

# 基于深度卷积神经网络的三维模型检索

刘 志 李江川

(浙江工业大学计算机科学与技术学院 杭州 310023)

**摘 要** 为了更有效地利用三维模型数据集进行特征的自主学习,提出一种使用自然图像作为输入源,以三维模型的较优视图集为基础,通过深度卷积神经网络的训练提取深度特征用于检索的三维模型检索方法。首先,从多个视点三维模型进行视图提取,并根据灰度熵的排序选取较优视图;然后,通过深度卷积神经网络对视图集进行训练,从而提取较优视图的深度特征并进行降维,同时,对输入的自然图像提取边缘轮廓图,经过相似度匹配获得一组三维模型;最后,基于检索结果中同类模型总数占检索列表长度的比例对列表进行重排序,从而获得最终的检索结果。实验结果表明,该算法能够有效利用深度卷积神经网络对三维模型的视图进行深度特征提取,同时降低了输入源的获取难度,有效提高了检索效果。

**关键词** 三维模型检索,视图选取,特征提取,深度学习

**中图分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.01.043

## 3D Model Retrieval Based on Deep Convolution Neural Network

LIU Zhi LI Jiang-chuan

(College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China)

**Abstract** In order to make use of 3D model data set for feature self-learning more effectively, this paper proposed a 3D model retrieval method, in which the natural images are as input sources, a better view set of 3D model is as basis, and the depth features obtained through training deep convolutional neural networks are applied for retrieval. Firstly, the 3D model views are extracted from multiple viewpoints and the optimal view is selected according to the order of gray entropy. Secondly, the view set is trained through deep convolution neural network, the depth features of the optimal view are extracted and its dimension is reduced. At the same time, edge contouring is extracted for the input natural image and a set of 3D models is gotten after similarity matching. Finally, according to the ratio of the total number of similar models to the length of the search list in the retrieval result, the retrieval list is reordered and the final result will be gained. Experimental results show that the algorithm can effectively use depth convolution neural network to extract the depth feature of the view of the 3D model, meanwhile reduce the difficulty of obtaining the input source, and improve the retrieval efficiency effectively.

**Keywords** 3D model retrieval, View selection, Feature extraction, Deep learning

## 1 引言

随着计算机图形处理能力和三维建模技术的日益增强,三维模型在工业设计、虚拟现实、医学诊断等诸多领域得到了广泛应用,三维模型的数量也随之呈爆发式增长。海量的数据为三维模型检索技术的发展带来了新的机遇与挑战。近年来,深度学习<sup>[1-2]</sup>在图像识别<sup>[3-4]</sup>、自然语言处理、立体视觉匹配<sup>[5-6]</sup>等领域取得了优异的成果。Kkizhevsky 等<sup>[7]</sup>于 2012 年在 ImageNet 图像数据库上采用深度卷积神经网络(Deep Convolutional Neural Network, DCNN)进行图像分类,取得了惊人的效果,使 Top5 分类错误率降低了 11%。Simonyan 等<sup>[8]</sup>于 2014 年提出了 VGG 网络结构模型,在 ILSVRC 2014

比赛中取得了图像分类第二名和图像定位第一名的成绩,并且验证了增加网络深度对提高训练效果的重要性,对其后的神经网络相关研究产生了较强的启发。深度卷积神经网络在基于内容的图像检索领域也得到了新的应用,Alzu'Bi 等<sup>[9]</sup>、Salvador 等<sup>[10]</sup>、刘冶等<sup>[11]</sup>国内外研究者基于自然图像数据集使用深度卷积神经网络进行训练,将全连接层获取到的特征用于检索中,实验证明该方法提取的特征可以对图像进行更为有效的表达,并能够显著提高检索效果,证明了深度卷积神经网络模型在图像检索领域的适用性。

在基于内容的三维模型检索领域,也有许多研究者跟进深度学习方法的研究,在基于示例模型的检索中, Zhu 等<sup>[12]</sup>使用三维模型的深度视图对深度贝叶斯网络进行有监

到稿日期:2017-11-30 返修日期:2018-02-15 本文受浙江省自然科学基金(LY16F020033)资助。

刘 志(1969—),女,博士,教授,CCF 会员,主要研究方向为三维模型检索、图像处理等,E-mail:lzhi@zjut.edu.cn(通信作者);李江川(1992—),男,硕士生,主要研究方向为三维模型检索。

督的微调,训练出自编码器来对深度视图进行特征表达,取得了良好的检索效果;Li 等<sup>[13]</sup>使用卷积神经网络对同一个三维模型的多个视图分别进行了特征提取,再将各个视图的特征压缩为一个 128 维的综合特征,用于对三维模型进行特征表达。以上基于示例模型的三维模型检索方法能够达到较高的检索精度,但在实际应用中,面临着合适的示例模型难以获取的问题。Wang 等<sup>[14]</sup>将深度学习应用于以手绘草图作为输入源的三维模型检索中,使用两个 Siamese CNN 分别针对草图和三维模型视图进行学习,并在网络结构末端进行融合,最终得到的特征对两个数据集都具有良好的表达效果。但随着复杂模型的日益增加,基于草图的三维模型检索面临着普通用户绘制的草图难以有效描述复杂模型的问题。

不同于使用手绘草图或示例模型作为输入的形式,本文以简单背景下的自然图像作为检索的输入源,对其提取轮廓

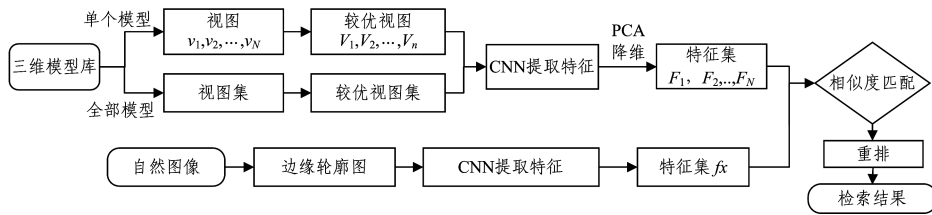


图 1 三维模型检索系统框架

Fig. 1 Framework of 3D model retrieval system

不同于以往的较优视图选取方法,本文通过计算三维模型视图的灰度熵来选取较优视图,并通过控制不同类别模型选取的较优视图数量来均衡各个分类下视图的总数,进而减少模型数量不均对神经网络的训练过程产生的影响,在此基础上通过深度卷积神经网络的构建对视图提取深度特征并将该特征用于检索。同时,该方法以网络上海量的自然图像为基础,进一步降低了输入源的获取难度。实验证明,本文方法能够有效提高三维模型检索的效果。

## 2 较优视图的提取

### 2.1 视图渲染方法

在基于视图的三维模型检索中,视图的渲染质量决定了视图所能携带的三维模型的信息量,进而直接影响检索的效果。本文选取 Decarlo 等<sup>[15]</sup>提出的闭合轮廓线与暗示轮廓相结合的渲染方法(以下简称混合轮廓线)对三维模型进行视图渲染。首先检测视点向量与模型表面法向量垂直的部分,并将其绘制出来得到闭合轮廓线,然后在其周围检测出接近人类可视曲率的部分线条,进一步绘制出暗示轮廓线,以获取最终的线条轮廓图。通过图 2 的对比可以看出,混合轮廓线能够对模型的形状信息进行有效的记录,并且在绘制外围轮廓的同时渲染出部分内部细节,使得模型表面的许多关键信息得以保留,取得了良好的渲染效果。



图 2 视图对比

Fig. 2 Comparison of view images

线条图,并与三维模型的二维视图进行相似度匹配,从而建立一个三维模型检索系统,该系统的框架如图 1 所示。首先,从多个预设视点渲染并提取三维模型的视图,根据数据集中该模型所属分类下的模型数量占模型总数的比例来确定该模型最终保留的较优视图数量,并依据视图灰度熵的大小去除冗余视图。其次,在 VGG16 网络结构中卷积模块的基础上构建一个由 3 个卷积层和 2 个全连接层构成的卷积神经网络,使用该网络结构,基于较优视图集进行充分训练后对视图提取深度特征,并通过 PCA 对高维特征进行降维。最后,在检索阶段对输入的自然图像进行边缘检测以获得轮廓线条图,通过上述神经网络进行特征提取,并与较优视图集进行相似度匹配,获取初始检索结果列表,依据列表中不同分类下的模型数量与列表长度的比例调整匹配值,得到最终的检索结果,以此构建一个三维模型检索系统。

而在输入源方面,本文选取简单背景下的自然图像作为输入,对其提取剪影轮廓,并结合 Canny 算子<sup>[16]</sup>提取混合边缘,然后通过形态学膨胀去除细小噪声,平滑物体边界,如图 3 所示。该方法能够有效控制噪声的产生,突出关键线条,提取到的线条轮廓图能够有效表述自然图像的轮廓特征,具有可观的辨识度。



图 3 自然图像与边缘轮廓图

Fig. 3 Natural image and edge line image

## 2.2 较优视图选择

在视图选取方式上,研究者通常选择用一个多面体包围三维模型,以多面体的顶点作为视图选取的视点,提取与顶点数量一致的视图;也有一些研究者通过网格显著度、可见面积比、最大投影面积法等不同的算法来选取较优视点。但前者获取到的视图冗余度过高,后者在不同类别的模型下提取的视图质量不一,难以找到具有普适性的视图选取方法。

在选择视图提取数量时,研究者往往会对每个模型提取相同数量的视图,但通常数据集中不同分类下的三维模型数量不一,因此会出现训练数据不均衡的问题,进而对基于较优视图集进行深度卷积神经网络的训练产生影响。

针对以上两点问题,文本提出了首先从预定义视点提取多张视图,然后根据数据集中每个分类下的模型数量与模型总量的比值确定该分类下模型的较优视图数量的视图选取方法。

在预定义视点的选取上,本文采用文献[17]中提到的视点球对三维模型进行包裹,使用6个正投影点和10个等距投影点从不同角度等距地对三维模型进行视图提取。

在视图质量的评价上,Li等[18]通过计算三维模型的视点熵,根据模型的投影面积结合面片数量定义较优视点,进而提取较优视图。受此启发,本文提出将提取后的视图按照灰度

熵排序,保留熵最大的 $n$ 个视图的方法,以选取较优视图,其中 $n$ 如式(1)所示:

$$n = N - a \frac{c_i}{\sum_{j=1}^t c_j} \quad (1)$$

其中, $N$ 代表固定视点数量,本文取值为16; $n$ 代表该类别下模型最终的视图数量; $a$ 是固定常数,用于控制视图数量,实验中取定值40; $c_i$ 代表 $i$ 类别下的模型数量; $t$ 代表数据集中模型的类别总数。

整体的选取流程如图4所示,具体步骤如下:

Step 1 使用如图4(a)所示的以三维模型的质心为球心的、包含视点 $p_1, p_2, \dots, p_N$ 的视点球对三维模型进行包裹。

Step 2 在每个视点下,使用混合轮廓线算法对三维模型进行视图渲染和提取,得到 $N$ 张二维视图( $v_1, v_2, \dots, v_N$ )。

Step 3 根据模型所属类别的数量,通过式(1)计算该类别下模型的较优视图数量 $n$ 。

Step 4 使用式(2)计算每张视图的灰度熵,并按照灰度熵的大小进行排序,对每个模型保留熵最大的前 $n$ 张视图,并将其作为该模型的较优视图( $V_1, V_2, \dots, V_n$ )。

$$H = -\sum_{i=0}^{255} p_i \log p_i \quad (2)$$

其中, $p_i$ 表示灰度值为 $i$ 的像素在整个视图中所占的比例, $H$ 代表灰度熵。

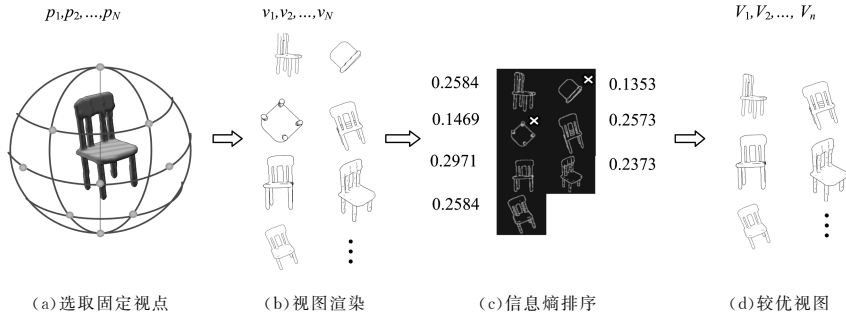


图4 视图提取过程

Fig. 4 Extraction process of view image

在视图数量的选择上,Hong等[19]在以示例模型作为输入源的三维模型检索中,对每个模型提取42张视图用于训练,取得了较高的检索精度。但在以自然图像作为输入源的三维模型检索领域,存在着人们对三维物体拍摄图像时选取的视角往往较为固定的现状,过多不同角度的视图会产生大量冗余,进而影响检索效果。因此,本文选择基于16个预设视点对模型进行视图渲染,并使用最小熵法去除三维模型表达效果较差的视图,同时针对不同类别下模型数量不均的状况,根据数据集中同类模型的数量选取该模型提取的视图总量。相较于主流的视图提取算法,本文方法降低了视图提取数量,同时均衡了不同类别下的较优视图总量,为下文卷积神经网络的训练提供了良好的数据基础。

## 3 深度特征提取

### 3.1 CNN-FLI网络结构设计

深度学习在基于视图的三维模型检索领域的应用面临着

以下两点问题:1)同一个三维模型在不同视点下提取的视图之间存在较大的差异,这会对训练过程产生一定的干扰;2)三维模型现有的数据规模远少于自然图像,因此直接基于三维模型视图集对卷积神经网络进行训练难以获得令人满意的效果。

而在深度卷积神经网络中,越靠近网络顶层提取到的特征越局部化,该部分特征的泛化能力越强,在多数情况下是适用于所有图像的;越靠近底层的部分获得的特征就越全局化,与使用的数据集的关系越紧密。利用这一特性,本文选取VGG16模型中的卷积模块,使用在ImageNet数据集中进行充分训练后的权重参数对其进行初始化,以此作为初步的特征提取器,用于通用的局部特征提取,在其后接上由3个卷积层、1个池化层、2个全连接层以及1个Softmax分类器组成的卷积神经网络CNN-FLI(Convolutional Neural Network for Lines based Image),以针对较优视图集进行训练。

本文使用的卷积神经网络结构的详细设计如图5所示。

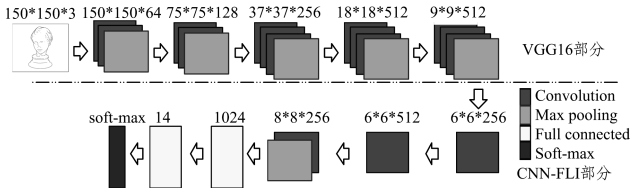


图 5 网络结构图

Fig. 5 Diagram of network structure

首先是卷积模块,本文将二维视图的面积统一转换为 $150 \times 150$ 进行输入,在每一个卷积层中,对前一层的输入 $x_i$ 使用卷积核 $W_{ij}$ 进行卷积,使用 $\text{ELU}^{[20]}$ 激活函数(见式(3))对结果进行处理后,产生新的输出 $y_j$ (见式(4))。

$$f(x) = \begin{cases} x, & x \geq 0 \\ \alpha(e^x - 1), & x < 0 \end{cases} \quad (3)$$

$$y_j^l = f(z^l) = f\left(\sum_{i=1}^M (x_i^{l-1} * W_{ij}^l + b_j^l)\right) \quad (4)$$

其中, $l$ 代表第 $l$ 个卷积层, $*$ 代表卷积操作, $M$ 代表该层的输入数量, $b_j$ 代表该层的偏置。在误差反向传播阶段,使用式(5)对实际输出与理想输出之间的误差进行计算:

$$E = \frac{1}{2m} \sum_{j=1}^m (y_j - o_j)^2 \quad (5)$$

其中, $y_j$ 代表实际输出, $o_j$ 代表理想输出, $m$ 代表样本数目。在此过程中,使用随机梯度下降法(SGD)对累计误差进行最小化。

在激活函数的选取上,相较于通常使用的 ReLU 函数,本文选取了 ELU 函数用于激活,避免了 Dead ReLU,即 ReLU 函数的使用可能导致一些神经元永远不被激活。同时,由于 ELU 函数的输出均值接近 0,因此使用该函数作为激活函数时,SGD 的收敛速度更快。

对于卷积模块中卷积核的设置,不同于 VGG16 的卷积层中卷积核大小全部为 $3 \times 3$ 的设置,本网络选择了 $1 \times 1$ , $3 \times 3$ , $1 \times 1$ 相交替的模式,其中 $1 \times 1$ 的卷积核的作用是在保持维度不变的情况下,对输入进行线性形变,之后再通过 ELU 激活函数对其进行非线性化,降低了层数的增加带来的过拟合倾向。在卷积层的最后设置了 $2 \times 2$ 的池化层,通过最大池化操作,达到了如下效果:

1)有效减少参数的数量。一个 $m \times m$ 的特征矩阵,经过面积为 $a \times a$ 的池化区域的池化后,参数数量会变成 $(m/a) \times (m/a)$ 。

2)使提取的特征对图像的平移、旋转、缩放具有一定的不变性。这是因为池化区域如果足够大,则会忽略掉区域内的一些细微变化,从而使得特征的泛化能力更强。

在全连接层的设计上,本文使用 1024 维的全连接层搭配与模型类别总数一致的最后一层全连接层作为 CNN-FLI 的全连接模块。考虑到线条图包含的信息量少于自然图像,因此相较于 VGG16 结构中的设置,本文将全连接层的参数从 4096 降低至 1024,并将三层全连接层结构减少至两层,以降低神经元数量,提高收敛速度。

### 3.2 神经网络训练

本文使用的深度卷积神经网络模型包括 VGG16 的卷积模块和 CNN-FLI 模块两部分,前者在 ImageNet 上已经经过

了充分训练,较高的学习率可能会破坏该部分权重,而 CNN-FLI 部分尚未初始化,因此将整个训练过程分成两部分进行。

首先是预训练阶段,此时固定 VGG16 卷积模块的权重参数,将其作为特征提取器,在此基础上对 CNN-FLI 的权重进行随机初始化,选择 RMSProp 最优化方法展开训练。以 $1e-3$ 为学习率进行充分训练后,CNN-FLI 会获得较好的初始权重。

其次是 fine-tune 部分,此时整个网络都获取到了经过训练的初始权重,已经具备了一定的特征提取能力。为了进一步提高该网络在线条视图集上的效果,本文将 VGG16 模块的最后 3 个卷积层与 CNN-FLI 进行同步 fine-tune。经过此阶段的训练,网络模型对二维线条图的特征提取能力会得到进一步的提升。而在 fine-tune 过程中,为了避免较高的学习率使训练效果产生震荡、难以收敛的情况出现,将学习率设置为 $1e-4$ 。实验证明,在该学习率下,loss 稳步下降,收敛速度较快,准确率稳定提升。

经过充分训练后,本文选取全连接层输出的 1024 维的特征代表二维视图,该特征对轮廓线条图具有较好的表达效果,但在检索领域需要保证一定的实时性,因此在高维特征的基础上,本文使用 PCA 将维度降低至 128 维,在保持检索精度稳定的前提下,减少了冗余数据,有效提高了检索效率。

### 3.3 检索结果重排序

重排序是指利用初始检索列表提供的信息,对检索结果进行二次排序,以优先展示与检索目标更为一致的检索结果的方法。本文依据同一模型的不同视图在检索列表中出现的次数以及列表中各个类别下的模型数量占列表长度的比例对检索结果进行重排序。

在三维模型视图的检索过程中,同一个模型的多个视图可能同时出现在结果列表中,通常的做法是仅保留该模型下匹配度最高的视图,而本文在此基础上将检索列表中每个模型被检索到的视图的数量纳入匹配度的计算中。首先,对输入图像的边缘轮廓图和模型的较优视图分别提取深度特征,并计算两者间的欧氏距离,进而得到匹配度。其次,按照匹配度排序获取初始检索列表 $L_0$ ,当同一三维模型的不同视图同时进入该列表时,选取匹配度最高的视图,并将其与所属的三维模型逐一对应,从而得到三维模型的候选列表 $L$ ,同时统计每个模型的视图在 $L_0$ 中出现的次数,当此为准对 $L$ 中模型的匹配度进行如下计算:

$$Score_1 = \begin{cases} Score_0(1 + (t-1) \times a_1), & t < 3 \\ Score_0(1 + 3a_1), & t \geq 3 \end{cases} \quad (6)$$

其中, $Score_0$ 代表结果列表中每个模型的匹配度; $a_1$ 为固定常数,经过实验选取定值 0.1; $t$ 为列表中模型的视图在 $L_0$ 中出现的次数,在 $t$ 的临界点取 3 时得到的结果最为理想,如表 1 所列。

表 1 不同临界值下的 mAP 值  
Table 1 mAP at different critical values

| 临界值 | 1     | 2     | 3     | 4     |
|-----|-------|-------|-------|-------|
| mAP | 0.543 | 0.546 | 0.535 | 0.534 |

许多三维模型具有对称的特点,因此同一模型在部分视

点下提取到的视图可能具有一定的相似性,因此当某三维模型的不同视图同时被检索到时,通过以上计算来相应地提高该模型的匹配度。

在相似度匹配过程中,初次检索得到的结果列表本身会提供一些信息,比如检索结果列表中各个类别下的模型数量占列表长度的比例 $r$ 。通常 $r$ 越高,输入图像与该类别下的视图越相似,因此我们对 $r$ 值较高的模型提高匹配度,在原有匹配度 $Score_0$ 经过式(6)的运算得到 $Score_1$ 之后,统计出 $L$ 中每个模型所属的类别,并进行如下计算:

$$Score = \begin{cases} Score_1(1+i \times a_2), & \frac{i}{length} > 0.2 \\ Score_1, & \frac{i}{length} \leq 0.2 \end{cases} \quad (7)$$

其中, $length$ 为候选列表的长度, $i$ 为同类模型的出现次数, $Score$ 为最终的匹配度。

在 $a_2$ 的取值上(见表2),过低的取值难以对检索序列产生影响,而过高的取值会使结果排序过于依赖模型所属的类别,忽视视图本身的特性。例如,检索列表中有两类模型交叉分布,如果其中一类模型的数量稍多,那么该类模型会集中占据检索列表最前面的位置,如果出现数量稍少的类别反而属于目标模型的情况,则会对结果产生较为严重的干扰。通过实验检测 $a_2$ 的变化对检索结果产生的影响,发现当 $a_2$ 的取值增加至0.1时, $mAP$ 不再随之提升,同时通过观察检索结果发现,在结果列表中出现分属两类的模型数量相似的情况时,两种类别下的模型依然会交错分布,因此选取0.1作为 $a_2$ 的最终取值。

表2 不同 $a_2$ 取值下的 $mAP$ 值  
Table 2  $mAP$  at different  $a_2$  values

| $a_2$ | 0     | 0.001 | 0.01  | 0.1   | 1     |
|-------|-------|-------|-------|-------|-------|
| $mAP$ | 0.503 | 0.513 | 0.535 | 0.543 | 0.538 |

三维模型从不同角度获得的视图姿态迥异,因此同类模型在检索结果中出现的次数过少,可能是该模型仅仅在某个特殊视角下提取的视图与输入图像相似,而该模型本身并非目标模型。本文算法对这部分模型不做处理,而如果列表中某分类下的模型多次出现,则相应提高该分类下模型的匹配度,使其排列次序更加靠前。

在重排序部分,本文通过调整结果列表的相对位置,得到更为合理的检索结果展现顺序,使最终的结果列表更符合检索预期。实验证明,通过这种方式对检索结果进行重新排序后, $mAP$ 会提升约4%。

## 4 实验分析

本文算法是在Ubuntu平台下通过Keras深度学习框架实现的。实验所用GPU为GTX965,使用CUDA配合cuDNN库进行加速。本文使用的三维模型数据是从SHREC14<sup>[21-22]</sup>数据集中选取得到的,该数据集中不同分类下的模型数量差距悬殊,例如在飞机分类下有超过300个模型,而面包、骆驼分类下的模型数量甚至不足10个。在卷积神经网络训练的过程中,这种数据不均衡的状况对训练效果会产生较大的干扰,通过数据增强或者调节视图提取数量也难以

有效进行均衡,因此我们选择了分类下模型数量超过90的衣柜、椅子、吉他、直升飞机等20个分类下的2676个模型用于实验。

### 4.1 实验设置

首先根据每个三维模型在16个预定义视点下渲染出二维线条图,通过灰度熵过滤机制对每个模型筛选出平均12张较优视图,共同组成较优视图集,在其中选取26066张视图作为训练集,7752张视图作为验证集;然后将图像尺寸统一调整为 $150 \times 150$ ,使用深度卷积神经网络进行训练,输出视图提取1024维的全连接层,并通过PCA将维度降低至128维,使用该特征描述二维视图。同时,组织人员依据类别搜集20幅/类的自然图像作为输入源,用于对本文算法的检索效果进行实验。

此处以直升飞机和摩托车两种分类为例,进行直观的检索效果展示,这两种分类下的三维模型较为复杂,渲染出来的视图包含了丰富的细节,对于普通用户来说,以手绘草图作为输入源往往着重于边缘的勾勒,容易忽视较为细致的部分,而这部分信息的损失会对检索结果造成一定的负面影响。本文选择以自然图像作为输入源,提取包含更多细节的线条轮廓图,使用本文算法进行检索后,得到了如图6所示的检索结果。

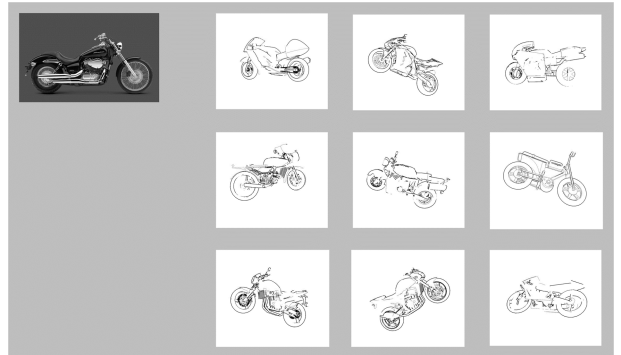


图6 摩托车分类下的检索结果

Fig. 6 Search results in motorcycles class

图6中得到的检索结果皆属于目标类别下的模型,而图7中出现的与检索目标不一致的模型的视图与输入图像在形状结构上也较为相似。此外,由于这两个模型所属的类别在检索列表中的占比较低,经过重排序后,其出现的位置也处于检索列表的末位。可以看到,本文算法在以自然图像作为输入源时,能够取得较为优秀的检索效果。

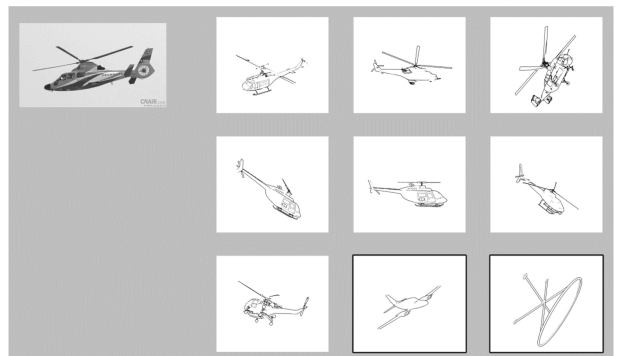


图7 直升飞机分类下的检索结果

Fig. 7 Search results in helicopter class

## 4.2 参数验证

在 PCA 降维过程中,本文通过实验量化了维度的不同对检索效果的影响,选取了多个特征维度进行对比,实验结果如表 3 所列。

表 3 不同维度下的  $mAP$  和  $AP$  值

Table 3  $mAP$  and  $AP$  under different dimensions

| 维度    | 1024  | 512   | 256   | 128   | 64    | 32    |
|-------|-------|-------|-------|-------|-------|-------|
| $mAP$ | 0.547 | 0.552 | 0.550 | 0.543 | 0.531 | 0.523 |
| $AP$  | 0.406 | 0.408 | 0.408 | 0.407 | 0.396 | 0.384 |

如表 3 所列,在特征维度降低至 128 之前,维度的降低对平均查准率的影响较小,波动幅度在 0.05 之内,而准确率相较 1024 维甚至有些微上升。因此,本文选取 128 维作为最终的特征维度,同时将降维前后以及检索列表重排序前后的平均查准率进行对比,结果如图 8 所示。

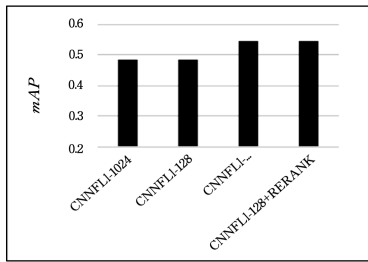


图 8 不同条件设置下的  $mAP$  对比

Fig. 8 Comparison of  $mAP$  under different conditions

可以看出,PCA 降维方法能够在有效保持检索的平均查准率的前提下,降低所使用的特征维数,从而提高检索效率,以满足大规模三维模型检索的需要。同时,我们也可以看出经过检索结果重排序后,检索的平均查准率有所提升。

## 4.3 对比实验

为了对本文检索方法进行评估与验证,本文选取文献[23]中使用的 Gabor 滤波器结合视觉词袋(Gabor-BOW)的算法、文献[24]中使用的 BF-SIFT 方法、Simonyan 等<sup>[7]</sup>提出的基于 ImageNet 数据集训练的 VGG16 网络模型(imgCNN)进行对比实验。在实验设置上,BF-SIFT 方法使用 sift 描述子结合视觉词袋进行特征提取,将词袋的聚类规模设置为 1000;Gabor-BOW 方法使用 Gabor 滤波器响应线条图边缘,并使用视觉词袋进行优化,词袋聚类规模与 BF-SIFT 保持一致;使用 imgCNN 替代本文的特征提取部分,在 ImageNet 预训练权重的基础上,基于本文数据集对全连接层进行重新训练,进而提取深度特征用于检索。各个分类查准率的对比情况如图 9 所示。

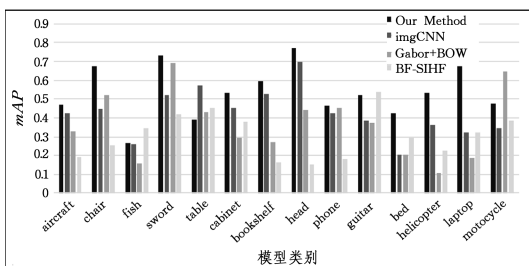


图 9 各个分类查准率的对比情况

Fig. 9 Comparison of  $mAP$  under different classes

通过对比发现,本文方法在椅子、头部、手机等大多数分类上取得了较好的检索效果,在鱼、床等分类上的效果一般,经过观察发现这些分类下线条提取的效果不够理想,因此影响到该分类下视图集的质量,下一步可以通过更为有效的轮廓提取方式进行相应的改进。而在平均查准率( $mAP$ )的对比中,本文方法的平均查准率达到了 0.543,高于基于 VGG16 方法的 0.395、BF-SIFT 算法的 0.293 以及 Gabor-BOW 算法的 0.319,获得了更为理想的检索效果。

同时,实验选取 Nearest Neighbor(NN),First-tier(FT),Second-tier(ST),E-Measures(E),Discounted Cumulated Gain(DCG),Average Precision(AP)共 6 个指标从不同角度对算法进行客观评估。

如表 4 所列,相较于其余 3 种算法,本文算法在整体上具有更优的检索表现。

表 4 不同指标下的量化比较

Table 4 Quantitative comparison under different indicators

|            | NN    | FT    | ST    | E     | DCG   | AP    |
|------------|-------|-------|-------|-------|-------|-------|
| Our Method | 0.473 | 0.321 | 0.493 | 0.126 | 0.622 | 0.407 |
| imgCNN     | 0.392 | 0.257 | 0.414 | 0.114 | 0.547 | 0.317 |
| Gabor+BOW  | 0.311 | 0.209 | 0.307 | 0.086 | 0.541 | 0.232 |
| BF-SIFT    | 0.243 | 0.168 | 0.319 | 0.080 | 0.463 | 0.209 |

**结束语** 当前研究者提出的三维模型检索方法通常以手绘草图或示例模型作为输入源,而本文选择简单背景下的自然图像作为检索的入口,利用互联网上海量的自然图像为检索者提供更丰富、更易获取的输入源,将三维模型检索问题转换为自然图像的线条轮廓图与三维模型的二维视图之间的匹配问题。同时,借助较优视图集的选取,通过深度卷积神经网络的有监督训练,有效地实现了三维模型检索方法。但当物体具有过多纹理或边缘特征模糊时,提取到的视图质量不够好,对训练效果有所影响。下一步的研究工作将会从优化视图渲染方法等方面展开,以进一步提高本文方法的应用范围以及检索效果。

## 参 考 文 献

- [1] SCHMIDHUBER J. Deep learning in neural networks: an overview[J]. Neural Networks the Official Journal of the International Neural Network Society, 2014, 61(1): 85-117.
- [2] SUN Z Y, LU C X, SHI Z Z, et al. Research and Progress of Deep Learning [J]. Computer Science, 2016, 43(2): 1-8. (in Chinese)  
孙志远, 鲁成祥, 史忠植, 等. 深度学习研究与进展[J]. 计算机科学, 2016, 43(2): 1-8.
- [3] LUO W, SCHWING A G, URTASUN R. Efficient Deep Learning for Stereo Matching[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016: 5695-5703.
- [4] ZHANG J, BAI C, NEZAN J F, et al. Joint motion model for local stereo video-matching method [J]. Optical Engineering, 2015, 54(12): 123108.
- [5] SHI M, XIE F, ZI Y, et al. Cloud detection of remote sensing images by deep learning[C]// Geoscience and Remote Sensing Symposium. IEEE, 2016: 701-704.

- [6] ZHANG J, SHANG J, ZHANG G. Verification for Different Contour Parameterizations Based on Integrated Satellite Observation and ECMWF Reanalysis Data[J]. *Advances in Meteorology*, 2017, 2017(1): 1-11.
- [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]// *International Conference on Neural Information Processing Systems*. Curran Associates Inc., 2012: 1097-1105.
- [8] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[EB/OL]. <https://cn.arXiv.org/abs/1409.1556>. 2014.
- [9] ALZU'BI A, AMIRA A, RAMZAN N. Content Based Image Retrieval with Compact Deep Convolutional Features[J]. *Neurocomputing*, 2017, 249(2): 95-105.
- [10] SALVADOR A, GIROINIETO X, MARQUES F, et al. Faster R-CNN Features for Instance Search[C]// *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE Computer Society, 2016: 394-401.
- [11] LIU Y, PAN Y, XIA R K, et al. FP-CNNH: A Fast Image Hashing Algorithm Based on Depth Convolution Neural Network [J]. *Computer Science*, 2016, 43(9): 39-46. (in Chinese)  
刘冶, 潘炎, 夏榕楷, 等. FP-CNNH: 一种基于深度卷积神经网络的快速图像哈希算法[J]. *计算机科学*, 2016, 43(9): 39-46.
- [12] ZHU Z, WANG X, BAI S, et al. Deep Learning Representation using Autoencoder for 3D Shape Retrieval[J]. *Neurocomputing*, 2016, 204(C): 41-50.
- [13] LI X X, CAO Q, WEI S. 3D object retrieval based on multiview convolutional neural networks[J]. *Multimedia Tools & Applications*, 2017, 76(19): 20111-201214.
- [14] 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*. 2015: 1875-1883.
- [15] DECARLO D, FINKELSTEIN A, RUSINKIEWICZ S, et al. Suggestive contours for conveying shape[C]// *ACM Siggraph*. ACM, 2003: 848-855.
- [16] DING L, GOSHTASBY A. On the Canny edge detector[J]. *Pattern Recognition*, 2001, 34(3): 721-725.
- [17] SANG M Y, SCHERER M, SCHRECK T, et al. Sketch-based 3D model retrieval using diffusion tensor fields of suggestive contours[C]// *International Conference on Multimedia 2010*. Firenze, Italy, DBLP, 2010: 193-200.
- [18] 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*. Eurographics Association, 2013: 49-56.
- [19] HONG Y, KIM J, HONG Y, et al. A 2D-View Depth Image and CNN-Based 3D Model Identification Method[J]. *Applied Sciences*, 2017, 7(10): 988-1001.
- [20] CLEVERT D, UNTERTHINER T, HOCHREITER S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)[EB/OL]. <https://cnarXiv.org/abs/1511.07289>. 2015.
- [21] LI B, LU Y, LI C, et al. A comparison of 3D shape retrieval methods based on a large-scale benchmark supporting multimodal queries [J]. *Computer Vision and Image Understanding*, 2015, 131(1): 1-27.
- [22] LI B, LU Y, LI C, et al. Shrec'14 track: extended large scale sketch-based 3D shape retrieval[C]// *Eurographics Workshop on 3D Object Retrieval*. 2014.
- [23] LIU Z, YIN S C, PAN X, et al. 3D model retrieval based on feature lines [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2016, 28(9): 1512. (in Chinese)  
刘志, 尹世超, 潘翔, 等. 基于特征线条的三维模型检索方法[J]. *计算机辅助设计与图形学学报*, 2016, 28(9): 1512.
- [24] HUA S, JIANG Q, ZHONG Q. 3D Model Retrieval Based on Multi-View SIFT Feature[M]// *Communication Systems and Information Technology*. Springer Berlin Heidelberg, 2011: 163-169.