

文章编号: 2095-2163(2019)03-0054-05

中图分类号: TP391.41

文献标志码: A

# 基于深度学习的三维模型检索研究

张 静, 曲志坚, 刘晓红

(山东理工大学 计算机科学与技术学院, 山东 淄博 255049)

**摘要:** 基于卷积神经网络在图像领域的发展, 本文研究了将卷积神经网络应用到三维模型检索领域。对三维模型预处理后, 选择6个投影角度把模型投影成6幅二维图像, 利用提取的视图作为神经网络的输入, 利用深度学习框架提取图像特征作为最终的模型描述符。之后通过比较2个模型多个视角的二维投影的相似度, 如果两者间相似, 则三维模型也是相似的, 再取多维视图的相似度平均值得到2个三维模型的最终相似度, 选择最终相似度最大的10个模型作为结果输出。充分利用二维图像领域性能优越的网络架构, 并且存在海量图像数据供深度学习模型进行预训练。从检索结果看出, 该方法能够得到较好的效果。

**关键词:** 卷积神经网络; 视图; 特征提取; 三维模型检索

## Research on 3D model retrieval based on deep learning

ZHANG Jing, QU Zhijian, LIU Xiaohong

(College of Computer Science and Technology, Shandong University of Technology, Zibo Shandong 255049, China)

**[Abstract]** Based on the development of Convolution Neural Network in the field of image, this paper studies the application of Convolution Neural Network to 3D model retrieval. After pre-processing the three-dimensional model, six projection angles are selected to project the model into six two-dimensional images. The extracted views are used as the input of neural network, and the image features are extracted by depth learning framework as the final model descriptors. Then the paper compares the similarity of the two models' two-dimensional projections from multiple perspectives. If the corresponding models are similar, then the three-dimensional models are similar. The paper takes the average similarity of the multi-dimensional views to get the final similarity of the two three-dimensional models, and select 10 models with the maximum similarity as the output of the results. It makes full use of the network architecture with superior performance in the field of two-dimensional images, and there is a huge amount of image data for depth learning model pre-training. From the retrieval results, it is demonstrated that this method can get better retrieval.

**[Key words]** CNN; view; feature extraction; 3D model retrieval

## 0 引言

随着计算机图形学的快速发展, 新的三维扫描设备和建模工具的出现使三维建模变得简单、快捷、高效; 显卡技术的不断进步则大幅提升了三维模型的显示和处理速度, 与此同时网络技术的深入系统研究更进一步推动了三维模型的传播和应用<sup>[1]</sup>。

互联网上已涌现出越来越多的三维模型<sup>[2]</sup>, 合理重复使用现有的三维模型, 对其进行编辑和修改以符合设计需要, 不仅能够节约成本、减少构建模型的工作量, 而且还将有助于三维模型的流通, 并对电子商务、虚拟现实等行业的发展有着重要的促进作用。这需要高效的三维模型检索系统作为支撑。根据文献[3-4]可知, 三维模型检索在各种领域已然得到广泛应用, 例如计算机辅助设计、数字娱乐、医疗诊断、电子商务和基于位置的手机应用。

三维模型检索是针对给定的查询模型, 检索出符合用户查询意图的模型集合的过程。为适应当前三维模型数量爆炸式增长和互联网搜索引擎规模与复杂性日趋增加的变化态势, 三维模型检索即已从基于文本关键字检索转变为基于内容的检索<sup>[3]</sup>。具体来说, 基于内容的三维模型检索首先从模型所包含的内容中用特征提取方法自动计算三维模型的特征, 然后在特征空间中计算待查询模型与目标模型之间的相似度, 实现对三维模型库的检索。该类检索方法的共同特点就是利用模型本身包含的低层特征, 然而人类是以视觉上的高层语义来理解模型内容的, 这就导致基于内容的检索效果尚未达到令人满意地步<sup>[5]</sup>。

三维模型特征提取作为三维模型检索的核心部分, 引起了国内外学者的广泛关注<sup>[6]</sup>。然而, 由于视点、光照强度、模型尺寸、模型样式等的变化, 却仍

**作者简介:** 张 静(1993-), 女, 硕士研究生, 主要研究方向: 深度学习、三维模型; 曲志坚(1980-), 男, 博士, 副教授, 主要研究方向: 机器学习技术; 刘晓红(1968-), 女, 教授, 主要研究方向: 图像处理技术。

收稿日期: 2019-01-21

然面临诸多颇具挑战性的困难。三维对象的表示对于三维模型识别和检索任务很重要。人工提取的特征主要用于表示3D对象。这些特征多是通过提取几何分布或直方图统计来获得,例如热核签名、旋转图像、FPFH<sup>[7-9]</sup>等。但是这种方法很难获得最佳的三维形状特征表示,而深度学习为特征学习提供了新的研究思路<sup>[10]</sup>。

## 1 算法流程

近年来,基于内容的图像检索方法在计算机视觉领域取得了突破性的进展<sup>[11]</sup>,这一切均需归功于深度卷积网络强大的非线性拟合能力。



图1 算法流程图

Fig. 1 Algorithm flow chart

### 1.1 预处理

三维模型所包含的信息自由度很高,其空间位置、尺寸和方位等信息都具有不确定性。用户在进行模型检索的时候,提供的模型的坐标系很有可能与数据库中的模型坐标系有着一定区别。例如示例模型和数据库中的模型在坐标系中的位置不同、或尺寸不一致、或对于坐标系旋转的角度不同。如果未做处理直接比较就可能导致本是相同的模型却被检索系统视为不同的模型,会影响检索系统检索的准确性<sup>[13]</sup>。

本文采用主成分分析法(Principle Component Analysis, PCA)对模型进行预处理。三维模型点集协方差矩阵计算对应的特征值 $\lambda_1$ 、 $\lambda_2$ 和 $\lambda_3$ ,使得 $\lambda_1 > \lambda_2 > \lambda_3$ ,对应的特征向量为 $I_1$ 、 $I_2$ 、 $I_3$ ,以 $I_1$ 、 $I_2$ 、 $I_3$ 为新坐标系,把三维模型调整到该新坐标系。

经过变换矩阵变换后的三维模型顶点分布最广泛的方向处于第一主轴上,是第一主成分,顶点分布次广泛的方向处于第二主轴上,是第二主成分,依此类推,得到第三主成分。这样就使模型分布最广泛的方向都分布在 $x$ 、 $y$ 、 $z$ 三个坐标轴上。

提取特征前对模型进行同向性处理,使得任意方向、任意比例的模型达到归一,提高系统的准确

性,变换矩阵如式(1)所示<sup>[5]</sup>:

$$\begin{cases} C_p = \left[ \sum_{i,j=1}^n (p_i - q_j) (p_i - q_j)^T \right]^{1/2}, \\ Q = C_p P. \end{cases} \quad (1)$$

### 1.2 提取三维模型视图

基于二维投影的三维模型特征识别的基本原理为:如果2个三维物体多个视角的二维投影看起来相似,那么两者在三维上也是相似的。分析可知,接下来则需提取三维模型的二维投影<sup>[14]</sup>。

观察者的位置(视点)确定坐标区的方向。依据方位角和仰角或者三维空间中的点来指定视点。

研究中,采用 $view(az, el)$ 来设置三维图的视角。其中,方位角 $az$ 是从负 $y$ 轴围绕 $z$ 轴水平旋转的角度(以度为单位)。正值表示按逆时针方向旋转视点。 $el$ 是视点的垂直仰角(以度为单位)。仰角为正值,对应于在对象上方移动;为负值,对应于在对象下方移动。观察视角的形象展示见图2。图2中,箭头指示正方向。

本文中选择了 $[-135, 45; 135, 45; 45, 135; 45, -135; 0, 45; 0, 90]$ 这6个角度的视图进行提取。

研究中给出了一个模型的6个视角的二维图像如图3所示。

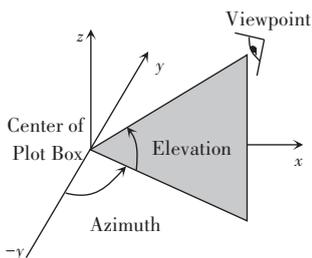


图2 观察视图角

Fig. 2 Viewing angle diagram

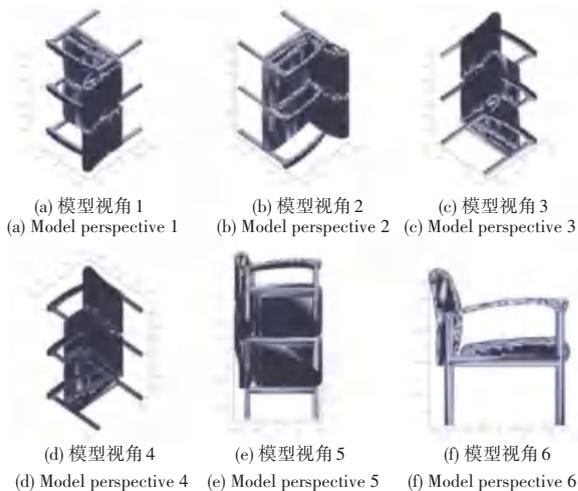


图3 不同视角的六视图

Fig. 3 Six views from different perspectives

依据前文所述,在将数据库中每一个三维模型成功提取视图后,本次研究即将所有三维模型的六视图作为卷积神经网络模型的输入。

### 1.3 卷积神经网络提取特征

卷积神经网络(Convolutional Neural Network, CNN)是人工神经网络中含有多个隐含层以及输入层的一种网络,其中的权值网络结构更为接近人脑的生物结构,同时也降低了模型的复杂度和权值的数目。这种设计减少了学习参数,与以往的神经网络相比较,在整体上提高了训练性能<sup>[15]</sup>。

CNN 分类的网络架构大体上可分为卷积层、池化层、全连接层和输出层。网络设计流程如图4所示。



图4 卷积神经网络流程图

Fig. 4 Convolutional Neural Network flowchart

由图4可知,设输入图像 $I$ ,其在CNN中的研发处理步骤可阐释解析如下。

(1)输入层(Image Layers):用于图像输入。本文建立的网络中,输入层大小为 $640 * 360 * 3$ 。

(2)卷积层(Convolutional Layers):用于图像的卷积计算,将图像数据进行映射,得到更高纬度的特征。在本次实验中,卷积核大小为 $5 * 5$ 。研究推得其数学公式可表示为:

$$f(x^l) = a^l[\delta(x^{l-1}) + b^l], \quad (2)$$

其中, $l$ 表示层数; $a^l$ 表示该层的激活函数; $\delta(I)$ 表示对图像的卷积操作; $b^l$ 表示该层的偏置; $x^l$ 表示经过卷积层操作后得到的特征图。

(3)池化层(Pooling Layers):用于卷积层输出的特征向量进行降维处理,以减少数据量,提高训练速度,加快收敛。研究推得其数学公式可表示为:

$$f(y^l) = \max(x^{l-1}) \text{ 或 } \text{avg}(x^{l-1}), \quad (3)$$

(4)正则(ReLU):用于在训练过程中对损失函数(loss function)进行修正,通过影响特征向量对每一个输入数据的修正,使得网络训练不至于过拟合或者欠拟合,使网络权重的值尽可能地小,具有较强的泛化能力。本此构建网络中,正则则为L2范数<sup>[16]</sup>。

(5)全连接层(Full Connect Layers):用于特征提取,利用激活函数直接得到网络训练后的特征。本实验中,全连接层为100维。研究推得其数学公式可表示为:

$$f(z^l) = a^l(W^l z^{l-1} + b), \quad (4)$$

其中, $W^l$ 表示对应的权重。

(6)输出层(Output Layer):对前一层的特征向量进行处理,得到最终分类,用于多分类器输出,将全连接层的输出映射到 $(0, 1)$ ,输出softmax值对应的类别标签。研究推得其数学公式可表示为:

$$\sigma(z_i^{l-1}) = \frac{e^{z_i}}{\sum_1^N e^{z_i}}. \quad (5)$$

至此,研究得出本文的网络部分结构,可做如下设计定义:

```
layers =
[ imageInputLayer([ 640 360 3]);
convolution2dLayer( 5, 20, 'Padding', [ 2 2 ],
'Stride', [ 1,1]);
reluLayer();
maxPooling2dLayer(2, 'Stride', 2);
fullyConnectedLayer( 100);
dropoutLayer(.4);
fullyConnectedLayer(35);
softmaxLayer();
classificationLayer(); ]
```

本文搭建的深度卷积神经网络,简称V9。在V9网络中没有过多的卷积和池化层,主要是由于数

据本身的特殊性,对此可描述为:数据本身纹理并不复杂,易于辨识,因而无需使用大规模的卷积层来提取更深层的特征,否则就会大大延长训练时间,增加网络的复杂度。

基于上述搭建模型,研究利用卷积神经网络提取每张图片的最终特征。V9 在图片数据集中训练好的效果作为三维模型每张视图的特征器。V9 网络结构如图 5 所示。提取得到的特征如图 6 所示。

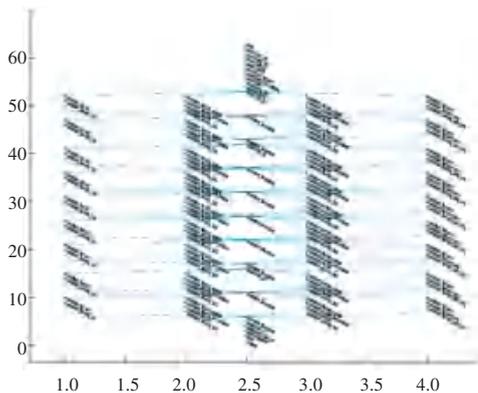


图5 网络结构图

Fig. 5 Network structure diagram

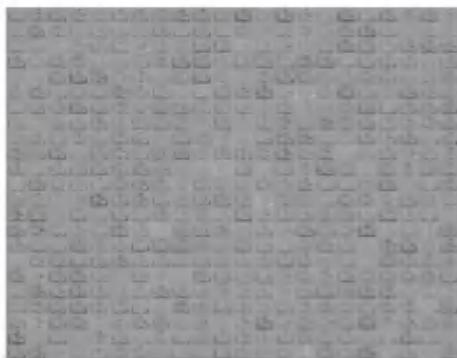


图6 提取的特征视图

Fig. 6 The extracted feature view

#### 1.4 相似度测量

对检索者提供的检索样本进行特征提取后,把提取到的特征与模型库里的模型特征进行比较,计算相似的程度,同时输出与输入样本特征最相似的前若干个模型。这个比较的过程就是相似性度量<sup>[17]</sup>。

基于卷积神经网络学习到的每张视图的特征,研究运算不同二维图像的相似度,利用相似度测量方法强调特征提取的效果。输出的三维模型的特征与输入样本的特征有较近的距离。通过距离度量函数来计算输入模型特征与模型库中所有模型特征向量之间的距离,并根据距离从小到大进行排序返回给用户。排序结果中,排位在前的模型具有较小的距离,因此与用户的输入模型越相似。

这是三维模型相似度计算过程中最常用的度量方法。距离计算中,最常用的就是欧式距离。假设任意 2 个模型的特征表示为  $X = (x_1, x_2, \dots, x_n)$  和  $Y = (y_1, y_2, \dots, y_n)$ , 其中  $n$  为特征向量的维数。由此可得 2 个特征间的欧式距离计算公式为:

$$D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (6)$$

先计算所有的相似度,然后取出每一个  $6 * 6$  方阵中主对角线的元素,作为每个模型对应的 6 个视图的相似度,再计算平均值得到最终的相似度,降序排列得到前 10 个相似度最大的检索模型。

## 2 实验分析与结果

为了更好地提取三维模型的特征,普林斯顿大学等机构提供了 2 个大规模的三维模型数据集,分别为 ShapeNet 数据集和 ModelNet 数据集。其中,ModelNet 数据集包含了来自 662 类的 127 915 个三维模型,相应内部子集 Model10 包含了来自 10 类的 4 899 个三维模型,ModelNet40 包含了来自 40 类的 12 311 个三维模型。本文选取 ModelNet40 的模型作为实验数据库。

实验平台和开发工具有:中央处理器(CPU)为 Intel(R) Xeon E5-2650 2.0 GHz,内存为 64 GB,显卡 1080Ti,编程环境为 Python3.2,基于 Tensorflow 框架搭建的实验验证本文方法的有效性。

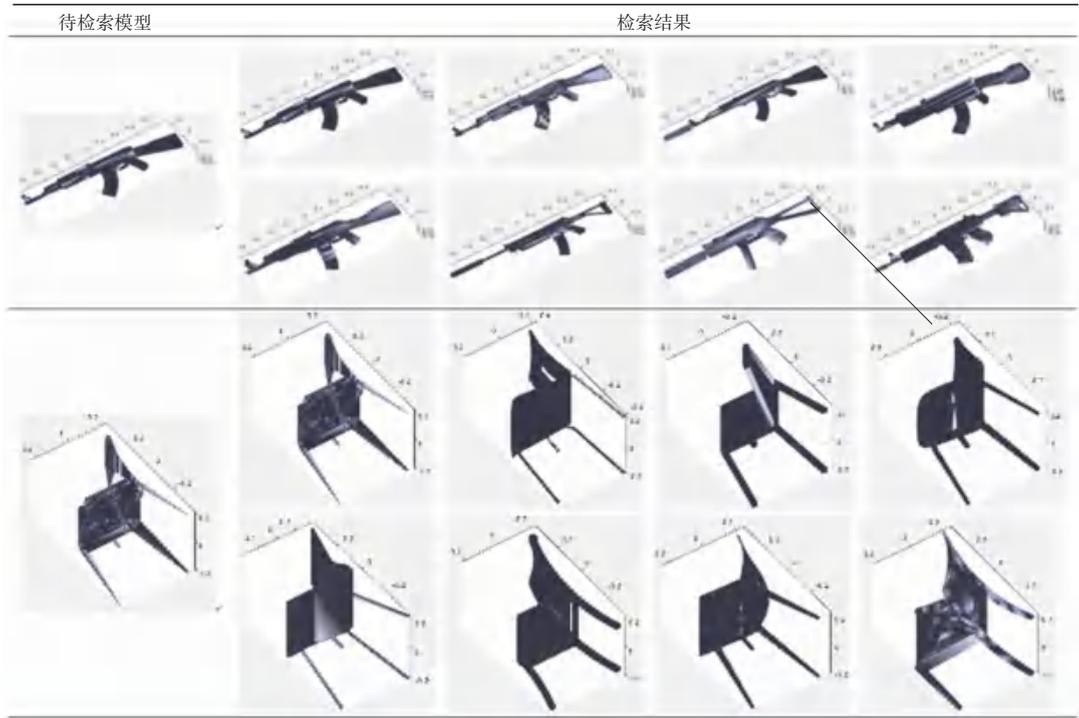
为了清晰观察检索效果,本文随机选取了数据集中的任意 2 个模型作为输入,计算选取模型与相应数据集中的其余模型之间的相似度,对最相似的 8 个模型进行可视化输出,位置最上面的表示相似程度越高,检索结果详见表 1。

实验结果证明了本文方法的有效性。从检索结果可看出,利用卷积神经网络提取特征的检索效果较好,所检索出的模型和待检索模型的相似度很高。

## 3 结束语

基于深度学习的提取算法,研究先将三维模型转换为一系列的二维图像,把三维模型从三维空间降到二维空间来进行处理,大大降低了特征处理和相似度度量的复杂度,且容易提取和计算。再采取深度学习方法提取图像特征作为最终的模型描述符。充分利用普林斯顿大学的数据集进行了模拟三维模型检索的可视化展示,可以看到,使用卷积神经网络在一定程度上能够得到不错的检索效果。在未来工作中,将继续探索三维模型的特征描述符,并尝试将本文中方法引入到实际的应用中。

表1 检索结果  
Tab. 1 Searched result



## 参考文献

- [1] 张晓东. 三维模型的形状特征提取方法研究[D]. 北京:中国石油大学, 2010.
- [2] TANGELDER J W H, VELTKAMP R C. A survey of content based 3D shape retrieval methods [J]. *Multimedia Tools and Applications*, 2008, 39(3): 441-471.
- [3] GAO Yue, ZHANG Hanwang, ZHAO Xibin, et al. Event classification in microblogs via social tracking [J]. *ACM Transactions on Intelligent System Technology (TIST)*, 2017, 8(3): 1-14.
- [4] ZHAO Sicheng, CHEN Lujun, YAO Hongxun, et al. Strategy for dynamic 3d depth data matching towards robust action retrieval [J]. *Neurocomputing*, 2015, 151(Part II): 533-543.
- [5] 李朋杰. 面向三维模型检索的特征提取算法研究[D]. 北京:北京邮电大学, 2012.
- [6] BIASOTTI S, CERRI A, BRONSTEIN A, et al. Recent trends, applications, and perspectives in 3D shape similarity assessment [J]. *Computer Graphics Forum*, 2016, 35(6): 87-119.
- [7] GAO Yue, WANG Meng, TAO Dacheng, et al. 3D object retrieval and recognition with hypergraph analysis [J]. *IEEE Transactions on Image Processing*, 2012, 21(9): 4290-4303.
- [8] GAO Yue, ZHEN Yi, LI Haojie, et al. Filtering of brand-related microblogs using social-smooth multiview embedding [J]. *IEEE Transactions on Multimedia*, 2016 18(10): 2115-2126.
- [9] ZHAO X, WANG N, ZHANG Y, et al. Beyond pairwise matching: Person reidentification via high-order relevance learning [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(8): 3701-3714.
- [10] 孙志军, 薛磊, 许阳明, 等. 深度学习研究综述[J]. *计算机应用研究*, 2012, 29(8): 2806-2810.
- [11] 席晓聪. 图像分类方法研究[D]. 济南:山东大学, 2013.
- [12] 郭裕兰. 三维形状数据的深度特征表示[J]. *CCF 计算机视觉专委简报*, 2017(2): 8-11.
- [13] 潘翔, 张三元, 张引, 等. 一种基于拓扑连接图的三维模型检索方法[J]. *计算机学报*, 2009, 27(9): 1250-1255.
- [14] WANG Dong, WANG Bin, ZHAO Sicheng, et al. View-based 3D object retrieval with discriminative views [J]. *Neurocomputing*, 2017, 252: 58-66.
- [15] HU Fan, XIA Guisong, HU Jingwen, et al. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery [J]. *Remote Sensing*, 2015, 7(11): 14680-14707.
- [16] YANG Songfan, RAMANAN D. Multi-scale recognition with DAG-CNNs [C]//2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 1215-1223.
- [17] 杨荣, 冯有前, 袁修久. 利用现有模型修复不完整三维模型[J]. *计算机辅助设计与图形学学报*, 2015, 27(1): 98-105.