

基于空洞卷积胶囊网络的激光雷达数据分类

王爱丽¹, 张宇泉¹, 吴海滨^{1*}, 姜开元¹, 岩堀 祐之²¹哈尔滨理工大学黑龙江省激光光谱技术及应用重点实验室, 黑龙江 哈尔滨 150080;²中部大学计算机科学学院, 日本 爱知 487-8501

摘要 激光雷达(LiDAR)作为获取地物物理属性的一种重要技术手段,被广泛应用于分类研究中。针对传统卷积神经网络中标量神经元无法表达特征位置信息,以及图像经过连续池化和降采样后出现空间分辨率下降、细节信息丢失这两个问题,提出胶囊网络和空洞卷积相结合的 LiDAR 数据分类算法。该算法以残差网络架构为基础,首先将 LiDAR 数据输入深度卷积神经网络中进行特征粗提取,然后采用奇偶混合扩张率的空洞卷积增大训练中特征图的感受野,同时捕获精细特征,最后通过胶囊网络进一步提取更详细的空间特征信息并进行分类输出。实验结果表明,所提方法在 Bayview Park 和 Recology 两个数据集上分别取得了 97.07%和 96.98%的总体分类精度,与其他 7 种分类算法相比具有更好的分类效果,由此证明所提方法可以提升 LiDAR 数据的分类性能。

关键词 遥感; 激光雷达; 深度学习; 残差网络; 空洞卷积; 胶囊网络

中图分类号 TP391

文献标志码

doi: 10.3788/CJL202148.1110003

1 引言

激光雷达(LiDAR)是一种主动式遥感技术^[1],是通过向目标发射激光束来获取高精度的地表地形数据,其获取的数据在地形测绘、三维场景重建、城市规划等领域应用广泛。LiDAR 系统获取的点云数据通过去噪和栅格化处理可衍生出 LiDAR 数字表面模型(DSM)^[2],该模型是一种包含地表建筑物、桥梁和树木等高度的地面高程模型,在区分不同高度的地物上有极大优势,更适合应用于地物分类研究。

目前,许多研究者将机器学习算法用于 LiDAR-DSM 数据分类问题研究中,Naidoo 等^[3]采用自动随机森林模型对高光谱和 LiDAR 衍生的结构参数进行分类。Khodadadzadeh 等^[4]将从 LiDAR 数据和高光谱数据中提取的多个特征进行集成,并使用混合逻辑回归子分类器完成了地物分类任务。Ghamisi 等^[5]提出了一种结合消光剖面 and 复合核支持向量机的方法,并用于 LiDAR 数据的分类。但是传统机器学习算法的特征提取是基于手

工设计的特征,特征设计过程繁琐。

近年来,采用多层非线性变换方式自动提取特征深度学习算法已成为图像处理领域的主流,因此研究者们也开始将深度学习方法应用于 LiDAR 数据的分类任务中。Ghamisi 等^[6]利用包含不同属性的消光剖面通过特征叠加将高光谱和 LiDAR 的特征进行融合,再利用卷积神经网络(CNN)对融合后的特征进行分类。Xia 等^[7]将高光谱数据的光谱信息和空间信息与 LiDAR-DSM 数据的高程信息相融合,并采用集成分类器对融合后的数据进行分类,提高了分类准确性。Wang 等^[8]使用形态学特征和多属性特征进行数据增强,结合 CNN,改善了 LiDAR 数据的分类性能。Wang 等^[9]提出结合空间变换网络和密集卷积神经网络的网络模型,该模型具有较强的特征表达能力,有效提高了 LiDAR 数据分类的精确度。Ge 等^[10]融合高光谱图像与 DSM 数据,提出了一种基于扩展属性、局部二值模式和核协同表示的新框架。

为了融合神经网络中浅层网络提取的局部低层特征与深层网络提取的全局高层语义特征,近几年

收稿日期: 2020-11-17; 修回日期: 2020-12-23; 录用日期: 2021-01-04

基金项目: 国家自然科学基金(61671190)

*E-mail: woo@hrbust.edu.cn

人们更多地将残差网络应用于遥感图像领域。汪鹏等^[11]在 ResNet 中嵌入了跳跃连接和协方差池化两个模块,融合了不同层次的多分辨率特征信息,并且利用高阶信息实现了更具代表性的特征学习。张祥东等^[12]主要针对高光谱图像中小样本数据分类精度低的问题,在残差模块中加入分支结构,实现了空间特征和光谱特征的多尺度特征融合,充分利用了高光谱图像中丰富的空谱信息。朱天佑等^[13]针对硬件资源和功耗的限制,提出了一种具有像素级输出的轻量级残差网络,通过引入残差连接,有效提高了检测精度并且将模型大小压缩为原来的 1/16。

上述研究证实了深度学习方法用于解决遥感图像分类问题的可行性和有效性。但是传统的 CNN 使用神经元作为处理数据的单元,每个神经元只能识别一种模式,对图像中物体的方向和位置信息不敏感。为克服 CNN 的这一缺点,Sabour 等^[14]提出了名为胶囊网络(CapsNet)的新型网络,目前许多研究者将胶囊网络应用在图像分类领域。Raza 等^[15]将注意力机制引入胶囊网络,以学习更丰富、更鲁棒的场景分类特征,用于解决遥感图像类内多样性和类间相似性高的问题。Wang 等^[16]将胶囊网络与三重生成对抗网络相结合,提高了胶囊网络在少量训练样本上的分类性能。Chang 等^[17]提出多路径胶囊网络,减小了图像背景多样化对图像分类的影响,提高了对于复杂场景图像的分类精度。Poleetti 等^[18]提出了一种基于光谱空间胶囊网络的 CNN 新结构,实现了高光谱图像的精确分类,同时降低了网络设计的复杂性。

本文提出一种融合胶囊网络与空洞卷积的 LiDAR 数据分类算法,称为空洞卷积胶囊网络(DCCN)。空洞卷积在不增加网络参数的情况下扩

增了卷积层的感受野,避免了采用池化层等传统方法扩增时引起的特征丢失现象。同时为克服 CNN 对方向位置等信息的不敏感性,模型分类部分采用胶囊网络。胶囊网络在其内部结构中使用胶囊作为数据处理的单元,并采用动态路由算法进行胶囊层之间的数据传递,故该网络具有比 CNN 更好的特征表达能力。实验结果表明,本文提出的 LiDAR 数据分类模型与传统分类方法(支持向量机,随机森林)和基于深度学习的分类方法(CNN、CapsNet、ResNet、Dilated-ResNet、ResCapsNet)相比,分类精度显著提高。

2 本文算法

本文所提的 DCCN 网络结构如图 1 所示。首先将 LiDAR 数据输入 DCCN 模型,该模型利用残差网络(ResNet)的前三层对 LiDAR 数据进行特征粗提取,后两层采用空洞卷积替代传统卷积,以获取不同尺度的 LiDAR 图像地物的细节特征。其中,采用奇偶混合扩张率的方法,避免了由空洞卷积导致的网格效应,如图 1 上半部分虚线框所示。鉴于胶囊网络将传统神经网络的每个神经元由标量转换成向量,对图像内在结构的建模呈现一定的语义关系,同时保留了图像空间信息的优良特性,将 ResNet 的输出送入胶囊网络,利用胶囊网络的卷积层捕获数据特征,并在卷积层中使用批量归一化(BN),加入 BN 层这一操作解决了初始化差的问题,同时可保证梯度传播到每一层。然后输出特征图并重整为若干个主胶囊,再经过动态路由得出耦合系数并完成从主胶囊层到数字胶囊层的转换,最后根据数字胶囊层中每个向量的长度得到各类地物数据的分类结果。

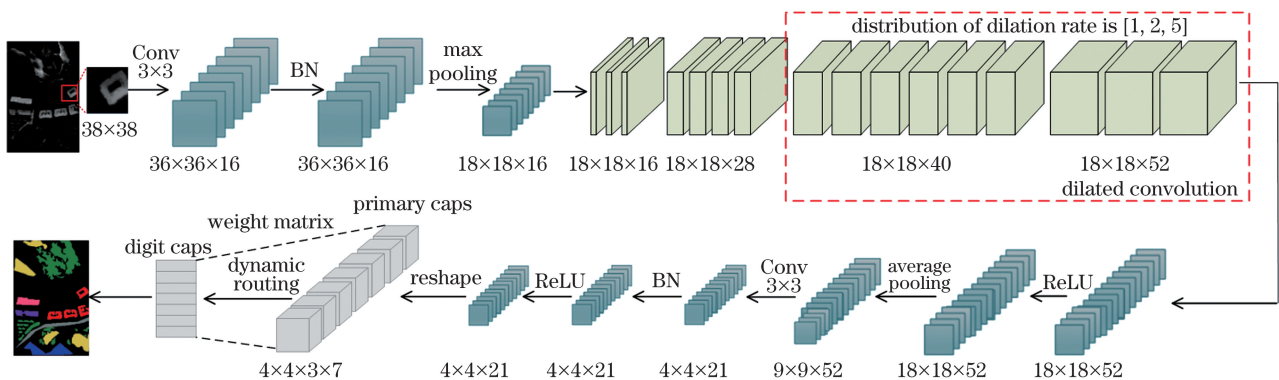


图 1 空洞卷积胶囊网络结构

Fig.1 Structure of dilated convolution capsule network

2.1 残差网络

相较于浅层神经网络,深度网络由于其高维的非线性操作,可以实现更抽象的图像特征提取与表示。但是当网络增加到一定深度时,靠前部分的隐藏层的权值更新变缓或者停滞,导致分类准确率达到饱和,如果继续加深网络,准确率反而会迅速下降。针对这一问题,He 等^[19]在 2015 年提出一种基于跳层连接的残差网络结构,该网络结构可以有效解决由于网络深度增加而导致的梯度下降问题。

残差模块的结构如图 2 所示,其中 x 为残差模块的输入, $F(x)$ 为残差映射, $H(x)$ 表示计算结果,残差模块将网络的恒等映射函数 $H(x)=x$ 训练转换为残差函数 $F(x)=H(x)-x$ 的训练,从而使得深度神经网络的权值收敛更有效。该方法既不会引入其他参数,也不会增加计算的复杂性。

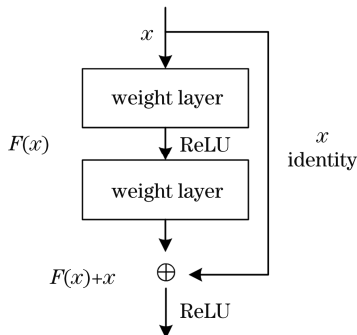


图 2 残差模块结构

Fig. 2 Residual block structure

残差结构的运算可表示为

$$F = W_2 \sigma(W_1 x), \quad (1)$$

$$y = F(x, W_i) + x, \quad (2)$$

$$y = F(x, W_i) + W_s x, \quad (3)$$

式中: F 为待训练的残差映射; σ 代表非线性函数 ReLU; W_1 和 W_2 为权重层的权重; x 和 y 分别表示残差模块的输入与输出; W_s 代表一个线性投影。输入 x 的维度和残差映射 F 的维度需保持一致,若输入 x 的维度和残差映射 F 的维度不一致时,需要在跳层连接上增加一个线性投影 W_s 来实现维度匹配。

2.2 空洞卷积设计及改进

空洞卷积^[20]是在普通卷积的基础上引入扩张率(r)的概念,具有扩展卷积核感受野的功能。通过设置不同的扩张率,可以捕获 LiDAR 图像的多尺度特征信息。空洞卷积可使卷积核的作用范围扩大。不同扩张率空洞卷积如图 3 所示。其中:图 3(a)对应扩张率为 1 的卷积,即传统 3×3 卷积核,感受野为 3×3 ;图 3(b)对应扩张率为 2 的卷积,虽然卷积核尺寸仍是 3×3 ,但感受野增大到 7×7 ;图 3(c)对应扩张率为 3 的卷积,其感受野为 13×13 。感受野的大小随着扩张率的增加而增大,但图 3(a)、图 3(b)、图 3(c) 中的参数量并没有增加,因此采用扩张卷积核代替传统卷积核可以获得更多的信息而不增加计算量。

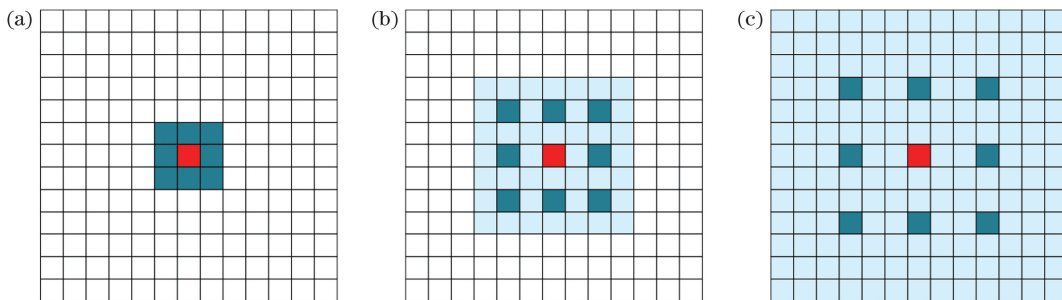


图 3 不同扩张率空洞卷积。(a) $r=1$;(b) $r=2$;(c) $r=3$

Fig. 3 Dilated convolution with different dilation rates. (a) $r=1$; (b) $r=2$; (c) $r=3$

在本模型中将空洞卷积引入 ResNet 中,目的是在不增加网络参数、不损失特征图分辨率的条件下,扩大特征图的感受野,捕获更丰富的上下文信息。空洞卷积核和感受野的计算式为

$$n = k + (k - 1) \times (r - 1), \quad (4)$$

$$l_m = l_{m-1} + \left[(f_m - 1) \times \prod_{i=1}^{m-1} S_i \right], \quad (5)$$

式中: k 表示原始卷积核大小; n 表示空洞卷积核大小; l_{m-1} 表示第 $(m-1)$ 层感受野大小; f_m 表示第 m

层卷积核的大小; l_m 表示经空洞卷积后的第 m 层感受野大小; S_i 表示第 i 层的步幅大小。

对于模型中扩张率的设计,如果仅多次叠加相同扩张率的卷积核,则会产生网格效应^[21]。网格效应会导致像素与像素之间不连续,存在一些空洞,遗漏一些像素,从而产生局部信息丢失,损坏了信息的连续性。这些问题都会降低分类模型的训练和测试精度。

针对网格效应,本文提出的 DCCN 模型中扩张率的设计借鉴了混合空洞卷积(HDC)^[22]思想,扩张

率设计须满足

$$r_{\max,i} = \max [r_{\max,i+1} - 2r_i, r_{\max,i+1} - 2(r_{\max,i+1} - r_i), r_i], \quad (6)$$

式中, r_i 是第 i 层的扩张率, $r_{\max,i}$ 是指 i 层的最大

扩张率。HDC 要求叠加卷积的扩张率不能是大于 1 的公约数。本文通过奇偶混合扩张率的方法对卷积核进行扩张, 将扩张率设置为 $[1, 2, 5]$ 循环结构, 可以覆盖图像上的每个像素点, 避免信息丢失, 如图 4 所示。

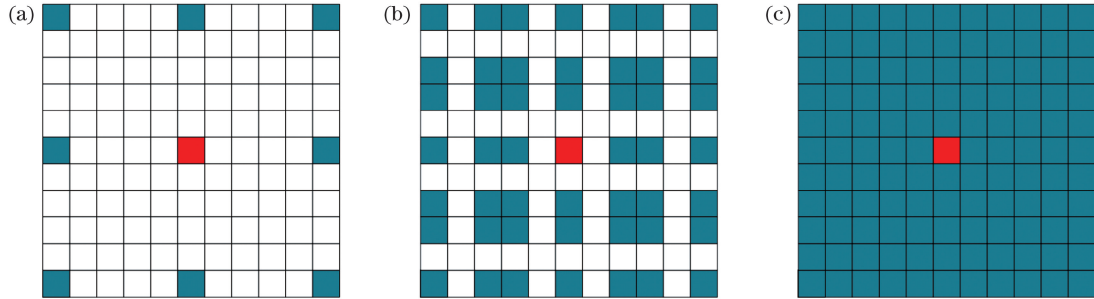


图 4 奇偶混合扩张率。(a) $r=5$; (b) $r=2$; (c) $r=1$

Fig. 4 Odd-even mixed dilation rates. (a) $r=5$; (b) $r=2$; (c) $r=1$

2.3 胶囊网络设计

与 CNN 使用神经元作为数据处理单元不同, 胶囊网络将识别同一物体的所有神经元组合在一起形成一个胶囊, 这些神经元包含特定图案中各种属性的所有信息, 可识别不同方向、不同位置的一类模式, 其输入和输出是包含每个特性的特征信息的向量。

胶囊网络主要由卷积层、主胶囊层和数字胶囊层组成, 如图 5 所示。卷积层旨在捕获输入数据的特征, 输出特征图。主胶囊层则对卷积层输出的特征图进行卷积处理, 并将生成的标量转换为向量, 至此, 胶囊网络则完成了标量到向量的转换, 即完成了

胶囊的构建。最后一层为数字胶囊层, 与全连接层类似, 其中 $\|L2\|$ 表示对数字胶囊层的向量求模, 利用输出向量模的大小来衡量某个类别的概率, 模最大的向量即为输出类别。胶囊网络采用 Squashing 非线性函数保证输出向量的长度在 $0 \sim 1$ 之间, 其表达式为

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \cdot \frac{s_j}{\|s_j\|}, \quad (7)$$

式中前一部分是输入向量 s 的缩放尺度, 后一部分是 s 的单位向量。该非线性函数既保留了输入向量的方向, 又将输入向量的模压缩到 $(0, 1)$ 之间。

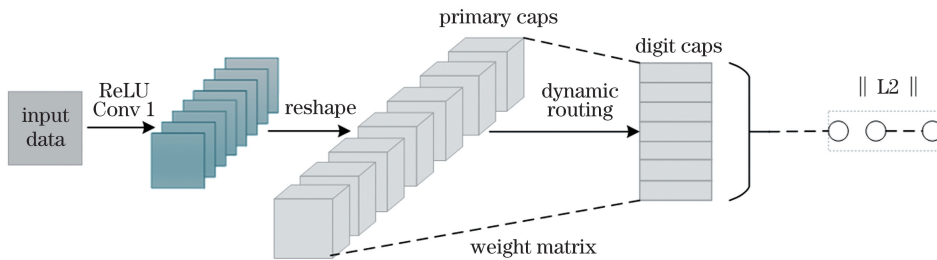


图 5 胶囊网络结构

Fig. 5 Capsule network structure

主胶囊层到数字胶囊层的数据传递是通过动态路由算法更新网络参数, 避免了由池化引入的损失。动态路由算法通过特定的路由协议在主胶囊层和数字胶囊层之间进行数据传播, 表达式为

$$s_j = \sum_i c_{ij} \hat{u}_{j|i}, \quad (8)$$

$$\hat{u}_{j|i} = W_{ij} u_i, \quad (9)$$

式中: s_j 为第 j 个胶囊的输入; $\hat{u}_{j|i}$ 为预测向量; c_{ij} 为耦合系数, W_{ij} 为权重矩阵, u_i 为上层胶囊网络

输出。

与全连接神经网络相似, 胶囊网络中第 j 个胶囊的输入 s_j 也为上层胶囊网络输出 u_i 的线性加权求和。在此基础上, 胶囊网络引入了一个耦合系数 c_{ij} 。权重矩阵是由误差反向传播进行更新, 耦合系数 c_{ij} 依赖动态路由算法进行更新, 更新表达式为

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}, b_{ij} + \hat{u}_{j|i} \cdot v_j \rightarrow b_{ij}. \quad (10)$$

利用预测向量 $\hat{u}_{j|i}$ 和输出向量 v_j 的内积来度量向

量间的一致性,并更新 b_{ij} ,利用 Softmax 更新耦合系数,进一步修正下一层胶囊的输入 s_j ,最后输出新的 v_j 。通过这种方式不断迭代更新一致性参数。胶囊层级间结构如图 6 所示。

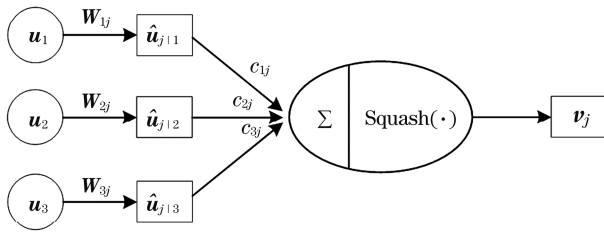


图 6 Capsule 层级间结构

Fig. 6 Capsule structure

图 1 下半部分为 DCNN 的胶囊网络部分,动态路由迭代次数设置为 3 时可以获得更好的分类结果,因此在本模型中,动态路由迭代次数设置为 3,主胶囊层卷积核大小设置为 3×3 ,通道数设置为 3,数字胶囊的数量设置为数据集中标记的类别数。本文采用的两个 LiDAR 数据集分别包含 7 个和 11 个地物类别,因此将数字胶囊的数量设置为 7 和 11。

3 实验结果及分析

3.1 实验数据集

为了验证本文所提模型对 LiDAR-DSM 数据分类的有效性,本研究使用的数据集是 2012 年 IEEE 国际遥感影像融合竞赛的 Bayview Park 和 Recology 公开数据集,二者均采集于美国旧金山城区^[23]。Bayview Park 数据集的尺寸为 $300 \text{ pixel} \times 200 \text{ pixel}$,空间分辨率为 1.8 m,包含 7 个有标记的地物类别。Recology 数据集的尺寸为 $200 \text{ pixel} \times 250 \text{ pixel}$,空间分辨率为 1.8 m,包含 11 个有标记的地物类别。真值由人工标记获得,该人工标记用作有监督分类的类别标签。图 7、图 8 分别显示了 Bayview Park 和 Recology 数据集的 DSM 灰度图

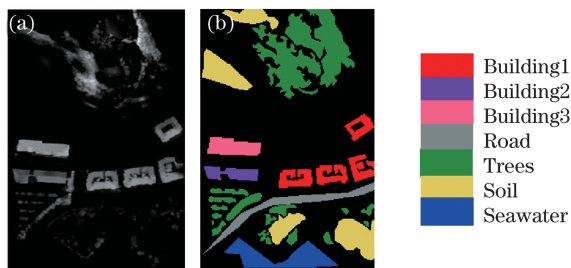


图 7 Bayview Park 数据集的 DSM 灰度图和真值图。

(a) DSM; (b) 真值图

Fig. 7 DSM and groundtruth map of Bayview Park dataset. (a) DSM; (b) groundtruth map

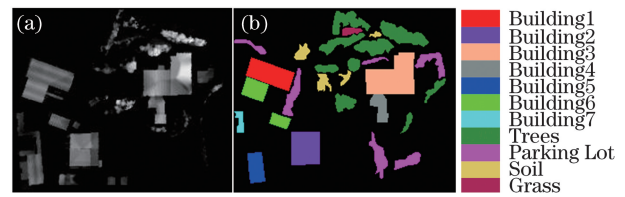


图 8 Recology 数据集的 DSM 灰度图和真值图。

(a) DSM; (b) 真值图

Fig. 8 DSM and groundtruth map of Recology dataset.

(a) DSM; (b) groundtruth map

和真值图,其中不同颜色代表不同地物类别。

3.2 实验平台参数设置

本文所有实验均采用以 TensorFlow 作为后端的 Keras 深度学习框架实现。实验环境为 Intel(R) Core(TM) i7-10750H CPU@2.60 GHz 处理器,16 GB 内存,Nvidia GeForce GTX 2060 显卡。为了尽可能减小实验的偶然误差,本文所有实验结果数据均为 10 次实验后的均值结果。在网络训练方面,所有实验都使用批处理方法。为了适应硬件资源,减少网络训练过程中每批次的计算量,输入数据的尺寸设置为 32×32 。在实验中发现,所有实验网络最多在训练 250 个 epoch 后即可达到稳定的收敛状态。为确保所有方法都能达到最佳分类效果,本文将训练 epoch 的最大数量设置为 300,并采用早停法来避免过拟合问题的发生。网络训练过程中,使用随机梯度下降(SGD)优化算法进行优化,动量设置为 0.9,学习率设置为 0.001。

实验从 LiDAR 原始数据中随机抽样以确保实验的可信度,抽取的样本总数为 5000。通过大量实验发现,当训练样本数大于 400 时,本研究中涉及的各种分类方法可以达到较好的分类效果。但当训练样本数达到 800 时,各分类方法的分类效果相对于训练样本数为 700 时提升不大,趋于稳定。为同时满足提高分类精度和节约计算资源的目的,本实验中训练集样本的数量分别设置为 400、500、600 和 700,其余样本作为测试样本。

本研究使用高斯核函数构造非线性的支持向量机(SVM)。为了保证 SVM 模型的分类效果,防止离群点过多造成分类精度下降,本研究采用的惩罚因子为 100,惩罚因子的作用是控制离群点引入的损失,惩罚因子越大,目标函数损失越大。对比实验中随机森林的子树数量,如果设置得太小,容易导致欠拟合,如果设置得过大,会增加训练和预测的时间,所以需要参数调优选择一个适中的数值。通过实验表明,随机森林子树数量为 30 时,可以得到较

好的分类结果。

3.3 分类评价指标

本文实验的分类结果评价指标采用了遥感数据分类问题中常用的总体分类精度(OA)、平均分类精度(AA)和 Kappa 系数(K)。OA 为所有被正确分类的样本数量与总体样本数量之比,体现了分类的总体情况。AA 为每类别的分类精度的均值,与 OA 不同,AA 能较好地衡量对小样本类别的分类效果。Kappa 系数既能衡量分类精度,又可以用于一致性检验,通常 Kappa 系数是[0, 1]范围内的一个数字,当 Kappa 系数在 0.81~1.00 之间时表示小样本类别几乎完全一致。

3.4 ResNet-34 模型设计

本研究将 ResNet-34 作为主要特征提取网络,其由 Conv1、Conv2_x、Conv3_x、Conv4_x、Conv5_x 五部分组成。为了提取到更密集的特征,避免图像信息丢失,ResNet-34 的 Conv4_x 和 Conv5_x 中的卷积替换为空洞卷积,既保持了特征图的空间分辨率又扩大了感受野范围。模型输入尺寸为 38×38,最终 Conv5_x 输出图像大小为 18×18,包含原始图像的大量特征,有利于获得较高的分类结果。由于本模型输入图像尺寸较小,将 Conv1 中卷积核的大小从 7 减小到 3,可使网络提取到更多细节信息。后四个部分分别由 3、4、6 三个恒等残差块组成。同时,将后四部分中恒等残差块的滤波器数量分别由 64、128、256、512 减少到 16、28、40 和 52,并且不使用输出分类层。表 1 为 ResNet-34 的参数设置。

表 1 ResNet-34 模型参数

Table 1 ResNet-34 model parameters

Layer name	Output size	ResNet-34 parameter setting
Conv 1	18×18	3×3, 16, stride 1
Conv2_x	18×18	3×3 max pooling, stride 2 (3×3, 16)×3
Conv3_x	18×18	(3×3, 28)×4
Conv4_x, distribution of dilation rate is [1,2,5]	18×18	(3×3, 40)×6
Conv5_x, distribution of dilation rate is [1,2,5]	18×18	(3×3, 52)×3
Output	9×9	average pooling, stride 2

3.5 扩张率的选取

在本文的实验中,分别选取扩张率分布为 [1,1,2]、[2,2,2]、[1,2,2]、[1,2,3]、[1,2,4]、[1,2,5]以及[1,2,6]共七种循环结构进行对比分析,空洞卷积层中不同扩张率分布对应的分类精度结果如图 9 所示。

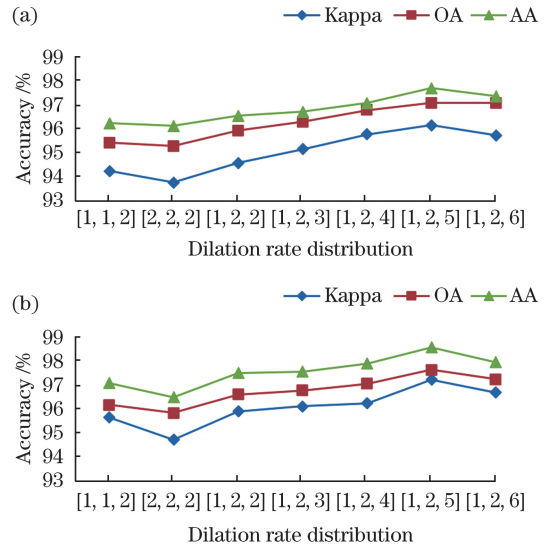


图 9 不同数据集的扩张率分布选取。(a)Bayview Park 数据集;(b)Recology 数据集

Fig. 9 Dilation rate distribution of different datasets.

(a) Bayview Park dataset; (b) Recology dataset

扩张率分布为[1,1,2]的感受野大小为 9×9,通过对比实验结果可以发现,单空洞卷积层与双普通卷积层组合的分类精度低于双空洞卷积层与单普通卷积层组合。在扩张率分布为[2,2,2]的条件下,虽然感受野范围增大到 13×13,但是分类精度却低于双普通卷积层组合的平均总体精度。这是因为纯空洞卷积层的叠加会导致较多的特征信息被遗漏,后五个实验为采用双空洞卷积层与单普通卷积层的组合,感受野分别为 11×11、13×13、15×15、17×17 以及 19×19。虽然[1,2,6]感受野范围最大,但是扩张率逐渐增大会导致输入采样的数据越来越稀疏,从而导致局部信息丢失,损坏信息的连续性。通过图 9 的实验结果可以发现,扩张率分布为[1,2,5]时,两个数据集能获得最优的分类结果。

3.6 不同模型分类结果分析

首先用训练样本数分别为 400、500、600、700 的训练样本进行实验,对比在不同方法下 OA, AA, Kappa 三个系数评价指标的值,由表 2 和表 3 可以发现,在训练样本数增加的过程中,各对比实验方法及本文所提算法的各项评价指标都在增大。

显然,与传统的 SVM 和随机森林等分类方法相比,深度学习方法的分类精度有了很大提高。由于采用了动态路由算法,与传统的 CNN 相比,CapsNet 的分类效果也得到了改善。由于 CapsNet 的卷积块中只有两个卷积层,对于 LiDAR 数据而言其捕获特征的能力受到限制,故本文所提模型在胶囊网络之前引入 ResNet 网络,ResNet 网络前三组卷积先对特征进行粗提取,后两组采用空洞卷积替代传统卷积捕获图像细节信息,弥补了胶囊网络捕获特征能力受限的不足。因此,本文所提模型拥有更强的特征表达能力和分类精度。

在所有实验中,当训练样本数为 700 时,本文所

提方法分类精度最高,在 Bayview Park 数据集中 OA 最高达到了 $97.07\% \pm 0.54\%$,与 SVM, Random Forest, CNN, CapsNet, ResNet, Dilated-ResNet, ResCapsNet 对比,分别提高了 20.32% , 6.51% , 6.35% , 6.34% , 2.48% , 1.23% , 0.65% 。在 Recology 数据集中 OA 最高达到了 $96.98\% \pm 0.76\%$,与 SVM, Random Forest, CNN, CapsNet, ResNet, Dilated-ResNet, ResCapsNet 对比,分别提高了 19.73% , 5.23% , 4.25% , 6.72% , 1.15% , 1.09% , 0.67% 。由此充分证实 DCCN 模型在 LiDAR 数据分类任务上的有效性。

表 2 Bayview Park 数据集的不同训练样本分类结果

Table 2 Classification results of different training samples on Bayview Park dataset

Training sample	Index	400	500	600	700
SVM	OA /%	72.30 ± 2.07	75.27 ± 1.36	75.36 ± 2.02	76.75 ± 0.76
	AA /%	76.98 ± 1.42	78.50 ± 2.10	78.86 ± 1.12	81.20 ± 2.27
	100K	65.22 ± 1.79	66.36 ± 0.95	67.65 ± 1.72	69.12 ± 2.23
Random Forest	OA /%	86.56 ± 0.75	87.34 ± 0.62	88.23 ± 0.34	90.56 ± 0.47
	AA /%	88.94 ± 1.52	89.12 ± 0.23	89.45 ± 0.38	90.14 ± 0.73
	100K	82.66 ± 0.24	83.61 ± 0.38	84.26 ± 0.39	86.73 ± 0.64
CNN	OA /%	87.23 ± 2.01	88.12 ± 0.96	88.52 ± 0.43	90.72 ± 1.69
	AA /%	88.70 ± 1.13	89.63 ± 2.71	89.96 ± 1.65	90.23 ± 0.68
	100K	83.26 ± 1.45	85.25 ± 1.46	86.23 ± 1.78	86.72 ± 2.34
CapsNet	OA /%	85.43 ± 1.12	87.25 ± 0.89	90.07 ± 1.08	90.73 ± 0.36
	AA /%	84.26 ± 1.78	88.26 ± 1.38	91.14 ± 1.35	91.86 ± 1.52
	100K	81.21 ± 0.81	83.19 ± 0.69	86.81 ± 1.45	86.92 ± 1.22
ResNet	OA /%	90.25 ± 1.73	92.16 ± 1.26	93.26 ± 1.21	94.59 ± 1.20
	AA /%	91.53 ± 1.38	93.23 ± 0.81	94.25 ± 1.06	95.86 ± 1.25
	100K	87.15 ± 1.49	89.46 ± 1.48	91.26 ± 1.77	93.49 ± 1.28
Dilated-ResNet	OA /%	91.32 ± 0.45	93.16 ± 0.75	93.89 ± 0.34	95.84 ± 1.25
	AA /%	92.67 ± 0.76	94.09 ± 1.06	95.27 ± 0.87	96.42 ± 1.08
	100K	88.46 ± 1.05	91.58 ± 0.57	92.65 ± 0.65	94.15 ± 1.34
ResCapsNet	OA /%	93.15 ± 0.52	94.79 ± 0.41	94.59 ± 0.73	96.42 ± 0.71
	AA /%	94.27 ± 0.43	95.42 ± 0.90	96.03 ± 0.75	97.01 ± 1.07
	100K	90.49 ± 1.00	92.48 ± 0.47	93.23 ± 0.75	94.99 ± 1.17
DCCN	OA /%	93.48 ± 0.39	94.51 ± 0.47	95.45 ± 0.60	97.07 ± 0.54
	AA /%	94.97 ± 0.44	95.39 ± 0.61	95.90 ± 0.74	97.70 ± 0.20
	100K	91.36 ± 0.57	92.79 ± 0.42	94.02 ± 0.72	96.14 ± 0.71

表 3 Recology 数据集的不同训练样本的分类结果
Table 3 Classification results of different training samples on Recology dataset

Training sample	Index	400	500	600	700
SVM	OA /%	72.68±1.89	76.83±0.26	76.94±2.25	77.25±0.86
	AA /%	77.20±1.22	78.69±1.87	78.73±1.08	81.28±2.33
	100K	67.22±1.79	68.42±0.98	68.82±1.59	69.78±2.24
Random Forest	OA /%	85.24±1.25	87.33±0.74	89.20±1.88	91.75±1.00
	AA /%	88.36±1.92	89.85±3.06	90.27±1.26	91.29±1.41
	100K	82.25±0.88	86.26±1.57	86.59±1.97	89.22±1.34
CNN	OA /%	86.26±1.48	88.24±0.98	90.52±0.68	92.73±1.86
	AA /%	89.16±2.84	90.15±0.29	90.67±1.24	92.44±2.34
	100K	83.18±1.59	86.78±0.67	86.83±0.88	90.15±2.17
CapsNet	OA /%	80.73±1.07	84.92±1.67	86.75±0.43	90.26±1.24
	AA /%	81.93±1.94	86.29±1.08	86.95±1.09	91.26±1.99
	100K	76.79±1.68	81.67±0.85	83.92±0.76	88.37±1.52
ResNet	OA /%	90.58±1.92	93.56±1.42	95.57±0.76	95.83±0.99
	AA /%	88.86±2.14	94.52±1.22	94.33±1.32	95.36±1.88
	100K	88.84±2.34	93.04±1.70	94.98±0.84	95.24±1.27
Dilated-ResNet	OA /%	92.07±0.98	93.75±0.34	94.87±0.54	95.89±0.76
	AA /%	93.76±0.67	94.88±0.96	95.98±1.25	96.34±0.38
	100K	90.67±0.34	92.39±1.07	94.77±1.15	95.09±0.47
ResCapsNet	OA /%	93.44±1.21	94.35±1.23	96.07±0.48	96.31±0.73
	AA /%	94.63±0.17	95.22±0.60	97.16±1.15	97.30±0.17
	100K	91.29±0.70	93.79±0.99	95.33±0.63	95.43±0.31
DCCN	OA /%	94.01±0.36	94.99±0.96	96.42±0.63	96.98±0.76
	AA /%	94.97±0.58	95.67±0.46	97.49±0.69	97.77±0.78
	100K	93.28±0.34	94.35±0.99	96.06±0.70	96.41±0.90

其次针对训练样本数为 700 时,两个数据集每一类的分类精度进行对比。通过表 4 和表 5 可以看出, Building 类的精度明显高于其他类别, 尤其高于

Soil, Seawater, Parking Lot, Grass, 这是因为 LiDAR 在获取信息时, 对高度较高的信息更加敏感。通过对比 Dilated-ResNet 与 DCCN 中每一类

表 4 训练样本为 700 时 Bayview Park 数据集的每个类别的分类结果

Table 4 Classification results of each class for 700 samples on Bayview Park dataset

unit: %

Classe	SVM	Random Forest	CNN	CapsNet	ResNet	Dilated-ResNet	ResCapsNet	DCCN
Building1	82.20±3.72	95.24±3.75	93.53±1.50	94.28±1.45	98.26±1.55	98.96±1.04	99.36±0.64	99.74±0.26
Building2	84.59±2.58	98.83±1.03	92.72±1.08	95.22±2.03	99.64±0.36	99.78±0.22	99.79±0.21	99.54±0.46
Building3	91.36±4.96	100	92.90±1.50	93.32±1.93	99.57±0.43	99.68±0.32	100	99.69±0.31
Road	81.60±4.43	82.59±6.35	91.36±1.58	94.90±1.20	96.38±2.65	97.09±2.83	98.12±1.22	98.82±1.18
Trees	83.78±1.69	90.43±1.20	86.79±1.82	92.82±1.79	97.65±0.82	98.06±0.76	98.67±0.88	98.39±0.36
Soil	62.23±2.27	86.78±0.52	85.59±1.71	83.56±0.80	86.95±1.62	87.47±1.52	89.32±2.06	90.98±1.39
Seawater	86.79±2.48	84.21±1.25	90.75±2.74	85.53±1.23	91.18±2.87	91.88±2.65	93.47±2.30	94.79±1.17

的分类结果,可以发现,对于 Recology 中 Parking Lot, Soil 和 Grass 类的分类精度提高显著,这充分证明胶囊网络对于空间特征信息的敏感性。由表 4 可知,在 7 个地物类别中, DCCN 在 Building1, Road, Soil 和 Seawater 四个类别上都取得了最高的分类精度。另外三类中 Building2 和

Building3 的分类精度分别为 99.54% 和 99.69%, 与最高值 100% 相当。Trees 的分类精度也达到 98.39%, 在以下 8 种分类方法中仅次于 ResCapsNet 方法, 排名第二。由表 5 可知, 在 11 个感兴趣的地物类别中, DCCN 获得了 6 个类别的最佳分类精度。

表 5 训练样本为 700 时 Recology 数据集的每个类别的分类结果

Table 5 Classification results of each class for 700 samples on Recology dataset unit: %

Classe	SVM	Random Forest	CNN	CapsNet	ResNet	Dilated-ResNet	ResCapsNet	DCCN
Building1	71.78±1.21	92.56±2.78	98.78±1.22	92.29±1.29	98.75±1.25	98.86±1.14	98.06±1.74	98.97±0.32
Building2	64.54±1.77	94.67±3.75	96.52±1.42	94.45±1.36	99.24±0.76	99.02±0.98	99.46±0.54	99.82±0.18
Building3	92.78±1.20	94.24±1.44	94.16±1.25	93.47±1.35	98.42±1.58	98.56±1.44	98.41±1.16	98.04±0.96
Building4	90.39±2.47	97.88±0.24	97.55±1.36	95.16±1.01	95.41±1.72	97.52±2.31	99.67±0.33	98.24±1.76
Building5	86.26±1.73	96.30±2.77	97.48±2.38	98.06±1.94	99.86±0.14	99.07±0.93	98.72±1.28	99.55±0.45
Building6	71.52±1.62	95.27±1.36	94.37±1.07	87.47±1.59	96.85±2.34	97.26±1.09	98.35±1.65	97.06±2.83
Building7	88.37±2.74	97.26±2.74	97.48±1.95	95.86±2.15	92.43±2.53	95.97±1.75	98.79±1.21	99.51±0.49
Trees	86.88±1.24	95.67±0.21	95.64±1.24	90.14±0.35	97.48±1.85	96.47±1.66	95.54±1.31	97.32±1.55
Parking Lot	62.77±1.92	76.21±0.18	77.86±1.87	83.39±0.14	89.45±1.75	89.29±1.06	89.28±1.27	91.31±0.58
Soil	81.83±3.22	73.16±0.32	73.29±1.72	75.41±1.45	88.69±2.47	90.66±2.54	95.68±2.43	96.92±3.08
Grass	97.77±1.37	98.26±1.25	97.26±1.58	98.26±1.42	92.54±2.46	93.48±2.49	95.52±2.61	99.61±0.39

模型的训练时间和测试时间是深度学习分类模型的重要评估指标,表 6 给出了 5 种网络模型的计算时间消耗情况。可以发现,单独使用胶囊网络进行 LiDAR 数据分类时,总体分类精度仅为 90% 左右,但是胶囊网络作为一种轻量型网络,在两个数据集的训练时间仅为 99.58 s 和 94.27 s,节省了大量的计算资源。本文保留了胶囊网络的优势,为增强轻量化胶囊网络的特征捕获能力,将融合了空洞卷积的残差网络提取特征作为胶囊网络的输入。虽然训练时间增加到了 518.65 s 和 586.55 s,但因其能大幅提高分类精度,时间的微弱劣势是可接受的。

最后,图 10、图 11 分别显示了两个 LiDAR 数据 DSM 真值图以及各分类方法分类结果的假彩色图。假彩色图作为一种主观评价指标,可以更加直观地展示分类效果。从图中分析可知,深度学习方法相比传统方法 SVM、随机森林而言,分类效果有很大的提升,错误分类的面积大大减少。本文提出

表 6 训练样本为 700 时 Bayview Park 数据集和 Recology 数据集计算时间对比

Table 6 Comparison of calculation time for 700 samples on Recology dataset and Bayview Park dataset

Dataset	Network	Train time /s	Test time /s	OA /%
Bayview Park	ResNet	125.15	1.82	94.51
	Dilated-ResNet	195.34	2.76	95.67
	CapsNet	99.58	2.18	90.73
	ResCapsNet	343.85	3.00	96.42
	DCCN	518.65	3.75	97.07
Recology	ResNet	196.34	2.67	95.83
	Dilated-ResNet	257.98	2.98	96.36
	CapsNet	94.27	1.39	90.26
	ResCapsNet	428.69	3.34	96.31
	DCCN	586.55	3.96	96.98

的 DCCN 分类边界更加平滑,已标记像元与背景像元的错分现象更少,更贴近真实地物分布。

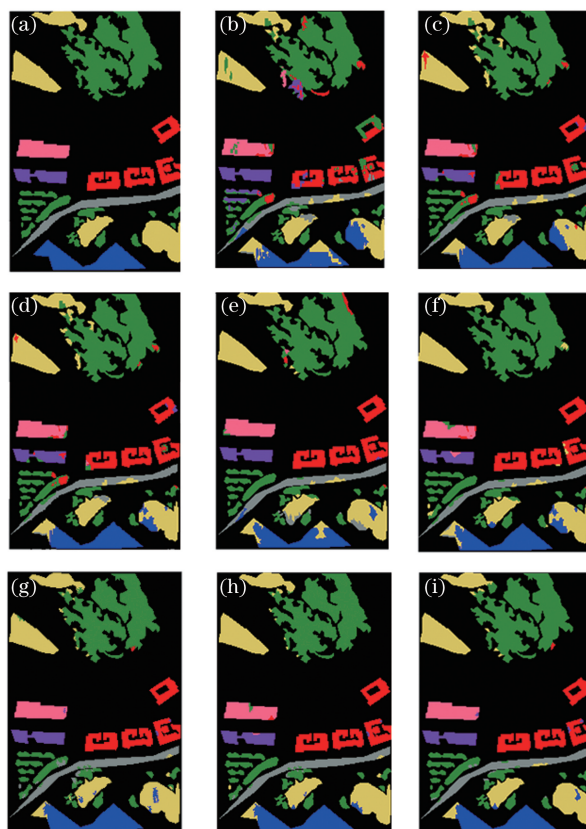


图 10 Bayview Park 数据集的分类结果。(a)真值图；(b)SVM；(c)随机森林；(d)CNN；(e)CapsNet；(f)ResNet；
(g)Dilated-ResNet；(h)ResCapsNet；(i)DCCN

Fig. 10 Classification results of Bayview Park dataset. (a) Groundtruth map; (b) SVM; (c) Random Forest; (d) CNN;
(e) CapsNet; (f) ResNet; (g) Dilated-ResNet; (h) ResCapsNet; (i) DCCN

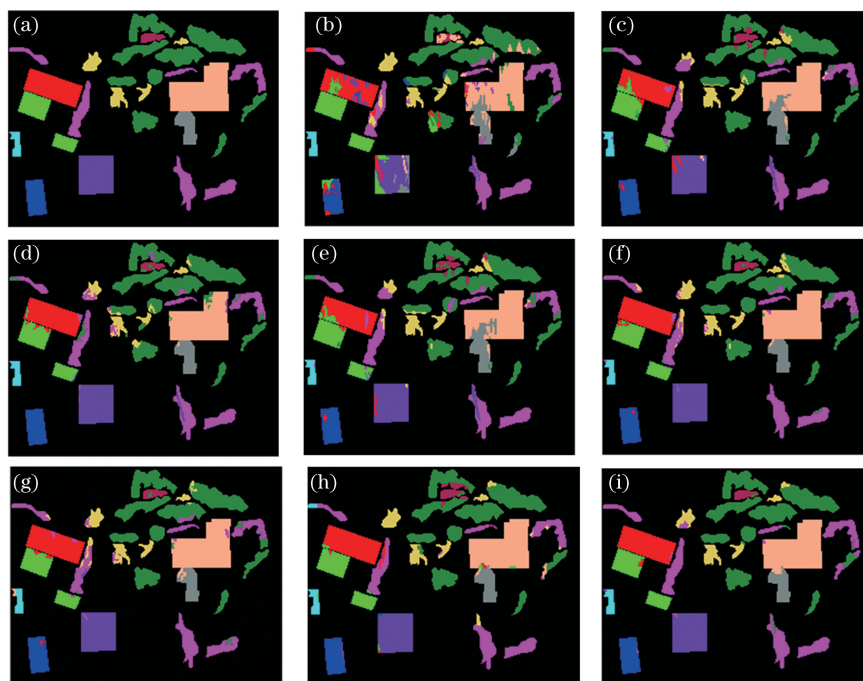


图 11 Recology 数据集的分类结果。(a)真值图；(b)SVM；(c)随机森林；(d)CNN；(e)CapsNet；(f)ResNet；
(g)Dilated-ResNet；(h)ResCapsNet；(i)DCCN

Fig. 11 Classification results of Recology dataset. (a) Groundtruth map; (b) SVM; (c) Random Forest; (d) CNN;
(e) CapsNet; (f) ResNet; (g) Dilated-ResNet; (h) ResCapsNet; (i) DCCN

4 结 论

研究了 LiDAR 数据的地物分类算法,提出了空洞卷积胶囊网络模型,该模型以深度卷积神经网络为基础,将空洞卷积模块引入 ResNet 网络中,针对空洞卷积的网格效应采用奇偶混合扩张率设计,并利用胶囊网络进一步提取目标精细的空间特征信息。本研究在两个有代表性的 LiDAR 数据集上进行实验,并与 7 种典型的分类算法进行对比,结果表明本文设计的网络分类精度较好,当训练样本数为 700 时,在 Bayview Park 和 Recology 数据集上 OA 分别达到了 97.07% 和 96.98%。

参 考 文 献

- [1] Hu H Y, Hui Z Y, Li N, et al. Airborne LiDAR point cloud classification based on multiple-entity eigenvector fusion [J]. Chinese Journal of Lasers, 2020, 47(8): 0810002.
胡海瑛, 惠振阳, 李娜, 等. 基于多基元特征向量融合的机载 LiDAR 点云分类 [J]. 中国激光, 2020, 47(8): 0810002.
- [2] Kong D M, Xu L J, Li X L, et al. A real-time method for DSM generation from airborne LiDAR data [C] // 2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), May 6-9, 2013, Minneapolis, MN, USA. New York: IEEE Press, 2013: 377-380.
- [3] Naidoo L, Cho M A, Mathieu R, et al. Classification of savanna tree species, in the greater Kruger national park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2012, 69: 167-179.
- [4] Khodadadzadeh M, Li J, Prasad S, et al. Fusion of hyperspectral and LiDAR remote sensing data using multiple feature learning [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8(6): 2971-2983.
- [5] Ghamisi P, Höfle B. LiDAR data classification using extinction profiles and a composite kernel support vector machine [J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(5): 659-663.
- [6] Ghamisi P, Höfle B, Zhu X X, et al. Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(6): 3011-3024.
- [7] Xia J S, Yokoya N, Iwasaki A, et al. Fusion of hyperspectral and LiDAR data with a novel ensemble classifier [J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(6): 957-961.
- [8] Wang A L, He X, Ghamisi P, et al. LiDAR data classification using morphological profiles and convolutional neural networks [J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5): 774-778.
- [9] Wang A L, Wang M H, Jiang K Y, et al. A novel LiDAR data classification algorithm combined densenet with STN [C] // 2019 IEEE International Geoscience and Remote Sensing Symposium, July 28-August 2, 2019, Yokohama, Japan. New York: IEEE Press, 2019: 2483-2486.
- [10] Ge C R, Du Q, Li W, et al. Hyperspectral and LiDAR data classification using kernel collaborative representation based residual fusion [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(6): 1963-1973.
- [11] Wang P, Liu R, Xin X J, et al. Scene classification of optical remote sensing images based on residual network [J]. Laser & Optoelectronics Progress, 2021, 58(2): 021001.
汪鹏, 刘瑞, 辛雪静, 等. 基于残差网络的光学遥感图像场景分类算法 [J]. 激光与光电子学进展, 2021, 58(2): 021001.
- [12] Zhang X D, Wang T J, Yang Y, et al. Classification of small-sized sample hyperspectral images based on multi-scale residual network [J]. Laser & Optoelectronics Progress, 2020, 57(16): 162801.
张祥东, 王腾军, 杨耘, 等. 基于多尺度残差网络的小样本高光谱图像分类 [J]. 激光与光电子学进展, 2020, 57(16): 162801.
- [13] Zhu T Y, Huang L F, Dong F, et al. Infrared-remote-sensing ship detection based on lightweight residual network [J]. Acta Optica Sinica, 2020, 40(1): 0111018.
朱天佑, 黄凌锋, 董峰, 等. 基于轻量级残差网络的红外遥感船只检测 [J]. 光学学报, 2020, 40(1): 0111018.
- [14] Sabour S, Frosst N, Hinton G E, et al. Dynamic routing between capsules [C] // Advances in Neural Information Processing Systems (NIPS), December 4-9, 2017, Long Beach, CA, USA. New York: Curran Associates, Inc., 2017: 3856-3866.
- [15] Raza A, Huo H, Sirajuddin S, et al. Diverse capsules network combining multiconvolutional layers for remote sensing image scene classification [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 5297-5313.
- [16] Wang X, Tan K, Du Q, et al. Caps-TripleGAN:

- GAN-assisted CapsNet for hyperspectral image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(9): 7232-7245.
- [17] Chang S W, Liu J. Multi-lane capsule network for classifying images with complex background [J]. *IEEE Access*, 2020, 8: 79876-79886.
- [18] Paoletti M E, Haut J M, Beltran R F, et al. Capsule networks for hyperspectral image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(4): 2145-2160.
- [19] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [20] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions [C] // International Conference on Learning Representations (ICLR), May 2-4, 2016, Caribe Hilton, San Juan, Puerto Rico. New York: IEEE, 2016: 1-13.
- [21] Yu F, Koltun V, Funkhouser T, et al. Dilated residual networks [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 636-644.
- [22] Wang P Q, Chen P F, Yuan Y, et al. Understanding convolution for semantic segmentation [C] // 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), March 12-15, 2018, Lake Tahoe, NV, USA. New York: IEEE Press, 2018: 1451-1460.
- [23] Debes C, Merentitis A, Heremans R, et al. Hyperspectral and LiDAR data fusion: outcome of the 2013 GRSS data fusion contest [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(6): 2405-2418.

LiDAR Data Classification Based on Dilated Convolution Capsule Network

Wang Aili¹, Zhang Yuxiao¹, Wu Haibin^{1*}, Jiang Kaiyuan¹, Iwahori Yuji²

¹ Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application, Harbin University of Science and Technology, Harbin, Heilongjiang 150080, China;

² Department of Computer Science, Chubu University, Aichi 487-8501, Japan

Abstract

Objective LiDAR, as an essential technical means of obtaining physical attributes of ground objects, is widely used in remote sensing image classification research. Recently, deep learning algorithms that use multilayer nonlinear transformation methods to automatically extract features have become mainstream in the field of image processing. Traditional convolutional neural networks (CNNs) use neurons as the unit for processing data. Each neuron can only recognize one pattern and is not sensitive to the direction and position information of objects in the image. To overcome this shortcoming of CNN, CapsNet is used in the field of image classification. However, the capsule network is a lightweight neural network, and its network structure indicates that the feature capture ability is still lacking compared with the deep network. Aiming at the above two problems, this study proposes a LiDAR data classification algorithm that combines capsule network and dilated convolution. The features extracted by the residual network combined with the dilated convolution are used as the input of the capsule network to enhance the feature capture capability of the lightweight capsule network. Dilated convolution can expand the receptive field of the convolutional layer without increasing network parameters. The capsule network uses capsules as a data processing unit in its internal structure and adopts a dynamic routing algorithm for data transfer between capsule layers. It has better feature expression capabilities than CNN. Thus, the proposed method can improve the classification performance of LiDAR data.

Methods The LiDAR data classification algorithm that combines the capsule network and dilated convolution is proposed. First, input the LiDAR data into the dilated convolution capsule network (DCCN) model, which uses the first three layers of the residual network to perform rough feature extraction on the LiDAR data, and the latter two layers use dilated convolution instead of traditional convolution to obtain detailed features of LiDAR images of different scales. Among them, the design of the dilation rate adopts the odd-even mixed method to avoid the gridding

effect. Since the capsule network converts each neuron of the traditional neural network from a scalar to a vector, the modeling of the internal structure of the image presents a certain semantic relationship. At the same time, the excellent characteristics of the spatial information of the image are retained. The feature extracted from the residual network combined with dilated convolution is used as the input of the capsule network, the convolutional layer of the capsule network is used to capture data features, and batch normalization is used in the convolutional layer to solve the problem of poor initialization and ensure gradient propagates to each layer. Then, output the feature map and reform it into several main capsules. The coupling coefficient is obtained through dynamic routing, and the conversion from the primary capsule layer to the digital capsule layer is completed. Finally, the classification results of various features are obtained according to the length of each vector in the digital capsule layer.

Results and Discussions The datasets used in this article are the Bayview Park and Recology public datasets of the 2012 IEEE International Remote Sensing Image Fusion Competition, both of which were collected in San Francisco, USA. The evaluation index of the classification result of the experiment adopts the overall classification accuracy (OA), average classification accuracy (AA), and Kappa coefficient commonly used in remote sensing image classification. To verify the advantages of the capsule network in time, the training and testing time of the model are used as essential evaluation indicators of the model. First, the proposed DCCN model is compared with the other seven classification algorithms when the training samples are 400, 500, 600, and 700. The results showed that the evaluation indicators of the proposed algorithm are higher than that of the other classification algorithms (Table 2, Table 3). When the number of training samples is 700, the classification accuracy of each category of the two datasets is compared. By comparing the classification results of each category of Dilated-ResNet and DCCN, the classification accuracy of the ground object category with a lower height is significantly improved, which fully proves the sensitivity of the capsule network to spatial feature information (Table 4, Table 5). When the capsule network is used to classify LiDAR data alone, although the overall classification accuracy is low, the training time for the two datasets is only 99.58 and 94.27 s, which saves computational cost. The proposed model uses the residual network extracted features combined with dilated convolution as the input of the capsule network. Although the training time is increased, the slight disadvantage of time is acceptable because it can significantly improve the classification accuracy (Table 6). Finally, the results of all classification algorithms are shown in pseudo-color maps. It can be seen that the DCCN classification boundary is smoother, the marked and background pixels are less misclassified, and they are closer to the real surface object distribution (Fig. 10, Fig. 11).

Conclusions This study proposes the DCCN model used for the ground object classification of LiDAR data, which introduces dilated convolution into ResNet. The dilation rate distribution of dilated convolution is designed using the odd-even mixed method to suppress the gridding effect. The capsule network is used to extract more detailed spatial feature information. This study conducts experiments on two representative LiDAR datasets and compares them with seven typical classification algorithms. The results showed that the designed network classification accuracy performs better. When the number of training samples is 700, the OA on the Bayview Park and Recology datasets reach 97.07% and 96.98%, respectively.

Key words remote sensing; LiDAR; deep learning; residual network; dilated convolution; capsule network

OCIS codes 280.3640; 100.4996; 100.2000