



中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION



基于树状多分支残差注意力网络的真实场景图像超分辨率重构

韩玉兰, 顾琴, 崔玉杰, 兰朝凤

引用本文:

韩玉兰, 顾琴, 崔玉杰, 兰朝凤. 基于树状多分支残差注意力网络的真实场景图像超分辨率重构[J]. 控制与决策, 2024, 39(10): 3442–3450.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.1004>

您可能感兴趣的其他文章

Articles you may be interested in

基于自注意力生成对抗网络的图像超分辨率重建

Image super-resolution reconstruction based on self-attention GAN

控制与决策. 2021, 36(6): 1324–1332 <https://doi.org/10.13195/j.kzyjc.2019.1290>

基于双分支特征融合的场景文本检测方法

A scene text detection based on dual-path feature fusion

控制与决策. 2021, 36(9): 2179–2186 <https://doi.org/10.13195/j.kzyjc.2020.0002>

一种基于多层语义特征的图像理解方法

An image understanding method based on multi-level semantic features

控制与决策. 2021, 36(12): 2881–2890 <https://doi.org/10.13195/j.kzyjc.2020.0927>

基于条件对抗生成孪生网络的目标跟踪

Conditional generative adversarial siamese networks for object tracking

控制与决策. 2021, 36(5): 1110–1118 <https://doi.org/10.13195/j.kzyjc.2019.1215>

Anchor-free的尺度自适应行人检测算法

Anchor-free scale adaptive pedestrian detection algorithm

控制与决策. 2021, 36(2): 295–302 <https://doi.org/10.13195/j.kzyjc.2020.0124>

基于树状多分支残差注意力网络的 真实场景图像超分辨率重构

韩玉兰[†], 顾琴, 崔玉杰, 兰朝凤

(哈尔滨理工大学 测控技术与通信工程学院, 哈尔滨 150000)

摘要: 真实场景图像超分辨率重构方法的研究进一步推动了超分辨率重构的应用, 成为研究的热点. 针对现有方法多采用单一输出代表高分辨率图像的高频特征细节, 难以保障稳定、准确高频细节的问题, 提出一种基于树状多分支残差注意力网络的真实场景图像超分辨率重构方法. 该方法通过树状结构形成多分支超分辨率重构网络, 增强特征表现能力, 进而丰富重构图像高频细节. 每条分支采用双通道残差策略对基础块进行连接, 允许更多低频特征通过. 进一步设计了基础块, 融入密集残差结构和注意力机制, 可以在加深网络的同时使网络在通道和空间上进行全局信息自适应调整. 面对树状分支的多个重构结果, 采用空间频率方法进行融合. 实验结果表明, 同当前先进的同类方法相比, 所提出方法具有更佳的重构效果.

关键词: 超分辨率; 真实场景; 注意力; 树状多分支结构; 双通道残差结构

中图分类号: TP391

文献标志码: A

DOI: 10.13195/j.kzyjc.2023.1004

引用格式: 韩玉兰, 顾琴, 崔玉杰, 等. 基于树状多分支残差注意力网络的真实场景图像超分辨率重构[J]. 控制与决策, 2024, 39(10): 3442-3450.

Real-world super-resolution based on residual attention network with tree based multi-branch structure

HAN Yu-lan[†], GU Qin, CUI Yu-jie, LAN Chao-feng

(School of Measurement and Control Technology and Communication Engineering, Harbin University of Science and Technology, Harbin 150000, China)

Abstract: The research on real-world super-resolution further promotes the application of super-resolution and becomes a hotspot. A real-world super-resolution method based on residual attention network with tree based multi-branch structure is proposed to solve the problem that a single output is difficult to ensure stable and accurate high-frequency details. Tree based structure is designed to form a multi-branch super-resolution network, which can enhance feature representation capability and enrich high-frequency details of the restored image. A dual residual path schema is proposed for basic blocks, which lets more low frequencies pass. The basic block adopts a dense residual module and attention mechanism to deepen the network and to adaptively adjust global information in both channel and space. Experimental results show that the proposed method can achieve better reconstruction than the current advanced similar methods.

Keywords: super-resolution; real-world; attention mechanism; tree network based multi-branch structure; dual channel residual structure

0 引言

单帧图像超分辨率 (super-resolution, SR) 重构是指从一幅低分辨率 (low resolution, LR) 图像中重构出同场景下的高分辨率 (high resolution, HR) 图像^[1]. 该技术通过图像处理方法实现, 无需更换或升级硬件设备, 在各领域具有广阔的应用前景. 目前, 随着深度学

习的发展, 基于深度学习的超分辨率重构方法克服了基于浅层学习提取纹理细节表达能力有限的底层特征来重构高分辨率图像的缺点, 具有较大的优势, 成为超分辨率重构领域的研究热点^[2-4].

2015年, Dong等^[5]率先将深度学习引入超分辨率重构问题中, 使图像重构效果得到了显著提升. 随后,

收稿日期: 2023-07-18; 录用日期: 2023-12-19.

基金项目: 国家自然科学基金项目(11804068); 黑龙江省自然科学基金项目(LH2020F033); 黑龙江省省属高等学校基本科研业务费项目(2020-KYYWF-0342).

[†]通讯作者. E-mail: hanyulan@hrbust.edu.cn.

大量研究者参与到基于深度学习的超分辨率重构研究中. Dong等^[6]对SRCNN的网络深度、卷积核大小等进行改进,以解决参数数量大的问题. Kim等^[7]借鉴了VGG-Net^[8]网络的架构并引入残差学习,提出了一个级联20层卷积层的残差网络. 受到DenseNet^[9]密集连接结构的启发,SRDenseNet^[10]通过密集跳跃连接有效结合了浅层和深层特征,缓解了梯度消失和卷积过程中图像特征逐级丢失的问题. Gao等^[11]利用不同大小的卷积核提取不同尺度的特征信息,解决了局部特征提取不足的问题. 这些模型虽然取得了较好的效果,但多数依赖于采用双三次线性插值方法构造的高低分辨率图像对,与实际退化模型之间存在较大差异,且重构的图像过于平滑,难以应用到真实场景中.

针对这一问题,更贴近实际的退化模型成为研究的焦点. Ji等^[12]通过估计模糊核以及真实噪声分布设计一种新的退化框架. MM-RealSR^[13]通过广义噪声和模糊核两个度量空间中的退化得分来构建现实场景下图像超分辨率的可调节交互机制. 面对真实场景图像退化过程的复杂多样,研究者们提出利用生成对抗网络来实现图像超分辨率重构. Ledig等^[14]首次将生成对抗网络引入图像超分辨率重构中,提出了SRGAN. ESRGAN^[15]在生成器上使用残差密集块来提升模型的效果. Wang等^[16]提出的Real-ESRGAN网络在合成数据集上训练,在预测模糊核过程考虑了振铃效应和压缩伪影. 基于GAN的方法取得了较好的视觉效果,但存在图像细节模糊、边缘结构扭曲的问题,并且这类退化模型受数据集的局限,无法很好地泛化到数据集以外的图像. 因而,真实场景下采集的高低分辨率图像对受到广泛关注. Cai等^[17]建立了RealSR数据集,为真实场景图像SR模型的训练和评估提供了通用基准. Wei等^[18]创建了一个大规模的真实数据集DRealSR. 同时,将可用算力的分配偏向于处理输入信号中信息量最大的部分是图像超分辨率重构的一种手段,因此将视觉注意力机制(attention mechanism)用于图像深层特征提取成为一个新的热点. Zhang等^[19]通过引入注意力机制来区别对待不同的信息通道,建立了一种基于残差通道注意力网络模型RCAN. Woo等^[20]提出了通道空间注意力模块(convolutional block attention module, CBAM),序列化地在通道和空间两个维度上产生注意力特征图信息. Li等^[21]提出的D2CSR算法引入了通道注意力机制提取图像深层特征,用半概率融合的方式复原更多高频细节.

注意力机制进一步提高了超分辨率的重构效果,但是以RCAN为代表的一类方法只注重通道注意力而忽略了空间注意力. 另外,多数方法使用单一输出来直接拟合所有的高频细节,难以保证准确、稳定的高频细节. 针对这一问题,本文提出树状多分支残差注意力网络(residual attention network with tree based multi-branch dtructure, RANTSMT)用于真实场景图像超分辨率重构. 主要工作包括:

1) 设计树状多分支超分辨率重构网络. 各分支分别输出具有不同高频成分的重构图像,多个重构图像通过空间频率融合法获得最终的重构图像,有效预防了单一输出高频成分容易缺失的问题.

2) 结合密集残差思想和通道空间注意力机制,提出密集残差通道空间注意力模块(residual dense convolutional block attention module, RDAM),增强特征提取能力的同时,可以在通道和空间上进行全局信息自适应调整.

3) 提出双通道残差结构,可以允许更多的低频特征通过,进一步增强特征传递.

1 网络模型

受注意力机制和D2CSR启发,本文提出一种树状多分支残差注意力网络用于图像超分辨率重构. 该网络整体框架如图1所示,分为浅层特征提取层、树状多分支SR层和特征融合层3个部分. 待重构的LR图像,首先经过卷积层提取浅层特征,然后提取的浅层特征经过树状多分支层提取图像深层特征,并重构获得多个含多种不同高频信息的HR图像,最后依据空间频率融合规则,对这些HR图像进行融合,获得重构的HR图像.

1.1 浅层特征提取

使用一个简单的卷积层提取LR的浅层特征 F_0 ,表达式为

$$F_0 = I_{LR} W_s^{3 \times 3} + b_s. \quad (1)$$

其中: I_{LR} 表示输入的低分辨率图像, $W_s^{3 \times 3}$ 表示提取浅层特征所使用的 3×3 的卷积核, b_s 表示偏置.

1.2 树状多分支结构设计

树状多分支结构是本文的重点. 由于网络不同深度、不同分支的特征图携带不同尺度的感受野信息和不同的高频信息,通过树状多分支特征重构获得的多个HR图像将具有多种不同的高频成分,将多个HR图像有机融合可以提高重构效果. 如图1所示,树状多分支SR层中,树状结构的每条分支主要包含深层特征提取模块和图像重构模块两个部分. 深层特

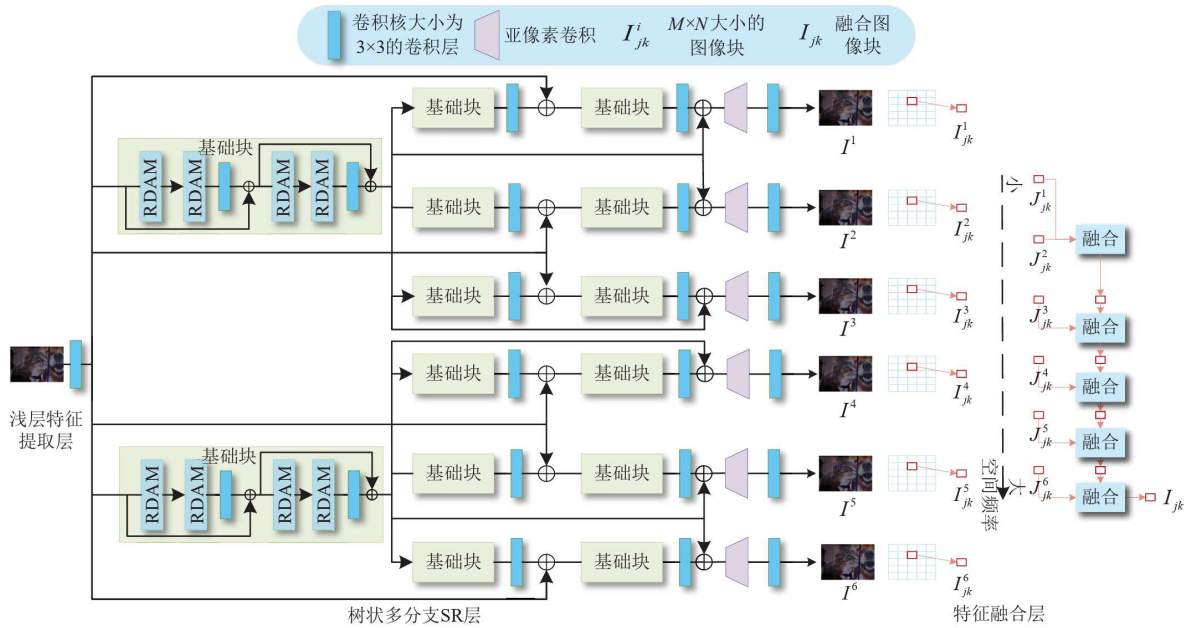


图1 RANTSM网络总体结构

征提取模块为由多个基础块构成的双通道残差结构,其中基础块由插入注意力机制的残差组组成,而每个残差组包含若干个密集残差通道空间注意力块RDAM,在获得重要特征信息的同时训练更深的网络。

1.2.1 RDAM模块

如图1所示,为了提取到更多有用的特征,在树状多分支SR层中,以引入注意力机制和密集残差结构的基础块为核心提取图像深层特征,其基础块为RDAM模块。RDAM的核心思想是将通道空间注意力机制加入密集残差块中,既可以保证密集连接机制能够缓解模型训练时出现梯度消失、加强特征传播和促进特征重用的优点,又可以学习不同特征层在不同位置、不同通道上的依赖性和差异,进而利用不同特征源进行残差学习使特征更加具有鉴别性。RDAM结构如图2所示,由3个密集残差块(residual dense block, RDB)和1个通道空间注意力模块CBAM组成,其中残差缩放系数 $\beta = 0.2$ 。

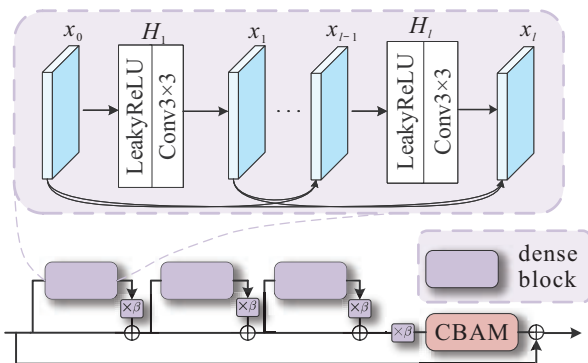


图2 密集残差通道空间注意力模块RDAM

密集连接模块(dence block)采用RCAN网络中的结构,由多个串联的卷积层组成,每层的输入都包含前面所有层输出的特征图。每层的输入输出关系可表示为

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \tag{2}$$

其中: $H_l(\cdot)$ 为非线性转化函数,使用包含conv和能增强残差块非线性特征学习能力的LeakyRelu组合操作;第 l 层的输入特征 $[x_0, x_1, \dots, x_{l-1}]$ 由第0层到第 $l-1$ 层的所有特征图进行张量拼接而成, x_l 为第 l 层的输出。

CBAM通道空间注意力机制如图3所示,由通道注意力和空间注意力两个模块组成。首先对输入特征进行通道注意力机制的处理,再对输出的通道注意力地图与输入特征加权后的特征层进行空间注意力机制的处理,从而得到最终的特征。

1.2.2 双通道残差结构

在卷积神经网络中,较浅的卷积层通常学到的是边缘和颜色等图像局部特征,而更深的卷积层则捕获更高级的语义信息。在浅层特征信息逐层映射为深层特征信息的过程中容易丢失部分浅层特征信息。另外,随着网络层数的加深,易产生模型退化问题。基本残差结构有效缓解了深度网络模型退化问题,并且采用跨层连接方式的深度网络能够被有效训练并取得较好的效果。受此启发,为了进一步增强特征传递,有利于特征重用,本文提出双通道残差结构,对如图1所示的树状分支结构中每一个分支所包含的3个基础块进行连接,具体如图4所示。

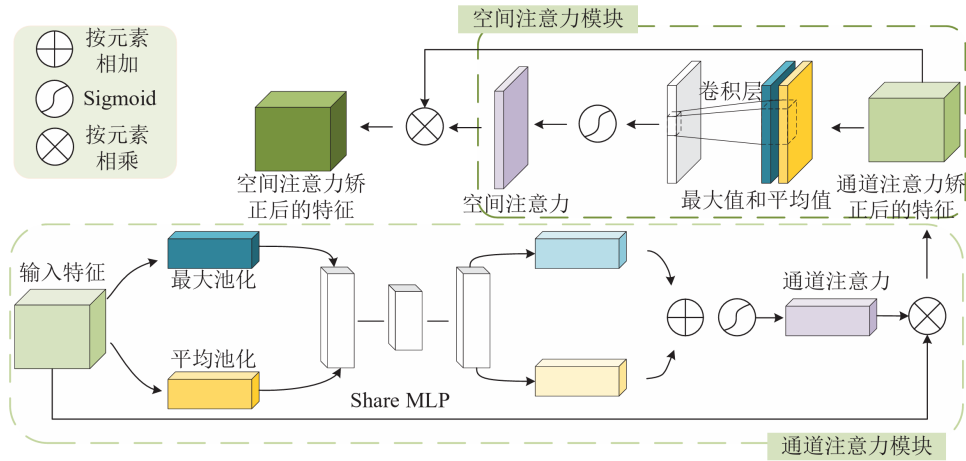


图3 通道空间注意力结构

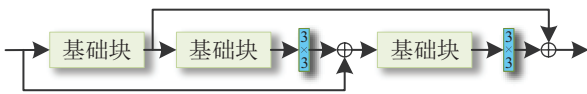


图4 双通道残差结构

该残差结构,第 m 个基础块的输入依赖之前所有基础块的输出,充分利用不同层次特征,结合图像位置、细节信息和图像整体语义信息来增强模型表现能力,训练过程中允许更多的低频成分通过.这个过程可以描述为

$$F_{m_o}^i = H_m^i(F_{m_{in}}^i), m = 1, 2, 3. \quad (3)$$

其中: $H_m^i(\cdot)$ 是第 i 条分支中第 m 个基础块的映射函数, $F_{m_{in}}^i$ 和 $F_{m_o}^i$ 分别为该基础块的输入和输出.

树状分支结构中,每条分支输出的特征表示为

$$F^i = H_1^i(F_0) + W_s^{3 \times 3} H_3^i(F_0 + W_s^{3 \times 3} H_2^i(H_1^i(F_0))). \quad (4)$$

其中: F_0 为输入LR图像的浅层特征, $W_s^{3 \times 3}$ 表示卷积核为 3×3 的卷积操作.

1.2.3 重构模块

如图1所示,图像重构部分由上采样层和卷积层构成.首先,通过亚像素卷积将输入的图像深层特征 F^i 映射到高分辨率空间.亚像素卷积是一种比较常用的上采样方法,与转置卷积相比,该方法不引入待学习参数,仅通过重新排列不同通道的特征图达到上采样的目的,速度较快,且上采样前不需要进行零填充.针对树状结构中的第 i 个分支,使用亚像素卷积对每一分支上的特征 F^i 做上采样,可获得与其对应的上采样后的特征图 F_{up}^i ,即

$$F_{up}^i = H_{up}(F^i), \quad (5)$$

其中 $H_{up}(\cdot)$ 表示使用亚像素卷积进行上采样操作.然后,使用一个卷积核大小为 3×3 的卷积层对上采样的特征 F_{up}^i 重构,获得重构后的图像

$$I^i = W_{rec}^{3 \times 3} F_{up}^i + b_{rec}, \quad (6)$$

其中 $W_{rec}^{3 \times 3}$ 和 b_{rec} 分别表示重构时的卷积核和偏置.

1.3 融合网络

对图像进行简单的融合常使用的是最大值法和加权平均法,这两种方法计算简单、运算速度快,能够较好地保留结果图像的整体效果,但是容易造成边缘、轮廓等细节信息严重丢失.本文特征融合层采用空间频率融合方法将多分支SR网络的多个预测结果融合:首先,将各分支的重构图像划分为 $M \times N$ 大小的图像块,并计算各块的空间频率,然后,根据空间频率对各分支重构图像相同位置的图像块进行融合,进而形成一个图像块.

对于任意图像块 $I_{jk}^i (i = 1, 2, \dots, P, P$ 为树状多分支SR层的输出分支数,本文取6)表示为第 i 个分支重构图像的 j 行 k 列的图像块,其空间频率表示为

$$SF_{jk}^i = \sqrt{(RF_{jk}^i)^2 + (CF_{jk}^i)^2}. \quad (7)$$

其中

$$RF_{jk}^i = \sqrt{\frac{1}{MN} \sum_{m=1}^M \sum_{n=2}^N [I_{jk}^i(m, n) - I_{jk}^i(m, n-1)]^2},$$

$$CF_{jk}^i = \sqrt{\frac{1}{MN} \sum_{m=2}^M \sum_{n=1}^N [I_{jk}^i(m, n) - I_{jk}^i(m-1, n)]^2}$$

分别为水平和垂直方向频率, $I_{jk}^i(m, n)$ 为 $M \times N$ 大小的图像块在位置 (m, n) 处的灰度值.

各分支重构图像相同位置的图像块,按其空间频率从小到大进行排序.假设各分支重构图像中第 j 行 k 列的图像块排序后表示为 $J_{jk}^1, J_{jk}^2, \dots, J_{jk}^P$, 从具有最小空间频率的图像块开始进行两两比较,按照下式的融合规则进行融合,得到最终的融合结果:

$$I^{jk} = F(J_{jk}^6, F(J_{jk}^5, F(J_{jk}^4, F(J_{jk}^3, F(J_{jk}^2, J_{jk}^1))))). \quad (8)$$

其中

$$F(x, y) = \begin{cases} x, & \text{SF}_x > \text{SF}_y + \text{th}; \\ y, & \text{SF}_x < \text{SF}_y + \text{th}; \\ \frac{x+y}{2}, & \text{otherwise} \end{cases}$$

表示图像块 x 、 y 的融合函数, SF_x 、 SF_y 表示图像块 x 、 y 的空间频率, th 为阈值. 需要说明的是, 本文 $P = 6$, 因此式(8)仅给出了本文的情况.

本文使用众数过滤器对融合结果进行验证并纠正, 设置 3×3 的单元格作为内核窗口. 通过遗传算法来寻找图像块的优化尺寸 $M \times N$, 最后通过对应的逆变换来获得最终的融合结果图像 I_{SR} .

1.4 损失函数

如图1所示, RANTSMM网络的损失主要来源于两个部分, 分别为树状多分支SR网络损失(以下简称树状多分支损失)和融合损失.

1.4.1 树状多分支损失

树状多分支损失函数由 L_2 损失和三元组损失(triplet loss) L_{Triplet} 组成. 为了使树状多分支网络中产生的 i 个重构结果在保证真实性的同时尽可能包含更丰富的高频细节, 需要令各分支的重构图像 I^i 与真实的HR图像 I_{HR} 更接近, 而各重构图像间有一定的差异. 因此, I^i 与 I_{HR} 之间使用 L_2 损失, 各分支重构图像 I^i 之间使用三元组损失. 在计算三元组损失时先将输入转化为能更好地表示纹理差异的残差域, 使网络关注于纹理空间的分化. 具体定义如下:

$$L_2 = \sum_{i=1}^P \|I^i - I_{\text{HR}}\|^2, \quad (9)$$

$$L_{\text{Triplet}} = \frac{\sum_{i=1}^P \sum_{j=1, j \neq i}^P \beta_{ij} \text{trip}(\text{res}_{I^i}, \text{zero}, \text{res}_{I^j})}{P(P-1)}. \quad (10)$$

其中: $\text{res}_{I^i} = |G(I^i) - G(I^{\text{HR}})|$; $\text{res}_{I^j} = |G(I^j) - G(I^{\text{HR}})|$; $G(x) = x - \mu_x / \sigma_x$, μ_x 和 σ_x 分别是 x 的均值和标准差; $\text{trip}(x, \text{zero}, y) = \max[d(x, \text{zero}) - d(x, y) + \text{margin}, 0]$, margin 为常数, $d(\cdot)$ 为欧氏距离计算算子. 设置衰减系数 $\beta_{ij} = \theta^{l-1}$, $l \in [1, L]$, $\theta \in (0, 1]$, zero 是零图. 输入彩色图片时, 使用 YCbCr 空间中的 Y 信道进行计算.

综合以上, 树状多分支损失描述为

$$L_T = L_2 + \alpha L_{\text{Triplet}}, \quad (11)$$

其中 $\alpha > 0$ 为平衡参数.

1.4.2 融合损失

融合损失函数使用 L_2 损失, 损失的优化目标是使生成的 I_{SR} 图像接近真实的HR图像 I_{HR} , 故融合损

失函数表示为

$$L_2^F = \|I_{\text{SR}} - I_{\text{HR}}\|^2. \quad (12)$$

2 实验与结果分析

2.1 实验设置

2.1.1 数据集

本文使用数据集 D2CRealSR^[21](<https://github.com/megvii-research/D2C-SR>) 中的 100 对图像对训练网络模型, 随机选取 20 对图像对作为验证集, 其余 80 对图像对作为训练集. 对于以上训练图像对, 在数据预处理阶段对其进行数据增强, 包括旋转、水平镜像和垂直翻转等操作, 最终参与训练的数据超过一万张. 其余测试集使用来自图像超分辨率重构任务场景的 4 个真实场景公开数据集 Set5^[22]、BSD100^[23]、RealSR^[17] 和 DRealSR^[18]. Set5 (<https://github.com/jbhuang0604/SelfExSR/tree/master/data>) 和 BSD100 (<https://github.com/niazwazir/BSD100>) 是图像超分辨率重构常用的两类测试集, 图像数量分别为 5 张和 100 张. 数据集 RealSR (<https://github.com/csjaic/RealSR/tree/master>) 是通过佳能 5D3 和尼康 D810 调整焦距获取同一场景下包 $\times 2$ 、 $\times 3$ 、 $\times 4$ 三种放大倍数的高低分辨率图像. DRealSR 数据集 (<https://github.com/xiezw5/Component-Divide-and-Conquer-for-Real-World-Image-Super-Resolution>) 与 RealSR 类似, 它由多种相机在自然场景下拍摄出 4 种不同分辨率的照片.

2.1.2 环境及参数设置

本文设置主分支数为 2, 每个主分支的子分支数为 3, 共计 6 个分支. 每个分支模块包括 2 个残差组, 每个残差组包含 2 个 RDAM 模块. 网络卷积层的卷积核大小为 3×3 , 步长为 1, 激活函数采用 LeakyReLU (leaky rectified linear unit). 使用 Adam (adaptive moment estimation) 算法作为树状多分支 SR 网络的优化器, 并设置指数衰减率为 0.9 和 0.999. 初始学习率设置为 10^{-4} , 而且在训练过程中不断下降, 每 2 000 步衰减一半. 先训练树状多分支 SR 网络直至稳定, 再进行融合网络的训练. 使用 NVIDIA2080Ti GPU 来训练网络, 软件环境基于 Python3.7, CUDA Toolkit 10.1, 网络模型在 Pytorch 框架下训练.

2.2 对比实验与结果分析

2.2.1 有参考图像验证实验

为了验证本文所提出方法的有效性, 本小节将在公开数据 Set5、BSD100、RealSR、DRealSR 测试集上分别进行 2 倍、4 倍和 8 倍超分辨率重构实验, 同时进行主客观分析, 并与双三次线性插值 (Bicubic)、

SRCNN^[5]、FSRCNN^[6]、VDSR^[7]、RealSR^[12]、MM-RealSR^[13]、ESRGAN^[15]、Real-ESRGAN^[16]、RCAN^[19]、D2C-SR^[21] 进行比较. 客观评价指标为结构相似度 (structural similarity, SSIM) 和峰值信噪比 (peak signal to noise ratio, PSNR).

表1~表3列出了在不同放大倍数下各方法的 PSNR 和 SSIM 的平均值, 其中重构效果最好的用粗体表示. 结果显示, Bicubic 方法由于缺乏局部差异的考虑, 在几种方法中表现最差; SRCNN、FSRCNN 方法重构效果优于 Bicubic, 但因网络设计简单提取不到过多图像深层特征而无法取得最佳重构效果. 与

简单的卷积神经网络方法相比, 受益于残差学习和注意力机制, VDSR 和 RCAN 取得了相对较好的重构效果. 值得注意的是, 使用 GAN 网络的方法在评价指标上并没有明显的提升, 但由于其判别器的优势可以生成视觉效果更好的重构图像. 使用真实数据集训练的模型均取得了很好的重构效果, 本文提出的模型表现最佳. 众所周知, 随着尺度放大因子的增加, 超分辨率重构的难度也在逐渐加大, 而本文方法在 8 倍重构时优势明显 (如表 3 所示). 由此可见, 本文提出的网络模型在高倍数超分辨任务上具有较强的重构能力.

部分重构结果如图 5~图 7 所示, Bicubic 方法重

表 1 不同方法各数据集的 PSNR 和 SSIM 均值比较 (×2)

算法	Set5		BSD100		RealSR		DRealSR	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
Bicubic	34.75	0.918	29.56	0.844	31.67	0.887	31.67	0.837
SRCNN ^[5]	36.66	0.924	31.36	0.887	32.34	0.875	32.13	0.859
FSRCNN ^[6]	37.00	0.946	31.53	0.892	32.52	0.893	32.34	0.876
VDSR ^[7]	37.74	0.951	31.90	0.880	32.79	0.903	32.65	0.869
RCAN ^[19]	38.17	0.934	32.41	0.858	33.83	0.923	32.93	0.879
ESRGAN ^[15]	37.75	0.943	32.03	0.881	33.80	0.922	32.70	0.889
Real-ESRGAN ^[16]	38.15	0.946	32.18	0.881	33.84	0.922	33.12	0.891
RealSR ^[12]	37.82	0.948	32.39	0.894	33.95	0.919	32.92	0.881
MM-RealSR ^[13]	37.69	0.943	32.21	0.883	33.89	0.920	33.34	0.889
D2CSR ^[21]	38.23	0.949	32.34	0.885	34.40	0.926	33.42	0.892
Ours	38.14	0.952	32.42	0.882	34.45	0.932	33.56	0.901

表 2 不同方法各数据集的 PSNR 和 SSIM 均值比较 (×4)

算法	Set5		BSD100		RealSR		DRealSR	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
Bicubic	28.42	0.810	25.96	0.667	27.24	0.764	29.56	0.820
SRCNN ^[5]	30.49	0.862	26.90	0.710	27.43	0.739	30.34	0.826
FSRCNN ^[6]	30.71	0.865	26.98	0.715	28.12	0.798	30.19	0.829
VDSR ^[7]	31.53	0.883	27.29	0.725	28.48	0.815	29.95	0.819
RCAN ^[19]	31.61	0.876	27.72	0.723	29.21	0.824	30.37	0.825
ESRGAN ^[15]	31.33	0.845	27.51	0.736	29.15	0.826	30.18	0.821
Real-ESRGAN ^[16]	31.55	0.861	27.75	0.772	29.59	0.835	30.65	0.846
RealSR ^[12]	31.53	0.865	27.78	0.743	29.76	0.833	30.45	0.832
MM-RealSR ^[13]	31.59	0.863	27.34	0.773	29.75	0.830	30.64	0.828
D2CSR ^[21]	31.67	0.875	27.23	0.775	29.72	0.831	30.80	0.825
Ours	31.68	0.887	27.79	0.780	29.78	0.836	30.84	0.838

表 3 不同方法各数据集的 PSNR 和 SSIM 均值比较 (×8)

算法	Set5		BSD100		RealSR		DRealSR	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
Bicubic	24.39	0.657	23.67	0.547	24.24	0.664	27.74	0.822
SRCNN ^[5]	25.33	0.689	24.13	0.565	24.34	0.689	27.98	0.834
FSRCNN ^[6]	25.41	0.682	24.21	0.567	24.72	0.673	28.34	0.839
VDSR ^[7]	25.72	0.711	24.37	0.576	25.38	0.697	28.46	0.844
RCAN ^[19]	25.65	0.693	24.19	0.583	25.82	0.711	30.26	0.868
ESRGAN ^[15]	26.29	0.743	24.56	0.595	26.25	0.716	29.67	0.847
Real-ESRGAN ^[16]	26.78	0.729	24.49	0.644	26.35	0.706	29.98	0.858
RealSR ^[12]	25.57	0.659	24.21	0.628	26.87	0.722	30.06	0.865
MM-RealSR ^[13]	26.76	0.762	24.45	0.639	26.74	0.719	30.12	0.865
D2CSR ^[21]	26.82	0.754	24.48	0.645	27.35	0.728	30.55	0.871
Ours	27.14	0.783	24.56	0.648	27.72	0.731	30.68	0.882



图 5 不同方法在 RealSR 数据集 Canon011 图像 4 倍超分辨率重构比较



图 6 不同方法在 RealSR 数据集 Nikon004 图像 4 倍超分辨率重构比较



图 7 不同方法在 D2CRealSR 数据集图像 8 倍超分辨率重构比较

构后的图像过于平滑,整体效果模糊,不能得到清晰的图像边缘和细节. RCAN 更倾向于恢复较为稠密的细节信息,容易导致图像模糊或不自然,并且随着放大倍数的提升视觉表现明显变差. RealSR 方法的重构结果产生了与真实纹理细节不相符的细节特征,对于字母很难捕捉纹理边缘细节. Real-ESRGAN 改善了图像模糊现象,获得了更清晰的图像,但是图像更加趋于动画感. 图 5 中可以看到对字母的还原丢失了图像的部分纹理信息. 与前几种方法相比, D2CSR 恢复了较多的高频细节,但边缘存在严重的伪影. 本文提出的方法能更好地拟合图像,在恢复纹理的同时还增强了重构图像的亮度,对边缘模糊的部分能重构较为清晰的边缘纹理,并且能去除边缘尖锐化现象,对高倍数的重构图像来说,强化了边缘细节的处理,

提高了视觉感知.

2.2.2 无参考图像验证实验

为了进一步验证本文方法的重构效果,本小节通过手机拍摄 20 张图像,构造一个低分辨率自建数据集,并在该数据集上进行无参考图像验证实验,这里采用自然图像质量评价器 (natural image quality evaluator, NIQE) 作为评价指标. NIQE 的值越低,图像失真程度越小,图像质量越好. 不同方法 8 倍超分辨率重构的 NIQE 结果如表 4 所示,部分重构结果如图 8 所示. 可以看出, Bicubic、SRCNN 和 FSRCNN 方法平均 NIQE 值较大,基本都在 10 以上. VDSR 和 RCAN 因引入了注意力机制和残差学习,一定程度上降低了 NIQE 值,并且视觉效果较好,减少了部分伪影. 虽然基于 GAN 网络的模型在 PSNR 和 SSIM 上表



图8 自制数据集不同方法图像8倍超分辨率重构比较

现不佳,但NIQE较低,使用真实数据集训练的模型表现基本高于其他模型,尤其是本文方法在无参考质量评价上取得了最佳的效果。

表4 自制数据集下不同方法NIQE均值比较(×8)

算法	NIQE
Bicubic	12.78
SRCNN ^[5]	11.34
FSRCNN ^[6]	10.83
VDSR ^[7]	11.04
RCAN ^[19]	9.62
ESRGAN ^[15]	8.61
Real-ESRGAN ^[16]	7.48
MM-RealSR ^[13]	6.98
D2CSR ^[21]	7.32
Ours	6.78

2.3 消融实验

2.3.1 模型结构分析

本文模型的核心为树状多分支SR网络,不同分支数和子分支数作为重要的参数将直接影响模型性能. 本文以RealSR数据集4倍超分辨率为例,保持其他参数不变,通过设置不同的主分支和子分支数量对树状多分支SR网络的宽度和深度进行实验评估,实验结果如表5所示. 需要说明的是,由于内存的使用大小与网络深度和宽度成正相关,基于内存的考虑,使用一个小的基线模型进行实验分析,模型包含一个残差组,每个残差组包含一个RDAM. 实验结果表明,使用2个主分支和3个子分支具有最佳重构效果。

表5 树状网络不同分支数对应的PSNR均值

子分支	主分支数		
	1	2	3
1	29.25	29.31	29.38
2	29.37	29.43	29.47
3	29.46	29.66	29.52

2.3.2 模型模块分析

本文提出了采用密集残差机制和注意力机制相结合的模块提取图像深层特征,为了研究模型中不同模块对最终重构结果的影响,本节以RealSR数据集4倍超分辨率为例,对密集残差机制和注意力机制的有效性进行分析. 实验结果如表6所示,其中×表示实验中没有使用该机制,√表示使用了相应机制. 实验结果表明,模型在同时加入密集残差机制和注意力机制时取得了最好的表现. 需要说明的是,本节仍然采用2.3.1所述的基线模型进行实验。

表6 含有不同模块对应的PSNR均值

密集残差机制	注意力机制	PSNR
×	×	27.32
×	√	29.12
√	×	28.98
√	√	29.66

3 结论

本文提出一种树状多分支残差注意力网络模型,实现了真实场景图像超分辨率重构. 为了获取准确、丰富的深层特征,设计了树状多分支重构网络,通过多分支重构结果融合增强特征表现能力,丰富重构图像的高频细节,同时利用密集连接和注意力机制加深网络的同时对特征图加权,改善了模型特征表达能力. 此外,为了更好地实现浅层特征重用,提出了双通道残差结构. 在多个数据集上的实验结果表明,相较于同类算法,本文方法在多个评价指标上均取得了较好的重构效果. 基于模型在移动端应用的考虑,接下来将进一步探索降低模型复杂度,在保证重构效果的同时构建更轻量化的模型。

参考文献(References)

[1] Yue L W, Shen H F, Li J, et al. Image super-resolution: The techniques, applications, and future[J]. Signal

- Processing, 2016, 128(11): 389-408.
- [2] 王雪松, 晁杰, 程玉虎. 基于自注意力生成对抗网络的图像超分辨率重建[J]. 控制与决策, 2021, 36(6): 1324-1332.
(Wang X S, Chao J, Cheng Y H. Image super-resolution reconstruction based on self-attention GAN[J]. Control and Decision, 2021, 36(6): 1324-1332.)
- [3] 宋昭漾, 赵小强, 惠永永, 等. 多级信息补偿的U型网络图像超分辨率重建算法[J]. 控制与决策, 2023, 38(9): 2479-2486.
(Song Z Y, Zhao X Q, Hui Y Y, et al. Image super-resolution reconstruction algorithm of U-shaped network based on multi-level information compensation[J]. Control and Decision, 2023, 38(9): 2479-2486.)
- [4] 吕恩辉, 王雪松, 程玉虎. 基于反卷积特征提取的深度卷积神经网络学习[J]. 控制与决策, 2018, 33(3): 447-454.
(Lv E H, Wang X S, Cheng Y H. Deep convolution neural network learning based on deconvolution feature extraction[J]. Control and Decision, 2018, 33(3): 447-454.)
- [5] Dong C, Loy C C, He K M, et al. Learning a deep convolutional network for image super-resolution[C]. European Conference on Computer Vision. Cham: Springer, 2014: 184-199.
- [6] Dong C, Loy C C, Tang X O. Accelerating the super-resolution convolutional neural network[C]. European Conference on Computer Vision. Cham: Springer, 2016: 391-407.
- [7] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[J/OL]. 2015, arXiv: 1511.04587.
- [8] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J/OL]. 2014, arXiv: 1409.1556.
- [9] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 4700-4708.
- [10] Tong T, Li G, Liu X J, et al. Image super-resolution using dense skip connections[C]. 2017 IEEE International Conference on Computer Vision. Venice, 2017: 4799-4807.
- [11] Gao F R, Wang Y, Yang Z Y, et al. Single image super-resolution based on multi-scale dense attention network[J]. Soft Computing, 2023, 27(6): 2981-2992.
- [12] Ji X Z, Cao Y, Tai Y, et al. Real-world super-resolution via kernel estimation and noise injection[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, 2020: 466-467.
- [13] Mou C, Wu Y Z, Wang X T, et al. Metric learning based interactive modulation for real-world super-resolution[C]. European Conference on Computer Vision. Cham: Springer, 2022: 723-740.
- [14] Ledig C, Theis L, Huszar F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 4681-4690.
- [15] Wang X T, Yu K, Wu S X, et al. ESRGAN: enhanced super-resolution generative adversarial networks[C]. European Conference on Computer Vision. Cham: Springer, 2019: 63-79.
- [16] Wang X T, Xie L B, Dong C, et al. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data[C]. 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal, 2021: 1905-1914.
- [17] Cai J R, Zeng H, Yong H W, et al. Toward real-world single image super-resolution: A new benchmark and a new model[C]. 2019 IEEE/CVF International Conference on Computer Vision. Seoul, 2019: 3086-3095.
- [18] Wei P X, Xie Z W, Lu H N, et al. Component divide-and-conquer for real-world image super-resolution[C]. European Conference on Computer Vision. Cham: Springer, 2020: 101-117.
- [19] Zhang Y L, Li K P, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]. European conference on computer vision (ECCV). Munich, 2018: 294-310.
- [20] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[C]. European Conference on Computer Vision. Cham: Springer, 2018: 3-19.
- [21] Li Y W, Huang H B, Jia L P, et al. D2C-SR: A divergence to convergence approach for real-world image super-resolution[C]. The 17th European Conference on Computer Vision. Tel Aviv, 2022: 379-394.
- [22] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding[C]. Proceedings of the British Machine Vision Conference 2012. Surrey, 2012: 1-10.
- [23] Arbelaez P, Maire M, Fowlkes C, et al. Contour detection and hierarchical image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5): 898-916.

作者简介

韩玉兰(1984—),女,讲师,硕士生导师,从事图像重构的研究, E-mail: hanyulan@hrbust.edu.cn;

顾琴(1997—),女,硕士生,从事图像重构的研究, E-mail: 18380126229@163.com;

崔玉杰(2000—),女,硕士生,从事图像重构的研究, E-mail: 1877314641@qq.com;

兰朝凤(1981—),女,副教授,博士生导师,从事AI语音处理技术的研究, E-mail: lanchaofeng@hrbust.edu.cn.