

基于光谱—空间注意力双边网络的高光谱图像分类

杨星, 池越, 周亚同, 王杨

河北工业大学 电子与信息工程学院, 天津 300401

摘要: 在过去几年里, 卷积神经网络已经在高光谱图像分类上取得良好的效果, 然而高光谱图像的高维性和卷积神经网络对所有波段的平等处理, 限制了这些方法性能。本文提出了一种端到端的光谱空间注意力双边网络 SSABN (Spectral-Spatial Attention Bilateral Network), 直接将原始图像 3D 块作为输入数据, 而不需要进行预处理。首先, 通过光谱空间注意力模块从原始数据中增强有用波段, 抑制无效波段。然后, 设计双边网络两条路径。其中, 空间路径用于提取空间信息, 上下文路径用于提供更大的感受野, 并通过特征融合模块有效的结合特征。实验结果表明, SSABN 在 3 个公开数据集上取得了更高的分类精度, 同时有效的减少训练时间。

关键词: 遥感, 卷积神经网络, 深度学习, 特征融合, Indian Pine 数据, Pavia University 数据, Salinas 数据集

中图分类号: P2

引用格式: 杨星, 池越, 周亚同, 王杨. 2023. 基于光谱—空间注意力双边网络的高光谱图像分类. 遥感学报, 27(11): 2565–2578

Yang X, Chi Y, Zhou Y T and Wang Y. 2023. Spectral-spatial attention bilateral network for hyperspectral image classification. National Remote Sensing Bulletin, 27(11): 2565–2578 [DOI: 10.11834/jrs.20210563]

1 引言

随着遥感技术的快速发展, 高光谱成像技术图像同时具有丰富光谱信息和详细的空间结构, 高光谱图像每个像素包含数百个连续光谱带, 其波长范围跨越可见光和红外波段。相较于其他遥感图像, 高光谱图像 HSI (Hyper Spectral Image) 由于其丰富的信息量, 对于目标物理、化学等特性有更好的表征能力 (Li 等, 2016), 高光谱图像被广泛应用于环境管理 (Chen 等, 2019)、农业和资源管理 (Murphy 等, 2018)、城市开发 (Ghamisi 等, 2015)、海洋观测 (曹引 等, 2019)、军事 (Shimoni 等, 2019) 等多种领域。高光谱图像分类 HSIC (Hyper Spectral Image Classification) 是 HSI 处理与应用中最为基础也是最为重要的一环, 为遥感数据的后续研究提供了科学依据。HSIC 的目的是将 HSI 中每一个像素分类到给定的地物类别中, 如植被和农作物。

在 HSIC 研究早期, 由于 HSI 光谱反射曲线良好的判别性, 不同地物类别的光谱曲线不相同。

研究者将传统机器学习方法加以改进并用于 HSIC, 包括支持向量机 SVM (Support Vector Machine) (Melgani 和 Bruzzone, 2004)、K 均值聚类 (Ma 等, 2010)、随机森林 (Ham 等, 2005)、极限学习机 (Zhou 等, 2015) 等。这些传统方法利用光谱曲线构造特征提取模型, 提取判别性特征。然而不同波段之间具有较强的相关性且存在大量的冗余信息, 过高的光谱分辨率所带来的更多计算量, 都可能导致维数灾难, 使得这些传统方法存在一些局限性。为了避免维数灾难, 常采用特征选择和特征提取两个方法对光谱波段进行降维。特征选择是选择一部分对结果更有效的光谱波段。Sun 和 Du (2018, 2019) 提出了一些波段选择方法来减少波段之间的冗余。然而, 对于不同类别的像素, 每个光谱波段有不同的贡献, 每个类都有它的最优分类光谱波段子集, 而这些子集对不同的类可能是不相同的。例如波段 1, 6, 10, ..., 对类别 1 是有用的, 而波段 2, 7, 9, ..., 对类别 2 是有用的。特征提取是通过线性或者非线性变换将这些光谱波段结合在一起。主成分分析 (PCA)

收稿日期: 2021-01-06; 预印本: 2021-03-06

基金项目: 河北省自然科学基金 (编号: F2019202364); 河北省教育厅资助重点项目 (编号: ZD2020304)

第一作者简介: 杨星, 研究方向为深度学习在高光谱图像上的应用。E-mail: yx374448357@163.com

通信作者简介: 池越, 研究方向为图像处理。E-mail: chiyeuexin@126.com

(Rodarmel 和 Shan, 2002)、独立成分分析 (ICA) (Wang 和 Chang, 2006) 和 Fisher 线性判别分析 (LDA) (Luo 等, 2015) 等是最常用的几种降维方法。主成分分析的方法是提取出前 k 个主成分作为光谱特征。但参数 k 需要通过专家经验手动设置, 合适的 k 会将所有有用的特征保存, 不合适的参数可能会丢失一些有用信息或者保留较多的冗余信息。传统的特征提取方法的缺点是需要手工制作特征, 依赖于设计者或领域专家的先验信息, 对于不同的数据集适应和泛化能力较差, 无法处理复杂和新的情况。

由于深度学习可以提取更高层次和更抽象的特征, 研究者们开始研究如何利用深度学习解决 HSIC。这些方法包括堆叠自编码器 SAEs (Stacked Autoencoders) (Deng 等, 2019; Mei 等, 2019a)、深度信念网络 DBNs (Deep Belief Networks) (Chen 等, 2015)、卷积神经网络 CNNs (Convolution Neural Networks) (Zhong 等, 2018; Wang 等, 2018; Mei 等, 2019b)、循环神经网络 RNNs (Recurrent Neural Networks) (Zhang 等, 2018; Mou 等, 2017) 和生成对抗网络 GAN (Generative Adversarial Networks) (Zhu 等, 2018)。Chen 等 (2014) 等首次使用多层 SAE 为 HSI 提取深度特征, 在深度网络最后一层加入 logistic 回归分类器, 利用标记样本对整个网络进行微调。Li (2015) 等使用单一限制玻尔兹曼机和多层 DBN 为 HSI 提取频谱空间特征。基于 SAE 和 DBN 的方法主要问题是将图像平展成向量, 无法在空间特征提取阶段考虑空间信息。而 CNN 可以很好的缓解上述问题, 因此将 CNN 引入 HSIC。Zhang 等 (2018) 等提出一种光谱-空间残差网络 (SSRN), 该网络使用光谱和空间残差块, 从 HSI 学习深度判别特征。然而其使用原始 HSI 三维块作为输入, 不加任何处理, 导致计算量巨大, 迭代速度较慢。而其他一些 CNN 的方法则是使用降维之后的三维块, 丢失一部分光谱信息。同时这些 CNN 对每个像素和光谱进行相同的处理, 而由于混合像素问题和冗余波段问题, 使得这样的处理方法同样具有局限性。

由于注意力机制可以自适应的抑制或者增强输入信息, 已经广泛应用于图像处理 (Vaswani 等, 2017), 很多学者都尝试在 HSIC 中结合注意力机制来关注更为重要的特征。Li (2020) 等设计 DBDA (Double-Branch Dual-Attention Mechanism) 两个分

支, 以通过不同注意力结构提取 HSI 中大量光谱和空间特征, 然而其复杂的结构导致需要更多的训练时间。Roy 等 (2020) 结合 SENet (Squeeze and Excitation Network) 和残差网络 (ResNet) 提出 FuSENet (Fused Squeeze-and-Excitation Network), 其通过 SENet 压缩和激发得到通道的权重值, 提取出每个特征图的重要性。Das 等 (2020) 设计一个 Multi-Receptive Lightweight Residual 模块 (GhoMR), 并提出 GhoMR-Net。GhoMR 模块通残差结构提取特征, 多接受域提取其权重, 以分层的方式提取更为重要的特征。这些注意力机制多关注于光谱或空间信息, 不能很好地体现 HSI 光谱合一的特点。

虽然已经有许多基于卷积神经网络的高光谱图像分类模型被提出, 但仍然存在一些问题。

(1) 如何自适应的强调对分类有效的光谱信息和空间信息;

(2) 如何充分提取 HSI 不同层次的特征, 并且从不同层次特征中提取出判别性特征;

(3) 如何设计一个端到端的网络, 在保证分类精度的同时, 有效加快网络训练速度。

针对这些问题, 我们借鉴注意力机制, 提出一种用于 HSIC 的端到端的空间光谱注意力双边网络 SSABN (Spectral-Spatial Attention Bilateral Network)。通过光谱注意力模块自适应的学习不同原始输入数据中每个光谱波段的权值, 将特征选择隐式的应用于光谱向量上。在空间特征方面, 由于与中心像素相同类别的周围像素对分类贡献大, 不同类别的周围像素贡献小, 空间注意力模块将根据像素对中心像素分类的贡献, 自适应的学习其重要性, 提高相同类别的周围像素权重, 削弱不同类别的周围像素权重。双边网络通过不同的卷积结构, 提取不同层次的特征, 通过特征融合模块将不同层次特征融合得到判别性的特征。

2 光谱空间注意力双边网络

本文提出一种光谱空间注意力双边网络模型用于高光谱图像分类。本节首先介绍提出的高光谱图像分类模型。其次介绍所提出双边网络中的特征融合模块, 并详细说明了光谱空间注意力模块。最后给出了该模型的损失函数及其优化方法。

2.1 提出模型概述

设高光谱数据集 $H \in R^{h \times w \times d}$, 其中空间维度

的高度和宽度用 h, w 表示, 光谱带的数量用 d 表示。假设数据集 H 包含 N 个标记像素 $U = \{u_1, u_2, \dots, u_n\} \in \mathbf{R}^{1 \times 1 \times d}$, 每个标记像素由 d 个光谱波段构成, 其对应的独热编码标签向量为 $V = \{v_1, v_2, \dots, v_n\} \in \mathbf{R}^{1 \times 1 \times k}$, 其中 k 是类别数。为了充分利用数据中的原始信息, 以标记像素 U 为中心的相邻立方体组成一组新的三维块 $W = \{w_1, w_1, \dots, w_n\}$ 。在本文中, 我们将每个 W 中的 w_i 输入到所提出的模型中, 对其中心对应像素 u_i 进行分类。

对高光谱数据标记后, 首先将所有可用的标

记数据随机分为训练、验证和测试数据集, 分别用 $w_{\text{train}}, w_{\text{val}}, w_{\text{test}}$ 表示, 对应的标签集为 $v_{\text{train}}, v_{\text{val}}, v_{\text{test}}$ 。然后使用 $w_{\text{train}}, w_{\text{val}}$ 优化和测试模型, 通过交叉验证找到模型的最佳参数。最后通过 w_{test} 对所有像素进行分类, 得到性能评价指标, 形成分类图。

图1给出了所提出模型的框架。该模型主要有4个模块。其中图1(A)表示光谱空间注意力模块用于寻找有效的光谱和空间部分。空间路径上下文路径为双边网络模型两条不同CNN路径, 分别由图1(B)和图1(C)所示。图1(D)所示的特征融合模块用于融合双边网络不同路径的输出特征。

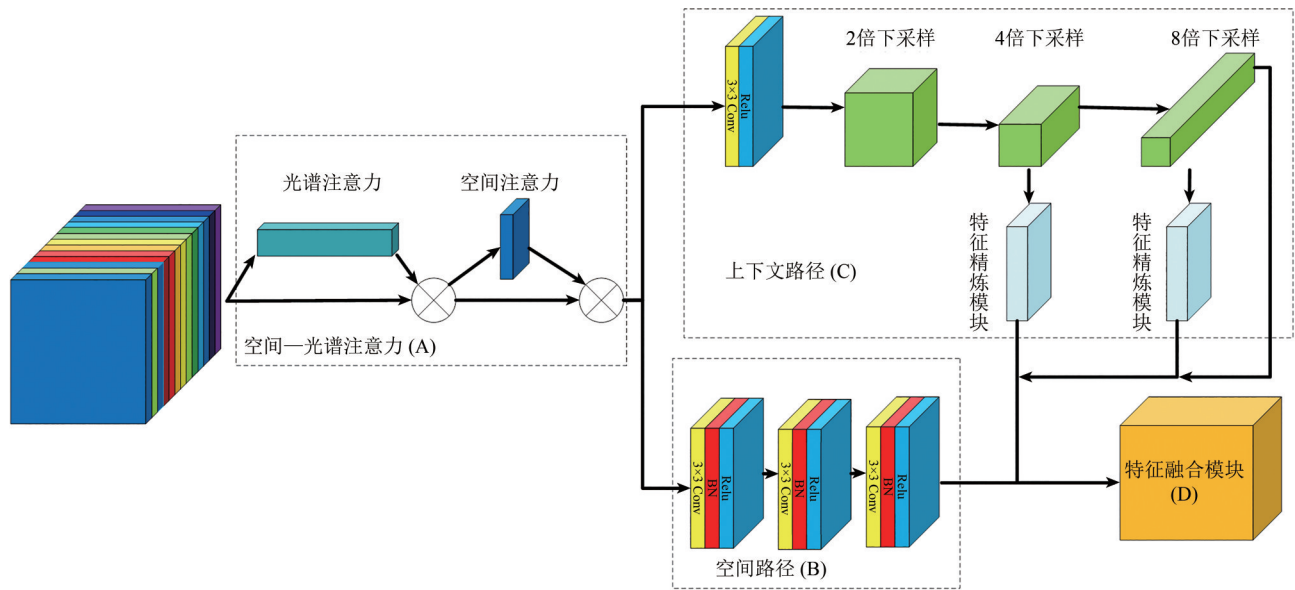


图1 光谱空间注意力双边网络结构

Fig. 1 Spectral-Spatial Attention Bilateral Network structure

空间路径共有3层, 每层均由二维卷积层、批归一化和Relu激活函数组成。每一层的卷积核大小为, 步长均为2, 因此输出特征图大小均为输入特征图大小的1/2。已有研究表明 (Luo等, 2017), 空间信息和感受野是获得高精度的关键, 该路径提取的输出特征图尺寸为原始图像的1/8, 能编码丰富的空间信息。

常用的增大感受野的方法, 如更大的卷积核、金字塔池化模块或者空间金字塔池化, 这些方法同时也带来大量的计算量和内存消耗, 导致较慢的迭代速度。本文中上下文路径采用一种类似VGG网络结构, 将其4倍下采样和8倍下采样结果通过注意力精炼模块 (图2) 提取特征。注意力精炼模块通过全局池化, 1×1 卷积, Batch Norm 和 Sigmoid 激活函数, 提取通道权重, 可以增强下采

样中更有效的特征。最后, 将两条路径输出特征通过特征融合模块提取判别性特征, 通过全连接层输出像素类别。

2.2 光谱空间注意力

光谱空间注意力如图1(A)所示, 分为光谱注意力和空间注意力, 目的是强调有助于最终分类的信息, 光谱注意力强调光谱波段即什么样的光谱是有意义的, 空间注意力则强调空间信息即在分类像素周围哪里的特征是有意义的。

光谱注意力模块输入是原始高光谱数据, 大小为 $w \times h \times d$ 。使用全局平均池化和最大池化两种方式提取互补的光谱全局特征, 得到两个 $1 \times 1 \times d$ 的光谱特征, 再将提取到的特征通过同一个多层感知机 MLP (Multi-Layer Perceptron) 提取权重信息,

将得到两个特征经过 Sigmoid 激活函数即可得到光谱权重系数 M_{sc} ，大小为 $1 \times 1 \times d$ 。通过权重系数对原始 HSI 进行加权即可得到光谱加权 HSI 特征 F' 。光谱注意力计算可概括为

$$M_{sc}(HSI) = \sigma(\text{MLP}(\text{AvgPool}(HSI))) + \sigma(\text{MLP}(\text{MaxPool}(HSI))) \quad (1)$$

式中，HSI 指原始高光谱图像， σ 指 Sigmoid 激活函数， M_{sc} 指权重系数。

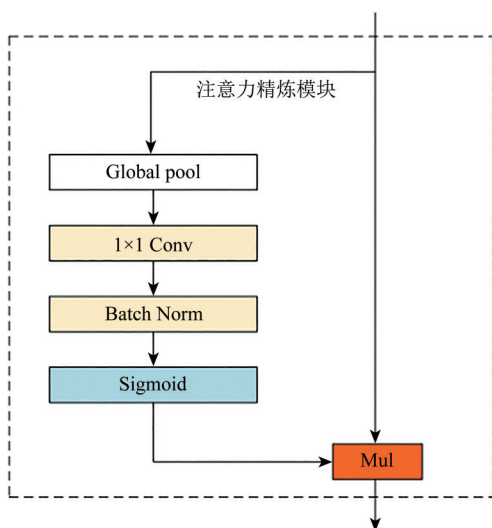


图2 注意力精炼模块

Fig. 2 Attention refinement module

空间注意力模块输入是 HSI 进行光谱加权的结果，大小为 $w \times h \times d$ 。与光谱注意力相似，空间注意力在通道维度分别进行全局平均池化和最大池化，并将得到大小为 $w \times h \times 1$ 结果进行拼接。然后经过一个 7×7 的卷积层，同样最后通过 Sigmoid 激活函数得到空间权重系数 M_{sa} ，空间权重系数 M_{sa} 与光谱加权特征 F' 相乘即可得到新特征。空间注意力计算可概括为

$$M_{sa}(F') = \sigma(f^{7 \times 7}(\text{AvgPool}(F'); \text{MaxPool}(F'))) \quad (2)$$

式中， F' 指原始高光谱图像光谱加权的结果， σ 指 Sigmoid 激活函数， $f^{7 \times 7}$ 指卷积滤波器大小为 7×7 的卷积层。

通过光谱空间注意力模块，可以将与待分类像素类别相同的更多信息进行增强，抑制对分类无效的信息，使得最终判别像素正确类别概率提高。

2.3 特征融合模块

双边网络不同路径由于结构不同，产生不同层次上特征，空间路径由于使用卷积加 Relu 函数，

得到的特征大多是丰富的细节信息；上下文路径通过多次下采样捕获更多的上下文信息。两种路径输出不同层次的特征，如果简单的进行拼接，无法得到更有效的特征，因此提出一个特征融合模块（图3）来解决特征融合问题。

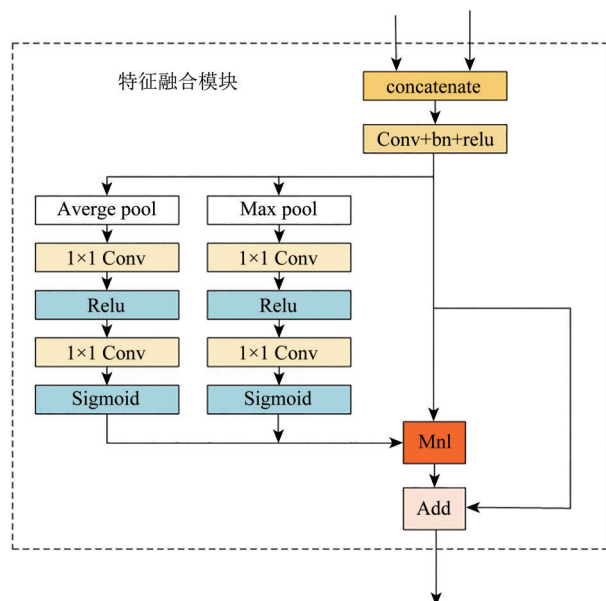


图3 特征融合模块

Fig. 3 Feature fusion module

对于两条不同路径的输出，将其串联起来，通过批归一化能平衡不同层次特征的尺度。然后将连接的特征通过一种与 SENet 类似的结构计算出一个权重向量，通过权重向量与原始特征向量相乘可将原始特征进行重新赋权，即完成了不同层次特征的选择和组合。类 SENet 结构由两条相似路径组成，为了充分利用特征信息，将特征通过最大池化和平均池化，然后通过 2 个 1×1 卷积限制模型复杂度，并在两个卷积中加入 Relu 函数增加其非线性，最后通过 sigmoid 函数得到权重向量。具体结构设计细节如图 3。

2.4 损失函数和优化器

卷积神经网络优化模型首先需要合适的目标函数，我们使用分类中最常用的交叉熵函数作为 SSABN 的损失函数，交叉熵的损失函数如下：

$$\text{Loss} = -\frac{1}{M} \sum_{m=1}^M \sum_{c=1}^C y_c^m \log(\hat{y}_c^m) \quad (3)$$

式中， y 和 $\hat{y}$ 为真实和预测标签， C 是类别的数量， M 是样本数量。模型中参数的更新则使用随机梯度下降。

3 HSI分类实验

在本节，首先介绍本文实验中使用的HSI数据集。然后对影响模型性能的一些超参数进行试验分析，之后由消融实验分析光谱空间注意力模块与特征融合模块的作用。最后将提出模型与和其

他现有模型进行比较，分析讨论实验结果。

3.1 数据集介绍

我们在实验中使用3个公开HSI数据IPs（Indian Pine），PU（Pavia University），SA（Salinas）来验证提出模型的有效性（表1）。

表1 本文所使用的3个数据集
Table 1 Three data sets used in this article

数据集	空间	光谱	类别	标记样本数	分辨率		波长/nm	传感器	拍摄时间
					光谱/nm	空间/m			
IP	145×145	200	16	10249	10	20	400—2500	AVIRIS	1992-06-01
PU	610×340	103	9	42776	3.74	1.3	430—860	ROSIS	2002-07-08
SA	512×217	204	16	54129	3.7	20	400—2500	AVIRIS	1998

（1）IP（Indian Pines）：这组数据拍摄时间为1992年6月，由机载可见光/红外成像分光计（AVIRIS）传感器在印第安纳州西北部的印第安松实验地点上空收集，该数据集像素大小为145×145，可用光谱波段数为200。该图像三分之一区域为森林或其他天然植物；接近三分之二的区域为农业作物，图中有一些农作物（玉米、大豆）覆盖率不足5%，是由于该图像拍摄时间为6月，这些农作物还处于生长初期；除此之外主要为高速公路和铁路线。现有的图像被标记为16个不同类别。表2列出了每个类的训练、验证和测试样本的数量。

表2 India Pine数据集的类名和每个类样本数
Table 2 Class name and sample number of each class for IP dataset

序号	类名	训练	验证	测试
1	Alfalfa	9	5	32
2	Corn-notill	285	143	1000
3	Corn-mintill	166	83	581
4	Corn	47	24	166
5	Grass-pasture	97	48	338
6	Grass-trees	146	73	511
7	Grass-pasture-mowed	6	3	19
8	Hay-windrowed	96	48	334
9	Oats	4	2	14
10	Soybean-notill	194	97	681
11	Soybean-mintill	491	246	1718
12	Soybean-clean	118	59	416
13	Wheat	41	20	144
14	Woods	253	126	886
15	Buildings-Grass-Trees-Drives	77	39	270
16	Stone-Steel-Towers	19	9	65
总计		2049	1025	7169

（2）PU（Pavia University）：这组数据拍摄时间为2002年7月，由意大利北部帕维亚大学的反射光学系统成像仪（ROSIS）采集，该数据集像素大小为610×340，可用光谱波段数为103。现有的图像被标记为9个不同类别。表3列出了每个类的训练、验证和测试样本的数量。

表3 Pavia University数据集的类名和每个类样本数
Table 3 Class name and sample number of each class for PU dataset

序号	类名	训练	验证	测试
1	Asphalt	1326	663	4642
2	Meadows	3730	1865	13054
3	Gravel	420	210	1469
4	Trees	613	306	2145
5	Painted metal sheets	269	134	942
6	Bare Soil	1006	503	3520
7	Bitumen	266	133	931
8	Self-Blocking Bricks	736	368	2578
9	Shadows	189	95	663
总计		8555	4277	29944

（3）SA（Salinas）：这组数据由AVIRIS传感器于1998年在Salinas山谷收集，该数据集像素大小为512×217，可用光谱波段数为224。现有地物大多为农作物，被分为16个地物类别。表4列出了每个类的训练、验证和测试样本的数量。

3个数据集有不同的可用标记样本数量。IP数据集有10249个标记样本，PU数据集有42766个标记样本，而SA数据集有54129个标记样本。对于3个数据集，我们随机选取20%，10%，70%的标记样本分别作为训练样本，验证样本和测试样本，详细数据如表2—4所示。

表 4 Salinas 数据集的类名和每个类样本数
Table 4 Class name and sample number of each class for SA dataset

序号	类名	训练	验证	测试
1	Brocoli_green_weeds_1	402	201	1406
2	Brocoli_green_weeds_2	745	372	2609
3	Fallow	395	198	1383
4	Fallow_rough_plow	279	139	976
5	Fallow_smooth	536	268	1874
6	Stubble	792	396	2771
7	Celery	716	358	2505
8	Grapes_untrained	2254	1127	7890
9	Soil_vinyard_develop	1240	620	4343
10	Corn_senesced_green_weeds	656	328	2294
11	Lettuce_roumaine_4wk	214	108	747
12	Lettuce_roumaine_5wk	385	193	1349
13	Lettuce_roumaine_6wk	183	91	642
14	Lettuce_roumaine_7wk	214	107	749
15	Vinyard_untrained	1453	727	5088
16	Vinyard_vertical_trellis	361	181	1265
总计		10825	5413	37891

3.2 实验环境配置

为了评估 SSABN 模型的性能，实验在 CPU 为 Intel Core i7-9750H 2.60 GHz*12，RAM 为 16 GB，GPU 为 NVIDIA GeForce 2060，RAM 为 6 GB 的个人计算机上进行。软件环境的系统为 Windows 10 家庭版，python 版本为 3.7.6，深度学习框架为 PyTorch，版本为 1.4.0。

实验的定量评估方法使用总体精度 OA (Overall Accuracy)，平均精度 AA (Average Accuracy) 和 Kappa 系数 (K) 作为评价指标。

3.3 超参数设置

我们分析了影响训练进度和分类性能的一些因素即超参数，如批量大小，卷积滤波器核数，空间输入大小和训练样本比例，选取其在验证数据集上分类性能最优的参数，作为最终的实验结果比较。在接下来的实验中，每个实验进行 200 个 epoch。

3.3.1 批量大小

神经网络的收敛速度和收敛性由学习率和批量大小决定，学习率控制着训练过程中梯度下降的步长，批量大小控制训练过程中每一个 epoch 中梯度下降的次数。我们使用多步下降策略控制学习率，对于批量大小则考虑以下集合 {8, 16, 32, 64, 128, 256}。结果如图 4 所示，在 IP 数据集上

最佳批量大小为 32，而在 PU 和 SA 数据集上则为 64。与 IP 数据集相比，PU 和 SA 数据集需要更大的批量大小，这是由于在相同学习率下，PU 和 SA 数据集训练样本数更大。

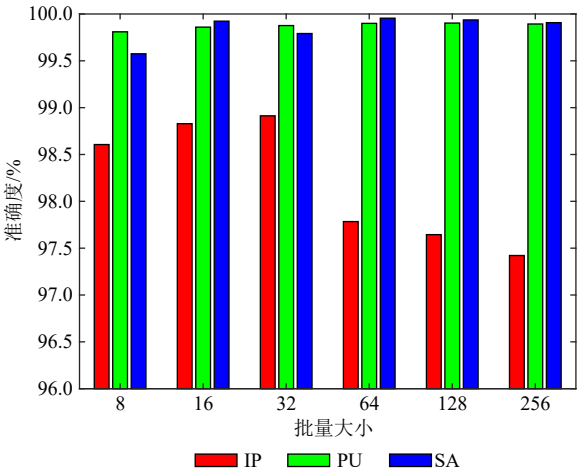


图 4 不同批量大小下的准确度

Fig. 4 Accuracy with different batch size

3.3.2 卷积滤波器核数

在双边网络中，上下文路径和空间路径的卷积滤波器的核数决定了其表示能力和计算消耗，设空间路径每层的核数均为 k_s ，上下文路径第一层同样为 k_s ，之后每次下采样将卷积滤波器的核数翻倍。为寻找 SSABN 模型最优滤波器核数 k_s ，我们考虑以下 k_s 集合 {8, 16, 24, 32}。如图 5 所示，卷积滤波器核数越大，在数据集 IP 上有更好的准确度，而在数据集 PU 和 SA 上，卷积滤波器数量为 24，模型达到最佳性能。

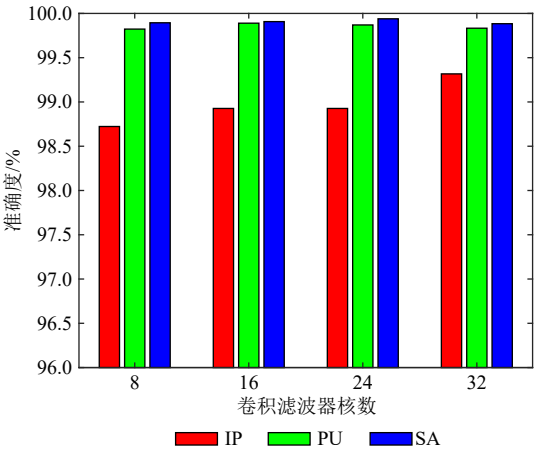


图 5 不同卷积滤波器核数下的准确度

Fig. 5 Accuracy with different kernel numbers of convolutional filters

3.3.3 空间输入大小

空间输入大小决定了有多少空间信息用于分类。随着空间输入大小的增加，可用的空间信息逐渐增加，同时也增加了一定的冗余信息。为评估空间输入大小对 SSABN 分类效果的影响，考虑 {5, 7, 9, 11, 13, 15, 17, 19, 21} 的空间输入大小集合。如图 6 所示，对于 3 个不同的数据集，当空间输入大小大于或等于 9×9 时，SSABN 表现稳健。

双边网络结构提取两种不同层次特征，其空间路径提取特征具有丰富的空间信息，上下文路径通过多次下采样提取全局特征。只有当两种不同层次特征有足够区分度时，经过特征融合模块的输出结果才具有足够的判别性和抽象性。由于上下文路径的最大下采样倍数为 8，当空间输入小于 8 时，4 倍下采样与 8 倍下采的空间特征相似。所以当空间输入大小大于 9×9 时，SSABN 表现稳健。

3.3.4 训练样本比例

卷积神经网络由于其结构，相较于传统方法有更强的拟合能力，当训练样本数量较少时更容

易产生过拟合。考虑分别使用 1%, 5%, 10%, 15% 和 20% 的标记像素作为训练集来训练 SSABN。如图 7 所示，随着训练样本数量比例增加到 15%，SSABN 在 3 个数据集上有更高的整体准确度。当训练样本比例为 1% 和 5% 时，SSABN 在 IP 数据集上略低于 SSRN 和 FDSSC，但随着训练样本的增加，SSABN 准确度略高于其他方法。这是由于 IP 数据集本身数据量较小，且部分类别样本数过小导致。

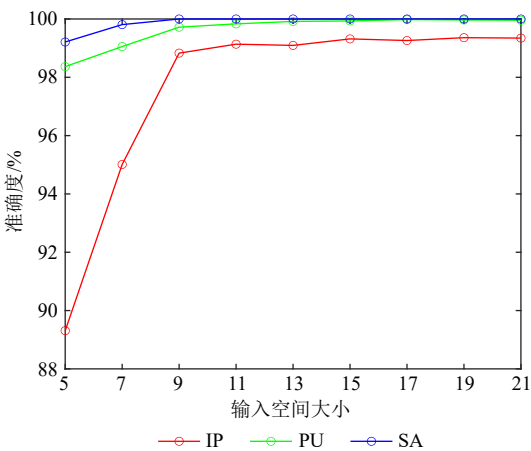


图 6 不同空间输入大小的 SSABN 的准确度

Fig. 6 Accuracy with different spatial input sizes

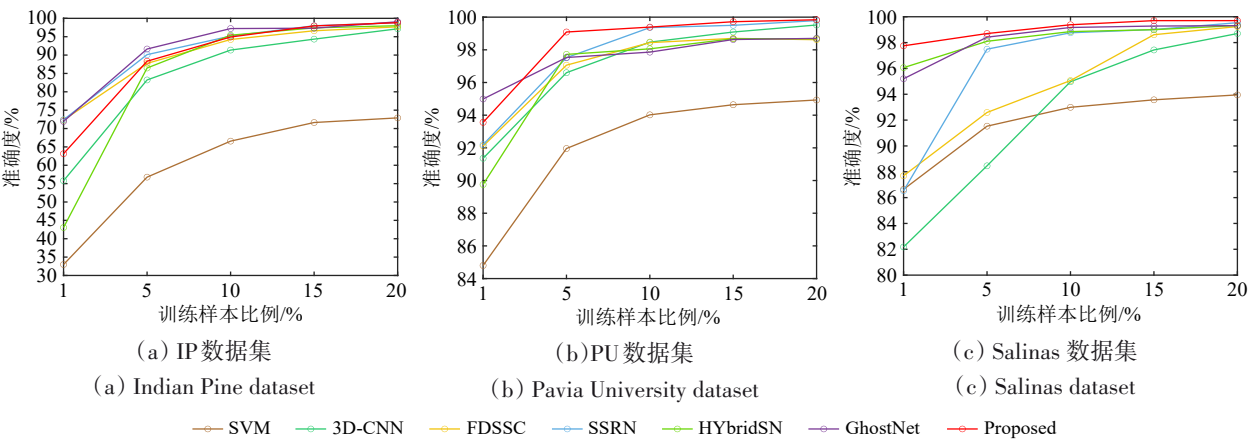


图 7 模型在不同训练样本比例的 IP, PU, SA 数据集上的准确度

Fig. 7 The accuracy of the algorithm on the IP, PU, and SA datasets of different training sample proportions

3.4 消融实验

3.4.1 光谱空间注意力

我们将 SSABN 与第一层没有光谱空间注意力模块的双边网络进行比较，结果如表 5。在 3 个数据集上，SSABN 分类准确度上均优于没有第一层注意力模块的双边网络。这些结果表明，第一层光谱空间注意力模块能有效选择有用的原始 HSI 数据进行后续的特征学习。

表 5 有/无光谱空间注意力模块的 OA

Table 5 OA of with/without spectral-spatial attention module

数据集	/%		
	IP	PU	SA
含注意力模块	98.32±1.36	99.63±0.24	99.83±0.17
不含注意力模块	96.54±1.44	95.63±0.52	94.51±0.74

我们将光谱—空间注意力模块与其他常用注意力模块：通道注意力，空间注意力，SENet 进行

比较。结果如表6所示，除了PU数据集，光谱—空间注意力在3个数据集上，比其他3个注意力模块都有更高的OA。而由于PU数据集具有1.3 m/pix的像素分辨率，这使得相邻像素对中心像素分类有更大影响，所以空间注意力比光谱空间注意力有更高的OA。

表6 不同注意力模块的OA
Table 6 OA of different attention module

数据集	IP	PU	SA
光谱—空间注意力	98.32±1.36	99.63±0.24	99.83±0.17
通道注意力	96.71±1.18	99.20±0.48	99.31±0.40
空间注意力	98.06±1.58	99.66±0.27	99.65±0.31
SENet	97.83±1.15	99.57±0.26	99.65±0.33

3.4.2 特征融合模块

我们将SSABN与没有特征融合模块的双边网络进行比较，结果如表7。在3个数据集中，含有特征融合模块分类准确度高于不含特征融合的网络，这说明特征融合模块能有效融合两条路径不同输出，产生更具判别性特征。

3.5 HSI分类实验

在本节中，我们将提出的SSABN与一种传统

机器学习方法：SVM（Melgani和Bruzzone，2004），以及3种基于深度学习的方法：3D-CNN，FDSSC（Wang等，2018），SSRN（Zhong等，2017），HYBridSN（Roy等，2020）以及GhostNet（Das等，2020）进行对比。我们进行10次重复实验，实验结果以平均值±标准偏差表示。

表7 特征融合模块对OA影响
Table 7 The influence of feature fusion module on OA

数据集	IP	PU	SA
SSABN	99.32	99.83	99.96
不含特征融合模块	97.66	99.70	99.08

（1）定量分析。不同方法的定量度量见表5—8。由表5—7可以看出，与传统算法相比，深度学习方法能在3个公开数据集上取得更好的性能。这是由于它们可以学习更多抽象的高级特性，而SVM仅仅使用原始数据的光谱特性进行分类。此外，光谱空间注意力模块能有效地提取光谱和空间信息，双边网络结构能有效地利用多尺度特征，使得SSABN的分类准确度优于其他几种深度学习方法。

上述实验证明，我们提出的方法在同等条件下

表8 不同模型在IP数据集上分类准确度
Table 8 Accuracy of different models on the IP dataset

序号	SVM	3DCNN	FDSSC	SSRN	HYBridSN	GhoMRNet	SSABN
OA	73.07±0.65	97.56±0.43	98.01±0.18	99.19±0.26	99.38±0.23	99.25±0.20	99.40±0.38
AA	52.90±3.37	99.23±0.19	87.82±0.39	98.93±0.59	99.48±0.26	98.64±0.26	99.49±0.22
κ	68.46±0.77	97.02±0.52	97.73±0.25	99.07±0.30	99.30±0.30	99.14±0.05	99.32±0.05
1	73.58	100.00	96.88	97.82	100.00	100.00	100.00
2	54.67	96.34	99.80	99.17	97.10	97.80	97.00
3	14.74	99.49	100.00	99.53	100.00	98.97	100.00
4	82.38	100.00	100.00	97.79	100.00	100.00	100.00
5	93.32	99.91	96.45	99.24	100.00	99.70	100.00
6	22.73	99.75	99.41	99.51	100.00	99.80	100.00
7	80.37	100.00	26.32	98.70	100.00	100.00	100.00
8	0.00	100.00	100.00	99.85	100.00	100.00	100.00
9	63.75	100.00	0.00	98.50	100.00	85.72	100.00
10	88.19	98.72	98.97	98.74	99.85	99.27	100.00
11	38.95	95.52	96.33	99.30	99.77	99.59	99.83
12	89.02	99.47	97.12	98.43	98.08	98.08	98.08
13	88.64	100.00	98.61	100.0	100.00	100.00	100.00
14	37.22	99.55	100.00	99.31	100.00	100.00	100.00
15	18.92	99.54	95.19	99.20	100.00	99.26	100.00
16	73.58	99.34	100.00	97.82	96.92	100.00	96.92

注：加粗表示最优效果。

可以达到较高的准确度，而一个好的方法应该平衡准确度和效率。在每个数据集中随机选取 20% 标记像素作为训练样本验证，其他参数选取算法的最优参数。由表 8 可以看到，SSABN 相比其他大部分深度学习方法有更快的迭代速度，这是由于光谱空间注意力模块是简单的轻量级模块，上下文路径模块由于下采样有快速的计算速度，而空间路径本身 2D 卷积结构计算量小于 SSRN 的 3D 卷积结构，使得 SSABN

拥有更快的计算效率和迭代速度。所提出 SSABN 训练时间略高于 HYBridSN，这是由于 HYBridSN 对原始 HIS 进行 PCA 降维，大大减少了计算量。

以上实验证明，我们提出的 SSABN 不仅在有较高的准确度的同时，有效减少了训练时间，能很好地提高准确度和效率，是一个较为优质的解决方案。

表 9 不同模型在 PU 数据集上分类准确度
Table 9 Accuracy of different models on the PU dataset

								/%
序号	SVM	3DCNN	FDSSC	SSRN	HYBridSN	GhoMRNet	SSABN	
OA	94.79±0.47	99.54±0.11	98.73±0.56	99.79±0.09	99.67±0.26	99.81±0.08	99.83±0.13	
AA	92.93±0.36	99.66±0.11	98.31±0.66	99.66±0.17	99.36±0.12	99.65±0.15	99.73±0.17	
κ	93.08±0.70	99.41±0.15	97.24±0.49	99.72±0.12	99.56±0.15	99.76±0.12	99.78±0.11	
1	94.48	99.36	98.82	99.92	100.00	99.94	100.00	
2	98.16	99.36	99.74	99.96	100.00	99.99	100.00	
3	82.13	99.69	98.14	98.46	99.80	98.43	98.98	
4	94.70	99.63	94.43	99.69	97.16	99.77	99.72	
5	98.51	99.95	99.14	99.99	100.00	100.00	100.00	
6	91.00	99.96	99.69	99.94	100.00	100.00	100.00	
7	86.84	100.00	96.68	99.82	100.00	99.36	100.00	
8	84.40	99.65	98.90	99.22	99.11	99.34	98.88	
9	99.47	99.38	89.54	99.95	98.19	100.00	100.00	

注：加粗表示最优效果。

表 10 不同模型在 SA 数据集上分类准确度
Table 10 Accuracy of different models on the SA dataset

								/%
序号	SVM	3DCNN	FDSSC	SSRN	HYBridSN	GhoMRNet	SSABN	
OA	94.14±0.58	97.56±1.25	99.04±0.47	99.60±0.22	99.88±0.09	99.93±0.7	99.96±0.04	
AA	97.14±0.86	99.23±1.97	99.54±0.18	99.84±0.57	99.90±0.04	99.95±0.05	99.97±0.03	
κ	93.47±0.65	97.02±1.39	98.93±0.53	99.56±0.25	99.87±0.03	99.92±0.08	99.95±0.05	
1	99.87	100.00	100.00	100.00	100.00	100.00	100.00	
2	99.81	96.34	100.00	100.00	100.00	100.00	100.00	
3	99.46	99.49	100.00	100.00	100.00	100.00	100.00	
4	98.83	100.00	100.00	100.00	100.00	99.80	100.00	
5	99.94	99.91	99.41	100.00	99.84	100.00	100.00	
6	99.69	99.75	100.00	100.00	100.00	100.00	100.00	
7	88.69	100.00	100.00	100.00	100.00	100.00	100.00	
8	99.76	97.94	97.22	98.90	100.00	99.91	99.92	
9	96.64	100.00	100.00	100.00	100.00	100.00	100.00	
10	98.59	98.72	99.13	100.00	98.91	100.00	100.00	
11	99.74	95.52	99.46	99.87	100.00	100.00	99.73	
12	99.18	99.47	100.00	100.00	99.85	100.00	100.00	
13	97.55	100.00	100.00	100.00	100.00	100.00	100.00	
14	77.49	99.55	99.73	100.00	100.00	99.73	100.00	
15	99.10	99.54	97.94	98.80	99.74	99.69	99.82	
16	99.87	99.34	99.68	99.92	100.00	100.00	100.00	

注：加粗表示最优效果。

(2) 定性分析。图 8—10 是原始图像的 groundtruth 图和不同方法在 3 个数据集上的分类结果可视化。从 3 个数据集的分类图可以看到，SSABN 有误差点，生成的分类图最为准确，特别是在两

类的分界线上，同时在没有标记类别的区域有更精细的分辨率。这是因为该模型利用空间信息来学习中心像素与周围像素的关系，所以能产生更为平滑的边界。

表 11 不同模型在 3 个数据集上迭代所需时间

Table 11 The time required for different models on the three datasets

	数据集	SVM	3D-CNN	FDSSC	SSRN	HYBridSN	GhoMRNet	SSABN
一个 epoch 所需时间/s	IP	2.56	20.58	125.89	61.09	9.15	9.36	5.37
	PU	2.32	40.58	438.57	205.68	11.03	29.31	16.24
	SA	7.67	93.89	636.57	298.24	13.58	36.22	21.28
训练所需时间/min	IP	—	35	208	101	15	16	9
	PU	—	70	730	342	18	49	27
	SA	—	155	1060	496	22	61	35

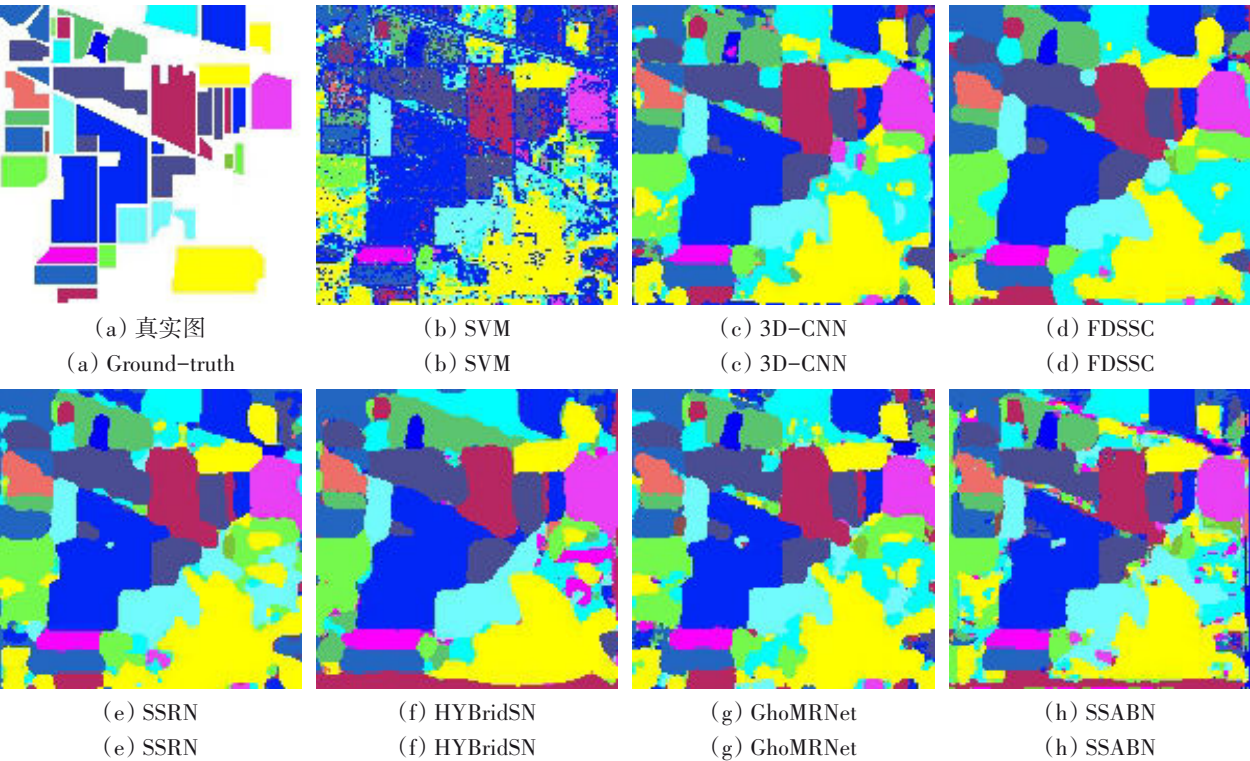
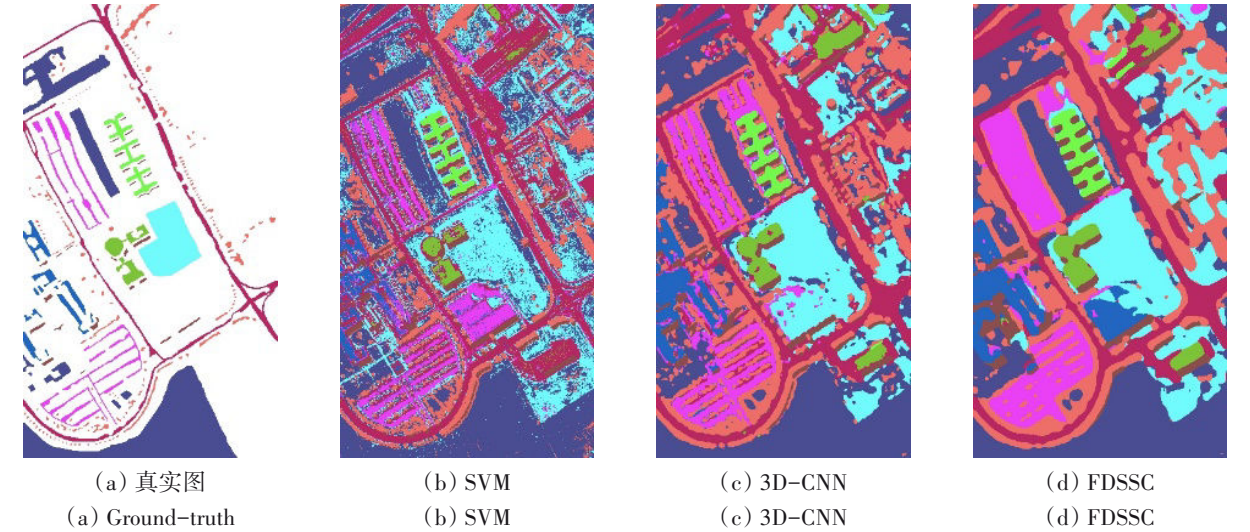


图 8 使用不同模型的 IP 数据分类图

Fig. 8 Classification maps for the IP dataset



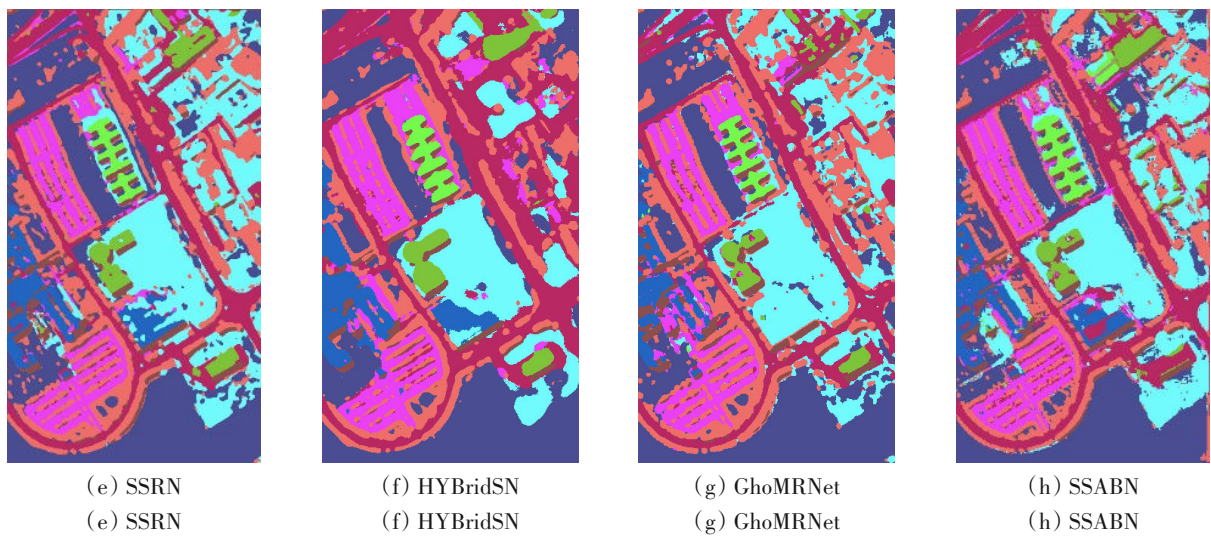


图9 使用不同模型的PU数据分类图
Fig. 9 Classification maps for the PU dataset

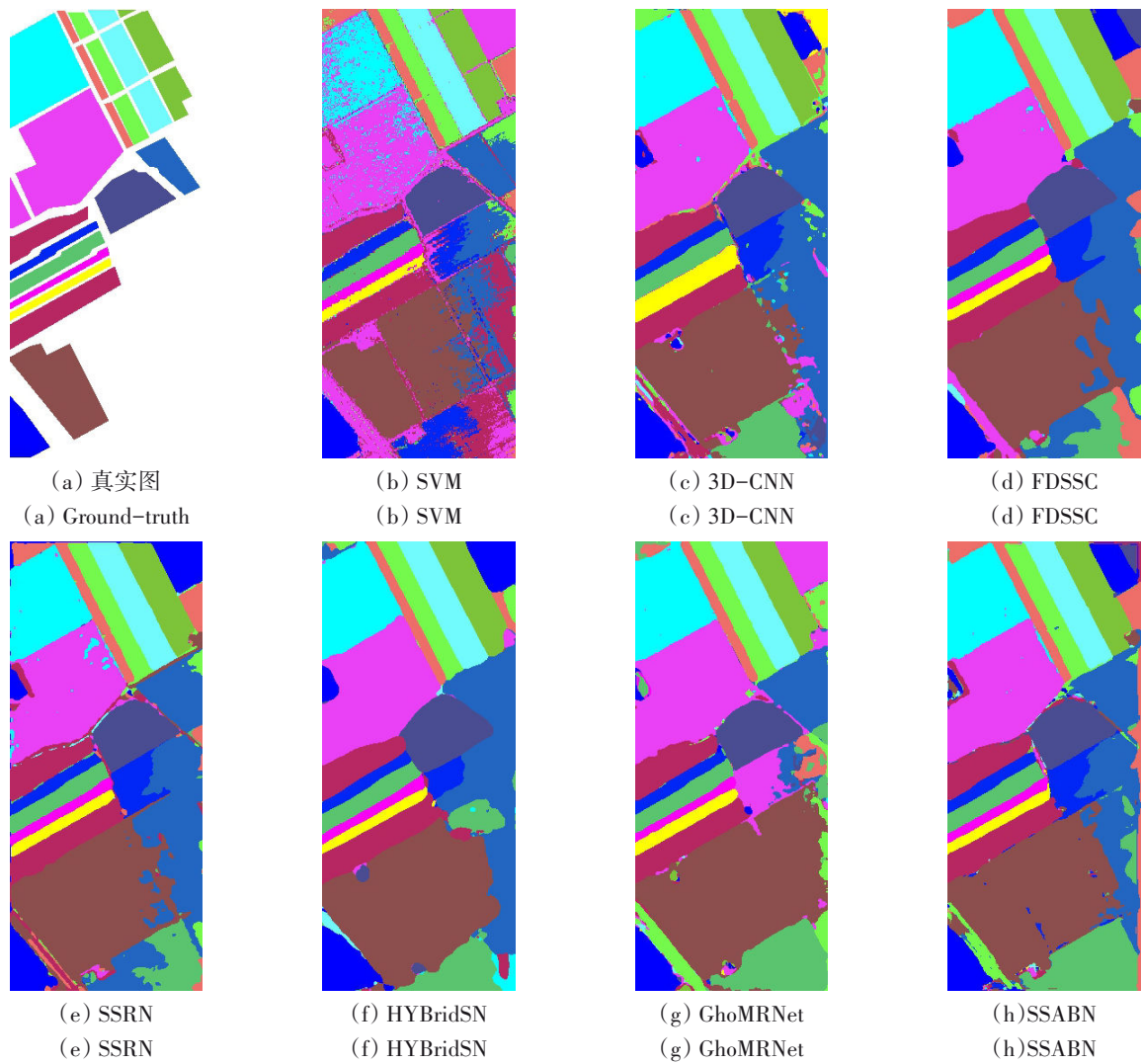


图10 使用不同模型的SA数据分类图
Fig. 10 Classification maps for the SA dataset

4 结 论

本文中提出一种用于HSIC的SSABN模型,该模型用双边网络作为基础网络结构学习光谱和空间特征,用光谱—空间注意力来进行光谱波段和空间信息的选择和增强,最后由特征融合模块进行特征融合输出判别性特征。

首先,原始的HSI块直接作为模型的输入,这是一个端到端的二维CNN网络框架,不用依靠领域知识和专家经验。其次,利用光谱注意力和空间注意力模块来增强和提取有效的波段和像素。消融实验表明,光谱空间注意力模块能有效地提取抽象性特征,增加分类精度。双边网络提取丰富信息后由特征融合模块提取判别性特征,并通过全连接层输出预测类别,通过熔断实验证明了特征融合模块的有效性。最后,在3个公开数据集上进行对比实验,由定量分析可以看出,SSABN网络模型均取得了更加好的分类精度同时也降低了计算消耗时间,结果可视化也表明,SSABN网络能有效地划分边界,且分类精度高于其他模型。

参考文献(References)

- Cao Y, Ye Y T, Zhao H L, Jiang Y Z and Wang H. 2019. Remote sensing of total suspended matter concentration and turbidity in a macrophytic lake. *Journal of Remote Sensing*, 23(6): 1253-1268 (曹引, 冶运涛, 赵红莉, 蒋云钟, 王浩. 2019. 草型湖泊总悬浮物浓度和浊度遥感监测. *遥感学报*, 23(6): 1253-1268) [DOI: 10.11834/jrs.20198144]
- Chen X, Lee H and Lee M. 2019. Feasibility of using hyperspectral remote sensing for environmental heavy metal monitoring. [DOI: 10.5194/isprs-archives-XLII-3-W7-1-2019]
- Chen Y S, Lin Z H, Zhao X, Wang G and Gu Y. 2014. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6): 2094-2107. [DOI: 10.1109/JSTARS.2014.2329330]
- Chen Y S, Zhao X and Jia X P. 2015. Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6): 2381-2392. [DOI: 10.1109/JSTARS.2015.2388577]
- Das A, Saha I and Scherer R. 2020. GhoMR: multi-receptive light-weight residual modules for hyperspectral classification. *Sensors*, 20(23): 6823. [DOI: 10.3390/s20236823]
- Deng C, Xue Y M, Liu X L, Li C and Tao D C. 2019. Active transfer learning network: a unified deep joint spectral-spatial feature learning model for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(3): 1741-1754. [DOI: 10.1109/TGRS.2018.2868851]
- Ghamisi P, Dalla Mura M and Benediktsson J A. 2015. A survey on spectral-spatial classification techniques based on attribute profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5): 2335-2353. [DOI: 10.1109/TGRS.2014.2358934]
- Ham J, Chen Y C, Crawford M M and Ghosh J. 2005. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3): 492-501. [DOI: 10.1109/TGRS.2004.842481]
- Li J L, Sun D W and Cheng J H. 2016. Recent advances in nondestructive analytical techniques for determining the total soluble solids in fruits: a review. *Comprehensive Reviews in Food Science and Food Safety*, 15(5): 897-911. [DOI: 10.1111/1541-4337.12217]
- Li R, Zheng S Y, Duan C X, Yang Y and Wang X Q. 2020. Classification of hyperspectral image based on double-branch dual-attention mechanism network. *Remote Sensing*, 12(3): 582. [DOI: 10.3390/rs12030582]
- Li T, Zhang J P and Zhang Y. 2015. Classification of hyperspectral image based on deep belief networks//2014 IEEE International Conference on Image Processing (ICIP). Paris: IEEE. [DOI: 10.1109/ICIP.2014.7026039]
- Luo R B, Liao W Z, Philips W and Pi Y G. 2015. An improved semi-supervised local discriminant analysis for feature extraction of hyperspectral image. 2015 Joint Urban Remote Sensing Event (JURSE). Lausanne: IEEE. [DOI: 10.1109/JURSE.2015.7120508]
- Luo W J, Li Y J, Urtasun R and Zemel R. 2017. Understanding the effective receptive field in deep convolutional neural networks//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc.
- Ma L, Crawford M M and Tian J W. 2010. Local manifold learning-based k -nearest-neighbor for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11): 4099-4109. [DOI: 10.1109/TGRS.2010.2055876]
- Mei S H, Ji J Y, Geng Y H, Zhang Z, Li X and Du Q. 2019a. Unsupervised spatial-spectral feature learning by 3D convolutional autoencoder for hyperspectral classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9): 6808-6820. [DOI: 10.1109/TGRS.2019.2908756]
- Mei X G, Pan E T, Ma Y, Dai X B, Huang J, Fan F, Du Q L, Zheng H and Ma J Y. 2019b. Spectral-spatial attention networks for hyperspectral image classification. *Remote Sensing*, 11(8): 963. [DOI: 10.3390/rs11080963]
- Melgani F and Bruzzone L. 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8): 1778-1790. [DOI: 10.1109/TGRS.2004.831865]

- Mou L C, Ghamisi P and Zhu X X. 2017. Deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7): 3639-3655. [DOI: 10.1109/TGRS.2016.2636241]
- Murphy R J, Whelan B, Chlingaryan A and Sukkarieh S. 2018. Quantifying leaf-scale variations in water absorption in lettuce from hyperspectral imagery: a laboratory study with implications for measuring leaf water content in the context of precision agriculture. *Precision Agriculture*, 20(4): 767-787. [DOI: 10.1007/s11119-018-9610-5]
- Rodarmel C and Shan J. 2002. Principal component analysis for hyperspectral image classification. *Surveying and Land Information Systems*, 62(2): 115-123.
- Roy S K, Dubey S R, Chatterjee S, Chaudhuri B B. 2020. FuSENet: fused squeeze-and-excitation network for spectral-spatial hyperspectral image classification. *IET Image Process*, 14(8): 1653-1661. [DOI: 10.1049/iet-ipr.2019.1462]
- Roy S K, Krishna G, Dubey S R and Chaudhuri B B. 2020. HybridSN: exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 17(2): 277-281 [DOI: 10.1109/LGRS.2019.2918719]
- Shimoni M, Haelterman R and Perneel C. 2019. Hypersectral imaging for military and security applications: combining myriad processing and sensing techniques. *IEEE Geoscience and Remote Sensing Magazine*, 7(2): 101-117. [DOI: 10.1109/MGRS.2019.2902525]
- Sun W and Du Q. 2018. Graph-regularized fast and robust principal component analysis for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, 56(6): 3185-3195. [DOI: 10.1109/TGRS.2018.2794443]
- Sun W W and Du Q. 2019. Hyperspectral band selection: a review. *IEEE Geoscience and Remote Sensing Magazine*, 7(2): 118-139. [DOI: 10.1109/MGRS.2019.2911100]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I. 2017. Attention is all you need//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc.
- Wang J and Chang C I. 2006. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6): 1586-1600. [DOI: 10.1109/TGRS.2005.863297]
- Wang W J, Dou S G, Jiang Z M and Sun L J. 2018. A fast dense spectral-spatial convolution network framework for hyperspectral images classification. *Remote Sensing*, 10(7): 1068. [DOI: 10.3390/rs10071068]
- Zhang X R, Sun Y J, Jiang K, Li C, Jiao L C and Zhou H Y. 2018. Spatial sequential recurrent neural network for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(11): 4141-4155. [DOI: 10.1109/JSTARS.2018.2844873]
- Zhong Z L, Li J, Luo Z M and Chapman M. 2018. Spectral-spatial residual network for hyperspectral image classification: a 3-d deep learning framework. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2): 847-858. [DOI: 10.1109/TGRS.2017.2755542]
- Zhou Y C, Peng J T and Chen C L P. 2015. Extreme learning machine with composite kernels for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6): 2351-2360. [DOI: 10.1109/JSTARS.2014.2359965]
- Zhu L, Chen Y S, Ghamisi P and Benediktsson J A. 2018. Generative adversarial networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(9): 5046-5063 [DOI: 10.1109/TGRS.2018.2805286]

Spectral-spatial attention bilateral network for hyperspectral image classification

YANG Xing, CHI Yue, ZHOU Yatong, WANG Yang

School of Electronics and Information Engineering, Hebei University of Technology, Tianjin 300401

Abstract: HypeSpectral Image Classification (HSIC) is a pixel-level classification problem, and it involves classifying each pixel in the hyperspectral image and confirming the pixel category. However, discriminative features in HSIC task are difficult to acquire and learn, and the extraction of sufficient and effective features directly affects the classification results. In the past few years, Convolutional Neural Networks (CNNs) have achieved better results in HSIC, but the high dimensionality of hyperspectral images and the equal processing of all bands by CNNs have limited the performance of CNN. This study proposes an end-to-end Spectral-Spatial Attention Bilateral Network (SSABN) for HSIC. The network directly uses 3D blocks of the original image as input data without the complicated preprocessing. First, the original data are processed through the spectral-spatial attention module to enhance the useful bands or pixels for classification and suppress invalid information. Then, the spatial and context paths of the bilateral network are designed. The spatial path has three layers, and each layer is composed of convolution, batch normalization, and Relu activation function to extract spatial information. The context path is

composed of three downsampling and attention refining modules. The downsampling is used to provide receptive field, and the attention refinement module is used to refine downsampling features. Finally, a feature fusion module is designed to fuse different levels of features through maximum pooling and average pooling for generating discriminative features. Compared with common CNN, SSABN can adaptively enhance effective information, extract more abstract discriminative features, and consume less training time. Experimental results show that SSABN has good fitting ability in different training sample ratios. In the results of ablation experiments, the accuracy of the spectral-spatial attention mechanism is 1%—2% higher than those of other mainstream attention mechanisms, and the feature fusion module can improve the discrimination of extracted features. In the experiments of three public datasets, the classification accuracy of SSABN is higher than 99%, and the training time is less than those of other methods. The classification performance of SSABN is better than those of other hyperspectral image classification algorithms, while reducing its training time can more effectively improve accuracy and efficiency.

Key words: remote sensing, convolutional neural network, deep learning, feature fusion, Indian Pine dataset, Pavia University dataset, Salinas dataset

Supported by Natural Science Foundation of Hebei Province (No. F201902364); Hebei Provincial Department of Education (No. ZD2020304)