# 基于Transformer的块内块间双聚合的单图像超分辨率重建网络

唐 述 曾琬凌 杨书丽 钟恒飞 陈 卓

（重庆邮电大学计算机科学与技术学院计算机网络和通信技术重庆市重点实验室 重庆 400065）

**摘 要** 近年来，基于深度学习的轻量级单幅图像超分辨率（Single Image Super-Resolution，SISR）重建网络已成为人们研究的热点.但是现有的轻量级方法在捕捉图像像素间长距离的全局依赖性方面存在显著局限，这主要是由于显式建模此类依赖关系所伴随的庞大计算复杂度所致.因此现有的轻量级SISR方法的性能仍有较大的提升空间.基于此，本论文提出了一种新颖的基于Transformer的块内块间双聚合的轻量级网络（Intra-block and Inter-block Dual Aggregation Network，IIDAN）来显式捕捉整幅图像中的全局依赖性，进而实现高质量的SISR.首先，在自然图像的非局部结构相似性的启发下，本论文提出了一种新颖的块内块间Transformer模块（Intra-block and Inter-block Transformer Module，IITM）.IITM通过交替地开发每个图像块内部的自注意力和不同图像块之间的自注意力实现了图像中局部特征的显式捕捉和图像中结构相似性的全局显式捕捉.其次，本论文还提出了一种信息交互机制（Information Interaction Mechanism，IIM）来分别对IITM中的两种自注意力进行对应信息的互补：IIM给块内自注意力（Intra-block Transformer，Intra-T）补充块间信息，使得Intra-T能够获得更多的全局结构信息；同时，IIM也给块间自注意力（Inter-block Transformer，Inter-T）补充局部信息，使得Inter-T能够获得更多的局部细节信息.实验结果表明，与近几年极具代表性的轻量级SISR方法相比，本论文提出的IIDAN能够重建出更高质量的超分辨率图像，同时具有更低的计算复杂度.

**关键词** 单幅图像超分辨率；轻量级；Transformer；全局的结构相似性；信息交互

**中图法分类号** TP391 **DOI号** 10.11897/SP.J.1016.2024.02783

# Intra-Block and Inter-Block Dual Aggregation Transformer for Single Image Super-Resolution

TANG Shu ZENG Wan-Ling YANG Shu-Li ZHONG Heng-Fei CHEN Zhuo

（*Chongqing Key Laboratory of Computer Network and Communications Technology*，*College of Computer Science and Technology*，*Chongqing University of Posts and Telecommunications*，*Chongqing* 400065）

**Abstract** The single Image super-resolution (SISR) reconstruction task is an ill-posed and challenging inverse problem, which is the research hotspot in low-level computer vision tasks. SISR attempts to reconstruct a clean high-resolution (HR) image with rich and natural texture details from its low-resolution (LR) version, which is crucial in various computer vision fields. Recently, lightweight networks for SISR have increased in popularity, and numerous lightweight SISR networks have been proposed for various practical applications. The landscape of deep learning has witnessed a significant surge in interest in lightweight SISR techniques, which have

收稿日期：2024-01-07；在线发布日期：2024-09-14. 本课题得到国家自然科学基金项目（No. 61601070）、重庆市自然科学基金面上项目（CSTB2023NSCQ-MSX0680）、重庆市教育委员会科学技术研究重大项目（KJZD-M202300101）、重庆邮电大学博士研究生创新人才项目（BYJS202217）资助.**唐 述**（通信作者），博士，副教授，中国计算机学会（CCF）会员，主要研究领域为低水平视觉任务、图像超分辨率重建、模糊图像复原. E-mail：tangshu@cqupt.edu.cn. **曾琬凌**（通信作者），硕士，主要研究领域为图像处理和深度学习，E-mail：S210231010@stu.cqupt.edu.cn. **杨书丽**，博士研究生，主要研究领域为图像超分辨率重建和深度学习. **钟恒飞**，硕士研究生，主要研究领域为计算机视觉和深度学习. **陈 卓**，硕士研究生，主要研究领域为图像处理和深度学习.

proven to be powerful tools for the enhancement of image quality. Despite their potential, these techniques often encounter a critical challenge: the difficulty in capturing the intricate, long-range interdependencies between pixels within an image. This limitation, primarily due to computational constraints, restricts the full realization of the capabilities of lightweight SISR algorithms, indicating a substantial area for improvement. In response to this challenge, we introduce a pioneering solution with the development of the Intra-block and Inter-block Dual Aggregation Network (IIDAN), a transformer-based, lightweight network architecture. Carefully designed, the IIDAN framework is engineered to explicitly capture the global dependencies that exist within images, thereby significantly enhancing the quality of SISR results. Our innovation is anchored in the understanding of the inherent non-local structural similarities present in natural images. Building upon this insight, we have crafted the Intra-block and Inter-block Transformer Module (IITM), a novel module that adeptly manages self-attention mechanisms at two distinct levels. The first level operates within a single block, referred to as the intra-block transformer (Intra-T), while the second level functions across different blocks, known as the inter-block transformer (Inter-T). By seamlessly alternating between these two attention mechanisms, the IITM integrates the extraction of complex local features with the recognition of broad global structural patterns, providing a comprehensive analysis of the image. Moreover, we have introduced the Information Interaction Mechanism (IIM) as a strategic enhancement to the IITM. This mechanism intelligently blends the strengths of intra-T and inter-T, using insights from inter-block information to enrich intra-block attention. This approach not only expands the scope of structural understanding but also ensures that a broader perspective does not compromise the detailed understanding of fine-grained details. Simultaneously, the inter-block attention is reinforced by the detailed local information from intra-block attention, ensuring a balanced and holistic approach to image analysis. The effectiveness of our IIDAN methodology is evidenced by a series of experiments. These experiments demonstrate that IIDAN not only stands its ground but also surpasses the most respected lightweight SISR methods of recent times. Our framework commendably strikes a balance between minimal parameterization and reduced computational complexity, consistently producing super-resolution images of exceptional quality. This achievement is a testament to the innovative design and meticulous implementation of IIDAN. In conclusion, the IIDAN presents a solution that is both computationally efficient and capable of generating high-fidelity super-resolution images. Its dual attention mechanism, complemented by the strategic Information Interaction Mechanism, positions the IIDAN as a leading contender in the pursuit of superior image quality enhancement.

**Keywords** single image super-resolution; lightweight; Transformer; global structural similarity; information interaction

## 1 引 言

单幅图像超分辨率(Single Image Super-Resolution, SISR)重建旨在从低分辨率(Low-Resolution, LR)图像中重建出对应的高分辨率(High-Resolution, HR)图像. SISR是一种非常具有挑战性且严重的病态问题,为了能够有效解决这一病态问题,人们提出了各种各样的方法. 在早期的研究中,研究人员们提出了基于插值、基于稀疏表示和基于正则化等一系列非网络的SISR方法. 在基于插值的方法中,2019年,Zheng等人[1]提出了一种基于加权的直接非线性回归的图像插值超分辨率方法. 该方法由于涉及大量训练图像块的聚类处理,因此对聚类算法的要求较高. 鉴于实际场景中复杂的非线性映射,Zheng等人分别从分类和回归两个角度来执行

SISR任务. 在基于稀疏表示的方法中，2021年，Peng等人[2]提出了一种基于全局梯度稀疏性、非局部低秩张量分解和超拉普拉斯先验的超光谱图像(Hyperspectral Image，HSI)超分辨率方法. 虽然该方法能够更好地捕捉HSI的空间和光谱相似性，但是该方法所需的数据量较大，计算复杂性较高. 2024年，Liao等人[3]提出了一种最小凹面惩罚的SISR模型(Minimax Concave Penalty Super-Resolution，MCPSR). MCPSR通过将最小凹面惩罚(minimax concave penalty，MCP)引入到图像超分辨率重建任务中来消除偏差，使得重建的图像更加准确和真实. 但是MCPSR同样存在计算资源消耗较大的问题. 在基于正则化的方法中，2020年，Li等人[4]提出了一种基于自适应范数的非局部自相似性正则化器. 该方法能够有效利用非局部自相似系数与单一自相似系数间的冗余信息来改善图像重建的质量. 然而该方法引入了较多的正则化器，不仅增加了超分辨率(Super-Resolution，SR)的复杂度，而且还会在迭代过程中造成计算误差的扩散，从而影响超分辨率图像的准确性. 同年，Tang等人[5]开发了一种基于联合正则化约束的图像超分辨率重建方法. 该方法的核心在于结合了1×1和2×2两种不同投影采集模式的系统矩阵来构建保真项，并利用块匹配和全变分(Total Variation，TV)正则化器来挖掘图像中的稀疏特性. 尽管早期的方法在处理SISR问题上取得了一定的成功，但是它们在性能、参数量和计算复杂度的最优化折中方面仍然存在较大的提升空间. 近年来，深度学习神经网络凭借强大的学习和拟合能力，已成为SISR的主流研究方法，尤其是基于卷积神经网络(Convolutional Neural Network，CNN[6])的SISR方法[7-11]和基于Transformer的SISR方法[12-16]. 例如，一些工作采用很深和很宽的卷积层和残差连接来提取LR图像中的局部信息，重建出较高质量的SR图像[10,17-19]. 还有一些工作通过开发全局的自注意力(Self Attention，SA)[17,20-22]来增强网络的表达能力. 虽然以上的基于CNN的方法[7-11]和基于Transformer的方法[12-16]能够显著提升SISR的性能，但是它们的参数量和计算复杂度都极其巨大，严重限制了这些方法在实际场景中的应用.

为了实现SISR在尽可能多的实际场景中的应用，尤其是在资源受限的边缘端设备中，轻量级的SISR网络已经成为人们研究的热点. 其中，基于CNN的方法[23-26]几乎都是通过减少卷积层的数量或残差块的数量，以及采用递归方式或参数共享等策略[8-9,27]来达到减少模型参数量的目的. 而基于Transformer的方法则几乎都是通过将SA限制在一个特定大小的窗口内而非整幅图像，并通过滑动窗口来达到降低计算复杂度的目的[12,15,28-33]. 然而，现有的轻量级SISR方法虽然能够显著降低模型的参数量和计算复杂度，但是它们都仅能显式捕捉局部/区域范围内的相互依赖性，而并不能显式地捕捉整幅图像范围内的全局依赖性，因为显式地捕捉整幅图像范围内的全局依赖性会带来巨大的计算代价. 因此现有的轻量级SISR方法的性能仍有较大的提升空间.

基于以上的分析，针对现有方法存在的缺陷，本论文提出了一种新颖的基于Transformer的块内块间双聚合轻量级网络(Intra-block and Inter-block Dual Aggregation Network，IIDAN). 本文提出的IIDAN能够在保证轻量级的基础上显式地捕捉整幅图像范围内的全局依赖性. 特别的，本论文首先提出了一种新颖的块内块间Transformer模块(Intra-block and Inter-block Transformer Module，IITM). IITM通过交替地开发每个图像块内部的自注意力(Intra-block Transformer，Intra-T)和不同块之间的自注意力(Inter-block Transformer，Inter-T)实现了对图像中局部特征相似性的显式捕捉和整幅图像范围内结构相似性的全局显式捕捉. 其次，本论文还提出了一种信息交互机制(Information Interaction Mechanism，IIM)来分别对IITM中的两种自注意力进行对应信息的补充：IIM给块内自注意力(Intra-T)补充块间信息，使得Intra-T能够获得更多的全局结构信息；同时，IIM也给块间自注意力(Inter-T)补充局部信息，使得Inter-T能够获得更多的局部细节. 综上所述，本论文提出的IIDAN的主要贡献如下：

(1)在自然图像非局部结构自相似性的启发下，本论文提出了一种新颖的Inter-T来实现全局范围内结构信息的显式捕捉和建模. 在提出的Inter-T中，本论文采用了大感受野的深度可分离卷积来提取每个图像块的结构信息，使得整幅图像中所有的结构信息被统计到一个更低的维度空间，因此全局范围内不同块之间的自注意力的空间复杂度和时间复杂度将会极大地降低. 也正因如此，本论文提出的IIDAN能够同时实现整幅图像的全局范围内相互依赖性的显式捕捉和轻量级.

(2)本论文还提出了一种IIM来分别对Intra-T

和Inter-T进行块间信息和局部信息的补充，从而进一步增强网络的特征捕捉和表达能力，有助于更高质量的SR重建.

(3)实验结果表明，与近几年极具代表性的轻量级SISR方法相比，本论文提出的IIDAN能够重建出更高质量的超分辨率图像，同时具有更低的计算复杂度.

## 2 相关工作

自2014年Dong等人[34]首次将CNN引入到SISR任务中以来，深度学习神经网络便凭借其强大的学习和拟合能力成为了图像超分辨率重建领域中最受欢迎的方法之一. 接下来，本论文就将对近年来极具代表性的SISR方法进行详细的论述.

### 2.1 经典的SISR方法

早期，研究者通过增加卷积层的数量和引入注意力机制等方式来提升SR重建的性能. 2018年，Zhang等人[10]通过运用大量的残差块和跳跃连接，提出了一种残差中的残差网络结构(Residual in Residual，RIR)和一种基于全局平均池化的通道注意力机制来实现单幅图像的超分辨率重建. 近年来，自注意力，也被称为非局部注意力(Non-local Attention，NLA)，迅速成为SISR领域中人们研究的热点[17,19-21,35-36]. 2020年，Zhou等人[17]利用非局部图卷积聚合模块，巧妙地为每个LR图像块找到多个HR图像块，并构建出LR-HR连接图，提出了一种跨尺度的图卷积SR网络. 2021年，Mei等人[19]提出了一种非局部的稀疏注意力(Non Local Sparse Attention，NLSA)模块来显式地捕捉较大范围内的特征相似性. 2023年，Mei等人[20]提出了一种新颖的金字塔注意力模型用于图像复原. Mei等人利用分块匹配的自注意力操作来获取不同尺度上的依赖关系. 2023年，Zhou等人[22]提出了一种多尺度共享的图像超分辨率方法(Multi-scale Shared Representation Acquisition，MSRA). Zhou等人设计了一种跨尺度匹配的自注意力卷积滤波器来捕捉图像中的多尺度特征. Xia等人[35]提出了一种基于核函数逼近和对比学习的SISR模型. Yang等人[36]将多尺度编码信息嵌入到注意力机制中，提出了一种多特征自注意力超分辨率网络.

虽然上述的方法能够重建出高质量的SR图像，但是它们的参数量和计算复杂度都极其巨大，严重限制了这些方法在实际场景中的应用，尤其是在资源受限的设备中.

### 2.2 轻量级的SISR方法

近年来，轻量级的SISR网络已逐渐成为人们研究的热点. 2022年，Chen等人[8]提出了一种基于多尺度递归反馈的轻量级SISR网络. Chen等人将递归学习用于多尺度投影组，利用高层次信息对低层次信息进行修正，能够有效细化浅层的特征. 2024年，Liu等人[9]提出了一种深度递归残差信道注意力网络. 该网络创建了一种通道特征融合模块，能够在有效融合不同特征层的同时减少网络的参数量. Ahn等人[23]通过参数共享策略提出了一种高效的残差块并构建了一种能够应用于移动场景的轻量级SISR网络(Cascading Residual Network，CARN). 2022年，Li等人[24]提出了一种轻量级的蓝图可分离残差网络. Li等人利用蓝图可分离卷积和一种高效的注意力模块来增强网络的表达能力. 2023年，Liu等人[27]提出了一种深度递归多尺度特征融合网络和一种渐进式特征融合技术来逐步利用多个尺度的特征. Hui等人[37]通过堆叠多个信息蒸馏模块，提出了一种轻量级的信息多蒸馏网络(Information Multi-distillation Network，IMDN). 2020年，Luo等人[38]提出了一种轻量级的图像超分辨率网络(Network with Lattice Block，LatticeNet)。Luo等人通过采用残差块、注意机制，以及反向特征融合策略来减少网络的参数量. 2022年，Kong等人[39]提出了一种改进的特征提取器和一种新颖的多阶段热启动训练策略，并创建了一种新颖的残差局部特征网络来高效复原图像的边缘和细节. 2023年，Park等人[40]通过构建块之间的残差自动连接，提出了一种适用于SISR的轻量级动态残差自注意力网络(Dynamic Residual Self-attention Network，DRSAN). 同年，Xie等人[41]开发了一种大核蒸馏块和一种大核注意力机制. 2024年，Zhang等人[42]开发了一种轻量级的稀疏注意力特征融合模块.

因为显式地开发了全局范围内的相互依赖特性，Transformer在高水平的视觉任务中获得了巨大的成功[28,43-47]，但也正因为全局范围内相互依赖性的显式开发，导致Transformer的计算负担极其巨大，因此，Transformer很难被直接应用到轻量级的SISR中. 2021年，Liang等人[12]提出了一种轻量级的Transformer：SwinIR来实现SISR. Liang等人通过仅在一个指定大小的窗口内进行相互依赖性的显式开发来达到减少计算负担的目的，并通过滑动窗口策略来隐式地捕捉全局信息. 2022年，Zhang等

人[15]提出了一种轻量级且高效力的SR长距离注意力网络(Lightweight Efficient Long-range Attention Network，ELAN-light). ELAN-light由移位卷积和分组多尺度注意力模块组成，其中的分组多尺度注意力模块将特征张量按通道划分成不同的组，每个组取不同大小的窗口，然后分别计算不同窗口内的自注意力. 因此ELAN-light仅能隐式地获取全局范围内的相互依赖性. 2022年，Lu等人[16]提出了一种快速且精确的轻量级SR网络(Efficient Super-resolution Transformer，ESRT). 虽然ESRT采用了CNN+Transformer的双骨干结构，但是该Transformer骨干仍然只考虑了区域范围内的相互依赖性. 2023年，Zhou等人[29]提出了一种基于维度置换的超分辨率网络(Super-resolution Transformer，SRFormer). SRFormer通过将空间维度的信息与通道维度的信息进行置换来降低计算复杂度. 同年，Wang等人[30]提出了一种轻量级的全方位聚合的SR网络(Omni for Super-resolution，Omni-SR). 在Omni-SR中，Wang等人提出了一种基于密集交互原理的全域自注意(Omni Self-attention，OSA)块来从空间和通道两个维度对像素进行交互的建模，以此来探索空间和通道之间的潜在相关性. 2024年，Wang等人[32]提出了一种内容感知混合器(Content-aware Mixer，CAMixer)来对图像中的不同成分进行分而治之的处理. 同年，Zamfir等人[33]提出了一种高效的图像超分辨率模型(Spstial Enhancement Expertise with a Mixture of Low-rank Experts，SeemoRe). SeemoRe通过对不同层进行不同信息提取的策略来提升SR性能. Zou等人[48]将高效的Transformer引入到轻量级的SISR任务中，提出了一种高效的特征传播策略. 2023年，Gu等人[49]将后向融合模块和递归Transformer结合到一起，提出了一种轻量级的SISR双分支网络(Dual Branch Network，DBNet). 2023年，Sun等人[50]提出一种轻量级的高效网络(Spatially-adaptive Feature Modulation Network，SAFMN). SAFMN在一个类似Transformer的模块中开发了一种空间自适应特征调制机制，并采用多尺度策略来获取不同尺度下的感受野，从而间接的提取长距离的特征. 同年，Chen等人[51]提出了一种面向SISR的多尺度余弦注意力Transformer网络(Multi-scale Cosine Attention Transformer Network，MCATN). 在MCATN中，Chen等人提出了一个残差多尺度Transformer群来捕获局部的特征信息，并通过多尺度对远程依赖关系进行建模.

通过上述的深入分析，我们不难发现，无论是基于CNN的方法[8-9,19,23,27-28,35-37,39-42]，还是基于Transformer的方法[12,15-16,28-30,32-33,48-50]，尽管这些轻量级SISR模型在减少模型参数量和降低计算复杂度方面展现出显著优势，但它们共同受限于仅能有效捕捉局部或区域内部的相互依赖关系，而未能实现对整幅图像全局范围内相互依赖性的显示建模. 这一局限性揭示了当前轻量级SISR方法在性能上仍有待进一步挖掘与提升的空间，迫切需要新的策略来克服这一障碍，以实现更为全面和高效的图像超分辨率重建.

## 3　本论文提出的轻量级IIDAN

本章节将对提出的轻量级IIDAN进行详细的论述. 首先介绍IIDAN的总体框架，然后详细介绍本论文提出的块内块间Transformer模块(IITM).

### 3.1　IIDAN的总体框架

本文提出的IIDAN的总体网络框架如图1所示，主要包括三个部分：浅层特征提取层(Shallow feature exaction)、深度特征提取层(Deep feature exaction)和图像重建部分(Image reconstruction). 首先将一幅LR图像，$\boldsymbol{I}^{LR}\in \mathrm{R}^{H\times W\times 3}$，输入到浅层特征提取层(即一个3×3的卷积层)提取该LR图像的浅层特征，得到$\boldsymbol{F}_s\in \mathrm{R}^{H\times W\times C}$($H$和$W$分别代表输入LR图像的高和宽，$C$代表特征通道的数量)：

$$\boldsymbol{F}_s = H_{conv3\times 3}\left(\boldsymbol{I}^{LR}\right) \tag{1}$$

其中，$H_{conv3\times 3}(\cdot)$代表3×3的卷积层. 然后$\boldsymbol{F}_s$将作为深度特征提取层的输入. 深度特征提取层是由$M$个残差块内块间Transformer组(Residuual Intra-Inter Transformer Group，RIITG)和一个1×1的卷积层组成，深度特征提取的公式为：

$$\boldsymbol{F}_m = H_{RIITG}\left(\boldsymbol{F}_{m-1}\right),\ m=1,2,\cdots,M \tag{2}$$

$$\boldsymbol{F}_d = H_{conv1\times 1}\left(\boldsymbol{F}_M\right)+\boldsymbol{F}_s \tag{3}$$

其中，$\boldsymbol{F}_{m-1}$、$\boldsymbol{F}_m$分别代表的是第$m$个RIITG模块的输入和输出(共$M$个RIITG模块). $H_{RIITG}(\cdot)$代表的是RIITG模块函数，$\boldsymbol{F}_d\in R^{H\times W\times C}$表示深度特征提取层的最终输出.

对于RIITG模块而言，每个RIITG模块中又包含了$N$个块内块间Transformer模块(IITM). 第$m$个RIITG的公式可表示为：

$$\boldsymbol{F}_{m,n} = H_{IITM}\left(\boldsymbol{F}_{m,n-1}\right),\ n=1,2,\cdots,N \tag{4}$$

$$F_m = H_{conv1\times1}(F_{m,N}) + F_{m-1} \tag{5}$$

其中，$H_{conv1\times1}(\cdot)$代表的是卷积核为1×1的卷积操作，$F_{m,n-1}$和$F_{m,n}$分别代表的是第$m$个RIITG模块中第$n$个IITM模块的输入和输出，$N$是每个RIITG模块中IITM模块的数量. $H_{IITM}(\cdot)$代表的是IITM模块函数. 同时，为了保证训练的稳定性，本论文采用了残差连接.

如图1所示，每个IITM模块中串联了两个Transformer模块：块内自注意力(Intra-T)和块间自注意力(Inter-T)能够分别实现对图像中局部特征相似性的显式捕捉和整幅图像范围内结构相似性的全局显式捕捉.

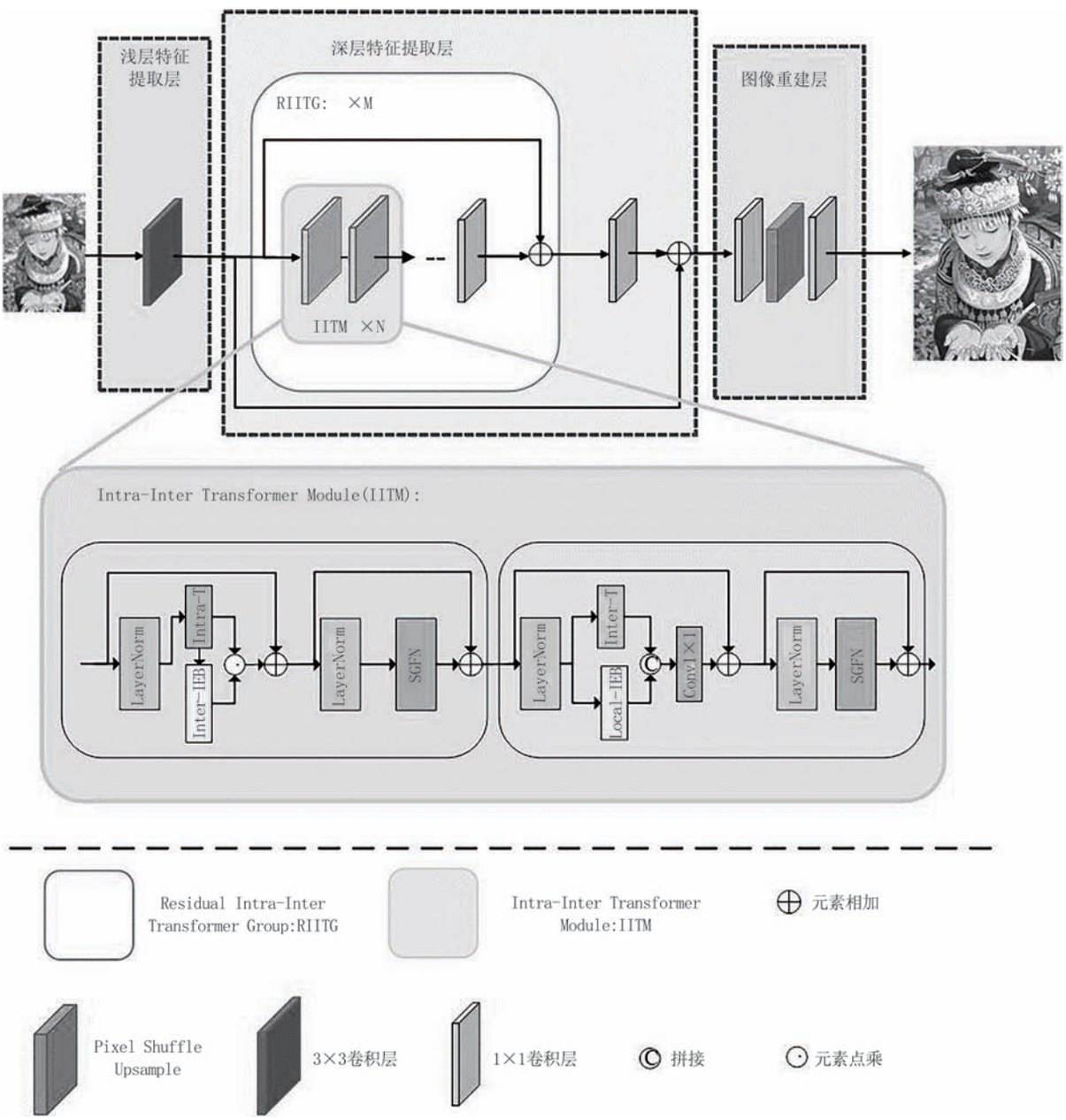



图1 本论文提出的IIDAN总体框架

在经过了深度特征提取层之后，将深度特征$F_d$通过图像重建部分重建出高分辨率图像$I_{SR}\in R^{H_{out}\times W_{out}\times 3}$($H_{out}$和$W_{out}$分别表示重建的SR图像的高和宽). 在图像重建部分，本论文采用像素混洗方式(Pixel Shuffle[52])对深度特征$F_d$进行上采样，并采用卷积层进行特征聚合. 具体的过程可由公式表示为

$$I_{SR} = H_{conv1\times1}\left(H_{up}\left(H_{conv1\times1}(F_d)\right)\right) \tag{6}$$

其中，$H_{up}(\cdot)$代表Pixel Shuffle[52]的上采样操作，$I_{SR}$表示IIDAN最终重建出的SR图像. 本论文采用平

均绝对误差损失(MAE LOSS)函数来优化网络的参数，该函数定义为

$$\mathcal{L}=\| \boldsymbol{I}_{SR}-\boldsymbol{I}_{HR} \|_1 \tag{7}$$

其中，$\boldsymbol{I}_{HR}$表示的是真实的高分辨率图像，$\|\cdot\|_1$代表的是$L_1$范数. 通过以上对IIDAN总体框架的论述可知，深度特征提取层中的IITM是IIDAN最重要的部分，也是本论文最主要的贡献和创新. 因此，接下来就将对深度特征提取层中的IITM进行详细论述.

### 3.2　块内块间Transformer模块(IITM)

如图1所示，在本论文提出的IITM中，首先通过交替地执行Intra-T和Inter-T来同时实现局部特征相似性和全局结构相似性的显式捕捉. 然后，通过一种IIM来分别对Intra-T和Inter-T进行块间信息和局部信息的补充.

#### 3.2.1　块内自注意力(Intra-T)

对于Intra-T，本论文采用Swin Transformer的思想来显式地获取指定窗口范围内的局部信息. 如图2(a)所示，给定输入特征$\boldsymbol{X}\in R^{H\times W\times C}$，先将$\boldsymbol{X}$划分为$h\times w$大小的非重叠块(即：$h\times w$大小的窗口)，$\boldsymbol{X}^i\in R^{hw\times C}$，$i=1,2,\cdots,L_{intra}$($L_{intra}=\dfrac{HW}{hw}$，即划分的块数)，并将$\boldsymbol{X}^i$通过一个线性层生成查询、键、值矩阵(分别表示为$\boldsymbol{Q}^i_{intra}$、$\boldsymbol{K}^i_{intra}$、$\boldsymbol{V}^i_{intra}\in\boldsymbol{R}^{hw\times d}$，$h$和$w$分别表示每个块的高度和宽度，$d=\dfrac{C}{s}$是每个头分到的通道数，共$s$个头)，可以公式化为

$$\boldsymbol{Q}^i_{intra}=\boldsymbol{X}^i\boldsymbol{P}_{Q_{intra}}, \tag{8}$$

$$\boldsymbol{K}^i_{intra}=\boldsymbol{X}^i\boldsymbol{P}_{K_{intra}}, \tag{9}$$

$$\boldsymbol{V}^i_{intra}=\boldsymbol{X}^i\boldsymbol{P}_{V_{intra}} \tag{10}$$

其中$\boldsymbol{P}_{Q_{intra}}$、$\boldsymbol{P}_{K_{intra}}$、$\boldsymbol{P}_{V_{intra}}\in R^{C\times d}$分别代表的是不同块之间共享的投影矩阵. 然后，在每个块的内部做自注意力. 以$\boldsymbol{X}^i$为例，$\boldsymbol{X}^i$的块内自注意力$\boldsymbol{X}^i_{intra}\in R^{h\times w\times d}$的计算公式为

$$\boldsymbol{X}^i_{intra}=softMax\left(\frac{\boldsymbol{Q}^i_{intra}\left(\boldsymbol{K}^i_{intra}\right)^T}{\sqrt{d}}+B\right)\boldsymbol{V}^i_{intra} \tag{11}$$

其中，$B$是可学习的相对位置编码. 因为有$s$个头，因此公式(11)将并行执行$s$次的块内自注意力计算. 在完成了所有头的自注意力计算之后，将每个头的$\boldsymbol{X}^i_{intra}$从通道维度上进行拼接得到特征$\boldsymbol{A}^i_{intra}\in R^{h\times w\times C}$，再将每个$\boldsymbol{A}^i_{intra}$拼接回原来的位置得到Intra-T的输出特征$\boldsymbol{A}_{intra}\in R^{H\times W\times C}$：

$$\boldsymbol{A}_{intra}=concat\left(\boldsymbol{A}^1_{intra},\boldsymbol{A}^2_{intra},\cdots,\boldsymbol{A}^{L_{intra}}_{intra}\right) \tag{12}$$

其中，$concat(\cdot)$表示拼接操作。很明显，Intra-T仅将注意力集中在一个指定大小($h\times w$)的块内，因此Intra-T仅能显式捕捉局部的块内特征相似性.

#### 3.2.2　块间自注意力(Inter-T)

最近，研究人员们发现，在一幅图像中，图像块水平上的匹配会比像素级水平的匹配更能获得图像的结构信息[17,20-21]，而且具有更强的噪声鲁棒性. 因此，在自然图像非局部结构自相似性的启发下，本论文提出了一种新颖的块间Transformer(Inter-T)来显式捕捉整幅图像范围内全局的结构相似性.

如图2(b)所示，本论文提出的Inter-T同样采用了多头注意力($s$为头的数量). 首先将输入的特征通过Split操作在通道维度上均分成两部分：$\boldsymbol{X}_1\in R^{H\times W\times\frac{C}{2}}$和$\boldsymbol{X}_2\in R^{H\times W\times\frac{C}{2}}$，并仅将$\boldsymbol{X}_1$输入到Inter-T中. 然后对输入的特征$\boldsymbol{X}_1\in R^{H\times W\times\frac{C}{2}}$进行不重叠的分块得到特征$\boldsymbol{X}^j_1\in R^{L_{inter}\times p^2_{inter}\times\frac{C}{2}}$($j=1,2,\cdots,s$)，每个块的大小为$p_{inter}\times p_{inter}$，共$L_{inter}=\dfrac{HW}{p^2_{inter}}$个块，并将$\boldsymbol{X}^j_1$通过一个线性层，生成多头的查询、键、值(表示为$Q^j_{inter}$、$K^j_{inter}$、$V^j_{inter}\in R^{L_{inter}\times p^2_{inter}\times d_1}$，$d_1=\dfrac{C}{2s}$为每个头的通道数). $Q^j_{inter}$、$K^j_{inter}$、$V^j_{inter}$可以公式化为

$$\boldsymbol{Q}^j_{inter}=\boldsymbol{X}^j_1\boldsymbol{P}_{Q_{inter}}, \tag{13}$$

$$\boldsymbol{K}^j_{inter}=\boldsymbol{X}^j_1\boldsymbol{P}_{K_{inter}}, \tag{14}$$

$$\boldsymbol{V}^j_{inter}=\boldsymbol{X}^j_1\boldsymbol{P}_{V_{inter}} \tag{15}$$

其中，$\boldsymbol{P}_{Q_{inter}}$、$\boldsymbol{P}_{K_{inter}}$、$\boldsymbol{P}_{V_{inter}}\in R^{\frac{C}{2}\times d_1}$分别表示不同块之间共享的投影矩阵. 接下来，本论文采用一个$p_{inter}\times p_{inter}$大小的深度可分离卷积对$\boldsymbol{Q}^j_{inter}$、$\boldsymbol{K}^j_{inter}$的块进行深度可分离卷积，将每个通道上的$p_{inter}\times p_{inter}$块聚合为一个像素点. 经过深度可分离卷积后，将得到特征$\boldsymbol{X}^j_Q$、$\boldsymbol{X}^j_K\in R^{L_{inter}\times d_1}$，此过程可用公式表示为

$$\boldsymbol{X}^j_Q=H_{DWConvp\times p}(\boldsymbol{Q}^j_{inter}) \tag{16}$$

$$\boldsymbol{X}^j_K=H_{DWConvp\times p}(\boldsymbol{K}^j_{inter}) \tag{17}$$

其中，$H_{DWConvp\times p}(\cdot)$表示的是核为$p_{inter}\times p_{inter}$的深度可分离操作. 显而易见，在经过了深度可分离卷积之后，$\boldsymbol{X}^j_Q$和$\boldsymbol{X}^j_K$中的每个像素点就聚集了对应图像块的结构信息，因此，我们直接对$\boldsymbol{X}^j_Q$、$\boldsymbol{X}^j_K$中的每个像素点执行逐像素点的自注意力计算，计算公式为

$$\boldsymbol{X}^j_{inter}=softMax\left(\frac{\boldsymbol{X}^j_Q\left(\boldsymbol{X}^j_K\right)^T}{\sqrt{d_1}}\right)\boldsymbol{V}^j_{inter} \tag{18}$$

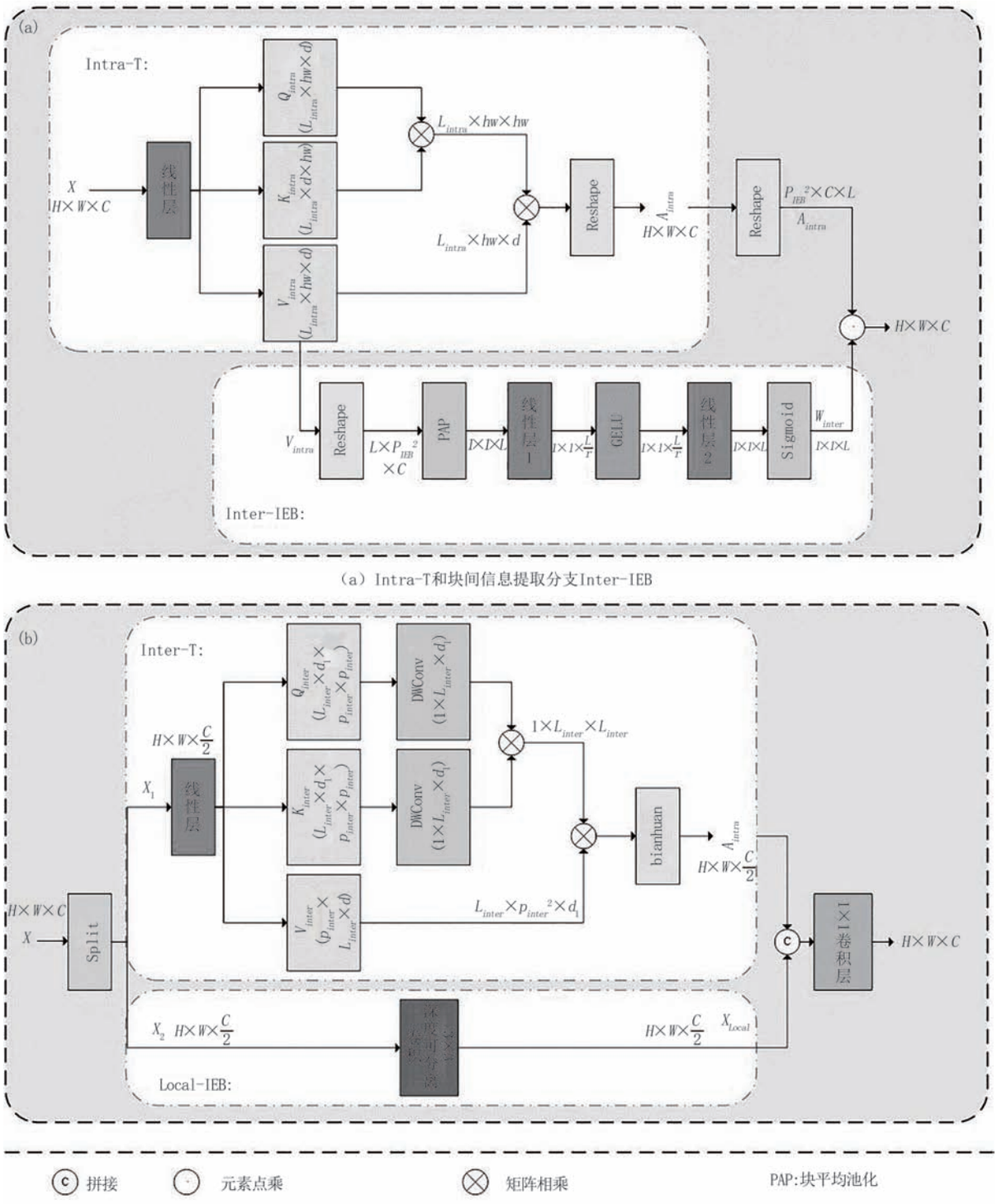



图2 本论文提出的块内块间Transformer模块

与Intra-T相同，公式(18)将并行执行$s$次自注意力计算，并将每个头部计算的自注意力结果$\boldsymbol{X}_{inter}^{j}\in R^{H\times W\times d_1}$从通道维度上进行拼接得到块间注意力特征$\boldsymbol{A}_{inter}\in R^{H\times W\times \frac{C}{2}}$：

$$\boldsymbol{A}_{inter}=concat\left(\boldsymbol{X}_{inter}^{1},\boldsymbol{X}_{inter}^{2},\cdots,\boldsymbol{X}_{inter}^{s}\right) \tag{19}$$

显而易见，本论文提出的Inter-T采用深度可分离卷积$H_{DWConv}(\cdot)$将每个通道上的$p_{inter}\times p_{inter}$块聚合为一个像素点，那么这个像素点就学到了对应的$p_{inter}\times p_{inter}$块中的结构信息，因此，每个通道上的$L_{inter}$个像素点就组成了能够表示该通道上整幅图像结构信息的结构矩阵，而且该结构矩阵的分辨率仅

为输入特征分辨率的$\frac{1}{p_{inter}^2}$. 由此可知，整幅图像中的所有结构信息被统计到一个更低的维度空间，也就使得全局范围内不同块之间的结构自注意力的空间复杂度和时间复杂度被极大地降低. 也正因如此，本论文提出的Inter-T能够同时实现显式的全局范围内相互依赖性的捕捉和轻量级.

3.2.3　信息交互机制(IIM)

虽然提出的Intra-T和Inter-T已经能够实现局部特征相似性和全局结构相似性的显式捕捉，但为了进一步增强IIDAN的表达能力，本论文又提出了一种IIM. 它能够通过集成块间信息来增强Intra-T的全局结构感知，同时辅以局部细节信息来弥补Inter-T自身可能忽略的局部特征. 这种互补性信息的融合不仅显著增强了网络的特征捕捉能力，还进一步提升了模型对复杂纹理和精细结构的表达能力，从而实现更高质量的SR重建. 本论文提出的IIM如图1和图2中的Inter-IEB、Local-IEB所示.

首先，针对Intra-T，本论文创建了一个块间信息提取分支(Inter-block Information Extraction Branch：Inter-IEB)来弥补Intra-T无法捕捉到块间信息的缺陷. 如图2(a)所示，Inter-IEB首先对输入特征$\boldsymbol{V}_{intra}\in R^{H\times W\times C}$进行不重叠的块平均池化操作(Patch Average Pooling，PAP)，每个块的大小为$p_{IEB}\times p_{IEB}$，如公式(21)所示：

$$\boldsymbol{X}_{pooling}^{l}=H_{pooling}(\boldsymbol{X}^{l}), \tag{20}$$

$$H_{pooling}(\boldsymbol{X}^{l})=\frac{1}{p_{IEB}\times p_{IEB}\times C}\sum_{i}^{p_{IEB}}\sum_{j}^{p_{IEB}}\sum_{k}^{C}\boldsymbol{X}^{l}(i,j,k) \tag{21}$$

其中，$H_{pooling}(\cdot)$代表的是PAP操作，$\boldsymbol{X}^{l}(i,j,k)$表示是第$l$个块上$(i,j,k)$位置上的像素值. 经过平均池化，得到向量$\boldsymbol{X}_{pooling}\in R^{1\times 1\times L}$，$(L=HW/p_{IEB}^2)$，因此，$\boldsymbol{X}_{pooling}$在通道维度上的每个值就代表了对应图像块的结构统计信息. 然后，将$\boldsymbol{X}_{pooling}$通过两个线性层和一个激活层对$\boldsymbol{X}_{pooling}$中各个块的信息进行融合和交互，此过程可以用公式表示为

$$\boldsymbol{X}_{IBI}=H_{Linear}^{2}\left(GELU\left(H_{Linear}^{1}\left(\boldsymbol{X}_{pooling}\right)\right)\right) \tag{22}$$

如公式(22)所示，第一个线性层$H_{Linear}^{1}(\cdot)$按缩减比$r$对$\boldsymbol{X}_{pooling}$进行通道维度的压缩，以实现不同块之间的信息融合，得到$\boldsymbol{X}_{pooling}^{1}\in R^{1\times 1\times \frac{L}{r}}$. 将$\boldsymbol{X}_{pooling}^{1}$经过GELU进行非线性激活后再通过第二个线性层$H_{Linear}^{2}(\cdot)$按比率$r$将$\boldsymbol{X}_{pooling}^{1}$的通道数又扩展成$L$：$\boldsymbol{X}_{IBI}\in R^{1\times 1\times L}$. 那么$\boldsymbol{X}_{IBI}$在通道维度上的每个值就是对应图像块与所有其他图像块进行了全局信息交互之后的结果，因此，Inter-IEB也就能很好地弥补Intra-T无法捕捉到全局块间信息的缺陷. 最后采用sigmoid函数得到每个块的权重$\boldsymbol{W}_{inter}\in R^{1\times 1\times L}$：

$$\boldsymbol{W}_{inter}=Sigmoid(\boldsymbol{X}_{IBI}) \tag{23}$$

因此，将$\boldsymbol{W}_{inter}$与$\boldsymbol{A}_{intra}$进行逐元素点乘也就实现了对$\boldsymbol{A}_{intra}$补充全局块间结构信息的目的.

接下来，对于Inter-T模块而言，因为Inter-T已经显式地捕捉了全局范围内的结构信息，因此，我们仅采用一个简单的3×3深度可分离卷积来提取局部信息，如图2(b)中的局部信息提取分支(Local Information Extraction Branch，Local-IEB)所示：

$$\boldsymbol{X}_{Local}=H_{DWConv3\times 3}(\boldsymbol{X}_{2}) \tag{24}$$

其中，$H_{DWConv3\times 3}(\cdot)$代表的是深度可分离卷积操作. 本论文采用拼接和1×1卷积来融合$\boldsymbol{X}_{Local}$和$\boldsymbol{A}_{inter}$，实现对$\boldsymbol{A}_{inter}$的局部信息的补充.

最后，在IITM中，本论文根据文献[53]的Simple Gate模块，设计了一种轻量级的前馈网络(Simple Gate Feed-forward Network，SGFN)，如图3所示.

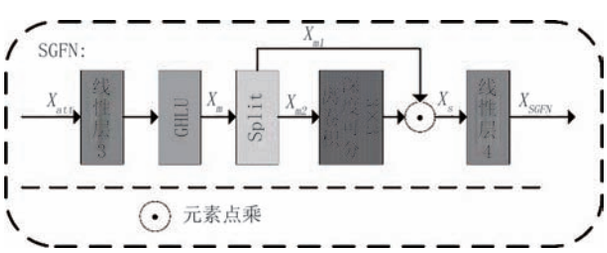



图3　简单门控前馈网络SGFN

SGFN先将输入的特征$\boldsymbol{X}_{att}\in R^{H\times W\times C}$通过一个线性层将通道数扩张为原来的4倍(即通道数由$C$扩张为$4C$)，再经过非线性层进行激活，然后通过Split操作将其中输入的特征沿着通道维度均分成两部分，一部分经过深度可分离卷积进行特征提取后与另一部分进行元素点乘，最后再次通过线性层将通道数变回$C$. 该过程可公式化为

$$\boldsymbol{X}_{m}=GELU\left(H_{Linear}^{3}(\boldsymbol{X}_{att})\right) \tag{25}$$

$$\boldsymbol{X}_{m1},\boldsymbol{X}_{m2}=H_{Split}(\boldsymbol{X}_{m}) \tag{26}$$

$$\boldsymbol{X}_{s}=\boldsymbol{X}_{m1}\cdot \mathrm{H}_{\mathrm{DWConv3\times 3}}(\boldsymbol{X}_{m2}) \tag{27}$$

$$\boldsymbol{X}_{SGFN}=H_{Linear}^{4}(\boldsymbol{X}_{s}) \tag{28}$$

其中，$H_{Split}(\cdot)$代表的是通道维度上的拆分操作. 输入特征$\boldsymbol{X}_{att}$经过线性层$H_{Linear}^{3}(\cdot)$和激活函数$GELU(\cdot)$之后变为$\boldsymbol{X}_{m}\in R^{H\times W\times 4C}$，然后通过Split操作得到$\boldsymbol{X}_{m1}$、$\boldsymbol{X}_{m2}\in R^{H\times W\times 2C}$. 将$\boldsymbol{X}_{m2}$经过一个深度

可分离卷积提取特征后与$\boldsymbol{X}_{m1}$点乘得到特征$\boldsymbol{X}_s \in R^{H\times W\times 2C}$，最后将$\boldsymbol{X}_s$通过$H^4_{Linear}(\cdot)$恢复通道数得到特征$\boldsymbol{X}_{SGFN} \in R^{H\times W\times C}$.

# 4 实验及分析

在本节中，我们进行了大量的实验，从定量（客观评价指标）和定性（主观视觉效果）两个方面来评价本论文提出的IIDAN的有效性和优越性. 同时，我们还实施了消融实验，以此来评估本论文提出的贡献点的有效性.

## 4.1 实验设置

### 4.1.1 IIDAN中的超参数

对于本论文提出的IIDAN而言，其超参数的设置为：$M$=6、$N$=3、$s$=6、$C$=60和$r$=8. 同时，对于Intra-T而言，将窗口大小设置为8×32（即$h$=8，$w$=32）；对于Inter-T而言，将其块的大小设置为8×8（即$p_{inter}$=8）；对于Inter-IEB而言，将块的数量$L$设置为64，那么$p_{IEB}$是可以随着输入图像的大小而改变的.

### 4.1.2 数据集和客观评价指标

本论文遵循大多数SISR任务的工作来训练和测试本论文提出的IIDAN. 具体来说，在800张图像的数据集DIV2K[54]上训练IIDAN，并在五个基准数据集：Set5[55]、Set14[56]、B100[57]、Urban100[58]和Manga109[59]上对其进行测试. 分别在×2、×3和×4三种放大因子下进行实验. 在定量评价方面，本论文采用峰值信噪比（Peak Signal to Noise Ratio，PSNR）和结构相似性指数（Structural Similarity Index Metrics，SSIM）[60]两种客观的评价指标来客观评估本论文提出的IIDAN的性能. 同时，在输入图像大小分别为3×320×180、3×426×240、3×640×360，以及分别放大4倍、3倍和2倍的情况下来计算提出的IIDAN的参数量和计算复杂度.

### 4.1.3 训练设置

在训练阶段，输入的LR块的大小被设置为64×64，批量大小设置为32，训练迭代次数为1 600 000次. 采用随机旋转90度、180度、270度和水平翻转来对训练数据集进行数据增强. 采用Adam[61]优化器，$\beta_1$=0.9、$\beta_2$=0.999. 初始学习率被设置为$5\times 10^{-4}$，并分别在500 000次、900 000次、1 200 000次、1 400 000次、1 500 000次和1 550 000次迭代时减半. 最后，本论文的所有实验都是在NVIDIA 3090 GPU和Pytorch深度学习框架[62]上进行训练和测试的.

## 4.2 消融实验

如前所述，本论文的主要贡献点为Inter-T和IIM. 因此，本章节中，我们将在Manga109[59]的测试集上对×4的SR结果进行消融实验，以此来验证本论文提出的Inter-T和IIM的有效性. 消融实验中所有模型的超参数和训练细节都一致.

### 4.2.1 Intrer-T的有效性消融实验

如前所述，本论文提出的Inter-T很好地实现了整幅图像全局范围内结构相似性的显式捕捉，能够增强网络的表达能力，重建出高质量的SR图像. 因此，为了能够准确评估提出的Inter-T的有效性，本论文首先创建了一个基线模型Baseline：（1）将IIDAN中的IIM移除，即同时去除掉Inter-IEB和Local-IEB；（2）将IIDAN中的Inter-T全部替换成Intra-T. Baseline如图4所示.

然后，本论文又创建了一个新的网络模型IIDAN-NoIIM：只将IIDAN中的IIM移除，其余成分保持不变. 由此可见，Baseline和IIDAN-NoIIM的区别就仅仅在于是否采用了Inter-T. 因此，比较Baseline和IIDAN-NoIIM之间的性能差异是能够准确评估Inter-T的有效性的. 如表1所示，虽然Baseline的参数量和计算复杂度略高于IIDAN-NoIIM，但是Baseline的PSNR和SSIM较IIDAN-NoIIM分别低了0.12 dB和0.0016. 表1能够很好地证明本论文提出的Inter-T的有效性.

为了进一步验证提出的Inter-T的有效性，本论文将Inter-T中不同块之间的自注意力进行了可视化，如图5所示. 图5（a）所示为一个参考块（右上方的红色方块）与整幅图像中所有其他块之间的自注意力地图，相似性越高的块颜色越深. 图5（b）所示为与参考块（右上方的红色方块）最相似的前五个块（绿色方块）. 从图5中可以看到，本论文提出的Inter-T能够在整幅图的全局范围内准确找到与参考块具有相似结构的块，因此图5也很好地证明了本论文的贡献：Inter-T能够在全局范围内实现结构相似性的准确显式捕捉.

### 4.2.2 IIM的有效性消融实验

为了能够准确评估本论文提出的IIM的有效性，我们在模型IIDAN-NoIIM上分别添加Inter-IEB和Local-IEB，得到两个新的模型：IIDAN-InterB（即仅在IIDAN-NoIIM的Intra-T中加入Inter-IEB）和IIDAN-LocalB（即仅在IIDAN-NoIIM的Inter-T中加入Local-IEB）. 表2所示为IIDAN-NoIIM、IIDAN-InterB、IIDAN-LocalB和IIDAN

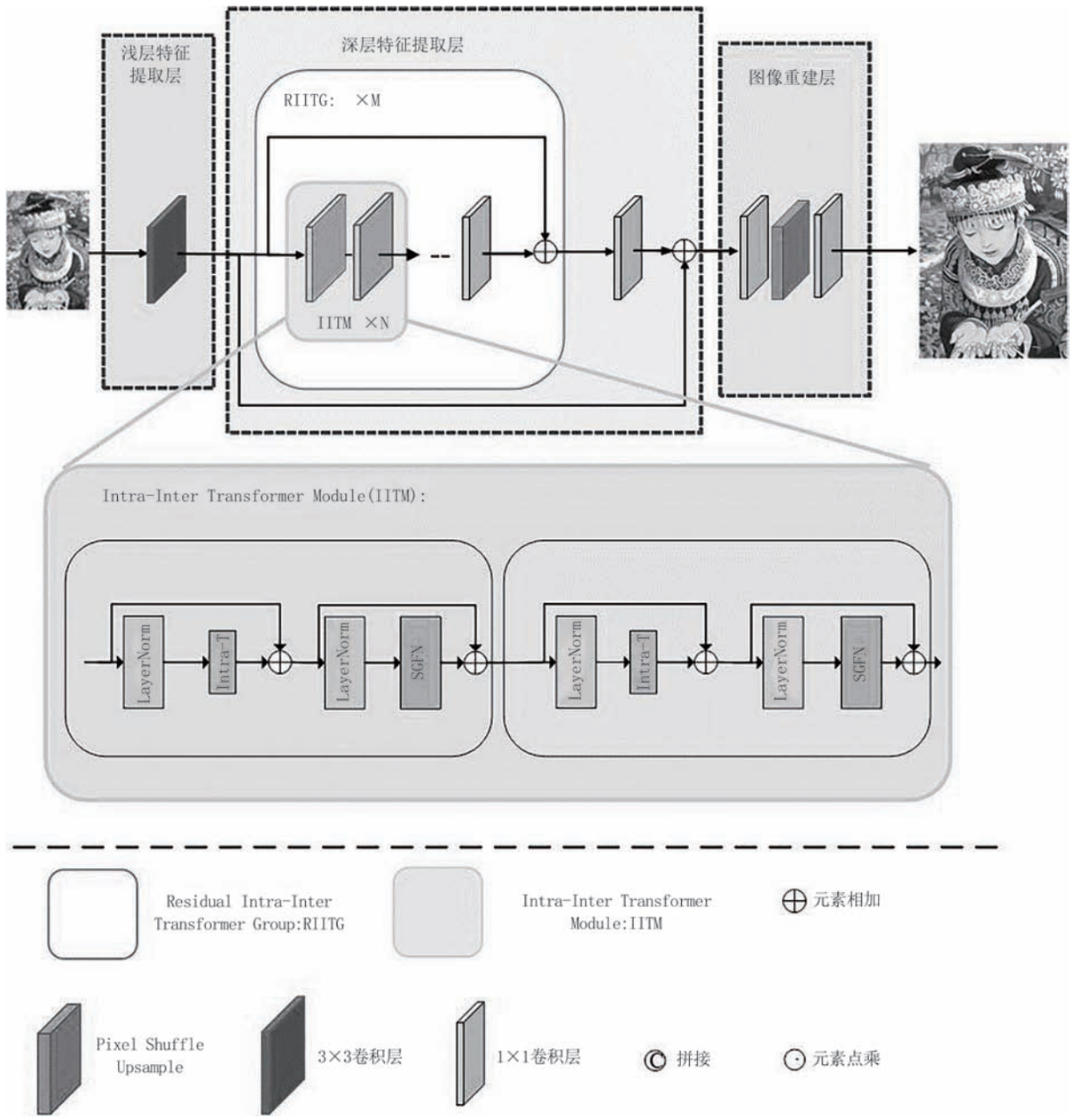



图4　基线模型(Baseline)

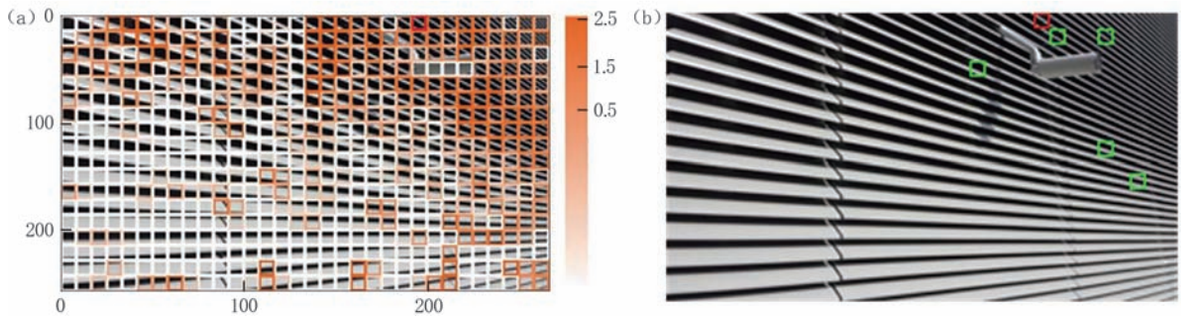
**表1　Inter-T的有效性消融实验**

| Model | Intra-T | Inter-T | Params (K) | FLOPs (G) | PSNR (dB) | SSIM |
|---|---|---|---|---|---|---|
| Baseline | √ | | 720.4 | 34.9 | 31.13 | 0.9152 |
| IIDAN-NoIIM | √ | √ | 723.0 | 34.8 | 31.25 | 0.9168 |

四种模型的参数量、计算复杂度、PSNR和SSIM的定量比较. 如表2所示，Inter-IEB和Local-IEB均能提升模型的性能，与IIDAN-NoIIM相比，Inter-IEB和Local-IEB分别将PSNR提升了0.14 dB和0.1 dB，分别将SSIM提升了0.0010和0.0005. 而当同时采用Inter-IEB和Local-IEB时，其性能提升是最大的：与IIDAN-NoIIM相比，PSNR和SSIM分别提升了0.22 dB和0.0013. 表2很好地证明了本论文提出的IIM的有效性.

4.2.3　IIDAN的各阶段消融模型可视化比较

为了能够更直观地证明本论文提出的创新点的有效性，本论文将消融实验中涉及到的模型的SR重建结果进行了主观视觉效果的可视化比较，即：Baseline模型、IIDAN-NoIIM模型、IIDAN-LocalB模型、IIDAN-InterB模型和IIDAN的SR重建结果的可视化比较. 比较结果如图6所示. 在可视化比较结果中可以很明显看到，随着创新点的逐步加入，SR重建结果的主观视觉效果呈现出明显的提升趋势，最终，本论文提出的IIDAN获得了最佳的可视




(a)参考块(红色方块)与其他块的注意力图谱的可视化图 (b)参考块(红色方块)在图像上与之最相似的五个块(绿色方块)

图5 本论文提出的Inter-T有效性实验

**表2 IIM的有效性消融实验**

| Model | Inter-T+Local-IEB | Intra-T+Inter-IEB | Params/K | FLOPs/G | PSNR/dB | SSIM |
|---|---|---|---|---|---|---|
| IIDAN-NoIIM | | | 723.0 | 34.8 | 31.25 | 0.9168 |
| IIDAN-InterB | | √ | 737.5 | 35.2 | 31.39 | 0.9178 |
| IIDAN-LocalB | √ | | 754.9 | 36.8 | 31.35 | 0.9173 |
| IIDAN | √ | √ | 769.2 | 37.3 | 31.47 | 0.9181 |

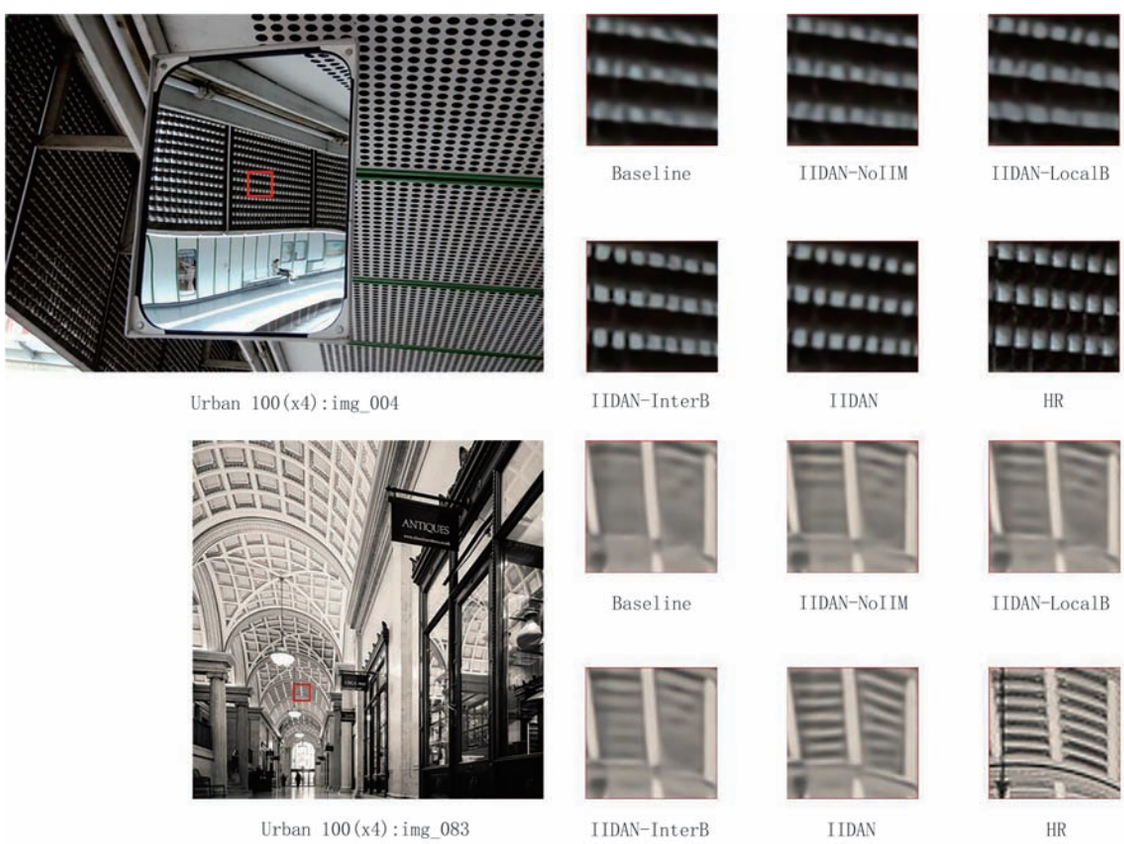



图6 消融实验各阶段模型的SR主观视觉效果比较图(红色矩形中的对比区域被放大在右侧)

化重建结果,最接近HR图像. 图6从主观的视觉方面证明了本论文提出的创新点的有效性.

### 4.3 与前沿方法的比较实验

#### 4.3.1 定量(客观评价指标)比较

为了验证本论文提出方法的优越性,在本小节中,本论文将提出的IIDAN与当前最先进的14种SISR方法进行比较:SwinIR-light[12]、ELAN-light[15]、ESRT[16]、MSRA[22]、CARN[23]、EMASRN[25]、SRFormer[29]、Omni-SR[30]、CAMixerSR[32]、SeemoRe[33]、IMDN[37]、LatticeNet[38]、DRSAN[40]和DBNet[49]. 表3~表5所示为在三种放大因子的五个基准测试集上的参数量、计算复杂度、平均PSNR值

**表3　不同的轻量级SISR方法在下采样因子为4倍下的平均PSNR值、平均SSIM值、模型参数量Params和计算复杂度FLOPs**

| Method | Publication Year | Scale | Params /K | FLOPs /G | Set5 | | Set14 | | B100 | | Urban100 | | Manga109 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | PSNR /dB | SSIM | PSNR /dB | SSIM | PSNR /dB | SSIM | PSNR /dB | SSIM | PSNR /dB | SSIM |
| SwinIR-light[12] | 2021 | | 930 | 63.6 | 32.44 | 0.8976 | 28.77 | 0.7858 | 27.69 | 0.7406 | 26.47 | 0.7980 | 30.92 | 0.9151 |
| ELAN-light[15] | 2022 | | 601 | 37.1 | 32.43 | 0.8975 | 28.78 | 0.7858 | 27.69 | 0.7406 | 26.54 | 0.7982 | 30.92 | 0.9150 |
| ESRT[16] | 2022 | | 751 | 45.8 | 32.19 | 0.8947 | 28.69 | 0.7833 | 27.69 | 0.7379 | 26.39 | 0.7962 | 30.75 | 0.9100 |
| MSRA[22] | 2023 | | 789 | 53.6 | 32.46 | 0.8984 | 28.86 | 0.7876 | 27.72 | 0.7419 | 26.65 | 0.8037 | 31.08 | 0.9157 |
| CARN[23] | 2018 | | 1592 | 90.9 | 32.13 | 0.8937 | 28.60 | 0.7806 | 27.58 | 0.7349 | 26.07 | 0.7837 | 30.47 | 0.9084 |
| EMASRN[25] | 2022 | | 546 | 1055.3 | 32.17 | 0.8948 | 28.57 | 0.7809 | 27.55 | 0.7351 | 26.01 | 0.7838 | 30.41 | 0.9076 |
| SRFormer-light[29] | 2023 | | 873 | 62.8 | 32.51 | 0.8988 | 28.82 | 0.7872 | 27.73 | 0.7422 | 26.67 | 0.8032 | 31.17 | 0.9165 |
| Omni-SR[30] | 2023 | ×4 | 792 | / | 32.49 | 0.8988 | 28.78 | 0.7859 | 27.71 | 0.7415 | 26.64 | 0.8018 | 31.02 | 0.9151 |
| CAMixerSR[32] | 2024 | | 765 | 44.6 | 32.51 | 0.8988 | 28.82 | 0.7870 | 27.72 | 0.7416 | 26.63 | 0.8012 | 31.18 | 0.9166 |
| SeemoRe-L[33] | 2024 | | 969 | 50,0 | 32.51 | 0.8990 | 28.92 | 0.7888 | 27.78 | 0.7428 | 26.79 | 0.8046 | 31.48 | 0.9181 |
| IMDN[37] | 2019 | | 715 | 40.9 | 32.21 | 0.8948 | 28.58 | 0.7811 | 27.56 | 0.7353 | 26.04 | 0.7838 | 30.45 | 0.9075 |
| LatticeNet[38] | 2020 | | 777 | 43.6 | 32.30 | 0.8962 | 28.68 | 0.7830 | 27.62 | 0.7367 | 26.25 | 0.7873 | / | / |
| DRSAN[40] | 2023 | | 730 | 57.6 | 32.25 | 0.8945 | 28.55 | 0.7817 | 27.59 | 0.7374 | 26.14 | 0.7875 | / | / |
| DBNet[49] | 2023 | | 832 | 51.8 | 32.29 | 0.8961 | 28.71 | 0.7834 | 27.66 | 0.7377 | 26.34 | 0.7909 | 30.83 | 0.9111 |
| IIDAN(ours) | / | | 769 | 37.3 | 32.59 | 0.9001 | 28.94 | 0.7888 | 27.79 | 0.7438 | 26.83 | 0.8070 | 31.47 | 0.9181 |

注：所有方法均是在DIV2K训练集上训练，在五个基准测试集上测试. 其中最好的性能和第二好的性能分别用红色和蓝色标记.

**表4　不同的轻量级SISR方法在下采样因子为3倍下的平均PSNR值、平均SSIM值、模型参数量Params和计算复杂度FLOPs**

| Method | Publication Year | Scale | Params /K | FLOPs /G | Set5 | | Set14 | | B100 | | Urban100 | | Manga109 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | PSNR /dB | SSIM | PSNR /dB | SSIM | PSNR /dB | SSIM | PSNR /dB | SSIM | PSNR dB | SSIM |
| SwinIR-light[12] | 2021 | | 918 | 111.0 | 34.62 | 0.9289 | 30.54 | 0.8463 | 29.20 | 0.8082 | 28.66 | 0.8624 | 33.98 | 0.9478 |
| ELAN-light[15] | 2022 | | 590 | 68.5 | 34.61 | 0.9288 | 30.55 | 0.8463 | 29.21 | 0.8081 | 28.69 | 0.8624 | 34.00 | 0.9478 |
| ESRT[16] | 2022 | | 770 | 69.3 | 34.42 | 0.9268 | 30.43 | 0.8433 | 29.15 | 0.8063 | 28.46 | 0.8574 | 33.95 | 0.9455 |
| MSRA[22] | 2023 | | 777 | 91.5 | 34.65 | 0.9291 | 30.60 | 0.8470 | 29.24 | 0.8093 | 28.86 | 0.8664 | 34.29 | 0.9489 |
| CARN[23] | 2018 | | 1592 | 118.8 | 34.29 | 0.9255 | 30.29 | 0.8407 | 29.06 | 0.8034 | 28.06 | 0.8493 | 33.50 | 0.9440 |
| EMASRN[25] | 2022 | | 427 | 853.5 | 34.36 | 0.9264 | 30.30 | 0.8411 | 29.05 | 0.8035 | 28.04 | 0.8493 | 33.43 | 0.9433 |
| SRFormer-light[29] | 2023 | | 861 | 105.0 | 34.67 | 0.9296 | 30.57 | 0.8469 | 29.26 | 0.8099 | 28.81 | 0.8655 | 34.19 | 0.9489 |
| Omni-SR[30] | 2023 | ×3 | 780 | / | 34.70 | 0.9294 | 30.57 | 0.8469 | 29.28 | 0.8094 | 28.84 | 0.8656 | 34.22 | 0.9487 |
| CAMixerSR[32] | 2024 | | 753 | 85.6 | 34.65 | 0.9295 | 30.62 | 0.8471 | 29.26 | 0.8093 | 28.81 | 0.8645 | 34.34 | 0.9491 |
| SeemoRe-L[33] | 2024 | | 959 | 87.0 | 34.72 | 0.9297 | 30.60 | 0.8469 | 29.29 | 0.8101 | 28.86 | 0.8653 | 34.53 | 0.9496 |
| IMDN[37] | 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[38] | 2020 | | 765 | 76.3 | 34.53 | 0.9281 | 30.39 | 0.8424 | 29.15 | 0.8059 | 28.33 | 0.8538 | / | / |
| DRSAN[40] | 2023 | | 750 | 78.0 | 34.47 | 0.9274 | 30.35 | 0.8422 | 29.11 | 0.8060 | 28.26 | 0.8542 | / | / |
| DBNet[49] | 2023 | | 826 | 70.2 | 34.46 | 0.9279 | 30.42 | 0.8427 | 29.18 | 0.8063 | 28.51 | 0.8571 | 33.99 | 0.9466 |
| IIDAN(ours) | / | | 757 | 68.0 | 34.78 | 0.9303 | 30.71 | 0.8489 | 29.31 | 0.8111 | 29.03 | 0.8686 | 34.57 | 0.9503 |

注：所有方法均是在DIV2K训练集上训练，在五个基准测试集上测试. 其中最好的性能和第二好的性能分别用红色和蓝色标记.

和平均SSIM值. 由表3～表5可见，首先，在所有的测试集和×2、×3、×4三种放大因子中，本论文提出的IIDAN能在绝大多数的情况下获得最高的平均PSNR和最高的平均SSIM，仅在放大4倍时，IIDAN在Manga109[59]上的平均PSNR位于第二高. 其次，在×2、×3、×4三种放大因子中，本论文提出的IIDAN在绝大多数的情况下都具有最低的计算复杂度：仅在放大4倍时，IIDAN的计算复杂度第二低；而在放大3倍和4倍时，IIDAN的计算复杂度都是最少的.

**表5 不同的轻量级SISR方法在下采样因子为2倍下的平均PSNR值、平均SSIM值、模型参数量Params和计算复杂度FLOPs**

| Method | Publication Year | Scale | Params /K | FLOPs /G | Set5 | | Set14 | | B100 | | Urban100 | | Manga109 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | PSNR /dB | SSIM | PSNR /dB | SSIM | PSNR /dB | SSIM | PSNR /dB | SSIM | PSNR /dB | SSIM |
| SwinIR-light[12] | 2021 | ×2 | 910 | 244.0 | 38.14 | 0.9611 | 33.86 | 0.9206 | 32.31 | 0.9012 | 32.76 | 0.9340 | 39.12 | 0.9783 |
| ELAN-light[15] | 2022 | | 582 | 168.4 | 38.17 | 0.9611 | 33.94 | 0.9207 | 32.30 | 0.9012 | 32.76 | 0.9340 | 39.11 | 0.9782 |
| ESRT[16] | 2022 | | 677 | / | 38.03 | 0.9600 | 33.75 | 0.9184 | 32.25 | 0.9001 | 32.58 | 0.9318 | 39.12 | 0.9774 |
| MSRA[22] | 2023 | | 769 | 196.0 | 38.23 | 0.9614 | 34.01 | 0.9211 | 32.33 | 0.9017 | 32.98 | 0.9358 | 39.24 | 0.9783 |
| CARN[23] | 2018 | | 1592 | 222.8 | 37.76 | 0.9590 | 33.52 | 0.9166 | 32.09 | 0.8978 | 31.92 | 0.9256 | 38.36 | 0.9765 |
| EMASRN[25] | 2022 | | / | / | / | / | / | / | / | / | / | / | / | / |
| SRFormer-light[29] | 2023 | | 853 | 236.0 | 38.23 | 0.9613 | 33.94 | 0.9209 | 32.36 | 0.9019 | 32.91 | 0.9353 | 39.28 | 0.9785 |
| Omni-SR[30] | 2023 | | 772 | / | 38.22 | 0.9613 | 33.98 | 0.9210 | 32.36 | 0.9020 | 33.05 | 0.9363 | 39.28 | 0.9784 |
| CAMixerSR[32] | 2024 | | 746 | 167.0 | 38.23 | 0.9613 | 34.00 | 0.9214 | 32.34 | 0.9016 | 32.95 | 0.9348 | 39.32 | 0.9781 |
| SeemoRe-L[33] | 2024 | | 931 | 197.0 | 38.27 | 0.9616 | 34.01 | 0.9210 | 32.35 | 0.9018 | 32.87 | 0.9344 | 39.49 | 0.9790 |
| IMDN[37] | 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[38] | 2020 | | 756 | 169.5 | 38.15 | 0.9610 | 33.78 | 0.9193 | 32.25 | 0.9005 | 32.43 | 0.9302 | / | / |
| DRSAN[40] | 2023 | | 850 | 196.3 | 38.13 | 0.9610 | 33.72 | 0.9189 | 32.24 | 0.9009 | 32.41 | 0.9312 | / | / |
| DBNet[49] | 2023 | | / | / | / | / | / | / | / | / | / | / | / | / |
| IIDAN(ours) | / | | 749 | 151.0 | 38.33 | 0.9622 | 34.12 | 0.9232 | 32.37 | 0.9028 | 33.07 | 0.9382 | 39.53 | 0.9797 |

注：所有方法均是在DIV2K训练集上训练，在五个基准测试集上测试. 其中最好的性能和第二好的性能分别用红色和蓝色标记.

**表6 与传统方法[4]在下采样因子为3倍下的定量比较**

| Method | Publication Year | Scale | Set5 | | Set14 | | B100 | | Urban100 | | Manga109 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | PSNR /dB | SSIM | PSNR /dB | SSIM | PSNR /dB | SSIM | PSNR /dB | SSIM | PSNR /dB | SSIM |
| 传统方法[4] | 2020 | ×3 | 31.53 | 0.8906 | 27.78 | 0.8125 | 28.29 | 0.7846 | 24.65 | 0.7468 | 24.14 | 0.7904 |
| IIDAN(ours) | / | | 34.78 | 0.9303 | 30.71 | 0.8489 | 29.31 | 0.8111 | 29.03 | 0.8686 | 34.57 | 0.9503 |

注：所有方法均是在五个基准测试集上测试.

在放大4倍时（如表3所示），虽然提出的IIDAN的参数量比EMASRN[25]多出了40.8%，但是IIDAN的计算复杂度却比它低了28倍之多，同时在Set5[55]数据集上提出的IIDAN的平均PSNR也比EMASRN[25]高出了0.49 dB. 虽然IIDAN与SeemoRe-L[33]在某些数据集上性能相似（放大因子为×4时的Set14数据集的SSIM值，以及Manga109[59]数据集的PSNR值和SSIM值），但是在绝大多数的数据集上，IIDAN的性能优于SeemoRe-L[33]，而且IIDAN的参数量和计算复杂度分别比SeemoRe-L[33]减少了20.6%和25.4%.

在×3的放大因子上（如表4所示），对于ELAN-light[15]和EMASRN[25]而言，提出的IIDAN的计算复杂度分别比它们低了0.7%和12倍，并且在Manga109[59]数据集上，IIDAN的平均PSNR比ELAN-light和EMASRN分别高出了0.57 dB和1.14 dB. 虽然提出的IIDAN与性能第二好的方法（SeemoRe-L[33]）相比，在某些数据集上性能提升不足0.1dB，但是IIDAN的计算复杂度和参数量却分别比SeemoRe-L[33]低了21%和21.8%.

在放大2倍时（如表5所示），提出的IIDAN的计算复杂度比ELAN-light[15]低了11.5%，同时在Manga109[59]数据集上IIDAN的平均PSNR比ELAN-light[15]高出了0.41 dB. 同样的，虽然提出的IIDAN与性能第二好的方法相比，在某些数据集上性能提升不足0.1 dB，但是相比这些性能第二好的方法，IIDAN的计算复杂度和参数量都是最低的. 因此，综上所述，本论文提出的IIDAN能够以更少的计算代价获得相似或者更优的性能. 表3-表5从客观的评价指标方面很好地证明了本论文提出的IIDAN的优越性.

此外，为了证明本论文提出的方法相较非网络的传统SISR方法的优越性，本论文还将IIDAN与传统的非网络方法[4]进行了比较. 由于文献[4]仅能进行3倍的超分辨率重建，因此，本论文仅在×3的放大因子上将提出的IIDAN与文献[4]进行了比较，

比较结果如表6所示，本论文提出的IIDAN在所有的数据集上都明显远优于传统的非网络方法[4]，进一步证明了本论文提出的IIDAN的优越性.

4.3.2　定性（主观视觉效果）比较

为了能够更加全面地证明本论文提出方法的优越性，除了客观的评价指标之外，本论文还在主观的视觉效果上将本论文提出的IIDAN与SwinIR-light[12]、ESRT[16]、CARN[23]、EMASRN[25]、SRFormer[29]、Omni-SR[30]、CAMixerSR[32]、IMDN[37]和LatticeNet[38]等方法进行了比较，比较结果如图8、图9所示（放大4倍的主观视觉比较结果）. 如图8、图9所示，文献[12，16，23，25，29，30，32，37，38]方法重建出的SR图像都会存在有一定程度的模糊伪影、失真或者不正确的边缘纹理等瑕疵，例如，在img_042中，现有方法对于纹理的细节恢复能力都较差，都会存在不同程度的失真（文献[16，23，25，29，32，37]）和模糊伪影[12,30,38]. 这些瑕疵同样存在于图像img_012、img_30、img_093和img_100中. 从图像comic中可以很明显看出，现有方法重建出的SR图像都存在较大程度的模糊（文献[12，16，23，25，29，30，32，37，38]）. 同时，在图像ppt3中，现有方法重建出的SR图像同样存在着类似的模糊瑕疵（字母D和O之间出现不同程度的粘连）. 相比之下，显而易见，本论文提出的IIDAN不仅能够有效消除以上的瑕疵，还能够重建出更多的结构信息和更精细准确的纹理细节，是最接近真实HR图像的方法（详见图8和图9以及其中的放大区域）.

此外，本论文还将提出的IIDAN与传统的非网络方法[4]进行了3倍上采样超分辨率的主观视觉图比较，比较结果如图7所示. 可以很明显看到，本论文提出的IIDAN能够重建出更清晰和准确的边缘和纹理细节：在baby图片中，传统方法[4]对睫毛的重建是模糊的，而本论文提出的IIDAN能够重建出更清晰的睫毛结构，更接近HR图片；在Manga109测试集的Akuhamu图片中，传统方法[4]对文字的重建是错误的，而本论文提出的IIDAN能够重建出准确的文字，几乎与HR图片无差别. 图7从主观的视觉方面进一步证明了本论文提出的IIDAN的优越性（详见图7以及其中的放大区域）.

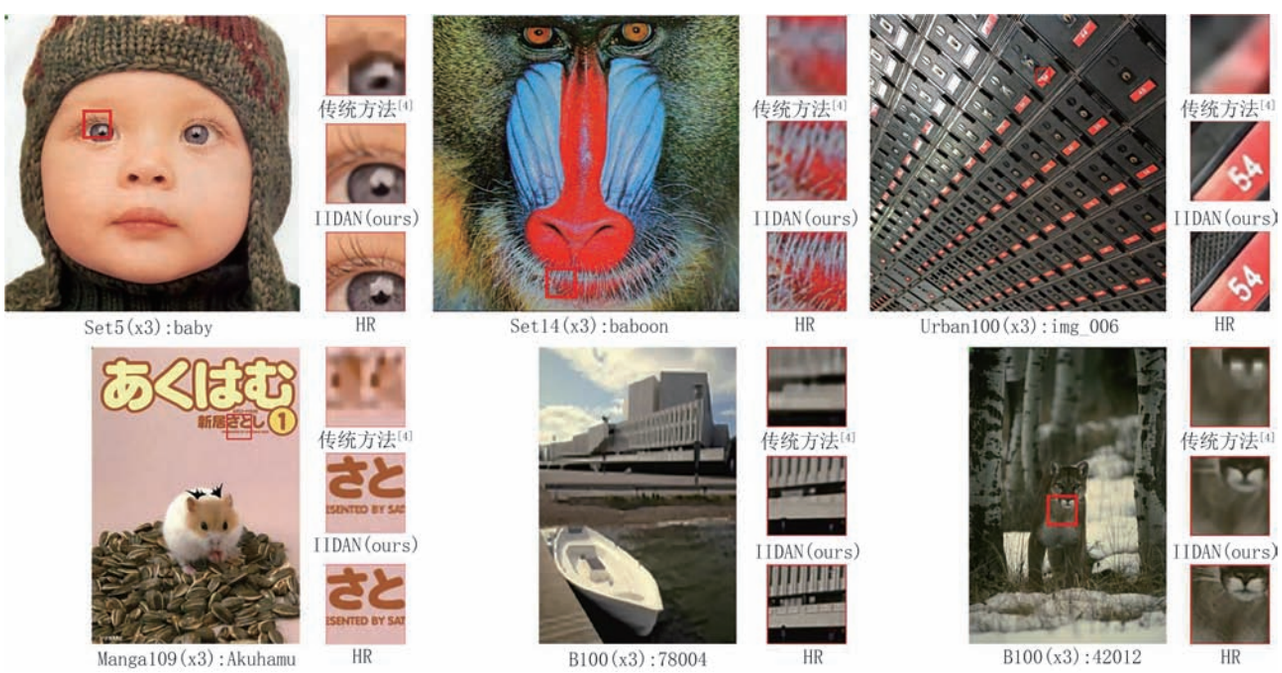



图7　传统方法[4]和本论文提出的IIDAN在×3上的SR主观视觉效果比较图（红色矩形中的对比区域被放大在右侧）

4.3.3　运行时间的比较

除了对方法性能的评估之外，本论文还评估了提出的IIDAN与SwinIR-light[12]、ELAN-light[15]、ESRT[16]、CARN[23]、EMASRN[25]、SRFormer[29]、Omni-SR[30]、CAMixerSR[32]、IMDN[37]、LatticeNet[38]和DRSAN[40]等方法，在放大因子为4、输入的LR图像辨率大小为320×180时的运行时间比较. 表7给出了每种方法采用的网络框架和运行时间的情况. 通过对表7的分析，可以很明显的看到：(1)基于Transformer的方法都会比基于CNN的方法消耗更

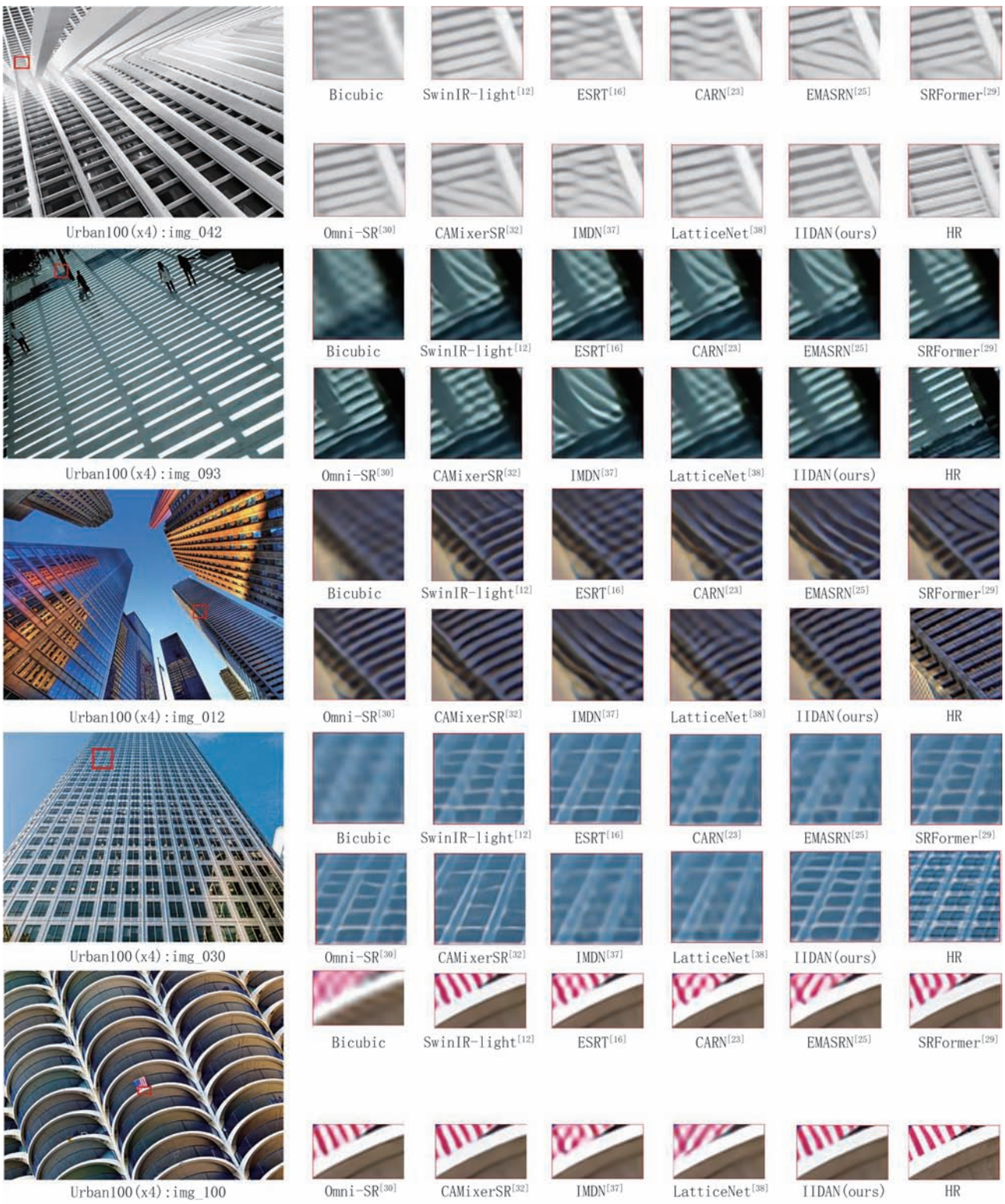



图8 现有的SOTA方法和本论文提出的IIDAN在×4上的SR主观视觉效果比较图(红色矩形中的对比区域被放大在右侧)

多的时间,这是因为自注意力相比卷积运算具有更多的乘法和加法的运算操作;(2)虽然本论文提出的IIDAN的计算复杂度在×2和×3下最低(如表4和表5所示),但是其运行时间却较多,仅比SwinIR-light[12]和Omni-SR[30]分别快了62 ms和7 ms. 这很可能是因为本论文提出的IIDAN采用了较多的深度可分离卷积,从而导致内存访问次数的增加[24,63],因此虽然执行乘法运算和加法运算的次数较少,但是整体的运行时间却反而增加较多,因为内存的速度远低于GPU和CPU的计算速度. 也正因如此,如何能够

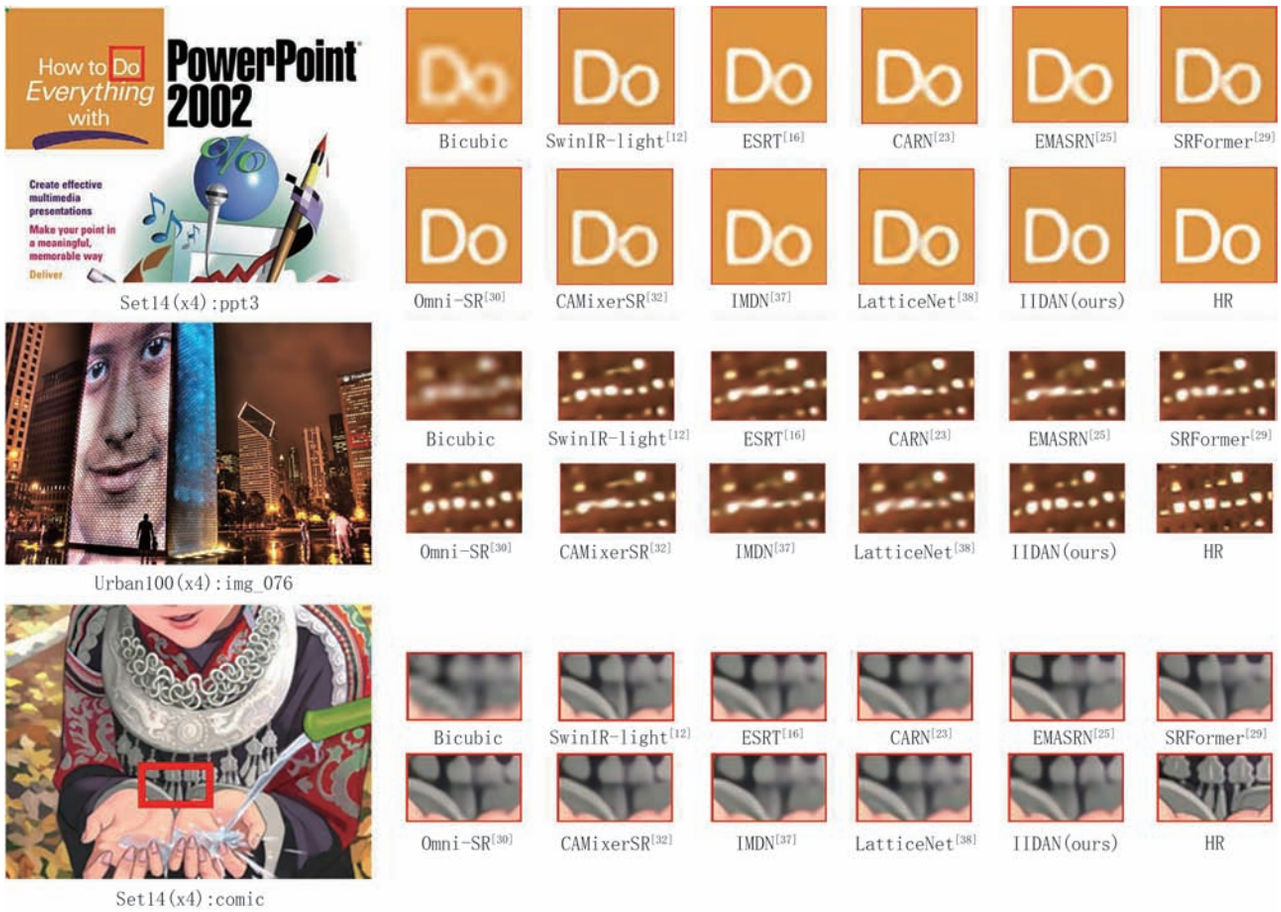



图9　现有的SOTA方法和本论文提出的IIDAN在×4上的SR主观视觉效果比较图(红色矩形中的对比区域被放大在右侧)

**表7　方法的运行时间和网络框架**

| Publication Year | Method | Running time/ms | Architecture |
|---|---|---|---|
| 2021 | SwinIR-light[12] | 271 | Transformer |
| 2022 | ELAN-light[15] | 170 | Transformer |
| 2022 | ESRT[16] | 95 | CNN+Transformer |
| 2018 | CARN[23] | 30 | CNN |
| 2022 | EMASRN[25] | 99 | CNN |
| 2023 | SRFormer[29] | 185 | Transformer |
| 2023 | Omni-SR[30] | 216 | Transformer |
| 2024 | CAMixerSR[32] | 156 | Transformer |
| 2019 | IMDN[37] | 19 | CNN |
| 2020 | LatticeNet[38] | 28 | CNN |
| 2023 | DRSAN[40] | 42 | CNN |
| | IIDAN(ours) | 209 | Transformer |

在保证高性能和低复杂度的同时进一步加速方法的执行时间便成为了未来研究工作的重点.

## 5　结　论

本论文提出了一种基于Transformer的块内块间双聚合的轻量级SISR网络(IIDAN),通过将图像的结构信息统计到一个更低的维度空间,实现了全局范围内相互依赖性的显式捕捉和轻量级.同时,本论文还提出了一种信息交互机制(IIM)来分别对块内自注意力和块间自注意力实行对应信息的补充,进一步增强了网络的特征捕捉和表达能力.实验结果表明,本论文提出的IIDAN能够重建出更高质量的超分辨率图像,同时具有更低的计算复杂度.然而,值得注意的是,因为采用了较多的深度可分离卷积,使得本论文的IIDAN的运行时间较长,因此,如何能够在保证高性能和低复杂度的同时进一步加速方法的执行时间便成为了未来研究工作的重点.

**致　谢**　首先,我们要感谢贵编辑部和所有的审稿人给本论文提出的宝贵意见.其次,我们还要感谢本论文所有作者的辛勤付出.

## 参　考　文　献

[1] Zheng J, Song W, Wu Y, et al. Weighted direct nonlinear regression for effective image interpolation. IEEE Access,