

# 语言指令驱动的对象点云三维可供性分割

杜佳璇<sup>1</sup>, 吴皓<sup>1</sup>, 马庆<sup>1</sup>, 田国会<sup>1</sup>, 赵志贤<sup>1</sup>, 冷述文<sup>2</sup>

<sup>1</sup>山东大学控制科学与工程学院, 中国济南市, 250061

<sup>2</sup>中国华能集团有限公司山东分公司, 中国济南市, 250014

**摘要:** 物体的抓取位置与任务类型密切相关。对于同一种物品, 不同用户需求可能对应不同的抓取方式。视觉可供性为操作行为提供了可靠的先验知识。现有方法通常从图像或视频中学习可供性, 但基于平面的可供性缺乏实现六自由度操作所需的空问信息。此外, 当前方法局限于预定义类别相关的可供性, 无法直接从用户指令中推断可供性。为解决上述问题, 提出了一项新任务: 语言指令驱动的三维物体可供性分割。为支持该研究, 构建了一个指令—可供性数据集。该数据集具有挑战性, 包含20类常见物体类别中的7190个物体实例, 并配有624条操作指令, 这些指令明确了相应的可供性。为评估模型对新指令的泛化能力, 数据集包括“已见”和“未见”两种设置。在此基础上, 设计了指令驱动的三维可供性分割网络, 该网络从点云中提取特征, 并逐层融合指令特征。依据给定用户指令, 模型能够在物体点云上直接分割出建议的操作区域, 从而指导最优抓取位姿的选择。实验结果表明, 该方法在“已见”和“未见”设置下均优于其他相关方法, 并展现出对多样化用户指令和未知可供性的泛化能力。

**关键词:** 视觉可供性; 点云分割; 开放语义; 多模态融合; 服务机器人  
<https://doi.org/10.1631/ENG.ITEE.2026.0044>