

基于多尺度特征融合轻量化网络的遥感影像场景分类

赵云飞¹, 李文², 马颂尧²

(1. 昆明理工大学 国土资源工程学院, 云南 昆明 650093; 2. 中国人民武装警察部队士官学校, 浙江 杭州 311403)

摘要: 针对现有遥感影像场景分类模型复杂度越来越高及特征提取冗余的问题, 提出一种多尺度特征融合轻量化网络MSA-ResNet。首先, 结合经典的ResNet模型与Swin-Transformer技术, 利用滑动窗口多头自注意力机制、多级特征堆叠与跨层连接策略, 强化模型对空间关系的理解和多尺度特征的融合能力; 其次, 在网络残差块内简化卷积模块、引入标签平滑策略, 显著降低模型参数量, 保持特征提取效率, 提升整体分类性能。在AID数据集上的实验表明, 所提模型的总体分类精度为92.97%, 相较于MSCP、SAFF、APDC、PSGAN等网络分别提升7.39%、1.14%、0.82%、0.65%, 证实了多尺度融合方式能在主干特征提取网络能力有限的情况下, 提升模型特征提取能力、特征效果和分类精度。

关键词: 遥感影像; 场景分类; 多头自注意力; 多尺度特征融合

DOI: 10.11907/rjdk.241415

中图分类号: TP311

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2025)006-0199-06



Remote Sensing Image Scene Classification Using a Lightweight Network with Multi-scale Feature Fusion

ZHAO Yunfei¹, LI Wen², MA Songyao²

(1. Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming 650093, China;

2. Chinese People's Armed Police Force Non commissioned Officer School, Hangzhou 311403, China)

Abstract: A multi-scale feature fusion lightweight network MSA ResNet is proposed to address the increasing complexity and redundant feature extraction of existing remote sensing image scene classification models. Firstly, by combining the classic ResNet model with Swin Transformer technology, the sliding window multi head self attention mechanism, multi-level feature stacking, and cross layer connection strategy are utilized to enhance the model's understanding of spatial relationships and its ability to fuse multi-scale features; Secondly, simplifying the convolution module within the residual blocks of the network and introducing label smoothing strategies significantly reduce the number of model parameters, maintain feature extraction efficiency, and improve overall classification performance. Experiments on the AID dataset show that the overall classification accuracy of the proposed model is 92.97%, which is 7.39%, 1.14%, 0.82%, and 0.65% higher than networks such as MSCP, SAFF, APDC, and PSGAN, respectively. This confirms that the multi-scale fusion method can improve the model's feature extraction ability, feature performance, and classification accuracy when the backbone feature extraction network has limited capabilities.

Key Words: remote sensing imagery; scene classification; multi-head self-attention; multi-scale feature fusion

0 引言

随着遥感技术发展, 遥感图像分辨率不断上升, 图像

包含的地物特征和空间信息将更丰富, 提升遥感场景分类准确率有助于提升农业检测准确率、土地规划效率及资源勘探能力, 但也给特征信息提取及遥感场景分类的准确性带来了挑战^[1-3]。

收稿日期: 2024-04-30

扫描二维码阅读全文:



作者简介: 赵云飞(1996-), 男, 昆明理工大学国土资源工程学院硕士研究生, 研究方向为遥感影像智能解译; 李文(1998-), 男, 硕士, 中国人民武装警察部队士官学校助教, 研究方向为遥感影像处理与分析; 马颂尧(1998-), 男, 中国人民武装警察部队士官学校助教, 研究方向为无线通信。

目前,遥感影像场景分类方法可分为基于模型驱动和数据驱动。支持向量机、随机森林等传统基于模型驱动的方法极其依赖先验知识,模型准确率和鲁棒性较差^[4,5]。随着神经网络兴起,基于数据驱动的遥感图像场景分类方法逐渐成为主流,最典型的是卷积神经网络(Convolutional Neural Network, CNN)^[6]。相较于传统方法, CNN 类方法能提取图像分层特征并自主学习图像特征,但受限于卷积算法的感受野,难以对高分辨率的复杂遥感图像进行全局建模。

国内外学者提出一系列方法提升遥感图像场景分类效果,较为常见的方法包括添加注意力机制、长短时记忆网络、Transformer 等^[7-9]。模型在大多数工作中会重新训练,未考虑数据的有限性和训练代价。为此, Li 等^[10]针对场景内分类问题,提出基于预训练 CNN 模型的多层特征融合策略。首先,通过预训练 CNN 模型初步提取特征;其次,将全连接层特征与通过 Fisher 核编码提取的中级特征表示融合;最后,利用谱回归判别分析法进行分类。然而上述方法计算成本较高,且不同预训练模型融合时可能会造成负面影响。

为了解决上述问题,许多学者设计了轻量化模型来降低模型计算代价。Hu 等^[11]基于二进制特征描述提出一种快速二进制编码方法。首先,通过卷积获取特征图;其次,利用哈希计算将二进制特征图转换为整数特征图;最后,将统计直方图作为图像场景的全局特征表示。同时,注意力机制可有效帮助模型抑制噪声、学习图像特征。时永欣等^[12]在模型中融合多尺度注意力并将其应用于多视角场

景分类。首先,设计两个网络分支提取地面和航空图像特征;其次,通过注意力模块学习特征通道的关联性,融合不同尺度的注意力;最后,使用支持向量机进行场景分类。

然而,基于 CNN 的模型通常难以获取长距离依赖关系。为此, Zhang 等^[13]设计了一种基于 Transformer 与 CNN 的模型 TRS, 首先将 BottleNeck 中的 3×3 卷积核替换为 MH-SA 层,其次通过卷积计算输入 Transformer 编码器,但计算复杂度较高^[14]。

为了解决上述遥感影像分类方面存在的问题,提出一种改进型 ResNet 网络结构——MSA-ResNet 网络模型。该模型融合了 Swin-Transformer 的滑动窗口多头自注意力机制,旨在复杂背景下提升高分辨率遥感影像场景的分类准确度。

1 本文方法

MSA-ResNet 网络模型由影像主干特征提取部分、多尺度特征融合部分及分类部分构成,如图 1 所示。首先,使用多级特征堆叠融合方式,在保证模型主干网络轻量化的同时,在一定程度上增强特征表征能力;其次,引入 Swin-Transformer 的高效自注意力模块,改造 ResNet 传统的卷积模块,以增强模型对空间关系的捕捉能力;再次,优化模型参数量、简化残差块内的卷积结构,在提升模型特征提取效率的同时抑制复杂度;接下来,通过跨层连接结构更有效地融合多尺度特征信息,提升分类任务的整体精度;最后,引入标签平滑提升整体的分类准确度。

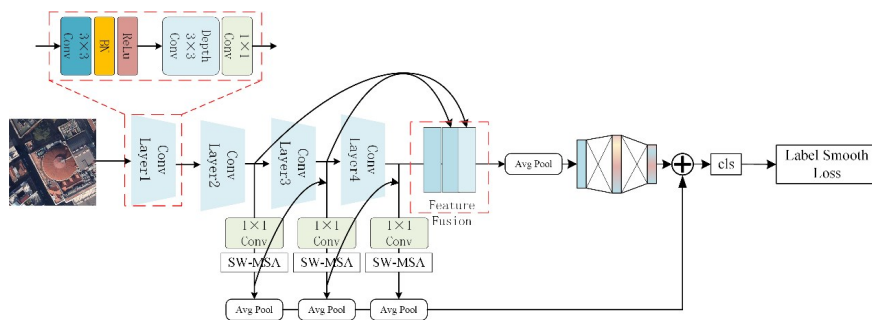


Fig. 1 MSA-ResNet network overall structure

图1 MSA-ResNet网络整体结构

1.1 主干特征提取模块

ResNet 网络模型的残差连接不仅能避免训练过程中的梯度消失和梯度爆炸等问题,还能使网络快速收敛,增强特征表达能力。考虑到网络的轻量化与特征提取准确度,本文基于 ResNet 网络模型改进得到主干网络,以更好地提取所需的多尺度特征。

在神经网络中,底层特征通常包含更多局部信息,高层特征则具有更抽象和更全局的语义信息。对于场景分类任务,局部信息可能并不十分重要,但全局信息能帮助网络进行准确分类。因此,本文仅使用更具有表征能力的后 3 层特征完成分类任务,以轻量化网络结构提高效率。

同时, ResNet 架构由多个残差块组成,这些残差块通常包含卷积层、批量归一化层和激活函数层。考虑到单一特征提取结果难以达到较好的分类效果,本文使用主干网络提取多个特征,将 ResNet 后 3 个残差模块提取的特征作为多尺度特征进行使用。

1.2 多尺度特征融合模块

在遥感影像场景分类任务中,单一特征的信息表达能力有限,无法完全捕捉影像中的复杂语义和结构信息,但多尺度特征融合可利用不同尺度下的特征信息,捕捉影像中不同层次的语义和结构信息,提升模型对场景的理解能力和鲁棒性。

多尺度特征融合模型利用跨层连接方式依次增强各层特征,在最后的特征提取层准确提取更有效的特征,以增加最终融合特征的表达。然而,简单利用多尺度特征容易导致模型在不同尺度下提取多余且错误的特征。多尺度特征融合模块的结构如图2所示。

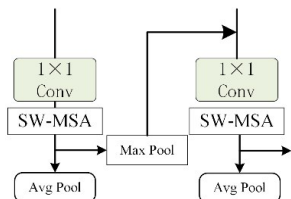


Fig. 2 Multi-scale feature fusion module structure

图2 多尺度特征融合模块结构

为了弥补主干网络特征提取能力弱的问题,本文采用了多级特征堆叠方式。首先,在主干特征提取后将多尺度特征分别进行堆叠;其次,将特征输入至两层 1×1 卷积进行特征融合,以获取更好的特征表现;最后,为了保证特征提取的准确性,利用基于滑动窗口的自注意力机制SW-MSA(Sliding Window Based Multi-head Self-Attention)对多尺度特征进行再提取,采用跨层连接下层与上层特征使特征提取更准确,在一定程度上减少周围背景特征对分类的影响。具体的数学表达式如式(1)所示。

$$M_i = \text{Attention}\{\text{Conv}(F_i + M_{i-1})\} \quad (1)$$

式中: F_i 为主干特征网络获取的特征; $\text{Conv}(\cdot)$ 为 1×1 卷积操作; $\text{Attention}(\cdot)$ 为滑动窗口多头自注意力模块; M_i 表示输出结果。

SW-MSA将图片划分为 $M \times M$ 的窗口,在每个窗口内分别计算注意力,相较于Transformer基于全局计算注意力方法而言,将注意力计算限制在每个窗口内以降低模型的计算复杂度,并增强模型对图像全局与局部信息的交互能力。具体计算过程为:首先,将影像 $P \in \mathbf{R}^{H \times W \times C}$ 划分为多个窗口,得到每个窗口的嵌入向量 $\mathbf{x}$,形状为 (N, d) ;其次,通过线性变换得到 Q (Query向量)、 K (Key向量)、 V (Value向量),通过Query向量与Key向量计算注意力特征图;最后,与Value向量相乘得到注意力。

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d} + B}\right) \times V \quad (2)$$

式中: K^T 为 K 的转置矩阵; $B \in \mathbf{R}^{M^2 \times M^2}$ 为相对位置偏置; d 表示 Q/K 的维度大小; M^2 表示切分影像窗口的大小。

1.3 分类模块

分类部分的操作为:将两个特征向量加和后,输入至分类器中得到分类结果。遥感图像通常在类别之间具有细微差异与较高的内部类别变异性,可导致模型在训练集上过于自信,尤其对小型或轻量级网络而言,无法达到深层网络那样利用多层卷积与激活等方式增加模型的非线性能力。

由于输出更平滑的概率分布可增加模型在未知数据

上的泛化能力,因此在计算损失函数中引入标签平滑方法,使用具有标签平滑后的交叉熵损失函数。具体计算公式如式(3)、式(4)所示。

$$V = V_1 + V_2 \quad (3)$$

$$\text{Loss} = -\sum_{i=1}^n y(x_i) \log(p(x_i)) * (1 - \varepsilon) - \frac{1}{n} \sum_{i=1}^n \log(x_i) * \varepsilon \quad (4)$$

式中: V_1 、 V_2 分别为主干特征提取的特征向量和经过多尺度特征融合的特征向量; $y(x_i)$ 为 x_i 真实标签; $p(x_i)$ 为 x_i 模型预测标签; ε 为平滑参数; n 为样本数。

2 实验结果与分析

2.1 实验环境

实验使用Adam优化器,权重衰减设置为0.001,利用余弦退火学习率调整策略动态调整学习率,并使用标签平滑交叉熵函数作为损失函数,将迭代次数设置为300,批处理大小设置为32,学习率设置为0.000 1。本文实验环境如表1所示。

Table 1 Experimental environment

表1 实验环境

操作系统	CPU/GHz	内存/GB	GPU/GB	深度学习框架
Ubuntu 20.04.3 LTS	I7-9750H 2.60	10	RTX3080 10	Pytorch v1.9

2.2 数据集

为了验证所提模型的性能与有效性,选取AID(Aerial Image Dataset)、NWPU45(NWPU-RESISC45 Dataset)两个公共数据库进行实验,如图3所示。

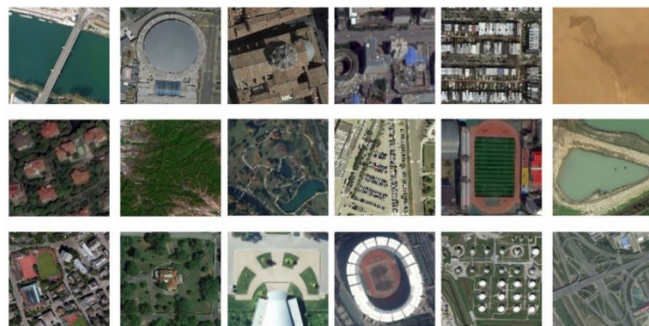


Fig. 3 Image samples in AID dataset

图3 AID数据集中的图像样本

AID数据集选自武汉大学和华中科技大学在2017年联合发布的遥感场景数据集,包括约10 000张来自于全世界各个国家和地区的影像,涵盖了30种不同的场景类别,每类包含220~420幅影像,每幅影像尺寸为600×600,拥有0.5~8 m的不同空间分辨率;NWPU45数据集选自西北工业大学创建的远程传感图像分类的公共基准数据集,聚焦于飞机、教堂、沙漠等45个类别,每个类别都有1 000张影像,拥有0.2~30 m不同空间分辨率,收录了超过100个国家和地区的图像,总计约有31 500张。需要注意的是,AID数据集主要用于网络模型的特征学习和验证测试,NWPU45数据集主要用于验证模型的泛化能力及鲁棒性。

2.3 评价指标

本文使用总体分类准确率(Overall Accuracy, OA)和混淆矩阵(ConfusionMatrix, CM)衡量所提方法的性能。其中,总体分类准确率是将正确分类的样本数量除以测试集中所有样本总数得到的比值,从宏观上体现分类模型的精确度,具体计算公式为:

$$OA = \frac{S}{N} \times 100\% \quad (5)$$

式中: S 为测试集中正确分类的样本数; N 为测试集的总样本数。

混淆矩阵以可视化形式展示不同类别之间的分类结果,每一类正确分类的数量由矩阵的对角线元素表示,非对角线元素则显示分类错误,并指出错误类别的具体类型。在混淆矩阵中,元素数值大小通过明暗不同的颜色表示,数值越大颜色越亮,以增强结果的直观性和辨识度。

2.4 实验结果

为了验证本文方法的有效性,在AID数据集、NWPU45数据集上进行一系列实验,以分析本文方法与现有典型分类模型的分类准确度与性能差异。首先,设定两个数据集的训练集比例分别为50%和20%;其次,比较本文方法在不同学习率下对模型性能的影响;再次,评估不同注意力机制模块对改进网络模型的影响;最后,将本文方法与已有方法进行比较。

实验算法包括:基础的VGG_VD16网络^[15];具有多尺度特征融合结构的MSCP网络模型^[16];DCCNN、APDC-Net等多分支网络结构;现阶段遥感影像场景分类的主流轻量化网络模型——PSGAN网络模型与LiT网络模型^[17-20]。

2.4.1 不同学习率

在进行遥感影像分类任务时,选择合适的批量大小和学习率对提升模型分类准确度影响重大。通常情况下,可运用学习率调度器在整个训练过程中获得最佳学习率,增强网络分类性能。本文采用余弦退火、固定步长与线性衰减调度器等学习率调整策略,设置不同的批量大小(64、32、16、8和4)在AID数据集上进行实验,实验结果如表2所示。

Table 2 Classification effect of different scheduling strategies

表2 不同调度策略的分类效果

批量大小	调度策略/%		
	余弦退火	固定步长	线性衰减
64	92.97	92.77	92.44
32	92.88	92.72	92.23
16	92.51	92.69	92.11
8	92.33	92.49	91.82
4	92.12	92.05	90.88

注:加粗数字代表同一批量大小下的最优分类结果。

由表2可知,在不同批量大小(64、32、16、8和4)下经过300次训练后,余弦退火调度器在分类精度上取得了较好的分类精度。当批量大小为16、8时,固定步长调度器的分类精度优于余弦退火,因为余弦退火调度器在处理较小批量时不够稳定,且固定步长调度器是按照预设的步

骤调整学习率,所以在较小的批量大小时表现更好。换言之,在较小的批量大小时梯度估计更不稳定,而固定步长调度器可保持固定的学习率,不会因批量大小变化而导致分类效果不稳定。

综上,较大的批量能提供更稳定的梯度估计,意味着模型在训练过程中能尽可能地避免训练的随机性和不稳定性。同时,较大的批量能更好地代表数据集的整体特征,使模型在面对未知样本时表现更好。此外,相较于其他学习率调度策略,余弦退火策略能更平滑地调整学习率,使模型在训练过程中避免局部最优解,从而在一定程度上提升模型的训练效果及泛化能力。需要注意的是,模型的表现会受到批量大小等参数影响。

2.4.2 不同注意力模块

为了探究不同注意力机制在遥感影像场景分类及特征再提取方面的有效性。在模型中添加SE模块和CBAM(Convolutional Block Attention Module)模块,使用通道增强、空间与通道混合注意力来提升模型的特征提取能力^[21,22]。本文将两种注意力机制替换网络模型中的SW-MSA模块,在两个数据集上分析不同注意力模块对网络模型分类效果的影响,实验结果如表3所示。

Table 3 Classification effectiveness of different attention mechanisms

表3 不同注意力机制的分类效果

注意力类别	AID数据集上的	NWPU45数据集上的
	分类精度/%	分类精度/%
×	89.07	85.21
SE	91.22	86.78
CBAM	91.97	87.12
SW-MSA	92.97	88.01

注:加粗数字代表同一数据集下的最优分类结果。

由表3可知,本文模型的注意力模块十分必要,仅使用通道注意力的SE模块都相较于不使用任何注意力时提升2.15%。在轻量级网络的情况下,注意力能极大程度增加模型特征提取能力,提升分类精度。CBAM模块相较于SE模块进行了更深入的集成和优化,混合注意力效果优于仅使用一种注意力方法,在AID、NWPU45数据集上的分类精度分别提升0.75%、0.34%。

随着注意力机制的增强和模块复合化,模型能更精细地调节特征提取焦点,不仅能识别出更具区分性的特征,还能在更广阔的上下文中有用地整合不同的尺度信息。

2.4.3 不同网络模型

选取8种典型的遥感影像分类网络模型与本文方法进行比较,实验如表4所示。在两个数据集中,MSA-ResNet相较于基准模型VGG_VD16的准确度分别提升3.33%、8.22%;相较于利用多尺度特征堆叠的MSCP模型提升0.61%、2.93%。SAFF模型在AID数据集中取得了较好效果,但SW-MSA的多头自注意力模块在面对更复杂的数据集时整体分类效果优于SAFF。DTDCNN模型虽使用了可变卷积来增强特征提取准确度,但整体分类效果不佳。

Table 4 Classification performance of different network models		
表 4 不同网络模型的分类效果		
网络模型	AID 数据集上的分 类精度/%	NWPU45 数据集上的 分类精度/%
VGG_VD16	89.64	79.79
MSCP	92.36	85.58
DCCNN	91.49	85.63
APDC-Net	92.15	87.84
SAFF	91.83	84.00
DTDCNN	89.26	84.12
PSGAN	92.17	88.47
LiT	93.25	83.58
MSA-ResNet	92.97	88.01

注:加粗数字代表同一数据集下的最优分类结果。

在 AID 数据集中,MSA-ResNet 相较于 DCCNN、APDC-Net、SAFF 模型、PSGAN 的分类准确度分别提升 1.48%、0.82%、1.14%、0.65%,但略逊于 LiT 网络。在较复杂的 NWPU45 数据集中,MSA-ResNet 的分类精度优于 DCCNN、APDC-Net 和 SAFF 模型,但略逊于 PSGAN 模型。

图 4 为网络模型在 AID 数据集上的混淆矩阵。由此可见,改进网络模型虽在整体上实现了准确分类,但在部分类别上存在混分类情况。这类混分类别遥感影像的背景复杂,目标地物具有多样性,存在明显的类间相似和类内多样特性。例如,Park、Resort 类别存在相似的背景地类,还发现一些类别被错误分类为 Resort,这也是轻量级网络难以避免的问题。

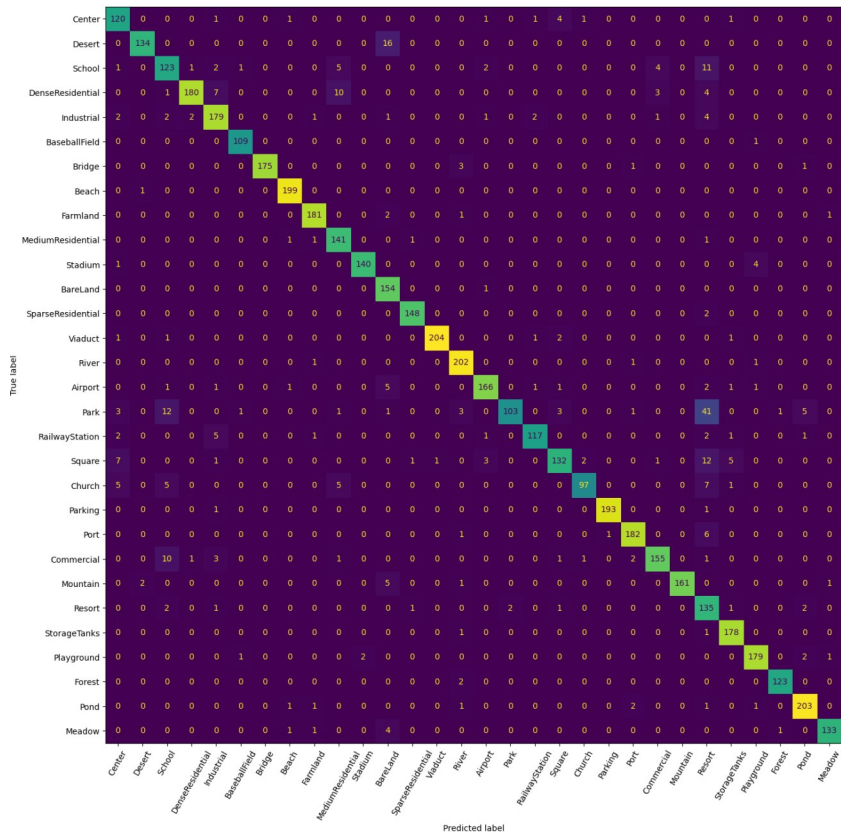


Fig. 4 Confusion matrix of MSA-ResNet network model on AID dataset
图 4 MSA-ResNet 网络模型在 AID 数据集上的混淆矩阵

虽然使用两种注意力模块都取得了较好的结果,但在 AID、NWPU45 数据集上的分类效果表现均逊色于 SW-MSA 模块,具体结果如表 3 所示。Transformer 的多头自注意力模块相较于两种常规注意力模块取得了更优的分类效果,因此在考虑参数数量的情况下,注意力模块的选取至关重要。不同网络模型的参数量如表 5 所示。

由此可知,MSA-ResNet 在精度上优于其他网络、参数量也更低,模型在保证较低复杂度的情况下能满足一定需求的分类精度。因此,综合考虑精度和参数数量的 MSA-ResNet 网络模型具有一定的可用性。

总而言之,PSGAN、LiT 模型与本文模型的分类精度基本持平,其余 6 种模型的表现较差。MSA-ResNet 模型在两

Table 5 Parameter quantities of different network models	
表 5 不同网络模型的参数量	
网络模型	参数量/million
VGG_VD16	138.0
MSCP	60.0
DCCNN	0.6
APDC-Net	0.6
SAFF	60.0
DTDCNN	23.5
PSGAN	25.6
LiT	39.9
MSA-ResNet	3.7

个数据集上都表现出较优的分类精度,表明在多尺度特征融合中增加注意力模块及利用连接方式来增强模型特征

提取的效果较好。在网络参数量方面,MSA-ResNet 模型能在参数量大幅降低的情况下取得较优的遥感影像分类效果。同时,MSCP、DCCNN 等模型相较于基准模型 VGG_VD16 具有更高效的特征提取能力,说明改进网络模型的架构设计,特别是将注意力机制与特征提取融入网络可显著提升模型对遥感图像的分类性能。

2.5 消融实验

MSA-ResNet 网络模型添加了 SW-MSA 模块与跨层连接方式,在 NWPU45 数据集上进行消融实验以验证各模块的有效性,实验结果如表 6 所示。

Table 6 Result of ablation experiment of MSA-ResNet

network model

表 6 MSA-ResNet 网络模型的消融实验结果

多级堆叠	SW-MSA	跨层连接	运行时间/s	OA/%
×	×	×	7.82	83.31
√	×	×	8.78	84.57
×	√	×	11.54	86.13
×	×	√	8.11	85.31
√	√	×	11.78	87.24
√	×	√	9.79	86.58
×	√	√	11.84	87.71
√	√	√	12.56	88.01

注:加粗数字代表最优值。

由表 6 可知,多级堆叠、SW-MSA 和跨层连接结构能提升模型性能,SW-MSA 模块对提升模型整体精度具有显著的积极作用。在不使用 3 个模块时,轻量型网络的特征提取能力较弱,无法取得较好的结果,当增加多级堆叠与跨层连接后精度提升 3.27%;使用 SW-MSA 后模型分类精度更优,虽然增加了部分训练时间但精度提升明显;当结合 3 种结构后,模型的整体分类精度进一步得到了提升。

综上,消融实验证明了结合多级堆叠、跨层结构和 SW-MSA 能有效提升模型的特征提取能力。

3 结语

本文提出一种多尺度特征融合轻量化网络 MSA-ResNet,将 Swin-Transformer 优秀的滑动窗口多头自注意力机制模块引入网络,还考虑了模型的复杂度。首先,通过改进 ResNet 残差块内的卷积模块,在保证特征提取能力的同时减少参数量,利用多级特征堆叠方式增强特征;其次,采用跨层连接关联多尺度特征,进一步提取多尺度特征来提升整体分类精度。

实验表明,本文网络在两个大型公开的数据集上能达到不错的精度,但还需要不断提升。后续,将考虑如何在增加少量参数量的情况下提升模型性能。

参考文献:

- [1] XU C, ZHU G, SHU J. Robust joint representation of intrinsic mean and kernel function of lie group for remote sensing scene classification[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 18(2): 796-800.
- [2] XU S J, OUYANG P Y, GUO X Y, et al. Multi-scale feature fusion dilated convolution ResNet remote sensing image building segmentation[J].

Optics and Precision Engineering, 2020, 28(7): 1588-1599.

徐胜军, 欧阳朴衍, 郭学源, 等. 多尺度特征融合空洞卷积 ResNet 遥感图像建筑物分割[J]. 光学精密工程, 2020, 28(7): 1588-1599.

- [3] CALVIN W M, LITTLEFIELD E F, KRATT C. Remote sensing of geothermal-related minerals for resource exploration in Nevada[J]. Geothermics, 2015, 53(3): 517-526.
- [4] NIU X, BAN Y. Multitemporal RADARSAT-2 polarimetric SAR data for urban land cover classification using support vector machine[J]. International Journal of Remote Sensing, 2013, 34(1): 1-26.
- [5] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [6] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(12): 2278-2324.
- [7] TONG W, CHEN W, HAN W, et al. Channel-attention-based DENSENet network for remote sensing image scene classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13(5): 4121-4132.
- [8] WANG Q, LIU S, CHANUSSOT J, et al. Scene classification with recurrent attention of VHR remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(2): 1155-1167.
- [9] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[DB/OL]. <https://arxiv.org/abs/1706.03762>.
- [10] LI E, XIA J, DU P, et al. Integrating multilayer features of convolutional neural networks for remote sensing scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 13(6): 411-416.
- [11] HU F, XIA G S, HU J, et al. Fast binary coding for the scene classification of high-resolution remote sensing imagery[J]. Remote Sensing, 2016, 8(7): 555-559.
- [12] SHI Y X, ZHOU W X, SHAO Z F. Multi-view remote sensing image scene classification integrating multiscale attention[J]. Journal of Wuhan University (Information Science Edition), 2024, 49(3): 366-375. 时永欣, 周维勋, 邵振峰. 融合多尺度注意力的多视角遥感影像场景分类[J]. 武汉大学学报(信息科学版), 2024, 49(3): 366-375.
- [13] ZHANG J, ZHAO H, LI J. TRS: transformers for remote sensing scene classification[J]. Remote Sensing, 2021, 13(6): 41-43.
- [14] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[DB/OL]. <https://arxiv.org/abs/1704.04861>.
- [15] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[DB/OL]. <https://arxiv.org/abs/1409.1556>.
- [16] NAN J H, FANG L, LI S, et al. Remote sensing scene classification using multilayer stacked covariance pooling[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 35(8): 1-12.
- [17] BI Q, QIN K, LI Z, et al. Multiple instance dense connected convolution neural network for aerial image scene classification[C]// IEEE International Conference on Image Processing, 2019: 2501-2505.
- [18] BI Q, QIN K, ZHANG H, et al. APDC-Net: attention pooling-based convolutional network for aerial scene classification[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 18(7): 1-5.
- [19] CHENG G, SUN X, LI K, et al. Perturbation-seeking generative adversarial networks: a defense framework for remote sensing image scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60(4): 1-11.
- [20] ZHANG X, SU Y, GAO L, et al. A lightweight transformer network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 128(61): 5517-5529.
- [21] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [22] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[DB/OL]. <https://arxiv.org/abs/1807.06521>.

(责任编辑:刘嘉文)