

## 基于胶囊网络的三维模型识别

曹小威, 曲志坚, 徐玲玲, 刘晓红\*

(山东理工大学 计算机科学与技术学院, 山东 淄博 255000)

(\* 通信作者电子邮箱 lxhsdut@163.com)

**摘要:**为解决传统卷积神经网络中大量池化层的引入导致特征信息丢失的问题,依据胶囊网络(CapsNet)使用向量神经元保存特征空间信息的特性,提出了一种用以识别三维模型的网络模型3DSPNCapsNet。使用新的网络结构,提取更具代表性的特征的同时降低了模型复杂度,并提出基于动态路由(DR)算法的DRL算法来优化胶囊权重的迭代计算过程。在ModelNet10上的实验结果表明,相比3DCapsNet以及VoxNet,该网络取得了更好的识别效果,在原始测试集上3DSPNCapsNet的平均识别准确率达到95%,同时验证了该网络对旋转三维模型的识别能力。适当扩展旋转训练集之后,所提网络对各角度旋转模型的平均识别率达到81%。实验结果表明,3DSPNCapsNet对三维模型及其旋转具有良好的识别能力。

**关键词:**胶囊网络;动态路由算法;池化;三维模型识别;旋转

**中图分类号:**TP391.4 **文献标志码:**A

### 3D model recognition based on capsule network

CAO Xiaowei, QU Zhijian, XU Lingling, LIU Xiaohong\*

(School of Computer Science and Technology, Shandong University of Technology, Zibo Shandong 250000, China)

**Abstract:** In order to solve the problem of feature information loss caused by the introduction of a large number of pooling layers in traditional convolutional neural networks, based on the feature of Capsule Network (CapsNet)——using vector neurons to save feature space information, a network model 3DSPNCapsNet (3D Small Pooling No dense Capsule Network) was proposed for recognizing 3D models. Using the new network structure, more representative features were extracted while the model complexity was reduced. And based on Dynamic Routing (DR) algorithm, Dynamic Routing-based algorithm with Length information (DRL) algorithm was proposed to optimize the iterative calculation process of capsule weights. Experimental results on ModelNet10 show that compared with 3DCapsNet (3D Capsule Network) and VoxNet, the proposed network achieves better recognition results, and has the average recognition accuracy on the original test set reached 95%. At the same time, the recognition ability of the network for the rotation 3D models was verified. After the rotation training set is appropriately extended, the average recognition rate of the proposed network for rotation models of different angles reaches 81%. The experimental results show that 3DSPNCapsNet has a good ability to recognize 3D models and their rotations.

**Key words:** Capsule Network (CapsNet); dynamic routing algorithm; pooling; 3D model recognition; rotation

## 0 引言

卷积神经网络(Convolutional Neural Network, CNN)不仅在传统的二维图像识别领域展现了强大的特征提取能力,在三维模型识别领域也被广泛应用:王慧<sup>[1]</sup>使用投影提取三维模型的形状特征,并使用一种可以同时计算三个侧面投影图的权值优化深度卷积网络,从而对三维模型信息进行深层次加工处理,进而对三维模型进行分类识别;侯宇昆<sup>[2]</sup>提取三维模型的12个角度视图以及全景式图,并将两种视图融合,输入到搭建好的13层的卷积神经网络中从而实现三维模型的识别,在ModelNet10上取得了94%的平均识别准确率;

Hegde等<sup>[3]</sup>融合了基于体素的V-CNN I、V-CNN II以及基于视图的MV-CNN(Multi-View Convolutional Neural Network),得到FusionNet,并在ModelNet10的识别中得到了93.1%的识别准确率;Bu等<sup>[4]</sup>首先将低层三维形状描述符编码成几何词袋,从中发现中层模式,探索词与词之间的几何关系。在此基础上,通过深度信念网络学习高级形状特征,实验证明该方法对三维模型分类任务具有较强的识别能力。尽管在卷积神经网络的设计中使用池化层在一定程度上增大了神经元的感受野以及降低了样本训练的开销,但是大量池化层的使用会造成原始数据结构信息的丢失,从而对最终的识别结果造成不良影响。

收稿日期:2019-10-16;修回日期:2019-12-13;录用日期:2019-12-17。

基金项目:国家自然科学基金资助项目(61473179);山东省高等学校科技计划项目(J16LN20);山东省自然科学基金资助项目(ZR2016FM18, ZR2017LF004);山东省高等学校青年创新团队发展计划项目(2019KJN048)。

作者简介:曹小威(1990—),男,河南项城人,硕士研究生,主要研究方向:三维模型识别;曲志坚(1980—),男,山东青岛人,副教授,博士,主要研究方向:机器学习;徐玲玲(1997—),女,山东菏泽人,硕士研究生,主要研究方向:推荐算法;刘晓红(1968—),女,黑龙江哈尔滨人,教授,硕士,主要研究方向:图像处理。

Hinton 等<sup>[5]</sup>于 2017 年提出的胶囊网络创新地使用向量胶囊取代传统神经网络的标量神经元,舍弃了池化层,并提出了动态路由(Dynamic Routing, DR)算法更新胶囊之间的连接权重,从而较好地保留了特征的空间信息,在 MINIST 手写数字识别上达到了 99.2% 识别准确率。自胶囊网络诞生之后,人们将其应用到了很多领域,如人类行为识别<sup>[6]</sup>、人工声音事件检测<sup>[7]</sup>、乳腺癌识别<sup>[8]</sup>等。在三维模型识别领域,Ahmad 等<sup>[9]</sup>于 2018 年提出了 3D Capsule Networks,证明了同样深度的胶囊网络比卷积网络,不仅对于二值体素化的三维模型具有更好的识别能力,而且在使用 40% 的训练集进行训练时,胶囊网络也能得到 91.37% 的识别率。Ryan<sup>[10]</sup>将二维胶囊网络加以改变,使之能够识别三维体素化数据,通过调优,在 ModelNet10 的识别中达到了 93% 的识别准确率。在验证旋转敏感性方面,通过将测试集的体素化数据进行转置并输入训练好的胶囊网络中进行识别之后,发现胶囊网络对该测试集中的模型并不能很好地识别。然而胶囊网络利用向量的方向存储特征的相对空间信息,理论上应该对旋转模型具有良好的识别能力。此外,为了保留了特征的空间信息,现有的基于胶囊网络的三维模型识别网络通常都舍弃了池化层的设计,但是池化层本身在特征降维、压缩数据、减少过拟合、提高模型容错性等方面具有优良表现,适当地引入一层步长和较小的池化层不会导致特征空间信息的大量丢失。

本文以 Ryan<sup>[10]</sup>模型为 Basemodel,引入 VoxNet<sup>[11]</sup>作对比,从提高识别率和探索旋转敏感性两个方面作了相关研究;在提高识别率方面,本文提出了一种新的基于胶囊网络的结构 3DSPNCapsNet,引入了一层步长和尺寸都较小的池化层;同

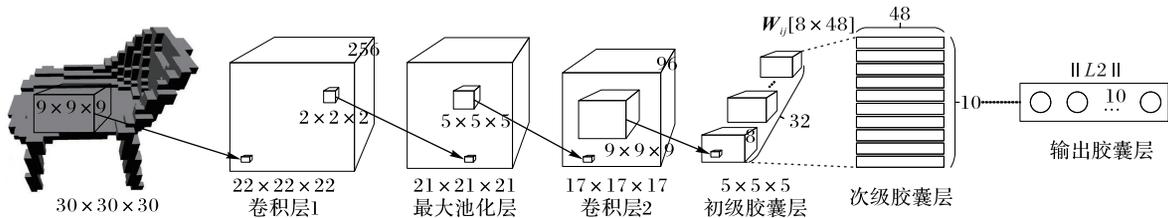


图1 3DSPNCapsNet 结构

Fig. 1 Structure of 3DSPNCapsNet

如图1,体素化后大小为  $30 \times 30 \times 30$  的三维模型数据首先被第一个由 256 个三维卷积核构成,大小为  $9 \times 9 \times 9$ 、步长为  $1 \times 1 \times 1$ 、激活函数为 Relu 的卷积层提取基础特征,得到  $22 \times 22 \times 22 \times 256$  的输出。经过大小为  $2 \times 2 \times 2$ 、步长为 2 的最大池化层处理,将得到的  $21 \times 21 \times 21 \times 256$  的结果通过第二个由大小为  $5 \times 5 \times 5$ 、步长为  $1 \times 1 \times 1$ 、96 个卷积核构成的卷积层进行卷积操作,得到  $17 \times 17 \times 17 \times 96$  的输出。然后使用 32 个步长为 2 的  $9 \times 9 \times 9 \times 256$  的卷积核对其进行 8 通道卷积,将每个 8 通道卷积结果封装成 8 维向量,并使用 Squash 激活函数处理胶囊,得到  $7 \times 7 \times 7 \times 32 \times 8$  的输出,构成了初级胶囊层。Squash 函数保证向量的长度在 0~1,从而使向量在高维空间中的方向体现实体的不同属性,其表达式如式(1):

$$u_j = \frac{\|s_j\|^2 s_j}{1 + \|s_j\|^2 \|s_j\|} \quad (1)$$

其中:  $u_j$  为输出向量,  $s_j$  为输入向量。公式右边的  $s_j / \|s_j\|$  是单

时改进了动态路由协议,将向量长度加入迭代过程中,使得网络模型在 ModelNet10 上的识别准确率得到提高;在探索旋转敏感性方面,本文对三维模型的旋转进行了可视化,并分别在原训练集和扩充了部分旋转模型的训练集上进行训练,在旋转测试集上测试,达到了良好效果。

## 1 3DSPNCapsNet

CapsNet 最早由 Hinton 提出,利用向量保存特征信息。用向量不同维度上的值分别表示不同的属性,向量的模表示这个实体出现的概率。Basemodel 是对 Hinton 提出的二维胶囊网络的三维扩展,在两个胶囊层之前仅仅使用了一个卷积层,并没有在网络设计中增加太多的层次结构。而传统的 GoogleNet 具有高达 22 层的深度网络结构。通常情况下,更深的结构意味着网络能提取到更具有代表性的特征,然而层数过多也容易导致过拟合、梯度爆炸等情况的出现。

基于此,本文提出了一种基于胶囊网络的网络结构 3DSPNCapsNet (3D Small Pooling No dense Capsule Networks),该网络由双 CNN 间连接一个最大池化层,以及初级胶囊层、次级胶囊层和输出胶囊层构成。与 Basemodel 的结构相比,3DSPNCapsNet 增加了一层池化层和卷积层,在提取到更具代表性的特征的同时,利用一层步长和大小都较小的池化层减少了冗余特征。虽然池化层的引入会在一定程度上损失部分信息,不过一层步长和较小的池化层损失的信息有限,而其在扩大感知野、减少过拟合等方面的作用,会使网络学习到的特征更具有代表性。将该网络优化参数配置后用于三维模型识别,其结构如图 1 所示。

位向量。如果  $s_j \rightarrow \infty$ , 则  $\frac{\|s_j\|^2}{1 + \|s_j\|^2} \rightarrow 1$ ; 如果  $s_j \rightarrow 0$ , 则

$$\frac{\|s_j\|^2}{1 + \|s_j\|^2} \rightarrow 0. \text{ 这种方式增加了模长较大的向量的权重}^{[12]}.$$

初级胶囊层的输出向量分别与  $8 \times 48$  的姿态矩阵  $W_{ij}$  点积之后,得到 48 维的向量,如式(2):

$$\hat{u}_{ji} = W_{ij} u_i \quad (2)$$

其中:  $u_i$  为初始胶囊层的输出向量;  $W_{ij}$  为姿态矩阵,由反向传播算法训练。点积之后的 48 维向量与次级胶囊层的 10 个  $1 \times 48$  的向量全连接,使用第 2 章的 DRL (Dynamic Routing-based algorithm with Length information) 算法对该过程的参数进行更新。得到的预测向量构成次级胶囊层,最后输出胶囊层通过计算次级胶囊层的向量长度,取其中模长最大的向量作为最终预测。

由于重构网络的损失权重只有 0.005,所以本文无论是训练还是预测过程,都舍弃了重构网络以及误差,从而降低了

模型复杂度,有助于提升模型的训练效率。对于有10个类别、批训练的大小为 $N$ 的数据集,新的误差计算方式如式(3)、(4):

$$loss = \frac{1}{N} \sum_{i=0}^N \sum_{k=0}^{10} L_k^i \quad (3)$$

$$L_k = T_k \max\left(0, m^+ - \|\mathbf{v}_k\|\right)^2 + \lambda(1 - T_k) \max\left(0, \|\mathbf{v}_k\| - m^-\right)^2 \quad (4)$$

根据Hinton提出的胶囊网络, $L_k$ 为每个胶囊的margin loss,其中 $m^- = 0.1$ 、 $m^+ = 0.9$ 、 $\lambda = 0.5$ 。当模型被正确识别为种类 $k$ 时, $T_k$ 为1,否则为0。分别计算每个胶囊的 $L_k$ ,再对所有胶囊的损失求和得到总体损失 $loss$ 。

## 2 DRL 算法

原有的动态路由算法仅仅使用余弦相似度来度量向量之间的相似度,但是这可能会导致网络训练时收敛速度较慢以及丢失部分信息。最近有一些胶囊网络的相关应用研究了新的动态规划算法,DRDL (Dynamic Routing-based algorithm with Direction and Length information)<sup>[13]</sup>就是其中的一个。DRDL在动态路由的基础上将向量长度和方向信息引入算法中。不过将该方法引入3DSPNCapsNet后,训练结果不收敛,所以本文在其基础上做了一些修改,舍弃了其中对聚类中心向量 $\mathbf{s}_i$ 取L2正则化以及使用预测向量 $\hat{\mathbf{u}}_{ji}$ 和目标向量 $\mathbf{v}_j$ 的模长之差更新 $b_{ij}$ 的步骤,提出了DRL算法。

定义多维空间的两个向量 $\mathbf{A} = (A_1, A_2, \dots, A_n)$ ,  $\mathbf{B} = (B_1, B_2, \dots, B_n)$ ,二者夹角的余弦:

$$\cos \theta = \frac{\sum_{i=1}^n (A_i \cdot B_i)}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (5)$$

同理,定义:

$$b_{ij} = \frac{\hat{\mathbf{u}}_{ji} \cdot \mathbf{v}_j}{\|\hat{\mathbf{u}}_{ji}\| \|\mathbf{v}_j\|} \quad (6)$$

其中: $\hat{\mathbf{u}}_{ji}$ 为初始胶囊层的第 $i$ 个胶囊的预测向量, $\mathbf{v}_j$ 为次级胶囊层的第 $j$ 个胶囊的输出向量, $b_{ij}$ 为两个胶囊之间的连接权重。此外,Squash只在最后一轮迭代的时候使用。算法如图2所示。

首先初始化 $b_{ij}$ 为0,其中 $b_{ij}$ 表示上一层的胶囊 $i$ 与本层的胶囊 $j$ 之间的相关性,其值越大,表示两胶囊之间相关性越强。然后开始迭代,每次迭代开始时都先判断当前迭代是不是最后一轮,如果不是,就利用Softmax函数得到 $c_{ij}$ ,Softmax函数如式(7):

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \quad (7)$$

$c_{ij}$ 为动态路由过程中的耦合系数,然后对式(2)中的预测向量 $\hat{\mathbf{u}}_{ji}$ 进行聚类,得到预测向量的聚类中心 $\mathbf{s}_j$ ,并使用式(8)对 $b_{ij}$ 进行更新:

$$b_{ij} = \frac{\hat{\mathbf{u}}_{ji} \cdot \mathbf{s}_j}{\left(\|\hat{\mathbf{u}}_{ji}\| \cdot \|\mathbf{s}_j\|\right)} \quad (8)$$

如果上一层的预测向量 $\hat{\mathbf{u}}_{ji}$ 的单位向量与本层的聚类中心向量 $\mathbf{s}_j$ 的单位向量的夹角在 $(-90^\circ, 90^\circ)$ ,则 $b_{ij}$ 为正;如果夹

角为 $\pm 90^\circ$ ,则 $b_{ij}$ 变为0;如果夹角在 $(90^\circ, 270^\circ)$ ,则 $b_{ij}$ 为负。如果是最后一轮迭代,计算出聚类中心 $\mathbf{s}_j$ 之后,通过Squash函数输出最终的预测向量。

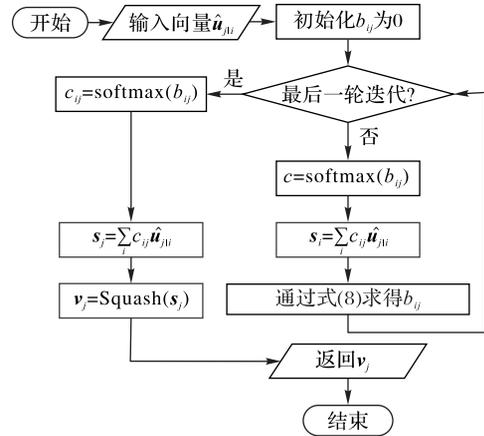


图2 DRL算法流程

Fig. 2 Flowchart of DRL algorithm

角为 $\pm 90^\circ$ ,则 $b_{ij}$ 变为0;如果夹角在 $(90^\circ, 270^\circ)$ ,则 $b_{ij}$ 为负。如果是最后一轮迭代,计算出聚类中心 $\mathbf{s}_j$ 之后,通过Squash函数输出最终的预测向量。

## 3 实验分析与结果

普林斯顿大学提供了两个大规模的标准三维模型数据集,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,其子集ModelNet10包含了来自10类的4 899个三维模型。本文选取ModelNet10的模型作为实验数据库。

实验平台和开发工具有:CPU为Intel Xeon E3-1231 3.40 GHz,内存为8 GB,显卡Quadro K620,编程环境为Python3.5,基于Keras搭建的实验验证本文方法的有效性。

### 3.1 3DSPNCapsNet在标准训练集和测试集上的性能比较

本节分别使用融合了DRL算法的3DSPNCapsNet以及使用原始DR算法的3DSPNCapsNet、Basemodel、VoxNet在标准的ModelNet10的训练集和测试集上训练和测试。将原始三维模型的.off文件体素化为.binvox文件,使用上采样解决类间不平衡问题之后,输入网络中使用5折交叉验证,每折训练5轮;学习率初始化为0.003,最后取交叉验证结果的平均值作为最终的识别率。训练期间根据输出胶囊的预测准确率对学习率进行动态调整,影响因子为0.5,学习率最低不小于0.0001;batch size取10,使用1个GPU加速计算。几种方法在标准训练集和测试集上的识别率折线图如图3,融合了DRL算法的3DSPNCapsNet的混淆矩阵如图4,Basemodel的混淆矩阵如图5,使用原始DR算法的3DSPNCapsNet的混淆矩阵如图6,VoxNet的混淆矩阵如图7所示。

实验结果显示,融合DRL算法的3DSPNCapsNet平均识别率达到95%、使用原始DR算法的3DSPNCapsNet为94%、Bsaemodel为93%、VoxNet为88%,平均识别准确率方面融合DRL算法的3DSPNCapsNet最高。从具体的模型识别分类来看,与Basemodel相比,融合DRL算法的3DSPNCapsNet除了sofa、table、toilet、bathtub略低1~3个百分点之外,monitor、chair、bed高了1~6个百分点,dresser、night\_stand、desk高了12~18个百分点。与VoxNet相比,除了table类的识别率略低于VoxNet,其他类别的模型识别率都取得了与VoxNet相同或

者更好的效果。平均识别率达到了95%以上。融合DRL算法的3DSPNCapsNet(以下简称3DSPNCapsNet)表现出了良好的特征提取能力和识别能力。

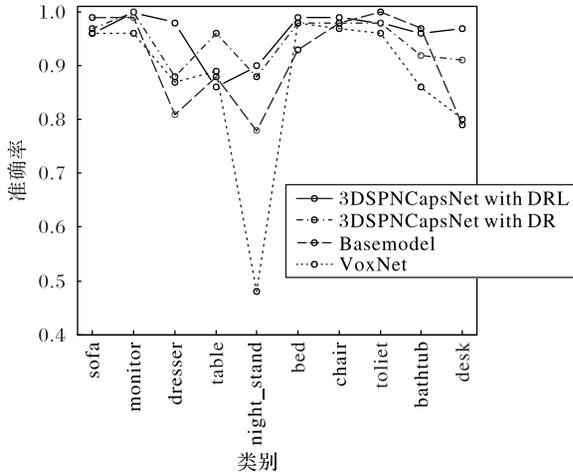


图3 使用DRL的3DSPNCapsNet与使用DR的3DSPNCapsNet、Basemodel、VoxNet的结果对比

Fig. 3 Results comparison of 3DSPNCapsNet with DRL, 3DSPNCapsNet with DR, Basemodel, VoxNet

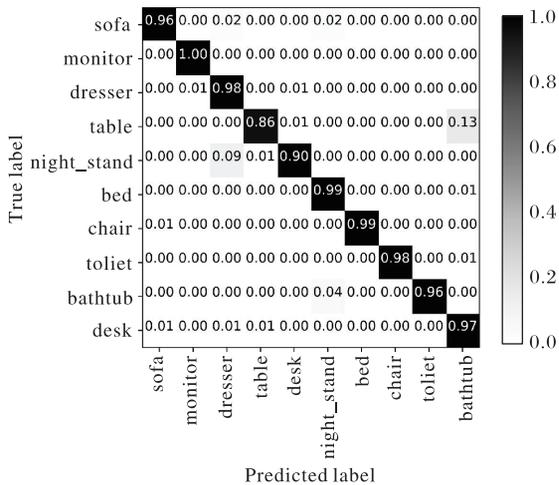


图4 3DSPNCapsNet混淆矩阵

Fig. 4 Confusion matrix of 3DSPNCapsNet

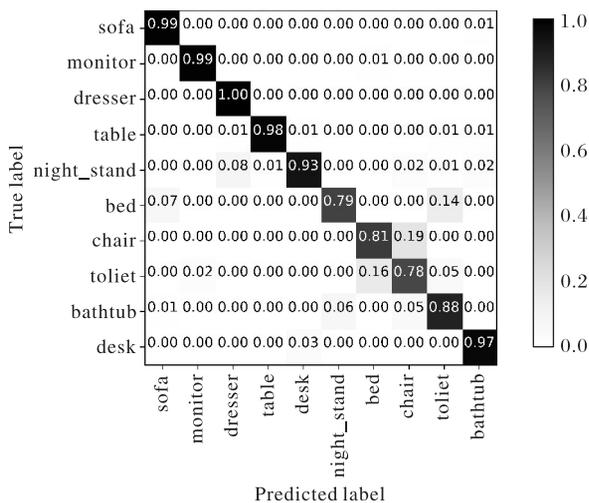


图5 Basemodel混淆矩阵

Fig. 5 Confusion matrix of Basemodel

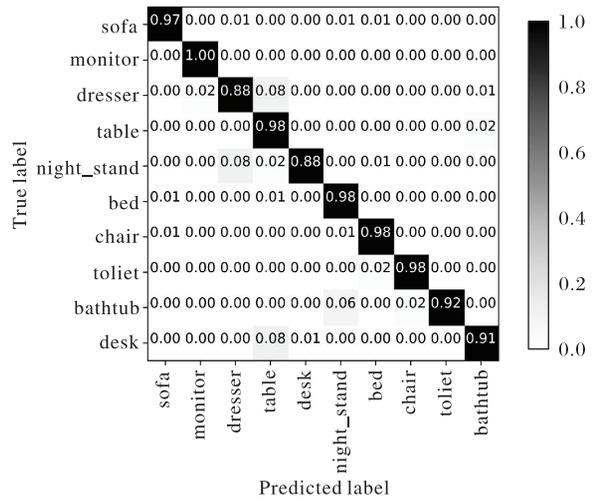


图6 3DSPNCapsNet with DR混淆矩阵

Fig. 6 Confusion matrix of 3DSPNCapsNet with DR

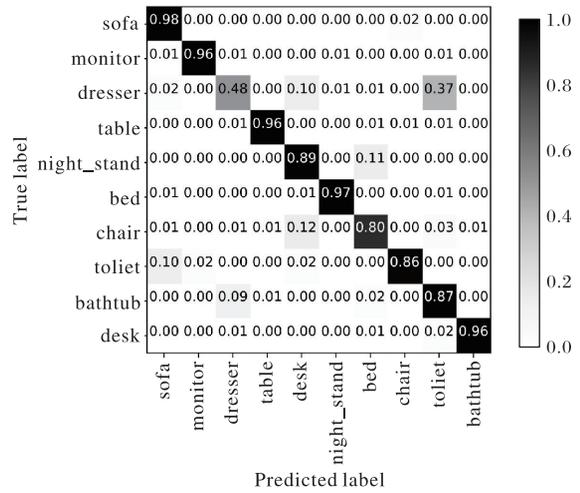


图7 VoxNet混淆矩阵

Fig. 7 Confusion matrix of VoxNet

### 3.2 3DSPNCapsNet在旋转敏感性方面的性能比较

#### 3.2.1 不增加旋转训练集时,三种方法对旋转的敏感性对比

胶囊网络通过向量胶囊的形式既保存了特征信息,又记录了不同特征之间的空间位置信息,理论上来说对于模型的旋转应该具有较高的识别准确率。为了验证这一特点,本文使用Meshlab将测试集的模型先绕z轴分别旋转30°、60°、90°、120°、150°、180°、210°、240°、270°、300°、330°,再体素化为.binbox文件,最后使用viewvox将其可视化。使用在原始的训练集上训练的模型分别在上述旋转测试集上测试。旋转体素化结果如图8,旋转测试结果如图9。

实验结果显示,当前实验条件下,3DSPNCapsNet对原始测试集模型识别率达到95%、Basemodel为93%、VoxNet为88%。然而3DSPNCapsNet对旋转模型的识别能力并没有达到预期效果,平均旋转识别率仅48%,与Basemodel的48%以及Voxnet的49%相似。这是由于网络没有学习到相关的旋转信息。将训练集进行扩展,重新训练网络之后再在旋转测试集上测试。

#### 3.2.2 增加旋转训练集时,三种方法对旋转的敏感性对比

将原始训练集分别旋转30°、60°、90°和原有的训练集共

同组成了新的训练集,在新的训练集上使用三种方法分别训练,并在旋转测试集测试。实验结果如图 10。

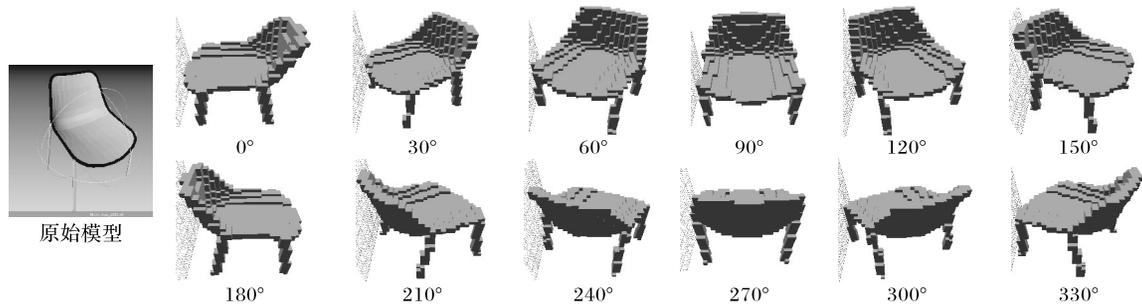


图 8 先绕z轴旋转,再体素化的测试集

Fig 8 Test set first rotated around the z-axis, then voxelized

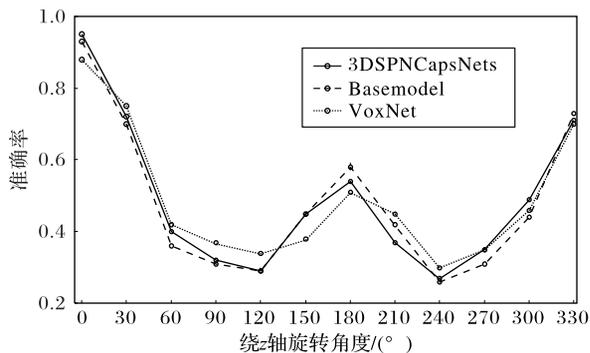


图 9 在原训练集训练、旋转测试集测试的准确率比较

Fig. 9 Test accuracy comparison of rotation test set with original training set training

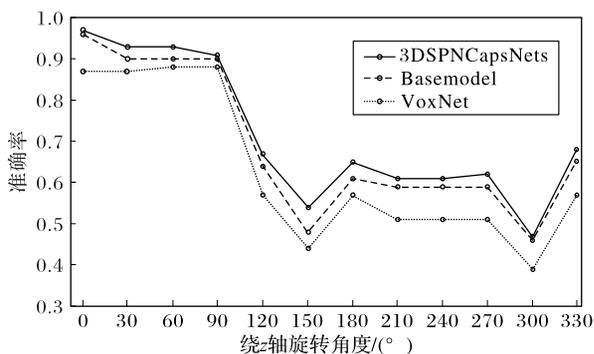


图 10 在旋转训练集训练、旋转测试集测试的准确率比较

Fig. 10 Test accuracy comparison of rotation test set with rotation training set training

由图 10 可以直观地看出,扩展了旋转训练集之后,三种方法对旋转三维模型都表现出了更好的识别能力,其中 3DSPNCapsNet 在 30°、60°、90° 这三个角度上的旋转识别率达到了 91% 以上, Basemodel 也达到了 90% 以上, VoxNet 达到 87% 以上。相对于没有增加旋转训练集的情况, 3DSPNCapsNet 对于旋转模型的平均识别率提升了 24 个百分点, Basemodel 提升了 21 个百分点, VoxNet 提升了 15 个百分点。虽然 3DSPNCapsNet 平均达到了 72% 的识别准确率,但是识别率的方差偏大,数据分布不均衡,在 30°、60°、90° 的识别率明显高于其他旋转角度的识别率。从实验结果还可以看出,三个模型在识别率的变化趋势上保持了一致性:在 120° 和 150° 上逐渐下降,这是因为旋转角度逐渐变大时,与原有模型的差异性也逐渐增加,导致识别率下降;而 180° 时识别率反而提高了,并在 210°、240°、270° 的旋转模型上保持了几乎一致的识别率。通过多次实验发现,这种情况出现的原因是三种

网络对于原始训练集具有最高的识别准确率。相对于其他旋转角度,三种网络对原始训练模型的中心对称模型也具有一定的识别能力,所以网络在旋转 300° 时识别率重新下降,而旋转 330° 时的识别率与 120° 的识别率相当,这是因为 120° 与 90° 的旋转训练集相差 30°, 同 330° 与原始训练集相差的角度一致。基于此,本文重新调整了旋转训练集的构成,将原始训练集分别旋转 90°、150°、240° 和原有的训练集共同组成了新的训练集,在新的旋转训练集上使用三种网络分别训练,并在旋转测试集测试。实验结果如图 11。

网络对于原始训练集具有最高的识别准确率。相对于其他旋转角度,三种网络对原始训练模型的中心对称模型也具有一定的识别能力,所以网络在旋转 300° 时识别率重新下降,而旋转 330° 时的识别率与 120° 的识别率相当,这是因为 120° 与 90° 的旋转训练集相差 30°, 同 330° 与原始训练集相差的角度一致。基于此,本文重新调整了旋转训练集的构成,将原始训练集分别旋转 90°、150°、240° 和原有的训练集共同组成了新的训练集,在新的旋转训练集上使用三种网络分别训练,并在旋转测试集测试。实验结果如图 11。

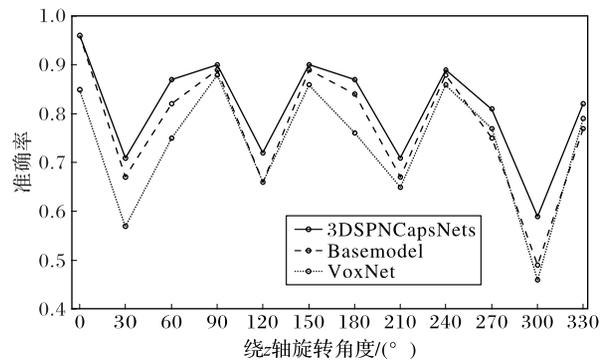


图 11 在新的旋转训练集训练、旋转测试集测试的准确率比较

Fig. 11 Test accuracy comparison of rotation test set with new rotation training set training

实验结果显示,调整了旋转训练集之后,三种方法在旋转模型的识别方面整体都有了较大的提升。3DSPNCapsNet 和 Basemodel 在原始测试集上的识别准确率都达到了 96%,在其他旋转测试集上,3DSPNCapsNet 的总体平均识别率达到了 81%,高于 Basemodel 的 78% 和 VoxNet 的 73%;相对于 30°、60°、90° 的旋转训练集实验,平均识别率提高了 9 个百分点, Basemodel 提高了 9 个百分点, VoxNet 提高了 10 个百分点,识别准确率方差较小,数据分布更加均衡。3DSPNCapsNet 整体对于旋转模型具有更好的识别能力。

### 3.3 池化层对识别率的影响

Hinton 提出胶囊网络的初衷是因为池化层的设计会导致特征信息的丢失,尤其是特征之间的空间位置信息很容易在池化的过程中被丢弃。基于此,本文实验比较分析了池化层对模型的旋转识别的影响。将 3DSPNCapsNet 的池化层舍弃,并与未舍弃池化层的网络分别在旋转训练集和测试集上训练和测试,并进行对比,实验结果如图 12。

图 12 显示,两种方法对于旋转都具有较好的识别能力,其中有池化层的网络在原始数据集上取得了更高的 96% 的

识别准确率,而除了在旋转 $210^\circ$ 的模型识别率上低了5个百分点,其他的角度都与无池化层的模型识别率相似,说明两种胶囊网络都具有良好的特征提取能力,而且一层大小和步长都较小的池化层的加入并不会导致特征信息的大量丢失。池化层在降维、实现非线性、扩大感知野等方面都有积极作用,所以不必在新架构的探索中完全舍弃池化层的设计,池化与胶囊网络相结合可能会有更好的效果。

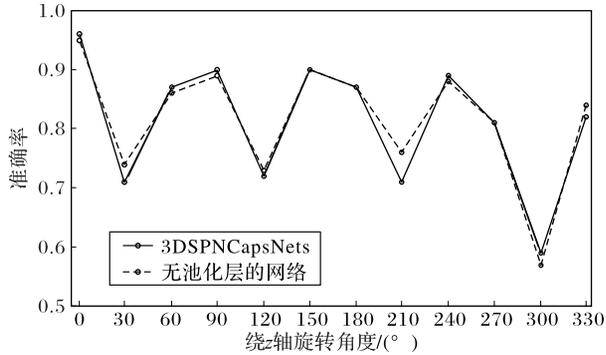


图 12 池化层对旋转准确率的影响

Fig. 12 Effect of pooling layer on rotation recognition accuracy

#### 4 结语

本文提出了一种基于胶囊网络的三维模型识别算法,通过在卷积网络中增加卷积层、池化层得到相对于原始网络结构更好的特征提取效果,舍弃重构网络减少参数数量,以及使用新的动态路由算法得到在原始测试集上95%的识别准确率。通过增加一定数量的旋转训练集,得到对旋转三维模型更有识别力的胶囊网络,并验证了一层大小和步长较小的池化层的引入不会损失太多的信息,反而有可能在一定程度上产生更好的结果。在后续的研究中,可以探索更深的网络结构来进一步提高识别准确率;另外还可以探索如何使胶囊网络学习到更具有旋转方面的代表性的特征,从而不需要通过扩充训练集的方式,就能对三维空间内任意方向任意角度的旋转都具有更好的识别能力。

#### 参考文献 (References)

- [1] 王慧. 基于深度学习的三维模型识别及检索研究[D]. 长春: 长春工业大学, 2018:37-43. (WANG H. Research on 3D model recognition and retrieval based on deep learning [D]. Changchun: Changchun University of Technology, 2018:37-43.)
- [2] 侯宇昆. 基于深度学习和视图的三维CAD模型分类技术研究[D]. 银川: 北方民族大学, 2017:19-42. (HOU Y K. 3D CAD model classification based on deep learning and views [D]. Yinchuan: North Minzu University, 2017:19-42.)
- [3] HEGDE V, ZADEH R. FusionNet: 3D object classification using multiple data representations[EB/OL]. [2016-11-27]. <https://arxiv.org/pdf/1607.05695.pdf>.
- [4] BU S, LIU Z, HAN J, et al. Learning high-level feature by deep belief networks for 3-D model retrieval and recognition[J]. IEEE Transactions on Multimedia, 2014, 16(8):2154-2167.
- [5] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2017:3856-3869.
- [6] ALGAMDI A M, SANCHEZ V, LI C T. Learning temporal information from spatial information using CapsNets for human action recognition[C]// Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2019: 3867-3871.
- [7] VESPERINI F, GABRIELLI L, PRINCIPI E, et al. Polyphonic sound event detection by using capsule neural networks[J]. IEEE Journal of Selected Topics in Signal Processing, 2019, 13(2): 310-322.
- [8] PENG C, ZHENG Y, HUANG D. Capsule network based modeling of multi-omics data for discovery of breast cancer-related genes[EB/OL]. [2019-06-20]. <https://ieeexplore.ieee.org/iel7/8857/4359833/08684326.pdf>.
- [9] AHMAD A, KAKILLIOGLU B, VELIPASALAR S. 3D capsule networks for object classification from 3D model data[C]// Proceedings of the 52nd Asilomar Conference on Signals, Systems, and Computers. Piscataway: IEEE, 2018: 2225-2229.
- [10] RYAN L. Ryanglambert/3d\_model\_retriever[EB/OL]. [2019-04-16]. [https://github.com/Ryanglambert/3d\\_model\\_retriever](https://github.com/Ryanglambert/3d_model_retriever).
- [11] MATURANA D, SCHERER S. VoxNet: a 3D convolutional neural network for real-time object recognition[C]// Proceedings of the 2015 IEEE/RSSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE, 2015: 922-928.
- [12] AFSHAR P, PLATANIOTIS K N, MOHAMMADI A. Capsule networks for brain tumor classification based on MRI images and coarse tumor boundaries[C]// Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2019:1368-1372.
- [13] ZHAO T, LIU Y, HUO G, et al. A deep learning iris recognition method based on capsule network architecture[J]. IEEE Access, 2019, 7: 49691-49701.

This work is partially supported by National Natural Science Foundation of China (61473179), the Shandong Higher Education Science and Technology Program (J16LN20), the Shandong Provincial Natural Science Foundation (ZR2016FM18, ZR2017LF004), the Youth Innovation Team Development Plan of Shandong Province Higher Education (2019KJN048).

**CAO Xiaowei**, born in 1990, M. S. candidate. His research interests include three-dimensional model recognition.

**QU Zhijian**, born in 1980, Ph. D., associate professor. His research interests include machine learning.

**XU Lingling**, born in 1997, M. S. candidate. Her research interests include recommendation algorithm.

**LIU Xiaohong**, born in 1968, M. S., professor. Her research interests include image processing.