

基于特征增强的轻量化可逆超分辨率网络

杨本臣, 李浩然, 金海波

(辽宁工程技术大学软件学院 葫芦岛 125000)
(yangbenchen@lntu.edu.cn)

摘要: 针对当前深度超分网络参数量非常庞大, 且推理速度缓慢, 而轻量化网络在复杂环境条件下无法表达图像深层特征的问题, 提出基于特征增强的轻量化可逆超分辨率网络。首先提出边缘特征残差, 配合所提的边缘相似损失指导模型重建, 增强重建图像对纹理轮廓的表达能力; 然后补充新的小波特征核, 使小波变换支持任意缩放因子的重建任务; 最后引入全局特征提取模块, 在特征图中嵌入自注意力机制, 提取全局特征。在缩放因子为 4 时的基准测试集 Set5 上的实验结果表明, 与 SwinIR-light 相比, 所提网络表现更优, PSNR 提升 0.41 dB, 参数量减少 244k, 推理时间缩短 49.05%。

关键词: 超分辨率; 小波变换; 边缘特征; 可逆神经网络
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2023-00319

A Lightweight Reversible Super-Resolution Network Based on Feature Enhancement

Yang Benchen, Li Haoran, and Jin Haibo

(School of Software, Liaoning Technical University, Huludao 125000)

Abstract: Deep super-resolution reconstruction networks have a large number of parameters and slow inference speed. Lightweight networks unable to express deep features of images under complex environmental conditions. Aiming at those issues, a lightweight reversible super-resolution network based on feature enhancement is proposed. Firstly, edge feature residuals are proposed, combined with the proposed edge similarity loss to guide model reconstruction and enhance the expression of texture contours. Then, add a new wavelet feature kernel to support the reconstruction task with any scaling factors. Finally, a global feature extraction module is introduced to embed a self attention mechanism in the feature map to extract global features. The proposed network showed better performance than SwinIR-light on the benchmark test Set5 with a scaling factor of 4. It improved the PSNR by 0.41 dB, decreased the number of parameters by 244k, and reduced the inference time by 49.05%.

Key words: super-resolution; wavelet transform; edge features; reversible neural network

图像单帧超分辨率(super-resolution, SR)重建任务广泛应用于视频监控^[1]、医疗^[2]和航空航天^[3]等领域。SR 重建就是将低分辨率(low-resolution, LR)的图

像通过 SR 重建算法形成高分辨率(high-resolution, HR)图像的过程^[4], SR 图像上得到的信息在实际应用中更有效、更有价值。

收稿日期: 2023-06-15; 修回日期: 2024-01-10. 基金项目: 国家自然科学基金(62173171). 杨本臣(1975—), 男, 博士, 副教授, 硕士生导师, CCF 会员, 主要研究方向为计算机图形学、深度学习、智能数据处理; 李浩然(2000—), 男, 硕士研究生, CCF 学生会员, 主要研究方向为图像超分辨率重建; 金海波(1983—), 男, 博士, 副教授, 硕士生导师, CCF 会员, 主要研究方向为计算机视觉, 复杂系统可靠性分析。

在深度 SR 网络模型中,卷积层和上采样层都有大量的参数需要学习,且随着网络的宽度和深度增加,学习参数庞大.网络需要大量数据和计算资源才能完成训练,否则网络会产生过拟合现象;通常,网络参数量增大与模型的内存消耗成正相关,亦会增加模型的部署和运维成本.早期,轻量化的深度 SR 网络由以卷积神经网络(convolutional neural network, CNN)为骨干的大型网络移除网络结构得到更少参数和层级的新网络,但其重建能力和视觉效果不如大型深度 SR 网络.当前,轻量 SR 网络大多依靠 Vision-Transformer^[5]的轻量化发展,虽然嵌入注意力机制增加了模型的重建能力,但大多数轻量化 SR 网络在重建结果和视觉效果上仍需加强.传统的深度 SR 重建网络发展越来越庞大,无法很好地满足实际需求,而大多数轻量化网络无法达到大网络的重建结果.

结合以上问题,本文提出基于特征增强的轻量化可逆 SR 网络(lightweight reversible SR network based on feature enhancement, ELRN).首先,提出边缘特征残差,与小波映射出的特征图相比,可以更直接地将边缘特征作用至局部特征提取的特征图上,并采用边缘相似度损失函数(edge similarity loss, ESloss),在复杂环境下指导模型学习到更多纹理特征;其次,补充新的 Harr 小波特征核,使得小波映射可以适用于任意缩放因子,沿用可逆残差网络中的可逆结构,修改可逆仿射运算,将残差结构修改为稠密卷积结构用作局部特征提取,亦可以保持在可逆结构中的双射性并使网络参数高效利用;最后,进行全局特征提取,向特征图中嵌入自注意力机制,利用隐层维度的变换与 1×1 卷积使得特征图前后维度保持一致.

1 相关工作

近年来,基于深度学习的 SR 重建算法快速发展. Dong 等^[6]提出基于 CNN 的 SR 重建网络,首次将深度学习引入到图像 SR 重建研究中,通过神经网络强大的学习能力,使深度学习可以处理 SR 重建任务,并通过插值将输入图像进行上采样,完成重建任务;但该网络只适用于单一尺度且过于依赖于图像区域的上下文联系. Kim 等^[7]借助深度残差网络^[8]的深度优势,提出更深层次的残差网络,增大特征提取的感受野,可以得到更多的上下文信息,以及全局特征;该网络也适用于多尺度放大,其使用梯度裁剪的训练技巧在一定程度上解

决了梯度爆炸,但相对于先前的工作,其参数量增加了 1 160%. Lim 等^[9]提出增强残差结构对残差结构进行最大程度的修改,去除了批量归一化并减少内存消耗,以此堆叠更多的残差卷积层提取更多特征,却增加了模型的计算负担. Ledig 等^[10]将生成对抗网络首次引入图像 SR 重建研究中,解决了生成 SR 图像中的纹理细节问题,使得图像更加自然,有更好的视觉效果;但生成网络的图像往往存在幻视或伪影的现象,且对抗生成网络常有易崩溃、难拟合等缺点,训练的终止条件不定,代价太高. Yang 等^[11]将 Vision-Transformer 引入图像 SR 重建研究中,其中嵌入注意力机制达到了更好的效果. Chen 等^[12]通过大型图像数据集进行预训练,提升 Transformer 的效果,但增大了模型参数量,且需要大量训练样本.

传统的神经网络往往需要构建不同的上下采样模块,如对称结构的 U-Net^[13]、编码器与解码器相结合的 VAE^[14]等. Gomez 等^[15]提出可逆残差网络,实现了网络的可逆操作,其上下采样共享权重参数,减少所用参数量及数据特征丢失,也能减少神经网络在反向传播过程中内存的消耗;但该网络依赖于残差块以及残差跳转实现的双射操作,与稠密网络结构^[16]相比,会造成有效特征的损失,在应用到 SR 重建任务时,保留有效特征的最大占比显得至关重要.

Ardiszone 等^[17]对可逆神经网络的逆向处理问题进行更详细的研究,将反向传播中混入随机噪声来拟合正向与逆向输出的特征分布相似度,其使用迁移学习中常见的最大均值差异(max mean discrepancy, MMD)损失^[18]衡量特征分布之间是否同源并计算它们之间的距离;但该损失在设计理念上与需要 SR 输出的理念相违背. Xiao 等^[19]将可逆神经网络应用到图像缩放研究中,通过对原可逆神经网络中的大部分仿射运算进行修改,使其更适用于图像缩放任务;但其依旧使用常规小波变换,没有直接与缩放因子产生联系. Jing 等^[20]将可逆神经网络算法应用到图像隐藏研究中,其中使用小波变换将数据图像进行下采样操作,该技术在以往图像处理任务中多有体现. Li 等^[21-22]采用小波变换处理图像的噪声干扰,其设计的小波变换层适用于 Harr 小波、Daubechies 小波和双正交小波,通过充分地抑制混叠效应,取得了更高的图像重建准确率,增强模型的鲁棒性. 上述 3 种常用离散的小波变换中, Harr 小波具有对称性,且滤波器长度为 2,可以更好地搭建小波特征核,合理地

选取特征分量, 增强输出特征可解释性. 由于具有完整对称性的小波在特征提取过程中可以有效地避免相位畸变, 即避免多余噪声产生, 因此本文使用 Harr 小波对图像进行上下采样, 并对其进行扩展.

常规的小波变换只支持缩放因子为 2^n 形式, 当缩放因子为 3 及其他奇数时并不适用, 本文对小波特征核进一步补充. 由于小波变换的映射图可以表现图像的边缘特征, 可间接地将其嵌入到特征图中, 因此提出直接使用 Canny 边缘提取算法^[23-24], 提取图像边缘特征到特征图中, 可增强图像重建的质量.

本文使用可逆神经网络结构和新的特征增强方案, 提出 ELRN. 特征增强方案包括边缘残差特征增强、小波映射特征增强和自注意力机制特征增强, 以提高重建图像对纹理轮廓特征的表达.

2 ELRN

2.1 网络结构

ELRN 分为全局特征提取模块(global feature extraction block, GFEB)和局部特征提取模块(local feature extraction block, LFEB)这 2 大部分. GFEB 中包含注意力特征提取层(attention feature extraction layer, AFEL); LFEB 中包含小波特征映射层(wavelet feature mapping layer, WFML)和深度特征提取层(deep feature extraction layer, DFEL). ELRN 整体架构如图 1 所示. 其中, $\oplus$ 表示残差运算.

从图 1 可以看出, 网络结构是可逆的, 其正向推理为特征提取, 逆向推理为图像重建, 即为传统非对称网络中的下上采样操作.

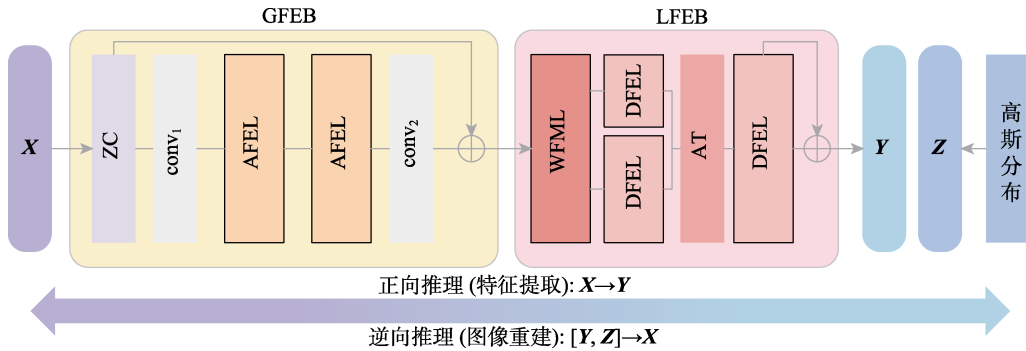


图 1 ELRN 整体架构

2.1.1 GFEB

GFEB 包含 AFEL, 并在前后使用 conv 操作调整隐层空间维度. GFEB 对数据 X 进行零均值化处理(zero-centered, ZC)并保存其均值 X_m , 使用 conv_1 将 ZC 后样本的隐层通道数提升至规定的特征嵌入维度并保存特征图 A , 提升维度旨在高维空间中找到易于多头自注意力机制标注的全局特征, 将转变维度后的特征图输入到 AFEL 中提取全局特征; 然后使用 conv_2 将其恢复到输入数据的原维度; 最后将输出与其均值进行残差求和得到全局特征提取结果 G . 该过程的映射关系可表示为

$$A = \text{conv}_1(\text{ZC}(X)) \quad (1)$$

$$G = \text{conv}_2([\text{AFEL}_0(A_0), \dots]) + X_m \quad (2)$$

2.1.2 LFEB

LFEB 包含 WFML 和 DFEL. LFEB 以提取的全局特征输出 G 作为输入, 通过 WFML 得到结果特征图 W . W 可拆分为 W^s 和 W^h , 前者保留原特征分布, 即主要特征; 后者在视觉中体现原特征分布的边缘特征. 分别将 W^s 和 W^h 代入到 DFEL 中进

行特征提取, 得到结果 y_1 ; 将 y_1 与 W^s 代入仿射变换(affine-transform, AT)得到结果 y_2 , 以充分利用原特征分布的映射特征图; 最后得到网络正向输出 Y . 该过程可以表示为 $W = \text{WFML}(G)$. 图像在经过小波映射后, 拆分出缩小维度的原特征图与映射特征图, 公式为

$$[W^s, W^h] = W \quad (3)$$

对式(3)进行特征提取, 得到输出结果

$$y_1 = \text{DFEL}(W^s) + \text{DFEL}(W^h).$$

AT 操作的结果为

$$y_2 = W^s \cdot e^{\text{Sigmoid}(S \cdot \text{DFEL}(y_1) - 1)} + \text{DFEL}(y_1).$$

AT 操作引入缩放因子最小因数 S 参与计算, 建立缩放因子与变换结果之间的直接联系. 最终拼接得到正向结果 $Y = [y_1, y_2]$.

2.2 正向特征提取

2.2.1 AFEB

为了使引入 Vision-Transformer 的空间维度可以适用于 SR 重建任务且仅用于特征提取, 将前向

传播中的平均池化和数据展平的操作重新调整, 以适用从词向量还原成输入特征图维度, 既可以

将注意力机制嵌入到特征提取中, 又不改变图像样本的原有维度. AFEB 结构如图 2 所示.

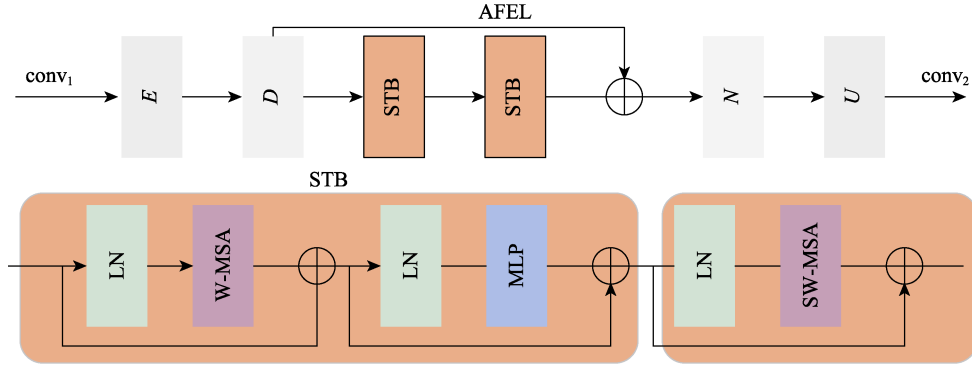


图2 AFEB 结构

式(1)(2)表示全局特征提取的过程. 对经过 conv_1 维度扩增后的特征图 X_{conv_1} 进行批量嵌入 E 和随机失活 D 操作, 得到 $A^{\text{emb}} = D(E(X_{\text{conv}_1}))$; 将其再代入到若干个 STB(swin Transformer block)中. 其包含层归一化(layer norm, LN)、多层感知机(multilayer perceptron, MLP)、窗口多头自注意力(window multi-head self-attention, W-MSA)和移动窗口的多头自注意力(shifted window multi-head self-attention, SW-MSA). 再将其输出特征图残差融合后进行归一化 N ; 最后数据样本进行批量反嵌入 U , 以恢复输入前的维度

$$A_{\text{STB}} = U \left(N \left(\left[\text{STB}_0(A_0^{\text{emb}}), \dots \right] \right) \right) + A^{\text{emb}}.$$

2.2.2 WFML

本文使用改进的 Harr 小波构建 WFML, 将不同的小波特征描述符构建成为方阵, 使用水平特征矩阵、竖直特征矩阵、交叉特征矩阵等作为映射卷积核进行卷积计算.

通过深入研究 2×2 的映射空间特征, 在保留原有小波特征映射的基础上, 提出新的小波特征核, 其结构如图 3 所示. 新的小波映射核与特征提取常用的细粒度卷积核维度相同. 其中, 2×2 特征核属于常规小波特征, 本文选取交叉、横向和纵向小波特征. 可以看出, 当缩放因子为 4 时, 使用 2 次 2×2 小波特征核进行卷积计算, 而之前此类任务的小波变换无法满足缩放因子为 3 或以 3 为最小因子的重建任务.

受到小波映射输出的限制, 原有的常规小波特征只能选择 3 种特征描述符, 所以图 3 中 2×2 常规特征核缺少交叉特征. 本文提出小波特征核 K_w 可同时描述更多特征, 并包含双向交叉特征.

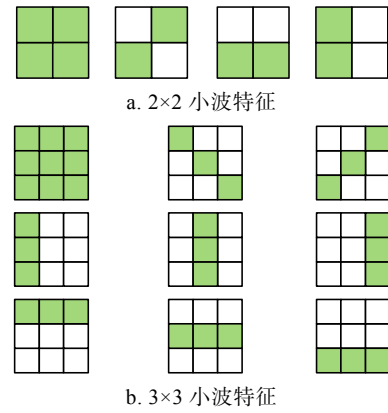


图3 小波特征核结构

通过小波特征核对特征图进行卷积计算, 输出特征图将会映射到不同小波特征的空间维度中. WFML 结构如图 4 所示. 其中, C_{fm} 表示特征图通道数, H_{fm} 表示特征图高度, W_{fm} 表示特征图宽度. 小波映射即特征图与小波特征核的卷积计算, 特征图被不同组的小波特征核分别映射到不同向量空间中, 输出特征图中包含原样本的映射特征图和其他被小波过滤到的少数特征. 将 GFEB 提取完的特征图 G 作为输入, 将 K_w 应用到卷积中, 得到特征图 $G' = K_w \otimes G$; 其中, $K_w \otimes G$ 表示以 K_w 为

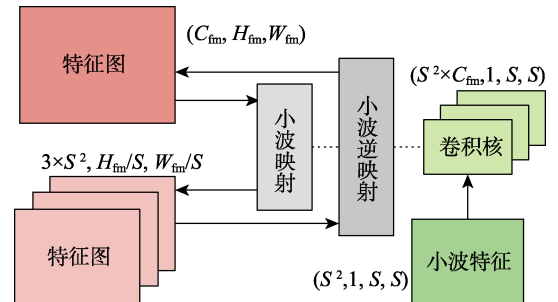


图4 WFML 结构

卷积核特征图 $\mathbf{G}$ 的卷积, $\mathbf{K}_w$ 表示所提小波特征核. 除将特征图映射到不同隐层空间之外, 这种细粒度分组卷积也可以进行下采样操作, 并将图像维度直接进行调整到指定缩放比例. 小波逆映射即特征图与小波特征核的反卷积计算, 将特征图上采样到输入特征图维度, 得到 $\mathbf{G}'_i = \mathbf{K}_w \hat{\otimes} \mathbf{G}'$; 其中, $\hat{\otimes}$ 表示反卷积.

2.2.3 DFEL

使用稠密网络结构对 WFML 结果特征图进行

卷积计算, 对输出通道统一处理, 并结合所有数据残差在最大限度地提取卷积计算过程中学习到的特征; 然后使用 ReLU 激活函数进行非线性激活处理, 输出特征提取结果. DFEL 结构如图 5 所示, 其中, 使用稠密网络结构累积迭代映射特征, 输出提取到的特征图 $\mathbf{D}_i$; 然后将 $\mathbf{D}_i$ 代入到边缘特征残差, 使用边缘特征提取算法提取边缘特征, 再与稠密卷积计算结果残差运算, 得到 DFEL 的最终结果 $\mathbf{D}$. 边缘特征提取及残差如图 6 所示.

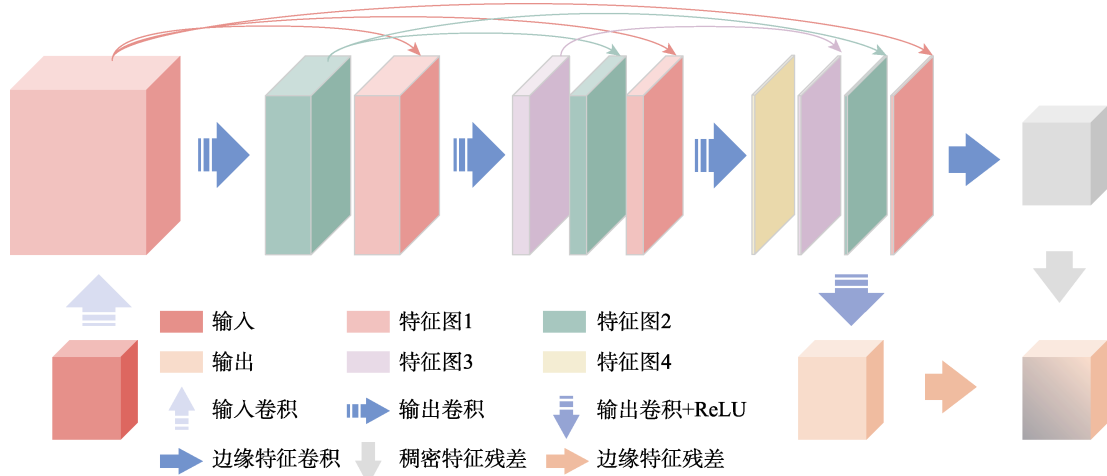


图 5 DFEL 结构

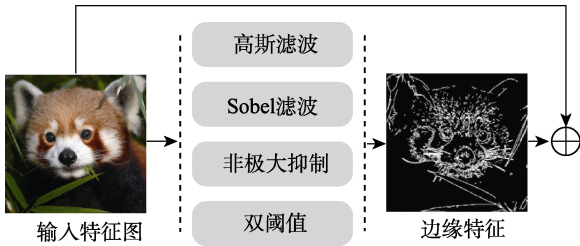


图 6 边缘特征提取及残差

将经过稠密结构特征提取后的 3 通道特征图输出结果进行边缘特征处理, 得到输出 $\mathbf{D}_i^e$ 并依次进行高斯降噪; 然后代入 Sobel 滤波器; 最后采用非最大抑制和边缘细化以双阈值限定输出, 将提取到的边缘特征以线性参数的形式代入到网络整体参数传播中. 深度边缘特征残差的输出结果可以表示为 $\mathbf{D} = [\mathbf{D}_0 + (\mathbf{W}_0 \mathbf{D}_0^e + \mathbf{b}_0), \dots]$. 其中, $\mathbf{W}$ 表示权重; $\mathbf{b}$ 表示偏置. 高斯降噪以横纵向对原样本进行卷积计算, 过滤掉随机分布和服从正态分布的噪声. 高斯函数可以表示为

$$G_{\text{dist}}(h, v) = \frac{1}{2\pi\sigma^2} e^{-\frac{h^2 + v^2}{2\sigma^2}} \quad (4)$$

Sobel 滤波器的本质是基于图像空间域进行卷

积计算, 平滑抑制噪声. 定义横向 Sobel 算子为

$$\mathbf{S}_h = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix} \quad (5)$$

纵向 Sobel 算子 $\mathbf{S}_v$ 是 $\mathbf{S}_h$ 的转置. 以式(4)(5)作为权重对样本进行横纵向的卷积计算, 得到横纵向的边缘梯度 $\mathbf{D}_i^{e_h}$ 与 $\mathbf{D}_i^{e_v}$. 其公式分别为

$$\mathbf{D}_i^{e_h} = \mathbf{S}_h \otimes (G_{\text{dist}}(h, 1) \otimes \mathbf{D}_i),$$

$$\mathbf{D}_i^{e_v} = \mathbf{S}_v \otimes (G_{\text{dist}}(1, v) \otimes \mathbf{D}_i).$$

其中, $\otimes$ 表示卷积核特征图的卷积计算. 然后在梯度方向上找出相邻像素点进行边缘细化; 最后使用双阈值 t_1 与 t_2 输出布尔特征图, 边缘特征过滤函数可以表示为

$$E(x) = \begin{cases} 1, & t_1 < x < t_2 \\ 0, & \text{otherwise} \end{cases}.$$

综上, $\mathbf{D}_i^e$ 可以表示为

$$\mathbf{D}_i^e = E\left(\sqrt{(\mathbf{D}_i^{e_h})^2 + (\mathbf{D}_i^{e_v})^2}\right).$$

2.3 逆向图像重建

已有的深度 SR 重建任务中, 上采样重建过程

通常需要额外设计新的网络模块或分支等, 而 ELRN 通过小波逆映射实现图像重建.

在计算过程中, 所有的正向四则运算在逆向推理中都要执行其相反运算, 依照图 1 逆向推理与图 4 中的小波逆向映射进行图像重建.

如图 1 的逆向过程所示, 以样本图像在正向输出结果 $\mathbf{Y}$ 为网络逆向输入特征, 公式为

$$[\mathbf{y}_1, \mathbf{y}_2] = \mathbf{Y}.$$

并从高斯分布中采样得到的相同维度噪声 $\mathbf{Z}$ 作为网络逆向推理的输入, 公式为

$$\hat{\mathbf{y}}_2 = \frac{\mathbf{Z} - \text{DFEL}(\mathbf{y}_1)}{e^{\text{Sigmoid}(S \cdot \text{DFEL}(\mathbf{y}_1) - 1)}}, \hat{\mathbf{y}}_1 = \mathbf{y}_1 - \text{DFEL}(\hat{\mathbf{y}}_2).$$

其中, $\hat{\cdot}$ 表示逆. 可逆神经网络结构中的逆向运算常将 $\mathbf{y}_2$ 隐式替换为高斯随机采样代入运算, 使特征分布中非重要特征进行部分回归预测, 以增强网络预测能力. 如图 4 的逆向过程所示, 经过 WFML 时, 使用小波逆映射进行逆向采样最终完成图像重建, 公式为 $\mathbf{X} = \text{WFML}([\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2])$.

3 实验及结果分析

3.1 实验设置

本文实验环境为 Intel Xeon Platinum 8358, RTX 3090, Ubuntu20.04, Python3.8, Pytorch1.11.0 和 CUDA11.3.

采用 SR 重建研究中的公共数据集 DIV2K^[25] 作为训练集; Set5^[26], Set14^[27], BSD100^[28] 和 Urban100^[29] 作为验证集和测试集.

ELRN 在前向模型数据推理中使用 L_2 损失, 以保证下采样的样本质量保持光滑稳定. 在逆向推理中使用 L_1 损失和 ES Loss, 其中, L_1 损失用于衡量图像重建质量; ES Loss 替代可逆神经网络中的 MMD 损失, 同样用于衡量特征分布之间的距离, 解决其无法指导参数向超越原图质量方向更新的不足. (1) 原标签按通道嵌入边缘特征得到新特征分布, 拉开特征分布之间的距离; (2) 计算最大均值差异, 与 MMD 不同, 此映射过后的核函数存在更多 1 组不同项, 且原特征分布处于新特征分布与输出特征分布之间, 避免产生同源问题的同时, 亦可以指导模型参数向更多边缘特征方向更新迭代. ES Loss 和总损失分别可以表示为

$$\begin{aligned} \text{ES Loss} &= E(G_f(x, x)) + E(G_f(y, y)) - \\ &E(G_f(x, y)) - E(G_f(y, x)), \end{aligned}$$

$$L_{\text{total}} = \lambda_1 L_1 + \lambda_2 L_2 + \lambda_3 \text{ES Loss}.$$

其中, E 表示期望, G_f 表示高斯滤波器.

实验采用自适应动量估计优化器 (adaptive momentum estimation, ADAM), 初始化学率学习率为 2×10^{-4} , 每 100 轮次学习率减半; batchsize 设置为 16. 损失函数权重 λ_i 的默认值为 S^2 , 1 和 1.

本文采用图像 SR 重建中常用的峰值信噪比 (peak signal-to-noise ratio, PSNR) 与结构相似度 (structural similarity index measure, SSIM) 作为评价指标. 其中, PSNR 用于测量原始图像与重建图像的差异性, 即比较像素最大可能值与原始图像和重建图像之间的均方误差, 取值范围为 1~100, 单位为 dB. SSIM 基于 2 个图像的内容来衡量相似性, 同时考虑图像中存在的亮度和颜色信息, 取值范围为 -1~1, 其值越接近 1 表示图像之间的相似度越高.

3.2 对比实验

将 ELRN 与其他深度 SR 重建模型进行对比实验, 包括 SRCNN^[6], VDSR^[7], CARN^[30], IMDN^[31], LatticeNet^[32], FDIWN^[33], LBNet^[34] 和 SwinIR-light^[35]. 分别以缩放因子 2, 3 和 4 进行对比, 结果如表 1 所示. 可以看出, 在缩放因子为 4 时, ELRN 以近似 FDIWN^[33] 的参数量得到了更好的重建结果, 且与参数量较高的 LBNet^[34] 相比, ELRN 表现出更好的重建结果; 缩放因子为 2, 3 时, ELRN 使用更少参数取得更好的重建结果.

为了验证模型在特殊情况下的重建能力. 以光线、物体多样化为前提, 分别将代表对光线充足、光线不足、颜色复杂多变的现实情况图像作为对比样本. 选取 Urban100 数据集集中的 img_049, img_012 和 img_081, 选取长、宽为 72×72 的样本即 HR, 对其进行缩放因子为 4 的插值下采样, 得到长、宽为 18×18 输入样本即 LR. 将 HR 与 LR 代入 VDSR^[7], CARN^[30], IMDN^[31], LatticeNet^[32], SwinIR-light^[35], FDIWN^[33], LBNet^[34] 和 ELRN 中进行实验, 结果如图 7 所示. 可以看出, 在复杂环境样本的实验中, ELRN 重建的视觉效果具有更强的纹理轮廓特征, 结果更加立体, 表现在图中书架的层次结构、光线的轮廓波动、车站站台字母的重建等.

进一步, 在模型性能上, 将 VDSR^[7], CARN^[30], SwinIR-light^[35], FDIWN^[33], LBNet^[34] 与 ELRN 进行对比, 在缩放因子为 4 的测试集 Set5 上, 对参数量与模型推理速度和性能的关系进行实验, 结果如图 8 所示. 可以看出, ELRN 的推理时间并非最好,

表 1 不同缩放因子下 9 种模型重建能力对比

方法	缩放因子	参数量	Set5		Set14		BSD100		Urban100	
			PNSR/dB	SSIM	PNSR/dB	SSIM	PNSR/dB	SSIM	PNSR/dB	SSIM
Bicubic			33.66	0.929 9	30.24	0.868 8	29.56	0.843 1	26.88	0.840 3
SRCNN ^[6]		57k	36.57	0.954 6	32.38	0.906 1	31.25	0.887 6	29.36	0.894 7
VDSR ^[7]		665k	37.47	0.959 7	33.08	0.913 7	31.88	0.897 2	30.78	0.914 9
CARN ^[30]		1 592k	37.87	0.960 6	33.55	0.917 7	32.15	0.899 7	31.97	0.926 8
IMDN ^[31]	×2	694k	37.99	0.961 1	33.62	0.918 0	32.20	0.900 7	32.18	0.928 6
LatticeNet ^[32]		756k	38.05	0.961 3	33.69	0.920 0	32.22	0.901 0	32.25	0.929 1
FDIWN ^[33]		629k	38.05	0.961 3	33.70	0.919 7	32.22	0.901 0	32.23	0.929 2
LBNet ^[34]		731k	38.06	0.961 3	33.63	0.917 9	32.18	0.900 4	32.29	0.929 0
SwinIR-light ^[35]		910k	38.12	0.961 5	33.84	0.920 6	32.32	0.902 1	32.78	0.934 2
ELRN(本文)		425k	38.21	0.969 5	35.07	0.952 4	35.03	0.956 6	32.24	0.947 8
Bicubic			30.39	0.868 2	27.55	0.774 2	27.21	0.738 5	24.46	0.734 9
SRCNN ^[6]		57k	32.68	0.909 5	29.25	0.820 6	28.33	0.785 9	26.05	0.795 9
VDSR ^[7]		665k	33.70	0.923 1	29.90	0.834 6	28.83	0.799 9	27.18	0.829 8
CARN ^[30]		1 592k	34.31	0.927 1	30.30	0.842 0	29.08	0.805 9	28.07	0.850 4
IMDN ^[31]	×3	703k	34.35	0.927 4	30.32	0.842 2	29.10	0.806 2	28.16	0.852 2
LatticeNet ^[32]		764k	34.39	0.927 8	30.31	0.842 0	29.10	0.806 5	28.19	0.851 5
FDIWN ^[33]		646k	34.50	0.928 5	30.37	0.843 2	29.13	0.807 7	28.27	0.854 9
LBNet ^[34]		736k	34.47	0.928 3	30.37	0.842 2	29.13	0.807 6	28.42	0.856 1
SwinIR-light ^[35]		918k	34.61	0.929 1	30.53	0.846 7	29.21	0.809 8	28.67	0.862 7
ELRN(本文)		489k	34.56	0.936 7	31.46	0.896 9	30.99	0.887 3	28.42	0.879 0
Bicubic			28.42	0.810 4	26.00	0.702 7	25.96	0.667 5	23.14	0.657 7
SRCNN ^[6]		57k	30.24	0.860 5	27.43	0.749 3	26.80	0.709 4	24.29	0.714 1
VDSR ^[7]		665k	31.42	0.885 9	28.15	0.771 1	27.28	0.727 4	25.22	0.754 5
CARN ^[30]		1 592k	32.13	0.895 6	28.59	0.781 7	27.59	0.737 5	26.06	0.784 6
IMDN ^[31]	×4	715k	32.18	0.895 4	28.57	0.781 5	27.56	0.737 0	26.03	0.783 9
LatticeNet ^[32]		777k	32.17	0.895 0	28.60	0.781 5	27.57	0.737 2	26.13	0.784 5
FDIWN ^[33]		664k	32.21	0.896 1	28.62	0.782 9	27.60	0.738 9	26.19	0.789 5
LBNet ^[34]		742k	32.27	0.896 6	28.66	0.783 6	27.62	0.739 9	26.26	0.790 7
SwinIR-light ^[35]		930k	32.42	0.898 2	28.74	0.785 9	27.69	0.742 4	26.47	0.798 2
ELRN(本文)		686k	32.83	0.906 0	29.67	0.835 0	29.04	0.810 2	26.73	0.816 8

注. 粗体表示最优值.

但也是可以接受的,且在相近数量级的参数量下,该模型的执行时间与性能达到良好的平衡.选取缩放因子为 4 时模型参数量并不占优,但重建能力要好于 LBNet^[34], FDIWN^[33]和 SwinIR-light^[35].

3.3 消融实验

为了验证 ELRN 中多种轻量特征与小波映射的有效性,对 GFEB, LFEB 中的 WFML,以及 DFEL 中的边缘特征残差和 ES Loss 进行消融实验.

本文设计以下消融实验:将 GFEB 从网络中去除,消除全局特征的提取;引入除常规 Harr 小

波以外的 2 种小波映射: Daubechies 小波和双正交小波^[21-22],并将其修改为普通维度映射,研究上述情况与所提 WFEL 对模型重建能力的影响;将 DFEL 的边缘特征残差和 ES Loss 去除.在缩放因子为 4 时以测试集 Set5 验证重建结果,实验结果如表 2 所示.不同组消融的重建结果如图 9 所示.从表 2 中可以看出: (1) 常规 Harr 小波表现最为稳定,其稳定性在一定程度上高于 ELRN,但上限有限,结果并不达标; Daubechies 小波与双正交小波表现相当,都在 8 000 次迭代处存在指标下降,结果并

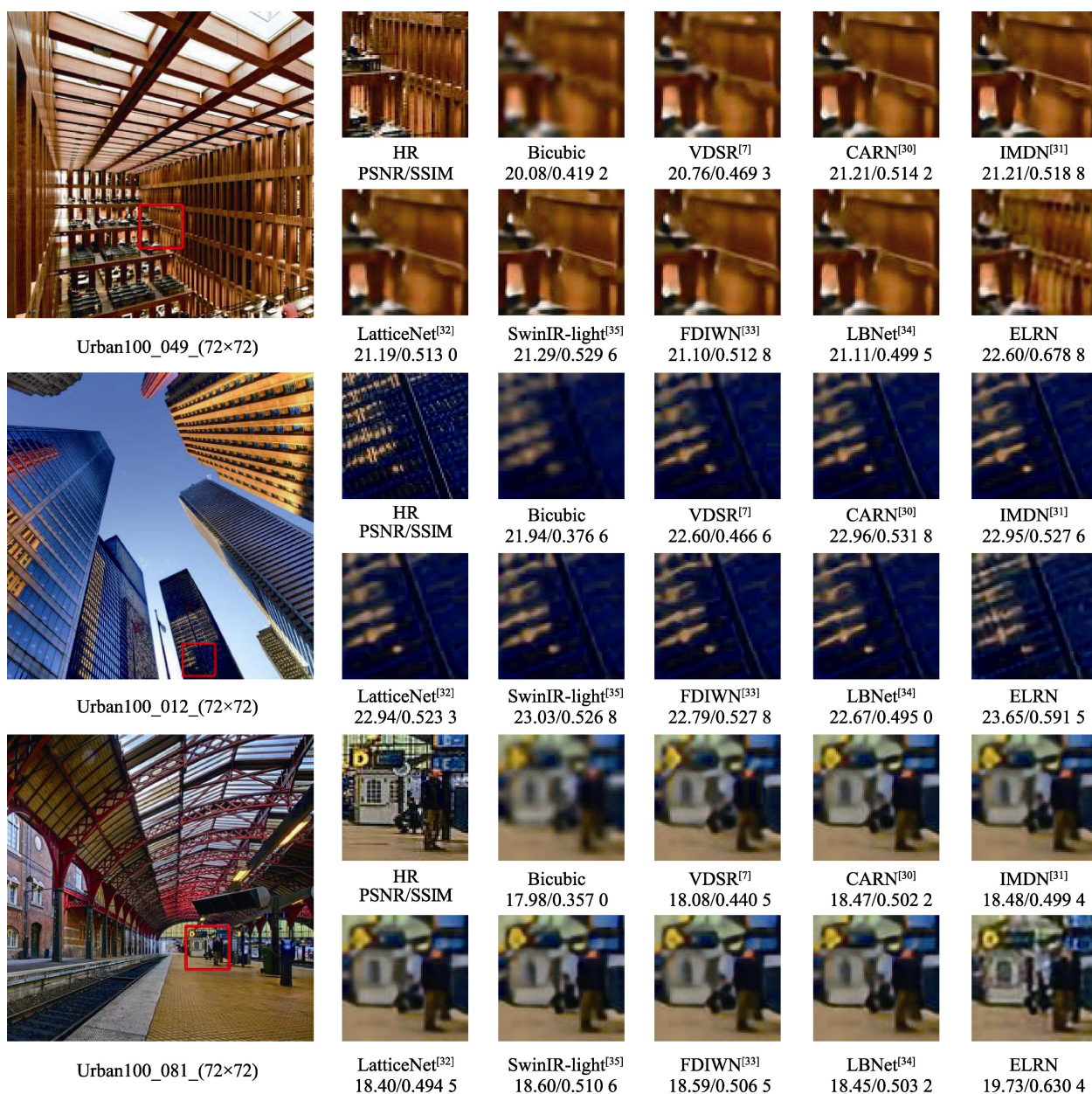


图 7 9 种模型在复杂环境下的重建结果对比

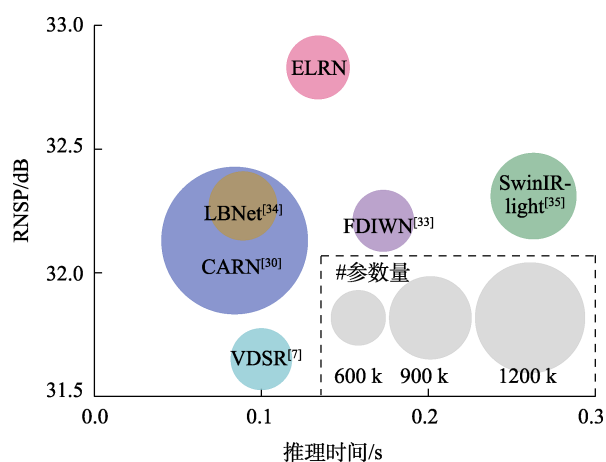


图 8 6 种模型性能对比

不理想. 这些结果也与理论相符合, 即具有完整对称性的小波在特征提取过程中可以有效地避免相位畸变和多余噪声产生. (2) 将 WFML 替换后, 降低了 53.24% 的重建能力; 去除 GFEB 时, 虽然参数有略微的降低, 但重建能力降低 1.28%, 这也证明全局特征的提取只增加额外 27.5k 的参数量; 将边缘特征与 ES Loss 移除后, 其重建能力降低了 2.35%.

实验结果表明, 在与 3 种常用小波的对比中, ELRN 的最终结果更好, 但稳定性不如常规 Harr 小波; 缺少 GFEB 的整体效果变差, 在模型即将收敛时, 产生偏低的异常点; 当移除边缘特征与 ES Loss 后, 性能指标上升不稳定, 常有下降趋势.

表 2 缩放因子为 4 时 Set5 测试集的消融实验对比

GFEB		LFEB				参数量	PSNR/dB	SSIM	
GFEB	GFEB*	小波	WFML		DFEL				
			WFML	WFML*	DFEL	DFEL*			
√		本文	√		√		685.8k	32.83	0.906 0
	√	本文	√		√		658.3k	32.41	0.893 8
√		Harr	√		√		685.6k	32.65	0.904 3
√		Daubechies	√		√		685.6k	32.58	0.899 1
√		双正交	√		√		685.6k	32.54	0.904 0
√				√	√		685.6k	15.35	0.411 4
√		本文	√			√	683.2k	32.06	0.896 4

注. GFEB*表示去除对应模块; WFML*表示使用卷积映射代替小波映射; DFEL*表示去除对应层; 本文表示所提小波映射.

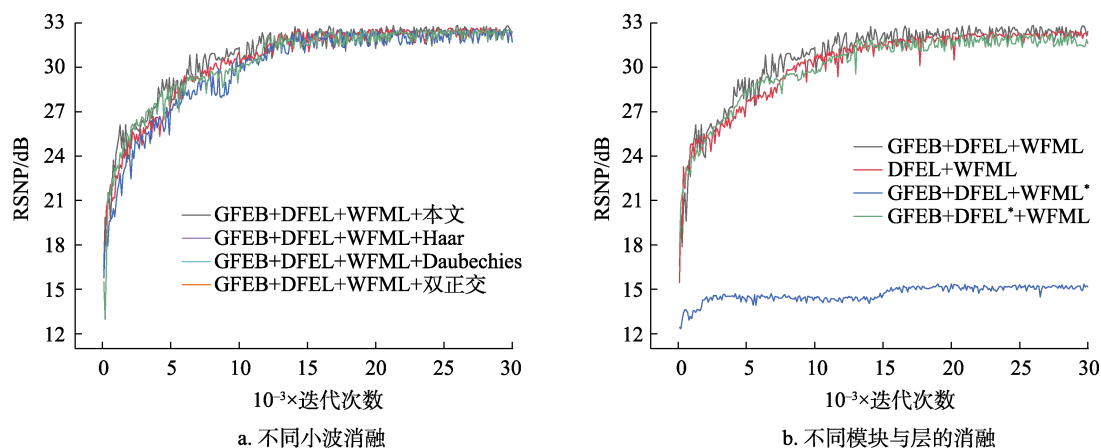


图 9 不同组消融的重建结果

4 结 语

针对深度 SR 模型参数量大、推理速度慢,而轻量化深度 SR 模型重建图像质量低,无法更好地重建高清图案的纹理轮廓特征的问题,本文提出基于特征增强的轻量化可逆 SR 网络 ELRN. 首先提出边缘特征残差及相应损失,加强网络对图像轮廓特征的表达;然后补充新的小波特征核,使小波变换支持任意缩放因子;最后嵌入自注意力机制提取全局特征. 实验结果表明,在复杂环境条件的图样比对中,ELRN 的重建结果可以清晰地辨别纹理特征.

但是,ELRN 仍存在一定局限性,这是由可逆神经网络结构对原空间分布的依赖所决定的,使得其更适用于网络传输类的应用场景. 未来,将优化可逆神经网络结构并寻找可学习的小波映射核.

参考文献(References):

- [1] Hiji M, Khan A, Alwakeel M M, *et al.* Intelligent image super-resolution for vehicle license plate in surveillance applications[J]. *Mathematics*, 2023, 11(4): Article No.892
- [2] Li Y, Sixou B, Peyrin F. A review of the deep learning methods for medical images super resolution problems[J]. *IRBM*, 2021, 42(2): 120-133
- [3] Wang P J, Bayram B, Sertel E. A comprehensive review on deep learning based remote sensing image super-resolution methods[J]. *Earth-Science Reviews*, 2022, 232: Article No.104110
- [4] Yang W M, Zhang X C, Tian Y P, *et al.* Deep learning for single image super-resolution: a brief review[J]. *IEEE Transactions on Multimedia*, 2019, 21(12): 3106-3121
- [5] Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16x16 words: transformers for image recognition at scale[OL]. [2023-06-15]. <https://arxiv.org/abs/2010.11929>
- [6] 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
- [7] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2016: 1646-1654
- [8] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2016: 770-778
- [9] Lim B, Son S, Kim H, *et al.* Enhanced deep residual networks

- for single image super-resolution[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Los Alamitos: IEEE Computer Society Press, 2017: 1132-1140
- [10] Ledig C, Theis L, Huszar F, *et al.* Photo-realistic single image super-resolution using a generative adversarial network[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 105-114
- [11] Yang F Z, Yang H, Fu J L, *et al.* Learning texture transformer network for image super-resolution[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 5790-5799
- [12] Chen H T, Wang Y H, Guo T Y, *et al.* Pre-trained image processing transformer[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 12294-12305
- [13] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[C] //Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Heidelberg: Springer, 2015: 234-241
- [14] Kingma D P, Welling M. Auto-encoding variational bayes[OL]. [2023-06-15]. <https://arxiv.org/abs/1312.6114>
- [15] Gomez A N, Ren M Y, Urtasun R, *et al.* The reversible residual network: backpropagation without storing activations[OL]. [2023-06-15]. <https://arxiv.org/abs/1707.04585>
- [16] Huang G, Liu Z, Van Der Maaten L, *et al.* Densely connected convolutional networks[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 2261-2269
- [17] Ardizzone L, Kruse J, Rother C, *et al.* Analyzing inverse problems with invertible neural networks[OL]. [2023-06-15]. <https://arxiv.org/abs/1808.04730>
- [18] Borgwardt K M, Gretton A, Rasch M J, *et al.* Integrating structured biological data by kernel maximum mean discrepancy[J]. *Bioinformatics*, 2006, 22(14): e49-e57
- [19] Xiao M Q, Zheng S X, Liu C, *et al.* Invertible image re-scaling[C] //Proceedings of the 16th European Conference on Computer Vision. Heidelberg: Springer, 2020: 126-144
- [20] Jing J P, Deng X, Xu M, *et al.* HiNet: deep image hiding by invertible network[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 4713-4722
- [21] Li Q F, Shen L L, Guo S, *et al.* Wavelet integrated CNNs for noise-robust image classification[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 7245-7254
- [22] Li Q F, Shen L L, Guo S, *et al.* WaveCNet: wavelet integrated CNNs to suppress aliasing effect for noise-robust image classification[J]. *IEEE Transactions on Image Processing*, 2021, 30: 7074-7089
- [23] Zhang C X, Zhang N, Yu W, *et al.* Improved canny-based algorithm for image edge detection[C] //Proceedings of the 36th Youth Academic Annual Conference of Chinese Association of Automation (YAC). Los Alamitos: IEEE Computer Society Press, 2021: 678-683
- [24] Canny J. A computational approach to edge detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, PAMI-8(6): 679-698
- [25] Agustsson E, Timofte R. NTIRE 2017 challenge on single image super-resolution: dataset and study[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Los Alamitos: IEEE Computer Society Press, 2017: 1122-1131
- [26] 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. Guildford: BMVA Press, 2012: 135.1-135.10
- [27] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations[C] //Proceedings of the 7th International Conference on Curves and Surfaces. Heidelberg: Springer, 2012: 711-730
- [28] 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
- [29] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 5197-5206
- [30] Ahn N, Kang B, Sohn K A. Fast, accurate, and lightweight super-resolution with cascading residual network[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 256-272
- [31] Hui Z, Gao X B, Yang Y C, *et al.* Lightweight image super-resolution with information multi-distillation network[C] //Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM Press, 2019: 2024-2032
- [32] Luo X T, Xie Y, Zhang Y L, *et al.* LatticeNet: towards lightweight image super-resolution with lattice block[C] //Proceedings of the 16th European Conference on Computer Vision. Heidelberg: Springer, 2020: 272-289
- [33] Gao G W, Li W J, Li J C, *et al.* Feature distillation interaction weighting network for lightweight image super-resolution[C] //Proceedings of the 36th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2022 : 661-669
- [34] Gao G W, Wang Z X, Li J X, *et al.* Lightweight bimodal network for single-image super-resolution via symmetric CNN and recursive transformer[OL]. [2023-06-15]. <https://arxiv.org/abs/2204.13286>
- [35] Liang J Y, Cao J Z, Sun G L, *et al.* SwinIR: image restoration using Swin transformer[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Los Alamitos: IEEE Computer Society Press, 2021: 1833-1844