引用格式：WANG Feifei, ZHAO Huijie, LI Na, et al. Spectral-spatial Attention Residual Networks for Hyperspectral Image Classification[J]. Acta Photonica Sinica, 2023, 52(12):1210002
汪菲菲,赵慧洁,李娜,等.基于光谱-空间注意力残差网络的高光谱图像分类[J].光子学报,2023,52(12):1210002

基于光谱-空间注意力残差网络的高光谱图像分类

汪菲菲[1,3],赵慧洁[1,2,3],李娜[1,2,3],李思远[4],蔡昱[5]
(1 北京航空航天大学 仪器科学与光电工程学院 精密光机电一体化技术教育部重点实验室,北京 100191)
(2 北京航空航天大学 人工智能研究院,北京 100191)
(3 北京航空航天大学 "空天光学—微波一体化精准智能感知"工信部重点实验室,北京 100191)
(4 中国科学院西安光学精密机械研究所 光谱成像技术重点实验室,西安 710119)
(5 中国运载火箭技术研究院,北京 100076)

摘　要:在高光谱图像分类任务中,引入注意力改变提取到的光谱和空间特征权重,有效突出重要特征,提高分类准确率。将注意力机制、残差网络和特征提取模块集成到分类框架中,引入中心区域光谱注意力机制,在避免干扰像素对波段权重影响的同时,利用周围像素增强中心像素波段,增强光谱特征的鲁棒性进而提取有效的光谱特征。并在此基础上提出了光谱-空间注意力残差网络,该网络可以从高光谱图像中连续提取到丰富的光谱特征和空间特征,并通过残差网络连接特征提取模块,缓解了精度下降问题,保证网络良好的分类性能。在4个公开数据集上,所提出的分类算法和其他算法相比,各项指标均达到最优。

关键词:光谱-空间特征;残差网络;高光谱分类;光谱注意力机制;空间注意力机制
中图分类号:TP391.41　　**文献标识码**:A　　**doi**:10.3788/gzxb20235212.1210002

0　引言

高光谱图像(Hyper Spectral Image, HSI)通过几十甚至上百个光谱通道来提供丰富的光谱信息,可用于对各地物类别进行准确分类[1]。高光谱图像分类是高光谱影像处理和应用领域的一个热点研究方向,分类模型通过分析每个像素的光谱信息与空间信息,对该像素所属类别进行预测,然后与实际地物进行对应比较,实现地物目标分类。深度学习由于其强大的特征学习能力成为高光谱分类的主流算法。

在基于深度学习的分类算法,根据是否提取到数据的空间信息,可分为基于光谱和基于光谱-空间融合的分类方法。基于卷积神经网络(Convolutional Neural Networks, CNN)的方法是一种带有卷积结构的前馈神经网络[2],是一类非常重要的高光谱地物分类方法。其中基于光谱的分类方法使用一维卷积操作提取到待分类像素的光谱信息进行分类。HU Wei等[3]利用一维卷积神经网络提取像素光谱信息来进行分类。MOU Lichao等[4]利用循环神经网络来进行高光谱图像分类,其本质上也是利用了一维卷积网络进行分类。基于光谱的方法虽然简单,但是其精度无法令人满意。高光谱的空间上下文信息也有助于提高分类精度,因此现在常见分类算法都是基于光谱-空间信息融合的。ZHONG Zilong等[5]提出了一种光谱-空间变换网络,由光谱特征提取模块和空间注意力模块组成,充分利用HSI的光谱-空间信息进行分类。而GHADERIZADEH S等[6]则是提出利用混合三维和二维卷积神经网络来进行高光谱分类,其中三维卷积有效地提取光谱-空间信息,并用二维卷积来增强空间信息。WU Hao等[7]将卷积神经网络和循环神经网络相

基金项目:国家自然科学基金(No. 61975004),预研项目(No. 6230111002)
第一作者:汪菲菲,wff1231@buaa.edu.cn
通讯作者:李娜,lina_17@buaa.edu.cn
收稿日期:2023-04-18;**录用日期**:2023-05-25

结合提出了卷积循环神经网络，利用卷积操作提取到高光谱图像的光谱-空间信息，然后利用循环神经网络进一步提取光谱-空间特征上下文信息。ZHONG Zilong等[8]提出了光谱空间残差网络，连续提取光谱信息和空间信息特征。SHI Yuetian等[9]提出了利用多角度平行特征编码的方式，通过增强局部空间特征的方式提高图像分类精度，并且该算法对图像旋转鲁棒。与此同时，在高光谱图像实际分类任务中存在光谱相似、类别易混等问题，注意力机制广泛应用于分类任务，XU Yue等[10]在三维光谱卷积模块中利用注意力机制进行光谱-空间特征选择和提取。YANG Kai等[11]提出了交叉注意力机制，该网络分为像素和图像块2个分支输入，并对像素分支网络采用光谱注意力机制提取光谱特征，并将该特征作用到图像块分支网络中。ZHENG Xiangtao等[12]提出了中心光谱注意力机制，将中心光谱像素值作为特征权重对光谱特征进行新的校正，但高光谱图像块不可避免地包含干扰像素，因此采用全局平均池化引入干扰像素类别对注意力权重的生成不利。FANG Shuai等[13]的研究表明了不同地物类别其分类所依靠的光谱波段并不相同，也说明不同类别的光谱冗余波段可能不同。为此，中心池化的操作可将中心像素值直接代替原有的全局平均池化后的像素，并根据该中心像素值生成光谱注意力权重。

尽管上述工作取得了不错的效果，但是还有如下问题：1）多数工作在使用光谱注意力机制后，直接进行了空间特征提取，没有单独提取光谱特征，或者是单独提取光谱特征时，默认光谱各维度同等重要；2）光谱注意力机制多采用全局特征或中心像素特征进行权重调整，前者引入了较多干扰像素；而后者忽略了周围相同类别对其的影响。

为了解决上述问题，本文提出了光谱-空间注意力残差网络（Spectral-Spatial Attention Residual Network，SSARN）来进行高光谱分类。该网络主要包括光谱特征学习、空间特征学习和分类器。其中，光谱特征学习部分包括光谱注意力模块和光谱残差网络模块；而空间特征学习部分包括空间注意力模块和空间残差网络模块。由于现有的光谱注意力模块通常采用全局平均池化或者中心池化来提取光谱特征，但是无论哪种方式都会丢失光谱特征，为此提出了一种新的光谱注意力机制，尽可能减少光谱信息损失。

1 光谱-空间注意力残差网络

高光谱图像是一个包含光谱信息和空间信息的三维立方体，基于此提出了一个用于高光谱图像分类的光谱-空间注意力残差网络（SSARN）。如图1所示，SSARN包括光谱特征学习、空间特征学习和分类器。其中，光谱特征学习部分包括光谱注意力模块和光谱残差网络模块；而空间特征学习部分包括空间注意力模块和空间残差网络模块。

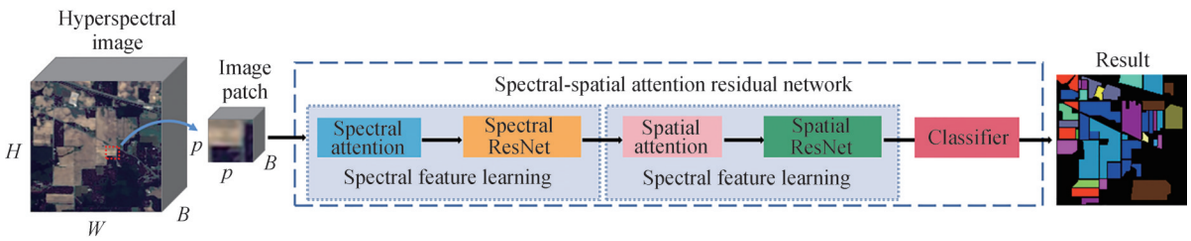

图1 光谱-空间注意力残差网络流程

Fig. 1 Flow chart of spectral-spatial attention residual network

首先高光谱图像会根据设定好的尺寸分割成图像块，然后这些图像块会被送入到网络中，根据各个模块提取图像特征，最后将特征输入到分类器中得到最终的分类结果。

1.1 中心区域光谱和空间注意力机制

1.1.1 中心区域光谱注意力机制

注意力机制的提出是为了节省资源，不需要让网络处理全部的输入信息，而是从这些信息中有选择地对与任务相关的信息进行计算[14]。根据处理任务时注意力机制作用的数据域位置不同，可分为光谱注意力机制和空间注意力机制。

光谱注意力机制在图像的光谱维度进行特征提取，也被称为通道注意力机制。图2所示就是一种光谱

注意力机制。由于高光谱图像包含几十甚至上百个光谱波段，而将全部波段放入网络中提取特征是不可行的，一方面需要大量的计算资源，另一方面这些波段和波段具有冗余关系[14]，可以用部分波段表征全部波段。主流方式用注意力模块重新调整各个波段的权重。该模块可以根据任务需要独立嵌入到任何网络中，自适应地生成注意力权重，即

$$\eta = f_{\mathrm{SpeA}}(X) \tag{1}$$

$$f_{\mathrm{SpeA}}(X) = \sigma(\mathrm{FC}(\mathrm{ave}(X))) \tag{2}$$

式中，权重参数η表示生成的每个波段的权重，$f_{\mathrm{SpeA}}(\cdot)$表示光谱注意力，X表示高光谱图像块，$\sigma(\cdot)$表示激活函数，FC$(\cdot)$表示全连接层，ave$(\cdot)$表示全局平均池化。权重越大的波段在后续特征学习时更容易得到神经网络的关注，提取更多的有利于高光谱分类的信息。通常，利用全局平均池化融合图像块的全部空间信息，然后对该信息利用全连接层和sigmoid函数来自适应地生成权重η参数。不同地物类别其分类所依靠的光谱波段不相同[14]，中心池化操作是将中心像素值直接代替原有的全局平均池化后的像素，并根据该中心像素值生成光谱注意力权重[11]。虽然该方法在一定程度上避免了干扰像素对权重的影响，但是也丢失了周围相同类别的光谱特征对中心像素光谱权重增强的作用。

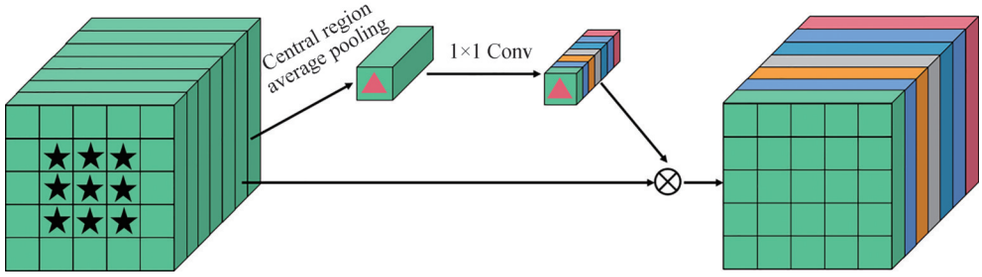

图2　中心区域光谱注意力机制结构

Fig. 2　The structure of the central region spectral attention mechanism

根据地理学第一定律[15]空间自相关性，待分类像素周围的像素可能属于同一类地物，因此周围的高光谱像素有可能会包含可用于提高分类结果的空间信息。所以一般在高光谱图像块中，周围像素与中心像素完全不同的概率较小，更多的是周围像素中包含了和中心像素相同的地物类别，并且越接近中心像素的区域，其包含相同类别的像素越多。为此，在现有的光谱注意力机制上提出了中心区域光谱注意力模块，在尽可能避免周围不同类别像素对中心像素干扰的同时，尽可能多利用周围相同类别像素波段对中心像素增强的作用。所提出的中心区域光谱注意力机制可以表示为

$$\tilde{x} = \mathrm{ave}(\mathrm{Center}_{3\times 3}(X)) \tag{3}$$

$$\eta = \sigma(\mathrm{conv}(\tilde{x})) \tag{4}$$

$$\tilde{X} = \eta \otimes X \tag{5}$$

式中，$\mathrm{Center}_{3\times 3}(X)$表示中心区域$3\times 3$范围的像素，conv$(\cdot)$表示卷积和激活函数的操作，$\otimes$表示卷积计算。如图2所示，选取中心区域像素，对这些像素求取平均值，获得中心区域像素平均值$\tilde{x}$。然后采用1×1卷积和激活函数从基于中心区域平均像素$\tilde{x}$生成注意力权重η。紧接着，利用该权重η与原始的图像块X进行卷积获得经过光谱注意力机制的高光谱图像块$\tilde{X}$。

1.1.2　空间注意力机制

空间注意力机制和光谱注意力机制的目的类似，都是将注意力转移到重要的部分，本质上是定位网络感兴趣的信息，抑制无用的信息。对于高光谱分类来说，空间包含的所有像素对中心像素的贡献并不是同等重要，只有能够帮助中心像素增加类间差异、缩小类内差异的像素才是网络需要关心的。空间注意力机制可以表示为

$$\delta = \mathrm{conv}[f_{\mathrm{m}}, f_{\mathrm{a}}] \tag{6}$$

$$\hat{X} = \delta \otimes X \tag{7}$$

式中，δ代表空间注意力权重，$[\cdot]$代表特征拼接，f_{m}，f_{a}分别代表最大池化和平均池化，X代表图像块（输入端）

或者是空间-光谱特征(在网络中),$\hat{X}$代表经过空间注意力机制后的空间特征。空间注意力机制如图3所示,将高光谱图像块X,经过池化层分别获取平均池化和最大池化特征,这两个特征在光谱维拼接后进行特征提取,输出经过注意力机制后的权重,最后和原始输入的空间特征进行卷积得到空间注意力机制后的空间特征。

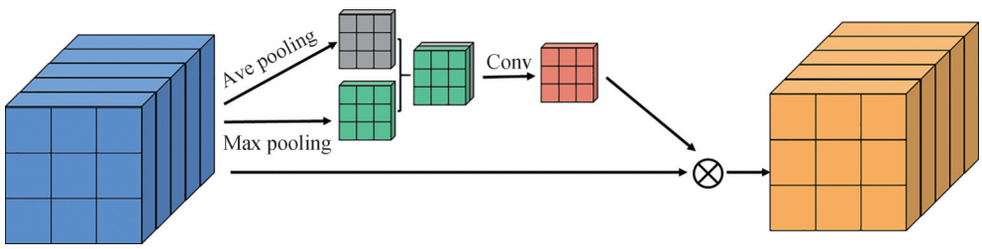

图3 空间注意力机制结构

Fig. 3 The structure of the spatial attention mechanism

1.2 光谱和空间残差网络模块

在深度学习中,神经网络层数的增加引发梯度下降,网络会发生退化现象,即训练集的损失会逐渐增大,浅层网络的精度反而优于深层网络,失去了深度学习的优势。其原因在于随着网络层数的递增,提取的特征所包含的图像信息越来越少,导致网络的分类精度下降。残差网络可进行图像识别任务[16-19],被广泛用于高光谱图像分类中[8,20-21],其由一系列残差单元组成,标准的残差单元可以表示为

$$x_{l+1}=x_l+F(x_l, W_l, b_l) \tag{8}$$

式中,x_{l+1}代表第$l+1$层特征,$F(x_l, W_l, b_l)$代表对第l层特征进行特征提取,(W_l, b_l)分别代表第l层网络参数,目的是让第l层和第$l+1$层的特征图保持大小一致,然后在输入输出前后增加一个恒等映射的跳跃连接,残差块的基本结构如图4所示。

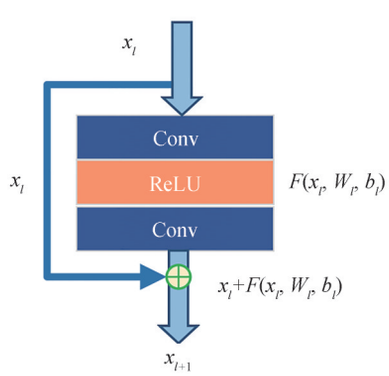

图4 残差块的基本结构

Fig. 4 The structure of the residual network

光谱特征学习的残差块如图5所示,残差块包括两个连续的卷积层和一个跳跃连接,跳跃连接可以保证第$p+2$层特征中包含有第p层的特征。对于第p层和第$p+1$层,分别使用尺寸为$1\times1\times m$的卷积核C_{p+1}和C_{p+2},并利用填充策略保持第$p+1$层和第$p+2$层的特征空间大小尺寸一致不变,即空间大小为$w\times w$。最后,利用残差函数对第p层和第$p+2$层进行连接。光谱残差网络模块结构可以表示为

$$X_{p+2}=X_p+F(X_p;\ r) \tag{9}$$

$$F(X_p;\ r)=X_{p+1}C_{p+2}+d_{p+2} \tag{10}$$

$$X_{p+1}=X_pC_{p+1}+d_{p+1} \tag{11}$$

式中,X_p代表第p层的特征,$F(\cdot)$代表特征提取的函数,$r=\{W_{p+1}, W_{p+2}, d_{p+1}, d_{p+2}\}$代表第$p+1$层和第$p+2$层卷积核和偏置参数的集合,$C$代表卷积核参数,$d$代表偏置参数。

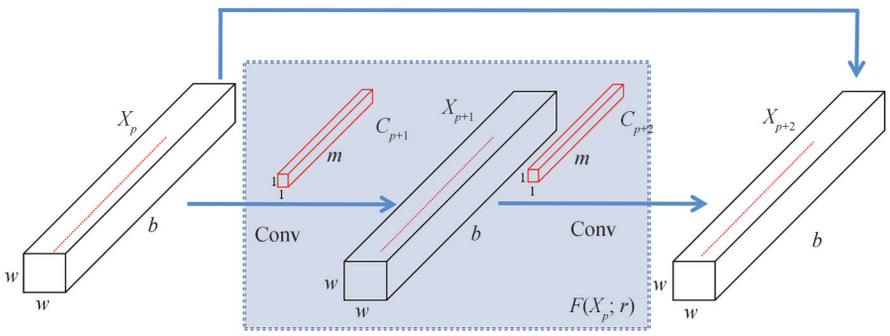

图5　光谱残差网络模块

Fig. 5　The spectral residual network module

空间残差网络模块如图6所示，残差块包括两个连续的卷积层和一个跳跃连接，跳跃连接可以保证第$q+2$层特征中包含有第q层的特征。对于第q层和第$q+1$层，分别使用尺寸为$a \times a \times b$的卷积核K_{q+1}和K_{q+2}，这些空间卷积核的光谱维度为b，等于输入特征图的光谱维度。利用填充策略保持第$q+1$层和第$q+2$层的特征空间大小尺寸一致不变，即空间大小为$w \times w$。最后，利用残差函数对第q层和第$q+2$层进行连接。因此，空间残差网络模块可以表示为

$$X_{q+2} = X_q + F(X_q; h) \tag{12}$$

$$F(X_q; h) = X_{q+1}K_{q+2} + l_{q+2} \tag{13}$$

$$X_{q+1} = X_q K_{q+1} + l_{q+1} \tag{14}$$

式中，X_q代表第q层的特征，$F(\cdot)$代表特征提取的函数，$h=\{K_{q+1}, K_{q+2}, l_{q+1}, l_{q+2}\}$代表第$q+1$层和第$q+2$层的卷积核和偏置参数，$K$代表卷积核参数，$l$代表偏置参数。

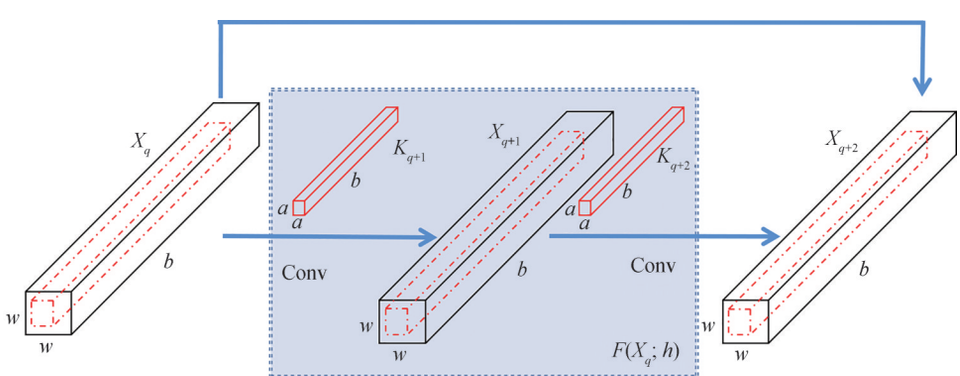

图6　空间残差网络模块

Fig. 6　The spatial residual network module

1.3　光谱-空间注意力残差网络

在上述内容基础上，提出了一个可以连续提取光谱和空间特征的高光谱分类网络，即图7所示的光谱-空间注意力残差网络（SSARN），该网络包括光谱特征学习模块、空间特征学习模块和分类器。其中，光谱特征学习模块包括光谱注意力和光谱残差网络；而空间特征学习模块包括空间注意力和空间残差网络。并且在网络中每个模块之间添加跳跃连接，将分层特征的表示层连接成为连续的残差块，以缓解精度下降的现象。

以Indian Pines（IP）数据集为例来解释所提出的SSARN网络。首先，将高光谱图像逐像素分割为一定尺寸的图像块，为方便说明，假定图像块尺寸大小为13×13，其光谱维度为200。该图像块经过中心区域光谱注意力后，光谱波段权重被重新调整，提高重要波段权重，降低不重要波段的权重。经过该注意力模块后，其图像块尺寸依然为13×13×200。中心区域选取范围为以中心像素为基准，周围3×3范围内为中心区域，一方面该范围内包含了一定相同类别的光谱信息，另一方面也尽可能减少不同类别像素的干扰。中心区域光谱注意力的计算方式如图2和式(3)～(5)所示。

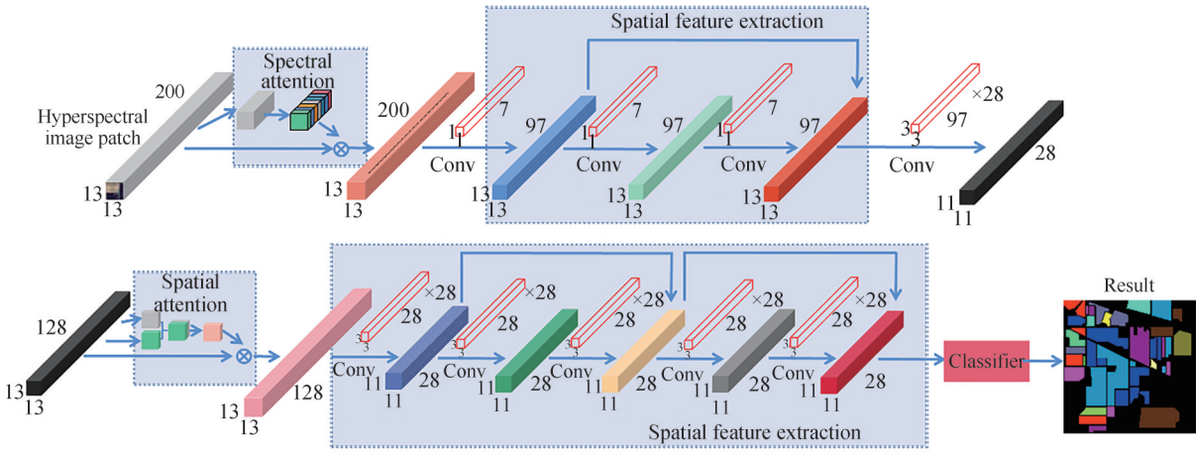

图 7 以 IP 数据集为例的 SSARN 流程

Fig. 7 The flow chart of SSARN with IP dataset as an example

光谱特征提取部分包括 1 个卷积层和 1 个光谱残差网络。在 HSI 中采用尺寸为 1×1×n 的三维卷积核提取光谱信息，不影响空间结构，保持了空间相关性。使用 1×1×7 的三维卷积作为光谱卷积核。该卷积核对经过中心区域光谱注意力机制的特征进行卷积，卷积步长为(1,1,2)。经过卷积层后，生成了 13×13×97 的光谱-空间特征。随后，该图像块被送入到光谱残差网络中提取光谱特征。光谱残差网络模块包含 2 个卷积层。在每个卷积层使用 1×1×7 的光谱卷积核来学习光谱特征。为了能够使用残差连接，需要保证输入和输出同样的尺寸，因此需要在卷积层中使用填充来保持相同的尺寸，填充尺寸统一为(0,0,3)。图像块经过光谱注意力和光谱残差网络模块后，网络已经提取到相应的光谱特征，该特征尺寸为 13×13×97，最后该特征输入到空间特征学习模块中。

空间特征学习模块包括 1 个空间注意力和 2 个空间残差网络模块。经过光谱特征学习后的空间-光谱特征输入到空间注意力模块中，进行空间权重重新校正，提高对中心像素的判别能力。空间注意力机制并不会改变特征的空间尺寸，因此经过空间注意力机制后的特征尺寸依然为 13×13×97。接着使用 28 个 13×13×97 的三维卷积核提取空间-光谱特征，同时降低空间尺寸和光谱尺寸；输出的光谱-空间特征为 11×11×28。在空间残差网络模块使用连续的二维卷积核提取空间判别特征，每层卷积均采用 28 个 3×3 的二维卷积核，同时为了保证残差网络模块前后尺寸统一，需要使用空间填充，填充尺寸为(1,1)。经过 4 个卷积层，2 个空间残差网络的特征学习，所提出的特征已经包含了丰富的光谱特征和空间特征。

将该特征放进分类器中，完成最后的分类任务。分类器包含平均池化层和全连接层，平均池化将提取 11×11×28 的光谱空间特征变成 1 个 1×1×28 的特征向量。接着全连接层根据每个数据集所包含的类别数生成一个输出向量，并选取最大值为预测结果。

2 实验结果

2.1 实验设置

本次实验选取的数据为三组公开的 Indian Pines (IP)数据集、Salinas (SA)数据集、Pavia University (PU)和 Houston 2013 标准划分数据集。各个数据集的假彩色图和真值图如图 8～11 所示。

IP 数据集每类随机选择 20% 的样本作为训练样本，SA 数据集每类随机选择 2% 的样本作为训练样本，PU 数据集每类随机选择 1% 的样本作为训练样本。随机按照比例选取样本，可以保留数据集本身的样本不均衡问题，有效验证算法在面对样本分布不均衡的性能。Houston 数据集有标准划分，因此按照标准划分进行训练和测试。各个数据集的训练样本和测试样本见表 1～4。

实验平台为 Pytorch 1.12，Python 3.9 和 Nvidia GTX 3090，24GB 图形处理器。所有算法的训练轮数设置为 100，每次训练输入 64 个图像块。损失函数、优化器都按照对比算法的最佳效果进行设置。所提出的光谱-空间注意力残差网络采用交叉熵损失函数，优化器为 Adam 优化器。初始学习率为 0.001，每 10 轮学习率调整为原来的 0.6 倍。

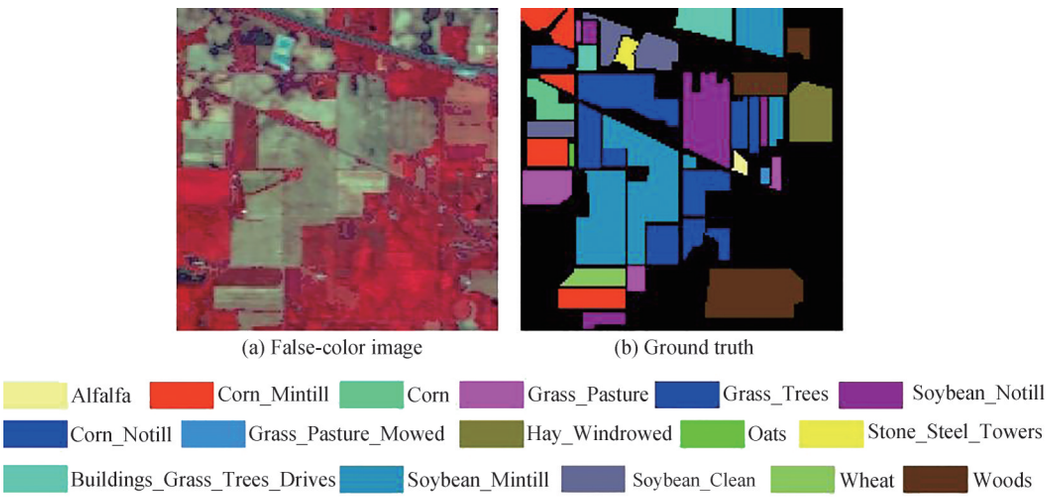

图 8　IP 数据集

Fig. 8　IP dataset

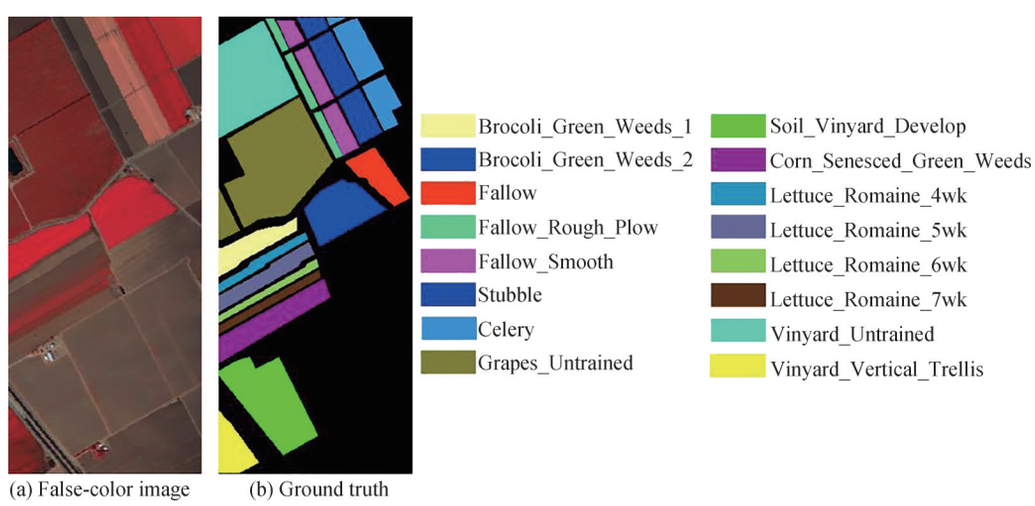

图 9　SA 数据集

Fig. 9　SA dataset

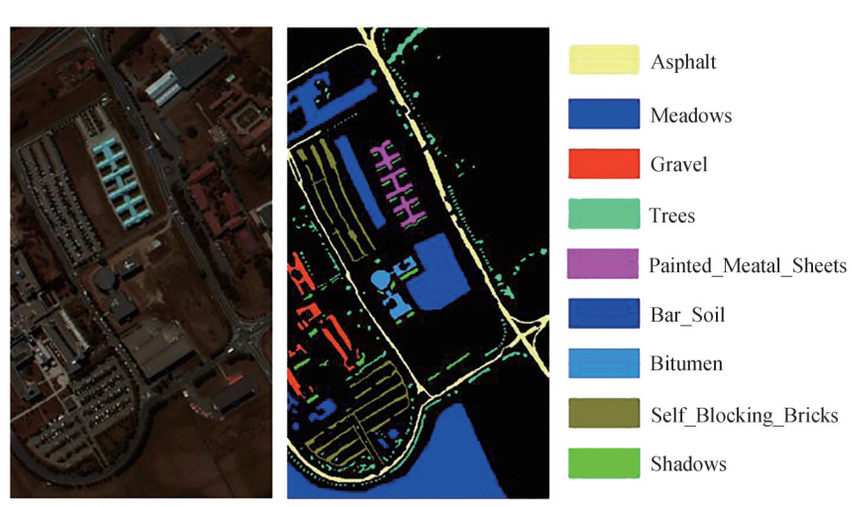

图 10　PU 数据集

Fig. 10　PU dataset

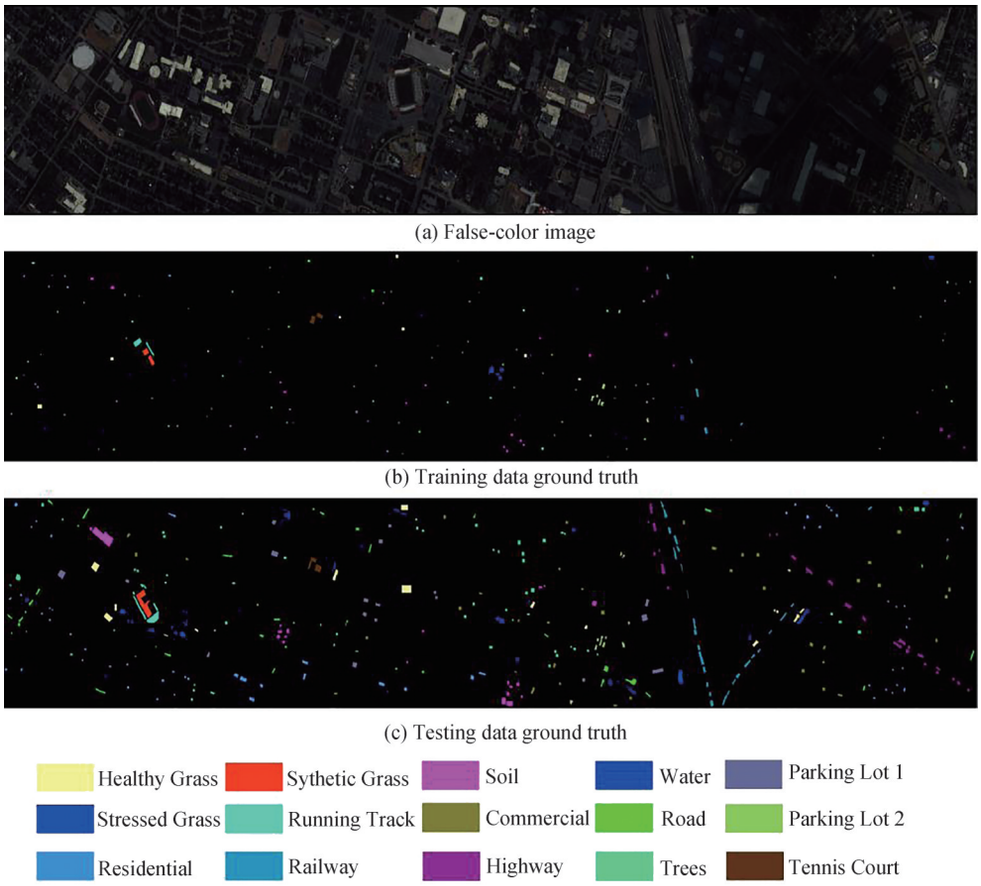

图 11 Houston 数据集

Fig. 11 Houston dataset

表 1 IP数据集的训练样本数量和测试样本数量

Table 1 The number of training and testing samples on IP dataset

Sample category	Sample name	Training samples	Test samples
0	Alfalfa	9	46
1	Corn-notill	285	1 428
2	Corn-mintill	166	830
3	Corn	47	237
4	Grass-pasture	96	483
5	Grass-trees	146	730
6	Grass-pasture-mowed	5	28
7	Hay-windrowed	95	478
8	Oats	4	20
9	Soybean-notill	194	972
10	Soybean-mintill	491	2 455
11	Soybean-clean	118	593
12	Wheat	41	205
13	Woods	253	1 265
14	Buildings-Grass-Trees-Drives	77	386
15	Stone-Steel-Towers	18	93
Total	——	2 045	10 249

表2 SA数据集的训练样本数量和测试样本数量

Table 2 The number of training and testing samples on SA dataset

Sample category	Sample name	Training samples	Test samples
0	Brocoli_green_weeds_1	40	2 009
1	Brocoli_green_weeds_2	74	3 726
2	Fallow	39	1 976
3	Fallow_rough_plow	27	1 394
4	Fallow_smooth	53	2 678
5	Stubble	79	3 959
6	Celery	71	3 579
7	Grapes_untrained	225	11 271
8	Soil_vinyard_develop	124	6 203
9	Corn_senesced_green_weeds	65	3 278
10	Lettuce_romaine_4wk	21	1 068
11	Lettuce_romaine_5wk	38	1 927
12	Lettuce_romaine_6wk	18	916
13	Lettuce_romaine_7wk	21	1 070
14	Vinyard_untrained	145	7 268
15	Vinyard_vertical_trellis	36	1 807
Total	——	1 076	54 129

表3 PU数据集的训练样本数量和测试样本数量

Table 3 The number of training and testing samples on PU dataset

Sample category	Sample name	Training samples	Test samples
0	Asphalt	66	6 631
1	Meadows	186	18 649
2	Gravel	20	2 099
3	Trees	30	3 064
4	Painted metal sheets	13	1 345
5	Bare Soil	50	5 029
6	Bitumen	13	1 330
7	Self-Blocking Bricks	36	3 682
8	Shadows	9	947
Total	——	423	42 776

评价指标为总体准确度（Overall Accuracy，OA）、平均准确度（Average Accuracy，AA）和Kappa系数。总体准确度（OA）表示正确分类的样本数占总样本数的比例，其公式为

$$\mathrm{OA}=\frac{\sum_i n_{ii}}{\sum_i N_i} \tag{15}$$

式中，n_{ij}代表图像中第i类样本预测标签为j的样本数目，n_{ii}代表i类样本中分类正确的样本数目，$N_i=\sum_j n_{ij}$代表第i类样本待分类样本的数目。

平均准确度（AA）表示每一类分类精度的平均值，其公式为

$$\mathrm{AA}=\frac{\sum_i \frac{n_{ii}}{N_i}}{k} \tag{16}$$

式中，k代表待分类样本的类别。

Kappa系数是用来衡量分类结果与真值地物之间一致性的指标。由于样本类别不均衡，OA、AA的指标会受到大样本精度影响。Kappa系数可以表示整个分类情况的偏差，代表分类与完全随机分类产生错误

表4 Houston数据集的训练样本数量和测试样本数量

Table 4 The number of training and testing samples on Houston dataset

Sample category	Sample name	Training samples	Test samples
0	Healthy Grass	198	1 053
1	Stressed Grass	190	1 064
2	Synthetic Grass	192	505
3	Trees	188	1 056
4	Soil	186	1 056
5	Water	182	143
6	Residential	196	1 072
7	Commercial	191	1 053
8	Road	193	1 059
9	Highway	191	1 036
10	Railway	181	1 054
11	Parking Lot 1	192	1 041
12	Parking Lot 2	184	285
13	Tennis Court	181	247
14	Running Track	187	473
Total	——	2 832	12 197

减少的比例，其公式为

$$\mathrm{Kappa}=\frac{\sum_i N_i\sum_i n_{ii}-\sum_i\left(\sum_j n_{ij}\cdot\sum_j n_{ji}\right)}{\left(\sum_i N_i\right)^2-\sum_i\left(\sum_j n_{ij}\cdot\sum_j n_{ji}\right)} \tag{17}$$

2.2 图像块尺寸

图像块尺寸选取过大，则需要较多的计算资源和时间成本。而图像块尺寸过小，又有可能使得网络不能够充分学习图像的空间特征，导致分类精度较低。因此，详细探索不同的图像块尺寸对总体分类准确度的影响。其分类结果见表5。

表5 不同图像块尺寸在四个数据集上的总体准确度

Table 5 The overall accuracy of the different size of the patch on the four datasets

Size	IP	SA	PU	Houston
9×9	99.57	98.23	98.20	81.04
11×11	99.69	99.02	98.46	85.35
13×13	**99.79**	99.69	**99.09**	85.75
15×15	99.50	99.65	98.87	85.78
17×17	99.51	99.68	**99.09**	**85.80**
19×19	99.38	**99.71**	98.95	84.16

通过表5可知，总体准确度总体上是根据尺寸大小先上升后下降。在IP数据集上，13×13的图像块精度最高；在SA数据集上，19×19的图像块精度最高，在较小尺寸的图像块上精度都有所下降。在PU数据集上，13×13和17×17的图像块精度一样，但在19×19时开始下降。在Houston数据集上，随着尺寸的增加，其精度不断提高，在17×17时达到最高精度。

对于IP数据集，其样本区域较为平滑，不同样本区域之间有交错但边缘区分较为明显，因此随着图像块尺寸的增大，其包含的空间信息越丰富，分类准确度也有所上升；当图像块尺寸超过一定尺寸时，有可能包含了更多的冗余空间信息，例如不属于同一类别的样本空间信息，反而会使分类精度下降。对于SA数据集，其样本区域较为规整，不同样本区域之间没有交错，当空间尺寸逐渐增大时，其精度会有提升。图像块尺寸越大，能提供的空间信息越丰富，越有利于提高分类精度。所以在图像块尺寸最大时，其精度最高。然

而，过大的尺寸会导致计算成本和计算资源成倍增长，因此需要平衡精度和计算资源来选取合适的图像块尺寸。对于PU数据集，其不同样本区域之间有交错。随着图像块尺寸增大，其总体准确度在上升，在尺寸为13×13时达到最大，后续基本保持不变。对于Houston数据集，各个样本区域比较分散，同一种样本分布也不集中；随着图像块尺寸增大，其包含的空间信息增多，总体分类精度在上升，尺寸在17×17时精度达到最高。而尺寸为13×13时，其精度比最高精度仅低了0.05%。

根据上述实验结果，从平衡计算资源和总体准确度出发，图像块尺寸统一为13×13。这样，一方面不需要过多的计算资源，另一方面还可以保持精度优势。

2.3 消融实验

为验证所提出的算法各个模块的有效性，在四个数据集上进行了消融实验，具体实验设置为：

基本网络：由1个光谱特征学习模块和2个空间特征学习模块构成。这些特征学习模块均采用了残差模块作为基础。

光谱注意力网络：由1个包含了中心光谱注意力机制的光谱特征学习模块和2个空间特征学习模块构成。也就是在基本网络的基础上，在光谱特征学习模块前加上中心光谱注意力机制。

光谱-空间注意力残差网络：由1个包含了中心光谱注意力机制的光谱特征学习模块和2个空间特征学习模块构成。在光谱特征提取结束后，空间特征学习前引入了空间注意力机制。

消融实验采取总体准确度（OA）作为评价指标，各个网络在四个数据集上的结果见表6。

表6 不同网络在四个数据集上的总体准确度

Table 6 The overall accuracy of the different network on the four datasets

Network	IP	SA	PU	Houston
Basic network	97.89	98.21	98.31	83.06
Spectral attention network	99.02	98.74	98.54	84.91
Spectral-spatial attention residual network	**99.79**	**99.69**	**99.09**	**85.75**

通过表6可以发现，相比基本网络，光谱注意力网络在IP、SA、PU和Houston数据集上，精度分别提升了1.13%、0.53%、0.23%和1.85%。说明光谱注意力机制可以有效地改变各个波段的权重，对分类结果影响较大的波段给予较高的权重，影响较小的波段给予较小的权重，而基本网络默认各个波段的权重相同，由于不同类别都有其容易识别的波段，而不是整个波段都可以用来进行分类[14]，意味着每个波段对待分类样本的影响程度不同。

光谱-空间注意力残差网络相比光谱注意力网络在IP、SA、PU和Houston数据集上，精度分别提升了0.77%、0.95%、0.55%和0.84%，比基本网络精度分别提升了1.9%、1.48%、0.78%和2.69%，说明空间信息对于分类结果的有一定影响。引入空间注意力机制可以有效地调整周围像素对中心像素的影响，具体来说，周围像素对待分类的中心像素有帮助时，其相应的权重就会提高，能有效地增强后续网络所提取的光谱-空间特征。而对待分类的中心像素没有帮助或者负面作用时，其权重则会降低。

综上，所提出的各个模块对最后的分类结果都有积极的影响，能够有效提高总体分类准确度。

2.4 实验结果

本次实验中，选取了2D CNN[22]、3D CNN[23]、HybridSN[24]、RIAN[12]、SSFTT[25]这5种方法作为对比算法，其中2D CNN、3D CNN、HybridSN、RIAN都是基于CNN的高光谱分类网络，而SSFTT是基于视觉变换网络（Vision Transformer，ViT）的高光谱分类网络，这些算法都是当前较为有代表性的算法。

2.4.1 不同训练比例对实验的影响

考虑到所用到的4个数据集中，只有Houston数据集给出了标准的训练集和测试集划分，其余3个数据集均没有标准划分，因此需要验证不同的训练集比例对各个算法的精度影响。在IP数据集中，训练集样本占全部样本的比例为5%、10%、15%和20%。在SA数据集中，训练集样本占全部样本的比例为0.5%、1%、1.5%和2%。在PU数据集中，训练集样本占全部样本的比例为0.3%、0.5%、0.7%和1%。各个算法在不同比例的训练集中的总体准确度表现如表7～9所示。

从表7～9中可以看出，随着训练比例提高，各个算法总体准确度都在上升。而SSARN在任何比例下都

表7 不同网络在IP数据集上的不同训练比例的总体准确度

Table 7 The overall accuracy of the different network with different training ratios on the IP datasets

IP dataset scale	5%	10%	15%	20%
2D CNN	65.49	71.18	80.17	82.85
3D CNN	73.61	84.20	90.79	93.23
HybirdSN	88.91	95.44	98.12	99.49
RIAN	87.65	93.87	96.75	97.82
SSFTT	94.98	98.19	99.30	99.45
SSARN	**96.09**	**98.56**	**99.43**	**99.79**

表8 不同网络在SA数据集上的不同训练比例的总体准确度

Table 8 The overall accuracy of the different network with different training ratios on the IP datasets

SA dataset scale	0.5%	1%	1.5%	2%
2D CNN	71.92	87.33	88.12	88.88
3D CNN	80.08	88.64	90.80	92.45
HybirdSN	93.27	95.55	98.17	98.44
RIAN	91.88	96.37	96.70	97.18
SSFTT	94.72	96.31	97.22	98.70
SSARN	**95.02**	**97.89**	**98.71**	**99.69**

表9 不同网络在PU数据集上的不同训练比例的总体准确度

Table 9 The overall accuracy of the different network with different training ratios on the IP datasets

PU dataset scale	0.3%	0.5%	0.7%	1%
2D CNN	76.13	82.92	84.86	89.22
3D CNN	76.94	82.21	85.10	86.24
HybirdSN	85.06	93.03	94.87	97.63
RIAN	76.82	89.98	91.74	94.03
SSFTT	86.99	94.78	96.65	97.24
SSARN	**93.93**	**97.46**	**98.15**	**99.09**

具有最高的精度，因此选择了各个算法精度最高的训练集比例，即IP数据集每类随机选择20%的样本作为训练样本，SA数据集每类随机选择2%的样本作为训练样本，PU数据集每类随机选择1%的样本作为训练样本作为统一比较的基础。

2.4.2 对比算法在各个数据集上的结果

表10展示了各个算法在IP数据集上的各类别准确度、总体准确度(OA)、平均准确度(AA)和Kappa值。表中所展示的Kappa值是在Kappa计算公式(17)的基础上乘以100进行展示。

通过表10可知，所提出的光谱-空间注意力残差网络SSARN，在AA、OA和Kappa系数上都取得了最佳的结果，并且在16个类别精度中有12个都达到了最好的效果，其中10个各类的精度为100%。这说明SSARN能够有效地学习不同类别的光谱特征和空间特征。在效果不好的4个类别中，其训练样本分别是285个、166个、194个和18个，相比类别最少的训练样本4个而言，其样本充足。也从侧面证明了SSARN可以有效地解决样本分布不均匀带来的在少样本上精度较差的效果。而对于上述4个效果较差的类别主要在两个不同样本区域的边缘，由于图像块包含了不同类别的样本，所学习的主要特征较少，最后分类的时候判断错误类别。

图12展示了各个算法在IP数据集上的分类效果。对比真值图(Ground Truth)，2D CNN和3D CNN分类效果较差，而HybridSN、RIAN、SSFTT、SSARN效果相对较好。2D CNN和3D CNN错误类别多集中在样本区域的内部，说明其对高光谱的空间特征没有有效地学习。而SSARN算法相比HybridSN、RIAN、SSFTT算法，判断错误的样本更少，更贴近真值图，说明该算法可以有效地学习高光谱图像的光谱特征和空间特征。

表10　不同算法在IP数据集上的类别准确度、OA、AA和Kappa

Table 10　The category accuracy, OA, AA and Kappa of the different algorithms on IP dataset

Sample category	2D CNN	3D CNN	HybridSN	RIAN	SSFTT	SSARN
0	43.48	84.78	**100.00**	95.62	95.65	**100.00**
1	75.28	91.81	99.72	96.29	**99.93**	99.86
2	74.22	88.67	**99.76**	96.63	99.40	99.28
3	59.92	91.14	94.94	94.94	97.89	**100.00**
4	93.17	94.62	98.14	95.03	99.38	**100.00**
5	98.08	99.45	**100.00**	99.45	99.45	**100.00**
6	57.14	60.71	96.43	92.86	**100.00**	**100.00**
7	93.31	**100.00**	**100.00**	99.79	**100.00**	**100.00**
8	50.00	95.00	**100.00**	95.00	**100.00**	**100.00**
9	76.95	90.43	98.66	97.48	**99.69**	99.38
10	85.17	93.36	99.71	98.45	99.02	**99.76**
11	62.56	82.97	99.33	96.46	98.82	**99.83**
12	99.02	**100.00**	**100.00**	98.54	99.51	**100.00**
13	95.02	97.15	**100.00**	99.84	100.00	**100.00**
14	80.83	96.37	**100.00**	97.41	100.00	**100.00**
15	78.49	93.55	**100.00**	98.92	98.92	**98.92**
AA	76.42	91.25	99.17	97.07	99.23	**99.81**
OA	82.85	93.23	99.49	97.82	99.45	**99.79**
Kappa	80.36	92.27	99.42	97.52	99.38	**99.76**

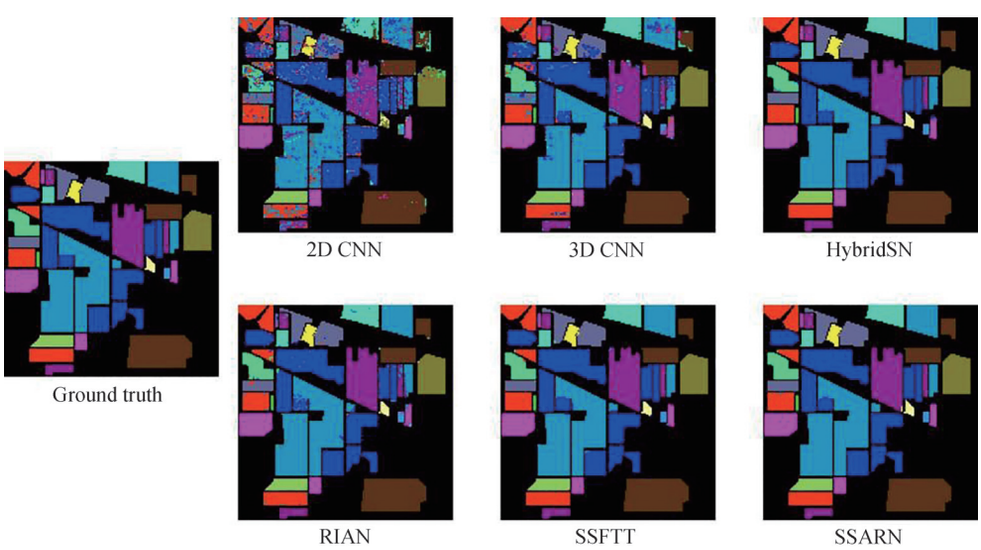

图12　各个对比算法在IP数据集的效果

Fig. 12　The visualization result of each algorithm on the IP dataset

表11展示了各个算法在SA数据集上的各类别准确度、总体准确度(OA)，平均准确度(AA)和Kappa值。

从表11可以看到，提出的SSARN在OA、AA和Kappa值上均达到了最优值，在16个类别精度中有12个都达到了最好的效果，其中8个类别精度为100%。这说明了该算法能够有效地学习不同类别的光谱特征和空间特征，而且面对不同数据集具有良好的泛化性。在效果相对不好的4个类别中，其精度也分别达到了98.64%、99.97%、98.93%和99.81%，相比最优效果，这4个类别精度仅仅低了0.86%，0.03%、0.25%和0.19%，差距并不明显。

图13展示了各个算法在SA数据集上的分类效果。对比真值图(Ground Truth)，2D CNN、3D CNN、

表11 不同算法在SA数据集上的类别准确度、OA、AA和Kappa

Table 11 The category accuracy, OA, AA and Kappa of the different algorithms on SA dataset

Sample category	2D CNN	3D CNN	HybridSN	RIAN	SSFTT	SSARN
0	98.41	96.67	**100.00**	99.85	**100.00**	**100.00**
1	97.26	99.60	**100.00**	99.76	**100.00**	**100.00**
2	94.33	99.44	**100.00**	99.54	99.80	**100.00**
3	98.57	98.42	**99.50**	98.28	98.14	98.64
4	97.87	97.80	98.02	98.47	97.42	**99.96**
5	98.36	**100.00**	**100.00**	99.92	**100.00**	**100.00**
6	97.26	97.09	**100.00**	99.89	**100.00**	99.97
7	79.3	86.86	96.73	92.91	97.32	**99.35**
8	99.19	97.11	**100.00**	98.94	99.90	**100.00**
9	79.44	92.04	98.87	**99.18**	96.71	98.93
10	92.13	91.67	**100.00**	96.44	98.60	99.81
11	99.95	99.48	99.90	99.95	99.90	**100.00**
12	95.63	99.78	99.56	98.91	**100.00**	**100.00**
13	95.70	97.01	98.88	97.66	**99.63**	**99.63**
14	70.20	77.45	95.18	94.30	97.84	**99.52**
15	93.97	93.69	99.45	97.12	99.28	**100.00**
AA	92.96	95.26	99.13	98.20	99.03	**99.74**
OA	88.88	92.45	98.44	97.18	98.70	**99.69**
Kappa	87.61	91.59	98.26	96.86	98.55	**99.65**

HybridSN、RIAN、SSFTT的分类效果都不如SSARN。SSARN分类错误的样本主要是第8类，会被错误地分为第14类，一方面是空间位置上这2类较近，另一方面其他算法错误的分类也集中在第8类，说明网络所提取第8类的光谱特征与第14类的光谱特征较为相近，进而出现了类别误判。

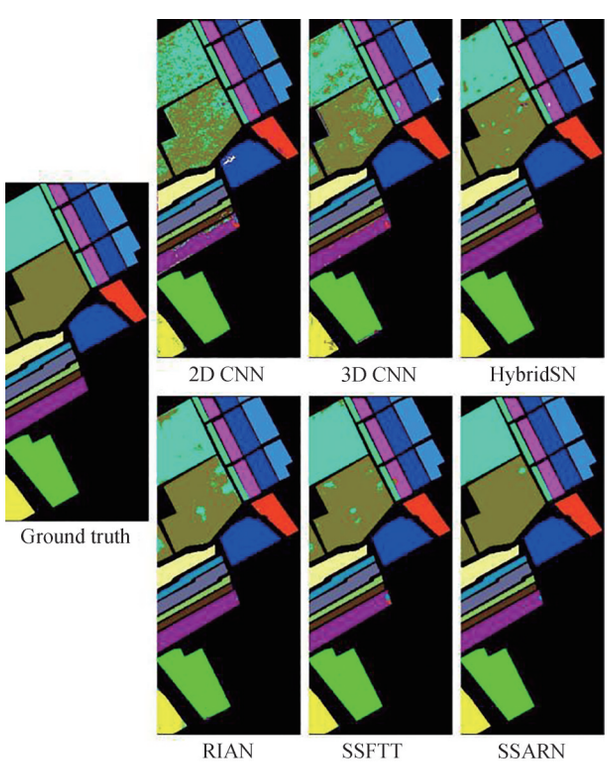

图13 各个算法在SA数据集的效果

Fig. 13 The visualization result of each algorithm on the SA dataset

表12展示了各个算法在PU数据集上的类别准确度、总体准确度（OA）、平均准确度（AA）和Kappa值。

表12　不同算法在PU数据集上的类别准确度、OA、AA和Kappa

Table 12　Category accuracy, OA, AA and Kappa of the different algorithms on PU dataset

Sample category	2D CNN	3D CNN	HybridSN	RIAN	SSFTT	SSARN
0	92.34	83.40	97.62	93.83	96.00	**100.00**
1	97.42	99.36	99.83	98.05	99.84	**99.98**
2	59.27	48.26	83.99	67.37	92.85	**98.71**
3	73.60	90.57	**97.55**	95.72	97.49	94.13
4	98.44	84.54	**100.00**	99.85	**100.00**	99.85
5	76.91	66.81	97.95	92.56	98.11	**100.00**
6	79.85	70.23	**97.89**	75.79	81.88	94.81
7	87.78	74.71	93.70	96.44	92.18	**97.58**
8	93.77	91.13	94.72	85.64	96.30	**97.99**
AA	84.37	78.78	95.92	89.47	94.96	**98.12**
OA	89.22	86.24	97.63	94.03	97.24	**99.09**
Kappa	85.47	81.46	96.85	92.08	96.34	**98.79**

从表12可以看到，SSARN在OA、AA和Kappa值上均达到了最优值，在9个类别精度中有6个都达到了最好的效果，其中2个类别精度为100%。这说明了该算法能够有效地学习不同类别的光谱特征和空间特征，而且面对不同数据集具有良好的泛化性。在效果相对不好的3个类别中，其精度也分别达到了94.13%、99.85%和94.81%，和最优的效果相比，分别低了3.42%、0.15%和3.08%，主要是第3类和第6类表现较差。分析其主要原因是第3类和第6类分散在全局中，集中区域较少，并且训练时选取的样本量也较少，因此网络提取特征时丢失了部分细节特征，从而导致其精度偏低。

图14展示了各个算法在PU数据集上的分类效果。对比真值图（Ground Truth），2D CNN、3D CNN、

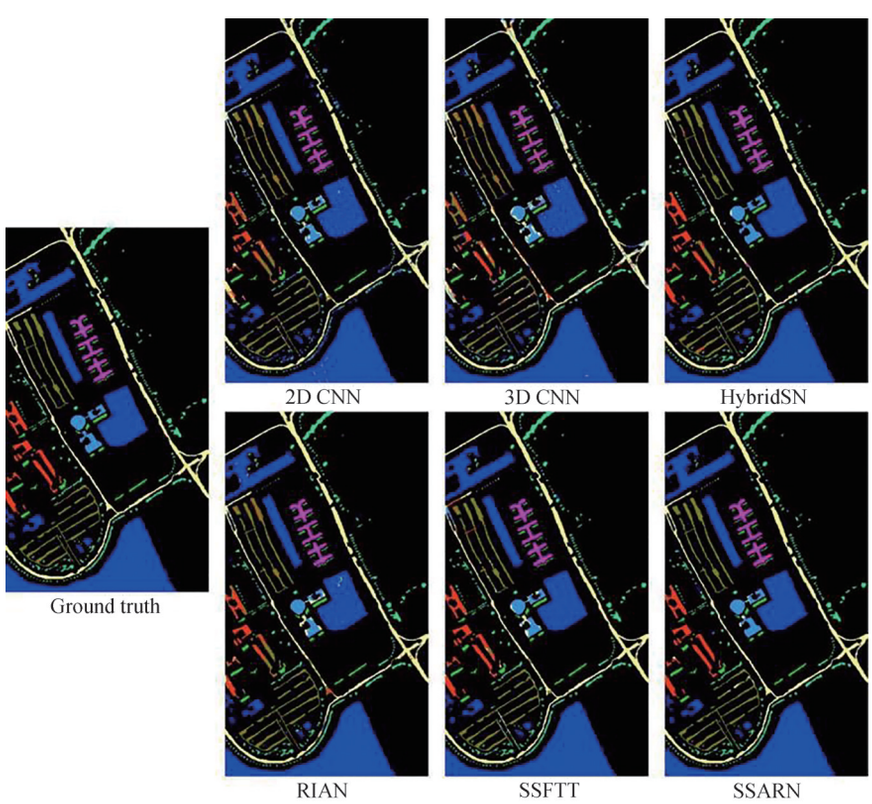

图14　各个算法在PU数据集的效果

Fig. 14　The visualization result of each algorithm on the PU dataset

HybridSN、RIAN、SSFTT的分类效果都不如SSARN。SSARN算法判断错误的样本更少，更贴近真值图。

表13展示了各个算法在Houston数据集上的类别准确度、总体准确度(OA)、平均准确度(AA)和Kappa值。

表13 不同算法在Houston数据集上的类别准确度、OA、AA和Kappa

Table 13 The category accuracy, OA, AA and Kappa of the different algorithms on Houston dataset

Sample category	2D CNN	3D CNN	HybridSN	RIAN	SSFTT	SSARN
0	82.72	82.53	72.93	81.29	82.53	**82.62**
1	84.21	82.05	81.96	58.55	84.77	**85.15**
2	97.82	92.28	85.15	88.32	85.94	**100.00**
3	91.29	91.57	72.25	80.21	**92.99**	91.67
4	98.20	99.24	98.49	83.62	99.72	**100.00**
5	94.41	92.31	77.62	60.84	94.41	**95.80**
6	75.75	75.19	66.33	66.51	83.86	**86.38**
7	66.95	56.51	73.31	45.11	65.43	**88.60**
8	73.47	66.95	50.05	58.83	74.41	**81.87**
9	44.79	50.77	**100.00**	20.17	51.74	47.88
10	78.18	73.91	**86.34**	35.48	74.38	81.50
11	77.91	72.33	80.31	51.30	78.48	**93.47**
12	84.21	81.75	65.61	45.61	**89.83**	85.26
13	98.79	96.76	**100.00**	76.52	99.19	**100.00**
14	**100.00**	82.45	**100.00**	86.68	94.93	**100.00**
AA	83.25	79.77	80.69	62.60	83.51	**88.01**
OA	79.92	76.95	79.41	60.66	80.66	**85.75**
Kappa	78.37	75.18	77.64	57.58	79.10	**84.57**

从表13可以看到，SSARN在OA、AA和Kappa值上均达到了最优值，在15个类别精度中有11个都达到了最好的效果，其中3个类别精度为100%。这说明该算法能够有效地学习不同类别的光谱特征和空间特征，而且面对不同数据集具有良好的泛化性。在效果相对不好的4个类别中，相比最优的效果，精度低了1.32%、52.12%、4.84%和4.57%。效果最差的是类别9，即高速公路(Highway)这一类别。绝大多数对比算法在这一类表现都很差，而对比算法HybirdSN在这一类达到了100%的精度，分析主要原因可能是数据预处理阶段，只有HybirdSN利用PCA进行数据降维，保留了主要的光谱特征，而其他算法都是在原始的光谱维度上进行特征学习，冗余的光谱波段会导致网络学习该类别的光谱能力下降，进而导致在该类别上精度下降较多。

图15展示了各个算法在Houston数据集上的分类效果。对比真值图(Ground Truth)，2D CNN、3D CNN、HybridSN、RIAN、SSFTT的分类效果都不如SSARN。SSARN分类错误的样本主要是第9类，其他算法错误的分类也集中在第9类。冗余的光谱

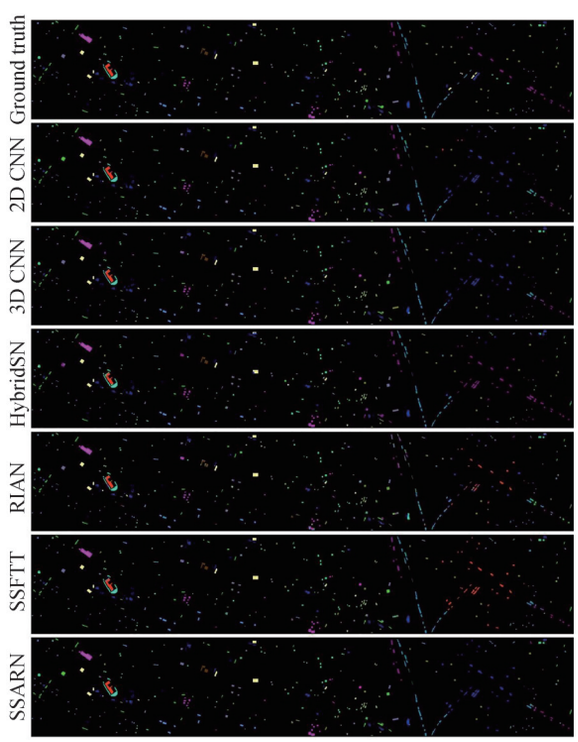

图15 各个对比算法在Houston数据集的效果

Fig. 15 The visualization result of each algorithm on the Houston dataset

特征会导致网络对某些类别产生过拟合，进而导致其分类精度下降。总体来看SSARN算法相比其他算法，判断错误的样本更少，更贴近真值图。

3 结论

本文提出了基于光谱-空间注意力残差网络（SSARN）的高光谱分类方法，该方法利用中心区域光谱注意力机制，在保留全部光谱信息的基础上，对光谱之间的权重进行了调整，避免了网络提取光谱特征时认为各个通道权重一致的问题。重新调整光谱权重后，采用了残差网络对光谱维度进行特征提取，一方面可以有效地提取和保留光谱信息，另一方面便于优化网络。在提取光谱特征后，利用空间注意力机制对空间-光谱特征进行学习，使后续的空间特征学习模块更多地关注输入中的相关空间特征，尽可能多地提取有用的空间特征来帮助分类。空间特征学习模块采用2个类似的残差特征提取模块，主要是因为空间信息相比光谱信息更多，需要更多的网络参数进行学习。在4个公开的数据集上，消融实验证明了各个模块的有效性。和常用以及最新算法相比，所提出的SSARN在所有数据集上都达到了最好效果，也证明了该网络的有效性和鲁棒性。

但是该算法面对分散样本时，其特征提取能力以及判别能力出现了下降，一方面是因为所提出的网络都是基于图像块输入的，对上下文信息的获取较差；另一方面类别分散在全局中，样本比例的不均衡性会导致网络提取特征时，分散类别样本特征权重较低，可能导致其重要的特征丢失。考虑到视觉变换模型对图像全局信息的把握能力更强，后续可以考虑用Transformer网络，并根据样本不均衡引入动态权重调整系数调整小样本的特征权重来解决全局分散样本精度较低的问题。

参考文献

[1] RASTI B, HONG Danfeng, HANG Renlong, et al. Feature extraction for hyperspectral imagery: the evolution from shallow to deep(overview and toolbox)[J]. IEEE Geoscience and Remote Sensing Magazine, 2020, 8(4): 60-88.

[2] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.

[3] HU Wei, HUANG Yangyu, WEI Li, et al. Deep convolutional neural networks for hyperspectral image classification[J]. Journal of Sensors, 2015, 2015: 1-12.

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

[5] ZHONG Zilong, LI Ying, MA Lingfei, et al. Spectral-spatial transformer network for hyperspectral image classification: a factorized architecture search framework[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-15.

[6] GHADERIZADEH S, ABASIMOGHADAM D, SHARIFI A, et al. Hyperspectral image classification using a hybrid 3D-2D convolutional neural networks[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 7570-7588.

[7] WU Hao, PRASAD S. Convolutional recurrent neural networks for hyperspectral data classifification[J]. Remote Sensing, 2017, 9(3): 298.

[8] ZHONG Zilong, LI J, LUO Zhiming, et al. Spectral-spatial residual network for hyperspectral image classification: a 3-D deep learning framework[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 56(2): 847-858.

[9] SHI Yuetian, FU Bin, WANG Nan, et al. Spectral-spatial residual network for hyperspectral image classification: a 3-D deep learning framework[J]. Drones, 2023, 7(4): 1-30.

[10] XU Yue, GONG Jianya, HUANG Xin, et al. Luojia-HSSR: a high spatial-spectral resolution remote sensing dataset for land-cover classification with a new 3D-HRNet[J]. Geo-spatial Information Science, 2022: 1-13.

[11] YANG Kai, SUN Hao, ZOU Chunbo, et al. Cross-attention spectral-spatial network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-14.

[12] ZHENG Xiangtao, SUN Hao, LU Xiaoqiang, et al. Rotation-invariant attention network for hyperspectral image classification[J]. IEEE Transactions on Image Processing, 2022, 31: 4251-4265.

[13] FANG Shuai, ZHANG Kun, ZHANG Jing, et al. Hyperspectral image classification with enhanced class separability[J]. Journal of Image and Graphics, 2021, 26(8): 1940-1951.
方帅，张坤，张晶，等. 增强类可分性的高光谱图像分类[J]. 中国图象图形学报, 2021, 26(8): 1940-1951.

[14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]. Advances in Neural Information Processing Systems, 2017, 30: 1-11.

[15] LUO Fulin, ZHANG Liangpei, ZHOU Xiaocheng, et al. Sparse-adaptive hypergraph discriminant analysis for

hyperspectral image classification[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 17(6): 1082-1086.
[16] VEITi A, WILBER M J, BELONGIE S. Residual networks behave like ensembles of relatively shallow networks[C]. Advances in Neural Information Processing Systems, 2016: 550-558.
[17] HUANG Gao, LIU Zhuang, MAATEN L D M, et al. Densely connected convolutional networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.
[18] HE Kaiming, ZHANG Xiangyu, REN Shaoping, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
[19] HE Kaiming, ZHANG Xiangyu, REN Shaoping, et al. Identity mappings in deep residual networks[C]. European Conference on Computer Vision, 2016: 630-645.
[20] PAOLETTI M E, HAUT J M, FERNANDEZ B R, et al. Deep pyramidal residual networks for spectral-spatial hyperspectral image classifification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(2): 740-754.
[21] WANG Wenju, DOU Shuguang, Jiang Zhongmin, et al. A fast dense spectral-spatial convolution network framework for hyperspectral images classifification[J]. Remote Sensing, 2018, 10(7):1068.
[22] LUO Yanan, ZOU Jie, YAO Chengfei, et al. HSI-CNN: a novel convolution neural network for hyperspectral image [C]. In Proceedings of the 2018 International Conference on Audio, Language and Image Processing, 2018: 464-469.
[23] LI Ying, ZHANG Haokui, SHEN Qiang. Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network[J]. Remote Sensing, 2017, 9(1): 67.
[24] ROY S K, KRISHNA G, DUBEY S R, et al. HybridSN: Exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 17(2): 277-281.
[25] SUN Le, ZHAO Guangrui, ZHENG Yuhui, et al. Spectral-spatial feature tokenization transformer for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-14.

Spectral-spatial Attention Residual Networks for Hyperspectral Image Classification

WANG Feifei[1,3], ZHAO Huijie[1,2,3], LI Na[1,2,3], LI Siyuan[4], CAI Yu[5]
(1 *Key Laboratory of Precision Opto-Mechatronics Technology, Ministry of Education, School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing* 100191, *China*)
(2 *Institute of Artificial Intelligence, Beihang University, Beijing* 100191, *China*)
(3 *Aerospace Optical-Microwave Integrated Precision Intelligent Sensing, Key Laboratory of Ministry of Industry and Information Technology, Beihang University, Beijing* 100191, *China*)
(4 *Key Laboratory of Spectral Imaging Technology, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an* 710119, *China*)
(5 *China Academy of Launch Vehicle Technology, Beijing* 100076, *China*)

Abstract: Hyperspectral image classification is a research hotspot in the field of hyperspectral image processing and application. Classification models predict the class of each pixel by analyzing the spectral and spatial information of each pixel and compare it to the actual features. In the hyperspectral classification task, the spatial context information of the data can be used to improve the classification accuracy, so this paper uses the powerful learning ability of 3D-CNN to extract effective spectral and spatial features into hyperspectral images, and then fuses the extracted spectra and spatial features to enhance the flow between different levels of the network, thereby improving the classification efficiency. Although CNN operations can mine deeper feature information as the network deepens, CNN is ineffective in modeling long-distance dependencies, so consider combining CNN with attention mechanisms. This combination can focus on the local position of the given information, assign corresponding weights to it, emphasize the key features in the feature map, adjust the global information of the attention statistics image through weight re-annotation, retain the features that are more conducive to the classification task, and improve the representation ability of extracted features. But the common attention mechanism is to calculate the average globally, that is, the pixel values of the entire image block, inevitably introducing information from different categories of pixels around it, which is not needed in classification tasks. Another spectral attention mechanism based on the center pixel provides weight values that ignore the effects of surrounding pixels in the same category. Therefore, a simple spectral attention mechanism in the central region is

proposed, in which the central region is selected with the central pixel as the reference and the surrounding 3×3 range as the central region, on the one hand, the range contains certain spectral information of the same category, and on the other hand, the interference of different categories of pixels is reduced as much as possible. The spectral attention mechanism in the central region can minimize the influence of interfering pixels on spectral features while extracting as many effective spectral features as possible. Based on the spectral attention mechanism of the central region, this paper proposes a spectral spatial attention residual network for hyperspectral classification, which mainly includes spectral feature learning, spatial feature learning and classifier. The network first selects appropriately sized image blocks from hyperspectral images and then classifies them. Starting from balancing computing resources and overall accuracy, experimental comparison shows that the size of the image patch is uniformly 13×13. The spectral feature learning part includes 1 frequency spectral attention module and 1 spectral residual network module. The spectral attention module adopts the central spectral attention mechanism, which can effectively suppress redundant bands and increase the weight of important bands. The spectral features after the attention mechanism will be extracted by the spectral residual network module, and more spectral features can be extracted. Convolution kernels of $1\times1\times n$ do not affect the spatial structure when extracting spectral features while maintaining spatial correlation. The spatial feature learning component includes 1 spatial attention module and 2 spatial residual network modules. The spatial attention module can obtain the important spatial information of the pixels to be classified, and use the spatial residual network to extract its spatial information. Add a hop connection between each module in the network to connect the presentation layer of the hierarchical features into a continuous residual block to mitigate the loss of accuracy. Finally, these rich spectral and spatial features are sent to the classifier to obtain the final classification result. The proposed algorithm is compared with the latest algorithm on four public datasets. Indicators and visualization results verify the superiority of the proposed algorithm.

Key words: Spectral-spatial feature; Residual network; Hyperspectral image classification; Spectral attention mechanism; Spatial attention mechanism

OCIS Codes: 100.4145; 100.6890; 200.3050; 330.5000

Foundation item: National Natural Science Foundation of China (No. 61975004), Pre-Research Project (No. 6230111002)