



小型微型计算机系统

Journal of Chinese Computer Systems

ISSN 1000-1220, CN 21-1106/TP

《小型微型计算机系统》网络首发论文

题目: 融合多路径与混合注意力的遥感图像超分辨率重建
作者: 李千, 赵逢禹
DOI: 10.20009/j.cnki.21-1106/TP.2021-0818
收稿日期: 2021-10-18
网络首发日期: 2022-05-13
引用格式: 李千, 赵逢禹. 融合多路径与混合注意力的遥感图像超分辨率重建[J/OL]. 小型微型计算机系统. <https://doi.org/10.20009/j.cnki.21-1106/TP.2021-0818>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

融合多路径与混合注意力的遥感图像超分辨率重建

李 千, 赵逢禹

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

E-mail: 1029988609@qq.com

摘 要: 针对遥感图像细节信息丰富、内容复杂、部分特征信息重建丢失以及现有遥感图像重建模型对感受野尺度关注不足带来的问题, 本文提出一种融合多路径与混合注意力机制的算法用于遥感图像的超分辨率重建。该算法首先利用浅层特征提取模块提取浅层特征图; 然后通过多层级特征提取模块进行深层特征的自适应学习; 接着使用全局特征融合模块进行多层级特征融合和全局残差学习; 最后经上采样模块和重建模块得到高分辨率图像。算法重点关注各层级特征提取和全局特征的融合, 并通过横向扩展网络以便于利用不同的卷积核来获得不同感受野, 同时结合混合注意力机制来关注高频信息, 最终取得不错的重建效果。实验表明, 该算法在峰值信噪比和结构相似度两个评价指标上相对于主流算法均有提升, 视觉效果上也能够恢复更多的细节信息。

关键词: 遥感图像; 超分辨率重建; 特征融合; 注意力机制; 卷积神经网络

中图分类号: TP391

文献标识码: A

Super-Resolution of Remote Sensing Images Based on Multi-path and Mixed Attention

LI Qian, ZHAO Feng-yu

(School of Optical-Electrical & Computer Engineering, University of Shanghai for Science & Technology, Shanghai 200093, China)

Abstract: Aiming at the problems caused by rich detail information, complex content, loss of partial feature information and insufficient attention to receptive field scale in existing remote sensing image reconstruction models, an algorithm integrating multi-path and mixed attention mechanism is proposed for super-resolution reconstruction of remote sensing images. Firstly, the shallow feature extraction module is used to extract the shallow feature map; Then, the multi-level feature extraction module is used for adaptive learning of deep features; Then, the global feature fusion module is used for multi-level feature fusion and global residual learning; Finally, the high-resolution image is obtained by up sampling module and reconstruction module. The algorithm focuses on the feature extraction at all levels and the fusion of global features, and expands the network horizontally to obtain different receptive fields by using different convolution kernels. At the same time, it focuses on high-frequency information by combining the mixed attention mechanism, and finally achieves good reconstruction results. Experiments show that the algorithm improves the two evaluation indexes of peak signal-to-noise ratio and structural similarity compared with the mainstream algorithms, and can recover more detailed information in visual effect.

Key words: remote sensing image; super-resolution reconstruction; feature fusion; attention; convolutional neural network

1 引言

遥感图像作为地表信息的关键数据, 不管是在民用领域还是军用领域都有着重要价值。高分辨率(High-Resolution, HR)的遥感图像不但具有丰富而细致的纹理细节, 而且其包含的关键信息在目标检测、物体识别、土地覆盖分类甚至是国防安全等工作中都有着无可替代的作用。然而, 由于图像传输条件和成像设备的限制^[1], 所获得的遥感图像通常是低分辨率(Low-Resolution, LR)图像, 并且通过提升硬件来增加相应分辨率需要付出大量时间成本和经济成本。因此, 采用超分辨率(Super-Resolution, SR)重建技术从软件的角度来

增强遥感图像的分辨率具有重要的意义。

超分辨率重建技术的通常定义是指在不改变硬件条件下, 结合一定的先验信息, 将某个给定场景的单张或者序列 LR 图像, 利用一定的图像处理方法, 重构得到该场景的 HR 图像^[2]。随着计算机视觉的发展, 卷积神经网络(Convolutional Neural Network, CNN)在自然图像超分辨率重建领域展现出巨大潜力。

Dong 等人首次将 CNN 应用在自然图像超分辨率重建任务上并提出了 SRCNN 算法^[3], 利用三层卷积完成图像特征提取, 相较于传统的方法获得了良好的重建效果。但由于

在图像特征提取操作前需要进行插值放大,增加了计算量,同时,插值过程引入的误差也会影响重构效果。针对插值放大带来的问题,Dong 等人进一步提出了 FSRCNN 算法^[4]。该算法利用反卷积层解决了插值问题,同时还增加网络深度和调整卷积核大小,获得了更好的重构效果和训练速度。为了提升上采样的效果,Shi 等人提出了 ESPCN 算法^[5],该算法直接在 LR 图像上进行卷积操作提取 LR 图像的特征,上采样则是通过亚像素的方式将 LR 空间中的特征扩展到 HR 空间,最后将卷积层得到的特征进行像素排列获得重建图像。网络中利用亚像素卷积层替代反卷积层,并证明是比反卷积更有效的上采样方式。Kim 等人发现深层网络比浅层网络的重建效果更好,并基于此提出了 VDSR 算法^[6],该算法构建了非常深的卷积神经网络以获得更大的感受野和更多的特征信息,并率先引入残差结构来缓解深度网络的梯度消失问题。Li 等人基于反馈机制提出了 SRFBN 算法^[7],该算法通过多次反馈连接不仅能达到加深网络的效果,还能实现参数的复用。国内学者程德强等人^[8]为了平衡网络深度和训练难度之间的关系,设计了一种多通道递归残差网络模型,利用递归方法复用残差网络块,并通过通道学习和交叉学习机制促进特征信息融合、提高网络重建性能。

上述基于 CNN 的方法在图像超分辨率重建任务上取得了很大的进步,但却没有考虑到特征之间的相关性。为了学习不同通道之间的特征相关性,Zhang 等人提出了 RCAN 算法^[9],该算法在超分辨率重建任务中引入通道注意力机制,提升了网络对有用信道的关注度,增强了网络的辨别学习能力。国内学者施举鹏等人^[10]提出了一种基于深度反馈注意

力的超分辨率算法,该算法以交替上下采样和通道注意力为基础,能够实现对冗余特征的过滤和提高参数的复用。此外,Li 等人提出 ADRD 算法^[11],该算法通过空间注意力模块和 DenseNet 来完成重建任务,加强了对空间特征相关性的关注。尽管这些基于注意力的方法在超分辨率重建任务上取得了显著的改善,但他们仍然有一些不足之处。

首先,图像的低频信息(平滑区域)容易恢复,而高频信息(边缘和轮廓)很难恢复。现有的大部分基于 CNN 的方法^[3-8]都对不同的空间区域进行平等处理,因此阻碍了高频信息的恢复。此外,融合多尺度的特征能够为重建出高质量细节提供更丰富的信息^[12]。虽然有些研究^{[13][14]}也考虑融合多尺度的特征,但对不同通道和不同空间区域上特征之间的相关性缺少关注,丢失了部分重要信息。

针对上述分析以及遥感图像超分辨率重建存在的问题,本文提出一种融合多路径与混合注意力的遥感图像超分辨率重建算法。算法以多层级特征提取及特征融合为网络主体结构,该结构可充分融合各级子网络输出的特征内容进行自适应地学习,增强特征的利用率以缓解图像中小目标特征丢失的问题。设计多路径融合混合注意力网络单元,通过横向扩展网络,利用不同尺度的卷积核获得不同的感受野,以通过这些并行处理获得更好的图像表征。在该网络单元中,引入一种混合注意力机制来学习特征之间不同的依赖关系,该混合注意力机制由空间注意力模块和区域级通道注意力模块组成。除了能够关注到空间特征以外,还能对每个通道上的不同区域计算注意力,以加强高频信息的重建。

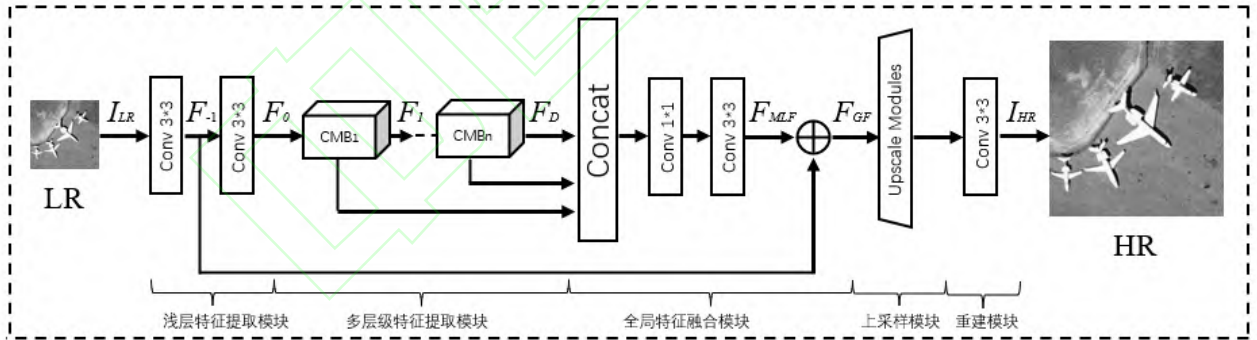


图1 网络总体结构

Fig.1 Overall Network Structure

2 网络结构

如图1所示,本文算法的网络结构包括五个模块,它们分别是浅层特征提取模块(Shallow Feature Extract Module, SFE)、多层级特征提取模块(Multi-level Feature Extraction Module, MFE)、全局特征融合模块(Global Feature Fusion Module, GFF)、上采样模块和重建模块。

2.1 浅层特征提取模块

浅层特征提取模块用于提取输入 LR 遥感图像的浅层特征。网络中使用两个 3×3 的卷积层来提取浅层特征。第一个卷积层从输入的 LR 中提取出特征 F_{-1} ,并将特征图扩展到 64 通道,这个操作可以描述为:

$$F_{-1} = C_{SFE1}(I_{LR}) \quad (1)$$

其中, $C_{SFE1}(\cdot)$ 为卷积操作,代表第一次浅层特征提取, I_{LR} 代表模型输入的低分辨率遥感图像。 F_{-1} 将在后续用于第二次浅层特征提取操作和全局的残差学习。第二个浅层特征提取操作可以描述为:

$$F_0 = C_{SFE2}(F_{-1}) \quad (2)$$

其中, $C_{SFE2}(\cdot)$ 代表第二次浅层特征提取的卷积操作,它的输出 F_0 将作为多层级特征提取模块(MFE)的输入。

2.2 多层级特征提取模块

多层级特征提取模块将进行深层特征的提取,该模块由

多个串行连接的级联多路径网络残差块(Cascaded Multi-path network residual Block, CMB)组成。假设模型中包含 D 个 CMB 块, 则第 i 个 CMB 块的输出可以描述为:

$$F_i = CMB_i(F_{i-1}) \quad (3)$$

其中 $i=1, \dots, D$, $CMB_i(\cdot)$ 代表第 i 个 CMB 块的特征提取操作。

如图 2 所示, 每个 CMB 块包含一系列的组合操作。首先连接若干个多路径混合注意力网络单元(Multi-path fusion mixed Attention Network unit, MAN), 然后经过一个 1×1 卷积进行特征压缩, 再经过注意力机制来加强高频信息的重建, 最后是一个残差运算。这部分的操作可以描述为:

$$\begin{aligned} F_i &= F_{i-1} \\ &+ A \left(C^{1 \times 1} \left(H_{MAN,n} \left(\dots \left(H_{MAN,1}(F_{i-1}) \right) \dots \right) \right) \right) \end{aligned} \quad (4)$$

其中, F_{i-1} 和 F_i 分别代表单个 CMB 块的输入和输出, $H_{MAN,i}(\cdot)$ 代表第 i 个 MAN 网络单元($i=1, \dots, D$), $C^{1 \times 1}$ 代表 1×1 的卷积层, A 代表混合注意力机制。

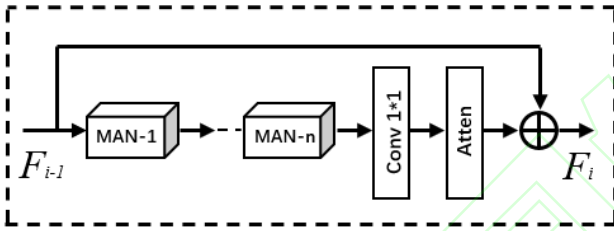


图 2 级联多路径残差块

Fig.2 Cascaded Multi-path network residual Block

2.3 全局特征融合模块

全局特征融合模块可充分利用各层级子网络输出的特征进行自适应学习, 提高特征利用率。该模块包含多层级特征融合(Multi-Level Feature Fusion, MLFF)和全局残差学习。

多层级特征融合(MLFF)首先将多个级联 CMB 块对应的层级特征抽取出来, 然后用 Concat 操作将这些特征图的通道拼接起来。特征融合操作由一个 1×1 卷积层和一个 3×3 卷积层完成, 1×1 的卷积层用于自适应地融合不同层级的特征, 并引入 3×3 的卷积层进一步抽取特征 F_{MLF} 用于后续的全局残差学习^[15]。这部分的操作可以描述为:

$$F_{MLF} = C^{3 \times 3}(C^{1 \times 1}(\text{Concat}(F_1, \dots, F_D))) \quad (5)$$

其中, $\text{Concat}(\cdot)$ 代表将特征图按照通道拼接, $C^{1 \times 1}$ 代表用于特征融合的 1×1 卷积层, $C^{3 \times 3}$ 代表特征抽取的 3×3 卷积层, F_{MLF} 代表多层级特征融合的输出。

全局残差学习用于获取最后用于上采样的特征图, 这个操作可以描述为:

$$F_{GF} = F_{MLF} + F_{-1} \quad (6)$$

其中, F_{-1} 代表提取的浅层特征图, F_{MLF} 为式 (5) 的输出, 其融合本文中级联 CMB 块的所有特征。

2.4 上采样模块

上采样模块会将低尺寸的特征图扩展至高尺寸的特征图, 本模块采用的上采样方法是基于 Shi 等人^[5]提出的亚像素算法实现。这个操作可以描述为:

$$F_{up} = C_{up}(F_{GF}) \quad (7)$$

其中, F_{GF} 代表全局特征, $C_{up}(\cdot)$ 代表上采样操作, F_{up} 代表上采样后的输出特征图。

2.5 重建模块

重建模块使用一个 3×3 的卷积层将上采样后学习到的特征图整合并形成最终的 HR 图像, 同时会将特征图的通道数恢复为 3。这个操作可以描述为:

$$I_{HR} = C_R(F_{up}) \quad (8)$$

其中, I_{HR} 代表最后模型输出的高分辨率遥感图像, $C_R(\cdot)$ 代表一个 3×3 的卷积层。

3 算法细节

本章将详细介绍在 CMB 块中出现的 MAN 网络单元, 以及在 CMB 块和 MAN 网络单元中均使用到的混合注意力机制。

3.1 多路径融合混合注意力网络单元

如图 3 所示, 多路径融合混合注意力网络单元(MAN)是一种多路网络结构, 该结构类似 Szegedy 等人^[16]在 GoogleNet 中提出的 Inception 结构。在这种结构中, 会对输入信息并行地执行多个卷积运算, 然后将所有结果拼接为一个特征图。因为大小不同的卷积核在图像中获取到感受野大小也不同, 所以通过这种多路结构能让网络自适应地学习以选择合适的卷积核, 从而提取到更关键的特征。

除了设计多路网络结构以外, 本文还加强了不同路径之间特征的复用, 这是 Inception 结构中没有考虑到的地方, 并且引入了混合注意力机制来关注不同通道上和不同空间上特征之间的相关性, 加强高频信息的重建。

MAN 网络单元主要包含三部分。第一部分为包含两个 (Conv 3×3 、Conv 5×5) 三层链式卷积的浅层多次提取操作, 第二部分为包含两个三层 (Conv 1×1 、Conv 3×3) 链式卷积的深层多次提取操作, 第三部分为特征图的拼接和压缩操作。模型将两个短跳连接加入到 MAN 网络单元中, 这能增加不同尺度特征的复用和扩大网络的感受野。此外, 还通过在

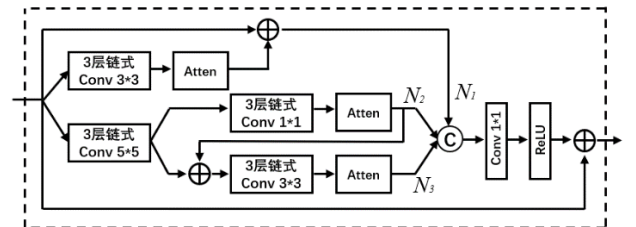


图 3 多路径融合混合注意力网络单元

Fig.3 Multi-path fusion mixed Attention Network unit

MAN 网络单元中设计长跳连接来缓解网络深度过深导致的梯度消失问题并提高特征的传递性能和利用效率。

当特征图经过浅层多次提取和深层多次提取后, 首先

用 Concat 将三个特征图的通道拼接起来(图 3 中 N_1 、 N_2 、 N_3)，再用 1×1 卷积将通道恢复至输入时的数量。为缓解网络过深导致梯度消失的问题和提高特征利用率，经 ReLU 激活函数后利用跳连接将输入前的特征和计算得到的特征进行残差运算，最终得到 MAN 网络单元的输出。

3.2 混合注意力机制

本节将介绍在 CMB 块和 MAN 网络单元中出现的混合注意力机制。由于遥感图像内容复杂、细节较多，为了能够获得更好的重建效果，本文在 RCAN^[9]中提出的通道注意力的基础上，结合深度可分离卷积(Depthwise separable Convolution, DC)，设计了一种混合注意力机制，用于充分学习特征图在空间上和通道上的特征依赖关系。深度可分离卷积首次在 Xception 模型^[17]中被提出，由于一个卷积核负责一个通道，一个通道只被一个卷积核卷积，故深度可分离卷积可以学习特征图的每个通道上的空间依赖关系，专注于空间特征的提取。

目前主流的通道注意力都是通过池化将特征图归一至一个点，这将导致丢失很多信息。因为就算是特征图的同一个通道，不同的位置获取的特征也是不同的，特别是对内容跨度大的遥感图像更是如此。考虑到这些不足，本文中采用

一种区域级的通道注意力，通过将输入特征图在空间维度上分成 $k \times k$ 个区域，这样就能单独对各个区域上的通道分配不同的注意力。区域级通道注意力通过对权重信息进行扩展能够更全面地表示特征信息，加强遥感图像高频信息的重建。此外，本文设计的混合注意力机制中利用最大池化和均值池化替代了 RCAN 中的单一池化，使用两种不同的池化方式能够更全面地聚合区域特征信息^[18]。

混合注意力机制如图 4 所示，包含空间注意力和区域级通道注意力两个部分。首先选用 5×5 的深度可分离卷积提取每个通道空间上的特征的依赖关系，在对特征图进行空间特征提取之后，经过 ReLU 激活后进入通道注意力部分操作。在通道注意力中分别使用 Mean 池化和 Max 池化将特征图的通道分成 $k \times k$ 个区域，然后将对应通道的值相加。图 4 中 W_d 代表降维的 1×1 卷积， W_u 代表升维的 1×1 卷积，其作用是完成通道间的复杂依赖拟合，在进行降维卷积 W_d 和升维卷积 W_u 之后分别用 ReLU 和 Sigmoid 对特征图进行操作。输出的注意力与输入特征图相乘之前，需要将输出的注意力通过最近邻插值操作上采样到与输入相同的尺寸，以完成对输入特征图上各个区域重点信息的关注。

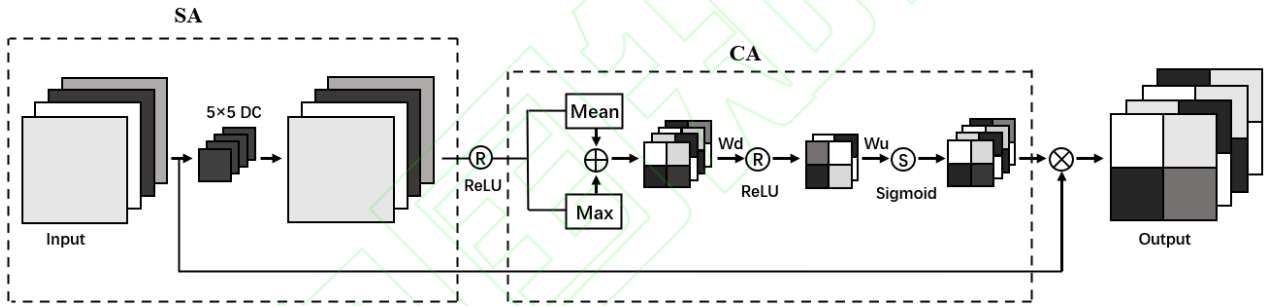


图 4 混合注意力机制

Fig.4 Mixed Attention Mechanism

4 实验

4.1 实验设置

为验证本文算法在遥感图像超分辨率重建任务上的有效性，基于 Pytorch 深度学习框架搭建了本文提出的网络模型。随后在 Google dataset of SIRI-WHU_earth、UCMerced、NWPU-RESISC45 以及 Christchurch Aerial Semantic Dataset 这 4 个数据集上将本文算法实验结果(MAN)和 Bicubic、FSRCNN^[4]、ESPCN^[5]、VDSR^[6]、SRFBN^[7]、RCAN^[9]、IRN^[19]这 7 种主流算法的实验结果进行比较。评价标准采用峰值信噪比(Peak Signal to Noise Ratio, PSNR)和结构相似度(Structural Similarity, SSIM)。

实验测试的硬件环境为搭载 AMD R-7 5800H 3.2GHz×8 处理器，内存为 32G，GPU 为 NVIDIA GeForce GTX 3060 6GB 的计算机。软件环境为 Ubuntu 16.04，PyCharm 开发环境，CUDA Toolkit 10.1。

本模型选择 ADAM 算法进行优化，优化器参数

$\beta_1=0.9$ ， $\beta_2=0.999$ ， $\epsilon=10^{-8}$ 。初始学习率为 10^{-4} ，每训练 100 个 epoch 学习率下降一半，共训练 300 个 epoch。经实验比较，CMB 的个数设置为 8，MAN 个数设置为 4 时效果较好。选择 L_1 ^{[20][21]} 为损失函数，其通过计算本模型输出图像与真实高分辨率图像之间的损失值来训练网络模型中的参数，模型训练目标为最小化损失函数：

$$L_1(\theta) = \frac{1}{N} \sum_{i=1}^N \|H_{Model}(I_{LR}^i) - I_{HR}^i\|_1 \quad (9)$$

其中， θ 表示整个网络中需要学习的所有参数， N 则代表每个 batch 中的图像数目。

4.2 数据集设置

数据集方面，Google dataset of SIRI-WHU_earth 中包含 2400 张遥感图像，UCMerced 中包含 1700 张遥感图像，NWPU-RESISC45 包含 800 张遥感图像，Christchurch Aerial Semantic Dataset 中包含 24 张超清的航拍图像。对于前三个数据集，按照 8: 1: 1 的比例划分训练集、验证集和测试集，

并通过将图像随机选择 90°、180°、270°和水平翻转来获得更多的训练数据。

对于 Christchurch Aerial Semantic Dataset 来说, 该数据中包含 4800×3600、4800×6624、6809×3600、6809×6624 这四种大小的分辨率。本文将这些图片沿水平和垂直方向以 406 为步长, 切割为 812×812 大小的图像, 整理得到 795 张图像, 并按照前文所述图像扩充方式和数据集划分方式进行处理。

最后, 本文使用双三次插值法将高分辨率图像分别进行 2 倍、3 倍和 4 倍下采样, 并将得到的低分辨率图像和对应的高分辨率图像构建为训练所需的数据格式。

4.3 消融实验分析混合注意力机制的有效性

为了验证上文中混合注意力机制的有效性, 本文进行了消融对比实验。在 MAN 网络单元中设计进行 3 种注意力结构的比较, 训练 100 个 epoch, 并基于 Christchurch Aerial Semantic Dataset 数据集比较 PSNR 值。

结构 A, 本文提出的混合注意力机制。

结构 B, RCAN 中的通道注意力机制。

结构 C, 将本文混合注意力机制中的区域级通道注意力替换为 RCAN 中的通道注意力, 其它部分不变。

消融实验结果如图 5 所示, 结构 B 和结构 C 的对比可以得出, 在通道注意力的基础上添加空间注意力能够提升遥感图像重建的效果。结构 A 和结构 C 的对比则得出, 在混合注意力中使用区域级的通道注意力重建效果更好。

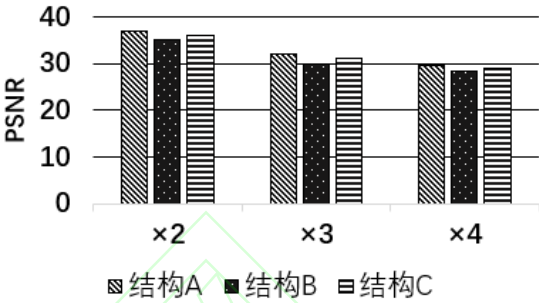


图 5 消融实验结果

Fig.5 Result of the ablation experiment

表 1 本文算法和其它算法的对比分析

Table 1 The comparative analysis of the algorithm in this paper and other algorithm

Scale	Method	Google dataset of SIRI-WHU_earth PSNR/SSIM	UCMerced PSNR/SSIM	NWPU-RESISC45 PSNR/SSIM	Christchurch Aerial Data PSNR/SSIM
×2	Bicubic	33.15/0.8891	32.04/0.9035	31.62/0.8798	34.77/0.9454
×2	FSRCNN	34.21/0.9125	34.86/0.9165	33.41/0.9177	37.73/0.9561
×2	ESPCN	34.45/0.9166	35.21/0.9253	33.66/0.9201	38.19/0.9413
×2	VDSR	34.71/0.9212	35.41/0.9314	33.81/0.9231	38.50/0.9661
×2	SRFBN	34.66/0.9186	35.38/0.9377	33.78/0.9217	38.94/0.9641
×2	RCAN	35.08/0.9257	35.83/0.9413	34.14/0.9275	39.13/0.9655
×2	IRN	35.01/0.9211	35.72/0.9423	33.89/0.9236	38.87/0.9591
×2	MAN(Ours)	35.14/0.9271	35.92/0.9455	34.21/0.9282	39.21/0.9687
×3	Bicubic	31.85/0.8231	29.15/0.8155	28.84/0.7839	31.74/0.8816
×3	FSRCNN	32.82/0.8498	31.08/0.8297	30.01/0.8155	32.60/0.9080
×3	ESPCN	33.05/0.8651	31.34/0.8407	30.14/0.8193	32.77/0.9244
×3	VDSR	33.17/0.8683	31.49/0.8483	30.56/0.8254	33.03/0.9242
×3	SRFBN	33.56/0.8694	31.55/0.8603	30.36/0.8199	34.11/0.8289
×3	RCAN	33.75/0.8749	31.88/0.8744	30.68/0.8390	34.15/0.9312
×3	IRN	33.61/0.8806	31.74/0.8761	30.51/0.8254	33.86/0.8237
×3	MAN(Ours)	33.70/0.8813	32.07/0.8807	30.68/0.8401	34.20/0.9391
×4	Bicubic	28.62/0.7742	27.22/0.7401	27.33/0.7168	28.54/0.8241
×4	FSRCNN	30.83/0.8275	28.46/0.7638	28.36/0.7516	30.42/0.8721
×4	ESPCN	31.25/0.8397	29.02/0.7718	28.63/0.7565	30.80/0.8852
×4	VDSR	31.29/0.8412	29.27/0.8046	28.66/0.7620	30.98/0.8820
×4	SRFBN	31.41/0.8466	29.24/0.7994	28.59/0.7581	31.24/0.9001
×4	RCAN	31.46/0.8467	29.49/0.8123	28.85/0.7715	31.20/0.8974
×4	IRN	31.51/0.8506	29.40/0.8007	28.64/0.7662	31.16/0.8899
×4	MAN(Ours)	31.55/0.8534	29.51/0.8134	28.96/0.7796	31.27/0.9007

4.4 本文算法和其它算法的对比分析

表 1 给出了各算法在 Google dataset of SIRI-WHU_earth、

UCMerced、NWPU-RESISC45 以及 Christchurch Aerial Semantic Dataset 数据集上的 PSNR 和 SSIM 结果值。可以看

到，双三次插值法的 PSNR 和 SSIM 值最低。FSRCNN 和 ESPCN 作为早期的超分辨率算法，相对于双三次插值法有一定的提高，但重建效果仍然一般。对于本文方法，当缩放因子为 3 时，除了在 Google dataset of SIRI-WHU_earth 上的 PSNR 略低以外，在其它测试集上都取得了最佳效果。例如，对于 PSNR 指标，在 Christchurch Aerial Semantic Dataset 测试集上，当缩放因子为 2、3、4 时本文方法分别为 39.21dB、34.20dB 和 31.27dB；而对于 SSIM 指标，当缩放因子为 2、3、4 时本文方法分别为 0.9687、0.9391 和 0.9007，可以看到本文方法均取得最好的重建效果。并且 Christchurch Aerial Semantic Dataset 单幅图像中包含的像素内容最多，图像中内容复杂，不同物体的尺寸差异较大，这说明本文提出的方法在对这类图像重建时有不错的效果。

4.5 视觉效果分析

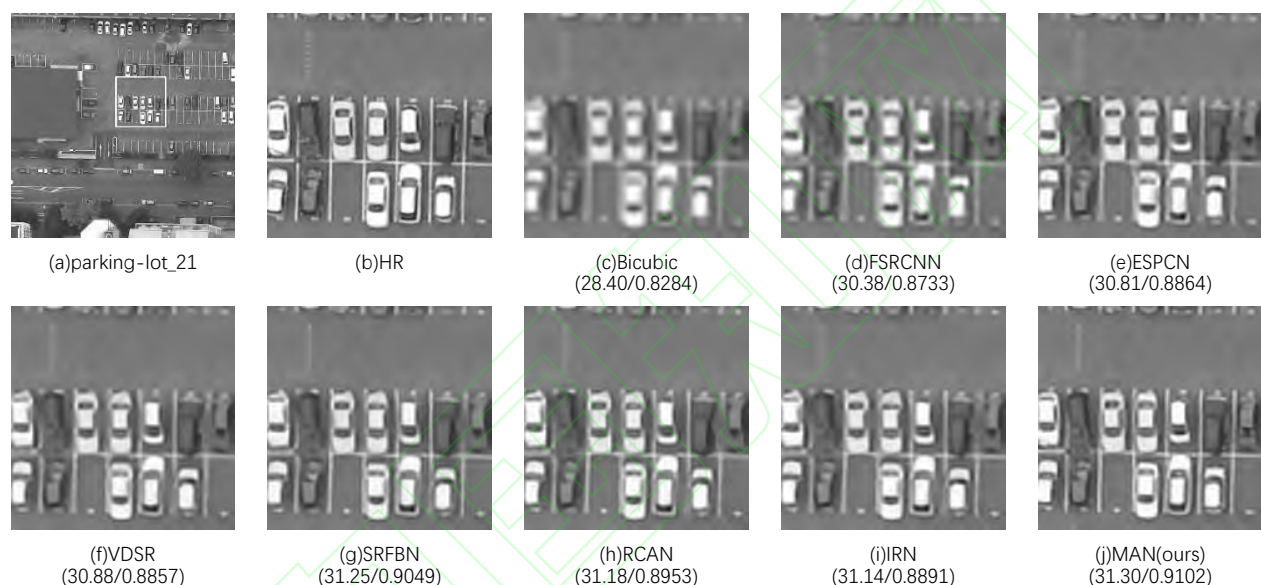


图 6 重建对比图

Fig.6 Comparison of reconstruction results

5 总结

本文提出了一种融合多路径与混合注意力网络的遥感图像超分辨率重建算法，通过引入 CMB 模块，能够将级联局部特征进行融合，充分提取不同层次的卷积层特征，关注空间差异较大的目标物。此外，设计 MAN 网络单元，该网络单元通过两层多次的特征提取能够更好地捕捉高频信息和不同尺度的特征，弥补细节丢失的不足。并在此基础上引入混合注意力机制，进一步对重要特征的重建进行关注。实验结果表明，本文方法在客观评价和主观评价上较之前的主流算法都有一定的提升。

虽然本文方法在遥感图像超分辨率重建中取得了不错的效果，但仍有许多问题需要完善。比如，由于本文在 MAN 网络单元中采用多路链式卷积的形式进行特征提取，并在其中应用 5×5 和 3×3 的大卷积核，所以在训练模型的时候相比其他同量级的方法占用更多的内存更多，同时也更加耗时。因此，如何解决这种多路级联网络带来的空间问题和时间问

除了通过 PSNR 和 SSIM 两项客观的量化评价指标来衡量本文模型与其它模型的效果以外，还通过视觉效果分析来对比本文模型与其它模型的重建效果。考虑到篇幅原因，此处只列举出在 Christchurch Aerial Semantic Dataset 数据集上缩放因子为 4 时在 parking-lot_21 上的视觉效果分析结果，如图 6 所示。

从图 6 中可以看出，Bicubic 算法所重建的图像较为模糊且丢失信息较多，重建效果最差。FSRCNN、ESPCN、VDSR、SRFBN、RCAN 和 IRN 算法相较于 Bicubic 算法有明显的重建效果提升，但是在细节上仍然有所缺失。而本文方法重建的细节更加丰富，与原图更加接近，视觉效果更好。当然，本文方法在采样倍数较高时与原图相比仍然有差距，但相对于其它方法，仍然取得最好的视觉效果。

题是今后需要研究的方向。

References:

- [1] Agustsson E, Timofte R. NTIRE 2017 challenge on single image super-resolution: methods and results [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1110-1121.
- [2] Wang Wei, Zhang Tong, Wang Xin. Survey of single image super-resolution based on deep learning [J]. Journal of Chinese Computer Systems, 2019, 40(9): 1891-1896.
- [3] Dong C, Loy C C, He K M, et al. Learning a deep convolutional network for image super resolution [C] // Proceedings of the 13th European Conference on Computer Vision, 2014: 184-199.
- [4] Dong C, Loy C C, Tang X O. Accelerating the super-resolution convolutional neural network [C] // Proceedings of the European Conference on Computer

Vision,2016:391-407.

[5] Shi W,Caballero J,Huszar F,et al.Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016:1874-1883.

[6] Kim J,Lee J K,Lee K M.Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016:1646-1654.

[7] Li Z,Yang J,Liu Z,et al.Feedback network for image super-resolution[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),2019.

[8] Cheng De-qiang,Guo Xin,Chen Liang-liang,et al.Image super-resolution reconstruction from multi-channel recursive residual network[J].Journal of the Image and Graphics,2021,26(3):605-618.

[9] Zhang Y,Li K,Li K,et al.Image super-resolution using very deep residual channel attention networks[C]//Proceedings of the European Conference on Computer Vision,2018:286-301.

[10] Shi Ju-peng,Li Jing,Chen Yan,et al.DFAN:a deep feedback attention network for image super-resolution[J].Journal of Chinese Computer Systems,2021,42(6):1206-1212.

[11] Li Z.Image super-resolution using attention based densenet with residual deconvolution[J].arXiv preprint arXiv:1907.05282,2019.

[12] Wang Zhi-hao,Jian Chen,Steven S H Hoi.Deep learning for image super-resolution:a survey[J].arXiv preprint arXiv:1902.06068,2019.

[13] Lim B,Son S,Kim H,et al.Enhanced deep residual networks for single image super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,2017:1132-1140.

[14] Ahn N,Kang B,Sohn K A.Fast,accurate,and lightweight super-resolution with cascading residual network[C]//Proceedings of the European Conference on Computer Vision,2018:252-268.

[15] Ledig C,Theis L,Huszar F,et al.Photo-realistic single image super-resolution using a generative adversarial network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:4681-4690.

[16] Szegedy C,Liu W,Jia Y,et al.Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2015:1-9.

[17] Chollet F.Xception:deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition,2017:1800-1807.

[18] Woo S,Park J,Lee J Y,et al.CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision,2018:3-19.

[19] Xie H P,Jiang H Y,Liu X Y,et al.Super resolution for remote sensing images via improved residual network[C]//Proceedings of the 5th International Conference on Mechanical,Control and Computer Engineering,2020:2295-2298.

[20] Lai W S,Huang J B,Ahuja N,et al.Deep laplacian pyramid networks for fast and accurate super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:5835-5843.

[21] Lai W S,Huang J B,Ahuja N,et al.Fast and accurate image super-resolution with deep laplacian pyramid networks[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2019,41(11):2599-2613.

附中文参考文献:

[2] 王 威,张 彤,王 新.用于图像超分辨率重构的深度学习方法综述[J].小型微型计算机系统,2019,40(9):1891-1896.

[8] 程德强,郭 昕,陈亮亮,等.多通道递归残差网络的图像超分辨率重建[J].中国图象图形学报,2021,26(3):605-618.

[10] 施举鹏,李 静,陈 琰,等.DFAN:一种基于深度反馈注意力网络的图像超分辨率方法[J].小型微型计算机系统,2021,42(6):1206-1212.