

基于扩张卷积注意力神经网络的高光谱图像分类

张祥东*, 王腾军, 朱劭俊, 杨耘

长安大学地质工程与测绘学院, 陕西 西安 710054

摘要 为了解决训练样本有限情况下高光谱图像分类精度低的问题,提出了一种结合扩张卷积与注意力机制的三维-二维串联卷积神经网络模型。首先,该模型以串联的三维-二维卷积神经网络作为基础结构,利用三维卷积同时提取高光谱图像的空谱特征,并采用二维卷积进一步提取高级空间语义信息;然后,通过引入扩张卷积增大卷积核感受野,构建了多尺度特征提取结构,实现了多尺度特征的融合;最后,利用注意力机制使网络关注重要的空谱特征,并抑制噪声和冗余信息。在两个常用数据集上对本文算法和四种基于深度学习的分类算法进行对比实验,结果表明,所提模型取得了最准确的分类结果,有效提高了训练样本有限条件下的分类精度。

关键词 图像处理; 高光谱图像分类; 卷积神经网络; 扩张卷积; 注意力机制

中图分类号 TP751

文献标志码 A

doi: 10.3788/AOS202141.0310001

Hyperspectral Image Classification Based on Dilated Convolutional Attention Neural Network

Zhang Xiangdong*, Wang Tengjun, Zhu Shaojun, Yang Yun

School of Geology Engineering and Geomatics, Chang'an University, Xi'an, Shaanxi 710054, China

Abstract In order to solve the problem of low classification accuracy of hyperspectral images in the case of limited training samples, we proposed a tandem three-dimensional(3D)-two-dimensional(2D) convolutional neural network model combining dilated convolution with attention mechanism. First, with the tandem 3D-2D convolutional neural network as the basic structure, the model used 3D convolution to simultaneously extract the spatial-spectral features of hyperspectral images and 2D convolution to further extract high-level spatial semantic information. Then, by introducing dilated convolution to enlarge the receptive field of the convolution kernel, we constructed a multi-scale feature extraction structure for the fusion of multi-scale features. Finally, the attention mechanism was applied to make the network pay attention to important spatial-spectral features and suppress noise and redundant information. Furthermore, we performed a comparative experiment between the proposed algorithm and four deep-learning-based algorithms on two common data sets. The results show that the proposed model achieves the most accurate classification results and effectively improves the classification accuracy under the condition of limited training samples.

Key words image processing; hyperspectral image classification; convolutional neural network; dilated convolution; attention mechanism

OCIS codes 100.2960; 100.4145; 100.4996

1 引言

随着高光谱成像技术的快速发展,高光谱图像

的空间分辨率和光谱分辨率不断提高,实现了光谱与图像的结合,具有“图谱合一”的特点。高光谱数据包含丰富的空间信息和光谱信息,其光谱维度由

收稿日期: 2020-07-06; 修回日期: 2020-08-10; 录用日期: 2020-09-08

基金项目: 长安大学中央高校基本科研业务费(300102269205, 300102269304)、国家重点研发计划项目(2018YFC1504805, 2019YFC1509201)

* E-mail: 18755150056@163.com

可见光到红外波长范围内的上百个连续波段组成,能够充分反映地表物体的光谱特征,大大提高地物的分辨能力。因此,高光谱图像被广泛应用于植被分类^[1]、精细化农业^[2-3]、环境监测^[4-5]和矿物填图^[6]等相关领域。

高光谱图像分类是利用其包含的空间信息和特有的光谱特征将图像中的每个像素划分到唯一的类别,是众多应用领域中的重要研究环节。由于高光谱数据存在光谱维度高、信息冗余和标签样本缺乏等问题,使得其在分类任务中面临着巨大的挑战^[7]。为了解决光谱维度高导致的分类精度随特征维数上升而下降的休斯现象,研究者们提出了主成分分析(PCA)^[8]、独立成分分析(ICA)^[9]和线性判别分析(LDA)^[10]等一系列降维方法,并采用支持向量机(SVM)^[11]、K-近邻法(K-NN)^[12]和随机森林(RF)^[13]等机器学习方法对高光谱图像进行分类。但是上述方法只考虑到了光谱信息,忽略了丰富的空间结构信息,没有取得理想的分类结果。

卷积神经网络(CNN)是通过模仿人类神经元的工作机制而特殊设计的人工神经网络,通常由卷积层、池化层、激活函数和全连接层组成。CNN通过局部连接、权重共享和空间降采样策略降低了普通神经网络的模型复杂度,并对平移、旋转和比例缩放等变形形式具有高度不变性,被广泛应用于图像分类、目标识别等领域。近年来,随着CNN的不断发展,基于CNN的高光谱图像分类方法已成为新的研究热点。2015年,Makantasis等^[14]提出了结合二维CNN(2D-CNN)和多层感知机的分类模型2D-CNN-MLP,利用2D-CNN提取空间特征,并通过多层感知机对提取后的特征进行聚合和分类,相比利用光谱信息的分类方法能够有效提高分类精度。2018年,李竺强等^[15]提出一种引入多标签条件随机场的三维卷积神经网络模型(3D-CNN-CRF),实现了空谱特征的同时提取,而且经过条件随机场优化后的分类精度得到了很大的提升。2020年,Roy等^[16]设计了一种同时采用三维卷积和二维卷积的混合网络模型(Hybrid-CNN),有效结合了3D-CNN和2D-CNN各自的优势,相比单独的3D-CNN模型能够降低网络的计算复杂度。为了避免池化层引起的特征信息丢失问题,颜铭靖等^[17]通过引入空洞卷积代替池化层,构建了三维空洞卷积残差神经网络(Dilated-3D-CNN),在不增加网络参数的情况下提高了高光谱影像地物分类的精度。虽然以上方法在高光谱数据集上取得了较高的分类精度,但都

存在着训练样本较少时分类精度较低的问题。

注意力机制最初被用于机器翻译等自然语言处理领域,直到Hu等^[18]提出的SENet获得了2017年ImageNet竞赛的分类任务冠军,才被广泛应用于计算机视觉领域。在用于图像处理的CNN模型中加入注意力模块能够实现类似于人类的视觉注意力机制,在关注重要信息的同时忽略低价值信息,提高有效信息的利用率。受Hybrid-CNN、Dilated-3D-CNN以及空间注意力模型^[19]的启发,本文设计了结合扩张卷积与注意力机制的串联3D-2D-CNN模型,以提高训练样本有限情况下高光谱图像的分类精度。本文的主要工作如下:

1) 设计了一种更加高效的3D-2D-CNN串联结构,其中3D-CNN实现了空间和光谱特征的同步提取,2D-CNN实现了高级抽象空间特征的提取,充分利用了高光谱图像中丰富的空间和光谱信息;

2) 基于扩张卷积构建了多尺度特征提取结构,实现了不同尺度下特征的融合,避免了单一尺度下特征信息的不完整表达;

3) 基于注意力机制构建了空谱注意力模块,使网络在训练过程中更多的关注有效的空谱特征,抑制空间邻域引入的噪声信息和波段间冗余信息;

4) 在两个公开的高光谱数据集上对本文方法与四种基于CNN的对比算法进行实验。结果表明,在训练样本极其有限的情况下,本文方法能够取得最高的分类精度。

2 研究方法

2.1 3D-2D-CNN 串联结构

早期用于图像识别的CNN采用的是二维卷积核,其对输入图像进行特征提取的过程是利用卷积核在图像上采用窗口滑动的方式依次提取局部特征,并将局部特征值通过激活函数映射到下一层的特征图中。2D-CNN中第*i*层的第*j*个特征图在坐标(*x*, *y*)处的特征值 v_{ij}^{xy} 可根据下式计算。

$$v_{ij}^{xy} = \delta \left(\sum_{k=1}^m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \omega_{ijk}^{pq} v_{(i-1)k}^{(x+p)(y+q)} + b_{ij} \right), \quad (1)$$

式中: P_i 和 Q_i 为第*i*层二维卷积核的大小; ω_{ijk}^{pq} 表示对*i*-1层第*k*个特征图进行卷积操作的卷积核; m 为*i*-1层特征图的个数; $v_{(i-1)k}^{(x+p)(y+q)}$ 为*i*-1层第*k*个特征图(*x*+*p*, *y*+*q*)位置的特征值; b_{ij} 为偏置; $\delta(\cdot)$ 表示激活函数。

由于2D-CNN在处理三维数据时具有一定的局限性,针对此问题文献^[20]提出了3D-CNN用

于识别视频数据中的人体动作。与 2D-CNN 类似,3D-CNN 利用三维卷积核在输入的立方体数据中采用滑动的方式依次进行卷积操作,第 i 层第 j 个特征图在坐标 (x, y, z) 处特征值 v_{ij}^{xyz} 的计算方式为

$$v_{ij}^{xyz} = \delta \left(\sum_{k=1}^m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} \omega_{ijk}^{pqr} v_{(i-1)k}^{(x+p)(y+q)(z+r)} + b_{ij} \right), \quad (2)$$

式中: P_i 、 Q_i 和 R_i 为第 i 层三维卷积核的大小。

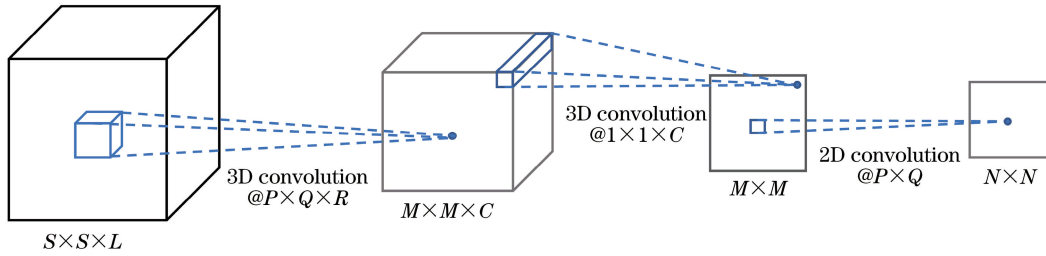


图 1 3D-2D-CNN 串联结构图

Fig. 1 Structure of tandem 3D-2D-CNN

对于形状为 $S \times S \times L$ 的高光谱立方体数据块,经过尺寸为 $P \times Q \times R$ 的三维卷积核提取联合的空谱特征,生成 $M \times M \times C$ 的特征图;再采用 $1 \times 1 \times C$ 的卷积核对其进行卷积操作,得到形状为 $M \times M$ 的二维特征图;最后利用尺寸为 $P \times Q$ 的二维卷积核在二维特征图上以窗口滑动的方式提取高级空间特征,经过非线性激活函数映射到形状为 $N \times N$ 的特征图中。3D-2D-CNN 串联结构将 3D-CNN 和 2D-CNN 进行转换和组合,直接以三维数据作为输入,在提取空谱特征的同时也能够提取更加抽象的空间特征,并具有计算复杂度低的特点。

2.2 扩张卷积

CNN 主要由交替的卷积层和池化层组成,其中卷积层的作用是提取局部特征,在卷积层后添加池

在高光谱图像分类任务中,3D-CNN 直接以原始图像作为网络输入,不需要复杂的预处理,能够同时提取空间特征和光谱特征,但具有计算复杂度高的问题。由于 2D-CNN 模型中采用的是二维卷积,相对 3D-CNN 包含较少的参数量,但是通常要采用主成分分析等降维方法来提取高光谱数据中具有代表性的波段。针对 2D-CNN 和 3D-CNN 单独使用时存在的不足,本文提出了一种 3D-2D-CNN 串联结构,如图 1 所示。

化层的目的是对特征图进行下采样操作,一方面,能够实现特征的压缩和聚合;另一方面,下采样减小了特征图尺寸,间接增大了后续卷积层的感受野,使其获取更大范围的特征信息。然而池化层在下采样的过程中会造成一部分信息的损失,对于样本缺乏的高光谱图像分类任务极其不利。针对此问题,本文在构造分类网络时选择不采用池化层,而是引入扩张卷积^[21],在不损失信息和不增加网络参数的同时扩大了卷积核的感受野。扩张卷积的结构简单,如图 2 所示,通过在标准卷积所有行和列的相邻权值间插入 $r-1$ 个权重为 0 的值, r 为扩张因子,从而实现感受野的扩张。

扩张卷积感受野的计算方式为

$$F_i = F_{i-1} + (k_i - 1) \times r_i \times \prod_{n=1}^{i-1} s_n, \quad (3)$$

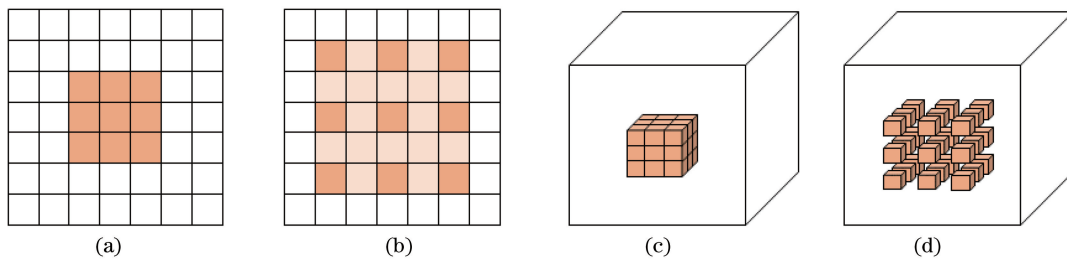


图 2 标准卷积与扩张卷积示意图。(a)二维标准卷积;(b)二维扩张卷积($r=2,2$);
(c)三维标准卷积;(d)三维扩张卷积($r=2,2,2$)

Fig. 2 Schematic of standard convolution and dilated convolution. (a) 2D standard convolution; (b) 2D dilated convolution ($r=2,2$); (c) 3D standard convolution; (d) 3D dilated convolution ($r=2,2,2$)

式中: F_i 为第 i 个卷积层中卷积核的感受野; k_i 表示卷积核的大小; r_i 为扩张因子; F_{i-1} 为第 $i-1$ 个卷积层中卷积核的感受野; s_n 为卷积步长。

结合扩张卷积和 Inception 网络^[22]中的宽度网络结构, 本文提出了多尺度特征融合结构, 包括多尺度空谱特征融合模块和多尺度空间特征融合模块, 如图 3 所示。对于 3D-CNN, 在网络的同一层分别设置三类卷积核: $3 \times 3 \times 3$ 、 $3 \times 3 \times 3 (r=2, 2, 2)$ 、连

续的 $3 \times 3 \times 3$ 和 $3 \times 3 \times 3 (r=2, 2, 2)$ 。其感受野分别为 $3 \times 3 \times 3$ 、 $5 \times 5 \times 5$ 、 $7 \times 7 \times 7$ 。不同的感受野意味着卷积核能够提取特征图中不同尺度的空谱特征, 将三类卷积核输出的特征图进行连接, 通过 $1 \times 1 \times 1$ 卷积实现空谱特征的融合。与 3D-CNN 类似, 在 2D-CNN 的同一层设置不同的二维卷积核: 3×3 、 $3 \times 3 (r=2, 2)$ 、连续的 3×3 和 $3 \times 3 (r=2, 2)$, 感受野分别为 3×3 、 5×5 、 7×7 。

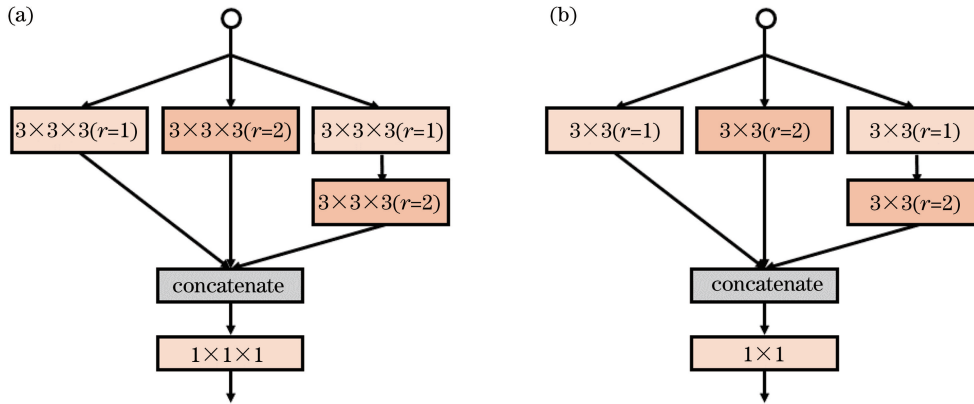


图 3 多尺度特征融合结构图。(a) 多尺度空谱特征融合模块; (b) 多尺度空间特征融合模块
Fig. 3 Multi-scale feature fusion structure. (a) Multi-scale spatial-spectral feature fusion module; (b) multi-scale spatial feature fusion module

2.3 注意力机制

本文提出的网络结构以高光谱图像的标注像元及其邻域组成的立方体数据块作为输入, 没有经过降维预处理。由于高光谱图像的光谱维度高, 波段间存在信息冗余, 在分类过程中会引入一部分噪声数据, 影响网络的分类性能; 而且像元的空间邻域大量重叠, 造成空间信息的冗余, 并增大模型的计算

量。受计算机视觉中注意力机制的启发, 本文构建了空谱注意力模块和空间注意力模块, 如图 4 所示, 注意力机制能使网络在训练的过程中关注重要的光谱和空间信息, 抑制噪声和冗余信息, 提高模型的性能。

在空谱注意力模块中, 对于三维输入 X , 首先经过沿着通道维的最大池化和平均池化得到特征压

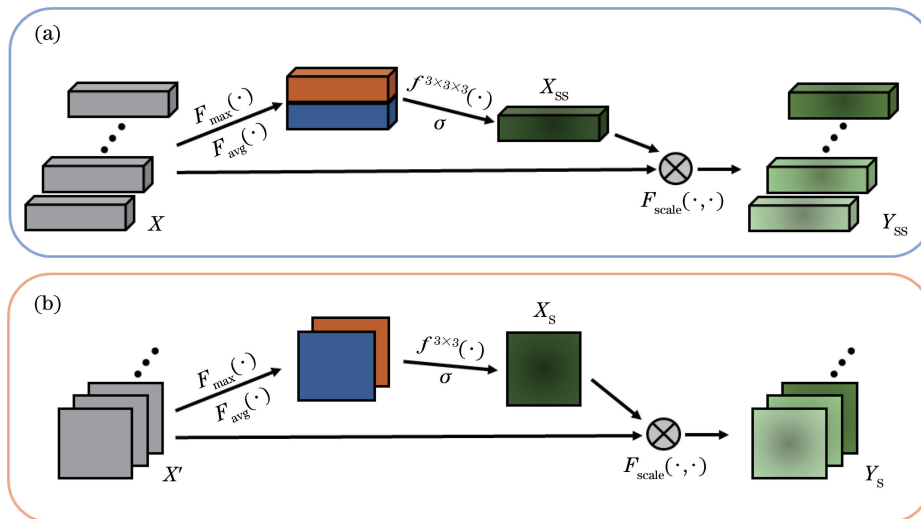


图 4 注意力模块结构图。(a) 空谱注意力模块; (b) 空间注意力模块
Fig. 4 Structure diagram of attention module. (a) Spatial-spectral attention module; (b) spatial attention module

缩后的两个全局特征描述图;然后,通过数量为 1 的 $3 \times 3 \times 3$ 卷积核对特征图进行融合,并通过 sigmoid 形式的门机制得到空间特征和光谱特征的权重 $X_{SS} \in (0, 1)$ 。通过网络在训练过程中不断学习,对于重要的空间和光谱信息,对应的权重接近于 1;对于不重要或者具有负面影响的噪声和冗余信息,对应的权重接近于 0。最后,将原始输入 X 与权重 X_{SS} 相乘,得到重标定后的特征图 Y_{SS} ,其过程可表示为

$$Y_{SS} = F_{scale} \{ \sigma [f^{3 \times 3 \times 3} (F_{max}, F_{avg})], X \}。 \quad (4)$$

空间注意力模块与空谱注意力模块结构相似,其原理可表示为

$$Y_S = F_{scale} \{ \sigma [f^{3 \times 3} (F_{max}, F_{avg})], X' \}, \quad (5)$$

式中: $F_{max}(\cdot)$ 和 $F_{avg}(\cdot)$ 分别为沿着通道维的最大池化和平均池化; $f^{3 \times 3 \times 3}(\cdot)$ 和 $f^{3 \times 3}(\cdot)$ 为普通的 $3 \times 3 \times 3$ 和 3×3 卷积; $F_{scale}(\cdot, \cdot, \cdot)$ 表示原始输入与权重的逐元素相乘; σ 为 hard-sigmoid 激活函数,相比 sigmoid 函数具有计算效率高的优势,其形式表示为

$$h - \text{sigmoid}(x) = \max \left[0, \min \left(1, \frac{2x + 5}{10} \right) \right]。 \quad (6)$$

2.4 本文网络结构

如图 5 所示,本文的网络结构主要由 3D-CNN、3D-CNN 与 2D-CNN 的转换以及 2D-CNN 这三个阶段组成。在 3D-CNN 阶段,以原始高光谱图像中标注像素及其邻域组成的尺寸为 $S \times S \times L$ 立方体数据块作为网络的输入,首先采用光谱步长为 2、卷积模式为 valid 的 $3 \times 3 \times 7$ 卷积核提取空间维和光谱维的局部特征;然后通过多尺度空谱特征融合模块和空谱注意力模块,实现不同尺度的空谱特征的融合、重要空间和光谱特征的关注以及噪声和冗余信息的抑制。在 3D-2D-CNN 转换阶段,采用卷积模式为 valid 的 $1 \times 1 \times C$ 卷积核将尺寸为 $M \times M \times C$ 的三维特征图转换为尺寸为 $M \times M$ 的二维特征图。在 2D-CNN 阶段,首先通过多尺度空间特征融合模块实现不同尺度的高级空间特征的融合,然后采用空间注意力模块使网络关注有效的空间特征,抑制无效的空间冗余信息,得到尺寸为 $N \times N$ 的特征图。采用平均池化对提取的高级抽象特征进行压缩和聚合,得到表示能力更强的特征;为了缓解过拟合问题,在全连接层前添加 dropout 层,最后通过 softmax 分类器得到分类结果。

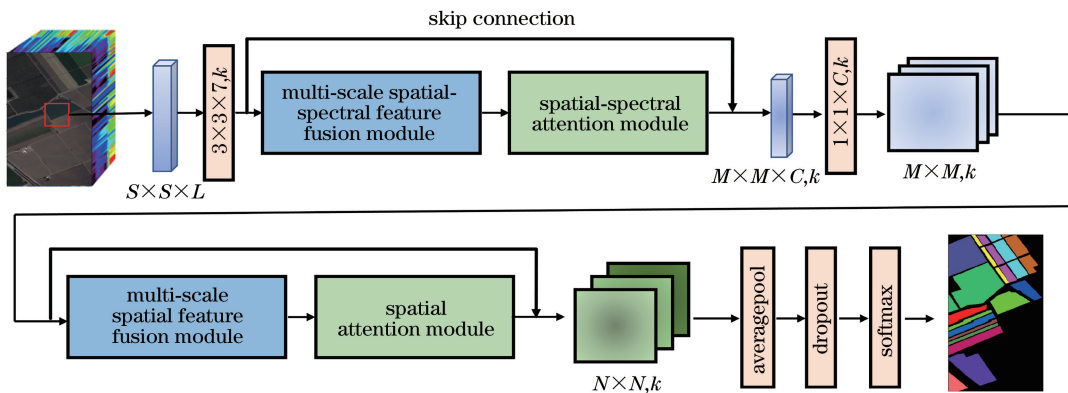


图 5 本文网络总体结构图

Fig. 5 Overall structure of proposed network

网络中所有卷积层具有相同的卷积核数目 k , 且在每个卷积层后都添加了批归一化层和 PReLU 激活函数,除了已说明的采用 valid 卷积模式的两个卷积层,其余的卷积层都采用 same 卷积模式。另外,在网络的 3D-CNN 和 2D-CNN 阶段添加了跳跃连接,跳跃连接的存在不仅改变了信息的传递方式,促进了网络的优化,而且能够实现浅层特征和深层特征的结合,提高特征信息的复用率。

3 实验数据与参数设置

本文选取 Pavia University (PU) 和 Salinas

(SA) 这两个常用于评价高光谱图像分类算法的基准数据集作为实验数据,使用总体精度(OA)、平均精度(AA)和 Kappa 系数作为分类精度评价指标。其中,OA 代表模型在测试集上预测正确的像元数量与所有像元数量的比值;AA 是指每一类预测正确的像元数量与每一类所有像元数量的比值的平均值;Kappa 系数表示模型的分类结果与完全随机的分类结果之间的差异,计算结果为 $-1 \sim 1$ 之间,通常介于 $0 \sim 1$ 之间,其值越大表示分类精度越高。每次实验均采用随机划分的样本对网络进行 20 次训练和测试,分类结果以平均值表示。本文所有实验

均在 Windows10 系统下使用开源的深度学习框架 Tensorflow 完成,硬件环境为 Intel Core i7-8700K、64 G 内存和 Nvidia GeForce GTX1080Ti 显卡。

3.1 高光谱数据集

PU 数据集是由机载反射光谱成像仪(ROSIS)于 2003 年在意大利的帕维亚城获取,空间分辨率为 1.3 m。该数据的尺寸为 610×340 ,剔除 12 个受噪声影响的波段后剩余 103 个可用光谱波段。图像共包含 2207400 pixel,已标记 9 种地物类别的像素只有 42776 个,其余的为背景像素。SA 数据集是由机载

可见红外成像光谱仪(AVIRIS)于 1998 年对美国加利福尼亚州的 Salinas 山谷拍摄的图像,空间分辨率为 3.7 m,剔除 20 个不能被水反射的波段后剩余 204 个有效波段。该图像的尺寸为 512×217 ,共包含 111104 个像素,其中 54129 个像素具有对应的类别标签,这些像素共分为 16 类。图 6 所示为两类数据集的伪彩图与标签图。为了测试本文所提模型在训练样本较少情况下的分类效果,对于 PU 和 SA 数据集中已标注类别的每类地物,分别选取 50 个样本作为训练集,50 个样本作为验证集,剩下的作为测试集。

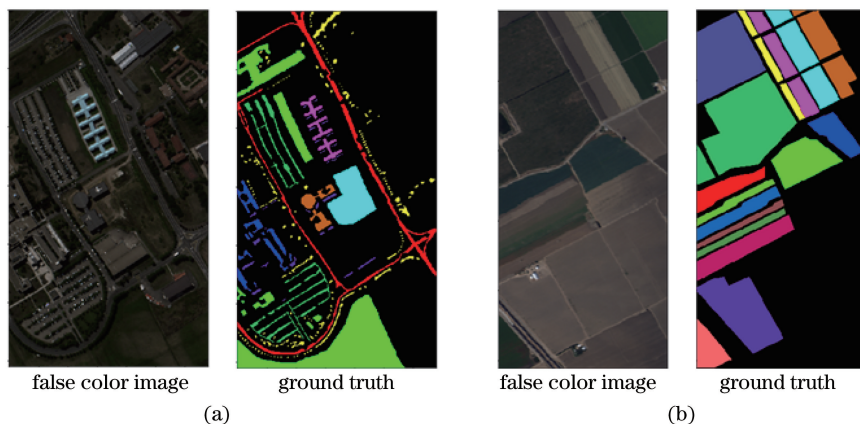


图 6 数据集伪彩图与地面真值图。(a) PU 数据集;(b) SA 数据集

Fig. 6 False color image and ground truth of data sets. (a) PU data set; (b) SA data set

3.2 参数设置

网络结构确定后,采用 He normal 对卷积层和全连接层的权值进行初始化,网络训练采用 Nadam 优化器优化网络参数,设置最大训练轮数为 100, batch size 为 32。学习率作为深度学习模型中重要的超参数,决定了损失函数能否收敛以及何时收敛到最小值。实验中采用取值为 0.0001、0.0003、0.0005、0.001、0.003 和 0.005 的学习率对网络进行测试,在 PU 和 SA 数据集上的测试结果表明,两类数据集对应的最优学习率均为 0.001。为了解决固定学习率在训练后期导致网络无法收敛的问题,在网络训练过程中加入了学习率衰减策略,当验证损失超过 5 轮训练不再降低时,触发学习率减半的

机制。

卷积核的数量决定了网络的特征表示能力,数量较少时模型的拟合能力不足,数量较多时模型拟合能力强,容易导致过拟合。为了便于参数的调整,本文提出的网络结构中所有卷积层卷积核数目相同。表 1 列出了在 PU 和 SA 数据集上,不同卷积核数目对应的总体精度、训练时间和测试时间。PU 和 SA 数据集上取得最佳分类精度对应的卷积核数目分别为 16 和 20,这是由于选取的训练样本较少,分类模型仅需要较少的参数即可完成拟合。当增加卷积核数目时,网络的参数量、训练以及测试时间也随之增加,而且使用过多的参数拟合小样本数据容易出现过拟合现象,导致分类精度降低。

表 1 不同卷积核数目的总体精度、训练及测试时间对比

Table 1 Comparison of overall accuracy, training time, and test time for different convolution kernel numbers

Kernel number	Pavia University			Salinas		
	OA /%	Training time /s	Test time /s	OA /%	Training time /s	Test time /s
16	97.78	118.03	29.71	96.48	243.42	66.38
20	96.87	130.56	34.19	97.28	324.71	76.93
24	97.17	149.93	38.34	96.26	373.57	88.24
28	96.25	185.74	43.35	95.88	383.83	98.84
32	97.05	165.58	46.76	96.12	416.47	107.36

网络中采用了 3 种方法避免训练时可能出现的过拟合问题:1)在全连接层前添加 Dropout 层,并设置神经元丢弃率为 0.5;2)在每个卷积层中使用 L2 正则化,将正则化参数 λ 设为 0.001;3)使用早停法,当在验证集上的损失超过 20 轮训练不再降低时,触发停止训练的机制,能够有效避免网络过度训练而导致的过拟合现象。

本文网络结构的输入为高光谱图像中的标注像元及其空间邻域内的所有像元组成的三维数据块,空间邻域的尺寸决定了网络所接受的信息量,从而影响最终的分类精度。为了选择合适的空间邻域尺

寸,选取了 5 种不同尺寸对网络进行训练和测试,结果如图 7 所示。从图中可以看出,当空间邻域尺寸依次增大时,分类精度也随之提高。这是由于较小的空间邻域包含的像元较少,而且会导致感受野不充分,使网络只能学习到小范围的局部信息,因此分类精度较低;而较大的空间邻域能够提供更多的空间和光谱信息,虽然会引入一部分噪声信息,但是通过网络中的注意力机制能够对其进行有效抑制,相比小尺寸的空间邻域取得了更好的分类效果。通过上述实验及分析,对于 PU 和 SA 数据集,均选择 19×19 作为输入立方体数据块的空间尺寸。

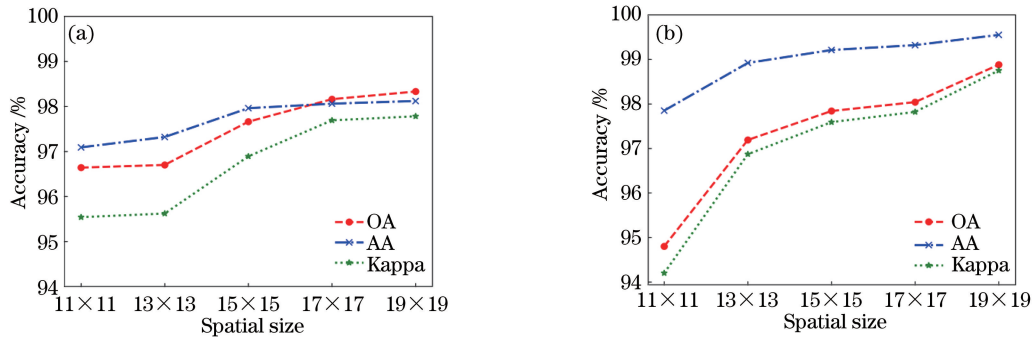


图 7 不同空间尺寸的分类精度对比。(a) PU 数据集;(b) SA 数据集

Fig. 7 Comparison of accuracy for different spatial sizes. (a) PU data set; (b) SA data set

4 实验结果与分析

4.1 消融实验

为了验证基于扩张卷积构造的多尺度特征提取融合模块以及注意力模块的有效性,在 PU 和 SA 数据集上进行了消融实验。实验中去除扩张卷积和注意力机制的 3D-2D-CNN 作为基准模型,为了

保证基准模型同样具有多尺度特征提取能力,保留了多尺度特征提取结构,并将其中的扩张卷积替换为两个连续的普通卷积。通过在基准模型中添加不同组合的扩张卷积和注意力机制,分别得到了两个对比模型和本文所提出的模型,表 2 中给出了不同模型结构的参数量和总体精度。

表 2 不同模型结构的参数量和总体精度对比

Table 2 Comparison of parameters and overall accuracy for different model architectures

Architecture	Pavia University		Salinas	
	Parameter	OA / %	Parameter	OA / %
3D-2D-CNN (baseline)	75596	98.11	139644	98.03
3D-2D-CNN+dilated convolution	56783	98.47	110402	98.42
3D-2D-CNN+attention	75643	98.51	139718	98.38
3D-2D-CNN+dilated convolution+attention	56857	98.73	110476	98.61

通过分析表 2 中的数据可以得出以下结论:1)基准模型中引入扩张卷积后,在 PU 和 SA 数据集上模型参数量分别减少了 24.89%和 20.94%,且精度分别提高了 0.36%和 0.39%;2)基准模型中添加注意力机制后,在模型参数几乎不变的情况下,在 PU 和 SA 数据集上的总体精度分别提高了 0.40%和 0.35%;3)基准模型中同时添加扩张卷积和注意

力机制,即所提出的模型在 PU 和 SA 数据集上参数量分别减少 24.79%和 20.89%的同时,总体精度分别提高了 0.62%和 0.58%。以上实验结果说明,在基准模型 3D-2D-CNN 中分别引入扩张卷积和注意力机制时,模型性能均能得到一定的提升,并且在同时添加扩张卷积和注意力机制的情况下能够获得最佳性能。

4.2 算法对比实验及分析

为了验证本文提出的扩张卷积注意力神经网络(3D-2D-ADCNN)的有效性,分别与基于 CNN 的 2D-CNN-MLP、3D-CNN-CRF、Hybrid-CNN 以及 Dilated-3D-CNN 模型进行对比实验。为保证实验的公平性,参与实验的对比算法按照最优参数进行配置,所有算法采用的训练集、验证集和测试集样本数量均相同。

表 3 和表 4 分别列出了不同算法在 PU 和 SA 数据集上的分类结果,训练样本数量均为 50。其中,2D-CNN-MLP、3D-CNN-CRF 和 Dilated-3D-CNN 这三类算法均只采用了 2D-CNN 或 3D-CNN 作为模型的基础网络,在训练样本较少的情况下分类精度接近。相比仅采用 2D-CNN 或 3D-CNN 的算法,

Hybrid-CNN 在两类数据集上的分类精度均有较大的提升,表明了当网络同时采用三维卷积和二维卷积时,能够更加充分地提取高光谱图像中的空谱特征。本文提出的 3D-2D-ADCNN 模型以串联的 3D-2D-CNN 作为网络主体,基于扩张卷积构建了多尺度特征融合结构,实现了空间特征和光谱特征的多尺度提取和融合,并通过引入注意力机制使网络通过学习自动关注重要空谱特征信息,提高了网络的表征能力。在 PU 数据集上,3D-2D-ADCNN 的 OA 相比 2D-CNN-MLP、3D-CNN-CRF、Dilated-3D-CNN 和 Hybrid-CNN 分别提高了 8.99%、10.83%、9.27% 和 3.86%;在 SA 数据集上,3D-2D-ADCNN 的 OA 相比 2D-CNN-MLP、3D-CNN-CRF、Dilated-3D-CNN 和 Hybrid-CNN 分别提高了 6.04%、7.23%、8.36% 和 2.13%。

表 3 PU 数据集上不同算法分类精度对比

Table 3 Comparison of classification accuracy for different algorithms on PU data set

unit: %

Category name	2D-CNN-MLP	3D-CNN-CRF	Hybrid-CNN	Dilated-3D-CNN	3D-2D-ADCNN
Asphalt	87.33	86.78	90.89	84.86	96.82
Meadows	90.46	86.51	96.31	87.67	99.47
Gravels	84.91	85.76	94.45	87.73	99.35
Trees	93.00	96.27	92.39	97.71	97.77
Painted-Metal-Sheets	99.72	99.90	99.78	99.75	99.92
Bare-Soil	88.46	85.18	99.06	85.94	100.00
Bitumen	94.67	93.23	99.24	89.33	99.84
Self-Blocking-Bricks	85.04	86.74	87.81	83.76	98.91
Shadows	98.94	99.11	96.87	99.82	98.57
OA	89.76	87.92	94.89	89.48	98.75
AA	91.39	91.05	95.20	89.89	98.96
Kappa	96.54	84.22	93.26	86.04	98.60

表 4 SA 数据集上不同算法分类精度对比

Table 4 Comparison of classification accuracy for different algorithms on SA data set

unit: %

Category name	2D-CNN-MLP	3D-CNN-CRF	Hybrid-CNN	Dilated-3D-CNN	3D-2D-ADCNN
Brocoli_green_weeds_1	99.98	99.39	99.99	99.97	100.00
Brocoli_green_weeds_2	99.48	98.92	99.96	99.70	100.00
Fallow	99.27	99.59	98.57	99.08	100.00
Fallow_rough_plow	99.30	99.16	99.47	99.47	99.61
Fallow_smooth	98.44	97.44	98.65	96.86	99.80
Stubble	99.76	99.94	99.79	99.76	100.00
Celery	99.61	99.60	99.84	99.60	100.00
Grapes_untrained	79.40	76.09	88.90	70.83	94.17
Soil_vinyard_develop	99.92	99.73	98.61	99.05	100.00
Corn_senesced_green_weeds	95.61	94.52	99.18	92.72	100.00
Lettuce_romaine_4wk	99.02	99.29	99.82	98.78	100.00
Lettuce_romaine_5wk	99.99	99.88	99.01	99.92	100.00
Lettuce_romaine_6wk	99.98	99.78	99.52	99.33	100.00
Lettuce_romaine_7wk	98.86	99.01	99.41	99.38	99.89
Vinyard_untrained	81.90	79.83	94.71	81.40	99.09
Vinyard_vertical_trellis	98.05	97.14	99.81	98.36	100.00
OA	92.57	91.38	96.48	90.25	98.61
AA	96.79	96.21	98.45	95.82	99.53
Kappa	91.73	90.40	96.09	89.17	98.45

图 8 和图 9 分别为不同算法在 PU 和 SA 数据集上的分类结果。从图中可以看出,相比其他对比算法,本文所提的 3D-2D-ADCNN 模型分类结果图中错分地物最少,整体较为平滑且只存在极少数噪声点,更加接近于地表真值图。另外,通过观察可以发现:在 PU 和 SA 数据集的分类图中,包含像素少的地物类别的分类效果普遍比包含像素多的地物

类别的分类效果好。这是因为实验中每个地物类别的训练样本数量固定为 50,当原始地物类别的像素较多时,训练集所占比例较小,网络不能充分学习到该地物的代表性特征,导致部分像素不能被正确分类。但是在由 3D-2D-ADCNN 模型得到的分类图中不存在上述现象,说明即使在训练样本较少的情况下,本文所提方法也能取得鲁棒的分类结果。

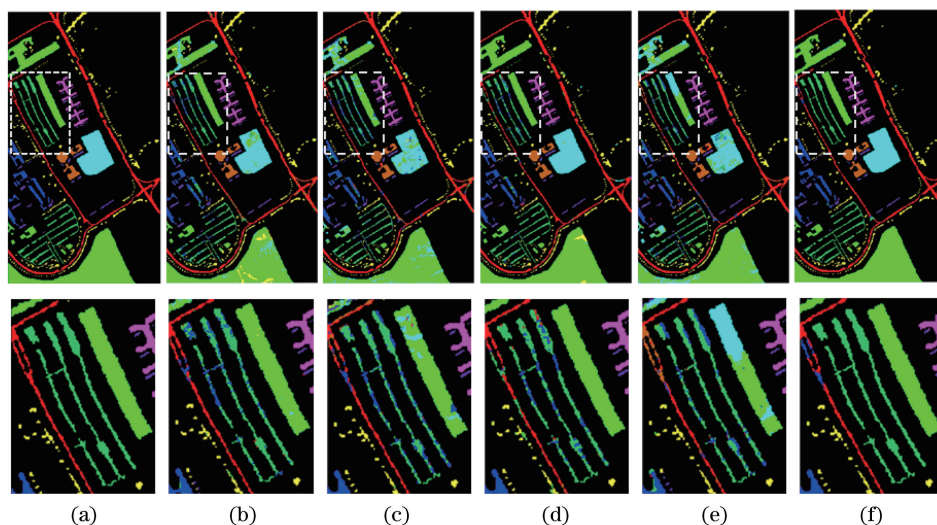


图 8 不同算法在 PU 数据集上的分类结果图及局部放大图。(a)地表真值图;

(b) 2D-CNN-MLP;(c) 3D-CNN-CRF;(d) Hybrid-CNN;(e) Dilated-3D-CNN;(f) 3D-2D-ADCNN

Fig. 8 Classification maps and partial enlarged maps with different algorithms on PU data set. (a) Ground truth image;

(b) 2D-CNN-MLP; (c) 3D-CNN-CRF; (d) Hybrid-CNN; (e) Dilated-3D-CNN; (f) 3D-2D-ADCNN

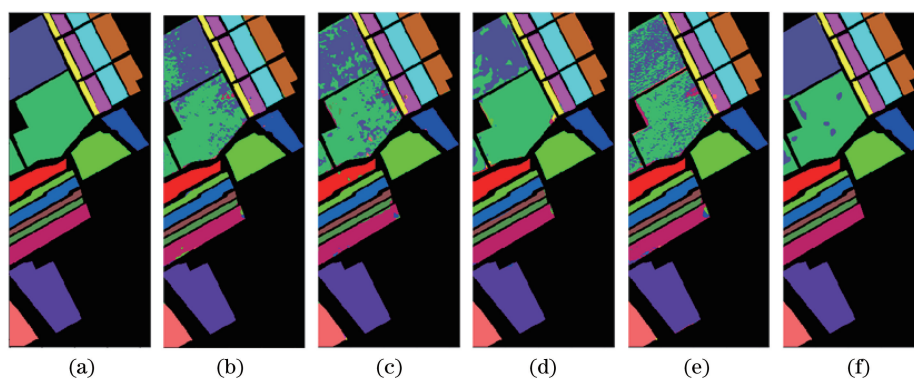


图 9 不同算法在 SA 数据集上的分类结果图。(a)地表真值图;(b) 2D-CNN-MLP;

(c) 3D-CNN-CRF;(d) Hybrid-CNN;(e) Dilated-3D-CNN;(f) 3D-2D-ADCNN

Fig. 9 Classification maps with different algorithms on SA data set. (a) Ground truth image; (b) 2D-CNN-MLP;

(c) 3D-CNN-CRF; (d) Hybrid-CNN; (e) Dilated-3D-CNN; (f) 3D-2D-ADCNN

表 5 和表 6 分别列出了不同算法在 PU 和 SA 数据集上的参数量、训练时间和测试时间。可以看出,本文所提出的 3D-2D-ADCNN 模型在使用扩张卷积代替普通卷积的情况下相比除 2D-CNN-MLP 外的其他模型具有较少的参数量,但是由于深度学习框架中三维扩张卷积的底层优化问题,导致最终

的训练和测试时间大于 3D-CNN-CRF 和 Hybrid-CNN。对于在网络中大量采用扩张卷积的 Dilated-3D-CNN,其训练时间和测试时间均大于其他模型。因此,在进一步的研究中,需要考虑从底层对三维扩张卷积的运算进行优化。

表 5 PU 数据集上不同算法的参数数量、训练及测试时间对比

Table 5 Comparison of parameters, training time, and test time for different algorithms on PU data set

Item	2D-CNN-MLP	3D-CNN-CRF	Hybrid-CNN	Dilated-3D-CNN	3D-2D-ADCNN
Parameter	33129	155017	4844793	413801	56857
Training time /s	8.91	103.30	63.75	671.89	122.22
Test time /s	2.17	13.04	7.85	72.36	51.93

表 6 SA 数据集上不同算法的参数数量、训练及测试时间对比

Table 6 Comparison of parameters, training time, and test time for different algorithms on SA data set

Item	2D-CNN-MLP	3D-CNN-CRF	Hybrid-CNN	Dilated-3D-CNN	3D-2D-ADCNN
Parameter	33556	192144	4845696	498912	110476
Training time /s	16.94	276.08	97.08	2122.97	327.33
Test time /s	2.70	27.69	9.70	170.50	122.61

为了进一步证明 3D-2D-ADCNN 在样本有限情况下的有效性,随机选取 PU 和 SA 数据集每类地物中的 10~50 个样本作为训练数据,且验证集和训练集样本数量相同。图 10 所示为不同算法在小样本条件下的总体分类精度,可以看出,即使在训练

样本极其有限的情况下,本文所提出的 3D-2D-ADCNN 模型仍然能够获得最高的分类精度。当每个类别的训练样本数量为 10 时,在 PU 和 SA 数据集上的总体精度均能达到 90%左右,证明了本文算法在小样本条件下的有效性和鲁棒性。

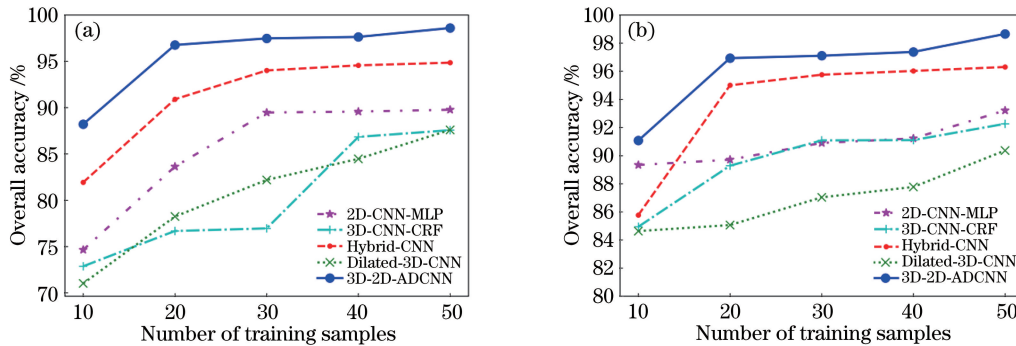


图 10 不同训练样本数目对应的总体精度。(a) PU 数据集;(b) SA 数据集

Fig. 10 Overall accuracy with different numbers of training samples. (a) PU data set; (b) SA data set

5 结 论

针对基于深度学习的高光谱图像分类算法在训练样本有限时分类精度低的问题,提出了结合三维卷积与二维卷积的 3D-2D-CNN 串联模型,在降低计算复杂度的同时提高了网络的特征提取能力。通过引入扩张卷积与注意力机制,分别构建了多尺度特征融合模块和注意力模块,实现了不同尺度下特征的融合以及重要空谱特征的关注,有效提高了模型的性能。在两类高光谱影像数据集上对本文算法和四种对比算法进行实验验证,结果表明,本文设计的网络模型在训练样本较少的条件下取得了精度最高的分类结果。

考虑到实际应用中样本标签的获取难度大且耗费时间多,未来的研究方向是结合弱监督学习与样本扩增技术,在使用更少的训练样本情况下获得更高的分类精度。

参 考 文 献

- [1] Adam E, Mutanga O, Rugege D. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review [J]. Wetlands Ecology and Management, 2010, 18(3): 281-296.
- [2] Liang L, Di L P, Zhang L P, et al. Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method [J]. Remote Sensing of Environment, 2015, 165:123-134.
- [3] Ershat Able, Baidengsha Maimaitiaili, Mamat Sawut, et al. Combined estimation of chlorophyll content in cotton canopy based on hyperspectral parameters and back propagation neural network[J]. Acta Optica Sinica, 2019, 39(9): 0930003. 依尔夏提·阿不来提, 白灯莎·买买提艾力, 买买提·沙吾提, 等. 基于高光谱和 BP 神经网络的棉花冠层叶绿素含量联合估算 [J]. 光学学报, 2019, 39

- (9): 0930003.
- [4] Yang X G, Yu Y. Estimating soil salinity under various moisture conditions: An experimental study [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(5): 2525-2533.
- [5] Govender M, Chetty K, Bulcock H. A review of hyperspectral remote sensing and its application in vegetation and water resource studies[J]. *Water SA*, 2009, 33(2): 145-151.
- [6] Yokoya N, Chan J, Segl K. Potential of resolution-enhanced hyperspectral data for mineral mapping using simulated EnMAP and sentinel-2 images [J]. *Remote Sensing*, 2016, 8(3): 172.
- [7] Du P J, Xia J S, Xue Z H, et al. Review of hyperspectral remote sensing image classification[J]. *Journal of Remote Sensing*, 2016, 20(2): 236-256. 杜培军, 夏俊士, 薛朝辉, 等. 高光谱遥感影像分类研究进展[J]. *遥感学报*, 2016, 20(2): 236-256.
- [8] Farrell M D, Mersereau R M. On the impact of PCA dimension reduction for hyperspectral detection of difficult targets [J]. *IEEE Geoscience and Remote Sensing Letters*, 2005, 2(2): 192-195.
- [9] Wang J, Chang C I. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2006, 44(6): 1586-1600.
- [10] Bandos T V, Bruzzone L, Camps-Valls G. Classification of hyperspectral images with regularized linear discriminant analysis [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2009, 47(3): 862-873.
- [11] Tarabalka Y, Fauvel M, Chanussot J, et al. SVM- and MRF-based method for accurate classification of hyperspectral images [J]. *IEEE Geoscience and Remote Sensing Letters*, 2010, 7(4): 736-740.
- [12] Ma L, Crawford M M, Tian J W. Local manifold learning-based k-nearest-neighbor for hyperspectral image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2010, 48(11): 4099-4109.
- [13] Wu F Y, Wang X, Ding J W, et al. Improved cascade forest deep learning model for hyperspectral imagery classification [J]. *Journal of Remote Sensing*, 2020, 24(4): 439-453. 武复宇, 王雪, 丁建伟, 等. 高光谱遥感影像多级联合森林深度网络分类算法[J]. *遥感学报*, 2020, 24(4): 439-453.
- [14] Makantasis K, Karantzalos K, Doulamis A, et al. Deep supervised learning for hyperspectral data classification through convolutional neural networks [C]//2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), July 26-31, 2015, Milan, Italy. New York: IEEE Press, 2015: 4959-4962.
- [15] Li Z Q, Zhu R F, Gao F, et al. Hyperspectral remote sensing image classification based on three-dimensional convolution neural network combined with conditional random field optimization[J]. *Acta Optica Sinica*, 2018, 38(8): 0828001. 李竺强, 朱瑞飞, 高放, 等. 三维卷积神经网络模型联合条件随机场优化的高光谱遥感影像分类[J]. *光学学报*, 2018, 38(8): 0828001.
- [16] 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*, 2020, 17(2): 277-281.
- [17] Yan M J, Su X Y. Hyperspectral image classification based on three-dimensional dilated convolutional residual neural network [J]. *Acta Optica Sinica*, 2020, 40(16): 1628002. 颜铭靖, 苏喜友. 基于三维空洞卷积残差神经网络的高光谱影像分类方法[J]. *光学学报*, 2020, 40(16): 1628002.
- [18] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [19] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [20] Ji S W, Xu W, Yang M, et al. 3D convolutional neural networks for human action recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1): 221-231.
- [21] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions [EB/OL]. (2016-04-30) [2020-07-06]. <https://arxiv.org/abs/1511.07122>.
- [22] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 1-9.