Computer Science and Application 计算机科学与应用, 2026, 16(1), 8-19
Published Online January 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.161002

Hans汉斯

基于混合感知与频率自适应门控网络的轻量图像超分辨率重建算法

庞梦鑫*，董智红#，曹　鹏，张鸣赟

北京印刷学院信息工程学院，北京

收稿日期：2025年12月2日；录用日期：2025年12月30日；发布日期：2026年1月7日

摘　要

基于Transformer的方法凭借其卓越的长距离依赖建模能力，在单图像超分辨率领域取得了显著进展。然而，现有的轻量级Transformer架构在追求计算效率时，往往通过通道压缩或稀疏窗口机制来降低计算负担，此类策略削弱了局部特征的空间连续性，且自注意力机制固有的低通滤波特性限制了网络对高频纹理细节的恢复能力。为了解决上述频率偏差与局部信息丢失的问题，本文提出了一种基于混合感知与频率自适应门控网络的轻量图像超分辨率重建算法HPG-SR。首先，本文设计了混合感知门控注意力模块，通过并行使用局部感知分支和可学习的门控机制，在保留大窗口全局感受野的同时，显式地强化局部高频细节。其次，本文提出了多尺度门控前馈网络，利用双路多尺度卷积和上下文门控替代传统的静态激活函数，增强了网络对不同频率特征的自适应选择能力。最后，提出了对比度感知特征细化模块，利用标准差统计量强化对纹理丰富区域的特征响应。在五个基准数据集上的广泛实验表明，HPG-SR在参数量和计算量相当的情况下，性能优于当前最先进的轻量级SR方法。特别是在纹理复杂的Urban100数据集上，该算法展现出了更佳的细节恢复能力。

关键词

图像超分辨率，轻量级Transformer，混合感知，门控机制

Lightweight Image Super-Resolution Reconstruction Algorithm Based on Hybrid Perception and Frequency-Adaptive Gating Network

*第一作者。
#通讯作者。

文章引用: 庞梦鑫, 董智红, 曹鹏, 张鸣赟. 基于混合感知与频率自适应门控网络的轻量图像超分辨率重建算法[J]. 计算机科学与应用, 2026, 16(1): 8-19. DOI: 10.12677/csa.2026.161002

Mengxin Pang*, Zhihong Dong#, Peng Cao, Mingyun Zhang

School of Information Engineering, Beijing Institute of Graphic Communication, Beijing

Received: December 2, 2025; accepted: December 30, 2025; published: January 7, 2026

Abstract

Transformer-based methods have achieved significant progress in single image super-resolution due to their superior ability to model long-range dependencies. However, existing lightweight Transformer architectures often employ channel compression or sparse window mechanisms to reduce computational burden, which inevitably weakens the spatial continuity of local features. Furthermore, the inherent low-pass filtering nature of the self-attention mechanism limits the network's capacity to recover high-frequency texture details. To address the issues of frequency bias and local information loss, this paper proposes a lightweight image super-resolution reconstruction algorithm based on a Hybrid Perception and Frequency-Adaptive Gating network, named HPG-SR. First, a Hybrid Perception Gated Attention module is designed. By utilizing a parallel local perception branch and a learnable gating mechanism, it explicitly enhances local high-frequency details while retaining the global receptive field of large windows. Second, a Multi-Scale Gated Feed-Forward Network is proposed, which employs dual-path multi-scale convolutions and context gating to replace traditional static activation functions, thereby enhancing the network's adaptive selection capability for features across different frequencies. Finally, a Contrast-Aware Feature Refinement module is introduced to strengthen feature responses in texture-rich regions using standard deviation statistics. Extensive experiments on five benchmark datasets demonstrate that HPG-SR outperforms state-of-the-art lightweight SR methods with comparable parameters and computational complexity. Particularly on the texture-complex Urban100 dataset, the proposed algorithm exhibits superior detail recovery capability.

Keywords

Image Super-Resolution, Lightweight Transformer, Hybrid Perception, Gating Mechanism

1. 引言

单图像超分辨率(Single Image Super-Resolution, SISR)旨在从退化的低分辨率(Low-Resolution, LR)图像中恢复出高分辨率(High-Resolution, HR)图像，是计算机视觉领域中一个经典的病态逆问题[1]。随着深度学习的快速发展，SR 技术已广泛应用于移动终端图像处理、医学影像增强、视频监控以及卫星遥感等领域。尽管近年来基于深度卷积神经网络(Convolutional Neural Networks, CNN)的方法取得了显著进展[2]-[4]，但随着移动设备对实时性和能效比的要求日益提高，如何在有限的计算资源下实现高质量的图像重建，成为了轻量级 SR 研究的核心挑战。

早期的 SR 方法主要依赖于精心设计的 CNN 架构，如残差学习[2]、密集连接[5]和注意力机制[3]等。然而，CNN 固有的局部归纳偏置限制了其感受野，使其难以有效建立长距离像素依赖关系，往往导致重建图像在重复性纹理和结构边缘处出现模糊或伪影。为了解决这一问题，基于 Transformer 的架构被引入

SR 领域[6] [7]。凭借自注意力机制(Self-Attention, SA)强大的全局建模能力，Transformer 类方法在性能上大幅超越了传统 CNN。

尽管基于 Transformer 的模型在客观指标上取得了显著提升，但现有的轻量级架构设计仍面临着在效率与高频细节保持之间取得平衡的难题。首先，自注意力机制本质上表现出低通滤波特性[8]。虽然基于大窗口或稀疏注意力的方法能够捕捉全局形状，但在计算注意力图的过程中，往往会平滑掉对于 SR 任务至关重要的高频纹理细节。其次，为了降低大窗口注意力的计算复杂度，现有的主流方法通常采用通道压缩、空间置换或稀疏采样等策略[9] [10]。这些操作虽然换取了更大的感受野，却削弱了局部特征的空间连续性，导致微小细节恢复模糊。最后，现有的前馈神经网络(Feedforward Neural Network, FFN)大多沿用简单的多层感知机(Multilayer Perceptron, MLP)结构，采用固定的激活函数。这种静态的特征变换方式缺乏对多尺度特征的自适应选择能力，限制了网络对复杂纹理特征的表达效率。

针对上述局限性，本文提出了一种混合感知与频率自适应门控网络 HPG-SR。其核心动机在于弥合全局建模与局部细节之间的差异，通过设计混合感知机制，使网络能够同时具备 Transformer 的长距离依赖捕捉能力和 CNN 的局部高频提取能力。具体而言，本文设计了混合感知门控注意力(Hybrid Perception Gated Attention, HPGA)，在保留大窗口自注意力的同时，使用并行的局部感知分支来显式地补偿高频信息，并通过可学习的门控系数实现两者的动态融合。此外，本文重构了前馈网络，提出了多尺度门控前馈网络(Multi-Scale Gated FFN, MSG-FFN)，利用双路不同尺度的卷积和上下文门控机制来增强特征的非线性表达。最后，为了进一步提升重建质量，本文在重建头之前加入了对比度感知特征细化(Contrast-Aware Feature Refinement, CAFR)模块，利用标准差池化来强化高对比度纹理区域的特征响应。本文的主要贡献总结如下：

本文提出了一种基于混合感知的轻量级超分辨率网络 HPG-SR。该网络摒弃了单纯追求大窗口或深层堆叠的传统思路，转而通过混合感知策略高效地平衡了全局结构与局部纹理的恢复。

本文设计了 HPGA 和 MSG-FFN。前者通过并行分支解决了自注意力的高频丢失问题，后者通过门控机制提升了特征变换的频率选择性。

本文提出了 CAFR 模块，利用统计特征显式地增强了网络对纹理丰富区域的关注度。

在五个基准数据集上的广泛实验表明，HPG-SR 在参数量和计算量相当的情况下，性能优于当前最先进的轻量级 SR 方法。特别是在纹理复杂的 Urban100 数据集上，本文的模型展现出了卓越的细节恢复能力。

2. 相关工作

2.1. 基于 CNN 的轻量级超分辨率

自 SRCNN [11]首次将卷积神经网络引入图像超分辨率任务以来，深度学习方法已逐渐主导了该领域。早期的工作如 EDSR [2]和 RCAN [3]通过堆叠深层残差块和通道注意力机制取得了卓越的性能，但其庞大的参数量和计算开销限制了在移动设备上的部署。为了解决这一问题，轻量级网络设计成为研究热点。

IDN [12]提出了信息蒸馏模块，通过分离保留特征和精炼特征来减少计算冗余。在此基础上，IMDN [13]进一步引入了信息多重蒸馏网络，通过逐步提取分层特征实现了更高效的重构。RFDN [14]则利用残差特征蒸馏块取代了 IMDN 中的通道分裂操作，进一步降低了模型复杂度。此外，LatticeNet [15]和 LAPAR [16]等方法通过引入晶格块或线性组装像素自适应回归，在保持轻量级的同时提升了重建质量。

尽管上述基于 CNN 的方法在轻量化方面取得了显著进展，但卷积操作固有的局部感受野限制了网络捕捉图像全局结构信息的能力，使得在恢复具有长距离依赖的重复纹理时仍存在局限。

2.2. 基于 Transformer 的图像超分辨率

近年来，Transformer 凭借其强大的全局建模能力在计算机视觉领域表现出巨大潜力[17]。IPT [6]首次将 Transformer 应用于底层视觉任务，但其需要海量数据进行预训练且计算量巨大。SwinIR [7]成功地将 Swin Transformer 的移位窗口注意力机制引入 SR，在性能上大幅超越了基于 CNN 的方法，证明了非局部先验的重要性。

为了进一步降低计算复杂度并适配轻量级场景，后续涌现了多种改进方案。ESRT [18]使用高效的 Transformer 和轻量级 CNN 混合架构来降低内存占用。ELAN [9]提出了一组高效的长距离注意力网络，通过移位卷积共享注意力计算来加速推理。SRFormer [10]则提出了一种置换自注意力机制，通过压缩通道并置换空间维度来在更大的窗口内计算注意力，从而平衡了感受野与计算量。

然而，现有的 Transformer 类方法通常面临着共同的挑战。自注意力机制往往表现出低通滤波特性，容易平滑高频信息。此外，为了追求效率而采用的通道压缩或稀疏采样策略，可能会破坏特征的空间连续性，导致局部微小纹理的丢失。如何在一个统一的框架内同时实现高效的全局建模和精细的局部高频恢复，仍是一个亟待解决的问题。

2.3. 混合感知与门控机制

为了结合卷积和 Transformer 的优势，混合架构逐渐受到关注。一些工作尝试在 Transformer 模块中并行或串行插入卷积层，以补充局部归纳偏置。例如，ACMix [19]探索了卷积和注意力在特征提取上的互补性。然而，直接的模块堆叠往往带来参数量的激增。

另一方面，门控机制作为一种动态特征选择手段，在图像复原中展现出巨大潜力。NAFNet [20]证明了通过简单的乘法门控替代复杂的非线性激活函数，可以显著提升去噪和去模糊的性能。门控机制允许网络根据上下文信息自适应地调节信息流，这对于处理不同频率的图像特征尤为重要。受此启发，本文通过提出 HPGA 和 MSG-FFN，旨在通过显式的局部与全局分支融合和频率自适应门控，克服现有轻量级 Transformer 在高频细节恢复上的不足。

3. 方法

3.1. 整体架构

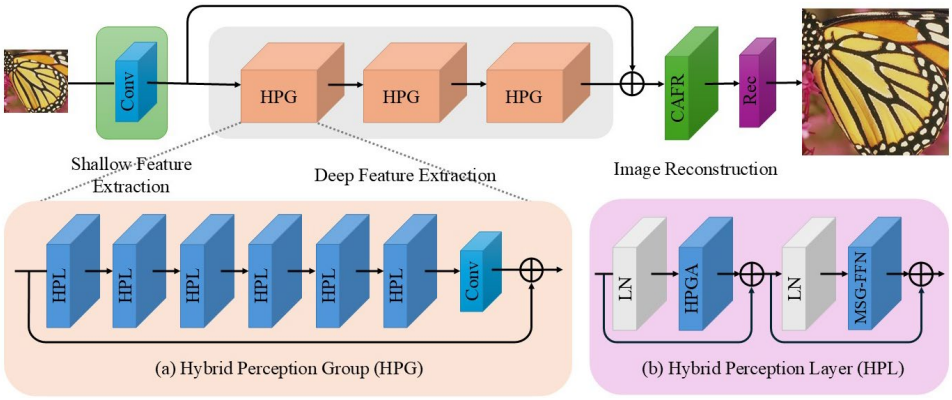

Figure 1. The overall architecture of the proposed HPG-SR. (a) The deep feature extraction stage is composed of N stacked Hybrid Perception Groups (HPG); (b) Each HPG contains multiple Hybrid Perception Layers (HPL), where each HPL consists of an HPGA module and an MSG-FFN module

图 1. 本文提出的 HPG-SR 整体架构图。(a) 深层特征提取由 N 个混合感知组(HPG)堆叠而成；(b) 每个 HPG 包含若干混合感知层(HPL)，每个 HPL 由 HPGA 和 MSG-FFN 模块组成

HPG-SR 旨在构建一个高效且能够精细恢复高频细节的轻量级超分辨率网络。如图 1 所示，网络主要包含三个阶段：浅层特征提取、深层特征提取和图像重建。

对于给定低分辨率输入图像 $I_{LR}\in\mathbb{R}^{H\times W\times C_{\text{in}}}$，首先使用一个 3×3 卷积层 $H_{SF}(\cdot)$将输入图像映射到特征空间，生成浅层特征 F_0：

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

其中 $F_0\in\mathbb{R}^{H\times W\times C}$，$C$ 为特征通道数。

随后，F_0 被送入深层特征提取模块 $H_{DF}(\cdot)$。该模块由 N 个堆叠的混合感知组(Hybrid Perception Group, HPG)组成。如图 1(a)所示，每个 HPG 包含若干个混合感知层(Hybrid Perception Layer, HPL)，如图 1(b)所示，每个 HPL 由一个 HPGA 模块和一个 MSG-FFN 模块组成。深层特征提取的输出 F_{DF} 包含了丰富的上下文信息。

为了进一步增强特征对纹理的响应，本文在深层特征提取之后使用了对比度感知特征细化模块 $H_{CAFR}(\cdot)$，并通过全局残差连接融合浅层特征：

$$F_{ref} = H_{CAFR}(F_{DF} + F_0) \tag{2}$$

最后，精炼后的特征 F_{ref} 通过包含上采样操作和卷积层的重建模块 $H_{Rec}(\cdot)$ 生成最终的高分辨率输出 I_{SR}：

$$I_{SR} = H_{\text{Rec}}(F_{ref}) \tag{3}$$

3.2. 混合感知门控注意力

标准的窗口自注意力虽然计算高效，但缺乏跨窗口的信息交互。SRFormer [10]提出的置换自注意力通过通道压缩和空间置换扩大了感受野，但本文观察到，对 K，V 矩阵的压缩和置换操作破坏了局部像素的空间邻域关系，导致高频纹理信息的丢失。为此，本文提出了 HPGA，通过并行双分支结构同时捕捉长距离依赖和局部高频细节。如图 2 所示，HPGA 主要由全局置换分支和局部感知分支两个分支组成。

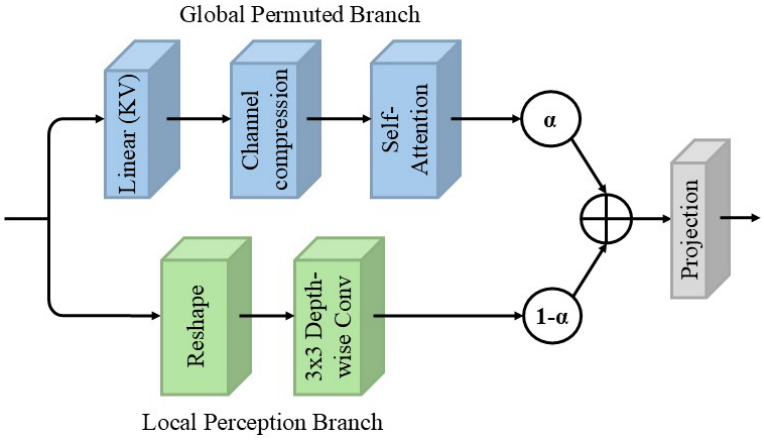

Figure 2. Structure of the Hybrid Perception Gated Attention (HPGA) module
图 2. 混合感知门控注意力(HPGA)模块结构

给定输入特征 $X\in\mathbb{R}^{H\times W\times C}$，首先将其重塑为非重叠窗口。为了降低计算量，本文采用线性投影层将键(Key)和值(Value)的通道维度压缩为原来的 $\frac{1}{R}$ (在本实验中设置压缩比 $R=2$)，并在空间维度

上进行置换操作以扩大感受野。令 Q， K， V 分别表示查询、键和值矩阵，全局注意力计算表示为：

$$\text{Attention}(Q,K,V)=\text{Soft max}\left(\frac{QK^T}{\sqrt{d_k}}+B\right)V \tag{4}$$

其中 $B\in\mathbb{R}^{M^2\times M^2}$ 是可学习的相对位置偏置参数(Learnable Relative Position Bias)，用于捕获序列内的相对距离信息， M 为窗口大小。 d_k 为缩放因子。该分支的输出记为 X_{global}。

为了补偿全局分支因压缩和置换造成的局部细节损失，本文加入了一个轻量级的局部感知路径。该路径直接在原始空间分辨率上操作，利用深度可分离卷积(Depthwise separable convolution, DSC)来提取局部的高频空间特征。假设 $F_{dw}(\cdot)$ 表示 3×3 的深度可分离卷积操作，局部特征 X_{local} 计算如下：

$$X_{local}=F_{dw}(X) \tag{5}$$

此操作在保持较低的计算成本，即参数量仅为 $C\times3\times3$ 的同时，能够有效地保留图像的边缘和纹理信息。

此外，为了让网络根据图像内容的频率特性动态平衡全局和局部信息，本文使用了一个可学习的门控标量 α。最终 HPGA 的输出 Y_{HPGA} 为：

$$Y_{HPGA}=\text{Proj}\left(\alpha\cdot X_{global}+(1-\alpha)\cdot X_{local}\right)+X \tag{6}$$

其中 $\text{Proj}(\cdot)$ 为线性投影层， α 初始化为 0.5 并参与端到端训练。这种设计使得 HPGA 能够在平坦区域更多地依赖全局信息，而在纹理丰富区域更多地利用局部细节。

3.3. 多尺度门控前馈网络

前馈网络负责特征的变换与非线性映射。现有的轻量级网络通常采用 MLP 的结构，或者仅添加简单的卷积。这种单一尺度的处理方式限制了特征的感受野多样性，且固定的激活函数缺乏对特征的选择性。受 Gated CNN 的启发，本文提出了 MSG-FFN 如图 3 所示，利用多尺度卷积和门控机制来增强特征表达。

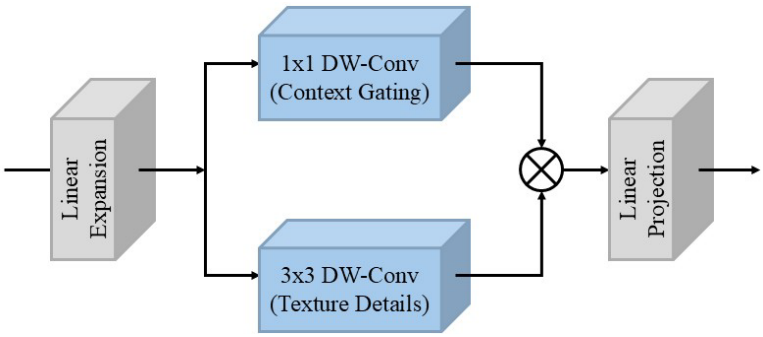

Figure 3. Structure of the Multi-Scale Gated Feedforward Network (MSG-FFN)
图 3. 多尺度门控前馈网络(MSG-FFN)结构

对于输入特征 X，MSG-FFN 首先通过一个线性层将通道扩展并分割为两部分 X_1 和 X_2：

$$[X_1,X_2]=Split(Linear_{in}(X)) \tag{7}$$

其中 $X_1,X_2\in\mathbb{R}^{H\times W\times C}$。

随后，本文在两个分支上分别应用不同尺度的深度可分离卷积。分支 1 使用 3×3 卷积提取空间细节，分支 2 使用 1×1 卷积作为上下文信息流：

$$Y_1 = DWConv_{3\times3}(X_1) \tag{8}$$

$$Y_2 = DWConv_{1\times1}(X_2) \tag{9}$$

最后，本文利用 Hadamard 积实现门控机制，利用 Y_2 来动态调制 Y_1 的特征响应，并通过输出线性层进行融合：

$$F_{MSG}(X) = Linear_{out}(Y_1 \odot Y_2) + X \tag{10}$$

这种设计不仅具备多尺度感受野，还通过门控机制替代了传统的非线性激活函数，能够更有效地过滤冗余信息并传递高频特征。

3.4. 对比度感知特征细化

在深层特征进入上采样模块之前，特征图中的不同通道往往包含不同类型的信息。传统的通道注意力仅利用全局平均池化(Global Average Pooling, GAP)来聚合统计信息：

$$z_{avg} = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{i,j} \tag{11}$$

然而，GAP 容易平滑掉纹理丰富的区域，导致网络无法区分平坦背景和高频震荡纹理，因为两者可能具有相同的均值。

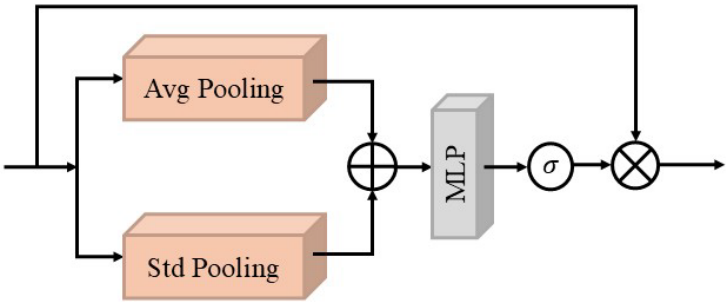

Figure 4. Structure of the Contrast-Aware Feature Refinement (CAFR) module
图 4. 对比度感知特征细化(CAFR)模块结构

为了解决这一问题，如图 4 所示，CAFR 模块使用了基于标准差的统计量。标准差能够衡量特征在空间维度上的变化剧烈程度，即对比度。对于第 c 个通道，其标准差统计量 z_{std}^{c} 计算如下：

$$z_{std}^{c} = \sqrt{\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(X_{i,j}^{c} - z_{avg}^{c}\right)^2 + \varepsilon} \tag{12}$$

本文将平均池化特征与标准差池化特征相加，形成对比度感知的特征描述符 z：

$$z = z_{avg} + z_{std} \tag{13}$$

随后，利用两层多层感知机和 Sigmoid 函数生成通道加权系数 w，并对原始特征进行重校准：

$$w = \sigma\left(MLP(z)\right) \tag{14}$$

$$F_{ref} = X \cdot w \tag{15}$$

CAFR 模块显式地增强了网络对高对比度、富含纹理的通道的关注，从而在上采样阶段能够生成更清晰的图像细节。

3.5. 损失函数

为了优化所提出的网络，本文采用 L_1 像素损失函数[21]。相较于 L_2 损失，L_1 损失在处理异常值时更为鲁棒，且更有利于产生锐利的边缘。给定 M 个训练图像对 $\left\{I_{LR}^{(i)}, I_{HR}^{(i)}\right\}_{i=1}^{M}$，损失函数定义为：

$$L = \frac{1}{M}\sum_{i=1}^{M}\left\| H_{HPG-SR}\left(I_{LR}^{(i)}\right) - I_{HR}^{(i)}\right\|_1 \tag{16}$$

其中 $H_{HPG-SR}(\cdot)$ 表示本文的网络模型。

4. 实验

4.1. 实验设置

为了保证比较的公平性，本文遵循广泛使用的实验协议。训练阶段仅使用 DIV2K 数据集[22]，包含 800 张高质量 2 K 分辨率的训练图像。在测试阶段，本文在五个标准的基准数据集上进行评估：Set5 [23]、Set14 [24]、BSD100 [25]、Urban100 [26]和 Manga109 [27]。

本文将峰值信噪比(Peak signal-to-noise ratio, PSNR)和结构相似性(Structure Similarity Index Measure, SSIM) [28]作为客观评估指标。所有指标均在转换到 YCbCr 色彩空间后的 Y 通道上进行计算。此外，本文也报告了模型的参数量和计算量 FLOPs 以评估模型的计算复杂度。FLOPs 是基于将低分辨率图像重建为1280×720 分辨率的高分辨率图像的计算量来统计的。实验基于 PyTorch 框架[29]，在 Ubuntu 22.04 操作系统上运行，硬件环境包含两块 NVIDIA RTX 4090 GPU。本文将低分辨率图像随机裁剪为 64×64 的 patch 作为输入，并采用随机旋转和水平翻转进行数据增强。使用 *Adam* 优化器[30]训练模型，参数设置为 $\beta_1 = 0.9$ 和 $\beta_2 = 0.99$。初始学习率设置为 2×10^{-4}，总迭代次数为 500,000 次，学习率在迭代次数达到设定点时分别减半。为了优化收敛，本文采用 L_1 像素损失函数。

4.2. 定量评估

如表 1 所示，本文将提出的 HPG-SR 与多种主流的轻量级 SR 方法进行了比较，包括基于 CNN 的方法 CARN 以及基于 Transformer 的方法 SwinIR-light、ELAN、SRFormer-light。

Table 1. Parameters, FLOPs, PSNR, and SSIM of different SR methods at scales ×2, ×3, and ×4
表 1. 不同 SR 方法在×2、×3、×4 尺度下的参数量、FLOPs、PSNR 和 SSIM

模型	缩放因子	参数量	FLOPs	Set5	Set14	BSD100	Urban100	Manga109
				PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
CARN	×2	1592 K	222.8 G	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256	38.36/0.9765
A²F	×2	1363 K	306.1 G	38.09/0.9607	33.78/0.9192	32.23/0.9002	32.46/0.9313	38.95/0.9772
SwinIR-light	×2	910 K	244.4 G	38.14/0.9611	33.86/0.9206	32.31/0.9012	32.76/0.9340	39.12/0.9783
ELAN-light	×2	582 K	168.4 G	38.17/0.9611	33.94/0.9207	32.30/0.9012	32.76/0.9340	39.11/0.9782
SRFormer-light	×2	853 K	236.3 G	38.23/0.9613	33.94/0.9209	32.36/0.9019	32.91/0.9353	39.28/0.9785
本文	×2	876 K	243.7 G	**38.24/0.9615**	**33.96/0.9211**	**32.43/0.9025**	**33.10/0.9367**	**39.42/0.9796**
CARN	×3	1592 K	118.8 G	34.29/0.9255	30.29/0.8407	29.06/0.8034	28.06/0.8493	33.50/0.9440
A²F	×3	1367 K	136.3 G	34.54/0.9283	30.41/0.8436	29.14/0.8062	28.40/0.8574	33.83/0.9463
SwinIR-light	×3	918 K	110.8 G	34.62/0.9289	30.54/0.8463	29.20/0.8082	28.66/0.8624	33.98/0.9478

续表

ELAN-light	×3	590 K	75.7 G	34.61/0.9288	30.55/0.8463	29.21/0.8081	28.69/0.8624	34.00/0.9478
SRFormer-light	×3	861 K	105.4 G	34.67/0.9296	30.57/0.8469	29.26/0.8099	28.81/0.8655	34.19/0.9489
本文	×3	881 K	117.1 G	**34.73/0.9304**	**30.61/0.8477**	**29.31/0.8115**	**28.94/0.8670**	**34.30/0.9503**
CARN	×4	1592 K	90.9 G	32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837	30.47/0.9084
A²F	×4	1374 K	77.2 G	32.32/0.8964	28.67/0.7839	27.62/0.7379	26.32/0.7931	30.72/0.9115
SwinIR-light	×4	930 K	63.6 G	32.44/0.8976	28.77/0.7858	27.69/0.7406	26.47/0.7980	30.92/0.9151
ELAN-light	×4	601 K	43.2 G	32.43/0.8975	28.78/0.7858	27.69/0.7406	26.54/0.7982	30.92/0.9150
SRFormer-light	×4	873 K	62.8 G	32.51/0.8988	28.82/0.7872	27.73/0.7422	26.67/0.8032	31.17/0.9165
本文	×4	889 K	65.3 G	**32.59/0.8997**	**28.86/0.7884**	**27.78/0.7435**	**26.79/0.8038**	**31.27/0.9181**

实验结果表明，本文的 HPG-SR 在所有测试数据集和所有缩放倍率下均取得了最佳性能。值得注意的是，在最具挑战性的包含大量建筑细节和重复纹理的 Urban100 数据集上，HPG-SR 展现出了显著的优势。在×2 超分辨率任务中，HPG-SR 的 PSNR 达到了 33.10 dB，相较于先进的 SRFormer-light [10]提升了 0.19 dB；在×4 任务中，PSNR 提升了 0.12 dB。这充分证明了本文提出的 HPGA 模块和 CAFR 模块在保留高频细节方面的有效性。

此外，本文的模型参数量和计算量在与基于 Transformer 的 SR 模型相当的前提下，取得了显著的性能提升。这表明 HPG-SR 成功地在模型复杂度和重建质量之间取得了更优的平衡。

4.3. 视觉质量评估

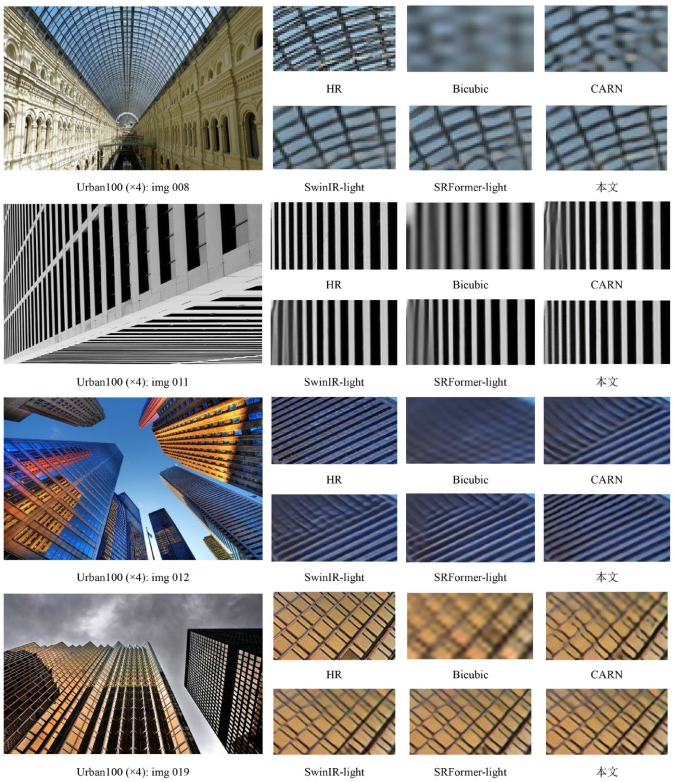

Figure 5. Visual comparison of ×4 super-resolution on Urban100 dataset
图 5. Urban100 数据集上×4 超分辨率的视觉对比

为了直观地展示 HPG-SR 的重建质量，本文在图 5 中展示了 Urban100 数据集上×4 超分辨率的视觉对比结果。

从视觉结果可以看出，基于 CNN 的方法在处理密集的网格和条纹图案时，往往会出现模糊或严重的混叠伪影和结构上的错乱。虽然 SRFormer 通过大窗口注意力改善了结构恢复，但在极细微的纹理处仍显得不够清晰，存在过度平滑的现象。

相比之下，HPG-SR 重建的图像纹理更加锐利，边缘更加清晰，且极大地减少了混叠效应。这种视觉上的提升得益于混合感知机制，局部感知分支保留了纹理的锐度，而门控机制有效地过滤了错误的频率响应，使得重建结果更接近真实图像。

5. 消融实验

为验证本文提出的各模块的有效性，本文通过逐步添加 HPGA，MSG-FFN 和 CAFR 模块进行了消融实验，得出的 PSNR 和 SSIM 结果为测试集上的平均值。如表 2 所示，展示了不同模块组合下的性能表现。其中，Model 1 作为基线模型(Baseline)，其架构移除了 HPGA 中的局部感知分支和门控机制，退化为标准的窗口自注意力；同时移除了 MSG-FFN 中的多尺度卷积与门控，退化为标准的前馈神经网络(Feed-Forward Network, FFN)，且未包含 CAFR 模块。

Table 2. Ablation study of core modules on Urban100 (×4) dataset
表 2. 在 Urban100 (×4)数据集上对核心模块的消融研究

Model	HPGA	MSG-FFN	CAFR	PSNR	SSIM
1				32.91	0.9353
2	✓			33.00	0.9360
3		✓		32.97	0.9358
4	✓	✓		33.07	0.9364
5	✓	✓	✓	**33.10**	**0.9367**

如表 2 模型 2 所示，使用本文提出的 HPGA 时，PSNR 提升至 33.00 dB。证明了并行加入的局部感知分支有效地补偿了由大窗口机制导致的细节丢失。局部卷积与全局注意力的互补性使得网络在保持长距离建模能力的同时，能够更敏锐地捕捉高频纹理。对比模型 1 和模型 3，使用 MSG-FFN 带来了 0.06 dB 的性能增益。这表明，相比于单一尺度的卷积和固定的 GELU 激活，使用多尺度感受野以及上下文门控机制，能够显著增强网络对特征的非线性变换能力，使其能够自适应地筛选有用的频率信息。当同时使用 HPGA 和 MSG-FFN 时，性能进一步提升至 33.07 dB，说明这两个模块在特征提取上是相互兼容的。在加入 CAFR 模块后，PSNR 最终达到了 33.10 dB。其中 CAFR 通过显式地对高对比度通道进行加权，成功地为重建模块提供了更优质的特征表示。

6. 结论

本文提出了一种高效的轻量级图像超分辨率网络 HPG-SR。针对现有窗口注意力机制在高频细节恢复上的局限性，本文探索了一种混合感知的设计范式。通过提出的 HPGA，本文成功地将 Transformer 的全局建模能力与 CNN 的局部纹理提取能力在特征层面进行了动态融合。同时，MSG-FFN 和 CAFR 模块的设计进一步增强了网络在特征变换过程中的频率自适应能力和纹理敏感度。

定性和定量的实验结果均表明，HPG-SR 在保持轻量级特性的同时，实现了先进的重建性能，特别是

在处理规则结构和高频纹理时优势明显。本文的工作证明了在轻量级模型设计中，显式地平衡全局与局部感知、关注频率特性是提升性能的关键。本文希望提出的混合感知与门控策略能为未来的高效超分辨率模型设计提供新的思路。

基金项目

北京市科技计划课题(Z241100007624008)；北京印刷学院信息与通信工程一级学科博士点培育项目(21090525004)；北京印刷学院科研平台建设项目(KYCPT202509)。

参考文献

[1] Yang, J.C, Wright, J., Huang, T.S., *et al*. (2010) Image Super-Resolution via Sparse Representation. *IEEE Transactions on Image Processing*, **19**, 2861-2873. https://doi.org/10.1109/tip.2010.2050625

[2] Lim, B., Son, S., Kim, H., Nah, S. and Lee, K.M. (2017) Enhanced Deep Residual Networks for Single Image Super-Resolution. 2017 *IEEE Conference on Computer Vision and Pattern Recognition Workshops* (CVPRW), Honolulu, 21-26 July 2017, 136-144. https://doi.org/10.1109/cvprw.2017.151

[3] Zhang, Y., Li, K., Li, K., *et al*. (2018) Image Super-Resolution Using Very Deep Residual Channel Attention Networks. *Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, 8-14 September 2018, 286-301.

[4] Ahn, N., Kang, B. and Sohn, K.A. (2018) Fast, Accurate, and Lightweight Super-Resolution with Cascading Residual Network. *Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, 8-14 September 2018, 252-268.

[5] Zhang, Y., Tian, Y., Kong, Y., Zhong, B. and Fu, Y. (2018) Residual Dense Network for Image Super-Resolution. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 2472-2481. https://doi.org/10.1109/cvpr.2018.00262

[6] Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., *et al*. (2021) Pre-Trained Image Processing Transformer. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 12294-12305. https://doi.org/10.1109/cvpr46437.2021.01212

[7] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L. and Timofte, R. (2021) SwinIR: Image Restoration Using Swin Transformer. 2021 *IEEE/CVF International Conference on Computer Vision Workshops* (*ICCVW*), Montreal, 11-17 October 2021, 1833-1844. https://doi.org/10.1109/iccvw54120.2021.00210

[8] Park, N. and Kim, S. (2022) How Do Vision Transformers Work? International Conference on Learning Representations (ICLR), 2022. https://openreview.net/forum?id=D78Go4hVcxO

[9] Zhang, X., Zeng, H., Guo, S. and Zhang, L. (2022) Efficient Long-Range Attention Network for Image Super-Resolution. In: *Lecture Notes in Computer Science*, Springer, 649-667. https://doi.org/10.1007/978-3-031-19790-1_39

[10] Zhou, Y., Li, Z., Guo, C.L., *et al*. (2023) SRFormer: Permuted Self-Attention for Single Image Super-Resolution. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Paris, 1-6 October 2023, 12780-12791.

[11] Dong, C., Loy, C.C., He, K. and Tang, X. (2015) Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**, 295-307. https://doi.org/10.1109/tpami.2015.2439281

[12] Hui, Z., Wang, X. and Gao, X. (2018) Fast and Accurate Single Image Super-Resolution via Information Distillation Network. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 723-731. https://doi.org/10.1109/cvpr.2018.00082

[13] Hui, Z., Gao, X., Yang, Y. and Wang, X. (2019) Lightweight Image Super-Resolution with Information Multi-Distillation Network. *Proceedings of the 27th ACM International Conference on Multimedia*, Nice, 21-25 October 2019, 2024-2032. https://doi.org/10.1145/3343031.3351084

[14] Liu, J., Tang, J. and Wu, G. (2020) Residual Feature Distillation Network for Lightweight Image Super-resolution. In: *Lecture Notes in Computer Science*, Springer, 41-55. https://doi.org/10.1007/978-3-030-67070-2_2

[15] Luo, X., Xie, Y., Zhang, Y., Qu, Y., Li, C. and Fu, Y. (2020) Latticenet: Towards Lightweight Image Super-Resolution with Lattice Block. In: *Lecture Notes in Computer Science*, Springer, 272-289. https://doi.org/10.1007/978-3-030-58542-6_17

[16] Li, W., Zhou, K., Qi, L., *et al*. (2020) Lapar: Linearly-Assembled Pixel-Adaptive Regression Network for Single Image Super-Resolution and beyond. *Advances in Neural Information Processing Systems*, **33**, 20343-20355.

[17] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al*. (2021) An Image Is Worth 16x16 Words: Transformers for Image

Recognition at Scale. *International Conference on Learning Representations*, Austria, 3-7 May 2021.

[18] Lu, Z.S., Li, J.C., Liu, H., Huang, C.Y., Zhang, L.L. and Zeng, T.Y. (2022) Transformer for Single Image Super-Resolution. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), New Orleans, 19-20 June 2022, 456-465. https://doi.org/10.1109/cvprw56347.2022.00061

[19] Pan, X., Ge, C., Lu, R., Song, S., Chen, G., Huang, Z., *et al*. (2022) On the Integration of Self-Attention and Convolution. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 805-815. https://doi.org/10.1109/cvpr52688.2022.00089

[20] Chen, L., Chu, X., Zhang, X. and Sun, J. (2022) Simple Baselines for Image Restoration. In: *Lecture Notes in Computer Science*, Springer, 17-33. https://doi.org/10.1007/978-3-031-20071-7_2

[21] Zhao, H., Gallo, O., Frosio, I. and Kautz, J. (2016) Loss Functions for Image Restoration with Neural Networks. *IEEE Transactions on Computational Imaging*, **3**, 47-57. https://doi.org/10.1109/tci.2016.2644865

[22] Agustsson, E. and Timofte, R. (2017) NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. 2017 *IEEE Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), Honolulu, 21-26 July 2017, 1122-1131. https://doi.org/10.1109/cvprw.2017.150

[23] Bevilacqua, M., Roumy, A., Guillemot, C. and Morel, M.A. (2012) Low-Complexity Single-Image Super-Resolution Based on Nonnegative Neighbor Embedding. *Proceedings of the British Machine Vision Conference* 2012, Surrey, 3-7 September 2012, 135.1-135.10. https://doi.org/10.5244/c.26.135

[24] Zeyde, R., Elad, M. and Protter, M. (2012) On Single Image Scale-Up Using Sparse-Representations. In: *Lecture Notes in Computer Science*, Springer, 711-730. https://doi.org/10.1007/978-3-642-27413-8_47

[25] Martin, D., Fowlkes, C., Tal, D., *et al*. (2001) A Database of Human Segmented Natural Images and Its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. *Proceedings of the 8th IEEE International Conference on Computer Vision*, Vancouver, 7-14 July 2001, 416-423.

[26] Huang, J.B., Singh, A. and Ahuja, N. (2015) Single Image Super-Resolution from Transformed Self-Exemplars. 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, 7-12 June 2015, 5197-5206. https://doi.org/10.1109/cvpr.2015.7299156

[27] Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., *et al*. (2017) Sketch-Based Manga Retrieval Using Manga109 Dataset. *Multimedia Tools and Applications*, **76**, 21811-21838. https://doi.org/10.1007/s11042-016-4020-z

[28] Wang, Z., Bovik, A.C., Sheikh, H.R., *et al*. (2004) Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, **13**, 600-612. https://doi.org/10.1109/tip.2003.819861

[29] Paszke, A., Gross, S., Massa, F., *et al*. (2019) PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, **32**, 8024-8035.

[30] Kingma, D.P. and Ba, J. (2014) Adam: A Method for Stochastic Optimization. International Conference on Learning Representations (ICLR), 2015.