

# 基于卷积神经网络的高效三维模型检索方法

汤磊, 丁博, 何勇军

(哈尔滨理工大学计算机科学与技术学院, 黑龙江哈尔滨 150080)

**摘要:** 目前基于视图的三维模型检索已经成为一个研究热点. 该方法首先将三维模型表示为二维视图的集合, 然后采用深度学习技术进行分类和检索. 但是现有的方法在精度和效率方面都有待提升. 本文提出了一种新的三维模型检索方法, 该方法包括索引建立和模型检索. 在索引建立阶段, 选择代表性视图输入到训练好的卷积神经网络 (Convolutional Neural Network, CNN) 中以提取特征和分类, 进而根据特征类别对特征进行组织以建立索引在检索阶段, 使用 CNN 和投票算法将输入模型的代表性视图分类为一个类别, 然后仅选择这个类别的特征而不是所有类别的特征进行相似度匹配, 因此减少了搜索空间. 此外, 随着用于检索的视图数量的逐渐增加, 一旦可以确定三维模型, 检索过程将提前终止. 实验的数据选用刚性三维模型数据集 ModelNet10, ModelNet40 和非刚性三维模型数据集 McGill10. 结果表明, 该方法在提升检索效率的同时, 确保检索准确率分别高达 94%、92% 和 100%.

**关键词:** 模型检索; 模型分类; 代表性视图; 卷积神经网络

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 0372-2112 (2021)01-0064-08

**电子学报 URL:** <http://www.ejournal.org.cn> **DOI:** 10.12263/DZXB.20191268

## An Efficient 3D Model Retrieval Method Based on Convolutional Neural Network

TANG Lei, DING Bo, HE Yong-jun

(Computer Science and Technology College, Harbin University of Science and Technology, Harbin, Heilongjiang 150080, China)

**Abstract:** Recently, 3D model retrieval based on views has become a research hotspot. In this method, 3D models are represented as a collection of 2D views, which allows deep learning techniques to be used for 3D model classification and retrieval. However, current methods need improvements on both accuracy and efficiency. We propose a 3D model retrieval method, which includes index building and model retrieval. In the index building stage, representative views are selected and input into a well-learned Convolutional Neural Network (CNN) for feature extraction and classification. Next, the features are organized according to their labels to build indexes. In the retrieval stage, the representative views of the input model are classified into a category with the CNN and voting algorithm, and then only the features of one category rather than all categories are chosen to perform similarity matching. In this way, the searching space for retrieval is reduced. In addition, the number of the used views for retrieval is gradually increased. Once there is enough evidence to determine a 3D model, the retrieval process will be terminated ahead of time. Experiments on the rigid 3D model datasets ModelNet10, ModelNet40, and the non-rigid 3D model dataset McGill10 show that the proposed method can improve the retrieval efficiency substantially while keeping high retrieval accuracy rates at 94%, 92% and 100%, respectively.

**Key words:** model retrieval; model classification; representative views; convolutional neural network

### 1 引言

近年来, 三维模型被广泛用于计算机辅助设计 (Computer Aided Design, CAD)、虚拟现实、3D 动画和影

视、医学诊断、3D 网络游戏、机械制造等领域. 模型设计是支撑这些应用的关键环节, 耗费了大量的人力物力. 根据产品之间的相似性, 最大限度地重用已有的设计资源来开发新产品, 不仅可以降低产品成本, 缩短设计

收稿日期: 2019-11-13; 修回日期: 2020-05-18; 责任编辑: 李勇锋

基金项目: 国家自然科学基金面上项目 (No. 61673142); 黑龙江省自然科学基金杰出青年项目 (No. JJ2019JQ0013); 黑龙江省普通本科高等学校青年创新人才项目 (No. UNPYSCT-2016034); 哈尔滨市杰出青年人才基金 (No. 2017RAYXJ013)

时间,而且还可以有效的提高产品可靠性,保证产品质量<sup>[1]</sup>.尤其是近年来,三维模型数量的飞速增长为模型增量设计提供了大量素材.三维模型数量的剧增带来了一个现实的问题,那就是如何在海量模型中快速搜索到需要的模型,这正是三维模型检索要解决的问题.

三维模型检索分为索引建立和模型检索两步.近年来,基于二维视图的索引建立是一个研究热点.这类方法从不同的角度对三维模型投影,得到一组二维视图.然后提取二维视图的特征表示来建立索引,达到了将三维模型检索转换成二维视图检索的目的.目前基于视图的三维模型检索方法的拍摄位置均匀分布在模型周围,产生了大量的投影视图,导致冗余信息多,严重影响了模型检索的精度和速度.为了提高视图的代表性,本文提出一种基于 K-means 的视图选择方法,以提高代表性视图的表示能力、减少投影视图的数量,进而提高检索效率.

选取到代表性视图后需要提取特征以建立索引.卷积神经网络(Convolutional Neural Network, CNN)是一种深度学习网络,利用卷积核来自动提取特征,并且所提取的特征具有平移、旋转和缩放不变性,适合二维视图的特征提取.基于 CNN 的检索方法能有效提升对视图的识别精度.因此,本文利用训练好的 CNN 提取代表性视图的特征,并将特征按类别建立索引,为提高模型的精度和效率提供了保证.

在模型检索方面,对三维模型投影、代表性视图选取和提取特征以后,需要与模型库中所有模型的索引进行相似度匹配,这种方式严重降低了检索效率.因此,我们采用先分类再检索的方式,减少了检索算法的搜索空间.与此同时,提出了一种基于视图递增的相似度匹配方法.该方法对于待检索模型,依次选择代表性视图进行相似度匹配,当可以确定所需模型时提前终止检索,从而进一步提高了检索效率.

## 2 相关工作

三维模型的检索方法有多种,其中基于视图的三维模型检索是一个研究热点,该方法概括起来主要分为基于低层次特征的检索方法和基于深度特征的检索方法.

### 2.1 基于低层次特征的检索方法

对于刚性三维模型,低层次特征通常是指提取二维视图的 Zernike 矩<sup>[3]</sup>或方向梯度直方图<sup>[4]</sup>等特征,并将其表示为一组多维向量.对于非刚性三维模型,热核特征(Heat Kernel Signature, HKS)是一种理想的描述符,它具有对模型的小扰动鲁棒性、等距变换的有效性和不变性等特征<sup>[5]</sup>.Aubry 等将波动方程引入到模型分析中,提出波核特征(Wave Kernel Signature, WKS)<sup>[6]</sup>.

单一特征难以完整的表征三维模型,Zeng 等将多种模型特征融合在一起,通过使用融合特征,提高模型检索准确率<sup>[7]</sup>.基于低层次特征的检索方法在提取特征时容易造成细节信息的丢失,降低了检索准确率.

### 2.2 基于深度特征的检索方法

基于深度特征的检索方法将三维模型表征为一组二维视图的集合,并以此为基础构建深度学习模型完成深度特征提取及模型分类.典型的投影方法有光场描述符(Light Field Descriptor, LFD)<sup>[8]</sup>.此方法采用均匀投影的方式,最后采用 100 幅视图表征三维模型.LFD 会产生大量冗余视图,不仅需要大量的存储空间,还降低了检索速度.针对该问题,多位学者对 LFD 进行了改进,如基于多视图的 Su-MVCNN<sup>[9]</sup>, Wang-MVCNN<sup>[10]</sup>, VS-MVCNN<sup>[11]</sup>,以及基于全景视图<sup>[12,13]</sup>的三维模型表征方法.LFD 除了会产生大量冗余视图之外,还存在另一个问题,就是笼统地对所有二维视图同等对待,忽略了不同的视图对表示三维模型的重要性不同.因此,出现了提取代表性视图的方法,包括加权二部图匹配(Weighted Bipartite Graph Matching, WBG)<sup>[14]</sup>和 Apriori 算法<sup>[15]</sup>等.白静等人首先采用深度学习模型 CaffeNet 作为弱分类器分类二维视图,然后利用加权投票方式作为强分类器完成三维模型分类<sup>[16]</sup>.该方法通过强分类器统计了视图属于各个类的概率,在刚性和非刚性数据库中检索效率均有所提高.

## 3 本文方法

### 3.1 总体框架

如图 1 所示,本文提出的三维模型检索方法共分为 3 步:(1)CNN 训练;(2)二维代表性视图选取和索引建立;(3)模型检索.在 CNN 训练阶段,首先将三维模型转换为二维视图,然后通过二维视图训练 CNN.在二维代表性视图选取和索引建立阶段,首先将三维模型转换为二维视图,然后采用 K-means 算法选择出二维代表性视图,最后将二维代表性视图作为训练好的 CNN 的输入,进行特征提取和索引建立.在模型检索阶段,输入可以是图片或是模型.如果是图片可以直接进行检索.如果是模型,则首先要产生二维代表性视图,然后通过代表性视图进行检索.本文的检索方法,采用先分类再检索的方式,每一个视图仅在一个类别中进行相似度匹配,并且提出了一种基于视图递增的相似度匹配方法,能有效提高检索效率.

### 3.2 CNN 训练

近年来,陆续出现了一系列高性能的卷积神经网络,如 VGG, GoogleNet, ResNet, DenseNet 等,其中 ResNet 和 DenseNet 相比于传统深度神经网络有较大创新,并且在 imageNet 数据集上分类表现突出.两者比较而言,

DenseNet 在训练数据较小时存在优势,而本文采用的 ModelNet 数据集模型数量较多,训练数据较大,因此选用 ResNet 网络.而 ResNet 网络又有 ResNet18, ResNet50, ResNet101 等不同规模.综合考虑三维模型检索的精度与效率,本文选择 ResNet50<sup>[18]</sup> 网络进行特征提取和分类.

### 3.3 索引建立

建立有效的索引对模型检索的效率提升至关重要.本文将二维视图采用 K-means 进行聚类,从而选取代表性视图.然后通过训练好的 CNN 为代表性视图建立索引.最终的目的在于提高模型检索的效率.

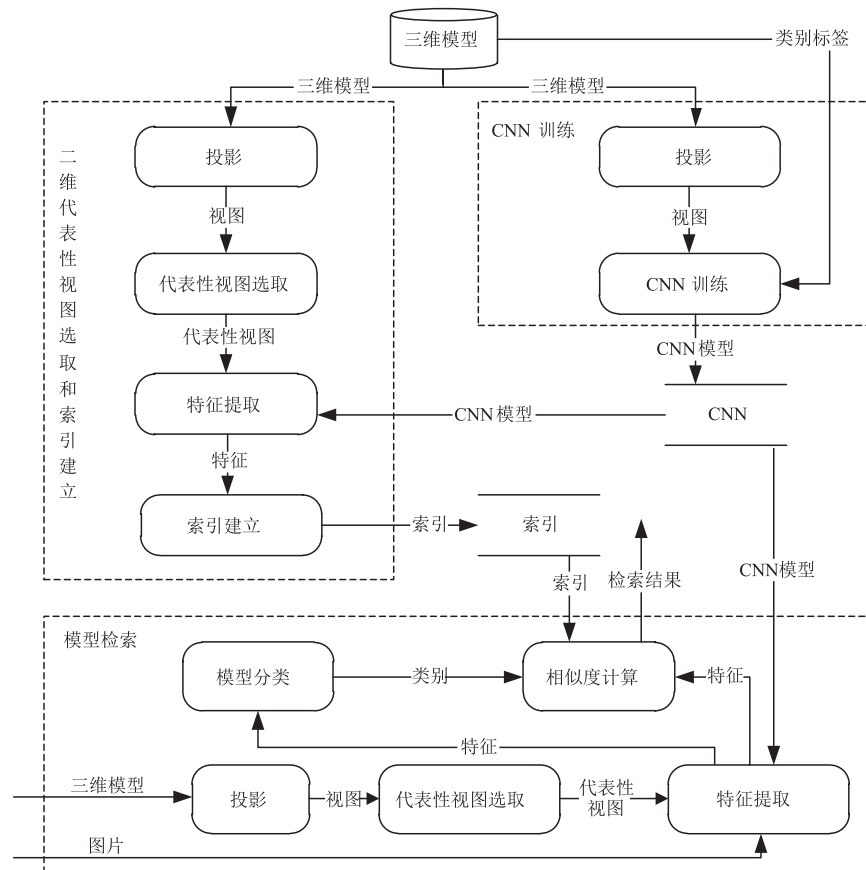


图1 检索过程

#### 3.3.1 基于 K-means 的代表性视图选取

二维视图的数量和投影角度对三维模型的有效表示具有重要影响.如果二维视图过多会增加检索的时间复杂度.本文提出了一种选取二维代表性视图的方法.该方法采用 K-means 算法对视图进行聚类,然后在每一类中选择一个代表性视图.对于不同的三维模型,通过 K-means 将得到不同数量的二维代表性视图.本文采用 K-means 获取代表性视图的方法如下:

**步骤 1** 采用改进的 LFD 的投影方法对每个三维模型产生 40 个二维视图.

**步骤 2** 用 K-means 算法对每个模型的视图进行聚类.当用 K-means 对二维视图进行聚类时,必须首先确定类别  $K$  的值.实验表明,10~20 个视图的匹配效果最好.因此, $K$  的取值范围大致为 10~20,然后用手肘法<sup>[19]</sup> 确定  $K$  的最终值.

**步骤 3** 选择最接近每个类中心的视图作为这个三维模型的代表性视图.将三维模型的二维视图划分为  $K$  个类别,在每个类别中通过欧式距离计算视图与该类中心的距离,取距离最小的视图为该类的代表性视图,最终得到  $K$  张二维视图.

#### 3.3.2 基于 CNN 的索引建立

该阶段首先通过本文提出的投影方法和视图选取方法生成三维模型的代表性视图,然后将代表性视图输入到训练好的 ResNet50 中,最后将输出的特征根据类别建立索引.全连接层输出的是代表性视图的深度特征. Softmax 层输出的是分类结果.索引建立过程如图 2 所示.

### 3.4 模型检索

模型检索的任务是根据输入,在模型库中找到相似或相同的模型.输入可以是图片或者三维模型.如果

输入是图片,则通过训练好的 CNN 确定类别并提取特征. 在一个类别中,通过下式得到检索结果:

$$I = \arg \min_{i,j} \text{dis}(\mathbf{W}, \mathbf{F}_{ij}) \quad (1)$$

$I$  表示检索结果,  $\text{dis}()$  表示计算欧氏距离,  $\mathbf{W}$  是输入图片的特征值,  $\mathbf{F}_{ij}$  是第  $i$  个模型第  $j$  个视图的特征值,  $1 \leq i \leq m$ ,  $m$  是一个类别中的模型个数,  $1 \leq j \leq n_i$ ,  $n_i$  为第  $i$  个模型的代表性视图个数, 模型  $i$  是检索到的结果.

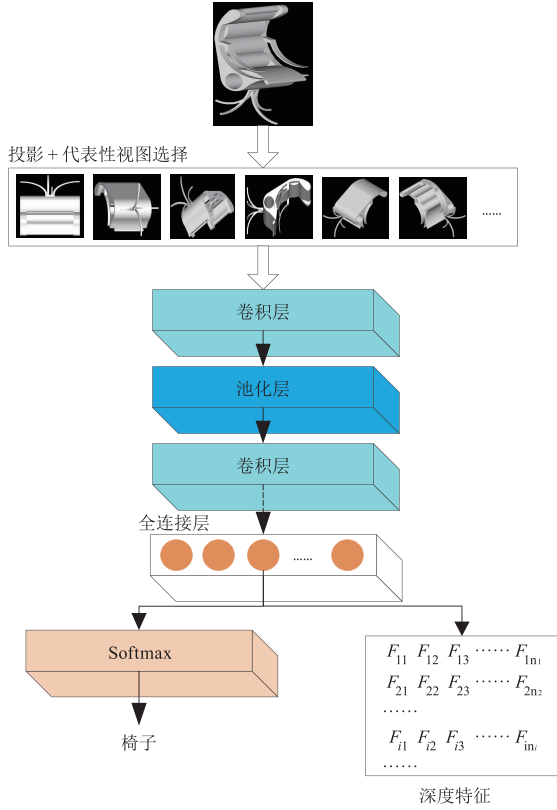


图2 索引建立过程

如果输入是三维模型,检索分为 3 步:(1)产生二维代表性视图;(2)将视图输入到 CNN 提取特征并分类. 考虑到分类器存在误差,导致输入模型的各个视图可能无法完全分类到同一个类别中,本文采用投票策略确定待检索模型的类别;(3)在确定的类别中进行相似度匹配. 为了提高匹配效率,本文提出了一种视图递增的相似度匹配方法.

$Category\_Vector$  表示类别向量,其中第  $c$  个元素表示被分到第  $c$  个类别中的视图数量.  $Category\_Vector$  初始化如下:

$$Category\_Vector = [0, 0, \dots, 0] \quad (2)$$

其中  $Category\_Vector$  为  $c$  维向量,表示模型库中的  $c$  个类别.

当一个代表性视图被分配到第  $c$  类时,这个向量更新如下:

$$Category\_Vector[c] = Category\_Vector[c] + 1 \quad (3)$$

最后,确定模型的类别如下:

$$C = \arg \max_c Category\_Vector[c] \quad (4)$$

$C$  表示分类结果. 若  $C$  不能唯一确定,则在剩下的视图中随机选择继续进行投票,直至  $C$  唯一确定.

通过投票策略将模型准确分类后,在同一类别中采用算法 1 实现检索. 为了提高检索效率,我们设计了灵活的检索策略:(1)如果一个输入视图和一个模型视图之间的距离足够小,也就是  $dis < \eta$  时,就可以确定这个模型是需要模型;(2)当依次选取的代表性视图达到 5 张,并且这 5 张视图都被分类到同一类的同一个模型中时,可以确定这个模型是所需模型;(3)如果代表性视图被匹配到同一类的不同模型中时,则计算累积距离值. 如果一个模型的累积距离值最小,则该模型就是所需模型.

#### 算法 1 检索策略

输入:代表性视图的特征值  $\mathbf{W}_l, l \in \{1, 2, 3, \dots, K\}$  ( $K$  表示模型的代表性视图数量)

$\mathbf{F}_{ij}$  表示第  $i$  个模型的第  $j$  个视图的特征,

$m$  表示类别中的模型数量,

$n_i$  表示第  $i$  个模型的代表性视图的数量,

$\eta$  表示最小距离值,

距离向量  $Distance\_Vector = [0, 0, \dots, 0]$ ,

计数向量  $count = [0, 0, \dots, 0]$ ,记录分类到每模型中的视图数量.

输出:最佳模型  $i_{searched}$

$i_{searched} = -1$ ;

$\eta = 1.5$ ;

for ( $1 \leq l \leq K$ )

{

$k_{min} = 0$ ;

$dis_{min} = 1000000$ ;

for ( $1 \leq i \leq m$ )

{

$dis = \min_j \text{dis}(\mathbf{W}_l, \mathbf{F}_{ij}) (j = 1, 2, \dots, n_i)$ ;

if ( $dis < \eta$ ) {  $i_{searched} = i$ ; return; }

$Distance\_Vector[i] = Distance\_Vector[i] + dis$ ;

if ( $dis < dis_{min}$ ) {  $k_{min} = i$ ;  $dis_{min} = dis$ ; }

}

$count(k_{min}) = count(k_{min}) + 1$ ;

for ( $1 \leq i \leq m$ )

{

if ( $count(i) = 5$ ) {  $i_{searched} = i$ ; return; }

}

if ( $i_{searched} = -1$ )  $i = \arg \min_i Distance\_Vector[i]$ ;

## 4 实验结果与分析

实验基于 MXNET 框架,使用 ResNet50 完成视图特征提取、索引建立及分类. 在 Intel i5 8400 + GTX 1060

的 PC 机上测试. 下面将分别从分类和检索两个方面对本文方法进行测评.

#### 4.1 分类实验

分类实验中, 本文选用了刚性数据集和非刚性数据集两种. 选取的刚性数据集为 ModelNet10 和 ModelNet40<sup>[20]</sup>, 非刚性数据集为 McGill<sup>[21]</sup>. 其中 ModelNet10 和 ModelNet40 分别有 4899 和 12311 个模型, 并且训练数据和测试数据分离已经分离, 分别使用 908 和 2468 个模型作为测试集, 3991 和 9843 个模型作为训练集. McGill 共有 255 个模型, 其中没有分离测试数据和训练数据, 本文按照测试集和训练集为 1:9、2:8、3:7 和 4:6 的比例随机划分.

##### 4.1.1 代表性视图选择

在本文的投影系统下, 每个三维模型可以得到 40 张二维视图. 为了提高分类和检索效率, 将投影得到的视图进行聚类, 用聚类后的代表性视图进行模型分类. 由于分类的准确性对检索影响比较大, 而不同的  $K$  值又会导致不同的分类准确率, 本文针对不同的  $K$  值对 McGill、ModelNet10 和 ModelNet40 进行实验, 不同  $K$  值下分类错误模型统计如表 1 所示.

表 1 不同  $K$  值下各数据集分类错误模型统计

$K$	5	10	20	30
McGill	0	0	0	0
ModelNet10	77	62	56	55
ModelNet40	212	191	182	183

从表 1 中可以看出,  $K$  值的选取对非刚性模型 McGill 数据集的影响不大. 而对刚性模型 ModelNet 而言,  $K$  值越大, 错分的模型就越少. 当  $K$  取 5 时, 错分模型最多, 而  $K$  取 20 或 30 时, 错分模型数量相差甚小. 由于视图数量越多, 分类和检索的效率越低. 因此, 从综合性能上考虑, 将  $K$  的取值范围设置为  $[10, 20]$ , 在区间内选择合适的  $K$  值进行分类和检索.

本文方法的目的在于在确保检索准确率的前提下, 尽可能地减少视图数目, 从而提高检索效率. 分析表 2 可以得出, 从模型分类准确率上看, 聚类对 McGill 数据集和 ModelNet10 数据集的模型分类没有影响, ModelNet40 聚类后模型分类准确率仅下降 0.9%. 从视图数量上看, 聚类可以大幅度降低视图的数量, 大约降为原来的三分之一. 由于视图数量越少, 分类和检索的

表 2 聚类前后测试集模型分类准确率

	McGill		ModelNet10		ModelNet40	
	聚类前	聚类后	聚类前	聚类后	聚类前	聚类后
视图数	960	362	36320	12742	98720	34526
准确率	100%	100%	94.10%	94.10%	92.90%	92.00%

效率越高, 下文中的实验默认采取的是聚类后数据, 即每个模型约 14 张二维视图.

##### 4.1.2 基于视图的分类方法对比

本文采用基于视图的三维模型分类方法, 该方法首先对模型的二维视图进行分类, 然后根据二维视图的分类结果进行投票, 最终得到模型类别. 在 CNN 训练阶段, 为了增加鲁棒性, 使用所有的二维视图进行训练. 在检索阶段为了提高效率, 采用代表性视图进行分类和相似度匹配. 在刚性数据集中, 表 3 是本文方法与其他基于视图的分类方法的准确率比较结果.

表 3 刚性三维模型分类准确率

方法	视图数	ModelNet10	ModelNet40
DeepPano <sup>[13]</sup>	1	88.66%	82.54%
PANORAMA-NN <sup>[12]</sup>	1	91.10%	90.70%
3D2SeqViews <sup>[22]</sup>	12	94.68%	91.64%
CNN-VOTE <sup>[16]</sup>	12	93.18%	91.82%
FusionNet <sup>[23]</sup>	60	93.11%	90.80%
VS-MVCNN <sup>[11]</sup>	80	93.50%	90.90%
本文	14	94.10%	92.00%

可以看出, 本文的分类方法在 ModelNet40 数据集上的准确率比其他算法都要好. 在 ModelNet10 数据集上准确率居于第二位, 比表中第一仅低 0.58%. 本文的检索方法是在分类的基础上进行检索, 因此分类的准确率会对最终的检索结果有所影响. 为了尽量减少模型被错分, 本文采用投票机制, 极大降低了因二维视图错分对模型分类产生的影响, 从而保证了检索准确率.

本文提出的分类算法能用较少的视图得到较高的分类准确率. 尤其是对非刚性数据集 McGill, 分类准确率达到 100%. 原因是非刚性模型类别间差距较大, 多视图更能体现其综合信息, 因此分类效果好. 但是对于某些刚性模型, 如 table 和 desk, 或者是 night\_stand 和 dresser 之间很容易混淆. 原因是其类别间的差距较小, 因此导致分类准确度有所下降.

#### 4.2 检索实验

本文的检索实验是基于分类实验之上的, 先分类再检索, 这样做的目的是使检索在同一类别中进行, 从而减小检索空间. 同时, 由于本文提出的分类方法分类准确率较高, 因此, 为三维模型的检索性能提供了保证. 下面从 ModelNet10、ModelNet40 和 McGill 对算法进行评价, 并对方法的准确率进行分析.

##### 4.2.1 刚性模型的检索实验

本文采用欧式距离计算二维视图间的相似度. 分别在 ModelNet10 和 ModelNet40 数据集上与 SPH<sup>[24]</sup>、

LFD<sup>[8]</sup>、3DShapeNets<sup>[25]</sup>、DeepPano<sup>[13]</sup> 和 PANORAMA-NN<sup>[12]</sup>检索算法进行对比. 采用平均准确率均值 (Mean Average Precision, MAP) 作为评价指标, 其结果如表 4 所示.

表 4 不同方法 MAP 对比

方法	数据集	
	ModelNet10	ModelNet40
SPH <sup>[24]</sup>	45.9%	34.4%
LFD <sup>[8]</sup>	49.8%	40.9%
3DShapeNets <sup>[25]</sup>	69.2%	59.9%
DeepPano <sup>[13]</sup>	84.2%	76.8%
PANORAMA-NN <sup>[12]</sup>	87.4%	83.5%
本文	94.1%	92.0%

我们方法的 MAP 值明显高于其他方法, 原因在于: (1) 采用先分类再检索的策略. 由于本文提出的分类方法, 分类准确率较高, 为检索准确率提供了保证; (2) 采用投票机制. 在刚性数据库中, 部分模型由于相似度高, 视图分类阶段容易出现错误, 但是通过投票机制可以使部分视图被错分的模型被正确分类. 因此, 本文提出的检索方法 MAP 居于第一位.

从图 3 和图 4 中可以看出, 本文的检索方法 P-R 曲线趋于平稳, 而其他算法的 P-R 曲线都是随着查全率的升高而查准率逐渐降低.

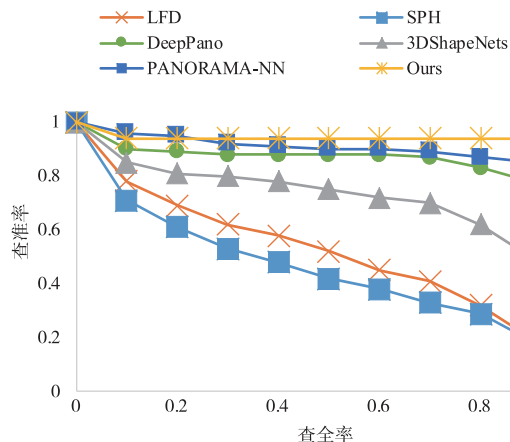


图3 ModelNet10 P-R曲线

#### 4.2.2 非刚性模型的检索实验

以最近邻方法 (Nearest Neighbor, NN)、第一层级 (First Tier, FT)、第二层级 (Second Tier, ST)、折扣的累积结果 (Discounted Cumulative Gain, DCG) 为评价指标对 McGill 的检索效果进行评价. 主要对比 HKS<sup>[5]</sup>、WKS<sup>[6]</sup>、CBoFHKS<sup>[26]</sup>、DASD<sup>[27]</sup> 和 MFFL<sup>[7]</sup>, 结果如表 5 所示.

从表 5 结果可以看出, 本文方法对非刚性模型的检索效果非常突出, 比刚性模型的检索效果还要好. 主

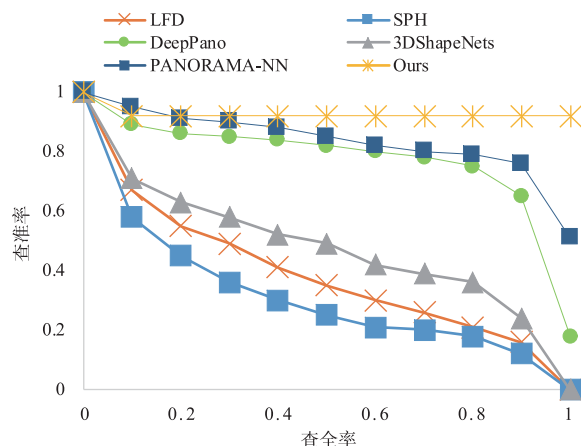


图4 ModelNet40 P-R曲线

要原因是非刚性模型分类准确率为 100%, 进而提升了检索准确率, 因此方法的各项指标值均为 1. 通过实验可以看出, 本文方法不仅适用于刚性模型和也适用于非刚性模型.

表 5 McGill 数据集评价指标

方法	评价标准			
	NN	FT	ST	DCG
HKS <sup>[5]</sup>	0.8190	0.6220	0.7440	0.8270
WKS <sup>[6]</sup>	0.9140	0.7750	0.8660	0.9140
CBoFHKS <sup>[26]</sup>	0.9010	0.7780	0.8760	0.8910
DASD <sup>[27]</sup>	0.9880	0.7820	0.8340	0.9550
MFFL <sup>[7]</sup>	0.9710	0.9050	0.9810	0.9630
本文	1.0000	1.0000	1.0000	1.0000

#### 4.2.3 检索效率分析

通过实验发现, 本文的时间消耗主要在视图与视图间的距离计算上. 以 ModelNet10 为例, 测试集中包含 908 个模型, 训练集中包含 3991 个模型. 每个模型产生 40 张视图, 那么测试集中包含 36320 张视图, 训练集中包含 159640 张视图. 如果直接用全部的视图进行相似度计算, 时间消耗巨大. 为了解决该问题, 本文提出了基于 K-means 的代表性视图选取方法和视图递增的匹配策略. 通过聚类, 每个模型的视图数减少到 14 张, 测试集的视图数由 36320 减少到 12742, 训练集的视图数由 159640 减少到 56613, 二者的视图数量约为聚类前的三分之一. 因此, 聚类可以有效减少冗余视图, 进而提高检索效率.

ModelNet10 训练集模型分为 10 类, 共 3991 个模型, 平均每类有 399 个模型. 在代表性视图进行相似度匹配后, 匹配次数从 638400 ( $40 \times 399 \times 40$ ) 减少到 78204 ( $14 \times 399 \times 14$ ), 即降低了 87.75%.

视图递增策略在聚类的基础上进一步提升了匹配

效率. 其中,  $\eta$  表示同一类别中模型间的相似程度. 不同数据集和不同投影点的选取都会影响  $\eta$  的取值. 本文的相似度匹配算法是在分类的基础上进行, 因此  $\eta$  的值对检索 P-R 曲线影响不大, 但就检索精度来讲,  $\eta$  越小精度越高. 本文取 1.5 为最终  $\eta$  值.

在  $\eta$  取 1.5 的情况下, 以 ModelNet10 为例, 采用本文算法 1, 平均每个模型的检索视图减少为 11 张. 因此, 每个模型进行相似度匹配的次数为 61446 ( $11 \times 399 \times 14$ ). 与仅采用代表性视图选择方法相比, 相似度匹配次数进一步减少了 21.43%.

## 5 结语

随着三维模型规模的不断增加, 检索精度和效率成为三维模型检索系统面临的最大挑战. 本文提出了一种高效的三维模型检索方法, 在索引建立和模型检索两个方面提升了检索性能. 首先通过聚类选取代表性视图, 通过建立高效的索引, 提高检索效率. 然后采用先分类再检索的检索方法, 减少模型匹配的搜索空间. 由于本文提出的分类方法准确率高, 因此确保了检索准确率. 最后采用视图递增的方法进行模型检索, 可以减少不必要的匹配运算, 进一步提高了检索效率. 由于用于表示模型的视图有足够的代表性, 使得所提出的方法在提高效率的同时, 也提高了检索精度. 实验表明, 本文方法在效率和精度上都优于目前方法, 同时适用于刚性三维模型和非刚性三维模型分类及检索.

## 参考文献

- [1] 李荫民, 薛凯心, 高赞. 基于残差网络的三维模型检索算法[J]. 计算机科学, 2019, 46(3): 148 - 153.  
LI Yin-min, XUE Kai-xin, GAO Zan. 3-D model retrieval algorithm based on residual network [J]. Computer Science, 2019, 46(3): 148 - 153. (in Chinese)
- [2] 杨军, 王亦民. 基于深度卷积神经网络的三维模型识别[J]. 重庆邮电大学学报(自然科学版), 2019, 31(2): 254 - 261.  
YANG Jun, WANG Yi-min. 3D model recognition based on depth convolution neural network [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2019, 31(2): 254 - 261. (in Chinese)
- [3] Qin F W, Li L Y, Gao S M, et al. A deep learning approach to the classification of 3D CAD models [J]. Journal of Zhejiang University: Science C, 2014, 15(2): 91 - 106.
- [4] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C]. San Diego, CA, USA: IEEE, 2005. 886 - 893.
- [5] Sun J, Ovsjanikov M, Guibas L. A concise and provably informative multi-scale signature based on heat diffusion [J]. Computer Graphics Forum, 2009, 28(5): 1383 - 1392.
- [6] Aubry M, Schlickewei U, Cremers D. The wave kernel signature: A quantum mechanical approach to shape analysis [A]. Proceedings of IEEE International Conference on Computer Vision Workshops [C]. Barcelona, Spain: IEEE, 2011. 1626 - 1633.
- [7] Zeng H, Liu Y, Li S, et al. Convolutional neural network based multi-feature fusion for non-rigid 3D model retrieval [J]. Journal of Information Processing Systems, 2018, 14(1): 176 - 190.
- [8] Chen D Y, Tian X P, Shen Y T, et al. On visual similarity based on 3D model retrieval [J]. Computer Graphics Forum, 2003, 22(3): 223 - 232.
- [9] Su H, Maji S, Kalogerakis E, et al. Multi-view convolutional neural networks for 3D shape recognition [A]. Proceedings of the IEEE International Conference on Computer Vision [C]. Santiago, Chile: IEEE, 2015. 945 - 953.
- [10] Wang C, Pelillo M, Siddiqi K, et al. Dominant set clustering and pooling for multi-view 3D object recognition [A]. Proceedings of the British Machine Vision Conference [C]. Heidelberg: Springer, 2017. 1 - 11.
- [11] Ma Y X, Zheng B, Guo Y L, et al. Boosting multi-view convolutional neural networks for 3D object recognition via view saliency [A]. Proceedings of the Chinese Conference on Image and Graphics Technologies [C]. Singapore: Springer, 2017. 199 - 209.
- [12] Sfikas K, Theoharis T, Pratikakis I. Exploiting the PANORAMA representation for convolutional neural network classification and retrieval [A]. Proceedings of Eurographics Workshop on 3D Object Retrieval [C]. Lyon, France: The Eurographics Association, 2017. 1 - 7.
- [13] 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.
- [14] Gao Y, Dai Q, Wang M, et al. 3D model retrieval using weighted bipartite graph matching [J]. Signal Processing: Image Communication, 2011, 26(1): 39 - 47.
- [15] 石民, 张树生, 李亮, 白晓亮. 基于二维典型视图的三维CAD模型检索算法[J]. 制造业自动化, 2012, 34(09): 82 - 84, 126.  
SHI Min, ZHANG Shu-sheng, LI Liang, BAI Xiao-liang. 3D CAD model retrieval using 2D characteristic views [J]. Manufacturing Automation, 2012, 34(09): 82 - 84, 126. (in Chinese)
- [16] 白静, 司庆龙, 秦飞巍. 基于卷积神经网络和投票机制的三维模型分类与检索[J]. 计算机辅助设计与图形学

- 学报,2019,31(2):303-314.
- BAI Jing, SI Qing-long, QIN Fei-wei. 3D model classification and retrieval based on CNN and voting scheme [J]. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(2):303-314. (in Chinese)
- [17] Phong B T. Illumination for computer generated pictures[J]. Communications of the ACM, 1975, 18(6):311-317.
- [18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep residual learning for image recognition [A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Las Vegas, NV, USA; IEEE, 2016. 770-778.
- [19] Bholowalia P, Kumar A. EBK-means: A clustering technique based on elbow method and k-means in WSN[J]. International Journal of Computer Applications, 2014, 105(9):17-24.
- [20] Shilane P, Min P, Kazhdan M, et al. The princeton shape benchmark [A]. Proceedings of the Shape Modeling Applications [C]. Genova, Italy; IEEE, 2004. 167-178.
- [21] Siddiqi K, Zhang J, Macrini D, et al. Retrieving articulated 3D models using medial surfaces[J]. Machine Vision and Applications, 2008, 19(4):261-275.
- [22] Han Z, Lu H, Liu Z, et al. 3D2SeqViews: Aggregating sequential views for 3D global feature learning by CNN with hierarchical attention aggregation [J]. IEEE Transactions on Image Processing, 2019, 28(3):3986-3999.
- [23] Hegde V, Zadeh R. Fusionnet: 3D object classification using multiple data representations [J]. arXiv Preprint, 2016, arXiv:1607.05695.
- [24] Kazhdan M, Funkhouser T, Rusinkiewicz S. Rotation invariant spherical harmonic representation of 3D shape descriptors [A]. Proceedings of Eurographics Symposium on Geometry Processing [C]. Granada, Spain; The Eurographics Association, 2003. 156-165.
- [25] Wu Z, Song S, Khosla A, et al. 3D shapenets: a deep representation for volumetric shapes [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Boston, MA, USA; IEEE, 2015. 1912-1920.
- [26] Lian Z, et al. SHREC'15 Track: non-rigid 3D shape retrieval [A]. Proceedings of the Eurographics Workshop on 3D Object Retrieval [C]. Zurich, Switzerland; The Eurographics Association, 2015. 107-120.
- [27] Xie J, Dai G, Zhu F, et al. DeepShape: deep-learned shape descriptor for 3D shape retrieval [J]. Pattern Analysis and Machine Intelligence, 2017, 39(7):1335-1345.

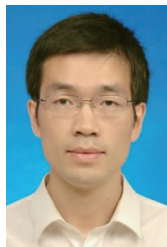
#### 作者简介



汤磊 男, 1995 年出生, 山东枣庄人. 硕士, 主要研究方向为三维模型检索、机器学习.  
E-mail: tanglei\_8520@163.com



丁博(通信作者) 女, 1983 年出生, 河北孟村人. 博士, 副教授, 硕士生导师, CCF 会员. 主要研究方向为三维模型检索、机器学习.  
E-mail: dingbo@hrbust.edu.cn



何勇军 男, 1980 年出生, 四川南充人. 博士, 教授, 博士生导师, CCF 会员. 主要研究方向为机器学习、模式识别.  
E-mail: holywit@163.com