用于图像超分辨率的傅里叶引导注意力上采样

Daejune Choi Youchan No Jinhyung Lee Duksu Kim
Korea University of Technology and Education (KOREATECH)
{ eowns13, bluebear102, refracta, bluekds } @koreatech.ac.kr

单图像超分辨率（SISR）旨在从低分辨率（LR）输入中重建高分辨率（HR）图像，恢复通常在降采样期间退化的精细纹理、锐利边缘和结构细节。受深度学习驱动，卷积神经网络(CNNs) [12–14, 24, 31, 34, 39, 41, 49, 57–59] 和基于 Transformer 的架构 [3, 4, 7, 8, 19, 29, 30, 53] 的最新进展显著推动了 SISR 的前沿，在广泛使用的基准测试上取得了令人印象深刻的性能。

尽管取得了这些进展，恢复高频细节如重复纹理、边缘和精细轮廓仍然充满挑战，因为这些特征在低分辨率图像中通常严重退化或完全缺失。单图像超分辨率（SISR）结果的质量关键取决于上采样模块是否能重建出合理的高频信息。传统方法如基于插值的方法、反卷积（deconvolution）以及广泛使用的子像素卷积（SPC）[47] 通常会引入空间伪影，尤其是棋盘格模式和频谱混叠，这是由于它们固有的空间和频率不敏感导致的（见第 2 节和图 ?? ）。

为了应对这些限制，我们提出了傅里叶引导注意力（FGA）上采样模块——这是一种轻量级、频率感知的解决方案，专门设计用于增强高频保真度并减少重建伪影（参见部分 3 ）。FGA 集成了：（1）基于傅里叶特征的多层感知器（FF-MLP），用于显式编码位置和光谱信息，（2）关联注意力层（CAL），通过跨分辨率注意力自适应地将高分辨率特征与低分辨率上下文对齐，以及（3）直接监督光谱保真的频域 L1 损失（FL1）。我们的广泛实验表明，FGA 在各种最先进的 SISR 骨干和不同基准上始终提高定量指标和感知质量，使其成为传统上采样技术的多功能和有效替代方案（参见部分 ?? ）。

1. 相关工作

1.1. 使用深度神经网络进行单图像超分辨率

深度神经网络在单图像超分辨率（SISR）领域取得了显著的进展，能够从低分辨率（LR）图像到高分辨率（HR）图像学习复杂的映射，并超越传统的插值方法。这些进展可以根据网络架构进行广泛分类，具体来说包括基于卷积神经网络（CNN）的方法、注意力机制和基于 Transformer 的模型。

早期的基于 CNN 的方法，如 SRCNN [13] ，通过超越传统插值方法，确立了深度学习在单帧图像超分辨率（SISR）中的有效性。随后的方法通过增加网络深度和引入高级卷积设计 [18, 27, 52, 58] 在此基础上进行了改进。值得注意的是，EDSR [31] 通过去除冗余组件如批归一化层，优化了残差学习，大大提高了效率和性能。同时，RDN [58] 引入了残差密集块（RDBs），利用密集连接的卷积层和连续内存机制进行丰富的局部特征提取和融合。然而，基于 CNN 的模型在恢复锐利、详细的图像重构所需的关键高频细节方面仍面临挑战。

注意机制通过使模型能够优先考虑显著特征、捕捉长距离依赖关系以及有效重建高分辨率图像，解决了这些限制。与纯 CNN 模型相比，增强注意力的模型能够更全面地捕捉全局上下文 [9, 12, 33, 39, 41, 57, 59] 。例如，RCAN [57] 利用深度残差内残差（RIR）架构搭配通道注意力，自适应地重新校准通道特征并强调信息丰富的区域，以实现出色的判别性学习。

最初为自然语言处理引入的 Transformer 架构，最近在单帧图像超分辨率中显示了有希望的结果。IPT [3] 采用基于 Transformer 的编码器-解码器框架进行多任务预训练，在各种低级视觉任务中取得了卓越的性能。基于 Swin Transformer [36] 的 SwinIR [30] ，采用固定窗口内的自注意力机制和移位窗口策略以扩大感受野，超过了许多基于 CNN 的方法。像 Uformer [53] 、CAT [7] 、HAT [4] 和 DRCT [19] 等模型通过多样的窗口注意力、特征聚合和密集连接进一步创新 Transformer 架构，实现了最先进的图像修复效果。

虽然近年来基于 Transformer 的 SISR 模型主要通过增强编码器结构来捕捉复杂的表示，我们提出的方法则创新性地专注于解码器。具体来说，我们通过将频域特征集成到解码阶段，增强了广泛采用的像素混洗方法，从而显著提高重建保真度和高频细节的恢复。

最近的单图像超分辨率（SISR）进展利用频域信息来提高重建质量，特别是通过小波变换和离散余弦变换（DCT）等技术明确地结合全局频率提示。例如，DWSR [16] 将小波变换应用于图像分解为低频和高频子带，并利用卷积神经网络（CNNs）有效结合这些子带以进行重建。WaveletSRNet [21] 引入了一个基于小波的损失函数，用于从低分辨率图像预测高分辨率系数，特别适用于人脸图像超分辨率。MWCNN [35] 在 U-Net 结构中集成了多级小波变换，增强了各种图像恢复任务的多尺度表示能力。此外，Dario 等人引入的傅里叶空间损失 [15] 直接在傅里叶域监督高频内容的恢

复，显著改善了图像的感知质量。WaveMixSR [23] 提出了一种基于小波分解的频率感知令牌混合策略，在 BSD100×2 基准测试 [37] 上实现了最新的结果。此外，ORDSR [17] 在卷积层中集成了 DCT 操作，促进了平滑的空间到频域的过渡。与之前的大多数将频率信息主要嵌入编码器模块的工作不同，我们的方法显著地将频域特征直接集成到解码器中。我们新颖的设计在解码阶段明确利用傅里叶特征，提供增强的重建保真度和更出色的高频细节恢复。

1.2. 单图像超分辨率网络中的上采样模块

上采样模块在单图像超分辨率（SISR）中起着关键作用，能够提升特征图的空间分辨率。上采样技术的选择显著影响 SISR 模型的质量和计算效率。一般来说，上采样方法可以分为基于插值的技术、反卷积层和像素重排的方法。

插值方法，如双线性或双三次插值，代表最简单的方式，即基于相邻像素强度生成缺失像素。Dong 等人 [13] 采用插值，然后使用卷积层来细化细节。类似地，Odena 等人 [42] 将最近邻插值与卷积层结合起来，从而实现更平滑的重建，并减少视觉伪影。反卷积层或转置卷积 [55]，通过适应训练数据中的滤波器，提供了一种可学习的上采样解决方案。然而，这些方法通常会出现棋盘伪影，需要精细的架构设计和额外的技术来缓解这些问题。

像素洗牌由 ESPCN [47] 引入，将特征图的通道重新排列为扩展的空间维度，使其在计算上高效，并在近年来的 SISR 模型中被广泛采用 [3, 4, 7, 8, 12, 19, 27, 29–31, 39, 41, 52, 57–59]。此外，最近的任意比例超分辨率技术已引入了创新的上采样模块 [5, 6, 20, 28, 40, 43, 51]，然而像素洗牌在固定比例超分辨率中仍然很普遍。

在这项工作中，我们专注于引入一个专为固定比例超分辨率网络设计的新型上采样模块，旨在克服现有方法的局限性，并进一步提高重建质量。

2. 现有上采样模块中伪影的分析

早期的 SISR 模型通常采用预上采样技术，这些技术在高分辨率域中进行特征提取 [13, 24, 25, 54]。然而，由于在大空间分辨率上操作，这些方法导致了大量的计算开销。因此，后上采样策略开始流行，突出显示了高效和有效的上采样模块在现代 SISR 流程中的关键作用 [3, 4, 7, 8, 12, 19, 27, 29–31, 39, 41, 52, 57–59]。

在 SISR 中的上采样模块可以分为传统的基于插值的方法和学习的方法，例如反卷积（转置卷积）和子像素卷积（PixelShuffle）。插值方法，比如双三次插值，通过基于预定义的核从邻近像素中估计新的像素值。虽然计算简单且高效，但这些方法固有地缺乏恢复高频细节的能力，从而导致输出结果光滑且模糊。[25, 26, 44, 47]

去卷积层利用可学习的核自适应地进行上采样，但由于核重叠不均，往往会引入棋盘状伪影，从而导致频谱混叠和不良的视觉网格模式 [42]。这样的伪影显著削弱了图像重建的准确性。

PixelShuffle (PS) [47]，或称子像素卷积，将低分辨率特征图的通道重新排列到空间维度上，从而产生高分辨率的输出。由于其计算效率和在端到端单图像超分辨率（SISR）框架中的无缝集成，PS 被广泛使用，如 EDSR [31]。然而，PS 可能会产生与频谱混叠 [1] 密切相关的棋盘状伪影。

图 ?? 在 EDSR 骨干网络中比较了上采样方法——基于插值的卷积（Interp+Conv）、PS 和反卷积（Deconv），并保持相同的网络组件和训练配置以保证公平。模型在 DIV2K 数据集上训练了 500k 次迭代，并在 Urban100 测试集 [22] 上进行了评估。

图 ?? 通过快速傅里叶变换（FFT）获得的频谱进行可视化，比较编码器输出（放大前）和放大后的特征图（放大后）。传统方法的频谱显示出明显的伪影，尤其是复制的频率模式，指示高频分量折叠到基带中，突出了频谱混叠的问题。这些频域伪影通常在空间上表现为棋盘格图案。图 ?? 通过基于 PCA 的编码器和放大特征图在 ×4 缩放下的降维可视化进一步说明了这些伪影。

这些观察强调了现有上采样方法的局限性，激励我们提出一种稳健的、频率感知的替代方案。如图 ?? 中的“我们的（FGA）”列所示，我们的方法有效地缓解了频谱混叠和棋盘伪影，实现了更清晰的空间特征表示和更一致的频率分布。

3. 傅里叶引导注意力上采样

在本节中，我们首先概述我们提出的上采样器的架构。然后，我们描述构成我们方法基础的关键组件和设计原则。

3.1. FGA 升频器概述

我们提出了一种傅里叶引导注意力 (FGA) 上采样器，通过利用频率感知表示和基于注意力的空间细化来增强高分辨率重建。如图 1 所示，FGA 模块位于主干网络之后，直接对其输出特征图进行操作。与传统的上采样方法不同，FGA 被明确设计用于通过频域编码和自适应相关建模来减轻频谱混叠并保留细粒度的纹理。

FGA 上采样器由两个主要组件组成：（1）基于傅里叶特征的多层感知机（FF-MLP），以及（2）相关注意层（CAL）。核心结构包括 N 个重复的阶段，每个阶段由一个卷积层、FF-MLP 和像素混洗单元组成，然后是一个 CAL 模块。最后应用一个卷积层来生成高分辨率输出。

FF-MLP 模块通过注入从空间坐标 (x, y) 导出的傅里叶位置编码来增强特征表示。这些编码特征与前一阶段的中间特征结合，并通过 MLP 处理，以提高空间辨别能力和对高频内容的敏感度。

CAL 模块通过计算上采样特征与骨干网络特征之间的查询-键相关性来执行基于注意力的精炼。这使得网络能够自适应地强调信息丰富的区域，并校正上采样过程中引入的空间不一致性。

这些组件共同使 FGA 能够有效地抑制混叠伪影并增强重建纹理的逼真度。整个模块是完全可微的，可以

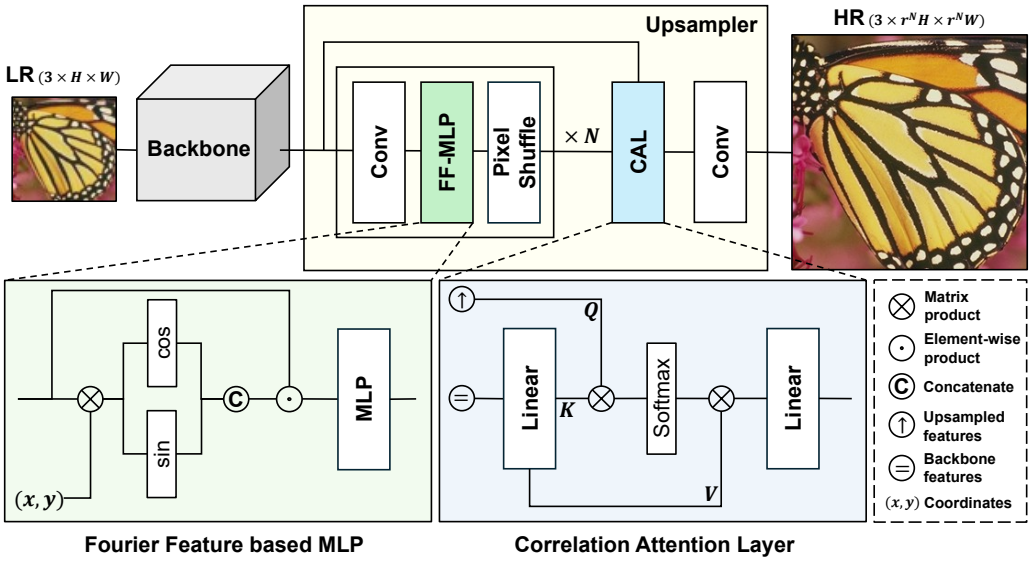


Figure 1. 提出的傅里叶引导注意力 (FGA) 上采样器概览。

端到端训练，因此与多种 SISR 骨干架构兼容。

3.2. 基于傅里叶特征的多层感知器

虽然像素洗牌上采样器效率很高，但本质上难以重建高频细节。因为在像素洗牌之前的卷积层无法识别它们生成的子像素的精确空间目的地，特征通道常常交错而没有位置区分。这种空间条件的缺失导致相同的高频成分在子像素间重复，产生了诸如重复纹理和混叠的伪影。

为了解决这些限制，我们通过浅层 MLP 增强了像素移动块，该 MLP 由显式位置信息引导。受隐式神经表示 [6, 40, 43] 中傅里叶特征位置编码及其在任意尺度超分辨率 [28] 中的适应启发，我们将归一化的二维坐标嵌入注入 MLP。与推理过程中嵌入高分辨率坐标的任意尺度方法不同，我们的设计针对固定尺度超分辨率进行了调整，通过预对齐每个子像素特征与其在输出网格中的指定空间位置来实现。

为了提高效率，我们将坐标编码过程与像素洗牌流水线对齐。与其为每个像素计算高分辨率级别的嵌入，我们将特征张量划分为不同的子像素组，并为每个组在低分辨率空间中分配一个唯一的坐标身份。这确保了位置条件而不会产生冗余计算。

坐标感知特征嵌入 令 $F \in \mathbb{R}^{\frac{H}{r}\times\frac{W}{r}\times C}$ 表示上采样前的骨干特征图，其中 r 是上采样因子。我们首先生成一个归一化的高分辨率坐标网格 $v \in \mathbb{R}^{H\times W\times 2}$，并应用 PixelUnshuffle 以这些坐标与亚像素布局对齐：

$$V = \text{PixelUnshuffle}(v), \quad V \in \mathbb{R}^{\frac{H}{r}\times\frac{W}{r}\times r^2\times 2} \tag{1}$$

每个 r^2 亚像素组都被分配了其自身的空间身份向量，我们通过傅里叶特征对此进行编码。这些坐标编码与主干特征通过元素级相乘融合，以生成频率感知表示：

$$F_{\text{ff}} = F \odot \begin{bmatrix}\cos(\pi \cdot FV)\\ \sin(\pi \cdot FV)\end{bmatrix} \tag{2}$$

然后将调制后的特征 F_{ff} 传递给一个 MLP：

$$F_{\text{MLP}} = \text{MLP}(F_{\text{ff}}) \tag{3}$$

最后，像素洗牌操作重建了上采样的特征图：

$$F_{\uparrow} = \text{PixelShuffle}(F_{\text{MLP}}), \quad F_{\uparrow} \in \mathbb{R}^{H\times W\times C} \tag{4}$$

该结构确保每个子像素通道基于其特定的空间目标进行调制，使 MLP 能够为不同空间位置学习不同的高频行为。因此，我们的方法能够有效地抑制混叠现象，并减少棋盘格伪影。

与 Transformer 风格的多头的关系 我们使用的子像素空间头与 transformers 中的多头注意力在概念上有相似之处。transformer 头学习独立的通道关系，而我们的空间头通过傅里叶编码学习位置依赖的基。每个头仅处理与特定子像素组对应的坐标特征，使得 MLP 能够产生空间上多样化的输出，并最小化子像素间的高频复制。

为了进一步缓解混叠伪影并纠正天真上采样引入的空间不一致性，我们引入了相关注意力层（CAL）。不同于典型的上采样方法仅依赖于在高分辨率空间中学习的特征，CAL 通过使高分辨率特征能够通过局部化注意力选择性地关注其对应的低分辨率特征，从而显式地重用原始的低分辨率上下文。该模块通过关注空间对齐和内容优化来补充 FF-MLP 的频率感知增强。

基于窗口的跨分辨率注意力 令 $F_{\text{LR}} \in \mathbb{R}^{\frac{H}{r}\times\frac{W}{r}\times C}$ 和 $F_{\text{HR}} \in \mathbb{R}^{H\times W\times C}$ 分别表示放大倍率为 r 时上采样前后的特征图。按照来自 SwinIR [30] 的基于窗口的注意力策略，我们将两个特征域分别划分为大小为 $(\frac{M}{r}, \frac{M}{r})$ 的 LR 域和大小为 (M, M) 的 HR 域的不重叠窗口。这样得到 $N = \frac{HW}{M^2}$ 对应窗口对，记为

$$F^w_{\text{LR}} \in \mathbb{R}^{N\times\frac{M^2}{r^2}\times C}, \quad F^w_{\text{HR}} \in \mathbb{R}^{N\times M^2\times C}.$$

。

在每个窗口内，我们执行交叉分辨率注意机制：HR 特征窗口 $w_\uparrow$ 充当查询，而对应的 LR 窗口 w 提供键和值。注意机制和随后的细化过程定义为：

$$\begin{aligned} Q &= w_\uparrow, \quad K = wP_K, \quad V = wP_V, \\ w_\uparrow &= \text{XRA}(w, w_\uparrow) + w_\uparrow, \\ w_\uparrow &= \text{MLP}(w_\uparrow) + w_\uparrow, \end{aligned} \tag{5}$$

，其中 P_K 和 P_V 是可学习的线性投影矩阵，XRA 表示交叉分辨率注意操作。该机制使得 HR 表示能够从其对应的 LR 窗口中整合结构指导，从而提高空间一致性并减轻上采样导致的伪影。

与 HR 域中的传统自注意力相比，CAL 通过利用 LR 特征的较小空间尺寸进行键值计算，提供了一种更高效的解决方案。标准自注意力（SA）的计算复杂度是：

$$\Omega(\text{SA}) = 4HWC^2 + 2M^2HWC, \tag{6}$$

，而 CAL 的复杂度变为：

$$\Omega(\text{CA}) = \left(1 + \frac{2}{r^2}\right) HWC^2 + \frac{2M^2}{r^2} HWC. \tag{7}$$

虽然 Swin Transformer [30] 通过堆叠多个移位窗口块来缓解局部注意力的有限感受野，但这样的堆叠在分辨率提高时会导致计算量呈二次增长——这使得它不适合轻量级 SISR 上采样器。

为了更有效地扩展感受野，我们采用了重叠窗口相关注意机制（OWCA） [4] ，该机制通过单个注意层扩大空间覆盖范围。具体来说，我们在上采样之前对低分辨率（LR）特征图应用重叠分区，使得每个 LR 窗口在使用比其高分辨率（HR）对应窗口更少像素的情况下能够访问更广泛的上下文。

对于重叠率 α ，LR 域的窗口大小变为方程 ?? 。尽管重叠增加了感受野，但它仅导致适度的计算开销：

$$\Omega(\text{OWCA}) = \left(1 + \frac{2}{r^2}\right) HWC^2 + (1+\alpha)^2 \cdot \frac{2M^2}{r^2} HWC. \tag{8}$$

该设计通过利用更广泛的 LR 上下文来指导 CAL 中的高分辨率细化，从而在效率和效果之间取得了平衡。

3.3. 频域 L1 损失

为了明确监督频谱保真度，我们提出了一种频域 L1 损失（$\mathcal{L}_{\text{freq}}$ ），该损失用于惩罚傅里叶频谱中振幅和相位的差异。这与我们减轻混叠和保留细节的目标一致。

令 $\hat{F}$ 和 F 分别表示预测图像和真实图像的离散傅里叶变换，逐通道计算：

$$\mathcal{L}_{\text{freq}} = \frac{2}{UVC} \sum_{c=1}^{C} \sum_{u=0}^{\lfloor U/2 \rfloor} \sum_{v=0}^{V-1} \left| \hat{F}_c(u,v) - F_c(u,v) \right|_1, \tag{9}$$

，其中 (U, V) 是光谱维度，$|\cdot|_1$ 表示实部和虚部的绝对差之和 [1] 。因子 2 补偿了实值图像中的厄米对称性。

通过同时约束幅度和相位，此损失实现了细粒度的频率一致性，减少了诸如振铃、棋盘效应和空间错位等伪影。由于仅评估频谱中非冗余的一半，它在计算上也保持高效。为了评估我们建议的 FGA 上采样器的影响，我们将其整合到五个广泛采用的超分辨率基础模型中（EDSR [31] , RCAN [57] , HAN [41] , NLSN [59] , 和 SwinIR [30] ），并在标准的 $\times 4$ 图像超分辨率任务上对每个模型进行基准测试。

在所有设置中，子像素卷积（SPC）被用作基线的上采样模块，从而能够与我们提出的 FGA 进行一致的比较。我们提出的 FGA 模块相比 SPC 引入了大约 0.3M 的额外参数，并且这种额外负担在所有骨干网络中都保持不变，这证明了性能的提升不是由于模型容量的增加。在 FGA 模块中的基于窗口的跨分辨率注意力（详见第 ?? 节）中，我们采用窗大小为 5×5 用于上采样前的特征和 4×4 用于上采样后的特征，因为根据经验评估（见 ?? ），此配置在性能和效率之间提供了最佳的平衡。

为了将上采样模块的影响与整体模型容量区分开来，我们在两种不同的容量设置下评估每个骨干网络（表 1 ）：

- 原始变体：全尺寸模型（12-40M 参数），从公开发布的检查点初始化。
- 轻量化变体（记为 $^\mu$ ）：减少容量版本（1-9M 参数），通过将通道宽度和深度减半而获得，并从头开始训练。

Table 1. 使用 SPC 或我们提出的 FGA 上采样器对每个骨干网的参数计数（以百万为单位）。每种模型都在轻量级（$^\mu$ ）和原始变体中进行评估。

Setting	EDSR	RCAN	HAN	NLSN	SwinIR
$^\mu$ + SPC	1.5	4.3	8.6	1.6	1.2
$^\mu$ + FGA	1.8	4.6	8.9	1.9	1.5
Original + SPC	38.7	15.6	16.1	39.7	11.9
Original + FGA	39.0	15.9	16.4	40.0	12.2

训练协议 除非另有说明，我们遵循每个主干的原始训练过程。EDSR $^\mu$ 、RCAN $^\mu$ 、HAN $^\mu$ 和 NLSN

[1]! $\left|\hat{F}_c(u,v) - F_c(u,v)\right|_1 = \left|\text{Re}\,\hat{F}_c(u,v) - \text{Re}\,F_c(u,v)\right| + \left|\text{Im}\,\hat{F}_c(u,v) - \text{Im}\,F_c(u,$

Table 2. 在六个基准数据集上进行的量化 × 4 SR 结果，以 PSNR (dB) / SSIM 的形式报告。每个骨干网都在其原始和轻量级变体（标记为 “$^{\mu}$ ”）中进行评估。在每种配置中，第一行使用 SPC 作为基线上采样器，而第二行使用我们提出的 FGA 模块。加粗数值表示在每种设置中 SPC 和 FGA 之间的优越性能。

Backbone	Upsampler	Set5		Set14		BSD100		Urban100		Manga109		DTD235	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
EDSR $^{\mu}$	SPC	32.12	0.8953	28.59	0.7822	27.58	0.7378	26.08	0.7860	30.41	0.9077	29.73	0.7599
	FGA (ours)	32.30	0.8968	28.68	0.7838	27.64	0.7392	26.27	0.7896	30.74	0.9106	29.80	0.7601
EDSR	SPC	32.46	0.8968	28.80	0.7876	27.71	0.7420	26.64	0.8033	31.02	0.9148	29.91	0.7664
	FGA (ours)	32.50	0.8995	28.80	0.7877	27.74	0.7437	26.67	0.8030	31.05	0.9158	29.89	0.7655
RCAN $^{\mu}$	SPC	32.34	0.8975	28.69	0.7851	27.65	0.7406	26.44	0.7972	30.78	0.9124	29.88	0.7649
	FGA (ours)	32.43	0.8984	28.76	0.7858	27.70	0.7414	26.52	0.7974	30.94	0.9135	29.90	0.7641
RCAN	SPC	32.63	0.9002	28.87	0.7889	27.77	0.7436	26.82	0.8087	31.22	0.9173	30.00	0.7681
	FGA (ours)	32.63	0.9007	28.88	0.7891	27.79	0.7449	26.83	0.8067	31.23	0.9161	30.02	0.7678
HAN $^{\mu}$	SPC	32.44	0.8988	28.76	0.7865	27.70	0.7419	26.55	0.8002	30.96	0.9142	29.92	0.7665
	FGA (ours)	32.52	0.8997	28.85	0.7878	27.75	0.7431	26.70	0.8024	31.21	0.9160	29.97	0.7660
HAN	SPC	32.64	0.9002	28.90	0.7890	27.80	0.7442	26.85	0.8094	31.42	0.9177	30.06	0.7703
	FGA (ours)	32.68	0.9013	28.93	0.7897	27.81	0.7454	26.91	0.8087	31.43	0.9181	30.06	0.7689
NLSN $^{\mu}$	SPC	32.30	0.8969	28.69	0.7841	27.65	0.7401	26.37	0.7937	30.75	0.9112	29.84	0.7653
	FGA (ours)	32.48	0.8990	28.81	0.7863	27.73	0.7425	26.64	0.7994	31.17	0.9155	29.96	0.7666
NLSN	SPC	32.59	0.9000	28.87	0.7891	27.78	0.7444	26.96	0.8109	31.27	0.9184	30.08	0.7715
	FGA (ours)	32.72	0.9021	28.98	0.7907	27.85	0.7471	27.09	0.8136	31.59	0.9207	30.11	0.7713
SwinIR $^{\mu}$	SPC	32.49	0.8994	28.86	0.7884	27.75	0.7441	26.66	0.8040	31.23	0.9172	29.96	0.7683
	FGA (ours)	32.59	0.9007	28.97	0.7905	27.82	0.7455	26.86	0.8064	31.55	0.9195	30.06	0.7693
SwinIR	SPC	32.92	0.9044	29.09	0.7950	27.92	0.7489	27.45	0.8254	32.03	0.9260	30.28	0.7789
	FGA (ours)	32.96	0.9052	29.13	0.7957	27.96	0.7510	27.47	0.8249	32.14	0.9261	30.29	0.7774

$^{\mu}$ 在 DIV2K 数据集上训练 [32]，批量大小为 16。SwinIR $^{\mu}$ 遵循其原始配置，并使用 DF2K（DIV2K + Flickr2K [48]），批量大小为 32。所有模型都在 RTX A6000 GPU 上训练 50 万次迭代，学习率在 250K、400K、450K 和 475K 迭代时衰减。评估在单个 RTX A6000 GPU 上进行。

我们在六个基准数据集上评估模型性能：Set5 [2]，Set14 [56]，BSD100 [37]，Urban100 [22]，Manga109 [38] 和 DTD235。DTD235 是从可描述纹理数据集（DTD [10]）中随机选择的 235 张图像的自定义子集，特别强调具有挑战性的高频纹理。

3.4. 结果

表 2 展示了在原始和轻量级（ $^{\mu}$ ）配置下，使用五种骨干模型进行 ×4 超分辨率的六个基准测试的 PSNR 和 SSIM 分数。对于每个设置，我们将基线 SPC 上采样器与我们提出的 FGA 模块进行比较。

FGA 在所有架构和数据集上始终优于 SPC，证实了其在提高重建质量时可以在不显著增加模型复杂性的情况下发挥作用。平均而言，FGA 为轻量级网络带来了 +0.14 dB 的 PSNR 提升，并为全容量（即原始）模型带来了 +0.12 dB 的提升。这一增益在高频基准测试如 Urban100 和 Manga109 上尤为显著，在这些测试中，细纹理恢复至关重要。例如，在 NLSN $^{\mu}$ 中用 FGA 替代 SPC 将 Urban100 上的 PSNR 从 26.37 dB 提升到 26.64 dB（+0.27 dB），并将 Manga109 上的 PSNR 从 30.75 dB 提高到 31.17 dB（+0.42 dB）。

这些结果强调了即使在仅增加 0.3M 参数的固定成本下，FGA 的益处也能可靠地随骨干大小扩展。在 PSNR 和 SSIM 上的持续改进验证了我们的频率感知上采样策略在增强结构保真度的同时有效地减少了伪影，无论底层网络能力如何。

图 2 在 × 4 倍图像上采样的情况下提供了跨数据集和主干网络的定性比较。在所有主干网络和数据集中，FGA 显著改善了重复结构、锐利轮廓和纹理等细节的重构，同时减少了使用 SPC 上采样器时常见的混叠伪影。无论是轻量级还是全容量主干网络，这些改进都很明显，强化了 FGA 在有效增强空间细节和减少混叠方面不受模型大小影响的能力。

3.5. 光谱特征分析

为了更好地理解 FGA 所提供的改进来源，我们重新审视了在超分辨率过程中产生的特征图的频域特性。根据第 2 节的分析，我们检查了上采样前后的中间表示，特别关注我们的 FGA 与 SPC 相比如何修改光谱结构。

图 ?? 展示了使用三种不同上采样策略的 EDSR 方法产生的 PCA 降维特征图和频谱图。频谱图表明 SPC 和去卷积层存在明显的混叠现象，包括重复的高频模式和棋盘伪影。相比之下，FGA 抑制了这些模式，产生的频谱与 GT 分布更为接近。这些观察表明，FGA 成功地减轻了频谱折叠并增强了局部结构。

为了确认这些改进在不同架构上具有普遍性，图

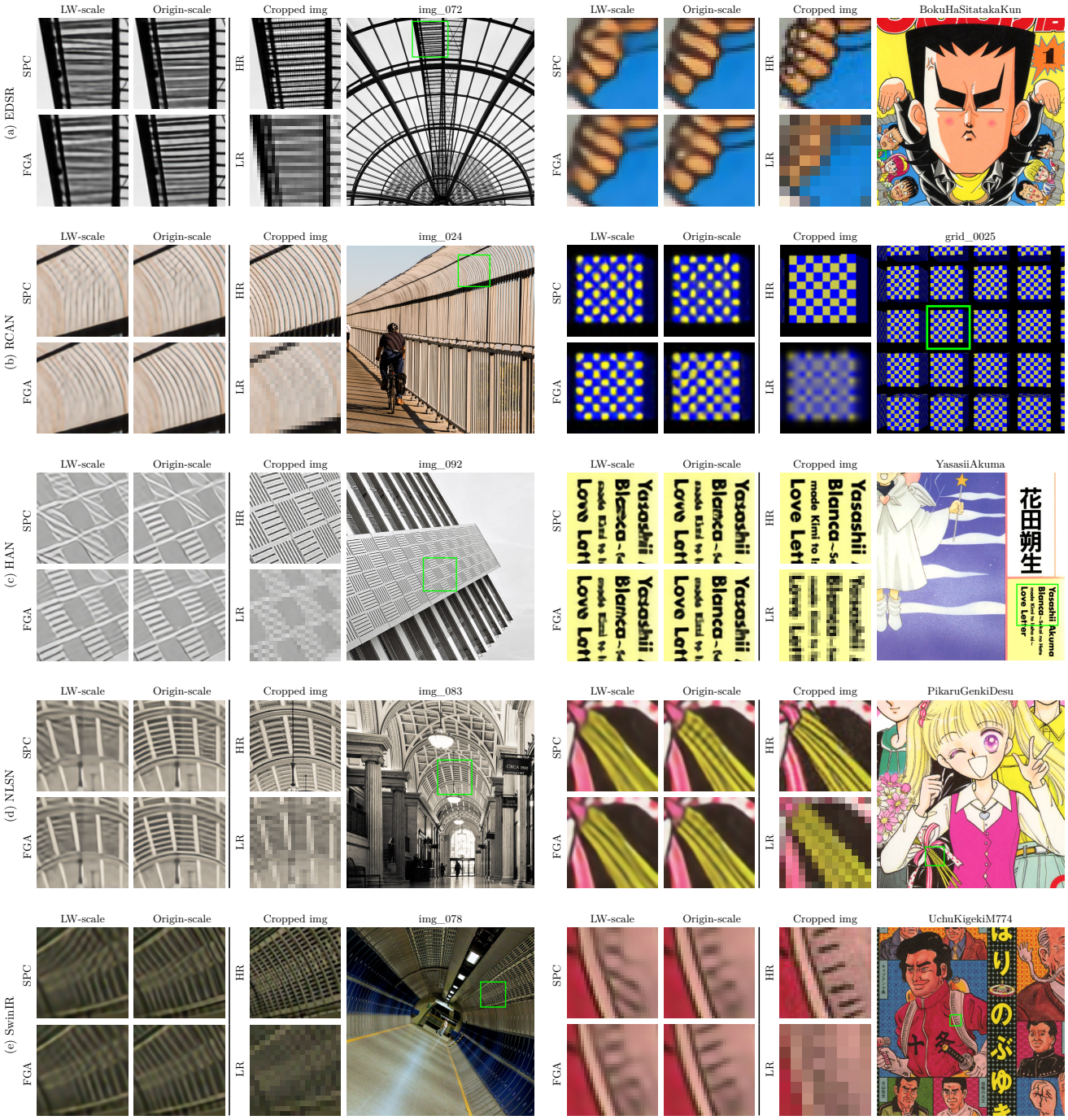


Figure 2. 在 Urban100、Manga109 和 DTD235 上的定性 ×4 超分辨率结果。每一行比较轻量级（LW-scale）和原始（Origin-scale）变体的五种 SR 骨干之间 SPC（顶部）和我们的 FGA（底部）。最右边的列显示完整的高分辨率参考；绿色框表示放大的区域。

3 对 SwinIR $^{\mu}$ 进行了相同的分析。即使在未见过的 Urban100 测试图像上，FGA 的后上采样频率图仍保留了更高的频率成分，并且显示出比 SPC 更少的混叠伪影。图 3b 的可视化结果也揭示，FGA 生成的空间特征图噪声更少，边缘过渡更清晰。

综上所述，这些结果提供了有力的视觉证据，表明所提出的 FGA 直接在特征层面减少了频谱混叠。通过在上采样之前显式编码空间频率和位置信息，FGA 促进了更一致的特征变换，从而转换为感知上更加清晰和结构上更加一致的重建。

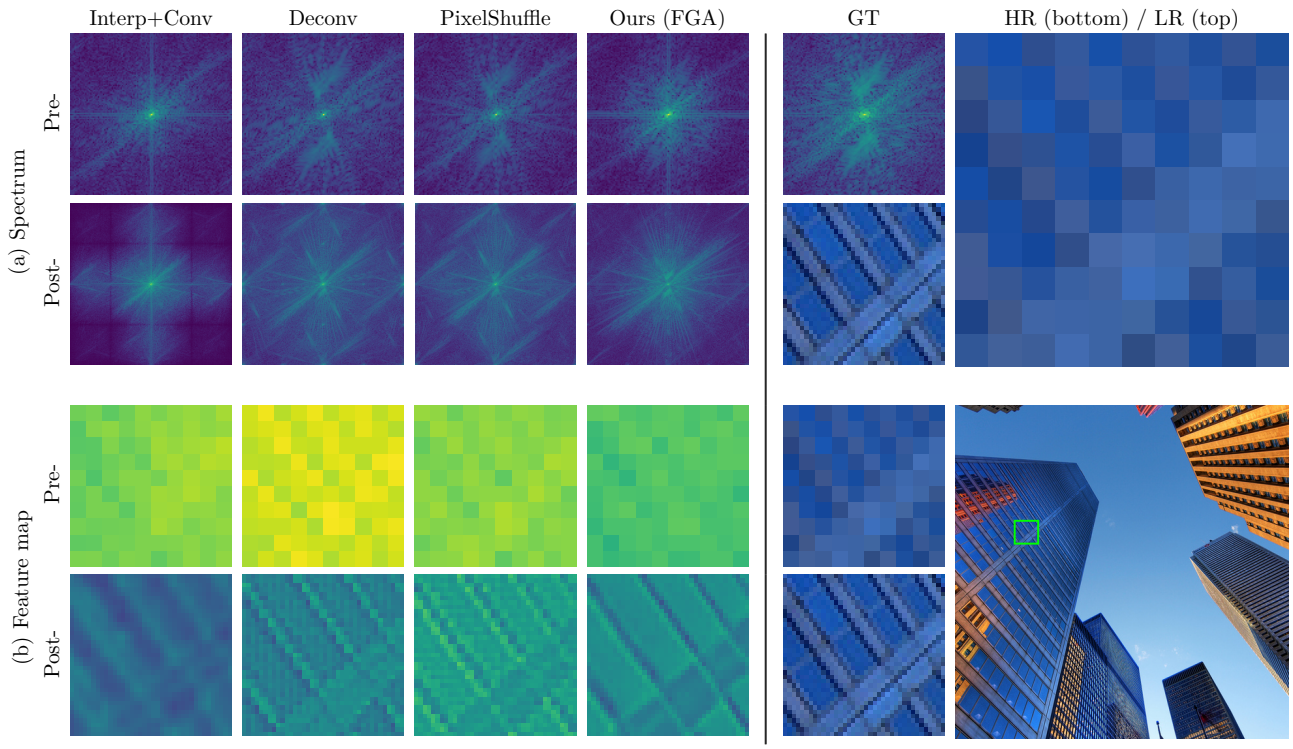


Figure 3. 在 Urban100 [22] 的 “img_012” 上使用 SwinIR $^{\mu}$ 骨干网进行不同上采样方法的特征和频谱可视化。输入经过中心裁剪至 128 × 128 像素并用 ×4 SR 推理进行处理。前两行显示上采样前后的特征图，而后两行显示它们对应的傅里叶频谱。

3.6. 通过 FRC 的定量频率分析

为了定量评估 SR 输出在频域的一致性，我们采用了傅里叶环相关（FRC），这最初是为了评估冷冻电镜中的分辨率而开发的 [45, 46] 。FRC 通过计算频率空间中同心环上的相关性来测量傅里叶域中两幅图像的结构相似性。给定两幅图像的 2D 傅里叶变换 $F_1(\boldsymbol{f})$ 和 $F_2(\boldsymbol{f})$ （例如，SR 输出和 GT），空间频率半径为 $q = \|\boldsymbol{f}\|$ 的 FRC 得分定义为：

$$\mathrm{FRC}(q) = \frac{\sum_{|\boldsymbol{f}|=q} F_1(\boldsymbol{f})\, F_2^*(\boldsymbol{f})}{\sqrt{\Big(\sum_{|\boldsymbol{f}|=q} |F_1(\boldsymbol{f})|^2\Big)\Big(\sum_{|\boldsymbol{f}|=q} |F_2(\boldsymbol{f})|^2\Big)}} \tag{10}$$

。根据先前的研究 [50] ，我们将 2D 频域量化为 $N = 64$ 个同心环，每个环大致包含相同数量的频域像素（见图 4c ），从而实现对重建图像和参考图像之间的精细和频率化的比较。

在所有测试的数据集中，FRC 曲线的前 75 % 保持接近 1.0，这反映出在主导全局结构的低到中频上有很强的一致性（例如，图 4a 和 4b ）。然而，在环的后 25 %（48–63）中出现了有意义的差异，这些对应于高频内容，例如纹理和精细细节。我们将此上四分位数中的平均 FRC 定义为一个标量指标：

$$\text{FRC-AUC} = \frac{1}{N_{\mathrm{HF}}} \sum_{i=i_{\mathrm{HF}}}^{N-1} \mathrm{FRC}_i, \quad i_{\mathrm{HF}} = \lceil 0.75N \rceil = 48, \tag{11}$$

其中 $N_{\mathrm{HF}} = 16$ ($N - i_{\mathrm{HF}}$)。该指标为 SR 重建中细节保真的简洁和针对性度量。

FRC-AUC 结果分析 表 3 汇总了所有评估的骨干网络及两种上采样器（SPC 对比 FGA）在六个基准测试中的 FRC-AUC 分数。在整体上，用我们提出的 FGA 替换 SPC 可一致地提高高频保真度。特别是在像 Urban100 和 Manga109 这样纹理丰富的数据集上，提升尤为显著。例如，NLSN $^{\mu}$ 在 Urban100 数据集上的分数从 0.2330 提高到 0.3017 (+29 %)，而 EDSR $^{\mu}$ 和 SwinIR $^{\mu}$ 也取得了类似的相对增益 (+26 %)。即使在完全容量的模型中，改进也不可忽视，例如，NLSN 从 0.3019 上升到 0.3178 (+5.3 %)。有趣的是，我们观察到 HAN 的 FRC-AUC 分数低于其轻量级对手 HAN $^{\mu}$ ，这归因于某些样本中的边缘光晕伪影（见 A ）。尽管如此，我们提出的 FGA 始终能够提升 HAN 和 HAN $^{\mu}$ 的 FRC-AUC 分数，显示出其在不同骨干变体中的稳健性。

平均而言，与 SPC 相比，FGA 在 Urban100 上的 FRC-AUC 提高了 +10 %，在 Manga109 上提高了 +7

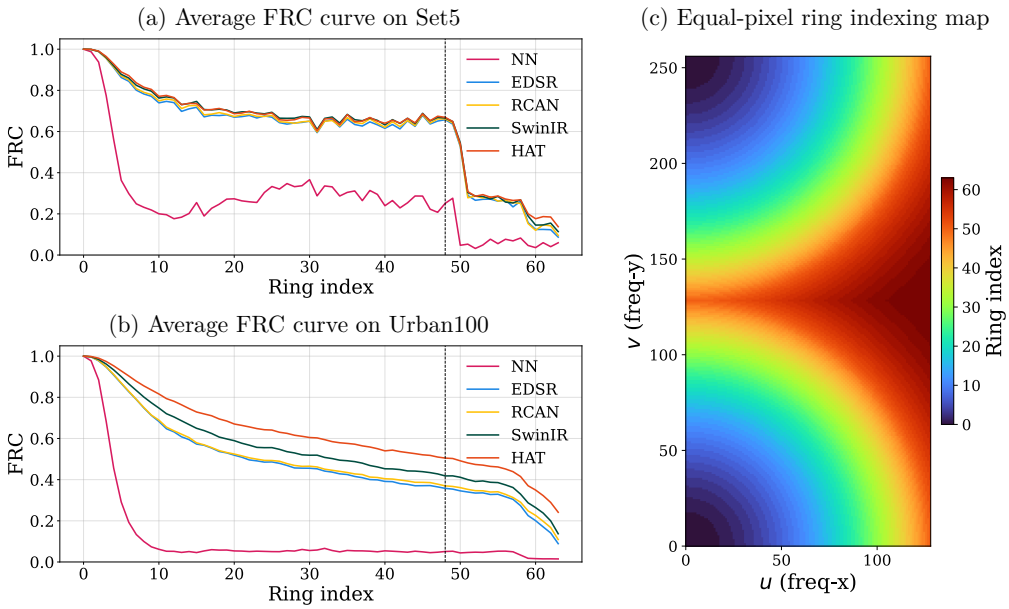


Figure 4. FRC 分析中使用的图和频率环索引，其中 NN 表示最近邻插值。(a)，(b)：Set5 和 Urban100 上的平均 FRC 曲线。(c)：用于光谱分箱的 64 个量化同心环的可视化；颜色表示环的索引。

Table 3. 通过在 × 个超分辨率基准上使用 FRC-AUC 测量的高频重建质量。对于每个主干网络，第一行显示的是使用基线上采样器（SPC）的结果，第二行则使用了提出的 FGA 模块。较高的分数表示对高频内容的更好保留。

Backbone	Upsampler	Set5	Set14	B100	U100	M109	D235
		FRC-AUC					
EDSR $^{\mu}$	SPC	0.2648	0.2093	0.2126	0.2226	0.2998	0.1625
	FGA (ours)	0.2941	0.2347	0.2294	0.2804	0.3313	0.1758
EDSR	SPC	0.2857	0.2281	0.2279	0.2809	0.3221	0.1745
	FGA (ours)	0.2920	0.2358	0.2362	0.3010	0.3341	0.1771
RCAN $^{\mu}$	SPC	0.2740	0.2186	0.2199	0.2534	0.3216	0.1691
	FGA (ours)	0.3020	0.2373	0.2339	0.2975	0.3323	0.1771
RCAN	SPC	0.2924	0.2367	0.2365	0.2962	0.3387	0.1788
	FGA (ours)	0.2947	0.2433	0.2389	0.3097	0.3424	0.1795
HAN $^{\mu}$	SPC	0.2824	0.2289	0.2299	0.2799	0.3287	0.1735
	FGA (ours)	0.3028	0.2406	0.2378	0.3119	0.3487	0.1796
HAN	SPC	0.2211	0.1815	0.1983	0.1810	0.3224	0.1564
	FGA (ours)	0.2953	0.2397	0.2373	0.3070	0.3450	0.1776
NLSN $^{\mu}$	SPC	0.2657	0.2155	0.2144	0.2330	0.3142	0.1670
	FGA (ours)	0.2963	0.2439	0.2366	0.3017	0.3439	0.1799
NLSN	SPC	0.2851	0.2354	0.2357	0.3019	0.3380	0.1769
	FGA (ours)	0.2999	0.2440	0.2415	0.3178	0.3502	0.1820
SwinIR $^{\mu}$	SPC	0.2655	0.1902	0.2209	0.2446	0.3342	0.1641
	FGA (ours)	0.2965	0.2286	0.2392	0.3080	0.3548	0.1740
SwinIR	SPC	0.2955	0.2238	0.2470	0.3371	0.3616	0.1805
	FGA (ours)	0.3028	0.2336	0.2492	0.3477	0.3715	0.1815

%，展示了其在高频内容保护上的优越性。有趣的是，使用 FGA 的轻量级变体通常优于使用较大 SPC 的对应物，例如，EDSR $^{\mu}$ + FGA 在 Set5、Manga109 和 DTD235 上的表现超过了完整的 EDSR + SPC。类似的跨尺度改进也出现在 RCAN 和 SwinIR 中。

这些结果证实，频率感知的 FGA 设计在不同的架构和模型容量中提供了稳定的光谱细节改进，验证了其作为插入式上采样替代方案的有效性。

高频段的视觉比较 我们在图 5 中进一步可视化了最高频分位数。这些图隔离了超分辨率模型差异最大的细节区域。在所有的模型中，FGA 产生的高频模式更密集、更一致，与 GT 参考紧密对齐，而 SPC 的输出则显得支离破碎或噪声较大。放大图显示，FGA 保持了细微的周期性并保持清晰的边缘过渡，而 SPC 通常会引入残留的振铃和分散的噪声。这些模式证实了我们的傅里叶引导设计在强化各种架构的细节保真度方面的作用，尽管仅增加了少量参数。

3.7. 消融研究

我们进行了一项消融研究，以量化我们 FGA 上采样器中各个组件的贡献，包括基于 MLP 的通道混合（MLP）、傅里叶特征位置编码（FF）、相关注意层（CAL）和频域 L1 损失（FL1）。通过有选择地禁用模块或将 FL1 替换为标准像素域 L1 损失，创建出不同的变体。

结果汇总于表 4。完整配置（使用 FL1 训练的 MLP + FF + CA）实现了最高性能，在 Set5 上达到了 32.30 dB 的 PSNR 和 0.8966 的 SSIM，同时获得了 0.2934 的 FRC-AUC。在 Urban100 上，它达到了 26.19 dB 的 PSNR，0.7867 的 SSIM，以及 0.2743 的 FRC-AUC。去除 FF 或 CAL 中的任何一个会导致性能显著下降：

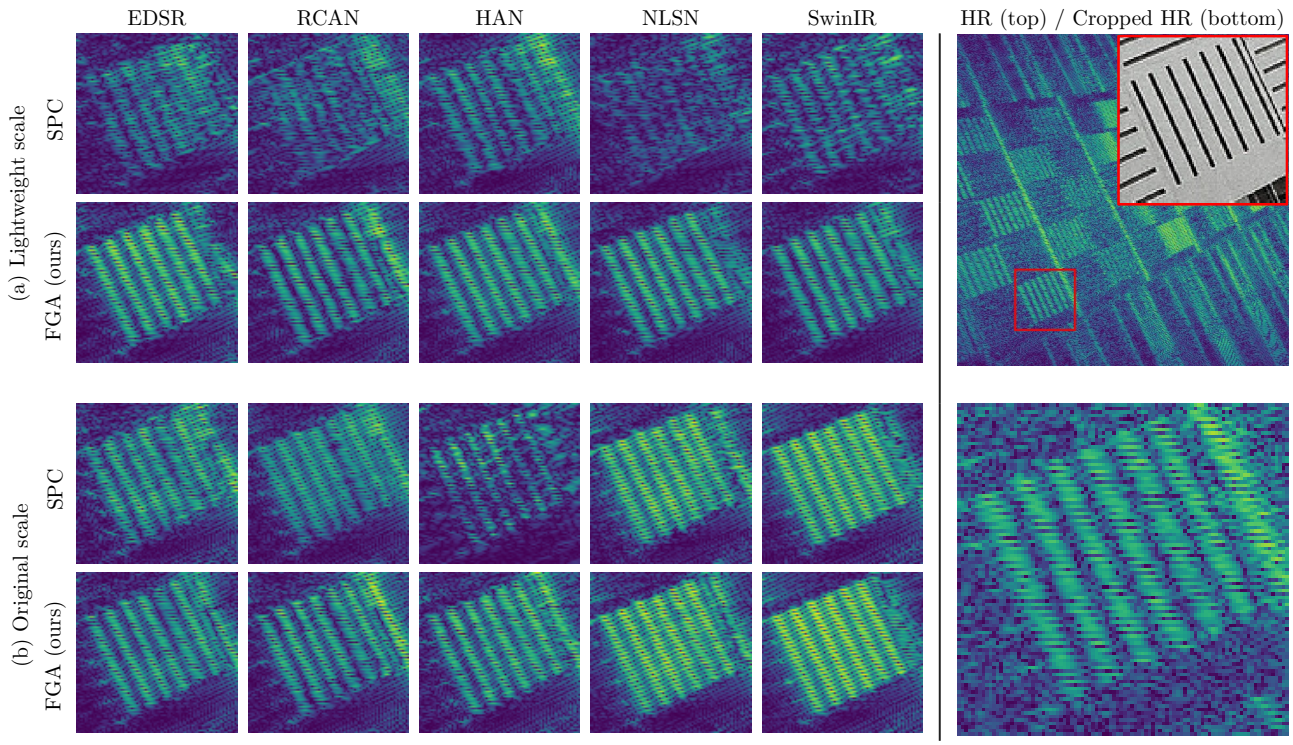


Figure 5. 从 Fourier 频谱的最高四分位部分（圆环 48–63）可视化的高频幅度图，用于 Urban100 中的 “img_092” [22] 。列展示了不同骨干网络的输出，紧随其后的是高分辨率地面实况图像和红框区域的放大视图。面板 (a) 显示了轻量模型的结果，而面板 (b) 则展示了各自论文中定义的原始规模模型。

Table 4. 使用 EDSR $^{\mu}$ 骨干网对提出的 FGA 上采样器进行消融研究。我们在 Set5 和 Urban100 上报告 PSNR、SSIM 和高频 FRC-AUC。

Module				Loss		Set5			Urban100		
Conv	MLP	FF	CAL	L1	FL1	PSNR	SSIM	FRC	PSNR	SSIM	FRC
✓	✓	✓	✓		✓	32.30	0.8966	0.2934	26.19	0.7867	0.2743
✓	✓	✓	✓	✓		32.18	0.8956	0.2792	26.09	0.7864	0.2616
✓	✓	✓		✓		32.16	0.8951	0.2806	26.04	0.7843	0.2527
✓	✓			✓		32.11	0.8948	0.2673	26.00	0.7834	0.2359
✓				✓		32.12	0.8947	0.2555	26.00	0.7833	0.2111

PSNR 下降 0.1-0.2 dB，FRC-AUC 下降 4-9 %。去除两个组件或用 L1 替代 FL1 会导致所有指标最显著的性能下降。

有趣的是，单独添加 MLP 而不添加 FF 带来的收益有限——其 PSNR (32.11 dB) 和 FRC-AUC (0.2673) 都比包含 FF 时更低——这表明仅靠通道混合无法恢复空间频率线索。总体而言，FF 提供了重要的位置感知光谱上下文，使得 CAL 和 MLP 能够有效地运作，而 FL1 进一步促进了高频保真。这些结果验证了所提组件的互补设计。

4. 结论

我们提出了一种新颖的频率引导注意力 (FGA) 上采样器用于图像超分辨率，该方法旨在通过明确的空间和频谱条件来抑制混叠并增强高频细节。通过注入傅里叶编码的子像素坐标并采用基于相关性的注意力，FGA 使得每个上采样像素能够学习与位置有关且对频率敏感的表示。

大量标准基准上的实验表明，FGA 在广泛的骨干和模型规模上连续优于传统的 SPC，在 PSNR、SSIM 和频域相似度方面取得了改善。我们进一步验证了它在轻量级和全容量网络中的泛化能力，并提供了频谱可视化，以解释 FGA 在减轻诸如棋盘格模式和频谱折叠等伪影方面的作用。消融研究强调了傅里叶特征、注意层和频域损失的互补贡献。

限制与未来工作 虽然 FGA 显著提高了重建质量，但相对于 SPC，它带来了显著的计算开销增加。如表 5 所示，根据输入分辨率的不同，FLOPs 和内存使用量增加约 2-4 × 。然而，我们发现为上采样模块分配容量比单纯扩展主干网络带来更大的性能提升。

尽管 FGA 引入的计算开销相对较小，仅占 EDSR 等大型 SR 网络总 FLOPs 的约 5 %，但减少傅里叶引导的空间调制的成本仍然是未来工作的一个重要方向。一种有前景的方法是利用硬件加速器，例如 Transformer

Engine [11]，这可以帮助减少延迟并提高吞吐量，尤其是对于实时或资源受限的应用。

Table 5. 骨干网络和上采样器在四个输入分辨率下的计算成本。

	GFLOPs				Memory (MB)			
Resolution	64^2	128^2	256^2	512^2	64^2	128^2	256^2	512^2
Backbone								
EDSR	157	628	2,513	10,052	173	245	469	1,432
RCAN	62	248	993	3,972	64	83	156	446
HAN	64	256	1,026	4,102	118	253	806	3,016
NLSN	171	682	2,729	10,917	350	911	3,168	12,180
SwinIR	49	176	681	2,701	118	245	739	2,708
Upsampler								
SPC	3	13	53	210	44	148	561	2,217
FGA	8	30	120	482	155	576	2,259	8,991

相比之下，如表格 5 所示，FGA 的内存消耗显著高于传统上采样器，如 SPC。虽然用于深度学习的现代 GPU 通常可以提供足够的内存来应对这种增加，但在资源有限的环境中，较高的内存使用仍可能成为瓶颈。因此，通过轻量化设计、注意力裁剪或优化数据流来减少内存开销，可能是未来改进的一个有价值的方向。

References

[1] Shashank Agnihotri, Julia Grabinski, and Margret Keuper. Improving feature stability during upsampling–spectral artifacts and the importance of spatial context. In European Conference on Computer Vision, pages 357–376. Springer, 2024.

[2] Marco Bevilacqua et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. pages 135–1, 2012.

[3] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 12299–12310, 2021.

[4] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 22367–22377, 2023.

[5] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 8628–8638, 2021.

[6] Zhiqin Chen. Im-net: Learning implicit fields for generative shape modeling. 2019.

[7] Zheng Chen, Yulun Zhang, Jinjin Gu, yongbing zhang, Linghe Kong, and Xin Yuan. Cross aggregation transformer for image restoration. In Advances in Neural Information Processing Systems, pages 25478–25490. Curran Associates, Inc., 2022.

[8] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xiaokang Yang, and Fisher Yu. Dual aggregation transformer for image super-resolution. In Proceedings of the IEEE/CVF international conference on computer vision, pages 12312–12321, 2023.

[9] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation, 2014.

[10] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[11] NVIDIA Corporation. Transformer Engine, 2025. Library for accelerating Transformer models on NVIDIA GPUs (FP8, FP16 mixed precision).

[12] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[13] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13, pages 184–199. Springer, 2014.

[14] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In Computer Vision – ECCV 2016, pages 391–407, Cham, 2016. Springer International Publishing.

[15] Dario Fuoli, Luc Van Gool, and Radu Timofte. Fourier space losses for efficient perceptual image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 2360–2369, 2021.

[16] Tiantong Guo, Hojjat Seyed Mousavi, Tiep Huu Vu, and Vishal Monga. Deep wavelet prediction for image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pages 104–113, 2017.

[17] Tiantong Guo, Hojjat Seyed Mousavi, and Vishal Monga. Adaptive transform domain image super-resolution via orthogonally regularized deep networks. IEEE transactions on image processing, 28(9):4685–4700, 2019.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[19] Chih-Chung Hsu, Chia-Ming Lee, and Yi-Shiuan Chou. Drct: Saving image super-resolution away from information bottleneck. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6133–6142, 2024.

[20] Xuecai Hu, Haoyuan Mu, Xiangyu Zhang, Zilei Wang, Tieniu Tan, and Jian Sun. Meta-sr: A magnification-arbitrary network for super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[21] Huaibo Huang, Ran He, Zhenan Sun, and Tieniu Tan. Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution. In Proceedings of the IEEE international conference on computer vision, pages 1689–1697, 2017.

[22] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[23] Pranav Jeevan, Akella Srinidhi, Pasunuri Prathiba, and Amit Sethi. Wavemixsr: Resource-efficient neural network for image super-resolution. In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pages 5884–5892, 2024.

[24] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[25] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[26] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 624–632, 2017.

[27] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4681–4690, 2017.

[28] Jaewon Lee and Kyong Hwan Jin. Local texture estimator for implicit representation function. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 1929–1938, 2022.

[29] Wenbo Li, Xin Lu, Shengju Qian, Jiangbo Lu, Xiangyu Zhang, and Jiaya Jia. On efficient transformer-based image pre-training for low-level vision, 2022.

[30] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, pages 1833–1844, 2021.

[31] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2017.

[32] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2017.

[33] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang. Non-local recurrent network for image restoration. In Advances in Neural Information Processing Systems. Curran Associates, Inc., 2018.

[34] Jie Liu, Wenjie Zhang, Yuting Tang, Jie Tang, and Gangshan Wu. Residual feature aggregation network for image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[35] Pengju Liu, Hongzhi Zhang, Kai Zhang, Liang Lin, and Wangmeng Zuo. Multi-level wavelet-cnn for image restoration. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pages 773–782, 2018.

[36] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pages 10012–10022, 2021.

[37] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, pages 416–423 vol.2, 2001.

[38] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. Multimedia Tools and Applications, 76(20): 21811–21838, 2017.

[39] Yiqun Mei, Yuchen Fan, and Yuqian Zhou. Image super-resolution with non-local sparse attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 3517–3526, 2021.

[40] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 4460–4470, 2019.

[41] Ben Niu, Weilei Wen, Wenqi Ren, Xiangde Zhang, Lianping Yang, Shuzhen Wang, Kaihao Zhang, Xiaochun Cao, and Haifeng Shen. Single image super-resolution via a holistic attention network. In Computer Vision – ECCV 2020, pages 191–207, Cham, 2020. Springer International Publishing.

[42] Augustus Odena, Vincent Dumoulin, and Chris Olah.

Deconvolution and checkerboard artifacts. Distill, 2016.

[43] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 165–174, 2019.

[44] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. IEEE Signal Processing Magazine, 20 (3):21–36, 2003.

[45] Peter B Rosenthal and Richard Henderson. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. Journal of molecular biology, 333(4):721–745, 2003.

[46] WO Saxton. The correlation averaging of a regularly arranged bacterial cell envelope protein. Journal of microscopy, 127(2):127–138, 1982.

[47] Wenzhe Shi, Jose Caballero, Ferenc Huszar, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[48] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pages 114–125, 2017.

[49] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao. Image super-resolution using dense skip connections. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.

[50] Marin Van Heel and Michael Schatz. Fourier shell correlation threshold criteria. Journal of structural biology, 151(3):250–262, 2005.

[51] Longguang Wang, Yingqian Wang, Zaiping Lin, Jungang Yang, Wei An, and Yulan Guo. Learning a single network for scale-arbitrary super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pages 4801–4810, 2021.

[52] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European conference on computer vision (ECCV) workshops, pages 0–0, 2018.

[53] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 17683–17693, 2022.

[54] Xin Yang, Haiyang Mei, Jiqing Zhang, Ke Xu, Baocai Yin, Qiang Zhang, and Xiaopeng Wei. Drfn: Deep recurrent fusion network for single-image super-resolution with large factors. IEEE Transactions on Multimedia, 21(2):328–337, 2018.

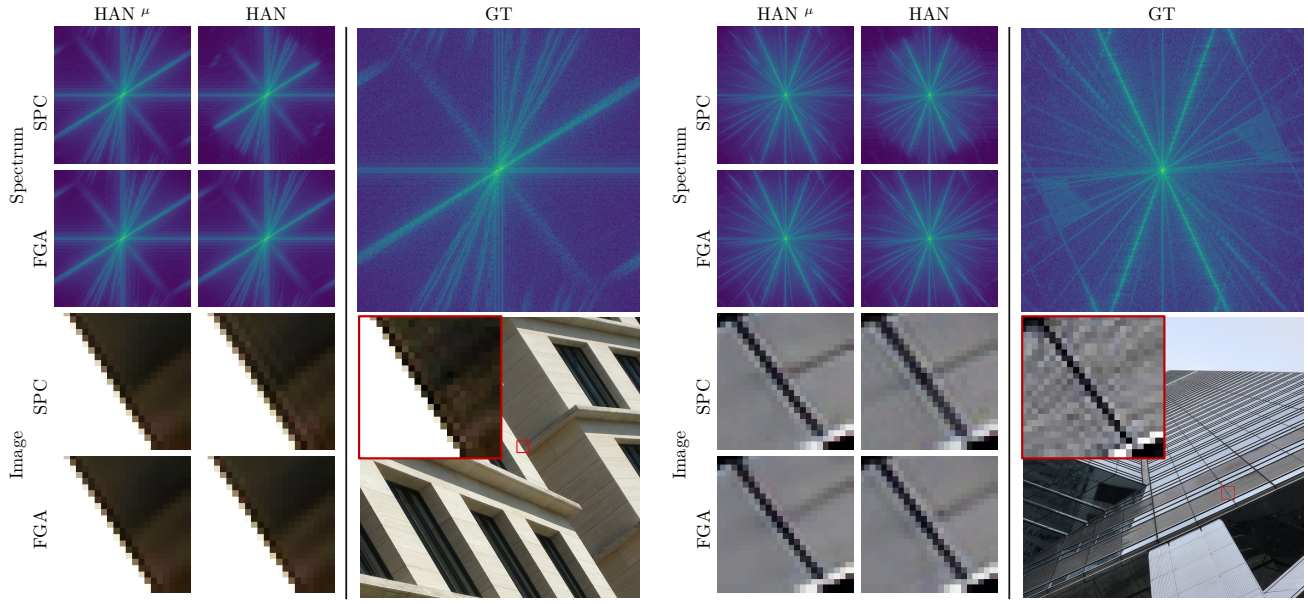
[55] Matthew D. Zeiler, Dilip Krishnan, Graham W. Taylor, and Rob Fergus. Deconvolutional networks. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 2528–2535, 2010.

[56] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In Curves and Surfaces, pages 711–730, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[57] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision (ECCV), 2018.

[58] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[59] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration, 2019.



(a) HAN restuls of img_025

(b) HAN restuls of img_059

Figure 6. 两个 Urban100 案例比较了 HAN 和 HAN $^{\mu}$ 在 SPC 和 FGA 下的表现。面板 (a) 和 (b) 分别对应 “img_025” 和 “img_059”。对于每个案例，上排显示傅里叶幅度谱，下排显示空间裁剪（右侧为真实值；红色方框表示放大区域）。对于空间裁剪，已增强对比度以使光晕伪影更加明显。

我们评估了交叉分辨率注意模块中使用的各种局部注意窗口大小，以识别在性能和平衡效率之间的最佳配置。使用 EDSR $^{\mu}$ 主干，我们在 Set5 和 Urban100 上测试了多种预上采样和后上采样窗口对，并且还对其在 512^2 分辨率输入上的 FLOPs 和内存使用情况进行了剖析（表 ?? ）。

在所有测试的设置中，$(5 \times 5,\ 4 \times 4)$ 重叠窗口配置实现了最佳的权衡，提供了改善的 PSNR/SSIM，同时计算成本仅有适度增加。因此，这种配置在主要实验中被采用为默认设置。

A. HAN 的 FRC-AUC 结果分析

虽然我们提出的 FGA 在大多数主干中一致提高了 FRC-AUC，但我们注意到在 HAN 主干中出现了一个有趣的异常。原始的 HAN 模型（从公开发布的检查点初始化）偶尔在边缘周围产生光晕或振铃伪影（图 6 ）。在频域中，这些伪影表现为非对角线或径向的脊，而在空间域中则表现为图像裁剪中的可见光晕。这样的伪影引入了过多的高频能量，偏离了真实情况，从而降低 FRC-AUC，即使 PSNR 和 SSIM 仍然很高。

相比之下，FGA 减少了多余的高频能量，使光谱更接近 GT，而从头开始训练的 HAN $^{\mu}$ 明显对这种伪影更具抵抗力。将 FGA 整合到任何一种变体中都能减少多余的高频能量，并使光谱与真实情况更加一致，表明频域关联为传统像素空间度量提供了一个补充视角。