文章编号　1004-924X(2025)12-1955-16

多维度聚合 Transformer 的图像超分辨率重建

陈清江，陈鹏民*

（西安建筑科技大学 理学院，陕西 西安 710055）

摘要：针对现有基于 Transformer 的图像超分辨率网络中感受野尺度单一以及未充分挖掘额外维度信息等问题，本文提出了一种多维度聚合 Transformer 网络。首先，通过构建多尺度交互调制模块，从低分辨率图像中提取多尺度特征，以增强信息流的丰富性。其次，设计了空间-通道交互模块，并将其集成于 Transformer 层中，利用四种形式的注意力机制充分提取关键特征并实现特征融合，从而提升模型性能。最后，提出了特征重用 Transformer 模块，深入挖掘各层特征之间的关联，精准提取并高效重用重要特征，进一步加强模型表现。实验结果表明，在五个基准测试集上，所提方法优于其他先进算法。在不同放大倍数的超分辨率任务中，相较于基于 Swin Transformer 的图像恢复方法，峰值信噪比和结构相似度分别平均提升了约 0.26 dB 和 0.002 4，且重建效果更加清晰。该方法有效克服了现有方法的不足，在超分辨率任务中展现出显著的性能提升和应用潜力。

关　键　词：图像超分辨率；Transformer；注意力机制；特征交互；特征重用；多尺度

中图分类号：TP394.1；TH691.9　　**文献标识码**：A

doi：10.37188/OPE.20253312.1955　　**CSTR**：32169.14.OPE.20253312.1955

MDAT：Multi-dimensional aggregation transformer for image super-resolution reconstruction

CHEN Qingjiang，CHEN Pengmin*

（*School of Science*，*Xi'an University of Architecture and Technology*，*Xi'an* 710055，*China*）

* *Corresponding author*，*E-mail*：409316492@*qq.com*

Abstract：To address the limitations of restricted receptive-field scales and insufficient exploration of additional dimensional information in existing Transformer-based image super-resolution networks，this paper proposed a multi-dimensional aggregation transformer network. First，a multi-scale interaction modulation module was designed to extract multi-scale features from low-resolution images，enhancing the diversity of information flow. Second，a spatial - channel interaction module was integrated into transformer layers，employing four types of attention mechanisms to fully extract key features and achieve effective feature fusion，thereby improving model performance. Third，a feature-reuse transformer module was proposed to explicitly model inter-layer feature relationships，enabling precise extraction and efficient reuse of important features. Experimental results demonstrate that the proposed method outperforms existing state-of-the-art algorithms on five benchmark datasets. Specifically，in super-resolution tasks with various magnification factors，it achieves an average improvement of 0.26 dB in peak signal-to-noise ratio and 0.002 4 in

收稿日期：2025-02-13；**修订日期**：2025-04-28.

基金项目：国家自然科学基金（No. 12101482）；陕西省自然科学基础研究计划（No. 2021JQ-495）

structural similarity index measure compared to Swin Transformer-based methods, producing clearer reconstruction results. These findings validate the effectiveness of the proposed approach and its strong potential for practical applications in image super-resolution tasks.

Key words: image super-resolution; transformer; attention mechanism; feature interaction; feature reuse; multi-scale

1 引　言

单幅图像超分辨率(Single Image Super-Resolution, SISR)是一项经典的低级视觉任务,在医学成像、数字摄影和降低图片传输成本等多个领域应用广泛,其目标是将低分辨率(Low-Resolution, LR)图像重建为细节更加丰富的高分辨率(High-Resolution, HR)图像。由于单一LR输入可能对应多个潜在的HR输出,SISR被视为一种典型的病态逆问题。为应对这一挑战,研究者提出了多种方法。

早期研究主要依赖于插值算法(如双三次插值)其核心思想是基于现有的LR像素值,通过插值算法推算出未知的像素值。尽管这些方法计算简单且效率较高,但由于缺乏对图像内容的深入理解,它们生成的HR图像在细节和纹理恢复上存在明显不足,难以满足高质量图像重建的需求。

随着深度学习的兴起,基于卷积神经网络(Convolutional Neural Networks, CNN)的方法逐渐成为主流。2014年,Dong等首次提出了SRCNN(Super-Resolution Convolutional Neural Network)[1],采用三层卷积结构直接学习LR图像到HR图像的映射,开启了CNN在SISR任务中的应用。随后,Lim等提出EDSR(Enhanced Deep Residual Networks for Super-Resolution)[2],通过移除批归一化层并加深网络结构,在图像重建精度上取得了显著提升。Zhang等在RDN(Residual Dense Network)[3]中引入密集残差块,有效增强了模型特征传递与重用的能力。为了实现模型轻量化,Hui等提出IDN(Information Distillation Network)[4],设计了信息蒸馏模块逐层提取关键特征,并通过级联融合的方式构建了高效的模型架构,进一步提升了模型的表达能力与推理效率。总体而言,基于CNN的方法主要侧重于精细化的架构设计,与早期的方法相比,表现更为优越。然而,这类方法受卷积操作本身的两大局限性所制约:其一,卷积操作为内容无关机制,即相同卷积核对不同图像区域的处理方式一致,难以适应图像内容的多样性;其二,卷积操作受限于局部感受野,在建模长程依赖关系时效率较低,难以充分捕获全局信息。

近年来,Transformer在自然语言处理领域的成功推动了其在计算机视觉中的广泛应用。Transformer的核心机制——自注意力(Self-Attention, SA)能够在序列任意位置间建立全局依赖关系,有效弥补了CNN在长程建模方面的不足。2021年,Liang等首次将Vision Transformer引入SISR任务中,提出了SwinIR(Image Restoration Using Swin Transformer)[5]。该方法采用基于位移的局部窗口自注意力机制,在保证计算效率的同时扩大了建模范围,在多个基准测试集上表现优于主流CNN方法。随后,Zhou等提出Restormer(Efficient Transformer for High-Resolution Image Restoration)[6],通过引入通道维度的自注意力机制,弥补了纯空间建模的不足。进一步地,Wang等提出Omni-SR(Omni Aggregation Network)[7],集成双维度交互模块以实现跨空间与通道的信息融合,进一步提升了模型性能。这些方法不断拓展Transformer在图像细节与全局依赖建模方面的能力,极大地推动了SISR的发展。

尽管上述方法在性能上取得了显著突破,但仍存在以下亟须解决的问题:(1)大多数方法采用固定窗口大小进行自注意力计算,将模型感受野限制在固定尺度内,难以适配复杂纹理与多尺度结构;(2)当前Transformer模块多聚焦于单一维度的信息交互,缺乏对空间、通道与层级之间联合依赖关系的有效建模。尤其在轻量化模型中,缺乏跨层级特征的重用机制,导致模型的表达能力受限。

为此,本文提出了一种多维度聚合Trans-

former（Multi-Dimensional Aggregation Transformer，MDAT）网络，旨在轻量化框架下全面建模多尺度特征、空间-通道交互以及层级特征依赖，以提升 Transformer 模型的性能。本文的主要工作如下：

（1）构建了多尺度交互调制模块，采用多分支卷积结构提取不同尺度的浅层特征，并结合中等尺度特征融合不同感受野下的信息，丰富了信息流，缓解了因固定窗口难以收集多尺度信息的问题。

（2）设计了空间-通道交互模块，通过空间和通道注意力交互分支的协同作用，动态聚合卷积和自注意力获取的特征，增强了模型的表征能力。

（3）提出了特征重用 Transformer 模块，利用自注意力挖掘层级间的特征依赖，实现重要特征的精准提取与高效重用，进一步提升了模型性能。

（4）在多个基准数据集上的实验结果表明，所提方法在保持较低模型复杂度的同时，性能优于现有先进方法。

2　网络模型

2.1　网络结构

本文网络主要包括三个阶段：浅层特征提取、深层特征提取和特征重构，如图 1 上半部分所示。

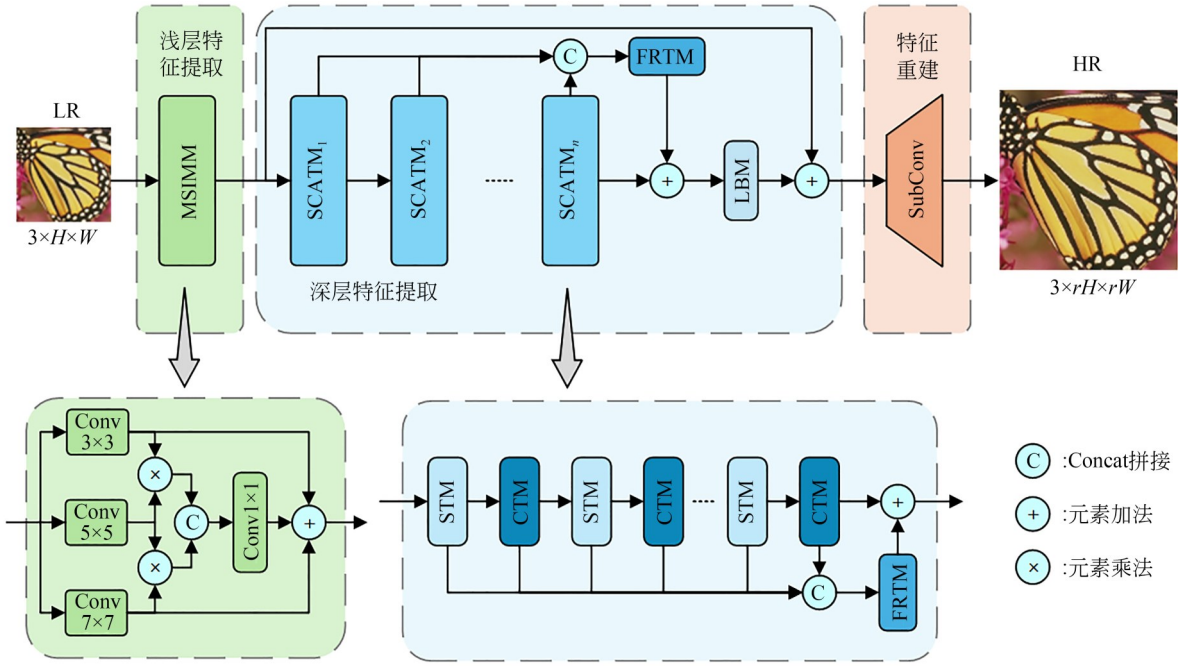

图 1　MDAT 的网络结构

Fig. 1　Network structure of MDAT

网络的总体流程：首先，利用多尺度交互调制模块对 LR 图像进行浅层特征提取：

$$F_0 = H_{\mathrm{MSIMM}}(I_{\mathrm{LR}}), \tag{1}$$

其中：$I_{\mathrm{LR}} \in \boldsymbol{R}^{3\times H\times W}$ 表示输入的 LR 图像，$\boldsymbol{R}$ 表示特征空间，$H\times W$ 表示空间大小，$H_{\mathrm{MSIMM}}(\cdot)$ 表示多尺度交互调制模块，$F_0 \in \boldsymbol{R}^{C\times H\times W}$ 表示浅层特征，C 表示扩展的通道数。

随后，多个空间-通道交替 Transformer 模块用于深入挖掘浅层特征，而特征重用 Transformer 模块则对拼接后多层特征进行筛选和整合，以进一步提升特征的表达能力：

$$F_i = H_{\mathrm{SCATM}_i}(F_{i-1}), \tag{2}$$

$$F_R = H_{\mathrm{FRTM}}(\mathrm{Cat}(F_1, F_2, \cdots, F_n)), \tag{3}$$

$$F_d = H_{\mathrm{LBM}}(F_n + F_R), \tag{4}$$

其中：$H_{\mathrm{SCATM}_i}(\cdot)$ 表示第 i 个空间-通道交替 Transformer 模块，$F_i \in \boldsymbol{R}^{C\times(HW)}$ 表示第 i 个空间-通道交

替 Transformer 模块提取的特征，$i=1,2,\cdots,n$；Cat(•)表示特征拼接操作，$H_{\mathrm{FRTM}}(\cdot)$表示特征重用 Transformer 模块，$F_R\in\boldsymbol{R}^{C\times H\times W}$表示重用特征；$H_{\mathrm{LBM}}(\cdot)$表示轻量化瓶颈模块，它主要由若干卷积和激活函数组成，用于简单的特征融合，$F_d\in\boldsymbol{R}^{C\times H\times W}$表示深层特征。

最后，通过全局残差连接将浅层特征与深层特征相加，再采用亚像素卷积(Sub-pixel Convolution, SubConv)上采样进行特征重构，最终生成超分辨率(Super-Resolution, SR)图像：

$$I_{\mathrm{SR}}=H_{\mathrm{SubConv}}(F_0+F_d),\tag{5}$$

其中：$H_{\mathrm{SubConv}}(\cdot)$表示亚像素卷积上采样，$I_{\mathrm{SR}}\in\boldsymbol{R}^{3\times rH\times rW}$表示输出的 SR 图像，$r$表示放大倍数。

2.2　多尺度交互调制模块

基于 Transformer 的主流 SR 方法普遍采用固定窗口进行局部自注意力计算，虽然在计算效率上具有优势，但也将感受野限制在局部小区域，从而削弱了网络对多尺度信息的捕获能力。

为了缓解以上问题，本文提出了多尺度交互调制模块(Multi-Scale Interaction Modulation Module, MSIMM)，作为 Transformer 主干前的引导模块，通过多尺度卷积提取、跨尺度交互调制和融合残差连接的方式，有效扩展感受野、增强尺度间语义协同与局部细节建模能力，从而提高网络整体性能。模块结构如图 1 左下部分所示。

多尺度特征提取与交互：输入特征首先通过三个并行的卷积(3×3，5×5，7×7)分别提取不同尺度特征：

$$F_S,F_M,F_L=\mathrm{Conv}_{3\times3}(F_{\mathrm{in}}),\mathrm{Conv}_{5\times5}(F_{\mathrm{in}}),\mathrm{Conv}_{7\times7}(F_{\mathrm{in}}),\tag{6}$$

其中，$F_S,F_M,F_L\in\boldsymbol{R}^{C\times H\times W}$分别表示小尺度特征、中等尺度特征和大尺度特征。

为了实现不同尺度特征间的有效交互，本文以中等尺度特征为中介，分别与小尺度特征和大尺度特征进行元素乘法操作，实现特征调制：

$$F_{M_1},F_{M_2}=F_M\otimes F_S,F_M\otimes F_L,\tag{7}$$

其中：$\otimes$表示元素乘法，$F_{M_1},F_{M_2}\in\boldsymbol{R}^{C\times H\times W}$表示交互后的特征。该操作使得中等尺度特征能够结合局部细节(来自 F_S)和大范围上下文信息(来自 F_L)，从而突出重要特征、抑制冗余信息，并增强跨尺度依赖性。

多尺度融合与残差连接：将交互后的特征进行通道拼接，并利用 1×1 卷积进行维度压缩与信息融合，得到融合特征：

$$F_{\mathrm{fusion}}=\mathrm{Conv}_{1\times1}(\mathrm{Cat}(F_{M_1},F_{M_2})),\tag{8}$$

其中，$F_{\mathrm{fusion}}\in\boldsymbol{R}^{C\times H\times W}$表示融合特征。

最后，为了稳定训练过程并提升信息流通效率，将融合特征与小尺度特征和大尺度特征分别进行残差连接，形成一个多尺度残差结构：

$$F_{\mathrm{out}}=F_{\mathrm{fusion}}+F_S+F_L.\tag{9}$$

MSIMM 通过多尺度特征的交互调制机制，在提取多尺度特征的同时，增强了不同尺度特征之间的交互与融合，从而提升了模型对复杂结构与局部细节的建模效果。

2.3　空间–通道交替 Transformer 模块

大多数基于 Transformer 的 SR 方法采用单一维度的特征聚合操作，并使用同质化的聚合方案，即通过级联多个包含空间自注意力和前馈神经网络的 Transformer 层来构建网络主框架。然而，这种方案忽略了通道维度的依赖建模以及层级维度的信息交互对模型性能的重要影响。

受 DAT(Dual Aggregation Transformer)[8]的研究启发，本文设计了空间-通道交替 Transformer 模块(Spatial-Channel Alternating Transformer Module, SCATM)，如图 1 右下部分所示。该模块通过交替堆叠空间 Transformer 模块(详见后文)和通道 Transformer 模块(详见后文)，来捕获空间和通道两个维度的特征，实现了空间与通道信息的互补建模，显著提升特征的表达能力。

具体地，输入特征首先通过空间 Transformer 模块以建模长程空间上下文信息，得到空间增强特征；然后将其传入通道 Transformer 模块以建模全局通道依赖，得到通道增强特征。随后，通道增强特征再次送入下一个空间 Transformer 模块，空间与通道建模过程交替进行，以逐步增强特征的表达能力。具体过程如下：

$$\begin{aligned}&F_{S_1}=H_{\mathrm{ST}_1}(F_{S_0}),F_{C_1}=H_{\mathrm{CT}_1}(F_{S_1})\\&F_{S_2}=H_{\mathrm{ST}_2}(F_{C_1}),F_{C_2}=H_{\mathrm{CT}_2}(F_{S_2})\\&\cdots\\&F_{S_i}=H_{\mathrm{ST}_i}(F_{C_i}),F_{C_i}=H_{\mathrm{CT}_i}(F_{S_i})\end{aligned},\tag{10}$$

其中：$F_{S_0}\in\boldsymbol{R}^{C\times(HW)}$表示输入特征，$H_{\mathrm{ST}_i}(\cdot)$和

$H_{\mathrm{CT}_i}(\cdot)$分别表示第i个空间 Transformer 模块和第i个通道 Transformer 模块，$F_{S_i}\in R^{C\times(HW)}$和$F_{C_i}\in R^{C\times(HW)}$分别表示由第$i$个空间 Transformer 模块提取的特征和第i个通道 Transformer 模块提取的特征，$i=1,2,\cdots,n$。

为进一步挖掘层级维度的特征依赖，SCATM 采用特征重用 Transformer 模块（详见后文）增强跨层特征交互。具体过程如下：

$$F_R=H_{\mathrm{FRTM}}(\mathrm{Cat}(F_{S_1},F_{C_1},\cdots,F_{S_n},F_{C_n})),\quad(11)$$

$$F_{\mathrm{SCATM}}=F_{C_n}+F_R,\quad(12)$$

其中：$H_{\mathrm{FRTM}}(\cdot)$表示特征重用 Transformer 模块，$F_R\in R^{C\times H\times W}$表示重用特征；$F_{\mathrm{SCATM}}\in R^{C\times H\times W}$表示空间-通道交替 Transformer 模块的输出。

SCATM 通过统一处理空间、通道与层级三大维度之间的信息交互，利用空间-通道交替结构有效融合局部与全局上下文信息。同时采用特征重用 Transformer 模块促进了层级间信息的交互，从而构建了跨维度协同优化的 Transformer 单元，实现了多个维度上的信息挖掘，进而提升了模型性能。

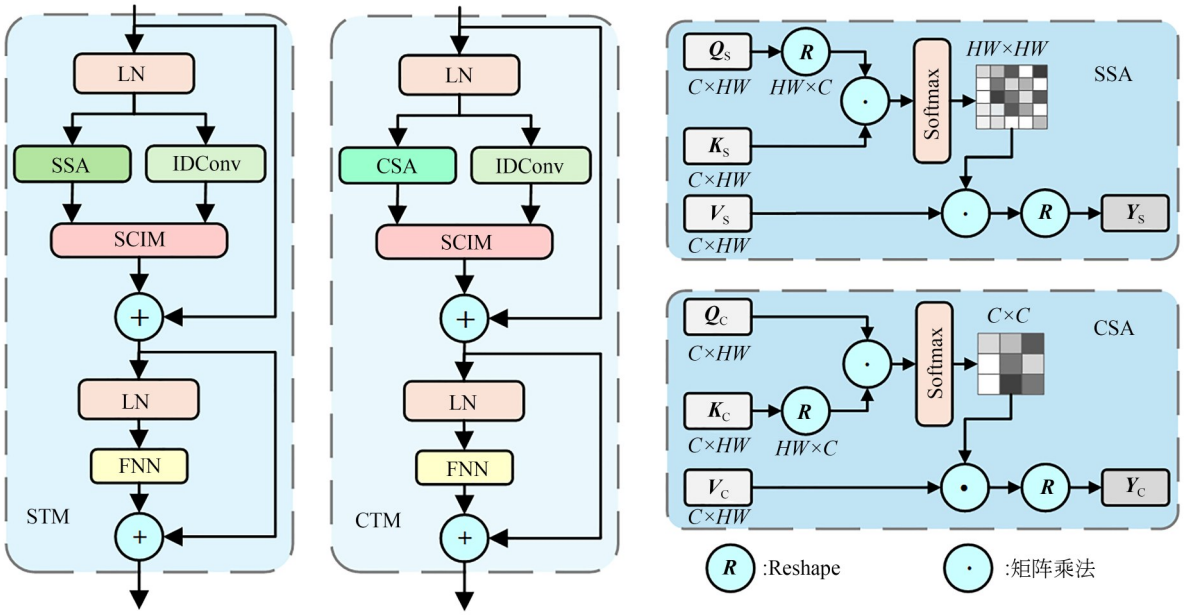

图 2　空间 Transformer 模块和通道 Transformer 模块的结构

Fig. 2　Structure of the spatial transformer module and channel transformer module

2.4　空间 Transformer 模块和通道 Transformer 模块

传统 Transformer 架构以空间自注意力与前馈神经网络为核心，擅长建模长程依赖关系，但在捕获局部细节方面存在不足。相比之下，CNN 在局部建模能力上更具优势。因此，融合 Transformer 的全局建模能力与 CNN 的局部感知能力成为提升特征建模质量的有效途径。

然而，二者在信息感受范围和表征方式上存在显著差异，直接融合可能导致特征颗粒度不匹配。为缓解该问题，本文在空间 Transformer 模块和通道 Transformer 模块中均引入大核分解卷积[9]与本文设计的空间-通道交互模块（详见后文），以统一特征颗粒度、提升融合效果。

空间 Transformer 模块（Spatial Transformer Module，STM）和通道 Transformer 模块（Channel Transformer Module，CTM）的具体结构如图 2 所示。两者的主要区别在于自注意力机制关注的维度不同：STM 主要聚焦于空间维度的上下文建模，而 CTM 则专注于通道维度的信息交互。

STM 的整体流程：首先对输入特征进行层归一化（LN）处理，以增强训练稳定性和特征分布一致性：

$$F_{\mathrm{in}}'=\mathrm{LN}(F_{\mathrm{in}}).\quad(13)$$

将归一化后的特征，分别送入空间自注意力（SSA）和大核分解卷积（IDConv）进行特征提取：

$$F_{\mathrm{SSA}}, F_{\mathrm{Conv}} = \mathrm{SSA}(F_{\mathrm{in}}'), \mathrm{IDConv}(F_{\mathrm{in}}'), \quad (14)$$

其中：SSA 通过局部窗口的自注意力机制，关注空间位置的长程依赖；IDConv 借助卷积结构增强局部感受能力，弥补自注意力机制对细节建模的不足。

随后，将注意力分支与卷积分支提取的特征送入空间-通道交互模块，以实现跨维度的信息融合：

$$F_{\mathrm{SA}} = \mathrm{SCIM}(F_{\mathrm{SSA}}, F_{\mathrm{Conv}}), \quad (15)$$

其中，SCIM(•,•)表示空间-通道交互模块。融合后的特征与输入特征进行残差连接后，输入前馈神经网络(FFN)进一步提升特征表达能力：

$$\begin{aligned} F_{\mathrm{mid}} &= F_{\mathrm{in}} + F_{\mathrm{SA}} \\ F_{\mathrm{FFN}} &= \mathrm{FFN}(\mathrm{LN}(F_{\mathrm{mid}}))' \end{aligned} \quad (16)$$

其中：$F_{\mathrm{mid}} \in \boldsymbol{R}^{C\times H\times W}$ 表示中间特征；$F_{\mathrm{FFN}} \in \boldsymbol{R}^{C\times H\times W}$ 表示前馈网络的输出。

最终，将前馈神经网络的输出再次与中间特征进行残差连接，作为 STM 的输出：

$$F_{\mathrm{out}} = F_{\mathrm{FFN}} + F_{\mathrm{mid}}, \quad (17)$$

CTM 的具体过程与 STM 类似，此处不再赘述。

STM 和 CTM 分别针对空间和通道两个维度设计自注意力机制，显式区分建模方向，避免信息混淆。两者均引入卷积增强局部感知能力，并通过空间-通道交互模块实现高效协同融合，有助于提升模型对全局上下文与局部细节信息的建模能力。

2.5　空间-通道交互模块

为了充分聚合卷积和自注意力获取的特征，本文在 STM 和 CTM 中，设计了空间-通道交互模块(Spatial-Channel Interaction Module，SCIM)，以实现空间维度与通道维度的双向交互。该模块不仅融合了 STM 中空间自注意力提取的特征和 CTM 中通道自注意力提取的特征，还通过独立的空间交互分支与通道交互分支，引入额外的交互式注意力机制，从而实现空间-通道、局部-全局四重信息的协同建模。

如图 3 所示，SCIM 由通道交互分支和空间交互分支组成，分别用于建模通道间的依赖关系和空间维度中的全局结构信息。输入特征分别经过两个分支提取交互注意力后，再与原始输入进行“蝴蝶结”式(即交叉乘法)的特征交互。最终通过加法整合为统一输出。该过程如公式(18)～式(20)所示：

$$F_C = \mathrm{CI}(F_{\mathrm{in}_1}), \quad (18)$$

$$F_S = \mathrm{SI}(F_{\mathrm{in}_2}), \quad (19)$$

$$F_{\mathrm{out}} = F_{\mathrm{in}_1} \otimes F_S + F_{\mathrm{in}_2} \otimes F_C, \quad (20)$$

其中：$F_{\mathrm{in}_1} \in \boldsymbol{R}^{C\times H\times W}$ 表示通道交互分支的输入，CI(•)表示通道交互分支，$F_C \in \boldsymbol{R}^{C\times 1\times 1}$ 表示通道注意力权重；$F_{\mathrm{in}_2} \in \boldsymbol{R}^{C\times H\times W}$ 表示空间交互分支的输入，SI(•)表示空间交互分支，$F_S \in \boldsymbol{R}^{1\times H\times W}$ 表示空间注意力权重；$F_{\mathrm{out}} \in \boldsymbol{R}^{C\times H\times W}$ 表示空间-通道交互模块的输出。

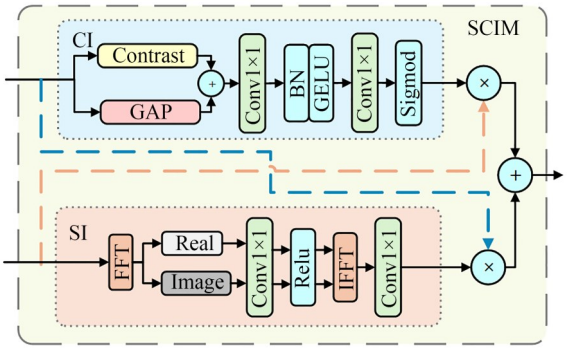

图 3　空间-通道交互模块的结构

Fig. 3　Structure of the spatial-channel interaction module

通道交互分支：该分支借鉴统计信息提取策略，结合图像对比度和全局平均特征，生成通道注意力权重。具体而言，输入特征先进行对比度计算和全局平均池化操作，两者相加后作为初步表示。随后，通过 1×1 卷积进行通道压缩，批归一化(BN)与 GELU 激活函数进行特征增强。最后，再次通过另一个 1×1 卷积恢复通道数，并用 Sigmoid 激活函数生成通道注意力权重。该过程如公式(21)～式(22)所示：

$$F_{\mathrm{mid}} = \mathrm{Conv}_{1\times 1}(\mathrm{CT}(F_{\mathrm{in}_1}) + \mathrm{GAP}(F_{\mathrm{in}_1})), \quad (21)$$

$$F_C = \mathrm{Sigmod}(\mathrm{Conv}_{1\times 1}(\mathrm{GELU}(\mathrm{BN}(F_{\mathrm{mid}})))), \quad (22)$$

其中：CT(•)表示对比度计算，GAP(•)表示全局平均池化操作，$F_{\mathrm{mid}} \in \boldsymbol{R}^{C/r\times H\times W}$ 表示中间特征，r 表示通道压缩系数。

空间交互分支：该分支通过引入傅里叶变换拓展传统空间注意力的建模范围。具体而言，首先将输入特征经过傅里叶变换映射到频域，分别提取其实部和虚部，随后通过参数共享的 1×1 卷

积与 Relu 激活函数生成频域注意力权重，最后经过傅里叶逆变换和 1×1 卷积还原为空间注意力权重。此方法在保留空间结构的同时扩大了感受野。该过程如式(23)～式(25)所示：

$$F_{re}, F_{im} = \mathrm{FFT}(F_{in_2}), \tag{23}$$

$$F_{re}', F_{im}' = \mathrm{Relu}(\mathrm{Conv}_{1\times1}(F_{re}, F_{im})), \tag{24}$$

$$F_S = \mathrm{Conv}_{1\times1}(\mathrm{IFFT}(F_{re}', F_{im}')), \tag{25}$$

其中：FFT(•)表示傅里叶变换，$F_{re}\in R^{C\times H\times W}$，$F_{im}\in R^{C\times H\times W}$分别表示频域特征的实部和虚部；$F_{re}'\in R^{C\times H\times W}$，$F_{im}'\in R^{C\times H\times W}$表示频域注意力权重；IFFT(•, •)表示傅里叶逆变换。

通过将 SCIM 放置于 STM 和 CTM 中，模型得以同时融合空间自注意力、通道自注意力、空间交互注意力和通道交互注意力这四种形式的注意力机制，实现了特征在多个维度上的交互与融合，从而有效提升图像重建过程中的细节还原能力与结构保持能力。

2.6　特征重用 Transformer 模块

对于网络层数较浅的轻量级网络而言，特征重用尤为关键。大多数现有网络在设计中未充分考虑这一点，或仅通过简单的残差连接实现特征重用，这在一定程度上限制了模型性能的提升。同时，目前的自注意力机制仅在空间和通道维度进行计算，忽略了层级维度的潜在依赖关系，这也成为性能优化的瓶颈。

针对上述问题，本文提出了特征重用 Transformer 模块(Feature Reuse Transformer Module，FRTM)，如图 4 所示。该模块通过对拼接后的各层特征，在层级维度上应用自注意力机制，显式建模不同层级间的关联性，并结合可学习权重对各层特征进行加权融合，从而实现更有效的特征重用与信息整合。

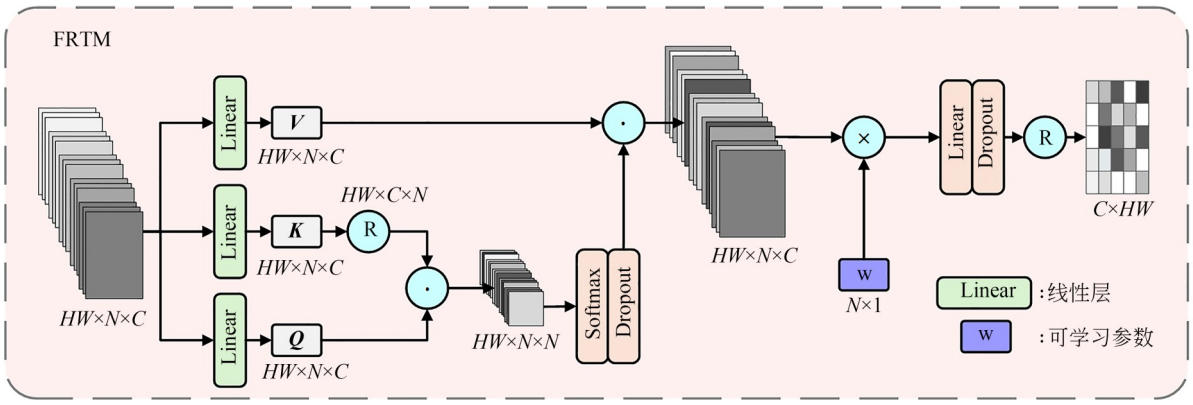

图 4　特征重用 Transformer 模块的结构

Fig. 4　Structure of the feature reuse transformer module

具体而言，首先将各个层级拼接后的特征通过线性层(Linear)转换为查询($\boldsymbol{Q}$)、键($\boldsymbol{K}$)和值($\boldsymbol{V}$)：

$$\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} = \mathrm{Linear}(F_{in}), \mathrm{Linear}(F_{in}), \mathrm{Linear}(F_{in}), \tag{26}$$

其中：$F_{in}\in R^{(HW)\times N\times C}$表示输入特征，$N$表示层数，Linear(•)表示线性层，$\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}\in R^{(HW)\times N\times C}$分别表示变换后的特征：查询、键和值。

为了挖掘层级之间的关系，$\boldsymbol{Q}$与$\boldsymbol{K}$在层级维度(N)上通过矩阵乘法实现自注意力计算。随后，使用 Softmax 激活函数生成层级注意力权重，并将其与$\boldsymbol{V}$相乘，得到层级自注意力的输出：

$$F_{SA} = \mathrm{Softmax}(\boldsymbol{Q}\cdot\boldsymbol{K}^{\mathrm{T}})\cdot\boldsymbol{V}, \tag{27}$$

其中：• 表示矩阵乘法，$(\cdot)^{\mathrm{T}}$表示矩阵的转置，$F_{SA}\in R^{(HW)\times N\times N}$表示层级自注意力权重。

最后，将自注意力输出与可学习参数w加权求和，并通过线性层转换为最终的输出：

$$F_{out} = \mathrm{Resh}(\mathrm{Linear}(F_{SA}\otimes w)), \tag{28}$$

其中：w表示可学习参数，Resh(•)表示形状重排操作，$F_{out}\in R^{C\times H\times W}$表示输出的重用特征。此外，本文在激活函数层和末端线性变换层之后引入了 Dropout 操作，用于抑制模型过拟合，提升泛化性能。

为了充分发挥该模块的作用，FRTM 被灵活嵌入到网络的多个关键位置，如图 1 所示。一方面，它被集成在 SCATM 的内部，用于模块内不

同阶段的特征重用；另一方面，它也应用于SCATM的外部，以实现跨模块间的信息流动。这种策略实现了模块内部和外部的双重特征重用，进一步提升了模型性能。

2.7 损失函数

与其他轻量级SR算法一致，本文选择L_1损失函数作为模型的优化目标，其表达式为：

$$L_1=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-f(x_i)\right|, \tag{29}$$

其中：n表示图像的数量；y_i表示HR图像；x_i表示LR图像；$f(x_i)$表示SR图像；$f(\cdot)$表示SR模型。

3 实验结果与分析

3.1 数据集和评价指标

本文以DF2K作为模型的训练集。DF2K由DIV2K和Flickr2K两个数据集整合而成，包含3550幅HR图像，其对应的LR图像是通过对HR图像进行双三次插值下采样生成的。为确保与其他研究的公平对比，实验选用了五个广泛使用的公开测试集：Set5，Set14，BSD100，Urban100以及Manga109。其中Set5与Set14多为基础图像，适合初步测试；BSD100则包含多样化的自然图像内容，可用于评估模型的通用性；Urban100包含大量城市建筑图像，具有丰富的高频细节，挑战性较高；而Manga109则以漫画风格为主，含有明确的线条和区域块，常用于测试模型在非自然纹理图像上的能力。通过这些具有代表性的数据集，能够全面评估所提方法在不同图像风格和复杂度下的表现。

在性能评价指标方面，本文与先前研究保持一致，采用峰值信噪比（Peak Signal-to-Noise Ratio，PSNR）和结构相似性（Structural Similarity，SSIM）作为客观质量的评价标准，其中PSNR的单位为：分贝（dB）。具体计算时，先将图像从RGB空间转换到YC_bC_r空间，并在Y通道上测量PSNR和SSIM。此外，为了更全面评估模型在实际应用中的性能，本文还引入了参数量（Number of Parameters，Params）和乘加运算量（Multiply-Add Operations，MACs），分别衡量模型的存储需求和计算复杂度，其单位分别为千（Kilo，K）和十亿（Giga，G）。

3.2 训练细节及参数设置

实验环境：本文所有实验均在以下硬件和软件环境下完成：操作系统为Ubuntu 20.04，处理器为AMD EPYC 7502（16核），系统内存为80 GB，显卡为NVIDIA GeForce RTX 4090（24 GB）。实验基于CUDA 12.2和PyTorch 2.3的深度学习框架，编译环境为PyCharm，编程语言为Python 3.10。

实验细节：本文模型主要由1个MSIMM，2个SCATM和1个FRTM组成，每个SCATM包含4个STM，4个CTM和1个FRTM，其中，STM采用窗口自注意力机制，窗口大小为8×32，自注意力头数设为8；CTM则采用全局自注意力机制，不使用窗口划分，同样设置8个自注意力头。训练过程中，图像被裁剪为64×64的小块作为输入，并采用数据增强策略，包括随机水平翻转和旋转。使用Adam优化器，其中$\beta_1=0.9$和$\beta_2=0.99$。输入批次大小为24，特征通道数为80。学习率采用多步衰减策略进行更新，初始学习率为2×10^{-4}，并在迭代次数为[250 K，400 K，450 K，475 K]时将学习率减半，训练总迭代次数为50 W。

3.3 对比试验

定量分析：为验证本文方法的优越性，将其与近年来图像超分辨率重建领域的多种先进方法进行对比。这些方法包括SRCNN（Super-Resolution Convolutional Neural Network），FSRCNN（Fast Super-Resolution Convolutional Neural Network）[10]，CARN（Cascading Residual Network）[11]，SwinIR-L（Lightweight Swin Transformer）[5]，DAT-L（Lightweight Dual Aggregation Transformer）[8]，Omni-SR，SwinIR-NG（N-gram in swin transformers）[12]，CRAFT（Cross-Refinement Adaptive Feature Modulation Transformer）[13]，MambaIR-light（Lightweight Image Restoration with State-Space Model）[14]，SRConvNet-L（Lightweight Transformer-Style ConvNet）[15]，HiT-SNG（Hierarchical Transformer）[16]。

图5展示了本文算法与其他9种先进算法在不同放大倍数下的性能对比。其中，PSNR经过标准化处理，本文算法以粗实线标注。从图中可以看出，本文算法在多个数据集上的PSNR表现

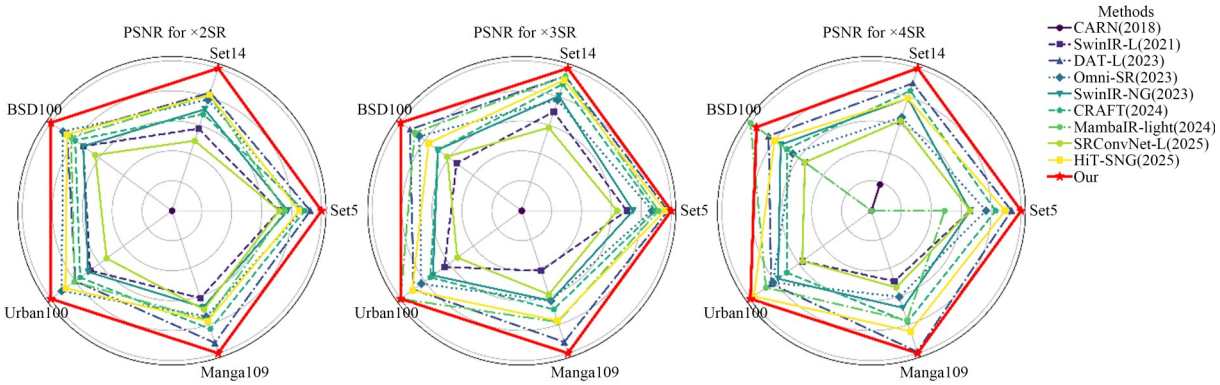

图 5　不同算法的性能对比

Fig. 5　Performance comparison of different algorithms

优于现有方法，尤其在高难度数据集 Urban100 和 Manga109 上，展现出显著的性能优势。与近期方法（如 DAT-L 和 Omni-SR）相比，不仅在复杂场景下表现卓越，在较易处理的数据集（如 Set5 和 Set14）上也取得了优异的性能。此外，无论在何种放大倍数下，本文方法的表现均衡，体现了良好的泛化能力和鲁棒性。

表 1～表 3 分别详细列出了在五个基准数据集上，不同算法在处理不同放大倍数的 SR 任务时的模型复杂度和性能分析结果。其中，MACs 是在输出分辨率为 1 280×720 的条件下计算的。为便于对比，性能指标中的最佳值以加粗形式标注。

从表 1～表 3 的数据可以看出，本文提出的方法在多个测试集上的 PSNR 和 SSIM 指标超越了当前最先进的算法。与 SwinIR-L 相比，本文方法在不同放大倍数的 SR 任务中，PSNR 平均提升约 0.26 dB，充分体现了其在图像重建质量上的卓越表现。同时，这一性能提升并未显著增加模型复杂度，证明了本文方法在效率与性能之间达成的良好平衡。这些定量结果充分验证了本文方法在 SR 任务中的高效性和竞争力。

定性比较：为了进一步评估本文方法在图像重建质量上的表现，与其他 8 种算法进行了视觉效果的对比。图 6 展示了五组不同算法在处理×4SR 任务时的可视化结果，图像 Image_062，Im-

表 1　在基准数据集上×2SR 任务的定量评估

Tab. 1　Quantitative evaluation of ×2 SR task on benchmark datasets

方法	模型复杂度	Set5	Set14	BSD100	Urban100	Manga109
	Params/MACs	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic	—	33.66/0.929 9	30.24/0.868 8	29.56/0.843 1	26.88/0.840 3	30.80/0.933 9
SRCNN(2016)	8/53	36.66/0.954 2	32.45/0.906 7	31.36/0.887 9	29.50/0.894 6	35.60/0.966 3
FSRCNN(2016)	13/6	37.00/0.955 8	32.63/0.908 8	31.53/0.892 0	29.88/0.902 0	36.67/0.971 0
CARN(2018)	1 592/222.8	37.76/0.959 0	33.52/0.916 6	32.09/0.897 8	31.92/0.925 6	38.36/0.976 5
SwinIR-L(2021)	910/244.4	38.14/0.961 1	33.86/0.920 6	32.31/0.901 2	32.76/0.934 0	39.12/0.978 3
DAT-L(2023)	763/173.0	38.25/0.961 5	34.01/0.921 5	32.35/0.902 1	32.91/0.934 8	39.51/0.979 0
Omni-SR(2023)	772/194.5	38.22/0.961 3	33.98/0.921 0	32.36/0.902 0	33.05/0.936 3	39.28/0.978 4
SwinIR-NG(2023)	1 181/274.1	38.17/0.961 2	33.94/0.920 5	32.31/0.901 3	32.78/0.934 0	39.20/0.978 1
CRAFT(2024)	737/176.6	38.23/0.961 5	33.92/0.921 1	32.33/0.901 6	32.86/0.934 3	39.39/0.978 6
MambaIR-light(2024)	1 363/278.9	38.16/0.961 0	34.00/0.921 2	32.34/0.901 7	32.92/0.935 6	39.31/0.977 9
SRConvNet-L(2025)	885/160	38.14/0.961 0	33.81/0.919 9	32.28/0.901 0	32.59/0.932 1	39.22/0.977 9
HiT-SNG(2025)	1 013/213.9	38.21/0.961 2	34.00/0.921 7	32.35/0.902 0	33.01/0.936 0	39.32/0.978 2
Ours	997/289.4	**38.29/0.961 6**	**34.11/0.921 9**	**32.39/0.902 4**	**33.16/0.936 9**	**39.60/0.979 1**

表 2　在基准数据集上×3SR 任务的定量评估

Tab. 2　Quantitative evaluation of ×3 SR task on benchmark datasets

方法	模型复杂度	Set5	Set14	BSD100	Urban100	Manga109
	Params/MACs	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic	—	30.39/0.868 2	27.55/0.774 2	27.21/0.738 5	24.46/0.734 9	26.95/0.855 6
SRCNN(2016)	8/53	32.75/0.909 0	29.30/0.821 5	28.41/0.786 3	26.24/0.798 9	30.48/0.911 7
FSRCNN(2016)	13/5	33.18/0.914 0	29.37/0.824 0	28.53/0.791 0	26.43/0.808 0	31.10/0.921 0
CARN(2018)	1 592/119.0	34.29/0.925 5	30.29/0.840 7	29.06/0.803 4	28.06/0.849 3	33.50/0.944 0
SwinIR-L(2021)	918/110.8	34.62/0.928 9	30.54/0.846 3	29.20/0.808 2	28.66/0.862 4	33.98/0.947 8
DAT-L(2023)	947/96.7	34.75/0.929 9	30.63/0.847 5	29.30/0.810 6	28.90/0.866 8	34.55/0.950 2
Omni-SR(2023)	780/88.4	34.70/0.929 4	30.57/0.846 9	29.28/0.809 4	28.84/0.865 6	34.22/0.948 7
SwinIR-NG(2023)	1 190/114.1	34.64/0.929 3	30.58/0.847 1	29.24/0.809 0	28.75/0.863 9	34.22/0.948 8
CRAFT(2024)	744/78.4	34.71/0.929 5	30.61/0.846 9	29.24/0.809 3	28.77/0.863 5	34.29/0.949 1
MambaIR-light(2024)	1 371/124.6	34.72/0.929 6	30.63/0.847 5	29.29/0.809 9	**29.00/0.868 9**	34.39/0.949 0
SRConvNet-L(2025)	906/74	34.59/0.928 8	30.50/0.845 5	29.22/0.808 1	28.56/0.860 0	34.17/0.947 9
HiT-SNG(2025)	1 021/99.5	34.74/0.929 7	30.62/0.847 4	29.26/0.810 0	28.91/0.867 1	34.38/0.949 5
Ours	1 007/129.8	**34.76/0.930 0**	**30.65/0.847 6**	**29.32/0.810 7**	**29.00**/0.868 3	**34.64/0.950 5**

表 3　在基准数据集上×4SR 任务的定量评估

Tab. 3　Quantitative evaluation of ×4 SR task on benchmark datasets

方法	模型复杂度	Set5	Set14	BSD100	Urban100	Manga109
	Params/MACs	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic	—	28.42/0.810 4	26.00/0.702 7	25.96/0.667 5	23.14/0.657 7	24.89/0.786 6
SRCNN(2016)	8/53	30.48/0.862 6	27.50/0.751 3	26.90/0.710 1	24.52/0.722 1	27.58/0.855 5
FSRCNN(2016)	13/4.6	30.72/0.866 0	27.61/0.755 0	26.98/0.715 0	24.62/0.728 0	27.90/0.861 0
CARN(2018)	1 592/91.0	32.13/0.893 7	28.60/0.780 6	27.58/0.734 9	26.07/0.783 7	30.47/0.908 4
SwinIR-L(2021)	930/63.6	32.44/0.897 6	28.77/0.785 8	27.69/0.740 6	26.47/0.798 0	30.92/0.915 1
DAT-L(2023)	910/78.4	32.57/0.899 2	28.87/0.788 0	27.75/0.743 1	26.65/0.803 5	31.37/**0.917 9**
Omni-SR(2023)	792/50.9	32.49/0.898 8	28.78/0.785 9	27.71/0.741 5	26.64/0.801 8	31.02/0.915 1
SwinIR-NG(2023)	1 201/64.4	32.44/0.898 0	28.83/0.787 0	27.73/0.741 8	26.61/0.801 0	31.09/0.916 1
CRAFT(2024)	753/46.5	32.52/0.898 9	28.85/0.787 2	27.72/0.741 8	26.56/0.799 5	31.18/0.916 8
MambaIR-light(2024)	1 280/42.9	32.36/0.898 4	28.53/0.789 5	**27.78/0.744 6**	26.68/0.805 7	31.17/0.917 6
SRConvNet-L(2025)	902/45	32.44/0.897 6	28.77/0.785 7	27.69/0.740 2	26.47/0.797 0	30.96/0.913 9
HiT-SNG(2025)	1 032/57.7	32.55/0.899 1	28.83/0.787 3	27.74/0.742 6	26.75/0.805 3	31.24/0.917 6
Ours	1 023/73.8	**32.60/0.899 7**	**28.91/0.788 5**	27.77/0.743 4	**26.77/0.806 3**	**31.38**/0.917 2

age_092 和 Image_011 选自高难度数据集 Urban100，图像 YumeiroCooking 选自 Manga109 数据集，图像 10805 选自 BSD100 数据集。其中，矩形框标注为局部放大区域，全局图的右下角为 HR 参考图。

第一组高楼建筑纹理重建：在规则网格纹理和直线边缘的恢复上，本文方法表现突出，清晰地重建了墙体的网格结构。相比之下，Bicubic 和 CARN 方法存在明显的纹理模糊现象，而 SwinIR-NG 在部分区域表现略显平滑。本文方法在边缘锐利度上显著优于 CRAFT 和 HiT-SNG，能够有效恢复建筑图像的细节；第二组几何建筑细节重建：本文方法能够精准恢复几何图案的完整结构和清晰边缘，成功减少了 Bicubic，CARN 和 SwinIR-L 方法中常见的模糊和锯齿伪影。同时，本文方法在黑白纹理区域的高对比度保持上，优

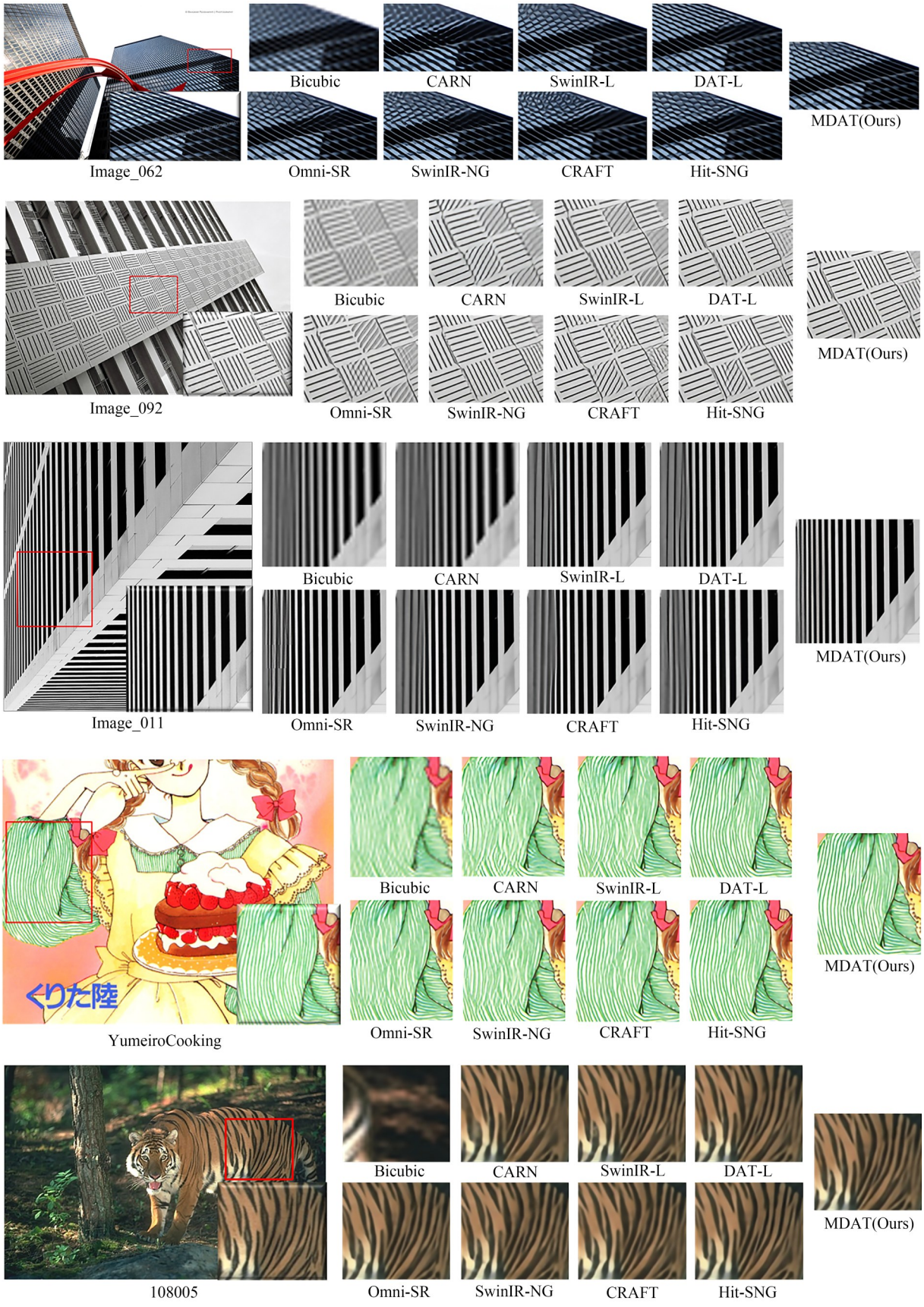

图 6　不同算法处理×4SR 任务的可视化

Fig. 6　Visualization of ×4SR task for different algorithms

于SwinIR-NG和CRAFT，避免了纹理信息的丢失；第三组高对比条纹区域重建：在条纹的连续性和锐利度上，本文方法的重建效果接近HR参考图，展现了优秀的细节恢复能力。相较而言，Bicubic和CARN方法的条纹出现了断裂，其他方法则在一定程度上表现出过度平滑。本文方法在条纹边界细节的处理上更加精细，成功避免了纹理溢出和模糊问题；第四组漫画图像重建：在漫画图像的重建中，本文方法展现了显著的优势，尤其是在绿色线条区域的细节恢复上，成功避免了Bicubic和CARN方法在高频细节上的模糊。SwinIR-L和SwinIR-NG虽然在图像质量上有所改进，但仍无法精确恢复漫画图像中复杂的线条和色块交界。相比之下，本文方法能够清晰地重建这些细节，并保持了漫画图像的独特风格；第五组自然图像重建：在自然图像的纹理和细节恢复上，本文方法展现了强大的能力，重建效果远超Bicubic和CARN方法。SwinIR-L在一定程度上恢复了老虎皮毛的纹理，但整体图像在细节和清晰度上较为平滑。CRAFT和HiT-SNG在高对比度区域表现较好，但无法完全保留自然场景中的精细纹理。本文方法在高频纹理和细节部分的恢复上，表现出了更高的精度，避免了其他方法常见的模糊问题。

可视化结果表明，本文方法在多种图像类型上均表现出色。无论是漫画图像、建筑细节，还是自然场景中的复杂纹理，均能有效恢复细节，保持较好的重建效果，展现出良好的泛化能力。

推理效率分析：为了评估所提方法在实际应用场景下的推理效率，本文在保持输出分辨率一致的前提下，测试了模型处理单帧图像的平均推理时间。表4给出了SwinIR-L与MDAT在不同放大倍数下单帧图像平均处理时间的对比结果。

从表4中可以看出，尽管MDAT在×2SR任务下略慢于SwinIR-L，但在×3SR和×4SR任务下表现出更优的推理速度，说明了MDAT在保持更优重建质量的同时具有良好的推理效率。

表4　单帧图像平均推理时间

Tab. 4　Per-frame average inference time

放大倍数	输出分辨率	单帧图像平均推理时间/s	
		SwinIR-L	MDAT(Ours)
×2	1 280×720	1.164 7	1.190 0
×3		0.510 9	0.501 8
×4		0.282 3	0.264 2

综上所述，本文方法在定量指标、主观视觉效果以及模型复杂度之间达到了较为均衡的表现，体现了较强的综合性能与实用价值。

3.4　消融实验分析

为了验证各模块的有效性，本文针对多尺度交互调制模块、空间－通道交互模块、特征重用Transformer模块及其组合效果，进行了消融实验。实验基于Urban100和Manga109数据集，并以×2SR任务为研究对象，可视化结果选取的局部图像来自Urban100数据集中的image_49。

多尺度交互调制模块的消融实验：为评估多尺度交互调制模块的贡献，设计了三个实验对象。首先，将多尺度交互调制模块替换为常用的单个3×3卷积，作为对照实验(Re-Conv3×3)。其次，将多尺度交互调制模块中并行的三个卷积统一替换为3×3卷积，作为第二个实验对象(Re-3*Conv3×3)。最后，使用MSIMM作为第三个实验对象(W-MSIMM)，实验结果如表5和图7所示。

实验结果表明，与其他两组实验对象相比，W-MSIMM在PSNR和SSIM指标上均取得了提升，同时视觉效果最接近HR图像，并且仅增加少量的模型复杂度，充分验证了该模块多尺度设计的有效性。

表5　多尺度交互调制模块的消融实验

Tab. 5　Ablation study of MSIMM

实验对象	模型复杂度	Urban100	Manga109
	Params/MACs	PSNR/SSIM	PSNR/SSIM
Re-Conv3×3	871/197.4	33.00/0.935 7	39.50/0.979 0
Re-3*Conv3×3	871/200.4	32.00/0.935 6	39.50/0.979 0
W-MSIMM	885/203.5	33.03/0.935 9	39.53/0.979 1

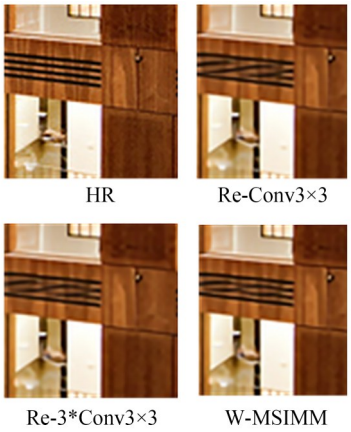

图 7　多尺度交互调制模块的消融实验可视化

Fig. 7　Visualization of ablation study on MSIMM

空间-通道交互模块的消融实验：针对空间-通道交互模块，设置了四种实验对象：移除 SCIM（WO-SCIM）、仅保留空间交互分支（W-SI）、仅保留通道交互分支（W-CI）以及完整 SCIM（W-SCIM），实验结果如表 6 和图 8 所示。

实验结果表明，SCIM 模块对模型性能具有显著贡献。当完全移除 SCIM（WO-SCIM）时，尽管模型性能和可视化效果保持在一定水平，但由于缺乏特征交互的支持，PSNR 和 SSIM 指标低于保留 SCIM 的情况。而单独使用空间交互分支（W-SI）或通道交互分支（W-CI）时，模型性能有所下降，同时重建的图像出现伪影。这些结果表明，SCIM 能够有效融合卷积和自注意力提取的特征，进而显著提升图像重建质量。

表 6　空间通道交互模块的消融实验

Tab. 6　Ablation study of SCIM

实验对象	模型复杂度	Urban100	Manga109
	Params/MACs	PSNR/SSIM	PSNR/SSIM
WO-SCIM	885/203.5	33.03/0.935 9	39.53/0.979 1
W-SI	891/205.0	32.95/0.935 3	39.51/0.979 0
W-CI	912/203.8	32.91/0.934 8	39.49/0.979 0
W-SCIM	919/205.3	33.11/0.936 7	39.58/0.979 1

图 8　空间-通道交互模块的消融实验可视化

Fig. 8　Visualization of ablation study on SCIM

特征重用 Transformer 模块的消融实验：特征重用 Transformer 模块主要聚焦于重用特征，减少信息丢失，残差连接也具有类似的功能。为了评估特征重用 Transformer 模块的作用，设计了移除 FRTM（WO-FRTM）、用多个残差连接替代 FRTM（W-Res）以及采用 FRTM（W-FRTM）三个实验对象，实验结果如图 9 和表 7 所示。

实验结果表明，FRTM 进一步提升了模型性能。当移除 FRTM（WO-FRTM）时，模型总体表现不错；将 FRTM 替换为多个残差连接（W-Res）

图 9　特征重用 Transformer 模块的消融实验可视化

Fig. 9　Visualization of ablation study on FRTM

表 7　特征重用 Transformer 模块的消融实验

Tab. 7　Ablation study of FRTM

实验对象	模型复杂度	Urban100	Manga109
	Params/MACs	PSNR/SSIM	PSNR/SSIM
WO-FRTM	919/205.3	33.11/0.936 7	39.58/0.979 1
W-Res	919/205.3	32.94/0.935 0	39.50/0.979 0
W-FRTM	997/289.4	33.16/0.936 9	39.60/0.979 1

后，尽管模型的参数量和计算复杂度与 WO-FRTM 相同，性能却有所下降，图像纹理出现扭曲；采用 FRTM（W-FRTM）的实验对象在两个数据集上的性能均达到最佳，同时重建图像的纹理更加分明。尽管 W-FRTM 的参数量和计算复杂度有所增加，但其性能提升和可视化效果证明了 FRTM 在特征重用和减少信息丢失方面的优异表现。此外，实验结果还验证了 FRTM 相比简单残差连接，能够更准确地重用重要特征，从而提升重建图像的质量。

核心模块的消融实验：为直观展示网络核心模块的相互作用，本文对三个核心模块的组合效果进行了消融实验。实验结果如表 8 和图 10 所示。

表 8　核心模块的消融实验

Tab. 8　Ablation study of Core Modules

MSIMM	SCIM	FRTM	模型复杂度	Urban100	Manga109
			Params/MACs	PSNR/SSIM	PSNR/SSIM
×	×	×	871/197.4	33.00/0.935 7	39.50/0.979 0
√	×	×	885/203.5	33.03/0.935 9	39.53/0.979 1
√	√	×	919/205.3	33.11/0.936 7	39.58/0.979 1
√	√	√	997/289.4	33.16/0.936 9	39.60/0.979 1

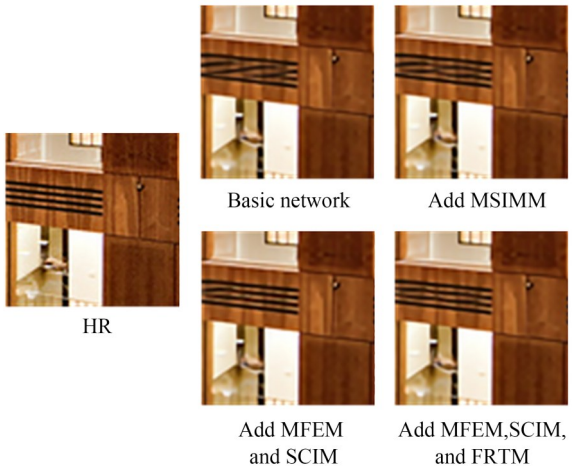

图 10　核心模块消融实验可视化

Fig. 10　Visualization of ablation study on core modules

实验结果表明，加入 MSIMM 后，模型复杂度小幅增加，精度实现了一定提升；进一步引入 SCIM，PSNR 和 SSIM 均显著提高；最终结合 FRTM，虽然带来明显复杂度提升，但是模型性能达到最佳，且整体复杂度仍处于轻量级模型范围内。整体数据结果验证了各模块在网络中的有效性。从视觉效果上来看，MSIMM 有效改善了伪影，SCIM 进一步去除了伪影，而 FRTM 增强了结构细节，使图像中部的两条线条更加清晰分明，进一步提升了重建质量。总而言之，MSIMM、SCIM 和 FRTM 三个模块相辅相成，实现了性能指标与视觉质量的双重优化。

4　结　论

本文提出了一种轻量化的多维度聚合 Transformer 网络。首先，设计了多尺度交互调制模块，有效缓解了传统 Transformer 网络因固定窗口大小难以捕获多尺度信息的问题，丰富了浅层特征的信息。其次，引入频域信息增强的空间注意力，结合对比度调整的通道注意力，构建了空间-通道交互模块，实现了自注意力与像素

注意力的互补，从而显著提升了模型性能。此外，设计的特征重用 Transformer 模块通过自注意力挖掘层级特征的依赖关系，实现了更为精确的特征重用，进一步提升了模型性能，同时为基于 Transformer 的网络设计提供了新的思路。实验结果表明，本文方法性能优于其他先进算法，在模型复杂度与性能之间达成了更优的平衡。在×4SR 任务上，相较于 CARN，在五个测试集上的 PSNR 和 SSIM 指标分别平均提升了 0.52 dB 和 0.010 8，并且重建的图像在纹理细节和边缘信息方面具有明显优势。消融实验验证了各个模块在网络结构中的作用。

作者贡献声明：

陈鹏民：提出研究方案，设计实验并主导论文撰写；负责数据分析与结果验证，多轮次修改与润色论文；

陈清江：参与论文审核与优化，协助实验分析，负责项目的组织协调与监督管理。

参考文献：

[1] DONG C, LOY C C, HE K M, *et al*. Image super-resolution using deep convolutional networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(2): 295-307.

[2] LIM B, SON S, KIM H, *et al*. Enhanced deep residual networks for single image super-resolution [C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*). *July* 21-26, 2017, *Honolulu*, *HI*, *USA*. IEEE, 2017: 1132-1140.

[3] ZHANG Y L, TIAN Y P, KONG Y, *et al*. Residual dense network for image super-resolution [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. *June* 18-23, 2018, *Salt Lake City*, *UT*, *USA*. IEEE, 2018: 2472-2481.

[4] HUI Z, WANG X M, GAO X B. Fast and accurate single image super-resolution *via* information distillation network [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. *June* 18-23, 2018, *Salt Lake City*, *UT*, *USA*. IEEE, 2018: 723-731.

[5] LIANG J Y, CAO J Z, SUN G L, *et al*. SwinIR: image restoration using swin transformer [C]. 2021 *IEEE/CVF International Conference on Computer Vision Workshops* (*ICCVW*). *October* 11-17, 2021, *Montreal*, *BC*, *Canada*. IEEE, 2021: 1833-1844.

[6] ZAMIR S W, ARORA A, KHAN S, *et al*. Restormer: efficient transformer for high-resolution image restoration [C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 18-24, 2022, *New Orleans*, *LA*, *USA*. IEEE, 2022: 5718-5729.

[7] WANG H, CHEN X H, NI B B, *et al*. Omni aggregation networks for lightweight image super-resolution [C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 17-24, 2023, *Vancouver*, *BC*, *Canada*. IEEE, 2023: 22378-22387.

[8] CHEN Z, ZHANG Y L, GU J J, *et al*. Dual aggregation transformer for image super-resolution [C]. 2023 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *October* 1-6, 2023, *Paris*, *France*. IEEE, 2023: 12278-12287.

[9] YU W H, ZHOU P, YAN S C, *et al*. InceptionNeXt: when inception meets ConvNeXt [C]. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 16-22, 2024, *Seattle*, *WA*, *USA*. IEEE, 2024: 5672-5683.

[10] DONG C, LOY C C, TANG X O. Accelerating the super-resolution convolutional neural network [C]. *Computer Vision-ECCV* 2016. *Cham*: *Springer International Publishing*, 2016: 391-407.

[11] AHN N, KANG B, SOHN K A. *Fast*, *Accurate*, *and Lightweight Super-Resolution with Cascading Residual Network* [M]. Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 256-272.

[12] CHOI H, LEE J, YANG J. N-gram in swin transformers for efficient lightweight image super-resolution [C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 17-24, 2023, *Vancouver*, *BC*, *Canada*. IEEE, 2023: 2071-2081.

[13] LI A, ZHANG L, LIU Y, *et al*. Feature modulation transformer: cross-refinement of global representation *Via* high-frequency prior for image super-resolution [C]. 2023 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *October*

1-6, 2023, *Paris*, *France*. IEEE, 2023: 12480-12490.

[14] GUO H, LI J M, DAI T, *et al*. *MambaIR: A Simple Baseline for Image Restoration with State-Space Model*[M]. Computer Vision-ECCV 2024. Cham: Springer Nature Switzerland, 2024: 222-241.

[15] LI F, CONG R M, WU J J, *et al*. SRConvNet: a transformer-style ConvNet for lightweight image super-resolution[J]. *International Journal of Computer Vision*, 2025, 133(1): 173-189.

[16] ZHANG X, ZHANG Y L, YU F. HiT-SR: *Hierarchical Transformer for Efficient Image Super-Resolution* [M]. Computer Vision-ECCV 2024. Cham: Springer Nature Switzerland, 2024: 483-500.

作者简介：

陈清江(1966－),男,河南信阳人,博士,教授,研究生导师,2006年于西安交通大学获得博士学位。主要从事小波分析、图像处理与信号处理研究。E-mail: chen66xanat@163.com

通讯作者：

陈鹏民(2000－),男,云南曲靖人,硕士研究生,2023年于西安建筑科技大学获得学士学位,主要从事小波分析与图像处理研究。E-mail: 409316492@qq.com