

# 结合多尺度边缘增强和轻量化 Transformer 的图像超分辨率重建网络<sup>①</sup>



王文佩, 赵宇峰, 徐 飞, 赵莎莎

(西安工业大学 计算机科学与工程学院, 西安 710021)

通信作者: 王文佩, E-mail: [wangwenpei@st.xatu.edu.cn](mailto:wangwenpei@st.xatu.edu.cn)

**摘 要:** 现有超分辨率网络 (super-resolution, SR) 通过简单堆叠多分支结构来捕捉多尺度特征, 导致网络推理速度缓慢, 且无法有效建模全局像素关联. 部分研究引入 Transformer 的自注意力机制来提升重建质量, 却导致计算复杂度大幅上升. 针对这些问题, 本文提出了一种结合多尺度边缘增强和轻量化 Transformer 的 SR 网络 (ECTL-SR). 具体而言, 提出一种轻量边缘导向卷积块有效捕捉并融合不同感受野下边缘细粒度特征, 同时引入结构重参数化技术来减少多分支冗余计算和内存开销. 此外, 将轻量型位置感知环形卷积嵌入改进的 Transformer 架构中来增强网络捕捉图像长距离依赖能力, 在低成本下实现感受野高效扩展. 实验结果表明, 该网络在性能与效率之间取得了良好平衡, 并在 Urban100 等多个基准数据集上优于现有 SR 方法, 展现出更优的重建效果.

**关键词:** 图像超分辨率; 边缘增强; 轻量化 Transformer; 结构重参数化; 位置感知环形卷积

引用格式: 王文佩, 赵宇峰, 徐飞, 赵莎莎. 结合多尺度边缘增强和轻量化 Transformer 的图像超分辨率重建网络. 计算机系统应用, 2026, 35(1): 117-128. <http://www.c-s-a.org.cn/1003-3254/10051.html>

## Image Super-resolution Reconstruction Network Combining Multi-scale Edge Enhancement and Lightweight Transformer

WANG Wen-Pei, ZHAO Yu-Feng, XU Fei, ZHAO Sha-Sha

(School of Computer Science and Engineering, Xi'an Technological University, Xi'an 710021, China)

**Abstract:** Existing super-resolution (SR) networks typically capture multi-scale features by stacking multi-branch structures, leading to slow inference and limited modeling of global pixel dependencies. Some studies introduce Transformer-based self-attention to enhance reconstruction quality but at the cost of increased complexity. To address these challenges, this study proposes a novel SR network that combines multi-scale edge enhancement with a lightweight Transformer (ECTL-SR). A lightweight edge-guided convolutional block effectively captures and fuses fine-grained edge features under different receptive fields, while structural re-parameterization reduces redundant computation and memory overhead. A lightweight position-aware circular convolution is embedded into a modified Transformer to boost the network's ability to model long-range dependencies and efficiently expand the receptive field at low cost. Experiments show that the proposed network achieves a good balance between performance and efficiency, outperforming existing SR methods on benchmark datasets such as Urban100 with superior reconstruction results.

**Key words:** image super-resolution; edge enhancement; lightweight Transformer; structural reparameterization; position-aware circular convolution

① 基金项目: 陕西省科技厅区域创新能力指导计划 (20122qfy01-14); 咸阳市科技计划重点研发项目 (2021ZDYF-NY-0019)

收稿时间: 2025-06-12; 修改时间: 2025-07-03, 2025-08-01; 采用时间: 2025-08-15; csa 在线出版时间: 2025-12-01

CNKI 网络首发时间: 2025-12-02

超分辨率技术通过恢复图像高频细节,从而增强下游任务的表现.其中,单图像超分辨率(single image super-resolution, SISR)作为计算机视觉的重要研究方向,在视频监控、医学影像分析和卫星遥感等领域具有广泛应用.卷积神经网络(CNN)凭借其局部连接性和平移不变性,在SISR任务中展现了卓越的性能.Dong等人<sup>[1]</sup>提出的SRCNN首次将CNN引入超分辨率任务,显著提升了传统方法的性能.EDSR<sup>[2]</sup>引入残差学习并去除批归一化,进一步提高了超分辨率的重建效果.随后RDN<sup>[3]</sup>通过引入残差密集块扩展网络深度至百层以上,RCAN<sup>[4]</sup>在此基础上结合残差结构与通道注意力机制,加深网络结构至400层以上,取得了当时的最先进性能.这些模型主要依赖堆叠具有单一固定感受野的卷积层来提升重建性能,无法捕捉多尺度特征,增加了计算复杂度.

近来研究逐渐聚焦于设计更高效的轻量级超分辨率网络,旨在显著降低计算开销和模型规模的同时,最大限度地保持重建质量.IMDN<sup>[5]</sup>提出了信息多重蒸馏机制,并结合对比度感知通道注意力(CCA),初步实现性能与效率平衡.RFDN<sup>[6]</sup>基于此提出了特征蒸馏连接(FDC)并引入了无额外参数开销的浅层残差块,保持优越性能并提升了模型运行效率.轻量SR受限于局部感受野难以捕捉图像的全局长程依赖关系,SISR领域引入ViT(vision Transformer),其采用全连接的自注意力机制,可直接建立跨任意距离像素的全局关联.为了兼顾局部细节捕捉与全局依赖建模,研究者将CNN与Transformer相结合,以提升超分辨率任务的整体性能与推理效率.SwinIR<sup>[7]</sup>基于swin Transformer架构采用动态窗口划分策略,通过局部自注意力在保持建模能力的同时大幅降低计算复杂度.LBNet<sup>[8]</sup>通过设计双路径特征蒸馏机制,结合可变形卷积与通道稀疏注意力模块,高效建模图像全局特征.然而自注意力机制需要计算每个像素对之间的相关性,导致计算开销呈二次方增长,忽视了推理速度对超分辨率重建的重要性.

为此,本文提出结合多尺度边缘增强和Transformer架构的轻量级超分辨率网络ECTL-SR(edge convolution and Transformer lightweight super-resolution),其中多尺度边缘特征增强(multi-scale edge-enhanced, MSEE)模块采用可参数化的多分支并行卷积核来提取多尺度边缘与局部细节纹理特征,并结合局部注意力机制实现跨粒度特征的自适应融合.本文还基于Trans-

former架构设计了轻量级全局特征提取模块(global feature extraction block, GFEB),引入轻量级全局位置感知环形(global position-aware circular, GCPA)卷积模块以替代计算成本高昂的自注意力机制进行全局特征建模,结合轻量通道注意力机制(ECA)<sup>[9]</sup>,自适应学习通道间的依赖关系,对经过全局建模与非线性变换后的特征进行细粒度增强,提升特征表达能力与重建质量.

本文的主要贡献如下.

(1)设计了一种显式建模图像的局部、边缘和全局信息的轻量超分辨率混合模型,在保持较低计算成本的前提下实现了优越的重建性能.

(2)提出一种高效的多尺度边缘特征提取与融合方法,通过可参数化的多分支卷积与局部注意力机制,实现不同感受野下纹理细节的充分捕捉与融合.

(3)为降低全局像素建模成本,在超分任务中引入位置感知环形卷积,实现对全局与区域特征的有效整合.

(4)在多个主流基准数据集和不同放大倍率的实验中,验证了所提模型在重建质量与推理效率之间的良好平衡,并取得了优于现有方法的性能表现.

## 1 相关工作

主流超分辨率方法通过多层卷积操作实现低分辨率(low resolution, LR)到高分辨率(high resolution, HR)的复杂非线性建模,FSRCNN<sup>[10]</sup>通过去除传统插值上采样操作并引入转置卷积来提升图像重建效果.Kim等人<sup>[11]</sup>随后提出VDSR,使用20个权重层在网络中多次级联小滤波器,通过大感受野捕获图像上下文信息.但深度卷积模型在追求高性能的同时,以高昂的内存消耗和较低的推理速度为代价.为了提高模型的实用性和部署可行性,研究者提出了IDN<sup>[12]</sup>,通过信息蒸馏模块和组卷积技术显著减少了模型参数和计算量.IMDN<sup>[5]</sup>在其基础上引入了逐级特征蒸馏和选择性核融合机制,进一步实现了轻量化.此外,很多网络通过参数共享策略来减少模型的参数量,DRCN<sup>[13]</sup>通过递归结构实现多层次特征提取,之后DRRN<sup>[14]</sup>结合残差学习机制,进一步提升了图像细节恢复能力.然而这些轻量级SR在计算与内存资源受限的边缘设备上,仍难以满足实时推理需求.ECBS<sup>[15]</sup>将重参数化技术应用于移动端的实时超分辨率任务,显著提升了模型的推理速度和计算效率.受此启发,本文提出一种可重参数化边缘增强卷积块(RepECB),在训练阶段引入边缘多分支

结构,融合不同感受野下的局部细粒度特征,以增强对图像边缘结构与局部高频细节的表征能力。

近年来研究者将 Transformer 的全局自注意力机制与 CNN 有效融合,大幅提升了超分辨率任务中跨像素长程依赖的建模能力与重建表现。代表性的 SwinIR<sup>[7]</sup>采用局部注意力机制结合 CNN 实现了卓越的重建质量。DAT<sup>[16]</sup>采用双聚合架构,交替采用空间窗口自注意力和通道自注意力机制,缓解了通道冗余问题。为缓解自注意力机制带来的高计算开销,ELAN<sup>[17]</sup>在非重叠特征组上采用多尺度窗口进行自注意力计算,并引入共享注意机制,构建了高效的远程注意块。SRFormer<sup>[18]</sup>重新设计了注意力结构,提出置换自注意力机制(PSA),将空间信息转移至通道维度,在显著降低计算复杂度的同时,性能超越 SwinIR。CAMixerSR<sup>[19]</sup>则依据图像纹理复杂度动态分配特征提取方式,在稀疏区域融合平滑窗口注意力与卷积注意机制,在简单区域采用高效卷积操作实现性能与效率的平衡。尽管上述方法在降低自注意力计算开销方面取得了显著进展,但其在高分辨率图像处理时仍然计算密集,导致推理延迟较高,从而制约了在资源受限设备上的高效部署与应用。

## 2 模型设计

本文提出的轻量级超分辨率网络通过浅层、局部与全局分层特征提取,融合多尺度信息,旨在低计算开销下实现边缘与高频纹理高性能重建。浅层特征通过跨阶段残差将低频信息直接传递至重建层,抑制上采样失真;局部特征引入可重参数化的边缘卷积算子来高效增强多尺度纹理边缘敏感性,结合轻量注意力机制动态校准权重,实现局部细节与边缘信息的自适应融合;深层特征提取模块将全局空间依赖分解为垂直和水平方向的全局环形卷积,并嵌入位置编码,在降低计算复杂度的同时保持位置敏感性,有效建立远距离

像素间的结构关联。

### 2.1 整体架构

所提出的 ECTL-SR 网络如图 1 所示,在浅层特征提取阶段,将输入低分辨率图像  $I_{LR} \in R^{H \times W \times 3}$  ( $H, W$  是图像的高和宽)在通道维度上复制 4 次,随后通过卷积层将输入图像空间映射到多维特征空间,生成初始特征  $F_0$ :

$$F_0 = M_F(\text{Concat}(I_{LR}^1, \dots, I_{LR}^4)) \quad (1)$$

其中,  $F_0 \in R^{H \times W \times C}$  表示提取的浅层特征,  $\text{Concat}(\cdot)$  表示通道维度的拼接操作,  $M_F(\cdot)$  是初步特征提取模块,即一个  $3 \times 3$  的卷积操作。

网络主体包含局部特征提取和全局特征提取两部分。局部特征提取由  $n$  个级联的 MSEE 模块构成,将获得的特征  $F_0$  传入模块细化特征,第  $n$  个 MSEE 的输出可以表示为:

$$F_m^n = \begin{cases} M_{LF}(F_0), & n = 1 \\ M_{LF}(F_m^{n-1}), & 1 < n \leq N \end{cases} \quad (2)$$

其中,  $F_m^n$  表示第  $n$  个 MSEE 的输出特征,  $M_{LF}(\cdot)$  为多尺度边缘强化块代表的函数映射,  $N$  表示串联的数量。所有 MSEE 模块的输出经过拼接实现融合,该模块输出可表示为:

$$F_{LF} = [F_m^1, F_m^2, \dots, F_m^N] \quad (3)$$

其中,  $F_{LF}$  表示通道拼接后的局部融合特征,这些特征随后经过  $N$  个级联的全局特征增强块 (GFEB) 建立远程依赖关系,从而将局部细节和全局结构相结合,该阶段的特征提取过程可表示为:

$$F_g^n = \begin{cases} M_{GF}(F_{LF}), & n = 1 \\ M_{GF}(F_g^{n-1}), & 1 < n \leq N \end{cases} \quad (4)$$

其中,  $F_g^n \in R^{H \times W \times C}$  为第  $n$  个 GFEB 的输出特征,  $M_{GF}(\cdot)$  则为全局特征提取块映射函数。

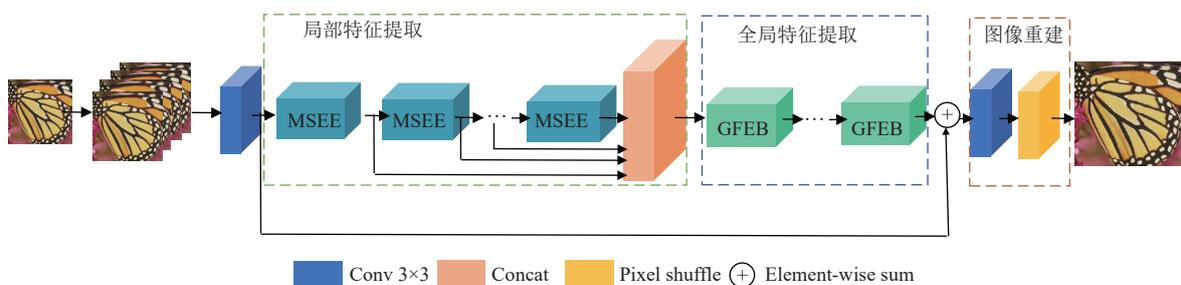


图 1 ECTL-SR 架构图

最后通过上采样重建模块生成高分辨率图像  $I_{SR}$  并利用全局残差连接保留  $F_0$  中的低频信息, 重建过程可表示为:

$$I_{SR} = H_{up}(F_g^n + F_0) \quad (5)$$

其中,  $H_{up}(\cdot)$  表示重建模块的操作, 即一层  $3 \times 3$  卷积和像素重排 (pixel shuffle)<sup>[20]</sup> 操作实现空间维度上采样.

### 2.2 多尺度边缘增强模块

现有方法通常依赖不同大小的固定卷积核提取局部多感受野信息, 边缘信息随着超分辨率网络的加深而退化, 导致重建图像出现模糊和锯齿伪影, 同时多次卷积操作限制了推理速度. 为此, 本文提出了一种基于结构重参数化的多尺度边缘特征增强模块 MSEE. 如图 2 所示, MSEE 模块主要由 3 个串联的边缘导向卷积块 (reparameterized edge-oriented convolution block, RepECB) 和 1 个轻量注意力模块 (lightweight attention module, LAM) 组成. RepECB 在训练阶段采用可重参数化的多分支卷积并行架构, 融合边缘检测算子<sup>[21]</sup>, 自适应捕捉图像不同尺度的局部细节、方向性纹理和高频边缘信息. 以第  $n$  个 MSEE 的第  $m$  个 RepECB 为例, 假设输入为  $F_{n,m-1}^e \in R^{H \times W \times C}$ , 训练时输入特征先通过 1 个标准  $3 \times 3$  卷积层进行基础特征处理, 该模块不

使用批归一化 (BN) 层, 以避免其平滑效应削弱对像素级对比度敏感的边缘信息<sup>[2]</sup>. 该分支的输出可以表示为:

$$F_n = \mathcal{K}_n * F_{n,m-1}^e + B_n \quad (6)$$

其中,  $F_n$  表示该分支的输出特征,  $\mathcal{K}_n$  表示  $3 \times 3$  卷积核的权重参数,  $B_n$  是偏置项,  $*$  表示卷积操作. 随后引入非对称卷积, 通过可学习的滤波器参数来捕捉水平方向和垂直方向细粒度特征:

$$F_{vert} = \sum_{i=0}^0 \sum_{j=-1}^1 \mathcal{K}_{vert}(i+1, j+1) \cdot F_{n,m-1}^e(x+i, y+j, c) + B_{vert} \quad (7)$$

其中,  $F_{vert}$  是经垂直方向卷积后的输出特征图,  $B_{vert}$  为对应的可学习偏置项,  $\mathcal{K}_{vert}$  则表示滤波器权重,  $F_{n,m-1}^e(\cdot)$  表示输入特征图在指定位置的空间像素值, 水平方向与之类似. 其次采用级联膨胀卷积分支, 通过分阶段特征变换实现特征的高效融合, 增强模型在复杂纹理区域中的建模能力.

$$F_{es} = \mathcal{K}_s * (\mathcal{K}_e * F_{n,m-1}^e + B_e) + B_s \quad (8)$$

其中,  $\mathcal{K}_e$  表示扩展滤波器权重,  $\mathcal{K}_s$  表示压缩滤波器权重,  $B_e$ 、 $B_s$  表示对应阶段的偏置项.

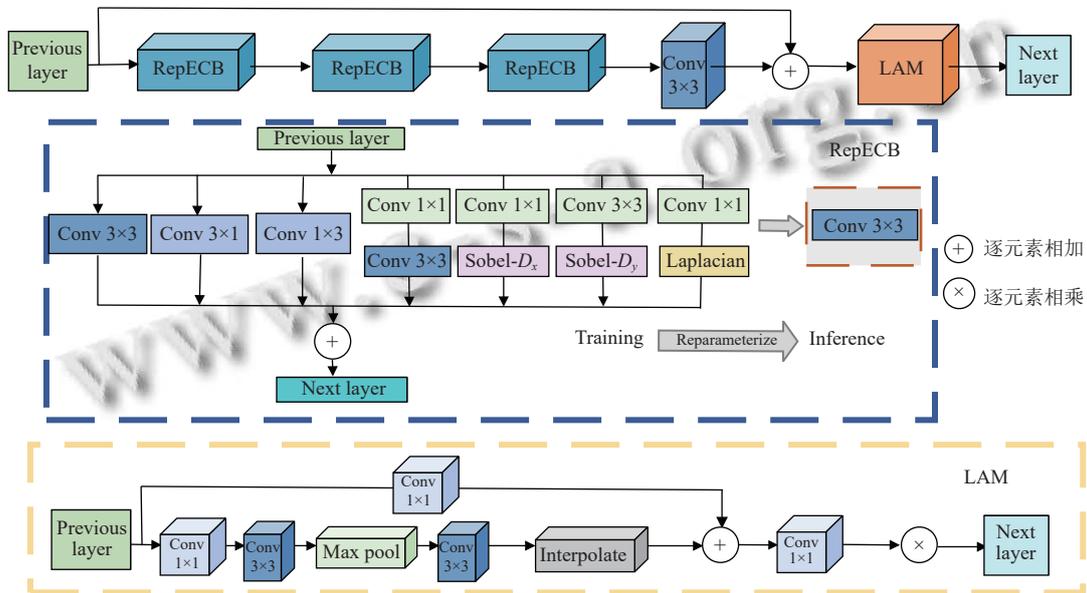


图 2 MSEE 模块图

为避免多尺度特征融合过程中局部细节的丢失, 训练阶段采用预定义的多路边缘滤波器分支, 其中的 Sobel 滤波器分支利用水平  $D_x$  与竖直方向  $D_y$  一阶差

分滤波器提取横向与纵向边缘特征, 并结合缩放因子自适应调整信息贡献度, 从而增强对细粒度纹理的敏感性.

$$\begin{cases} F_{D_x} = (S_{D_x} \cdot D_x) \otimes (\mathcal{K}_x * F_{n,m-1} + B_x) + B_{D_x} \\ F_{D_y} = (S_{D_y} \cdot D_y) \otimes (\mathcal{K}_y * F_{n,m-1} + B_y) + B_{D_y} \end{cases} \quad (9)$$

其中,  $F_{n,m-1}$  是输入特征,  $S_{D_x}$ 、 $S_{D_y}$  是通道缩放因子,  $\mathcal{K}_x$ 、 $\mathcal{K}_y$  和  $B_x$ 、 $B_y$  分别是  $1 \times 1$  滤波器和 DWConv 滤波器学习可训练的权重参数和偏置项;  $\otimes$  和  $*$  分别表示 DW-Conv 与常规卷积,  $\cdot$  表示通道级广播乘法.  $B_{D_x}$  和  $B_{D_y}$  是输出偏置项.

最后利用 Laplacian 分支的拉普拉斯滤波器  $D_{lap}$  计算邻域像素的二阶差分来捕捉纹理锐度和细小边缘等高频信息.

$$F_{lap} = (S_{lap} \cdot D_{lap}) \otimes (\mathcal{K}_{lap} * F_{n,m-1}^e + B_l) + B_{lap} \quad (10)$$

其中,  $S_{lap}$  和  $B_{lap}$  分别为 DWConv<sup>[22]</sup> 的缩放因子与偏置,  $\mathcal{K}_{lap}$  和  $B_l$  是  $1 \times 1$  卷积的权重和偏置. 训练阶段所有上述分支的输出通过深度卷积层和标准卷积进行特征融合, 最终输出为  $F_{n,m}^e \in R^{H \times W \times C}$  并传递至下一层.

为解决多分支结构带来的推理冗余与部署复杂性, RepECB 引入结构重参数化, 利用卷积的线性可加性, 在推理时将所有并行分支的参数等效转换为单一的  $3 \times 3$  卷积, 从而在不影响模型性能的情况下, 实现高效推理. 参数化过程如图 3 所示.

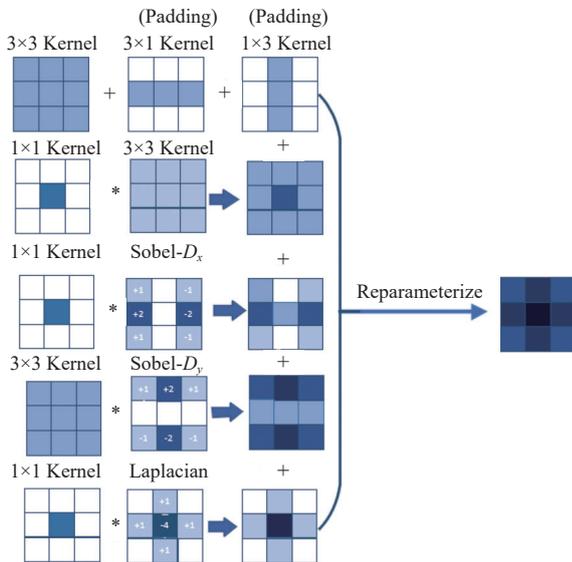


图 3 RepECB 重参数化

根据文献[23]将  $1 \times 1$  卷积、非对称卷积通过零填充与核相加操作实现结构对齐. 对于由  $1 \times 1$  扩展卷积与  $3 \times 3$  压缩卷积串联构成的级联分支, 其参数也可以在数学上等效地合并为单一正态卷积.

$$\begin{cases} K_{es} = perm(K_e) * K_s \\ B_{es} = K_s * rep(B_e) + B_s \end{cases} \quad (11)$$

其中,  $K_{es}$  是生成的标准卷积核,  $B_{es}$  是标准卷积的偏置项,  $perm(\cdot)$  是维度置换操作, 将扩展滤波器与输出通道维度匹配,  $rep(\cdot)$  是将偏置扩展空间广播操作.

边缘滤波器分支的 DWConv 在每个通道上独立运算, 所以需构建一个稀疏的标准卷积核  $K_{D_x}$  来等效进行重参数化. 一个标准卷积在计算第  $i$  个输出通道时, 会聚合所有输入通道的信息; 而 DWConv 在计算第  $i$  个输出通道时, 仅处理来自第  $i$  个输入通道的信息. 权重张量必须满足输入通道索引  $j$  与输出通道索引  $i$  不相等时, 它们之间的连接权重必须为 0, 这使得该等效核在通道维度上呈现出对角的稀疏结构. 对于输入通道  $j$  和输出通道  $i$  的空间滤波器, 等效卷积核定义如下:

$$K_{D_x}[i, j, :, :] = \begin{cases} (S_{D_x} \cdot D_x)[i, 1, :, :], & \text{if } i = j \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

$K_{D_y}$  和  $K_{lap}$  同理可得. 最终将所有模块参数整合重排, 合并各分支输出得到单一标准卷积层输出项. 随后引入轻量注意力模块 LAM 自适应校准权重, 增强关键区域的响应. 输入特征通过标准滤波降低计算量:

$$F_l = M_{Lam}(F_{n,3}^e) \quad (13)$$

其中,  $F_l \in R^{H \times W \times \frac{C}{2}}$  表示压缩后的特征,  $M_{Lam}(\cdot)$  表示通道压缩. 随后结合双路径提升多尺度边缘信息的融合效果: 细节路径使用  $F_{detail} = M_{Lam}(F_l)$  保持原始分辨率下的局部细节信息; 全局路径通过级联操作提取全局结构信息.

$$F_{con} = F_{Interp}(F_e(F_{MP}(F_l))) \quad (14)$$

其中,  $F_{MP}(\cdot)$  表示对输入特征的池化操作,  $F_e(\cdot)$  表示滤波器提取局部空间信息操作,  $F_{Interp}(\cdot)$  表示双线性插值操作. 最后经  $1 \times 1$  卷积调整并通过 Sigmoid 函数生成空间注意力图  $F_A$ .

$$\begin{cases} F_A = \sigma(\text{Conv}_{1 \times 1}(F_{detail} + F_{con})) \in [0, 1]^{H \times W \times C} \\ F'_A = X \odot F_A \end{cases} \quad (15)$$

其中,  $F'_A$  是注意力加权特征,  $\odot$  表示逐元素乘法,  $\sigma(\cdot)$  是 Sigmoid 函数. 在模块中采用残差连接结构避免梯度消失, 增强训练稳定性.

$$F'_l = F_l + F'_A \quad (16)$$

## 2.3 轻量化 Transformer 模块

Vision Transformer (ViT) 的成功很大程度上归功

于其自注意力令牌混合机制. 本文提出一种全局循环位置感知模块 GCPA, 以具有线性复杂度的环形卷积替代了计算成本高达二次复杂度的自注意力机制, 其采用正交的环形卷积结合动态生成位置感知矩阵, 实现全局空间位置感知与动态权重分配. 研究表明<sup>[24]</sup>, 在 MetaFormer 架构中采用池化或 MLP 等简单操作作为令牌混合器, 也能获得与自注意力模块可比的性能. 受此启发, 将 GCPA 嵌入 MetaFormer 架构替代计算复杂的自注意力来提取全局特征, 并在前馈网络 (FFN) 后融合高效通道注意力机制 (ECA) 自适应校准通道间特

征依赖, 保留局部位置敏感性, 构建轻量特征提取模块 GFEB, 如图 4 所示. 传统局部卷积核受限于固定滑动窗口, 难以有效捕捉空间长程依赖. 为了突破此局限, GCPA 采用水平与垂直方向的环形卷积高效扩展了全局感受野. 当一个标准的局部卷积核移动到原始特征图的边界时, 环形卷积的感受野会包含另一端的像素, 从而以计算成本很低的局部卷积操作实现了对整个输入序列信息的高效聚合. 其次, 为了补偿环形卷积所造成的位置信息模糊, 引入了动态生成的位置编码, 以保持和注入关键的空间信息.

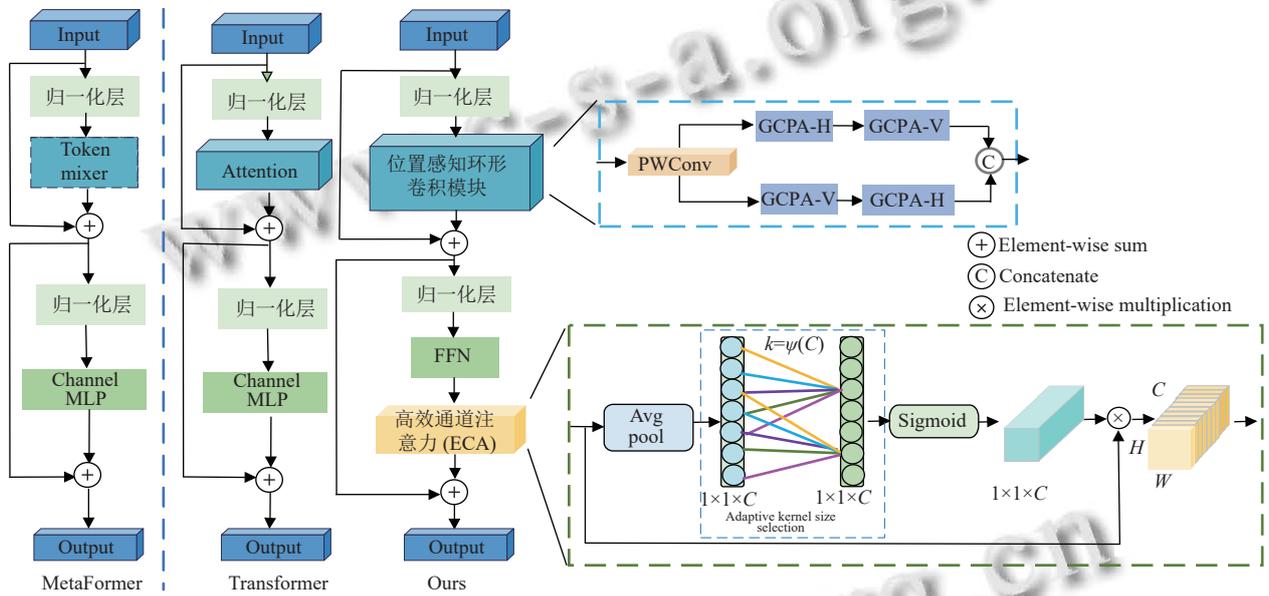


图 4 MetaFormer 与 Transformer 架构对比及 GFEB 模块结构

具体实现依赖于水平方向和垂直方向上的位置感知卷积, 分别在空间域中覆盖同一行和同一列的所有像素, 捕获不同方向的特征, 同时环形卷积采用了循环边界条件机制, 通过将特征图的边界像素与相对侧的像素连接起来, 当卷积核滑动到特征图的边界时, 它不会受到传统卷积中的零填充影响, 而是循环到另一端继续卷积操作, 从而避免了边界信息的丢失.

如图 5 所示, 以 GCPA-V 为例说明其实现过程, 输入特征图  $x \in R^{H \times W \times C}$  通过双线性插值函数  $F(\cdot)$  调整位置嵌入大小, 生成对应一维垂直位置编码向量  $pe^v$ . 之后将  $pe^v$  的每一个行向量  $pe_i^v$  通过扩展函数  $EV(\cdot)$  沿水平方向复制  $w$  次 ( $w$  是特征图的宽度) 并堆叠起来, 生成一个  $h \times w$  大小的位置编码矩阵  $PE$ , 该矩阵的每一行都包含了对应垂直位置  $i$  的编码信息, 并覆盖了整个特

征图宽度.

$$pe^v = F(\tilde{pe}^v) = [pe_0^v, pe_1^v, \dots, pe_{h-1}^v]^T \quad (17)$$

$$pe_e^v = EV(pe^v, w) \quad (18)$$

为了使卷积核能够根据高度方向的不同位置生成动态、位置敏感的特征表达, 同样通过插值函数  $F(\cdot)$  生成动态的卷积核  $k_v$ , 最后将输入特征  $x$  与生成的位置信息  $pe_e^v$  相结合, 确保了卷积操作能够与位置编码同步适配, 形成位置感知的输入特征  $x_p$ .

$$k_v = F(\tilde{k}) = [k_0^v, k_1^v, \dots, k_{h-1}^v] \quad (19)$$

$$x_p = x + pe_e^v \quad (20)$$

GCPA-H 的设计与实现与 GCPA-V 完全对称, 仅将操作方向转为水平, 并生成对应的行位置编码和动态卷积核, 以捕捉行内的全局上下文信息. 最后结合这

两种方向提取的全局特征送入前馈网络 (FFN) 层, 通过轻量高效的 ECA 通道注意力机制<sup>[9]</sup> 来自适应地捕

获通道间的特征依赖关系, 实现局部-全局感知特征的高效融合。

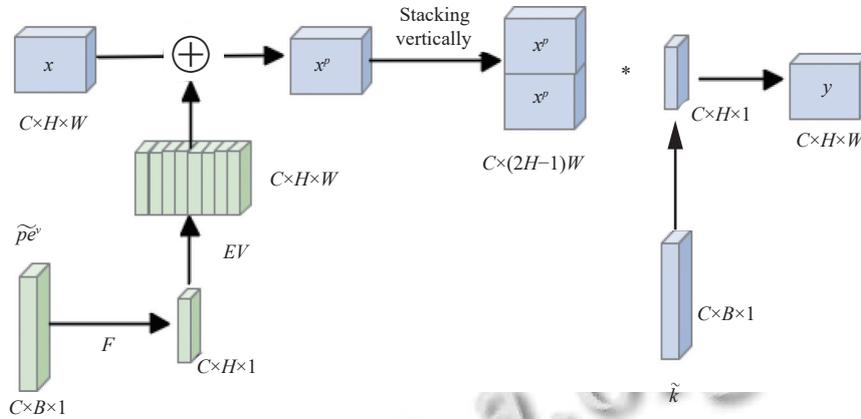


图5 GCPA-V 流程

### 3 实验

#### 3.1 数据集与评估指标

为实现公平对比, 本文采用两种训练设置: 仅使用 DIV2K 数据集 (800 张图像) 训练, 及使用 DF2K (即 DIV2K 与 Flickr2K<sup>[25]</sup> 共计 3 450 张图像) 训练, 采用 DF2K 训练的模型在文中用  $\uparrow$  符号标记. 为了验证模型性能并与当前先进方法进行对比, 在 5 个标准基准数据集 Set5<sup>[26]</sup>、Set14<sup>[27]</sup>、BSD100<sup>[28]</sup>、Urban100<sup>[29]</sup> 和 Manga109<sup>[30]</sup> 上进行了测试评估.

评估指标包括图像重建质量和计算复杂度这两个方面. 本文使用峰值信噪比 (PSNR) 和结构相似性 (SSIM)<sup>[31]</sup> 来定量评估重建质量. 这两个指标均基于 YCbCr 颜色空间中的亮度通道 (Y) 进行计算, 更高的 PSNR 和 SSIM 值表明更好的重建效果. 为了全面衡量模型效率, 评估模型的参数数量 (Params) 和乘加运算次数 (Multi-Adds). Multi-Adds 计算基于生成 1280 $\times$ 720 分辨率 RGB 高分辨率图像所需的运算量. 同时引入实际推理延迟 (latency) 作为实用性评估指标, 以直接反映模型在硬件上的运行速度, latency 越低表明推理速度越快.

#### 3.2 实验设置

在训练过程中, 本文均使用标准的双三次 (bicubic) 插值方法对原始高分辨率图像进行降采样, 以生成对应的低分辨率图像. 随后从 LR 图像中随机裁剪出 64 $\times$ 64 大小的图像块作为网络输入, 并采用随机数据增强策略, 包括 90 $^\circ$ 、180 $^\circ$ 、270 $^\circ$  旋转及水平翻转. 模型训练采用 Adam 优化器<sup>[32]</sup> 来最小化  $L_1$  损失, 参数设置为

$\beta_1=0.9$ ,  $\beta_2=0.999$ ,  $\varepsilon=1E-8$ . 总训练周期为 1200 个 epoch, 批处理大小设置为 16. 初始学习率为 5E-4, 并在每 200 轮后减半, 1000 轮后保持固定不变. 本文提出的 ECTL-SR 模型主要结构参数设置如下: 包含了 4 个 MSEE 模块, 其中每个 MSEE 模块由 3 个 RepECB 模块串联构成. 全局特征提取模块 (GFEB) 的数量设置为 2. 模型实现基于 PyTorch 1.7 框架<sup>[33]</sup>, 为了确保推理延迟基准测试的一致性和可复现性, 所有测量均在单线程环境下进行.

#### 3.3 对比实验

##### 3.3.1 重建性能定量比较

本文在 5 个公开的基准测试数据集与多个先进的 SISR 方法进行了定量比较, 包括经典 CNN 基线模型如 EDSR<sup>[2]</sup>、IMDN<sup>[5]</sup>、RLFN<sup>[34]</sup> 以及最新轻量混合型模型 SwinIR<sup>[7]</sup>、SRFormer<sup>[18]</sup>、ELAN<sup>[17]</sup> 和 ESRT<sup>[35]</sup>, 如表 1 所示.

在 Urban100 的 2 倍任务上, ECTL-SR 与高性能重量级 SwinIR 模型相比, PSNR 比 SwinIR 高出 0.48 dB, 同时参数量相对减少了约 30.9%, 计算量更是降低为 SwinIR 的 1/6. 这表明 ECTL-SR 在大幅降低模型复杂度的同时, 甚至实现了性能反超.

当使用更大规模的 DF2K 训练集时, 模型在所有测试集上的性能均取得了一致且显著的提升, 其中 Urban100 数据集包含大量相似且具有规律性的图像块, 模型所引入的长距离依赖建模机制, 能够有效捕捉并利用这些全局的、重复的结构信息, 从而在细节恢复上展现出比其他方法更显著的优势.

表1 不同 SISR 方法在×2、×3、×4 尺度下于 5 个基准数据集上的定量比较

Scale	Method	Params (k)	Multi-Adds (G)	PSNR (dB)/SSIM				
				Set5	Set14	BSD100	Urban100	Manga109
×2	EDSR	1363	356.9	37.99/0.9604	33.57/0.9175	32.16/0.8994	31.98/0.9272	38.54/0.9769
	IMDN	695	54.2	38.00/0.9605	33.63/0.9177	32.19/0.8996	32.17/0.9283	38.88/0.9774
	SwinIR	896	210.7	38.14/0.9611	33.86/0.9206	32.31/0.9012	32.76/0.9340	39.12/0.9783
	SRFormer	734	91.6	<u>38.23/0.9613</u>	<u>33.94/0.9209</u>	<u>32.36/0.9019</u>	<u>32.91/0.9353</u>	<u>39.28/0.9785</u>
	RLFN	643	89.6	38.07/0.9607	33.82/0.9188	32.23/0.9124	32.28/0.9288	—/—
	ELAN	632	173.8	38.11/0.9610	33.72/0.9190	32.30/0.8962	32.76/0.9340	39.11/0.9779
	ESRT	751	53.8	38.21/0.9602	32.94/0.9163	32.08/0.9103	32.19/0.9176	39.13/0.9760
	ECTL-SR	619	34.5	<b>38.31/0.9616</b>	<b>34.13/0.9226</b>	<b>33.16/0.9387</b>	<b>33.24/0.9379</b>	<b>39.45/0.9781</b>
	ECTL-SR↑	619	34.5	<b>38.54/0.9673</b>	<b>34.28/0.9283</b>	<b>33.35/0.9426</b>	<b>34.67/0.9452</b>	<b>39.92/0.9864</b>
×3	EDSR	1526	343.7	34.37/0.9270	30.28/0.8417	29.07/0.8052	28.15/0.8527	33.45/0.9439
	IMDN	703	52.4	34.36/0.8947	30.32/0.8563	29.09/0.8047	28.17/0.8519	33.61/0.9445
	SwinIR	892	196.7	34.62/0.9289	30.54/0.8463	29.20/0.8082	28.66/0.8624	33.98/0.9478
	SRFormer	778	83.7	<u>34.67/0.9296</u>	<u>30.57/0.8466</u>	29.26/0.8173	28.81/0.8655	<u>34.19/0.9489</u>
	RLFN	643	87.3	34.52/0.9227	30.41/0.8348	29.34/0.8237	28.57/0.8624	—/—
	ELAN	625	168.3	34.61/0.9288	30.55/0.8463	29.21/0.8081	28.62/0.8642	34.02/0.9478
	ESRT	751	51.2	33.97/0.9176	29.52/0.8362	<u>29.65/0.8079</u>	<u>28.89/0.8675</u>	33.67/0.9483
	ECTL-SR	637	27.8	<b>34.76/0.9302</b>	<b>30.70/0.8504</b>	<b>29.88/0.8210</b>	<b>29.31/0.8711</b>	<b>34.56/0.9504</b>
	ECTL-SR↑	637	27.8	<b>34.82/0.9425</b>	<b>32.14/0.8627</b>	<b>30.56/0.8374</b>	<b>32.01/0.8972</b>	<b>35.94/0.9586</b>
×4	EDSR	1518	321.2	32.09/0.8938	28.58/0.7813	27.57/0.7357	26.04/0.7838	30.45/0.9075
	IMDN	715	47.2	32.04/0.8879	28.52/0.7811	27.69/0.7406	26.47/0.7982	30.92/0.9151
	SwinIR	899	165.3	32.44/0.8973	28.77/0.7858	27.43/0.7345	26.64/0.7998	31.09/0.8973
	SRFormer	734	72.8	<u>32.82/0.9029</u>	<b>29.01/0.7919</b>	27.85/0.7472	<u>27.20/0.8189</u>	<b>31.75/0.9137</b>
	RLFN	643	84.6	32.61/0.8918	28.83/0.7870	<u>27.72/0.7870</u>	26.63/0.8012	—/—
	ELAN	618	152.7	32.43/0.8965	28.78/0.7858	27.69/0.7406	26.54/0.7982	30.92/0.9150
	ESRT	751	49.8	32.29/0.8960	28.54/0.7832	<u>27.85/0.7379</u>	26.84/0.7975	30.75/0.9102
	ECTL-SR	604	23.4	<b>33.52/0.8975</b>	<b>28.95/0.7890</b>	<b>28.97/0.8182</b>	<b>28.36/0.8428</b>	<b>31.45/0.9180</b>
	ECTL-SR↑	604	23.4	<b>33.86/0.9012</b>	<b>29.08/0.7912</b>	<b>29.23/0.8314</b>	<b>28.77/0.8589</b>	<b>33.12/0.9312</b>

注: 加粗数据表示各数据集上性能指标最优, 下划线数据表示性能指标次优

本文可视化了在 Set5 上各方法的 PSNR、SSIM 指标与参数数量的关系图。在大多数场景下, ECTL-SR 模

型都能够以显著更低的资源消耗实现与现有先进 SR 网络相当甚至更高的 PSNR 和 SSIM 值, 如图 6 所示。

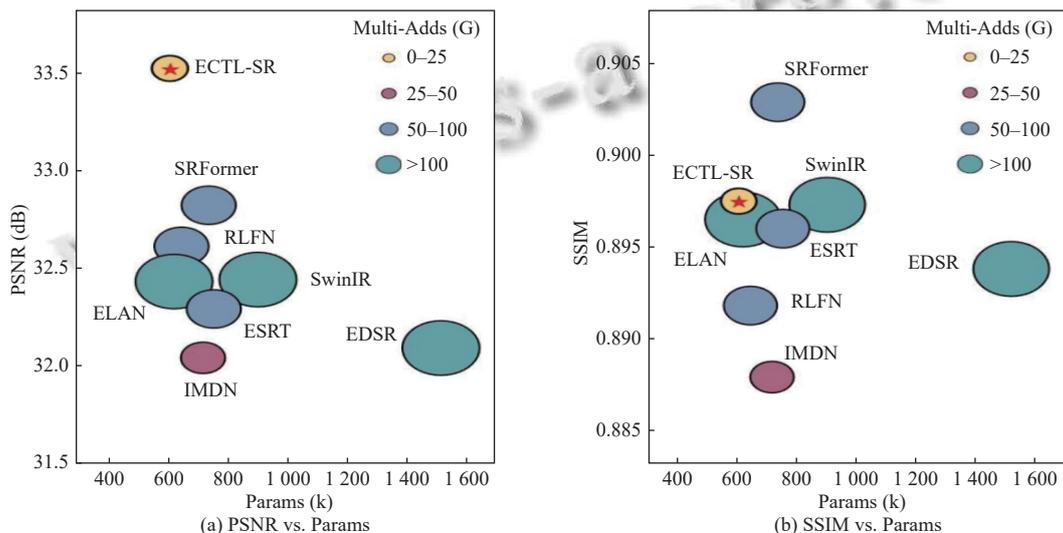


图6 不同方法在 Set5 数据集×4 放大尺度下的 PSNR、SSIM 与参数数量的关系

### 3.3.2 重建性能定性比较

为了直观评估模型重建图像的视觉质量, 选取

Urban100 数据集集中的 img012、img005 和 img002 这 3 个图像进行对比。如图 7 所示, 展示了 ECTL-SR 与其

他几种代表性轻量级 SISR 方法在 Urban100 数据集上×4 放大尺度下的视觉比较结果。

通过观察可以发现本文提出的 ECTL-SR 模型生成更清晰、更锐利的窗户边缘,成功地恢复了规整且

连续的网格线条,线条边缘清晰,没有明显的断裂或模糊。在重建图像时能够恢复出更丰富的细粒度细节,伪影更少,进一步验证模型对边缘结构信息的精确捕捉以及远距离像素依赖关系的有效建模。

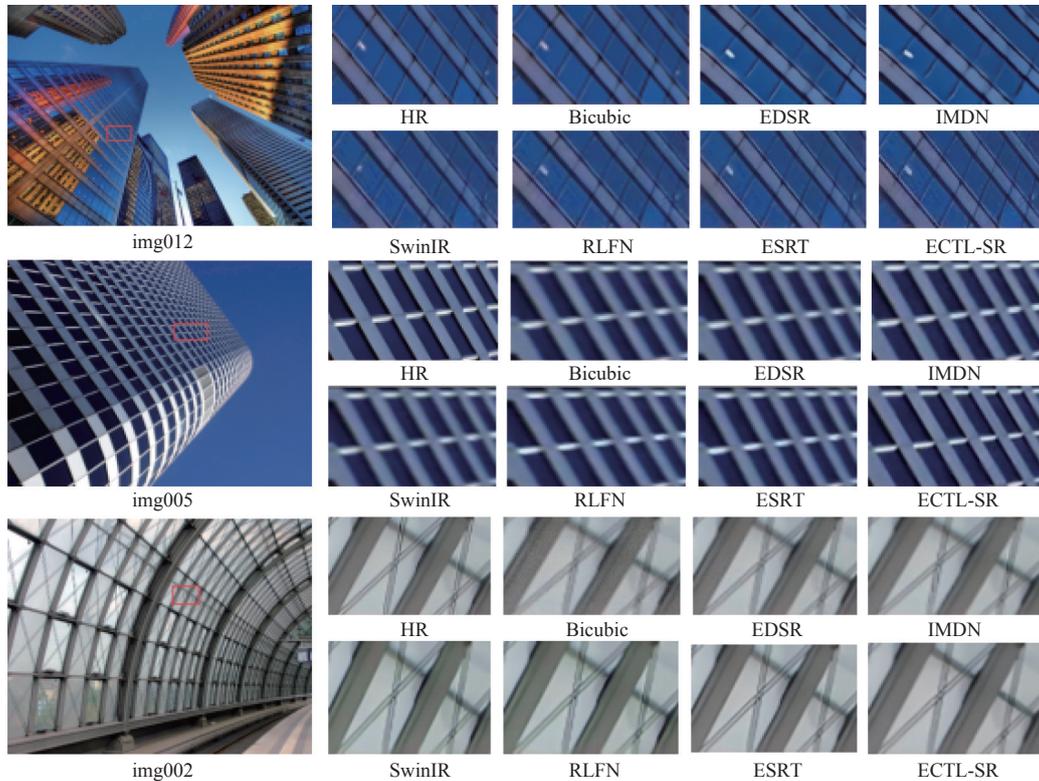


图7 不同方法在 Urban100 数据集×4 放大尺度下生成的超分重建结果对比

### 3.3.3 网络收敛性分析

为了验证本文 ECTL-SR 网络架构的训练稳定性与收敛性能,我们对其在训练过程中的关键性能指标进行了可视化分析,结果如图8所示。

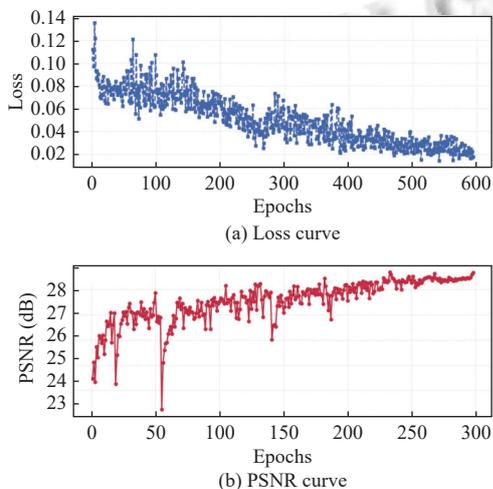


图8 网络收敛性

综合损失函数与 PSNR 的收敛过程,网络具有良好的训练稳定性和收敛特性,训练过程无梯度消失或爆炸等异常现象,模型能够最终收敛并达到理想的性能水平。训练过程中的损失曲线呈现出清晰的下降趋势,并在后期趋于平缓,证明了模型的有效收敛。与此同时,PSNR 性能曲线的持续上升也验证了模型重建质量在稳步提高。

### 3.3.4 推理速度与边缘设备部署

尽管模型参数量是衡量其复杂度的理论指标,但模型在硬件上的实际运行速度同样依赖于该硬件的计算架构与内存带宽。为了对 ECTL-SR 的真实效率进行评估,我们分别在高性能计算平台与资源受限的边缘计算平台上,对其推理延迟进行了详尽的测试。

#### (1) 高性能 GPU 平台

在 NVIDIA RTX 3060 Ti GPU 上,对×4 和×2 超分辨率任务进行了基准测试。我们测量了在 Urban100 数

据集各模型处理输入以生成 1280×720 分辨率图像时, 重复运行 100 次后的平均推理时间 (latency). 如表 2 所示, ECTL-SR 模型在所有对比模型中展现出了最快的推理速度. 与同样追求高性能的 SwinIR 相比, ECTL-SR 的速度提升了近一倍, 同时参数量更少. 相较于同样采用重参数化高效设计的 ECBSR<sup>[42]</sup>相比, ECTL-SR 在×2 和×4 放大尺寸下的速度分别快了约 1.75 倍和 2.14 倍.

表 2 不同 SR 模型在高性能 GPU 推理速度对比

Method	Params (k)	Latency (ms)	
		×2	×4
ECBSR	576	14.49	21.03
SwinIR	899	16.92	19.81
IMDN	715	28.53	30.15
VDSR	665	42.57	26.14
Ours	604	<b>8.29</b>	<b>9.84</b>

### (2) 边缘计算平台

我们将本文模型及多个具有代表性的轻量级 SOTA 方法部署在了主流的边缘计算平台上, 验证其实用性, 方便之后集成到无人机或智能监控摄像头. 测试平台为 NVIDIA Jetson Nano, 搭载 Maxwell 架构 GPU, 0.5 TFLOPs FP16 算力, 在 Ubuntu 操作系统下, 利用 NVIDIA TensorRT 对模型进行推理加速优化. 为确保结果的稳定与公平, 所有模型的延迟测试均在单线程、预热后的环境下, 连续运行 100 次后取平均值. 表 3 展示了各模型在 Jetson Nano 上的性能对比. 实验结果清晰地表明, 我们的 ECTL-SR 模型在资源受限的边缘设备上同样展现出了卓越的运行效率, 其推理延迟远低于其他对比方法.

表 3 不同 SR 模型在边缘部署推理速度对比

Method	Params (k)	Latency (ms)	
		×2	×4
ESRT	751	28.63	30.75
RLFN	643	29.12	32.38
IMDN	715	35.74	39.52
Ours	604	<b>10.63</b>	<b>16.15</b>

## 3.4 消融实验

为了验证 ECTL-SR 模型关键组件的有效性, 我们进行了详细的消融实验. 实验在×4 超分辨率任务上进行, 使用 DIV2K 数据集训练, 所有变体均采用与 ECTL-SR 完全相同的训练策略. 将 MSEE 和 GFEB 替换为简单的标准残差卷积块, 作为基线模型 (baseline), 并通过调整网络层数确保参数量尽可能一致. 消融实验结果如表 4 所示.

表 4 ECTL-SR 关键组件消融实验结果

Model	Block structure	Params (k)	Urban100	Manga109
			PSNR (dB)/ SSIM	PSNR (dB)/ SSIM
1	Baseline	569	27.46/0.8327	28.46/0.8920
2	Baseline+RepECB	578	27.83/0.8362	28.94/0.8931
3	Baseline+MSEE	586	28.05/0.8379	29.05/0.8936
4	Baseline+MSEE+GCPA	593	28.21/0.8403	30.29/0.8943
5	Baseline+MSEE+ GFEB-Basic	598	28.28/0.8412	30.32/0.9012
6	ECTL-SR (MSEE+ GFEB-MetaFormer)	604	28.36/0.8428	31.45/0.9180

(1) 多尺度边缘增强模块 (MSEE): 通过将 MSEE 内部的标准卷积块替换为核心的 RepECB 组件, 实验结果如表 4 中模型 2 所示, 在 Urban100 和 Manga109 数据集上的 PSNR 分别提升了 0.37 dB 和 0.48 dB, 这表明相较于简单卷积块, RepECB 结合标准卷积、方向性卷积和边缘算子的多分支结构, 提高了多尺度感受野和对结构边缘的感知能力, 能够更好地提取局部多尺度边缘特征. 在此基础上, 引入通道级注意力机制 LAM 构建出完整 MSEE 模块. LAM 根据每个通道的重要性自适应调整特征, 从而突出关键信息, 性能进一步提升了 0.22 dB 和 0.11 dB.

(2) 全局特征提取模块 (GFEB): 高质量超分辨率重建不仅关注局部高频细节的保真度, 还依赖于全局结构提供的上下文信息. 为此模型引入 GFEB 模块, 在 MSEE 基础上引入 GCPA 提取全局特征, 如表 4 中模型 4 所示, 实验发现性能提升且并未显著增加额外的参数量. 实验结果表明加入 ECA 后, 在 Urban100 和 Manga109 数据集上的 PSNR 分别进一步提升 0.07 dB 和 0.03 dB. 之后对比了基于 MetaFormer 架构的 ECTL-SR 模型与简单串联模块的 GFEB-Basic 来验证 MetaFormer 架构的有效性, 结果显示前者性能高出 0.08 dB 和 1.13 dB.

为了进一步验证本文提出的 RepECB 模块设计的有效性, 我们在 Urban100 上进行了多项消融实验, 如表 5 所示. 通过移除 RepECB 中的 1×3 和 3×1 方向性卷积分支, 转化为标准的 ECB, PSNR 下降了 0.30 dB, 表明结合在特定轴向的特征上可以更有效地捕捉图像中的边缘和纹理细节, 特别是. 此外, 我们还设计了不使用重参数化的 N-RepECB, 并与 RepECB 进行了对比. 结果表明尽管两者的性能几乎相同, RepECB 降低了约 55.6% 的推理延迟. 证明该模块能够在不牺牲精度的情况下, 显著提升模型的实际推理效率.

表5 RepECB 内部消融实验结果

Block	Params (k)	PSNR (dB)/SSIM	Latency (ms)
ECB <sup>[9]</sup>	447.09	28.06/0.8428	20.85
N-RepECB	513.68	28.24/0.8423	22.16
RepECB	498.42	28.36/0.8428	9.84

#### 4 结束语

针对单图像超分辨率中普遍存在的细节恢复不足、长距离依赖建模效率低和计算开销大等问题, 本文提出了一种高效轻量级的超分辨率网络 ECTL-SR. 该网络通过多尺度边缘增强模块 (MSEE) 精确捕捉并融合多层次细节, 其核心 RepECB 通过结构重参数化技术解耦训练复杂性与推理效率. 轻量化全局特征提取块 (GFEB) 利用位置感知环形卷积 (GCPA) 改进 Transformer 架构, 显著降低计算负担的同时增强全局上下文建模能力. 实验结果表明, ECTL-SR 在多个基准数据集上实现了卓越的重建质量, 尤其在结构保持和边缘纹理细节恢复方面优于现有轻量级方法, 并在模型复杂度与推理速度之间取得了良好平衡. 未来, 我们计划探索量化感知训练 (QAT) 和硬件感知结构化剪枝等压缩技术, 以进一步提升模型效率.

#### 参考文献

- Dong C, Loy CC, He KM, *et al.* Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(2): 295–307. [doi: [10.1109/TPAMI.2015.2439281](https://doi.org/10.1109/TPAMI.2015.2439281)]
- Lim B, Son S, Kim H, *et al.* Enhanced deep residual networks for single image super-resolution. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Honolulu: IEEE, 2017. 1132–1140.
- Zhang YL, Tian YP, Kong Y, *et al.* Residual dense network for image super-resolution. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 2472–2481.
- Zhang YL, Li KP, Li K, *et al.* Image super-resolution using very deep residual channel attention networks. *Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018. 294–310.
- Hui Z, Gao XB, Yang YC, *et al.* Lightweight image super-resolution with information multi-distillation network. *Proceedings of the 27th ACM International Conference on Multimedia*. Nice: ACM, 2019. 2024–2032.
- Liu J, Tang J, Wu GS. Residual feature distillation network for lightweight image super-resolution. *Proceedings of the 2020 European Conference on Computer Vision (ECCV)*. Glasgow: Springer, 2020. 41–55.
- Liang JY, Cao JZ, Sun GL, *et al.* SwinIR: Image restoration using swin Transformer. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021. 1833–1844.
- Gao GW, Wang ZX, Li JC, *et al.* Lightweight bimodal network for single-image super-resolution via symmetric CNN and recursive Transformer. *Proceedings of the 31st International Joint Conference on Artificial Intelligence*. Macao: IJCAI, 2022. 913–919.
- Wang QL, Wu BG, Zhu PF, *et al.* ECA-Net: Efficient channel attention for deep convolutional neural networks. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020. 11531–11539.
- Dong C, Loy CC, Tang XO. Accelerating the super-resolution convolutional neural network. *Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016. 391–407.
- Kim J, Lee JK, Lee KM. Accurate image super-resolution using very deep convolutional networks. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 1646–1654.
- Hui Z, Wang XM, Gao XB. Fast and accurate single image super-resolution via information distillation network. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 723–731.
- Kim J, Lee JK, Lee KM. Deeply-recursive convolutional network for image super-resolution. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 1637–1645.
- Tai Y, Yang J, Liu XM. Image super-resolution via deep recursive residual network. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 2790–2798.
- Zhang XD, Zeng H, Zhang L. Edge-oriented convolution block for real-time super resolution on mobile devices. *Proceedings of the 29th ACM International Conference on Multimedia*. Chengdu: ACM, 2021. 4034–4043.
- Chen Z, Zhang YL, Gu JJ, *et al.* Dual aggregation Transformer for image super-resolution. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision*. Paris: IEEE/CVF, 2023. 12278–12287.

- 17 Zhang XD, Zeng H, Guo S, *et al.* Efficient long-range attention network for image super-resolution. Proceedings of the 17th European Conference on Computer Vision. Tel Aviv: Springer, 2022. 649–667.
- 18 Zhou YP, Li Z, Guo CL, *et al.* SRFormer: Permuted self-attention for single image super-resolution. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023. 12734–12745.
- 19 Wang Y, Liu Y, Zhao SJ, *et al.* CAMixerSR: Only details need more “attention”. Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024. 25837–25846.
- 20 Shi WZ, Caballero J, Huszár F, *et al.* Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 1874–1883.
- 21 Paris S, Hasinoff SW, Kautz J. Local Laplacian filters: Edge-aware image processing with a Laplacian pyramid. Communications of the ACM, 2015, 58(3): 81–91. [doi: [10.1145/2723694](https://doi.org/10.1145/2723694)]
- 22 Sandler M, Howard A, Zhu ML, *et al.* Mobilenetv2: Inverted residuals and linear bottlenecks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 4510–4520.
- 23 Ding XH, Zhang XY, Ma NN, *et al.* RepVGG: Making VGG-style ConvNets great again. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13728–13737.
- 24 Yu WH, Luo M, Zhou P, *et al.* MetaFormer is actually what you need for vision. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 10809–10819.
- 25 Agustsson E, Timofte R. NTIRE 2017 challenge on single image super-resolution: Dataset and study. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu: IEEE, 2017. 1122–1131.
- 26 Bevilacqua M, Roumy A, Guillemot C, *et al.* Low-complexity single-image super-resolution based on nonnegative neighbor embedding. Proceedings of the 2012 British Machine Vision Conference. Surrey: BMVA, 2012. 1–10.
- 27 Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations. Proceedings of the 7th International Conference on Curves and Surfaces. Avignon: Springer, 2012. 711–730.
- 28 Martin D, Fowlkes C, Tal D, *et al.* 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: IEEE, 2001. 416–423.
- 29 Huang JB, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 5197–5206.
- 30 Matsui Y, Ito K, Aramaki Y, *et al.* Sketch-based manga retrieval using Manga109 dataset. Multimedia Tools and Applications, 2017, 76(20): 21811–21838. [doi: [10.1007/s11042-016-4020-z](https://doi.org/10.1007/s11042-016-4020-z)]
- 31 Gao XB, Lu W, Tao DC, *et al.* Image quality assessment based on multiscale geometric analysis. IEEE Transactions on Image Processing, 2009, 18(7): 1409–1423. [doi: [10.1109/TIP.2009.2018014](https://doi.org/10.1109/TIP.2009.2018014)]
- 32 Wang Z, Bovik AC, Sheikh HR, *et al.* Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13(4): 600–612. [doi: [10.1109/TIP.2003.819861](https://doi.org/10.1109/TIP.2003.819861)]
- 33 Paszke A, Gross S, Chintala S, *et al.* Automatic differentiation in PyTorch. Proceedings of the 31st Conference on Neural Information Processing Systems. Long Beach: NIPS, 2017. 1–4.
- 34 Kong FY, Li MX, Liu SW, *et al.* Residual local feature network for efficient super-resolution. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans: IEEE, 2022. 765–775.
- 35 Lu ZS, Li JC, Liu H, *et al.* Transformer for single image super-resolution. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans: IEEE, 2022. 456–465.

(校对责编: 张重毅)