

控制与决策

Control and Decision

基于浅层特征调制的轻量级单幅图像超分辨率重建

程德强, 王子强, 张皓翔, 寇旗旗, 钱建生, 江鹤

引用本文:

程德强, 王子强, 张皓翔, 等. 基于浅层特征调制的轻量级单幅图像超分辨率重建[J]. *控制与决策*, 2024, 39(12): 4045-4054.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.1714>

您可能感兴趣的其他文章

Articles you may be interested in

[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>

[基于多层次特征的机械臂单阶段抓取位姿检测](#)

Single-stage grasp pose detection of manipulator based on multi-level features

控制与决策. 2021, 36(8): 1815-1824 <https://doi.org/10.13195/j.kzyjc.2019.1840>

[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>

基于浅层特征调制的轻量级单幅图像超分辨率重建

程德强¹, 王子强¹, 张皓翔¹, 寇旗旗², 钱建生¹, 江鹤^{1†}

(1. 中国矿业大学 信息与控制工程学院, 江苏 徐州 221116;

2. 中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116)

摘要: 浅层特征在超分辨率重建网络中扮演关键角色, 其中蕴含丰富的图像细节, 对准确估计深层特征具有明显参考价值. 然而, 研究者常常忽视浅层特征, 过度依赖深层模块堆叠和拓扑结构优化, 进而造成信息冗余. 为此, 提出一种轻量级超分辨率重建网络, 旨在探索浅层特征与深层特征的映射机制, 以提升重建质量. 首先, 通过利用浅层特征生成特征掩码, 引导深层特征的生成过程; 其次, 采用基于注意力机制的特征选择模块, 动态生成特征权重信息; 最后, 设计双分支特征增强学习模块, 平衡输出特征权重并增强特征融合能力, 进一步提升重建性能. 实验结果表明, 所提出的算法在国际通用数据集上显著提升了峰值信噪比和结构相似度指标, 同时具有较小的模型参数数量和卓越的视觉表现. 这些结果验证了所提出的轻量级超分辨率重建网络的有效性和优越性.

关键词: 超分辨率; 浅层特征调制; 特征掩码; 动态注意力; 双分支; 轻量级

中图分类号: TP391

文献标志码: A

DOI: 10.13195/j.kzyjc.2023.1714

引用格式: 程德强, 王子强, 张皓翔, 等. 基于浅层特征调制的轻量级单幅图像超分辨率重建[J]. 控制与决策, 2024, 39(12): 4045-4054.

Lightweight single image super-resolution based on shallow feature modulation

CHENG De-qiang¹, WANG Zi-qiang¹, ZHANG Hao-xiang¹, KOU Qi-qi², QIAN Jian-sheng¹, JIANG He^{1†}

(1. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China; 2. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: Shallow features play a pivotal role in super-resolution reconstruction networks, as they encompass intricate image details and offer explicit reference value for precise estimation of deep features. Nonetheless, researchers frequently disregard shallow features and excessively depend on deep module stacking and topology optimization, leading to redundant information. Consequently, this study introduces a lightweight super-resolution reconstruction network that endeavours to explore the mapping mechanism between shallow and deep features to enhance reconstruction quality. Firstly, shallow features are leveraged to generate feature masks that guide the generation process of deep features. Then, an attention-based feature selection module dynamically generates feature weight information. Finally, a dual-branch feature enhancement learning module is devised to balance the output feature weights and augment the feature fusion capability, thereby further improving the reconstruction performance. Experimental findings substantiate that the proposed algorithm significantly enhances the peak signal-to-noise ratio and structural similarity metrics on a widely used international dataset, while concurrently exhibiting a limited number of model parameters and exemplary visual performance. These outcomes validate the effectiveness and superiority of the lightweight super-resolution reconstruction network proposed in this study.

Keywords: super-resolution; shallow feature modulation; feature masks; dynamic attention; dual branch; lightweight

0 引言

单幅图像超分辨率 (single image super-resolution, SISR) 被广泛应用于计算摄影等领域^[1-2],

其目标是学习低分辨率 (low resolution, LR) 图像与高分辨率 (high resolution, HR) 图像间的映射关系. 近年来, 深度学习技术在 SISR 中取得了显著进展. Dong

收稿日期: 2023-12-11; 录用日期: 2024-04-24.

基金项目: 国家自然科学基金项目 (52204177, 52304182); 济宁市重点研发计划项目 (2023KJHZ007).

责任编辑: 张文安.

[†]通讯作者. E-mail: jianghe@cumt.edu.cn.

等^[3]率先将卷积神经网络 (convolutional neural networks, CNN) 用于SISR, 提出了SRCNN. 为了减少计算成本, Dong等^[4]删除了SRCNN中的预处理操作, 提出了它的轻量级版本. Lim等^[5]提出了增强深度超分辨率网络EDSR, 删除了SRResNet^[6]特征提取模块中的批归一化层, 显著提升了网络推理速度. Zhang等^[7]引入通道注意力机制, 构建了RCAN, 使网络能够有针对性地关注不同通道的特征, 提升了CNN的表达能力. 程德强等^[8]提出的多递归残差网络结构MCSR, 通过加大网络的深度, 获得了更好的重建效果. 赵小强等^[9]通过构建类金字塔结构, 学习多尺度信息用于图像重建. 程德强等^[10]提出一种基于多尺度空间注意力引导的图像超分辨率重建网络. 随着对注意力机制研究的加深, 不少研究工作^[11-13]关注到自注意力机制对解决长距离特征依赖的优势, 能进一步提高网络重建图像的能力, 但此项技术算力消耗巨大, 增加了网络的训练难度. 因此, 轻量级网络成为图像超分辨率领域的另一个研究热点. 轻量化模型因其容易部署、处理速度高等优点, 被广泛应用于现实场景中. Ahn等^[14]提出级联残差网络CARN, 在保持较少参数的同时实现了出色的性能. Hui等^[15]提出了一种信息蒸馏网络IMDN, 通过逐步提取分层特征来减少网络参数量. Liu等^[16]将其中的通道分割操作替换成卷积核大小为 1×1 的标准卷积, 提出了残差特征蒸馏网络RFDN, 降低了网络的参数量, 并提高了网络的性能. 宋昭漾等^[17]提出了一种U型网络加强对不同层次信息的利用. 为了加强特征的复用, 程德强等^[18]提出了一种多尺度密集连接网络. Huang等^[19]通过密集连接使得图像的早期特征得到重用, 但密集连接会使得网络训练的过程中出现大量冗余计算, 降低模型的推理速度. 为了缓解这个问题, Huang等^[20]在训练过程中对不重要的权重进行剪枝, 减少了冗余计算. 但无论如何, 上述网络利用浅层特征的方式都是通过跳跃连接, 且重用全部的浅层特征, 这无疑是低效的. 鉴于此, 本文提出基于浅层特征调制的轻量级图像超分辨率重建网络, 其主要工作可被归纳为: 1) 提出了浅层特征调制层 (shallow feature modulation layer, SFML), 建立了浅层特征与深层特征间高效的映射关系. 浅层特征通过SFML生成特征掩码矩阵 (α, β) , 完成对深层特征的调制. 2) 构建了基于注意力机制的动态特征选择模块, 模块动态生成两个注意力特征权重, 细化深层特征的纹理信息. 3) 探索了深层特征双分支结构, 配合特征选择模块生成的注意力权重, 融合双分支的特征

信息, 进一步增强网络的特征提取能力.

1 本文算法

本文所提出的浅层特征调制网络 (shallow feature modulation network, SFMN) 如图1(a)所示, 主要由3个部分组成: 浅层特征提取、深层特征提取和图像重建.

1.1 浅层特征提取

浅层特征提取由卷积核大小为 3×3 的标准卷积完成, 此过程可以表述为

$$F_s = H_{\text{conv3}}(I^{\text{LR}}). \quad (1)$$

其中: F_s 为提取到的图像浅层特征, $H_{\text{conv3}}(\cdot)$ 为卷积核大小为 3×3 的标准卷积的非线性映射关系, I^{LR} 为输入的低分辨率图像.

1.2 双分支深层特征提取

浅层特征往往带有图像轮廓、边缘等对图像重建任务具有重要意义的信息, 而目前主流网络的关注点在于优化网络拓扑, 忽视了对浅层特征的利用. 针对这一问题, 本文提出浅层特征调制层, 通过浅层特征指导深层特征的生成, 在网络保持轻量的情况下, 高效地利用浅层特征, 提高图像重建质量, 并设计基于浅层特征调制层的双分支深层特征提取模块. 考虑到网络的轻量化, 对每个模块的输入输出分别使用两个卷积核大小为 1×1 的标准卷积, 完成特征通道降、升维操作, 以降低网络的参数量.

1.2.1 浅层特征调制层

本文利用常见的特征生成网络 (由3个 3×3 的卷积和3个ReLU^[21]激活函数构成) 来进行浅层特征提取. 如图1(b)所示, 浅层特征调制层由4组 3×3 卷积和LReLU^[22]激活函数的组合 (下述为调制操作) 组成, 输入的浅层特征分别通过上下两组调制操作, 生成特征掩码矩阵 (α, β) . 通过特征掩码对完成对输入的深层特征进行调制, 即调制后的特征为 $\alpha \otimes$ 深层特征 $\oplus \beta$. 由于特征掩码矩阵是根据浅层特征生成, 因此无需进行预定义值, 而可以随着网络训练过程的推进, 自适应地对深层特征进行优化.

$$\alpha = \tau(H_{\text{conv3}}(\tau(H_{\text{conv3}}(F_s))))m, \quad (2)$$

$$\beta = \tau(H_{\text{conv3}}(\tau(H_{\text{conv3}}(F_s))))n, \quad (3)$$

$$F_m = \text{SFML}(F_d | (\alpha, \beta)) = \alpha \otimes F_d \beta. \quad (4)$$

其中: F_s 、 F_d 和 F_m 分别是提取到的浅层特征、待调制的深层特征以及调制后的深层特征, $\text{SFML}(\cdot)$ 为浅层特征调制层所学习到的非线性映射关系, $H_{\text{conv3}}(\cdot)$ 为 3×3 卷积, $\tau(\cdot)$ 为LReLU激活函数, \otimes 代表元素相乘,

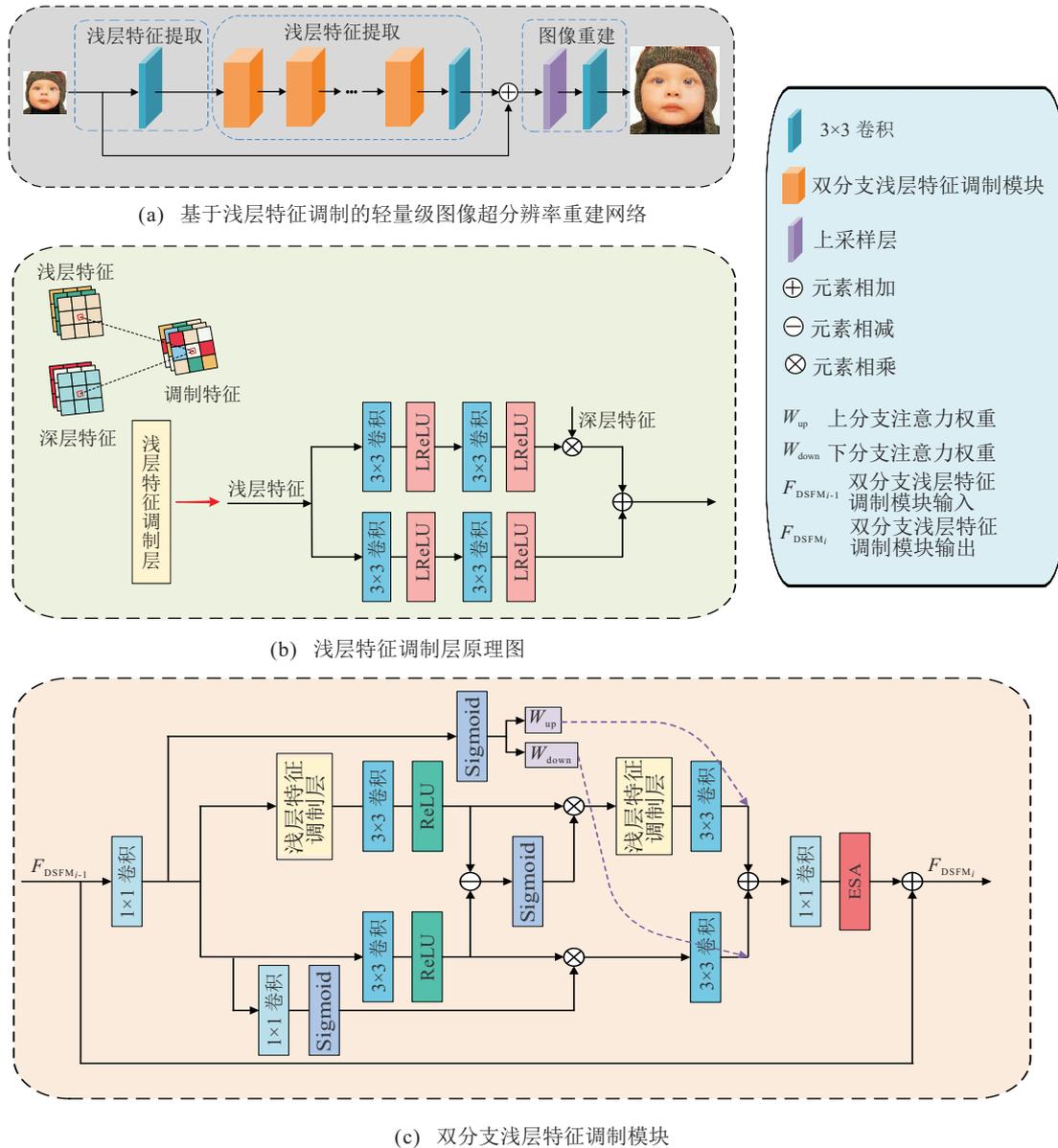


图1 基于浅层特征调制的轻量级图像超分辨率重建网络及实现细节

⊕代表元素相加。

1.2.2 深层特征提取模块

为了增强网络特征提取能力,提高训练过程中对浅层特征的高效利用,本文基于浅层特征调制层设计一个双分支模块用于提取深层特征。相比于单一分支的结构,双分支的特征提取结构强化了对不同深层特征的高效利用,从而提高网络重建能力。

图1(c)为双分支浅层特征调制模块。包含上下两个分支结构,上分支由2个浅层特征调制层、2个3×3的卷积以及一个ReLU激活函数组成,下分支仅由2个3×3的卷积和一个ReLU激活函数构成。

模块上下分支分别采用两种方式校准输出:上分支目的在于高效复用浅层特征,利用浅层特征对深层特征进行调制。为此,采用元素相减的方式进一步放大与下分支所学特征的差异,并用此差异进行上

分支特征校准。下分支则使用标准卷积操作完成深层特征提取,并采用自校准的方式建立特征之间的远程依赖关系。

规定第*i*个双分支浅层特征调制(dual-branch shallow feature modulation, DSFM)模块的输入为 $F_{DSFM_{i-1}}$, F_{SFML_1} 、 F_{conv3} 分别为上分支经过一次浅层特征调制所输出的特征和经过1次3×3卷积操作后得到的输出特征。为减少网络参数量,每个DSFM模块的输入都先经过1×1卷积进行降维操作来降低网络参数。

$$F_{SFML_1} = \sigma(H_{conv3}(H_{SFML}(H_{conv1}(F_{DSFM_{i-1}})))) \quad (5)$$

$$F_{conv3} = \sigma(H_{conv3}(H_{conv1}(F_{DSFM_{i-1}}))) \quad (6)$$

$$F_{diff} = F_{SFML_1} \ominus F_{conv3} \quad (7)$$

其中:两分支的特征差异 F_{diff} 用于对上分支进行校

准, $H_{\text{conv1}}(\cdot)$ 和 $H_{\text{conv3}}(\cdot)$ 分别为卷积核为 1×1 、 3×3 的标准卷积所学习的非线性映射关系, $H_{\text{SFML}}(\cdot)$ 为浅层特征调制层所学习的非线性映射关系, $\sigma(\cdot)$ 为 ReLU 激活函数, \ominus 为元素相减.

为了更好地融合上下两分支的输出特征, 本文设计一个基于注意力机制的特征选择模块, 通过该模块输出注意力权重, 通过动态权重来约束上下两分支的输出. 受到 A2N^[23] 的启发, 本文遵循了约束动态权重值的方法, 将两个注意力权重之和设置为 1.

$$W_{\text{up}} = \delta(H_{\text{conv1}}(F_{\text{DSFM}_{i-1}})), \quad (8)$$

$$W_{\text{down}} = 1 \ominus W_{\text{up}}. \quad (9)$$

其中: W_{up} 、 W_{down} 分别表示用于上下两分支的注意力权重, $\delta(\cdot)$ 表示 sigmoid 激活函数. 则第 i 个 DSFM 的输出 F_{DSFM_i} 可以表述为

$$F_{\text{up}} = H_{\text{conv3}}(H_{\text{SFML}}(\delta(F_{\text{diff}}) \otimes F_{\text{SFML}_1})), \quad (10)$$

$$F_{\text{down}} = H_{\text{conv3}}(\delta(H_{\text{conv1}}(F_{\text{DSFM}_{i-1}})) \otimes F_{\text{conv1}}), \quad (11)$$

$$F_{\text{DSFM}_i} = H_{\text{ESA}}(H_{\text{conv1}}(W_{\text{up}} \otimes F_{\text{up}} + W_{\text{down}} \otimes F_{\text{down}})) + F_{\text{DSFM}_{i-1}}. \quad (12)$$

其中: F_{up} 、 F_{down} 分别为上下两分支的输出, $H_{\text{ESA}}(\cdot)$ 为文献[24]中提出的增强空间注意力.

文献[24]表明, ESA 机制通常放在特征提取模块的末尾, 使网络关注更感兴趣的区域得到具有代表性的特征, 从而提高网络的学习能力. ESA 的轻量化结构是其被广泛使用的原因之一, 通过跨步卷积和通道降维操作, ESA 会大大降低其参数量和计算成本.

1.3 图像重建

如图 1(a) 所示, 图像重建部分包含上采样层和重建层. 上采样层采用图像超分辨率中常用的亚像素卷积, 它可以在不引入额外参数的情况下, 将低分辨率的特征映射到高分辨空间, 其原理如图 2 所示.

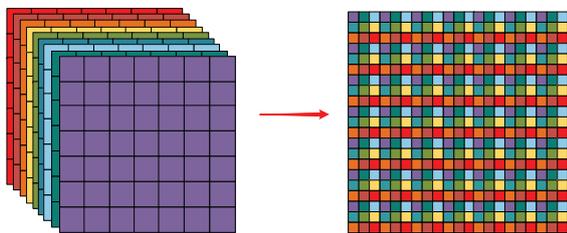


图 2 亚像素卷积示意图

1.4 损失函数

SISR 中常用的损失函数为 L_1 损失, 其数学表达式为

$$L_1 = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N |I^{\text{SR}}(x, y) - I^{\text{HR}}(x, y)|. \quad (13)$$

其中: $I^{\text{SR}}(x, y)$ 和 $I^{\text{HR}}(x, y)$ 分别代表重建图像和真实高分辨率图像, M 、 N 分别是图像的宽和高, L_1 则为计算出的 L_1 损失值.

2 实验结果与分析

2.1 数据集及实验环境

本文使用 DIV2K 数据集^[25] 作为训练数据集, 其中有 800 张训练图像、100 张验证图像、100 张测试图像. 本文采用 5 个基准数据集进行测试, 即: Set5^[26]、Set14^[27]、B100^[28]、Urban100^[29] 以及 Manga109^[30]. 本文采用随机 90° 和水平翻转的方式增强训练图像, 并将训练图像随机裁剪为 48×48 的图像片, 以获得足够的训练样本. 训练过程中, 每批次随机输入 16 个图像片. 网络训练采用 Adam 优化器, 其参数设置为 $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$. 通道数设置为 64, 初始学习率设置为 2×10^{-4} , 每 200 个迭代周期学习率减半, 总共运行 1 000 个迭代周期. 模型训练采用 Pytorch 深度学习框架, 在 NVIDIA 3090GPU 上进行训练与测试. 本文利用图像的 Y 通道来测试峰值信噪比 (peak signal to noise ratio, PSNR) 和结构相似度 (structural similarity, SSIM). 它们的计算公式如下:

$$\text{PSNR} = 10 \times \lg \left(\frac{\text{MAX}^2}{\text{MSE}} \right), \quad (14)$$

$$\text{SSIM} = \frac{(2\mu_1\mu_2 + C_1)(2\sigma_{12} + C_2)}{(\mu_1^2 + \mu_2^2 + C_1)(\sigma_1^2 + \sigma_2^2 + C_2)}. \quad (15)$$

其中: MAX 为 HR 图像中最大像素值, MSE 为均方误差, μ_1 和 μ_2 分别是 HR 图像和 SR 图像的亮度均值, σ_1 和 σ_2 分别为 HR 图像和 SR 图像的方差, σ_{12} 为 HR 图像和 SR 图像的协方差, C_1 和 C_2 为稳定计算所用常数项.

2.2 定量比较

为了验证本文算法的有效性, 将本文模型 SFMN 与主流的轻量级网络: EDSR-baseline^[5]、IMDN^[15]、RFDN^[16]、SMSR^[31]、LBNet^[32]、ShuffleMixer^[33]、RLFN^[34]、ARRFN^[35]、PILN^[36]、DRSAN^[37] 分别在客观评价指标和主观视觉效果两方面进行对比. 为了充分验证 SFMN 网络模型的效果, 分别在 3 种不同缩放因子下进行实验. 表 1 展示了不同超分辨率重建算法在 5 个基准数据集上取得的 PSNR 和 SSIM. 由表 1 中数据可以看出, 本文提出的 SFMN 在保持低参数量的情况下, 达到了较好的重建效果. 与 2023 年模型 PILN^[36]、DRSAN^[37] 相比, 除了在缩放因子为 2 的结果上本文算法有些许指标不是最优, 但均保持在次优水平. 这是由于相比于其他模型, 本文算法加强了对浅层特征的利用, 从而能在下采样倍数更高的低

分辨率图像中捕获更多的特征信息. 从3倍和4倍缩放因子下的重建指标中可以看出, 本文算法均优于其他算法. 在纹理细节信息较多的Urban100数据集上, 本文算法相比于次优算法分别提升0.03 dB/0.0015和0.11 dB/0.0017. 由表1可以看出, 本文算法在4倍

缩放因子下的结果取得最优效果, 并较之次优算法有着较大的提升. 在5个基准数据集Set5、Set14、B100、Urban100和Manga109上, 本文算法相比于次优算法分别提升: 0.11 dB/0.0009、0.07 dB/0.0012、0.04 dB/0.0009、0.11 dB/0.0017和0.18 dB/0.0018.

表1 3种缩放因子下网络在5个基准数据集上的测试指标对比

缩放因子	模型	参数量	Set5	Set14	B100	Urban100	Manga109
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
× 2	EDSR-baseline	1 370 K	37.99/0.960 4	33.57/0.917 5	32.16/0.899 4	31.98/0.927 2	38.54/0.976 9
	IMDN	694 K	38.00/0.960 5	33.63/0.917 7	32.19/0.899 6	32.17/0.928 3	38.88/0.977 4
	RFDN	534 K	38.05/0.960 6	33.68/0.918 4	32.16/0.899 4	32.12/0.927 8	38.88/0.977 3
	SMSR	985 K	38.00/0.960 1	33.64/0.917 9	32.17/0.899 0	32.19/0.928 4	38.76/0.977 1
	LBNNet	731 K	38.05/0.960 7	33.65/0.917 6	32.15/0.899 3	32.30/0.928 3	38.88/0.977 5
	ShuffleMixer	394K	38.01/0.960 6	33.63/0.918 0	32.17/0.899 5	31.89/0.925 7	38.83/0.977 4
	RLFN	527 K	38.07/0.960 7	33.72/0.918 7	32.22/0.900 0	32.33/0.929 9	—
	ARRFN	1 003 K	38.01/0.960 6	33.66/0.917 9	32.20/0.899 9	32.27/0.929 5	—
	PILN	580 K	38.08/0.960 7	33.72/0.918 1	<u>32.23/0.900 3</u>	<u>32.38/0.930 6</u>	39.92/0.977 1
	DRSAN	870 K	38.13/0.960 6	<u>33.72/0.918 9</u>	32.24/0.900 9	32.41/0.931 2	—
SFRN	939 K	<u>38.12/0.960 9</u>	<u>33.71/0.919 2</u>	<u>32.24/0.900 3</u>	<u>32.33/0.930 6</u>	<u>38.85/0.977 7</u>	
× 3	EDSR-baseline	1 555 K	34.37/0.927 0	30.28/0.841 7	29.09/0.805 2	28.15/0.852 7	33.45/0.943 9
	IMDN	703 K	34.36/0.927 0	30.32/0.841 7	29.09/0.804 6	28.17/0.851 9	33.61/0.944 5
	RFDN	541 K	34.41/0.927 3	30.34/0.842 0	29.09/0.805 0	28.21/0.852 5	33.67/0.944 9
	SMSR	993 K	34.40/0.927 0	30.33/0.841 2	29.10/0.805 0	28.25/0.853 6	33.68/0.944 5
	LBNNet	742 K	34.47/0.927 7	30.38/0.841 7	<u>29.13/0.806 1</u>	<u>28.42/0.855 9</u>	<u>33.82/0.946 0</u>
	ShuffleMixer	415K	34.40/0.927 2	30.37/0.842 3	29.12/0.805 1	28.08/0.849 8	33.69/0.944 8
	ARRFN	1 013 K	34.38/0.927 2	30.36/0.842 2	29.09/0.805 0	28.22/0.853 3	—
	PILN	588 K	34.39/0.926 9	30.34/0.841 5	29.08/0.804 8	28.09/0.850 0	33.68/0.944 6
	DRSAN	768 K	<u>34.50/0.927 8</u>	<u>30.39/0.843 7</u>	<u>29.13/0.806 5</u>	<u>28.35/0.856 6</u>	—
	SFRN	1 123 K	34.60/0.928 6	30.48/0.845 0	29.18/0.807 3	28.45/0.858 1	33.93/0.947 0
× 4	EDSR-baseline	1 518 K	32.09/0.893 8	28.58/0.781 3	27.57/0.735 7	26.04/0.784 9	30.35/0.906 7
	IMDN	715 K	32.21/0.894 8	28.58/0.781 1	27.56/0.735 3	26.04/0.783 8	30.45/0.907 5
	RFDN	550 K	32.24/0.895 2	28.61/0.781 9	27.57/0.736 0	26.11/0.785 8	30.58/0.908 9
	SMSR	1 006 K	32.12/0.893 2	28.55/0.780 8	27.55/0.735 1	26.11/0.786 8	30.54/0.908 5
	LBNNet	742 K	32.29/0.896 0	<u>28.68/0.783 2</u>	<u>27.62/0.738 2</u>	<u>26.27/0.790 6</u>	<u>30.76/0.911 1</u>
	ShuffleMixer	411 K	32.21/0.895 3	28.66/0.782 7	27.61/0.736 6	26.08/0.783 5	30.65/0.909 3
	RLFN	543 K	32.24/0.895 2	28.62/0.781 3	27.60/0.736 4	26.17/0.787 7	—
	ARRFN	1 024 K	32.22/0.895 2	28.60/0.781 7	27.57/0.735 5	26.09/0.785 8	—
	PILN	600 K	32.22/0.894 9	28.62/0.781 3	27.59/0.736 5	26.19/0.787 8	30.54/0.908 6
	DRSAN	747 K	<u>32.30/0.895 4</u>	<u>28.66/0.783 8</u>	27.61/0.738 1	<u>26.26/0.792 0</u>	—
SFRN	1 086 K	32.41/0.896 9	28.75/0.785 0	27.66/0.739 1	26.38/0.793 7	30.94/0.912 9	

注: 粗体为最优, 下划线为次优, — 为该模型未进行测试

2.3 主观效果对比

在主观视觉效果方面, 本文选取了纹理细节丰富的图片进行比较, 分别是4倍的缩放因子情况下, Urban100数据集上的Img033、Img093和Img100. 从图3(a)可以看出, 本文算法能够较好地保持大楼线条走向, 边缘更加清晰, 其余算法重建出的质量都存在线条扭曲的现象. 图3(b)更体现出本文算法的优势, 可以看出, 本文算法能够很好地重建出图中白色横线, 与原始真值图像保持一致, 而其他算法则存在将线条方向错误重建的问题. 从图3(c)中可以看出, 本文网络重建出的图像所受到的影响最小, 最接近原始图像, 图片中的“旗帜”线条较好地维持成垂直方

向. 因此, 得益于本文所提出浅层特征调制方法, 使得网络模型能够更好地处理图像的纹理细节, 图像纹理更细腻, 更接近真值图像, 视觉效果更好.

2.4 算法创新性和合理性

在SISR中, 大部分网络的工作集中在深层特征的优化上, 如图4(a)所示. 通过改进深层特征提取模块的结构和个数, 达到优化网络性能的目的, 而很少有人关注并有效利用浅层特征. 本文通过提出浅层特征调制层, 利用浅层特征生成掩码矩阵, 指导生成深层特征, 实现了对浅层特征的高效利用, 解决了上述盲点问题.

为验证浅层特征对深层特征具有指导意义, 本文

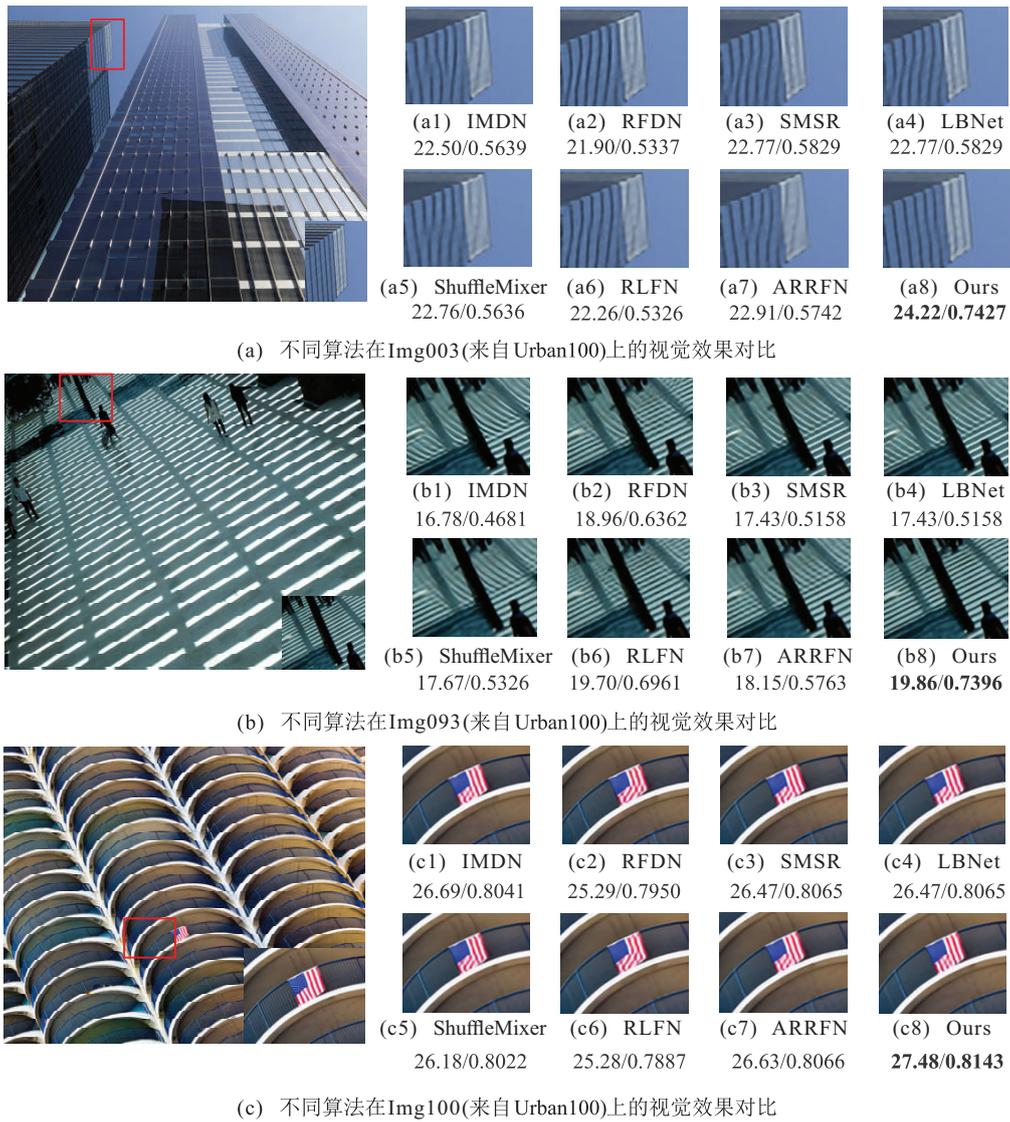


图3 标准测试集下4倍放大视觉效果比较

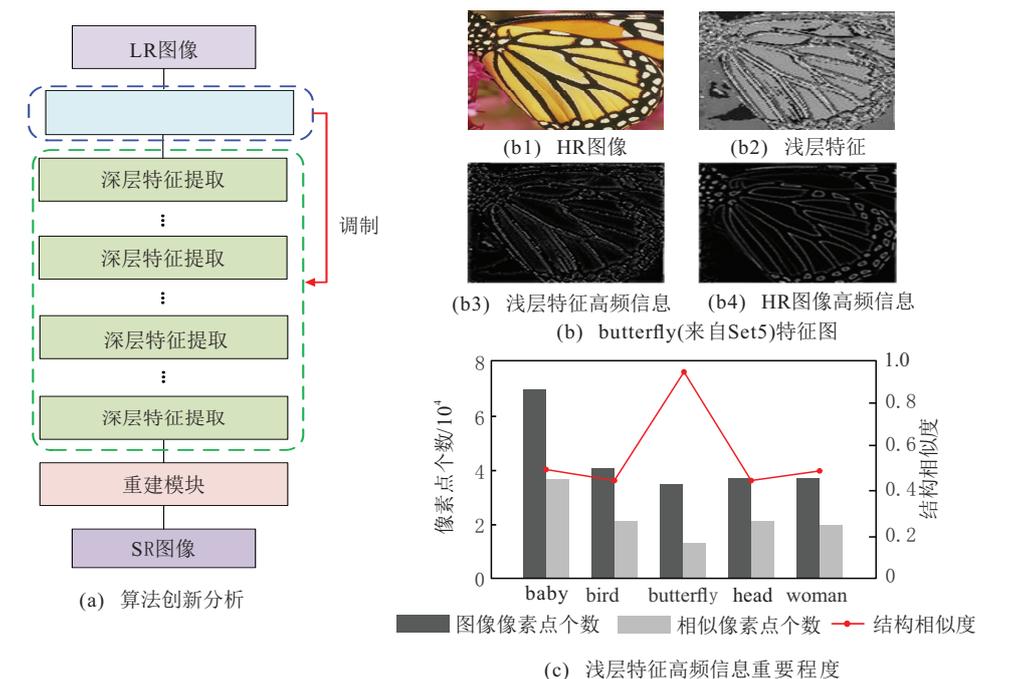


图4 算法创新性与合理性分析

对特征进行可视化操作. 如图4(b)所示, 浅层特征为图像通过一个 3×3 的卷积提取到的特征, 浅层特征和真实HR图像的高频信息, 均是经过高斯核函数提取的浅层特征得到的. 可以看出, 浅层特征的高频信息和真实图像的高频信息之间存在相似性. 为了进一步量化这种相似性, 本文选取Set5数据集为测试集, 提取Set5数据集中图像的浅层特征, 通过高斯核函数得到浅层特征的高频信息, 并将浅层特征高频信息与HR图像高频信息进行逐像素点相减并取绝对值. 若得到的结果位于 $0 \sim 5$ 之间, 则认为在该位置上, 浅层特征高频信息与HR图像高频信息相似. 如图4(c)所示, Set5数据集中的5张图片所提取出的浅层特征相似程度分别为: baby 58.24%、bird 46.58%、butterfly 35.89%、head 53.47%以及woman 45.94%, 平均相似程度为48.02%. 图4(c)中的红色折线部分为浅层特征高频信息与HR图像高频信息之间的结构相似度, 其计算结果依次为: 0.5148、0.4314、0.8390、0.4319、0.5117, 平均结构相似度为0.5457. 可以认为, 浅层特征的高频信息中, 含有大量与最终

真实图像相似的信息, 高效地对浅层特征进行利用, 可以在不需要堆叠网络层并减少网络参数的同时, 提高图像重建的质量. 这些量化数据表明本文所提出浅层特征对深层特征具有指导意义.

2.5 消融实验

本小节通过消融实验验证浅层特征调制层、基于注意力特征选择模块以及校准方法的有效性. 先作以下说明, DSFM为原网络的双分支特征调制模块, DSFM_woSFML为DSFM去除特征调制层后的模块; DSFM_woATT为DSFM去除注意力特征选择后的模块, DSFM_woCAL为DSFM去除两种校准方式后的模块. 本节分别将DSFM_woSFML、DSFM_woATT以及DSFM_woCAL替换原网络中的DSFM构建出3个网络进行训练, 并在纹理细节较多的Urban100数据集上进行二倍上采样重建. 训练过程中, 学习率设置为 1×10^{-4} , 训练批次设置为16, 图片大小设置为 50×50 , 模块数为5, 训练轮数为300轮. 实验结果如表2所示.

表2 不同方法之间有效性对比

模块名	参数量	浅层特征调制	注意力特征选择	分支校准	PSNR/dB	SSIM
DSFM	401 K	✓	✓	✓	31.21	0.9192
DSFM_woSFML	374 K	×	✓	✓	31.10	0.9177
DSFM_woATT	401 K	✓	×	✓	31.13	0.9178
DSFM_woCAL	396 K	✓	✓	×	31.15	0.9180

由表2可以看出, 对比DSFM和DSFM_woSFML可以得到, 网络参数量在仅增加27K的情况下, PSNR和SSIM分别提升了0.11dB, 0.0015, 表明增加浅层特征的利用有益于图像重建任务, 经过浅层特征指导生成的深层特征能够更好地用于图像重建; 基于注意力特征选择模块采用适合轻量级的像素注意力^[38](pixel attention, PA), 几乎未增加网络参数, 但动态权重平衡了两分支的输出特征, 高效地进行特征融合, 使得网络性能得到提升. 对比DSFM和DSFM_woATT可以看出, PSNR和SSIM分别提升0.08dB和0.0014. 根据DSFM和DSFM_woCAL对比可以看出, 有校准的模块比无校准的模块在PSNR上提升0.06dB, 在SSIM上提升0.0012. 因此, 可以认为两种不同的校准方式, 进一步放大了两分支特征之间的差异, 使得两分支可以学习到不同层次的特征信息, 提高网络重建性能. 本文采用图1(c)所示的双分支浅层特征调制模块结构, 为确定最优的模块数目做出以下对比实验. 表3展示了不同模块个数堆叠出的网络结构参数, 对不同模块数目的网络进行训练, 设置训练学习率为 1×10^{-4} , 训练批次为8, 图片大小设

置为 50×50 , 训练轮数为200轮, 并在4倍缩放因子下分别在Urban100数据集上进行测试.

表3 模块数量、网络参数对网络性能的影响

模块数量	参数量	PSNR/dB	SSIM
10 blocks	0.64 M	25.23	0.7553
12 blocks	0.84 M	25.29	0.7566
14 blocks	0.74 M	25.31	0.7589
16 blocks	0.94 M	25.36	0.7607
18 blocks	1.04 M	24.27	0.7237

注: **粗体**为最优

由表3中可以看出, 网络参数量增长与堆叠的模块数成正相关关系, 每增加两个模块会增加约0.2M的参数量, 但网络的拟合泛化能力并不与模块数量成正相关. 当模块数在10~16之间时, 网络性能与模块数成正比, 但继续增加模块会使得网络性能下降, 从16增加到18后, 模型PSNR下降1.09dB, SSIM下降0.037. 因此, 为保证网络的轻量化, 实验中, 选取16个双分支浅层特征调制模块进行深层特征的提取操作.

轻量级图像超分辨率网络一般采用减少通道数的方式减少网络的参数量. 为确定本文网络最优通

道数,即在保持低参数量的同时,保持重建性能,对通道数目进行消融实验对比. 实验设置模块数为3,学习率为 1×10^{-4} ,训练批次为8,训练轮数为200轮. 考虑到通道数对网络性能存在较大的影响,实验过程中并不考虑较低的通道数目,因此通道数分别设置为48、64和96,并在4倍缩放因子下分别在Urban100数据集上进行测试. 实验结果如表4所示,其中粗体为最优.

表4 通道数对网络性能的影响

通道数量	参数量	PSNR/dB	SSIM
48	0.18 M	25.01	0.721 7
64	0.30 M	25.14	0.722 6
96	0.67 M	25.25	0.723 1

由表4可以看出,通道数和网络参数量及性能呈

正相关. 通道数为64的网络相比于通道数为48的网络,PSNR提升0.13 dB,SSIM提升0.000 9. 同样,通道数为96的网络的客观指标PSNR和SSIM,相比于通道数为64的网络,分别提升0.11 dB和0.000 5. 综合考虑网络的性能的参数量,本文确定模块数目为16,通道数确定为64.

为了进一步验证浅层特征对于深层特征的生成具有指导意义. 本文做出以下实验,如图5所示. 首先,通过有无浅层特征调制操作,验证浅层特征对于指导深层特征生成的必要性;其次,将浅层特征调制层利用跳跃连接替代,验证所提出的浅层特征调制操作的高效性. 统一确定模块个数为5,学习率为 1×10^{-4} ,训练批次为8,训练轮数为200轮,并在4倍缩放因子下在4个标准数据集上进行测试.

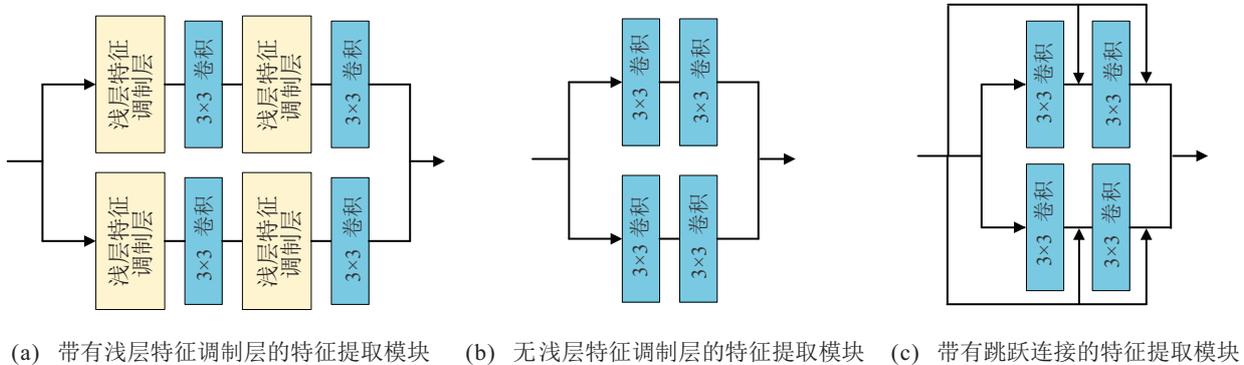


图5 浅层特征调制层消融实验

如图5所示,(a)、(b)、(c)三个子图所对应的模块名分别为:with_mod、without_mod和with_skip. 表5为(a)、(b)、(c)子图模块测试所得客观指标,从with_mod和without_mod指标可看出,当存在浅层特征调制时,由于调制层存在卷积操作,网络参数量增加31 K,计算量增加0.53 G,但网络性能得到大幅度提

高. 在4个基准数据集Set5、Set14、B100、Urban100上的PSNR分别增加1.21 dB、0.79 dB、0.42 dB、0.88 dB,SSIM分别提高0.027 1、0.016 3、0.010 2、0.035 1,表明了浅层特征对于深层特征具有指导意义,经过调制后的深层特征包含更加丰富的纹理信息,能够更好地用于图像重建任务.

表5 缩放因子为4时不同模块在4个基准数据集下的指标对比

模块名	参数量/KB	浮点计算次数/G	Set5	Set14	B100	Urban100
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
with_mod	399	1.03	31.24/0.879 0	27.98/0.763 0	27.16/0.718 3	24.96/0.741 4
without_mod	368	0.5	30.03/0.851 9	27.19/0.746 4	26.74/0.708 1	24.08/0.706 3
with_skip	378	1.03	30.71/0.869 9	27.65/0.757 6	26.99/0.715 1	24.58/0.727 6

注:粗体为最优

从without_mod和with_skip的对比中可看出,通过使用浅层特征到深层特征的跳跃连接,可以在牺牲网络参数和计算量的情况下,通过特征复用提高网络性能. 但通过对比with_mod和with_skip的实验结果能够看出,网络所增加的参数量相同,但使用本文所提出的浅层特征调制模块可以进一步提升网络的性

能指标,在数据集Set5、Set14、B100、Urban100上的PSNR分别增加0.53 dB、0.33 dB、0.17 dB、0.38 dB,SSIM分别提高0.009 1、0.005 4、0.003 2、0.013 8,所重建出的图像质量更好. 为了能够更加直观地看出重建图像的差异,本文选取Urban100数据集的Img092和B100数据集的148026进行主观视觉对

比,从图6中可以看到,with_mod网络重建出的图像最好,能够更好地保持图像边缘纹理信息。

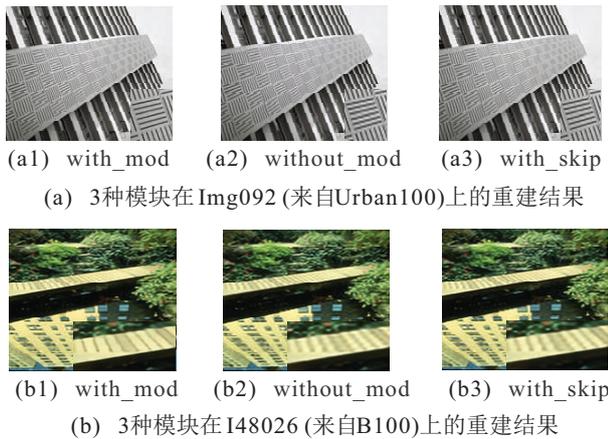


图6 2种网络4倍重建图像对比

3 结论

本文提出了一种基于浅层特征调制的轻量级网络,用于图像超分辨率重建。首先,设计了浅层特征调制层,引导深层特征生成,以改善浅层特征未能充分利用的问题,并丰富图像特征的纹理细节信息;其次,构建了双分支结构和注意力机制引导生成的权重,高效地融合多路输出特征。在5个基准数据集(Set5、Set14、B100、Urban100、Manga109)上进行测试,缩放因子为4。相比于次优算法,本文算法在峰值信噪比上的提升为0.11 dB、0.07 dB、0.04 dB、0.11 dB、0.18 dB;结构相似度提升为0.0009、0.0012、0.0009、0.0017、0.0018。

尽管本文所提出的模型表现优异,但仍存在局限性。本文的调制模型是一阶的,即调制特征和输入特征之间存在线性映射关系。显然,这种方式未必是最优的。未来将深入探索浅层特征与深层特征之间的高效映射关系,在保持网络轻量化的同时,追求更优的重建性能。

参考文献(References)

- [1] Jiang H, Asad M, Liu J J, et al. Single image detail enhancement via metropolis theorem[J]. *Multimedia Tools and Applications*, 2024, 83(12): 36329-36353.
- [2] Jiang H, Cai H K, Yang J. Learning in-place residual homogeneity for image detail enhancement[C]. *IEEE International Conference on Acoustics, Speech and Signal Processing*. Calgary, 2018: 1428-1432.
- [3] Dong C, Loy C C, He K M, et al. Image super-resolution using deep convolutional networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(2): 295-307.
- [4] Dong C, Loy C C, Tang X O. Accelerating the super-resolution convolutional neural network[M].

- Computer Vision—ECCV 2016. Cham: Springer International Publishing, 2016: 391-407.
- [5] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Honolulu, 2017: 1132-1140.
- [6] Ledig C, Theis L, Huszár 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: 105-114.
- [7] Zhang Y L, Li K P, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]. *Computer Vision—ECCV 2018*. Cham: Springer International Publishing, 2018: 294-310.
- [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.
(Cheng D Q, Guo X, Chen L L, et al. Image super-resolution reconstruction from multi-channel recursive residual network[J]. *Journal of Image and Graphics*, 2021, 26(3): 605-618.)
- [9] 赵小强, 王泽, 宋昭漾, 等. 基于类金字塔图残差网络的图像超分辨率重建[J]. *控制与决策*, 2024, 39(3): 786-794.
(Zhao X Q, Wang Z, Song Z Y, et al. Pyramid-like graph residual network for image super-resolution reconstruction[J]. *Control and Decision*, 2024, 39(3): 786-794.)
- [10] 程德强, 王培杰, 董彦强, 等. 基于多尺度空间注意力引导的图像超分辨率重建网络[J]. *北京航空航天大学学报*, DOI: 10.13700/j.bh.1001-5965.2023.0547.
(Cheng D Q, Wang P J, Dong Y Q, et al. Image super-resolution reconstruction based on multi-scale spatial attention guidance network[J/OL]. *Journal of Beijing University of Aeronautics and Astronautics*, DOI: 10.13700/j.bh.1001-5965.2023.0547.)
- [11] 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.)
- [12] Zhou Y P, Li Z, Guo C L, et al. SRFormer: Permuted self-attention for single image super-resolution[J/OL]. 2023, arXiv: 2303.09735.
- [13] Liu Q G, Gao P, Han K, et al. Degradation-aware self-attention based transformer for blind image super-resolution[J/OL]. 2023, arXiv: 2310.04180.
- [14] Ahn N, Kang B, Sohn K A. Fast, accurate, and lightweight super-resolution with cascading residual network[M]. *Computer Vision—ECCV 2018*. Cham: Springer International Publishing, 2018: 256-272.
- [15] Hui Z, Gao X B, Yang Y C, et al. Lightweight image super-resolution with information multi-distillation network[C]. *Proceedings of the 27th ACM International*

- Conference on Multimedia. New York, 2019: 2024-2032.
- [16] Liu J, Tang J, Wu G S. Residual feature distillation network for lightweight image super-resolution[C]. Computer Vision—ECCV 2020 Workshops. Cham: Springer International Publishing, 2020: 41-55.
- [17] 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.)
- [18] 程德强, 赵佳敏, 寇旗旗, 等. 多尺度密集特征融合的图像超分辨率重建[J]. 光学精密工程, 2022, 30(20): 2489-2500.
(Cheng D Q, Zhao J M, Kou Q Q, et al. Multi-scale dense feature fusion network for image super-resolution[J]. Optics and Precision Engineering, 2022, 30(20): 2489-2500.)
- [19] 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: 2261-2269.
- [20] Huang G, Liu S C, van der Maaten L, et al. CondenseNet: An efficient DenseNet using learned group convolutions[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 2752-2761.
- [21] Nair V, Hinton G E. Rectified linear units improve restricted Boltzmann machines[C]. International Conference on Machine Learning. Haifa, 2010: 807-814.
- [22] Xu B, Wang N Y, Chen T Q, et al. Empirical evaluation of rectified activations in convolutional network[J/OL]. 2015, arXiv: 1505.00853.
- [23] Chen H Y, Gu J J, Zhang Z. Attention in attention network for image super-resolution[J/OL]. 2021, arXiv: 2104.09497.
- [24] Liu J, Zhang W J, Tang Y T, et al. Residual feature aggregation network for image super-resolution[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, 2020: 2356-2365.
- [25] Agustsson E, Timofte R. NTIRE 2017 challenge on single image super-resolution: Dataset and study[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, Piscataway: IEEE, 2017: 1122-1131.
- [26] Higham D J, Higham N J. MATLAB guide[M]. The 3rd Edition. Philadelphia: Society for Industrial and Applied Mathematics, 2016.
- [27] 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. Surrey, 2012: 1-10.
- [28] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations[C]. Curves and Surfaces. Berlin, 2012: 711-730.
- [29] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]. Proceedings of 8th IEEE International Conference on Computer Vision. ICCV. Vancouver, 2001: 416-423.
- [30] Matsui Y, Ito K, Aramaki Y, et al. Sketch-based manga retrieval using manga109 dataset[J]. Multimedia Tools and Applications, 2017, 76(20): 21811-21838.
- [31] Wang L G, Dong X Y, Wang Y Q, et al. Exploring sparsity in image super-resolution for efficient inference[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, 2021: 4915-4924.
- [32] Gao G W, Wang Z X, Li J C, et al. Lightweight bimodal network for single-image super-resolution via symmetric CNN and recursive transformer[J/OL]. 2022, arXiv: 2204.13286.
- [33] Sun L, Pan J S, Tang J H. ShuffleMixer: An efficient convnet for image super-resolution[J]. Advance in Neural Information Processing Systems, 2022 35: 17314-17326.
- [34] Kong F Y, Li M X, Liu S W, et al. Residual local feature network for efficient super-resolution[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans, 2022: 765-775.
- [35] Qin J H, Zhang R M. Lightweight single image super-resolution with attentive residual refinement network[J]. Neurocomputing, 2022, 500: 846-855.
- [36] Qin J Y, Chen L H, Jeon S, et al. Progressive interaction-learning network for lightweight single-image super-resolution in industrial applications[J]. IEEE Transactions on Industrial Informatics, 2023, 19(2): 2183-2191.
- [37] Park K, Soh J W, Cho N I. A dynamic residual self-attention network for lightweight single image super-resolution[J]. IEEE Transactions on Multimedia, 2021, 25: 907-918.
- [38] Zhao H Y, Kong X T, He J W, et al. Efficient image super-resolution using pixel attention[C]. Computer Vision—ECCV 2020 Workshops. Cham: Springer International Publishing, 2020: 56-72.

作者简介

程德强(1979—), 男, 教授, 博士生导师, 主要研究方向为机器视觉、模式识别, E-mail: chengdq@cumt.edu.cn;

王子强(2000—), 男, 硕士生, 主要研究方向为图像超分辨率, E-mail: wangziqiang@cumt.edu.cn;

张皓翔(1994—), 男, 硕士生, 主要研究方向为图像超分辨率、模式识别, E-mail: zhanghx@cumt.edu.cn;

寇旗旗(1988—), 男, 副教授, 硕士生导师, 主要研究方向为图像修复、三维重建, E-mail: kouqiqi@cumt.edu.cn;

钱建生(1964—), 男, 教授, 博士生导师, 主要研究方向为机器视觉、模式识别, E-mail: qianjsh@cumt.edu.cn;

江鹤(1990—), 男, 讲师, 硕士生导师, 主要研究方向为图像修复、三维重建, E-mail: jianghe@cumt.edu.cn.