

引用格式: CHENG Yin Zhu, LIU Song, WANG Nan, et al. Multi-scale Remote Sensing Image Classification Based on Weighted Feature Fusion[J]. Acta Photonica Sinica, 2023, 52(11):1110002

程寅翥,刘松,王楠,等.基于加权特征融合的多尺度遥感影像分类[J].光子学报,2023,52(11):1110002

基于加权特征融合的多尺度遥感影像分类

程寅翥^{1,2},刘松^{1,2},王楠^{1,2},师悦天^{1,2},张耿¹

(1 中国科学院西安光学精密机械研究所 光谱成像技术重点实验室,西安 710119)

(2 中国科学院大学,北京 100049)

摘要:针对普通的三维卷积神经网络(3D CNN)从一个尺度上提取特征,会丢失部分细节信息,且对小样本任务表现一般的问题,本文提出了一种三支路的 3D CNN,从不同尺度上提取特征后进行加权特征融合,从而获取了更为全面的特征;并引入数据增强技术,从而改善了小样本情形下的分类性能。现有特征融合方法通常对各个支路直接进行拼接,本文采用加权拼接的特征融合方法,将各特征分别乘以一个加权系数后再进行拼接,该系数通过模拟退火算法求取。本文方法在公开数据集 Indian Pines, Pavia University, Salinas 等上采用 10% 的数据进行训练,分别得到了 98.60%、99.83%、99.97% 的总体准确率,与各类对比方法相比,提升了高光谱遥感影像分类问题的准确率。

关键词:光学遥感影像;深度学习;图像分类;卷积神经网络;特征融合;多尺度

中图分类号:P407.8

文献标识码:A

doi:10.3788/gzxb20235211.1110002

0 引言

遥感影像分类是遥感影像处理中的关键分支,为精细农业、矿物勘探、城市规划以及其它工农业应用提供了重要依据。近年来,光谱成像技术从多光谱技术发展到高光谱技术,丰富的谱段信息对遥感影像分类算法提出了更高的要求。

许多基于传统方法的高光谱影像分类算法,例如超像素方法^[1]、扩展形态学特征方法^[2]、基于组合核的空-谱联合分类算法^[3]、基于支持向量机和图割的分类算法^[4],在各个高光谱公开数据集上均取得了一定的效果。这些方法往往采用人工设计的某种特征提取方法,结合经典的分类器,如支持向量机^[5]、逻辑回归^[6]、随机森林^[7]等进行分类。

近年来,受益于硬件条件的提升和算法的更新迭代,各类深度学习算法层出不穷,并被研究人员引入高光谱遥感影像分类领域。这类方法为端到端的分类,特征提取和分类器设计合二为一,且采用多层网络进行特征提取,在误差反向传播的过程中不断优化网络参数,从而能够提取图像块更为深层的信息,大幅改进了原有算法的性能。2014年,CHEN Yushi等^[8]将 GEOFFREY Hinton 等提出的深度学习概念引入高光谱影像分类,并验证了堆栈自编码器在遥感影像分类领域的有效性。后来,深度信念网络^[9]、孪生神经网络^[10]、长短时记忆网络^[11]、胶囊网络^[12]、门控循环单元^[13]、Transformer^[14]等神经网络结构被相继引入高光谱影像分类领域,并取得了优良的性能。卷积神经网络(Convolutional Neural Network, CNN)是一种用于图像处理的神经网络,其通过模拟人类视觉过程,提取感受野的特征,并通过共享权值减少参数量。在高光谱影像分类领域, CNN 能够天然地融合空间维度和光谱维度的特征,基于 CNN 的高光谱影像分类算法因而在各个公开数据集上取得了比许多其他算法更好的效果。在此基础上,SHI Yuetian等^[15]提出了利用多角度平行特征编码的方式,通过增强局部空间特征的方式提高图像分类精度,并且该算法对图像旋转是鲁棒的。三维卷积

基金项目:基础加强计划重点项目(No. 2022-JCJQ-ZD-215-03),自主部署项目(No. S22-037)

第一作者:程寅翥, cyz19970209@163.com

通讯作者:张耿, gzhang@opt.ac.cn

收稿日期:2023-04-04;录用日期:2023-05-22

<http://www.photon.ac.cn>

神经网络(Three-dimensional Convolutional Neural Network, 3D CNN)^[16]是一种常用于高光谱影像处理的卷积神经网络结构,与常见的2D CNN不同,3D CNN的卷积核为一个立方体,能比2D CNN更好地提取光谱维度的特征,从而在各个高光谱公开数据集上取得了比2D CNN更为优秀的性能指标。

尽管3D CNN方法在目前的高光谱影像分类领域取得了一定的效果,但是该方法仍然存在许多问题。传统的3D CNN通常从单一的尺度提取图像块,往往会损失一定的局部信息;单纯通过增加模型的深度来提高参数量,往往会导致梯度消失等问题^[17],且容易导致过拟合。据此,本文提出了一种多支路3D CNN,三条支路分别采用三种不同的3D CNN结构设计,从不同的尺度上提取了图像块的特征,从而既能获取全局信息又能兼顾局部信息,同时增加了参数量,提升了模型拟合能力。另外,为了缓解过拟合现象,引入了数据增强,并在训练阶段使用了丢弃法,本文引入了一种基于模拟退火算法的加权特征融合方法,提升了算法的性能。为了验证本文方法的有效性,分别在公开数据集 Indian Pines, Pavia University, Salinas 等上采用10%的有标签数据进行训练,分别得到了98.60%、99.83%、99.97%的总体准确率,达到了目前的领先水平。进一步的实验结果表明,当有标签的数据比例从10%下降到2%时,本文方法相比对比方法的分类总体准确率更高。此外,还对本文方法进行了实用性验证,说明了本文方法具有一定的应用价值。

1 本文方法

如图1所示,对于输入的高光谱数据立方体,本文首先采用主成分分析对数据进行降维,降维后光谱的维数选取为40。将数据立方体分解为许多 $19 \times 19 \times 40$ 的图像块,并以其中中心像素的标签作为整个图像块的标签,之后采用旋转 $90^\circ, 180^\circ, 270^\circ$ 的方法进行数据增强。在特征提取阶段,采用了三支路并联的三维卷积神经网络,分别从 $2 \times 2, 4 \times 4, 6 \times 6$ 三种空间尺度来提取特征。在训练阶段,采用Adam优化器对三条支路的参数分别进行优化,并选用交叉熵损失函数。在测试阶段,对从三条支路上提取到的特征进行加权特征融合。在分类器方面,本文选择了逻辑回归分类器,这是一种经典而有效的分类器,对于中小规模的数据集具有不劣于全连接神经网络的效果^[18]。

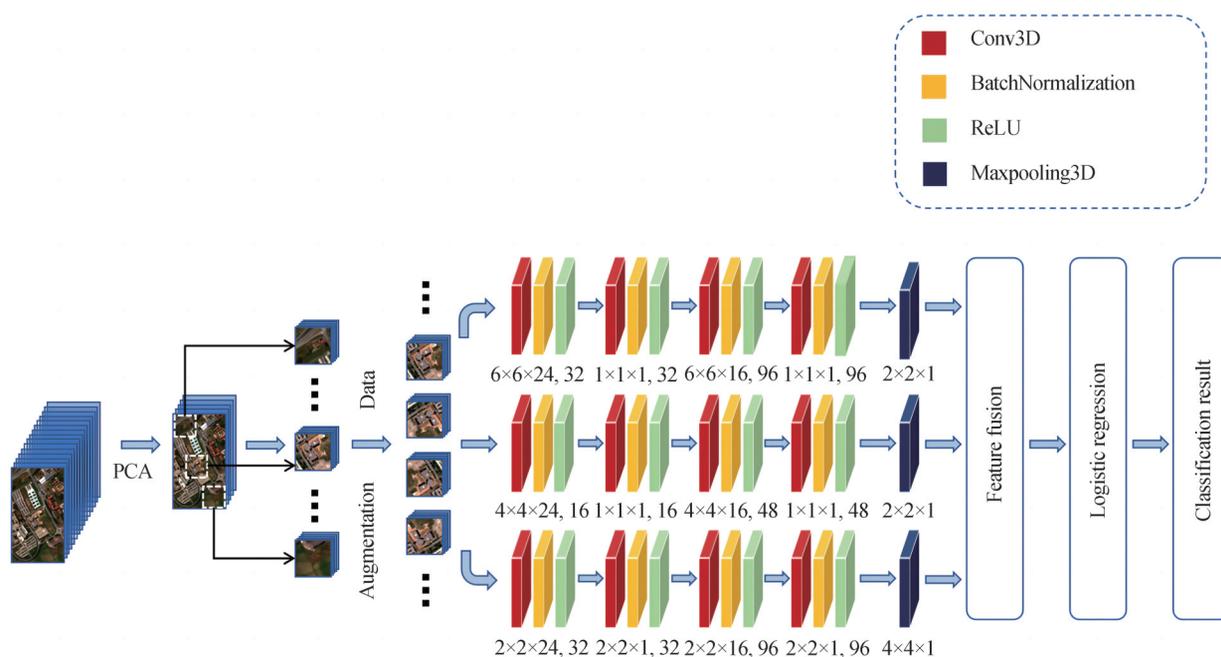


图1 针对高光谱遥感影像分类的多支路3D CNN结构
Fig.1 multi-branch 3D CNN framework for HSIC

1.1 预处理

可以将高光谱数据集描述为一个数据立方体 $X \in \mathbb{R}^{n_1 \times n_2 \times n_{\text{bands}}}$ 。 $n_1 \times n_2$ 为像素规模,代表空间信息。而 n_{bands} 为波段数目,代表光谱信息。高光谱影像分类的要求是逐像素的分类,需要根据每个像素点的光谱信

息以及上下文信息给出其类别。假设该数据集一共有 n_{classes} 类,则需要建立一个从高光谱数据立方体 $X \in \mathbb{R}^{n_1 \times n_2 \times n_{\text{bands}}}$ 到集合 $Y = \{1, 2, \dots, n_{\text{classes}}\}^{n_1 \times n_2}$ 的映射。

对于预处理模块。HUGHES^[19]现象指出,随着光谱波段范围增大和光谱分辨率的提高,分类准确率出现先上升后下降的现象。其出现的原因主要是:遥感影像各个谱段的相关性较高,因而有一定的冗余度,干扰了后续图像处理的过程。为此,研究人员提出了一系列的算法^[20],如主成分分析、奇异值分解、线性判别分析、独立成分分析,以及各类波段选择方法等用于降低数据维度,缓解 HUGHES 现象。根据 MUHAMMAD A 等^[21]的对比实验,在常见高光谱公开数据集下,对于多数分类算法,使用主成分分析算法进行降维的效果最好,故本文选取主成分分析作为降维算法。

将光谱数据中心化,并将其存放到一个 $m \times n$ 行, n_{bands} 列的矩阵 A 中^[22]。协方差矩阵 $A^T A$ 的特征值分解为

$$A^T A = P \Lambda P^T \quad (1)$$

式中, Λ 对角线元素按照模的大小由大到小排列,取 P 的前若干列组成 W ,则主成分分析降维后的数据 A' 为

$$A' = A W \quad (2)$$

经由主成分分析,数据矩阵 A 被投影到 $A W$,光谱由 n_{bands} 维降到了 40 维。

为了将数据立方体输入 3D CNN 中,以每个像素为中心,取其周围 $19 \times 19 \times 40$ 的立方体为图像块(边缘部分不足的,使用零填充补齐),从而得到一系列 3D 图像块,以这些 3D 图像块中心像素的标签作为整个图像块的标签进行训练。

LI Wei 等^[23]的研究表明,将图像块绕着 Z 轴旋转 90° 、 180° 、 270° ,可以有效缓解高光谱影像分类过程中的过拟合现象。具体做法如图 2 所示。

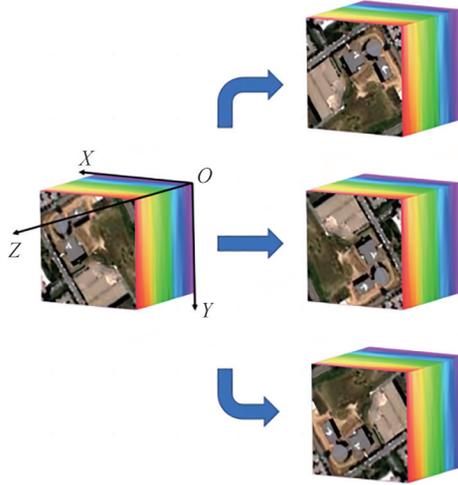


图 2 高光谱遥感影像的旋转数据增强方法
Fig.2 Rotational data augmentation for HSIC

1.2 三支路 3D CNN 结构

3D CNN 包含输入层、三维卷积层、激活函数、池化层、全连接层、输出层等。如图 3 所示,三维卷积核与图像块进行点乘,并沿着三个互相正交的方向滑动,从而形成一幅三维特征图。一个卷积层包含多个三维卷积核,从而形成多幅三维特征图。这些特征图经由参数 b 进行偏置,再通过激活函数 F 实现非线性映射。若第 l 层共有 d_l 幅特征图,且卷积核的长、宽、高分别为 $2\delta+1$, $2\gamma+1$, $2\xi+1$,则第 i 层的第 j 幅特征图中,经过偏置和激活函数后,其 (x, y, z) 处的值 $v_{i,j}^{x,y,z}$ 表示为

$$v_{i,j}^{x,y,z} = F(b_{i,j} + \sum_{\tau=1}^{d_{l-1}} \sum_{\lambda=-\xi}^{\xi} \sum_{\rho=-\gamma}^{\gamma} \sum_{\sigma=-\delta}^{\delta} w_{i,j,\tau}^{\sigma,\rho,\lambda} v_{i-1,\tau}^{x+\sigma,y+\rho,z+\lambda}) \quad (3)$$

选取 ReLU 函数为激活函数,可以减少计算量,并缓解梯度消失问题,该函数表达式为

$$\text{ReLU}(x) = \begin{cases} x, & x \geq 0 \\ 0, & \text{others} \end{cases} \quad (4)$$

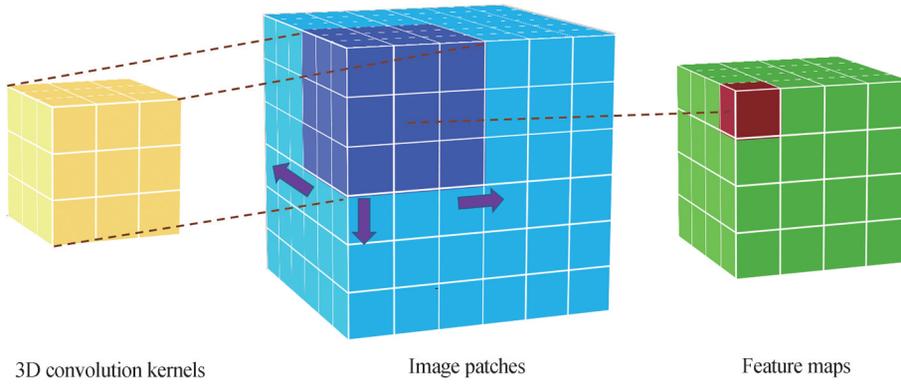


图3 3D卷积层示意图

Fig.3 Schematic diagram of 3D convolution layer

池化层起到降采样的作用,本文方法选择最大池化层(Max pooling)。该层将输入特征图分为许多矩体,并取每个矩体内的最大值作为该区域的输出。从而降低了参数量,缓解了过拟合。全连接层是神经网络中起到分类器作用的层,该层的神经元密集连接。

在训练阶段,采用三条支路分开训练的方法,三条支路的网络结构图如图4所示。

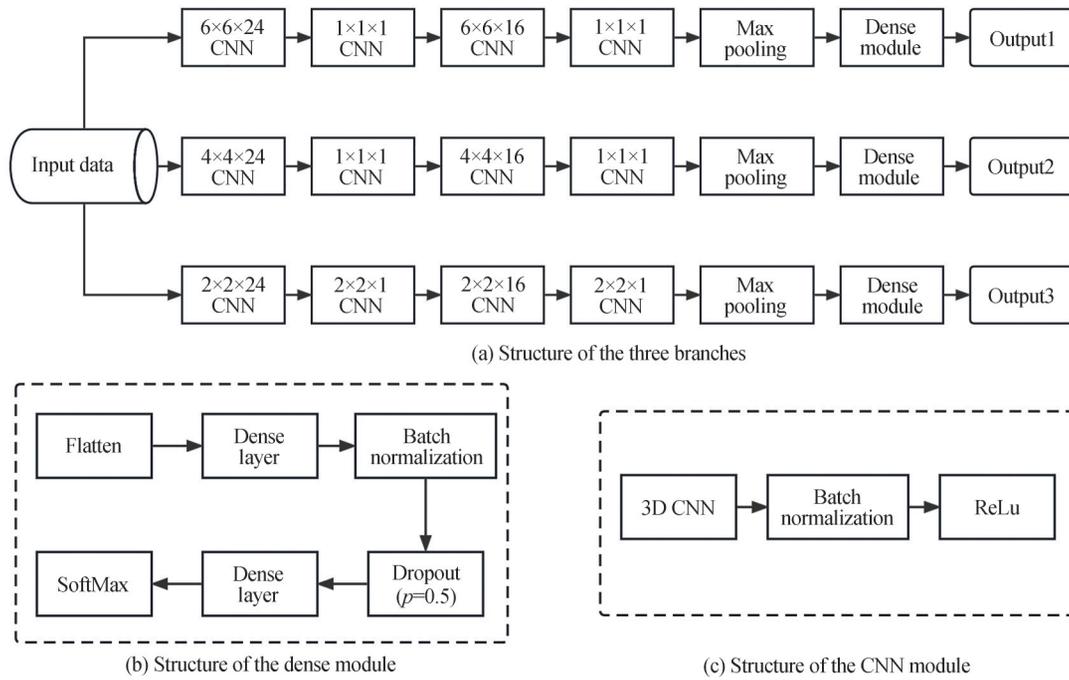


图4 训练阶段三条支路的网络结构图

Fig.4 The network structure diagram of the three branches in the training phase

为了解决常规3D CNN通常从一个尺度上提取特征,往往会丢失部分局部信息的问题,本文方法将数据输入三条3D CNN支路中,分别从2x2,4x4,6x6三种空间尺度来提取特征。在网络结构中用到了批量标准化^[24](Batch Normalization, BN),以缓解梯度消失问题,加快网络收敛速度。在全连接层模块中使用了丢弃法^[25](Dropout),将输入全连接层的数据的各个分量以0.5的概率置为0,以缓解过拟合现象。

每批次数据的大小(Batchsize)为 n_{batch} 的情形下,若输入 Batch normalization 层的数据为 $x_i (i=1, 2, \dots, n_{batch})$, Batch normalization 单元的表达式为

$$\mu = \frac{1}{n_{\text{batch}}} \sum_{i=1}^{n_{\text{batch}}} x_i \quad (5)$$

$$\sigma^2 = \frac{1}{n_{\text{batch}}} \sum_{i=1}^{n_{\text{batch}}} (x_i - \mu)^2 \quad (6)$$

$$\tilde{x}_i = \frac{x_i - \mu}{\sigma} \quad (7)$$

经由上述变换,每批次内的数据均值变为0,方差变为1。再经由下列的变换,得到 Batch normalization 层的输出(γ 和 β 是可学习参数)

$$y_i = \gamma \tilde{x}_i + \beta \quad (8)$$

通过 softmax 层产生输出,若输入 softmax 层的数据为 $x_t (t=1, 2, \dots, n_{\text{class}})$, 其类别为 c 的概率为

$$p(\text{label} = c) = \frac{\exp(x_c)}{\sum_{t=1}^{n_{\text{class}}} \exp(x_t)} \quad (9)$$

采用交叉熵损失函数^[26],该损失函数下网络收敛较快。交叉熵损失函数表达式如下(n_{batch} 为该批次样本总数, n_{class} 为类别数, I_{ic} 在样本 i 被预测正确时为1,否则为0)

$$L = -\frac{1}{n_{\text{batch}}} \sum_{i=1}^{n_{\text{batch}}} \sum_{c=1}^{n_{\text{class}}} I_{ic} \log(p(\text{label}_i = c)) \quad (10)$$

采用 Adam 优化器^[27]进行优化,这是一种自适应学习率的优化器,是带动量的随机梯度下降算法和 RMSProp 算法的结合。实验表明,该优化器在现有的多数数据集上可以使得网络收敛速度较快。

1.3 基于模拟退火算法的加权特征融合

在测试阶段,通过三条 3D CNN 支路提取特征,并进行特征融合后,采用逻辑回归分类器进行分类。现有的特征融合方法通常将三条支路的特征直接拼接,而本文提出了一种加权的特征融合方法。本文特征融合的表达式为(\otimes 为克罗内克积; $\alpha_1, \alpha_2, \alpha_3$ 为加权系数; f_1, f_2, f_3 为三条支路的特征; f 为融合后的特征)

$$f = (\alpha_1, 0, 0)^T \otimes f_1 + (0, \alpha_2, 0)^T \otimes f_2 + (0, 0, \alpha_3)^T \otimes f_3 \quad (11)$$

式中, $\alpha_1, \alpha_2, \alpha_3$ 最优的取值可以采用模拟退火算法得到。模拟退火算法^[28]是对固体退火过程的模拟,以 $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ 取值为模拟退火算法的状态向量;将模拟退火算法的能量函数,即损失函数设为 $L(\alpha)$ (本实验中以验证集准确率的负值为损失函数)。温度为 T 时,由状态 $(\alpha_1, \alpha_2, \alpha_3)$ 转移到新状态 $(\alpha'_1, \alpha'_2, \alpha'_3)$ 的概率由以下的 Metropolis 准则求得

$$P_{ij} = \min \left(1, \exp \left(\frac{L(\alpha_1, \alpha_2, \alpha_3) - L(\alpha'_1, \alpha'_2, \alpha'_3)}{T} \right) \right) \quad (12)$$

由马尔科夫链^[29]相关理论,随着 $T \rightarrow 0$,该随机过程在各温度下达到平衡状态后, $L(\alpha_1, \alpha_2, \alpha_3) \rightarrow L_{\min}$,此时取到了最优的加权系数。具体而言,模拟退火算法由外循环和内循环组成,其算法过程描述为

输入: i (外循环当前轮次), j (内循环当前轮次), n_1 (外循环最大迭代次数), n_2 (内循环最大迭代次数), α (状态向量), $\text{new}(\alpha)$ (状态更新准则), $L(\alpha)$ (损失函数), $T(i)$ (温度函数)。

执行:

$\alpha \leftarrow \text{random}$ (随机初始化)

$i \leftarrow 0$ (初始化外循环当前轮次)

$j \leftarrow 0$ (初始化内循环当前轮次)

while $i < n_1$ do(控制外循环)

 while $j < n_2$ do(控制内循环)

$\alpha_{\text{new}} \leftarrow \text{new}(\alpha)$ (产生新状态)

 if $L(\alpha_{\text{new}}) < L(\alpha)$ (Metropolis 准则)

$\alpha \leftarrow \alpha_{\text{new}}$

else if $U[0, 1] < \exp\left(\frac{L(\alpha) - L(\alpha_{\text{new}})}{T(i)}\right)$ ($U[0, 1]$ 表示 $[0, 1]$ 区间上的均匀分布随机数)

$\alpha \leftarrow \alpha_{\text{new}}$

end if

end if

$j \leftarrow j+1$

end while

$j \leftarrow 0$

$i \leftarrow i+1$

end while

输出: $\alpha, L(\alpha)$ (返回损失函数最小值及相应最优解)

1.4 逻辑回归分类器

采用逻辑回归算法对样本进行分类。该分类器通过逻辑斯蒂克函数刻画样本属于某一类的概率。假设共有 n_{class} 类样本, 记为 $1, \dots, c, \dots, n_{\text{class}}$ 。

设第 c 类样本数为 n_c , 该类中各个样本的特征为 $f_i (i=1, 2, \dots, n_c)$, 若样本 f_i 的标签为 c , 则 y_i 取 1, 否则 y_i 取 0。在上述符号规定下, 逻辑回归分类器的似然函数为

$$L(\omega_c, b_c) = \prod_{i=1}^{n_c} \left(\frac{1}{1 + e^{-(\omega_c^\top f_i + b_c)}} \right)^{y_i} \left(1 - \frac{1}{1 + e^{-(\omega_c^\top f_i + b_c)}} \right)^{1 - y_i} \quad (13)$$

通过最大似然估计, 可以得到参数 ω_c 和 b_c 。样本属于第 c 类的概率表示为

$$p(\text{label} = c | f) = \frac{1}{1 + e^{-(\omega_c^\top f + b_c)}} \quad (14)$$

取概率最大的那一类作为分类结果即可。

$$\text{label}_{\text{pred}} = \underset{i=1, \dots, n_{\text{class}}}{\text{argmax}} p(\text{label} = i | f) \quad (15)$$

2 实验与分析

2.1 数据集介绍

为了验证本文方法的有效性, 在 Indian Pines, Pavia University, Salinas 三个公开数据集上进行了实验, 下面分别对这些数据集进行介绍:

1) Indian Pines: Indian Pines (IP) 数据集是最早的高光谱影像分类测试数据。采用机载可见红外成像光谱仪 (Airborne Visible Infrared Imaging Spectrometer, AVIRIS) 收集数据, 光谱仪波长范围为 $0.4 \sim 2.5 \mu\text{m}$, 以 220 个连续波段对地物进行连续成像, 其中 20 个噪声严重的波段被剔除。该数据集包含 16 种地物, 像素规模为 145×145 , 其中有标签的样本数目为 10 249, 具有典型的样例不均衡的特点, 且由于空间分辨率较低, 为 20 m, 有严重的混合像元现象, 具有一定的难度。

2) Pavia University: Pavia University (UP) 数据集是德国机载反射光学光谱成像系统 (ROSIS-03) 于 2003 年在意大利帕维亚城市获取的高光谱数据的一部分。光谱成像仪在 $0.43 \sim 0.86 \mu\text{m}$ 波长范围内对 115 个波段进行连续成像, 其中 12 个波段由于受噪声影像严重而被剔除, 所得图像的空间分辨率为 1.3 m, 像素规模为 610×340 。该数据集共有 9 类地物, 其中有标签的样本数目为 42 776。

3) Salinas: Salinas (SV) 数据集由 AVIRIS 成像光谱仪采集, 其拍摄地点是加州 Salinas Valley。这个数据集的空间分辨率是 3.7 m, 像素规模为 512×217 , 共有 224 个波段, 剔除部分噪声波段后剩余了 204 个波段。该数据集共有 16 类地物, 其中有标签的样本数目为 54 129。

表 1 展示了上述数据集各类样本的数目。

表1 高光谱遥感数据集样本分布
Table 1 Sample distribution of hyperspectral remote sensing dataset

Indian Pines		Pavia University		Salinas	
Class	Samples	Class	Samples	Class	Samples
Asphalt	46	Asphalt	6 631	Brocoli-green-weeds-1	2 009
Corn-notill	1 428	Meadows	18 649	Brocoli-green-weeds-2	3 726
Corn-mintill	830	Gravel	2 099	Fallow	1 976
Corn	237	Trees-painted	3 064	Fallow-rough-plow	1 394
Grass-pasture	483	Metal-sheets	1 345	Fallow-smooth	2 678
Grass-trees	730	Bare soil	5 029	Stubble	3 959
Grass-pasture-mowed	28	Bitumen	1 330	Celery	3 579
Hay-windrowed	478	Blocking-bricks	3 682	Grapes-untrained	11 271
Oats	20	Shadows	947	Soil-vinyard-develop	6 203
Soybean-notill	972			Corn-senesced-green-weeds	3 278
Soybean-mintill	2 455			Lettuce-romaine-4wk	1 068
Soybean-clean	593			Lettuce-romaine-5wk	1 927
Wheat	205			Lettuce-romaine-6wk	916
Woods	1 265			Lettuce-romaine-7wk	1 070
Building-grass-Trees-drives	386			Vinyard-untrained	7 268
Stone-steel-towers	93			Vinyard-vertical-trellis	1 807
Total	10 249	Total	42 776	Total	54 129

2.2 实验环境与参数设置

实验硬件条件为:处理器为 AMD Ryzen 9 5900HX;内存为 DDR4 3 200 MHz 16 GB×2;显卡为 NVIDIA GeForce RTX 3080,显存为 16 G。软件环境为 windows11 下的 python 3.8+tensorflow 2.3.0+numpy 1.23.5。

在参数设置方面,采用 Adam 优化器的默认参数进行训练;BatchSize 设置为 100;最大训练轮数设置为 150,以使得模型收敛;每训练三轮进行一次验证,仅仅保存损失最小的模型;Dropout 的概率设置为 0.5,以缓解过拟合。训练集、验证集、测试集分别随机取总样本量的 9%,1% 和 90%(部分算法不需要验证集,则将训练集和验证集合二为一)。

2.3 性能指标

本实验采用的性能指标有总体准确率(Overall Accuracy, OA),平均准确率(Average Accuracy, AA),Kappa 系数。下列各式中, c 为类别数, N_i 为第 i 类样本的数目, m_{ij} 表示第 j 类样本被分为 i 类的数量, m_{ij} 各元素组成了一个混淆矩阵。

1) OA 为正确分类的样本数与总样本数 N 的比值,表示为

$$OA = \frac{1}{N} \sum_{i=1}^c m_{ii} \quad (16)$$

2) AA 为各类分类准确率的平均值

$$AA = \frac{1}{c} \sum_{i=1}^c \frac{m_{ii}}{N_i} \quad (17)$$

3) Kappa 系数是基于混淆矩阵的精确度量,其公式为

$$Kappa = \frac{N \sum_{i=1}^c m_{ii} - \sum_{i=1}^c (\sum_{t=1}^c m_{it} \sum_{s=1}^c m_{si})}{N^2 - \sum_{i=1}^c (\sum_{t=1}^c m_{it} \sum_{s=1}^c m_{si})} \quad (18)$$

2.4 分类结果

实验对比了本文方法与支持向量机(Support Vector Machine, SVM)、长短时记忆网络(Long Short-Term Memory, LSTM)、门控循环单元(Gated Recurrent Unit, GRU)、深度语义卷积神经网络(Going

Deeper With Contextual CNN, CDCNN)、双支路双注意力机制网络(Double-Branch Dual-Attention Mechanism-Network, DBDA)、双支路多注意力机制网络(Double-Branch Multi-Attention Mechanism Network, DBMA)等方法。这些对比算法均在文献中被证明在高光谱遥感影像分类任务中有效,采用文献中的默认参数将这些方法与本文方法进行对比实验,证明本文方法的有效性。

对比方法分别为:1) SVM^[30]:基于最大分类间隔的分类器;原文作者在经典SVM的框架下优化了其核函数,并在算法中引入了光谱角以提升分类性能;2) LSTM^[31]:通过输入门、遗忘门和输出门控制网络信息流动的一种循环神经网络,其输出由当前输入和历史输入所决定,相比于传统的循环神经网络,LSTM能够有效地解决长时间序列中的梯度消失和梯度爆炸等问题,由于引入了遗忘门,其能够忽视历史输入中的次要信息,从而获得比传统的循环神经网络更为优良的性能;3) GRU^[31]:通过更新门和重置门控制网络信息流动的一种循环神经网络,其中,更新门有利于捕捉时间序列里的长期依赖关系,而重置门有利于捕捉时间序列里的短期依赖关系;4) CDCNN^[32]:一种包含多尺度滤波和残差单元的2D CNN网络,能利用相邻像素的局部空间光谱关系很好提取各个像素的上下文信息,原文作者引入了多尺度空间滤波器组提取空间特征,与光谱特征进行联合推理,实现空谱联合分类;5) DBDA^[33]:双分支的3D CNN网络,原文作者引入了自己设计的注意力机制,且采用了Mish激活函数代替常用的ReLU激活函数;6) DBMA^[34]:双分支的3D CNN网络,两条支路分别注重提取空间特征和光谱特征,从而缓解两方面特征的相互干扰,且仿照稠密连接网络引入了三维稠密连接模块,并引入了空间注意力机制和通道注意力机制。

实验在IP、UP、SV三个公开数据集上进行,分类结果由表2~表4,图5~图7所示。

表2 IP数据集使用10%有标签样本的分类结果表
Table 2 Classification results for the IP dataset using 10% of the available labeled data

Class	SVM/%	LSTM/%	GRU/%	CDCNN/%	DBDA/%	DBMA/%	Ours/%
Asphalt	73.17	46.34	51.21	51.75	60.00	93.82	95.37
Corn-notill	78.37	81.78	78.52	77.64	95.10	97.03	97.66
Corn-mintill	70.55	71.35	61.98	77.13	96.57	97.11	99.95
Corn	72.30	61.03	72.30	78.92	98.98	98.58	99.23
Grass-pasture	90.80	81.14	88.04	91.86	96.83	97.03	94.93
Grass-trees	90.11	95.89	89.64	95.45	97.60	98.66	99.26
Grass-pasture-mowed	68.00	80.00	68.00	54.20	58.82	80.01	97.40
Hay-windrowed	97.67	99.53	97.67	90.46	96.71	99.90	100.00
Oats	61.11	77.77	50.00	63.27	40.00	81.54	97.22
Soybean-notill	71.31	75.88	73.82	81.34	88.24	95.30	98.37
Soybean-mintill	83.30	74.78	78.81	79.97	96.66	94.06	98.73
Soybean-clean	73.78	30.14	80.71	64.98	90.72	95.49	98.56
Wheat	95.68	97.83	98.37	97.89	100.00	98.46	97.89
Woods	95.43	96.13	94.29	93.99	96.72	97.36	99.73
Building-grass-trees-drivers	63.11	45.53	58.50	83.31	96.56	93.64	98.17
Stone-steel-towers	88.10	88.09	91.66	92.39	93.23	96.89	95.89
OA	82.05	77.65	80.43	81.35	95.07	96.02	98.60
AA	79.55	75.20	77.10	79.66	87.67	94.68	98.02
Kappa	79.47	74.51	77.66	78.67	94.38	95.55	98.41

1) Indian Pines数据集

随机选取Indian Pines数据集中922个样本作为训练集,103个样本作为验证集,9224个样本作为测试集,并进行五次重复实验,得到的分类结果如表2和图5所示,每类最好的分类结果加粗。从图5可以看出,SVM、GRU、LSTM、CDCNN这些方法的分类结果中有较多噪声,分类性能不佳;各类基于3D CNN的方法,如DBDA和DBMA等可以很好地融合空间信息和光谱信息,取得了更好的分类性能。相比于对比方法,本文方法的分类结果图与真值图最为接近。本文提出的方法引入了旋转数据增强,并从多尺度上提

表3 UP数据集使用10%有标签样本的分类结果表
Table 3 Classification results for the UP dataset using 10% of the available labeled data

Class	SVM/%	LSTM/%	GRU/%	CDCNN/%	DBDA/%	DBMA/%	Ours/%
Asphalt	93.37	93.12	92.69	95.48	99.04	99.60	99.91
Meadows	97.81	97.54	97.11	98.77	99.86	99.86	100.00
Gravel	78.00	75.91	79.17	95.73	96.19	99.45	99.08
Trees-painted	95.32	92.93	95.30	98.85	99.01	99.04	99.28
Metal-sheets	99.42	99.63	99.75	99.83	99.93	99.61	99.88
Bare soil	90.37	86.86	88.17	95.25	99.93	99.90	100.00
Bitumen	87.97	83.96	87.05	93.30	100.00	100.00	99.83
Blocking-bricks	88.61	87.24	85.95	89.37	98.27	97.64	99.66
Shadows	99.94	99.88	99.71	99.07	99.08	99.42	99.53
OA	94.09	93.01	93.23	96.57	99.33	99.53	99.83
AA	92.31	90.79	91.65	96.18	99.03	99.39	99.68
Kappa	92.16	90.71	91.02	95.46	99.11	99.38	99.78

表4 SV数据集使用10%有标签样本的分类结果表
Table 4 Classification results for the SV dataset using 10% of the available labeled data

Class	SVM/%	LSTM/%	GRU/%	CDCNN/%	DBDA/%	DBMA/%	Ours/%
Brocoli_g_w_1	99.56	99.69	99.86	79.94	100.00	99.23	100.00
Brocoli_g_w_2	99.94	99.63	99.88	92.88	99.98	99.99	100.00
Fallow	99.49	99.69	99.71	98.60	99.92	98.52	100.00
Fallow_r_p	99.52	99.56	99.60	98.74	99.35	98.84	99.92
Fallow_s	99.59	98.28	99.04	89.00	99.44	99.55	99.96
Stubble	99.92	99.84	99.89	99.87	99.92	100.00	100.00
Celery	99.66	99.78	99.76	99.72	99.92	99.94	100.00
Grapes_u	90.20	87.56	87.11	85.78	96.09	99.80	99.95
Soil_v_d	99.91	99.02	99.67	99.79	99.94	99.84	100.00
Corn_s_g_w	97.69	95.74	97.67	97.62	99.80	98.84	99.97
Lettuce_r_4wk	99.06	98.02	98.59	93.85	100.00	99.19	100.00
Lettuce_r_5wk	99.77	96.86	99.42	99.45	99.97	99.73	100.00
Lettuce_r_6wk	99.39	98.24	99.21	98.79	100.00	99.27	100.00
Lettuce_r_7wk	97.92	96.62	97.82	97.76	99.77	99.74	99.83
Vinyard_u	71.24	82.15	82.55	83.56	97.86	94.59	99.96
Vinyard_v_t	99.20	98.83	98.86	93.05	86.77	99.96	99.80
OA	93.74	94.18	94.55	91.99	97.45	98.89	99.97
AA	97.00	96.85	97.42	94.27	98.67	99.18	99.96
Kappa	93.02	93.53	93.93	91.09	97.16	98.77	99.96

取了特征,能更准确地对高光谱影像进行分类。从表2可以看出,本文方法对于IP数据集16个类别中的13个类别,均取得了最高的准确率,对于Hay-windrowed类别取得了100%的准确率。本文方法的OA,AA,Kappa等性能指标均优于其他对比方法。相比于次优的DBMA方法,本文方法的OA,AA,Kappa指标分别提升了2.69%、3.53%、2.99%。

2) Pavia University数据集

随机选取Pavia University数据集中3850个样本作为训练集,428个样本作为验证集,38498个样本作为测试集,并进行五次重复实验,得到的分类结果如表3和图6所示,其中每类最好的分类结果加粗。从图6可以看出,相比于对比方法,本文方法的分类结果图与真值图最为接近。从表3中可以看出,本文方法对于UP数据集9个类别中的5个类别取得了最高的准确率,对于Meadows类别和Bare Soil类别取得了100%

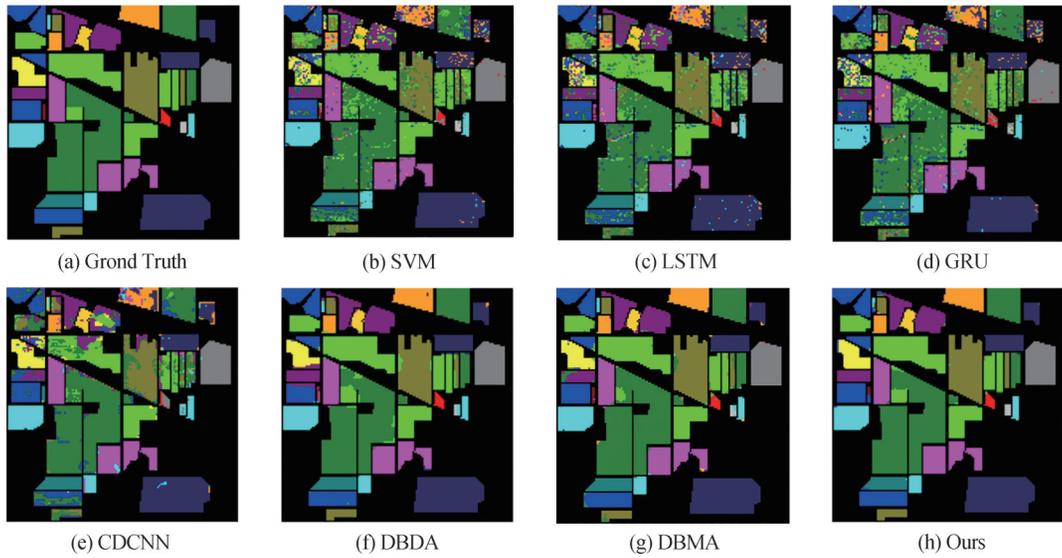


图5 使用IP数据集10%有标签样本的分类结果

Fig.5 Classification results for the IP dataset using 10% of the available labeled data

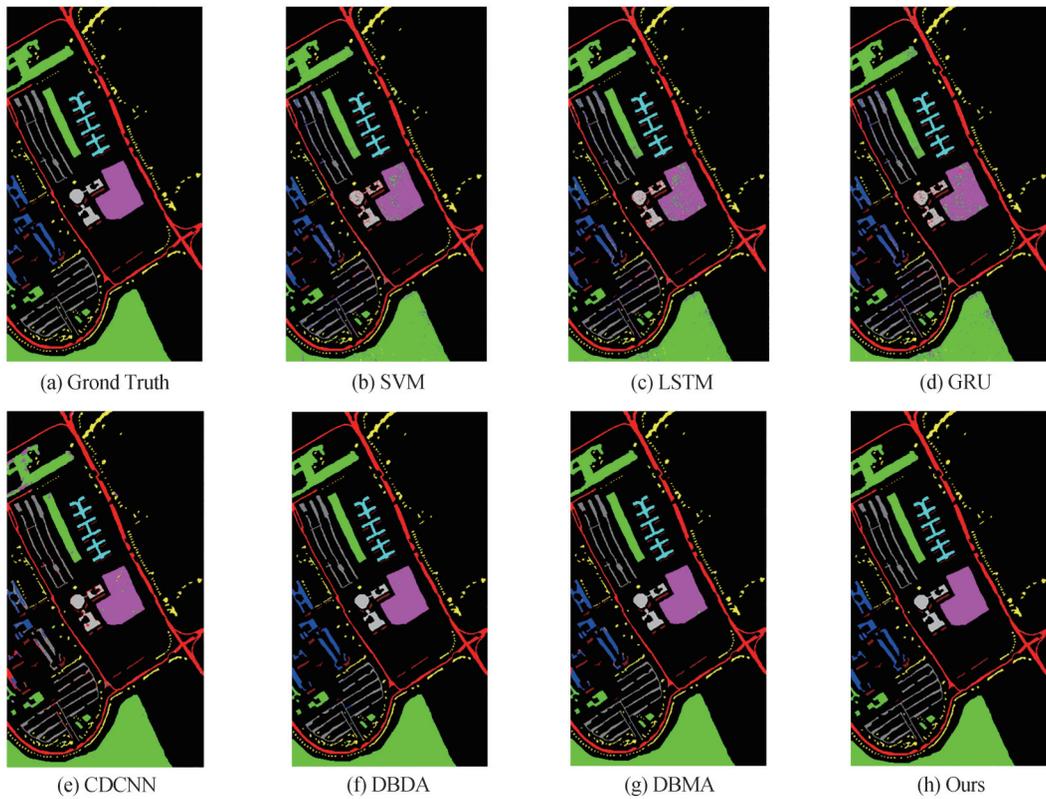


图6 使用UP数据集10%有标签样本的分类结果

Fig.6 Classification results for the UP dataset using 10% of the available labeled data

的准确率。本文方法的OA、AA、Kappa等性能指标均优于其他对比方法。相比于次优的DBMA方法,本文的方法的OA、AA、Kappa指标分别提升了0.30%、0.29%、0.40%。

3) Salinas数据集

随机选取 Salinas 数据集中 4 872 个样本作为训练集, 541 个样本作为验证集, 48 416 个样本作为测试集, 并进行五次重复实验, 得到的分类结果如表 4 和图 7 所示, 其中每类最好的分类结果加粗。从图 7 可以看出, 相比于对比方法, 本文方法的分类结果图与真值图最为接近。从表 4 中可以看出, 本文方法对于 SV 数据集 16 个类别中的 15 个类别, 均取得了最高的准确率, 对于其中 8 个类别取得了 100% 的准确率。本文方法

的 OA、AA、Kappa 等性能指标均优于其他对比方法。相比于次优的 DBMA 方法,本文方法的 OA、AA、Kappa 指标分别提升了 1.09%、0.79%、1.20%。

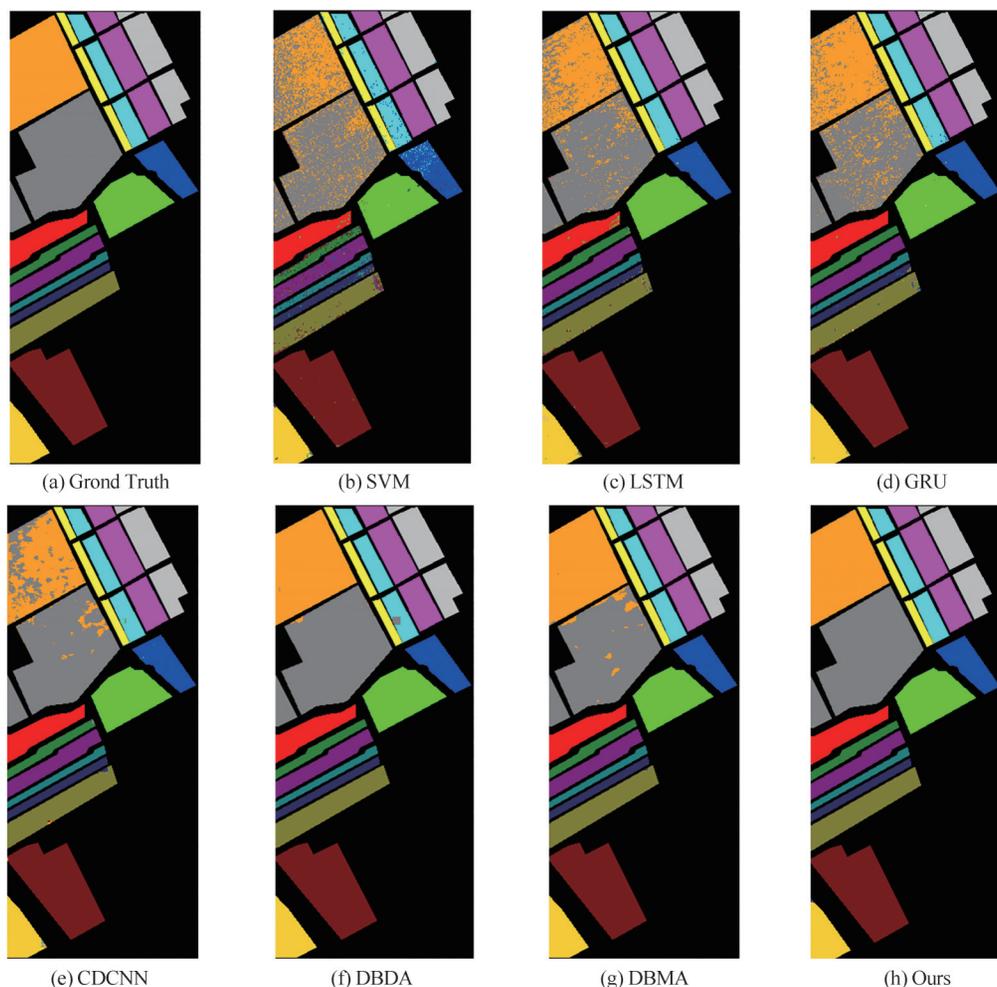


图7 使用SV数据集10%有标签样本的分类结果

Fig.7 Classification results for the SV dataset using 10% of the available labeled data

4) 训练时有标签样本占比与总体准确率的关系

为了进一步探究有标签样本占比与总体准确率的关系,并验证本文方法各个模块的有效性,在 Indian Pines、Pavia University、Salina 数据集上进行了一系列消融实验。消融后方法分别为:多支路无数据增强 3D CNN、单支路有数据增强 3D CNN、单支路无数据增强 3D CNN。IP 数据集上的分类结果如图 8(a)所示,UP 数据集上的分类结果如图 8(b)所示,SV 数据集上的分类结果如图 8(c)所示。横轴表示训练时用到的有标签样本的比例,纵轴表示测试集上的总体准确率(OA)。

从图 8 可以看出,对于 IP、UP、SV 三个数据集,本文提出的方法与消融后方法相比,在训练时采用的有标签样本比例逐渐减少的情形下仍具有更高的总体准确率。即使只采用 2% 的有标签样本进行训练,本文方法在这三个数据集上仍能获得较好的分类结果,而消融后的方法受样本量减少的影响更大。该实验说明了无论是数据增强,还是多支路结构,均能有效提升 3D CNN 对于遥感影像分类问题的准确率。且当两者共同运用时,分类准确率会有进一步地提升。

5) 小样本情形下的对比实验

为了进一步验证本文方法相比于对比方法在小样本情形下的优势,将本文方法与前文实验中较优的 DBDA、DBMA 方法进行进一步对比。采用 2% 的有标签数据对 IP、UP、SV 三个数据集进行分类,其分类结果如表 5 所示。

从表 5 可以看出,在小样本情形下,本文方法比对比方法分类准确率明显更高。这验证了本文方法在小

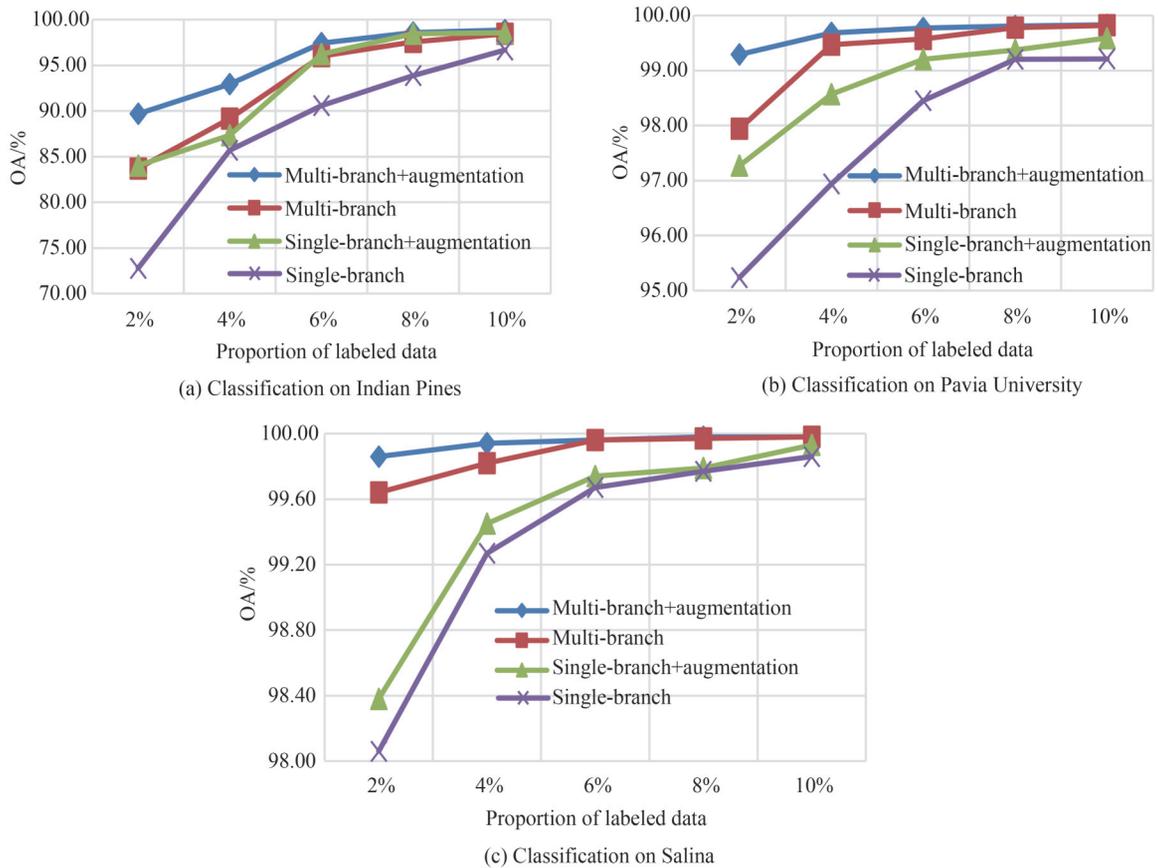


图8 训练时有标签样本比例与分类结果的关系

Fig.8 The relationship between the proportion of the available labeled data and classification results

表5 采用2%有标签样本训练时各算法的分类总体准确率

Table 5 The overall classification accuracy of each algorithm when using 2% labeled samples for training

Algorithm	IP/%	UP/%	SV/%
DBDA	87.29	97.81	96.28
DBMA	78.35	97.85	98.14
OURS	89.68	99.29	99.86

样本情形下,相比于对比方法的优势。

6) 实用性验证

为了验证本文方法的实用性,采用本文方法对无人机机载高光谱相机采集的数据进行了分类。该高光谱相机光谱范围为400~1 000 nm,波段数为176,空间分辨率为0.05 m。成像区域为我国山东烟台沿海区域,共有5类地物。应用本文方法,采用1%的有标签样本进行训练,得到了99.64%的总体准确率。该数据集伪彩色图如图9(a)所示,真值图如图9(b)所示,分类结果如图9(c)所示。从图中可以看出,本文方法的分

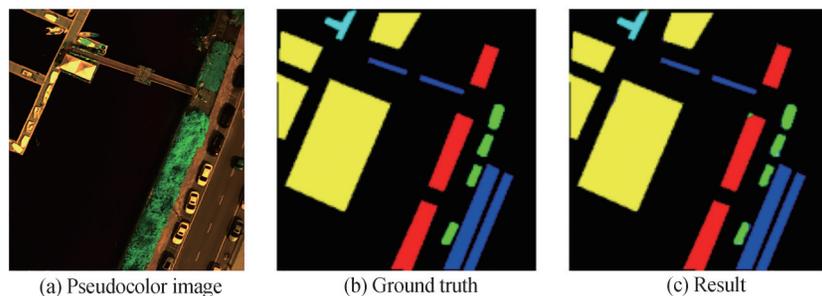


图9 机载高光谱数据集伪彩色图及其分类结果

Fig.9 Airborne hyperspectral dataset and its classification result map

类结果与真值图较为接近,分类错误较少。该实验表明,本文方法具有一定的应用价值。

3 结论

针对普通的3D CNN从一个尺度上提取特征,会丢失部分细节信息,且对于小样本任务表现一般的问题,本文提出了一种三支路的3D CNN,从不同尺度上提取特征后进行加权特征融合,从而获取了更为全面的特征;并引入旋转数据增强技术,从而改善了小样本情形下的分类性能。实验结果表明,从多个尺度上对高光谱数据进行特征提取,可以获得更为丰富的空间信息,提升遥感影像分类的性能指标。采用合适的数据增强技术,也可以提升遥感影像分类的性能指标。本文方法与各类对比方法相比,在各类数据集,各种训练样本比例的情形下均有优势,说明了模型具有较强泛化能力,且能很好地适应小样本情形下的高光谱影像分类问题。

由于本文采用了多个3D CNN支路,导致训练时间和测试时间偏长,今后可以考虑采用各类并行计算方法以缩短训练时间;采用各类模型压缩方法如知识蒸馏、剪枝以缩短测试时间。此外,可以考虑在本文方法中引入一些卷积神经网络的最新成果如可变型卷积、Transformer等,进一步提高分类准确率。

参考文献

- [1] REN X, MALIK J. Learning a classification model for segmentation[C]. IEEE International Conference on Computer Vision, IEEE Computer Society, 2003, 2: 10.
- [2] BENEDIKTSSON J A, PALMASON J A, SVEINSSON J R. Classification of hyperspectral data from urban areas based on extended morphological profiles[J]. IEEE Transactions on Geoscience and Remote Sensing, 2005, 43(3): 480-491.
- [3] CAMPS-VALLS G, GOMEZ-CHOVA L, MUÑOZ-MARÍ J, et al. Composite kernels for hyperspectral image classification[J]. IEEE Geoscience and Remote Sensing Letters, 2006, 3(1): 93-97.
- [4] TARABALKA Y, RANA A. Graph-cut-based model for spectral-spatial classification of hyperspectral images[C]. 2014 IEEE Geoscience and Remote Sensing Symposium, IEEE, 2014: 3418-3421.
- [5] CHEN P H, LIN C J, SCHÖLKOPF B. A tutorial on ν -support vector machines[J]. Applied Stochastic Models in Business and Industry, 2005, 21(2): 111-136.
- [6] CRAMER J S. The origins of logistic regression[M]. SSRN Electronic Journal, 2002.
- [7] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45: 5-32.
- [8] CHEN Y, LIN Z, 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.
- [9] LI T, ZHANG J, ZHANG Y. Classification of hyperspectral image based on deep belief networks[C]. 2014 IEEE International Conference on Image Processing, IEEE, 2014: 5132-5136.
- [10] KALANTARI L, GADER P, GRAVES S, et al. Evaluating similarity measures for hyperspectral classification of tree species at Ordway-Swisher Biological Station[C]. 2014 IEEE Geoscience and Remote Sensing Symposium, 2014: 2691-2694.
- [11] LIU Q, ZHOU F, HANG R, et al. Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification[J]. Remote Sensing, 2017, 9(12): 1330.
- [12] ARUN P V, BUDDHIRAJU K M, PORWAL A. CapsuleNet-based spatial - spectral classifier for hyperspectral images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(6): 1849-1865.
- [13] PAN E, MEI X, WANG Q, et al. Spectral-spatial classification for hyperspectral image based on a single GRU[J]. Neurocomputing, 2020, 387: 150-160.
- [14] HE X, CHEN Y, LIN Z. Spatial-spectral transformer for hyperspectral image classification[J]. Remote Sensing, 2021, 13(3): 498.
- [15] SHI Yuetian, FU Bin, WANG Nan, et al. Spectral-spatial attention rotation-invariant classification network for airborne hyperspectral images[J]. Drones, 2023, 7(4): 1-30.
- [16] KAMNITSAS K, LEDIG C, NEWCOMBE V F J, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation[J]. Medical Image Analysis, 2017, 36: 61-78.
- [17] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks[C]. Computer Vision-ECCV 2016: 14th European Conference, 2016: 630-645.
- [18] ISSITT R W, CORTINA-BORJA M, BRYANt W, et al. Classification performance of neural networks versus logistic regression models: evidence from healthcare practice[J]. Cureus, 2022, 14(2): e22443.
- [19] MA W, GONG C, HU Y, et al. The Hughes phenomenon in hyperspectral classification based on the ground spectrum of grasslands in the region around Qinghai Lake[C]. International Symposium on Photoelectronic Detection and Imaging 2013: Imaging Spectrometer Technologies and Applications, 2013, 8910: 363-373.

- [20] KHODR J, YOUNES R. Dimensionality reduction on hyperspectral images: a comparative review based on artificial datasets [C]. 2011 4th International Congress on Image And Signal Processing, 2011, 4: 1875-1883.
- [21] AHMAD M, SHABBIR S, RAZA R A, et al. Hyperspectral image classification: Artifacts of dimension reduction on hybrid CNN[J]. arXiv preprint arXiv:2101.10532, 2021.
- [22] DU Q, FOWLER J E. Hyperspectral image compression using JPEG2000 and principal component analysis[J]. IEEE Geoscience and Remote Sensing Letters, 2007, 4(2): 201-205.
- [23] LI W, CHEN C, ZHANG M, et al. Data augmentation for hyperspectral image classification with deep CNN[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 16(4): 593-597.
- [24] SANTURKAR S, TSIPRAS D, ILYAS A, et al. How does batch normalization help optimization?[J]. Advances in Neural Information Processing Systems, 2018: 31.
- [25] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [26] ZHANG Z, SABUNCU M. Generalized cross entropy loss for training deep neural networks with noisy labels [J]. Advances in Neural Information Processing Systems, 2018: 31.
- [27] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [28] SEXTON R S, DORSEY R E, JOHNSON J D. Optimization of neural networks: a comparative analysis of the genetic algorithm and simulated annealing[J]. European Journal of Operational Research, 1999, 114(3): 589-601.
- [29] ALDOUS D. On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing[J]. Probability in the Engineering and Informational Sciences, 1987, 1(1): 33-46.
- [30] MERCIER G, LENNON M. Support vector machines for hyperspectral image classification with spectral-based kernels [C]. IEEE International Geoscience and Remote Sensing Symposium. Proceedings, 2003; 288-290.
- [31] PAOLETTI M E, HAUT J M, PLAZA J, et al. Deep learning classifiers for hyperspectral imaging: a review[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 158: 279-317.
- [32] LEE H, KWON H. Going deeper with contextual CNN for hyperspectral image classification[J]. IEEE Transactions on Image Processing, 2017, 26(10): 4843-4855.
- [33] LI R, ZHENG S, DUAN C, et al. Classification of hyperspectral image based on double-branch dual-attention mechanism network[J]. Remote Sensing, 2020, 12(3): 582.
- [34] MA W, YANG Q, WU Y, et al. Double-branch multi-attention mechanism network for hyperspectral image classification [J]. Remote Sensing, 2019, 11(11): 1307.

Multi-scale Remote Sensing Image Classification Based on Weighted Feature Fusion

CHENG Yin Zhu^{1,2}, LIU Song^{1,2}, WANG Nan^{1,2}, SHI Yuetian^{1,2}, ZHANG Geng¹
 (1 *Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China*)
 (2 *University of Chinese Academy of Sciences, Beijing 100049, China*)

Abstract: Remote sensing image classification is a key branch in remote sensing image processing, which provides an important basis for agricultural, industrial, and military applications. With the development of remote sensing satellite, spectral imaging technique has also developed from multi-spectral technology to hyperspectral technology. Rich spectral information puts forward higher requirements for remote sensing image classification. Many hyperspectral image classification algorithms based on traditional methods, such as superpixel methods, extended morphological feature methods, space-spectral joint classification algorithms based on combined kernels, and classification algorithms based on support vector machines and graph cuts, have achieved certain results. In recent years, benefiting from the improvement of hardware conditions and the update and iteration of algorithms, various deep learning methods have emerged one after another, and have been introduced into the field of hyperspectral image classification by researchers, further improving the accuracy of hyperspectral classification. Autoencoder, Convolutional Neural Network (CNN), and capsule neural network have all been experimentally verified to be effective in this field. Different from the common two-dimensional convolutional neural network, the convolution kernel of the three-dimensional convolutional neural network (3D CNN) is a cube, which can naturally integrate the features of spatial dimension and spectral dimension, and has achieved state-of-the-art performance in the

field of remote sensing image classification. Conventional 3D CNN usually extracts data cube features from a single scale, which often loses certain local information; excessively increasing the depth of the model will lead to overfitting problems; limited by actual conditions, it is often difficult to obtain hyperspectral data with a large number of labels, while conventional 3D CNN does not perform well for few-sample situations (for example, the total sample size is only a few hundred). To solve these problems, a multi-branch 3D CNN is proposed in this paper, and the three branches are designed with three different 3D CNN structures. For the input hyperspectral data image set, this paper first utilizes the principal component analysis method to reduce the dimensionality of the data, and the dimensionality of the spectral dimension is selected as 40 after dimensionality reduction. The data cube is decomposed into many $19 \times 19 \times 40$ image patches, and the label of its central pixel is used as the label of the image patch, and then the method of rotating 90 degrees, 180 degrees, and 270 degrees is introduced for data augmentation. In the feature extraction stage, a three-dimensional convolutional neural network connected in parallel with three branches is employed to extract features from three spatial scales of 2×2 , 4×4 , and 6×6 . In the training phase, Adam optimizer is used to optimize the parameters of the three branches, respectively, and the cross-entropy loss function is adopted. In order to alleviate overfitting, the dropout unit and Batch Normalization are introduced. In the test phase, the features extracted from the three branches are combined by weighted connection, and the optimal weighting coefficient is optimized by utilizing simulated annealing algorithm. In terms of classifiers, the logistic regression classifier is adopted, which has performance not inferior to fully connected neural networks for small-sized and medium-sized data sets. In order to verify the effectiveness of the method in this paper, 10% of the labeled data were used for training on public datasets such as Indian Pines, Pavia University, and Salinas, the overall accuracy of 98.60%, 99.83%, and 99.97% were respectively obtained. Our method outperforms the comparative methods such as support vector machine, 2D CNN, and conventional 3D CNN. Moreover, the overall accuracy of the method in this paper is studied when the amount of data in the training phase gradually decreases, and compared with the single-branch 3D CNN with data augmentation, multi-branch 3D CNN without data augmentation, single-branch 3D CNN without data augmentation. The method in this paper is also compared with the DAMA and DBDA methods in the case of a few samples. Compared with the comparison method, the performance of the method in this paper still maintains a high classification accuracy when the sample size is small. In addition, a practical test was carried out using the method in this paper. A series of experimental results show that compared with various comparison methods, the method proposed in this paper has a good classification accuracy and has high application value for hyperspectral image classification problems.

Key words: Optical image sensing images; Deep learning; Image classification; Convolutional neural network; Feature fusion; Multiscale

OCIS Codes: 100.4996; 120.0280; 150.1135; 330.6180