ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 36,No. 1,Jan. 2021,pp. 1－21
DOI:10. 16337/j. 1004-9037. 2021. 01. 001

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于深度学习的三维模型检索算法综述

刘安安，李天宝，王晓雯，宋　丹

(天津大学电气自动化与信息工程学院,天津300072)

摘　要:近年来,深度学习被广泛应用于各个领域并取得了显著的进展,如何利用深度学习高效管理呈爆炸式增长的三维模型一直是一个研究热点。本文介绍了发展至今主流的基于深度学习的三维模型检索算法,并根据实验得出的算法性能评估分析了其优缺点。根据检索任务的不同,可将主要的三维模型检索算法分为两类:(1)基于模型的三维模型检索方法,即检索对象与被检索对象都是三维模型,按照对三维模型的表示方式不同,可进一步分为基于体素、基于点云和基于视图的方法;(2)基于二维图像的跨域三维模型检索方法,即检索对象是二维图像,被检索对象是三维模型,包括基于二维真实图像和基于二维草图的三维模型检索方法。最后,对基于深度学习的三维模型检索算法目前存在的问题进行分析和讨论,并展望未来发展的新方向。

关键词:三维模型检索;深度学习;特征表示;度量学习;域适应

中图分类号:TP391　　**文献标志码**:A

Review of 3D Model Retrieval Algorithms Based on Deep Learning

LIU Anan, LI Tianbao, WANG Xiaowen, SONG Dan

(School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: In recent years, deep learning has been widely used and achieved significant development in various fields. How to utilize deep learning to effectively manage the explosive increasing 3D models becomes a hot topic. This paper introduces the mainstream algorithms for deep learning based 3D model retrieval and analyzes the advantages and disadvantages according to the experimental performance. In terms of the retrieval tasks, 3D model retrieval algorithms are classified into two categories: (1) Model-based 3D model retrieval algorithms require that both query and gallery are 3D models. It can be further divided into voxel-based method, point cloud-based method and view-based method in regard of different representations of 3D models. (2) For 2D image-based cross-domain 3D model retrieval algorithms, the query is 2D image while the gallery is 3D model. It can be classified to 2D real image-based method and 2D sketch-based method. Finally, we analyze and discuss existing issues of deep learning based 3D model retrieval methods, and predict possible promising directions for this research topic.

Key words: 3D model retrieval; deep learning; feature representation; metric learning; domain adaptation

---

基金项目:国家自然科学基金(61772359, 61902277)资助项目;天津市新一代人工智能重大专项(19ZXZNGX00110, 18ZXZNGX00150)资助项目;中国博士后科学基金(2020M680884)资助项目。
收稿日期:2020-12-14;**修订日期**:2021-01-04

引 言

随着三维建模技术和计算机图形学的发展,三维模型广泛应用于CAD、VR/AR和自动驾驶等领域。与此同时,3D重建和3D打印等技术的不断革新,使得三维模型的生成过程变得更加容易。由于三维模型的数量剧增,对模型数据库的管理和对模型的检索匹配工作已经无法由人工完成。为了减少重复性工作,三维模型的各个领域都出现了各自的专业数据库和对应的检索技术,如何对三维模型进行高效准确的检索和管理得到了越来越多研究者的关注[1]。

虽然基于文本的搜索引擎非常常见,但是对于三维模型之类的高维数据,首先很难用文本信息概括其所包含的大量信息,其次来自于各个领域工程师的文本信息由于语言文化和行业习惯的差异,容易产生误差和混乱,仅仅利用文本或者标签对三维模型进行检索很难达到较高的准确性,因此需要基于模型的三维模型检索技术对三维模型进行精准的检索和匹配[1],即根据三维模型本身的信息来对模型进行匹配。基于模型的三维模型检索方法,在处理三维模型数据时需要先对模型进行特征提取,将三维模型转换为各种形式的三维模型特征描述子,根据三维模型描述子表示形式的不同,可以进一步将三维模型检索方法分为3类:基于体素的方法、基于点云的方法和基于视图的方法。基于体素的方法[2-3]和基于点云的方法[4-5]都是直接在原始的三维模型上,通过构建相应的特征提取网络提取三维模型的高阶全局特征。基于视图的方法[6-9]利用虚拟相机对三维模型进行渲染,利用得到的一组二维视图来表示原始的三维模型,由于深度学习在二维图像处理上的技术非常成熟,基于视图的三维模型算法也取得了很好的效果。

尽管基于模型的三维模型检索方法性能良好,完整三维模型实例的获取和标注仍然是一个尚未完善的工作。随着深度神经网络以及大量二维图片数据集的提出,二维图像的分类和识别技术已经非常成熟,基于二维图像的跨域三维模型检索算法利用域适应等技术在大量有标记的二维图像上学习得到相关知识并协助处理三维模型也是近年来的一个热点问题[10-12]。基于二维图像的三维模型检索方法,即检索对象是二维图像、被检索对象是三维模型,其二维图像和三维模型之间存在的模态差异给三维模型检索带来了很大的挑战。根据二维图像的表现方式不同,可以将算法分为2类:基于二维真实图像的方法和基于二维草图的方法。基于二维真实图像的方法[11]即在真实的图片上提取特征,图像来源通常是拍摄的照片。基于二维草图的方法[12]即在已知的草图上提取特征。和基于模型的三维模型检索方法相比,基于二维图像的三维模型检索方法在特征匹配上面临着更大的挑战,但在检索对象上表现为更易于获取,在应用中更具有普遍性。

本文首先全面介绍了近年来的三维模型检索算法,然后根据三维模型检索任务的不同,对不同的三维模型检索算法进行分类,并对各个类别下的算法进行比较和分析。其中,本文针对基于模型的三维模型检索算法着重分析了三维模型在体素、点云和视图等不同表示形式下的优缺点。针对基于二维真实图像的跨域三维模型检索算法,本文限定了三维模型不含标签的场景,重点分析了无监督迁移学习在无监督跨域三维模型检索中的性能。针对基于草图的三维模型检索算法,本文着重关注了二维草图和三维模型在巨大模态差异下的跨域检索和匹配问题。最后,本文对三维模型检索算法目前存在的问题进行分析和总结,并对该领域新的发展方向进行了探讨和展望,为该领域下的后续研究提供了新的思路。

1 基于深度学习的三维模型检索算法分类

三维模型检索是计算机视觉领域最为热门的研究方向之一,科研人员围绕它开展了许多研究工作。根据三维模型检索任务的不同,本文将基于深度学习的三维模型检索算法分为基于模型的三维模型检索算法和基于二维图像的三维模型检索算法。

1.1 基于模型的三维模型检索方法

近年来,随着深度神经网络的快速发展和大规模三维模型数据集的出现,利用深度学习方法对三维模型进行学习和表示成为了三维模型研究领域的热门研究方法。三维模型的表现形式包括体素、点云和视图等,因此现阶段根据对三维模型的表示方式不同,相关研究大致可分为基于体素、基于点云、基于视图和混合的三维模型检索方法。

基于体素的三维模型检索方法通常使用密集而规整的三维网格表示三维模型,对其进行三维卷积和池化等操作,从而提取高阶特征用于学习表示。Qi等[13]提出了多方位体素卷积神经网络(Multi-orientation volumetric convolutional neural network, MO-VCNN),该网络使用权值共享的3D卷积神经网络获取不同方向的三维体素特征,然后执行最大池化操作聚合特征,最终将聚合之后的特征输入到另一个三维卷积神经网络进行预测,取得了不错的效果。为了降低计算成本、提高效率,Li等[14]提出使用场探测滤波器,通过学习改变探测滤波器的形状,将其自适应地分布在三维空间中。

基于点云的三维模型检索方法指的是从三维模型表面进行三维坐标采样构成三维点集用于学习表示。三维点云处理的挑战主要源于点集的非结构化和不规则性。Klokov等[15]提出了k-D网络,首先用k-维树(k-dimensional trees, k-D trees)表示点云,然后根据数据结构对点云进行处理。受形状匹配算法[16]的启发,Xie等[17]使用自注意力模块来为点云的形状进行上下文建模。

基于视图的三维模型检索方法往往先为每个三维模型渲染一组二维视图图像,再利用二维卷积神经网络提取相关特征。Bai等[18]对相应视图集之间的豪斯多夫距离进行分析,提出了基于视图的三维模型检索系统GIFT。Su等[6]提出了多视图卷积神经网络(Multi-view CNN, MVCNN),最先使用一个共享的卷积神经网络对视图图像进行特征提取,然后跨视图维度进行最大池化来聚合特征。然而,由于该网络在最大池操作中丢弃了非最大值,可能无法充分利用多视图特性。针对这个缺陷,Wang等[19]提出了一种基于优势集合思想的视图聚类和池化算法,在聚类后的多个视图集内进行池化操作,获得了不错的实验效果。还有学者尝试使用长短期记忆网络(Long short-term memory, LSTM)来聚合多视角图像的特征[20-21]。此外,Chen等[22]提出使用循环注意力模型自动选择视图序列进行准确的三维形状识别。Sfikas等[23]提出全景卷积神经网络(PANORAMA-based CNN, PANORAMA-CNN),即使用一组基于全景的卷积神经网络来对三维形状进行学习表示。

混合方法通常指对两种或两种以上的三维模型的表示进行处理。Lu等[24]尝试分析三维模型基于模型特征和视图特征之间的相关性。You等[25]引入点-视图关系网络(Point-view relation network, PVR-Net)融合点云特征和多视角特征,取得了很好的效果。文献[26]中使用FusionNet对三维模型基于体素特征和视图特征进行融合从而对三维模型进行识别和检索。Rahman等[27]提出将三维模型的深度图和彩色图进行融合,来对三维模型进行更好的表示。

1.2 基于图像的三维模型检索方法

除了基于模型的三维模型检索方法外,基于图像进行三维模型检索也是一个可选择的研究思路。基于二维图像的三维模型检索方法,即检索对象是二维图像、被检索对象是三维模型,二维图像和三维模型之间存在的模态差异给三维模型检索带了很大的挑战。根据二维图像表现方式的不同,可以将算法分为两大类:基于二维真实图像的方法和基于二维草图的方法。

基于二维真实图像的三维模型检索方法中,无监督跨域三维模型检索方法吸引了众多研究者的关注,即在源域的标签已知、而目标域的标签未知的情况下设计算法对三维模型进行检索和识别。Zhou等[10]提出两级嵌入对齐网络(Dual-level embedding alignment network, DLEA),采用对抗性域适应算法和类别中心对齐来联合约束模型的训练。Li等[11]提出多视图多分布学习方法(Multi-view multi-dis-

tribution learning, MVML),通过学习两个耦合子空间将源域特征和目标域特征映射到公共域进行学习。域对抗神经网络(Domain-adversarial neural network, DANN)[28]采用对抗性域适应算法,对源域数据和目标域数据进行对齐。

基于草图的三维模型检索方法的难点在于解决二维草图与三维模型之间存在的巨大模态差异问题。Zhu等[29]提出使用跨域神经网络(Cross-domain neural network,CDNN)来缩小二维草图与三维模型的差异。Daras等[30]提出了一种支持多媒体查询的三维模型检索系统,该系统将三维模型投影到一组二维图像中,通过从二维图像中提取特征来确定不同模型之间的相似度。Bronstein等[31]将在二维计算机视觉中流行的特征袋(Bag-of-Feature, BoF)方法应用于三维模型检索中。更进一步,Eitz等[32]将BoF与基于Gabor的局部线性特征(Gabor local line based feature, GALIF)结合,用于基于草图的三维形状检索研究。除了BoF编码算法外,局部约束线性编码(Locality-constrained linear coding, LLC)[33]是另一种在图像分类中广泛应用的编码方法,它可以保留图像的局部特性。Biasotti等[34]将LLC方法应用于三维模型检索。此外,Xie等[35]从三维模型的不同二维投影中聚合有效特征。Tasse等[36]提出了一种新的跨域检索方法,将不同模态的样本嵌入语义特征向量中进行特征学习。

2 基于模型的三维模型检索方法

按照三维模型的表示形式,可以将三维模型处理算法分为基于体素的方法、基于点云形式的方法和基于视图的方法。本节详细对比分析了几种代表性的方法:(1)基于体素的三维模型检索算法包括3DshapeNets[3]和深度局部特征聚合网络(Deep local feature aggregation network, DLAN)[37];(2)基于点云的三维模型检索算法包括PointNet[4]和PointNet++[5];(3)基于视图的三维模型检索算法包括MVCNN[6]、RotationNet[38]、组合视图卷积神经网络(Group-view convolutional neural network, GVCNN)[7]和基于视图的图卷积神经网络(View-based graph convolutional network, View-GCN)[8]。

2.1 基于体素的三维模型检索算法

2.1.1 3D ShapeNet网络

基于体素的方法[2-3]利用三维空间中体素的集合来表示三维模型,然后在体素的基础上建立神经网络提取特征并进行三维模型的识别和检索。

Wu等[3]提出的3D ShapeNets网络,在三维体素网格上用二元变量的概率分布来表示三维几何模型,如图1所示。对每个三维模型,建立一个30×30×30大小的三维网格,将每个三维网格表示为二值化向量:1表示体素在该网格内,0表示体素没有在该网格内。3D ShapeNets提出了一种深度卷积网络来学习三维模型在三维网格中的体素化特征,该网络共有5层网络结构:第1层卷积层有48个卷积

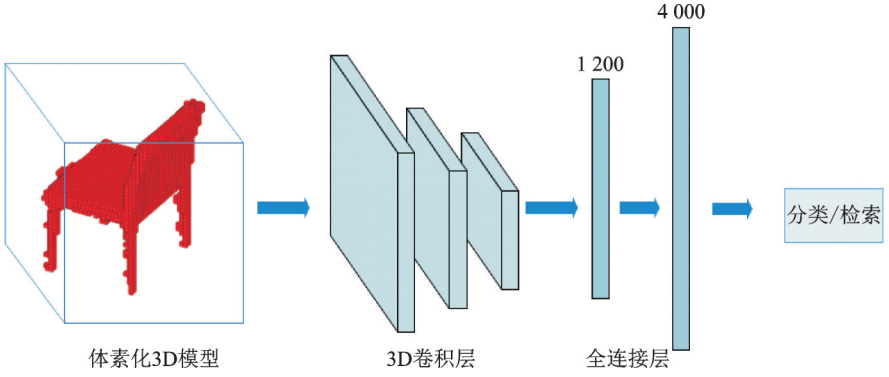


图1 3D ShapeNets结构示意图

Fig.1 Schematic of the 3D ShapeNets structure

核,每个卷积核大小为6,卷积步长为2;第2层卷积层有160个卷积核,每个卷积核大小为5,卷积步长为2;第3层卷积层有512个卷积核,每个卷积核大小为4;第4层是全连接层,有1 200个隐藏单元;第5层是最后一层,包含4 000个隐藏单元,这一层的输出作为整个三维模型的特征。

基于体素的三维模型处理算法将三维模型规范化处理,可以利用3D卷积神经网络进行特征学习。30×30×30大小的网格在神经网络中所需要的参数量与一个165像素×165像素大小的二维图像很接近,但是其计算量和内存消耗量随着分辨率的3次方增长,因此3D ShapeNets网络结构在分辨率较高的三维模型数据集中表现较差。

2.1.2 DLAN网络

局部特征聚合是二维图像或三维模型研究中一种非常流行和行之有效的方法。近年来,研究发现基于端到端的深度三维卷积神经网络(3D-deep CNN, 3D-DCNN)在三维模型的分类和检索任务上取得了不错的效果[3]。然而基于3D-DCNN的方法也存在不足:一方面是对三维旋转缺乏不变性,另一方面在应用过程中形状量化过于粗糙,缺失详细的几何特征。

Furuya等[37]提出了一种深度局部特征聚合网络DLAN。该网络结构可以分为局部特征提取模块和特征聚合模块。如图2所示,局部特征聚合模块对局部三维特征进行提取和编码,利用三维卷积层和全连接层生成中层局部特征。特征聚合模块对局部三维特征进行聚合,通过池化层将局部特征提取模块计算得到的中层局部特征集聚合为每个三维模型的单个特征。最后,利用全连接层对聚合后的整体特征进行降维或压缩,利用最终的特征进行三维模型检索。

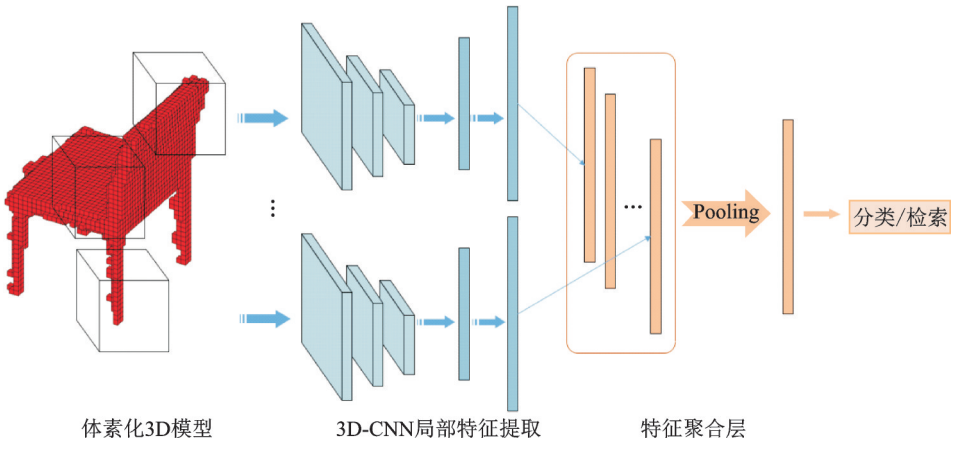


图2 DLAN结构示意图

Fig.2 Schematic of DLAN structure

DLAN采用了先从局部提取特征、再将特征聚合的思想,能学习得到三维模型更细节的信息,并且相对于3DshapeNets,DLAN可以处理更高分辨率的三维模型结构。

2.2 基于点云的三维模型检索算法

点云是一种重要的几何结构数据表示类型,用三维欧氏空间里点的集合来表示三维模型,点云数据相对于其他类型的数据有几条非常重要的性质:

(1) 无序性。点云是三维空间中一系列没有顺序的点的集合,处理点云数据的模型需要对点的不同排列表示不变性。

(2) 相关性。一个三维模型由三维空间中一定数量的点构成,每个点之间存在一定的相互关系,需要同时考虑局部特征和全局特征。

(3) 不变性。点云数据需要对几种空间变换保持不变性,比如全部点云数据整体的平移和旋转等。

由于点云数据的不规则性和无序性,难以用传统的2D深度网络来处理三维点云数据。常见的方法将无序的点云数据转换为三维网格中的规则化的体素来处理,以便神经网络学习过程中的权值共享等操作,但是这种方法降低了三维模型的分辨率并且忽略了很多几何结构信息。

2.2.1 PointNet网络

PointNet网络[4]是一种直接在无序点云数据集上提取特征的新型神经网络,充分考虑了点云特征的无序性、每个点之间的关联性以及整体点云的置换不变性,达到了很好的分类和检索效果。点云数据可以表示为一系列点的集合$\{P_i|i=1,\cdots,n\}$,每个点P_i的数据为其三维坐标(x,y,z)以及一些额外信息,比如颜色等。

如图3所示,PointNet网络结构以n个点的三维坐标作为输入,针对点之间的相关性,设计了局部信息转换模块;针对点的无序性,利用最大池化层作为对称函数来聚合整个三维模型的特征为

$$f(x_1,x_2,\cdots,x_n)=g(h(x_1),h(x_2),\cdots,h(x_n)) \tag{1}$$

式中:x_i表示第i个点;h表示特征提取函数;g为对称函数,例如最大池化层和平均池化层等,用来将每个点的特征聚合成全局特征。最终PointNet网络会得到一个全局的整体特征,利用这个特征可以进行分类和检索任务。

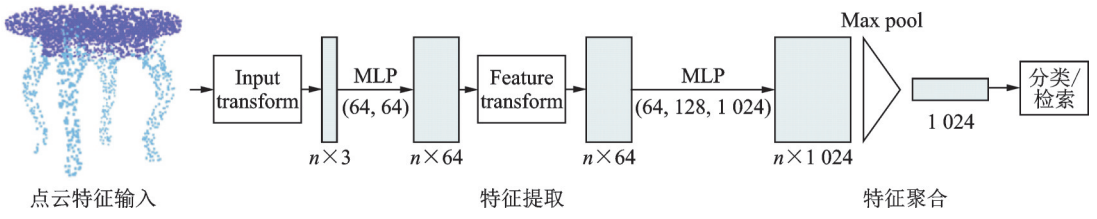


图3 PointNet结构示意图

Fig.3 Schematic of the PointNet structure

2.2.2 PointNet++网络

PointNet只是学习每个点的空间编码,然后将所有的点聚合为模型整体特征,不能很好地捕捉三维模型的局部结构变化,限制了其在复杂场景下进行三维模型识别的泛化能力和对细粒度三维模型的判断能力,并且PointNet是在均匀采样密度的点集上进行训练,这也会造成在复杂应用场景下的性能下降。PointNet++[5]网络在其基础上进行改进,提出了一个层级化的特征学习网络,可以对三维模型的局部特征进行很好的学习,并且可以处理不同尺度的点云特征。

如图4所示,PointNet++参考卷积神经网络使用不同尺度卷积核的思想,从不同尺度提取点云特性,然后聚合得到最终的整体特征。PointNet++主要包含3种模块:采样模块、分组模块和特征提取模块。

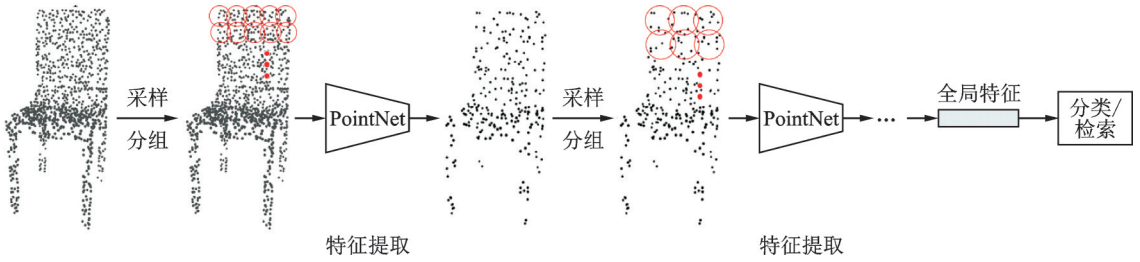


图4 PointNet++结构示意图

Fig.4 Schematic of the PointNet++ structure

在特征提取模块中,用PointNet提取局部特征,不同特征提取模块之间共享参数。采样模块旨在选取采样中心点,将点云数据划分为局部点云集,在采样时用最远点采样(Farthest point sampling,FPS)方法来选取中心点,即先随机选择一个点,然后选择第2个离该点最远的点,循环此操作,直到选出需要的部分点云集数。分组模块确定局部点云集的尺度,在这里认为不同的区域可能会有不同的密度,所以通过选取三维几何空间尺度来确定采样的范围,即在中心点周围的规定尺度的三维球体空间中对点进行采样,这样可以更好地提取局部信息。

2.3 基于视图的三维模型检索算法

如图5所示,基于视图的三维模型检索算法主要的思想是使用虚拟相机将三维模型转换为二维投影视图图像,进而利用深度神经网络提取视图的高阶特征,然后将多个视图的特征融合为一个全局描述子,从而进一步进行检索或分类。研究人员提出了各种算法对视图特征进行融合:MVCNN直接用池化操作对视图进行融合;RotationNet考虑了视图的姿态信息;GVCNN对不同视图进行分组;View-GCN利用图卷积网络对视图进行聚合。得益于深度学习在二维图像处理领域高度成熟的相关技术,以及二维图像庞大的带标签数据集可以对模型进行有效的预训练,基于视图的三维模型检索算法取得了很好的效果。

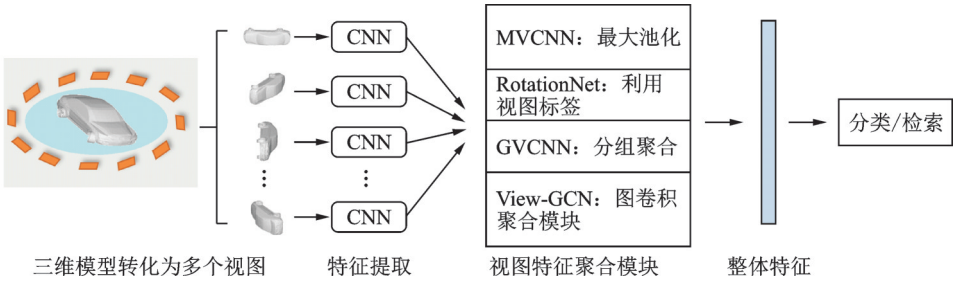


图5 基于视图的三维模型检索算法结构示意图

Fig.5 Schematic of the view-based 3D model retrieval algorithm

2.3.1 MVCNN网络

基于视图的三维模型处理方法,针对每个三维模型生成一组视图数据,常规的方法是为每个视图提取得到一个2D图像描述符,然后根据投票或对齐方案直接使用单个描述符来执行识别任务。然而将视图与三维模型进行对齐比较困难。与上述简单的方法相比,结合来自多个视图特性的聚合表示更为可取,因为它产生了一个全局的、更有代表性的描述符来表示3D形状。

Su等提出的MVCNN[6]框架,首先将三维模型投影成多个二维视图,然后利用卷积神经网络分别对每个视图进行特征提取,再利用最大池化层将多个视图聚合成为一个包含整个三维模型的全局特征,最后利用这个全局特征进行进一步的检索或分类。非常简单的网络架构取得了较好的效果。

2.3.2 RotationNet网络

MVCNN在将多个视图特征聚合为一个视图特征时,每个视图的地位是平等的,这可能忽视了一定的有效信息。RotationNet网络[38]在MVCNN的基础上将视图的姿态标签作为潜在变量,在训练过程中利用未对齐数据集对每个视图的姿态进行无监督自对齐优化。

一个训练样本由M个视图$\{x_i\}_{i=1}^{M}$组成,其类别标签为$y\in\{1,\cdots,N\}$,其视图标签为$v_i\in\{1,\cdots,M\}$,M表示每个三维模型的视图数,N表示三维模型数据集的类别数。在利用了视图的姿态作为隐藏信息来辅助模型进行训练后,RotationNet取得了优于MVCNN的检索和分类效果。

2.3.3 GVCNN网络

更进一步,Feng等研究发现三维模型的多个视图之间存在一定的相关性,因此在视图和全局特征之间设计了一个对视图进行组合的模块,提出了GVCNN网络[7],将不同视角下卷积神经网络提取得到的视图特征按照其判别性强弱进行分组,对于N个视图特征$V=\{v_1,v_2,\cdots,v_n\}$,分别送入设计好的计算权重的全连接层,得到输出值$O=\{o_1,o_2,\cdots,o_n\}$,每个视图的判别性强弱定义为

$$\xi(v_i)=\mathrm{sigmoid}(\log(\mathrm{abs}(o_i))) \tag{2}$$

式中abs(·)为绝对值函数。按照式(2)计算的权重对N个视图进行分组,然后在每一组内进行特征池化操作,最后再在每个组间对特征进行聚合最终的三维模型全局特征,能够更充分地利用三维模型多个视图之间的相关性。

2.3.4 View-GCN网络

同样为了对多视图特征进行更加有效的聚合,Wei等提出了View-GCN网络[8],将图卷积网络[39]应用在多视图特征聚合上。根据生成二维视图时虚拟相机的三维空间位置建立图结构,更好地保留了视图之间的几何结构信息,并且设计了局部图卷积模块和非局部信息传递模块,兼顾了相邻视图和偏远节点视图之间的信息传递,也取得了很好的分类和检索效果。

3 基于二维图像的三维模型检索算法

3.1 基于二维真实图像的三维模型检索算法

近年来三维模型数据大量涌现,但是只有很少的数据得到了有效的标记,绝大多数是无标记的数据,而经过深度神经网络[40-42]多年的发展,二维图像处理技术已经非常完善,并且有充足的有标记数据[43]可以利用。那么一个非常直观的想法就是利用迁移学习相关算法将从标签丰富的二维图像中学习得到的知识迁移到无标签的三维模型数据中去,促进三维模型的大数据管理。

本节详细对比分析了基于二维真实图像的跨域三维模型检索算法:流形嵌入分布对齐网络(Manifold embedded distribution alignment, MEDA)[44],联合域适应网络(Joint adaptation network, JAN)[45],深度域对抗网络(Domain-adversarial neural network, DANN)[46]和两级嵌入对齐网络(Dual-level embedding alignment network, DLEA)[10]。

图6展示了基于二维图像的无监督跨域三维模型检索算法的基本框架:对于二维真实图像直接用卷积神经网络提取特征,对于三维模型采用基于视图的三维模型表示形式提取高阶特征,然后对源域、目标域数据进行跨域特征学习。

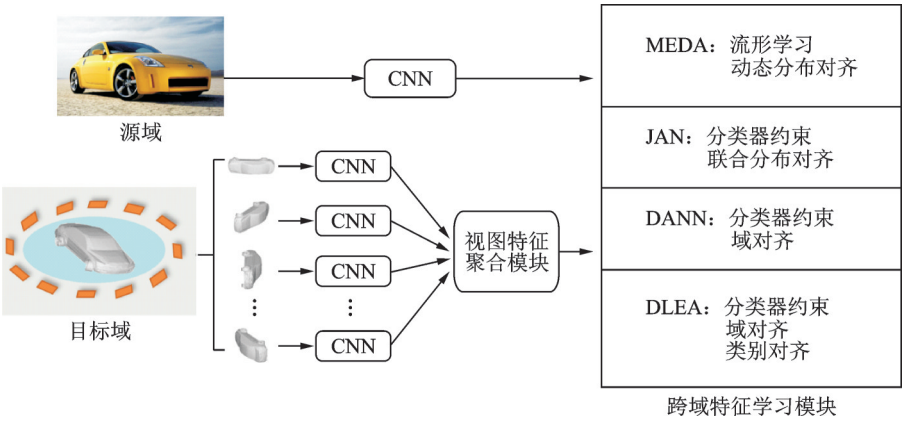


图6 基于二维真实图像的跨域三维模型检索算法结构示意图

Fig.6 Schematic of the 2D real image-based cross-domain 3D model retrieval algorithm

3.1.1 MEDA网络

MEDA网络[44]利用MVCNN方法提取二维图像特征和三维模型特征,然后将原始空间转化为高维的高斯曼流形,学习其流形中的领域不变分类器,进一步实现流形域自适应的动态分布对齐。

原始特征空间中不同模态之间存在模态差异,根据流形假设嵌入在流形空间的点和他们的相邻节点往往具有相似的性质,因此MEDA提出流形特征变换,以此来减小不同域之间的数据漂移,然后动态衡量边缘分布和条件分布的重要性,学习得到最终的模型。

3.1.2 JAN网络

Long等提出的联合域适应网络JAN[45]利用联合最大平均误差准则(Joint maximum mean discrepancy, JMMD),通过对齐域适应网络多个特定层的联合分布来优化模型。通常进行域适应时只针对最后的特征输出进行约束,而忽略了中间层的信息,因此当网络的层数比较多时效果不是很理想。JAN综合网络输出层和网络中间层的特征,利用JMMD准则对模型进行整体约束,即

$$\min \frac{1}{n_s}\sum_{i=1}^{n_s} J(f(x_i), y_i) + \lambda \hat{D}_{\xi}(S, T) \tag{3}$$

式中:第一项表示交叉熵分类损失,用源域带标签的数据进行约束;第二项$\hat{D}_{\xi}$表示针对网络输出层和中间层的联合对齐损失函数。这样从适应网络的多个层次对模型进行约束,可以达到较好的迁移效果。

3.1.3 DANN网络

2016年,Ganin等提出的深度域对抗神经网络DANN[46]首次介绍了深度对抗性域适应算法,在解决无监督跨域学习等方面取得了显著的效果。在处理基于图像的无监督三维模型检索算法时,将有标签的二维数据集$\{x_s^{(i)}, y_s^{(i)}\}_{i=1}^{n_s}$作为源域$\{x_s^{(i)} \in X_s, y_s^{(i)} \in Y_s\}$,将无标签的三维模型多视图数据集$\{x_t^{(i)}\}_{i=1}^{n_t}$作为目标域$\{x_t^{(i)} \in X_t\}$。源域二维图像数据直接用卷积神经网络提取特征,目标域三维模型多视图数据用卷积神经网络分别对多个视图进行提取特征,利用最大池化操作将多视图特征进一步聚合为全局特征描述子。

设计如下损失函数对模型进行约束

$$E(\theta_f, \theta_y, \theta_d) = L_y(\theta_f, \theta_y) - \lambda L_d(\theta_f, \theta_d) \tag{4}$$

式中:L_y表示分类器的损失函数;L_d表示域判别器的损失函数;$\theta_f, \theta_y, \theta_d$分别表示特征提取网络、分类器网络和域判别网络的参数,用对抗性训练[47]的方式优化模型为

$$(\hat{\theta}_f, \hat{\theta}_y) = \arg\min E(\theta_f, \theta_y, \hat{\theta}_d) \tag{5}$$

$$\hat{\theta}_d = \arg\max E(\hat{\theta}_f, \hat{\theta}_y, \theta_d) \tag{6}$$

经过上述优化后的模型,可以学习得到域一致性的特征,最大可能地忽略域差异,更好地对齐二维真实图像与三维模型数据的特征分布,达到更好的分类和检索效果。

3.1.4 DLEA网络

Zhou等提出的DLEA[10]网络将有标签的二维数据集$\{x_s^{(i)}, y_s^{(i)}\}_{i=1}^{n_s}$作为源域$\{x_s^{(i)} \in X_s, y_s^{(i)} \in Y_s\}$,将无标签的三维模型多视图数据集$\{x_t^{(i)}\}_{i=1}^{n_t}$作为目标域$(x_t^{(i)} \in X_t\}$。源域二维图像数据直接用卷积神经网络提取特征,目标域三维模型多视图数据,用卷积神经网络分别对多个视图进行提取特征,并利用注意力模块[48]对多个视图赋予合适的权值,进一步聚合多个视图为全局特征描述子,期望多个视图中与二维图像角度相近的视图可以获得更大的权值。

模型训练过程中,首先利用源域有标签的二维图像数据集训练特征提取器和分类器,用常用的交叉熵损失约束模型进行训练

$$L_{\mathrm{C}}(X_s, Y_s)=E_{(x,y)\sim D_s}[J(f(x),y)] \tag{7}$$

式中：D_s表示源域；f表示特征提取网络。

然后利用对抗域适应算法在领域级别用L_{DC}损失函数进行约束，在类别级别用L_{SM}损失函数进行约束，将源域数据和目标域数据进行对齐

$$L_{\mathrm{DC}}(X_s, X_t)=E_{x\sim D_s}[\log(1-D(G(x)))]+E_{x\sim D_t}[\log(D(G(x)))] \tag{8}$$

式中：G表示分类器；D代表域判别器。通过式(8)损失函数的约束能够消除源域二维图像数据集与目标域三维模型数据集的整体域差异，使得源域与目标域的特征分布更加相似。

$$L_{\mathrm{SM}}(X_s, Y_s, X_t)=\sum_{k=1}^{K}\phi(C_{\mathrm{S}}^k, C_{\mathrm{T}}^k) \tag{9}$$

式中：K表示类别数；C_{S}^k, C_{T}^k分别表示第k类的源域类中心和目标域类中心。通过式(9)的损失函数能够精确地在类别级别对二维图像数据与三维模型数据进行特征对齐[49]。

最终训练好的网络能消除二维图像与三维模型之间的域差异，将在二维图像中学习到的信息很好地用在处理三维模型数据的过程中。

3.2 基于二维草图的三维模型检索算法

基于草图的三维模型检索的目标是根据输入的草图检索得到一系列三维模型，该方案更简单、更直观，便于用户更好地学习和使用，有较为广泛的应用，如基于草图的三维建模、基于草图的三维动画设计等。然而由于人类手绘的三维模型比较简单，与精确建模的三维模型在细节、精度等方面存在巨大的模态差异，基于草图的三维模型检索存在巨大的挑战，吸引了众多研究者的关注[50-53]。

本节详细分析对比了基于二维草图的以下4种三维模型检索算法：跨域流形排序算法(Cross-domain manifold ranking, CDMR)[52]，基于孪生神经网络的方法(Siamese-based CNN, Siamese-CNN)[54]，深度相关度量学习方法(Deep correlated metric learning, DCML)[55]及其变体深度整体相关度量学习方法(Deep correlated holistic metric learning, DCHML)[56]。

3.2.1 CDMR算法

2013年，Furuya等提出了一种基于度量学习的跨域流形排序算法CDMR[52]，如图7所示。CDMR算法在二维草图特征和三维模型特征之间构建一个流形，即一个图结构W，图结构的大小为：$(N_s+N_m)\times(N_s+N_m)$，其中N_s为二维草图样本的个数，N_m为三维模型样本的个数。W_{ij}表示第i个二维草图与第j个三维模型的相似度，在计算了两个样本的特征距离$d(i,j)$之后，其相似度可以表示为

$$w_{ij}=\begin{cases}\exp(-d(i,j)/\sigma) & i\neq j\\ 0 & \text{其他}\end{cases} \tag{10}$$

式中：参数σ控制相关性，在计算草图到草图、草图到三维模型、三维模型到三维模型的相似度时，采用不同的参数σ。

在计算完跨域流形之后，采用流形排序算法在跨域流形图中计算检索过程中样本的相似性为

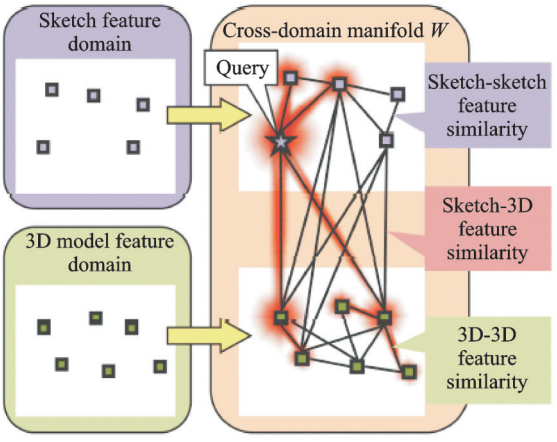


图7 CDMR结构示意图[52]

Fig.7 Schematic diagram of CDMR structure[52]

$$F=(I-\alpha S)^{-1}Y \tag{11}$$

式中:I表示单位矩阵;S表示归一化的拉普拉斯矩阵$S=D^{-\frac{1}{2}}WD^{-\frac{1}{2}}$;$Y$表示初始指示矩阵,是一个对角矩阵。则$F$表示最终的相似性矩阵,较大的$F_{ij}$值表示较大的相似度。

3.2.2 Siamese-CNN网络

2015年Wang等提出Siamese-CNN[54]网络,其结构如图8所示。该网络利用孪生网络[57]的思想,构建二维草图和三维模型视图的样本对,然后对二维图像和三维模型设计完全相同的特征提取网络结构,利用孪生网络定义损失函数,对两个网络的输出进行相似性约束,从而缩小二维草图与三维模型之间的域差异。

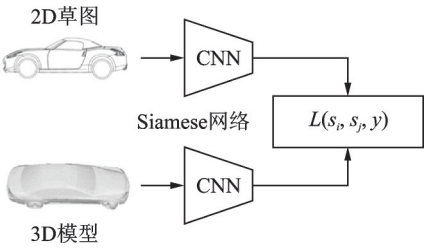


图8 Siamese-CNN结构示意图

Fig.8 Schematic diagram of Siamese-CNN structure

孪生网络通常采用一对样本作为输入,损失函数定义如下

$$L(s_i,s_j,y)=(1-y)\alpha D_w^2+y\beta e^{rD_w} \tag{12}$$

式中:s_i,s_j表示2个不同的样本;y表示相似性标签;α、β、r表示超参数;D_w表示两个样本在特征空间的距离。在Siamese-CNN中,针对跨域三维模型检索算法设计了损失函数为

$$L(s_1,s_2,v_1,v_2,y)=L(s_1,s_2,y)+L(v_1,v_2,y)+L(s_1,v_1,y) \tag{13}$$

式中:s_1、s_2表示不同的二维草图,v_1、v_2表示不同的三维模型视图,并且s_1、v_1来自同一类,s_2、v_2来自同一类,因此一个y就可以表明4个样本之间的关系。利用式(13)可以约束神经网络进行训练,以缩小域差异,进行更好的跨域三维模型检索。

3.2.3 DCML方法

Dai等在2017年提出了一种基于深度相关度量学习的方法DCML[55]。如图9所示,DCML对二维草图源域和三维模型目标域各训练一个深度神经网络,用同一组联合损失函数来约束每个域内最大化类间相似性及最小化类内相似性,同时减小不同域间的整体域差异,最终学习两个深度的非线性变换,将两个域的特征映射到同一个非线性特征空间中,在相同的特征空间中对二维草图和三维模型进行检索和匹配。

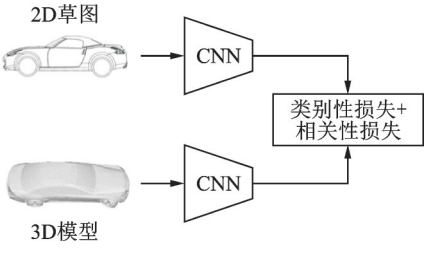


图9 DCML结构示意图

Fig.9 Schematic diagram of DCML structure

经过各自的深度神经网络,源域二维草图和目标域三维模型分别得到了各自的特征:$F_s=\{x_1,x_2,x_3,\cdots\}$,$F_t=\{y_1,y_2,y_3,\cdots\}$,联合损失函数包括两部分:判别性损失和相关性损失,即

$$L=\alpha L^d+(1-\alpha)L^c+\lambda\left(\left\|\boldsymbol{W}^s\right\|_F^2+\left\|\boldsymbol{W}^t\right\|_F^2\right) \tag{14}$$

式中:等号右侧第3项是用来防止过拟合的正则化项;判别性损失L^d用于在源域中构建正样本和负样本,利用三元组损失[58]对模型进行约束;相关性损失L^c用来约束不同域之间的分布一致性,首先在源域与目标域的所有样本中构建正样本和负样本,设计三元组损失对模型进行约束,然后对源域和目标域中相同类的样本在特征空间最小化其欧式距离来进行相似性约束。

由于二维草图与三维模型之间存在巨大的域差异,直接对神经网络最后输出层的特征进行约束很

难达到理想的效果，Dai等在DCML的基础上对神经网络隐藏层的特征进行进一步约束，提出DCHML算法[56]，达到了更好的效果。

4 实验分析

4.1 数据集

(1)ShapeNet Core55[59]：大规模三维模型数据集包含55个类别，共计51 162个三维模型样本。将数据集划分为训练集、验证集和测试集，大约分别占70%，20%和10%。ShapeNet Core55数据集示例如图10所示。

(2)ModelNet40[3]：大规模三维模型数据集包含40个类别，共计12 311个三维模型，将数据集划分为训练集和测试集，大约分别占80%和20%。ModelNet40数据集示例如图11所示。

图10 ShapeNet Core55数据集示例

Fig.10 Example of the ShapeNet Core55 dataset

图11 ModelNet40数据集示例

Fig.11 Example of the ModelNet40 dataset

(3)SHREC'14 LSSTB[12]：二维草图数据集包含13 680张草图，共分为171类，每类80张草图，每张草图都在三维模型数据集中有对应的模型。三维模型数据集包含8 987个三维模型，共分为171类，每类平均大约53个样本。SHREC'14 LSSTB数据集示例如图12所示。

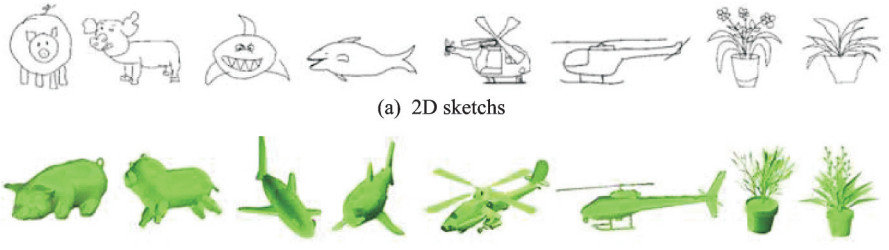


图12 SHREC'14 LSSTB数据集示例

Fig.12 Examples of SHREC'14 LSSTB dataset

(4)MI3DOR[11]：基于真实图像的三维模型检索数据集，二维真实图像数据集包含21个类别，每类1 000个样本，共计21 000个样本，将数据集划分为训练集和测试集，分别占50%和50%；三维模型数据集包含21个类别，共计7 690个样本，样本类别分布不完全均衡，训练集包括3 842个样本，测试集包括3 848个样本。MI3DOR数据集示例如图13所示。

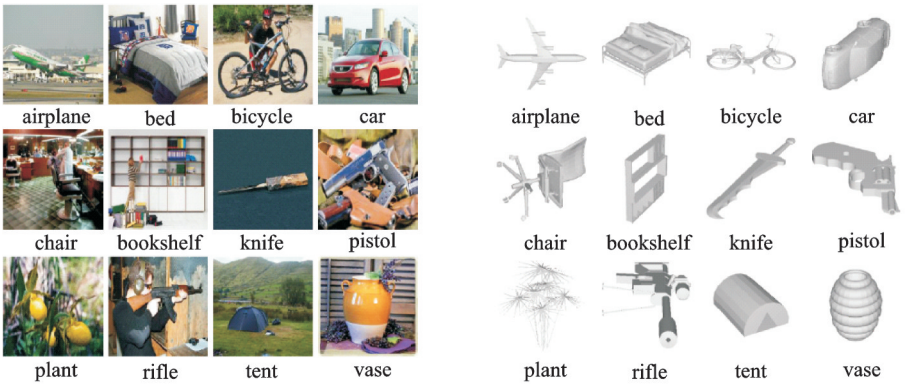


图13 MI3DOR数据集示例

Fig.13 Examples of MI3DOR dataset

4.2 相似性及检索效果评测指标

4.2.1 相似性度量

在三维模型检索过程中,利用各种特征提取方法对三维模型提取高阶全局特征,然后利用提取得到的特征向量进行模型间的匹配和检索。对特征向量进行检索和匹配的最直观的做法是对特征向量之间的相似度进行判断,可以用以下度量方式对相似度进行量化。

(1)闵可夫斯基距离,表达式为

$$D_{\mathrm{p}}(\boldsymbol{X},\boldsymbol{Y})=\left(\sum_{i=1}^{N}\left\|x_i-y_i\right\|^{p}\right)^{\frac{1}{p}} \tag{15}$$

式中:$\boldsymbol{X}$、$\boldsymbol{Y}$表示不同三维模型的特征向量;N表示特征向量的维度;当$p=1$时,式(15)为曼哈顿距离;当$p=2$时,式(15)为常见的欧几里得距离。

(2)余弦相似度,表达式为

$$D_{\mathrm{c}}(\boldsymbol{X},\boldsymbol{Y})=\frac{\boldsymbol{X}\cdot\boldsymbol{Y}}{\|\boldsymbol{X}\|\|\boldsymbol{Y}\|} \tag{16}$$

由上述距离计算得到相似度矩阵:$\boldsymbol{D}$:$P\times P\rightarrow\mathbf{R}^{+}\cup\{0\}$,其中$P$为特征空间的一个子空间。

则对任意的特征点$x,y\in P$,有如下性质:①$D(x,y)\geqslant 0$;②$D(x,y)=D(y,x)$;③$D(x,y)=0\Leftrightarrow x=y$。

4.2.2 检索效果评价指标

根据特征向量计算得到的相似度矩阵,有如下几个测评指标来对三维模型检索的效果进行综合评价。

(1)平均召回率(Average recall, AR)

$$\mathrm{Recall}=\frac{N_{\mathrm{C}}}{N_{\mathrm{All}}} \tag{17}$$

$$\mathrm{AR}=\sum_{i=1}^{N_{\mathrm{Class}}}\mathrm{Recall}(i) \tag{18}$$

式中:N_{C}表示检索到的正确的模型的数量;N_{All}为数据集中所有相关物体的数量;Recall(i)为第i类模型的召回率。

(2)平均精确率(Average precision, AP)

$$\text{Precision}=\frac{N_{\text{C}}}{N_{\text{RAll}}} \tag{19}$$

$$\text{AP}=\sum_{i=1}^{N_{\text{Class}}}\text{Precision}(i) \tag{20}$$

式中:N_{RAll}表示所有被召回的模型的数量;Precision(i)表示第i类模型的精确率。

(3)平均精确率的均值(mean Average precision, mAP)。对于每一个查询样本都会有一个AP值,综合所有的AP值计算所有数据集的mAP指标,假设Q为所有查询样本的个数,则有

$$\text{mAP}=\frac{1}{Q}\sum_{j=1}^{Q}\text{AP}(j) \tag{21}$$

(4)最近邻(Nearest neighbor, NN),表示检索列表中第一个检索结果的检索精度。

(5)第一层级(First tier, FT),表示前T个检索结果的检索精度,T表示整个数据集中相关样本的个数。

(6)第二层级(Second tier, ST),表示前$2T$个检索结果的检索精度。

(7)F度量(F-measure, F),联合评价检索结果的精准率和召回率,表达式为

$$F=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{22}$$

(8)折损累计增益(Discounted cumulative gain, DCG),对检索结果排名靠前的样本赋予较大的权重,从而综合评测,表达式为

$$\text{DCG}=\sum_{i=1}^{p}\frac{\text{rel}_i}{\log_2(i+1)} \tag{23}$$

式中:rel_i表示第i个样本的相关度。在三维模型检索中,通常1表示相关,即检索正确;0表示无关,即检索错误。

(9)归一化折损累计增益(Normalized discounted cumulative gain, NDCG),用DCG的最大值IDCG对DCG进行归一化处理,即

$$\text{IDCG}=\sum_{i=1}^{\text{REL}}\frac{\text{rel}_i}{\log_2(i+1)} \tag{24}$$

$$\text{NDCG}=\frac{\text{DCG}}{\text{IDCG}} \tag{25}$$

式中:REL表示检索结果列表按照相关性大小降序排列,即检索列表中都是正确的结果。

(10)平均归一化修正检索排序(Average normalized modified retrieval rank, ANMRR)是一种基于排序的度量,考虑了检索对象中相关对象的排序信息。

$$\text{ARR}(q)=\sum_{i=1}^{\text{Ng}(q)}\frac{r(i)}{\text{Ng}(q)} \tag{26}$$

式中:Ng(q)表示数据集中检索样本q的所有相关样本的个数;$r(i)$表示检索q检索到的第i个样本的主观排序,即

$$r(i)=\begin{cases}r(i) & r(i)\leqslant K(q)\\ K & r(i)>K(q)\end{cases} \tag{27}$$

式中:K定义为$\min(4\times\text{Ng}(q),2\times\text{GTM})$,其中$\text{GTM}=\max(\text{Ng}(q))$,则有

$$\mathrm{NMRR}(q)=\frac{\mathrm{ARR}(q)-\frac{\mathrm{Ng}(q)}{2}-0.5}{K(q)-\frac{\mathrm{Ng}(q)}{2}+0.5} \tag{28}$$

$$\mathrm{ANMRR}=\frac{1}{Q}\mathrm{NMRR}(q) \tag{29}$$

式中Q代表查询样本的总数。

值得注意的是,上述检索指标中只有ANMRR值越小表示算法性能越好,其他检索指标均为越大表示算法性能越好。

4.3 三维模型检索算法对比及实验结果

三维模型检索算法对比如表1所示。

表1 三维模型检索算法对比

Table 1 Comparison of 3D model retrieval algorithms

算法	Query			Gallery	3D模型表示形式	有无监督	域适应
	2D图像	2D草图	3D模型	3D模型			
3DshapeNet[3]			√	√	体素	有	无
DLAN[37]			√	√	体素	有	无
PointNet[4]			√	√	点云	有	无
PointNet++[5]			√	√	点云	有	无
MVCNN[6]			√	√	视图	有	无
RotationNet[38]			√	√	视图	有	无
GVCNN[7]			√	√	视图	有	无
View-GCN[8]			√	√	视图	有	无
MEDA[44]	√			√	视图	无	流形学习
JAN[45]	√			√	视图	无	分布对齐
DANN[46]	√			√	视图	无	对抗域适应
DLEA[10]	√			√	视图	无	对抗域适应+类别对齐
CDMR[52]		√		√	视图	有	度量学习
Siamese-CNN[54]		√		√	视图	有	Siamese网络
DCML[55]		√		√	视图	有	深度度量学习
DCHML[56]		√		√	视图	有	深度度量学习

4.3.1 基于模型的三维模型检索算法对比及实验性能分析

采用ModelNet40数据集评估上述模型的分类效果和检索效果,利用ShapeNet Core55数据集来评估上述算法的检索效果。ModelNet40数据集和ShapeNet Core55数据集都提供标准的三维obj格式文件,可以自行转换为体素数据、点云数据或者多视图数据。表2中,ShapeNet Core55数据集每类样本的数目并不均衡,因此利用ShapeNet Core55数据集对算法的检索性能进行分析时,为了更好地对不同类别的模型检索结果进行综合评价,采用了两种不同的评测指标计算方法:宏观平均和微观平均。宏观平均指的是在整个数据集上对所有样本计算各种检索指标(F, mAP, NDCG),每个三维模型样本在计算检索指标时占相同地位。微观平均指的是首先在每个类别内计算检索指标(F, mAP, NDCG),然后对每个类别的检索指标取平均值,每个类别在计算检索指标时占相同地位。

表2 基于模型的三维模型检索算法在ModelNet40和ShapeNet Core55数据集上的性能比较

Table 2 Performance comparison of the model-based 3D model retrieval algorithms on ModelNet40 and ShapeNet Core55 datasets

算法	ModelNet40		ShapeNet Core55					
	分类	检索	检索(微观)			检索(宏观)		
	ACC	mAP	*F*	mAP	NDCG	*F*	mAP	NDCG
3DshapeNet[3]	0.847	0.492						
DLAN[37]			0.712	0.663	0.762	0.505	0.477	0.563
PointNet[4]	0.892							
PointNet++[5]	0.919							
MVCNN[6]	0.901	0.795	0.764	0.735	0.815	0.575	0.566	0.640
RotationNet[38]	0.924		0.798	0.772	0.865	0.590	0.583	0.656
GVCNN[7]	0.931	**0.857**						
View-GCN[8]	**0.976**		**0.806**	**0.784**	**0.852**	**0.611**	**0.602**	**0.665**

上述基于模型的三维模型处理算法都借鉴了深度学习的思想,采用不同的三维模型处理方式(体素、点云、多视图),设计了有效的深度网络结构,提取三维模型高阶特征进行三维模型检索或分类。3DshapeNets将三维模型数据体素化,每个三维模型表示为一个30×30×30的三维网格;DLAN先用小的体素网格分别提取三维模型的局部信息然后聚合为整体信息;PointNet和PointNet++设计了深度网络结构直接在点云数据上提取特征;MVCNN、RotationNet、GVCNN和View-GCN选取多视图数据作为输入数据,设计不同的多视图融合方式,提取高阶全局信息。分析实验结果可以观察得到:

(1) 3DshapeNets和DLAN是基于体素的三维模型处理算法,采用3D-CNN结构提取体素化的三维模型数据并处理,由于其计算量和内存占用量随着三维模型网格分辨率的3次方增长,难以处理高分辨率的三维模型,实验性能低于基于点云的和基于视图的方法。

(2) PointNet和PointNet++两种基于点云的三维模型处理算法,充分考虑了点云特征的无序性、平移旋转不变性和点之间的相互关联性,取得了令人满意的实验效果。PointNet++进一步考虑了点云尺度变化,设计了层级化的点云特征提取结构,取得了更好的实验效果。

(3) 基于视图的方法,得益于卷积神经网络在二维图像上的成熟技术,性能普遍偏高;并且不同视图之间存在关联信息,RotationNet利用了视图的姿态信息作为隐藏信息参与模型的训练,GVCNN根据视图的判别性对视图进行分组,View-GCN更好地利用了视图相互之间的结构信息,都取得了更好的三维模型检索效果。

4.3.2 基于二维真实图像的三维模型检索算法对比及实验性能分析

基于二维真实图像的跨域三维模型检索算法,在目标域无标签信息的情况下,充分利用二维图像的标签信息,结合迁移学习相关算法,协助处理三维模型数据,有重要的现实意义。

采用MI3DOR数据集来对上述算法进行性能评估,结果如表3所示。MI3DOR数据集分为21个类别,其训练集中包含10 500个二维真实图像和3 842个三维模型,其测试集中包含10 500个二维真实图像和3 848个三维模型。三维模型数据提供三维obj格式文件和多视图文件。

针对MI3DOR数据集,在进行特征提取时上述算法都选择AlexNet[40]网络作为基本网络架构,AlexNet网络包含5个卷积层$conv_1$~$conv_5$和3个全连接层fc_6~fc_8。对于二维真实图像,直接用AlexNet提取特征,采用其fc_7层的输出作为二维图像的高阶特征;对于三维模型,首先利用AlexNet对

三维模型的每个视图提取特征,采用fc_7的输出作为每个视图的高阶特征,然后利用多视图特征融合方法,将每个三维模型不同视图的高阶特征融合为三维模型的全局高阶特征。

表3 基于二维真实图像的三维模型检索在MI3DOR数据集上的性能比较

Table 3 Performance comparison of the 2D real image-based 3D model retrieval algorithms on MI3DOR dataset

算法	NN	FT	ST	F	DCG	ANMRR
MEDA[44]	0.430	0.344	0.501	0.046	0.361	0.646
JAN[45]	0.446	0.343	0.495	0.085	0.364	0.647
DANN[46]	0.650	0.505	0.643	0.112	0.542	0.474
DLEA[10]	**0.764**	**0.558**	**0.716**	**0.143**	**0.597**	**0.421**

表3中JAN,DANN,DLEA分别设计了相应的域适应模块和损失函数,对AlexNet特征提取网络进行约束和优化。而MEDA在进行域适应时直接利用AlexNet提取得到的特征进行相应的跨域学习,并不对AlexNet网络进行训练。分析实验结果可以得到:

(1) MEDA是一种非深度的迁移学习方法,将特征映射到流形空间,以此来减小不同域之间的数据差异,然后进行自适应的分布适配,取得了不错的实验效果。

(2) JAN考虑了深度域适应网络的多个特定层,设计联合分布进行对齐损失,能够更好地对深度网络进行约束,取得很好的实验效果。

(3) DANN和DLEA都采用了对抗性域适应的思想,能很好地减小域差异,学习得到域不变性特征,显著提高模型的性能;并且DLEA增加了类别中心对齐的约束,对深度域适应算法有很好的增益效果。

4.3.3 基于二维草图的三维模型检索算法对比及实验性能分析

采用SHREC'14 LSSTB数据集对上述算法进行性能评估,结果如表4所示。SHREC'14 LSSTB数据集包含171个类别,其中二维草图数据集包含13 680个样本,每个类别80张草图,其中三维模型数据集包含8 978个样本,每个类别的样本数分布不均衡,平均每类含有53个样本。三维模型数据提供off格式文件。

表4 基于二维草图的三维模检索算法在SHREC'14 LSSTB数据集上性能比较

Table 4 Performance comparison of the 2D sketch-based 3D model retrieval algorithms on SHREC'14 LSSTB dataset

算法	NN	FT	ST	F	DCG	mAP
CDMR[52]	0.109	0.057	0.089	0.041	0.328	0.054
Siamens-CNN[54]	0.239	0.212	0.316	0.140	0.495	0.228
DCML[55]	0.351	0.276	0.335	0.174	0.500	0.282
DCHML[56]	**0.403**	**0.329**	**0.394**	**0.201**	**0.544**	**0.336**

针对基于草图的跨域三维模型检索问题,上述方法分别从度量学习、深度神经网络和深度度量学习等方向给出了解决方案。CDMR设计了一种基于BoF和GALIF的特征表示方式BF-fGALIF来表示草图的特征,采用了BF-DSIFT方法[60]来表示多视图三维模型的特征。而Siamese-CNN,DCML和DCHML都采用深度神经网络的方法来提取二维草图和三维模型的多视图特征。观察实验结果可以发现:

(1) CDMR设计跨域流形结构,采用流形排序算法,对草图和三维模型的相似性进行计算和更新,对跨模态检索有一定的增益效果。

(2) Siamese-CNN采用孪生网络的思想,巧妙地利用孪生网络对输入的二维草图和三维模型的相似度进行判断,取得了较好的模型检索效果。

(3) 深度学习能显著提升模型的性能,DCML和DCHML在深度学习的基础上加入度量学习相关算法,显著地提高了模型的性能;而DCHML在深度网络的中间层也采用度量学习的思想进行约束,取得了更好的实验效果。

4.3.4 不同类别的三维模型检索算法对比及实验性能分析

根据三维模型检索的任务不同,即三维模型检索时查询样本是三维模型还是二维图像,本文将三维模型检索算法分为基于模型的三维模型检索算法和基于二维图像的三维模型检索算法,其中基于二维图像的三维模型检索算法包括基于二维真实图像的方法和基于二维草图的方法。根据表1的算法对比和表2~4的实验结果可以分析得到:

(1) 基于模型的三维模型检索算法,体素、点云和多视图的三维模型表达方式各有优势,得益于深度学习的发展,三维模型检索处理技术得到了很大的提高。

(2) 基于二维真实图像的三维模型检索算法,将二维真实图像作为查询样本,在现实生活中有很大应用价值,二维图像也包含极为丰富的信息,能够充分利用二维图像的信息,对协助处理三维模型有很大的现实意义。

(3) 基于草图的三维模检索算法,由于二维草图过于简单、携带的信息量较少,而三维模型大都是精确建模的结果,二维草图与三维模型之间尚存在巨大的模态差异,具有很大的挑战性,实验性能还有较大的提升空间。

5 结束语

三维模型检索时待检索样本是三维模型,查询样本可以是三维模型,也可以是二维图像。据此本文对基于深度学习的三维模型检索方法进行分类:基于模型的三维模型检索算法和基于二维图像的三维模型检索算法。

针对基于模型的三维模型检索算法,本文着重分析了三维模型不同表示形式的方法:基于体素的方法、基于点云的方法和基于视图的方法,各种方法都有一定的优缺点。将三维模型进行体素化即将三维模型转换为结构化的数据,可以直接利用三维卷积神经网络来处理,但是三维体素网格占的内存太多,现有的计算力限制了高分辨率的三维模型体素化操作。点云类型数据能尽可能全面地表示三维模型的信息,但是点云数据是非结构化的,无法像二维图像一样用一个矩阵来表示,只能用点的坐标来表示三维模型,此外由于三维模型在空间的姿态是任意的,只将物体进行旋转平移等操作时物体本身并不会发生变化,其三维坐标却发生了大范围的迁移。在实际应用场景中,点的个数会有很大范围的误差,好的算法需要能够处理不同尺度的点云数据。基于视图的方法虽然依赖于深度学习在处理二维图像方面的强大能力而取得了成功,但是同样由于内存的限制,无法无限制地使用各个视觉的图像,而固定数量的视图可能无法全面表示三维模型,同时各个视图之间还存在一定的信息冗余。除此之外,使用二维视图数据不可避免地会损失一些三维结构化的信息,尤其是在实际应用场景下,当三维模型比较复杂时很难使用固定数量的视图对三维模型进行有效的表示。如果能更好地利用各种三维模型表示形式的优点,同时减少信息冗余,可能会为三维模型处理带来新的启发。

针对基于二维图像的三维模型检索算法,本文对二维图像做了二维真实图像和二维草图的区分。二维真实图像中携带着大量的信息,各种深度学习方法都在探索如何更全面地提取图像中丰富的信息,均取得了很好的效果。并且二维图像数据集非常完善,有丰富的标注和完善的信息用来训练各种模型,利用迁移学习的思想,以二维真实图像为源域、三维模型为目标域,将从二维图像中学习得到的信息迁移到三维模型中去,有重要的现实意义。因此,本文详细介绍了基于二维真实图像的无监督跨域三维模型检索算法,算法中限定了三维模型不含标签的场景,重点分析了无监督迁移学习在无监督跨域三维模型检索中的性能。本文发现二维图像与三维模型之间的模态差异很难完全消除,但是各种深度域适应算法还是能取得较好的实验效果,对三维模型检索任务有很好的指导作用。基于二维草图的三维模型检索算法,在实际应用中能更方便直观地提升三维模型建模和分析的效率,三维模型结构复杂,假如能够通过简单的草图或者现场根据回忆画出示意图,就可以很好地检索出想要的三维模型,

对三维模型建模人员以及广大用户都有很好的帮助作用。基于此,本文着重关注了基于草图的三维模型检索算法下二维草图和三维模型的跨域检索和匹配问题,发现深度学习和度量学习都有一定的增益效果。但是,由于二维草图过于简单,与复杂的三维模型存在巨大的模态差异,因此基于草图的三维模型检索算法目前仍存在较大的挑战,将是下一步的研究重点。

参考文献:

[1] CHEN D Y, TIAN X P, SHEN Y T, et al. On visual similarity based 3D model retrieval[C]//Proceedings of Computer Graphics Forum. Oxford, UK: Blackwell Publishing, Inc, 2003, 22(3): 223-232.

[2] MATURANA D, SCHERER S. Voxnet: A 3D convolutional neural network for real-time object recognition[C]//Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2015: 922-928.

[3] WU Z, SONG S, KHOSLA A, et al. 3D shapenets: A deep representation for volumetric shapes[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2015: 1912-1920.

[4] QI C R, SU H, MO K, et al. Pointnet: Deep learning on point sets for 3D classification and segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2017: 652-660.

[5] QI C R, YI L, SU H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[C]//Proceedings of Advances in Neural Information Processing Systems. Vancouver BC, Canada: [s.n.], 2017: 5099-5108.

[6] SU H, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional neural networks for 3D shape recognition[C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2015: 945-953.

[7] FENG Y, ZHANG Z, ZHAO X, et al. GVCNN: Group-view convolutional neural networks for 3D shape recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 264-272.

[8] WEI X, YU R, SUN J. View-GCN: View-based graph convolutional network for 3D shape analysis[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 1850-1859.

[9] LIU A A, NIE W Z, GAO Y, et al. View-based 3-D model retrieval: A benchmark[J]. IEEE Transactions on Cybernetics, 2017, 48(3): 916-928.

[10] ZHOU H, LIU A A, NIE W. Dual-level embedding alignment network for 2D image-based 3D object retrieval[C]//Proceedings of the 27th ACM International Conference on Multimedia. Nice, France: ACM, 2019: 1667-1675.

[11] LI W, LIU A, NIE W, et al. SHREC 2019-monocular image based 3D model retrieval[C]//Proceedings of Eurographics Workshop on 3D Object retrieval. [S.l.]: Eurographics Association, 2019: 103-110.

[12] LI B, LU Y, LI C, et al. SHREC'14 track: Extended large scale sketch-based 3D shape retrieval[C]//Proceedings of Eurographics Workshop on 3D Object retrieval. Strasbourg, France: Eurographics Association, 2014: 121-130.

[13] QI C R, SU H, NIEBNER M, et al. Volumetric and multi-view CNNs for object classification on 3D data[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2016: 5648-5656.

[14] LI Y, PIRK S, SU H, et al. FPNN: Field probing neural networks for 3D data[J]. Advances in Neural Information Processing Systems, 2016, 29: 307-315.

[15] KLOKOV R, LEMPITSKY V. Escape from cells: Deep kd-networks for the recognition of 3D point cloud models[C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2017: 863-872.

[16] BELONGIE S, MALIK J, PUZICHA J. Shape context: A new descriptor for shape matching and object recognition[J]. Advances in Neural Information Processing Systems, 2000, 13: 831-837.

[17] XIE S, LIU S, CHEN Z, et al. Attentional shape context net for point cloud recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 4606-4615.

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

[19] WANG C, PELILLO M, SIDDIQI K. Dominant set clustering and pooling for multi-view 3D object recognition [EB/OL]. (2019-06-04)[2020-12-10]. http://arXiv preprint arXiv:1906.01592.

[20] DAI G, XIE J, FANG Y. Siamese CNN-BiLSTM architecture for 3d shape representation learning[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. [S.l.]: AAAI, 2018: 670-676.

[21] MA C, GUO Y, YANG J, et al. Learning multi-view representation with LSTM for 3-D shape recognition and retrieval[J]. IEEE Transactions on Multimedia, 2018, 21(5): 1169-1182.

[22] CHEN S, ZHENG L, ZHANG Y, et al. VERAM: View-enhanced recurrent attention model for 3D shape classification[J]. IEEE Transactions on Visualization and Computer Graphics, 2018, 25(12): 3244-3257.

[23] SFIKAS K, PRATIKAKIS I, THEOHARIS T. Ensemble of panorama-based convolutional neural networks for 3D model classification and retrieval[J]. Computers & Graphics, 2018, 71: 208-218.

[24] LU K, HE N, XUE J, et al. Learning view-model joint relevance for 3D object retrieval[J]. IEEE Transactions on Image Processing, 2015, 24(5): 1449-1459.

[25] YOU H, FENG Y, ZHAO X, et al. PVRnet: Point-view relation neural network for 3D shape recognition[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2019, 33: 9119-9126.

[26] HEGDE V, ZADEH R. Fusionnet: 3D object classification using multiple data representations [EB/OL].(2016-11-27)[2020-12-10].http:// arXiv preprint arXiv:1607.05695.

[27] RAHMAN M M, TAN Y, XUE J, et al. 3D object detection: Learning 3D bounding boxes from scaled down 2D bounding boxes in RGB-D images[J]. Information Sciences, 2019, 476: 147-158.

[28] GANIN Y, LEMPITSKY V. Unsupervised domain adaptation by backpropagation[EB/OL]. (2015-02-27) [2020-12-10]. https://arxiv.org/abs/1409.7495.

[29] ZHU F, XIE J, FANG Y. Learning cross-domain neural networks for sketch-based 3D shape retrieval[C]// Proceedings of Thirtieth AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2016: 3683-3689.

[30] DARAS P, AXENOPOULOS A. A 3D shape retrieval framework supporting multimodal queries[J]. International Journal of Computer Vision, 2010, 89(2/3): 229-247.

[31] BRONSTEIN A M, BRONSTEIN M M, GUIBAS L J, et al. Shape google: Geometric words and expressions for invariant shape retrieval[J]. ACM Transactions on Graphics, 2011, 30(1): 1-20.

[32] EITZ M, RICHTER R, BOUBEKEUR T, et al. Sketch-based shape retrieval[J]. ACM Transactions on Graphics, 2012, 31 (4): 1-10.

[33] WANG J, YANG J, YU K, et al. Locality-constrained linear coding for image classification[C] // Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2010: 3360-3367.

[34] BIASOTTI S, CERRI A, AONO M, et al. Retrieval and classification methods for textured 3D models: A comparative study [J]. The Visual Computer, 2016, 32(2): 217-241.

[35] XIE J, DAI G, ZHU F, et al. Learning barycentric representations of 3D shapes for sketch-based 3D shape retrieval[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2017: 5068-5076.

[36] TASSE F P, DODGSON N. Shape2vec: Semantic-based descriptors for 3D shapes, sketches and images[J]. ACM Transactions on Graphics, 2016, 35(6): 1-12.

[37] FURUYA T, OHBUCHI R. Deep aggregation of local 3D geometric features for 3D model retrieval[C]// Proceedings of BMVC. [S.l.]: BMVC Press, 2016, 7: 8.

[38] KANEZAKI A, MATSUSHITA Y, NISHIDA Y. RotationNet: Joint object categorization and pose estimation using multiviews from unsupervised viewpoints[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 5010-5019.

[39] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[EB/OL]. (2017-02-27)[2020-12-10]. http://arXiv preprint arXiv:1609.02907.

[40] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[41] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL].(2015-04-10) [2020-12-10]. http:// arXiv preprint arXiv:1409.1556.

[42] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2016: 770-778.

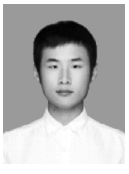
[43] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]// Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2009: 248-255.

[44] WANG J, FENG W, CHEN Y, et al. Visual domain adaptation with manifold embedded distribution alignment[C]//

Proceedings of the 26th ACM International Conference on Multimedia. Seoul, Korea: ACM, 2018: 402-410.
[45] LONG M, ZHU H, WANG J, et al. Deep transfer learning with joint adaptation networks[C]//International Conference on Machine Learning. [S.l.]: PMLR, 2017: 2208-2217.
[46] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks[J]. The Journal of Machine Learning Research, 2016, 17(1): 2096-2030.
[47] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]// NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge, MA, USA: MIT Press, 2014: 2672-2680.
[48] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. (2018-02-04)[2020-12-10]. http://arXiv preprint arXiv:1710.10903.
[49] XIE S, ZHENG Z, CHEN L, et al. Learning semantic representations for unsupervised domain adaptation[C]// Proceedings of International Conference on Machine Learning. Stockholm, Sweden: [s.n.], 2018: 5423-5432.
[50] DAI W, LIANG S. Cross-modal guidance network for sketch-based 3D shape retrieval[C]//Proceedings of 2020 IEEE International Conference on Multimedia and Expo (ICME). [S.l.]: IEEE, 2020: 1-6.
[51] GONG B, LIU J, WANG X, et al. Learning semantic signatures for 3D object retrieval[J]. IEEE Transactions on Multimedia, 2012, 15(2): 369-377.
[52] FURUYA T, OHBUCHI R. Ranking on cross-domain manifold for sketch-based 3D model retrieval[C]// Proceedings of 2013 International Conference on Cyberworlds. [S.l.]: IEEE, 2013: 274-281.
[53] LI B, LU Y, GODIL A, et al. A comparison of methods for sketch-based 3D shape retrieval[J]. Computer Vision and Image Understanding, 2014, 119: 57-80.
[54] WANG F, KANG L, LI Y. Sketch-based 3D shape retrieval using convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2015: 1875-1883.
[55] DAI G, XIE J, ZHU F, et al. Deep correlated metric learning for sketch-based 3D shape retrieval[C]// Proceedings of Thirty-First AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2017.
[56] DAI G, XIE J, FANG Y. Deep correlated holistic metric learning for sketch-based 3D shape retrieval[J]. IEEE Transactions on Image Processing, 2018, 27(7): 3374-3386.
[57] CHOPRA S, HADSELL R, LECUN Y. Learning a similarity metric discriminatively, with application to face verification [C]// Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). [S.l.]: IEEE, 2005, 1: 539-546.
[58] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2015: 815-823.
[59] SAVVA M, YU F, SU H, et al. Shrec16 track: Largescale 3D shape retrieval from Shapenet core55[C]//Proceedings of the Eurographics Workshop on 3D Object Retrieval. Lisbon, Portugal: [s.n.], 2016.
[60] FURUYA T, OHBUCHI R. Dense sampling and fast encoding for 3D model retrieval using bag-of-visual features[C]// Proceedings of the ACM International Conference on Image and Video Retrieval. New York, NY, USA: ACM, 2009: 1-8.

作者简介:

刘安安(1982-),男,教授,博士生导师,研究方向:计算机视觉、机器学习、三维模型检索,E-mail:liuanan-tju@163.com。



李天宝(1996-),男,硕士研究生,研究方向:三维模型检索。

王晓雯(2000-),女,本科,研究方向:机器学习。

宋丹(1992-),**通信作者**,女,讲师,硕士生导师,研究方向:计算机图形学、三维模型检索,E-mail:dan.song@tju.edu.cn。

(编辑:张黄群)