

基于多尺度局部与全局特征融合的轻量级超分辨率网络

冯俊杰, 洪智勇*, 熊利平*, 劳雪颖

五邑大学电子与信息工程学院, 广东 江门

收稿日期: 2026年3月9日; 录用日期: 2026年3月29日; 发布日期: 2026年4月9日

摘要

当前图像超分辨率(SR)的研究重点方向是开发一种平衡图像重建性能和低资源消耗量的轻量级模型。Transformer擅长全局特征建模, 但资源消耗量和参数量极高, 同时欠缺探索精细局部细节的能力。而卷积神经网络(CNN)资源消耗低, 但缺乏捕获全局特征的能力。为了解决这些问题, 本文提出了一种基于纯CNN的模型, 称为多尺度局部与全局特征融合网络(MLGFFN)。MLGFFN采用多尺度大核卷积块和局部特征提取块双分支分别捕获全部和局部特征信息, 并且改进了特征融合方式。此外, MLGFFN还引入了特征增强前馈网络进一步细化前一级的特征图, 对其中的重要特征信息进行强调, 同时对冗余特征进行抑制。大量实验表明, 本文提出的MLGFFN方法优于现有的轻量级SR模型, 在轻量化设计和图像重建性能之间取得了良好的平衡。特别是在 $\times 4$ 放大因子的Set14数据集上, MLGFFN的PSNR指标比SMSR模型高出0.21 dB, 而参数量和FLOPs仅为SMSR的37.38%和40%。

关键词

轻量级超分辨率, 大核卷积, 特征融合, 多尺度

Lightweight Super-Resolution Network Based on Multi-Scale Local and Global Feature Fusion

Junjie Feng, Zhiyong Hong*, Liping Xiong*, Xueying Lao

School of Electronics and Information Engineering, Wuyi University, Jiangmen Guangdong

Received: March 9, 2026; accepted: March 29, 2026; published: April 9, 2026

*通讯作者。

文章引用: 冯俊杰, 洪智勇, 熊利平, 劳雪颖. 基于多尺度局部与全局特征融合的轻量级超分辨率网络[J]. 图像与信号处理, 2026, 15(2): 235-247. DOI: 10.12677/jisp.2026.152020

Abstract

Current research in Image Super-Resolution (SR) focuses on developing lightweight models that achieve a balance between reconstruction performance and low resource consumption. While Transformer excels at global feature modeling, it suffers from excessive computational overhead and parameter counts, while lacking the ability to explore fine-grained local details. Conversely, Convolutional Neural Networks (CNNs) offer lower resource consumption but are deficient in capturing global features. To address these issues, this paper proposes a pure CNN-based model termed the Multi-scale Local and Global Feature Fusion Network (MLGFFN). MLGFFN utilizes a dual-branch architecture, comprising Multi-scale Large Kernel Convolution blocks and Local Feature Extraction blocks, to capture global and local feature information, respectively, while incorporating an improved feature fusion mechanism. Furthermore, a Feature Enhancement Feed-forward Network (FEFN) is introduced to further refine the feature maps from previous stages, emphasizing critical information while suppressing noise. Extensive experiments demonstrate that the proposed MLGFFN outperforms existing lightweight SR models, achieving a superior balance between lightweight design and image reconstruction performance. Notably, on the Set14 dataset with a $\times 4$ upscaling factor, MLGFFN achieves a PSNR improvement of 0.21 dB over SMSR model, while utilizing only 37.38% of the parameters and 40% of the FLOPs required by SMSR.

Keywords

Lightweight Super-Resolution, Large Kernel Convolution, Feature Fusion, Multi-Scale

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

图像超分辨率(Super-Resolution, SR)是指将输入的低分辨率(Low-Resolution, LR)图像重建为高分辨率(High-Resolution, HR)图像。自从 SRCNN [1]首次将深度学习引入超分辨率以来, SR 图像重建研究取得了突破性进展。随着强调全局特征的 Transformer [2]模型的出现, 基于 Transformer 的 SR 模型的图像重建性能实现了显著的飞跃[3] [4]。然而, 这种性能的提升往往带来极高的计算资源消耗量, 并且 Transformer 中的 Self-Attention 机制在探索精细局部细节方面的能力有所欠缺。

针对这些限制, SMFANet [5]提出了一种自调制特征聚合模块用于协同利用局部和全局特征双分支的交互, 局部细节分支引入了 3×3 的深度可分离卷积弥补捕捉高频局部信息方面的不足, 实现了更准确的图像重建。但是图像局部细节中的边缘细节和轮廓都是具有方向性的, 单一卷积核难以表示这些特征信息。AIDN [6]使用非对称的卷积, 从水平和垂直方向进行特征提取来识别图像不同区域的特征, 同时保持了较低的数量和计算量。此外, PLKSR [7]采用 17×17 大核卷积有效地捕获了全局特征信息, 但同时导致模型参数量和计算资源消耗量急剧增加。MAN [8]提出一种 MLKA 模块, 使用多尺度的大核注意力机制, 在不同的粒度级别上获得丰富的注意力图。然而, 并行的大核卷积操作只能粗略地提取多尺度特征, 无法充分利用不同尺度下的上下文信息。在视觉任务中, 图像的信息特征的尺寸差异很大, 上述的方法都只考虑了单一感受野下的特征表示, 无法区分和表达不同尺寸区域的特征信息。

特征融合可以同时利用全局信息和局部信息理解特征图全局上下文结构, 既能保留精细的边缘细节

信息,又能保持整体结构的稳定性。最近研究中利用双分支探索局部和全局特征的方法[5][9]-[12]均使用了简单的逐元素相加的融合方式,没有引入额外的计算资源消耗量,保持了模型轻量高效的特点。OSFFNet[13]从网络的多个中间层收集特征,将来自不同阶段的特征图进行逐元素相加并进行整合,综合利用局部细节和图像整体结构。然而,来自全局特征和局部特征信息的感受野显著不同,基础的加法融合操作无法应对这种内在的失配问题。同时又忽略了特征通道之间的依赖关系,缺乏通道间信息交互,无法有效地区分不同区块信息。

为了同时获得全局特征和局部特征的捕获能力,解决特征融合失配问题,以及保持 CNN 模型的低资源消耗优势,本文提出了一种新颖的基于纯 CNN 的轻量级 SR 网络,称为多尺度局部与全局特征融合网络(Multi-scale Local-Global Feature Fusion Network, MLGFFN),主要框架见图 1。其包含两个核心模块,称为多尺度异构特征融合块(Multi-scale Asymmetric Feature Fusion Block, MAFFB)和特征增强前馈网络(Feature Enhancement Feed-forward Network, FEFN)。MAFFB 模块采用局部特征分支 LFEB 和全局特征分支 MLKB 进行特征提取,然后通过 Fusion 融合模块赋予模型通道信息交互能力。尽管本文注意到 Res2Net[14]提出了一种级联的多尺度残差连接模块,而非像传统网络那样依赖于堆叠多个卷积层,从而在更细的粒度上表示多尺度特征。但与 Res2Net 不同,MLKB 使用了级联残差连接的多个不同尺寸的大核非对称深度可分离卷积替代掉 Res2Net 中 3×3 的卷积层,有效地扩大了感受野,同时对不同尺寸的信息区域特征分别进行表达。LFEB 采用与 MLKB 异构的非对称卷积构造方式,从水平和垂直上独立地提取具有方向性的局部细节纹理。FEFN 进一步细化了前一级模块提取出的代表性特征。基于上述设计,与其他轻量级 SR 方法相比,所提出的 MLGFFN 实现了卓越的图像重建性能,同时保持了相对较低的模型复杂度。大量实验表明,本文提出的 MLGFFN 在计算效率和重建性能之间实现了有利的平衡。

本文的主要贡献总结如下:

- 1) 开发了多尺度局部与全局特征融合网络,MLGFFN 高效地提取具有丰富信息的多尺度特征,其中 MLKB 分支用于探索全局信息,LFEB 分支用于捕获局部特征,Fusion 融合模块可以同时利用局部特征信息和全局特征信息。
- 2) 提出了 FEFN 模块,进一步细化从空间和通道维度上的特征,强调重要信息以及抑制冗余特征。
- 3) 本文在对公共基准数据集上提出的多尺度局部与全局特征融合网络 MLGFFN 进行了定量和定性评估,结果表明,MLGFFN 方法在降低模型复杂度和图像重建性能方面达到了较好的平衡。

2. 相关工作

2.1. 轻量级图像超分辨率模型

近年来,轻量级超分辨率网络的发展因其在移动设备和嵌入式系统等资源受限场景中的实际应用而受到越来越多的关注。人们提出了许多轻量级 SR 模型来降低模型复杂度和计算资源消耗量,同时保持令人满意的图像重建性能。SRCNN[1]首先利用端到端的卷积神经网络(CNN)结构学习低分辨率 LR 到高分辨率 HR 的映射关系,相较于传统方法在速度和效果上均有提升。FSRCNN[15]采用后上采样机制,取代了 SRCNN 在网络外部放大图片尺寸所产生的巨大计算成本。DCRN[16]结合了递归连接和优化的残差学习过程,在降低模型复杂度的同时提升了图像重建任务的性能。CARN[17]在残差网络上使用分组卷积和级联机制来提高效率。RCAN[18]通过引入通道注意力机制构建极深的网络结构,有效地处理特征通道之间的相互依赖关系。MRDN[19]提出了一种轻量级多阶段残差蒸馏块,允许信息在不同的深度上进行前向传播和跳跃连接,更好地捕获和融合不同尺度的特征。RMDN[20]结合了多分支并行特征提取和特征蒸馏,以低计算成本获得丰富的多尺度特征,并通过像素归一化来提高训练的稳定性。LKDN[21]把大核注意力机制引入到蒸馏网络架构,有效地扩大感受野,从而帮助模型捕捉图像的长程依赖关系和更丰

富的全局上下文信息。GPDN [22]引入了特征通道注意力模块，以精确地过滤和增强对恢复高分辨率图像至关重要的像素特征。

2.2. 大核卷积网络

自从 VGG [23]被提出以来， 3×3 等小型卷积核由于其高效率 and 轻量级的性质而得到了广泛的应用。然而随着 Transformer [2]模型的出现，强调全局感受野对视觉任务的重要性逐渐被认可。许多研究人员发现使用更大的卷积核可以减少模型的网络深度，同时能够获得与基于 Transformer 的 SR 模型相当的性能。ConvNeXt [24]使用大卷积核来获得更大的感受野，并实现与 Swin-Transformer [25]相当的性能。RepLKNet [26]使用深度卷积和重新参数化将内核扩展到 31×31 。SegNeXt [27]通过融合具有 7、11 和 21 大内核大小的深带状卷积分支来增强其捕获多尺度上下文信息的能力。VAN [28]提出了一种新的大核注意力(Large Kernel Attention, LKA)模块，将自注意力机制的长距离依赖建模能力与卷积网络的效率和结构优势相结合。VAPSR [29]将 LKA 模块引入到轻量级图像超分辨率，证明了 LKA 在保持模型轻量化的同时，能够显著提升重建质量，特别是在纹理细节丰富的场景。MAN [8]提出多尺度大内核注意力(MLKA)来改进 LKA 模块，将多尺度特征提取与精细的注意力机制相结合，实现了对不同尺度信息的自适应选择和聚合，使模型更擅长处理复杂、多尺度的细节信息。LKDN [21]将 LKA 应用于有效的信息蒸馏框架并实现了高效的性能。

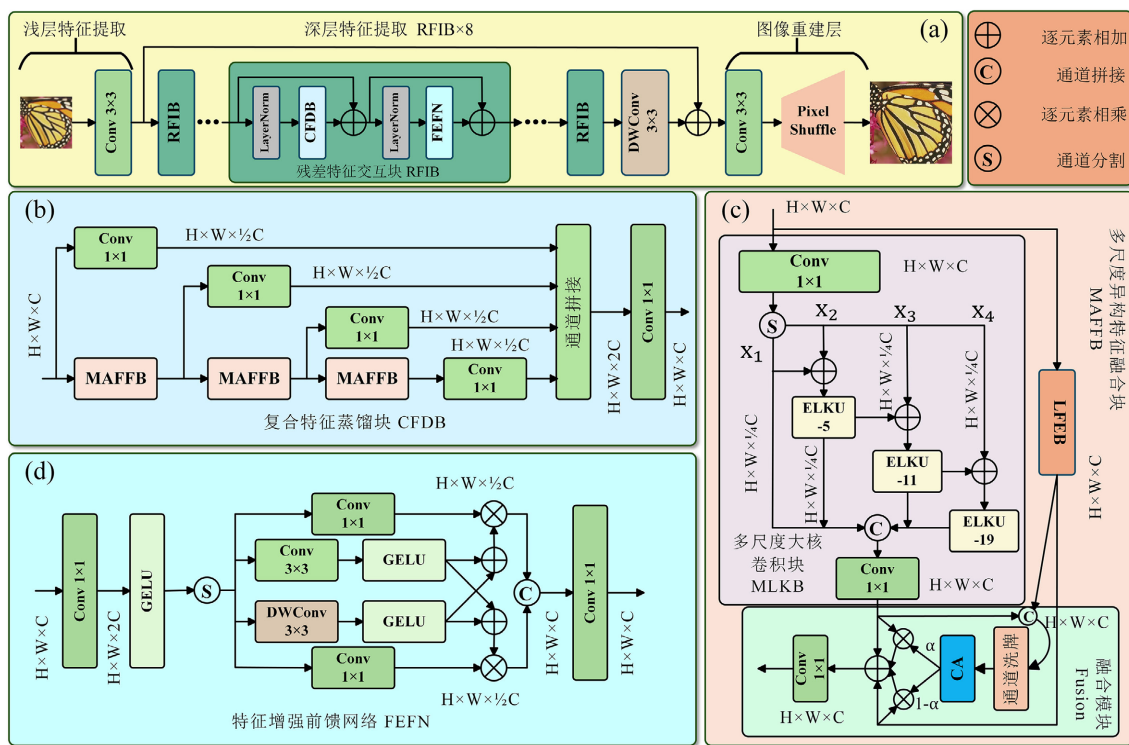


Figure 1. Overview of the Multi-scale Local-Global Feature Fusion Network (MLGFFN), including the architectures of the Composite Feature Distillation Block (CFDB), Feature Enhancement Feed-forward Network (FEFN), and Multi-scale Asymmetric Feature Fusion Block (MAFFB)

图 1. 多尺度局部与全局特征融合网络(MLGFFN)总览，包含了复合特征蒸馏块(CFDB)的结构、特征增强前馈网络(FEFN)的结构和多尺度异构特征融合块(MAFFB)的结构

3. 提出的方法

本节详细介绍多尺度局部与全局特征融合网络(MLGFFN)中每个组件的结构。图 1(a)展示了 MLGFFN

的整体结构, 包含了用于浅层特征提取的 3×3 卷积层、用于深层特征提取的 8 个堆叠的残差特征交互块和一个带有 3×3 卷积层以及 PixelShuffle 层[30]的图像重建层。图 1(b)和图 1(d)分别展示了复合特征蒸馏块(CFDB)和特征增强前馈网络(FEFN)的组织结构, 残差特征交互块(RFIB)由其组成。图 1(c)展示了复合特征蒸馏块的核心组件多尺度异构特征融合块的结构。

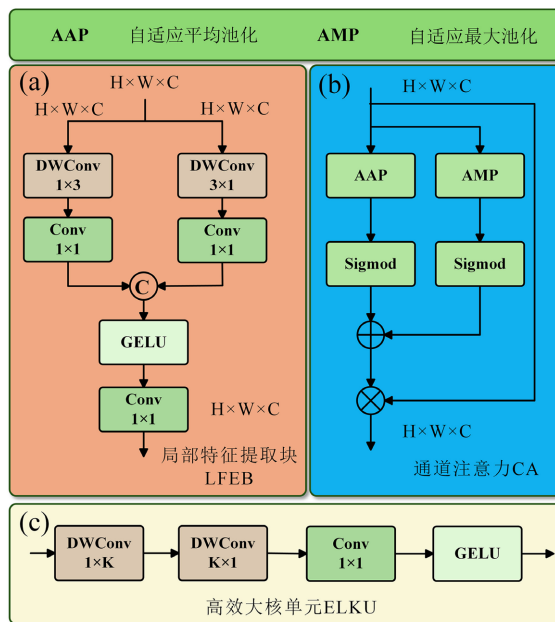


Figure 2. Architectures of the Local Feature Extraction Block (LFEB), Channel Attention (CA) module, and Efficient Large Kernel Unit (ELKU)

图 2. 局部特征提取块(LFEB)、通道注意力(CA)模块和高效大核单元(ELKU)的结构

3.1. 多尺度异构特征融合块

为了高效提取和融合图像的全局结构信息与局部细节纹理, 本文设计了多尺度异构特征融合块 (Multi-scale Asymmetric Feature Fusion Block, MAFFB)。如图 1(c)所示, MAFFB 采用双分支结构提取全局和局部特征, 其中多尺度大核卷积块(Multi-scale Large Kernel Block, MLKB)是用于捕捉大范围的上下文依赖关系的全局特征分支, 而局部特征提取块(Local Feature Extraction Block, LFEB)则是专注于提取高频细节的局部特征分支, 见图 2(a)。这两个分支的输出通过一个特征融合模块进行整合, 用于获得全局信息和局部信息自适应融合的特征图。

多尺度大核卷积块(MLKB)通过级联的多个卷积核大小不同的高效大核单元(Efficient Large Kernel Unit, ELKU)模块, 见图 2(c), 提取具有丰富的全局上下文特征。级联残差连接构建了一个信息流动的通道, 不同尺度的特征能够相互增强和融合, 为模型提供更丰富的特征信息表达。大核卷积核可以提供较大的感受野, 有助于理解特征图的长期依赖关系和结构信息, 保持重建图像的整体一致性。同时, ELKU 使用串联的非对称深度可分离大核卷积可以有效减少模型的参数量。本文选用了三个大小分别为 5、11 和 19 的 ELKU 模块。MLKB 具体实现如下:

$$\begin{aligned}
 X_1, X_2, X_3, X_4 &= S(\text{Conv}_{1 \times 1}(F_{in})) \\
 X_{2,3,4} &= \text{ELKU}_{5,11,19}(X_{2,3,4} + X_{1,2,3}) \\
 X_{out} &= \text{Conv}_{1 \times 1}(\text{Concat}(X_1, X_2, X_3, X_4))
 \end{aligned} \tag{1}$$

其中, $S(\cdot)$ 为通道分离操作, $Concat(\cdot, \cdot)$ 为通道拼接操作。 $ELKU_K$ 表示为使用卷积核尺寸为 K 的多尺度高效大核单元, 其实现如下所示:

$$\begin{aligned} X &= DWConv_{1 \times K} (DWConv_{K \times 1} (F_{in})) \\ F_{out} &= GELU (Conv_{1 \times 1} (X)) \end{aligned} \quad (2)$$

其中, $GELU$ [31] 为高斯误差线性单元激活函数(Gaussian Error Linear Unit), $DWConv_{1 \times K}$ 是核为 $1 \times K$ 的深度可分离卷积。局部特征提取块(LFEB)由并联的 1×3 和 3×1 尺寸的非对称深度可分离卷积构成, 用于高效捕捉特征图的局部高频信息, 如边缘、纹理和细节, 这些信息是提升图像感知质量的关键, 见图 2(a)。LFEB 与 ELKU 使用了异构的非对称卷积结构, LFEB 使用并行的非对称卷积设计, 独立地关注水平和垂直方向上的纹理信息, 在捕捉方向性的细节上有着不错的表现。最后用一个带有 GELU 激活函数的核为 1×1 的卷积增强具有丰富信息量的局部特征细节。LFEB 的实现如下:

$$\begin{aligned} Y_1 &= Conv_{1 \times 1} (DWConv_{1 \times 3} (F_{in})) \\ Y_2 &= Conv_{1 \times 1} (DWConv_{3 \times 1} (F_{in})) \\ Y_{out} &= Conv_{1 \times 1} (GELU (Concat (Y_1, Y_2))) \end{aligned} \quad (3)$$

最后 MAFFB 对来自上一层级输出的全局和局部信息特征图进行自适应特征融合。实现如下:

$$\begin{aligned} \alpha &= CA (F_{Shuffle}^{group} (Concat (X, Y))) \\ X_{fuse} &= \alpha * X \\ Y_{fuse} &= (1 - \alpha) * Y \\ F_{out} &= Conv_{1 \times 1} (X + Y + X_{fuse} + Y_{fuse}) \end{aligned} \quad (4)$$

其中, $*$ 为逐元素相乘(Element-wise Product)操作, $F_{Shuffle}^{group}(\cdot)$ 为通道洗牌操作[32]。CA 为通道注意力(Channel Attention, CA)模块, 见图 2(b)。CA 模块引入了自适应平均池化(Adaptive Average Pooling, AAP)和自适应最大池化(Adaptive Max Pooling, AMP)操作, 没有引入额外的参数量。双池化赋予特征图对重要细节(如清晰的点、精确的边缘)通道更高的权重, 同时抑制冗余特征信息的通道。CA 模块的实现如下所示:

$$F_{out} = F_{in} * (\sigma(AAP(F_{in})) + \sigma(AMP(F_{in}))) \quad (5)$$

其中, AAP 为自适应平均池化, AMP 为自适应最大池化, σ 为 Sigmoid 激活函数。

3.2. 复合特征蒸馏块

为了同时聚合局部信息和全局信息, 本文设计了一个复合特征蒸馏块(Composite Feature Distillation Block, CFDB)。CFDB 将多个多尺度异构特征融合块(MAFFB)嵌入到特征蒸馏的主分支中, 将 MAFFB 提取到的全局和局部融合信息以及通过多个 1×1 卷积层提取的全局信息进行通道拼接聚合。给定输入 F_{in} , CFDB 的过程可以表示为:

$$\begin{aligned} X_d^1, X_m^1 &= Conv_{1 \times 1}^1 (F_{in}), MAFFB_1 (F_{in}) \\ X_d^2, X_m^2 &= Conv_{1 \times 1}^2 (X_m^1), MAFFB_2 (X_m^1) \\ X_d^3, X_m^3 &= Conv_{1 \times 1}^3 (X_m^2), MAFFB_3 (X_m^2) \\ X_d^4 &= Conv_{1 \times 1}^4 (X_m^3) \end{aligned} \quad (6)$$

其中, X_m^i 、 X_d^i 分别代表主分支(m)和蒸馏分支(d)的第 i 层输出。最后本文将蒸馏步骤产生的所有蒸馏分

支和主分支拼接在一起，再通过 1×1 卷积聚合特征的信息，实现如下：

$$F_{out} = Conv_{1 \times 1} \left(Concat \left(X_d^1, X_d^2, X_d^3, X_d^4 \right) \right) \quad (7)$$

3.3. 特征增强前馈网络

上节所述的复合特征蒸馏块(CFDB)对于全局局部融合信息和全局信息进行蒸馏聚合，本文设计了特征增强前馈网络(Feature Enhancement Feed-forward Network, FEFN)对聚合后的信息进行增强。常规前馈网络(Feed-Forward Network, FFN) [2]对图像每个像素位置进行相同的操作，缺乏重要信息和非重要信息的侧重，因此本文对 FFN 进行改进。FEFN 的实现如下：

$$\begin{aligned} X_1, X_2, X_3, X_4 &= S \left(GELU \left(Conv_{1 \times 1} \left(F_{in} \right) \right) \right) \\ X'_1 &= Conv_{1 \times 1} \left(X_1 \right) \\ X'_2 &= DWConv_{3 \times 3} \left(X_2 \right) \\ X'_3 &= Conv_{3 \times 3} \left(X_3 \right) \\ X'_4 &= Conv_{1 \times 1} \left(X_4 \right) \\ Y_1 &= Concat \left(X'_1, X'_2 + X'_3 \right) \\ Y_2 &= Concat \left(X'_4, X'_2 + X'_3 \right) \\ F_{out} &= Conv_{1 \times 1} \left(Y_1 * Y_2 \right) \end{aligned} \quad (8)$$

3.4. 残差特征交互块

如上文所述，本文用多个堆叠的残差特征交互块(Residual Feature Interaction Block, RFIB)作为深层特征提取。为了增加模型训练的稳定性，本文使用了层归一化(Layer Normalization, LN) [33]和残差连接，RFIB 的过程可以表示为：

$$\begin{aligned} X &= LN \left(CFDB \left(F_{in} \right) \right) + F_{in} \\ F_{out} &= LN \left(FEFN \left(X \right) \right) + X \end{aligned} \quad (9)$$

4. 实验

4.1. 数据集处理及实现细节

4.1.1. 数据集

本文采用 DIV2K 数据集[34]作为训练集和验证集，包含了各种复杂场景的高分辨率图片，其中训练集图片有 800 张，验证集图片有 100 张。为了有效评估本文模型的能力，本文选用了常用的五个测试数据集进行测试，包括 Set5 [35]、Set14 [36]、BSD100 [37]、Urban100 [38]和 Manga109 [39]。本文使用峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)和结构相似性指数(Structure Similarity Index Measure, SSIM)来评估恢复图像的质量。这两种指标都是在转换为 YCbCr 颜色空间的图像的 Y 通道上计算的。

4.1.2. 实现细节

在训练期间，本文将输入的 LR 图像进行随机水平翻转和旋转的数据增强，并将训练集图像随机裁剪成大小为 64×64 像素的图像块。本模型采用与 SMFANet [5]方法相同的损失函数即 L1 Loss 和 FFT Loss 损失函数进行训练，并由 Adam 优化器[40]进行优化，其中 $\beta_1 = 0.9$ ， $\beta_2 = 0.99$ 。实验的迭代次数设置为 500,000 次。初始学习率为 1×10^{-3} ，最小学习率设置为 1×10^{-6} ，并使用余弦退火方案[41]进行更新。所有的实验均在 NVIDIA GeForce RTX 3090 GPU 上使用 PyTorch 框架进行。

4.2. 与先进方法对比

4.2.1. 定量分析

为了充分评估模型的性能, 本文与 SMSR [42]、ShuffleMixer [43]、SAFMN [44]、SMFANet [5]、CFM [45]、MAN-tiny [8]、MSWSR [46]和 ESDAN [47]方法进行比较。表 1 展示了各种方法在基准数据集上 $\times 2$ 、 $\times 3$ 和 $\times 4$ 放大因子的定量比较。除了 PSNR 和 SSIM 指标之外, 本文还列出了参数数量(#Params)和浮点运算次数(#FLOPs)用于评估方法的模型复杂度。为了公平比较, 本文在将 LR 图像超分辨率至 1280×720 像素的设置下进行所有的模型复杂度计算。

如表 1 所示, 本文提出的 MLGFFN 方法在实验中的所有放大倍率下, 均在所有基准数据集上取得了最优异的性能指标, 充分证明了其优秀的图像重建能力。例如, 在 $\times 4$ 放大因子的 Set14 数据集上, MLGFFN 的性能比 SMSR 方法高出 0.21 dB, 而参数数量和 FLOPs 仅为 SMSR 方法的 37.38%和 40%。这说明了 MLGFFN 在有效降低参数数量和计算消耗成本的同时能获取更多的特征信息。而与 CFM 相比, MLGFFN 以仅仅增加了 18.99%参数量的代价, 获得了更优的性能, 在 $\times 4$ 放大因子的所有数据集上比 CFM 平均高出了 0.276 dB, 同时 FLOPs 下降了 26.96%。而在 $\times 2$ 放大因子和 $\times 3$ 放大因子的所有数据集上分别比 CFM 平均高出 0.324 dB 和 0.272 dB。MLGFFN 方法在图像重建性能与资源消耗量之间取得了优秀的平衡。

Table 1. Performance of different methods on various datasets

表 1. 不同方法在不同的数据集上的表现

倍率	方法	#Params (K)	#FLOPs (G)	Set5	Set14	BSD100	Urban100	Manga109
				PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
$\times 2$	SMSR	985	132	38.00/0.9601	33.64/0.9179	32.17/0.8990	32.19/0.9284	38.76/0.9771
	ShuffleMixer	394	91	38.01/0.9606	33.63/0.9180	32.17/0.8995	31.89/0.9257	38.83/0.9774
	SAFMN	228	52	38.00/0.9605	33.54/0.9177	32.16/0.8995	31.84/0.9256	38.71/0.9771
	SMFANet	186	41	38.04/0.9606	33.65/0.9186	32.19/0.8999	32.22/0.9283	38.98/0.9776
	CFM	298	71	37.97/0.9608	33.62/0.9182	32.17/0.9003	31.94/0.9266	38.64/0.9773
	MAN-tiny	134	30	37.93/0.9604	33.48/0.9171	32.13/0.8993	31.75/0.9250	38.58/0.9770
	MSWSR	312	-	38.01/0.9601	33.71/0.9193	32.22/0.9003	32.29/0.9301	38.86/0.9774
	ESDAN	541	97.4	38.07/0.9607	33.67/0.9183	32.22/0.9003	32.30/0.9296	39.02/0.9777
	MLGFFN	364	65.2	38.15/0.9611	33.81/0.9199	32.33/0.9017	32.52/0.9320	39.15/0.9781
$\times 3$	SMSR	993	68	34.40/0.9270	30.33/0.8412	29.10/0.8050	28.25/0.8536	33.68/0.9445
	ShuffleMixer	415	43	34.40/0.9272	30.37/0.8423	29.12/0.8051	28.08/0.8498	33.69/0.9448
	SAFMN	233	223	34.34/0.9267	30.33/0.8418	29.08/0.8048	27.95/0.8474	33.52/0.9437
	SMFANet	191	19	34.46/0.9275	30.39/0.8432	29.13/0.8059	28.25/0.8525	33.84/0.9454
	CFM	319	34	34.37/0.9272	30.33/0.8419	29.09/0.8060	28.02/0.8495	33.49/0.9441
	MAN-tiny	141	14	34.24/0.9259	30.25/0.8405	29.04/0.8046	27.85/0.8465	33.28/0.9426
	MSWSR	307	-	34.40/0.9277	30.35/0.8437	29.12/0.8067	28.22/0.8548	33.68/0.9454
	ESDAN	546	45.6	34.47/0.9280	30.38/0.8429	29.14/0.8065	28.29/0.8547	33.85/0.9459
	MLGFFN	369	29.3	34.56/0.9285	30.46/0.8450	29.21/0.8083	28.45/0.8578	33.98/0.9472
$\times 4$	SMSR	1006	42	32.12/0.8932	28.55/0.7808	27.55/0.7351	26.11/0.7868	30.54/0.9085
	ShuffleMixer	411	28	32.21/0.8953	28.66/0.7827	27.61/0.7366	26.08/0.7835	30.65/0.9093
	SAFMN	240	14	32.18/0.8948	28.60/0.7813	27.58/0.7359	25.97/0.7809	30.43/0.9063
	SMFANet	197	11	32.25/0.8956	28.67/0.7825	27.61/0.7371	26.19/0.7861	30.72/0.9097

续表

CFM	316	23	32.17/0.8955	28.62/0.7824	27.58/0.7376	25.99/0.7826	30.38/0.9073
MAN-tiny	150	8.4	32.07/0.8930	28.53/0.7801	27.51/0.7345	25.84/0.7786	30.18/0.9047
MSWSR	316	-	32.26/0.8966	28.67/0.7843	27.62/0.7379	26.17/0.7896	30.60/0.9092
ESDAN	554	26.1	32.26/0.8959	28.66/0.7834	27.62/0.7373	26.22/0.7897	30.72/0.9109
MLGFFN	376	16.8	32.38/0.8972	28.76/0.7850	27.68/0.7391	26.34/0.7916	30.96/0.9123

4.2.2. 定性分析

本文选取了具有丰富细节纹理的 Urban100 数据集[38], 将 MLGFFN 方法的可视化结果与 SAFMN [44]、CFM [45]、ShuffleMixer [43]、MAN-tiny [8]、SMFANet [5]和 ESDAN [47]方法的可视化结果在 $\times 4$ 放大因子上进行对比。图 3 列出了 Urban100 数据集中的 img019 和 img084 两张图片的可视化对比结果。不难看出, 对于 img019 图片的图像重建情况, 其他方法对不同尺寸的区域恢复能力差, 不能很好地理解多尺度信息。然而, MLGFFN 通过级联的不同尺寸的大核卷积有效地捕获多尺度信息, 能正确地恢复不同尺寸区域的信息, 而通道注意力和通道洗牌操作能很好地理解区域之间的交互关系, 避免了区块黏连现象。而对于 img084 图片的细节纹理恢复情况来看, 其他方法重建后的图像都出现了 HR 图像中所没有的交叉线, 而本文的 MLGFFN 方法重建出来的图像并未出现错误的线条, 这得益于本文局部特征提取块 (LFEB) 对方向性细节纹理的理解能力。图 3 展示了 MLGFFN 在主观视觉质量方面表现更好, 证明了本文提出的 MLGFFN 图像重建方法的有效性。

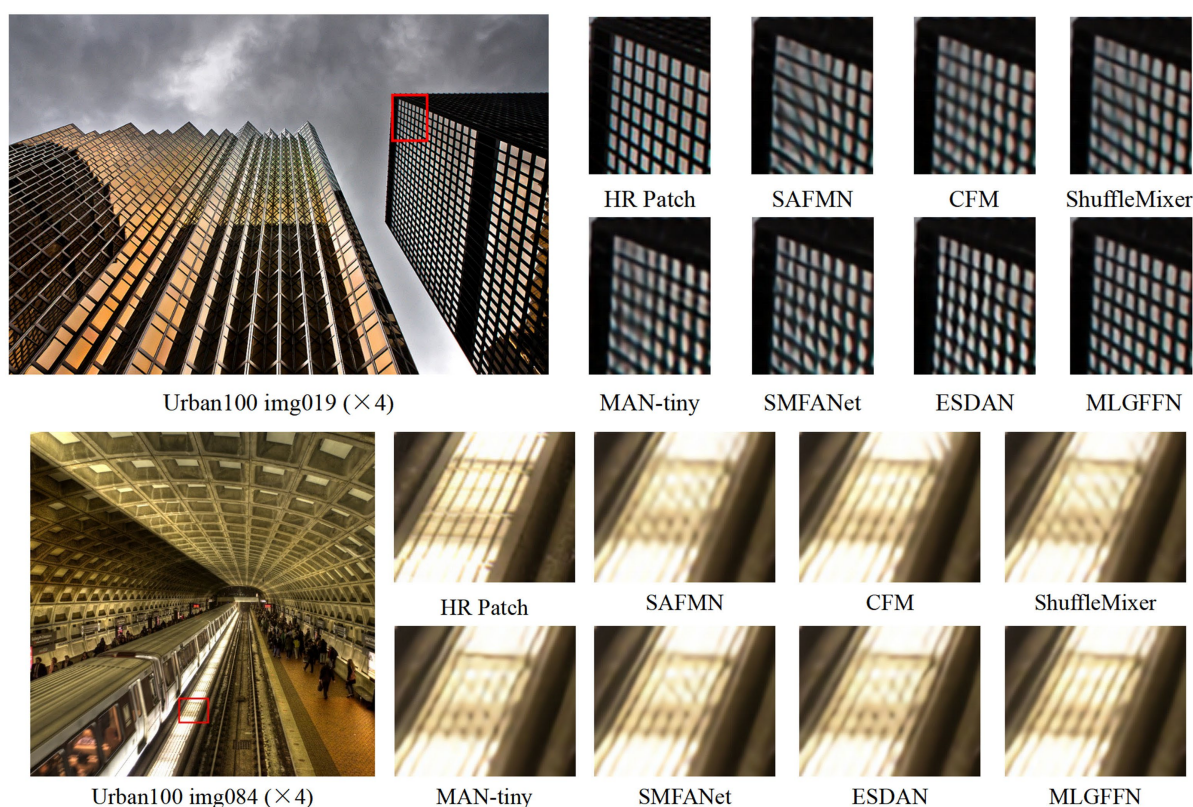


Figure 3. Visual results of different methods on Urban100 image019 and Urban100 image084 ($\times 4$)

图 3. 各个方法在 Urban100 image019 及 Urban100 image084 图像上的可视化结果($\times 4$)

4.3. 消融实验

在本节中，本文进行了广泛的消融研究，以分析和评估所提出的 MLGFFN 中每个模块的效果，见表 2。本文基于×4 的 MLGFFN 模型实现了所有消融实验，并使用 DIV2K 数据集[34]对其进行训练以进行公平比较。表中的定量消融结果均是在 Set5 数据集[35]和 Manga109 数据集[39]上测量的。

本文所提出的 MLGFFN 模型的核心模块为多尺度异构特征融合块(MAFFB)和特征增强前馈网络(FEFN)，为了验证其有效性，本文分别移除两个模块，与 MLGFFN 基准模型进行性能对比。表 2 显示，移除 MAFFB 在×4 的 Set5 和 Manga109 数据集上分别降低了 1.38 dB 和 1.28 dB，而移除 FEFN 模块在×4 的 Set5 和 Manga109 数据集上分别降低了 0.38 dB 和 0.44 dB。实验结果表明，多尺度异构特征融合块(MAFFB)和特征增强前馈网络(FEFN)在 MLGFFN 网络中的必要性。

此外，本文提出的 MAFFB 中包含了一个用于探索全局信息的 MLKB、一个用于探索局部信息的 LFEB 以及一个用于融合局部和全局信息的融合模块。由表 2 可以看出，本文对 MLKB 模块或者 LFEB 模块的禁用都导致网络性能的大幅度下降，而把融合模块换成简单的特征相加融合，则会让模型在×4 的 Set5 和 Manga109 数据集上降低 0.19 dB 和 0.15 dB 精度。融合模块通过通道注意力 CA 模块和通道洗牌操作增强通道信息的交互能力，从表 2 中可以看出当融合模块禁用 CA 和通道洗牌操作后都出现了一定程度的精度下降。本文的 MLKB 模块相比 Res2Net 采用了不同尺寸的大核非对称的深度可分离卷积，表 2 显示将 MLK-K 模块替换成 3×3 卷积后，在×4 的 Set5 和 Manga109 数据集上的精度分别下降了 0.26 dB 和 0.21 dB，这得益于 MLK-K 模块捕获不同尺寸区域的特征信息的能力。

FEFN 模块相比 FFN 模块，对重要信息和非重要信息有所侧重，对来自上层融合后的特征信息进行增强。本文把 FEFN 模块替换成 FFN 模块后，网络性能在×4 的 Set5 和 Manga109 数据集上分别降低了 0.11 dB 和 0.14 dB，这主要是 FFN 模块对每个位置的特征信息都进行相同的操作，缺乏有效的重要信息筛选能力。

Table 2. Ablation study results of MLGFFN on Set5 and Manga109 (×4)

表 2. MLGFFN 在 Set5 及 Manga109 数据集上的消融实验结果(×4)

消融实验	变化		#Params (K)	#FLOPs (G)	Set5 PSNR/SSIM	Manga109 PSNR/SSIM
MLGFFN	-		376	16.84	32.38/0.8972	30.96/0.9123
核心模块	MAFFB	FEFN	-			
	×	√	312	14.02	31.00/0.8663	29.68/0.8973
	√	×	122	5.52	32.11/0.8925	30.52/0.9084
MAFFB	w/o MLKB + Fusion		210	9.36	31.24/0.8702	29.84/0.9012
	w/o LFEB + Fusion		257	11.51	31.56/0.8806	30.24/0.9056
	Fusion → +		343	15.36	32.19/0.8932	30.81/0.9113
	w/o Channel Attention		376	16.80	32.29/0.8952	30.87/0.9113
	w/o Channel Shuffle		376	16.84	32.31/0.8958	30.88/0.9119
	MLK-K → 3×3 卷积		406	18.30	32.12/0.8920	30.75/0.9106
FEFN	FEFN → FFN		345	15.46	32.27/0.8948	30.82/0.9109

5. 结论

本文提出了一种简单而有效的 MLGFFN 方法，有效地解决了图像超分辨率的轻量化和性能平衡问

题。在 MLGFFN 方法中,提出了融合局部特征和全局特征的多尺度异构特征融合块(MAFFB)以及用于增强重要信息特征的特征增强前馈网络(FEFN)。在 MAFFB 模块中,本文设计了多尺度大核卷积块(MLKB)模块和局部特征提取块(LFEB)模块用于捕获全局特征和局部特征。大量的实验结果表明,本文所提出的 MLGFFN 方法有卓越的图像重建能力,同时还保持了 CNN 的低资源消耗特性。本文提出的 MLGFFN 方法设计的局部特征提取分支轻量同时能够表达图像的方向性细节纹理。但由于设计比较简洁,在图像重建质量上还有可进步的空间。未来的工作可以使用小波变换对特征图的高频和低频分量进行分离,以达到更好的图像重建质量。

参考文献

- [1] Dong, C., Loy, C.C., He, K. and Tang, X. (2016) Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**, 295-307. <https://doi.org/10.1109/tpami.2015.2439281>
- [2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- [3] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L. and Timofte, R. (2021) SwinIR: Image Restoration Using Swin Transformer. 2021 *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, 11-17 October 2021, 1833-1844. <https://doi.org/10.1109/iccvw54120.2021.00210>
- [4] Zhou, Y., Li, Z., Guo, C., Bai, S., Cheng, M. and Hou, Q. (2023) SRFormer: Permuted Self-Attention for Single Image Super-Resolution. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 12780-12791. <https://doi.org/10.1109/iccv51070.2023.01174>
- [5] Zheng, M., Sun, L., Dong, J. and Pan, J. (2024) SMFANet: A Lightweight Self-Modulation Feature Aggregation Network for Efficient Image Super-Resolution. In: *European Conference on Computer Vision*, Springer, 359-375. https://doi.org/10.1007/978-3-031-72973-7_21
- [6] Zong, Z., Zha, L., Jiang, J. and Liu, X. (2022) Asymmetric Information Distillation Network for Lightweight Super Resolution. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, New Orleans, 19-20 June 2022, 1249-1258. <https://doi.org/10.1109/cvprw56347.2022.00131>
- [7] Lee, D., Yun, S. and Ro, Y. (2024) Partial Large Kernel CNNs for Efficient Super-Resolution. arXiv: 2404.11848.
- [8] Wang, Y., Li, Y., Wang, G. and Liu, X. (2024) Multi-Scale Attention Network for Single Image Super-Resolution. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, 17-18 June 2024, 5950-5960. <https://doi.org/10.1109/cvprw63382.2024.00602>
- [9] Yu, Z., Chen, L., Zeng, Z., Yang, K., Luo, S., Chen, S., et al. (2024) LGFN: Lightweight Light Field Image Super-Resolution Using Local Convolution Modulation and Global Attention Feature Extraction. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, 17-18 June 2024, 6712-6721. <https://doi.org/10.1109/cvprw63382.2024.00665>
- [10] Zhao, L., Wang, Y., Qing, Y., Zeng, D. and Xu, L. (2026) MCFINet: A Cost-Efficient Multi-Channel Feature Integration Network for Surface Scenarios Image Super-Resolution. *ACM Transactions on Multimedia Computing, Communications, and Applications*, **22**, 1-17. <https://doi.org/10.1145/3777465>
- [11] Zhang, A., Guo, B., Liu, X. and Liu, W. (2025) HIEN: A Hybrid Interaction Enhanced Network for Horse Iris Super-Resolution. *Applied Sciences*, **15**, Article No. 7191. <https://doi.org/10.3390/app15137191>
- [12] Gao, T. and Liu, Y. (2025) PECNet: A Lightweight Single-Image Super-Resolution Network with Periodic Boundary Padding Shift and Multi-Scale Adaptive Feature Aggregation. *Symmetry*, **17**, Article No. 1833. <https://doi.org/10.3390/sym17111833>
- [13] Wang, Y. and Zhang, T. (2024) Ossfnet: Omni-Stage Feature Fusion Network for Lightweight Image Super-Resolution. *Proceedings of the AAI Conference on Artificial Intelligence*, **38**, 5660-5668. <https://doi.org/10.1609/aaai.v38i6.28377>
- [14] Gao, S., Cheng, M., Zhao, K., Zhang, X., Yang, M. and Torr, P. (2021) Res2net: A New Multi-Scale Backbone Architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 652-662. <https://doi.org/10.1109/tpami.2019.2938758>
- [15] Dong, C., Loy, C.C. and Tang, X. (2016) Accelerating the Super-Resolution Convolutional Neural Network. In: *European Conference on Computer Vision*, Springer International Publishing, 391-407. https://doi.org/10.1007/978-3-319-46475-6_25
- [16] Feng, Z., Lai, J., Xie, X. and Zhu, J. (2018) Image Super-Resolution via a Densely Connected Recursive Network.

- Neurocomputing*, **316**, 270-276. <https://doi.org/10.1016/j.neucom.2018.07.076>
- [17] Ahn, N., Kang, B. and Sohn, K. (2018) Fast, Accurate, and Lightweight Super-Resolution with Cascading Residual Network. In: Ferrari, V., et al., Eds., *Computer Vision—ECCV 2018*, Springer International Publishing, 256-272. https://doi.org/10.1007/978-3-030-01249-6_16
- [18] Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B. and Fu, Y. (2018) Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In: Ferrari, V., et al., Eds., *Computer Vision—ECCV 2018*, Springer International Publishing, 294-310. https://doi.org/10.1007/978-3-030-01234-2_18
- [19] Yang, X., Guo, Y., Li, Z., Zhou, D. and Li, T. (2022) MRDN: A Lightweight Multi-Stage Residual Distillation Network for Image Super-Resolution. *Expert Systems with Applications*, **204**, Article ID: 117594. <https://doi.org/10.1016/j.eswa.2022.117594>
- [20] Gao, X., Zhou, Y., Wu, S., Wu, X., Wang, F. and Hu, X. (2024) Residual Multi-Branch Distillation Network for Efficient Image Super-Resolution. *Multimedia Tools and Applications*, **83**, 75217-75241. <https://doi.org/10.1007/s11042-024-18504-5>
- [21] Xie, C., Zhang, X., Li, L., Meng, H., Zhang, T., Li, T., et al. (2023) Large Kernel Distillation Network for Efficient Single Image Super-Resolution. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Vancouver, 17-24 June 2023, 1283-1292. <https://doi.org/10.1109/cvprw59228.2023.00135>
- [22] Hong, Z., Liang, G. and Xiong, L. (2025) Gradient Pooling Distillation Network for Lightweight Single Image Super-Resolution Reconstruction. *PeerJ Computer Science*, **11**, e2679. <https://doi.org/10.7717/peerj-cs.2679>
- [23] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv: 1409.1556.
- [24] Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T. and Xie, S. (2022) A ConvNet for the 2020s. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 11966-11976. <https://doi.org/10.1109/cvpr52688.2022.01167>
- [25] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 10012-10022. <https://doi.org/10.1109/iccv48922.2021.00986>
- [26] Ding, X., Zhang, X., Han, J. and Ding, G. (2022) Scaling up Your Kernels to 31×31 : Revisiting Large Kernel Design in CNNs. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 11963-11975. <https://doi.org/10.1109/cvpr52688.2022.01166>
- [27] Cheng, M., Guo, M., Hou, Q., Hu, S., Liu, Z. and Lu, C. (2022) SegNeXt: Rethinking Convolutional Attention Design for Semantic Segmentation. In: *Advances in Neural Information Processing Systems 35*, Neural Information Processing Systems Foundation, Inc., 1140-1156. <https://doi.org/10.52202/068431-0084>
- [28] Guo, M.H., Lu, C.Z., Liu, Z.N., Cheng, M.M. and Hu, S.M. (2023) Visual Attention Network. *Computational Visual Media*, **9**, 733-752. <https://doi.org/10.1007/s41095-023-0364-2>
- [29] Zhou, L., Cai, H., Gu, J., Li, Z., Liu, Y., Chen, X., et al. (2023) Efficient Image Super-Resolution Using Vast-Receptive-Field Attention. In: *European Conference on Computer Vision*, Springer, 256-272. https://doi.org/10.1007/978-3-031-25063-7_16
- [30] Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A.P., Bishop, R., et al. (2016) Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 1874-1883. <https://doi.org/10.1109/cvpr.2016.207>
- [31] Hendrycks, D. (2016) Gaussian Error Linear Units (GELUs). arXiv: 1606.08415.
- [32] Zhang, X., Zhou, X., Lin, M. and Sun, J. (2018) ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 6848-6856. <https://doi.org/10.1109/cvpr.2018.00716>
- [33] Ba, J.L., Kiros, J.R. and Hinton, G.E. (2016) Layer Normalization. arXiv: 1607.06450.
- [34] Timofte, R., Agustsson, E., Van Gool, L., Yang, M.H. and Zhang, L. (2017) Ntire 2017 Challenge on Single Image Super-Resolution: Methods and Results. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, 21-26 July 2017, 114-125.
- [35] Bevilacqua, M., Roumy, A., Guillemot, C. and Morel, M.A. (2012) Low-Complexity Single-Image Super-Resolution Based on Nonnegative Neighbor Embedding. In: *Proceedings of the British Machine Vision Conference 2012*, BMVA Press, 135.1-135.10. <https://doi.org/10.5244/c.26.135>
- [36] Zeyde, R., Elad, M. and Protter, M. (2012) On Single Image Scale-Up Using Sparse-Representations. In: *International Conference on Curves and Surfaces*, Springer, 711-730. https://doi.org/10.1007/978-3-642-27413-8_47
- [37] Martin, D., Fowlkes, C., Tal, D. and Malik, J. (2001) A Database of Human Segmented Natural Images and Its Application

-
- to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. *Proceedings 8th IEEE International Conference on Computer Vision. ICCV 2001*, Vol. 2, 416-423. <https://doi.org/10.1109/iccv.2001.937655>
- [38] Huang, J., Singh, A. and Ahuja, N. (2015) Single Image Super-Resolution from Transformed Self-Exemplars. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 7-12 June 2015, 5197-5206. <https://doi.org/10.1109/cvpr.2015.7299156>
- [39] Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., *et al.* (2016) Sketch-Based Manga Retrieval Using Manga109 Dataset. *Multimedia Tools and Applications*, **76**, 21811-21838. <https://doi.org/10.1007/s11042-016-4020-z>
- [40] Kingma, D.P. (2014) Adam: A Method for Stochastic Optimization. arXiv: 1412.6980.
- [41] Loshchilov, I. and Hutter, F. (2016) SGDR: Stochastic Gradient Descent with Warm Restarts. arXiv: 1608.03983.
- [42] Wang, L., Dong, X., Wang, Y., Ying, X., Lin, Z., An, W., *et al.* (2021) Exploring Sparsity in Image Super-Resolution for Efficient Inference. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 4917-4926. <https://doi.org/10.1109/cvpr46437.2021.00488>
- [43] Pan, J., Sun, L. and Tang, J. (2022) ShuffleMixer: An Efficient Convnet for Image Super-Resolution. *Advances in Neural Information Processing Systems* 35, New Orleans, 28 November-9 December 2022, 17314-17326. <https://doi.org/10.52202/068431-1259>
- [44] Sun, L., Dong, J., Tang, J. and Pan, J. (2023) Spatially-Adaptive Feature Modulation for Efficient Image Super-Resolution. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 13190-13199. <https://doi.org/10.1109/iccv51070.2023.01213>
- [45] Wu, Z., Liu, W. and Huang, D. (2024) When Handcrafted Filter Meets CNN: A Lightweight Conv-Filter Mixer Network for Efficient Image Super-Resolution. *Proceedings of the 2024 International Conference on Multimedia Retrieval*, Phuket, 10-14 June 2024, 722-730. <https://doi.org/10.1145/3652583.3658003>
- [46] Song, W., Yan, X., Guo, W., Xu, Y. and Ning, K. (2025) MSWSR: A Lightweight Multi-Scale Feature Selection Network for Single-Image Super-Resolution Methods. *Symmetry*, **17**, Article No. 431. <https://doi.org/10.3390/sym17030431>
- [47] Zhang, C., Tu, X., Cui, Z., Gu, X., Li, K. and Lu, Y. (2025) A General Lightweight Image Super-Resolution with Sharpening Enhancement and Double Attention Network. *Scientific Reports*, **15**, Article No. 40848. <https://doi.org/10.1038/s41598-025-24493-8>