摘要

点云是点的集合，刻画了物理世界的三维结构，对智能体的环境感知至关重要。点云语义分割为点云中的每一个点赋予类别，使之具备语义信息。分割后的点云在智能体的决策、规划和控制任务中发挥了重要作用。近年来，随着深度学习技术的快速发展，涌现出了一批数据驱动的三维点云语义分割方法。这些方法强烈依赖大规模有标记的语义分割数据，并要求训练与测试样本满足独立同分布假设。然而，现实场景无法满足上述要求。一方面，三维点云标注的复杂繁琐使得数据标注资源十分受限；另一方面，模型通常部署于开放场景中，场景之间数据分布存在显著差异，使得点云语义分割方法在开放场景中的应用面临严峻挑战。因此，本文聚焦于低标签条件下开放场景中三维点云语义分割方法的研究。

针对上述问题，本文首先在独立同分布假设条件下，构建了胜任源域数据的点云语义分割模型，而后拓展该模型至开放场景中。对于源域语义分割模型的构建，充分利用了大规模无标记数据及小规模数据标注资源，探究了三维点云语义分割任务中的低标签学习方法。在此基础上，对于模型向开放场景的拓展，依次引入领域偏移及类别偏移，探究了三维点云语义分割任务中的领域自适应及开集识别方法。所提出的方法以电力巡检为应用场景，亦可以拓展至其它应用场景中。本文的创新贡献如下：

第一，提出了一种基于多模态掩码自编码的自监督预训练方法，通过刻画无标记数据的分布规律及多模态数据间的配对关系，提取三维点云的通用特征，为语义分割模型的有监督优化提供良好的初值。该方法依据点云与图像间的配对关系，以互补掩蔽的方式，掩蔽点云与图像中互补的区域，并以重建原始输入为目标训练模型的骨干网络。本文首次构建了电力巡检场景点云语义分割数据集PowerLineSeg，该数据集上自注意力层的激活可视化表明了配对图像块与点云簇之间激活响应强烈。并且相比于基线方法，该方法语义分割的平均交并比大幅提高了2.02%。

第二，提出了一种基于场景剖分主动学习的点云语义分割方法，通过评估无标记样本的期望标注价值，自适应挑选并标注最具代表性的关键样本，以样本挑选的方式最优化地分配有限的数据标记资源。该方法同时在时域与空域细化了关键样本的挑选粒度。在空间维度，将场景级点云动态剖分为区域级，并在区域级挑选关键样本。在时间维度，将有限的数据标记资源摊派至数轮迭代中，随着迭代的进行，场景剖分及价值评估逐步变得更加可靠。在PowerLineSeg数据集上的实验结果显示，在平均交并比相当的情况下，该方法节省了85%的数据标记资源。

第三，提出了一种预训练视觉模型引导的跨领域点云语义分割方法，通过引入预

训练视觉模型作为外部知识源，增强领域之间的关联性并缩小领域差异，缓解了开放场景中的领域偏移问题。该方法以预训练视觉模型的预测为引导，生成目标域场景伪标签，通过跨领域表征对齐，同时从全局及逐类别表征维度缩小领域差异。在此基础上，进一步通过跨模态表征对齐，迁移预训练视觉模型中的知识至点云语义分割模型中。在PowerLineSeg数据集上的实验结果显示，相比于直接迁移源域模型至目标域，该方法在目标域场景点云语义分割任务上的平均交并比大幅提高了16.08%。

第四，提出了一种基于多模态对比关联的开集点云语义分割方法，通过建模开集类别表征空间及多模态语义关联，赋予模型识别开集目标的能力，缓解了开放场景中类别偏移的问题。该方法用对比学习损失替换了用于闭集分类的交叉熵损失，通过度量学习建模开集类别表征空间。在此基础上，利用“点云与图像”、“图像与文本”之间的配对关系，以图像为中间媒介，通过多模态去偏对比学习桥接文本与点云的表征空间。在PowerLineSeg数据集上的实验结果显示，在无需新增三维标注的情况下，开集类别语义分割的平均交并比大幅提高12.01%。

综上所述，本文通过自监督预训练及关键样本有监督微调，构建了胜任源域数据分布的点云语义分割模型。在此基础上，通过预训练视觉模型引导的领域自适应及多模态对比关联开集识别，进一步拓展模型至面临领域偏移及类别偏移的开放场景中。所提出的方法在全国多个地市的电力巡检业务中试点应用，大幅提高了电力巡检业务中点云语义分割的数据处理效率，为点云语义分割方法向开放场景电力巡检的拓展应用奠定了基础。

关键词：点云语义分割，自监督预训练，主动学习，领域自适应，开集识别

Point Cloud Semantic Segmentation in the Open-World

Yuheng Lu (Computer Application Technology)
Directed by Prof. Xiaodong Xie

ABSTRACT

Point clouds represent 3D structures that are essential for an intelligent agent to perceive its surroundings. Semantic segmentation assigns categories to points, providing semantic context. Segmented point clouds are essential for agent decision making, planning, and control. Recent advances in deep learning have spawned a variety of 3D point cloud semantic segmentation methods that rely on extensive annotated data following the I.I.D. assumption. However, Real-world scenarios often fail to meet these requirements due to labor-intensive annotation and variable data distribution. This thesis focuses on semantic segmentation of open-world point clouds under low-label conditions.

This thesis addresses the aforementioned challenges by first building a source domain model following the I.I.D assumption and then extending this model to the open-world. To construct source-domain models, we explore low-label learning methods that leverage large-scale unlabeled data together with limited data annotations. Building on this foundation, we sequentially introduce domain shift and category shift to transfer source-domain models to the open-world via exploratory domain adaptation and open-set recognition methods. The proposed method, originally devised for power inspection, is applicable to a wider range of applications. The innovative contributions encompass:

This thesis proposes a self-supervised learning method based on multi-modal masked autoencoders, which extract general point cloud features by characterizing unlabeled data patterns and point cloud-image pairing relationships. Complementary masking is applied to both the point cloud and the image, and the backbone network is trained to reconstruct the original input. Additionally, we create the PowerLineSeg dataset for power inspection point cloud semantic segmentation. Visualization of the self-attention layer shows strong activation between paired images and point clouds. Compared to the baseline method, our approach achieves a 2.02% improvement in average intersection and union (mIoU).

This thesis proposes a scene partition-based active learning approach for point cloud

semantic segmentation. It adaptively selects and labels key samples by evaluating expected label values of unlabeled data, improving sample selection in both temporal and spatial domain. Spatially, it divides scene-level point clouds into regions, while temporally, it iteratively allocates limited labeling resources, enhancing scene partitioning and value evaluation over time. Results on PowerLineSeg show that the proposed method reduces 85% labeled data with similar mIoU.

This thesis proposes a domain adaptation method for point cloud semantic segmentation, guided by a pre-trained visual model. Leveraging the visual model as external knowledge, it enhances domain correlation to address domain shift challenges. The method utilizes visual model predictions to generate pseudo-labels for the target domain and mitigates domain gap through cross-domain representation alignment. Additionally, it transfers external knowledge to the semantic segmentation model via cross-modal representation alignment. Results on PowerLineSeg demonstrate that the proposed method yields a significant 16.08% improvement in mIoU compared to direct transfer.

This thesis proposes an open-set point cloud semantic segmentation method based on multi-modal contrastive learning. It first models the open-set representation and then correlates the multi-modal representation, enabling the model to identify open-set categories, addressing the category shift challenges. Specifically, this method replaces the cross-entropy loss with a contrastive loss, aiding in the modeling of open-set representation through metric learning. Additionally, it leverages the pairing relationship between ”point cloud and image” and ”image and text,” using images as intermediaries to connect text and point cloud representations through de-biased multi-modal contrastive learning. Results on PowerLineSeg show a notable 12.01% improvement in mIoU for open-set semantic segmentation, all without the need for additional annotations.

In summary, this thesis initially constructs the source domain model through self-supervised pretraining and supervised fine-tuning on key samples. Subsequently, this model is adapted to the open-world via domain adaptation and open-set semantic segmentation. The proposed method has been successfully piloted and implemented in power inspection around China. Moreover, it significantly enhances the efficiency of point cloud semantic segmentation for power inspection, establishing a solid foundation for its application in open-world power inspection scenarios.

KEYWORDS: Point Cloud Semantic Segmentation, Self-supervised Pre-training, Active Learning, Domain Adaptation, Open-Set Recognition