

DOI:10.19651/j.cnki.emt.2518792

基于空频特征调制的轻量级超分辨率网络设计^{*}

顾羽舟¹ 李娇¹ 郭爱英² 吴昊辰¹ 陆俊宇¹

(1. 上海大学机电工程与自动化学院微电子研究与开发中心 上海 200444; 2. 上海大学微电子学院 上海 201899)

摘要:为了克服现有基于自注意力机制的Transformer超分辨率模型在计算复杂度和局部细节捕捉方面的局限性,提出了一种优化的轻量级超分辨率网络结构,旨在协同利用全局、非局部和局部特征来实现更高效的重建。首先,构建了包含动态条带注意力与无偏差动态频域感知的空频特征聚合层用以捕捉全局与非局部特征,确保网络能充分地恢复图像特征。然后,构建了局部细节增强层以对局部上下文编码并进行通道混合,确保图像的细节恢复。最后,由多个空频特征调制块逐级提取特征并进行上采样重建以得到最终的超分辨率图像。本算法在Set14、BSD100、Urban100等5个超分辨率领域的公共数据集上进行了对比分析,其中,在2倍重建任务上,较同为轻量级超分辨率网络的ShuffleMixer相比,在减少了24.2%的FLOPs并使用更小训练集的同时,PSNR与SSIM在Urban100上分别高出了0.54 dB和0.0055。实验表明,提出的模型在轻量级超分辨率任务中表现优异,并在性能和复杂度之间取得了良好的平衡。

关键词:超分辨率;卷积神经网络;大核注意力;频域学习

中图分类号: TP391.4; TN791 文献标识码: A 国家标准学科分类代码: 520.60

Lightweight super resolution network based on spatial-frequency feature modulation

Gu Yuzhou¹ Li Jiao¹ Guo Aiying² Wu Haochen¹ Lu Junyu¹

(1. Microelectronics Research & Development Center, School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China; 2. School of Microelectronics, Shanghai University, Shanghai 201899, China)

Abstract: To overcome the limitations of existing Transformer-based super-resolution models, which rely on self-attention mechanisms and face challenges in computational complexity and local detail capture, an optimized lightweight super resolution network is proposed. The network aims to efficiently utilize global, non-local, and local features for enhanced reconstruction. First, a spatial-frequency feature aggregation layer, incorporating dynamic strip attention and unbiased dynamic frequency awareness, is used to capture global and non-local features, ensuring that the network can effectively recover image feature. Then, to ensure the restoration of image details, a local detail enhancement layer is constructed to encode local context and perform channel mixing. Finally, multiple spatial-frequency feature modulation blocks progressively extract features and perform up-sampling reconstruction to produce the final super-resolution image. The proposed algorithm was benchmarked on five public super-resolution datasets, including Set14, BSD100, and Urban100. Under the $\times 2$ reconstruction, it reduces FLOPs by 24.2% and requires a smaller training dataset compared with ShuffleMixer, another lightweight super-resolution network, while attaining gains of 0.54 dB in PSNR and 0.0055 in SSIM on the Urban100. Experiments show that the proposed network excels in lightweight super-resolution tasks, achieving a good balance between performance and complexity.

Keywords: super-resolution; convolutional neural network; large kernel attention; frequency learning

0 引言

是一个典型的病态问题,旨在将低分辨率(low resolution, LR)图像提升至所需的高分辨率(high resolution, HR)并尽可能还原其高频细节。自Dong等^[1]提出第一个基于卷

单图超分辨率(single image super-resolution, SISR)

积神经网络(convolution neural network, CNN)的 SISR 网络 SRCNN 以来,出现了许多优秀的端到端 SISR 方法^[2-3]。近年来,越来越多复杂且对硬件要求较高的算法被提出,如非局部注意力和 Transformer 模型所使用的自注意力机制等^[4-5],以及一些混合注意力机制^[6]。尽管这些模型在性能上表现优异,但其高昂的计算复杂度和庞大的资源需求,限制了其在实际场景中的应用。

为了缓解高计算成本这一问题,Shi 等^[7]利用亚像素卷积在进行特征处理后做可学习的上采样操作,由该方法抽象出的后采样型超分辨率网络架构一直沿用至今。Hui 等^[8]基于信息提取机制提出了具有对比度感知注意力层的信息多重蒸馏块。周登文等^[9]提出了一种多尺度的特征融合模块以提取多种尺度的特征,并通过通道搅乱注意力模块促进特征通道之间的信息流动以增强特征选择能力。杨胜荣等^[10]提出一种结合信息蒸馏及双链路上采样的超分辨率重建算法,首先通过多尺度信息蒸馏模块对特征进行多维度提取,其次蒸馏机制将多尺度特征进行选择性提炼,并将蒸馏出的部分特征利用多层次注意力机制进行全局复用,不仅降低了网络参数,还能获取更丰富的上下文信息。Sun 等^[11]采用了一种基于多尺度表示的特征调制机制,通过利用非局部特征关系动态选择代表性特征,取得了显著成果。李轩等^[12]针对轻量级网络存在的特征提取方式单一、中间层特征提取不充分等问题,利用多尺度卷积、通道压缩与坐标注意力等手段,提出了一种高效的通道可分离残差网络。然而,尽管这些轻量化算法在计算效率上表现优异,但由于其感受野受限,难以充分捕捉全局上下文信息,因此在复杂场景下的重建效果仍有待提升。

针对上述问题,本文提出了以空频特征调制块(spatial-frequency feature modulation block, SFFMB)为主体的轻量级超分辨率网络 SFFMNet。其中的空频特征聚合层(spatial-frequency feature aggregation layer, SFFAL)利用大核可分离卷积与核选择机制构成的动态条带注意力(dynamic strip attention, DSA)进行非局部特征提取,同时在 FFT 变换后的频域空间中利用提出的无偏差动态频域感知(unbiased dynamic frequency awareness, UDFA)从全局角度进行细节恢复。之后,通过由部分卷积与门控机制构成的局部细节增强层(local detail enhancement layer, LDEL)来实现跨通道交互并对局部特征进行建模,保证重建后的图像在指标与视觉效果上更加精准与自然。

1 基于空频特征调制的轻量级超分辨率网络 SFFMNet 设计

1.1 网络结构

所提出的基于空频特征调制的轻量级超分辨率网络 SFFMNet 如图 1(a)所示。其主要由 SFFMB 模块作为主体,同时为降低内存消耗^[13],网络采用多个 SFFMB 级联的方式构成,根据经验值,本文选择 8 个 SFFMB 级联作

为网络主体。

对于输入的 LR 图像,首先使用 3×3 的卷积作为浅层特征提取器将输入转换至特征空间。浅层特征经由网络主体提取更深、更抽象的高级特征。受 Transformer 架构在特征提取方面的启发,本文提出的 SFFMB 采用相似的设计原则,将层归一化(layer normalization, LN)、SFFAL、LN 和 LDEL 顺序堆叠以进行特征提取,并在前一组 LN、DSA 和后一组 LN、LDEL 之间添加残差连接来促进信息的高效传递和提升训练时的稳定性。该模块可表示为:

$$N_1 = LN(X) \quad (1)$$

$$Z = SFFAL(N_1) + X \quad (2)$$

$$N_2 = LN(Z) \quad (3)$$

$$\hat{Z} = LDEL(N_2) + Z \quad (4)$$

其中, $LN(\cdot)$ 代表对输入进行层归一化, $SFFAL(\cdot)$

与 $LDEL(\cdot)$ 为本文所提出的两个模块, X 与 \hat{Z} 为 SFFMB 的输入与输出,其余变量为中间变量。

之后,对深层特征利用上采样模块重建 HR 结果。为了保持网络的轻量化,仅使用一个 3×3 卷积层将通道维度转换为适应上采样比例的特定尺寸,并使用亚像素卷积操作进行放大。

最后,利用由 L1 像素损失和基于快速傅里叶变换(fast fourier transform, FFT)的频率损失构成的联合损失 $L = \|I^{SR} - I^{HR}\|_1 + \lambda \|F(I^{SR}) - F(I^{HR})\|_1$ 来训练 SFFMNet,其中 I^{SR} 和 I^{HR} 分别为超分辨率图像和原始高分辨率图像, F 为快速傅里叶变换, λ 为平衡参数并根据经验设置为 0.01。

1.2 空频特征聚合层

为了解决自注意力机制及其变体计算复杂度高的问题,本文设计如图 1(b)所示的轻量级空频特征聚合层(SFFAL)以高效的建模由空域和频域表示的非局部信息。尽管本文的主要动机在于利用频谱图的全局属性,但空域信息的利用仍是必要的。因此,分别提出了动态条带注意力分支(DSA)和无偏差动态频域感知(UDFA)分支。

1) 动态条带注意力

相关文献的表明,具有大核注意力(large kernel attention, LKA)的 CNN 模型可以在性能方面与基于 Transformer 的模型竞争^[14-15]。本工作受上述文献启发提出了一种动态条带注意力,以在空域分支中作为自注意力机制的替代。DSA 的结构如图 2(a)所示,其主要由条带大核注意力(large strip kernel attention, LSKA)与自适应核选择机制构成。

考虑到大卷积核虽然能直接带来感受野的提升,但参数与计算量也会成倍的增长。本工作在核层面与卷积层面均进行了等效分解,如图 2(b)所示。对于本文提出的条带大核注意力 LSKA,其参数量与浮点运算量为:

$$Param = 2 \times (2d - 1) \times C + 2 \times \frac{k}{d} \times C + C^2 \quad (5)$$

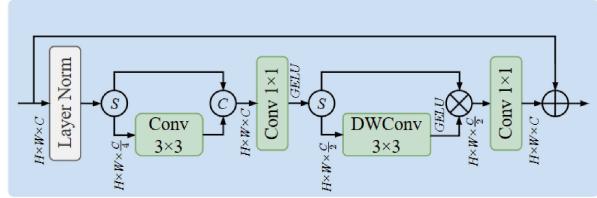
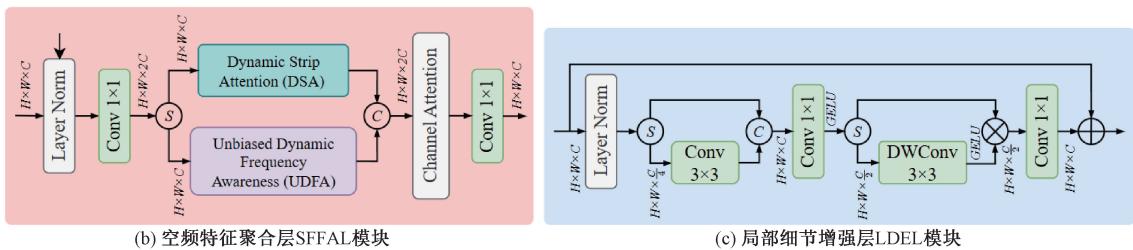
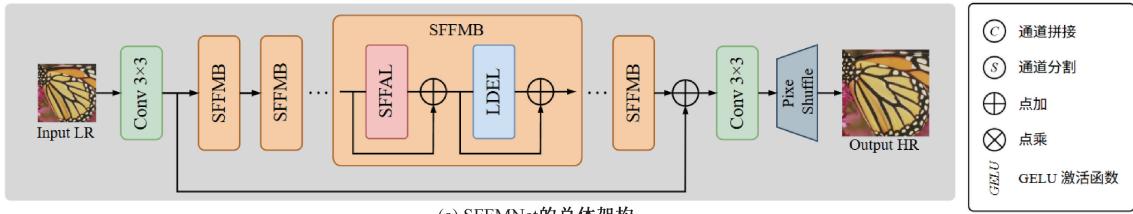


图 1 SFFMNet 总览

Fig. 1 Overview of SFFMNet

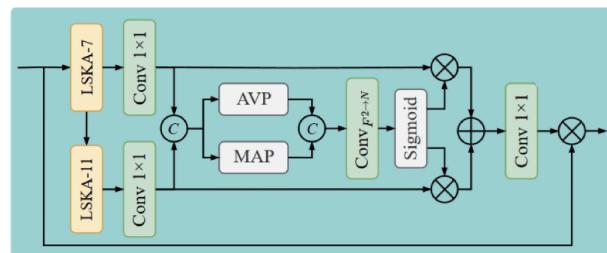
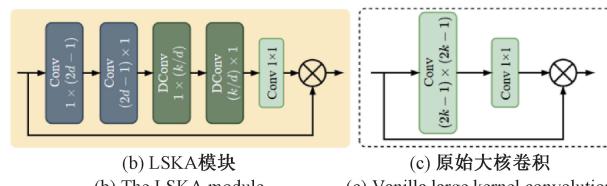
(a) DSA模块构成
(a) The DSA module

图 2 DSA 模块概览

Fig. 2 Overview of the DSA module

FLOPs =

$$H \times W \times (2 \times (2d-1) \times C \times C + 2 \times \frac{k}{d} \times C + C^2) \quad (6)$$

其中, k 为最终等效的卷积核大小, d 为卷积核的膨胀系数, C 、 H 和 W 分别为特征的通道数、高度和宽度。LSKA 相较图 2(c) 所示的原始大核卷积可以分别节省 $\frac{(2d-1)}{2}$ 和 $\frac{k}{2d}$ 的参数量和 FLOPs, 因此该方法不仅能够有效减轻计算压力,还能保留大卷积核的原始特性。

此外,由于不同类型的图像区域对上下文信息的依赖程度可能存在显著差异,本文利用了自适应核选择技术^[16],对于具有不同卷积核的多个分支,沿通道维度选择性的将它们融合在一起,让模型不仅能学习到不同通道间的关系,还能学习到不同大小卷积核之间的关系。此处,将

其与 LSKA 结合在一起以形成 DSA,如图 2(a)所示,以使得超分辨率模型可以利用不同感受野的上下文信息来实现更有效的恢复。

具体来说,在 DSA 包含两个卷积分支,前支使用 3×3 的原始条状卷积与膨胀率为 2 的 3×3 的条带卷积来实现 7×7 的感受野,后支接收前支输出的特征图并使用 3×3 的原始条状卷积与膨胀率为 2 的 5×5 条带卷积来实现 11×11 的感受野,后支所输出的特征图具有等效 17×17 的感受野。这使网络可以提取更大范围的特征,从而获得丰富的上下文信息。该过程可以表示如下:

$$Y = \text{Conv}_{(2d-1) \times 1}(\text{Conv}_{1 \times (2d-1)}(F_{in})) \quad (7)$$

$$Y = D\text{Conv}_{\left[\frac{k}{d}\right] \times 1}(D\text{Conv}_{1 \times \left[\frac{k}{d}\right]}(Y)) \quad (8)$$

$$\text{Att} = \text{Conv}_{1 \times 1}(Y) \quad (9)$$

$$Z = \text{Att} * Y \quad (10)$$

其中, F_{in} 与 Y 分别为 LSKA 的输入与输出, $\text{Conv}_{1 \times 1}$ 为逐点卷积, Att 代表注意力图, Z 则是最终的输出。之后,两组 DSA 的输出经 1×1 卷积降维并拼接后进行平均与最大池化以提取空间关系描述符,用卷积运算获得空间维度上的特征映射以转换为空间注意力图。对于每个空间注意力图再应用 Sigmoid 函数来获取每个 DSA 的通道权重。该过程可表示为:

$$\hat{Y}_1 = \text{Conv}_{1 \times 1}(F_{DSA_1}), \hat{Y}_2 = \text{Conv}_{1 \times 1}(F_{DSA_2}) \quad (11)$$

$$\tilde{Y} = \text{Concat}(\hat{Y}_1, \hat{Y}_2) \quad (12)$$

$$\hat{Y} = \text{Concat}(P_{avg}(\tilde{Y}), P_{max}(\tilde{Y})) \quad (13)$$

$$\hat{Y}_{DSA_1}, \hat{Y}_{DSA_2} = \text{Split}(\sigma(\text{Conv}_{7 \times 7}(\hat{Y}))) \quad (14)$$

随后,通过对权重和相应的特征图进行逐个元素乘以来获得加权的 DSA。最后,这些特征图由 1×1 卷积层融合,产生最终的空间注意力特征 Z 。调制后的结果 Z_s 为原始输

入 F 和注意力图 Y 的乘积:

$$Y = \text{Conv}_{1 \times 1}(\hat{Y}_1 \hat{Y}_{\text{DSA}_1} + \hat{Y}_1 \hat{Y}_{\text{DSA}_2}) \quad (15)$$

$$Z_s = F * Y \quad (16)$$

2) 无偏差动态频域感知

本文提出无偏差动态频域感知的动机在于:在其他频域工作中,如 FFC^[17]与 FFT-ReLU^[18]等,卷积算子被直接应用于频谱,然而频谱的每个元素都分别由实数和虚数组成,使用点卷积来跨双通道的处理会导致数据实部和虚部之间的交换,从根本上破坏了图像的空间结构和特征定位。因此,本文将频域特征的实部与虚部隔离后,用不同的卷积核来进行频域变换。此外,为了有效地提取图像退化后的频率特征,本文提出了两个核心模块:其一,考虑到频谱中的每个元素都会影响图像的整体结构,本文设计了频谱位置信息嵌入(spectrum position information embedding, SPIE)模块来保持重建时图像的原始结构;其二,提出了频域动态卷积(spectrum dynamic convolution, SDC)来对各频率分量进行不同的变换以增强建模的灵活性。由这两个核心模块构成的无偏差动态频域感知(UDFA)模块,如图 3 所示。

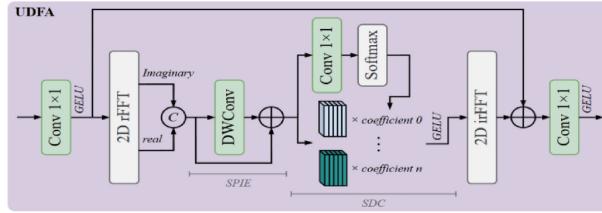


图 3 UDFA 模块

Fig. 3 The UDFA module

具体来说,输入特征首先通过 2 D-rFFT 获得频域特征,分别提取实部与虚部后进行通道维度上地拼接,产生的联合特征会被依次送入上文的两个核心模块,如图 3 所示。SPIE 以深度卷积进行建模并与残差相加以获得包含位置信息编码的频域特征,如式(17)所示。

$$F_{\text{SPIE}} = F_c + \text{DWConv}_{3 \times 3}(F_c) \quad (17)$$

其中, F_{SPIE} 为 SPIE 输出的结果, F_c 为实部与虚部拼接后的特征。

SDC 则先以点卷积与 Softmax 操作获取动态系数 $\text{coefficient}_0, \dots, \text{coefficient}_n$, 该系数与对应的卷积核相乘后,生成一系列可学习的卷积核,这些新卷积核将作为特定频域分量的实际卷积参数。其可以被表示为:

$$f_{\text{sdm}}^{u,v} = \sum_{i=1}^n \text{coefficient}_i^{u,v} \times \text{conv}_i \quad (18)$$

$$F_{\text{SDC}} = f_{\text{sdm}}^{u,v}(F_{\text{SPIE}}) \quad (19)$$

其中, u, v 分别代表频谱图中的坐标, $f_{\text{sdm}}^{u,v}$ 代表频域动态卷积操作, F_{SDC} 为 SDC 输出的结果。

1.3 局部细节增强层

上节所述的空频特征聚合层(SFFAL)侧重于捕获全局信息,缺乏对图像局部细节的捕获能力。因此,本文设计

了局部细节增强层(LDEL)来实现跨通道交互的同时对局部特征进行建模。考虑到常规的 FFN 缺乏空间维度上的信息交换且通道中存在大量冗余信息,本文利用部分卷积与门控机制对 FFN 进行了改进。

如图 1(c)所示,LDEL 首先进行部分卷积,将特征分为 $\{F_1^L, F_2^L\}$ 两部分,对 F_1^L 使用一个原始 3×3 卷积和 GELU 以强化特征中的局部信息。随后,经过处理的 F_1^L 和 F_2^L 在通道维度上拼接,并通过带有 GELU 的 1×1 卷积执行跨通道交互,同时扩展特征维度至隐藏空间。之后对特征进行门控,具体为扩展后的特征被再次分为两部分 X'_1 和 X'_2 ,其中 X'_1 经过一个 3×3 深度卷积和 GELU 以进一步编码局部特征。然后, X'_1 和 X'_2 通过元素乘法融合。最后,使用一个 1×1 卷积还原特征维度。LDEL 可以被表示为:

$$X' = \text{GELU}(\text{Conv}_{1 \times 1}(P\text{Conv}(F_{in}^L))) \quad (20)$$

$$Y = X'_1 \cdot DW\text{Conv}(X'_2), X' = [X'_1, X'_2] \quad (21)$$

$$Z_{\text{LDEL}} = \text{GELU}(\text{Conv}_{1 \times 1}(Y)) \quad (22)$$

其中, F_{in}^L 为 LDEL 的输入特征图, $[\cdot, \cdot]$ 表示通道分割, $P\text{Conv}(\cdot)$ 表示部分卷积, $DW\text{Conv}(\cdot)$ 表示深度卷积,其余元素为中间变量。通过通道分割和信息聚合,LDEL 在保留重要信息的同时减少冗余,简化了通道维度上的无用特征,进而提升了计算效率并与原始卷积构成的前馈网络相比拥有相近的性能。

2 分析与讨论

2.1 实验数据集与实现方法

1) 数据集

网络采用 DIV2K 数据集^[19]训练,并在常用的测试集上进行评估,包括 Set5、Set14、B100、Urban100 和 Manga109。评估时,SR 图像被转换至 YCbCr 颜色空间,并在图像的 Y 通道上计算 PSNR 和 SSIM,以评估恢复图像的质量。

2) 实现方法

在训练过程中,对训练集图像随机裁剪为大小 64×64 的图像块,并对 LR 图像进行随机水平翻转和旋转作为基本训练输入。进行实验的网络由 8 个 SFFMB 级联构成,采用 L1 Loss 和 FFT Loss 组成的联合损失函数进行训练,并通过 Adam 优化器进行优化, $\beta_1 = 0.9, \beta_2 = 0.99$ 。初始学习率为 5×10^{-4} ,最大小学习率为 1×10^{-6} ,并通过余弦退火策略进行更新。所有实验的迭代次数设置为 600 000。所有实验均在 NVIDIA GeForce RTX4070 GPU 上使用 PyTorch 框架进行。

2.2 与最新方法的比较

1) 定量比较

为了全面评估 SFFMNet 的性能,本节将 SFFMNet 与最新的基于 CNN 的各类 SR 方法以及部分基于 Transformer 的方法进行比较,包括 SAFMN^[11]、ShuffleMixer^[20]、

LAPAP^[21]、FDIWN^[22]、IMDN^[23]、EDSR^[24]与 ESTR^[25]。表 1 展示了在基准数据集上对于 $\times 2$ 、 $\times 3$ 和 $\times 4$ 放大因子的定量

比较结果,此外还列出了所使用的训练数据集、参数数量(Params)和浮点运算次数(FLOPs)。

表 1 不同数据集上,对于放大倍率 2、3 和 4 的平均 PSNR/SSIM 结果

Table 1 Average PSNR/SSIM for scaling factors 2, 3, and 4 across the various datasets

放大倍率	算法	训练	# Params	# FLOPs	Set5	Set14	BSD100	Urban100	Manga109
		数据集	/K	/G	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
$\times 2$	SAFMN	DF2K	228	52	38.00/0.960 5	33.54/0.917 7	32.16/0.899 5	31.84/0.925 6	38.71/0.977 1
	ShuffleMixer	DF2K	394	91	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
	LAPAP	DF2K	548	171	38.01/0.960 5	33.62/0.918 3	32.19/0.899 9	32.10/0.928 3	38.67/0.977 2
	FDIWN	D2K	629	—	38.07/0.960 8	33.75/0.920 1	32.23/0.900 0	32.40/0.930 5	38.85/0.977 4
	IMDN	D2K	694	132	38.00/0.960 5	32.63/0.908 8	31.53/0.892 0	29.88/0.902 0	36.67/0.969 4
	EDSR	D2K	1370	316	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
	ESTR	D2K	678	1116	38.03/0.960 0	33.75/0.918 4	32.25/0.900 1	32.58/0.931 8	39.12/0.977 4
	SFFMNet(Ours)	D2K	444	69	38.10/0.961 3	33.82/0.919 6	32.25/0.900 1	32.43/0.931 2	38.94/0.977 6
$\times 3$	SAFMN	DF2K	233	23	34.34/0.926 7	30.33/0.841 8	29.08/0.804 8	27.95/0.847 4	33.52/0.943 7
	ShuffleMixer	DF2K	415	43	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
	LAPAP	DF2K	594	114.4	34.36/0.926 7	30.34/0.842 1	29.11/0.805 4	28.15/0.852 3	33.51/0.944 1
	FDIWN	D2K	645	51	34.52/0.928 1	30.42/0.843 8	29.14/0.806 5	28.36/0.856 7	33.77/0.945 6
	IMDN	D2K	703	72	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
	EDSR	D2K	1555	160	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
	ESTR	D2K	770	835	34.42/0.926 8	30.43/0.843 3	29.15/0.806 3	28.46/0.857 4	33.95/0.945 5
	SFFMNet(Ours)	D2K	449	31	34.54/0.928 5	30.55/0.844 0	29.18/0.806 5	28.38/0.853 6	33.80/0.946 1
$\times 4$	SAFMN	DF2K	240	14	32.18/0.894 8	28.60/0.781 3	27.58/0.735 9	25.97/0.780 9	30.43/0.906 3
	ShuffleMixer	DF2K	411	28	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
	LAPAP	DF2K	659	95	32.15/0.894 4	28.61/0.781 8	27.61/0.736 6	26.14/0.787 1	30.42/0.907 4
	FDIWN	D2K	664	28	32.23/0.895 5	28.66/0.782 9	27.62/0.738 0	26.28/0.791 9	30.63/0.909 8
	IMDN	D2K	715	41	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
	EDSR	D2K	1518	114	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
	ESTR	D2K	752	298	32.19/0.894 7	28.69/0.783 3	27.69/0.737 9	26.39/0.796 2	30.75/0.910 0
	SFFMNet(Ours)	D2K	456	18	32.33/0.896 5	28.73/0.784 5	27.64/0.738 8	26.32/0.784 5	30.72/0.911 5

所提出的 SFFMNet 在所有基准数据集上的性能均优于其他 CNN 方法。例如,在 Urban100 数据集上,SFFMNet 用更少的参数、FLOPs 与更低的训练成本以 0.33 dB 与 0.03 dB 的优势超越了 LAPAP 和 FDIWN,对于 SAFMN 与 ShuffleMixer 这种更小的模型,SFFMNet 以少量的参数与 FLOPs 为代价分别取得了 0.59 dB 和 0.58 dB 的性能提升。这主要得益于 SFFMB 模块的非局部建模能力。另外,在与 Transformer 网络的对比中,除保持与 ESTR 近似性能的同时,在 $\times 2$ 、 $\times 3$ 与 $\times 4$ 的放大因子上,所分别产生的 FLOPs 仅为 ESTR 的 6.18%、3.71% 与 6.04%,这证明了本文所提出的空频特征聚合层在超分辨率性能上的有效性,并且其很好地避免了自注意力机制带来的计算压力。

2) 定性比较

除了定量评估所提出的方法,本节将进一步展示模型的可视化结果。图 4 显示了视觉比较的结果,本文提出的方法在恢复平行直线和网格图案方面比列出的方法更准确,模糊和伪影更少。例如,在 Urban100 的 img011 图像中,

SFFMNet 可以恢复更多的高质量平行直线。在 img097 中,SFFMNet 几乎复原了远处建筑中所有的窗格且结构正确。在 Set14 的 barbara 图像中,其余方法均没有正确的恢复头巾的纹理,而 SFFMNet 做到了完全的修复。这证实了 SFFMNet 在主观视觉质量方面表现更好,并证明了本文通过利用空频特征进行自适应特征调制的方法的有效性。

2.3 实验结果与分析

局部归因图(local attribution map, LAM)^[26]表明在恢复过程中红色像素与矩形位置补丁之间的显著相关性。此处,分别选取了高效超分模型中具有代表性的 3 个模型进行 LAM 比较,分别是轻量级模型 SAFMN、中等模型 ShuffleMixer 和 IMDN。图 5 展示了 LAM 比较的结果并标注了相应的扩散指数(diffusion index, DI)值,更高的 DI 值表示模型能涉及更广的像素范围。数据结果表明,本文的网络可以探索更多的非局部信息以实现准确的图像超分辨率率。

此外,本文还从特征图与其功率谱密度(power spectral density, PSD)的角度对模型进行了解释。如图 6

所示,从空域特征图的角度可以观察到 SFFAL 的输出更加关注图像的整体结构与周期性纹理。而 LDEL 的输出则是更关注图像的局部细节,雕像和墙壁上的纹理共同印证了这一点。

对于 PSD 图,将频谱进行了周期性的位移以使得低频成分移至中心位置。可以观察到,SFFAL 激活了更多的低频成分,而 LDEL 则在其基础上增强了更多的高频表示。

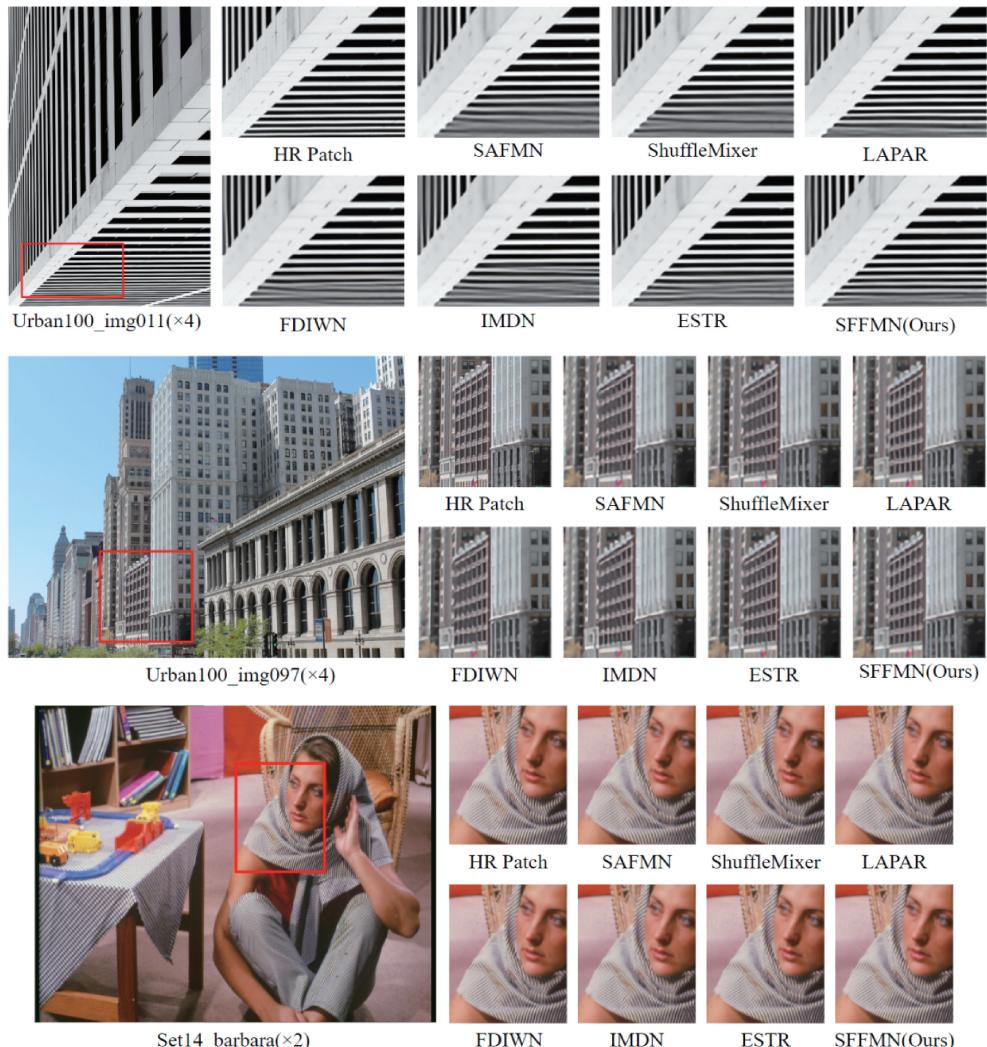


图 4 Urban100 与 Set14 数据集图像在细节区域内的视觉比较结果

Fig. 4 Visual comparison results in detail regions of images from the Urban100 and Set14 datasets

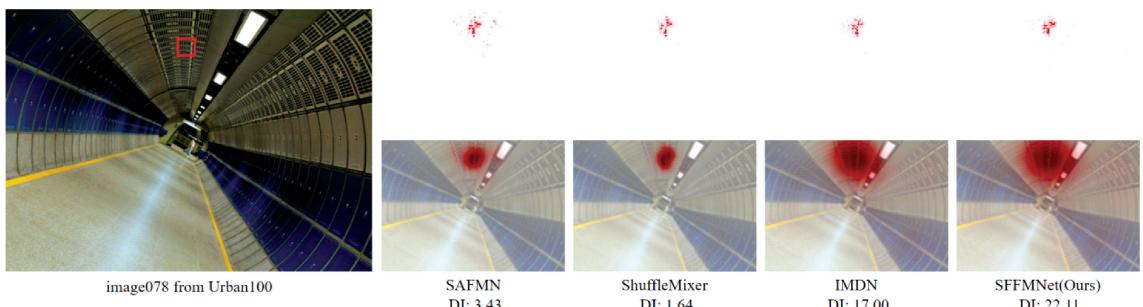


图 5 LAM 图分析结果与扩散指数

Fig. 5 Analysis results of LAM maps and diffusion indices

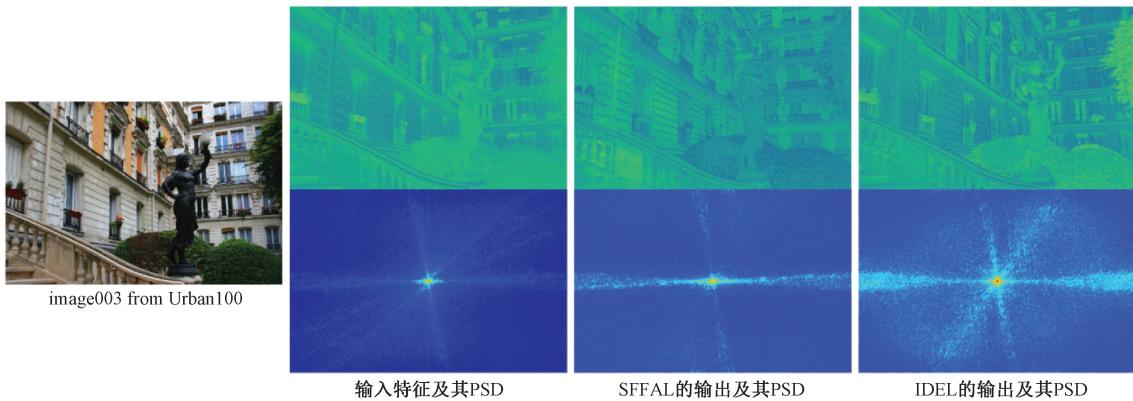


图 6 特征图与其功率谱密度分析

Fig. 6 Analysis of feature maps and corresponding power spectral density

2.4 消融实验

本节对所提出的网络进行了广泛的消融研究,以分析和评估所网络中每个模块的效果。此处,本工作基于 $\times 4$ 的SFFMNet模型进行所有消融实验,并使用DIV2K数据集进行训练,以确保公平的比较。消融结果如表2所示,分别在Set5、Urban100和Manga109数据集上进行性能评估。

LDEL模块通过跨通道交互机制增强局部细节捕捉能力,其移除导致Urban100数据集PSNR值下降0.31 dB;

SFFAL模块通过空频域特征融合以捕捉非局部信息,禁用后使Set5和Urban100的PSNR分别骤降1.76 dB和0.53 dB;UDFA模块的频域处理单元(含SPIE与SDC组件)通过建立全局频域依赖性优化特征表达,当替换其核心SDC组件为FFT-ReLU流^[16]并且不分离实部与虚部时,模型在Set5和Urban100的性能显著退化。实验数据证实了各模块与本文提出的频域特征处理方法在特征建模中的必要性。

表2 消融实验的方法及其在Set5和Urban100上 $\times 4$ 的结果Table 2 Ablation-study configurations and their $\times 4$ results on the Set5 and Urban100

消融实验	模块变化				# Params	# FLOPs	Set5	Urban 100
	DSA	SPIE	SDC	LDEL	/K	/G	PSNR/SSIM	PSNR/SSIM
w/o LDEL	✓	✓	✓		381	13.8	31.95/0.892 0	26.01/0.775 2
w/o SFFAL				✓	93	5.2	31.47/0.883 7	25.79/0.766 9
w/o DFA	✓			✓	164	9.0	32.11/0.894 4	26.05/0.781 0
SPIE+FFT-ReLU	✓	✓	FFT-ReLU	✓	281	12.9	31.81/0.884 2	25.64/0.765 0
w/o SPIE	✓		✓	✓	445	17.7	32.27/0.896 1	26.28/0.783 8
Baseline	✓	✓	✓	✓	456	18.0	32.33/0.896 5	26.32/0.784 5

3 结论

本文提出了一种用于轻量级单图超分辨率的空频特征调制网络SFFMNet,旨在高效地实现对单个图像的超分辨率。

该网络主要由多个SFFMB模块级联构成,每一个SFFMB模块均采用了全局到局部的特征提取策略,包含用于进行非局部建模的SFFAL与局部信息增强的LDEL模块。在SFFAL中,利用空域和频域来捕捉图像的非局部信息并进行融合,使网络能以较低计算复杂度对图像特征建立长程依赖关系,而在LDEL模块中,则利用部分卷积和门控模块,在减少参数量的同时,既能执行通道混合又能有效捕获局部上下文信息。大量实验表明,基于CNN的SFFMNet在重建性能和计算效率之间实现了良好的

平衡。

本文的研究成果是为单图超分辨率的实现提供了一种有效的解决方案,有望在未来的超分辨率应用中发挥一定作用。然而,同其他神经网络架构设计中的关键概念一样,频域特征在神经网络中的作用仍未被充分探索。因此,对于频域特征的可解释性研究以及开发更高效的频域特征提取方法,将成为未来计算机视觉领域的重要方向之一。

参考文献

- [1] DONG C, LOY C C, HE K, et al. Image super-resolution using deep convolutional networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(2): 295-307.
- [2] 丁子轩, 俞雷, 张娟, 等. 基于深度残差自适应注意力网

- 络的图像超分辨率重建[J]. 计算机工程, 2023, 49(5): 231-238.
- DING Z X, YU L, ZHANG J, et al. Image super-resolution reconstruction based on depth residual adaptive attention network [J]. Computer Engineering, 2023, 49(5): 231-238.
- [3] NIU B, WEN W, REN W, et al. Single image super-resolution via a holistic attention network [C]. Springer International Publishing, 2020: 191-207.
- [4] MEI Y, FAN Y, ZHOU Y, et al. Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 5690-5699.
- [5] LIANG J, CAO J, SUN G, et al. Swinir: Image restoration using swin transformer[C]. IEEE/CVF International Conference on Computer Vision, 2021: 1833-1844.
- [6] 杨同, 魏为民, 符程程, 等. 基于 CBAM-SRResNet 的水下图像超分辨率重建研究[J]. 电子测量技术, 2025, 48(1): 20-28.
YANG T, WEI W M, FU CH CH, et al. Research on underwater image super-resolution reconstruction based on CBAM-SRResNet [J]. Electronic Measurement Technology, 2025, 48(1): 20-28.
- [7] SHI W, CABALLERO J, HUSZÁR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1874-1883.
- [8] HUI Z, GAO X, YANG Y, et al. Lightweight image super-resolution with information multi-distillation network[C]. 27th ACM International Conference on Multimedia, 2019: 2024-2032.
- [9] 周登文, 李文斌, 李金新, 等. 一种轻量级的多尺度通道注意图像超分辨率重建网络[J]. 电子学报, 2022, 50(10): 2336-2346.
ZHOU D W, LI W B, LI J X, et al. Image super-resolution reconstruction based on lightweight multi-scale channel attention network[J]. Acta Electronica Sinica, 2022, 50(10): 2336-2346.
- [10] 杨胜荣, 车文刚, 高盛祥. 多尺度信息蒸馏的轻量级图像超分辨率算法[J]. 厦门大学学报(自然科学版), 2023, 62(4): 654-664.
YANG SH R, CHE W G, GAO SH X. Super-resolution algorithm for lightweight image based on multi-scale information distillation [J]. Journal of Xiamen University (Natural Science), 2023, 62(4): 654-664.
- [11] SUN L, DONG J, TANG J, et al. Spatially-adaptive feature modulation for efficient image super-resolution[C]. IEEE/CVF International Conference on Computer Vision, 2023: 13190-13199.
- [12] 李轩, 刘小祎. 通道可分离残差网络的图像超分辨率重建[J]. 电子测量技术, 2023, 46(6): 84-90.
LI X, LIU X Y. Image super-resolution reconstruction based on channel-separable residual network [J]. Electronic Measurement Technology, 2023, 46(6): 84-90.
- [13] DU Z, LIU D, LIU J, et al. Fast and memory-efficient network towards efficient image super-resolution[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 853-862.
- [14] CUI Y, REN W, KNOLL A. Omni-kernel modulation for universal image restoration[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(12): 12496-12509.
- [15] 陈佳艺, 黄晓宇, 吴胜昔, 等. 基于重参数化大核卷积的高分辨率姿态估计[J]. 华东理工大学学报(自然科学版), 2025, 51(3): 341-352.
CHEN J Y, HUANG X Y, WU SH X, et al. High-resolution pose estimation based on reparameterized large kernel convolution [J]. Journal of East China University of Science and Technology, 2025, 51(3): 341-352.
- [16] LI Y, HOU Q, ZHENG Z, et al. Large selective kernel network for remote sensing object detection[C]. IEEE/CVF International Conference on Computer Vision, 2023: 16794-16805.
- [17] ZHANG D, HUANG F, LIU S, et al. Swinfir: Revisiting the swinir with fast fourier convolution and improved training for image super resolution [J]. ArXiv preprint arXiv:2208.11247, 2022.
- [18] MAO X, LIU Y, LIU F, et al. Intriguing findings of frequency selection for image deblurring[C]. AAAI Conference on Artificial Intelligence, 2023, 37(2): 1905-1913.
- [19] AGUSTSSON E, TIMOFTE R. Ntire 2017 challenge on single image super-resolution: Dataset and study [C]. IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017: 1110-1121.
- [20] SUN L, PAN J, TANG J. Shufflemixer: An efficient convnet for image super-resolution[J]. Advances in Neural Information Processing Systems, 2022, 35: 17314-17326.
- [21] LI W, ZHOU K, QI L, et al. Lapar: Linearly-

- assembled pixel-adaptive regression network for single image super-resolution and beyond[J]. Advances in Neural Information Processing Systems, 2020, 33: 20343-20355.
- [22] GAO G, LI W, LI J, et al. Feature distillation interaction weighting network for lightweight image super-resolution[C]. AAAI Conference on Artificial Intelligence, 2022, 36(1): 661-669.
- [23] HUI Z, GAO X, YANG Y, et al. Lightweight image super-resolution with information multi-distillation network[C]. 27th ACM International Conference on Multimedia, 2019: 2024-2032.
- [24] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution[C]. IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017: 136-144.
- [25] LU Z, LI J, LIU H, et al. Transformer for single image super-resolution[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 457-466.
- [26] GU J, DONG C. Interpreting super-resolution networks with local attribution maps[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.

作者简介

顾羽舟,硕士研究生,主要研究方向为深度学习、神经网络硬件加速、集成电路设计。

E-mail:xaviergu.cn@gmail.com

李娇(通信作者),博士,教授,主要研究方向为深度学习与集成电路设计。

E-mail:ljjiaoshu@shu.edu.cn

郭爱英,博士,教授,主要研究方向为显示驱动与模拟电路设计。

E-mail:gayshh@shu.edu.cn

吴昊辰,硕士研究生,主要研究方向为深度学习、神经网络硬件加速、集成电路设计。

E-mail:wuhaochen@shu.edu.cn

陆俊宇,硕士研究生,主要研究方向为三维片上网络。

E-mail:284089474@qq.com