

基于深度可分离卷积的轻量化残差网络高光谱影像分类

程榕杰¹, 杨耘^{1,2*}, 李龙威¹, 王燕婷¹, 王家禹¹

¹长安大学地质工程与测绘学院, 陕西 西安 710054;

²自然资源部矿山地质灾害成灾机理与防控重点实验室, 陕西 西安 710054

摘要 针对传统深度残差网络在对高光谱图像进行特征提取和分类过程中因参数量大导致的训练时间长的问题, 提出一种基于深度可分离卷积的轻量化残差网络模型(DSC-Res14)。该模型首先基于一层三维卷积层对经主成分分析方法降维后的高光谱影像进行光谱特征和空间特征初提取; 其次, 引入3个不同尺度的三维深度可分离卷积残差层对影像的深层语义特征进行提取, 减少了网络训练参数量, 增强了网络对高维、多尺度空间特征信息的表达能力。经在公开的 Indian Pines 和 Pavia University 标准数据集上进行实验验证, 结果表明: 所提模型在两个数据集上的分类精度分别为 99.46% 和 99.65%; 对比同类模型, 所提模型在保证较高分类精度的同时, 参数量和计算量小, 训练时间短, 并具有良好的鲁棒性。

关键词 遥感; 高光谱影像; 深度可分离卷积; 空-谱特征; 轻量化

中图分类号 TP751.1

文献标志码 A

DOI: 10.3788/AOS221848

1 引言

土地覆盖/土地利用遥感分类是对遥感影像中的每个像元赋予唯一的类别标识^[1]。利用高光谱遥感技术可获得高光谱分辨率的遥感影像, 从而获得地物丰富的光谱特征和空间特征, 因此高光谱遥感技术对土地覆盖/土地利用等地物精细分类识别具有重要意义^[2]。近年来, 随着高光谱遥感信息提取技术在精细农业、自然资源调查与监测等领域的应用日益广泛, 高光谱遥感影像分类等解译任务面临着新的挑战^[3]。

传统基于像素的高光谱影像分类算法(如波谱角分类法等), 受制于高光谱影像高维特征和非线性的冗余信息, 需要进行人工降维、光谱特征选择及分类等操作, 此类算法精度低且无法实现端对端的影像分类^[4-5]。早期的机器学习算法如反向传播(BP)神经网络和支持向量机(SVM)虽然能够达到较高的精度, 但存在收敛速度慢、特征提取能力有限和自动化程度低等问题^[6-7]。近年来深度学习算法的理论发展迅速, 如卷积神经网络(CNN)^[8]、深度置信网络(DBN)^[9]和堆叠自编码器(SAE)^[10]等典型的深度学习网络模型为高光谱影像特征提取及分类提供了新的解决方案。

早期的 CNN 仅能通过一维卷积提取影像光谱特

征进行高光谱影像像素分类。为进一步挖掘影像的光谱和空间特征, 有学者提出二维及三维 CNN 网络。其中, 二维 CNN 与主成分分析(PCA)法等降维算法联合, 有效地减少了高光谱影像在光谱维度的冗余信息, 取得了良好的分类效果^[11]。文献[12-13]提出用三维卷积代替二维卷积直接对高维影像进行空-谱特征提取, 提高了 CNN 网络分类的准确性。此外, 针对上述单一卷积核网络结构在分类精度方面的不足, 有学者将一维 CNN、二维 CNN 和三维 CNN 子网络进行组合, 并提出一些组合 CNN 网络, 如组合一维 CNN 与二维 CNN 的双通道卷积神经网络(DC-CNN)^[14]、组合二维 CNN 与三维 CNN 的混合光谱卷积神经网络(HybridSN)^[15], 这些组合网络改善了单一卷积结构的分类效果。

自 2015 年 ResNet^[16]在 ImageNet 图像识别竞赛中展现出优异性能起, 残差结构被引入到遥感影像分类问题中, 以解决深层网络结构带来的模型退化问题, 如: Lee 等^[17]和刘冰等^[18]成功搭建了引入残差结构的 Res-2D-CNN 和 Res-3D-CNN 模型; 邓子青等^[19]和陈文豪等^[20]将残差结构分别与多尺度卷积核和注意力机制结合, 进一步提高了残差网络对于高光谱影像的分类精度。但是, 深层的网络结构会导致模型训练的时

收稿日期: 2022-10-19; 修回日期: 2022-11-29; 录用日期: 2022-12-12; 网络首发日期: 2023-01-06

基金项目: 陕西省自然科学基金基础研究计划(2022JM-163)、中央高校基本科研业务费(300102269205)

通信作者: *yangyunbox@chd.edu.cn

间成本大幅增加,如何在 不影响深度学习模型性能的情况下进行模型压缩与加速成为研究热点,常见的轻量化研究方法有参数剪枝、模型量化和知识蒸馏等^[21]。目前,国内外研究人员提出了 SqueezeNet^[22]、ShuffleNet 以及 MobileNet^[23] 等轻量化网络模型。其中,MobileNet 将深度可分离卷积(DSC)代替传统卷积核,节省了大量参数,为深度模型轻量化提供了新思路。刘启超等^[24] 构建了由光谱维一维卷积和空间维二维卷积的分离卷积结构构成隐层单元的深层稠密网络,在小样本高光谱图像分类问题上取得了优异成果。

针对前人研究中仍存在的“深度残差网络模型参数量大,导致训练耗时长、计算代价高”的问题,本文设计并构建了一种结合多尺度残差结构和深度可分离卷积的轻量化残差网络(DSC-Res14),在保证网络具有较高分类精度的同时,提高了网络训练的效率,以期为进一步推动高光谱遥感影像的智能化信息提取提供新的思路。

2 相关理论基础

2.1 深度可分离卷积

随着深度学习网络模型结构的加深,网络的参数量不断增大,计算代价逐渐增加,导致所需的存储空间变大、网络训练时间变长。因此,这类网络对服务器算力资源的要求较高。为此,在保证网络准确率的同时

使模型结构更加轻量化、训练速度更快是深度网络走向更广泛应用的关键。

标准的三维 CNN 的卷积核沿着长、宽和通道高度 3 个方向移动,神经网络第 i 层第 j 个特征图在 (x, y, z) 处的点值 v_{ij}^{xyz} 计算公式为

$$v_{ij}^{xyz} = f \left(\sum_{k=1}^m \sum_{l=0}^{L_i-1} \sum_{w=0}^{W_i-1} \sum_{h=0}^{H_i-1} \omega_{ijk}^{lwh} v_{(i-1)k}^{(x+l)(y+w)(z+h)} + b_{ij} \right), \quad (1)$$

式中: L_i 、 W_i 和 H_i 为三维卷积核的长、宽、高; m 为 $i-1$ 层特征图数量; $v_{(i-1)k}^{(x+l)(y+w)(z+h)}$ 为 $i-1$ 层中第 k 个特征图 $((x+l), (y+w), (z+h))$ 位置的值; ω_{ijk}^{lwh} 为与 $i-1$ 层第 k 个特征图连接的卷积核; b_{ij} 为偏置; $f(\cdot)$ 为激活函数。

标准的三维卷积运算涉及的参数众多,计算代价高,而在 MobileNet^[23] 中提出的深度可分离卷积为深度模型轻量化提供了新的方法。不同于标准的三维卷积原理,深度可分离卷积是将三维卷积运算分离成深度卷积和点卷积两种运算。第一步是深度卷积,即在不改变通道深度的情况下利用 $w \times w$ 的二维卷积核对输入图像进行分组卷积,分别提取各个通道的空间特征;第二步是点卷积,利用 $1 \times 1 \times h$ 的核函数逐点提取光谱特征,增加特征图数量。深度可分离卷积本质上是对三维卷积核的分解,以深度可分离卷积代替传统 $3 \times 3 \times 3$ 卷积核,其框架如图 1 所示。

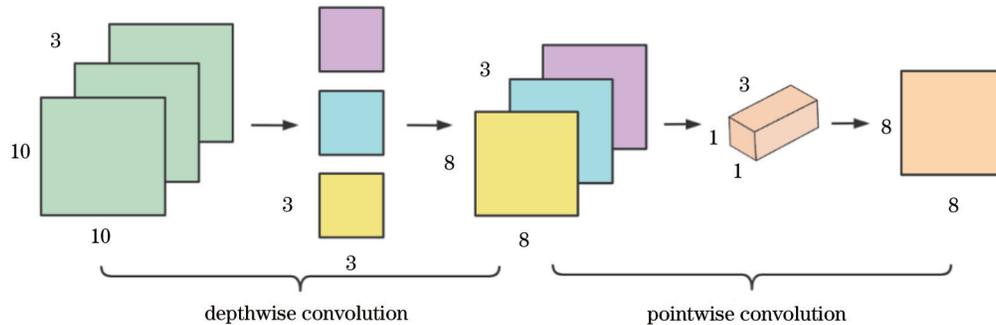


图 1 深度可分离卷积结构
Fig. 1 Structure of depthwise separable convolution

假定三维卷积核的大小为 $3 \times 3 \times 3$, 输入特征图通道数为 3, 输出特征图数量为 N , 则深度可分离卷积的深度卷积和逐点卷积的卷积核大小分别为 3×3 和 $1 \times 1 \times 3$, 其中三维卷积的参数量为 $3 \times 3 \times 3 \times 3 \times N$, 而深度可分离卷积参数量为 $3 \times 3 \times 3 \times 3 + 1 \times 1 \times 3 \times 3 \times N$, 则两种卷积的参数量之比为

$$\frac{3 \times 3 \times 3 \times 3 + 1 \times 1 \times 3 \times 3 \times N}{3 \times 3 \times 3 \times 3 \times N} = \frac{1}{N} + \frac{1}{9}. \quad (2)$$

式(2)表明,当输出特征图数量 N 增大时,深度可分离卷积与标准三维卷积相比,有更少的参数量和计算量,并且内核移动更少。

综上所述,深度可分离卷积所涉及的参数量少,并且保持了传统三维卷积核同时提取影像的空-谱联合

特征的优势。

2.2 残差结构

网络模型的深度与结构是高光谱影像分类精度的重要影响因素,随着深度学习网络层数的不断增加,网络所提取的特征信息将更加丰富。理论上,网络层数越多,学习效果越好。然而,已有研究^[16]表明,随着卷积层和池化层的叠加,网络模型出现了梯度消失和梯度爆炸的现象,预测效果越来越差。为了解决由网络加深导致的模型退化问题,He 等^[16] 在传统卷积层的基础上引入了恒等链接,其残差结构如图 2 所示。

当网络加深到某一深度时状态已经达到最优,继续加深网络结构不仅会增加计算量,网络的退化现象也不可避免。从图 2 可以看出,残差结构可以通过恒

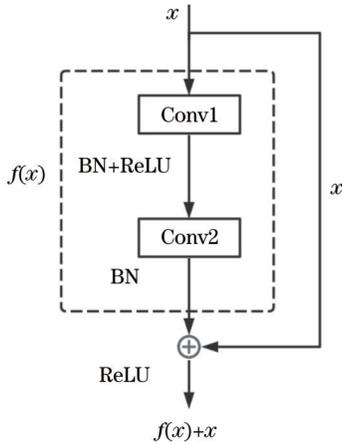


图 2 文献[16]模型的残差结构

Fig. 2 Residual structure in the model of Ref. [16]

等链接在神经网络卷积层求解函数 $f(x)=0$, 直接跳过深层卷积层进行恒等映射, 最终的输出等于上层结构的输入 x , 从而保证了网络的最优状态。此外, 恒等链接也使得网络较一般 CNN 收敛更快, 因此在图像识

别与分类中能取得更好的效果。

3 基于深度可分离卷积的轻量化残差网络

3.1 深度可分离卷积残差模块设计

传统的残差网络如 ResNet18、ResNet34 和 ResNet50 等, 都是由二维 CNN 构成的, 并且处理的影像通常为三通道光学影像, 不能很好地应用于光谱维度高的高光谱影像分类任务。因此, 所提模型使用三维深度可分离卷积代替二维卷积核, 每个残差结构包含了两个深度可分离卷积层。传统残差网络中, 每个残差层由多个残差结构组成, 除第一层外每层的首个残差结构为虚线链接结构, 会对输入特征图的尺寸进行缩减并增加特征图数量。深度可分离卷积虚线链接结构如图 3 所示: 残差层内的其余残差结构为恒等链接结构, 不改变特征图的尺寸与维度。为了保证特征图尺寸在恒等链接中的不变性, 卷积过程中对边缘进行填充, 得到的深度可分离卷积恒等链接结构如图 4 所示。

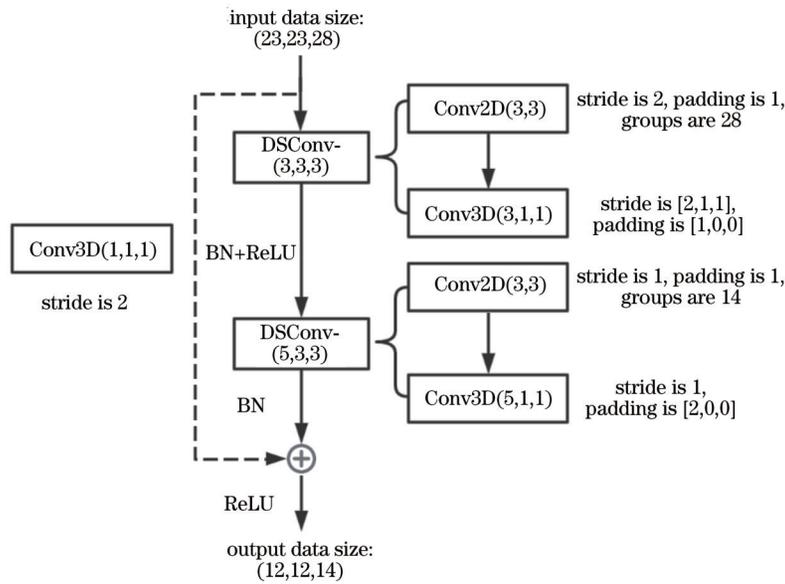


图 3 深度可分离卷积虚线链接结构

Fig. 3 Depthwise separable convolution dotted shortcuts structure

3.2 基于深度可分离卷积的轻量化残差网络

在 3.1 节设计的残差结构基础上, 提出了基于深度可分离卷积的轻量化深度残差网络模型 (DSC-Res14), 其结构如图 5 所示。该模型将经过主成分分析方法降维的图像块作为输入数据, 先通过一个卷积层进行特征初提取, 再经过 3 个残差层进一步提取光谱特征和空间特征, 最后经过一个全链接层将一维特征向量输入到分类器进行高光谱影像逐像素分类。

如图 5 所示, 每个残差层包含两个残差结构, 不同于传统残差网络使用单一尺度卷积核, 为了提取多尺度特征以提高分类精度, 各层残差单元卷积核大小依

次为 $7 \times 3 \times 3$ 、 $5 \times 3 \times 3$ 和 $3 \times 3 \times 3$ 。为了减少网络训练参数, 各层残差单元中使用深度可分离卷积代替三维卷积核, 先由 3×3 的二维分组卷积提取各个通道的空间特征, 再由一维点卷积提取光谱特征。在每个卷积层后加入批量归一化 (BN) 层, 使输入特征保持相同分布, 并使用 ReLU 激活函数, 以加快网络收敛速度, 缓解梯度消失问题。

传统的深度学习模型常采用 Softmax 作为网络最后的分类层, 为了进一步优化分类效果, 本文以 LogSoftmax 函数为分类函数, 其定义式为

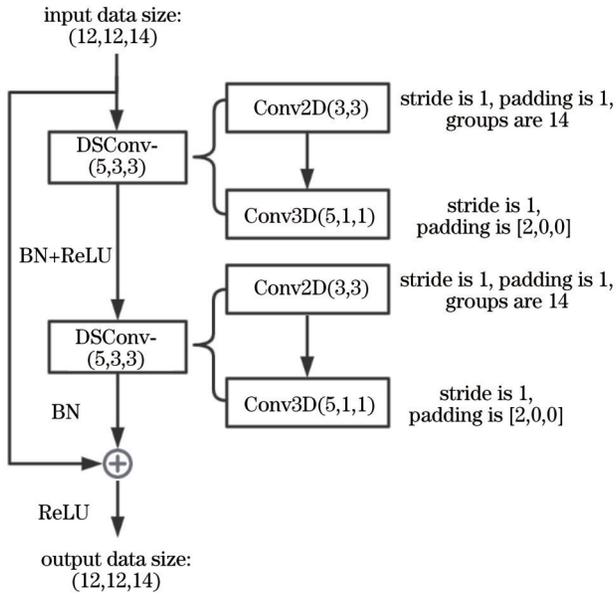


图 4 深度可分离卷积恒等链接结构

Fig. 4 Depthwise separable convolution identity shortcuts structure

$$\text{LogSoftmax}(z_i) = \log \left[\frac{\exp(z_i - c)}{\sum_{j=1}^n \exp(z_j - c)} \right], \quad (3)$$

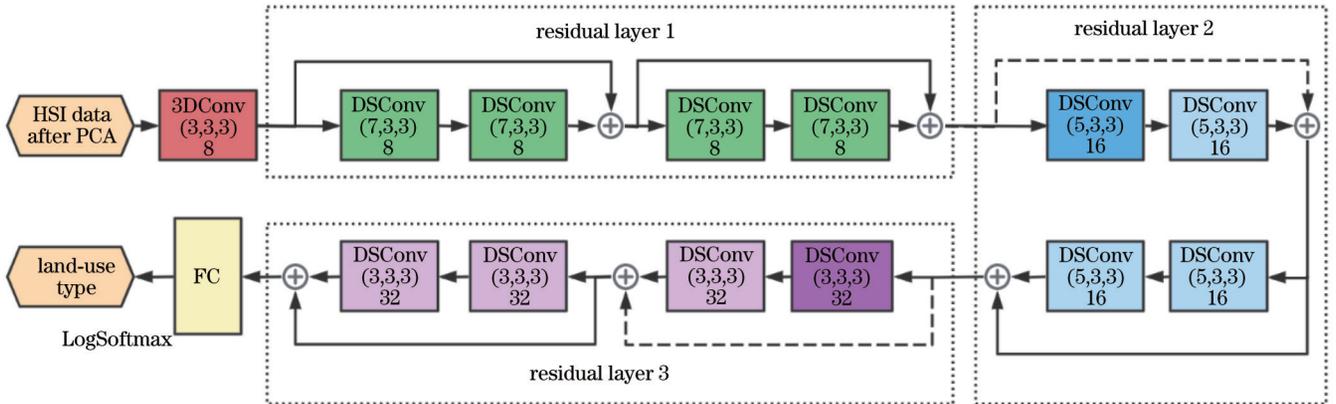


图 5 所提网络结构

Fig. 5 Proposed network structure

表 1 Indian Pines 数据集地物编号及样本数量

Table 1 Category No. and sample size in Indian Pines dataset

No.	Category	Sample size	No.	Category	Sample size
1	Alfalfa	46	9	Oats	20
2	Corn-notill	1428	10	Soybean-notill	972
3	Corn-mintill	830	11	Soybean-mintill	2455
4	Corn	237	12	Soybean-clean	593
5	Grass-pasture	483	13	Wheat	205
6	Grass-trees	730	14	Woods	1265
7	Grass-pasture-mowed	28	15	Buildings-grass-trees-drives	386
8	Hay-windrowed	478	16	stone-steel-towers	93

式中： z_i 为第 i 个类别的输入变量； z_j 为第 j 个类别的输入变量； n 为类别数量； c 为类别输入变量中的最大值。在取对数前后 Softmax 都呈单调变化，但它们对损失函数相对值的影响会发生变化，使用 LogSoftmax 函数将更严厉地惩罚似然空间中的错误，并能有效防止数字下溢，得到更加稳定的结果。

4 实验结果与分析

4.1 实验数据

为了验证所提模型的有效性，选取 Indian Pines (IP) 和 Pavia University (PU) 这两个公开的高光谱数据集进行分类实验，两个数据集的地物编号及样本数量如表 1 和表 2 所示。

Indian Pines 高光谱数据的尺寸为 $145 \text{ pixel} \times 145 \text{ pixel}$ ，空间分辨率为 20 m ，包括农作物、树林等 16 类植物共 10249 个样本；光谱范围为 $0.4 \sim 2.5 \mu\text{m}$ ，经校正后剩余 200 个波段以供研究。Pavia University 高光谱数据集是由德国的机载反射光学光谱成像仪在意大利帕维亚大学所在区域获取到的。该数据集的尺寸为 $610 \text{ pixel} \times 340 \text{ pixel}$ ，空间分辨率为 1.3 m ，光谱范围为 $0.43 \sim 0.86 \mu\text{m}$ ，包含 103 个波段，共有 42776 个地物像素点，包含 9 类地物。

表 2 Pavia University 数据集的地物编号及样本数量
Table 2 Category No. and sample size in Pavia University dataset

No.	Category	Sample size	No.	Category	Sample size
1	Asphalt	6631	6	Bare soil	5029
2	Meadows	18649	7	Bitumen	1330
3	Gravel	2099	8	Self-blocking bricks	3682
4	Trees	3064	9	Shadows	947
5	Painted metal sheets	1345			

4.2 实验环境及参数分析

本实验是基于 Pytorch 深度学习框架、Python 3.7 语言环境,在 GPU 为 NVIDIA GeForce RTX 3050、CPU 为 Intel i5-12400F 的硬件环境下完成。

输入网络的样本波段是经主成分分析后选取的前 30 个波段,Indian Pines 数据集的训练集、验证集和测试集的样本数量比例为 3:1:6,空间邻域大小为 25 pixel×25 pixel,batchsize 选为 32。Pavia University 数据集的训练集、验证集和测试集的样本数量比例为 1:1:8,空间邻域大小为 25 pixel×25 pixel,batchsize 为 32。该模型采用 Adam 优化器优化网络参数,学习率设为 0.001,激活函数统一使用 ReLU 函数以提高计算效率并加快函数收敛速度,损失函数选择交叉熵损失函数,最后由 LogSoftmax 函数进行分类。在实验过程中,最大迭代次数为 100 次,选取在验证集上精度最高的模型作为最优模型。

在精度评定方面,采用总体精度(OA)、平均精度(AA)和 kappa 系数作为分类精度评价指标。为了消除偶然误差对结果的影响,对于不同参数模型均进行连续 5 次重复训练和测试,对实验结果取平均值作为最终的结果。

4.2.1 不同权重衰减参数对分类结果的影响

深层的网络模型和少量的、分布不均衡的训练样本容易使网络出现过拟合现象,最直观的体现就是网络模型在训练集上表现优异,但对验证集和测试集的分类精度差,使模型的泛化能力弱。为缓解分类中的过拟合情况,引入 L2 正则化技术对特征权重进行衰减,以降低特征权值偏大对模型的影响。不同权重衰减参数对应的分类精度如图 6 所示。

从图 6 可以看到,L2 正则化对于两个数据集的优化效果差异较大,当权重衰减参数为 0.001 时,所提模型在两个数据集上的分类效果最好。Indian Pines 数据集各个类别的训练样本数量不均衡,个别类别的训练样本数量仅为个位数,这就造成模型对于小样本类别提取的特征图像数量相对较少,但预测时更加偏向样本数量大的类别,从而导致错分。在引入 L2 正则化后,分类精度明显提升。Pavia University 数据集的样本类别较少,样本数据充足,L2 正则化仍可小幅提升分类精度。

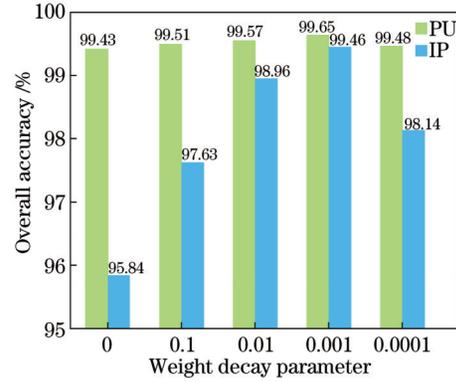


图 6 不同权重衰减参数的分类精度

Fig. 6 Classification accuracy of attenuation parameters with different weights

4.2.2 batchsize 对分类结果的影响

batchsize 表示单次传递给程序用于训练的数据数量。较大的 batchsize 可以提高内存利用率以及大矩阵乘法的并行化效率、增强模型的收敛性并缩短每个 epoch 的迭代时间。若参数设置过大,会占用较大的内存容量,并易使下降方向难以变化而陷入局部最优。为了选取合适的 batchsize,设置了 32、64、128、256 和 512 5 个数值,通过对比其精度来选取合适的参数。

由图 7 可知:当 batchsize 设置为 32 时,模型在两个数据集上精度最高;随着 batchsize 的增大,分类精度不断降低,当 batchsize 大于 128 时,分类精度迅速降低;当 batchsize 设置为 512 时,模型在 Pavia University 数据集上的精度仅为 88.34%。

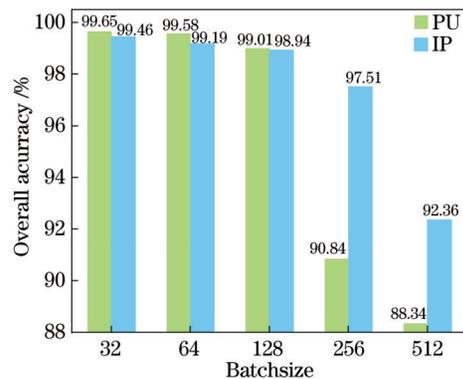


图 7 不同 batchsize 的分类精度

Fig. 7 Classification accuracy of different batchsize

4.3 多尺度残差单元的影响

所提模型 DSC-Res14 相较于传统深度残差模型的单一卷积结构,使用了 3 个尺度的卷积核残差单元,目的是通过提取多尺度特征来增强高维特征的信息表达,从而提高模型分类精度。为了证明多尺度卷积核

残差单元的有效性,分别设置了 DSC($3 \times 3 \times 3$)、DSC($3 \times 3 \times 5$)和 DSC($3 \times 3 \times 7$)的单一尺度深度可分离卷积核残差结构,并将它们与所提模型进行对比,网络结构和各项参数设置均相同,分类精度如图 8 所示。

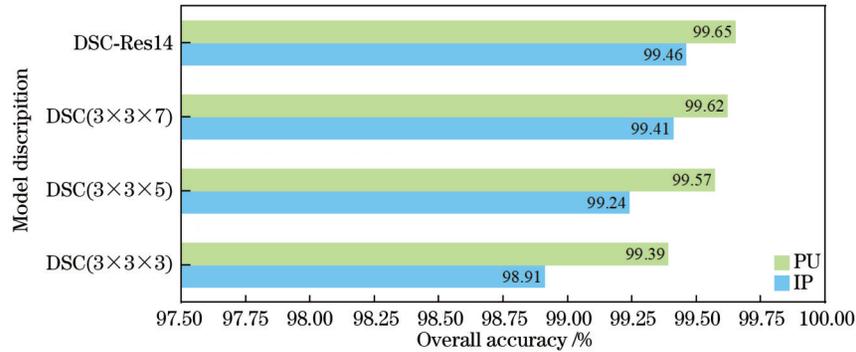


图 8 不同卷积核大小下的所提模型分类精度

Fig. 8 Classification accuracy of proposed model with different convolution kernel sizes

从图 8 可以看出,随着卷积核大小的不断增加,所提模型在两个数据集上分类效果不断提高。大尺度卷积核带来的精度提升在 Indian Pines 数据集上较为明显,其对于 Pavia University 数据集也有小幅提升。实验结果表明所使用的多尺度卷积核残差结构的分类效果优于单一尺度卷积核残差结构。

4.4 算法对比分析与验证

为了验证所提模型 DSC-Res14 的分类精度及速度,将其与 3DCNN-DSC^[25]、2D-Res-CNN^[17]、3D-Res-CNN^[18]和应用传统三维卷积核但与所提模型网络结构相同的 Res14 进行对比分析,每个网络均按照最优参数配置,样本比例相同,除 3D-Res-CNN 和 3DCNN-DSC 迭代 200 次外,其余网络均迭代 100 次,并保存指定迭代次数中验证集精度最高的模型结果。各个网络的卷积层参数量(Conv parameter)、总参数量(parameter number)、浮点计算次数(FLOPs)以及在两个数据集上训练时间如表 3 所示。

为了减少偶然误差的影响,对每个模型进行连续

表 3 各个网络的卷积层参数量、总参数量、FLOPs 和训练时间对比

Model	3DCNN-DSC	2D-Res-CNN	3D-Res-CNN	Res14	DSC-Res14
Conv parameter	200228	611840	111424	152232	20482
Parameter number	200228	611840	136016	281272	149522
FLOPs / 10^6	206	2	291	347	50
Time (IP) /s	300.30	21.32	1165.10	1170.32	394.71
Time (PU) /s	1913.00	25.97	997.12	1830.24	641.63

10 次实验并对各个类别的精度、总体精度、平均精度和 kappa 系数取平均值作为最终结果,各个模型在 Indian Pines 和 Pavia University 数据集的分类精度如表 4 和表 5 所示。

上述两个数据集的分类结果如图 9 和图 10 所示。对比 3DCNN-DSC、2D-Res-CNN 和 3D-Res-CNN 模

表 4 Indian Pines 测试集上不同算法的分类精度对比

Table 4 Comparison of classification accuracy of different algorithms on the Indian Pines test set unit: %

No.	3DCNN-DSC	2D-Res-CNN	3D-Res-CNN	Res14	DSC-Res14
1	92.31	69.81	61.61	100.00	96.92
2	91.22	95.18	97.52	99.08	99.21
3	97.41	90.58	96.49	99.56	99.36
4	97.59	89.75	98.62	99.72	99.45
5	98.41	98.85	98.58	99.53	99.23
6	98.95	99.29	97.50	99.19	99.28
7	100.00	67.78	63.36	100.00	100.00
8	95.96	96.15	99.31	100.00	100.00
9	80.00	69.17	58.33	96.03	94.67
10	98.90	94.98	97.05	99.97	99.71
11	97.86	91.67	92.37	99.33	99.50
12	89.62	96.92	91.25	99.71	99.47
13	98.80	99.10	98.16	100.00	100.00
14	98.82	97.81	98.75	99.97	99.75
15	99.33	93.74	98.99	99.91	99.49
16	89.74	94.31	53.54	98.56	98.41
OA	96.46	84.41	96.09	99.51	99.46
AA	95.31	90.32	87.59	99.41	99.03
Kappa	95.96	93.61	95.53	99.10	99.38

表 5 Pavia University 测试集上不同算法的分类精度对比
Table 5 Comparison of classification accuracy of different algorithms on the Pavia University test set unit: %

No.	3DCNN-DSC	2D-Res-CNN	3D-Res-CNN	Res14	DSC-Res14
1	95.52	93.08	98.17	99.71	99.77
2	99.92	97.75	99.16	99.93	99.83
3	100.00	87.51	99.55	99.62	99.54
4	99.75	99.19	98.13	98.37	98.92
5	99.82	99.95	99.83	100.00	99.47
6	92.49	95.14	99.98	99.98	99.98
7	97.86	93.23	98.25	99.57	99.47
8	94.73	83.66	96.60	98.38	99.31
9	99.73	97.80	99.05	99.18	98.68
OA	97.65	94.90	98.79	99.61	99.65
AA	97.75	94.15	98.75	99.42	99.44
Kappa	96.90	93.23	98.39	99.49	99.53

型,可以看到:1)2D-Res-CNN相较于其他网络,分类精度最低,其原因是使用二维卷积结构不能有效提取光谱特征用于分类,因此在高光谱分类中表现差。2)3DCNN-DSC和3D-Res-CNN模型在Indian Pines数据集上,前者对于每个类别的分类准确性差异小,其平均精度显著高于后者,而在总体精度和kappa系数的差异较小;在Pavia University数据集上两者差异较为明显,3D-Res-CNN的3项精度评价指标均显著高

于3DCNN-DSC,其原因是该数据集样本数量大、类别少,深度残差结构可以有效对类别进行区分。

在分类精度上,应用传统三维卷积的Res14在两个数据集上精度都达到99.5%以上,验证了所提模型结构的合理性。在Indian Pines数据集上,引入深度可分离卷积的DSC-Res14模型较Res14在样本数量较少的类别Alfalfa和Oats上分类精度下降明显,下降幅度分别达3.08%和1.36%,总体精度小幅(0.05%)下降,但对比前3个模型仍有明显优势。在Pavia University数据集上,DSC-Res14在5个模型中表现最佳,其各项精度均比Res14高,其中总体精度高出0.04%,平均精度高出0.02%,kappa系数上升了0.04%,具有良好的鲁棒性。较为均衡且充足的样本使得DSC-Res14不仅没有因参数的减少出现精度下滑,反而较传统三维卷积残差网络的分类效果有小幅提升。

将各个模型的训练时间、参数量以及FLOPs进行对比,结果显示:2D-Res-CNN模型由于结构简单,参数量和FLOPs最小,可以快速地完成模型训练,但是这也导致模型没有充分提取地物特征,分类精度不佳;3D-Res-CNN模型在训练过程中所需的最大迭代次数为200次,FLOPs较大,训练时间长,总体精度在Pavia University数据集上达到98.79%;所提模型DSC-Res14无论是在训练速度上还是分类精度上均具有最佳表现。对比同样引入深度可分离卷积的3DCNN-DSC,所提模型的参数量更少,深层残差结构

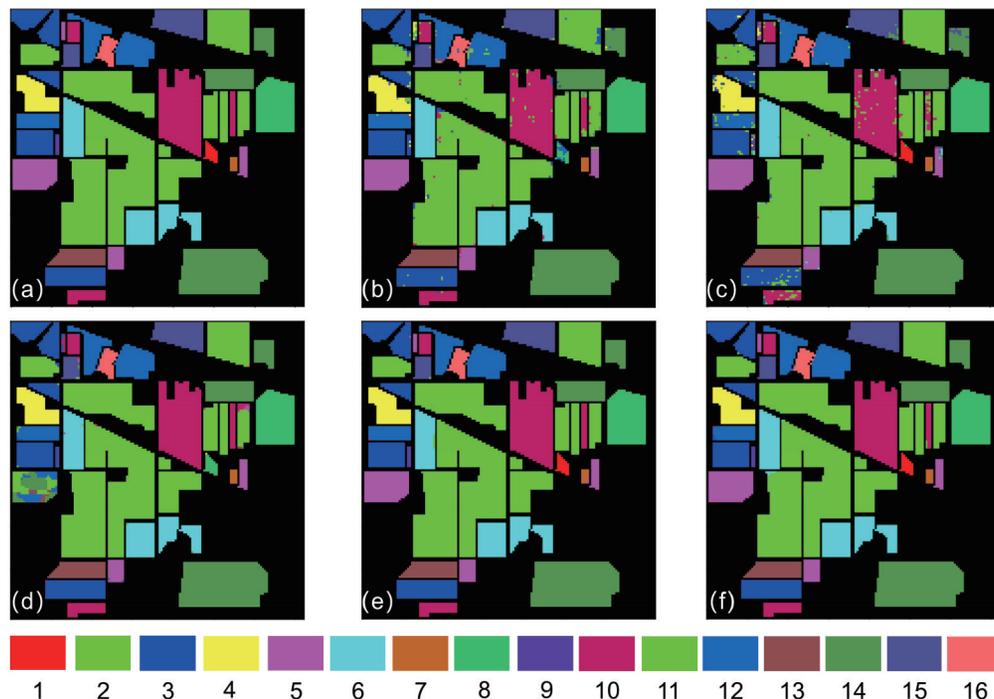


图9 不同模型在数据集Indian Pines的分类结果。(a)地面实况;(b)3DCNN-DSC;(c)2D-Res-CNN;(d)3D-Res-CNN;(e)Res14;(f)DSC-Res14

Fig. 9 Classification results of different models on the Indian Pines dataset. (a) Ground truth; (b) 3DCNN-DSC; (c) 2D-Res-CNN; (d) 3D-Res-CNN; (e) Res14; (f) DSC-Res14

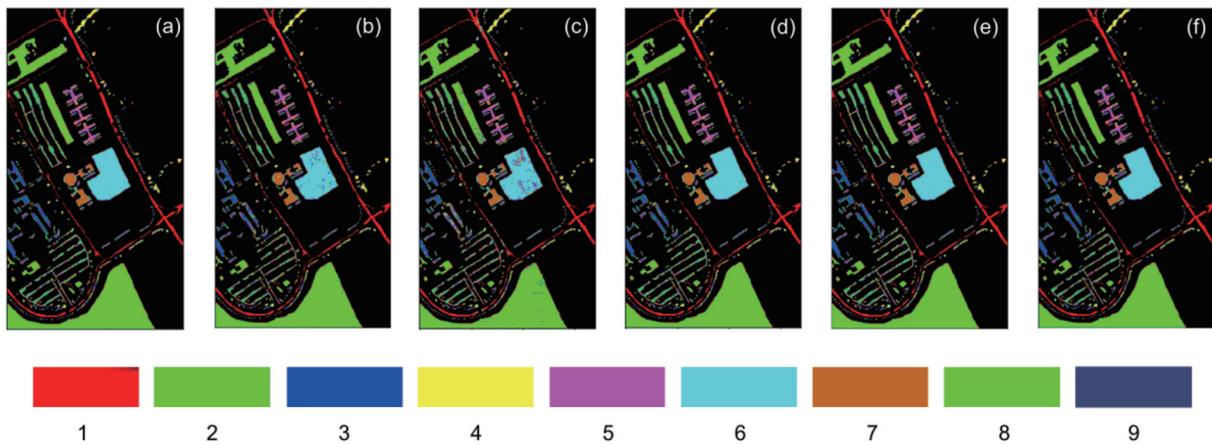


图 10 不同模型在数据集 Pavia University 的分类结果。(a)地面实况;(b) 3DCNN-DSC;(c) 2D-Res-CNN;(d) 3D-Res-CNN;(e) Res14;(f) DSC-Res14

Fig. 10 Classification results of different models on the Pavia University dataset. (a) Ground truth; (b) 3DCNN-DSC; (c) 2D-Res-CNN; (d) 3D-Res-CNN; (e) Res14; (f) DSC-Res14

也使得分类精度更高。在 Indian Pines 数据集上, grass-pasture-mowed、oats 和 stone-steel-towers 3 个类别训练样本呈小样本特性,其他模型对这 3 类样本的分类精度均较差,而所提模型对各个类别的分类精度更加均衡,平均精度达到 99.03%,有效克服了样本概率分布不均的问题。

上述对比分析证明了所提模型结构的合理性和优越性。更重要的是,所提模型在保证高分类精度的同时,其卷积层参数数量与 FLOPs 仅为 Res14 的 1/7,训练时间约为其 1/3,证明了所提模型 DSC-Res14 是一种高效、轻量化的深度残差网络。

5 结 论

提出一种基于深度可分离卷积和多尺度的深度残差网络模型(DSC-Res14),该模型有效地改善了高光谱遥感影像分类任务中需要使用深层网络结构来保证精度所带来的模型参数量大、训练时间长的情况。先对输入数据进行特征初提取,再将其输入到 3 个用深度可分离卷积代替传统三维卷积的多尺度特征提取残差层,最后经一层全连接层和 LogSoftmax 分类函数完成分类任务。经过对 Indian Pines 和 Pavia University 两个数据集的分类效果评价,以及与 3DCNN-DSC、2D-Res-CNN 和 3D-Res-CNN 模型的对比,所提模型的总分类精度较其余 3 个模型均有所提高,分别达到了 99.46% 和 99.65%,在各项指标中均为最佳。所提模型在保证高的总分类精度的情况下,大大减少了需训练的参数数量,训练时间明显缩短,性能优于同类网络。

对比传统三维卷积残差网络,当数据集各个类别样本数量分布差异大、出现小样本的情况时,引入深度可分离卷积会导致模型精度小幅下降。对于如何优化小样本条件下的分类效果,解决深度可分离卷积在减

少参数数量的同时导致模型分类精度降低的问题,还需进一步研究与改进。

参 考 文 献

- [1] 杜培军,夏俊士,薛朝辉,等. 高光谱遥感影像分类研究进展[J]. 遥感学报, 2016, 20(2): 236-256.
Du P J, Xia J S, Xue Z H, et al. Review of hyperspectral remote sensing image classification[J]. Journal of Remote Sensing, 2016, 20(2): 236-256.
- [2] 童庆禧,张兵,张立福. 中国高光谱遥感的前沿进展[J]. 遥感学报, 2016, 20(5): 689-707.
Tong Q X, Zhang B, Zhang L F. Current progress of hyperspectral remote sensing in China[J]. Journal of Remote Sensing, 2016, 20(5): 689-707.
- [3] 胡杰,张莹,谢仕义. 国产遥感影像分类技术应用研究进展综述[J]. 计算机工程与应用, 2021, 57(3): 1-13.
Hu J, Zhang Y, Xie S Y. Summary of research progress on application of domestic remote sensing image classification technology[J]. Computer Engineering and Applications, 2021, 57(3): 1-13.
- [4] 王斌,范冬林. 深度学习在遥感影像分类与识别中的研究进展综述[J]. 测绘通报, 2019(2): 99-102, 136.
Wang B, Fan D L. Research progress of deep learning in classification and recognition of remote sensing images[J]. Bulletin of Surveying and Mapping, 2019(2): 99-102, 136.
- [5] 杨耘,李龙威,高思岩,等. 基于 YOLOv3 网络训练优化的高分辨率遥感影像目标检测[J]. 激光与光电子学进展, 2021, 58(16): 1601002.
Yang Y, Li L W, Gao S Y, et al. Objects detection from high-resolution remote sensing imagery using training-optimized YOLOv3 network[J]. Laser & Optoelectronics Progress, 2021, 58(16): 1601002.
- [6] 付伟锋,邹维宝. 深度学习在遥感影像分类中的研究进展[J]. 计算机应用研究, 2018, 35(12): 3521-3525.
Fu W F, Zou W B. Review of remote sensing image classification based on deep learning[J]. Application Research of Computers, 2018, 35(12): 3521-3525.
- [7] 王振武,孙佳骏,于忠义,等. 基于支持向量机的遥感图像分类研究综述[J]. 计算机科学, 2016, 43(9): 11-17, 31.
Wang Z W, Sun J J, Yu Z Y, et al. Review of remote sensing image classification based on support vector machine[J]. Computer Science, 2016, 43(9): 11-17, 31.
- [8] 张号逵,李映,姜晔楠. 深度学习在高光谱图像分类领域的研

- 究现状与展望[J]. 自动化学报, 2018, 44(6): 961-977.
- Zhang H K, Li Y, Jiang Y N. Deep learning for hyperspectral imagery classification: the state of the art and prospects[J]. Acta Automatica Sinica, 2018, 44(6): 961-977.
- [9] 刘大伟, 韩玲, 韩晓勇. 基于深度学习的高分辨率遥感影像分类研究[J]. 光学学报, 2016, 36(4): 0428001.
- Liu D W, Han L, Han X Y. High spatial resolution remote sensing image classification based on deep learning[J]. Acta Optica Sinica, 2016, 36(4): 0428001.
- [10] Chen Y S, Lin Z H, Zhao X, et al. Deep learning-based classification of hyperspectral data[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(6): 2094-2107.
- [11] Makantasis K, Karantzalos K, Doulamis A, et al. Deep supervised learning for hyperspectral data classification through convolutional neural networks[C]//2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), July 26-31, 2015, Milan, Italy. New York: IEEE Press, 2015: 4959-4962.
- [12] Li Y, Zhang H K, Shen Q. Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network[J]. Remote Sensing, 2017, 9(1): 67.
- [13] 李竺强, 朱瑞飞, 高放, 等. 三维卷积神经网络模型联合条件随机场优化的 hyperspectral 遥感影像分类[J]. 光学学报, 2018, 38(8): 0828001.
- Li Z Q, Zhu R F, Gao F, et al. Hyperspectral remote sensing image classification based on three-dimensional convolution neural network combined with conditional random field optimization[J]. Acta Optica Sinica, 2018, 38(8): 0828001.
- [14] Zhang H K, Li Y, Zhang Y Z, et al. Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network[J]. Remote Sensing Letters, 2017, 8(5): 438-447.
- [15] Roy S K, Krishna G, Dubey S R, et al. HybridSN: exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(2): 277-281.
- [16] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [17] Lee H, Kwon H. Going deeper with contextual CNN for hyperspectral image classification[J]. IEEE Transactions on Image Processing, 2017, 26(10): 4843-4855.
- [18] 刘冰, 余旭初, 张鹏强, 等. 联合空-谱信息的高光谱影像深度三维卷积网络分类[J]. 测绘学报, 2019, 48(1): 53-63.
- Liu B, Yu X C, Zhang P Q, et al. Deep 3D convolutional network combined with spatial-spectral features for hyperspectral image classification[J]. Acta Geodaetica et Cartographica Sinica, 2019, 48(1): 53-63.
- [19] 邓子青, 王阳, 张兵, 等. 多尺度特征融合残差网络的高光谱地物分类[J]. 激光与光电子学进展, 2022, 59(18): 1810014.
- Deng Z Q, Wang Y, Zhang B, et al. Hyperspectral image classification based on multi-scale feature fusion residual network[J]. Laser & Optoelectronics Progress, 2022, 59(18): 1810014.
- [20] 陈文豪, 何敬, 刘刚. 引入注意力机制的卷积神经网络高光谱图像分类[J]. 激光与光电子学进展, 2022, 59(18): 1811001.
- Chen W H, He J, Liu G. Hyperspectral image classification based on convolution neural network with attention mechanism[J]. Laser & Optoelectronics Progress, 2022, 59(18): 1811001.
- [21] 高晗, 田育龙, 许封元, 等. 深度学习模型压缩与加速综述[J]. 软件学报, 2021, 32(1): 68-92.
- Gao H, Tian Y L, Xu F Y, et al. Survey of deep learning model compression and acceleration[J]. Journal of Software, 2021, 32(1): 68-92.
- [22] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size[EB/OL]. (2016-02-24)[2022-09-08]. <https://arxiv.org/abs/1602.07360>.
- [23] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17) [2022-08-06]. <https://arxiv.org/abs/1704.04861>.
- [24] 刘启超, 肖亮, 刘芳, 等. SSCDenseNet: 一种空-谱卷积稠密网络的高光谱图像分类算法[J]. 电子学报, 2020, 48(4): 751-762.
- Liu Q C, Xiao L, Liu F, et al. SSCDenseNet: a spectral-spatial convolutional dense network for hyperspectral image classification[J]. Acta Electronica Sinica, 2020, 48(4): 751-762.
- [25] Li W M, Chen H H, Liu Q, et al. Attention mechanism and depthwise separable convolution aided 3DCNN for hyperspectral remote sensing image classification[J]. Remote Sensing, 2022, 14(9): 2215.

Lightweight Residual Network Based on Depthwise Separable Convolution for Hyperspectral Image Classification

Cheng Rongjie¹, Yang Yun^{1,2*}, Li Longwei¹, Wang Yanting¹, Wang Jiayu¹

¹School of Geological Engineering and Geomatics, Chang'an University, Xi'an 710054, Shaanxi, China;

²Key Laboratory of Disaster Mechanism and Prevention of Mine Geological Disasters, Ministry of Natural Resources, Xi'an 710054, Shaanxi, China

Abstract

Objective Due to the high spectral resolution, hyperspectral remote sensing imaging technology can describe rich spectral features of ground objects, which is of great significance for the fine classification and recognition of ground objects. In the feature extraction and classification of hyperspectral images, the traditional deep learning network model employs the deep network structure to improve the classification accuracy. However, with the superposition of the convolution layer and the pooling layer, the phenomenon of gradient vanishing and gradient explosion appears in the model, which exerts adverse

effects on the classification. Although some scholars have proposed the residual network with the identity link to solve the model degradation caused by the deepening of the network model, it still has the shortcomings of large reference quantity and high time cost. To this end, a lightweight multi-scale residual network model (DSC-Res14) based on depthwise separable convolution is designed and built in this paper, which not only ensures the high classification accuracy of the model but also improves the model training efficiency. This innovative study provides a new solution to further promote the intelligent information extraction of hyperspectral remote sensing images.

Methods In this paper, a lightweight residual network (DSC-Res14) is proposed based on three-dimensional depthwise separable convolution instead of the traditional two-dimensional convolution, so that the long training time caused by large parameters in feature extraction and classification of hyperspectral images by traditional depth residual network is solved, and the performance of object classification using hyperspectral images is improved. For the proposed model with the input using image blocks after reduction dimensionally by principal component analysis, a convolution layer is first employed for initial feature extraction, and both spectral and spatial features of the image blocks are further extracted by three residual layers, each of which contains two residual structures. Finally, a full-connected layer is adopted to provide an input of one-dimensional feature vector for a classifier for pixel-to-pixel classification of hyperspectral images. To reduce network training parameters, this paper leverages the depthwise separable convolution for each residual structure in the residual layer. For the depthwise separable convolution operation, two-dimensional grouping convolution is utilized to extract spatial features from each channel, and then one-dimensional point convolution is employed to extract spectral features. After each convolution layer, batch normalization layer is added to keep the same distribution of input features, and the ReLU activation function is adopted to accelerate the network convergence speed and alleviate gradient disappearance.

Results and Discussions To verify the classification accuracy and speed of the proposed DSC-Res14 model, the paper compares this model with other three similar models and the Res14 model which employs the traditional 3D convolution kernel but has the same network structure as the proposed model. In terms of classification accuracy, the overall accuracy of Res14 on two public standard datasets has reached more than 99.5%, indicating the rationality of the network structure in this paper. For the categories with a small number of samples in the Indian Pines dataset, the classification accuracy of the DSC-Res14 model after introducing depthwise separable convolution is slightly decreased compared with Res14, but it still has obvious advantages over other similar models. In the Pavia University dataset, the accuracy indexes of the DSC-Res14 model proposed in this paper are all superior to Res14. The overall accuracy (OA) is 0.04% higher, the average accuracy (AA) 0.02% higher, and the kappa coefficient 0.04% higher, which shows the best performance among the similar models involved in the comparison. Under the conditions of relatively balanced and sufficient samples, the proposed DSC-Res14 model not only avoids a decline in classification accuracy with a reduction of the network parameters and an optimization of network structure, but also slightly improves the classification accuracy, compared with the traditional 3D convolution residual network. In contrast to similar models likely with depthwise separable convolution, the parameter number of the proposed model is smaller and the deep residual structure also leads to higher classification accuracy. For the three categories with fewer training samples of the Indian Pines dataset, the classification accuracy of other models is poor, but the classification accuracy of each category in the proposed model becomes better and more balanced with an average accuracy of 99.03%, which indicates the ability to deal with uneven sample probability distribution.

From a comparative analysis above, the conclusions are as follows. The introduction of depthwise separable convolution makes the parameter number of convolution layers and floating point operations (FLOPs) of the proposed DSC-Res14 model in the paper only 1/7 of that of the Res14 model, and the training time is about 1/3 of that of the Res14 model with ensuring high classification accuracy. The proposed model is proven to be a lightweight and efficient depth residual network.

Conclusions In this paper, a lightweight deep residual network model based on depthwise separable convolution is proposed to address the issue of large parameter size and longer training time caused by a deep network structure for improving classification accuracy using hyperspectral remote sensing images. Firstly, both spectral and spatial features of dimensional-reduced hyperspectral images by principal component analysis method are extracted through a three-dimensional convolution layer of the proposed network. Then, three 3D depthwise separable convolution residual layers with different spatial scales are introduced to extract deep semantic features of the given images. This reduces the number of training parameters of the network and enhances the expression ability in high-dimensional and multi-scale spatial features of the image. Experiments on the public Indian Pines and Pavia University datasets show that the classification accuracy of the proposed model is 99.46% and 99.65%. Compared with similar models, this model guarantees high classification accuracy and has fewer parameters and lower computation costs, shorter training time, and better robustness.

Key words remote sensing; hyperspectral image; depthwise separable convolution; spatial-spectrum feature; lightweight