摘要

三维重建是计算机视觉和计算机图形学的重要研究方向，在虚拟现实、增强现实、自动驾驶、工业检测等多个领域具有广泛应用。本文聚焦于多视角单物体三维体素重建任务，此任务旨在从多个视角的二维图像重建物体的三维体素模型。现有方法主要依赖深度学习对输入图像进行特征提取并进行融合，尽管取得了不错的结果，但仍然面临以下挑战：(1) 现有方法主要利用图像特征进行三维重建，对视角信息的利用较为有限，未能充分利用图像特征中隐含的视角信息来辅助多视角的融合和重建；(2) 由于输入的多视角图像通常无法覆盖物体的完整形态，导致对不可见区域的形状推理存在较大不确定性，尤其是在视角数量较少时，给定的语义信息较少，重建的质量下降明显；(3) 现有方法大多仅基于视觉信息，未充分利用多模态信息，如文本描述，从而限制了三维重建的性能和泛化性。

针对上述问题，本文提出两种基于先验知识的多视角单物体三维体素重建方法，以提升重建精度和模型的泛化能力。

首先，本文提出了基于视角信息和形状先验的多视角单物体三维体素重建方法。该方法利用数据集中已有的视角参数信息进行监督，显式建模输入图像的视角信息，并利用视角信息指导特征融合，增强不同视角特征的一致性。同时，为了提升模型在输入视角较少时的性能，本文引入了形状先验，建立了一个形状先验字典，通过从字典中检索形状相似的三维模型，提供额外的结构信息，缓解输入视角稀疏导致的信息缺失问题。实验表明，该方法基于视角信息和形状先验能够有效提升各个视角数量下的重建精度，在输入视角较多时，视角信息能够提供非常大的帮助，而在输入视角较少时，形状先验能够提供更多的先验信息帮助网络重建三维模型。

此外，本文还探索了基于多模态先验信息的多视角单物体三维体素重建方法。通过利用多模态大模型作为先验的模型，为三维重建模型提供更多的语义信息。具体来讲，本文利用大规模预训练的视觉-语言模型从输入图像生成文本描述，并提取文本特征，与视觉特征进行跨模态融合，以提供额外的语义信息辅助重建。文本描述能够提供关于物体类别、结构的高层次语义信息，进一步提升模型对复杂形状的建模能力。实验结果表明，多模态信息的引入能够有效改善重建效果，尤其当文本质量越高，重建效果越好。

综上所述，本文提出的两种基于先验知识的多视角单物体三维体素重建方法能够有效的弥补图像信息的不足，从而提升三维重建的性能，为以后的研究提供新的思路。

关键词：三维重建，形状先验，视角信息，多模态学习