中图法分类号:TP18; TP75　　**文献标识码**: A　　**文章编号**: 1006-8961(2026)03-0927-17

论文引用格式:Yang L X, Bao Y J, Zhang R and Yang S Y. 2026. Lightweight Res-3D-CNN with Transformer embedding for cross-domain hyperspectral image classification. Journal of Image and Graphics, 31(3):0927-0943(杨丽霞, 鲍雅君, 张瑞, 杨淑媛. 2026. 面向跨域高光谱图像分类的嵌入Transformer层的轻量型Res-3D-CNN. 中国图象图形学报, 31(3):0927-0943)[DOI:10.11834/jig.250020]

面向跨域高光谱图像分类的嵌入Transformer层的轻量型Res-3D-CNN

杨丽霞[1,2],鲍雅君[1],张瑞[1,2*],杨淑媛[3]

1. 宁夏大学数学统计学院,银川　750021; 2. 宁夏数学基础学科研究中心,银川　750021;
3. 西安电子科技大学人工智能学院,西安　710071

摘　要:**目的**　跨域分类是高光谱图像分类的主要挑战之一。结合域适应与少样本学习的跨域少样本学习(cross domain few shot learning,CDFSL)方法已广泛应用于跨域高光谱图像分类(cross domain hyperspectral image classification,CD-HIC)问题。由于光谱序列编码的复杂度和类间的光谱相似性,现有的CDFSL方法大多使用卷积神经网络(convolutional neural network, CNN)或其他优秀的空间特征提取器来获取空间信息,以提高分类精度。然而,提取空间特征通常伴随着地面物体分布和类别边界的扭曲。为解决该问题,本文提出了嵌入Transformer层的轻量型Res-3D-CNN(lightweight Res-3D-CNN with Transformer layer embedding,LRCT)作为CD-HIC的特征提取器。LRCT能在提取空间信息的同时获取光谱的长期依赖性,从而显著提高光谱特征方法的判别性能。**方法**　CNN中的卷积(Conv)通过局部感受野的权重共享机制捕捉图像高频特征。而Transformer可通过自注意力机制建模特征间的长程依赖关系,并自适应聚焦关键区域。此外,Transformer表现出低通滤波特性,主要捕获图像的低频全局信息。基于Conv和Transformer的互补特性,将Transformer层嵌入Res-3D-CNN构建轻量型双流特征提取网络,分别对源域和目标域进行特征提取,通过CDFSL框架迁移源域通用特征,实现目标域少样本场景下的高精度分类。**结果**　以Chikusei数据为源域, Indian Pines、Salinas和Pavia University为目标域进行验证。实验结果表明,在每类仅有5个标记样本时,目标域上的总体精度分别达到71.01%、92.06%和84.14%。相较于主流的CDFSL方法,基于LRCT网络的CDFSL(LRCT network based CDFSL, LRCT-CDFSL)方法在各个目标域上均展现出更优的分类性能。**结论**　LRCT-CDFSL结合了残差三维卷积神经网络(residual 3-dimensionl CNN,Res-3D-CNN)、Transformer网络、域适应和少样本学习方法的优势,使CD-HIC精度提升。

关键词:高光谱图像分类(HIC);跨域分类;少样本学习(FSL);域适应;残差三维卷积神经网络(Res-3D-CNN);Transformer

Lightweight Res-3D-CNN with Transformer embedding for cross-domain hyperspectral image classification

Yang Lixia[1,2], Bao Yajun[1], Zhang Rui[1,2*], Yang Shuyuan[3]

1. School of Mathematics and Statistics, Ningxia University, Yinchuan　750021, China; 2. Ningxia Basic Science Research Center of Mathematics, Yinchuan　750021, China; 3. School of Artificial Intelligence, Xidian University, Xi'an　710071, China

收稿日期:2025-01-14;**修回日期**:2025-06-13;**预印本日期**:2025-09-23

*通信作者:张瑞　guanyue_002@163.com

基金项目:国家自然科学基金项目(62466047,62361051);宁夏自然科学基金项目(2024AAC03007)

Supported by: National Natural Science Foundation of China (62466047, 62361051); Natural Science Foundation of Ningxia, China (2024AAC03007)

Abstract: Objective Benefiting from their high-resolution spectral information and large-scale spatial information, hyperspectral images (HSIs) have demonstrated exceptional capabilities in numerous remote sensing applications. Over the past few decades, hyperspectral image classification (HIC) has attracted considerable research attention. Many machine learning-based HIC methods have been proposed; however, these approaches require a sufficient number of labeled samples to achieve ideal classification accuracy. Unfortunately, the high cost and effort associated with labeling HSIs often results in a scarcity of labeled data for many newly acquired HSIs. Therefore, researchers have introduced the cross-domain HIC (CD-HIC) mechanism, which uses a hyperspectral image with sufficient labeled samples (source domain) to assist in classifying a hyperspectral image with limited labeled samples (target domain). However, cross-domain classification is one of the major challenges in HIC due to feature and class distribution differences between source and target domains. The cross-domain few shot learning (CDFSL) methods, which integrated domain adaptation with few-shot learning, have been widely applied to the CD-HIC problem. Owing to the difficulty of spectral sequence encoding and the spectral similarity between classes, most existing CDFSL methods use convolutional neural network (CNN) or other remarkable spatial feature extractors to obtain spatial information, thereby improving classification accuracy. However, extracting spatial features often leads to distortion in the distribution of ground objects and their class boundaries. Aiming to address this issue, a lightweight Res-3D-CNN with embedded Transformer layers (LRCT) has been designed for feature extraction in CD-HIC. LRCT effectively captures long-term dependencies of the spectrum while simultaneously extracting spatial information, thereby notably improving the performance of spectral feature-based methods. **Method** In this study, a simple and effective deep learning network is proposed for feature extraction from HSIs. In CNNs, the convolution (Conv) captures high-frequency features of images by employing a weight-sharing mechanism within local receptive fields. In contrast, Transformers model long-range dependencies between features using self-attention mechanisms and adaptively focus on key areas. Moreover, the Transformer exhibits low-pass filtering characteristics, which primarily captures the low-frequency global information of images. Considering the complementary characteristics of Conv and Transformer, the Transformer layer is embedded into Res-3D-CNN to establish a lightweight dual-stream feature extraction network to perform feature extraction on the source and target domains. Furthermore, the CDFSL method is adopted to learn general information from the extracted features of source class data, which helps in target class data prediction with only very few or no labeled data. This approach helps achieve outstanding classification performance in the subsequent CD-HIC. The LRCT-CDFSL method comprises four main aspects as follows: 1) Data preprocessing: using a mapping layer to combine the dimensions of the original his; 2) Feature learning based on lRCT: employing a deep neural network for feature learning to enhance the representation capability for HSI data; 3) Few-shot learning: enhancing intra-class compactness and inter-class separability by calculating the Euclidean distance between labeled and unlabeled samples, thereby effectively adapting to scenes with limited samples; 4) Domain adaptation: using domain adaptation techniques to enable the feature extractor to generate highly generalized features, thereby improving the generalization capability of the model across different domains. **Result** Using Chikusei data as the source domain, and Indian Pines, Salinas, and Pavia University data as the target domains, extensive experiments are conducted to validate the performance of the proposed LRCT-CDFSL model and ensure fair comparison with advanced methods. Moreover, the effectiveness of the proposed LRCT-CDFSL is respectively tested by using the Indian Pines, Salinas, and Pavia University datasets as the target domain. The experimental results demonstrate that the LRCT-CDFSL method achieves faster and more accurate classification performance compared to existing methods. When only five labeled samples per class are available, the LRCT-CDFSL method achieves overall accuracy (OA) scores of 71.01% and 92.06%, and 84.14% on the respective target domain datasets. Compared to current mainstream cross-domain few-shot HIC methods, the LRCT-CDFSL method shows superior classification performance across various target domain datasets. Specifically, LRCT-CDFSL improves OA by 7.57%, 3.35%, and 2.77%, respectively, and reduces training time by 36%, 37%, and 30%, respectively. **Conclusion** A deep transfer learning network called LRCT network is introduced by embedding Transformer layer into a residual three-dimensional CNN (Res-3D-CNN). In Res-3D-CNN, a 1×1 convolution kernel is incorporated to adjust the number of channels in the feature map, reduce the model parameters, and accelerate the training. In addition, after $n \times n$ convolutional kernels, 1×1 convolutional kernels are then introduced to expand the number of channels of the feature map, thereby addressing the problem of reducing the feature map area

caused by the convolution kernel. Additionally, the network shows excellent performance in few-shot learning and domain adaptation using convolutional kernels of different scales for feature extraction. Meanwhile, the Transformer layer is used to extract the local-global semantic information of HSI. Experimental results reveal that LRCT effectively captures spatial-spectral features through a combination of local-global information, fully representing the local-global semantic information of HSIs and enabling strong classification performance in the subsequent CD-HIC task. However, within the LRCT-CDFSL framework, the metric-based few-shot learning method, which emphasizes the relationships between samples, has not yet been fully explored. Aiming to further enhance performance in CD-HIC tasks, future studies may explore the integration of cross-attention learning techniques into the LRCT-CDFSL architecture. This enhancement is expected to improve the generalization capability and adaptability of the model across diverse domains.

Key words: hyperspectral image classification (HIC); cross-domain classification; few-shot learning (FSL); domain adaptation; residual 3-dimensional convolutional neural network (Res-3D-CNN); Transformer

0 引　言

高光谱图像分类(hyperspectral image classification, HIC)是高光谱图像处理中的核心任务之一(Chen等, 2015; Zhao和Du, 2016)。传统的HIC方法通过特征提取、特征选择获取高光谱图像(hyperspectral image, HSI)中的空间特征和光谱特征,然后用K最近邻算法(Cover和Hart, 1967)、最大似然方法(Fisher等, 1922)、支持向量机(Chandra和Bedi, 2021)等合适的分类算法实现对地物的准确分类。尽管这些方法在一定程度上取得了一些成效,但无法做到端到端地学习,因此无法提取更高级、更抽象的深层特征,在处理复杂的HSI时难以获得令人满意的分类精度。因此,如何有效地挖掘这些图像中的光谱和空间信息以更全面地捕捉图像特征成为一个关键问题。

深度学习(deep learning, DL)凭借卓越的特征提取能力,在语音识别、语言处理和图像识别等领域取得了显著成果(Marmanis等, 2016)。例如,He等人(2017)提出了一种多尺度三维卷积神经网络(convolutional neural network, CNN),联合利用HSI的二维多尺度空间特征和一维光谱特征在HIC中取得了良好的效果。然而,随着网络的深入,会出现梯度消失和梯度爆炸等问题。为了解决这些问题,Zhong等人(2018)将残差结构引入光谱模块和空间模块,提出了一种空—谱残差网络(residual network, ResNet)。ResNet允许网络学习残差(跳跃连接),即输入到输出之间的差异,而不是整个映射函数,从而更容易地训练深层网络。然而,由于CNN、ResNet等网络主干的限制,这些方法很难捕获光谱特征的序列属性。

Transformer依赖于注意力机制来捕捉序列数据中元素之间的相关性,从而能够捕获数据中的复杂关系和依赖关系,不仅可以捕获光谱特征的序列属性,而且克服了传统模型在处理长距离依赖和并行计算方面的局限性(Vaswani等, 2017; 游雪儿 等, 2024; 胡帅 等, 2024)。由于其结构简单、易于理解和实现,以及强大的泛化能力,一些研究人员将Transformer应用于HIC。例如,Hong等人(2022)提出了一种基于纯视觉Transformer(vision Transformer, ViT)的框架,该框架使用分组谱嵌入模块学习局部详细的光谱表示。虽然Transformer在捕获光谱特征方面表现出色,但它失去了捕获局部语义特征的能力,并且对HSI空间信息的利用不足。因此,Sun等人(2022)将Transformer与CNN相结合,提出了一种空谱特征标记化Transformer。类似地,Zhong等人(2022)提出了一个空谱Transformer网络,使用堆叠混合CNN提取HSI的空谱特征。然后,用高斯分布加权标记模块使特征与原始样本保持一致,有利于Transformer学习光谱信息。总之,CNN和Transformer结合起来有利于特征的充分提取,在HIC任务中具有潜在的优势。

尽管DL在HSI领域取得了显著成果,然而训练这类模型通常需要大量有标签数据(Tun等, 2021; Zheng等, 2017)。少样本学习方法(few shot learning, FSL)能够在相对较少的标注样本情况下,实现比DL更好的学习性能(Zeng等, 2022),但它们通常假设源类和目标类的数据来自分布相同的域。然而,实际情况中,源类和目标类数据可能来自分布差异显著的域。为了提高模型的泛化性能,使从源域训练的模型同时适应新类别和新域,研究者通常将

基于元学习的FSL方法与域适应方法(Wang等,2022)相结合,旨在在元学习训练(Garcia和Bruna,2018)的过程中尽可能地减少两域数据分布差异(王浩宇 等,2023)。例如,Li等人(2022)提出了跨域少样本学习(cross domain few shot learning,CDFSL)方法,首次在一个统一的框架中解决了FSL和域适应问题。通过采用条件对抗性域适应策略,成功地克服了域漂移问题,实现了域分布的对齐。Zhang等人(2024)则通过基于图神经网络(graph neural network,GNN)的信息传播和图对齐策略,将图信息聚合CDFSL与基于图信息聚合的域对齐相结合,克服了Li等人(2022)方法中仅依赖局部空间信息进行域对齐的局限性。Liu等人(2024)在Li等人(2022)的基础上提出了多视图关系学习,采用对比学习方法学习类级样本关系,获得更具可判别性的样本特征,使得样本之间的关系得到充分的探索和利用。

以上方法使用残差3维卷积神经网络(residual 3-dimensionl CNN,Res-3D-CNN)提取源域和目标域的空—谱特征,尽管它们能有效捕捉图像的三维特征,从而提高分类精度,但Res-3D-CNN参数较多导致模型训练时间较长,在处理三维数据方面需要大量计算资源。此外,Res-3D-CNN存在不能处理序列数据和捕获长距离依赖关系等问题。为了更有效地应对这些挑战,本文基于Res-3D-CNN和Transformer的互补特性,设计了一种嵌入Transformer层的轻量型Res-3D-CNN(lightweight Res-3D-CNN with Transformer layer embedding, LRCT)网络,分别对源域和目标域进行全特征提取,在获取空间信息的同时获取光谱的长期依赖性,从而显著提高了光谱特征方法的判别性能。此外,本文采用CDFSL方法从源类数据提取的特征中学习通用信息,以助于对只有少量或者没有标记的目标类数据进行预测,从而提出了基于LRCT网络的CDFSL模型(LRCT network based CDFSL, LRCT-CDFSL),应用于跨域高光谱图像分类(cross domain hyperspect image classification, CD-HIC)问题。

本文的主要贡献如下:1)提出了一种轻量型Res-3D-CNN网络。通过引入1×1卷积核,以调整特征图通道数,减小模型参数,提高训练速度。同时,在卷积核$n\times n$后引入卷积核1×1以扩充特征图的通道数,弥补由于卷积核导致的特征图面积减小的问题。2)将Transformer层嵌入Res-3D-CNN,设计了一种LRCT网络,提取到具有局部—全局联合的空间—光谱特征,该特征能够充分表达HSI的局部—全局语义信息,从而在后续的CD-HIC中获得出色的分类性能。3)真实高光谱数据上的实验表明,本文提出的网络能有效减少模型的参数,减轻模型的计算负担。

1 基于LRCT-CDFSL的高光谱图像分类

1.1 基本思路和总体分析

LRCT-CDFSL旨在解决标记数据稀缺且源于不同领域的场景中少样本学习的挑战性问题。LRCT-CDFSL的训练过程如图1所示,主要分为4个阶段。

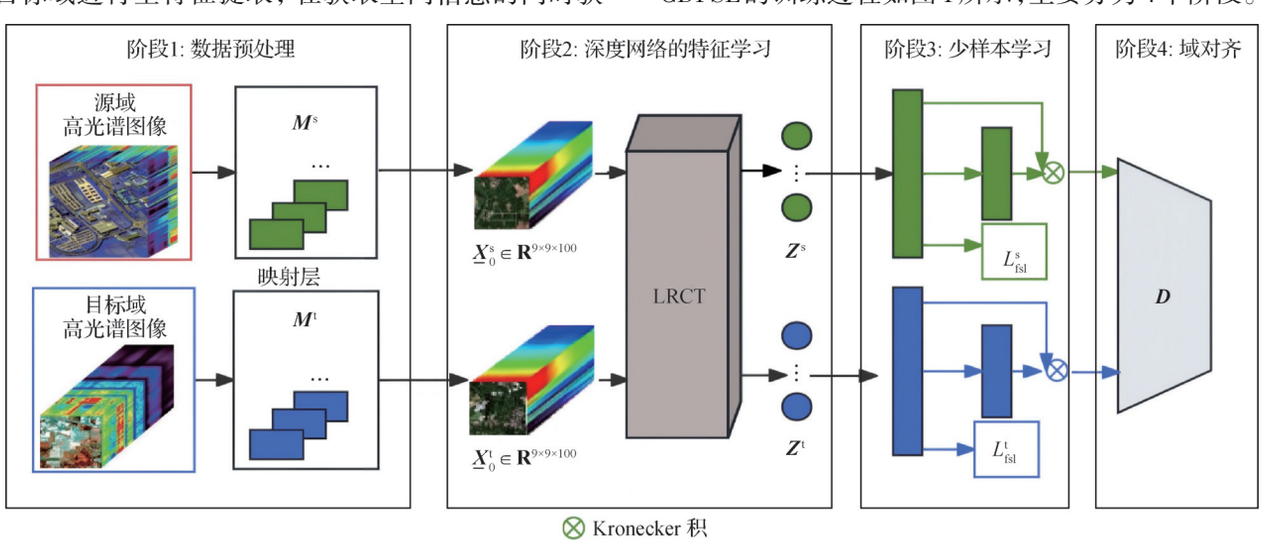


图1 LRCT-CDFSL模型

Fig. 1 LRCT-CDFSL model

首先，利用映射层 $\boldsymbol{M}^{s}$ 和 $\boldsymbol{M}^{t}$ 分别对源域和目标域原始 HSI 进行预处理，得到维度统一的两域输入。其次，将统一维度的样本输入到深度特征提取器，LRCT 分别对源域和目标域进行全特征提取，学习到表示性和判别性强的高级特征，使相同类别的样本尽可能接近，而不同类别的样本尽可能远离。再次，分别基于源域和目标域数据学得的特征，执行源域和目标域的 FSL，分别得到源域小样本分类损失 L_{fsl}^{s} 和目标域小样本分类损失 L_{fsl}^{t}。最后，借助条件域鉴别器减缓域漂移，实现域分布的对齐，并计算得出域判别损失 L_d。因此，具有域适应的源域 FSL 的损失函数为

$$L_{\mathrm{total}}^{s} = L_{\mathrm{fsl}}^{s} + L_d \tag{1}$$

同样，具有域适应的目标域 FSL 的损失函数为

$$L_{\mathrm{total}}^{t} = L_{\mathrm{fsl}}^{t} + L_d \tag{2}$$

最后，给出了 LRCT-CDFSL 的总损失函数为

$$L_{\mathrm{total}} = L_{\mathrm{total}}^{s} + L_{\mathrm{total}}^{t} \tag{3}$$

训练时，式(1)和式(2)交替更新，直至模型收敛。

1.1.1　映射层

首先利用映射层 $\boldsymbol{M}^{s}$ 和 $\boldsymbol{M}^{t}$ 使来自源域和目标域的样本的输入维数相等，然后进行特征提取。映射层通过一个 2D-CNN 实现。设 $\underline{\boldsymbol{I}} \in \mathbf{R}^{n \times n \times B}$ 为映射层的输入 HSI 立方体，其中 $n \times n$ 为空间维，B 是谱带数目。映射层的特征输出为

$$\underline{\boldsymbol{I}}' = \underline{\boldsymbol{I}} \times_3 \boldsymbol{T} \tag{4}$$

式中，张量 $\underline{\boldsymbol{I}} \in \mathbf{R}^{n \times n \times B}$ 与矩阵 $\boldsymbol{T} \in \mathbf{R}^{100 \times B}$ 作模态-3 乘法($\times_3$)，乘积 $\underline{\boldsymbol{I}}' \in \mathbf{R}^{n \times n \times 100}$ 为变换后的数据集。

1.1.2　特征提取模块

在 CD-HIC 中 Res-3D-CNN 具备一种独特的优势，即能够捕获 HSI 的深度空间—光谱特征。但是 Res-3D-CNN 在表示序列数据时存在一定的局限性。因此，本文提出了将 Transformer 层嵌入轻量型 Res-3D-CNN 的 LRCT 特征提取器 F，如图 2 所示，其中，图的下方是 Transformer 的模块。特征提取器 F 由一个残差块 block、一个 Transformer 模块和 Maxpooling 组成。残差块主要包括 2 个 1 × 1 的卷积层、1 个 $n \times n$ 卷积层和 1 个修正线性单元(Rectified Linear Unit，ReLU)激活层，如图 3 所示。

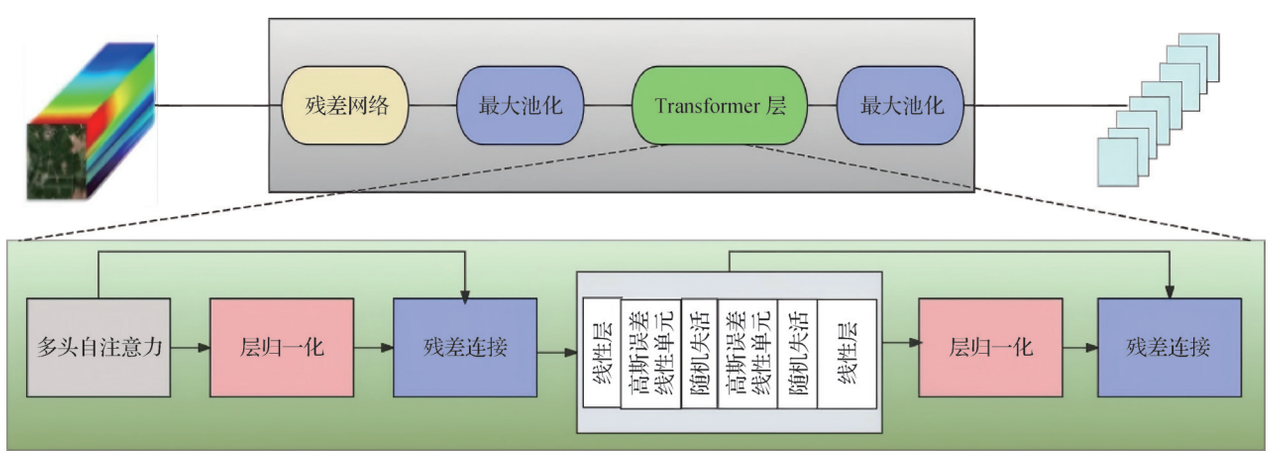


图 2　LRCT 的简要示意图

Fig. 2　LRCT model

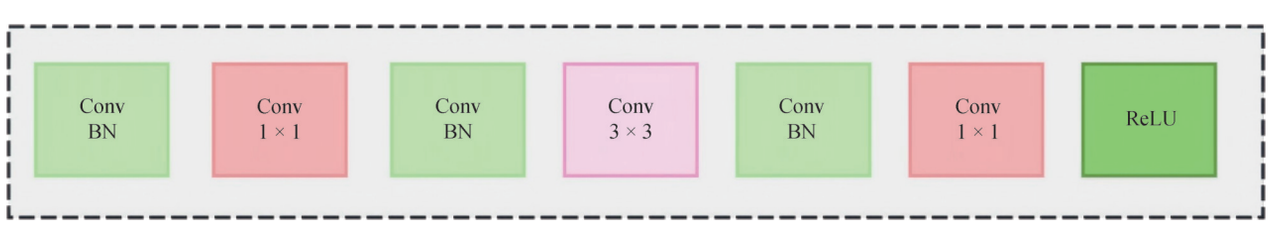


图 3　残差网络的结构

Fig. 3　Structure of ResNet

以源域数据为例，将样本 $\underline{\boldsymbol{X}}_j \in \mathbf{R}^{n\times n\times b}$ 经过映射层 $\boldsymbol{M}^{\mathrm{s}}$、残差块Block和Maxpool，可得

$$\tilde{\underline{\boldsymbol{X}}}_j = f_{\mathrm{maxpool}}\left(f_{\mathrm{Block}}\left(\boldsymbol{M}^{\mathrm{s}}\left(\underline{\boldsymbol{X}}_j\right)\right)\right) \tag{5}$$

式中，$\tilde{\underline{\boldsymbol{X}}}_j \in \mathbf{R}^{b'\times n'\times n'}$，其中，$b'$为通道数，$n'\times n'$为空间大小。

在输入Transformer模块之前，先将 $\tilde{\underline{\boldsymbol{X}}}_j \in \mathbf{R}^{b'\times n'\times n'}$ 展开成 $\hat{\boldsymbol{X}}_j \in \mathbf{R}^{b'\times (n'\cdot n')}$，然后通过以下过程实现多头自注意力（multi-head self-attention，MHSA）机制，提取样本 $\underline{\boldsymbol{X}}_j \in \mathbf{R}^{n\times n\times b}$ 具有局部—全局联合的空谱特征。

1）将 $\hat{\boldsymbol{X}}_j$ 通过线性层向 h 个不同的方向投影，为 h 个注意力头各自生成不同的查询 $\boldsymbol{Q}$（query）、键 $\boldsymbol{K}$（key）和值 $\boldsymbol{V}$（value）矩阵。具体为

$$\begin{cases}\boldsymbol{Q}_{ij} = \boldsymbol{X}\boldsymbol{W}_{ij}^{Q}\\ \boldsymbol{K}_{ij} = \boldsymbol{X}\boldsymbol{W}_{ij}^{K}\\ \boldsymbol{V}_{ij} = \boldsymbol{X}\boldsymbol{W}_{ij}^{V}\end{cases} \quad i = 1,2,\cdots,h \tag{6}$$

式中，$\boldsymbol{W}_{ij}^{Q}, \boldsymbol{W}_{ij}^{K}, \boldsymbol{W}_{ij}^{V}(i=1,2,\cdots,h)$ 是投影参数矩阵。

2）对每个头独立计算自注意力值，具体为

$$head_{ij} = f_{\mathrm{Attention}}(\boldsymbol{Q}_{ij},\boldsymbol{K}_{ij},\boldsymbol{V}_{ij}) = f_{\mathrm{softmax}}\left(\frac{\boldsymbol{Q}_{ij}\boldsymbol{K}_{ij}^{\mathrm{T}}}{\sqrt{d_{\boldsymbol{K}_{ij}}}}\right)\boldsymbol{V}_{ij} \quad i = 1,2,\cdots,h \tag{7}$$

式中，矩阵 $\boldsymbol{Q}_{ij}$ 和 $\boldsymbol{K}_{ij}$ 用于计算注意力权重，而 $\boldsymbol{V}_{ij}$ 矩阵计算加权平均后的输出，伸缩参数 $d_{\boldsymbol{K}_{ij}}$ 表示组成 $\boldsymbol{K}_{ij}$ 的键向量的维数。

3）用Concat对每个注意力头的结果进行拼接，然后将拼接后的结果与投影矩阵 $\boldsymbol{W}_j^0$ 进行相乘，从而得到多头自注意力计算结果。具体为

$$MHSA(\boldsymbol{Q}_{ij},\boldsymbol{K}_{ij},\boldsymbol{V}_{ij}) = f_{\mathrm{Concat}}(head_{1j},\cdots,head_{hj})\boldsymbol{W}_j^0 \tag{8}$$

4）用dropout层随机丢弃一定比例的神经元，以减少过拟合风险。然后，经过Maxpool进行降维得到最终的特征提取器输出的结果。具体为

$$\boldsymbol{z}_j = f_{\mathrm{Maxpool}}(f_{\mathrm{dropout}}(MHSA(\boldsymbol{Q}_{ij},\boldsymbol{K}_{ij},\boldsymbol{V}_{ij}))) \tag{9}$$

式中，$\boldsymbol{z}_j$ 代表源域样本经过具有参数 φ 的特征提取器 F_φ 提取的具有局部—全局联合的空谱特征。

用同样的方法可以得到目标域样本经过特征提取器 F_φ 提取的具有局部—全局联合的空谱特征。

1.1.3 少样本学习模块

FSL的目标是利用以前的经验知识，在有很少标记数据的情况下解决新任务。它通过利用包含大量标记样本的可用数据集，学习出具有良好泛化能力、能够快速适应新任务的模型（Ye等，2023）。

该方法同时在源域和目标域中交替执行FSL，以挖掘源类的可迁移知识并学习目标类的判别嵌入模型。以源域FSL为例，从 C_{s} 个源类中随机抽取 C 个类组成一个集，然后从所选类中选取 K 个样本作为支持集 $\boldsymbol{S}_{\mathrm{s}} = \left\{\left(\boldsymbol{x}_l, y_l\right)\right\}_{l=1}^{C\times K}$，这称为 C-way K-shot 任务。同时，从相同类中选取 N 个样本作为查询集 $\boldsymbol{Q}_{\mathrm{s}} = \left\{\left(\boldsymbol{x}_u, y_u\right)\right\}_{u=1}^{C\times N}$。利用基于度量的元学习方法——原型网络完成小样本分类任务。在少样本训练阶段，采用基于距离度量结果的非参数softmax优化网络参数。以源域为例，查询集 $\boldsymbol{Q}_{\mathrm{s}} = \left\{\left(\boldsymbol{x}_u, y_u\right)\right\}_{u=1}^{C\times N}$ 中查询样本 $\boldsymbol{x}_u$ 的类分布为

$$P(y_u = k \mid \boldsymbol{x}_u \in \boldsymbol{Q}_{\mathrm{s}}) = \frac{\mathrm{e}^{-d(\boldsymbol{z}_u^{\mathrm{s}},\boldsymbol{c}_k)}}{\sum_{k=1}^{C}\mathrm{e}^{-d(\boldsymbol{z}_u^{\mathrm{s}},\boldsymbol{c}_k)}} \tag{10}$$

式中，$d(\cdot)$ 为欧氏距离，$\boldsymbol{c}_k$ 为支持集中第 k 类的嵌入特征的均值，C 为每个episode中类的个数，$\boldsymbol{x}_u$ 为查询集的样本，y_u 为 $\boldsymbol{x}_u$ 的标签，$\boldsymbol{z}_u^{\mathrm{s}}$ 是查询集样本经过特征提取器 F_φ 提取的具有局部—全局联合的空谱特征。

根据查询样本 $\boldsymbol{x}$ 的负对数概率及其真实类标号 k，源域集和目标集中所有查询样本的分类损失可表示为

$$L_{\mathrm{fsl}}^{\mathrm{s}} = E_{S_{\mathrm{s}},Q_{\mathrm{s}}}\left[-\sum_{(\boldsymbol{x},y)\in Q_{\mathrm{s}}}\log P_\varphi(y=k\mid\boldsymbol{x})\right] \tag{11}$$

$$L_{\mathrm{fsl}}^{\mathrm{t}} = E_{S_{\mathrm{t}},Q_{\mathrm{t}}}\left[-\sum_{(\boldsymbol{x},y)\in Q_{\mathrm{t}}}\log P_\varphi(y=k\mid\boldsymbol{x})\right] \tag{12}$$

式中，E 为数学期望，$\boldsymbol{S}_{\mathrm{s}}$ 和 $\boldsymbol{Q}_{\mathrm{s}}$、$\boldsymbol{S}_{\mathrm{t}}$ 和 $\boldsymbol{Q}_{\mathrm{t}}$ 分别为源域、目标域的少样本数据集的支持集和查询集。

1.1.4 域适应模块

域判别是将一个域的特征与另一个域的特征区分开来，并由此产生判别器损失 L_d，其目的是使特征提取器能够产生更广义的特征。为了减少源分布 $P_{\mathrm{s}}(\boldsymbol{x})$ 和目标分布 $P_{\mathrm{t}}(\boldsymbol{x})$ 之间的域偏移，Li等人（2022）在DCFSL中提出一种条件域鉴别器。本文工作沿用其域鉴别器 D，采用条件对抗域的适应策略来对齐源和目标域的全局数据分布。

分类器的预测结果 g 包含判别性信息，可将其与特征表示 f 的对抗性适应过程相结合，假设 f 与 g 的维度分别为 d_f 和 d_g。DCFSL采用随机化多线性映射替代原始映射，多线性映射 $T_\otimes(f,g)$ 可通过点积

$T_{\odot}(f,g)=\frac{1}{\sqrt{d}}(\boldsymbol{R}_f f)\odot(\boldsymbol{R}_g g)$ 近似，其中，$\odot$ 表示逐元素乘积，$\boldsymbol{R}_f\in\mathbf{R}^{d\times d_f}$ 和 $\boldsymbol{R}_g\in\mathbf{R}^{d\times d_g}$ 为随机矩阵（训练时仅采样一次并固定），且 $d\ll d_f\times d_g$。矩阵 $\boldsymbol{R}_f$ 和 $\boldsymbol{R}_g$ 的元素服从对称单峰分布（如均匀分布或高斯分布）。最终采用以下条件化策略，具体为

$$T(h)=\begin{cases}T_{\otimes}(f,g) & d_f\times d_g\leqslant 1\,024\\ T_{\odot}(f,g) & \text{其他}\end{cases}\tag{13}$$

式中，1 024 为深度网络中单元数的上限。若多线性映射 $T_{\otimes}$ 的维度超过 1 024，则使用随机多线性映射 $T_{\odot}$。令 $h=(f,g)$ 为 f 和 g 的联合变量；则学习域不变嵌入特征的损失函数定义为

$$L_d=\min_D\max_T L=-E_{\boldsymbol{x}_i^{\mathrm{s}}\sim P_{\mathrm{s}}(\boldsymbol{x})}\log\left[D(T(h_i^{\mathrm{s}}))\right]-E_{\boldsymbol{x}_j^{\mathrm{t}}:P_{\mathrm{t}}(\boldsymbol{x})}\log\left[1-D(T(h_j^{\mathrm{t}}))\right]\tag{14}$$

域判别器的网络结构如图 4 所示。本文采用多层感知机（multi-layer perception，MLP）堆叠在多线性映射之后，包含 5 个全连接（fully connection，FC）层。除最后一层外，每层后均添加 ReLU 非线性激活函数和 dropout 层，最终通过 softmax 函数预测输入数据来自源域或目标域。

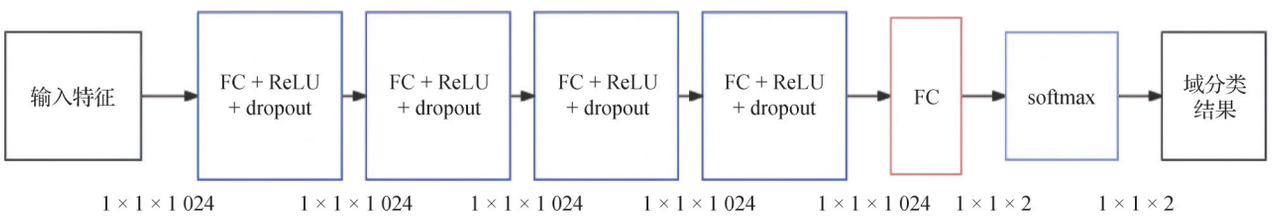


图 4　域判别器 D 的网络结构

Fig. 4　The network of domain discriminator D

1.1.5　测试

在测试阶段，目标域中未标记样本的分类过程如图 5 所示。首先，利用目标映射层对目标域数据集中有标注样本集和待分类的未标记样本集进行降维处理。其次，通过已训练好的 LRCT 来提取 HSI 的空间—光谱嵌入特征，以获取数据的深层次信息。最后，用最近邻分类器对待分类的未标记样本集进行分类，得到的分类精度作为 LRCT-CDFSL 的评估结果。

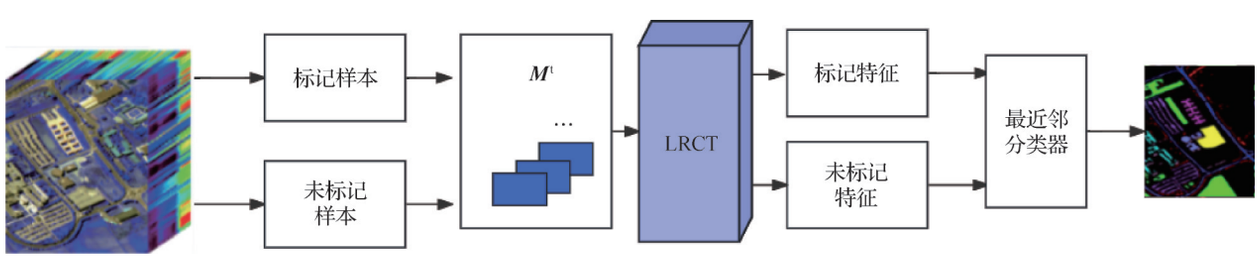


图 5　测试模型

Fig. 5　Testing model

1.2　算法流程

输入：源域原始高光谱数据 $\underline{\boldsymbol{X}}^{\mathrm{s}}$ 及其标签 $\boldsymbol{Y}^{\mathrm{s}}$，目标域原始高光谱数据 $\underline{\boldsymbol{X}}^{\mathrm{t}}$，训练总轮数和批（batch）大小分别为 K 与 n_{b}。

输出：目标域预测标签 $\boldsymbol{Y}^{\mathrm{t}}$。

1）数据预处理。根据式（4）对源域数据 $\underline{\boldsymbol{X}}^{\mathrm{s}}$ 和目标域数据 $\underline{\boldsymbol{X}}^{\mathrm{t}}$ 进行映射，获得维度统一的高光谱数据 $\underline{\boldsymbol{X}}_0^{\mathrm{s}}$ 和 $\underline{\boldsymbol{X}}_0^{\mathrm{t}}$。

2）网格初始化。随机初始化 LRCT-CDFSL 网络参数。

3）迭代训练。

for $k=1$ to：

（1）批采样。从 $\underline{\boldsymbol{X}}_0^{\mathrm{s}}$ 和 $\underline{\boldsymbol{X}}_0^{\mathrm{t}}$ 中分别随机采样 n_b 个样本，构成源域批数据 $\underline{\boldsymbol{X}}_{0(n_b)}^{\mathrm{s}}$ 和目标域批数据 $\underline{\boldsymbol{X}}_{0(n_b)}^{\mathrm{t}}$。

（2）特征提取。将 $\underline{\boldsymbol{X}}_{0(n_b)}^{\mathrm{s}}$ 和 $\underline{\boldsymbol{X}}_{0(n_b)}^{\mathrm{t}}$ 输入特征提取器 LRCT，提取深度特征 $\boldsymbol{Z}^{\mathrm{s}}$ 和 $\boldsymbol{Z}^{\mathrm{t}}$。

（3）少样本学习与分类损失计算。先将 $\boldsymbol{Z}^{\mathrm{s}}$ 和 $\boldsymbol{Z}^{\mathrm{t}}$ 输入少样本学习模块，然后根据式（10）计算查询集 $\boldsymbol{Q}_{\mathrm{s}}$ 中查询样本 $\boldsymbol{x}_u$ 的类分布 P；最后根据式（11）和式（12）分别计算源域和目标域中所有查询样本的分

类损失 $L_{\mathrm{fsl}}^{\mathrm{s}}$ 和 $L_{\mathrm{fsl}}^{\mathrm{t}}$。

(4)源域监督与域对齐。将 $\boldsymbol{Z}^{\mathrm{s}}$ 输入分类器和域对齐模块,之后结合源域标签 $\boldsymbol{Y}^{\mathrm{s}}$,计算源域分类与域对齐损失 L_d。

(5)目标域伪标签生成。将 $\boldsymbol{Z}^{\mathrm{t}}$ 输入分类器,得到目标域批数据的伪标签 $\tilde{\boldsymbol{Y}}^{\mathrm{t}}$。

(6)损失计算与反向传播。利用 $L_{\mathrm{fsl}}^{\mathrm{s}}$、$L_{\mathrm{fsl}}^{\mathrm{t}}$ 和 L_d,根据式(3)计算总损失 L_{total}。执行反向传播以优化LRCT-CDFSL网络参数。

(7)结束迭代训练。

end for。

4)模型预测。将预处理后的目标域数据 $\underline{\boldsymbol{X}}_0^{\mathrm{t}}$ 输入训练完成的LRCT-CDFSL网络,得到最终的目标域预测结果 $\boldsymbol{Y}^{\mathrm{t}}$。

2 实 验

2.1 数据集

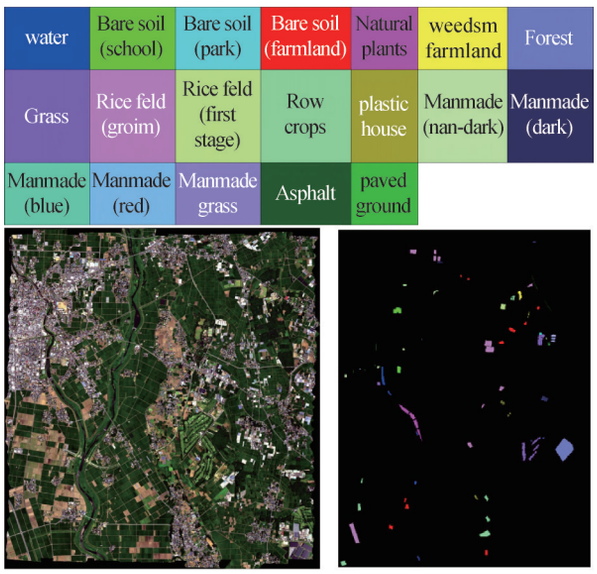
在本文中,实验数据来自4个真实的高光谱数据集,本节对这4个数据集分别进行简单的介绍。

1)Chikusei数据集(王肖,2021),该数据集由Hyperspectral Visible/Near-Infrared Cameras光谱成像仪在日本筑西(Chikusei)市采集所得。该数据集包含343～1 018 nm范围内的128个光谱带,空间像素为2 517×2 335。空间分辨率为2.5 m。它包含19个类,其假彩色图像和相应的真值图如图6所示。

2)印第安森林(Indian Pines,IP)数据集(Feng等,2019)。该数据集由机载可视/红外成像光谱仪在美国印第安纳州采集。该数据大小为145×145像素。波长范围为0.4～2.5 μm,光谱波段为224个,其中有效波段为200个,空间分辨率为20 m,它包含了16个地物种类,表1给出了这个数据集的概述。

3)帕维亚大学(Pavia University,UP)数据集(Peng等,2023)。该数据集是由反射光学系统成像光谱仪在意大利帕维亚市采集得到的高光谱数据组成。该数据集的有用光谱波段数目是103。图像的空间分辨率为1.3 m,总共有610×340像素,其中包含了9种类型的地物。表2提供了对这个数据集的概述。

4)萨利纳斯(Salinas)数据集(Shen和Ma,2019)。该数据集是由机载可见红外成像光谱仪在萨利纳斯山谷上空收集的,该数据集包含224个光谱反射波段,最终保留204个波段用于实验。图像大小为512×217像素,空间分辨率为3.7 m。该数据集共包含16个地物种类。表3提供了这个数据集的概述。



(a) 假彩色图像　　(b) 真值图

图6 Chikusei数据集

Fig. 6 Chikusei dataset

((a) false-color image;(b) groud-truth map)

表1 IP数据集的类别及样本个数

Table 1 The type and number of samples of the IP dataset

类别	类别名称	样本个数
C1	Alfalfa	46
C2	Corn-notill	1 428
C3	Corn-mintill	830
C4	Corn	237
C5	Grass-pasture	483
C6	Grass-trees	730
C7	Grass-pasture-mowed	28
C8	Hay-windrowed	478
C9	Oats	20
C10	Soybean-notill	972
C11	Soybean-mintill	2 455
C12	Soybean-clean	593
C13	Wheat	205
C14	Woods	1 265
C15	Buildings-Grass-Trees-Drives	1 136
C16	Stone-Steel-Towers	93

表2　UP数据集的类别及样本个数

Table 2　The type and number of samples of the UP dataset

类别	类别名称	样本个数
C1	Asphalt	6 631
C2	Meadows	18 649
C3	Gravel	2 099
C4	Trees	3 064
C5	Painted metal sheets	1 345
C6	Bare Soil	5 029
C7	Bitumen	1 330
C8	Self-Blocking Bricks	3 682
C9	Shadows	947

表3　Salinas数据集的类别及样本个数

Table 3　The type and number of samples of the Salinas dataset

类别	类别名称	样本个数
C1	Brocoli_green_weeds_1	2 009
C2	Brocoli_green_weeds_2	3 726
C3	Fallow	1 976
C4	Fallow_rough_plow	1 394
C5	Fallow_smooth	2 678
C6	Stubble	3 959
C7	Celery	3 579
C8	Grapes_untrained	11 271
C9	Soil_vinyard_develop	6 203
C10	Corn_senesced_green_weeds	3 278
C11	Lettuce_romaine_4wk	1 068
C12	Lettuce_romaine_5wk	1 927
C13	Lettuce_romaine_6wk	916
C14	Lettuce_romaine_7wk	1 070
C15	Vinyard_untrained	7 268
C16	Vinyard_vertical_trellis	1 807

2.2　实验设置

本文所用的实验数据集为Chikusei、Indian Pines、Salinas和Pavia University。在实验中，以Chikusei数据集为源域、其他数据集分别为目标域进行实验。为确保实验的公正性，在所有对比实验中，均从每类中随机选取5个有标记的目标域样本进行训练，选择9×9大小的邻域作为输入立方体的空间尺寸，其他实验条件保持一致。训练过程共进行10 000次迭代，学习率设定为0.001。为确保结果的稳定性，每个实验都独立重复10次。

使用Python3.7的PyTorch-1.13.1框架，在硬件配置为Intel（R）Xeon（R）W-2295 CPU，NVIDIA GeForce RTX 3090和64 G的Ubuntu系统上完成。

2.3　对比实验

为验证提出的LRCT-CDFSL方法的性能，本文选择了5种经典的HIC方法进行对比。具体包括支持向量机（support vector machine，SVM）和4种经典的跨域少样本HIC方法：深度跨域少样本学习（deep cross-domain few-shot learning，DCFSL）（Li等，2022）、图信息聚合跨域少样本学习（graph information aggregation cross-domain few-shot learning，Gia-CFSL）（Zhang等，2024）、跨域少样本的多视图关系学习（multi-view relation for cross-domain few-shot learning，MvR-CFSL）（Peng等，2023）和注意力跨域少样本学习（attentive cross-domain few-shot learning，ACDFSL）（Basnet等，2023）。此外，比较了LRCT-CDFSL方法的两种消融方法：基于Transformer网络的CDFSL（Transformer network based CDFSL，TN-CDFSL）以及基于轻量型Res-3D-CNN网络的CDFSL（lightweight Res-3D-CNN based CDFSL，LRC-CDFSL）的性能。选择以Chikusei数据集为源域，IP、Salinas和UP数据集分别为目标域进行CD-HIC。

1）对LRCT-CDFSL方法和7种对比算法的分类性能进行定量比较。采用每类的分类精度以及总体精度（overall accuracy，OA）、平均分类精度（average accuracy，AA）和Kappa系数等3个性能指标评估实验结果。所有算法在3个数据集上的分类结果分别如表4—表6所示。可以看出，SVM因为提取的是浅层次的特征，且只进行了单场景学习，因此其分类精度较低。而深度模型DCFSL、Gia-CFSL、MvR-CFSL、ACDFSL、TN-CDFSL、LRC-CDFSL等能提取深层次的空间—光谱特征，且进行了跨域少样本学习，因此在CD-HIC任务上的分类精度始终比SVM高。值得注意的是，在所有方法中，LRCT-CDFSL方法在IP、Salinas和UP数据集上均取得了最高的OA和Kappa系数，同时在IP和Salinas数据集上均取得了最高的AA，并在这两个数据上有一半的类别取得最高的分类精度。此外，LRCT-CDFSL在大多数情形下标准差最低，显示出其更好的稳健性。这些实验结果

表4　不同方法在IP数据集的分类精度对比

Table 4　Classification accuracy comparison of different methods on IP dataset

/%

类别	SVM	DCFSL	Gia-CFSL	MvR-CFSL	ACDFSL	TN-CDFSL	LRC-CDFSL	LRCT-CDFSL
1	74.15±19.35	95.12±6.17	90.98±12.25	99.02±2.24	95.12±6.17	90.49±13.60	95.12±7.07	**100.00±0.00**
2	34.80±**7.40**	40.07±9.99	44.03±13.66	54.00±14.65	40.07±9.99	43.75±13.46	56.07±11.55	**61.86**±10.58
3	38.58±10.88	48.32±11.37	44.18±8.73	57.27±8.21	48.32±11.37	44.15±10.28	54.15±8.84	**62.02±6.59**
4	53.88±11.49	74.18±12.51	77.93±8.88	77.84±12.20	74.18±12.51	72.37±14.39	75.17±14.36	**79.91±7.96**
5	68.51±16.60	73.33±7.35	71.76±7.10	78.70±8.38	73.33±7.35	74.39±8.02	73.39±7.18	**84.02±3.87**
6	70.26±12.12	83.54±7.45	82.37±5.99	**91.64±3.92**	83.54±7.45	88.69±6.74	90.03±4.98	88.72±7.63
7	90.43±7.88	96.96±5.52	98.26±5.22	98.26±5.22	96.96±5.52	98.26±3.98	98.70±2.78	**100.00±0.00**
8	71.84±10.44	80.57±13.79	85.12±9.04	89.18±17.05	80.5±13.79	90.19±10.39	92.39±9.41	**93.49±5.09**
9	89.33±7.17	100.00±0.00	97.33±3.27	98.67±2.67	**100.00±0.00**	96.67±6.15	**100.00±0.00**	99.33±2.00
10	43.42±11.47	60.64±7.17	58.70±8.43	65.14±8.85	60.64±7.17	57.79±10.09	65.76±11.36	**69.31±2.26**
11	32.61±9.02	59.40±12.77	61.07±13.58	61.79±12.17	59.40±12.77	59.91±16.00	**65.10±6.52**	60.53±8.47
12	26.73±**5.50**	45.32±11.13	42.02±8.89	**54.00**±8.65	45.32±11.13	38.72±8.89	53.35±17.07	52.70±12.36
13	87.50±7.40	97.20±4.80	98.35±2.72	**99.60**±0.77	97.20±4.80	98.45±1.96	93.65±6.63	99.35±**0.55**
14	57.76±13.68	83.26±9.24	79.24±13.06	89.01±9.78	83.26±9.24	86.98±8.99	**90.00**±5.70	84.29±**5.51**
15	29.79±8.51	67.72±10.83	64.33±**6.31**	67.93±16.52	67.72±10.83	67.03±13.93	**70.29**±16.52	69.74±12.11
16	87.73±3.82	99.32±1.16	97.61±4.61	99.77±0.68	99.32±1.16	99.77±0.68	97.39±2.69	**100.00±0.00**
OA	45.72±2.66	63.44±2.20	63.20±3.13	69.73±3.01	65.81±2.57	64.36±4.35	70.37±**1.52**	**71.01**±3.02
AA	59.83±2.64	75.31±1.59	74.58±1.82	80.11±2.49	77.63±1.75	75.48±20.63	79.41±2.07	**81.58±1.49**
Kappa	39.74±2.72	58.87±2.25	58.45±3.23	66.00±3.25	61.60±2.72	59.89±4.57	66.55±**1.76**	**67.42**±3.32

注：加粗字体表示各行最优结果。

充分验证了LRCT-CDFSL方法的有效性。这是因为Transformer层和Res-3D-CNN中Conv层的结合能更好地学习到具有局部—全局的多尺度特征表示，能为后续的CD-HIC提供支持。

2）采用分类效果图对LRCT-CDFSL方法和7种对比算法的分类性能进行视觉比较。图7—图9分别展示了所有算法在3个数据集上的分类效果图。从图7—图9可以发现，基于光谱信息的SVM分类结果最差，存在明显的椒盐噪声和误分类。而同时考虑了空谱信息的DCFSL、Gia-CFSL、MvR-CFSL、ACDFSL和LRC-CDFSL方法明显降低了椒盐噪声和误分类，但是仍有许多误分类，如IP数据上的woods、Salinas数据上的Grapes_untrained和Vinyard_ untrained以及UP数据上的Meadows类别。得益于Transformer的自注意力机制，LRCT-CDFSL能够在输入序列中建立起全局的依赖关系，准确预测woods、Grapes_untrained、Vinyard_ untrained和Meadows等地物类别。LRCT-CDFSL在分类边界的处理上展现出了显著优势，其分类边界更为清晰、平滑。

3）实验验证了LRCT-CDFSL方法的CD-HIC效率。表7和表8给出了不同算法在3个数据集上的训练时间和测试时间。可以看出，相较于其他方法，Gia-CFSL的训练时间最长，这主要是因为在训练过程中计算了庞大的节点数量，以及图之间的连接关系极为复杂，因此导致训练时间显著增加。而DCFSL、MvR-CFSL、ACDFSL方法中采用基于Res-3D-CNN的特征提取方法，Res-3D-CNN在处理三维数据时，虽然能够捕捉到丰富的空间—光谱信息，但其复杂的网络结构和庞大的计算量使得训练过程耗时较长。

LRC-CDFSL和LRCT-CDFSL方法中通过引入卷积核调整特征图通道数，减少模型参数，从而提高了训练速度，在IP、UP和Salinas这3个实验数据集上

表 5　不同方法在 Salinas 数据集的分类精度对比

Table 5　Classification accuracy comparison of different methods on Salinas dataset

/%

类别	SVM	DCFSL	Gia-CFSL	MvR-CFSL	ACDFSL	TN-CDFSL	LRC-CDFSL	LRCT-CDFSL
1	96.60±1.64	99.02±0.77	99.18±0.94	99.00±1.17	99.69±0.53	99.60±0.60	98.73±0.98	**99.70±0.41**
2	93.42±3.08	99.31±1.06	98.83±0.87	99.56±0.66	99.24±1.81	99.65±0.58	98.92±1.81	**100.00±0.01**
3	85.37±12.5	90.65±9.18	88.91±10.0	92.89±8.00	90.66±10.2	88.47±12.05	98.14±1.69	**99.01±0.63**
4	98.88±0.91	99.16±1.00	99.04±1.03	99.47±0.83	99.37±0.67	99.04±1.03	98.71±0.60	**99.79±0.16**
5	95.02±3.79	92.03±3.61	89.14±5.75	91.28±3.87	90.73±4.23	90.00±5.19	92.32±**2.15**	**95.32**±2.86
6	97.59±0.91	99.19±0.89	98.74±1.60	**99.92±0.17**	99.36±1.05	99.63±0.45	**99.92**±0.19	99.64±0.85
7	98.21±2.83	99.24±0.85	93.53±17.50	**99.63±0.46**	99.35±1.09	96.59±8.02	98.83±0.69	97.68±1.01
8	61.56±14.81	74.72±12.56	76.59±10.11	80.08±7.27	74.49±8.39	71.41±8.26	**83.69±3.58**	81.91±3.92
9	94.81±2.07	99.14±1.04	97.01±3.62	**99.49±0.70**	98.65±2.39	98.64±2.20	98.57±1.58	97.75±1.46
10	76.87±8.61	83.94±7.31	80.86±9.64	82.91±12.59	85.34±9.43	85.46±6.99	**92.45±2.35**	91.91±4.46
11	92.33±5.12	98.19±1.95	96.10±3.71	98.84±0.97	97.17±2.62	96.80±2.37	98.88±0.54	**99.76±0.38**
12	97.71±1.94	97.46±5.72	**99.42**±0.98	98.06±3.32	98.02±2.56	98.92±1.64	98.91±**0.69**	98.98±1.44
13	98.22±0.73	98.85±1.26	97.32±2.88	97.82±3.22	98.75±0.66	99.15±0.63	**99.31±0.41**	99.12±0.49
14	87.98±3.14	98.54±1.18	97.27±2.60	97.88±2.46	**99.08±0.66**	99.06±0.84	98.44±1.35	97.11±1.80
15	56.28±16.23	74.17±7.15	74.18±5.41	70.56±16.07	77.84±5.92	75.57±9.06	76.78±7.75	**79.04±5.84**
16	84.48±8.66	89.36±6.26	84.20±9.39	90.56±6.23	90.53±5.47	93.30±5.03	93.72±3.87	**99.12±0.67**
OA	81.42±2.26	88.71±2.74	87.83±2.64	89.51±1.86	89.21±1.93	88.13±1.61	91.86±1.12	**92.06±0.27**
AA	88.46±1.28	93.31±1.64	91.89±2.10	93.62±1.16	93.64±1.35	93.21±8.63	95.40±**0.92**	**95.99**±6.22
Kappa	79.37±2.48	87.46±3.00	86.48±2.92	88.33±2.09	88.03±2.12	86.84±1.76	90.95±0.92	**91.17±0.29**

注：加粗字体表示各行最优结果。

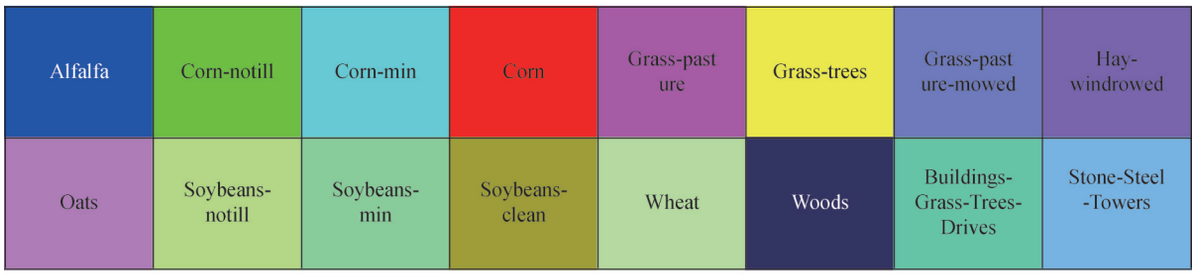
表 6　不同方法在 UP 数据集的分类精度对比

Table 6　Classification accuracy comparison of different methods on UP dataset

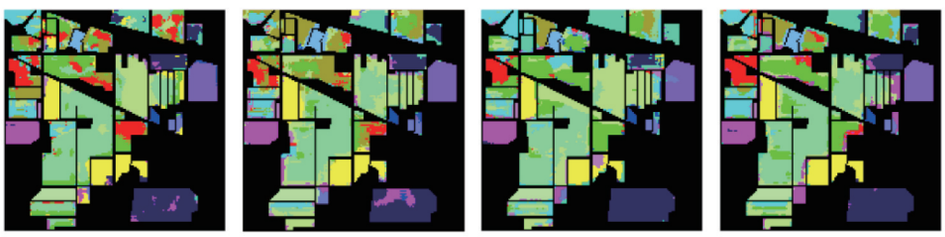
/%

类别	SVM	DCFSL	Gia-CFSL	MvR-CFSL	ACDFSL	TN-CDFSL	LRC-CDFSL	LRCT-CDFSL
1	62.78±7.38	78.01±12.02	76.33±11.27	76.68±8.56	78.51±9.64	73.55±19.44	**83.48±5.54**	77.80±9.06
2	66.10±13.90	86.30±6.36	87.14±5.31	80.22±8.75	84.25±8.51	81.90±10.35	85.57±6.93	**95.56±1.47**
3	61.46±18.65	63.00±10.84	64.60±11.48	63.5±14.24	62.94±14.43	55.71±14.60	68.56±11.89	**81.71±5.27**
4	80.84±13.85	**92.96**±4.53	92.77±3.55	87.63±8.09	92.82±**3.51**	90.43±4.18	85.81±9.08	77.97±8.17
5	98.88±0.80	98.84±1.12	98.03±2.32	99.89±0.24	99.42±0.75	99.49±0.44	99.10±1.21	**99.95±0.13**
6	54.14±13.01	73.38±8.79	73.97±7.46	62.09±14.69	**79.63**±9.00	72.96±12.91	76.16±13.99	77.44±**4.83**
7	74.87±11.68	79.07±8.50	79.33±8.92	87.27±5.40	80.85±8.73	73.83±11.53	85.23±4.74	**87.72±4.45**
8	71.24±9.23	64.31±11.37	67.54±13.83	**86.71±6.52**	66.00±16.37	57.22±7.39	75.08±7.72	43.54±7.31
9	99.85±0.12	97.93±3.27	**99.00±0.81**	95.63±5.37	98.12±2.27	97.26±4.31	95.84±6.27	94.75±3.10
OA	67.49±6.01	81.37±3.00	81.89±2.16	78.99±3.31	81.49±3.61	77.40±4.26	83.06±2.71	**84.14±1.46**
AA	74.46±2.38	81.53±0.82	82.08±2.06	82.18±2.13	82.50±1.77	78.04±14.88	**83.87**±1.93	81.83±**1.84**
Kappa	59.09±6.53	75.81±3.50	76.52±2.56	73.00±3.75	76.16±4.11	70.89±4.86	77.93±3.41	**78.84±1.94**

注：加粗字体表示各行最优结果。



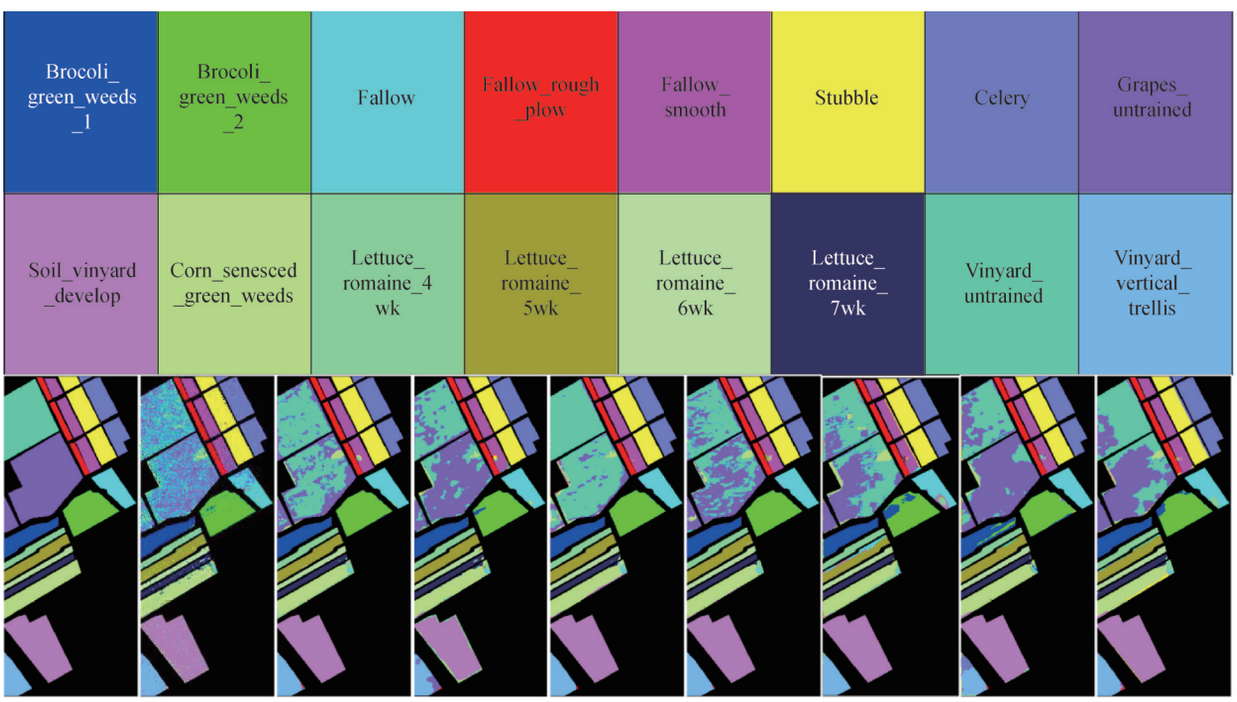
(a) 原图 (b) SVM (c) DCFSL (d) Gia-CFSL (e) MvR-CFSL



(f) ACDFSL (g) TN-CDFSL (h) LRC-CDFSL (i) LRCT-CDFSL (本文)

图7 不同分类方法在IP数据集上的分类效果图

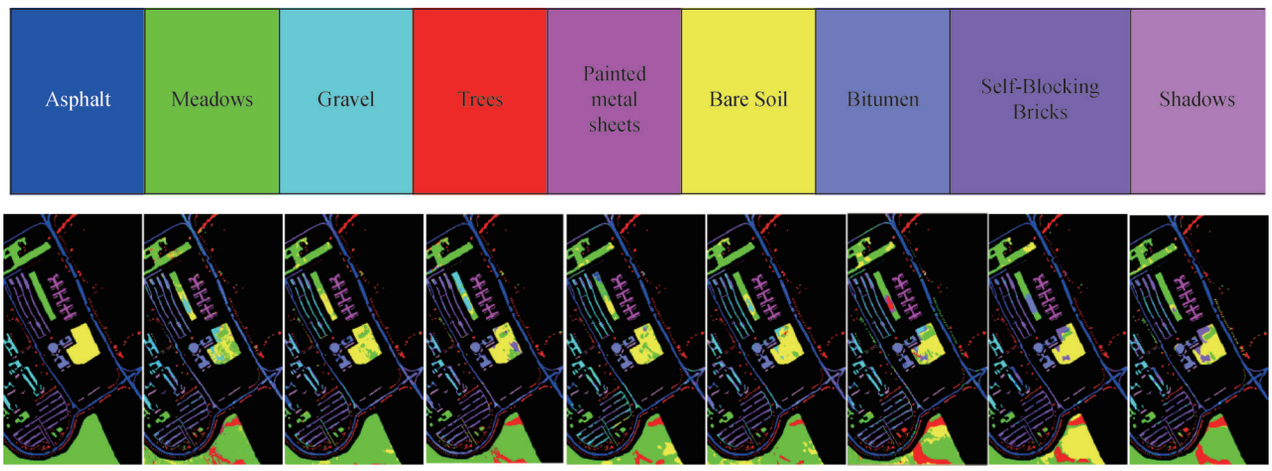
Fig. 7 Classification maps of different methods on IP dataset ((a) ground-truth; (b) SVM; (c) DCFSL; (d) Gia-CFSL; (e) MvR-CFSL; (f) ACDFSL; (g) TN-CDFSL; (h) LRC-CDFSL; (i) LRCT-CDFSL(ours))



(a) 原图 (b) SVM (c) DCFSL (d) Gia-CFSL (e) MvR-CFSL (f) ACDFSL (g) TN-CDFSL (h) LRC-CDFSL (i) LRCT-CDFSL (本文)

图8 不同分类方法在Salinas上的分类效果图

Fig. 8 Classification maps of different methods on Salinas dataset ((a) ground-truth; (b) SVM; (c) DCFSL; (d) Gia-CFSL; (e) MvR-CFSL; (f) ACDFSL; (g) TN-CDFSL; (h) LRC-CDFSL; (i) LRCT-CDFSL(ours))



(a) 原图　(b) SVM　(c) DCFSL　(d) Gia-CFSL　(e) MvR-CFSL　(f) ACDFSL　(g) TN-CDFSL　(h) LRC-CDFSL　(i) LRCT-CDFSL (本文)

图 9　不同分类方法在 UP 上的分类效果图

Fig. 9　Classification maps on UP dataset ((a) ground-truth; (b) SVM; (c) DCFSL; (d) Gia-CFSL; (e) MvR-CFSL; (f) ACDFSL; (g) TN-CDFSL; (h) LRC-CDFSL; (i) LRCT-CDFSL(ours))

的训练时间均显著低于其他方法。而 TN-CDFSL 中采用的 Transformer 网络没有经过残差层和池化层，其网络规模较大、参数较多，所以耗费的训练时间比 LRC-CDFSL 和 LRCT-CDFSL 方法长。

表 7　不同方法在 IP、UP、Salinas 上的训练时间对比

Table 7　Comparison of training time on IP, UP, and Salinas datasets

/s

方法	数据集		
	IP	UP	Salinas
DCFSL	1 250.35	757.97	1 281.72
Gia-CFSL	3 117.61	1 935.96	3 286.92
MvR-CFSL	1 771.35	837.98	1 549.40
ACDFSL	1 372.50	855.59	1 411.86
TN-CDFSL	1 135.79	782.77	1 427.84
LRC-CDFSL	**760.14**	**489.74**	816.44
LRCT-CDFSL	799.31	527.35	**811.64**

注：加粗字体表示各列最优结果。

4）折线图 10—图 12 分别描述了 7 种跨域小样本学习算法在 IP、UP 和 Salinas 数据集上，随着每类训练样本量增加时，整体精度（OA）的变化趋势。

在 IP 数据集上，随着每类训练样本量的增加，所有模型的 OA 均持续上升。如图 10 所示，随着每类训练样本数目由 1 增加至 5，LRCT-CDFSL 的 OA 从 42.38% 增至 71.01%，始终领先其他模型；LRC-CDFSL（70.37%）与 MvR-CFSL（69.73%）次之，但它们在低样本量阶段提升较慢；DCFSL（63.44%）、Gia-

表 8　不同方法在 IP、UP、Salinas 上的测试时间对比

Table 8　Comparison of testing time on IP, UP, and Salinas datasets

/s

方法	数据集		
	IP	UP	Salinas
DCFSL	0.71	2.30	3.76
Gia-CFSL	0.70	2.30	3.69
MvR-CFSL	0.98	2.46	4.63
ACDFSL	0.82	2.70	4.33
TN-CDFSL	7.44	26.71	43.20
LRC-CDFSL	0.78	2.92	3.96
LRCT-CDFSL	**0.66**	**2.23**	**3.24**

注：加粗字体表示各列最优结果。

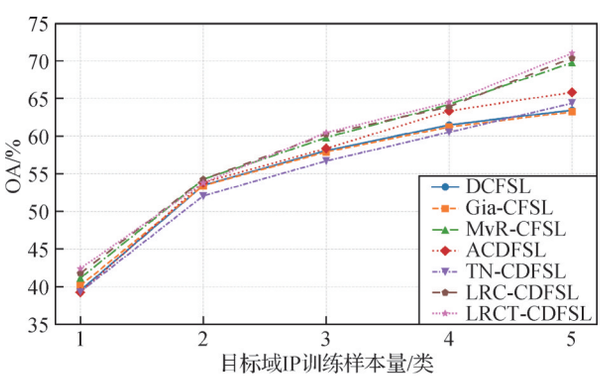


图 10　各算法在 IP 数据集上不同训练样本量下所获 OA 值

Fig. 10　OA accuracy of algorithms on IP dataset with different number of training samples

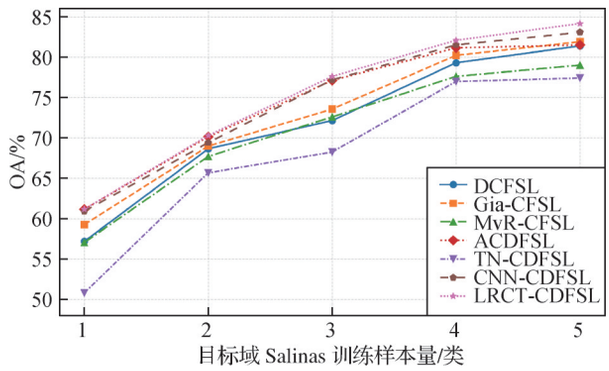


图11　各算法在UP数据集上不同训练样本量下所获OA值

Fig. 11　OA accuracy of algorithms on UP dataset with different number of training samples

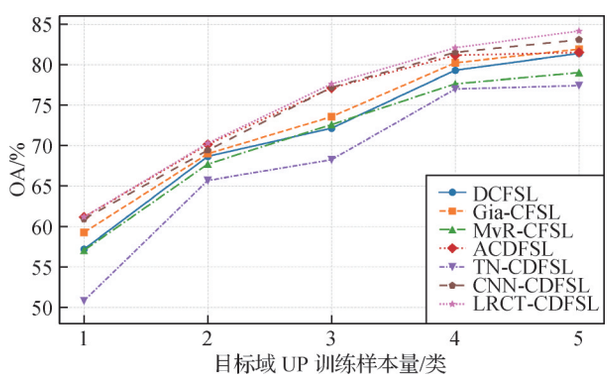


图12　各算法在Salinas数据集上不同训练样本量下所获OA值

Fig. 12　OA accuracy of algorithms on Salinas dataset different with number of training samples

CFSL(63.20%)和ACDFSL(65.81%)增幅平缓,后者在每类4个训练样本时跃升5%(58.34%→63.33%);TN-CDFSL增速最低(每类训练样本数目由2到3时仅增加4.63%),最终达到64.36%。LRCT-CDFSL在不同训练样本量下均保持最高增长速率与最终性能,验证了其在小样本场景下的鲁棒性。

在UP数据集上,随着每类训练样本量的增加,所有模型的OA均持续提升。如图11所示,LRCT-CDFSL保持最优,每类训练样本数目为1～5时,OA从61.17%稳步增至84.14%,尤其在训练样本数目4～5阶段增速最快(+2.07%);LRC-CDFSL以83.06%紧随其后,每类训练样本数目达到3后表现强劲(77.17%→83.06%)。ACDFSL在每类训练样本数目为3时突增6.98%(70.12%→77.10%),最终达到81.49%。DCFSL与Gia-CFSL分别以81.37%和81.89%收尾,但低样本量阶段(每类训练样本数为1～2时)提升平缓(均<10%)。MvR-CFSL增速最低(每类5个训练样本时为78.99%),而TN-CDFSL从50.81%逐步增至77.40%,每类训练样本从2个到3个时仅增2.57%,凸显其小样本适应性瓶颈。LRCT-CDFSL方法全程领先,验证了其鲁棒性。

在Salinas数据集中,随着每类训练样本量的增加,所有模型的OA均显著提升。如图12所示,LRCT-CDFSL全程领跑,训练样本从每类1个增加至5个时,从72.64%稳步攀升至92.06%,尤其在每类训练样本数由4至5阶段增速最快(+2.96%);LRC-CDFSL以91.86%紧随其后,每类5个训练样本时反超其他模型(88.68%→91.86%)。MvR-CFSL与ACDFSL表现接近,最终分别达到89.51%和89.21%,其中ACDFSL在每类4个训练样本时突增4.00%(84.55%→88.55%)。DCFSL与Gia-CFSL增速平缓,在每类训练样本数由1到2阶段提升均超过8%。TN-CDFSL低样本量表现最弱(每类1个训练样本时仅为69.54%),但后期增长强劲(每类训练样本数目由3～5阶段累计+4.31%),最终达到88.13%。LRCT-CDFSL在全部样本规模下的持续领先(最高增幅达19.42%)验证了其跨域小样本学习的综合优势。

5)LRCT-CDFSL是嵌入Transformer层的轻量型Res-3D-CNN。它提取HSI具有局部—全局联合的空间—光谱信息,并将其运用CDFSL框架实现跨域高光谱图像分类目的。本文采用消融实验验证其各关键部分在跨域高光谱图像分类中所起的作用。如表9所示,可得如下结论:基于纯轻量型Res-3D-CNN的CDFSL算法(LRC-CDFSL)在IP和Salinas中较纯Transformer网络的CDFSL算法(TN-CDFSL)性能强,这主要得益于Res-3D-CNN提取深度空间—光谱高频特征;但是数据集并非均为高频特征,光谱的低频特征也同样不可或缺。这3种算法中,LRCT-CDFSL在3个实验数据上的分类精度均最高,说明Transformer网络对于需要建立长距离依赖性和低频光谱数据中具有出色表现,将其嵌入Res-3D-CNN能提升跨域分类精度。LRCT-CDFSL吸收了以上两种算法的多种优势,从而在大多场景下性能得以提升。此外,表7和表8表明,在TN-CDFSL、LRC-CDFSL和LRCT-CDFSL 3种算法中,TN-CDFSL总体耗时最严重,LRC-CDFSL算法比较省时,而LRCT-CDFSL在提高性能的前提下计算效率相较于TN-CDFSL方法有很大提高,在IP和UP数据集上的训练耗时略高于LRC-CDFSL算法;LRCT-CDFSL在Salinas数据集上

表 9　LRCT-CDFSL 关键组成部分的消融实验结果

Table 9　Ablation study of key components of LRCT-CDFSL

/%

方法	数据集		
	IP	UP	Salinas
TN-CDFSL (Transformer)	64.36±4.35	77.40±4.26	88.13±1.61
LRC-CDFSL (lightweight Res-3D-CNN)	70.37±**1.52**	83.06±2.71	91.86±1.12
LRCT-CDFSL (Hybrid)	**71.01**±3.02	**84.14±1.46**	**92.06±0.27**

注：加粗字体表示各列最优结果。

的训练耗时以及所有数据集上的测试耗时均略低于 LRC-CDFSL。总而言之，综合考虑分类精度和效率，可以发现 LRCT-CDFSL 算法优于其他两种算法。

2.4　参数分析

1）为了探究空间窗口大小 $T \times T$ 对模型分类性能的影响，分别选取 T = 3, 5, 7, 9, 11, 13 进行实验。对每种情形分别进行 10 次独立实验，然后取平均结果。图 13 中，在 Salinas 数据集和 IP 数据集上，可以清晰地观察到一种趋势：随着空间窗口的大小 T 的不断增大，LRCT-CDFSL 的 OA 整体呈现出明显递增的趋势。这是因为随着空间窗口尺寸的增大，LRCT-CDFSL 所捕获的空间信息和光谱信息越来越丰富，这有助于 HSI 的分类。在 Pavia University 数据集上，LRCT-CDFSL 的 OA 整体呈现出先上升后下降的趋势，这是因为当空间邻域大于 9 × 9 时，空间邻域中的不同类像素对分类产生的负面效应强于空间窗口增大所包含的空谱信息对 HIC 产生的正面效应。总之，在空间窗口大小为 9 × 9 的时候，LRCT-CDFSL 的 OA 已经达到了相对理想的精度，同时，该空间窗口下的计算复杂度相对较低。因此当空间窗口大小为 9 × 9 时，该模型在性能和效率之间达到平衡。

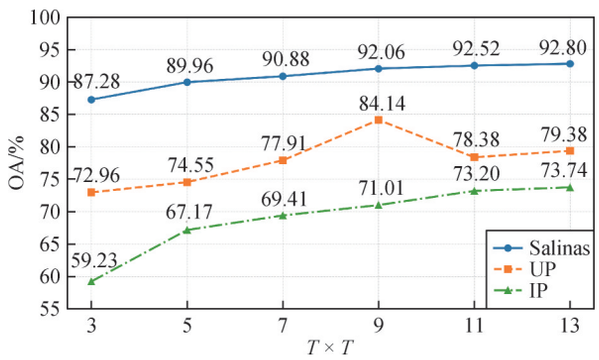


图 13　LRCT-CDFSL 在不同窗口大小下所获 OA 值

Fig. 13　OA accuracy of LRCT-CDFSL under different window sizes

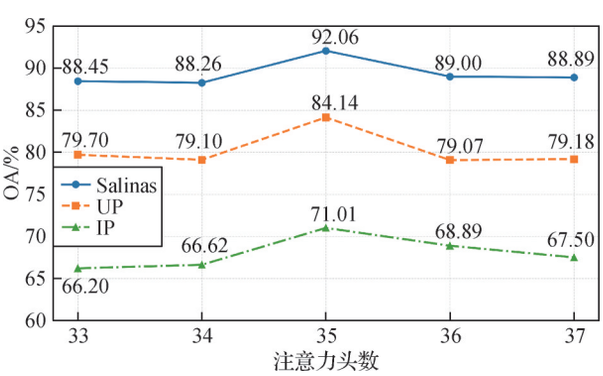


图 14　LRCT-CDFSL 在不同注意力头数目下所获 OA 值

Fig. 14　OA accuracy of LRCT-CDFSL under different numbers of attention heads

2）为探究 Transformer 模块中注意力头数目对模型分类性能的影响，分别选取注意力头数 33～37 进行实验。对每种情形分别进行 10 次独立实验，然后取平均结果。图 14 中，在所有数据集中随着注意力头数的增加，LRCT-CDFSL 的 OA 整体呈现出先上升后下降的趋势，顶点均在注意力头数目为 35 处，与其余处的 OA 差异明显。所以，在注意力头数为 35 时，LRCT-CDFSL 的 OA 已经达到了相对理想的精度。

3）为探究 Transformer 层数对模型分类性能的影响，分别选取 1～5 层进行实验。对每种情形分别进行 10 次独立实验，然后取平均结果。图 15 的所有数据集随着 Transformer 层数的增加，LRCT-CDFSL 的 OA 整体走向基本相同，显然单层 Transformer 的 LRCT-CDFSL 性能最佳。

3　结　论

针对现有 CD-HIC 方法所面临光谱相似性及地物分布扭曲的挑战，本文提出 LRCT-CDFSL 框架并做出验证。主要结论如下：

1）本文的主要工作在于设计了一种用于 CD-HIC 的特征提取网络 LRCT，该网络通过将 Trans-

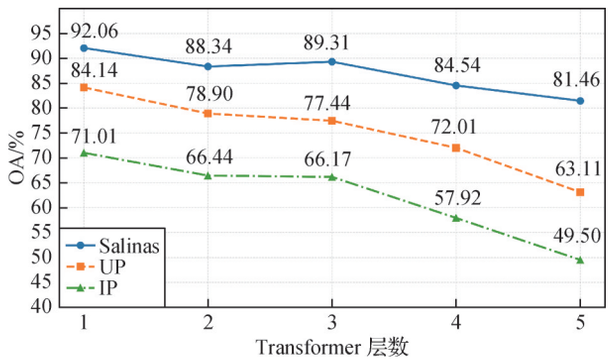


图 15　LRCT-CDFSL在不同Transformer层数下所获OA值

Fig. 15　OA accuracy of LRCT-CDFSL under different numbers of Transformer layers

former层嵌入轻量化的Res-3D-CNN架构中实现突破：首先，在Res-3D-CNN中策略性引入1×1卷积核灵活调控特征图通道数，显著减少参数量以缓解传统模型的计算负担；其次，充分利用三维卷积擅长提取局部高频空谱细节和Transformer建模光谱长程依赖、聚焦关键区域并捕获低频全局信息的互补特性，融合生成判别力更强的“局部—全局联合”空谱特征，有效应对光谱相似性及地物分布扭曲的挑战；最后，将LRCT作为特征提取器嵌入CDFSL框架，结合映射层预处理、原型网络少样本学习及条件对抗域适应技术，构建了从特征提取到跨域少样本分类的端到端优化模型LRCT-CDFSL。

2）在以Chikusei为源域，IP、Salinas和UP为目标域的跨域实验中，LRCT-CDFSL的OA分别达71.01%、92.06%、84.14%，验证了其融合特征在极端样本稀缺与域漂移下的高分类能力。与最优对比方法相比，LRCT-CDFSL在IP、Salinas和UP的OA分别提升1.28%、2.55%、2.25%，训练时间减少47.6%～72.8%，且消融实验证实其融合策略（优于纯卷积或纯Transformer模型）兼具性能与效率优势。

3）然而，由于LRCT过分依赖简单度量学习而忽略高阶交互，且其深层依赖建模受限，使得本文方法存在样本关系建模不足和Transformer层数增加反而导致性能下降的不足。在未来工作中，计划深入考量交叉注意力学习深化样本关系建模，探索多源域适应增强泛化性，设计自适应Conv-Transformer架构，以及应用模型压缩技术进一步提升模型性能。

综上，LRCT-CDFSL通过轻量级双流特征提取网络与CDFSL框架，显著提升跨域少样本高光谱分类的精度与效率，为标记稀缺和域差异问题提供有效解决方案，未来将聚焦深化样本关系建模、增强架构鲁棒性及拓展场景应用。

参考文献（References）

Basnet R, Goperma R and Zhao L. 2023. Attentive cross-domain few-shot learning and domain adaptation in HSI Classification//Proceedings of 2023 TENCON IEEE Region 10 Conference (TENCON). Chiang Mai, Thailand: IEEE: 220-225 [DOI: 10.1109/TENCON58879.2023.10322397]

Chandra M A and Bedi S S. 2021. Survey on SVM and their application in image classification. International Journal of Information Technology, 13(5): 1-11 [DOI: 10.1007/s41870-017-0080-1]

Chen Y S, Zhao X and Jia X P. 2015. Spectral-spatial classification of hyperspectral data based on deep belief network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(6): 2381-2392 [DOI: 10.1109/JSTARS.2015.2388577]

Cover T and Hart P. 1967. Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1): 21-27 [DOI: 10.1109/TIT.1967.1053964]

Feng F, Wang S T, Wang C Y and Zhang J. 2019. Learning deep hierarchical spatial-spectral features for hyperspectral image classification based on residual 3D-2D CNN. Sensors, 19(23): #5276 [DOI: 10.3390/s19235276]

Fisher R A. 1922. Statistical methods for research workers. London: Oliver & Boyd

Garcia V and Bruna J. 2018. Few-shot learning with graph neural networks [EB/OL]. [2025-01-12]. http://arxiv.org/pdf/1711.04043v3.pdf

He M Y, Li B and Chen H H. 2017. Multi-scale 3D deep convolutional neural network for hyperspectral image classification//Proceedings of 2017 IEEE International Conference on Image Processing (ICIP). Beijing, China: IEEE: 3904-3908 [DOI: 10.1109/ICIP.2017.8297014]

Hong D F, Han Z, Yao J, Gao L R, Zhang B, Plaza A, et al. 2022. SpectralFormer: rethinking hyperspectral image classification with transformers. IEEE Transactions on Geoscience and Remote Sensing, 60: #5518615 [DOI: 10.1109/TGRS.2021.3130716]

Hu S, Gao F, Gong Z R, Tao S E, Shangguan X Y and Dong J Y. 2024. Parallel channel shuffling and Transformer-based denoising for hyperspectral images. Journal of Image and Graphics, 29(7): 2063-2074 (胡帅, 高峰, 龚卓然, 陶盛恩, 上官心语, 董军宇. 2024. 基于Transformer和通道混合并行卷积的高光谱图像去噪. 中国图象图形学报, 29(7): 2063-2074) [DOI: 10.11834/jig.230381]

Li Z K, Liu M, Chen Y S, Xu Y M, Li W and Du Q. 2022. Deep cross-domain few-shot learning for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 60:

#5501618［DOI：10.1109/TGRS.2021.3057066］

Liu C，Yang L W，Li Z，Yang W，Han Z G，Guo J Z，et al. 2024. Multi-level relation learning for cross-domain few-shot hyperspectral image classification. Applied Intelligence，54(5)：4392-4410［DOI：10.1007/s10489-024-05384-3］

Marmanis D，Datcu M，Esch T and Stilla U. 2016. Deep learning earth observation classification using ImageNet pretrained networks. IEEE Geoscience and Remote Sensing Letters，13(1)：105-109［DOI：10.1109/LGRS.2015.2499239］

Peng Y S，Liu Y R，Tu B and Zhang Y W. 2023. Convolutional Transformer-based few-shot learning for cross-domain hyperspectral image classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing，16：1335-1349［DOI：10.1109/JSTARS.2023.3234302］

Shen D and Ma L. 2019. Cross-domain extreme learning machine for classification of hyperspectral images//IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium. Yokohama，Japan：IEEE：3305-3308［DOI：10.1109/IGARSS.2019.8898437］

Sun L，Zhao G R，Zheng Y H and Wu Z B. 2022. Spectral-spatial feature tokenization Transformer for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing，60：#5522214［DOI：10.1109/TGRS.2022.3144158］

Tun N L，Gavrilov A，Tun N M，Trieu D M and Aung H. 2021. Hyperspectral remote sensing images classification using fully convolutional neural network//Proceedings of 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering（ElConRus）. St. Petersburg，Russia：IEEE：2166-2170［DOI：10.1109/ElConRus51938.2021.9396673］

Vaswani A，Shazeer N，Parmar N，Uszkoreit J，Jones L，Gomez A N，et al. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach，USA：Curran Associates Inc.：6000-6010

Wang B Q，Xu Y，Wu Z B，Zhan T M and Wei Z H. 2022. Spatial-spectral local domain adaption for cross domain few shot hyperspectral images classification. IEEE Transactions on Geoscience and Remote Sensing，60：#5539515［DOI：10.1109/TGRS.2022.3208897］

Wang H Y，Cheng Y H and Wang X S. 2023. Correlation subdomain alignment network based cross-domain hyperspectral image classification method. Journal of Image and Graphics，28(10)：3255-3266（王浩宇，程玉虎，王雪松. 2023. 关联子域对齐网络的跨域高光谱图像分类. 中国图象图形学报，28(10)：3255-3266）［DOI：10.11834/jig.220763］

Wang X. 2021. Research on Hyperspectral and Multispectral Image Fusion Algorithm Based on Convolutional Neural Network. Hefei：Hefei University of Technology（王肖. 2021. 基于卷积神经网络的高光谱与多光谱图像融合算法研究. 合肥：合肥工业大学）［DOI：10.27101/d.cnki.ghfgu.2021.000660］

Ye Z，Wang J，Liu H，Zhang Y and Li W. 2023. Adaptive domain-adversarial few-shot learning for cross-domain hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing，61：#5532017［DOI：10.1109/TGRS.2023.3334289］

You X E，Su Y C，Jiang M Y，Li P F，Liu D S and Bai J Y. 2024. Deep embedded Transformer network with spatial-spectral information for unmixing of hyperspectral remote sensing images. Journal of Image and Graphics，29(8)：2220-2235（游雪儿，苏远超，蒋梦莹，李朋飞，刘东升，白晋颖. 2024. 深度嵌套式Transformer网络的高光谱图像空谱解混方法. 中国图象图形学报，29(8)：2220-2235）［DOI：10.11834/jig.230393］

Zeng Q J，Geng J，Jiang W，Huang K and Wang Z Y. 2022. IDLN：iterative distribution learning network for few-shot remote sensing image scene classification. IEEE Geoscience and Remote Sensing Letters，19：#8020505［DOI：10.1109/LGRS.2021.3109728］

Zhang Y X，Li W，Zhang M M，Wang S，Tao R and Du Q. 2024. Graph information aggregation cross-domain few-shot learning for hyperspectral image classification. IEEE Transactions on Neural Networks and Learning Systems，35(2)：1912-1925［DOI：10.1109/TNNLS.2022.3185795］

Zhao W Z and Du S H. 2016. Spectral-spatial feature extraction for hyperspectral image classification：a dimension reduction and deep learning approach. IEEE Transactions on Geoscience and Remote Sensing，54(8)：4544-4554［DOI：10.1109/TGRS.2016.2543748］

Zheng Z Z，Zhang Y M，Li L T，Zhu M C，He Y，Li M Q，et al. 2017. Classification based on deep convolutional neural networks with hyperspectral image//2017 IEEE International Geoscience and Remote Sensing Symposium（IGARSS2017）. Fort Worth，USA：IEEE：1828-1831［DOI：10.1109/IGARSS.2017.8127331］

Zhong Z L，Li J，Luo Z M and Chapman M. 2018. Spectral-spatial residual network for hyperspectral image classification：a 3-D deep learning framework. IEEE Transactions on Geoscience and Remote Sensing，56(2)：847-858［DOI：10.1109/TGRS.2017.2755542］

Zhong Z L，Li Y，Ma L F，Li J and Zheng W S. 2022. Spectral-spatial Transformer network for hyperspectral image classification：a factorized architecture search framework. IEEE Transactions on Geoscience and Remote Sensing，60：#5514715［DOI：10.1109/TGRS.2021.3115699］

作者简介

杨丽霞，女，副教授，主要研究方向为机器学习及图像处理。E-mail：yanglixia1982@163.com

张瑞，通信作者，男，副教授，主要研究方向为图像处理。E-mail：guanyue_002@163.com

鲍雅君，女，硕士研究生，主要研究方向为机器学习及图像处理。E-mail：1826692512@qq.com

杨淑媛，女，教授，主要研究方向为机器学习、信号处理及图像处理。E-mail：syyang@xidian.edu.cn