

## 基于多种强调机制的深度点云网络改进研究

刘慧, 田帅华

引用本文

刘慧, 田帅华. 基于多种强调机制的深度点云网络改进研究[J]. 计算机科学, 2023, 50(6A): 220400164-7.

LIU Hui, TIAN Shuaihua. [Study on Improvement of Deep Point Cloud Network Based on Multiple Emphasis Mechanisms](#) [J]. Computer Science, 2023, 50(6A): 220400164-7.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

**Similar articles recommended (Please use Firefox or IE to view the article)**

### [基于机器视觉的驾驶员视野盲区安全预警方法研究](#)

Study on Safety Warning Method of Driver's Blind Area Based on Machine Vision

计算机科学, 2023, 50(6A): 220700141-7. <https://doi.org/10.11896/jsjcx.220700141>

### [基于注意力机制和可变形卷积的路面裂缝检测](#)

Pavement Crack Detection Based on Attention Mechanism and Deformable Convolution

计算机科学, 2023, 50(6A): 220300214-6. <https://doi.org/10.11896/jsjcx.220300214>

### [基于动态金字塔和子空间注意力的图像超分辨率重建网络](#)

Image Super-resolution Reconstruction Network Based on Dynamic Pyramid and Subspace Attention

计算机科学, 2022, 49(11A): 210900202-8. <https://doi.org/10.11896/jsjcx.210900202>

### [基于机器视觉的Web应用页面元素识别及可视化脚本生成](#)

Web Application Page Element Recognition and Visual Script Generation Based on Machine Vision

计算机科学, 2022, 49(11): 65-75. <https://doi.org/10.11896/jsjcx.220200122>

### [三维激光雷达点云空间多特征分割](#)

Spatial Multi-feature Segmentation of 3D Lidar Point Cloud

计算机科学, 2022, 49(8): 143-149. <https://doi.org/10.11896/jsjcx.210300275>

# 基于多种强调机制的深度点云网络改进研究

刘 慧 田帅华

北京建筑大学电气与信息工程学院 北京 100044

**摘 要** 机器视觉是机器人从复杂空间环境中识别工作对象的关键技术。在机器人系统中常用的 Kinect 深度相机或激光扫描传感器能够获取目标的三维信息,这使得机器人完成更加复杂的如组装、拆卸、抓取等工作任务成为可能。但是,这也对机器人系统处理三维信息的能力如三维定位、工作对象尺寸测量、估计等提出更高要求。以 PointNet 网络为基础,分析了软阈值挤压激励、通道门控、注意力等机制的主要特征强调机理,分别采用软阈值挤压激励、通道门控、注意力网络对 PointNet 网络进行改进,并在斯坦福大学公开的 ShapeNet 数据集上进行实验验证。结果表明,3 种强调机制对原网络的改进,使三维点云的分割精度(均交并比)较 PointNet 原网络分别提高了 0.24%,0.68%,0.93%。该改进方法为后续解决机器人在组装、拆卸、抓取等任务中对工作对象的尺寸精确估计奠定了基础。

**关键词:** 机器视觉;三维点云;挤压激励;通道门控;注意力模块

**中图分类号** TP391

## Study on Improvement of Deep Point Cloud Network Based on Multiple Emphasis Mechanisms

LIU Hui and TIAN Shuaihua

School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

**Abstract** Machine vision is a key technology for robots to identify working objects from complex spatial environments. Kinect depth cameras or laser scanning sensors commonly used in robotic systems are capable of acquiring three-dimensional information about the target, which makes it possible for robots to perform more complex work tasks such as assembly, disassembly, and grasping. However, this also places higher demands on the robot system's ability to process 3D information such as 3D localization, work object size measurement, and estimation. We analyze the main feature emphasis mechanisms of soft threshold squeeze-and-excitation, channel-wise gated, and attention mechanisms based on PointNet networks, and improve PointNet networks by using soft threshold squeeze-and-excitation, channel-wise gated, and attention networks, respectively, and experimentally validate them on the publicly available ShapeNet dataset from Stanford University. Experimental results show that the improvement of original network by the three emphasis mechanisms improves segmentation accuracy (mean intersection and merge ratio) of 3D point clouds by 0.24%, 0.68%, and 0.93%, respectively, in comparison with original PointNet network. The improved method lays foundation for the subsequent solution of accurate estimation for the size of working objects in tasks such as assembly, disassembly and grasping by robots.

**Keywords** Machine vision, 3D point cloud, Squeeze-and-excitation, Channel-wise gated, Attention module

## 1 引言

自 2010 年微软推出 Kinect 视觉传感器之后<sup>[1]</sup>, Kinect 极强的探测能力便使得它在机器人视觉领域得到广泛应用。结合视频及图像的检测、识别等技术,搭载该传感器的机器人给人类生产生活带来诸多便利。例如,在教育领域, Kinect 的 2.5 维信息获取能力(二维图像+深度信息)结合虚拟现实技术,实现教学内容传递<sup>[2]</sup>;在餐饮方面,机器人通过实时监测老年人饮食摄入,为老人提供合理饮食安排<sup>[3]</sup>;在制造装备领域,在机器人上搭载视觉 Kinect 传感器,可以检测工作中物体的位置、形状和尺寸<sup>[4]</sup>;在运输领域,机器人通过人机交互系统,突破无人送货技术<sup>[5]</sup>;在医疗行业, Kinect 传感器用于医学成像,结合神经网络等技术实现对病人病情的检查

和诊断<sup>[6]</sup>。由此可见,传感器技术发展极大地提高了信息获取能力,但是仍然要依赖机器视觉等后处理软技术才能实现上述复杂的应用场景。

机器视觉技术是利用视觉传感器来探测和获取机器人所处的环境信息<sup>[7]</sup>,进而利用数据处理技术帮助机器人系统完成工作任务。国外的机器视觉技术发展较早,20 世纪 70 年代, Marr 等学者系统地创立了计算机视觉体系,为机器视觉技术发展奠定了基础。1988 年, John 等提出主动视觉概念<sup>[8]</sup>,机器视觉技术了飞速发展的阶段。直到 20 世纪 90 年代,三维视觉技术开始蓬勃发展。三维视觉技术可以类似于人的视觉功能测量复杂的空间环境的三维信息。2018 年, Kim 等利用移动机器人视觉系统获取物体点云数据并进行配准<sup>[9]</sup>。同年, Emrah 等利用基于机器视觉技术的伺服目标

基金项目:国家自然科学基金(62176018)

This work was supported by the National Natural Science Foundation of China(62176018).

通信作者:刘慧(liuhui@bucea.edu.cn)

行为控制器,通过高斯函数算法加强了移动机器人的高精度控制,更好应用到了机器人视觉目标检测,不过对检测环境中复杂多目标的应用还不够完善<sup>[10]</sup>;2019年,Khairudin等利用视觉跟踪系统并使用图像处理方法来控制机器人头部,使得机器人更好地对物体识别和定位<sup>[11]</sup>;2020年,Anna等利用机器人视觉系统完成苹果干采摘任务,对果园苹果进行检测,通过YOLOv3算法提高了机器人对苹果的识别率<sup>[12]</sup>。

我国的机器视觉技术起步较晚,在21世纪初期发展起来。例如,东南大学自动化学院测控重点实验室 Zheng等将机器视觉技术与工业机器人相结合,通过快速教学方法实现了焊接和喷涂任务<sup>[13]</sup>;西安理工大学机电工程学院 Tang等利用双目视觉系统研究机器人图像特征提取与匹配方法,通过加速鲁棒特征算法进入,使得搬运机器人定位精准、搬运高效<sup>[14]</sup>;重庆邮电大学工业物联网与网络控制教育部重点实验室 Li等利用机器人视觉系统检测障碍物,通过改进的YOLOv3算法提高移动机器人局部动态避障能力,并在实际环境中进行了可行性和有效性的验证<sup>[15]</sup>。

在机器人三维视觉中,三维点云的高质量获取是关键。加拿大伦敦西安大略大学 Chaudhury等通过机载3D激光扫描仪从多个视图捕获植物的表面点云数据,并通过高质量的三维点云重建算法获取植物的表面积和体积,准确分析出植物在不同环境下的生长规律<sup>[16]</sup>;澳大利亚国立大学 Cheraghian等采用3D深度相机获取物体三维点云信息,试图利用零镜头学习方法对系统中首次出现的对象类别实现分类,并提出评价这类系统的标准协议<sup>[17]</sup>;摩洛哥穆莱伊斯梅尔大学 Mahdaoui等使用3D设备获取密集三维点云数据,使用聚类和香农熵算法对冗余点云数据优化,获取了高质量的点云<sup>[18]</sup>。在我国,研究人员也开展了机器人利用三维点云开展工作的研究。重庆市勘测院梁建国等利用投影分布熵的方法对地面激光点云自动配准,取得了良好的配准效果<sup>[19]</sup>;北京信息科技大学自动化学院兰文昊等通过计算机视觉技术,结合深度学习算法对三维点云进行配准,提高了配准精度<sup>[20]</sup>。

综上,三维点云的处理结果直接关系到机器人的工作能力。如通过机器人完成物体的拆卸、组装、抓取,完成墙壁刷漆或检验产品的合格率等,都需要通过三维点云精准定位物体部件。如图1所示,物体拆卸组装系统包括物体检测系统、物体定位系统和机器人拆卸组装系统。物体通过 Kinect 识别后,经过计算机处理完成机器人对物体的拆卸组装工作。因此,机器人更精确地识别物体的类别部件,并准确处理信息、三维定位和自动抓取,是提高工作效率的关键。随着深度学习的流行,便有许多学者试着将深度学习和三维点云分割结合起来进行特征提取训练。但传统的卷积神经网络需先将三维点云数据处理为多视图或体素化网格的形式才能对其训练,对三维数据进行转换的过程会造成空间信息特征的丢失,导致三维点云分割准确率低下等问题。近几年,基于原始点云数据的点云分割方法渐渐成为了点云分割的主流发展方向。关键三维数据集 ShapeNet<sup>[21]</sup>的发布使得三维深度学习热潮激增<sup>[22]</sup>。2017年,Qi等首次提出 PointNet<sup>[23]</sup>网络,奠定了基于三维原始点云网络模型研究的基础。虽然原始三维点云数据的噪声、稀疏和有序等对其分割具有一定难度,但这样做可以充分利用点云的原始属性和空间特征信息等特性,从而避免了对其转换数据格式带来信息丢失的问题。渐渐地,通过

深度学习对三维点云数据直接处理成为主流。

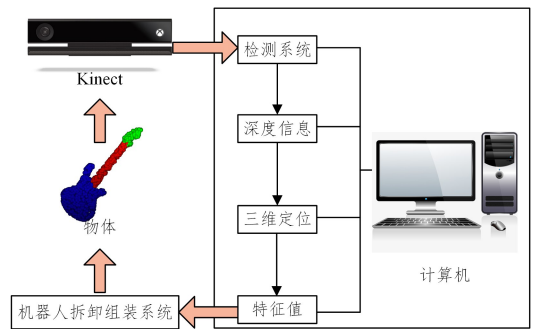


图1 物体拆卸组装系统结构示意图

Fig.1 Schematic diagram of the structure of object disassembly and assembly system

源于深度学习中图像分类技术的思想启发,2017年 Hu等<sup>[24]</sup>提出挤压激励网络(Squeeze-and-Excitation Networks, SENet),并对挤压激励网络引入了软阈值思想,改善了卷积之间的通道关系,提高了数据特征提取的性能;2018年 Woo等<sup>[25]</sup>提出一种卷积块的注意力模块,其可以添加到所有常规的卷积层中,显著提高了数据在通道和空间维度上的特征提取能力;2021年 Li等<sup>[26]</sup>提出一种通道门控机制,根据输入动态选择通道特征,以抑制不太相关的通道并增强检测数据的泛化性,提高了数据的鲁棒性。本文结合上述不同的强调机制对 PointNet 三维点云分类分割网络进行改进,实现三维点云的高精度分类分割,为后续解决机器人在组装、拆卸、抓取等任务中对工作对象的尺寸估计奠定技术基础。论文采用斯坦福大学公开的 ShapeNet 数据集作为训练样本进行实验验证。针对三维点云数据进行物体部件分割训练,通过应用上述强调机制的网络,显著改善了原始三维点云数据分类分割训练的性能。实验结果表明,本文所提出的强调机制改进方法,如结合软阈值的挤压激励网络、通道门控机制以及注意力模块,使得三维点云的分割精度(均交并比)较原 PointNet 网络分别提高了 0.24%,0.68%,0.93%。

## 2 PointNet 网络模型

PointNet 网络是一个直接对点云数据进行处理的深度学习网络模型,它是使用对称函数来解决点云的顺序问题,具体表达式如下所示:

$$f(x_1, x_2, \dots, x_N) = g(h(x_1), h(x_2), \dots, h(x_N)) \quad (1)$$

其中, $x_i (i=1, 2, \dots, N)$ 表示点云数据的第  $i$  个点;函数  $h(x_i) (i=1, 2, \dots, N)$ 表示第  $i$  个点的局部特征,通常, $h(\cdot)$ 被称作多层感知机(MultiLayer Perceptron, MLP)<sup>[27]</sup>;  $g(\cdot)$ 表示点集全局特征的对称函数,由单变量函数和最大池化函数的组合实现,是通过最大池化(max-pooling)操作将集合中所有点的局部特征聚合而来。

PointNet 网络模型如图2所示。输入为  $N \times 3$  的点云数据, $N$ 代表点云的个数, $3$ 代表点云有3个维度  $X, Y, Z$ 。输入的点云数据经过 T-Net 网络(如图2中黑色虚线框中所示),使点云数据特征统一对齐。整个 PointNet 网络中,存在两个 T-Net 模块。第一个 T-Net 是一个  $3 \times 3$  的旋转矩阵,该旋转矩阵将原始点云数据变换到标准三维坐标系下。第二个 T-Net 是一个  $64 \times 64$  的旋转矩阵,它对第一 MLP 的输出数据进行操作,将输出特征数据变换到标准 64 维特征坐标系下。

第一多层感知机为两层的卷积神经网络,每层 64 个滤波器,得到  $N \times 64$  的特征数据;第二多层感知机为三层的卷积神经网络,分别为 64,128,1024 个滤波器,得到  $N \times 1024$  的点云

特征;第三多层感知机为三层的全连接神经网络,分别有 512,256, $k$  个节点,实现  $k$  个分类。在第二多层感知机和第三多层感知机之间通过最大池化操作,得到全局特征。

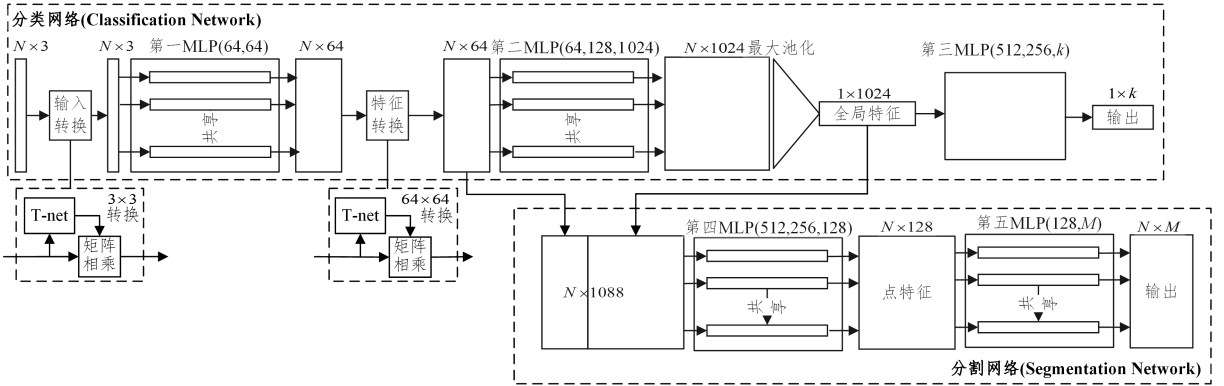


图 2 PointNet 网络模型结构

Fig. 2 PointNet network model structure

将第二个 T-Net 输出的  $N \times 64$  局部特征和最大池化后的 1024 个全局特征组合,得到  $N \times 1088$  的局部+全局数据特征,经过第四和第五层感知机得到分割结果。其中,第四多层感知机为三层的卷积神经网络,分别为 512,256,128 个滤波器;第五多层感知机也为两层的卷积神经网络,分别为 128 和  $M$  个滤波器,实现  $M$  个部件分割。

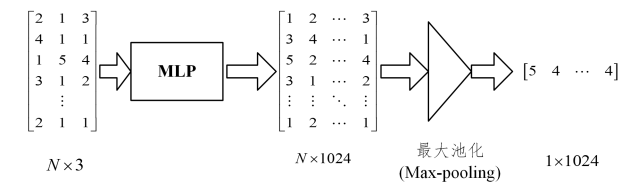


图 3 max-pooling 在 PointNet 网络模型结构中的示意图

Fig. 3 Schematic diagram of max-pooling in the structure of PointNet network model

PointNet 网络最主要的两个模块是 T-Net 和 max-pooling。1)T-Net 模块对数据进行空间或特征对齐操作,但是特征空间中的变换矩阵比空间变换矩阵维数高得多,这大大增加了优化的难度。因此,在 softmax<sup>[28]</sup> 训练损失中增加了一个正则化项。将特征变换矩阵限制为接近正交矩阵。

### 3 改进的 PointNet 网络模型

#### 3.1 多种强调机制

##### 3.1.1 结合软阈值的挤压激励网络

$$L_{reg} = \| I - AA^T \|_F^2 \quad (2)$$

$A$  是第二个 T-Net 网络输出的旋转矩阵。正交变换不会丢失数据信息,通过添加正则化项使得优化变得更加稳定。2)Max-pooling 模块不受点云顺序影响,将最显著的特征提取出来。如图 3 所示,输入点云为  $N \times 3$ ,经过多层感知后,为  $N \times 1024$ ,最大池化提取显著特征后变为  $1 \times 1024$  的点云全局特征。

挤压激励网络通过建模通道之间明确的相互依赖关系,自适应地重新校准通道内部的特征响应。将这种思想引入深度学习三维点云分类和分割网络中,通过建模特征之间的相互依赖关系,自适应地重新修正特征提取结果,来提高分类和分割准确度。挤压激励网络分为挤压模块和激励模块,该网络的具体流程如图 4 所示。

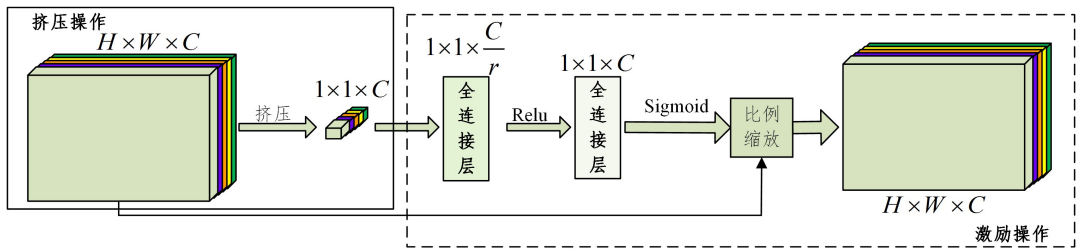


图 4 挤压激励(SE)模块具体流程图

Fig. 4 Specific flow chart of squeeze-and-excitation(SE) module

设挤压激励网络输入的是高  $H$ 、宽  $W$ 、通道数为  $C$  的数据。挤压激励网络对各特征进行重新评估,并根据评估结果修正特征提取的结果。

数据进行重新修正,该过程称为激励(如图 4 虚线框中所示)。

该过程如下:首先通过挤压操作将  $C$  个通道的  $H \times W$  的图像数据挤压成  $C$  个描述符,它们构成当前数据的全局特征,该过程称为挤压(如图 4 实线框中所示);然后通过节点数为  $C \rightarrow$  节点数为  $C/r \rightarrow$  节点数为  $C$  的全连接网络(称为降维系数),构建特征之间的相互依赖关系,经过全连接网络调整,获得新的全局特征,用该全局特征对输入

2019 年,Zhao 等<sup>[29]</sup>提出一种深度残差收缩网络,在收缩模块引入软阈值函数,用来解决数据去噪问题。软阈值函数<sup>[30]</sup>的表达式如下:

$$y = \begin{cases} x - \tau, & x \leq -\tau \\ 0, & |x| \leq \tau \\ x + \tau, & x \geq \tau \end{cases} \quad (3)$$

软阈值函数对  $x$  的偏导表达式如下:

$$f(x) = \frac{\partial y}{\partial x} = \begin{cases} 1, & x \leq -\tau \\ 0, & |x| \leq -\tau \\ 1, & x \geq \tau \end{cases} \quad (4)$$

其中,  $\tau$  是预先选定的阈值。而阈值的量化采用固定阈值法, 其定义为  $\tau = \sigma 2 \log(N)$ , 这里  $N$  为信号采样的数目,  $\sigma$  为噪声的标准偏差, 本文令  $\tau = 1$ 。

图 5 为本文提出的深度收缩网络模块, 该模块引入挤压激励(SE)网络结构(如图中黑色虚线框内所示)及软阈值特性, 忽略残差, 利用挤压激励网络学习训练软阈值, 用软阈值提高梯度的反向传播能力。具体过程为: 把上述 SE 模块高  $H$  变为 1、宽仍然为  $W$ 、通道数为  $C$  的特征数据经过两层标准化处理(BN)、ReLU 激活函数以及卷积操作之后, 输入到

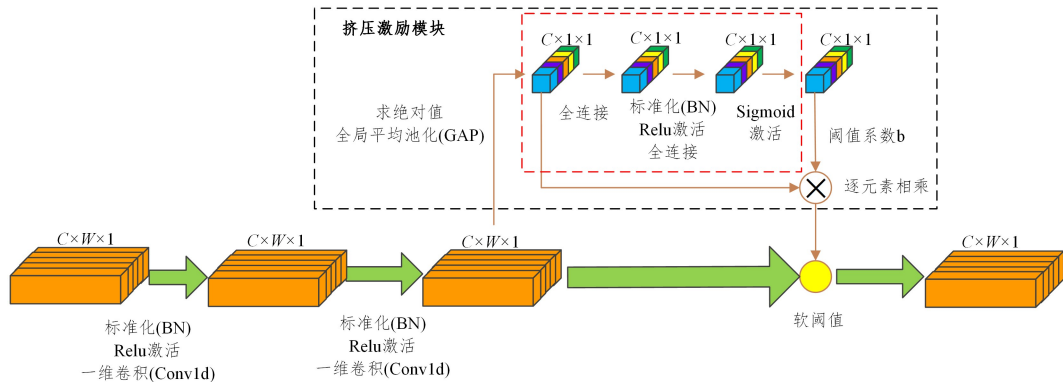


图 5 深度收缩模块

Fig. 5 Deep shrink module

### 3.1.2 注意力机制

文献[25]提出一种卷积块的注意力模块(Convolutional Block Attention Module, CBAM), 将注意力关注在通道和空间维度上, 其结构如图 6 所示。通过通道注意力机制和空间注意力机制, 将卷积之后重要的三维数据特征提取出来, 以提高准分类分割精度。

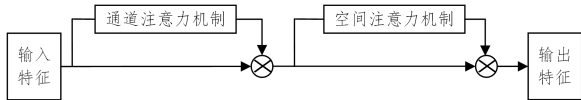


图 6 注意力模块

Fig. 6 Attention module

图 6 中通道注意力机制可以使用公式表示为:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ = \sigma(W_1(W_2(F_{avg}^C)) + W_1(W_0(F_{max}^C))) \quad (5)$$

其中,  $\sigma$  表示 sigmoid 函数,  $W_0 \in \mathbb{R}^{C/r \times C}$ ,  $W_1 \in \mathbb{R}^{C \times C/r}$  ( $r$  为减少率)。将输入的特征  $F$  分别经过全局最大池化和全局平均池化, 得到两个  $1 \times 1 \times C$  的特征, 接着再将它们分别送入一个两层的神经网络(MLP), 第一层神经元个数为  $C/r$ , 激活函数为 ReLU, 第二层神经元个数为  $C$ , 这个两层的神经网络(MLP)是共享的。然后, 将 MLP 输出的特征进行加和操作, 再经过 sigmoid 激活操作, 生成最终的通道注意力特征, 即  $M_c$ 。最后, 将  $M_c$  和输入特征  $F$  做乘法操作( $\otimes$ ), 生成空间注意力机制需要的输入特征  $F'$ 。空间注意力机制可以使用公式表示为:

$$M_s(F') = \sigma(f^{7 \times 7}[AvgPool(F'), MaxPool(F')]) \\ = \sigma(f^{7 \times 7}[F'_{avg}^S, F'_{max}^S]) \quad (6)$$

其中,  $\sigma$  表示 sigmoid 函数,  $f^{7 \times 7}$  表示滤波器大小为  $7 \times 7$  的

图 5 的黑色虚线框内(通过挤压激励操作, 高和宽变为  $1 \times 1$ , 通道为  $C$ )。经过求绝对值、全局平均池化(Global Average Pooling, GAP)<sup>[31]</sup>后, 得到全局特征。全局特征通过全连接网络(如图 5 中红色虚线框)及 Sigmoid 激活函数得到归一化的阈值系数  $b$ , 将全局特征和阈值系数  $b$  相乘获得软阈值, 该软阈值用来提高梯度的反向传播能力。在该过程中, GAP 可以有效防止过拟合。深度收缩网络融合了 SE 模块和软阈值函数, 将 SE 模块中的“重新加权”操作替换成了“软阈值”操作。该收缩网络模块中的软阈值应用相当于引入了注意力机制。注意力机制能够判断全局特征与当前任务的相关性, 并根据相关性抑制或者放大特征。将此强调机制应用到 PointNet 网络中, 分类分割效率精度显著提升。

卷积运算。将上述经过通道注意力机制生成的特征  $F'$  作为空间注意力机制的输入特征。首先做一个基于通道的全局最大池化和全局平均池化, 得到两个  $H \times W \times 1$  的特征, 然后将这两个特征基于通道做通道之间的拼接操作, 经过一个  $7 \times 7$  卷积操作降维为一个通道, 即  $H \times W \times 1$ 。再经过 sigmoid 函数生成空间注意力特征, 即  $M_s$ 。最后将  $M_s$  和空间注意力机制的输入特征  $F'$  做乘法操作( $\otimes$ ), 得到最终生成的特征, 用于后续处理操作。本文主要做的改进是将二维卷积进行一维卷积, 即使用强调机制 MLP, 显著提高了 PointNet 网络三维点云数据在通道和空间维度上的特征提取能力。

### 3.1.3 通道门控机制

文献[26]提出了一种通道门控机制, 其原理是根据输入动态选择通道特征, 以抑制不太相关的通道并增强检测数据的泛化性。本文参考该论文的 3 种不同结构的门控机制, 提出了效果较好的跳连接的双层网络门控模块, 如图 7 所示。

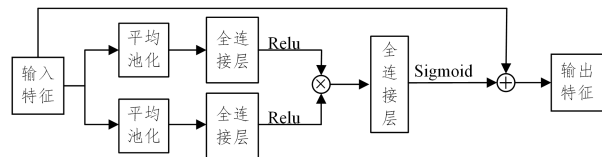


图 7 双层网络门控模块

Fig. 7 Double-layer network gate control module

图 7 可以用公式表示为:

$$\text{第一层: } L_1(x) = \delta(W_1^T F(x)) \quad (7)$$

$$\text{第二层: } L_2(x) = \delta(W_2^T F(x)) \quad (8)$$

$$\text{输出: } a = \sigma(W_3^T [L_1(x) \oplus L_2(x)]) \quad (9)$$

其中,  $\delta$  表示 ReLU 函数,  $\sigma$  表示 sigmoid 函数, 全连接层  $W_1$

和  $W_2$  为  $W_1, W_2 \in \mathbb{R}^{C \times C/r}, W_3 \in \mathbb{R}^{2C/r \times C}$  ( $r$  为减少率)。

$F(x) = \frac{1}{D \times T} \sum_{d=1}^D \sum_{t=1}^T x$  ( $D$  为频谱,  $T$  为时间), 即采用平均池化将频率和时间特征压缩到同一通道上。本文对通道门控机制进行改进, 将输入特征  $x$  设为相同, 两层网络分别经过平均池化 ( $F(x)$ ) 和全连接层 ( $W_1, W_2$ ) 相同的处理 (第二层可作为参考层), 再经过 ReLU 函数、相乘操作 ( $\otimes$ ) 以及全连接层 ( $W_3$ ) 和 sigmoid 函数之后得到  $a$ , 最后与输入特征  $x$  相加 ( $\oplus$ ) 得到输出特征。本文利用 PointNet 网络特性, 使用一维卷积将双层输入特征全部设为上一层的输出, 加入到 PointNet 网络中, 可以更好地捕获通道信息, 加强三维点云的数据特征提取能力。上述机制的有效性都在第 4 节通过实验得到了验证。

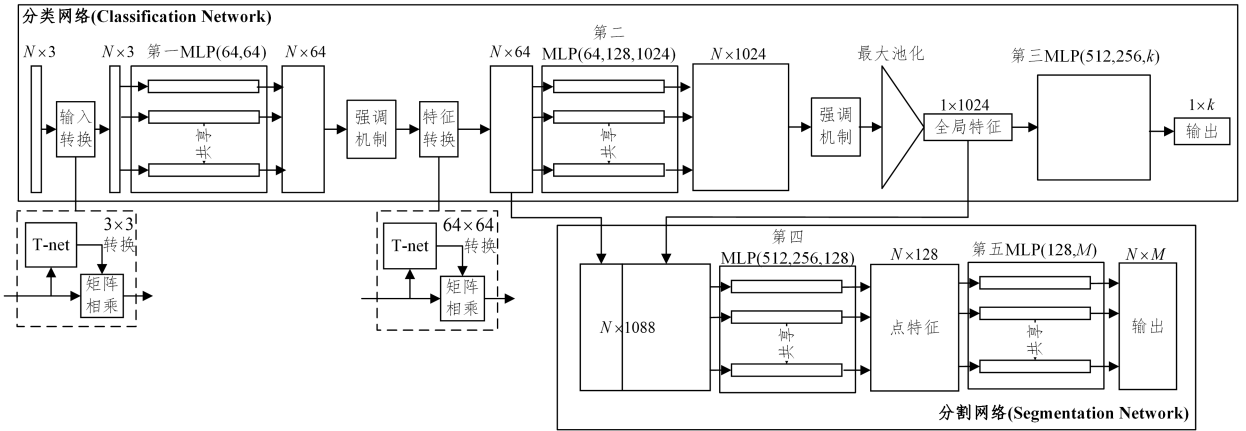


图 8 改进的 PointNet 网络模型

Fig. 8 Improved PointNet network model

## 4 实验结果及对比分析

实验所用 GPU 配置为 NVIDIA Quadro P1000, 4 GB 显存, 系统环境为 Windows10, 在 Python3.8 下搭建 CUDA10.2, CUDNN 7.6.5, PyTorch1.8.1 深度学习环境。

### 4.1 实验数据及评价标准

为验证本文提出的网络架构在处理三维点云分割任务上的效果, 采用公开数据集 ShapeNet 进行实验。该数据集共有 16 个类别、16 881 个形状, 包含了汽车、皮包、飞机、台灯、桌子等常见物体。每个类别中都单独标注了该类别物体的组件结构 (如皮包有提手和包身), 该数据集的 16 个类别最终一共被分出了 50 个组件结构。点云分类分割评价标准采用平均交并比 miou (mean intersection-over-union), 其表达式如下:

$$miou = \frac{1}{k+1} \sum_{i=0}^k \frac{T_p}{T_p + F_p + F_N} \quad (10)$$

其中,  $k$  代表分割的类别数,  $T_p$  代表分类正确的正样本数量,  $F_p$  代表分类错误的正样本数量,  $F_N$  代表分类错误的负样本数量。

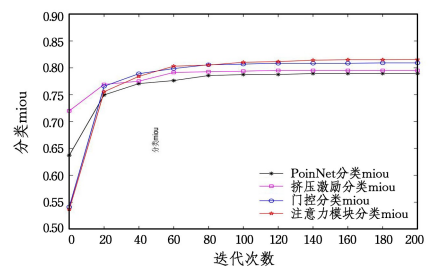
### 4.2 实验结果及分析

网络训练时, 超参数设置为: 初始学习率  $lr=0.001$ , 初始衰减率  $dr=0.5$ , 衰减速度  $=1/20$  次, 通道尺寸  $batch\_size=8$ 。优化算法为 Adam, 该算法中的参数设置为: 动量  $momentum=0.1$ , 迭代次数  $epoch=251$ 。每次每个通道输入样本 2048 个, 每次迭代训练的点数  $8 \times 2048$  个, 对所有训练样本总共迭代 251 次。

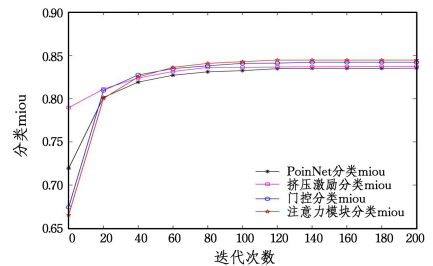
## 3.2 改进 PointNet 网络模型

根据上述原理将融合软阈值的挤压激励网络、通道门控机制以及注意力模块应用到 PointNet 网络第一和第二多层感知机之后, 保持上层与下层通道相同, 通过其网络学习非线性变换去判别点云数据特征, 可以有效消除次要点云数据特征, 并且该网络的阈值是在网络架构中自动学习训练的, 真正实现了端到端<sup>[32]</sup>的学习模型。图 8 给出了强调机制改进的 PointNet 网络架构, 整个 PointNet 网络中存在两个 T-Net 模块。第一个 T-Net 是一个  $3 \times 3$  的旋转矩阵, 该旋转矩阵将原始点云数据变换到标准三维坐标系下。第二个 T-Net 是一个  $64 \times 64$  的旋转矩阵, 它对第一多层感知机的输出数据进行操作, 将输出特征数据变换到标准 64 维特征坐标系下。3 种强调机制改进的 PointNet 网络分类分割实验结果将在下一节给出。

为了验证本文改进网络的分类、分割效果, 采用均交并比 miou 作为评价标准。图 9 给出了迭代次数从 1 到 200 的训练过程中, PointNet 网络和本文通过结合软阈值的挤压激励网络、通道门控机制以及注意力模块进行改进后的 3 种网络均交并比 miou。其中, 图 9(a) 为分类结果的均交并比 miou, 图 9(b) 为分割结果的均交并比 miou。



(a) 分类均交并比 miou 对比



(b) 分割均交并比 miou 对比

图 9 均交并比 miou 对比

Fig. 9 Comparison of miou

从图 9 中可以清楚地看到, 无论是分类还是分割结果, 当

迭代次数足够多时, PointNet 网络和改进后的 3 种网络训练的 miou 均能趋于稳定, 但是改进后的 3 种网络的分类分割均交并比 miou 都比 PointNet 网络有所提高; 当迭代次数较少时, 3 种改进的网络可以先于 PointNet 网络达到较高的均交并比。具体的 PointNet 网络及改进后的网络分类和分割均

交并比 miou 结果如表 1 所列。可以看出, 结合软阈值的挤压激励网络、通道门控机制以及注意力模块分割的均交并比 miou 较 PointNet 网络分别提高了 0.24%, 0.68%, 0.93%, 验证了结合软阈值的挤压激励网络、通道门控机制以及注意力模块 3 种改进网络的有效性。

表 1 16 个类别部件分割均交并比及其平均均交并比

Table 1 16 categories of parts segmentation miou and their average miou

(单位: %)

Model	Airplane	Bag	Cap	Car	Chair	Earphone	Guitar	Knife	Lamp	Laptop	Motorbike	Mug	Pistol	Rocket	Skateboard	Table	mean
PointNet	81.7	74.7	84.5	73.4	89.6	68.7	90.1	86.9	79.5	95.3	56.4	91.9	78.4	53.7	74.1	81.9	83.53
挤压激励	81.6	77.7	83.1	74.7	90.1	72.2	90.5	87.4	80.2	95.2	62.1	92.5	78.3	54.2	74.6	81.9	83.77 (↑ 0.24)
通道门控	82.3	82.4	86.3	75.8	90.3	75.6	91.0	86.7	81.2	95.4	62.6	93.9	81.5	54.3	74.5	81.9	84.21 (↑ 0.68)
注意力模块	82.8	78.5	90.2	75.3	90.4	79.2	91.1	86.7	81.7	95.5	66.1	95.0	82.2	56.2	74.9	81.8	84.46 (↑ 0.93)

图 10 给出了飞机、桌子、椅子、吉他、台灯和刀等典型物体的点云分割结果, 不同的颜色表示不同的部件类别, 例如飞机的机身、机翼和机尾的重要部位在点云中能够清晰区分。

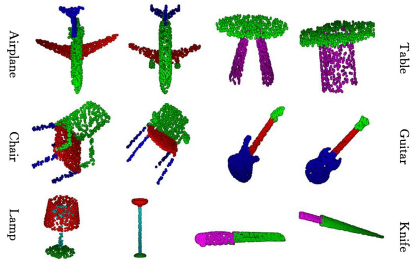


图 10 ShapeNet 数据集上部分类别部件分割效果图

Fig. 10 Segmentation effect diagram of some categories of component on ShapeNet dataset

**结束语** 本文在 PointNet 网络基础上, 通过结合软阈值的挤压激励网络、通道门控机制以及注意力模块进行改进, 并通过大量实验验证改进后网络对三维点云分类分割方法的实用性, 通过强调机制 PointNet 网络架构, 实现物体三维点云的分类分割处理。采用 ShapeNet 数据集验证改进网络的分类分割性能, 大量实验结果表明, 本文引入结合软阈值的挤压激励网络、通道门控机制以及注意力模块之后, 三维点云的分割精度(均交并比)较 PointNet 分别提高了 0.24%, 0.68%, 0.93%, 改进网络在稳定性、分类分割准确率、分类分割速度、分类分割均交并比等方面均有所提高。但是, 该网络仍然有很多地方需要完善, 未来的工作重心是研究网络的优化问题, 探索提高局部信息分割均交并比的处理方法, 进一步提高三维点云分割准确度, 以解决机器人准确测量和估计工作对象的尺寸等技术问题。

## 参考文献

- [1] XU X, MCGORRY R W. The validity of the first and second generation Microsoft Kinect for identifying joint center locations during static postures[J]. Appl. Ergon., 2015, 49: 47-54.
- [2] ZHOU Y, YU Z, XU X D, et al. Practice research of classroom teaching system based on Kinect[C]// 15th Int. Conf. Comput. Sci. Educ. (ICSE 2020). 2020: 572-575.
- [3] CUNHA A, PÁDUA L, COSTA L, et al. Evaluation of MS Ki-

nect for Elderly Meal Intake Monitoring[C]// Procedia Technol. 2014: 1383-1390.

- [4] CARUSO L, RUSSO R, SAVINO S. Microsoft Kinect V2 vision system in a manufacturing application[J]. Robot. Comput. Integr. Manuf., 2017, 48: 174-181.
- [5] BIERMANN H, PHILIPSEN R, BRELL T, et al. Users' Expectations, Fears, and Attributions Regarding Autonomous Driving: A Comparison of Traffic Scenarios[M]. Springer International Publishing, 2021.
- [6] JAWAID I, QURESHI J K. Advancements in medical imaging through Kinect: A review[C]// 2017 Int. Symp. Wirel. Syst. Networks (ISWSN 2017). 2017: 1-5.
- [7] FERNANDES A O, MOREIRA L F E, MATA J M. Machine vision applications and development aspects[C]// IEEE Int. Conf. Control Autom. (ICCA). 2011: 1274-1278.
- [8] ALOIMONOS J, WEISS I, BANDYOPADHYAY A. Active vision[J]. Int. J. Comput. Vis., 1988, 1(4): 333-356.
- [9] KIM P, CHEN J, CHO Y K. SLAM-driven robotic mapping and registration of 3D point clouds[J]. Autom. Constr., 2018, 89: 38-48.
- [10] DÖNMEZ E, KOCAMAZ A F, DIRIK M. A Vision-Based Real-Time Mobile Robot Controller Design Based on Gaussian Function for Indoor Environment[J]. Arab. J. Sci. Eng., 2018, 43(12): 7127-7142.
- [11] KHAIRUDIN M, CHEN G D, WU M C, et al. Control of a movable robot head using vision-based object tracking[J]. Int. J. Electr. Comput. Eng., 2019, 9(4): 2503-2512.
- [12] KUZNETSOVA A, MALEVA T, SOLOVIEV V. Using YOLOv3 algorithm with pre-And post-processing for apple detection in fruit-harvesting robot[J]. Agronomy, 2020, 10(7).
- [13] ZHENG F, FANG F, MA X. Trajectory Sampling and Fitting Restoration Based on Machine Vision for Robot Fast Teaching[C]// Proc. 15th IEEE Conf. Ind. Electron. Appl. (ICIEA 2020). 2020: 604-609.
- [14] TANG B, JIANG L. Binocular stereovision omnidirectional motion handling robot[J]. Int. J. Adv. Robot. Syst., 2020, 17(3): 1-11.
- [15] LI Y, LIU Y. Vision-based Obstacle Avoidance Algorithm for Mobile Robot[C]// Proc. -2020 Chinese Autom. Congr. (CAC 2020). 2020: 1273-1278.

- [16] CHAUDHURY A. Machine Vision System for 3D Plant Phenotyping[J]. IEEE/ACM Trans. Comput. Biol. Bioinforma. ,2018, 16(6):2009-2022.
- [17] CHERAGHIAN A, RAHMAN S, PETERSSON L. Zero-shot learning of 3d point cloud objects[C]// Proc. 16th Int. Conf. Mach. Vis. Appl. (MVA 2019). 2019.
- [18] MAHDAOUI A. 3D point cloud simplification based on the clustering algorithm and introducing the Shannon's entropy[C]// Thirteenth International Conference on Machine Vision. SPIE, 2021,11605:174-182.
- [19] LIANG J G, CHEN M L, MA H. Registration of Terrestrial Laser Scanning Data Based on Projection Distribution Entropy[J]. Laser & Optoelectronics Progress, 2019,56(13):131501.
- [20] LAN W H, LI N, TONG Q. Improved3-D Point Cloud Registration Algorithm with Oriented Bounding Box[J]. Computer Engineering and Applications, 2022, 58(14):177-184.
- [21] CHANG A X, FUNKHOUSER T, GUIBAS L, et al. Shapenet: An information-rich 3d model repository [J]. arXiv: 1512.03012, 2015.
- [22] GUO Y, WANG H, HU Q, et al. Deep learning for 3d point clouds: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(12):4338-4364.
- [23] QI C R, SU H, MO K, et al. PointNet: Deep learning on point sets for 3D classification and segmentation [C] // Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognition (CVPR 2017). 2017:77-85.
- [24] HU J. Squeeze-and-Excitation\_Networks\_CVPR\_2018\_paper.pdf[C]//CVPR. 2018:7132-7141.
- [25] WOO S, PARK J, LEE J, et al. CBAM: Convolutional Block Attention Module[C]//ECCV. 2018:3-19.
- [26] LI X, WU X, LU H, et al. Channel-wise gated res2net: Towards robust detection of synthetic speech attacks [J]. arXiv: 2107.08803, 2021.
- [27] TOLSTIKHIN I O, HOULSBY N, KOLESNIKOV A, et al. Mlp-mixer: An all-mlp architecture for vision[J]. Advances in Neural Information Processing Systems, 2021, 34:24261-24272.
- [28] LIU W, WEN Y, YU Z, et al. Large-margin softmax loss for convolutional neural networks[J]. arXiv:1612.02295, 2016.
- [29] ZHAO M, ZHONG S, FU X, et al. Deep Residual Shrinkage Networks for Fault Diagnosis[J]. IEEE Trans. Ind. Informatics, 2020, 16(7):4681-4690.
- [30] PENG Y H. De-noising by modified soft-thresholding[J]. IEEE Asia-Pacific Conf. Circuits Syst. ,2000, 41(3):760-762.
- [31] LIN M, CHEN Q, YAN S. Network in network(2nd)[C]//Int. Conf. Learn. Represent. ICLR 2014-Conf. Track Proc. 2014:1-10.
- [32] SALTZER J H, REED D P, CLARK D D. End-to-end arguments in system design[J]. ACM Trans. Comput. Syst. , 1984, 2(4): 277-288.



**LIU Hui**, born in 1978, Ph.D, associate professor. Her main research interests include indoor 3D positioning, TomoSAR 3D reconstruction and radar signal processing, 3D point cloud classification and microwave remote sensing image processing.