更直观地展示本文方法的性能优势,本文在M-OWOD任务上对比了本文方法与OrthoDet和RandBox的预测结果.图3中展示了部分典型场景的检测结果,黄色的检测框代表的是已知类别,绿色的检测框代表的是未知类别.其中,在第1列棒球的检测中,OrthoDet和RandBox方法未能正确识别目标,本文方法能够准确预测其类别为棒球.在第4列电话的检测中,OrthoDet和RandBox方法将其误判为背景,本文方法能够将其正确识别为未知类别.

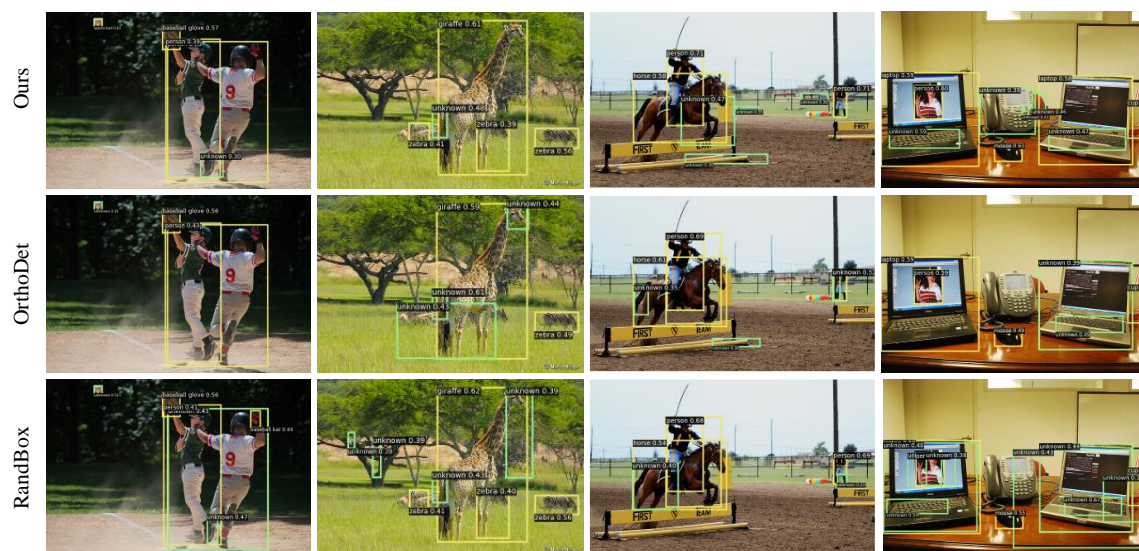


图3 不同方法在M-OWOD任务上的检测结果可视化

## 5 结论

本文提出了一种基于因果提示蒸馏的开放世界目标检测方法.针对开放场景中未知类别语义信息缺失导致的监督偏差问题,该方法从因果干预视角出发,设计了创新性解决方案.通过构建结构因果模型,本文分析了语义偏差对模型学习的影响,并提出了因果提示

和因果蒸馏两种核心机制.因果提示学习通过CLIP生成未知类别的语义向量,显式干预因果路径以增强模型对未知目标的语义感知能力;因果蒸馏策略通过分离教师模型对已知与未知类别的指导信息,设计双重KL散度损失函数,有效缓解知识传递中的偏差问题.实验结果表明,本文方法在多个开放世界目标检测基

准数据集上均取得了性能提升,在已知类别的平均检测精度和未知类别的召回率上均优于现有主流方法,验证了方法的有效性和鲁棒性.尽管如此,随着开放世界场景中未知类别数量的不断增加,如何有效处理复杂类别关系,以及进一步提升模型的计算效率,仍然是需要深入研究的重要方向.

#### 参考文献

- [1] VOULODIMOS A, DOULAMIS N, DOULAMIS A, et al. Deep learning for computer vision: A brief review[J]. *Computational Intelligence and Neuroscience*, 2018, 2018(1): 1-13.
- [2] ZHAO J Q, DING Z Y, ZHOU Y, et al. OrientedFormer: An end-to-end transformer-based oriented object detector in remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5640816.
- [3] YANG S, LU H M, LI J R. Multifeature fusion-based object detection for intelligent transportation systems[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(1): 1126-1133.
- [4] ZHAO J Q, WANG H Z, ZHOU Y, et al. Spatial-channel enhanced tr-ansformer for visible-infrared person re-identification[J]. *IEEE Transactions on Multimedia*, 2023, 25: 3668-3680.
- [5] MULLAPPILLY S S, GEHLOT A S, ANWER R M, et al. Semi-supervised open-world object detection[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, 38(5): 4305-4314.
- [6] WU Y, CHEN Y P, WANG L J, et al. Large scale incremental learning[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 374-382.
- [7] WANG X, CHEN G Y, QIAN G W, et al. Large-scale multi-modal pre-trained models: A comprehensive survey[J]. *Machine Intelligence Research*, 2023, 20(4): 447-482.
- [8] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//Proceedings of the International Conference on Machine Learning. New York: ICML, 2021: 8748-8763.
- [9] LI J M, ZHANG Y N, QIANG W W, et al. Disentangle and remerge: Interventional knowledge distillation for few-shot object detection from a conditional causal perspective[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, 37(1): 1323-1333.
- [10] ANTONELLI S, AVOLA D, CINQUE L, et al. Few-shot object detection: A survey[J]. *ACM Computing Surveys*, 2022, 54(11): 1-37.
- [11] GOU J P, YU B S, MAYBANK S J, et al. Knowledge distillation: A survey[J]. *International Journal of Computer Vision*, 2021, 129(6): 1789-1819.
- [12] ZHANG D, ZHANG H, TANG J, et al. Causal intervention for weakly-supervised semantic segmentation[J]. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020, 33: 655-666.
- [13] ZENG Y, ZHANG X S, LI H, et al. X<sup>2</sup>-VLM: All-in-one pre-trained model for vision-language tasks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(5): 3156-3168.
- [14] QIAN Q, HU J H. Online Zero-shot Classification with CLIP[M]//Computer Vision - ECCV 2024. Cham: Springer Nature Switzerland, 2024: 462-477.
- [15] WANG Y Q, YAO Q M, KWOK J T, et al. Generalizing from a few examples[J]. *ACM Computing Surveys*, 2021, 53(3): 1-34.
- [16] ZHAO J Q, FU A, ZHOU Y, et al. Fine-grained semantic oriented embedding set alignment for text-based person search[J]. *Image and Vision Computing*, 2024, 152: 105309.
- [17] SCHEIRER W J, DE REZENDE ROCHA A, SAPKOTA A, et al. Toward open set recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(7): 1757-1772.
- [18] FENG G C, DESAI D, PASQUALI S, et al. Open set recognition for random forest[C]//Proceedings of the 5th ACM International Conference on AI in Finance. New York: ACM, 2024: 45-53.
- [19] YANG Z, YUE J, GHAMISI P, et al. Open set recognition in real world[J]. *International Journal of Computer Vision*, 2024, 132(8): 3208-3231.
- [20] JOSEPH K J, KHAN S, KHAN F S, et al. Towards open world object detection[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 5826-5836.
- [21] MA Y Q, LI H N, ZHANG Z G, et al. Annealing-based label-transfer learning for open world object detection[C]//

- 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 11454-11463.
- [22] WANG Y H, YUE Z Q, HUA X S, et al. Random boxes are open-world object detectors[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2023: 6210-6220.
- [23] SUN Z C, LI J H, MU Y D. Exploring orthogonality in open world object detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2024: 17302-17312.
- [24] JUDEA P. Causality: Models, Reasoning and Inference[M]. New York: Cambridge University Press, 2000.
- [25] 邵志文, 陈必宽, 祝汉城, 等. 基于因果干预的无偏面部动作单元识别[J]. 电子学报, 2024, 52(10): 3312-3321.
- SHAO Z W, CHEN B K, ZHU H C, et al. Causal intervention for unbiased facial action unit recognition[J]. Acta Electronica Sinica, 2024, 52(10): 3312-3321. (in Chinese)
- [26] KHATTAK M U, RASHEED H, MAAZ M, et al. MaPLE: Multi-modal prompt learning[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 19113-19122.
- [27] NADKARNI P M, OHNO-MACHADO L, CHAPMAN W W. Natural language processing: An introduction[J]. Journal of the American Medical Informatics Association, 2011, 18(5): 544-551.
- [28] 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.
- [29] 王楠井, 刘阿建, 梁凤梅, 等. 基于图像内容理解的判别性类别提示学习[J]. 电子学报, 2025, 53(2): 493-502.
- WANG N J, LIU A J, LIANG F M, et al. Discriminative category prompt learning based on image content understanding[J]. Acta Electronica Sinica, 2025, 53(2): 493-502. (in Chinese)
- [30] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[EB/OL]. (2015-05-09)[2025-05-27]. <https://arxiv.org/abs/1503.02531v1>.
- [31] YANG Z D, LI Z, JIANG X H, et al. Focal and global knowledge distillation for detectors[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 4633-4642.
- [32] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft Coco: Common Objects in Context[M]. Berlin: Springer International Publishing, 2014.
- [33] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [34] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization[EB/OL]. (2019-01-04)[2025-05-27]. <https://arxiv.org/abs/1711.05101v3>.
- [35] YANG S, SUN P, JIANG Y, et al. Objects in semantic topology[C]//International Conference on Learning Representation. Washington DC: ICLR, 2022: 1.
- [36] GUPTA A, NARAYAN S, JOSEPH K J, et al. OW-DETR: Open-world detection transformer[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 9225-9234.
- [37] WU Z H, LU Y, CHEN X Y, et al. UC-OWOD: Unknown-classified open world object detection[C]//Computer Vision - ECCV 2022. Cham: Springer Nature Switzerland, 2022: 193-210.
- [38] ZOHAR O, WANG K C, YEUNG S. PROB: Probabilistic objectness for open world object detection[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 11444-11453.
- [39] MA S L, WANG Y F, WEI Y, et al. CAT: LoCalization and IdentificAtion cascade detection transformer for open-world object detection[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 19681-19690.
- [40] SHMELKOV K, SCHMID C, ALAHARI K. Incremental learning of object detectors without catastrophic forgetting[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 3420-3429.
- [41] PENG C, ZHAO K, LOVELL B C. Faster ILOD: Incremental learning for object detectors based on faster RCNN[J]. Pattern Recognition Letters, 2020, 140: 109-115.
- [42] KANG M X, ZHANG J P, ZHANG J M, et al. Alleviating catastrophic forgetting of incremental object detection via within-class and between-class knowledge distillation[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2023: 18848-18858.

## 作者简介



**赵佳琦** 男,1988年5月出生于江苏省徐州市.中国矿业大学副教授、硕士生导师.主要研究方向为人工智能、计算机视觉、无人驾驶等.

E-mail: jiaqizhao@cumt.edu.cn



**杜文亮** 男,1989年2月生.中国矿业大学计算机科学与技术学院讲师.主要研究方向为遥感图像处理、人工智能、机器学习、模式识别、医学图像处理.

E-mail: wldu@cumt.edu.cn



**王平安** 男,2000年9月出生于河南省焦作市.中国矿业大学计算机科学与技术学院硕士研究生.主要研究方向为计算机视觉、目标检测.

E-mail: pinganwang@cumt.edu.cn



**姚睿** 男,1982年7月生.中国矿业大学计算机科学与技术学院教授、博士生导师.主要研究方向为计算机视觉、模式识别、深度学习、人工智能等.

E-mail: ruiyao@cumt.edu.cn



**周勇** 男,1974年9月生.中国矿业大学计算机科学与技术学院教授,博士生导师.主要研究方向为机器学习、人工智能、数据科学与工程等.

E-mail: yzhou@cumt.edu.cn



**刘兵** 男,1981年8月生.中国矿业大学计算机科学与技术学院副教授.主要研究方向为人工智能、图像处理和矿山智能化.

E-mail: liubing@cumt.edu.cn