

DOI:10.19651/j.cnki.emt.2314321

高光谱图像去噪的稀疏空谱 Transformer 模型^{*}

杨智翔 孙玉宝 白志远 来鸿康

(南京信息工程大学计算机学院网络空间安全学院 南京 210044)

摘要: 现阶段 Transformer 模型的应用提升了高光谱图像去噪的性能,但原始 Transformer 模型对图像空间-光谱耦合关联性的利用仍存在不足;对空间特征的处理存在过于平滑,容易丢失小尺度结构的现象;同时在光谱维度上也过于关注全部通道特征,缺乏对不同光谱波段间差异性的利用;为了应对这些问题,本文提出了一种新的稀疏空谱 Transformer 模型,提升了对空谱耦合关联性的利用。在空间维度,引入局部增强模块增强空间特征细节,应对过平滑问题;同时在光谱维度上提出了 Top-k 稀疏自注意力机制,自适应选择前 K 个最相关的光谱通道特征进行特征交互,从而能够有效捕获空谱特征。最终通过稀疏空谱 Transformer 的层级残差连接实现高光谱图像的去噪。在 ICVL 数据集上分别对高斯噪声和复杂噪声进行去噪处理,峰值信噪比分别达到 40.56 dB 和 40.19 dB,证明了本文提出的稀疏空谱 Transformer 模型优越的性能。

关键词: 高光谱图像去噪;空间-光谱联合特征;稀疏 Transformer

中图分类号: TP751 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Hyperspectral image denoising with sparse spatial-spectral transformer

Yang Zhixiang Sun Yubao Bai Zhiyuan Luan Hongkang

(School of Computer Science & School of Cyberspace Security, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract: The application of Transformer models has improved the performance of hyperspectral image denoising. However, the original Transformer model still falls short in effectively leveraging the spatial-spectral coupling in HSIs. It tends to excessively smooth spatial features, leading to the loss of small-scale structures. Moreover, it overly emphasizes all spectral channel features, neglecting the differences between different spectral bands. In order to solve these problems, this paper introduces a novel Sparse Spatial-Spectral Transformer model, enhancing the utilization of spatial-spectral coupling. In the spatial dimension, a local enhancement module is introduced to refine spatial feature details and deal with over-smoothing problem. Simultaneously, in the spectral dimension, a Top-k sparse self-attention mechanism is proposed, which adaptively selects the top-K most relevant spectral channel features for feature interaction, effectively capturing spatial-spectral characteristics. Ultimately, hyperspectral image denoising is achieved through hierarchical residual connections with the Sparse Spatial-Spectral Transformer. On the ICVL dataset, denoising performance for both Gaussian noise and complex noise attains peak signal-to-noise ratios of 40.56 dB and 40.19 dB, respectively, demonstrating the superior performance of the proposed Sparse Spatial-Spectral Transformer model in this paper.

Keywords: hyperspectral image denoising; spatial-spectral joint feature; sparse Transformer

0 引言

高光谱图像(hyperspectral image, HSI)是由光谱仪捕获的一种三维图像数据,对比彩色图像具有更精细的光谱分辨率,能够获得场景中各空间位置处物体反射的近似光

谱曲线,可用于判别地物类型。因此高光谱图像可以广泛应用于遥感^[1]、材料识别^[2]、农业^[3]、医疗诊断^[4]和分类^[5]等领域。然而在光谱仪捕获过程中,由于低光照、光子效应、大气干扰以及传感器电路影响等原因,高光谱图像常常会受到多种类型噪声的干扰与污染,这对高光谱图像后续

收稿日期:2023-08-08

* 基金项目:国家自然科学基金(62276139, U2001211)项目资助

分析产生了不利影响。因此,针对高光谱图像的去噪处理是后续高光谱图像进一步分析应用的关键预处理过程。

高光谱图像去噪旨在从含噪图像中恢复出干净的无噪声图像,是一个典型的反问题。为此,先验正则约束的变分模型得到广泛研究。结合高光谱图像一些内在的物理特性,代表性的先验正则约束主要包括低秩先验^[6],非局部相似性^[7],空间-光谱相关性^[8]和全局光谱相关性^[9]等。其中 BM3D^[10]主要关注的是空间维度上的非局部相似性,没有考虑到光谱维度的特征信息。LLRT^[11]利用非局部低秩张量来建模干净高光谱图像,还有基于字典学习的 TDL^[12]以及基于阈值滤波的 BM4D^[13]。尽管这些方法有效的利用了高光谱图像空间和光谱的潜在特征并取得了不错的去噪结果,但是这些方法的去噪性能最终取决于手动设置的先验信息与高光谱图像固有物理特性的匹配程度。在另一个方面,基于先验约束进行高光谱图像去噪时需要进行多次迭代优化处理,时间复杂度高。

近些年随着深度学习的发展,基于卷积神经网络(convolutional neural networks, CNN)的高光谱图像去噪方法展现出了极为优秀的去噪性能。Yuan 等^[14]提出了一种基于卷积神经网络的高光谱图像去噪方法 HSID-CNN,结合残差连接,端到端的学习原始的无噪声干净图像。QRNN3D^[9]运用 3 维卷积和准循环池来捕获高光谱图像空间和光谱特征的相关性,提升高光谱图像去噪的效果。同时,自然语言领域兴起的 Transformer 模型也被应用于视觉任务^[15],Wang 等^[16]提出了一个基于 Transformer 架构的图像去噪方法 U-former,能捕捉图像空间上的全局依赖关系。TRQ3D^[17]将 U-former 与 QRNN3D 结构进行结合,分别对高光谱图像空间和光谱特征进行提取再融合处

理。高光谱图像的空谱耦合关联性是极为关键的先验特性。然而,现有的 Transformer 结构对空谱耦合关联性的利用仍存在不足。且存在对空间特征处理过于平滑,容易丢失小尺度结构的现象。同时,在光谱维度上也过于关注全部通道特征,缺乏对不同光谱波段间差异性的利用。

为了解决上述问题,本文提出了一种稀疏空间-光谱 Transformer 模型,主要由空间维度 Transformer 和光谱维度 Transformer 两个部分组成。在空间维度 Transformer 中,通过基于移动窗口的自注意力模块捕获高光谱图像的非局部空间相似性,并提出局部增强模块对空间局部特征进行补充融合,提升对空间边缘细节处理的能力。其次,在光谱维度 Transformer 中提出用 Top-k 稀疏自注意力代替原始自注意力对光谱波段进行选择,利用特征最相关的前 K 个波段进行辅助去噪,并通过一个光谱通道混合模块提升光谱维度 Transformer 探索全局光谱相关性的能力。最后,通过在 ICVL 数据集上的实验结果表明,本文提出的模型在高斯噪声和复杂混合噪声条件下的去噪性都优于当前的先进方法。同时,消融实验也验证了本文提出的稀疏空谱 Transformer 模型的有效性。

1 稀疏空谱 Transformer 去噪网络

如图 1 所示,本文提出的稀疏空谱 Transformer 去噪网络由 6 个稀疏空谱去噪模块组成,每一个模块包含 6 个 Transformer 子模块。每个 Transformer 子模块都包含了本文提出的稀疏空间-光谱自注意力模块。通过这样的层级结构,网络模型能够在一个统一的维度下对高光谱图像的空间-光谱特征相关性进行处理。准确地恢复干净高光谱图像,提高去噪结果的质量和细节保留能力。

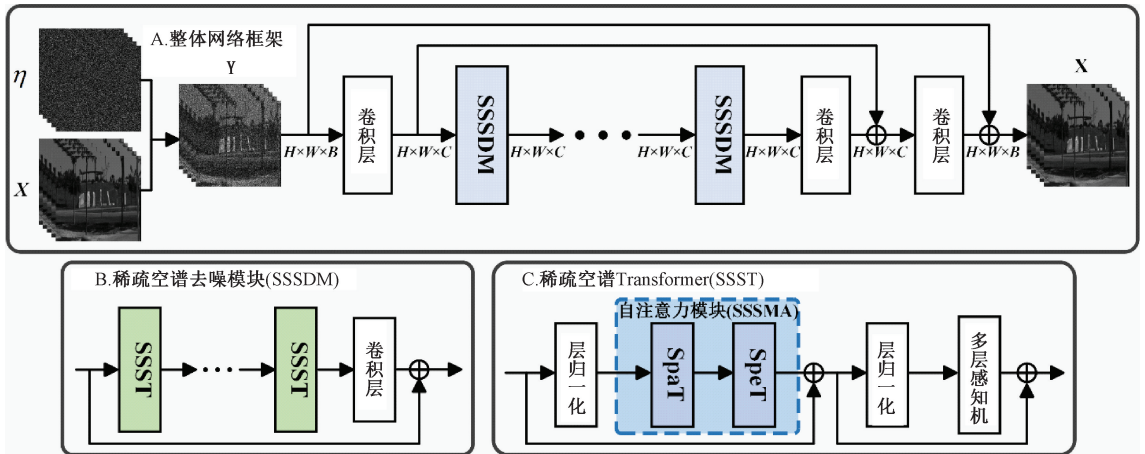


图 1 总体网络框架

具体而言,噪声高光谱图像可以看作是在干净的高光谱图像上添加各种类型的噪声。可以通过一个公式来建模这一图像退化过程:

$$Y = X + \eta \tag{1}$$

其中, η 表示高光谱图像采集过程中存在的各种类型

的噪声, $X \in \mathbb{R}^{H \times W \times B}$ 表示干净的高光谱图像, $Y \in \mathbb{R}^{H \times W \times B}$ 表示被加性噪声 η 破坏的噪声图像,其中 H 和 W 分别代表空间维度的高度和宽度。 B 表示光谱维度的波段数,在进入稀疏空谱去噪模块(SSSDM)之前,首先需要通过一个 3×3 卷积层从噪声图像 $Y \in \mathbb{R}^{H \times W \times B}$ 中提取浅层

特征 $F_0 \in \mathbb{R}^{H \times W \times C}$ 。然后浅层特征以固定的特征尺寸输入到稀疏空谱去噪模块。

$$F_t = \text{SSSDM}_t(F_{t-1}), t = 1, 2, \dots, 6 \quad (2)$$

其中, SSSDM_t 代表第 t 层的稀疏空谱去噪模块。每个稀疏空谱去噪模块的尾部都拥有一个 3×3 卷积层, 并采用了跳跃连接的方式, 从上一个模块中添加残差特征防止网络过拟合。为了从最终的图像特征 $F_T \in \mathbb{R}^{H \times W \times C}$ 中恢复出干净高光谱图像, 通过一次跳跃连接和两个 3×3 卷积层将浅层特征连接起来。最后再通过全局跳跃连接噪声图像, 从噪声图像中恢复得到干净图像。

1.1 稀疏空谱 Transformer(SSST)

为了防止了网络陷入秩崩溃。即在无残差连接和多层感知机(multi-layer perceptron, MLP)架构的情况下, 自注意力网络的表达能力随深度增加而呈双指数形式衰减, 导致模型丧失对输入图像特征中的大部分内容的建模能力。因此, 本文提出的稀疏空谱 Transformer 遵循了 VIT^[18] 的基础结构, 在经过自注意力模块处理之后进行残差连接、层归一化以及 MLP 处理层等结构。

具体而言, 假设 $Z_{l-1} \in \mathbb{R}^{H \times W \times C}$ 代表第 l 个稀疏空谱 Transformer 的输入特征, 则稀疏空谱 Transformer 的处理过程可以表示为:

$$Z'_l = \text{SSSMA}(\text{LN}(Z_{l-1})) + Z_{l-1} \quad (3)$$

$$Z_l = \text{MLP}(\text{LN}(Z'_l)) + Z'_l \quad (4)$$

其中, Z'_l 和 Z_l 分别代表稀疏空谱多头自注意力模块和稀疏空谱 Transformer 的输出。

稀疏空谱多头自注意力模块主要包含空间维度 Transformer 和光谱维度 Transformer 两个结构。假定稀疏空谱多头自注意力模块的输入特征为 $Z^{\text{in}} \in \mathbb{R}^{H \times W \times C}$, 在经过层归一化之后首先在空间维度上进行窗口划分

作, 每个窗口的大小为 $M \times M$, 从而将整个输入特征划分为 $\frac{HW}{M^2}$ 个不重叠的块, 可以表示为 $\{Z_1^{\text{in}}, \dots, Z_i^{\text{in}}, \dots, Z_I^{\text{in}}\}$,

其中每一个块的维度为 $Z_i^{\text{in}} \in \mathbb{R}^{M^2 \times C}$ 。划分后, 每个特征块通过空间维度 Transformer 来进行高光谱图像的非局部空间特征相似性的提取。这一过程可以表述为:

$$\{Z_i^{\text{in}}\} = \text{WinPartition}(Z^{\text{in}}), i = 1, \dots, I \quad (5)$$

$$Z_i^{\text{spa}} = \text{SpaT}(Z_i^{\text{in}}), i = 1, \dots, I \quad (6)$$

$$Z^{\text{spa}} = \text{WinReverse}(\{Z_i^{\text{spa}}\}), i = 1, \dots, I \quad (7)$$

通过窗口反向恢复操作将空间维度 Transformer 输出的特征块恢复到与输入特征一样的大小, 得到最终的输出为 $Z^{\text{spa}} \in \mathbb{R}^{H \times W \times C}$ 。然后本文将得到的输出特征传递给位于光谱维度 Transformer。其中包含有一个光谱通道混合模块, 利用 Top-k 稀疏自注意力模块获取最相关波段特征信息进行辅助去噪。这个过程可以简单表示为:

$$Z^{\text{cm}} = \text{SpeT}(\text{SCM}(Z^{\text{spa}})) \quad (8)$$

1.2 空间维度 Transformer(SpaT)

在实际应用 Transformer 对高光谱图像去噪时, 尽管全局空间自注意力可以在大视野中带来空间上的长程依赖, 但由于其计算复杂度随着空间大小的增加而显著增大, 因此本文采用窗口移位的自注意力方法来获取临近像素点以外的非局部空间上的相似性信息。但是, 实验发现在运用 Transformer 的自注意力模块提取空间上长程依赖的时候, 不能完全忽略卷积操作在恢复图像边缘细节上的优势。因此, 如图 2 所示, 本文在原始自注意力机制中引入一个局部增强模块。该模块将图像边缘细节引入到自注意力运算中, 能加强网络对图像边缘细节的建模能力, 让网络捕获到更多的图像细节信息。

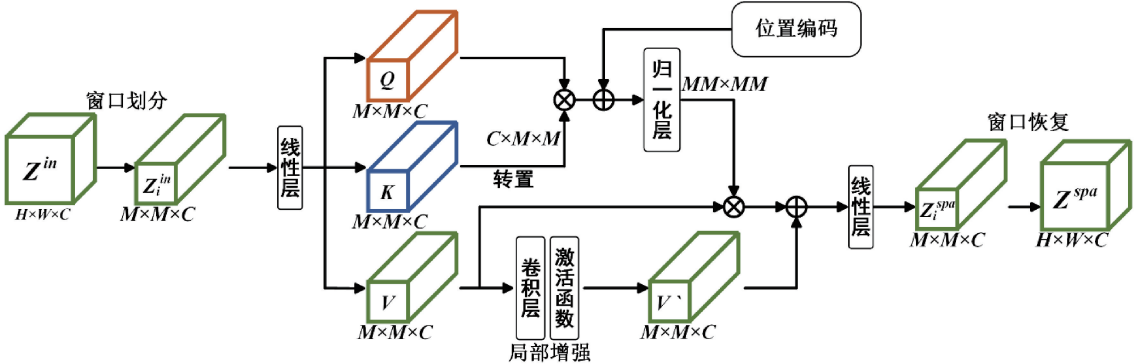


图 2 空间维度 Transformer(SpaT)

具体而言, 将输入的图像特征块 $Z_i^{\text{in}} \in \mathbb{R}^{M^2 \times C}$ 通过线性投影得到查询 Q_i^{spa} 键 K_i^{spa} 值 $V_i^{\text{spa}} \in \mathbb{R}^{M^2 \times C}$ 三个向量, 具体可以表示为:

$$Q_i^{\text{spa}} = Z_i^{\text{in}} W_q^{\text{spa}}, K_i^{\text{spa}} = Z_i^{\text{in}} W_k^{\text{spa}}, V_i^{\text{spa}} = Z_i^{\text{in}} W_v^{\text{spa}} \quad (9)$$

其中, $W_q^{\text{spa}}, W_k^{\text{spa}}, W_v^{\text{spa}}$ 是大小为 $C \times C$ 的权重。后续采用多头机制对来自不同窗口子空间的信息进行联合组装。

$Q_i^{\text{spa}}, K_i^{\text{spa}}$ 和 V_i^{spa} 被分成 N 个头, 分别作为 $Q_{ij}^{\text{spa}}, K_{ij}^{\text{spa}}$ 和 V_{ij}^{spa} 。因此, 每个头 $\text{head}_{ij}^{\text{spa}}$ 的非局部空间自注意力矩阵计算如下:

$$A_{ij}^{\text{spa}} = \text{SoftMax}(Q_{ij}^{\text{spa}} K_{ij}^{\text{spa}T} / \sqrt{d} + B) \quad (10)$$

$$\text{head}_{ij}^{\text{spa}} = A_{ij}^{\text{spa}} V_{ij}^{\text{spa}} \quad (11)$$

其中, d 是 Q_{ij}^{spa} 的维数, 具体地说就是 C/N 。而 B 是

相对偏差。此外,本文通过在每个多头自注意力模块之间进行空间移位操作,以获得非局部的空间信息。具体地说,移位操作是通过在窗口划分之前将输入特征进行移位 $\left\lfloor \frac{M}{2} \right\rfloor, \left\lfloor \frac{M}{2} \right\rfloor$ 大小的像素来实现的。

同时,为了实现对高光谱图像局部细节信息的补充,局部增强模块将线性投影得到的值 $V_i^{spa} \in \mathbb{R}^{M^2 \times C}$ 作为输入,接入一个卷积层和激活函数,与一般的 Transformer 自注意力结构不同,这种设计有助于引入局部窗口的细节信息,实现对 Transformer 结构在局部细节信息提取上的补充。这个过程可以表示如下:

$$head_{ij}^{lcm} = head_{ij}^{spa} + \text{RELU}(\text{Conv}(V_i^{spa})) \quad (12)$$

空间自注意力运算后,将所有的多头输出 $head_{ij}^{lcm}$ 嵌入

在一起并进行线性运算,得到空间维度 Transformer 的输出 $Z_i^{spa} \in \mathbb{R}^{M \times M \times C}$ 。再按位置信息进行反向窗口恢复,得到最终的输出 $Z^{spa} \in \mathbb{R}^{H \times W \times C}$ 并向后传递。

1.3 光谱维度 Transformer (SpeT)

在经过空间维度 Transformer 之后,已经获得了输入高光谱图像的空间特征信息。但它仍然缺乏光谱特征信息,所以需要对光谱维度进行特征提取。为了对高光谱图像进行全局光谱特征相关性的提取,本文提出了光谱维度 Transformer。同时,为了防止不相关特征信息的干扰,本文新提出了 Top-k 稀疏自注意力结构。通过 Top-k 算法对不同波段间的特征信息进行自适应的选择,选择与目标波段特征信息最相关的前 K 个波段进行辅助去噪。最后再通过线性层得到输出特征,如图 3 所示。

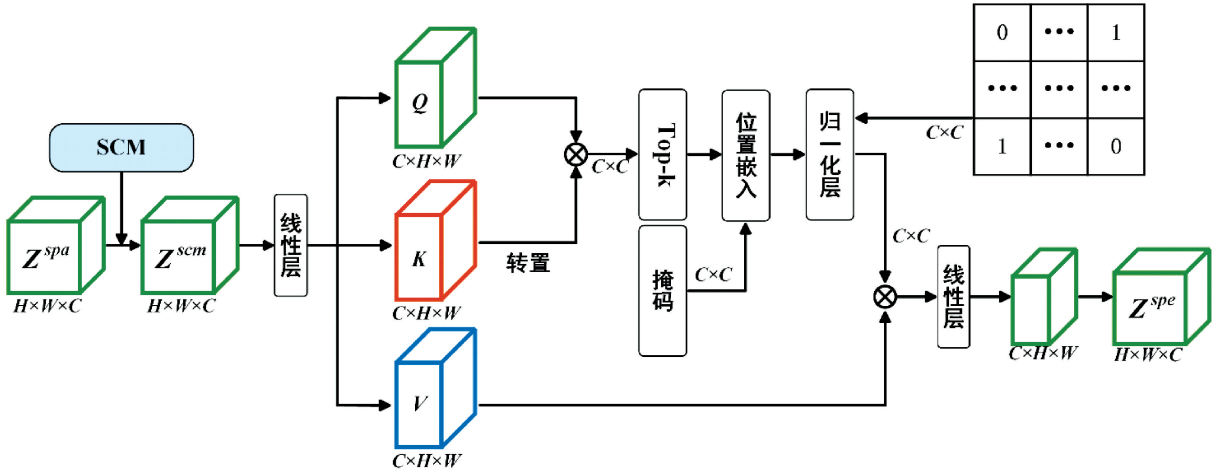


图 3 光谱维度 Transformer (SpeT)

具体而言,光谱维度 Transformer 首先经过一个光谱通道混合模块(SCM),如图 4 所示。为提升不同波段间光谱特征的表达能力,保证 Top-k 算法能够选择到最相关的波段。本文提出的光谱通道混合模块主要包含动态混合(DM)和静态注意力映射(SM)两个部分,整个过程可以简单的表示为:

$$Z^{scm} = \text{SM}(\text{GELU}(\text{DM}(Z^{spa}))) \quad (13)$$

其中, Z^{spa}, Z^{scm} 分别表示输入输出特征图。不同于以往的通道注意力机制,本文提出了利用逐元素动态权重通道混合的方法来增强全局波段间的特征相关性。具体地,首先通过线性投影计算每个输入特征图的缩放权重,其中 $W_1 \in \mathbb{R}^{C \times C}$ 。然后,本文针对每个特征信息的位置进行逐元素乘法。再用缩放权重重新缩放输入特征,其中 $W_2 \in \mathbb{R}^{C \times 2C}$ 。动态混合过程可以表示为:

$$Z^{dm} = (W_2(W_1 Z^{spa} \odot Z^{spa} + Z^{spa})) \quad (14)$$

其中, \odot 表示矩阵点乘,即矩阵间逐一对应元素相乘。静态注意力映射与以往的通道式门控模块只控制个别通道向后续传递信息不同,本文提出的静态注意力映射通过静态注意力映射图 $W_3 \in \mathbb{R}^{2C \times C}$ 混合所有通道的信息。因

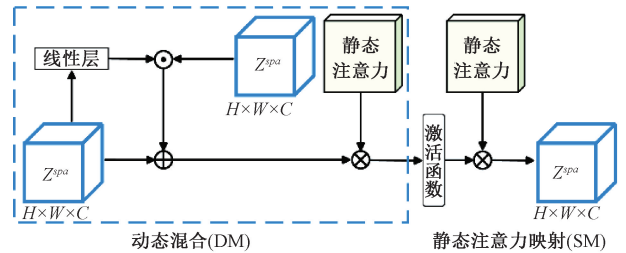


图 4 光谱通道混合模块 (SCM)

为高光谱图像具有全局光谱相关性,需要对全部的波段信息进行处理。其中注意力图映射图 W_1, W_2, W_3 是与整个网络训练共同联合学习的。静态注意力映射表示为:

$$Z^{scm} = Z^{dm} W_3 \quad (15)$$

在经过光谱通道混合模块之后,将得到的高光谱图像特征通过 Top-k 稀疏自注意力模块进行光谱特征的提取。具体而言,首先将输入图像 $Z^{scm} \in \mathbb{R}^{H \times W \times C}$ 转置为 $Z^T \in \mathbb{R}^{C \times HW}$ 。然后通过线性投影得到查询 $Q^{spe} \in \mathbb{R}^{C \times HW}$ 键 $K^{spe} \in \mathbb{R}^{C \times HW}$ 和值 $V^{spe} \in \mathbb{R}^{C \times HW}$ 3 个向量。具体可以表示为:

$$Q^{\text{spe}} = W_q^{\text{spe}} Z^T, K^{\text{spe}} = W_k^{\text{spe}} Z^T, V^{\text{spe}} = W_v^{\text{spe}} Z^T \quad (16)$$

其中, $W_q^{\text{spe}}, W_k^{\text{spe}}, W_v^{\text{spe}}$ 是大小为 $C \times C$ 的权重。然后 Top-k 稀疏自注意力模块的具体实现可以表示为:

$$\text{Att}^{\text{spe}} = \text{SoftMax}(\Gamma_\kappa(K^{\text{spe}T} Q^{\text{spe}} / \sqrt{d})) V^{\text{spe}} \quad (17)$$

其中, Γ_κ 选择算子的一个可学习的选择算子, 具体实现如下:

$$\Gamma_\kappa(S) = \begin{cases} S_{ij}, & S_{ij} \in \text{top-k}(\text{row } j) \\ 0, & \text{其他} \end{cases} \quad (18)$$

意味着从自注意力矩阵 A_j^{spe} 的每一行特征信息中, 选择与当前行最相关的前 K 个特征信息进行保留, 其余的权重将设置为 0。这种做法可以帮助网络选择特征信息最相关的光谱通道进行辅助去噪。

具体而言, 针对线性投影得到的查询 Q^{spe} 和键 K^{spe} , 需要先计算查询-键对整体的相似度, 并得到一个自注意力矩阵 $A_j^{\text{spe}} \in \mathbb{R}^{C \times C}$ 。但与一般自注意力模块不同的是, 本文利用 Top-k 算法来屏蔽掉这个自注意力矩阵中不相关的冗余特征信息。需要注意的是, 与随机丢弃策略(dropout)不同, 本文采用的 Top-k 方法包含一个可学习的 Top-k 算子, 可以自适应的选择 K 值大小, 而这个 K 值的取值范围是手动设定的一个范围。自适应的 K 值选取可以动态的控制不同波段选取与自身最相关的特征信息的数量。这种方法的主要目的是选出与每一个波段的特征最相关的前 K 个波段来进行辅助去噪, 并忽略掉不相关的波段信息。然后, 再对自注意力矩阵进行 SoftMax 处理与后续处理。以往直接对自注意力矩阵进行 SoftMax 处理会导致自注意力矩阵中高频信息被减弱, 得到过度平滑的结果, 导致特征提取被干扰。同时, 光谱维度 Transformer 也进行与空间维度 Transformer 中类似的操作, 采用了多头机制。最后, 本文将多头注意力的所有输出连接起来, 然后通过线性投影得到最终输出结果 $Z^{\text{spe}} \in \mathbb{R}^{H \times W \times C}$ 。值得强调的是, 与 Z^{spa} 相比, Z^{spe} 具有更多的光谱细节, 并且也还保留了关键的空间信息。

2 实验设置

2.1 数据集

本文主要在 ICVL^[19] 数据集上评估本文提出的方法。ICVL 数据集由 201 幅图像组成, 空间分辨率为 1392×1300 , 光谱波段范围为 $400 \sim 700 \text{ nm}$ 。本文遵循了 QRNN3D 工作中的数据集设置方法, 随机选择其中的 100 个高光谱图像作为训练集, 50 个高光谱图像作为测试集, 从而确保实验的真实性。为了实验能顺利进行, 本文将训练集的图像大小统一裁剪到 1024×1024 , 并归一化到 $[0, 1]$ 。然后, 本文为了扩大训练集, 通过裁剪将训练集中的高光谱图像处理成多个重叠的图像块。每个图像块的空间分辨率为 64×64 , 光谱分辨率保持不变, 还采用了旋转等方式进一步增强训练数据集。至于测试数据集, 为了更好的视觉效果, 每个高光谱图像被裁剪成包含主要区域

的 $512 \times 512 \times 31$ 大小的图像块。同时测试数据集也进行了归一化处理。

2.2 实验细节

本文使用 Adam 优化器^[20] 对网络进行优化, 参数由 Xavier 初始化进行初始化。批量大小被设置为 8, 训练周期被设置为 100。初始学习速率设置为 1×10^{-4} , 并在 60 个周期后调整为 1×10^{-5} 。在网络模型中, 本文使用均方误差 (MSE Loss) 作为模型的损失函数, 均方误差损失函数在图像去噪任务中被广泛应用。通过计算预测输出与真实标签之间的均方误差, 从而衡量预测结果与真实目标之间的差异。本文的目标是最小化均方误差损失, 以保证网络能够学习到对高光谱图像进行去噪的最优输出。

本文对比的深度学习方法 (QRNN3D^[9]、NSSNN^[21] 和 TRQ3D^[17]) 和本文提出的方法都采用 PyTorch 实现, 并用 GeForce RTX 3090 运行。传统的方法, 包括 BM4D^[13]、LLRT^[11]、NMoG^[22] 和 NG-MEET^[23] 等, 在 MATLAB 环境下实现, 并在 Intel Core i5-12400 CPU 上运行。这些对比方法中涉及的所有参数都按照参考文献中的描述进行了优化分配或自动选择。

2.3 噪声类型的设置

遵循高光谱图像以前的去噪工作^[9, 14] 中的设置, 模拟了高斯噪声和复杂噪声两种主要噪声类型。

高斯噪声: 高斯噪声是具有不同方差的零均值独立同分布 (independent and identically distributed, IID) 高斯噪声, 所有波段上的噪声水平相同。评估了方差为 30, 50 和 70 以及 Blind (方差范围从 $10 \sim 70$) 等 4 种条件下的高斯噪声。

复杂噪声: 在 QRNN3D^[9] 这项工作之后, 为了能够有效检测复杂噪声情况下的去噪效果, 主要将复杂噪声模拟为只含有非 IID 高斯噪声 (Non-IID)、非 IID 高斯噪声与截止噪声的混合 (Deadline)、非 IID 高斯噪声与条纹噪声的混合 (Stripe)、非 IID 高斯噪声与脉冲噪声的混合 (Impulse) 以及以上所有复杂噪声的混合 (Mixture) 等五种情况。

2.4 评估标准

为了定量地评价本文提出的方法, 本文使用了 3 个性能指标, 包括峰值信噪比 (peak signal to noise ratio, PSNR)^[24]、结构相似度 (structural similarity, SSIM)^[24] 和光谱角制图 (spectral angle mapper, SAM)^[25]。较大的 PSNR 和 SSIM 值意味着较好的去噪性能, 而较小的 SAM 值表明去噪后图像较高的光谱保真度。

3 实验结果

3.1 高斯噪声下的实验结果

在高斯噪声条件下, 本文在干净高光谱图像中加入具有不同方差 σ 的零均值加性高斯白噪声以产生噪声图像。本文的方法和对比方法在 ICVL 数据集上的定量结果如

表 1 所示。本文的方法优于所有的对比方法(最好的结果已经加粗以方便查看)。在 $\sigma = 50$ 的噪声水平下,本文的方法在 PSNR 上提高了 0.62 dB。此外,通过对比发现以往的深度学习方法在处理高强度噪声下的高光谱图像时可以取得比传统方法更好的效果,但在处理低强度噪声下的高光谱图像时表现一般。但本文提出的方法在低噪声下仍然取得了比较好的效果,显示出了本文提出的模型去噪性能优越。为了更直观的展示本文所提出的方法的去噪性能,在图 5 中展示了在高斯噪声方差为 50 的强度下不同方法去噪之后的高光谱图像。本文选取了高光谱图像的第 20 波段作为可视化图像并进行了局部放大处理。可

以看出,BM4D 方法导致空间不连续,丢失了许多高频方向细节。LLRT 获得了相对较好的效果,但它们仍然丢失了图像的精细纹理。QRNN3D 和 TRQ3D 的结果在树叶边缘附近有明显的伪影。但本文提出的方法在边缘细节处处理的更为细致,验证了局部增强模块在边缘细节上处理的优越性。同时,通过对 ICVL 数据集中 nachal_0823-1040.mat 图像在高斯噪声强度为 Blind 的条件下指定像素点位置(216,216)的光谱曲线进行对比,如图 6 所示。本文提出的方法在大部分波段都是最契合干净高光谱图像的光谱曲线。证明在光谱维度上本文的方法也得到了最满意的恢复效果。

表 1 高斯噪声

噪声强度	评价指标	噪声图像	BM4D ^[13]	TDL ^[12]	LLRT ^[11]	NG Meet ^[23]	QRNN3D ^[9]	TRQ3D ^[17]	NSSNN ^[21]	本文
30	PSNR	18.59	36.79	39.03	40.28	42.04	41.40	41.95	42.18	42.79
	SSIM	0.647	0.989	0.994	0.993	0.997	0.995	0.997	0.997	0.998
	SAM	0.559 1	0.083 5	0.058 9	0.052 7	0.043 1	0.044 6	0.040 3	0.040 8	0.038 6
50	PSNR	14.15	33.96	36.31	36.72	39.33	39.21	39.89	39.94	40.56
	SSIM	0.422	0.979	0.990	0.983	0.994	0.994	0.995	0.996	0.996
	SAM	0.777 8	0.114	0.079 8	0.077 1	0.058 1	0.057 3	0.051 5	0.052 7	0.042 9
70	PSNR	11.23	32.14	34.77	35.11	37.80	37.24	38.32	38.36	38.76
	SSIM	0.279	0.968	0.986	0.977	0.984	0.984	0.994	0.994	0.995
	SAM	0.930 2	0.139 9	0.094 9	0.090 6	0.067 2	0.073 2	0.062	0.063 3	0.060 5
Blind	PSNR	17.59	36.74	38.93	40.24	42.01	41.59	41.96	42.16	42.68
	SSIM	0.587	0.989	0.994	0.993	0.997	0.995	0.997	0.997	0.998
	SAM	0.659 1	0.083 9	0.059 9	0.053 2	0.044 2	0.045 6	0.040 5	0.040 9	0.038 4

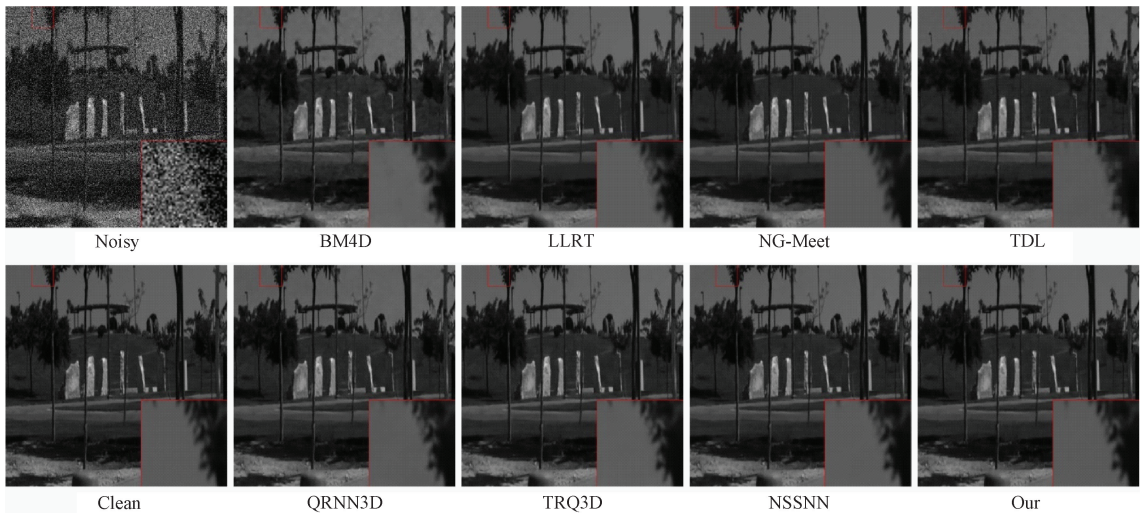


图 5 ICVL 数据集在高斯噪声强度为 50 的条件下不同去噪方法的可视化效果对比

3.2 复杂噪声下的实验结果

复杂噪声主要分为非 IID 高斯噪声、截止噪声、脉冲噪声以及它们的混合噪声等 5 种噪声类型。本文的方法和

对比方法在 ICVL 数据集上的定量结果如表 2 所示。在混合噪声强度下的去噪效果的可视化图如图 7 所示。可以得出结论,在复杂噪声环境下,本文的方法优于其他方法。

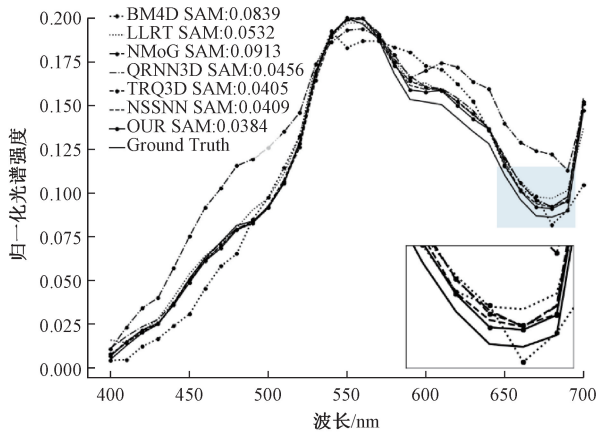


图 6 高斯噪声(Blind)条件下去噪后光谱曲线对比图

在最复杂的混合噪声条件下提升了 0.6 dB。在复杂噪声的环境下,肉眼可见的发现传统方法的去噪效果没有基于

深度学习的方法效果好。同时,本文发现在含有脉冲噪声的条件下尤为明显,在脉冲噪声或混合噪声条件下,传统方法几乎都无法恢复出一个干净的高光谱图像,优化后的图像中存在很明显的脉冲噪声。而深度学习方法却取得了相当好的效果。经过推测,脉冲噪声和混合噪声的存在会使得高光谱图像在空间维度和光谱维度上失去一定的关键信息。如果没有干净的高光谱图像作为指导,手工设置的先验信息不足以在噪声图像的基础上很好地进行优化处理。而本文的方法不仅关注高光谱图像的非局部空间相似性,而且通过 Top-k 稀疏自注意力利用相对干净的波段进行辅助去噪,在混合噪声(Mixture)条件下,对利用不同去噪方法去噪后得到的高光谱图像在指定像素点位置的光谱曲线进行分析。如图 8 所示,本文提出的稀疏空谱 Transformer 模型得到的光谱曲线是最契合干净高光谱图像的光谱曲线。证明了本文的方法无论是在空间维度还是光谱维度下都是最好的方法。

表 2 复杂噪声

噪声类型	评价指标	噪声图像	BM4D ^[13]	LLRT ^[11]	NMoG ^[22]	QRNN3D ^[9]	TRQ3D ^[17]	NSSNN ^[21]	本文
Non-IID	PSNR	18.81	34.74	31.46	32.01	42.88	43.49	43.58	44.31
	SSIM	0.608	0.975	0.925	0.946	0.995	0.998	0.998	0.998
	SAM	0.585 6	0.115 3	0.183 6	0.150 1	0.038 9	0.033 6	0.036 1	0.033 6
Deadline	PSNR	17.52	30.93	28.64	31.08	42.05	42.75	43.20	43.57
	SSIM	0.554	0.946	0.895	0.945	0.991	0.997	0.998	0.998
	SAM	0.647 4	0.145 2	0.202 8	0.151 8	0.040 3	0.035 9	0.036 9	0.035 3
Impulse	PSNR	17.46	28.47	25.34	29.21	42.23	43.22	43.59	44.00
	SSIM	0.536	0.811	0.744	0.867	0.991	0.997	0.998	0.998
	SAM	0.595 9	0.227 7	0.309 8	0.180 4	0.043 3	0.034 6	0.035 6	0.034 5
Stripe	PSNR	18.06	34.23	29.86	33.83	42.29	42.95	43.43	43.98
	SSIM	0.581	0.975	0.903	0.982	0.996	0.997	0.998	0.998
	SAM	0.607 7	0.114 7	0.210 5	0.111 3	0.039 7	0.035 3	0.035 9	0.033 5
Mixture	PSNR	14.05	28.97	26.52	26.19	39.14	39.37	39.60	40.19
	SSIM	0.390	0.911	0.860	0.868	0.995	0.995	0.995	0.996
	SAM	0.776 8	0.167 8	0.281 6	0.256 1	0.051 5	0.051 6	0.052 9	0.048 2

3.3 消融实验

为了验证本文所提出的模块的有效性,本文在 ICVL 数据集上针对在高斯噪声强度为 $\sigma = 50$ 的情况下对本文所提出的各个模块进行了消融实验。成分分析如表 3 所示。本文研究了稀疏空谱 Transformer 模块中的各个子组件的实际影响,包括验证空间维度 Transformer(SpaT)的局部增强模块、光谱通道混合模块(SCM)以及光谱维度 Transformer(SpeT)的 Top-k 稀疏自注意力模块。在去除局部增强模块的原始空间维度 Transformer 的情况下,PSNR 降低了 0.73 dB。当同时去除光谱维度 Transformer 下的光谱通道混合模块和

Top-k 稀疏自注意力的情况下,PSNR 降低了 0.94 dB,而且在去除光谱通道混合模块,单独添加 Top-k 稀疏自注意力模块的情况下,PSNR 降低了 0.35 dB。可以看出,光谱通道混合模块在计算量不显著增加的同时显著提升了去噪性能,验证了光谱通道混合模块的有效性。另一方面,Top-k 稀疏自注意力虽然增加了一定的参数量,但在和空间维度 Transformer 联合处理之后得到了很不错的效果。在平衡计算成本和去噪效果的考虑下,本文提出的各个模块在经过合理的联合处理之后达到了最好的性能,证明了本文提出的稀疏空谱 Transformer 模块的有效性。

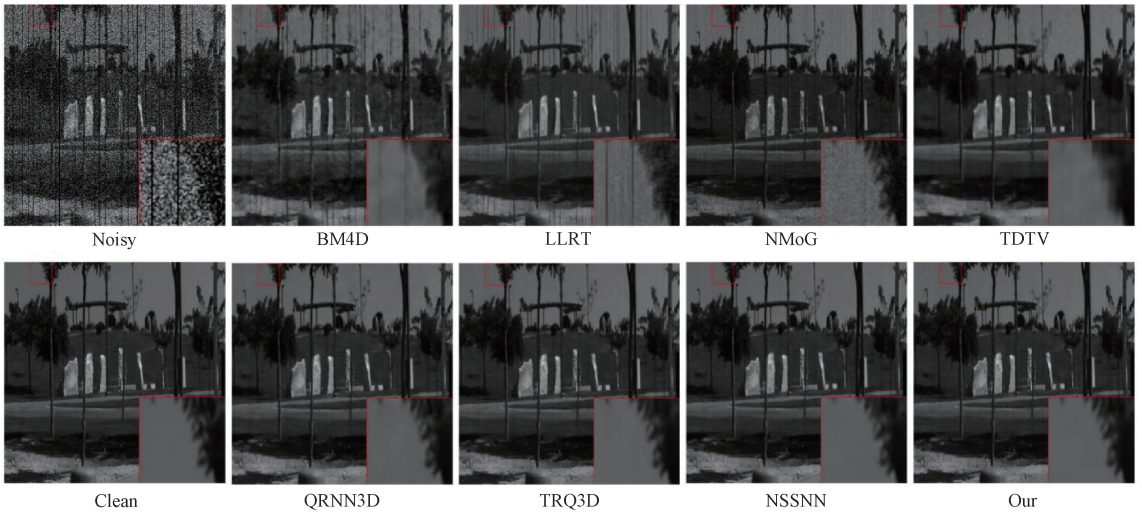


图 7 ICVL 数据集在复杂噪声(Mixture)的条件下不同去噪方法的可视化效果对比

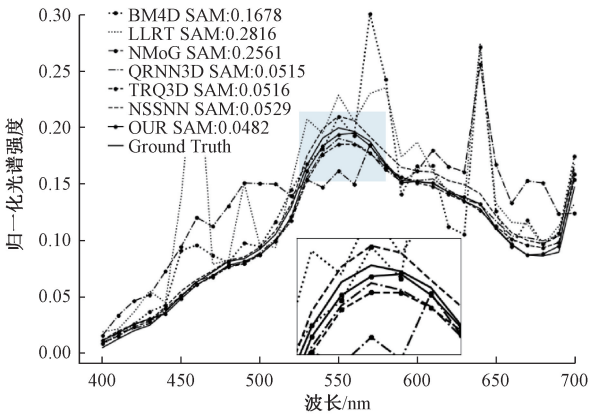


图 8 复杂噪声(Mixture)条件下去噪后光谱曲线对比图

表 3 稀疏空间-光谱自注意力模块子组件消融实验

SpaT	SCM	SpeT	Params/M	PSNR/dB
✓	×	×	1.09	39.62
×	×	✓	2.03	39.75
×	✓	✓	2.52	39.83
✓	×	✓	2.45	40.21
✓	✓	✓	2.94	40.56

4 结 论

通过高斯噪声和复杂噪声条件下的对比,本文提出的用于高光谱图像去噪的稀疏空谱 Transformer 模型不仅仅保留了高光谱图像的空间特征,还在保留空间特征的基础上对高光谱图像的光谱特征也进行了很好的建模,有效的利用了空谱耦合关联性。本文的主要贡献是对 Transformer 中的自注意力模块进行改进,提出了局部增强模块和 Top-k 稀疏自注意力结构。从而提升 Transformer 对高光谱图像空间与光谱特征的联合提取能

力。通过局部增强模块解决 Transformer 在提取特征时过平滑的问题。在通过空间维度 Transformer 提取高光谱图像的非局部空间相似性的同时,利用 Top-k 稀疏自注意力模块代替 Transformer 的原始自注意力模块,通过保留特征更相关的前 K 个光谱通道信息进行辅助去噪,提升在光谱维度的去噪性能。通过空间维度 Transformer 和光谱维度 Transformer 的联合处理同时考虑到高光谱图像的非局部空间相似性和全局光谱相关性,充分利用了高光谱图像的空谱耦合相关性。通过对比传统方法和其他先进的深度学习方法在各种噪声退化条件下的定量和可视化分析,证明了本文的方法的优越性。

参考文献

- [1] IM J, JENSEN J R. Hyperspectral remote sensing of vegetation[J]. Geography Compass, 2010, 2(6):1943-1961.
- [2] 王敏,陈金勇,孙康,等. 基于神经网络的高光谱图像异常目标检测方法[J]. 国外电子测量技术, 2019, 38(1): 96-101.
- [3] KERSTING K, XU Z, WAHABZADA M, ET AL. Pre-symptomatic prediction of plant drought stress using dirichlet-aggregation regression on hyperspectral images[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2012, 26(1): 302-308.
- [4] FEI B. Hyperspectral imaging in medical applications[M]. Data Handling in Science and Technology, 2019.
- [5] 刘雪峰,刘佳明,付民. 生成对抗网络扩充样本用于高光谱图像分类[J]. 电子测量技术, 2022, 45(3): 146-152.
- [6] ZHAO X L, ZHANG H, JIANG T X, et al. Fast algorithm with theoretical guarantees for constrained low-tubal-rank tensor recovery in hyperspectral images

- denoising[J]. *Neurocomputing*, 2020, 413: 397-409.
- [7] MAGGIONI M, KATKOVNIK V, EGIAZARIAN K, et al. Nonlocal transform-domain filter for volumetric data denoising and reconstruction[J]. *IEEE Transactions on Image Processing*, 2012, 22(1): 119-133.
- [8] PENG Y, MENG D, XU Z, et al. Decomposable nonlocal tensor dictionary learning for multispectral image denoising [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 2949-2956.
- [9] WEI K, FU Y, HUANG H. 3-D quasi-recurrent neural network for hyperspectral image denoising[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 32(1): 363-375.
- [10] DABOV K, FOI A, KATKOVNIK V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering [J]. *IEEE Transactions on Image Processing*, 2007, 16(8): 2080-2095.
- [11] CHANG Y, YAN L, ZHONG S. Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 4260-4268.
- [12] PENG Y, MENG D, XU Z, et al. Decomposable nonlocal tensor dictionary learning for multispectral image denoising [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 2949-2956.
- [13] MAGGIONI M, KATKOVNIK V, EGIAZARIAN K, et al. Nonlocal transform-domain filter for volumetric data denoising and reconstruction [J]. *IEEE Transactions on Image Processing*, 2012, 22(1): 119-133.
- [14] YUAN Q, ZHANG Q, LI J, et al. Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 57(2): 1205-1218.
- [15] DAI Z, CAI B, LIN Y, et al. Up-detr: Unsupervised pre-training for object detection with transformers[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 1601-1610.
- [16] WANG Z, CUN X, BAO J, et al. Uformer: A general u-shaped transformer for image restoration [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 17683-17693.
- [17] PANG L, GU W, CAO X. TRQ3DNet: A 3D quasi-recurrent and transformer based network for hyperspectral image denoising[J]. *Remote Sensing*, 2022, 14(18): 4598.
- [18] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [C]. *International Conference on Learning Representations*, 2020.
- [19] ARAD B, BEN-SHAHAR O. Sparse recovery of hyperspectral signal from natural RGB images[C]. *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, 2016: 19-34*.
- [20] KINGA D, ADAM J B. A method for stochastic optimization[C]. *International Conference on Learning Representations(ICLR)*, 2015, 5: 6.
- [21] FU G, XIONG F, LU J, et al. Nonlocal spatial-spectral neural network for hyperspectral image denoising[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-16.
- [22] CHEN Y, CAO X, ZHAO Q, et al. Denoising hyperspectral image with non-iid noise structure[J]. *IEEE Transactions on Cybernetics*, 2017, 48(3): 1054-1066.
- [23] HE W, YAO Q, LI C, et al. Non-local meets global: An integrated paradigm for hyperspectral denoising [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 6868-6877.
- [24] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: From error visibility to structural similarity [J]. *IEEE Transactions on Image Processing*, 2004, 13(4): 600-612.
- [25] YUHAS R H, BOARDMAN J W, GOETZ A F H. Determination of semi-arid landscape endmembers and seasonal trends using convex geometry spectral unmixing techniques[C]. *JPL, Summaries of the 4th Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop, 1993*.

作者简介

杨智翔, 硕士研究生, 主要研究方向为图像处理。

E-mail: 810412861@qq.com

孙玉宝, 教授, 博士, 主要研究方向为模式识别、高光谱遥感影像处理与分析等。

E-mail: sunyb@nuist.edu.cn

白志远, 硕士研究生, 主要研究方向为文本生成图像。

E-mail: 973295946@qq.com

栾鸿康, 硕士研究生, 主要研究方向为图像处理。

E-mail: 913876018@qq.com