# 基于多尺度特征自适应调制的单图像超分辨率网络

沈伟露,刘 杰*,唐 杰,武港山
(南京大学计算机学院,江苏南京 210023)

**摘 要:** 基于Transformer的图像恢复方法在单图像超分辨率(Single Image Super-Resolution,SISR)任务中展现了卓越的性能,这得益于其自注意力(Self-Attention,SA)机制能够有效捕捉非局部信息,从而实现更高质量的高分辨率(High Resolution,HR)图像重建. 然而,SA机制中的矩阵乘法操作需要消耗大量计算资源,这使得基于Transformer的模型通常难以适配计算能力和内存受限的低功耗设备. 此外,SA机制的低通特性限制了其捕获高频局部细节的能力,从而导致平滑的重建结果. 为了解决以上问题,本文提出了一种基于多尺度特征自适应调制的单图像超分辨率网络(Multi-scale Feature Adaptive Modulation Network,MFAMNet),其核心是多尺度特征自适应调制(Multi-scale Feature Adaptive Modulation,MFAM)模块,该模块通过下采样操作获取不同尺度的低频内容,计算输入特征的全局方差来调制处理后的低频特征,然后使用调制后的特征自适应地聚合输入特征,从而实现对非局部信息的高效建模. 在聚合输入特征之后,本文引入通道注意力机制,从通道维度对融合特征进行细化处理,以增强所有通道间公共信息的提取能力,同时实现跨通道权重的动态重分配. 此外,由于MFAM从远程角度处理输入特征,因此需要补充局部上下文信息. 为此,本文还设计了空间增强模块(Spatial Enhancement Module,SEM),作为复杂自注意机制的有效替代方案,以显著提高空间局部聚合能力,在空间和通道维度上进一步细化从MFAM输出的特征. 大量实验表明:所提出的MFAMNet,在公共基准数据集上实现了重建性能和计算效率之间的更佳权衡. 特别是在4倍超分辨率(Super-Resolution,SR)下,与最新的自调制特征聚合网络(Self-Modulation Feature Aggregation Network,SMFANet)相比,MFAMNet在五个公共测试集上的平均性能提高了0.15 dB,而模型复杂性,如每秒浮点运算次数(FLoating-point Operations Per second,FLOPs),与前者几乎相同.

**关键词:** 单图像超分辨率(SISR);轻量化;自注意力(SA);特征调制;多尺度特征
**基金项目:** 国家自然科学基金(No.62402211);江苏省自然科学基金(No.BK20241248)
**中图分类号:** TP391.41 **文献标识码:** A **文章编号:** 0372-2112(2025)07-2324-18
**电子学报URL:**http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20250032

# Multi-Scale Feature Adaptive Modulation for Single Image Super-Resolution

SHEN Wei-lu, LIU Jie*, TANG Jie, WU Gang-shan
(*School of Compute Science*, *Nanjing University*, *Nanjing*, *Jiangsu* 210023, *China*)

**Abstract:** Transformer-based image restoration methods have demonstrated remarkable performance in single image super-resolution tasks, owing to their self-attention (SA) mechanism, which effectively captures non-local information, thereby achieving higher-quality high-resolution image reconstruction. However, the matrix multiplication operations in the self-attention mechanism consume substantial computational resources, making Transformer-based models generally challenging to deploy on low-power devices with limited computational capabilities and memory. Additionally, the low-pass characteristics of the SA mechanism restrict its ability to capture high-frequency local details, leading to overly smooth reconstruction results. To address these issues, we propose a multi-scale feature adaptive modulation network (MFAMNet) for single image super-resolution, whose core is the multi-scale feature adaptive modulation (MFAM) module. This module obtains low-frequency content at different scales through downsampling operations, computes the global variance of the input features to modulate the processed low-frequency features, and then adaptively aggregates the input features using the modulated features, thereby achieving efficient modeling of non-local information. After aggregating the input features, we intro-

收稿日期:2025-01-08;录用日期:2025-07-15;责任编辑:覃怀银
*通讯作者:刘杰

duce a channel attention mechanism to refine the fused features from the channel dimension, enhancing the extraction of shared information across all channels while dynamically reallocating cross-channel weights. Furthermore, since MFAM processes input features from a long-range perspective, it is necessary to supplement local contextual information. To this end, we also design a spatial enhancement module (SEM) as an effective alternative to complex self-attention mechanisms, significantly improving spatial local aggregation capabilities and further refining the features output from MFAM in both spatial and channel dimensions. Extensive experiments demonstrate that the proposed MFAMNet achieves a better trade-off between reconstruction performance and computational efficiency on public benchmark datasets. Notably, in 4× super-resolution, self-modulation feature aggregation network (MFAMNet) improves the average performance by 0.15 dB compared to the state-of-the-art self-modulation feature aggregation network (SMFANet) on five public test sets, while maintaining nearly the same model complexity, e.g., floating-point operations per second (FLOPs).

Key words: single image super-resolution (SISR); lightweight; self-attention (SA); feature modulation; multi-scale feature

Foundation Item(s): National Natural Science Foundation of China (No.62402211); Natural Science Foundation of Jiangsu Province (No.BK20241248)

## 1 引言

单图像超分辨率(Single Image Super-Resolution,SISR)是计算机视觉领域的一项基本任务,旨在从给定退化的低分辨率(Low Resolution,LR)图像中恢复丢失的细节,重建高分辨率(High Resolution,HR)图像. 随着流媒体平台的迅速崛起和超高清设备的快速发展,SISR因其能够以较低的媒体传输成本提供愉悦的观看体验而受到学术界和工业界的广泛关注.

如图1所示,在深度学习出现之前,研究人员主要通过传统的基于插值、稀疏表示、正则化和学习的方法来解决SISR的问题. 传统的插值方法包括最近邻插值、双线性和双三次插值(Bicubic)[1]. 最近邻插值是指为需插值的每个位置选择最近像素的值,而不考虑任何其他像素. 因此,这种方法非常快,但通常会产生低质量的块状结果. 双线性插值首先对图像的一个轴进行线性插值,然后对另一轴进行线性插值,由于它产生感受野大小为2×2的二次插值,因此它在保持相对较快速度的同时,表现出比最近邻插值更佳的性能. 类似地,Bicubic[1]在两个轴上分别执行三次插值. 与双线性插值相比,Bicubic考虑了4×4个像素,结果更平滑,伪影更少,但速度要低得多. 文献[2,3]提出了一种与稀疏表示相关的超分辨率(Super-Resolution,SR)技术,通过考虑LR和HR特征,共享相似的重构稀疏分量作为学习的HR字典原子的稀疏线性组合,以检索HR属性. 他们使用的优化函数需要L1范数的规律性,这使得SR技术因每个LR特征输入的高强度计算而恶化.

基于正则化的方法假设检测到的LR图像是由多个退化组件产生的,即模糊、变形、下采样,并专注于重建的限制. 因此,选定的HR图像的下采样版本必须与相应的LR图像相同. 不同的先验知识,如边缘导向先验[4,5]和相似性冗余[6],通常用于这种形式的SR方法,以获得明确解. 然而,基于正则化的方法通常会产生包含更多锐度和不需要的边缘的HR图像,并在显著边缘处产生振铃伪影. 相比之下,基于学习的SR方法,又称基于实例的SR方法,因其计算速度快、性能优异而备受关注. 这些方法通常利用机器学习算法,从大量训练示例中分析LR和相应的HR之间的统计关系. Markov随机场[7]方法最早由Freeman等人采用,利用丰富的真实世界图像合成视觉上令人愉悦的图像纹理. 文献[8]提出的邻域嵌入方法,利用LR和HR之间相似的局部几何形状来恢复HR图像块(patch). 随机森林[9]也被用于提高重建性能. 许多工作将基于重建的方法的优点与基于学习的方法相结合,进一步减少外部训练样例[10~12]引入的伪影.

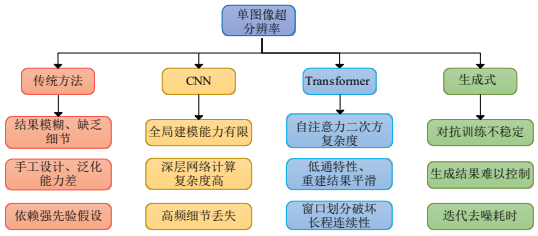



图1　SISR方法存在的问题

SR任务是基于将LR图像映射到HR图像的观测模型的合理假设或先验知识,通过融合LR图像来恢复原始HR图像的逆问题. SR的基本重建约束是:恢复后的图像在应用相同的生成模型后,应该能够再现观察到的LR图像. 然而,由于LR图像数量不足、配准条件不良以及未知的模糊算子,SR图像重建通常是一个严重不适定的问题,且从重建约束中得到的解不是唯一的. 因此,传统的方法很难妥善解决这个问题.

在过去的十年中,深度学习彻底改变了SISR领域. 人们开发了各种卷积神经网络(Convolutional Neural Network,CNN)[1~3,13~27]来解决这个问题. 作为CNN的基本运算,卷积算子具有平移不变性且感受野有限,限制

了其对非局部信息建模的能力. 许多强大的SR网络已经实现了图像的高质量恢复. 然而,为了追求更佳的性能,大多数网络使用具有巨大计算复杂度的更大模型,不断地加深网络的层数. 例如,残差通道注意力网络(Residual Channel Attention Networks,RCAN)[22]是一种代表性的基于CNN的图像SR网络,具有15.59 M参数,深度超过400层. 尽管这些复杂的SR网络提高了图像重建的质量,但由于模型容量不断升级和密集的计算需求,它们无法在计算资源有限的边缘设备上部署. 因此开发一种高效且有效的SR方法来估计HR图像,对于在这些平台或产品上更佳的视觉显示是非常有意义的. 为了减轻繁重的计算负担,人们使用了各种方法,包括高效的模块设计[28~31]、知识蒸馏[32]、神经架构搜索[33,34]以及结构重新参数化[35]等.

近年来,生成式人工智能在图像重建方面取得了显著的成果. 生成式对抗网络(Generative Adversarial Network,GAN)通过生成器和判别器的对抗训练,能够生成更具视觉真实感的细节,重建出逼真的HR图像,显著提升了图像重建的质量. 超分辨率生成对抗网络(Super-Resolution Generative Adversarial Networks,SRGAN)[36]首次将GAN引入图像SR,通过感知损失和对抗损失的结合,突破传统基于均方误差方法的局限性,生成了具有更佳视觉效果的HR图像. 而增强型超分辨率生成对抗网络(Enhanced Super-Resolution Generative Adversarial Networks,ESRGAN)[37]在SRGAN的基础上改进模型,在网络架构、损失函数等方面进行优化,显著提升了图像SR的效果,使生成的图像质量更高、细节更丰富、视觉效果更自然. 然而,GAN容易出现模式崩溃、计算量庞大、有时无法收敛,并且存在训练稳定性问题.

扩散模型的出现标志着包括SR在内的图像生成任务的重大转变,挑战了生成对抗网络长期以来的主导地位. 扩散模型通过逐步去除噪声来生成高质量的图像,为图像SR提供了新的解决方案,在SR任务中展现出巨大潜力,其生成结果与人类评估者的定性判断高度吻合. SR3(Super-Resolution via iterative refinement)[38]是首个将扩散模型应用于SR的模型,通过预测噪声逐步恢复HR图像,在自然图像和人脸SR中表现出高逼真度. DAll-E[39]和视觉变换器(Stable Diffusion)[40]的出现更是让扩散模型在多个方面超越生成对抗网络. 但是扩散模型通常需要数百甚至数千次的迭代去噪步骤来生成高质量的HR图像. 尽管有研究尝试通过单步推理来加速扩散模型,但单步推理的效果仍然有限. 此外,为了生成高质量的图像,扩散模型通常需要较大的模型规模和大量的参数,并且依赖大量的训练数据来学习图像的复杂分布.

最近的视觉Transformer(Vision Transformer,ViT)[41]在高级和低级视觉任务上都取得了令人印象深刻的成功[42~52],远远超过了CNN. ViT中的自注意力(Self-Attention,SA)机制可以有效地对非局部信息进行建模,这对于高质量重建至关重要. 然而,SA机制需要较高的计算资源和大量的内存消耗,并且不利于高效的SR设计. 为了降低计算成本,人们开发了基于窗口的SA[33,35,39,53]、转置SA[54]和权重复用策略[55]来减轻计算负担. 然而,这些方法仍然需要很长时间来学习图像SR的特征依赖性. 此外,最近的研究[56,57]表明:ViT有利于学习低频分量,从而获得平滑的重建结果.

因此,这促使我们思考:能否开发一种高效的特征调制模块,以类似SA的方式探索非局部信息,同时对局部细节进行建模以实现高效的图像SR. 为此,本文提出了一个多尺度特征自适应调制(Multi-scale Feature Adaptive Modulation,MFAM)模块,通过下采样操作获取不同尺度的低频内容,引入输入特征的全局方差来调制处理后的低频特征,然后使用调制后的特征自适应地聚合输入特征. 在聚合输入特征之后,本文使用通道注意力从通道维度对融合的特征进行处理. 此外,由于MFAM模块从远程角度处理输入特征,因此需要补充局部上下文信息. 为此,本文还设计了空间增强模块(Spatial Enhancement Module,SEM),作为复杂SA机制的有效替代方案,以显著提高空间局部聚合能力,在空间和通道维度上进一步细化从MFAM输出的特征. 本文将提出的MFAM模块和SEM模块制定为端到端可训练网络,称为MFAMNet,以解决SISR. 如图2所示,图中的圆圈大小表示模型的每秒浮点运算次数(FLoating-point Operations Per second,FLOPs),本文提出的MFAMNet在计算效率和重构性能之间实现了更佳的平衡.

综上所述,本文提出的MFAMNet的主要贡献如下:

(1)本文开发了一个高效的MFAM模块提取多尺度特征.

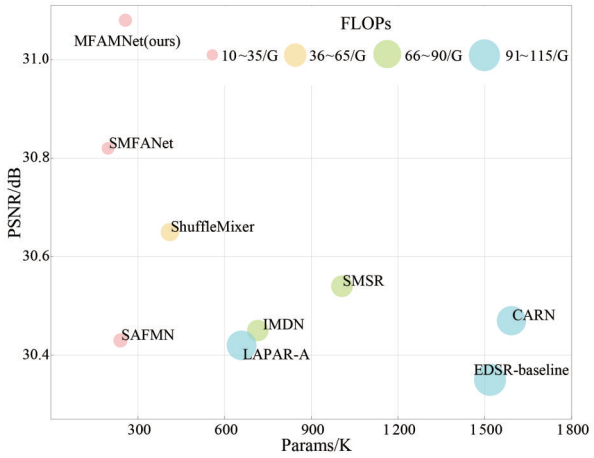
(2)本文提出了一种SEM,利用大核条纹卷积,有效提取并融合局部空间信息以进一步完善MFAM在空间和通道维度上的特征.

(3)本文在基准数据集上定量和定性地评估了所提出的方法,大量的实验结果表明:与近几年代表性的轻量级SISR方法相比,本文提出的MFAMNet能够重建出更高质量的SR图像,同时具有较低的参数量和更低的计算复杂度.

## 2 相关工作

### 2.1 基于CNN的SR

经典的插值算法,例如线性或双三次上采样,首先通过在LR的相邻像素之间插入零来创建HR图像,然




注:与MFAMNet比较的方法包括级联残差网络(CascAding Residual Network,CARN)[13]、EDSR-baseline[18]、信息多蒸馏网络(Information Multi-Distillation Network,IMDN)[16]、线性组装像素自适应回归网络-A(Linearly-Assembled Pixel-Adaptive Regression Network-A,LAPAR-A)[58]、基于稀疏性的多分支超分辨率(Sparse-based Multi-branch Super-Resolution,SMSR)[59]、ShuffleMixer[21]、空间自适应特征调制网络(Spatially-Adaptive Feature Modulation Network,SAFMN)[20]和自调制特征聚合网络(Self-Modulation Feature Aggregation Network,SMFANet)[60],圆圈大小表示模型的FLOPs数量.

图2　MFAMNet模型与其他最先进的轻量级方法在Manga109数据集×4 SR上的模型复杂性和性能比较

后使用低通滤波器来保留输入图像的内容信息[30]. 基于CNN的图像SR技术相较于传统的插值算法,通过端到端训练学习输入图像和目标输出之间的非线性映射,取得了显著的性能提升. 超分辨率卷积神经网络(Super-Resolution Convolutional Neural Network,SRCNN)[14]是首个使用CNN来解决图像SR问题的模型,它直接将LR图像映射到HR图像,性能优于传统手工设计的方法. 随后,快速超分辨率卷积神经网络(Fast Super-Resolution Convolutional Neural Network,FSRCNN)[15]和高效亚像素卷积神经网络(Efficient Sub-Pixel Convolutional Neural network,ESPCN)[19]采用后上采样策略进一步提高了效率,而VDSR(Very Deep Super-Resolution)[61]利用全局残差学习有效解决深度网络的训练难题,使得后续的CNN方法可以构建更深、更大的网络以捕获更多信息,从而更好地恢复图像. 深度递归残差网络(Deep Recursive Residual Network,DRRN)[62]集成了局部残差学习和全局残差连接,以减轻训练难度并增强高频细节. 增强型深度残差网络(Enhanced Deep residual networks,EDSR)[18]将模型大小增加到43 MB,显著提升了重建性能,而RCAN[22]基于通道注意力和密集连接构建了超过400层的模型. 然而,随着模型复杂度的增加,这些大型模型的高计算成本限制了它们在资源受限设备上的实际应用.

## 2.2　基于Transformer的SR

视觉Transformer(ViT)[41]利用SA机制探索全局信息进行图像重建,并取得了巨大成功. 图像处理变换器(Image Processing Transformer,IPT)[42]首先引入了标准视觉Transformer来解决图像SR问题. 然而,ViT的SA机制消耗大量资源,因此提出了各种注意力变体,通过将SA限制在局部区域,并引入更高程度的局部性偏差,可以显著减轻计算负担. 用于图像恢复的Swin变换器(Swin transformer for Image Restoration,SwinIR)[26]结合了局部窗口SA和受Swin Transformer设计启发的移位机制,并且优于基于CNN的大型模型. 边缘学习加速网络(Edge Learning Accelerator Network,ELAN)[30]提出了一种分组SA模块并共享权重以降低复杂性. 高效单图像超分辨率变换器(Efficient Single-image super-Resolution Transformer,ESRT)[63]通过分割或缩小尺寸来减少特征维度,提高计算效率. 轻量级图像超分辨率的全聚合网络(Omni aggregation networks for lightweight image Super-Resolution,Omni-SR)[64]对不同轴上的像素交互进行建模,创建通用相关性,而SRFormer[50]通过SA机制的排列,采用大窗口SA来优化计算效率. 尽管这些基于Transformer的方法在图像重建方面取得了进展,但它们通常需要更高的计算资源,即使模型容量较小.

此外,最近的研究工作[56,57]指出:ViTs具有低通滤波器的性质,从而产生平滑的重建结果. 因此,许多基于ViT的方法[65~67]利用卷积算子来增强局部细节,以实现更高质量的图像重建. 尽管这些方法集成了卷积和Transformer结构的优点,并利用局部和非局部信息来实现高质量的图像重建结果,但它们仍然依赖于基于窗口的注意力变体,并且跨窗口的特征交互需要大量时间来执行.

## 2.3　高效的图像SR

为了在图像SR中实现高效的重建性能和模型复杂性之间的平衡,许多基于CNN的方法被提出以减轻计算负担. FSRCNN[15]和ESPCN[19]通过后上采样方式减少了预定义输入的计算负担,而CARN[13]应用组卷积和级联残差网络来提高效率. IMDN[16]采用信息多重蒸馏块对特征映射进行逐步分割、细化和聚合,显著减少了模型参数,其改进工作残差特征蒸馏网络(Residual Feature Distillation Network,RFDN)[29]在AIM 2020高效SR挑战赛中获胜. ShuffleMixer[21]为轻量级SR设计引入了大内核卷积,而蓝图可分离残差网络(Blueprint Separable Residual Network,BSRN)[17]提出了基于蓝图可分离卷积的模型来降低模型复杂性.

除了这些方法,模型量化、结构重新参数化或知识蒸馏等技术也被用来压缩或加速训练有素的深度模

型，神经架构搜索（Neural Architecture Search，NAS）也被用于搜索图像SR的良好约束架构．值得注意的是，深度神经网络的效率可以通过参数数量、FLOPs、激活、内存消耗和推理运行时间等不同指标来衡量．尽管这些方法在不同方面都取得了效率的提升，但在重建性能与模型效率之间仍有进一步优化的空间．

最近的研究趋势之一是引入非局部特征调制来提取代表性特征以进行重建．大感受野像素注意力网络（Vast-receptive-field pixel attention network，VapSR）[68]在注意力分支中使用大内核卷积来扩大感受野，SAFMN[20]通过引入不同的下采样率来获得多尺度空间特征，然后聚合这些特征以生成调制图，而多级分散残差网络（Multi-level Dispersion Residual Network，MDRN）[69]采用多个跨步卷积和轮询操作的并行实现来捕获更精细的全局结构信息．这些方法以较低的计算成本对非局部特征交互进行建模并实现良好的性能，但仅使用聚合的非局部特征执行重建可能会忽略局部特征，导致SR图像的局部细节（例如角和边缘）中的伪影．为了避免这个问题，提出了一种高效的MFAM模块，该模块可以协同建模局部和非局部特征，以实现更准确的重建．该模块的引入旨在提高图像SR技术在实际应用中的可行性，同时保持高质量的图像重建效果，进一步探索在保持模型效率的同时提高重建性能的可能性．

## 3 基于多尺度特征自适应调制的单图像超分辨率网络

图3展示了本文提出的基于多尺度特征自适应调制的单图像超分辨率网络（Multi-scale Feature Adaptive Modulation Network，MFAMNet）的网络架构．其整体架构如图3（$a$）所示，它继承自BSRN[17]的结构，由浅层特征提取、深层特征提取、多层特征融合和重建四个阶段组成．在预处理阶段，输入图像首先被复制$k$次．然后将这些图像连接在一起，表达式为

$$I_{\mathrm{LR}}^{k}=\mathrm{Concat}\left(I_{\mathrm{LR}}\right) \tag{1}$$

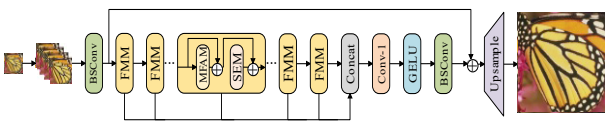
其中，$\mathrm{Concat}\left(I_{\mathrm{LR}}\right)$ 表示沿通道维度的连接操作；$I_{\mathrm{LR}}$ 表示LR的图像；$k$ 表示需连接的 $I_{\mathrm{LR}}$ 的数量．在接下来的浅层特征提取部分，将输入图像映射到更高维的特征空间中：

$$F_0=\mathrm{BSconv}\left(I_{\mathrm{LR}}^{k}\right) \tag{2}$$

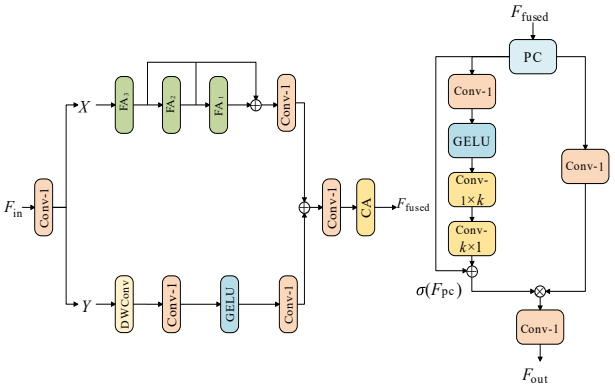
其中，$F_0$ 表示提取的浅层特征；BSconv 表示浅层特征提取模块，具体来说，使用蓝图可分离卷积（Blueprint Separable Convolution，BSConv）[17]来实现浅层特征提取．然后使用多个堆叠的特征混合模块（Feature Mixing Modules，FMM）模块进行深层特征提取，逐渐细化提取的特征：

$$F_l=\mathrm{FMM}_l\left(F_{l-1}\right),l=1,2,\cdots,n \tag{3}$$

其中，$\mathrm{FMM}_l$ 表示第 $l$ 个FMM模块，其中包含MFAM和SEM两个子模块；$F_{l-1}$ 和 $F_l$ 分别表示第 $l$ 个FMM模块的输入和输出．




（$a$）MFAMNet的整体结构




（$b$）MFAM模块　　（$c$）SEM的结构

图3　MFAMNet的网络架构

在FMM中，MFAM模块和SEM模块之间的协同作用对于实现高质量的图像SR重建至关重要．这两个模块通过互补的方式，分别从全局和局部角度对特征进行增强和细化，从而显著提升了模型的重建性能．MFAM模块通过MFAM机制，有效地捕捉了图像中的非局部信息．它利用全局方差计算和特征调制，将不同尺度的低频特征与全局统计信息相结合，从而增强了模型对长距离依赖关系的建模能力．这种全局信息的捕捉有助于模型理解图像的整体结构和上下文，为高质量的SR重建提供了重要的背景信息．而SEM模块则专注于提取和融合局部空间信息．通过部分卷积和条纹卷积的设计，SEM模块能够有效地捕捉图像中的局部细节和纹理．它通过使用条纹卷积将大内核卷积分解为两个连续的卷积，不仅减少了参数量和计算复杂度，还扩大了感受野，从而更好地提取局部特征．这种局部细节的增强有助于模型在重建过程中保留更多的细节信息，避免模糊和失真．在训练过程中，MFAM模块和SEM模块通过端到端的训练方式协同优化．MFAM模块负责提供全局信息的指导，而SEM模块则负责细化局部特征．这种协同优化方式使得两个模块能够相互补充，共同提升模型的重建性能．通过这种方式，模型不仅能够更好地理解图像的整体结构，还能够更准确地重建图像中的细节和纹理．

为了充分利用所有层的特征，将不同层生成的特

征通过1×1卷积和GELU激活进行融合和映射. 然后,使用BSConv来细化特征. 多层特征融合公式为

$$F_{\text{fused}}=\text{Fusion}\left(\text{Concat}\left(F_1,F_2,\cdots,F_n\right)\right) \tag{4}$$

其中, $F_{\text{fused}}$ 表示融合之后的特征; Fusion 表示特征融合模块. 最后利用残差连接,对得到的融合特征和浅层特征求和,使用PixelShuffle层重建SR图像:

$$I_{\text{SR}}=\text{Rec}\left(F_{\text{fused}}+F_0\right) \tag{5}$$

其中, $I_{\text{SR}}$ 表示SR的图像;Rec 表示重建模块. 继之前的工作[20]之后,本文使用平均绝对误差(Mean Absolute Error, MAE)损失和基于快速傅里叶变换(Fast Fourier Transform, FFT)的频率损失函数的组合进行优化,其定义为

$$L=\left\| I_{\text{SR}}-I_{\text{HR}}\right\|_1+\lambda\left\| F\left(I_{\text{SR}}\right)-F\left(I_{\text{HR}}\right)\right\|_1 \tag{6}$$

其中, $I_{\text{HR}}$ 表示高质量的真实图像;1 表示L1范数;$\lambda$ 表示权重参数;$F$ 表示快速傅里叶变换.

### 3.1 MFAM模块

探索非局部信息对于图像SR重建至关重要,因为它通过整合图像大范围的信息,允许模型捕捉到长距离依赖关系,从而更好地重建高质量的细节和纹理. 传统的CNN通常依赖于局部感受野来提取特征,虽然有效,但在处理全局上下文信息时存在一定的局限性. 非局部信息能够覆盖更广泛的区域,帮助模型理解和重建图像中的全局结构,尤其是在处理复杂场景或LR图像时,能够提供更丰富的上下文支持.

最近基于ViT的SR方法[42,43,45,48,49]利用各种SA机制来探索非局部信息,并实现令人印象深刻的重建性能. 然而,这些SA变体的计算成本很高,并且对局部细节进行建模的能力有限,因为它们的低通滤波器性质使它们优先捕获低频信息. 与SA机制相比,本文提出了一种轻量级替代方案,从多尺度特征表示中学习长程依赖关系,以便可以得到更有用的特征,从而更好地探索HR图像重建. 如图3(*b*)所示,本文开发了一个轻量级MFAM模块,可以协作建模局部和非局部特征以实现精确重建.

MFAM模块的核心设计思路是通过多尺度特征提取与融合、全局方差计算、特征调制和局部上下文补充,实现了对非局部信息和局部细节的有效建模. 这种设计不仅提高了模型对全局信息的捕捉能力,还通过调制机制和局部特征的补充,确保了重建图像的高质量和计算效率. 其中多尺度特征提取与融合主要通过不同尺度的下采样操作,获取不同分辨率的低频特征. 这些低频特征能够覆盖更广泛的区域,从而捕捉到全局信息. 在特征提取过程中,通过计算输入特征图的全局方差,作为对全局信息的统计描述. 全局方差能够反映特征图中像素值的离散程度,提供关于图像整体特性的线索,增强模型对全局信息的感知能力,能够有效表征非局部结构的显著性. 将全局方差与低频特征结合,通过1×1卷积进行融合,增强模型对非局部信息的建模能力. 这种调制机制帮助模型更好地捕捉全局依赖关系. 通过全局方差调制后的低频特征作为调制器,与原始输入特征进行逐元素相乘. 通过逐元素相乘,确保了非局部信息能够有效地影响每个像素的局部特征表示,使得模型在处理局部细节时能够更好地利用全局上下文信息,增强输入特征图,以更好地捕捉全局信息.

具体来说,给定输入特征 $F_{\text{in}}\in\mathbb{R}^{H\times W\times C}$,其中 $H\times W$ 表示空间大小,$C$ 是通道数,首先对归一化的 $F_{\text{in}}$ 应用1×1的卷积来扩展通道数,然后分割通道.

$$\{X,Y\}=S\left(\text{Conv}_{1\times1}\left(\left\| F_{\text{in}}\right\|_2\right)\right) \tag{7}$$

其中, $X\in\mathbb{R}^{H\times W\times C}$, $Y\in\mathbb{R}^{H\times W\times C}$; $S$ 表示通道分割操作; $\text{Conv}_{1\times1}$ 表示 $1\times1$ 卷积;2表示L2归一化. 其次,对于输入 $X$,本文应用三个顺序的特征聚合模块(Feature Aggregation, FA),这三个FA模块唯一的区别就是分别对应三个不同的下采样尺度,以此来提取不同尺度下的特征. 同时,前一个FA模块提取的特征能够直接作为下一个FA模块的输入,这样可以更好地利用前面提取到的信息,逐步对特征进行细化和完善,有利于学习到更丰富、更具代表性的特征,最后通过一个 $1\times1$ 的卷积融合三个不同尺度的特征.

$$X_{\text{down8}}=\text{FA}_3(X) \tag{8}$$

$$X_{\text{down4}}=\text{FA}_2\left(X_{\text{down8}}\right) \tag{9}$$

$$X_{\text{down2}}=\text{FA}_1\left(X_{\text{down4}}\right) \tag{10}$$

$$X_{\text{fused}}=\text{Conv}_{1\times1}\left(X_{\text{down8}}+X_{\text{down4}}+X_{\text{down2}}\right) \tag{11}$$

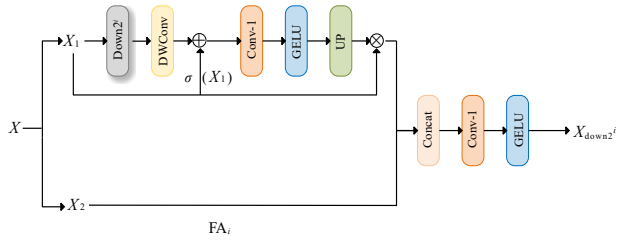
具体而言,如图4所示,FA模块首先通过通道分割操作将输入$X$分成两份,其中$X_1\in\mathbb{R}^{H\times W\times C/4}$, $X_2\in\mathbb{R}^{H\times W\times 3C/4}$. 对于分量$X_1$,通过降采样操作获得低频分量,并将其输入3×3深度卷积以生成非局部结构信息:

$$\left\{X_1,X_2\right\}=S(X) \tag{12}$$

$$X_s=\text{DWconv}\left(\text{Down}_{2^i}\left(X_1\right)\right) \tag{13}$$

其中,DWconv 表示一个 $3\times3$ 的深度卷积层; $\text{Down}_{2^i}$ 表示缩放因子为 $2^i$ 的自适应最大池化. 通过多尺度下采样操作,可以生成三种不同分辨率的特征图. 由于更小的特征图(比如8倍下采样的特征图)有着更大的感受野,因此在应用相同大小的卷积核时,不同尺度的特征图中的每个像素的感受野都是不同的. 其中8倍下采样特征图的感受野最大. 通过引入分割通道机制,可以有效降低特征通道的数量,从而显著减少模型的参数量和计算复杂度,实现更高效的模型架构设计.

为了嵌入调制非局部表示 $X_1$ 的全局描述,本文引




注：$i$表示每个FA模块不同的下采样尺度.

图4 FA模块

入$X_1$的方差作为空间信息的统计散度，并通过1×1卷积将其与非局部表示$X_s$合并，以增强模型对非局部信息的探索能力，进而获得更精确的空间注意力分布：

$$X_m=\mathrm{Conv}_{1\times1}\left(X_s+\sigma\left(X_1\right)\right) \tag{14}$$

其中，$\sigma$表示求$X_1$的方差，这种方差调制机制有助于更好地探索非局部信息. 最后，本文使用调制特征聚合输入特征$X_1$来提取具有代表性的结构信息并将它与$X_2$合并. 为了确保特征图在通道维度上的可拼接性，本文对经过下采样的特征进行上采样，使其恢复到原始空间分辨率($H\times W$)，从而实现特征图的尺度一致性对齐：

$$X_l=\mathrm{UP}\left(\mathrm{Gelu}\left(X_m\right)\right)\odot X_1 \tag{15}$$

$$X_{\mathrm{down2}^i}=\mathrm{Gelu}\left(\mathrm{Conv}_{1\times1}\left(\mathrm{Concat}(X_l,X_2)\right)\right) \tag{16}$$

其中，UP表示最近邻上采样操作；Gelu表示GELU激活函数；$\odot$表示元素积操作. 值得注意的是，本文并未将合并之后的上采样结果直接作为最终输出，而是将其作为特征调制器，与原始输入特征进行逐元素相乘. 具体而言，该调制器首先经过GELU激活函数处理，随后与输入特征进行乘积运算. 在SR等低级视觉任务中，需严格避免空间信息的损失. 尽管在获取多尺度特征时采用了可能丢失细节信息的下采样操作，但通过将其作为门控机制的权重而非模块的最终输出，有效缓解了细节信息丢失的问题.

而对于另一个输入$Y$，保持其原始分辨率不变，采用直接处理策略. 具体来说，本文使用一个核大小为3×3的扩展深度卷积，对输入特征的局部上下文信息进行编码，然后本文使用两个1×1的卷积和一个隐藏的GELU激活来生成增强的局部特征$Y_d$：

$$Y_d=\mathrm{Conv}\left(\mathrm{Gelu}\left(\mathrm{Conv}\left(\mathrm{DWconv}\left(Y\right)\right)\right)\right) \tag{17}$$

最后，本文利用加法将$X_l$和$Y_d$融合在一起，并将它们输入1×1卷积和通道注意力模块，促进通道之间的信息交互，以形成MFAM模块的输出. 该过程可以表述为

$$F_{\mathrm{fused}}=\mathrm{CA}\left(\mathrm{Conv}\left(X_{\mathrm{fused}}+Y_d\right)\right) \tag{18}$$

其中，CA表示通道注意力；$F_{\mathrm{fused}}$表示MFAM模块的特征融合输出结果.

### 3.2 SEM

MFAM主要致力于提取全局上下文信息，其有效性可通过空间局部信息的补充而得到进一步提升. 传统的卷积操作是提取局部上下文信息的常用方法，通过滑动窗口的方式对局部区域进行特征提取. 然而，标准卷积操作在处理大感受野时需要大量的参数和计算资源，这限制了其在轻量级模型中的应用. 此外，卷积操作通常难以捕捉长距离的空间依赖关系. Transformer中的空间注意力机制通过计算所有位置之间的SA权重来建模全局依赖关系. 尽管这种方法能够有效捕捉长距离依赖，但其计算复杂度较高，且对局部细节的建模能力有限. 此外，空间注意力机制的二次方计算复杂度使得其在处理HR图像时，面临内存和计算效率的挑战. 为了解决传统的卷积操作和Transformer中的空间注意力机制的局限性，本文同时利用这两者的优点，设计了SEM，通过直接计算大核卷积的输出和值表示之间的Hadamard乘积来近似模拟Transformer中的空间注意力机制，克服了Transformer中空间注意力机制二次方计算复杂度的缺点. 同时当使用大内核卷积时，使用条纹卷积将大内核卷积分解为两个连续的卷积，在保持大内核感受野的同时，比直接使用大内核卷积的参数量更少，计算效率更高，克服了传统的卷积操作直接使用大内核卷积需要大量的参数和计算资源的缺陷. 此外，在通过条纹卷积得到注意力图之后，本文还融合了输入特征的方差，与MFAM模块中的思路一样，提供关于图像整体特性的线索，进一步增强模型输入特征图.

本文提出的SEM如图3($c$)所示，具体来说，本文首先使用一个部分卷积来细化MFAM输出的特征：

$$F_{\mathrm{pc}}=\mathrm{PC}\left(F_{\mathrm{fused}}\right) \tag{19}$$

其中，PC表示部分卷积(Partial Convolution)；$F_{\mathrm{fused}}$表示经过部分卷积细化后的特征. 由于不同通道之间的特征图存在高度相似性，本文采用了一种简化的部分卷积机制，旨在同时减少计算冗余和内存访问开销. 该机制仅在部分输入通道上执行常规卷积操作以提取空间特征，而其余通道则保持原始状态不变. 这种设计不仅显著降低了计算复杂度，还有效减少了内存带宽需求，从而在保持特征提取能力的同时提升了整体效率.

随后，本文使用1×1卷积、GELU激活函数和一个条纹卷积(Striped Convolution)生成注意力图$A$，以此来简化键$K$和查询$Q$之间相似性矩阵$\boldsymbol{A}$的计算. 与MFAM中类似，在计算注意力图时也引入了方差：

$$\boldsymbol{A}=\mathrm{Striped}\left(\mathrm{Gelu}\left(\mathrm{Conv}\left(F_{\mathrm{pc}}\right)\right)\right)+\sigma\left(F_{\mathrm{pc}}\right) \tag{20}$$

$$F_{\mathrm{out}}=\mathrm{Conv}\left(A\odot\mathrm{Conv}\left(F_{\mathrm{pc}}\right)\right) \tag{21}$$

其中，Striped表示条纹卷积；$F_{\mathrm{out}}$表示SEM的最终输

出. 通过使用条纹卷积将大内核卷积分解为两个连续的卷积,在保持大内核感受野的同时,比直接使用大内核卷积的参数量更少,计算效率更高.

## 4　实验

### 4.1　数据集和实现细节

为了与最先进的方法进行公平比较,本文使用常用的DIV2K+Flickr2K(DF2K)数据集训练SR模型,在常用的测试数据集上评估了本文的方法,包括Set5、Set14、B100、Urban100和Manga109. 本文将图像变换到YCbCr颜色空间,并计算图像$Y$通道上的峰值信噪比(Peak Signal-to-Noise Ratio,PSNR)和结构相似性指数(Structural Similarity Index Measure,SSIM)来评估恢复图像的质量. 在训练过程中,本文将图片随机裁剪成大小为64×64的图像块,并使用LR图像的随机水平翻转和旋转对训练输入进行数据增强. 模型使用Adam优化器进行优化,$\beta_1$=0.9,$\beta_2$=0.99. 本文将初始学习率设置为$1\times10^{-3}$,最小学习率为$1\times10^{-5}$,并通过余弦退火方案进行更新. MFAMNet由8个FMM、36个通道组成. 所有实验均使用PyTorch框架进行,并且迭代次数设置为1 000 000.

### 4.2　与先进方法的比较

#### 4.2.1　定量比较

为了全面评估本文方法的性能,本文将MFAMNet与基于CNN的轻量级SR方法进行比较,包括FSRCNN[15]、VDSR[61]、CARN[13]、EDSR-baseline[18]、IMDN[16]、LAPAR-A[58]、SMSR[59]、ShuffleMixer[21]、SAFMN[20]、SMFANet[60]、MSGN-S(Multi-Scale Gated Network for Single image super-resolution)[70]、SeemoRe-T(See more details for Real-Time image super-resolution)[71]、EARFA-light(Entropy Attention and Receptive Field Augmentation-light)[72]、SRConvNet(Super-Resolution Convolutional neural Network)[73]、MELTN(Multi-scale Enhanced Large-kernel attention Transformer Network)[74]和分组残差特征网络(Group Residual Feature Network,GRFN)[75]. 表1报告了基准数据集上×2、×3和×4放大因子的定量比较,其中PSNR和SSIM指标表示在该数据集上的平均值,红色和蓝色分别表示最佳的性能和第二佳的性能. 除了PSNR和SSIM指标之外,本文还列出了参数数量(#Params)和FLOPs(#FLOPs). 为了公平比较,本文在将LR图像SR至1 280×720像素的设置下,使用fvcore库(即fvcore.nn.flop_count_str)计算所有评估方法的模型复杂度. 受益于MFAM模块的非局部信息建模能力,与之前基于CNN的方法相比,MFAMNet可以有效地探索更多信息. 表1表明:本文的MFAMNet在所有基准数据集上都取得了更佳的性能,并且具有极低的参数量和FLOPs数.

**表1　不同方法在公共基准数据集上的定量比较**

| 算法 | 尺度 | #Params/K | #FLOPs/G | Set5 | Set14 | B100 | Urban100 | Manga109 |
|---|---|---|---|---|---|---|---|---|
| | | | | PSNR/SSIM | PSNR/SSIM | PSNR/SSIM | PSNR/SSIM | PSNR/SSIM |
| Bicubic | ×2 | — | — | 33.66/0.929 9 | 30.24/0.868 8 | 29.56/0.843 1 | 26.88/0.840 3 | 30.80/0.933 9 |
| FSRCNN | | 12 | 6 | 37.00/0.955 8 | 32.63/0.908 8 | 31.53/0.892 0 | 29.88/0.902 0 | 36.67/0.969 4 |
| VDSR | | 665 | 613 | 37.53/0.958 7 | 33.03/0.912 4 | 31.90/0.896 0 | 30.76/0.914 0 | 37.22/0.972 9 |
| CARN | | 1 592 | 223 | 37.76/0.959 0 | 33.52/0.916 6 | 32.09/0.897 8 | 31.92/0.925 6 | 38.36/0.976 5 |
| EDSR-baseline | | 1 370 | 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 |
| IMDN | | 694 | 161 | 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 |
| LAPAR-A | | 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 |
| SMSR | | 985 | 132 | 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 |
| ShuffleMixer | | 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 |
| SAFMN | | 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 |
| SMFANet | | 186 | 41 | 38.08/0.960 7 | 33.65/0.918 5 | 32.22/0.900 2 | 32.20/0.928 2 | 39.11/0.977 9 |
| SeemoRe-T | | 220 | — | 38.06/0.960 8 | 33.65/0.918 6 | 32.23/0.900 4 | 32.22/0.928 6 | 39.01/0.977 7 |
| EARFA-light | | 199 | — | 38.05/0.960 8 | 33.65/0.918 8 | 32.23/0.900 5 | 32.28/0.929 8 | 39.10/0.978 1 |
| MSGN-S | | 387 | 82 | 38.10/0.960 8 | 33.68/0.918 6 | 32.22/0.900 3 | 32.21/0.928 8 | — |
| SRConvNet | | 387 | 74 | 38.00/0.960 5 | 33.58/0.918 6 | 32.16/0.899 5 | 32.05/0.927 2 | 38.87/0.977 4 |
| GRFN | | 528 | — | 38.10/0.960 9 | 33.77/0.919 3 | 32.28/0.900 9 | 32.52/0.931 7 | 39.14/0.978 1 |
| MELTN | | 639 | 148 | 38.03/0.960 3 | 33.69/0.918 6 | 32.20/0.900 1 | 32.21/0.927 6 | 38.76/0.977 3 |
| MFAMNet(ours) | | 245 | 48 | 38.12/0.961 4 | 33.80/0.919 8 | 32.28/0.900 9 | 32.44/0.930 5 | 39.28/0.978 3 |
| Bicubic | ×3 | — | — | 30.39/0.868 2 | 27.55/0.774 2 | 27.21/0.738 5 | 24.46/0.734 9 | 26.95/0.855 6 |
| FSRCNN | | 12 | 5 | 33.16/0.914 0 | 29.43/0.824 2 | 28.53/0.791 0 | 26.43/0.808 0 | 30.98/0.921 2 |
| VDSR | | 665 | 613 | 33.66/0.921 3 | 29.77/0.831 4 | 28.82/0.797 6 | 27.14/0.827 9 | 32.01/0.931 0 |

续表

| 算法 | 尺度 | #Params/K | #FLOPs/G | Set5 PSNR/SSIM | Set14 PSNR/SSIM | B100 PSNR/SSIM | Urban100 PSNR/SSIM | Manga109 PSNR/SSIM |
|---|---|---|---|---|---|---|---|---|
| CARN | ×4 | 1 592 | 119 | 34.29/0.925 5 | 30.29/0.840 7 | 29.06/0.803 4 | 28.06/0.849 3 | 33.50/0.944 0 |
| EDSR-baseline | | 1 555 | 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 |
| IMDN | | 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 |
| LAPAR-A | | 594 | 114 | 34.36/0.926 7 | 30.34/0.841 2 | 29.11/0.805 4 | 28.15/0.852 3 | 33.51/0.944 1 |
| SMSR | | 993 | 68 | 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 |
| ShuffleMixer | | 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 |
| SAFMN | | 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 |
| SMFANet | | 191 | 19 | 34.42/0.927 4 | 30.41/0.843 0 | 29.16/0.806 5 | 28.22/0.852 3 | 33.96/0.946 0 |
| SeemoRe-T | | 225 | — | 34.46/0.927 6 | 30.44/0.844 5 | 29.15/0.806 3 | 28.27/0.853 8 | 33.92/0.946 0 |
| EARFA-light | | 203 | — | 34.48/0.928 0 | 30.44/0.843 8 | 29.16/0.806 7 | 28.29/0.854 9 | 33.94/0.946 6 |
| MSGN-S | | 408 | 39 | 34.47/0.928 0 | 30.45/0.844 4 | 29.16/0.807 1 | 28.31/0.854 5 | — |
| SRConvNet | | 387 | 33 | 34.40/0.927 2 | 30.30/0.841 6 | 29.07/0.804 7 | 28.04/0.850 0 | 33.56/0.944 3 |
| GRFN | | 539 | — | 34.55/0.928 1 | 30.45/0.843 9 | 29.19/0.806 8 | 28.40/0.856 8 | 34.01/0.946 9 |
| MELTN | | 640 | 66 | 34.46/0.927 8 | 30.42/0.844 2 | 29.16/0.806 7 | 28.33/0.856 0 | 33.83/0.945 3 |
| MFAMNet(ours) | | 250 | 22 | 34.56/0.928 8 | 30.52/0.845 6 | 29.21/0.808 0 | 28.46/0.857 1 | 34.19/0.947 6 |
| Bicubic | ×4 | — | — | 28.42/0.810 4 | 26.00/0.702 7 | 25.96/0.667 5 | 23.14/0.657 7 | 24.89/0.786 6 |
| FSRCNN | | 12 | 5 | 30.71/0.865 7 | 27.59/0.753 5 | 26.98/0.715 0 | 24.62/0.728 0 | 27.90/0.851 7 |
| VDSR | | 665 | 613 | 31.35/0.883 8 | 28.01/0.767 4 | 27.29/0.725 1 | 25.18/0.752 4 | 28.83/0.880 9 |
| CARN | | 1 592 | 91 | 32.13/0.893 7 | 28.60/0.780 6 | 27.58/0.734 9 | 26.07/0.783 7 | 30.47/0.908 4 |
| EDSR-baseline | | 1 518 | 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 |
| IMDN | | 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 |
| LAPAR-A | | 659 | 94 | 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 |
| SMSR | | 1 006 | 42 | 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 |
| ShuffleMixer | | 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 |
| SAFMN | | 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 |
| SMFANet | | 197 | 11 | 32.25/0.895 6 | 28.71/0.783 3 | 27.64/0.737 7 | 26.18/0.786 2 | 30.82/0.910 4 |
| SeemoRe-T | | 232 | — | 32.31/0.896 5 | 28.72/0.784 0 | 27.65/0.738 4 | 26.23/0.788 3 | 30.82/0.910 7 |
| EARFA-light | | 209 | — | 32.33/0.896 4 | 28.68/0.783 2 | 27.64/0.738 2 | 26.20/0.788 9 | 30.75/0.911 5 |
| MSGN-S | | 404 | 25 | 32.37/0.897 0 | 28.74/0.784 5 | 27.66/0.738 7 | 26.27/0.789 1 | — |
| SRConvNet | | 382 | 22 | 32.18/0.895 1 | 28.61/0.781 8 | 27.57/0.735 9 | 26.06/0.784 5 | 30.35/0.907 5 |
| GRFN | | 554 | — | 32.30/0.896 5 | 28.74/0.784 5 | 27.65/0.737 8 | 26.33/0.791 5 | 30.84/0.912 1 |
| MELTN | | 641 | 37 | 32.35/0.895 9 | 28.77/0.783 1 | 27.64/0.738 1 | 26.17/0.789 9 | 30.74/0.910 8 |
| MFAMNet(ours) | | 257 | 12 | 32.41/0.897 9 | 28.79/0.786 2 | 27.70/0.740 0 | 26.37/0.792 0 | 31.08/0.913 9 |

在4倍SR下，尽管MFAMNet的参数量相较于SMFANet增加了60 KB，但其PSNR性能在五个基准数据集上相较于SMFANet平均提升了0.15 dB. 特别是在Urban100和Manga109数据集上，MFAMNet分别实现了0.19 dB和0.26 dB的PSNR提升，同时FLOPs与SMFANet几乎相同. 与SAFMN相比，MFAMNet在五个数据集的每个数据集上均实现了至少0.12 dB的PSNR增益. 此外，与传统的Bicubic方法相比，MFAMNet在PSNR性能上表现出显著优势，即便在PSNR提升幅度最小的B100数据集上，也实现了至少1.74 dB的增益. 与GRFN模型相比，MFAMNe的参数量约等于其1/2，但是却实现了更好的性能，例如在Set5和Mange109两个数据集的PSNR上，分别实现了0.11 dB和0.24 dB的性能提升. 与基于Transfomer的方法MELTN相比，本文的方法只用了约等于其1/3的参数量就实现了全面的性能领先.

### 4.2.2 定性比较

将本文提出的MFAMNet的可视化结果与基于CNN的方法在×2 Set14、×3 Set5和×4 Urban100数据集上进行比较，包括CARN[13]、IMDN[16]、ShuffleMixer[21]、SAFMN[20]和SMFANet[60]. 如图5~图7所示，本文列出了经过评估的基于CNN的方法对应的子图，并且在每个子图下方标记相应的PSNR/SSIM. 特别是在×4 Urban100数据集上，如

图7所示,其他方法存在模糊伪像和扭曲的线条. 相比之下,本文的方法可以恢复具有更清晰边界、更准确的平行直线和网格图案. 同时,还证明了本文通过利用非局部特征交互进行自适应特征调制的方法的有效性.

#### 4.2.3　内存和运行时间比较

为了充分证明本文提出的MFAMNet的效率,本文进一步将本文的方法与基于CNN的方法在图形处理单元(Graphics Processing Unit,GPU)内存消耗和4倍SR上的推理时间方面进行比较,包括CARN[13]、EDSR-baseline[18]、IMDN[16]、LAPAR-A[58]、ShuffleMixer[21]和SAFMN[20]. 表2显示了GPU内存和推理时间的比较. 其中#GPU内存表示推理阶段的最大GPU内存消耗,由torch.cuda.max_memory_allocated函数导出. #Avg.Time是500张320×180像素的LR图像的平均运行时间. 通过对表2的分析,本文的MFAMNet比除SAFMN之外的所有列出的CNN方法具有更少GPU内存消耗. 与ShuffleMixer[21]相比,MFAMNet的GPU内存消耗降低了65%,同时运行速度提升了9%. 此外,相较于LPARA-A[58],MFAMNet的GPU内存消耗减少了约91%,而运行速度则提高了1.5倍. 这些定量和定性比较结果表明:本文方法比先进的SR模型获得了更好的结果,但复杂性显著降低.

#### 4.2.4　LAM比较

局部归因图(Local Attribution Maps,LAM)是一种用于解释和分析SR网络的技术,旨在识别输入图像中哪些像素对SR网络输出的特定区域有显著影响. LAM的原理基于积分梯度方法(Integral Gradient Method),通过计算从基线输入(如模糊图像)到实际输入图像之间的路径上的梯度积分来确定每个输入像素的重要性. 模糊图像作为基线输入,表示高频成分(如边缘和纹理)的"缺失",而渐进式模糊函数作为路径函数,通过逐渐改变模糊核实现从基线输入到实际输入的平滑过渡. LAM通过简单的算子或滤波器(如梯度检测器)来量化局部区域中特定特征(如边缘和纹理)的存在,并特别关注难以重建的区域,以分析SR网络如何利用信息来提升性能. 对于给定的输入图像,LAM计算从基线输入到实际输入的路径上的积分梯度,以此确定每个像素对SR结果的贡献程度. LAM的作用不仅在于提供了一种可视化工具,帮助研究者理解SR网络的工作原理和关注区域,还在于通过分析LAM结果,揭示SR

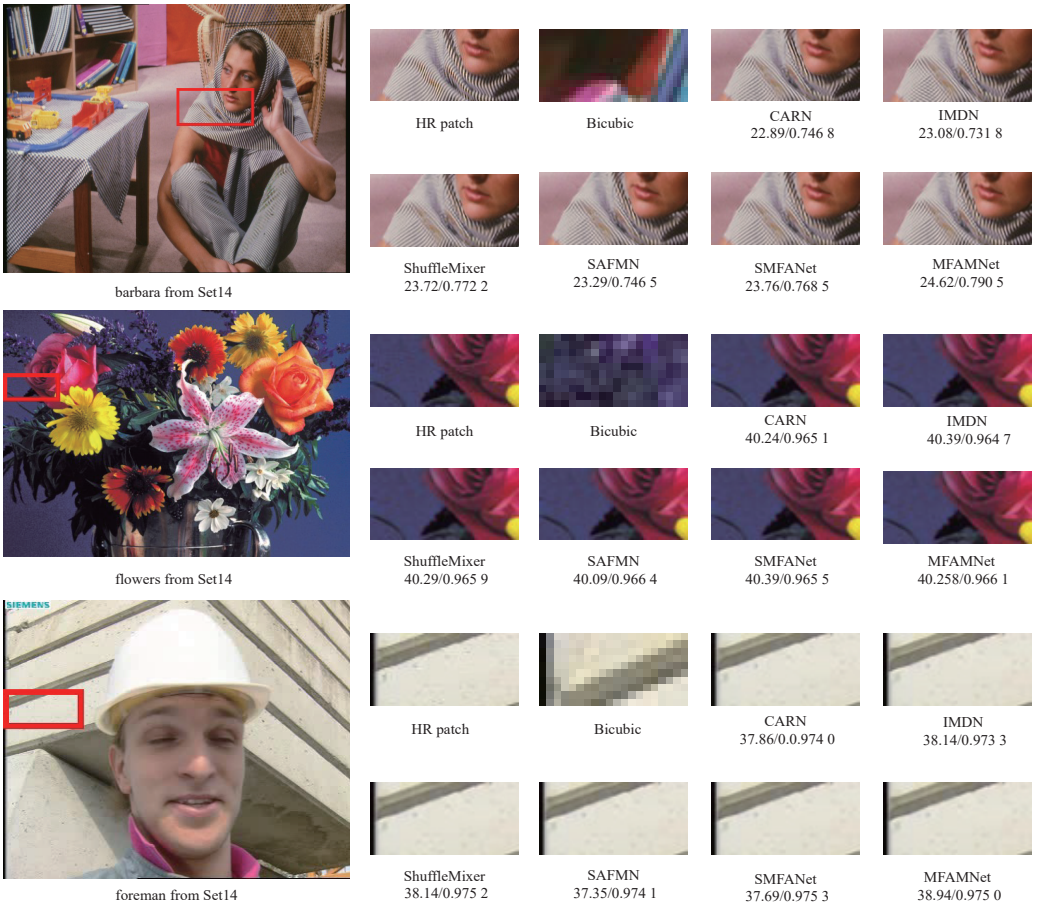



图5　Set14数据集上×2 SR的视觉对比

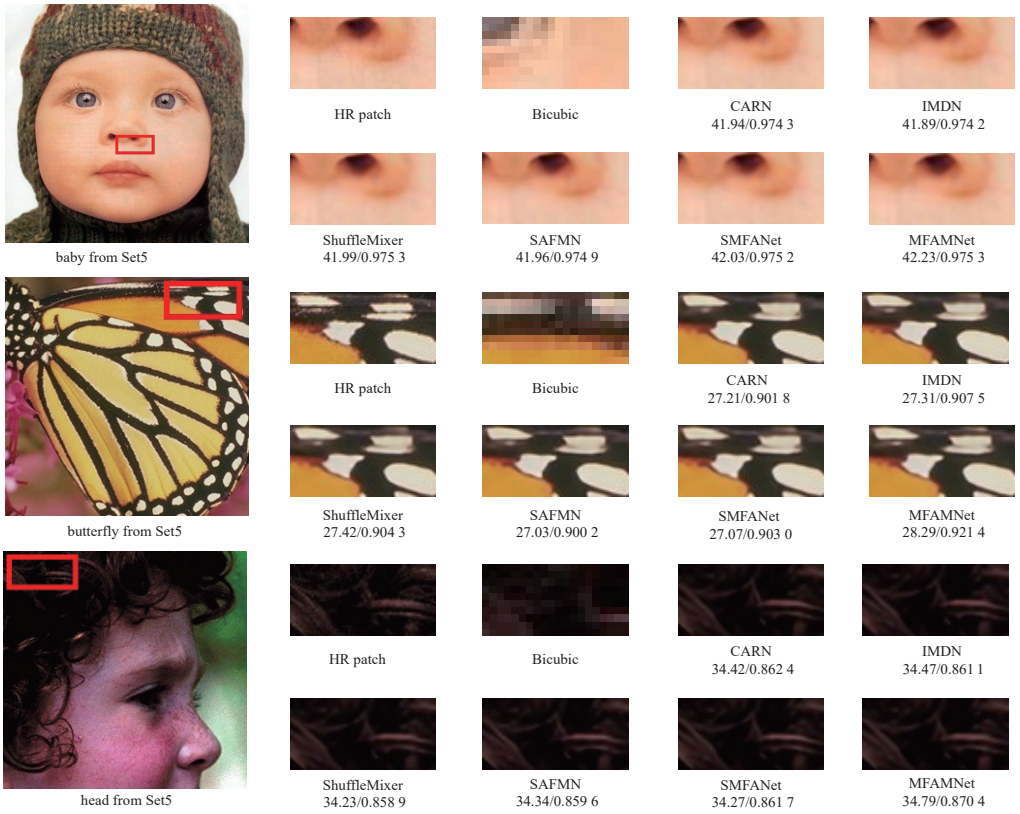



图 6 Set5 数据集上×3 SR 的视觉对比

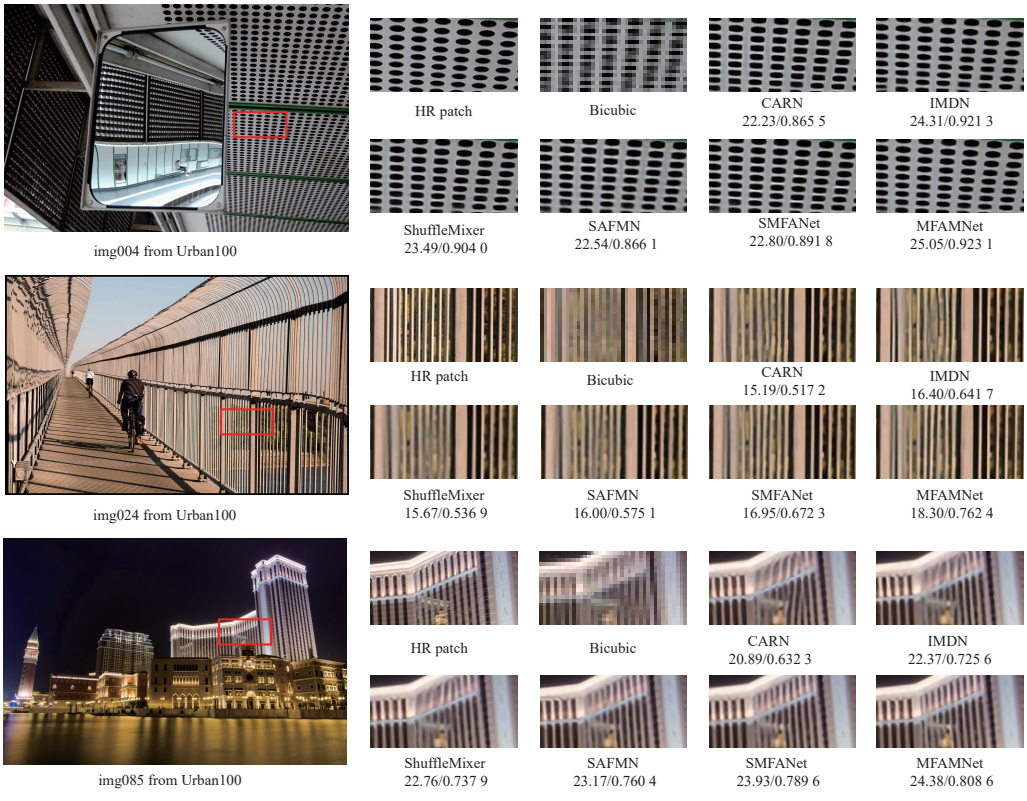



图 7 Urban100 数据集上×4 SR 的视觉对比

表2　不同方法的复杂度比较

| Methods | #GPU Mem/M | #Avg.Time/ms |
|---|---|---|
| CARN | 677.00 | 18.67 |
| EDSR-baseline | 480.12 | 16.31 |
| IMDN | 196.88 | 8.73 |
| LPARA-A | 1 808.23 | 30.03 |
| ShuffleMixer | 466.06 | 22.17 |
| SAFMN | 63.64 | 10.49 |
| MFAMNet | 162.40 | 20.02 |

网络在信息利用上的模式,从而指导设计更高效的网络架构.

LAM[76]表明在恢复过程中红色像素和矩形位置块之间存在显著相关性,如图8所示. 将本文的方法与基于CNN方法的LAM结果进行比较,包括LAPAR-A[58]、ShuffleMixer[21]、SAFMN[20]和SMFANet[60],并在图8中每个子图下方标记相应的扩散指数(DI)值,DI值越大表示涉及的像素范围越广. 与SAFMN[11]相比,MFAMNet的扩散指数增加了48%,而相较于LAPAR-A[58],MFAMNet的扩散指数更是显著提高了4.33倍. 这些结果表明:所提出的MFAMNet可以探索更多的非局部信息,以实现准确的图像SR.

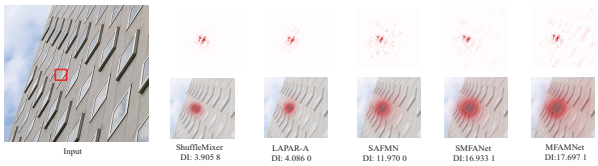



图8　LAM和扩散指数(DIs)的比较

## 5　分析与讨论

### 5.1　理论分析

#### 5.1.1　特征提取的优化

MFAMNet通过MFAM模块和SEM实现了高效的特征提取. MFAM模块通过不同下采样率(2、4、8倍)将输入特征分解为多级低频分量,利用深度卷积在大感受野下提取全局结构信息. 例如,8倍下采样特征的单个像素对应原始图像8×8区域的综合语义,使得模型能够捕捉长距离依赖关系,避免传统卷积因局部感受野限制导致的结构失真. 全局方差计算和特征调制机制能够增强模型对全局信息的感知能力. 全局方差反映了输入特征图的统计分布,通过将全局方差与低频特征结合,模型能够更好地利用全局信息来增强局部特征的表达能力. SEM模块通过大核条纹卷积和部分卷积,在原始分辨率下提取高频细节. 条纹卷积在感受野保持不变的情况下,极大地降低了计算复杂度和模型参数量,有效平衡细节提取与计算成本. 根据信息论可知,图像可分解为低频语义层和高频细节层的叠加. MFAMNet通过MFAM和SEM两个模块分别建模这两层信息,避免传统单尺度模型因跨尺度信息混杂导致的特征冗余,提升信息提取效率.

#### 5.1.2　信息融合的高效性

FA模块计算输入特征的全局方差,作为特征分布的统计量表征. 方差较高的区域对应图像中像素值变化剧烈的部分,通常包含关键细节信息,能够增强调制权重,迫使模型重点关注这些区域的特征聚合. 而方差较低的部分对应平滑区域,语义上多为背景,特征冗余度高,通过方差机制可以抑制冗余响应. 人类视觉系统会自动聚焦于场景中的“显著性区域”,忽略冗余背景. MFAM的方差调制通过强化高方差区域的特征聚合权重,模拟了这一机制,使模型优先处理对重建质量影响最大的信息.

MFAM的通道注意力机制通过全局平均池化将空间信息压缩为通道级统计量,捕捉跨通道依赖关系,动态抑制无关通道或噪声通道,增强语义相关通道的权重,减少通道间的信息冗余. SEM利用大核条纹卷积模拟得到的空间注意力图结合局部方差生成,突出高频细节所在的空间位置. 通道注意力提供语义层面的“内容筛选”,解决“哪些特征更重要”的问题,空间注意力实现几何层面的“位置校准”,解决“重要特征在哪里”的问题,两者结合形成“语义-几何”双约束的特征融合机制,提升重建精度.

#### 5.1.3　计算效率的提升

MFAM模块的子模块FA模块将输入通道分割为$C/4$(多尺度分支)和$3C/4$(原始分辨率分支),仅对小部分通道进行下采样处理. 例如,当$C$=36时,多尺度分支的通道数仅为9,大幅降低FLOPs. 这种设计通过减少参与下采样的通道数,在保留多尺度语义信息的同时,避免了全通道处理带来的冗余计算.

SEM采用部分卷积仅对部分通道执行卷积操作,减少了内存访问开销. 同时,采用条纹卷积将11×11核分解为1×11和11×1的连续卷积,计算量从标准大核卷积的$121C^2$降至$22C^2$,降幅达82%. 相较于原始大核,条纹卷积在保持相同感受野的前提下,通过一维卷积的顺序执行,将二维空间的密集计算转化为水平和垂直方向的稀疏计算,显著降低了乘加运算次数.

#### 5.1.4　模型复杂度与性能的权衡

SA通过计算全局像素对的相似性矩阵实现非局部建模,其本质是一种动态加权求和,但二次方的复杂度问题使得SA机制在处理HR图像时面临显著的计算和内存挑战. MFAM模块则通过多尺度下采样和方差调制模拟非局部依赖,利用不同尺度的低频特征聚合替代全局像素交互,复杂度与图像大小呈线性关系,避免了SA的二次增长缺陷. SA机制的低通滤波特性更易捕

捉低频语义，限制了其对高频局部细节的捕捉能力，导致重建结果过于平滑，缺乏细节. 而MFAM通过原始分辨率分支（$Y$路径）保留高频细节，结合SEM模块实现高低频解耦，能够同时捕捉全局信息和局部细节. 这种设计使得MFAM模块在建模长距离依赖关系的同时，能够保留更多的局部细节，更适合SR任务对细节重建的需求.

传统空间注意力通过局部卷积生成注意力图，仅捕捉局部邻域的空间关系；MFAM通过多尺度下采样（最大8倍）扩大建模范围，等效于捕捉（8×8）像素的全局空间关系，兼具局部细节和全局结构建模能力. 空间注意力依赖局部特征强度生成权重，而MFAM引入全局方差作为引导信号，从特征分布的统计层面强化显著性区域，避免局部噪声干扰.

尽管与SA和空间注意力相比，MFAM模块有种种优势，但是SA的全局动态建模能力在复杂场景中更具优势，而MFAM的多尺度聚合属于“伪非局部”建模，对极长距离依赖的捕捉能力较弱. 此外，MFAM模块的多尺度下采样可能导致极高频细节的轻微损失，未来可通过引入可变形卷积或动态尺度选择机制进一步优化.

### 5.2 消融实验

本节进行了广泛的消融研究，以分析和评估所提出的MFAMNet中每个组件的效果. 本文基于×4 MFAMNet模型实现了所有消融实验，并使用DF2K数据集对其进行训练以进行公平比较. 表3中的定量消融结果是在Urban100和Manga109数据集上测量的.

本文所提出的MFAM模块提取多尺度特征以提高重建精度. 为了证明其有效性，本文首先移除MFAM模块，将其与基线MFAMNet进行比较. 如表3所示，Urban100和Manga109数据集上的PSNR值分别下降了0.66 dB和0.86 dB. 这些结果表明了MFAM的重要性. 此外，由于所提出的MFAM模块主要包含一个用于处理不同分辨率特征的分支和一个用于处理原始分辨率特征的分支，这里分别用 $X$ 和 $Y$ 分别表示这两个分支. 本文对这些组件进行了消融研究，以证明它们在图像SR上的有效性. 如果没有 $X$ 分支，在Urban100和Manga109数据集上观察到PSNR性能显著下降0.29 dB和0.41 dB. 这些结果表明：$X$ 分支可以有效地提取和融合不同分辨率的特征. 至于 $Y$ 分支，移除后，模型在Urban100和Manga109上的性能仅达到26.15 dB和30.82 dB，这也说明了原始分辨率的特征对于单图SR同样很重要. 由于$X$分支利用不同的下采样尺度来提取多尺度特征，因此本文还对不同的下采样分支进行了消融实验，包括单独移除某个下采样尺度和组合移除两个尺度. 无论是单独移除某一个下采样尺度，还是同时移除两个下采样尺度，模型在Urban100和Manga109的PSNR性能都有所下降. 例如，同时移除下采样2倍和8倍，模型在Urban100的PSNR下降了0.1 dB，这充分表明了三个不同尺度下的特征对于单图SR都很重要. 此外，本文还比较了单独移除某个下采样尺度和组合移除两个尺度提取的特征图，如图9所示，原始网络提取出的特征图能看到连续且完整的高频结构，结构清晰统一，既有全局轮廓，又有细密纹理，这些是还原HR细节的关键，而移除某个下采样尺度或组合移除两个尺度的特征图，由于丢失某一层级信息，出现细节断裂、纹理模糊、结构破碎、语义混乱等问题.

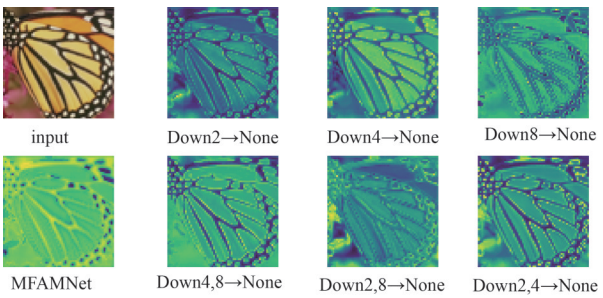



图9 部分消融实验的中间特征图比较

SEM利用大核条纹卷积，有效提取并融合局部空间信息，以进一步完善MFAM在空间和通道维度上的特征. 为了证明该模块的有效性，本文分别进行了移除SEM和将SEM替换成前馈神经网络（Feed-Forward Network，FFN）[77]的消融实验. 由表3可知，与基线相比，移除SEM导致PSNR值在Urban100和Manga109数据集上分别下降了0.19 dB和0.32 dB. 而将SEM替换成FFN[77]之后，虽然参数量减少了35 K，但是在Urban100和Manga109数据集上的PSNR值却降低了0.17 dB和0.19 dB. 这些结果证明SEM可以有效地提取局部特征，进一步对MFAM提取的特征进行补充和完善.

在MFAM和SEM两个模块中，本文均使用了方差机制. 为了证明其有效性，本文分别移除了这两个模块中的方差操作. 表3表明：在移除方差操作之后，模型参数量和FLOPs几乎没有减少，而PSNR值却降低了. 具体来说，MFAM模块在移除方差操作之后，在Urban100和Manga109数据集上的PSNR值分别下降了0.06 dB和0.05 dB. 而SEM模块在移除方差操作之后，在Urban100和Manga109数据集上的PSNR值分别下降了0.03 dB和0.08 dB. 这些结果表明：方差调制机制可以在几乎不增加模型复杂度的情况下增强模型对全局和局部信息的提取能力.

表3　MFAMNet在Urban100和Manga109数据集上的消融实验

| Ablation | Variant | #Params/K | #FLOPs/G | Urban100 | Manga109 |
|---|---|---|---|---|---|
| Baseline | MFAMNet | 257 | 12 | 26.37/0.792 0 | 31.08/0.913 9 |
| MFAM | MFAM→None | 85 | 5 | 25.81/0.775 5 | 30.22/0.903 9 |
| | *X*→None | 153 | 9 | 26.08/0.783 3 | 30.67/0.909 3 |
| | *Y*→None | 188 | 8 | 26.15/0.785 4 | 30.82/0.910 6 |
| | Down2→None | 245 | 12 | 26.33/0.791 7 | 31.03/0.913 4 |
| | Down4→None | 245 | 12 | 26.32/0.791 1 | 31.05/0.913 4 |
| | Down8→None | 245 | 12 | 26.30/0.790 3 | 31.04/0.913 5 |
| | Down2,4→None | 222 | 11 | 26.28/0.790 1 | 30.98/0.912 8 |
| | Down2,8→None | 222 | 11 | 26.27/0.789 4 | 30.99/0.912 9 |
| | Down4,8→None | 222 | 11 | 26.29/0.789 7 | 30.98/0.912 7 |
| | Variance→None | 257 | 12 | 26.31/0.791 5 | 31.03/0.913 5 |
| SEM | SEM→None | 201 | 10 | 26.12/0.783 9 | 30.76/0.909 7 |
| | SEM→FNN | 222 | 11 | 26.20/0.786 1 | 30.89/0.911 2 |
| | Variance→None | 257 | 12 | 26.34/0.791 6 | 31.00/0.912 9 |

## 6　结论

本文提出了一种简单而高效的深度CNN模型,旨在解决图像SR任务中的效率与性能平衡问题. 所提出的MFAMNet方法创新性地引入了基于多尺度特征表示的调制机制,通过自适应地捕捉远程依赖关系,显著提升了模型对全局信息的提取能力. 为了进一步增强局部上下文信息的利用,本文设计了一种SEM,该模块能够有效地编码空间局部上下文,使得模型在图像处理过程中能够更加精确地复原出细腻的纹理和清晰的边缘细节. 本文在多个常用基准数据集上对所提出的方法进行了全面的定性和定量评估. 大量实验结果表明:MFAMNet模型在重建性能和计算效率之间实现了更优的权衡,不仅显著提升了SR图像的质量,还保持了较低的计算复杂度,为实际应用提供了可行的解决方案.

## 参考文献


[1] KEYS R. Cubic convolution interpolation for digital image processing[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1981, 29(6): 1153-1160.

[2] YANG J C, WRIGHT J, HUANG T S, et al. Image super-resolution via sparse representation[J]. IEEE Transactions on Image Processing, 2010, 19(11): 2861-2873.

[3] YANG J C, WRIGHT J, HUANG T, et al. Image super-resolution as sparse representation of raw image patches[C]// 2008 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2008: 1-8.

[4] XU H T, ZHAI G T, YANG X K. Single image super-resolution with detail enhancement based on local fractal analysis of gradient[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(10): 1740-1754.

[5] WANG L F, XIANG S M, MENG G F, et al. Edge-directed single-image super-resolution via adaptive gradient magnitude self-interpolation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(8): 1289-1299.

[6] ZHANG K B, GAO X B, TAO D C, et al. Single image super-resolution with non-local means and steering kernel regression[J]. IEEE Transactions on Image Processing, 2012, 21(11): 4544-4556.

[7] FREEMAN W T, JONES T R, PASZTOR E C. Example-based super-resolution[J]. IEEE Computer Graphics and Applications, 2002, 22(2): 56-65.

[8] CHANG H, YEUNG D Y, XIONG Y M. Super-resolution through neighbor embedding[C]//Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. Piscataway: IEEE, 2004: I.

[9] SCHULTER S, LEISTNER C, BISCHOF H. Fast and accurate image upscaling with super-resolution forests[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 3791-3799.

[10] YU J F, GAO X B, TAO D C, et al. A unified learning framework for single image super-resolution[J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25(4): 780-792.

[11] DENG C, XU J, ZHANG K B, et al. Similarity constraints-based structured output regression machine: An approach to image super-resolution[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(12): 2472-2485.

[12] YANG W M, TIAN Y P, ZHOU F, et al. Consistent cod-

ing scheme for single-image super-resolution via independent dictionaries[J]. IEEE Transactions on Multimedia, 2016, 18(3): 313-325.

[13] 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.

[14] 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.

[15] 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.

[16] 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: ACM, 2019: 2024-2032.

[17] LI Z Y, LIU Y Q, CHEN X Y, et al. Blueprint separable residual network for efficient image super-resolution[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2022: 832-842.

[18] 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. Piscataway: IEEE, 2017: 1132-1140.

[19] SHI W Z, CABALLERO J, HUSZÁR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1874-1883.

[20] SUN L, DONG J X, TANG J H, et al. Spatially-adaptive feature modulation for efficient image super-resolution[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2023: 13144-13153.

[21] 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.

[22] ZHANG Y L, LI K P, LI K, et al. Image Super-resolution using Very Deep Residual Channel Attention Networks[M]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 294-310.

[23] 高丹丹, 周登文, 王婉君, 等. 特征频率分组融合的轻量级图像超分辨率重建[J]. 计算机辅助设计与图形学学报, 2023, 35(7): 1020-1031.
GAO D D, ZHOU D W, WANG W J, et al. Lightweight super-resolution via grouping fusion of feature frequencies[J]. Journal of Computer-Aided Design & Computer Graphics, 2023, 35(7): 1020-1031. (in Chinese)

[24] 吴靖, 叶晓晶, 黄峰, 等. 基于深度学习的单帧图像超分辨率重建综述[J]. 电子学报, 2022, 50(9): 2265-2294.
WU J, YE X J, HUANG F, et al. A review of single image super-resolution reconstruction based on deep learning[J]. Acta Electronica Sinica, 2022, 50(9): 2265-2294. (in Chinese)

[25] 缪永伟, 张新杰, 任瀚实, 等. 基于通道多尺度融合的场景深度图超分辨率网络[J]. 计算机辅助设计与图形学学报, 2023, 35(1): 37-47.
MIAO Y W, ZHANG X J, REN H S, et al. A channel multi-scale fusion network for scene depth map super-resolution[J]. Journal of Computer-Aided Design & Computer Graphics, 2023, 35(1): 37-47. (in Chinese)

[26] 周登文, 刘子涵, 刘玉铠. 基于像素对比学习的图像超分辨率算法[J]. 自动化学报, 2024, 50(1): 181-193.
ZHOU D W, LIU Z H, LIU Y K. Pixel-wise contrastive learning for single image super-resolution[J]. Acta Automatica Sinica, 2024, 50(1): 181-193. (in Chinese)

[27] 王云涛, 赵蔺, 刘李漫, 等. 基于组-信息蒸馏残差网络的轻量级图像超分辨率重建[J]. 自动化学报, 2024, 50(10): 2063-2078.
WANG Y T, ZHAO L, LIU L M, et al. G-IDRN: A group-information distillation residual network for lightweight image super-resolution[J]. Acta Automatica Sinica, 2024, 50(10): 2063-2078. (in Chinese)

[28] 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. Piscataway: IEEE, 2022: 765-775.

[29] LIU J, TANG J, WU G S. Residual feature distillation network for lightweight image super-resolution[M]//Computer Vision - ECCV 2020 Workshops. Cham: Springer International Publishing, 2020: 41-55.

[30] MICHELINI P N, LU Y H, JIANG X Q. Edge-SR: Super-resolution for the masses[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2022: 4019-4028.

[31] ZHAO H Y, KONG X T, HE J W, et al. Efficient Image Super-resolution using Pixel Attention[M]//Computer Vision - ECCV 2020 Workshops. Cham: Springer Interna-

tional Publishing, 2020: 56-72.

[32] HE Z B, DAI T, LU J, et al. Fakd: Feature-affinity based knowledge distillation for efficient image super-resolution[C]//2020 IEEE International Conference on Image Processing. Piscataway: IEEE, 2020: 518-522.

[33] CHU X X, ZHANG B, MA H L, et al. Fast, accurate and lightweight super-resolution with neural architecture search[C]//Proceedings of the 25th International Conference on Pattern Recognition. Piscataway: IEEE, 2021: 59-64.

[34] SONG D H, XU C, JIA X, et al. Efficient residual dense block search for image super-resolution[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12007-12014.

[35] ZHANG X D, ZENG H, ZHANG L. Edge-oriented convolution block for real-time super resolution on mobile devices[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 4034-4043.

[36] 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. Piscataway: IEEE, 2017: 105-114.

[37] WANG X T, YU K, WU S X, et al. ESRGAN: Enhanced super-resolution generative adversarial networks[M]//Computer Vision-ECCV 2018 Workshops. Cham: Springer International Publishing, 2019: 63-79.

[38] SAHARIA C, HO J, CHAN W, et al. Image super-resolution via iterative refinement[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(4): 4713-4726.

[39] RAMESH A, PAVLOV M, GOH G, et al. Zero-shot text-to-image generation[C]//International Conference on Machine Learning. Cambridge: PMLR, 2021: 8821-8831.

[40] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 10674-10685.

[41] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[EB/OL]. (2021-06-03) [2024-10-31]. https://arXiv.org/abs/2010.11929.

[42] CHEN H T, WANG Y H, GUO T Y, et al. Pre-trained image processing transformer[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 12294-12305.

[43] CHOI H, LEE J, YANG J. N-gram in swin transformers for efficient lightweight image super-resolution[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 2071-2081.

[44] LI M, MA B, ZHANG Y L. Lightweight image super-resolution with pyramid clustering transformer[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, PP(99): 1.

[45] LIANG J Y, CAO J Z, SUN G L, et al. SwinIR: Image restoration using swin transformer[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops. Piscataway: IEEE, 2021: 1833-1844.

[46] LIU J, CHEN C, TANG J, et al. From coarse to fine: Hierarchical pixel integration for lightweight image super-resolution[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(2): 1666-1674.

[47] WANG H, CHEN X H, NI B B, et al. Omni aggregation networks for lightweight image super-resolution[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 22378-22387.

[48] ZHANG A P, REN W Q, LIU Y, et al. Lightweight image super-resolution with superpixel token interaction[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2024: 12682-12691.

[49] ZHANG X D, ZENG H, GUO S, et al. Efficient long-range attention network for image super-resolution[M]//Computer Vision - ECCV 2022. Cham: Springer Nature Switzerland, 2022: 649-667.

[50] ZHOU Y P, LI Z, GUO C L, et al. SRFormer: Permuted self-attention for single image super-resolution[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2024: 12734-12745.

[51] 唐述, 曾琬凌, 杨书丽, 等. 基于Transformer的块内块间双聚合的单图像超分辨率重建网络[J]. 计算机学报, 2024, 47(12): 2783-2802.
TANG S, ZENG W L, YANG S L, et al. Intra-block and inter-block dual aggregation transformer for single image super-resolution[J]. Chinese Journal of Computers, 2024, 47(12): 2783-2802. (in Chinese)

[52] 毕修平, 陈实, 张乐飞. 轻量级图像超分辨率的蓝图可分离卷积Transformer网络[J]. 中国图象图形学报, 2024, 29(4): 875-889.
BI X P, CHEN S, ZHANG L F. Blueprint separable convolution Transformer network for lightweight image su-

per-resolution[J]. Journal of Image and Graphics, 2024, 29(4): 875-889. (in Chinese)

[53] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2022: 9992-10002.

[54] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: Efficient transformer for high-resolution image restoration[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 5718-5729.

[55] ZHANG J N, PENG H W, WU K, et al. MiniViT: Compressing vision transformers with weight multiplexing[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 12135-12144.

[56] DONG J X, PAN J S, YANG Z B, et al. Multi-scale residual low-pass filter network for image deblurring[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2024: 12311-12320.

[57] PARK N, KIM S. How do vision transformers work? [EB/OL]. (2022-06-08)[2024-10-31]. https://arXiv.org/abs/2202.06709.

[58] 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.

[59] 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. Piscataway: IEEE, 2021: 4915-4924.

[60] ZHENG M J, SUN L, DONG J X, et al. SMFANet: A lightweight self-modulation feature aggregation network for efficient image super-resolution[M]//Computer Vision-ECCV 2024. Cham: Springer Nature Switzerland, 2024: 359-375.

[61] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1646-1654.

[62] TAI Y, YANG J, LIU X M. Image super-resolution via deep recursive residual network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2790-2798.

[63] LU Z S, LI J C, LIU H, et al. Transformer for single image super-resolution[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2022: 456-465.

[64] WANG H, CHEN X H, NI B B, et al. Omni aggregation networks for lightweight image super-resolution[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 22378-22387.

[65] CHEN X Y, WANG X T, ZHOU J T, et al. Activating more pixels in image super-resolution transformer[C]// 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 22367-22377.

[66] LI A, ZHANG L, LIU Y, et al. Feature modulation transformer: Cross-refinement of global representation via high-frequency prior for image super-resolution[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2024: 12480-12490.

[67] LI Y W, FAN Y C, XIANG X Y, et al. Efficient and explicit modelling of image hierarchies for image restoration[C]// 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 18278-18289.

[68] ZHOU L, CAI H M, GU J J, et al. Efficient image super-resolution using vast-receptive-field attention[M]//Computer Vision - ECCV 2022 Workshops. Cham: Springer Nature Switzerland, 2023: 256-272.

[69] MAO Y Y, ZHANG N H, WANG Q, et al. Multi-level dispersion residual network for efficient image super-resolution[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2023: 1660-1669.

[70] MIAO X, LI S J, LI Z, et al. Multi-scale gated network for efficient image super-resolution[J]. The Visual Computer, 2025, 41(2): 1227-1239.

[71] ZAMFIR E, WU Z, MEHTA N, et al. See more details: Efficient image super-resolution by experts mining[C]// Proceedings of the 41st International Conference on Machine Learning. Cambridge: PMLR, 2024: 1-16.

[72] ZHAO X L, LI L Z, XIE C X, et al. Efficient single image super-resolution with entropy attention and receptive field augmentation[C]//Proceedings of the 32nd ACM International Conference on Multimedia. New York: ACM, 2024: 1302-1310.

[73] LI F, CONG R M, WU J J, et al. SRConvNet: A transformer-style ConvNet for lightweight image super-resolution[J]. International Journal of Computer Vision, 2025, 133(1): 173-189.

[74] CHANG K R, SUN J, YANG B, et al. A multi-scale enhanced large-kernel attention transformer network for lightweight image super-resolution[J]. Signal, Image and Video Processing, 2025, 19(3): 204.

[75] YANG X, HONG C M, ZHANG P P. GRFN: A group residual feature network for lightweight image super-resolution[J]. Circuits, Systems, and Signal Processing, 2025, 44(5): 3513-3533.

[76] GU J J, DONG C. Interpreting super-resolution networks with local attribution maps[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 9195-9204.

[77] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017). Long Beach: Curran Associates, Inc., 2017: 5998-6008.

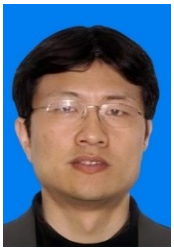
## 作者简介

**沈伟露**　男,2001年7月生,江苏淮安人.现为南京大学计算机学院硕士研究生.主要研究方向为计算机视觉、图像超分辨率.

E-mail: 522023330079@nju.edu.cn

**刘　杰**　男,1992年8月生,重庆人.现为南京大学计算机学院MCG实验组助理研究员.主要研究方向为视频/图像增强、视频插帧、姿态估计.

E-mail: jieliu@smail.nju.edu.cn



**唐　杰**　男,1971年4月生,江苏南京人.现为南京大学计算机学院副教授、博士生导师.主要研究方向为对象建模、大规模并行计算.

E-mail: tangjie@nju.edu.cn

**武港山**　男,1967年2月生,江苏盐城人.现为南京大学计算机学院教授、博士生导师.主要研究方向为媒体内容分析、多媒体信息检索.

E-mail: gswu@nju.edu.cn