

# 基于交错组卷积与稀疏全局注意力的轻量级图像超分辨率重建

李想, 张娟\*

上海工程技术大学电子电气工程学院, 上海 201620

**摘要** 卷积神经网络已在图像超分辨率领域得到广泛应用, Transformer 近年来在该类图像处理任务中的扩展更是具有里程碑的意义, 然而这些大型网络具有过多的参数量和计算量, 其在部署和应用上存在很大局限性。考虑到上述发展现状, 提出一种基于交错组卷积与稀疏全局注意力的轻量级图像超分辨率重建网络, 该网络引入了以交错组卷积为主的特征提取模块, 对 Transformer 的多头自注意力机制进行优化, 设计了一种稀疏全局注意力机制以增强特征学习能力, 并提出了一种多尺度特征重构模块来提高重建效果。实验结果表明: 相比其他几种基于神经网络的方法, 所提方法的 PSNR、SSIM、参数量、计算量等性能指标都表现较好。而与基于 Transformer 的方法相比, 所提方法在 PSNR、SSIM 指标上平均提高 0.03、0.0002, 在参数量、计算量、运行时间上平均降低  $2.66 \times 10^6$ 、 $130 \times 10^9$ 、930 ms。

**关键词** 图像超分辨率; 交错组卷积; 注意力机制; 轻量化网络; Transformer; 多尺度特征重建

中图分类号 TP391.41

文献标志码 A

DOI: 10.3788/LOP231061

## Lightweight Image Super-Resolution Based on Shuffle Group Convolution and Sparse Global Attention

Li Xiang, Zhang Juan\*

School of Electrical and Electronic Engineering, Shanghai University of Technology, Shanghai 201620, China

**Abstract** Convolutional neural networks have been widely used in the field of image super-resolution, and the expansion of the transformer in such image processing tasks is a milestone in recent years. However, these large networks have excessive parameters and entail a large amount of computation, limiting their deployment and application. Given the above development status, a network based on staggered group convolution and sparse global attention lightweight image super-resolution reconstruction is proposed. A staggered group convolution feature extraction module is introduced in the network and in the transformer to improve attention mechanism optimization, and a sparse global attention mechanism is designed to enhance the feature learning ability. A multiscale feature reconstruction module is put forward to improve the reconstruction effect. The experiments show that compared with several other methods based on deep neural networks, the proposed method performs better in the peak signal to noise ratio (PSNR), structural index similarity (SSIM), parameter quantity, amount of calculation, and other performance indicators. Compared with the Transformer-based method, the proposed method has an average increase of 0.03 and 0.0002 in PSNR and SSIM, respectively, and an average decrease of  $2.66 \times 10^6$ ,  $130 \times 10^9$ , and 930 ms in parameter quantity, amount of calculation, and running time, respectively.

**Key words** image super-resolution; shuffle group convolution; attention mechanism; lightweight network; Transformer; multiscale feature reconstruction

## 1 引言

近年来基于深度学习的单幅图像超分辨率重建算法(SISR)因其端到端的训练方式和强大的特征表达能力, 在诸如图像检测<sup>[1]</sup>、光场图像<sup>[2]</sup>、医学影像<sup>[3]</sup>等领域都得到了广泛的应用。因此大多数关于 SISR 的研究都从常规方法转向了基于深度学习的方法。Dong 等<sup>[4]</sup>首次为 SISR 算法提出了一个 3 层的卷积神经网络模型。超分辨率卷积神经网络(SRCNN)使用端到端的卷积神经网络架构将低分辨率图像(LR)重建成清

晰图像。因此大多数关于 SISR 的研究都从常规方法转向了基于深度学习的方法。Dong 等<sup>[4]</sup>首次为 SISR 算法提出了一个 3 层的卷积神经网络模型。超分辨率卷积神经网络(SRCNN)使用端到端的卷积神经网络架构将低分辨率图像(LR)重建成清

收稿日期: 2023-04-10; 修回日期: 2023-05-11; 录用日期: 2023-05-29; 网络首发日期: 2023-06-10

通信作者: \*zhang-j@foxmail.com

晰的高分辨率图像(HR),该算法相比于常规算法在重建图像质量上有着显著提升。此后,许多研究都基于深度学习尝试构建更加高效的网络来提升图像重建的效果。

在降低网络参数量上,Li等<sup>[5]</sup>构建了一个U形特征提取模块,并融入深度可分离卷积进一步减少了模型在卷积运算中的参数量与计算量。Kong等<sup>[6]</sup>使用3个卷积层进行残差局部特征学习提高特征聚合效率,并提出了一种多阶段热启动训练策略。

在提高网络学习能力上,Zhang等<sup>[7]</sup>提出了用于图像超分辨率重建的残差密集网络,该网络利用级联卷积神经网络从LR图像中提取丰富的特征。Lim等<sup>[8]</sup>提出了超分辨率重建的增强深度残差网络,该网络优化了常规残差网络并用不同的缩放因子重建HR。

而近年来Transformer已是自然语言处理领域的必选模型,Vision Transformer<sup>[9]</sup>作为其改进结构应用到各类视觉任务中,随后关于Transformer在图像处理中的研究成为热点。Liu等<sup>[10]</sup>提出的Swin Transformer使用了一个层级结构来学习不同层次的特征,并使用滑动窗口操作来降低参数量。Liang等<sup>[11]</sup>提出的SwinIR网络将Swin Transformer应用到了SISR任务中,对比卷积神经网络架构的模型取得更好的性能指标。

此外,注意力机制在超分辨率重建中的应用也取得优异的效果,该机制能够对输入特征赋予不同的权重,帮助网络提取特征中局部的关键信息。Zhang等<sup>[12]</sup>提出了超深度残差注意力超分辨率重建网络,该网络结合残差结构以及通道注意力,在增强通道信息交互的同时保留更多的原始图像信息,进而提高网络的表重建征能力。Niu等<sup>[13]</sup>设计了一种层注意力机制和通道-空间注意机制以模拟不同层、通道和位置之间的整体关系,获得更好的超分效果。Dai等<sup>[14]</sup>提出了一种基于二阶特征统计的二阶注意力网络,该网络拥有高效的长距离信息捕获能力。Bin等<sup>[15]</sup>提出了高效的全局注意力机制,聚合所有重要的相关特征,保持非局部模的稀疏性。结合当前研究现状可知该方向的研究仍有很大的进步空间。

基于深度学习的图像超分辨率重建模型需要大量计算资源,这使得该类研究难以落地于一些低端设备和应用场景。此外,在模型算法的设计中不断加深神经网络的层数,可以增强网络的特征提取能力,但同时也会存在大量的冗余特征,难以重点捕获关键信息。因此需引入注意力机制、多尺度特征融合等可解释性高、复杂度低的模型来优化策略。

考虑到上述问题,提出了一种交错组卷积和稀疏全局注意力(SGA)机制的图像超分辨率重建(SSIR)网络。具体研究有以下两点:

1) 针对大型网络轻量化问题,研究设计了交错组卷积构成的轻量级注意力特征提取(LFEB)模块作

为网络主干,在降低参数量和计算量的情况下进行特征学习。

2) 针对提升网络特征学习能力的问题,研究设计了一种SGA机制提升网络的特征捕获效率,并在网络的图像重构部分设计了多尺度特征重建模块(MFRM)以多尺度卷积学习重建特征,重建图像能包含不同尺度下的特征信息。

## 2 模型结构及原理

### 2.1 总体网络结构

所提网络的结构主要由浅层特征提取、深层特征提取以及图像重构3个部分串联构成,如图1所示。令 $I_{LR}$ 图像作为输入, $I_{SR}$ 图像作为输出,如图1所示。

#### 1) 浅层特征提取

浅层特征与输入图像较为相似,包含更多原始的纹理细节,将SSIR网络的特征提取过程拆分为浅层特征提取和深层特征提取两个部分,该部分学习到的浅层特征将作为后续深层特征的输入,将其直接作用于图像重构部分来恢复图像原始的低频信息。该部分采用 $3 \times 3$ 的卷积来提取 $I_{LR}$ 的浅层特征,具体可表示为

$$\mathbf{F}_{SF} = H_{SF}(I_{LR}), \quad (1)$$

式中: $H_{SF}$ 为卷积层组成的浅层特征提取模块; $\mathbf{F}_{SF}$ 为提取出的浅层特征,该浅层特征将作为深层特征提取部分的输入。

#### 2) 深层特征提取

研究设计了一种LFEB模块作为深层特征提取的网络主干,该模块结合交错组卷积以及SGA机制,在简化计算复杂度和参数量的情况下进行特征的深度提取,整个深层特征提取的过程可表示为

$$\mathbf{F}_{DF} = H_L(\mathbf{F}_{SF}), \quad (2)$$

式中: $H_L$ 为该网络的深层特征提取部分,该部分共由 $n$ 个LFEB模块构成; $\mathbf{F}_{DF}$ 为提取到的深层特征,作为图像重建部分的输入。其内部的结构将在下一节详细阐述。

#### 3) 图像重构

在图像重构部分,为尽可能减少参数量,首先采用了简单便捷的双线性插值获取输入图像的粗略重建特征,然后设计了一种MFRM,该模块运用不同尺寸的卷积进行特征重建并将其从特征空间转换到图像空间,有效提升网络的重建效果。整个图像重构过程可表示为

$$\mathbf{I}_{SR} = H_{MR}(\mathbf{F}_{DF}) + \uparrow_b(I_{LR}), \quad (3)$$

式中: $\mathbf{I}_{SR}$ 为最终重建的高清图像; $H_{MR}$ 为MFRM; $\uparrow_b$ 为双线性插值; $I_{LR}$ 为原始低分辨率图像。最终的重建结果将由双线性插值的结果和MFRM的重建结果叠加构成。

损失函数在基于深度学习的超分辨率重建网络中的作用是通过训练迭代不断缩小重建图像与真实标签

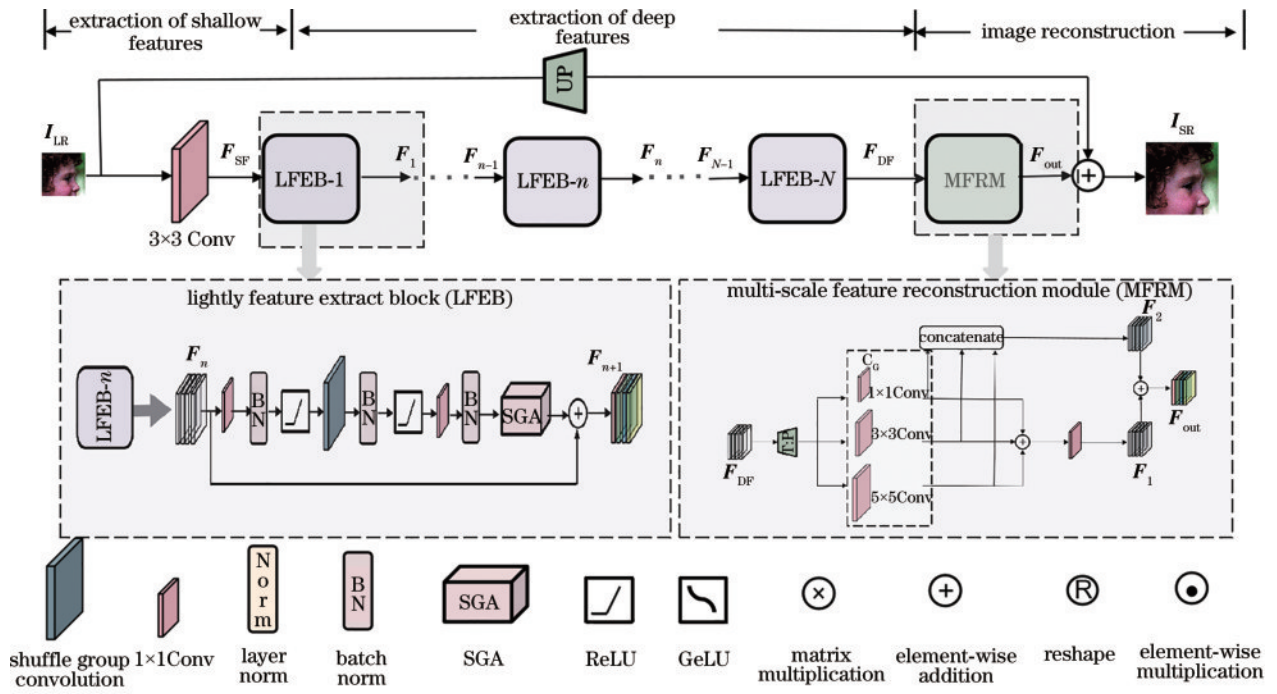


图1 总体网络结构

Fig. 1 Overall network structure

图像之间的损失值。在之前的研究中,已有多种损失函数应用到图像超分辨率,例如L1损失<sup>[16]</sup>、交叉熵损失<sup>[17]</sup>、对抗损失<sup>[18]</sup>和感知损失<sup>[19]</sup>等,所提网络选择L1损失函数来计算SSIR网络重建的图像与真实的高清图像 $I_{HR}$ 之间的差异。该过程可表示为

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \|H_{SSIR}(I_{LR}^i) - I_{HR}^i\|, \quad (4)$$

式中: $L(\theta)$ 为该网络的损失值, $I_{LR}^i$ 、 $I_{HR}^i$ 分别为训练批次中的输入低分辨率图像和对应的真实标签图像; $H_{SSIR}$ 为整个网络的特征映射过程; $N$ 为批处理大小。

## 2.2 轻量级特征提取模块

大部分基于深度学习的图像超分辨率重建算法

中,主要计算量都来源于不断堆叠的卷积层和Transformer中的多头自注意力(MSA)机制。因此,为构建参数量更少的特征提取模块,LFEB模块设计成串联的方式,在主干部分引入一种交错组卷积取代常规卷积运算。深层特征提取部分总共由 $N$ 个LFEB模块构成,第 $n+1$ 个LFEB模块的输入输出可表示为

$$F_{n+1} = L_n \{ L_{n-1} [ \dots L_1 (F_{SF}) \dots ] \}, n = 3, \dots, N, \quad (5)$$

从内部结构的角度来分析LFEB模块的运算过程,由图1可知,该模块主要由 $1 \times 1$ 卷积层、批标准化(BN)、ReLU激活函数、交错组卷积、SGA机制、残差连接等几个部分组合构成,该结构可表示为

$$F_{n+1} = F_n + \text{SGA} \left\{ \text{BN} \left\{ C_{1 \times 1} \left\{ \gamma \left\{ \text{BN} \left\{ C_{SG} \left\{ \gamma \left\{ \text{BN} \left[ C_{1 \times 1} (F_n) \right] \right\} \right\} \right\} \right\} \right\} \right\} \right\}, \quad (6)$$

式中: $F_n$ 为输入特征; $F_{n+1}$ 为输出特征; $C_{1 \times 1}$ 为 $1 \times 1$ 卷积层;BN为批标准化; $\gamma$ 为ReLU激活函数; $C_{SG}$ 为交错组卷积;SGA为SGA机制。LEFB的整个运算机制由上述几个计算单元串联构成。交错组卷积和SGA机制负责对 $F_n$ 进行深度特征提取, $1 \times 1$ 卷积负责调整输入输出特征的通道数,便于特征在网络中能够正确地前向传播,加入激活函数能够在抑制噪声的同时,增强网络的表征能力,BN则用于加速网络的训练过程,而使用残差连接让原始输入特征与深度提取后的特征线性叠加,能够缓解梯度消失、过拟合等现象。

## 2.3 交错组卷积及参数量分析

常规卷积操作中,每个卷积核都需要学习一组权

重,这可能会导致卷积核的数量非常大,从而增加计算量和存储成本。而组卷积则是将卷积核分成若干组,每组卷积核只学习一部分权重,以此减少参数量和计算量。

研究所引入的交错组卷积由两次组卷积和通道交错(洗牌)两部分构成。其中,组卷积将输入特征在通道维度上拆分为不同的组,再分别对每一组输入进行卷积运算,这使得同组特征在通道维度上将存在大量相同的信息,且各组特征之间无法进行信息交互,所以组卷积虽然能够有效减少计算量和参数量,但在分组运算过程中的通道阻断现象使得学习到的特征十分有限,这直接影响模型性能。为在组卷积中增强特征之

间的通道相关性,引入通道洗牌机制,该过程能均匀打乱不同特征之间的通道并进行重组,使重组之后的每组特征都能包含其他组特征的通道信息。然后,引入  $1 \times 1$  的卷积核将通道洗牌的特征进行第二次分组卷积,以此恢复被打乱的通道,最后将每组特征进行拼接得到输出特征。交错组卷积的流程和数理运算原理具体可表示为

$$C_{SG}(\mathbf{f}) = \odot \left\{ C_2 \left\{ C_{sf} \left[ C_1(\mathbf{f}) \right] \right\} \right\} = \mathbf{U} \mathbf{W}^d \mathbf{U}^T \mathbf{W}^p \mathbf{f}, \quad (7)$$

式中:  $\mathbf{f}$  为输入特征;  $C_1$ 、 $C_2$  分别为第一次和第二次分组卷积;  $C_{sf}$  为通道洗牌;  $\odot$  为通道拼接, 整个交错组卷积由这几个单元串联构成; 从数理运算的角度分析,  $\mathbf{U}$ 、 $\mathbf{U}^T$  分别为排序矩阵及其转置矩阵, 负责将特征进行重排序;  $\mathbf{W}^p$ 、 $\mathbf{W}^d$  为一组对角矩阵, 该组对角矩阵分别由第一次和第二次分组卷积的卷积核构成。

交错组卷积和常规卷积的运算过程示意如图 2 所

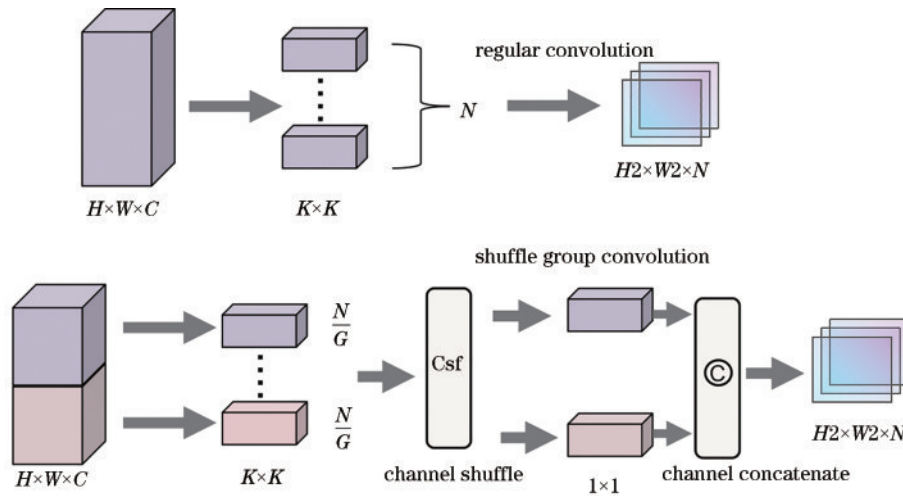


图 2 交错组卷积与常规卷积的运算过程对比示意图

Fig. 2 Schematic diagram of the comparison between the operation process of shuffle group convolution and conventional convolution

## 2.4 SGA 机制

基于 Transformer 的方法凭借 MSA 机制能够在捕获全局特征上有良好的效果。诸如文献 [11] 和文献 [20] 中的 MSA 机制都使用 Softmax 函数进行归一化, 保留所有的查询值 (Query) 和索引值 (Key) 之间的相关性, 然而并非所有的查询值和索引值都有实际相关性。而使用 Softmax 函数对特征进行归一化后将影响特征聚合, 增大计算量的同时, 捕获许多冗余的特征。因此, 在 MSA 机制的基础上进行改进, 提出了一种 SGA 机制, 该机制将原先的 depth-wise 卷积替换为交错组卷积, 从而在不增加参数量的情况下加强通道间的信息交互。此外, 由于 ReLU 函数能够保留正特征而去除负特征, 因此用 ReLU 函数替换原先的 Softmax 函数以此提升网络的特征捕获效率。SGA 机制结构如图 3 所示, 其中左半部分为注意力特征聚合部分。记  $\mathbf{F}_m$  为输入特征图, 经过  $1 \times 1$  卷积以及交错组卷积后, 分别获得查询特征 ( $\mathbf{Q} \in \mathbb{R}^{H \times W \times C}$ )、索引特征 ( $\mathbf{K} \in \mathbb{R}^{H \times W \times C}$ )、内

容特征 ( $\mathbf{V} \in \mathbb{R}^{H \times W \times C}$ ), 经过维度变换后分别得到  $\mathbf{Q} \in \mathbb{R}^{H \times W \times C}$ 、 $\mathbf{K}' \in \mathbb{R}^{H \times W \times C}$ 、 $\mathbf{V}' \in \mathbb{R}^{H \times W \times C}$ 。首先获得估计特征  $\mathbf{F}_A$ , 再通过  $\mathbf{F}_A$  计算注意力聚合特征  $\mathbf{F}'_{SGA}$ , 整个过程可表示为

$$\mathbf{F}_A = \gamma \left( \frac{\mathbf{Q}'^T \mathbf{K}'}{\alpha} \right), \quad (11)$$

$$\mathbf{F}'_{SGA} = C_{1 \times 1} \left[ \mathbf{R}(\mathbf{F}_A \times \mathbf{V}') \right] + \mathbf{F}_m, \quad (12)$$

式中:  $\alpha$  为可学习的权重参数;  $\mathbf{R}$  为维度调整。  $\mathbf{F}_A$  与  $\mathbf{V}'$  相乘并经过维度调整与  $1 \times 1$  卷积后, 与  $\mathbf{F}_m$  残差相连获得  $\mathbf{F}'_{SGA}$ 。SGA 机制右半部分为文献 [20] 中的门控前向传播结构, 同样将原先的 depth-wise 卷积替换为交错组卷积, 公式化过程可表示为

$$G(\mathbf{F}'_{SGA}) = \lambda \left\{ C_{SG} \left[ C_{1 \times 1}(\mathbf{F}'_{SGA}) \right] \right\} \cdot \left\{ C_{SG} \left[ C_{1 \times 1}(\mathbf{F}'_{SGA}) \right] \right\}, \quad (13)$$

$$\mathbf{F}_{SGA} = \mathbf{F}'_{SGA} + \mathbf{N} \left[ G(\mathbf{F}'_{SGA}) \right], \quad (14)$$

式中:  $G$  为控制通道信息流的门控机制 (Gating

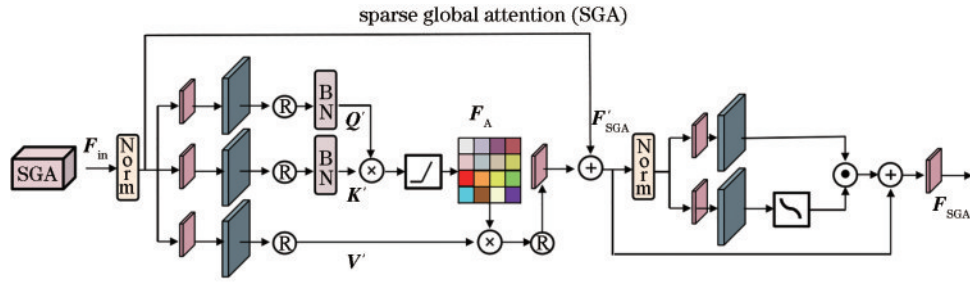


图 3 SGA 机制结构图

Fig. 3 Structure diagram of SGA mechanism

mechanism), 该结构使用平行分支, 其中一条分支加入 GeLU 函数  $\lambda$ ;  $N$  为层归一化。两条分支的结果逐元素相乘后使得每层特征都能与其他层特征在细节上互补, 丰富上下文信息。最后, 与  $F'_{SGA}$  相加输出最终的注意力特征  $F_{SGA}$ 。

## 2.5 MFRM

先前研究表明融合不同尺度的特征信息有助于网络重建出细节更加丰富的重建图像, 使得超分辨率网络的性能得以提高。例如, 较小尺度的特征可以包含图像的细节纹理信息, 而较大尺度的特征可以包含更多的全局结构信息, 同时融合后的特征可以综合考虑不同尺度条件下的信息, 从而提高网络模型的鲁棒性。因此研究设计了一种 MFRM, 便于从低分辨率的特征图中重建出更清晰的图像。MFRM 首先采用亚像素卷积<sup>[21]</sup>对深层特征  $F_{DF}$  进行上采样, 在扩大其分辨率的同时对通道信息进行重组, 然后采用多尺度的卷积层进行特征融合。这些卷积层主要采用 3 种不同大小的卷积核进行卷积运算, 以此来聚合不同尺度的空间特征。

具体而言, 将这些不同尺度的卷积称为卷积群 ( $C_G$ ), 该卷积群由  $5 \times 5$  卷积、 $3 \times 3$  卷积和  $1 \times 1$  卷积组成, 其中卷积核较小的卷积层和卷积核较大的卷积层分别用于重点捕获短距离局部信息和长距离全局信息。在特征融合部分, 同时引入了拼接操作和逐元素求和操作来对特征图进行融合, 以此让不同尺度的特征信息进行互补, 从而输出更加完善的重建结果。整个 MFRM 的过程可表示为

$$C_G = [C_{1 \times 1}(\cdot), C_{3 \times 3}(\cdot), C_{5 \times 5}(\cdot)], \quad (15)$$

$$F_1 = C_{1 \times 1} [C_{1 \times 1}(\cdot) + C_{3 \times 3}(\cdot) + C_{5 \times 5}(\cdot)] [\uparrow_p(F_{DF})]. \quad (16)$$

MFRM 从整体上看是将两个分支的输出结果进行融合, 结构示意图如图 1 所示, 其中经过深层特征提取阶段所得到的特征  $F_{DF}$  将作为此模块的输入, 主分支经过不同尺度的卷积运算, 随后通过  $1 \times 1$  的卷积调整从而输出一个分支的重建结果  $F_1$ 。

下式为另一分支的运算过程, 相较于第一个分支, 该分支将经过 3 种不同尺度的卷积结果进行拼接得到  $F_2$ 。

$$F_2 = \odot \{C_G [\uparrow_p(F_{DF})]\}, \quad (17)$$

$$F_{Out} = F_1 + F_2. \quad (18)$$

$F_1$  与  $F_2$  经过线性叠加后获得重建结果  $F_{Out}$ , 该重建结果再与经过残差连接与插值采样的原始特征叠加, 得到最终的重建图像  $I_{SR}$ 。

## 3 实验验证分析

### 3.1 实验数据与评价标准

在训练阶段研究从 DIV2K<sup>[22]</sup> 中选择 800 张图像作为模型训练的训练集。在测试阶段研究选择了 4 个基准数据集作为测试集: Set5<sup>[23]</sup>、Set14<sup>[24]</sup>、Urban100<sup>[25]</sup> 和 BSD100<sup>[26]</sup>。将重建结果从 RGB 空间转换为 YCbCr 空间, 用结构相似性指数 (SSIM) 和峰值信噪比 (PSNR) 评估实验结果。

### 3.2 实验设置

在训练过程中, 对 800 张训练图像进行数据增强, 这些图像水平翻转并随机旋转  $90^\circ$ 、 $180^\circ$ 、 $270^\circ$ 。在每个训练批次中, 将训练批次大小设置为 32, 使用 ADAM 优化器优化训练过程。此外, 初始学习率为  $5 \times 10^{-4}$ ,  $\beta_1$ 、 $\epsilon$  分别设置为 0.9 和 0.999。整个网络总共迭代 1200 次。所有实验均在 Ubuntu 16.04、CUDA Toolkit 10.1、PyTorch 0.4.1、Python 3.7 和 Titan Xp GPU 等环境中实现。

### 3.3 客观指标评估

为评估该网络性能指标, 将所提的 SSIR 网络与近几年的其他方法 (RDN<sup>[7]</sup>、EDSR<sup>[8]</sup>、RCAN<sup>[12]</sup>、HAN<sup>[13]</sup>、SAN<sup>[14]</sup>、SwinIR<sup>[11]</sup> 和 ENLCN<sup>[15]</sup>) 进行比较。比例因子为 2、3、4 时的实验结果如表 1~表 3 所示。

在实验中引入参数量 (Parameter quantity)、计算量 (Flops)、时延 (Latency) 等 3 项指标, 并将各方法得出的结果进行对比, 如表 4 所示。其中, 参数量通常用 Float32 表示, 计算量表示每秒浮点运算次数, Latency 指计算机系统中的延迟时间, 即从发出指令到执行完成所需的时间。在各比例因子下, 使用  $1280 \times 720$  的图像作为输入对比各方法的 3 项指标。

由表 1~表 3 可知, 在 PSNR、SSIM 这两个主要的客观性能评价指标上, SSIR 网络在不同比例因子、不同测试集上都优于其他方法, 与基于 Transformer 的方

表 1 比例因子为 2 时不同测试集上的对比试验

Table 1 Comparative tests on different test sets when the scale factor is 2

Method	Set5		Set14		Urban100		B100	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
EDSR	38.11	0.9602	33.92	0.9195	32.93	0.9351	32.32	0.9013
RDN	38.24	0.9614	34.01	0.9212	32.89	0.9353	32.34	0.9017
RCAN	38.27	0.9614	34.12	0.9216	33.34	0.9384	32.41	0.9027
SAN	38.31	0.9620	34.07	0.9213	33.10	0.9370	32.42	0.9028
HAN	38.27	0.9614	34.16	0.9217	33.35	0.9385	32.41	0.9027
SwinIR	38.35	0.9620	34.14	0.9227	33.40	0.9393	32.44	0.9030
ENLCN	38.37	0.9618	34.17	0.9229	33.56	0.9398	32.49	0.9032
SSIR	38.37	0.9620	34.16	0.9227	33.38	0.9387	32.58	0.9039

表 2 比例因子为 3 时不同测试集上的对比试验

Table 2 Comparative tests on different test sets when the scale factor is 3

Method	Set5		Set14		Urban100		B100	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
EDSR	34.65	0.9280	30.52	0.8462	28.80	0.8653	29.25	0.8093
RDN	34.71	0.9296	30.57	0.8468	28.80	0.8653	29.26	0.8093
RCAN	34.74	0.9299	30.65	0.8482	28.09	0.8702	29.32	0.8111
SAN	34.75	0.9300	30.59	0.8476	28.93	0.8671	29.33	0.8112
HAN	34.75	0.9299	30.67	0.8483	29.10	0.8705	29.32	0.8110
SwinIR	34.89	0.9312	30.77	0.8503	29.29	0.8744	29.37	0.8124
ENLCN								
SSIR	34.87	0.9312	30.83	0.8505	29.33	0.8757	29.37	0.8125

表 3 比例因子为 4 时不同测试集上的对比试验

Table 3 Comparative tests on different test sets when the scale factor is 4

Method	Set5		Set14		Urban100		B100	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
EDSR	32.46	0.8968	28.80	0.7876	26.64	0.8033	27.71	0.7420
RDN	32.47	0.8990	28.81	0.7871	26.61	0.8028	27.72	0.7419
RCAN	32.63	0.9002	28.87	0.7889	26.82	0.8087	27.77	0.7436
SAN	32.64	0.9003	28.92	0.7888	26.79	0.8068	27.78	0.7436
HAN	32.64	0.9002	28.90	0.7890	26.85	0.8094	27.80	0.7442
SwinIR	32.72	0.9021	28.94	0.7914	27.07	0.8164	27.83	0.7459
ENLCN	32.67	0.9004	28.94	0.7892	27.12	0.8184	27.82	0.7452
SSIR	32.74	0.9021	28.94	0.7911	27.12	0.8169	27.91	0.7465

表 4 各方法参数量、计算量、运行时间对比

Table 4 Comparison of reference quantity, calculation quantity and running time of each method

Method	Scale factor is 2			Scale factor is 3			Scale factor is 4		
	Parameter quantity / $\times 10^6$	Flops / $\times 10^9$	Latency / ms	Parameter quantity / $\times 10^6$	Flops / $\times 10^9$	Latency / ms	Parameter quantity / $\times 10^6$	Flops / $\times 10^9$	Latency / ms
EDSR	40.73	9387	1143	43.86	4470	573	43.09	2895	360
RDN	22.12	5098	846	22.30	2282	406	22.27	1310	243
RCAN	15.44	3530	3530	15.62	1586	367	15.59	918	223
SAN	15.86	3050		15.89	1620		15.86	937	
HAN	63.60	14551	2278	64.34	6534	1014	64.19	3776	628
SwinIR	11.75	2301	2913	11.93	1026	1238	11.90	584	645
SSIR	9.24	1987	1245	9.18	1004	427	9.18	532	332

法 (SwinIR) 以及 ENLCN 等方法相比, PSNR、SSIM 平均提升 0.03、0.0002 和 0.008、0.0004。

在参数量、计算量与运行时间方面, 由表 4 可知, SSIR 网络的参数量和计算量要优于其他基于深度神经网络的方法。而与 SwinIR 进行对比, 二者在各项指标上比较接近, SSIR 网络在参数量和计算量上略微小于 SwinIR, 平均降低  $2.66 \times 10^6$ 、 $130 \times 10^9$ , 平均运行时间上提高了 930 ms。总体来看, 在运行响应时间上 SSIR 网络虽未能在各方法中达到最快, 但综合参数量和性能指标来看, SSIR 网络在整体上拥有更高的性价比。

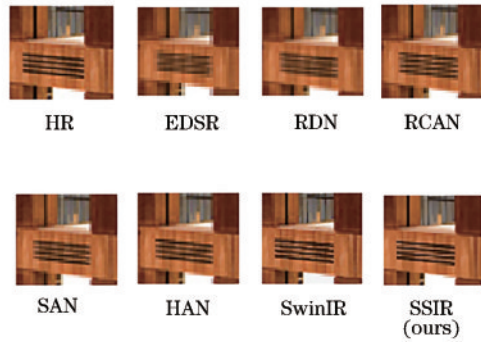


图 4 img-49 重建结果示例

Fig. 4 Example of reconstruction results to img-49

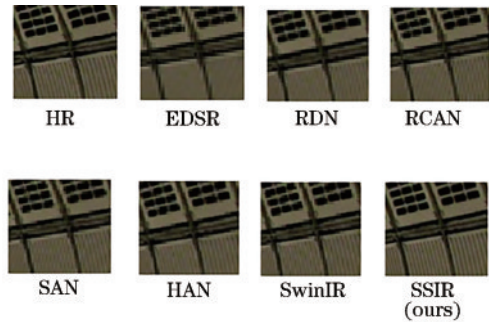
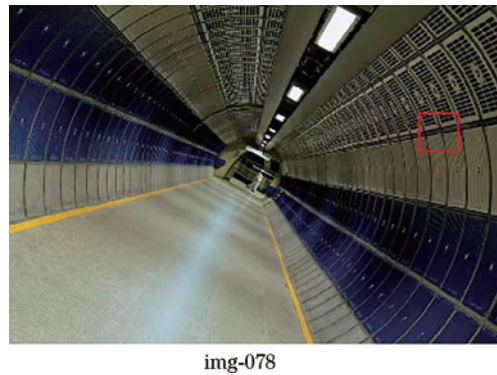


图 5 img-78 重建结果示例

Fig. 5 Example of reconstruction results to img-78

从 RealSR<sup>[27]</sup> 数据集中进行真实数据集的重建, 结果如图 6 所示。可观察到 SSIR 网络在纹理细节的重建上优于其他方法。这验证了该网络在参数量较少的前提下, 仍能取得良好的重建效果, 并在真实场景中拥有良好的鲁棒性。

### 3.5 消融实验

通过消融实验验证 LFEB 模块中交错组卷积、SGA 机制及 MFRM 的有效性。所有消融实验都在比例因子为 2 的 Set5 数据集上进行。

为验证 SSIR 网络中 LFEB 模块所使用的交错组卷积具有合理性, 研究将其与常规卷积进行消融实验。具体方法为将原始的 SSIR 网络作为对照组, 将 SSIR

### 3.4 主观视觉评估

合成数据集上的测试结果, 如图 4、图 5 所示。从 Urban100 数据集中选择了 image-049、image-078 等图片进行主观视觉评估实验, 并将 SSIR 网络与 6 种方法在重建视觉效果上进行比较。可直观看出, 通过 EDSR、RDN 重建的图像不能产生完整的线条, 存在部分伪影。RCAN、SAN、HAN 相较于前二者能取得更好的重建效果, 但在恢复主体纹理和轮廓方面仍然有着较为明显的缺陷, SwinIR 相比于其他方法能够恢复更多细节, 但仍包含少量的噪声, 而所提的 SSIR 网络可以获得细节更清晰的结果。

网络中 LFEB 模块的交错组卷积替换为常规的  $3 \times 3$  卷积作为实验组, 保持数据集和实验环境相同, 将训练出的模型在同一测试集上进行客观评价指标的评估, 实验结果如表 5 所示。由表 5 可知, 将交错组卷积替换为常规卷积之后, 网络计算量提升近 1.6 倍, 但 PSNR、SSIM 两个指标仅略微提升 0.04、0.0001, 因此在特征提取中采用交错组卷积替代常规卷积, 保证模型具有较好性能, 也显著降低模型的参数量。

SGA 机制作为 Transformer 中 MSA 机制的改进, 对该模块进行消融实验。将融入 SGA 机制的原始 LFEB 模块作为实验组、将 SGA 机制移除后, 用 MSA 机制加入 LFEB 模块作为对照组, 此外再设置一组空

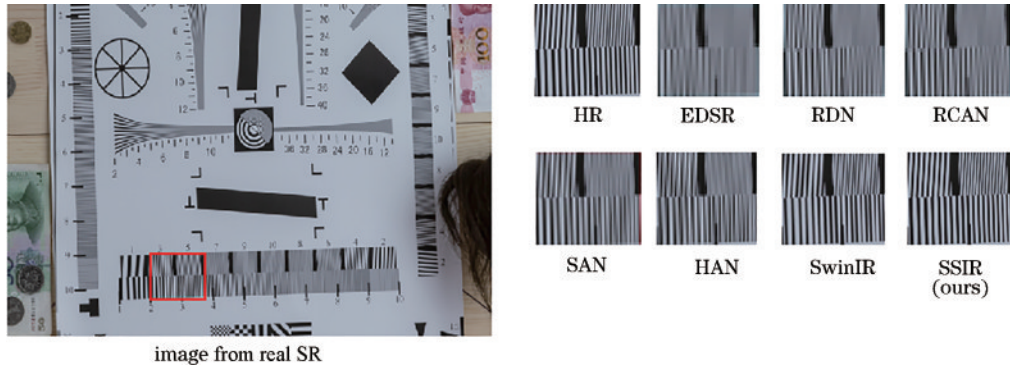


图 6 真实数据集重建结果示例

Fig. 6 Example of reconstruction results to real dataset

表 5 交错组卷积的消融实验

Table 5 Ablation experiments for the shuffle group convolution

Method	Parameter quantity / $\times 10^6$	Flops / $\times 10^9$	PSNR	SSIM
Shuffle group convolution	9.24	1897	38.37	0.9620
$3 \times 3$ convolution	14.86	3015	38.41	0.9621

白对照组,即 LFEB 模块中不加入任何注意力机制,在同一测试集上分析其量化结果。消融实验结果如表 6 所示。由表 6 可知,在参数量和计算量上,SGA 机制少于加入 MSA 机制的 LFEB 模块,分别降低  $3.44 \times 10^6$ 、 $361 \times 10^9$ ,而大于空白对照组,分别增加  $2.52 \times 10^6$ 、 $525 \times 10^9$ 。在 PSNR、SSIM 两个性能指标方面,融入 SGA 机制的 LFEB 模块要全面优于两个对照组 (LEFB、LFEB+MSA),分别提高 0.27、0.0019 和 0.08、0.0006。因此,实验数据证明,SGA 机制在参数量和计算量小幅提高的情况下,能有效提升网络性能。

表 6 SGA 机制的消融实验

Table 6 Ablation experiments for the SGA mechanism

Method	PSNR	SSIM	Parameter quantity / $\times 10^6$	Flops / $\times 10^9$
LFEB+SGA	38.37	0.9620	9.24	1897
LFEB+MSA	38.29	0.9614	12.68	2258
LEFB	38.10	0.9601	6.72	1372

MFRM 的消融实验结果如表 7 所示。为验证 MFRM 在图像重建阶段所起到的作用,研究在该阶段引入双线性插值和反卷积这两种上采样的方式来替换 MFRM,在网络中进行图像重建,并在统一数据集上进行性能测试,同样以 PSNR、SSIM 作为客观评价指标进行性能对比。由表 7 可知,MFRM 加入网络后,

表 7 MFRM 的消融实验

Table 7 Ablation experiment on the number of MFRM

Module	PSNR	SSIM
Interpolation	38.05	0.9560
Dconv	38.18	0.9613
MFRM	38.37	0.9620

显著提升网络模型性能,与双线性插值、反卷积两种上采样方式相比,PSNR、SSIM 分别提高 0.22、0.0060 和 0.19、0.0007,该结果验证了 MFRM 的加入能够融合多尺度特征信息,帮助网络取得更好的重建效果。

## 4 结 论

针对当前超分辨率网络参数量和计算量过多、深层网络特征提取能力不高等问题,提出了一种基于交错组卷积和 SGA 机制的 SSIR 网络,该网络在深层特征提取部分采用交错组卷积,融入 SGA 机制以增强维度信息间的交互,并在图像重建部分设计了一种 MFRM 用于融合不同尺度的重建特征。通过对比实验以及消融实验证明,所提的 SSIR 网络显著减少参数量和计算量的同时仍能保持较好的性能。

基于深度学习的图像超分辨率算法以数据作为基础,因此在未来的工作中,研究将尝试引入更多的评价指标和场景数据集来更全面地评价超分辨率效果,以此来提高模型的泛化能力以及实际应用价值。

## 参 考 文 献

[1] 葛鑫鑫,崔海华,徐振龙,等. 金属发动机叶片微小缺陷超分辨图像重建方法[J]. 光学学报, 2023, 43(2): 0210001.  
Ge X X, Cui H H, Xu Z L, et al. Super-resolution image reconstruction method for micro defects of metal engine blades[J]. Acta Optica Sinica, 2023, 43(2): 0210001.

[2] 徐欣宜,邓慧萍,向森,等. 基于特征交互融合与注意力的光场图像超分辨率[J]. 激光与光电子学进展, 2023, 60(14): 1410017.  
Xu X Y, Deng H P, Xiang S, et al. Light field image super-resolution based on feature interaction fusion and attention mechanism[J]. Laser & Optoelectronics Progress, 2023, 60(14): 1410017.



- [3] 潘梦雪, 曲宁, 夏业儒, 等. 基于深度学习的核磁图像超分辨重建[J]. 激光与光电子学进展, 2022, 59(22): 2217001.  
Pan M X, Qu N, Xia Y R, et al. Super-resolution reconstruction of nuclear magnetic resonance images based on deep learning[J]. Laser & Optoelectronics Progress, 2022, 59(22): 2217001.
- [4] Dong C, Loy C C, He K M, et al. Learning a deep convolutional network for image super-resolution[M]// Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8692: 184-199.
- [5] Li B, Wang B, Liu J B, et al. S-LWSR: super lightweight super-resolution network[J]. IEEE Transactions on Image Processing, 2020, 29: 8368-8380.
- [6] Kong F Y, Li M X, Liu S W, et al. Residual local feature network for efficient super-resolution[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 19-20, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 765-775.
- [7] 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. New York: IEEE Press, 2018: 2472-2481.
- [8] 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. New York: IEEE Press, 2017: 1132-1140.
- [9] Dong X Y, Bao J M, Chen D D, et al. CSwin transformer: a general vision transformer backbone with cross-shaped windows[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 12114-12124.
- [10] Liu Z, Lin Y T, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2022: 9992-10002.
- [11] 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. New York: IEEE Press, 2021: 1833-1844.
- [12] Zhang Y L, Li K P, Li K, et al. Image super-resolution using very deep residual channel attention networks[M]// Ferrari V, Hebert M, Sminchisescu C, et al. Computer Vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 294-310.
- [13] Niu B, Wen W L, Ren W Q, et al. Single image super-resolution via a holistic attention network[M]//Bartoli A, Fusiello A. Computer vision-ECCV 2020 workshops. Lecture notes in computer science. Cham: Springer, 2020: 191-207.
- [14] Dai T, Cai J R, Zhang Y B, et al. Second-order attention network for single image super-resolution[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 11057-11066.
- [15] Xia B, Hang Y C, Tian Y P, et al. Efficient non-local contrastive attention for image super-resolution[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(3): 2759-2767.
- [16] Peng X J, Xu D, Kong L Y, et al.  $L_1$ -norm loss based twin support vector machine for data recognition[J]. Information Sciences, 2016, 340/341: 86-103.
- [17] Li X L, Zhang X T, Huang W, et al. Truncation cross entropy loss for remote sensing image captioning[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(6): 5246-5257.
- [18] 刘羿漩, 葛广英, 齐振岭, 等. 基于改进 DCGAN 的刺绣图像修复的研究[J]. 激光与光电子学进展, 2023, 60(20): 2010005.  
Liu Y X, Ge G Y, Qi Z L, et al. Research on embroidery image restoration based on improved DCGAN[J]. Laser & Optoelectronics Progress, 2023, 60(20): 2010005.
- [19] Rad M S, Bozorgtabar B, Marti U V, et al. SROBB: targeted perceptual loss for single image super-resolution [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2020: 2710-2719.
- [20] 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. New York: IEEE Press, 2022: 5718-5729.
- [21] Ruan H H, Tan Z Y, Chen L T, et al. Efficient sub-pixel convolutional neural network for terahertz image super-resolution[J]. Optics Letters, 2022, 47(12): 3115-3118.
- [22] Agustsson E, Timofte R. NTIRE 2017 challenge on single image super-resolution: dataset and study[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1122-1131.
- [23] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding[C]//Proceedings of the British Machine Vision Conference 2012, September 3-7, 2012, Surrey, UK. London: British Machine Vision Association, 2012: 135.1-135.10.
- [24] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations[M]//Boissonnat J D, Chenin P, Cohen A, et al. Curves and surfaces. Lecture notes in computer science. Heidelberg: Springer, 2012, 6920:

- 711-730.
- [25] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 5197-5206.
- [26] Arbeláez P, Maire M, Fowlkes C, et al. Contour detection and hierarchical image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5): 898-916.
- [27] Cai J R, Zeng H, Yong H W, et al. Toward real-world single image super-resolution: a new benchmark and a new model[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2020: 3086-3095.