

# 全局-局部注意力特征重用高光谱图像超分辨率网络

王思泽, 关欣, 李锐\*

天津大学微电子学院, 天津 300072

**摘要** 为了解决高光谱图像超分辨率重建中物质本征光谱表达能力不足、图像尺度变化过程中存在细节信息损失的问题,提出一种全局-局部注意力特征重用网络。首先,通过多节点特征重用增强网络的多尺度信息提取能力。其次,引入局部注意力,利用空间注意力机制聚焦重点空间信息,并通过通道注意力机制增强特异性光谱提取能力。最后,设计全局修正模块,根据原始多光谱图像空间信息丰富与高光谱图像光谱保真度高的特性进一步补偿处理过程中损失的空间与光谱维度信息,提高网络可靠性。选取 CAVE 和 Harvard 数据集进行训练与测试,并与多种先进方法进行定量与定性评估。结果表明,所提网络能够重建出更高分辨率的高光谱图像,更好完成高光谱超分辨率任务。

**关键词** 高光谱图像; 超分辨率; 注意力机制; 特征重用; 卷积神经网络

**中图分类号** TP751 **文献标志码** A

**DOI:** 10.3788/AOS230613

## 1 引言

高光谱图像不同于常见的多光谱图像,其包含几十个甚至更多的窄波段,从而反映物体的空间信息与光谱信息<sup>[1]</sup>。空间信息被用来反映被测目标的位置与外形,光谱信息则利用不同物质成分间的光谱吸收特性差异,反映目标内部结构与物质成分信息<sup>[2]</sup>。与多光谱图像相比,高光谱图像能够有效提升物体检测、识别能力,因此存在许多工业应用场景,如医学检测<sup>[3]</sup>、光学气体成像<sup>[4]</sup>、矿产勘探<sup>[5]</sup>等。

由于高光谱图像传感器接受能量有限,光谱波段的增加导致每个波段的光量子减少,降低了图像空间分辨率,从而出现大量混合像元<sup>[6]</sup>。因此,研究人员使用超分辨率技术对高光谱图像进行后处理,旨在保证光谱不失真的前提下从低分辨率图像中恢复高分辨率细节信息。单张高光谱图像的超分辨率方法中,SSPSR<sup>[7]</sup>通过渐进上采样与权重共享方法提升采样效果。同时,MCNet<sup>[8]</sup>、SFCSR<sup>[9]</sup>、CoarSR<sup>[10]</sup>等采用2D/3D卷积混合技术平衡参数与性能,并不断更新特征融合方式以获得更好的重建效果。但是,上述方法输入信息形式单一,在高比例因子放大任务中难以实现高质量重建。而基于融合的高光谱图像超分辨率能够利用多光谱图像内空间增量信息,更好地适用于图像

的高比例因子重建。Zhang等<sup>[11]</sup>提出一种使用像素感知细化的无监督递归高光谱图像超分辨率方法,利用中间重建结果进行自我监督学习。Wang等<sup>[12]</sup>将融合问题表述为一个优化问题,通过求解 Sylvester 方程,估计正则化参数,重建出最终的高分辨率图像。Zhu等<sup>[13]</sup>受到小波分解的图像融合启发,对特征图进行零均值归一化,并通过渐进方式对零均值特征进行融合训练。Wang等<sup>[14]</sup>利用线性退化模型与神经网络输出设计光谱与空间域中的深度先验正则化方法,优化重建结果。Hu等<sup>[15]</sup>结合 Transformer 突破卷积核大小限制,加强全局信息与长距离信息捕捉能力。Huang等<sup>[16]</sup>联合利用光谱正则化与物理成像模型,将融合问题转化为可微优化问题。

上述方法大多利用1D/2D卷积分别对光谱与空间特征进行提取,但是高光谱图像信息复杂度高、光谱维度提取难度大<sup>[17]</sup>,可能会造成网络对光谱信息的提取能力不足,从而导致重建效果下降。而3D卷积的使用会引入大量参数,增加网络训练压力。复杂的光谱还容易造成训练过程中的信息损失,仅采用大跨度的残差连接<sup>[18]</sup>一定程度上可以缓解该问题,但忽略了不同输入图像的信息差异,难以有针对性地利用原始输入信息补偿网络。同时,许多网络结构信息重用方式过于复杂,造成信息冗余,增大模型训练压力。

收稿日期: 2023-03-02; 修回日期: 2023-04-27; 录用日期: 2023-05-31; 网络首发日期: 2023-06-28

基金项目: 国家自然科学基金(61471263, 61872267, U21B2024)、天津市自然科学基金(16JCZDJC31100)、天津大学创新基金(2021XZC-0024)

通信作者: \*liaiqng@tju.edu.cn

为了解决上述问题,本文提出一种简单有效的全局-局部注意力特征重用神经网络。通过简单特征重用连接减小模型训练压力,利用不同粒度的细节信息实现高光谱图像的高质量重建。采用局部注意力块,利用通道空间注意力机制局部聚焦重点空间特征,并补充强化光谱信息提取能力。最后设置全局修正模块,有针对性地利用初始输入的多光谱图像与高光谱图像分别对空间与光谱维度进行信息更新,减少提取过程中的信息损失,进一步提高网络可靠性。

## 2 网络结构

### 2.1 整体架构

假设高分辨率高光谱图像  $X \in \mathbb{R}^{HW \times L}$ , 则多光谱图像  $Y \in \mathbb{R}^{HW \times l}$  与低分辨率高光谱图像  $Z \in \mathbb{R}^{hw \times L}$  的退化模型为

$$\begin{cases} Y = XR + N_1 \\ Z = DBX + N_2 \end{cases}, \quad (1)$$

式中:  $R \in \mathbb{R}^{L \times l}$ 、 $D \in \mathbb{R}^{hw \times HW}$ 、 $B \in \mathbb{R}^{HW \times HW}$  分别为光谱响应矩阵、下采样矩阵与模糊矩阵;  $N_1$ 、 $N_2$  为噪声;  $H(h)$ 、 $W(w)$ 、 $L(l)$  分别为图像高度、图像宽度与光谱波段总数,且  $H = rh$ 、 $W = rw$ 、 $l \ll L$ 、 $r$  为放大倍数。

所提全局-局部注意力特征重用网络 (LGAR-Net) 为了保证网络使用的广泛性与便捷性,输入为最

常见的多光谱图像,即 RGB 图像  $I_{\text{MSI}} \in \mathbb{R}^{3 \times W \times H}$  与经过双三次插值放大后的低分辨率高光谱图像  $I_{\text{LHSI}} \in \mathbb{R}^{L \times W \times H}$ , 网络结构如图 1 所示。高光谱图像数据复杂,包含更多的细节信息,直接将完整图片作为输入可能会导致细节信息关注不足,降低重建效果。而渐进式网络<sup>[19-21]</sup>在图像超分辨率领域已经被证明是一种强有力的方式,因此 LGAR-Net 设计 3 个渐进阶段,并在不同的网络阶段以光谱跨步衰减的方式嵌入  $I_{\text{LHSI}}$ , 从第 1 阶段到第 3 阶段,抽取光谱波段数量分别为完整波段数量的 1/4、1/2 和 1。先通过较少波段进行浅层构建,后续加入更多波段对细节不断微调,使网络从高级配置开始逐步细化,实现更精准的重建。

具体而言, LGAR-Net 中每个渐进阶段包含一个特征重用模块 (FRM) 和一个局部注意力块 (LA)。通过 FRM 进行深度特征提取,随后通过 LA 进一步探究光谱和与空间信息的关系来加强网络表征能力。随后将渐进阶段结果输入全局修正模块 (GC) 中,利用原始输入的信息相关性对结果进行修正。最后在加入原始输入的长跳转连接后经过两层卷积进行图像重建,获得高分辨率高光谱图像  $I_{\text{SHSI}} \in \mathbb{R}^{L \times W \times H}$ , LGAR-Net 通过简单的渐进式网络结构,利用全局-局部注意力机制融合两种输入图像,从而重建出高质量的高光谱图像。

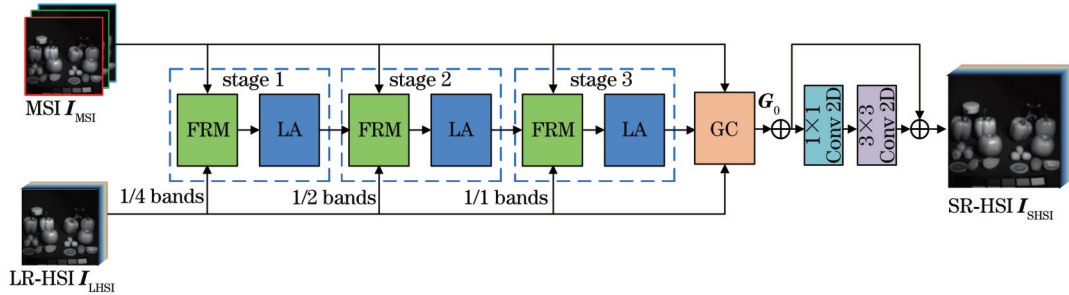


图 1 全局-局部特征重用网络架构

Fig. 1 Global-local feature reuse network architecture

### 2.2 特征重用模块

高光谱图像需要对光谱维度和空间维度同时进行提取。最直观的 3D 卷积虽然有所成效但是模型负担较大,因此使用  $1 \times 1$  卷积对所有的光谱波段  $s_1, s_2, \dots, s_l$  通过线性组合权重  $\omega_i = (\omega_{i,1}, \omega_{i,2}, \dots, \omega_{i,l})$  得到重建的光谱波段  $x'_i$ :

$$x'_i = \omega_{i,1}s_1 + \omega_{i,2}s_2 + \dots + \omega_{i,l}s_l. \quad (2)$$

通过不断更新可学习参数权重向量  $\omega_i$ , 将相关信息嵌入重建光谱波段中,从而更好地利用相邻波段之间的光谱增量信息。并使用  $3 \times 3$  卷积提取空间特征,通过不断训练强化空间分辨率。设计的 FRM 如图 2 所示,其中,包含结构相似的提取模块与融合模块,第  $q$  个提取模块输出  $B_q$  可表示为

$$B_q = \begin{cases} \sigma_r \left\{ \beta \left\{ C_{3 \times 3} \left[ C_{1 \times 1} (B_{q-1}) \right] \right\} \right\}, & q \neq 0 \\ \text{CAT} (I_{\text{LHSI}}, M, F_x), & q = 0 \end{cases}, \quad (3)$$

式中:  $C_{3 \times 3}(\cdot)$  表示  $3 \times 3$  的 2D 卷积;  $C_{1 \times 1}(\cdot)$  表示  $1 \times 1$  的 2D 卷积;  $\sigma_r(\cdot)$  表示 ReLU 激活函数;  $\beta(\cdot)$  表示批归一化;  $\text{CAT}(\cdot)$  表示级联;  $M$  表示多光谱图像原始输入  $I_{\text{MSI}}$  通道整形后的结果;  $F_x$  为上一渐进阶段的输出。

为了充分保留数据特征的浅层信息与深层信息,需要综合利用多个不断叠加的提取模块的结果。许多网络采用 DenseNet<sup>[22]</sup> 等结构进行特征重用,但是密集连接带来的输入通道线性增长会导致较高的模型参数与能耗,因此采用简单的聚合连接将  $N$  个提取模块在通道维度级联,再输入融合模块进行特征融合与整形,级联结果  $S_0$  可表示为

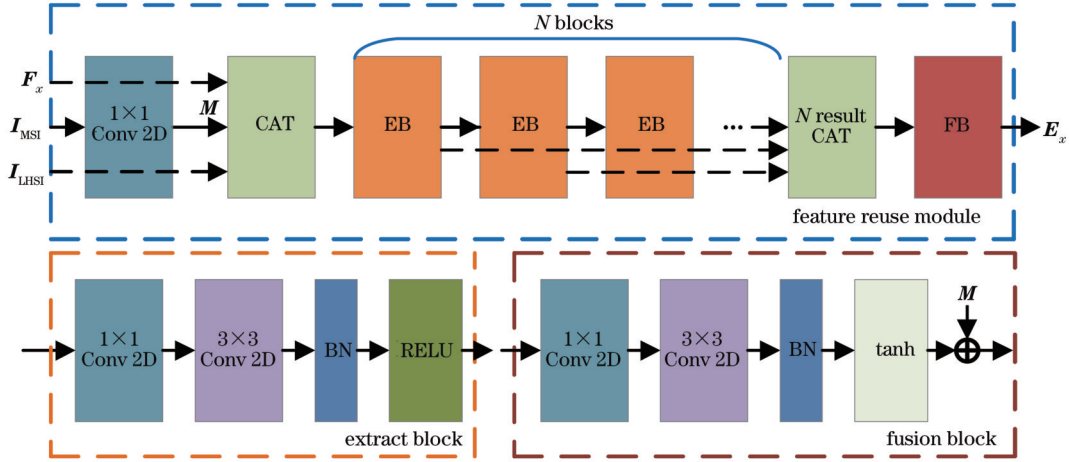


图 2 特征重用模块结构

Fig. 2 Feature reuse module architecture

$$S_0 = \text{CAT}(B_N, B_{N-1}, \dots, B_1). \quad (4)$$

最后通过与提取模块相类似的融合模块,对多节点重用特征进行融合,利用多粒度的信息提升图像重建质量。如果用  $\sigma(\cdot)$  表示 tanh 激活函数,则 FRM 模块输出  $E_x$  可表示为

$$E_x = \sigma\left\{\beta\left\{C_{3 \times 3}\left[C_{1 \times 1}(S_0)\right]\right\}\right\} + M. \quad (5)$$

### 2.3 局部注意力块

高光谱图像的相邻光谱波段之间存在强相关性的特点已经被广泛用于高光谱图像的重建和分析<sup>[23]</sup>。然而高光谱图像结构更加复杂,仅仅使用 1D 卷积很难表达这种强相关性,会导致光谱维度提取能力不足。受到光谱空间注意力机制<sup>[24]</sup>启发,引入通道注意力机制自适应重新分配光谱维度的特征,以更好学习光谱间的增量信息。具体结构如图 3 中 channel attention 模块所示,其输出权重矩阵  $\omega_c$  的表达式为

$$\omega_c = \sigma\left\{C_{1 \times 1}\left\{\sigma\left(C_{1 \times 1}\left[P_{\text{avg}}^s(E_x)\right]\right)\right\}\right\} + C_{1 \times 1}\left\{\sigma\left(C_{1 \times 1}\left[P_{\text{avg}}^s(E_x)\right]\right)\right\}, \quad (6)$$

式中:  $\sigma(\cdot)$  表示 Sigmoid 函数;  $P_{\text{avg}}^s(\cdot)$  与  $P_{\text{max}}^s(\cdot)$  分别表示空间维度的平均池化与最大池化。先通过池化操作获得全局内容的嵌入向量,并减少通道层与恢复通道层增加网络非线性能力,最后利用 Sigmoid 函数学习光谱之间相互作用,得到通道的变换系数矩阵。

同时,为了进一步提高模型特征提取能力,引入空间注意力机制聚焦空间信息,实现更好的重建效果。具体结构如图 3 中 spatial attention 模块所示,其输出权重表达式  $\omega_s$  为

$$\omega_s = \sigma\left\{C_{7 \times 7}\left\{\text{CAT}\left[P_{\text{avg}}^c(\omega_c E_x), P_{\text{max}}^c(\omega_c E_x)\right]\right\}\right\}. \quad (7)$$

利用  $P_{\text{avg}}^c(\cdot)$  与  $P_{\text{max}}^c(\cdot)$  所代表的通道平均池化与最大池化,将多个通道压缩成单通道特征图,级联两种方法的结果,并利用卷积核大小为  $7 \times 7$  的 2D 卷积提取空间关键信息,得到空间变换系数矩阵。

整体 LA 模块如图 3 所示,利用上述两种注意力对输入特征图像进行光谱与空间维度的权重重新分配,以获取更加接近真实的特征信息。模块输出  $F_{x+1}$  可以描述为

$$F_{x+1} = \omega_s(\omega_c E_x). \quad (8)$$

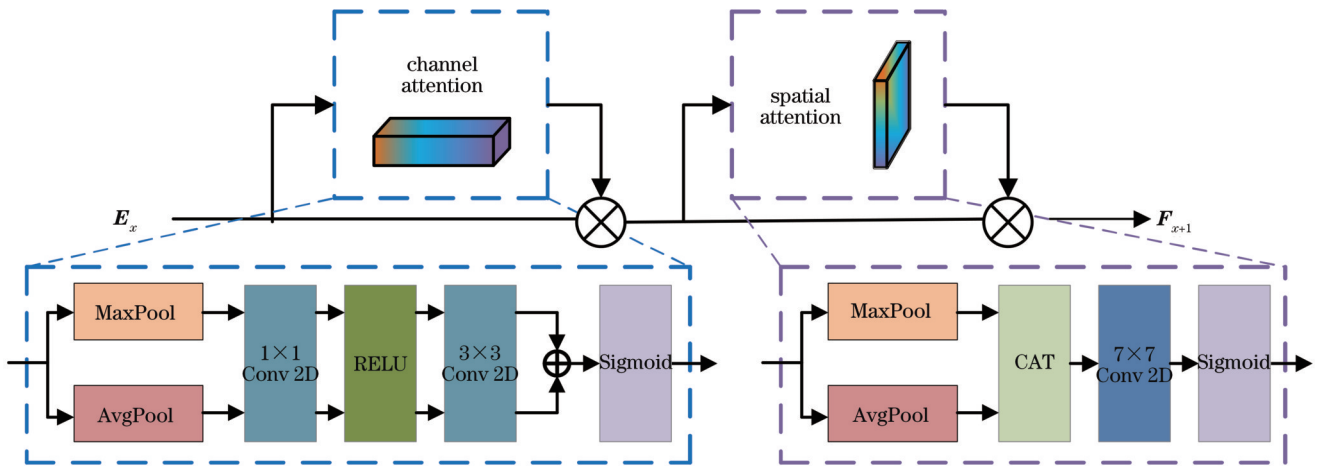


图 3 局部注意力块结构

Fig. 3 Local attention module architecture

### 2.4 全局修正模块

事实上,在基于融合的高光谱图像超分辨率方法中,大部分的光谱信息来自原始输入  $I_{LHSI}$ ,但是其通过 3 次渐进阶段逐步输入,不可避免地会出现信息损失。而大部分的空间信息来自原始输入  $I_{MSI}$ ,但在训练中其空间信息需要被平均到与  $I_{LHSI}$  相同数量的光谱中,会造成一定程度的失真。同时,随着网络深度的叠加,光谱与空间都会不可避免地再次出现一定程度的信息损失。此前工作大多采用全局跳转连接保证网络训练的可靠性,但是这样的方法只是简单的线性相

加,无法专注于高光谱图像本身的结构特点。因此提出了全局修正模块,具体结构如图 4 所示,该模块利用全局注意力机制对两个原始输入进行权重学习,有针对性地利用通道注意力机制聚焦低分辨率高光谱图像中原始光谱信息,提升光谱信息可靠性,同时利用空间注意力机制聚焦包含更多空间特征的多光谱图像,以保证空间信息的完整性。该模块通过不断更新权重对渐进阶段的训练结果进行补偿优化,解决网络退化问题的同时进一步加强光谱与空间维度的信息提取能力。

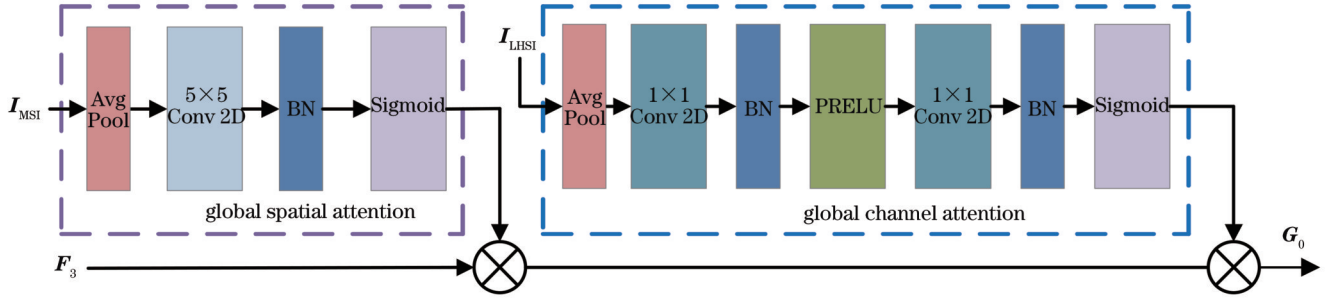


图 4 全局修正模块结构

Fig. 4 Global correction module architecture

由于原始输入结构比较简单且模块主要起修正作用,因此选择较为简单的注意力机制结构以降低网络参数与训练压力。训练权重  $\omega_c^g$ 、 $\omega_s^g$  可分别表示为

$$\omega_c^g = \sigma \left\{ \beta \left\{ C_{1 \times 1} \left\{ \sigma_p \left\{ C_{1 \times 1} \left[ P_{avg}^s(I_{LHSI}) \right] \right\} \right\} \right\} \right\}, \quad (9)$$

$$\omega_s^g = \sigma \left\{ \beta \left\{ C_{5 \times 5} \left[ P_{avg}^c(I_{MSI}) \right] \right\} \right\}, \quad (10)$$

式中:  $\sigma_p(\cdot)$  为 PRELU 激活函数。整个模块的输出  $G_0$  表示为

$$G_0 = \omega_c^g \left( \omega_s^g F_3 \right). \quad (11)$$

据经过退化模型生成模型所需要的多光谱图像与低分辨率高光谱图像输入。经过前馈网络训练后,计算损失函数,当取得更小的损失值时,对模型参数进行保存,并采用反馈网络不断更新模型参数直到最后一次迭代训练。而在测试阶段,测试数据通过预处理,得到高、低分辨率高光谱图像,并将低分辨率图像输入已经加载最优模型参数的网络中,输出重建图像。最后对重建图像与高分辨率高光谱图像进行误差计算,分析网络性能。

## 3 实验分析与讨论

### 3.1 实验准备

网络技术流程图如图 5 所示,在训练阶段,训练数

#### 3.1.1 数据集

CAVE 数据集<sup>[25]</sup>:该数据集是使用冷却式 CCD 摄像机获得的多种真实物体和材料的数据集,数据集光谱范围从 400~700 nm,间隔为 10 nm。该数据集包含

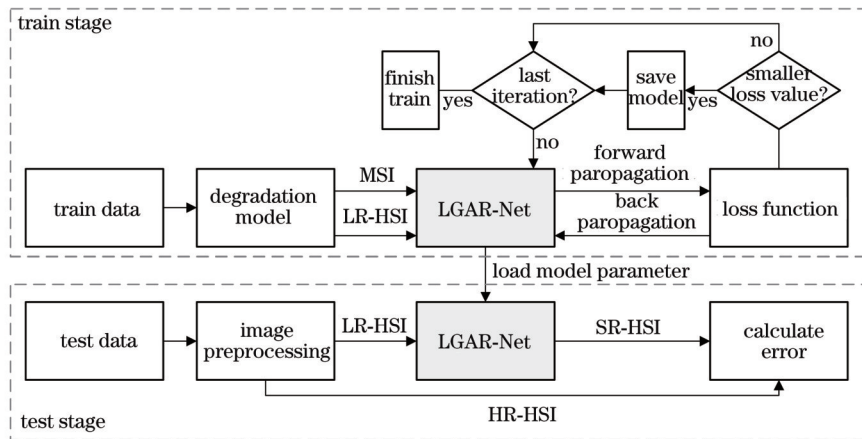


图 5 技术流程图

Fig. 5 Technical flowchart

32 个场景,分为 5 个部分:材料、皮肤和头发、真假、食物与饮料、颜料。该数据集所有高光谱图像大小为  $512 \times 512 \times 31$ , 每个波段存储为 16-bit 灰度 PNG 图像。

Harvard 数据集<sup>[26]</sup>:该数据集由 Nuance FX 相机在 400~700 nm 的波长范围内获得。该数据集由日光照明下 77 张室内或室外的真实场景的高光谱图像组成。高光谱图像大小为  $1040 \times 1392 \times 31$ 。不同于 CAVE 数据集,该数据存储在 .mat 文件。

### 3.1.2 数据处理

由于难以获得成对的高、低分辨率高光谱图像作为数据集,且数据样本较少,因此需要采用数据增强技术进行处理。对 CAVE 数据集内每张图像进行随机裁剪,获得 400 个大小为  $128 \times 128 \times 31$  的图像补丁,对于 Harvard 数据集,选择左上角  $1024 \times 1024$  区域作为待测数据集后,进行同样操作。随后对每一个图像补丁使用一个大小为  $3 \times 3$ 、标准差为 0.5 的高斯卷积核进行空间模糊,并通过下采样,获得空间大小为  $32 \times 32$  的低空间分辨率高光谱图像补丁。对于多光谱图像,采用尼康 D700 相机光谱响应矩阵  $P$  模拟光谱退化,通过对高空间分辨率高光谱图像光谱下采样获得输入的 RGB 图像。对于 CAVE 数据集,选取其中 20 张作为训练集,12 张作为测试集。而对于 Harvard 数据集,选取其中 50 张作为训练集,27 张作为测试集。

### 3.1.3 实验环境与参数设置

实验部分的具体环境为:Ubuntu 16.04 的操作系统, Intel Core i9-9900X 的 CPU 和 NVIDIA RTX2080 Ti 的 GPU。此外,所提 LGAR-NET 模型使用 PyTorch 深度学习框架搭建而成,并选取  $L_1$  函数作为损失函数。

在本实验中, LGAR-Net 模型使用 Adam 优化器来更新神经网络权重,其具体参数设置为  $\beta_1 = 0.9$ 、 $\beta_2 = 0.999$ 。优化器的初始学习率设置为 0.0002。总轮数设置为 300,批大小(batch size)设置为 4。

### 3.1.4 评价指标

为了评估各种方法的性能,4 个广泛使用的定量图像指标被采用,包括平均峰值信噪比(PSNR)、平均结构相似性(SSIM)、光谱角度制图(SAM)和相对无量纲绝对误差(ERGA)。对于高光谱图像超分辨率任

务,  $I_{SR}$  表示重建的高分辨率高光谱图像,  $I_{HR}$  表示真实的高分辨率高光谱图像。PSNR 和 SSIM 取得的值越大,说明图像重建效果越好; SAM 和 ERGA 取得的值越小,说明重建图像越接近真实图像。

$$R_{PSNR} = \frac{1}{L} \sum_{l=1}^L 10 \times \lg \left\{ \frac{X_{\max, l}^2}{\frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W [I_{SR}(h, w, l) - I_{HR}(h, w, l)]^2} \right\}, \quad (12)$$

$$R_{SSIM}(I_{SR}, I_{HR}) = \frac{1}{L} \sum_{l=1}^L \frac{(2\mu_{SR}\mu_{HR} + c_1)(2\sigma_{SR,HR} + c_2)}{(\mu_{SR}^2 + \mu_{HR}^2 + c_1)(\sigma_{SR}^2 + \sigma_{HR}^2 + c_2)}, \quad (13)$$

$$R_{SAM} = \frac{1}{WH} \sum_{w=1}^W \sum_{h=1}^H \arccos \left( \frac{\langle I_{SR}(h, w), I_{HR}(h, w) \rangle}{\|I_{SR}(h, w)\|_2 \|I_{HR}(h, w)\|_2} \right), \quad (14)$$

$$R_{ERGA} = \frac{100}{r} \sqrt{\frac{1}{L} \sum_{l=1}^L \frac{X_{\max, l}^2}{\mu_{SR}^2 \sum_{h=1}^H \sum_{w=1}^W [I_{SR}(h, w, l) - I_{HR}(h, w, l)]^2}}, \quad (15)$$

式中:  $H$  和  $W$  表示图像的高度和宽度;  $L$  表示图像的波段数;  $r$  为放大倍数;  $X_{\max, l}$  表示第  $l$  个波段的像元最大值;  $\mu_{SR}$  和  $\mu_{HR}$  表示图像  $I_{SR}$  和  $I_{HR}$  的平均值;  $\sigma_{SR,HR}$  表示图像  $I_{SR}$  和  $I_{HR}$  的协方差;  $\sigma_{SR}^2$  和  $\sigma_{HR}^2$  表示图像  $I_{SR}$  和  $I_{HR}$  的方差;  $c_1$  和  $c_2$  为常数;  $\langle \cdot, \cdot \rangle$  表示点乘,  $\|\cdot\|_2$  表示二阶范数。

## 3.2 实验分析

### 3.2.1 提取模块数量

对于 LGAR-Net, 特征重用模块中的提取模块的数量对网络结构的设计至关重要。因此,为了评估  $N$  对网络性能的影响,确定最佳的数值,分别设定  $N$  为 6~10,对网络进行训练。实验结果如表 1 所示,当  $N$  小于 8 时,特征提取力度不足,难以达到理想指标。而  $N$  大于 8 时,各项指标数值相当,因此为了平衡参数与性能, LGAR-Net 中的特征提取模块数量  $N$  设置为 8。

表 1 提取模块  $N$  的数量对网络性能指标的影响

Table 1 Influence of the number of module  $N$  on the network performance index

Index	6	7	8	9	10
PSNR $\uparrow$ /dB	49.933	50.665	51.244	50.322	50.285
SSIM $\uparrow$	0.9954	0.9963	0.9964	0.9960	0.9960
SAM $\downarrow$	2.027	1.940	1.703	1.751	1.765
ERGA $\downarrow$	0.512	0.409	0.392	0.455	0.438

### 3.2.2 消融实验分析

基本网络通过 1D 卷积进行光谱特征提取, 2D 卷积进行空间特征提取。LGAR-NET 模型主要通过特征重用设计、局部注意力块和全局修正模块改善基本网络性能, 其消融实验同样在比例因子  $\times 4$  的 CAVE 数据集上进行, 实验结果如表 2 所示。

从表 2 可以看出, 基本网络的性能最差, 表明基本

表 2 各部分消融研究

Table 2 Ablation study about the components

Component	Different Combinations				
FRM	×	√	√	√	√
LA	×	×	√	×	√
GC	×	×	×	√	√
PSNR ↑ /dB	49.865	50.387	50.571	50.435	51.244
SSIM ↑	0.9952	0.9956	0.9960	0.9960	0.9964
SAM ↓	2.007	1.925	1.795	1.728	1.703
ERGA ↓	0.595	0.547	0.485	0.477	0.392

表 3 CAVE 数据集不同比例因子对比算法结果

Table 3 Comparison of algorithm results with different scale factors on CAVE dataset

Scale	Metric	CSTF	UAL	TSFN	PZNet	Fusformer	DHIF	LGAR-Net
$\times 4$	PSNR /dB	42.300	43.964	48.853	50.288	49.106	50.169	<b>51.244</b>
	SSIM	0.9644	0.9920	0.9948	0.9950	0.9952	0.9952	<b>0.9964</b>
	SAM	7.716	3.042	1.869	1.982	1.997	1.833	<b>1.703</b>
	ERGA	2.093	1.629	0.461	0.530	0.521	0.470	<b>0.392</b>
$\times 8$	PSNR /dB	41.717	42.585	49.266	49.654	49.108	49.733	<b>50.162</b>
	SSIM	0.9665	0.9910	0.9943	0.9945	0.9944	<b>0.9947</b>	<b>0.9947</b>
	SAM	7.502	3.899	2.311	2.189	2.269	2.108	<b>1.965</b>
	ERGA	1.170	0.474	0.386	0.319	0.378	0.335	<b>0.307</b>
$\times 16$	PSNR /dB	39.824	40.105	46.533	46.994	46.322	46.865	<b>47.128</b>
	SSIM	0.9633	0.9886	0.9922	0.9920	0.9920	0.9922	<b>0.9930</b>
	SAM	7.329	4.257	2.921	2.881	2.961	2.848	<b>2.768</b>
	ERGA	0.892	0.381	0.311	0.287	0.324	0.265	<b>0.246</b>

LGAR-Net 对比其他网络, 4 项指标均有明显改善, 尤其是 SAM, 比次优结果 DHIF 降低 0.13, 证实了其强大的光谱表征能力。同时, PSNR、SSIM 分别提升 1 dB、0.0012, 反映其优秀的空间重建能力。这些提升主要来自于局部注意力模块对光谱维度的进一步提取, 提升对光谱间映射关系的学习。同时利用原始输入对全局结果进行修正, 对空间与光谱维度的表征能力都有所增强。同时表 3 中也给出了比例因子分别为  $\times 8$ 、 $\times 16$  时的比对结果, 相对于小因子结果, 各项指标性能均有所下降, 但是 LGAR-NET 仍然取得了最优值。比对结果进一步说明了, LGAR-NET 在不同比例因子下均能发挥良好的作用。

为了进一步评估 LGAR-Net 方法的网络泛化性, 在 Harvard 数据集上也在比例因子为  $\times 4$ 、 $\times 8$ 、 $\times 16$  的任务中进行比对, 具体如表 4 所示。LGAR-Net 仍然

网络没有充分利用空间光谱特征, 特征提取力度不足。加入简单的特征重用连接后, 通过重用深层提取与浅层提取的特征信息, 性能有所提升。进一步引入局部注意力块时, SAM 指标显著下降, 说明该模块可有效提升光谱间映射关系。补充光谱维度特征提取后, 其余指标均有所优化, 说明空间注意力机制也在一定程度上加强网络空间表征能力。在加入重用连接的同时加入全局修正模块, 各项指标也均有所提升, 说明其可提高信息流通度, 并有效利用原始输入对网络训练结果进行修正。最后同时加入 3 个组件时的最优实验结果表明, 特征重用连接、局部光谱注意块、全局修正模块均是 LGAR-Net 模型的关键组成部分。

### 3.2.3 定量实验评估

表 3 给出了比例因子为  $\times 4$  时所有对比算法在 PSNR、SSIM、SAM 和 ERGA 的具体结果。从结果可以看出, CSTF、UAL 效果均不理想, 存在较大偏差, 其他 4 种对比算法效果相当, 但各有侧重。

取得最优结果, 表明该网络具有良好的泛化能力。Harvard 数据集相较于 CAVE 数据集图片数量更多、包含信息更丰富, 因此挑战性更大。而 LGAR-Net 仍能保持较大优越性, 说明 LGAR-Net 各项核心部件的必要性与有效性。

### 3.2.4 定性实验评估

为了进一步评估 LGAR-Net 模型的性能, 直观对比重建图像与真实图像之间的差异, 图 6 给出了基于 CAVE 数据集中“oil\_painting”图像的第 10 个波段各网络结果与真值的绝对误差映射图。差异越小, 图像颜色越接近黑色, 差异越大, 颜色越接近白色。从图中可以看出: CSTF 网络结果轮廓最清晰, 效果最差; UAL 有较大改善, 但仍存在清晰边缘; TSFN 进一步淡化轮廓, 但原图中蓝色背景处, 仍有较多纹理; PZNet 效果与其他不同, 内部存在很多凸起状纹理,

表 4 Harvard 数据集不同比例因子对比算法结果

Table 4 Comparison of algorithm results with different scale factors on Harvard dataset

Scale	Metric	CSTF	UAL	TSFN	PZNet	Fusformer	DHIF	LGAR-Net
×4	PSNR /dB	42.362	44.853	48.993	49.382	48.856	49.263	49.962
	SSIM	0.9722	0.9865	0.9874	0.9873	0.9873	0.9873	0.9882
	SAM	8.257	4.762	3.488	3.250	3.422	3.488	3.085
	ERGA	5.693	1.855	1.732	1.640	1.662	1.539	1.218
×8	PSNR /dB	41.233	43.368	48.661	48.864	48.203	48.835	49.210
	SSIM	0.9706	0.9837	0.9850	0.9849	0.9850	0.9850	0.9862
	SAM	9.214	5.011	3.763	3.901	3.955	3.858	3.317
	ERGA	3.847	1.025	0.839	0.807	0.822	0.796	0.752
×16	PSNR /dB	39.822	42.563	47.299	47.262	46.958	47.235	47.386
	SSIM	0.9521	0.9788	0.9803	0.9801	0.9803	0.9803	0.9812
	SAM	9.621	5.812	4.344	4.219	4.425	4.268	4.105
	ERGA	2.102	0.842	0.662	0.678	0.666	0.655	0.602

且对大边缘重建效果较差;Fusformer与DHIF结果相类似,但Fusformer中出现更多的局部高亮,且整体白色更亮;而所提LGAR-Net相较于其他网络,场景边

缘较浅甚至没有边缘。由此可见,LGAR-Net能够有效地融合两种不同的信息,并对空间信息进行高效探索。

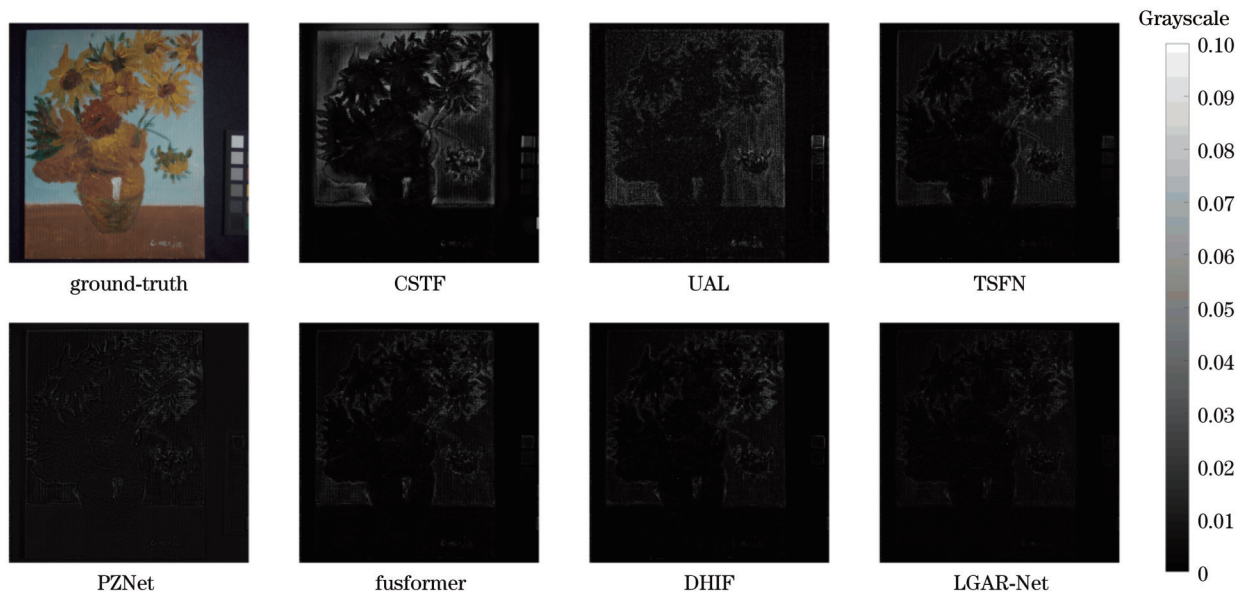


图 6 CAVE 数据集的 oil\_painting 图像的绝对误差图比较

Fig. 6 Comparison of absolute error maps of "oil\_painting" images from the CAVE dataset

除了可视化方法分析像元绝对误差,需要对像元的光谱变化进行对比,帮助在光谱维度下进行定性分析。为了更好地适用于高光谱图像能够反映物质信息的特性,在CAVE数据集“paint\_ms”图片上选择4个代表不同物质的典型像元位置,通过对应位置的光谱曲线反映重建图像是否存在较大范围的光谱失真,如图7所示。其中,横坐标为波段,纵坐标为反射率,图内分别对第10~20波段、第20~30波段结果进行放大,方便进一步观察与对比。在大多数情况下,LGAR-Net的结果更加接近真实图像,与定量评估中SAM结果保持一致,表明其能够有效地避免光谱

失真。

## 4 结 论

目前现有的基于融合方式的高光谱图像超分辨率网络虽然都兼顾了空间与光谱信息的提取,但都更注重空间信息,缺少对光谱信息保真的重视,且无法有针对性地补充网络中的损失信息。因此,提出一种全局-局部注意力特征重用网络——LGAR-Net,使用多重注意力与简单特征重用实现更好的高光谱图像超分辨率效果。网络针对高光谱图像特点,加强光谱特征提取能力,并有针对性地进行信息补偿。在不同放大因

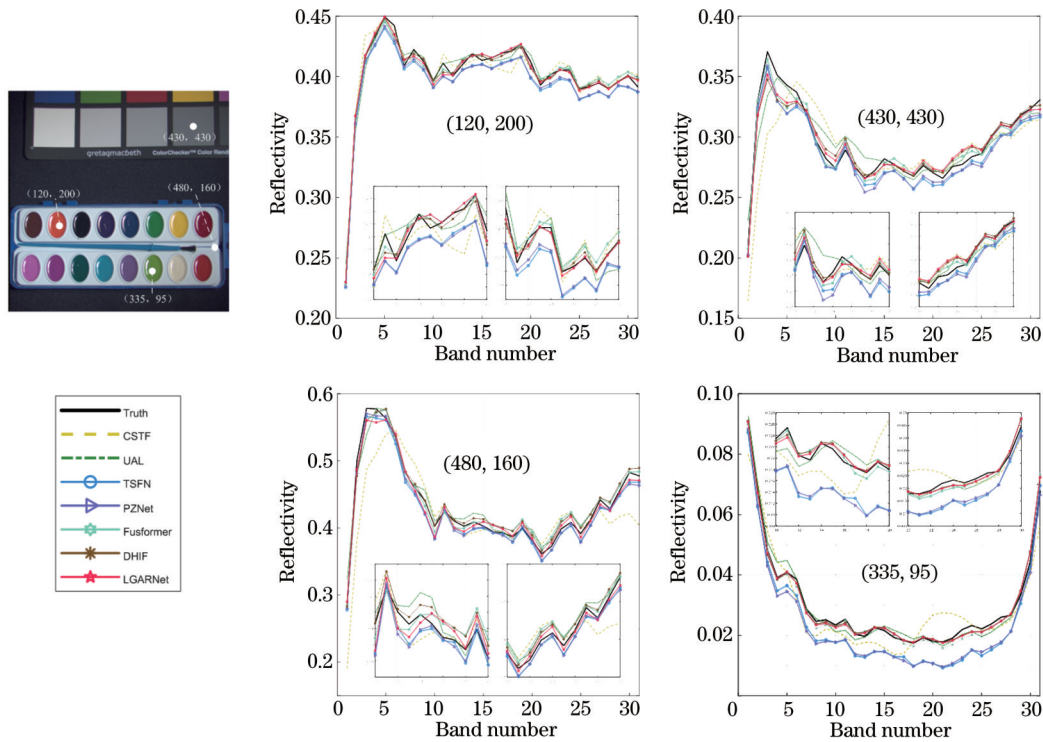


图 7 CAVE 数据集 "paints\_ms" 图片上不同位置像元光谱比较

Fig. 7 Comparison of pixel spectra at different locations in CAVE data set "paints\_ms"

子的任务下,对比最新方法,LGAR-Net均取得了最优指标。所提网络具有复杂度低、超分辨率精度高等优点。目前输入网络的低分辨率图像与多光谱图像大多是由高分辨率的高光谱图像退化获得的人工图像,真实场景中的噪声、运动模糊等因素造成的图像退化过程更加复杂,需要进一步探讨与研究。

## 参 考 文 献

- [1] 刘嘉敏, 郑超, 张丽梅, 等. 基于图像重构特征融合的高光谱图像分类方法[J]. 中国激光, 2021, 48(9): 0910001.  
Liu J M, Zheng C, Zhang L M, et al. Hyperspectral image classification method based on image reconstruction feature fusion[J]. Chinese Journal of Lasers, 2021, 48(9): 0910001.
- [2] Wang Q, Lin J Z, Yuan Y. Salient band selection for hyperspectral image classification via manifold ranking[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(6): 1279-1289.
- [3] Calin M A, Parasca S V, Savastru D, et al. Hyperspectral imaging in the medical field: present and future[J]. Applied Spectroscopy Reviews, 2014, 49(6): 435-447.
- [4] Liu K, Su H B, Li X K. Estimating high-resolution urban surface temperature using a hyperspectral thermal mixing (HTM) approach[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2016, 9(2): 804-815.
- [5] Thai B, Healey G. Invariant subpixel target identification in hyperspectral imagery[J]. Proceedings of SPIE, 1999, 3717: 14-24.
- [6] 沈柄志, 聂若梅, 蒋海鹏, 等. 基于混合卷积网络的高分辨率高光谱图像分类研究[J]. 激光与光电子学进展, 2022, 59(24): 2428006.  
Shen B Z, Nie R M, Jiang H P, et al. Research on high resolution hyperspectral image classification based on hybrid convolution network[J]. Laser & Optoelectronics Progress, 2022, 59(24): 2428006.
- [7] Jiang J J, Sun H, Liu X M, et al. Learning spatial-spectral prior for super-resolution of hyperspectral imagery[J]. IEEE Transactions on Computational Imaging, 2020, 6: 1082-1096.
- [8] Li Q A, Wang Q, Li X L. Mixed 2D/3D convolutional network for hyperspectral image super-resolution[J]. Remote Sensing, 2020, 12(10): 1660.
- [9] Wang Q, Li Q, Li X L. Hyperspectral image super-resolution using spectrum and feature context[J]. IEEE Transactions on Industrial Electronics, 2021, 68(11): 11276-11285.
- [10] Li Q, Wang Q, Li X L. Exploring the relationship between 2D/3D convolution for hyperspectral image super-resolution[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(10): 8693-8703.
- [11] Zhang L, Nie J T, Wei W, et al. Unsupervised adaptation learning for hyperspectral imagery super-resolution[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 3070-3079.
- [12] Wang X H, Chen J, Wei Q, et al. Hyperspectral image super-resolution via deep prior regularization with parameter estimation [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(4): 1708-1723.
- [13] Zhu Z Y, Hou J H, Chen J, et al. Hyperspectral image super-resolution via deep progressive zero-centric residual learning[J]. IEEE Transactions on Image Processing, 2021, 30: 1423-1438.
- [14] Wang X H, Chen J, Richard C. Hyperspectral image super-resolution with deep priors and degradation model inversion[C]// ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 23-27, 2022, Singapore, Singapore. New York: IEEE Press, 2022: 2814-2818.
- [15] Hu J F, Huang T Z, Deng L J, et al. Fusformer: a transformer-based fusion network for hyperspectral image super-resolution [J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 6012305.



- [16] Huang T, Dong W S, Wu J J, et al. Deep hyperspectral image fusion network with iterative spatio-spectral regularization[J]. IEEE Transactions on Computational Imaging, 2022, 8: 201-214.
- [17] Hong D F, Gao L R, Yao J, et al. Graph convolutional networks for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(7): 5966-5978.
- [18] 张祥东, 王腾军, 朱劭俊, 等. 基于扩张卷积注意力神经网络的高光谱图像分类[J]. 光学学报, 2021, 41(3): 0310001. Zhang X D, Wang T J, Zhu S J, et al. Hyperspectral image classification based on dilated convolutional attention neural network[J]. Acta Optica Sinica, 2021, 41(3): 0310001.
- [19] Ahn N, Kang B, Sohn K A. Image super-resolution via progressive cascading residual network[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 18-22, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 904-9048.
- [20] Lin D Y, Xu G L, Xu W J, et al. SCRSR: an efficient recursive convolutional neural network for fast and accurate image super-resolution[J]. Neurocomputing, 2020, 398: 399-407.
- [21] Xin J W, Li J, Jiang X R, et al. Wavelet-based dual recursive network for image super-resolution[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(2): 707-720.
- [22] Wang Z R, Liu G L, Tian G H. A parameter efficient human pose estimation method based on densely connected convolutional module[J]. IEEE Access, 2018, 6: 58056-58063.
- [23] Qi J H, Wan P C, Gong Z Q, et al. A self-improving framework for joint depth estimation and underwater target detection from hyperspectral imagery[J]. Remote Sensing, 2021, 13(9): 1721.
- [24] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [25] Yasuma F, Mitsunaga T, Iso D, et al. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum[J]. IEEE Transactions on Image Processing, 2010, 19(9): 2241-2253.
- [26] Chakrabarti A, Zickler T. Statistics of real-world hyperspectral images[C]//Proceeding of 2011 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2011, Colorado Springs, CO, USA. New York: IEEE Press, 2011: 193-200.

## Hyperspectral Image Super-Resolution Network of Local-Global Attention Feature Reuse

Wang Size, Guan Xin, Li Qiang\*

*School of Microelectronics, Tianjin University, Tianjin 300072, China*

### Abstract

**Objective** Hyperspectral images usually need to sacrifice spatial resolution to improve spectral resolution, which will lead to the emergence of a large number of mixed pixels and seriously affect the performance of subsequent applications. Convolutional neural networks (CNNs) can maximize the spatial resolution of hyperspectral images on the premise of ensuring spectral information integrity by fusing multispectral images. The 2D convolution scheme adopts 1D convolution and 2D convolution respectively for feature extraction of spectral and spatial information. However, 1D convolution can only take global spectral information into account and lacks attention to complementary spectral information between adjacent pixels, which easily results in insufficient spectral feature extraction. 3D convolution will introduce a large number of network parameters, limiting the depth and width of the design. Meanwhile, most network designs pay more attention to spatial feature extraction but ignore spectral dimension information to easily cause spectral confusion. In addition, as the network deepens, both spatial and spectral dimensions will lose information. The residual connection can alleviate this problem to a certain extent by ignoring the information differences among different input images, and it is difficult to employ the original input information to compensate the network. Therefore, hyperspectral image super-resolution needs to enhance the extraction of spectral information and improve the spatial resolution of images. In addition, the design should strengthen the adaptability to hyperspectral images to ensure that the network can accurately take advantage of the characteristics of different input images.

**Methods** To solve the insufficient utilization of intrinsic spectral features and supplement information more effectively in the fusion-based hyperspectral image super-resolution method, we propose a global-local attention feature reuse network (LGAR-Net). The network adopts low-resolution hyperspectral images and multispectral images with bicubic interpolation as the original input and designs a progressive structure. The progressive network leverages a few bands for the initial build first and gradually adds more band information to fine-tune the details for more accurate reconstruction effects. The network optimization features a reuse mechanism to preserve multi-scale spatial information while reducing the parameter number. Each progressive stage contains local attention blocks which employ spatial attention to enhance spatial information extraction and channel attention to supplement spectral information representation ability. Finally, we design a

global correction module. According to the characteristics of a high spatial abundance of multispectral images and high spectral fidelity of hyperspectral images, The module adopts a global attention mechanism to focus information of different dimensions on the two kinds of original inputs to supplement targeted global information and ensure network stability.

**Results and Discussions** To achieve the balance between performance and parameters, we design the module performance experiments to determine the specific number of extraction modules in feature reuse modules (Table 1). At the same time, we perform an ablation on the feature reuse connection, local attention blocks, and global correction module to determine the effectiveness of each core module (Table 2). In the comparative experiment, LGAR-Net is compared with other six representative advanced algorithms through quantitative evaluation, and two datasets of CAVE and Harvard are selected for evaluation. In the CAVE dataset results, LGAR-Netnet  $\times 4$  magnification results in PSNR and SSIM, SAM, and EGRA respectively reach 51.244 dB, 0.9644, 1.703, and 0.392, and prove the network advancement (Table 3). Additionally, LGAR-Net yields the best performance in both  $\times 8$  and  $\times 16$  magnification tasks, which verifies its strong adaptability to different magnification factors. LGAR-Net still achieves the best results by comparing the results of  $\times 4$ ,  $\times 8$ , and  $\times 16$  magnification in the Harvard dataset, which further proves the network generalization (Table 4). We carry out some qualitative experiments to further evaluate the model performance. The absolute error map is employed to reflect the differences between the reconstructed image and the real image (Fig. 6). In addition, we compare the spectral curve to reflect the spectral distortion condition (Fig. 7). The results of qualitative experiments also prove the superior performance of LGAR-Net.

**Conclusions** In this paper, we propose a hyperspectral image super-resolution network named LGAR-Net to obtain high-resolution hyperspectral images by integrating information from low-resolution hyperspectral images and multispectral images. The network refines the reconstruction effect continuously through the progressive network and adopts the feature reuse mechanism to retain multiple granularity information. Local attention is utilized to enhance spectral information extraction, and global attention is to make information compensation by the original input image characteristics for strengthening the network adaptation to hyperspectral images. In addition, the optimal number of modules is analyzed in the network design experiments, and the effectiveness of innovation points is proven by the ablation experiment. In the comparative experiment, LGAR-Net conducts quantitative and qualitative evaluations alongside other six methods on the CAVE and Harvard datasets. Across various magnifications, LGAR-Net consistently achieves outstanding results, demonstrating its effectiveness and advanced capabilities.

**Key words** hyperspectral images; super-resolution; attention mechanism; feature reuse; convolutional neural network