

面向遥感图像场景分类的轻量级沙漏密集网络

刘向举, 吴文彦, 蒋社想

安徽理工大学 计算机科学与工程学院, 安徽 淮南 232001

摘要:目的 在具有复杂空间结构和地理布局的遥感图像场景分类任务中,深度卷积神经网络(CNNs)虽然具有更好的分类性能,但是通常具有高复杂度,存在不适用于移动或嵌入式设备等问题,针对此,提出一种新的轻量级沙漏密集网络(LHD-NET),以实现分类精度和模型复杂性之间的良好平衡。方法 首先通过具有特征补偿机制的浅层混合下采样结构提取浅层信息,在保证信息充分提取的同时减少后续层的参数数量,从而在保持模型轻量级的同时提高性能;然后在沙漏结构间采用密集连接以提高特征复用,在一定程度上避免了梯度消失,促进了信息传递;最后利用一个具有较高语义信息的卷积层特征来指导多层特征聚合,以此来提高分类器的性能,同时训练过程中采用基于标签平滑的交叉熵损失函数,对真实标签进行平滑处理,相比于普通交叉熵损失函数能够有效提高鲁棒性和减轻模型过拟合问题。结果 实验结果表明:该模型在 5.4 M 参数量下取得了显著的分类性能,在 UC Merced Land-Use、SIRI-WHU 和 NWPU-RESISC45 3 个公开遥感数据集上分别取得了 99.19%、97.75% 和 92.38% 的平均分类准确率。结论 通过实验结果可证明所提模型能够在少量参数下便取得较好分类性能,相较于深度神经网络在保持高分类精度的前提下能显著降低模型参数量,对遥感图像场景分类任务及模型轻量化具有一定的参考价值。
关键词:遥感图像;场景分类;轻量级;卷积神经网络

中图分类号:TP391.4 文献标识码:A doi:10.16055/j.issn.1672-058X.2025.0004.003

Lightweight Hourglass Dense Network for Remote Sensing Image Scene Classification

LIU Xiangju, WU Wenyan, JIANG Shexiang

School of Computer Science and Engineering, Anhui University of Science and Technology, Anhui Huainan 232001, China

Abstract: Objective In the task of remote sensing image scene classification with complex spatial structures and geographical layouts, although deep convolutional neural networks (CNNs) have better classification performance, they often have high complexity and are not suitable for mobile or embedded devices. Therefore, a new lightweight hourglass dense network (LHD-NET) was proposed to achieve a good balance between classification accuracy and model complexity. **Methods** Firstly, shallow information was extracted through a shallow mixed downsampling structure with a feature compensation mechanism. This structure can reduce the number of parameters in subsequent layers while ensuring sufficient information extraction, thus improving performance while keeping the model lightweight. Then, dense connections were used between hourglass structures to improve feature reuse and to some extent avoid gradient disappearance, which promoted information transfer. Finally, a convolutional layer feature with high semantic information was used to guide multi-layer feature aggregation, so as to improve the performance of the classifier. Meanwhile, during the training process, the cross-entropy loss function based on label smoothing was employed to smooth the true labels,

收稿日期:2023-07-21 修回日期:2023-10-18 文章编号:1672-058X(2025)04-0017-10

基金项目:安徽省重点实验室项目(ZKSYS202204)。

作者简介:刘向举(1978—),男,硕士,副教授,从事物联网、智能控制和软件定义网络研究。

通信作者:吴文彦(1999—),男,硕士研究生,从事图像分类研究。Email:1774964485@qq.com。

引用格式:刘向举,吴文彦,蒋社想.面向遥感图像场景分类的轻量级沙漏密集网络[J].重庆工商大学学报(自然科学版),2025,42(4):17-26.

LIU Xiangju, WU Wenyan, JIANG Shexiang. Lightweight hourglass dense network for remote sensing image scene classification [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(4): 17-26.

which could effectively improve robustness and alleviate overfitting compared with ordinary cross-entropy loss functions.

Results Experimental results showed that the model achieved significant classification performance with only 5.4 M parameters, obtaining average classification accuracies of 99.19%, 97.75%, and 92.38% on three publicly available remote sensing datasets, namely UC Merced Land-Use, SIRI-WHU, and NWPU-RESISC45, respectively.

Conclusion Experimental results demonstrate that the proposed model can achieve good classification performance with a small number of parameters. Compared with deep neural networks, the proposed model significantly reduces the number of model parameters while maintaining high classification accuracy, providing certain reference values for remote sensing image scene classification tasks and model lightweighting.

Keywords: remote sensing image; scene classification; lightweight; convolutional neural network

1 引言

随着遥感成像技术的飞速发展,利用卫星或无人机拍摄的大量高分辨率遥感影像可以提供有关地表的丰富细节信息,例如空间布局、形状和纹理。这些信息作为重要的数据来源被广泛应用于土地利用分类^[1]、地理图像检索^[2]等诸多领域。作为遥感图像理解的一项基础性和挑战性任务,遥感图像场景分类已经成为近年来研究的热点之一。然而,遥感场景图像复杂的空间格局和地理结构给场景分类带来了很大的困难,因此,有效理解场景图像的语义内容和提取重要特征就显得尤为重要。

为解决信息提取问题,传统方法主要依赖于一些手工设计的视觉特征,例如颜色直方图^[3]、定向梯度直方图^[4]或尺度不变特征变换^[5]等,并试图提取有区别的场景表示进行分类。然而手工设计的视觉特征需要研究人员具有领域专业知识和经验,且花费大量时间和精力来进行特征提取,这些缺点使得手工特征提取难以适应现代大规模、多样化的数据。

近年来,以卷积神经网络(Convolutional Neural Network, CNN)为代表的深度学习技术由于其强大的特征学习能力,在计算机视觉领域取得了重大突破。与手工制作的特征相比,深度学习中的神经网络模型可以自动地学习特征,并且自动提取复杂、抽象的高层次特征。卷积神经网络作为深度学习技术中最主流的模式之一,已成为遥感图像场景分类的主流方法。如刘等^[6]提出基于 CNN 的多分辨率特征融合网络(Multi-Resolution Feature Network, MRFN),在城市用地分类上取得了较好效果,但是分类场景较为单一,不能充分体现模型泛化能力;岳等^[7]在 MRFN 基础网络上,融合多尺度特征 CNN 和自注意力机制提出多通道子注意力网络,通过捕获局部纹理信息和全局语义信息,进一步提升了分类性能;Wang 等^[8]提出场景的多尺度表示,构建全局-局部双流架构提取全局-局部特征,显著提高了特征表示能力;但是上述岳和 Wang 二位学者的研究

方法模型复杂度较高,无法适应资源受限的移动终端和边缘设备。因此,针对如何在模型轻量化和分类精度这二者间取得平衡这一问题,研究者们进一步提出了轻量化卷积神经网络。其中, Sandler 等^[9]以深度可分离卷积为基础构建了逆置残差结构,极大地减少了网络参数数量,并在一定程度上提高了精度;Tav 等^[10]通过 NAS 神经网络结构搜索(Neural Architecture Search, NAS)合适的图像输入分辨率、网络深度、通道数和宽度,并进行合理化配置来实现网络轻量化;Zhang 等^[11]使用通道重排和分组卷积的方法来减少模型的大小。上述几种网络模型虽然足够轻量化,但是在具有类内差异性大、类间相似性显著的遥感场景分类任务中表现一般。受此启发,Zhang 等^[12]使用 MobileNetV2 作为基础网络,并引入扩张卷积和信道注意力来提取鉴别特征,进一步提高 CNN 在场景分类中的性能,但是未能考虑浅层特征充分提取而造成重要信息丢失的问题,同时降低了模型鲁棒性;Shi 等^[13]结合深度可分离卷积提出了基于分支特征融合的轻量级 CNN 分类方法(LCNN-BFF),模型参数量显著降低,但是未考虑进行特征复用,导致较早层的参数更新较慢和学习得不够充分,造成网络退化和过拟合问题。

针对上述问题,本文提出了一种轻量级遥感图像场景分类方法。该方法首先通过并行的浅层混合下采样模块(Shallow Hybrid Downsampling Module, SHDM),在减小特征图大小的同时充分提取具有丰富边缘、纹理等基础特征的浅层信息;其次,利用轻量级沙漏密集卷积块(Sand Glass Dense Module, SGDM)充分捕获地物目标的特征信息,其密集连接可以缓解梯度和浅层信息丢失问题,避免网络退化;最后,通过多层特征融合器(Multi-Layer Feature Fusion Generator, MFFG)进行加权特征聚合,抑制无关信息,提高分类器的准确性。通过在 UC Merced_LandUse、SIRI-WHU 和 NWPU-RESISC45 3 个数据集上的大量实验证明了该方法的轻量高效性。

2 模型设计

2.1 模型整体结构

针对遥感图像空间结构复杂、类内差异高、类间相似性大的特点,提出了一个新的轻量级沙漏密集聚合网络(LHDA-NET),该网络模型的架构设计如图1所示。首先,通过两组浅层混合下采样模块(SHDM)充分提取输入图像的浅层特征,再由标准卷积和深度可分离卷积相结合的混合卷积模块进一步对输入的浅层特征进行融合和提取;接下来,利用4个连续堆叠的沙漏密集卷积块(SGDM)提取更丰富的条纹、纹路和形状等

中层纹理特征。为避免梯度消失或梯度爆炸的问题,在SGDM内部采用密集跳跃连接;同时,在第一到第三组SGBM模块中,采用步长为2的 3×3 标准卷积进行下采样,以此减少计算量和内存消耗,提高模型的泛化能力;随后再次使用混合卷积块,由顺序 1×1 卷积、 3×3 卷积和深度可分离卷积组成,以提取深层特征;然后将不同层的特征图送入多层特征融合器(MFFG)得到3组特征图 F_1 、 F_2 、 F_3 ,将其进行拼接融合得到多层融合特征;最后通过全局平均池化和Softmax分类器实现遥感场景图像分类。

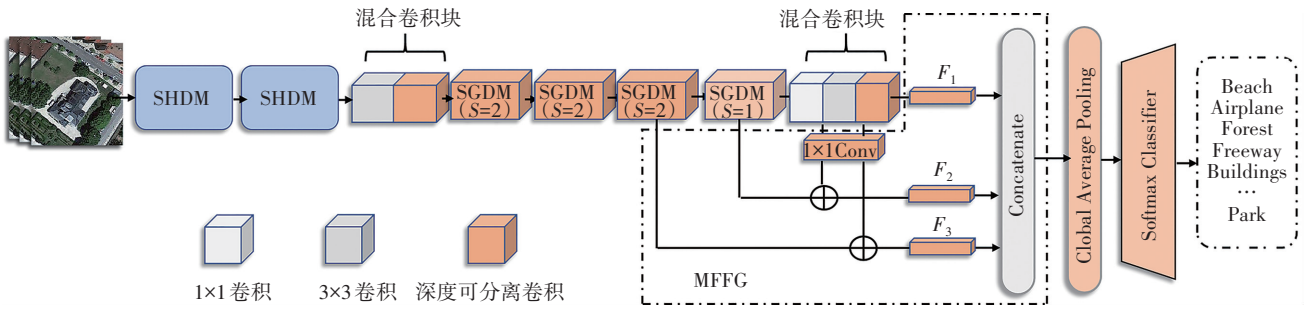


图1 模型整体结构

Fig. 1 The overall structure of the model

2.2 浅层混合下采样模块

下采样是将卷积特征图缩小到一定比例,在保留图像主要特征的同时缩减图像的空间大小,减少计算量。下采样方法主要有池化下采样和卷积下采样,通常来说,较小的卷积网络最好使用池化下采样,在较深的卷积网络中,多层叠加卷积可以帮助网络学习更复杂的非线性特征,因为每个卷积层可以捕获更多的抽象特征。因此,本文提出了基于池化和卷积的浅层混合下采样模块(SHDM),如图2所示。该模块由3个下采样分支组成,其中,分支1和分支2均由最大池化进行下采样,再分别通过 3×3 卷积和 1×1 卷积进行特征提取和升维得到 $p_1(x)$ 、 $p_2(x)$,然后将它们融合为 $f_1(x) = p_1(x) + p_2(x)$;由于池化会导致一些关键特征信息的丢失,不利于浅层特征的提取,分支3使用两个 3×3 卷积进行特征信息提取,其中第一个卷积步长为2用于下采样,然后将卷积下采样得到的特征 $f_2(x)$ 和 $f_1(x)$ 拼接为高维特征,即 $y(x) = \text{Concat}(f_1(x), f_2(x))$ 。通过并行的混合下采样方法,在减少特征大小和保证信息完整性的同时,可以增加网络的多样性,提高其学习能力和泛化性能;最后,通过混合卷积块对浅层特征进一步提取和融合。为保证网络的轻量化,混合卷积块(Depthwise Separable Convolution, DSC)使用MobileNet^[9]中的深度可分离卷积操作,由深度卷积(DepthWise Convolution, DWConv)和逐点卷积(PointWise Convolution, PWConv)组成,相比于普通卷

积,深度可分离卷积更轻量化。假设输入特征图 X 为 $C_{in} \times H \times W$,输出特征图为 $C_{out} \times H_0 \times W_0$,卷积核为 $D_K \times D_K$,则深度可分离卷积的参数数量如式(1)所示:

$$C_{DSC} = C_{DW} + C_{PW} = C_{in} \times D_K \times D_K + C_{in} \times C_{out} \quad (1)$$

而普通卷积的参数数量 C_{normal} 如式(2)所示:

$$C_{normal} = C_{in} \times C_{out} \times D_K \times D_K \quad (2)$$

深度可分离卷积和普通卷积的参数数量的比值 M 为

$$M = \frac{C_{DSC}}{C_{normal}} = \frac{C_{in} \times D_K^2 + C_{in} \times C_{out}}{C_{in} \times C_{out} \times D_K^2} = \frac{1}{C_{out}} + \frac{1}{D_K^2} \quad (3)$$

随着输出通道数 C_{out} 的增加, $\frac{1}{C_{out}}$ 的值可以忽略不计,则由式(3)可知,深度可分离卷积的参数数量是普通卷积的 $\frac{1}{D_K^2}$,达到使卷积网络轻量化的效果。

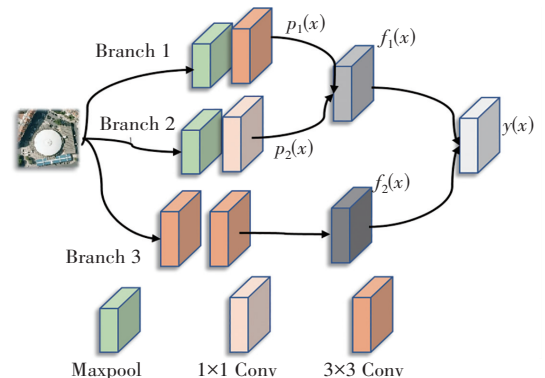


图2 浅层混合下采样模块

Fig. 2 Shallow hybrid downsampling module

2.3 沙漏密集卷积块和多层特征融合

2.3.1 沙漏密集卷积块

为应对遥感场景图像复杂的空间结构,提出了沙漏密集卷积块(SGDM),结构主要由沙漏瓶颈块(Sand Glass Bottleneck Block)和通道注意力模块(SE Module)^[14]组成。其中,本文采用的沙漏瓶颈块是轻量级网络 MobileNet^[15]中的骨干模块,相比于图 3(a)所示的逆置残差结构^[16],沙漏瓶颈结构能够在高维度进行恒等映射并进行空间变换,可以有效缓解信息丢失与梯度混淆。因为在逆置残差块中,中间扩展层编码的特征映射首先要投影到低维的残差块中,由于通道压缩,可能会有信息丢失和梯度混淆的风险;而 MobileNet 将 3×3 深度卷积置于两个 1×1 卷积之外执行两次深度卷积,从而确保可以在更高的维度上进行特征处理和表达,其结构如图 3(b)所示。

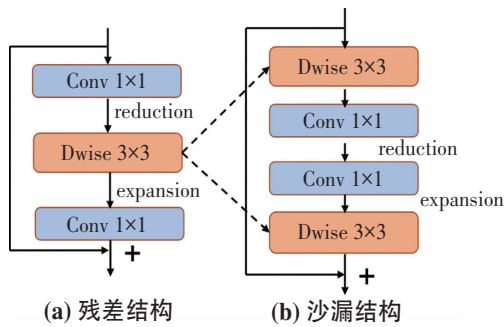


图 3 结构对比

Fig. 3 Comparison of structures

SE 注意力模块是特征通道领域的一种注意力机制,如图 4 所示。该通道注意力主要分为 Squeeze 和 Excitation 两个阶段。在 Squeeze 阶段,特征图经由全局平均池化(Global Average Pooling, GAP)操作进行压缩,得到输出为 $1 \times 1 \times C$ 的向量;在 Excitation 阶段,使用全

连接层将 Squeeze 阶段的输出映射到一个新的向量,并使用一个 sigmoid 函数将该向量中的每个元素压缩到 0 和 1 之间;最后,将这些元素乘以原始的特征图,以加权每个通道的特征图。

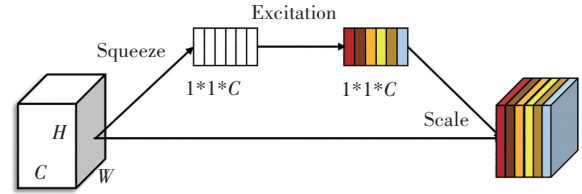


图 4 挤压-激励模块

Fig. 4 Squeeze-excitation module

结合上述两种模块的优点,SGDM 将沙漏瓶颈结构和通道注意力块交替堆叠,如图 5 所示(其中, C 和 C' , 分别表示输入、输出通道数, H 、 W 表示长和宽两个维度)。模型可以学习到更为复杂的特征,并且容易从输入数据中提取到重要信息。同时,沙漏瓶颈结构可以有效地进行梯度传递,而通道注意力能够自适应地调整每个通道的权重,对每个通道进行加权,以加强重要特征的影响,从而提高网络的泛化能力并加速训练过程,使其可以更好地适应不同的任务和数据集。此外,采用 add 方式的密集连接(Dense Connection)将上一层的所有输入连接到下一层的所有输出,这有助于更快地传递梯度,避免梯度稀释和消失,从而更好地利用之前所有层的特征,同时还能降低模型的计算成本和内存占用量。这种方法使模型更加紧凑和高效,具有更好的性能和泛化能力,同时也有助于降低过拟合风险,特别是对于遥感图像这样的复杂场景。为了在减小计算量和图像尺寸的同时避免信息丢失,模块最后采用了混合下采样结构,包括过渡层和步长为 2 的卷积层。其中,过渡层由最大池化和 1×1 卷积实现。

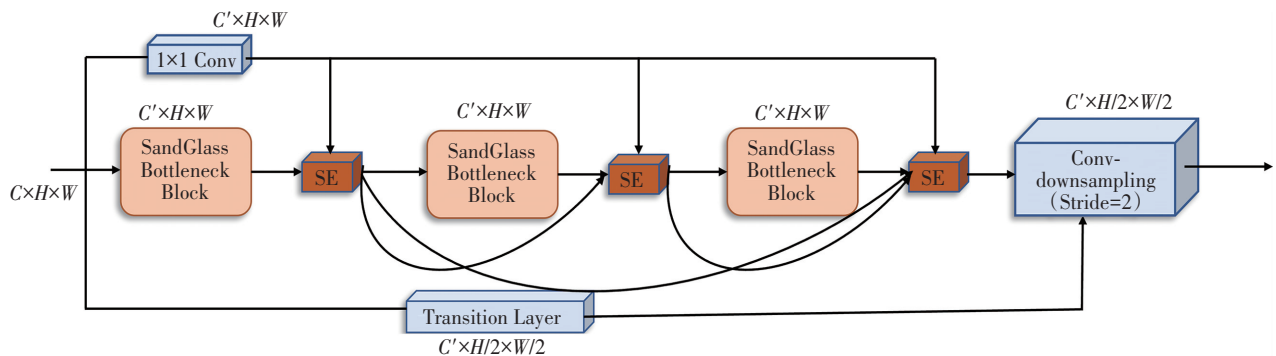


图 5 沙漏密集卷积块

Fig. 5 Hourglass dense module

2.3.2 多层特征融合器

一般来说,在遥感图像场景分类任务中,神经网络

较深层的特征可以更好地描述给定场景的语义信息,而较低层的特征具有丰富的外观信息,因此,聚合不同

层的特征可以获得更全面的场景表示。文献[17-18]也证明了融合多个卷积特征可以显著提升场景分类的性能。但是直接聚合不同层的特征而不考虑它们之间的差异,例如语义差异、空间差异等,可能会导致模型判别能力的下降。针对此问题,本文提出多层特征融合器(MFFG),具体思想是将每一个融合分支的特征与主干分支最后一个卷积层的特征进行加权求和,通过对不同特征层的加权来生成新的特征表示。因为最后一个卷积层的特征可以很好地表示给定场景的语义区域,利用最后一个卷积层特征来指导多层特征融合,从而可以获得更丰富、更准确的特征表示。如图1所示,在MFFG中,主干流经过 1×1 卷积将通道升维后,与两个分支流分别聚合得到加权求和特征图 F_2 和 F_3 ,最后将主干流特征图 F_1 与 F_2 、 F_3 进行特征拼接,形成一个更高维的特征向量,送入分类器进行分类。这种方式可以在一定程度上弥补不同层之间的特征差异,提高模型表达能力。

2.4 基于标签平滑的交叉熵损失函数

本文使用的损失函数为基于标签平滑^[19]的交叉熵损失函数(Cross-Entropy Loss with Label Smoothing),公式如式(4)所示。相比于普通的交叉熵损失函数,标签平滑的交叉熵损失函数在计算损失值时对真实标签进行了平滑处理,以缓解模型对训练数据的过拟合问题,同时使得模型更加关注数据中真实的分布情况,从而更好地学习到数据中的特征,提高模型的分类准确率。具体而言,该损失函数会将真实标签的概率值从1降低到一个小于1的数,并将剩余概率均匀地分配给其他类别。这样做可以使模型更加关注正确标签上的信息,并且鼓励模型在多个可能性之间进行均衡选择,从而提高模型的鲁棒性和精度。

$$L(x, y) = -\frac{1-\alpha}{K-1} \sum_{i=1}^K \tilde{y}_i \log(\hat{y}_i) \quad (4)$$

式(4)中, $\hat{y}_i$ 表示softmax函数的输出经过取对数后得到的值, $\tilde{y}$ 表示加权混合后的标签,公式为

$$\tilde{y} = (1-\alpha)y + \frac{\alpha}{K} \quad (5)$$

式(5)中, y 为真实标签, α 是平滑参数,本文中平滑参数的值设为0.1, K 是要进行分类的类别数。

3 实验与结果分析

3.1 实验数据集与预处理

为了验证所提出的模型在遥感场景分类任务上的分类性能,本文使用3个公共数据集:UC Merced_LandUse(UCM)^[20]、SIRI-WHU^[21]和NWPU-RESISC45(NWPU)^[22]。其中,UCM^[20]数据集是评价遥感图像场景分类性能的经典数据集,包括21类不同类型的地物,共计2100张图片,每类100张图片,图像像素大小为 256×256 ,部分场景的示例如图6(a)所示;SIRI-WHU^[21]数据集由武汉大学RE-IDEA团队设计,数据集由12个类别组成,共2400张图片,每类200张图片,图像像素大小为 200×200 ,部分场景的示例如图6(b)所示;NWPU^[22]包括45个类别的31500张遥感图像,每个类别包含700张图片,每张图片像素大小为 256×256 ,数据集涵盖从自然环境到城市景观的各种场景,部分场景实例如图6(c)所示。为防止在实验过程中出现过拟合,采用如下的数据增强方式:随机旋转,对输入图像进行 $0\sim 270^\circ$ 的随机旋转;随机偏移,对输入图像的长度和宽度进行随机偏移,偏移系数为0.2;随机缩放,对图像进行随机比例缩放,缩放比例为0.8;随机垂直水平翻转,对图像进行随机比例翻转。



图6 部分数据集样本

Fig.6 Partial dataset samples

3.2 实验环境及参数设置

实验环境: Windows11 操作系统、i5-12490F CPU、NVIDIA GeForce RTX 3060 GPU、内存为 16G,深度学习框架采用 Pytorch 框架。

在实验中,采用随机梯度下降(Stochastic Gradient Descent, SGD)作为模型的优化器,初始学习率设置为 0.02,动量设为 0.9,权重衰减设为 $1e-3$ 。学习率衰减方法采用余弦退火, batch_size 为 16,模型迭代次数为 200 个 epoch,损失函数为基于标签平滑的交叉熵损失函数。

3.3 评价标准

实验过程采用总体分类精度 OA(Overall Accuracy)、混淆矩阵 CM(Confusion Matrix)和参数量来进行性能评估。为保证实验结果的公平公正性,根据文献[7-8]的数据集划分方法,从 3 个公开数据集中随机选取数据进行训练和测试。在 UCM^[20]数据集中,随机选取 80%的场景图像作为训练集,其余作为测试集;在 SIRI-WHU^[21]数据集中,随机选取 50%和 80%的场景图像作为训练集,剩余 50%和 20%作为测试集;在 NWPU^[22]数据集中,随机选取 10%和 20%的场景图像作为训练集,剩余的 90%和 80%作为测试集。在相同的实验环境和参数设置下,所有实验均重复进行 5 次,取平均值和标准差作为最终的性能指标。

3.4 数据集实验结果及分析

为验证所提 LHD-NET 模型在 UCM^[20]、SIRI-WHU^[21]和 NWPU^[22]这 3 个遥感图像场景分类数据集上的性能,在上述 3 个公开数据集上进行充分实验,并与现有算法展开对比,分别如表 1—表 3 所示。

如表 1 所示,所提模型 LHD-NET 与一些现有方法在 UCM^[20]数据集上的分类效果进行对比,结果显示:无论是在准确率还是在模型参数量上都具有显著优势,LHD-NET 的准确率达到了 99.19%,超出了所有比较方法。本文方法比 VGG16-DCNN^[24]高 0.26%,比 DSFATN^[25]高 0.94%,比 AlexNet-SCCOV^[27]高 1.15%,并且相比这些方法,参数量显著减少;同时,从表中可以看出,LHDA-NET 只有 0.11%的准差,在所有比较方法中最小,表明所提出模型性能的稳定性和一致性较好,拥有较好的鲁棒性。其原因是 LHD-NET 的沙漏密集卷积模块在保证较少参数量的同时,内部采用密集连接方式,在避免梯度消失的前提下进行了充分的特征复用和信息融合,使得模型在处理不同类型和程度的干扰时,依然能够保持较好的性能表现和稳定性。

表 1 不同方法在 UCM 数据集上的对比(80%训练集)

Table 1 Comparison of different methods on the UCM dataset(80% training set)

分类方法	准确率/%	参数量/M
ReNnet 50-WSPM-CRC ^[23]	97.95	23
VGG16-DCNN ^[24]	98.93±0.10	130
AlexNet-DCNN ^[24]	96.67±0.10	60
DSFATN ^[25]	98.25	143
ADFF ^[26]	98.81±0.51	—
AlexNet-SCCOV ^[27]	98.04±0.23	6
VGG16-SCCOV ^[27]	99.05±0.25	13
ECNN ^[28]	98.00±0.18	—
LHD-NET	99.19±0.11	5.4

表 2 展示了将所提模型 LHD-NET 和一些现有方法在 SIRI-WHU^[21]数据集上的比较结果:在 80%和 50%两个训练比率下,LHD-NET 分别达到了 97.75%和 95.83%,与一些大规模方法相比,如 Siamese AlexNet^[30]、Siamese VGG16^[30]和文献[7]的方法,在 80%的训练数据下,准确率分别提高了 8.79%、0.45%和 3.75%,并且参数量仅为其 9%、4.15%和 7.97%;与轻量级方法 MRFN^[6]相比,LHDA-NET 模型虽然在参数量上高于 MRFN^[6]方法 0.17M,但是在少量参数增加的前提下,取得了十分可观的性能提升,准确率相比 MRFN^[6]提升了 13.58%;并且与其他方法相比,LHD-NET 模型在保持模型轻量化的同时,在准确率上也是最优的。

表 2 不同方法在 SIRI-WHU 数据集上的对比

Table 2 Comparison of different methods on the SIRI-WHU dataset

分类方法	准确率/%		参数量/M
	50%训练比率	80%训练比率	
BOVW ^[29]	—	73.93	—
MRFN ^[6]	—	84.17	5.23
Siamese AlexNet ^[30]	83.25	88.96	60
Siamese VGG16 ^[30]	94.50	97.30	130
Siamese RestNet ^[30]	95.75	97.50	26
文献[7]	—	94.00	67.71
LHD-NET	95.83±0.17	97.75±0.31	5.4

NWPU^[22]数据集在空间分辨率、视角、光照、背景等方面具有多样性,有较高的类内多样性和类间相似性,且拥有高达 3 万多张遥感图像,是一个极具挑战性的数据集,因此把它作为本文方法的最后一个性能评估数据集,以此验证 LHD-NET 模型在较大规模数据集上的效果。表 3 列出了将本文模型 LHD-NET 和其他

一些优秀方法在 NWPU^[22]数据集上的分类效果。从表 3 中可以看出:所提模型 LHD-NET 在 10%和 20%的训练比率下分别达到了 88.46%和 92.38%的分类准确率,对比比较的方法,LHD-NET 模型在 20%的训练数据集下达到了最佳的分类效果;VGG16-DCNN^[24]方法的准确率虽然在 10%的训练数据下比本文方法略高 0.76%,但 LHD-NET 的参数数量是 VGG16-DCNN^[24]的 1/24,相比而言,本文模型在更小的计算资源消耗下,在性能方面与 VGG16-DCNN^[24]相近,因此更具有竞争优势。

表 3 不同方法在 NWPU 数据集上的对比
Table 3 Comparison of different methods on the NWPU dataset

分类方法	准确率/%		参数量 /M
	10%训练比率	20%训练比率	
Siamese VGG16 ^[30]	—	90.06	130
Siamese ResNet ^[30]	—	92.11	26
VGG16-DCNN ^[24]	89.22±0.50	91.89±0.22	130
Alexnet-DCNN ^[24]	85.56±0.20	87.24±0.12	60
ADFF ^[26]	86.01±0.15	88.79±0.17	—
LASC-CNN ^[31]	81.37	84.30	—
Alexnet-SCCOV ^[27]	84.33±0.26	87.30±0.23	6
ECNN ^[28]	87.21	89.93	—
LHD-NET	88.46±0.23	92.38±0.12	5.4

3.5 混淆矩阵可视化分析

混淆矩阵广泛应用于图像分类领域,它总结了所有基于实际类别和预测类别的分类结果,并展示了分类器对每个类别的分类效果,是评估分类模型性能的常用工具。在本文的混淆矩阵中(图 7),矩阵的每行代表预测标签,每列代表真实标签,对角线上的数值为分类正确的百分比,并且对角线上的颜色越深,分类准确率越高。鉴于 NWPU^[22]数据集类别过多,本文仅在 UCM^[20]和 SIRI-WHU^[21]两个公开数据集上进行混淆矩阵可视化实验。

图 7 是 LHD-NET 模型在 UCM^[20]数据集上获得的混淆矩阵。从图中可以看出:除了在类别 6(密集住宅)和类别 4(建筑)、类别 18(稀疏住宅)和类别 12(中型住宅)容易混淆外,几乎所有的类别分类精度均能达到 100%。在一些类别上,分类错误可能有以下两个原因:一是部分场景由多个不同的土地利用单元组成,即类内差异较高,使得网络模型难以捕获正确特征;二是部分场景表现出不同的空间布局特征,增大了模型分类的难度。尽管如此,在类别 6(密集住宅)和类别 4(建筑)等易混淆的场景,依旧保持较好的分类效果(90%

以上)。为探明发生混淆的原因,各给出混淆类的一份错分类样本(密集住宅—建筑、稀疏住宅—中型住宅)。观察图 8 中分类错误的样本可知:密集住宅和建筑在空间布局上相似,稀疏住宅和中型住宅在地物类型上相似,导致模型分类时发生混淆。但 LHD-NET 模型依旧保持较高分类性能,说明 LHD-NET 模型能够充分捕获到遥感影像的关键特征,从而能够适应类内差异和类间相似度高的样本,并取得较好的分类效果。

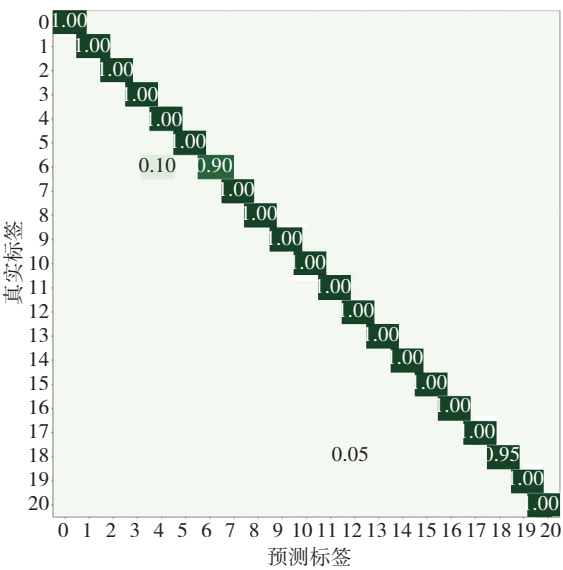


图 7 UCM 数据集的混淆矩阵

Fig. 7 Confusion matrix for the UCM dataset

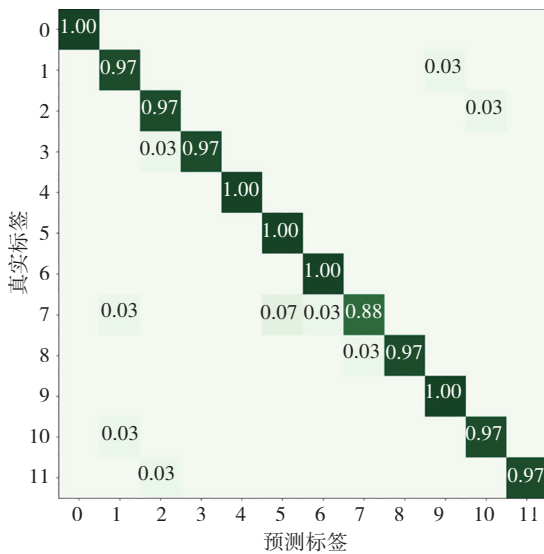


图 8 错分类样本案例

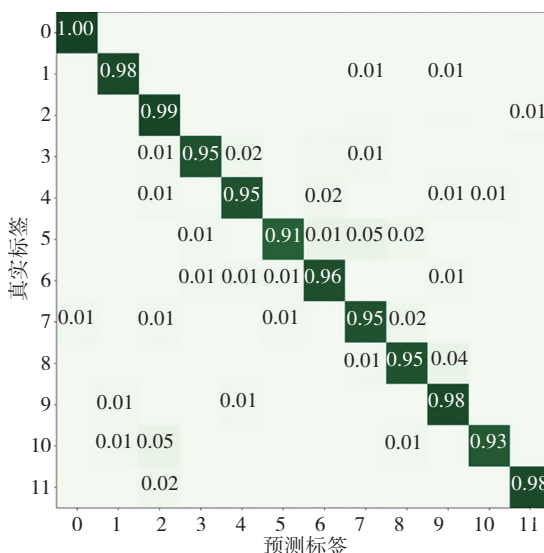
Fig. 8 Misclassified sample cases

图 9(a)和图 9(b)为在 SIRI-WHU^[21]数据集上的混淆矩阵。由图 9(a)可知:在 80%训练数据下,有 11 个场景均达到 97%以上的识别率。由图 9(b)可知:在

50%的训练数据下,有 11 类场景准确率均达到 95%以上。说明本文模型在该数据集上有较好的预测性能。但是,从图 9(a)中可知,在类别 7(公园)上的错分类比例为 12%,其中 7%被错分类为类别 5(草地)。对数据样本进行分析后发现,类别 7 和类别 5 具有相近的纹理特征和颜色,并且二者在地物类型上也具有较高的相似度,这使得这两个类别的类间相似性较大,增大了模型的判别难度。同时类别 7(公园)包括了大量类别 5(草地)的特征,这也造成了在少部分样本上的错误预测。



(a) 训练比率为 80%



(b) 训练比率为 50%

图 9 不同训练比率下 SIRI-WHU 数据集的混淆矩阵

Fig. 9 Confusion matrix of SIRI-WHU dataset with different training ratios

3.6 消融实验

本文在 UCM^[20]、SIRI-WHU^[21] 和 NWPU^[22] 3 个公开数据集上进行了充分实验,定量地评价多层特征融

合器(MFFG)以及基于标签平滑的交叉熵损失函数中平滑因子的选择对模型性能的影响。

为验证本文提出的多层特征融合器的有效性,比较了以下特征融合方法,即 LHD-BS(无融合)、LHD-DA(直接融合)以及 LHD-NET(加权融合)。具体来说,LHD-BS 表示仅使用单一的特征层,并未融合多层特征,以便比较其与多层特征融合方法之间的差异;LHD-DA 表示多个特征层组合在一起,但没有采用任何特殊的加权或组合方法将多个卷积特征映射直接聚合;LHD-NET 是本文提出的多层特征加权融合方法,对分支流上的卷积特征进行加权聚合。由表 4 可知:相比于 LHD-BS 和 LHD-DA 两种方法,LHD-NET 在 3 个数据集上的准确率均有所提升。其中,在 UCM 数据集上相比前两种方法准确率分别提升了 0.34% 和 0.17%;在 SIRI-WHU 数据集上分别提升了 0.19% 和 0.14%;NWPU 数据集上分别提升了 0.17% 和 0.25%。通过实验数据表明:本文提出的多层特征加权融合方法相比单一特征和直接融合方法更具有优势,这说明在复杂的遥感图像场景分类任务中,对不同层特征进行加权融合的确可以提高模型预测准确性。

表 4 不同特征融合方法比较

Table 4 Comparison of different feature fusion methods /%

融合方法	准确率		
	UCM	SIRI-WHU	NWPU
LHD-BS	98.85±0.17	97.56±0.29	92.21±0.16
LHD-DA	99.02±0.12	97.61±0.21	92.13±0.17
LHD-NET	99.19±0.11	97.75±0.31	92.38±0.12

本文在实验过程中采用基于标签平滑的交叉熵损失函数,其中的平滑因子是一个超参数。平滑参数的取值范围一般在 0~1 之间,较小的平滑参数可以提高模型的预测准确率,但也可能会使模型过拟合;较大的平滑参数可以降低模型的过拟合风险,但可能会导致模型预测准确率降低。当平滑参数为 0 时,不会进行标签平滑,此时损失函数退化为传统的交叉熵损失函数;但是,若平滑参数为 1,会导致过于平滑,使得模型预测结果过于保守。在不同的数据集和神经网络模型上,平滑参数的取值都有所不同,因此本文对平滑参数的选择进行了分析实验,其取值分别为 0、0.001、0.01、0.1、0.2、0.3,并在 3 个公开数据集上进行选择实验。结果如表 5 所示:当平滑参数逐渐增大,在 3 个数据集上准确率基本上呈现先下降后上升再下降的态势;由于平滑参数 0.001 过小,导致模型过拟合,所以相比普

通交叉熵损失函数(平滑参数为 0),其分类精度略有下降;在平滑参数取值逐渐增大时,模型的泛化能力逐渐提升,性能逐渐提升,在平滑参数取 0.1 时均达到最高性能;之后继续增大参数取值,导致模型欠拟合,分类精度又逐渐下降。因此,通过实验证明:本文所提出的模型下,当平滑参数设置为 0.1 时,在 3 个数据集上获得最佳结果。

表 5 不同平滑参数取值时的准确率分析
Table 5 Accuracy analysis with different values of smoothing parameters /%

数据集	平滑参数取值					
	0	0.001	0.01	0.1	0.2	0.3
UCM	98.81	98.80	98.32	99.19	98.60	98.57
SIRI-WHU	97.46	97.20	97.29	97.75	97.71	97.29
WNPU	91.97	91.89	92.33	92.38	91.90	91.41

3.7 鲁棒性测试

为测试 LHD-NET 模型的鲁棒性,本文在 AID、RSSCN7、RSC11 和 WHU-RS19 这 4 个公开的遥感图像场景分类数据集上进行了广泛的实验。其中训练集和测试集的比率均为 5:5,所有实验均随机进行 5 次,取平均值和标准差作为最终结果。结果如表 6 所示。

表 6 模型在不同数据集下的测试性能
Table 6 Model performance tested with different datasets

数据集	图像数量	场景类别数量	平均准确率/%
AID	10 000	30	95.97±0.14
RSSCN7	2 800	7	93.56±0.32
RSC11	1 232	11	93.33±0.37
WHU-RS19	1 005	19	92.81±0.44

通过表 6 可知:LHD-NET 模型在上述 4 个公开数据集上依旧取得优异的分类性能,说明本文模型拥有较好的鲁棒性,能够适应不同的遥感场景分类任务。

4 结 论

本文提出一种轻量级沙漏密集聚合网络(LHD-NET)对遥感场景图像进行分类。在所提出的 LHD-NET 模型中,设计了浅层混合下采样模块提取通用的特征信息,接着利用轻量级沙漏密集卷积块捕获更加精细化的细粒度特征,最后通过指导多层特征聚合,实现场景分类。通过在 3 个数据集上不同训练比例的实验结果可知,所提模型可以以更少的参数量获得更高的分类精度。但是,在卷积神经网络中,由于卷积操作

只关注局部区域信息,难以捕获图像中的全局信息,因此,在未来的工作中,尝试将卷积神经网络和 Transformer 进行结合,从局部-全局来提取图像信息。此外,由于 Transformer 模型具有更高的参数量和计算量,使得复杂度增加,平衡 CNN-Transformer 模型的性能和复杂度和也是未来的工作重点之一。

参考文献(References):

[1] MOU L, GHAMISI P, ZHU X X. Deep recurrent neural networks for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(7): 3639-3655.

[2] 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.

[3] SWAIN M J, BALLARD D H. Color indexing[J]. International Journal of Computer Vision, 1991, 7(1): 11-32.

[4] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, 2005, 1: 886-893.

[5] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[6] 刘美, 卿粼波, 韩龙玫, 等. 基于遥感影像和神经网络的城市用地功能分类[J]. 太赫兹科学与电子信息学报, 2021, 19(1): 132-137.

LIU Mei, QING Lin-bo, HAN Long-mei, et al. Urban land use classification based on remote sensing images and neural network[J]. Journal of Terahertz Science and Electronic Information Technology, 2021, 19(1): 132-137.

[7] 岳泓光, 韩龙玫, 王正勇, 等. 基于多通道自注意力网络的遥感图像场景分类[J]. 四川大学学报(自然科学版), 2023, 60(2): 97-105.

YUE Hong-guang, HAN Long-mei, WANG Zheng-yong, et al. Remote sensing scene classification based on multi-channel self-attention network[J]. Journal of Sichuan University (Natural Science Edition), 2023, 60(2): 97-105.

[8] WANG Q, HUANG W, XIONG Z, et al. Looking closer at the scene: Multiscale representation learning for remote sensing image scene classification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 33(4): 1414-1428.

[9] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 4510-4520.

- [10] TAV M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//International Conference on Machine Learning. PMLR, 2019: 6105–6114.
- [11] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 6848–6856.
- [12] ZHANG B, ZHANG Y, WANG S. A lightweight and discriminative model for remote sensing scene classification with multidilation pooling module[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(8): 2636–2653.
- [13] SHI C, WANG T, WANG L. Branch feature fusion convolution network for remote sensing scene classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 5194–5210.
- [14] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 7132–7141.
- [15] ZHOU D, HOU Q, CHEN Y, et al. Rethinking bottleneck structure for efficient mobile network design [C]//Computer Vision-ECCV 2020. Cham: Springer International Publishing, 2020: 680–697.
- [16] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 4510–4520.
- [17] XU K, HUANG H, LI Y, et al. Multilayer feature fusion network for scene classification in remote sensing[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(11): 1894–1898.
- [18] SUN H, LI S, ZHENG X, et al. Remote sensing scene classification by gated bidirectional network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 58(1): 82–96.
- [19] LUKASIK M, BHOJANAPALLI S, MENON A, et al. Does label smoothing mitigate label noise?[C]//International Conference on Machine Learning. PMLR, 2020: 6448–6458.
- [20] YANG Y, NEWSAM S. Bag-of-visual-words and spatial extensions for land-use classification[C]//Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information systems. ACM, 2010: 270–279.
- [21] ZHU Q, ZHONG Y, ZHAO B, et al. Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(6): 747–751.
- [22] 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.
- [23] LIU B D, MENG J, XIE W Y, et al. Weighted spatial pyramid matching collaborative representation for remote-sensing-image scene classification[J]. Remote Sensing, 2019, 11(5): 518–527.
- [24] CHENG G, YANG C, YAO X, et al. When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(5): 2811–2821.
- [25] GONG X, XIE Z, LIU Y, et al. Deep salient feature based anti-noise transfer network for scene classification of remote sensing imagery[J]. Remote Sensing, 2018, 10(3): 410.
- [26] LI B, SU W, WU H, et al. Aggregated deep fisher feature for V-HR remote sensing scene classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(9): 3508–3520.
- [27] HE N, FANG L, LI S, et al. Skip-connected covariance network for remote sensing scene classification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 31(5): 1461–1474.
- [28] 余东行, 张保明, 赵传, 等. 联合卷积神经网络与集成学习的遥感影像场景分类[J]. 遥感学报, 2020, 24(6): 717–727.
YU Dong-xing, ZHANG Bao-ming, GUO Hai-tao, et al. Scene classification of remote sensing image using ensemble convolutional neural network[J]. Journal of Remote Sensing (Chinese), 2020, 24(6): 717–727.
- [29] LIU Y, ZHONG Y, FEI F, et al. Scene classification based on a deep random-scale stretched convolutional neural network[J]. Remote Sensing, 2018, 10(3): 444–453.
- [30] LIU X, ZHOU Y, ZHAO J, et al. Siamese convolutional neural networks for remote sensing scene classification[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(8): 1200–1204.
- [31] YUAN B, LI S, LI N. Multiscale deep features learning for land-use scene recognition[J]. Journal of Applied Remote Sensing, 2018, 12(1): 015010–015022.

责任编辑:李翠薇