

联合分离卷积与密集连接轻量级神经网络的高光谱图像分类

宋廷强¹, 宗达¹, 刘童心¹, 范海生², 黄腾杰², 蒋晓旭², 王浩宇¹

1. 青岛科技大学 信息科学技术学院, 青岛 266061;

2. 珠海欧比特宇航科技股份有限公司 人工智能研究院, 珠海 519080

摘要: 针对高光谱遥感图像空间分辨率低, 标注训练样本困难的问题, 本文提出一种基于分离卷积 (Separable convolution) 与密集连接 (Dense connection) 的轻量级神经网络 SDLN 模型。该模型基于 DenseNet 的思想, 同时采用计算量更少的分离卷积代替 3D 卷积, 根据算法提取的光谱信息和空间信息, 结合目标及周围像素信息推断其中心像素内容, 实现对单像素的分类。基于 IP、PU 和 KSC 这 3 个广泛使用的高光谱数据集进行实验, 按照分层抽样的方法, 每个类别选取少量样本作为训练集, 分类精度分别达到了 97.4%、97.6%、99.2%, 与 SSRN、SVM-RBF、MDGCN、DBDA 及 pResNet 多种先进分类算法对比, 分类精度提高且时间成本降低。

关键词: 高光谱图像分类, 深度学习, 轻量级网络, 密集连接, 可分离卷积

中图分类号: P2

引用格式: 宋廷强, 宗达, 刘童心, 范海生, 黄腾杰, 蒋晓旭, 王浩宇. 2022. 联合分离卷积与密集连接轻量级神经网络的高光谱图像分类. 遥感学报, 26(11): 2317-2328

Song T Q, Zong D, Liu T X, Fan H S, Huang T J, Jiang X X and Wang H Y. 2022. Hyperspectral image classification of the deep neural network based on 3D convolution and dense connection. National Remote Sensing Bulletin, 26(11): 2317-2328 [DOI:10.11834/jrs.20210313]

1 引言

高光谱图像中的光谱信息和空间信息可以充分反映地物的物理结构和化学成分, 单个像素点包含的信息远大于高分辨率图像或者雷达图像, 对于不同的地物类型的敏感度更高, 因此更适合用于地物的分类, 并已广泛应用于地物监测 (Prey 等, 2018)、农作物疾病检测 (Zheng 等, 2018)、资源勘测 (Kirsch 等, 2018) 等多个领域。

高光谱分类任务需对高光谱图像的每一个像素点给出其所属类别。影响分类准确率的原因主要有两个方面: 首先是受成像传感器的影响, 空间分辨率低导致单个像素包含多种地物, 出现混合像元 (Chang 等, 1998) 的情况, 天气和大气气体的变化也会影响地物的光谱特性 (周雨霁和田庆久, 2008); 其次是高光谱图像的高维性, 在样

本有限的情况下, 当用于分类的特征超过一定限度时, 分类精度反而会下降 (孙伟伟等, 2018)。

在早期的高光谱图像分类中, 人们更关注于图像的光谱信息, 使用支持向量机 (Archibald 和 Fann, 2007)、随机森林 (Joelsson 等, 2005) 等方法进行分类, Melgani 和 Bruzzone (2004) 提出了基于支持向量机 SVM (Support Vector Machine) 的分类方法, 但是该方法更适用于解决二分类问题, 对于多分类问题和高维度的高光谱数据, 矩阵的存储和计算需要占用大量的机器内存和计算资源。Li 等 (2019a) 对 SVM 进行了改进, 基于 boost 算法使用多核 SVM 进行训练, 具有更好的分类性能。

传统机器学习 (Li 等, 2015) 方法依赖于复杂的特征提取工程, 采用主成分分析 PCA (Principal Component Analysis) 或者其他方法进行降维处理, 特征提取的好坏直接影响分类器的效果。随

收稿日期: 2020-08-11; 预印本: 2021-01-11

基金项目: 山东省重点研发计划(公益类专项)项目(编号:2019GGX101047);“珠海一号”卫星大数据云服务平台与应用示范基金(编号:ZH0111-0405-170027-P-WC)

第一作者简介: 宋廷强, 研究方向为数据、人工智能、遥感的理论和应用。E-mail: 116638307@qq.com

通信作者简介: 宗达, 研究方向为图像处理、人工智能、高光谱应用。E-mail: 594880373@qq.com

着深度学习的发展,卷积神经网络在特征提取与特征融合方面显示出了强大的优势,端到端的训练过程舍弃了特征工程,并在多个领域出色的完成了任务 (Redmon 和 Farhadi, 2018)。同时近年来深度学习也成为高光谱图像分类的研究热点 (Ball 和 Wei, 2018; Gao 等, 2018; Makantasis 等, 2015; 张康 等, 2018), 光谱信息与空间信息融合 (Mei 等, 2017; Fu 等, 2015) 成为高光谱图像分类的常用方法。

3D 卷积 (He 等, 2017) 被验证适用于高光谱图像的分类任务。Liu 等 (2018) 提出了高光谱图像分类的新思路, 基于 3D 卷积搭建网络结构, 通过迁移学习和数据增强来解决样本小的问题, 这种训练方式取得了较好的效果, 但是以 3D 卷积搭建的神经网络信息传递的效率低, 3D 卷积的卷积核带来了大量的参数; Jiao 等 (2017) 通过计算机视觉中语义分割的方法来完成高光谱图像分类任务, 首先利用 PCA 降维和空谱特征融合方法对输入图像处理, 得到图像的空间特征, 然后使用全卷积网络 FCN (Fully Convolutional Networks) 进行分类, 具有创新性, 但是这种方法无可避免的丢失了大量信息, 也舍弃深度学习端到端的优势; Paoletti 等 (2019) 提出基于残差结构的神经网络, 但是相比于密集连接, 残差结构虽然保留了信息传递但是增加了大量的网络参数, 对于高光谱小样本来说增大了过拟合的风险; 王雷全等 (2021) 提出使用双向 RNN 提取图像的光谱信息不适用于高光谱图像, 因为像素中光谱信息相互独立, 使用卷积神经网络提取光谱变化信息效果要优于 RNN。

由于高光谱图像空间分辨率通常较低, 标注困难; 小样本类别地物的占地面积小, 数据量少, 不足以提供大量训练数据 (崔宾阁 等, 2017)。高光谱分类需要算法增强信息提取能力, 同时防止小训练样本造成的过拟合情况。为解决上述问题, 本文提出了一种基于可分离卷积 (Separable convolution) 和密集连接 (Dense connection) 的轻量级卷积神经网络 (Light Network), 简称 SDLN, 在光谱信息提取部分使用密集连接, 提高特征图的复用, 加入归一化层和 Dropout 降低过拟合的概率, 使用可分离卷积替换 3D 卷积减少网络参数, 联合空间光谱信息, 在小训练样本的情况下也可实现较高的分类精度。

2 相关工作

2.1 可分离卷积

卷积神经网络 (Krizhevsky 等, 2012) 是一类主要包含卷积计算的前馈神经网络统称, 按计算方法分为: 一维卷积、二维卷积和三维卷积。一维卷积用于处理序列信号的特征, 二维卷积用于提取图像空间特征, 三维卷积根据卷积核的设置可以同时提取空间信息与序列信息, 而高光谱图像可以被视为具有空间信息与光谱信息的三维立体数据块。许多研究者使用 3D 卷积作为基础结构 (Rao 等, 2020), 但是这种方法需要更多的计算, 影响分类效率。

假设输入数据为 $(N, C_{in}, H_{in}, W_{in}, D_{in})$, 3D 卷积核大小为 (K_h, K_w, K_d) , 个数为 K_c , 空洞率 (dilation) 为 (d_1, d_2, d_3) , 步长 (Stride) 为 (S_1, S_2, S_3) , 填充 (padding) 为 (P_1, P_2, P_3) 则输出特征图的大小和卷积计算量 FLOPS 如下:

$$H_{out} = \frac{D_{in} + 2 \times P_1 - d_1 \times (K_h - 1) - 1}{S_1} + 1 \quad (1)$$

$$W_{out} = \frac{H_{in} + 2 \times P_2 - d_2 \times (K_w - 1) - 1}{S_2} + 1 \quad (2)$$

$$D_{out} = \frac{D_{in} + 2 \times P_3 - d_3 \times (K_d - 1) - 1}{S_3} + 1 \quad (3)$$

$$FLOPS = K_c \times K_h \times K_w \times K_d \times K_c \times H_{out} \times W_{out} \times D_{out} \quad (4)$$

Howard 等 (2017) 提出专注于移动端和嵌入式设备中的轻量级 CNN 网络 MobileNet, 使用深度可分离卷积替换传统的卷积操作, 将标准卷积分解为 depthwise convolution (逐通道卷积) 和 pointwise convolution (逐点卷积), 如图 1 所示, 首先使用空间卷积对输入进行特征提取, 同时保持通道维度不变, 然后使用深度卷积进行通道合并, 在有同样效果的情况下大大降低了模型参数量。

受 MobileNet 启发, 本文将分离卷积引入高光谱分类任务, 使用分离卷积代替 3D 卷积, 将传统的 3D 卷积核 (K_h, K_w, K_d) 替换为 $(K_h, K_w, 1)$ 和 $(1, 1, K_d)$, 保证有相同输出的同时减少了计算量。分离卷积的计算量如下:

$$D_{in} + K_c \times 1 \times 1 \times K_d \times H_{out} \times W_{out} \times D_{out} \quad (5)$$

假设卷积之后的输出与输入相同, 则分离卷积与 3D 卷积计算量之比如下:

$$\frac{\text{depthwise} + \text{pointwise}}{\text{conv}} = \frac{K_c \times K_h \times K_w \times 1 \times K_c \times H \times W \times D + K_c \times 1 \times 1 \times K_d \times K_c \times H \times W \times D}{K_c \times K_h \times K_w \times K_d \times K_c \times H \times W \times D} = \frac{1}{K_d} + \frac{1}{(K_h \times K_w)} \quad (6)$$

由此可得, 卷积核越大时, 分离卷积的计算量相对于3D卷积越少。

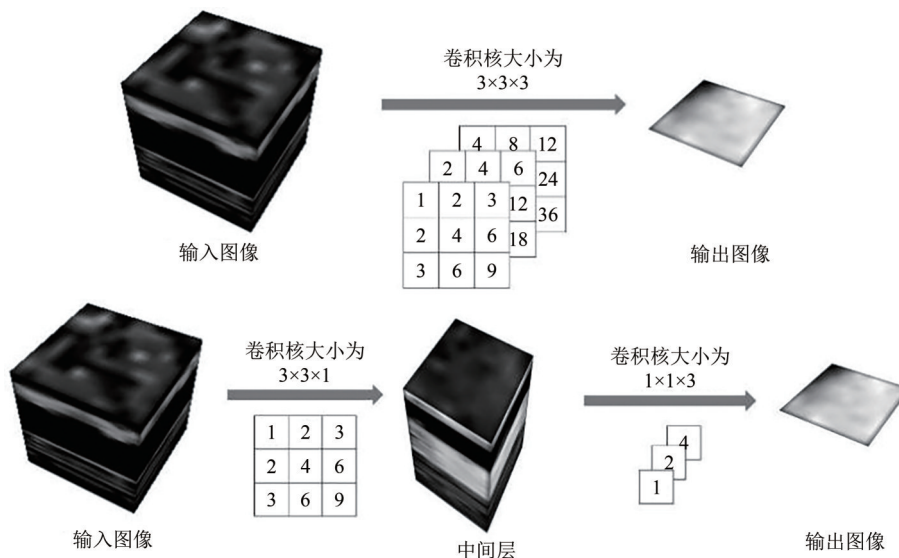


图1 使用分离卷积代替3D卷积

Fig. 1 Use the separable convolution instead of 3D convolution

2.2 密集连接

高光谱图像一般包括数百个波段, 地物的类别越多, 就需要更加细致、更加丰富的特征提取, 对于深度学习来说是一个挑战。传统的卷积神经网络为了提高性能一般是加深网络层数, 例如VGG (Visual Geometry Group Network) (Simonyan and Zisserman, 2014), 或者加宽网络结构Inception (Szegedy 等, 2015), 但是这种方法给网络带来了更多的参数, 对于高光谱图像数据难标注、数据小的问题, 容易出现过拟合的现象。

Huang 等 (2017) 提出了密集连接卷积神经网络 (DenseNet), 解决了这一问题, 如图2所示, 该网络结构的思想是尽可能的保留输入层到输出层的信息传递, 在每一层的输入之前, 将之前所有层的输出进行连接, 作为一个整体的输入连接到下一层。 $layer_m$ 表示第 m 层输出, H_m 表示第 m 层的卷积等操作, 计算过程如下:

$$layer_m = H_m(layer_0 + layer_1 + \cdots + layer_{m-1}) \quad (7)$$

2.3 Dropout

Dropout是在神经网络训练时正则化的一种手

段 (Srivastava 等, 2014), 如图3所示, 它的原理是使某一层的每个神经元以 P 的概率失活, 把输入 x 通过修改后的网络进行前向传播, 然后把损失结果通过修改后的网络后向传播, 在第二次前向传播的过程中, 激活全部神经元, 再次使每个神经元以 P 的概率失活, 不断重复以上过程。这种方法减少了神经元之间的相互作用, 避免对某一特征的过度依赖。

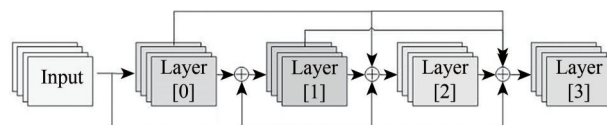


图2 密集连接结构

Fig 2 Dense block

由于采用大量的训练参数和有限的训练样本, 使得基于深度学习高光谱分类中经常存在过拟合问题。因此, 本文在网络中加入了Dropout来防止这种现象。

2.4 归一化层 (Batch Normalization)

神经网络训练的过程本质是拟合训练数据的过程, 当第一层参数出现变化时势必会影响后面每一层的数据与参数变化, 这种现象称为内部协

变量偏移 (Internal Covariate Shift) (Santurkar 等, 2018); 造成这种现象的原因是在网络训练时不同批次的数据分布可能差距过大, 网络为了拟合不同分布的数据会不断波动, 导致网络难收敛, 泛化能力差, BN 层 (Ioffe 和 Szegedy, 2015) 的提出解决了这个问题, 它的作用是将不同的数据分布拉成均值为 0, 方差为 1 的标准正态分布, 加快了网络收敛, 同时缓解神经网络中的“梯度弥散”问题。以一个批次 (batch) 激活层函数的输出作为输入, $B = \{x_1, \dots, x_m\}$; 计算过程如下:

$$\mu_B \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad (8)$$

$$\sigma_B^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (9)$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (10)$$

$$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv BN_{\gamma, \beta}(x_i) \quad (11)$$

2.5 网络结构

网络结构如图 4 所示, 首先对全部数据进行归

一化处理, 降低不同维度数据之间的差异性, 提高网络梯度下降求最优解的速度, 假设图像光谱波段数为 N , 以 $11 \times 11 \times N$ 的数据块作为输入数据。网络结构整体分为光谱信息提取, 光谱压缩, 空间信息与光谱信息融合, 投票求解共 4 个部分。

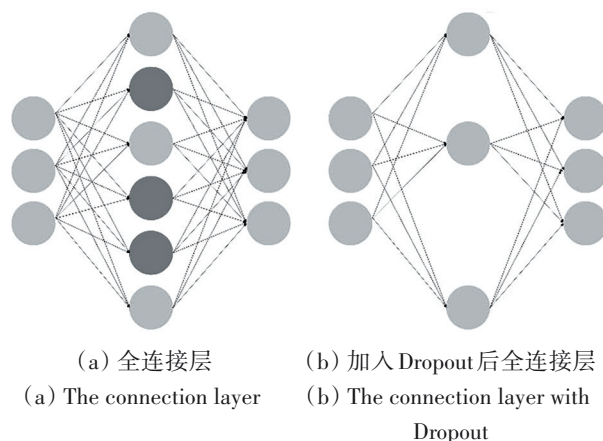


图 3 全连接层和加入 Dropout 后全连接层

Fig. 3 The connection layer and the connection layer with Dropout

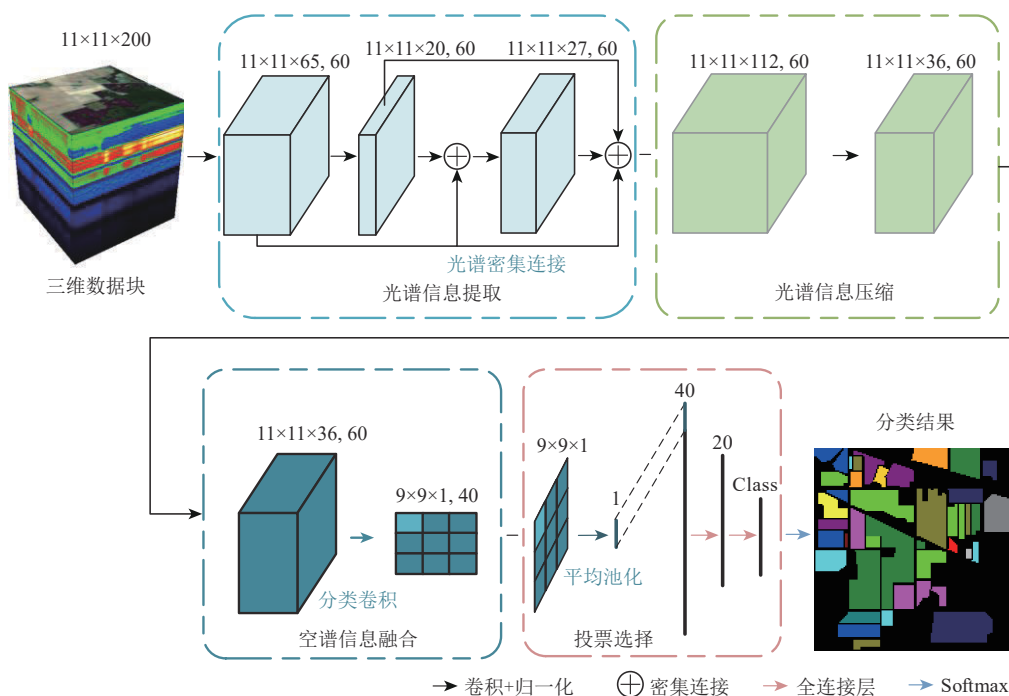


图 4 网络结构

Fig. 4 The network structure

在光谱信息提取部分, 目的是对像素点的光谱信息进行深层次提取, 密集连接支持特征图的复用, 使用小量卷积核可以生成大量深层特征, 搭建 3 层卷积层, 每层使用 $1 \times 1 \times 7$ 的卷积核进行光谱信息提取, 并保持空间信息的独立性, 个

数设置为 60。结合 DenseNet 思想, 将第一层网络输出与第二层网络输出进行光谱维度拼接, 输入到第三层网络; 将第一层、第二层与第三层的输出同样进行维度拼接输入到光谱压缩部分。

在光谱压缩部分, 使用 $1 \times 1 \times 7$ 的卷积核,

设置步长为3，对光谱维进行压缩，通过减少特征图的大小，降低更深层网络的参数量。

在空间信息与光谱信息融合部分，目的是以 $3 \times 3 \times K$ 的感受野同时整合数据的空间信息和光谱信息，其中 K 的值等于输入特征图的光谱维度。使用3D卷积会引入大量参数，因此使用可分离卷积代替3D卷积，将 $3 \times 3 \times K$ 的卷积核分解为 $3 \times 3 \times 1$ 的空间卷积和 $1 \times 1 \times K$ 的光谱卷积，此时输出40张 $9 \times 9 \times 1$ 的特征图。

投票的含义是指如果大部分像素点的输出为同一个值，则所有值的平均值也会拉到这个值的附近。因此在投票求解部分，使用无参的全局平均池化，将 $9 \times 9 \times 1$ 的特征图投票得出 $1 \times 1 \times 1$ 的输出值，将其他总共40个输出值拼接输入两个全连接层，通过softmax输出分类。为了避免归一化层和Dropout层同时使用带来的方差偏移（Variance Shift）（Li 等，2019b），BN层放在卷积层之后，Dropout设置在网络最深层的全连接层之后。

3 实验和结果分析

3.1 数据集描述

在本文中，基于Indian Pines（IP）、Pavia University（PU）和Kennedy Space Center（KSC）作为研究对象，数据集从http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes [2020-08-11] 获得。

（1）Indian Pines（IP）：Indian Pines数据集（图5）是由机载可见红外成像光谱仪（AVIRIS）于1992年对美国印第安纳州一块区域进行成像，空间分辨率为20 m，包括苜蓿、免耕玉米地、玉米幼苗、玉米、草地—牧场、草地—树木、修剪后草地、干草—料堆、燕麦、免耕大豆地、大豆幼苗、修剪后的大豆地、小麦、树林、建筑—草—树木—机器、石头—钢—塔等16个地物类型（图5），成像波长400—2500 nm，共220个波段，但由于第104—108波段、第150—163波段和第220波段不能被水反射，因此在实验中剔除这20个波段，选取其他200个波段作为研究对象。

（2）Pavia University（PU）：Pavia University数据集（图6）是由德国的机载反射光学光谱成像仪ROSIS-03（Reflective Optics Spectrographic Imaging System）于2003年对意大利北部帕维亚上空飞行

时捕捉到的场景，空间分辨率为1.3 m，包括柏油路、草地、碎石地、树木、金属板、裸地、沥青路、瓷砖路、阴影等9个地物类型（见图6），成像波长430—860 nm，共115个波段，在实验中剔除了12个噪声波段，选取其他103个波段作为研究对象。

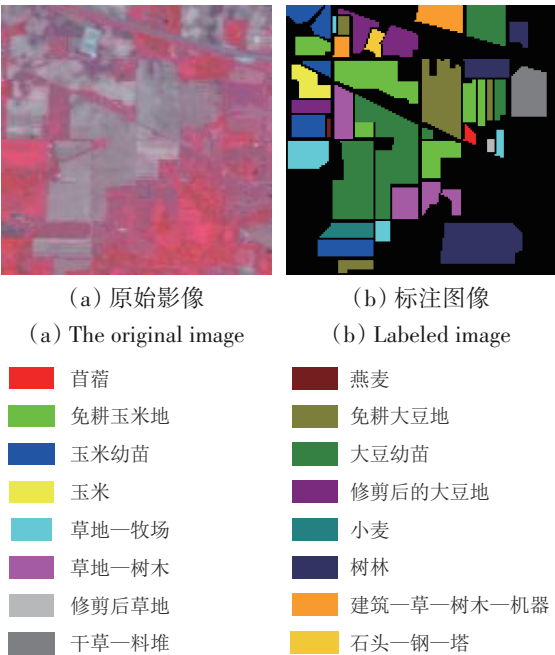


图5 Indian Pains数据集

Fig. 5 Indian Pains data set

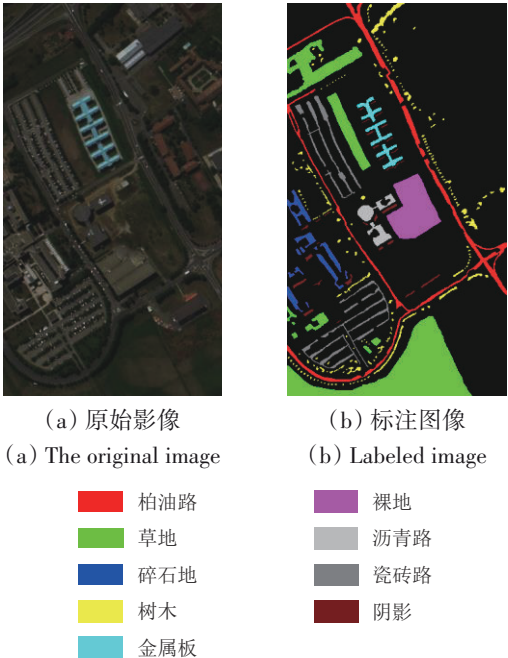


图6 Pavia University数据集

Fig. 6 Pavia University dataset

（3）Kennedy Space Center（KSC）：Kennedy Space Centery数据集是由机载可见红外成像光谱仪

(AVIRIS) 于 1996 年佛罗里达州肯尼迪航天中心 (KSC) 成像取得, 空间分辨率 18 m, 如图 7 所示, 由于某些植被类型的光谱特征相似, 很难区分这种环境下的土地覆盖。为了进行分类, 定义了 13 个类别, 代表该环境中出现的各种土地覆盖类型, 涵盖多种湿地及沼泽, 共 13 种类型 (见图 7), 图像成像波长 400—2500 nm, 共 224 波段, 剔除噪声波段, 选用剩余的 176 个波段作为研究对象。

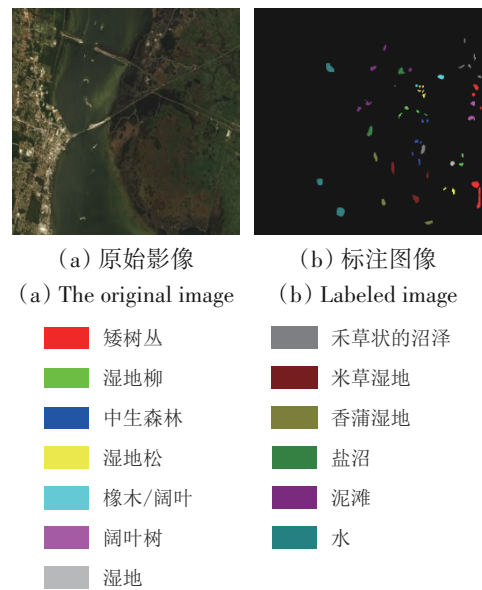


图 7 Kennedy Space Center 数据集
Fig. 7 Kennedy Space Center data set

3.2 实验设置

所有的实验基于 Pytorch 框架, CPU 为 Intel Xeon Silver 4114、Memory 为 256 G、GPU 为 NVIDIA GeForce TITAN Xp。对于本文提出的方法, 初始学习率设为 0.0005, 采用学习率自动下降的方法, 迭代次数设为 200, 每个批次训练集大小为 32; 其它方法按照论文作者所设置的参数; 同时, 为了减少训练样本带来的随机性, 每个试验重复 10 次取平均值作为最终试验结果。

实验采用分层抽样的方式, 每个类别按一定比例抽取训练样本, 舍弃每个类别抽取相同数量训练样本的抽样方式, 虽然会增加样本不均衡问题, 但是这种方法会更接近地面真实数据分布。

3.3 用于对比试验的分类器描述

为了证明比较实验的有效性, 本文选取了经典的机器学习算法和多种先进的深度学习方法作对比试验:

- (1) SSRN: Zhong 等 (2018) 提出的 3D 空谱联合深度学习网络, SOTA 算法;
- (2) SVM-RBF: 基于高斯核函数的支持向量机;
- (3) MDGCN: Wan 等 (2020) 提出的多尺度动态图卷积网络用于高光谱图像分类;
- (4) DBDA: Li 等 (2020) 提出基于空间注意力机制与光谱注意力机制的双分支神经网络, 通过搭建光谱分支与空间分支提取信息后进行信息融合;
- (5) pResNet: Paoletti 等 (2019) 通过使用金字塔形残差块搭建深层卷积网络模型。

3.4 实验及结果

为验证分类效果, 本实验采用总体精度 OA (Overall Accuracy)、平均精度 AA (Average Accuracy) 和 Kappa 系数 (Kappa) 进行评价。

第 1 个实验采用 Indian Pines 数据集, 按照分层抽样的方法, 每个类别选取 5% 的数据作为训练集, 实验结果如表 1 和图 8 所示。

表 1 Indian Pines 数据集不同方法分类精度
Table 1 Accuracy of Indian Pines data set classification by different methods

类别	SSRN/ %	SVM/ %	MDGCN/ %	DBDA/ %	pResNet/ %	SDLN/ %
1	98.059	35.166	85.714	95.541	77.150	96.183
2	95.188	62.788	90.598	93.988	65.809	96.925
3	92.568	68.520	90.166	90.311	61.052	96.541
4	93.907	52.792	94.737	96.240	60.418	96.613
5	99.475	86.194	97.921	97.506	79.161	97.612
6	97.560	85.503	96.038	99.093	82.329	98.678
7	92.777	75.544	75.000	76.515	56.801	77.959
8	98.734	89.322	99.982	99.161	90.745	99.976
9	89.025	59.873	62.963	92.144	43.889	79.744
10	90.690	68.687	89.666	95.804	67.697	95.080
11	94.028	69.971	99.636	96.692	73.765	98.109
12	95.473	61.247	95.564	95.797	60.223	97.147
13	99.676	87.175	99.001	99.626	79.505	99.001
14	97.166	89.366	94.540	98.573	88.670	99.120
15	93.230	69.364	98.714	95.886	68.498	96.969
16	94.064	98.610	94.252	96.269	89.124	94.461
OA	94.834	73.749	94.855	95.510	72.793	97.393
AA	95.101	72.508	91.530	94.946	71.552	95.007
Kappa	0.9411	0.6986	94.141	0.9490	0.6881	0.9702

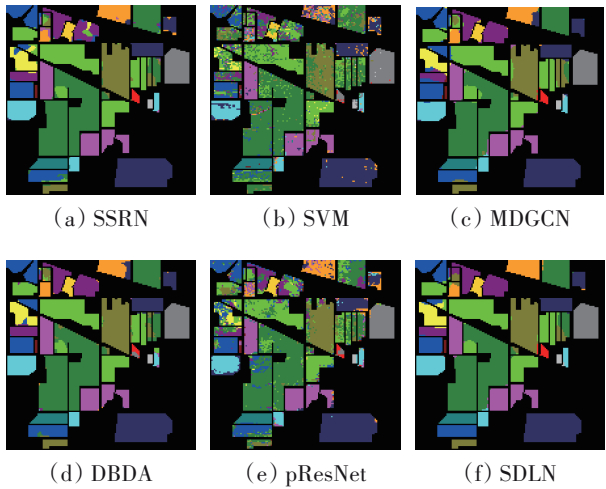


图8 不同方法的Indian Pines分类结果图

Fig. 8 Results of Indian Pines data set classification by different methods

由表1给出的实验结果可以看出，本文提出的网络对于IP数据集有更好的分类性能，使用5%数据集训练，整体精度达到97.4%，平均精度达到95.0%，Kappa达到0.97，相比于双分支双注意力机制网络，OA提高1%，Kappa提高0.21，相比于支持向量机，整体性能提高20%以上；SDLN相比于pResNet、SVM的结果基本消除了“椒盐现象”。

第2个实验采用Pavia University数据集，按照分层抽样的方法，每个类别选取0.5%的数据作为训练集，实验结果如表2和图9所示。

表2 Pavia University数据集不同方法分类精度
Table 2 Accuracy of Pavia University data set classification by different methods

类别	SSRN/ %	SVM/ %	MDGCN/ %	DBDA/ %	pResNet/ %	SDLN/ %
1	95.707	80.418	96.047	93.568	78.042	97.466
2	97.865	83.945	95.644	98.768	87.518	99.276
3	82.091	54.416	99.563	91.290	43.050	95.003
4	98.835	93.146	97.518	97.765	91.741	98.188
5	99.865	96.850	99.333	98.821	93.221	98.381
6	93.118	82.341	99.930	96.812	71.358	99.392
7	91.323	40.515	99.982	99.814	53.459	99.007
8	84.157	71.137	80.678	86.098	64.363	88.689
9	99.068	99.957	99.889	98.477	86.400	98.268
OA	94.571	81.688	95.139	96.081	78.992	97.578
AA	93.559	78.081	96.510	95.824	74.350	97.074
Kappa	0.9280	0.7481	0.93495	0.9479	0.7172	0.9678

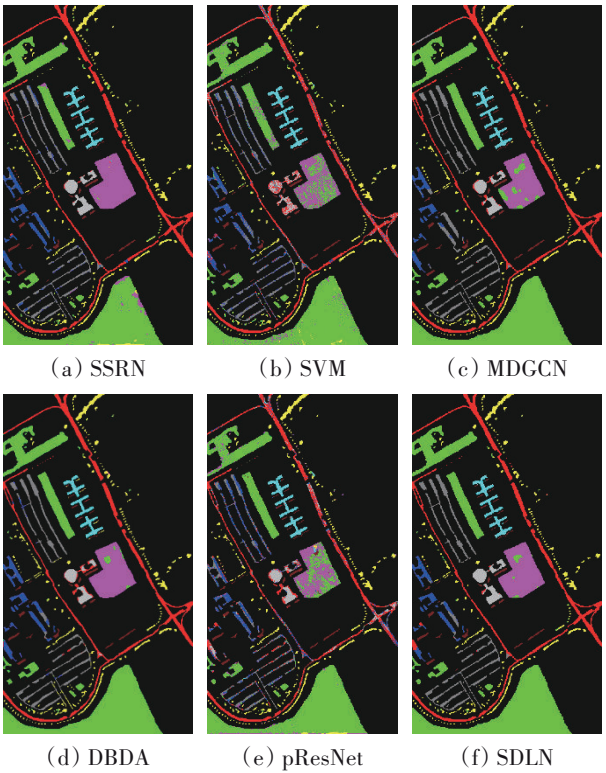


图9 不同方法的Pavia University分类结果图

Fig. 9 Results of Pavia University data set classification by different methods

从表2给出的实验数据中可以看出，在使用0.5%数据集训练，OA、AA、Kappa分别达到97.6%，97.1%，0.968，优于其他方法，草地、裸土、沥青路这3种地物的分类准确率高高于99%，柏油路、树木、金属板、阴影这4种地物的分类准确率高高于97%；第8类是瓷砖路，5种方法对该类的分类精度不高，错分的情况比较多，本文提出的方法取得了最好的结果，相比pResNet网络提高23.8%。

第3个实验采用Kennedy Space Center数据集，按照分层抽样的方法，每个类别选取5%的数据作为训练集，实验结果如表3和图10所示。

从表3给出的实验数据中可以看出，在使用5%数据集训练，OA、AA、Kappa分别达到99.2%，98.8%，0.991，同时在沼泽、硬木、米草湿地、香蒲湿地、水这5个类别中分类准确率达到到了100%；从图10可以看出，本文方法的分类结果更接近地面的真实值。

3.5 消融实验

为了验证密集连接和可分离卷积的有效性，删除了密集连接、使用3D卷积替换可分离卷积并进行了实验。由于删除了密集连接，网络输出特

征图的第3个光谱维度减少，深层的卷积核大于特征图的维度，因此对卷积核进行了缩放处理。实验结果见图 11，从中可以得出，加入密集连接和可分离卷积使网络平均有 2.7% 以上的精度提高。

表 3 Kennedy Space Center 数据集不同方法分类精度
Table 3 Accuracy of Kennedy Space Center data set classification by different methods

类别	SSRN/ %	SVM/ %	MDGCN/ %	DBDA/ %	pResNet/ %	SDLN/ %
1	99.288	92.535	99.236	99.462	97.400	99.857
2	96.727	87.377	98.155	95.589	83.854	98.519
3	87.589	73.414	90.443	94.768	62.179	95.803
4	84.444	53.638	84.670	87.295	55.113	97.024
5	80.502	65.847	85.512	88.836	48.258	94.911
6	96.459	66.756	98.564	99.616	74.216	99.622
7	89.117	75.000	92.434	92.072	72.641	100
8	98.167	87.482	98.851	98.416	79.154	100
9	98.674	88.440	100	99.978	80.926	100
10	98.551	97.841	100	100	87.398	100
11	99.196	96.041	100	99.695	99.499	98.713
12	99.550	94.552	98.732	99.668	88.522	99.780
13	99.988	99.709	100	100	99.642	100
OA	96.552	88.270	97.381	97.946	85.510	99.212
AA	94.481	82.972	95.892	96.569	79.139	98.787
Kappa	0.9616	0.8693	97.083	0.9771	0.8386	0.9912

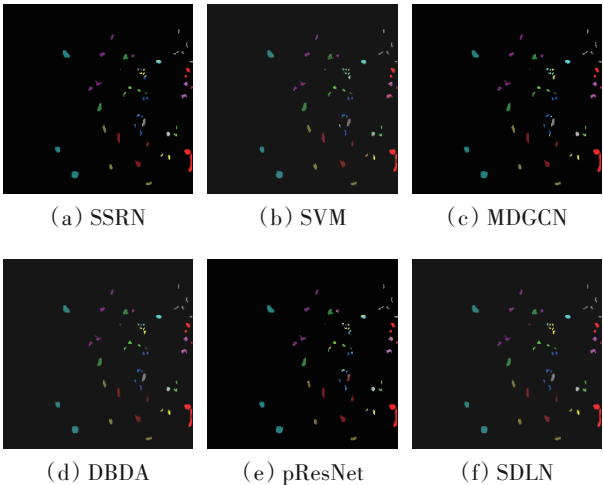


图 10 不同方法的 Kennedy Space Center 分类结果图
Fig. 10 Results of Kennedy Space Center data set classification by different methods

3.6 不同数量训练样本对结果的影响

本文对算法进行了进一步的评估，在 IP、UP、

KSC 这 3 个数据集上，设置不同大小的训练数据，测试了不同算法的泛化能力。

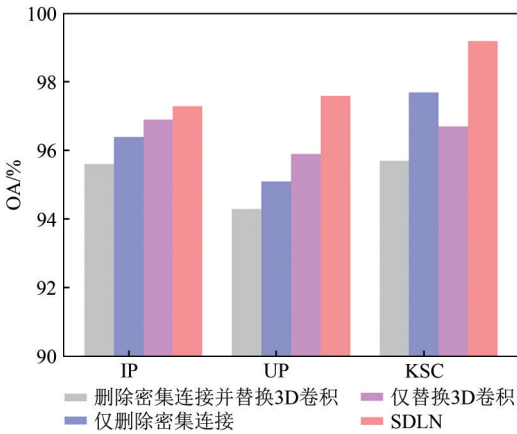


图 11 密集连接和可分离卷积的有效性
Fig. 11 Effectiveness of the Dense Connection and Separable Convolution

从图 12 给出的分析结果中可以看出，随着数据集的增加，各种方法的整体精度、平均精度、Kappa 系数都不断提高，精度差距缩小。本文提出的网络即使在小样本的情况下，也取得较好的分类表现，达到了最高分类精度。对于 IP 数据集，使用 3% 数据作为训练，总体分类精度可以达到 96%，使用 5% 作为训练，总体分类精度可以达到 97%；对于 Pavia University 数据集，该数据集的空间分辨率高，混合像元少，使用 0.6% 的数据足以使整体分类精度达到 98%，使用更大的数据对提升分类精度效果不大；对于 KSC 数据集，本文提出的网络在 OA、AA 指标上都优于其他算法，使用 5% 数据训练，OA 达到 99.2%。

3.7 测试不同大小训练样本输入对结果的影响

在 IP 数据集进行输入数据大小对分类精度的测试，Batchsize 为 32，Epoch 为 200，将全部数据集按 5%：95% 的比例划分为训练集、测试集。如图 13 所示，左侧代表实验指标（Kappa×100），右侧代表训练所需时间（单位为秒）。结果表明，增大输入数据的尺寸能使数据带有更多的空间信息，也会使每一层的特征图尺寸增加，造成网络参数增加，训练时间加长。当输入数据为 9 × 9 和 11 × 11 时，网络能达到精度与效率的平衡；因为 IP 数据集的空间分辨率为 20 m，当输入数据为 13 × 13 时，太大的空间信息对于网络分类精度的提升并没有帮助，效果反而下降。

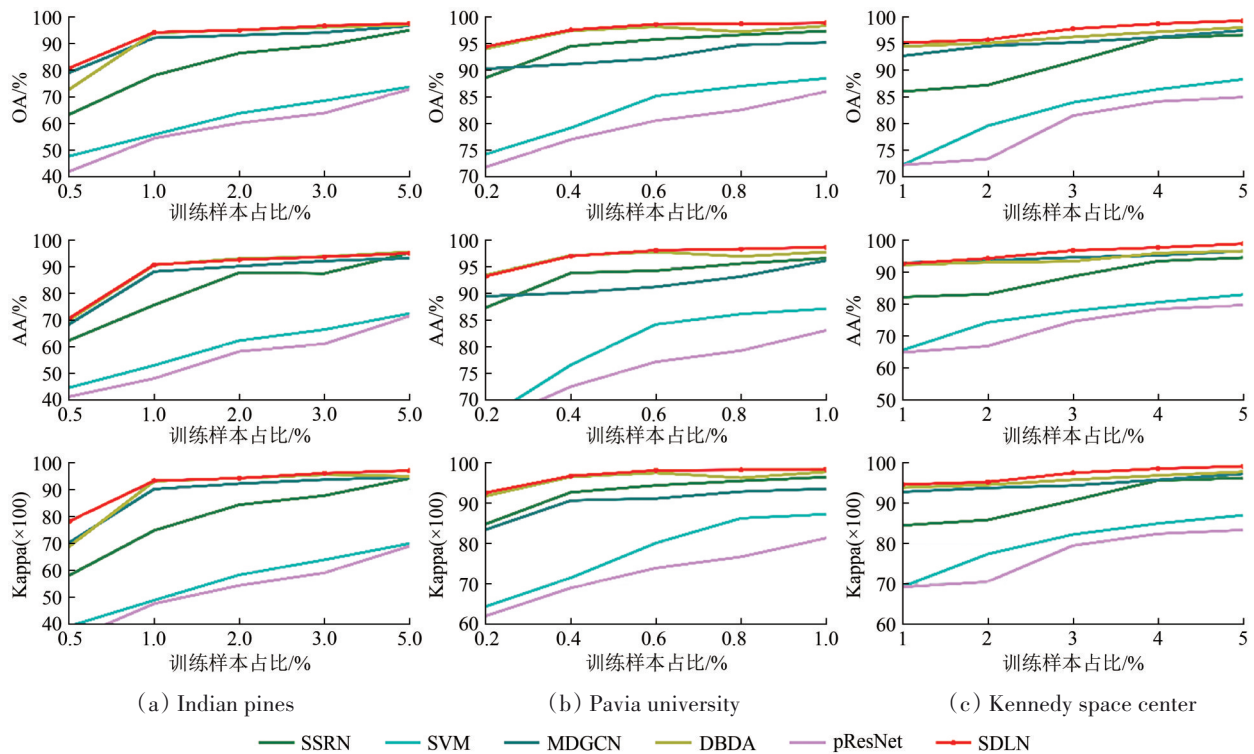


图 12 SSRN, SVM, MDGCN, DBDA, pResNet 和本文方法在 IP, PU, KSC 中不同训练集大小下的 OA、AA、Kappa 结果

Fig. 12 The OA, AA, Kappa results of SVM, MDGCN, DBDA, pResNet and SDLN methods with varying proportions of training samples on the IP, PU

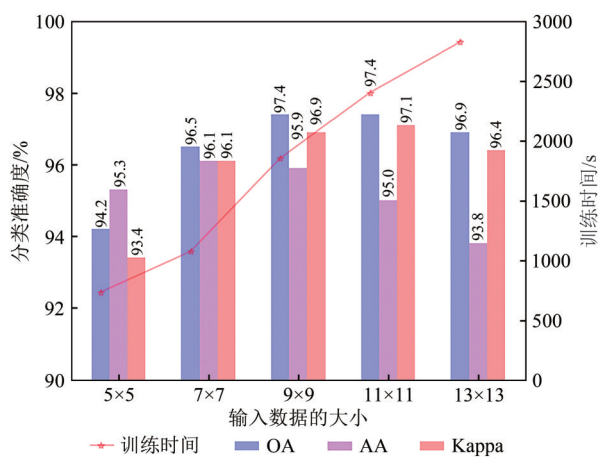


图 13 不同大小输入数据分类精度与训练所需时间

Fig. 13 Classification accuracy and training time of input data of different sizes

3.8 比较各个算法的运行时间

对 IP 数据集测试了不同算法训练测试所需的时间, Batchsize 为 16, Epoch 为 200, 输入数据设定为 5×5 的数据块, 将全部数据集按 5%: 95% 的比例划分为训练集、测试集。如图 14 所示, 柱状图代表训练所需时间, 折线图代表测试所需时

间 (单位为秒)。

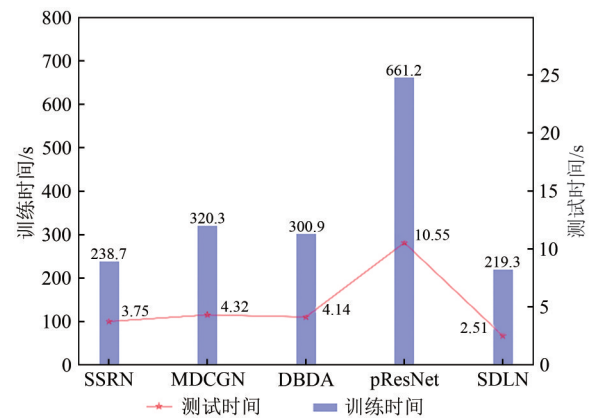


图 14 不同算法训练测试所需时间

Fig. 14 Time required for training and testing of different algorithms

由图 14 可知, 在相同数据大小输入下, 本文提出的网络在速度上达到了最优, 网络使用浅层密集连接, 减少了卷积核参数量, 增加了信息传递效率, 在空间特征提取部分使用无参数的全局平均池化, 减少了训练参数量。相比 pResNet 网

络, 训练时间缩短 66.8%, 测试时间缩短 76.5%。

4 结 论

为了提高小样本下高光谱图像分类的精度和速度, 采用光谱密集连接方式增加了特征图的复用, 提高了算法的光谱特征提取能力, 使用分离卷积代替 3D 卷积提高算法运行效率, 基于 IP、PU 和 KSC 这 3 个数据集, 与 SSRN 等先进深度学习进行精度对比、消融实验、运行时间对比等多个实验, 证明了本文提出的算法可以有效提高高光谱图像的分类精度, 在精度和效率上都获得了理想的效果, 相比于 3D 卷积, 深度可分离卷积能够有效提高算法运行效率, 密集连接能够更好的提取地物的光谱信息, 提高分类精度。

同时也研究了训练数据大小与输入数据大小对分类精度的影响, 随着训练数据的增加或者输入数据的增大, 网络的分类精度提升, 但是超过一定界限, 过多的训练数据和过大的输入数据会带来冗余信息增加, 使分类精度降低。

通过实验发现在高光谱数据集和真实的遥感图像中普遍存在样本不均衡的问题, 每个地物的数量是不同的, 但是数量少的地物并不意味着重要程度低。不平衡的样本类别会造成分类模型的效果下降, 分类器会偏向于学习大的类别, 小类别数据的分类精度降低, 因此下一步将重点研究高光谱图像中类别不均衡问题, 提高小类别数据的分类效果。

志 谢 感谢珠海欧比特人工智能研究院提供的技术支持和山东省重点研发计划的基金资助, 在此表示衷心的感谢!

参考文献 (References)

- Archibald R and Fann G. 2007. Feature selection and classification of hyperspectral images with support vector machines. *IEEE Geoscience and Remote Sensing Letters*, 4(4): 674-677 [DOI: 10.1109/LGRS.2007.905116]
- Ball J E and Wei P. 2018. Deep learning hyperspectral image classification using multiple class-based denoising autoencoders, mixed pixel training augmentation, and morphological operations// *IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium*. Valencia: IEEE: 6903-6906 [DOI: 10.1109/IGARSS.2018.8519368]
- Chang C I, Zhao X L, Althouse M L G and Pan J J. 1998. Least squares subspace projection approach to mixed pixel classification for hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 36(3): 898-912 [DOI: 10.1109/36.673681]
- Cui B G, Ma X D and Xie X Y. 2017. Hyperspectral image de-noising and classification with small training samples. *Journal of Remote Sensing*, 21(5): 728-738 (崔宾阁, 马秀丹, 谢小云. 2017. 小样本的高光谱图像降噪与分类. *遥感学报*, 21(5): 728-738) [DOI: 10.11834/jrs.20176239]
- Fu W, Li S T and Fang L Y. 2015. Spectral-spatial hyperspectral image classification via superpixel merging and sparse representation// *2015 IEEE International Geoscience and Remote Sensing Symposium*. Milan: IEEE: 4971-4974 [DOI: 10.1109/IGARSS.2015.7326948]
- Gao Q S, Lim S and Jia X P. 2018. Hyperspectral image classification using convolutional neural networks and multiple feature learning. *Remote Sensing*, 10(2): 299 [DOI: 10.3390/rs10020299]
- He M Y, Li B and Chen H H. 2017. Multi-scale 3D deep convolutional neural network for hyperspectral image classification// *2017 IEEE International Conference on Image Processing (ICIP)*. Beijing: IEEE: 3904-3908 [DOI: 10.1109/ICIP.2017.8297014]
- Huang G, Liu Z, Laurens V and Weinberger K. Q. 2017. Densely connected convolutional networks// *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1063-6919 [DOI: 10.1109/CVPR.2017.243]
- Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T, Andreetto M and Adam H. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *ArXiv Preprint ArXiv: 1704.04861* [DOI: 10.48550/arXiv.1704.04861]
- Ioffe S and Szegedy C. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ArXiv Preprint ArXiv: 1502.03167* [DOI: 10.48550/arXiv.1502.03167]
- Jiao L C, Liang M M, Chen H, Yang S Y, Liu H Y and Cao X H. 2017. Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(10): 5585-5599 [DOI: 10.1109/TGRS.2017.2710079]
- Joelsson S R, Benediktsson J A and Sveinsson J R. 2005. Random forest classifiers for hyperspectral data// *Proceedings of IEEE International Geoscience and Remote Sensing Symposium, 2005 (IGARSS '05)* [DOI: 10.1109/IGARSS.2005.1526129]
- Kirsch M, Lorenz S, Zimmermann R, Tusa L, Möckel R, Hödl P, Booyens R, Khodadadzadeh M and Gloaguen R. 2018. Integration of terrestrial and drone-borne hyperspectral and photogrammetric sensing methods for exploration mapping and mining monitoring. *Remote Sensing*, 10(9): 1366-1366 [DOI: 10.3390/rs10091366]
- Krizhevsky A, Sutskever I and Hinton G E. 2012. ImageNet classification with deep convolutional neural networks// *Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe, Nevada: Curran Associates Inc.: 1097-1105

- [DOI: 10.1145/3065386]
- Li F, Lu H C and Zhang P P. 2019a. An innovative multi-kernel learning algorithm for hyperspectral classification. *Computers and Electrical Engineering*, 79: 106456-106464 [DOI: 10.1016/j.compeleceng.2019.106456]
- 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 W, Chen C, Su H J and Du Q. 2015. Local binary patterns and extreme learning machine for hyperspectral imagery classification. *IEEE Transactions on Geoscience and Remote Sensing*, 53(7): 3681-3693 [DOI: 10.1109/TGRS.2014.2381602]
- Li X, Chen S, Hu X L and Yang J. 2019b. Understanding the disharmony between dropout and batch normalization by variance shift// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA: IEEE: 2677-2685 [DOI: 10.1109/CVPR.2019.00279]
- Liu X F, Sun Q Q, Meng Y, Fu M and Bourennane S. 2018. Hyperspectral image classification based on parameter-optimized 3D-CNNs combined with transfer learning and virtual samples. *Remote Sensing*, 10(9): 1425-1425 [DOI: 10.3390/rs10091425]
- Makantasis K, Karantzalos K, Doulamis A and Doulamis N. 2015. Deep supervised learning for hyperspectral data classification through convolutional neural networks//2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Milan: IEEE: 4959-4962 [DOI: 10.1109/IGARSS.2015.7326945]
- Mei S H, Ji J Y, Hou J H, Li X and Du Q. 2017. Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8): 4520-4533 [DOI: 10.1109/TGRS. 2017. 2693346]
- 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]
- Paoletti M E, Haut J M, Fernandez-Beltran R, Plaza J, Plaza A J and Pla F. 2019. Deep pyramidal residual networks for spectral-spatial hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2): 740-754 [DOI: 10.1109/TGRS. 2018.2860125]
- Prey L, von Bloh M and Schmidhalter U. 2018. Evaluating RGB imaging and multispectral active and hyperspectral passive sensing for assessing early plant vigor in winter wheat. *Sensors*, 18(9): 2931 [DOI: 10.3390/s18092931]
- Rao M, Tang P and Zhang Z. 2020. A developed siamese CNN with 3D adaptive spatial-spectral pyramid pooling for hyperspectral image classification. *Remote Sensing*, 12(12): 1964 [DOI: 10.3390/rs12121964]
- Redmon J and Farhadi A. 2018. YOLOv3: An Incremental Improvement. *ArXiv E-prints* [DOI: 10.48550/arXiv.1804.02767]
- Santurkar S, Tsipras D, Ilyas A, Madry A. 2018. How does batch normalization help optimization? //NeurIPS 2018 [DOI: 10.48550/arXiv.1805.11604]
- Simonyan K and Zisserman A. 2014. Very Deep Convolutional Networks for Large-scale Image Recognition//2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [DOI: 10.48550/arXiv.1409.1556]
- Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1): 1929-1958 [DOI: 10.1109/cisp.2015.7407967]
- Sun W W, Zhang D F, Yang G and Li W Y. 2018. Band selection for hyperspectral imagery based on weighted probabilistic archetypal analysis. *Journal of Remote Sensing*, 22(1): 110-118 (孙伟伟, 张殿发, 杨刚, 李巍岳. 2018. 加权概率原型分析的高光谱影像波段选择. *遥感学报*, 22(1): 110-118) [DOI: 10.11834/jrs.20186446]
- Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. 2015. Going deeper with convolutions//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA: IEEE: 1-9 [DOI: 10.1109/CVPR. 2015.7298594]
- Wan S, Gong C, Zhong P, Du B, Zhang L F and Yang J. 2020. Multi-scale dynamic graph convolutional network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5): 3162-3177 [DOI: 10.1109/TGRS.2019.2949180]
- Wang L Q, Zhao X and Qin Z C. 2021. Hyperspectral image classification with spectral-spatial consistency regularization. *Journal of China Academy of Electronics and Information Technology*, 16(8): 789-796 (王雷全, 赵欣, 秦智超. 2021. 基于光谱-空间一致性正则化的高光谱图像分类. *中国电子科学研究院学报*, 16(8): 789-796) [DOI: 10.3969/j.issn.1673-5692.2021.08.007]
- Zhang K, Hei B Q, Zhou Z and Li S Y. 2018. CNN with coefficient of variation-based dimensionality reduction for hyperspectral remote sensing images classification. *Journal of Remote Sensing*, 22(1): 87-96 (张康, 黑保琴, 周壮, 李盛阳. 2018. 变异系数降维的CNN高光谱遥感图像分类. *遥感学报*, 22(1): 87-96) [DOI: 10.11834/jrs.20187075]
- Zheng Q, Huang W J, Cui X M, Dong Y Y, Shi Y, Ma H Q and Liu L Y. 2018. Identification of wheat yellow rust using optimal three-band spectral indices in different growth stages. *Sensors*, 19(1): 35 [DOI: 10.3390/s19010035]
- 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 J and Tian Q J. 2008. Image quality evaluation of EO-1 hyperion sensor. *Geo-information Science*, 10(5): 678-683 (周雨霁, 田庆久. 2008. EO-1 Hyperion 高光谱数据的质量评价. *地球信息科学*, 10(5): 678-683) [DOI: 10.3969/j.issn.1560-8999.2008. 05.020]

Hyperspectral image classification of the deep neural network based on 3D convolution and dense connection

SONG Tingqiang¹, ZONG Da¹, LIU Tongxin¹, FAN Haisheng², HUANG Tengjie²,
JIANG Xiaoxu², WANG Haoyu¹

1. College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China;

2. Zhuhai Orbita Aerospace Science & Technology Co., Ltd. Artificial Intelligence Research Institute, Zhuhai 519080, China

Abstract: With the progress of deep learning, researchers are increasingly paying attention to its application in hyperspectral image classification. Many experiments are conducted to achieve a trade-off between accuracy and efficiency to improve the feature extraction performance of neural networks toward small training sample sets.

This work has proposed a high-speed and high-precision neural network structure based on spatial spectral information. A cascaded neural network for spectral spatial information extraction is constructed by combining the idea of DenseNet and adopting dilated convolutions instead of 3D convolutions as the main calculation method. The whole network structure is divided into four components: spectral information extraction, spectral compression, fusion of spatial and spectral information, and voting solution.

Three convolutional layers are built in the spectral information extraction component. In each layer, $1 \times 1 \times 7$ convolution kernels are used to extract spectral information and maintain the independence of spatial information. The number of kernels is set to 60. In light of the DenseNet idea, the network outputs of the first and second layers are dimensionally split in spectrum and inputted into the third layer. The outputs of the first, second, and third layers are also dimensionally split and inputted into the spectral compression component.

In the spectral compression component, a $1 \times 1 \times 7$ convolution kernel is used with a step size set to three. The spectral dimension is compressed, and the number of parameters of the deeper network is lessened by reducing the size of the feature map.

In the spatial and spectral information fusion component, the goal is to fuse spatial information for the first time with 3×3 receptive fields and integrate the spectral information of the data. Separable convolutions are adopted instead of traditional 3D convolutions, and the $3 \times 3 \times K$ convolution kernel is decomposed into a $3 \times 3 \times 1$ convolution and a $1 \times 1 \times K$ convolution. The value of K is equal to the spectral dimension of the input feature map. Then, $40 \times 9 \times 9 \times 1$ feature maps are outputted.

Voting means that if the output of most pixels is the same value, then the average value of all values will also be pulled near this certain value. In the voting solution component using parameter-free global average pooling, the $9 \times 9 \times 1$ feature maps are voted to obtain $1 \times 1 \times 1$ output values. These 40 output values are spliced into the fully connected layer, and the classification results are outputted through Softmax.

A series of experiments were carried out on the Indian Pines and Pavia University and Kennedy Space Center datasets. In the IP data set, the average accuracy reaches 95.0%, the overall accuracy 97.4%, and Kappa 0.97 by training with 5% data sets. In the UP data set, OA, AA, and Kappa reach 97.6%, 97.1%, and 0.97, respectively, by training with a 0.5% data set. The overall accuracy in the KSC data set can reach 99.2%. The network has been proven to strong feature extraction and classification ability.

This method effectively improves the classification accuracy of hyperspectral images in the case of small sample sets and studies the effect of training and input data sizes on the classification accuracy. The classification accuracy of the network is improved with the increase in the training or input data. However, redundant information generated by a large amount of training data and excessive input data does not help improve the classification performance.

Key words: hyperspectral image classification, deep learning, lightweight network, dense connection, separable convolution

Supported by 2019 Shandong Provincial Key Research and Development Program Project (No. 2019GGX101047); Cloud Service Platform and Applications based on ZHUHAI No.1 Constellation & Remote Sensing Big Data Project (No. ZH0111-0405-170027-P-WC)