基于深度学习的三维数据分析理解方法研究综述

李海生[1),2)] 武玉娟[1),2)] 郑艳萍[1),2)] 吴晓群[1),2)] 蔡 强[1),2)] 杜军平[3)]

[1)](北京工商大学计算机与信息工程学院 北京 100048)

[2)](食品安全大数据技术北京市重点实验室 北京 100048)

[3)](北京邮电大学计算机学院 北京 100876)

摘 要 基于深度学习的三维数据分析理解是数字几何领域的一个研究热点.不同于基于深度学习的图像分析理解,基于深度学习的三维数据分析理解需要解决的首要问题是数据表达的多样性.相较于规则的二维图像,三维数据有离散表达和连续表达的方法,目前基于深度学习的相关工作多基于三维数据的离散表示,不同的三维数据表达方法与不同的数字几何处理任务对深度学习网络的要求也不同.本文首先汇总了常用的三维数据集与特定任务的评价指标,并分析了三维模型特征描述符.然后从特定任务出发,就不同的三维数据表达方式,对现有的基于深度学习的三维数据分析理解网络进行综述,对各类方法进行对比分析,并从三维数据表达方法的角度进一步汇总现有工作.最后基于国内外研究现状,讨论了亟待解决的挑战性问题,展望了未来发展的趋势.

关键词 三维数据分析理解;深度学习;单个模型;场景模型;特征提取

中图法分类号 TP391 **DOI 号** 10.11897/SP.J.1016.2020.00041

A Survey of 3D Data Analysis and Understanding Based on Deep Learning

LI Hai-Sheng[1),2)] WU Yu-Juan[1),2)] ZHENG Yan-Ping[1),2)] WU Xiao-Qun[1),2)]
CAI Qiang[1),2)] DU Jun-Ping[3)]

[1)](*School of Computer and Information Engineering, Beijing Technology and Business University, Beijing* 100048)

[2)](*Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing* 100048)

[3)](*School of Computer Science, Beijing University of Posts and Telecommunications, Beijing* 100876)

Abstract 3D data analysis and understanding based on deep learning is a research hotspot in the field of digital geometry. Increasingly rich three-dimensional data, including single models and scene models, encourages us to use these abundant data to effectively process and analyze digital geometric models, such as 3D object classification, 3D object recognition, 3D shape retrieval, 3D shape segmentation, 3D shape matching and 3D modeling. How to obtain the information we need by analyzing large-scale 3D data is the key to solve tasks related to the fields of computer graphics and computer vision. With the successful application of deep learning in computer vision, it is imperative to extend it to the field of digital geometry processing. 3D data processing method based on deep learning is data-driven. It is no longer limited to a single three-dimensional model. Instead, a set of three-dimensional models is analyzed. Unlike image analysis and understanding based on deep learning, the key problem that needs to be solved of 3D data analysis and

收稿日期:2018-06-19;在线出版日期:2019-07-09.本课题得到国家自然科学基金(61877002,61532006,61602015)、北京市自然科学基金(4172013)和北京市教委科研团队建设项目(PXM2019_014213_000007)资助.**李海生**,博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为计算机图形学、数字几何处理和科学可视化.E-mail:lihsh@btbu.edu.cn.**武玉娟**,硕士研究生,中国计算机学会(CCF)会员,主要研究方向为计算机图形学和数字几何处理.**郑艳萍**,硕士研究生,中国计算机学会(CCF)会员,主要研究方向为计算机图形学和数字几何处理.**吴晓群**,博士,副教授,中国计算机学会(CCF)会员,主要研究方向为计算机图形学、数字几何处理和图像处理.**蔡 强**,博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为计算机图形学、数字几何处理和科学可视化.**杜军平**,博士,教授,中国计算机学会(CCF)杰出会员,主要研究领域为运动图像处理、社交网络分析与搜索和多源数据融合与大数据挖掘.

understanding based on deep learning is the diversity of data. Compared with regular two-dimensional images, the expression of three-dimensional data is diverse. The current related work is mostly based on the discrete representation of three-dimensional data. Different three-dimensional data representation methods and different digital geometry processing tasks have different requirements for deep learning networks. The method based on deep learning can extract feature mapping relationships and semantic correlations between these three-dimensional objects. The characteristics of the 3D model are learned, so that the attributes of the 3D model and the relationships between them can be effectively derived. There are some methods including extracting high-level features based on low-level features, structured representation of 3D data, dimensionality reduction for 3D data, fusion of multimodal features, and the method based on manifold and so on. In this paper, a comprehensive and in-depth review of 3D data processing based on deep learning is provided. We first summarize the commonly used 3D datasets and evaluation indicators of specific tasks, and investigate the 3D model feature descriptors. Then, starting from specific tasks, the existing three-dimensional data analysis and understanding network based on deep learning is reviewed according to different input data representation. Then we further summarize existing work from the perspective of 3D data representation. Meanwhile, the comparison, advantages and disadvantages of various methods are summarized. Finally, based on the research status at home and abroad, the problems existing in the existing research are expounded, and the future development trend is forecasted.

Keywords 3D data analysis and understanding; deep learning; single model; scene model; feature extraction

1 引 言

人类的视觉感知是立体的,相较二维数据,三维数据具有更强的细节感与质感[1],在文物保护[2]、城市规划、工程制图、生物医学和影视娱乐等领域具有广泛的应用[1].随着互联网的发展,越来越多的三维模型库可以在网上免费获取[3],三维扫描技术和虚拟现实技术的发展也使得三维模型的数量和质量不断提高[4].三维数据的日益丰富,包括单个模型和场景模型,鼓励研究者们利用这些丰富的数据来对数字化几何模型进行有效的处理与分析,比如:三维物体分类、三维物体识别、三维模型检索、三维模型分割、三维模型匹配和三维场景重建等[5].如何对数量庞大的三维数据进行快速有效且鲁棒的处理、分析与理解,已成为数字几何领域的研究热点[4].

近几年,深度学习在二维视觉领域得到了较为广泛的应用,尤其是在二维图像特征提取方面表现出了卓越的性能.自动提取到的二维图像特征在大多数图像分析和理解任务中相较于传统解决方案取得了很好的效果.深度学习模型的核心思想是采用数据驱动的方式,通过多层非线性运算单元,将低层运算单元的输出作为高层运算单元的输入,从原始数据中提取由低层到高层、由一般到抽象的特征[6-7].与传统的三维数据分析理解方法相比,基于深度学习的方法是数据驱动的,它不再局限于单个三维模型,而是通过分析一组三维模型,提取这组三维物体之间的特征映射关系及语义相关性,能够从数据计算模型中学习得到三维模型的特征,从而有效推导出三维模型的属性及其之间的关系,而不依赖于硬编码规则[5].深度学习在计算机视觉领域表现出了良好的效果.将深度学习推广到数字几何处理领域是目前的一个研究热点[8].Mitra 等人[9]于 2013 年总结了三维结构感知形状处理的关键概念和方法,并提出这些方法受限于高质量三维模型和大规模三维数据集的可用性.随着数字三维模型数量的不断增长,Xu 等人[10]于 2016 年总结了数据驱动的三维模型和场景的处理算法.Ioannidou 等人[11]于 2017 年综述了对三维数据应用深度学习的方法.Bronstein 等人[12]于 2017 年总结了将(结构化的)深度神经网络模型推广到非欧几里得(如图和流形)的几何深度学习方法.本文对现有的基于深度

学习进行三维数据分析与理解，并对三维模型检索/分类、分割、识别和生成等特定任务的方法进行了综述.

本文第 2 节汇总目前常用的三维数据集，对其模型数量、分类情况以及模型描述等进行对比与阐述，并总结特定任务的评价指标；第 3 节调研现有的三维特征描述符，对其优缺点进行总结；第 4 节从特定任务出发，就三维数据不同的表达方式，综述目前基于深度学习的三维数据分析理解网络，进行对比分析，并且从三维数据表达方式角度进一步汇总现有代表性的工作；第 5 节总结目前亟待解决的挑战性问题，讨论未来发展趋势；第 6 节对全文进行总结.

2　三维数据集及特定任务评价指标

2.1　三维数据集

为了使广大研究者方便利用日益丰富的三维数据来解决数字几何处理中的相关问题，各个机构提供了一系列标准的三维模型库，促进了数据驱动的数字几何分析. 目前，广泛使用的三维模型库有 McGill 数据集[13]、ModelNet 数据集、ShapeNet 数据集、ShapeGoogle 数据集[14]、SHREC 数据集[15-17]、LSUN 数据集[18]、IKEA 数据集[19]等，以上数据集的详细描述如表 1 所示.

表 1　常用三维数据集

名称	提供机构	描述	模型总数	分类情况
McGill[13]	麦吉尔大学	数据集规模较小，但所选取的不同姿势的各类模型较有代表性	255 个非刚性三维模型	分为 10 个类
ModelNet①	普林斯顿大学	可用于三维模型分类，识别和检索的较大数据集，有 ModelNet10 和 ModelNet40 两个子集	127 915 个模型	分为 662 个类
ShapeNet②	普林斯顿大学、斯坦福大学、丰田工业大学芝加哥分校	为每个模型提供了丰富的语义注释，子集 ShapeNet Core 包含 51 300 个模型，55 个类	约 300 万个模型	其中 22 万个模型被分为 3135 个类
ShapeGoogle[14]		包含分类明确的用来做检索的非刚性三维模型和一些不属于任何类的无关模型，用来扩充数据集和测试检索算法的性能	596 个模型用于检索，463 个无关模型	检索部分分为 10 个类
SHREC 2007(水密)③	欧洲图形学会	标准的非刚性三维模型数据集，模型为水密的三角网格，不含拓扑结构误差	400 个模型	平均分成 20 个类
SHREC 2011(非刚性)[15]	欧洲图形学会	标准的非刚性三维模型数据集，模型为水密的三角网格，不含拓扑结构误差	600 个姿态各异的非刚性三维模型	平均分成 30 个类
SHREC 2014(人)[16] 真实数据集	欧洲图形学会	都是属于“人”这个类的模型，按照性别、年龄、身材等分类	400 个模型	平均分成 40 个类
SHREC 2014(人)[16] 合成数据集	欧洲图形学会	都是属于“人”这个类的模型，按照性别、年龄、身材等分类	300 个模型	平均分成 15 个类
SHREC 2015(非刚性)[17]	欧洲图形学会	标准的非刚性三维模型数据集，模型为水密的三角网格，不含拓扑结构误差	1200 个模型	平均分成 50 个类
LSUN[18]	普林斯顿大学	大规模场景理解数据集	约 990 万个标记图像	分为 10 个场景类别，20 个对象类别
IKEA[19]	麻省理工学院	包含代表典型室内场景的三维模型和可用模型注释的图像	759 个图像和 219 个三维模型	

2.2　特定任务评价指标

2.2.1　分类/检索

三维模型分类根据三维模型内在几何特征和语义特征，计算未知模型与已知类别的模型的相似度，确定未知模型所属类别. 该任务在与理解三维场景有关的应用中起着重要的作用，如机器人室内导航[20]、虚拟现实[21]、增强现实[22]和自动驾驶[23]等. 三维模型分类信息在各种特征值优劣的比较、三维模型检索效果的评价、三维模型库的组织等方面有重要作用[24]. 如今，在线社交网络吸引了越来越多的关注，用户可以分享和消费各种各样的多媒体内容. 随着三维模型的迅猛增长，如何快速有效地从三维模型库中检索到所需模型成为数字几何处理领域

① ModelNet. http://modelnet.cs.princeton.edu/
② ShapeNet. https://www.shapenet.org/
③ SHREC 2007. http://watertight.ge.imati.cnr.it/

的研究热点[24].三维数据量的快速增长以及互联网搜索引擎的复杂性使得三维模型检索已从基于文本的检索转变为基于内容的检索[25].基于内容的检索其关键在于三维模型特征描述符的提取,之后通过计算特征描述符的相似度完成三维模型检索.

三维模型分类通常使用准确率(accuracy)作为算法性能评价的重要指标.假设数据集中第 i 类模型数量为 P,非 i 类模型数量为 N,正确识别为 i 类的模型数量为 TP,正确识别为非 i 类的模型数量为 TN,则

$$accuracy_i=\frac{TP+TN}{P+N}$$
$$mean\ accuracy=\frac{\sum accuracy_i}{P+N} \tag{1}$$

目前,三维模型检索任务通常使用 Shilane 等人[26]提出的普林斯顿基准(Princeton Shape Benchmark,PSB)来进行算法性能的评估.PSB 主要包含以下内容:

(1)PR 曲线和平均精度 mAP(Average Precision).PR 曲线由查准率 P(Precision rate)和查全率 R(Recall rate)之间的函数关系生成,查准率指的是三维模型检索结果中同类三维模型的比率,查全率指的是三维模型检索结果中同类三维模型占整个数据集中该类三维模型总量的比率.mAP 是 PR 曲线与坐标轴形成区域的面积大小,在一定的范围内,查准率与查全率呈现出反比关系,mAP 值越大,三维模型检索性能越好.

(2)最近邻方法 NN(Nearest Neighbor).若对某类三维模型进行检索,检索后返回的三维模型总数为 N,其中相关的三维模型数量为 K,则 NN 的计算公式为

$$NN=K/N \tag{2}$$

(3)第一层级 FT(First Tier)和第二层级 ST(Second Tier).若检索 C 类三维模型,数据集中的所有相关的三维模型总量为 $|C|$,检索后返回的相关三维模型的数量为 K,则

$$\begin{cases}FT=K/(|C|-1)\\ST=K/(2(|C|-1))\end{cases} \tag{3}$$

(4)E_度量(E_Measure,E).考虑前 32 个检索返回的结果,计算 P-R(Precision-Recall),则

$$E=2/(1/P+1/R) \tag{4}$$

(5)折扣的累积结果 DCG(Discounted Cumulative Gain).DCG 表明了检索后正确匹配的三维模型的顺序,其定义如下所示:

$$DCG_i=\begin{cases}G_i, & i=1\\ DCG_{i-1}+\dfrac{G_i}{\log_2(i)}, & i>1\end{cases}$$
$$DCG=\frac{DCG_k}{1+\sum_{j-2}^{|C|}\frac{1}{\log_2(j)}} \tag{5}$$

上述定义默认在检索结果的排序列表中,几乎不考虑排在该列表后面的三维模型.假设把有序列表 R 转换成列表 G,如果模型 R_i 是正确匹配的,则 G_i 的值为 1,否则为 0.

2.2.2 分割

三维模型分割是一项细粒度三维识别任务,给定三维模型,对点云模型或网格模型的每个点或面分配部分类别标签,例如椅子腿、椅子背、杯盖、杯柄等[27].被广泛认可的 PSB 衡量分割质量的指标包含以下内容[28]:

(1)切割差异 CD(Cut Discrepancy)是对分割边界互相接近程度的度量,是通过将一个分割到另一个分割的测地距离相加来完成的.假设有算法分割 S_a 和人工分割(真实值)S_g,C_1 和 C_2 分别是它们分割边界上的点集,则 CD 定义如下:

$$CD(S_1,S_2)=\frac{mean\{d_G(p_1,C_2),\forall p_1\in C_1\}+mean\{d_G(p_2,C_1),\forall p_2\in C_2\}}{avgRadius} \tag{6}$$

$$d_G(p_1,C_2)=\min\{d_G(p_1,p_2),\forall p_2\in C_2\} \tag{7}$$

(2)汉明距离 HD(Hamming Distance)是计算 S_a 和 S_g 两个分割的差异,定义如下:

$$D_H(S_a,S_g)=\frac{1}{2}\left(\frac{D_H(S_a\to S_g)}{\|S\|}+\frac{D_H(S_g\to S_a)}{\|S\|}\right) \tag{8}$$

其中,$\|S\|$ 表示整个模型的基数(顶点数或面数),$D_H(S_a\to S_g)$ 是汉明距离的方向函数,定义如下:

$$D_H(S_a\to S_g)=\sum_i\|R_a^i\backslash R_g^{i_t}\| \tag{9}$$

"\"表示集合差异运算符.R_a^i 是来自 S_a 的第 i 个分段,$R_g^{i_t}$ 是从 S_g 得到的最接近 R_a^i 的分段:

$$i_t=\max_k\|R_a^i\cap R_g^k\| \tag{10}$$

(3)边缘索引 RI(Rand Index).假设 N 是三维网格模型的面片数,则

$$RI(S_a,S_g)=\binom{2}{N}^{-1}\sum_{i,j,i<j}[C_{ij}P_{ij}+(1-C_{ij})(1-P_{ij})] \tag{11}$$

其中,$C_{ij}P_{ij}=1$ 表明面 i 和 j 在分割 S_a 和 S_g 中具有相同的标签,$(1-C_{ij})(1-P_{ij})$ 表明面 i 和 j 在分割 S_a 和 S_g 中具有不同的标签.

(4)一致性误差 *CE*(Consistency Error). 分为全局一致性误差 *GCE*(Global Consistency Error)和局部一致性误差 *LCE*(Local Consistency Error),定义如下:

$$GCE(S_a,S_g)=\frac{1}{N}\min\left\{\sum_i L_{3D}(S_a,S_g,f_i),\sum L_{3D}(S_g,S_a,f_i)\right\},$$

$$LCE(S_a,S_g)=\frac{1}{N}\sum_i \min\{L_{3D}(S_a,S_g,f_i),L_{3D}(S_g,S_a,f_i)\},$$

$$L_{3D}(S_a,S_g,f_i)=\frac{\|R(S_a,f_i)\backslash R(S_g,f_i)\|}{\|R(S_a,f_i)\|} \tag{12}$$

其中,$L_{3D}(S_a,S_g,f_i)$是局部细化误差,$R(S,f_i)$是分割 S 中包含 f_i的分段.

2.2.3　识别/物体检测

物体检测是计算机视觉领域研究的一个热点,它能够使计算机模拟人类的视觉检测机制,在具有冗余繁杂信息的场景中提取出人们感兴趣的物体[29].

三种被普遍使用的物体检测评价指标是:PR 曲线、*F*-Measure 和 *mAP*,PR 曲线和 *mAP* 的定义和 2.2.1 节中相似. 通常,精度(precision)和召回度(recall)都不能全面评估物体检测方法,*F*-Measure 将精度和召回度结合到单个度量中,通过加权调和平均值得到非负权重 β^2:

$$F_\beta=\frac{(1+\beta^2)\,precision\times recall}{\beta^2\,precision+recall} \tag{13}$$

其中,β 经常被设置为 1 以给予每个物体相同的重要性,此时 $F_\beta=F_1$[30].

2.2.4　生成/合成/重建

近年来,人们可以有更多的方法得到三维模型. 利用 CATIA、UG、MAYA 等三维建模软件,用户能通过输入符合其规则的参数或者交互操作构造三维模型,但是一般建模软件操作比较复杂,尤其对于自由曲面的构造更为繁琐;利用三维扫描仪可以得到真实场景和物体的点云信息,但其造价昂贵,且输出数据量庞大. 如何利用深度学习技术实现简便快速直观地构造三维模型是当前计算机视觉和计算机图形学领域研究的重点之一,集中在三维重建、三维生成和三维形状合成等方面.

三维重建是通过分析三维物体在真实世界中的图像等二维投影,学习其深度和颜色信息,推断出该物体的几何结构及空间位置,再现模型或场景的形状和外观[31]. 三维形状合成则通过检索和组合数据库中三维模型和其部件来合成新的三维模型. 基于生成式对抗网络 GAN 的三维生成可以实现用含有噪音的数据生成三维模型.

定性实验是研究三维重建、三维生成、三维形状合成时最直观的算法性能展示方式,通过图形化的输出展示结果与输入信息的对比,能较为直接地反映算法的优劣情况,如图 1 所示. 定量评价便于进行不同方法间的横向对比和大规模数据集上的客观性能度量. 交并比(Intersection-over-Unidon,IoU)是指重建结果与其真实值(ground truth)之间的交集占他们并集的比率,IoU 的值越高表示重建、生成效果越好[32].

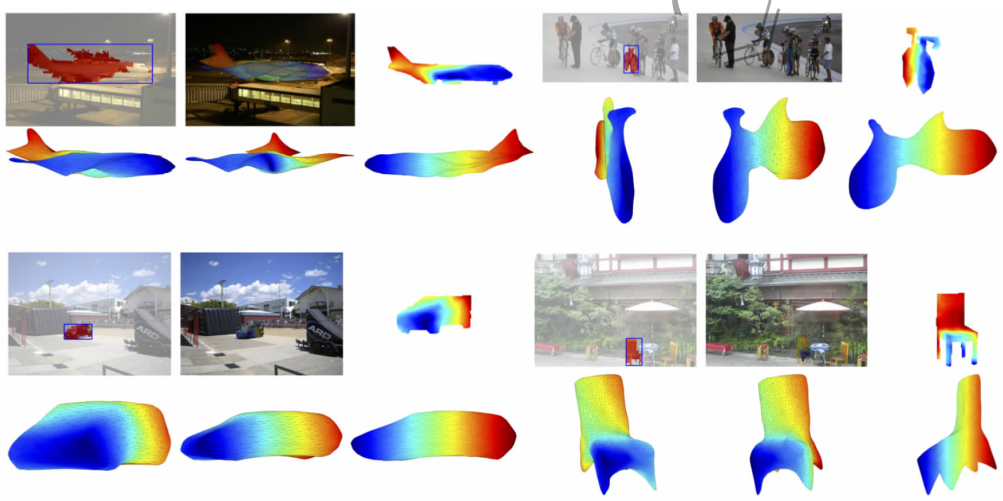

图 1　Kar 等部分重建结果[33]

3　三维形状特征描述符

三维形状特征描述符是三维模型简洁而紧凑的表示,蕴含着三维模型的几何本质. 处理三维模型分类、检索以及分割等任务的关键是提取三维模型信息丰富且具有辨别力的特征描述符[34]. 一个理想的三维形状特征描述符首先应具有对刚性变换(主要

有平移、旋转、缩放和映射等)的不变性,并且应该为“相似”模型设计“近似”描述符,以实现在分类和检索时缩短类内距,增大类间距.三维形状特征描述符的另外一个重要特性是针对姿势变化的不变性,姿势不变的特征描述符应该能够识别不同姿势的相同模型,例如站立、坐姿的人[35].表 2 中总结了被广泛使用的三维形状特征描述符.

表 2 常用三维形状特征描述符

类型	三维特征描述符	优点	缺点
统计数据	形状分布	从几何信息分布的角度对三维模型的结构特征进行描述,直观、易于理解	往往聚焦于三维模型的某一种几何特性,描述不够充分,且可能忽略了一些重要的局部信息
	测地距离		
	扩散距离		
	曲率加权距离		
	点密度		
	扩散高斯图		
	形状直方图		
图像	正交投影	将复杂的三维问题简化到相对成熟的二维图像领域,降低问题难度且符合人的视觉特征	容易丢失三维模型的空间几何信息
	旋转图像		
	光场		
	二维切片		
拓扑	多分辨率 Reeb 图	描述了三维模型各部分间的连通性,有效表达三维模型的全局特性和局部细节,对三维模型的旋转、拉伸等变换有较好的不变性	对拓扑噪声敏感,对模型质量要求较高,且计算复杂度高
	三维骨架		
信号分析	三维傅里叶变换	从频域的角度分析三维模型,可以有效地描述三维模型的内蕴属性	可能忽略空间相关性信息,且计算量较大
	球面调和分析		
	谱分析		

由于三维数据的复杂性以及表达的多样性,通常用户设定的三维形状特征描述符很难完整地捕捉三维数据的特征信息,其表达能力有限.不同的描述符有不同的优缺点和不同的适用场景,很难有统一的描述符适用于所有场景.因此,传统方法需针对特定场景设计特定的描述符.基于深度学习的方法是数据驱动的,可以直接通过数据学习得到描述能力更加强大的特征,且不依赖于场景,可直接用于数字几何领域相关任务.

4 特定任务及数据输入的三维数据分析理解网络

近年来,深度学习技术在图像识别和自然语言处理等领域取得显著进展,但传统的深度学习模型将输入设计为一维或二维数据,为使其能够在三维领域得到有效利用,研究者们提出了许多方法.本文从特定任务及数据输入的表达方式两个角度对三维数据分析理解网络进行了汇总,具体如表 3 所示.

目前提出的方法可以概括为以下三种:(1)基于低层特征提取高层特征的方法.先提取三维模型传统手工设计的特征描述符(称为低层特征),通过深度学习方法进行训练,学习得到新的抽象特征(称为高层特征),将其作为三维模型最终的特征描述符;(2)非本征方法.将三维模型变换到图像区域、体素区域、基元区域或点云等欧式空间,以适应传统的深度学习方法;(3)本征方法.设计能处理三维数据的深度学习模型,通常将三维模型看作二维流形或由点组成的图.本小节以具体的处理任务为线索总结了现有方法,结合三维数据的表示方式对这些方法进行分类与对比.本文进一步从数据表示方法的角度总结了目前的代表性工作,如表 4 所示.

表 3 特定任务及数据输入的三维数据分析理解网络汇总

任务	基于低层特征提取高层特征的方法	非本征方法					本征方法
		图像	体素	点云	基元	多模态	
分类/检索	DeepSD[36], Xie 等人[37], Ghodrati 等人[38], Dai 等人[39]	MVCNN[40], DeepPano[41], GIFT[42], Guo 等人[43], geometry images[44]	O-CNN[8], 3D ShapeNets[45], FPNN[46], Qi 等人[47], OctNet[48], VoxNet[49]	PointNet[27], PointNet++[50], PointCNN[51],	—	FusionNet[52](多视图和体素), Bu 等人[53](多视图和体素)	GCNN[54], Boscaini 等人[55]
分割	Guo 等人[56]	Huang 等人[57]	—	SPLANet[58], PCNN[59], Wang 等人[60], MDGNN[61]	—	—	—
识别/物体检测	—	SeeThrough[62], Im2CAD[63]	Bu 等人[53]	—	—	—	—
生成/合成/重建	—	Ben-Chen 等人[35], PointOutNet[64], MarrNet[65], Han 等人[66], PrGAN[67]	3D-GAN[68], FrankenGAN[69], Adaptive O-CNN[70]	Han 等人[66]	Li 等人[71], Tulsiani 等人[72]	—	—

表 4　基于三维模型表示方法的工作汇总

表示方法	一般方法	代表工作
体素	将三维模型表示为在三维体素网格上的二维变量的分布概率，即若体素在三维表面内则值为 1，否则为 0，是二维到三维的直接推广.将三维模型进行体素化表示后即可利用深度神经网络进行训练学习.	VoxNet[49]，3D ShapeNets[45]，FPNN[46]，Qi 等人[47]，OctNet[48]，O-CNN[8]，FusionNet[52]，Bu 等人[53]，3D-GAN[68]，FrankenGAN[69]，Adaptive O-CNN[70]
点云	直接将 3D 数据看成点的集合.将每个点看作一个神经元节点，节点包含点的坐标等信息，然后利用深度神经网络提取点的特征.基于点云的方法其主要思想是设计深度学习网络直接对点云数据进行分析处理.这类方法主要工作是对原始点云数据做一定处理或变换使其作为深度学习网络输入时能够得到有效的特征.	PointNet[27]，PointNet++[50]，PointCNN[51]，SPLANet[58]，PCNN[59]，Wang等人[60]，MDGNN[61]，Han 等人[66]
图像	先将三维模型通过投影等方法得到二维图像，再利用图像领域的方法进行处理，可以较好地利用现有的图像处理方法.	MVCNN[40]，DeepPano[41]，GIFT[42]，Guo 等人[43]，geometry images[44]，Huang 等人[57]，SeeThrough[62]，Im2CAD[63]，Kar 等人[35]，PointOutNet[64]，MarrNet[65]，Han 等人[66]，PrGAN[67]，FusionNet[52]，Im2struct[73]
基元(部件)	三维形状还可以表示成一些基本单元(部件)的组合，如矩形块、圆柱、圆锥、球等，然后利用网络来学习这些基本体块的参数对象来抽象复杂形状.	Li 等人[71]，Tulsiani 等人[72]，SCORES[74]
本征方法	本征方法直接将三维形状看成二维流形(Manifold)或由点组成的图(Graph)，顶点之间的距离不再是欧氏距离，然后直接将卷积定义在这样的数据结构上.	GCNN[54]，Boscaini 等人[55]
基于低层特征提取高层特征的方法	将三维模型传统低层特征与深度学习技术进行结合.首先提取三维模型一定数量的低层特征，然后通过深度学习方法进行训练，得到高层特征，将其作为三维模型最终的特征描述符，可用于三维模型检索与分类等.	DeepSD[36]，Xie 等人[37]，Ghodrati 等人[38]，Dai 等人[39]，Guo 等人[56]

4.1　分类/检索

4.1.1　基于低层特征提取高层特征的方法

该类方法的主要思路是将三维模型传统低层特征与深度学习技术进行结合.首先提取三维模型一定数量的低层特征，然后通过深度学习方法进行训练，得到新的抽象特征，即高层特征，将其作为三维模型最终的特征描述符，用于三维模型检索与分类等.

Fang 等人[36]于 2015 年基于深度神经网络提出了深度模型描述符(Deep Shape Descriptor，DeepSD).首先提取三维模型谱特征热核特征描述符(Heat Kernel Signature，HKS)；之后对模型顶点进行聚类，提出了热模型描述符(Heat Shape Descriptor，HeatSD)；最后将 HeatSD 特征作为深度神经网络的输入学习得到 DeepSD.该方法利用计算几何、计算机视觉和深度学习等多个研究领域的现有技术，不仅能够处理三维几何数据的复杂性，而且还能应对三维模型的结构变化和不一致性.

Xie 等人[37]于 2015 年基于有识别力的深度自编码器(discriminative deep auto-encoder)提出了一种三维模型高层特征学习方法.首先提取多尺度 HKS 特征，并将其作为自编码器的输入；之后将 Fisher 判别标准施加于隐含层中的神经元，学习得到三维模型的特征；最后连接所有自编码器隐含层中的神经元得到向量，并将该向量作为最终的模型特征描述符.

Ghodrati 等人[38]于 2016 年提出了一种多级特征学习方法.首先提取三维模型的谱图小波特征(Spectral Graph Wavelets Signature，SGWS)矩阵，其中每列表达了三维模型的局部信息；然后基于上述局部描述符应用 BoF(Bag of Feature，特征袋)框架构造中间层特征描述符，并且在 BoF 框架中，使用局部约束线性编码(Locality-constrained Linear Coding，LLC)的方式完成特征编码，引入双调和距离(biharmonic distance)来度量 BoF 向量间的空间关系；最后通过深度自编码器学习高层特征，形成简洁、几何信息丰富且计算效率高的深度模型感知描述符.

Dai 等人[39]于 2016 年提出了提取简洁且数据驱动的三维模型特征描述符的特征学习框架.首先使用尺度不变的热核特征描述符(Scale-Invariant Heat Kernel Signature，SIHKS)描述模型顶点，使用 LLC 编码模型顶点，得到三维模型的全局表示；然后使用多对一编码器(Many-to-One encoder，MOencoder)进一步学习高层特征，将隐含层作为三维模型最终的特征描述符.MOencoder 可以将同类三维模型从相同目标输出，不同类三维模型从不同目标输出，因此最终的特征描述符对三维模型结构变化具有鲁棒性[34].

4.1.2 非本征方法

Su 等人[40]于 2015 年提出了用于三维模型识别的多视图卷积神经网络(Multi-View Convolutional Neural Netwoks,MVCNN),首先通过相机从三维模型 12 个不同的视角得到一组视图,将其作为第一层卷积神经网络(Convolutional Neural Netwoks,CNNs)的输入,得到基于视图的特征.然后对其进行池化操作,之后作为第二层 CNNs 的输入,最终得到简洁的三维模型特征描述符,用于三维模型分类.流程如图 2 所示.

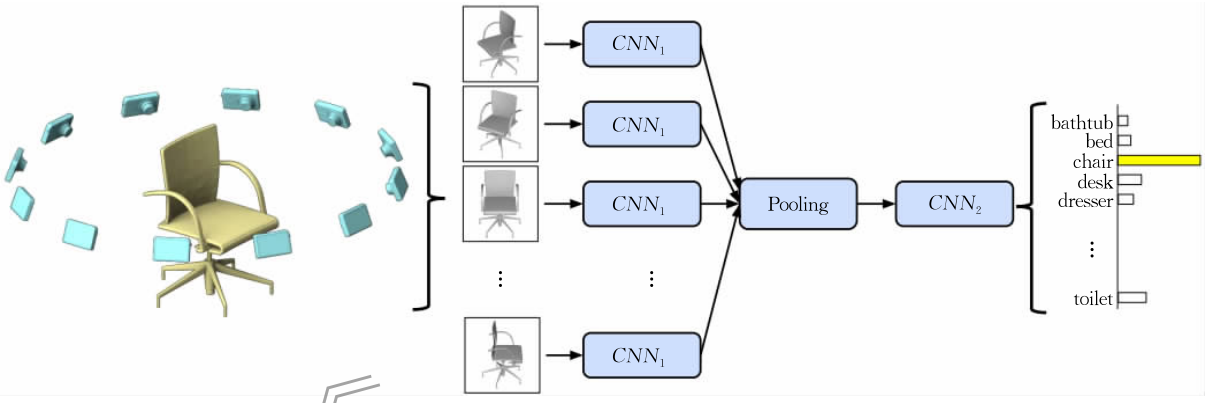

图 2 基于 Multi-View CNN 的三维模型识别流程[40]

Shi 等人[41]于 2015 年提出了用于三维模型识别任务的深度全景表示网络(Deep Panoramic,Deep-Pano).首先将每个三维模型转换成全景图,即围绕其主轴进行投影,得到一组视图.然后用经过改动的 CNNs 直接从这些视图学习三维模型的深度表示.与传统的 CNNs 不同,该网络在卷积层和全连接层之间加入了特殊的最大池化层,使最终学习到的特征对主轴旋转具有不变性.其网络结构如图 3 所示.整个网络的输入是三维模型的一组全景图,输出为该三维模型的类别概率,可以从图 3 中的 RWMP(Row-Wise Max-Pooling)、fc1 和 fc2 层中提取三维模型特征.其中,RWMP 即所加入的特殊的最大池化层,最终提取到的特征具有主轴旋转不变性.

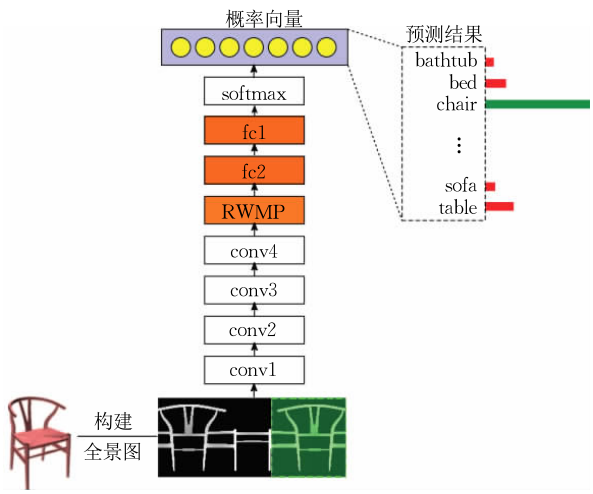

图 3 DeepPano 网络结构[41]

Bai 等人[42]于 2017 年提出了基于三维模型投影图像的搜索引擎 GIFT.它利用 GPU 加速的高效投影和视图特征提取,通过第一倒排文件(the First Inverted File,F-IF)加速视图匹配的过程,结合第二倒排文件(the Second Inverted File,S-IF)来有效地进行基于上下文的重排序,其中第二倒排文件在特征流形中得到了三维模型的局部分部信息.GIFT 的结构如图 4 所示.

Guo 等人[43]于 2016 年提出了一种由分类损失和三重态损失共同监督的深度嵌入网络,其关键思想是利用 CNNs 来学习针对多视图三维模型检索的有效三维模型描述符.该方法将高维图像空间映射到低维特征空间,其中特征的欧几里得距离直接对应于图像的语义相似性,即同一类的图像比不同类的图像更加接近.该方法中,每个三维模型用一组深度特征表示,于是多视图三维模型检索被转换为集合到集合的匹配问题,其在 SHREC2015 数据集[17]上的实验结果证明了该方法的有效性.

Sinha 等人[44]于 2016 年提出了“几何图像(geometry images)”这一概念,利用卷积神经网络学习几何图像特征,获得三维模型的内蕴表达.该方法首先将三维网格通过保面积参数化方法映射到球面上,得到参数化后的球形三维网格,然后将其映射到八面体上,沿着八面体的边剪开、铺平,得到二维平面;然后采用主曲率或 HKS 来保存三维模型的几何信息获得二维图像;最后使用

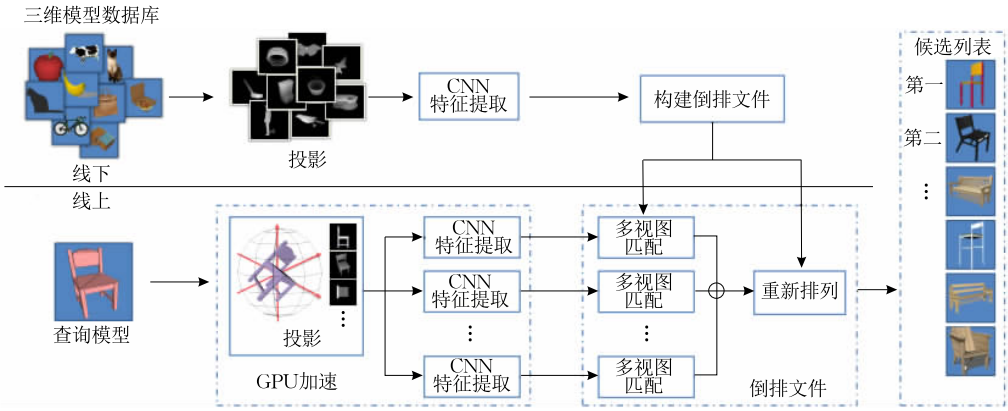

图 4　GIFT 结构图[42]

卷积神经网络学习几何图像特征. 在 TOSCA[75] 和 SHREC2011[15] 数据集上的实验结果表明几何图像能够灵活编码模型的内蕴属性，且具有等距变换不变性. 该方法在降维的同时能够保留三维模型丰富的几何信息，且可以使用现有的深度学习网络直接对几何图像进行处理，学习三维模型特征，提高特征描述符的表达能力.

Wu 等人[76] 于 2018 年提出了一种基于结构感知的噪声数据合并技术，将其作为标准聚类和降维的预处理，显著提高了许多非线性降维和聚类算法在复杂场景下的性能. 在对三维模型进行降维时，该算法具有很好的借鉴作用.

Wu 等人[45] 于 2015 年提出了 3D ShapeNets，用于体素化表示的三维模型的深度表示. 其主要思想是使用卷积深度信念网络(Convolutional Deep Belief Netwoks，CDBNs)将三维模型表示为在三维体素网格上的二维变量的分布概率，若体素在三维表面内则值为 1，否则为 0；然后使用卷积深度信念网络学习三维体素与标签之间的联合分布. 3D ShapeNets 的网络框架如图 5 所示. 为了方便描述，图中的每一个卷积层只画出了一个卷积核.

3D ShapeNets 模型的一个非常有意义的应用是从深度图中推断出完整的三维体素，而不需要明确物体的类别，也不需要事先对其建模. 它不是通过得到的零散的深度图来进行三维物体的形状建模，而是直接构建出初始的三维体素，以此来获得复杂的三维形状. 除此之外，它还可以针对一些缺失体素的三维模型的识别任务计算其信息增益. 这就使得在第一个视图的类别识别不够充分的时候可以选择后续的观察视角来进行识别. 3D ShapeNets 部分应用如图 6 所示.

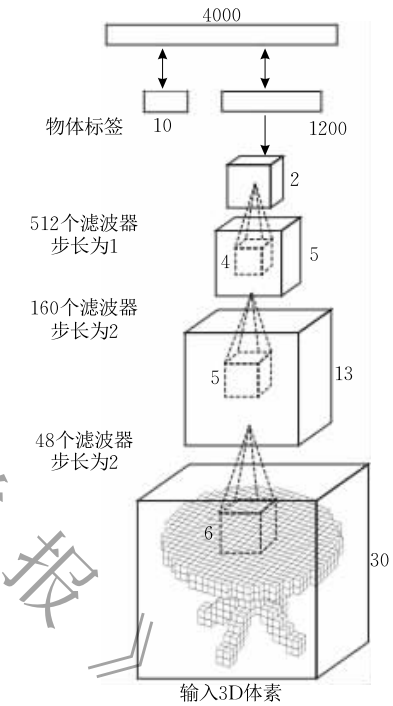

图 5　3D ShapeNets 网络框架[45]

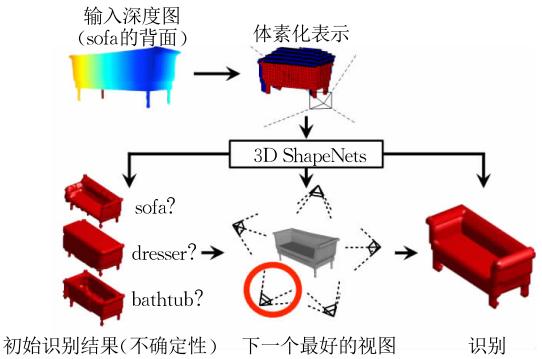

图 6　3D ShapeNets 的应用[45]

Li 等人[46] 于 2016 年提出了对三维数据的场探索网络 FPNN. 为了解决三维体素表达带来的稀疏性问题，该方法将三维模型表示成体素场，采用场探

索滤波器(Field Probing Filter)来学习特征,代替了卷积神经网络中的卷积层.在 ModelNet40 数据集上的实验结果表明该方法可以提取鲁棒的全局特征描述符,同时可以显著减少计算时间.

Qi 等人[47]于 2016 年提出一种添加辅助任务的处理体素数据的 CNN 网络体系框架.网络由三个 *mlpconv* 层和一个全连接层组成,其中,*mlpconv* 层是由 *ReLU* 层交织的三个 *conv*(卷积)层组成,将其用于提取局部块的特征,可以提高对模型更抽象表达的逼近.图 7 中 *mlpconv* 下的五个数字分别是第一个 *conv* 层的通道数、卷积核大小和步长以及第二个和第三个 *conv* 层的通道数,且默认第二个和第三个 *conv* 层的卷积核大小和步长为 1.例如,*mlpconv*(48,6,2;48;48)是由 *conv*(48,6,2)+*ReLU*、*conv*(48,1,1)+*ReLU* 和 *conv*(48,1,1)+*ReLU* 三个单元组成的.网络在全连接层分成两个部分,如图 7 所示,该网络结构的右下方分支输入为整个三维模型,完成经典的分类任务,右上方分支完成辅助任务,即将 512×2×2×2 的四维张量切分成 512 维的 2×2×2=8 个向量,并为每个向量设置分类任务.实验结果表明通过添加辅助任务在有效防止网络过拟合的同时,可以更好地挖掘每个模型的局部特征信息.

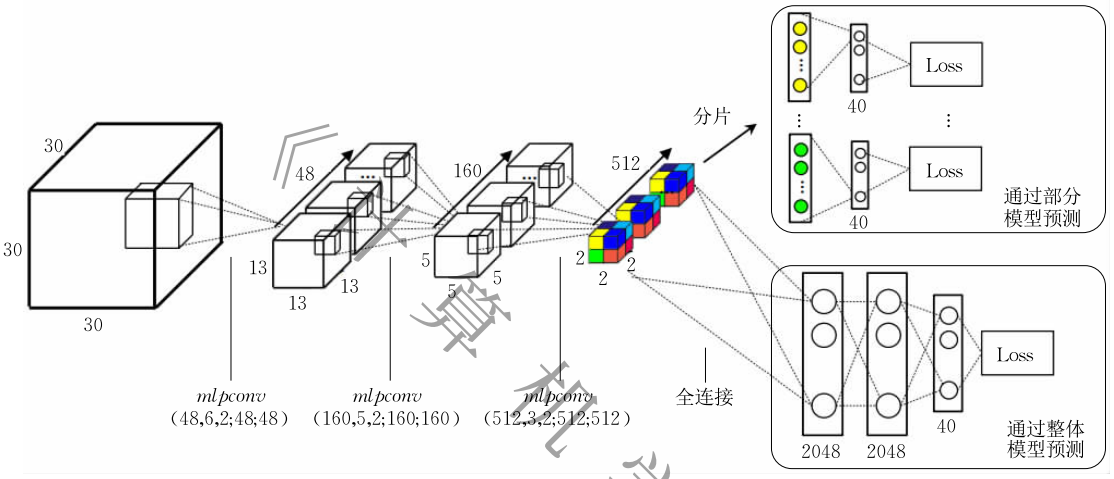

图 7 添加辅助任务的体素 CNN 网络结构[47]

Riegler 等人[48]于 2017 年提出了学习高分辨率三维模型表示框架 OctNet,可以在不影响三维模型分辨率的情况下实现更深的三维卷积网络.其主要思想是利用输入三维数据的稀疏性,使用一组不平衡的八叉树对三维空间进行分层划分,每个八叉树根据输入数据的密度来对三维空间进行划分,直到达到最优的分辨率.八叉树叶节点的大小各不相同,例如:一个深度为 3 的树的一个空叶节点可能包含多达 $8^3=512$ 个体素,并且八叉树每个叶节点存储池化后的特征表示.卷积网操作定义在这些树结构上.该方法利用输入三维数据的稀疏性,能够实现更有效的内存使用和计算.

Wang 等人[8]于 2017 年提出了基于八叉树的卷积神经网络(O-CNN)用于三维形状的分析.该方法首先将点云数据转换为八叉树,同时在八叉树的最深叶子结点存储平均法向信息;然后将平均法向信息作为 O-CNN 网络的输入;最后输出三维模型的类别概率,并将该类别概率存储在八叉树结构中.基于八叉树结构表示的三维模型与基于体素表示的三维模型如图 8 所示,其中,左边是真实模型,中间是基于体素表示的模型,右边是基于八叉树表示的模型.可以看出,基于八叉树结构的三维模型更加接近于真实模型.

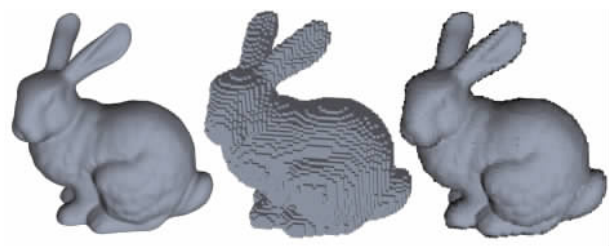

图 8 体素化与八叉树结构表示的三维模型[8]

O-CNN 网络反复在八叉树数据结构上由底向上应用卷积和池化操作来计算输出三维模型的类别概率.其中,使用 ReLU 函数来激活输出,使用 BN 算法[77](Batch Normalization,BN)来减少网络训练过程中中间层数据分布发生的变换;将"卷积+BN+*ReLU*+池化"作为一个基单元 U,一个 d 层(d 为八叉树层数)的 O-CNN 网络表示为 O-CNN(d),其定义形式如下:

输入$\rightarrow U_d \rightarrow U_{d-1} \rightarrow \cdots \rightarrow U_2$.

通过定义八叉树的深度来定义 O-CNN 网络的层数，即八叉树的深度等于 O-CNN 网络的层数. O-CNN 前端网络结构如图 9 所示.

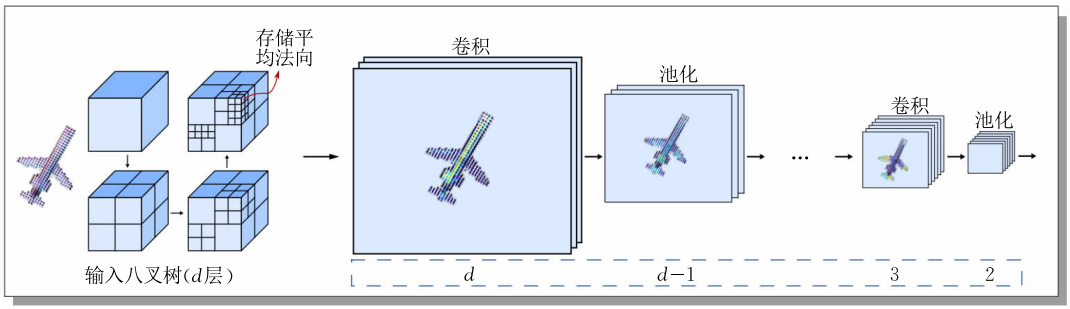

图 9　基于八叉树的卷积神经网络结构[8]

O-CNN 的核心是利用八叉树将原始三维数据进行稀疏表达并将 CNN 计算限制在八叉树的最深叶子结点上. 实验证明基于 O-CNN(6)的三维模型检索结果明显优于体素化及 PointNet 等方法的检索结果.

三维点云数据正在迅速增长. 无论是源于 CAD 模型还是来自 LiDAR 传感器或 RGBD 相机的扫描点云，无处不在. 三维点云的识别与分析在智能驾驶、无人机、机器人操纵等方面具有重要的作用. 基于点云的方法，其主要思想是设计深度学习网络直接对点云数据进行分析处理.

点云是几何数据结构的重要表达方式之一，它在计算机里的表示是 N 个无序的点，每个点以三维的坐标来表示. 由于点云数据的不规则性和无序性，直接将卷积操作应用在点云上会导致三维形状信息的丢失，造成计算结果的不准确. 因此，这类方法主要工作是对原始点云数据做一定处理或变换使其作为深度学习网络输入时能够得到有效的特征. 由于点云数据的不规则性，之前大部分的研究是将其转换为规则的图像集或者将其体素化，然后应用深度学习框架进行进一步处理. 但这样就会引起三维信息描述不准确和计算量太大的问题. 因此 Qi 等人[27]于 2017 年提出了在点云数据上直接进行深度学习的网络框架 PointNet，并用于三维点云分类和分割. 其网络结构如图 10 所示.

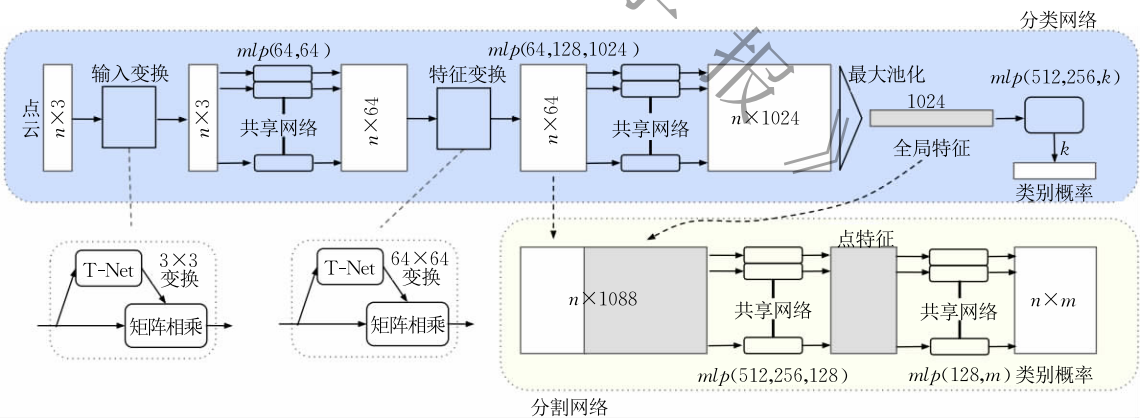

图 10　PointNet 网络结构图[27]

PointNet 网络将点云作为输入，分类任务输出为三维点云的类别标签，分割任务输出为分段点的类别标签. 该方法的关键是使用一个简单的对称函数，称为最大池化. 整个网络可以学习到一组优化标准，通过该优化标准可以选择点云的兴趣点或信息点，并且对其选择原因进行编码. 最后全连接层将这些学习到的优化值汇总成整个三维点云的全局特征描述符，用于三维点云分类与分割.

由于 PointNet 最终得到的是三维点云的全局特征描述符，它的基本思想是学习每一个点的空间编码，然后将所有的单个点特征汇总成全局点云特征. 无法得到度量空间中的局部结构，限制了它对细粒度模型的识别能力和对复杂场景的泛化能力. 因此，Qi 等人[50]于 2017 年提出了 PointNet++，一种对点云数据进行直接特征学习的分层网络框架，该方法解决了 PointNet 所存在的问题.

PointNet++首先通过底层空间的距离度量将点集划分为重叠的局部区域,然后利用 PointNet 提取从这些小邻域中得到的几何结构的局部特征,接着将这些局部特征进一步分成更大的单元并对其进行处理,产生更高层的特征.以上过程重复进行直到获得整个点集的特征.PointNet++网络结构及其在三维点云分类和分割的应用如图 11 所示.

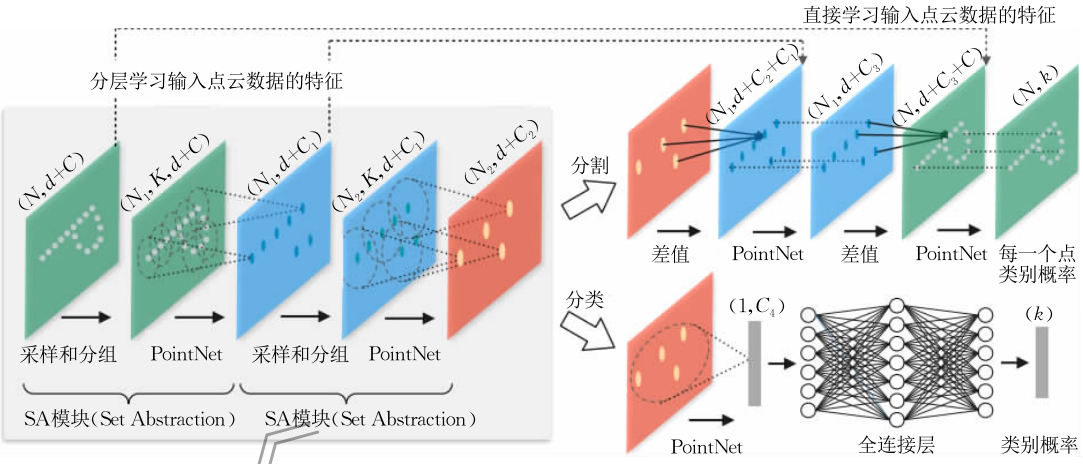

图 11 PointNet++网络结构及其在分割与分类上的应用[50]

PointNet++网络是一个分层的结构,解决了 PointNet 最大池化的过程中丢失很多信息导致网络不能用于深层学习的问题.例如:PointNet 网络可以将 1024 个点通过最大池化得到 1 个点,在这个过程中就会丢失很多信息.而 PointNet++不是一步将 1024 个点压缩成一个点,而是先将其根据空间距离进行分组,将其压缩成若干组,再利用 PointNet 网络对这若干个组进行处理.即 PointNet++可以将输入的点集 N 分成 M 个小的点集,每一个小的点集用 PointNet 进行处理得到一个特征,最终可以得到 M 个经过 PointNet 处理得到的特征.最后再用一层 PointNet 网络把这 M 个特征汇总起来,得到整个点集的特征.

Li 等人[51]于 2018 年提出了点云特征学习框架 PointCNN.首先从输入的原始点云数据学习 X 变换,然后使用它来加权排列输入点的特征.当输入点的顺序发生变化时,这组权值 X 能够相应地变化,使加权排列后的特征近似不变.该操作可以把输入点的形状信息编码到特征中,同时把输入特征的顺序归一化到某种潜在一致的模式.

Hegde 等人[52]于 2016 年提出一种三维卷积神经网络 FusionNet,结合三维数据体素表示和多视图表示学习三维模型新的特征,获得比单独使用其中任一个表示更好的分类结果.该方法核心在于融合三个卷积网络,分别为 V-CNN1、V-CNN2 和 MVCNN(基于 ALexNet[78]并在 ImageNet① 上进行预训练),三个网络在评分层(the scores layers)融合,评分层在找到类别预测前给出分数的线性组合,网络结构如图 12 所示.

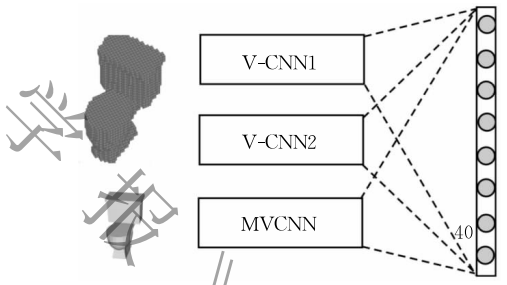

图 12 FusionNet 融合卷积网络得到物体标签的最终决策[52]

其中,V-CNN1 和 V-CNN2 是用来处理体素数据的网络.V-CNN1 由三个三维卷积层和两个全连接层组成,最终的全连接层被用作分类器.卷积层中使用的卷积核可以找到整个模型的体素相关性,并在 ReLU 层引入了非线性.池化层确保神经元学习到的冗余信息不会影响模型的大小.使用训练集中所有模型的 60 个方向来训练 V-CNN1.V-CNN2 引入 GoogLeNet[79]的初始模块,该模块可以拼接不同大小的卷积核处理结果,使网络可以学习到不同尺度的特征.V-CNN2 由两个这样的初始模块和一个卷积层,两个全连接层组成.

Bu 等人[53]于 2017 年提出了融合多模态特征

① ImageNet. http://www.image-net.org/

的方法，提高了单一特征的鉴别能力. 主要实现过程是利用卷积神经网络和卷积深度信念网络分别提取三维模型的视觉信息和几何信息，然后利用两个独立的深度信念网络(Deep Belief Netwoks，DBNs)从几何特征和视觉特征中学习高层特征. 最后，使用经过训练的受限玻尔兹曼机(Restricted Boltzmann Machine，RBM)来发掘不同模态之间的深度相关性. 该方法的流程如图 13 所示.

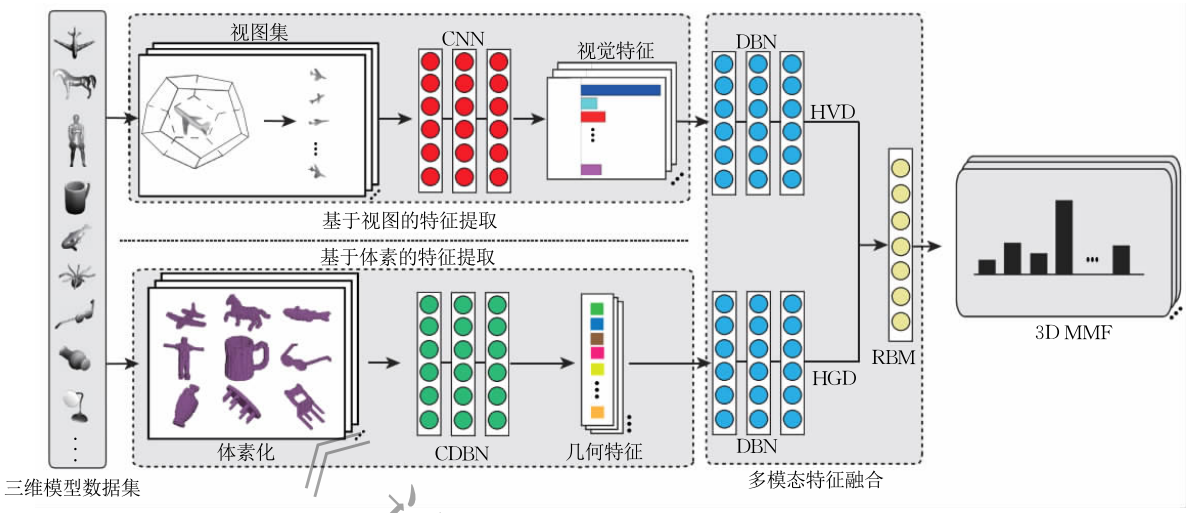

图 13　融合多模态特征算法流程图[53]

它主要包含以下三个部分：

(1) 基于视图的特征学习. 首先将三维模型表示成多视图；接下来，由于 CNNs 在计算机视觉领域具有非常好的视觉特征提取能力，所以用这些多视图去训练 CNNs，可以提取三维模型的视觉特征.

(2) 基于几何的特征学习. 由于卷积运算具有旋转和平移不变的优点，并且将其集成到了神经网络中实现了权重共享，减少了参数数量，从而提高了训练效果. 因此，使用 CDBNs 去学习几何特征. 首先将三维模型转化为容易输入到 CDBNs 中的体素形式，然后去学习三维模型的几何表示.

(3) 多模态特征融合. 以上两部分可以得到三维模型不同方面的信息，使用 DBNs 来得到它们的高层表示，分别称之为高层视觉描述符(High-level Visual Descriptor，HVD)和高层几何描述符(High-level Geometric Descriptor，HGD). 然后利用 RBM 将两种模态的高层特征进行关联，挖掘其非线性信息，生成更好的具有代表性的特征描述符，称之为 3D 多模态特征(3D Multi-Modality Feature，3D MMF).

4.1.3　本征方法

Masci 等人[54]于 2015 年提出了黎曼流形上的测地卷积神经网络(Geodesic Convolutional Neural Networks，GCNN)，它将卷积神经网络 CNN 推广到了非欧几里得流形. GCNN 的构建是基于极坐标系的局部测地系统，以此来提取每一个点的小块，然后将其输送给级联的滤波器，对其进行线性与非线性的操作. 其中，滤波器和线性组合权重的系数是优化的变量，通过学习这些变量来最小化特定任务的损失函数. 在流形上局部测地极坐标系的构建以及 GCNN 的网络结构分别如图 14(a)和图 14(b)所示. 其实验结果表明，由 GCNN 学习得到的三维特征描述符能有效地用于三维模型检索任务.

Boscaini 等人[55]于 2015 年在 GCNN 的基础上提出了使用局部谱卷积网络去学习形变模型的特定类描述符，其构建是基于局部频域分析，即窗口傅里叶变换在流形上的推广，用于提取一些密集内蕴描述符的局部行为. 然后将所得到的频域表示输送给一组滤波器，它的系数是通过学习得到的，用于最小化特定任务的损失.

Cohen 等人[80]提出了球面卷积神经网络，可鲁棒地分析球面图像，并不会受到曲面失真的影响. 球面 CNN 具有针对旋转的“等变”特性，它意味着该网络学习到的内部表征会与输入信息同步旋转. 该方法在三维模型分类等任务上验证了其有效性.

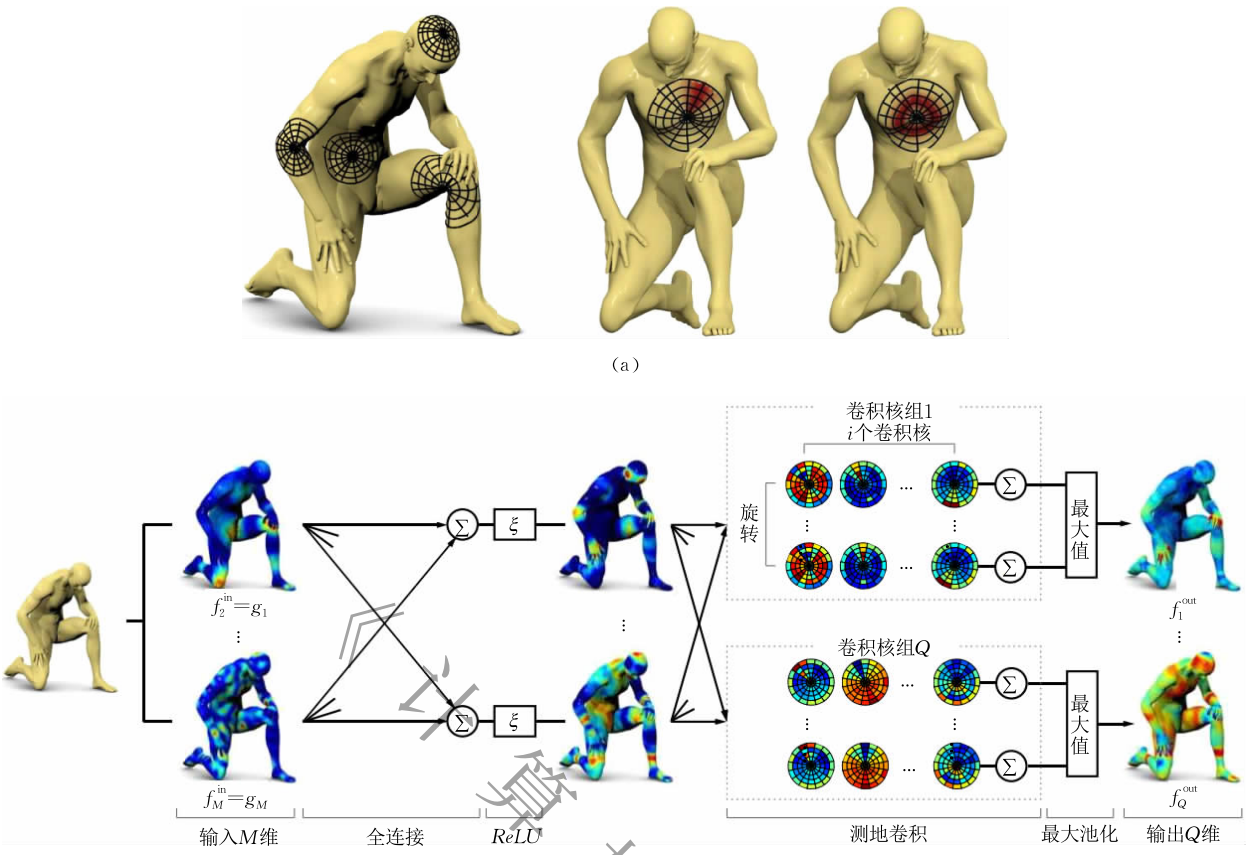

图 14 GCNN 局部测地极坐标系的构建以及网络结构[54]

4.2 分割

4.2.1 基于低层特征提取高层特征的方法

Guo 等人[56]于 2015 年提出使用 CNN 来适应不同的三维网格模型，以解决使用传统手工设计特征用于分割特定三维模型泛化能力不足的问题. 该方法首先提取七种传统几何特征，包括曲率[81](Curvature，CUR)、主成分分析特征[82](Principal Components Analysis，PCA)、形状直径函数[83](Shape Diameter Function，SDF)、内侧面距离[84](Distance from medial surface，DIS)、平均测地距离[85](Average Geodesic Distance，AGD)、形状上下文特征[86](Shape Context，SC)和旋转图像特征[87](Spin Image，SI)，这些特征组成二维特征矩阵，用来表示三维网格模型上的每个三角面片. 然后通过 CNN 从低层特征中学习得到更加紧凑有效的高层特征. 最后使用传统优化算法改善三角面片标签的局部平滑度，完成三维模型的分割. 其实验结果表明，该方法相比传统手工设计特征能获得更有效且稳健的分割效果，但该方法不能很好地识别并精细分割出三维模型中的微小部分.

4.2.2 非本征方法

Khoury 等人[88]于 2017 年提出一种用于非结构化点云的局部几何特征学习方法，该方法首先学习围绕点的局部几何特征的高维表示，然后训练一个深层网络将这个初始表示嵌入一个紧凑的欧几里得空间，获得最终的紧凑几何特征(Compact Geometric Features，CGF).

Su 等人[58]于 2018 年提出一种用于处理点云的深度学习网络架构 SPLANet. 该方法使用稀疏的双边卷积层(Bilateral Convolutional Layers，BCLs)作为构建块，这些层通过使用索引结构仅在点阵的占用部分上应用卷积来保持效率，并允许灵活的点阵结构规范，从而实现分层和空间感知特征学习，以及联合二维和三维推理. 由三维分割实验结果表明，SPLANet 能有效地实现点云处理.

Atzmon 等人[59]于 2018 年提出一种将 CNN 应用于点云的点卷积神经网络(Point Convolutional Neural Network，PCNN). 首先，点云上的函数在空间上扩展为连续体素上的函数；然后将连续体素卷积应用于该函数；最后再将结果限制回点云. 该方法

使基于图像的标准 CNN 能适应点云数据，实验结果表明该方法能有效地对三维模型进行分割，且在三维模型分类任务中也能获得较高准确率.

不同于 PCNN[59]，Wang 等人[60]于 2018 年提出可以对非网格结构化数据进行操作的参数连续卷积，其关键思想是设计了跨域整个连续向量空间的参数化核函数，利用该方法来学习任意数据结构，只要元素件的支持关系是可计算的. 该方法在对室内和室外场景的点云分割任务以及驾驶场景的激光雷达运动估计任务上验证了其有效性.

Huang 等人[57]于 2018 年提出一种基于多视图卷积网络，可以生成局部的基于点的形状描述符. 该方法首先以二维图像的形式处理三维模型的渲染视图，然后通过在大量分割三维模型部分的非刚性对齐来自动生成合成密集点对应数据集，最终网络能有效地编码多尺度的局部内容和细粒度的表面细节. 由于引入了多尺度，基于视图的投影表示，该方法得到的局部描述符能有效地应用于三维模型分割任务中，也能在关键点检测等任务中获得较好效果.

4.2.3　本征方法

Poulenard 等人[61]于 2018 年提出的多向测地神经网络(Multi-Directional Geodesic Neural Networks, MDGNN)实现卷积层的跨层传播和关联方向信息，从而更好地捕捉不同点的对应关系，以获得更好的分割结果.

4.3　识别/物体检测

Maturana 等人[49]2015 年提出了用于实时物体检测的三维卷积神经网络 VoxNet. 其主要思想是首先将点云数据进行体素化表示，然后对其进行卷积、池化和全连接等操作，最终得到有效的特征. 在物体检测任务中，VoxNet 较传统的机器学习方法体现出了很好的效果. VoxNet 的网络框架如图 15 所示.

其中，$Conv(f,d,s)$表示大小为 d 的 f 个卷积核，并且步长为 s；$Pool(m)$表示在区域 m 进行池化；$Full(n)$表示有 n 个输出的全连接层.

Song 等人[89]于 2016 年提出针对 RGB-D 图像中三维物体检测的方法. 该方法将三维体素场景作为输入，使用三维区域提议网络(Region Proposal Network, RPN)学习几何形状中的对象，并联合 RPN 和物体识别网络(Object Recognition Network, ORN)用以提取三维几何特征和二维颜色特征. 该方法通过训练两个不同尺度的 RPN 和用于复原三维边界框的 ORN 来处理各种大小的物体. 实验结果表明，该方法不仅准确率有所提升，处理速度也比 Depth-RCNN[90]和传统形状滑动方法[91]快得多.

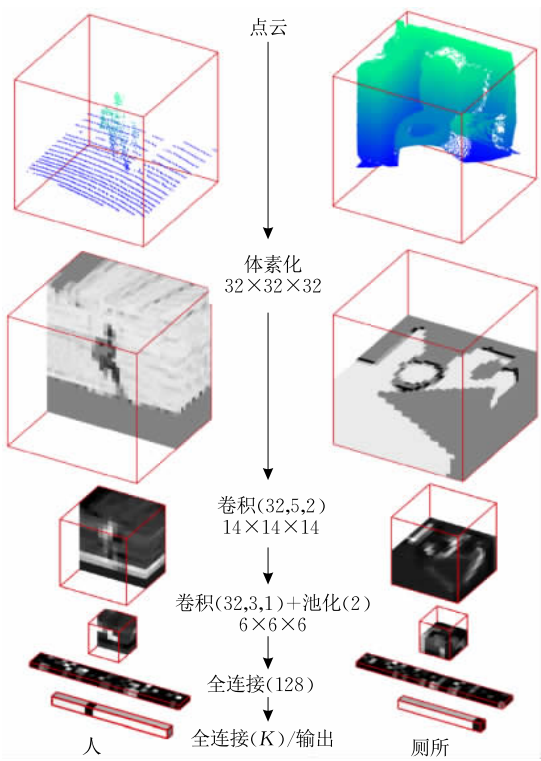

图 15　VoxNet 网络结构[49]

Hueting 等人[62]于 2017 年提出在高遮挡图像中识别物体的方法 SeeThrough，该方法考虑整体上下文的三维信息，并利用物体通常在典型场景中共同出现的事实来解决图像中的高遮挡问题. SeeThrough 首先使用在真实室内注释图像上训练的神经网络来提取二维关键点，并将它们输入到三维候选物体生成阶段. 然后，使用从大型三维场景数据库发现的成对共现统计来解决这些候选物体之间的全局选择问题. 迭代该过程，允许基于已发现物体的附近位置递增地检测具有低关键点响应的候选物体.

Izadinia 等人[63]于 2017 年提出 Im2CAD 系统，实现给定二维图像和家具 CAD 模型数据库，重建图像中描绘的场景，并且是由从数据库中检索得到的三维模型组成. 该系统流程图如图 16 所示. 该方法将物体检测器集中在单个对象中(如椅子、桌子、沙发、书架、床、床头柜、柜子和窗户等). 使用 Faster-RCNN[92]深度网络通过以下两个步骤进行检测：首先产生物体区域提议，然后使用深度卷积层计算每个物体的类别概率，完成系统中的物体检测任务.

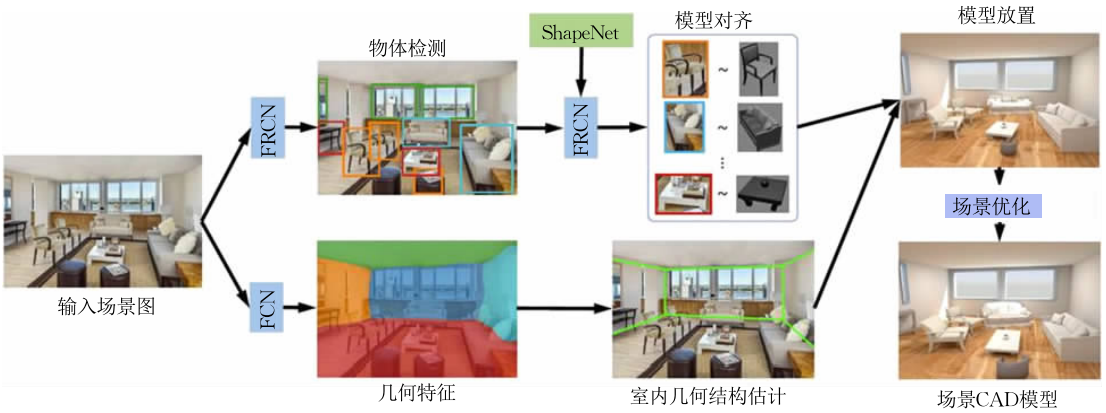

图 16 Im2CAD 系统流程图[63]

4.4 生成/合成/重建

Kar 等人[33]于 2015 年提出以图像作为输入的三维重建方法. 利用估计的实例分割和预测的视点来生成完整的三维网格和高频 2.5D 深度图,如图 17 所示. 学习物体检测数据集中可用的二维注释和形状轮廓,通过自动对象分割进行驱动,使用自底向上模块补充模型高频细节从而实现三维重建.

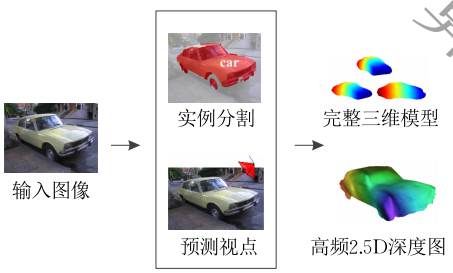

图 17 三维模型重建过程[33]

它包含以下三个部分:

(1)初始化. 首先检测并分割图像中的物体,预测模型的视点(旋转矩阵,$\boldsymbol{V}$)和子类别. 在特定比例的边界框中学习模型,即所有图像在训练期间先调整到特定宽度. 在给定预测的边界框内,相应地缩放该预测子类别由学习得到的平均形状(mean Shape,S). 最后,平均形状根据预测的视点旋转并放置到预测边界框的中心.

(2)形状推测. 初始化后,通过对固定的 S,$\boldsymbol{V}$ 优化方程:

$$\min_{S,\boldsymbol{V},\alpha} E_{\text{tot}}(\bar{S},\boldsymbol{V},\alpha)$$
$$\text{subject to: } S^i=\bar{S}+\sum_k \alpha_{ik}V_k \tag{14}$$

来求解变形权重 α(初始化为 0)以及所有相机投影参数(尺度、平移距离和旋转角度). 优化结果是根据推测的模型轮廓,视点和形状结合输入的类别信息,实现模型自顶向下的重建.

(3)自下而上的形状细化. 利用形状、反射率和光照之间的统计规律来恢复高频形状信息,实现模型与图像中物体的对齐.

Fan 等人[64]于 2017 年提出了点集生成网络 PointOutNet,实现从单视图进行三维重建,其网络输入是单张图像(RGB 或 RGB-D),由该图像和推断的视点位置决定三维空间中的点位置,输出用点云表示的三维模型.

Wu 等人[65]于2017年提出MarrNet,首先从RGB图像恢复物体法向、深度和轮廓草图,然后从 2.5D 草图中复原出三维模型,最后利用重投影一致性损失函数来确保估计的三维模型与 2.5 维草图对齐. MarrNet 的结构如图 18 所示.

Li 等人[71]于 2017 年提出一种基于递归神经网络(Recursive Neural Net,RvNN)的自动编码器,可以在两个三维形状之间进行结构混合. 该方法将平坦,无标签的任意部分布局映射到紧凑的编码,使用生成对抗网络[93](Generative Adversarial Network,GAN)进一步调整所学习的双向映射,以合成模型,从中对新结构进行采样. 最后,在第二个训练模块中生成细粒度的几何零件,合成的三维模型更加精确.

Tulsiani 等人[72]于 2017 年提出一个学习框架,通过学习三维体素基元组装对象来抽象复杂形状. 该方法通过无监督学习和训练 CNN,能够以刚性变换的长方体等简单基元来组装三维对象. 其在单个图像中重建的实验表明该方法能生成基础形状的抽象,并在部分形状处理任务上验证了该方法的有效性.

Wu 等人[68]于 2016 年结合体素卷积网络和生成对抗网络提出三维生成对抗网络(3D Generative-Adversarial Network,3D-GAN),实现从概率空间生成三维体素模型. 该方法中,生成器建立从低维概

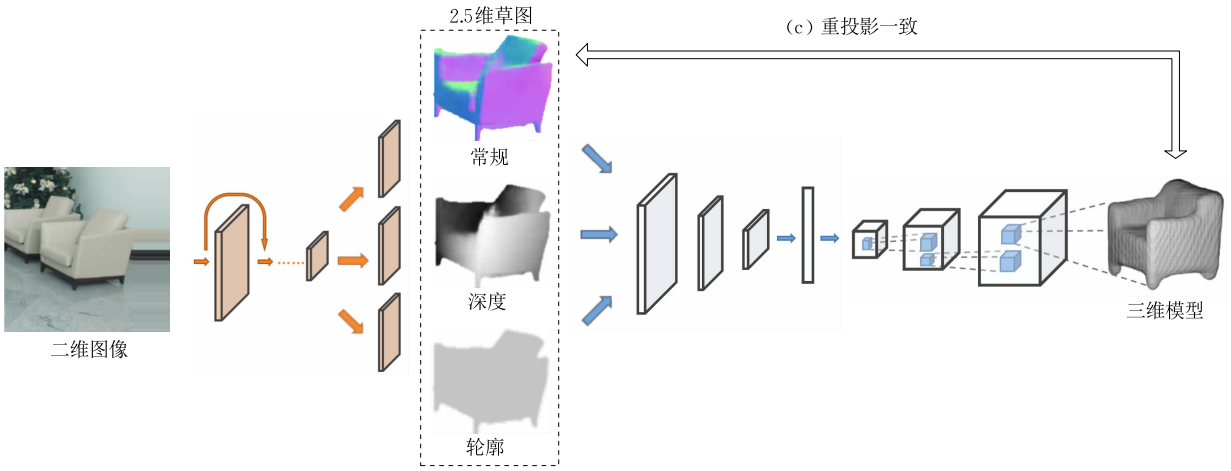

图 18　MarrNet 网络结构[65]

率空间到三维模型的映射，以实现在没有参考图像和 CAD 模型的情况下对模型进行采样，探索三维流形. 鉴别器判别输入的三维模型是生成模型或真实模型，并将结果返回到生成器，以此指导生成器的训练. 其实验结果表明该方法能有效地生成三维模型，基于其生成器对三维模型的理解，通过生成器提取的三维模型特征也能有效地用于三维模型分类和检索任务. Smith 等人[94]结合 GAN 的改进方案 WGAN-GP[95]和 3D-GAN 提出了 3D-IWGAN，能够生成和重建高质量的三维模型，同时训练过程更加稳定. Gadelha 等人[67]提出 PrGAN，在生成器和鉴别器中增加投影模块，使生成模型的投影视图与输入的二维图像相匹配，从而实现以完全无监督的方式重建三维模型.

Han 等人[96]于 2017 年提出一种恢复三维模型缺失部分的方法. 该方法结合了三维全卷积网络(3D Fully Convolutional，3DFCN)和长短期记忆上下文融合模块(Long Short-Term Memorized Context Fusion module，LSTM-CF)来推断和完善输入的不完整模型的全局结构. 然后利用块级三维 CNN(patch-level 3D CNN)由低分辨率体素表示进行局部精细化，完成端到端的恢复全局结构和局部细节.

Han 等人[66]于 2017 年提出一个基于深度学习的三维人脸和漫画建模草图系统. 其利用深度回归网络从二维草图推断出三维人脸模型，用户也可对初始三维人脸模型进行绘制和修改.

Kelly 等人[69]于 2018 年提出的 FrankenGAN 通过级联的 GANs 生成纹理和几何细节，从而实现在风格示例图像的指导下，由粗糙的建筑体素模型构建出具有真实性的完整城市场景.

Wang 等人[70]于 2018 年提出基于自适应八叉树的卷积神经网络(Adaptive O-CNN)，该方法通过平面片引导的分区策略来生成适应性的八叉树，并且使用平面片对不同层节点内的三维形状进行建模，获得更稀疏的三维表达. 基于自适应八叉树构造的三维解码器能生成更高质量的三维模型，同时降低了存储器需求和计算成本.

Niu 等人[73]于 2018 年提出的 Im2struct 能够从单张 RGB 图像中恢复三维形状的结构，该方法将三维形状表示为多个长方体的组合，通过卷积递归自动编码器从图像中识别各种形式和尺度的物体，再由解码器恢复各长方体的相互关系，包括连通性和对称性，从而获得最终的三维形状结构.

Zhu 等人[74]于 2018 年提出 SCORES 用以实现由输入子结构合成三维模型，该方法首先学习有效子结构的模型，然后训练编码器网络进行子结构调整，学习的子结构先验用于指导迭代优化，获得一组能构成连贯形状的子结构，即最终的合成三维模型.

Wang 等人[97]于 2018 年提出基于图卷积神经网络(Graph Convolutional Neural Network，GCN)的 Pixel2Mesh，利用从输入图像中提取的特征，通过逐渐变形网格表示的椭球体来生成正确的三维形状，从而实现端到端的从单张彩色图像重建三角形网格表达的三维模型. 三维网格是顶点、边和面的几何，该方法使网络以较少数量的顶点开始训练，学习将顶点分布到最具代表性的位置，并在前向传播时增加顶点数量从而添加局部细节. 训练过程中，该方法利用四种损失来约束输出形状和变形过程，其中，斜面损失(the chamfer loss)约束网格顶点的位置，法线损失(normal loss)控制模型表面法线的一致

性,拉普拉斯正则化(Laplacian regularization)保持在变形期间相邻顶点的相对位置不变,以及边缘长度正则化(edge length regularization)防止异常值.

5 亟待解决的问题及发展趋势

数据驱动的方法在发现三维形状之间几何、结构和语义关系中扮演着越来越重要的角色.近年来,越来越多的研究者尝试将深度学习技术应用于三维数据的分析与理解,并取得了一系列突破性进展.在前文综合介绍各类深度学习分析理解三维数据方法及对比基础上,本节将讨论亟待解决的挑战性问题和发展趋势.

5.1 亟待解决的问题

近年来,由于深度学习的广泛使用,三维模型处理技术迅速发展,算法性能不断提升,但仍然存在许多难题与挑战.主要有以下几个方面:

(1)缺乏大规模用于深度学习训练的数据集

深度学习技术在图像处理领域广泛使用并取得突破性效果的重要条件之一是大规模图像数据集的出现(例如:ImageNet 数据集,共 14197122 张图像).随着 3D 扫描设备以及建模技术的发展,我们可以获得越来越多的三维数据,但其数量远远不及文本、图像、视频等数据.例如:现有较大的三维模型数据集 ModelNet 数据集和 ShapeNet 数据集合起来仅约 300 万个模型,远少于图像数据.在较小的三维数据集上训练得到的深度模型往往很容易发生过拟合,也很难将其泛化到其它数据集,制约了深度学习模型的性能发挥.对于特定任务上的深度学习模型训练,往往需要相关任务的数据集,比如:场景生成问题,往往需要训练数据集包含场景模型,而现阶段标准的三维场景数据集较小,通过扫描设备得到的场景数据含有较多的噪声,在前期还需要有很多的降噪工作.虽然目前有一些三维模型的增广方法,但相较于文本、图像等来说还是比较少,如何扩展三维模型增广方法,获得更多的三维模型用于网络训练,也是目前存在的一个问题.

(2)三维模型表示的多样性与复杂性

相较于基于深度学习的图像分析理解,基于深度学习的三维数据分析理解关键在于输入数据的复杂性.二维图像的表达(像素)是规则的,可以直接对其做卷积等操作.但三维数据的表达具有多样性,目前的相关工作多基于离散表示(多视图、体素、点云等),不同的表达方法对深度学习网络的要求不同.对于多视图,该类方法先将三维模型通过投影等方法得到二维图像,再利用图像领域的方法进行处理,可以较好地利用现有的图像处理方法,目前处理多视图的方法主要有:MVCNN[40]等.对于体素,体素是规则的,可以直接将图像的相关处理方法进行扩展,即将二维扩展到三维,例如:将二维卷积操作扩展为三维卷积操作,该类方法与图像处理方法较为类似,目前处理体素的网络主要有:VoxNet[49]、3D-GAN[68]等.三维模型的体素表达虽然能够编码更多的形状信息,但由于计算能力的局限性,这种方法使得处理高分辨率三维数据受到了限制.目前有一些基于稀疏体素的工作,如 O-CNN[8]等.这类方法可以有效改善基于体素的方法,在处理高分辨率三维数据时可以减少时间和空间消耗.但这类方法的本质仍然是将二维图像的处理方法拓展到三维,并且对于更容易获得的数据输入形式,如点云并不奏效.对于点云,由于点云具有无序性,直接对点云进行卷积等操作首先需要解决该问题.目前有一些专门处理点云数据的网络,如 PointCNN[51]、PointNet[27]、PointNet++[50]等.但像这种直接处理三维模型的方法仍然较少,很难全面高效地对三维模型的特征进行概括,限制了深度学习技术在数字几何处理和分析领域的应用范围[98].

(3)三维深度学习网络训练过程耗时

三维数据的复杂性使得深度学习网络在处理大规模三维模型时需要较长的训练时间.尤其是基于体素的方法,网络训练耗时最久.目前使用稀疏体素和网格的表示方法虽然可以减少一些空间和时间消耗,但相较于文本和图像来说仍然十分耗时.这种情况由于 GPU 的快速发展和采用得到了一定程度的缓解,但仍然无法满足需求.因此,是否存在使网络训练高效的三维模型表达方式,如何改进或设计高效的深度学习网络,从而解决三维模型训练过程耗时的问题仍是一个值得关注的研究热点.

5.2 发展趋势

根据三维数据的复杂性以及亟待解决的问题,主要有以下发展趋势:

(1)引入计算共形几何相关理论

所有三维几何处理的问题都能通过计算共形几何转换为球面、欧式平面和二维双曲空间等标准空间中的二维问题.将三维网格模型参数化到规则的二维平面上可以获取模型对应的几何图像,从而使三维空间的复杂问题简化到二维图像处理领域中.基于多视图的三维数据分析理解方法在转换二维图

像的过程中会丢失重要的几何信息，并且改变了三维形状的局部和全局结构. 不同于三维模型的投影视图，几何图像是一种特殊的表面参数化. 三维模型被采样到一个类似于图像的普通二维网格中，在降维的同时可以保留三维模型丰富的几何信息，并且能够直接使用现有的深度学习网络框架基于几何图像提取三维模型特征，完成数字化几何模型的处理与分析.

(2) 融合二维图像与三维形状的数据驱动方法

三维模型具有结构复杂、格式多样的特点，单模态数据往往只在某些种类的模型处理上表现突出，多模态数据融合的网络框架可以突破该局限. 每种表示形式对三维模型信息的表达角度和程度不同，有其各自的适用范围和优缺点. 虽然从实验结果上，基于多模态数据的方法获得了较好的效果，但没有从几何意义和数学证明上对这种方法的合理性给出具有信服力的解释. 如何选择和选择哪些单模态数据进行融合，以及在融合过程中如何设置其权重都需要进一步的讨论研究.

(3) 融合先验知识的深度学习网络研究

基于深度学习的三维数据分析理解相较于手工设计的方法，是数据驱动的，通过深度神经网络可以自动学习三维数据的特征，避免手工设计的描述符的缺点. 但目前普遍认为深度学习方法仍旧是一个“黑箱”，没有坚实的理论基础. 加之三维数据量的不足，利用深度学习进行三维处理的过程中可以融合先验知识，将先验知识与深度学习网络进行结合是将来的研究热点之一. 比如：将传统手工设计的特征描述符与深度学习模型进行结合. 针对于特定任务，可以将有用的先验知识加入到网络设计中，使其适合小量的数据集，减少网络对数据的依赖. 例如，在三维重建任务时，可以将描述该场景的语义信息和三维数据或二维图像作为不同网络的输入，加入有用的语义信息相当于增加了一个新的通道，每个网络可以做一部分任务，得到不同的特征. 在一些视觉应用中，深度学习网络隐含地构建概率模型，存在一种方向是可以直接用概率的工具，例如最优传输理论及其各种降维近似，直接取代神经网络，从而使得黑箱透明①.

6　总　结

深度学习已成功广泛应用于计算机视觉、语音处理和自然语言处理等领域. 近年来，越来越多的研究团队将深度学习应用于数字几何处理领域并取得了瞩目的效果. 本文综述了基于深度学习的三维数据分析与理解的研究现状、存在的问题及发展趋势. 从特定的任务出发，如三维模型检索/分类、三维模型分割、三维物体识别/检测、三维模型生成/合成/重建，根据三维数据的表达方式，如基于低层特征提取高层特征的方法、本征方法(体素、点云等)、非本征方法(基于流形的方法)，介绍了处理不同输入数据表达的深度学习网络. 本文讨论了目前亟待解决的挑战性问题：(1) 三维数据量的不足限制了深度学习模型的性能，使其容易发生过拟合，并且在特定数据集上训练的模型很难迁移到其它数据集上；(2) 不同于二维图像，三维深度学习模型与三维数据的表达密切相关，在作为深度学习模型输入之前，需要对其做相应的处理，使其能够作为网络的最终输入，解决由于三维数据复杂性带来的问题，而且不同的表达方法也会使网络的结构千差万别；(3) 目前处理三维数据的深度学习模型的时间和空间消耗对于高分辨率三维模型仍然很有限制. 基于上述的问题，发展趋势主要有：(1) 引入计算共形几何相关理论，将三维数据降维到二维的同时可以保留丰富的形状信息，之后可以直接运用在图像上的处理方法；(2) 二维图像辅助的三维数据分析理解，可以将二维图像与三维数据用不同的深度学习网络进行学习，对得到的特征进行融合，作为最终的特征描述符用于三维语义相关的任务；(3) 将先验知识融入三维深度学习网络，当三维数据量不足时，可以先提取三维模型的低层特征，再利用浅层网络进行高层特征学习，可以在避免过拟合的同时学习高层特征. 针对于特定任务，也可以将对其有用的信息和三维数据同时作为一个网络的输入，将最终的优化目标设置为损失函数.

参　考　文　献

[1] Yang Yu-Bin, Lin Hui, Zhu Qing. Content-based 3D model retrieval: A survey. Chinese Journal of Computers, 2004, 27(10): 1297-1310(in Chinese)
(杨育彬，林珲，朱庆. 基于内容的三维模型检索综述. 计算机学报，2004，27(10)：1297-1310)

[2] Qiu Z, Zhang T. Key techniques on cultural relic 3D reconstruction. Acta Electronica Sinica, 2008, 36(12): 2423-2427

[3] Wang W, Liu X, Liu L. Shape matching and retrieval based on multiple feature descriptors. Computer Aided Drafting,

① https://m.sohu.com/n/481614827/

Design and Manufacturing, 2013, 23(1): 71-78

[4] Xie Zhi-Ge, Wang Yue-Qing, Dou Yong, et al. 3D feature learning via convolutional auto-encoder extreme learning machine. Journal of Computer-Aided Design and Computer Graphics, 2015, 27(11): 2058-2064(in Chinese)
(谢智歌, 王岳青, 窦勇等. 基于卷积-自动编码机的三维形状特征学习. 计算机辅助设计与图形学学报, 2015, 27(11): 2058-2064)

[5] Li H S, Zheng Y P, Wu X Q, et al. 3D model generation and reconstruction using conditional generative adversarial network. International Journal of Computational Intelligence Systems, 2019, 12(2): 697-705

[6] Sun Zhi-Jun, Xue Lei, Xu Yang-Ming, et al. Overview of deep learning. Application Research of Computers, 2012, 29(8): 2806-2810(in Chinese)
(孙志军, 薛磊, 许阳明等. 深度学习研究综述. 计算机应用研究, 2012, 29(8): 2806-2810)

[7] Liu Quan, Zhai Jian-Wei, Zhang Zong-Zhang, et al. A survey on deep reinforcement learning. Chinese Journal of Computers, 2018, 41(1): 1-27(in Chinese)
(刘全, 翟建伟, 章宗长等. 深度强化学习综述. 计算机学报, 2018, 41(1): 1-27)

[8] Wang P S, Liu Y, Guo Y X, et al. O-CNN: Octree-based convolutional neural networks for 3D shape analysis. ACM Transactions on Graphics, 2017, 36(4): 72

[9] Mitra N, Wand M, Zhang H R, et al. Structure-aware shape processing//Proceedings of the SIGGRAPH Asia 2013 Courses. Hong Kong, China, 2013: 1

[10] Xu K, Kim V G, Huang Q, et al. Data-driven shape analysis and processing. Computer Graphics Forum, 2017, 36(1): 101-132

[11] Ioannidou A, Chatzilari E, Nikolopoulos S, et al. Deep learning advances in computer vision with 3D data: A survey. ACM Computing Surveys, 2017, 50(2): 20

[12] Bronstein M M, Bruna J, LeCun Y, et al. Geometric deep learning: Going beyond euclidean data. IEEE Signal Processing Magazine, 2017, 34(4): 18-42

[13] Siddiqi K, Zhang J, Macrini D, et al. Retrieving articulated 3-D models using medial surfaces. Machine Vision and Applications, 2008, 19(4): 261-275

[14] Bronstein A M, Bronstein M M, Guibas L J, et al. Shape Google: Geometric words and expressions for invariant shape retrieval. ACM Transactions on Graphics, 2011, 30(1): 623-636

[15] Lian Z, Godil A, Bustos B, et al. Shape retrieval on non-rigid 3D watertight meshes//Proceedings of the 4th Eurographics Conference on 3D Object Retrieval. Llandudno, UK, 2011: 79-88

[16] Pickup D, Sun X, Rosin P L, et al. Shape retrieval of non-rigid 3D human models. International Journal of Computer Vision, 2016, 120(2): 169-193

[17] Lian Z, Zhang J, Choi S, et al. SHREC'15 track: Non-rigid 3D shape retrieval//Proceedings of the Eurographics Workshop on 3D Object Retrieval. Zurich, Switzerland, 2015: 257-266

[18] Yu F, Seff A, Zhang Y, et al. Construction of a large-scale image dataset using deep learning with humans in the loop. arXiv: 1506.03365, 2015

[19] Lim J J, Pirsiavash H, Torralba A. Parsing IKEA objects: Fine pose estimation//Proceedings of the IEEE International Conference on Computer Vision. Washington, USA, 2013: 2992-2999

[20] Wang C, Meng L, She S, et al. Autonomous mobile robot navigation in uneven and unstructured indoor environments. arXiv: 1710.10523, 2017

[21] Anbarjafari G, Haamer R, Lusi L, et al. 3D face reconstruction with region based best fit blending using mobile phone for virtual reality based social media. arXiv: 1801.01089, 2017

[22] Ge X, Pan L, Li Q, et al. Multipath cooperative communications networks for augmented and virtual reality transmission. IEEE Transactions on Multimedia, 2017, 19(10): 2345-2358

[23] Siam M, Mahgoub H, Zahran M, et al. MODNet: Motion and appearance based moving object detection network for autonomous driving//Proceedings of the International Conference on Intelligent Transportation Systems. Hawaii, USA, 2018: 2859-2864

[24] Li H S, Liu X, Lai L, et al. An area weighted surface sampling method for 3D model retrieval. Chinese Journal of Electronics, 2014, 23(3): 484-488

[25] Li H, Zhao T, Li N, et al. Feature matching of multi-view 3D models based on hash binary encoding. Neural Network World, 2017, 27(1): 95

[26] Shilane P, Min P, Kazhdan M, et al. The Princeton shape benchmark//Proceedings of the Conference on the Shape Modeling Applications. Washington, USA, 2004: 167-178

[27] Qi C R, Su H, Kaichun M, et al. PointNet: Deep learning on point sets for 3D classification and segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 77-85

[28] Chen X, Golovinskiy A, Funkhouser T. A benchmark for 3D mesh segmentation. Proceedings of the ACM Transactions on Graphics, 2009, 28(3): 73

[29] Lin Hua-Feng, Li Jing, Liu Guo-Dong, et al. Saliency detection method using adaptive background template and spatial prior. Acta Automatica Sinica, 2017, 43(10): 1736-1748(in Chinese)
(林华锋, 李静, 刘国栋等. 基于自适应背景模板与空间先验的显著性物体检测方法. 自动化学报, 2017, 43(10): 1736-1748)

[30] Cheng G, Han J. A survey on object detection in optical remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 117: 11-28

[31] Gomes L, Bellon O R P, Silva L. 3D reconstruction methods for digital preservation of cultural heritage: A survey. Pattern Recognition Letters, 2014, 50: 3-14

[32] Choy C B, Xu D, Gwak J Y, et al. 3D-R2N2: A unified approach for single and multi-view 3D object reconstruction//Proceedings of the European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 628-644

[33] Kar A, Tulsiani S, Carreira J, et al. Category-specific object reconstruction from a single image//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 1966-1974

[34] Li Hai-Sheng, Sun Li, Wu Yu-Juan, et al. Survey on feature extraction techniques for non-rigid 3D shape retrieval. Journal of Software, 2018, 29(2): 483-505(in Chinese)
(李海生，孙莉，武玉娟等. 非刚性三维模型检索特征提取技术研究. 软件学报，2018，29(2)：483-505)

[35] Ben-Chen M, Gotsman C. Characterizing shape using conformal factors//Proceedings of the Eurographics Workshop on 3D Object Retrieval. Crete, Greece, 2008: 1-8

[36] Fang Y, Xie J, Dai G, et al. 3D deep shape descriptor//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 2319-2328

[37] Xie J, Fang Y, Zhu F, et al. DeepShape: Deep learned shape descriptor for 3D shape matching and retrieval//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 1275-1283

[38] Ghodrati H, Hamza A B. Deep shape-aware descriptor for nonrigid 3D object retrieval. International Journal of Multimedia Information Retrieval, 2016, (3): 1-14

[39] Dai G, Xie J, Zhu F, et al. Learning a discriminative deformation-invariant 3D shape descriptor via many-to-one encoder. Pattern Recognition Letters, 2016, 83: 330-338

[40] Su H, Maji S, Kalogerakis E, et al. Multi-view convolutional neural networks for 3D shape recognition//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2016: 945-953

[41] Shi B G, Bai S, Zhou Z, et al. DeepPano: Deep panoramic representation for 3D shape recognition. IEEE Signal Processing Letters, 2015, 22(12): 2339-2343

[42] Bai S, Bai X, Zhou Z X, et al. GIFT: Towards scalable 3D shape retrieval. IEEE Transactions on Multimedia, 2017, 19(6): 1257-1271

[43] Guo H, Wang J, Gao Y, et al. Multi-view 3D object retrieval with deep embedding network. IEEE Transactions on Image Processing, 2016, 25(12): 5526-5537

[44] Sinha A, Bai J, Ramani K. Deep learning 3D shape surfaces using geometry images//Proceedings of the European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 223-240

[45] Wu Z R, Song S, Khosla A, et al. 3D ShapeNets: A deep representation for volumetric shapes//Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 1912-1920

[46] Li Y, Pirk S, Su H, et al. FPNN: Field probing neural networks for 3D data//Proceedings of the Advances in Neural Information Processing Systems. Barcelona, Spain, 2016: 307-315

[47] Qi C R, Su H, Nießner M, et al. Volumetric and multi-view CNNs for object classification on 3D data//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 5648-5656

[48] Riegler G, Ulusoy A O, Geiger A. OctNet: Learning deep 3D representations at high resolutions//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 3

[49] Maturana D, Scherer S. VoxNet: A 3D convolutional neural network for real-time object recognition//Proceedings of the IEEE International Conference on Intelligent Robots and Systems. Hamburg, Germany, 2015: 922-928

[50] Qi C R, Yi L, Su H, et al. PointNet++: Deep hierarchical feature learning on point sets in a metric space//Proceedings of the Advances in Neural Information Processing Systems. California, USA, 2017: 5105-5114

[51] Li Y, Bu R, Sun M, et al. PointCNN. arXiv: 1801.07791, 2018

[52] Hegde V, Zadeh R. FusionNet: 3D object classification using multiple data representations. arXiv: 1607.05695, 2016

[53] Bu S H, Wang L, Han P, et al. 3D shape recognition and retrieval based on multi-modality deep learning. Neurocomputing, 2017, 259: 183-193

[54] Masci J, Boscaini D, Bronstein M, et al. Geodesic convolutional neural networks on Riemannian manifolds//Proceedings of the IEEE International Conference on Computer Vision Workshop. Santiago, Chile, 2015: 832-840

[55] Boscaini D, Masci J, Melzi S, et al. Learning class-specific descriptors for deformable shapes using localized spectral convolutional networks. Computer Graphics Forum, 2015, 34(5): 13-23

[56] Guo K, Zou D, Chen X. 3D mesh labeling via deep convolutional neural networks. ACM Transactions on Graphics, 2015, 35(1): 3

[57] Huang H, Kalogerakis E, Chaudhuri S, et al. Learning local shape descriptors from part correspondences with multiview convolutional networks. ACM Transactions on Graphics, 2018, 37(1): 6

[58] Su H, Jampani V, Sun D, et al. SPLATNet: Sparse lattice networks for point cloud processing//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 2530-2539

[59] Atzmon M, Maron H, Lipman Y. Point convolutional neural networks by extension operators. arXiv: 1803.10091, 2018

[60] Wang S, Suo S, Pokrovsky A, et al. Deep parametric continuous convolutional neural networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 2589-2597

[61] Poulenard A, Ovsjanikov M. Multi-directional geodesic neural networks via equivariant convolution//Proceedings of the SIGGRAPH Asia 2018 Technical Papers. Tokyo, Japan, 2018: 236

[62] Hueting M, Kim V, Yumer E, et al. SeeThrough: Finding chairs in heavily occluded indoor scenes. arXiv: 1710.10473, 2017

[63] Izadinia H, Shan Q, Seitz S M. Im2CAD//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 2422-2431

[64] Fan H, Su H, Guibas L. A point set generation network for 3D object reconstruction from a single image//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 2463-2471

[65] Wu J, Wang Y, Xue T, et al. MarrNet: 3D shape reconstruction via 2.5D sketches//Proceedings of the Advances in Neural Information Processing Systems. California, USA, 2017: 540-550

[66] Han X, Gao C, Yu Y. DeepSketch2Face: A deep learning based sketching system for 3D face and caricature modeling. ACM Transactions on Graphics, 2017, 36(4): 126

[67] Gadelha M, Maji S, Wang R. 3D shape induction from 2D views of multiple objects//Proceedings of the IEEE International Conference on 3D Vision. Qingdao, China, 2017: 402-411

[68] Wu J, Zhang C, Xue T, et al. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling //Proceedings of the Advances in Neural Information Processing Systems. Barcelona, Spain, 2016: 82-90

[69] Kelly T, Guerrero P, Steed A, et al. FrankenGAN: Guided detail synthesis for building mass models using style-synchonized GANs. arXiv: 1806.07179, 2018

[70] Wang P S, Sun C Y, Liu Y, et al. Adaptive O-CNN: A patch-based deep representation of 3D shapes//Proceedings of the SIGGRAPH Asia 2018 Technical Papers. Tokyo, Japan, 2018: 217

[71] Li J, Xu K, Chaudhuri S, et al. GRASS: Generative recursive autoencoders for shape structures. ACM Transactions on Graphics, 2017, 36(4): 52

[72] Tulsiani S, Su H, Guibas L J, et al. Learning shape abstractions by assembling volumetric primitives//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 2635-2643

[73] Niu C, Li J, Xu K. Im2struct: Recovering 3D shape structure from a single RGB image//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 4521-4529

[74] Zhu C, Xu K, Chaudhuri S, et al. SCORES: Shape composition with recursive substructure priors//Proceedings of the SIGGRAPH Asia 2018 Technical Papers. Tokyo, Japan, 2018: 211

[75] Bronstein A M, Bronstein M M, Kimmel R. Efficient computation of isometry-invariant distances between surfaces. SIAM Journal on Scientific Computing, 2006, 28(5): 1812-1836

[76] Wu S H, Bertholet P, Huang H, et al. Structure-aware data consolidation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(10): 2529-2537

[77] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv: 1502.03167, 2015

[78] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks//Proceedings of the Advances in Neural Information Processing Systems. California, USA, 2012: 1097-1105

[79] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 1275-1283

[80] Cohen T S, Geiger M, Köhler J, et al. Spherical CNNs. arXiv: 1801.10130, 2018

[81] Gal R, Cohen-Or D. Salient geometric features for partial shape matching and similarity. ACM Transactions on Graphics, 2006, 25(1): 130-150

[82] Kalogerakis E, Hertzmann A, Singh K. Learning 3D mesh segmentation and labeling. ACM Transactions on Graphics, 2010, 29(4): 102

[83] Shapira L, Shalom S, Shamir A, et al. Contextual part analogies in 3D objects. International Journal of Computer Vision, 2010, 89(2-3): 309-326

[84] Liu R, Zhang H, Shamir A, et al. A part-aware surface metric for shape analysis//Proceedings of the Computer Graphics Forum. Oxford, UK, 2009, 28(2): 397-406

[85] Hilaga M, Shinagawa Y, Kohmura T, et al. Topology matching for fully automatic similarity estimation of 3D shapes //Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. Los Angeles, USA, 2001: 203-212

[86] Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 24(4): 509-522

[87] Johnson A E, Hebert M. Using spin images for efficient object recognition in cluttered 3D scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, (5): 433-449

[88] Khoury M, Zhou Q Y, Koltun V. Learning compact geometric features//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 153-161

[89] Song S, Xiao J. Deep sliding shapes for amodal 3D object detection in RGB-D images//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 808-816

[90] Gupta S, Arbeláez P, Girshick R, et al. Aligning 3D models to RGB-D images of cluttered scenes//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 4731-4740

[91] Song S, Xiao J. Sliding shapes for 3D object detection in depth images//Proceedings of the European Conference on Computer Vision. Zurich, Switzerland, 2014: 634-651

[92] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2015: 91-99

[93] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2014: 2672-2680

[94] Smith E, Meger D. Improved adversarial systems for 3D object generation and reconstruction. arXiv: 1707.09557, 2017

[95] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of Wasserstein GANs//Proceedings of the Advances in Neural Information Processing Systems. California, USA, 2017: 5767-5777

[96] Han X, Li Z, Huang H, et al. High-resolution shape completion using deep neural networks for global structure and local geometry inference//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017: 85-93

[97] Wang N, Zhang Y, Li Z, et al. Pixel2Mesh: Generating 3D mesh models from single RGB images//Proceedings of the European Conference on Computer Vision. Munich, Germany, 2018: 52-67

[98] Xia Qing, Li Shuai, Hao Ai-Min, et al. Deep learning for digital geometry processing and analysis: A review. Journal of Computer Research and Development, 2019, 56(1): 155-182(in Chinese)
(夏清，李帅，郝爱民等. 基于深度学习的数字几何处理与分析技术研究进展. 计算机研究与发展，2019，56(1)：155-182)

LI Hai-Sheng, Ph. D., professor. His research interests include computer graphics, digital geometry processing and visualization.

WU Yu-Juan, M. S. candidate. Her research interests include computer graphics and digital geometry processing.

ZHENG Yan-Ping, M. S. candidate. Her research interests include computer graphics and digital geometry processing.

WU Xiao-Qun, Ph. D., associate professor. Her research interests include computer graphics, digital geometry processing and image processing.

CAI Qiang, Ph. D., professor. His research interests include computer graphics, digital geometry processing and visualization.

DU Jun-Ping, Ph. D., professor. Her research interests include motion image processing, social network analysis and search, multi-source data fusion and big data mining.

Background

With the development of Internet, more and more 3D models can be acquired on the web for free. The increasing abundance of 3D data, including single models and scene models, encourages researchers to utilize these data to effectively process and analyze digital geometric models. Since deep learning has shown good performance in the field of computer vision, it is a hot research topic to extend deep learning to the field of digital geometric processing. Present works is mainly carried out from the following aspects due to the complexity of 3D data: extracting high-level features based on low-level features, structured representation of 3D data, dimensionality reduction for 3D data, fusion of multi-modal features, and the method based on manifold.

In recent years, authors have made a breakthrough in some aspects including multi-view feature learning of 3D models, 3D model retrieval based on view matching, 3D model shape feature extraction, 3D model multi-feature fusion, and 3D model retrieval based on point cloud features. In the aspect of multi-view feature learning of 3D models, authors conducted research on traditional feature extraction methods (low-level features), secondary feature extraction methods (middle-level features), and deep learning feature extraction methods (high-level features) to complete low-level feature extraction and middle-level features. By using the proposed three-dimensional model multi-view feature extraction algorithm, the accuracy of three-dimensional model retrieval is effectively improved. In terms of 3D model retrieval techniques based on view matching, the authors propose a view rendering mechanism for external view camera arrays that project three-dimensional models into binary images, depth images, and contour images respectively. In addition, we designed feature extraction algorithms for three types of views to extract more diverse shape feature information, and improved the original binary matching algorithm for image set three-dimensional reconstruction. Besides, we have participated in the SHREC competition many times and achieved good results.

This work was partially supported by the National Natural Science Foundation of China (Nos. 61877002, 61532006, 61602015), the Beijing Natural Science Foundation (No. 4172013) and the Beijing Education Commission Research Team Construction Project (No. PXM2019_014213_000007).