中图法分类号:TP391　　**文献标识码**:A　　**文章编号**:1006-8961(XXXX)XX-0001-13

论文引用格式:Wu Siqi, Liu Wei, Chen Weidong. XXXX. Lightweight Image Super-Resolution Network with Sparse and Permuted Self-Attention. Journal of Image and Graphics, XX(XX):0001-0013(吴思琪, 柳薇, 陈卫东. XXXX. 轻量级稀疏置换自注意力图像超分辨率网络. 中国图象图形学报, XX(XX):0001-0013)[DOI:10.11834/jig.250519]

轻量级稀疏置换自注意力图像超分辨率网络

吴思琪,柳薇,陈卫东

华南师范大学计算机学院,广州　510631

摘　要:**目的**　图像超分辨重建是计算机视觉领域中的一个典型低层视觉任务,能够为目标检测、图像分割等高层任务提供更清晰更结构化的输入。基于CNN的图像超分辨率模型注重恢复图像的纹理和边缘信息,而基于Transformer的方法能建模全局上下文信息,但是存在注意力权重冗余问题。针对这两种模型的优缺点,本文设计了一种轻量级图像超分辨率网络。**方法**　首先改进了传统的Transformer,提出了一种稀疏置换自注意力机制,在扩大窗口的同时解决冗余问题。在此基础上,我们基于CNN构建高频信息增强模块加强模型对局部细节信息的重建。在得到两种结构提取的特征后,我们提出一种双分支特征融合模块对全局特征和局部特征进行高效融合。**结果**　本文方法在5个公开数据集上与11种先进超分辨方法进行了对比实验。结果表明,在保证模型轻量化的前提下,稀疏置换自注意力网络(Sparse and Permuted Self-Attention Network,SPSANet)在不同放大倍率和数据集上均取得最优或次优性能。当放大倍率为3时,在Urban100和Manga109数据集上的峰值信噪比(peak signal-to-noise ratio,PSNR)分别较最新的SOTA(state of the art)方法提升了0.15dB和0.25dB。主观视觉效果显示,SPSANet在复杂纹理和细节丰富的场景中重建的图像更加清晰、自然。**结论**　本文提出的轻量级稀疏置换自注意力图像超分辨率网络能够在保持较低参数量与计算复杂度的同时,在多个数据集上取得优异的重建效果,展现出良好的泛化性与应用价值。

关键词:深度学习;图像超分辨率;稀疏置换自注意力;高频信息增强;双分支特征融合

Lightweight Image Super-Resolution Network with Sparse and Permuted Self-Attention

Wu Siqi, Liu Wei, Chen Weidong

School of Computer Science, South China Normal University, Guangzhou 510631, China

Abstract: **Objective** Image super-resolution (SR) is a fundamental low-level vision task in computer vision that aims to reconstruct high-resolution (HR) images from low-resolution (LR) inputs, thereby enhancing image detail quality and sharpness. High-quality reconstruction not only improves human visual perception but also provides clearer and more structured input features for high-level vision tasks such as object detection, semantic segmentation, and face recognition. Therefore, SR technology has broad application prospects in fields such as intelligent surveillance, medical imaging, autonomous driving, and satellite remote sensing. Convolutional neural networks (CNNs) have been widely used for SR due to their ability to capture local textures and edges. However, their limited receptive fields hinder the modeling of long-range dependencies, which are essential for preserving global structural consistency. Transformer-based methods, by contrast, leverage self-attention mechanisms to capture global context, achieving superior reconstruction performance. Despite this advantage, conventional Transformers suffer from high computational costs and redundant attention weight, limiting their applicability in lightweight or real-time scenarios. To address these challenges, we propose a lightweight SR network

收稿日期:2025-10-23;**修回日期**:2025-12-19

that integrates the advantages of both CNN and Transformer architectures. The network integrates efficient local feature extraction with global context modeling, achieving a balance between reconstruction quality and computational efficiency. **Method** We propose SPSANet, a lightweight image super-resolution network based on Sparse and Permuted Self-Attention, composed of shallow feature extraction, deep feature extraction, and high-resolution reconstruction. In the shallow feature extraction stage, a standard ×33 convolution captures basic texture information, and shallow features are forwarded to the reconstruction stage via a long skip connection to provide residual guidance. The deep feature extraction stage consists of four Sparse Permuted Self-Attention Groups, each containing six Sparse Permuted Self-Attention Blocks and a convolutional layer. Each block integrates a Sparse and Permuted Self-Attention Module (SPSAM) that expands the receptive field while reducing attention redundancy, a High-Frequency Enhancement Module (HFEM) for refining texture and edge details, and a Dual-Branch Feature Fusion Module (DBFFM) that effectively fuses global and local features through spatial-channel interactions. Finally, shallow and deep features are fused via a global residual connection, and a pixel shuffle operation generates the final high-resolution output. Our method is implemented on the PyTorch framework and trained using an NVIDIA RTX 3090 GPU. The training dataset is DIV2K, which contains 800 images. For the ×2 super-resolution (SR) task, the model is trained for a total of 500K iterations using the Adam optimizer with an initial learning rate of 2×10^{-4}. The input image patch size is fixed at 64×64 and the batch size is set to 16, and a MultiStepLR scheduler is applied to halve the learning rate at iterations [250K, 400K, 450K, 475K]. For the ×3 and ×4 SR tasks, the model is initialized with the pretrained weights from the ×2 model, while the total number of training iterations is reduced by half. **Result** The proposed method was evaluated on five public benchmark datasets and compared with eleven state-of-the-art super-resolution approaches. Experimental results demonstrate that under the constraint of lightweight design, SPSANet achieves either the best or second-best performance across different upscale factors and datasets, showing strong generalization ability and stability. Specifically, when the upscale factor is ×3, SPSANet surpasses the latest SOTA methods by 0.15 dB and 0.25 dB in peak signal-to-noise ratio (PSNR) on the Urban100 and Manga109 datasets, respectively, and also achieves corresponding improvements in structural similarity index (SSIM). Furthermore, when the self-ensemble strategy is applied during the testing phase, the model performance is further enhanced. Under the same scaling factor, the PSNR increases by an additional 0.11 dB and 0.23 dB on Urban100 and Manga109, respectively, indicating that the proposed method maintains strong robustness under multi-view inference. To further investigate the contribution of each component, ablation studies were conducted on the SPSAM, the HFEM, and the DBFFM. The results show that removing any of these modules leads to a noticeable degradation in performance. Among them, SPSAM plays a key role in global dependency modeling and attention sparsification, HFEM effectively enhances texture restoration quality, and DBFFM significantly improves the fusion between global and local features. In terms of visual quality, SPSANet is capable of reconstructing clearer, more natural, and sharper image textures, particularly in scenes with complex structures and abundant details. It preserves the structural consistency and perceptual realism of the original images more effectively than existing approaches. **Conclusion** In this work, we propose SPSANet, a lightweight image super-resolution network that integrates a Sparse and Permuted Self-Attention mechanism to efficiently capture long-range dependencies while minimizing redundant computations. By establishing inter-layer attention connections, SPSANet enables more focused feature learning, ensuring that the network prioritizes tokens most critical to the reconstruction process. Additionally, the proposed High-Frequency Enhancement Module enhances the recovery of fine textures and structural details, while the Dual-Branch Feature Fusion Module effectively aligns and integrates global and local representations. Through extensive experiments, SPSANet demonstrates superior reconstruction performance and visual fidelity compared with existing lightweight SR approaches, especially in challenging scenarios with intricate patterns and rich textures. Despite these promising results, opportunities for further improvement remain. Future work will aim to enhance model compactness and generalization through knowledge distillation, hybrid convolution-transformer integration, and adaptation to broader low-level vision tasks such as image denoising and deblurring. Overall, SPSANet provides a practical and interpretable solution for lightweight Transformer-based image restoration, offering insights for the design of efficient and high-performing SR networks.

Key words: deep learning; image super-resolution; sparse and permuted self-attention; high-frequency information enhancement; dual-branch feature fusion

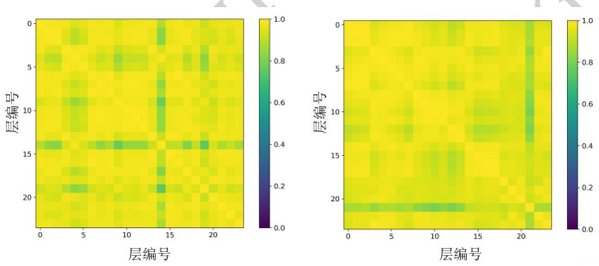
0 引 言

图像超分辨率(super-resolution,SR)旨在从低分辨率的输入重建出高分辨率图像,高效的图像超分辨率技术不仅能提升图像的视觉质量,还能增强图像中可供分析和识别的有效信息。在医学影像诊断、遥感测绘、视频监控和自动驾驶等对图像清晰度要求较高的领域,SR技术都有重要的应用价值。

早期的SR研究主要依赖卷积神经网络CNN,通过堆叠卷积层来学习低分辨率到高分辨率的映射。尽管CNN在SR任务上取得满意的效果,但受限于卷积核的局部感受野,难以充分建模全局上下文信息。为了突破这一瓶颈,许多研究开始探索具备长程依赖建模能力的结构,其中Transformer凭借自注意力(Self-Attention,SA)机制的全局建模优势,逐渐成为SR任务中的关键架构。基于Transformer的SR方法能够在更大范围内捕获像素间的相关性,从而有效提升图像细节的恢复精度与视觉质量。

在本文,我们旨在设计一种面向轻量化SR任务的高效Transformer架构,在保证计算高效的同时进一步提升重建性能。首先我们对基于Transformer架构的SRFormer(Zhou等,2023)网络内部各层的注意力图进行了相似性评估。具体而言,我们在Urban100数据集上,对每一层的注意力权重在"注意力头维度"和"窗口维度"上分别求平均,以得到该层的整体注意力响应。随后,将这些注意力权重展开为一维向量,并计算不同层之间注意力权重的两两余弦相似度。结果如图1所示:无论是×2还是×4超分辨率任务,各层注意力分布都表现出高度一致性。基于这一观察,我们提出了一种稀疏置换自注意力(Sparse and Permuted Self-Attention,SPSA)机制,能捕获更大范围的像素依赖关系,并利用层间注意力权重相似实现稀疏注意力。具体来说,SPSA将空间信息重新置换到通道维度,在不增加额外计算量的前提下将窗口大小扩大至原来的4倍。因为全连接计算放大了较小相似度token的权重,产生冗余和不相关的特征表示(Chen X等,2023),我们将当前层与前一层的注意力权重进行逐元素相乘,并将得到的结果归一化,从而自适应地增强连续相似token的响应,抑制无关或冗余特征,实现注意力的稀疏化与有效化。同时,为了弥补Transformer表征高频信息的不足,我们设计了一个高频细节增强(High Frequency Enhancement,HFE)模块,以强化局部纹理与边缘特征的表征能力。为了更有效的融合Transformer分支的全局特征和卷积分支的局部特征,我们进一步提出了一种双分支特征融合机制(Dual-Branch Feature Fusion,DBFF),用于缓解全局与局部特征之间的对齐偏差问题,从而实现信息的高效互补。

本文的主要贡献包括4个方面:1)提出稀疏置换自注意力机制,在扩大感受野的同时,有效缓解了注意力权重的冗余问题。2)设计高频信息增强模块,强化了模型对局部高频特征的建模能力。3)构建双分支特征融合机制,实现全局与局部特征的高效融合与对齐。4)提出一种轻量化Transformer架构,在保证计算高效的同时,进一步提升了SR任务的重建性能。



(a)×2余弦相似度 (b)×4余弦相似度
((a)×2 Cosine similarity;(b)×4 Cosine similarity)

图1 基于Urban100计算的余弦相似度

Fig. 1 Cosine similarity calculated based on Urban100

1 相关工作

1.1 基于CNN的图像超分辨率模型

CNN驱动的图像超分辨率模型从2014年以来快速发展。SRCNN(Dong等,2014)模型是首次将深度卷积网络直接用于单图像超分辨率的端到端的方法,它把经过双三次插值放大的低分辨率图像作为网络输入,学习从粗略高分辨率到精细高分辨率的像素级映射,与传统方法相比效果更好。自此之后,研究人员通过研究如何堆叠更深的网络架构来提高性能。但随着网络层数的提高,模型会带来梯度消失等问题。VDSR(Kim等,2016)提出了一种由20层卷积构成的深度超分辨率网络,将网络学习到的残

差和插值的低分辨率图像直接相加得到高分辨图像，该模型通过引入残差学习有效地解决了梯度消失问题，从而加速模型收敛。为了减少计算开销和内存消耗，研究人员提出了基于蒸馏的方法，主要分为基于信息蒸馏和基于知识蒸馏。IDN（Hui等，2018）提出信息蒸馏块来分流并蒸馏特征通道，从而减少冗余计算并提升信息利用率，其中增强单元专注于提升局部表示能力以恢复高频细节，而压缩单元则负责将连续块中的重要信息压缩并传递到后续层级，二者协同作用使得模型在显著降低计算成本的同时仍能保持较高的重建性能。FAKD（He等，2020）采用一种基于知识蒸馏框架，通过将教师模型的结构化知识迁移到轻量学生模型来压缩计算量和内存开销，为了保留教师模型的结构化表示能力，FAKD侧重蒸馏特征图的二阶统计信息。（Wu等，2022）提出了一种新的跨尺度耦合上采样机制，用于替代传统固定比例的上采样层，使模型能够灵活地应对任意放大倍数的重建任务，同时引入跨尺度卷积结构在不同尺度上并行提取图像特征，有效增强了模型在连续放大场景下的重建质量与泛化能力。MLFN（Song等，2025）通过使用多尺度可分离大核卷积提高感受野从而增强全局特征建模能力，在有效控制计算复杂度的同时实现高质量的图像超分辨率重建。

1.2 基于Transformer的图像超分辨率模型

基于CNN的图像超分辨率模型局部感知能力强，但难以捕获长距离依赖关系，而Transformer得益于窗口机制增大了感受野，广泛应用于图像恢复任务。SwinIR（Liang等，2021）以Swin Transformer为基础，通过引入窗口划分和移位机制，有效扩大了感受野并展现了优异性能。为了解决SA计算成本高的问题，ELAN（Zhang等，2022）使用移位卷积提取局部信息，同时在不同尺度窗口上对不重叠的特征组计算注意力，将二者串联组成高效的长距离注意力网络，并通过注意力共享策略进一步提高计算效率。为了解决SA只在单一维度建模，导致空间与通道之间交互不足的问题，OminSR（Wang等，2023）提出两个关键改进，一是使用空间注意力和通道注意力同时在两个维度建模像素间的关系，二是设计多尺度交互机制，增强局部、中观和全局特征的协调作用，从而获得丰富的多尺度特征表达。ATD（Zhang等，2024）引入自适应标记字典，用于从训练数据中学习先验知识，并在测试阶段自适应地优化以适配输入图像。该机制能为特征提供全局信息并根据相似性分组，从而实现更有效的特征融合。该方法同时提出了基于类别的自注意力机制，利用相似标记间的关系增强特征表达。HiT-SR（Zhang等，2024）将传统的固定小窗口替换为分层可扩展窗口，同时还设计了空间通道关联机制，以线性复杂度高效融合多层窗口中的空间和通道特征，在性能与效率之间取得了很好平衡。BSTN（Bi等，2024）以蓝图可分离卷积为核心，构建了高效的蓝图前馈网络和蓝图多头自注意力，同时结合移位卷积设计了移位通道注意力，取得了很好的性能。ESC（Lee等，2025）提出通过共享的大核卷积和动态卷积模拟SA的长距离依赖建模和动态加权特性，大幅减少了对传统自注意力及其高内存操作的依赖，同时保持了Transformer的强大特征表达能力。该方法还成功将闪存注意力（Flash Attention，FA）机制应用于轻量级超分辨率模型，将窗口扩大至32×32，有效缓解了注意力机制的内存瓶颈。（Li等，2025）在深层特征提取阶段引入了跨尺度Transformer结构，使网络能够在不同尺度之间建立更丰富的长程依赖关系。同时，他们结合通道增强机制，进一步提升特征在通道维度的表达能力。在特征融合阶段，方法采用自适应通道加权策略，根据当前内容自动调节各尺度特征的重要性。

2 本文方法

2.1 网络结构

如图2所示，本文提出的SPSANet网络结构主要由浅层特征提取层、深层特征提取层和高分辨率图像重建层三个关键部分组成。浅层特征提取层包含一个3×3的卷积，它将输入的低分辨率图像$\boldsymbol{I}_{\mathrm{LR}} \in \mathbf{R}^{H\times W\times 3}$转化为$\boldsymbol{F}_{\mathrm{SF}} \in \mathbf{R}^{H\times W\times C}$。在获得浅层特征之后，$\boldsymbol{F}_{\mathrm{SF}}$被送入深层特征提取层获得深层特征表示$\boldsymbol{F}_{\mathrm{DF}} \in \mathbf{R}^{H\times W\times C}$，其中深层特征提取层由M个稀疏置换注意力组（Spare and Permuted Self-Attention Group，SPSAG）组成，每个SPSAG包含N个稀疏注意力块（Spare and Permuted Self-Attention Block，SPSAB）和一个卷积。为了进一步整合深层特征信息，在这部分的末尾加入一个3×3卷积。最后，使用全局残差连接将浅层特征$\boldsymbol{F}_{\mathrm{SF}}$和深层特征$\boldsymbol{F}_{\mathrm{DF}}$进行相加，并将相加后的特征送入图像重建层得到最终

的高分辨率输出图像，其中图像重建层包含3×3卷积和亚像素卷积。

2.2 稀疏置换自注意力

本文提出的稀疏置换自注意力不仅为模型提供了强大的全局特征建模能力，而且解决了注意力权重的冗余问题。如图4(a)所示，给定输入的特征$\boldsymbol{X}_{\text{in}} \in \mathbf{R}^{H\times W\times C}$和一个token压缩比例$r$，首先将输入划分为$N$个互不重叠的窗口，然后将窗口展开得到$\boldsymbol{X} \in \mathbf{R}^{N\times S^2\times C}$，其中$S$是窗口的边长。接着分别通过

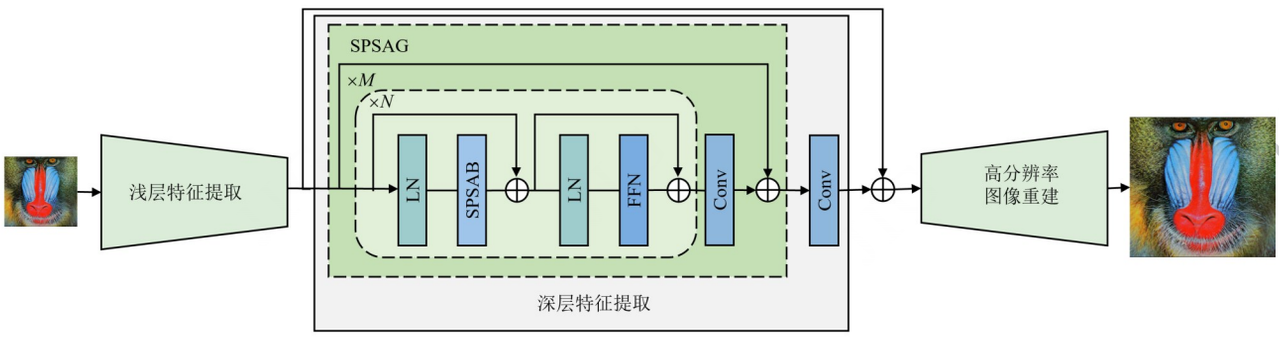


图2 网络结构

Fig. 2 Network Architecture

三个线性变化层来生成查询$\boldsymbol{Q} \in \mathbf{R}^{N\times S^2\times C}$，键$\boldsymbol{K} \in \mathbf{R}^{N\times S^2\times C/r^2}$和值$\boldsymbol{V} \in \mathbf{R}^{N\times S^2\times C/r^2}$，其中$\boldsymbol{Q}$保持与$\boldsymbol{X}$相同的通道维度，$\boldsymbol{K}$和$\boldsymbol{V}$的通道维度压缩为$C/r^2$，具体实现中我们设置$r=2$。为了在不增加计算量的前提下让更多的token参加注意力计算，将$\boldsymbol{K}$和$\boldsymbol{V}$的空间维度上相邻的4个token置换到通道维度上，具体的实现如图3所示，得到置换后的特征$\boldsymbol{K}_{\text{p}} \in \mathbf{R}^{N\times S^2/r^2\times C}$和$\boldsymbol{V}_{\text{p}} \in \mathbf{R}^{N\times S^2/r^2\times C}$(Zhou等，2023)。然后以$\boldsymbol{Q}$和$\boldsymbol{K}_{\text{p}}$为输入计算注意力分数：

$$\text{Score}(\boldsymbol{Q},\boldsymbol{K}_{\text{p}}) = \text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}_{\text{p}}^{\text{T}}}{\sqrt{d_k}} + B\right) \tag{1}$$

其中d_k是缩放系数，用来稳定注意力计算。B是相对位置偏置，通过插值的方法在窗口大小不一致的情况下进行相对位置对齐。以往方法的注意力权重仅由当前层的token相似度计算得到，而这种标准相似度计算难以强化高相关token的同时抑制低相关token，从而造成冗余(Long等，2025)。为了解决该问题，本文使用乘法继承的方式跨层计算注意力权重，定义如下：

$$\boldsymbol{S}_{\text{L}} = \text{Norm}(\boldsymbol{S}_{\text{L}-1} \odot \boldsymbol{S}_{\text{L}}^{\text{cal}}) \tag{2}$$

其中$\boldsymbol{S}_{\text{L}}^{\text{cal}}$表示根据公式(1)计算得到的第L层的注意力权重，$\boldsymbol{S}_{\text{L}-1}$为前一层的注意力权重，Norm(·)表示归一化操作，⊙表示逐元素相乘。如图3所示，第L层的注意力权重$\boldsymbol{S}_{\text{L}}$由当前层计算的注意力权重$\boldsymbol{S}_{\text{L}}^{\text{cal}}$与前一层的注意力权重$\boldsymbol{S}_{\text{L}-1}$逐元素相乘后归一化得到。如图1所示，不同层之间的注意力权重极其相似，随着网络层数的增加，那些相关性较低的token会在连续的注意力乘积过程中逐渐被抑制，而在多层中保持高度相似的关键token则会在归一化后获得更大的权重，从而达到稀疏的作用。

2.3 高频信息增强

Transformer结构更善于捕捉低频信息(Li等，2023)，与卷积结构相比，构建高频表征的能力有限，所以本文加入卷积分支来增强模型对图像高频信息的学习。如图4(b)所示，具体结构由三部分组成，首先采用1×1卷积实现通道间的线性映射，用于特征压缩与重组。接着利用3×3深度可分离卷积提取局部空间信息，从邻域中捕获高频细节，并通过GELU激活函数引入非线性变换，增强特征的表达能力。最后使用1×1卷积融合通道信息并恢复通道维度。

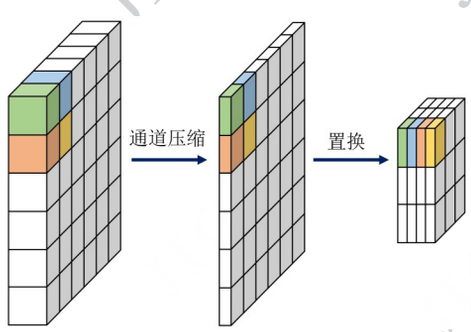


图3 压缩置换操作

Fig. 3 Compression Permutation Operation

2.4 双分支特征融合

由于来自Transformer的全局特征和来自卷积的局部特征之间存在特征对齐偏差(Mao等,2021;Peng等,2021),两者直接融合可能会引起歧义,所以本文引入双分支特征融合机制,用于更有效地融合全局和局部信息(Chen等,2023)。如图4(c)所示,融合机制包含两种操作:空间交互和通道交互。如图4(d)和4(f)所示,对于输入特征$\boldsymbol{X}_{\text{attn}} \in \mathbf{R}^{H\times W\times C}$和$\boldsymbol{X}_{\text{conv}} \in \mathbf{R}^{H\times W\times C}$,分别经过空间交互S-I和通道交互C-I生成空间注意力图和通道注意力图,所得到的注意力图用于对分支特征的自适应加权,具体的过程如下:

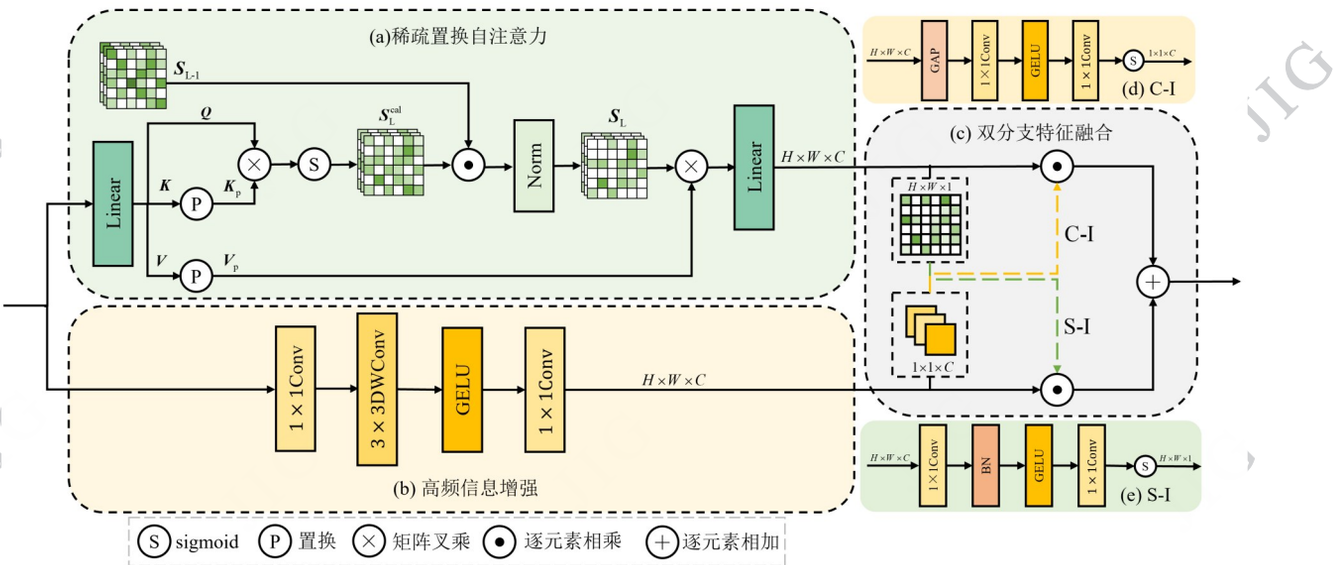


图4 稀疏置换注意力块

Fig. 4 Sparse permuted Self-Attention block

$$\begin{aligned}\boldsymbol{S-Map} &= \mathrm{f}(\mathrm{W}_2\sigma(\mathrm{N}(\mathrm{W}_1\boldsymbol{X}_{\text{attn}}))),\\ \boldsymbol{C-Map} &= \mathrm{f}(\mathrm{W}_4\sigma(\mathrm{W}_3\,\mathrm{GAP}(\boldsymbol{X}_{\text{conv}}))),\\ \boldsymbol{X}_{\text{out}} &= \boldsymbol{S-Map}\odot\boldsymbol{X}_{\text{conv}} + \boldsymbol{C-Map}\odot\boldsymbol{X}_{\text{attn}}\end{aligned} \tag{3}$$

其中f(·)表示sigmoid激活函数,W(·)表示逐点卷积,N(·)表示批量归一化,σ(·)为GELU激活函数,GAP(·)表示全局平均池化,⊙表示逐元素相乘,$\boldsymbol{S-Map}$为空间注意力图,$\boldsymbol{C-Map}$为通道注意力图,$\boldsymbol{X}_{\text{out}}$为输出。Transformer分支提取的全局特征通过空间交互生成空间注意力图,该注意力图与卷积分支的局部特征进行加权融合,从而有效缓解了两者之间的特征对齐偏差问题。另一方面,在Transformer中,由查询$\boldsymbol{Q}$和键$\boldsymbol{K}$点积生成的注意力权重会被所有通道共享(Chen等,2022),这限制了模型在通道维度上的特征建模能力。为此,我们引入通道交互机制,通过生成通道注意力图并对Transformer提取的特征进行加权,使不同通道的特征权重得到自适应调整,从而克服了权值共享带来的限制。通过结合空间与通道两个层面的双向交互,网络在局部细节保持与全局依赖建模之间实现了充分融合。

3 实 验

3.1 实验设置

3.1.1 数据集和评估指标

本文采用DIV2K(Lim等,2017)数据集作为训练集,包含900张图片,并通过对高分辨率图像进行双三次插值降采样获得低分辨率图像。为了评估所提出方法的性能,本文在五个基准数据集上进行测试,分别为Set5(Bevilacqua等,2012)、Set14(Zeyde等,2012)、BSD100(Martin等,2001)、Urban100(Huang等,2015)和Manage109(Matsui等,2017)。实验在YCbCr图像空间中的Y通道上评估,以峰值信噪比(Peak signal-to-noise ratio,PSNR)和结构性相似度(structural similarity index,SSIM)作为评价标准。

3.1.2 实现细节

SPSANet网络深层提取部分由4个SPSAG组成,其中每组包含6个SPSAB,通道数、注意力头数

和窗口大小分别设置为60、6和16。模型在×2 SR任务上共迭代训练500K次，优化器采用Adam，参数设置为$\beta_1 = 0.9$、$\beta_2 = 0.99$。训练的图像块大小固定为64×64，初始学习率为2×10^{-4}，并使用MultiStepLR学习率调度器，在迭代次数[250000，400000，450000，475000]时将学习率减半，批量大小设置为16。对于×3和×4任务，模型以×2的预训练权重初始化，同时将迭代次数减半。本文采用L1损失函数来计算重建高分辨率图像和真实图像之间的绝对误差。在推理阶段，我们进一步引入自集成策略（Lim等，2017），使模型增强为SPSANet+。与基础版本SPSANet仅依赖单次前向推理不同，SPSANet+会对输入进行八种不同的数据增强，如翻转、旋转和转置等，然后分别送入同一个模型推理，

表1　不同方法的2倍超分辨率客观评价结果

Table 1　Results of objective results of ×2super-resolution for different methods

方法	参数量/k	Mult-Adds/G	Set5		Set14		BSD100		Urban100		Manage109	
			PSMR	SSIM	PSMR	SSIM	PSMR	SSIM	PSMR	SSIM	PSMR	SSIM
EDSR（Lim等，2017）	1370	316.3	37.99	0.9604	33.57	0.9175	32.16	0.8994	31.98	0.9272	38.54	0.9769
CARN（Ahn等，2018）	1592	222.8	37.76	0.9590	33.52	0.9166	32.09	0.8978	31.92	0.9256	38.36	0.9765
IMDN（Hui等，2019）	694	158.8	38.00	0.9605	33.63	0.9177	32.19	0.8996	32.17	0.9283	38.88	0.9774
LatticeNet（Luo等，2020）	756	169.5	38.06	0.9607	33.70	0.9187	32.20	0.8999	32.25	0.9288	–	–
SwinIR-It（Liang等，2021）	878	195.6	38.14	0.9611	33.86	0.9206	32.31	0.9012	32.76	0.9340	39.12	0.9783
LBNet（Gao等，2022）	731	153.2	38.05	0.9607	33.65	0.9177	32.16	0.8994	32.30	0.9291	38.88	0.9775
ESRT（Lu等，2022）	677	191.4	38.03	0.9600	33.75	0.9184	32.25	0.9001	32.58	0.9318	39.12	0.9774
SwinIR-NG（Choi等，2023）	1181	274.1	38.17	0.9612	33.94	0.9205	32.31	0.9013	32.78	0.9340	39.20	0.9785
SRFormer-It（Zhou等，2023）	853	197.5	38.23	0.9613	33.94	0.9209	<u>32.36</u>	0.9019	32.91	0.9353	39.28	0.9785
BSTN（Bi等，2024）	736	163.6	38.17	0.9616	33.87	0.9201	32.28	0.9010	32.60	0.9318	–	–
MLFN-L（Song等，2025）	1071	–	38.15	0.9612	33.85	0.9197	32.33	0.9013	32.58	0.9331	39.06	0.9782
SPSANet（Ours）	1063	247.0	<u>38.25</u>	<u>0.9616</u>	<u>33.95</u>	<u>0.9214</u>	32.35	<u>0.9020</u>	<u>32.99</u>	<u>0.9363</u>	<u>39.34</u>	<u>0.9787</u>
SPSANet+（Ours）	1063	247.0	**38.28**	**0.9617**	**34.02**	**0.9217**	**32.38**	**0.9023**	**33.14**	**0.9372**	**39.48**	**0.9790**

注：加粗和下划线字体分别表示各列最优和次优结果

再将结果逆变换并求平均，得到更高质量的输出。

3.2　对比实验

我们将提出的方法与当前轻量级图像超分辨率领域内的先进方法进行比较，这些方法中有基于CNN的模型EDSR（Lim等，2017）、CARN（Ahn等，2018）、IMDN（Hui等，2018）、LatticeNet（Luo等，2020）、MLFN-L（Song等，2025）和基于Transformer的模型SwinIR（Liang等，2021）、LBNet（Gao等，2022）、SwinIR-NG（Choi等，2023）、SRFormer-It（Zhou等，2023）、ESRT（Lu等，2022）、BSTN（Bi等，2024）。这些为我们验证所提方法的性能优势提供了可靠的对比基线。

3.2.1　客观实验结果

各方法在五个基准测试集上的参数量，乘加运算量和测试结果如表1~3所示，其中参数量和乘加运算量均在将图片放大至1280×720测试得到。从结果来看本文的方法在数据集Urban100上超过次优模型0.08~0.15dB，在数据集Manga100上超过次优模型0.06~0.25dB，并且参数量和计算量保持较

表 2　不同方法的3倍超分辨率客观评价结果

Table 2　Results of objective results of ×3 super-resolution for different methods

方法	参数量/k	Mult-Adds/G	Set5		Set14		BSD100		Urban100		Manage109	
			PSMR	SSIM	PSMR	SSIM	PSMR	SSIM	PSMR	SSIM	PSMR	SSIM
EDSR(Lim等,2017)	1555	160.2	34.37	0.9270	30.28	0.8417	29.09	0.8052	28.15	0.8527	33.45	0.9439
CARN（Ahn 等，2018）	1 592	118.8	34.29	0.9255	30.29	0.8407	29.06	0.8034	28.06	0.8493	33.50	0.9440
IMDN(Hui等,2019)	703	71.5	34.36	0.9270	30.32	0.8417	29.09	0.8046	28.17	0.8519	33.61	0.9445
LatticeNet（Luo 等，2020）	765	76.3	34.53	0.9281	30.39	0.8424	29.15	0.8059	28.33	0.8538	–	–
SwinIR-It(Liang 等，2021)	886	87.2	34.62	0.9289	30.54	0.8463	29.20	0.8082	28.66	0.8624	33.98	0.9478
LBNet(Gao等,2022)	736	68.4	34.47	0.9277	30.38	0.8417	29.13	0.8061	28.42	0.8559	33.82	0.9460
ESRT(Lu等,2022)	770	96.4	34.42	0.9268	30.43	0.8433	29.15	0.8063	28.46	0.8574	33.95	0.9455
SwinIR-NG(Choi等，2023)	1190	114.1	34.64	0.9293	**30.58**	0.8471	29.24	0.8090	28.75	0.8639	34.22	0.9488
SRFormer-It（Zhou等,2023）	862	87.8	34.67	0.9296	30.57	0.8469	29.26	0.8099	28.81	0.8655	34.19	0.9489
BSTN(Bi等,2024)	736	72.2	34.68	0.9296	30.54	0.8451	29.23	0.8081	28.62	0.8604	–	–
MLFN-L（Song 等，2025）	1079	–	34.63	0.9292	30.55	0.8464	29.24	0.8077	28.69	0.8627	34.05	0.9481
SPSANet(Ours)	1071	108.8	34.67	0.9300	30.56	0.8484	29.27	0.8103	28.96	0.8689	34.47	0.9500
SPSANet+(Ours)	1071	108.8	**34.76**	**0.9305**	30.57	**0.8492**	**29.32**	**0.8111**	**29.10**	**0.8707**	**34.64**	**0.9510**

注:加粗和下划线字体分别表示各列最优和次优结果

少的增长。这主要是因为稀疏置换自注意力在扩大窗口的同时减少注意力权重的冗余,增大相似token权重的同时抑制不相关token的贡献,同时构建卷积模块弥补细节信息,并对局部和全局信息进行有效融合,提升模型对图像细节和整体结构的感知能力。对于引入自集成策略的SPSANet+,性能会进一步提升,这些结果均证明本文提出的方法是高效的。

3.2.2　主观实验结果

图5展示了在Urban100数据集上不同放大倍率下的主观重建结果。可以观察到,传统的轻量化方法如CARN(Ahn等,2018)和IMDN(Hui等,2018)在细节恢复方面存在明显不足,容易产生模糊或伪影;SwinIR-light(Liang 等，2021)与 NGSwin(Choi 等，2023)在纹理和结构保持上有一定改进,但仍存在细节缺失和边缘不够锐利的问题。相比之下,本文方法在多个样例中均能重建出更清晰的纹理和更准确的结构边缘,例如在img012的建筑窗格、img044的吊顶格栅以及img024的栏杆阴影等场景中,本文方法能够有效抑制模糊、恢复出连续且平滑的线条结构,视觉效果更接近真实图像GT。这些结果表明,所提出的SPSANet模型在保持轻量化的同时,能够显著提升复杂纹理与高频细节的重建能力,主观视觉质量明显优于现有SR方法。

3.3　消融实验

本节进一步评估所提方法中各组件的作用,并与基线模型进行对比。为确保公平性,我们在与所

提SPSANet相同的训练设置下,对所有模型在×2 SR任务上进行了训练,总共迭代250K次。实验采用Urban100数据集作为测试集,因为该数据集包含丰

富的结构信息,能够充分验证模型的重建能力。本文以SRFormer(Zhou等,2023)作为基线模型进行实验,如表4所示,在基线模型的基础上加入稀疏置换自注意力模块。该模块能够在不增加参数量和计算量的前提下,有效提升模型性能。在相同训练与测试设置下,模型在测试集上的PSNR提升了

表 3　不同方法的4倍超分辨率客观评价结果

Table 3　Results of objective results of ×4 super-resolution for different methods

方法	参数量/k	Mult-Adds/G	Set5		Set14		BSD100		Urban100		Manage109	
			PSMR	SSIM	PSMR	SSIM	PSMR	SSIM	PSMR	SSIM	PSMR	SSIM
EDSR(Lim等,2017)	1518	114.0	32.09	0.8938	28.58	0.7813	27.57	0.7357	26.04	0.7849	30.35	0.9067
CARN（Ahn 等，2018）	1592	90.0	32.13	0.8937	28.60	0.7806	27.58	0.7349	26.07	0.7837	30.47	0.9084
IMDN(Hui等,2019)	752	40.9	32.21	0.8948	28.58	0.7811	27.56	0.7353	26.04	0.7838	30.45	0.9075
LatticeNet（Luo 等，2020）	777	43.6	32.18	0.8943	28.61	0.7812	27.57	0.7355	26.14	0.7844	–	–
SwinIR-It(Liang 等，2021)	897	49.6	32.44	0.8976	28.77	0.7858	27.69	0.7406	26.47	0.7980	30.92	0.9151
LBNet(Gao等,2022)	742	38.9	32.29	0.8960	28.68	0.7832	27.62	0.7382	26.27	0.7906	30.76	0.9111
ESRT(Lu等,2022)	751	67.7	32.19	0.8947	28.69	0.7833	27.69	0.7379	26.39	0.7962	30.75	0.9100
SwinIR-NG(Choi等，2023)	1201	63.0	32.44	0.8980	<u>28.83</u>	0.7870	27.73	0.7418	26.61	0.8010	31.09	0.9161
SRFormer-It（Zhou等,2023）	873	52.7	**32.51**	0.8988	28.82	0.7872	27.73	0.7422	26.67	0.8032	31.17	0.9165
BSTN(Bi等,2024)	751	41.8	<u>32.50</u>	0.8985	28.78	0.7852	27.71	0.7403	26.49	0.7965	–	–
MLFN-L（Song 等，2025）	1071	–	32.37	0.8975	**28.85**	0.7864	27.73	0.7410	26.52	0.7985	30.95	0.9153
SPSANet(Ours)	1082	65.1	32.39	<u>0.8992</u>	28.70	<u>0.7888</u>	<u>27.74</u>	<u>0.7432</u>	<u>26.80</u>	<u>0.8083</u>	<u>31.33</u>	<u>0.9183</u>
SPSANet+(Ours)	1082	65.1	**32.51**	**0.9002**	28.72	**0.7899**	**27.79**	**0.7441**	**26.91**	**0.8103**	**31.56**	**0.9202**

注:加粗和下划线字体分别表示各列最优和次优结果

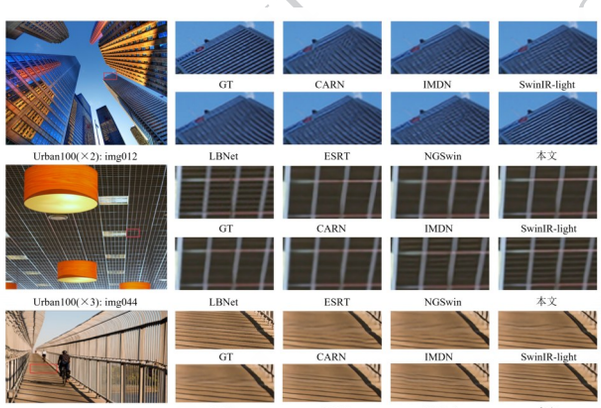


图 5　Urban100上不同放大倍数超分辨率可视化对比

Fig. 5　Comparison of super-resolution visualization at different magnifications on Urban100

0.11dB,表明所提出的稀疏机制能够在不牺牲效率的情况下提高特征表达能力。为了进一步探究性能提升的来源,我们分析了稀疏机制是否通过减少注意力权重冗余、增强注意力集中性来实现性能改进。我们计算注意力权重的熵来衡量注意力集中程度(Ghader等,2017),定义如下:

$$\text{Entropy}_{\text{attn}} = -\frac{1}{B}\sum_{b}\frac{1}{H}\sum_{i,j}\textbf{att}\,\boldsymbol{n}_{i,j}^{b,h}\log\textbf{att}\,\boldsymbol{n}_{i,j}^{b,h} \quad (4)$$

其中 $\textbf{att}\,\boldsymbol{n}_{i,j}^{b,h}$ 表示在第 b 个batch、第 h 个注意力头中查询 i 与键 j 之间的注意力得分。较低的熵值意味着注意力更集中在少数关键区域,而较高的熵值则说明模型从较多位置提取信息,注意力较为分散。我们在Urban100数据集上对100张测试图像进行了统计分析,结果如图6所示。可以观察到,稀疏置换注意力的注意力熵值整体呈逐层递减趋势,说明随着网络层数的加深,模型的注意力逐渐聚焦于少量具有判别性的区域。这一特性有助于模型在高层语义特征提取中更加聚焦关键结构,从而提升重建质量。相比之下,基线模型的注意力熵值显著更高,说明其在特征提取过程中更均匀地关注所有token,导致模型可能受到无关区域的干扰,从而影响整体性能。

如表4所示,在模型中加入高频信息增强模块

表 4　不同模型在Urban100(×2)数据集的结果对比

Table 4　Comparison of result with different model settings on Urban100(×2) dataset

Baseline	稀疏置换注意力	高频信息增强	拼接加卷积	双分支特征融合	参数量/k	Mult-Adds/G	PSNR	SSIM
√					853	197.5	32.52	0.9322
√	√				853	197.5	32.63	0.9332
√	√	√	√		1218	284.0	32.74	0.9340
√	√	√		√	1063	247.0	32.79	0.9343

后,测试集上的PSNR值得到了显著提升。这表明该模块能够有效强化图像中的高频成分,使模型在细节重建和边缘恢复方面具备更强的表达能力。为了进一步分析该模块对特征提取过程的影响,我们对Transformer分支与卷积分支提取的中间特征进行了可视化。具体做法是:对网络中所有层的特征图进行平均处理。可视化结果如图7所示。从图中可以观察到,卷积分支的特征图颜色更亮,局部细节更加锐利和清晰,边缘和纹理也更加突出,特征分布更加集中在局部区域;Transformer分支的特征图颜色更暗,特征分布更加平滑均匀,显示出更好的全局一致性。结合两者可以发现,Transformer分支与卷积分支在特征建模上具有明显的互补性。总体而言,高频信息增强模块提升了模型的高频响应能力,使网络在图像超分辨率任务中能够生成更加清晰、自然且细节丰富的重建结果。

本文进一步测试并比较了不同特征融合策略对模型性能的影响。如表4所示,采用双分支特征融合机制后,PSNR较传统的先拼接再使用卷积恢复通道数的方式提升了0.05dB,同时模型的参数量与计算量均有所下降。这充分验证了所提出融合策略的有效性与高效性。

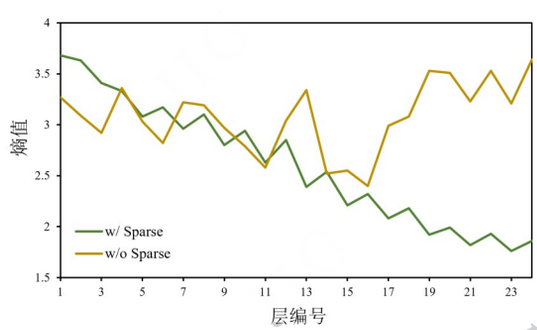
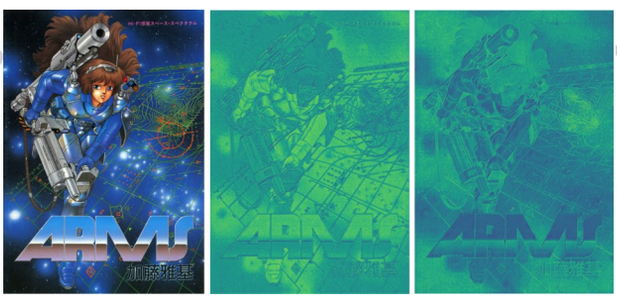


图 6　不同注意力机制各层的熵值对比

Fig. 6　Comparison of entropy values at each layer of different attention mechanisms



(a)低分辨率图像 (b)卷积分支特征图 (c)注意力分支特征图

((a)LR image;(b)feature map of conv;(c)feature map of transformer))

图 7　不同分支特征图对比

Fig. 7　Comparison of feature maps of different branches

3.4　LAM局部归因图分析

为了分析模型在图像重建过程中所利用的像素范围,我们采用LAM(Local Attribution Maps)进行可视化比较。如图8所示,与SwinIR-It相比,本文提出的SPSANet与SRFormer均表现出更高的扩散指数,因为我们在保证计算量与参数量不增加的前提下,通过扩大自注意力窗口有效提升了远程依赖的建模能力。进一步地,相较于SRFormer,本文的方法能够参考更广的像素区域,这得益于所设计的双分支特征融合模块。该模块中的通道交互机制可生成包含全局语义信息的通道注意力图,从而显著增强了模型的感受野。

4　结　论

本文提出了一种高效的稀疏置换自注意力机制,在扩大注意力窗口的同时,通过建立层间注意力的关联,有效减少了注意力权重的冗余,使模型能够更加聚焦于与重建任务最相关的token。在此基础上,我们进一步设计了高频细节增强模块来加强模型对图像纹理和边缘等高频信息的恢复能力。为了

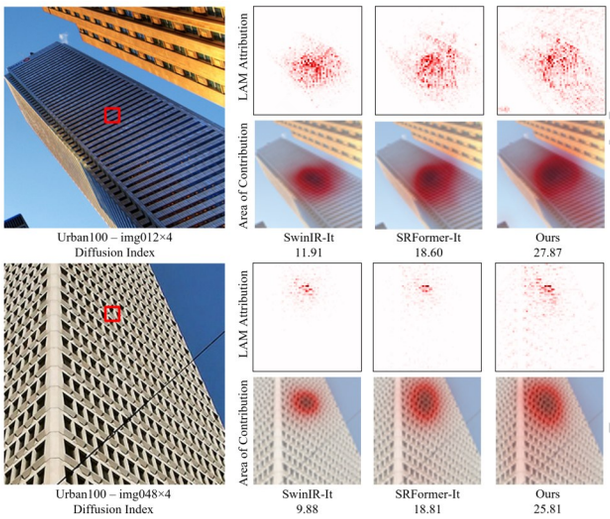


图 8　LAM 比较结果

Fig. 8　LAM comparison results

充分融合全局与局部特征，我们还提出了双分支特征融合模块，有效地解决了特征对齐问题，进一步提升了重建效果。大量实验结果表明，SPSANet 在轻量级图像超分辨率任务上取得了领先性能，同时重建图像在视觉效果上明显优于现有方法。特别是在结构复杂、纹理丰富的场景中，SPSANet 能更好地恢复细节与层次信息。

然而，我们也注意到，尽管 SPSANet 在重建质量上取得了显著提升，但在模型轻量化方面仍有进一步优化的空间。因此未来可以从以下几个方面展开，一是引入信息蒸馏机制，通过教师网络引导学生网络学习关键特征，以减少参数冗余；二是融合高效卷积结构以替代部分 Transformer 模块实现性能和效率的平衡；三是探索其他任务泛化能力，将本文的方法扩展到图像去噪、去模糊等其他低层视觉任务中，以验证其通用性和稳定性。

参考文献(References)

Ahn N, Kang B and Sohn K A. 2018. Fast, accurate, and lightweight super-resolution with cascading residual network// Proceed- ings of the 15th European Conference on Computer Vision. Munich, Germany: Springer:256-272[DOI:10.1007/978-3- 030-01249-6_16]

Bevilacqua M, Roumy A, Guillemot C and Alberi-Morel M L.2012. Low-complexity single image super-resolution based on nonnegative neighbor embedding//Proceedings of 2012 British Machine Vision Conference. Surrey, UK: BMVA Press: #135 [DOI: 10.5244/c. 26.135]

Bi X P, Chen S and Zhang L F.2024. Blueprint separable convolution Transformer network for lightweight image super-resolution. Journal of Image and Graphics, 29(04):0875-0889 (毕修平,陈实, 张乐飞. 2024. 轻量级图像超分辨率的蓝图可分离卷积 Transformer 网络. 中国图象图形学报, 29(04): 0875-0889) [DOI:10.11834/jig.230225]

Chen Q, Wu Q M, Wang J, Hu Q H, Hu T, Ding E, Cheng J and Wang J D, 2022. MixFormer: Mixing Features across Windows and Dimensions// Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE: 5239-5249 [DOI: 10.1109/ CVPR52688.2022.00518]

Chen X, Li H, Li M Q and Pan J S, 2023. Learning A Sparse Transformer Network for Effective Image Deraining// Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE: 5896-5905 [DOI: 10.1109/CVPR52729.2023.00571]

Chen Z, Zhang Y L, Gu J J, Kong L H, Yang X K and Yu F, 2023. Dual Aggregation Transformer for Image Super-Resolution// Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 12278-12287 [DOI: 10.1109/ICCV51070.2023.01131]

Choi H, Lee J and Yang J. 2023. N-Gram in swin Transformers for efficient lightweight image super-resolution//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 2071-2081 [DOI: 10.1109/CVPR52729.2023.00206]

Dong C, Loy C C, He K M and Tang X O. 2014. Learning a deep convolutional network for image super resolution// Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: IEEE:184-199 [DOI:10.1007/978-3- 319-10593-2_13]

Gao G W, Wang Z X, Li J C, Li W J, Yu Y and Zeng T Y. 2022. Lightweight bimodal network for single-image super-resolution via symmetric CNN and recursive Transformer//Proceedings of the 31st International Joint Conference on Artificial Intelligence. Vienna, Austria: IJCAI.org:913-919[DOI:10. 24963/ijcai.2022/128]

Ghader H and Monz C. 2017. What does Attention in Neural Machine Translation Pay Attention to? //Proceedings of the Eighth International Joint Conference on Natural Language Processing. Taipei, Taiwan: Asian Federation of Natural Language Processing: 30 – 39

He Z B, Dai T, Lu J, Jiang Y and Xia S T, 2020. Fakd: Feature-Affinity Based Knowledge Distillation for Efficient Image Super-Resolution// 2020 IEEE International Conference on Image Processing. Anchorage, Alaska, USA: IEEE: 518-522 [DOI: 10.1109/ICIP40778.2020.9190917]

Huang J B, Singh A and Ahuja N.2015. Single image super-resolution from transformed self-exemplars//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 5197-5206[DOI:10.1109/cvpr.2015. 72 99156]

Hui Z, Gao X B, Yang Y C and Wang X M, 2019. Lightweight Image Super-Resolution with Information Multi-distillation Net- work // Proceedings of the 27th ACM International Conference on Multimedia. New York, USA: Association for Computing Machinery: 2024-2032[DOI:10.1145 /3343 031. 3351084]

Hui Z, Wang X M and Gao X B, 2018. Fast and Accurate Single Image Super-Resolution via Information Distillation Network// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT: IEEE: 723-731[DOI: 10.1109/CVPR.2018.00082]

Kim J, Lee J K, and Lee K M, 2016. Accurate Image Super-Resolution Using Very Deep Convolutional Networks// Proceedings of 2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE: 1646-1654 [DOI: 10.1109/CVPR.2016.182]

Lee D, Yun S and Ro Y, 2025. Emulating Self-attention with Convolution for Efficient Image Super-Resolution// Proceed- ings of 2025 IEEE/CVF International Conference on Com- puter Vision. Honolulu, Hawai'i.

Li A, Zhang L, Liu Y and Zhu C, 2023. Feature Modulation Transformer: Cross-Refinement of Global Representation via High -Frequency Prior for Image Super-Resolution// Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 12480-12490 [DOI: 10.1109/ ICCV51070.2023.01150]

Li Y, Dong S H, Zhang J W, Zhao R and Zheng Y H.2025. Cross-scale Transformer image super-resolution reconstruction with fusion channel attention. Journal of Image and Graphics, 30(3): 0784-0797 (李焱, 董仕豪, 张家伟, 赵茹, 郑钰辉. 2025. 融合通道注意力的跨尺度Transformer图像超分辨率重建. 中国图象图形学报, 30(3):0784-0797)[DOI: 10.118 34/jig.240279]

Liang J Y, Cao J Z, Sun G L, Zhang K, van Gool L and Timofte R. 2021. SwinIR: image restoration using swin Transformer//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal, Canada: IEEE: 1833-1844[DOI:10.1109/ICCVW54120.2021.00210]

Lim B, Son S, Kim H, Nah S and Lee K M, 2017. Enhanced Deep Residual Networks for Single Image Super-Resolution// Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, HI, USA: IEEE: 1132-1140 [DOI:10.1109/CVPRW.2017.151]

Long W, Zhou X Y, Zhang L H, and Gu S H, 2025. Progressive Focused Transformer for Single Image Super-Resolution//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE: 2279-2288 [DOI: 10.1109/CVPR52734.2025.00218]

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//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans, USA: IEEE: 456-465 [DOI: 10.1109/ CVPRW56347.2022.00061]

Luo X T, Xie Y, Zhang Y L, Qu Y Y, Li C H and Fu Y, 2020. LatticeNet: Towards Lightweight Image Super-Resolution with Lattice Block//Proceedings of 2020 European Conference on Computer Vision. Cham: Springer: 272-289 [DOI: 10.1007/ 978-3-030-58542-6_17]

Mao M Y, Gao P, Zhang R R, and Zheng H H, Ma T, Peng Y, Ding E, Zhang B C and Han S M. 2021. Dual-stream Network for Visual Recognition// Proceedings of the 35st International Conference on Neural Information Processing Systems. Online: Curran Associates Inc: 25346-25358

Martin D, Fowlkes C, Tal D and Malik J.2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological stati- stics//Proceedings of the 8th IEEE International Conference on Computer Vision. Vancouver, Canada: IEEE: 416-423 [DOI: 10.1109/iccv. 2001.937655]

Matsui Y, Ito K, Aramaki Y, Fujimoto A, Ogawa T, Yamasaki T and Aizawa K. 2017. Sketch-based manga retrieval using manga- 109 dataset. Multimedia Tools and Applications, 76(20): 218 1 1-21838[DOI:10.1007/s11042-016-4020-z]

Peng Z L, Huang W, Gu S Z, Xie L X, Wang Y W, Jiao Y B and Ye Q X 2021. Conformer: Local Features Coupling Global Representations for Visual Recognition// Proceedings of 2021 IEEE/ CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE:357-366 [DOI:10.1109/ ICCV48922.2021.00042]

Song X G, Zhang P F, Liu W B, Lu X F and Hei X H.2025.Image super resolution reconstruction method based on multiscale large-kernel attention feature fusion network. Journal of Image and Graphics, 30(4):084-1099 (宋霄罡, 张鹏飞, 刘万波, 鲁晓锋, 黑新宏. 2025.多尺度大核注意力特征融合网络的图像超分辨率重建.中国图象图形学报, 30(4): 1084-1099)[DOI: 10.11834/jig.240 042]

Wang H, Chen X H, Ni B B, Liu Y T and Liu J F, 2023. Omni Aggregation Networks for Lightweight Image Super Resolution//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE: 22378-22387 [DOI:10.1109/CVPR52729.2023. 02143]

Wu H L, Li W Y and Zhang L B.2022. Cross-scale coupling network for continuous-scale image super-resolution. Journal of Image and Graphics, 27(05):1604-1615 (吴瀚霖, 李宛谕, 张立保. 2022. 跨尺度耦合的连续比例因子图像超分辨率. 中国图象图形学报, 27(05): 1604-1615)[DOI:10.11834/jig.210815]

Zeyde R, Elad M and Protter M.2012. On single image scale-up using sparse-representations//Proceedings of the 7th International Conference on Curves and Surfaces. Avignon, France: Springer:711-730 [DOI:10.1007/978-3-642-27413-8_47]

Zhang L H, Li Y, Zhou X Y, Zhao X R and Gu S H, 2024. Transcend-

ing the Limit of Local Window: Advanced Super-Resolution Transformer with Adaptive Token Dictionary//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE: 2856-2865 [DOI: 10. 1109/CVPR52733.2024.00276]

Zhang X, Zhang Y and Yu F, 2024. HiT-SR: Hierarchical Transformer for Efficient Image Super-Resolution //Proceedings of 2024 European Conference on Computer Vision. Milan, Italy: Springer: 483 - 500 [DOI: 10.1007/978-3-031-73661-2_27]

Zhang X D, Zeng H, Guo S, and Zhang L, 2022. Efficient Long-Range Attention Network for Image Super-Resolution//Proceedings of 2022 European Conference on Computer Vision. Tel Aviv, Israel: Springer: 649-667[DOI:10.1007/978-3-031-19790-1_ 39]

Zhou Y P, Li Z, Guo C L, Bai S, Chen M M and Hou Q B. 2023. SRFormer: Permuted Self-Attention for Single Image Super- Resolution//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 12734-12745 [DOI: 10.1109/ICCV51070.2023.01174]

作者简介

吴思琪，女，硕士研究生，主要研究方向为深度学习图像处理。E-mail: 2024023468@m.scnu.edu.cn

柳薇，通信作者，女，副教授，主要研究方向为多媒体信息处理、机器学习。E-mail: liuwei@m.scnu.edu.cn

陈卫东，男，教授，主要研究方向为图论与复杂网络、组合优化、算法与计算复杂性。Email: chenwd@m.scnu.edu.cn