E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2025)04-0922-31

论文引用格式:Liu C, Cao T, Kang W X, Jiang Z H, Yang C H, Gui W H and Liang X J. 2025. Single-view 3D object reconstruction based on deep learning: survey. Journal of Image and Graphics, 30(4):0922-0952(刘草, 曹婷, 康文雄, 蒋朝辉, 阳春华, 桂卫华, 梁骁俊. 2025. 深度学习邻域单视图三维物体重建研究综述. 中国图象图形学报, 30(4):0922-0952)[DOI:10.11834/jig.240389]

深度学习领域单视图三维物体重建研究综述

刘草[1],曹婷[2*],康文雄[1],蒋朝辉[3],阳春华[3],桂卫华[3],梁骁俊[2]

1. 华南理工大学未来技术学院,广州　510006; 2. 鹏城实验室,深圳　518000; 3. 中南大学自动化学院,长沙　410083

摘　要:从单个视图恢复物体三维结构信息是计算机视觉的重要研究方向,在工业生产、医疗诊断和虚拟现实等领域发挥重要作用。传统单视图三维物体重建方法需要结合几何模板和几何假设以完成特定场景对象的三维重建任务。而当前基于深度学习的单视图三维物体重建方法通过数据驱动的方式,在重建对象适用范围和重建模型鲁棒性等方面取得进展。本文首先讨论近年来单视图三维物体重建领域常用的数据集与评价指标。然后围绕基于深度学习的单视图三维物体重建领域,对有监督学习单视图三维物体重建、无监督学习单视图三维物体重建和半监督学习单视图三维物体重建等相关研究工作进行系统性地分析和总结。最后,对基于深度学习的单视图三维物体重建方法未解决难题进行总结,并展望未来可能的发展趋势与关键技术。

关键词:深度学习;三维物体重建;单视图;有监督学习;无监督学习;半监督学习

Single-view 3D object reconstruction based on deep learning: survey

Liu Cao[1], Cao Ting[2*], Kang Wenxiong[1], Jiang Zhaohui[3], Yang Chunhua[3], Gui Weihua[3], Liang Xiaojun[2]

1. School of Future Technology, South China University of Technology, Guangzhou　510006, China;
2. Peng Cheng Laboratory, Shenzhen　518000, China; 3. School of Automation, Central South University, Changsha　410083, China

Abstract: Single-view 3D object reconstruction seeks to leverage the 2D structure of a single-view image to reconstruct the 3D shape of an object, facilitating subsequent tasks such as 3D object detection, 3D object recognition, and 3D semantic segmentation. Single-view 3D object reconstruction has emerged as a pivotal topic in computer vision, with wide-ranging applications in industrial production, medical diagnostics, virtual reality, and other fields. Traditional single-view 3D object reconstruction methods rely on a combination of geometric templates and geometric assumptions to complete the 3D reconstruction of specific scene objects. However, the traditional methods based on geometric templates are tailored to specific objects, limiting their generality and scalability. Meanwhile, the traditional methods based on geometric assumptions require strong prior conditions for the object, which limits the reconstruction quality of different changing scenes. Current single-view 3D object reconstruction methods based on deep learning have made significant progress in terms of the applicability of reconstructed objects and the robustness of reconstructed models through data-driven approaches. To further understand their current development, this work systematically analyzes and summarizes single-view 3D object reconstruction

收稿日期:2024-07-11;**修回日期**:2024-10-03;**预印本日期**:2024-10-10

*通信作者:曹婷　caot@pcl.ac.cn

基金项目:鹏城实验室重大攻关项目;国家自然科学基金项目(62103206,62376100);中国博士后科学基金资助项目(2021M701804)

Supported by: Major Key Project of Peng Cheng Laboratory; National Natural Science Foundation of China(62103206, 62376100); China Postdoctoral Science Foundation(2021M701804)

methods based on deep learning from following aspects: commonly used datasets and evaluation indicators, method classification and improvement innovation, problem challenges, and development trends in single-view 3D object reconstruction. First, this work focuses on the commonly used datasets and evaluation indicators in single-view 3D object reconstruction. These datasets are the foundation of 3D reconstruction methods based on deep learning and can be divided into three categories: red-green-blue-depth (RGBD) datasets, which contain object depth information and are commonly used for algorithm testing and evaluation; synthetic datasets, which contain large-scale rendering object images and 3D shape data and are commonly used for algorithm training and evaluation; and real-scene datasets, which contain a limited number of real object images and 3D shape and are commonly used for algorithm evaluation. Evaluation indicators, mainly including distance evaluation indicators and classification evaluation indicators, can quantitatively demonstrate algorithm performance. Distance evaluation indicators are mainly used to evaluate the shape distance between the reconstructed 3D model and the ground truth 3D model. The smaller the value, the closer the overall shape of the reconstructed 3D model is to the ground truth 3D model. Classification evaluation indicators are mainly used to evaluate the accuracy of the 3D shape classification of each point in the 3D space. The larger the value, the more accurate the reconstructed 3D model. Second, this work analyzes the field of single-view 3D object reconstruction based on deep learning and systematically summarizes the research works related to supervised, unsupervised, and semi-supervised learning single-view 3D object reconstructions. Supervised learning single view 3D object reconstruction methods mainly focus on reconstruction resolution in the early stage. With the improvement of 3D representations, especially the application of implicit 3D representation, the high-resolution reconstruction of object details has become possible. Subsequent works have improved and innovated various aspects, such as input image, encoding and decoding, prior knowledge, and general structure, to further solve reconstruction problems, such as unknown perspectives, key details, and generalized shapes. Unsupervised learning single view 3D object reconstruction methods mainly focus on improving the rendering process in the early stage, laying the foundation for unsupervised learning. Subsequent works have improved and innovated from the perspectives of rendering image quantity, image feature attributes, and additional prior knowledge to further solve various problems, such as lighting and background interference. Semi-supervised learning single view 3D object reconstruction methods are mainly divided into 2D and 3D labeled data-based methods. The former proposes a small-sample data training paradigm and a general data training paradigm to overcome challenges such as difficulty in 3D annotation, deviation, and inconsistency between annotation data and test data. The latter enhances the robust generalization performance of reconstruction through semantic and perspective information. The above three learning methods have their own advantages and disadvantages in terms of technical frameworks. Supervised learning methods utilize 3D labeled data for learning and reconstruction, resulting in high reconstruction quality; however, they are limited by the high cost of data annotation. Unsupervised learning methods can directly use 2D images to learn and reconstruct, effectively reducing training costs; however, the reconstruction quality is not stable enough. Semi-supervised learning methods propose a paradigm for joint learning of labeled and unlabeled data to address the problems of high data annotation cost and unstable reconstruction quality, combining the advantages of the two methods mentioned above. These methods have been widely studied. This work summarizes the unresolved challenges from the perspectives of data, training paradigms, evaluation metrics, and reconstruction performance and proposes future development trends and key technologies of single-view 3D object reconstruction methods based on deep learning. For the difficulty in collecting data from wild objects, studies must focus on how to use internet object image data to build datasets and develop efficient interactive annotation tools to reduce data collection costs and annotation costs. For the insufficient learning of local object structures, studies must focus on the training paradigm guided by prior knowledge of local object structures to enhance the accuracy and reliability of single-view 3D object reconstruction. For the limited reconstruction performance of few-shot 3D annotated data, studies must design the optimal combination of different tasks and develop multitask learning methods to obtain effective semantic information of objects and supplement effective object reconstruction supervision information. For the neglected local structure assessment of existing evaluation indicators, studies must design reconstruction evaluation indicators that focus on the reconstruction results of local structures to further guide high-precision reconstruction optimization. For the problems of long training optimization cycles and limited object categories faced by existing methods, studies must focus on 3D foundation models that achieve universal category object shape reconstruction

to promote the development and application of single-view 3D object reconstruction methods.

Key words: deep learning; 3D object reconstruction; single-view; supervised learning; unsupervised learning; semi-supervised learning

0 引 言

三维物体重建是计算机视觉领域的重要研究方向,旨在将物体的表面形态、结构以及纹理等信息快速、准确地还原为三维模型,从而实现对物体的多维度理解与展示,在工业生产(何瑞清 等,2022)、医疗诊断(Alsadoon 等,2024)和虚拟现实(卢经纬 等,2023)等领域得到了广泛研究与应用。其中单视图三维物体重建是三维重建的重要分支之一,能够仅利用单个视角图像实现物体的三维形状恢复。与激光雷达(Park 等,2020)、结构光相机(Wang 等,2022)和多视图三维物体重建(周晓清 等,2023)等方法相比,单视图三维物体重建方法面临单个视图信息有限的挑战,但具有设备简单和数据易于处理等方面优势。传统单视图三维物体重建方法包括利用几何模板和几何假设进行三维模型重建(Kim 等,2016;Liu 等,2015)。常用的几何模板包括圆柱体和物体形状等,基于几何模板的方法往往只适用于具体形状的对象,限制了方法的通用性和可扩展性。常见的几何假设包括对称性假设、拓扑约束、体积约束、形状约束和光源假设等。基于几何假设的方法需要针对重建对象的较强先验条件,限制了方法应用到实际场景的重建质量。

基于深度学习的单视图三维物体重建方法发展迅速,可以通过大规模数据驱动的方式自适应不同复杂变化,克服传统方法受限于几何模板和几何假设等问题,实现鲁棒单视图三维物体重建(Bednarik 等,2018;Pumarola 等,2018)。本文针对基于深度学习的单视图三维物体重建的关键方法和最新前沿工作进行系统性的分析和总结。当前单视图三维物体重建领域的相关综述文献(Fahim 等,2021;Samavati 和 Soryani,2023;杨航 等,2023),主要根据三维模型的表现形式:体素(Wu 等,2015)、点云(Charles 等,2017)、网格(Wang 等,2018a)、占有函数(Peng 等,2020)、符号距离函数(Chabra 等,2020)和神经辐射场(Mildenhall 等,2020),对相关单视图三维物体重建方法进行分类,总结不同三维模型的表现形式在单视图三维物体重建领域的研究方向与难题。而本文聚焦于单视图三维物体重建方法的监督学习方式,并按照相关工作关注的问题与研究优化方向进一步细分与详细总结;此外,本文对每类方法的优缺点与定性、定量对比结果进行分析,总结提炼面临的研究难题和可能的改进方向,以启发读者在现有基础上,做出进一步的突破。

1 概 述

按监督学习方式的不同,基于深度学习的单视图三维物体重建方法可分为有监督、无监督和半监督 3 大类;根据研究工作关注的问题,相关方法可更进一步细分成 15 小类。根据上述分类,本文调研了近年来单视图三维物体重建相关文献,涵盖了该领域的早期工作、研究创新过程中的重要工作,以及直到 2024 年的近期前沿工作;并从中选取国内外权威期刊与会议的相关论文进行详细的分析归纳,以确保其时效性和质量。

为了直观展示相关发展历程,本文对单视图三维物体重建方法进行归纳,如表 1 所示。表中对每类方法的关键论文、来源、年份、细分类、技术突破、关注问题和进展相关的研究工作进行了总结。

有监督单视图三维物体重建方法前期主要关注重建分辨率问题,随着模型表示形式的改进,特别是隐式表示形式的应用,物体细节高分辨率重建成为可能;后续一些工作对输入图像、编解码、先验知识和通用结构等各个环节进行改进创新,以进一步解决未知视角、关键细节和泛化形状等重建难题。

无监督单视图三维物体重建方法前期主要关注渲染过程改进问题,为无监督学习奠定基础。后续工作从渲染学习图像数量、渲染特征属性和额外先验知识等角度出发进行了改进创新,以进一步解决单视图缺乏形状约束、图像背景和光照干扰等难题。

半监督单视图三维物体重建方法主要分为二维标签数据和三维标签数据等发展方向。利用三维标签数据的相关工作提出小样本三维标注数据训练与通用数据预训练的训练范式,以克服三维标注难、标

注数据与测试数据偏差、不一致等难题；利用二维标签数据的相关工作通过语义信息和视角信息，进一步提升了重建的鲁棒泛化性能。

根据表 1 归纳的单视图三维物体重建相关方法，本文总结提炼了 3 种监督学习单视图三维物体重建方法的技术框架。

表 1　现有单视图三维物体重建方法归纳

Table 1　Overview of existing single view 3D object reconstruction methods

类型	关键论文	来源	年份	相关研究工作
有监督	Tatarchenko 等人（2017）	ICCV	2017	Riegler 等人（2017）、Häne 等人（2017）、Wang 等人（2018b）、Victoria 等人（2023）
	Mescheder 等人（2019）	CVPR	2019	Park 等人（2019）、Chen 等人（2019）、Koneputugodage 等人（2023）
	Zhang 等人（2018）	NIPS	2018	Wu 等人（2018）、王年等人（2020）、张豪等人（2021）、Zhang 等人（2024）
	Li 和 Kuang（2021）	ICCST	2021	朱育正等人（2021）、白静等人（2022）、Jin 等人（2022）、李雷等人（2022）、吴繁和贺赛先（2023）、万骏辉等人（2024）
	Li 和 Zhang（2021）	CVPR	2021	王年等人（2020）、Hsu 等人（2020）、Zhao 等人（2021）、梁春阳等人（2023）、Wimbauer 等人（2023）
	Yang 等人（2021b）	CVPR	2021	Pinheiro 等人（2019）、Michalkiewicz 等人（2020）、Nie 等人（2023a）
	Gan 等人（2023）	Expert Systems with Applications	2023	Wang 和 Fang（2020）、Zhou 等人（2021）、Kuo 等人（2022）
无监督	Liu 等人（2019）	ICCV	2019	Niemeyer 等人（2020）、Zubić 和 Liò（2021）、Wang 等人（2023a）、Liu 等人（2024）
	Ho 等人（2021）	ICCV	2021	Yao 等人（2020）、Duggal 和 Pathak（2022）、Chen 等人（2024）
	Rebain 等人（2022）	CVPR	2022	Kato 和 Harada（2019）、Zhou 等人（2020a）、Kim 和 Chun（2023）
	Monnier 等人（2022）	ECCV	2022	Chen 等人（2020b）、Hu 等人（2021）、Huang 等人（2023）、Dundar 等人（2023）、Aygün 和 Mac Aodha（2024）
	Wu 等人（2023b）	IEEE Transactions on Pattern Analysis and Machine Intelligence	2023	Yao 等人（2021）、Mei 等人（2022）、Nie 等人（2023b）、刘晓楠等人（2024）
半监督	Piao 等人（2019）	ICCV	2019	Gao 等人（2020）、Xing 等人（2022）、Roessle 等人（2022）、Chung 等人（2024）
	Kohli 等人（2020）	3DV	2020	Yue 等人（2020）、Xu 等人（2021a，2021b）
	Laradji 等人（2021）	ICCVW	2021	Yang 等人（2018）、Shen 等人（2023）
	Yang 等人（2021a）	IROS	2021	Li 和 Wu（2021）、Melas-Kyriazi 等人（2023）
	Lai 等人（2024）	IEEE Transactions on Multimedia	2024	Elich 等人（2022）、Alwala 等人（2022）、Yang 等人（2023）、Nam 等人（2023）

注：ICCV：IEEE / CVF International Conference on Computer Vision；CVPR：IEEE / CVF Computer Vision and Pattern Recognition Conference；NIPS：International Conference on Neural Information Processing Systems；ICCST：International Conference on Culture-oriented Science and Technology；ECCV：European Conference on Computer Vision；3DV：International Conference on 3D Vision；ICCVW：IEEE/CVF International Conference on Computer Vision Workshops；IROS：IEEE/RSJ International Conference on Intelligent Robots and Systems。

续表1　现有单视图三维物体重建方法归纳

Table 1　Overview of existing single view 3D object reconstruction methods

类型	关键论文	细分类	技术突破	关注问题
有监督	Tatarchenko等人(2017)	基于模型表示形式改进	提出非平衡八叉树结构重建架构	局部复杂结构更高分辨率重建
	Mescheder等人(2019)		提出隐式表示形式重建架构	物体任意分辨率重建
	Zhang等人(2018)	基于图像数据信息增强	预测完整视角深度信息,与图像信息融合重建	物体未知视角的自适应重建
	Li和Kuang(2021)	基于注意力机制	利用自注意力Transformer重建	物体关键细节区域重建
	Li和Zhang(2021)	基于局部特征优化	构建物体与图像投影关系,提取图像局部特征优化重建	物体视角可见区域精准重建
	Yang等人(2021b)	基于先验知识融合	通过记忆网络检索相似物体形状信息,与图像特征融合重建	物体整体形状鲁棒重建
	Gan等人(2023)	基于通用三维结构学习	图像信息指导几何基元细粒度优化	物体未知类别泛化重建
无监督	Liu等人(2019)	基于渲染过程改进	提出可微渲染重建架构	渲染前向与后向过程一致性
	Ho等人(2021)	基于多视图增强学习	翻转视图交叉渲染,利用多视图一致性重建	单视图缺乏形状约束
	Rebain等人(2022)	基于单视图对抗训练	构建先潜在特征向量映射神经辐射场,再投影图像对抗训练范式	不可见区域二维监督重建
	Monnier等人(2022)	基于图像特征解耦学习	利用同一类别不同实例图像,解耦学习物体光照、纹理、光照、视角和背景等属性	二维监督信息包含背景、光照等干扰
	Wu等人(2023b)	基于先验知识优化	提出近似形状对称学习重建架构	渲染学习形状先验不足
半监督	Piao等人(2019)	基于部分三维形状优化	使用小样本三维形状二维图像联合训练	小样本三维形状数据高质量重建
	Kohli等人(2020)	基于语义分割任务联合优化	融合语义分割监督训练任务与新视图合成无监督训练任务	物体三维结构语义感知重建
	Laradji等人(2021)	基于视角信息学习优化	预训练视角判别模块,获得视角伪标签训练重建	物体视角信息缺失
	Yang等人(2021a)	基于预训练模型优化	提取域不变图像特征进行重建	训练数据和测试数据域偏差
	Lai等人(2024)		提出跨类别知识转移重建架构	训练数据和测试数据类别不一致

有监督单视图三维物体重建方法主要基于编码器提取图像特征,解码器恢复物体形状的重建架构,使用3D形状重建损失进行有监督学习,如图1所示。在展示有监督方法编解码重建架构的基础上,本文归纳展示了后续的6个细分类研究工作的关键改进创新,以及与基础框架的输入、编码、解码和输出等环节的对应关系。

无监督单视图三维物体重建主要基于图像渲染学习的重建架构,无需标签数据,仅使用2D图像重建损失进行无监督学习,如图2所示。在展示无监督方法渲染重建架构的基础上,本文归纳展示了后续的5个细类研究工作的关键改进创新,以及与基础框架的输入、编码、解码和渲染等环节的对应关系。

半监督单视图三维物体重建方法主要在部分有监督、无监督重建架构基础上,利用二维、三维标签预测损失进行半监督学习,如图3所示。在展示半监督方法重建架构的基础上,本文归纳展示了后续

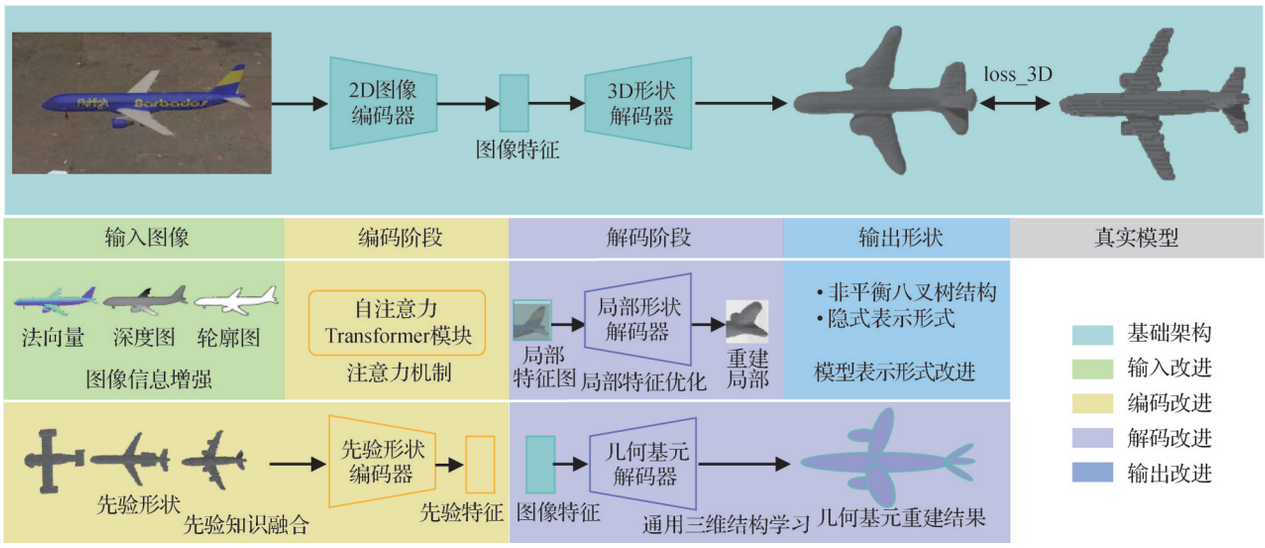

图 1　有监督学习单视图三维物体重建方法框架

Fig. 1　The framework of the supervised learning single view 3D object reconstruction

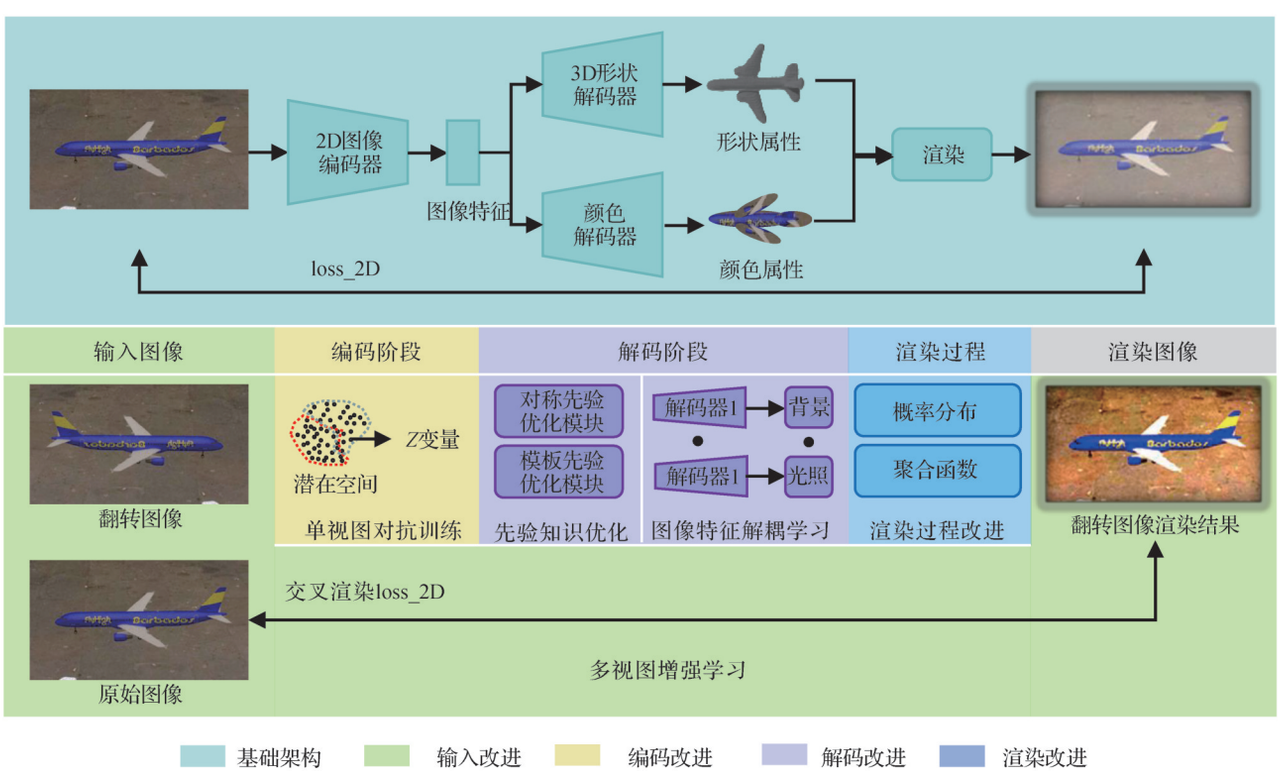

图 2　无监督学习单视图三维物体重建方法框架

Fig. 2　The framework of the unsupervised learning single view 3D object reconstruction

的 4 个细类研究工作的关键改进创新，以及与基础框架的输入、编码和解码等环节的对应关系。

上述 3 种监督学习方法技术框架各有优缺点。例如有监督学习方法的优点在于能够利用三维标注数据学习重建，重建质量高，但其受限于较高的数据标注成本；而无监督学习方法能够直接使用二维图像学习重建，有效降低了训练成本，但重建质量不够稳定；半监督学习方法针对前两者数据标注成本高与重建质量不稳定的问题，提出了标注数据与未标注数据联合学习的范式，目前得到广泛的研究。

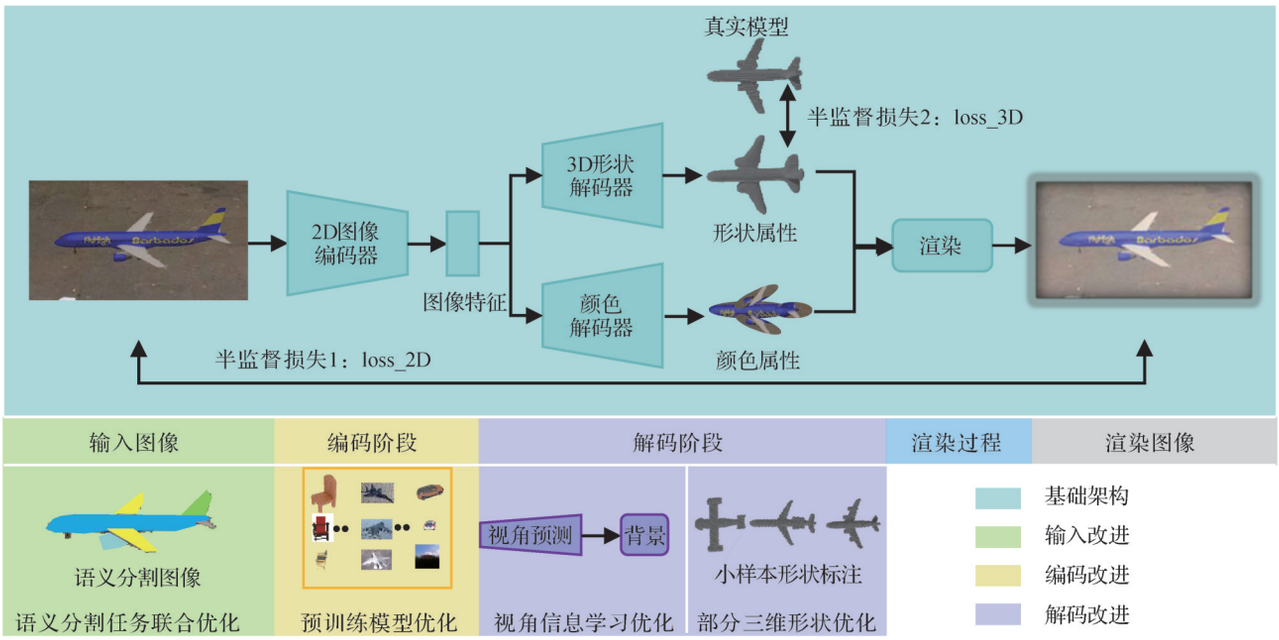

图3　半监督学习单视图三维物体重建方法框架

Fig. 3　The framework of the semi-supervised learning single view 3D object reconstruction

2　数据集与评价指标

2.1　数据集

基于深度学习的单视图三维物体重建方法需要高质量的数据集进行优化训练和测试评估。高质量的数据集为算法模型提供大规模训练与测试数据，直接影响模型的重建性能。单视图三维重建数据集分为RGBD数据集、合成数据集和真实场景数据集。

Nyuv2（Silberman 等，2012）、kitti（Geiger 等，2013）、SUN RGB-D（Song 等，2015）和 ScanNet（Dai 等，2017）等属于RGBD数据集，包含RGB图像和深度图像数据。RGBD数据集虽然不是专门设计用于三维物体重建方法的训练优化，但是常用于测试评估，检验基于深度学习的单视图三维物体重建方法的泛化能力。

ModelNet（Wu 等，2015）、ShapeNet（Chang 等，2015）、ABC（a big CAD model dataset for geometric deep learning）（Koch 等，2019）和 PartNet（Mo 等，2019）等属于合成数据集，包含三维模型和对应的渲染图像数据。合成数据集具有对象类别丰富、数据规模大和标注质量高等优点，常用于三维物体重建的训练优化和测试评估。

ikea（Lim 等，2013）、PASCAL3D+（Xiang 等，2014）、ObjectNet3D（Xiang 等，2016）和 Pix3D（Sun 等，2018a）等属于真实场景数据集，包含真实场景RGB图像和三维模型。真实场景数据集对图像和三维模型进行了精确对齐标注，适用于三维物体重建方法面向真实场景的测试评估。

2.2　评价指标

单视图三维物体重建任务中常用评价指标包括倒角距离（chamfer distance，CD）、EMD（earth mover's distance）、交并比（intersection over union，IoU）和 F_1-Score 等，可以分为距离评价指标和分类评价指标两类。CD 和 EMD 评价指标主要用于评估重建三维模型与真实三维模型之间的形状距离，距离值越小，代表重建三维模型与真实三维模型的整体形状越接近。IoU、F_1-Score 评价指标主要用于评估重建三维模型中每一个点的三维形状分类准确程度，IoU 和 F_1-Score 数值越大，则重建三维模型越准确。利用重建模型和真实模型计算评价指标，可以定量评价算法的重建性能。

3　有监督学习单视图三维物体重建

3.1　基于模型表示形式改进的单视图三维物体重建

体素、网格和点云是重要的三维重建表示形式，

随着分辨率的增加，存在模型计算量大和占用内存多等问题。基于模型表示形式改进的单视图三维物体重建方法用于平衡模型分辨率和计算资源，克服三维模型表示分辨率引起的计算资源需求过高的问题，以更精细的方式捕捉物体的形状和细节。

1）第 1 类模型表示形式改进方法采用非平衡八叉树结构作为三维重建表示方式。Riegler 等人（2017）提出一种稀疏三维数据的深度学习表示方法 OctNet（octree net），使用一组非平衡八叉树对空间进行分层划分表示三维形状。相比较于体素结构将三维空间均匀划分，OctNet 只在空间占有区域生成新的下一层八叉树节点，在保证高分辨率的情况下实现内存分配和计算集中到关键的区域，在处理稀疏形状时具有计算效率优势。Tatarchenko 等人（2017）提出一种基于非平衡八叉树结构的深度卷积解码器架构，学习预测八叉树的结构和节点的占用状态以表示物体三维形状。与基于体素、网格等表示形式的解码器相比，该方法在初始低分辨率三维模型的基础上，对预测为占有状态的局部区域进行进一步细分，从而在有限的内存预算下能够表示更高分辨率的输出。

上述方法随着三维模型分辨率的提高，需要学习预测大量的非表面节点，但非表面节点通常在较低分辨率学习的阶段就能够被正确地预测与表示形状。因此，训练过程进行高分辨率的非表面节点学习，浪费了大量可以用于进一步提高表面节点分辨率的计算资源。针对以上问题，Häne 等人（2017）提出了一种基于分层表面预测的单视图三维物体重建方法。使用八叉树结构将三维空间的分辨率进行分层，将每一层的八叉树节点按照自由区域、边界区域和占有区域进行分类。其中自由区域代表预测为没有物体占有的区域，占有区域与之相反。而边界区域为物体的边界，通常需要高分辨率表示以捕捉物体几何形状细节。因此，该方法只对每一层中预测为边界区域的节点进行更高分辨率节点划分。

对于占有物体空间区域，采用自适应采样的方式构建八叉树：平坦区域采用较大的八叉树划分，而局部细节区域采用更精细的八叉树表示，是进一步提高计算效率和优化存储的重要方向（Wang 等，2017）。Wang 等人（2018b）提出了一种自适应八叉树的卷积神经网络，由自适应 O-CNN（octree-based convolutional neural network）编码器和解码器组成，实现三维空间中重建物体区域的不同分辨率八叉树适应性表示。上述采用非平衡八叉树结构的三维模型表示形式减少了重建所需计算资源，在 128 × 128 × 128 等低分辨率三维重建领域得到广泛应用（Victoria 等，2023）。但随着重建物体细节要求提升，非平衡八叉树方法面临着有限的内存和较高的分辨率的冲突挑战。

2）第 2 类则是基于隐式表示的三维模型重建方法。Mescheder 等人（2019）提出一种基于连续三维空间分类学习的隐式三维物体重建网络 OccNet（occupancy networks），对三维空间中的随机采样点进行空间占有状态预测训练，从而获得空间中任意分辨率位置的分类能力。与非平衡八叉树结构等体素方法相比，OccNet 通过连续三维空间占有函数学习的方式克服了分辨率上限问题，同时具有计算与存储优势；与点云和网格等形状边界顶点数量有限的方法相比，OccNet 采用一种多分辨率等值面提取的方法，自适应选取合适的分辨率提取三维物体表面，实现更精细化的表面形状细节恢复。

除了占有状态学习方法，符号距离场函数学习方法同样被广泛应用到基于隐式表示的三维模型重建方法中。Park 等人（2019）提出了一种连续的符号距离函数表示方法 DeepSDF（deep learning continuous signed distance functions for shape representation），可以从局部缺失和有噪声的三维输入数据中实现高质量的形状表示、插值和补全。符号距离函数表示方法可以预测空间中任意位置到物体表面边界距离，同时距离的符号可以用于确定该点位置处于形状的内部还是外部。

与占有状态学习方法对三维空间进行占有分类相似，符号距离场函数学习方法的一种特殊情况是只考虑形状内外分类的学习。例如 Chen 和 Zhang（2019）提出了一种隐式空间解码器 IM-NET（implicit fields learning network），输入三维空间点坐标和对应形状编码器提取的特征向量，输出该点坐标对应的内外分类状态，利用内外区域的边界提取方法实现最终的三维模型表示。

基于隐式表示的三维模型重建方法，通过学习预测三维空间中的连续函数来生成任意分辨率的三维模型，展现了实现高精度单视图三维物体重建的巨大潜力。上述方法在整体形状学习与高分辨率表示方面具有优势，但隐式表示是连续函数形式，无法

直接学习物体边界，存在飞机机翼、汽车后视镜等局部区域表面轮廓不准确的问题，如图4所示。因此，如何准确定位基于隐式表示的物体边界，是重要的研究方向。例如Koneputugodage等人(2023)将非平衡八叉树结构与隐式表示结构相结合，一方面使用隐式表示结构学习高分辨率三维形状，另一方面使用非平衡八叉树结构学习准确的表面轮廓。

图4　基于隐式表示的重建方法结果示例
(Chen和Zhang，2019)

Fig. 4　Example of reconstruction results of implicit 3D representation (Chen and Zhang, 2019)

3.2　基于图像数据信息增强的单视图三维物体重建

3D-R2N2(Choy等，2016)等早期单视图三维物体重建方法在重建性能方面容易受到单个输入图像信息有限问题的干扰。基于图像数据信息增强的方法通过额外的图像处理过程，扩展了输入图像的信息维度，从而学习精细的三维形状信息和重建精细几何细节。当前基于图像数据信息增强的方法从输入图像的视角信息有限、分辨率等因素出发，设计对应的图像处理过程以提升单视图三维物体重建的精度。

从图像恢复整体三维模型的本质是对图像进行编码分类，然后解码映射到对应的已训练的三维形状。因此，当输入未训练的视角图像时，重建模型的精度往往有所下降。针对单视角的局限性问题，Zhang等人(2018)提出了一种单视图三维物体重建方法GenRe(generalizable single-image 3d reconstruction)。通过不可见表面深度图补全训练等方式获得完整视角的深度信息，然后利用完整视角的深度信息学习三维形状的映射。与只使用单视图进行三维形状映射学习的方法相比，GenRe使用了更全面的信息学习与三维形状的映射关系，提升了未知视角的自适应重建能力。

GenRe方法增加了更为全面的深度图信息，但直接利用深度图进行重建的方式忽略了原本输入图像丰富的细节纹理特征，对最终的重建精度有一定的影响。针对输入的单视角图像重建整体形状受其他视角图像信息缺失限制的问题，王年等人(2020)提出了一种端到端的视图感知三维物体重建网络。首先将单视图输入到多邻近视图合成网络，生成结构细节信息更为丰富的多视图数据；然后在3D重建子网络中，采用自适应融合方式将不同视角的图像特征合并，最后解码生成更为精准的三维形状。

输入图像的分辨率同样是影响重建模型精度的重要因素。由于低分辨率图像存储的局部细节信息模糊，在训练过程中容易引入不确定性，导致细节区域重建效果不佳。针对输入的单视图分辨率影响局部细节重建学习的问题，张豪等人(2021)提出了一种融合超分辨率、投影和对抗生成网络等模块的单视图三维物体重建方法。使用超分辨率模块对输入图像进行细节信息增强，降低对输入图像的要求，提升模型对于采集图像过程中对焦、摇晃等外界因素干扰的鲁棒性。与Wu等人(2018)提出的ShapeHD(shape high quality with fine details)方法相比，该方法实现了低分辨率输入图像中飞机、变电箱等边缘细节区域的重建。

基于图像数据信息增强的方法，提升了面对视角信息有限、分辨率低等多样化数据的整体形状高精度重建能力，在采集图像设备受限等场景具有重要应用价值(Zhang等，2024)。但是，上述方法的性能易受分辨率增强、深度图预测补全的结果误差影响。因此，如何设计合理的框架克服信息增强中的误差干扰，是值得进一步研究改进的方向。例如Zhang等人(2023)提出了一种名为Go-slam(global optimization for simultaneous localization and mapping)的框架，通过优化重建任务和姿态估计任务的结果一致性来缓解重建误差与失真。

3.3　基于注意力机制的单视图三维物体重建

前两种类别方法优化了输出和输入两个环节，而特征学习环节是连接输入和输出的重要桥梁，特征学习表征能力关系着局部细节重建精度。由于三维结构中不同区域的重建难度不同，复杂细节区域特征和简单平滑区域特征相互干扰，影响最终的三维重建训练效果的问题。因此，基于注意力机制

的方法对输入图像的复杂细节关键区域进行特征学习，通过注意力机制的方式引导网络优化难度更高的复杂区域三维形状，从而进一步提高最终单视图三维物体重建精度。当前基于注意力机制的方法主要致力于解决复杂背景干扰去除和细节结构感知等问题。

当前基于编解码结构的单视图三维物体重建方法在面向真实场景采集的数据时，编码对象结构特征往往会受到复杂背景区域干扰，导致解码训练过程不稳定。针对复杂背景特征干扰训练效果的问题，Li 和 Kuang（2021）提出了一种名为 3D-VRVT（3D voxel reconstruction from a single image with vision Transformer）的单视图三维物体重建网络。首先将输入图像分割成均匀大小的局部区域；然后输入基于自注意力机制的 Transformer 编码器中学习提取重建对象局部区域的特征，抑制背景区域特征学习；最后编码生成三维体素模型，实现复杂背景真实图像中高质量生成目标对象三维形状。

在去除背景干扰的基础上，进行输入图像中目标的关键区域特征学习，是提升单视图三维物体重建精度的重要途径。朱育正等人（2021）提出了一种基于注意力机制的单视图三维物体重建方法。将编码器提取的各层特征输入到 CBAM（convolutional block attention module）（Woo 等，2018）注意力卷积模块中，其中浅层特征包含了输入图像丰富的细节纹理信息，通过 CBAM 可以提升表征关键图像细节的能力，而深层特征往往包含重建对象的整体结构信息。最后经过解码器引导模板局部结构区域细粒度变形，从而提升重建三维模型质量。该方法提升了关键区域的特征表达能力，但对关键区域进一步实现空间结构感知仍然是一大挑战。针对重建结果出现的断裂、缺失等空间结构不完整情况，白静等人（2022）提出了一种基于结构和目标感知的三维物体重建方法。融合注意力机制构建结构感知编码器，从水平方向和垂直方向捕捉输入特征图的结构信息。然后通过压缩激励机制对其进行加权，获得既包含局部细节又包含全局信息的增强特征图，从而提升对细节区域的重建性能。

当前单视图三维物体重建方法还存在边角和细节过于平滑等挑战。针对三维物体重建过程中边角细节感知难题，Jin 等人（2022）提出了一种 IMC-NET（implicit field with corner attention network）方法。IMC-NET 引入一个边角注意分支，并利用 3D 边角关键点数据集，进行有监督训练来增强三维形状的边角感知能力。在此基础上聚合模块用于融合多分支的预测结果，从而进一步提升最终重建精度。

上述方法利用边角等具有明显位置信息特征区域进行重建优化，然而对于复杂结构的模糊区域重建优化仍然是一大挑战。针对复杂结构区域形状概率预测性能增强的难题，李雷等人（2022）提出了一种融合深度强化学习算法进行概率预测增强的单视图三维物体重建方法。首先使用多分支网络模块对输入图像进行编解码，其中一个分支获得原始的预测概率分布，另一个分支输入到邻域路由机制聚合模糊概率点区域信息；最后结合深度强化学习算法进行模糊概率调整，提升模型的复杂拓扑结构高保真度重建能力。

IMC-NET 等基于注意力机制的单视图三维物体重建方法，引入额外注意力分支结构提升重建精度，但这也增加了网络的参数量和运算量，特别是对于高分辨率场景，计算复杂度是一个重要挑战。真实应用场景往往对算法的参数量和运算量有所限制，所以平衡单视图三维物体重建方法的精度与运算量，是未来值得深入研究的潜在方向。例如，吴繁和贺赛先（2023）提出一种轻量化三维物体重建方法。设计注意力模块，为图像空间信息分配不同的权重，在较低的模型参数量和运算量下具有更好的三维重建效果。万骏辉等人（2024）首先检测目标位置，然后利用 Transformer 对目标区域提取物体细节局部特征。该方法仅需要关注目标区域，降低了算法的参数量和运算量。

3.4 基于局部特征优化的单视图三维物体重建

基于局部特征优化的方法旨在提高物体视角可见区域局部细节重建能力。当前基于局部细节特征融合优化的方法包括全局特征与局部特征融合恢复形状、全局特征与局部特征分离恢复形状等方式。

全局特征与局部特征融合恢复形状方式主要针对三维形状中复杂结构细节的重建问题。当前单视图三维物体重建方法利用全局特征恢复成三维形状。而全局特征主要存储深层的整体形状信息，局部结构特征信息未能得到充分利用。为解决潜在空间向量缺乏足够细节表征能力的问题，Hsu 等人（2020）提出了一种单视图三维物体重建方法。在图像编码提取特征阶段采用空洞卷积，以增强局部细

节特征表征能力；在潜在空间向量解码恢复三维形状阶段采用多层级输入图像与潜在空间向量连接融合进行上采样，以增强浅层和中层结构特征表征能力，从而提高结构细节重建精度。

上述方法（Hsu等，2020）增强了细节表征能力，但特征与三维结构细节之间缺乏直接映射学习。而隐式表示三维物体重建方法具有直接关联三维空间位置信息和输入图像特征的特性，为进一步优化局部细节结构提供了可能性。针对局部特征映射优化三维局部结构难题，Zhao等人（2021）提出了一种预测无符号距离场从单个图像重建服装三维模型的方法。在提取输入图像特征的基础上，利用无符号距离场的梯度方向获得准确的投影方向，从而融合对应的局部图像特征和三维局部结构特征，在解码阶段联合优化最终的无符号距离场。

全局特征与局部特征分离恢复形状是应用到复杂结构细节重建的另外一种重要方式。由于全局特征主要用于整体形状重建，而局部特征用于捕捉局部结构，两者融合训练过程中存在一定的冲突，Wang等人（2020）提出了一种基于有符号距离场的单视图三维物体重建方法DISN（deep implicit surface network）。DISN方法提取输入图像的全局特征表征整体形状，提取对应空间位置的局部特征捕捉孔洞等局部细节，然后将两者融合用以恢复最终的细粒度三维形状。梁春阳等人（2023）提出了一种基于稀疏特征改进的单视图三维物体重建方法。该方法利用全局特征预测有符号距离函数的高频残差，表示重建物体的整体形状。而稀疏特征用于预测有符号距离函数的低频部分，表示重建物体的表面细节。这两部分相加得到最终的符号距离函数，解决了表面凹凸不平的局部细粒度重建难题。

与直接将全局特征和局部特征融合优化方式相比，上述方法保留了全局特征和局部特征各自的信息维度，转而在最终的预测结果侧进行融合优化，一定程度上抑制全局特征和局部特征的相互干扰。然而，DISN等方法在训练过程以融合后的整体形状损失为主，复杂结构细节产生的损失较小，限制了重建细节结构的进一步优化。针对当前损失函数忽略局部细节特征优化的问题，Li和Zhang（2021）提出一种单视图三维物体重建网络D2IM-Net（detail disentangled implicit fields learning network）。D2IM-Net将全局特征和局部特征分别解码生成粗糙的整体形状和细粒度的局部表面细节结构，然后使用粗糙整体形状损失、表面细节损失和融合整体结构损失联合优化，从而保证局部细节特征的表征性能，实现复杂结构细节的进一步重建优化。

基于局部细节特征融合优化的方法展现了高精度三维物体重建的巨大潜力。但此类方法通常需要进行输入图像的视角估计，以获得图像局部特征与三维形状的位置对应关系，实现最终局部细节结构优化。而真实场景往往缺乏多样化且准确的视角信息训练数据。因此，如何克服视角估计性能的制约，是提升基于局部细节特征融合优化的方法实际应用性能的重要研究方向。例如与上述需要进行视角估计重建统一位姿三维形状的方法相比，Wimbauer等人（2023）提出一种预测图像中每个位置的体密度的方法，仅用当前视角方向完成几何结构的预测。

3.5 基于先验知识融合的单视图三维物体重建

上述基于局部特征优化的方法利用图像局部细节信息增强重建性能，但单视角仅能提供有限信息。而基于先验知识融合的方法主要从重建对象本身的特性出发，融合目标物体三维形状知识，采用更为全面和多维度的信息，以实现鲁棒泛化单视图三维物体重建。基于先验知识融合的方法主要包括融合形状几何先验和融合多任务一致性先验等方式。

融合形状先验知识的方法，使用丰富多样化的合成数据集的形状先验知识，与有限的训练数据形状信息相融合，以提升单视图三维物体重建的泛化性能。例如Pinheiro等人（2019）提出了一种基于两种数据对抗性训练的单视图三维物体重建方法。首先使用大规模合成数据集的三维形状，训练编码器E^*与解码器D^*，使其获得多样化的三维形状先验知识。然后在单视图三维物体重建训练阶段，将目标物体特征输入到解码器D^*中进行三维形状恢复。为优化图像特征与形状先验知识的融合，使用鉴别器D_{shape}强迫提取的图像特征接近于三维形状特征分布。此外，为消除真实图像和合成图像的差异影响，使用另一个鉴别器D_{img}进行对抗性训练。

上述方法（Pinheiro等，2019）将图像特征与预训练的先验形状知识进行融合，但是目标对象与形状先验知识缺乏直接映射学习，限制了鲁棒泛化性能的进一步提升。针对使用形状先验知识有效重建新对象的难题，Michalkiewicz等人（2020）提出了3种方法从大规模基类数据和小样本重建对象数据学习类

别全局形状。这3种方法的核心思想是训练一个隐式组成结构捕捉三维形状的多尺度信息，并学习形状先验特征组合范式，用以反映三维形状的结构变化。在重建阶段，将图像的特征与形状先验组合特征输入到解码器，从而完成新对象的鲁棒泛化单视图重建任务。Yang等人(2021b)提出了一种名为Mem3D (single-view 3D object reconstruction from shape priors in memory)的单视图三维物体重建方法，由图像编码器、记忆网络、LSTM(long short-term memory)形状编码器和形状解码器等部分组成。图像编码器提取带有噪声真实场景图像特征；接着通过记忆网络检索相似的先验3D形状，使用形状编码器提取先验形状细节特征；最后将带有噪声真实场景图像特征与先验形状细节特征融合，通过形状解码器重建物体形状。

先验形状组合方法实现三维形状与输入图像的直接映射关系，但输入图像的形状类别与先验形状可能存在冲突，这对目标形状的细节恢复存在一定的影响。针对形状先验知识与目标形状相适应的问题，Nie等人(2023a)提出了一种单视图三维物体重建方法CPG3D(cross-modal priors guided 3D object reconstruction)。首先利用图像检索形状近似三维模型；然后提取输入图像和形状先验的特征，并通过跨模态注意模块进一步融合；最后进行解码完成形状重建，并采用优化模块对形状进行细粒度优化，以消除先验形状和目标形状差异造成的干扰。

上述基于先验形状融合的方式提升了单视图三维物体重建的泛化性能，但由于合成数据集往往难以覆盖真实场景的复杂的形状变化，泛化重建范围存在一定的限制。针对先验形状知识的范围限制问题，Zhou等人(2020b)提出了一种基于反射对称性几何先验知识的单视图三维物体重建方法。该方法从三维形状中较为普遍的对称几何关系出发，首先从图像中检测对称对象的三维镜像平面，然后通过计算相对于对称平面的像素级对应关系来恢复深度图，拓宽了单视图三维物体重建方法的泛化范围。

几何先验知识相较于形状先验具有更广泛适用范围，但不同目标形状适用的几何先验关系存在不同。因此，先验知识的适用范围是制约重建性能的重要因素，如何设计融合当前场景更有效的先验知识，是未来的重要研究方向。例如图像的目标检测、深度估计等一系列视觉感知任务的预测结果并不是独立的，具有跨任务的一致性先验知识。Zamir等人(2020)提出了一种基于跨任务一致性学习的单视图重建方法，通过跨任务一致性约束的训练，减少不同任务之间冲突与矛盾的预测结果。

3.6 基于通用三维结构学习的单视图三维物体重建

一部分单视图三维物体重建方法可归类为识别模式，利用类别先验形状生成模型，不适用于未知类别重建任务。而基于通用三维结构学习的方法旨在学习与类别解耦的三维结构知识，实现重建模式。

Zhou等人(2021)提出一种数据分散分数，证明识别模式还是重建模式取决于训练数据分散程度，提高训练形状分散程度更利于学习与类别解耦的三维结构知识。但数据集中形状类别是有限的，进一步提高形状分散程度需要昂贵的数据采集标注成本。而通用三维结构学习方法可以在有限的分散数据中提升重建泛化性能。例如Wang和Fang(2020)提出了一种通用三维结构学习和重建的方法GSIR (generalizable 3D shape interpretation and reconstruction)。在三维重建任务的基础上，GSIR方法引入三维结构信息预测任务，包括预测三维形状中各个组合的长方体的位置、方向和大小等。然后基于预测的三维结构微调恢复完整三维模型，增强了重建过程与通用三维结构的联系与可解释性。

上述基于长方体结构信息预测的方法提升了单视图三维物体重建泛化性能，但此类方法通常以重建对象为中心。即训练过程中不同视角的单视图输入、输出的重建对象为固定方向，这为当前视角信息的学习引入了干扰与偏差。而Kuo等人(2022)提出了一种以视图为中心的、基于几何基元的单视图三维物体重建方法。该方法主要分为3个阶段进行三维重建学习：第1阶段进行深度估计任务学习，获得当前视角的三维结构信息；第2阶段使用深度图全局特征预测几何基元的重心，局部特征预测几何基元的半径与角度；第3阶段结合预测的深度图和几何基元，优化网格变形获得视角一致三维模型。

基于通用三维结构学习的方法从单视图三维物体重建的识别与重建模型区别出发，通过数据集优化和三维形状通用结构信息学习等方式，进一步提高了单视图三维物体重建方法的泛化性能。但通用结构学习面临着无法覆盖重建形状所有结构的挑战。例如Interpretation(Wang和Fang，2020)、GenRe

(Zhang 等, 2018)和 GSIR(generalizable 3D shape interpretation and reconstruction)(Wang 和 Fang, 2020)等方法虽然预测了输入图像的整体形状,但桌腿、扶手和靠背等细节结构重建存在优化需求,如图5所示。因此,如何增强输入图像的局部细节结构和通用形状的对齐映射关系,是进一步提高通用结构学习效果的重要方向。例如Gan等人(2023)提出一种自适应映射网络SOM,对重建对象的先验平均形状进行优化,以获得输入图像形状的近似值,再融合输入图像特征进一步重建细粒度三维结构。

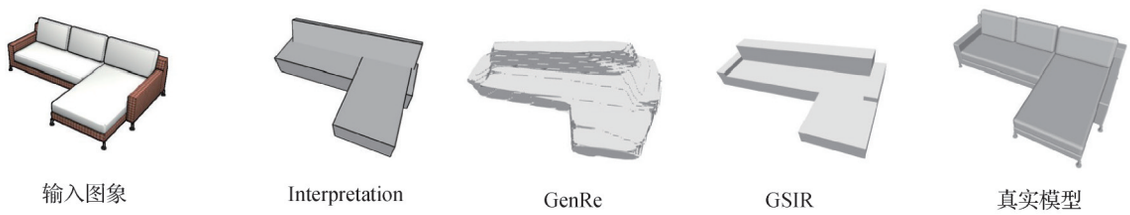

图5 通用三维结构学习方法重建结果对比

Fig. 5 Reconstruction results comparison of universal 3D structure learning methods

3.7 有监督学习单视图三维物体重建方法小结

与无监督学习、半监督学习单视图三维物体重建方法相比,有监督学习单视图三维物体重建方法利用三维形状标注数据监督训练,改进输入、输出、特征学习和形状优化等各个环节,具有重建性能优势。表2总结了上述有监督学习单视图三维物体重建改进方法的设计思路和优缺点。从各类方法所面临的挑战出发,对物体形状边界确定、信息增强误差干扰等问题继续深入研究,具有进一步完善有监督学习单视图三维物体重建效果的巨大潜力。

表2 有监督学习单视图三维物体重建方法总结

Table 2 Summary of supervised learning single view 3D object reconstruction methods

方法类别	设计思路	优点	缺点
基于模型表示形式改进	采用非平衡八叉树结构,基于隐式表示形式	减少高分辨率重建所需计算资源	分辨率限制,物体边界定位难题
基于图像数据信息增强	深度、视角和分辨率等信息增强	提高不同质量数据重建的鲁棒性	信息增强误差干扰
基于注意力机制	引导学习前后景和细节结构等关键区域	增强关键区域结构学习性能	计算复杂度高
基于局部特征优化	使用局部特征优化局部细节质量	提升局部细节细粒度重建质量	视角估计性能制约
基于先验知识融合	几何先验和形状先验学习	弥补数据驱动学习的知识局限性	先验知识适用范围问题
基于通用三维结构学习	数据集离散优化和几何基元学习	确保重建三维结构的可信度与合理性	三维结构覆盖范围问题

表3展示了部分有监督学习单视图三维物体重建方法相关方法与3D-R2N2(Choy 等, 2016)、Pixel2Mesh(Wang 等, 2018a)、AtlasNet(Groueix 等, 2018)和IMNET(learning implicit fields for generative shape modeling)(Chen和Zhang, 2019)等方法的定量对比结果。由于当前有监督学习单视图三维物体重建方法主要都以编码器解码器为网络框架基础,局部结构重建精度易受到单视图不可见区域结构因素干扰。一方面编码器提取的图像特征不足,仅包含当前视角下可见区域的局部结构信息;另一方面解码器还原不可见区域结构产生的损失更大,最终训练过程往往偏向于还原整体形状而忽略了局部结构质量。因此,不可见区域结构干扰是所有有监督学习单视图三维物体重建方法面临的共同挑战,限制着重建精度的进一步提升。

当前3D-R2N2(Choy 等, 2016)、Pix2Vox(Pixel2Voxel)(Xie 等, 2019)和Umiformer(Unified Transformer)(Zhu 等, 2023)等方法融合不同视角的单视图重建结果可以提升部分不可见区域重建性能,但部分遮挡区域结构仍然重建难度高。而Wang等人(2024)提出一种单视图三维物体重建方法Slice3D,从输入视图预测不同方向的物体切片图

表 3　有监督学习单视图三维物体重建方法定量对比

Table 3　Quantitative comparison results of single view 3D reconstruction methods with supervised learning

方法	数据集	评价指标	对比方法	对比结果	实验结果
Tatarchenko 等人（2017）	ShapeNet-all	IoU ↑	3D-R2N2	0.560	0.596
Sun 等人（2018b）	4 categories (car, chair, table and guitar) of ShapeNet	IoU ↑	3D-R2N2	0.235	0.298
王年等人（2020）	ShapeNet Core	IoU ↑	Pixel2Mesh	0.473	0.594
Chen 等人（2019）	5 categories (Plane Car Chair Rifle Table) of ShapeNet Core	LFD ↓	AtlasNet	3 804	3 541.4
王年等人（2020）	ShapeNet Core	IoU ↑	3D-R2N2	0.597 6	0.668 9
Hsu 等人（2020）	ShapeNet	IoU ↑	3D-R2N2	0.560	0.677
沈伟超等人（2021）	ShapeNet	IoU ↑	3D-R2N2	0.560	0.618
Zhao 等人（2021）	Deep Fashion3D	CD ↓ /P2S ↓	Pixel2Mesh	4.266/5.33	0.621/0.839
张豪等人（2021）	ShapeNet	CD ↓	AtlasNet	0.134	0.107
Yang 等人（2021b）	ShapeNet	IoU ↑	3D-R2N2	0.560	0.729
朱育正等人（2021）	ShapeNet	IoU ↑	3D-R2N2	0.493	0.593
Li 和 Kuang（2021）	ShapeNet	IoU ↑	3D-R2N2	0.560	0.654
白静等人（2022）	ShapeNet Core	IoU ↑	3D-R2N2	0.560	0.687
Jin 等人（2022）	Shape-CornerNet	IoU ↑	IMNET	0.527	0.541
梁春阳等人（2023）	4 categories (Plane Car Chair Sofa) of ShapeNet Core	CD ↓	3D-R2N2	0.484	0.462
吴繁和贺赛先（2023）	ShapeNet	IoU ↑	3D-R2N2	0.560	0.664

注：“↑”表示值越大越优；“↓”表示值越小越优。

像。Slice3D 基于当前视角二维图像学习物体遮挡区域的二维结构信息，降低了二维图像直接学习三维不可见区域结构的难度，为解决不可见区域结构重建难题提供了可能的研究方向。

4　无监督学习单视图三维物体重建

4.1　基于渲染过程改进的单视图三维物体重建

针对渲染逆向过程不可导问题，Loper 等人（2014）设计了一种近似的可微渲染方法 OpenDR（open differentiable renderer），使用图像颜色的梯度近似作为像素颜色和坐标的梯度，迭代优化实现从单个二维图像中恢复三维形状。虽然近似的可微渲染方法实现了单视图三维物体重建，但存在渲染前向过程和后向过程不一致性的问题。因此，Liu 等人（2019）设计一种基于真正可微渲染框架的单视图三维物体重建方法 SoftRas（soft rasterizer）。传统渲染过程中，像素由离散的三维形状位置和外观信息决定。而 SoftRas 方法通过概率分布模块和聚合函数模块，将像素和所有三维形状位置和外观信息联系起来，从而实现更加平滑的可微渲染效果，增强反向过程恢复三维形状的性能。

上述 OpenDR、SoftRas 等可微分渲染的方法预测的三维形状主要基于点云、网格表示方式，存在离散化和低分辨率的问题。Niemeyer 等人（2020）提出了一个隐式形状和纹理表示的可微渲染方法。与上述方法需要初始化点云、网格表示的方式不同，该方法直接从二维图像中学习隐式的形状和问题表示。首先将采样的像素点 u 投影到三维空间中，并沿相机原点到这一点的射线上以固定步长预测占有概率；然后从占有网络提取深度信息，输入到纹理预测网络得到二维渲染图像，使用渲染重建损失反向传导优化得到一个连续的三维形状隐式模型，实现了与有监督方法相当的重建性能。

二维图像损失优化过程如何确保三维结构的准确性是另一个重要难题和挑战。Zubić和Liò(2021)提出了一种基于三维点云轮廓优化的单视图三维物体重建方法。首先输入单幅图像,预测初始的三维点云和视角信息;然后与上述渲染过程不同,该方法使用基于生成对抗网络(generative adversarial network,GAN)的纹理映射获得轮廓图像;最后,为了在轮廓图像优化过程中增强对三维结构的学习,该方法在轮廓重建损失的基础上,设计结构不变性损失以确保预测的三维点云均匀可靠,从而提高重建的性能。

基于渲染过程改进的方法是重要的无监督单视图三维物体重建方式,奠定了使用单幅二维图像学习三维形状的基础。SoftRas方法在整体形状重建取得了与有监督学习方法Pixel2Mesh相当的效果,而且能够恢复桌腿等局部细节结构。但当前方法重建过程中需要去除背景,在实际应用场景中往往面临着背景干扰的挑战难题。因此,设计合理的抑制背景干扰的可微分渲染框架,是进一步提升其适用范围和泛化能力的重要研究方向。目前一些研究工作(Wang等,2023;Liu等,2024)引入多阶段重建策略,首先利用SAM(segment anything model)预训练模型(Kirillov等,2023)等实现前景对象高质量自动分割,再进一步重建目标对象三维结构。

4.2 基于多视图增强学习的单视图三维物体重建

在基于渲染过程改进的方法中,三维结构监督信息的缺失可能导致重建结果的不稳定。而基于多视图增强学习的方法,在训练过程中使用同一对象的不同视角数据学习三维形状,可以在一定程度上约束重建结果,减少重建中出现的错误。Tulsiani等人(2022)引入不同视图的可微光线一致性,从而提高三维物体重建的准确性和稳定性。然而该方法在重建过程需要从不同视角采集目标对象的多幅图像,并不适用于单视图物体重建场景。

Ho等人(2021)提出了一种多视图学习的改进方法。在训练阶段,使用同一对象多视图数据集学习更高精度无监督单视图重建,在推理重建阶段只需要物体的单幅图像,拓展了其应用场景范围。该方法使用来自同一类别的图像来训练单视图三维物体重建网络,在可微分渲染框架常用的重建像素误差的基础上,一方面通过投影生成不同视图的重建误差学习该类别对象的通用三维结构知识;另一方面引入反照率损失进一步提升重建的合理性和质量。

在多视图优化的基础上,Yao等人(2020)提出了一种仅需前后两个视图训练的单视图三维物体重建方法,在推理重建阶段恢复输入图像的前后视图以得到对应的三维形状,进一步降低了训练数据的要求。该方法首先预测输入图像对应的深度图、法向量图和轮廓图;然后对深度图和法向量图进行对称性预测以丰富当前视角的三维结构信息;最后假设物体的前后视图可以展现物体的绝大部分三维结构信息,通过前视图和后视图的轮廓一致性,预测对应的后视图并与前视图融合重建出物体三维模型。

上述方法需要同一物体的不同视图训练,而Duggal和Pathak(2022)从同一类别对象不同视角变形一致性的角度出发,提出了一种基于拓扑感知变形的方法,旨在学习三维形状与二维图像的密集对应关系。该方法主要由基于LSTM的可微渲染器、拓扑感知变形网络和规范形状生成器模块等组成。对于输入图像对应的物体,使用拓扑感知变形网络学习语义结构信息,指导物体三维坐标点的位置变换。然后利用可微渲染器和规范形状生成器模块预测对应的渲染图像与基于SDF(signed distance function)的隐式三维形状表示。最后使用对应的渲染图像RGB损失和三维形状的轮廓损失优化,以获得同一类别不同实例图像的泛化重建性能。

上述方法通过同一物体类别的不同视图数据训练,学习提取重建对象的多视角结构能力,进一步增强了无监督三维形状重建的合理性。但在实际应用场景中,采集同一物体类别大规模多视角数据往往存在限制。因此,设计克服同一对象数据依赖的多视图信息增强方法,是进一步拓展基于多视图增强学习方法应用范围的重要研究方向。Chen等人(2024)提出一种基于高质量背面视图生成的单视图物体三维重建方法Recon3D。与使用正反视图训练方法(Yao等,2020)不同,Recon3D不需要正面与背面视图对数据,而是利用Zero 1-to-3(Liu等,2023b)生成模型强大的零样本新视图生成能力。Recon3D使用Zero 1-to-3预测与正面视图轮廓一致的背面视图,减少多视角数据的依赖;然后使用正面视图与预测的背面视图指导NeRF(neural radiance fields for view synthesis)渲染,并引入点云渲染生成高保真视图一致的几何结构与细节纹理。

4.3 基于单视图对抗训练的单视图三维物体重建

基于多视图增强学习的方法在真实场景应用中面临多视角数据有限的难题。而生成对抗网络方法为从一幅视图学习全面视角结构信息提供了可行方案。Kato 和 Harada（2019）提出了一种基于单视图对抗训练重建方法，训练生成器获得不同视角的三维形状，然后使用鉴别器进行对抗性判别训练，从而在保证已观察视图重建性能的基础上，进一步提升对于未观测视角区域三维结构重建的合理性。

上述方法通过对抗训练方式将有限的多视图的三维结构信息和整体形状信息联系起来，从而进一步提升对未观察区域的重建效果。而 Zhou 等人（2020a）提出一种仅利用单幅图像学习不同视角结构的人脸三维重建方法，克服需要同一个对象多视图数据的限制。首先给定一个 Pa 视角的输入图像，学习三维网格形状和纹理，并将其旋转到随机视角 Pb；然后将 Pb 视角的三维网格形状和纹理旋转回初始视角 Pa，渲染图像在保留部分已观察区域局部纹理信息的基础上，引入了其他视角的结构纹理信息；最后通过高分辨率生成对抗网络 GAN 将渲染图像转换回真实图像，使用对抗性损失等训练调整输入图像的三维形状。该方法不依赖于配对数据或任何类型的标签，实现了仅使用单幅人脸图像即可完成人脸三维重建任务。

与随机转换不同视角学习方式相对应，标准视角学习再投影到输入视角的对抗训练方式，同样是仅使用单幅图像学习全面视角信息的可行方案。Rebain 等人（2022）提出了一种基于神经辐射场的生成式三维模型的学习方法 LOLNeRF（learn from one look）。基于神经辐射场的三维重建方法与分辨率解耦，且在纹理等细节方面具有优势。Nerfies（Park 等，2021）等基于神经辐射场的方法需要多视图进行重建，而 LOLNeRF 方法通过结合生成对抗网络的方式实现单视图重建，使用标准姿态下的潜在空间向量生成神经辐射场，然后投影生成不同姿态图像，以对抗训练模型，在新视图合成和深度估计等方面取得良好效果。

基于单视图对抗训练的方法在有限图像的基础上，进一步增强对三维形状未观察区域的重建能力，为提升无监督单视图三维物体重建的整体形状质量提供了可行方案。当前方法使用生成图像视角的合理性鉴别的方式，减少了对同一对象不同视角数据的依赖，但鉴别器的性能受训练数据的多样性限制。例如合成数据集的对象种类丰富，可以提供多样化的二维视图训练鉴别器，从而提升生成器生成不同视图的合理性；而真实场景数据集的对象种类有限，限制了鉴别器的训练效果，导致不同视图的学习生成效果下降。

因此，训练数据多样性不足是当前生成对抗网络方法面临的重要挑战和待解决的问题。例如 Kim 和 Chun（2023）提出一种迁移学习重建方法 DATID-3D（diversity-preserved domain adaptation using text-to-image diffusion for 3d generative model），通过文本引导三维结构重建域自适应的方式，在不需要额外图像的情况下实现重建性能提升。

4.4 基于图像特征解耦学习的单视图三维物体重建

二维图像的纹理光照特征和不变的形状特征耦合是干扰三维模型重建性能的重要因素。因此，基于图像特征解耦学习的方法，例如 DIB-R（interpolation-based differentiable renderer）（Chen 等，2020b）添加专门网络模块分别学习形状、光照和纹理等信息，以排除干扰并提高重建性能。

DIB-R 方法依赖视角信息，而真实场景数据往往没有视角标注，这限制了 DIB-R 应用范围。Monnier 等人（2022）提出了一种基于同一类别不同实例原始图像的单视图三维物体重建方法。在上述可微分渲染框架方法学习形状、纹理的基础上，进一步学习视角姿态和背景模型。其中学习背景模型减少对象轮廓注释的依赖。学习视角姿态从而将三维形状与当前视角图像更好地对应优化。Hu 等人（2021）提出了一种自监督网格重建方法，在物体视角预测学习的基础上渲染不同视角的物体图像，提取新渲染图像的三维属性以学习属性变换的一致性，进一步提升二维监督信息学习三维形状的能力。不同于上述视角姿态学习的改进方向，Dundar 等人（2023）提出一种两阶段形状纹理解耦学习的单视图三维物体重建方法，第 1 阶段进行形状学习，第 2 阶段进行纹理学习，并通过鉴别器保证形状与纹理的一致性学习质量。

与刚性物体不同，动物等重建对象的关节区域可移动，Aygün 和 Mac Aodha（2024）提出一种跨类别的单视图物体三维重建方法 SAOR。该方法首先预测物体基础形状，然后将其分割成不同的组成部件，

对每个组成部件进行位移形变优化，在此基础上进一步学习纹理与视角姿态。为提升关节等移动区域重建性能，引入了跨类别的关节区域形变一致性损失。

通过对形状、背景和纹理等特征分离学习，图像特征解耦学习方法提高了基于可微分渲染框架方法的重建精度。但上述方法在物体缺失等场景、形状特征学习面临挑战。如图6所示，单视图中仅包含动物头部信息，物体重建失败，最后恢复了训练类别的平均形状（Wu等，2023a）。因此，如何设计合理的三维形状特征优化方法，是进一步提升重建性能的重要方向。例如Huang等人（2023）提出了一种单视图三维物体重建方法ShapeClipper，在视角姿态估计的基础上，利用二维图像的几何约束优化对应视角区域的三维形状结构，提高其学习全局形状结构和局部几何细节的能力。

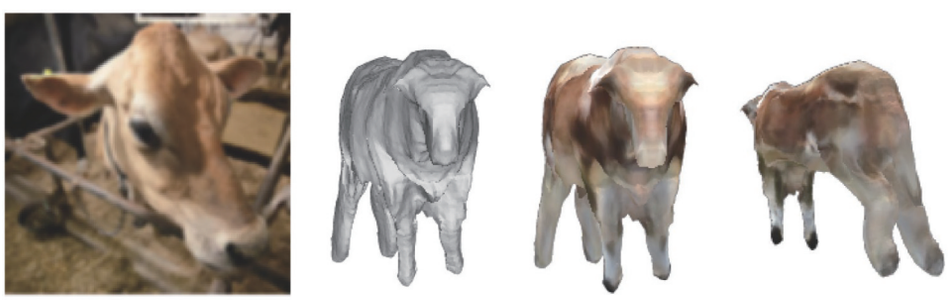

图6　重建失败示例（Wu等，2023a；Aygün和Mac Aodha，2024）

Fig. 6　Examples of failed reconstruction result（Wu et al.，2023a；Aygün and Mac Aodha，2024）

4.5　基于先验知识优化的单视图三维物体重建方法

三维形状先验知识能够提供更多维度的特征信息，以适应大尺度变形对象重建的挑战。例如Mei等人（2022）提出了一种基于模板的重建方法，将模板变形渲染图像任务和长度测量任务联合训练，在鱼类重建数据集上达到最先进的精度。但该方法需要专门类别三维形状模型，增加了数据标注成本，且对于结构变化对象的重建存在限制。

Yao等人（2021）提出了一种基于三维形状基元自监督学习的单视图三维物体重建方法。首先使用椭球体、圆柱体和长方体等基础形状预训练，以获得形状基元特征的学习与重建能力；然后训练一个单视图重建网络，提取输入图像的组合形状特征，使用预训练形状解码器恢复出对应三维形状基元；最后组合成完整的形状渲染出对应图像，通过常用的渲染重建图像损失监督优化重建效果。但上述方法需要预训练，而对称性先验知识可以利用物体本身结构信息，进一步完善重建形状细节。例如Wu等人（2023b）提出了一种基于人脸近似对称先验知识的单视图三维物体重建方法。由于在实际人脸数据中，头发、光照等存在随机的不对称，直接学习近似对称知识存在挑战。该方法分解学习对应的深度、反射率和光照等因素，实现重建性能提升。

上述基于先验知识优化的方法引入先验知识，进一步提升了三维形状的学习能力。但是当前方法面临着先验知识引入的误差干扰难题。虽然形状基元联合表征了三维形状，但较难覆盖所有结构，例如车轮、桌腿和机翼等形状。因此，设计克服先验知识引入误差的策略，是进一步提升当前方法性能的重要方向。例如Nie等人（2023b）提出一种学习三维物体先验分布和形状的方法，通过分类预测、三维目标检测和模板形状重建等多任务训练的方式，缓解单任务学习噪声以实现重建质量的提升。刘晓楠等人（2024）提出一种带深度监督的神经辐射场虚拟视点画面合成方法。该方法在RGB图像监督的基础上，引入二维密集深度值监督，有效降低体渲染过程的深度信息不准确的影响。

4.6　无监督学习单视图三维物体重建方法小结

表4总结了上述无监督学习单视图三维物体重建改进方法的设计思路和优缺点。三维形状渲染过程改进是确保无监督学习方法可行性的基础，后续几类方法则在视图信息增强、图像特征学习和形状监督优化等各个环节进行了优化。从各类方法面临的挑战出发，对背景干扰、多视角数据限制等问题继续深入研究，具有进一步完善无监督学习单视图三维物体重建效果的巨大潜力。

表4　无监督学习单视图三维物体重建方法总结

Table 4　Summary of unsupervised learning single view 3D reconstruction methods

方法类别	设计思路	优点	缺点
基于渲染过程改进	设计不同模型表示形式的可微渲染框架	增强二维信息监督学习三维结构的可解释性	背景干扰
基于多视图增强学习	多视图学习提取重建对象的完整结构能力	提升物体不可见区域结构的重建性能	多视角数据限制
基于单视图对抗训练	单视图对抗训练方式增强不同视角形状准确性	减少同一对象不同视角训练数据的依赖	训练数据多样性限制
基于图像特征解耦学习	学习分离物体形状特征和其他耦合的特征	提高形状与纹理学习重建性能	形状特征优化问题
基于先验知识优化	使用几何基元等先验知识进行形状优化	在二维监督学习基础上补充三维结构监督信息	先验知识引入误差

表5展示了部分无监督学习单视图三维物体重建方法相关方法与ShapeNet 3D（Tulsiani等，2022）、NMR（Kato等，2018）、SoftRas（Liu等，2019）、DVR（differentiable volumetric rendering）（Niemeyer等，2020）、3D Supervised（Hu等，2021）和Cat3D（category-guided 3D）（Huang等，2022b）等方法的定量对比结果。与有监督学习单视图三维物体重建方法相比，上述无监督学习单视图三维物体重建方法利用二维图像数据监督训练，保证重建性能的同时具有数据成本优势。但当前无监督学习单视图三维物体重建方法主要基于可微分渲染框架。在二维图像数据训练过程中，输入图像仅能提供单个视角区域的局部形状的监督信息，这导致单视图到整体形状的映射学习面临着困难和挑战，一定程度上限制了重建性能的进一步提升。而随着生成模型Zero 1-to-3（Liu等，2023b）、One-2-3-45（Liu等，2024）和One-2-3-45++（Liu等，2023a）等生成模型的发展，基于当前视图可以零样本学习生成几何纹理一致的物体多视角图像。因此，基于多视角图像生成的物体三维重建为解决单视图监督信息局限问题提供了可能的研究方向。Szymanowicz等人（2024）提出一种单视图三维物体重建方法Splatter Image，通过高斯溅射方式实现不同视角图像渲染和准确三维形状学习。

表5　无监督学习单视图三维物体重建方法定量对比

Table 5　Quantitative comparison results of single view 3D reconstruction methods with unsupervised learning

方法	数据集	评价指标	对比方法	对比结果	实验结果
Tulsiani等人（2022）	PASCAL VOC	IoU ↑	ShapeNet 3D（有监督）	0.510	**0.540**
Chen等人（2020b）	ShapeNet	IoU ↑	NMR（无监督）	0.570	**0.612**
Niemeyer等人（2020）	ShapeNet	CD ↓	Pixel2Mesh（有监督）	**0.210**	0.239
Yao等人（2020）	ShapeNet Core	CD ↓	Pixel2Mesh（有监督）	0.022 1	**0.020 6**
Hu等人（2021）	ShapeNet	IoU ↑	3D Supervised（有监督）	**0.472**	0.465
Yao等人（2021）	ShapeNet	IoU ↑	SoftRas（无监督）	0.469	**0.529**
Monnier等人（2022）	ShapeNet	IoU ↑	DVR（无监督）	0.166	**0.201**
Duggal和Pathak（2022）	ShapeNet	EMD ↓	SoftRas（无监督）	0.776	**0.598**
Huang等人（2023）	Pix3D	CD ↓	Cat3D（无监督）	0.679	**0.618**

注：“↑”表示值越大越优，“↓”表示值越小越优；加粗字体表示各行最优结果。

5 半监督学习单视图三维物体重建

5.1 基于部分三维形状优化的单视图三维物体重建

三维形状数据能够提供精准的三维结构信息，但数据标注成本较高、数据较为稀少。而基于部分三维形状优化的方法，使用三维形状标注数据和未标注二维图像数据联合训练，增强三维形状数据稀缺场景的应用泛化能力。例如Piao等人(2019)提出一种半监督单视图三维人脸重建方法，在合成人脸数据使用三维形状标注数据训练的基础上，真实人脸数据使用重建人脸形状和转换图像人脸边缘的误差来训练优化。在40%以下的三维形状标注数据的情况下，该方法实现了比其他方法更优的性能，表明小样本三维形状和二维图像联合训练的有效性。

半监督方式可以约束人脸形状的重建在合理区间内，但照明等因素制约着人脸重建性能。因此，Gao等人(2020)提出一种改进的半监督人脸重建方法。该算法利用大量不同反照率和照明条件的二维图像数据，以及部分三维形状标注数据进行联合训练，使得模型学习区分输入图像的身份、表情、姿态和光照等因素，从而提高整体的单视图人脸重建性能。

上述方法使用三维形状和二维图像数据联合训练，从而实现半监督学习。而先使用部分三维形状数据预训练，再使用预训练模型指导未标注数据训练是另外一种重要的半监督学习方式。例如，Xing等人(2022)提出了一种半监督框架的单视图三维物体重建方法SSP3D(semi-supervised single-view 3D reconstruction via prototype shape priors)。首先，使用三维形状标注数据训练一个教师网络。然后在未标注数据训练阶段，使用教师网络生成伪三维形状标注数据。为了生成更可靠的三维形状，使用一个原型注意模块进行形状先验融合。最后，对未标注数据生成的三维形状进行对抗性训练，以保证其生成三维形状的合理性。

基于部分三维形状优化的单视图三维物体重建方法，一方面充分挖掘利用三维形状数据的监督信息，另一方面有效分解学习二维图像的关键特征，具有重要的实际场景应用价值。但与直接使用二维图像数据进行无监督学习方法相比，当前方法重建性能更依赖于三维标注数据。在三维标注数据数量减少时，重建性能下降明显，如表6所示。因此，设计小样本三维形状训练优化方案，是当前方法进一步提升性能的潜在研究方向。目前已有三维物体重建工作(Roessle等，2022；Chung等，2024)，在小样本数据重建场景中，引入深度预测任务以获得额外的几何结构信息，再进行三维形状和深度一致性学习以优化重建结果。

表6 不同比例三维标注数据IoU对比(Xing等，2022)

Table 6 IoU comparison with different scales 3D annotated data (Xing et al., 2022)

/%

方法	标注比例1% 标注(301个) 图像(30 596幅)	标注比例5% 标注(1 527个) 图像(30 596幅)	标注比例10% 标注(3 060个) 图像(30 596幅)	标注比例20% 标注(6 125个) 图像(30 596幅)
Supervised (ICCV'19)	41.13	50.32	53.99	58.06
mean-Teacher (NeurIPS'17)	43.36 (↑ 2.23)	51.92 (↑ 1.60)	55.93 (↑ 1.94)	58.88 (↑ 0.82)
mixMatch (NeurIPS'19)	41.77 (↑ 0.64)	51.23 (↑ 0.91)	52.62 (↓ 1.37)	57.43 (↓ 0.63)
FixMatch (NeurIPS'20)	42.44 (↑ 1.31)	51.89 (↑ 1.57)	55.79 (↑ 1.80)	59.63 (↑ 1.57)
SSP3D (Xing等，2022)	**46.99** (↑ **5.86**)	**55.23** (↑ **4.91**)	**58.98** (↑ **4.99**)	**61.64** (↑ **3.58**)

注：加粗字体表示各列最优结果。“↑”和“↓”分别表示相对Supervised方法的IoU提升和降低。

5.2 基于语义分割联合优化的单视图三维物体重建

过少的三维形状会降低上述基于部分三维形状优化方法的重建性能。而基于语义分割任务联合优化的半监督单视图三维物体重建方法，可微分渲染框架结合语义分割任务学习的结构语义信息，增强

对输入图像的三维形状的感知重建性能。Kohli等人(2020)提出了一种多模态的场景表征网络。首先对SRN(scene representation network)(Sitzmann等，2020)进行新视图合成训练，以获得三维感知能力。每个训练对象都由对应的潜在空间向量z_i表示；然后进行半监督训练，冻结SRN的潜在空间向量和网络权重，使用30个人工标注的语义标签对训练一个线性分割分类器，以增强SRN对于目标对象的结构组成感知能力。

多模态的场景表征网络方法在训练完成后，输入测试图像，SRN能够在新视图合成的基础上，实现对应视图的结构语义分割。结合多视角图像和结构语义分割图像，能够重建出目标对象的三维形状以及对应语义结构分割结果。在机器人抓取物体等实际应用场景中，机器人必须同时对对象的三维几何形状、外观和语义结构进行推理，以选择最佳的抓取点。多模态的场景表征网络方法通过二维标签监督学习实现了三维的语义结构分割，具有降低人工标注成本和提升三维感知水平的重要意义。

上述方法(Kohli等，2020)使用语义分割任务提升了重建三维形状的结构感知能力。而Yue等人(2020)提出了一种语义单目深度估计方法，直接使用语义分割任务指导单目深度估计的学习。该方法在输入图像预测初始深度图的基础上，对输入图像进行语义分割并预测包含语义结构信息的深度图，最后融合的深度图中，不同对象之间的边线更加清晰准确，深度估计的性能得到了提升。

基于语义分割联合优化的半监督单视图三维物体重建方法，使用无监督方式进行单视图三维物体重建，同时融合二维语义结构信息以增强性能。与直接标注三维形状数据相比，进行语义分割标注成本更低；但是由于不同重建对象之间的结构复杂多样性，语义分割标注面临着难以明确定义好所有二维图像元素语义类别的问题。因此，解决不同二维图像语义类别的标注训练一致性问题，是进一步提升当前方法性能的潜在方向。Xu等人(2021a，b)通过多视角图像进行无监督的协同分割，以挖掘出多视角图像之间的共有语义信息来构建自监督损失，增强了三维物体重建的稳定性和可靠性。

5.3 基于视角信息学习优化的单视图三维物体重建

无监督学习方法降低了三维形状标注数据成本，但重建过程需要专门对应的视角信息(Liu等，2019)，而在真实应用场景中可能存在视角信息缺失等情况。针对视角信息限制问题，基于视角信息学习优化的方法旨在使用视角信息进行半监督学习，以克服视角信息依赖。

例如Yang等人(2018)提出一种基于视角信息数据的半监督方法。首先使用具有视角标签的同类别数据对，进行无监督单视图三维物体重建训练，以增强不同视角三维重建生成形状的合理性。在有视角信息数据训练完成后，输入未视角标记的图像，通过随机视角渲染的图像进一步训练，以确保缺少视角信息数据的三维重建性能。该方法在仅50%数据具有视角信息的情况下，实现完全使用视角信息进行单视图三维物体重建相当性能。但随着视角信息占比逐渐减少，重建性能下降明显。

针对视角信息依赖问题，Laradji等人(2021)提出了一种改进的SoftRas重建方法。视角判别模块通过大规模带有视角信息的合成数据集进行训练，获得判别两个图像之间的视角相似程度的能力。然后寻找输入图像与合成数据集中最相近的视角，以此视角信息作为输入图像的伪标签进行可微分渲染训练。在合成数据集的实验结果表明，该方式克服了视角依赖，保持了较好的重建性能。

半监督视角学习和近似视角学习等方法降低了对于视角信息的依赖，但主要使用合成数据进行视角信息学习，真实场景视角信息学习难度更高。因此，进一步设计真实场景数据视角信息解耦的重建学习方案，具有重要的研究和应用价值。Shen等人(2023)提出一种从真实场景图像恢复三维形状的方法Anything-3D。该方法首先分割单视图中的目标对象前景，然后利用BLIP(bootstrapping language-image pre-training for unified vision-language understanding and generation)模型(Li等，2022)生成目标对象的纹理文本描述，最后使用扩散模型(Seo等，2024)生成多视角图像来完成物体重建。

5.4 基于预训练模型优化的单视图三维物体重建

上述方法通过挖掘学习图像多样化信息，增强其三维重建学习能力。而基于预训练模型优化方法，对三维标签丰富的物体进行重建预训练，提升对无标签数据三维感知性能。

5.4.1 域自适应单视图三维物体重建

域自适应方法适用于预训练的物体类别与无标

签物体类别相同的场景。例如合成三维模型数据集的样本数量较多且标注准确，但实际应用场景数据和训练的合成数据存在域偏差，如图7所示。

(a) 合成物体　(b) 真实物体

图7　域偏差数据示例(Yang等，2021a)

Fig. 7　Examples of domain bias data(Yang et al. , 2021a)
((a) render objects; (b) real objects)

MarrNet(Wu等，2017)方法首先使用ShapeNet数据集进行预训练，再使用重投影一致性损失对真实物体重建性能进行微调。该方法引入了深度图等2. 5D图像学习作为中间过程，将三维重建过程拆分成2. 5D图像学习和2. 5D图像到3D的映射过程。与直接映射相比，MarrNet将单幅二维图像转换生成多幅不同的2. 5D图像，利用合成物体和真实物体图像对应的深度、边缘轮廓风格统一性，提高模型对域偏差数据的重建性能。

MarrNet方法通过深度图预测等中间过程削弱了输入图像域偏差的影响。但是与输入图像相比，深度图等缺乏颜色与细节纹理信息，这对最终的重建泛化能力造成一定的影响。而域不变特征学习的方式能够在保留输入图像细节特征信息的同时，克服域偏差问题。Yang等人(2021a)提出了一种动态域自适应网络OccNet-DDA(occupancy networks-dynamic domain adaptation)，以提取域不变图像特征进行三维重建。首先将真实图像和合成图像经过同一编码器提取特征，并使用全局域判别器D_g和局部域判别器D_l进行区分。然后缩小真实图像和合成图像的全局、局部特征差距，优化编码器以域不变特征。最后将域不变特征输入到3D MGCN模块中，实现对域偏差的鲁棒泛化三维形状生成。

DDA(dynamic domain adaptation)方法通过域不变特征学习的方式提升了单视图三维物体重建的泛化性能。但在数据类内形状不足时，测试图像的域不变特征往往偏向于训练先验形状。为了解决拓扑变化结构的域自适应难题，Li和Wu(2021)提出了一种动态多分支信息融合网络DmifNet(dynamic multi-branch information fusion network)。以编解码结构作为主干网络，一方面在各个层级的特征层中引入分支以学习多样化结构特征表示；另一方面与MarrNet方法类似，从输入图像中提取与类别解耦的边缘几何和角信息，并设计一个专门的分支学习从边缘几何和角信息重建对应的三维形状。最后动态融合所有分支结果，实现面向细节结构拓扑变化物体重建。

上述基于域自适应的方法，通过中间过程学习、域不变特征学习和多分支信息融合等方式，实现了鲁棒泛化性能。但由于真实场景数据无法提供准确的三维形状数据，域自适应的重建性能存在限制，如图8所示。OccNet-DDA恢复更偏向于图像形状的三维模型，但仍然面临局部形状平滑和细节缺乏等挑战。因此，如何克服真实场景数据形状监督信息不足，是进一步提升域自适应重建性能的重要研究方向。Melas-Kyriazi等人(2023)提出一种基于几何一致性调节的重建方法，训练优化生成形状与输入图像的对齐过程，从而提升真实场景数据形状重建的自适应能力。

输入图像　OccNet (Mescheder等，2019)　OccNet-DA (Yang等，2021a)　OccNet-DDA (Yang等，2021a)　真实模型

图8　局部形状平滑缺乏细节挑战

Fig. 8　Smooth local shape and lack of detail challenge

5. 4. 2　域泛化单视图三维物体重建

域泛化方法适用于预训练的物体类别与无标签物体类别不同的场景，利用合成数据集的三维形状知识，提高对未知数据类别的泛化重建能力。例如

Elich 等人(2022)提出了一种基于 Deep SDF 预训练模型的新型可微分渲染框架方法。该方法第 1 阶段使用三维形状数据集训练 Deep SDF 网络。第 2 阶段，将解码器 Φ 与其他模块组成可微分渲染框架，使用深度图等二维图像进行监督训练。其中，预训练的解码器 Φ 可以提供与物体类别解耦的三维空间点表面距离信息，实现可微分渲染重建泛化性能优化。

Alwala 等人(2022)提出了一种基于有标签的合成形状集合和无标签的真实图像集合数据的多阶段训练方法。首先在第 1 阶段，使用合成数据集渲染生成同一实例的多个视图以及精确相机姿态，以新视图渲染合成任务进行训练获得三维重建能力。然后在第 2 阶段，将真实图像集合输入到预训练模型，以粗糙估计视角信息进行视图渲染，使用图像重建损失对预训练模型进行微调，学习通用类别的三维形状先验知识。最后，将所有类别的预训练模型提取成一个统一的重建模型，实现跨类别的三维重建。该方法可以实现超过 150 个类别的三维重建，但是基于二维图像训练的三维重建性能存在不足，制约了最后通用类别三维重建的质量。

基于三维形状监督学习可以保证三维重建的质量，但面临过拟合训练类别物体形状问题，未知类别重建存在挑战。针对此问题，Lai 等人(2024)提出跨类别知识转移网络 CCKTN(cross-category knowledge transferring network)，在图像到点云映射学习过程中，引入共享知识库存储通用的点云三维结构知识，用以引导点云自动编码器重建未知类别形状。Yang 等人(2023)提出一种域泛化的单视图三维物体重建方法 GenMesh(generalization mesh)，将二维图像到三维形状的映射过程分解为二维图像到三维点云的映射，以及三维点云到网格模型的映射。其中点云到网格的映射过程专注于学习几何变换，与重建物体类别解耦，从而增强了对未知类别物体的重建能力。此外，在图像映射点云阶段，引入局部特征采样策略来捕获不同类别物体间共享的局部几何图形，以增强模型的泛化；在点云映射网格阶段，引入多视图轮廓损失来监督表面生成过程，这提供了额外的正则化，进一步缓解了过拟合问题。

基于预训练模型优化的域泛化方法通过常见类别单视图三维物体重建模型微调的方式，提升了重建未知类别三维形状的性能，实现图像物体的大体形状重建。但面对复杂几何结构的物体，域泛化方法往往只能预测粗糙形状，无法捕捉到细节，如图 9 所示。因此，对物体的个性化形状结构精细重建，是潜在的未来研究方向。例如 Nam 等人(2023)提出一种三维人体重建方法 CycleAdapt。该方法由人体重建网络 HMRNet(human mesh reconstruction network)和人体形状去噪网络 MDNet(human motion denoising network)组成。与上述从图像到三维形状的端到端三维重建方法不同，该方法引入人体形状去噪网络 MDNet，输入重建的带噪声的人体形状，恢复细粒度去噪人体形状，通过三维到三维的去噪训练学习精细形状结构。

图 9　域泛化失败示例(Lai 等，2024)

Fig. 9　Examples of failed domain generalization result (Lai et al. , 2024)

5.5　半监督单视图三维物体重建方法小结

半监督学习单视图三维物体重建方法主要在有监督学习和半监督学习方法的基础上，通过部分三维形状数据或者二维标签数据学习的方式优化了重建性能和适用范围。

表 7 总结了上述半监督学习单视图三维物体重建改进方法的设计思路和优缺点。从各类方法面临的挑战出发，对小样本三维标注数据重建性能不足、语义类别标注训练一致性等问题继续深入研究，在进一步提升半监督学习单视图三维物体重建效果方面具有广阔的研究空间和巨大潜力。

表 8 展示了部分半监督学习方法在 ShapeNet 数据集的测试效果。与 DRC(differentiable ray consistency)(Tulsiani 等，2018)、SoftRas(Liu 等，2019)和 Pix2Vox(Xie 等，2019)等经典 Baseline 方法相比，半

表7　半监督学习单视图三维物体重建方法总结

Table 7　Summary of semi-supervised learning single view 3D reconstruction methods

方法类别	设计思路	优点	缺点
基于部分三维形状优化	三维形状和二维图像数据联合训练	提高稀疏三维标注场景的重建性能	小样本三维标注数据重建性能不足
基于语义分割任务联合优化	语义分割任务学习结构语义信息	增强三维形状与二维图像的一致性	语义类别标注训练一致性问题
基于视角信息学习优化	半监督视角学习和近似视角学习等	减轻对物体视角信息的依赖	真实场景数据视角信息学习问题
基于预训练模型优化	中间过程学习、域不变特征学习等方式 常见类别预训练与重建类别微调	抑制真实场景数据域偏差干扰 提高未知数据类别的泛化重建能力	真实场景数据形状监督信息不足 个性化形状结构精细重建问题

表8　半监督学习方法重建结果

Table 8　Reconstruction results of semi-supervised methods

方法	标注比例	IoU ↑	对比方法	IoU ↑
Yang等人(2018)	50%	0.596 0	DRC(Tulsiani等,2018)	0.573 0
Laradji等人(2021)	8%	0.556 0	SoftRas(Liu等,2019)	0.420 0
Xing等人(2022)	5%	0.552 3	Pix2Vox(Xie等,2019)	0.503 2
Yang等人(2023)	50%	0.815 0	OccNet(Mescheder等,2019)	0.653 1

注:“↑”表示值越大越优。

监督学习方法在标注数据训练的基础上,添加未标注的数据提升了重建性能。其中标注数据的比例问题是当前所有半监督学习方法面临的核心难题。标注数据样本量不足往往会影响预训练模型的重建性能,同时未标注数据缺乏其他有效的监督信息,进而限制了半监督单视图三维物体重建方法的整体性能。因此,如何完成小样本标注数据的半监督单视图三维物体重建,克服标注数据样本少以及未标注数据缺乏监督信息的影响,是未来值得深入研究的关键问题。

三维高斯溅射3DGS(3D Gaussian splatting)(Kerbl等,2023;Lee等,2024),以其基于点云的显式表示的优势,为未标注二维图像学习显式三维结构提供了新的方向。例如Zou等人(2024)提出一种基于点云和三维高斯混合表示学习的单视图物体重建方法。进行物体三维点云学习,基于点云表示学习高斯特征,实现二维图像的快速渲染。该方法通过三维点云形状和二维渲染图像的联合训练优化,实现了已标注类别物体的高质量重建以及未标注类别物体的鲁棒泛化重建。

6　结　语

基于深度学习的单视图三维物体重建方法得到了深入研究与发展,结合本文对单视图三维物体重建方法的回顾以及在当前方法存在的难题,本文展望了未来可能的发展趋势与关键技术,具体如下:

1)真实场景的单视图三维物体重建数据集构建。当前基于深度学习的单视图三维物体重建方法主要使用ShapeNet等合成数据集进行训练与验证,应用到真实场景时受复杂的背景、光照和遮挡等因素影响,可能会出现较大的性能下降。而真实场景数据集中,ikea、PASCAL3D+和ObjectNet3D等数据集主要用于三维物体重建性能测试评估任务。Pix3D数据集可用于三维物体重建训练与验证,但是数据集中物体图像和三维模型数量不足。由于真实场景数据集难以获取,可以研究利用互联网丰富的图像数据资源减少数据采集成本(Li和Snavely,2018;Fang等,2021),开发高效交互式标注工具减少标注成本(Chen等,2020a),构建具有代表性的真

实场景数据集，以便于提升单视图三维物体重建方法在真实场景的应用性能。

2）数据驱动与局部结构先验知识结合的单视图三维物体重建方法。基于深度学习的单视图三维物体重建方法利用采集的二维图像与三维模型数据，通过数据驱动的方式学习三维物体重建，单视图数据本身信息有限问题可能导致未观察区域结果错误，进而影响实际应用效果。而几何模板、几何假设以及几何形状等先验知识在有监督学习、无监督学习和半监督学习方法中得到广泛应用，对于重建结果能够起到约束作用。但主要优化了整体形状重建效果，物体的细节结构恢复存在挑战。因此，提炼总结实际应用场景的局部结构先验知识，研究数据驱动与局部结构先验知识结合的方法，开发从数据学习重建能力过程中加入局部结构先验知识指导的训练范式，能够提高单视图三维物体重建准确可靠性。

3）基于多任务联合优化的小样本半监督学习单视图三维物体重建方法。有监督学习方法在真实场景中面临数据标注专业知识水平要求高时间成本大问题，半监督学习方法在真实场景中面临缺乏准确三维结构监督信息问题。因此，结合小样本标注数据和未标注二维图像数据，开展半监督单视图三维物体重建研究具有重要意义。当前半监督单视图三维物体重建方法使用基于标注数据的预训练模型来提供未标注数据的三维物体重建优化监督信息。而标注数据样本量不足往往会影响预训练模型的重建性能，同时未标注数据缺乏其他有效的监督信息，进而限制了半监督单视图三维物体重建方法的整体性能。而结合实际应用场景中的多任务特点，研究合理的多任务联合训练网络结构和损失函数，以及设计不同的任务的最优组合，有望提升小样本半监督单视图三维物体重建性能；同时在重建任务的基础上，通过多任务学习方式可以获得重建对象更多的有效语义信息，提升智能感知水平（Zheng等，2023）。

4）单视图三维物体重建形状局部细节质量评价指标。物体的局部细节存在一定拓扑变化现象，对基于深度学习的单视图三维物体重建方法的训练学习造成了一定的困难。当前基于深度学习的单视图三维物体重建方法往往将全局特征和局部特征联合训练以提升模型的重建性能。但是在测试评价阶段，使用的CD等指标常用于评判全局的重建质量，忽略了局部细节重建质量评价的问题。因此，研究局部细节质量评价指标，设计合理的测试评判模型性能体系，指导单视图三维物体重建方法的高质量优化，以适应实际应用场景的高精度要求。

5）基于大模型的单视图三维物体重建方法与应用研究。在实际应用场景中重建对象类别多，当前重建算法泛化能力不足，需要设计个性化三维物体重建模型。针对不同对象不同场景的三维物体重建模型，从数据采集标注、模型训练开发到重建性能升级等各个环节的工作需要耗费大量资源。而预训练大模型结合小样本学习微调的方式，展现出实现通用模型的可能和突破当前深度学习方法泛化能力和精度的潜力（Xu等，2024）。研究基于大模型的单视图三维物体重建方法，开发数据处理、模型训练、部署应用、迭代拓展的全流程应用方案，可以推动单视图三维物体重建方法开发应用的“工业化”升级。

参考文献（References）

Alsadoon A, AlSallami N, Rashid T A, Gosper J J, Prasad P W C and Haddad S. 2024. RETRACTED ARTICLE: DVT: a recent review and a taxonomy for oral and maxillofacial visualization and tracking based augmented reality: image guided surgery. Multimedia Tools and Applications, 83(1): 685-729 [DOI: 10.1007/s11042-023-15581-w]

Alwala K V, Gupta A and Tulsiani S. 2022. Pretrain, self-train, distill: a simple recipe for supersizing 3D reconstruction//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 3763-3772 [DOI: 10.1109/CVPR52688.2022.00375]

Aygün M and Mac Aodha O. 2024. SAOR: single-view articulated object reconstruction//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 10382-10391 [DOI: 10.1109/CVPR52733.2024.00988]

Bai J, Meng Q L, Xu H, Fan Y F and Yang Z Y. 2022. ST-Rec3D: a structure and target-aware 3D reconstruction. Journal of Graphics, 43(3): 469-477 (白静, 孟庆亮, 徐昊, 范有福, 杨瞻源. 2022. ST-Rec3D: 基于结构和目标感知的三维重建. 图学学报, 43(3): 469-477) [DOI: 10.11996/JG.j.2095-302X.2022030469]

Bednarik J, Fua P and Salzmann M. 2018. Learning to reconstruct texture-less deformable surfaces from a single view//Proceedings of 2018 International Conference on 3D Vision (3DV). Verona, Italy: IEEE: 606-615 [DOI: 10.1109/3DV.2018.00075]

Chabra R, Lenssen J E, Ilg E, Schmidt T, Straub J, Lovegrove S and Newcombe R. 2020. Deep local shapes: learning local SDF priors for detailed 3D reconstruction//Proceedings of the 16th European Conference on Computer Vision (ECCV). Glasgow, UK: Springer:

608-625［DOI：10.1007/978-3-030-58526-6_36］

Chang A X, Funkhouser T, Guibas L, Hanrahan P, Huang Q X, Li Z M, Savarese S, Savva M, Song S R, Su H, Xiao J X, Yi L and Yu F. 2015. ShapeNet：an information-rich 3D model repository［EB/OL］.［2024-07-11］. https://arxiv.org/pdf/1512.03012.pdf

Charles R Q, Su H, Kaichun M and Guibas L J. 2017. PointNet：deep learning on point sets for 3D classification and segmentation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition（CVPR）. Honolulu, USA：IEEE：77-85［DOI：10.1109/CVPR.2017.16］

Chen R Y, Yin M H, Shen J W and Ma W. 2024. Recon3D：high quality 3D reconstruction from a single image using generated back-view explicit priors//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops（CVPRW）. Seattle, USA：IEEE：2802-2811［DOI：10.1109/CVPRW63382.2024.00286］

Chen W F, Qian S Y, Fan D, Kojima N, Hamilton M and Deng J. 2020a. OASIS：a large-scale dataset for single image 3D in the wild//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. Seattle, USA：IEEE：676-685［DOI：10.1109/CVPR42600.2020.00076］

Chen W Z, Gao J, Ling H, Smith E J, Lehtinen J, Jacobson A and Fidler S. 2020b. Learning to predict 3D objects with an interpolation-based differentiable renderer//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada：Curran Associates Inc.：#862

Chen Z Q and Zhang H. 2019. Learning implicit fields for generative shape modeling//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. Long Beach, USA：IEEE：5932-5941［DOI：10.1109/CVPR.2019.00609］

Choy C B, Xu D F, Gwak J Y, Chen K and Savarese S. 2016. 3D-R2N2：a unified approach for single and multi-view 3D object reconstruction//Proceedings of the 14th European Conference on Computer Vision（ECCV）. Amsterdam, the Netherlands：Springer：628-644［DOI：10.1007/978-3-319-46484-8_38］

Chung J, Oh J and Lee K M. 2024. Depth-regularized optimization for 3D Gaussian splatting in few-shot images//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops（CVPRW）. Seattle, USA：IEEE：811-820［DOI：10.1109/CVPRW63382.2024.00086］

Dai A G L, Chang A X, Savva M, Halber M, Funkhouser T and Nießner M. 2017. ScanNet：richly-annotated 3D reconstructions of indoor scenes//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition（CVPR）. Honolulu, USA：IEEE：2432-2443［DOI：10.1109/CVPR.2017.261］

Duggal S and Pathak D. 2022. Topologically-aware deformation fields for single-view 3D reconstruction//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. New Orleans, USA：IEEE：1526-1536［DOI：10.1109/CVPR52688.2022.00159］

Dundar A, Gao J, Tao A and Catanzaro B. 2023. Fine detailed texture learning for 3D meshes with generative models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(12)：14563-14574［DOI：10.1109/TPAMI.2023.3319429］

Elich C, Oswald M R, Pollefeys M and Stuckler J. 2022. Semi-supervised learning of multi-object 3D scene representations［EB/OL］.［2024-07-11］. https://arxiv.org/pdf/2010.04030.pdf

Fahim G, Amin K and Zarif S. 2021. Single-view 3D reconstruction：a survey of deep learning methods. Computers and Graphics, 94：164-190［DOI：10.1016/j.cag.2020.12.004］

Fang Q, Shuai Q, Dong J T, Bao H J and Zhou X W. 2021. Reconstructing 3D human pose by watching humans in the mirror//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. Nashville, USA：IEEE：12809-12818［DOI：10.1109/CVPR46437.2021.01262］

Gan Y S, Chen W H, Yau W C, Zou Z Y, Liong S T and Wang S Y. 2023. 3D SOC-net：deep 3D reconstruction network based on self-organizing clustering mapping. Expert Systems with Applications, 213：#119209［DOI：10.1016/j.eswa.2022.119209］

Gao Z P, Zhang J Y, Guo Y D, Ma C, Zhai G T and Yang X K. 2020. Semi-supervised 3D face representation learning from unconstrained photo collections//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops（CVPRW）. Seattle, USA：IEEE：1426-1435［DOI：10.1109/CVPRW50498.2020.00182］

Geiger A, Lenz P, Stiller C and Urtasun R. 2013. Vision meets robotics：the KITTI dataset. The International Journal of Robotics Research, 32(11)：1231-1237［DOI：10.1177/0278364913491297］

Groueix T, Fisher M, Kim V G, Russell B C and Aubry M. 2018. A papier-mache approach to learning 3D surface generation//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA：IEEE：216-224［DOI：10.1109/CVPR.2018.00030］

Häne C, Tulsiani S and Malik J. 2017. Hierarchical surface prediction for 3D object reconstruction//Proceedings of 2017 International Conference on 3D Vision（3DV）. Qingdao, China：IEEE：412-420［DOI：10.1109/3DV.2017.00054］

He R Q, Jiang C H, Liu J S, Cao T and Gui W H. 2022. Blast furnace material particle size detection based on concave-convexity and density-guided point cloud segmentation//Proceedings of the China Automation Congress（CAC）. Xiamen, China：Chinese Society of Automation：141-146（何瑞清, 蒋朝辉, 刘金狮, 曹婷, 桂卫华. 2022. 基于凹凸性与密度引导点云分割的高炉炉料粒度检测//2022 中国自动化大会论文集. 厦门：中国自动化学会：141-146）［DOI：10.26914/c.cnkihy.2022.053812］

Ho L N, Tran A T, Phung Q and Hoai M. 2021. Toward realistic single-view 3D object reconstruction with unsupervised learning from multiple images//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision（ICCV）. Montreal, Canada：IEEE：12580-12590［DOI：10.1109/ICCV48922.2021.01237］

Hsu C H, Chiu C T and Kuan C Y. 2020. Fast single-view 3D object reconstruction with fine details through dilated downsample and multi-path upsample deep neural network//Proceedings of 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain: IEEE: 1653-1657 [DOI: 10.1109/ICASSP40776.2020.9053607]

Hu T, Wang L W, Xu X G, Liu S and Jia J Y. 2021. Self-supervised 3D mesh reconstruction from single images//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 5998-6007 [DOI: 10.1109/CVPR46437.2021.00594]

Huang Z X, Jampani V, Thai A, Li Y Z, Stojanov S and Rehg J M. 2023. ShapeClipper: scalable 3D shape learning from single-view images via geometric and CLIP-based consistency//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 12912-12922 [DOI: 10.1109/CVPR52729.2023.01241]

Huang Z X, Stojanov S, Thai A, Jampani V and Rehg J M. 2022b. Planes vs. chairs: category-guided 3D shape learning without any 3D cues//Proceedings of the 17th European Conference on Computer Vision (ECCV). Tel Aviv, Israel: Springer: 727-744 [DOI: 10.1007/978-3-031-19769-7_42]

Jin J C, Xu H Q, Ji P L and Leng B. 2022. IMC-NET: learning implicit field with corner attention network for 3D shape reconstruction//Proceedings of 2022 IEEE International Conference on Image Processing (ICIP). Bordeaux, France: IEEE: 1591-1595 [DOI: 10.1109/ICIP46576.2022.9897709]

Kato H and Harada T. 2019. Learning view priors for single-view 3D reconstruction//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 9770-9779 [DOI: 10.1109/CVPR.2019.01001]

Kato H, Ushiku Y and Harada T. 2018. Neural 3D mesh renderer//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3907-3916 [DOI: 10.1109/CVPR.2018.00411]

Kerbl B, Kopanas G, Leimkuehler T and Drettakis G. 2023. 3D Gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics (TOG), 42(4): #139 [DOI: 10.1145/3592433]

Kim G and Chun S Y. 2023. DATID-3D: diversity-preserved domain adaptation using text-to-image diffusion for 3D generative model//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 14203-14213 [DOI: 10.1109/CVPR52729.2023.01365]

Kim Y M, Cho J and Ahn S C. 2016. 3D modeling from photos given topological information. IEEE Transactions on Visualization and Computer Graphics, 22(9): 2070-2081 [DOI: 10.1109/TVCG.2015.2505307]

Kirillov A, Mintun E, Ravi N, Mao H Z, Rolland C, Gustafson L, Xiao T T, Whitehead S, Berg A C, Lo W Y, Dollár P and Girshick R. 2023. Segment anything [EB/OL]. [2024-07-11]. https://arxiv.org/pdf/2304.02643.pdf

Koch S, Matveev A, Jiang Z S, Williams F, Artemov A, Burnaev E, Alexa M, Zorin D and Panozzo D. 2019. ABC: a big CAD model dataset for geometric deep learning//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 9593-9603 [DOI: 10.1109/CVPR.2019.00983]

Kohli A P S, Sitzmann V and Wetzstein G. 2020. Semantic implicit neural scene representations with semi-supervised training//Proceedings of 2020 International Conference on 3D Vision (3DV). Fukuoka, Japan: IEEE: 423-433 [DOI: 10.1109/3DV50981.2020.00052]

Koneputugodage C H, Ben-Shabat Y and Gould S. 2023. Octree guided unoriented surface reconstruction//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 16717-16726 [DOI: 10.1109/CVPR52729.2023.01604]

Kuo Y L, Ko W J, Chiu C Y and Chiu W C. 2022. Improving single-view mesh reconstruction for unseen categories via primitive-based representation and mesh augmentation//Proceedings of 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Kyoto, Japan: IEEE: 2001-2008 [DOI: 10.1109/IROS47612.2022.9982024]

Lai L, Chen J and Wu Q Y. 2024. Zero-shot single-view point cloud reconstruction via cross-category knowledge transferring. IEEE Transactions on Multimedia, 26: 1448-1459 [DOI: 10.1109/TMM.2023.3282467]

Laradji I, Rodríguez P, Vazquez D and Nowrouzezahrai D. 2021. SSR: semi-supervised soft rasterizer for single-view 2D to 3D reconstruction//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal, Canada: IEEE: 1427-1436 [DOI: 10.1109/ICCVW54120.2021.00164]

Lee J C, Rho D, Sun X Y, Ko J H and Park E. 2024. Compact 3D Gaussian representation for radiance field//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 21719-21728 [DOI: 10.1109/CVPR52733.2024.02052]

Li J N, Li D X, Xiong C M and Hoi S. 2022. BLIP: bootstrapping language-image pre-training for unified vision-language understanding and generation//Proceedings of the 39th International Conference on Machine Learning. Baltimore, USA: PMLR: 12889-12901

Li L and Wu S P. 2021. DmifNet: 3D shape reconstruction based on dynamic multi-branch information fusion//Proceedings of the 25th International Conference on Pattern Recognition (ICPR). Milan, Italy: IEEE: 7219-7225 [DOI: 10.1109/ICPR48806.2021.9411960]

Li L, Xu H and Wu S P. 2022. Fuzzy probability points reasoning for 3D reconstruction via deep deterministic policy gradient. Acta Automatica Sinica, 48(4): 1105-1118 (李雷, 徐浩, 吴素萍. 2022. 基于DDPG的三维重建模糊概率点推理. 自动化学报, 48(4):

1105-1118)[DOI: 10.16383/j.aas.c200543]

Li M Y and Zhang H. 2021. D2IM-Net: learning detail disentangled implicit fields from single images//Proceedings of 2021 IEEE/CVF IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 10241-10250 [DOI: 10.1109/CVPR46437.2021.01011]

Li X and Kuang P. 2021. 3D-VRVT: 3D voxel reconstruction from a single image with vision transformer//Proceedings of 2021 International Conference on Culture-oriented Science and Technology (ICCST). Beijing, China: IEEE: 343-348 [DOI: 10.1109/ICCST53801.2021.00078]

Li Z Q and Snavely N. 2018. MegaDepth: learning single-view depth prediction from internet photos//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE: 2041-2050 [DOI: 10.1109/CVPR.2018.00218]

Liang C Y, Tang H M, Xi J R and Liu X. 2023. Improved single-view surface reconstruction based on sparse feature. Application Research of Computers, 40(3): 925-931, 937 (梁春阳, 唐红梅, 席建锐, 刘鑫. 2023. 基于稀疏特征改进的单视图表面重建. 计算机应用研究, 40(3): 925-931), 937) [DOI: 10.19734/j.issn.1001-3695.2022.06.0320]

Lim J J, Pirsiavash H and Torralba A. 2013. Parsing IKEA objects: fine pose estimation//Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney, Australia: IEEE: 2992-2999 [DOI: 10.1109/ICCV.2013.372]

Liu M H, Shi R X, Chen L H, Zhang Z Y, Xu C, Wei X Y, Chen H S, Zeng C, Gu J Y and Su H. 2023a. One-2-3-45++: fast single image to 3D objects with consistent multi-view generation and 3D diffusion//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 10072-10083 [DOI: 10.1109/CVPR52733.2024.00960]

Liu M H, Xu C, Jin H A, Chen L H, T M V, Xu Z X and Su H. 2024. One-2-3-45: any single image to 3D mesh in 45 seconds without per-shape optimization//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: #976

Liu R S, Wu R D, Van Hoorick B, Tokmakov P, Zakharov S and Vondrick C. 2023b. Zero-1-to-3: zero-shot one image to 3D object//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 9264-9275 [DOI: 10.1109/ICCV51070.2023.00853]

Liu S C, Chen W K, Li T Y and Li H. 2019. Soft rasterizer: a differentiable renderer for image-based 3D reasoning//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 7707-7716 [DOI: 10.1109/ICCV.2019.00780

Liu X N, Chen C Y, Hu X J and Yu H Y. 2024. Virtual viewpoint image synthesis using neural radiance fields with depth information supervision. Journal of Image and Graphics, 29(7): 2035-2045 (刘晓楠, 陈纯毅, 胡小娟, 于海洋. 2024. 带深度信息监督的神经辐射场虚拟视点画面合成. 中国图象图形学报, 29(7): 2035-2045) [DOI: 10.11834/jig.221188]

Liu Z, Tang S H and Wang Y B. 2015. Single view reconstruction based on marker and geometric constraints//Proceedings of the 8th International Congress on Image and Signal Processing (CISP). Shenyang, China: IEEE: 1058-1062 [DOI: 10.1109/CISP.2015.7408036]

Loper M M and Black M J. 2014. OpenDR: an approximate differentiable renderer//Proceedings of the 13th European Conference on Computer Vision (ECCV). Zurich, Switzerland: Springer: 154-169 [DOI: 10.1007/978-3-319-10584-0_11]

Lu J W, Guo C, Dai X Y, Miao Q H, Wang X X, Yang J and Wang F Y. 2023. The ChatGPT after: opportunities and challenges of very large scale pre-trained models. Acta Automatica Sinica, 49(4): 705-717 (卢经纬, 郭超, 戴星原, 缪青海, 王兴霞, 杨静, 王飞跃. 2023. 问答ChatGPT之后: 超大预训练模型的机遇和挑战. 自动化学报, 49(4): 705-717) [DOI: 10.16383/j.aas.c230107]

Mei J, Yu J X, Romain S, Rose C, Magrane K, LeeSon G and Hwang J N. 2022. Unsupervised severely deformed mesh reconstruction (DMR) from a single-view image for longline fishing//Proceedings of 2022 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). Taipei City, China: IEEE: 1-6 [DOI: 10.1109/ICMEW56448.2022.9859312]

Melas-Kyriazi L, Rupprecht C and Vedaldi A. 2023. PC^2: projection-conditioned point cloud diffusion for single-image 3D reconstruction//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 12923-12932 [DOI: 10.1109/CVPR52729.2023.01242]

Mescheder L, Oechsle M, Niemeyer M, Nowozin S and Geiger A. 2019. Occupancy networks: learning 3D reconstruction in function space//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 4455-4465 [DOI: 10.1109/CVPR.2019.00459]

Michalkiewicz M, Parisot S, Tsogkas S, Baktashmotlagh M, Eriksson A and Belilovsky E. 2020. Few-shot single-view 3-D object reconstruction with compositional priors//Proceedings of the 16th European Conference on Computer Vision (ECCV). Glasgow, UK: Springer: 614-630 [DOI: 10.1007/978-3-030-58595-2_37]

Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2020. NeRF: representing scenes as neural radiance fields for view synthesis//Proceedings of the 16th European Conference on Computer Vision (ECCV). Glasgow, UK: Springer: 405-421 [DOI: 10.1007/978-3-030-58452-8_24]

Mo K C, Zhu S L, Chang A X, Yi L, Tripathi S, Guibas L J and Su H. 2019. PartNet: a large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 909-918 [DOI: 10.1109/CVPR.2019.00100]

Monnier T, Fisher M, Efros A A and Aubry M. 2022. Share with thy neighbors: single-view reconstruction by cross-instance consistency//Proceedings of the 17th European Conference on Computer Vision (ECCV). Tel Aviv, Israel: Springer: 285-303 [DOI: 10.1007/978-3-031-19769-7_17]

Nam H, Jung D S, Oh Y and Lee K M. 2023. Cyclic test-time adaptation on monocular video for 3D human mesh reconstruction//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 14783-14793 [DOI: 10.1109/ICCV51070.2023.01362]

Nie W Z, Jiao C Q, Chang R H, Qu L and Liu A A. 2023a. CPG3D: cross-modal priors guided 3D object reconstruction. IEEE Transactions on Multimedia, 25: 9383-9396 [DOI: 10.1109/TMM.2023.3251697]

Nie Y Y, Dai A G L, Han X G and NieBner M. 2023b. Learning 3D scene priors with 2D supervision//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 792-802 [DOI: 10.1109/CVPR52729.2023.00083]

Niemeyer M, Mescheder L, Oechsle M and Geiger A. 2020. Differentiable volumetric rendering: learning implicit 3D representations without 3D supervision//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 3501-3512 [DOI: 10.1109/CVPR42600.2020.00356]

Park J J, Florence P, Straub J, Newcombe R and Lovegrove S. 2019. DeepSDF: learning continuous signed distance functions for shape representation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 165-174 [DOI: 10.1109/CVPR.2019.00025]

Park K, Kim S and Sohn K. 2020. High-precision depth estimation using uncalibrated LiDAR and stereo fusion. IEEE Transactions on Intelligent Transportation Systems, 21(1): 321-335 [DOI: 10.1109/TITS.2019.2891788]

Park K, Sinha U, Barron J T, Bouaziz S, Goldman D B, Seitz S M and Martin-Brualla R. 2021. Nerfies: deformable neural radiance fields//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 5845-5854 [DOI: 10.1109/ICCV48922.2021.00581]

Peng S Y, Niemeyer M, Mescheder L, Pollefeys M and Geiger A. 2020. Convolutional occupancy networks//Proceedings of the 16th European Conference on Computer Vision (ECCV). Glasgow, UK: Springer: 523-540 [DOI: 10.1007/978-3-030-58580-8_31]

Piao J, Qian C and Li H S. 2019. Semi-supervised monocular 3D face reconstruction with end-to-end shape-preserved domain transfer//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 9397-9406 [DOI: 10.1109/ICCV.2019.00949]

Pinheiro P O, Rostamzadeh N and Ahn S. 2019. Domain-adaptive single-view 3D reconstruction//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 7637-7646 [DOI: 10.1109/ICCV.2019.00773]

Pumarola A, Agudo A, Porzi L, Sanfeliu A, Lepetit V and Moreno-Noguer F. 2018. Geometry-aware network for non-rigid shape prediction from a single view//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE: 4681-4690 [DOI: 10.1109/CVPR.2018.00492]

Rebain D, Matthews M, Yi K M, Lagun D and Tagliasacchi A. 2022. LOLNeRF: learn from one look//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 1548-1557 [DOI: 10.1109/CVPR52688.2022.00161]

Riegler G, Ulusoy A O and Geiger A. 2017. OctNet: learning deep 3D representations at high resolutions//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 6620-6629 [DOI: 10.1109/CVPR.2017.701]

Roessle B, Barron J T, Mildenhall B, Srinivasan P P and Nießner M. 2022. Dense depth priors for neural radiance fields from sparse input views//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 12882-12891 [DOI: 10.1109/CVPR52688.2022.01255]

Samavati T and Soryani M. 2023. Deep learning-based 3D reconstruction: a survey. Artificial Intelligence Review, 56(9): 9175-9219 [DOI: 10.1007/s10462-023-10399-2]

Seo J, Jang W, Kwak M S, Kim H, Ko J, Kim J, Kim J H, Lee J and Kim S. 2024. Let 2D diffusion model know 3D-consistency for robust text-to-3D generation//Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: ICLR

Shen Q H, Yang X Y and Wang X C. 2023. Anything-3D: towards single-view anything reconstruction in the wild [EB/OL]. [2023-04-19]. https://arxiv.org/pdf/2304.10261.pdf

Shen W C, Ma T S, Wu Y W and Jia Y D. 2021. Component-aware high-resolution 3D object reconstruction. Journal of Computer-Aided Design and Computer Graphics, 33(12): 1887-1898 (沈伟超, 马天朔, 武玉伟, 贾云得. 2021. 组件感知的高分辨率三维物体重建方法. 计算机辅助设计与图形学学报, 33(12): 1887-1898) [DOI: 10.3724/SP.J.1089.2021.18805]

Silberman N, Hoiem D, Kohli P and Fergus R. 2012. Indoor segmentation and support inference from RGBD images//Proceedings of the 12th European Conference on Computer Vision (ECCV). Florence, Italy: Springer: 746-760 [DOI: 10.1007/978-3-642-33715-4_54]

Sitzmann V, Zollhofer M and Wetzstein G. 2020. Scene representation networks: continuous 3D-structure-aware neural scene representations//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #101

Song S R, Lichtenberg S P and Xiao J X. 2015. SUN RGB-D: a RGB-D scene understanding benchmark suite//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: 567-576 [DOI: 10.1109/CVPR. 2015. 7298655]

Sun X Y, Wu J J, Zhang X M, Zhang Z T, Zhang C K, Xue T F, Tenenbaum J B and Freeman W T. 2018a. Pix3D: dataset and methods for single-image 3D shape modeling//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE: 2974-2983 [DOI: 10.1109/CVPR.2018.00314]

Sun Y B, Liu Z W, Wang Y and Sarma S E. 2018b. Im2Avatar: colorful 3D reconstruction from a single image [EB/OL]. [2024-07-11]. https://arxiv.org/pdf/1804.06375.pdf

Szymanowicz S, Rupprecht C and Vedaldi A. 2024. Splatter image: ultra-fast single-view 3D reconstruction//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 10208-10217 [DOI: 10.1109/CVPR52733.2024.00972]

Tatarchenko M, Dosovitskiy A and Brox T. 2017. Octree generating networks: efficient convolutional architectures for high-resolution 3D outputs//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 2107-2115 [DOI: 10.1109/ICCV.2017.230]

Tulsiani S, Efros A A and Malik J. 2018. Multi-view consistency as supervisory signal for learning shape and pose prediction//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE: 2897-2905 [DOI: 10.1109/CVPR.2018.00306]

Tulsiani S, Zhou T H, Efros A A and Malik J. 2022. Multi-view supervision for single-view reconstruction via differentiable ray consistency. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12): 8754-8765 [DOI: 10.1109/TPAMI.2019.2898859]

Victoria C, Torres F, Garduño E, Cosío F A and Gastelum-Strozzi A. 2023. Real-time 3D ultrasound reconstruction using octrees. IEEE Access, 11: 78970-78983 [DOI: 10.1109/ACCESS.2023.3298887]

Wan J H, Liu X P, Chen L L, Ao S, Zhang P and Guo Y L. 2024. Geometric attribute-guided 3D semantic instance reconstruction. Journal of Image and Graphics, 29(1): 218-230 (万骏辉, 刘心溥, 陈莉丽, 敖晟, 张鹏, 郭裕兰. 2024. 几何属性引导的三维语义实例重建. 中国图象图形学报, 29(1): 218-230) [DOI: 10.11834/jig.230106]

Wang J R and Fang Z Y. 2020. GSIR: generalizable 3D shape interpretation and reconstruction//Proceedings of the 16th European Conference on Computer Vision (ECCV). Glasgow, UK: Springer: 498-514 [DOI: 10.1007/978-3-030-58601-0_30]

Wang K Q, Kemao Q, Di J L and Zhao J L. 2022. Deep learning spatial phase unwrapping: a comparative review. Advanced Photonics Nexus, 1(1): #014001 [DOI: 10.1117/1.APN.1.1.014001]

Wang N, Hu X Y, Zhu F and Tang J. 2020. Single-view 3D Reconstruction Algorithm Based on View-aware. Journal of Electronics and Information Technology, 42(12): 3053-3060 (王年, 胡旭阳, 朱凡, 唐俊. 2020. 基于视图感知的单视图三维重建算法. 电子与信息学报, 2020, 42(12): 3053-3060) [DOI: 10.11999/JEIT190986]

Wang N Y, Zhang Y D, Li Z W, Fu Y W, Liu W and Jiang Y G. 2018a. Pixel2Mesh: generating 3D mesh models from single RGB images//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 55-71 [DOI: 10.1007/978-3-030-01252-6_4]

Wang P S, Liu Y, Guo Y X, Sun C Y and Tong X. 2017. O-CNN: octree-based convolutional neural networks for 3D shape analysis. ACM Transactions on Graphics (TOG), 2017, 36(4): #72 [DOI: 10.1145/3072959.3073608]

Wang P S, Sun C Y, Liu Y and Tong X. 2018b. Adaptive O-CNN: a patch-based deep representation of 3D shapes. ACM Transactions on Graphics (TOG), 2018, 37(6): #217 [DOI: 10.1145/3272127.3275050]

Wang W Y, Xu Q G, Ceylan D, Mech R and Neumann U. 2020. DISN: deep implicit surface network for high-quality single-view 3D reconstruction//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #45

Wang Y A, He X Y, Peng S D, Lin H T, Bao H J and Zhou X W. 2023. AutoRecon: automated 3D object discovery and reconstruction//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 21382-21391 [DOI: 10.1109/CVPR52729.2023.02048]

Wang Y Z, Lira W, Wang W Q, Mahdavi-Amiri A and Zhang H. 2024. Slice3D: multi-slice, occlusion-revealing, single view 3D reconstruction//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 9881-9891 [DOI: 10.1109/CVPR52733.2024.00943]

Wimbauer F, Yang N, Rupprecht C and Cremers D. 2023. Behind the scenes: density fields for single view reconstruction//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 9076-9086 [DOI: 10.1109/CVPR52729.2023.00876]

Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]

Wu F and He S X. 2023. Single-view 3D reconstruction based on attentional mechanisms. Laser Journal, 44(1): 109-114 (吴繁, 贺赛先. 2023. 基于注意力机制的单视图三维重建. 激光杂志, 44(1): 109-114) [DOI: 10.14016/j.cnki.jgzz.2023.01.109]

Wu J J, Wang Y F, Xue T F, Sun X Y, Freeman W T and Tenenbaum J B. 2017. MarrNet: 3D shape reconstruction via 2.5D sketches//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates

Inc.: 540-550

Wu J J, Zhang C K, Zhang X M, Zhang Z T, Freeman W T and Tenenbaum J B. 2018. Learning shape priors for single-view 3D completion and reconstruction//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 673-691 [DOI: 10.1007/978-3-030-01252-6_40]

Wu S Z, Li R N, Jakab T, Rupprecht C and Vedaldi A. 2023a. MagicPony: learning articulated 3D animals in the wild//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 8792-8802 [DOI: 10.1109/CVPR52729.2023.00849]

Wu S Z, Rupprecht C and Vedaldi A. 2023b. Unsupervised learning of probably symmetric deformable 3D objects from images in the wild (Invited Paper). IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4): 5268-5281 [DOI: 10.1109/TPAMI.2021.3076536]

Wu Z R, Song S R, Khosla A, Yu F, Zhang L G, Tang X O and Xiao J X. 2015. 3D ShapeNets: a deep representation for volumetric shapes//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE: 1912-1920 [DOI: 10.1109/CVPR.2015.7298801]

Xiang Y, Kim W, Chen W, Ji J W, Choy C, Su H, Mottaghi R, Guibas L and Savarese S. 2016. ObjectNet3D: a large scale database for 3D object recognition//Proceedings of the 14th European Conference on Computer Vision (ECCV). Amsterdam, The Netherlands: Springer: 160-176 [DOI: 10.1007/978-3-319-46484-8_10]

Xiang Y, Mottaghi R and Savarese S. 2014. Beyond PASCAL: a benchmark for 3D object detection in the wild//Proceedings of 2014 IEEE Winter Conference on Applications of Computer vision (WACV). Steamboat Springs, USA: IEEE: 75-82 [DOI: 10.1109/WACV.2014.6836101]

Xie H Z, Yao H X, Sun X S, Zhou S C and Zhang S P. 2019. Pix2Vox: context-aware 3D reconstruction from single and multi-view images//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 2690-2698 [DOI: 10.1109/ICCV.2019.00278]

Xing Z, Li H D, Wu Z X and Jiang Y G. 2022. Semi-supervised single-view 3D reconstruction via prototype shape priors//Proceedings of the 17th European Conference on Computer Vision (ECCV). Tel Aviv, Israel: Springer: 535-551 [DOI: 10.1007/978-3-031-19769-7_31]

Xu H B, Zhou Z P, Qiao Y, Kang W X and Wu Q X. 2021a. Self-supervised multi-view stereo via effective co-segmentation and data-augmentation//Proceedings of the 35th Conference on Artificial Intelligence (AAAI). AAAI: 3030-3038 [DOI: 10.1609/aaai.v35i4.16411]

Xu H B, Zhou Z P, Wang Y L, Kang W X, Sun B G, Li H and Qiao Y. 2021b. Digging into uncertainty in self-supervised multi-view stereo//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 6058-6067 [DOI: 10.1109/ICCV48922.2021.00602]

Xu J L, Cheng W H, Gao Y M, Wang X T, Gao S H and Shan Y. 2024. InstantMesh: efficient 3D mesh generation from a single image with sparse-view large reconstruction models [EB/OL]. [2024-04-14]. https://arxiv.org/pdf/2404.07191.pdf

Yang C, Xie H S, Tian H H and Yu Y L. 2021a. Dynamic domain adaptation for single-view 3D reconstruction//Proceedings of 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Prague, Czech Republic: IEEE: 3563-3570 [DOI: 10.1109/IROS51168.2021.9636343]

Yang G D, Cui Y, Belongie S and Hariharan B. 2018. Learning single-view 3D reconstruction with limited pose supervision//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 90-105 [DOI: 10.1007/978-3-030-01267-0_6]

Yang H, Chen R, An S P, Wei H and Zhang H. 2023. The growth of image-related three dimensional reconstruction techniques in deep learning-driven era: a critical summary. Journal of image and Graphics, 28(8): 2396-2409 (杨航, 陈瑞, 安仕鹏, 魏豪, 张衡. 2023. 深度学习背景下的图像三维重建技术进展综述. 中国图象图形学报, 28(8): 2396-2409) [DOI: 10.11834/jig.220376]

Yang S, Xu M, Xie H Z, Perry S and Xia J H. 2021b. Single-view 3D object reconstruction from shape priors in memory//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 3151-3160 [DOI: 10.1109/CVPR46437.2021.00317]

Yang X H, Lin G S and Zhou L P. 2023. Single-view 3D mesh reconstruction for seen and unseen categories. IEEE Transactions on Image Processing, 32: 3746-3758 [DOI: 10.1109/TIP.2023.3279661]

Yao C H, Hung W C, Jampani V and Yang M H. 2021. Discovering 3D parts from image collections//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 12961-12970 [DOI: 10.1109/ICCV48922.2021.01274]

Yao Y, Schertler N, Rosales E, Rhodin H, Sigal L and Sheffer A. 2020. Front2Back: single view 3D shape reconstruction via front to back prediction//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 528-537 [DOI: 10.1109/CVPR42600.2020.00061]

Yue M, Fu G Y, Wu M and Wang H Q. 2020. Semi-supervised monocular depth estimation based on semantic supervision. Journal of Intelligent and Robotic Systems, 100(2): 455-463 [DOI: 10.1007/s10846-020-01205-0]

Zamir A R, Sax A, Cheerla N, Suri R, Cao Z J, Malik J and Guibas L J. 2020. Robust learning through cross-task consistency//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 11194-11203 [DOI: 10.1109/CVPR42600.2020.01121]

Zhang H, Zhang Q, Li Y X and Shao S Y. 2021. Research on 3D model

reconstruction based on deep learning. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 33(2): 289-295 (张豪, 张强, 李勇祥, 邵思羽. 2021. 基于深度学习的三维模型重构研究. 重庆邮电大学学报(自然科学版), 33(2): 289-295)

Zhang T, Bian N J, Liu Q Y and Du Y. 2024. 3D super-resolution reconstruction of porous media based on GANs and CBAMs. Stochastic Environmental Research and Risk Assessment, 38(4): 1475-1504 [DOI: 10.1007/s00477-023-02639-2]

Zhang X M, Zhang Z T, Zhang C K, Tenenbaum J B, Freeman W T and Wu J J. 2018. Learning to reconstruct shapes from unseen classes//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates Inc.: 2263-2274

Zhang Y M, Tosi F, Mattoccia S and Poggi M. 2023. GO-SLAM: global optimization for consistent 3D instant reconstruction//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 3704-3714 [DOI: 10.1109/ICCV51070.2023.00345]

Zhao F, Wang W H, Liao S C and Shao L. 2021. Learning anchored unsigned distance functions with gradient direction alignment for single-view garment reconstruction//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 12654-12663 [DOI: 10.1109/ICCV48922.2021.01244]

Zheng S H, Bao Z P, Hebert M and Wang Y X. 2023. Multi-task view synthesis with neural radiance fields//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 21481-21492 [DOI: 10.1109/ICCV51070.2023.01969]

Zhou H, Liu J H, Liu Z W, Liu Y and Wang X G. 2020a. Rotate-and-render: unsupervised photorealistic face rotation from single-view images//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 5910-5919 [DOI: 10.1109/CVPR42600.2020.00595]

Zhou X Q, Wang X, Zheng J and Bai X. 2023. Adaptive spatial sparsification for efficient multi-view stereo matching. Acta Electronica Sinica, 51(11): 3079-3091 (周晓清, 王翔, 郑锦, 百晓. 2023. 基于自适应空间稀疏化的高效多视图立体匹配. 电子学报, 51(11): 3079-3091) [DOI: 10.12263/DZXB.20230353]

Zhou Y C, Liu S C and Ma Y. 2020b. Learning to detect 3D reflection symmetry for single-view reconstruction [EB/OL]. [2020-06-17]. https://arxiv.org/pdf/2006.10042.pdf

Zhou Y F, Shen Y R, Yan Y J, Feng C and Yang Y Q. 2021. A dataset-dispersion perspective on reconstruction versus recognition in single-view 3D reconstruction networks//Proceedings of 2021 International Conference on 3D Vision (3DV). London, United Kingdom: IEEE: 1331-1340 [DOI: 10.1109/3DV53792.2021.00140]

Zhu Y Z, Zhang Y P and Feng Q S. 2021. Colorful 3D reconstruction from single image based on deep learning. Laser and Optoelectronics Progress, 58(14): 199-207 (朱育正, 张亚萍, 冯乔生. 2021. 基于深度学习的单视图彩色三维重建. 激光与光电子学进展, 58(14): 199-207) [DOI: 10.3788/LOP202158.1410010]

Zhu Z W, Yang L Y, Li N, Jiang C H and Liang Y Y. 2023. UMIFormer: mining the correlations between similar tokens for multi-view 3D reconstruction//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 18180-18189 [DOI: 10.1109/ICCV51070.2023.01671]

Zou Z X, Yu Z P, Guo Y C, Li Y G, Liang D, Cao Y P and Zhang S H. 2024. Triplane meets Gaussian splatting: fast and generalizable single-view 3D reconstruction with transformers//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 10324-10335 [DOI: 10.1109/CVPR52733.2024.00983]

Zubić N and Liò P. 2021. An effective loss function for generating 3D models from single 2D image without rendering//Proceedings of the 17th IFIP WG 12.5 International Conference on Artificial Intelligence Applications and Innovations (AIAI). Hersonissos, Greece: Springer: 309-322 [DOI: 10.1007/978-3-030-79150-6_25]

作者简介

刘草,男,博士研究生,主要研究方向为计算机视觉、深度学习与三维重建。E-mail: ftliuc@mail.scut.edu.cn

曹婷,通信作者,女,助理研究员,主要研究方向为应用于复杂工业环境的智能检测技术、图像处理和点云处理。E-mail: caot@pcl.ac.cn

康文雄,男,教授,主要研究方向为图像处理与模式识别、计算机视觉。E-mail: auwxkang@scut.edu.cn

蒋朝辉,男,教授,主要研究方向为智能传感与检测技术、图像处理与智能识别和人工智能与机器学习。E-mail: jzh0903@csu.edu.cn

阳春华,女,教授,主要研究方向为复杂工业过程建模与优化控制、智能自动化系统与装置。E-mail: ychh@csu.edu.cn

桂卫华,男,中国工程院院士,主要研究方向为复杂工业过程建模、优化与控制应用和故障诊断与分布式鲁棒控制。E-mail: gwh@csu.edu.cn

梁骁俊,男,副研究员,主要研究方向为人工智能应用、智能控制与智能系统、工业监测技术、数字孪生仿真建模和工业互联网边缘计算。E-mail: liangxj@pcl.ac.cn