



计算机应用研究  
Application Research of Computers  
ISSN 1001-3695, CN 51-1196/TP

## 《计算机应用研究》网络首发论文

题目: 基于多模态融合的三维模型检索算法研究  
作者: 王亚, 郑博文, 张欣  
DOI: 10.19734/j.issn.1001-3695.2020.02.0086  
收稿日期: 2020-02-12  
网络首发日期: 2020-09-30  
引用格式: 王亚, 郑博文, 张欣. 基于多模态融合的三维模型检索算法研究[J/OL]. 计算机应用研究. <https://doi.org/10.19734/j.issn.1001-3695.2020.02.0086>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 基于多模态融合的三维模型检索算法研究

王 亚<sup>1</sup>, 郑博文<sup>1†</sup>, 张 欣<sup>2</sup>

(1. 天津大学 自动化与信息工程学院, 天津 300072; 2. 中国电子技术标准化研究院, 北京 100007)

**摘要：**随着计算机技术的发展，三维模型在多媒体和计算机视觉领域获得了广泛的关注。三维模型检索和分类成为当今研究热点问题。为了获得更好的三维模型检索分类性能，该文基于深度学习模型研究了多模态信息融合对三维模型的特征描述，在训练步骤提出相关性损失函数来指导不同模态之间的训练，提取更稳健的特征向量。最后将融合特征应用于三维模型的检索和分类，在 ModelNet40 数据集上进行了三维模型分类任务和检索任务评估。实验结果和与现有方法进行了对比证明了该方法的优越性，这为三维模型检索分类领域提出了一种新的思路。

**关键词：**三维模型；多模态；深度学习

**中图分类号：**TP37      doi: 10.19734/j.issn.1001-3695.2020.02.0086

## 3d model retrieval algorithm based on multimodal fusion

Wang Ya<sup>1</sup>, Zheng Bowen<sup>1†</sup>, Zhang Xin<sup>2</sup>

(1. School of Electrical & Information Engineering, Tianjin University, Tianjin 300072, China; 2. China Electronics Standardization Institute, Beijing 100007, China)

**Abstract:** With the development of computer technology, 3D models gained wide attention in the fields of multimedia and computer vision. 3D model retrieval and classification became a research hotspot. In order to get better performance of 3D model retrieval and classification, this paper proposed the characterization of 3D models by multi-modal information fusion, which based on deep-learning model. The method could consider the correlation of different modalities in the training step for extracting a more robust feature vector, which benefit from the proposed correlation loss function. In addition, this paper applied fusion features to 3D model retrieval and classification. This paper evaluates the proposed method on the ModelNet40 dataset for 3D models classification task and retrieval, the comparison between the experimental results and the existing methods proves the superiority of this method, which provides a new idea for the field of 3D model retrieval and classification.

**Key words:** 3d models; multi-modal; deep-learning

## 0 引言

随着数字化技术和计算机视觉技术的发展，三维模型被广泛的应用于人们的日常生活之中，例如计算机辅助设计，医学诊断，生物信息学，3D 打印，医学成像和数字娱乐等等。如何理解三维模型、识别三维模型愈发称为人们研究的热点。目前已经提出了多种表示三维模型的深度学习模型，例如多视图、点云、草图和全景图等。

目前相关领域涌现了许多优秀的深度学习模型：MVCNN<sup>[1]</sup>通过提取三维模型在不同视角下的二维视图，结合多视图特征信息组合成单一简洁的三维特征描述符。PointNet<sup>[2]</sup>使用对称函数和最大池化去标记点云的关键点来获取点云的全局特征，PointNet++<sup>[3]</sup>能够在不同尺度提取局部特征，通过多层网络结构得到点云的深层特征。PANORAMA<sup>[4]</sup>通过使用 SYMPAN 方法连续姿势归一化三维模型获得特征表示，全景视图由 3 通道图像组成，包含空间分布图，法线的偏差图和法线的偏差梯度图<sup>[5]</sup>。草图模型利用草图信息来表示三维模型。Yoon 等<sup>[6]</sup>提出了一种新颖的网络来提取三维模型表示的草图信息。可以有效地处理由于比例变化引起的一些形状信息变化。然而，这些方法只关注单模态的三维数据而忽略训练步骤中这些模态之间的相关性。为了综合不同模态网络的优势，获取更具鲁棒性的特征，本文提出利用多种模态联合来获得三维模型的特征。

在本文之前就有相关学者进行了多模态的研究，PVNet<sup>[7]</sup>

提取模型的二维关键点，利用二维和三维的对应，通过 PnP 计算出物体的 6D 特征。晋忠孝等人<sup>[8]</sup>提出通过语义分割网络对目标物体提取 RGB 图片和点云数据，并进行融合特征预测目标物体的 6D 姿态。PVRNet<sup>[9]</sup>利用模型的点云和单一视图之间的相关性指导来自点云的特征和视图的特征的融合。这些方法在三维模型检索和分类领域都有良好的表现。然而均只能解决特定模态之间的融合，本文提出了一个更具普适性的框架，可以综合不同模态的特性，以获取更具鲁棒性的特征表示。

本文认为，一个三维模型可以通过不同的模态进行表示，因此可以联合不同的模态来学习三维模型的特征描述符。虽然特征向量是通过不同的模态学习得到的，但均表示的是同一个三维模型，故而包含有相同或者相似的部分。因此，这些特征向量可以在训练步骤中彼此指导以获得更稳健的特征表示。为了证明这一想法，本文提出了一种用于三维模型分类和检索的新型多模态联合网络。首先，提取每个三维模型三个模态的特征。其中点云模态用来描述结构信息、多视图模态用来描述视觉信息，PANORAMA 模态用来表示三维模型的表面信息；其次，本文提出了一种新的相关性损失函数，可以有效地减轻不同模态的特征分布差异以获取更稳健的特征表示，并在训练步骤实现不同模态的相互指导，提高了学习效率；最后，本文提出了一种基于统计思想的融合方法以融合不同模型分类的结果，用于最终的三维模型分类和检索。

收稿日期：2020-02-12；修回日期：2020-05-17

**作者简介：**王亚(1995-)，男，河北唐山人，硕士研究生，主要研究方向为计算机视觉；郑博文(1994-)，男(通信作者)，山西太原人，硕士研究生，主要研究方向为计算机视觉(zbw19940723@163.com)；张欣(1993-)，女，山东济南人，工程师，硕士研究生，主要研究方向为智能制造、人工智能。

## 1 基于多模态融合的三维模型检索算法

### 1.1 数据处理

本文提出的多模态网络中单一模态的选取是为了获取三维模型更多样的特征表示。其中点云模态是对原始三维模型的特征提取方法, 描述了三维模型的结构特征; 多视图模态是对三维模型的二维视图采样进行特征提取, 描述了模型的视觉特征; 全景图模态, 是基于三维模型的全景图的特征提取, 描述了模型的表面信息。这三种信息包含了当今三维模型研究的主流特征, 对三种特征进行多模态训练, 以期获得更具鲁棒性的特征表示。

本文算法研究的是基于多模态的网络融合, 因此根据不同模态网络的需求, 进行数据的格式化处理。

多视图数据(MV 模态): 采用 NPCA<sup>[10]</sup>中的方法对三维模型进行标准化。之后利用 OpenGL<sup>[11]</sup>开发的可视化工具从三维模型中提取一组视图, 具体方法是以三维模型的 Z 轴为中心, 每间隔 30° 对其进行视图采样, 每一个三维模型都可以获得一个 12 张视图的视图集来表示三维模型的视觉信息。

点云数据(PC 模态): 采用 MeshLab<sup>[12]</sup>将 PLY 模型转换成点云数据(PCD)。由于数据集中模型的大小不均匀, 模型表面的网格密度也不同。因此先需要加载模型, 通过使用螺旋细分算法<sup>[13]</sup>对网格细分, 并间接增加点云点数。之后通过对模型进行采样, 每个模型采样获得 1024 个点的点云数据<sup>[14]</sup>。

全景视图数据(PV 模态): 三维模型的全景视图可以将三维模型的表面信息转换为二维图像信息<sup>[4,15,16]</sup>。全景视图通过将三维模型置于到半径为 R, 高度为 H = 2R 圆柱体的中心, 将其表面投影到圆柱体侧表面上获得的。三维模型轴线平行于 X、Y、Z 的主轴之一都可以获取初始全景图。其中 R 的值设定为 3 \* dmax, dmax 是模型表面与其质心的最大距离, 采样速率为 128。圆柱形投影用于捕捉三维模型表面上的两种不同特征:

a). 模型表面在三维空间中的位置信息作为模型的空间分布图(SDM)。

b). 模型表面的方向信息作为法线偏差图(NDM)。

对 NDM 图求解梯度获得表面方向的梯度图, 也就是 Magnitude of Gradient。将以上三个单通道黑白图像作为 RGB 图像的三个通道, 获得三维模型的 3-channel 图。因此对于每个三维模型, 可以得到 12 个不同的视图, 如图 1 所示。

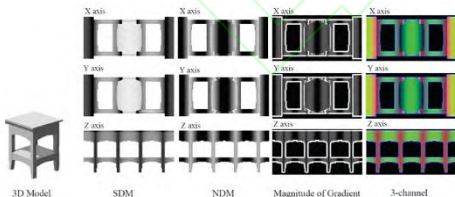


图 1 三维模型的 12 张全景视图

Fig. 1 12 panoramic views of the 3D model

### 1.2 多模态联合网络

基于多模态三维模型数据, 本文提出了一种新的多模态联合网络。图 2 显示了本文方法的详细框架。所提出的网络由三个模态组成: 点云模态, 利用流行的 PointNet 模型进行特征提取; 多视图模态, 利用 MVCNN 模型来实现特征提取功能, 全景图模态, 利用 PANORAMA-MVCNN 学习三维模型的特征向量。本文认为: 基于不同模态的特征应该是相似的, 描述的是相同 3D 模型。为了证明该假设, 设计的三个模态网络输出三维模型特征向量的维度相同为  $f_c \in \mathbb{R}^{1024}$ 。本文所提出的方法同时训练这三个神经网络, 训练时包含两种损失函数: 每个模态的传统判别损失函数和多模态相互指导的相关损失函数。

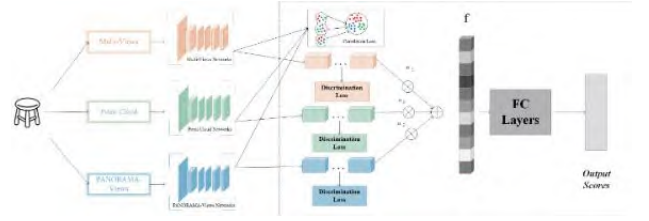


图 2 多模态融合算法系统流程图

Fig. 2 Flow chart of multimodal fusion algorithm system

在传统的单模态网络训练步骤中, 损失函数的作用在于最小化所提取特征的类内距离, 并且使所提取特征的类间距离最大化, 这种损失被称为交叉熵损失, 该损失函数的定义如下:

$$L_d = -\sum_{j=1}^n y_j \log \prod_{i=1}^K p_j(\beta_j | \beta_1, \beta_2, \dots, \beta_K) \quad (1)$$

其中  $y_j = [y_1, y_2, \dots, y_K]$  是来自  $j$  类样本的真实标签,  $k$  是所有类别的总数,  $p_j$  是 softmax 层输出的类别分布概率,  $\beta_j$  的概率预测是根据模态的特征定义的, 如下所示。

$$p_j(\beta_j | \beta_1, \beta_2, \dots, \beta_K) = \frac{e^{\beta_j}}{\sum_{i=1}^K e^{\beta_i}} \quad (2)$$

目前许多流行的算法都应用了交叉熵损失<sup>[17~19]</sup>, 并且获得了优异的结果。在本文的方法中, 对于每个模态网络都保留了原网络模型中的交叉熵损失函数, 以保证原本网络结构的优异性能, 此外引入一种新的相关性损失函数, 以确保在训练过程中多个模态之间可以相互指导, 提高网络训练的学习速度, 并提高最终特征向量的鲁棒性。本文提出的相关性损失函数公式如下:

$$L_c(M_i, M_j) = \|\xi(f_{M_i}) - \xi(f_{M_j})\|_2 \quad (3)$$

在这里, 本文将两个不同特征向量之间的距离度量的 2 范数用来表示两个不同模态特征向量的相关性, 其中  $f$  表示由不同模态网络提取的特征向量,  $M$  的下标代表的是第 1, 2, 3 个模态的数据,  $\xi = \text{sigmoid}(\log(\text{abs}(\bullet)))$  是一个归一化激发函数<sup>[20,21]</sup>。在训练过程中, 相关性损失的值逐渐减小, 表明在训练过程中不同模态特征相互指导。这会加快训练的收敛速度, 获得更具有鲁棒性的特征向量。以模态  $M_1$  为例, 基于这种相关性损失函数的设计, 不同模态网络的最终损失函数如下:

$$L_{M_1} = L_{d,M_1} + L_c(M_1, M_2) + L_c(M_1, M_3) \quad (4)$$

其中  $L_{d,M_1}$  是基于单模态的交叉熵损失,  $L_c(M_1, M_2)$  和  $L_c(M_1, M_3)$  分别代表了模态  $M_1$  与模态  $M_2$  和  $M_3$  的相关性损失。最后通过随机梯度下降的反向传播优化这三个单模态网络。

### 1.3 多模态特征融合

正如 1.2 所述, 本文的多模态网络包含基于点云、视图和 PANORAMA 三种。通过三模态卷积神经网络的联合学习, 可以得到三种基于不同模态的三维模型的特征。区别于传统的采用池化操作<sup>[1,3,22~25]</sup>的特征融合方法, 本文实验中基于统计方法, 采用加权融合方法融合三个特征向量。本方法的框架如图 3 所示。

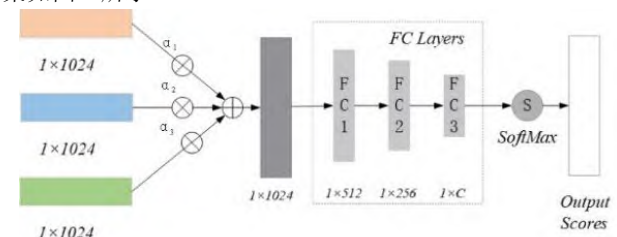


图 3 特征融合框架图

Fig. 3 Feature fusion frame



具体公式如下:

$$f = \sum_{i=1}^3 \alpha_i \xi(f_{M_i}) \quad (5)$$

$$\sum_{i=1}^3 \alpha_i = 1 \quad (6)$$

其中  $f$  代表的是在不同模态下提取的特征向量,  $\alpha_i$  表示的是不同模态的权重, 对加权融合特征的特征输入到全连接层 (FC 层), 全连接层的维度依次为 512, 256, C。C 代表数据集类别的数量, 最后通过一个 softmax 层获得三维模型的分类概率分布。

## 2 实验与结果分析

### 2.1 实验数据集

为了评估本文提出的分类算法, 本文在公开的流行数据集 ModelNet 上进行评测。该数据集有两个版本组成, 分为: ModelNet10 和 ModelNet40。其中 ModelNet10 包含 4899 个 CAD 模型, 包含 10 个类别。此外该数据集的训练集和测试集是官方规定好的, 分别有 3991 和 908 个模型。ModelNet40 包含 12311 个 CAD 模型, 拥有 40 个类别。ModelNet40 训练集和测试集分别有 9843 个 2468 个。该数据集经过人工处理, 剔除了错误分类的模型。值得一提的是 ModelNet10 模型经过平移和旋转方面的标准化, 而 ModelNet40 的模型没有。

### 2.2 相关性损失函数有效性实验

在本文中, 提出了一种新的应用于全局损失的相关性损失函数, 可以实现每个模态在训练中进行相互的指导, 以此获得更健壮的特征表示。比较了包含相关损失函数作用下与传统单模态网络交叉熵损失函数下的收敛趋势。实验结果如图 4 所示。这里只展示了 100 次迭代的收敛趋势, 本实验中的三个模态在前 100 次迭代的收敛趋势可以明显看出, 在全局损失中加入相关性损失函数, 判决损失的收敛速度明显加快。此外最终的分类结果也明显优于传统的交叉熵损失网络的结果。相关结果如表 1 所示, 该实验证明了本文的方法的可靠性和优越性。

表 1 不同模型分类准确率比较

Tab. 1 Comparisons of different model classification accuracy

method	classification accuracy	
	ModelNet10	ModelNet40
MV	89.01%	87.23%
PC	93.28%	89.70%
PV	87.33%	83.10%
MV + PC	92.29%	91.28%
MV + PV	90.41%	89.38%
PC + PV	93.61%	89.95%
MV + PC + PV	93.83%	92.46%

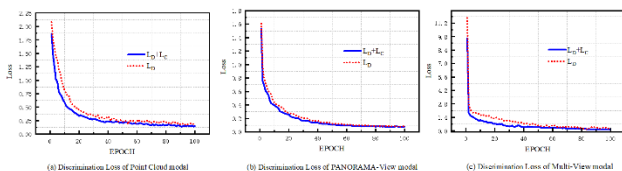


图 4 多模态网络 loss 对比图

Fig. 4 Comparison chart multimodal network loss

表 1 中的数据将单一模态的特征和任意双模态特征融合网络与本文多模态网络的分类准确率进行对比。其中 MV 表示单一多视图模态, PC 表示单一点云模态, PV 表示单一的全景图模态。其中“+”表示是两种模态的融合, 融合方法与三模态融合方法相同, 且使用了相同定义的相关性损失函数。

### 2.3 多模态特性相似度度量对比实验

在上文 1.2 节中提到, 对于不同模态的特征向量采用 2 范数进行度量。为验证 2 范数进行相似度度量的合理性, 本节采用对比实验, 将 2 范数距离 (ED), 余弦距离 (CS), 曼哈顿距离 (MD) 应用于所提方法中, 实验结果如表 2 所示。

表 2 不同相似度度量方法对比

Tab. 2 Comparisons of different similarity measurements

methods	classification(acc)	retrieval(mAP)
ED	93.8%	89.8%
CS	93.7%	89.7%
MD	93.5%	89.7%

实验结果表明不同度量方式, 在方法中效果相差不大, 这也说明了本文提出的网络的鲁棒性。其中 2 范数度量的结果较好, 因此本文采用了 2 范数作为相似度度量方式。

### 2.4 多模态信息融合实验

根据前两节的分析, 为了利用每种模态的优势, 本文基于统计的方法, 采用加权融合的方式对不同模态的特征进行融合。为了定义不同模态网络的权重, 本文对不同的融合参数值进行采样以找到最佳值。统计实验结果如下图 5 所示, 该图显示了不同权重下的分类的准确率。根据实验结果本文将融合方法中的参数设置为 0.7, 0.2, 0.1, 分别对应点云模态, 多视图模态和全景图模态的权重。本实验表明, 点云模态具有最大的权重。

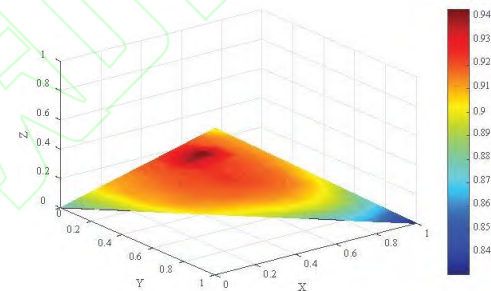


图 5 不同模态权重下的分类准确率图

Fig. 5 Classification accuracy graphs with different modal weights

同时, 本文对不同的特征融合算法组合进行了对比评测, 并在表中展示了对比实验的结果。如表 3 所示, 其中“Late Fusion”表示三个模态网络的特征拼接后的分类结果。“Our”表示本文权重融合方法的分类结果。由表可以发现: 本文的方法在平均分类准确率和全局分类准确率方面明显优于其他方法。也就是说本融合方法可以有效地利用每个模态数据的优势来获得最佳性能。

表 3 不同特征融合方法在分类任务上的比较

Tab. 3 Comparisons of different feature fusion methods on classification

methods	mean class accuracy	overall class accuracy
MV	85.68%	87.23%
PC	87.27%	89.70%
PV	82.30%	83.10%
Late Fusion	90.53%	92.46%
本文算法	92.24%	93.82%

### 2.5 分类和检索对比实验

为验证本文所提算法的优越性, 本文在 ModelNet40 数据集上与当今流行的算法进行对比实验。

在实验中, 将本文算法与基于不同表示的各种算法进行了比较, 包括基于体素的算法模型<sup>[26]</sup>, 基于多视图的人工描述符<sup>[27,28]</sup>, 基于多视图数据的深度学习模型<sup>[1,22]</sup>, 基于 PANORAMA-Views<sup>[23]</sup>的深度学习模型和基于点云的算法模型<sup>[2,3,29-31]</sup>。

分类和检索实验具体结果如表 4。

在分类任务中,可以看出,本文提出的算法的效果最好,分类准确率为 93.8%。均高于当今流行的单一模态下的各个算法。同时本文与双模态的 PVNet 算法进行对比,依然具备 0.6%的领先优势,与 PVRNet 对比有 0.2%的领先优势。

表 4 在 ModelNet40 数据集上的分类准确率的比较

Tab. 4 Comparisons of classification accuracy on modelnet40

method	pre train	classification(overall acc)	retrieval(mAP)
(1)SPH <sup>[28]</sup>	-	68.2%	33.3%
(2)LFD <sup>[27]</sup>	-	75.5%	40.9%
(3)3D ShapeNets <sup>[26]</sup>	ModelNet40	77.3%	49.2%
(4)VoxNet <sup>[32]</sup>	ModelNet40	83.0%	-
(5)Primitive-GAN <sup>[33]</sup>	ModelNet40	86.4%	-
(6)VRN <sup>[34]</sup>	ModelNet40	91.3%	-
(7)LP-3DCNN <sup>[35]</sup>	ModelNet40	92.1%	-
(8)LDGCNN <sup>[36]</sup>	-	92.9%	-
(9)MVCNN-MultiRes <sup>[22]</sup>	-	91.4%	-
(10)PVRNet <sup>[9]</sup>	ModelNet40	93.6%	90.5%
(11)MVCNN,12× <sup>[1]</sup>	ImageNet1K	89.9%	70.1%
(12)MVCNN,metric,12× <sup>[1]</sup>	ImageNet1K	89.5%	80.2%
(13)MVCNN,80× <sup>[1]</sup>	ImageNet1K	90.1%	70.4%
(14)MVCNN,metric,80× <sup>[1]</sup>	ImageNet1K	90.1%	79.5%
(15)PointNet <sup>[2]</sup>	-	89.2%	-
(16)PointNet++ <sup>[3]</sup>	-	90.7%	-
(17)KD-Network <sup>[29]</sup>	-	91.8%	-
(18)PointCNN <sup>[30]</sup>	-	91.8%	-
(19)DGCNN <sup>[31]</sup>	-	92.2%	-
(20)PANORAMA-NN <sup>[4]</sup>	-	90.7%	83.4%
(21)PVNet <sup>[7]</sup>	ImageNet1K	93.2%	89.5%
(22)本文算法	ImageNet1K & ModelNet40	93.8%	89.8%

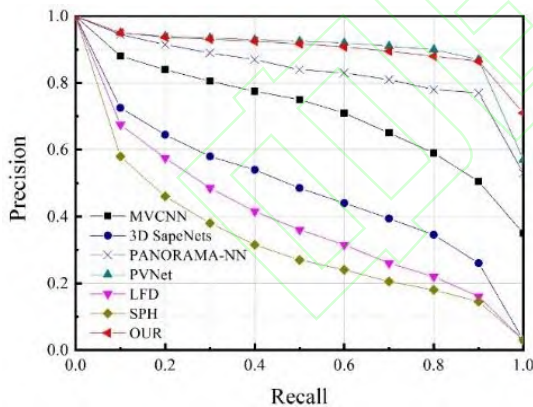


图 6 PR 曲线对比图

Fig. 6 Comparison chart of PR curve

针对本文提出方法取得这样的性能分析如下。首先,本文提出了一种新的相关性损失函数,在训练时可以利用不同模态的优势来相互指导模型网络的学习,提高训练步骤中的特征学习效率。其次,本文所提出的基于统计方法的多模态特征融合方法可以在最终的分类和检索问题中综合利用不同模态的优势。相关实验也证明了本文的方法的优越性。

### 3 结束语

本文提出了一种新的多模态联合网络,它可以联合不同的模态数据进行三维模型分类和检索。在本框架中,引入一种新的相关性损失函数以利用不同模态网络的优势来相互指导进行特征学习,并有效提高学习速度和单一模态网络性能。此外,本文提出了一种新的基于统计思想的多模态特征融合

在检索任务中,本文将式(5)中的融合特征  $f$  作为三维模型的特征向量。采用欧几里德距离度量三维模型之间的相似度。所有检索实验的 PR 曲线如图 6 所示。结果表明,本文提出的方法效果最优,平均精度均值(mAP)达到了 89.8%,证明了本文方法在三维模型检索情形下的有效性。

方法,通过定义了每个模态特征的权重,以发挥每个模态网络的优势,为每个三维模型获得更强大的特征表示。本文提出的网络已经通过实验结果和与 ModelNet 数据集上最先进模型的比较,结果证明了本文方法的有效性。此外本文还研究了模型的不同模态组合的对比实验结果,以证明本文网络的稳健性。

随着三维模型领域单一模态特征表示在三维模型检索分类领域的发展,检索效果和分类准确性提升进入瓶颈。除了提出新的特征描述算法,很多学者采取加深网络的深度以获得更优秀的特征表示。本文提出了一种具有鲁棒性的多模态融合框架,可以利用不同模态的优势进行特征训练,是一种新的思路。

### 参考文献:

- [1] Su Hang, Maji S, Kalogerakis E, *et al.* Multi-view convolutional neural networks for 3D shape recognition [C]// Proc of IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ: IEEE Press, 2015: 945-953.
- [2] Alberto G G, Francisco G D, José G R, *et al.* PointNet: a 3D convolutional neural network for real-time object class recognition [C]// Proc of International Joint Conference on Neural Networks (IJCNN). Piscataway, NJ: IEEE Press, 2016.
- [3] Qi Charles R, Yi Li, Su Hao, *et al.* PointNet++: deep hierarchical feature learning on point sets in a metric space [J]. 2017.
- [4] Sfikas K, Pratikakis I, Theoharis T. Ensemble of PANORAMA-based convolutional neural networks for 3D model classification and retrieval [J]. Computers & Graphics, 2017: S0097849317301978.

- [5] Shi Baoguang, Bai Song, Zhou Zhichao, *et al.* DeepPano: deep panoramic representation for 3-D shape recognition [J]. IEEE Signal Processing Letters, 2015, 22 (12): 2339-2343.
- [6] Gang-Joon Y, Sang-Min Y. Sketch-based 3D object recognition from locally optimized sparse features [J]. Neurocomputing, 2017: 556-563.
- [7] You Haoxuan, Feng Yifan, Ji Rongrong, *et al.* PVNet: a joint convolutional network of point cloud and multi-view for 3D shape recognition [J]. ACM Transactions on Graphics, 2018.
- [8] 晋忠孝. 基于 3D 多视图的物体识别及姿态估计方法 [D]. 安徽: 中国科学技术大学, 2019. (Jin Zhongxiao. Object recognition and 6D pose estimation method based on 3D multi-view data generation [D]. Anhui: University of Science and Technology of China, 2019.)
- [9] You Haoxuan, Feng Yifan, Zhao Xibin, *et al.* PVRNet: point-view relation neural network for 3D shape recognition [EB/OL]. (2018-12-02) [2020-04-28]. <https://arxiv.org/pdf/1812.00333.pdf>.
- [10] Papadakis P, Pratikakis I, Perantonis S, *et al.* Efficient 3D shape matching and retrieval using a concrete radialized spherical projection representation [J]. Pattern Recognition, 2007, 40 (9): 2437-2452.
- [11] Dan Z. Study on interior decoration system design based on 3D scene modeling technology [C]// Proc of International Conference on Smart Grid and Electrical Automation (ICSGEA). Piscataway, NJ: IEEE Press, 2017: 380-383.
- [12] Callieri M, Dellepiane M, Cignoni P, *et al.* MeshLab as a complete open tool for the integration of photos and color with high-resolution 3D geometry data [J]. Computer Applications & Quantitative Methods in Archaeology, 2012.
- [13] Dyn N, Levine D, Gregory J A. A butterfly subdivision scheme for surface interpolation with tension control [J]. ACM Transactions on Graphics, 1990, 9 (2): 160-169.
- [14] 黄思奇, 夏海英. 基于点云面的区域性三维重建及点云拼接 [J]. 计算机应用研究, 2016, 033 (12): 3909-3911, 3919. (Huang Siqi, Xia Haiying. Regional 3D reconstruction and point cloud mosaicing based on point cloud surface [J]. Application Research of Computers, 2016, 033 (12): 3909-3911, 3919.)
- [15] Yang Wenyan, Qian Yanlin, Kamarainen J K, *et al.* Object detection in equirectangular panorama [C]// Proc of the 24th International Conference on Pattern Recognition (ICPR). Piscataway, NJ: IEEE Press, 2018: 2190-2195.
- [16] Yoshimi K, Takahashi K. Warping approach for rearview pedestrian detection with fish eye cameras [C]// Proc of IEEE International Conference on Vehicular Electronics and Safety (ICVES). Piscataway, NJ: IEEE Press, 2017: 121-126.
- [17] Nápoles G, Concepción L, Falcon R, *et al.* On the accuracy-convergence tradeoff in sigmoid fuzzy cognitive maps [J]. IEEE Transactions on Fuzzy Systems, 2017, 99: 1-1.
- [18] Nápoles G, Papageorgiou E, Bello R, *et al.* On the convergence of sigmoid Fuzzy Cognitive Maps [J]. Information Sciences, 2016: 154-171.
- [19] Gonzalo Nápoles, Bello R, Vanhoof K. Learning stability features on sigmoid fuzzy cognitive maps through a swarm intelligence approach [C]// Proc of the 18th Iberoamerican Congress on Pattern Recognition (CIARP 2013). Berlin: Springer, 2013: 270-277.
- [20] Elfving S, Uchibe E, Doya K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning [J]. Neural Netw, 2017, 107: 3-11.
- [21] Jebadurai J, Peter J D. SK-SVR: sigmoid kernel support vector regression based in-scale single image super-resolution [J]. Pattern Recognition Letters, 2017, 94.
- [22] Qi Charles R, Su Hao, Niebner M, *et al.* Volumetric and multi-view CNNs for object classification on 3D data [C]// Proc of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR2016). Piscataway, NJ: IEEE Press, 2016: 5648-5656.
- [23] Sfikas K, Theoharis T, Pratikakis I. Exploiting the PANORAMA representation for convolutional neural network classification and retrieval [C]// Proc of the 10th Eurographics Workshop on 3D Object Retrieval 2017. Amsterdam, NLD: 3DOR Press, 2017.
- [24] Monauni, Max. Agility enablers in production networks-pooling and allying of manufacturing resources [J]. Procedia CIRP, 2014, 17: 657-662.
- [25] Yu Ye, Jin Qiang, Chen C W. FF-CMnet: a CNN-based model for fine-grained classification of car models based on feature fusion [C]// Proc of IEEE International Conference on Multimedia and Expo (ICME). Piscataway, NJ: IEEE Press, 2018: 1-6.
- [26] Wu Zhirong, Song Shuran, Khosla A, *et al.* 3D ShapeNets: a deep representation for volumetric shape modeling [C]// Proc of the 28th IEEE Conference on Computer Vision and Pattern Recognition (CVPR2015). Piscataway, NJ: IEEE Press, 2015: 1912-1920.
- [27] Chen Dingyun, Tian Xiaopei, Shen Yute, *et al.* On visual similarity based 3D model retrieval [J]. Computer Graphics Forum, 2003, 22 (3): 223-232.
- [28] Kobbelt L, Schröder P, Kazhdan M, *et al.* Rotation invariant spherical harmonic representation of 3D shape descriptors [J]. Eurographics Symposium on Geometry Processing, 2003, 43 (2): 156-164.
- [29] Klovov R, Lempitsky V. Escape from Cells: Deep Kd-networks for the recognition of 3D point cloud models [C]// Proc of IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ: IEEE Press, 2017: 863-872.
- [30] Li Yangyan, Bu Rui, Sun Mingchao, *et al.* PointCNN: convolution on X-transformed points [C]// Proc of Conference and Workshop on Neural Information Processing Systems (NIPS). 2018.
- [31] Wang Yue, Sun Yongbin, Liu Ziwei, *et al.* Dynamic graph CNN for learning on point clouds [J]. ACM Transactions on Graphics, 2018, 38 (5).
- [32] Maturana D, Scherer S. VoxNet: a 3D convolutional neural network for real-time object recognition [C]// Proc of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway, NJ: IEEE Press, 2015: 922-928.
- [33] Khan S H, Guo Y, Hayat M, *et al.* Unsupervised primitive discovery for improved 3D generative modeling [C]// Proc of the 32th IEEE Conference on Computer Vision and Pattern Recognition (CVPR2019). Piscataway, NJ: IEEE Press, 2019: 9731-9740.
- [34] Brock A, Lim T, Ritchie J M, *et al.* Generative and discriminative voxel modeling with convolutional neural networks [J]. Computer Science, 2016.
- [35] Kumawat S, Raman S. LP-3DCNN: unveiling local phase in 3D convolutional neural networks [C]// Proc of the 32th IEEE Conference on Computer Vision and Pattern Recognition (CVPR2019). Piscataway, NJ: IEEE Press, 2019: 4898-4907.
- [36] Zhang Kuangen, Hao Ming, Wang Jing, *et al.* Linked dynamic graph CNN: learning on point cloud via linking hierarchical features [EB/OL]. (2019-08-06) [2020-03-30]. <https://arxiv.org/pdf/1904.10014.pdf>.