

李美钰, 丁纪峰, 熬英健, 等. 基于注意力机制的倒残差图像超分辨率重建模型[J]. 智能计算机与应用, 2025, 15(12): 31-38. DOI: 10.20169/j.issn.2095-2163.251205

基于注意力机制的倒残差图像超分辨率重建模型

李美钰, 丁纪峰, 熬英健, 张昊然, 金宏轩

(大连民族大学 信息与通信工程学院, 辽宁 大连 116605)

摘要: 为了应对在图像超分辨率任务中, 图像特征提取不充分、图像超分辨率模型性能表现不佳的问题, 本文提出了一种基于注意力机制的倒残差特征提取网络的图像超分辨率模型, 该模型由3个部分组成: 浅层特征提取模块、深层特征提取模块和高质量图像重建模块。特别地, 浅层特征提取模块依然使用一个卷积层从输入图像中提取浅层特征; 深层特征提取模块由4个残差倒置注意力模块和9个残差的Swin Transformer特征提取模块(STFEB)并联组成, 每个模块具有6个Swin Transformer层以及一个残差连接, 高质量图像重建模块采用上采样机制。大量的实验证明, 模型在不同放大倍数情形下均表现良好, 与现有的模型相比更具有竞争力。

关键词: 图像超分辨率; 注意力机制; 特征提取; 倒残差

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2025)12-0031-08

A super-resolution algorithm for inverted residual images based on attention mechanism

LI Meiyu, DING Jifeng, AO Yingjian, ZHANG Haoran, JIN Hongxuan

(School of Information and Communication Engineering, Dalian Minzu University, Dalian 116605, Liaoning, China)

Abstract: In order to address the issues of insufficient image feature extraction and poor performance of image super-resolution models in image super-resolution tasks, an image super-resolution model based on an attention mechanism based inverted residual feature extraction network is proposed. This model consists of three parts: shallow feature extraction, deep feature extraction, and high-quality image reconstruction. Specifically, the shallow feature extraction module still uses a convolutional layer to extract shallow features from the input image, while the deep feature extraction module consists of four residual inverted attention modules and nine parallel residual Swin Transformer Feature Extraction Block (STFEB), each module having six Swin Transformer layers along with a residual connection. The high-quality image reconstruction module adopts an upsampling mechanism. A large number of experiments have shown that the model performs well under different magnification conditions and exhibits competitive performance compared to existing methods.

Key words: image super-resolution; attention mechanism; feature extraction; Inverse residual

0 引言

图像超分辨率一直是数字图像处理领域中的热点问题, 逐渐被应用到各个领域, 如监控设备、遥感图像、医学影像、雷达成像、显微成像等。超分辨率重建方法主要分为基于插值的方法、基于重建的方法、基于学习的方法。

2014年Dong等^[1]提出的超分辨率卷积神经网络(Super-Resolution Convolutional Neural Network, SRCNN)首次将卷积神经网络引入图像超分辨率领

域, 取得了非常优异的效果, 开启了基于深度学习超分辨率重建的大门; 2016年, Dong等^[2]在SRCNN的基础上解决了非线性映射计算代价太高的问题, 提出了快速超分辨率卷积神经网络(Fast Super-Resolution Convolutional Neural Network, FSRCNN), 进一步证实了卷积神经网络在图像超分辨率重建领域中相比于其他模型具有更好的效果; Kim等^[3]提出的具有更深层次的网络结构的图像超分辨率(Very Deep Super-Resolution, VDSR)网络, 将残差的思想融入到图像超分辨率重建中, 通过残差块解决

作者简介: 李美钰(1998—), 女, 硕士研究生, 主要研究方向: 图像处理。

通信作者: 丁纪峰(1978—), 男, 硕士, 副教授, 硕士生导师, 主要研究方向: 计算机视觉。Email: 1486757892@qq.com。

收稿日期: 2024-03-13

图像超分辨率网络随着网络变深梯度消失和梯度爆炸的问题;Shi 等^[4]提出的高效亚像素卷积神经网络(Efficient Sub-Pixel Convolutional Neural Network, ESPCN),使用亚像素卷积上采样的方式,增加了预上采样机制在图像超分辨率网络的可行性。

现有模型存在特征信息挖掘不足、算法复杂度高及训练不稳定的问题,本文提出一种基于注意力机制的倒残差模型和 Swin Transformer 模型相结合的图像超分辨率重建模型(RISR)。该模型由3个部分组成:浅层特征提取模块、深层特征提取模块和高质量图像重建模块。与流行图像超分辨率模型相比,RISR有如下优点:(1)图像内容和注意力权重之间基于内容的交互,可以解释为空间变化卷积;(2)通过移位窗口机制实现了长程相关性建模。

1 设计路线

随着神经网络深度的增加,输入图像的特征集中与模型的不同方面,这些层次特征能够提供重要信息来恢复重建图像的细节。然而,现有的一些基于神经网络的方法未能充分利用这些层次特征,导致重建图像的质量不佳。针对现有的一些方法没有充分利用低像素图像的信息,造成低性能、很少能挖掘层间特征的内在相关性,造成表达能力不强问题,本文提出了一种基于注意力机制的倒残差模型和 Swin Transformer 模型相结合的图像超分辨率重建模型(RISR),该模型基于 Swin Transformer 模块、注意力机制和倒残差结构,能够充分利用层次特征,提高图像的重建质量。

1)构建了一个基于注意力机制的倒残差模块,融入通道注意力模块,加强了网络对图像的通道细节的感知能力,能够更好地捕捉和学习图像中的高频信息,从而提高网络对图像特征的提取和表达能力;

2)构建了一个深层特征提取模块,该模块由多个注意力倒残差模块和 Swin Transformer 特征提取模块组成。在深度特征提取模块中,注意力倒残差模块和 Swin Transformer 特征提取模块支持模块间的连续记忆,充分利用低像素图像的信息,允许注意力倒残差模块的输出和 Swin Transformer 特征提取模块的输出特征通过残差连接在一起,提高网络性能。

对于图像超分辨率重建任务而言,准确地恢复出缺失的高频信息对于提升图像的质量至关重要。RISR 通过使用注意力倒残差模块和 Swin

Transformer 模块进行全局残差连接,可以更好地从低分辨率图像中提取并恢复高频信息,从而获得更加逼真的高分辨率图像。

基于注意力机制的倒残差图像超分辨率重建模型如图1所示,本文模型的3个组成部分分别是浅层特征提取模块、深层特征提取模块以及图像高质量重建模块。浅层特征提取模块对输入的低分辨率图像进行初步的特征提取,以帮助后续模块更好地进行特征学习和提取;深层特征提取模块利用本文提出的注意力倒残差模块和 Swin Transformer 特征提取模块同时对先前提取的浅层特征进行更深层次的学习,从而产生更全面、更准确的深层特征;注意力倒残差模块、Swin Transformer 特征提取模块与深层特征提取的输入构成全局残差学习,得到特征映射模块的输出;图像高质量重建模块将得到的输入特征进行放大,并重建为相应的高分辨率图像,完成图像超分辨率任务。该模型通过使用这三大模块,能够将低分辨率图像重建为相应的高分辨率图像。

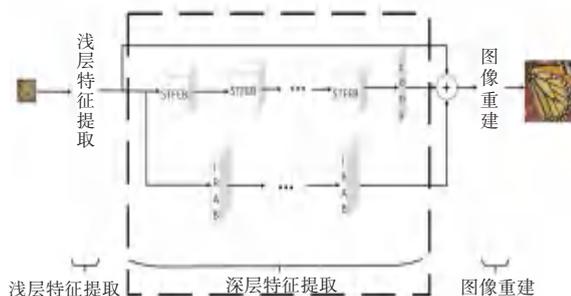


图1 基于注意力机制的倒残差图像超分辨率重建模型

Fig. 1 A super-resolution reconstruction model for inverted residual images based on attention mechanism

2 详细设计

2.1 浅层特征提取模块

浅层特征提取模块依然使用一个卷积层从输入图像中提取浅层特征 F_0 :

$$F_0 = H(I_{LR}) \quad (1)$$

其中, I_{LR} 表示低分辨率图像, H 表示浅层特征提取操作。

提取的浅层特征作为后续模块的输入。使用单层 3×3 卷积网络从低质量图像中提取浅层特征图像,输入图像 $C_{in} = 3$, 输出特征通道 $C = 96$, 步长 $stride = 1$, 填充 $padding = 1$ 的 3×3 卷积网络。卷积层在前期的视觉处理中往往能起到较好的效果,同时也是从图像空间映射到多维特征空间的手段。

经过浅层特征提取模块提取之后,得到的浅层

特征,作为后续模块即特征映射模块的输入。

2.2 深层特征提取模块

本文构建了一个深层特征提取模块,如图 2 所示。该模块由 4 个倒残差注意力模块(Inverted Residual Attention Block, IRAB)和 9 个 Swin Transformer 特征提取模块组成。在深层特征提取模块中,充分利用不同网络的层次特征,提高层次特征的利用率;通过跳跃连接对 IRAB 和 STFEB 的输出与输入特征实现全局残差学习,增强网络提取与表达能力。

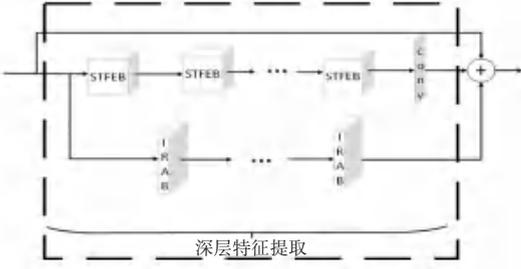


图 2 深层特征提取模块

Fig. 2 Deep feature extraction module

在深层特征提取模块中,支持上下两层之间的连续记忆,允许前一层的 STFEB 输出和当前 IRAB 中卷积层的输出通过局部特征融合连接在一起。

从浅层特征中提取深层特征,如下式所示:

$$F_{DF} = H_{DF}(F_0) \quad (2)$$

其中, H_{DF} 是深度特征提取模块,包含 M 个 STFEB、 N 个 IRAB、1 个特征融合模块、1 个 3×3 卷积层和 1 个长跳跃连接。

中间特征 $F_1, \dots, F_M; F_1, \dots, F_N$ 和输出的深度特征 F_{DF} 被逐块提取,公式如下:

$$F_i = H_{STFEB_i}(F_{i-1}), i = 1, 2, \dots, M \quad (3)$$

$$F_j = H_{IRAB_j}(F_{j-1}), j = 1, 2, \dots, N \quad (4)$$

最后,STFEB 使用 1 个 3×3 卷积层进行降维,去除冗余信息,通过跳跃连接和浅层特征 F_0 构成全局残差学习,具体计算如下式:

$$F_{DF} = H_{CONV}(F_i) + F_T + F_j \quad (5)$$

其中, H_{conv} 表示 3×3 卷积层操作, F_0 表示浅层特征提取模块的输出。

2.2.1 倒残差注意力模块(IRAB)

倒残差注意力模块就是普通倒残差模块和注意力机制相连的模块,倒残差注意力模块如图 3 所示。

倒残差注意力模块采用 1×1 的卷积核进行升高通道维度的操作,通过卷积核大小为 3×3 的深度可分离卷积模块进行卷积,再通过 1×1 的卷积进行降低通道维度的处理,形成两头小中间大的结构。

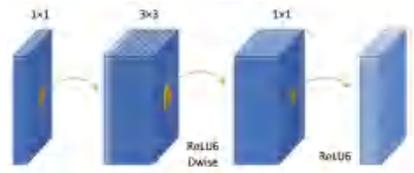


图 3 倒残差注意力模块

Fig. 3 Inverse Residual Attention Module

在高维进行卷积操作获得的特征信息比在低维进行卷积操作所获得的特征信息更加丰富。在升维卷积操作之后要经过归一化层和 ReLU6 激活函数,抑制非常大的值输出,可以适应于低精度的移动端。

注意力机制(Attention Mechanism)是受到人眼视觉研究的启发而提出的,通过模拟人眼对事物的观察顺序来区别对待不同的信息或特征。在神经网络中,注意力机制通常采用网络模块的形式,通过学习训练得到的权重矩阵,重新校准特征图中不同区域对网络模型的重要程度,将关键特征标识出来,有选择性地特征学习。

为了使网络更加专注于关键的通道信息,本文中每个 IRAB 模块中的 SE 通道注意力机制是一种有效的图像特征加权手段,能够帮助模型更好地感知和利用输入图像或特征的通道信息,从而提高模型的性能。

首先,从浅层特征提取层提取的特征信息通过给定第 x 个倒残差注意力模块在倒残差结构中进行投影卷积升维,经过深度卷积进行投影卷积降维,经过注意力机制进行特征提取,由倒残差机制提取中间特征 $F_{x,1}, F_{x,2}, \dots, F_{x,R}$, 如下式:

$$F_{x,i} = H_{inverted}(F_{x,i-1}), i = 1, 2, \dots, R \quad (6)$$

其中, $H_{inverted}$ 表示在第 x 个 IRAB 中的第 i 个倒残差操作。

其次,在每个 IRAB 中添加了一个注意力模块,IRAB 的输出如下式:

$$F_{x,out} = H_{attention}(F_{x,R}) + F_{y,0} \quad (7)$$

其中, $H_{attention}$ 表示第 x 个 IRAB 中的注意力模块进行特征提取。

2.2.2 Swin Transformer 特征提取模块(STFEB)

由于视觉实体规模变化较大,且相对于文本单词,图像具有更高的分辨率和更大的像素密度。由于庞大的参数量和计算量,Transformer 模型在视觉任务中难以发挥最佳性能。为了解决这一问题,Liu 等^[5]在 2021 年提出了 Swin Transformer 模型,通过连续的局部窗口来构建全局特征表示,既能保持准确性,又大大减少了计算量和内存消耗,可作为各种

视觉任务的主干网络。

Swin Transformer 将标准 Transformer 中的多头自注意力模块 (MSA) 替换为基于移位窗口的多头自注意力模块 (W-MSA), 即将图像划分为不重叠的窗口, 在每个窗口内进行自注意力计算。假设每个窗口包含 $M \times M$ 个图像块, 一幅图像有 $h \times w$ 个图像块, 则 MSA 和 W-MSA 的计算复杂度分别如下式:

$$\Omega(\text{MSA}) = 4hwC^2 + 2(hw)^2C \quad (8)$$

$$\Omega(\text{W-MSA}) = 4hwC^2 + 2M^2hwC \quad (9)$$

其中, hw 表示图像块, C 表示图像的通道。

MSA 的计算复杂度与图像块数量 $h \times w$ 成平方关系, 而 W-MSA 的计算复杂度则与图像块数量成线性相关, 从而大大降低了模型的计算复杂度。

为了加强互不重叠窗口之间的信息交流和相关性, Swin Transformer 还提出了移位窗口划分的自注意力 (SW-MSA) 方法来实现跨窗口交互。该方法通过将上一层划分的窗口整体向左上平移像素, 建立上一层相邻不重合窗口的相关性, 同时扩大了感受野。Swin Transformer 的网络结构如图 4 所示。

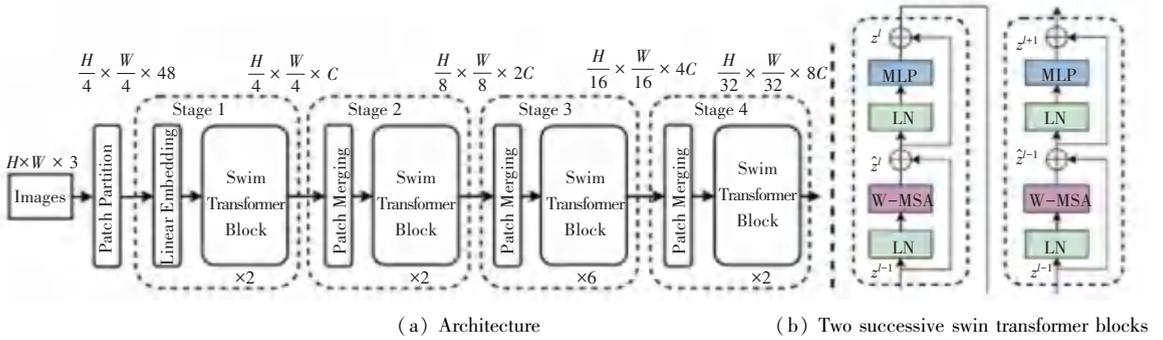


图 4 Swin Transformer 网络结构

Fig. 4 Swin Transformer network architecture

本文的 Swin Transformer 特征提取模块 (STFEB) 受上述启发, 由 Swin Transformer 层 (STL) 和卷积层组成, 模块结构如图 5 所示。

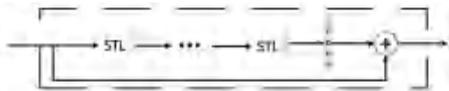


图 5 Swin Transformer 特征提取模块结构

Fig. 5 Swin Transformer feature extraction module structure

给定第 y 个 STFEB 的输入特征 $F_{y,0}$, 首先由 Swin Transformer 层提取中间特征 $F_{y,1}, F_{y,2}, \dots, F_{y,L}$, 如下式:

$$F_{y,j} = H_{\text{STL}}(F_{y,j-1}), j = 1, 2, \dots, L \quad (10)$$

其中, H_{STL} 表示在第 y 个 STFEB 中的第 j 个 Swin Transformer 层。

其次, 本文添加了一个卷积层, 之后进行残差连接, 允许不同级别特征的聚合。浅层特征主要包含低频信息, 而深层特征则专注于恢复丢失的高频信息; 采用一个长距离连接, 将低频信息直接传输给重建模块, 可以帮助深度特征提取模块专注于高频信息, 稳定训练。在图像超分辨率任务中, 将特征上采样, 实现重建。STFEB 的输出如下式:

$$F_{y,\text{out}} = H_{\text{CONV}}(F_{y,L}) + F_{y,0} \quad (11)$$

其中, H_{CONV} 表示第 y 个 STFEB 中的卷积层。

Swin Transformer 层 (STL) 是基于原 Transformer 层的标准多头自注意力, 是局部注意力和窗口机制

的转移, Swin Transformer 层结构如图 6 所示。

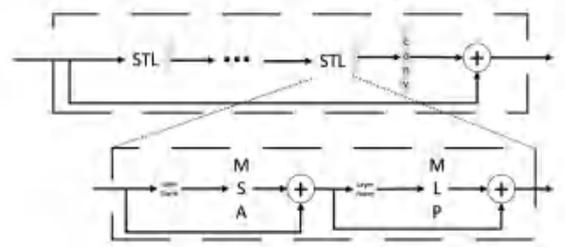


图 6 Swin Transformer 层结构

Fig. 6 Swin Transformer layer structure

给定大小的输入, 模块首先通过将输入划分为不重叠的局部窗口, 将输入重塑特征; 其次, 分别计算每个窗口的标准自关注即局部关注。一个局部窗口特征计算如下式:

$$\begin{cases} \mathbf{Q} = \mathbf{X}\mathbf{P}_Q \\ \mathbf{K} = \mathbf{X}\mathbf{P}_K \\ \mathbf{V} = \mathbf{X}\mathbf{P}_V \end{cases} \quad (12)$$

其中, $\mathbf{P}_Q, \mathbf{P}_K, \mathbf{P}_V$ 是在不同窗口共享的投影矩阵; $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 分别为查询向量、键向量、值向量; \mathbf{X} 表示局部窗口特征。

注意矩阵由局部窗口内的自注意机制计算得到, 如下式:

$$\text{attention} = (\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}(\mathbf{Q}\mathbf{K}^T/\sqrt{d} + \mathbf{B})\mathbf{V} \quad (13)$$

其中, B 为可学习的相对位置编码。

并行执行注意函数 h 次, 并将多头自注意 (MSA) 的结果连接起来。

多层感知器 (MLP) 具有两个完全连接的层, 这两个层之间具有非线性, 用于进一步的特征转换。在 MSA 和 MLP 之前都添加了 LayerNorm (LN) 层, 两个模块都使用了剩余连接, 如下式:

$$X = \text{MSA}(\text{LN}(X)) + X \quad (14)$$

$$X = \text{MLP}(\text{LN}(X)) + X \quad (15)$$

其中, MSA、MLP、LN 分别表示多头自注意力机制、多层感知器、归一化操作。

3 实验与分析

3.1 实验环境与参数设置

实验的硬件环境: 16 GB 运行内存, CPU 为 Intel Core i5, GPU 为 NVIDIA RTX 2080, 16 GB 显存。软件环境: Window10 操作系统、Anaconda2021、CUDA11.3、Pytorch1.9、Python3.9。

在浅层特征提取中, 输入图像通道为 3, 输出特征通道为 96, 步长为 1, 填充单层卷积网络。残差结构的个数设置为 16。实验训练采用 Adam 优化器, 参数 β_1 设置为 0.9, β_2 设为 0.99; epoch 设为 1 000, batch_size 设为 16, 初始学习率设置为 10^{-4} , 指数衰减率设置为 0.5, 每经过 200 次学习率衰减为原来的一半。

3.2 数据集

本文选择数据集 DIV2K 作为训练数据来训练本文提出的重建模型。选择 Set5, Set14, BSD100 和 Urban100 标准基准数据集进行测试。

3.3 结果对比分析

为了证明本文所提模型的有效性, 将训练模型重建图像的评估结果与以下经典模型评估结果进行比较: SRCNN^[6], FSRCNN^[7], VDSR^[8], SRGAN^[9], Memnet^[10], LESRCNN^[11] 和 IMDN^[12], 所有在 Bicubic 退化模型下的各基准数据集的对比结果见表 1。

表 1 各基准数据集的结果比较

Table 1 Comparison of results between benchmark dataset

模型	倍数	PSNR/SSIM			
		Set5	Set14	BSD100	Urban100
Bicubic	×2	33.66/0.929 9	30.24/0.868 8	29.56/0.834 1	26.88/0.840 3
SRCNN	×2	36.66/0.954 2	32.45/0.960 7	31.36/0.887 9	29.05/0.894 6
FSRCNN	×2	37.00/0.955 8	32.63/0.908 8		
VDSR	×2	37.53/0.958 7	33.05/0.913 0	31.90/0.896 0	30.76/0.914 0
SRGAN	×2				
LESRCNN	×2	37.65/0.959	33.32/0.915	31.95/0.896	31.45/0.921
MemNet	×2	37.78/0.960	33.28/0.914	32.08/0.898	31.31/0.920
IMDN	×2	38.00/0.961	33.63/0.918	32.19/0.900	32.17/0.928
RISR	×2	38.02/0.960	33.55/0.917	32.16/0.899	32.08/0.928
Bicubic	×3	30.49/0.929 9	30.24/0.868 8	29.56/0.738 5	26.88/0.840 3
SRCNN	×3	32.75/0.954 2	32.45/0.960 7	28.41/0.786 3	26.24/0.798 9
FSRCNN	×3	33.16/0.914 0	29.43/0.824 2		
VDSR	×3	33.67/0.921 0	29.78/0.832 0	28.83/0.799 0	27.14/0.829 0
SRGAN	×3				
LESRCNN	×3	33.93/0.923	30.12/0.838	28.91/0.801	27.70/0.841
MemNet	×3	34.09/0.925	30.00/0.838	28.96/0.800	27.56/0.838
IMDN	×3	34.36/0.927	30.32/0.842	29.09/0.805	28.17/0.852
RISR	×3	34.41/0.925	30.29/0.848	29.08/0.804	28.24/0.867
Bicubic	×4	28.42/0.810 4	26.00/0.702 7	25.96/0.667 5	23.24/0.657 7
SRCNN	×4	30.48/0.826 8	27.65/0.751 3	26.90/0.722 1	24.52/0.721 1
FSRCNN	×4	30.71/0.865 7	27.59/0.753 5		
VDSR	×4	31.35/0.883 0	28.11/0.762 4	27.29/0.716 7	25.18/0.896 0
SRGAN	×4	29.40/0.847 2	26.02/0.739 7	25.16/0.668 8	
LESRCNN	×4	31.88/0.890	28.44/0.777	27.45/0.731	25.77/0.773
MemNet	×4	31.74/0.892	28.26/0.772	27.40/0.728	25.50/0.763
IMDN	×4	32.21/0.895	28.58/0.781	27.56/0.735	26.04/0.787
RISR	×4	32.24/0.895	28.60/0.782	27.60/0.739	26.06/0.797

由表 1 可见,本文的 RISR 模型与其他模型相比,可以取得更好或相当的性能。对于 3 倍放大超分辨率, RISR 在 Set5 上峰值信噪比 (PSNR) 表现最好,结构相似性 (SSIM) 指标在 Set14 和 BSD100 上表现出最好结果。特别是对 4 倍 SR, RISR 在大多数基准数据集上的性能都远远优于其他模型,也证明了本文的模型具有更强的泛化能力。

对于图像的视觉质量比较,本文将重建的超分辨率图像与其他模型重建的图像进行视觉效果对比,如图 7 所示。

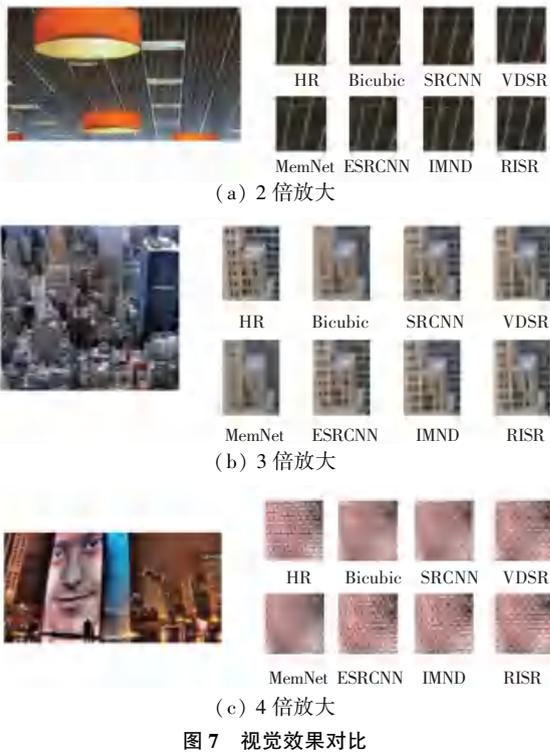


图 7 视觉效果对比

Fig. 7 Visual effect comparison

对于 2 倍放大,本文模型的重建图像有更高的对比度;对于 3 倍放大,大多数模型都容易产生伪影或生成的轮廓不够清晰,本文模型恢复了更准确的细节和更尖锐的轮廓;对于 4 倍放大,本文模型实现了其他模型无法恢复的线细节,可以清楚地看到其他模型生成的图片中液晶显示屏的失真,而本文模型恢复了接近真实的线轮廓。此外,本文模型重建的超分辨率图像具有更清晰的边缘和形状,进一步证明了本文模型可以提取更复杂的特征,能重建出更接近真实的图像。

3.4 消融实验

通过进一步的实验证明了所提方法的有效性,分别研究深层特征提取网络内部组件注意力机制倒残差倒置模块和 Swin Transformer 特征提取模块数量对实验结果的影响。

3.4.1 IRAB 模块对模型的影响

实验采用 DIV2K 数据集作为训练集。在 Set5、Set14、BSD100 测试集上,比例因子为 $\times 2$ 、 $\times 3$ 和 $\times 4$,不同数量的 IRAB 模块对模型的影响见表 2~表 4。

表 2 在 Set5 测试集上 IRAB 数量对模型的影响

Table 2 Impact of the number of IRABs on the model on the Set5 test set

倍数	IRAB	Set5
$\times 2$	2	37.88/0.959
	4(本文)	38.02/0.963
	6	38.00/0.961
$\times 3$	2	34.27/0.923 2
	4(本文)	34.41/0.932
	6	34.39/0.927
$\times 4$	2	31.99/0.880
	4(本文)	32.19/0.900
	6	32.12/0.887

表 3 在 Set14 测试集上 IRAB 数量对模型的影响

Table 3 Impact of the number of IRABs on the model on the Set14 test set

倍数	IRAB	Set14
$\times 2$	2	33.11/0.912
	4(本文)	33.73/0.921
	6	33.45/0.916
$\times 3$	2	30.06/0.835
	4(本文)	30.35/0.848
	6	30.30/0.841
$\times 4$	2	28.28/0.773
	4(本文)	28.60/0.782
	6	28.49/0.778

表 4 在 BSD100 测试集上 IRAB 数量对模型的影响

Table 4 Impact of the number of IRABs on the model on the BSD100 test set

倍数	IRAB	BSD100
$\times 2$	2	31.77/0.888
	4(本文)	32.20/0.901
	6	32.01/0.893
$\times 3$	2	28.57/0.801
	4(本文)	29.11/0.820
	6	29.09/0.815
$\times 4$	2	26.96/0.730
	4(本文)	27.58/0.750
	6	27.43/0.744

对于 2 倍、3 倍、4 倍的图像超分辨率任务,在 Set5、Set14、Urban100 这 3 个基准数据集上,当 IRAB 模块数量为 4 时的 PSNR 和 SSIM 指标总是优于该模块使用其他数量时的指标。这表明 RISR 的 IRAB 模块对提升图像超分辨率的性能指标有效,并且该模块数量为 4 时可以使 RISR 提高性能。

在放大倍数为 2,3,4 时 IRAB 数量对模块的影响如下:

在 2 倍放大时,在模块数为 4 之前,随着模块数

的增加,模型性能逐渐提高。在模块数为 4 之后,模型性能开始下降,为了提高模型性能,在其余实验中选择 IRAB 模块数为 4 的网络。

3 倍放大和 4 倍放大时,与二倍放大时类似,在

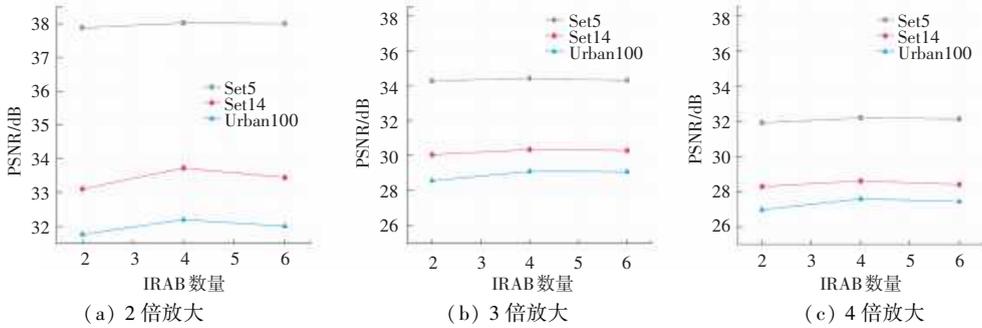


图 8 在放大倍数为 2,3,4 时 IRAB 数量对模块的影响

Fig. 8 Impact of IRAB quantity on modules at magnifications of 2, 3, and 4

总之,选择 IRAB 模块数量为 4 时,在放大倍数为 2,3,4 时,本文提出的重建模型可以有效提升模型的性能。

3.4.2 STFEB 模块对模型的影响

实验采用 DIV2K 数据集作为训练集。在 Set5, Set14, BSD100 上,比例因子为 $\times 2$ 、 $\times 3$ 和 $\times 4$,不同数量的 IRAB 模块对模型的影响结果见表 5~表 7, PSNR, SSIM 值越大表示重建结果越好。

表 5 Set5 测试集上 STFEB 数量对模型的影响

Table 5 Impact of the number of STFEBs on the model on the Set5 test set

倍数	STFEB	Set5
$\times 2$	6	37.85/0.958
	9(本文)	38.02/0.963
	12	37.96/0.959
$\times 3$	6	34.11/0.907
	9(本文)	34.41/0.932
	12	34.35/0.921
$\times 4$	6	31.87/0.889
	9(本文)	32.19/0.900
	12	32.04/0.893

表 6 Set14 测试集上 STFEB 数量对模型的影响

Table 6 Impact of the number of STFEBs on the model on the Set14 test set

倍数	STFEB	Set14
$\times 2$	6	33.28/0.911
	9(本文)	33.73/0.921
	12	33.59/0.914
$\times 3$	6	29.88/0.833
	9(本文)	30.35/0.848
	12	30.01/0.839
$\times 4$	6	28.27/0.771
	9(本文)	28.60/0.782
	12	28.44/0.774

模块数为 4 之前,模块的数量增加也带来了性能增益,但随着模块数量的增加,性能增益逐渐趋于最高。为了提高模型的性能,选择该模块数量为 4,如图 8 所示。

表 7 BSD100 测试集上 STFEB 数量对模型的影响

Table 7 Impact of the number of STFEBs on the model on the BSD100 test set

倍数	STFEB	BSD100
$\times 2$	6	31.65/0.886
	9(本文)	32.20/0.901
	12	32.03/0.890
$\times 3$	6	28.65/0.789
	9(本文)	29.11/0.820
	12	29.06/0.811
$\times 4$	6	26.96/0.721
	9(本文)	27.58/0.750
	12	27.48/0.743

STFEB 模块的消融实验结果如图 9 所示。使用数量为 9 的 STFEB 模块搭建的网络在 2 倍放大、3 倍放大和 4 倍放大的超分辨率任务上相比于使用模块数量为 6,12 的网络,PSNR 和 SSIM 都有所提升,表明残差倒置模块对于提升图像超分辨率的性能指标也是有效的。

在放大倍数为 2 时,在模块数为 9 之前,随着模块数的增加,模型性能逐渐提高;在模块数为 9 之后,模型性能开始下降,为了提高模型性能,本文在其余实验中选择 STFEB 模块数为 9 的网络。

放大倍数为 3 和 4 时,STFEB 数量对模型性能的影响与放大倍数为 2 时类似,在模块数为 9 之前,模块的数量增加也带来了性能增益,但随着模块数量的增加,性能增益逐渐趋于最高。为了提高模型的性能,本文选择该模块数量为 9。

总之,选择 STFEB 模块数量为 9 时,在放大倍数为 2,3,4 时,可以有效提升模型的性能。在 DIV2K 上训练本文提出的模型用于图像超分辨率,并在 Set5、Set14、Urban100 数据集上进行测试时,发现这些超参数的选择对于模型的性能至关重要。通过选择适当的超参数,可以获得更好的图像超分辨率效果。

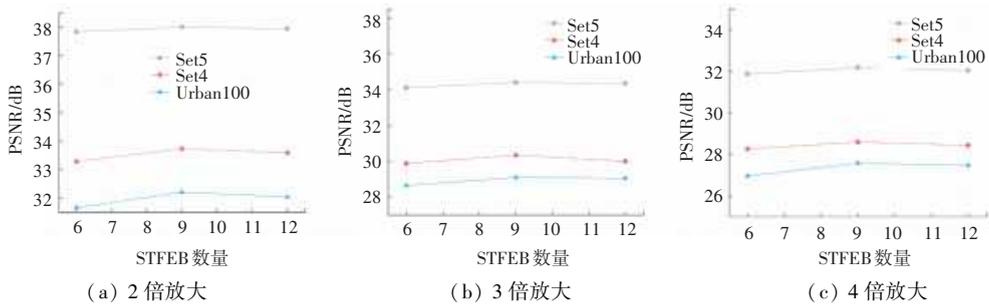


图9 放大倍数为2,3,4时STFEb数量对模块的影响

Fig. 9 Impact of STFEb quantity on modules at magnifications of 2, 3, and 4

4 结束语

本文提出了一种基于深度学习的图像超分辨率重建模型。该模型由3部分组成:浅层特征提取、深层特征提取和图像重建模块,有效地挖掘了浅层、深层的特征信息,并进行了高效的多尺度特征融合,提高了超分辨率任务的性能。该模型引入注意力机制、残差倒置模块来进一步提取特征,避免信息丢失,并通过多尺度特征融合增强特征表示,有效地从低分辨率图像中学习高频信息,强化网络对细节特征的提取。大量实验表明,本文提出的模型在经典图像超分辨率重建任务中,峰值信噪比和结构相似性较高,得到的重构图像纹理细节更清晰,提升了图像的视觉效果,证明了模型有效性。

参考文献

- [1] DONG C, LOY C C, HE K, et al. Learning a deep convolutional network for image super-resolution [C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2014: 184-199.
- [2] DONG C, LOY C C, TANG X. Accelerating the super-resolution convolutional neural network [C]// Proceedings of the 14th European Conference on Computer Vision. Cham: Springer, 2016: 391-407.
- [3] KIM J, LEE J K, LEE K. Accurate image super-resolution using very deep convolutional networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 1646-1654.
- [4] 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. Piscataway, NJ: IEEE, 2016: 1874-1883.
- [5] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2021: 10012-10022.
- [6] DONG C, LOY C C, HE K, et al. Learning a deep convolutional network for image super-resolution [C]//Proceedings of European Conference Computer Vision. Cham: Springer, 2014: 184-199.
- [7] CHAO D, CHEN CHANGE L, TANG X. Accelerating the super-resolution convolutional neural network [C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2016: 391-407.
- [8] KIM J, LEE K M. Accurate image super-resolution using very deep convolutional networks [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 1646-1654.
- [9] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C]//Proceedings of 2017 IEEE Conference on Compute Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2017: 105-114.
- [10] TAI Yang, YANG Jian, LIU Xiaoming, et al. MemNet: A persistent memory network for image restoration [C]//Proceedings of IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 4549-4557.
- [11] TIAN Chunwei, ZHUGE Ruibin, WU Zhihao, et al. Lightweight image super-resolution with enhanced CNN [J]. Knowledge-based systems, 2020, 205(12): 106235.
- [12] HUI Z, GAO X, YANG Y, et al. Lightweight image super-resolution with information multi-distillation network [C]// Proceedings of the 27th ACM International Conference on Multi-media (MM'19). New York: ACM, 2019: 2024-2032.