

中图法分类号: TP751 文献标识码: A 文章编号: 1006-8961(2024)07-2063-12

论文引用格式: 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(07):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(07):2063-2074)[DOI: 10. 11834/jig. 230381]

基于Transformer和通道混合并行卷积的高光谱图像去噪

胡帅^{1,2}, 高峰^{1,2*}, 龚卓然¹, 陶盛恩^{1,2}, 上官心语^{1,2}, 董军宇^{1,2}

1. 中国海洋大学计算机科学与技术学院, 青岛 266100; 2. 中国海洋大学三亚海洋研究院, 三亚 572025

摘要: 目的 高光谱图像因设备及环境因素容易受到噪声污染, 导致图像的可见性和分析精度降低, 因此高光谱图像去噪任务已经成为遥感图像处理领域国内外研究热点。当前的高光谱图像去噪方法主要面临两个难题: 1) 对特征的全局信息利用不足。当前基于卷积神经网络的方法受限于卷积核的大小, 难以捕获特征的全局信息; 2) 卷积神经网络和Transformer在结构上存在差异, 导致两者难以融合, 因此, 需要考虑合理的特征交互方式, 来平衡局部和全局特征提取之间的关系。**方法** 针对上述问题, 本文提出了基于Transformer和通道混合并行卷积的高光谱图像去噪模型, 包括3个模块: 通道混合特征提取模块、基于块下采样的全局增强模块和自适应双向特征融合模块。通过这3个模块的相互作用, 可以充分结合全局和局部的特征信息, 处理不同区域中的噪声和纹理差异, 有效提高模型对空间细节信息的恢复能力。**结果** 实验在2个数据集上与主流的5种方法进行比较, 在Pavia数据集中设置不同高斯噪声强度的情况下, 相比于性能第2的模型, 峰值信噪比(peak signal-to-noise ratio, PSNR)值最大提高了0.4 dB; 在ICVL数据集中设置各种混合噪声的情况下, 相比于性能第2的模型, PSNR最大提高了2.18 dB。同时可视化的去噪结果图像体现了本文所提出的去噪模型的优异性能。**结论** 本文方法在各种噪声情况下均具有较好的去噪效果, 显著优于当前主流方法, 能够有效去除高光谱图像中噪声, 同时保留图像丰富的纹理信息。

关键词: 高光谱图像去噪; 通道混合; Transformer; 特征融合; 全局注意力

Parallel channel shuffling and Transformer-based denoising for hyperspectral images

Hu Shuai^{1,2}, Gao Feng^{1,2*}, Gong Zhuoran¹, Tao Shengen^{1,2}, ShangGuan Xinyu^{1,2}, Dong Junyu^{1,2}

1. School of Computer Science and Technology, Ocean University of China, Qingdao 266100, China;

2. Sanya Oceanographic Institution, Ocean University of China, Sanya 572025, China

Abstract: Objective With the increasing availability and advancement of hyperspectral imaging technology, hyperspectral images have become an invaluable resource in various fields, including agriculture, environmental monitoring, and remote sensing. However, these images are often prone to noise contamination, which can significantly degrade their quality and hinder accurate analysis and interpretation. As a result, denoising hyperspectral images has become a crucial task in the field of remote sensing image processing, attracting significant attention from researchers worldwide. The challenges associ-

收稿日期: 2023-06-19; 修回日期: 2023-10-25; 预印本日期: 2023-11-02

* 通信作者: 高峰 gaofeng@ouc.edu.cn

基金项目: 新一代人工智能国家科技重大专项项目(2022ZD0117202); 青岛市自然科学基金项目(23-2-1-222-zyyd-jch)

Supported by: National Science and Technology Major Project (2022ZD0117202); Natural Science Foundation of Qingdao, China (23-2-1-222-zyyd-jch)

ated with denoising hyperspectral images are multifaceted. First, the inherent characteristics of hyperspectral data, such as high dimensionality and complex spectral information, pose significant difficulties for traditional denoising approaches. The presence of noise in hyperspectral images can obscure valuable information embedded within the spectral bands, making it essential to develop advanced denoising techniques that can effectively restore the original signal while preserving the rich texture and spatial details. Furthermore, the development of deep learning techniques, particularly convolutional neural networks (CNNs), has revolutionized the field of image processing, including denoising tasks. CNN-based approaches have shown promising results in denoising various types of images. However, when it comes to hyperspectral data, traditional CNN architectures face limitations in capturing the global contextual information necessary for accurate denoising. The fixed-size receptive fields of CNNs restrict their ability to exploit the spatial and spectral correlations present in hyperspectral images, thereby reducing their overall denoising performance. To overcome these limitations, recent research has explored the integration of Transformers, which were originally designed for natural language processing tasks, into the field of computer vision, including hyperspectral image denoising. Transformers are capable of capturing long-range dependencies and global contextual information, making them an attractive alternative to CNNs for denoising tasks. However, directly applying Transformer-based models to hyperspectral data requires careful consideration of the specific challenges posed by the unique characteristics of hyperspectral images.

Method In this study, we propose a novel denoising model for hyperspectral images that combines the strengths of Transformers and parallel convolution operations. Our model comprises three key modules: channel shuffling module, block downsampling global enhancement module, and adaptive bidirectional feature fusion module. These modules work synergistically to address the challenges encountered in denoising hyperspectral images. The channel shuffling module exploits the inter-channel relationships within hyperspectral data by incorporating channel-mixing operations. By fusing information across different spectral channels, the module enhances the representation power of the network and enables more comprehensive feature extraction. This approach effectively addresses the limitation of traditional CNN-based methods in fully utilizing the global information available in hyperspectral images, ultimately improving the model's denoising performance. In the block downsampling global enhancement module, we leverage a block downsampling strategy to capture global contextual information. By reducing the spatial resolution of the input hyperspectral image, the module enlarges the receptive fields, allowing the model to incorporate larger-scale information during the denoising process. This mechanism enhances the model's understanding of the overall structure of the image, facilitating more effective noise suppression and accurate restoration of spatial details. The adaptive bidirectional feature fusion module is designed to strike a balance between local and global feature extraction, leveraging the complementary strengths of CNNs and Transformers. This module introduces a mechanism for adaptively fusing features from local and global contexts, enabling the model to effectively combine local details with global information. By considering the intricate relationship between spatial and spectral features, our proposed approach improves the denoising performance and preserves the rich texture information inherent in hyperspectral images.

Result To evaluate the effectiveness of our proposed model, extensive experiments were conducted on publicly available hyperspectral image datasets, including ICVL and Pavia. Experimental results demonstrated the superior denoising performance of our approach compared with that of current state-of-the-art methods. Our model consistently outperformed existing techniques in various noise scenarios, effectively removing noise while preserving the fine spatial details and rich texture information of hyperspectral images. The experimental evaluation involved quantitative metrics such as peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and spectral angle mapping (SAM). Our proposed model achieved significantly higher PSNR values and SSIM scores compared with the baseline methods, indicating improved denoising accuracy and visual quality of the restored images. In addition, the SAM values obtained using our model were consistently lower, indicating higher spectral similarity. Moreover, we conducted a comprehensive analysis of the computational efficiency of our model. With the increasing volume and complexity of hyperspectral data, developing denoising methods that are computationally efficient without sacrificing performance is crucial. Our proposed model demonstrated competitive computational efficiency, making it practical for real-world applications that involve large-scale hyperspectral image processing.

Conclusion The success of our denoising model can be attributed to the synergistic combination of the Transformer-based architecture and the channel-mixing parallel convolution operations. The Transformer module enables effective capture of global contextual information, facili-

tating better understanding of the relationships between spectral bands and spatial features. By incorporating channel-mixing operations, our model exploits the inter-channel correlations and enhances the discriminative power of feature extraction, resulting in improved denoising performance. Furthermore, our model's ability to handle diverse noise scenarios and maintain image quality can be attributed to the adaptive bidirectional feature fusion module. This module intelligently combines local and global features, enabling effective noise suppression while preserving the fine details and texture information specific to different regions of the hyperspectral images. The adaptability of the feature fusion mechanism ensures robust denoising performance across various noise levels and image characteristics. In conclusion, this study presents a novel denoising model for hyperspectral images based on the integration of Transformers and channel-mixing parallel convolution. The proposed model effectively addresses the limitations of traditional approaches in utilizing global information and captures the complex spatial-spectral correlations inherent in hyperspectral data. Experimental results demonstrate its superior denoising performance compared with that of state-of-the-art methods, with improved accuracy and preservation of fine details and texture information. The model's computational efficiency further enhances its practicality for real-world applications. Future research directions may include exploring additional mechanisms for adaptive feature fusion and investigating the model's performance on other hyperspectral image processing tasks such as classification and segmentation.

Key words: hyperspectral image denoising; channel shuffling; Transformer; feature fusion; global attention

0 引言

高光谱遥感图像是使用成像光谱仪同时对地物数以百计的光谱波段成像而获得的数据,包含地物的空间结构信息和光谱信息(Yu 等, 2023; Hao 等, 2023; Gelvez-Barrera 等, 2022)。近年来,随着高光谱成像技术的不断发展,高光谱图像已经成为遥感和地球科学领域中非常重要的数据类型之一(Zhao 等, 2022; Li 等, 2022; Xiong 等, 2022)。然而,在获取高光谱图像的过程中,受到内部设备和外部环境的影响,会引入多种类型的噪声,从而降低高光谱数据的质量和可靠性,影响数据的进一步分析和应用(分类、解混和目标检测等)。因此,高光谱去噪任务已成为遥感数据应用领域的研究热点。

针对高光谱图像去噪任务,国内外学者提出了大量方法。最初,高光谱图像去噪研究是在自然图像去噪的基础上进行的,Chen 等人(2022)采用一种逐光谱通道的灰度图像去噪方法,这种方法主要基于滤波算子,其原理是通过非局部均值滤波或傅里叶变换来分离干净图像和噪声信号。除了传统的滤波算子外,研究者们也提出了多种基于小波收缩、稀疏近似和协同过滤等技术的高光谱去噪方法。Othman 和 Qian(2006)提出了一种利用光谱域和空间域特征不相似性的小波收缩方法,结合了光谱信息和空间信息来进行高光谱去噪。Maggioni 等人(2013)提出了一种基于非局部变换的滤波器,用于对三维

体积数据进行去噪。利用非局部相似性来构建一组权重以表示不同小块之间的相关性,以此来增强噪声的去除效果。尽管这些方法在高光谱去噪方面取得了一定的进展,但是它们只能处理单一类型的噪声,难以去除高光谱图像中的混合噪声。

随着海量遥感数据的获取和计算机硬件平台的快速发展,深度学习技术已经成为解决高光谱图像修复领域的主流方法。卷积神经网络(convolutional neural network, CNN)因其局部相关性和平移不变性,在图像处理领域表现出优秀的性能。基于卷积神经网络的方法在图像去噪任务上取得了非常好的效果。Yuan 等人(2019)提出了一种基于残差卷积神经网络的高光谱图像去噪方法(deep combined spatial-spectral convolutional neural network, HSID-CNN),该方法采用多尺度特征提取和多层次特征表示来捕获多尺度空间光谱特征,表现出优秀的性能。Wei 等人(2021)提出了一种基于 3D 卷积和循环交替的高光谱图像去噪算法(3D quasi-recurrent neural network, QRNN3D)。其中基于 3D 卷积的模块(3D quasi-recurrent unit, QRU3D)用于提取空间和光谱结构的关联特征,其利用交替循环的结构提取光谱之间的相关信息。Shi 等人(2021)使用双注意力去噪网络对高光谱图像进行去噪,该网络设计了两个并行分支,分别用于处理空间信息和光谱信息,并充分利用空间信息与光谱信息之间的全局相关性和局部相关性来提升去噪效果。

尽管当前基于深度学习的高光谱图像去噪方法

取得了良好的性能,但是仍面临两个关键难题:

1)对特征的全局信息利用不足。卷积操作具有局部相关性,受限于卷积核的大小,只利用感受野范围内的信息来帮助图像像素恢复,而感受野之外的信息没有充分利用,难以捕获全局的特征信息。Transformer得益于其强大的自注意力机制(Vaswani等,2017;Wang等,2023;Zhang等,2023),可以提取到全局信息,结合CNN模型和Transformer模型获取图像的局部信息和全局信息,可以进一步强化模型的特征表达能力。

2)CNN和Transformer结构存在差异,导致两者难以融合。由于Transformer和CNN在结构上存在明显差异,图像在两个模型之间的特征图结构不同,因此需要考虑合理的特征交互方式,使模型既能保留图像本身结构,又能充分利用长短距离之间的依赖关系。

针对上述问题,本文提出了基于Transformer和通道混合并行的高光谱图像去噪模型(parallel channel shuffling former, PSFormer)。为了充分利用高光谱图像的光谱信息,设计了一个通道混合特征提取模块(channel shuffling module, CSM),该模块主要是通过多尺度通道混合来提取高光谱图像的局部特征。同时,为了提取全局信息,设计了基于块下采样的全局增强模块(block-downsampling global enhancement module, DGEM),该模块是基于移动窗口的Transformer实现的,在窗口内进行自注意力计算,为了整合邻近像素之间的信息,同时减少计算量,针对窗口内的自注意力采用划分块的形式,可以获取更广的感受范围,整合更远距离的特征信息。

最后为了更好地整合全局增强模块和通道混合特征提取模块的特征,使得局部信息和全局特征得以充分利用,提出了自适应双向特征融合模块(adaptive bidirectional feature fusion module, ABFM)。该模块由两个部分组成:全局特征融合模块(global feature fusion module, GFM)和局部特征融合(local feature fusion module, LFM)模块,该模块可以更好地整合全局增强模块和通道混合特征提取模块的特征,使得局部信息和全局特征得以充分利用。

为验证模型的有效性,在ICVL和Pavia两个高光谱数据集上进行了实验。结果显示,本文提出的模型与国内外主流方法相比,具有优异的去噪性能。

本文的主要贡献如下:1)提出通道混合特征提取模块。将3个尺度提取到的特征拼接作为模块的输出,多尺度通道混合模块的分组通道混合在减少计算量的同时保证了通道之间的信息流动。2)提出基于块下采样的全局增强模块。对窗口内的自注意力采用划分块的形式逐步下采样,维持通道数不变的同时,可以整合更远距离的特征信息。3)提出自适应双向特征融合模块。通道混合特征提取模块能够有效捕获局部细节信息,基于块下采样的全局增强模块能获取全局特征,使用自适应双向特征融合模块,能够使得两边的信息得以交互。

1 本文方法

1.1 总体框架

本文提出的PSFormer整体流程如图1所示,整个架构由两个并行的分支组成:第1个分支由4个通

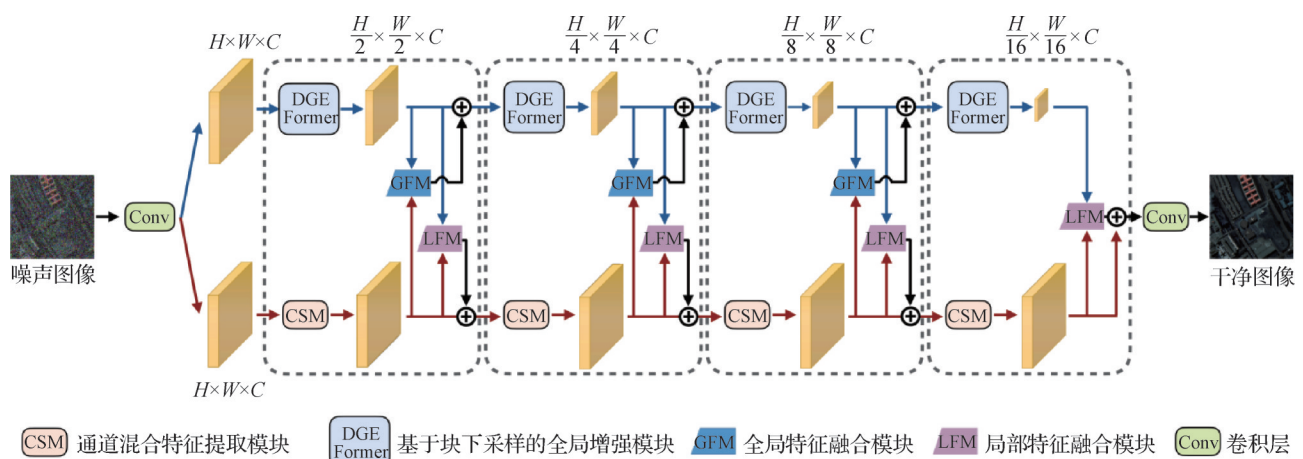


图1 基于Transformer和通道混合并行的高光谱图像去噪网络架构

Fig. 1 Hyperspectral image denoising network based on Transformer and parallel channel shuffling

道混合特征提取模块组成,用于提取近邻空间和光谱之间的依赖关系,这些特征提取模块的输入输出特征图大小相同;第2个分支则由4个基于块下采样的全局增强模块组成,以获取更长距离的依赖关系。通过全局增强模块对特征图进行处理,可以将其尺寸大小降至原来的 $1/2 \times 1/2$,使得在执行窗口注意力时能获取更大的感受野。在这两个分支之间,搭建了基于注意力的自适应双向特征融合模块,该模块由全局特征融合和局部特征融合组成,以促进局部信息和全局信息之间的交互。在前3个阶段,自适应双向特征融合模块发挥双向交流的作用,而在最后一个阶段完成时,仅有局部特征融合模块发挥作用,以整合长短距离依赖。最后,在并行架构的首尾分别使用卷积层,帮助噪声图像的特征转化以及生成干净图像。这些设计能够提高模型的鲁棒性和准确性,提升了去噪效果。

1.2 通道混合特征提取模块

高光谱图像与普通相机获取的自然图像不同,

其通道数能达到数十甚至上百个(Li等,2023;Maffei等,2020)。因此,对其进行去噪处理需要更复杂的网络模型和更多的参数。本文提出通道混合特征提取模块 CSM 对去噪网络模型进行优化,以提高效率并减少冗余连接,其结构可用图2表示。

多尺度通道混合模块分组后打乱组内通道,让不同组内的信息得以混合,同时利用不同大小的分组方式获取多尺度关系。设定用 g 来表示分组的组数,在 $g = 4$ 的分组通道混合中,含有 64 个通道的特征图被平均分为 4 组,每一组 16 个通道,每一组都各自经过一次卷积核大小为 3×3 的卷积层和 Leaky Relu 激活函数层后,4 个组的特征图进行通道混合,每个组的特征信息都混合到其他组中,4 个组分别经过卷积层,并且与模块输入相加。

多尺度通道混合中分别设定了 3 个尺度来提取多尺度的特征信息,当 $g = 1$ 时,认为不对其进行通道混合操作,用于捕获整体的全局特征。当 $g = 2$ 时,是指将通道分成两组进行混合,这种混合方式可

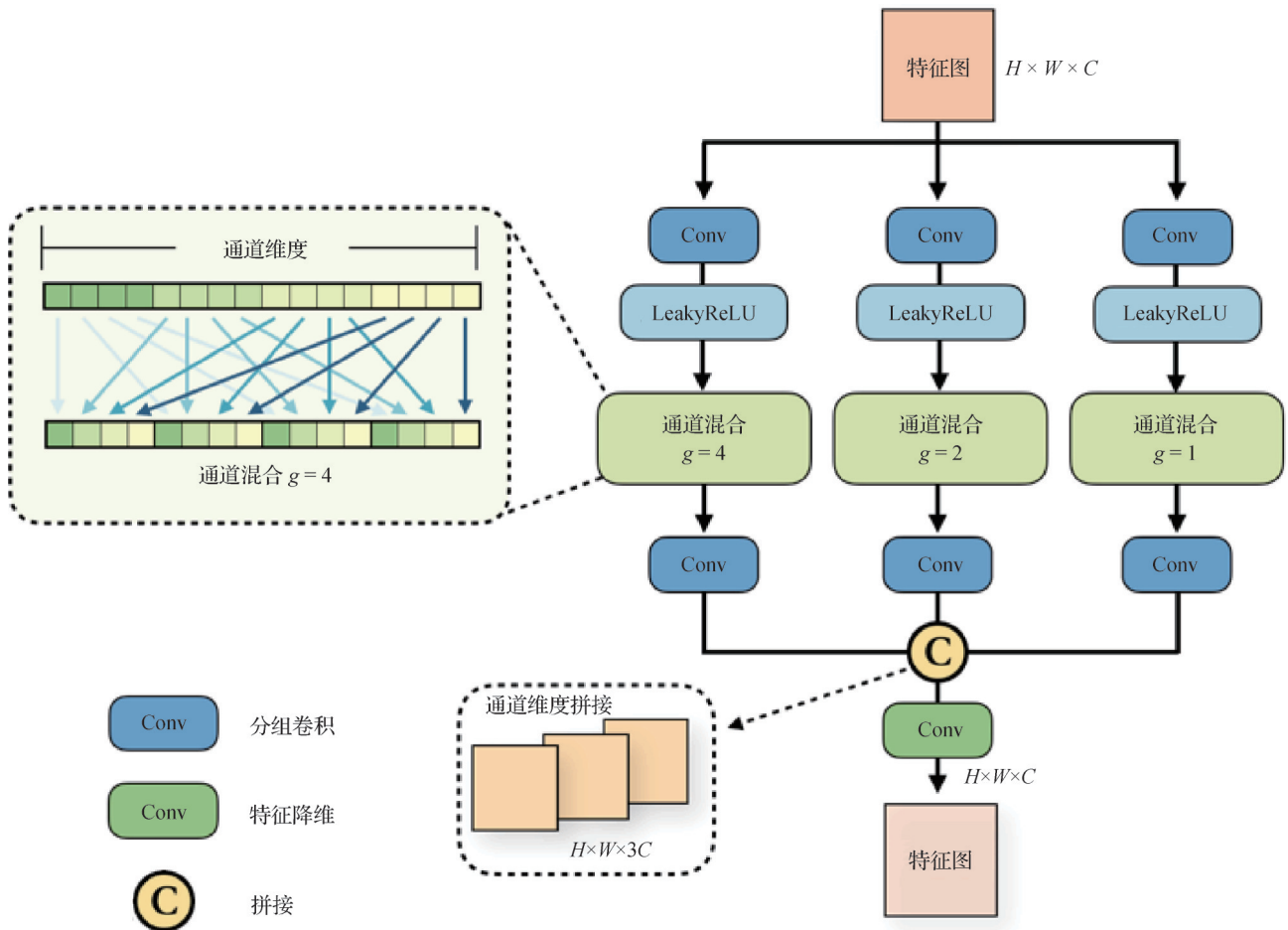


图2 通道混合特征提取模块

Fig. 2 Channel shuffling feature extraction module

以在一定程度上增加模型对多尺度特征的感知能力,当 $g = 4$ 时则是将通道混合成4组进行混合,可以更细致地考虑多个尺度的特征信息。每组通道可以对应不同的感受野和语义信息。这种混合方式能够更全面地捕获不同尺度下的目标、纹理和结构信息。将3个尺度提取到的特征拼接作为模块的输出,多尺度通道混合模块的分组通道混合在减少计算量的同时保证了通道之间的信息流动。

1.3 基于块下采样的全局增强模块

为了整合邻近像素之间的信息,获取更广的感受范围,同时减少计算量,本文提出了基于块下采样的全局增强模块DGEM,主要是针对窗口内的自注意

力采用划分块的形式,逐步进行下采样,维持通道数不变,整合更远距离的特征信息。该模块如图3所示。

输入特征 X 按照窗口划分后大小为 $H \times W \times C$,设定块(patch)的大小为 2×2 ,在经过下采样模块时,特征图先按照patch大小划分为块,每个块大小为 $2 \times 2 \times C$,随后通过全连接下降到 $1 \times 1 \times C$ 。将每一个块作为一个向量(token),在窗口内做自注意力得到输出大小为 $H/2 \times W/2 \times C$ 。设定窗口大小为 16×16 ,初始特征大小为 $128 \times 128 \times 64$,则4个基于块下采样的全局增强模块的输入特征大小分别是 $128 \times 128 \times 64, 64 \times 64 \times 64, 32 \times 32 \times 64$ 和 $16 \times 16 \times 64$,窗口数量分别是64、16、4和1。

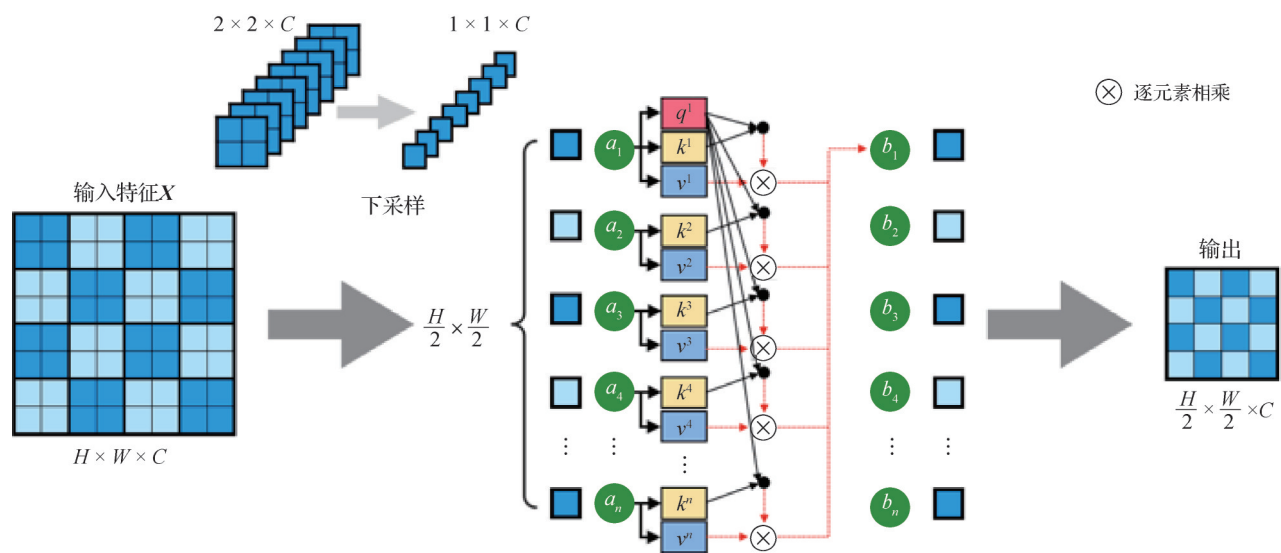


图3 基于块下采样的全局增强模块

Fig. 3 Global enhancement module based on block-downsampling

1.4 自适应双向特征融合模块

为了更好地整合全局增强模块和通道混合特征提取模块的特征,使得局部信息和全局特征得以充分利用,本文提出了自适应双向特征融合模块ABFM,其由两个部分组成:全局特征融合模块和局部特征融合模块,结构如图4所示。

通道混合特征提取模块能够有效捕获局部细节信息,基于块下采样的全局增强模块能获取全局特征,使用自适应双向特征融合模块,能够使得两边的信息得以交互。其中,全局特征融合模块的作用是将通道混合特征提取模块提取到的局部特征细节通过注意力融合的方式,整合到全局信息中;局部特征融合模块则与之相反。

在全局特征融合模块中有两个输入 X 和 Z ,分

别代表局部信息特征图和全局信息特征图。考虑将局部信息整合到全局信息中,认为全局信息 Z 作为查询向量 Q ,局部信息 X 作为关键值向量 K 和值向量 V 。对于全局信息 Z ,对其进行线性映射生成查询向量 Q 。为了减少计算负担,与关键值向量 X 相乘之后送入softmax对每个位置的特征值进行归一化处理,生成对应权重,再与值向量做加权求和处理,得到的相关特征图做线性映射后与全局信息 Z 相加,得到新的全局信息 Z' 。过程可用式(1)表示为

$$Z' = GFM(Z, X) = W^o(\text{softmax}(W^o(Z) \otimes X) \otimes X) \oplus Z \quad (1)$$

式中, W^o 表示线性变换和Dropout的组合操作, W^o 表示线性变换, \otimes 为逐元素相乘, \oplus 为逐元素相加。

局部特征融合模块则与之相反。希望将全局特

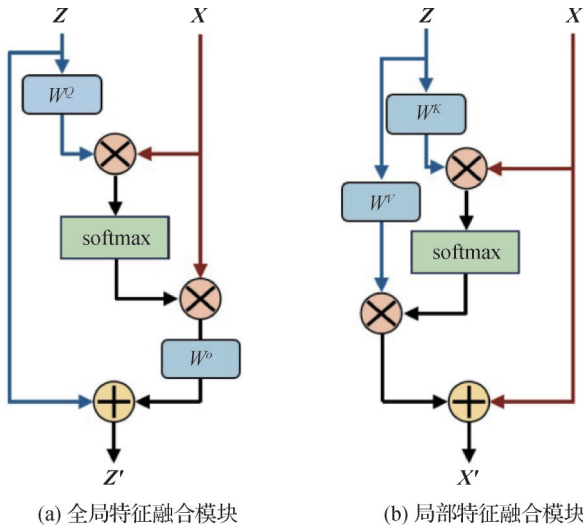


图 4 自适应双向特征融合模块

Fig. 4 Adaptive bidirectional feature fusion module

- (a) global feature fusion module;
(b) local feature fusion module)

征融合到局部信息中,将局部信息作为查询,全局特征作为关键值向量和值向量。通过 LFM 模块,可以将长距离的上下文表征融合到局部特征中,其过程可表示为

$$X' = LFM(Z, X) = \text{softmax}(W^K(Z) \otimes X) \otimes W^V(Z) \oplus X \quad (2)$$

式中, W^K 、 W^V 均表示线性变换。

1.5 损失函数

目前,大部分损失函数关注图像之间的像素差异,而忽略了整体风格的偏移程度。事实上,人们在观察图像时,往往对图像中的亮度颜色变化更为敏感。有研究认为,比较两幅图像的相似程度,不应只从像素相近出发,更应该关注两幅图像的高层信息,如内容和全局结构。

在图像去噪方向,重建后图像的优劣应该从像素级差异和视觉效果、图像内容上进行评判。有研究提出了使用感知损失来对图像去噪模型进行优化,具体来说,感知损失是通过将两幅图像送入预训练好的网络中,得到对应的感知特征图,这相当于对重建前后的图像进行了特征提炼,在感知域中计算重建前后的损失,会使得重建后的图像在视觉、图像内容上呈现更好的效果。感知损失可表示为

$$L_p^j(X, Y) = \frac{1}{C_j H_j W_j} \left\| \phi_j(X) - \phi_j(Y) \right\|_2^2 \quad (3)$$

式中, X 和 Y 分别代表去噪前后的图像, $\phi_j(\cdot)$ 表示预训练好的模型的第 j 层, $C_j H_j W_j$ 是该层产生特征图的

大小,对重建前后每层产生的特征图进行对应欧氏距离计算,得到每层的感知损失,最后将所有层损失相加。本节在 1.3 节的基础上引入了感知损失,帮助模型更好地拟合去噪过程,提升图像恢复的效果。本节所使用的损失函数定义为

$$L(X, Y) = L_{\text{Char}}(X, Y) + \lambda_p L_p(X, Y) \quad (4)$$

式中, $L_{\text{Char}}(\cdot)$ 是 Charbonnier 损失函数, λ_p 是感知系数,经过实验验证后,设定为 0.05。

2 实验结果与分析

2.1 实验平台与数据

本文实验使用平台具体为: Intel(R) Xeon(R) E5 CPU 处理器,配有 NVIDIA 2080Ti 显卡, 32 GB 内存。实验使用 PyTorch 深度学习框架,初始学习率设置为 10^{-4} ,在训练过程中采用动态调整策略,每训练 50 轮,将学习率衰减为原来的 1/2,总共训练 150 轮。

本文使用 ICVL 数据集 (Arad 和 Ben-Shahar, 2016) 和 Pavia University 数据集对模型进行训练和测试。ICVL 数据集包含 201 幅高光谱图像,这些图像是利用 Specim PS Kappa DX4 高光谱相机,在旋转的空间扫描仪上采集得到的。每幅图像都具有 1392×1300 像素的空间分辨率,并涵盖了 $400 \text{ nm} \sim 1000 \text{ nm}$ 的 519 个光谱波段。数据集中提供了 31 个光谱波段的数据图像,覆盖了从 $400 \text{ nm} \sim 700 \text{ nm}$ 的光谱范围。为了网络模型得到更好的训练,ICVL 数据集作为主要训练集,被切割为 $128 \times 128 \times 31$ 的尺寸大小,切割步长为 128,共获得 19 790 幅图像,按照约 8:2 的比例,其中 15 760 幅图像用于训练,4 030 幅图像用于测试。Pavia Center/University 数据集包含 Pavia Center 和 Pavia University 两幅图像,该数据集是通过 ROSIS 传感器在意大利北部帕维亚上空飞行采集而得。其中 Pavia Center 数据是一幅空间分辨率为 1096×715 像素的高光谱图像,其光谱波段是 102, Pavia University 数据的空间分辨率为 610×340 像素,光谱波段为 103,其实验设置与 ICVL 数据集基本相同。

2.2 实验设置

在传感器捕获过程中,高光谱图像会受到各种噪声的影响,包括高斯噪声、条带噪声和脉冲噪声等。这些噪声的叠加也会导致混合噪声的污染。为

了全面展示各种高光谱噪声污染情况,本文设置了以下不同的噪声污染情况:

1)不同水平的高斯噪声。对所有的光谱通道添加零均值的高斯噪声,设定4个噪声水平,高斯噪声强度分别是 $\sigma = 30$ 、 $\sigma = 50$ 、 $\sigma = 70$ 和 $\sigma \in [10, 70]$ 。对于前3个设定,每个通道的噪声水平都是相等的,第4个设定中每个通道的 σ 是在10到70之间随机取值。

2)不同类型的复杂噪声。在高斯噪声的基础上添加椒盐噪声、条纹噪声。同样设定4个噪声水平:

设定1,为所有的光谱通道随机添加零均值的高斯噪声,每个通道噪声水平为30~70的随机数。

设定2,在设定1的基础上,在1/3的通道上随机选择10%~20%的列添加条纹噪声。

设定3,在设定1的基础上,在1/3的通道上增加10%~70%的随机椒盐噪声。

设定4,前3个设定噪声的混合噪声。

2.3 参数分析

2.3.1 全局增强模块多头自注意力中head的选择

在基于块下采样的全局增强模块中,使用了多头自注意力,目的是获得token之间不同维度的关系。head数量意味着token被映射到对应数量的子空间中,其取值不同也会影响模型最终的去噪结果。本文通过设置 $\text{head} = [1, 2, 4, 8, 16]$,针对高斯强度 $\sigma = 30$ 的噪声图像进行去噪,分析head的取值对模型去噪的影响,结果如图5所示。从图中可以看出,当设置head数量为8时,本方法取得最优的去噪性能,因此本文将head数量设置为8。

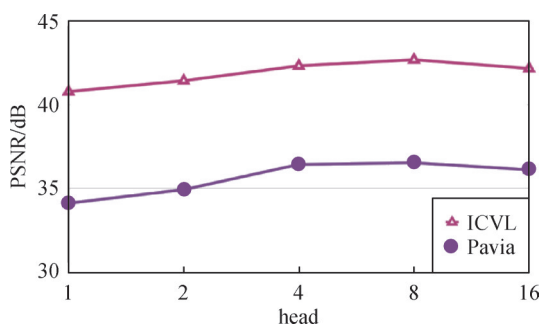


图5 多头自注意力中head取值对去噪性能的影响

Fig. 5 The relationship between the number of head and PSNR

2.3.2 损失函数中感知系数 λ_p 的取值

为使得重建后图像的效果更优,去噪性能更佳,本文在损失函数部分引入了带有系数的感知损失函

数。其中感知系数 λ_p 的取值会对模型重建图像和去噪性能产生影响。针对感知系数 λ_p ,本文设置实验找到最佳取值。从ICVL数据集和Pavia数据集中分别选取部分图像,添加强度 $\sigma = 30$ 的高斯噪声,送入模型中对网络参数进行训练,训练过程中分别使用不同的 λ_p 取值的损失函数,验证不同的 λ_p 取值对网络性能的影响。其中, λ_p 的取值由0.01逐步增加到0.1。如图6所示,当 λ_p 的取值在0.01到0.05之间时,PSNR逐渐增加,去噪后的图像质量逐渐增强,当感知系数 λ_p 超过0.05时,PSNR开始呈下降趋势,去噪后的图像出现局部细节模糊的情况。因此,本文将损失函数中感知系数 λ_p 设定为0.05。

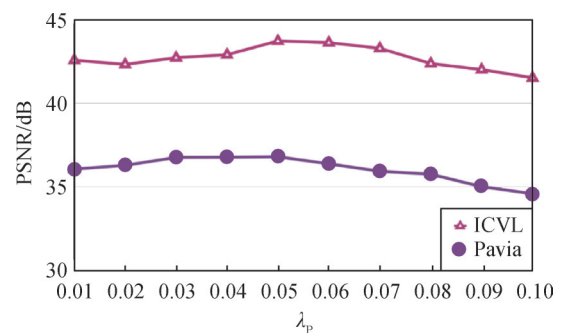


图6 损失函数中感知系数 λ_p 的取值对去噪性能的影响

Fig. 6 The value of perception coefficient λ_p in loss function

2.4 消融实验

为了验证本文设计的通道混合特征提取模块(CSM)、基于块下采样的全局增强模块(DGEM)和自适应双向特征融合模块(ABFM)的有效性,本文进行了消融实验以测试3个模块对于高光谱图像去噪的提升效果。不同于对比实验中使用全部数据用于训练,在消融实验部分,本文从ICVL数据集中选择了40幅具有代表性的图像,对每幅图像按照 $128 \times 128 \times 31$ 的大小、步长为128进行切割,其中80%用于模型训练,20%用于测试,且设定噪声水平为高斯噪声 $\sigma = 30$ 。

实验结果如表1所示。通过结果可以看出,3个模块对去噪效果均有贡献,当网络模型包含3个模块时,去噪性能最佳。一方面,通道混合特征提取模块提取局部特征,保证了重建图像在细节方面的处理;另一方面,基于块下采样的全局增强模块利用自注意力机制增加了长距离特征依赖,进一步提高网络去噪性能。自适应双向特征融合模块可以更好地融合前两个模块提取到的局部特征和全局特征,从

表 1 本文设计的 3 个模块对网络模型去噪能力的影响
Table 1 The impact of the three designed modules on the denoising ability of the network model

Index	PSNR/dB	SSIM	SAM
CSM	43.07	0.968	0.071
CSM+DGEM	43.41	0.969	0.083
CSM+DGEM+GFM	43.52	0.969	0.075
CSM+DGEM+LFM	43.57	0.970	0.069
CSM+DGEM+ABFM	43.81	0.972	0.067

注:加粗字体表示各列最优结果。

而进一步提升网络的去噪性能。

2.5 对比实验分析

为了验证本文提出的基于 Transformer 和通道混合并行的高光谱图像去噪模型的有效性,与其他经典的高光谱图像去噪方法进行对比,包括 BM4D (block-matching 4D filtering) (Maggioni 等, 2013)、LRMR (low rank matrix recovery) (Zhang 等, 2014)、NGMeet (non-local meets global) (He 等, 2022)、HSID-CNN (spatial-spectral deep residual convolutional neural network) (Yuan 等, 2019) 和 QRNN3D (3D quasi-recurrent neural network) (Wei 等, 2021)。

为保证实验的有效性和准确性,所有方法使用的数据集均保持一致,并且均通过对去噪后的图像进行视觉效果比较,量化指标计算与对比,得出各个算法与本文所提方法之间的差异。客观指标采用高光谱图像领域常见的评价指标,包括 PSNR、SSIM 和 SAM。实验由两个部分组成:

1)不同水平的高斯噪声。在 Pavia 数据集上,针对不同水平的高斯噪声进行了对比实验。使用 Pavia University 数据集在已经训练好的模型上进行微调,使用 Pavia Center 数据集做测试。表 2 展示了高斯噪声设置下不同去噪方法在 Pavia 数据集上的去噪结果。由表中数据容易得出,在多数情况下,本文提出的方法在 PSNR、SSIM、SAM 三个评价指标取得了最优的结果。图 7 展示了 Pavia 数据集中 $\sigma = 30$ 的高斯噪声情况下不同去噪算法的去噪可视化结果。实验结果表明,相较于传统方法,基于深度学习的方法在高光谱图像去噪方面具有较好的性能和优势。受益于通道混合和基于块下采样全局增强的并行架构,以及自适应的双向融合模块,本文所提出的 PSFormer 模型更好地学习到噪声特征,具有更广泛的适用性。

表 2 高斯噪声设置下不同去噪方法在 Pavia 数据集上的去噪结果

Table 2 The denoising results of different denoising methods on the Pavia dataset under Gaussian noise settings

方法	$\sigma = 30$			$\sigma = 50$			$\sigma = 70$			盲噪声		
	PSNR/dB	SSIM	SAM	PSNR/dB	SSIM	SAM	PSNR/dB	SSIM	SAM	PSNR/dB	SSIM	SAM
Noisy	18.58	0.155	0.777	14.15	0.068	1.020	11.23	0.037	1.155	16.58	0.213	0.638
LRMR	29.13	0.777	0.307	27.02	0.643	0.452	23.23	0.431	0.525	24.96	0.667	0.339
BM4D	33.97	0.837	0.203	31.54	0.713	0.387	29.03	0.703	0.354	30.55	0.749	0.229
NGMeet	32.88	0.739	0.299	32.01	0.729	0.239	29.77	0.743	0.302	31.44	0.787	0.271
HSID-CNN	34.02	0.864	0.192	32.26	0.731	0.258	31.91	0.718	0.223	33.81	0.791	0.250
QRNN3D	35.19	0.876	0.112	32.33	0.834	0.130	32.13	0.799	0.140	34.67	0.858	0.203
PSFormer (本文)	35.36	0.879	0.116	33.25	0.835	0.128	32.53	0.810	0.138	34.99	0.861	0.124

注:加粗字体表示各列最优结果。

2)不同类型的复杂噪声。本文在 ICVL 数据集上进行了复杂噪声的实验。图 8 展示了在复杂噪声的情况下不同去噪算法的去噪结果图像,其中图像尺寸为 512×512 像素,取 [10, 20, 30] 通道组成伪彩色图像便于展示。从图中容易看出,传统方法如 BM4D 对复杂噪声的去噪效果不足,结果仍然存在

一部分噪声,不能达到预期效果。而 NGMeet 虽然去除了大部分噪声,但在重建图像时出现了平滑过度的问题,丢失了部分图像细节。基于深度学习的方法表现更加优异,对比 HSID 和 QRNN3D,本方法在有效去除大部分噪声的同时,更加关注图像微小的局部细节,保存了绝大部分图像信息,有效恢复了原

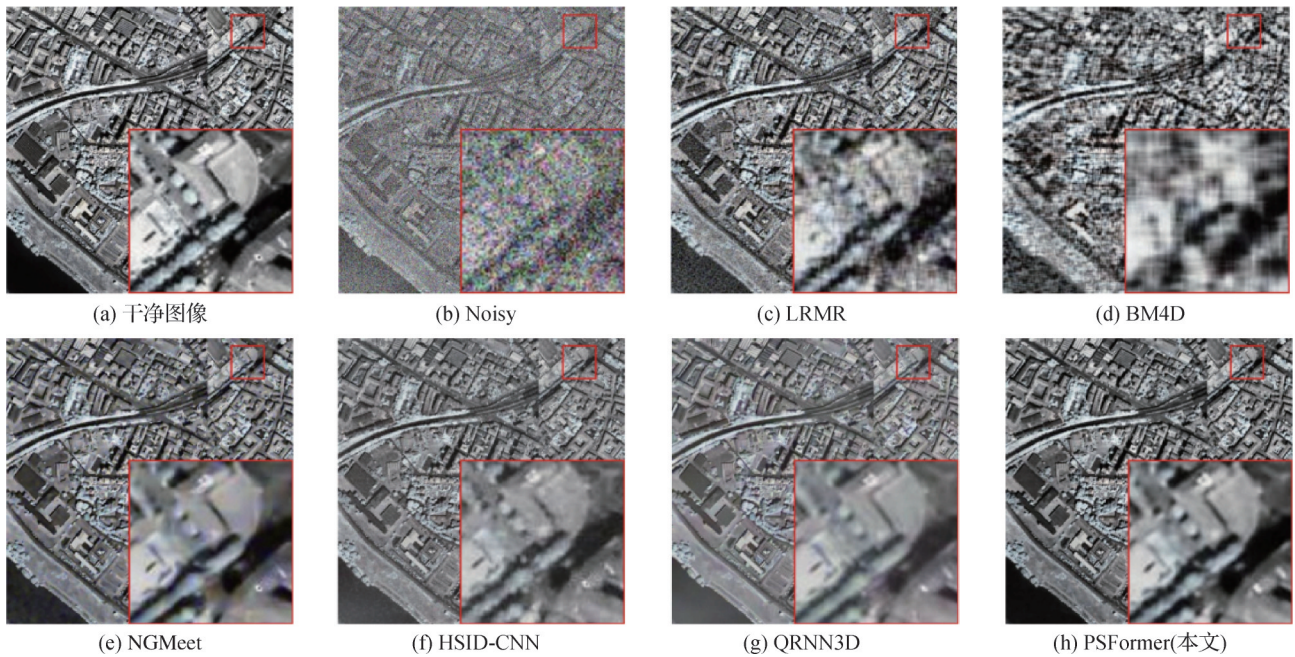


图7 Pavia数据集中 $\sigma = 30$ 的高斯噪声情况下不同去噪算法的去噪结果图像

Fig. 7 Image denoising results of different denoising algorithms in the Pavia dataset with Gaussian noise of $\sigma = 30$
(a) ground truth; (b) Noisy; (c) LRMR; (d) BM4D; (e) NGMeet; (f) HSID-CNN; (g) QRNN3D; (h) PSFormer)

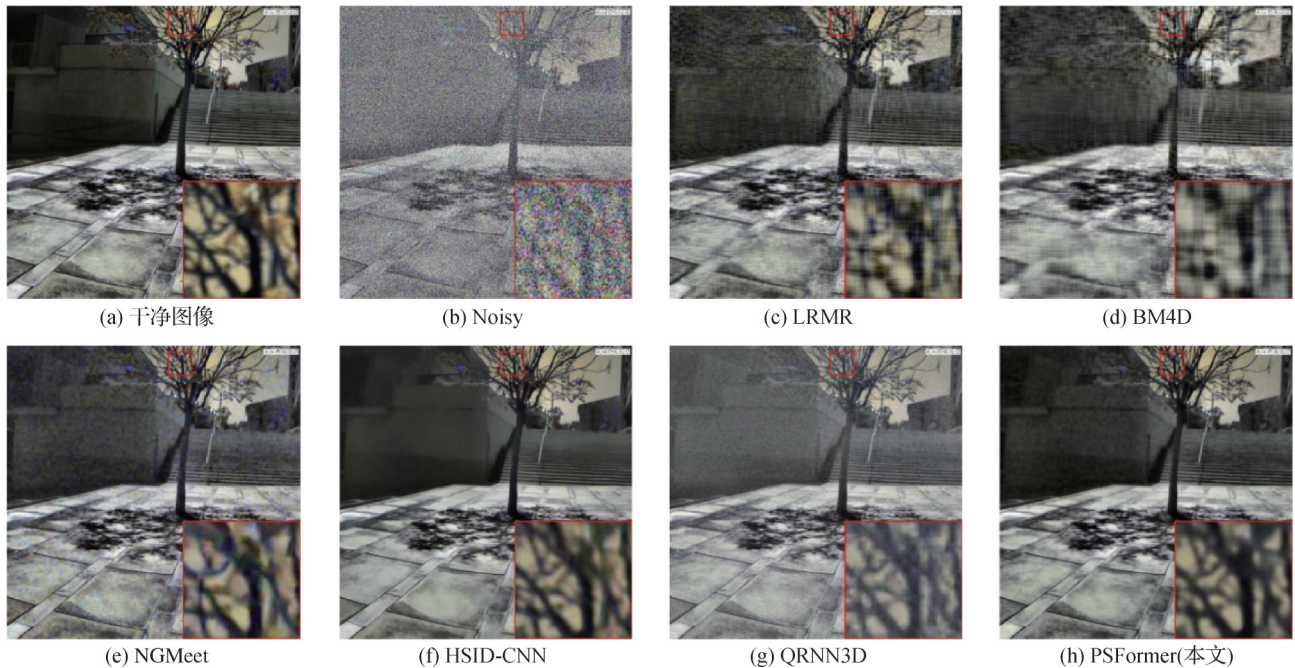


图8 ICVL数据集中复杂噪声情况下不同去噪算法的去噪结果图像

Fig. 8 Image denoising results of different denoising algorithms under complex noise conditions in the ICVL dataset
(a) ground truth; (b) Noisy; (c) LRMR; (d) BM4D; (e) NGMeet; (f) HSID-CNN; (g) QRNN3D; (h) PSFormer)

始图像中的细节特征,其去噪后的图像与原始图像更为接近。表3展示了复杂噪声设置下不同去噪方法在ICVL数据集上的去噪结果,从表3数据中可知,本文提出的基于PSFormer在4项复杂噪声设置中均取得了最佳的指标。

3 结论

为了解决当前高光谱图像去噪方法对远距离依赖利用不足的问题,本文设计了基于Transformer和

表 3 复杂噪声设置下不同去噪方法在 ICVL 数据集上的去噪结果

Table 3 The denoising results of different denoising methods on the ICVL dataset under complex noise settings

方法	设定 1			设定 2			设定 3			设定 4		
	PSNR/dB	SSIM	SAM	PSNR/dB	SSIM	SAM	PSNR/dB	SSIM	SAM	PSNR/dB	SSIM	SAM
Noisy	17.96	0.130	0.983	15.92	0.108	1.004	14.58	0.104	0.991	12.77	0.101	1.030
LRMR	27.58	0.650	0.253	26.80	0.512	0.373	26.38	0.483	0.354	23.52	0.426	0.565
BM4D	36.79	0.928	0.196	31.76	0.749	0.199	27.20	0.585	0.415	25.78	0.434	0.257
NGMeet	36.03	0.938	0.143	35.71	0.914	0.131	33.60	0.916	0.152	32.79	0.771	0.191
HSID-CNN	39.20	0.971	0.071	38.63	0.969	0.089	38.12	0.948	0.101	37.41	0.913	0.105
QRNN3D	40.19	0.969	0.067	39.08	0.955	0.079	38.47	0.935	0.093	37.31	0.948	0.086
PSFormer (本文)	42.09	0.972	0.061	41.82	0.968	0.067	40.65	0.956	0.073	38.54	0.944	0.081

注:加粗字体表示各列最优结果。

通道混合并行的高光谱图像去噪模型 PSFormer, 该模型通过通道混合特征提取模块和基于块下采样的全局增强模块, 有效结合了卷积神经网络和 Transformer 模型, 通过捕捉高光谱图像中的长距离依赖关系和强化近距离特征, 使得模型具有更高的去噪性能, 从而得到更高质量的图像。为了解决 CNN 和 Transformer 结构之间的差异, 本文提出了自适应双向特征融合模块。该模块更好地整合了全局增强模块和通道混合特征提取模块的特征, 使局部信息和全局特征得到充分利用, 同时也有助于生成远距离和近距离的特征。在 ICVL 和 Pavia 数据集上的实验结果表明, 本文方法明显优于 BM4D、LRMR、NGMeet、HSID-CNN 和 QRNN3D 等主流方法。同时, 也验证了本方法在不同强度, 不同类型的噪声水平下都具有较好的去噪效果和泛化能力。

参考文献 (References)

- Arad B and Ben-Shahar O. 2016. Sparse recovery of hyperspectral signal from natural RGB images//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 19-34 [DOI: 10.1007/978-3-319-46478-7_2]
- Chen Y, Huang T Z, He W, Zhao X L, Zhang H Y and Zeng J S. 2022. Hyperspectral image denoising using factor group sparsity-regularized nonconvex low-rank approximation. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #5515916 [DOI: 10.1109/TGRS.2021.3110769]
- Gelvez-Barrera T, Arguello H and Foi A. 2022. Joint nonlocal, spectral, and similarity low-rank priors for hyperspectral-multispectral

image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #5537112 [DOI: 10.1109/TGRS.2022.3203294]

- Hao J L, Xue J Z, Zhao Y Q and Chan J C W. 2023. Transformed structured sparsity with smoothness for hyperspectral image deblurring. *IEEE Geoscience and Remote Sensing Letters*, 20: #5500105 [DOI: 10.1109/LGRS.2022.3230205]
- He W, Yao Q M, Li C, Yokoya N, Zhao Q B, Zhang H Y and Zhang L P. 2022. Non-local meets global: an iterative paradigm for hyperspectral image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44 (4): 2089-2107 [DOI: 10.1109/TPAMI.2020.3027563]
- Li H, Zhao X L, Lin J and Chen Y. 2022. Low-rank tensor optimization with nonlocal plug-and-play regularizers for snapshot compressive imaging. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 581-593 [DOI: 10.1109/JSTARS.2021.3136217]
- Li Y Z, Chong Y W, Pan S M and Ding Y. 2023. First-order smoothing-based deep graph network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61: #5504716 [DOI: 10.1109/TGRS.2023.3250715]
- Maffei A, Haut J M, Paoletti M E, Plaza J, Bruzzone L and Plaza A. 2020. A single model CNN for hyperspectral image denoising. *IEEE Transactions on Geoscience and Remote Sensing*, 58 (4): 2516-2529 [DOI: 10.1109/TGRS.2019.2952062]
- Maggioni M, Katkovnik V, Egiazarian K and Foi A. 2013. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Transactions on Image Processing*, 22 (1): 119-133 [DOI: 10.1109/TIP.2012.2210725]
- Othman H and Qian S E. 2006. Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage. *IEEE Transactions on Geoscience and Remote Sensing*, 44 (2): 397-408 [DOI: 10.1109/TGRS.2005.860982]
- Shi Q, Tang X P, Yang T R, Liu R and Zhang L P. 2021. Hyperspectral

- image denoising using a 3-D attention denoising network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(12): 10348-10363 [DOI: 10.1109/TGRS.2020.3045273]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I. 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 X H, Zhao K Y, Zhao X Y and Li S Y. 2023. TriTF: a triplet Transformer framework based on parents and brother attention for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61: #5507213 [DOI: 10.1109/TGRS.2023.3260969]
- Wei K X, Fu Y and Huang H. 2021. 3-D quasi-recurrent neural network for hyperspectral image denoising. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1): 363-375 [DOI: 10.1109/TNNLS.2020.2978756]
- Xiong F C, Zhou J, Zhao Q L, Lu J F and Qian Y T. 2022. MAC-Net: model-aided nonlocal neural network for hyperspectral image denoising. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #5519414 [DOI: 10.1109/TGRS.2021.3131878]
- Yu D B, Li Q W, Wang X L, Zhang Z L, Qian Y X and Xu C. 2023. DSTrans: dual-stream Transformer for hyperspectral image restoration//Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, USA: IEEE: 3728-3738 [DOI: 10.1109/WACV56688.2023.00373]
- Yuan Q Q, Zhang Q, Li J, Shen H F and Zhang L P. 2019. Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2): 1205-1218 [DOI: 10.1109/TGRS.2018.2865197]
- Zhang B, Chen Y X, Rong Y, Xiong S W and Lu X Q. 2023. MATNet: a combining multi-attention and Transformer network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61: #5506015 [DOI: 10.1109/TGRS.2023.3254523]
- Zhang H Y, He W, Zhang L P, Shen H F and Yuan Q Q. 2014. Hyperspectral image restoration using low-rank matrix recovery. *IEEE Transactions on Geoscience and Remote Sensing*, 52(8): 4729-4743 [DOI: 10.1109/TGRS.2013.2284280]
- Zhao B, Ulfarsson M O, Sveinsson J R and Chanussot J. 2022. Hyperspectral image denoising using spectral-spatial transform-based sparse and low-rank representations. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #5522125 [DOI: 10.1109/TGRS.2022.3142988]

作者简介

胡帅,男,硕士研究生,主要研究方向为高光谱图像修复和深度学习。E-mail: hushuai_ouc@qq.com

高峰,通信作者,男,副教授,主要研究方向为多媒体计算与高光谱图像处理。E-mail: gaofeng@ouc.edu.cn

龚卓然,女,硕士研究生,主要研究方向为高光谱图像去噪。E-mail: gongzhuoran_anais@163.com

陶盛恩,男,硕士研究生,主要研究方向为计算机视觉和深度学习。E-mail: 547196205@qq.com

上官心语,女,硕士研究生,主要研究方向为图像处理和深度学习。E-mail: 18342535177@163.com

董军宇,男,教授,主要研究方向为计算机视觉与模式识别。E-mail: dongjunyu@ouc.edu.cn