引文格式:李熙瑶,李家艺,丁建勋,等.基于改进残差结构的空-谱融合轻量网络用于大范围农作物分类[J].武汉大学学报(信息科学版),2025,50(5):937-948.DOI:10.13203/j.whugis20230008

Citation: LI Xiyao, LI Jiayi, DING Jianxun, et al. Spatial-Spectral Fusion Lightweight Network Based on Improved Residual Structure for Large-Scale Crop Classification[J]. Geomatics and Information Science of Wuhan University, 2025, 50(5):937-948. DOI:10.13203/j.whugis20230008

基于改进残差结构的空-谱融合轻量网络用于大范围农作物分类

李熙瑶[1] 李家艺[1,2] 丁建勋[1] 敦力民[3] 于笑雨[3]
彭 敏[3] 黄 昕[1,4]

1 武汉大学遥感信息工程学院,湖北 武汉,430079
2 湖北珞珈实验室,湖北 武汉,430079
3 沈阳市勘察测绘研究院有限公司,辽宁 沈阳,110004
4 武汉大学测绘遥感信息工程全国重点实验室,湖北 武汉,430079

摘 要:现有高光谱图像分类网络在小范围、高空间分辨率的基准数据集测试时常采用大量的网络参数拟合数据特征,使其在分类方面展示出良好性能。然而实际生产任务中使用的卫星高光谱数据通常分辨率较低、覆盖范围大,过多的网络参数容易导致过拟合问题。针对上述问题,提出了一种新的轻量化卷积网络,并使用具有32波段、10 m分辨率的珠海一号高光谱影像对中国沈阳市进行主要农作物(玉米、水稻)的提取。该网络分别针对影像的空间特性和光谱特性设计网络组件,再设计自适应融合的模块组件来综合空间特征与光谱特征,实现高光谱影像分类。其中,空间特征学习模块设计改进多尺度残差映射块,可以在尺度各异的多分支卷积核中自适应地检测空间特征,再进行基于注意力机制的空间特征优化。光谱特征学习模块则考虑图像中丰富的光谱信息,引入时序卷积网络的思想,通过在光谱维度的连续卷积块学习图像的精细光谱特征。这种端到端的分类网络通过卷积同时提取空间信息与光谱信息,从而实现精确推理。该方法在测试集上的总体精度和平均交并比分别为92.3%和77.9%,玉米、水稻和其他类的单类$F1$分数分别达到85.1%、81.95%和94.9%,优于所有对比网络。通过在珠海一号高光谱影像数据集上的实验证明,所提方法比其他现有网络取得了更好的分类结果。这种端到端的轻量级分类网络通过不同的卷积块同时提取空间和频谱信息,有效地利用了图像的空间上下文信息和光谱通道之间的相关性。

关键词:深度学习;高光谱影像;农作物分类;残差结构;空-谱融合

中图分类号:P237 **文献标识码**:A **收稿日期**:2023-05-25

DOI:10.13203/j.whugis20230008 **文章编号**:1671-8860(2025)05-0937-12

Spatial-Spectral Fusion Lightweight Network Based on Improved Residual Structure for Large-Scale Crop Classification

LI Xiyao[1] *LI Jiayi*[1,2] *DING Jianxun*[1] *DUN Limin*[3] *YU Xiaoyu*[3] *PENG Min*[3] *HUANG Xin*[1,4]

1 School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
2 Hubei Luojia Laboratory, Wuhan 430079, China
3 Shenyang Geotechnical Investigation & Surveying Research Institute Co. Ltd, Shenyang 110004, China
4 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

Abstract: **Objectives**: Deep learning methods with massive parameters have shown good performance in classifying hyperspectral image with small size and high resolution. Meanwhile, there are at risk of overfitting and difficulty in training, due to the low resolution and large size (i.e., diversity) of satellite hyperspectral data in practical applications. Thus, we proposes a new lightweight convolutional network to extract

基金项目:国家重点研发计划(2022YFB3903701);国家自然科学基金(42271328, 42071311)。

第一作者:李熙瑶,硕士,主要研究方向为基于深度学习的遥感信息提取。Lixiyao99@whu.edu.cn

通信作者:李家艺,副教授。zjjerica@whu.edu.cn

the main crops (corn, rice) in Shenyang city, China, by use of orbita hyperspectral satellite (OHS) images with 32 spectral bands ranging from 0.4-1.0 μm and 10-meters spatial resolution. **Methods:** First, an improved multi-scale residual mapping blocks and a spatial feature optimization block based on attention mechanism are designed to formulate the spatial feature learning module. Then, spectral feature learning module with three successive 1D convolution blocks is conducted to mine the rich spectra information of hyperspectral images. Finally, a fusion classification module is built to classify pixels by integrating weighted spatial features and spectral features. **Results:** Four recent networks including PSPNet, U-Net, MAP-Net and FreeNet are considered for comparison. From the perspective of model size, the proposed network is much lighter than PSPNet, UNet, MAP-Net and equal to FreeNet, which is benefited from the use of 1D convolution and 2D convolution in two learning modules that avoid excessive parameters caused by 3D convolution. From the perspective of accuracy, considering the overall assessment metrics and the $F1$-score of each category, the proposed method achieves the best performance over the recent four networks. The overall accuracy and mean intersection over union of our method on the test set are 92.3% and 77.9%, respectively, which is superior to other networks. Our method also has a better classification result in single category, the $F1$-score of corn, rice and other each reach 85.1%, 81.95% and 94.9%. Expecially for rice, the $F1$-score of our method is 11.5%, 4.2%, 13.6% and 12.0% higher than PSPNet, U-Net, MAP-Net and FreeNet respectively. **Conclusions:** Experiments on the data set show that our method achieves better classification results than other existing networks. This end-to-end lightweight classification network extracts spatial and spectral information simultaneously by different convolutional blocks. The idea of combining spatial and spectral features can effectively utilize the spatial context information of images and the correlation among spectral channels to improve classification accuracy.

Key words: deep learning; hyperspectral images; crop classification; residual structure; spatial-spectral fusion

高光谱遥感是遥感领域的一项重要前沿技术,结合先进的成像技术和光谱技术,高光谱图像(hyperspectral image,HSI)能够提供丰富的地表光谱信息。HSI在地物类型的准确识别与分类方面具有广泛的应用,如地质矿产定量反演[1]、农作物精细分类[2]、森林病虫害监测[3]、环境监测[4]等。高光谱图像分类的目标是为高光谱图像中的每个像素分配一个唯一的语义标签,是高光谱遥感中一个基本但具有挑战性的任务。

深度网络具有自动学习特征的强大能力,近年来在多个领域与应用中展现出优异的性能,推动了其在图像分类[5-6]、变化检测[7-8]、自然语言处理[9-10]等方面的应用。在高光谱遥感图像分类领域,深度学习方法也已被广泛应用。为了挖掘丰富的光谱信息,一维卷积[11-12]和递归神经网络[13-14]被应用于从序列光谱特征中学习判别特征。但基于像素的输入忽略了相邻像素之间的关系,无法有效利用影像的空间信息[15]。因此,一些研究者尝试对高光谱图像的局部空间邻域信息或全局空间信息进行保留,通过二维卷积等方法提取重要的空间特征,与光谱特征联合分类[15-16]。二维卷积常应用于语义分割任务中,其中一些经典的网络,如基于UNet[17]、PSPNet[18]等网络的方法[19-20]也被应用于高光谱影像分类中。二维卷积在语义分割任务中的成熟应用使其在高光谱影像的空间特征提取方面具有较大的优势,如残差结构的使用可以有效改善深度网络中梯度消失的问题[21];不同卷积核获得的多尺度特征可以提高模型在小样本数据集上的表达能力[22];编解码结构可以考虑到全局空间信息,实现端到端推理[23]等。另外,除分别获取光谱特征和空间特征最后进行联合分类的方法外,三维卷积常被用来同时挖掘高光谱立方体的空谱信息[24-27]。但其在光谱域与空间域同时进行卷积的操作会导致模型的计算复杂度与参数量大大增加[16],容易造成模型过拟合。上述对高光谱影像分类的探索大多在公开的小范围基准数据集上进行实验与测试,如印第安纳州西北部农场数据集(145×1 145,20 m)、帕维亚大学数据集(610×340,1.3 m)和萨利纳斯数据集(512×217,3.7 m)等。在这些数据集中,网络可以通过大量的参数来拟合数据特征,通常能够获得非常高的分类精度[22-25]。但实际生产任务中,在使用大范围、较低空间分辨率的高光谱影像进行分类时精度有时偏低[11,28]。除了分辨率的影响,这也可能和网络复杂度高、参数量多有关,使得在训练大量样本

时较为困难。因此,面向大范围的高光谱影像深度学习方法分类,设计同时考虑到光谱和空间特征的轻量化网络是十分必要的。本文使用珠海一号高光谱影像为研究数据,其空间分辨率适中、光谱信息丰富、数据量充足,基于此提出一种基于多分支卷积的空-谱融合高光谱分类网络用于中国沈阳市两类农作物玉米、一季稻的高光谱遥感影像分类。

1　方法介绍

1.1　基本框架

本文所提出的基于卷积的高光谱分类网络结构如图1所示。该模型采用分别获取特征的思路,主要由3部分组成,分别为空间特征学习模块、光谱特征学习模块以及融合分类模块。其中空间特征学习模块仅通过9个二维卷积核提取空间特征;光谱特征学习模块通过三层一维卷积提取特征,在保留主要光谱信息的同时降低了计算量。该模型可以同时保留较为完整的光谱信息和空间信息,避免了三维卷积结构的高参数量和复杂度,达到了轻量级的效果。

图1所示的框架中,输入为高光谱图像块 $X\in\mathbf{R}^{B\times C\times H\times W}$,其中 B 表示输入网络的批次大小;C 表示光谱通道数;H 表示影像高度;W 表示影像宽度。首先,将 X 输入具有分支结构与残差连接的空间特征学习模块,利用改进多尺度残差映射块对 X 进行特征提取。然后将特征输入基于注意力机制的空间特征优化块得到表征性更强的空间特征。另外将 X 输入到光谱特征学习模块,通过在光谱维度上的卷积来提取光谱特征。最后将该特征与空间特征输入融合分类模块,得到高和宽与输入一致、通道数为类别数(N)的分类概率图。

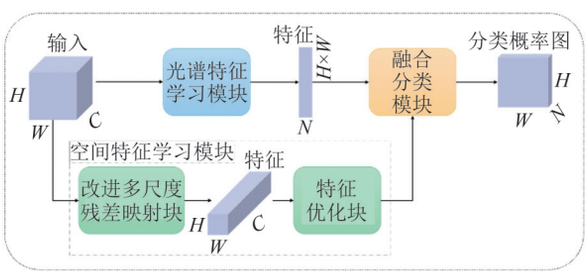


图1　基于卷积的高光谱图像分类框架

Fig. 1　Convolutional Hyperspectral Image Classification Framework

1.2　空间特征学习模块

空间特征学习模块主要由改进多尺度残差映射块与基于注意力机制的空间特征优化块组成。其中,多尺度卷积核是为了自适应地检测空间图像特征。下文首先介绍单一尺度下的改进残差结构,然后将其扩展到多尺度中。

1.2.1　改进多尺度残差映射块

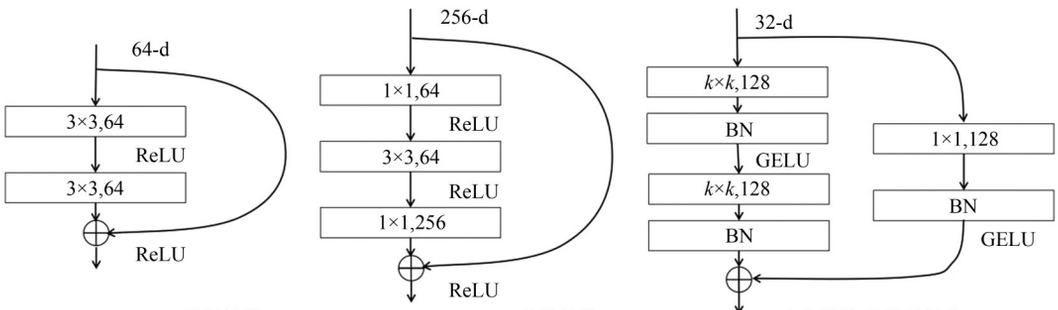
残差结构最早由何凯明等[5]在残差神经网络(residual neural network,ResNet)中提出。残差结构有助于解决梯度消失和梯度爆炸问题,使得网络在训练的同时能保证良好的信息传递,不少研究者尝试将残差结构应用于高光谱分类的网络设计当中[26,29-30]。

经典的残差结构如图2(a)、2(b)所示。其中,ResNet浅层网络使用图2(a)所示结构,其主干由两个3×3卷积构成,捷径分支直接将输入与主干输出进行逐像素相加,再通过一层整流线性单元(rectified linear unit,ReLU)激活函数得到特征。ResNet深层网络中使用图2(b)所示结构,其主干在3×3卷积前后加入1×1卷积进行降维、升维,最终得到高维特征。本文设计的残差结构如图2(c)所示,其中 k 表示卷积核大小。其主干由两个卷积层、两个BN(batch normalization)层构成,两层之间加入高斯误差线性单元(Gaussian error linear unit,GELU)激活函数进行自适应Dropout。捷径分支由一层1×1卷积和BN层及GELU激活函数构成。1×1卷积用来对原始输入进行升维,从而与分支末提取的特征相加,来获得更加多样的特征表示。

本文设计的残差结构中使用GELU替代传统的ReLU。ReLU激活函数与GELU激活函数的示意图如图3所示。ReLU的计算公式为:

$$y=\mathrm{ReLU}(x)=\max(0,x) \tag{1}$$

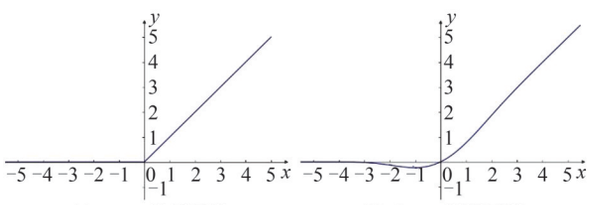
式中,x 表示输入的值;y 表示经过ReLU激活函数后的输出值。ReLU公式表明,如果输入的值小于0,则令输出等于0;若输入的值大于0,则令输出等于输入。因此ReLU激活函数的微分值为0或1,可以对输入信号进行选择性响应并且不会



(a) ResNet 18、34残差结构 (b) ResNet 101、154残差结构 (c) 本文设计的残差结构

图2 残差结构

Fig. 2 Residual Structure



(a) ReLU激活函数 (b) GELU激活函数

图3 ReLU与GELU激活函数图示

Fig. 3 Diagram of ReLU and GELU Activation Functions

产生过小值,能够有效解决梯度消失问题。

虽然ReLU已经成为了神经网络中最常用的激活函数之一,但ReLU由于其特性并不适用于低方差中心且梯度较大的输入。原因是较大的梯度容易使整个输入的分布发生变化,一旦高斯分布中心变为负值,会产生过多负数输入导致“死亡ReLU问题”。该问题指大部分输入经过ReLU函数后变为0,不能反向传播得到梯度,导致网络的很多参数无法更新。GELU替换ReLU可以提升网络精度,是更好的激活函数选择[31]。GELU激活函数的计算公式为:

$$y=\mathrm{GELU}(x)=0.5x\left(1+\tanh\left(\sqrt{2/\pi}\left(x+0.044\,715x^3\right)\right)\right) \tag{2}$$

式中,tanh表示双曲正切函数;x表示输入的值;y表示经过GELU激活函数后的输出值。GELU可以看作双曲正切函数与近似数值的组合。不同于ReLU会确定性地将输入乘以0或1,GELU会在考虑到输入自身分布的情况下将0或1与输入相乘。选择0或1取决于当前输入有多大的概率大于其余输入,当输入越小时,被丢弃的可能性越高。因此GELU既保持了随机性,又建立了与输入的依赖关系,可以有效解决梯度消失和梯度爆炸的问题。

此外,原始残差结构中主干和捷径分支都没有标准化处理过程,这导致训练过程中容易产生分布偏差导致模型训练困难并影响其后激活函数的性能。为了改进该弊端,本文设计的残差结构在GELU激活函数前加入BN层进行正则化处理。其计算公式如下:

$$\mu_B=\frac{1}{m}\sum_{i=1}^{m}x_i \tag{3}$$

$$\sigma_B^2=\frac{1}{m}\sum_{i=1}^{m}\left(x_i-\mu_B\right)^2 \tag{4}$$

$$\hat{x}_i=\frac{x_i-\mu_B}{\sqrt{\sigma_B^2+\epsilon}} \tag{5}$$

$$y_i=\gamma\hat{x}_i+\beta\equiv\mathrm{BN}_{\gamma,\beta}(x_i) \tag{6}$$

式中,μ_B表示一批样本的均值;σ_B^2表示样本的方差;$\hat{x}_i$表示经过标准化处理的样本;y_i表示经过仿射变换的输出;x_i表示输入样本;m表示一个batch中样本的个数;ϵ是避免分母为零的常数;γ和β分别表示缩放因子和平移因子。

基于上述改进残差结构,引入不同尺寸卷积核来将其扩展到多尺度结构中。采用多尺度卷积核是为了自适应地检测不同尺度的空间图像特征。图4为改进多尺度残差映射块示意图。该模块采取分支形式由3个上述残差结构组成,其中每个分支具有不同尺度的卷积核,分别为3×3、5×5、7×7。利用3种不同大小的卷积核可以使网络自适应地提取不同尺度的空间特征。该模块将原始高光谱图像块$X\in\boldsymbol{R}^{B\times C\times H\times W}$分别输入3个分支中,其中$B=4$,$C=32$,$H=256$,$W=256$。首先计算得到每个分支的尺度特征$F_3$、$F_5$、$F_7$,然后进行拼接操作,得到$F$作为该模块的最终特征,计算公式如下:

$$f_i^1=\mathrm{BN}(\mathrm{Conv}_{i\times i}(\sigma(\mathrm{BN}(\mathrm{Conv}_{i\times i}(X))))) \tag{7}$$

$$f_i^2=\sigma(\mathrm{BN}(\mathrm{Conv}_{1\times 1}(X))) \tag{8}$$

$$F_i=f_i^1+f_i^2 \tag{9}$$

$$F=\mathrm{Concat}(F_3,F_5,F_7) \tag{10}$$

式中,f_i^1表示主干分支得到的特征;f_i^2表示捷径分支得到的特征;F_i表示加和后的分支特征;F表示将

分支特征进行拼接后得到的特征;$\mathrm{Conv}_{i\times i}(\cdot)$为具有$i\times i$卷积核的二维CNN层,$i$表示卷积核大小,$i\in[3,5,7]$;BN(·)表示批量归一化层;$\sigma$为GELU激活函数;Concat(·)表示将每个分支特征在光谱维度进行拼接,将特征空间扩展为$\boldsymbol{R}^{B\times C'\times H\times W}$,本文中$C'=3\times 128=384$。

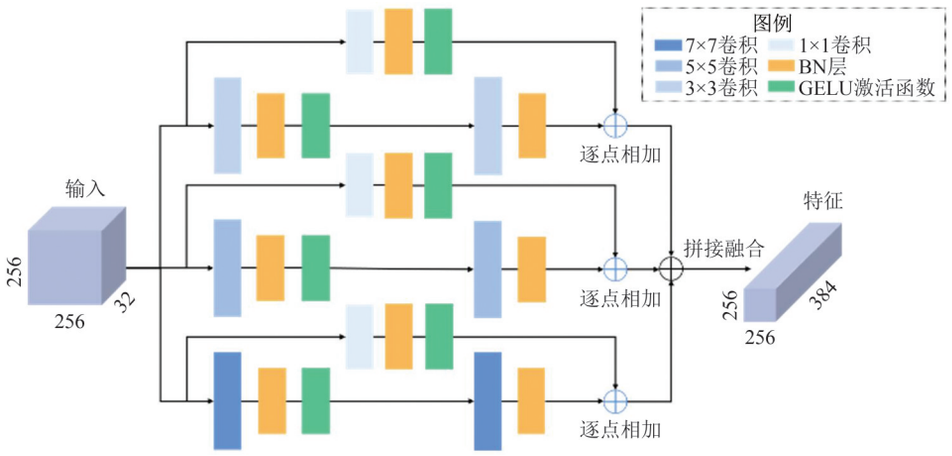


图4　改进多尺度残差映射块示意图

Fig. 4　Diagram of Improved Multi-scale Residual Mapping Block

输入X经过改进多尺度残差映射块后,得到通道维为384的高维特征。该特征将输入到空间特征优化块对每一通道赋予相应的权重,使网络有效地保留重要通道信息来提升网络的分类精度。

1.2.2　基于注意力机制的空间特征优化块

空间特征优化块本质上是一种使用注意力机制的模块,它以通道域为注意力域,通过计算每个通道的均值学习权重来表示该通道的重要程度。这种模块结构简单、使用方便,能够有效地对特征图与全局空间的依赖关系进行建模。实际应用中,该模块可以对重要的特征图进行突出展示,以提高模型的准确性,在卷积神经网络分类中得到了广泛应用[23,29,32-33]。本文使用的基于注意力机制的空间特征优化块结构见图5。

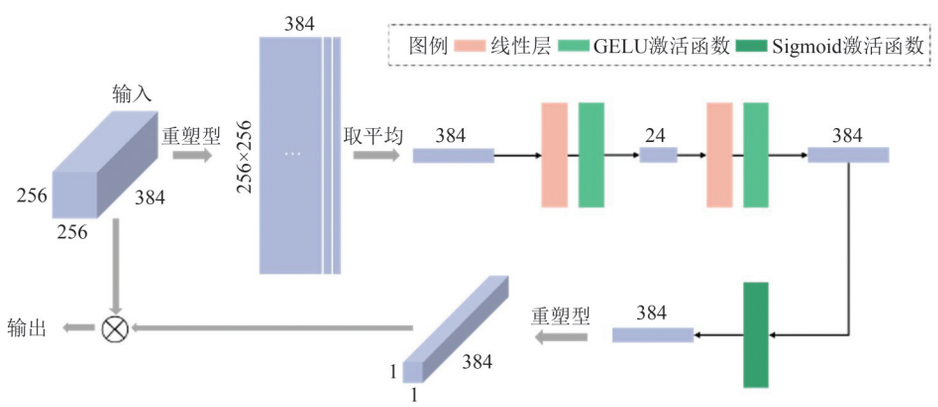


图5　基于注意力机制的空间特征优化块示意图

Fig. 5　Diagram of Spatial Feature Optimization Block Based on Attention Mechanism

空间特征优化块将改进多尺度残差映射块得到的特征$F\in\boldsymbol{R}^{C'\times H\times W}$作为输入,对$F$首先进行高度和宽度维的展平操作,得到二维矩阵$\boldsymbol{M}\in\boldsymbol{R}^{C'\times(H\times W)}$。然后计算通道级均值:

$$Q(k)=\frac{1}{H\times W}\sum_{i=0}^{H}\sum_{j=0}^{W}M(k,i,j)\tag{11}$$

式中,$Q(k)$表示第k个通道的均值,$k\in[0,C')$,$Q\in\boldsymbol{R}^{C'}$。接着通过两层全连接层和一层Sigmoid激活函数对提取到的光谱信息进行拟合:

$$Y=\mathrm{Sigmoid}(W\cdot\sigma(WQ))\tag{12}$$

式中,Y表示拟合后的光谱信息,$Y\in\boldsymbol{R}^{C'}$;W表示全连接层;σ表示GELU激活函数。最后将调整大小后的拟合特征与F相乘进行加权,得到该模块的输出:

$$Z=F\cdot Y\tag{13}$$

式中,$Z\in\boldsymbol{R}^{C'\times H\times W}$。

1.3　光谱特征学习模块

本文光谱特征学习模块主要由三层一维卷

积组成,先将影像的空间形状展平来使卷积仅聚焦于光谱信息,接着通过在频谱维度的卷积提取特征。该模块类似于时序分类网络提取数据的时序信息[34],其结构示意图见图6。一维卷积通常用于识别序列数据中的局部模式,因此对于高光谱数据可以较好地捕捉相邻光谱间的相关性,同时具有参数量少的优势。该模块中一维卷积的卷积核大小为7,填充边长设为3,步长为1。

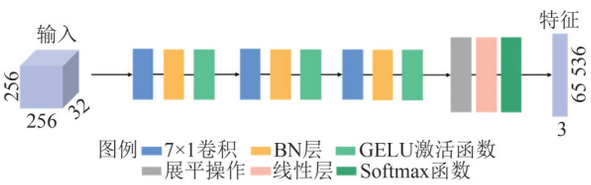


图6 光谱特征学习模块结构

Fig. 6 Structure of Spectral Feature Extraction Module

该模块首先将原始输入$X\in\boldsymbol{R}^{C\times H\times W}$输入到由7×1一维卷积、BN及GELU组成的三层卷积块中,然后将特征进行展平输入全连接层和最后一层Softmax,通过下式得到特征$L\in\boldsymbol{R}^{(H\times W)\times N}$:

$$L=g(W\cdot A+b) \tag{14}$$

式中,W和b分别表示全连接层的权重系数与偏置值;$g(\cdot)$表示Softmax激活函数,用来对线性组合后的输出进行非线性激活。

1.4 融合分类模块

融合分类模块用于融合空间特征学习模块和光谱特征学习模块获得的特征并进行最终分类。本文融合分类模块结构示意图如图7所示。通过空间特征学习模块网络将原始输入变化为深度特征并进行加权。通过光谱特征学习模块,得到每个像素的光谱特征。融合分类模块首先将空间深度特征经过BN和GELU进行正则化,再通过1×1卷积进行降维。1×1卷积在减少计算量的同时将特征向量由特征空间映射到样本类别空间,得到H和W与输入相同、通道维与类别数(N)相同的特征图。另外,融合分类模块对光谱特征进行重塑型操作,同样得到H和W与输入相同、通道维与类别数(N)相同的光谱特征图。最后,该模块将两个特征图进行相加操作再输入Softmax层,得到每个像素的分类概率结果。

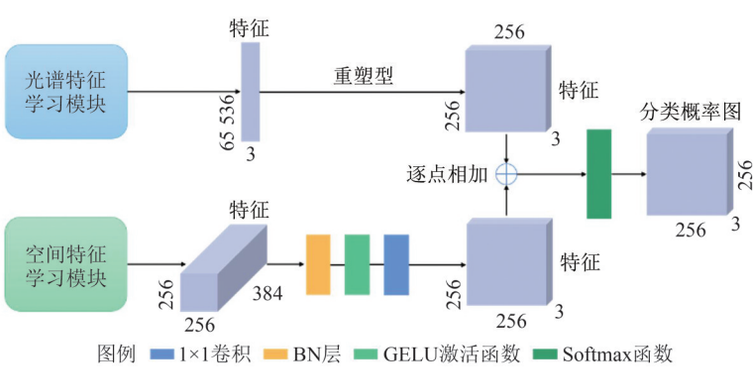


图7 融合分类模块结构

Fig. 7 Structure of Classification Module

1.5 网络训练

本文采用Adam(adaptive moment estimation)算法[35]作为优化器,采用余弦退火学习率进行学习率的调节。训练时将初始学习率设置为0.001,批次大小设置为4,运行100个epochs。本文采用分类任务中常用的交叉熵损失函数作为Loss函数。Loss函数的作用是判断网络的输出与期望的输出之间的接近程度,即计算输出与标签之间的差异,再利用反向传播梯度来更新参数。

2 实验与分析

2.1 实验数据集

沈阳位于中国东北地区南部、辽宁中部,位于41°11′52″N~43°02′25″N、122°25′13″E~123°48′30″E之间,属于温带半湿润大陆性气候。主要作物类型为粮食作物,占农作物总播种面积的80%。粮食作物中又以玉米和稻谷为主要类型,占粮食作物总播种面积的72%和23%[36]。

本文数据源为珠海一号高光谱卫星(orbita hyperspectral satellite,OHS)。珠海一号OHS为珠海欧比特宇航科技股份有限公司投资建设并运营的商业遥感微纳卫星星座之一。该卫星采用推扫成像方式,空间分辨率10 m,光谱分辨率为2.5 nm,包含32个波谱范围在0.4~1.0 μm的光谱通道。本文使用的是03组高光谱卫星OHS-3B获取的影像。OHS影像首先需要经过波段组合及辐射定标、大气校正、正射有理多项式系数

校正、投影转换、拼接及裁剪等数据预处理过程,其主要使用OHS数据处理软件及ENVI(the environment for visualizing images)软件实现。表1为该卫星各通道中心波长信息。

本文数据集影像由OHS于2021年8月在沈阳市获取的9幅影像制成,总体大小为17 028×11 810。数据集农作物标签由沈阳市勘察测绘研究院有限公司提供,由测绘人员2020年实地调查采集制成。标签分布范围达12 942 km^2,原始分辨率为1 m,需要重采样至10 m。重采样后玉米、一季稻、其他三类分别占总样本的25.4%、2.9%、71.7%。研究区影像及标签分布见图8。

为了验证本文方法的有效性,将整个沈阳市数据集划分为256×256大小互不相交的训练集、验证集、测试集,比例为7:1:2。最终得到训练集2 101张、验证集391张、测试集835张。在划分数据的过程中,剔除无效数据大于20%的影像。

表1　OHS-3B卫星各通道中心波长/nm

Table 1　Center Wavelength of Each Channel of OHS-3B/nm

通道	中心波长	通道	中心波长	通道	中心波长	通道	中心波长
B01	443	B09	580	B17	709	B25	833
B02	466	B10	596	B18	730	B26	850
B03	490	B11	620	B19	746	B27	865
B04	500	B12	640	B20	760	B28	880
B05	510	B13	665	B21	776	B29	896
B06(B)	531	B14(R)	670	B22	780	B30	910
B07	550	B15	686	B23	806	B31	926
B08(G)	560	B16	700	B24	820	B32	940

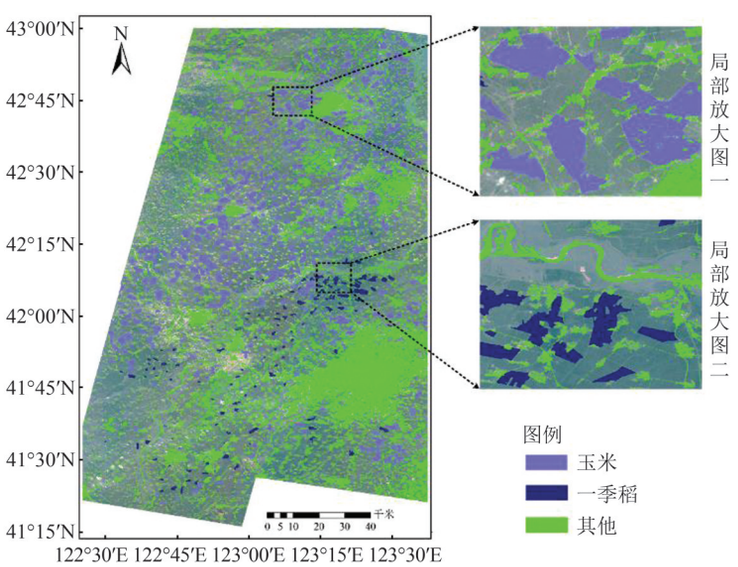


图8　研究区高光谱影像及标签分布

Fig. 8　HSI and Label Distribution in the Study Area

2.2　实验环境与设置

本文方法使用PyTorch深度学习框架,通过Python语言编程实现,需使用NVIDIA GeForce RTX 3090显卡。使用珠海一号OHS高光谱影像对沈阳市玉米、一季稻两种作物类型进行识别,并且将本文方法与近年来提出的U-Net[17]、PSPNet[18]、FreeNet[23]、MAP-Net[37]4种网络进行对比来评估所提出方法的性能。采用$F1$分数来评价每种作物的分类效果,采用总体精度(overall accuracy,OA)、Kappa系数、平均交并比(mean intersection over union,mIoU)、$F1$分数来评价整体分类效果。OA表示正确分类的像元数与总像元个数的比值,即准确率;Kappa系数常用于一致性检验和分类精度评价,它可以衡量数据集分类的偏向性,值越高,偏向性越低;mIoU是一种语义分割标准度量,它表示所有类别交集和并集之比的平均值;$F1$分数是一种分类问题衡量指标,是精确率和召回率的调和平均数,适合用于单类精度评价。4种精度指标值越高,表明分类效果越好。

2.3　实验结果分析

本节对融合分类模块的融合策略进行了对比,对提出的改进多尺度残差映射块及基于注意力机制的空间特征优化块和光谱特征学习模块进行了消融实验。

2.3.1　融合策略对比

为了分析空-谱融合的影响,本文尝试了4种融合策略(见表2):策略1为特征层融合,在融合分类模块进行特征融合,输出一个分类结果并计算损

失;策略2为决策层融合,表示将空间特征学习模块和光谱特征学习模块的分类结果分别输出并计算损失,然后直接相加进行反向传播;为了对比空-谱特征的重要性程度,策略3在策略2的基础上将各模块分类损失采用2:1加权;而策略4是在策略2的基础上将各模块分类损失进行1:2加权。

表2 不同融合策略在测试集上的分类结果/%

Table 2 Classification Results on the Test Set of Different Fusion Strategies/%

融合策略	F1分数			OA	Kappa	mIoU
	玉米	一季稻	其他			
策略1	87.0	75.8	94.1	91.5	81.3	75.6
策略2	86.9	74.5	93.9	91.3	80.9	74.9
策略3	86.4	66.9	94.0	90.9	80.3	71.6
策略4	73.4	20.0	89.2	83.0	65.4	50.0

由表2可知,4种融合策略中,策略2、3、4分别计算两模块的分类损失再进行加权组合的方法精度更低,并且策略3的精度高于策略4。这表明在本文分类任务中:(1)空间特征学习模块对农作物的解译更重要;(2)本文提出的特征融合模块可以有效地将空间特征和光谱特征进行融合;(3)特征融合模块也可以解决多个任务的损失由于权重设置不合理导致的模型退化。因此,本文最终采用了策略1作为空-谱模块融合的方法。

2.3.2 消融实验

为了验证改进多尺度残差映射块的有效性,在原网络基础上将该模块替换为2个对比模块(见表3):对比模块1:使用ResNet浅层网络中使用的经典残差结构代替本文残差结构,卷积核大小仍采用多尺度核;对比模块2:使用ResNet浅层网络中使用的经典残差结构代替本文残差结构,所有卷积核都采用3×3大小。

表3 改进多尺度残差映射块消融实验结果/%

Table 3 Ablation Study Results of Improved Multi-scale Residual Mapping Blocks/%

模型	F1分数			OA	Kappa	mIoU
	玉米	一季稻	其他			
对比模块1	69.3	50.5	88.1	81.6	59.0	50.9
对比模块2	65.2	20.0	86.8	78.9	52.5	44.1
改进多尺度残差映射块	73.5	51.4	89.4	83.0	62.9	55.2

由表3可知,本文提出的改进多尺度残差映射块的测试集整体精度和单类精度均优于两种对比模块。与对比模块1相比,实验证明添加的BN层和GELU激活函数有效提升了测试集精度,验证了本文提出的改进残差映射块的有效性。对比模块2在对比模块1的基础上将多尺度进行了消融,实验结果表明,该模块效果最差,验证了将本文提出的改进残差映射块扩展到多尺度卷积核上的有效性。

为了验证基于注意力机制的空间特征优化块和光谱特征学习模块的有效性,构建了3个对比网络模型(见表4):模型1为在原网络基础上去掉基于注意力机制的空间特征优化块;模型2在原有网络基础上去掉光谱特征学习模块;模型3在原有网络基础上同时去掉基于注意力机制的空间特征优化块和光谱特征学习模块。

由表4可知,本文算法的分类结果在整体精度和单类精度上均要优于3种消融模型。与去除空间特征优化块的模型1相比,本文算法玉米、一季稻、其他类别的单类精度分别提升了6.4%、26.1%、1.9%,整体精度OA、Kappa、mIoU分别提升了3.7%、8.7%、13.6%。与去除光谱特征学习模块的模型2相比,本文算法玉米、一季稻、其他类别的单类精度分别提升了4.9%、15.8%、1.7%,整体精度OA、Kappa、mIoU分别提升了3.1%、6.7%、9.4%。与同时去除空间特征优化块和光谱特征学习模块的模型3相比,本文算法的精度优势最为明显,玉米、一季稻、其他类别的单类精度分别提升了14.5%、57.1%、6.4%,整体精度OA、Kappa、mIoU分别提升了10.4%、21.1%、27.2%。实验结果证明了基于注意力机制的空间特征优化块与光谱特征学习模块的有效性。另外,由模型1与模型2的精度对比可以看出,基于注意力机制的空间特征优化块对模型的贡献要大于光谱特征学习模块。

表4 消融模型在测试集上的分类结果/%

Table 4 Classification Results on the Test Set of Ablation Models/%

模型	F1分数			OA	Kappa	mIoU
	玉米	一季稻	其他			
模型1	80.6	49.7	92.2	87.8	72.6	62.0
模型2	82.1	60.0	92.4	88.4	74.6	66.2
模型3	72.5	18.7	87.7	81.1	60.2	48.4
本文算法	87.0	75.8	94.1	91.5	81.3	75.6

进一步分析3种消融模型和本文算法的实验效果,采用梯度加权类激活映射(gradient-weighted class activation mapping,Grad-CAM)方法[38]进行特征可视化。Grad-CAM是一种卷积神经网络解释方

法,通过构建热力图的形式直观展示神经网络学习到的特征,即关注区域是哪里。图 9 展示了 4 种模型通过 Grad-CAM 方法得到的 3 种地类的热力图。由图 9 可以看出,对每种类别,本文算法得到的热力图都更加清晰,与影像上的类别区域重叠范围较为符合,证明与对比模型相比,本文方法使得这部分影像激活值更大,更加有效地关注到了对分类最具贡献性的区域。另外,由可视化结果可以看出,3 种对比模型中模型 2 效果最好,模型 1 次之,模型 3 最差,与定量的精度结果相符。

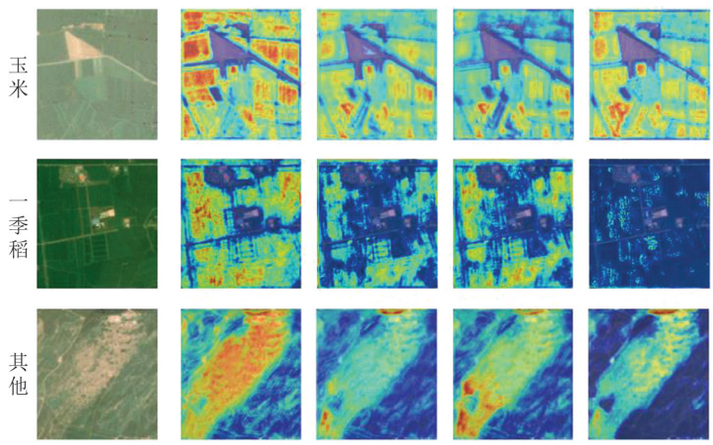


图 9　消融模型 Grad CAM 可视化热力图

Fig. 9　Grad CAM Visualization Heat Map of Ablation Models

基于注意力机制的空间特征优化块以通道域为注意力域首先计算每个特征的均值来学习权重,再将该权重赋予各通道,能够有效地改善各通道特征与全局空间的依赖关系,排除了冗余信息对分类的影响。光谱特征学习模块通过一维卷积来识别每个像素光谱数据中的局部模式,能够捕捉相邻光谱间的相关性,实现农作物光谱特征的有效提取。空间特征优化块对空间信息的压缩和重新利用与光谱特征学习模块对光谱信息的提取于本文的农作物分类任务而言都十分重要。而基于注意力机制的空间特征优化块对模型的贡献大于光谱特征学习模块,这可能是由于本文使用的高光谱数据通道数较少,分辨率相对适中,因此空间信息在分类时占据优势。

2.3.3　不同算法模型对比

本文提出的基于卷积的高光谱图像分类算法与 4 种对比算法在本文数据集测试集上的分类精度见表 5。由表 5 可以看出,本文算法在测试集上的 OA、mIoU 均优于其他对比算法,Kappa 优于 PSPNet。对于每个类别的单类精度,本文算法提取玉米的效果优于 PSPNet,提取一季稻的效果优于所有对比算法,提取其他类别的效果优于 U-Net、PSPNet、MAP-Net,与 FreeNet 精度相当。一季稻在整体样本中的占比最少,仅包含 2.9%,因此一季稻的单类分类更具挑战性。本文算法实现了对该类别的最优提取,其 $F1$ 分数相较 U-Net、PSPNet、MAP-Net、FreeNet 分别提升了 4.2%、11.5%、13.6%、12.0%。这表明虽然 FreeNet 获得了较高的 Kappa 值,但这种高精度很大程度上依赖于对优势样本的分类,因此 FreeNet 对于样本量最少的一季稻分类效果较差,而本文算法可以对所有类别实现较好的分类效果。

表 5　不同算法模型在测试集上的分类结果/%

Table 5　Classification Results on the Test Set of Different Models/%

模型	$F1$ 分数			OA	Kappa	mIoU
	玉米	一季稻	其他			
PSPNet	83.4	70.4	93.6	90.1	77.6	71.2
U-Net	85.8	77.7	94.5	91.2	81.5	76.3
MAP-Net	88.0	68.3	94.8	91.8	81.3	73.5
FreeNet	87.1	69.9	94.9	91.9	82.1	73.7
本文算法	85.1	81.9	94.9	92.3	80.7	77.9

将 5 种网络在测试集上的分类结果进行可视化,挑选两张结果图进行展示,如图 10、图 11 所示。由可视化结果可以看出,与所有其他对比网络的分类结果相比,本文方法分类错误的像素更少、边界更加清晰。另外还可以看出,对于一些较小地物如细长的街道或田间小路,本文方法有着更好的识别效果。因此,综合考虑 5 种网络在测试集上的整体精度、单类精度以及可视化分类结果图可以看出,本文提出的基于卷积的高光谱图像分类算法可以更加有效地提取高光谱遥感

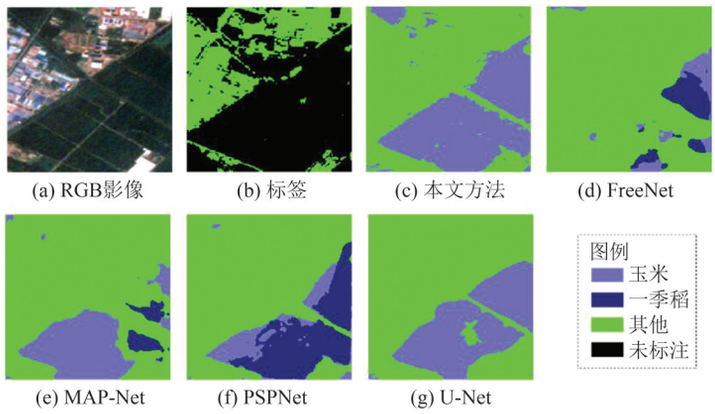


图10 某张测试集影像1的RGB图、标签及5种网络预测结果图

Fig. 10 RGB, Label and Prediction Results of Five Networks on Image 1 in the Test Set

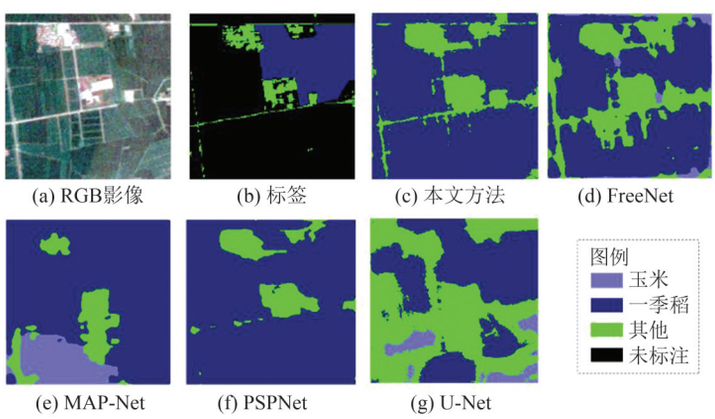


图11 某张测试集影像2的RGB图、标签及5种网络预测结果图

Fig. 11 RGB, Label and Prediction Results of Five Networks on Image 2 in the Test Set

影像的光谱-空间特征,从而提升网络的分类性能。

为了对不同算法模型的复杂程度进行对比,表6列举了4种对比算法及本文算法的参数量、计算量和推理速度。参数量表示模型需要训练的参数总数,即网络在反向传播过程中可训练权重参数的数量,由网络结构决定。计算量表示浮点运算数,通过理论上的计算量来衡量计算复杂度。由表6可观察到本文算法与FreeNet使用的参数数量远小于PSPNet、U-Net和MAP-Net,但本文算法与UNet的计算量FLOPs大于其他网络。然而FLOPs并不是一个直接衡量模型速度或者大小的指标,如ShuffleNet V2[39]与MobileNet V2[40]具有相同的FLOPs,但速度前者明显快于后者[39],因此添加了推理速度作为更直观的网络速度指标(推理速度表示一张影像需要的推理时间)。由表6可以看出,本文算法在推理速度上仍具优势。由此可得:本文设计的轻量化网络在参数量较少即模型较为简单、推理速度较快的同时,实现了更优异的分类性能。

表6 不同算法模型的参数量、计算量与推理速度

Table 6 Parameters, Computation and Inference Speed of Different Algorithm Models

模型	参数量/MB	计算量FLOPs/MB	推理速度/ms
PSPNet	46.74	497.44	12.85
U-Net	34.54	1 043.05	13.35
MAP-Net	23.57	354.86	11.40
FreeNet	2.50	253.87	11.71
本文算法	2.87	1 245.65	10.60

3 结 语

本文提出了一种基于多分支卷积的空-谱融合高光谱分类网络,其空间、光谱特征联合的思路可以有效地利用影像的空间上下文信息及光谱通道间的相关性,从而提高分类精度。本文算法主要由空间特征学习模块、光谱特征学习模块、融合分类模块3个模块组成。其中空间特征学习模块的改进多尺度残差映射块通过分支结构和改进后的残差结构来获取不同尺度且鲁棒性更强的空间特征;而基于注意力机制的空间特

征优化块可以使网络在学习时集中于那些更具判别力的特征。光谱特征学习模块通过在光谱维进行卷积来获得光谱特征。最后由融合分类模块将两类特征融合,从而获得更具表征性的分类特征,提高算法的分类性能。将所提出的方法与其他4种对比方法用于沈阳市高光谱影像上两种农作物的提取,实验结果表明,本文算法可以实现最佳的高光谱图像分类。和其他对比方法相比,本文算法在具有较少模型参数量和较短推理速度的同时,获得了最佳的测试集OA和mIoU。同时,该方法也在两类地物的分类上获得了最佳单类精度,尤其在一季稻类别上具有最大的优势。但本文方法仍有改进空间,比如在消融实验中,光谱特征学习模块的贡献较小,表明没有最大化地利用频谱信息。后续可以考虑使用不同尺度的卷积核或者使用循环神经网络等其他网络进行光谱特征提取。另外,本文采用单独获得空间、光谱特征最后融合成深度特征的方法进行分类,这虽然可以减少模型的参数量,但会导致两类特征间没有依赖关系。后续将考虑设计轻量级三维卷积网络来同时提取两种特征。

参 考 文 献

[1] LI N, DONG X F, GAN F P. Research on Inversion of Mineral Content Information Based on Hyperspectral Remote Sensing[C]//2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS,Brussels, Belgium, 2021.

[2] YANG L X, ZHANG R, BAO Y J, et al. Kernel Tensor Sparse Coding Model for Precise Crop Classification of UAV Hyperspectral Image[J]. *IEEE Geoscience and Remote Sensing Letters*, 2023, 20: 5511405.

[3] 张凝, 杨贵军, 赵春江, 等. 作物病虫害高光谱遥感进展与展望[J]. 遥感学报, 2021, 25(1): 403-422.
ZHANG Ning, YANG Guijun, ZHAO Chunjiang, et al. Progress and Prospects of Hyperspectral Remote Sensing Technology for Crop Diseases and Pests[J]. *National Remote Sensing Bulletin*, 2021, 25(1): 403-422.

[4] 张立强, 李洋, 侯正阳, 等. 深度学习与遥感数据分析[J]. 武汉大学学报(信息科学版), 2020, 45(12): 1857-1864.
ZHANG Liqiang, LI Yang, HOU Zhengyang, et al. Deep Learning and Remote Sensing Data Analysis[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(12): 1857-1864.

[5] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016.

[6] ZHANG K, GUO L R, GAO C, et al. Pyramidal RoR for Image Classification[J]. *Cluster Computing*, 2019, 22(2): 5115-5125.

[7] PENG D F, ZHANG Y J, GUAN H Y. End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++ [J]. *Remote Sensing*, 2019, 11(11): 1382.

[8] CHEN H, QI Z P, SHI Z W. Remote Sensing Image Change Detection with Transformers [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 5607514.

[9] DEVLIN J, CHANG M, LEE K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [C]//Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, USA, 2019.

[10] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need [C]//International Conference on Neural Information Processing Systems (NIPS'17), AngelesLos, USA, 2017.

[11] LI W, WU G D, ZHANG F, et al. Hyperspectral Image Classification Using Deep Pixel-Pair Features [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(2): 844-853.

[12] HU W, HUANG Y Y, WEI L, et al. Deep Convolutional Neural Networks for Hyperspectral Image Classification [J]. *Journal of Sensors*, 2015, 2015(1): 258619.

[13] MOU L C, 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.

[14] ZHANG X R, SUN Y J, JIANG K, et al. Spatial Sequential Recurrent Neural Network for Hyperspectral Image Classification[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018, 11(11): 4141-4155.

[15] 高奎亮, 刘冰, 余旭初, 等. 面向高光谱影像分类的网络结构自动搜索方法[J]. 武汉大学学报(信息科学版), 2024, 49(2): 225-235.
GAO Kuiliang, LIU Bing, YU Xuchu, et al. Automatic Network Structure Search Method for Hyperspectral Image Classification[J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(2): 225-235.

[16] HANG R L, LIU Q S, HONG D F, et al. Cascaded Recurrent Neural Networks for Hyperspectral Image

Classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(8): 5384-5394.

[17] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]//Medical Image Computing and Computer-Assisted Intervention — MICCAI, Cham, Switzerland, 2015.

[18] ZHAO H S, SHI J P, QI X J, et al. Pyramid Scene Parsing Network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017.

[19] 王克奇, 彭熙雯, 张怡卓, 等. 基于改进U-Net的高光谱农林植被分类方法[J]. 森林工程, 2022, 38(1): 58-66.
WANG Keqi, PENG Xiwen, ZHANG Yizhuo, et al. A Hyperspectral Classification Method for Agro-forestry Vegetation Based on Improved U-Net[J]. *Forest Engineering*, 2022, 38(1): 58-66.

[20] PAOLETTI M E, HAUT J M, FERNANDEZ-BELTRAN R, et al. Deep Pyramidal Residual Networks for Spectral - Spatial Hyperspectral Image Classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(2): 740-754.

[21] SONG W W, LI S T, FANG L Y, et al. Hyperspectral Image Classification with Deep Feature Fusion Network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(6): 3173-3184.

[22] GONG Z Q, ZHONG P, YU Y, et al. A CNN with Multiscale Convolution and Diversified Metric for Hyperspectral Image Classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(6): 3599-3618.

[23] ZHENG Z, ZHONG Y F, MA A L, et al. FPGA: Fast Patch-Free Global Learning Framework for Fully End-to-End Hyperspectral Image Classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(8): 5612-5626.

[24] PRAVEEN B, MENON V. Dual-Branch-AttentionNet: A Novel Deep-Learning-Based Spatial-Spectral Attention Methodology for Hyperspectral Data Analysis[J]. *Remote Sensing*, 2022, 14(15): 3644.

[25] TINEGA H C, CHEN E Q, MA L, et al. HybridGBN-SR: A Deep 3D/2D Genome Graph-Based Network for Hyperspectral Image Classification[J]. *Remote Sensing*, 2022, 14(6): 1332.

[26] ROY S K, MANNA S, SONG T C, et al. Attention-Based Adaptive Spectral - Spatial Kernel ResNet for Hyperspectral Image Classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(9): 7831-7843.

[27] SHI C, PUN C M. Superpixel-Based 3D Deep Neural Networks for Hyperspectral Image Classification[J]. *Pattern Recognition*, 2018, 74: 600-616.

[28] 宫正. 基于深度学习的高光谱遥感影像降维及地物分类算法研究[D]. 长春: 吉林大学, 2022.
GONG Zheng. Research on Dimension Reduction and Ground Object ClassificationAlgorithm of Hyperspectral Remote Sensing lmages Based on Deep Learning[D]. Changchun: Jilin University, 2022.

[29] 袁芊芊, 谢维信. 基于空谱注意力机制及预激活残差网络的高光谱图像分类算法[J]. 信号处理, 2022, 38(12): 2594-2605.
YUAN Qianqian, XIE Weixin. Hyperspectral Image Classification Based on Spatial Spectral Attention and Pre-Activation Residual Networks[J]. *Journal of Signal Processing*, 2022, 38(12): 2594-2605.

[30] 杨赵. 基于深度学习的空谱联合高光谱图像分类技术研究[D]. 西安: 中国科学院西安光学精密机械研究所, 2021.
YANG Zhao. Research on Spectral - Spatial Joint Hyperspectral Image Classification Technology Based on Deep Learning[D]. Xi'an: Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, 2021.

[31] LIU Z, MAO H Z, WU C Y, et al. A ConvNet for the 2020s[EB/OL]. 2022: 2201.03545. https://arxiv.org/abs/2201.03545v2.

[32] HU J, SHEN L, SUN G. Squeeze-and-Excitation Networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018.

[33] 沈金悦. 基于卷积神经网络和注意力机制的高光谱遥感图像分类研究[D]. 青岛: 青岛科技大学, 2022.
SHEN Jinyue. Research on the Classification of Hyperspectral Remote Sensing Images Based on Convloutional Neural Network and Attention Mechanism[D]. Qingdao: Qingdao University of Science & Technology, 2022.

[34] PELLETIER C, WEBB G, PETITJEAN F. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series[J]. *Remote Sensing*, 2019, 11(5): 523.

[35] KINGMA D, BA J. Adam: A Method for Stochastic Optimization[C]//The 3rd International Conference on Learning Representations (ICLR 2015), DiegoSan, USA, 2015.

[36] 沈阳市统计局. 沈阳统计年鉴2021[EB/OL]. (2021-11)[2022-08-18]. http://tjj.shenyang.gov.cn/sjfb/ndsj/.

(下转第972页)