文章编号　1004-924X(2023)05-0667-30

三维补全关键技术研究综述

肖海鸿[1]，吴秋遐[2]，李玉琼[3]，康文雄[1*]

(1. 华南理工大学 自动化科学与工程学院，广东 广州 510640；
2. 华南理工大学 软件学院，广东 广州 510006；
3. 中国科学院 力学研究所 国家微重力实验室，北京 100190)

摘要：从部分观测信息中推断出完整三维形状与语义场景信息对自动驾驶、机器人视觉、元宇宙生态体系构建等而言是至关重要的，因此，主要围绕三维形状补全、三维场景补全和三维语义场景补全任务而展开的三维补全技术被广泛研究。本文围绕上述三维补全任务，对近年来的相关研究工作进行了系统性的分析和总结。首先，针对三维形状补全任务，对基于传统方法的形状补全和基于深度学习的形状补全这两个方面的研究进展进行了综述。其次，针对三维场景补全任务，对基于模型拟合的场景补全和基于生成式的场景补全方法这两个方面的研究进展进行了综述。再次，针对三维语义场景补全任务，深入分析了场景补全和语义分割两大任务之间的耦合特性，并根据输入数据的不同类型，对基于深度图的语义场景补全方法、基于深度图联合彩色图像的语义场景补全方法、以及基于点云的语义场景补全方法这三个方面的研究进展进行了综述。最后，对三维补全任务目前面临的主要问题及未来发展趋势进行了分析和展望，旨在为三维视觉中这一新兴领域的相关研究者提供一些有益的参考。

关　键　词：形状补全；场景补全；语义场景补全；三维视觉

中图分类号：TP391.4　　**文献标识码**：A　　**doi**：10.37188/OPE.20233105.0667

Key techniques for three-dimensional completion: a review

XIAO Haihong[1], WU Qiuxia[2], LI Yuqiong[3], KANG Wenxiong[1*]

(1. *School of Automation Science and Engineering, South China University of Technology, Guangzhou* 510640, *China*;
2. *School of Software of Engineering, South China University of Technology, Guangzhou* 510006, *China*;
3. *National Microgravity Laboratory, Insituate of Mechanics, Chinese Academy of Sciences, Beijing* 100190, *China*)
* *Corresponding author, E-mail: auwxkang@scut.edu.cn*

Abstract: The inference of complete three-dimensional (3D) shape and semantic scene information from partial observations is crucial for various applications, such as autonomous driving, robotic vision, and metaverse ecosystem construction. Research on 3D completion has primarily focused on 3D-shape, 3D-scene, and 3D-semantic scene completion. In this paper, we systematically summarize and analyze recent

收稿日期：2022-09-10；**修订日期**：2022-10-14.

基金项目：国家自然科学基金项目(No. 61976095, No. 61575209)；广东省自然科学基金项目(No. 2022A1515010114)；中央高校基本科研业务费专项资金项目(No.2022ZYGXZR099)

relevant studies concerning these 3D completion tasks. First, for 3D-shape completion, the research progress is reviewed from two aspects: traditional shape completion and deep learning-based shape completion. Second, for 3D-scene completion, the research progress is reviewed from two aspects: the scene completion method based on model fitting and the scene completion method based on a generative approach. For 3D-semantic scene completion, the coupling characteristics between the two tasks of scene completion and semantic segmentation are analyzed, and the research progress is reviewed from three aspects: the depth map-based semantic scene completion method, the depth map-based semantic scene completion method with color images, and the point cloud-based semantic scene completion method, according to the different forms of input data. Finally, we analyze the current problems and future development trends of 3D completion tasks, aiming to provide a reference for related studies in this emerging field in 3D vision.

Key words: shape completion; scene completion; semantic scene completion; 3D vision

1 引　言

近年来,随着深度学习和传感器技术的快速发展,三维视觉受到了学术界和工业界的广泛关注,在目标检测[1]、语义分割[2]、三维重建[3]等领域都取得了突破性的进展。然而,一个固有的问题仍然存在,即由于物体遮挡、表面反射、材料透明、视角变换和传感器分辨率的限制,传感器在真实场景下所获取的数据并不完整,阻碍了下游任务的研究进展。在无人驾驶领域,三维补全技术可为环境感知任务提供精确的物体识别和跟踪信息[4]。在生产制造领域,三维补全技术可为机械臂抓取任务提供准确的物体位姿信息[5]。在文物保护领域,三维补全技术可为数字化的文物鉴定、检测和修复提供依据[6]。此外,三维补全技术还可为虚拟数字人的重建[7]和元宇宙生态体系的构建[8]奠定基础,如图1所示。理解三维环境是人类的一种自然能力,人们可以利用学到的先验知识估计出缺失区域的几何和语义信息,然而,这对计算机而言是比较困难的[9]。

针对上述问题,研究人员开展了一系列围绕三维形状补全、三维场景补全和三维语义场景补全方面的研究工作。其中,三维形状补全可以提高场景理解的准确度,其目的是根据已观测到的局部形状恢复出物体的完整几何形状,其补全对象通常是单个物体[10]。传统的三维形状补全方法主要是通过几何对称性[11-12]、表面重建[13-16]、模板匹配[17-20]等方式进行补全。近年来,随着深度学习的发展,基于学习的形状补全工作[21-24]取得重要进展。然而,基于学习的形状补全方法目前大多是在合成数据集上进行,在真实场景下的补全效果仍然存在较大的提升空间。

三维场景补全是形状补全的扩展,需要在扫描的场景中对缺失部分进行补全[25],其核心在于补全后场景细粒度的保持。相较于形状补全,场景补全具有补全面积大和补全对象多的特点[25]。当缺失区域较小的时候,可以采用平面拟合[26]和插值[27]的方法。当缺失区域较大的时候,这类方法难以达到令人满意的结果。因此,一些工作试图通过模型拟合的方法[28-30]来得到干净而紧凑的场景表示。最近,利用深度神经网络直接作用于整个场景的生成补全方法[31-32]显示出了很大的研究潜力。然而,这类方法忽略了语义信息对场景补全的辅助,当补全的场景过于复杂时,其精度会有所下降。

三维语义场景补全是在场景补全的几何基础上同时估计出场景的语义信息。事实证明,语义信息和几何信息是相互交织耦合的[33]。换句话说,当在未完整观测一个物体的情况下,已知它的语义信息有助于估计出它可能占据的场景区域。如图2所示,看到桌子后面的椅子顶部,推断出椅子的座位和腿的存在。同理,已知一个对象的完整几何信息,有助于识别其语义类别。然而,语义场景补全是相对复杂的,表现为数据的稀疏性和没有真实完整的地面参考值(通过多帧融合形成的参考值仅能提供较弱的监督信号)。相较于形状补全,语义场景补全需要深入了解整个场景,严重依赖于学习到的先验知识来解决歧义性。伴随着大规模语义场景数据集[34-37]的出现,基于深度学习的语义场景补全方法[38-41]相继

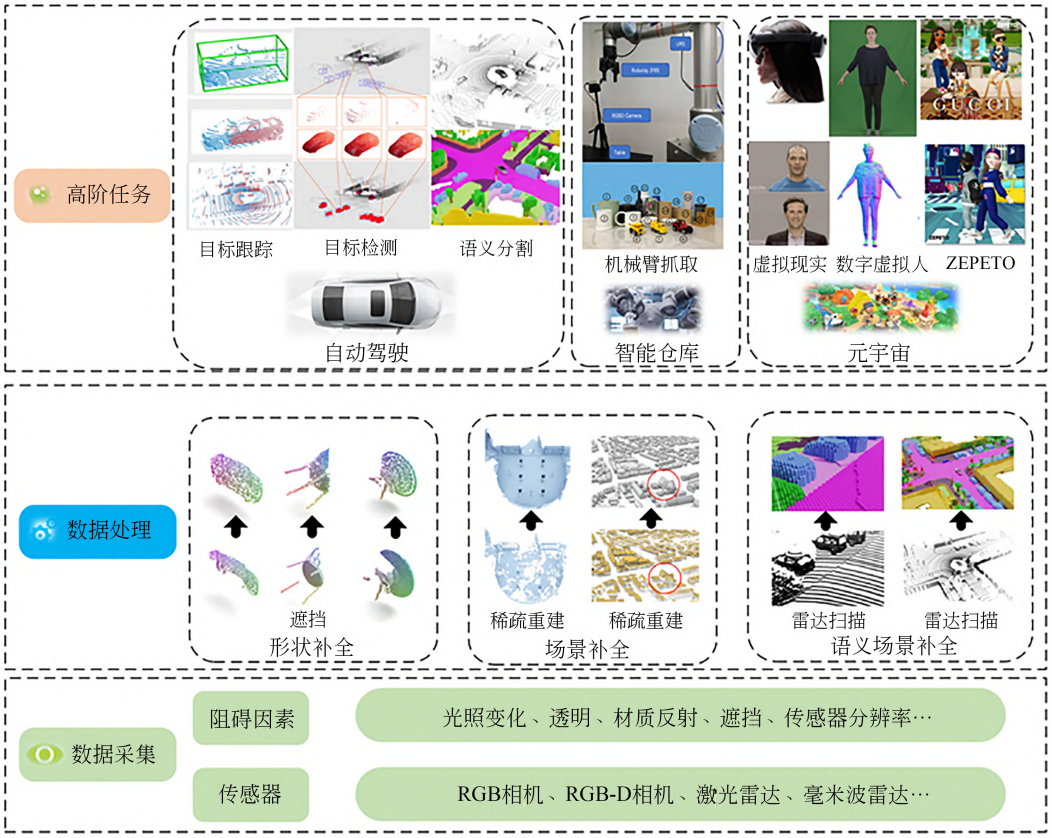

图 1　研究目的导向图

Fig. 1　Research purpose oriented map

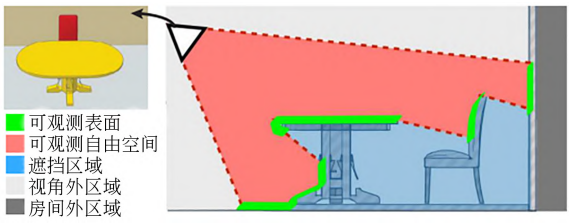

图 2　场景信息观测视角图

Fig. 2　Scene information observation perspective map

被提出并取得不错的结果。然而，现有的方法在物体几何细节、模型内存占用、场景不确定性估计等方面还存在诸多不足。

在过去几年，关于三维视觉的相关工作，如三维深度学习[42]、三维目标检测[43]、三维语义分割[44]、三维重建[45]、实时重建[46]等方面都有相对应的综述，但系统总结三维补全的工作几乎没有，而与本文并行的工作[47]也仅是总结基于点云输入的形状补全。本文将系统地介绍国内外在三维形状补全、三维场景补全和三维语义场景补全这三方面所展开的相关研究工作，并选取其中部分具有代表性的算法进行客观评价和归纳总结。最后，本文讨论了该领域目前存在的问题并展望了未来的发展趋势。希望本文能够对刚进入这一新兴领域的研究者起到导航的作用，同时，也希望能够对相关领域的研究者提供一些参考和帮助。

本文的后续内容安排为：第 2 节整理三维补全相关数据集和评价指标；第 3 节根据模型构建过程中有无神经网络的参与，将现有的形状补全算法分为传统方法和基于深度学习方法两大类并进行梳理与小结；第 4 节分别从模型拟合和生成式的角度梳理了场景补全任务中具有代表性的算法并进行小结；第 5 节根据输入数据的不同类型，分别从深度图输入、深度图联合彩色图像输入、点云输入三方面梳理了语义场景补全任务中具有代表性的算法并进行小结；第 6 节讨论了三维补全领域存在的问题，并对未来可能的发展方向进行展望；第 7 节对本文内容进行总结。

2　数据集和评价指标

2.1　数据集

随着传感器技术的突破和三维视觉的快速发展，三维补全相关数据集的广泛获取成为可能。为了便于研究者能够直接展开相关工作，本文汇总了常用的数据集，并根据数据集的类型不同，将其划分为合成数据集和真实数据集。其中，合成数据集包括：ShapeNet-Part（ShapeNet[48]子集）、SUNCG[33]、Fandisk[49]、Raptor[49]和NYUCAD[31]，真实数据集包括：KITTI[50]、ScanNet[51]、Matterport3D[52]、DFAUST[52]、MHAD[53]、NYUv2[54]、tabletop[31]、Semantic KITTI[35]和SemanticPOSS[37]。数据集的详细介绍如表1所示。

表1　三维补全相关数据集

Tab. 1　3D completion related datasets

数据集	年份	类型	来源	简要描述
ShapeNet-Part	2015	合成	—	ShapeNet-Part[1]数据集是ShapeNet数据集[48]的子集，ShapeNet是大规模CAD模型注释的形状存储库。不同补全方法使用数据集的类别数目（8类、16类、34类、55类）有所不同，但都来自ShapeNet数据集，为了简明，统一用ShapeNet-Part表示。数据形式包括网格、体素和点云。点云是CAD模型经采样得到，采样方法包括最远点采样（Farthest Point Sampling，FPS）、均匀采样（Uniform Sampling，US）和泊松圆盘采样（Poisson Disk Sampling，PDS）。
KITTI[50]	2012	真实	LiDAR传感器＋灰度相机＋彩色相机＋GPS	KITTI数据集是面向自动驾驶的场景点云数据集。它包含7 481个图像点云对用于训练，7 518个图像点云对用于测试，共包含80 256个标记对象。
ScanNet[51]	2017	真实	RGB-D深度相机	ScanNet数据集是RGB-D室内视频数据集，包括场景实例分割、语义分割标注和相机姿态信息，共采集21个类别对象，1 513个场景数据。
Matterport3D[52]	2017	真实	3D扫描仪	Matterport3D数据集包含10 800张尺寸相同的全景图（RGB＋深度图像）。
DFAUST[52]	2017	真实	3D扫描仪	DFAUST数据集是以60帧每秒速度捕获运动人体的高分辨率4D扫描数据集。
MHAD[53]	2013	真实	RGB-D深度相机	MHAD数据集提供来自2个视角RGB-D相机获取的人体动作数据。
SUNCG[33]	2017	合成	—	SUNCG数据集是一个合成的具有密集体素标注的三维室内场景数据集，包含超过45 622个室内场景。
NYUv2[54]	2012	真实	RGB-D深度相机	NYUv2数据集是由Kinect深度相机记录的室内场景数据集。由26个场景类中的464个不同室内场景中拍摄获取的RGB和深度图像组成，包含894个不同的语义类别，共1 449张拥有语义标签标注的RGB-D图片以及407 204张未标注图片。
tabletop[31]	2016	真实	GB-D深度相机	tabletop数据集包含90个桌面对象的完整场景。
Fandisk[49]	2001	合成	万维网	Fandisk数据集[2]来源于万维网上下载的三维模型库中的Fandisk模型，经过采样得到点云数据。
Raptor[49]	2001	合成	万维网	Raptor数据集[2]来源于万维网上下载的三维模型库中的Raptor模型，经过采样得到点云数据。

续表 1　三维补全相关数据集

Tab. 1　3D completion related datasets

数据集	年份	类型	来源	简要描述
NYUCAD[31]	2013	合成	—	NYUCAD数据集是NYUv2数据集的衍生版，将网格标注数据渲染为深度图。
Semantic KITTI[35]	2019	真实	LiDAR-64	Semantic KITTI数据集是基于KITTI Vision Benchmark里程计数据集的大型户外点云数据集，包括城内交通、住宅区、高速路和乡村道路场景，提供23 201个完整的3D扫描场景用于训练，20 351个场景用于测试。
SemanticPOSS[37]	2020	真实	LiDAR-40	SemanticPOSS数据集包含2 988个复杂的雷达扫描场景，有大量的动态实例。

注：1：尽管在不同的点云补全方法中，数据集名称有所不同。如：在PCN中，数据集为PCN数据集，在TopNet中，数据集为Completion3D数据集，在PointTr中，数据集为ShapeNet34和ShapeNet55数据集，在VRCNet中，数据集为MVP数据集，但是它们均来源于ShapeNet合成数据集，只是类别数目不同和虚拟相机的视角不同。为了简明，统一用ShapeNet-Part表示。ShapeNet数据集地址：https://shapenet.org/。2：Fandisk数据集和Raptor数据集中原始的3D模型来自于网络下载，如：https://www.viewpoint.com。

2.2　评价指标

2.2.1　形状补全结果评价指标

由于三维形状表示的不同形式，补全结果的评价标准也是不同的。针对体素网格形式的补全评价标准，主要采用一定分辨率下的误差率衡量补全性能的好坏（误差率即补全形状和真值之间的差异体素网格数除以真值体素网格总数）[55]。针对三角网格或多边形网格形式的补全评价标准，主要是通过计算补全网格顶点和真值网格顶点之间的平均欧几里得距离（Euclidean Distance，ED）进行评估[56]。针对点云表示的补全评价标准，大多数方法采用补全点云与真值点云之间的倒角距离（Chamfer Distance，CD）[10]或地球移动距离（Earth Mover′s Distances，EMD）[10]进行评估，其公式定义如下：

$$d_{\mathrm{CD}}(P_1,P_2)=\frac{1}{P_1}\sum_{a\in P_1}\min_{b\in P_2}\|a-b\|_2^2+\frac{1}{P_2}\sum_{b\in P_2}\min_{a\in P_1}\|a-b\|_2^2, \tag{1}$$

$$d_{\mathrm{EMD}}(P_1,P_2)=\min_{\varphi:P_1\to P_2}\sum_{a\in P_1}\|a-\varphi(a)\|_2, \tag{2}$$

其中：P_1和P_2分别表示生成的点云和真实完整点云，a和b分别表示P_1和P_2中的点，φ表示双向映射函数。

Shu等[57]提出弗雷歇点云距离（Fr′echet Point Cloud Distance，FPD）用于衡量补全点云与真实点云之间相似度，其公式定义如下：

$$d_{\mathrm{FPD}}(P_1,P_2)=\|\boldsymbol{m}_{P_1}-\boldsymbol{m}_{P_2}\|_2^2+\mathrm{Tr}(\boldsymbol{\Sigma}_{P_1}+\boldsymbol{\Sigma}_{P_2}-2(\boldsymbol{\Sigma}_{P_1}\boldsymbol{\Sigma}_{P_2})^{\frac{1}{2}}), \tag{3}$$

其中：$\boldsymbol{m}_{P1}$和$\boldsymbol{m}_{P2}$分别表示生成点云和真实点云的特征向量，$\boldsymbol{\Sigma}_{P1}$和$\boldsymbol{\Sigma}_{P2}$分别表示生成点云和真实点云的协方差矩阵，$\mathrm{Tr}(\boldsymbol{A})$表示矩阵$\boldsymbol{A}$的主对角线元素之和。

Wu等[58]提出密度感知倒角距离（Density-Aware Chamfer Distance，DCD），和原始CD相比，它对密度分布的一致性更加敏感，而和EMD相比，它更擅长于捕捉局部细节，其公式定义如下：

$$d_{\mathrm{DCD}}(P_1,P_2)=\frac{1}{2}\left(\frac{1}{P_1}\sum_{a\in P_1}\left(1-\frac{1}{n_{b_{\min}}}e^{-\alpha\|a-b_{\min}\|_2}\right)+\frac{1}{P_2}\sum_{b\in P_2}\left(1-\frac{1}{n_{a_{\min}}}e^{-\alpha\|b-a_{\min}\|_2}\right)\right), \tag{4}$$

$$b_{\min}=\min_{b\in P_2}\|a-b\|,\ a_{\min}=\min_{a\in P_1}\|b-a\|_2$$

其中：α表示温度标量，P_1和P_2分别表示生成的点云和真实完整点云。

Chen等[59]提出视觉相似度评价指标——光场描述符（Light Field Descriptor，LFD）。它的

原理是对3D形状渲染的2D视图通过Zernike矩阵和傅里叶变换(Fourier Transform,FT)进行相似度分析。对于一些无监督的点云补全方法,由于没有真值参考,使用单向倒角距离(Unidirectional Chamfer Distance, UCD)[60]或单向豪斯多夫距离(Unidirectional Hausdorff Distance, UHD)[60]进行评估。

2.2.2 场景补全结果评价指标

在场景补全任务中,由于最终输出场景的表示形式不同,因此也有不同的评价标准。大多数方法输出的场景表示是TSDF[25]编码的矩阵,因此常用L_1距离[32]作为评价标准。其中,一些方法的输出是网格或点云。因此,可以使用CD[10]、EMD[10]、FPD[57]和DCD[58]作为评价指标。对于基于模型拟合的场景补全方法,其评价标准大多采用模型对齐精度[28]作为评价指标。Dahnert等[30]使用混淆分数(Confusion Score, CS)衡量嵌入空间的学习程度,以此进一步衡量补全模型和CAD模型之间的平衡程度。

2.2.3 语义场景补全结果评价指标

在三维语义场景补全任务中,评价指标是相对统一的,为预测结果和真值结果之间的交并比(Intersection over Union, IoU)[33]或平均交并比(Mean Intersection over Union, mIoU)[33],其公式定义如下:

$$I_{\mathrm{IoU}}=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}+N_{\mathrm{FN}}}, \tag{5}$$

$$I_{\mathrm{mIoU}}=\frac{1}{C}\sum_{c=1}^{C}\frac{N_{\mathrm{TP}_c}}{N_{\mathrm{TP}_c}+N_{\mathrm{FP}_c}+N_{\mathrm{FN}_c}}, \tag{6}$$

其中:N_{TP}表示“正阳性”即预测已占用体素结果中的预测正确的样本数量,N_{FP}表示“假阳性”即预测错误的样本数量,N_{FN}表示“假阴性”即未被检测到的已占用体素数量,C表示类别。

3 三维形状补全

3.1 基于传统的形状补全方法

3.1.1 基于对称的方法

对称是自然界广泛存在的一种现象,对称性是重要的科学思维方法之一,最初的形状补全方法主要是利用物体或空间呈现的几何对称性[11-12]恢复缺失区域的重复结构。该方法假设了缺失的几何部分在现有的部分观测信息中具有重复结构,对于大部分呈现立体对称结构的简单物体是有效的。然而,对称性假设并不适用于自然界中的所有物体。

3.1.2 基于表面重建的方法

现有的表面重建方法主要分为插值和拟合两种方式[13]。插值是将表面上集中的数据点作为初始条件,通过不同算法执行插值操作得到密集表面。拟合是利用采样点云直接重建近似表面,通常以隐式形式表示。

Lee等[14]提出基于多层B样条的快速分散数据插值算法。该算法在插值效果和计算时间上都具有较好的优势。但在选点方式和定义权重上存在一定的困难,导致重建的表面存在不连续情况。Price等[15]采用分形插值方法来重建三维曲面。与传统插值方法相比,分形插值在拟合具有分形特征或较为复杂的事物时具有优势。但它计算复杂,并且分形的参数H较难估计。泊松表面重建[13]是一种隐函数表面重建方法,它通过平滑滤波指示函数构建泊松方程,将表面重建问题等同为泊松方程的求解问题。通过对该方程进行等值面提取,得到具有几何实体信息的表面模型;其构建的表面能容忍一定程度的噪声,但存在过度平滑问题。针对泊松表面重建的过度平滑问题,Kazhdan等[16]通过引入样本点的位置约束,将其表示为屏蔽泊松方程进行求解,生成更为贴合的表面,但该方法比较依赖于准确的点云法向量。尽管以上基于插值和拟合的表面重建方法都取得了较好的结果,但这类方法通常用于孔洞修复,存在补全面积小的限制。

3.1.3 基于模板匹配的方法

基于模板匹配的形状补全方法主要包括部分形状匹配方法[17]和整体形状匹配方法[18]。部分形状匹配主要是在预先定义的大型形状模型库中寻找能够最佳拟合的对象部件,然后将它们组装起来获得完整形状。整体形状匹配是在模型库中直接检索完整的最佳拟合形状。其中,寻优算法的选取对最终模板匹配的结果起到至关重要的作用。

Rock等[19]提出模型变形的匹配方法,其核心是将从数据库中检索到的候选模型执行非刚性

曲面对齐使其形状变形以拟合输入。Sun等[20]进一步提出基于补丁的检索-变形方法。该方法首先从输入形状中选择候选补丁，其次，对检索到的候选对象执行变形操作并缝合成完整形状。该方法可以重建在拓扑结构上不同于训练数据的形状。尽管基于模板匹配的方法取得了较好的补全结果。但这类方法通常存在寻优速度慢和对噪声比较敏感的问题。同时，它依赖于较大的模型库来覆盖补全的全部形状，这在真实世界中往往是不切实际的。

3.2　基于深度学习的形状补全方法

目前常用的三维数据表示形式包括点云[61]、体素[62]和网格[63]。尽管最近基于深度隐式表示的方式，如占用网络(Occupancy Networks)[64]、连续符号距离函数(Sign Distance Function，SDF)[65]和神经辐射场(Neural Radiance Field，NeRF)[66]在三维重建和三维语义场景补全任务中有相关的工作。但在形状补全任务中，目前大多数补全方法都依赖于点云的数据形式，这不仅与点云自身的特性有关，即存储空间小且表征能力强，还与点云数据集相对容易获取有关。

基于学习的形状补全方法根据其算法原理，可归纳为6种主要类型：基于逐点的多层感知机(Multi-layer Perceptron，MLP)方法[10,21-24,67-68]、基于卷积的方法[69-72]、基于图的方法[56,73-76]、基于生成对抗的方法[77-81]、基于Transformer的方法[82-83]和其他方法[65,84-85]，其发展历程如图3所示。下面，本文将对其中具有代表性的一些算法进行介绍和总结。

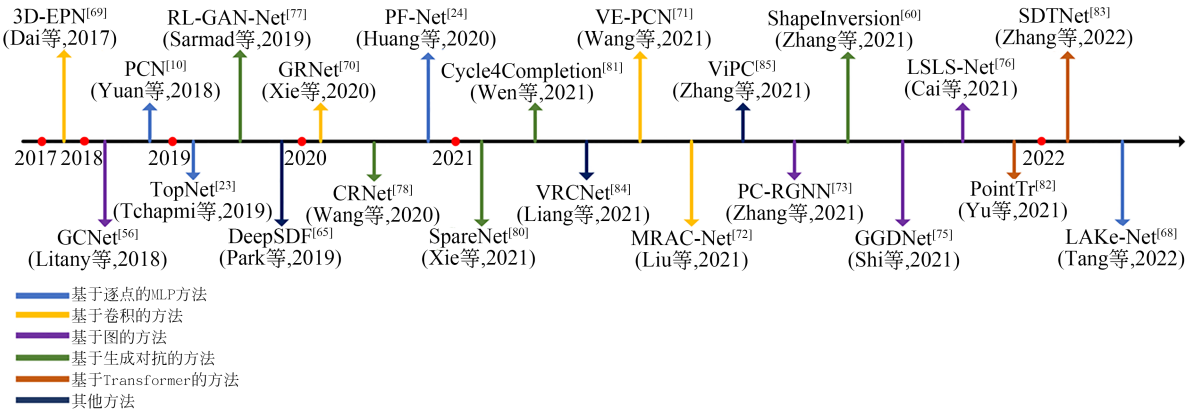

图3　基于深度学习的三维形状补全方法发展历程

Fig. 3　Development history of 3D shape complementation methods based on deep learning

3.2.1　基于逐点的MLP方法

点云是三维形状补全任务中最常使用的数据形式。尽管它具有存储空间小且表征能力强的优点，但是它的无序性和不规则性也给特征提取带来了巨大挑战。PointNet[61]是首个被提出直接作用于点云数据的深度学习网络，它采用MLP对每个点进行独立地特征提取。然后，使用最大池化(Max Pooling)函数得到点云的全局特征。得益于这项工作的启发，基于逐点的MLP形状补全算法相继被提出[10,21-24,67-68]。

作为点云形状补全的开创性工作，PCN[10]遵循编码器-解码器范式来完成点云补全任务。编码器主要由堆叠的MLP层构成，解码器包括全连接层解码器[21]和折叠解码器[22]两部分。其中，全连接层解码器负责估计点云的几何形状，而折叠解码器负责近似出局部几何形状的光滑表面。尽管该算法能够获得较好的完整点云和稠密点云，基于折叠的二维网格变形操作在某种程度上会限制三维点云的几何表达。Tchapmi等[23]提出分层的树结构点云补全网络TopNet，其核心思想是采用基于MLP的树解码器生成结构化的完整点云。该算法允许网络学习任意的拓扑结构，而不是强制执行某一种拓扑结构。然而，该算法需要足够的冗余空间来学习任意的体系结构，因此，解码器的容量在某种程度上会限制学习到的拓扑结构。

Huang等[24]提出点云分形网络PF-Net。首先，该算法使用FPS采样方法将输入点云下采样

为不同分辨率的点云。其次，使用提出的组合多层感知机（Combined Multi-layer Perception，CMLP）分别进行特征提取并融合成全局特征向量。最后，将得到的全局特征向量输入到点云金字塔解码器（Point Pyramid Decoder，PPD）进行多阶段预测。该算法采用的多分辨率特征提取方法能够更好地捕获输入点云的局部特征。然而，PF-Net仅预测缺失的点云，对已有的部分不进行预测，导致生成的点云和已有的点云在拼接时存在间隙。Liu等[67]提出两阶段的稠密点云补全算法MSN。首先，该算法通过自动编码器预测一个完整但粗粒度的点云。其次，通过采样算法将粗粒度的预测和输入点云合并生成致密点云。值得一提的是，为了防止参数化表面元素的重叠，文中提出扩展惩罚损失来引导表面元素集中在一个局部区域。Tang等[68]提出基于关键点-骨架-形状预测的点云补全算法LAKe-Net，该算法主要包括3个步骤：(1)使用非对称关键点定位器（Asymmetric Keypoint Locator，AKL）定位出输入点云和完整点云中对齐的关键点；(2)利用基于几何先验的关键点生成表面骨架来充分显示拓扑信息；(3)使用递归细化模块辅助点云骨架的精细化完成。该算法严重依赖于缺失和完整的形状匹配对进行监督训练。在某些情况下，完整的点云数据是无法获取的，从而限制了其在实际场景下的适用性。

尽管基于逐点的MLP形状补全算法表现出不错的性能，但仍然存在着以下局限：

(1)基于逐点的MLP算法大多沿用PointNet的特征提取思路，而这种方式是独立的处理每个点，忽略了相邻点之间的几何关系。

(2)一些方法采用了由粗到细的点云生成策略，但是它们对形状的高频信息并不敏感，难以对复杂的拓扑结构进行友好生成。

3.2.2 基于卷积的方法

卷积神经网络（Convolutional Neural Network，CNN）[86]近年来在视觉图像领域取得了巨大的成功，其相关工作也启发了研究者使用体素来表示三维形状。相较于点云的无序形式，体素更贴近于规则像素的表达方式，同时也更容易使用CNN进行特征提取和学习。

3D-EPN[69]使用三维卷积层组成的编码-解码器网络预测部分输入的完整形状，但是随着分辨率的提升，计算量会呈指数增加，给网络的训练带来了极大挑战。Xie等[70]提出网格残差网络GRNet，通过将无序点云转为规则网格的中间表示，然后利用3DCNN进行特征提取和中间数据生成，最后将生成的网格单元再次转化为点云形式。此外，该算法设计了立方特征采样（Cubic Feature Sampling，CFA）层来提取相邻点信息和上下文信息。然而，该方法存在以下2个缺点：(1)点云体素化的过程不可避免的导致信息丢失；(2)体素表示仅适用于低分辨率的形状重建。

Wang等[71]提出基于体素的多尺度点云补全网络VE-PCN。相比于GRNet采用逆体素点云化策略生成粗糙点云，VE-PCN增加了边生成器（Edge Generator）将补全对象的高频结构信息注入到形状补全分支中，并取得较好的补全结果。需要注意的是，这里的高频结构信息指代三维对象的边缘结构信息[71]。Liu等[72]提出多分辨率各向异性卷积网络MRAC-Net。文中设计了一种多分辨率各向异性卷积编码器（Anisotropic Convolutional Encoder，ACE）提取三维对象的局部和全局特征，以提高网络对语义和几何信息的理解能力。此外，该网络提出的组合金字塔解码器能够分层输出不同分辨率的完整结构点云，实现更好的监督。

尽管基于卷积的形状补全算法均表现出不错的性能，但是仍然存在着以下局限：

(1)内存随分辨率呈立方增加，现有的网络算法依旧局限于相对较低的分辨率。

(2)使用体素的中间表示会不可避免的导致细节丢失。

3.2.3 基于图的方法

点云作为一种无序的非欧几里德结构数据，无法直接将经典的CNN应用于点云学习，点云中的拓扑信息由点之间的距离隐式表示。因此，一种可行的思路是将点云中的点看作图顶点，使用图卷积网络（Graph Convolutional Network，GCN）[87]提取邻域顶点间的结构信息。动态图CNN（Dynamic Graph CNN，DGCNN）[87]使用一种可插拔的边卷积（EdgeConv.）模块动态地捕获点云的邻域特征，该工作也启发了后续基于图卷积的形状补全工作[56,73-76,88]。

Litany等[56]提出基于可变形的形状补全方法GCNet，其核心是通过一个带有图卷积的变分自动编码器（Variational Autoencoder，VAE）来学习完整真实形状的潜在空间表示。然而，该方法假定所有的形状都与一个共同的参考形状相对应，从而限制了对某些类别形状的适用性。Zhang等[73]提出3D目标检测网络PC-RGNN。他们首次使用点云补全技术辅助三维目标检测任务，设计了一种基于注意力的多尺度图卷积（Attention Based Multi-scale Graph Convolution，AMS-GCN）模块来编码点之间的几何关系，增强对应特征的传递。在点云生成阶段，该方法沿用了PF-Net的思路，采用PPD生成多阶段的完整点云，在补全数据集和检测数据集上均表现良好。Pan[74]提出具有图卷积的边缘感知点云补全网络ECG。该网络包括两个阶段，第一阶段生成粗糙的骨架，以方便捕获有用的边缘特征；第二阶段采用图卷积层次编码器来传播多尺度边缘特征，以实现局部结构的细化。为了在上采样时保留局部几何细节，作者进一步提出边缘感知特征扩展（Edge-aware Feature Expansion，EFE）模块来平滑上采样点的特征。实验结果表明，该算法在稠密点云的生成方面具有一定的优势。

Shi等[75]提出一种以输入数据和中间生成为控制点和支撑点的图引导变形网络GGD-Net，通过利用网格变形方法模拟最小二乘的拉普拉斯变形过程，这为建模几何细节的变化带来了自适应。据公开文献[75]，这是第一个通过使用GCN引导变形操作来模拟传统图形算法优化的点云补全工作，在室内和室外数据集上均表现良好。Cai等[76]提出无监督点云补全方法LSLS-Net。他们认为不同遮挡程度的缺失点云共享统一完整的潜在空间编码，其核心思想是引入遮挡码对潜在空间的统一编码进行掩码，再通过解码器对掩码的潜在编码解码成不同遮挡比例的残缺点云。编码器主要包括多个EdgeConv层，解码器主要由多层MLP组成。尽管该方法在泛化性上取得了较好的结果，但是该方法设计的解码器较为简单，在补全结果的细节性上还有待提升。

基于图的形状补全算法在邻域特征提取上表现出良好的性能，但是仍然存在着以下局限：

（1）基于图的形状补全方法大多是采用K近邻（K-nearest Neighbor，KNN）算法选取每个点的n个最近点作为它的邻居集合，然后利用图滤波操作来学习这些点的表示。然而，n的取值会极大地影响网络的性能。此外，KNN算法对点云的密度分布非常敏感。

（2）基于图的算法是相对耗费时间的，当点云数据更大或者堆叠的图模块更多时，其内存消耗更为明显。因此针对点云的图浓缩（Graph Condensation）[88]技术是值得探讨的。

3.2.4　基于生成对抗的方法

生成对抗网络（Generative Adversarial Network，GAN）[89]创新性地采用了相互对抗的网络框架，通过生成模型和判别模型进行最小化和最大化博弈学习不断提升数据的生成能力。为了提升点云的生成质量，相关研究者结合GAN来完成形状补全任务。

Sarmad等[77]提出将自动编码器（Autoencoder，AE）、GAN和强化学习（Reinforcement Learning，RL）相结合的点云补全网络RL-GAN-Net。通过RL代理优化GAN的潜在变量输入，并使用预训练解码器对GAN生成的潜在全局特征向量解码为完整点云。然而，多阶段训练过程增加了网络的复杂性。此外，基于RL的代理控制难以找到最优的潜在变量输入。Wang等[78]提出级联细化补全网络CRNet，该方法遵循由粗到细的生成策略。在第1阶段采用PCN的特征提取方式通过全连接层生成粗糙点云，在第2阶段引入条件迭代细化子网络生成高分辨的点云。为了提升生成点云的逼真性，文中提出了块判别器（Patch Discriminator）来保证每个区域都是真实的。此外，该方法加入类别的平均形状先验信息来提升补全结果的完整性，但同时也降低了类内补全结果的多样性。此外，由于块之间的互斥性易导致生成点云的不均匀分布。

Hu等[79]将点云的补全问题转化为深度图补全问题。通过将点云从固定视角渲染成8个多视图，并执行每个视图的补全。值得一提的是每个视图的补全并不是独立的，而是利用所有视图的信息来辅助每一个视图的补全。此外，为了提升深度图补全的逼真性，采用深度图判别器对补全结果和真实点云的渲染结果进行真假判断。然而，该方法缺乏对点云的直接监督，通过渲染的

方式会导致信息的丢失。Xie等[80]提出基于风格生成和对抗渲染的点云补全网络SpareNet。该算法分别从特征提取、点云生成和优化三个方面进行了改进。针对特征提取部分,引入通道注意力的边卷积(Channel-attentive EdgeConv, CA-EdgeConv)模块来增强点云的局部特征提取能力。针对点云生成部分,通过将学习到的特征作为样式码(Style Code)来提高折叠生成能力。为了进一步优化生成质量,引入了可微分对抗渲染器来提升点云的视觉逼真度。该方法在ShapeNet-part和KITTI数据集上均表现良好。

Wen等[81]提出双向循环的无监督点云补全网络Cycle4Completion,与现有的无监督形状补全方法不同[60,76],之前的方法都只考虑从缺失点云到完整点云的正向对应关系,而该算法同时考虑了正向和逆向的对应关系。此外,该算法中判别器的输入是潜在表示而不是点云。潜在表示在这里代表一个完整点云的特征向量,根据这个特征向量能够恢复出点云结构。然而,双向循环网络需要单独建模,这对训练过程提出了较大的挑战。Zhang等[60]提出无监督形状反演补全网络ShapeInversion,首次将GAN逆映射(GAN Inversion)引入到点云补全任务中。类比GAN逆映射在二维图像修复中的应用,文中提出了k-mask退化函数将生成的完整点云转化为与输入点云对应的残缺点云。利用GAN提供的先验知识,ShapeInversion在多个数据集上表现出优异的结果,甚至超过了部分有监督方法。然而,该方法需要额外的预训练生成模型,降低了其在实际情况下的适用性。

尽管基于生成对抗的形状补全算法在相关数据集上均表现良好,但是仍然存在着以下限制:

(1)虽然相比于直接训练GAN生成完整点云的方式,在潜在空间表示上训练GAN会相对容易。但是,训练GAN需要达到纳什均衡,因此其训练过程充满着不稳定性。

(2)在无监督形状补全方法中,一些算法需要借助额外的预训练生成模型,这会大大降低算法在实际情况下的适用性。

3.2.5 基于Transformer的方法

近两年,Transformer[90]在自然语言处理、计算机视觉和语音处理领域取得了巨大成功,吸引了研究者的广泛关注。原始的Transformer模型主要包括编码器和解码器,其中编解码器主要由多头注意力(Multi-head Self-attention, MSA)模块和前馈神经网络(Feed-forward Network, FFN)组成;而解码器的内部结构与编码器类似,在MSA模块和FFN模块之间额外插入了一个交叉注意力(Cross-attention, CA)模块。受此启发,Zhao等[91]提出Point transformer框架,在点云分类和语义分割任务上达到了当时的最先进水平。几乎同一时间,Guo等[92]提出了Point cloud transformer网络,在点云分类、法向量估计和语义分割任务上均表现优异。

Yu等[82]首次将Transformer应用到点云补全任务中,即PointTr。该方法将无序点云表示为一组带有位置嵌入的无序点组,从而将点云转换为一系列点代理,并使用几何感知的Transformer编码-解码器生成缺失部分的点代理(Point Proxy)。最后,基于生成的点代理结合折叠网络生成细粒度的缺失点云。然而,Transformer模型的二次方计算量需要极大的显存和内存占用。Zhang等[83]提出具有骨架-细节Transformer的点云补全框架SDTNet,该方法遵循由粗到细的生成策略。该算法探索了局部模块和骨架点云之间的相关性,有效地恢复出点云细节。此外,文中引入了一种选择性注意力机制(Selective Attention Mechanism, SAM),在显著降低Transformer记忆容量的同时而不影响整体网络性能。

尽管基于Transformer的形状补全方法在相关数据集上表现优异,但仍然存在着以下局限:

(1)Transformer的二阶计算量和内存复杂度极大地限制了它的可适用性。

(2)由于Transformer的计算复杂度会随着上下文长度的增加而增长,这使其难以有效地建模长期记忆。

(3)Transformer对形状补全的增益需要更多的训练数据作为基础。

3.2.6 其他方法

Pan等[84]提出变分关系点云补全网络VRC-Net,它由概率建模子网络和关系增强子网络级联而成。在第1阶段,通过重建路径引导补全路径学习生成粗粒度完整点云,实现从高层次的特

征分布到低层次的信息流动。在第2阶段，通过并联的多尺度自注意力模块增强点云的细节生成，该算法显著提升了点云的细节生成能力。

Zhang等[85]提出视觉引导的跨模态点云补全网络ViPC。不同于现有的算法仅依赖部分点云作为输入，该算法从额外输入的单视图中挖掘缺失点云的全局结构信息作为引导。此外，该算法引入动态偏移预测器（Dynamic Offset Predictor，DOP）和差分精调策略（Differential Refinement Strategy，DRS）对低质量点进行维精调，对高质量点执行重度精炼，并在所提出的ShapeNet-ViPC数据集上取得了最好的结果。

Park等[65]提出基于学习的深度隐式形状补全算法DeepSDF。该方法利用连续SDF生成像水一样密集的封闭形状表面，不仅具有良好的视觉效果，需要的内存空间也大幅降低，为在复杂形状的生成方面提供了新的思路。

3.3　分析与小结

分析不同类型的三维形状补全方法，并根据表2、表3和图4所示的部分方法对比结果，得出下列结论：

（1）在基于深度学习的三维形状补全工作中，点云因其存储空间小、表征能力强的特点，成为广泛使用的三维数据表示形式。因此，基于点云的深度学习补全算法也成为当今的研究热点之一。

（2）之前的形状补全方法严重依赖于形状匹配对的形式进行监督训练，在域内数据集上能够表现出较好的结果，当扩展到其他部分形状数据集或真实世界所观测的部分数据时，模型泛化性还存在较大的提升空间。同时，考虑到在真实情

表2　基于深度学习的三维形状补全主要方法对比

Tab. 2　Comparison of the main methods of 3D shape completion based on deep learning

3D形状补全	代表性方法	描述及局限性	评估数据集	是否监督
基于点	PCN[10]	点云补全开创性工作，遵循编码器-解码器范式，细节生成能力欠缺	ShapeNet-Part（point cloud）、KITTI	是
	TopNet[23]	引入树形解码器进行点云生成，细节信息丢失，需要足够冗余空间	ShapeNet-Part（point cloud）	是
	PF-Net[24]	引入多分辨率编码器和金字塔解码器，仅补全缺失部分，缺乏泛化性	ShapeNet-Part（point cloud）	是
	LAKe-Net[68]	引入非对称关键点定位器和递归细化模块生成细粒度点云，缺乏泛化性	ShapeNet-Part（point cloud）	是
	3D-EPN[69]	分辨率和计算资源呈现正相关，体素表达缺乏细节纹理	ShapeNet-Part（mesh）	是
基于卷积	GRNet[70]	引入体素作为点云的中间表示，点云体素化过程不可避免导致信息丢失	ShapeNet-Part（point cloud）、KITTI	是
	VE-PCN[71]	引入高频边缘结构信息注入到形状补全分支，计算量大	ShapeNet-Part（point cloud）、KITTI	是
	MRAC-Net[72]	引入各向异性卷积编码器同时提取全局特征和局部特征，计算量大	ShapeNet-Part（point cloud）	是
基于图	GCNet[56]	引入带有图卷积的变分自动编码器学习潜在空间表示，缺乏泛化性	DFAUST、MHAD	是
	PC-RGNN[73]	引入基于注意力的多尺度图卷积模块辅助三维检测，缺乏泛化性	KITTI	是
	GGDNet[75]	引入图引导变形模块优化点云补全任务，缺乏泛化性	ShapeNet-Part（mesh）、KITTI	是
	LSLS-Net[76]	引入不同缺失点云共享完整编码的掩码机制，缺乏细节性、忠实性	ShapeNet-Part（point loud）、KITTI	否

续表 2　基于深度学习的三维形状补全主要方法对比

Tab. 2　Comparison of the main methods of 3D shape completion based on deep learning

3D形状补全	代表性方法	描述及局限性	评估数据集	是否监督
基于生成对抗	RL-GAN-Net[77]	引入AE、RL和GAN多阶段协作，补全细节不足	ShapeNet-Part（point cloud）	是
	CRNet[78]	引入级联细化策略和块判别器提升点云的生成质量，缺乏泛化性	ShapeNet-Part（point cloud）	是
	Cycle4.[81]	引入双向循环的无监督点云补全算法，训练过程不易，缺乏细节性	ShapeNet-Part（point cloud）、KITTI	否
	ShapeInve.[60]	将GAN逆映射引入到点云补全，需要额外的生成模型，缺乏泛化性	ShapeNet-Part（point cloud）、KITTI、Matter-port3D	否
基于Transformer	PointTr[82]	引入Transformer进行位置编码，计算量大、部署较难、泛化性差	ShapeNet-Part（point cloud）	是
	SDTNet[83]	引入骨架-细节Transformer，遵循由粗到细生成策略，缺乏泛化性	ShapeNet-Part（point cloud）、KITTI	是
	VRC-Net[84]	引入概率建模和关系增强子网络，细节生成能力提升，缺乏泛化性	ShapeNet-Part（point cloud）、KITTI、ScanNet	是
其他	ViPC[85]	引入额外的单张图像信息辅助点云补全，缺乏泛化性	ShapeNet-ViPC	是
	DeepSDF[65]	引入连续符号距离函数生成密集的形状表面，推理速度慢	ShapeNet-Part（mesh）	是

表 3　Completion3D数据集上部分方法的定量结果

Tab. 3　Quantitative results of partial methods on the Completion3D dataset

代表性方法	是否监督	飞机	橱柜	汽车	椅子	台灯	沙发	桌子	船舰	平均值
PCN[10]	是	9.79	22.70	12.43	25.41	22.72	20.26	20.27	11.73	18.16
TopNet[23]	是	9.29	18.79	11.57	18.44	14.69	18.63	13.45	8.65	14.19
ECG[74]	是	4.99	15.09	8.95	12.86	10.65	12.90	10.03	6.08	10.19
MSN[67]	是	4.91	13.04	10.87	10.62	11.75	11.90	8.72	9.53	10.17
GRNet[70]	是	6.13	16.90	8.27	12.23	10.22	14.93	10.08	5.86	10.57
CRNet[70]	是	3.38	13.17	8.31	10.62	10.00	12.86	9.16	5.80	9.16
VRCNet[84]	是	3.94	10.93	6.44	9.32	8.32	11.35	8.60	5.78	8.09
PointTr[82]	是	4.77	10.45	8.68	9.39	7.77	10.83	7.91	7.19	8.37
SDTNet[83]	是	4.60	10.05	8.16	9.15	8.12	10.65	7.64	7.66	8.25
Cycle4.[81]	是	5.23	14.77	12.41	18.09	17.32	21.06	18.90	11.54	14.92
ShapeInve.[60]	是	5.65	16.11	13.05	15.42	18.06	24.64	16.27	10.13	14.91
LSLS-Net[76]	是	3.90	13.50	8.70	13.90	15.80	14.80	17.10	10.00	12.21

注：1. 评价指标为倒角距离CD$\times 10^4$。2. Completion3D数据集属于ShapeNet-part范畴。

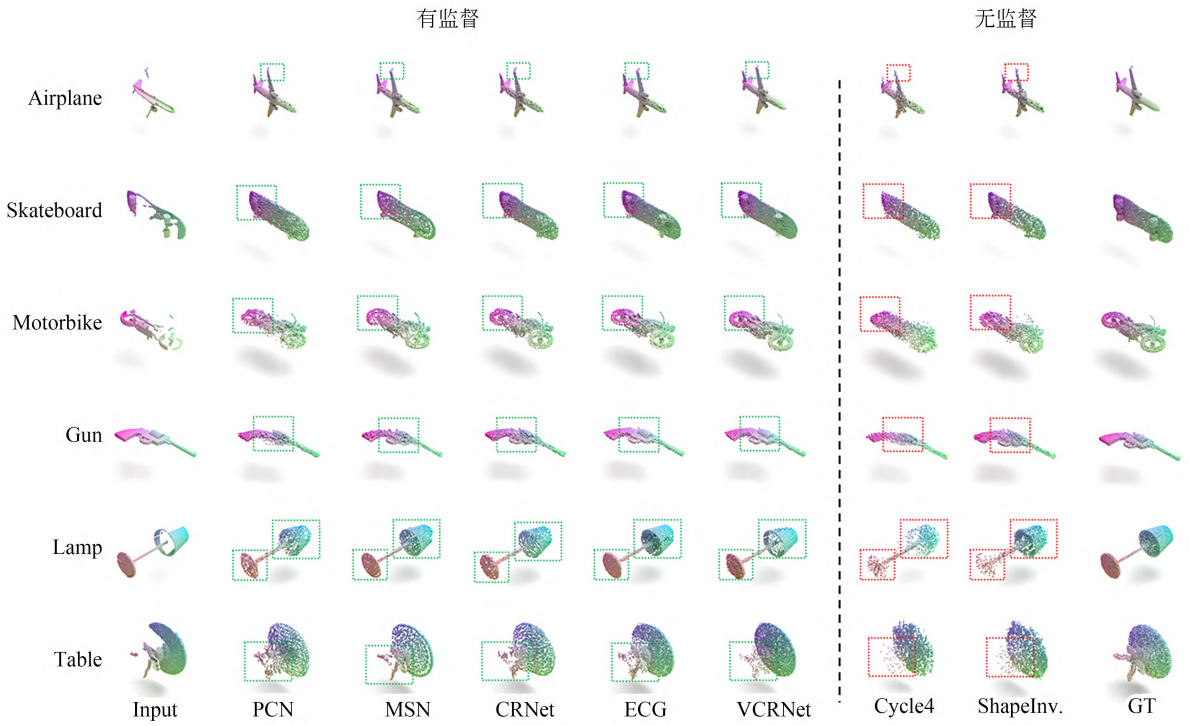

图 4　部分形状补全方法结果对比

Fig. 4　Comparison of the results of some shape completion methods

况下获取的数据是没有对应真值的。因此,无需匹配对的无监督形状补全方法仍然是进一步研究的方向。

(3)现有方法难以对形状细节进行精细补全。同时,很少有方法考虑补全结果的忠实性,即补全生成的点能否忠实地落在真值参考点或面上。在最近的深度隐式重建[65]和点云上采样[93]工作中,忠实性问题有被提到。因此,点云补全的忠实性也是形状补全任务需要考虑的因素。

4　三维场景补全

围绕三维形状补全的研究已经有较多的工作,但关于场景补全的工作仍然较少。一方面原因在于相较于形状补全,场景补全具有补全面积大和补全对象多的特点[28]。另一方面在于场景补全任务希望补全的缺失内容与现有场景内容的语义信息是一致的[32],而这也是场景补全的主要挑战。

尽管场景补全面临着以上挑战,但其中仍不乏优秀的研究工作,根据场景补全算法所遵循的主要策略,可将其归纳为 2 种主要类型:基于模型拟合的场景补全方法[28-30,94]和基于生成式的场景补全方法[25,31-32,95-98],其发展历程如图 5 所示。下面,本文将对发展进程中具有代表性的算法进行介绍和总结。

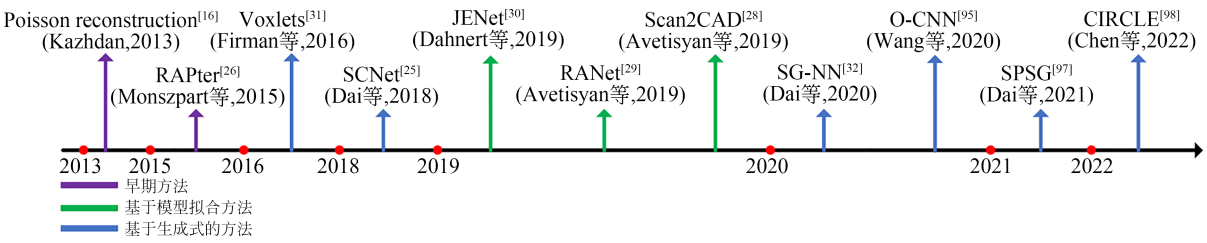

图 5　三维场景补全方法发展历程

Fig. 5　Development history of 3D scene complementary methods

4.1　基于模型拟合的场景补全方法

针对场景缺失区域较小时，可以通过平面拟合[26]和表面插值[16]这类早期方法进行补全。然而，这与艺术家所需求的精细化场景模型相比是远远不够的。一个可行的思路是通过从预先创建的形状数据库中检索一组CAD模型，并将它们与不完整扫描场景中的形状对象进行对齐、替换，以此来得到干净而紧凑的场景表示，如图6所示。

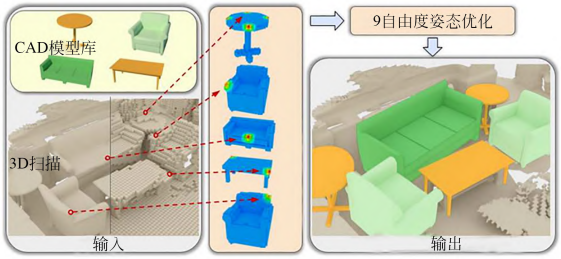

图6　模型拟合场景

Fig. 6　Model fitting scenario

Avetisyan等[28]提出CAD模型对齐的场景补全算法Scan2CAD。首先，将RGB-D扫描的场景数据通过体融合(Volumetric Fusion)方式[99]转换成有符号距离场表示，并使用Batty提供的SDFGen工具包计算CAD模型的无符号距离场。其次，使用3DCNN学习场景对象和CAD模型对象之间的嵌入关系，并预测出对应的热图。最后，基于对应的热图，通过变分优化公式(Variational Optimization Formulation，VOF)优化对齐的结果。该算法在Scan2CAD基准上超越了基于手工特征的方法和基于CNN的方法。Avetisyan等[29]提出一种端到端的CAD模型检索对齐算法RALNet。该算法提出了可微概率对齐策略和对称几何感知策略，并通过全卷积网络(Fully Convolutional Network，FCN)一次对齐场景中检测到的所有对象，在速度上具有较大的提升。Dahnert等[30]提出联合嵌入的场景补全方法，简称JENet。利用堆叠沙漏方法(Stacked Hourglass Approach)从扫描场景中分离出对象并将其转化成类似CAD模型的表示形式，以学习一个共享的嵌入空间用于CAD模型检索。该算法在实例检索精度方面比当时最先进的CAD模型检索算法提高12%。Zeng等[94]提出基于数据驱动的三维匹配描述符3DMatch。该算法通过学习局部空间块的描述符来建立局部三维数据的对应关系。为了获取训练数据，提出了一种自监督的特征学习方法在现有的RGB-D重建结果中获取大量的对应关系。实验结果表明该描述符不仅在重建的局部几何匹配上表现良好，还可以泛化到不同的任务和空间尺度中。

尽管基于模型拟合的场景补全方法取得了一定的进展，然而，这类方法存在固有的自身局限性，主要包括两个方面：

(1)模型库中模型并不能包括真实场景中的所有对象。

(2)模型拟合方法对场景中的实例对象进行补全，但对场景中的背景信息通常不进行补全，例如墙壁和地面。

4.2　基于生成式的场景补全方法

近两年，基于深度学习从部分RGB-D观测信息中生成完整场景的方法显示出较大的研究前景。其中，基于截断符号距离函数(Truncated Signed Distance Function，TSDF)[25]的体素编码是常用的数据处理形式和场景输出表征形式，如图7所示。

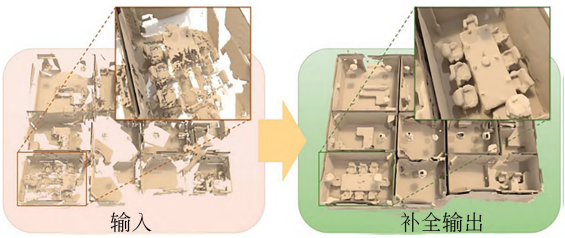

图7　场景生成补全

Fig. 7　Scene generation complementary

Dai等[25]提出能够处理任意比例大小的场景补全网络SCNet。首先，将RGB-D观测的局部场景深度图通过体融合方法生成TSDF编码的场景表示；其次，利用3DCNN进行场景生成补全。其算法补全过程遵循由粗到细的策略，在补全质量和处理速度方面都有大幅度的提升。Firman等[31]提出一种结构化预测的场景补全算法Voxlets。算法核心是使用结构化的随机森林(Structured Random Forest，SRF)从局部观测的深度图中估计出周围表面形状。然而，该算法补全的场景较小，仅适用于桌面大小的场景。Wang等[95]提出基于八叉树卷积神经网络(Oc-

tree-based Convolutional Neural Networks，O-CNN）的场景补全算法。该算法以类似U-Net[100]的结构进行特征提取，并引入以输出为引导的跳跃连接方式来更好地保持输入数据的几何信息。值得一提的是，该算法具有较高的计算效率，并支持深层次的O-CNN结构，在形状补全数据集和场景补全数据集上取得了较好的实验结果。Azinović等[96]同时使用NeRF和TSDF实现高质量的场景表示。该方法具有两个优势：(1)虽然目前使用NeRF的体渲染新视图合成方法显示出了良好的结果，但是NeRF不能重建实际的表面，当使用标记立方体（Marching Cube，MC）提取曲面时，基于密度的曲面体积表示会导致伪影。因此，该方法使用隐式函数来表示场景曲面。在这里，隐式函数为截断符号距离函数。(2)该方法整合了深度先验信息，并提出了姿态和相机细化技术来改善重建质量，在真实数据集ScanNet上取得了较好的场景表示结果。Han等[101]提出基于深度强化学习的场景表面生成算法。该算法创新性地引入了深度强化学习策略来确定场景补全的最优视点序列。此外，为了保证不同视点之间的一致性和更好地利用上下文信息，该算法进一步提出了体素引导的视图补全框架产生高分辨率的场景输出。

尽管上述方法在大规模域内数据集，如SUNCG[33]、ShapeNet[48]和NYUv2[54]上，取得了较好的补全效果，但扩展到其他观测的不完整场景数据集时，由于数据集间的域差距，其补全的效果仍然是有限的。同时，在大部分真实场景下，是没有与之对应的完整场景真实值的。为了解决上述有监督方法的缺陷，一些无监督的场景补全方法被提出。

Dai等[32]首次提出自监督的场景补全算法SG-NN，该算法直接在不完整场景数据上进行训练，其核心思想是在RGB-D扫描的场景信息中移除部分图像以此得到更加不完整的场景信息。然后，通过在这两个不同程度的缺失场景中构建自监督信号进行训练，并最终得到以TSDF表示的高分辨率场景。受SG-NN的启发，Dai等[97]提出一种能够同时补全场景几何信息和颜色信息的自监督算法SPSG。值得一提的是该算法对于几何信息和颜色信息的推断不是依赖于模型补全的3D损失，而是依赖于在模型渲染所得到的2D图像上进行监督引导，这样充分利用了原始RGD-D扫描的高分辨率图像信息。Chen等[98]介绍了一种基于点云中间表示的场景补全框架CIRCLE。该算法首先将RGB-D深度图在已知相机位姿的情况下转化为点云数据，其转换过程遵循Kinectfusion[102]。其次，分别使用Point Encoder、UNet和SDF Decoder进行特征提取和几何补全。最后，使用可微分隐式渲染（Differentiable Implicit Rendering，DIR）模块进行补全细化。该算法不仅具有更好的重建质量，而且在速度上比第2名快10～50倍。

尽管以上无监督场景补全方法在真实数据集上取得了令人振奋的结果，但是他们在复杂场景的生成方面仍存在不足。由于缺乏先验信息的引导，在面对更加复杂的场景时，不同对象的生成结果具有歧义性。

4.3　分析与小结

对比分析不同类型的三维场景补全方法，并根据图8和表4、表5中的部分方法对比结果，得出下列结论：

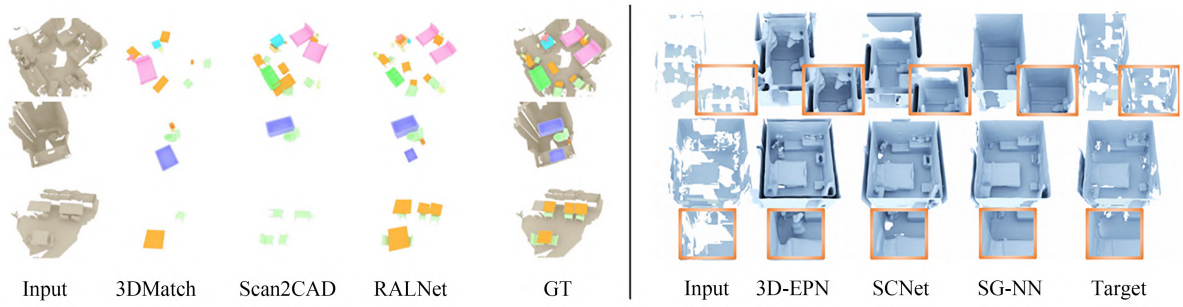

图8　部分场景补全方法结果对比

Fig. 8　Comparison of the results of some scenario complementary methods

表 4　三维场景补全主要方法对比

Tab. 4　Comparison of the main methods of 3D scene completion

3D场景补全	代表性方法	输入数据	描述及局限性	数据集	是否监督
早期方法	RAPter[26]	Points	引入平面排列规则，速度慢，补全面积小	Real-world Scans	是
	Poisson[16]	Points	引入带筛选的泊松方程，补全面积小，补全质量较低	Fandisk、Raptor	是
模型拟合	Scan2CAD[28]	RGB-D→SDF	引入三维卷积学习对应嵌入关系，存在自身局限制，仅能粗略补全	ShapeNet、ScanNet	是
	RALNet[29]	RGB-D→TSDF	引入全卷积网络预测 9 自由度对齐，模型存在自身局限性	ShapeNet、ScanNet	是
	JENet[30]	RGB-D→Occ. Grid	引入沙漏网络进行场景对象分离学习共享嵌入空间，存在局限性	ShapeNet、ScanNet	是
	SCNet[25]	RGB-D→TSDF	遵循由粗到细的场景补全策略，面对其他数据集泛化性差	ScanNet、SUNCG	是
	Voxlets[31]	Depth	引入结构化随机森林，适用于桌面大小场景，场景补全面积受限	NYUv2、tabletop	是
生成式	O-CNN[95]	Points	引入输出引导的跳跃连接策略，缺乏泛化性	ShapeNet、SUNCG	是
	SG-NN[32]	RGB-D→TSDF	引入作用于真实数据的自监督场景补全方法，补全分辨率受限	Matterport3D	否
	SPSG[97]	RGB-D→TSDF	同时补全场景几何和颜色信息的自监督方法，补全分辨率受限	ShapeNet、Matterport3D	否
	CIRCLE[98]	RGB-D→Points	引入 SDF 解码器和可微分隐式渲染，缺乏语义信息辅助	Matterport3D	否

表 5　SUNCG 数据集上部分方法的定量结果

Tab. 5　Quantitative results of partial methods on the SUNCG dataset

代表性方法	L_1误差(整体)	L_1误差(未观测空间)	L_1误差(目标)	L_1误差(预测)
Poisson[16]	0.53	0.51	1.70	1.18
3D-EPN[69]	0.25	0.30	0.65	0.47
SCNet[25]	0.18	0.23	0.53	0.42
SG-NN[32]	0.15	0.16	0.50	0.28

注：L_1距离以 5 cm 体素分辨率为单位。

(1)对于三维场景补全任务，三维 TSDF 矩阵是常用的场景表示形式。相较于模型拟合场景补全方法的自身局限性，基于生成式的场景补全方法表现出更好的优势。

(2)在场景补全任务中，由于合成数据集与真实数据集之间存在域差距，采用直接作用于真实数据集的无监督场景补全方法取得了令人振奋的结果。因此，基于无监督的场景补全方法仍然是接下来的重要研究方向。

(3)现有的场景补全方法大多在室内场景数据集上进行补全，在室外场景上的补全工作相对较少，希望在之后的研究中能有更多关于室外场景的补全工作。

(4)以上的场景补全方法没有考虑语义信息对场景补全的辅助，当补全的场景过于复杂时，补全的精度会下降，因此将语义信息和几何信息

相结合的方式也是进一步研究的方向。这方面的工作在接下来的语义场景补全任务中将会介绍。此外，颜色信息也是场景补全任务中需要考虑的重要因素。

5　三维语义场景补全

三维场景的全面理解对许多应用领域而言都是至关重要的，如机器人感知、自动驾驶、数字孪生等。较早的场景理解工作大多是从语义分割或场景补全的角度分别展开研究。然而，文献[33]表明语义分割和场景补全并不是相互独立的，其语义信息和几何信息是相互交织耦合的，是相互促进的，并由此引出了语义场景补全(Semantic Scene Completion，SSC)的概念。语义场景补全是指从局部观测信息中推断出场景的完整几何信息与语义信息，实现与现实世界更好地交互。

目前，三维语义场景补全方法根据输入数据的不同类型，可以归纳为3种主要类型：基于深度图的语义场景补全方法[33,38-41,103-105]、基于深度图联合彩色图像的语义场景补全方法[106-113]和基于点云的语义场景补全方法[114-117]，其发展历程如图9所示。下面，本文将对研究发展进程中具有代表性的算法进行介绍和总结。

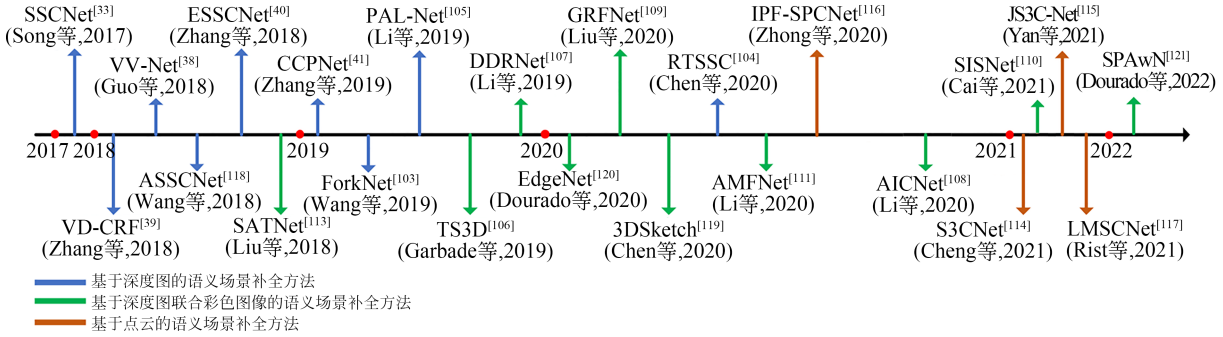

图9　三维语义场景补全方法发展历程

Fig. 9　Development history of 3D semantic scene complementation methods

5.1　基于深度图的语义场景补全方法

Song等[33]开创性地提出语义场景补全网络SSCNet。该网络以单张深度图作为输入，使用扩展上下文卷积模块同时进行场景的体素网格占用和语义标签预测。该算法对深度图的体素编码采用翻转的截断符号距离函数(Flipped Truncated Signed Distance Function，f-TSDF)。普通的TSDF容易在离物体表面较远的地方出现强梯度，基于投影的截断符号距离函数(Projective Truncated Signed Distance Function，p-TSDF)有严重的视角依赖性，而f-TSDF在离物体较近的表面进行强梯度引导，如图10所示。该算法在其提出的SUNCG数据集[33]上取得了当时最好的结果。

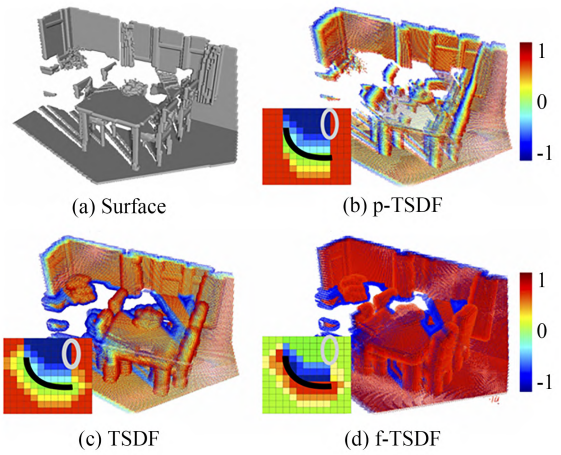

图10　TSDF变体

Fig. 10　TSDF variants

Guo等[38]提出视图-体素卷积网络VV-Net，该网络将2DCNN与3DCNN相结合。相较于SSCNet直接使用3DCNN对TSDF编码的体素网格进行特征提取，VV-Net先使用2DCNN从深度图中提取几何特征并投影为三维体素网格，从而降低了一定的计算量。Zhang等[39]将密集条件随机场(Conditional Random Field，CRF)引入SSC任务中，首先将深度图通过f-TSDF编码为

体素矩阵。其次，将SSCNet输出的概率图与经过CRF处理后的深度图相结合，组成VD-CRF模型。该算法分别在SUNCG、NYUv2和NYUCAD数据集上验证了其有效性，并分别取得了2.5%、3.7%和5.4%的提升。Zhang等[40]提出通过高效的空间分组卷积(Spatial Group Convolution，SGC)来加速密集任务的计算。为了避免3DCNN过大的计算量，目前常用的方法是通过稀疏卷积网络[118]或闵可夫斯基卷积网络[119]进行特征提取。然而，与这些方法不同，SGC是沿着空间维度来创建组，同时使每个组中的体素网格更加稀疏，进一步降低网络的计算量。为了便于对比分析，文中将该方法简称为ESSCNet。Zhang等[41]提出级联上下文金字塔网络CCPNet，该算法不仅改进了金字塔上下文中的标签一致性，还提出了基于引导的残差细化(Guided Residual Refinement，GRR)模块渐进式地恢复场景的精细化结构，在SUNCG和NYUv2数据集上取得了有竞争力的结果，尤其在场景细节的生成方面更具优势。

Wang等[103]提出多分支结构的语义补全网络ForkNet，该网络包括1个共享的编码器分支和3个独立的解码器分支，3个分支分别预测不完整的表面几何形状、完整的几何体积和完整的语义体积。此外，该方法还引入多个判别器来提升语义场景补全任务的准确性和真实性。Chen等[104]提出一种融合特征聚合策略(Feature Aggregation Strategy，FAS)与条件预测模块(Conditioned Prediction Module，CPM)的实时语义场景补全算法RTSSC。首先，该方法通过具有扩张卷积的编码器来获得较大的感受野。其次，利用分阶段FAS融合全局上下文特征。最后，采用逐步CPM进行最终结果预测。该算法在单张1080Ti GPU上实现了110 FPS的速度。Li等[105]提出具有位置重要性感知损失的语义场景补全网络PAL-Net。该算法通过考虑局部各向异性来确定场景内不同位置的重要性，有利于恢复对象的边界信息和场景角落信息。实验表明所提出的位置重要性感知损失在训练过程中收敛速度更快，可以取得更好的性能。

尽管以上基于深度图的语义场景补全算法取得了不错的结果，但RGB图像包含的丰富颜色信息和纹理信息并没有被充分地利用。接下来，将介绍基于深度图联合彩色图像的语义场景补全方法。

5.2 基于深度图联合彩色图像的语义场景补全方法

RGB图像具有丰富的颜色信息和纹理信息，可以作为深度图的重要补充，进一步提升三维语义场景补全的性能。

Garbade等[106]提出基于双流卷积的语义场景补全网络TS3D，该方法首先使用Resnet101[122]对RGB图像进行语义分割。其次，将图像的语义分割结果映射到由深度图生成的3D网格上，得到不完整语义体。最后，使用具有上下文感知的3DCNN推断出完整的语义场景信息。实验表明，引入RGB图像作为输入可以显著地提高SSC任务，在NYUv2数据集上相较于第2名提升了9.4%。Li等[107]提出一种轻量级的维度分解残差网络DDRNet。该方法通过引入维度分解残差(Dimensional Decomposition Residual，DDR)模块降低网络的参数。同时，使用多尺度融合策略提升网络对不同大小物体的适应能力。相较于SSCNet算法，该方法仅使用了21%的参数量。Li等[108]提出各向异性卷积的语义场景补全网络AICNet。相较于标准3DCNN的固定感受野，该算法使用提出的各向异性卷积(Anisotropic Convolution，AIC)模块将三维卷积分解为三个连续的一维卷积实现各向异性的三维感受野，每个一维卷积的核大小是自适应的。实验表明，叠加多个AIC模块，可以进一步提升该算法在SSC任务上的性能。

Liu等[109]提出首个使用门控循环单元(Gated Recurrent Unit，GRU)的语义场景补全的网络GRFNet。该方法根据DDRNet网络进行扩展，改进了多尺度融合策略，并构建具有自主选择和自适应记忆保存的多模态特征融合模块。此外，通过引入非显著性参数融合不同层级特征并进一步提出多阶段的融合策略。该算法在SSC数据融合方面显示出优越的性能。Cai等[110]提出场景到实例与实例到场景的迭代语义补全网络SISNet。具体而言，场景到实例指通过编码实例对象的上下文信息，将实例对象与场景解耦，以此得到更多细节信息的高分辨率对象。而实例

到场景指将细粒度的实例对象重新集成到场景中，从而实现更精确的语义场景完成。该算法在合成数据集和真实数据集上均表现出良好的性能。

Li 等[111]提出基于注意力机制的多模态融合网络 AMFNet。该算法使用 2D 分割结果指导 SSC 任务。值得一提的是，相较于以前直接从深度图中提取几何信息的方法，该方法先将深度图转换成 3 通道的 HHA 编码格式再进行特征提取。HHA 编码图像[112]的三个通道依次代表水平视差、高于地面的高度和像素的局部表面法线与重力方向的倾角，如图 11 所示。该算法在 SUNCG 和 NYUv2 数据集上分别有 2.5% 和 2.6% 的相对增益。Liu 等[113]提出一种解纠缠的语义场景补全网络 SATNet。该方法首先使用编码-解码网络结构得到语义分割图像。其次，通过 2D 到 3D 重投影变换得到不完整场景的语义体素表示。最后，通过 3DCNN 得到完整场景的语义体素表示。实验结果表明该算法在合成数据集和真实数据集上均表现出良好的性能。

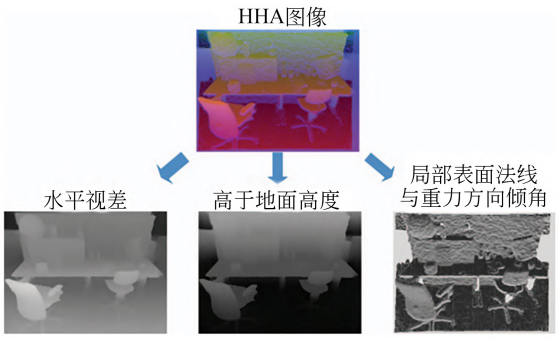

图 11　HHA 图像

Fig. 11　HHA map

尽管基于深度图联合彩色图像的语义场景补全方法具有较好的性能。然而，基于体素的三维表示仍然受到分辨率和内存的限制，当在室外场景下时，其内存缺陷尤其明显。接下来，将介绍基于点云的语义场景补全方法。

5.3　基于点云的语义场景补全方法

Cheng 等[114]提出基于点云输入的语义场景补全网络 S3CNet。由于雷达点云具有的稀疏性导致直接提取其空间特征较为困难。因此，该算法首先将点云依次通过球形投影算法、扩展算法[123]、基于修改的 f-TSDF 编码方法得到高效的稀疏三维张量。其次，将稀疏 3D 张量投影的鸟瞰图进行语义分割。最后，将得到的 2D 分割结果用于强化 3D SSC。值得一提的是该方法对体素的融合是动态的，抵消了对 3D 卷积的显著内存要求。

Yan 等[115]提出上下文形状先验的稀疏雷达点云语义分割框架 JS3C-Net。该方法将雷达获取的多帧进行配准合并，其合并结果不仅可以作为场景补全任务的参考真值，还可以捕获那些显著对象形状的先验信息，而这些得到的形状先验信息有利于进一步的语义分割任务。此外，该算法还引入了点云-体素交互（Point-voxel Interaction，PVI）模块，用于语义分割和语义场景补全之间的隐式信息融合。该算法分别在 SemanticKITTI 和 SemanticPOSS 基准上提升了 4% 和 3%。Zhong 等[116]提出一种融合 RGB 图像纹理信息与点云几何信息的场景补全网络 IPF-SPC-Net。该算法首先使用 2D 分割网络得到语义分割图像。其次，将分割图像的语义信息重投影到对应的点云上，得到包含语义信息的不完整场景点云。最后，再通过基于点云的观测编码器和遮挡解码器得到完整的语义场景补全点云。实验结果表明该算法在场景补全和语义场景补全任务上均表现良好。Rist 等[117]提出基于局部深度隐式函数的语义场景补全网络 LMSCNet。该方法与之前的场景补全方法不同，采用非体素化的连续场景表示，并引入自由空间信息作为监督信号，在室外场景数据集 Semantic KITTI 上得了较好的实验结果。然而，该方法在不确定性估计方面仍然还有提升的空间。

尽管上述基于点云的语义场景补全方法在大规模室外场景数据集上表现出了良好的性能，但这方面的研究工作仍然较少。此外，现有的点云处理方法尚没有统一认可的特征提取范式，还没有一种点云特征提取算法能够像 CNN 或 Transformer 在图像领域获得那么高的认可度。尽管近几年基于点云的特征提取方法蓬勃发展，但是人们依然还是使用相对较早的点云特征提取算法，如：PointNet[61]和 DGCNN[87]。

5.4　分析与小结

分析不同类型的语义场景补全方法，并根据表 6、表 7 和图 12 所示的部分方法对比结果，得出

下列结论：

（1）现有的语义场景补全方法大多是通过3DCNN对体素网格表示的三维空间进行特征提取。尽管其中有些算法使用了稀疏卷积、闵可夫斯基卷积或空间分组卷积来降低参数量，但当场景规模足够大时或者针对室外场景时，其内存和显存上的消耗仍然是致命的。

（2）最近，基于局部深度隐式的语义场景补全工作被提出，然而模型预测的不确定性问题还有待进一步的解决。

表6　三维语义场景补全主要方法对比

Tab. 6　Comparison of the main methods of 3D semantic scene completion

3D语义场景补全	代表性方法	描述及局限性	数据集评估	是否监督
深度图	SSCNet[33]	引入扩展3D卷积和翻转f-TSDF编码，输出分辨率受限	NYUv2、SUNCG	是
	VV-Net[38]	将2D卷积提取的几何特征作为先验信息，输出分辨率较低	NYUv2、SUNCG	是
	VD-CRF[39]	引入密集条件随机场进行有效推理，输出分辨率较低	NYUv2、SUNCG、NYUCAD	是
	ASSCNet[118]	引入多个对抗损失函数学习特征关联，输出分辨率较低	NYUv2、SUNCG	是
	ESSCNet[40]	引入高效空间分组卷积降低网络参数量，输出分辨率较低	NYUv2、SUNCG	是
	CCPNet[41]	引入级联金字塔策略和基于引导的残差细化模块，输出分辨率较低	NYUv2、SUNCG	是
	ForkNet[103]	引入多分支结构生成器和多个判别器模块，输出分辨率较低	NYUv2、SUNCG	是
	RTSSC[104]	引入分阶段的特征聚合策略与条件预测模块，输出分辨率较低	NYUv2、SUNCG、NYUCAD	是
	PAL-Net[105]	引入位置重要性感知损失函数，输出分辨率较低	NYUv2、NYUCAD	是
	TS3D[106]	引入双流卷积网络结构，计算量大，在室外场景受限	NYUv2、NYUCAD	是
	DDRNet[107]	引入轻量级的维度分解残差模块降低网络参数，室外场景受限	NYUv2、NYUCAD	是
	AICNet[108]	引入各向异性卷积模块获取自适应感受野，室外场景受限	NYUv2、NYUCAD	是
深度图＋RGB图像	GRFNet[109]	构建具有自主选择和自适应记忆保存的特征融合模块，室外场景受限	NYUv2、NYUCAD	是
	3DSketch[119]	引入深度信息的几何嵌入策略，室外场景受限	NYUv2、SUNCG、NYUCAD	是
	SISNet[110]	引入场景到实例与实例到场景的迭代策略，室外场景受限	NYUv2、SUNCG、NYUCAD	是
	AMFNet[111]	引入注意力机制的多模态融合策略，室外场景受限	NYUv2、SUNCG	是
点云	S3CNet[114]	点云的语义场景补全，可用于室外场景，缺乏RGB纹理	Semantic KITTI	是
	JS3C-Net[115]	引入上下文形状先验信息，可用于室外，缺乏RGB纹理	Semantic KITTI、SemanticPOSS	是
	IPF-SPCNet[116]	融合RGB图像纹理与点云几何信息，室外场景待验证	NYUv2、NYUCAD	是
	LMSCNet[117]	使用局部深度隐式函数构建场景，可用于室外，不确定性估计待提升	Semantic KITTI	是

表 7　NYUv2 数据集上部分方法的定量结果

Tab. 7　Quantitative results of partial methods on the NYUv2 dataset

代表性方法	分辨率	天花板	地板	墙	窗	椅子	床	沙发	桌子	电视	家具	电视	mIoU
SSCNet[33]	240×60	15.1	94.7	24.4	0.0	12.6	32.1	35.0	13.0	7.8	27.1	10.1	24.7
ESSCNet[40]	240×60	17.5	75.4	25.8	6.7	15.3	53.8	42.4	11.2	0	33.4	11.8	26.7
DDRNet[107]	60×60	21.1	92.2	33.5	6.8	14.8	48.3	42.3	13.2	13.9	35.3	13.2	30.4
VV-Net[38]	120×60	19.3	94.8	28.0	12.2	19.6	57.0	50.5	17.6	11.9	35.6	15.3	32.9
AICNet[108]	60×60	23.2	90.8	32.3	14.8	18.2	51.1	44.8	15.2	22.4	38.3	15.7	33.3
TS3D[106]	240×60	9.7	93.4	25.5	21.0	17.4	55.9	49.2	17.0	27.5	39.4	19.3	34.1
ForkNet[103]	80×80	36.2	93.8	29.2	18.9	17.7	61.6	52.9	23.3	19.5	45.4	20.0	37.1
CCPNet[41]	240×240	23.5	96.3	35.7	20.2	25.8	61.4	56.1	18.1	28.1	37.8	20.1	38.5
3DSketch[119]	60×60	43.1	93.6	40.5	24.3	30.0	57.1	49.3	29.2	14.3	42.5	28.6	41.1
GRFNet[109]	60×60	24.0	91.7	33.3	19.0	18.1	51.9	45.5	13.4	13.3	37.3	15.0	32.9
ISNet[110]	60×60	54.7	93.8	53.2	41.9	43.6	66.2	61.4	38.1	29.8	53.9	40.3	52.4

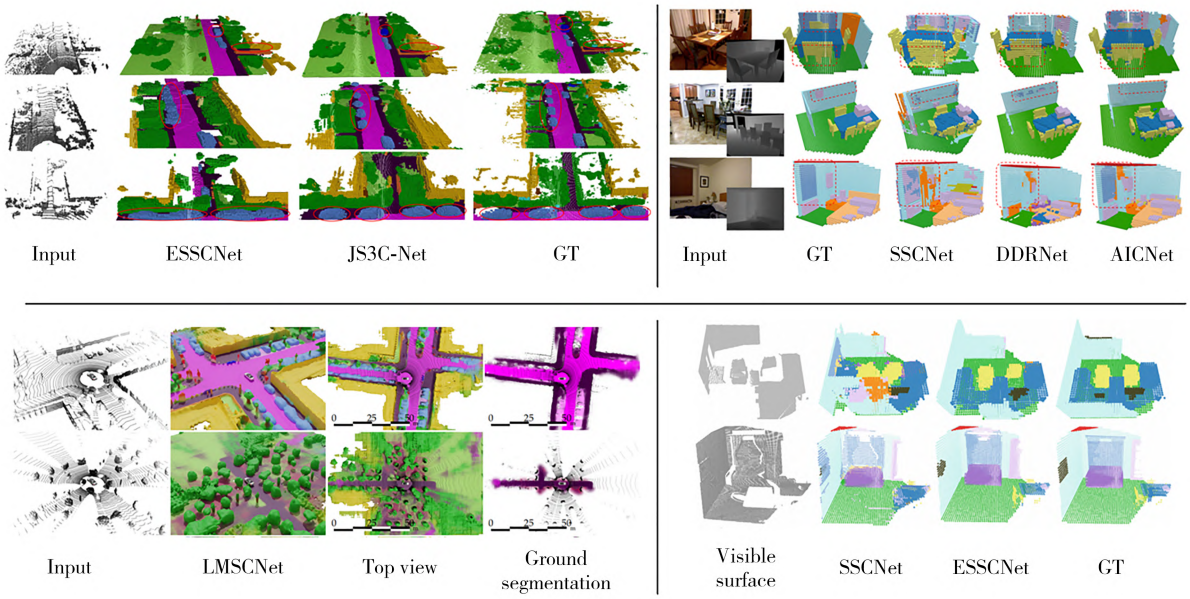

图 12　部分语义场景补全方法结果对比

Fig. 12　Comparison of the results of some semantic scene completion methods

6　面临问题与研究趋势

尽管目前围绕三维形状补全、三维场景补全和三维语义场景补全的研究取得了一定的成果，但在现有的方法中，还存在一些亟待解决的问题，本节将对此进行深入分析，并从技术角度对三维补全未来的发展趋势进行展望。

6.1　三维补全面临的主要问题

(1)几何细节丢失问题：由于局部观测信息缺乏鲁棒的几何约束，现有的三维补全方法在对个体形状或大面积场景进行补全时，往往会丢失细节或无法预测正确的几何信息。尽管已有方法采用注意力机制[124]或 Transformer 模型[82-83]实现细节生成，但其二阶计算量和内存复杂度极大地限制了它的可适用性。

(2)模型泛化性不足问题：现有的大部分三维补全方法存在泛化性差的问题，这主要可以归纳为两方面的因素，(a)依赖于缺失-完整匹配对

的监督训练方式易导致模型过拟合和泛化性差。(b)不同合成数据集之间以及合成数据集与真实数据集之间存在域差距(Domain Gap)。

(3)计算资源受限问题:计算资源包括内存资源和显存资源,其受限的主要因素来源于场景表征的方式和场景数据的规模。由于大部分方法依赖于三维TSDF矩阵或大规模点云来表征场景信息,虽然可以直接或间接使用3DCNN进行特征提取,但是其计算量随分辨率呈立方增加。尽管有研究者通过和空间分组卷积[40]、稀疏卷积[125]和闵可夫斯基卷积[126]来缓解三维卷积参数量的问题,但当场景规模足够大时,其计算资源的缺陷仍然是致命的。此外,部分研究者使用图卷积神经网络对点云进行特征提取,但是这类方法需要耗费较长的模型训练时间和推理时间,对一些计算成本敏感和实时性较高的应用并不友好。

(4)实例区分模糊问题:针对三维形状补全或部分场景补全任务,本文关注到大部分方法都遵循编码器-解码器的范式,而这种范式易导致补全的不同实例存在区分性不足的问题。不同实例包括相同对象类别下的实例和不同对象类别下的实例。

(5)数据集类别不平衡问题:深度学习能够在大规模均衡数据集上取得显著成绩。但现有的三维补全数据集,特别是室外数据集,其类别分布存在严重的不平衡,如:Semantic KITTI数据集[35]。这导致样本量少的类别包含的特征过少,模型学习效果大打折扣,难以完成高质量的补全任务。

6.2 未来的研究方向

针对上述三维补全研究存在的主要问题,并结合实际的应用场景和当下的研究热点,本文提出未来可能的研究方向:

(1)针对几何细节丢失问题,可以从以下两个层面展开研究:(a)常见的三维数据表示形式包括点云、体素和网格。点云表示具有存储空间小和表征能力强的特点,但它不能描述拓扑结构,亦不能产生水密的表面[65]。体素表示易受分辨率和计算存储空间的限制[69]。网格表示易受固定拓扑结构的限制[75]。相较于点云、体素和网格的离散表示形式,隐式函数能够支持以任意分辨率的形状恢复,处理不同的拓扑结构,并且输出结果是连续的。因此,基于隐式函数的三维补全是值得进一步探讨的研究方向。(b)通过引入几何先验信息、语义先验信息、颜色先验信息和场景图信息解开复杂场景中不同对象之间的纠缠和构建不同语义对象之间的关联,从而进一步提升几何信息预测的正确性,是值得进一步探讨的研究方向。

(2)针对模型泛化不足问题,可以从以下三个方面进行探索:(a)由于在真实世界中收集大量完整的3D数据是耗时甚至是不现实的,因此,无需匹配对的无监督三维补全方法仍是接下来值得探讨的方向。(b)对于合成数据集与真实数据集之间存在的域差距,使用无监督域自适应方法缩小域差距是值得进一步探讨的方向。(c)现有的三维补全方法依赖于已经对齐的训练数据进行训练,其测试的数据也需满足与训练数据一致的对齐要求,否则训练好的模型在扩展到未对齐场景中时无法实现有效的补全。因此,实现三维补全算法在未对齐情况下的局部观测输入补全是值得进一步探讨的研究方向。

(3)针对计算资源受限问题,可以分别从以下两个角度深入挖掘:(a)使用隐式函数表征场景信息可以大幅降低内存的占用。但在场景补全任务中,模型预测的不确定性问题还有待进一步的解决。因此,结合概率统计学知识提升补全场景质量是很有价值的研究方向。(b)针对一些计算成本敏感的应用,通过模型压缩[127]或图浓缩技术[88]开发更轻便的实时应用模型是一个有趣和值得探讨研究方向。

(4)针对实例区分模糊问题,可以从以下两个方面进行探索:(a)从对比学习和实例持续学习的角度展开研究,实现不同实例的可区分性是值得进一步探讨的研究方向。(b)引入额外的语义信息来指导三维补全任务是值得进一步探讨的研究方向。例如,如果知道一个椅子缺失腿的数量是4而不是3,那么模型在面对数据分布偏差时将提升预测结果的可靠性。

(5)针对数据集类别的不平衡问题,采用类别再平衡策略和主动学习策略缓解数据集类别的不平衡是值得进一步探讨的方向。

(6)现有的三维补全方法还停留在相对独立

的领域展开研究，结合具体应用场景的工作相对较少。尽管已有相关工作将三维补全技术应用于目标检测这类高阶任务。然而，在高精度地形图构建、数字虚拟人重建、机械臂精确抓取等应用领域，三维补全作为一种辅助技术的潜力还未充分挖掘。因此，基于三维补全技术结合具体应用领域的研究是值得进一步探讨的方向。

7　结　论

三维补全是计算机视觉研究的基础性课题，可以指导多种下游高阶视觉任务的学习，且具有重要的理论意义和广阔的应用前景，已成为计算机视觉领域的研究热点。本文分别从三维形状补全、三维场景补全和三维语义场景补全三方面对近年来的相关研究工作进行了梳理和小结，讨论了现有的三维补全方法所存在的问题，并从技术角度提出了未来的研究趋势。总而言之，深度学习为解决三维补全问题提供了新的技术，取得了较为显著的成果，但将其应用到真实场景中仍然存在很多问题。后续可以在计算资源、模型泛化性、补全质量等方面开展进一步的研究，这对于促进三维视觉领域的发展具有重要的意义。

参考文献：

[1] 陈俊英，白童垚，赵亮. 互注意力融合图像和点云数据的3D目标检测[J]. 光学 精密工程，2021，29(9)：2247-2254.
CHEN J Y，BAI T Y，ZHAO L. 3D object detection based on fusion of point cloud and image by mutual attention[J]. *Opt. Precision Eng.*，2021，29(9)：2247-2254.(in Chinese)

[2] 杨军，张敏敏. 利用模型相似性的三维模型簇协同分割[J]. 光学 精密工程，2021，29(10)：2504-2516.
YANG J，ZHANG M M. Co-segmentation of three-dimensional shape clusters by shape similarity[J]. *Opt. Precision Eng.*，2021，29(10)：2504-2516.(in Chinese)

[3] 刘东生，陈建林，费点，等. 基于深度相机的大场景三维重建[J]. 光学 精密工程，2020，28(1)：234-243.
LIU D S，CHEN J L，FEI D，*et al*. Three-dimensional reconstruction of large-scale scene based on depth camera[J]. *Opt. Precision Eng.*，2020，28(1)：234-243.(in Chinese)

[4] YIN T W，ZHOU X Y，KRÄHENBÜHL P. Center-based 3D object detection and tracking[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Nashville*，*TN*，*USA*. IEEE，2021：11779-11788.

[5] CHEN H S，WANG P C，WANG F，*et al*. EPro-PnP：generalized end-to-end probabilistic perspective-n-points for monocular object pose estimation[C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *New Orleans*，*LA*，*USA*. IEEE，2022：2771-2780.

[6] ASSAEL Y，SOMMERSCHIELD T，PRAG J. Restoring ancient text using deep learning：a case study on Greek epigraphy[C]. *Proceedings of the* 2019 *Conference on Empirical Methods in Natural Language Processing and the* 9*th International Joint Conference on Natural Language Processing* (*EMNLP-IJCNLP*). *Hong Kong*，*China. Stroudsburg*，*PA*，*USA*：*Association for Computational Linguistics*，2019.

[7] XIU Y L，YANG J L，TZIONAS D，*et al*. ICON：implicit clothed humans obtained from normals[C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *New Orleans*，*LA*，*USA*. IEEE，2022：13286-13296.

[8] MYSTAKIDIS S. Metaverse[J]. *Encyclopedia*，2022，2(1)：486-497.

[9] ROLDAO L，DE CHARETTE R，VERROUST-BLONDET A. 3D semantic scene completion：a survey[J]. International Journal of Computer Vision，2022：1-28.

[10] YUAN W T，KHOT T，HELD D，*et al*. PCN：point completion network[C]. 2018 *International Conference on* 3*D Vision* (3*DV*). *Verona*，*Italy*. IEEE，2018：728-737.

[11] MITRA N J，GUIBAS L J，PAULY M. Partial and approximate symmetry detection for 3D geometry[J]. *ACM Transactions on Graphics*，2006，25(3)：560-568.

[12] MITRA N J, PAULY M, WAND M, *et al.*. Symmetry in 3d geometry: Extraction and applications[J]. *Computer Graphics Forum*, 2013, 32(6): 1-23.

[13] KAZHDAN M, BOLITHO M, HOPPE H. Poisson surface reconstruction[C]. *Proceedings of the fourth Eurographics symposium on Geometry processing*, 2006, 7.

[14] LEE S, WOLBERG G, SHIN S Y. Scattered data interpolation with multilevel B-splines[J]. *IEEE Transactions on Visualization and Computer Graphics*, 1997, 3(3): 228-244.

[15] PRICE J R, HAYES M H. Resampling and reconstruction with fractal interpolation functions[J]. *IEEE Signal Processing Letters*, 1998, 5(9): 228-230.

[16] KAZHDAN M, HOPPE H. Screened Poisson surface reconstruction[J]. *ACM Transactions on Graphics*, 2013, 32(3): 1-13.

[17] SHEN C H, FU H, CHEN K, *et al.*. Structure recovery by part assembly[J]. *ACM Transactions on Graphics* (*TOG*), 2012, 31(6): 1-11.

[18] LI Y Y, DAI A, GUIBAS L, *et al.* Database-assisted object retrieval for real-time 3D reconstruction[J]. *Computer Graphics Forum*, 2015, 34(2): 435-446.

[19] ROCK J, GUPTA T, THORSEN J, *et al.* Completing 3D object shape from one depth image[C]. 2015 *IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA*. IEEE, 2015: 2484-2493.

[20] SUN B, KIM V G, AIGERMAN N, *et al.* PatchRD: detail-preserving shape completion by Learning patch retrieval and Deformation[C]. *Computer Vision-ECCV* 2022, 2022: 503-522.

[21] ACHLIOPTAS P, DIAMANTI O, MITLIAGKAS I, *et al.* Learning representations and generative models for 3d point clouds[C]. *Proceedings of the International Conference on Machine Learning*, 2018: 40-49.

[22] YANG Y Q, FENG C, SHEN Y R, *et al.* FoldingNet: point cloud auto-encoder via deep grid deformation[EB/OL]. 2017: *arXiv*: 1712.07262[cs.CV]. https://arxiv.org/abs/1712.07262

[23] TCHAPMI L P, KOSARAJU V, REZATOFIGHI H, *et al.* TopNet: structural point cloud decoder[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Long Beach, CA, USA*. IEEE, 2019: 383-392.

[24] HUANG Z T, YU Y K, XU J W, *et al.* PF-net: point fractal network for 3D point cloud completion[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Seattle, WA, USA*. IEEE, 2020: 7659-7667.

[25] DAI A, RITCHIE D, BOKELOH M, *et al.* ScanComplete: large-scale scene completion and semantic segmentation for 3D scans[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA*. IEEE, 2018: 4578-4587.

[26] MONSZPART A, MELLADO N, BROSTOW G J, *et al.* RAPter: rebuilding man-made scenes with regular arrangements of planes[J]. *ACM Transactions on Graphics*, 2015, 34(4): 103:1-103:12.

[27] ZHAO W, GAO S M, LIN H W. A robust hole-filling algorithm for triangular mesh[J]. *The Visual Computer*, 2007, 23(12): 987-997.

[28] AVETISYAN A, DAHNERT M, DAI A, *et al.* Scan2CAD: learning CAD model alignment in RGB-D scans[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Long Beach, CA, USA*. IEEE, 2019: 2609-2618.

[29] AVETISYAN A, DAI A, NIESSNER M. End-to-end CAD model retrieval and 9DoF alignment in 3D scans[C]. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *Seoul, Korea* (*South*). IEEE, 2019: 2551-2560.

[30] DAHNERT M, DAI A, GUIBAS L, *et al.* Joint embedding of 3D scan and CAD objects[C]. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *Seoul, Korea* (*South*). IEEE, 2019: 8748-8757.

[31] FIRMAN M, AODHA O M, JULIER S, *et al.* Structured prediction of unobserved voxels from a single depth image[C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA*. IEEE, 2016: 5431-5440.

[32] DAI A, DILLER C, NIESSNER M. SG-NN: sparse generative neural networks for self-supervised scene completion of RGB-D scans[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA*. IEEE, 2020: 846-855.

[33] SONG S R, YU F, ZENG A, *et al*. Semantic scene completion from a single depth image[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA*. IEEE, 2017: 190-198.

[34] ARMENI I, SAX S, ZAMIR A R, *et al*. Joint 2D-3D-semantic data for indoor scene understanding [EB/OL]. 2017: *arXiv*: 1702.01105 [cs. CV]. https://arxiv.org/abs/1702.01105

[35] BEHLEY J, GARBADE M, MILIOTO A, *et al*. SemanticKITTI: a dataset for semantic scene understanding of LiDAR sequences [C]. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South)*. IEEE, 2019: 9296-9306.

[36] GRIFFITHS D, BOEHM J. SynthCity: a large scale synthetic point cloud[EB/OL]. 2019: *arXiv*: 1907.04758[cs. CV]. https://arxiv.org/abs/1907.04758

[37] PAN Y C, GAO B, MEI J L, *et al*. SemanticPOSS: a point cloud dataset with large quantity of dynamic instances[C]. 2020 *IEEE Intelligent Vehicles Symposium. Las Vegas, NV, USA*. IEEE, 2020: 687-693.

[38] GUO Y X, TONG X. View-volume network for semantic scene completion from a single depth image [EB/OL]. 2018: *arXiv*: 1806.05361 [cs. CV]. https://arxiv.org/abs/1806.05361

[39] ZHANG L. Semantic scene completion with dense CRF from a single depth image[J]. *Neurocomputing*, 2018, 318: 182-195.

[40] ZHANG J H, ZHAO H, YAO A B, *et al*. Efficient semantic scene completion network with spatial group convolution[C]. *Computer Vision-ECCV* 2018, 2018: 733-749.

[41] ZHANG P P, LIU W, LEI Y J, *et al*. Cascaded context pyramid for full-resolution 3D semantic scene completion[C]. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South)*. IEEE, 2019: 7800-7809.

[42] GUO Y L, WANG H Y, HU Q Y, *et al*. Deep learning for 3D point clouds: a survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(12): 4338-4364.

[43] ARNOLD E, AL-JARRAH O Y, DIANATI M, *et al*. A survey on 3D object detection methods for autonomous driving applications[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 20(10): 3782-3795.

[44] XIU H Y, VINAYARAJ P, KIM K S, *et al*. 3D semantic segmentation for high-resolution aerial survey derived point clouds using deep learning [C]. *SIGSPATIAL '18: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 2018: 588-591.

[45] JIN Y W, JIANG D Q, CAI M. 3D reconstruction using deep learning: a survey[J]. *Communications in Information and Systems*, 2020, 20(4): 389-413.

[46] BRESSON G, ALSAYED Z, YU L, *et al*. Simultaneous localization and mapping: a survey of current trends in autonomous driving [J]. *IEEE Transactions on Intelligent Vehicles*, 2017, 2(3): 194-220.

[47] FEI B, YANG W D, CHEN W M, *et al*. Comprehensive review of deep learning-based 3D point cloud completion processing and analysis [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(12): 22862-22883.

[48] CHANG A X, FUNKHOUSER T, GUIBAS L, *et al*. . Shapenet: An information-rich 3d model repository[EB/OL]. 2015: arXiv: 1512.03012[CS. CV]. http://arXiv.org/abs/1512.03012.

[49] OSADA R, FUNKHOUSER T, CHAZELLE B, *et al*. Matching 3D models with shape distributions[C]. *Proceedings International Conference on Shape Modeling and Applications. Genova, Italy*. IEEE, 2001: 154-166.

[50] GEIGER A, LENZ P, STILLER C, *et al*. Vision meets robotics: the KITTI dataset[J]. *The International Journal of Robotics Research*, 2013, 32(11): 1231-1237.

[51] DAI A, CHANG A X, SAVVA M, *et al*. Scan-

Net: richly-annotated 3D reconstructions of indoor scenes[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA*. IEEE, 2017: 2432-2443.

[52] CHANG A, DAI A, FUNKHOUSER T, *et al*. Matterport3D: learning from RGB-D data in indoor environments[C]. 2017 *International Conference on 3D Vision (3DV). Qingdao, China*. IEEE, 2017: 667-676.

[53] BOGO F, ROMERO J, PONS-MOLL G, *et al*. Dynamic FAUST: registering human bodies in motion[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA*. IEEE, 2017: 5573-5582.

[54] OFLI F, CHAUDHRY R, KURILLO G, *et al*. Berkeley MHAD: a comprehensive multimodal human action database[C]. 2013 *IEEE Workshop on Applications of Computer Vision. Clearwater Beach, FL, USA*. IEEE, 2013: 53-60.

[55] WANG W Y, HUANG Q G, YOU S Y, *et al*. Shape inpainting using 3D generative adversarial network and recurrent convolutional networks[C]. 2017 *IEEE International Conference on Computer Vision. Venice, Italy*. IEEE, 2017: 2317-2325.

[56] LITANY O, BRONSTEIN A, BRONSTEIN M, *et al*. Deformable shape completion with graph convolutional autoencoders[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA*. IEEE, 2018: 1886-1895.

[57] SHU D, PARK S W, KWON J. 3D point cloud generative adversarial network based on tree structured graph convolutions[C]. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South)*. IEEE, 2019: 3858-3867.

[58] WU T, PAN L, ZHANG J, *et al*. Density-aware Chamfer Distance as a Comprehensive Metric for Point Cloud Completion[EB/OL]. 2021: arXiv: 2111.12702 [CS. CV]. http://arXiv. org/abs/2111.12702.

[59] CHEN Z Q, ZHANG H. Learning implicit fields for generative shape modeling[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA*. IEEE, 2019: 5932-5941.

[60] ZHANG J Z, CHEN X Y, CAI Z A, *et al*. Unsupervised 3D shape completion through GAN inversion[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA*. IEEE, 2021: 1768-1777.

[61] CHARLES R Q, HAO S, MO K C, *et al*. PointNet: deep learning on point sets for 3D classification and segmentation[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA*. IEEE, 2017: 77-85.

[62] WU Z R, SONG S R, KHOSLA A, *et al*. 3D ShapeNets: a deep representation for volumetric shapes[C]. 2015 *IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA*. IEEE, 2015: 1912-1920.

[63] MASCI J, BOSCAINI D, BRONSTEIN M M, *et al*. Geodesic convolutional neural networks on Riemannian manifolds[C]. 2015 *IEEE International Conference on Computer Vision Workshop. Santiago, Chile*. IEEE, 2015: 832-840.

[64] MESCHEDER L, OECHSLE M, NIEMEYER M, *et al*. Occupancy networks: learning 3D reconstruction in function space[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA*. IEEE, 2019: 4455-4465.

[65] PARK J J, FLORENCE P, STRAUB J, *et al*. DeepSDF: learning continuous signed distance functions for shape representation[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA*. IEEE, 2019: 165-174.

[66] MILDENHALL B, SRINIVASAN P P, TANCIK M, *et al*. NeRF: representing scenes as neural radiance fields for view synthesis[C]. *Computer Vision-ECCV* 2020, 2020: 405-421.

[67] LIU M, SHENG L, YANG S, *et al*. Morphing and sampling network for dense point cloud completion[C]. *Proceedings of the AAAI conference on artificial intelligence*, 2020, 34(07): 11596-11603.

[68] TANG J S, GONG Z J, YI R, *et al*. LAKe-net: topology-aware point cloud completion by localizing aligned keypoints[C]. 2022 *IEEE/CVF Con-*

ference on Computer Vision and Pattern Recognition (*CVPR*). *New Orleans*, *LA*, *USA*. IEEE, 2022: 1716-1725.

[69] DAI A, QI C R, NIEßNER M. Shape completion using 3D-encoder-predictor CNNs and shape synthesis[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition. Honolulu*, *HI*, *USA*. IEEE, 2017: 6545-6554.

[70] XIE H Z, YAO H X, ZHOU S C, *et al*. *GRNet: Gridding Residual Network for Dense Point Cloud Completion*[M]. Computer Vision-ECCV 2020. Cham: Springer International Publishing, 2020: 365-381.

[71] WANG X G, ANG M H, LEE G H. Voxel-based network for shape completion by leveraging edge generation[C]. 2021 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *Montreal*, *QC*, *Canada*. IEEE, 2021: 13169-13178.

[72] LIU S, LI D D, HUANG W H, *et al*. MRAC-net: multi-resolution anisotropic convolutional network for 3D point cloud completion[C]. *PRICAI* 2021: *Trends in Artificial Intelligence*, 2021: 403-414.

[73] ZHANG Y N, HUANG D, WANG Y H. PC-RGNN: point cloud completion and graph neural network for 3D object detection[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(4): 3430-3437.

[74] PAN L. ECG: edge-aware point cloud completion with graph convolution[J]. *IEEE Robotics and Automation Letters*, 2020, 5(3): 4392-4398.

[75] SHI J Q, XU L Y, HENG L, *et al*. Graph-guided deformation for point cloud completion[J]. *IEEE Robotics and Automation Letters*, 2021, 6(4): 7081-7088.

[76] CAI Y J, LIN K Y, ZHANG C, *et al*. Learning a structured latent space for unsupervised point cloud completion[C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *New Orleans*, *LA*, *USA*. IEEE, 2022: 5533-5543.

[77] SARMAD M, LEE H J, KIM Y M. RL-GAN-net: a reinforcement learning agent controlled GAN network for real-time point cloud shape completion[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Long Beach*, *CA*, *USA*. IEEE, 2019: 5891-5900.

[78] WANG X G, ANG M H, LEE G H. Cascaded refinement network for point cloud completion[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Seattle*, *WA*, *USA*. IEEE, 2020: 787-796.

[79] HU T, HAN Z Z, SHRIVASTAVA A, *et al*. Render4Completion: synthesizing multi-view depth maps for 3D shape completion[C]. 2019 *IEEE/CVF International Conference on Computer Vision Workshop* (*ICCVW*). *Seoul*, *Korea* (*South*). IEEE, 2019: 4114-4122.

[80] XIE C L, WANG C X, ZHANG B, *et al*. Style-based point generator with adversarial rendering for point cloud completion[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Nashville*, *TN*, *USA*. IEEE, 2021: 4617-4626.

[81] WEN X, HAN Z Z, CAO Y P, *et al*. Cycle4Completion: unpaired point cloud completion using cycle transformation with missing region coding[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Nashville*, *TN*, *USA*. IEEE, 2021: 13075-13084.

[82] YU X M, RAO Y M, WANG Z Y, *et al*. PoinTr: diverse point cloud completion with geometry-aware transformers[C]. 2021 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *Montreal*, *QC*, *Canada*. IEEE, 2021: 12478-12487.

[83] ZHANG W, ZHOU H, DONG Z, *et al*. Point cloud completion via skeleton-detail transformer[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[84] PAN L, CHEN X Y, CAI Z A, *et al*. Variational relational point completion network[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Nashville*, *TN*, *USA*. IEEE, 2021: 8520-8529.

[85] ZHANG X C, FENG Y T, LI S Q, *et al*. View-guided point cloud completion[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern*

Recognition (*CVPR*). *Nashville*, *TN*, *USA*. IEEE, 2021: 15885-15894.

[86] KOUSHIK J. Understanding convolutional neural networks [EB/OL]. 2016: arXiv: 1605.09081 [CS.CV]. http://arXiv.org/abs/1605.09081.

[87] WANG Y, SUN Y B, LIU Z W, *et al*. Dynamic graph CNN for learning on point clouds[J]. *ACM Transactions on Graphics*, 2019, 38(5): 1-12.

[88] JIN W, ZHAO L, ZHANG S, *et al*.. Graph condensation for graph neural networks [EB/OL]. 2021:arXiv: 2110.07580[CS.CV]. http://arXiv.org/abs/2110.07580.

[89] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks [C]. *International Conference on Machine Learning. PMLR*, 2017: 214-223.

[90] HAN K, WANG Y, CHEN H, *et al*. A survey on vision transformer [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[91] ZHAO H S, JIANG L, JIA J Y, *et al*. Point transformer [C]. 2021 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *Montreal*, *QC*, *Canada*. IEEE, 2021: 16239-16248.

[92] GUO M H, CAI J X, LIU Z N, *et al*. PCT: point cloud transformer[J]. *Computational Visual Media*, 2021, 7(2): 187-199.

[93] LI R H, LI X Z, HENG P A, *et al*. Point cloud upsampling via disentangled refinement [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Nashville*, *TN*, *USA*. IEEE, 2021: 344-353.

[94] ZENG A, SONG S R, NIEßNER M, *et al*. 3DMatch: learning local geometric descriptors from RGB-D reconstructions[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition. Honolulu*, *HI*, *USA*. IEEE, 2017: 199-208.

[95] WANG P S, LIU Y, TONG X. Deep octree-based CNNs with output-guided skip connections for 3D shape and scene completion [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*). *Seattle*, *WA*, *USA*. IEEE, 2020: 1074-1081.

[96] AZINOVIĆ D, MARTIN-BRUALLA R, GOLDMAN D B, *et al*. Neural RGB-D surface reconstruction [C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *New Orleans*, *LA*, *USA*. IEEE, 2022: 6280-6291.

[97] DAI A, SIDDIQUI Y, THIES J, *et al*. SPSG: self-supervised photometric scene generation from RGB-D scans [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Nashville*, *TN*, *USA*. IEEE, 2021: 1747-1756.

[98] CHEN H X, HUANG J H, MU T J, *et al*. CIRCLE: convolutional implicit reconstruction and-Completion forLarge-scale indoor scene[C]. *Computer Vision-ECCV* 2022, 2022.

[99] CURLESS B, LEVOY M. A volumetric method for building complex models from range images [C]. *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. 1996: 303-312.

[100] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]. *Medical Image Computing and Computer-Assisted Intervention-MICCAI* 2015, 2015: 234-241.

[101] HAN X G, ZHANG Z X, DU D, *et al*. Deep reinforcement learning of volume-guided progressive view inpainting for 3D point scene completion from a single depth image[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Long Beach*, *CA*, *USA*. IEEE, 2019: 234-243.

[102] NEWCOMBE R A, IZADI S, HILLIGES O, *et al*. KinectFusion: Real-time dense surface mapping and tracking [C]. 2011 10*th IEEE International Symposium on Mixed and Augmented Reality. Basel*, *Switzerland*. IEEE, 2011: 127-136.

[103] WANG Y D, TAN D J, NAVAB N, *et al*. ForkNet: multi-branch volumetric semantic completion from a single depth image [C]. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *Seoul*, *Korea* (*South*). IEEE, 2019: 8607-8616.

[104] CHEN X K, XING Y J, ZENG G. Real-time semantic scene completion via feature aggregation

and conditioned prediction[C]. 2020 *IEEE International Conference on Image Processing. Abu Dhabi, United Arab Emirates*. IEEE, 2020: 2830-2834.

[105] LI J, LIU Y, YUAN X, *et al*. Depth based semantic scene completion with position importance aware loss[J]. *IEEE Robotics and Automation Letters*, 2020, 5(1): 219-226.

[106] GARBADE M, CHEN Y T, SAWATZKY J, *et al*. Two stream 3D semantic scene completion [C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Long Beach, CA, USA*. IEEE, 2019: 416-425.

[107] LI J, LIU Y, GONG D, *et al*. RGBD based dimensional decomposition residual network for 3D semantic scene completion [C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA*. IEEE, 2019: 7685-7694.

[108] LI J, HAN K, WANG P, *et al*. Anisotropic convolutional networks for 3D semantic scene completion [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA*. IEEE, 2020: 3348-3356.

[109] LIU Y, LI J, YAN Q, *et al*. 3D gated recurrent fusion for semantic scene completion[EB/OL]. 2020: arXiv: 2002.07269[CS. CV]. http://arXiv.org/abs/2002.07269.

[110] CAI Y J, CHEN X S, ZHANG C, *et al*. Semantic scene completion via integrating instances and scene in-the-loop[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA*. IEEE, 2021: 324-333.

[111] LI S Q, ZOU C Q, LI Y P, *et al*. Attention-based multi-modal fusion network for semantic scene completion[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 11402-11409.

[112] ZHANG W, LIU G L, TIAN G H. HHA-based CNN image features for indoor loop closure detection [C]. 2017 *Chinese Automation Congress (CAC). Jinan, China*. IEEE, 2017: 4634-4639.

[113] LIU S, HU Y, ZENG Y, *et al*. See and think: Disentangling semantic scene completion[J]. *Advances in Neural Information Processing Systems*, 2018, 31.

[114] CHENG R, AGIA C, REN Y, *et al*. S3CNet: A Sparse Semantic Scene Completion Network for LiDAR Point Clouds[J]. *Conference on Robot Learning. PMLR*, 2021: 2148-2161.

[115] YAN X, GAO J T, LI J, *et al*. Sparse single sweep LiDAR point cloud segmentation via learning contextual shape priors from scene completion [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(4): 3101-3109.

[116] ZHONG M, ZENG G. *Semantic Point Completion Network for 3D Semantic Scene Completion* [M]. ECAI 2020. IOS Press, 2020: 2824-2831.

[117] RIST C B, EMMERICHS D, ENZWEILER M, *et al*. Semantic scene completion using local deep implicit functions on LiDAR data[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(10): 7205-7218.

[118] WANG Y D, TAN D J, NAVAB N, *et al*. Adversarial semantic scene completion from a single depth image[C]. 2018 *International Conference on 3D Vision (3DV). Verona, Italy*. IEEE, 2018: 426-434.

[119] CHEN X K, LIN K Y, QIAN C, *et al*. 3D sketch-aware semantic scene completion via semi-supervised structure prior[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA*. IEEE, 2020: 4192-4201.

[120] DOURADO A, DE CAMPOS T E, KIM H, *et al*. EdgeNet: Semantic scene completion from a single RGB-D image[C]. 2020 25*th International Conference on Pattern Recognition (ICPR).Milan, Italy*.IEEE, 2021: 503-510.

[121] DOURADO A, GUTH F, CAMPOS DE T. Data augmented 3D semantic scene completion with 2D segmentation priors[C]. 2022 *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA*. IEEE, 2022, pp: 687-696.

[122] HE K M, ZHANG X Y, REN S Q, *et al*. Deep residual learning for image recognition[C]. 2016

IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. IEEE, 2016: 770-778.

[123] KU J, HARAKEH A, WASLANDER S L. In defense of classical image processing: fast depth completion on the CPU[C]. 2018 15*th Conference on Computer and Robot Vision* (*CRV*). *Toronto, ON, Canada*. IEEE, 2018: 16-22.

[124] WEN X, LI T Y, HAN Z Z, *et al*. Point cloud completion by skip-attention network with hierarchical folding[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Seattle, WA, USA*. IEEE, 2020: 1936-1945.

[125] GRAHAM B, ENGELCKE M, VAN DER MAATEN L. 3D semantic segmentation with submanifold sparse convolutional networks[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA*. IEEE, 2018: 9224-9232.

[126] CHOY C, GWAK J, SAVARESE S. 4D spatio-temporal ConvNets: minkowski convolutional neural networks[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Long Beach, CA, USA*. IEEE, 2019: 3070-3079.

[127] SUN X L, HASSANI A, WANG Z Y, *et al*. DiSparse: disentangled sparsification for multi-task model compression[C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *New Orleans, LA, USA*. IEEE, 2020: 12372-12382.

作者简介：

肖海鸿(1995－)，男，四川达州人，博士研究生，2021年于南京农业大学获得硕士学位，主要从事三维视觉与场景表示学习方面的研究。E-mail: auhhxiao@mail.scut.edu.cn

通讯作者：

康文雄(1976－)，男，湖南新化人，教授，博士生导师，2003年于西北工业大学获得硕士学位，2009年于华南理工大学获得博士学位，主要从事图像处理、模式识别和计算机视觉等方面的研究。E-mail: auwxkang@scut.edu.cn