

基于深度学习的数字几何处理与分析技术研究进展

夏清 李帅 郝爱民 赵沁平

(虚拟现实技术与系统国家重点实验室(北京航空航天大学) 北京 100083)

(xiaqing@buaa.edu.cn)

Deep Learning for Digital Geometry Processing and Analysis: A Review

Xia Qing, Li Shuai, Hao Aimin, and Zhao Qinping

(State Key Laboratory of Virtual Reality Technology and Systems (Beihang University), Beijing 100083)

Abstract With the rapid development of various hardware sensors and reconstruction technologies, digital geometric models have become the fourth generation of digital multimedia after audio, image and video, and have been widely used in many fields. Traditional digital geometry processing and analysis are mainly based on manually defined features that can only be valid for specific problems or under specific conditions. The deep learning, especially the neural network model, in the success of natural language processing and image processing demonstrates its powerful ability as a feature extraction tool for data analysis, and is therefore gradually used in the field of digital geometry processing. In this paper, we review the works of digital geometry processing and analysis based on deep learning in recent years, carefully analyze the research progress of shape matching and retrieval, shape classification and segmentation, shape generation, shape completion and reconstruction and shape deformation and editing, and also point out some existing problems and a few possible directions of future works.

Key words computer graphics; digital geometry processing and analysis; deep learning; neural networks; research progress review

摘要 随着各种硬件传感器以及重建技术的快速发展,数字几何模型成为继音频、图像、视频之后的第4代数字媒体,并在多个领域得到广泛应用。传统的数字几何分析和处理方法主要建立在手工定义的模型特征之上,这类方法只对特定问题或者在特定条件下才有效。而深度学习,尤其是神经网络模型,在自然语言处理和图像处理方面的成功,展示了它作为数据特征提取工具的强大能力,因此越来越多地被用在数字几何处理领域。对近年来基于深度学习的数字几何处理与分析技术进行了综述,重点分析了模型匹配与检索、模型分类与分割、模型生成、模型修复与重建以及模型变形与编辑中的相关技术国内外最新研究进展,并指出了存在的主要问题和发展方向。

关键词 计算机图形学;数字几何处理与分析;深度学习;神经网络;研究进展综述

中图法分类号 TP391

20世纪70年代至今,随着各种硬件传感器以及重建技术的快速发展,不同种类的数字多媒体数据大规模涌现,并先后经历了由低维到高维、由简单到复杂的发展演变过程^[1-3]。早期的数字多媒体主要

以一维的音频等数据为主,之后语音识别、自然语言处理等技术开始迅速发展。到80、90年代,随着手持数码相机和摄像机的普及,开始出现二维数字图像、视频等数字多媒体数据,大量方便的数据获取方式

以及后来互联网的爆炸式发展推动了图像处理、视频分析技术的快速进步。而后随着激光、结构光等三维扫描设备以及 Kinect 等消费级别的动作捕捉设备的逐渐普及,数字多媒体数据的描述能力进一步增强,此时三维几何模型成为继音频、图像、视频之后的第 4 代数字媒体,并在工业制造、数字娱乐、计算机辅助设计、逆向工程、生物医药、数字文化遗产保护、虚拟现实等方面得到广泛应用。

传统的数字几何分析和处理主要建立在手工定义的模型特征之上,只能针对特定的问题或者满足特定的条件才能有效,而深度学习,尤其是神经网络模型,在自然语言处理和图像处理方面的成功,展示了它作为数据分析特征提取工具的强大能力,因此逐渐被用来提高传统数字几何处理方法的性能,并且取得了很多不同方向上的应用成果。但是三维几何模型不像语音、图像那样具有标准的表示形式,通常难以直接应用于神经网络的训练当中,因此衍生了很多不同的针对不同几何模型表示形式的处理方法,甚至是针对性的网络结构,来将数字几何处理和深度学习技术结合起来。

本文立足于数字几何模型的分析和处理,对近年来基于深度学习技术的数字几何处理工作进行了系统梳理。具体内容安排如下:第 1 节介绍常见的深度学习模型,主要是神经网络模型,以及各种模型的典型应用场景;第 2 节结合深度学习技术介绍了数字几何模型的主要表示方式,以及对应的处理方法;第 3 节对具体的数字几何处理应用方向的相关工作进行了介绍,主要有数字几何模型的匹配与检索、分类与识别、生成、修复与重建、变形与编辑等;第 4 节指出了存在的问题和发展方向。

1 相关深度学习模型

深度学习模型不同于传统的机器学习模型,它基于神经网络模型,通过训练调整神经网络的参数,得到每一层的参数值,每层代表对输入数据不同层次的表征,以此来将原始数据自动地转化成最为简单的特征表示。常见的神经网络模型包括了深度神经网络、卷积神经网络、对抗生成网络、循环神经网络、置信神经网络、递归神经网络、自编解码器等,本节将对这些常见网络结构及其应用进行简要介绍。需要注意的是,这些神经网络并非严格神经网络分类,比如卷积神经网络实际上是深度神经网络的一种,对抗生成网络是一种学习的思想,其具体的实现

可以是卷积神经网络,本节只是根据各种典型网络的特点对其进行分别介绍。

1.1 深度神经网络

深度神经网络是最为典型的深度学习模型,其他深度学习模型都是在此基础上演变扩展而来。深度神经网络中最基本的单元是神经元,神经元模型模拟生物神经网络中“兴奋”传导的机制:当某神经元的电位达到一定阈值后被激活从而继续向其他神经元传递“兴奋”,McCulloch 和 Pitts 据此提出了 M-P 神经元模型^[4],该模型被广泛运用至今。图 1 所示为一个典型的神经元结构示意图。

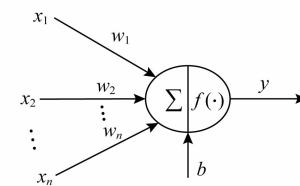


Fig. 1 A typical neuron in neural networks

图 1 神经网络中的典型神经元结构

由 2 层神经元组成的神经网络称为感知机,然而感知机往往不能解决非线性可分问题,因此多层神经网络应运而生。如图 2 所示,多层次网络除了输入和输出层之外,网络内部还具有多个隐藏层。为使多层次网络满足特定任务需求,基于链式求导法则的误差逆传播算法(back propagation, BP)^[5]被用于网络参数的训练更新。而随着神经网络的层数增加,其参数量也相应增长,从而其拟合现实生活中某些复杂问题的函数关系的能力也不断增强。

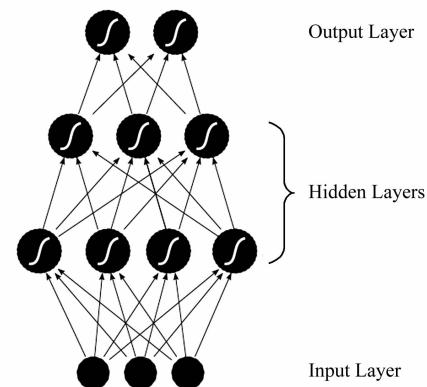


Fig. 2 A typical multi-layer neural network model

图 2 典型的多层次神经网络模型结构

近年来,随着 GPU 硬件设备计算能力的大幅提高以及任务数据的不断增长,深度神经网络训练的时间消耗被显著缩短,并且网络过拟合风险也显著减小。因此深度神经网络在图像分类、目标检测识别、

图像分割以及图像生成等任务中都得到了广泛的应用。Krizhevsky 等人^[6]运用深层神经网络 AlexNet 在 ImageNet^[7] 图像分类任务上取得重要成功,之后涌现出了大量的如 VGG^[8], GoogleNet^[9], ResNet^[10]

等优秀的深层网络结构。如图 3 所示为 AlexNet 的网络结构,相比于浅层神经网络, AlexNet 的网络更深,也更加复杂,因此具有更好的图像描述和区分能力。

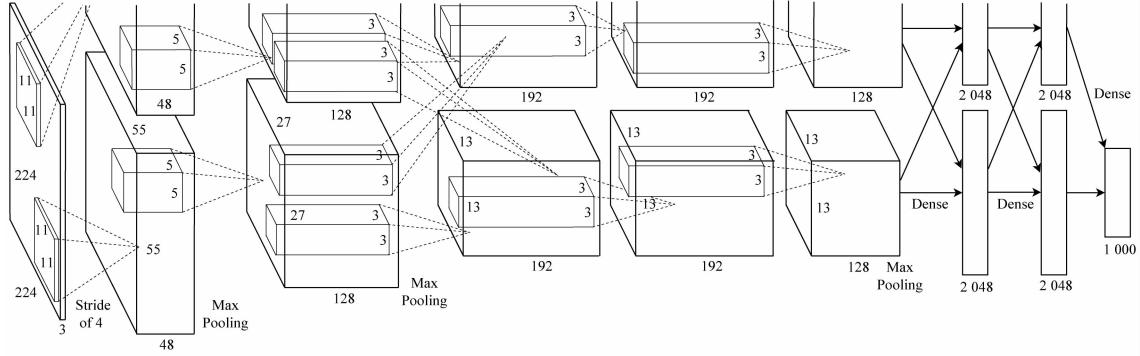


Fig. 3 Architecture of AlexNet^[6]

图 3 AlexNet 结构示意图^[6]

在深度神经网络的结构中,若均采用全连接方式,将使得网络的时间与空间效率十分低下。因此一些优秀的深度神经网络采用“权值共享”策略,以此节约网络模型的存储空间并且提高网络运行效率,这一策略在卷积神经网络(convolutional neural network, CNN)^[11] 中有着最为重要的体现。

1.2 卷积神经网络

卷积神经网络(CNN)是专门针对图像设计的深度学习模型,通过卷积操作和权值共享来减少深度神经网络中的连接数量,从而降低深度网络的复

杂度。当前主流的视觉处理任务,如检测、分割、估计等大都采用卷积神经网络进行处理。

检测任务是众多图像处理任务,如人脸识别、实例分割、姿态估计等任务的基础,一个好的检测方法对后续任务的准确性、有效性、实时性起着至关重要的作用。

Girshick 等人提出的著名的 R-CNN^[12] 网络使用传统方法提取候选框,并用卷积神经网络提取特征,对物体通过滑动窗口的方式实现检测分类,其具体处理流程如图 4 所示:

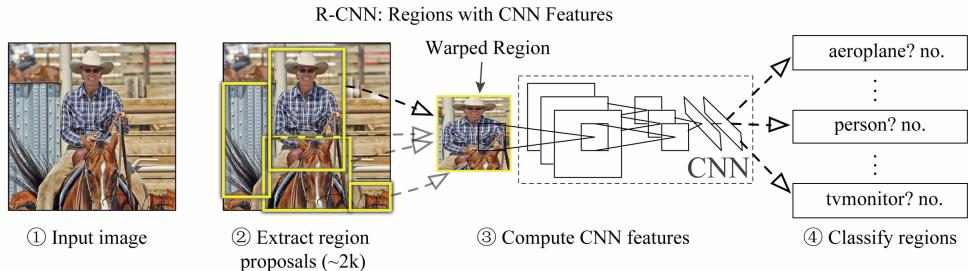


Fig. 4 Object detection based on R-CNN^[12]

图 4 基于 R-CNN 的目标检测示意图^[12]

之后 Girshick^[13] 又对 R-CNN 候选区域提取方法进行了改进,采用卷积方法提取候选区域,提出了 Fast R-CNN,不仅提高了检测的准确率,同时也取得了更快的运行速度。后来 Ren 等人^[14] 进一步复用提取候选区域的特征用于候选的分类任务,提出了 Faster R-CNN。借助于 region proposal network (RPN), Faster R-CNN 将特征在提取兴趣区域(region of interest, ROI)和分类 2 个阶段中进行复用,减少了冗余计算。

除了检测任务之外,卷积神经网络还可以用于图像分割任务,最初 Long 等人^[15] 将用于分类任务的卷积神经网络最后的全连接层替换为卷积层,从而提出了全卷积神经网络(fully convolutional networks, FCN),用于图像分割,其结构如图 5 所示。之后几乎所有的图像分割网络都是从 FCN 的基础上演变而来,比如 U-Net^[16], SegNet^[17], Deeplab^[18-20], PSPNet^[21] 等。

针对估计任务,卷积神经网络也表现出优越的

能力. 例如人体骨架估计任务, Newell 等人提出了一种端到端(end-to-end)的基于卷积神经网络的多人骨架识别模型 Associative Embedding^[22]. 如图 6 所示.

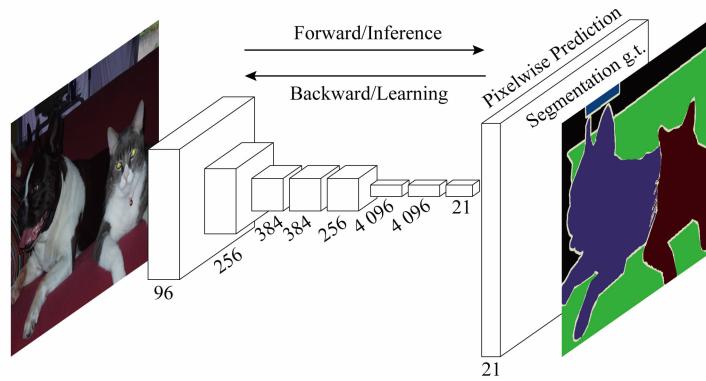


Fig. 5 Architecture of FCN^[15]

图 5 全卷积神经网络结构示意图^[15]

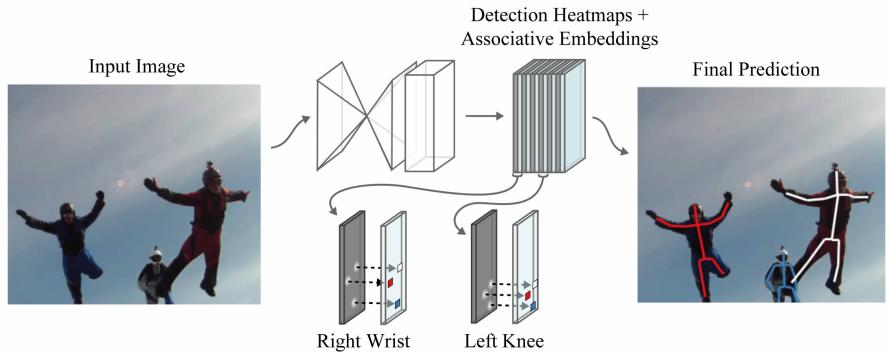


Fig. 6 Human skeleton estimation based on CNN^[22]

图 6 基于卷神经网络的人体骨架估计^[22]

1.3 对抗生成网络

对抗生成网络(generative adversarial network, GAN)^[23]最早由 Goodfellow 等人提出, 是一个通过对抗过程估计生成模型的框架, 利用了二元零和博弈的思想. 具体如图 7 所示, 框架中同时训练 2 个模

所示, 该方法通过卷积神经网络同时预测骨架关键点的热力图(heat maps)及其所对应人的标签的热力图, 从而实现对多人骨架的快速预测与正确归类.

Detection Heatmaps +
Associative Embeddings

Final Prediction

型: 捕获数据分布的生成模型 G 和估计样本来自训练数据的概率的判别模型 D, G 的训练程序是将 D 错误的概率最大化, 这个框架对应一个最大值集下限的双方对抗游戏. 在 G 和 D 由多层感知器定义的情况下, 整个系统可以用反向传播进行训练. 在训练

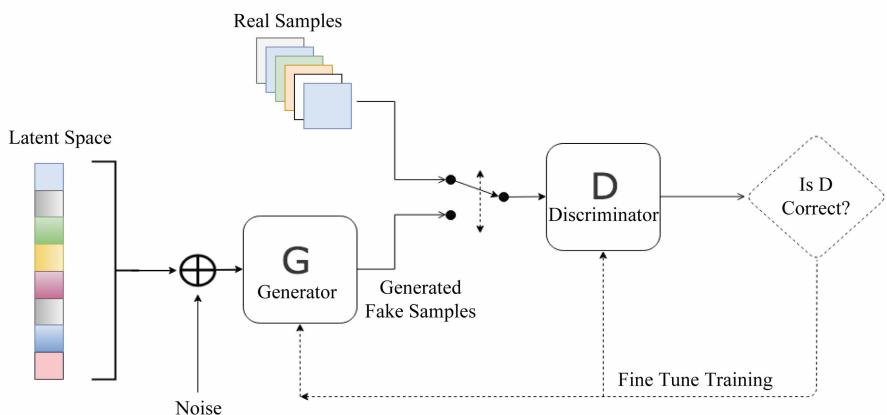


Fig. 7 Illustration of GAN

图 7 对抗生成网络结构示意图

或生成样本期间,不需要任何马尔可夫链或展开的近似推理网络。

目前GAN最常使用的方面就是图像生成,如超分辨率任务、语义分割、图像风格转换等。在超分辨率任务方面,Ledig等人^[24]提出SRGAN,可以使用GAN将低分辨率图片一一对应地生成高分辨率图片,训练需要高分辨率图片,将高分辨率图片用双线性插值法生成低分辨率图片后,分别作为真实数据和虚假数据输入到GAN网络中,最终训练得到的模型可以将低分辨率图片生成为高分辨率图片。

在图像风格转换方面,Gatys等人^[25]提出使用GAN的图像风格转换网络,可以将纹理转换问题由专有的固定的非参方法变为提取高层抽象特征的深度神经网络方法,从而生成与已有艺术绘画风格类似的高质量的新图像,如图8所示。

在语义分割方面,Couprie等人^[26]提出了使用GAN的语义分割模型,将传统的分割过程看作一个生成过程,在零和博弈的框架下,生成器可以生成难以被判别器区分的语义分割图片,其处理流程如图9所示。

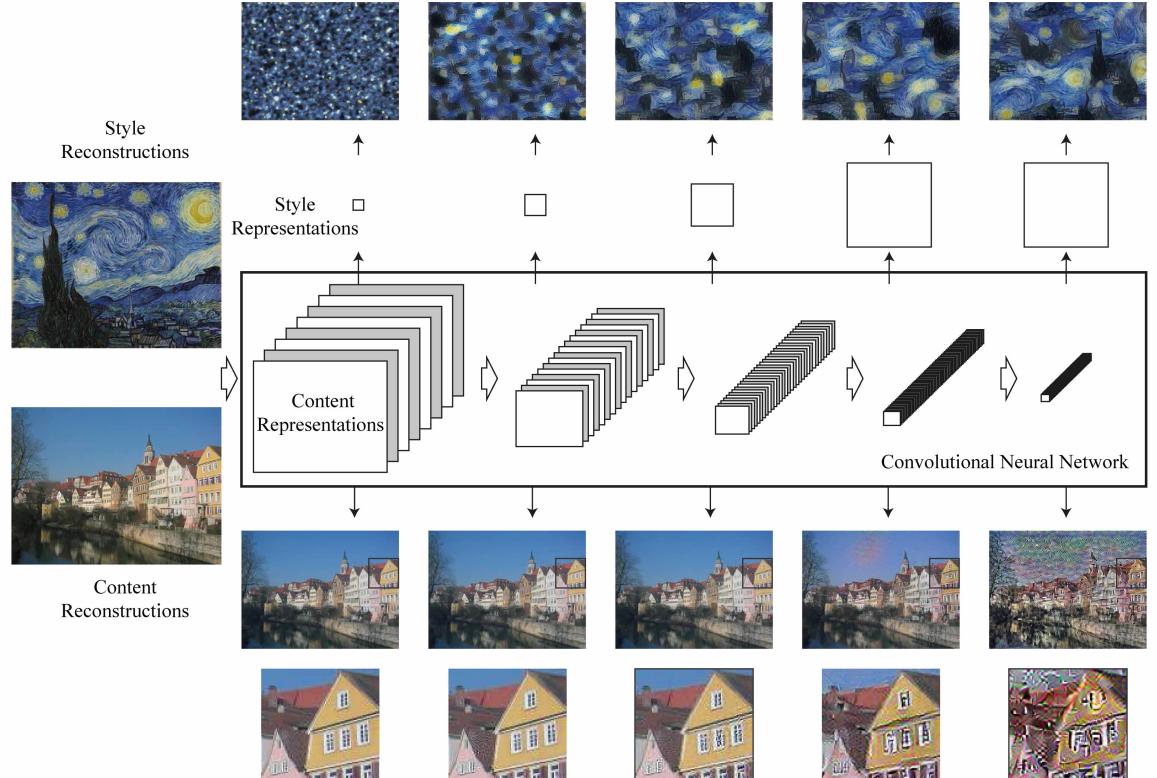


Fig. 8 Image style transfer based on GAN^[25]

图8 基于GAN的图像风格转换^[25]

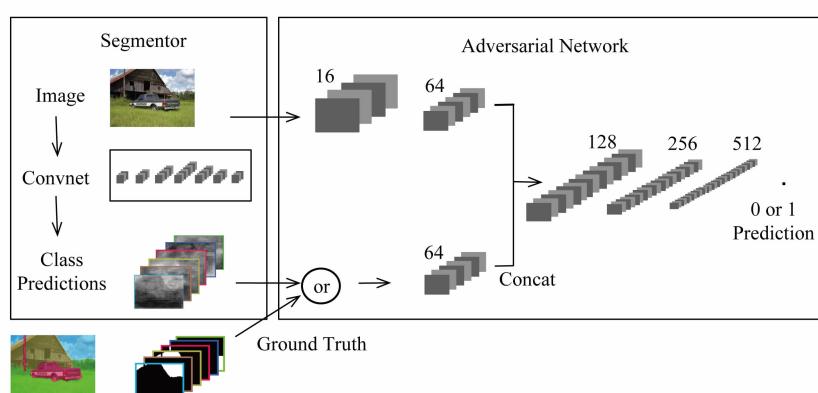


Fig. 9 Semantic segmentation based on GAN^[26]

图9 基于GAN的语义分割模型^[26]

1.4 递归神经网络

递归神经网络(recurrent neural network, RNN)是一种比较具有代表性的循环网络。1.1~1.3节提到的多种神经网络模型中,数据和信息的传递是以模型的“数据流”为方向的,也就是单一传递方向。在RNN中,核心思想是利用顺序信息,即对模型序列中的每个元素执行相同的任务,其输出取决于先前的计算,网络展开形式表示如图10所示:

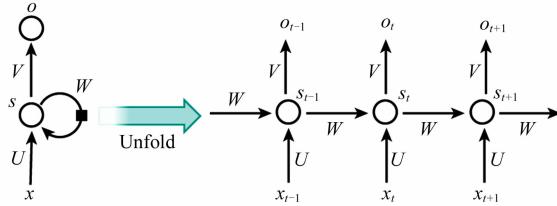


Fig. 10 Architecture of RNN

图 10 递归神经网络结构

递归神经网络模型考虑的是具体任务的循环模式,若任务为5个单词序列的预测,则该网络将展开为5层神经网络,每个单词为一层,每层的输入为当前层显性和隐性的输入。特别的隐藏状态S可以是网络当前的运算记忆,代表所有先前时间步骤中发生的事件信息。与传统深度神经网络不同的是,RNN中的主要参数U,V,W在所有步骤中共享相同的参数。这反映了每一步运算都是在执行相同任务的事实,只是使用不同的输入。这种网络训练方式大大减少了需要学习的参数总数。

递归循环网络主要应用在自然语言处理领域,包括语言模型生成文本、机器翻译和语音识别等。语言模型是对序列语句的可能性预测,即给定一系列单词,预测给定前一个单词的当前单词概率。Mikolov等人^[27]提出对长短不一的递归网络进行混合构建语言模型,从而对语言模型进行了完善。

Liu等人^[28]通过构建不同长度RNN序列和输出的关系的重复利用,模拟了翻译过程中的不同语序问题。语音识别是通过处理给定的声学信号输入序列,利用网络模型对一系列语音片段进行预测。Graves等人^[29]介绍了一种语音识别系统,直接对音频数据进行文本转换,不需要中间的语音表示。该系统基于一个深层双向递归神经网络结构和CTC模型记性标签的预测判断^[30],如图11所示:

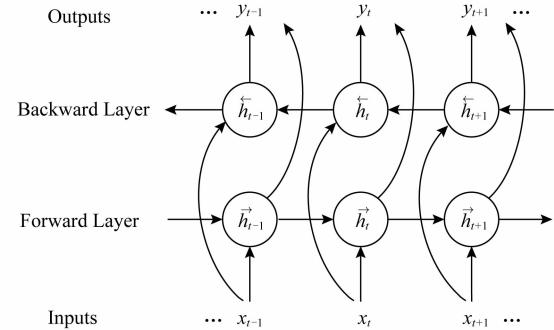


Fig. 11 Bidirectional RNN^[29]

图 11 双向递归神经网络^[29]

1.5 其他深度网络模型

深度神经网络模型的结构还有很多,除了递归神经网络之外,深度置信网络(deep belief network, DBN)^[31]也是广为应用的一种网络结构模型。如图12所示,DBN由受限玻尔兹曼机(restricted Boltzmann machine, RBM)堆叠而成,前一个受限玻尔兹曼机的输出作为下一个的输入,由此进行深度迭代。受限玻尔兹曼机是一种具有随机性的生成神经网络结构,它本质上是一种由具有随机性的一层可见神经元和一层隐藏神经元所构成的无向图模型。它只有在隐藏层和可见层神经元之间有连接,可见层神经元之间以及隐藏层神经元之间都没有连接。在DBN的训练过程中,依赖于“贪婪思想”,即在每一层的

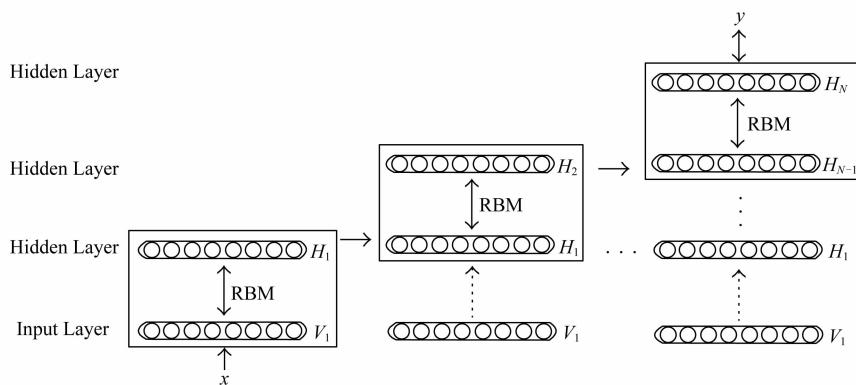


Fig. 12 Architecture of DBN^[31]

图 12 深度置信网络结构示意图^[31]

RBM 网络确保自身层内的权值对该层特征向量映射达到最优,自底向上地对每一个 RBM 进行参数的学习。最后在整体网络的最上层设置反向传播层,对整个结构进行自顶向下的参数优化。这是一种双向的训练加微调的模型训练方法。

自编码器(AutoEncoder)也是神经网络的一种,通过训练后能够尝试将输入复制到输出。自编码器由编码器和解码器 2 部分组成,编码器将输入编码成隐藏空间表示,解码器通过重构来自隐藏空间的表示获得输出。与其他神经网络关注输出层和错误率不同的是,自编码器关注的是隐藏空间。单纯地复制输入到输出并没有太大用处,实际应用中往往通过添加其他约束,使得自编码器能够学习到数据的有用特性。一种常见的约束是限制隐藏空间的维度,这种情况下,自编码器被称为欠完备的。通过训练欠完备的表示,可以迫使自编码器学到数据中最有代表性或者说最为显著的特征,其中一个比较典型的应用就是用于图像的去噪,如图 13 所示:

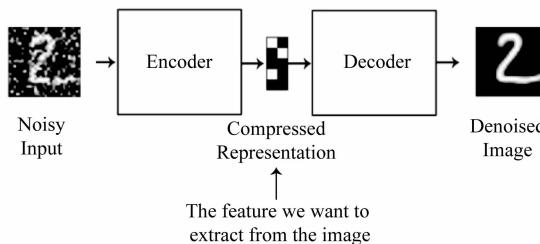


Fig. 13 Illustration of AutoEncoder

图 13 自编码器示意图

自编码器虽然结构简单,但是在机器学习领域有重要的作用。Rumelhart 和 Hinton 等人^[32]最早提出自编码器作为学习特征表示的方法。目前自编码器已经被成功应用于降维和信息检索任务。降维是将特征映射到更低的维度进行表示,是深度学习最早的一批应用之一,例如,Rumelhart 和 Hinton 等人^[32]在 2006 年的工作。数据的低维表示可以在一些深度学习任务中用于提高性能,根据 Salakhutdinov^[34] 和 Tieleman^[35] 的观察,低维空间的特征表示能够将语义相关的样本映射到相邻的位置,在深度学习任务中能够提高泛化能力。相比降维,信息检索从自编码器的研究中获益更多,它要求从数据集合中查询类似的条目。当使用自编码器对数据进行降维表示时,信息检索任务不仅能像其他任务一样获得降维表示所带来的益处,还使得低维空间的某些搜索变得极为有效。这种用降维表示实现信息检索的方法已经被广泛应用于文本输入^[34]和图像输入^[34,36]等研究中。

神经网络结构还有很多,由于篇幅关系,本节只介绍了当前多种典型的网络结构,其他网络结构不再赘述。

2 面向深度学习的几何数据表示

数字几何模型的表示方法主要有两大类:实体表示和边界表示。实体表示包括实体几何、点云、体网格以及体素等;边界表示包括表面网格、参数曲面、细分曲面、隐式曲面等。不是所有的表示形式都适用于深度学习,目前只有图像、体素、点云和网格等几种形式可以用于神经网络模型,如图 14 所示:

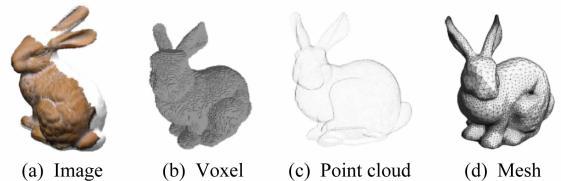


Fig. 14 Popular representations of digital geometric models for deep learning

图 14 适用于深度学习的数字几何模型常用表示方法

2.1 体素

几何模型的体素表示是最接近于图像的表示形式,它具有标准的定义域,因此可以直接应用图像分析和处理中常用的卷积神经网络。这种体素方法的优点是可以表示任意拓扑,允许较大的形状变化。Wu 等人^[37]首先将深度模型引入到三维体素表示的模型分析,他们利用一个三维深度置信网络——3D ShapeNets,直接从原始数据中学习不同种类的模型和模型任意姿态的特征分布,从而实现了三维物体的自动识别和修复。类似地,Maturana 等人^[38]和 Qi 等人^[39]采用同样的三维体素表示,利用 CNN 实现了超过 State-of-the-art 的高准确度物体识别和模型分类,图 15 所示为其网络结构和网络不同层所学习到的特征。尽管利用三维体素表示几何模型比较简单直观,而且可以直接利用卷积神经网络进行分析处理,但是随着模型体积的增大、模型的分辨率增加,受限于计算机内存和显存大小,深度学习模型通常只能对低分辨率模型进行分析,难以处理高精度的模型。

而将体素与八叉树数据结构进行融合的方法正好可以从一定程度上减轻内存消耗的问题。八叉树模型是一种具有代表性的栅格数据模型。根据八叉树不同的存储方式,将其分为规则八叉树、线性八叉树以及一对八式八叉树等。不同的存储结构的空间

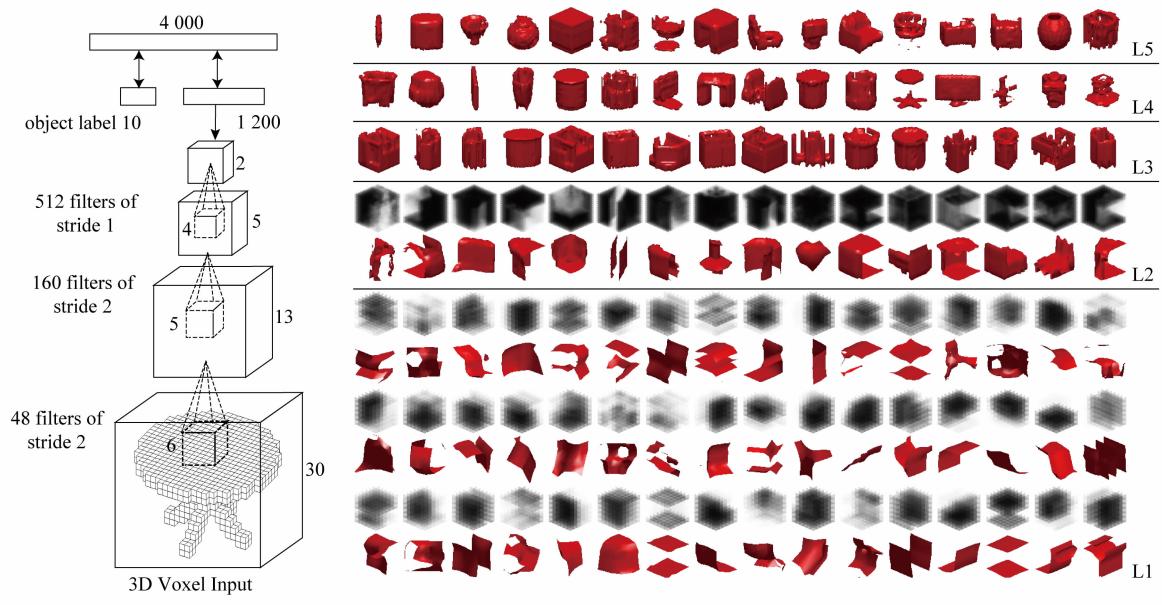


Fig. 15 3D object recognition based on voxel representation and deep model^[37]

图 15 基于体素表示和深度模型的三维物体识别^[37]

利用率及运算操作的方便性是不同的。Steinbrücker 等人^[40]通过八叉树数据自适应离散体素来弥补一般体素分辨率低这一缺点,在 CPU 上实现了实时的三维体映射。Riegler 等人^[41]提出了 OctNet 网络用于稀疏三维数据的深度学习表示,它使用混合网格-八叉树结构,将八叉树(Octree)结构深度限制在

一定的层数范围内,优化了网络性能,可以训练更深的神经网络而且处理的模型分辨率也得到进一步提高,基于八叉树结构的数据表示方法如图 16 所示。Wang 等人^[42]采用类似的表示方法提出了 O-CNN 来对三维模型进行分析。Tatarchenko 等人^[43]提出了一种新的卷积解码器结构,用于生成以八叉树表示

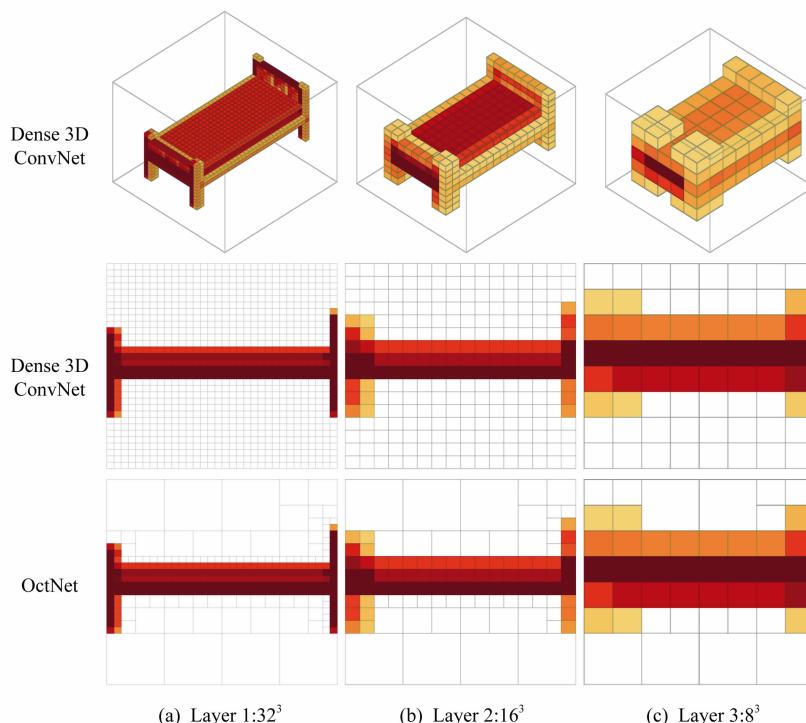


Fig. 16 Voxel representation based on Octree^[41]

图 16 基于八叉树结构的体素表示^[41]

的高分辨率三维输出。Häne 等人^[44]在三维重建中应用八叉树结构体素表示来进行三维几何预测。

2.2 多视角图像

由于三维模型的表示通常没有二维图像那样标准的定义域,体素表示方法又受限于存储往往丢掉模型上的几何细节,因而另一种比较直观的处理方法就是将三维模型投影为二维图像,但是投影会造成遮挡、不可见而引发数据的丢失,因此需要在多个视角下进行投影,用多个视角下的图像来表示三维模型。多视角图像通常是在以物体中心为球心的球面的各个位置上利用虚拟相机对物体成像,或是直接对物体进行投影。这样生成的一组图像具有更多的特征信息,对物体的描述更加完整。

在多视角图像生成的基础上,使用经典、成熟的二维图像深度学习网络进行训练,自然可以较好地

从各个视角图像中提取到三维物体的特征。图像与图像之间也可以直接用于比较,因此不同于以其他数据表示形式开展深度学习的研究中人们关注如何从数据中提取特征,在多视角图像方面人们更加关注如何将多视角下图像的特征信息进行融合,获得更好的网络模型,比如 Guo 等人^[45]利用多视角下的二维图像标签来指导三维模型的标记,并取得了很好效果。Su 等人^[46]提出了一个由卷积神经网络并联和串联组成的网络模型用于实现利用多视角图像对三维物体进行分类。如图 17 所示,网络首先由一组共享参数的卷积神经网络从不同视角图像提取特征。在特征融合方面,该模型并未简单采取求和的思想,而是利用一个卷积神经网络学习如何将多个特征图进行融合。该网络在模型分类方面的性能要明显优于基于体素表示的卷积神经网络。

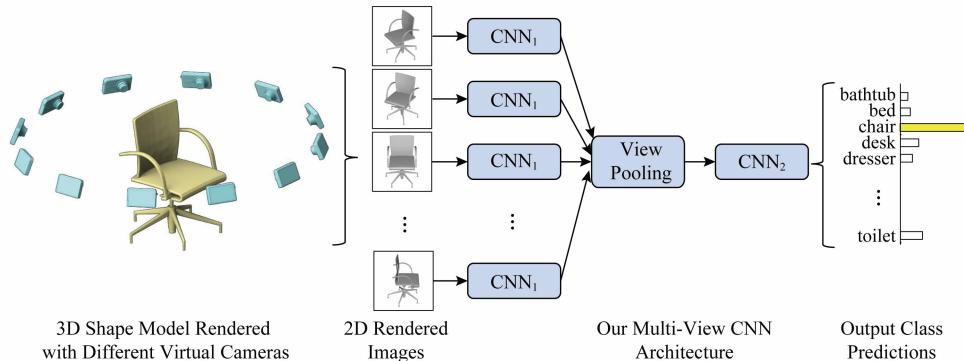


Fig. 17 Learning multi-view feature fusion using CNN^[46]

图 17 利用卷积神经网络学习多视图的特征融合^[46]

Huang 等人^[47]将多视角图像聚焦于模型的局部点,并且采用以模型局部点为中心、多种半径的球面为成像位置,使得获得的多视角图像既包含局部点的局部特征,又具有全局特征,如图 18 所示。由此使用卷积神经网络获取的局部点描述子也兼具局部与全局性。描述子可以用于模型匹配、结构分割以及

语义分割。Kalogerakis 等人^[48]将多视角图像与深度图结合,提出了一种利用全卷积网络与条件随机场对三维模型进行分割的方法。该方法同样采用局部与全局的多视图生成思路,并一同生成了深度图。在生成多视图的过程中记录了二维图像中每个像素对应的三维模型三角面片。利用多视角渲染图像与

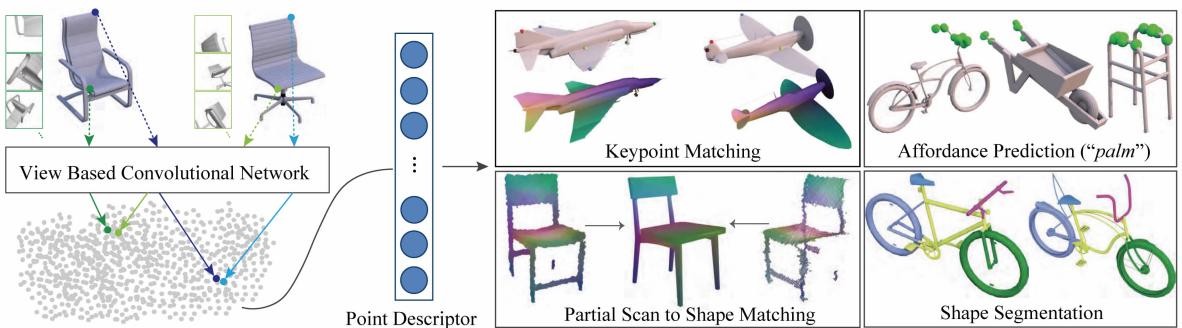


Fig. 18 Extraction of local shape descriptor using images from local and global views^[47]

图 18 使用局部与全局视图获取模型局部点的描述子^[47]

深度图像进行学习,获得图像中一个区域对应不同特征类型的置信度,再将图中的特征识别结果投射

到三维表面。最后利用条件随机场实现模型分割,其分割结果如图 19 所示:

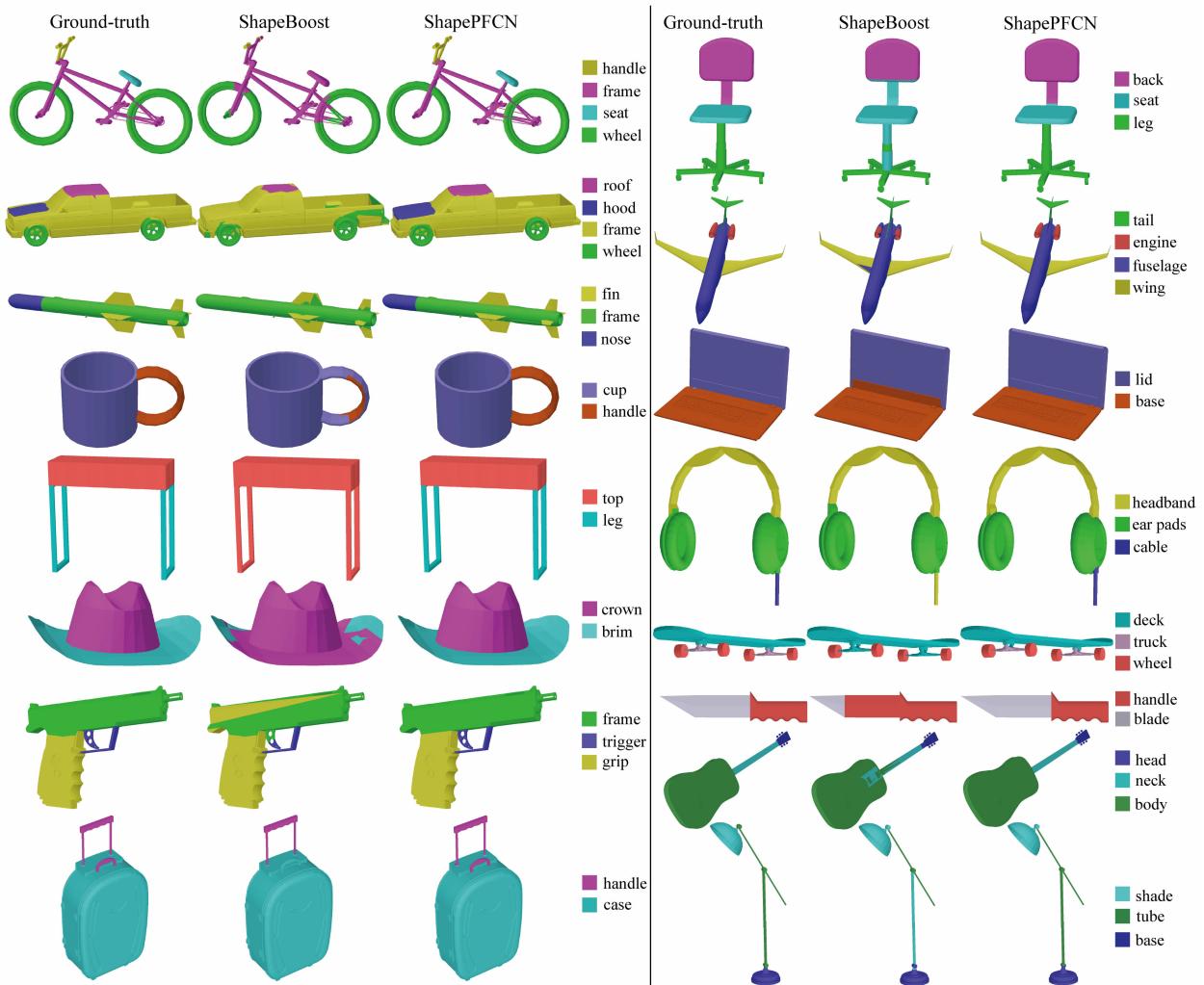


Fig. 19 3D shape segmentation results based on FCN and multi-view CNN^[48]

图 19 基于全卷积神经网络和多视角视图的三维模型分割结果^[48]

2.3 点 云

三维几何模型的另一种常见表示方法就是点云,点云通常是由激光雷达扫描得到的数据,由于其数据量庞大,且点云数据固有的无序性、无连接性以及特征旋转不变性等特点,没有图像体素那样标准的数据表示结构,无法直接将数据向量化,所以难以直接应用于深度学习模型。

Qi 等人^[49]首先提出了 PointNet 卷积网络方法,如图 20 所示,它首先通过一个空间变换网络解决了点云旋转问题,在网络中对每个点进行一定程度的特征提取后,采用 MaxPooling 对点云整体提取出全局特征,解决了点云的无序性问题。PointNet 最终可以提取出点云系统固定维数的特征,实现了点云的特征向量化,可以用于三维物体的分类、分割

等应用。之后他们又对其进行进一步改进,提出了更新的 PointNet++^[50],解决了 PointNet 无法很好地捕捉由度量空间引起的局部结构问题,可以更好地适应精细场景的识别,对复杂场景具有更强的泛化能力。

由于点云这种不规则的数据形式(顶点集合)和社交网络(个体集合)比较类似,因此在非规则数据分析领域常用的图结构也可以用来对点云进行表示,从而利用图卷积神经网络(graph convolutional network, GCN)^[51-53]对点云进行分析。Qi 等人^[54]在点云的基础上根据 K 近邻的方式建立图结构,提出了 3DGNN,图中的结点由从二维图像中提取的特征来进行表示,从而实现语义分割。由于不同的点云模型会造成不同的最近邻图结构,所以无法直接

应用图拉普拉斯(graph Laplacian)的谱分析,因此Yi等人^[55]在频谱域中利用函数映射方法将这些不同的图结构对齐到一个标准的空间,提出了SyncSpecCNN,可以用于多种不同的应用任务,比

如三维模型分割、三维关键点检测等,都取得了State-of-the-art的效果,其网络结构如图21所示。其他类似的针对图神经网络进行改进的还有很多,比如RGCNN^[56],DynGCNN等^[57]。

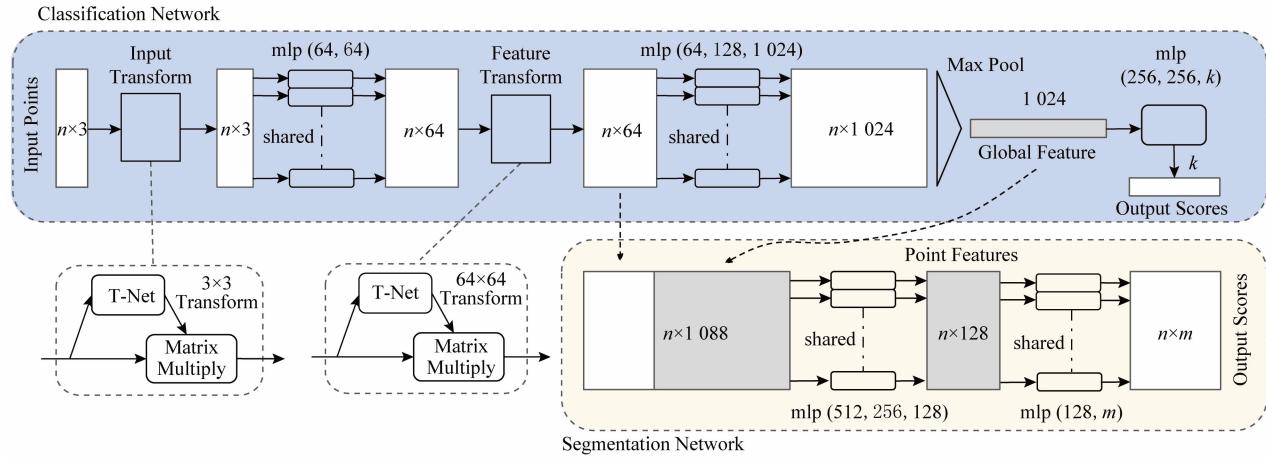


Fig. 20 Architecture of PointNet^[49]

图 20 PointNet 网络结构^[49]

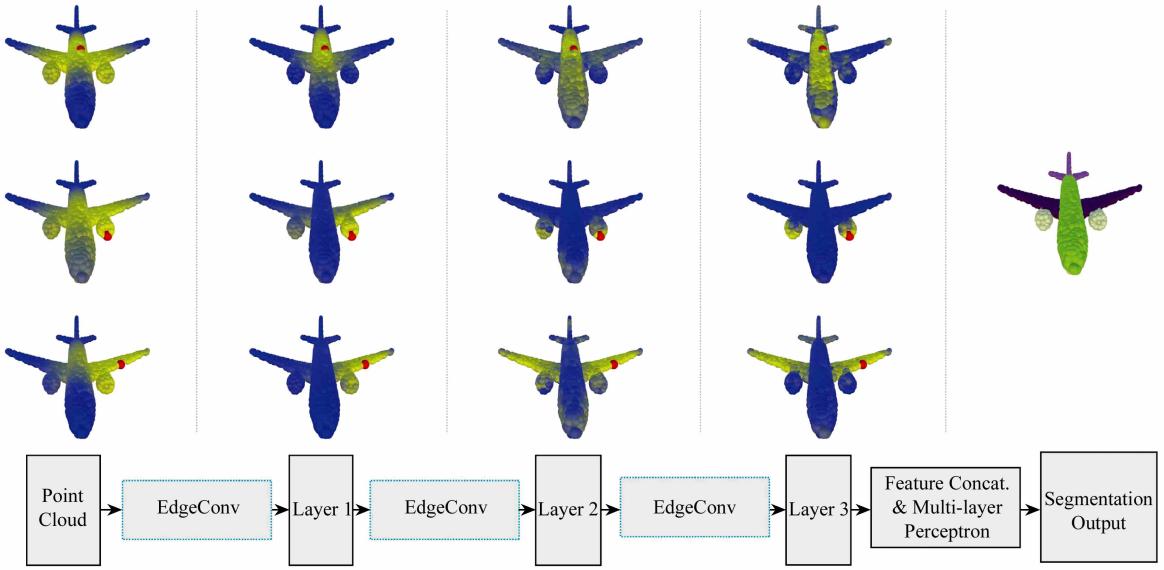


Fig. 21 Architecture of DynGCNN^[57]

图 21 DynGCNN 结构示意图^[57]

2.4 网格

网格模型是目前使用最为广泛的几何模型表示方式,但是它不具备标准的定义域,很难直接应用于深度模型,所以通常需要手动地在模型表面建立标准定义域,或者手动提取模型上顶点的特征用于神经网络的训练。由于网格表示实际上也是一种图结构(顶点和边),因此可以用2.3节中的图卷积神经网络来进行分析和处理^[57-62]。另一种方式就是在网格表面建立标准的定义域,比如Masci等人^[63-64]利

用测地线的性质在网格模型表面建立如图22所示的局部极坐标系,从而可以在网格上实现与图像卷积类似的卷积操作,这样就将传统的CNN扩展到了非欧的几何流形上,可以利用CNN的数据抽象能力提取模型的各种几何特征,应用于不同的场景。在此基础上,Boscaini等人^[65]利用各项异性的扩散核函数进行了进一步改进,并用于学习模型之间的对应关系,得到了更好的效果。在文献[66]中,同一作者采用加窗傅立叶变换方法建构了空间-频域上的

CNN 架构. Monti 等人^[67]提出了一种更加通用的框架来利用深度模型处理非欧数据,例如图结构数据和几何流形,而之前所述的这些模型实际上都可以看作这种框架的某种特殊情况,因此更具通用性.

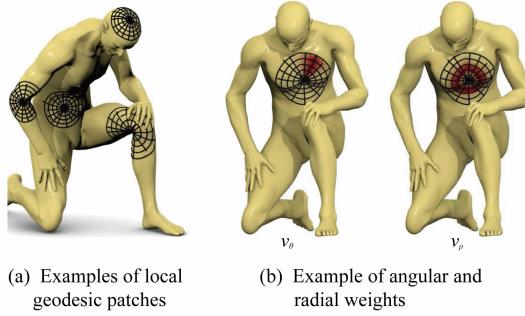


Fig. 22 Geodesic polar system built on mesh surface^[63]

图 22 网格模型表面建立测地线极坐标系^[63]

除了在网格表面建立标准的定义域之外,还有

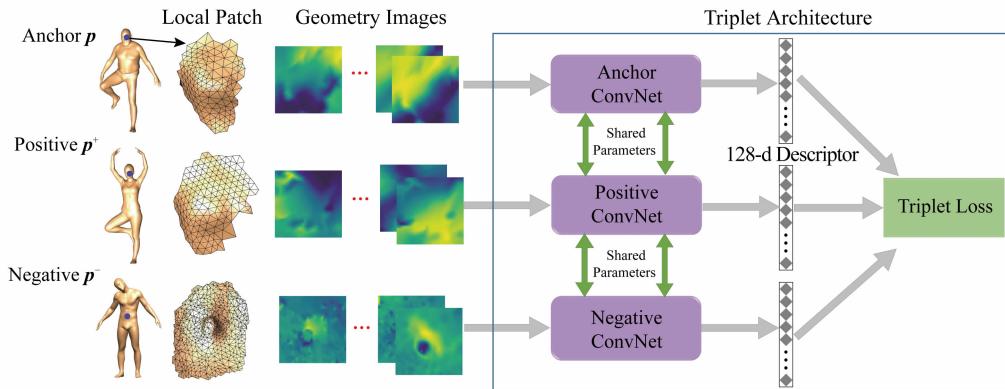


Fig. 23 Extraction of high-level features using local low-level features^[72]

图 23 利用网格上低层次的局部特征抽取高层次特征^[72]

3 基于深度学习的数字几何处理

3.1 模型匹配与检索

三维模型的匹配和检索是数字几何分析领域研究最为广泛的方向之一. 传统的匹配和检索方法^[71-72]都是通过手工设计的描述子来实现三维模型上的顶点之间相似性比对,通常称为模型的匹配或者模型的对应,而对三维模型整体而言的相似性比对往往应用于模型的检索. 深度学习技术已经在自然语言处理、图像处理领域证明了其强大的特征抽取和描述能力,因此很多研究者开始研究如何利用深度学习技术来提高模型匹配和检索的准确率,扩展其应用场景.

Wei 等人^[73]首次采用深度学习方法来学习三维人体扫描数据之间的对应关系,这种方法仅仅需要

一种常见的针对网格表示的深度学习方法是利用传统的手工定义的描述子来表示模型的局部特征,然后利用神经网络对这些低层次的特征进行进一步地抽象提取,从而实现不同的应用任务,同时可以取得更好的表现. Guo 等人^[68]根据 7 种传统的几何特征对网格模型上的顶点进行描述,然后将这种描述子转换为二维矩阵,利用经典的 CNN 结构对这些特征进行进一步抽象提取,实现了三维模型的语义标记,相比传统方法取得了巨大的准确度提升. Chen 等人^[69]使用热核特征作为描述子描述每个点的特征,并使用三联网络的方法训练得到具有更好区分特性的每个点的特征,从而对不同模型上的点进行匹配. Wang 等人^[70]利用网格模型上的局部特征描述子定义几何图像(geometry image),然后利用如图 23 所示的三联神经网络来学习模型之间的对应关系.

部分几何信息,不要求被扫描的人体数据具有相似的朝向.与传统的模型匹配对应方法不同的是,该方法不是直接训练网络来求解匹配问题,而是利用卷积神经网络来解决一个人体区域的分割问题,从而提高学到的描述子在区域边界的平滑性,提高匹配的准确度,其对应结果如图 24 所示. Zeng 等人^[74]

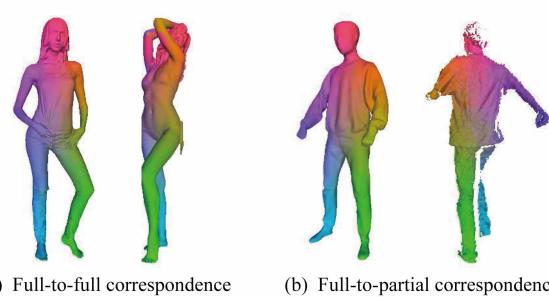


Fig. 24 Illustration of shape correspondence^[73]

图 24 模型对应结果示意图^[73]

提出了另一种数据驱动的三维关键点描述子,用于匹配更加复杂的真实世界的深度扫描图像。如图 25 所示,Zeng 等人使用三维体素表示的卷积神经网络结构学习局部体素块的特征来建立不完整的扫描数

据之间的对应关系,这种方法不仅可以在新的场景中建立局部的几何对应,而且可以用于不同的任务和空间尺度。这种基于体素表示的网络结构的缺陷是由于内存消耗的原因导致它们的分辨率比较低。

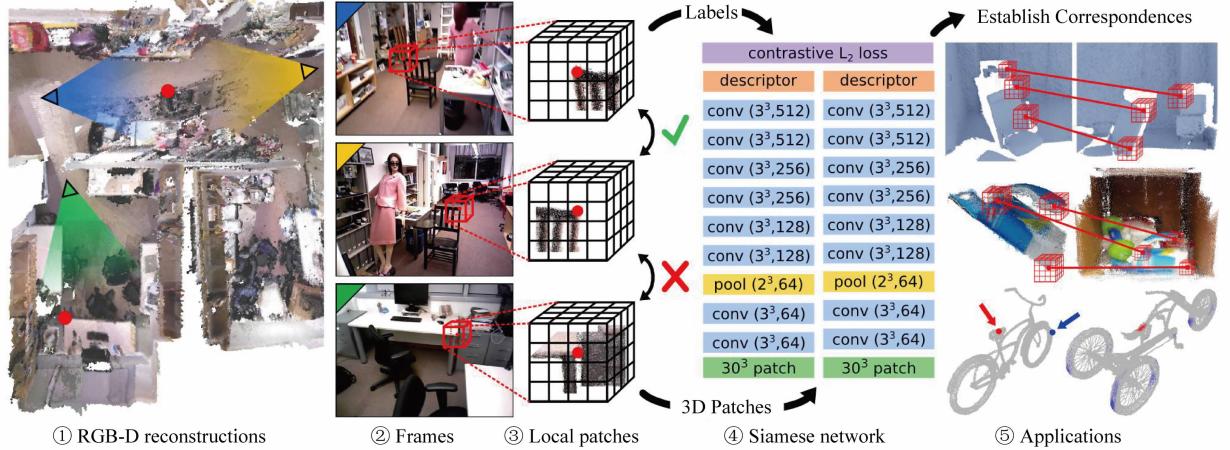


Fig. 25 Key-point matching based on voxel representation^[74]

图 25 基于体素表示的关键点描述子匹配方法^[74]

为了解决卷积神经网络不能直接应用于网格表示的三维模型的问题,Bronstein 等人提出了基于测地线的卷积神经网络(geodesic convolutional neural network, GCNN)^[63,65,75],如图 26 所示。这是将传统的卷积神经网络向非欧几何流行进行的扩展,通过在模型表面利用测地线建立标准的上下文结构,使得图像上的卷积操作可以在网格表面进行,从而可以利用神经网络结构对传统的手工定义特征描述方法进行抽象,从而获得更加高层次的特征描述子,可以用于建立模型之间的对应关系,而且其准确率远远超过了传统的描述方法。与这类直接学习对应的方法不同,Litany 等人^[76]利用深度残差网络以定义在模型上的稠密描述作为输入来学习模型之间的函数映射和线性算子,而这种映射关系又提供了一

种紧致的对应表示。

Wang 等人^[72]提出了一种新颖的深度学习框架,能够推导出基于网格表示的三维模型的局部描述子。不同于之前的基于卷积神经网络的方法来提取基于多视角图片或者直接提取形状的内在属性,这种方法参数化了多个尺度上的在关键点附近的局部信息,之后利用三重网络来进行学习,通过最小化三重损失函数从而得到模型之间的对应关系。类似地,如图 27 所示,Chen 等人^[69]根据传统模型上的扩散理论和热核函数提取多尺度几何特征描述,然后利用三重网络进一步抽象已有的几何特征描述,学习到新的特征描述具有更强的描述能力,可以获得更高的匹配准确性。

Huang 等人^[47]介绍了一种新的局部描述子,如

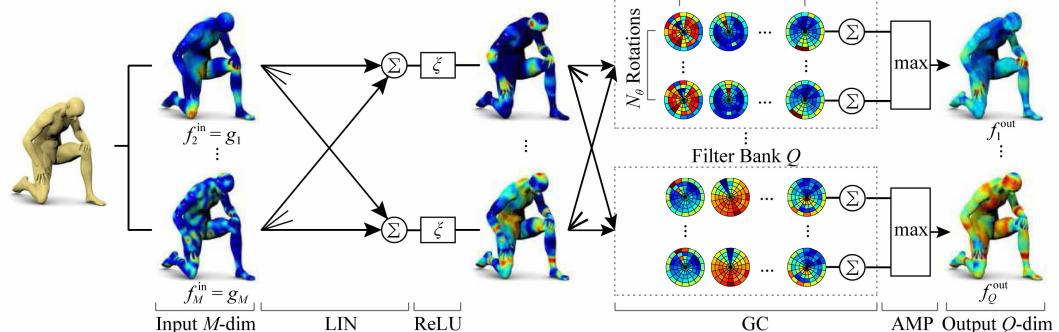
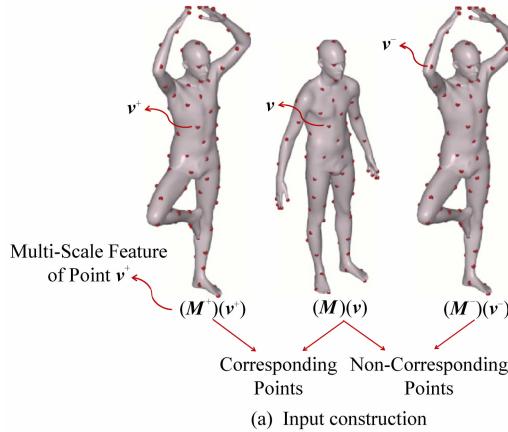


Fig. 26 Shape correspondence based on Geodesic CNN^[63]

图 26 基于测地线卷积神经网络的模型对应方法^[63]

图 28 所示,通过采用多视角和多尺度的处理,并通过比较传统的二维神经网络方式进行特征提取,其训练的基本条件是不同模型上几何或语义相似的点应该被嵌入相近的特征空间当中。这种描述子不仅可以用于模型匹配,还可以用于分割和模型部件的功能型预测等多种应用场景。Furuya 等人^[77]提出

了一种基于局部的模型检索网络结构,叫做 PWRE-net。利用深度神经网络,PWRE-net 从大量的局部-整体模型对中学习到一个共同的统一嵌入空间,这个空间中的局部模型和它对应的整体模型具有密切的关系,从而实现利用局部模型来进行三维几何模型的准确检索。



(a) Input construction

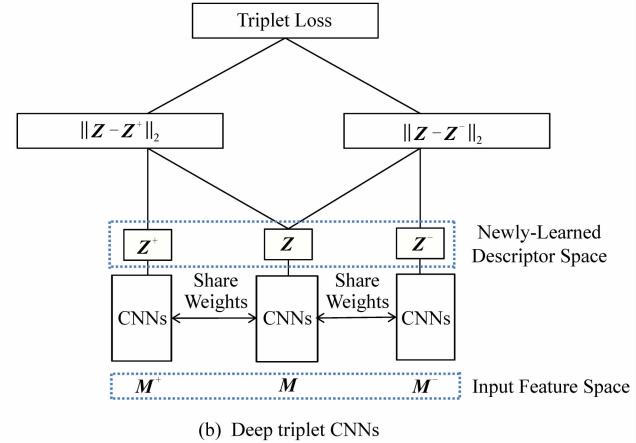
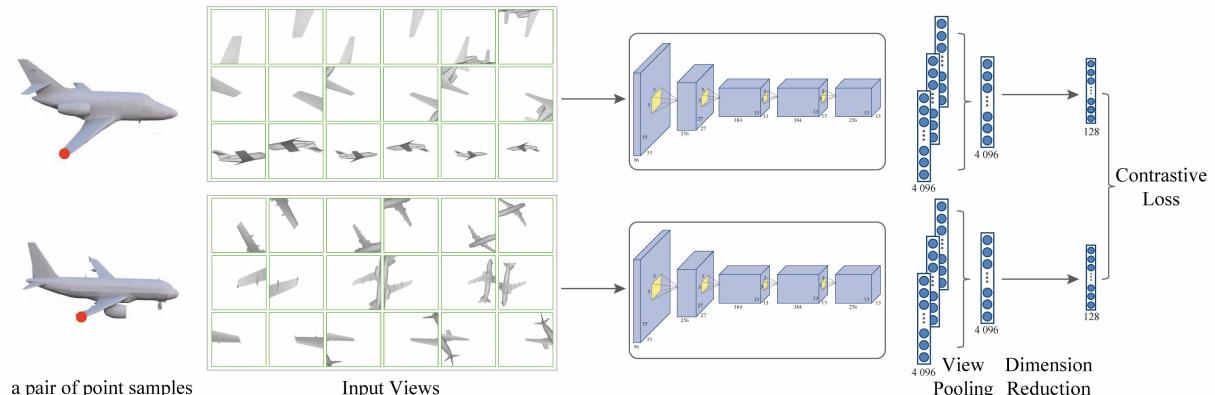
Fig. 27 3D shape matching based on deep triplet CNN^[69]图 27 基于深度三重网络的三维模型匹配方法^[69]Fig. 28 Multi-scale shape matching based on multi-view images^[47]

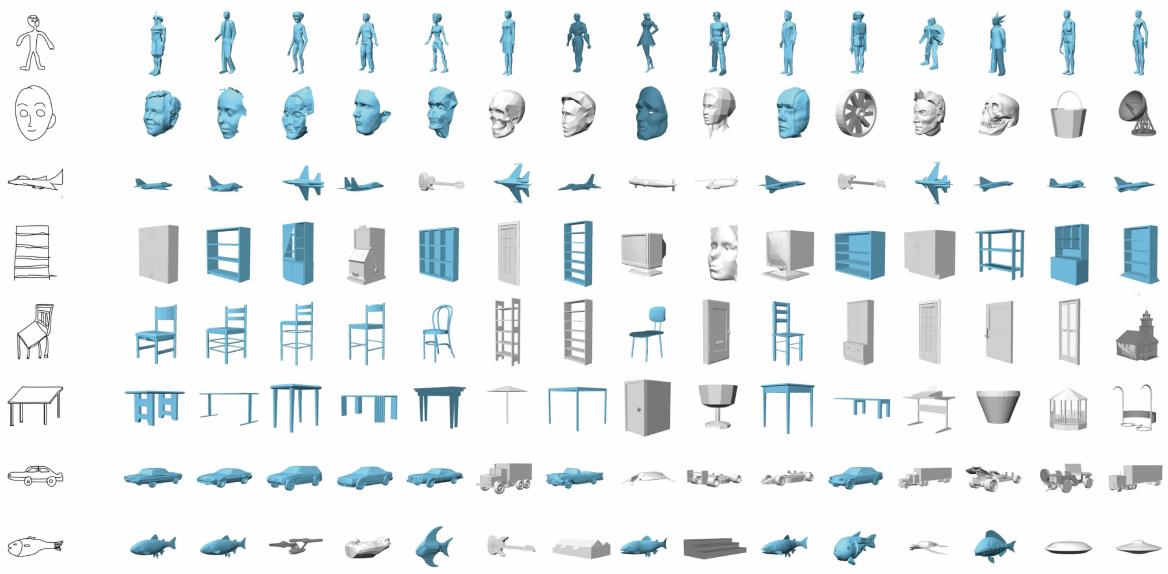
图 28 基于多视角多尺度的模型匹配

基于整体三维模型的描述可以用于模型检索,Zhu 等人^[78]利用自编码方式进行特征学习,通过将三维模型投影到二维空间,融合多个视角下二维图像中学习到的特征描述,实现三维模型的检索。Xie 等人^[79]提出了深度非线性度量学习方法,首先利用局部约束的线性编码对模型顶点进行编码,然后将编码系数直方图作为全局三维模型描述子,再用深度度量网络学习三维描述子到非线性特征空间的映射关系,从而实现三维模型特征描述的特征抽取,可以直接用于模型检索,具有很高的检索准确率。此外还有基于草图的模型检索方法^[80],即输入用户描绘的简单的草图,系统返回相似的三维模型,其结

果如图 29 所示。Wang 等人^[81]利用 2 个卷积神经网络,一个用于训练草图,一个用于训练多视角投影,然后实现了基于草图的模型检索。

3.2 模型分类与分割

与模型的匹配和检索类似,分类和分割也是深度学习应用于三维模型的典型方向,因为分类与分割同样需要对三维模型进行特征提取和描述。分割问题的特征描述是针对顶点或局部区域而言,分类问题是针对整个模型,传统的模型分割或分类方法都是依赖于手工定义的描述子和特定的聚类方法^[82-83],具有一定的局限性,而深度学习技术正好具有强大的数据抽象和描述能力。最初尝试利用深度学习进行



(a) Sketch queries

(b) Retrieved examples

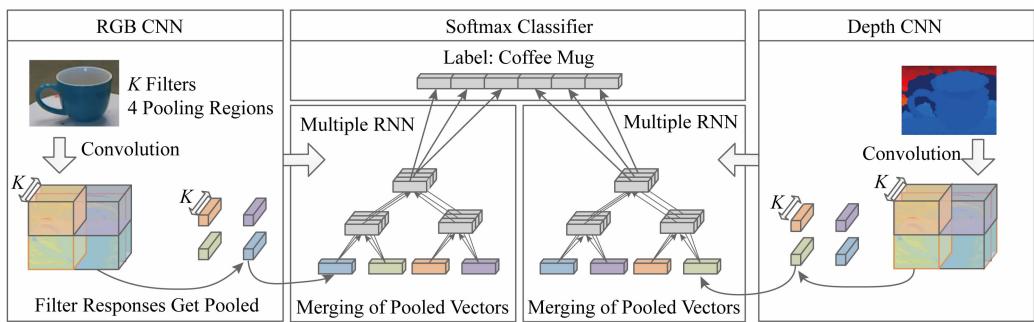
Fig. 29 Shape retrieval based on sketches^[81]图 29 基于草图的模型检索方法^[81]

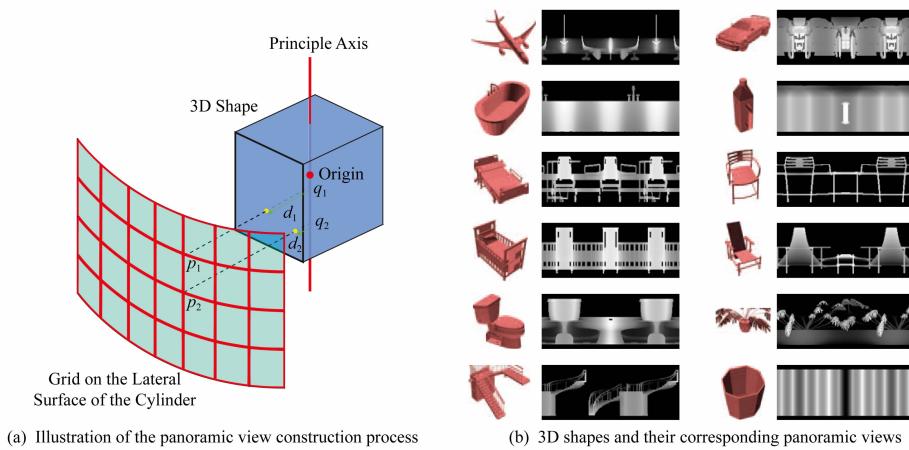
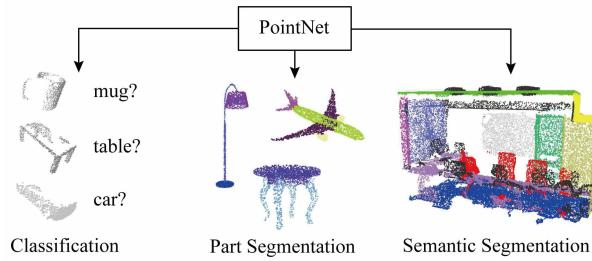
分类的方法结合了卷积和递归神经网络,用来学习和抽象特征,实现了如图 30 所示的 RGB-D 图像的分类^[84]. Gupta 等人^[85]对 RGB-D 的信息进行了扩展,不仅为每个体素提供了颜色和深度信息,还为它们编码了距离地面的高度和重力的角度,这一方法取得了显著效果. Eitel 等人^[86]融合了以上 2 种方法,提出了一种融合后的神经网络. 这些方法都是针对 2.5 维的物体进行研究. Wu 等人^[87]提出了对体素进行分类和检索的体积架构,通过学习不同类物体的复杂的三维形状分布,实现物体的分类和分割. 类似地, Maturana 等人^[88]提出了自己的针对体素的网络架构,并且可以实现实时的物体识别.

另一个主要的分类方法是将三维物体投影到二维平面上进行卷积. Shi 等人^[89]将三维物体投影到包围它的圆柱体上形成全景图,并对这个全景图进行卷积操作,从而进行二维上的特征提取来实现三

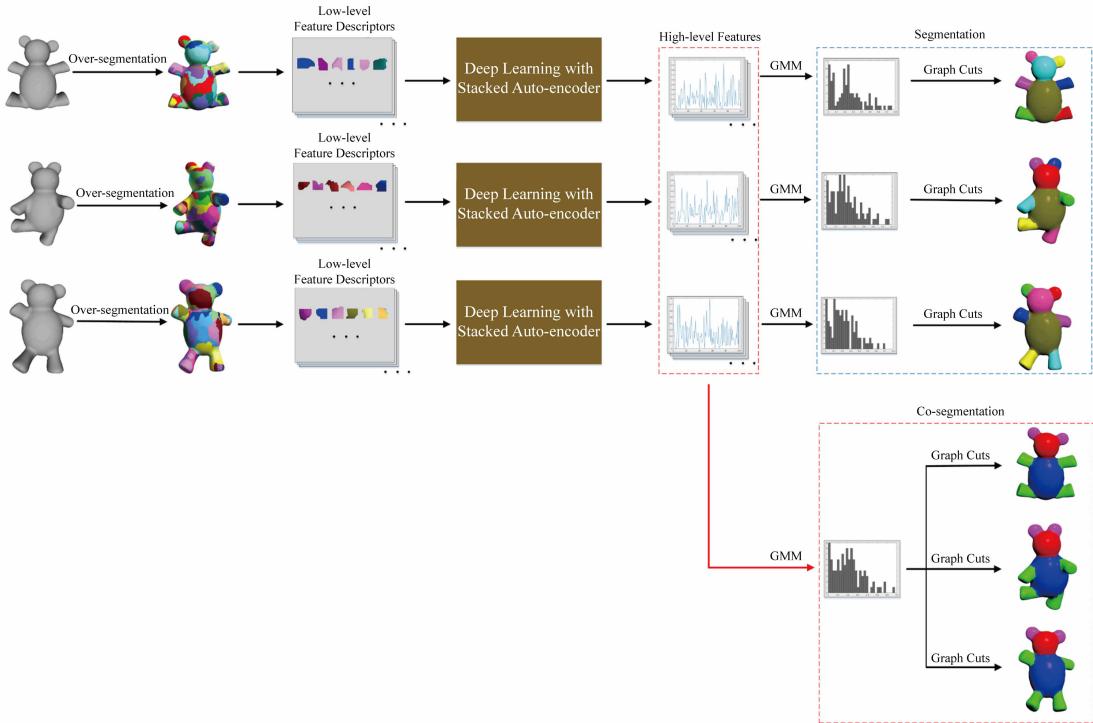
维模型的分类,具体过程如图 31 所示. Su 等人^[46]通过联合训练多视图的方式对物体进行分类,取得了较好的效果. Hu 等人^[90]通过三联神经网络进行度量学习,并通过生成与体素模型相对应的使用场景的方法来揭示物体的功能. 这种方法虽然简单,但是不能将学到的特征投射到模型上,因此很难进行分割.

Qi 等人^[49]将机器学习领域扩展到了点云,他们将每一个点输入多层感知机单独学习一个特征,并最后用取最值等对称性的操作获得全局特征表示从而进行分类,如图 32 所示,这种方法可以解决点云无序性的特点. 紧接着的 PointNet++^[50]采取多层次采样的方式,保证了网络对不同大小模型的鲁棒性,并且使用跳跃的连接方式,在特征传导过程中,将每一层学到的信息拼接在每个点的信息上,从而实现物体分割.

Fig. 30 3D object recognition based on RNN^[84]图 30 基于递归网络的三维物体识别^[84]

Fig. 31 3D object classification based on panorama projected images^[89]图 31 基于全景图投影的三维物体分类^[89]Fig. 32 3D shape classification and segmentation based on PointNet^[49]图 32 基于 PointNet 的三维模型分类和分割^[49]

最早利用深度学习技术对网格模型进行分割的方法由 Guo 等人^[68]提出。他们先利用多种手工定义的特征描述子将模型的顶点描述成一个向量, 然后将这个一维向量变换为二维的矩阵形式, 这样就可以直接利用图像中的卷积神经网络进行特征提取, 并实现了语义标记。Shu 等人^[91]将三维模型分解成多个小块, 然后利用深度学习从低层次手工定义的描述子抽取更高层次的语义描述, 再用聚类方法实现网格模型的分割和共分割, 如图 33 所示。之后 Wang 等人^[92]设计了一种形状全卷积网络 (shape

Fig. 33 (Co-)Segmentation model based on deep learning^[91]图 33 基于深度学习的模型(共)分割^[91]

fully convolutional network, SFCN),在这种网络中他们定义了基于图的卷积和池化操作,将图像分割中广泛应用的全卷积网络应用于三维网格模型的分割,取得了很好的分割结果。

除了外观以外,还可以从风格的角度对三维模型进行分类。物体的功能与其主要部分的总体形状和每个部分的排列密切相关,而风格与这些部分的几何

细节密切相关^[93],如图 34 所示为基于风格的三维模型分类.Lim 等人^[94]通过分析研究多视图的方式进行分类,每次输入一个一正一反的三元组,通过三联的神经网络学习他们的距离。学习的目标应满足 2 个条件:相似风格的物体足够近和不相似的物体足够远,对于任意一个模型,所有和它相似的模型都比和它不相似的模型要远,从而生成一个均匀的空间。

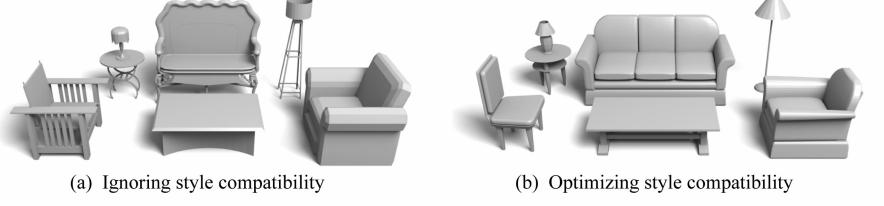


Fig. 34 3D object classification based on styles^[93]

图 34 基于风格的三维模型分类^[93]

3.3 模型生成

相比于三维几何模型,深度学习在二维图像领域应用较早。同时,多视角图像又是三维几何模型的一种有效的表示方式。利用多视角深度图像或利用多视角图像生成物体的可见外壳都可以有效地实现三维模型的重建。因此部分学者将三维几何模型生成问题转化为模型的多视角图像生成问题并开展研究。Hinton 等人^[95]提出了用于实现变换操作的自编解码器,可以对输入图像进行小幅度的视角变换。Flynn 等人^[96]训练了一种端到端的深度神经网络,可以保留场景相邻视图中共有的像素并对未知像素进行预测,网络可以用于真实场景相邻视图间的插

值。如图 35 所示,Alexey 等人^[97]在解码器中使用了上采样层与卷积层组合形成的“上卷积层”,使得解码器具有生成高分辨率图像、大幅度的视角变换、对模型进行缩放变换、颜色变换的能力。

空间变换网络^[98]通过在现有的卷积神经网络中引入一种可微的图像采样模块,使得网络具有利用数据的空间信息、对特征进行空间变换的能力。Zhou 等人^[99]使用卷积神经网络对像素点在相邻视图中的坐标变换进行预测,并能根据多视图输入对生成视图进行优化。如图 36 所示,Park 等人^[100]首先利用自编解码器推断在输入和输出视角图像中均可见的部分,并将对应的像素移动到目标视角图像

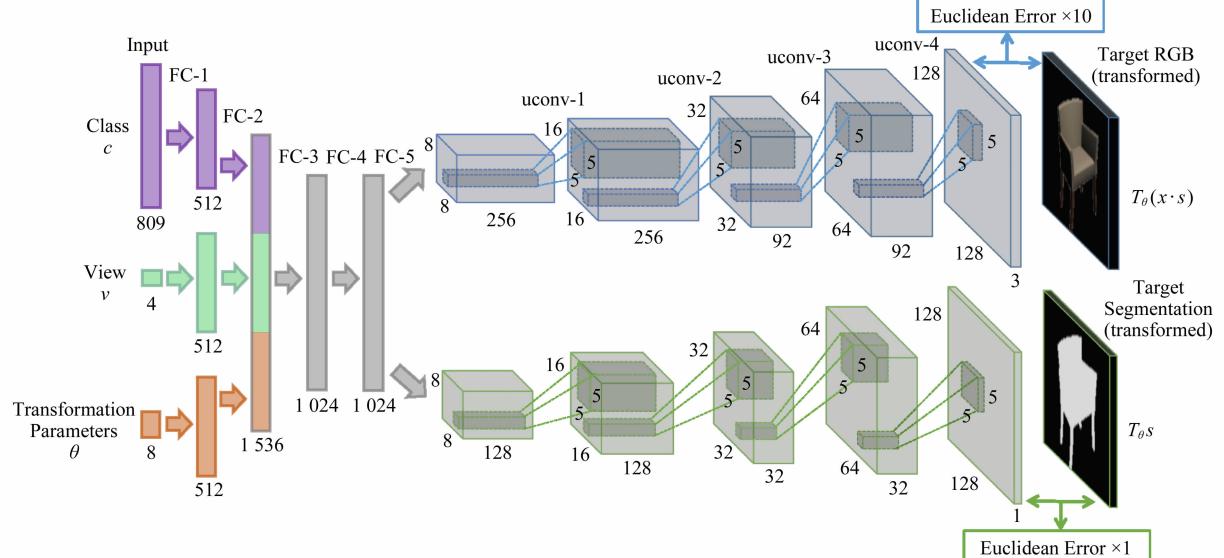


Fig. 35 Shape generation based on CNN^[97]

图 35 基于卷积神经网络的模型生成^[97]

中的位置;然后将输出图像剩余的生成问题转换为图像填充问题,利用神经网络根据中间结果来推断输入图像中不可见的部分。神经网络本身也利用对抗网络和特征识别网络作为损失函数来减少合成结果中的畸变、模糊和空洞等问题。

由于几何模型的体素表示可以直接应用于卷积神经网络,通过训练网络使其能够直接将模型在低维空间中的编码解码成模型的体素表示,从而可以免去由编码到多视角图像再到多视角图像建模这一过程。Girdhar 等人^[101]使用自编解码器构建了一个端到端的网络,分别从模型的单张图像以及体素获取表示编码,将 2 种表示编码的欧氏距离作为训练的损失函数,同时使用已有的模型体素对解码器进

行监督学习。由此产生的表示编码不仅能从二维图像中推理出来,而且可用于生成三维模型。对抗生成网络在图像生成方面取得了理想的效果,如图 37 所示,Wu 等人^[102]将其应用于三维体素,获取到模型样本在高维特征空间的分布,并将这一分布映射到低维空间中。对抗生成网络无需借助参考图像或已有模型,可以根据输入的低维特征向量直接输出三维体素模型。由于网络具有探索流形空间的能力,不仅能够生成全新的模型,也可对已有模型进行分类、索引、合成等应用。在此基础上人们将八叉树数据结构应用于体素表示以提高生成模型的精细程度^[42-43],但生成模型的分辨率依旧难以达到多视角图像的水准。

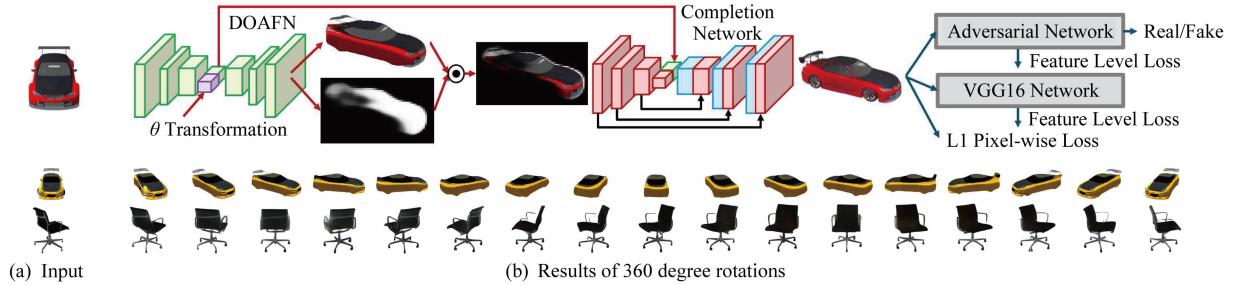


Fig. 36 Multi-view images generated from a single view image^[100]

图 36 由单一视图生成模型的多视角图像^[100]

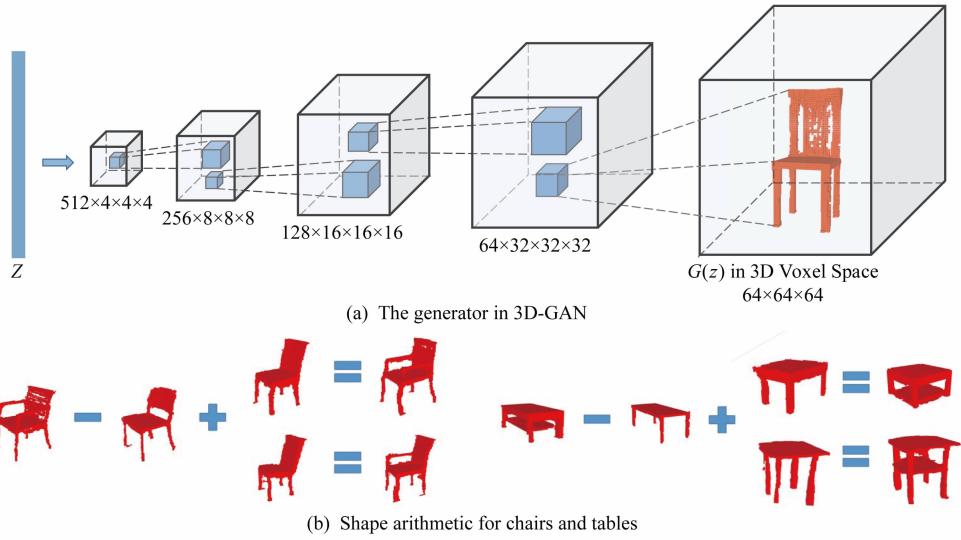


Fig. 37 Shape synthesis using GAN^[102]

图 37 使用对抗生成网络进行模型合成^[102]

Xu 等人^[103]利用基于递归神经网络的自编码器将模型的整体结构参数(模型零部件包围盒参数、对称性参数等)按层级编码成一个低维的特征向量。并利用对抗生成网络训练对应的解码器,从而组成一个整体结构与特征向量间的双向转换器。通过训

练使得解码器可将输入的随机噪声拟合成可表示合理模型结构的低维特征向量。最后再利用深度神经网络实现由结构参数到具体模型零部件的转换,与自编解码器构成完整的模型生成管线,在模型分类、局部匹配、模型生成、形状插值等问题上都有不俗的表现。

Wang 等人^[104]提出了一种由单视图图像生成网格模型的方法。如图 38 所示,利用从输入图像中提取的特征逐步对一个椭球进行变形从而产生目标模型。对于三维网格模型使用的非规则数据形式——图,该方法采用图卷积网络来处理椭球逐步变形获得的三维网格,使用卷积神经网络来提取二维图像特征,

并利用一个投影层使得网格中的每个顶点能获取相应的二维图像的特征。通过构建图的上池化层,使得网格顶点的数量逐渐由少到多,网格形状由粗糙到精细,既保留了全局信息,又具有细节表达,同时使得整个变形过程更加稳定。此外还有利用过程建模和神经网络从草图中重建并合成三维模型的方法^[105]。

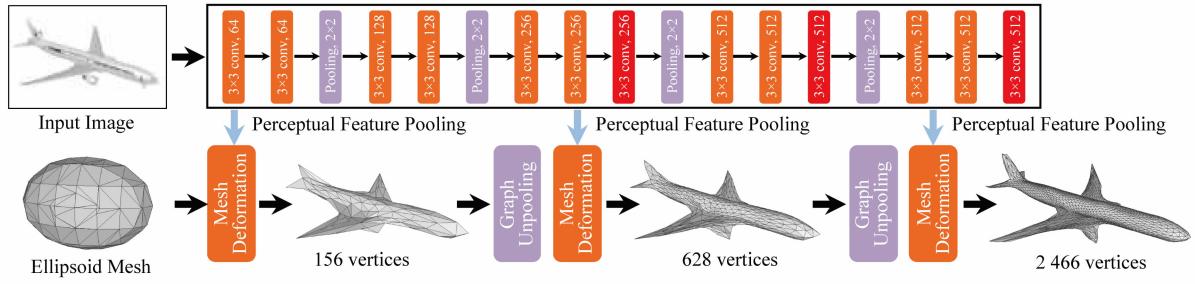


Fig. 38 Deform an coarse ellipsoid mesh into a refined mesh using CNN^[104]

图 38 卷积网络处理椭球网格变形生成精细模型^[104]

3.4 模型修复与重建

随着消费级激光扫描设备(比如微软的 Kinect 和英特尔的 RealSense)精度和效率的不断提高,几何模型的获取越来越容易。海量几何模型的涌入,为深度学习在几何处理与分析领域提供了坚实的数据基础。然而由于光照、遮挡、环境等外在因素影响,直接获取的原始几何数据不完整,需要对获取的原始几何数据进行修复与重建。传统的模型修复与重建算法^[106-108]主要依赖于较强的假设和结构先验知识,只能处理特定场景下的模型修复与重建,无法应用于大量的几何模型。近年来,随着深度学习在音频、图像、视频领域的突破性进展,研究人员已经开始将深度学习迁移到几何模型上,尤其是几何模型的修复与重建。与传统算法不同,基于深度学习的方法直接从海量数据中挖掘模型的结构信息,不再受限于具体的应用场景以及强假设先验信息。

相比于图像规则的定义域,几何模型的表示形式与具体的应用场合有关,进而也阻碍了深度学习

在几何模型上的直接推广。最直观的方法就是将几何数据用体素表示,然后将 2D 上的操作直接扩展到 3D 上。受图像修复中的自编码网络启发,Dai 等人^[109]提出了一种基于体素的自编码网络来对缺失模型进行修复与重建。如图 39 所示,在训练阶段,借助 3D 卷积操作将输入的残缺几何模型映射到高维空间(编码),然后再通过 3D 反卷积操作重构完整几何模型(解码)。作者通过编解码网络学习几何模型潜在的概率分布,可以有效地对缺失部分进行修复。而且分类网络语义信息,可以更精准地修复缺失部分。受制于内存限制,重构几何模型的分辨率低,无法表示更为精细模型。为此,Wang 等人^[110]提出了一种全自动的模型修复网络架构。与 Dai 等人相同,Wang 等人也采用自编码网络架构来重构几何模型。为了保证重构几何模型在功能和结构上的准确性,Wang 等人利用判别器网络来约束自编码网络生成的几何模型。为增加输出模型的分辨率,Wang 等人借助循环卷积网络自动重构出更高分辨

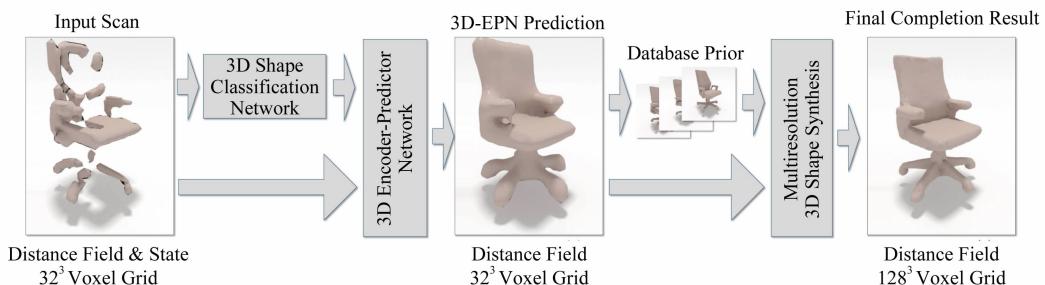


Fig. 39 Shape restoration and reconstruction based on AutoEncoder^[109]

图 39 基于自编码网络的模型修复与重建^[109]

率的几何模型。虽然这些方法能够有效地完成模型的修复与重建,但由于这些方法主要关注几何模型全局结构,无法有效地修复局部几何细节。

为了解决上述问题,Han 等人^[111]提出一种基于全局结构指导的局部结构修复网络架构。如图 40 所示,该文首先利用 LSTM 网络架构提取模型的全局结构信息,然后在全局结构指导下逐一对缺失区域局部结构进行修复。这种方法有效地摆脱了内存

对于模型分辨率的限制,可以达到很高的分辨率,完成全局结构修复的同时也能保证局部几何细节的修复。然而对于纹理级的几何细节,这样的分辨率还不足以表示。为了进一步扩大生成体素模型的分辨率,Cao 等人^[112]借助级联网络架构对缺失模型由粗糙到精细的逐层修复。如图 41 所示,该方法能够将生成模型分辨率提高到极高的水平,可以有效处理精细的几何结构。

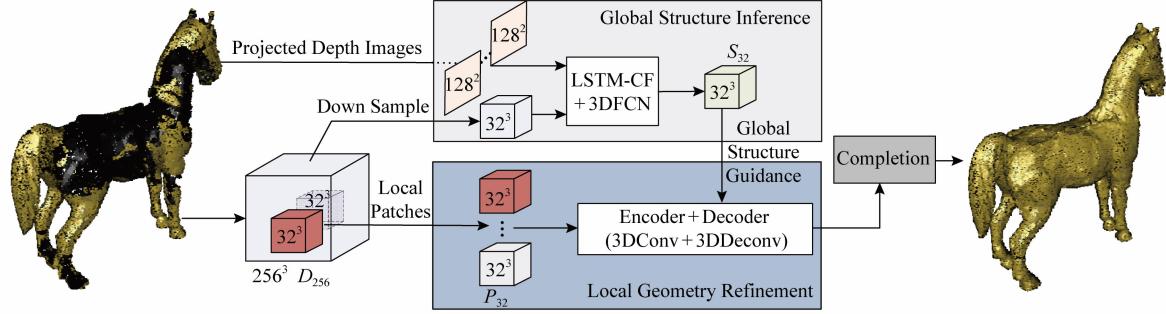


Fig. 40 Global structure guided shape restoration and reconstruction^[111]

图 40 全局结构指导下的模型修复与重建^[111]

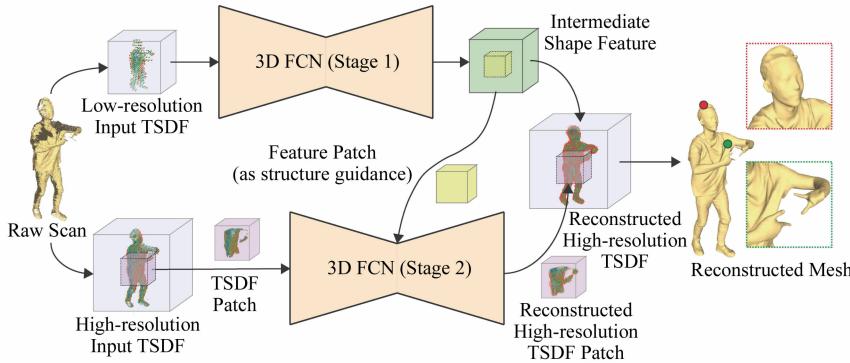


Fig. 41 Shape restoration and reconstruction based on cascaded FCN^[112]

图 41 基于级连网络架构的模型修复与重建^[112]

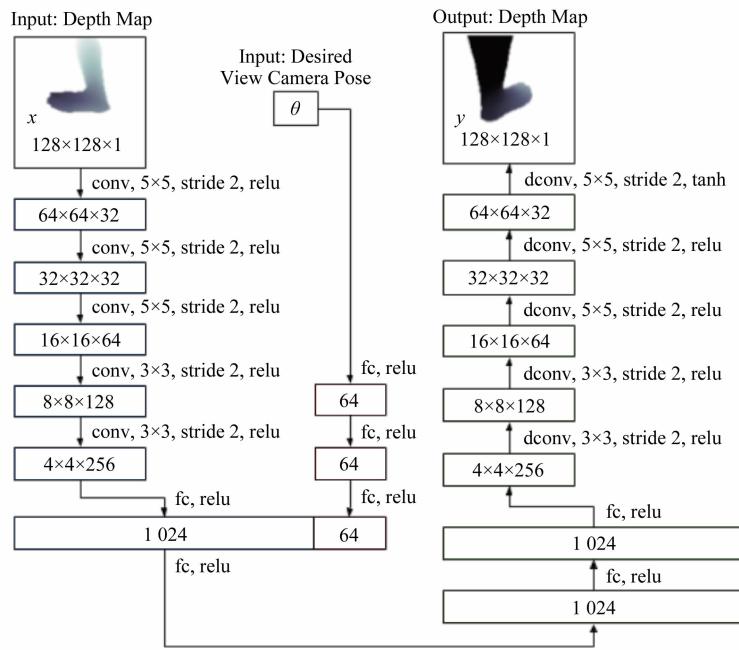
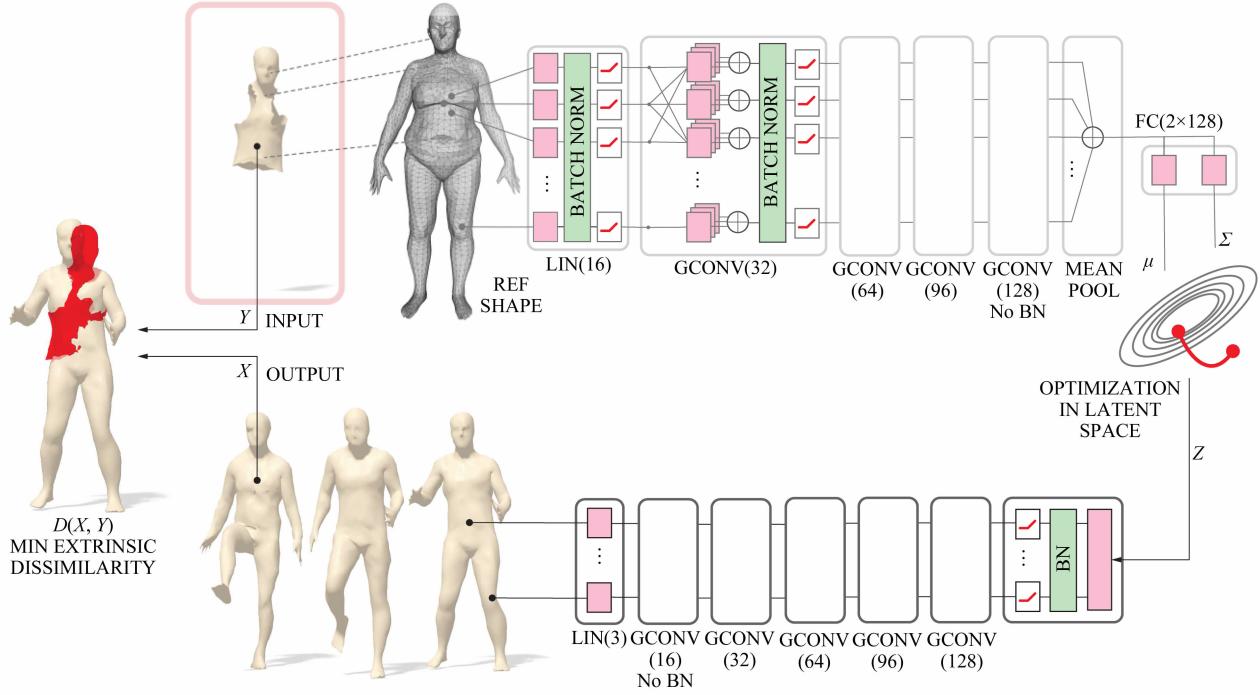
尽管体素有着规则的定义域,也很容易将 2D 的操作直接扩展到 3D 域上,但是体素表示效率较低且受制于内存限制。相比于体素这种中间表示形式,点云则获取直接且存储更为容易。尽管 PointNet^[49]解决了深度学习在点云上的应用,但受限于深度学习输入输出固定长度的约束,这类方法目前还无法更好地服务于破損点云模型的修复与重建。为此,Lunscher 等人^[113-114]利用破損几何模型深度图来完成点云模型修复。

如图 42 所示,其核心思想深度图的修复,然后对不同角度深度图进行融合来完成点云模型的修复与重建。与体素表示相比,该方法不受制于内存限制且数据获取容易,可以有效地应用于服饰定制领域。

另一种常见的模型表示形式就是多边形网格,尽管基于图的卷积神经网络已经被广泛应用,但是受制于网络连接关系等信息,还无法直接处理网格模型的修复。为了解决无法直接重构出缺失区域连接关系信息不同的问题,Google 团队^[115]借助自编码图卷积网络来处理可变形模型的修复与重建。如图 43 所示,首先借助自编码网络架构,学习几何模型潜在的特征空间;然后借助外部约束来重构模型。与一般的模型修复不同,该网络结构对于缺失部分可以给出多种选择,很好地满足用户的不同需求。

3.5 模型变形与编辑

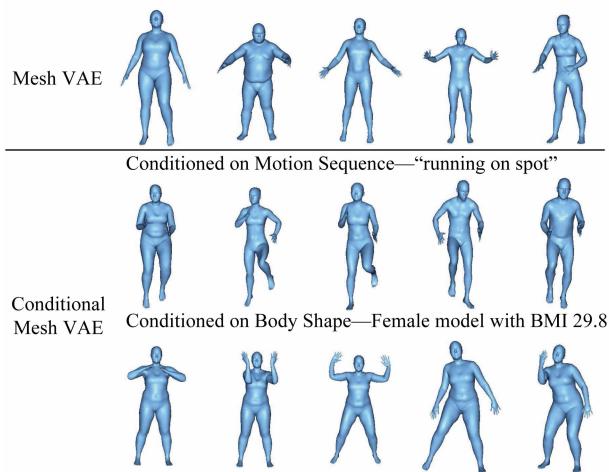
总的来说,构造新的三维模型依旧是一项比较费时费力的工作,而通过对现有模型进行变形和编辑

Fig. 42 Point cloud restoration using geometric depth map^[113]图 42 利用几何模型深度图来完成点云模型修复^[113]Fig. 43 Shape restoration and reconstruction based on Graph CNN^[115]图 43 基于图卷积架构的模型修复与重建^[115]

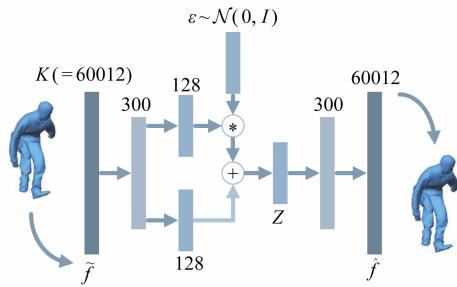
产生新模型的方法显然要高效很多,这也是当前计算机动画生成和制作的主要手段。将深度学习运用到模型变形和编辑会显著提升三维模型数据集构造、三维模型动画制作和模型设计等各项工作的效率。

一类常用的模型编辑方法就是模型变形,图 44 给出了一些使用深度学习技术产生的模型变形结

果。Tan 等人^[116]提出了一种新的网格变分自动编码器(mesh VAE)框架,如图 45 所示,该框架利用深度神经网络来探索三维模型的潜在形状空间,并且能够通过对现有模型的变形生成原始数据集中不存在的新模型。为了有效表示任意的网格变形,该框架使用了变分自动编码器(VAE)^[117]和旋转不变网格

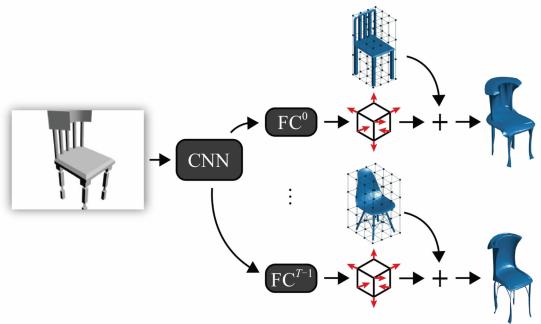
Fig. 44 Shape deformation based on deep learning^[116]图 44 基于深度学习的模型变形^[116]

差分(RIMD)^[118]的曲面表示方法，并使用了全连接的网络架构和简单的基于均方差(MSE)的重构损失函数。该文最后给出证明这项工作在新模型生成、模型插值、模型空间嵌入和模型潜在空间探索等方面都能取得明显优于传统方法所产生的结果。

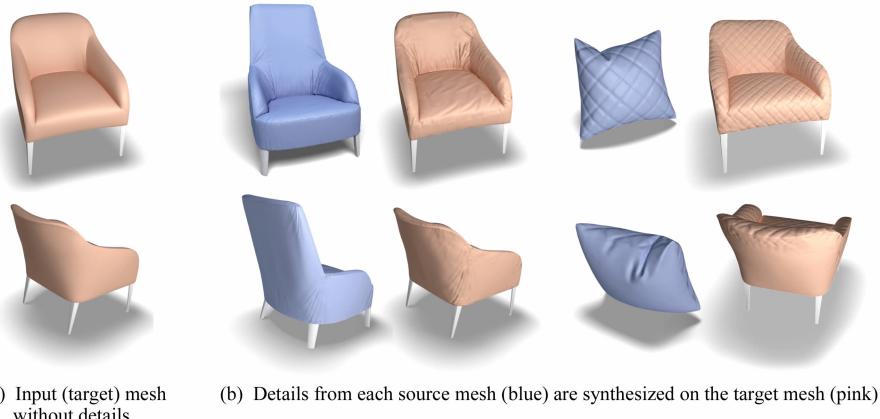
Fig. 45 Shape deformation based on mesh VAE^[116]图 45 基于变分自编码结构的模型变形^[116]

变形传递是模型变形的重要研究课题。建立可靠的模型对应关系是现有变形传递方法的一个基本问题，但解决这个问题需要一定的专业知识，尤其是当源和目标形状差别比较大的时候需要大量的人工操作来选择变形对应的关键点。Yang 等人^[119]提出了一种新的可以自动选择关键点的变形转移方法。该文采用双调和权重形状变形^[120-121]解决变形传递问题，并改进了聚类和误差成本来优化变形传递。这种方法可以在源模型上自动产生一组关键点，用户只需要在目标形状上指定相应的点，这样可以极大地减少变形传递所需要的时间和专业知识。该文给出的实验结果表明，该方法优于目前最先进的变形传递方法，而且这种方式自动选择的关键点比普通用户选择的更合理。

虽然计算机图形学中使用最广泛的图形表示形式是多边形网格，特别是三角形网格。但是这种参数化表示在机器学习领域还有很大的发展空间，因为机器学习主要需要规则表示的输入和输出数据，如体素和点云，而网格是高度非结构化的，这就使通过神经网络从 2D 图形实现相应的 3D 重建变得非常挑战。Jack 等人^[122]利用模型变形的优势，如图 46 所示，通过探索广泛运用于 3D 网格建模的自由变形技术(FFD)^[123]提出了一种新颖的从单张图片重构 3D 网格模型的方法。该方法同时学习对模板网格的选择和变形。该文使用了轻量级的 CNN 推断多个模板的多维 FFD 参数并且学习将大尺度变形应用于拓扑不同的模板，以产生具有相似形状的预测网格。该文通过实验证明小型 CNN 可以很好地学习到网格变形信息，而且这种方法重构出的模型比基于标准体素和点云方法产生的模型具有更加精细的几何结构。

Fig. 46 Shape editing based on free-form deformations learning using CNN^[122]图 46 基于神经网络自由变形学习的模型编辑^[122]

此外，风格迁移也是一种常用的模型编辑方法。图 47 给出了一些使用风格转移进行模型编辑的结果。风格迁移是指将源模型的样式应用于目标模型，同时保留目标模型的几何结构。这一过程的主要挑战是正确区分模型的风格和几何结构。Ma 等人^[124]提出了一种可以在 2D 和 3D 对象之间无缝地迁移各种样式属性的框架。该文使用了几何形状类比的思想。Berkiten 等人^[125]提出了一种基于几何细节的风格迁移方法，该文将单个高质量模型的细节转移到各种易于生成和广泛可用的低分辨率网格。该文的主要创新之处就在于使用度量学习来找出能够最好地预测目标网格和源网格哪些区域具有相似细节的几何特征的组合。然后使用学习到的度量来评估源和目标模型的相似性，指导目标纹理合成。Lun 等人^[126]通过一系列元素级别的操作将示例模型样式

Fig. 47 Style transfer for 3D models^[125]图 47 三维模型的风格迁移^[125]

传递给目标模型。该文引入了跨结构元素兼容性度量,以此来评估每个操作对编辑形状的影响。系统递增地更新目标形状,逐步提升其与示例样式的相似性,同时严格遵循兼用性度量的约束。该文评估了包含家具、灯具和餐具在内的一系列人造物体模型,均能得到符合期望的结果。

4 总结与展望

自从 Guo 等人^[68]首次将深度学习技术引入数字几何处理和分析领域之后,基于神经网络模型的几何处理和分析方法大量涌现,如第 3 节所述,相比于传统的分析和处理方法,利用神经网络模型强大的数据抽象特征提取能力,深度学习技术极大地提高了诸如模型匹配与检索、模型分类与分割、模型生成、模型修复与重建以及模型变形和编辑的性能和效率。但可以看到的是,当前深度学习技术在数字几何处理领域的应用还远不如其在自然语言处理和图像处理中那样成熟,造成这样的原因大致有 3 个方面:

1) 没有统一的数据表示。由于三维模型很难像语音图像那样表示成统一的标准数据表示方式,当前不同的三维模型表示方法各有特点,同时又无法避免地具有各自的缺点,限制了直接迁移语音图像领域成熟应用的各种模型结构到三维模型上的可能性。实际上,当前基于深度学习技术的数字几何分析和处理方法仍然是在特定的数据表示的基础上,针对某些常见的特定任务进行处理,相较于传统的分析处理方法只是从根据几何物理理论的手工定义转变到了采用神经网络进行自动地特征提取和分析,像在传统机器学习领域常见的迁移学习(transfer learning)^[127]、增强学习(reinforcement learning)^[128]等

一些已经在语音和图像相关的深度学习领域得到很好应用的技术,还未能向几何处理和分析领域迁移。

2) 缺乏大规模公开数据集。尽管三维模型的获取已经比较方便,出现了像 ShapeNet^[129]这样具有一定规模的公开数据集,但是其从规模上仍然难以和公开的图像数据集如 ImageNet^[130], COCO^[131], PASCAL^[132]等相提并论,而深度学习技术要真正应用到实际产品中必须有大量的训练数据支持,这就使得当前深度学习技术在数字几何处理领域的应用仍然停留在实验室里,很难像语音图像那样走进普通人的生活当中。另外针对训练数据不足的问题,自然语言处理、计算机视觉领域都有很多人工设定或者自学习的数据增广方式,这不仅可以在一定程度上增加数据规模,同时也可提高方法的泛化能力。目前针对三维模型的数据增广方法还比较缺乏,实际上三维模型的插值和生成已经研究多年,如何从一个有限的模型数据集生成更多的模型已经有了很多的成熟方法,将这些方法引入数据增广或者与深度学习相结合应该可以更好地提高算法的能力。

3) 网络结构缺乏针对性。当前用于几何处理和分析的神经网络主要是从语音和图像领域借鉴而来,都需要先将三维模型转换为特定的网络适用的数据表示形式(特征、图像、体素等),而像 PointNet^[49], GCCN^[63]等专门为三维模型设计的网络还比较少,很难全面高效地对三维模型的特征进行概括分析,限制了深度技术在数字几何处理和分析领域的应用范围。所以尽管深度学习技术在数字几何处理领域中已经有了一定的研究成果,但要真正推广应用,还有很长的路要走。

本文全面总结了近年来深度学习技术在数字几何处理和分析领域的研究进展,希望有助于读者对

这一研究方向的发展和面临的问题有更加全面的了解，并对未来的研究工作有一定启发和推动作用。

参 考 文 献

- [1] Schröder P, Sweldens W. Digital geometry processing [C] // Proc of the 6th Annual Symp on Frontiers of Engineering. Washington, DC: National Academies Press, 2001: 41–44
- [2] Hu Shimin, Yang Yongliang, Lai Yukun. Research progress of digital geometry processing [J]. Chinese Journals of Computers, 2009, 32(8):1451–1469 (in Chinese)
(胡事民, 杨永亮, 来煜坤. 数字几何处理研究进展[J]. 计算机学报, 2009, 32(8):1451–1469)
- [3] Zhao Qinping. A survey on virtual reality [J]. Science in China Series F: Information Sciences, 2009, 39(1): 2–46 (in Chinese)
(赵沁平. 虚拟现实综述[J]. 中国科学: 信息科学, 2009, 39(1): 2–46)
- [4] McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity [J]. The Bulletin of Mathematical Biophysics, 1943, 5(4): 115–133
- [5] Werbos P. Beyond regression: New tools for prediction and analysis in the behavioral science [D]. Cambridge, MA: Harvard University, 1974
- [6] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C] // Proc of the 25th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2012: 1097–1105
- [7] Alex B, Jia D, Li F F. Large scale visual recognition challenge 2010 [EB/OL]. [2010-09-16]. <http://www.image-net.org/challenges/LSVRC/2010/index>
- [8] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014
- [9] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 1–9
- [10] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770–778
- [11] LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series [J]. The Handbook of Brain Theory and Neural Networks, 1995, 3361(10): 1995: 255–258
- [12] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 580–587
- [13] Girshick R. Fast R-CNN [C] //Proc of the IEEE Conf on Computer Vision. Piscataway, NJ: IEEE, 2015: 1440–1448
- [14] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [C] //Proc of the 28th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2015: 91–99
- [15] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 3431–3440
- [16] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation [C] //Proc of the 18th Int Conf on Medical Image Computing and Computer-assisted Intervention. Berlin: Springer, 2015: 234–241
- [17] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation [J]. arXiv preprint arXiv:1511.00561, 2015
- [18] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFS [J]. arXiv preprint arXiv:1412.7062, 2014
- [19] Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834–848
- [20] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation [J]. arXiv preprint arXiv:1706.05587, 2017
- [21] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 2881–2890
- [22] Newell A, Huang Z, Deng J. Associative embedding: End-to-end learning for joint detection and grouping [C] //Proc of the 30th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2017: 2277–2287
- [23] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [C] //Proc of the 27th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2014: 2672–2680
- [24] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 4681–4690
- [25] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 2414–2423
- [26] Luc P, Couprie C, Chintala S, et al. Semantic segmentation using adversarial networks [J]. arXiv preprint arXiv:1611.08408, 2016
- [27] Mikolov T, Karafiat M, Burget L, et al. Recurrent neural network based language model [C] //Proc of the 11th Annual Conf of the International Speech Communication Association. Grenoble, France: International Speech Communication Association, 2010: 1045–1048

- [28] Liu S, Yang N, Li M, et al. A recursive recurrent neural network for statistical machine translation [C] //Proc of the 52nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2014: 1491–1500
- [29] Graves A, Jaitly N. Towards end-to-end speech recognition with recurrent neural networks [C/OL] //Proc of the 31st Int Conf on Machine Learning. JMLR.org, 2014: 1764–1772. [2018-10-22]. <https://dl.acm.org/citation.cfm?id=3045089>
- [30] Graves A, Gomez F. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks [C] //Proc of the 23rd Int Conf on Machine Learning. New York: ACM, 2006: 369–376
- [31] Wang H, Cai Y, Chen L. A vehicle detection algorithm based on deep belief network [J]. The Scientific World Journal, 2014, 2014: 647380
- [32] Rumelhart D E, Hinton G E, Williams R J. Learning internal representations by error propagation, ICS-8506 [R/OL]. California Univ San Diego La Jolla Inst for Cognitive Science, 1985 [2018-10-22]. <https://dl.acm.org/citation.cfm?id=104293>
- [33] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504–507
- [34] Salakhutdinov R, Hinton G. Semantic hashing [J]. International Journal of Approximate Reasoning, 2009, 50(7): 969–978
- [35] Tieleman T. Training restricted Boltzmann machines using approximations to the likelihood gradient [C] //Proc of the 25th Int Conf on Machine Learning. New York: ACM, 2008: 1064–1071
- [36] Weiss Y, Torralba A, Fergus R. Spectral hashing [C] //Proc of the 22nd Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2009: 1753–1760
- [37] Wu Z, Song S, Khosla A, et al. 3D ShapeNets: A deep representation for volumetric shapes [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 1912–1920
- [38] Maturana D, Scherer S. VoxNet: A 3D convolutional neural network for real-time object recognition [C] //Proc of IEEE/RSJ Int Conf on Intelligent Robots and Systems. Piscataway, NJ: IEEE, 2015: 922–928
- [39] Qi C R, Su H, Nießner M, et al. Volumetric and multi-view CNNs for object classification on 3D data [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 5648–5656
- [40] Steinbrücker F, Sturm J, Cremers D. Volumetric 3D mapping in real-time on a CPU [C] //Proc of Int Conf on Robotics and Automation. Piscataway, NJ: IEEE, 2014: 2021–2028
- [41] Riegler G, Ulusoy A O, Geiger A. OctNet: Learning deep 3D representations at high resolutions [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017, 3577–3586
- [42] Wang P S, Liu Y, Guo Y X, et al. O-CNN: Octree-based convolutional neural networks for 3D shape analysis [J]. ACM Transactions on Graphics, 2017, 36(4): 72
- [43] Tatarchenko M, Dosovitskiy A, Brox T. Octree generating networks: Efficient convolutional architectures for high-resolution 3D outputs [C] //Proc of the IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 2088–2096
- [44] Häne C, Tulsiani S, Malik J. Hierarchical surface prediction for 3D object reconstruction [C] //Proc of the 5th Int Conf on 3D Vision. Piscataway, NJ: IEEE, 2017: 412–420
- [45] Guo K, Chen X, Zhou B, et al. Image-guided 3D model labeling via multiview alignment [J]. Graphical Models, 2018, 96: 30–37
- [46] Su H, Maji S, Kalogerakis E, et al. Multi-view convolutional neural networks for 3D shape recognition [C] //Proc of the IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2015: 945–953
- [47] Huang H, Kalogerakis E, Chaudhuri S, et al. Learning local shape descriptors from part correspondences with multiview convolutional networks [J]. ACM Transactions on Graphics, 2018, 37(1): 6
- [48] Kalogerakis E, Averkiou M, Maji S, et al. 3D shape segmentation with projective convolutional networks [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 3779–3788
- [49] Qi C R, Su H, Mo K, et al. PointNet: Deep learning on point sets for 3D classification and segmentation [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 652–660
- [50] Qi C R, Yi L, Su H, et al. PointNet++: Deep hierarchical feature learning on point sets in a metric space [C] //Proc of the 31st Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2017: 5099–5108
- [51] Gori M, Monfardini G, Scarselli F. A new model for learning in graph domains [C] //Proc of IEEE Int Joint Conf on Neural Networks. Piscataway, NJ: IEEE, 2005, 2: 729–734
- [52] Scarselli F, Gori M, Tsotsos J, et al. The graph neural network model [J]. IEEE Transactions on Neural Networks, 2009, 20(1): 61–80
- [53] Bruna J, Zaremba W, Szlam A, et al. Spectral networks and locally connected networks on graphs [J]. arXiv preprint arXiv:1312.6203, 2013
- [54] Qi X, Liao R, Jia J, et al. 3D graph neural networks for RGB-D semantic segmentation [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 5199–5208
- [55] Yi L, Su H, Guo X, et al. SyncSpecCNN: Synchronized spectral CNN for 3D shape segmentation [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 6584–6592
- [56] Te G, Hu W, Guo Z, et al. RGCNN: Regularized graph CNN for point cloud segmentation [J]. arXiv preprint arXiv: 1806.02952, 2018

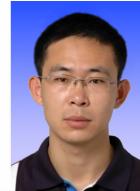
- [57] Wang Y, Sun Y, Liu Z, et al. Dynamic graph CNN for learning on point clouds [J]. arXiv preprint arXiv: 1801.07829, 2018
- [58] Henaff M, Bruna J, LeCun Y. Deep convolutional networks on graph-structured data [J]. arXiv preprint arXiv: 1506.05163, 2015
- [59] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering [C] //Proc of the 30th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2016: 3844–3852
- [60] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks [J]. arXiv preprint arXiv: 1609.02907, 2016
- [61] Atwood J, Towsley D. Diffusion-convolutional neural networks [C] //Proc of the 30th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2016: 1993–2001
- [62] Duvenaud D K, Maclaurin D, Iparraguirre J, et al. Convolutional networks on graphs for learning molecular fingerprints [C] //Proc of the 29th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2015: 2224–2232
- [63] Masci J, Boscaini D, Bronstein M M, et al. Geodesic convolutional neural networks on riemannian manifolds [C] // Proc of the IEEE Conf on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE, 2015: 37–45
- [64] Masci J, Boscaini D, Bronstein M, et al. ShapeNet: Convolutional neural networks on non-euclidean manifolds [R]. CoRR abs/1501.06297, 2015
- [65] Boscaini D, Masci J, Rodolà E, et al. Learning shape correspondence with anisotropic convolutional neural networks [C] //Proc of the 30th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2016: 3189–3197
- [66] Boscaini D, Masci J, Melzi S, et al. Learning class-specific descriptors for deformable shapes using localized spectral convolutional networks [J]. Computer Graphics Forum, 2015, 34(5): 13–23
- [67] Monti F, Boscaini D, Masci J, et al. Geometric deep learning on graphs and manifolds using mixture model CNNs [C] // Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 5425–5434
- [68] Guo K, Zou D, Chen X. 3D mesh labeling via deep convolutional neural networks [J]. ACM Transactions on Graphics, 2015, 35(1): 3
- [69] Chen M, Wang C, Qin H. Jointly learning shape descriptors and their correspondence via deep triplet CNNs [J]. Computer Aided Geometric Design, 2018, 62: 192–205
- [70] Wang H, Guo J, Dongming Y, et al. Learning 3D keypoint descriptors for non-rigid shape matching [C] //Proc of the European Conf on Computer Vision. Berlin: Springer, 2018: 3–19
- [71] Yang Shen, Qi Yue, Shen Yukun, et al. Rapid and automatic method for 3D scanned data registration [J]. Journal of Software, 2010, 21(6): 1438–1450
(杨梦, 齐越, 沈旭昆, 等. 一种快速的三维扫描数据自动配准方法[J]. 软件学报, 2010, 21(6): 1438–1450)
- [72] Wan Lili, Zhao Qiping, Hao Aimin. A method of 3D model retrieval by the spatial distributions of components [J]. Journal of Software, 2007, 18(11): 2902–2913
(万丽莉, 赵沁平, 郝爱民. 一种基于部件空间分布的三维模型检索方法[J]. 软件学报, 2007, 18(11): 2902–2913)
- [73] Wei L, Huang Q, Ceylan D, et al. Dense human body correspondences using convolutional networks [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 1544–1553
- [74] Zeng A, Song S, Nießner M, et al. 3Dmatch: Learning local geometric descriptors from RGB-D reconstructions [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 199–208
- [75] Bronstein M M, Bruna J, LeCun Y, et al. Geometric deep learning: Going beyond euclidean data [J]. IEEE Signal Processing Magazine, 2017, 34(4): 18–42
- [76] Litany O, Remez T, Rodolà E, et al. Deep functional maps: Structured prediction for dense shape correspondence [C] // Proc of the IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 5660–5668
- [77] Furuya T, Ohbuchi R. Learning part-in-whole relation of 3D shapes for part-based 3D model retrieval [J]. Computer Vision and Image Understanding, 2018, 166: 102–114
- [78] Zhu Z, Wang X, Bai S, et al. Deep learning representation using autoencoder for 3D shape retrieval [J]. Neurocomputing, 2016, 204: 41–50
- [79] Xie J, Dai G, Zhu F, et al. Deep nonlinear metric learning for 3-D shape retrieval [J]. IEEE Transactions on Cybernetics, 2018, 48(1): 412–422
- [80] Dai G, Xie J, Zhu F, et al. Deep correlated metric learning for sketch-based 3D shape retrieval [C] //Proc of the 31st AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2017: 4002–4008
- [81] Wang F, Kang L, Li Y. Sketch-based 3D shape retrieval using convolutional neural networks [C] //Proc of the IEEE Con on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 1875–1883
- [82] Song Y, Chen X, Li J, et al. Embedding 3D geometric features for rigid object part segmentation [C] //Proc of the IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 580–588
- [83] Wang X, Zhou B, Wang Z, et al. Efficiently consistent affinity propagation for 3D shapes co-segmentation [J]. The Visual Computer, 2018, 34(6/8): 997–1008
- [84] Socher R, Huval B, Bath B, et al. Convolutional-recursive deep learning for 3D object classification [C] //Proc of the 25th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2012: 656–664

- [85] Gupta S, Girshick R, Arbeláez P, et al. Learning rich features from RGB-D images for object detection and segmentation [C] //Proc of the European Conf on Computer Vision. Berlin: Springer, 2014: 345–360
- [86] Eitel A, Springenberg J T, Spinello L, et al. Multimodal deep learning for robust RGB-D object recognition [C] //Proc of IEEE/RSJ Int Conf on Intelligent Robots and Systems. Piscataway, NJ: IEEE, 2015: 681–687
- [87] Wu Z, Song S, Khosla A, et al. 3D ShapeNets: A deep representation for volumetric shapes [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 1912–1920
- [88] Maturana D, Scherer S. VoxNet: A 3D convolutional neural network for real-time object recognition [C] //Proc of IEEE/RSJ Int Conf on Intelligent Robots and Systems. Piscataway, NJ: IEEE, 2015: 922–928
- [89] Shi B, Bai S, Zhou Z, et al. Deeppano: Deep panoramic representation for 3-D shape recognition [J]. IEEE Signal Processing Letters, 2015, 22(12): 2339–2343
- [90] Hu R, Yan Z, Zhang J, et al. Predictive and generative neural networks for object functionality [J]. Computer Graphics Forum, 2018, 37(4): 603–624
- [91] Shu Z, Qi C, Xin S, et al. Unsupervised 3D shape segmentation and co-segmentation via deep learning [J]. Computer Aided Geometric Design, 2016, 43: 39–52
- [92] Wang P, Gan Y, Shui P, et al. 3D shape segmentation via shape fully convolutional networks [J]. Computers & Graphics, 2018, 70: 128–139
- [93] Liu T, Hertzmann A, Li W, et al. Style compatibility for 3D furniture models [J]. ACM Transactions on Graphics, 2015, 34(4): 85
- [94] Lim I, Gehre A, Kobbelt L. Identifying style of 3D shapes using deep metric learning [J]. Computer Graphics Forum, 2016, 35(5): 207–215
- [95] Hinton G E, Krizhevsky A, Wang S D. Transforming auto-encoders [C] //Proc of the 21st Int Conf on Artificial Neural Networks. Berlin: Springer, 2011: 44–51
- [96] Flynn J, Neulander I, Philbin J, et al. Deepstereo: Learning to predict new views from the world's imagery [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 5515–5524
- [97] Dosovitskiy A, Springenberg J T, Tatarchenko M, et al. Learning to generate chairs, tables and cars with convolutional networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 692–705
- [98] Jaderberg M, Simonyan K, Zisserman A, et al. Spatial transformer networks [C] //Proc of the 28th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2015: 2017–2025
- [99] Zhou T, Tulsiani S, Sun W, et al. View synthesis by appearance flow [C] //Proc of the European Conf on Computer Vision. Berlin: Springer, 2016: 286–301
- [100] Park E, Yang Jimei, Yumer E, et al. Transformation-grounded image generation network for novel 3D view synthesis [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 702–711
- [101] Girdhar R, Fouhey D F, Rodriguez M, et al. Learning a predictable and generative vector representation for objects [C] //Proc of the European Conf on Computer Vision. Berlin: Springer, 2016: 484–499
- [102] Wu J, Zhang C, Xue T, et al. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling [C] //Proc of the 29th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2016: 82–90
- [103] Li J, Xu K, Chaudhuri S, et al. Grass: Generative recursive autoencoders for shape structures [J]. ACM Transactions on Graphics, 2017, 36(4): 52
- [104] Wang N, Zhang Y, Li Z, et al. Pixel2Mesh: Generating 3D mesh models from single RGB images [C] //Proc of the European Conf on Computer Vision. Berlin: Springer, 2018
- [105] Huang H, Kalogerakis E, Yumer E, et al. Shape synthesis from sketches via procedural models and convolutional networks [J]. IEEE Transactions on Visualization and Computer Graphics, 2017, 23(8): 2003–2013
- [106] Ju T. Fixing geometric errors on polygonal models: A survey [J]. Journal of Computer Science and Technology, 2009, 24(1): 19–29
- [107] Lu F, Zhou B, Zhang Y, et al. Real-time 3D scene reconstruction with dynamically moving object using a single depth camera [J]. The Visual Computer, 2018, 34(6/8): 753–763
- [108] Lu F, Zhou B, Lu F, et al. Reconstructing non-rigid object with large movement using a single depth camera [J]. Computer Aided Geometric Design, 2018, 64: 15–26
- [109] Dai A, Qi C R, Nießner M. Shape completion using 3D-encoder-predictor CNNs and shape synthesis [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 5868–5877
- [110] Wang W, Huang Q, You S, et al. Shape inpainting using 3D generative adversarial network and recurrent convolutional networks [J]. arXiv preprint arXiv: 1711.06375, 2017
- [111] Han X, Li Z, Huang H, et al. High-resolution shape completion using deep neural networks for global structure and local geometry inference [C] //Proc of the IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 85–93
- [112] Cao Y P, Liu Z N, Kuang Z F, et al. Learning to reconstruct high-quality 3D shapes with cascaded fully convolutional networks [C] //Proc of the European Conf on Computer Vision. Berlin: Springer, 2018: 616–633

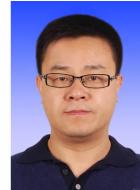
- [113] Zelek J, Lunscher N. Point cloud completion of foot shape from a single depth map for fit matching using deep learning view synthesis [C] //Proc of the IEEE Int Conf on Computer Vision Workshops. Piscataway, NJ: IEEE, 2017: 2300–2305
- [114] Lunscher N, Zelek J S. Deep learning anthropomorphic 3D point clouds from a single depth map camera viewpoint [C] //Proc of the IEEE Int Conf on Computer Vision Workshops. Piscataway, NJ: IEEE, 2017: 689–696
- [115] Litany O, Bronstein A, Bronstein M, et al. Deformable shape completion with graph convolutional autoencoders [J]. arXiv preprint arXiv:1712.00268, 2017
- [116] Tan Q, Gao L, Lai Y K, et al. Variational autoencoders for deforming 3D mesh models [C] //Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 5841–5850
- [117] Kingma D P, Welling M. Auto-encoding variational Bayes [J]. arXiv preprint arXiv:1312.6114, 2013
- [118] Gao L, Lai Y K, Liang D, et al. Efficient and flexible deformation representation for data-driven surface modeling [J]. ACM Transaction on Graphics, 2016, 35(5): 158:1–158:17
- [119] Yang J, Gao L, Lai Y K, et al. Biharmonic deformation transfer with automatic key point selection [J]. Graphical Models, 2018, 98: 1–13
- [120] Jacobson A, Baran I, Popovic J, et al. Bounded biharmonic weights for real-time deformation [J]. ACM Transaction on Graphics, 2011, 30(4): 78:1–78:8
- [121] Jacobson A, Baran I, Kavan L, et al. Fast automatic skinning transformations [J]. ACM Transactions on Graphics, 2012, 31(4): 77
- [122] Jack D, Pontes J K, Sridharan S, et al. Learning free-form deformations for 3D object reconstruction [J]. arXiv preprint arXiv:1803.10932, 2018
- [123] Sederberg T W, Parry S R. Free-form deformation of solid geometric models [J]. ACM SIGGRAPH Computer Graphics, 1986, 20(4): 151–160
- [124] Ma C, Huang H, Sheffer A, et al. Analogy-driven 3D style transfer [J]. Computer Graphics Forum, 2014, 33(2): 175–184
- [125] Berkiten S, Halber M, Solomon J, et al. Learning detail transfer based on geometric features [J]. Computer Graphics Forum, 2017, 36(2): 361–373
- [126] Lun Z, Kalogerakis E, Wang R, et al. Functionality preserving shape style transfer [J]. ACM Transactions on Graphics, 2016, 35(6): 209
- [127] Pan S J, Yang Q. A survey on transfer learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345–1359
- [128] Sutton R S, Barto A G, Bach F. Reinforcement Learning: An Introduction [M]. Cambridge, MA: MIT Press, 1998
- [129] Chang A X, Funkhouser T, Guibas L, et al. ShapeNet: An information-rich 3D model repository [J]. arXiv preprint arXiv:1512.03012, 2015
- [130] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge [J]. International Journal of Computer Vision, 2015, 115(3): 211–252
- [131] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common objects in context [C] //Proc of the European Conf on Computer Vision. Berlin: Springer, 2014: 740–755
- [132] Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes (VOC) challenge [J]. International Journal of Computer Vision, 2010, 88(2): 303–338



Xia Qing, born in 1989. PhD. His main research interests include computer graphics, digital geometry processing, virtual surgery simulation, medical image analysis and machine learning.



Li Shuai, born in 1981. PhD, associate professor, PhD supervisor. His main research interests include computer graphics, pattern recognition, computer vision, and medical image processing.



Hao Aimin, born in 1968. PhD, professor, PhD supervisor, associate director of the State Key Laboratory of Virtual Reality Technology and Systems. His main research interests include virtual reality, computer simulation, computer graphics, geometric modeling, image processing and computer vision.



Zhao Qinping, born in 1948. PhD, professor, PhD supervisor. Member of the China Academy of Engineering, chief scientist of the State Key Laboratory of Virtual Reality Technology and System, president of China Simulation Federation. His main research interests include virtual reality, computer simulation, and computer vision.