

doi: 10.7690/bgzd.2024.04.008

基于注意力机制的残差网络超分辨率重建方法

常哲¹, 陈鹏云¹, 李佳成¹, 马英琪²

(1. 中北大学航空宇航学院, 太原 030051; 2. 中北大学机电工程学院, 太原 030051)

摘要: 针对现有方法在图像超分辨率重建中高频信息遗漏的问题, 提出一种基于注意力机制的多层嵌套残差网络超分辨率重建方法, 对不同的频次信息采用不同的特征提取方案。利用跨越残差网络结构将前端的特征信息利用恒等映射连接直接传送到后端的注意力机制模块, 添加注意力机制的多层嵌套残差网络来捕捉原始图像中隐含的特征信息, 通过深度并行残差网络结构融合图像特征信息。实验结果表明, 改进后算法可有效提高图像超分辨率重建精度。

关键词: 图像处理; 多层嵌套残差网络; 注意力机制; 超分辨率重建

中图分类号: TP399 **文献标志码:** A

Residual Network Super-resolution Reconstruction Method Based on Attention Mechanism

Chang Zhe¹, Chen Pengyun¹, Li Jiacheng¹, Ma Yingqi²

(1. School of Aerospace Engineering, North University of China, Taiyuan 030051, China;

2. School of Mechanical and Electrical Engineering, North University of China, Taiyuan 030051, China)

Abstract: In order to solve the problem of missing high-frequency information in image super-resolution reconstruction by existing methods, this paper proposes a multi-layer nested residual network super-resolution reconstruction method based on attention mechanism, which uses different feature extraction schemes for different frequency information. The front-end feature information is directly transmitted to the back-end attention mechanism module by using the identity mapping connection across the residual network structure, and the hidden feature information in the original image is captured by adding the multi-layer nested residual network of the attention mechanism, and the image feature information is fused through the deep parallel residual network structure. Experimental results show that the improved algorithm can effectively improve the accuracy of image super-resolution reconstruction.

Keywords: image processing; multilayer nested residual network; attention mechanism; super-resolution reconstruction

0 引言

传统的单幅图像超分辨率重建 (single image super resolution reconstruction, SISR) 方法主要基于插值和重建, 但随着计算机技术的发展, 基于学习的 SISR 方法成为现有研究的热点。Dong 等^[1]提出 SRCNN 基础网络框架。基于此种框架, 卢正浩等^[2]在网络中加入注意力机制提升特征提取能力。Song 等^[3]提出 AdderNets 算法来降低运行内存和成本, 提高特征提取能力。Deng 等^[4]提出耦合反馈神经网络, 以解决曝光异常的输入图像的重建问题。Chen 等^[5]利用局部隐式函数在图像放大任意倍数后仍保持较高分辨率水平。原铭等^[6]提出了基于多尺度的密集残差网络, 在提取不同尺度特征信息的同时, 还能够提取不同感受野下的特征信息。陈玉立等^[7]为解决无法有效提取关键信息问题, 提出一种基于注意力机制和生成对抗网络的短期航迹预测模型,

实现了数据分布的较高精准拟合。Shang 等^[8]交替使用不同的上采样方法并将感受野块加入网络以提高可分辨率性。Hwang 等^[9]提出一种分层压缩技术, 实现低保存率下的收敛。Jiang 等^[10]设计了双流残差块和简化上采样模块, 一定程度上平衡了特征提取能力和轻量级性能之间的矛盾。

随着网络深度的增加, 网络很可能会出现梯度消失的情况, 从而使模型训练失败^[11]。同时, 结构的加宽和加深也容易导致网络性能被图像的低频信息所妨碍, 不能更好地读取不同的频次信息, 重建效果往往不佳。在这种方式下, 映射图像会舍弃掉大量高频信息, 导致重建图像质量不高; 而且对于网络提取的特征在空间和通道间内一视同仁, 这种处理方式并不能很好地应对不同的特征映射。大部分的 SISR 算法都是在中间的特征提取层更新算法结构, 整体仍为基础架构上的单层网络, 常用的单

收稿日期: 2023-12-10; 修回日期: 2024-01-17

基金项目: 国家自然科学基金(51909245, 62003314); 山西省重点研发计划(202202020101001); 2022年中北大学研究生科技立项资助项目(20221876)

第一作者: 常哲(1997—), 女, 辽宁人, 硕士。

层残差块传递的特征信息有限，而且对于提取到的特征信息一视同仁，没有侧重与辨识。

针对以上问题，笔者提出一种基于注意力机制的多层嵌套残差网络超分辨率重建方法。对于不同的频次信息采用不同的特征提取方案，首先提取原始特征，通过跨越残差网络结构，把前端的特征信息利用恒等映射连接直接传送到后端的注意力机制模块，将多层次的注意力残差块通过不断地累积相加，以提取更深层次的特征信息，最终通过一个全局长连接将输入进行信息融合，得到重建图像。笔者的主要工作如下：

1) 在基础架构的单层网络框架上，更新特征提取层的算法，将注意力残差块分层级联，采用大循环套小循环的模式进行嵌套，在大循环中，累计每个小循环的输出结果作为残差叠加。同时，添加深度并行残差网络结构，融合局部残差学习信息和全

局信息，提升训练速度和重建效果。

2) 在新建立起的深层残差块内添加通道注意力机制，即在小循环中，累计每一层注意力残差块输出结果，并且在该支路的最后使用卷积层整合信息，这种嵌套方法能尽可能多地提取出隐含的残差高频细节信息，使嵌套的网络模块能更大程度地捕捉不同层次的高频残差特征。

1 算法基本结构

深度学习 SISR 基础网络框架如图 1 所示。

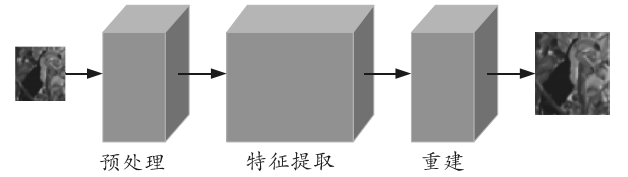


图 1 深度学习 SISR 基础网络框架

如图 2 所示，本文中算法的结构分为 3 个模块。

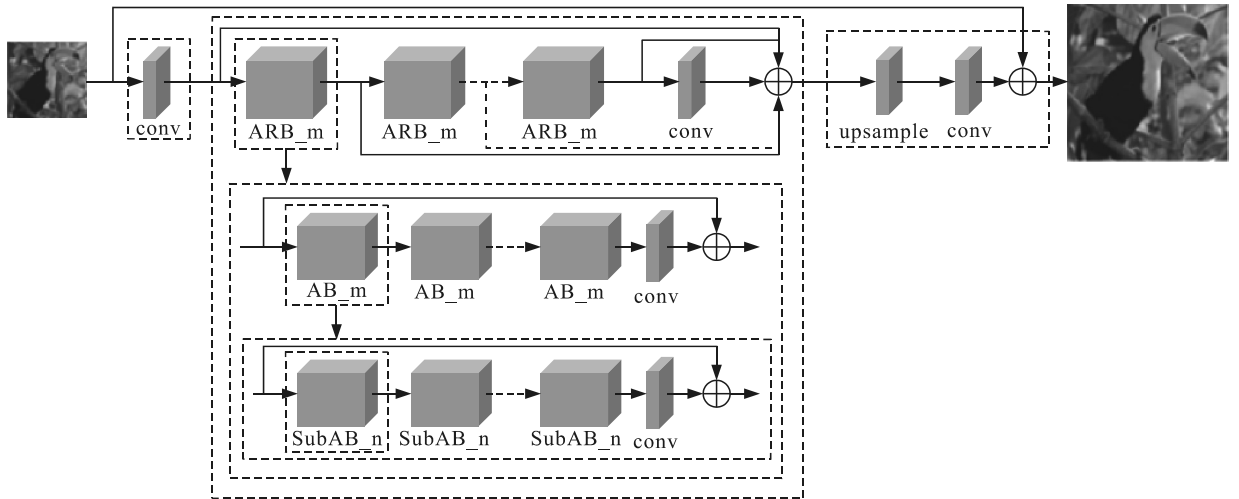


图 2 本文中算法网络结构

在提取浅层特征的模块中， I_L 为网络输入， I_H 为网络输出，使用卷积得到浅层特征 F_{SF} ，即：

$$F_{SF} = H_{SF}(I_L)。 \quad (1)$$

式中 $H_{SF}(\cdot)$ 为单层卷积映射。

在提取多层注意力特征的模块中，将 F_{SF} 作为输入，通过多层注意力特征提取以后得到高维特征 F_{GF} ，即：

$$F_{SF} = H_{FJS}(H_{TCRF}(H_{OCRF}(H_{ZCRF}(H_{CA}(F_{SF}))))); \quad (2)$$

$$F_{GF} = H_{MCRF}(F_{SF})。 \quad (3)$$

式中： $H_{MCRF}(\cdot)$ 为多层注意力特征提取模块映射关系； $H_{CA}(\cdot)$ 为通道注意力机制映射关系； $H_{ZCRF}(\cdot)$ ， $H_{OCRF}(\cdot)$ 和 $H_{TCRF}(\cdot)$ 为多层次残差层； $H_{FJS}(\cdot)$ 为密集连接。

在用于重建的模块中，转变后的特征 F_{GF} 作为

输入之一，首先要通过上采样操作，本文中选择亚像素卷积。像素重排是亚像素上卷积操作的核心，能够使网络在重建速率和精度方面达到较好效果，而且这种上采样操作的学习参数比反卷积操作的学习参数要少^[12]。

之后通过一个卷积层，计算如下：

$$T_H = H_{SR}(F_{GF}) = H_{CR}(H_{SM}(F_{GF}))。 \quad (4)$$

式中： T_H 为输出三通道彩色图像； $H_{SR}(\cdot)$ 为重建模块的映射关系； $H_{SM}(\cdot)$ 为亚像素上采样操作； $H_{CR}(\cdot)$ 为最后一步卷积操作。

融合需要的另外一种输入是源输入信息，表示为：

$$I_H = H_{FUS}(T_H, I_L)。 \quad (5)$$

式中 H_{FUS} 为长跳跃连接融合。

2 多层注意力特征提取

2.1 多层嵌套残差网络

一个残差块可以表示为:

$$y_i = h(x_i) + \mathcal{F}(x_i, W_i); \quad (6)$$

$$x_{i+1} = \sigma(y_i). \quad (7)$$

式中: $h(x_i)$ 为直接映射部分; $\mathcal{F}(x_i, W_i)$ 为残差部分; $\sigma(*)$ 为 Relu 激活函数。

不考虑升维或者降维的情况, 这时残差块可以表示为:

$$x_{i+1} = x_i + \mathcal{F}(x_i, W_i). \quad (8)$$

对于一个更深的层 L , 其与 l 层的关系可以表示为:

$$x_L = x_l + \sum_{i=l}^{L-1} \mathcal{F}(x_i, W_i). \quad (9)$$

根据 BP 中使用的导数的链式法则, 损失函数 ε 关于 x_l 的梯度可以表示为:

$$\frac{\partial \varepsilon}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_l} \frac{\partial x_l}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_l} \left(1 + \frac{\partial}{\partial x_l} \sum_{i=l}^{L-1} \mathcal{F}(x_i, W_i) \right). \quad (10)$$

也就是说在残差网络中不会出现梯度消失的问题, 并且 L 层的梯度可以直接传递到任一比它浅的 l 层^[13]。

每个 ARB 模块的输入都与输出进行密集连接以融合信息, 能够更好地利用起分层提取到的特征, 强化整体的采集能力。BN 层通过采取归一化操作^[14], 把分布逐渐混乱的神经元输入强制拉回到标准正态分布^[15], 从而控制梯度爆炸防止梯度消失, 加快网络训练和收敛的速度, 同时又具有正则化效果以防止模型过拟合^[16], 能够在一定的程度上提高模型训练精度。但是在实际的 SISR 任务中, 好的方法不一定适用于所有问题。针对一般情况而言, BN 层的规范化手段会破坏原始空间的表征, 模型的收敛会变简单, 但不见得最终结果一定最好。本文中省去了 BN 层。结构设计如图 3 所示。

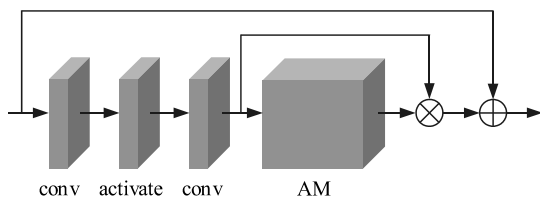


图 3 本文中局部残差网络结构

2.2 通道注意力机制

注意力机制借鉴了人眼对画面的关注度原理, 一经提出便得到广泛关注和采用。以 SE 块^[17]为例,

它的出现是为了解决在卷积池化过程中不同通道所占重要性不同而带来的损失问题, 通过网络的全局损失函数自适应地矫正特征相应强度^[18]。

假设输入为 $H*W*C$ 的特征图, $H*W$ 表示高和宽, C 表示通道数, 主要分为 2 个阶段^[19]。

第 1 阶段: 采用 2 种池化操作, 进行平均池化时, 令 $U=[u_1, u_2, \dots, u_c, \dots, u_C]$, 所有通道汇总为 $Z \in R^c$, Z 的第 c 个通道表示为:

$$F_{\text{avg}}^c = H_{\text{avg}}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j). \quad (11)$$

式中: $u_c(i, j)$ 为输入特征映射中第 c 个通道中位置 (i, j) 的像素值; F_{avg}^c 为所有通道中的第 c 个通道; $H_{\text{avg}}(*)$ 为全局平均池化操作。

同时进行的还有将输入特征连接到最大池化层, 表示为:

$$F_{\text{max}}^c = H_{\text{MP}}(F_{\text{SF}}^c). \quad (12)$$

式中 $H_{\text{MP}}(*)$ 为最大池化操作。

将特征向量 F_{avg}^c 和 F_{max}^c 进行特征拼接得到 2 维矩阵 F_C , 表示为:

$$F_C = C(F_{\text{avg}}^c, F_{\text{max}}^c). \quad (13)$$

式中 $C(*)$ 为特征拼接操作。

第 2 阶段: 归一化权重作用到输入特征中, 表示为:

$$F_E = f(W_s(\sigma(W_U(W_V F_C))))). \quad (14)$$

式中: $f(*)$ 为 Sigmoid 激活函数; W_s 为 1 维卷积参数; W_U 和 W_V 是卷积核的权重参数。

最后, 利用新旧信息相乘得到最终的处理结果。表示为:

$$F_{\text{CA}} = F_E \cdot u_c. \quad (15)$$

式中 u_c 和 F_E 分别为第 c 个通道的原始特征和处理后的特征。本文中通道注意力机制的实现原理如图 4 所示。

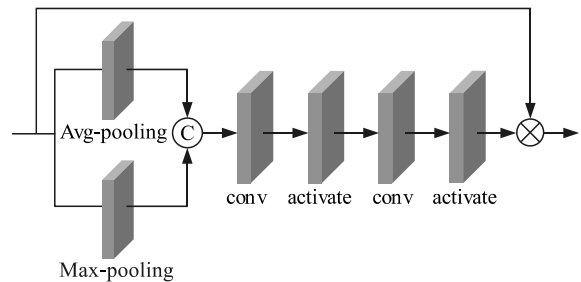


图 4 本文中通道注意力机制

2.3 损失函数

本文中采用 L_1 损失函数:

$$L(\Theta) = \sum_{i=1}^n |H_{\text{NET}}(I_L^i) - I_H^i|。 \quad (16)$$

式中： Θ 为网络中所有参数； $H_{\text{NET}}(I_L^i)$ 为通过本文中网络结构得到的高分辨率图像； I_L^i 和 I_H^i 分别为第 i 对重建图像和原始的高分辨率图像。

3 实验和结果分析

3.1 实验条件设置

实验使用 Windows10 操作系统，利用深度学习框架 TensorFlow 搭建模型，并通过 CUDA Toolkit 10.0 和 cuDNN 7.6.5 加速，使用 ADAM 优化器 ($\alpha=10^{-3}$, $\beta_1=0.9$, $\beta_2=0.999$) 更新参数。在整个模型中，特征融合部分和通道注意力机制内通道压缩和扩增瓶颈层采用 $1*1$ 的卷积核，其余卷积层则采用 $3*3$ 的卷积核，压缩倍数为 16，激活函数只有通道注意力机制处最后一层使用 Sigmoid 函数激活，其余都采用 Relu 函数激活，使用补零策略。仿真硬件平台如表 1 所示。

表 1 仿真硬件平台

硬件	型号/版本
CPU	11 th Gen Inter(R) Core (TM) i5-11500
GPU	NVIDIA T600
内存/GB	16

实验采用的训练数据集是包含了丰富的场景和边缘及纹理细节的超分辨率修复或重建常用数据集 DIV2K^[20]。测试集采用 Set5^[21]、Set14^[22] 以及开源航拍图像。Set5 数据集包含 5 张图像；Set14 数据集包含 14 张图像，比 Set5 数据集的细节信息更多；几张开源航拍图像进行不同算法重建以后将局部放大，用作算法细节比较。

本文中采用峰值信噪比 (peak signal to noise ratio, PSNR) 和结构相似度 (structural similarity index, SSIM) 指标来评价重建质量。计算公式如下：

峰值信噪比定义为：

$$\text{PSNR} = 10 \lg \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right) = 20 \lg \left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}}} \right)。 \quad (17)$$

结构相似性定义为：

$$\text{SSIM} = \frac{(2\mu_a\mu_b + C_1)(2\sigma_{ab} + C_2)}{(\mu_a^2 + \mu_b^2 + C_1)(\sigma_a^2 + \sigma_b^2 + C_2)}。 \quad (18)$$

式中：表现亮度的参数 μ_a 和 μ_b 分别为图像 A 和 B 的均值；表现对比度的参数 σ_a^2 和 σ_b^2 分别为图像 A 和 B 的方差；表现结构的参数 σ_{ab} 为图像 A 和 B 的协方差； C_1 和 C_2 是常数。结果越接近 1 表示重建

效果越好。

3.2 消融实验

3.2.1 嵌套残差网络层数的确定

为确定多层嵌套残差网络对性能的影响，分别对层数为 2, 3, 4 的模型训练 1 000 个 epoch，在 set5 测试集上放大 4 倍操作得到测评指标，结果如表 2 所示。由表中数据可知，随着层数的增加，测评指标的数据也会有所增加。层数为 3 时要比层数为 2 时的 PSNR 增长 0.43 dB，但层数为 4 时仅比层数为 3 时增长 0.08 dB，提升百分比并不高，最终选定层数为 3 的网络模型。

表 2 嵌套残差网络层数的选取

层数	PSNR/dB	SSIM
2	30.19	0.830 7
3	30.62	0.851 1
4	30.70	0.852 7

3.2.2 通道注意力机制有效性验证

为验证通道注意力机制对网络性能的影响，分别对有无通道注意力机制的模型训练 1 000 个 epoch，在 set5 测试集上放大 4 倍操作得到测评指标，结果如表 3 所示。由表中数据可知，添加通道注意力机制时要比不加时的 PSNR 增长 0.39 dB，有助于提升网络性能。

表 3 通道注意力机制的有效性验证

通道注意力	PSNR/dB	SSIM
×	30.23	0.834 4
√	30.62	0.851 1

3.2.3 不同池化方式的有效性验证

为验证通道注意力机制中池化方式对网络性能的影响，分别对不同池化方式的模型训练 1 000 个 epoch，在 set5 测试集上放大 4 倍操作得到测评指标，结果如表 4 所示。由表中数据可知，使用最大池化和平均池化组合时比单独使用时的 PSNR 分别增长 0.12 和 0.08 dB，有助于提升图像重建质量。

表 4 不同池化方式的有效性验证

最大池化	平均池化	PSNR/dB	SSIM
√	×	30.50	0.847 9
×	√	30.54	0.848 6
√	√	30.62	0.851 1

3.3 实验结果分析

将本文中算法与 SRResnet^[23]、EDSR^[24]、VDSR^[25]、RDN^[26]、PAN^[27] 和 CFSRCN^[28] 等方法进行比较。放大因子为 4 时，几种算法对 Set5 和 Set14

数据集的评测指标平均值如表 5 和 6 所示。为叙述方便，将本文中算法称为 Ours。从测试集和开源航

拍图像中选取部分图像进行重建测试，实验结果如图 5—9 以及表 5、6 所示。

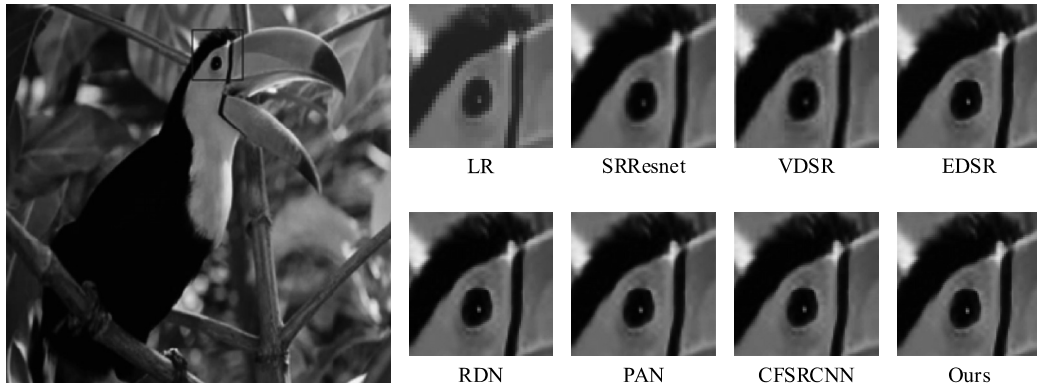


图 5 Set5 测试集 Bird 的 4 倍超分辨率重建可视化比较

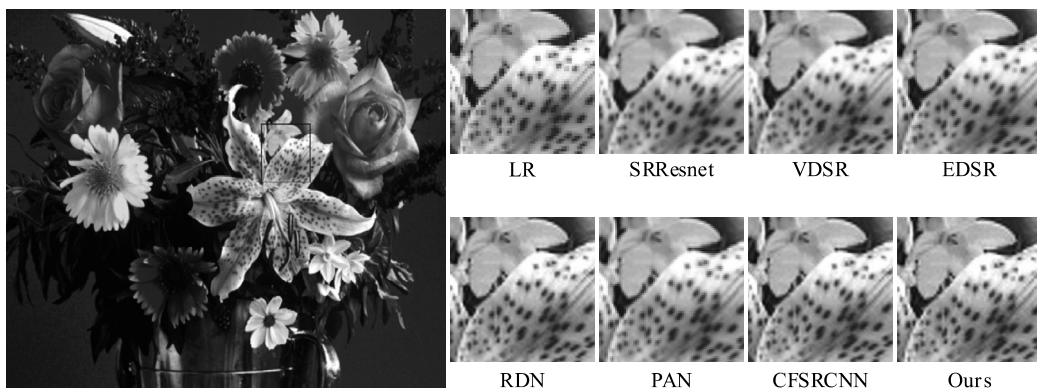


图 6 Set14 测试集 Flower 的 4 倍超分辨率重建可视化比较

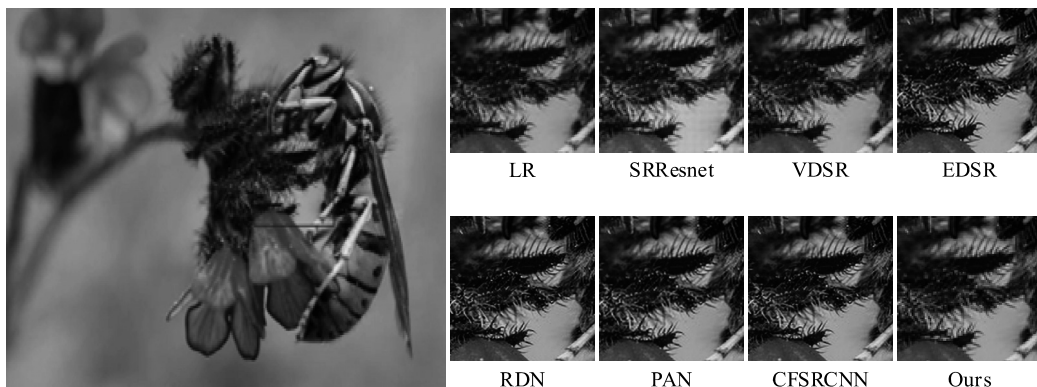


图 7 开源航拍图像 Bee 的 4 倍超分辨率重建可视化比较

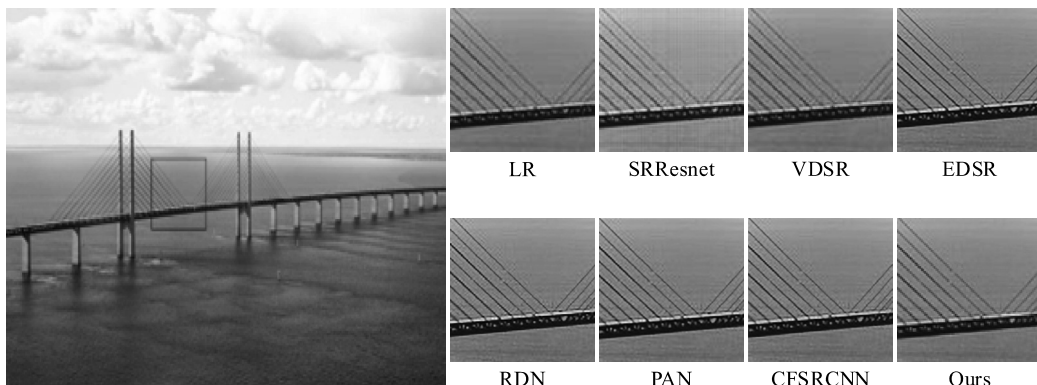


图 8 开源航拍图像 Bridge 的 4 倍超分辨率重建可视化比较

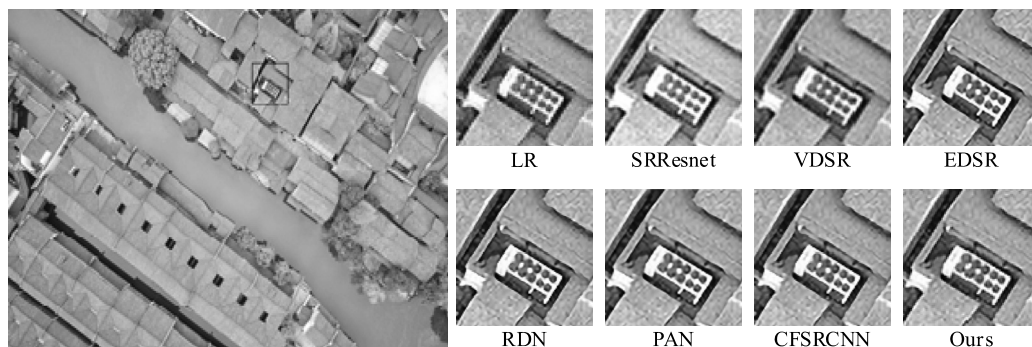


图 9 开源航拍图像 Wuzhen 的 4 倍超分辨率重建可视化比较

表 5 不同 SISR 方法的平均 PSNR 结果

算法	Set5	Set14
SRResnet	28.30	26.47
VDSR	29.77	26.46
EDSR	30.58	26.90
RDN	30.60	26.91
PAN	30.51	26.79
CFSRCNN	30.49	26.78
Ours	30.62	26.93

表 6 不同 SISR 方法的平均 SSIM 结果

算法	Set5	Set14
SRResnet	0.801 5	0.721 9
VDSR	0.825 5	0.738 7
EDSR	0.850 7	0.743 3
RDN	0.850 8	0.743 5
PAN	0.849 6	0.741 6
CFSRCNN	0.848 4	0.740 2
Ours	0.851 1	0.744 0

图 5—9 展示了几张测试图像使用本文中算法和其他几种算法在放大 4 倍时的重建效果。从视觉效果上来看，SRResnet 重新划分了像素块网格，展现出一定的细节恢复；VDSR 采用了深的网络，颜色更加鲜明；EDSR 增强了相邻像素块之间的对比度用以体现颜色的边缘；RDN 在 EDSR 的基础上表现更佳；PAN 和 CFSRCNN 重建效果良好，但本文中算法重建后的线条边缘更加细腻，纹理更加丰富，重建效果更优。从表 5 和 6 数据也可表明本文中提出的方法具有更佳的重建精度。

4 结束语

笔者提出一种基于注意力机制的多层残差网络超分辨率重建方法，通过嵌套的方式增加网络层数，使用密集连接融合图像，添加通道注意力机制以提高网络获取关键特征的能力，获取更清晰的图像。实验结果证明：本文中提出的重建算法的测评数据对比另外几种算法有更佳的重建精度，对不同场景图像的超分辨率重建均具有较好效果。未来将更注重信息特征以获得更强的识别学习能力，研究更深层次的网络结构以提高重建精度，优化残差阶段和上采样阶段以获得更轻、更快的结构。

参考文献：

- [1] DONG C, LOY C C, HE K, et al. Learning a deep convolutional network for image super-resolution[C]//Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13. Springer International Publishing, 2014: 184-199.
- [2] 卢正浩, 刘丛. 多尺度特征复用混合注意力网络的图像重建[J]. 中国图象图形学报, 2021, 26(11): 2645-2658.
- [3] SONG D, WANG Y, CHEN H, et al. Addsr: Towards energy efficient image super-resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2021: 15648-15657.
- [4] DENG X, ZHANG Y, XU M, et al. Deep coupled feedback network for joint exposure fusion and image super-resolution[J]. IEEE Transactions on Image Processing, 2021, 30: 3098-3112.
- [5] CHEN Y, LIU S, WANG X. Learning continuous image representation with local implicit image function[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. IEEE, 2021: 8628-8638.
- [6] 原铭, 李凡, 李华锋, 等. 基于多级注意力密集残差网络的单幅图像超分辨率重建[J]. 光学技术, 2022, 48(3): 357-363.
- [7] 陈玉立, 佟强, 谌彤童, 等. 基于注意力机制和生成对抗网络的飞行器短期航迹预测模型[J]. 计算机应用, 2022, 42(10): 3292-3299.
- [8] SHANG T, DAI Q, ZHU S, et al. Perceptual extreme super-resolution network with receptive field block[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 440-441.
- [9] HWANG J, UDDIN A F M S, BAE S H. A layer-wise extreme network compression for super resolution[J]. IEEE Access, 2021, 9: 93998-94009.
- [10] JIANG Y, LIU Y, ZHAN W, et al. Lightweight dual-stream residual network for single image super-resolution[J]. IEEE Access, 2021, 9: 129890-129901.