

# 基于多尺度特征提取和坐标注意力的光学遥感图像超分辨率重建<sup>①</sup>



肖子安<sup>1,2</sup>, 张静<sup>1</sup>, 苑馨方<sup>1</sup>, 朱家佳<sup>1</sup>, 李晓辉<sup>1</sup>, 米琳<sup>1</sup>, 窦帅<sup>1</sup>

<sup>1</sup>(中国科学院 空天信息创新研究院 中国科学院定量遥感信息技术重点实验室, 北京 100094)

<sup>2</sup>(中国科学院大学 电子电气与通信工程学院, 北京 100049)

通信作者: 李晓辉, E-mail: xhli@aoe.ac.cn

**摘要:** 本文针对现有光学遥感图像超分辨率重建模型对感受野尺度关注不足和对特征通道信息提取不充分带来的问题, 提出了一种基于多尺度特征提取和坐标注意力的光学遥感图像超分辨率重建模型. 该重建模型基于深度残差网络结构, 在网络的高频分支中设计了多个级联的多尺度特征和坐标注意力模块 (multi-scale feature & coordinate attention block, MFCAB), 对输入的低分辨率光学遥感图像的高频特征进行充分发掘: 首先, 在 MFCAB 模块中引入 Inception 子模块, 使用不同尺度的卷积核捕捉不同感受野下的空间特征; 其次, 在 Inception 子模块后增加坐标注意力子模块, 同时关注通道与坐标两个维度, 以获得更好的通道注意力效果; 最后, 对各 MFCAB 模块提取的特征进行多路径融合, 实现多重多尺度空间信息与通道注意信息的有效融合. 本文模型在 NWPU4500 数据集上 2 倍、3 倍放大中 PSNR 值达到 34.73 dB 和 30.12 dB, 较 EDSR 分别提升 0.66 dB 和 0.01 dB, 在 AID1600 数据集上 2 倍、3 倍、4 倍放大中 PSNR 值达到 34.71 dB、30.58 dB、28.44 dB, 较 EDSR 分别提升 0.09 dB、0.03 dB、0.04 dB. 实验结果表明, 该模型在光学遥感图像数据集上的重建效果优于主流的图像超分辨率重建模型.

**关键词:** 遥感图像; 超分辨率重建; 深度学习; 注意力机制; 残差网络

引用格式: 肖子安, 张静, 苑馨方, 朱家佳, 李晓辉, 米琳, 窦帅. 基于多尺度特征提取和坐标注意力的光学遥感图像超分辨率重建. 计算机系统应用, 2023, 32(7): 105-112. <http://www.c-s-a.org.cn/1003-3254/9168.html>

## Super-resolution Reconstruction of Remote Sensing Images Based on Multi-scale Feature Extraction and Coordinate Attention

XIAO Zi-An<sup>1,2</sup>, ZHANG Jing<sup>1</sup>, YUAN Xin-Fang<sup>1</sup>, ZHU Jia-Jia<sup>1</sup>, LI Xiao-Hui<sup>1</sup>, MI Lin<sup>1</sup>, DOU Shuai<sup>1</sup>

<sup>1</sup>(Key Laboratory of Quantitative Remote Sensing Information Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China)

<sup>2</sup>(School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** Considering the problems caused by insufficient attention to receptive field scale and inadequate extraction of feature channel information in existing super-resolution reconstruction models for optical remote sensing images, this study proposes a new super-resolution reconstruction model for optical remote sensing images, which is based on multi-scale feature extraction and coordinate attention. On the basis of the deep residual network structure, some cascaded multi-scale feature & coordinate attention blocks (MFCABs) are designed in the high-frequency branch of the network to fully explore the high-frequency features of the input low-resolution images. Firstly, the Inception submodule is introduced into MFCABs to capture spatial features under different receptive fields by convolution kernels of different scales. Secondly, the coordinate attention submodule is added after the Inception submodule, and attention is paid to the

① 基金项目: 国家重点研发计划 (2021YFC3000302)

收稿时间: 2022-12-28; 修改时间: 2023-01-20, 2023-02-13; 采用时间: 2023-02-20; csa 在线出版时间: 2023-05-22

CNKI 网络首发时间: 2023-05-24

channel and coordinate dimensions to obtain a better channel attention effect. Finally, the features extracted by each MFCAB are fused in multiple paths to realize the effective fusion of multi-scale spatial information and multi-channel attention information. In the double and triple magnification of the MFCAB model on the NWPU4500 dataset, the PSNR reaches 34.73 dB and 30.12 dB, respectively, which is 0.66 dB and 0.01 dB higher than EDSR. In the double, triple, and quadruple magnification of the model on the AID1600 dataset, the PSNR reaches 34.71 dB, 30.58 dB, and 28.44 dB, respectively, which is 0.09 dB, 0.03 dB, and 0.04 dB higher than EDSR. The experimental results show that the reconstruction effect of this model on the optical remote sensing image datasets is better than the mainstream super-resolution image reconstruction model.

**Key words:** remote sensing image; super-resolution (SR) reconstruction; deep learning; attention mechanism; residual network (ResNet)

## 1 引言

光学遥感图像在测绘、地球资源调查、地质灾害调查与救治、生态环境监测等领域发挥着十分重要的作用<sup>[1]</sup>。高分辨率光学遥感图像包含了更为丰富的地物信息,对后续目标检测、识别、解译等遥感图像高级视觉任务的分析处理精度与效率的提升具有十分重要的意义。超分辨率 (super-resolution, SR) 重建技术通过软件方式提升图像的分辨率,能够在一定程度上规避硬件系统升级的弊端,如成本高、研制周期长等<sup>[2]</sup>,使图像的解译能力和应用效能得到进一步增强,具有极高的研究意义与实用价值。

Dong 等<sup>[3]</sup>将卷积神经网络 (convolutional neural network, CNN) 首次应用于自然图像的超分辨率重建,提出了 SRCNN (super-resolution CNN),使用卷积神经网络进行超分辨率重建。但是由于 SRCNN 中使用双三次插值的方法来放大图像,重建效果不够理想。因此,在 SRCNN 的基础上 Dong 等<sup>[4]</sup>提出了改进的模型 FSRCNN (fast SRCNN),引入了反卷积层对图像进行放大,并引入了更小卷积核和更多的卷积层,取得了更好的效果和更快的收敛速度。Shi 等<sup>[5]</sup>随后提出了 ESPCN (efficient sub-pixel convolutional neural network),采用亚像素卷积层代替反卷积层,获得了更好的重建效果。Kim 等<sup>[6]</sup>设计了包含 20 个卷积层的 VDSR (very deep super-resolution) 以获得更深层次的特征信息,并首次将残差网络 (residual networks, ResNet)<sup>[7]</sup>引入了超分辨率重建领域,有效解决了深度网络的梯度消失问题。Lim 等<sup>[8]</sup>针对超深网络结构进一步改进,提出了增强深度残差网络 (enhanced deep residual networks, EDSR)。EDSR 使用更多的卷积层提取特征,去掉批规范化 (batch

normalization, BN) 层以提升效率,采用 L1 范数作为损失函数,获得了比 VDSR 更好的重建效果。

光学遥感图像相比自然图像,除具有目标场景多、结构复杂、小目标密集分布的特点,还因其远距离俯视成像的特点,图像中的目标与背景耦合度高,难以区分。因此,应用于自然图像的超分辨率重建模型需要结合光学遥感图像的上述特点进行针对性优化与验证,才能获得适用于光学遥感图像的超分辨率重建模型,有效提升图像的重建效果与信息提取能力。Lei 等<sup>[9]</sup>基于 SRCNN 提出 LGCNet (local-global combined network),将网络提取的局部与全局特征联合起来进行超分辨率重建,获得了明显优于传统浅层机器学习方法的重建效果。李欣等<sup>[10]</sup>构建了包含 4 个卷积层的 CNN 模型,在前 3 层中添加参数修正线性单元,对局部响应归一化层进行优化,模型在 Pleiades 卫星影像及 AID 遥感数据集上的抗过拟合能力和训练效率均有所提高。李强等<sup>[11]</sup>基于生成对抗网络,针对遥感图像的特点,使用多尺度卷积级联增强全局特征,并利用多尺度感受野模块与密集残差模块为生成网络的细节特征提取获取了更多纹理信息。潘梦迪<sup>[12]</sup>针对重建图像缺乏细节信息的问题,提出了一种基于密集剩余残差网络的机载遥感图像超分辨率重建模型,密集和残差连接的引入使得信息能够在深度网络中更好的流动。印珏泽等<sup>[13]</sup>结合对偶回归约束映射空间和通道注意力机制,提出了基于对偶回归和残差注意力机制的图像超分辨率重建网络,实现了较好的超分辨率重建效果。上述研究从不同角度出發,将视觉领域深度学习及自然图像超分辨率重建的先进技术引入光学遥感图像超分辨率重建领域,进行了诸多有益的尝试,使光学遥感图

像超分辨率重建的研究不断深入,取得了很大进展.但是,坐标注意力机制在遥感图像超分辨率领域中的应用尚不广泛,现有模型多使用通道注意力机制,对感受野尺度和通道内的信息关注有限.

本文针对现有光学遥感图像超分辨率重建模型对感受野尺度关注不足和对特征通道信息提取不充分的问题,提出了一种基于多尺度特征提取和坐标注意力的光学遥感图像超分辨率重建模型.该重建模型基于深度残差网络结构,在网络的高频分支中设计了用于提取多尺度特征与坐标注意力的残差模块——MFCAB,利用更深的网络层次、融合多尺度空间特征与坐标注意特征的多级残差模块,对输入的低分辨率光学遥感图像的高频信息从空间、通道、位置等角度提取其显著性特征,对图像的高频结构与特征进行充分发掘;最后,对各残差模块提取的特征进行多路径融合,实现不同深度的MFCAB模块提取信息的有效融合.

## 2 深度学习图像超分辨率重建特征提取

### 2.1 多尺度特征提取

光学遥感图像具有目标尺度差异大、细节丰富、结构复杂、主体与背景区分不明显等特点,传统超分辨率重建模型采用单一尺度的卷积核提取图像目标空间分布特征,对图像高频结构与空间特征的发掘并不充分.针对这一问题,本文在残差网络设计中引入Inception结构<sup>[14]</sup>,如图1所示.Inception结构采用3个不同大小的卷积核和1个最大池化并,能够获得不同尺度感受野下的空间特征;在并行的分支中,首先通过1×1卷积进行维度调整,然后经过不同尺度的卷积核和池化操作,最后将提取的多尺度空间特征通过Concat操作特征融合,有利于网络在不同深度层次中获取更完整的目标空间结构特征描述.

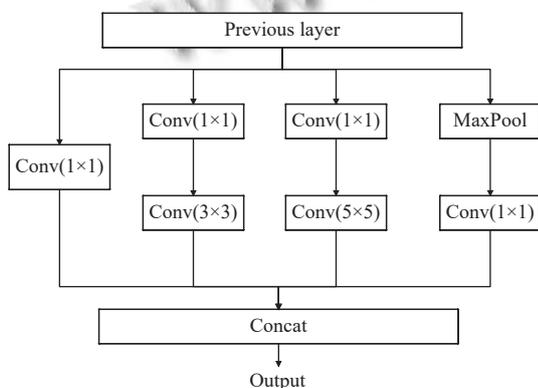


图1 Inception 结构描述

### 2.2 坐标注意力机制

2017年, Hu等<sup>[15]</sup>将通道注意力机制加入深度残差网络中,提出了SENet (squeeze-and-excitation networks),通过挤压-激励 (squeeze-and-excitation, SE) 操作,把每一通道特征图通过全局池化计算为一个标量作为该特征图的权重;而后通过全连接-Sigmoid 激活-全连接层,融合各个通道的特征信息,根据生成的通道权重张量赋予每个通道不同的权重. SENet对每一通道特征图进行全局池化,将其全局空间信息压缩成一个标量,但没有考虑通道特征图内部特征之间的位置关系和方向关系,这些信息的丢失可能会导致权重信息不准确、特征提取不充分. Hou等<sup>[16]</sup>在移动网络设计中提出了坐标注意力 (coordinate attention, CA) 机制,将同一通道特征图的位置信息嵌入到通道注意力提取过程中,能够保留特征的方向敏感和坐标敏感信息. CA机制网络结构如图2所示,主要分为两部分:坐标信息嵌入和坐标注意力生成.

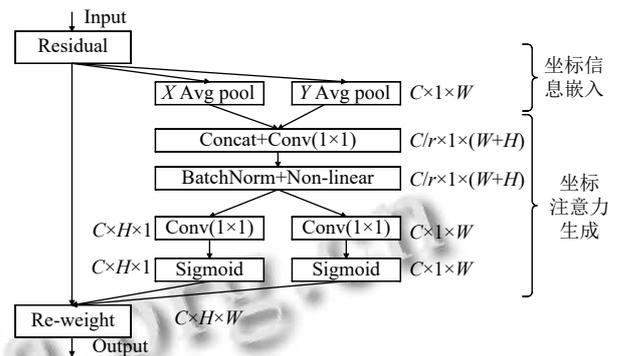


图2 CA 模块网络结构图

在坐标信息嵌入部分,对于H行×W列×C通道的输入特征,CA机制对其每一通道的特征图分别在水平和垂直方向并行使用1维平均池化核进行编码操作,可以表示为式(1).其中, $x_c(h, i)$ 和 $x_c(j, w)$ 分别表示第c通道特征图 $x_c$ 在高度为h处的水平方向特征和宽度为w处的垂直方向的特征; $z_c^w$ 和 $z_c^h$ 分别表示 $x_c$ 经过1维池化后水平方向和垂直方向的张量.

$$\begin{cases} z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \\ z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \end{cases} \quad (1)$$

在坐标注意力生成部分,CA机制首先将坐标信

息嵌入输出的聚合特征图进行 Concat 连接, 并使用  $1 \times 1$  卷积、BatchNorm (BN) 层和非线性激活 (non-linear) 层, 这一过程表示为式 (2):

$$f = \delta(\text{Conv}^{1 \times 1}([\mathbf{Z}^h, \mathbf{Z}^w])) \quad (2)$$

其中,  $\mathbf{Z}^h$  和  $\mathbf{Z}^w$  分别表示聚合的水平方向和垂直方向特征图,  $\delta(\cdot)$  为非线性函数,  $[\cdot]$  表示 Concat 操作,  $\text{Conv}^{1 \times 1}(\cdot)$  为  $1 \times 1$  卷积操作,  $f \in R^{C/r \times (H+W)}$  为水平方向和垂直方向编码生成的中间特征,  $R$  为实数空间,  $r$  为缩减因子, 主要用于减少计算参数量。

将中间特征  $f$  分割为两个独立的特征张量  $f^h \in R^{C/r \times H}$  和  $f^w \in R^{C/r \times W}$ 。然后将  $f^h$  和  $f^w$  分别经过  $1 \times 1$  卷积调整维度, 从  $C/r$  个通道恢复到  $C$  个通道, 再经过 Sigmoid 激活函数得到水平方向和垂直方向的注意力权重  $g^h$  和  $g^w$ , 表示为式 (3):

$$\begin{cases} g^h = \sigma(\text{Conv}^{1 \times 1}(f^h)) \\ g^w = \sigma(\text{Conv}^{1 \times 1}(f^w)) \end{cases} \quad (3)$$

其中,  $\sigma(\cdot)$  表示 Sigmoid 函数。

最后将权重  $g^h$  和  $g^w$  第  $c$  通道的特征张量  $g_c^h$  和  $g_c^w$  与  $x_c$  在对应的坐标位置相乘, 得到最终的赋予了通道

以及特征图内部位置方向注意信息的输出特征图  $y_c$ , 表示为式 (4):

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (4)$$

本文在残差模块中引入以上坐标注意力机制, 在获取特征图通道间权重信息的同时, 还考虑特征图内与方向相关的位置信息的影响, 从而可以更准确表达特征与通道注意力, 提取更准确的高频信息。

### 3 模型设计

#### 3.1 网络框架

本文设计的基于多尺度特征提取与坐标注意力机制的光学遥感图像超分辨率重建模型结构如图 3 所示。该模型采用深度残差网络架构, 分为低频和高频两个分支, 分别对输入的低分辨率光学遥感图像  $I_{LR}$  的低频信息与高频结构特征进行发掘与重构, 最后联合两路分支输出高分辨率重建图像  $I_{SR}$ 。其中, 低频分支包含 2 个部分: 浅层特征提取、上采样与重建; 高频分支包含 4 个部分: 浅层特征提取、多级特征提取、多路径融合、上采样与重建。

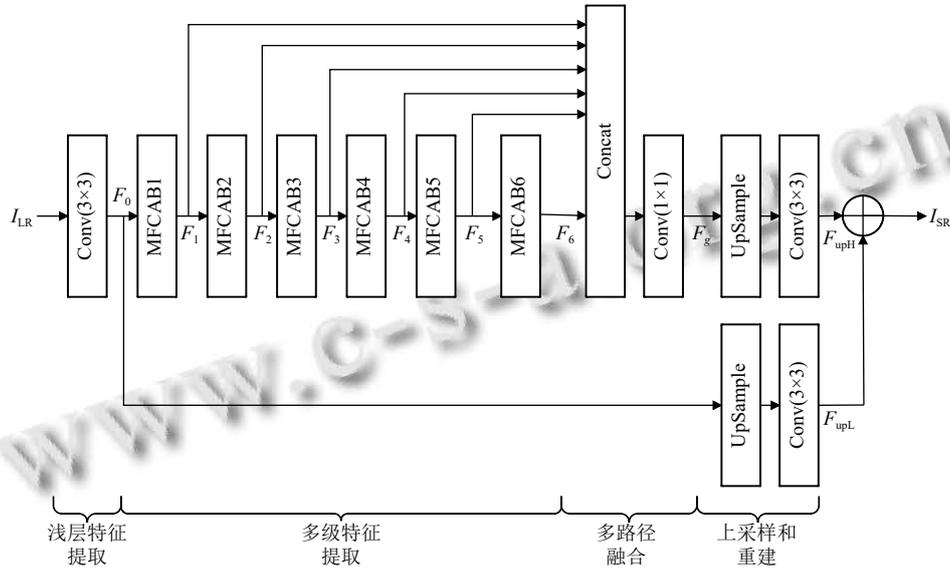


图 3 基于多尺度特征提取与坐标注意力机制的光学遥感图像超分辨率重建模型框架

#### 3.2 浅层特征提取

浅层特征提取将输入图像  $I_{LR}$  经过一个卷积层, 通过卷积生成特征图  $F_0$ 。该过程可以表示为式 (5)。其中,  $\text{Conv}^{3 \times 3}(\cdot)$  代表一个  $3 \times 3$  的卷积层。

$$F_0 = \text{Conv}^{3 \times 3}(I_{LR}) \quad (5)$$

#### 3.3 多级特征提取

光学遥感图像内容复杂且细节多, 为了更好地提取图像中的细节和低频特征信息, 本文基于 EDSR 的残差模块设计, 在模型高频分支设计了 6 个级联的多尺度特征提取和坐标注意力模块 MFCAB, 用于对高频

特征信息进行多级特征提取。

设模型中第  $x$  ( $x=1, 2, \dots, 6$ ) 个 MFCAB 的输出为  $F_x$ , 表示为:

$$F_x = MFCAB_x(\dots MFCAB_1(F_0)\dots) \quad (6)$$

其中,  $MFCAB_x(\cdot)$  代表第  $x$  个 MFCAB 模块的提取操作。

MFCAB 模块包含 5 个残差子模块 (residual block, RB)、1 个 Inception 子模块和 1 个 CA 子模块, 结构如图 4 所示。

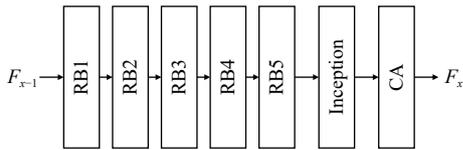


图 4 MFCAB 模块结构图

对于进入主网络高频分支的 MFCAB 模块的高频特征, 首先经过 5 个 RB 子模块进行特征提取, 然后经过 Inception 子模块进行多尺度空间特征提取与整合, 最后进入 CA 子模块, 获取包含特征位置信息的通道权重。该部分的操作表示为:

$$\begin{aligned} F_x &= MFCAB_x(F_{x-1}) \\ &= A(RB_5(\dots RB_1(F_{x-1})\dots)) \end{aligned} \quad (7)$$

其中,  $RB_i(\cdot)$  代表经过第  $i$  ( $i=1, 2, \dots, 5$ ) 个 RB 子模块,  $F_{x-1}$  和  $F_x$  分别代表第  $x$  个 MFCAB 模块的输入和输出,  $A(\cdot)$  代表经过 Inception 和 CA 子模块操作。

MFCAB 中的 RB 子模块由 1 个卷积层、1 个非线性激活层、1 个卷积层组成, 输入和输出有一条残差连接, RB 子模块结构如图 5 所示。

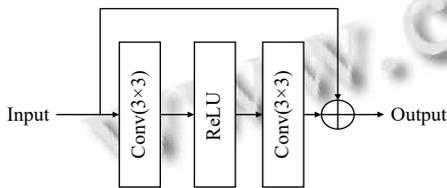


图 5 RB 模块结构图

在 Inception 子模块中 (如图 1), 特征图在并行的分支中经过不同尺度的卷积核或池化操作提取多尺度空间特征, 最后通过 Concat 操作进行特征融合。在 CA 子模块中, 首先进行坐标信息嵌入, 使用 1 维平均池化核分别沿垂直和水平方向将输入特征聚合为两个单独的方向感知特征张量; 然后进行坐标注意力生成, 将提取出的两个张量进行 Concat 连接, 再经过 BN 层

和非线性激活层, 最后将生成的张量分割为纵横方向的张量并展开, 为输入特征图的对应位置添加注意力权重。

本文 CA 模块核心伪代码如算法 1。

算法 1. CA 模块模型构建和前传函数

模型构建 def \_\_init\_\_():

输入: 输入特征通道数  $inC$ , 输出特征通道数  $outC$ , 缩减因子  $r$

输出: CA 模块结构

```
1. #分别添加纵横两个方向平均池化层
self.pool_h = nn.AdaptiveAvgPool2d((None, 1))
self.pool_w = nn.AdaptiveAvgPool2d((1, None))
2. #根据缩减因子 r 计算压缩后的通道数 tempC, 添加卷积层 conv1
3. #添加 BN 层和非线性激活层, 激活函数为 h_swish
4. #分别添加处理纵横两个分割后张量的卷积层 conv_h 和 conv_w
self.conv_h = nn.Conv2d(tempC, outC, kernel_size=1, stride=1, padding=0)
self.conv_w = nn.Conv2d(tempC, outC, kernel_size=1, stride=1, padding=0)
```

前传函数 def forward(self, x):

输入: 待计算的特征矩阵  $x$

输出: 计算后的特征矩阵  $out$

```
1. #获取输入特征矩阵维度 n, c, H, W
n, c, H, W = x.shape
2. #分别沿纵横两个方向进行池化, 获得张量 x_h 和 x_w
x_h = self.pool_h(x)
x_w = self.pool_w(x)
3. #调整 x_w 中的 H 和 W 的顺序, 方便后续进行 Concat 操作
x_w = x_w.permute(0, 1, 3, 2)
4. #对 x_h 和 x_w 进行 Concat 操作, 得到 x_cat
x_cat = torch.cat([x_h, x_w], dim=2)
5. #x_cat 通过 conv1 调整通道数, 再通过 BN 层和非线性激活层
x_cat = self.act1(self.bn1(self.conv1(x_cat)))
6. #对 x_cat 进行 split 操作, 分割为两个张量, 并将分割后的两个张量分别通过 conv_h 和 conv_w 恢复为原通道数, 而后再分别通过 Sigmoid 函数, 获得 out_h 和 out_w
x_h, x_w = torch.split(x_cat, [H, W], dim=2)
x_w = x_w.permute(0, 1, 3, 2)
out_h = torch.Sigmoid(self.conv_h(x_h))
out_w = torch.Sigmoid(self.conv_w(x_w))
7. #将原始输入的特征张量 x 和 out_w、out_h 相乘, 对原始输入的特征张量 x 的不同坐标位置赋权重
out = x * out_w * out_h
8. #返回结果
return out
```

### 3.4 多路径融合

多路径融合将高频分支中各个 MFCAB 模块输出的特征进行 Concat 合并与  $1 \times 1$  卷积降维, 在反向传播的过程中实现特征自适应融合学习, 提高特征注意力,

适应不同复杂程度的重建任务. 这部分的操作可以表示为:

$$F_g = \text{Conv}^{1 \times 1}(\text{Concat}(F_1, F_2, \dots, F_6)) \quad (8)$$

其中,  $\text{Concat}(\cdot)$  表示多路径特征合并操作,  $\text{Conv}^{1 \times 1}(\cdot)$  表示  $1 \times 1$  卷积操作,  $F_g$  表示最终的全局特征融合后的输出.  $F_1-F_6$  分别表示子 MFCAB 模块的输出.

### 3.5 上采样与重建

上采样是将低分辨率小尺寸的特征图重排为高分辨率大尺寸的特征图. 本文采用文献 [5] 提出的亚像素卷积方式进行上采样. 这个操作可以描述为:

$$F_{\text{upH}} = C_{\text{up1}}(F_g) \quad (9)$$

$$F_{\text{upL}} = C_{\text{up2}}(F_0) \quad (10)$$

其中,  $F_{\text{upH}}$  和  $F_{\text{upL}}$  分别表示上采样后输出的高频特征图和低频特征图;  $C_{\text{up1}}(\cdot)$  和  $C_{\text{up2}}(\cdot)$  分别表示高频分支与低频分支的上采样操作.

重建部分通过一个  $3 \times 3$  的卷积层将多通道特征图生成 3 通道 SR 图像, 最终的重建图像  $I_{\text{SR}}$  由低频重建图和低频重建图相加后获得, 如式 (11):

$$I_{\text{SR}} = \text{Conv}^{3 \times 3}(F_{\text{upH}}) + \text{Conv}^{3 \times 3}(F_{\text{upL}}) \quad (11)$$

## 4 实验验证及结果分析

### 4.1 损失函数

本模型使用平均绝对误差 (mean absolute error, MAE) 作为损失函数, 也被称为 L1 范数损失. 该损失函数计算重建图像的预测值  $f(x_i)$  和真实图像的  $y_i$  之间距离的均值. 损失函数在模型训练中不断减小, 从而获得更好的重建效果. MAE 损失函数如式 (12):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - f(x_i)| \quad (12)$$

图 6 为在 NWPU4500 数据集上训练过程中的损失函数收敛图, 放大因子  $\text{scale}=2$ . 从图中可以看出, 随着训练的进行, 损失函数不断降低, 在 300 个 epoch 处损失函数达到收敛.

### 4.2 实验环境和数据集

为分析和验证本文提出模型对于光学遥感图像超分辨率重建任务的适用性与有效性, 选用通用的 NWPU-RESISC45、AID 光学遥感图像数据集进行了消融实验和重建效果对比实验.

实验环境配置: GPU 为 Quadro RTX 8000 (48 GB), 软件环境为 Ubuntu 18.04, PyTorch 1.9.0, CUDA 11.1.

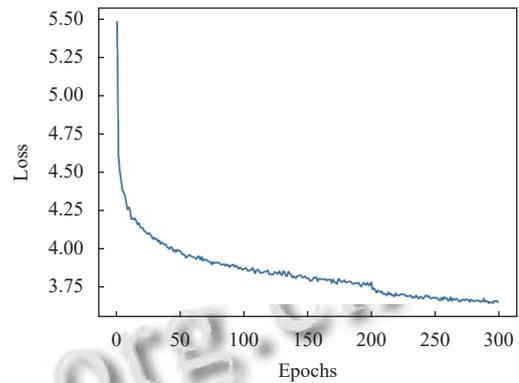


图 6 L1 损失函数收敛图

模型选择 L1 范数作为损失函数, 采用 ADAM 优化器, 参数为  $\text{beta1}=0.9$ ,  $\text{beta2}=0.999$ ,  $\text{epsilon}=1\text{E}-8$ ,  $\text{batch}=16$ , 初始学习率  $1\text{E}-4$ , 经过 200 个 epoch 学习率下降一半, 一共训练 300 个 epoch.

数据集方面, 选择通用的 NWPU-RESISC45 和 AID 数据集进行模型训练优化与测试. 在 NWPU-RESISC45 数据集的 45 个场景中分别各自随机选取 100 幅、4 幅、4 幅图像构成训练集 (4 500 幅)、验证集 (180 幅) 和测试集 (180 幅); AID 数据集采用相同方式构成训练集 (1 600 幅)、测试集 (200 幅)、验证集 (200 幅). 下文分别以 NWPU4500 和 AID1600 代指这两个数据集. 在训练时, 对训练集图像进行随机的  $90^\circ$  旋转、水平翻转、垂直翻转来获得更多的训练数据, 训练时输入样本尺寸为  $48 \text{ pixel} \times 48 \text{ pixel}$ .

超分辨率重建图像的客观评价指标采用通用的峰值信噪比 (peak signal to noise ratio, PSNR) 和结构相似度 (structural similarity, SSIM).

### 4.3 消融实验

为了验证本文模型的有效性, 在 NWPU4500 数据集上进行了消融实验 (放大因子  $\text{scale}=2$ ), 测试添加 Inception 子模块和 CA 子模块对超分辨率重建的有效性. 消融实验结果如表 1 所示.

从表 1 可知, 单独添加 Inception 或 CA 子模块时, PSNR 和 SSIM 均有所提高, 表明采用多尺度特征提取或坐标注意力机制对提升重建图像质量是有效的. 当同时使用 Inception 和 CA 子模块时, PSNR 和 SSIM 均较不添加或单独添加时提升更为明显, 表明多尺度空间特征提取与坐标注意力机制结合使用, 对于高频特

征提取能起到相互促进的效果, 将使光学遥感图像的超分辨率重建效果最优。

表1 消融实验

Inception	CA	PSNR (dB)	SSIM
×	×	34.35	0.943
√	×	34.46	0.944
×	√	34.39	0.944
√	√	<b>34.73</b>	<b>0.947</b>

#### 4.4 超分辨率重建效果对比

表2给出了本文模型与 Bicubic、FSRCNN、VDSR、EDSR 等主流的超分辨率重建模型在 NWPU4500 和 AID1600 遥感图像数据集上的超分辨率重建效果 (放大因子 scale 分别为 2、3、4) 对比。

从表中可以看出, 深度学习模型的超分辨率重建效果均显著优于传统的双三次插值模型 Bicubic, 显示出深度学习模型在光学遥感图像超分辨率重建方面的巨大潜力。而 EDSR 与本文模型由于采用了更深和更宽的网络结构, 超分辨率重建图像的质量较 VDSR、SRCNN 提升更为明显。由于在网络的高频分支使用了多尺度特征提取和坐标注意力机制, 除了在 NWPU4500 数据集上的 4 倍放大因子下指标略低, 本文模型在 2 个数据集上 3 种放大因子下的超分辨率重建质量均

优于 EDSR, 显示出本文模型对于光学遥感图像超分辨率重建具有更好的适用性。由于 NWPU4500 数据集的图像初始尺寸为  $256 \times 256$ , 在  $scale=4$  时下采样后图像尺寸仅为  $64 \times 64$ , 而所需训练样本尺寸为  $48 \times 48$ , 产生的样本变化有限, 可能限制了重建效果的提升。

表2 超分辨率重建效果对比

数据集	scale	指标	Bicubic	FSRCNN	VDSR	EDSR	本文模型
NWPU4500	×2	PSNR (dB)	30.82	32.51	32.85	34.07	<b>34.73</b>
		SSIM	0.876	0.916	0.921	0.940	<b>0.947</b>
	×3	PSNR (dB)	27.78	29.16	29.36	30.11	<b>30.12</b>
		SSIM	0.764	0.823	0.829	<b>0.857</b>	<b>0.857</b>
	×4	PSNR (dB)	26.12	27.03	27.29	<b>28.05</b>	28.04
		SSIM	0.676	0.724	0.736	<b>0.785</b>	0.784
AID1600	×2	PSNR (dB)	31.53	33.30	33.63	34.62	<b>34.71</b>
		SSIM	0.882	0.922	0.926	0.939	<b>0.940</b>
	×3	PSNR (dB)	28.25	29.60	29.81	30.55	<b>30.58</b>
		SSIM	0.769	0.826	0.830	0.859	<b>0.860</b>
	×4	PSNR (dB)	26.52	27.45	27.62	28.40	<b>28.44</b>
		SSIM	0.682	0.733	0.736	0.787	<b>0.788</b>

#### 4.5 视觉效果对比

除通过客观量化评价指标分析本文模型的重建效果, 图7列出了本文模型与其他典型模型重建图像的视觉效果对比。因篇幅所限, 仅列举出各模型对 NWPU4500 数据集中一幅图像, 放大因子  $scale=4$  时的重建效果。

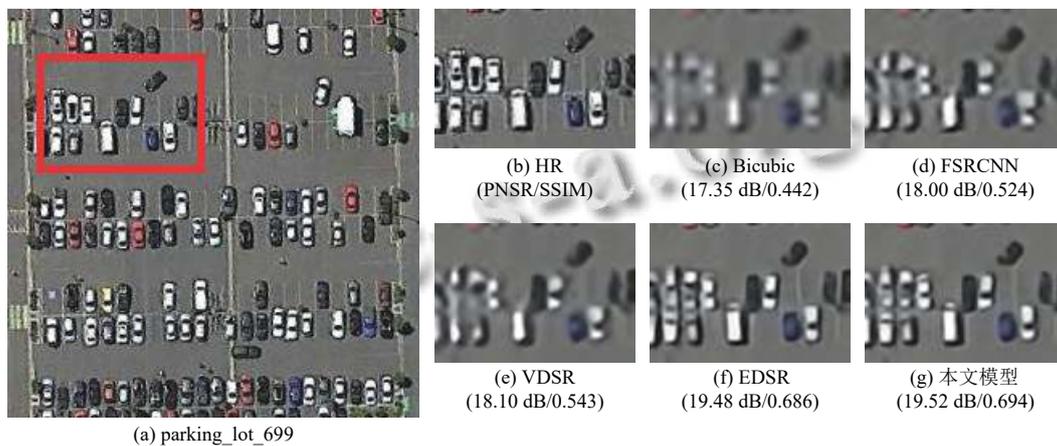


图7 视觉效果对比

## 5 结论与展望

本文提出了一种基于多尺度特征提取和坐标注意力的光学遥感图像超分辨率重建模型, 在深度残差网络的高频分支引入多尺度特征提取和通道注意力机制构建了 MFCAB 模块, 实现对通道特征图进行不同尺度的空间域信息提取, 并通过对通道中坐标的权重关

系进行自适应学习, 能够更好地适应遥感图像物体跨度大、细节信息多的特点, 更好地提取高频信息, 对更重要的特征进行关注; 最后将多个 MFCAB 提取的特征进行多路径融合, 加强了各个网络块之间的联系, 充分利用特征信息。

本文模型在 NWPU4500 和 AID1600 两个光学遥

感图像数据集上进行了测试与验证,实验结果显示出模型的重建效果优于主流的图像超分辨率重建模型。后续将重点针对大放大因子(scale=4)下的模型优化、模型参量优化、模型泛化验证等方面开展进一步研究和测试。

### 参考文献

- 1 王润生,熊盛青,聂洪峰,等. 遥感地质勘查技术与应用研究. 地质学报, 2011, 85(11): 1699–1743.
- 2 李胜楠. 基于深度学习的图像超分辨率算法研究 [硕士学位论文]. 秦皇岛: 燕山大学, 2021. [doi: 10.27440/d.cnki.gysdu.2021.001318]
- 3 Dong C, Loy CC, He KM, *et al.* Learning a deep convolutional network for image super-resolution. Proceedings of the 13th European Conference on Computer Vision. Zurich: Springer, 2014. 184–199.
- 4 Dong C, Loy CC, Tang XO. Accelerating the super-resolution convolutional neural network. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 391–407.
- 5 Shi WZ, Caballero J, Huszár F, *et al.* Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 1874–1883.
- 6 Kim J, Lee JK, Lee KM. Accurate image super-resolution using very deep convolutional networks. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 1646–1654.
- 7 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.
- 8 Lim B, Son S, Kim H, *et al.* Enhanced deep residual networks for single image super-resolution. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu: IEEE, 2017. 1132–1140.
- 9 Lei S, Shi ZW, Zou ZX. Super-resolution for remote sensing images via local-global combined network. IEEE Geoscience and Remote Sensing Letters, 2017, 14(8): 1243–1247. [doi: 10.1109/LGRS.2017.2704122]
- 10 李欣. 基于深度学习的单幅遥感图像超分辨重建 [硕士学位论文]. 北京: 中国科学院大学(中国科学院遥感与数字地球研究所), 2018.
- 11 李强,汪西原,何佳玮. 基于生成对抗网络的遥感图像超分辨率重建改进算法. 激光与光电子学进展, 2023, 60(10): 432–439.
- 12 潘梦迪. 基于深度学习的机载遥感图像超分辨率重建 [硕士学位论文]. 哈尔滨: 哈尔滨工程大学, 2020.
- 13 印瑛泽,周宁宁. 基于对偶回归和注意力机制的图像超分辨率重建网络. 计算机系统应用, 2023, 32(2): 111–118. [doi: 10.15888/j.cnki.csa.008939]
- 14 Szegedy C, Liu W, Jia YQ, *et al.* Going deeper with convolutions. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 1–9.
- 15 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City: IEEE, 2018. 7132–7141.
- 16 Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13708–13717.

(校对责编:牛欣悦)