



引用格式:刘春容, 宁 芊, 雷印杰, 等. 改进残差神经网络在遥感图像分类中的应用[J]. 科学技术与工程, 2021, 21(31): 13421-13429.
Liu Chunrong, Ning Qian, Lei Yinjie, et al. Application of improved residual network in sensing image classification[J]. Science Technology and Engineering, 2021, 21(31): 13421-13429.

改进残差神经网络在遥感图像分类中的应用

刘春容¹, 宁 芊^{1,2*}, 雷印杰¹, 陈炳才³

(1. 四川大学电子信息学院, 成都 610065; 2. 新疆师范大学物理与电子工程学院, 乌鲁木齐 830054;
3. 大连理工大学计算机科学与技术学院, 大连 116024)

摘 要 针对传统卷积神经网络随着深度加深而导致网络退化以及计算量大等问题, 提出一种改进残差神经网络的遥感图像场景分类方法。该方法以残差网络 ResNet50 作为主框架, 在残差结构中引入深度可分离卷积和分组卷积, 减少了网络的参数量和计算量, 加快模型收敛的同时也提升了分类精度。此外在网络中嵌入多尺度 squeeze and excitation block 模块对通道特征进行重校准, 提取出更加重要的特征信息, 进一步提升了网络的分类性能。在航空图像数据集(aerial image dataset, AID)和 UCMerced_Land Use 两个公开数据集上的分类精度分别为 91.92% 和 93.52%, 相比常规残差网络分类精度分别提高了 3.38% 和 10.24%, 证明所提方法在遥感图像场景分类任务中的可行性和有效性。
关键词 遥感图像; 场景分类; 残差神经网络; 分组卷积; 深度可分离卷积; 多尺度缩聚与激发模块
中图法分类号 TP391.41; 文献标志码 A

Application of Improved Residual Network in Sensing Image Classification

LIU Chun-rong¹, NING Qian^{1,2*}, LEI Yin-jie¹, CHEN Bing-cai³

(1. College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China;
2. College of Physics and Electronic Engineering, Xinjiang Normal University, Urumqi 830054, China;
3. College of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China)

[Abstract] In terms of that the deepening of traditional convolution neural network led to various problems, including the degradation of the network and the large amount of calculation, a remote sensing image scene classification method was proposed to improve the residual neural network. In this method, the residual network ResNet50 was taken as the main framework, and the deep separable convolution and grouping convolution were introduced into the residual structure, and thus the amount of network parameters and calculation were reduced, the convergence of the model was sped up and the classification accuracy was improved. Furthermore, the multi-scale squeeze and excitation block module was embedded into the network to recalibrate the channel features and extract more important feature information, and the classification performance of the network was further improved. The classification accuracy of AID(aerial image dataset) and UCMerced_Land Use reached 91.92% and 93.52%, respectively, which were 3.38% and 10.24% higher than that of the conventional residual network. It was proved that the proposed method in remote sensing image scene classification task is feasible and effective.
[Keywords] remote sensing image; scene classification; residual neural network; grouping convolution; depthwise separable convolution; multi-scale squeeze and excitation block

随着遥感技术的飞速发展, 获取大量高分辨率的遥感图像越来越容易, 这些高质量的遥感图像在实际生产生活中有着广泛的应用, 如精细农业、森林火灾检测、城市规划^[1]等。高分辨率遥感图像包含了非常丰富的地物纹理信息和空间语义信息, 如何将大量的遥感图像进行准确分类是遥感图像解析的重要研究内容。

遥感图像场景分类方法根据特征提取的层次

可分为低层特征处理和中层特征处理两大类方法。低层特征通常采用场景图像的颜色^[2]、方向梯度^[3]、密度特征^[4]、特征点^[5]、变换域的纹理^[6]等进行描述。这些低层特征不能很好地对中层语义进行描述, 难以跨越低层到高层语义的鸿沟, 泛化能力差。目前遥感图像分类算法主要集中于中层语义特征建模, 如视觉词袋(bag of visual words, BoVW)模型^[7]、空间金字塔匹配(spatial pyramid

收稿日期: 2021-03-06; 修订日期: 2021-08-12

基金项目: 国家自然科学基金(U1903215); 新疆维吾尔自治区区域协同创新专项(科技援疆计划)(2019E0214)

第一作者: 刘春容(1997—), 女, 汉族, 四川遂宁人, 硕士研究生。研究方向: 深度学习。E-mail: 2921253287@qq.com。

*通信作者: 宁 芊(1969—), 女, 汉族, 四川成都人, 博士, 教授。研究方向: 模式识别、智能系统。E-mail: ningq@scu.edu.cn。

matching, SPM) 模型^[8]、概率潜在语义分析 (probabilistic latent semantic analysis, pLSA) 模型^[9] 等方法, 虽然这些方法能一定程度地缩小低层到高层语义的鸿沟, 但其对图像深层特征的提取和场景的尺度变化缺乏有效的处理措施, 难以在复杂场景分类中取得较高精度。

近年来随着深度学习的快速发展, 以卷积神经网络为代表的深度神经网络^[10] 在图像识别、自然语言处理等领域中取得了较好的结果, 也有不少研究人员将其应用到遥感图像场景分类任务中^[11]。经典卷积神经网络模型有 VGG^[12] (visual geometry group)、AlexNet^[13]、GoogleNet^[14]、ResNet^[15] 等。Tianshun Z^[16] 等对 AlexNet 模型进行了改进, 将改进后的模型与迁移学习结合进行遥感图像分类实验, 在不增加训练时间的同时提升了分类精度。Muhammad^[17] 等选择预训练的 VGGNet 作为深度特征提取器, 采用典型相关分析对不同层的特征进行融合, 该方法在遥感图像场景分类任务中取得了较好的结果。Li 等^[18] 提出了一种自适应多尺度深度融合残差网络, 自适应特征融合模块通过学习权值来突出有用信息, 有效地抑制了无用信息。王改华等^[19] 提出了一种基于多分辨率和残差注意力机制相结合的图像分类模型, 通过在 DenseNet 网络稠密块中添加多分辨率因子来控制瓶颈层层数, 此外还在模型中引入了残差注意力机制, 能有效地从大量信息中筛选出少量重要的信息。张桐等^[20] 提出了一种深度多分支特征融合网络的方法进行遥感图像场景分类, 利用多分支网络结构提取中、高、低三个层次的特征信息, 将三个层次的特征进行基于拆分-融合-聚合的分组融合, 提升了模型的表征能力。王雨滢^[21] 等提出一种深度学习和支持向量机相结合的图像分类模型, 基于实际图像改进卷积神经网络, 并提取训练集的图像特征, 通过使用训练集的深度特征来训练支持向量机 (support vector machine, SVM) 分类器。史文旭^[22] 等提出了一种基于卷积神经网络 (convolutional neural network, CNN) 的多尺度方法结合反卷积网络的特征提取算法并对腺癌病理图像进行分类, 利用反卷积操作实现不同尺度特征的融合, 然后利用 Inception 结构不同尺度卷积核提取多尺度特征, 最后通过 Softmax 方法对图像进行分类。蓝洁等^[23] 提出一种基于跨层精简双线性池化的深度卷积神经网络模型, 首先根据 Tensor Sketch 算法计算出多组来自不同卷积层的精简双线性特征向量, 然后将归一化后的特征向量级联送至 softmax 分类器, 最后引入成对混淆对交叉熵损失函数进行正则化以优化网络。

正如上述所言, 越来越多的深度学习方法被运用到遥感图像分类任务中, 这些方法与传统机器学习方法相比有了很大提升, 但针对遥感图像空间信息复杂, 图像中关键物体小并且尺度变化大等问题, 这些方法还不能够对遥感图像信息进行很好的描述。同时随着神经网络的加深, 网络变得更加复杂和难以训练, 导致分类准确率饱和甚至下降。

针对上述问题, 提出一种改进的 ResNet50 网络模型, 在残差块中引入分组卷积和可分离卷积, 以减少网络的参数量和计算量, 加速模型的收敛。同时在每一个大的卷积组后嵌入多尺度缩聚与激发模块 (squeeze and excitation block, SE block), 将不同尺度的特征进行融合, 提高通道间的依赖性, 并以此进行模型训练, 以验证本文方法对遥感图像的分类有效性。

1 相关工作

1.1 残差神经网络

随着深度学习神经网络层数的不断增加, 网络的学习能力越来越强, 但相对的网络的收敛速度会减慢, 梯度在传播过程中会消失, 导致无法对前面网络层的权重进行有效调整。传统卷积神经网络, 网络中除了第一层外, 每一层的输入都来源于上一层的输出, 而残差神经网络采用跳跃式结构, 使得深度残差网络可以越过中间几层直接将参数传递给后面的层, 降低了网络的复杂度, 解决了深层次网络的退化问题, 促进了网络性能的提升。

残差神经网络的网络结构如图 1 所示, 该网络借鉴了高速网络的跨层连接思想。在图 1 所示的残差单元结构中, x 为网络的输入, $H(x)$ 为最优解映射, $F(x)$ 代表残差项, 直接把输入 x 传到输出作为初始结果, 输出 $H(x) = F(x) + x$, 当 $F(x) = 0$ 时, $H(x) = x$, 即恒等于映射。残差神经网络的训练目标就是要使得残差项 $F(x) = H(x) - x$ 的结果逼近于 0。让网络学习 $F(x) = 0$ 相比于更新该网络的参数来学习 $H(x) = x$ 更简单, 该层学习 $F(x) = 0$

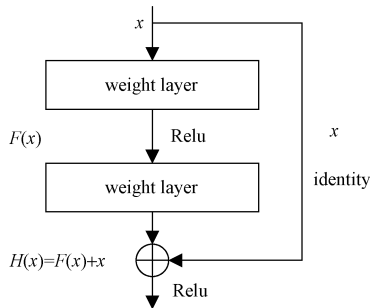


图 1 残差网络单元示意图

Fig. 1 Schematic diagram of residual network unit

的更新参数能够更快收敛。该种结构的残差网络与没有采用跳跃式结构的网络模型相比,其输入的数据更加清晰,能最大程度的保留数据的准确性。

1.2 分组卷积

分组卷积最早在 AlexNet 中出现,由于计算资源有限,通常将卷积操作分配给多个图形处理器 (graphics processing unit, GPU) 分别进行处理,最后将多个 GPU 的结果进行融合。分组卷积的原理如图 2 所示,假设输入尺寸为 $H \times W \times C$, 卷积核的个数为 N , 大小为 $k \times k$, 将输入特征图分为 G 个组,在图 2 中 $G = 2$, 则每一组的输入特征图数量为 C/G , 每组的卷积核个数为 N/G , 每组卷积核只与其所对应组的输入进行卷积。

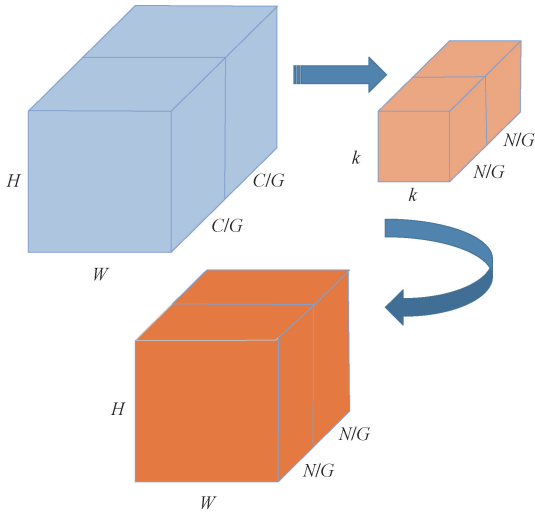


图2 分组卷积
Fig. 2 Grouping convolution

对于普通卷积来说,卷积核的总参数数量为

$$p_1 = k \times k \times N \times C \tag{1}$$

分组卷积所对应的卷积核参数总量为

$$p_2 = k \times k \times (N/G) \times (C/G) \times G \tag{2}$$

从计算结果可以看出,经过分组卷积可将总参数量降为原来的 $1/G$, 其用少量的参数量和计算量就能产生大量的特征图。

1.3 深度可分离卷积

深度可分离卷积分为深度卷积 (depthwise convolution) 和逐点卷积 (pointwise convolution) 两个过程,在进行卷积操作时将通道和空间区域分开考虑,对于不同的输入通道采取不同的卷积核进行卷积。深度卷积过程如图 3 所示,假设输入大小为 $H \times W \times C$, C 为通道数,在图 3 中 $C = 3$ 。假设卷积核个数为 m , 大小为 $k \times k \times 1$ 。则在深度卷积中,将输入分为了 C 组,每一组做 $k \times k \times 1$ 的卷积,每一个卷积核只提取其所在通道的空间特征。

逐点卷积的原理如图 4 所示,其将经过逐通道

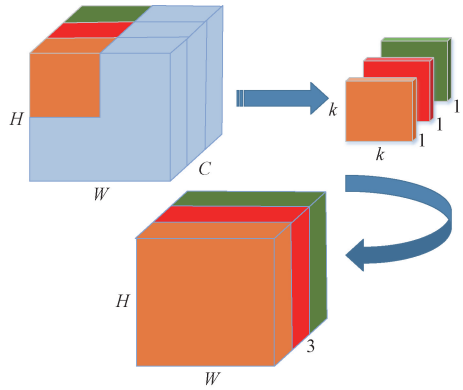


图3 深度卷积
Fig. 3 Depthwise convolution

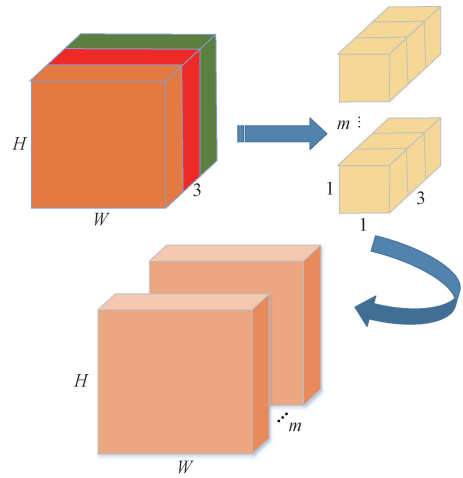


图4 逐点卷积
Fig. 4 Pointwise convolution

卷积后的特征做 m 个普通的 1×1 卷积,经过逐通道卷积核逐点卷积后的输出为 $H \times W \times m$ 。

对于普通卷积来说,在相同参数设置的情况下输出为 $H \times W \times C$, 其计算量为

$$p = H \times W \times C \times m \times k \times k \tag{3}$$

深度卷积的计算量为

$$p_1 = H \times W \times C \times k \times k \tag{4}$$

逐点卷积的计算量为

$$p_2 = H \times W \times C \times m \tag{5}$$

则深度可分离卷积相当于将普通卷积的计算量压缩为

$$\frac{p_1 + p_2}{p} = \frac{1}{m} + \frac{1}{k \times k} \tag{6}$$

从上述计算结果可以看出,深度可分离卷积相比于普通卷积,在深度一样的情况下大大减少了模型的计算量。

1.4 SENet 网络结构

SENet (squeeze and excitation networks) 是一种轻量级的网络结构,它通过对特征通道间的依赖关

系进行建模来获取每个特征通道的重要程度,然后依照这个重要程度提升有用的特征通道并抑制对当前任务不重要的特征。SENet 对输入特征进行 Squeeze 和 Excitation 两个重要的操作,通过 Squeeze 操作对输入的每个通道进行全局平均池化得到通道描述子,Excitation 操作利用两层全连层为每个通道生成权重,最后将 Excitation 的输出权重与先前特征进行重新加权,完成在通道维度上对原始特征的校准。

2 网络结构设计

2.1 改进残差神经单元

残差神经网络由多个 bottleneck block 残差块和 convolution block 残差块堆叠构成。bottleneck 残差块的主路径包含三个卷积层,第一部分 1×1 的卷积核用于减少特征通道的数量;第二部分 3×3 的卷积核用于进行特征的提取,同时减少了参数的数量;第三部分的 1×1 卷积核可以保证输出通道数等于输入通道数,在保证数据可靠性的同时增加了网络的深度,可以对数据进行充分的提取,进而提高模型的性能。同时残差块中引入了 BN(batch normalization)算法,加快了网络的收敛。convlution 残差块结构与 bottleneck 残差块结构相似,区别之处在于其分支上增加了一个 1×1 的卷积层,用于调整输入特征的尺寸。

对残差神经单元进行了改进,改进后的 bottleneck block 和 convolution block 结构分别如图 5、图 6 所示。

bottleneck block 中第一部分和第二部分的普通卷积替换为深度可分离卷积,第三部分普通卷积替

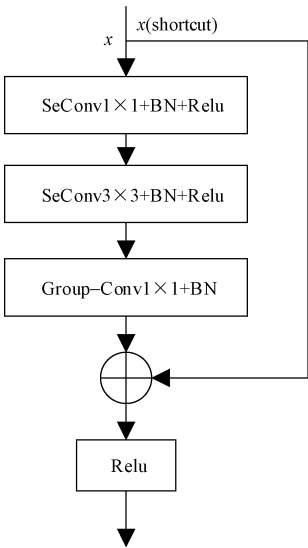


图 5 改进后 bottleneck block 结构(SGBB)
Fig. 5 Structure of improved bottleneck block(SGBB)

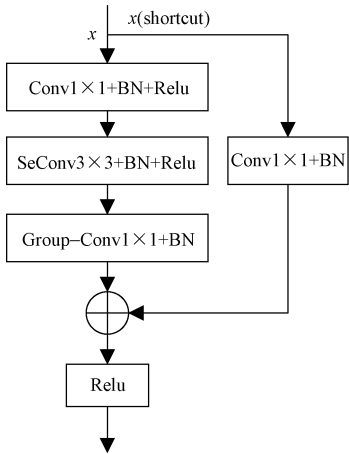


图 6 改进后 convolution block 结构(SGCB)
Fig. 6 Structure of improved convolution block(SGCB)

换为分组数为 2 的分组卷积。convolution block 中的第一部分保持不变,第二部分卷积采用深度可分离卷积,第三部分卷积采用分组数为 2 的分组卷积,分支上的卷积保持不变。改进后的残差块大大减少了整个网络的参数量和计算量,加快了网络的收敛,在通道数不变的情况下加强了模型的特征提取能力。

2.2 多尺度 SE block

在常规 SE block 中的全局平均池化前先对输入分别做 1×1 和 3×3 的卷积,将两组卷积结果相连接,再进行 Squeeze 和 Excitation 操作。其结构如图 7 所示。多尺度 SE block 有效地将不同尺度的特征进行融合,通过全连层和非线性激活层得到多尺度特征图的权重,然后将学习得到的权重与输入特征图相乘,通过对通道特征的重校准提升了有效特征的权重,增强了网络的特征表达能力,使得遥感图像场景分类精度得到提升。

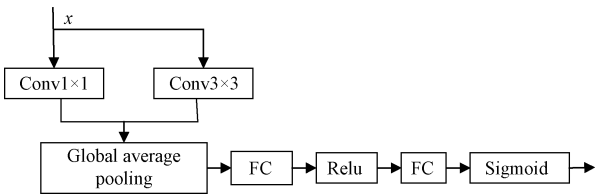


图 7 多尺度 SE block(MSEB)
Fig. 7 Multi-scale SE block(MSEB)

2.3 改进 ResNet50 网络结构

采用 ResNet50 作为基础模型,其 bottleneck block 和 convolution block 分别使用上文所对应的改进结构,即 SGBB 和 SGCB,本文中对网络的层数进行了调整,在 conv3_x 和 conv4_x 上分别去掉了 1 个和 3 个 bottleneck block 残差块,由于分组卷积和深度可分离卷积的引入,大大减少了网络的参数量和

计算量,有利于加快模型的收敛。同时在每一个大的卷积组后嵌入多尺度 SE 模块(MSEB),自动获取每个遥感图像特征通道的重要程度,有效增强有用特征,抑制用处不大的特征,进一步提升了网络的性能和泛化能力。改进后的网络结构如图 8 所示。

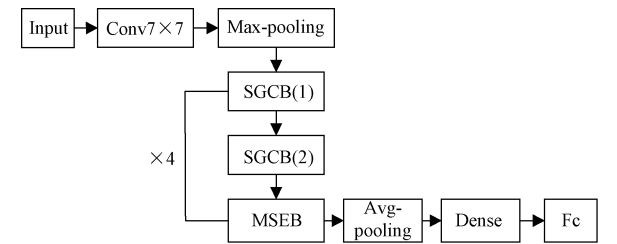


图 8 本文网络结构

Fig. 8 Network structure of the proposed method

3 实验与分析

3.1 实验数据集

实验数据集 1 为航空图像数据集(aerial image dataset, AID^[24]),该数据集由华中科技大学与武汉大学联合制作并于 2017 年发布,其中包含机场、森林、学校等 30 类遥感场景图像,每类遥感图像的样本数量为 200~420 张不等,尺寸均为 600×600 像素,该数据集共计 10 000 张图像。各类图像部分示例如图 9 所示。

实验数据集 2 为 UCMerced_Land Use (UCMLU)^[25]数据集,选自美国地质调查局国家城市地图航空遥感图像,包含建筑、飞机、棒球场、海滩等 21 类遥感场景图像,每类包含 100 张大小为 256×256 像素的图像,该数据集共计 2 100 副图像,空间分辨率为 0.3 m,各类图像部分示例如图 10 所示。

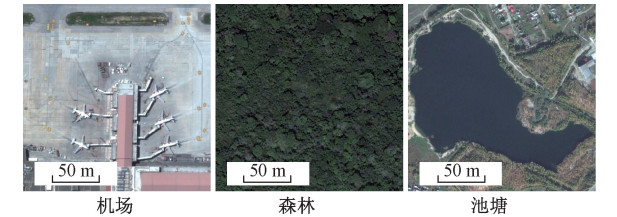


图 9 AID 数据集部分场景图像

Fig. 9 Part of the scene image of AID dataset

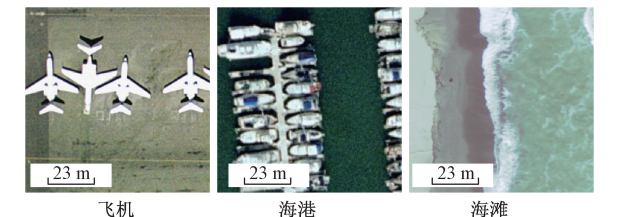


图 10 UCMLU 数据集部分场景图像

Fig. 10 Part of the scene image of UCMLU dataset

在以上两个数据集中,分别从每类遥感场景图像中随机选取 80% 的图像作为训练数据集,剩余 20% 的图像作为测试数据集。同时为了达到更好的测试效果,对两个数据集都采用在线数据增强的方式进行数据集扩充,增强方式包括随机旋转、随机缩放、随机裁剪、垂直变换等操作,并将所有图像尺寸统一为 256×256 像素大小。

3.2 实验环境及参数设置

本文中采用 Adam 优化器对模型进行优化训练,初始学习率设置为 0.001,当测试集的损失值持续 5 个 epoch 不再下降时,将学习率调整为原来的 0.1 倍,学习率的最小值设置为 0.000 01。更小的 batch_size 有助于增强模型的泛化能力,将 batch_size 设置为 20,网络一共经过 100 个 epoch 进行充分训练,损失函数采用交叉熵。

本文中实验的环境为:Windows10 操作系统,2.90 GHz Intel(R) Core(TM) i7-10700 CPU,内存为 16 GB,GPU 使用 NVIDIA GeForce RTX 2070,显存为 8 GB。神经网络训练使用 keras 框架。

3.3 评价标准

本文中采用总体分类精度(overall accuracy)、混淆矩阵(confusion matrix)和每秒浮点运算次数(floating-point operations per second,FLOPS)这三种评价指标对所提出方法的分类性能进行分析评估。总体分类精度指被正确分类的样本个数占总样本数的比值,能够很好地表征分类精度。混淆矩阵能更直观地展现每一类场景图像的分类准确率以及该类被错分为其他类的情况。计算量指浮点运算数,可以用来衡量模型的复杂度,精度越高,计算量越小,则网络的性能越好。本文中所有分类准确率值均为 5 次独立重复实验结果的平均值。

3.4 实验结果分析

3.4.1 网络分组数对比实验

为了更合理地选择分组卷积的分组数,本文将网络分组数分别设置为 2、4、8 组进行对比实验,在 AID 和 UCMLU 两个数据集上的实验结果分别如表 1 所示。

表 1 分组数对 AID、UCMLU 数据集分类精度的影响		
Table 1 Influence of grouping number on classification accuracy of AID dataset		
数据集	分组数	准确率/%
AID	2	91.92±0.39
	4	91.79±0.50
	8	89.38±0.79
UCMLU	2	93.52±0.77
	4	91.14±0.51
	8	89.28±0.72

从实验结果可以看出,分组数为 2 相比分组数为 4 和 8 在两个数据集上都能取得更高的分类精度,因此本文将分组数设置为 2。

3.4.2 AID 数据集结果分析

为了分析本文改进残差神经单元和引入多尺度 SE block 对分类精度的影响,分别对传统 CNN (ResNet50)、在残差结构引入分组卷积和可分离卷积的方法(SG_ResNet50)以及在 SG_ResNet50 基础上嵌入多尺度 SE block 的方法在 AID 数据集上进行了实验。三种方法平均分类精度随迭代次数的变化关系如图 11 所示。从图 11 中可以看出,使用未做任何改进的 ResNet50 做全训练的分类精度最低;对残差单元结构进行改进后,由于引进了分组卷积和可分离卷积,减少了网络的参数和计算量,改进后的模型 SG_ResNet50 比 ResNet50 准确率增长速度更快,模型更快收敛,并且分类准确率达到 90.42%,高于 ResNet50 准确率 1.88%;最后在 SG_ResNet50 的基础上嵌入多尺度 SE block,分类精度有了进一步提升,达到了 91.92%,说明多尺度 SE block 的引入有效地提升了模型性能。

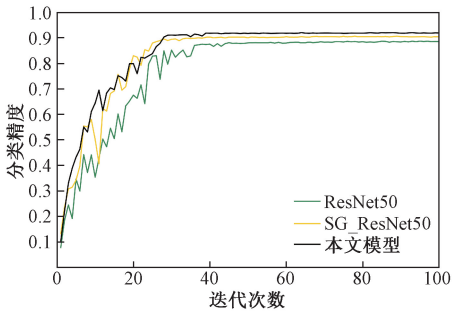


图 11 ResNet50、SG_ResNet50、本文模型分类精度对比
Fig. 11 Comparison of classification accuracy of ResNet50, SG_ResNet50 and the proposed model

表 2 中给出了 ResNet50、SG_ResNet50、SG_ResNet50_SE 和本文模型四种网络结构的性能对比,从结果可以看出,SG_ResNet50 相比于 ResNet50 的计算量降低了 60%,平均精度提升了 1.88%,表明在残差结构中引入分组卷积和深度可分离卷积大大降低了网络的计算量,同时分类准确率有了较大的提升。在 SG_ResNet50 的基础上引入多尺度 SE block,在计算量几乎不变的情况下分类精度有了进一步提升(提升了 1.5%)。同时本文模型的平均精度比 SG_ResNet50_SE 提高了 0.6%,表明引入多尺度 SE 的模型效果优于标准 SE 模型。

在测试集上进行 5 次独立重复试验平均分类准确率的混淆矩阵如图 12 所示,总体平均分类准确率为 91.92%。由图 12 可以看出,30 类场景图像中有 22 类的分类精度均高于 90%,说明本文提出的方法

对复杂场景图像具有较好的分类性能。同时从实验结果可以看出学校和广场的分类精度最低,分别只有 78% 和 67%,其中学校容易被划分为工业区和密集住宅区,广场容易被划分为中心区域和公园,说明本文所提出的方法对存在二义性的场景图像分类性能还有待进一步提高。

为了验证本文方法的优势,将本文方法在 AID 数据集上的分类精度与近几年具有代表性的几种方法进行对比,各方法的分类精度如表 3 所示。由表 3 可以看出,文提出方法的分类精度相比于中层特征提取方法 BoVW (SIFT) 提升了 23.55%,相对于以深度卷积神经网络为特征提取框架的 VGG-VD-16 和 GoogLeNet,分类精度分别提升了 2.28% 和 5.53%。同时与未做任何改进的 ResNet50 相比提升了 3.38%,证明本文所提出的模型在遥感图像场景分类中有一定优越性。

表 2 AID 数据集上不同模型性能对比 Table 2 Performance comparison of dataset AID by different models		
模型	计算量/GB	平均精度/%
ResNet50	0.35	88.54
SG_ResNet50	0.14	90.42
SG_ResNet50_SE	0.15	91.32
本文模型	0.17	91.92

表 3 AID 数据集上不同方法的平均分类准确率 Table 3 Average classification accuracy of dataset AID by different methods	
方法	准确率/%
BoVW (SIFT) [25]	68.37 ± 0.40
IFK (LBP) [25]	69.22 ± 0.72
VGG-VD-16 [25]	89.64 ± 0.36
GoogLeNet [25]	86.39 ± 0.55
ResNet50	88.54 ± 0.72
SG_ResNet50	90.42 ± 0.97
本文方法	91.92 ± 0.39

3.4.3 UCM 数据集结果分析

为了进一步验证本文提出模型的有效性,在 UCMLU 数据集上进行了实验,21 类场景图像分类结果如图 13 所示,总体平均分类精度达到了 93.52%。由图 13 可以看出,大部分场景图像类别分类准确率在 90% 以上,其中农业、棒球场、海滩等九类场景图像的分类准确率达到 100%,可见该方法对纹理差异较小的场景类别分类准确率较高。

为了更加直观地分析每一类场景图像的分类效果,图 14 给出了每一类场景分类准确率柱状图。从图 14 中可以看出仍有部分类别场景图像分类准确率较低,如高速公路和立交桥,其分类准确率分

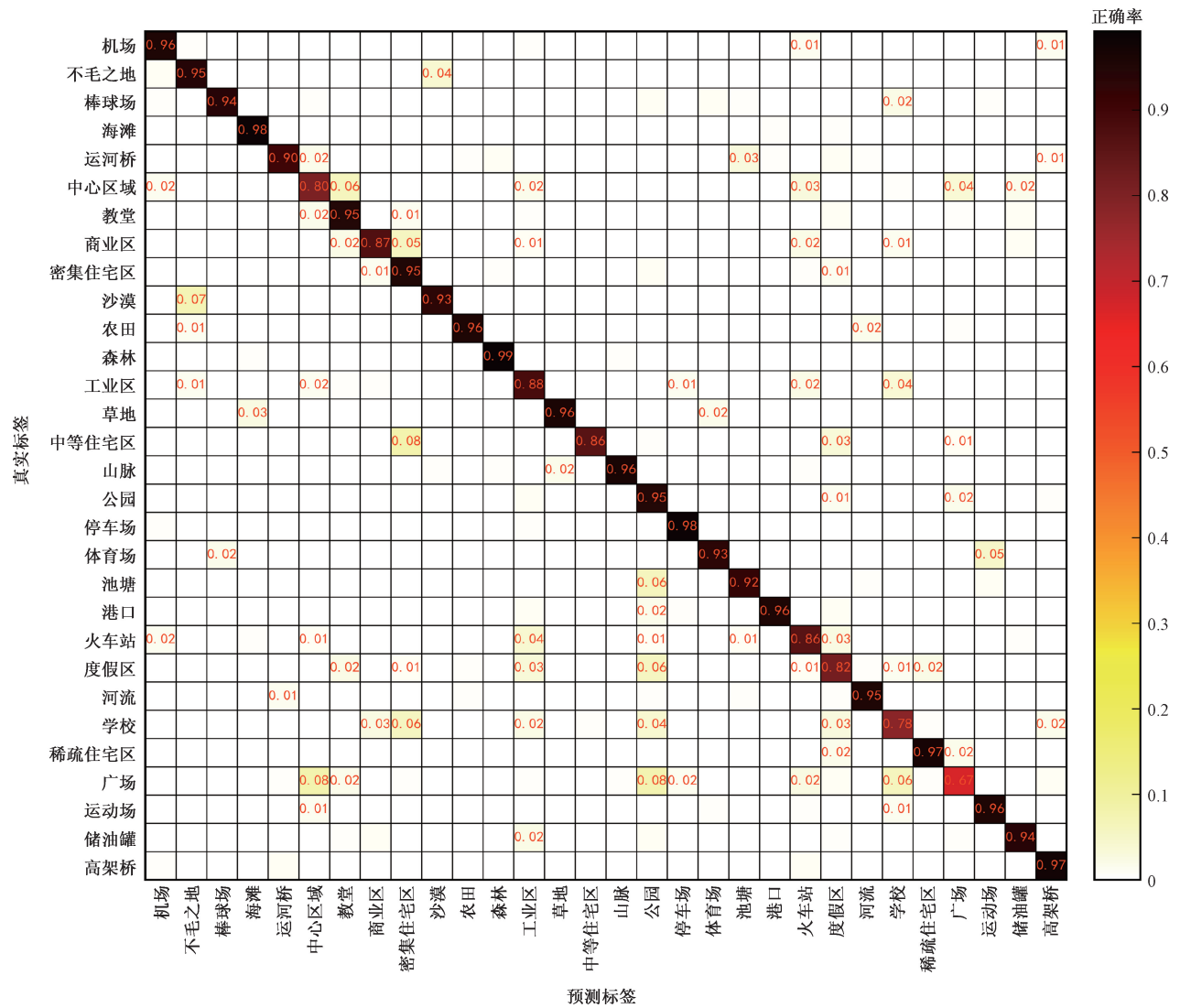


图 12 AID 场景分类混淆矩阵

Fig. 12 Classification confusion matrix for AID scene

别只有 78% 和 84%。结合图 13 的混淆矩阵分析可知,高速公路分类准确率较低是因为其和十字路口以及立交桥的场景十分相似,易造成误分。此外,中等密度住宅区与密集住宅区地物信息高度相似,造成部分中等密集住宅区被错分为密集住宅区。

UCMLU 数据集上不同模型的性能对比如表 4 所示,可以看出本文提出的模型相比于未做任何改进的残差网络大大降低了网络的计算量,同时拥有更高的分类准确率,体现了其强大的特征提取能力。

表 4 UCMLU 数据集上不同模型性能对比		
Table 4 Performance comparison of dataset UCMLU by different models		
模型	计算量/GB	平均精度/%
ResNet50	0.35	83.28
SG_ResNet50	0.14	91.47
SG_ResNet50_SE	0.15	92.99
本文模型	0.17	93.52

表 5 列出了本文提出方法在 UCMLU 数据集上的分类精度与一些现有方法的对比结果,由表 5 可以看出,传统的低中层特征提取方法如视觉词袋 (bag of visual word, BOVW) 的分类精度最低,而卷积神经网络由于其强大的特征提取和学习能力,使

表 5 UCMLU 数据集上不同方法的平均分类准确率	
Table 5 Average classification accuracy of dataset UCMLU by different methods	
方法	准确率/%
BOVW ^[7]	76.81 ± 0.41
BOVW + SCK ^[7]	77.71 ± 0.63
SVM ^[26]	78.57
SVM-LDA ^[26]	80.33 ± 0.71
EML ^[27]	86.86
ResNet50	83.28 ± 0.88
SG_ResNet50	91.47 ± 0.44
本文方法	93.52 ± 0.77

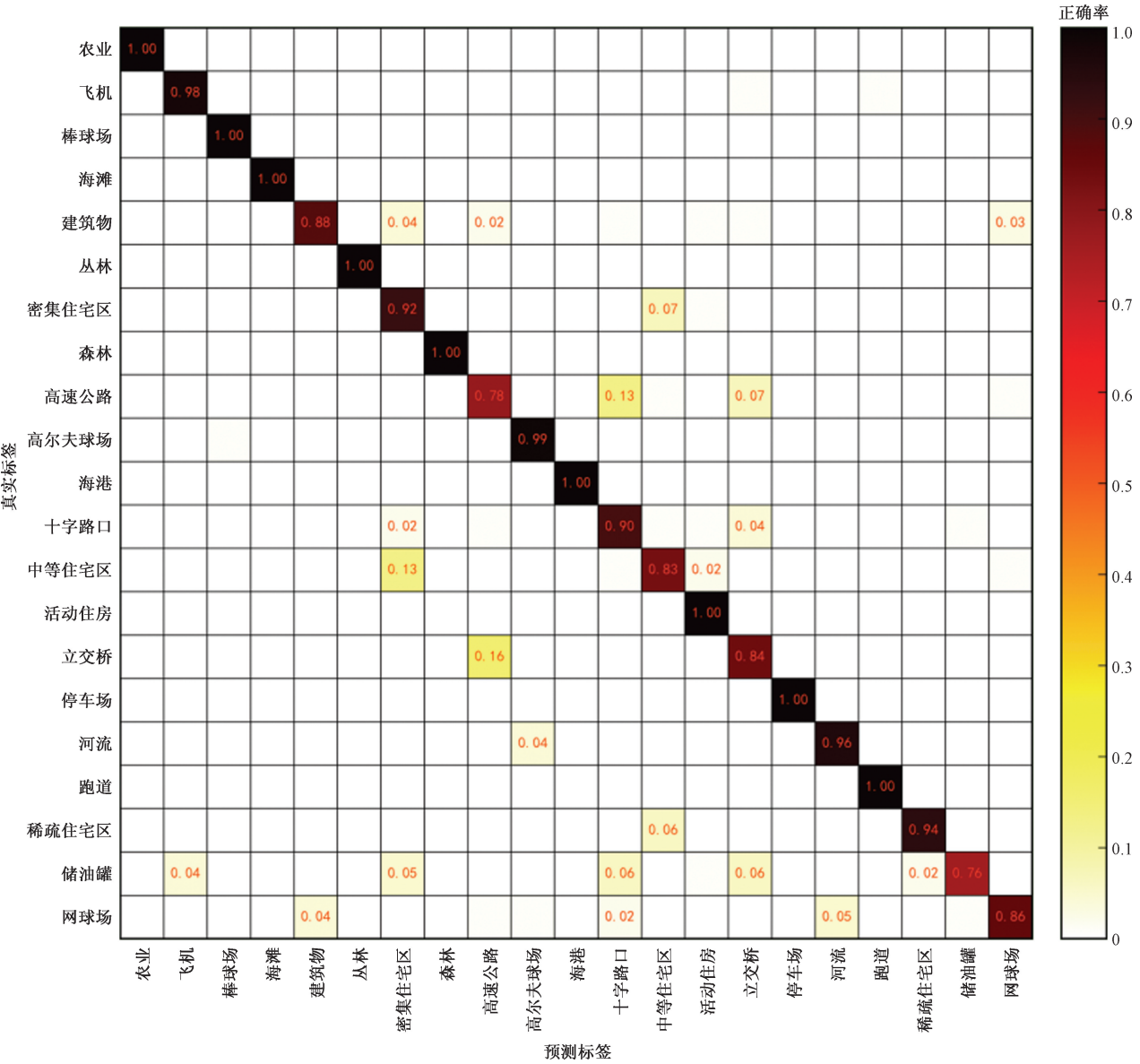


图 13 UCMLU 场景分类混淆矩阵
Fig. 13 Classification confusion matrix for UCMLU scene

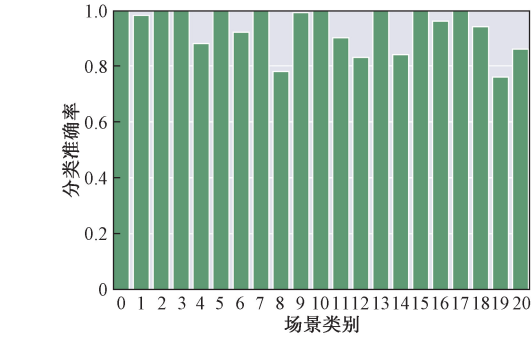


图 14 UCMLU 每类场景图像准确率柱状图
Fig. 14 Histogram of image accuracy of each scene in UCMLU
用其进行场景图像分类相比于传统方法有进一步提升。而本文提出的改进残差网络模型与以下几种较为流行的方法相比得到了最高分类精度

(93.52%), 相对于直接使用 ResNet50 进行分类训练, 本文方法的分类精度提升了 10.24%。

4 结论

针对遥感场景图像空间信息复杂、图像中关键物体小并且尺度变化大等问题, 提出一种改进残差网络的场景图像分类方法。该方法以残差神经网络 ResNet50 为基础框架, 在残差单元中引入分组卷积和可分离卷积, 有效减少了网络的参数量和计算量。同时引入多尺度 SE block 模块, 通过对多尺度特征的校准提升了有效特征的权重, 进一步提升了模型的性能。在 AID 和 UCMerced_Land Use 两个数据集上的分类精度分别达到了 91.92% 和 93.52%, 相比于常规的残差网络 ResNet50 分别提升了

3.38%和10.24%,证明本文方法在遥感场景图像分类上的可行性和有效性。但本文方法对地物信息及其相似的场景类别分类效果一般,如何对模型进行改进以适应场景信息更丰富、规模更大的数据集是下一步研究方向。

参 考 文 献

- [1] Cheng G, Han J, Lu X. Remote sensing image scene classification: benchmark and state of the art [J]. *Proceedings of the IEEE*, 2017, 105(10): 1865-1883.
- [2] van de Sande K E A, Gevers T, Snoek C G M. Evaluating color descriptors for object and scene recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(9): 1582-1596.
- [3] Zhou L, Dewen H U, Zhou Z T. Scene recognition combining structural and textural features [J]. *Science China (Information Sciences)*, 2013(7): 225-238.
- [4] Zhang F, Du B, Zhang L. Saliency-guided unsupervised feature learning for scene classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 53(4): 2175-2184.
- [5] Zhu X, Ma C, Liu B, et al. Target classification using SIFT sequence scale invariants [J]. *Journal of Systems Engineering & Electronics*, 2012, 23(5): 633-639.
- [6] Qian X, Guo D, Hou X, et al. HWVP: hierarchical wavelet packet descriptors and their applications in scene categorization and semantic concept retrieval [J]. *Multimedia Tools & Applications*, 2014, 69(3): 897-920.
- [7] Yang Y, Newsam S. Bag-of-visual-words and spatial extensions for land-use classification [C]// *Sigspatial International Conference on Advances in Geographic Information Systems*. New York: ACM, 2010: 270.
- [8] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid pooling for recognizing natural scene categories [C]// *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR06)*, New York: IEEE, 2006: 2169-2178.
- [9] Bosch A, Zisserman A, Muñoz X. Scene classification via pLSA [J]. *Lecture Notes in Computer Science*, 2006, 3954: 517-530.
- [10] Schmidhuber J. *Deep learning in neural networks* [M]. Amsterdam: Elsevier Science Ltd, 2015.
- [11] Cheng G, Ma C, Zhou P, et al. Scene classification of high resolution remote sensing images using convolutional neural networks [C]// *IGARSS 2016—2016 IEEE International Geoscience and Remote Sensing Symposium*. New York: IEEE, 2016: 767-770.
- [12] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. *Computer Science*, 2014: 1409-1556.
- [13] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]// *Advances in Neural Information Processing Systems (NIPS)*. Cambridge: MIT Press, 2012: 1097-1105.
- [14] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions [C]// *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 2015: 1-9.
- [15] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]// *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas: IEEE, 2016: 770-778.
- [16] Zhou T, Dang P, Xie H. Research on remote sensing image classification based on improved AlexNet network model [J]. *Bjng Surveying and Mapping*, 2018, 32(11): 1263-1266.
- [17] Muhammad U, Wang W, Chattha S P, et al. Pretrained VGGNet Architecture for Remote Sensing Image Scene Classification [C]// *2018 24th International Conference on Pattern Recognition (ICPR)*. Piscataway: IEEE, 2018: 1622-1627.
- [18] Li G, Li L, Zhu H, et al. Adaptive multi-scale deep fusion residual network for remote sensing image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(11): 8506-8521.
- [19] 王改华, 郭 钊, 周志刚, 等. 基于卷积神经网络的灾难场景图像分类 [J]. *科学技术与工程*, 2021, 21(17): 7217-7223. Wang Gaihua, Guo Zhao, Zhou Zhigang, et al. Image classification of disaster scenes based on convolutional neural network [J]. *Science Technology and Engineering*, 2021, 21(17): 7217-7223.
- [20] 张 桐, 郑恩让, 沈钧戈, 等. 基于深度多分支特征融合网络的光学遥感场景分类 [J]. *光子学报*, 2020, 49(5): 166-177. Zhang Tong, Zhang Enrang, Shen Junge, et al. Remote sensing image scene classification based on deep multibranch feature fusion network [J]. *Acta Photonica Sinica*, 2020, 49(5): 166-177.
- [21] 王雨滢, 赵庆生, 梁定康. 基于深度神经网络的电气设备图像分类 [J]. *科学技术与工程*, 2020, 20(23): 9491-9496. Wang Yuying, Zhao Qingsheng, Liang Dingkan. Electrical equipment image classification based on deep learning network [J]. *Science Technology and Engineering*, 2020, 20(23): 9491-9496.
- [22] 史文旭, 江金洪, 鲍胜利. 改进卷积神经网络在腺癌病理图像分类中的应用 [J]. *科学技术与工程*, 2019, 19(35): 279-285. Shi Wenxu, Jiang Jinhong, Bao Shengli. Application of improved convolutional neural network in pathological image classification of glandular [J]. *Science Technology and Engineering*, 2019, 19(35): 279-285.
- [23] 蓝 洁, 周 欣, 何小海, 等. 基于跨层精简双线性网络的细粒度鸟类识别 [J]. *科学技术与工程*, 2019, 19(36): 240-246. Lan Jie, Zhou Xin, He Xiaohai, et al. Fine-grained bird recognition based on cross-layer compact bilinear network [J]. *Science Technology and Engineering*, 2019, 19(36): 240-246.
- [24] Xia G S, Hu J, Hu F, et al. AID: a benchmark dataset for performance evaluation of aerial scene classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 55(7): 3965-3981.
- [25] Castelluccio M, Poggi G, Sansone C, et al. Land use classification in remote sensing images by convolutional neural networks [J]. *Acta Ecologica Sinica*, 2015, 28(2): 627-635.
- [26] Zhang F, Du B, Zhang L. Saliency-guided unsupervised feature learning for scene classification [J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2015, 53(4): 2175-2184.
- [27] Yang C W, Liu H P, Wang S C, et al. Remote sensing image classification using extreme learning machine-guided collaborative coding [J]. *Multidimensional Systems & Signal Processing*, 2017, 28(3): 835-850.