

doi:10.3969/j.issn.1673-9833.2025.06.006

# 基于细化聚合多频特征的图像超分辨率研究

吴大荣, 胡仕刚

(湖南科技大学 信息与电气工程学院, 湖南 湘潭 411201)

**摘要:** 基于 Transformer 的方法在提取全局上下文方面表现优异, 且在单图像超分辨率 (SISR) 方面拥有显著效果, 但因其主要功能是捕获全局特征, 这使得它更侧重于捕获低频信息, 从而忽略了对于高频特征的提取。为解决这一问题, 提出了一种集成卷积和 Transformer 结构优势的多频特征聚合网络 (MFAN)。该网络由 3 个重要模块组成: 用于提取全局上下文的耦合自注意力 Transformer (CSAT)、用于提取并增强高频信息的高频增强模块 (HFEM), 以及用于细化全局特征的细化融合模块 (RFM)。通过实验得知, 与其他 SR 方法相比, 所提出的 MFAN 显著提高了分辨的视觉效果和图像质量。

**关键词:** 超分辨率; 耦合自注意力机制; 高频增强; 细化融合模块; 多频特征聚合网络 (MFAN)

**中图分类号:** TN911.73 **文献标志码:** A **文章编号:** 1673-9833(2025)06-0037-07

**引文格式:** 吴大荣, 胡仕刚. 基于细化聚合多频特征的图像超分辨率研究 [J]. 湖南工业大学学报, 2025, 39(6): 37-43.

## Research on Image Super Resolution Based on Refinement Aggregation of Multi Frequency Features

WU Darong, HU Shigang

(School of Information and Electrical Engineering, Hunan University of Science and Technology,  
Xiangtan Hunan 411201, China)

**Abstract:** Given its good performance in extracting global context as well as its significant performance in single image super-resolution (SISR), the Transformer based method focuses more on capturing low-frequency information thus neglecting the extraction of high-frequency features, due to the fact that the main function of Transformer is for global feature capturing. In view of a solution of this issue, a multi-frequency feature aggregation network (MFAN) has thus been proposed with the advantages of convolution and transformer structures integrated together. This network consists of three important modules: the coupled self-attention transformer (CSAT) for extracting global context, the high-frequency enhancement module (HFEM) for extracting and enhancing high-frequency information, and the refinement fusion module (RFM) for refining global features. It is found that, compared with other SR methods, the proposed multi-frequency feature aggregation network is characterized with a significant improvement of the visual resolution and image quality based on experimental results.

**Keywords:** super-resolution; coupling self-attention mechanism; high-frequency enhancement; fusion module refinement; multi-frequency feature aggregation network (MFAN)

收稿日期: 2024-11-26

作者简介: 吴大荣, 男, 湖南科技大学硕士生, 主要研究方向为深度学习, E-mail: 3306220653@qq.com

通信作者: 胡仕刚, 男, 湖南科技大学教授, 博士, 硕士生导师, 主要研究方向为深度学习, E-mail: hsg99528@126.com

## 1 研究背景

单图像超分辨率 (single image super-resolution, SISR), 即从已知低分辨率 (low resolution, LR) 图像中恢复其对应高分辨率图像, 该任务在计算机视觉领域尤为经典。由于超分辨率拥有恢复图像纹理细节功能, 被广泛用于医学图像增强<sup>[1]</sup>、监控视频<sup>[2]</sup>、面部识别等领域。然而单个图像提供的先验信息有限, 因此对于一个给定的 LR 图像存在多个解, 这是典型的不适定问题。近年来, 研究人员提出了各种基于卷积神经网络 (convolutional neural network, CNN) 和视觉 Transformer 模型来解决这一问题。

1) 基于 CNN。基于深度学习的超分辨率方法比传统算法具更好的性能, 其中图像超分辨率的卷积神经网络 (convolutional network for image super-resolution, SRCNN)<sup>[3]</sup> 的开创性工作, 在 SISR 方面取得了重大进展, 导致各种基于 CNN 的工作被提出。例如, Kim J. 等<sup>[4]</sup> 提出了一种利用深度网络级联 20 层的超分辨率 (super-resolution, SR) 方法, 获得了良好效果。为了获得更好的重构性能, Li F. 等<sup>[5]</sup> 介绍了一种涉及预定义滤波器的方法, 利用递归扩张残差网络构建了空间调制机制, 在聚合高频信息时有效利用了大区域上下文信息。Hui Z. 等<sup>[6]</sup> 提出了一种选择性融合方法, 通过引入生成对抗网络, 利用级联信息蒸馏块构建一个有效模型以挖掘丰富的多尺度信息, 追求更好的结构保存 SR 重建结果。上述基于 CNN 的方法取得了较好性能, 但因 CNN 接受域有限, 无法捕获全局上下文信息, 不利于细小纹理的恢复。

2) 视觉 Transformer。Transformer 中的自注意 (self attention, SA) 机制捕获了基本信息间的依赖关系, 可计算单位信道和像素间的权值, 并对信道和空间维度<sup>[7]</sup> 的长期依赖关系建模。Chen H. T. 等<sup>[8]</sup> 提出了

一种预训练图像处理 Transformer, 即在特征图处理阶段引入一个 Transformer 模块, 研究结果表明, 预训练机制可显著提高低水平任务性能。Yang F. Z. 等<sup>[9]</sup> 提出了一种用于重构 SR 的纹理 Transformer, 其采用跨尺度堆叠方法建立深度特征间的对应关系, 并传输精确的纹理特征。Wang Z. D. 等<sup>[10]</sup> 利用 Transformer 构建了一个分层编解码器网络, 通过在多个尺度上恢复调制器调整解码器的层次特征。然而这些网络只考虑图像的全局信息, 而卷积机制和注意机制<sup>[11]</sup> 之间的特征提取能力存在差异, 使得有效地聚合局部特征和全局特征变得困难。

为解决上述问题, 本文结合 CNN 和 Transformer 的优势, 提出了一种多频特征聚合网络 (multi-frequency feature aggregation network, MFAN) 模型。在考虑通道和空间域关键信息时, 利用 Transformer 较强的全局建模能力自适应捕获全局信息和多尺度特征。具体来说, 本文提出了一种耦合自注意变换器 (coupled self-attention Transformer, CSAT), 它自适应地加权空间维度和信道维度分支特征, 将空间信息和信道信息聚合到一单一注意力模块中, 以实现强大的全局特征表示。此外, 设计了一个用于提取高频信息的高频增强模块 (high-frequency enhancement module, HFEM), 其通过融合基于通道注意的局部特征提取分支和基于空间注意的高频信息增强分支增强高频信息。最后, 为防止多种频率特征信息混合导致重叠和信息冗余, 设计了一个细化融合模块 (refine the fusion module, RFM), 其在融合不同层次特征信息及在图像中捕获细节方面起重要作用。

## 2 多频特征聚合网络

本文提出的 MFAN 结构如图 1 所示。

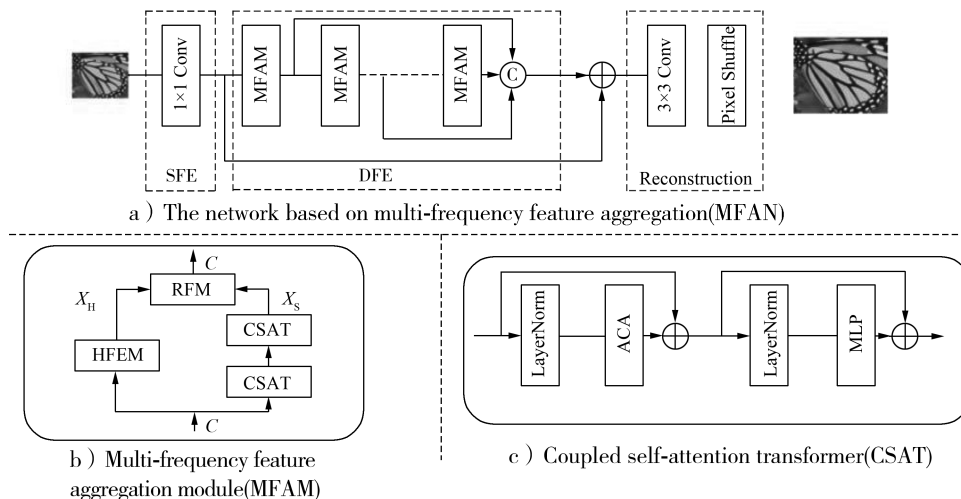


图 1 多频特征聚合网络的整体结构及各模块细节

Fig. 1 Overall structure of MFAN with its module details

图1所示网络除了拥有良好的全局建模能力, 还可捕获高频图像信息。MFAM由3个关键部分组成: CSAT、HFEM及RFM。

### 2.1 网络架构

如图1a所示, MFAN包括浅层特征(shallow features, SF)提取层、深度特征提取层和图像重建层。首先, 给定一个LR输入图像,  $I_{LR} \in \mathbf{R}^{H \times W \times C}$ , 通过一个SF提取层提取SF,  $F_0 \in \mathbf{R}^{H \times W \times C}$ 。其中,  $C$ 为输入图像特征通道数, 通常为3, 故本文选择3;  $H$ 和 $W$ 分别为输入图像的高度和宽度。此操作表示为 $F_0 = H_{SF}(I_{LR})$ , 其中 $H_{SF}(\cdot)$ 为SF提取层。

上述操作将LR图像从低维空间升级到高维空间。然后, 将得到的SF作为下一层的输入信息, 使用深度特征提取层从 $F_0$ 中提取深度特征 $F_d \in \mathbf{R}^{H \times W \times C}$ , 该模块由 $d$ 个深度特征提取模块组成, 可以被表述为 $F_d = H_{MFAM}^d(H_{MFAM}^{d-1} \cdots (H_{MFAM}^1(F_0)))$ , 式中 $F_d \in \mathbf{R}^{H \times W \times C}$ , 为从第 $d$ 个多频特征聚合模块(multi-frequency feature aggregation module, MFAM)中提取的中间特征, 本实验中,  $d$ 被设为4。该MFAM模块由两个CSAT堆栈组成。具体来说, CSAT是基于自适应耦合注意(adaptive coupling attention, ACA)

构建的Transformer模块。MFAM和CSAT的具体结构分别如图1b和c所示。最后, 将得到的融合特征 $F_{FF}$ 作为重构层输入, 重构层由 $3 \times 3$ 卷积层和一个亚像素(pixel shuffle, PS)卷积层组成。PS卷积层具更大的接受域, 这提供了额外的上下文来生成更真实的细节, 帮助生成更准确的高分辨率图像。重建阶段表述为 $I_{SR} = H_{PS}(F_0 + F_{FF})$ , 式中 $H_{PS}(\cdot)$ 为重建层, 用所有特征信息重建图像, 生成最终结果 $I_{SR}$ 。

### 2.2 耦合自注意变换器

为了构建一种能够在空间和通道维度上同时聚集全局特征和局部特征的自适应交互作用网络, 本文提出了一种ACA, 其设计的主要目的, 是在实现空间和通道之间的信息交换时, 还能够解决Transformer对于局部信息处理的不足。因此, 特地引入了局部特征, 并且将其与全局信息自适应融合。具体的设计细节见图2。首先, 利用窗口内的自注意机制作为一个分支, 计算注意矩阵来获取全局特征。然后, 将 $V$ 值通过深度卷积作为另一个分支。为使这两个特征能更好地融合, 在融合前分别在通道和空间维度进行交互作用。

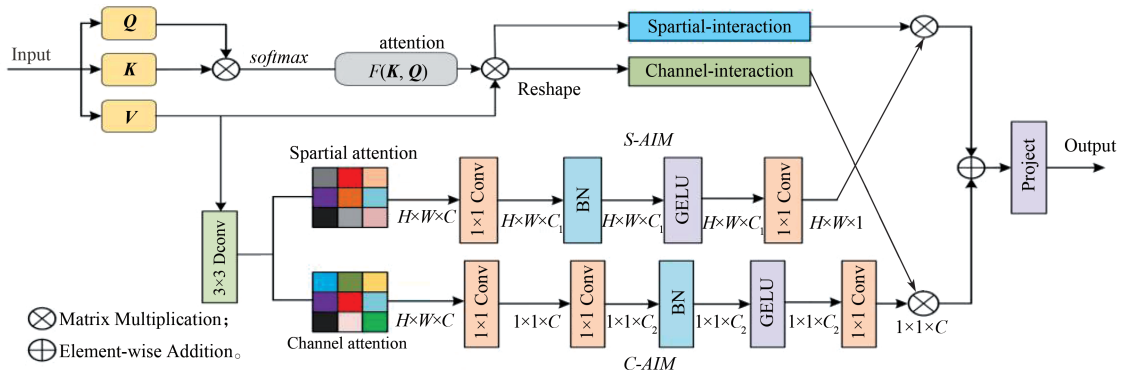


图2 自适应耦合注意模块

Fig. 2 Adaptive coupling attention module

1) 空间和通道相互作用。空间交互利用图像中的空间相关性提高超分辨率性能, 而通道交互通过组合和调整图像的不同通道增强图像细节和纹理。如图2, 首先使用卷积层降低输入特征图维数, 从而降低计算复杂度和参数数量; 然后引入批归一化(batch normalization, BN)层标准化上一层输出, 不仅可加快训练过程, 且使其更稳定; 接下来用GELU(Gaussian error linear unit)函数作为非线性激活函数, 以高斯方式平滑输入, 帮助学习更复杂的模式; 最后再次使用卷积层将特征映射到单通道输出。与空间交互不同, 信道交互模块首先执行自适应平均池化操作, 将输入特征映射调整为输出大小, 实现了全局信息的

聚合和压缩。

2) 自适应交互模块。为解决自注意中局部特征处理不足的问题, 将 $V$ 值进行深度卷积处理。然而, 该卷积分支无法有效地聚合通道和空间维度信息。所以引入了一个自适应交互模块, 从空间维度或信道维度自适应地调整这两个分支的特征, 使它们能够有效聚合。在自适应交互模块处理下, 这时的ACA有两个优点。首先, 该方法可以更好地集成局部和全局特征。与其他简单的特征融合方法不同, 该方法通过自适应交互调整局部和全局特征, 实现了更完美的特征融合。其次, 在其独特的耦合注意机制下, 它不仅提高了通道的交互能力, 而且提高了空间表征能力,

从而增强了块内通道与空间特征之间的交互作用。

如图2所示,对于给定的输入特征  $X \in \mathbf{R}^{H \times W \times C}$ ,使用线性投影生成查询( $Q$ )、键( $K$ )和值( $V$ )矩阵。该流程定义如下:  $Q = XW_Q$ ,  $K = XW_K$ ,  $V = XW_V$ , 其中  $W_Q$ 、 $W_K$ 、 $W_V \in \mathbf{R}^{H \times W \times C}$  为省略偏差的线性投影。

由于自注意的重点是捕获全局特征,因此引入了  $3 \times 3$  深度卷积和 ReLU 激活函数,对3个关键值( $K$ 、 $V$ 、 $Q$ )的信道空间进行编码,并将局部性引入注意机制中。应用额外的注意后,增加BN,进一步突出感兴趣的区域,获得更好的注意系数。最后,将从不同分支得到的注意系数与低水平特征图像逐像素相乘,以分配注意权重来定位裂纹区域。给定输入  $X$ , ACA 块被定义如下:  $A = \text{softmax}(QK^T / \sqrt{d} + B)V$ ,  $A_c = CI(A)$ ,  $A_s = SI(A)$ ,  $V_c = C-AIM(CA(Dconv(V)))$ ,  $V_s = S-AIM(SA(Dconv(V)))$ ,  $C_{map} = V_c \otimes A_c$ ,  $S_{map} = V_s \otimes A_s$ ,  $X_{out} = C_{map} \oplus S_{map}$ 。以上各式中:  $d$  为查询和键的维数;  $B$  为相对位置编码;  $CI(\cdot)$ 、 $SI(\cdot)$  分别为通道和空间交互;  $Dconv(\cdot)$  为深度卷积操作;  $CA(\cdot)$ 、 $SA(\cdot)$  分别为通道和空间注意机制;  $C-AIM(\cdot)$  和  $S-AIM(\cdot)$  分别为通道和空间自适应交互。

如图1b, CSAT 为 MFAM 网络中构建全局特征提取层的核心,由 ACA、前馈网络(feedforward network, FFN)和 LayerNorm 层(LN)组成。给定一个 CSAT 块的输入  $X_{in} \in \mathbf{R}^{H \times W \times C}$ , 该块被定义为

$$\begin{cases} X'_{in} = ACA(LN(X_{in})) + X'_{in}, \\ X_{out} = FFN(LN(X'_{in})) + X'_{in}. \end{cases}$$

式中:  $ACA(\cdot)$  为自适应耦合注意操作;  $LN(\cdot)$  为 LayerNorm 层;  $FFN(\cdot)$  为前馈网络。

### 2.3 高频增强模块

HFEM 的目的是增强高频信息。如图3所示,首先经过包含 ReLU 激活函数的两个标准卷积层。紧随其后的是基于通道注意的局部特征提取分支和基于空间注意的高频信息增强分支。具体来说,先将输入特征分成两部分,然后分别由两个分支进行处理。为增强模型捕获更多高频特性信息的能力,通道注意模块利用了特征通道间的相互依赖关系。它通过应用全局平均池化操作聚合特征映射,从而生成通道描述符。这些描述符封装了信道特征响应的全局分布,允许网络的较低层利用来自全局接受域的信息。经过两次卷积后,将执行压缩和恢复操作。每个通道的权值描述符通过激活函数得到。为了提取高频信息,引入空间注意模块,用高频空间信息补充通道注意。它降低了信道本身的维数,得到了最大池和平均池的结果。然

后将这些结果连接成一个特征图,该特征图通过一个卷积层捕获二维空间中的高频信息。最后将两分支输出连接并输入  $1 \times 1$  卷积层,以彻底融合信息。

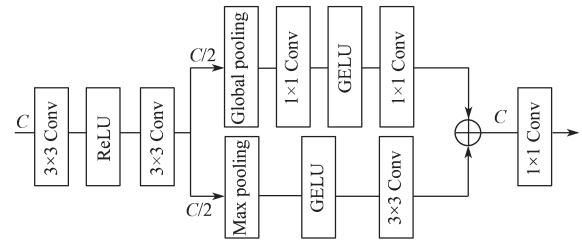


图3 高频增强模块

Fig. 3 High frequency enhancement module

### 2.4 细化融合模块

经过 CSAT 及高频增强模块后,由于多种注意机制的混合导致了特征重叠和信息冗余。因此,本文设计一个 RFM (见图4) 解决这一问题,其可融合不同层次的特征信息及细化从图像中捕获的特征。

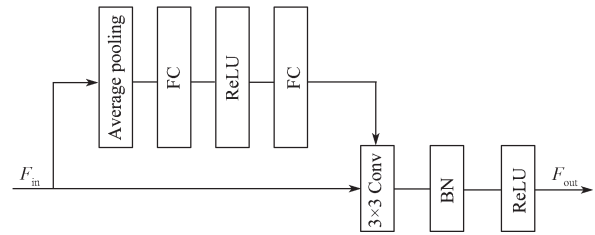


图4 细化融合模块

Fig. 4 Refinement fusion module

如图4所示, RFM 包含一个轻量级细化分支,该分支从输入处平均池化开始,然后是两个完全连接的 FC 层。在这些 FC 层间插入一个 ReLU 激活层。第一个 FC 层降低了信道维数,而第二个 FC 层却扩展了它。这种细化方法促进了图像中不同特征信息的融合,增强了模型的特征表达能力。接着将细化后的特征与原始特征通过一个卷积融合,再经过归一化层和 ReLU 激活层,得到细化融合后的特征。细化融合块采用细化调整,结合不同尺度的上下文信息,减少特征冗余的同时提高模型性能。

## 3 实验

### 3.1 实验设置

本实验中选用 DIV2K 为训练数据集,包括 800 张训练图像、100 张验证图像和 100 张测试图像。数据增强方法为旋转  $90^\circ$ 、 $180^\circ$ 、 $270^\circ$  和水平翻转。为实现测试目的,在 5 个 SR 基准数据集 (Set5、Set14、BSD100、Urban100 和 Manga109) 上将提出的 MFAN 模型与其他高级 SR 方法进行比较。以峰值信噪比 (peak signal to noise ratio, PSNR) 和结构

相似度 (structural similarity, SSIM) 评价重建图像质量。对每个训练批, 从 LR 图像中随机裁剪 16 个不同大小的补丁作为输入。使用 Adam 优化器, 通过设置  $\beta_1=0.9$ 、 $\beta_2=0.999$ 、 $\varepsilon=1E-8$  训练 MFAN, 共有 300 个训练时代。学习速率初始化为  $5E-4$ , 每 50 个周期减少一半。为保证模型性能和参数数量, 设定 MFAN, 包括 4 个多频特征聚合模型。实验部分使用 PyTorch 构建模型, 并在一个 RTX 3090 GPU 上进行训练。

### 3.2 与当前艺术水平的比较

为了直观地演示 MFAN 性能, 将其与其他先进的 SR 方法进行了比较, 包括 SRCNN<sup>[3]</sup>、极深的超分辨率 (very deep super-resolution, VDSR)<sup>[4]</sup>、

MemNet<sup>[12]</sup>、EDSR-baseline<sup>[13]</sup>、用于多重退化的单卷积超分辨率网络 (single convolutional super-resolution network for multiple degradations, SRMDNF)<sup>[14]</sup>、信息蒸馏网络 (information distillation network, IDN)<sup>[15]</sup>、级联剩余网络 (cascading residual network, CARN)、信息多蒸馏网络 (information multi-distillation network, IMDN)、像素注意力 (pixel attention, PAN)、高效超分辨率变压器 (efficient super-resolution transformer, ESRT)<sup>[7]</sup> 和蓝图可分残差网络 (blueprint separable residual network, BSRN)<sup>[16]</sup>。并且在表 1 和表 2 中提供了定量比较结果和可视化比较结果, 表 1 中加粗数据为最好的结果。

表 1 在 5 个基准数据集上对最先进的 SR 算法进行定量比较的结果

Table 1 Quantitative comparison results of state-of-the art SR algorithm on five benchmark datasets

method	scale	Set5		Set14		BSD100		Urban100		Manga109	
		PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
SRCNN	×2	36.66	0.954 2	32.45	0.906 7	31.36	0.887 9	29.50	0.894 6	35.60	0.966 3
VDSR	×2	37.53	0.958 7	33.03	0.912 4	31.90	0.896 0	30.76	0.914 0	37.22	0.975 0
MemNet	×2	37.78	0.959 7	33.28	0.914 2	32.08	0.897 8	31.31	0.919 5	37.72	0.974 0
EDSR-baseline	×2	38.11	0.960 1	33.52	0.919 5	32.22	0.900 3	32.23	0.925 2	38.03	0.979 5
SRMDNF	×2	37.79	0.960 0	33.32	0.915 0	32.05	0.898 1	31.33	0.920 1	38.05	0.975 2
IDN	×2	37.83	0.960 0	33.30	0.914 8	32.08	0.898 5	31.27	0.919 6	38.01	0.974 9
CARN	×2	37.76	0.959 0	33.52	0.916 6	32.09	0.897 8	31.92	0.925 6	38.12	0.975 6
IMDN	×2	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
PAN	×2	38.00	0.960 5	33.59	0.918 1	32.18	0.899 7	32.01	0.927 3	38.70	0.977 3
ESRT	×2	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
BSRN	×2	38.10	0.961 0	33.74	0.919 3	32.24	0.900 6	32.34	0.930 3	38.88	0.977 3
MFAN(ours)	×2	<b>38.11</b>	<b>0.961 2</b>	<b>33.78</b>	<b>0.920 2</b>	<b>32.28</b>	<b>0.901 1</b>	<b>32.62</b>	<b>0.932 1</b>	<b>39.16</b>	<b>0.978 2</b>
SRCNN	×3	32.75	0.909 0	29.30	0.821 5	28.41	0.786 3	26.24	0.798 9	30.48	0.911 7
VDSR	×3	33.66	0.921 3	29.77	0.831 4	28.82	0.797 6	27.14	0.827 9	32.01	0.934 0
MemNet	×3	34.09	0.924 8	30.00	0.835 0	28.96	0.800 1	27.56	0.837 6	32.51	0.936 9
EDSR-baseline	×3	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
SRMDNF	×3	34.12	0.925 4	30.04	0.838 2	28.97	0.802 5	27.57	0.839 8	33.00	0.940 3
IDN	×3	34.11	0.925 3	29.99	0.835 4	28.95	0.801 3	27.42	0.835 9	32.71	0.938 1
CARN	×3	34.29	0.925 5	30.29	0.840 7	29.06	0.903 4	28.06	0.849 3	33.50	0.939 2
IMDN	×3	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
PAN	×3	34.40	0.927 1	30.36	0.842 3	29.11	0.805 0	28.11	0.851 1	33.61	0.944 8
ESRT	×3	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
BSRN	×3	34.46	0.927 7	30.47	0.844 9	29.18	0.806 8	28.39	0.856 7	34.05	0.947 1
MFAN(ours)	×3	<b>34.52</b>	<b>0.928 1</b>	<b>30.49</b>	<b>0.845 1</b>	<b>29.19</b>	<b>0.807 3</b>	<b>28.46</b>	<b>0.857 6</b>	<b>34.07</b>	<b>0.948 1</b>
SRCNN	×4	30.48	0.862 6	27.50	0.751 3	26.90	0.710 1	24.52	0.722 1	27.58	0.855 5
VDSR	×4	31.35	0.883 8	28.01	0.767 4	27.29	0.725 1	25.18	0.752 4	28.83	0.887 0
MemNet	×4	31.74	0.889 3	28.26	0.772 3	27.40	0.728 1	25.50	0.763 0	29.42	0.894 2
EDSR-baseline	×4	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
SRMDNF	×4	31.96	0.892 5	28.35	0.778 7	27.49	0.733 7	25.68	0.773 1	30.09	0.902 4
IDN	×4	31.82	0.890 3	28.25	0.773 0	27.41	0.729 7	25.41	0.763 2	29.41	0.894 2
CARN	×4	32.13	0.893 7	28.60	0.780 6	27.58	0.734 9	26.07	0.783 7	30.47	0.907 9
IMDN	×4	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
PAN	×4	32.13	0.894 8	28.61	0.782 2	27.59	0.736 3	26.11	0.785 4	30.51	0.909 5
ESRT	×4	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
BSRN	×4	32.35	0.896 6	28.73	0.784 7	27.65	0.738 7	26.27	0.790 8	30.84	0.912 3
MFAN(ours)	×4	<b>32.42</b>	<b>0.900 1</b>	<b>28.82</b>	<b>0.786 3</b>	<b>27.73</b>	<b>0.741 2</b>	<b>26.46</b>	<b>0.796 1</b>	<b>30.91</b>	<b>0.914 2</b>

1) 定量结果。表 1 显示了 ×2、×3 和 ×4 下的 5 个基准数据集的图像 SR 结果。如表 1 所示, MFAN 在所有数据集上的表现都显著优于其他方法。这些比较结果表明, 与以往基于 CNN 和 Transformer





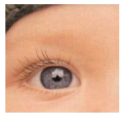


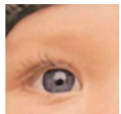





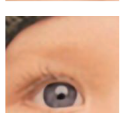
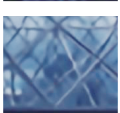

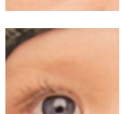


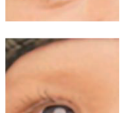





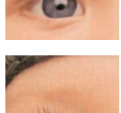

的方法相比, 充分整合局部细节和全局特征、充分融合多频特征, 可以有效地提高图像重建质量。

2) 视觉效果。表 2 为 MFAN 和其他方法间的可视化比较结果, 其原始图片 “img014” “baby” “223061”

分别来自 Urban100、Set5 和 BSD100。经放大处理，可看到 MFAN 重建的 SR 图像的条纹和线比其他方法的更接近原始 HR 图像的条纹和线。这种视觉比较可进一步证明 MFAN 的有效性。

表 2 MFAN 在 BSD100, Urban100 和 Set5 × 4 数据集上与其他 SR 方法的可视化比较结果

Table 2 Visual comparison results of MFAN with other SR methods on BSD100, Urban100, and Set5 × 4 datasets

方法	img014	baby	223061
			
HR			
VDSR			
IDN			
IMDN			
PAN			
ESRT			
BSRN			
MFAN			

总的来说，定量结果和主观视觉效果都表明，MFAN 优于其他已有先进的超分辨率模型。

### 3.3 消融研究

对 MFAN 中每个模块进行消融实验，以验证其有效性。为了科学有效地进行比较，所有模型都有一致的基本架构和参数设置，用基准数据集 DIV2K 训练 MFAN 模型，并对两个数据集进行 × 4SR 测试：

BSD100 和 Manga100，且进行 300 次迭代，结果见表 3。

表 3 使用 × 4SR 的 3 个基准数据集上 DBFA 中不同基本块的 PSNR/SSIM 比较结果

Table 3 PSNR/SSIM comparison results of different basic blocks in the MFAN on the 2 benchmark datasets with SR

scale	different components			BSD100		Manga109	
	CSAT	HFEM	RFM	PSNR/dB	SSIM	PSNR/dB	SSIM
× 4	✓	✓	✓	27.73	0.741 2	30.91	0.914 2
	✓	✓		27.62	0.738 9	30.79	0.910 2
	✓		✓	27.67	0.739 5	30.83	0.911 9
		✓	✓	27.60	0.738 7	30.46	0.908 6

由表 3 可知，MFAN 通过充分提取高频和低频信息，进行细化融合特征处理，提高了图像重建性能。为评估 CSAT 的有效性，将其替换为普通 Transformer。数据显示，使用 CSAT 比传统 Transformer 具更好的 PSNR 和 SSIM 值，表明 CSAT 在重建图像方面更有效。接着将 HFEM 模块替换为具激活函数的 3 × 3 卷积层，导致模型性能显著下降。这表明 HFEM 模块可捕获不同频率特征，并增强高频特征的提取。最后，RFM 模块通过融合高频信息与全局特征并细化融合特征。实验结果表明，去掉 RFM 模块后，网络性能明显下降，说明 RFM 对多频率特征的融合有良好效果。通过实验结果可得出，在替换或删除本模型结构后，性能都得到了大幅削弱，因此证明了模型的有效性。

### 3.4 复杂性分析

在图像超分辨率研究领域中，除了优化模型的性能外，还应考虑模型的复杂度和计算量。因此，实验还进一步展示了 MFAN 与其它方法在计算参数与性能间的关系。图 5 显示了 Set5 上的 PSNR 与不同模型参数数量的比较结果。比较模型包括 VDSR、DRCN、SRMDNF、CARN、ESRT 和 MFAN。

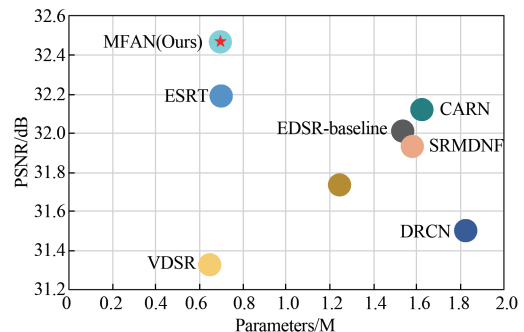


图 5 在 Set5 数据集上与其他网络的性能和模型复杂度比较结果

Fig. 5 Comparison results of performance and model complexity with other networks on the Set5 dataset

如图 5 所示，与 VDSR 和 ESRT 方法相比，MFAN 在相同参数的情况下提高了网络性能。与其他

方法相比, MFAN 模型的参数在远小于它们的情况下取得了更显著的 SR 性能。这一结果表明, MFAN 在模型的复杂性和性能间取得了更好的平衡。

## 4 结语

本文提出了一种细化冗余特征并融合高频信息的 MFAN 模型, 用于图像的全局远程依赖建模。具体来说, 本研究使用耦合自注意 Transformer 为基本模块, 以增强块内通道与空间特征的交互, 激活更多像素, 提高图像超分辨率效果。此外, 设计了用于提取高频信息的高频增强模块, 即通过在 Transformer 模块旁并联一个高频增强模块, 以弥补 Transformer 无法提取高频特征的缺点, 突出更重要的信息。然后使用 RFM 融合不同层次特征信息, 细化从图像中捕获到的特征, 以便更好地聚合多频特征。实验结果表明, 所提模型比之前的方法具有更好的性能。

然而与基于 CNN 方法相比, 仍有相对较高的计算复杂度。因此, 未来的工作将集中于提高多频特征聚合模型的计算速度, 并尽可能降低计算复杂度。

### 参考文献:

- [1] 柏正尧, 陶劲宇. 采用伪 3D 卷积网络的脑部 MRI 图像超分辨率重建 [J]. 计算机辅助设计与图形学学报, 2022, 34(2): 208-216.  
BAI Zhengyao, TAO Jinyu. Super-Resolution Reconstruction of Brain MR Images Using Pseudo-3D Convolutional Network[J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(2): 208-216.
- [2] 王博琛. 基于稀疏字典学习的遥感图像超分辨率重建研究 [D]. 西安: 西北大学, 2019.  
WANG Bochen. Research on Super-Resolution Reconstruction of Remote Sensing Images Based on Sparse Dictionary Learning[D]. Xi'an: Northwest University, 2019.
- [3] DONG C, LOY C C, HE K M, et al. Learning a Deep Convolutional Network for Image Super-Resolution[C]//Computer Vision-ECCV 2014: 13th European Conference. Zurich: Springer International Publishing, 2014: 184-199.
- [4] 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 (CVPR). Las Vegas: IEEE, 2016: 1646-1654.
- [5] LI F, BAI H H, ZHAO Y. Detail-Preserving Image Super-Resolution via Recursively Dilated Residual Network[J]. Neurocomputing, 2019, 358, 285-293.
- [6] HUI Z, LI J, GAO X B, et al. Progressive Perception-Oriented Network for Single Image Super-Resolution[J]. Information Sciences, 2021, 546: 769-786.
- [7] LU Z S, LI J C, LIU H, et al. Transformer for Single Image Super-Resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 457-466.
- [8] CHEN H T, WANG Y H, GUO T Y, et al. Pre-Trained Image Processing Transformer[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 12299-12310.
- [9] YANG F Z, YANG H, FU J L, et al. Learning Texture Transformer Network for Image Super-Resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 5791-5800.
- [10] WANG Z D, CUN X D, BAO J M, et al. Uformer: A General U-Shaped Transformer for Image Restoration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 17683-17693.
- [11] 陈纯毅, 吴欣怡, 胡小娟, 等. 多尺度注意力融合的图像超分辨率重建 [J]. 中国光学, 2023, 16(5): 1034-1044.  
CHEN Chunyi, WU Xinyi, HU Xiaojuan, et al. Image Super-Resolution Reconstruction with Multi-Scale Attention Fusion[J]. Chinese Optics, 2023, 16(5): 1034-1044.
- [12] TAI Y, YANG J, LIU X M, et al. MemNet: A Persistent Memory Network for Image Restoration[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 4539-4547.
- [13] LIM B, SON S, KIM H, et al. Enhanced Deep Residual Networks for Single Image Super-Resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu: IEEE, 2017: 136-144.
- [14] ZHANG K, ZUO W M, ZHANG L. Learning a Single Convolutional Super-Resolution Network for Multiple Degradations[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 3262-3271.
- [15] HUI Z, WANG X M, GAO X B. Fast and Accurate Single Image Super-Resolution via Information Distillation Network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 723-731.
- [16] LI Z Y, LIU Y Q, CHEN X Y, et al. Blueprint Separable Residual Network for Efficient Image Super-Resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 833-843.

(责任编辑: 廖友媛)