

中图法分类号: TP751 文献标识码: A 文章编号: 1006-8961(2024)08-2220-16

论文引用格式: 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(08):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(08):2220-2235)[DOI:10.11834/jig.230393]

# 深度嵌套式Transformer网络的高光谱图像空谱解混方法

游雪儿<sup>1</sup>, 苏远超<sup>1\*</sup>, 蒋梦莹<sup>2</sup>, 李朋飞<sup>1</sup>, 刘东升<sup>3</sup>, 白晋颖<sup>3</sup>

1. 西安科技大学测绘科学与技术学院, 西安 710054; 2. 西安交通大学电子与信息学部, 西安 710049;  
3. 航天宏图信息技术股份有限公司, 北京 100195

**摘要:** **目的** 基于深度学习的解混方法在信息挖掘和泛化性能上优于传统方法, 但主要关注光谱信息, 对空间信息的利用仍停留在滤波、卷积的表层处理。这使得构建解混网络时需要堆叠多层网络, 易丢失部分图像信息, 影响解混准确性。Transformer网络因其强大的特征表达能力广泛应用于高光谱图像处理, 但将其直接应用于解混学习容易丢失图像局部细节。本文基于Transformer网络提出了改进方法。**方法** 本文以TNT(Transformer in Transformer)构架为基础提出了一种深度嵌套式解混网络(deep embedded Transformer network, DETN), 通过内外嵌入式策略实现编码器中局部与整体空间信息共享, 不仅保留了高光谱图像的空间细节, 而且在编码器中只涉及少量卷积运算, 大幅度提升了学习效率。在解码器中, 通过一次卷积运算来恢复数据结构以便生成端元与丰度, 并在最后使用Softmax层来保障丰度的物理意义。**结果** 最后, 本文分别采用模拟数据集和真实高光谱数据集进行对比实验, 在50 dB模拟数据集中平均光谱角距离和均方根误差取得最优值, 分别为0.038 6和0.004 5, 在真实高光谱数据集Samson、Jasper Ridge中取得最优平均光谱角距离, 分别为0.119 4, 0.102 7。**结论** 实验结果验证了DETN方法的有效性和优势, 并且能为实现深度解混提供新的技术支撑和理论参考。

**关键词:** 遥感图像处理; 高光谱遥感; 混合像元分解; 深度学习; Transformer网络

## Deep embedded Transformer network with spatial-spectral information for unmixing of hyperspectral remote sensing images

You Xueer<sup>1</sup>, Su Yuanchao<sup>1\*</sup>, Jiang Mengying<sup>2</sup>, Li Pengfei<sup>1</sup>, Liu Dongsheng<sup>3</sup>, Bai Jinying<sup>3</sup>

1. College of Geomatics, Xi'an University of Science and Technology, Xi'an 710054, China; 2. Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China; 3. PIESAT Information Technology Co., Ltd., Beijing 100195, China

**Abstract: Objective** In hyperspectral remote sensing, mixed pixels often exist due to the complex surface of natural objects and the limitation of spatial resolution of instruments. Mixed pixels typically refer to the situation where a pixel in the hyperspectral images usually contains multiple spectral features, which hinders the application of hyperspectral images

收稿日期: 2023-06-29; 修回日期: 2024-02-05; 预印本日期: 2024-02-12

\* 通信作者: 苏远超 suyeh3@xust.edu.cn

基金项目: 国家自然科学基金项目(42001319); 教育部产学研协同育人项目(220802313200859); 陕西省教育厅科研计划项目(21JK0762)

Supported by: National Natural Science Foundation of China(42001319); University-Industry Collaborative Education Program of Ministry of Education of China(220802313200859); Scientific Research Program of the Education Department of Shaanxi Province, China(21JK0762)

in various fields such as target detection, image classification, and environmental monitoring. Therefore, the decomposition (unmixing) of mixed pixels is a main concern in the processing of hyperspectral remote sensing images. Spectral unmixing aims to overcome the limitations of image spatial resolution by extracting pure spectral signals (endmembers) representing each land cover class and their respective proportions (abundances) within each pixel. It is based on a spectral mixing model at the sub-pixel level. The rise of deep learning has brought many advanced modeling theories and architecture tools to the field of hyperspectral mixed pixel decomposition and has also spawned many deep learning-based unmixing methods. Although these methods have advantages over traditional methods in terms of information mining and generalization performance, deep networks often need to combine multiple layers of stacked network layers to achieve optimal learning outcomes. Therefore, deep networks may cause damage to the internal structure of the data during the training process, which leads to the loss of important information in hyperspectral data and affects the accuracy of unmixing. In addition, most existing deep learning-based unmixing methods focus only on spectral information, but the exploit of spatial information is still limited to surface processing stages such as filtering and convolution. In recent years, autoencoder has been one of the research hotspots in the field of deep learning, and many variant networks based on autoencoder networks have emerged. Transformer is a novel deep learning network with an autoencoder-like structure. It has garnered considerable attention in various fields such as natural language processing, computer vision, and time series analysis due to its powerful feature representation capability. The Transformer, as a neural network primarily based on the self-attention mechanism, can better explore the underlying relationships among different features and more comprehensively aggregate the spectral and spatial correlations of pixels. This way enhances the ability of abundance learning and improves the accuracy of unmixing. Although the Transformer network has recently been used to design unmixing methods, using unsupervised Transformer models directly to obtain features can lose many local details and cause difficulty in exploiting the long-range dependency properties of Transformers effectively. **Method** To address these limitations, the study proposes a deep embedded Transformer network (DETN) based on the Transformer-in-Transformer architecture. This network adopts an autoencoder framework that consists of two main parts: node embedding (NE) and blind signal separation. In the first part, the input hyperspectral image is first uniformly divided twice, and the divided image patches are mapped into sub-patch sequences and patch sequences through linear transformation operations. Then, the sub-patch sequences are processed through an internal Transformer structure to obtain pixel spectral information and local spatial correlations, which are then aggregated into the patch sequences for parameter and information sharing. Finally, the local detail information in the patch sequences is retained, and the patch sequences are processed through an external Transformer structure to obtain and output pixel spectral information and global spatial correlation information containing local information. In the second part, the input NE is first reconstructed into an abundance map and smoothed during this process using a single layer of 2D convolution layer to eliminate noise. A SoftMax layer is used to ensure the physical meaning of the abundances. Finally, a single-layer 2D convolution layer is used to reconstruct the image, which optimizes and estimates the endmembers in the convolution layer. **Result** To evaluate the effectiveness of the proposed method, experiments are conducted using simulated datasets and some real hyperspectral datasets, including the Samson dataset, the Jasper Ridge dataset, and a part of the real hyperspectral farmland data in Nanchang City, Jiangxi Province, obtained by the Gaofen-5 satellite provided by Beijing Shengshi Huayao Technology Co., Ltd. In addition, resources from the ZY1E satellite provided by Beijing Shengshi Huayao Technology Co., Ltd. are used to obtain partial hyperspectral data of the Marseille Port in France for comparative experiments with different methods. The experimental results are quantitatively analyzed using spectral angle distance (SAD) and root mean square error (RMSE). In addition, the method evaluates the proposed DETN compared with several state-of-the-art deep learning-based unmixing algorithms: fully strained least squares (FCLS), deep autoencoder networks for hyperspectral unmixing (DAEN), autoencoder network for hyperspectral unmixing with adaptive abundance smoothing (AAS), the untied denoising autoencoder with sparsity (uDAS), hyperspectral unmixing using deep imageprior (UnDIP), and hyperspectral unmixing using Transformer network (DeepTrans-HSU). Results demonstrate that the proposed method outperforms the compared methods in terms of spectral angle distance (SAD), root mean square error (RMSE), and other evaluation metrics. **Conclusion** The proposed method effectively captures and preserves the spectral information of pixels at local and global levels, as well as the spatial correlations among pixels. This method results in accurate extraction of endmembers that

match the ground truth spectral features. Moreover, the method produces smooth abundance maps with high spatial consistency, even in regions with hidden details in the image. These findings validate that the DETN method provides new technical support and theoretical references for addressing the challenges posed by mixed pixels in hyperspectral image unmixing.

**Key words:** remote sensing image processing; hyperspectral remote sensing; hyperspectral unmixing; deep learning; Transformer network

## 0 引言

高光谱图像因其“图谱合一”及光谱分辨率高的优点已广泛应用于环境监测、精准农业、医学检测等诸多领域的重要工作中(曲海成等, 2015; Gao等, 2022b; Su等, 2023)。然而复杂的自然地物表面以及仪器空间分辨率的限制使得高光谱图像中的一个像元通常包含多个光谱特征,这就导致了“混合像元”现象,给遥感图像解译带来严重问题(童庆禧等, 2016; 张兵, 2016)。混合像元分解(解混)是处理高光谱遥感图像中混合像元问题的重要技术,旨在突破图像空间分辨率限制,依据光谱混合模型从亚像元角度获取能够代表每类地物的纯净光谱信号(端元)和每类端元在像元中的占比(丰度)(Su等, 2021)。通常,解混的过程主要包括3个部分:估算端元数量、提取端元以及估算丰度。在解混的过程中,首先需要确定其所依赖的光谱混合模型,即线性光谱混合模型(linear spectral mixture model, LSMM)或非线性光谱混合模型(nonlinear spectral mixture model, NSMM)(张兵, 2016)。目前,许多有效的解混算法使用线性光谱混合模型简化高光谱图像中冗余的信息,使得解混过程相对简单、高效。

传统的基于线性光谱混合模型进行解混的方法主要有基于几何、统计或稀疏回归的方法(Miao和Qi, 2007; 汪顺清等, 2023)。其中,基于几何的解混算法主要有 Nascimento和Dias(2005)提出的顶点成分分析(vertex component analysis, VCA)、Winter(1999)提出的N-FINDR以及Boardman等人(1995)提出的纯像元指数法(pure pixel index, PPI)等。以上这些方法都基于纯像元假设,通过几何计算从图像中识别出纯净像元,作为解混估算的端元。然而在实际条件下,成像区域中复杂的地物场景使得高光谱图像中存在许多混合度较高的像元,纯净的像元往往难以获取。此外,基于统计以及稀疏回归的方法同样能够较好地处理高度混合像元(宋晓瑞

等, 2020)。在基于统计的解混方法中,学者们通常基于贝叶斯框架利用统计假设和先验知识来提高解混性能(Eches等, 2010)。而基于稀疏回归是一种半监督解混方法,通过已知光谱库稀疏回归筛选出最优端元子集作为解混结果以提高解混精度(Su等, 2018; Ozkan等, 2019)。例如, Qian等人(2011)在非负矩阵分解法的基础上,引入了 $L_{1/2}$ 正则化来约束丰度,并提出了一种新的方法用于解混工作,即 $L_{1/2}$ 稀疏约束非负矩阵分解法( $L_{1/2}$  sparsity constrained non-negative matrix factorization,  $L_{1/2}$  SC-NMF)。

近年来,深度学习的兴起为高光谱遥感混合像元领域带来了诸多先进建模理论和构架工具,也催生出许多基于深度学习的解混方法。在学者们尝试构建深度网络用于解混的初期, Guo等人(2015)首次成功将深度学习的理论应用于解混中,结合降噪自动编码器与非负稀疏自动编码器,提出了一种自动编码器层叠的混合像元分解方法(unmixing using autoencoder cascade)。自此,基于深度学习的解混方法中出现了很多以自动编码器网络为构架的变体网络。在解混学习设计中,自动编码器网络通常包括编码器和解码器两部分,编码器将输入的高光谱数据映射到一个低维表示(中间隐层),该表示通常用于输出丰度。而解码器则将低维表示映射回原始数据的高维空间,重构图像。在解码器中,通常会将权重设计为端元,初始化并通过优化损失函数反向梯度传递,进而得到输出端元。Su等人(2018)提出了栈式非负稀疏自编码(stacked nonnegative sparse autoencoders, SNSA)用于解混,该方法能够在无需纯净像元为先验条件的情况下,针对图像中的异常像元问题提升非负稀疏自动编码器的解混性能; Qu和Qi(2019)提出了编码器和解码器使用独立权重的稀疏去噪网络,在去噪自动编码器的基础上运用 $l_{2,1}$ 范数等正则化对编码器和解码器施加不同的约束,使得解混网络取得了良好的效果; Ozkan等人(2019)通过在稀疏自动编码器网络的基础上引入额外的层、指标以及新的损失函数来改进原始的自动

编码器网络以获得具有更强稀疏性的端元与丰度; Su 等人(2019)提出了一种基于深度自编码网络(deep autoencoder network, DAEN)的解混方法,构造了堆叠的自动编码器网络来处理像元异常值和噪声,并使用变分自动编码器(variational autoencoder, VAE)来约束丰度以提高解混精度;而 Borsoi 等人(2020)则使用 VAE 来学习光谱变化率并生成模型,以进一步学习高光谱影像图像中的端元变化率,从而引导网络估算的端元以及丰度具有更高的精度;同样地, Jin 等人(2022)对自动编码器网络进行了改进,提出了一种双流自动编码器网络,通过引入另一个具有相同结构的自动编码器来增强网络获取端元光谱信息的能力,从而能够将解混学习过程中所需的空谱信息与光谱信息充分结合。

相比于传统的解混方法,基于深度学习的解混方法在图像信息的挖掘方面更能发挥其优势。例如, Palsson 等人(2021)在自动编码器的模型中引入了卷积神经网络(convolution neural network, CNN),该方法将高光谱图像分块进行解混学习,保留了图块的空间特征;此外,在感知机制对空间信息挖掘的启发下, Gao 等人(2022a)提出了循环一致性的解混网络(cycle-consistency unmixing network, CyCU-Net),该网络通过学习两个端到端的级联自动编码器,从而减少了图像重建时的损失,提高了解混算法的精确性; Fang 等人(2022)提出了贝叶斯全卷积解混网络,该方法考虑了图像的空间信息、噪声等因素对端元的影响,从而增加了网络解混学习过程中的鲁棒性;最近, Qi 等人(2023)提出了一种多视图空间—光谱双流网络(multiview spatial-spectral two-stream network, MSSS-Net)用于解混,该方法使用两个端到端的循环神经网络(recurrent neural networks, RNNs)协同运作,充分挖掘图像的空间信息,并结合光谱信息得到高质量的解混结果。

如何有效地利用高光谱图像的空间信息和光谱信息,是提高解混精度的关键。由上述解混算法可知,现有基于深度学习的大多数解混方法能够比较深层地挖掘并利用图像光谱维度的信息,而对于空间维信息方面的利用却仍然停留在以滤波、卷积为主的表层处理阶段。并且在这些解混方法中,仅使用单一的基于深度学习网络进行解混容易丢失部分解混所需的特征信息。例如,仅使用单一 CNN 网络进行解混学习时,通常容易为了获取图像的空间信

息而忽视光谱信息,影响解混的精度等(Tu 等, 2022)。所以,这些方法在构建深度网络时往往需要结合多层堆叠的网络层以达到较为理想的学习效果。具有丰富地物信息的高光谱数据在被层层网络剔除冗余信息,从而高效捕获解混所需有效信息的同时,却存在对数据自身内部结构造成损坏的情况,导致处理高光谱数据时丢失部分重要信息,进而影响到解混的准确性。Transformer 是一种具有自动编码器结构的新型深度学习网络,其凭借强大的特征表达能力在自然语言处理、计算机视觉、时间序列分析等领域得到广泛关注(Sun 等, 2023)。作为一种主要基于自注意机制的神经网络,Transformer 能够更好地挖掘不同特征的潜在关系,更充分地聚合像元的光谱信息与空间相关性。Transformer 通过获取并分析高光谱图像中像元光谱信息的差异性和丰富度,从而区分不同的地物或材质,提高解混网络识别并选择端元的准确性。与此同时,Transformer 能够基于图像全局对像元空间相关性进行建模,通过考虑端元周围像元的特征信息,可以更准确地估算丰度,从而提升解混的精度(Ghosh 等, 2022)。但是,在解混学习中,若以无监督的方式直接使用传统的 Transformer 模型来获取特征信息会受到高光谱图像中大量冗余信息的影响,从而需要大量的数据进行解混训练(Sun 等, 2022)。一般情况下,人们会将 Transformer 模型与多层 CNN 网络进行堆栈结合,从而达到 CNN 的空间特征提取能力和 Transformer 的光谱建模能力相结合的目的。这种结构可以在高光谱图像解混中保留空间和光谱信息之间的关联,并利用 CNN 获取局部相关性特征来辅助 Transformer 模型以达到较为理想的解混训练效果。然而,这种方式若过度关注全局光谱与空间信息则容易破坏数据内部的空间结构。并且在数据量较少以及堆叠多个模型层使得参数数量过多的情况下,这种 CNN 与 Transformer 结合的方式可能会在学习过程中产生过拟合现象,导致 Transformer 本身的长距离依赖特性难以有效发挥,最后对解混精度造成影响(Tu 等, 2023)。此外,由于传统的 Transformer 本身对于图像分割的粒度不够精细,相关方法通常仅简单地将高光谱图像同构区域的图块进行分割以获取图块间像元的同质性特征,却忽略了图块内部像元之间的局部空间相关性。

针对传统的 Transformer 本身在图像处理方面存

在的问题, Han 等人(2021)提出了一种称为 TNT (Transformer in Transformer) 的新型 Transformer 网络, 该网络能够将图像更精细地分割为局部小块, 同时将图块间位置编码共享, 从而能够提供更多图像的细节特征。此外, 由于该网络架构针对局部图块进行挖掘并保留细节特征的能力较强, 使得 Transformer 长距离依赖特性得到有效发挥, 从而能够减少其受到图像中冗余信息的影响, 更加准确地在大量图像数据中获取所需的特征信息, 高效地达到较为理想的训练效果。

在解混方法设计方面, 从信号学角度出发, 无监督解混可以看做是一个盲信号分离的过程。而 Transformer 模型更适于构建无监督解混学习网络, 其自动编码器构架中的编码器部分能够从输入的高光谱图像中获取盲信号分离所需的特征信息, 并且其长距离依赖特性能更好地从全局角度获取像元之间的空间相关性。本文从空谱协同角度出发, 在 TNT 策略的启发下提出一种深度嵌套式解混网络 (deep embedded Transformer network, DETN)。DETN 使用内部和外部嵌套式 Transformer 结构将图像二次精细分割, 保留了传统 Transformer 模型能够有效获取高光谱图像中每个像元的光谱信息以及全局空间相关性优势的同时, 不仅能够弥补单一使用传统 Transformer 模型对图像局部相关性特征挖掘的不足, 而且整个解混学习过程中仅仅涉及少量二维卷积运算, 减少了结合传统 Transformer 模型与堆栈 CNN 网络层获取局部相关性特征发生过拟合的概率。此外, 在 DETN 方法中, 内部 Transformer 结构获取的局部像元光谱信息以及空间相关性直接聚合于外部 Transformer 结构中进行信息与位置编码共享, 使得编码器输出包含全局特征信息的矩阵维度与外部 Transformer 结构中图块表示维度相同, 减少参数设置且显著提高了解混的学习性能。

## 1 基于深度嵌套式 Transformer 网络解混方法与原理

### 1.1 线性光谱混合模型

深度嵌套式 Transformer 网络基于线性光谱混合模型进行解混, 在该模型中假设在太阳辐射所对应的物理区域内, 电磁波仅与一种地物表面发生作用, 而不考虑地物之间的相互作用(彭倩等, 2017)。由

此可得, 一个像元所对应的混合像元是由各个端元线性混合而成, 则在线性光谱混合模型中得到的光谱反射率为

$$X = WA + N \quad (1)$$

式中,  $X = [x_1, x_2, \dots, x_f] \in \mathbf{R}^{k \times f}$  是具有  $f$  个像元,  $k$  个波段的高光谱数据集,  $W = [w_1, w_2, \dots, w_c] \in \mathbf{R}^{k \times c}$  表示端元矩阵,  $c$  表示端元数,  $A = [a_1, a_2, \dots, a_f] \in \mathbf{R}^{c \times f}$  表示丰度矩阵,  $N \in \mathbf{R}^{k \times f}$  为  $X$  中存在的额外噪声矩阵。通常式(1)满足 3 个物理约束条件: 1) 端元矩阵应是非负矩阵; 2) 丰度矩阵应是非负矩阵; 3) 丰度向量的各元素之和为 1, 即:  $W \geq 0$ ;  $A \geq 0$ ;  $\mathbf{1}_c^T A = \mathbf{1}_f^T$ , 其中  $\mathbf{1}_c$  和  $\mathbf{1}_f$  为两个元素全为 1 的向量。

### 1.2 DETN 方法解混框架

本文提出的 DETN 方法主要由两个部分组成, 即节点表示和盲信号分离。第 1 部分为编码器中的节点表示, 其中一个节点表示一个像元, 该部分主要用于获取像元之间的空间相关性特征; 第 2 部分为解码器中的盲信号分离, 用于最终输出端元与丰度。在第 1 部分中, 首先将输入的高光谱图像两次均匀分割, 并且通过线性变换运算将分割后的图块映射为子图块序列以及图块序列; 其次将子图块序列通过内部 Transformer 结构获取像元光谱信息与局部空间相关性, 并且将其聚合于图块序列中, 实现参数与信息共享; 最后保留图块序列中的局部细节信息, 并将图块序列通过外部 Transformer 结构获取并输出像元的光谱信息以及包含局部信息的全局空间相关性特征信息。在第 2 部分中, 首先将输入的节点表示重构为丰度图, 使用单层  $3 \times 3$  大小的二维卷积层在此过程中进行平滑处理以消除噪声, 并且通过 softmax 层来保障丰度的物理意义。最后使用单层  $1 \times 1$  大小的二维卷积层来重构图像, 在卷积层中优化并估算出端元。

DETN 方法执行解混的工作流程如图 1 所示, 使用的部分关键符号如表 1 所示。

### 1.3 编码器与节点表示

在提出的深度嵌套式 Transformer 中, 采用了内部 Transformer 结构与外部 Transformer 结构相同的嵌套式网络构架, 增强了编码器聚合图像光谱信息与空间相关性的能力。由编码器获得的信息极大地提高了盲信号分离过程中估算端元的准确性以及获取丰度图的质量。

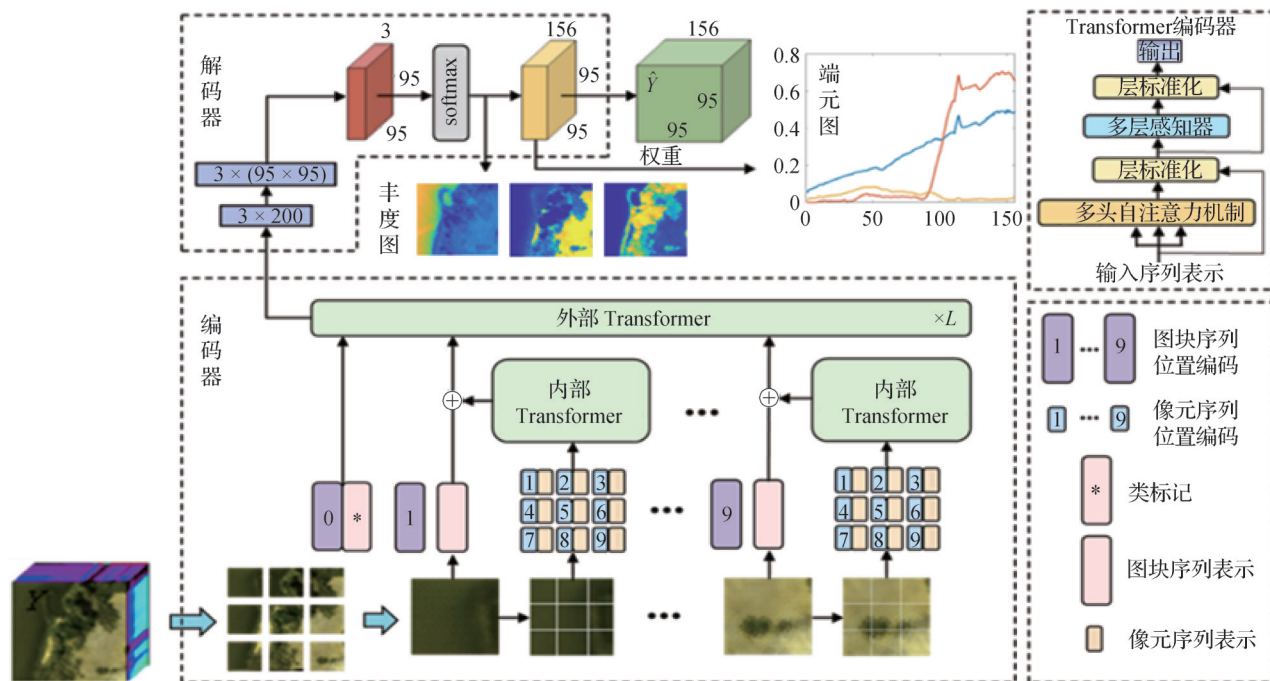


图1 DETN方法解混流程图

Fig. 1 Flowchart of DETN method for unmixing

表1 本文使用的部分关键符号说明

Table 1 Explanation of key notations in this paper

符号	说明
$X \in \mathbf{R}^{k \times f}$	高光谱数据集, 具有 $f$ 个像元, $k$ 个波段
$W \in \mathbf{R}^{k \times c}$	端元矩阵, 具有 $c$ 个端元
$A \in \mathbf{R}^{c \times f}$	丰度矩阵
$U^i \in \mathbf{R}^{m \times h}$	第 $i$ 个图块中的子图块表示序列, 包含 $m$ 个维度为 $h$ 的子图块表示
$U'^i \in \mathbf{R}^{m \times h}$	使用 1D 位置编码的子图块表示序列
$Q_u, K_u, V_u \in \mathbf{R}^{m \times h}$	内部 Transformer 模块中多头自注意层中的查询(query)矩阵、键值(key)矩阵、值项(value)矩阵
$\alpha_l \in \mathbf{R}^{m \times h}$	内部 Transformer 模块输出的子图块表示序列
$Y_l \in \mathbf{R}^{(n+1) \times d}$	图块表示序列, 包含 $n$ 个维度为 $d$ 的图块表示
$Y'_l \in \mathbf{R}^{(n+1) \times d}$	使用 1D 位置编码的图块表示序列
$Q_y, K_y, V_y \in \mathbf{R}^{(n+1) \times d}$	外部 Transformer 模块中多头自注意层中的查询(query)矩阵、键值(key)矩阵、值项(value)矩阵
$\delta_l \in \mathbf{R}^{(n+1) \times d}$	外部 Transformer 模块输出的图块表示序列

在编码器中, 将输入的高光谱图像通过滑动窗口均匀分割成若干图块与子图块, 具体为

$$\begin{aligned} X &\rightarrow [X^1, X^2, \dots, X^n] \\ X^i &\rightarrow [X^{i,1}, X^{i,2}, \dots, X^{i,m}] \end{aligned} \quad (2)$$

式中,  $X$  被均匀划分为  $n$  个图块, 分割后的图块表示为  $X^i$ 。将  $X^i$  进一步划分为  $m$  个子图块, 则在子图块  $X^{i,m}$  中,  $i$  为每个图块的索引,  $i = 1, 2, \dots, n$ ;  $j$  为每个子图块在图块中的索引,  $j = 1, 2, \dots, m$  (即

$X^{i,j}$  为第  $i$  个图块中的第  $j$  个子图块)。

接下来, 将子图块序列通过向量化运算并且线性投影到低维空间, 则子图块的表示序列为  $U^i = [U^{i,1}, U^{i,2}, \dots, U^{i,m}] \in \mathbf{R}^{m \times h}$ ,  $h$  为每个子图块序列的维度。在子图块表示中运用 1D 位置编码保留子图块内部局部像元空间相关性, 具体为

$$U'^i = U^i + U_{\text{pos}} \quad (3)$$

式中,  $U_{\text{pos}} \in \mathbf{R}^{b \times h}$ , 为子图块位置编码。然后, 将子图

块序列  $U^i$  通过层归一化(layer normalization, LN)运算并输入内部 Transformer 结构中,通过多头自注意(multi-head self-attention, MSA)层以获取所需的子图块内部局部像元的光谱以及空间信息。在 MSA 层中,子图块序列通过线性变换分为 3 个部分,具体为

$$Q_u = W_q U^i, K_u = W_k U^i, V_u = W_v U^i \quad (4)$$

式中,  $Q_u \in \mathbf{R}^{m \times h}$  为第 1 个线性层中的查询(query)矩阵,  $W_q \in \mathbf{R}^{m \times m}$  为  $Q_u$  的权重;  $K_u \in \mathbf{R}^{m \times h}$  为第 2 个线性层中的键值(key)矩阵,  $W_k \in \mathbf{R}^{m \times m}$  为  $K_u$  的权重;  $V_u \in \mathbf{R}^{m \times h}$  为第 3 个线性层中的值项(value)矩阵,  $W_v \in \mathbf{R}^{m \times m}$  为  $V_u$  的权重。在 MSA 层中,  $Q_u$  与  $K_u$  之间的相关性通过 softmax 函数进行计算并作为该层的总权重,则子图块序列内部自注意力为

$$f_{\text{MSA}}(\text{LN}(U^i)) = f_{\text{MSA}}(Q_u, K_u, V_u) = f_{\text{softmax}}\left(\frac{Q_u K_u^T}{\sqrt{h}}\right) V_u \quad (5)$$

式中,  $\text{LN}()$  为归一化层。

将子图块序列的自注意力通过线性转换并输出,为了进一步挖掘图像局部隐藏的深层次细节信息,将子图块序列的自注意力输入多层感知(multi-layer perceptron, MLP)层,具体为

$$U_l^i = U_{l-1}^i + f_{\text{MSA}}(\text{LN}(U_{l-1}^i)) \quad (6)$$

$$U_l^i = U_l^i + f_{\text{MLP}}(\text{LN}(U_l^i)) \quad (7)$$

式中,  $U_{l-1}^i$  为输入第 1 个内部 Transformer 结构中的子图块序列,  $U_l^i \in \mathbf{R}^{m \times h}$  为 MSA 层的输出序列,  $U_l^i \in \mathbf{R}^{m \times h}$  为 MLP 层的输出子图块表示。将整个内部与外部嵌入式 Transformer 结构堆叠  $L$  次以达到最佳训练效果,其中,内部与外部 Transformer 结构皆堆叠  $L$  次,  $l = 1, 2, \dots, L$  为内部与外部 Transformer 结构堆叠次数的索引。最后,内部 Transformer 结构输出的子图块表示序列为  $\alpha_l = [U_l^1, U_l^2, \dots, U_l^n] \in \mathbf{R}^{m \times h}$ 。

然后,将保留图像局部细节信息的序列  $\alpha_l$  聚合于外部 Transformer 结构中,得到  $Y_l^1, \dots, Y_l^n \in \mathbf{R}^{n \times d}$ , 为了进一步挖掘所需的图像全局信息,将外部 Transformer 结构中的图块序列中的第 2 个矩阵定义为类标记(class tokens)矩阵  $Y_{\text{cls}} \in \mathbf{R}^{1 \times d}$ ,  $Y_{\text{cls}}$  能够学习各类地物光谱信息以及同类地物像元之间的空间相关性。将图块序列通过向量化运算并且线性投影到低维空间,则图块的表示序列为  $Y_l =$

$[Y_{\text{cls}}, Y_l^1, \dots, Y_l^n] \in \mathbf{R}^{(n+1) \times d}$ ,  $d$  为每个图块表示的维度。同样在图块序列中,运用 1D 位置编码共享子图块序列中位置编码的权值并且获取全局像元空间相关性

$$Y_l' = (Y_{\text{cls}} \parallel Y_{\text{patch}}) + Y_{\text{pos}} = (Y_{\text{cls}} \parallel Y_{\text{patch}}') \quad (8)$$

式中,  $Y_{\text{pos}} \in \mathbf{R}^{n \times d}$  为图块序列的位置编码,  $Y_{\text{patch}} \in \mathbf{R}^{n \times d}$  为图块序列,  $Y_{\text{cls}}', Y_{\text{patch}}'$  为位置编码后的类标记矩阵以及图块序列。同样,将图块表示序列  $Y_l'$  通过层归一化运算并输入外部 Transformer 结构中,通过多头自注意层以获取所需的图块之间全局像元的光谱以及空间信息。多头自注意层通过线性变换将图块序列分为 3 个部分,即

$$Q_y = W_Q Y_{\text{cls}}', K_y = W_K Y_l', V_y = W_V Y_l' \quad (9)$$

式中,  $Q_y \in \mathbf{R}^{n \times d}$  为第 1 个线性层中的查询(query)矩阵,作用于  $Y_{\text{cls}}'$ ,  $W_Q \in \mathbf{R}^{n \times n}$  为  $Q_y$  的权重;  $K_y \in \mathbf{R}^{n \times d}$  为第 2 个线性层中的键值(key)矩阵,  $W_K \in \mathbf{R}^{n \times n}$  为  $K_y$  的权重;  $V_y \in \mathbf{R}^{n \times d}$  为第 3 个线性层中的值项(value)向量序列,  $W_V \in \mathbf{R}^{n \times n}$  为  $V_y$  的权重,第 2 个与第 3 个线性层皆作用于  $Y_l'$ 。与内部 Transformer 结构相同,图块序列的自注意力在 MSA 层中计算,通过线性转换并输出。为了进一步挖掘图块之间全局像元的深层次空间相关性,将图块序列的自注意力输入 MLP 层,即

$$Y_l'' = Y_{l-1}' + f_{\text{MSA}}(\text{LN}(Y_{l-1}')) \quad (10)$$

$$Y_l' = Y_l'' + f_{\text{MLP}}(\text{LN}(Y_l'')) \quad (11)$$

式中,  $Y_{l-1}'$  为输入第 1 个外部 Transformer 结构的图块序列,  $Y_l''$  为 MSA 层的输出序列,  $Y_l'$  为 MLP 层的输出序列。最后,外部 Transformer 结构输出的图块表示序列为  $\delta_l = [Y_{\text{cls}}', Y_l^1, \dots, Y_l^n] \in \mathbf{R}^{(n+1) \times d}$ 。

#### 1.4 解码器进行盲信号分离

在提出的 DETN 方法中,编码器中的外部 Transformer 结构最后输出的  $Y_{\text{cls}}' \in \mathbf{R}^{1 \times d}$  中具有解码器进行盲信号分离所需图像的空间信息。在解码器中,为了将  $Y_{\text{cls}}'$  重构为大小为  $\hat{A} \in \mathbf{R}^{c \times f}$  的丰度图,使用了二维卷积层对编码器输入的信息进行平滑处理,消除噪声,并且使用了 softmax 层来约束丰度矩阵,以确保丰度矩阵中的向量“非负”且“和为 1”的约束。其中,使用二维卷积层消除噪声的同时,为了能够尽量保留编码器输出图像的光谱信息和空间信息,本文使用卷积核为  $1 \times 1$ 、步长为 1 的单层二维卷积层,在卷积层中使用了 ReLU(rectified linear unit)激活

函数,能够确保输入的信息不被过度压缩。同时,在卷积层中设置输入与输出的通道数为端元数不变,卷积填充大小为 $1 \times 1$ ,确保输入与输出图像大小相同且能够获取并保留足够的特征信息。最后,在解码器中采用单层1的二维卷积层来重建图像 $\hat{X} \in \mathbf{R}^{k \times f}$ ,并且利用纯端元假设的端元提取方法(例如VCA)实现端元初始化作为卷积层的权值,通过梯度反向传播对权值进行更新,从而计算出端元矩阵 $\hat{W} \in \mathbf{R}^{k \times c}$ 。

### 1.5 损失函数的构建

为了更好地训练深度嵌套式Transformer网络,本研究将重构误差(reconstruction error, RE)损失与光谱角距离(spectral angle distance, SAD)损失进行加权和计算网络的总体损失,具体为

$$F_{\text{RE}}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{f} \sum_{i=1}^f (\hat{X}_i - X_i)^2 \quad (12)$$

$$F_{\text{SAD}}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{c} \sum_{s=1}^c \arccos \frac{\langle X_s, \hat{X}_s \rangle}{\|X_s\|_2 \|\hat{X}_s\|_2} \quad (13)$$

$$F_{\text{total}} = \varphi F_{\text{RE}} + \rho F_{\text{SAD}} \quad (14)$$

式中, $\varphi, \rho$ 为正则化权重参数。在以上计算中,RE损失通过均方误差(mean square error, MSE)目标函数进行计算。MSE通过学习输入图像的光谱重建,有助于充分地保留解混前后图像的光谱细节信息,能够更准确地区分图像中不同地物或材质的类别。但是,MSE对于端元光谱的绝对变化幅度非常敏感,可能会导致不稳定的解混学习结果。SAD损失通过光谱角距离度量了输入图像的光谱相似性,从光谱形状的角度强调了光谱特征信息的匹配,有助于解混网络更准确地估算不同类别端元的光谱信息。将MSE损失与SAD损失进行加权使用,降低了RE损失对端元光谱信息绝对变化幅度的过度敏感,加快模型的收敛速度,帮助解混网络更快地学习到正确的端元与丰度。本文相关权重参数设置为:在模拟数据实验中, $\varphi = 1 \times 10^3, \rho = 5 \times 10^{-2}$ ;在Samson数据实验中, $\varphi = 5 \times 10^4, \rho = 3 \times 10^{-2}$ ;在Jasper Ridge数据实验中, $\varphi = 5 \times 10^3, \rho = 5 \times 10^{-2}$ ;在农田数据实验中, $\varphi = 5 \times 10^3, \rho = 3 \times 10^{-2}$ 。

## 2 实验结果与分析

为了验证本文无监督算法的有效性,采用模拟

高光谱数据以及真实高光谱数据通过与一些经典解混算法进行对比实验,包括:全约束最小二乘(fully constrained least squares, FCLS) (Heinz 和 Chein-I-Chang, 2001)、DAEN (Su 等, 2019)、AAS (Hua 等, 2021)、uDAS (Qu 和 Qi, 2019)、UnDIP (Rasti 等, 2022)和DeepTrans-HSU (Ghosh 等, 2022)。此外,对不同算法的性能进行定性和定量分析,进一步说明提出方法的前沿性与可行性。

本文提出的DETN方法首先通过VCA算法初始化端元,然后使用Adam优化器实现梯度下降,并且分别对于不同的数据集设置不同的学习率衰减参数 $\gamma$ 以及迭代次数来优化训练。在使用模拟数据集的实验中,设置 $\gamma = 3 \times 10^{-3}$ ,迭代次数为200次;在Samson数据集中,设置 $\gamma = 4 \times 10^{-3}$ ,迭代次数为200次;在使用Jasper Ridge数据集的实验中,设置 $\gamma = 6 \times 10^{-3}$ ,迭代次数为300次;在使用高分5号农田数据集的实验中,设置 $\gamma = 4 \times 10^{-3}$ ,迭代次数为200次;在使用ZY-1 02D高光谱数据集的实验中,设置 $\gamma = 5 \times 10^{-3}$ ,迭代次数为300次。

为了准确地对图像的重构误差、丰度估算误差和端元估计误差的精度进行定量分析,引入了以下几种方法进行衡量。采用SAD评估图像的重构误差以及端元的估算误差 $f_{\text{SAD}}$ ,采用均方根误差(root mean square error, RMSE)评定丰度估算的误差( $f_{\text{RMSE}}$ ),以上评估方法的结果皆以平均值进行展示。具体为

$$f_{\text{SAD}}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{f} \sum_{i=1}^f \arccos \frac{\langle \mathbf{x}_i, \hat{\mathbf{x}}_i \rangle}{\|\mathbf{x}_i\|_2 \|\hat{\mathbf{x}}_i\|_2} \quad (15)$$

$$f_{\text{RMSE}}(\mathbf{A}, \hat{\mathbf{A}}) = \sqrt{\frac{1}{f} \sum_{i=1}^f (\mathbf{a}_i - \hat{\mathbf{a}}_i)^2} \quad (16)$$

$$f_{\text{SAD}}(\mathbf{W}, \hat{\mathbf{W}}) = \frac{1}{c} \sum_{s=1}^c \arccos \frac{\langle \mathbf{w}_s, \hat{\mathbf{w}}_s \rangle}{\|\mathbf{w}_s\|_2 \|\hat{\mathbf{w}}_s\|_2} \quad (17)$$

式中, $\mathbf{x}_i$ 是输入的高光谱图像, $\hat{\mathbf{x}}_i$ 是重构的高光谱图像; $\mathbf{a}_i$ 是参考的丰度, $\hat{\mathbf{a}}_i$ 是估算的丰度; $\mathbf{w}_s$ 是参考的端元, $\hat{\mathbf{w}}_s$ 是估算的端元。由此可得,SAD与RMSE的值越小表示精度越高。

### 2.1 使用模拟数据的验证实验

通过使用模拟数据能够在实验阶段获得精确的参考光谱,这对于验证算法的有效性至关重要。在本文的模拟实验部分,从USGS光谱库中随机选取4种矿物光谱作为真实的端元进行实验,这些真实

的端元作为基准用于测试并评估 DETN 的解混结果。该模拟数据的像素为  $26 \times 26$ , 包含了 198 个不同波段, 从而确保了高光谱数据的丰富性与细致度。在实验过程中, 分别对 3 种不同噪声水平 (20 dB, 30 dB, 50 dB) 的模拟数据进行解混。通过这种方式, 能够评估本文提出的 DETN 方法在不同噪声条件下的鲁棒性和学习效率。

图 2 以直观的方式展示了在 50 dB 噪声条件下, 采用不同算法对模拟数据进行解混处理后获得的丰度对比结果图。从图中可以观察到, 本文方法与真实情况 (ground truth) 之间的吻合度显著高于其他对比算法, 体现出 DETN 方法的优越性和准确性。

为了进一步验证这一点, 在实验中进行了定量分析。通过对图 2 展示的不同解混效果进行 10 次重复实验, 并计算平均值, 利用 SAD 和 RMSE 两个指标对比了不同方法的性能, 结果如表 2 所示。可以看出, DETN 方法的 SAD 和 RMSE 平均值均低于其他对比算法的相应值。这一定量分析结果进一步证实了相比于其他算法, DETN 方法能够获得更加精准和稳定的解混效果。

综上所述, 通过图 2 和表 2 的结果展示可以得出结论, 即在不同噪声环境下进行解混任务时, DETN 方法相较于其他同类算法, 能够实现更为精确和可靠的解混效果。

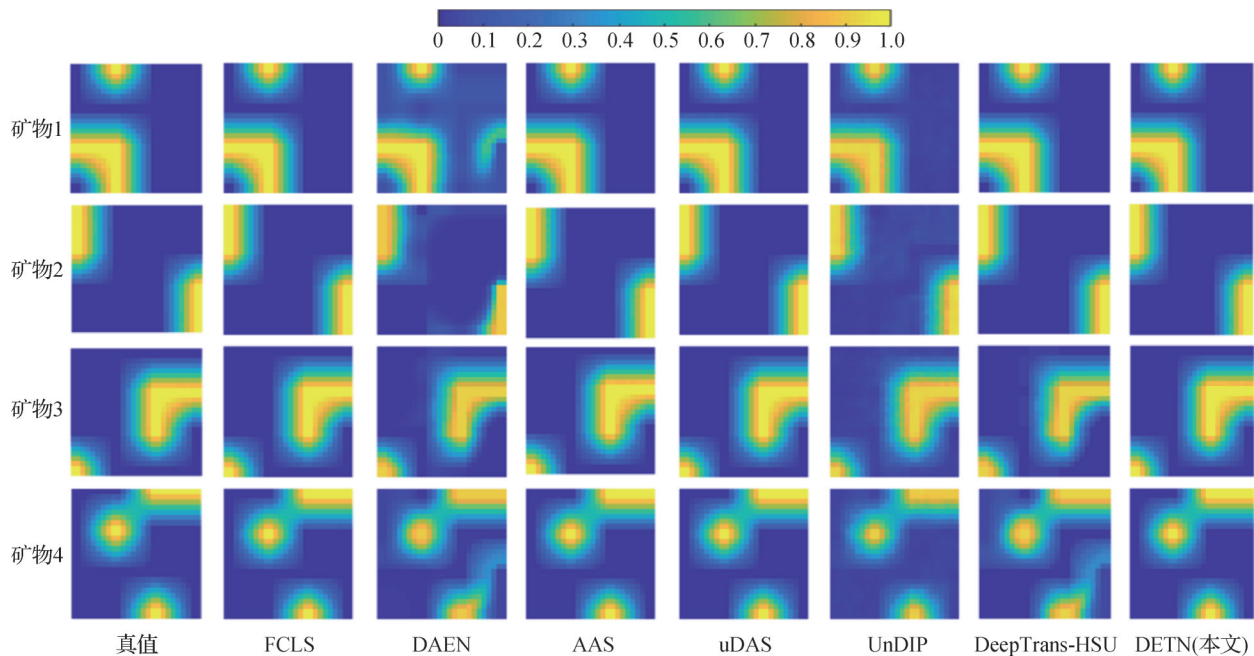


图 2 不同方法对模拟数据解混后的丰度图对比 (50 dB)

Fig. 2 Comparison of abundance maps obtained by different unmixing methods on the synthetic dataset (50 dB)

表 2 不同方法使用模拟数据实验的精度评定

Table 2 Accuracy evaluation of different unmixing methods using simulated dataset

方法	20 dB			30 dB			50 dB		
	SAD( $X, \hat{X}$ )	RMSE( $A, \hat{A}$ )	SAD( $W, \hat{W}$ )	SAD( $X, \hat{X}$ )	RMSE( $A, \hat{A}$ )	SAD( $W, \hat{W}$ )	SAD( $X, \hat{X}$ )	RMSE( $A, \hat{A}$ )	SAD( $W, \hat{W}$ )
FCLS	0.145 3	0.124 5	0.242 1	0.062 5	0.022 1	0.261 7	0.037 4	0.005 6	0.054 3
DAEN	0.132 6	0.121 7	0.231 5	0.058 3	0.011 7	0.247 8	0.026 5	0.007 5	0.041 2
AAS	0.132 8	0.121 1	0.231 1	0.057 6	0.012 3	0.235 5	0.019 8	0.007 1	0.041 0
uDAS	0.133 5	0.123 2	0.232 4	0.058 9	0.013 4	0.244 3	0.031 6	0.007 3	0.042 8
UnDIP	0.132 1	0.120 6	0.230 2	0.056 4	0.011 8	0.242 8	0.025 7	0.006 8	0.038 7
DeepTrans-HSU	0.130 9	0.117 6	0.228 7	0.055 2	0.010 5	0.241 7	0.019 8	0.005 9	0.039 8
DETN(本文)	<b>0.128 7</b>	<b>0.112 3</b>	<b>0.211 9</b>	<b>0.052 6</b>	<b>0.009 7</b>	<b>0.240 9</b>	<b>0.018 4</b>	<b>0.004 5</b>	<b>0.038 6</b>

注: 加粗字体表示各列最优结果。

2.2 使用真实数据的验证实验

2.2.1 实验1 在高光谱数据集 Samson 的验证实验

高光谱数据集 Samson 的原始图像数据具有  $952 \times 952$  个像元, 光谱范围覆盖  $0.4 \sim 0.9 \mu\text{m}$ , 包含了 156 个波段, 且具有高达  $3.13 \text{ nm}$  的光谱分辨率。实验中, 选取原始数据中像元大小为  $95 \times 95$  的图像数据, 该数据受噪声等干扰因素影响较少, 其中包括水体、植被和土壤 3 种端元。图 3 和图 4 分别展示了不同方法对 Samson 数据集解混后得到的端元光谱曲线对比和丰度图对比。可以看出, 本文提出的

表3 不同方法使用 Samson 数据集实验的精度评定  
Table 3 Accuracy evaluation of different unmixing methods using Samson dataset

方法	SAD( $X, \hat{X}$ )	RMSE( $A, \hat{A}$ )	SAD( $W, \hat{W}$ )
FCLS	0.124 3	0.178 9	0.144 5
DAEN	0.121 1	0.167 0	0.132 6
AAS	0.117 6	0.165 4	0.127 8
uDAS	0.123 5	0.168 5	0.136 9
UnDIP	0.115 2	0.164 2	0.125 9
DeepTrans-HSU	0.107 3	0.163 9	0.123 3
DETN(本文)	<b>0.104 9</b>	<b>0.161 2</b>	<b>0.119 4</b>

注:加粗字体表示各列最优结果。

DETN 方法较接近真实情况, 且能够获得较好的解混效果。不同方法使用 Samson 数据集实验的精度评定如表 3 所示, 可以看出, 本文方法通过 SAD、

RMSE 计算的精度结果明显优于其他方法, 说明

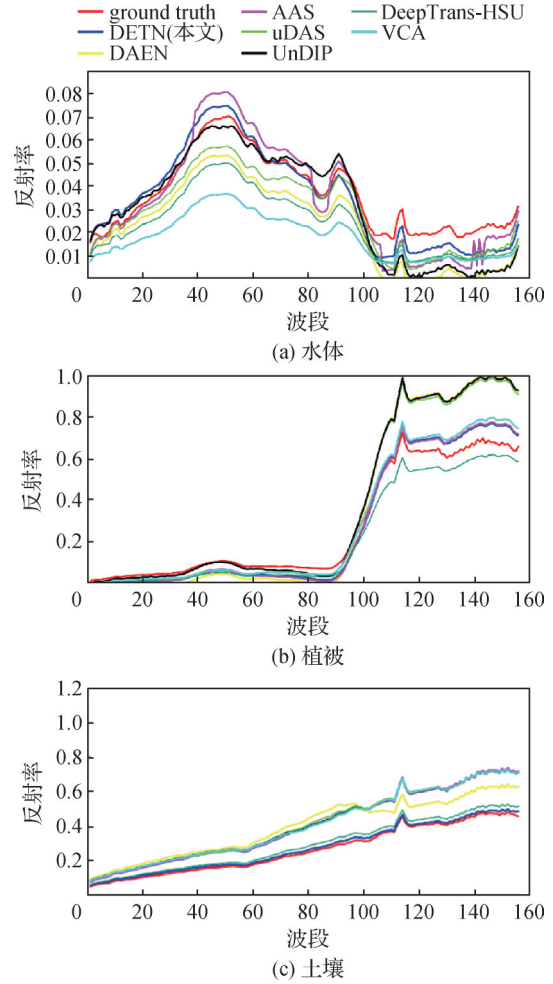


图3 不同方法对 Samson 数据集解混后的端元光谱曲线对比  
Fig. 3 Comparison of endmember spectral curves after unmixing the Samson dataset using different methods ((a)water; (b)vegetation; (c)soil)

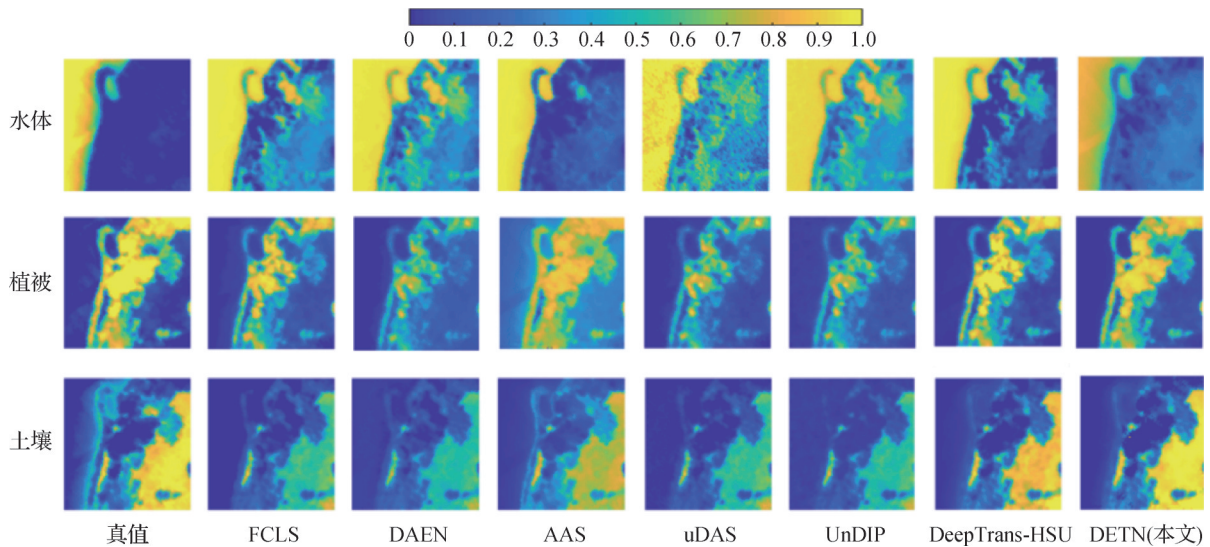


图4 不同方法对 Samson 数据解混后的丰度图对比

Fig. 4 Comparison of abundance maps after unmixing the Samson dataset using different methods

DETN方法能够获得较高精度的解混结果。

2.2.2 实验2 在高光谱数据集Jasper Ridge的验证实验

高光谱数据集Jasper Ridge由位于美国加利福尼亚州中部碧玉岭上的机载可见光/红外成像光谱仪 (airborne visible infrared imaging spectrometer,

AVIRIS)传感器采集,原始图像像素为512×217,包含224个波段,光谱范围介于0.4~2.5 μm之间。为排除水汽密度以及大气效应等因素的影响,实验选取了原始数据中100×100像素的图像部分,并保留其中的198个波段。这部分数据包括4种不同的端元:土壤、道路、植被和水体。图5和图6分别展示了

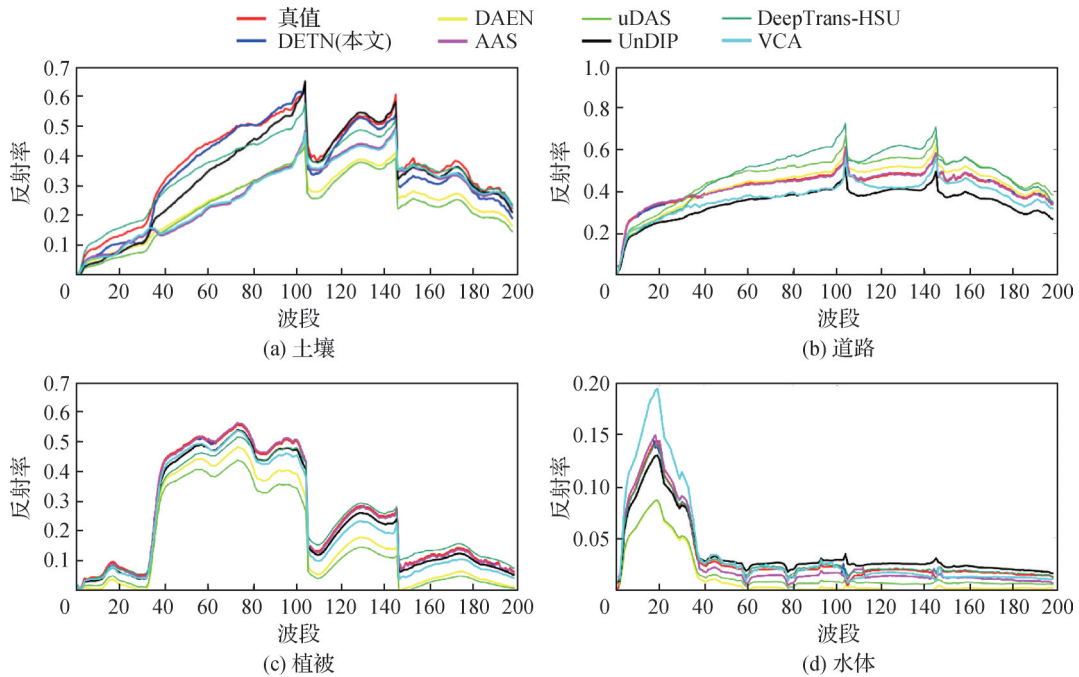


图5 不同方法对Jasper Ridge数据集解混后得到的端元光谱曲线对比

Fig. 5 Comparison of endmember spectral curves after unmixing the Jasper Ridge dataset using different methods ((a)soil;(b)road;(c)vegetation;(d)water)

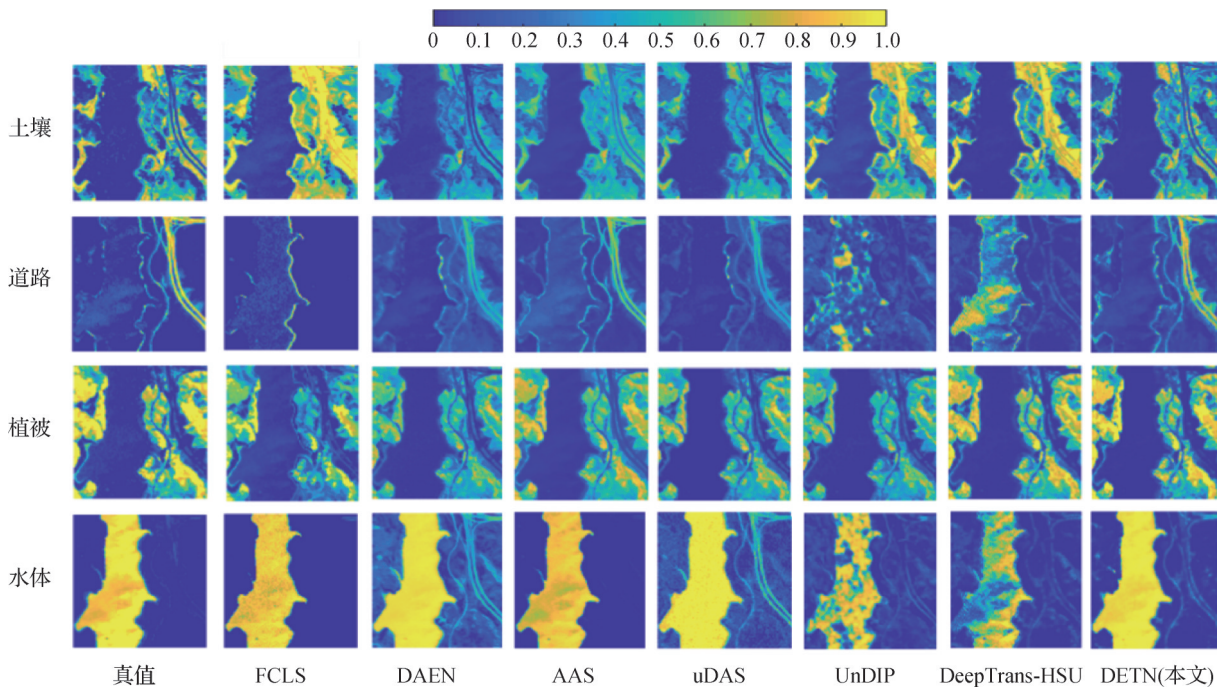


图6 不同方法对Jasper Ridge数据解混后得到的丰度图对比

Fig. 6 Comparison of abundance maps after unmixing the Jasper Ridge dataset using different methods

不同方法对 Jasper Ridge 数据集解混后得到的端元光谱曲线对比和丰度图对比。可以看出,本文提出的 DETN 方法较接近真实情况,且能够获得较好的解混效果。表4展示了不同方法使用 Jasper Ridge 数据集实验的精度评定。可以看出,本文方法在 SAD、RMSE 评价指标上计算得到的精度明显高于其他方法,这进一步验证了 DETN 方法的有效性。

表4 不同方法使用 Jasper Ridge 数据集实验的精度评定

Table 4 Accuracy evaluation of different unmixing methods using Jasper Ridge dataset

方法	SAD( $X, \hat{X}$ )	RMSE( $A, \hat{A}$ )	SAD( $W, \hat{W}$ )
FCLS	0.225 6	0.345 1	0.124 5
DAEN	0.179 4	0.278 0	0.116 6
AAS	0.154 3	0.276 4	0.110 1
uDAS	0.180 8	0.277 5	0.108 2
UnDIP	0.143 8	0.275 4	0.106 5
DeepTrans-HSU	0.129 9	0.273 2	0.105 8
<b>DETN(本文)</b>	<b>0.125 4</b>	<b>0.270 9</b>	<b>0.102 7</b>

注:加粗字体表示各列最优结果。

### 2.2.3 实验3 在南昌真实高光谱农田数据的验证实验

随着中国高光谱遥感技术的高速发展,高光谱成像技术被确定为我国未来重大科技专项规划的重点发展方向之一(王建宇和李春来,2021)。除了 Samson 和 Jasper Ridge 数据集,实验使用了北京盛世华遥科技有限公司提供的高分5号卫星获取的江西南昌部分真实高光谱农田数据。图7展示了该数据真彩色图像。由于受传感器分辨率、复杂地形等多方面条件的影响,获得的高分5号卫星影像在几何以及波段辐射值上都有所偏差。因此,选取了像元大小为  $100 \times 100$  的图像,并且经过辐射校正等预处理保留共150个波段,图像中包含3种端元:水体、农作物和裸土。

图8和图9分别展示了不同方法使用农田数据估算的端元光谱曲线对比和丰度图对比。虽然该农田数据没有可供参考的 ground truth 数据,但通过图8与图9呈现不同方法的解混对比图,并且结合图7进行定性分析,可以观察到 DETN 方法解混获得的丰度图更符合该数据真彩色合成图像中地物真实分布的情况。

### 2.2.4 实验4 法国马赛港高光谱数据的验证实验

实验使用了北京盛世华遥科技有限公司提供的



图7 江西南昌部分农田数据集真彩色图像

Fig. 7 Partial farmland dataset of true-color composite images in Nanchang, Jiangxi Province

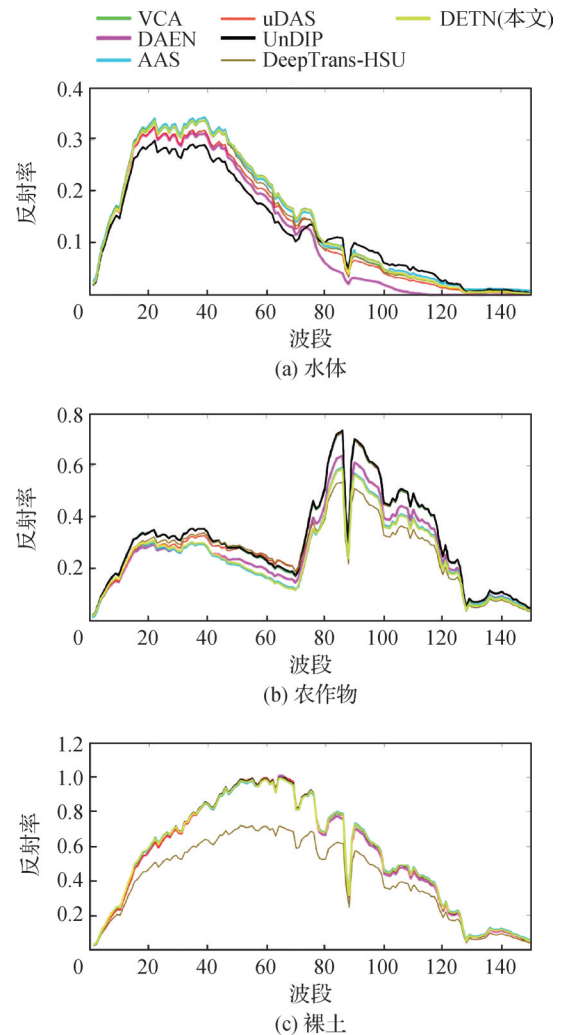


图8 不同方法对农田数据解混后的端元光谱曲线对比

Fig. 8 Comparison of endmember spectral curves after unmixing the farmland dataset using different methods ((a)water; (b)farmland; (c)soil)

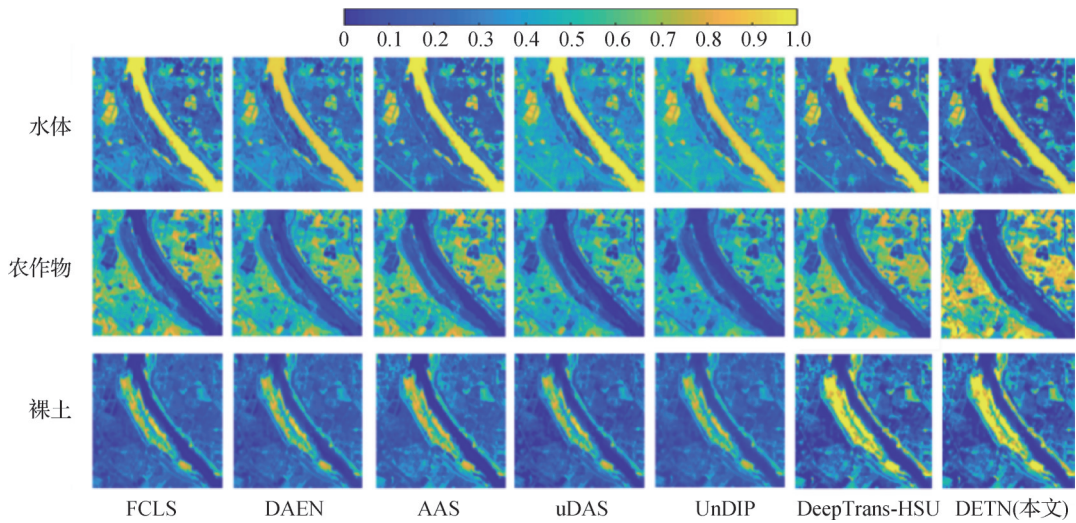


图9 不同方法对农田数据解混后得到的丰度图对比

Fig. 9 Comparison of abundance maps after unmixing the farmland dataset using different methods

资源1号02D(ZY1E)卫星获取的部分法国马赛港高光谱数据,图10展示了该数据真彩色图像。同样地,为了消除复杂成像条件所带来的影像偏差,选取了 $100 \times 100$ 像素的图像,并且经过辐射校正等预处理保留了共166个波段,图像中包含3种端元:水体、土壤和植被。图11和图12分别展示了不同方法对资源1号02D高光谱数据解混后得到的端元光谱曲线对比和丰度图对比。与高分5号农田数据相同,尽管该高光谱数据没有可供参考的ground truth数据,但是通过图11与图12呈现不同方法的解混对比图效果,并结合图10进行定性分析,可以看出提出的DETN方法解混得到的丰度图同样更加符合该数据真彩色合成图像中地物真实分布的情况。



图10 资源1号02D卫星高光谱数据真彩色图像

Fig. 10 True-color image of ZY1E satellite hyperspectral data

由实验1和实验2说明本文方法在使用解混验证实验常用的经典高光谱数据集进行解混时,与一

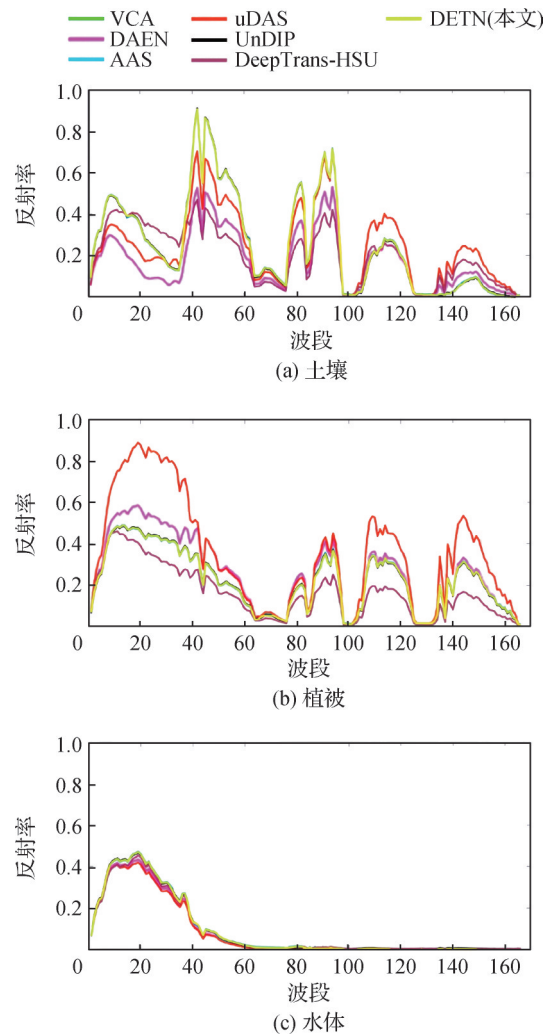


图11 不同方法对资源1号02D卫星高光谱数据解混后得到的端元光谱曲线对比

Fig. 11 Comparison of endmember spectral curves after unmixing the hyperspectral data from the ZY1E satellite using different methods((a)soil;(b)vegetation;(c)water)

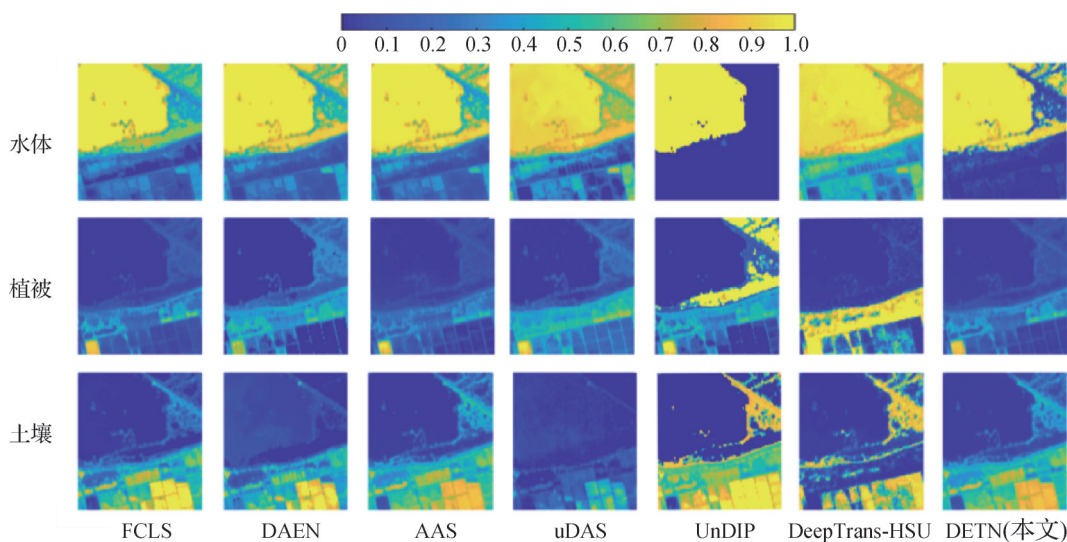


图12 不同方法对资源1号02D卫星高光谱数据解混后得到的丰度图对比

Fig. 12 Comparison of abundance maps after unmixing the hyperspectral data from the ZY1E satellite using different methods

些先进的解混算法相比能够获得较为理想的解混效果。而上述实验3和实验4则进一步说明该方法使用解混验证实验较少使用的国产高光谱数据集进行解混时,同样具有可行性以及有效性。

### 3 结论

本文提出的DET N方法以TNT网络构架为基础来实现高光谱图像的空一谱解混,能够同时保留像元的光谱信息以及图块内部像元之间的空间相关信息,获得了更多高光谱图像丰富的细节特征。在DET N中,不仅能够弥补仅使用传统Transformer模型进行解混学习容易忽略图块局部像元之间的空间相关性的问题,并且使用内外嵌套式Transformer结构减少结合Transformer模型与堆栈CNN网络层获取局部相关性特征容易发生过拟合的概率。此外,在DET N方法中,内部Transformer结构获取的局部像元空一谱信息直接聚合于外部Transformer结构中进行信息与位置编码共享,使得编码器输出包含局部像元之间的细节性空一谱信息以及完整的全局性空一谱信息,减少参数设置且显著提高了解混的学习性能。从而使得编码器能够为解码器更新、获取端元以及丰度提供丰富且有效的高光谱图像空一谱信息,最终获得较为理想的解混效果。最后,通过使用模拟数据集以及多个真实数据集充分验证本文DET N方法的有效性。

由于目前所研究的方法在挖掘图像的空间信息

方面较为突出,接下来,将在该方法的基础上更深入地挖掘像元之间的内在关系,并针对图像中像元可能存在的光谱变异性问题进行探究。

### 参考文献(References)

- Boardman J W, Kruse F A. and Green, R O. 1995. Mapping Target Signatures via Partial Unmixing of Aviris Data//Proceeding of the 5th Annual JPL Airborne Earth Science Workshop. [s.l.]: AVIRIS Workshop
- Borsoi R A, Imbiriba T and Bermudez J C M. 2020. Deep generative endmember modeling: an application to unsupervised spectral unmixing. *IEEE Transactions on Computational Imaging*, 6: 374-384 [DOI: 10.1109/TCL.2019.2948726]
- Eches O, Dobigeon N, Mailhes C and Tourmeret J Y. 2010. Bayesian estimation of linear mixtures using the normal compositional model. Application to hyperspectral imagery. *IEEE Transactions on Image Processing*, 19(6): 1403-1413 [DOI: 10.1109/TIP.2010.2042993]
- Fang Y, Wang Y X, Xu L L, Zhuo R M, Wong A and Clausi D A. 2022. BCUN: Bayesian fully convolutional neural network for hyperspectral spectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #5523714 [DOI: 10.1109/TGRS.2022.3151004]
- Gao L R, Han Z, Hong D F, Zhang B and Chanussot J. 2022a. CyCU-Net: cycle-consistency unmixing network by learning cascaded autoencoders. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #5503914 [DOI: 10.1109/TGRS.2021.3064958]
- Gao L R, Wang Z C, Zhuang L N, Yu H Y, Zhang B and Chanussot J. 2022b. Using low-rank representation of abundance maps and non-negative tensor factorization for hyperspectral nonlinear unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 60:

- #5504017 [DOI: 10.1109/TGRS.2021.3065990]
- Ghosh P, Roy S K, Koirala B, Rasti B and Scheunders P. 2022. Hyperspectral unmixing using Transformer network. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #5535116 [DOI: 10.1109/TGRS.2022.3196057]
- Guo R, Wang W and Qi H R. 2015. Hyperspectral image unmixing using autoencoder cascade//7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS). Tokyo, Japan: IEEE: 1-4 [DOI: 10.1109/WHISPERS.2015.8075378]
- Han K, Xiao A, Wu E H, Guo J Y, Xu C J and Wang Y H. 2021. Transformer in Transformer//Proceedings of the 35th Conference on Neural Information Processing Systems. [s.l.]: [s.n.]: 15908-15919
- Heinz D C and Chein-I-Chang. 2001. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(3): 529-545 [DOI: 10.1109/36.911111]
- Hua Z Q, Li X R, Qiu Q H and Zhao L Y. 2021. Autoencoder network for hyperspectral unmixing with adaptive abundance smoothing. *IEEE Geoscience and Remote Sensing Letters*, 18(9): 1640-1644 [DOI: 10.1109/LGRS.2020.3005999]
- Jin Q W, Ma Y, Mei X G and Ma J Y. 2022. TANet: an unsupervised two-stream autoencoder network for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #5506215 [DOI: 10.1109/TGRS.2021.3094884]
- Nascimento J M P and Dias J M B. 2005. Vertex component analysis: a fast method to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4): 898-910 [DOI: 10.1109/TGRS.2005.844293]
- Miao L, and Qi H. 2007. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 45(3): 765-777 [DOI: 10.1109/TGRS.2006.888466]
- Ozkan S, Kaya B and Akar G B. 2019. EndNet: sparse autoencoder network for endmember extraction and hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1): 482-496 [DOI: 10.1109/TGRS.2018.2856929]
- Palsson B, Ulfarsson M O and Sveinsson J R. 2021. Convolutional autoencoder for spectral-spatial hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 59(1): 535-549 [DOI: 10.1109/TGRS.2020.2992743]
- Peng Q, Zhang B, Sun X, Gao L R and Yu W B. 2017. Hyperspectral unmixing based on spatial and spectral preprocessing prior and constrained non-negative matrix factorization. *Journal of Image and Graphics*, 22(4): 542-550 (彭倩, 张兵, 孙旭, 高连如, 于文博. 2017. 结合空间光谱预处理和约束非负矩阵分解的高光谱图像混合像元分解. *中国图象图形学报*, 22(4): 542-550) [DOI: 10.11834/jig.20170414]
- Qi L, Chen Z W, Gao F, Dong J Y, Gao X B and Du Q. 2023. Multiview spatial-spectral two-stream network for hyperspectral image unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 61: #5502016 [DOI: 10.1109/TGRS.2023.3237556]
- Qian Y T, Jia S, Zhou J and Robles-Kelly A. 2011. Hyperspectral unmixing via  $L_{1/2}$  sparsity-constrained nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11): 4282-4297 [DOI: 10.1109/TGRS.2011.2144605]
- Qu H C, Ji R Q, Liu W J and Liang X J. 2015. Acceleration of hyperspectral image endmember extraction based on MapReduce pattern. *Journal of Image and Graphics*, 20(7): 973-980 (曲海成, 籍瑞庆, 刘万军, 梁雪剑. 2015. MapReduce 模式下高光谱图像端元提取算法加速. *中国图象图形学报*, 20(7): 973-980) [DOI: 10.11834/jig.20150714]
- Qu Y and Qi H R. 2019. uDAS: an untied denoising autoencoder with sparsity for spectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 57(3): 1698-1712 [DOI: 10.1109/TGRS.2018.2868690]
- Rasti B, Koirala B, Scheunders P and Ghamisi P. 2022. UnDIP: hyperspectral unmixing using deep image prior. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #5504615 [DOI: 10.1109/TGRS.2021.3067802]
- Song X R, Wu L D and Meng X L. 2020. Unsupervised hyperspectral unmixing based on robust non-negative matrix factorization. *Journal of Image and Graphics*, 25(4): 801-812 (宋晓瑞, 吴玲达, 孟祥利. 2020. 利用稳健非负矩阵分解实现无监督高光谱解混. *中国图象图形学报*, 25(4): 801-812) [DOI: 10.11834/jig.190354]
- Su Y C, Gao L R, Jiang M Y, Plaza A, Sun X and Zhang B. 2023. NSCKL: normalized spectral clustering with kernel-based learning for semisupervised hyperspectral image classification. *IEEE Transactions on Cybernetics*, 53(10): 6649-6662 [DOI: 10.1109/TCYB.2022.3219855]
- Su Y C, Li J, Plaza A, Marinoni A, Gamba P and Chakravorty S. 2019. DAEN: deep autoencoder networks for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7): 4309-4321 [DOI: 10.1109/TGRS.2018.2890633]
- Su Y C, Marinoni A, Li J, Plaza J and Gamba P. 2018. Stacked non-negative sparse autoencoders for robust hyperspectral unmixing. *IEEE Geoscience and Remote Sensing Letters*, 15(9): 1427-1431 [DOI: 10.1109/LGRS.2018.2841400]
- Su Y C, Xu X, Li J, Qi H R, Gamba P and Plaza A. 2021. Deep autoencoders with multitask learning for bilinear hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 59(10): 8615-8629 [DOI: 10.1109/TGRS.2020.3041157]
- 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]
- Sun Y F, Liu B, Wang R R, Zhang P Q and Dai M F. 2023. Spectral-

- spatial MLP-like network with reciprocal points learning for open-set hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61: #5513218 [DOI: 10.1109/TGRS.2023.3280183]
- Tong Q X, Zhang B and Zhang L F. 2016. Current progress of hyperspectral remote sensing in China. *Journal of Remote Sensing*, 20(5): 689-707 (童庆禧, 张兵, 张立福. 2016. 中国高光谱遥感的前沿进展. *遥感学报*, 20(5): 689-707) [DOI: 10.11834/jrs.20166264]
- Tu B, Liao X L, Li Q M, Peng Y S and Plaza A. 2022. Local semantic feature aggregation-based Transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #5536115 [DOI: 10.1109/TGRS.2022.3201145]
- Tu B, Ren Q, Li Q M, He W Q and He W. 2023. Hyperspectral image classification using a superpixel-pixel-subpixel multilevel network. *IEEE Transactions on Instrumentation and Measurement*, 72: #5013616 [DOI: 10.1109/TIM.2023.3271713]
- Wang J Y and Li C L. 2021. Development and prospect of hyperspectral imager and its application. *Chinese Journal of Space Science*, 41(1): 22-33 (王建宇, 李春来. 2021. 高光谱遥感成像技术的发展与展望. *空间科学学报*, 41(1): 22-33) [DOI: 10.11728/cjss2021.01.022]
- Wang S Q, Yang J X, Shao Y T and Xiao L. 2023. Non-negative sparse component decomposition based modeling and robust unmixing for hyperspectral images. *Journal of Image and Graphics*, 28(2): 613-627 (汪顺清, 杨劲翔, 邵远天, 肖亮. 2023. 高光谱图像非负稀疏分量分解建模与鲁棒性解混方法. *中国图象图形学报*, 28(2): 613-627) [DOI: 10.11834/jig.211054]
- Winter M E. 1999. N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data//*Proceedings of SPIE 3753, Imaging Spectrometry V*. Denver, USA: SPIE: 266-275 [DOI: 10.1117/12.366289]
- Zhang B. 2016. Advancement of hyperspectral image processing and information extraction. *Journal of Remote Sensing*, 20(5): 1062-1090 (张兵. 2016. 高光谱图像处理与信息提取前沿. *遥感学报*, 20(5): 1062-1090) [DOI: 10.11834/jrs.20166179]

## 作者简介

游雪儿,女,硕士研究生,主要研究方向为深度学习和混合像元分解。E-mail: xeryou2022@163.com

苏远超,通信作者,男,副教授,硕士生导师,主要研究方向为高光谱遥感数据信息提取、遥感智能解译、深度学习。

E-mail: suych3@xust.edu.cn

蒋梦莹,女,博士研究生,主要研究方向为深度学习、遥感图像处理、药品智能检测、机器学习。

E-mail: mengyingjiangdut@foxmail.com

李朋飞,男,教授,博士生导师,主要研究方向为地貌遥感、激光点云处理与分析、机器学习。E-mail: pengfeili@xust.edu.cn

刘东升,男,高级工程师,主要研究方向为遥感图像处理、遥感软件开发。E-mail: ldsh1201@163.com

白晋颖,男,高级讲师,主要研究方向为遥感图像处理、大数据云计算。E-mail: baijinying@piesat.cn