液晶与显示
Chinese Journal of Liquid Crystals and Displays

文章编号:1007-2780(XXXX)XX-0001-10

基于小波变换与Transformer的图像超分辨率增强算法研究

黄佳怡[1],景明利[1,2*],焦　龙[3],章　成[1]
(1.西安石油大学 电子工程学院,陕西 西安 710065;
2.西安石油大学 计算机学院,陕西 西安 710065;
3.西安石油大学 化学化工学院,陕西 西安 710065)

摘要:针对现有Transformer的超分辨率方法存在的局部纹理和高频细节建模能力不足的问题,设计了一种基于小波变换与Transformer的高效混合网络W-EMT,在Transformer框架中引入离散Haar小波变换,对输入特征进行多级频域分解,实现多尺度频域建模。在局部特征建模过程中,构建局部小波混合器,将特征分解为不同频率子带,并对低频结构信息和高频纹理信息进行差异化增强,再通过逆小波变换恢复特征。在全局特征建模过程中,引入条纹窗口自注意力机制,通过沿水平和垂直方向建立长距离依赖关系,在降低计算复杂度的同时保持全局上下文信息建模能力,实现局部与全局特征的协同学习。实验采用DIV2K和Flickr2K数据集进行训练,并在Set5、Set14、BSD100、Urban100和Manga109数据集上进行测试,以峰值信噪比(PSNR)和结构相似性(SSIM)作为评价指标。实验结果表明,W-EMT在各测试数据集上均取得稳定性能提升,在×4放大任务中,相较EMT在Urban100数据集上的PSNR提升0.06 dB、SSIM提升0.0011,在复杂纹理场景中表现出更优的高频细节重建能力,同时在其他数据集上保持一致性增益,验证了该方法在图像超分辨率重建中的有效性。

关　键　词:图像超分辨率;小波变换;Transformer;多尺度特征;注意力机制
中图分类号:TP391.4　　**文献标识码**:A　　**doi**:10.3724/CJLCD.2026-0128　　**CSTR**:32172.14.CJLCD.2026-0128

Research on an image super-resolution algorithm based on wavelet and transformer

HUANG Jiayi[1], JING Mingli[1,2*], JIAO Long[3], ZHANG Cheng[1]

(1. *School of Electronic Engineering, Xi'an Shiyou University, Xi'an* 710065, *China*;
2. *School of computer Science, Xi'an Shiyou University, Xi'an* 710065, *China*;
3. *Chemistry and Chemical Engineering, Xi'an Shiyou University, Xi'an* 710065, *China*)

Abstract: To address the limitations of existing Transformer-based image super-resolution methods in modeling local textures and high-frequency details, an Efficient Mixed Transformer with Wavelet (W-EMT) is proposed by integrating wavelet transform into the Transformer architecture. The discrete Haar wavelet transform is introduced into the Transformer framework to perform multi-level frequency

收稿日期:2026-07-06;**修订日期**:2026-10-19.
基金项目:国家自然科学基金(No.22373075)
Supported by National Natural Science Foundation of China (No.22373075)
*通信联系人,E-mail:mljingsy@xsyu.edu.cn

decomposition of input features, enabling multi-scale frequency-domain representation learning. In local feature modeling, a local wavelet mixer is constructed to decompose features into different frequency subbands, where low-frequency structural information and high-frequency texture information are enhanced using a differentiated strategy, followed by feature reconstruction via the inverse wavelet transform. In global feature modeling, a stripe-window self-attention mechanism is introduced to capture long-range dependencies along the horizontal and vertical directions. The proposed mechanism maintains global contextual modeling capability while reducing computational complexity and enables collaborative learning of local and global features. The model is trained on the DIV2K and Flickr2K datasets and evaluated on the Set5, Set14, BSD100, Urban100, and Manga109 benchmark datasets using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) as evaluation metrics. Experimental results demonstrate that W-EMT achieves consistent performance improvements across all test datasets. For ×4 image super-resolution, W-EMT improves PSNR by 0.06 dB and SSIM by 0.0011 on the Urban100 dataset compared with EMT, demonstrating superior high-frequency detail reconstruction in complex texture regions and verifying the effectiveness of the proposed method.

Key words: image super-resolution; wavelet transform; transformer; multi-scale features; attention mechanism

1 引　　言

图像质量与空间分辨率的提升是成像研究的重要目标,相关探索包括神经阵列超构成像[1]和多焦点超透镜辅助的显微超分辨成像[2]。单幅图像超分辨率(Single Image Super-Resolution, SISR)旨在从低分辨率图像中重建高分辨率图像,以恢复更多纹理细节和边缘信息,增强其视觉效果,广泛应用于医疗影像分析、遥感图像处理及视频监控等领域[3]。

图像超分辨率重建传统方法主要分为三类:插值、重建、学习[4]。插值方法通过已知像素估计未知像素实现图像放大,典型方法包括最近邻插值、双线性插值和双三次插值。其中,双三次插值因能够较好保持图像边缘信息而被广泛应用。

重建方法将超分辨率重建视为逆问题,通过引入边缘、梯度或稀疏性等先验约束恢复高分辨率图像,代表方法包括最大后验概率估计和迭代反向投影等。然而,该类方法操作方便、重建速度较快,但是所引入的先验知识有限,在重建后会丢失很多图像细节,效果并不理想[5]。

学习方法通过训练样本建立低分辨率与高分辨率图像之间的映射关系,典型方法包括邻域嵌入和稀疏表示[6]。相比传统方法,其能够恢复更多高频细节,但性能受特征表示能力和训练数据规模的影响。

神经网络已用于光学系统参数反演[7]及光学图像处理[8]等研究,也为图像超分辨率重建提供了新的解决思路。基于卷积神经网络(Convolutional Neural Network, CNN)的方法通过学习低分辨率与高分辨率图像间的复杂非线性映射关系提升图像重建质量[9]。超分辨率卷积神经网络(Super-Resolution Convolutional Neural Network, SRCNN)[10]首次将这一思路付诸实践,通过三层卷积模拟稀疏编码流程,在多个基准集上提高了模型性能。超深层超分辨率网络(Very Deep Super-Resolution, VDSR)[11]加深至20层并引入残差学习;增强型深度超分辨率网络(Enhanced Deep Residual Networks, EDSR)[12]去除批归一化以降低显存,支持更宽更深结构;超分辨率生成对抗网络(Super-Resolution Generative Adversarial Network, SRGAN)[13]借助生成对抗网络与感知损失提升感知质量;残差通道注意力网络(Very Deep Residual Channel Attention Network, RCAN)[14]则通过残差-残差结构与通道注意力机制,在超400层的深度下增强高频细节重建。然而,CNN方法仍受限于局部感受野,难以建模长距离依赖关系。

近年来,Transformer及其混合架构已应用于缺陷检测[15]和目标跟踪[16-17]等视觉任务,并借

助自注意力机制建模超分辨率重建中的长距离依赖。Image Restoration Using Swin Transformer（SwinIR）[18]利用移位窗口自注意力在线性复杂度下实现跨窗口信息交互，在多项任务中取得领先；Efficient Super-Resolution Transformer（ESRT）[19]通过高效多尺度注意力降低开销；N-Gram Swin Transformer（NGswin）[20]融合N-gram思想增强局部建模，Efficient Mixed Transformer（EMT）[21]混合自注意力机制与像素混合器以建模全局和局部特征，Large Kernel Modulation Network（LKMN）[22]利用大核卷积和跨门控调制机制增强特征表征能力，Multi-scale Mixed Transformer（MMT）[23]结合多尺度特征融合与高效混合注意力机制降低计算复杂度。Selective Aggregation Transformer（SAT）[24]通过密度驱动的Token聚合实现高效全局建模。Multi-dimensional Aggregation Transformer（MDAT）[25]通过多尺度调制、空间-通道交互及特征重用增强特征表达。尽管上述方法不断增强特征建模能力，如何进一步兼顾长距离依赖建模与局部高频细节恢复，仍值得研究。

高低频协同有助于结构保持与细节恢复。多聚焦图像融合方法CNN and Transformer Cross Fuse（CTCFuse）[26]结合CNN与Transformer提取高低频特征，以低频信息引导高频细节融合；小波变换则通过多分辨率分解显式分离高低频信息，支持图像结构与细节建模[27]。Deep Wavelet Super-Resolution（DWSR）[28]利用卷积神经网络预测高分辨率图像的小波系数，实现了频域特征重建。Wavelet-based Residual Attention Network（WRAN）[29]利用小波变换实现高低频信息分离，并分别进行特征学习，以降低模型训练难度。wavelet-based dual recursive network（WSR）[30]基于小波变换的稀疏表示特性，在小波域内设计对偶递归网络学习不同频率子带特征。然而，现有小波变换相关方法在高频细节增强与全局建模协同方面仍存在不足，难以同时兼顾局部纹理恢复与长距离依赖建模能力。

针对上述问题，本文提出一种基于小波变换与Transformer的高效混合网络（Efficient Mixed Transformer with Wavelet，W-EMT）。该方法将小波变换的多尺度分解特性引入Transformer，采用小波混合器（Wavelet Mixer）和条纹窗口自注意力，对图像的高低频特征分别进行处理，实现高效的超分辨率重建。

2 基于小波变换与Transformer的高效混合网络

2.1 网络整体框架

高效混合网络包含三个模块：浅层特征提取模块、深层频率感知特征提取模块以及图像重建模块，网络框架如图1所示。

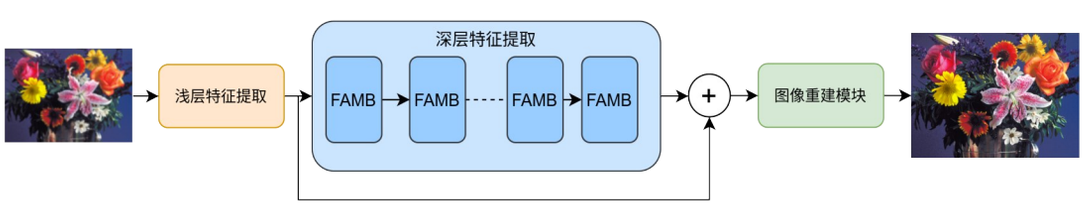


图1 W-EMT网络的整体框架示意图

Fig. 1 Overall architecture of the W-EMT network

设输入低分辨率图像为I_{LR}，输出超分辨率图像为I_{SR}。浅层特征提取过程可表示为：

$$F_0 = H_{SF}(I_{LR}), \tag{1}$$

其中，$H_{SF}(\cdot)$表示浅层特征提取函数，F_0为输入图像经初步映射后得到的浅层特征表示。

深层频率感知特征提取模块由N个频率感知混合块（Frequency-Aware Mixed Block，FAMB）堆叠而成：

$$F_D = H_{DF}(F_0) = H_{B^N}(H_{B^{N-1}}(\cdots H_{B^1}(F_0)\cdots)), \tag{2}$$

其中，$H_{DF}(\cdot)$表示深层特征增强映射，$H_B^i(\cdot)$表示第i个频率感知混合块。

浅层特征与深层特征进行逐元素相加得到融合特征，再经图像重建模块生成高分辨率图像：

$$F_R = F_0 + F_D, \tag{3}$$

$$I_{SR}=H_{REC}(F_R), \tag{4}$$

其中，$H_{REC}(\cdot)$表示图像重建函数。

2.2 频率感知混合块

FAMB是本文网络的核心组成单元，其主要作用是在保持较低计算复杂度的同时，实现全局结构关系建模与局部纹理细节增强。如图2所示，每个FAMB包含L个处理层，由两种不同类型的特征混合器构成：全局频谱学习器（Global Spectrum Learner，GSL）和局部小波混合器（Local Wavelet Mixer，LWM）。

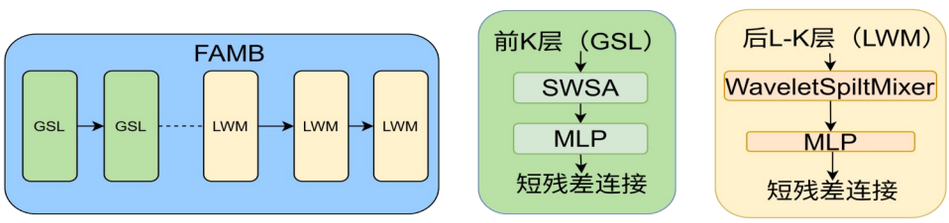


图2 频率感知混合块内部结构

Fig. 2 Internal structure of a frequency-aware mixer block

前K层采用基于窗口的移位自注意力机制作为GSL，用于捕获跨窗口的全局频谱依赖关系；后L－K层采用小波分解混合器作为LWM，用于在频率域内精细化处理局部特征。两种特征混合器均采用残差连接的形式：

$$X_{i+1}=X_i+M_i(X_i), \tag{5}$$

其中，$M_i()$表示第i层的特征混合器，X_i为当前层的输入特征。

在FAMB中，自注意力层与局部小波混合器依次作用于同一特征流。前者通过窗口内的特征交互建立跨区域联系；后者对输入特征进行离散Haar小波分解，分别处理低频结构与高频细节，再经逆小波变换融合为空间域特征，传递至后续层。

2.3 局部小波混合器

局部小波混合器利用离散Haar小波变换（Discrete Haar Wavelet Transform，DWT）对输入特征进行多尺度频域分解，将其划分为低频子带和多个高频子带，分别采用差异化增强策略对低频结构信息和高频细节信息进行特征学习，最后通过逆离散Haar小波变换（Inverse Discrete Haar Wavelet Transform，IDWT）将增强后的各频率子带重新融合，恢复到原始空间分辨率，实现频域信息与空间信息的有效交互，为后续特征提取提供更加丰富的局部表示。其中，介于整体结构与细微纹理之间的中频信息可由不同尺度的细节特征表示，并在子带处理与融合过程中参与图像重建。局部小波混合器结构如图3所示。

2.4 全局频谱学习器

图像中的目标轮廓、纹理延展和方向性结构通常具有明显的各向异性特点，仅采用局部方形窗口进行注意力计算，往往难以充分捕获沿水平或垂直方向延伸的长距离相关信息，且会增加一

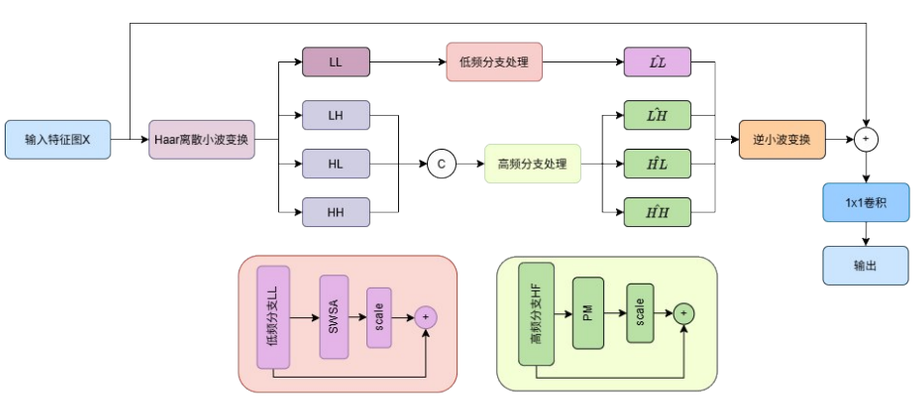


图3 局部小波混合器结构示意图

Fig. 3 Schematic diagram of a local wavelet mixer

定的计算冗余。因此，为增强长距离依赖关系的建模能力，引入条纹窗口自注意力机制，其结构如图4所示，通过分别沿水平和垂直方向设计条状窗口，并在不同窗口间进行跨区域特征交互，从而在有效降低计算复杂度的同时更有针对性地增强模型对方向性结构的建模能力。

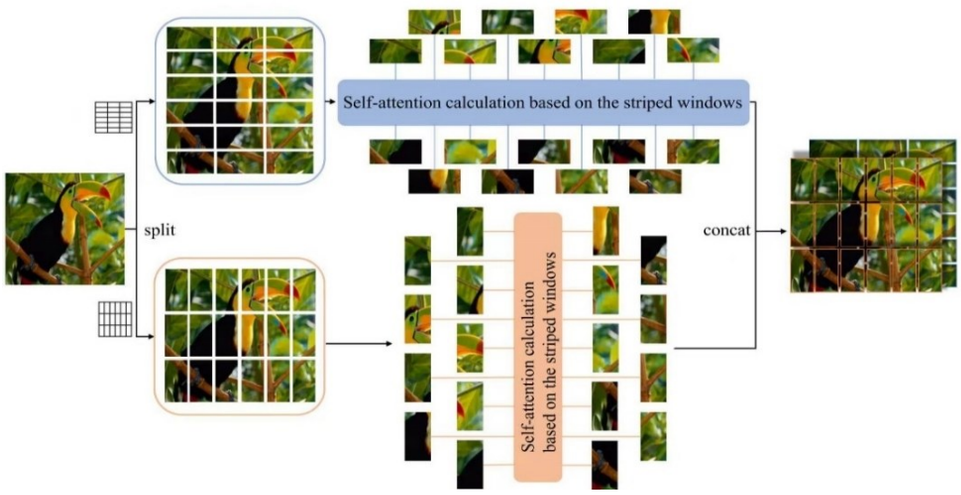


图4 条纹窗口自注意力机制

Fig. 4 Striped window self-attention mechanism

设输入特征为$X\in\mathbb{R}^{C\times H\times W}$，将其按水平方向和垂直方向对特征图进行划分，并在不同方向的条纹窗口内分别进行注意力计算。对第n个方向窗口中的特征X_n'执行线性映射，可得查询矩阵与数值矩阵：

$$Q_n=X_n'W_Q,\ V_n=X_n'W_V\ , \tag{6}$$

其中，W_Q和W_V分别表示对应的线性投影矩阵。通过缩放点积计算注意力响应：

$$F_n=\mathrm{Softmax}\left(\frac{Q_nQ_n^T}{\sqrt{d}}\right)V_n\ , \tag{7}$$

其中，d为缩放因子。最后将不同方向窗口的结果进行拼接，得到全局建模输出：

$$F=[F_1;\ F_2;\ \cdots;\ F_n]\ . \tag{8}$$

2.5 损失函数

图像恢复领域常用的损失函数有L1、L2两种损失函数，与L_2损失相比，L_1损失在图像恢复任务中通常更有利于缓解过度平滑现象，并能够获得更加稳定的边缘与纹理恢复效果。因此采用L_1损失函数对模型训练进行监督。设训练样本集为$\{I_{LR}^i,I_{HR}^i\}_{i=1}^N$，其中$I_{HR}^i$表示与低分辨率图像对应的高分辨率真实图像，$f(\cdot)$表示网络前向映射函数。则损失函数定义为：

$$\mathcal{L}=\frac{1}{N}\sum_{i=1}^{N}\left\| f\left(I_{LR}^i\right)-I_{HR}^i\right\|_1\ . \tag{9}$$

3 实验结果与分析

3.1 实验环境

本文采用DF2K数据集进行训练，并在Set5、Set14、BSD100、Urban100和Manga109五个公开基准数据集上进行测试与评估。在性能评价方面，本文采用峰值信噪比（Peak Signal-to-Noise Ratio，PSNR）和结构相似性（Structural Similarity Index Measure，SSIM）作为主要评价指标，并在YCbCr颜色空间的Y通道上进行计算。

本文基于PyTorch实现，以EMT为基础框架，在局部特征混合层中引入小波增强模块。主要超参数设置如下：特征维度60，网络块数6，块内层数6，注意力头数3，FAMB中的L、K分别为6和2，Batch Size设置为8，优化器采用Adam，学习率4×10^{-4}，损失函数采用L_1损失进行监督训练。

3.2 定量结果分析

为全面评估所提方法在图像超分辨率重建任务中的有效性，本文将所提出的W-EMT模型与多种代表性的超分辨率方法在×4放大倍率下进行定量比较。实验采用PSNR和SSIM作为评价指标，其中最优结果和次优结果分别以加粗和下划线标注，具体结果见表1。

从表中可以看出，W-EMT在×4放大倍率

表1 不同方法在×4超分辨率任务上的PSNR/SSIM对比

Tab. 1 Comparison of PSNR/SSIM for different methods on ×4 super-resolution tasks

Method	Scale	Set5		Set14		BSD100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
LAPAP-A[31]	x4	32.15	0.8944	28.61	0.7818	27.61	0.7366	26.14	0.7871	30.42	0.9074
EDSR[12]		32.42	0.8968	28.80	0.7876	27.71	0.7420	26.64	0.8033	31.02	0.9148
NGswin[20]		32.33	0.8963	28.78	0.7859	27.66	0.7396	26.45	0.7963	30.80	0.9128
ESRT[19]		32.19	0.8947	28.69	0.7833	27.69	0.7379	26.39	0.7962	30.75	0.9100
SwinIR[18]		32.44	0.8976	28.77	0.7858	27.69	0.7406	26.47	0.7980	30.92	0.9151
LKMN[22]		32.40	0.8976	28.78	0.7853	27.70	0.7399	26.35	0.7918	30.97	0.9126
SRFormer[32]		32.51	0.8988	28.82	0.7872	27.73	0.7422	26.67	0.8032	31.17	0.9165
SP-MoMamba[33]		32.35	0.8970	28.77	0.7850	27.69	0.7398	26.40	0.7939	31.01	0.9160
EMT[21]		32.57	**0.8998**	28.89	**0.7889**	27.77	0.7433	26.76	0.8063	31.34	0.9176
W-EMT		**32.59**	0.8997	**28.91**	**0.7889**	**27.78**	**0.7438**	**26.82**	**0.8074**	**31.36**	**0.9178**

下表现出具有竞争力的性能，相比于主流轻量级超分方法在PSNR和SSIM指标上均有明显提升，展现了本文网络在处理不同类型图像内容时具有较好的泛化能力。且本模型在Urban100等数据集上表现更为出色。这验证了本文提出的频率分解机制在捕获多尺度频域特征和增强局部纹理信息方面的有效性。

3.3 定性结果分析

图5展示了不同方法在多个测试集上的可视化结果。相比对比方法，W-EMT通过多频信息协同建模，在保持整体结构一致性的同时增强了高频细节表达能力，能够恢复更加清晰的边缘结

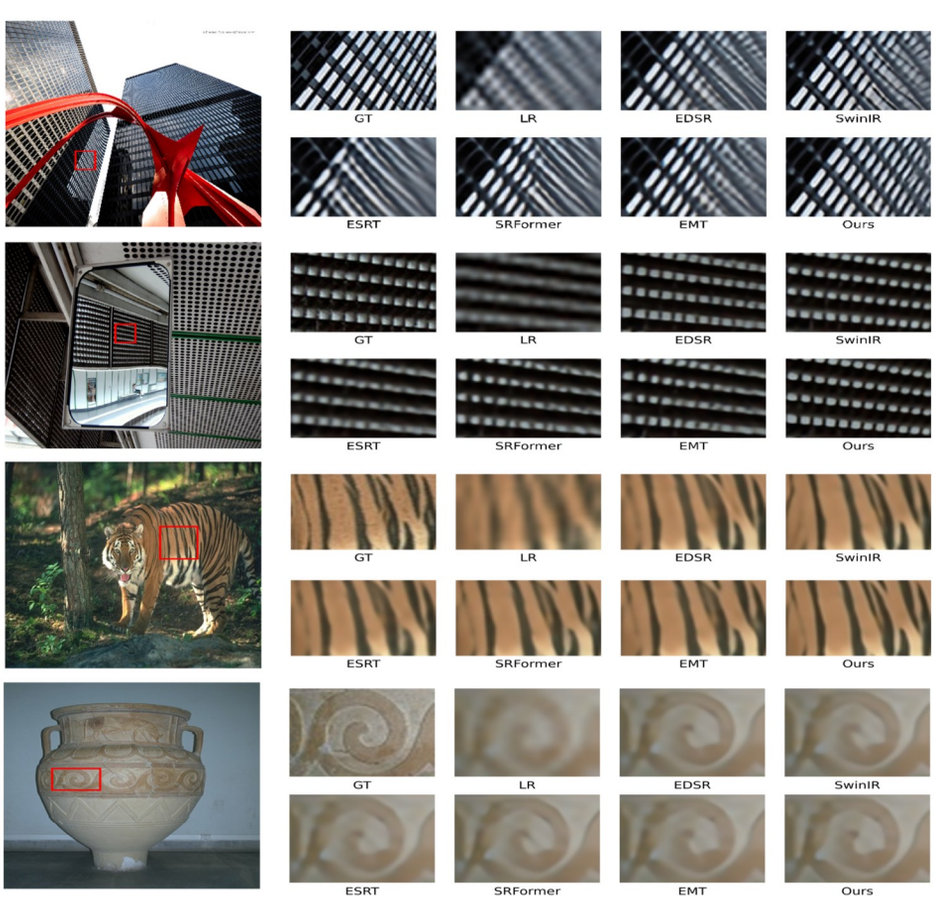


图5 本文与其它方法在Urban和BSD100数据集的可视化对比

Fig. 5 Visual comparison of this method with other methods on the Urban and BSD100 datasets

构和纹理细节，尤其在复杂纹理区域表现出更强的重建能力。

3.4 消融实验

3.4.1 实验设置

为验证局部小波混合器中低频分支与高频分支的有效性，本文在×4超分任务上设计了五种不同配置进行消融实验。Baseline配置未启用低频和高频分支，保留原有GSL数量；HF仅启用高频分支增强局部纹理细节，并用LWM替换4层GSL；LF仅启用低频分支强化低频结构信息，也替换4层GSL；Dual同时启用低频和高频分支，仅替换2层GSL；Full为完整模型，低频和高频分支均启用，共替换4层GSL。

3.4.2 消融结果分析

表2给出了不同配置下的定量实验结果对比。从表中可以看出，引入低频或高频分支均能带来一定性能提升，说明频域信息分解对于提升特征表达能力具有积极作用。而启用双分支结构可获得较好结果，说明低频结构建模与高频细节增强之间具有良好的互补关系。当采用完整结构（Full）时，在Set14、Urban100及Manga109等数据集上取得最佳结果，表现出更强的细节恢复能力与结构重建能力。该结果表明，多频信息协同建模能够有效提升网络对不同频率特征的表达与融合能力，从而在整体上进一步增强图像超分辨率重建性能。

表2 不同配置在×4超分辨率任务上的PSNR/SSIM结果

Tab. 2 PSNR/SSIM results for different configurations on the ×4 super-resolution task

模型配置	Set5		Set14		BSD100		Urban100		Manga109	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Baseline	32.57	**0.8998**	28.89	**0.7889**	27.77	0.7433	26.76	0.8063	31.34	0.9176
HF	32.55	0..8995	28.87	0.7883	27.75	0.7432	26.75	0.8059	31.34	0.9175
LF	32.57	0.8996	28.88	0.7885	27.76	0.7434	26.78	0.8070	31.36	0..9177
Dual	32.56	0.89965	28.87	0.7884	27.76	0.7433	26.75	0.8069	31.35	0.9176
Full	**32.59**	0.8997	**28.91**	**0.7889**	**27.78**	**0.7438**	**26.82**	**0.8074**	**31.36**	**0.9178**

图6展示了不同消融配置的重建视觉效果对比。从整体结果可以看到，仅使用单一分支时，模型在不同程度上存在重建不足的问题。相比之下，同时引入低频与高频分支能够实现多频信息的协同建模，在保持图像整体结构完整性的同时，有效增强纹理细节与边缘清晰度，使重建结

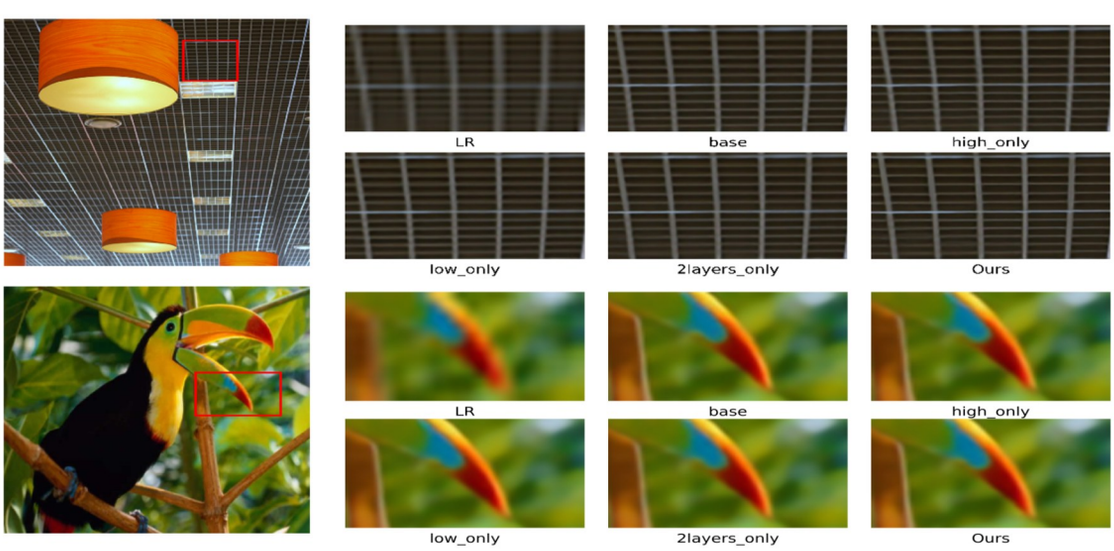


图6 各消融结果在Urban100和set5数据集上的可视化对比

Fig. 6 Visual comparison of ablation results on the Urban100 and set5 datasets

果在主观视觉上更加自然。

4 结 论

针对现有Transformer方法存在的局部纹理和高频细节建模能力不足的问题。本文提出一种基于小波变换与Transformer的高效混合网络W-EMT。通过构建局部小波混合器和全局频谱学习器,实现了频域特征增强与长距离依赖建模。实验结果表明,W-EMT在Set5、Set14、BSD100、Urban100和Manga109等公开数据集上均取得了稳定性能提升,其中在×4放大任务中,Urban100数据集上的PSNR达到26.82 dB,SSIM达到0.8074,相较于EMT均有一定提升,说明该方法能够有效增强复杂纹理和高频细节的重建能力。虽然W-EMT能恢复更多细节信息,但在小波分解中采用最简单的基函数,对复杂结构或真实退化图像的自适应表达能力有限,且小波混合器在多频带分支处理与融合过程中仍引入一定计算冗余。未来可进一步研究轻量化网络设计、可学习小波分解以及真实场景超分辨率任务,并探索该方法在视频超分辨率、图像去噪和图像去模糊等视觉任务中的应用潜力。

参 考 文 献:

[1] DUN X, ZHANG J, CHEN F, *et al*. Neural array meta-imaging[J]. *eLight*, 2025, 5: 33.

[2] REN W, LIANG Q X, XI P. Multifocal metalenses drive super-resolution imaging of brain organoids[J]. *Light: Advanced Manufacturing*, 2025, 6: 71.

[3] WU G, JIANG J, JIANG J, *et al*. Transforming image super-resolution: a convformer-based efficient approach[J]. *IEEE Transactions on Image Processing*, 2024, 33: 6071-6082.

[4] 郑龙光,朱波,李国梁,等. 基于高频融合注意力的轻量级图像超分辨率重建[J]. 液晶与显示,2026,41(2):280-293.
ZHENG G L, ZHU B, LI G L, *et al*. Lightweight image super-resolution reconstruction based on high-frequency fusion attention [J]. *Chinese Journal of Liquid Crystals and Displays*, 2026,41(2):280-293. (in Chinese)

[5] 刘郭琦,刘进锋,朱东辉. 基于生成对抗网络的图像超分辨率重建算法[J]. 液晶与显示,2021,36(12):1720-1727.
LIU G Q, LIU J F, ZHU D H. Image super-resolution reconstruction algorithm based on generative adversarial networks [J]. *Chinese Journal of Liquid Crystals and Displays*, 2021,36(12):1720-1727. (in Chinese)

[6] LIU Y, XIONG Z, YUAN Y, *et al*. Distilling knowledge from super-resolution for efficient remote sensing salient object detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-16.

[7] LIU S Y, ZHONG H T, YU X C, *et al*. Deciphering optical coupled resonant systems with physics-data co-driven deep neural networks[J]. *Light: Science & Applications*, 2026, 15: 279.

[8] DONG Y, BAI Y, ZHANG Q, *et al*. High-throughput optical neuromorphic graphic processing at millions of images per second[J]. *eLight*, 2025, 5: 29.

[9] MEI Y, FAN Y, ZHOU Y. Image super-resolution with non-local sparse attention[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021: 3517-3526.

[10] DONG C, LOY C C, HE K, *et al*. Learning a deep convolutional network for image super-resolution[C]// European conference on computer vision. Cham: Springer International Publishing, 2014: 184-199.

[11] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C]// Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 1646-1654.

[12] 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, 2017: 136-144.

[13] LEDIG C, THEIS L, HUSZAR F, *et al*. Photo-realistic single image super-resolution using a generative adversarial network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2017:

4681-4690.

[14] ZHANG Y, LI K, LI K, *et al*. Image super-resolution using very deep residual channel attention networks[C]// Proceedings of the European conference on computer vision (ECCV), 2018: 286-301.

[15] JEONG H, YOON C H, LIM H S, *et al*. MT-former: multi-task hybrid transformer and deep support vector data description to detect novel anomalies during semiconductor manufacturing[J]. *Light: Advanced Manufacturing*, 2025, 6: 32.

[16] 刘时, 朱明. 基于Swin-Transformer改进的目标跟踪算法[J]. 液晶与显示, 2024,39(11):1569-1580.
LIU S, ZHU M. Improved target tracking algorithm based on Swin-Transformer[J]. *Chinese Journal of Liquid Crystals and Displays*, 2024, 39(11): 1569-1580. (in Chinese)

[17] 王鑫, 陈志旺, 卫燕侨, 等. 基于Transformer与跟踪轨迹的目标跟踪算法[J]. 液晶与显示, 2025,40(11):1710-1728.
WANG X, CHEN Z W, WEI Y Q, *et al*. Object tracking algorithm based on Transformer and tracking trajectory [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(11): 1710-1728. (in Chinese)

[18] LIANG J, CAO J, SUN G, *et al*. Swinir: Image restoration using swin transformer[C]//Proceedings of the IEEE/CVF international conference on computer vision, 2021: 1833-1844.

[19] LU Z, LI J, LIU H, *et al*. Transformer for single image super-resolution[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022: 457-466.

[20] CHOI H, LEE J, YANG J. N-gram in swin transformers for efficient lightweight image super-resolution[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023: 2071-2081.

[21] ZHENG L, ZHU J, SHI J, *et al*. Efficient mixed transformer for single image super-resolution[J]. *Engineering Applications of Artificial Intelligence*, 2024, 133: 108035.

[22] HU Q, TANG Y, ZHANG X. Large Kernel Modulation Network for Efficient Image Super-Resolution[J]. *arXiv preprint* arXiv:2508.11893, 2025.

[23] ZHANG C, JING M L, JIAO L, *et al*. Multiscale Mixed Transformer for Single Image Super-Resolution[J]. *Concurrency and Computation: Practice and Experience*, 2026, 38(6): e70649.

[24] TRAN D P, DO T, WAZIR S, *et al*. SAT: Selective Aggregation Transformer for Image Super-Resolution[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2026: 4982-4992.

[25] 陈清江,陈鹏民. 多维度聚合Transformer的图像超分辨率重建[J]. 光学精密工程,2025,33(12):1955-1970.
CHEN Q J, CHEN P M. MDAT: Multi-dimensional aggregation transformer for image super-resolution reconstruction[J]. *Optics and Precision Engineering*,2025,33(12):1955-1970. (in Chinese)

[26] 王宇轩,夏振平,罗格等. 基于CNN和Transformer架构的多聚焦图像融合研究及应用[J]. 光学精密工程,2025, 33(22):3577-3591.
WANG Y X, XIA Z P, LUO G, *et al*. Research and application of multi-focus image fusion based on CNN and Transformer architectures[J]. *Optics and Precision Engineering*,2025,33(22):3577-3591. (in Chinese)

[27] 董本志,于明聪,赵鹏. 基于小波域的图像超分辨率重建方法[J]. 液晶与显示,2021,36(2):317-326.
DONG B Z, YU M C , ZHAO P. Wavelet-based image super-resolution reconstruction method [J]. *Chinese Journal of Liquid Crystals and Displays*, 2021,36(2):317-326. (in Chinese)

[28] GUO T, SEYED MOUSAVI H, HUU VU T, *et al*. Deep wavelet prediction for image super resolution[C]. Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017: 104-113.

[29] XUE S, QIU W, LIU F, *et al*. Wavelet-based residual attention network for image super-resolution [J]. *Neurocomputing*, 2020, 382: 116-126.

[30] XIN J, LI J, JIANG X, *et al*. Wavelet-based dual recursive network for image super-resolution [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 33(2): 707-720.

[31] LI W, ZHOU K, QI L, *et al*. Lapar: Linearly-assembled pixel-adaptive regression network for single image super-

resolution and beyond[C]. Advances in Neural Information Processing Systems, 2020, 33: 20343-20355.

[32] ZHOU Y, LI Z, GUO C L, *et al*. Srformer: Permuted self-attention for single image super-resolution[C]// Proceedings of the IEEE/CVF international conference on computer vision, 2023: 12780-12791.

[33] ZOU W, CUI Y, WANG Y, *et al*. SP-MoMamba: Superpixel-driven Mixture of State Space Experts for Efficient Image Super-Resolution[J]. *arXiv preprint* arXiv:2605.25892, 2026.

作者简介：



黄佳怡，女，硕士研究生，2023年于陕西理工大学获得学士学位，主要从事图像处理方面的研究。E-mail: 24212030465@stumail.xsyu.edu.cn



景明利，男，博士，副教授，2014年于西安交通大学获得博士学位，主要从事图像处理、机器学习等方面的研究。E-mail: mljingsy@xsyu.edu.cn