

多尺度特征融合残差网络的高光谱地物分类

邓子青¹, 王阳¹, 张兵¹, 丁召¹, 边历峰², 杨晨^{1*}

¹贵州大学大数据与信息工程学院半导体功率器件可靠性教育部工程研究中心, 贵州 贵阳 550025;

²中国科学院苏州纳米技术与纳米仿生研究所, 江苏 苏州 215123

摘要 为充分提取高光谱图像(HSI)的光谱空间信息特征,实现HSI的高精度地物分类,提出端到端的多尺度特征融合残差(MFFI)模块。该模块结合了3D多尺度卷积、特征融合以及残差连接3种手段,实现了HSI多尺度光谱空间特征的联合提取。因模块具有端到端特性,可通过堆叠多个MFFI模块得到具有提取深层特征能力的MFFI网络。该网络在Salinas、Indian Pines和University of Pavia 3个HSI数据集的平均总体准确率为99.73%,平均准确率为99.84%,平均卡帕系数为0.9971。结果表明:MFFI模块可以有效提取不同类型地物数据集的光谱空间特征,并取得良好的分类结果。

关键词 高光谱图像分类; 残差结构; 多尺度特征融合; 光谱空间特征提取; 卷积神经网络

中图分类号 P407.8

文献标志码 A

DOI: 10.3788/LOP202259.1810014

Hyperspectral Image Classification Based on Multi-Scale Feature Fusion Residual Network

Deng Ziqing¹, Wang Yang¹, Zhang Bing¹, Ding Zhao¹, Bian Lifeng², Yang Chen^{1*}

¹Engineering Research Center of Semiconductor Power Device Reliability, Ministry of Education, College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, Guizhou, China;

²Suzhou Institute of Nano-Tech and Nano-Bionics, Chinese Academy of Sciences, Suzhou 215123, Jiangsu, China

Abstract In order to fully extract the spectral-spatial features of hyperspectral image (HSI) and to achieve high-precision ground object classification of HSI, an end-to-end multi-scale feature fusion identity (MFFI) block is proposed. This block combines 3D multi-scale convolution, feature fusion and residual connection. Through this block, multi-scale spectral-spatial joint features of HSI can be extracted. Because of the end-to-end feature of the block, the final MFFI network can be obtained by stacking multiple MFFI blocks. The average overall accuracy of 99.73%, average accuracy of 99.84%, and Kappa coefficient of 0.9971 are obtained on three HSI datasets: Salinas, Indian Pines and University of Pavia. The results show that the proposed MFFI block can effectively extract the spectral-spatial features of different types of ground object datasets and achieve satisfactory classification results.

Key words hyperspectral image classification; residual structure; multi-scale feature fusion; spectral-spatial feature extraction; convolutional neural network

1 引言

高光谱图像(HSI)由成像光谱仪采集,每个像素含有上百个光谱带,大量光谱带信息为地物精确分类提供了可能。目前HSI被广泛应用于地物分类、矿物勘测、军事探测等领域,光谱高维性使其在这些领域表现优异,但如何结合高维度光谱数据实现HSI的精确分类是一个具有挑战性的课题。

HSI的分类研究主要以传统方法和深度学习方法为主。经典传统算法如支持向量机(SVM)^[1]、字典学习(K-SVD)^[2]、亚像素到超像素的决策融合(Sps-Dc)^[3]、多分支注意力(DBMAM)^[4],主要根据测试样本和训练样本之间的光谱相似性来对图像进行分类。经典传统算法在HSI分类中取得了长足的进展,但HSI存在同谱异物和异物同谱的现象,所以仅根据光谱相似性进一步提高分类精确度非常具有挑战性。

收稿日期: 2021-06-15; 修回日期: 2021-07-21; 录用日期: 2021-08-10

基金项目: 国家自然科学基金(62065003,61827823)、贵州大学人才引进项目(贵大人基合 2013)

通信作者: *eliot.c.yang@163.com

卷积神经网络 (Convolutional neural network, CNN) 由于具有图像特征自动提取以及参数自我学习更新等优点, 在 HSI 地物分类领域应用广泛。2017 年 Lee 等^[5]使用多尺度卷积去探索 HSI 的局部光谱空间联系以增强网络表达能力。同年, Zhong 等^[6]使用 3D 卷积先后提取 HSI 的空间和光谱信息, 同时利用光谱空间综合信息提升网络的分类精度。2018 年 Wang 等^[7]在 Zhong 等^[6]的基础上添加密集连接结构, 虽然进一步增加了网络模型参数, 但取得了更好的分类效果。2019 年 Mughees 等^[8]使用深度置信网络去自适应提取有效光谱特征信息, 提升了分类的精度。2020 年 Chen 等^[9]使用 1D 和 2D 卷积来代替 3D 卷积提取光谱和空间特征, 在保证精度的前提下加速训练过程。2021 年 Alotaibi 等^[10]和 Shi 等^[11]采用残差连接和带融合的思想加深网络的深度和融合带信息以实现优良分类精度。综上, 目前基于 CNN 实现 HSI 分类的常用策略是使用多尺度手段进行特征融合和分阶段提取光谱空间信息, 将这两种常规架构引入 CNN 作为 HSI 分类的手段是有效的。

不同于其他 HSI 分类网络分别考虑光谱和空间信息, 以及在单通道内使用多尺度手段, 本文将并行多尺度手段与光谱空间信息联合提取相结合, 构成一个端到端的多尺度特征融合残差 (MFFI) 模块, 从而实现 HSI 的多尺度光谱空间特征提取, 且由于该模块具有端到端特性, 可通过叠加该模块提升网络深度, 增强网络的深层特征提取能力, 从而提高分类精度。

2 基本原理

2.1 MFFI 网络

3D 卷积可提取空间和光谱等信息, 而多尺度能感受不同的细节特征。将多尺度与 3D 卷积思想相融合构成一个端到端的 MFFI 模块, 实现 HSI 空间光谱综合特征的提取。通过堆叠所提模块组成一个 MFFI 网络, 其网络结构如图 1 所示。

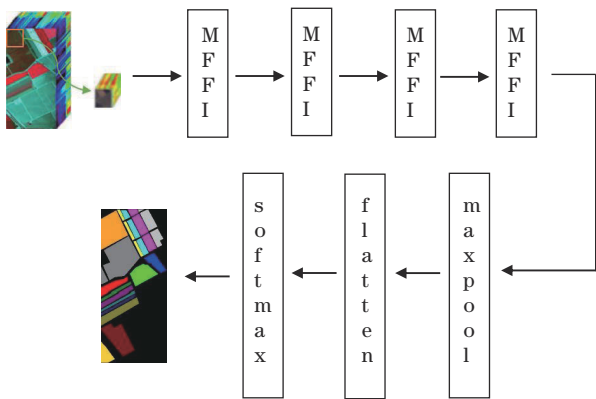


图 1 MFFI 网络结构图

Fig. 1 Structure of MFFI network

首先, 将待分类像素和以其为中心的周围像素组成的 HSI 像素块送入到 MFFI 模块中。该模块由降维、并行多尺度特征提取、特征融合以及残差连接 4 种手段组成, 主要对 HSI 数据进行光谱降维和多尺度的光谱空间特征提取, 实现冗余信息的消除和综合特征的提取。

之后, 将综合特征图送入最大池化层 (maxpool)、平坦层 (flatten) 转化成特征向量, 最后使用 softmax 分类器对特征向量进行分类, 其分类器可表示为

$$y^L = \frac{1}{\sum_{j=1}^m \exp(\mathbf{w}_j^T f^L)} \begin{bmatrix} \exp(\mathbf{w}_1^T f^L) \\ \exp(\mathbf{w}_2^T f^L) \\ \vdots \\ \exp(\mathbf{w}_m^T f^L) \end{bmatrix}, \quad (1)$$

式中: L 为数组 \mathbf{y} 中的第 L 个元素; j 为当前神经元层数; T 为转置操作符; \mathbf{w} 为权重; l 为神经元层数; f 为神经元值; m 为分类问题中的类别数; $\exp(\mathbf{w}_j^T f^L)$ 为输入的数组值; 输出 y^L 为每个样本的向量, 其中向量值表示样本对应类别的可能性, 将最大值对应分类作为样本最终分类标签。

该网络通过引入 MFFI 模块提取 HSI 光谱空间综合特征, 并在特征提取阶段就使用多尺度手段以整合不同感受野上的特征, 模块所提取的综合特征将更具可辨识度。

2.2 端到端的 MFFI 模块

在图 1 结构中, MFFI 模块是网络提取多尺度光谱空间融合特征的关键, 该模块主要由降维、多尺度特征提取和残差特征融合 3 个部分组成, 如图 2 所示。图 2 中, feature maps 代表图像特征图, 3D CONV 代表 3D 卷积, RELU 代表激活函数, ADD 代表逐像素加法, BN 代表批归一化操作, residual connection 代表残差连接。

在降维部分, 为避免多尺度带来的信息冗余, 使用 24 个尺寸为 (1, 1, 7)、步长为 (1, 1, 2) 的 3D 卷积进行输入图像的降维。该阶段采用的光谱维步长为 2, 故可将经此处理后的光谱维度降至输入的一半。采用下采样卷积的降维方式, 除可达到诸如主成分分析 (PCA) 等传统方法类似的降维效果外, 还能实现通道信息的整合。

在多尺度特征提取部分, 为实现同时提取多个尺度上的光谱空间信息, 将降维后图像输入并行多尺度 3D 卷积层。其中 3D 卷积的数量为 24, 尺寸分别为 (1, 1, 1)、(3, 3, 3)、(5, 5, 5), 均使用等尺寸 (same) 填充方式。卷积层输出得到 3 个尺度下的光谱空间综合特征。其中, (1, 1, 1) 卷积用于提取图像全局信息, 而 (3, 3, 3)、(5, 5, 5) 则能提供图像不同感受野下的局部信息, 三者互相补充。

在残差特征融合部分, 将 3 个尺度下的空间光谱综合特征图作逐元素相加融合, 并将降维后图像与融

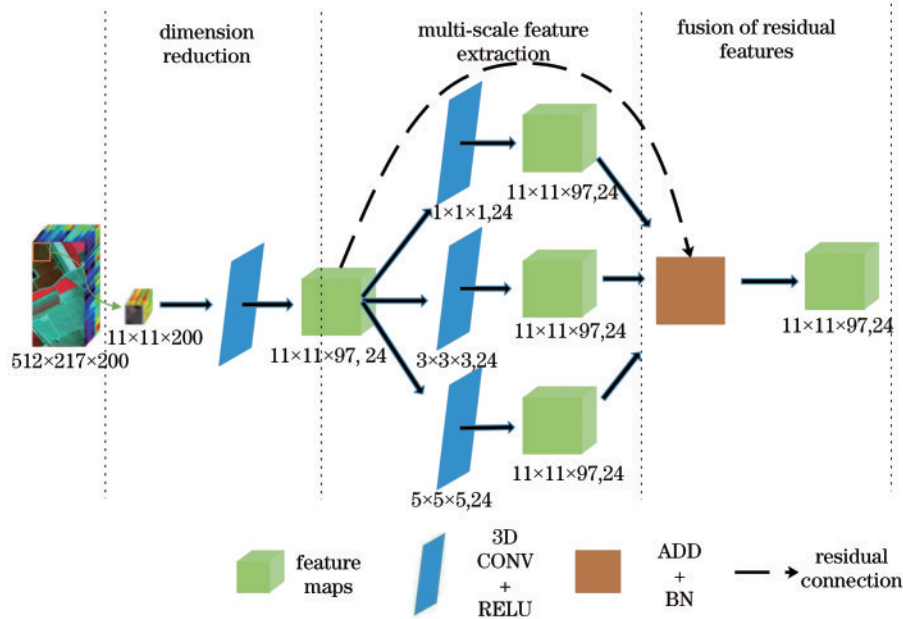


图2 MFFI模块
Fig. 2 MFFI block

合特征图作残差连接,得到最终的光谱空间融合特征图。

对比分别提取光谱、空间的特征和单一使用多尺度提取特征,MFFI模块利用3D卷积直接提取HSI光谱空间综合特征,并在提取过程中结合多尺度手段提取HSI不同尺度下的细节特征。此架构的优势是同时考虑空间与空间、光谱与光谱、空间与光谱的总体关系,避免分阶段提取光谱和空间的特征时先后特征的互相干扰,而在特征提取阶段使用多尺度手段则能保证网络感知原始图像的各细节特征。

2.3 多尺度3D卷积

图2中的MFFI模块采用了多尺度3D卷积结构以提取不同尺度上的光谱空间信息。其中,3D卷积,即卷积核(kernel),可在3个维度上进行移动,常被用于视频数据处理^[12]、三维对象识别^[13]和HSI分类^[14-16]。将其与多尺度结构相结合,有利于提取多尺寸的3D特征,其基本原理如图3所示。

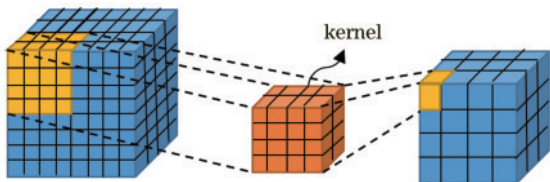


图3 3D卷积
Fig. 3 3D convolution

由图3可知,当输入一个3D数据时,将其与3D卷积核进行卷积得到3D输出特征图,可将其公式化^[11],表示为

$$F_j^l = \sum_{i \in M} F_i^{l-1} \cdot k_{ij}^l + b_j^l, i = 1, 2, \dots, p, j = 1, 2, \dots, q, \quad (2)$$

式中: i 为第*i*个神经元; M 为总神经元数; F_j^l 为第*l*层卷积层的输出特征图; F_i^{l-1} 为*l*-1层的输出特征图; k_{ij}^l 为卷积层的3D滤波器; b_j^l 为*j*层输出特征图的偏置量; p 为输入特征图的数量; q 为输出特征图的数量。

2.4 残差连接

在图2多尺度特征提取部分,引入了残差连接。残差网络(resnet)是2016年由He等^[17]提出的网络架构,该网络有效缓解了网络层数加深时梯度消失的问题,使网络能学习到更深层次的特征,其结构如图4所示。

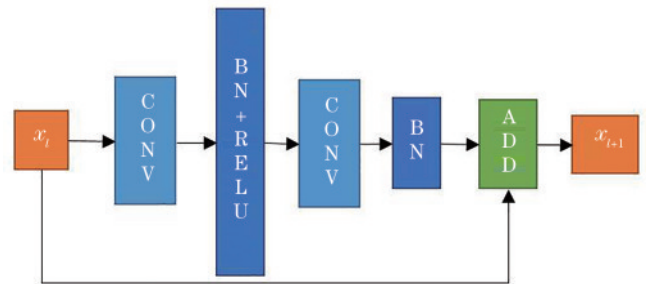


图4 基本残差结构
Fig. 4 Basic residual structure

图4中, x_l, x_{l+1} 表示*l*与*l*+1层的特征图。残差结构^[17]可以表示为

$$x_{l+1} = f(x_l) + x_l, \quad (3)$$

式中, f 为卷积操作。在残差连接中,输入图会在一系列卷积、归一化等操作后与其本身相加融合。残差连

接的优点是在原神经网络结构的基础上,添加跨层跳转连接,形成残差元(identity block),使网络具有包含大量浅层网络特征的可能性。在数学层面上,神经网络的反向传播是一个链式求导的过程,而残差连接会使反向传播从连乘变成连加,从而有效缓解梯度消失的问题。

假设从全连接层的 $l+1$ 层到 l 层反向传播,公式表示为

$$\frac{\partial j}{\partial z^l} = \frac{\partial j}{\partial z^{l+1}} \frac{\partial z^{l+1}}{\partial z^l} = \frac{\partial j}{\partial z^{l+1}} \sigma'(z^l)(w^l)^T, \quad (4)$$

式中: $\frac{\partial j}{\partial z^l}$ 为对 l 层的网络参数求偏导; z 为神经元; σ' 为求偏导。由于残差连接增加了一项 ∂z^l , 所以通过反向求导公式可以求得有残差连接时的公式, 表示为

$$\frac{\partial j}{\partial z^l} = \frac{\partial j}{\partial z^{l+1}} \frac{\partial z^{l+1}}{\partial z^l} = \frac{\partial j}{\partial z^{l+1}} \sigma'(z^l)(w^l + 1)^T. \quad (5)$$

在增加残差连接后, 权重值 w^l 额外增加 1, 而在网络加深到一定程度时, 权重值会变得非常小, 所以额外加 1 可以提升梯度从而避免梯度消失, 达到使网络模型更容易训练的效果。

3 分析与讨论

3.1 实验数据

实验使用 Salinas (SA)、Indian Pines (IN) 和 University of Pavia (UP) 3 种传统高光谱数据集作为测试对象。UP 数据集于 2001 年在帕维尔大学采集(使用光学成像光谱仪), 其中包含树 (Trees)、柏油路 (Asphalt)、草地 (Meadows) 等 9 种城市土地覆盖类型, 其空间分辨率为 $610 \text{ pixel} \times 340 \text{ pixel}$, 有效波段数为 103 个, 范围从 430 nm 到 860 nm。IN 数据集在印第安纳州西北部采集(使用红外成像光谱仪), 其中包括由谷物 (Corn)、燕麦 (Oats)、森林 (Woods) 等 16 种植被类别, 其空间分辨率为 $145 \text{ pixel} \times 145 \text{ pixel}$, 有效波段数为 200 个, 范围从 400 nm 到 2500 nm。SA 数据集利用美国加利福尼亚州萨利纳斯山谷上方的机载可见光/红外成像光谱仪 (AVIRIS) 采集, 由芹菜 (Celery)、耕地 (Fallow)、残茎 (Stubble) 等 16 种地物覆盖构成, 其空间分辨率为 $512 \text{ pixel} \times 217 \text{ pixel}$, 有效波段数为 204 个。实验测试硬件配置为 Intel(R) Core™ i5-3470 处理器和 Tesla K40c 12GB 显存显卡, 代码基于 TensorFlow 深度学习框架开发。其中 UP 和 SA 使用 10% 样本训练, IN 使用 20% 样本进行训练; 学习率设

置为 0.0003; 衡量分类精度的指标使用 OA、AA、K, 其中 OA 是对分类结果精度的总体评价, AA 是所有分类精度的平均值, K 反映分类结果与真实结果之间的吻合程度。

3.2 分类结果

为验证网络分类有效性, 在 3 种不同地物类型数据集上进行分类测试, 表 1 为在 IN、UP、SA 的网络分类结果。

表 1 3 种数据集分类结果
Table 1 Classification results of three datasets

Dataset	OA / %	AA / %	K
IN	99.60	99.55	0.9955
UP	99.85	99.80	0.9980
SA	99.95	99.89	0.9995

由表 1 可知, 所提网络架构可以有效提取 HSI 数据集的特征并加以分类, 而且对不同地物类型的 HSI 数据具有精准性, 所取得各项精度均能达到 99.5% 以上。

3.3 网络参数分析

为测试 MFFI 层数对网络分类精度的影响, 依次验证不同 MFFI 层数下网络在各数据集上的表现, 如表 2 所示。

表 2 叠加不同层模块数据集的平均精度
Table 2 Average accuracy of three datasets with different number of blocks

MFFI number	OA / %	AA / %	K
One MFFI	98.88	99.11	0.9869
Two MFFIs	98.42	96.85	0.9821
Three MFFIs	99.61	99.71	0.9955
Four MFFIs	99.64	99.64	0.9958
Five MFFIs	99.59	99.58	0.9959

由表 2 可知, 网络层数的增加可在一定程度上提升各项分类精确度。网络在经过 3 层 MFFI 模块后已经可以较好地提取数据集的深层特征, 进一步加深网络只会增加训练时间。综合考虑训练时间和分类精度, 本网络选择配置 4 层 MFFI 模块。

除网络深度外, 输入 HSI 的块大小也是影响网络分类的重要因素。输入块增大, 所包含的光谱空间信息会增加, 信息的冗余度也会上升。为平衡两者关系对块尺寸进行对照测试, 精度对比如表 3 所示 (N 为光谱通道数, $s \cdot \text{epoch}^{-1}$ 为每个训练迭代所需时间)。

表 3 不同输入块大小数据集的平均精度以及训练时间
Table 3 Average accuracy and average training time of three datasets with different input sizes

Input size	OA / %	AA / %	K	Training time / ($s \cdot \text{epoch}^{-1}$)
$7 \times 7 \times N$	99.64	99.64	0.9958	43
$9 \times 9 \times N$	99.75	99.74	0.9971	56
$11 \times 11 \times N$	99.77	99.75	0.9973	71
$13 \times 13 \times N$	99.78	99.69	0.9974	132

由表 3 可知,随着网络输入块的增大,网络的分类精度提高、训练时间延长。为达到较为平衡的网络性能,对不同块大小下网络的分类精度以及训练时间进行对比分析。可以看出,网络在输入块尺寸为 $11 \times 11 \times N$ 时,在精度与训练时间两者间有较好的折中。相较于 $9 \times 9 \times N$,其训练精度提高 0.02 个百分点,每 epoch 训练时间仅增加 15 s;而当尺寸增加到 $13 \times 13 \times N$ 时,在精度仅提高 0.01 个百分点情况下,却需多付

出 $61 \text{ s} \cdot \text{epoch}^{-1}$ 的训练代价。因此,选择 $11 \times 11 \times N$ 的输入块大小对网络性能进行测试。

3.4 方法比较

为验证所提网络的有效性,将所提 MFFI 网络与传统的 SVM^[1]及经典的深度学习网络模型——光谱空间残差网络(SSRN)^[4]、堆叠自动编码网络(SAE)^[18]、3D 卷积神经网络(3D CNN)^[19]——进行对比。在 IN、UP、SA 数据集获得的测试结果如表 4~表 6 所示。

表 4 IN 数据集上各网络精确度以及训练时间对比

Table 4 Classification result and training time of different networks on IN dataset

Results	SVM	SAE	3D CNN	SSRN	Proposed
OA / %	81.67	92.99	97.08	99.19	99.61
AA / %	79.84	89.76	95.09	98.93	99.55
K	0.7876	0.9218	0.9674	0.9907	0.9955
Training time / (s·epoch ⁻¹)	6	20	34	92.8	41

表 5 UP 数据集上各网络精确度以及训练时间对比

Table 5 Classification result and training time of different networks on UP dataset

Results	SVM	SAE	3D CNN	SSRN	Proposed
OA / %	90.08	86.10	98.85	99.79	99.82
AA / %	92.99	76.09	98.40	99.66	99.79
K	0.8721	0.8109	0.9847	0.9972	0.9976
Training time / (s·epoch ⁻¹)	4	25	38	61.2	50

表 6 SA 数据集上各网络精确度以及训练时间对比

Table 6 Classification result and training time of different networks on SA dataset

Results	SVM	SAE	3D CNN	SSRN	Proposed
OA / %	88.16	92.01	94.19	99.46	99.95
AA / %	93.29	96.56	95.65	99.77	99.89
K	0.8680	0.9110	0.9353	0.9940	0.9995
Training time / (s·epoch ⁻¹)	6	46	72	302	132

从表 4~表 6 可以看出,本文所提出的网络模型无论是在分类城市地物方面,还是分类自然地物方面,与其他算法相比均能达到较好的分类效果。其中传统 SVM 分类算法所需训练时间最短,但不能获得较高精度的分类结果。对于经典的 SAE 和 3D CNN 算法而言,虽然其运算时间相较 SVM 有一定增加,但在各数据集的分类精度均有明显提高。不同于 SAE 和 3D CNN 这两种通用图像分类网络,SSRN 针对 HSI 分类而提出,故网络取得了高精度

的分类结果,但所需训练时间大幅增加。对比 SSRN,本文所提网络精确度提高,训练时间也显著减少。

图 5~图 7 为在 IN、SA、UP 数据集中不同算法下分类的实际效果。可以看出所提网络分类图与其他网络相比分类更加精确。在 IN 分类图上各分类区域类别块更加平滑且噪点少;在 SA 分类图上各分类区域内部基本无错误分类空洞;在 UP 数据集上各分类区域内的噪点减少,基本与标签图一致。

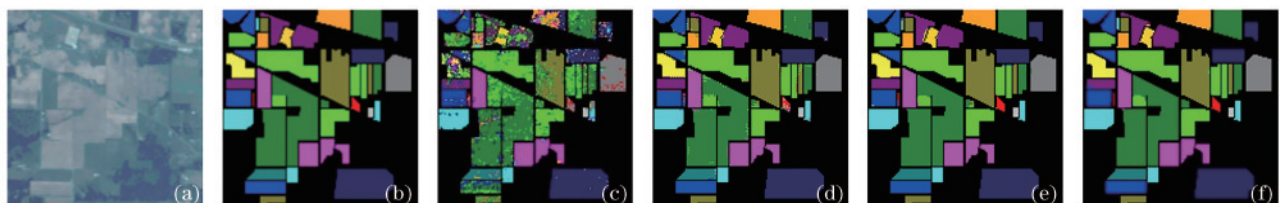


图 5 IN 数据集上的分类结果。(a) 样本影像;(b) 真实标签;(c) SVM 分类结果;(d) 3D CNN 分类结果;(e) SSRN 分类结果;(f) 所提网络分类结果

Fig. 5 Classification maps for IN dataset. (a) Picture of samples; (b) ground-truth label; (c) classification map of SVM; (d) classification map of 3D CNN; (e) classification map of SSRN; (f) classification map of proposed network

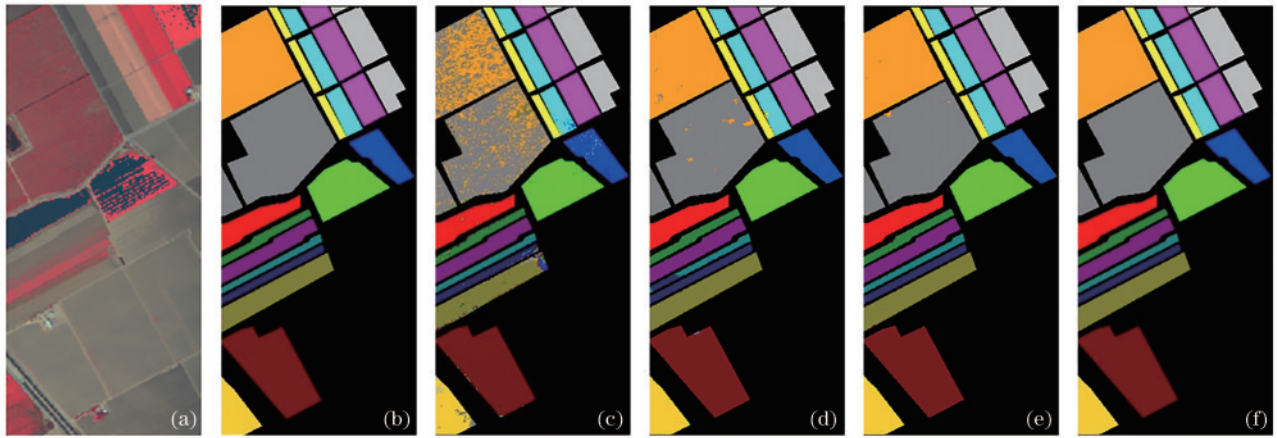


图 6 SA 数据集上的分类结果。(a) 样本影像; (b) 真实标签; (c) SVM 分类结果; (d) 3D CNN 分类结果; (e) SSRN 分类结果; (f) 所提网络分类结果

Fig. 6 Classification maps for SA dataset. (a) Picture of samples; (b) ground-truth label; (c) classification map of SVM; (d) classification map of 3D CNN; (e) classification map of SSRN; (f) classification map of proposed network

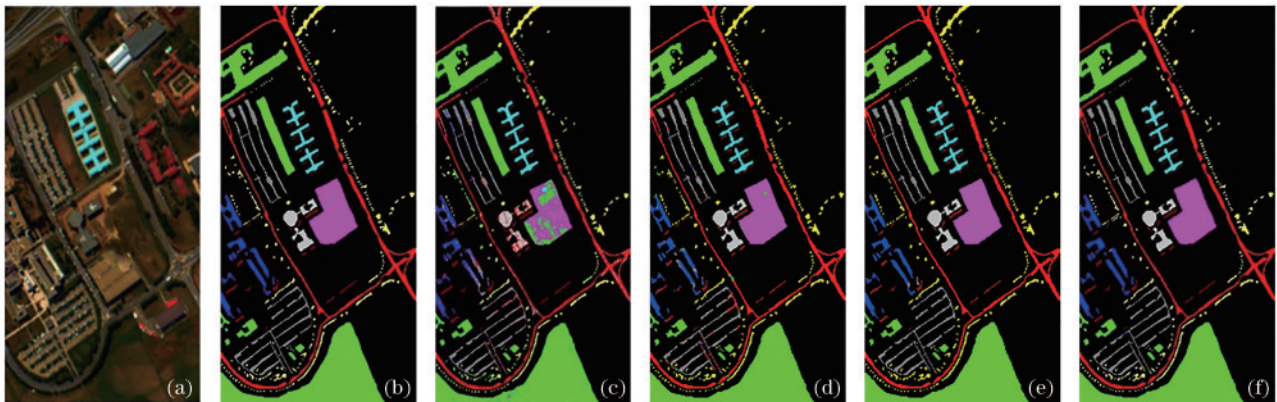


图 7 UP 数据集上的分类结果。(a) 样本影像; (b) 真实标签; (c) SVM 分类结果; (d) 3D CNN 分类结果; (e) SSRN 分类结果; (f) 所提网络分类结果

Fig. 7 Classification maps for UP dataset. (a) Picture of samples; (b) ground-truth label; (c) classification map of SVM; (d) classification map of 3D CNN; (e) classification map of SSRN; (f) classification map of proposed network

4 结 论

HSI 拥有丰富的空间和光谱信息,充分结合两种信息是提升网络分类精度的关键。为实现 HSI 光谱空间综合信息的充分利用,所提 MFFI 模块在特征提取阶段同时考虑光谱和空间的特征关系,还结合多尺度手段提升网络的感受野。结果表明,基于 MFFI 模块的网络架构对 HSI 数据集的综合特征把握更加准确。所提网络在 3 个不同地物类型数据集上各项平均准确率均能达到 99.5% 以上,能很好地提取不同地物类型数据集的特征,且具有较高的分类精度。综上,所提网络可有效提取 HSI 的多尺度光谱空间联合特征,且其端到端特性,使其可通过叠加模块加深网络深度,增强网络的深层特征提取能力,从而取得高精度的分类结果。后续研究考虑在网络中引入注意力机制以实现有效特征的高效提取,从而提升训练效率。

参 考 文 献

- [1] Waske B, van der Linden S, Benediktsson J A, et al. Sensitivity of support vector machines to random feature selection in classification of hyperspectral data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2010, 48(7): 2880-2889.
- [2] Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation[J]. IEEE Transactions on Signal Processing, 2006, 54(11): 4311-4322.
- [3] Lu T, Li S T, Fang L Y, et al. From subpixel to superpixel: a novel fusion framework for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(8): 4398-4411.
- [4] Ma W P, Yang Q F, Wu Y, et al. Double-branch multi-attention mechanism network for hyperspectral image classification[J]. Remote Sensing, 2019, 11(11): 1307.
- [5] Lee H, Kwon H. Going deeper with contextual CNN for hyperspectral image classification[J]. IEEE Transactions on Image Processing, 2017, 26(10): 4843-4855.

- [6] Zhong Z L, Li J, Luo Z M, et al. Spectral-spatial residual network for hyperspectral image classification: a 3-D deep learning framework[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(2): 847-858.
- [7] Wang W J, Dou S G, Jiang Z M, et al. A fast dense spectral-spatial convolution network framework for hyperspectral images classification[J]. *Remote Sensing*, 2018, 10(7): 1068.
- [8] Mughees A, Tao L M. Multiple deep-belief-network-based spectral-spatial classification of hyperspectral images[J]. *Tsinghua Science and Technology*, 2018, 24(2): 183-194.
- [9] Chen L L, Wei Z H, Xu Y. A lightweight spectral-spatial feature extraction and fusion network for hyperspectral image classification[J]. *Remote Sensing*, 2020, 12(9): 1395.
- [10] Alotaibi B, Alotaibi M. A hybrid deep ResNet and inception model for hyperspectral image classification[J]. *PFJ-Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 2020, 88(6): 463-476.
- [11] Shi C, Zhang J, You Z Z, et al. 3D convolutional neural networks with image fusion for hyperspectral image classification[M]//Liu Q, Liu X D, Shen T, et al. The 10th international conference on computer engineering and networks. *Advances in intelligent systems and computing*. Singapore: Springer, 2021, 1274: 419-428.
- [12] Jiang Y J, Liu X. Image/video aesthetic quality management based on fusing 3D CNN and shallow features[J]. *Future Generation Computer Systems*, 2021, 118: 118-123.
- [13] Anvarov F, Kim D H, Song B C. Action recognition using deep 3D CNNs with sequential feature aggregation and attention[J]. *Electronics*, 2020, 9(1): 147.
- [14] 毕晓君, 周泽宇. 基于双通道 GAN 的高光谱图像分类算法[J]. *光学学报*, 2019, 39(10): 1028002.
- Bi X J, Zhou Z Y. Hyperspectral image classification algorithm based on two-channel generative adversarial network[J]. *Acta Optica Sinica*, 2019, 39(10): 1028002.
- [15] 闫苗, 赵红东, 李宇海, 等. 基于卷积神经网络的高光谱遥感地物多分类识别[J]. *激光与光电子学进展*, 2019, 56(2): 021702.
- Yan M, Zhao H D, Li Y H, et al. Multi-classification and recognition of hyperspectral remote sensing objects based on convolutional neural network[J]. *Laser & Optoelectronics Progress*, 2019, 56(2): 021702.
- [16] 马永杰, 刘培培. 基于 DenseNet 进化的卷积神经网络图像分类算法[J]. *激光与光电子学进展*, 2020, 57(24): 241001.
- Ma Y J, Liu P P. Convolutional neural network based on DenseNet evolution for image classification algorithm[J]. *Laser & Optoelectronics Progress*, 2020, 57(24): 241001.
- [17] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [18] Chen Y S, Lin Z H, Zhao X, et al. Deep learning-based classification of hyperspectral data[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(6): 2094-2107.
- [19] Chen Y S, Jiang H L, Li C Y, et al. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(10): 6232-6251.