融合信息熵和 CNN 的基于手绘的三维模型检索

刘玉杰[1]， 宋 阳[1]， 李宗民[1]， 李 华[2,3]

(1. 中国石油大学计算机与通信工程学院，山东 青岛 266580；
2. 中国科学院计算技术研究所智能信息处理重点实验室，北京 100190；
3. 中国科学院大学，北京 100190)

摘 要：基于手绘草图的三维模型检索(SBSR)已成为三维模型检索、模式识别与计算机视觉领域的一个研究热点。与传统方法相比，基于卷积神经网络(CNN)的三维深度表示方法在三维模型检索任务中性能优势非常明显。本文提出了一种基于手绘图像融合信息熵和 CNN 的三维模型检索方法。首先，通过计算模型投影图的信息熵得到模型的代表性视图，并将代表性视图经过边缘检测等处理得到三维模型投影图的轮廓图像；然后，将轮廓图像和手绘草图输入到 CNN 中提取特征描述子，并进行特征匹配。本文方法在 Shape Retrieval Contest (SHREC) 2012 数据库和 SHREC 2013 数据库上进行实验。实验证明，该方法的效果较其他传统方法检索准确度更高。

关 键 词：三维模型检索；卷积神经网络；代表性视图；信息熵
中图分类号：TP 391 **DOI：**10.11996/JG.j.2095-302X.2018040735
文献标识码：A **文 章 编 号：**2095-302X(2018)04-0735-07

Sketch-Based 3D Shape Retrieval with Representative View and Convolutional Neural Network

LIU Yujie[1], SONG Yang[1], LI Zongmin[1], LI Hua[2,3]

(1. College of Computer & Communication Engineering, China University of Petroleum, Qingdao Shandong 266580, China;
2. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology Chinese Academy of Sciences, Beijing 100190, China;
3. University of Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Sketch-based shape retrieval (SBSR) has become a hot research spot in the field of model retrieval, pattern recognition, and computer vision. 3D deep representation based on convolutional neural network (CNN) enables significant performance improvement over state-of-the-arts in task of 3D shape retrieval. Motivated by this, in this paper a sketch-based 3D model retrieval algorithm by utilizing entropy representative views and CNN feature matching is proposed. The representative views are obtained by viewpoint entropy. And the representative views are processed by edge detection to get the contour image of 3D model projection. The CNN descriptors extracted as features for representative view of each object. And the method of feature matching is based on CNN descriptors. Our experiments on Shape Retrieval Contest (SHREC) 2012 database and SHREC 2013 database demonstrate that our method is better than state-of-the-art approaches.

收稿日期：2017-11-09；**定稿日期：**2018-01-23
基金项目：国家自然科学基金项目(61379106，61379082，61227802)；山东省自然科学基金项目(ZR2013FM036，ZR2015FM011)
第一作者：刘玉杰(1971–)，男，辽宁沈阳人，副教授，博士。主要研究方向为计算机图形图像处理、多媒体数据分析、多媒体数据库。
E-mail：50312700@qq.com
通信作者：李宗民(1965–)，男，山东聊城人，教授，博士，博士生导师。主要研究方向为计算机图形学、图像处理、科学计算可视化。
E-mail：lizongmin@upc.edu.cn

Keywords: 3D shape retrieval; convolutional neural network; representative view; entropy

1 相关工作

由于三维模型广泛应用于计算机辅助设计、机器人自动化、自动驾驶等领域，在计算机视觉和计算机图形学的发展中，如何从数据集中准确地获取三维模型成为一个重要课题。在过去的十几年中，许多研究者认识到仅依靠基于关键字的文本检索技术进行检索已不能满足需要，试图利用三维模型作为输入从三维数据库中检索相似的模型。虽然三维建模技术和三维扫描设备的发展使得三维模型获取变得容易，但过程仍然复杂。随着触屏手机、平板电脑等智能移动终端的普及，手绘图的获取越来越容易，这也从一定程度上拓宽了基于手绘草图的三维模型检索(sketch-based shape retrieval，SBSR)的应用范围，并提高了 SBSR 技术在学术研究领域的关注度。

SBSR 技术发展至今，学者们发现与基于模型的三维模型检索相比，手绘草图更容易获得。但根据手绘草图检索数据库中的三维模型是一个具有挑战性的问题，由于手绘草图的多变性和随机性，基于手绘草图的三维模型检索的现有方法精度普遍较低。大多数算法其目的是使三维模型或模型在某个视角与手绘草图在特征空间上更接近，图 1 为 SHREC 2012 数据库中部分手绘草图和相应的 3D 模型。因为手绘草图和三维模型属于不同的域空间，两者在高层视觉感知上有明显的差异，这种域差异直接削弱了基于底层图像特征设计的特征描述子的有效性，SBSR 技术力求建立手绘图像与三维模型之间相似度度量关系，所以在过往的研究中，通常的解决方法是通过投影过程将三维模型投影成二维图像，再对投影图像进行边缘检测等处理生成轮廓图，并将其看作是一种特殊形式的手绘图[1]。

SBSR 技术需要解决一个重要问题即寻找具有描述力强而鲁棒性好的特征描述子。基于草图的三维模型检索方法有：基于局部特征的方法、基于全局特征的方法，基于深度特征的方法。傅立叶描述子特征[2]，Zernike moments 特征[3]，Shape Context 特征[4]，方向梯度直方图(histogram of oriented gradients，HOG)特征[5]在图像检索、三维模型检索领域中显示了良好性能并被广泛应用于 SBSR 技术中。有研究者在图像、三维模型

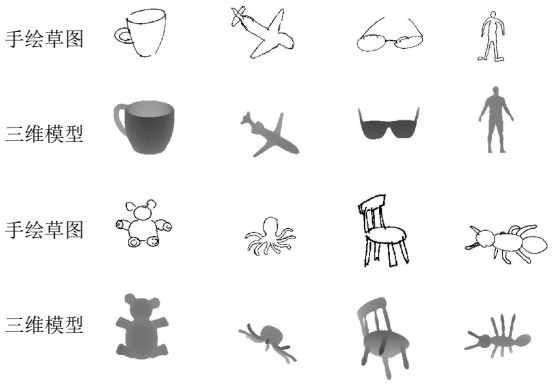

图 1 数据库中手绘草图与其对应的模型

检索技术的基础上，结合手绘图的特点，对传统特征描述子进行改进以实现基于手绘的三维模型检索。另外，随着深度学习(Deep Learning)的迅速发展其特征学习方法备受关注，并在学术界和工业界掀起了一股研究浪潮。Deep Learning 利用大规模训练数据通过深层人工神经网络的方式来学习特征，经过多层的卷积和池化等操作后提取图像的内在信息。Deep Learning 在二维图像和三维模型识别领域得到广泛的应用，在手绘图像分类检索领域也取得了很大的进展。

文献[6-7]通过提取 SIFT 特征并使用改进后的词包模型提出 CDMR-BF-fGALIF+CDMR-BF-fDSIFT 方法，利用流形排序的方法优化检索结果，取得了比较好的效果。文献[8-9]提出了通过 Shape Context 特征的基于手绘草图三维模型检索方法，该算法选取了一组与手绘图相似的模型投影图，再对其进行特征距离计算。文献[10-12]利用 HOG、Gabor Filter 局部特征，并结合词包模型实现检索算法，目标是找到模型的最佳视图(即，手绘草图由该视图角度绘制)，然而，一般计算得到的最佳视图往往和手绘草图的角度不同。2015 年 WANG 等[13]提出了一种新的基于卷积神经网络(convolutional neural network，CNN)的 Siamese 网络算法，前提是数据库中三维模型均为竖直方向，取两个间隔大于 45°视图取代计算最佳视图的方法，并将其命名为极简视图方法(minimalism approach)，取得了良好的效果。但随着三维模型数据库的模型数量不断扩大也给该方法带来不确定性，使得到的极简视图与手绘图的绘图角度不一定相同，而且该方法仍存在检索精度不高的问题。

2　基于代表性视图和卷积神经网络的 SBSR 方法

本文提出了通过计算模型投影图的熵值选取模型的代表性视图，进而利用 CNN 提取得到的三维模型与手绘草图的特征进行匹配，其流程如图 2 所示。

2.1　数据预处理

手绘草图的三维检索面临的最大挑战是手绘草图和三维模型之间的语义鸿沟，本文通过对三维模型数据进行处理克服这一问题，实现手绘草图和三维模型在图像域上的统一。利用文献[14]中提出的方法将一个三维模型包围在正十二面体中，在各顶点处进行投影，并在每个顶点位置旋转 5 次，此时三维模型可以由 100 个大小为 225×225 像素的深度图像表示。本文利用 Canny 边缘检测算法对深度图像进行边缘提取，不同于自然场景图片，手绘图和三维模型投影图的背景较为纯净，经过边缘提取的深度图像，在视觉上接近于手绘草图的轮廓图。

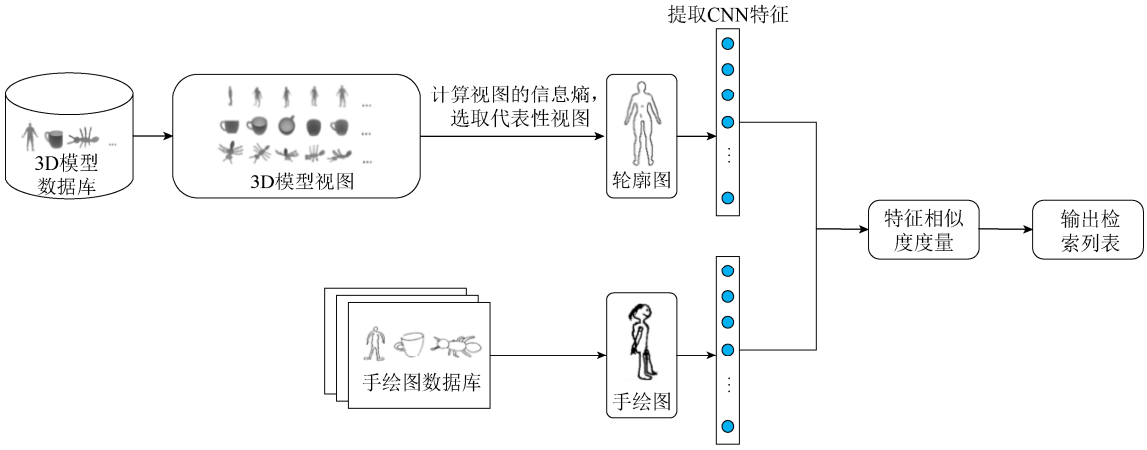

图 2　本文方法的流程图

2.2　代表性视图的选取

三维模型中最具代表性的视图应尽可能多地包含模型的信息，且最能帮助描述模型。换言之，代表性视图是获得模型最多信息的视图。绘制手绘草图需根据这些视图来选择绘制视角，而且冗余的投影图会给检索带来很大地挑战，利用文献[13]方法尝试解决这个问题。该方法假设在数据库中的三维模型均为竖直的且间隔大于 45°，随机采样两个视图，并以此代表模型，但垂直方向随机选择的视点可能与手绘草图的绘制角度不同，从而影响特征匹配的正确性。本文方法受文献[15]的启发，提出了基于视图信息熵的三维模型视图的度量方法。熵是对系统或对象混乱程度的度量，即表示其平均信息含量。离散随机变量 $X=\{a1, a2, \cdots, an\}$的熵定义为

$$H(X)=-\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) \tag{1}$$

式(1)为有 n 个投影图的三维模型，视图的熵值是在视点为中心的方向上投影面相对面积的概率分布。因此，可将视图的熵定义为

$$H=-\sum_{i=1}^{n}\frac{A_i}{S}\log_2\frac{A_i}{S} \tag{2}$$

其中，A_i 为第 i 个投影图中投影区域的面积；S 为投影窗口的总面积，即 $S=A_0+\sum_{i=1}^{n}A_i$，A_0 为投影区域的背景区域面积。

通过计算视图的信息熵，解决了模型的代表性视图选取问题。如图 3 所示，在三维模型的投影图中，视图的熵值越大包含的信息越多，越易分辨出模型的类别。只选取 1 个代表性的视图不能保证获得足够的信息。实验证明，选择 6 个代表性视图时，检索的效果最佳。因此，本文选择了 6 个熵值最大的视图作为代表性视图。

2.3　CNN 特征提取

通过对手绘草图和三维模型的数据预处理，三维模型的投影图被转换成与手绘图类似的轮廓图像；之后的任务是提取适合 SBSR 并且具有高描述力的描述特征。本文利用 CNN 提取特征，采用经典的深度网络框架 AlexNet[16]，AlexNet 是通过 ILSVRC 2012 Image Net 数据集训练的。此数据集

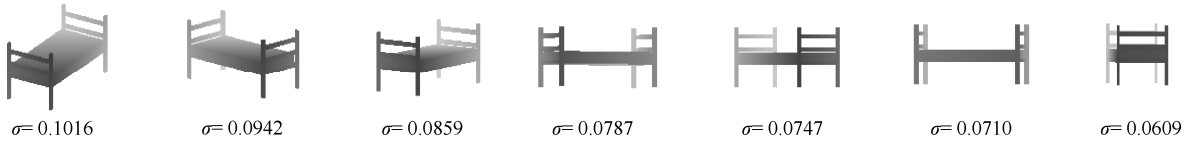

图 3 三维模型不同视图的熵值

由 1 300 万个数据组成，利用深度神经网络提取特征的方法，主要包括：

步骤 1. 对 AlexNet 的网络结构和参数进行修改；

步骤 2. 输出网络的全连接层作为输入对象的特征。

AlexNet 在二维图像领域是一个可靠的 CNN。然而，与 AlexNet 庞大的二维图像训练集不同，手绘草图数据有其自身的特点，比如背景纯净，以线条为主。步骤 1 中，以 SHREC 2013 数据库中的训练集对深度 CNN 进行微调。本文中，考虑到 Sketch-a-Net[17]在手绘草图像领域的优势，借鉴其网络设计方法，修改 Alexnet 结构。CNN 成功地使用了极小的 3×3 滤波器，相比文献[17]发现较大的滤波器更适合手绘草图的处理。第一卷积层的参数可能是最敏感的，因此，在第一个卷积层中，使用大小为 15×15 的滤波器取代原有的 11×11 的滤波器，在之后的卷积层中，均使用大小为 3×3 的滤波器，本文选择最后一个连接层 FC_7 作为输入对象的特征。表 1 展示了修改后的网络结构和参数；基于深度神经网络的手绘草图和模型投影图的特征匹配过程如图 4 所示。

表 1 修改后的神经网络结构和参数

网络层	卷积核大小	步长	扩充边缘	输出
input	-	-	-	225×225
conv1	15×15	3	0	64×71×71
pooling1	3×3	2	0	64×35×35
conv2	5×5	1	0	128×31×31
pooling2	3×3	2	0	128×15×15
conv3	3×3	1	1	256×15×15
conv4	3×3	1	1	256×15×15
conv5	3×3	1	1	256×15×15
pooling3	3×3	2	0	256×7×7
conv (=FC)_a	7×7	1	0	4096×1×1
conv (=FC)_b	1×1	1	0	4096×1×1
conv (=FC)_c	1×1	1	0	90×1×1

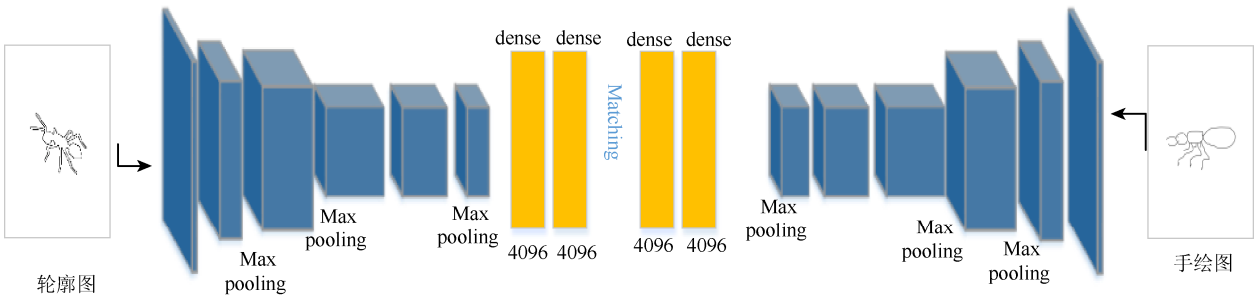

图 4 基于深度神经网络框架的手绘草图和模型投影图的特征匹配

2.4 相似度度量

本文用欧式距离度量手绘草图和三维模型投影图之间的相似性。其中，N 为三维模型代表性视图的集合，在本文方法中 n=6。$D(f_S, f_i)$ 表示手绘草图与投影图之间的欧氏距离，相似性度量为

$$d_m = \min(D(f_s, f_i)),\ i \in N\{P_1, P_2, \cdots, P_n\} \quad (3)$$

其中，f_s 为手绘草图的特征；f_i 为三维模型投影图的特征。d_m 值越小，表明两幅图像越相似。

3 实 验

3.1 数据集与评价指标

为了验证本文所提出方法的有效性，分别在 SHREC 2012 和 SHREC 2013 两个标准数据库上进行了实验并评估检索的性能。

SHREC 2012 数据集包含基础版和扩展版两个版本，分 20 类，每类有 20 个模型，共计 400 个模型。在实验中，使用的是基础版本数据集，其包括手绘图和三维模型两部分，手绘草图分为 13 个类，共有 250 幅图像。三维模型分为 13 个类，每类 20 个模型。

SHREC 2013 是基于大规模手绘草图的三维模型数据库。该数据集中包含 1 258 个三维模型和 7 200 个手绘草图，分为 90 个类。在每个手绘草图类中，50 个草图用于训练，其余 30 个草图用于测试。值得注意的是，其是一个类间模型数量非常不平衡的数据集。在不同的模型类中模型的数量变化很大，这对实验结果造成了很大的挑战。例如，airplane 类有 184 个模型，axe 类只有 4 个模型。评价过程中，本文采用三维检索领域基本的评价指标 Nearest Neighbor (NN)、First Tier (FT)、Second Tier (ST)、E-Measure (E)、Discounted Cumulated Gain (DCG)、mean Average Precision (mAP)和 Precision-Recall (P-R)曲线图。其中，NN 表示返回的第一个模型属于目标类的比例，FT 表示返回的前 C–1 (C 为目标类模型的数量)个模型属于目标类的比例，ST 表示返回的前 2 (C–1)个模型属于目标类的比例，E、DCG 是综合查全率和查准率的指标，mAP 反映平均检索精度，P-R 曲线图则能体现总体检索效率。

3.2　实验结果与评价

本文在 SHREC 2012 和 SHREC 2013 两个标准数据库与几种传统经典方法进行对比。其在 Caffe 框架下实现对网络的修改及微调，训练过程在 GeForce GTX TITAN GPU 上完成，用 Caffe 的 MATLAB 接口实现输出全连接层作为输入对象的特征。

选取不同数量的代表性视图对检索结果有着很大的影响，代表性视图选取得过少对模型的描述将不够完整，选取得过多会产生许多有歧义性的投影图，很难分辨出物体。图 5 是在 SHREC 2012 数据集上完成。实验结果表明，选取 6 个代表性视图时，平均 mAP 值达到最高点，之后基本保持平稳，所以本文采用 6 个熵值最大的投影图来描述。另外，为了验证熵值最大的方式选取代表性视图的合理性，在 SHREC 2013 数据集上进行了前 6 个熵值最大值、前 6 个熵值最小值、随机选取 6 个视图对比实验。图 6 为以不同方式选取代表性视图在检索实验中得到的 PR 曲线，实验证明选取熵值最大的 6 个视图时，检索精度最高。

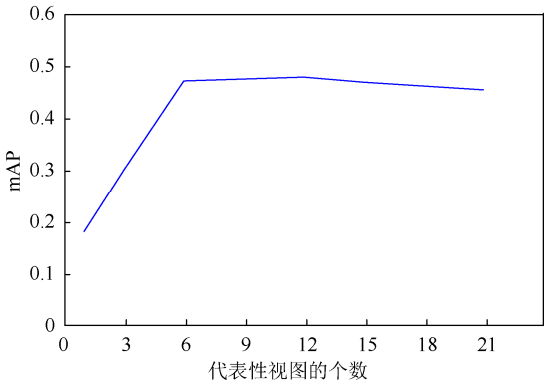

图 5　在 SHREC 2013 数据库上选取不同数量的代表性视图对检索精度的影响

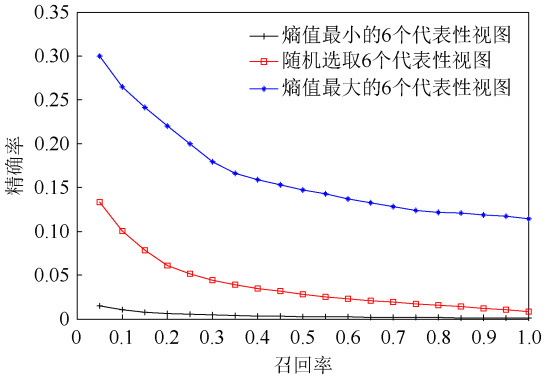

图 6　以不同方式选取代表性视图的 PR 曲线结果对比

本文以特征向量之间的欧氏距离作为代表性视图与手绘草图的相似度度量方式，CNN 的全连接层作为输入数据的特征，与其他几种经典的方法进行对比，SHREC 2012 数据集中的部分检索结果如图 7 所示，本文方法在形状比较简单的图像上检索精度较高(如图 7 第 2、7 行)，且在一些歧义比较大的手绘图像，出现部分误检(如图 7 第 3、6 行)，误检结果在检索列表中用方框标出。这些误检结果的出现主要是因为手绘草图的自身歧义性造成的，比如蜘蛛、章鱼、蚂蚁在手绘图像表达上区分度较小。在该数据集实验中，本文的方法与传统的特征进行了对比，其中，利用 SIFT 特征描述三维模型和手绘图像的方法结果表现较好，直接使用 HOG 特征表达力较弱，从表 2 中可以看出，本文方法的各项指标均优于其他方法。其中 NN 评价标准比传统特征中表现最好的方法提升了 12.5%，并且在 FT、ST 等指标上都有显著提高。

在 SHREC 2013 数据集中，本文方法与几种经典方法进行对比，见表 3。图 8 是本文与其他方法在 PR 曲线上的比较，实验结果显示，本文的方法优于其他方法。由于 SHREC 2013 数据库属于较大

规模的基于手绘草图的三维模型数据库，模型数量的增加和手绘图的多变性，致使现有的方法在该数据库上的检索精度均不理想。另一个重要的原因是该数据库各类中的三维模型数量相差很大且及不平衡，如：airplane 有 184 个模型，而 bed 仅有 4 个模型，模型数量太少，会极大影响检索的查全率和查准率。

表 2 在 SHREC 2012 数据库中各方法检索结果对比

方法	NN	FT	ST	E	DCG	mAP
本文方法	**0.701**	**0.621**	**0.811**	**0.652**	**0.823**	**0.644**
HOG-SIL[5]	0.188	0.123	0.223	0.139	0.466	0.143
CDMR[18]	0.668	0.585	0.770	0.554	0.801	0.625
BF-fGALIF + BF-fDSIFT[12]	0.508	0.317	0.461	0.316	0.648	0.338
SBR-VC NUM 100[4]	0.576	0.372	0.519	0.360	0.682	0.392

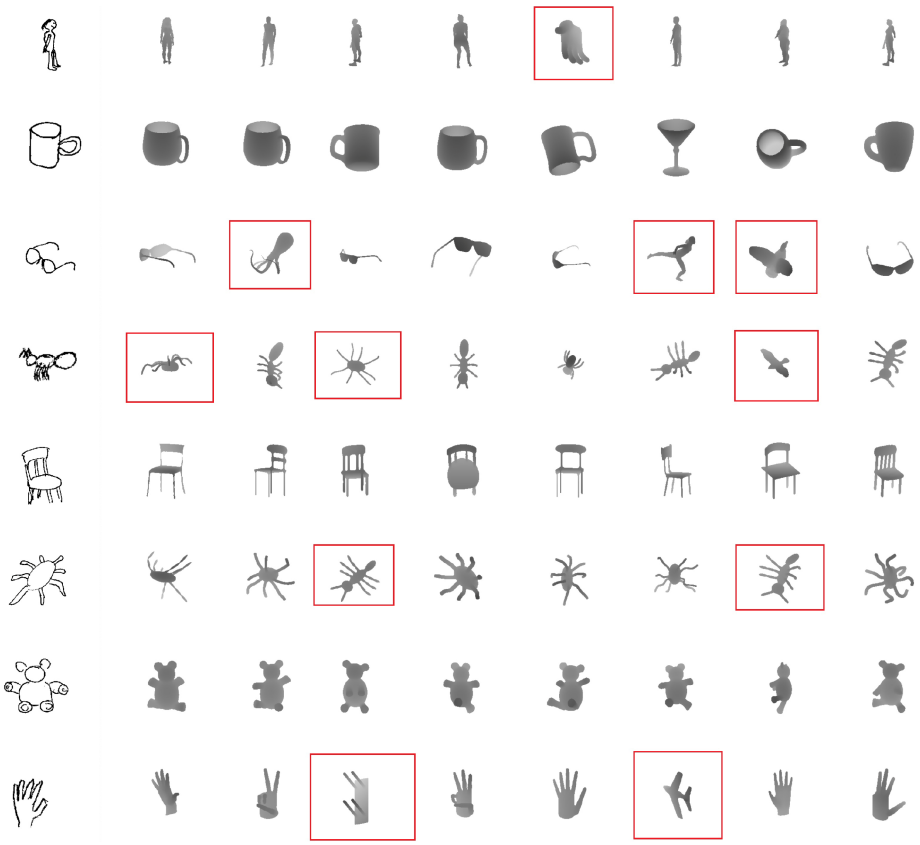

图 7 本文方法在 SHREC 2012 数据库上的部分检索结果(框内的模型是误检模型)

表 3 在 SHREC 2013 数据库中各方法检索结果对比

方法	NN	FT	ST	E	DCG	mAP
本文方法	**0.417**	**0.410**	**0.586**	**0.339**	**0.610**	**0.472**
Siamese CNNs[13]	0.405	0.403	0.548	0.287	0.607	0.469
HOG-SIL[5]	0.110	0.069	0.107	0.061	0.307	0.084
CDMR[18]	0.279	0.203	0.296	0.166	0.458	0.246
SBR-VC[4]	0.161	0.097	0.149	0.085	0.349	0.113

针对以上提出的 SHREC 2013 数据库中各类模型数量不平衡的问题，本文评估了该数据库中类间模型数量的不齐对实验精度的影响，在实验中，选取了该数据库中的 9 类数据以 3 项指标对其进行评估。实验结果见表 4，可以看出，模型数量较多的模型类 3 项指标精度高于模型数量较少的类。由此可见，模型数量少的类对精度影响较大，尤其是 FT、ST 评价指标。

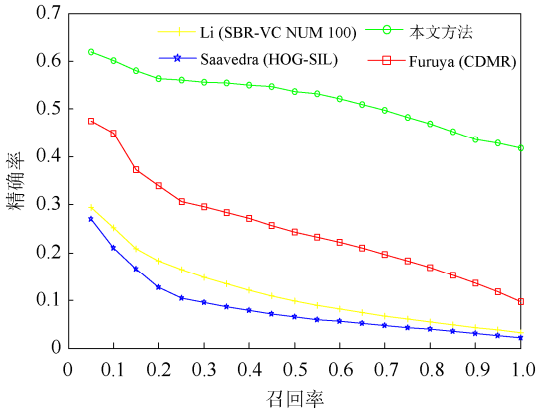

图 8 本文方法在 SHREC 2013 数据库上与其他方法比较的 PR 曲线

表 4　SHREC 2013 数据库中部分类的检索结果对比

类	类内模型个数	NN	FT	ST
airplane	184	0.892	0.868	0.956
bee	4	0.201	0.162	0.223
bed	8	0.296	0.198	0.360
mailbox	7	0.301	0.256	0.410
potted-plant	51	0.625	0.606	0.752
race-car	33	0.412	0.398	0.684
sword	31	0.212	0.209	0.364
table	63	0.891	0.808	0.920
vase	22	0.600	0.536	0.692

4　结 束 语

随着大数据时代的来临，图像获取变得越来越简单，获取方式变得越来越多样化。基于手绘草图的多媒体检索技术受到了广泛的关注。本文针对手绘草图与三维模型之间存在的语义差异，提出了一种基于熵值计算对三维投影获得的轮廓视图选取代表性视图，进而利用 CNN 提取特征进行相似性匹配。在 SHREC 2012 和 SHREC 2013 两个标准数据库上对本文方法进行了验证，实验结果证明相同评价标准下本文方法在检索精度上高于其他算法。进一步研究发现，现有的深度神经网络对手绘草图这一类特殊图像的描述力不足，在 SHREC 2013 这种具有挑战性的数据库中的检索精度还是较低。下一步工作将主要研究基于深度神经网络的手绘草图特征提取问题。

参 考 文 献

[1] LI B, LU Y, GODIL A, et al. A comparison of methods for sketch-based 3D shape retrieval [J]. Computer Vision & Image Understanding, 2014, 119(2): 57-80.

[2] ZHANG D, LU G. A comparative study of Fourier descriptors for shape representation and retrieval [C]// Proceedings of 5th Asian Conference on Computer Vision (ACCV). Berlin: Springer, 2002: 646-651.

[3] KHOTANZAD A, HONG Y H. Invariant image recognition by zernike moments [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 12(5): 489-497.

[4] LI B, LU Y, JOHAN H. Sketch-based 3D model retrieval by viewpoint entropy-based adaptive view clustering [C]//Eurographics Workshop on 3D Object Retrieval. New York: ACM Press, 2013: 49-56.

[5] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]//Computer Vision and Pattern Recognition (CVPR), 2005. New York: IEEE Press, 2005: 886-893.

[6] LI B, SCHRECK T, GODIL A, et al. SHREC'12 track: sketch-based 3D shape retrieval [C]//Eurographics Conference on 3D Object Retrieval. New York: ACM Press, 2012: 109-118.

[7] FURUYA T, OHBUCHI R. Dense sampling and fast encoding for 3D model retrieval using bag-of-visual features [C]//ACM International Conference on Image and Video Retrieval. New York: ACM Press, 2009: 26.

[8] LI B, JOHAN H. Sketch-based 3D model retrieval by incorporating 2D-3D alignment [J]. Multimedia Tools & Applications, 2013, 65(3): 363-385.

[9] LI B, LU Y, GODIL A, et al. SHREC'13 track: large scale sketch-based 3D shape retrieval [C]// Eurographics Workshop on 3D Object Retrieval. New York: ACM Press, 2013: 89-96.

[10] EITZ M, HILDEBRAND K, BOUBEKEUR T, et al. Sketch-based 3D shape retrieval [C]//SIGGRAPH 2010. New York: ACM Press, 2010: 5.

[11] EITZ M, HILDEBRAND K, BOUBEKEUR T, et al. Sketch-based image retrieval: benchmark and bag-of-features descriptors [J]. IEEE Transactions on Visualization & Computer Graphics, 2011, 17(11): 1624-1636.

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

[13] WANG F, KANG L, LI Y. Sketch-based 3D shape retrieval using convolutional neural networks [C]// Computer Vision and Pattern Recognition (CVPR), 2015. New York: IEEE Press, 2015: 1875-1883.

[14] CHEN D Y, TIAN X P, SHEN Y T, et al. On visual similarity based 3D model retrieval [C]//Computer Graphics Forum. Oxford: Blackwell Publishing, Inc, CiNii, 2003: 223-232.

[15] FEIXAS M, SBERT M, HEIDRICH W. Viewpoint selection using viewpoint entropy [C]//Vision Modeling and Visualization Conference. New York: ACM Press, 2001: 273-280.

[16] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]//International Conference on Neural Information Processing Systems. New York: ACM Press, 2012: 1097-1105.

[17] YU Q, YANG Y, LIU F, et al. Sketch-a-Net: a deep neural network that beats humans [J]. International Journal of Computer Vision, 2017, 122(3): 411-425.

[18] UIJLINGS J R, SANDE K E, GEVERS T, et al. Selective search for object recognition [J]. International Journal of Computer Vision, 2013, 104(2): 154-171.