

# 高频信息强化的双分支高光谱图像超分辨率网络<sup>①</sup>



侯钧译, 杨 锦, 边太成, 朱习军

(青岛科技大学 信息科学技术学院, 青岛 266061)

通信作者: 朱习军, E-mail: zhuxj990@163.com

**摘 要:** 高光谱图像 (hyperspectral image, HSI) 的窄光谱波段为许多视觉任务提供了丰富信息, 但也给特征提取带来了挑战. 尽管许多研究者提出了各种深度学习方法, 但尚未充分结合这些架构的优势. 因此, 本文提出了一种基于高频信息强化的双分支高光谱图像超分辨率网络 (HFEDB-Net), 将卷积神经网络 (convolutional neural network, CNN) 的图像空间特征提取优势与 Transformer 的自适应能力和长距离依赖提取优势相结合, 有效地提取了 HSI 的空间和光谱信息. HFEDB-Net 由高频信息强化分支和主干分支组成. 在高频信息强化分支中, 通过拉普拉斯金字塔提取低分辨率和高分辨率 HSI 的高频信息, 并将结果作为高频分支的输入和标签, 采用光谱强化 Transformer 来作为该分支的方法. 在主干分支中, 使用结合通道注意力的 CNN 充分提取空间特征和光谱信息. 最后将两个分支的结果通过 CNN 进行结合以得到最终的重建图像. 此外, 采用多头注意力和多尺度策略分别改进了 Transformer 的注意力机制和编码器层, 以更好地提取 HSI 的空间和光谱信息. 实验结果表明, HFEDB-Net 在两个公开数据集上的定量评价指标和视觉效果上优于当前主流方法.

**关键词:** 高光谱图像; 超分辨率重建; 自注意力机制; 神经网络

引用格式: 侯钧译, 杨锦, 边太成, 朱习军. 高频信息强化的双分支高光谱图像超分辨率网络. 计算机系统应用, 2024, 33(10): 217-227. <http://www.c-s-a.org.cn/1003-3254/9643.html>

## High-frequency Enhanced Dual-branch Hyperspectral Image Super-resolution Network

HOU Jun-Yi, YANG Jin, BIAN Tai-Cheng, ZHU Xi-Jun

(School of Information Science and Technology, Qingdao University of Science & Technology, Qingdao 266061, China)

**Abstract:** The narrow spectral bands of hyperspectral images (HSI) provide rich information for many visual tasks, but also pose challenges for feature extraction. Despite various deep learning methods proposed by researchers, the advantages of these architectures are not fully combined. Therefore, this study proposes a high-frequency enhanced dual-branch hyperspectral image super-resolution network (HFEDB-Net) that effectively extracts spatial and spectral information of HSI by integrating the image spatial feature extraction advantage of convolutional neural network (CNN) with the adaptive capability and long-distance dependency extraction advantage of Transformers. HFEDB-Net consists of a high-frequency information enhancement branch and a backbone branch. In the high-frequency information enhancement branch, the high-frequency information of low-resolution and high-resolution HSI is extracted by using Laplacian pyramids, and the results serve as the input and label for the high-frequency branch. A spectral-enhanced Transformer is employed as the feature extraction method for this branch. In the backbone branch, a CNN with channel attention is utilized to extract spatial features and spectral information comprehensively. Finally, the results from both branches are combined through CNN to obtain the final reconstructed image. Additionally, the attention mechanism and encoder layers of the Transformer are respectively improved by using multi-head attention and multi-scale strategies to

① 基金项目: 青岛市科技惠民示范专项 (23-2-8-smjk-20-nsh); 山东省产教融合研究生联合培养示范基地项目 (2020-19)

收稿时间: 2024-04-03; 修改时间: 2024-04-29; 采用时间: 2024-05-09; csa 在线出版时间: 2024-08-28

CNKI 网络首发时间: 2024-08-29

better extract spatial and spectral information from HSI. Experimental results demonstrate that HFEDB-Net outperforms current state-of-the-art methods in terms of quantitative evaluation metrics and visual effects on two public datasets.

**Key words:** hyperspectral image (HSI); super-resolution reconstruction; self-attention mechanism; neural network

高光谱图像 (hyperspectral image, HSI) 凭借丰富连续的窄光谱波段信息, 在许多视觉领域如矿产勘察<sup>[1]</sup>、蔬菜新鲜检测<sup>[2]</sup>等有着不可替代的优势和巨大的应用潜力. 但由于成像硬件的限制, 高光谱图像的高空间分辨率和丰富光谱信息往往无法兼得, 进而导致其无法得到许多对空间分辨率要求较高的计算机视觉任务的青睐. 所以如何从软件方面提升高光谱图像的空间分辨率 (hyperspectral image super-resolution, HSISR) 已然成为一个重要的研究课题.

传统的提升图像空间分辨率的方法包括插值法 (如: 双三次插值、线性插值、双线性插值、最近邻插值)、稀疏表示法等. 这些传统机器学习方法往往对处理图像的先验假设较简单, 无法充分考虑和学习到图像的复杂空间信息规律, 致使最终重建效果差强人意. 随着硬件计算力的不断提升, 越来越多复杂高效的深度学习方法被应用于图像超分辨率任务, 如卷积神经网络<sup>[3]</sup>、Transformer<sup>[4]</sup>和生成式网络<sup>[5]</sup>等, 不断提高图像重建质量. 但若将这些方法不加特定改动调整直接用于 HSISR 任务上不但会带来较差的空间重建效果, 光谱维度的规律也会被破坏. 这是因为相对于普通 RGB 图像的超分辨率重建而言, HSISR 的目标不仅是从小空间分辨率图像 (low resolution image, LR) 中恢复出足够多的空间细节以得到高空间分辨率图像 (high resolution image, HR), 还要保证 HR 的光谱维度规律的重建质量. 此外, HSI 的空间信息在光谱维度上分布稀疏, 需要使用特定的方法策略才能达到较好的效果.

HSISR 方法可以被划分为两大类, 分别是图像融合高光谱图像超分辨率和单幅高光谱图像超分辨率方法. 基于融合的高光谱图像超分辨率方法通过利用高光谱低空间分辨率图像的光谱信息和其对应的相同场景的高空间分辨率低光谱图像的丰富空间信息, 生成高光谱高空间分辨率图像. 采用的主流方法包括深度学习<sup>[6]</sup>、矩阵分解<sup>[7]</sup>和稀疏表示方法<sup>[8]</sup>等. 但基于融合的方法所需的高适配的同场景信息难以获得, 所以许多研究者更倾向于单幅高光谱图像超分辨率重建方法.

单幅高光谱图像超分辨率重建方法需要应对数据集规模小且数据维度大带来的困难, 许多研究者想到利用数据降维方法来降低数据维度<sup>[9]</sup>. 但是对于 HSI 来说, 因为 HSI 独特的光谱-空间特性, 在空间区域冗余的信息在光谱维度上确差异明显, 在光谱维度上冗余的信息在空间上却不一定相同. 因此, 传统的数据降维方法无法很好地处理 HSI 数据的特殊性.

为了解决这个问题, 一些研究者提出了基于光谱-空间联合降维的方法, 以更好地保留 HSI 数据的特征<sup>[10,11]</sup>. 这些方法将光谱和空间维度的信息进行联合建模, 并通过联合降维来减少数据维度. 例如, 一种常用的方法是将 HSI 数据表示为张量形式, 其中光谱维度和空间维度被同时考虑. 然后应用张量分解或张量低秩近似等技术, 将张量数据进行降维, 同时保留光谱和空间维度的相关性. 但数据降维依旧存在对 HSI 的有效信息进行丢失的情况, 进而影响到重建结果.

另有研究者从深度学习模型架构角度入手, 提出特定的光谱-空间特征提取策略并取得了可观的结果<sup>[12]</sup>. 受以上思路的启发, 本文提出了基于光谱强化 Transformer 和注意力卷积的特征强化高光谱图像超分辨率网络, 结合了 Transformer 捕捉高频特征和卷积神经网络空间特征提取能力的优势, 采用了双分支的网络架构, 增强对 HSI 高频信息重建能力的同时也保证了光谱-空间信息重建能力. 本文主要贡献如下:

(1) 提出了具有额外监督的光谱强化 Transformer 高频特征分支, 使用拉普拉斯金字塔提取后的输入和标签的高频信息分别作为分支的输入与标签, 进而加强对 HSI 细节的重建能力.

(2) 提出了 Multi-head ESSA 注意力机制, 并结合提出的多尺度编码器层, 加强了对高频多样化特征的提取能力和对上下文特征信息的利用.

(3) 提出了双分支高频信息增强高光谱图像超分辨率网络 (high-frequency enhanced dual-branch hyperspectral image super-resolution network, HFEDB-Net), 充分提取了 HSI 的空间和光谱信息, 在两个公开数据集上都获得了最先进的重建效果.

## 1 相关工作

### 1.1 基于卷积神经网络的高光谱图像超分辨率重建

卷积神经网络 (convolutional neural network, CNN) 因具备的平移不变性和尺度不变性使得其备受许多研究者的青睐, 在近年来产生了大量的不同架构如残差连接、密集连接、金字塔、U-Net 网络等, 不断刷新各个计算机视觉任务的记录, 而对于 HSISR 任务也不例外. Liebel 等人<sup>[13]</sup>提出了基于 CNN 的高光谱超分辨率图像重建方法. 他们将 SRCNN<sup>[14]</sup>引申到高光谱超分辨率任务领域, 通过学习低分辨率和高分辨率图像之间的映射关系, 实现了高质量的超分辨率重建. 然而其网络深度有限, 限制了对高光谱图像的进一步表达能力. 之后 Li 等人<sup>[15]</sup>提出了基于深度残差网络的高光谱超分辨率图像重建方法. 他们引入了残差学习的思想, 设计了一个具有多个残差块的网络结构. 通过逐层的特征提取和重建, 他们有效地提高了超分辨率图像的质量. 但由于 2D 卷积无法捕捉通道间关系的限制, 导致重建图像的光谱失真情况比较严重.

为解决上述问题, 研究者们提出了一些方法, 例如, 使用 3D 卷积替代 2D 卷积, 进行高光谱图像的特征提取<sup>[16]</sup>, 或引入通道注意力机制以增强整个网络对跨通道信息的关注<sup>[17]</sup>, 从而提升对 HSI 光谱特征的表征能力. 然而, 卷积神经网络依靠滑动窗口机制来进行特征映射, 导致其在捕捉长距离依赖关系和适应数据方面的能力有限, 仍难以对复杂光谱信息进行利用和重建.

### 1.2 视觉 Transformer

Transformer 最初由 Vaswani 等人<sup>[18]</sup>提出, 该模型在序列到序列的自然语言处理领域取得了重要的突破. 随后, Dosovitskiy 等人<sup>[19]</sup>提出了 Vision Transformer (ViT) 将 Transformer 应用于图像处理领域, 并实现了端到端的视觉处理方法. 这引发了对 ViT 的广泛关注, 开始吸引许多计算机视觉任务研究者的兴趣.

在图像超分辨率任务中, Liang 等人<sup>[20]</sup>提出了一个强大的图像超分辨率基准模型 SwinIR, 该模型利用 ViT 架构进行深层特征提取, 包括多个 SwinIR 层和残差连接, 具有出色的性能和泛化能力. 接着, Li 等人<sup>[21]</sup>提出了分层 Swin Transformer (HST) 网络来恢复低分辨率压缩图像, 该网络通过联合捕获分层特征表示并使用 Swin Transformer 来增强每个尺度的表示能力.

为了进一步提高恢复效果, Chen 等人<sup>[22]</sup>提出了一种新的混合注意力 Transformer (HAT), 该模型结合了通道注意力和基于窗口的自注意力方案, 以充分利用它们的互补优势, 激活更多的输入像素.

最早将 ViT 引入 HSISR 任务的是 Hu 等人<sup>[23]</sup>提出的 Fuserformer 模型, 该模型在高光谱图像超分辨率任务中取得了巨大的成功, 然而, 其没有在注意力机制本身角度对 HSISR 任务进行适配, ESSAformer<sup>[24]</sup>提出了一种针对 HSI 的高效注意力机制, 充分考虑了高光谱领域的特点, 并引入了通道级归纳偏置, 进而提高了重建图像的光谱质量. 但其使用的注意力机制在映射  $Q$ 、 $K$ 、 $V$  三向量时的能力有限, 缺乏特征提取的多样性, 并且针对图像空间信息的利用不够充分, 缺少对上下文信息的交互.

为了解决上述存在的问题, 本文提出了一种基于高频信息强化的双分支高光谱图像超分辨率网络 HFEDB-Net. 通过结合 Transformer 和卷积神经网络, HFEDB-Net 能够有效地重建高光谱图像的高频信息和光谱-空间细节. 通过使用拉普拉斯金字塔提取输入和标签的高频信息, 并构建了基于多头 ESSA 的多尺度 Transformer 高频特征提取分支, 从而有效提升了对高频特征的提取能力和上下文特征信息的利用效果.

## 2 网络结构

### 2.1 整体架构

基于深度学习的 HSISR 任务本质是通过损失函数 Loss 训练出一组能从低分辨率高光谱图像  $S_{LR}$  到超分辨率高光谱图像的属于映射函数  $F$  的参数  $\theta$ , 并通过反向传播使重建的图像尽可能地与其对应的真实高分辨率高光谱图像  $S_{HR}$  相同, 用公式表达为:

$$\theta^* = \arg \min_{\theta} (Loss(F(S_{LR}; \theta), S_{HR})) \quad (1)$$

其中,  $F(S_{LR}; \theta)$  表示通过将具有参数  $\theta$  的映射函数  $F$  应用于低分辨率高光谱图像  $S_{LR}$  而获得的重建高光谱图像. 目标是找到最小化损失函数  $Loss$  的最佳参数  $\theta^*$ . 因所提网络是双分支结构, 且分支包含独立监督, 所以将所提出的方法的目标进一步精确为:

$$\begin{cases} \theta_1^* = \arg \min_{\theta_1} (Loss_1(F_1(S_{LR}, F_2(S_{sharpLR}; \theta_2); \theta_1), S_{HR})) \\ \theta_2^* = \arg \min_{\theta_2} (Loss_2(F_2(S_{sharpLR}; \theta_2), S_{sharpHR})) \end{cases} \quad (2)$$



其中,  $\theta_1^*$  表示将主干分支的映射函数  $F_1$  应用于参数为  $\theta_1$  的低分辨率高光谱图像  $S_{LR}$  得到的重建高光谱图像, 其中  $F_1$  输入的额外参数为高频分支映射函数  $F_2$  的输出. 损失函数  $Loss_1$  测量的是重建的高光谱图像和真实高分辨率高光谱图像之间的差异, 而损失函数  $Loss_2$  衡量的是  $S_{LR}$  通过拉普拉斯算出的高频图像与真实图像算出的高频图像  $S_{sharpHR}$  之间的差异. 在提出的 HFEDB-Net 网络中我们使用总变差损失函数<sup>[25]</sup>作为  $Loss_1$ , 使用 L1 损失函数作为  $Loss_2$ . 总变差损失函数公式为:

$$\mathcal{L}_{SSTV} = \frac{1}{N} \sum_{n=1}^N (\nabla_h F(S_{LR}^n)_1 + \nabla_w F(S_{LR}^n)_1 + \nabla_c F(S_{LR}^n)_1) \quad (3)$$

其中,  $\nabla_h$  与  $\nabla_w$  表示为所提出的网络生成最终的重建高光谱图像在高和宽两个维度上的梯度. 而  $\nabla_c$  则表示最终结果在光谱维度上的梯度. 通过将总变差损失函数作为  $Loss_1$  可以有效地保持重建图像的结构与边缘的连续性和一致性, 并减少重建图像中的伪影和失真现象, 进而提高重建图像的质量. 在高频分支所使用的拉普拉斯金字塔原理为将输入图像减掉其下采样后紧接着上采样处理后的图像, 用公式表达为:

$$I_{sharp} = I - Up(Down(I)) \quad (4)$$

其中,  $I$  与  $I_{sharp}$  分别表示输入图像与输出的高频信息图像,  $Up$  与  $Down$  分别表示插值上采样和插值下采样操作, 在提出的网络中, 我们使用的插值方法为双线性插值法. 提出的网络总结构如图 1 所示.

### 2.2 主干分支

为了提取 HSI 的空间和光谱特征, 主干分支由残差空间提取块与通道注意力空间-光谱块组成. 在空间提取块中, 使用  $3 \times 3$  卷积来学习图像的空间上不同特征, 结合激活函数来为模块增加更多的非线性, 同时利用残差机制加强了特征的传递且缓解梯度爆炸和消失的问题. 值得注意的是, 使用单次上采样操作会不可避免地造成上下文细节的丢失情况, 所以在主干分支中我们将亚像素卷积上采样操作放在了每 3 个连续的空间提取块之后, 通过采用逐步上采样策略, 避免了只使用一次上采样带来的噪声影响过大问题. 且亚像素卷积具有重建质量高、参数量小和有效利用低分辨率信息的优点, 与反卷积上采样相比, 重建图像也会更少出现伪影的情况. 空间提取块如图 1 左上方所示.

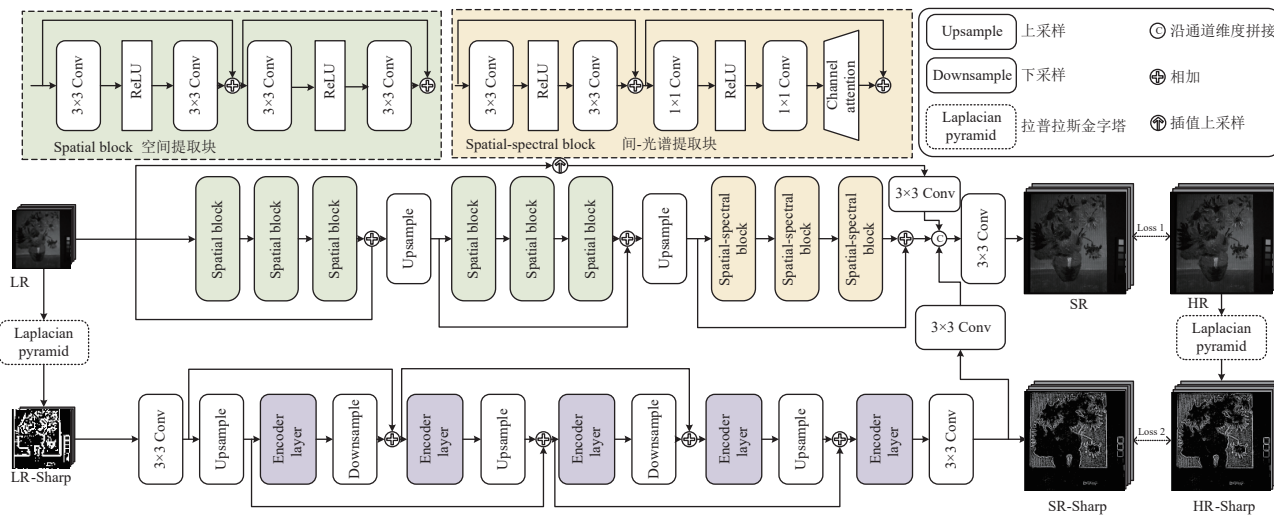


图 1 HFEDB-Net 网络总架构

对于空间-光谱提取块而言, 本文引入擅长捕捉通道之间关系的通道注意力机制. 同样结合  $1 \times 1$  卷积和激活函数, 增强了本块对空间和光谱特征的表达能力.

### 2.3 Transformer 高频特征分支

因 HSI 空间信息在光谱维度分布稀疏且不规律的特性, 卷积神经网络难以对其较高频的特征进行有效

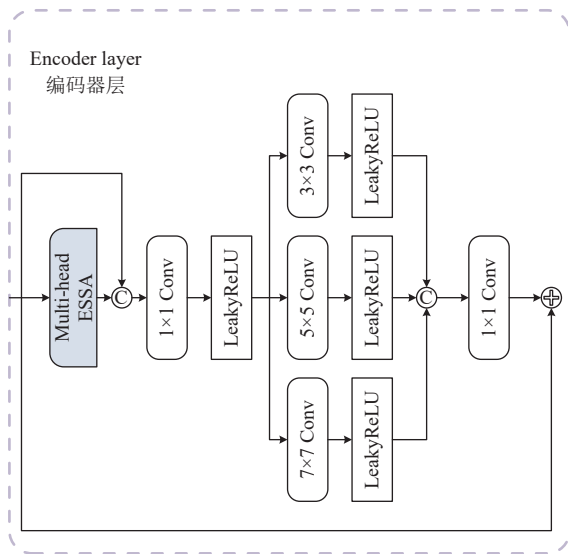
重建. 为了解决这个问题, 我们提出了具有额外监督的 HSI 高频特征超分辨率分支, 利用具有强自适应能力和擅于捕捉特征长距离依赖关系的 Transformer 方法. 本文设计了多头 ESSA 注意力机制, 采用光谱相关系数 (SCC) 用于替代原始的注意力矩阵, 并将归纳偏好引入模型中以便于训练. 并且将  $Q$ 、 $K$ 、 $V$  这 3 个向量

的相乘顺序进行变换,进而大幅减少了计算复杂度。

然而,ESSA无法有效地捕捉细粒度细节并充分利用多尺度的上下文信息。这是因为其仅在一个空间尺度上进行操作,并仅使用单个注意力头,从而无法完全捕捉复杂HSI中存在的多样视觉模式。

为了解决这些问题并进一步增强Transformer分支的性能,本文设计了多头ESSA和多尺度编码层。通过引入多头注意力机制,图像多样化的模式可以被分支所捕捉到,且能够同时关注不同的区域。此外,编码层提取的多尺度特征还更好的整合了不同抽象层次的上下文信息。

多头策略使得高频特征分支能够学习多组注意力权重,每组权重分别用于学习特定的高频特征模式规律、关注不同的空间特征位置。而多尺度策略使得分支网络能够在不同空间分辨率上捕捉丰富的上下文信息,有助于充分理解图像的高频局部细节和全局结构。



改进后的ESSA用公式表示如下:

$$\begin{cases} ESSA_i(Q_i, K_i, V_i) = \varphi(Q_i)(\varphi(K_i)^T V_i) \\ Z = \text{Concat}(ESSA_1, \dots, ESSA_h)W_o \end{cases} \quad (5)$$

其中,  $Q_i$ ,  $K_i$  和  $V_i$  分别表示第  $i$  个 ESSA 注意力头的查询向量, 键向量和值向量。  $Z$  表示改进后的 ESSA 注意力的输出向量。  $ESSA_i(Q_i, K_i, V_i)$  表示第  $i$  个 ESSA 注意力头的输出向量。  $h$  表示注意力头的数量。  $W_o$  是输出投影矩阵。

另外,我们在高频特征分支中还采用了逐步优化的结构策略,通过多次迭代优化模型输出,逐步提升上采样结果的质量。每一轮迭代都可以对先前的输出进行细化和改进,从而逐渐减小上采样后图像中的伪影和信息损失。通过逐步优化,高频特征分支可以更好地适应数据的细节和复杂性,提高最终超分辨率重建图像的准确性和真实性。高频特征分支的编码器层和多头ESSA注意力如图2所示。

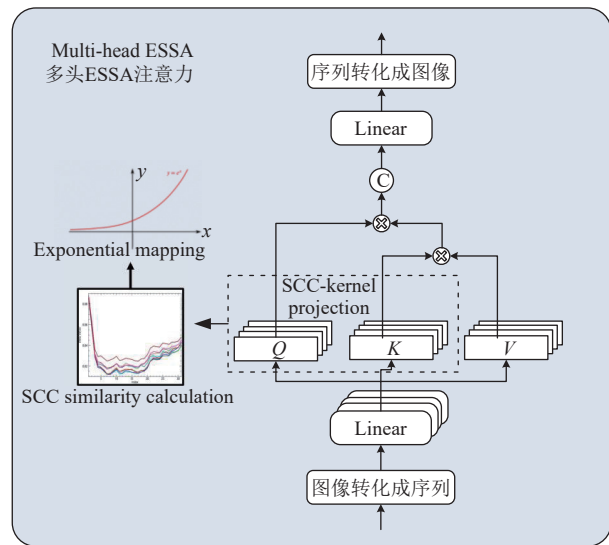


图2 高频分支中的编码器层与多头ESSA注意力机制

### 3 实验分析

#### 3.1 数据集

本文使用两个公开可用的高光谱数据集 Harvard 自然高光谱图像数据集<sup>[26]</sup>和 Chikusei 遥感高光谱图像数据集<sup>[27]</sup>进行了放大4倍和8倍的对比与消融实验,证明了提出的HFEDB-Net的有效性。

Harvard数据集包含50个真实世界高光谱图像,每个图像的尺寸为1392×1040×31。所有的31个光谱波段是由420–720 nm范围内每隔10 nm采样得到。其

中我们将40个图像用作训练集,10个图像用作测试集。为了适应模型训练,我们将训练集进行裁块处理。在放大因子为4的情况下,将每张训练集裁剪为64×64的互相重叠的图像块(重叠32个像素),通道数保持不变;最后将其降采样为1/4大小作为模型的输入。在放大因子为8的情况下,将每张数据集裁成128×128的重叠图像块(重叠64个像素),通道数保持不变;最后将其降采样为1/8大小作为模型的输入。Harvard数据集包含室内与室外的真实场景图像,拍摄的物体和场

景复杂多样, 蕴含丰富的空间和光谱信息。

Chikusei 数据集是一张高光谱遥感图像, 图像尺寸为  $2517 \times 2335 \times 128$ , 光谱范围为  $363\text{--}1018\text{ nm}$ 。由于图像中存在边缘信息缺失不连续的情况, 所以我们首先将图像从中间裁剪为  $2304 \times 2048$  的子区域。然后, 从 Chikusei 图像的顶部选择 4 个尺寸为  $512 \times 512 \times 128$  的区域作为测试集, 剩余部分作为训练集。同样, 裁块处理策略与 Harvard 相同。与 Harvard 等自然高光谱图像数据集相比, Chikusei 数据集具有更少数据、更多噪声、较小目标以及更多光谱波段的特点, 使针对其的重建任务更具挑战性。

### 3.2 评价指标和实验设置

在对实验结果进行定量图像质量评价时, 我们选择使用以下 6 个主流评价指标: 互相关性度量 (cross correlation, CC)、结构相似性 (structural similarity, SSIM)、峰值信噪比 (peak signal-to-noise ratio, PSNR)、相对全局精度误差 (error relative global accuracy, ERGAS)、均方根误差 (root mean square error, RMSE) 和光谱角映射度量 (spectral angle mapper, SAM) 全面地评估了各个方法的性能。另外, 为进一步从主观的角度评价重建图像的质量, 我们还将重建结果用多种方式进行呈现。

所提出的 HFEDB-Net 网络相关参数设置如下: 多头 ESSA 中的注意力头数量设置为 3, 编码器层中的多尺度模块采用  $3 \times 3$  卷积核、 $5 \times 5$  卷积核与  $7 \times 7$  卷积核的组合。多头 ESSA 中每个注意力头的  $Q$ 、 $K$  和  $V$  向量映射长度都设置为与输入向量长度相同。

### 3.3 对比方法设置

为了证明提出的工作的优越性和有效性, 我们在两个数据集的两种放大因子上对比了 4 种先进的高光谱图像超分辨率方法, 从可视化和定量指标分析等方面对各个方法的重建结果进行分析。以下是我们使用的对比方法。

(1) SSPSR<sup>[28]</sup>: SSPSR 采用了一种有效的重叠分组策略的卷积神经网络, 强化了对光谱信息的捕捉能力, 从而实现了显著的重建效果。

(2) MCNET<sup>[29]</sup>: MCNET 是一种混合卷积网络, 结合了 2D 和 3D 卷积, 能更有效地提取 HSI 图像的光谱空间特征。

(3) PDE-Net<sup>[30]</sup>: 一种使用概率启发的高光谱图像编码嵌入来进行高分辨率高光谱图像的迭代优化网络。

(4) ESSAformer: 使用了一种高效的基于光谱相关

系数 SCC 核的自注意力 (ESSA), 用于替代原始的注意力矩阵的 Transformer 方法。

### 3.4 实验结果与分析

我们将提出的方法在两个数据集上与其余先进方法进行对比, 从定量评价指标表 1 可以得到, HFEDB-Net 在两个公开数据集的 4 倍和 8 倍放大倍数下都取得了最好的结果。其中在 Harvard 数据集上的优势最大, 尤其是在 4 倍放大倍数的条件下, MPSNR 指标要高于 ESSAformer 0.4 dB 左右。这是因为所提出的方法利用构造基于 Transformer 的额外监督高频分支来加强对高频信息的捕捉和重建, 同时从捕捉空间细节能力、上下文信息能力以及对多样特征的提取能力方面改进了 ESSA 注意力机制和编码器层, 进一步提升了总体网络的重建能力, 保证了重建图像的质量。

SSPSR、MCNet 和 PDE-Net 都为基于卷积神经网络的方法, 从表 1 数据来看, 他们在大多数情况下要明显弱于基于 Transformer 方法的 ESSAformer, 这是因为卷积神经网络的特性, 其特有的平移不变性的归纳偏置使得卷积神经网络很难构建长距离的依赖关系, 而长距离的依赖关系对于高光谱高维数据来说是非常重要的。相反, Transformer 利用自注意力机制, 能够对输入的各个位置进行全局范围的关联和交互, 从而更好地捕捉 HSI 中的长距离依赖关系, 这使得基于 Transformer 的方法如 ESSAformer 在处理高光谱高维数据时具有优势。因此, 相对于基于卷积神经网络的方法, ESSAformer 能够更好地建模 HSI 图像中的空间和光谱信息, 从而实现更准确的重建结果。

但是从空间提取能力角度分析, 卷积神经网络的空间提取能力要强于 Transformer, 这是因为卷积神经网络局部感知和权值共享的特性。卷积层通过滑动窗口的方式对图像的局部进行映射感知, 从而有效地捕捉到图像中的空间特征。且卷积层中的权重共享也大幅度减少了网络的参数量, 进而一定程度上提高了网络的重建能力。所以同样在表 1 中可以看到, 在放大因子为 8 的情况下, HSI 空间信息的重建难度加大, ESSAformer 的重建效果要略低于基于卷积神经网络的 MCNet 和 PDE-Net。MCNet 通过混合使用 2D 与 3D 卷积来对高光谱图像的高维特征进行捕捉, 与单使用 3D 卷积的方法相比, 减少了网络的参数量。PDE-Net 则采用其提出的特征嵌入策略来对高光谱图像进行重建, 有效地保护了光谱信息, 避免了光谱失真现象的发生。各个方法的重建可视化结果见图 3、图 4。



表1 在 Harvard 与 Chikusei 两个数据集上放大倍数为 4 倍和 8 倍的重建结果

数据集	Models	倍数	平均PSNR (dB)↑	SAM↓	CC↑	RMSE↓	平均SSIM↑	ERGAS↓
Harvard	Bicubic	4	38.6821	3.2790	0.9650	0.0165	0.9207	4.7858
	SSPSR	4	40.7853	3.2051	0.9757	0.0131	0.9390	3.8914
	MCNet	4	40.9431	3.2828	0.9740	0.0126	0.9415	3.6942
	PDE-Net	4	40.8323	3.2152	0.9710	0.0127	0.9437	3.7114
	ESSAformer	4	41.0057	3.1942	0.9778	0.0130	0.9474	3.7558
	HFEDB-Net (Ours)	4	41.4032	3.1871	0.9803	0.0121	0.9531	3.3204
	Bicubic	8	35.4501	3.7950	0.9305	0.0238	0.8503	3.3666
	SSPSR	8	36.5656	3.6729	0.9422	0.0226	0.8612	3.2451
	MCNet	8	37.0585	3.5897	0.9475	0.0219	0.8893	3.0766
	PDE-Net	8	37.1495	3.6161	0.9483	0.0203	0.8982	3.0400
	ESSAformer	8	37.1083	3.8227	0.9488	0.0203	0.8965	3.0272
	HFEDB-Net (Ours)	8	37.4399	3.5761	0.9492	0.0198	0.9017	2.6629
Chikusei	Bicubic	4	37.3527	3.3750	0.9223	0.0159	0.9049	6.9682
	SSPSR	4	39.8653	2.7329	0.9478	0.0123	0.9191	5.6006
	MCNet	4	40.1363	2.6457	0.9448	0.0125	0.9528	5.2429
	PDE-Net	4	39.7341	2.7127	0.9469	0.0129	0.9493	5.3087
	ESSAformer	4	40.2490	2.6210	0.9502	0.0120	0.9597	5.1767
	HFEDB-Net (Ours)	4	40.5448	2.4314	0.9533	0.0118	0.9628	4.9720
	Bicubic	8	34.5766	4.9608	0.8311	0.0220	0.8240	4.7482
	SSPSR	8	35.3497	4.8404	0.8547	0.0215	0.8220	4.4871
	MCNet	8	35.4594	4.6012	0.8740	0.0208	0.8562	4.3052
	PDE-Net	8	35.4812	4.5691	0.8517	0.0201	0.8505	4.2571
	ESSAformer	8	35.4591	4.5704	0.8625	0.0204	0.8537	4.2830
	HFEDB-Net (Ours)	8	35.6280	4.4521	0.8748	0.0196	0.8633	4.0241

注: 表中“↑”表示数值越高则模型在该评价指标上的表现越好, 反之, “↓”表示数值越低则模型在该评价指标上的表现越好

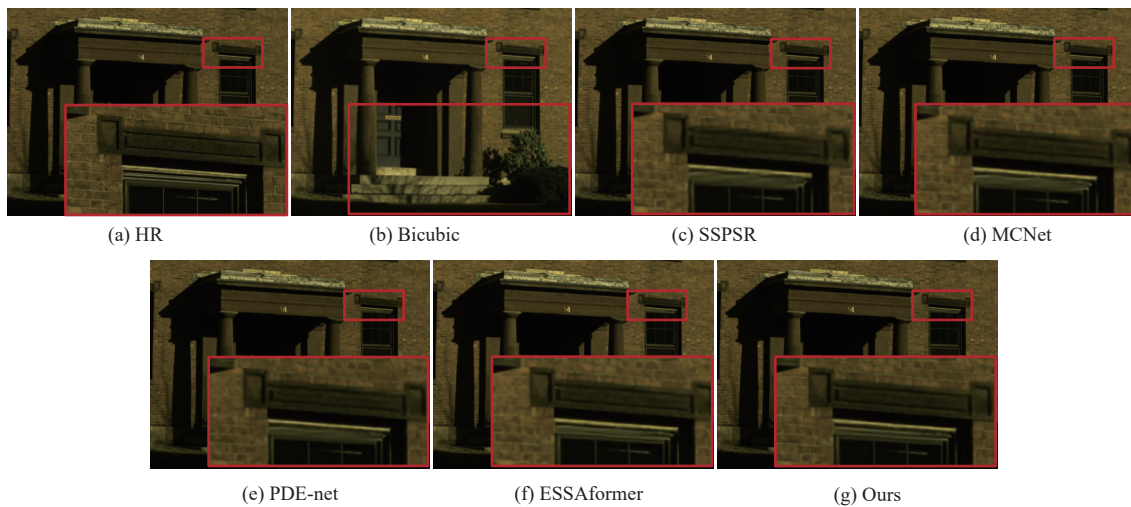


图3 Image1 重建图像的伪彩色图 (波段 31, 21, 11)

从重建效果对比图中不难看出, 所提出的方法的超分辨率图像与其他方法相比更接近标签真实图像 HR. 另外值得注意的是, 在高光谱图像的高频细节区域如砖块细节纹理和书本排列缝隙, HFEDB-Net 重建的质量要明显优于其他方法. 因为在高频分支中, HFEDB-Net 利用具有强自适应能力和捕捉特征长距离依赖关系的 Transformer 方法, 并利用提出的多头

ESSA 注意力和多尺度编码层, 丰富了对上下文信息和对空间细节的特征提取, 使得重建图像的质量进一步的提升.

为全面地对比各个方法的重建结果, 我们从光谱角度, 对比不同方法重建图像的质量. 我们利用平均误差图和平均误差曲线来从光谱角度对重建结果进行分析. 其中平均误差图的原理是通过将每个方法重建后

的图像与标签真实图像做差, 并将结果沿着光谱维度求平均, 最后生成平均误差热力图, 图上越明亮越清晰

的部分代表误差越大. 反之, 重建质量越高的部分则越暗无细节. 平均误差图热力图可见图 5.

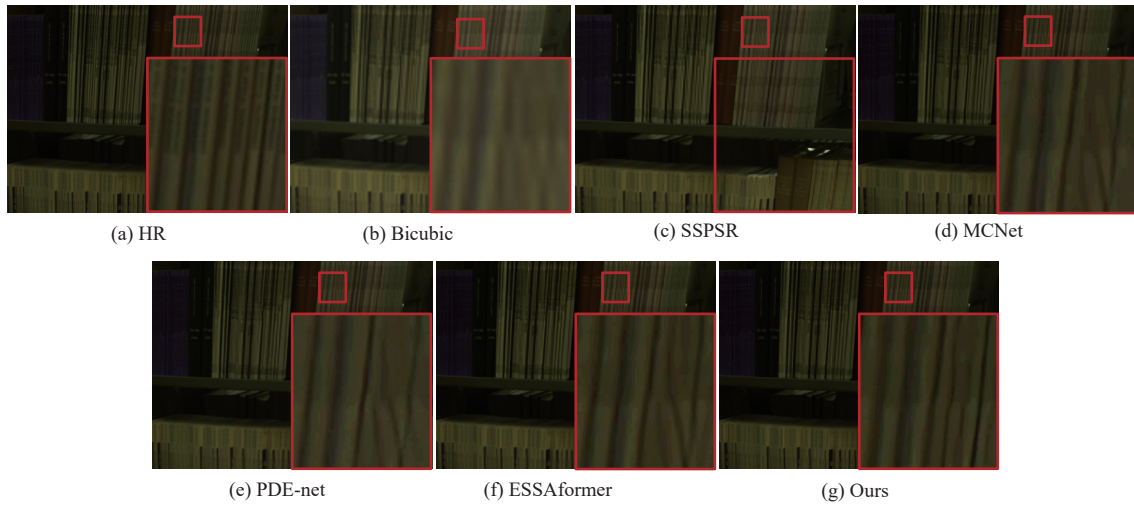


图 4 Image2 重建图像的伪彩色图 (波段 31, 21, 11)

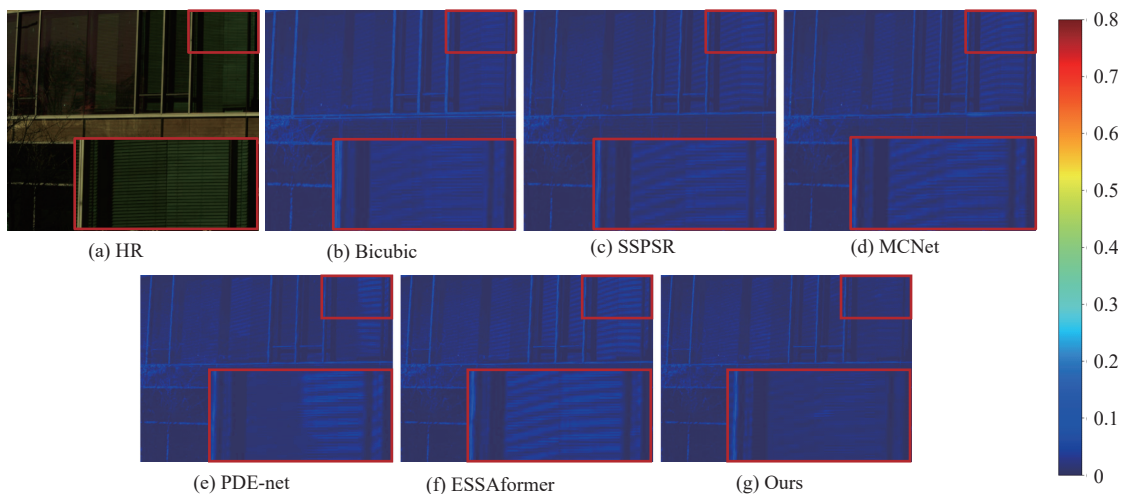


图 5 Image3 的平均误差热力图

从图 5 中可以看出, 提出的方法在光谱维度依然要比其他方法的重建质量要高, 并且从图 5 中观察到在图像的高频细节部分, 如百褶窗的纹理, 提出的方法的重建质量在空间和光谱两个维度都有明显的优势. 平均误差曲线图的原理为将重建后的图像与标签真实图像做差, 在空间维度求平均, 进而在每个光谱维度得到一个平均误差值, 将所有维度的平均误差值算出作为折线图的点, 最终得到平均误差折线图. 其中折线中的值越高说明重建后图像在该通道与真实标签的差距越大, 反之折线值越低则说明重建后图像在该通道与

对应该通道的标签越接近. 平均误差折线图如图 6 所示.

从平均误差折线图也可以看出, 提出的方法在光谱维度依然要比其他方法的重建质量要高. 这是因提出的方法有效地利用并结合了卷积的空间提取能力和 ESSA 注意力对光谱的长距离依赖规律的捕捉.

同为基于 Transformer 方法的 ESSAformer 在光谱维度也同样有着不俗的表现, 其平均误差折线仅次于提出的方法排在第 2 名. 这证明了 ESSA 有效地捕捉了高光谱图像的光谱信息规律.



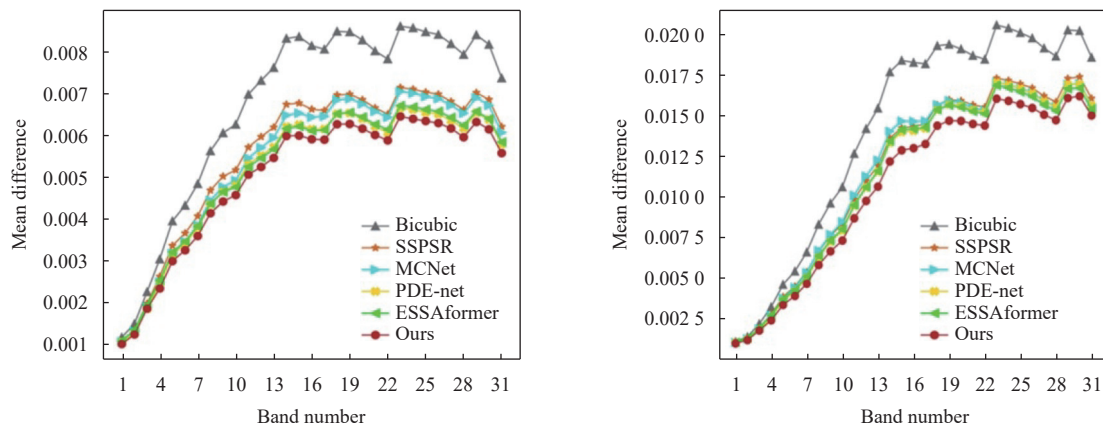


图6 Image1与Image2的平均误差折线图

### 3.5 消融实验与分析

我们进行了不同组合的消融实验来证明所提出方法每个部分的贡献与有效性. 其中主要针对的部分为

高频分支所使用的方法、主干分支所使用的方法以及改进后的多头 ESSA 与原始 ESSA 的差别. 消融实验结果如表 2 所示.

表2 在4倍放大因子下Harvard数据集上不同模块的消融实验结果

模型	平均PSNR (dB)↑	SAM↓	CC↑	RMSE↓	平均SSIM↑	ERGAS↓
完整HFEDB-Net方法	41.4032	3.1871	0.9803	0.0121	0.9531	3.3204
高频分支换成卷积神经网络	40.8274	4.2154	0.9726	0.0167	0.9357	3.6786
主干分支换成Transformer	40.8323	4.1334	0.9713	0.0154	0.9316	3.7868
高频分支与主干分支互换	40.6573	4.7265	0.9697	0.0177	0.9285	3.9618
多头ESSA替换为原始ESSA	41.0691	3.5556	0.9752	0.0137	0.9427	3.5248
去掉编码层多尺度结构	41.1582	3.6487	0.9775	0.0143	0.9486	3.4439

从表 2 中可以看到, 完整的 HFEDB-Net 方法在 4 倍放大因子 Harvard 数据集上的重建能力是最好的, 而在将基于 Transformer 的高频分支与基于卷积神经网络的主干分支进行方法互换后对重建能力的影响最大, 结合单独将高频分支换成卷积神经网络以及单独将主干分支换成 Transformer 的结果, 我们可以得出以下结论: 针对图像高频信息的处理, 从自适应能力和捕捉长距离依赖, 来看 Transformer 有一定的优势. 而对于充分提取图像空间特征的目标, 卷积神经网络却是一个更好的选择.

同样从表 2 中显示, 对于 ESSA 和编码层的多头注意力机制和多尺度结构这两个改进方向是正确且有效的, 并且多头注意力机制的作用相对而言更大, 这是因为对于高光谱图像这种高维数据, 单头注意力机制难以捕捉足够的特征和关系.

## 4 结论与展望

本文提出了一种基于高频信息强化的 Transformer 与卷积神经网络的双分支高光谱图像超分辨率网络

HFEDB-Net. 通过结合 Transformer 和卷积神经网络的优势, 双分支网络架构实现了对 HSI 高频信息和光谱-空间信息的有效重建. 本文通过使用拉普拉斯金字塔提取输入和标签的高频信息, 并构建了具有额外监督的光谱强化 Transformer 高频特征分支, 从而增强了对 HSI 空间细节的重建能力. 其次, 通过设计了多头 ESSA 注意力与多尺度编码层, 有效提升了对高频特征的特征提取能力和对上下文特征信息的利用. 在两个公开数据集的两种放大倍数下, 主观可视化和客观定量指标上均实现了最好的重建结果.

接下来可以开展的工作方向可以从光谱一致性以及高光谱空间信息的去噪方面来继续研究.

### 参考文献

- 车永飞, 叶发旺, 张明林. 高光谱数据与航空伽玛能谱数据在砂岩型铀矿勘查中的应用. 地质论评, 2023, 69(S1): 341-342. [doi: 10.16509/j.georeview.2023.s1.152]
- 邵慧, 杨丽娟, 王程, 等. 利用高光谱图像实现生菜失水率和新鲜度预测. 光电子·激光, 1-13. <http://kns.cnki.net/kcms/>

- [detail/12.1182.o4.20240325.1857.008.html](#). [2024-04-02].
- 3 Yamanaka J, Kuwashima S, Kurita T. Fast and accurate image super resolution by deep CNN with skip connection and network in network. Proceedings of the 24th International Conference on Neural Information Processing. Guangzhou: Springer, 2017. 217–225. [doi: [10.1007/978-3-319-70096-0\\_23](#)]
  - 4 Yang FZ, Yang H, Fu JL, *et al.* Learning texture Transformer network for image super-resolution. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 5790–5799. [doi: [10.1109/CVPR42600.2020.00583](#)]
  - 5 Wang XT, Yu K, Wu SX, *et al.* ESRGAN: Enhanced super-resolution generative adversarial networks. Proceedings of the 2018 European Conference on Computer Vision (ECCV) Workshops. Munich: Springer, 2018. 63–79. [doi: [10.1007/978-3-030-11021-5\\_5](#)]
  - 6 Xu YH, Hou JY, Zhu XJ, *et al.* Hyperspectral image super-resolution with ConvLSTM skip-connections. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1–16. [doi: [10.1109/TGRS.2024.3401843](#)]
  - 7 Liu JJ, Wu ZB, Xiao L, *et al.* A truncated matrix decomposition for hyperspectral image super-resolution. IEEE Transactions on Image Processing, 2020, 29: 8028–8042. [doi: [10.1109/TIP.2020.3009830](#)]
  - 8 Dian RW, Li ST, Fang LY, *et al.* Hyperspectral image super-resolution via local low-rank and sparse representations. Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium. Valencia: IEEE, 2018. 4003–4006. [doi: [10.1109/IGARSS.2018.8519213](#)]
  - 9 Patel RC, Joshi MV. Super-resolution of hyperspectral images using compressive sensing based approach. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2012, I-7: 83–88. [doi: [10.5194/isprsannals-I-7-83-2012](#)]
  - 10 Hong DF, Yokoya N, Chanussot J, *et al.* Joint and progressive subspace analysis (JPSA) with spatial-spectral manifold alignment for semisupervised hyperspectral dimensionality reduction. IEEE Transactions on Cybernetics, 2021, 51(7): 3602–3615. [doi: [10.1109/TCYB.2020.3028931](#)]
  - 11 Huang H, Shi GY, He HB, *et al.* Dimensionality reduction of hyperspectral imagery based on spatial-spectral manifold learning. IEEE Transactions on Cybernetics, 2020, 50(6): 2604–2616. [doi: [10.1109/TCYB.2019.2905793](#)]
  - 12 Jia JR, Ji LY, Zhao YC, *et al.* Hyperspectral image super-resolution with spectral-spatial network. International Journal of Remote Sensing, 2018, 39(22): 7806–7829. [doi: [10.1080/01431161.2018.1471546](#)]
  - 13 Liebel L, Körner M. Single-image super resolution for multispectral remote sensing data using convolutional neural networks. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2016, XLI-B3: 883–890. [doi: [10.5194/isprs-archives-XLI-B3-883-2016](#)]
  - 14 Dong C, Loy CC, He KM, *et al.* Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295–307. [doi: [10.1109/TPAMI.2015.2439281](#)]
  - 15 Li YS, Hu J, Zhao X, *et al.* Hyperspectral image super-resolution using deep convolutional neural network. Neurocomputing, 2017, 266: 29–41. [doi: [10.1016/j.neucom.2017.05.024](#)]
  - 16 Mei SH, Yuan X, Ji JY, *et al.* Hyperspectral image spatial super-resolution via 3D full convolutional neural network. Remote Sensing, 2017, 9(11): 1139. [doi: [10.3390/rs9111139](#)]
  - 17 Shi Z, Chen C, Xiong ZW, *et al.* Deep residual attention network for spectral image super-resolution. Proceedings of the 2018 European Conference on Computer Vision (ECCV) Workshops. Munich: Springer, 2018. 214–229.
  - 18 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
  - 19 Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. Proceedings of the 9th International Conference on Learning Representations. OpenReview.net, 2021.
  - 20 Liang JY, Cao JZ, Sun GL, *et al.* SwinIR: Image restoration using swin Transformer. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 1833–1844.
  - 21 Li BC, Li X, Lu YT, *et al.* HST: Hierarchical swin Transformer for compressed image super-resolution. Proceedings of the 2022 European Conference on Computer Vision (ECCV) Workshops. Tel Aviv: Springer, 2022. 651–668.
  - 22 Chen XY, Wang XT, Zhang WL, *et al.* HAT: Hybrid attention Transformer for image restoration. arXiv:2309.05239, 2023.
  - 23 Hu JF, Huang TZ, Deng LJ, *et al.* Fusformer: A Transformer-based fusion network for hyperspectral image

- super-resolution. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 6012305. [doi: [10.1109/LGRS.2022.3194257](https://doi.org/10.1109/LGRS.2022.3194257)]
- 24 Zhang MJ, Zhang C, Zhang QM, *et al.* ESSAformer: Efficient Transformer for hyperspectral image super-resolution. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision*. Paris: IEEE, 2023. 23016–23027.
- 25 Aggarwal HK, Majumdar A. Hyperspectral image denoising using spatio-spectral total variation. *IEEE Geoscience and Remote Sensing Letters*, 2016, 13(3): 442–446. [doi: [10.1109/LGRS.2016.2518218](https://doi.org/10.1109/LGRS.2016.2518218)]
- 26 Chakrabarti A, Zickler T. Statistics of real-world hyperspectral images. *Proceedings of the 2011 Conference on Computer Vision and Pattern Recognition*. Colorado Springs: IEEE, 2011. 193–200. [doi: [10.1109/CVPR.2011.5995660](https://doi.org/10.1109/CVPR.2011.5995660)]
- 27 Yokoya N, Iwasaki A. Airborne hyperspectral data over Chikusei. Technical Report, SAL-2016-05-27. Tokyo: The University of Tokyo, 2016.
- 28 Jiang JJ, Sun H, Liu XM, *et al.* Learning spatial-spectral prior for super-resolution of hyperspectral imagery. *IEEE Transactions on Computational Imaging*, 2020, 6: 1082–1096. [doi: [10.1109/TCI.2020.2996075](https://doi.org/10.1109/TCI.2020.2996075)]
- 29 Li Q, Wang Q, Li XL. Mixed 2D/3D convolutional network for hyperspectral image super-resolution. *Remote Sensing*, 2020, 12(10): 1660. [doi: [10.3390/rs12101660](https://doi.org/10.3390/rs12101660)]
- 30 Hou JH, Zhu ZY, Hou JH, *et al.* Deep posterior distribution-based embedding for hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, 2022, 31: 5720–5732. [doi: [10.1109/TIP.2022.3201478](https://doi.org/10.1109/TIP.2022.3201478)]

(校对责编: 张重毅)