摘要

点云数据通过捕捉环境中的三维结构信息表示现实场景，在现实生活中被广泛应用。在自动驾驶、机器人导航和抓取等领域，准确检测识别场景中的各个目标是理解点云场景的重要环节。然而，传统的点云检测识别模型基于人工标注的点云场景数据集进行有监督训练，并且依赖固定通道的分类器识别具体目标类别，在面临类别偏移现象，即应用场景具有新类别时的泛化能力不足，在实际部署过程中性能会降低。

开放词汇训练方法借助视觉语言模型的零样本学习能力增强模型对类别偏移的适应性，通过将视觉数据与具有相同语义的文本描述相关联，扩展模型对新类别的识别能力。然而具有通用能力的大规模预训练模型大多基于图像和文本数据训练，可处理的数据与点云数据之间存在模态差异，增加了应用开放词汇训练方法的复杂性。现有的方法训练时使用的多模态数据内容和形式较为单一，缺乏细节和多样性，限制了视觉语言模型的零样本能力向点云空间的有效迁移，从而影响了点云模型的整体性能。尽管人工标注的点云场景数据集能够提升模型的准确性，但其泛化性和方法的可用性受限于标注的规模。为了缓解上述问题，本文从两方面着手提出了无监督的训练方法：1）整合多种大规模预训练的基础模型，增强训练数据，提高视觉数据和文本描述的关联的准确性；2）改进迁移学习策略，以通用图像目标检测分割模型的输出特征作为中介，融合多任务学习和辅助网络提升点云模型输出的特征表示的质量。本文的主要贡献如下：

本文首先提出一种多基础模型增强的无监督点云目标检测训练方法。本方法采用了通用分割模型在点云场景样本对应的真实图像上分割出目标区域，以跨模态地监督点云模型在点云场景样本中生成类别无关的目标级区域提议。进一步地，本方法提示生成式预训练的基础模型生成多样化、高质量的图像和文本，在图像数据和文本描述之间建立准确的语义关联，经过进一步训练增强点云数据和文本描述的语义关联。本方法所训练的模型在点云场景数据集的目标检测基准上的性能提升验证了其有效性。

为了准确检测目标的位置和边界，本文提出了一种基于多任务学习的点云实例分割方法，改进了视觉语言模型和点云模型之间的迁移学习过程。该方法引入图像上的通用对象级模型提升点云模型的检测识别能力。其中，目标检测和实例分割的联合训练利用了两种任务之间的相似性和互补性，提升点云模型检测目标准确边界的能力。由于上述图像模型是将二维分割模型的输出特征和文本描述的编码结果映射到共享的语义空间，该方法通过点云数据和图像数据之间的跨模态对齐，在点云数据和文本描述之间建立语义关联。该过程以通用图像模型作为中介，优化了点云模型的特征表示。

该方法还引入了点云模型的时间平均模型以稳定训练。在常用的点云场景数据集的实例分割基准上的性能提升和新类别上的可视化结果验证了本方法的有效性。

关键词：开放词汇，点云目标检测，实例分割，无监督，基础模型