文章编号:1671-4598(2022)03-0156-05 DOI:10.16526/j.cnki.11—4762/tp.2022.03.026 中图分类号:TP3 文献标识码:A

基于深度卷积神经网络的点云三维目标识别方法研究

李豪杰，杨海清
（浙江工业大学 信息工程学院，杭州 310012）

摘要：为了提高对三维点云目标的识别精确度，提出一种基于深度卷积神经网络（CNN，convolutional neural network）的点云目标识别模型；针对已有的深度卷积点云目标识别网络无法有效提取点云局部拓扑特征的问题，采用迭代最远点采样（FPS，terative farthest point sampling）结合方向卷积编码方式来捕获局部形状特征；并引入空间变换网络（STN，spatial transform network）使点云数据能够自适应进行空间变换和对齐，以解决点云数据旋转性会造成目标识别结果不稳定的问题；实验结果表明：文中提出的点云目标识别方法有效提高了识别精度度，相较于 PointNet 在 ModelNet40 和 ShapeNetCore 两个数据集上分别提高 1.2%和 1.4%。

关键词：三维点云；目标识别；深度卷积神经网络；方向卷积编码；空间变换网络

Research on 3D Object Recognition Method for Laser Point Cloud Based on Deep Convolution Neural Network

LI Haojie，YANG Haiqing
（School of information engineering，Zhejiang University of Technology，Hangzhou 310012，China）

Abstract：In order to improve the accuracy of 3-D point cloud target recognition，a target recognition model for point cloud based on the Convolutional Neural Network（CNN）is proposed. Aiming at the problem that the existing deep convolutional point cloud target recognition network can not effectively extract the local topological features of point cloud，combined with directional convolutional coding，Iterative Farthest Point Sampling（FPS）is used to capture the local shape features. In view of the instability of target recognition results caused by the rotation of point cloud data，the introduction of Spatial Transform Network（STN）enables point cloud data to self-adaptively perform spatial transformation and alignment. The experimental results show that the point cloud target recognition method proposed in this paper effectively improves the recognition accuracy，which respectively increases by 1.2% and 1.4% compared with PointNet on ModelNet40 and ShapeNetCore data sets.

Keywords：3-D point cloud；target recognition；deep convolution neural network；directional convolution coding；space transformation network

0 引言

三维激光扫描技术被广泛应用于自动驾驶[1]、建筑施工[2]以及遥感测绘等领域[3]，其通过三维激光扫描仪或者三维激光雷达等扫描设备来获取被测物体空间和表面纹理等信息，具有分辨率高，采集速度快以及非接触式等优点。三维扫描设备得到的是被测物体的三维点云信息，其抗干扰能力强，不受外界因素影响，从而更有利于目标识别与姿态估计[4]。然而，传统基于三维点云数据的目标识别方法有着计算数据量大以及速度慢等问题[5]，因此，对基于激光点云数据的三维目标识别方法进行研究具有十分重要的现实意义。

近年来，学术界对基于点云数据的三维目标识别方法进行了很多相关研究。在以深度神经网络为代表的深度学习还没有兴起之前，计算机视觉领域主要采用传统的机器学习方法对点云目标进行分类和检测，如支持向量机（SVM，support vector machine）[6]和决策树（Decision Tree）[7]等。文献［8］采用基于点云梯度的局部最优分割方法对激光雷达扫描的点云目标进行梯度分割以提取出障碍物轮廓，并根据障碍物三维点云数据特征，利用基于核的支持向量机有效完成了障碍物分类。然而，传统机器学习算法的鲁棒性能和泛化能力较差，无法满足现实生活中复杂场景下应用的需要。因此，研究人员将研究重点转向了对点云数据特性的挖掘上，引入降维思想将三维点云转

收稿日期:2021-08-19； **修回日期**:2021-09-24。
基金项目:浙江省自然科学基金(LY13F010008)；浙江省科技计划项目(2015F50009)。
作者简介:李豪杰(1994-)，男，硕士研究生，主要从事深度学习方向的研究。
杨海清(1971-)，男，博士，副教授，主要从事深度学习方向的研究。
引用格式:李豪杰，杨海清.基于深度卷积神经网络的点云三维目标识别方法研究[J].计算机测量与控制，2022，30(3)：156-160.

化为深度图，借助图像中的关键点检测方法提取点云特征点。文献［9］利用三维激光雷达传感器的隐式拓扑将三维点云目标映射到二维图像上，并提出了一种基于深度直方图的移动对象半自动分割方法和引入变分图像修复方法来重建被物体遮挡的区域，实际三维激光雷达街道场景验证了该算法的有效性。

随着人工智能时代的到来，深度学习算法在目标识别任务中取得了广泛的应用和突破性的进展[10]。典型的卷积结构需要高度规则的输入数据格式，所以无序散乱点云数据首先需要转换为常规的三维体素网格或图像集合[11]，这会导致数据不必要的损失，且带来额外的工作量。针对上述问题，文献［12］提出了一种三维点云目标分类和语义分割神经网络－Pointnet，该模型保证了输入点云的排列不变性，学习每个输入点对应的空间编码并通过对称函数得到全局特征，Pointnet 为从目标分类到场景语义分析应用程序提供了统一的体系结构，测试效果证明该模型的有效性和优越性。

针对 Pointnet 网络无法获得空间点局部结构信息的问题，文中将方向卷积编码方式应用到深度卷积网络中，通过对由 FPS 算法选取的局部区域中心点进行 3 次方向编码卷积来捕获局部形状特征。同时，针对点云数据旋转性会造成目标识别结果不稳定的问题，引入空间变换网络（STN，spatial transform network）来使点云目标识别模型具有空间不变性，从而进一步提高目标识别精确度和鲁棒性。在 ModelNet40、ShapeNetCore 数据集上的实验结果表明了文中提出的点云目标识别方法的有效性和优越性。

1　三维点云数据特征

1.1　旋转性

在获取点云数据时，三维扫描设备的旋转会导致不同时刻采集的同一点云目标的空间坐标信息（x，y，z）发生旋转变化，如图 1 所示。

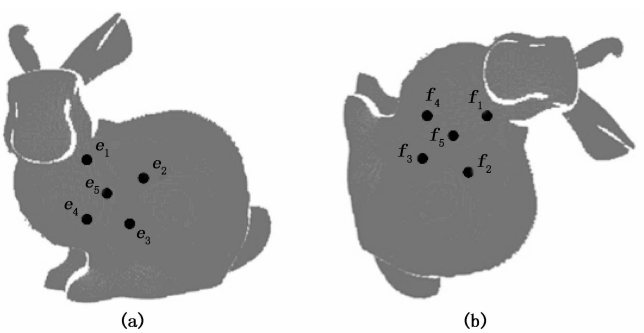

图 1　点云旋转性

虽然图 1 中的（a）和（b）是相同点云，但坐标值因为经过旋转变化而发生改变，对应的卷积操作结果分别入式（1）和式（2）所示：

$$G_a = Conv(K,[e_1,e_2,e_3,e_4,e_5]) \tag{1}$$

$$G_b = Conv(K,[f_5,f_4,f_3,f_2,f_1]) \tag{2}$$

其中：G 为卷积结果，$e_1 \sim e_5$ 和 $f_1 \sim f_5$ 为输入点云的各个点的坐标信息，$Conv$（·）为卷积操作。

根据式（1）～（2）的计算结果，$G_a \neq G_b$。虽然（a）和（b）是相同的点云，但是因为经过旋转，坐标发生改变，卷积结果也不同，所以卷积操作对点云的旋转变化不具有鲁棒性。文献［12］采用 T-net 姿态对齐网络对点云的旋转特征进行学习，但是受限于训练数据规模，以及点云旋转特征难以捕捉等问题，模型效果还有待提高。因此，文中提出一种空间变换网络来更好地解决点云旋转性问题。

1.2　排列不变性

三维点云通常呈无规则随机分布，且点与点之间没有顺序之分，具有排列不变性，每一组点云数据可以有 $N!$（N 为点数）种排列方式，即相同的点云可以有 $N!$ 种矩阵表示，如图 2 所示。

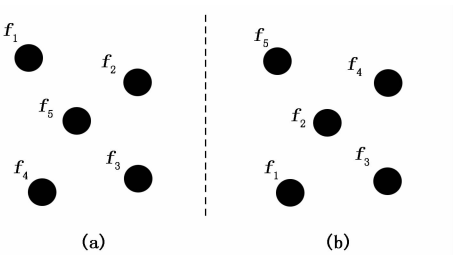

图 2　点云排列不变性

图 2 中，(a) 和 (b) 是相同的点云数据，但矩阵表示不同。假设分别对 (a)、(b) 中的点云数据进行卷积操作，如式（3）～（4）所示：

$$G_a = Conv(K,[f_1,f_2,f_3,f_4,f_5]) \tag{3}$$

$$G_b = Conv(K,[f_5,f_4,f_3,f_2,f_1]) \tag{4}$$

其中：G 为卷积结果，$f_1 \sim f_5$ 为输入点云的各个点的坐标信息，$Conv$（·）为卷积操作。

根据式（3）～（4）的计算结果，$G_a \neq G_b$。虽然（a）和（b）是相同的点云数据，但是因为矩阵表示不同，卷积结果也不相等，所以卷积操作无法保证点云的排列不变性。因此，将点云数据直接输入到传统卷积神经网络的进行学习的方法存在困难。为了解决上述问题，文中采用 max-pooling 池化操作解决点云数据排列不变性问题。

2　三维点云目标识别模型

2.1　空间变换网络（STN）

为了使三维点云目标识别模型具有空间不变性，即对于发生旋转变换的点云输入，模型仍能够对其正确进行分类，本文采用空间变换网络（STN，spatial transform networks）来自适应三维点云的旋转变换，将数据进行空间变换和对齐[13]。STN 由本地化网络（Localisation network）、网格生成器（grid generator）及采样器（sampler）3 个部分构成，如图 3 所示。

本地化网络以 feature map 或者是点云数据为输入，输出为空间变换所需的参数 θ，变换矩阵可以为任意形式。网

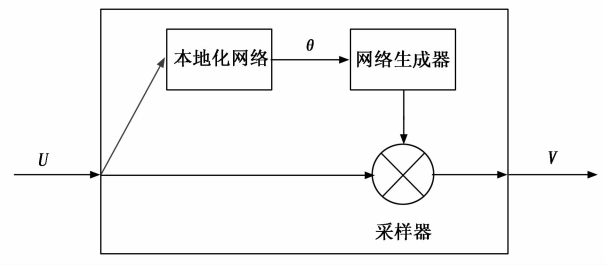

图 3 STN 结构

格生成器通过 θ 和定义的空间变换方式得出输出 V 与输入 U 的映射 $T(\theta)$，即实现点云坐标的对应关系，如式（5）所示：

$$\begin{pmatrix} x_i^s \\ y_i^s \\ z_i^s \end{pmatrix} = T_\theta(G_i) = A_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ z_i^t \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ \theta_{31} & \theta_{32} & \theta_{33} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ z_i^t \end{pmatrix} \tag{5}$$

其中：(x_i^s, y_i^s, z_i^s) 为输入的点云或者 feature map 中每个点的位置坐标，(x_i^t, y_i^t, z_i^t) 为经过空间变换后的每个点的位置坐标，A_θ 为映射关系。

输出 V 的所有坐标点是先定义好的，根据 A_θ 和 V 中每个坐标就可以计算出输入 U 的坐标，为了使求得的 U 中的坐标为整数，利用双线性差值法进行取值，采样器根据该坐标点获取到 U 中的特征，并将其填充到输出 V 中，如式（6）所示：

$$V_i = \sum_n \sum_m U_{nm} * \max(0, 1 - |x_i^s - m|) * \max(0, 1 - |y_i^s - n|) \tag{6}$$

STN 可用于输入层，也可插入到卷积层或者其它层的后面，不需要改变原 CNN 模型的内部结构[14]。

2.2 深度卷积点云特征提取网络

将 STN 应用到传统的深度卷积神经网络中，以避免三维点云旋转性造成的网络识别结果不稳定，并采用 max-pooling 差异化对称函数来解决因点云数据排列不变性导致的点云数据无法

直接输入到传统 CNN 网络的问题，搭建的深度 CNN 点云特征提取网络入图 4 所示。

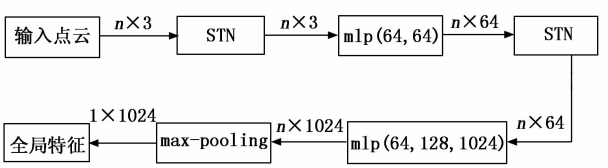

图 4 点云特征提取网络结构

网络中重复的 mlp 是通过共享权重的卷积实现的，第一层是 1×3 卷积核（对应三维坐标输入），之后都是 1×1 大小的卷积核。经过两个空间变换网络和两个 mlp 之后，将原始输入的三维特征映射到高维空间，通过 max-pooling 层得到 1×1 024 的全局特征。最后经过全连接层得到 k 个 score，连接 Softmax 输出层得到分类结果。

针对图 4 中点云特征提取网络无法提取点云局部拓扑特征的问题，引入了方向卷积编码方法，如图 5 所示。

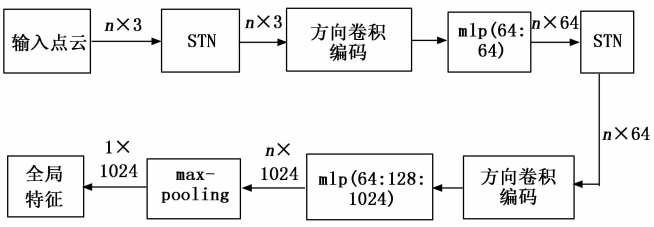

图 5 点云局部特征提取网络

首先，采用 FPS 采样方法[15]选取局部区域中心点，具体过程为：先随机选择一个点，再以离此点最远的点为起点继续迭代，直至获得需要的点数，该方法相比随机采样能够更完整得通过区域中心点采样到全局点云。

然后，对中心点分别沿 X，Y 和 Z 轴 3 个方向进行 3 级编码卷积来捕获局部形状特征，将中心点的特征放入张量 $M \in R^{2\times2\times2\times d}$，三阶段定向卷积如式（7）～（9）所示：

$$M_1 = g[Conv_x(A_x, M)] \in R^{2\times2\times d} \tag{7}$$

$$M_2 = g[Conv_y(A_y, M_1)] \in R^{2\times d} \tag{8}$$

$$M_3 = g[Conv_z(A_z, M_2)] \in R^{1\times d} \tag{9}$$

A_x，A_y，A_z 是要优化的卷积权重，$Conv_x$，$Conv_y$ 和 $Conv_z$ 是沿 X，Y 和 Z 轴方向的卷积，g 是激活函数。经过方向编码卷积后，每个点被表示为能够以方向编码方式表示中心点周围的形状图案的 d 维度的矢量。

2.3 点云目标分类

Softmax 回归是逻辑函数在多类分类问题上的推广[16]，其可以将一个含任意实数的 K 维向量映射到到另一个 K 维实向量中，并保证向量中每个元素值都在（0，1）之间且所有元素的和为 1。文中采用 Softmax 回归函数对上一节深度卷积网络提取的全局特征进行处理，得到每一个类别的概率值，如图 6 所示。

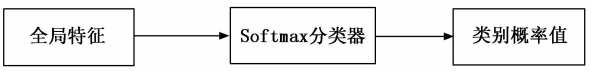

图 6 点云目标分类结构

对于给定样本，Softmax 回归预测的是属于某一类别的概率如式（10）所示：

$$p(y = c \mid x) = \text{softmax}(w_c^T x) = \frac{e^{w_c^T x}}{\sum_{c=1}^{C} e^{w_c^T x}} \tag{10}$$

其中：x 为样本，c 是类别，w_c 是第 c 类的权重向量。

3 试验结果与分析

实验的硬件环境为 Intel Xeon W-2123 处理器，Tesla v100 32G 显存显卡。

3.1 三维点云目标识别

实验使用 ModelNet40[17] 和 ShapeNetCore[18] 两个形状分类基准数据集来验证模型的性能。ModelNet40 数据集是

由 40 类人造目标产生的 12 311 个数据样本组成，该数据集将其中 9 843 个数据模型作为训练集，剩余 2 468 个数据作为测试集。而 ShapeNetCore 数据集更为丰富，有高达 51 300 数据样本总数，分为 55 个类别，数据集按 70%/10%/20%的比例划分为训练机、验证集和测试集。

网络模型叠加不同区域的局部特征获得的全局特征可进一步用于目标识别，如图 5 网络结构所示。在网格表面区域均匀的选取 1 000 个点，将每个点的空间坐标归一化后输入到网络模型中，同时为了提高模型的泛化能力，采用数据增强[19]的方法，对点云数据绕 Z 轴随机旋转一定角度，并在每个点的空间坐标加上均值为 0 标准差为 0.02 的高斯噪声。

在相同数据集上使用基于体素化、多视角二维图像的传统点云数据处理方法，以及 PointNet 基准点云处理卷积网络来和文中提出的直接输入数据的三维点云目标识别模型作对比，得到的结果如表 1 和表 2 所示。

表 1　不同模型在 ModelNet40 上的目标分类效果比较

模型	数据类型	不同类平均准确率/%	测试集准确率/%
VoxNet[20]	体素	83.0	85.9
Subvolume[21]	体素	86.0	87.7
MVCNN[22]	图像	86.5	88.3
PointNet[12]	点云	86.2	88.1
文中模型	点云	88.1	89.3

表 2　不同模型在 ShapeNetCore 上的目标分类效果比较

模型	数据类型	不同类平均准确率/%	测试集准确率/%
VoxNet	体素	84.1	86.7
Subvolume	体素	87.3	88.7
MVCNN	图像	87.5	89.5
PointNet	点云	87.4	89.0
文中模型	点云	89.7	90.4

从表 1～2 可以看出，文中提出的点云识别模型相比于基于体素的方法与基准点云模型取得了最佳准确率，文中所提点云卷积网络模型直接采用点云数据输入，避免了复杂的手工提取特征过程。相比于点云基准网络 PointNet，文中所提模型加入了方向卷积编码模块具备利用点云局部特征的能力，在两个数据集上准确率分别提升了 1.2%、1.4%。

3.2　网络模型结构验证分析

三维点云目标经过刚性旋转后的坐标信息将发生改变，为了克服点云目标旋转所造成的识别结果不稳定，在已有的深度卷积网络结构中引入 STN 方法来自适应的将数据进行空间变换和对齐，和其它对无序点云输入数据处理方法在 ModelNet40 数据集上进行对比，结果如表 3 所示。

表 3　STN 网络对模型准确率的影响

STN 网络位置	准确率/%
None	86.6
输入层	87.9
中间特征层	87.1
中间特征层＋正则化	87.5
输入层＋特征层＋正则化	89.3

从表 3 中可以看出，在输入层之后加入 STN 之后，模型的识别率提高了 1.3%。此外，对 STN 加入正则化约束以后，识别率进一步提高了 0.4%。本次实验输入的点云数据仅包含（x，y，z）三维坐标信息，在处理更高维的点云输入数据（如包含 RGB 颜色信息）时，采用正则化约束的识别效果将有更大的提升。

由于实际采集的点云数据的点云位置信息易受到腐蚀，部分点云数据会产生丢失，从而造成点云密度分布不均，因此需要进一步验证文中所提模型对采样数据的鲁棒性。在 ModelNet40 数据集上测试不同实验条件下的分类精确度，如图 7 所示。实验条件为随机删除一定比例的采样点数。

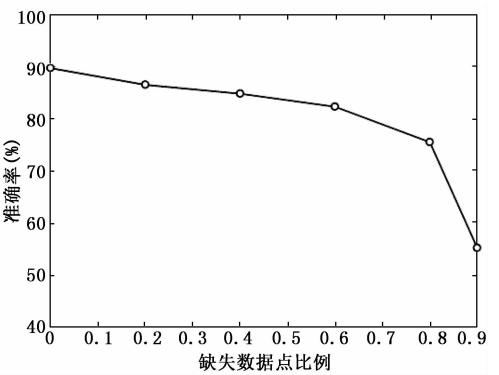

图 7　不同数据缺失比例下的识别率变化

从图 7 可以看出，在 60%采样点丢失的情况下，文中所提模型分类准确率仅下降了 7.4%，实验结果表明了点云方向编码卷积网络模型在处理输入数据缺失和不均匀时有较强的鲁棒性。

3.3　网络模型可视化

深度卷积点云目标识别网络实质上将输入的低维特征（$N*3$）映射到高维特征（$N*1\ 024$），再采取对称函数（max-pooling）来综合得到全局特征，整个网络结构如图 5 所示。为便于分析，将进入对称函数前一层的特征进行可视化，具体做法是标注在每个维度上取最大值的点云数据点坐标，得到这些标注的关键点坐标与原始点云目标如图 8 所示。

所标注的关键点决定了最后网络输出的全局特征，从图 8 中可以看出它们描绘了一个点云目标的大致骨架结构，这样意味着即使一些非关键点数据的缺失也不会影响到网

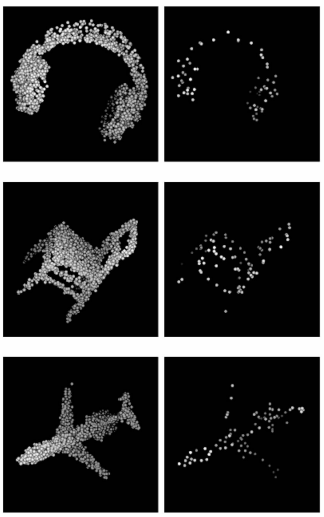

图 8 输入点云目标（左）与网络模型提取的关键点（右）对比图

络的最终判断，也证明文中提出的卷积点云目标识别网络的鲁棒性。

4 结束语

鉴于已有的深度卷积点云目标识别模型无法有效提取点云局部拓扑特征，文中通过 FPS 算法选取局部区域中心点，并对中心点进行 3 次方向编码卷积来捕获点云目标局部形状特征。同时，采用空间变换网络来解决因点云数据旋转性会造成导致的目标识别结果不稳定问题，进一步提高了目标识别精确度和鲁棒性。文中提出的点云目标识别方法有效提高了识别精度，相较于 PointNet 在 ModelNet40 和 ShapeNetCore 两个数据集分别提高 1.2%和 1.4%。

参考文献：

[1] 叶语同，李必军，付黎明. 智能驾驶中点云目标快速检测与跟踪 [J]. 武汉大学学报（信息科学版），2019，44（1）：139－144.

[2] 杨 林. 三维激光扫描技术在建筑工程施工变形监测中的应用研究 [D]. 天津：天津大学，2016.

[3] REN H，LIU Z，HAN Z，et al. Application of 3D Laser Scanning Technology in the Surveying and Mapping Goaf [J]. Gold Science & Technology，2013，21（3）：64－68.

[4] WANG H，CHENG W，LUO H，et al. 3-D Point Cloud Object Detection Based on Supervoxel Neighborhood with Hough Forest Framework [J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing，2017，8（4）：1570－1581.

[5] 杨永光. 基于点云的目标检测方法研究 [D]. 南京：南京邮电大学，2020.

[6] 雷 东. 基于联合准则与支持向量机的室内点云目标识别 [D]. 秦皇岛：燕山大学，2018.

[7] 雷 钊，习晓环，王 成，等. 决策树约束的建筑点云提取方法 [J]. 激光与光电子学进展，2018，55（8）：26－32.

[8] CHENG J，XIANG Z，CAO T，et al. Robust vehicle detection using 3d lidar under complex urban environment [C] // 2014 IEEE International Conference on Robotics and Automation（ICRA）. IEEE，2014：691－696.

[9] BIASUTTI P，AUJOL J F，BRéDIF M，et al. Range-Image：Incorporating sensor topology for LiDAR point cloud processing [J]. Photogrammetric Engineering & Remote Sensing，2018，84（6）：367－375.

[10] 柳碧辉，王培元. 基于深度学习的三维目标识别技术现状 [J]. 兵器装备工程学报，2021，42（8）：140－146.

[11] PAPON J，ABRAMOV A，SCHOELER M，et al. Voxel cloud connectivity segmentation-supervoxels for point clouds [C] // Proceedings of the IEEE conference on computer vision and pattern recognition，2013：2027－2034.

[12] QI C R，SU H，MO K，et al. Pointnet：Deep learning on point sets for 3D classification and segmentation [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition（CVPR）. New York：IEEE Press，2017：77－85.

[13] JADERBERG M，SIMONYAN K，ZISSERMAN A. Spatial transformer networks [J]. Advances in neural information processing systems，2015，28：2017－2025.

[14] 周飞燕，金林鹏，董 军. 卷积神经网络研究综述 [J]. 计算机学报，2017，40（6）：1229－1251.

[15] QI C R，YI L，SU H，et al. Pointnet＋＋：Deep hierarchical feature learning on point sets in a metric space [C] // The 24th Annual Conference on Neural Information Processing Systems. Cambridge：MIT Press，2017：5105－5114.

[16] WANG H，CHEN Y，LI Y. Face recognition method based on principal component analysis and Softmax regression model [J]. Journal of Hefei University of Technology（Natural Science），2015，38（6）：759－763.

[17] CHEN S，ZHENG L，YAN Z，et al. VERAM：View-Enhanced Recurrent Attention Model for 3D Shape Classification [J]. IEEE Transactions on Visualization & Computer Graphics，2018，25（12）：3244－3257.

[18] SAVVA M，YU F，SU H，et al. Large-scale 3D shape retrieval from ShapeNet Core55：SHREC'17 track [C] // Proceedings of the Workshop on 3D Object Retrieval，2017：39－50.

[19] 基于 BPL 数据增强的手写公式识别 [D]. 哈尔滨：哈尔滨工业大学，2018.

[20] MATURANA D，SCHERER S. Voxnet：A 3d convolutional neural network for real-time object recognition [C] // 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems（IROS）. IEEE，2015：922－928.

[21] QI C R，SU H，NIENER M，et al. Volumetric and multi-view cnns for object classification on 3d data [C] // Proceedings of the IEEE conference on computer vision and pattern recognition，2016：5648－5656.

[22] SU H，MAJI S，KALOGERAKIS E，et al. Multi-view convolutional neural networks for 3d shape recognition [C] // Proceedings of the IEEE international conference on computer vision，2015：945－953.