

基于分组 ConvLSTM 和 Transformer 的双分支 遥感高光谱图像超分辨率网络^①



边太成, 杨 锦, 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: 遥感高光谱图像超分辨率 (remote sensing hyperspectral image single super-resolution, HSISR) 任务近年来已取得可观进展, 其中使用深度卷积神经网络 (convolutional neural network, CNN) 技术的方法得到广泛运用. 然而, 大多数基于 CNN 的超分辨模型往往会忽略遥感高光谱图像的光谱结构, 同时由于卷积网络受卷积核大小限制, 长距离的特征依赖关系被忽略, 进而影响了重建的精度. 为了解决这些问题, 本文提出了一个基于分组 ConvLSTM 和 Transformer 的双分支遥感高光谱图像超分辨率网络 (dual-branch remote sensing hyperspectral image super-resolution network based on grouped ConvLSTM and Transformer, DGCTNet), 该方法结合了 Transformer 捕捉长距离依赖关系和卷积长短时记忆网络 (ConvLSTM) 对提取序列性特征的优势, 在提取空间特征的同时保持了光谱的有序性, 增强了重建图像的效果. 此外, DGCTNet 还设计了边缘学习网络, 将边缘信息扩散到图像空间中. 同时为重新校准光谱响应, 加入提出的双组级通道注意力机制 (dual-group level channel self-attention, DSA). 在 Houston 数据集上的实验表明, DGCTNet 方法在定量评价指标和多种场景下的视觉质量上, 都优于当前最先进的对比模型.

关键词: 遥感高光谱图像; 超分辨率; 边缘增强; ConvLSTM

引用格式: 边太成, 杨锦, 朱习军. 基于分组 ConvLSTM 和 Transformer 的双分支遥感高光谱图像超分辨率网络. 计算机系统应用, 2025, 34(3):286-295. <http://www.c-s-a.org.cn/1003-3254/9783.html>

Dual-branch Remote Sensing Hyperspectral Image Super-resolution Network Based on Grouped ConvLSTM and Transformer

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

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

Abstract: Remote sensing hyperspectral image single super-resolution (HSISR) tasks have made considerable progress in recent years. Methods using deep convolutional neural network (CNN) technology are widely employed. However, most CNN-based super-resolution models tend to ignore the spectral structure of remote sensing hyperspectral images. Meanwhile, due to the limitation of convolutional networks by the size of convolutional kernels, long-distance feature dependencies are ignored, which in turn affects the reconstruction accuracy. To solve these problems, this study proposes a dual-branch remote sensing hyperspectral image super-resolution network based on grouped ConvLSTM and Transformer (DGCTNet), which combines the advantages of Transformer in capturing long-distance dependencies and ConvLSTM in extracting sequential features. It enhances the reconstructed image effect by extracting spatial features while maintaining spectral orderliness. In addition, DGCTNet also designs an edge learning network to diffuse edge information into the image space. At the same time, to recalibrate the spectral response, the proposed dual-group level channel self-attention mechanism (DSA) is added. Experiments on the Houston dataset show that the proposed DGCTNet

^① 基金项目: 国家自然科学基金青年项目 (32301702); 山东省自然科学基金青年项目 (ZR2021QC120)

收稿时间: 2024-08-20; 修改时间: 2024-09-19; 采用时间: 2024-09-24; csa 在线出版时间: 2025-01-16

CNKI 网络首发时间: 2025-01-17

method outperforms the current state-of-the-art comparison models in terms of quantitative evaluation metrics and visual quality in a wide variety of scenarios.

Key words: remote sensing hyperspectral image; super-resolution; edge enhancement; ConvLSTM

单幅高光谱图像 (hyperspectral image single, HIS) 经常被应用于各种场景中, 包括考古^[1]、农业和水资源控制^[2]以及军事防御^[3]。近年来, 由于场景级别逐渐转变为像素级别任务, 对识别精度的要求大大提高, 需要遥感卫星捕捉更高空间分辨率和更丰富光谱细节的图像。然而, 由于物理设备的限制和成像时间的限制, 遥感卫星无法捕获具有高空间分辨率的高光谱图像。因此提高 HSI 的空间分辨率 (hyperspectral image super-resolution, HSISR) 具有重要意义。

现有的单幅图像超分辨率重建 (super-resolution reconstruction, SR) 方法可以分为两类^[4], 包括传统方法和深度学习方法。传统的图像 SR 主要有基于插值^[5]、基于正则化^[6]和基于稀疏表示^[7]的方法。这些方法在恢复图像信息时, 利用不同先验知识, 从有限信息中进行重建。但是由于经过人工筛选的特征表示能力有限, 无法恢复真实图像更多细节, 导致图像重建的质量较低。

近年来, 深度学习已成为实现单图像超分辨率的一种有效方法。Dong 等^[8]提出了第 1 个应用于超分辨率的卷积神经网络 SRCNN 证明了 CNN 在 LR 到 HR 图像信息的非线性映射关系方面优于传统方法。在后续提出的各类神经网络模型中, 残差网络因其在图形处理方面的出色表现^[9], 已被广泛应用于单图像超分辨率任务, 取得了惊人的重建效果。这主要归因于其多层次特征的表示和特征重用能力。随后, 研究者们开始使用 3D 卷积进行光谱信息的学习。3DFCNN^[10]网络利用 3D 卷积直接提取相邻像素的空间上下文和相邻波段的光谱相关性以减轻生成图像的光谱失真。三维网络中, 通常存在着大量的参数和复杂的计算需求。为解决这些问题, Sun 等^[11]设计一种混合卷积网络 (MCNet) 通过 2D/3D 可分离卷积的方式降低计算复杂度。

与传统方法相比, 基于深度学习的方法有着强大的非线性建模能力和自适应特征学习能力。因为在处理复杂地理环境、捕捉细微细节以及适应不同数据变化方面具有明显优势, 所以通常在 HSI 超分辨率重建任务中表现良好。目前, 基于深度学习的现有方法在处理高光谱图像空间超分辨率重建任务上仍存在两个重要挑战, 即数据的高维性和被过度侵蚀的空间信息。

一方面, HSI 的光谱维度虽然有助于反映地物微小的空间形态特性, 但却增加了特征提取的难度。因此, Li 等^[12]提出一种分组深度递归残差网络 (GDRRN), 该方法残差网络中嵌入分组递归模块以描述复杂特征信息。空间光谱先验网络 (SSPSR)^[13]的空间光谱特征提取块利用高效残差学习和注意机制, 充分挖掘了空间光谱先验信息。DualSR^[14]采用双级设计, 粗级阶段通过交替光谱融合机制逐带超分辨图像, 精细阶段使用增强反投影方法来学习空间-光谱一致性, 提升了重建性能。上述方法及其变体都没完全考虑到高光谱图像的光谱独立性, 进而导致重建图像的光谱信息紊乱。

另一方面, 高光谱图像具有广泛的覆盖范围和多样的地面场景, 但其空间信息被过度侵蚀和受到复杂退化过程的影响使得从低分辨率输入中恢复细节和锐利边缘对于神经网络 (CNN) 更加困难。为了获得清晰且易于区分的细节边缘, 研究人员进一步引入了不同的处理策略来独立学习图像边缘。Hui 等^[15]提出的细节分量预测模型去获得细节增强分量, 从而恢复图像细节信息。TE-ResUNet^[16]采用纹理增强模块来增强低层特征的纹理细节。Zhang 等^[17]设计边缘增强高效网络, 使用边缘增强卷积块, 通过多阶梯度提取和其他特征增强分支方式丰富网络的边缘感知能力。TE-SAGAN^[18]采用了残差密集生成器来改善生成的边缘轮廓和纹理, 并使用知觉损失用来巩固中高频率的细节。上述的边缘增强方法假设输入图像的边缘是清晰的和容易推断的, 使用这些假设对图像边缘进行增强。因此导致高光谱图像空间信息被过度侵蚀, 边缘细节信息模糊且噪声污染严重, 很难从这些先验中恢复真实边缘细节信息。

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^[19]。ConvLSTM 在降水预测、动作识别、交通预测等序列任务中均展现了良好的性能。此外, 在高光谱图像分类任务中 ConvLSTM 也展示了其在处理高光谱图像相关任务中捕捉和利用正则光谱信息的显著能力, 并成功应用于该分类任务中, 如植被分类、矿物分类等, 他们将 ConvLSTM 作为 HSI 特征提取的基本单元, 旨在同时捕捉 HSI 的空间信息和光谱信息。在类似的思想

下, Lu 等^[20]提出了 ConvBiLSTM 网络用于高光谱图像超分辨率重建, 采用逐波段传播方法将卷积和递归神经网络相结合, 以改善光谱保真度. 然而 ConvLSTM 在一定程度上对图像空间上的噪声是敏感的. 由于 ConvLSTM 通过卷积操作处理时空数据, 卷积层对输入数据的每个位置都有局部感知和特征提取的能力. 因此, 如果图像空间上存在较强的噪声或干扰, ConvLSTM 可能会受到影响, 并在处理中保留这些噪声信息. 受以上思路的启发, 本文提出了基于光谱强化 Transformer 和 ConvLSTM 的特征强化高光谱图像超分辨率网络, 采用了双分支的网络架构, 同时加入了双组级通道注意力机制, 在增强对 HSI 高频信息重建能力的同时也保证了光谱-空间信息的重建能力. 本文主要贡献如下.

(1) 提出了一种基于分组 ConvLSTM 和 Transformer

的双分支高光谱图像超分辨率网络 (DGCTNet), 结合了 Transformer 捕捉长距离依赖特征和 ConvLSTM 对提取序列性特征的优势, 增强了重建图像的效果.

(2) 提出了一种有效实用的边缘学习模块来增强图像细节信息, 既能生成接近真实边缘细节信息, 又能有效去除噪声. 同时加入提出的双组级通道注意力保持了光谱信息的有序性.

(3) 在 Houston 上进行的实验表明, DGCTNet 在高光谱图像超分辨率重建方面明显优于其他方法.

1 本文方法

1.1 DGCTNet 网络模型

图 1 为基于分组 ConvLSTM 和 Transformer 的双分支高光谱图像超分辨率网络 (DGCTNet) 的总结构.

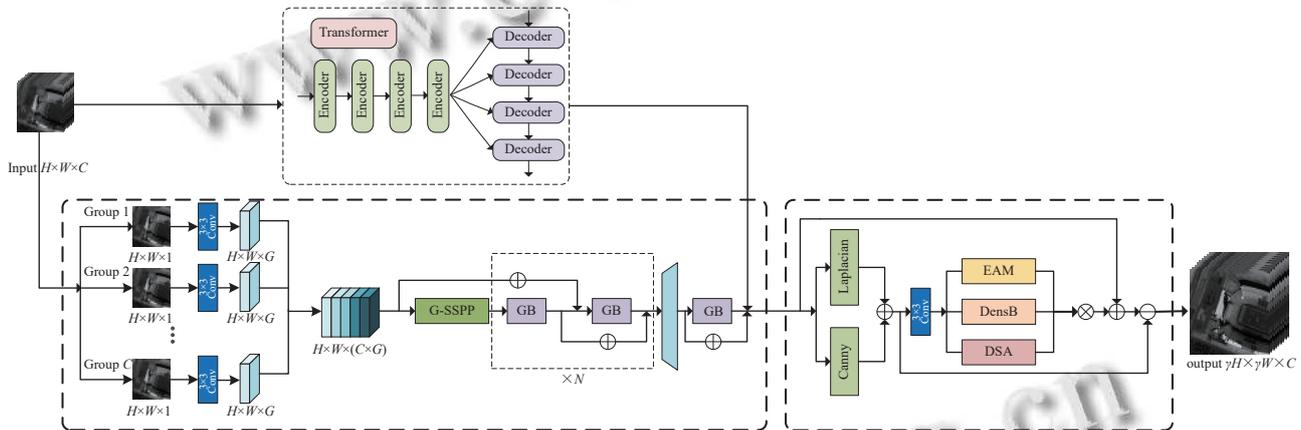


图 1 总网络结构示意图

该模型主要由两部分组成: 特征提取模块和边缘增强模块. 特征提取模块可以分为两个分支结构: Transformer 高频特征提取分支和基于 ConvLSTM 的特征提取分支. 由于遥感图像的尺寸较大且维度较高, 普通的卷积网络对于捕捉长距离信息依赖度低, 本文利用 Transformer 网络对图像中的高频信息进行提取. 由于在 Transformer 特征提取过程中会破坏光谱信息规律, 因此在第 2 个分支中采用分组 ConvLSTM 网络对特征进行提取有效地避免了光谱规律的丢失, 同时加入提出的分组空间-光谱金字塔池化模块 (G-SSPP) 进行光谱信息的提取, 最后将两个分支进行特征融合以保证光谱信息的有序性.

第 2 分支中的边缘增强模块是由一个双路并行的边缘提取子模块构成, 每一路边缘提取子模块都是由多个密集块、用于清除噪声和伪影的边缘注意力模

块、双组级通道注意力 (DSA) 组成.

输入的低分辨率图像 (LR) 首先进入两个分支分别进行提取, 第 1 分支经过 Transformer 特征提取得到高频特征图像 I_{Base}^* , 第 2 分支经过 ConvLSTM 网络特征提取后与图像 I_{Base}^* 融合生成中间高分辨率图像 (HR) I_{Base} , 然后通过边缘增强模块得到一个纯净无噪声的图像边缘 I_{Edge}^* , 最后将图像边缘 I_{Edge}^* 与 I_{Base} 结合得到最终的超分辨率图像 (SR).

本文的目标是使提出的模型学到从 LR 到高空间分辨率图像 (HR) 的映射 F , 如下所示:

$$I_{HR} = F(I_{LR}; \theta) \quad (1)$$

其中, $I_{LR} \in R_{H \times W \times C}$ 和 $I_{HR} \in R_{H \times W \times C}$ 分别表示低分辨率高光谱图像 (LRHSI) 和高分辨率高光谱图像 (HRHSI), H 表示每个通道图像的高, W 表示每个通道图像的宽, C 表示图像的通道数, γ 表示放大因子, F 表示映射网络.

1.2 特征提取模块

我们使用了两分支特征提取方法对图像进行重建. 第1个分支为 Transformer 特征提取模块, 传统的卷积神经网络在提取对于遥感高光谱这种大图像时很难对图像整体的相关特征进行提取, Transformer 方法凭借其对长距离依赖特征的捕捉, 弥补了这一缺陷. LR 图像首先通过 Transformer 网络进行特征提取.

$$I_{Base}^* = T(I_{LR}) \quad (2)$$

其中, I_{Base}^* 表示由 Transformer 网络提取到的高频特征信息图像, T 表示 Transformer 网络的映射, I_{LR} 表示输入的低分辨率图像.

图像首先被输入到编码器中由图像映射到连续的向量空间中, 编码器由多个相同的层级组成, 每个层级包括一个多头自注意力层和一个前馈神经网络层. 其

中, 多头自注意力层用于捕捉序列中的依赖关系, 前馈神经网络层用于对位置向量进行非线性变换. 然后通过解码器利用编码器输出的向量来生成目标图像.

经过 Transformer 网络提取过特征的虽然会得到较好的空间信息, 但是由于卷积过程中通道顺序被打乱, 经过特征提取图像的光谱信息会相对较差, 因此在第2分支采用提出的基于 ConvLSTM 的分组分支, 该分支可以充分利用图像的空间和光谱信息, 通过以波段为单位的分组策略减少计算负担并保护光谱信息. 同时利用多尺度策略对不同尺度的特征信息进行提取, 以此来对图像信息进行深度挖掘, 进而恢复出更多的信息. 由图1可知基于 ConvLSTM 的分组模块, 该模块由分组空间-光谱金字塔池化模块 (G-SSPP) 以及残差分组 ConvLSTM (ResGB) 组成. G-SSPP 如图2所示.

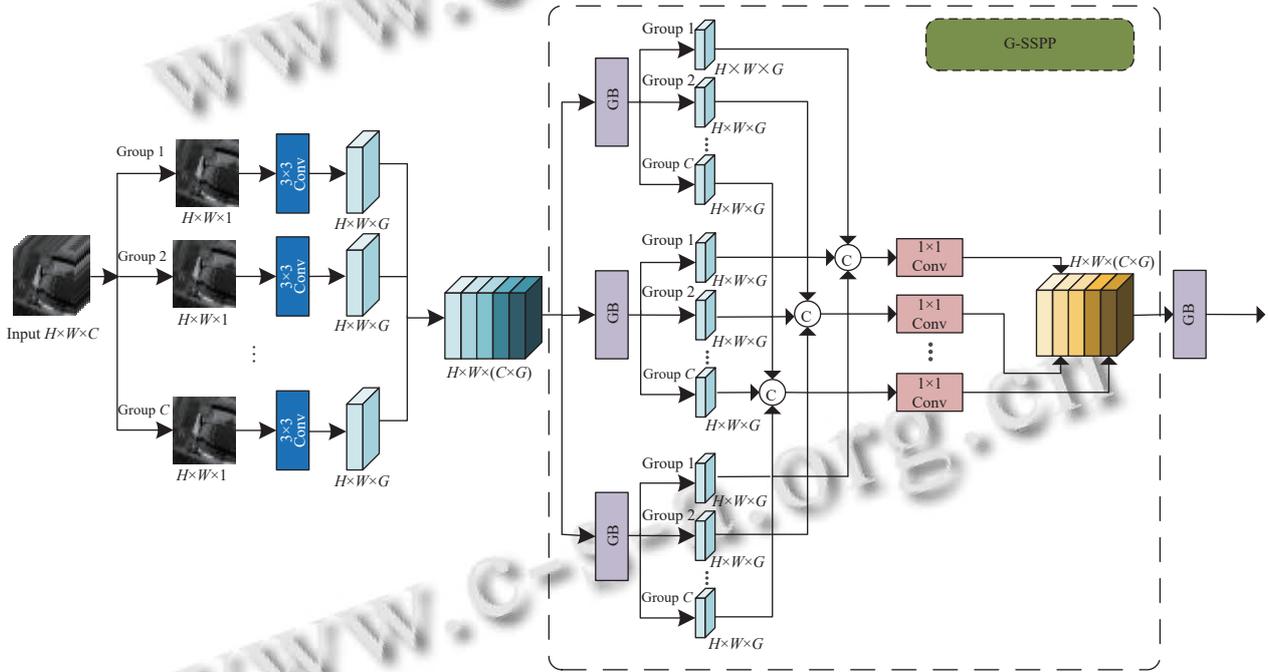


图2 G-SSPP 模块

首先, 将 $I_{LR} = [I_{LR}^1, \dots, I_{LR}^g, \dots, I_{LR}^C]$ 通过分组卷积层 F_{GConv} 进行浅层特征提取, 表示为:

$$F_0^g = F_{GConv}(I_{LR}^g) \quad (3)$$

其中, $I_{LR}^g \in R_{H \times W \times 1}$ 与 $F_0^g \in R_{H \times W \times k}$ 分别表示 I_{LR} 的第 g 个波段和 F_{GConv} 第 g 组的输出. k 表示每个组内的通道数. 因此, 该层的总输出可以看作 $F_0 = [F_0^1, \dots, F_0^g, \dots, F_0^C]$, $F_0 \in R_{H \times W \times (k \times C)}$.

其次, 被分组后的特征 F_0 将进入 G-SSPP 模块进

行光谱信息的修正, 该模块使用具有多尺度的分组双向 ConvLSTM 方法 F_{GB} , 以在不同尺度上提取空间特征并在波段组之间传播. 在获取不同尺度下 F_{GB} 的输出之后, 它们以组间对应的方式进行拼接. 然后, 将拼接结果通过 1×1 卷积进行分组融合. 最后, 融合的输出与进行残差连接. 上述 G-SSPP 模块可以表示为:

$$F_{ms} = F_{BF}[F_{GB^{r1}}(F_0), F_{GB^{r2}}(F_0), F_{GB^{r3}}(F_0)] + F_0 \quad (4)$$

其中, F_{GB^r} 表示具有多尺度的方法. 在本节中, $r1$ 、 $r2$ 和 $r3$ 分别设置为 4、8 和 16. F_{BF} 表示融合 1×1 卷积

操作. 为了进一步提取丰富的详细特征, 接着将 F_{ms} 输入到由 n 个连续方法组成的 ResGB 块中. 输出通过残差连接与结合, 得到第 2 分支特征提取的最终输出:

$$F_a = F_{GBn}(F_{GBn-1}(\dots F_{GB1}(F_{ms})\dots)) + F_{ms} \quad (5)$$

其中, F_{GBn} 表示第 n 个连续的 F_{GB} 方法.

本文使用基于 Bi-ConvLSTM 的网络进行特征提取, 在本阶段使用共享的 ConvLSTM 单元. ConvLSTM 单元具有同时捕捉空间和光谱信息的能力. 此外, 对前向和后向输出的融合方式也需特别指出, 分组共享 Bi-ConvLSTM 方法通过将对应的组沿着通道拼接, 然后再通过 1×1 卷积降维到原来组的大小来进行前向后向的特征融合. 此做法可以避免直接相加的融合方式造成信息丢失. 同时引入残差机制, 有效避免梯度消失和爆炸的发生. 在特征提取的最后, 输出特征图通过分组的 PixelShuffle 方法将尺寸放大到目标大小.

$$F_{UP} = F_{GU}(F_a) \quad (6)$$

其中, F_{GU} 与 $F_{UP} \in R_{\gamma H \times \gamma W \times (t \times C)}$ 表示分组 PixelShuffle 方法及其对应的输出. 最后将两个分支提取的特征相融合得到中间的 HR 图像 I_{Base} .

$$I_{Base} = F_{UP} + I_{Base}^* \quad (7)$$

其中, F_{UP} 和 I_{Base}^* 分别代表由基于 ConvLSTM 的分组模块和 Transformer 特征提取模块特提取到的特征图像.

1.3 边缘增强网络

特征提取后得到空间信息丰富并且光谱信息有序的中间 HR 图像 I_{Base} . 然而受到卷积网络和损失函数的限制, 其边缘信息相对其他空间信息更为平滑且模糊. 为了得到更加锐化的边缘信息, 受 Jiang 等^[21]的启发本文设计了边缘学习模块, 对经过空间和光谱信息重建部分重建的中间图像 I_{Base} 进行边缘特征提取. 边缘学习网络由边缘特征提取和边缘特征学习组成, 结构如图 3 所示.

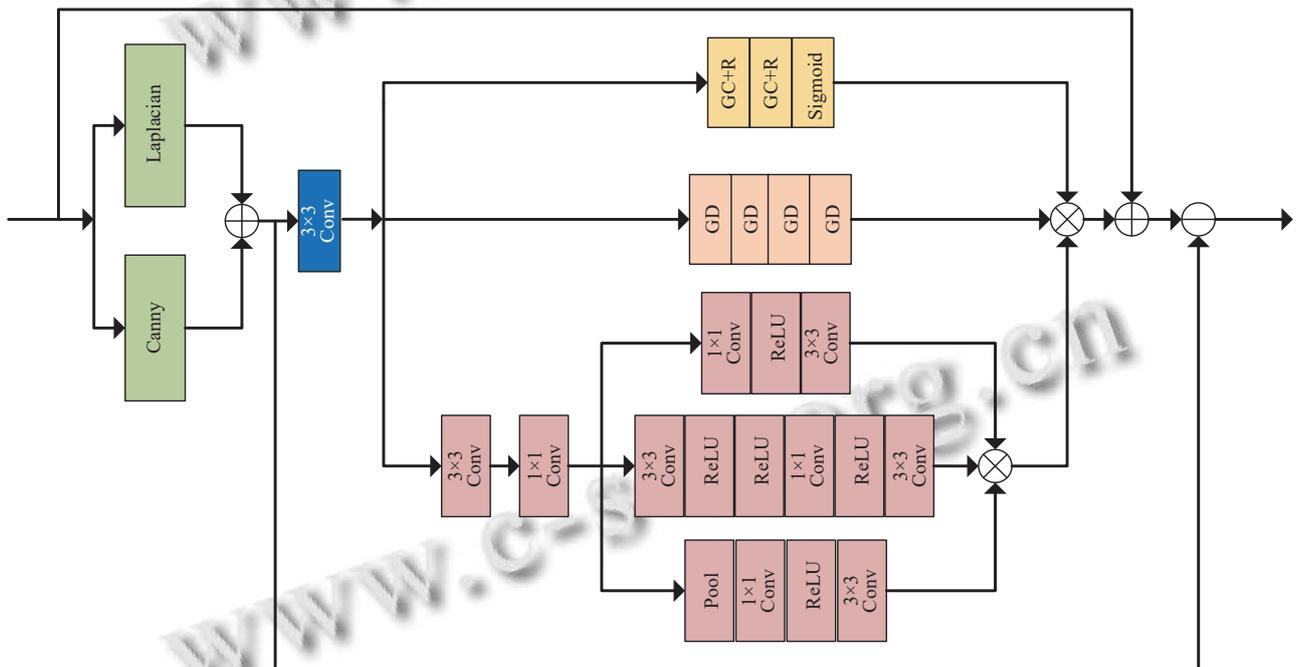


图 3 边缘学习网络

首先, 边缘特征提取使用 Laplacian 算子和 Canny 算子从中间结果中分别提取边缘粗边缘和细边缘. 其次, 使用边缘特征学习对边缘信息进行应映射, 为了能够过滤噪声更专注的学习边缘信息, 我们设计了边缘注意力机制 (edge attention mechanism, EAM) 自适应地学习真实边缘特征的权重矩阵. 最后再加入提出的双组注意力机制来完成对光谱的有序排列.

作为典型的方法, Laplacian 算子可以突出灰度变

化最剧烈的区域, 但对噪声敏感它会生成较粗的边缘. Canny 算法对噪声有较好的抵抗能力, 它通常能够产生 Laplacian 算子捕捉不到细致的边缘. 这两个算子被广泛应用于之前的边缘检测和提取. 两个算法捕捉的边缘如图 4 所示. 图 4 可以看出两算法提取到的边缘信息的侧重点有所不同. Laplacian 算子提取的边缘虽然有噪声, 但是可以提取到灰度变化剧烈边缘. Canny 算法提取到的边缘信息偏向结构化, 有助于重建图像

的结构特征. 因此, 我们的边缘特征提取采用 Laplacian 和 Canny 相结合的方式捕捉 I_{Base} 中较为全面的边缘细节信息. 边缘特征提取整个过程可表示为:

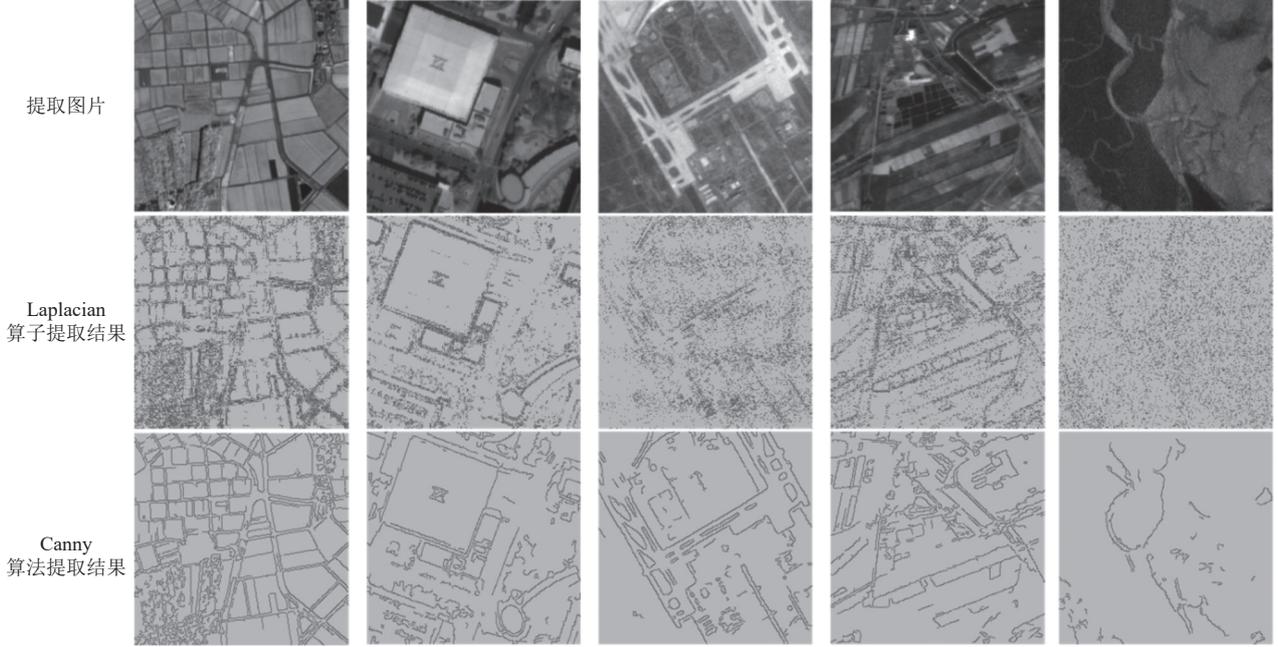


图4 边缘提取可视化

边缘特征学习包括低级边缘特征提取块、Dens-Block 块、边缘注意力机制 (EAM) 和双组级通道注意力机制 (DSA). 低级边缘特征提取块对 I_{Edge} 中的边缘细节信息进行表示. DensBlock 负责探索深度特征. 同时, 边缘注意机制分支学习真实边缘权重, 检测并挑选真实边缘, 对噪声点进行剔除. 双组级通道注意力着重于对光谱信息的修正. 边缘特征学习过程可表示为:

$$I_{\text{Edge}}^* = \text{Ad}(\text{EAM}(\text{DensB}(I_{\text{Edge}})) \otimes \text{DensB}(\text{DensB}(I_{\text{Edge}})) \otimes \text{DSA}(\text{DensB}(I_{\text{Edge}})), I_{\text{Edge}}) \quad (9)$$

其中, ConvB 表示 6 个 3×3 卷积层和 ReLU, DensB 表示密集连接残差块, EAM 表示边缘注意机制操作, DSA 表示双组级通道注意力, Ad 表示将两个特征图相加, \otimes 表示逐元素相乘, I_{Edge}^* 表示学习到的边缘特征. 然后, 将 I_{Edge}^* 添加到 I_{Base} , 并将其中的 I_{Edge} 成分减去, 最终得到具有清晰和锐利边缘的高空间分辨率空间, 可表示为:

$$SR = \text{Subtrac}(\text{Add}(I_{\text{Base}}, I_{\text{Edge}}^*), I_{\text{Edge}}) \quad (10)$$

其中, Subtrac 表示第 1 项逐元素相减第 2 项, SR 表示具有接近真实边缘高空间分辨率空间.

1.4 损失函数

根据相关研究表明, 在处理图像异常值时, Char-

$$I_{\text{Edge}} = L(I_{\text{Base}}) + C(I_{\text{Base}}) \quad (8)$$

其中, L 表示 Laplacian 算子, C 表示 Canny 算子, I_{Edge} 表示从 I_{Base} 中提取的较为全面的边缘细节信息.

bonnier 损失函数在超分辨率任务中的表现优于常用的 L_1 和 L_2 损失函数. 因此我们选择 Charbonnier 损失作为组合损失函数的一部分. 该损失函数表示如下:

$$\mathcal{L}_C(\Theta) = \frac{1}{N} \sum_{n=1}^N \sqrt{I_{\text{HR}}^n - F_{\text{DGCT}}(I_{\text{LR}}^n)^2 + \epsilon^2} \quad (11)$$

其中, F_{DGCT} 是提出的 DGCTNet. LR 与 HR 分别表示低空间分辨率高光谱图像和高空间分辨率高光谱图像, N 表示每个 epoch 的训练批次, ϵ 代表 Charbonnier 损失函数中的常数, 本文将其设置为原始论文中相同的值: 0.001. Θ 表示 DGCTNet 的参数集合.

此外, 仅使用 Charbonnier 损失函数来调节模型可能会导致超分辨率图像中因光谱维度信息和噪声未被充分约束引起的图像粗糙问题. 为了解决这些问题, 将总变差损失函数 (SSTV) 作为总损失函数的另一个组成部分. 总变差函数的公式如下:

$$\mathcal{L}_{\text{SSTV}} = \frac{1}{N} \sum_{n=1}^N (\nabla_h F_{\text{DGCT}}(I_{\text{LR}}^n)_1 + \nabla_w F_{\text{DGCT}}(I_{\text{LR}}^n)_1 + \nabla_c F_{\text{DGCT}}(I_{\text{LR}}^n)_1) \quad (12)$$

其中, ∇_h 、 ∇_w 和 ∇_c 分别表示 DGCTNet 输出的垂直、水平和光谱梯度函数. 本文使用的完整损失函数表示为:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_C + \lambda \mathcal{L}_{\text{SSTV}} \quad (13)$$

其中, 参数 λ 代表控制重建图像锐化程度的权重, 在本文中将其设置为0.001.

2 实验结果与分析

本节通过大量实验对提出的网络进行测试, 并与目前主流的高光谱超分辨率重建方法的结果进行对比分析. 为验证模块有效性, 进一步对提出的网络关键模块进行消融实验.

2.1 数据集和评价指标

使用一个遥感高光谱数据集对所提出的网络进行验证: Houston 数据集. Houston 数据集是由高光谱图像分析实验室和国家航空激光测绘中心 (the Hyperspectral Image Analysis Laboratory and the National Center for Airborne Laser Mapping) 使用 ITRES CASI 1500 光谱成像仪捕获的, 涵盖了休斯顿大学校园及周边城区的高光谱图像数据, 在 2017 年 2 月 16 日发布. 空间分辨率为 1 m, 覆盖了 380–1 050 nm 范围内的 48 个波段. 图像的空间尺寸为 4172×1202 像素, 本实验中使用了 Houston 数据集中两幅真实遥感图像数据. 数据集具有植被、郊区和各种建筑等典型地表覆盖类

型的区域, 还有平原、山区等多种地理形态.

本文使用 5 个常用定量图像质量指标对所提出方法的性能进行评估, 包括光谱角映射器 (SAM)、相对全局二维误差 (ERGAS)、交叉验证 (CC)、结构相似性 (SSIM) 和峰值信噪比 (PSNR). 为更好地描述模型在测试集上的整体性能, 分别计算 PSNR 和 SSIM 作为所有波段的平均值, 分别表示为 MPSNR 和 MSSIM. 上述 5 个指标的最佳值分别为 0、0、1、1 和 $+\infty$.

2.2 实验细节

每个数据集分为训练数据和测试数据. 具体来说, 我们将场景划分为两个不重叠的区域, 其中 80% 的图像用于训练, 20% 的图像用于构成测试集. 此外, 对数据集进行裁块处理, 提取重叠块作为训练数据集的参考高分辨率高光谱图像. 当放大倍数为 2 时, 提取的重叠块大小为 64×64 像素, 重叠为 16 像素. 当放大倍数为 4 时, 使用同样大小的重叠块, 但重叠像素为 32 像素. 实验过程选择使用 Adam 优化器, 搭载在配置为: NVIDIA GeForce A4000 16 GB 的显卡上进行训练.

2.3 对比结果与分析

本节将本文方法与流行的高光谱图像超分辨率重建方法进行对比. 实验结果如表 1 所示.

表 1 不同算法在 Houston 数据集上测试的定量评价结果

Scale	Metric	Bicubic ^[5]	DualSR ^[14]	SSPSR ^[13]	MCNet ^[11]	GDRRN ^[12]	3DFCNN ^[10]	DGCTNet
×2	MPSNR (dB)↑	34.8654	36.9123	37.2469	39.5021	40.5243	<u>40.6165</u>	41.1389
	MSSIM↑	0.9576	0.9723	0.9812	0.9841	0.9850	<u>0.9858</u>	0.9862
	SAM↓	2.0910	1.71274	1.5498	1.3289	1.2476	<u>1.2011</u>	1.1855
	ERGAS↓	0.6486	0.5698	0.4401	0.3534	0.3201	<u>0.3171</u>	0.3102
	CC↑	0.8235	0.8992	0.9088	0.9365	0.9438	<u>0.9479</u>	0.9500
×4	MPSNR (dB)↑	30.2019	34.9539	34.4398	35.8470	35.9960	<u>35.9210</u>	36.5923
	MSSIM↑	0.8587	0.9465	0.9525	0.9549	0.9651	<u>0.9715</u>	0.9752
	SAM↓	2.5380	2.0492	1.7901	1.7130	<u>1.4421</u>	1.5333	1.4238
	ERGAS↓	1.7147	1.3001	0.9030	0.8460	<u>0.4231</u>	0.8518	0.4199
	CC↑	0.8649	0.8955	0.9030	0.9131	0.9125	<u>0.9120</u>	0.9142

表 1 总结了在 Houston 的 3 幅图像数据集上进行 6 种方法对比实验的平均定量重建性能, 最好结果以粗体突出显示, 次好结果以下划线标识. 从表 1 可以看出, DGCTNet 方法在两种尺度的所有指标上, 均获得最佳值, 优于其他方法. 这是由于 DGCTNet 可在长距离特征提取的同时保持光谱的有序性和学习真实有效的边缘信息. SSPSR 和 DualSR 由于模型深度较浅, 表达能力有限, 导致重建结果较差. 虽然 MCNet 通过递归结构增加深度, 但其通过降维步骤丢失了图像细节特征. 相比之下, 复杂度相对较高的 3DFCNN 和 GDRRN 模

型表现出明显优势, 定量结果均得到大幅提高.

图 5 和图 6 分别是 Houston 数据集两幅图像的不同重建方法在×2 尺度上的可视化, 以 3 波段显示. 图 5 图 6 中红色矩形部分包含丰富边缘和纹理信息, 是 HSISR 重建的一个难点. 图 5 中地物较小, 纹理信息丰富. 对于图 5 红色矩形区域, DGCTNet 仍然可以重建出清晰且没有伪影的图像. 图 6 地物较大, 边缘结构信息较为丰富. 具体来看, 图 6 红色矩形区域, GDRRN、MCNet 和 DualSR 在恢复时提供了过平滑的轮廓, 而 3DFCNN 和 SSPSR 的重建结果造成了图像畸变. 总的

来说,提出的 DGCTNet 方法可以提高图像的清晰度,具有很强的适应性.图 6 右侧内容以边缘结构不规则的自然森林为主.3DFCNN、GDRNN 和 DualSR 重建

的边缘信息模糊,MCNet 和 SSPSR 重建结果中存在伪影和畸变.而 DGCTNet 的重建结果与原始图像最为接近,这证明本文方法更适用于重建不规则自然图像.

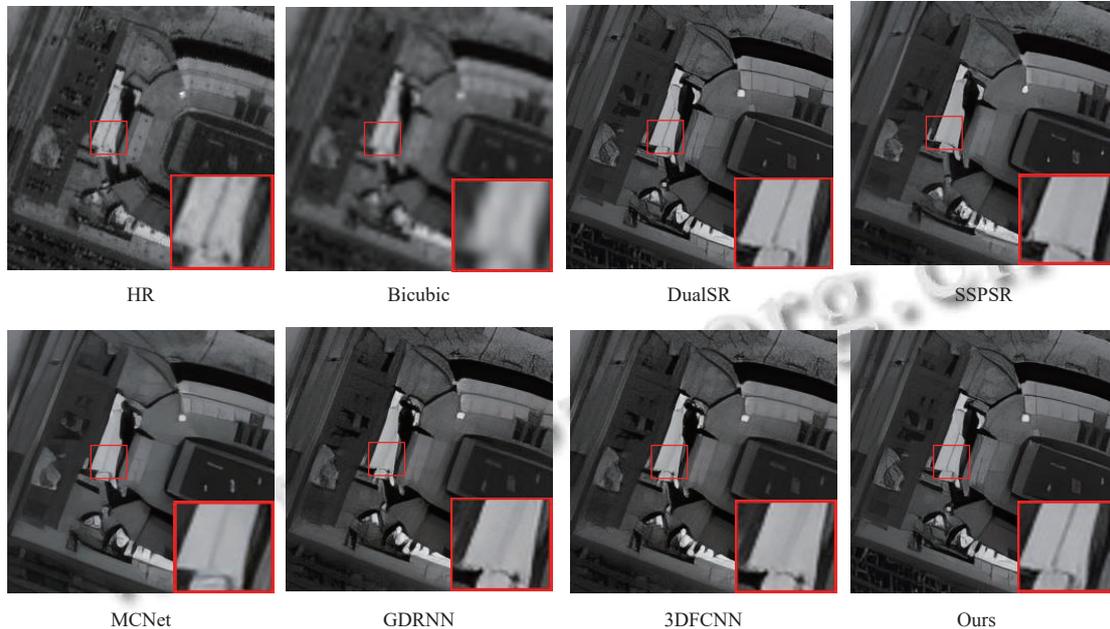


图 5 2 倍放大因子下 Houston 数据集在不同方法下的可视化

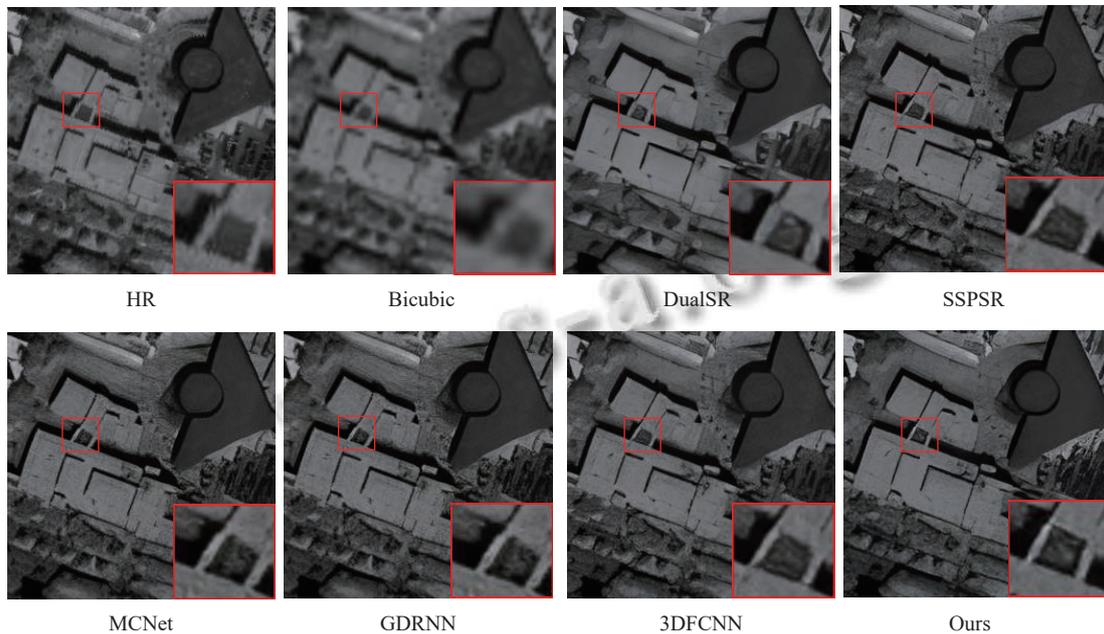


图 6 2 倍放大因子下 Houston 数据集在不同方法下的可视化

为了评估重建结果的质量,图 7 展示了 Houston 重建图像 4 倍放大因子下的平均误差曲线图.该曲线图的原理是通过将重建后的图像与真实标签图像进行差异计算,并在空间维度上求取平均值,然后在每个光谱维度上得到一个平均误差值.将所有维度的平均误

差值绘制为折线图的数据点,最终得到平均误差折线图.图中的折线值越高,表示重建后图像与真实标签之间的差距越大;反之,折线值越低表示重建后图像与相应通道的标签越接近.

Houston 数据集数据量大,具有多种地表覆盖类型

和多种地理形态的多种场景特点. 提出的方法可以很好地重建这些有规则的人工图案和无规则的自然形态.

至此可以得出结论, DGCTNet 适用于处理多种场景数据, 具有很强的鲁棒性.

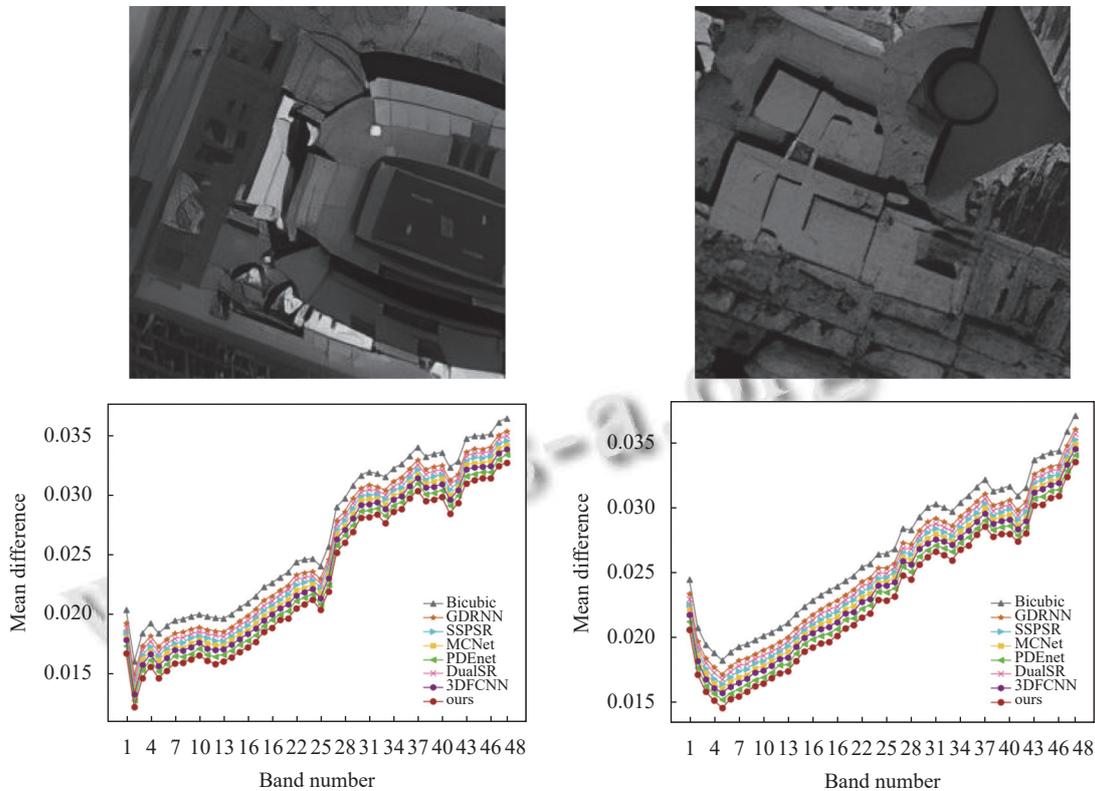


图7 4倍放大因子下 Houston 数据集的平均误差曲线

2.4 消融实验

为了验证提出的方法的有效性, 我们在 Houston 数据集上进行了消融实验, 图像上采样倍数为 2 和 4. 对所提出网络的每一个组件进行单独的测试, 并与完整的网络进行比较, 结果如表 2 所示.

表 2 在 Houston 数据集上的模块消融实验结果

Models	d	CC \uparrow	SAM \downarrow	ERGAS \downarrow	MPSNR (dB) \uparrow	MSSIM \uparrow
DGCTNet	2	0.9500	1.1855	0.3102	41.1389	0.9862
去掉Transformer	2	0.9417	1.2019	0.5137	40.4192	0.9625
去掉分组共享 Bi-ConvLSTM	2	0.9281	1.4891	0.6129	39.5740	0.9541
去掉边缘学习网络	2	0.9412	1.3012	0.5001	40.2600	0.9578
去掉通道注意力	2	0.9301	1.3001	0.4916	40.4793	0.9611
DGCTNet	4	0.9142	1.4238	0.4199	36.6380	0.9752
去掉Transformer	4	0.9019	1.6121	0.9698	34.5427	0.9425
去掉分组共享 Bi-ConvLSTM	4	0.8914	1.9510	2.2930	32.9891	0.9022
去掉边缘学习网络	4	0.9001	1.5673	1.2662	34.3863	0.9389
去掉通道注意力	4	0.9092	1.6264	1.9775	34.5626	0.9500

如表 2 所示, 提出的 DGCTNet 网络的每个模块都能够提高其在 HSISR 任务上的表现. 分组共享 Bi-

ConvLSTM 对网络性能的影响最大, 这是因为分组共享 Bi-ConvLSTM 通过在组内传输过程中共享 ConvLSTM 单元有效的捕捉了空间信息, 减少了对图像噪声的敏感性, 提高了学习特征的泛化性. Transformer 分支可以捕获全局的信息, 对于高频信息的提取也更为精准. 所提出的边缘学习模块专注于将图像的边缘进行锐化, 通过 Laplacian 算子和 Canny 算子将图像的边缘细节保留并进行特征提取后原特征图相融合, 使得图像的边缘更加锐利. 并且在边缘学习模块加入双组级通道注意力, 在将其去掉后网络重建效果会变差, 这是因为通道注意力会更加关注通道之间的权重, 在对特征进行利用时发挥作用, 增强了对有效特征的利用.

3 结论与展望

本文提出了一种基于 ConvLSTM 和 Transformer 的双分支高光谱图像超分辨率网络. 该网络充分利用 ConvLSTM 保护光谱维度顺序的能力及 Transformer 长距离捕捉信息的能力. 利用边缘注意力机制和双组级通道注意力机制, 在将图像边缘锐化的同时保持了

光谱的有序性。

对于高光谱图像超分辨率任务,目前仍存在局限性。模型在单个高光谱数据集上训练,难以将单个训练好的模型应用于其他数据集。这是因为不同高光谱数据集由不同摄像机收集,从而具有不同光谱响应函数等基本信息。因此,当应用模型于不同的高光谱数据集时,仍需要调整输入输出并重新训练网络。

参考文献

- 谭克龙, 万余庆, 杨一德, 等. 高光谱遥感考古探索研究. 红外与毫米波学报, 2005, 24(6): 437-440. [doi: 10.3321/j.issn.1001-9014.2005.06.009]
- 杨含成. 高光谱遥感技术在智慧农业中的应用研究. 南方农机, 2022, 53(20): 156-158. [doi: 10.3969/j.issn.1672-3872.2022.20.046]
- 许洪, 王向军. 多光谱、超光谱成像技术在军事上的应用. 红外与激光工程, 2007, 36(1): 13-17. [doi: 10.3969/j.issn.1007-2276.2007.01.004]
- Haris M, Shakhnarovich G, Ukita N. Recurrent back-projection network for video super-resolution. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 3892-3901.
- 熊卫辉. 基于稀疏表示和字典学习的单帧图像超分辨率算法研究 [硕士学位论文]. 昆明: 昆明理工大学, 2013.
- 路庆春. 正则化超分辨率图像重建算法研究 [硕士学位论文]. 合肥: 中国科学技术大学, 2010.
- Yang JC, Wright J, Huang TS, *et al.* Image super-resolution via sparse representation. IEEE Transactions on Image Processing, 2010, 19(11): 2861-2873. [doi: 10.1109/TIP.2010.2050625]
- 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]
- He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770-778.
- 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]
- Sun L, Li Y, Zheng M, *et al.* MCNet: Multiscale visible image and infrared image fusion network. Signal Processing, 2023, 208: 108996. [doi: 10.1016/j.sigpro.2023.108996]
- Li Y, Zhang L, Ding C, *et al.* Single hyperspectral image super-resolution with grouped deep recursive residual network. Proceedings of the 4th IEEE International Conference on Multimedia Big Data. Xi'an: IEEE, 2018. 1-4.
- 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]
- 呼亚萍, 孔韦韦, 李萌, 等. 基于边缘检测全变分模型的图像去噪方法. 现代电子技术, 2021, 44(5): 52-56. [doi: 10.16652/j.issn.1004-373x.2021.05.011]
- Hui YM, Wang J, Shi Y, *et al.* Low light image enhancement algorithm based on detail prediction and attention mechanism. Entropy, 2022, 24(6): 815. [doi: 10.3390/e24060815]
- Pan XR, Yang LN, Sun X, *et al.* Research on the extraction of hazard sources along high-speed railways from high-resolution remote sensing images based on TE-ResUNet. Sensors, 2022, 22(10): 3784. [doi: 10.3390/s22103784]
- Zhang TL, Chen HZ, Chen S, *et al.* Edge-enhanced efficient network for remote sensing image super-resolution. International Journal of Remote Sensing, 2022, 43(14): 5324-5347. [doi: 10.1080/01431161.2022.2128924]
- Xu YY, Luo W, Hu AN, *et al.* TE-SAGAN: An improved generative adversarial network for remote sensing super-resolution images. Remote Sensing, 2022, 14(10): 2425. [doi: 10.3390/rs14102425]
- 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]
- Lu JJ, Tang CL. Short-time prediction of image motion trajectories based on convolutional long and short-term memory neural networks. Proceedings of the 5th IEEE International Conference on Civil Aviation Safety and Information Technology. Dali: IEEE, 2023. 1-6. [doi: 10.1109/ICCSIT58768.2023.10351678]
- Jiang K, Wang ZY, Yi P, *et al.* Edge-enhanced GAN for remote sensing image super-resolution. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(8): 5799-5812. [doi: 10.1109/TGRS.2019.2902431]

(校对责编: 张重毅)