

## 融合多维度 CNN 的高光谱遥感图像分类算法

刘金香<sup>1</sup>, 班伟<sup>1</sup>, 陈宇<sup>1</sup>, 孙亚琴<sup>1</sup>, 庄会富<sup>1</sup>, 富尔江<sup>2</sup>, 张克非<sup>1,2,3\*</sup>

<sup>1</sup>中国矿业大学环境与测绘学院, 江苏 徐州 221116;

<sup>2</sup>北星空间信息技术研究院, 江苏 南京 210000;

<sup>3</sup>皇家墨尔本理工大学空间研究中心, 维多利亚 墨尔本, 澳大利亚 3001

**摘要** 随着计算机深度学习技术的不断发展,卷积神经网络(Convolutional Neural Network,CNN)作为代表算法,为高光谱图像分类提供了新的解决方案。但是,因为数据量迅速膨胀,复杂模型的分类精度和速度并不令人满意。因此提出了一种面向高光谱图像分类的轻量级融合 CNN 算法 3D-2D-1D CNN。该算法融合不同维度的 CNN,联合空间信息和光谱信息进行高光谱遥感图像分类。利用 Indian Pines、Pavia University、Salinas Scene 和 WHU-Hi-Han Chuan 数据集对该算法进行测试,结果显示,总体分类精度分别达到 99.65%、99.95%、100% 和 99.85%;与 3D CNN 算法和 3D-2D CNN 算法相比,训练时间和测试时间都明显降低。研究表明,所提算法能够在保证高分类精度的前提下有效地提高数据的分析速度。所提算法综合利用三种 CNN 对高光谱空-谱联合信息的抽象表达能力,有效促进了 CNN 在高光谱遥感图像分类领域中的应用。

**关键词** 遥感;高光谱图像分类;3D-2D-1D 卷积神经网络;深度学习

中图分类号 TP751

文献标志码 A

doi: 10.3788/CJL202148.1610003

### 1 引言

高光谱遥感技术利用高光谱传感器,以连续细分的波段对目标区域同时成像,得到含有空-谱信息的高光谱图像<sup>[1]</sup>。随着高光谱成像卫星的兴起,美国已初步建成高光谱遥感卫星体系,其他国家也加速研制和发射高光谱卫星,我国已发射了高分五号、珠海一号等卫星<sup>[2]</sup>。高光谱分辨率全色谱和高光空间分辨率载荷是高光谱的主要发展方向<sup>[3]</sup>,在土地监测<sup>[4-5]</sup>、环境保护<sup>[6]</sup>、防震减灾<sup>[7]</sup>等领域发挥着重要的作用。现在,针对高光谱图像的处理方法包括图像分类<sup>[8]</sup>、数据降维<sup>[9]</sup>、混合像元分解<sup>[10]</sup>和目标探测<sup>[11]</sup>等,其中图像分类是基础方法之一。然而,高光谱图像的高维特性、波段间的高度相关性和光谱混合等特征使得其分类面临巨大挑战<sup>[12]</sup>。随着深度学习新技术的出现,高光谱图像分类得到了突破性的进展<sup>[12]</sup>,典型方法就是卷积神经网络

(Convolutional Neural Network,CNN)。

CNN 已被广泛应用于视觉领域中,例如图像分类<sup>[13-15]</sup>、物体检测<sup>[16-17]</sup>、语义分割<sup>[18]</sup>、癌症分类<sup>[19]</sup>、深度估计<sup>[20]</sup>和面部防欺骗<sup>[21]</sup>等。常见的基于 CNN 的高光谱分类方法有三种,其中 1D CNN 提取光谱信息,2D CNN 提取空间信息<sup>[22]</sup>,3D CNN<sup>[23]</sup>通常由三个维度的卷积核组成,不仅能够顾及二维平面的纹理空间特征,还能提取不同光谱通道之间的特征。Li 等<sup>[24]</sup>利用 3D CNN 构建端到端的深度学习网络,在不进行任何预处理和后处理的情况下,该网络相较于 2D CNN 取得了更高的精度。Leng 等<sup>[25]</sup>通过 3D CNN 提取出局部区域的空-谱特征,提高了分类精度。虽然 3D CNN 在空-谱信息融合方面效果较好,但对比 1D CNN 和 2D CNN,网络计算成本和参数数目增加,这不利于大规模模型的训练<sup>[22]</sup>。

针对以上问题,学者们提出了多种解决方法。

收稿日期: 2020-12-23; 修回日期: 2021-01-18; 录用日期: 2021-02-25

基金项目: 国家自然科学基金青年项目(42004015,41804013)、江苏省自然科学基金青年基金项目(BK20190644, BK20180666)、徐州市重点研发计划(KC19111)、江苏省双创团队和双创人才项目

通信作者: \*profkzhang@gmail.com

Roy 等<sup>[26]</sup>提出了 3D-2D CNN 的融合算法,将多波段的光谱图像按光谱顺序排列成类似于扑克牌的数据立方体,先利用 3D CNN 在三维空间提取空-谱联合信息,再利用 2D CNN 进一步学习每张光谱图像中更多抽象的空间信息。与 3D CNN 相比,该方法在保持分类精度的同时大幅度提高了计算效率。Zhang 等<sup>[27]</sup>提出的基于注意力机制的新型 3D-2D CNN 和 Han 等<sup>[28]</sup>设计的双流 CNN 也是对空间特征进行进一步细化。Zhang 等<sup>[29]</sup>提出的 3D-1D CNN 同样先通过 3D CNN 获得空-谱高级语义信息,然后从数据立方体中任一组子像元中提取数百个光谱读数以构成一条光谱曲线,将其转换为 1D CNN 的输入并进行卷积运算,以学习更加抽象的光谱信息,解决了 CNN 在林木分类中性能较差的难题。陈亨等<sup>[22]</sup>提出了分解式 3D CNN,通过将一次三维卷积运算拆分成一次二维卷积和二次一维卷积运算来降低成本,但是没有保持 3D CNN 融合空-谱信息的特点,而且需要较高的迭代次数才能达到目标精度。总之,2D CNN 不能提取良好的光谱区分特征,3D CNN 的计算效率低,在具有相似纹理特征的高光谱分类上表现较差<sup>[26]</sup>。改进的 3D-2D CNN 没有充分考虑光谱特征优势,改进的 3D-1D CNN 不能有效兼顾空间特点。而在实际环境中,高光谱数据通常包含多种地物类别,一些类别具有明显的

空间特征差异,一些类别在光谱特征中更容易被区分,这是我们提出混合 CNN 模型的动机。

本文提出了一种融合 3D-2D-1D CNN 的新运算方法,将不同维度的 CNN 进行有机的融合,以减少 3D CNN 运算的计算量,并且充分利用高光谱的空-谱联合信息,提高了模型的适用性和计算效率。

## 2 基本原理

本文提出的 3D-2D-1D CNN 模型如图 1 所示,主要思想是在 3D CNN 之后使用 2D CNN 进行空间信息的提取,然后将输出的信息进行 1D CNN 运算,进一步提取高维光谱信息,最后进行分类处理。

3D CNN 是在高光谱立方体中,运用三维卷积核进行卷积计算,卷积层的每一个特征图都会与多个光谱图相连,因此能够实现空-谱联合信息的同时提取<sup>[30]</sup>。分解式 3D CNN 如图 2(a)所示,分别进行二维和一维卷积运算来模拟三维卷积过程,空间和光谱信息彼此独立<sup>[22]</sup>。本文所提算法利用 3D CNN 提取空-谱联合信息,并将之后的部分三维卷积层过程替换为计算量少的二维卷积和一维卷积过程,根据卷积运算的规则,该算法能够提高运算效率,如图 2(b)所示。下面将对本方法的处理过程予以介绍。

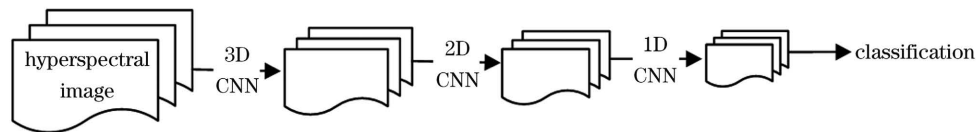


图 1 3D-2D-1D CNN 模型的卷积计算过程

Fig. 1 Convolution calculation process of 3D-2D-1D CNN model

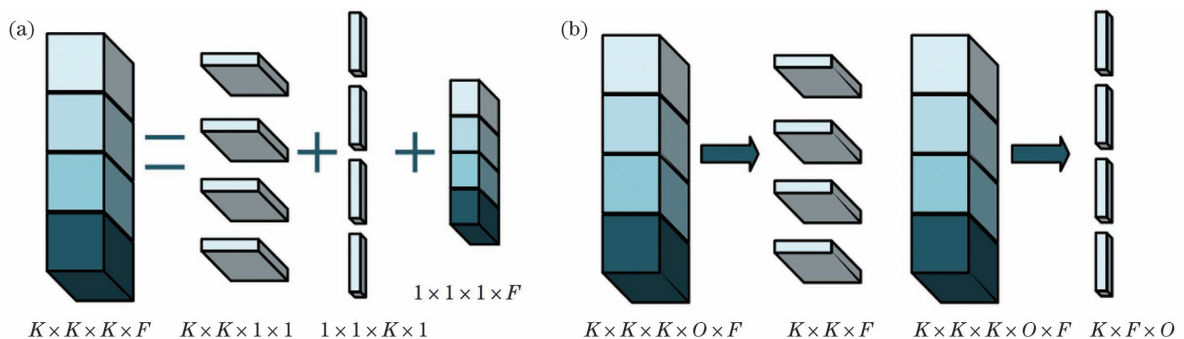


图 2 分解式 3D CNN 和 3D-2D-1D CNN 的处理过程简图。(a)分解式 3D CNN;(b)3D-2D-1D CNN

Fig. 2 Process diagrams of decomposed 3D CNN and 3D-2D-1D CNN. (a) Decomposed 3D CNN; (b) 3D-2D-1D CNN

假设参与本次运算的卷积核的个数为  $F$ ,单个卷积核的大小为  $K \times K \times K$ ,因为每一层不会仅仅只有一个卷积运算,所以我们假设该层采用了  $O$  个卷积核来进行运算。因此,卷积核的运算量为  $K \times$

$K \times K \times O$ 。假设卷积运算的步长为 1,并且采用填充的策略保证输入输出特征图的高  $H$  和宽  $W$  相等,输入图片的大小为  $H \times W \times F \times D$ ,输出特征图的大小为  $H \times W \times D \times O$ ,其中  $D$  是光谱图像的通

道数。如果我们认为每个卷积核的计算成本为 1 个单位时间,那么对于常规的三维卷积,一次卷积操作的计算量  $C_1$  可以认为是

$$C_1 = H \times W \times D \times K \times K \times K \times O \times F. \quad (1)$$

对于模型而言,假设最优的卷积操作的层数为  $n$ ,那么模型的总计算量  $C_2$  为

$$C_2 = nC_1 = H \times W \times D \times K \times K \times K \times O \times F. \quad (2)$$

对于本文所提的 3D-2D-1D CNN 模型,将  $n$  层三维卷积运算改为  $n_1$  层三维卷积运算、 $n_2$  层二维卷积运算和  $n_3$  层一维卷积运算 ( $n = n_1 + n_2 + n_3$ ,  $n_1, n_2, n_3 \geq 1$ ),此时模型的总计算量  $C_3$  为

$$C_3 = n_1 C_1 + n_2 C_{2D} + n_3 C_{1D}, \quad (3)$$

式中: $C_{2D}$  是每层二维卷积运算的计算量,  $C_{2D} = H \times W \times F \times D \times K \times K$ ;  $C_{1D}$  是每层一维卷积运算的计算量,  $C_{1D} = H \times W \times F \times D \times K \times O$ 。那么,改进卷积与原三维卷积的计算量之比  $R$  为

$$R = \frac{C_2}{C_3} = \frac{n \times K \times O}{n_1 \times K \times O + n_2 \times K + n_3 \times O} = \frac{(n_1 + n_2 + n_3) \times K \times O}{n_1 \times K \times O + n_2 \times K + n_3 \times O} = 1 + \frac{n_2 K (O - 1) + n_3 O (K - 1)}{n_1 \times K \times O + n_2 \times K + n_3 \times O}, \quad (4)$$

式中: $O, K > 1$ ,所以  $R > 1$ 。例如:对于模型成立的最简单的三层网络,  $n_1, n_2, n_3 = 1, n = 3$ ,任意取合理数值  $K = 5, O = 10$ ,那么  $R = 2.3$ 。由此可见,将卷积网络进行融合计算可以有效地减少计算量。

### 3 分析与讨论

本实验主要是在网络结构和超参数设置相同的情况下,比较支持向量机(SVM)、3D CNN、3D-2D CNN 和 3D-2D-1D CNN 的测试精度、训练时间、测试时间等,分析 3D-2D-1D CNN 在高光谱图像分类处理方面的优势。为了对所提 3D-2D-1D CNN 模型的性能进行充分的验证,实验还对比了所提模型的中间输出值 3D-2D CNN(新)特征图分类效果。

#### 3.1 实验数据

本文采用 4 个开源的高光谱数据集 Indian Pines、Pavia University、Salinas Scene 和 WHU-Hi-Han Chuan 来测试和验证数据。前三个数据集在高光谱分类算法的验证方面被广泛使用(代码 [https://](https://github.com/eecn/Hyperspectral-Classification)

[github.com/eecn/Hyperspectral-Classification](https://github.com/eecn/Hyperspectral-Classification), 数据 [http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes)), WHU-Hi-Han Chuan 数据集是武汉大学采集的高空间、高光谱分辨率数据,可作为分类研究的基准数据集。

Indian Pines 数据集由 AVIRIS 传感器收集,图像对应的地面范围为  $3.2 \text{ km} \times 3.2 \text{ km}$ ,图像大小为  $145 \text{ pixel} \times 145 \text{ pixel}$ ,波长范围为  $0.4 \sim 2.5 \mu\text{m}$ ,空间分辨率为  $20 \text{ m}$ ,实验中使用 200 个波段的数据。拍摄地点为美国印第安纳州西北部的 Indian Pines 测试区域,地表覆盖类型混合了林地、道路、农田、房屋建筑等,如表 1 及图 3(a)所示。该数据集具有标记样本分布不均衡且部分类别样本较少的特点。

Pavia University 数据集由 ROSIS 卫星采集,图像大小为  $610 \text{ pixel} \times 610 \text{ pixel}$ ,实验中采用了 103 个波段的数据。场景包括草地、树木、沥青等 9 个类别,具体的地物种类及数目如表 1 及图 3(b)所示。该数据集与 Indian Pines 数据集相比,空间分辨率不同,地面场景不同,各类样本数量较多,成像质量较好。

Salinas Scene 数据集由 AVIRIS 传感器收集,拍摄地点在美国加利福尼亚州,图像大小为  $512 \text{ pixel} \times 217 \text{ pixel}$ ,波长范围为  $0.4 \sim 2.5 \mu\text{m}$ ,空间分辨率为  $3.7 \text{ m}$ 。与 Indian Pines 一样,采用 200 个波段进行分类。数据集包括蔬菜、休耕田和葡萄园等 16 个常见的农业区种类,如表 1 及图 3(c)所示。该数据集含有较多样本,且样本之间具有较明显的差异性,常被用来测试高光谱分类模型。以上三个数据集联合应用,能更好地验证所提模型。

WHU-Hi-Han Chuan 数据集由 Nano-Hyperspec 成像传感器采集,拍摄地点在湖北省汉川市,图像尺寸为  $1217 \text{ pixel} \times 303 \text{ pixel}$ ,空间分辨率为  $0.109 \text{ m}$ ,波长范围为  $0.4 \sim 1000 \mu\text{m}$ ,实验采用了 274 个波段的数据。研究区域是一个城乡的边缘地带,有建筑物、水和耕地,其中有 7 种作物(草莓,豇豆,大豆,高粱,菠菜,西瓜和蔬菜),如表 1 及图 3(d)所示,详细介绍见 [http://rsidea.whu.edu.cn/resource\\_WHUHi\\_sharing.htm](http://rsidea.whu.edu.cn/resource_WHUHi_sharing.htm)。

#### 3.2 训练过程

本实验的参数采用以往文献中常用的较优参数<sup>[27,31-33]</sup>,在 Pavia University、Salinas Scene 和 WHU-Hi-Han Chuan 数据集中,数据预处理使用

表 1 Indian Pines、Pavia University、Salinas Scene 和 WHU-Hi-Han Chuan 数据集

Table 1 Datasets of Indian Pines, Pavia University, Salinas Scene, and WHU-Hi-Han Chuan

Indian Pines		Salinas Scene		WHU-Hi-Han Chuan		Pavia University	
Class	Number of samples	Class	Number of samples	Class	Number of samples	Class	Number of samples
Alfalfa	46	Brocoli green weeds 1	2009	Strawberry	44735	Asphalt	6631
Corn notill	1428	Brocoli green weeds 2	3726	Cowpea	22753	Meadows	18649
Corn mintill	830	Fallow	1976	Soybean	10287	Gravel	2099
Corn	237	Fallow rough plow	1394	Sorghum	5353	Trees	3064
Grass pasture	483	Fallow smooth	2678	Water spinach	1200	Shadows	947
Grass trees	730	Stubble	3959	Watermelon	4533	Bare soil	5029
Grass pasture mowed	28	Celery	3579	Greens	5903	Self blocking bricks	3682
Hay windrowed	478	Grapes untrained	11271	Trees	17978	Bitumen	1330
Oats	20	Soil vinyard develop	6203	Grass	9469	Painted metal sheets	1345
Soybean notill	972	Vinyard untrained	7268	Red roof	10516		
Soybean mintill	2455	Lettuce romaine 4wk	1068	Gray roof	16911		
Soybean clean	593	Lettuce romaine 5wk	1927	Plastic	3679		
Wheat	205	Lettuce romaine 6wk	916	Bare soil	9116		
Woods	1265	Lettuce romaine 7wk	1070	Road	18560		
Buildings grass trees drives	386	Corn-senesced green weeds	3278	Bright object	1136		
Stone steel towers	93	Vinyard vertical trellis	1807	Water	75401		

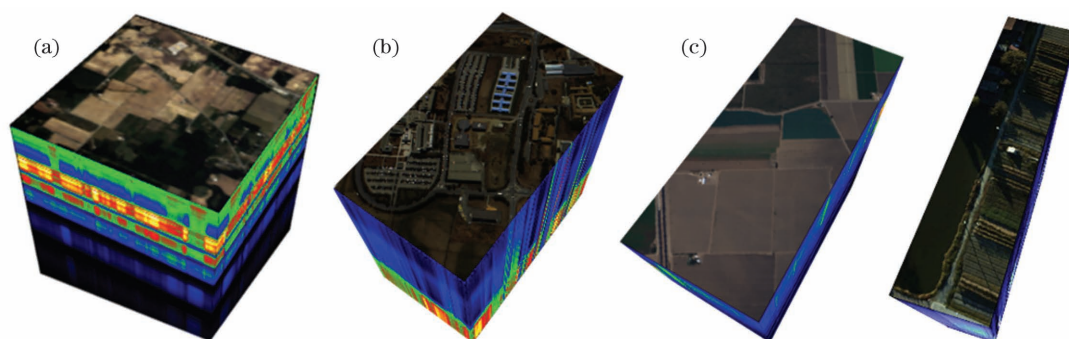


图 3 实验采用的高光谱数据。(a)Indian Pines;(b)Pavia University;(c)Salinas Scene;(d)WHU-Hi-Han Chuan

Fig. 3 Hyperspectral data used in experiment. (a) Indian Pines; (b) Pavia University; (c) Salinas Scene; (d) WHU-Hi-Han Chuan

主成分分析(PCA)进行降维处理,主成分的个数为 15,在卷积分析过程中窗口尺寸为  $25 \text{ pixel} \times 25 \text{ pixel}$ ,训练样本与测试样本的数量比例为 3 : 7,每批次训练数量设置为 20。而在 Indian Pines 数据集中,数据预处理同样使用 PCA 进行降维处理,主成分个数为 30,其余参数与 Pavia University 等数据集的参数设置相同。所有实验均在 Ubuntu 18 操作系统、NVIDIA GeForce RTX 2080Ti 显卡环境下运行。根据分类结果,实验选择 0.001 的最佳学习率。

以 Indian Pines 数据集为例,所建模型的详细

信息如表 2 所示,包括层类型、输出图尺寸和参数数量。所提模型包括 2 个三维卷积层、1 个二维卷积层和 1 个一维卷积层,综合利用了三种维度 CNN 的自动学习不同特征的能力。首先使用 2 个三维卷积层,卷积核的尺寸分别为  $8 \times 3 \times 3 \times 7 \times 1$  和  $16 \times 3 \times 3 \times 5 \times 8$ ,其中  $16 \times 3 \times 3 \times 5 \times 8$  表示 8 个三维输入特征图的 16 个三维卷积核的尺寸为  $3 \times 3 \times 5$ , $3 \times 3$  是空间尺寸,5 为光谱尺寸;然后采用一个尺寸为  $32 \times 3 \times 3 \times 320$  的二维卷积层,其中 32 是二维卷积核的数量, $3 \times 3$  表示卷积核的空间尺寸,320 是输入特征图的数量;最后使用的一维卷积层尺寸为

64×3×608, 64 是卷积核的数量, 3 表示卷积核的光谱尺寸, 608 是输入特征光谱的数量。在第一层中存在数量较多的参数, 最后一层的节点数为 16, 这与 Indian Pines 数据集的类型数相同, 该数据集可训练参数的总数为 529024。所有的权值都是随机初始化和训练的, 使用 Adam 优化器, 并在不进行批量归一化和数据扩充处理的情况下进行网络训练。

表 2 Indian Pines 数据集的卷积训练模型

Table 2 Convolution training model of Indian Pines dataset

Layer (type)	Output shape	Number of parameters
Input_1 (Input layer) (None, 25, 25, 30, 1)		0
Conv3d (Conv3D) (None, 23, 23, 24, 8)		512
Conv3d_1 (Conv3D) (None, 21, 21, 20, 16)		5776
Reshape (Reshape) (None, 21, 21, 320)		0
Conv2d (Conv2D) (None, 19, 19, 32)		92192
Reshape_1 (Reshape) (None, 19, 608)		0
Conv1d (Conv1D) (None, 17, 64)		116800
Flatten (Flatten) (None, 1088)		0
Dense (Dense) (None, 256)		278784
Dropout (Dropout) (None, 256)		0
Dense_1 (Dense) (None, 128)		32896
Dropout_1 (Dropout) (None, 128)		0
Dense_2 (Dense) (None, 16)		2064
Total number of parameters: 5361913		

### 3.3 实验结果对比分析

本实验采用总体分类精度 (Overall Accuracy, OA)、平均分类精度 (Average Accuracy, AA) 和 KAPPA 系数三种指标来衡量分类的精度。OA 代表测试集总数中预测正确的比例, AA 代表分类精

表 3 各个模型在 Indian Pines、Pavia University、Salinas Scene 和 WHU-Hi-Han Chuan 数据集上的分类精度

Table 3 Classification accuracies of each model in Indian Pines, Pavia University, Salinas Scene and WHU-Hi-Han

Chuan datasets

unit: %

Dataset	Accuracy of classification	SVM	2D CNN	3D CNN	3D-2D CNN	3D-2D CNN (new)	3D-2D-1D CNN
Indian Pines	OA	69.675	89.560	96.962	99.331	94.328	99.652
	AA	51.652	94.443	97.638	98.141	93.483	98.974
	KAPPA	64.498	84.364	96.526	99.238	93.512	99.603
Pavia University	OA	71.690	97.262	98.834	99.930	99.953	99.947
	AA	55.553	98.655	98.456	99.882	99.909	99.883
	KAPPA	57.041	91.957	98.456	99.907	99.938	99.929
Salinas Scene	OA	93.418	95.238	99.937	100.000	100.000	100.000
	AA	96.784	99.993	99.895	100.000	100.000	100.000
	KAPPA	92.663	93.103	99.929	100.000	100.000	100.000
WHU-Hi-Han Chuan	OA	81.575	99.271	99.917	99.956	99.953	99.849
	AA	62.336	98.205	99.800	99.802	99.872	99.816
	KAPPA	78.136	99.146	99.903	99.948	99.945	99.823

度的平均值。因为在实际问题中, 各类样本数量不均衡, 模型分类很容易偏向大类别而放弃小类别, 此时需要 KAPPA 系数进行一致性检验, 其范围为 -1~1, 通常落在 0~1 之间, 值越接近 1 表示一致性越高。

各个模型在 Indian Pines、Pavia University、Salinas Scene 和 WHU-Hi-Han Chuan 数据集上的分类精度如表 3 所示, 可以看出, 本文所提 3D-2D-1D CNN 在所有对比算法中精度都是最优的。另外, 一般的分类算法 SVM 的分类精度较低, 机器学习算法的整体分类精度较高。在 Indian Pines 数据集中, 3D CNN 与 2D CNN 相比, 分类精度有了较大提升, 从 89.56% 提升到 96.96%。本文所提模型 3D-2D-1D CNN 与 3D CNN 模型相比, 精度从 96.96% 提升到 99.65%, 表明所提模型具有较高的分类精度和更好的三维信息提取能力。为了充分证明所提 3D-2D-1D CNN 模型的精度优势, 将模型的中间输出值 3D-2D CNN(新)进行分类计算, 结果显示分类精度保持稳定。由此可见, 所提融合算法能够实现多维度卷积特征的互补, 在三个数据集上都取得了最高的分类精度。

各个深度学习模型在 Pavia University、Salinas Scene 和 WHU-Hi-Han Chuan 数据集上的分类精度都较高, 这是因为这些数据集的样本分布较为均衡, 样本之间的空-谱信息有明显差异。Indian Pines 数据集中每一个地物样本的正确率如表 4 所示, 可见存在一些差异较小的样本, 例如免耕大豆、少耕大豆和大豆, 进一步表明所提模型在不同的数据下都有稳定的性能。

本文所提 3D-2D-1D CNN 模型在 Indian Pines 数据集上的分类精度如表 4 所示,除“Grass pasture mowed”较低外,其他 15 类样本都达到了 98% 以上的分类精度,整体优于对比的其他所有模型。“Grass pasture mowed”类在各个模型中的分类精度十分不稳定,这是因为该类别在 3D-2D CNN(新)上的分类精度较高,其具有丰富的空间信息,光谱信息特征比空间信息特征弱,而所提模型比其他模型更多利用了光谱信息,最终该类别的分类精度降低。“Corn notill”、“Corn mintill”、“Corn”等植被具有丰富的光谱信息,所以分类精度与“Grass pasture

mowed”相反,在 3D-2D CNN(新)上的分类精度较低,在 3D-2D-1D CNN 上的分类精度有较大提升。“Buildings grass trees drives”和“Stone steel towers”等建筑物混合地物也是光谱一维特征具有显著性差异的代表。由于不同成分对光谱的吸收不同,光谱信息能充分反映样品内部的物理结构、化学成分的差异。例如高光谱具有丰富的一维光谱信息,在不同植被、矿物质的准确识别和精细分类方面具有优势;而对于“Grass pasture mowed”等纹理特征显著的地物,使用高光谱信息进行分类时效果较差。

表 4 各个模型对 Indian Pines 数据集中每个地物样本的分类精度

Table 4 Classification accuracy of each model for each ground object sample in Indian Pines dataset unit: %

No.	Class	SVM	2D CNN	3D CNN	3D-2D CNN	3D-2D CNN(new)	3D-2D-1D CNN
1	Alfalfa	52.174	15.217	100.000	87.500	87.500	100.000
2	Corn notill	84.804	71.919	96.700	98.500	90.600	99.600
3	Corn mintill	73.855	50.723	81.928	100.000	87.608	100.000
4	Corn	64.135	25.316	100.000	100.000	83.133	98.795
5	Grass pasture	89.648	65.010	98.225	100.000	97.633	99.408
6	Grass trees	96.027	87.945	99.804	100.000	99.609	99.609
7	Grass pasture mowed	71.429	21.429	100.000	95.000	100.000	90.000
8	Hay windrowed	89.331	61.088	100.000	100.000	100.000	100.000
9	Oats	45.000	5.000	100.000	92.857	92.857	100.000
10	Soybean notill	82.510	57.819	95.000	100.000	88.235	100.000
11	Soybean mintill	89.939	84.236	99.767	99.011	99.651	99.767
12	Soybean clean	77.572	45.363	97.831	98.313	91.807	99.277
13	Wheat	97.561	52.195	100.000	99.301	98.601	98.601
14	Woods	94.704	91.700	98.871	99.774	94.018	100.000
15	Buildings grass trees drives	69.430	40.933	94.074	100.000	93.704	98.519
16	Stone steel towers	79.570	50.538	100.000	100.000	90.769	100.000

为了进一步证明所提模型能更多利用光谱信息,对比了 Indian Pines 数据集的光谱一维特征与空间二维特征的差异性,分别绘制了各类别与其他类别的皮尔逊积矩相关系数图,类别编号与表 4 中编号相同,如图 4 所示。图 4(a)为利用光谱平均值计算的光谱相关系数,可见“Grass pasture mowed”与其他类别的相关系数最高,最难被区分;以  $3 \times 3$  矩阵为卷积核进行 4 次卷积运算,利用提取的空间特征计算空间相关系数,结果如图 4(b)所示,可见该类别的相关系数处于中等位置,表明所提模型在该类别上的分类精度较低的原因是提取了更抽象的光谱信息。

Indian Pines 数据集分类的测试结果图如图 5 所示,类别编号与表 4 中编号相同,可见在对比的 CNN 模型中,都存在一些分类精度较低的类别,其中圆圈所示的错分类明显。而本文所提模型解决了该问题,整体分类精度较高。模型前 100 次迭代过程中的损失和精度如图 6 所示,可以看出,模型在 20 次迭代时达到了收敛,表明模型具有很快的收敛速度。

同时,为了更好说明所提算法的分类性能,实验分析了各个数据集的训练时间和测试时间,结果如表 5 所示。可以看出,深度学习算法比一般分类算法 SVM 算法消耗更多的计算资源以获得精度的提

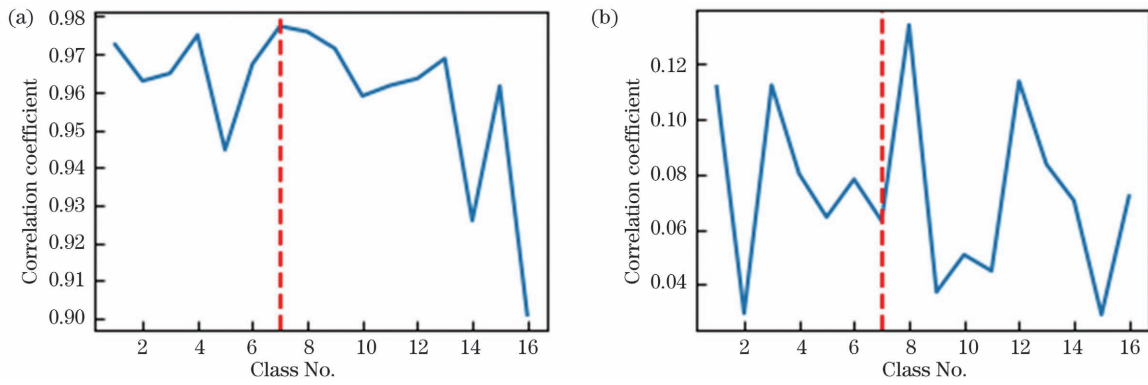


图 4 Indian Pines 数据集的光谱和空间特征相关系数图。(a)光谱特征相关系数;(b)空间特征相关系数

Fig. 4 Correlation coefficient graphs of spectral and spatial features of Indian Pines dataset. (a) Correlation coefficient of spectral features; (b) correlation coefficient of spatial features

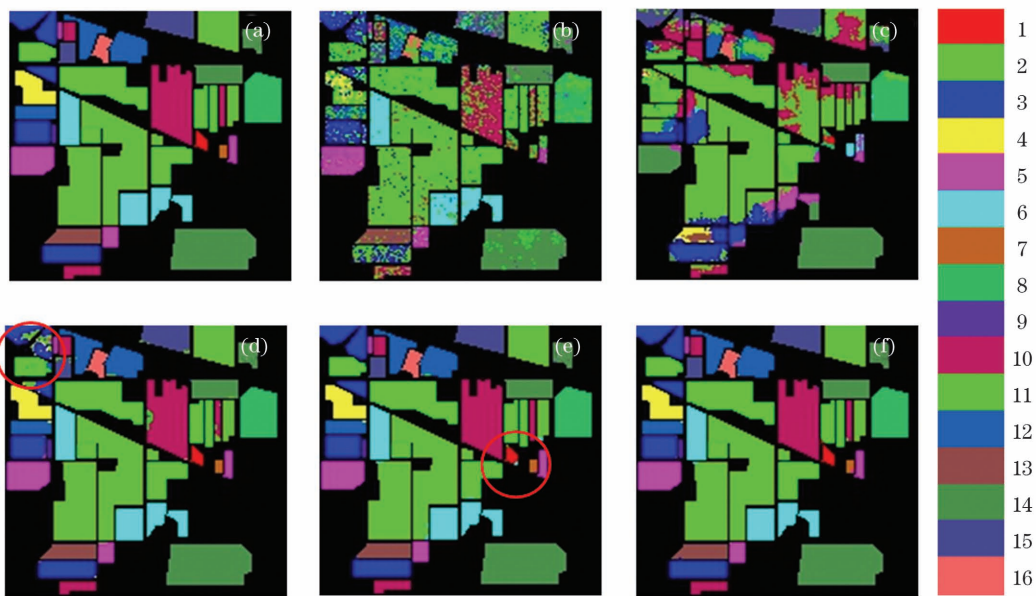


图 5 各个模型在 Indian Pines 数据集上的分类测试结果图。(a)地面真实;(b)SVM;(c) 2D CNN;(d) 3D CNN;(e) 3D-2D CNN;(f) 3D-2D-1D CNN

Fig. 5 Test results of each model classification in Indian Pines dataset. (a) Ground truth; (b) SVM; (c) 2D CNN; (d) 3D CNN; (e) 3D-2D CNN; (f) 3D-2D-1D CNN

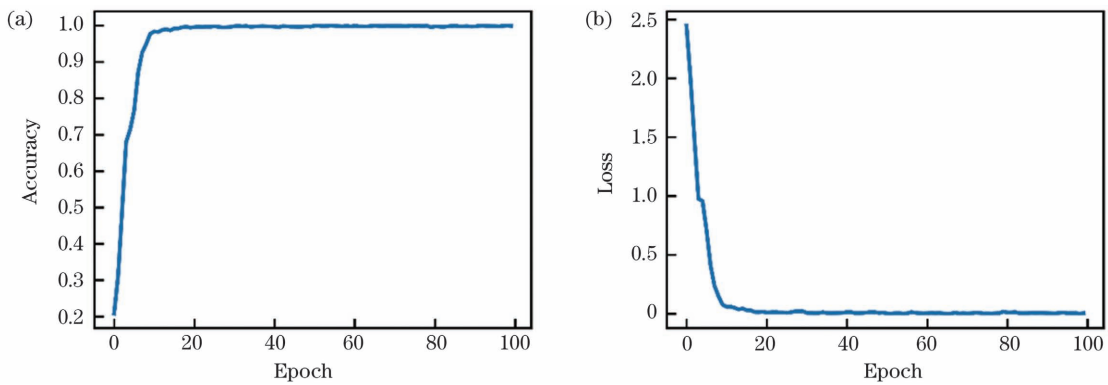


图 6 所提模型在前 100 次迭代过程中的损失和总体分类精度。(a)总体分类精度;(b)损失

Fig. 6 Overall classification accuracy and loss of proposed model in 100 epochs. (a) Overall classification accuracy; (b) loss

表 5 各个模型在 Indian Pines、Pavia University、Salinas Scene 和 WHU-Hi-Han Chuan 数据集中的训练时间和测试时间  
Table 5 Training time and test time of each model in Indian Pines, Pavia University, Salinas Scene, and WHU-Hi-Han Chuan datasets

Dataset	Classification performance	SVM	2D CNN	3D CNN	3D-2D CNN	3D-2D CNN(new)	3D-2D-1D CNN
Indian Pines	Training time /s	1.18	901.20	1477.93	968.71	642.97	600.77
	Test time /s	0.98	6.51	18.04	20.04	14.28	13.37
Pavia University	Train time /s	2.23	180.28	953.88	725.11	564.86	564.32
	Test time /s	4.38	4.37	21.20	23.29	20.17	19.81
Salinas Scene	Train time /s	3.41	572.69	1155.50	920.70	706.34	664.93
	Test time /s	8.34	5.5	25.93	29.42	25.43	22.83
WHU-Hi-Han Chuan	Training time /s	678.335	1396.175	7935.224	5372.430	3410.025	3228.423
	Test time /s	862.016	27.298	148.706	149.471	129.047	114.472

升;而对于卷积神经网络,3D CNN 与 2D CNN 相比,模型更加复杂,所以训练时间大幅增加;本文所提模型 3D-2D-1D CNN 与 3D CNN 相比,分类精度更高,而且节省了较多计算资源;与 3D-2D CNN 相比,3D-2D-1D CNN 模型分类速度优势同样明显。另外,WHU-Hi-Han Chuan 数据集的数据量大,本文所提模型分类速度提升得最快,表明该模型在大量数据分析中有较大的应用潜力。

以上实验表明,本文所提模型在分类精度和速度方面与对比算法相比都是最优的。由于所提模型分层次使用三种维度的卷积神经网络,因此融合模型能够分阶段提取空-谱特征、空间特征和光谱特征,提升了 3D CNN 对高光谱数据特征的提取能力,同时有效利用了 2D CNN 和 1D CNN 的速度优势,能有效降低计算成本,并保持较高的分类精度。

## 4 结 论

针对 3D CNN 计算成本过高的情况,提出了一种 3D-2D-1D CNN 融合模型。该模型能有效减少网络的参数和计算成本,并具有较高的分类精度。通过 Indian Pines、Pavia University、Salinas Scene 和 WHU-Hi-Han Chuan 数据集的实验,得到以下结论。

首先,在相同参数设置下,3D-2D-1D CNN 不仅能够达到甚至超过 3D CNN 的分类精度,而且能够大幅度降低时间成本和计算成本,是一种更有效率的网络结构。

其次,相比于 SVM、2D CNN 等方法,所提融合模型同时利用了空-谱联合的互补信息,因此在总体分类精度、平均分类精度和 Kappa 系数三种评价指标上都有更加优异的表现。

虽然针对 3D CNN 的计算成本问题进行了改进,但是没有对高维光谱特征进行有效的特征降维和特征转换,下一步将着重研究如何利用 CNN 方法有效提取出光谱信息中有助于提高分类精度的特征。

## 参 考 文 献

- [1] Bioucas-Dias J M, Plaza A, Camps-Valls G, et al. Hyperspectral remote sensing data analysis and future challenges[J]. IEEE Geoscience and Remote Sensing Magazine, 2013, 1(2): 6-36.
- [2] Wang J G, Guan S. Research on development of hyper-spectral imaging satellite [J]. Electro-Optic Technology Application, 2020, 35(3): 1-7.  
王君光, 关松. 高光谱成像卫星发展研究[J]. 光电技术应用, 2020, 35(3): 1-7.
- [3] Fauvel M, Tarabalka Y, Benediktsson J A, et al. Advances in spectral-spatial classification of hyperspectral images[J]. Proceedings of the IEEE, 2013, 101(3): 652-675.
- [4] Tan K, Wang H M, Chen L H, et al. Estimation of the spatial distribution of heavy metal in agricultural soils using airborne hyperspectral imaging and random forest[J]. Journal of Hazardous Materials, 2020, 382: 120987.
- [5] Li W, Dou Z G, Cui L J, et al. Suitability of hyperspectral data for monitoring nitrogen and phosphorus content in constructed wetlands [J]. Remote Sensing Letters, 2020, 11(5): 495-504.
- [6] Zhang Y S, Wu L, Ren H Z, et al. Mapping water quality parameters in urban rivers from hyperspectral images using a new self-adapting selection of multiple artificial neural networks[J]. Remote Sensing, 2020, 12(2): 336.
- [7] Veraverbeke S, Dennison P, Gitas I, et al.



- Hyperspectral remote sensing of fire: state-of-the-art and future perspectives [J]. *Remote Sensing of Environment*, 2018, 216: 105-121.
- [8] Li S T, Song W W, Fang L Y, et al. Deep learning for hyperspectral image classification: an overview [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(9): 6690-6709.
- [9] Luo F L, Zhang L P, Du B, et al. Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(8): 5336-5353.
- [10] Sun L, Wu F Y, Zhan T M, et al. Weighted nonlocal low-rank tensor decomposition method for sparse unmixing of hyperspectral images [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020, 13: 1174-1188.
- [11] Vincent F, Besson O. One-step generalized likelihood ratio test for subpixel target detection in hyperspectral imaging [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(6): 4479-4489.
- [12] Zhang H K, Li Y, Jiang Y N. Deep learning for hyperspectral imagery classification: the state of the art and prospects[J]. *Acta Automatica Sinica*, 2018, 44(6): 961-977.  
张号逵, 李映, 姜晔楠. 深度学习在高光谱图像分类领域的研究现状与展望[J]. *自动化学报*, 2018, 44(6): 961-977.
- [13] Lee J Y, Kim J S, Kim T Y, et al. Detection and classification of intracranial haemorrhage on CT images using a novel deep-learning algorithm [J]. *Scientific Reports*, 2020, 10(1): 20546.
- [14] Singh D, Kumar V, Vaishali, et al. Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks[J]. *European Journal of Clinical Microbiology & Infectious Diseases*, 2020, 39(7): 1379-1389.
- [15] Liu J M, Yang S, Huang H. Hyperspectral remote sensing image classification based on local reconstruction Fisher analysis[J]. *Chinese Journal of Lasers*, 2020, 47(7): 0710001.  
刘嘉敏, 杨松, 黄鸿. 基于局部重构 Fisher 分析的高光谱遥感影像分类[J]. *中国激光*, 2020, 47(7): 0710001.
- [16] Liu L X, He D, Li M Z, et al. Identification of Xinjiang jujube varieties based on hyperspectral technique and machine learning[J]. *Chinese Journal of Lasers*, 2020, 47(11): 1111002.  
刘立新, 何迪, 李梦珠, 等. 基于高光谱技术与机器
- 学习的新疆红枣品种鉴别[J]. *中国激光*, 2020, 47(11): 1111002.
- [17] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [18] Liu L L, Wu F X, Wang Y P, et al. Multi-receptive-field CNN for semantic segmentation of medical images[J]. *IEEE Journal of Biomedical and Health Informatics*, 2020, 24(11): 3215-3225.
- [19] Masood A, Sheng B, Yang P, et al. Automated decision support system for lung cancer detection and classification via enhanced RFCN with multilayer fusion RPN [J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(12): 7791-7801.
- [20] Abdulwahab S, Rashwan H A, Garcia M Á, et al. Adversarial learning for depth and viewpoint estimation from a single image [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(9): 2947-2958.
- [21] Chen F M, Wen C, Xie K, et al. Face liveness detection: fusing colour texture feature and deep feature[J]. *IET Biometrics*, 2019, 8(6): 369-377.
- [22] Chen H, Deng F. Hyperspectral remote sensing image classification based on decomposed three-dimensional convolutional neural network[J]. *Science of Surveying and Mapping*, 2020, 45(8): 96-102, 129.  
陈亨, 邓非. 分解式三维卷积神经网络的高光谱遥感影像分类[J]. *测绘科学*, 2020, 45(8): 96-102, 129.
- [23] Sabokrou M, Fayyaz M, Fathy M, et al. Deep-cascade: cascading 3D deep neural networks for fast anomaly detection and localization in crowded scenes [J]. *IEEE Transactions on Image Processing*, 2017, 26(4): 1992-2004.
- [24] Zhang H K, Li Y. Spectral-spatial classification of hyperspectral imagery based on deep convolutional network [C] // 2016 International Conference on Orange Technologies (ICOT), December 18-20, 2016, Melbourne, VIC, Australia. New York: IEEE Press, 2016: 44-47.
- [25] Leng J B, Li T, Bai G, et al. Cube-CNN-SVM: a novel hyperspectral image classification method[C]// 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI), November 6-8, 2016, San Jose, CA, USA. New York: IEEE Press, 2016: 1027-1034.
- [26] 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.
- [27] Zhang J, Wei F Y, Feng F, et al. Spatial-spectral feature refinement for hyperspectral image classification based on attention-dense 3D-2D-CNN [J]. Sensors, 2020, 20(18): 5191.
- [28] Han M X, Cong R M, Li X Y, et al. Joint spatial-spectral hyperspectral image classification based on convolutional neural network[J]. Pattern Recognition Letters, 2020, 130: 38-45.
- [29] Zhang B, Zhao L, Zhang X L. Three-dimensional convolutional neural network model for tree species classification using airborne hyperspectral images[J]. Remote Sensing of Environment, 2020, 247: 111938.
- [30] Kong F Q, Zhou Y B, Shen Q, et al. End-to-end multispectral image compression using convolutional neural network[J]. Chinese Journal of Lasers, 2019, 46(10): 1009001.
- 孔繁镛, 周永波, 沈秋, 等. 基于卷积神经网络的端到端多光谱图像压缩方法[J]. 中国激光, 2019, 46(10): 1009001.
- [31] 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.
- [32] Hamida A B, Benoit A, Lambert P, et al. 3-D deep learning approach for remote sensing image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(8): 4420-4434.
- [33] Liu B, Yu X C, Zhang P Q, et al. A semi-supervised convolutional neural network for hyperspectral image classification[J]. Remote Sensing Letters, 2017, 8 (9): 839-848.

## Multi-Dimensional CNN Fused Algorithm for Hyperspectral Remote Sensing Image Classification

Liu Jinxiang<sup>1</sup>, Ban Wei<sup>1</sup>, Chen Yu<sup>1</sup>, Sun Yaqin<sup>1</sup>, Zhuang Huifu<sup>1</sup>, Fu Erjiang<sup>2</sup>,  
Zhang Kefei<sup>1,2,3\*</sup>

<sup>1</sup> School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China;

<sup>2</sup> Bei-Stars Geospatial Information Innovation Institute, Nanjing, Jiangsu 210000, China;

<sup>3</sup> Space Research Centre, RMIT University, Victoria, Melbourne 3001, Australia

### Abstract

**Objective** Now hyperspectral images have high spatial and spectral resolution, and play an important role in the fields of land monitoring, environmental protection, earthquake prevention, and disaster reduction. However, the high dimensionality and large data volume of hyperspectral data bring many problems (e.g., strong correlation among bands, mixed redundant pixels, and data information redundancy) to hyperspectral classification. With the continuous development of deep learning technology, the convolutional neural network (CNN), as one of its representative algorithms, provides a new solution for hyperspectral image classification. There are three common hyperspectral classification methods based on the CNN network. Among them, 1D CNN extracts spectral information, and 2D CNN extracts spatial information. In contrast, 3D CNN is usually composed of three-dimensional convolution kernels, which can extract two-dimensional spatial features and one-dimensional spectral features at the same time. Although 3D CNN has a better effect in spatial-spectral information fusion, this model is more complex, which increases the cost of network calculation and the number of parameters. With the rapid expansion of data volume, the classification accuracy and speed of a complex model are not satisfactory. Here we propose a lightweight fusion CNN algorithm, 3D-2D-1D CNN, for hyperspectral image classification. This algorithm organically integrates CNNs of different dimensions, reduces the calculation amount of 3D CNN operations, and makes full use of the hyperspectral spatial-spectral joint information. We hope that our basic strategy and findings can be helpful to improve the applicability and computational efficiency of the model.

**Methods** A hybrid algorithm 3D-2D-1D CNN model (Fig. 1) is described as follows. Firstly, the hyperspectral data is processed by 3D CNN. In the hyperspectral cube, a three-dimensional convolution kernel is used for convolution calculation. Each feature map of the convolutional layer is connected to multiple spectrograms, so the information of

the fused spatial and spectral features can be extracted at the same time. After that, 2D CNN is used to extract spatial information, and subsequently the output information is subjected to 1D CNN operations to further extract high-dimensional spectral information. Different feature information is extracted by performing two-dimensional and one-dimensional convolution operations, respectively. Finally, classification is performed according to the extracted feature maps. The algorithm proposed here retains the spatial-spectral joint information extraction of 3D CNN, and replaces part of the subsequent three-dimensional convolution layer process with a 2D convolution process and a 1D convolution process with less calculation. According to the rules of convolution operation, the proposed algorithm improves operation efficiency (Fig. 2).

**Results and Discussions** The 3D-2D-1D CNN proposed here has the best accuracy among all three algorithms for comparison (Table 3). The classification accuracy of machine learning algorithms is higher than that of general classification algorithms (such as SVM). The classification accuracy of the intermediate output value of the proposed 3D-2D CNN model (new) remains stable, indicating that the fusion algorithm can achieve the complementarity of multi-dimensional convolution features. The 3D-2D-1D CNN model has high classification accuracy on data with large differences in spectral characteristics (Fig. 4), which proves that the model can extract more abstract spectral information. The model has fast convergence, reaching convergence in 20 iterations (Fig. 6). Deep learning algorithms consume more computing resources than the general classification algorithm SVM. Compared with 2D CNN, 3D CNN has a more complex model, so the training time is greatly increased (Table 5). Compared with the existing basic 3D CNN model and the existing improved algorithm 3D-2D CNN, the 3D-2D-1D CNN model has the highest classification accuracy and the fastest calculation speed. As the amount of data in the dataset increases, the speed of the proposed model increases fast, indicating that the model can be applied in large-scale data analysis with a great potential. In general, the model proposed here is the optimal in terms of classification accuracy and speed.

**Conclusions** In view of the high computational cost of 3D CNN, a 3D-2D-1D CNN fusion model is established. The model can effectively reduce the network parameters and calculation cost, and has a high classification accuracy. Firstly, the network structure of 3D-2D-1D CNN can not only reach or exceed the classification accuracy of 3D CNN under the same parameter settings, but also it can greatly reduce the time cost and calculation cost, which is a more efficient network structure. Moreover, the model has a great data advantage in big data calculation. Second, compared with SVM, 2D CNN and other methods, the fusion model proposed here uses the complementary information of spatial-spectral combination at the same time, so it is the most excellent in overall classification accuracy, average classification accuracy, and Kappa coefficient. Finally, because the proposed model uses 3D CNN hierarchically, the fusion model can extract spatial-spectral features, spatial features, and spectral features in stages. The model improves the ability of 3D CNN to extract the features of hyperspectral data, and at the same time effectively utilizes the speed advantages of 2D CNN and 1D CNN, so it can effectively reduce the calculation cost and maintain a high classification accuracy.

**Key words** remote sensing; hyperspectral image classification; 3D-2D-1D convolutional neural network; deep learning

**OCIS codes** 280.4788;120.0280; 110.4234; 100.4996