第 50 卷 第 1 期 2024 年 1 月
自 动 化 学 报
ACTA AUTOMATICA SINICA
Vol. 50, No. 1 January, 2024

基于像素对比学习的图像超分辨率算法

周登文 [1] 刘子涵 [1] 刘玉铠 [1]

摘 要 目前, 深度卷积神经网络 (Convolutional neural network, CNN) 已主导了单图像超分辨率 (Single image super-resolution, SISR) 技术的研究, 并取得了很大进展. 但是, SISR 仍是一个开放性问题, 重建的超分辨率 (Super-resolution, SR) 图像往往会出现模糊、纹理细节丢失和失真等问题. 提出一个新的逐像素对比损失, 在一个局部区域中, 使 SR 图像的像素尽可能靠近对应的原高分辨率 (High-resolution, HR) 图像的像素, 并远离局部区域中的其他像素, 可改进 SR 图像的保真度和视觉质量. 提出一个组合对比损失的渐进残差特征融合网络 (Progressive residual feature fusion network, PRFFN). 主要贡献有: 1) 提出一个通用的基于对比学习的逐像素损失函数, 能够改进 SR 图像的保真度和视觉质量; 2) 提出一个轻量的多尺度残差通道注意力块 (Multi-scale residual channel attention block, MRCAB), 可以更好地提取和利用多尺度特征信息; 3) 提出一个空间注意力融合块 (Spatial attention fuse block, SAFB), 可以更好地利用邻近空间特征的相关性. 实验结果表明, PRFFN 显著优于其他代表性方法.

关键词 图像超分辨率, 卷积神经网络, 对比学习, 注意力机制

引用格式 周登文, 刘子涵, 刘玉铠. 基于像素对比学习的图像超分辨率算法. 自动化学报, 2024, **50**(1): 181−193

DOI 10.16383/j.aas.c230395

Pixel-wise Contrastive Learning for Single Image Super-resolution

ZHOU Deng-Wen[1] LIU Zi-Han[1] LIU Yu-Kai[1]

Abstract Deep convolutional neural network (CNN) has achieved great success in single image super-resolution (SISR). However, SISR is still an open issue, and reconstructed super-resolution (SR) images often suffer from blurring, loss of texture details and distortion. In this paper, a new pixel-wise contrastive loss is proposed to improve the fidelity and visual quality of SR images by making the pixels of SR images as close as possible to the corresponding pixels of the original high-resolution (HR) images and away from the other pixels in the local region. We also propose a progressive residual feature fusion network (PRFFN) with combined contrastive loss, and the main contributions include: 1) A general pixel-wise loss function based on contrastive learning is proposed, which can improve the fidelity and visual quality of SR images; 2) A lightweight multi-scale residual channel attention block (MRCAB) is proposed, which can better extract and utilize multi-scale feature information; 3) A spatial attention fusion block (SAFB) is proposed, which can better utilize the correlation of neighboring spatial features. The experimental results demonstrate that PRFFN significantly outperforms other representative methods.

Key words Image super-resolution, convolutional neural network (CNN), contrastive learning, attention mechanism

Citation Zhou Deng-Wen, Liu Zi-Han, Liu Yu-Kai. Pixel-wise contrastive learning for single image super-resolution. *Acta Automatica Sinica*, 2024, **50**(1): 181−193

单图像超分辨率 (Single image super-resolution, SISR)[1] 是计算机视觉中一个基本任务, 旨在从低分辨率 (Low-resolution, LR) 图像, 恢复出对应的高分辨率 (High-resolution, HR) 图像, 在诸如遥感成像[2]、视频监控[3] 和医学成像[4] 中, 应用广泛. SISR 是一个病态的逆问题, 因为许多 HR 图像可退化为相同的 LR 图像, 需要提供图像的先验知识, 约束超分辨率 (Super-resolution, SR) 图像的解空间. SISR 仍然是计算机视觉中开放性的研究问题, 重建的 SR 图像往往会出现模糊、纹理细节丢失和失真等问题.

早期的 SISR 是基于插值的方法, 如双线性插值和双三次 (Bicubic) 插值. 基于插值的方法仍被广泛使用, 具有很低的计算复杂度, 但不能恢复 LR 图像中丢失的图像细节. 基于实例学习的方法[5−7] 旨在通过训练图像, 学习 LR 图像和 HR 图像之间的映射关系, 改进了基于插值的方法. 但是, 基于实例学习的方法往往优化困难, 并具有较高的推理复杂度. 目前, 深度卷积神经网络 (Convolutional neural

收稿日期 2023-06-27 录用日期 2023-10-15
Manuscript received June 27, 2023; accepted October 15, 2023
本文责任编委 陈胜勇
Recommended by Associate Editor CHEN Sheng-Yong
1. 华北电力大学控制与计算机工程学院 北京 102206
1. School of Control and Computer Engineering, North China Electric Power University, Beijing 102206

networks, CNN) 技术[8−13] 直接端到端地学习 LR 和 HR 图像之间的映射关系, 显著提高了 SISR 性能, 并主导了目前 SISR 技术的研究. 基于 CNN 的 SISR 方法研究主要是探索新的 SR 网络架构, 损失函数广泛使用均方误差 (Mean squared error, MSE)[12] 和平均绝对误差 (Mean absolute error, MAE)[8], 但这些传统的逐像素损失生成的 SR 图像是潜在 SR 输出图像的平均[14], 导致输出的图像被过度平滑. 为了解决这个问题, Johnson 等[15] 提出感知损失. 感知损失不是在图像空间度量逐像素的损失, 而是在预训练的 VGG (Oxford visual geometry group) 网络[16] 特征空间度量逐像素损失. 感知损失能改进 SR 图像的感知质量, 但降低了 SR 图像保真度. 重要的是, 感知损失也不能阻止 SR 图像的模糊. Wang 等[17] 提出一个对比自蒸馏 (Contrastive self-distillation, CSD) 网络, 引入一个基于对比学习的损失函数. CSD 损失与感知损失类似, 也在预训练 VGG 网络的特征空间逐像素比较损失. 以教师子网络输出的 SR 图像作为正样本, 学生子网络输出的 SR 图像作为锚, 从同一个批次中采样 K 个图像 (除锚外), 通过双三次上采样到与输出 SR 图像相同的分辨率作为负样本. CSD 损失使锚更靠近正样本, 并远离负样本, 进一步改进了学生子网络输出的 SR 图像的视觉质量. 与感知损失相比, CSD 损失除限制了学生子网络输出的上界 (正样本) 外, 也限制了下界 (负样本), 以减小解空间. 但 CSD 损失也有与感知损失类似的保真度低问题. 另外, CSD 损失使用双三次上采样图像作为负样本, 是一个较弱的下界. 受 CSD 启发, 提出一个新的基于对比学习的逐像素损失函数 L_{cntr}. L_{cntr} 作用在图像空间, 原 HR 图像与输出的 SR 图像对应的像素分别作为正样本和锚, HR 图像上正样本邻近的像素作为负样本. 类似地, L_{cntr} 也让锚更靠近正样本, 远离负样本. 常用损失及其组合的峰值信噪比 (Peak signal to noise ratio, PSNR)[18]、结构相似性 (Structural similarity, SSIM)[18] 和视觉效果见图 1, 其中 L_1 为 MAE 损失, L_{perc} 为感知损失[15], L_{tex} 为纹理损失[19], L_{CSD} 为 CSD 损失[17], L_{adv} 为对抗损失[20]. 原 L_{CSD} 正样本是教师网络输出的 SR 图像, 本文替换为 HR 图像. 计算 L_{tex} 的图像块大小为 48×48 像素. 可以看出, 本文的逐像素损失函数 L_{cntr} 可与其他损失组合使用, 显著改进 SR 图像保真度和视觉质量. 另外, 本文提出一个新的网络架构, 称为渐进残差特征融合网络 (Progressive residual feature fusion network, PRFFN). PRFFN 应用扩张卷积[21], 扩展极深残差通道注意力网络 (Very deep residual channel attention network, RCAN)[11] 的基本构件 —— 残差通道注意力块 (Residual channel attention block, RCAB)[11], 不增加参数并融合多尺度特征, 称为多尺度残差通道注意力块 (Multi-scale residual channel attention block, MRCAB). PRFFN 以 MRCAB 为基本构件, 并运用空间注意力机制[22], 渐进地融合 MRCAB 的输出特征. 本文的 L_{cntr} 损失与 PRFFN 网络架构组合, 能够获得先进的 SR 性能, 一些代表性方法的 PSNR 和参数量见图 2. L_{cntr} 是通用的, 可与其他 SR 网络架构协作使用.

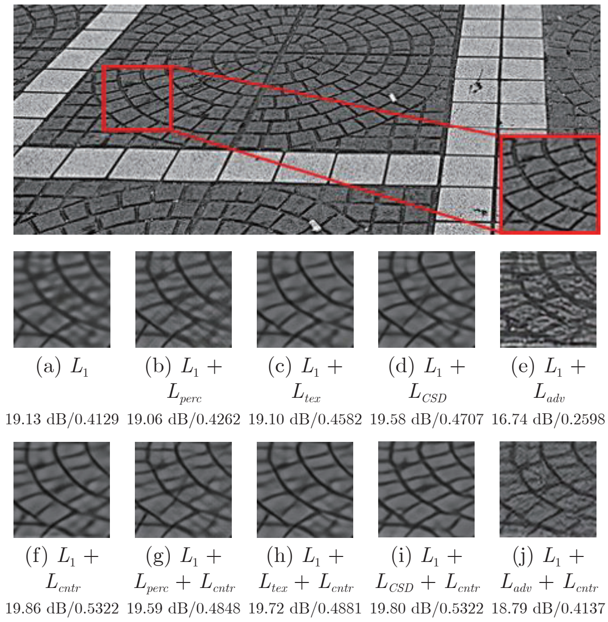

图 1 不同损失及其组合的 PSNR/SSIM 和视觉效果

Fig. 1 PSNR/SSIM and visual effects for different losses and their combinations

本文的主要贡献有: 1) 提出一个通用的基于对比学习的逐像素损失函数 L_{cntr}, 能够显著改进 SR

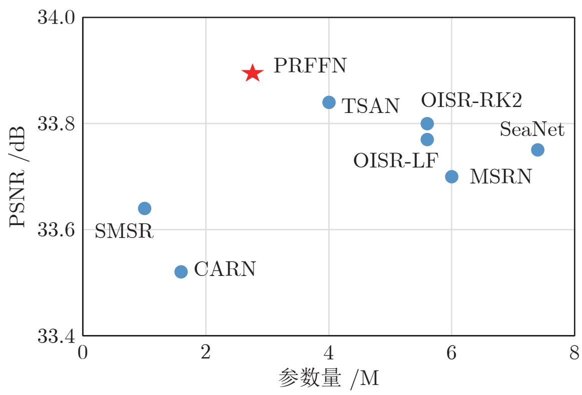

图 2 在 Set14 数据集上, 不同 SISR 方法 2 倍 SR 结果的平均 PSNR 值和参数量

Fig. 2 Average PSNRs and parameter counts for 2 times SR models for each state-of-the-art SISR method on the Set14 dataset

图像的视觉质量; 2) 提出一个新的 SR 网络架构 PRFFN, 主要组件是 MRCAB 和空间注意力融合块 (Spatial attention fuse block, SAFB), MRCAB 可以更好地提取和利用多尺度特征信息, 而 SAFB 可以更好地利用邻近特征的相关性; 3) 实验结果表明, PRFFN 组合 L_{cntr} 取得了有竞争力的 SR 性能.

1 相关工作

1.1 图像超分辨率的网络架构

2015 年, Dong 等[8] 提出第 1 个基于 CNN 的 SISR 方法 —— 超分辨率卷积神经网络 (Super-resolution convolutional neural network, SRCNN) 后, 深度卷积神经网络技术已主导了当前 SR 技术研究. SRCNN 只有 3 个卷积层, Kim 等[9] 提出极深超分辨率网络 (Very deep super-resolution network, VDSR), 通过引入残差学习, 加深了网络 (有 20 个卷积层), 改进了性能. Lim 等[10] 提出增强的深度超分辨率网络 (Enhanced deep super-resolution network, EDSR), 通过堆积残差块, 进一步加深了网络 (有 65 个卷积层), 改进了性能. Zhang 等[11] 提出 RCAN 方法, 在残差块中引入通道注意力 (Channel attention, CA) 机制, 有超过 400 个卷积层, 显著改进了 SISR 方法性能. Tong 等[12] 提出超分辨率稠密连接网络, 使用了稠密连接. Zhang 等[23] 提出残差稠密网络, 同时使用残差稠密连接. 稠密连接和残差稠密连接比残差连接可以更好地利用深度卷积层的分层特征. 同样, 注意力机制在 SR 网络中也得到了广泛关注, 通道注意力机制[11] 考虑不同通道特征之间的依赖, 显著提高了模型的表示能力和 SR 性能. Zhang 等[13] 提出一种非局部注意力块, 以建模长距离像素之间的依赖, 可以很好地捕捉空间注意力, 进一步增强特征的表示能力. Niu 等[24] 提出整体注意力网络, 进一步组合了层注意力、通道和空间注意力, 以建模层、通道和空间位置之间的整体依赖. Li 等[25] 提出多尺度残差网络 (Multi-scale residual network, MSRN), 采用局部多路径学习, 同时提取多个尺度特征, 改进了特征的表示能力. 视觉 Transformer 能够建模全局像素之间的依赖, 在图像分类、物体检测和分割等高级视觉任务中, 取得了巨大的成功. Liang 等[26] 将 Swin Transformer[27] 引入图像恢复, 取得了最先进的性能. 这些代表性的 SISR 方法虽然有好的 SR 性能, 但模型参数量较大, 需要较大的内存和较强的计算能力, 限制了它们在资源受限设备上的应用. 目前, 网络架构设计的趋势是设计更轻量的网络模型, 找到网络模型复杂度和性能之间最优的平衡[28]. 本文模型的基本构件设计受 Zhang 等[11] 和 Li 等[25] 启发, 相较于 RCAN 的基本构件 RCAB, 通过引入扩张卷积, 可以提取多尺度特征[25] 且不增加参数量; 同时, 设计一个空间注意力融合块, 可以有效地融合这些多尺度的特征.

1.2 图像超分辨率的损失函数

目前, 基于 CNN 的 SR 技术研究大多聚焦于网络架构设计, 但合适的损失函数对 SR 模型的性能也至关重要. 最广泛使用的损失函数是逐像素的 L_1 损失 (即平均绝对误差) 和 L_2 损失 (即均方误差). 这些逐像素的误差度量返回许多可能解的平均值[14], 会导致 SR 图像出现模糊、过度平滑和不自然的外观等问题, 尤其是在信息丰富的区域. 为了改进 SR 图像的感知质量, Johnson 等[15] 提出感知损失, 使用预训练的图像分类网络, 在特征空间度量高级感知和语义差异. 与 L_1 等像素损失不同, 感知损失鼓励输出的 SR 图像与目标图像具有相似特征表示, 而不是迫使其像素匹配. 考虑到 SR 图像应与目标图像具有相同的纹理, Sajjadi 等[19] 提出一种基于块的纹理损失, 图像纹理视为不同特征通道之间的相关性, 其定义基于预训练分类网络特征的 Gram 矩阵. 纹理损失迫使 SR 图像应与目标图像之间局部纹理相似. Ledig 等[20] 提出对抗性损失, 鼓励 SR 图像逼近自然图像流形. 这些损失函数改进 SR 图像的视觉质量, 但是保真度低, 仍不能阻止图像模糊.

1.3 对比学习

对比学习[29−30] 旨在学习一个嵌入的空间, 使得相似的样本彼此靠近, 不同的样本彼此远离. 对比学习可用于监督的学习环境, 也可用于无监督的学习环境, 在图像分类和图像聚类等高级视觉任务中, 应用非常成功, 但很少用于低级视觉任务. 高级视觉任务的对比学习技术可能不适用于低级视觉任务, 因为前者更需要全局视觉表示, 而后者更需要丰富的纹理细节[31]. 对比学习需要考虑度量相似性的隐空间和正/负样本的选取两个重要因素. 在低级视觉任务中, 使用的隐空间大都是预训练网络 (例如 VGG) 的特征空间. 采用简单方法选取正/负样本, 例如以输入的低质量图像作为负样本, 以输入的高质量图像作为正样本. Wu 等[32] 提出图像去雾方法, 使用 VGG 特征空间真实图像作为正样本, 退化的雾图像作为负样本, 估计的图像作为锚. 在 VGG 特征空间中, 让锚靠近正样本, 远离负样本. Wang 等[17] 提出一个对比自蒸馏 SR 网络, 同时使用 VGG 特征空间, 其中教师子网络输出的 SR 图像作为正样本, 学生子网络输出的 SR 图像作为锚,

从同一个批次中采样 K 个图像 (除锚外), 通过双三次插值上采样到与输出 SR 图像相同的分辨率, 作为负样本. 在盲 SR 中, Wang 等[33] 使用一个编码器子网络, 学习不同退化的抽象表示. 假定相同图像中的块退化相同, 不同图像中的块退化不同, 以分别选择为正/负样本. Wu 等[31] 提出一个基于对比学习的 SR 框架, 以判别器子网络的特征空间作为隐空间. 通过对真实图像施加一些轻微的模糊, 生成多个难的负样本并简单锐化真实图像, 生成多个信息丰富的正样本. Wu 等[31] 的隐空间对退化更敏感, 其负样本试图促使 SR 图像远离平滑的结果, 而其正样本试图迫使 SR 图像吸收更多的细节信息. 本文受到 CSD 的启发, CSD 虽能改进视觉结果, 但保真度低, 正/负样本限定的上/下界距离大, 即上/下界较弱. 本文直接使用图像空间训练 HR 图像和输出的 SR 图像对应的像素, 分别作为正样本和锚; HR图像上正样本邻近的像素, 作为负样本.

2 本文方法

本文的渐进残差特征融合网络架构如图 3 中模块 1) 所示, 主要包括浅层特征提取块 (Shallow feature extraction block, SFEB)、特征映射块 (Feature mapping block, FMB) 和上采样块三部分. 其中, SFEB 仅包含一个 3×3 的卷积层, 提取浅层特征信息; FMB 包含 N (实验中, $N=10$) 个渐进特征融合组 (Progressive feature fuse group, PFFG); 上采样块使用亚像素卷积[34], 使用 L_1 损失和本文的像素级对比损失 L_{cntr}. 假定输入的 LR 图像为 I_{LR}, 首先, 经过 SFEB 提取浅层特征:

$$F_0 = f_{SFEB}\left(I_{LR}\right) \tag{1}$$

式中, $f_{SFEB}(\cdot)$ 是浅层特征提取函数, F_0 是其输出. 然后, F_0 再通过 FMB 进行深层特征提取:

$$F_n = f_{PFFG}^{n}\left(F_{n-1}\right),\ n=1,\ 2,\ \cdots,\ N \tag{2}$$

式中, $f_{PFFG}^{n}(\cdot)$ 是 FMB 中第 n 个 PFFG 函数, F_n 是其输出, F_{n-1} 是其输入即第 $n-1$ 个 PFFG 的输出, N 是 FMB 中 PFFG 的个数. 接着, 深层残差特征 F_N 组合浅层特征 F_0, 通过上采样块重建目标 SR 图像:

$$I_{SR} = f_{UP}\left(F_n + F_0\right) \tag{3}$$

式中, $f_{UP}(\cdot)$ 是上采样块函数. 最后, 输出 SR 图像 I_{SR}.

2.1 渐进特征融合组

FMB 包含多个 PFFG, PFFG 中特征融合设计主要受到 Liu 等[35] 启发. 他们的研究表明, 网络中不同深度的残差特征逐渐集中于输入图像的不同方面, 对于重建空间细节非常有用. 基于 CNN 的 SISR 模型大都只是将残差学习作为缓解训练难度的策略. SISR 模型堆叠残差块, 残差特征传播到下一个块之前, 与恒等特征融合, 致使后面的残差块只馈入了复杂的融合特征, 忽略了充分利用更清洁

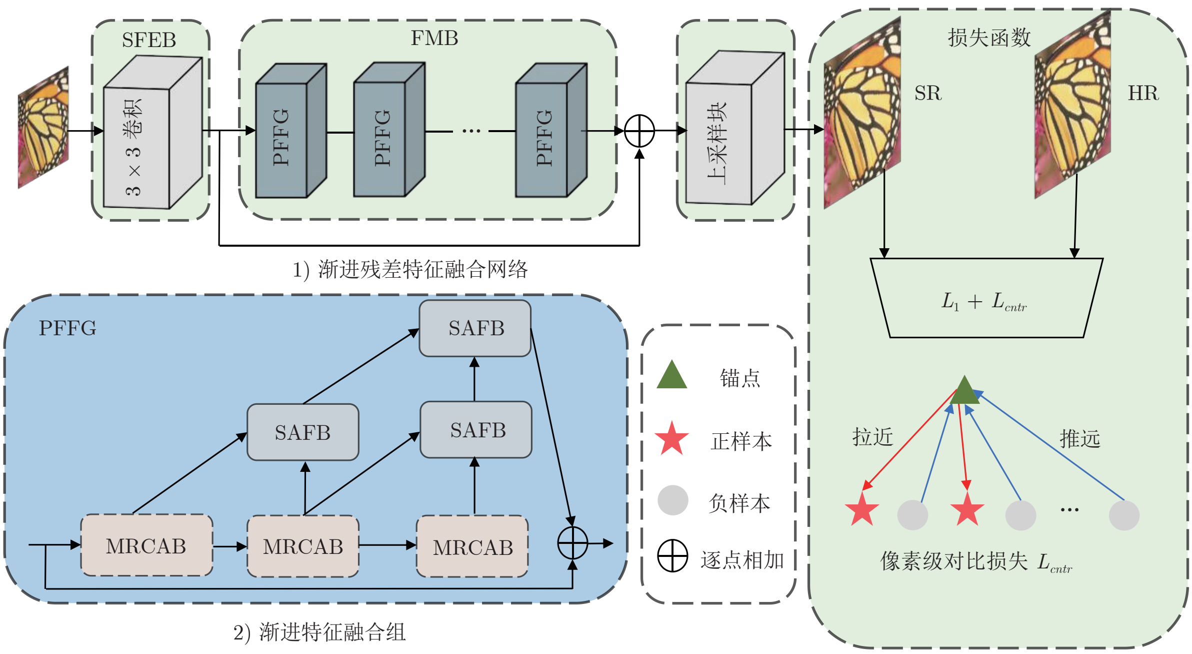

图 3 网络架构细节

Fig.3 Network architecture details

的残差特征. Liu 等[35] 设计的残差模块包含多个残差块, 并把各个残差块输出的特征拼接在一起, 馈入 1×1 卷积层进行融合.

PFFG 由 MRCAB 和 SAFB 组成, 每个 MRCAB 的输出特征是下一个 MRCAB 的输入, 且与后续所有 MRCAB 输出特征, 通过 SAFB 进行分层融合, 以强化中间残差特征, 见图 3 中模块 2). 每个 PFFG 使用了 3 个 MRCAB, 残差分支的特征融合可以分为 3 个步骤: 1) 从左到右, 每两个相邻的 MRCAB 输出特征通过 SAFB 进行融合; 2) 步骤 1) 输出的融合特征从左到右使用 SAFB, 对每两个相邻的特征进行融合, 其余类推; 3) 步骤 2) 输出的融合特征作为 PFFG 的残差分支输出, 与 PFFG 的输入特征求和. 每个 PFFG 有 3 个 MRCAB, 第 n 个 PFFG 中第 m ($m=1, 2, 3$) 个 MRCAB 的输出为 $F_{n,m}^{MR}$, 则第 n 个 PFFG 中第 1 个 MRCAB 的输入是前一个 PFFG 的输出 F_{n-1} (第 1 个 PFFG 中, 第 1 个 MRCAB 的输入为 SFEB 的输出 F_0); 第 n 个 PFFG 中第 m ($m=2, 3$) 个 MRCAB 的输入为前一个 MRCAB 的输出 $F_{n,m-1}^{MR}$. F_n 可表示为:

$$\begin{cases}F_{n,m}^{MR}=f_{MRCAB}\left(F_{n-1}\right), & m=1\\ F_{n,m}^{MR}=f_{MRCAB}\left(F_{n,m-1}^{MR}\right), & m=2, 3\end{cases} \tag{4}$$

式中, $f_{MRCAB}(\cdot)$ 是 MRCAB 函数; $F_{n,m}^{MR}$ 是第 n ($n=1, 2, \cdots, N$) 个 PFFG 中, 第 m ($m=1, 2, 3$) 个 MRCAB 的输出:

$$F_n=f_{SAFB}(f_{SAFB}(F_{n,1}^{MR}, F_{n,2}^{MR}), f_{SAFB}(F_{n,2}^{MR}, F_{n,3}^{MR}))+F_{n-1} \tag{5}$$

式中, $f_{SAFB}(\cdot)$ 是 SAFB 块函数.

2.2　多尺度残差通道注意力块

本文的 MRCAB 主要受 Zhang 等[11] 和 Li 等[25] 启发. Li 等[25] 开发了一个多尺度的残差块 (Multi-scale residual block, MSRB). MSRB 包含两个分支, 一个分支使用 3×3 卷积, 另一个分支使用 5×5 卷积. 两个分支输出的不同尺度特征拼接在一起, 再分别通过 3×3 和 5×5 卷积. 通过实验可以发现, 更多尺度可以更好地利用特征信息, 因此本文的 MRCAB 使用三个分支, 如图 4 所示. 为了减少参数量, 一个分支使用 3×3 卷积, 其余两个分支使用 3×3 扩张卷积, 扩张率分别为 2 和 4. 另外, 为了简化 MSRB, 本文把三个分支输出的不同尺度特征简单地求和, 进行融合. 在 MRCAB 的后部, 引入了 Zhang 等[11] 的通道注意力机制, 通过自适应地伸缩各个通道, 以建模各个通道之间的依赖性. 第 n 个 PFFG 中, 第 m ($m=1, 2, \cdots, M$) 个 MRCAB 的第 k 个分支 ($k=1, 2, 3$) 可表示为:

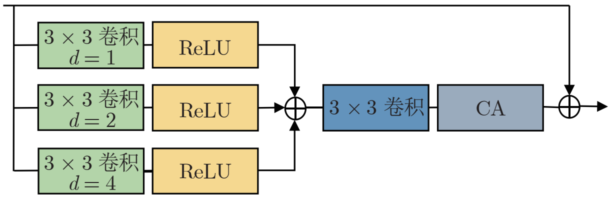

图 4　多尺度残差通道注意力块

Fig.4　Multi-scale residual channel attention block

$$F_{S_m}^k=\sigma(DC_{3\times3}^d(F_{n,m-1}^{MR})) \tag{6}$$

式中, $F_{S_m}^k$ 是输出的特征, $\sigma(\cdot)$ 是 ReLU 非线性激活函数, $DC_{3\times3}^d(\cdot)$ 是 3×3 的扩张卷积 (扩张率 $d=1, 2, 4$), $F_{n,0}^{MR}=F_{n-1}$.

第 n 个 PFFG 中, 第 m ($m=1, 2, \cdots, M$) 个 MRCAB 的输出可表示为:

$$F_{n,m}^{MR}=\mathrm{CA}\left(\mathrm{Conv}_{3\times3}\left(\sum_{k=1}^{3}F_{S_m^k}\right)\right)+F_{n,m-1}^{MR} \tag{7}$$

式中, CA $(\cdot)$ 是 RCAN 中通道注意力函数.

2.3　空间注意力融合块

考虑到 MRCAB 中使用了 Zhang 等[11] 的通道注意力建模特征通道之间的依赖关系, 在融合两个 MRCAB 的输出特征时, 使用空间注意力, 进一步建模特征像素之间的依赖关系. 两个 MRCAB 的输出特征拼接在一起, 使用 1×1 卷积进行融合, 生成两个特征通道, 再使用 Sigmoid 函数, 将通道像素值转换到 (0, 1) 之间, 作为输入特征的权重. 学习输入特征像素之间的依赖关系, 如图 5 所示. 假定输入的特征分别为 F_A 和 F_B, SAFB 中特征融合过程可表示为:

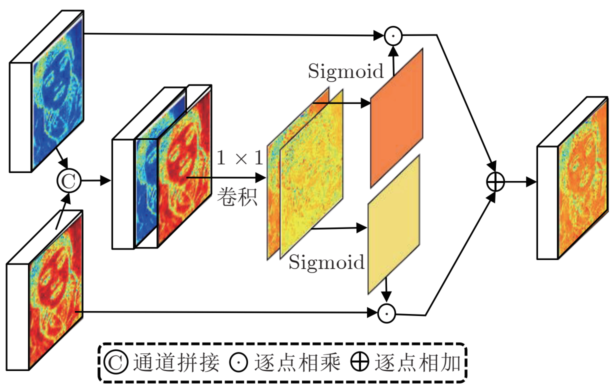

图 5　空间注意力融合块

Fig.5　Spatial attention fusion block

$$\begin{cases} W_A,\, W_B = f_{split}(\mathrm{Conv}_{1\times1}([F_A,\, F_B])) \\ F_{fuse} = \mathrm{Sigmoid}\,(W_A)\cdot F_A + \mathrm{Sigmoid}\,(W_B)\cdot F_B \end{cases} \tag{8}$$

式中, F_{fuse} 为 SAFB 输出的融合特征, $\mathrm{Conv}_{1\times1}(\cdot)$ 为 1×1 卷积函数, $f_{split}(\cdot)$ 把两个通道分裂为矩阵 W_A 和 W_B, $\mathrm{Sigmoid}(\cdot)$ 为 Sigmoid 函数, "$\cdot$" 为逐像素相乘.

SAFB 利用空间注意力机制, 渐进地融合分层的特征信息, 可以更好地利用和融合多样化的特征信息.

3 损失函数

本文的核心工作是引入像素级对比损失, 使被恢复的像素更靠近原 HR 图像中对应像素, 远离其他邻近像素, 以提高 SR 图像的保真度和清晰度. 对比学习框架[29, 32, 36−37] 主要是为高层语义理解任务设计的, 在低级视觉中应用的潜力还没有得到充分探索. 对于给定的锚点, 对比学习的目标是将锚点拉向正样本, 远离负样本. 基于 CNN 的 SISR 方法虽然显著改进了之前的 SISR 方法, 但如何恢复 SR 图像中丢失、模糊或失真的细节, 仍然是有待解决的问题. 目前, 已有一些损失函数 (如感知损失 L_{perc} [15]、纹理损失 L_{tex} [19] 和对抗损失 L_{adv} [20] 等) 改进了被广泛使用的 L_1 或 L_2 损失, 但仍不能令人满意. L_{perc} 和 L_{tex} 仍然不能阻止模糊, L_{adv} 虽然能生成清晰的细节, 但生成的细节严重失真和不自然. 为了解决这个问题, 本文提出像素级对比损失 L_{cntr}. L_{cntr} 可以与 L_1 、L_2 、L_{perc} 、L_{tex} 和 L_{adv} 等一个或多个结合使用, 进一步改进 SR 图像的保真度和清晰度.

优化 L_1 或 L_2 损失函数是将待恢复的 SR 图像中, 每个像素逐渐逼近原 HR 图像中对应的像素. 遗憾的是, SR 图像恢复是一个病态逆问题, 同一个低分辨率图像可以对应多个高分辨率图像, 恢复的 SR 图像是许多高分辨率图像的平均, 导致过度平滑. 本文的像素级对比损失 L_{cntr} 将 SR 图像中的像素作为锚点, HR 图像中对应的像素作为正样本, 与该正样本像素值相同的像素也作为正样本, 其他像素作为负样本. L_{cntr} 与 L_1 或 L_2 损失函数联合使用, 增加一个附加的约束, 让锚点接近正样本, 远离负样本, 具体过程见图 6. 不同的负样本可以自适应地推远锚点. L_{cntr} 的形式类似于 InfoNCE[38] 损失, 它是基于 Softmax 的分类损失, 在小批量样本中, 分类正样本和负样本. 原 HR 图像分为许多大小相等的区域 (假定共有 Q 个区域), 在每个局部区域计算对比损失. 在给定 HR 图像中, 第 q 个局部区域假定包含 S 个像素, 第 i 个像素 (记为 $x^q_{HR,\,i}$) 视为

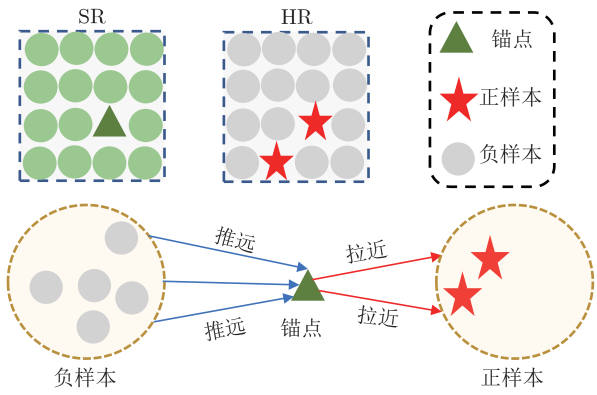

图 6 像素级对比损失

Fig. 6 Pixel-wise contrastive loss

正样本, 其他具有相同值的像素, 也视为正样本, 该区域对比损失计算为:

$$L^q_{cntr} = -\frac{1}{S}\sum_{i=1}^{S}\left[\ln \frac{\sum\limits_{j=1}^{S_1}\exp\left(-\left\|x^q_{SR,\,i} - x^q_{P,\,j}\right\|_1\right)}{\sum\limits_{j=1}^{S}\exp\left(-\left\|x^q_{SR,\,i} - x^q_{HR,\,j}\right\|_1\right)}\right] \tag{9}$$

式中, L^q_{cntr} 为第 q 个区域对比损失, 即该区域中每个像素视为正样本时的平均值. $x^q_{SR,\,i}$ 为 SR 图像第 q 个区域中, 第 i 个锚点; HR 图像第 q 个区域中, 与之对应的 $x^q_{HR,\,i}$ 是正样本. 在该区域中, 与 $x^q_{HR,\,i}$ 值相同的像素也视为正样本, 记为 $x^q_{P,\,j}$, 总共 S_1 个 $(j = 1,\, 2,\, \cdots,\, S_1)$. 总逐像素对比损失 (表示为 L_{cntr}) 是 Q 个区域中, 逐像素对比损失 L^q_{cntr} 之和, 即:

$$L_{cntr} = \sum_{q=1}^{Q} L^q_{cntr} \tag{10}$$

网络模型总损失计算为:

$$L_C = \lambda_C L_{cntr} + L_1 \tag{11}$$

式中, λ_C 是系数, 以平衡 L_{cntr} 和 L_1 损失.

4 实验分析

4.1 实验设置

本文使用 DIV2K 数据集[39] 训练模型, 其中 800 个图像作为训练图像, 5 个图像用于验证集, 标记为 DIV2K_val5. 为了增强训练图像, 训练图像进行了 90° 旋转和随机的水平翻转. 测试中, 使用 Set5[40]、Set14[6]、B100[41]、Urban100[42] 和 Manga-109[43] 5 个标准测试数据集; 对原 HR 图像进行双三次下采样, 以获得对应的 LR 图像. 与其他方法类似, 使用 PSNR 和 SSIM 评估模型的客观性能, PSNR 和 SSIM 均在 YCbCr 空间[23] 的 Y 通道上进行计

算. 训练中, 批大小设置为 16, LR 图像片尺寸设置为 48×48 像素. 使用 ADAM 优化器, 设置 $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$. 初始学习率设置为 10^{-4}, 每 200 个迭代周期减半. 使用 PyTorch 框架实现模型, 并在 NVIDIA GeForce RTX 2080Ti GPU 上训练.

4.2　实验分析

4.2.1　消融实验

为了验证本文网络架构和像素级对比损失的有效性, 进行了 5 组对比实验: 1) 把网络架构中的 MRCAB 替换为 RCAB, 仅使用 L_1 损失, 不包含 SAFB 块和像素级对比损失 L_{cntr}, 以该模型作为基准, 称为 PRFFN0; 2) 在 PRFFN0 基础上, 增加类似于残差特征聚合 (Residual feature aggregation, RFA)[35] 的特征融合方法, 与 RFA 块的区别是将其残差子块替换为 RCAB, 这个模型称为 PRFFN1; 3) 将 PRFFN1 中的 RFA 特征融合方法替换为 SAFB, 进行特征融合, 这个模型称为 PRFFN2; 4) 将 PRFFN2 中的 RCAB 替换为 MRCAB, 这个模型称为 PRFFN3; 5) 在 PRFFN3 中, 增加像素级对比损失 L_{cntr}, 这个模型称为 PRFFN 即本文模型. 训练 1 000 个迭代周期, 5 个模型在 DIV2K_val5 验证集上 3 倍超分辨率的平均 PSNR 和参数量如表 1 所示. 表 1 中, "✓"表示在训练时使用该损失, "—"表示训练时不使用该损失. 由表 1 可以看出, 基准模型 PRFFN0 的平均 PSNR 为 32.259 dB; RFA 特征融合改进了 PRFFN0 的 PRFFN1, 平均 PSNR 增加了 0.048 dB. 相较于 RFA, 本文 SAFB 使用了空间注意力机制, 增强了特征空间依赖关系之间的建模, 优于 RFA 中特征图的简单拼接, 使 PRFFN2 较 PRFFN1 模型, 参数量减少了 55 K, 平均 PSNR却增加了 0.035 dB; 相较于 RCAB, 本文 MRCAB使用了扩张卷积, 在不增加模型参数量的情况下, 扩大了特征感受野, 且融合了多个尺度扩张卷积特征, 使 PRFFN3 较 PRFFN2, 平均 PSNR 增加了 0.022 dB; 在 L_1 损失的基础上, 增加像素级对比损失 L_{cntr}, 提高了 SR 图像的保真度, 使 PRFFN 较 PRFFN3, 平均 PSNR 增加了 0.087 dB.

4.2.2　对比损失的设计与分析

1) 对比损失的有效性. 为了验证对比损失的有效性, 实验比较了 L_1 、L_{perc} 、L_{CSD} 和 L_{cntr} 损失及其组合对应的 PSNR 和学习的感知图像块相似性 (Learned perceptual image patch similarty, LPIPS)[44] 度量. LPIPS 被认为更符合人类的视觉感知, 视觉感知质量随 LPIPS 的降低而增加, 而 PSNR 值越高, 则图像的保真度越高. 训练 1 000 个迭代周期, 实验结果如表 2 所示, 表 2 中"✓"表示在训练时, 使用该损失; "—"表示训练时, 不使用该损失. L_{cntr} 与其他损失组合时, 可改进 PSNR 和 LPIPS 值. 当 L_{cntr}与 L_1组合时, 获得了最好的 PSNR 值 (32.451 dB); 当 L_{cntr} 与 L_1 、L_{CSD} 组合时, 获得了最好的 LPIPS 值 (0.0613). 由表 2 可知, L_{CSD} 作用于图像特征空间, 获得了更好的 LPIPS; L_{cntr} 作用于像素空间, 获得了更好的 PSNR, 即保真度更高. L_{cntr} 组合 L_{CSD} 虽获得了最好的 LPIPS, 但 PSNR 有所降低.

表 1　DIV2K_val5 验证集上, 不同模型, 3 倍 SR 的平均 PSNR 和参数量

Table 1　The average PSNRs and parameter counts of 3 times SR for different models on the DIV2K_val5 validation data set

模型	L_1	L_{cntr}	参数量 (K)	PSNR (dB)
PRFFN0	✓	—	2 975	32.259
PRFFN1	✓	—	3 222	32.307
PRFFN2	✓	—	3 167	32.342
PRFFN3	✓	—	3 167	32.364
PRFFN	✓	✓	3 167	32.451

表 2　DIV2K_val5 验证集上, 不同损失函数及其组合, 3 倍 SR 的平均 PSNR 和 LPIPS 结果

Table 2　The average PSNRs and LPIPSs of 3 times SR for different losses and their combinations on the DIV2K_val5 validation data set

L_1	L_{perc}	L_{CSD}	L_{cntr}	PSNR (dB)	LPIPS
✓	—	—	—	32.364	0.0978
✓	—	—	✓	**32.451**	0.0969
✓	✓	—	—	32.236	0.0672
✓	✓	—	✓	32.305	0.0656
✓	—	✓	—	32.387	<u>0.0624</u>
✓	—	✓	✓	<u>32.432</u>	**0.0613**

2) 不同比例对比损失影响. 如式 (11) 所示, 比例系数 λ_C 可以调节对比损失的比例. 训练 1 000 个迭代周期, 在 DIV2K_val5 验证集上, 不同 λ_C, 3 倍 SR 的平均 PSNR 结果见表 3. 由表 3 可以看出, 过大或过小的 λ_C 使 PSNR 结果恶化. 当 $\lambda_C = 10^{-1}$ 时, 获得了最好的 PSNR 性能.

3) 对比损失的通用性. 本文的对比损失 L_{cntr} 可以与 L_1 、L_{perc} 和 L_{CSD} 等损失组合使用, 从而改进 SR 图像的保真度. L_{cntr} 也可用于其他模型, 可以改进 SR 的性能. 以著名的 EDSR、RCAN 和 SwinIR-light[26] 为例, 训练 1 000 个迭代周期, 原模型和增加 L_{cntr} 损失后的平均 PSNR 和 SSIM 结果

表 3 DIV2K_val5 验证集上, 不同 λ_C , 3 倍 SR 的平均 PSNR 结果

Table 3 The average PSNRs of 3 times SR for different λ_C on the DIV2K_val5 validation data set

λ_C	PSNR (dB)
10^{-2}	32.386
10^{-1}	32.451
1	32.381
10	32.372

如表 4 所示, 表 4 中“✓”表示在训练时使用该损失, “—”表示训练时不使用该损失, “↑”代表指标提升. 由表 4 可以看出, L_{cntr} 损失可以显著改进 SR 图像视觉效果. SwinIR-light 增加 L_{cntr} 后, 在 Urban-100 数据集中, img004 图像的 3 倍 SR 结果如图 7 所示. 由图 7 可以看出, 使用 L_{cntr} 损失后, SR 图像的视觉质量得到显著改进.

表 4 DIV2K_val5 验证集上, 不同模型包含与不包含 L_{cntr} 损失, 3 倍 SR 的平均 PSNR 和 SSIM 结果

Table 4 The average PSNRs and SSIMs of 3 times SR for different models with and without L_{cntr} loss on the DIV2K_val5 validation data set

模型	L_{cntr}	PSNR (dB)	SSIM
EDSR	—	32.273	0.9057
	✓	32.334 (↑0.061)	0.9067 (↑0.0010)
SwinIR-light	—	32.442	0.9062
	✓	32.489 (↑0.047)	0.9069 (↑0.0007)
RCAN	—	32.564	0.9088
	✓	32.628 (↑0.064)	0.9096 (↑0.0008)

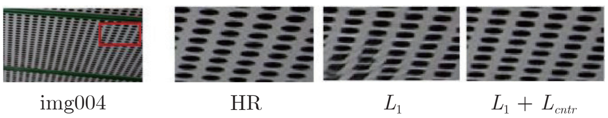

图 7 Urban100 数据集中, SwinIR-light 使用不同损失函数, img004 图像的 3 倍 SR 结果

Fig. 7 The 3 times SR results of SwinIR-light using different losses on the img004 image in the Urban100 data set

4) 不同大小局部区域对对比损失的影响. 计算对比损失的方形局部区域, 可用区域中像素个数 S 度量. 本文考虑 4×4 像素、8×8 像素和 16×16 像素的局部区域, S 分别为 16、64 和 256. 在 DIV2K_val5 验证集上, 训练 1 000 个迭代周期, 不同 S 值的 3 倍 SR 的平均 PSNR 结果见表 5. 可以看出, 更大的局部区域 (即更大的 S 值) 负样本更多, 有利于改进 PSNR 性能. 但在训练时, 更大的 S 值需要更多的时间和更大的内存量 (见表 6). 考虑到相较于 $S=64$, 在 $S=256$ 时, PSNR 改进很小, 因此本文选择 $S=64$.

表 5 DIV2K_val5 验证集上, 不同大小局部区域, 3 倍 SR 的平均 PSNR 结果

Table 5 The average PSNRs of 3 times SR for different size local regions on the DIV2K_val5 validation data set

S	PSNR (dB)
16	32.425
64	32.451
256	32.458

表 6 3 倍 SR 训练 10 个迭代周期, 训练占用的内存和使用的训练时间

Table 6 For 3 times SR, 10 epochs, comparing the memory and time used by training

L_{cntr}	S	内存 (MB)	时间 (s)
—	—	4126	1962
✓	16	4836	2081
✓	64	5216	2219
✓	256	7541	2893

5) 对比学习训练的计算代价. 4 组对比实验分别为不使用 L_{cntr} 损失以及 $S=16$、64 和 256 三种情形. 对于 3 倍 SR 训练 10 个迭代周期, 训练占用的内存和使用的训练时间如表 6 所示. 表 6 中, “✓”表示在训练时使用该损失, “—”表示训练时不使用该损失. 由表 6 可以看出, 计算代价没有显著增加. 当 $S=64$ 时, 内存增加了约 21%, 计算时间仅增加了约 11%.

4.2.3 PFFG 实验

在网络架构中, 共由 10 个 PFFG 组成, 每个 PFFG 包含 3 个 MRCAB 和 3 个 SAFB, 见图 3 中模块 2).

1) 扩张卷积的不同扩张率对性能的影响. 与 RCAN 中的 RCAB 相比, MRCAB 没有增加参数, 并且通过多分支提取了多尺度特征信息, 并通过扩张卷积 (或空洞卷积) 扩大了感受野. MRCAB 中, 3 个分支扩张卷积的扩张率组合分别设为 1, 2, 3、1, 2, 4 和 1, 2, 5. 考虑只有 1 个分支使用常规卷积即扩张率为 1, 此时 MRCAB 退化为 RCAB; 考虑 2 个分支, 扩张率组合为 1, 2 的情形, 训练 1 000 个迭代周期, 不同分支和不同扩张率组合, 3 倍 SR 的平均 PSNR 结果见表 7. 由表 7 可见, 不同尺度的特征融合可以改进性能. 扩张率更大, 特征感受野大, 对性能有益. 但过大的扩张率, 也是有害的. MRCAB 三个分支的扩张率设置为 1, 2, 4 较为合理.

表 7　DIV2K_val5 验证集上, MRCAB 不同分支和不同扩张率组合, 3 倍 SR 的平均 PSNR 结果

Table 7　The average PSNRs of 3 times SR for the different branches of MRCAB with different dilation rate combinations on the DIV2K_val5 validation data set

不同的扩张率卷积组合	PSNR (dB)
1	32.375
1, 2	32.370
1, 2, 3	32.392
1, 2, 4	32.451
1, 2, 5	32.415

4.3　与先进方法比较

1) 客观比较. 本文与代表性的双三次上采样、快速超分辨率卷积神经网络 (Fast super-resolution convolutional neural network, FSRCNN)[45]、稀疏掩码超分辨率网络 (Sparse mask super-reolution, SMSR)[46]、自适应级联注意力网络 (Adaptive cascading attention network, ACAN)[47]、权重自适应超分辨率网络 (Adaptive weighted super-resolution network, AWSRN)[48]、深度递归卷积网络 (Deeply-recursive convolutional network, DRCN)[49]、级联残差网络 (Cascading residual network, CARN)[50]、OISR-RK2 (Apply ODE-inspired schemes to super-resolution network designs-Runge-Kutta)[51]、OISR-LF (Apply ODE-inspired schemes to super-resolution network designs-leapfrog)[51]、MSRN、软边缘辅助网络 (Soft-edge assisted network, SeaNet)[52] 和两阶段注意力网络 (Two-stage attentive network, TSAN)[53] 方法进行比较. 在 5 个标准测试数据集上, 2 倍、3 倍和 4 倍 SR 的平均 PSNR 和 SSIM 结果见表 8. 与最先进的 TSAN 相比, 本文模型参数和计算量大约减少了 1/3, 推理时间大约减少了一半, 但性能更好. 例如, 对于 2 倍 SR, 本文方法与 TSAN 相当; 对于 3 倍和 4 倍 SR, 在所有测试数据

表 8　5 个标准测试数据集上, 不同 SISR 方法的 2 倍、3 倍和 4 倍 SR 的平均 PSNR 和 SSIM 结果

Table 8　The average PSNRs and SSIMs of 2 times, 3 times, and 4 times SR for different SISR methods on five standard test data sets

放大倍数	方法	参数量 (K)	计算量 (G)	推理时间 (ms)	Set5 PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM	Urban100 PSNR/SSIM	Manga109 PSNR/SSIM
2	FSRCNN	13	6.0	26	37.00/0.9558	32.63/0.9088	31.53/0.8920	29.88/0.9020	36.67/0.9694
	SMSR	985	224.1	536	38.00/0.9601	33.64/0.9179	32.17/0.8990	32.19/0.9284	38.76/0.9771
	ACAN	800	2108.0	—	38.10/0.9608	33.60/0.9177	32.21/0.9001	32.29/0.9297	38.81/0.9773
	AWSRN	1397	320.5	506	38.11/0.9608	33.78/0.9189	32.26/0.9006	32.49/0.9316	38.87/0.9776
	DRCN	1774	17974.0	—	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133	37.55/0.9732
	CARN	1592	222.8	352	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256	38.36/0.9765
	OISR-RK2	4971	1145.7	715	38.12/0.9609	33.80/0.9193	32.26/0.9007	32.48/0.9317	38.79/0.9773
	OISR-LF	4971	1145.7	722	38.12/0.9609	33.78/0.9196	32.26/0.9007	32.52/0.9320	38.80/0.9774
	MSRN	6078	1356.8	810	38.08/0.9607	33.70/0.9186	32.23/0.9002	32.29/0.9303	38.69/0.9772
	SeaNet	7471	3709.1	1920	38.08/0.9609	33.75/0.9190	32.27/0.9008	32.50/0.9318	38.76/0.9774
	TSAN	3989	1013.1	1183	**38.22/0.9619**	33.84/**0.9218**	**32.32/0.9015**	**32.77/0.9345**	—/—
	PRFFN	2988	656.4	792	38.18/0.9611	**33.90**/0.9207	32.30/0.9012	32.75/0.9337	**39.02/0.9777**
3	FSRCNN	13	4.6	14	33.16/0.9140	29.43/0.8242	28.53/0.7910	26.43/0.8080	30.98/0.9212
	SMSR	993	100.5	281	34.40/0.9270	30.33/0.8412	29.10/0.8050	28.25/0.8536	33.68/0.9445
	ACAN	1115	1051.7	—	34.46/0.9277	30.39/0.8435	29.11/0.8065	28.28/0.8550	33.61/09447
	AWSRN	1476	150.6	263	34.52/0.9281	30.38/0.8426	29.16/0.8069	28.42/0.8580	33.85/0.9463
	DRCN	1774	17974.0	—	33.85/0.9215	29.89/0.8304	28.81/0.7954	27.16/0.8311	32.31/0.9328
	CARN	1592	118.8	177	34.29/0.9255	30.29/0.8407	29.06/0.8034	27.38/0.8493	33.50/0.9440
	OISR-RK2	5640	578.6	366	34.55/0.9282	30.46/0.8443	29.18/0.8075	28.50/0.8597	33.80/0.9442
	OISR-LF	5640	578.6	367	34.56/0.9284	30.46/0.8450	29.20/0.8077	28.56/0.8606	33.78/0.9441
	MSRN	6078	621.2	476	34.38/0.9262	30.34/0.8395	29.08/0.8041	28.08/0.8554	33.44/0.9427
	SeaNet	7397	3233.2	994	34.55/0.9282	30.42/0.8444	29.17/0.8071	28.50/0.8594	33.73/0.9463
	TSAN	4174	565.6	650	34.64/0.9282	30.52/0.8454	29.20/0.8080	28.55/0.8602	—/—
	PRFFN	3167	312.5	441	**34.67/0.9288**	**30.54/0.8460**	**29.23/0.8084**	**28.65/0.8621**	**34.03/0.9473**

表 8 5 个标准测试数据集上, 不同 SISR 方法的 2 倍、3 倍和 4 倍 SR 的平均 PSNR 和 SSIM 结果 (续表)

Table 8 The average PSNRs and SSIMs of 2 times, 3 times, and 4 times SR for different SISR methods on five standard test data sets (continued table)

放大倍数	方法	参数量 (K)	计算量 (G)	推理时间 (ms)	Set5 PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM	Urban100 PSNR/SSIM	Manga109 PSNR/SSIM
4	FSRCNN	13	4.6	8	30.71/0.8657	27.59/0.7535	26.98/0.7150	24.62/0.7280	27.90/0.8517
	SMSR	1006	57.2	192	32.12/0.8932	28.55/0.7808	27.55/0.7351	26.11/0.7868	30.54/0.9085
	ACAN	1556	616.5	—	32.24/0.8955	28.62/0.7824	27.59/0.7379	26.31/0.7922	30.53/0.9086
	AWSRN	1587	91.1	188	32.27/0.8960	28.69/0.7843	27.64/0.7385	26.29/0.7930	30.72/0.9109
	DRCN	1774	17974.0	—	31.56/0.8810	28.15/0.7627	27.23/0.7150	25.14/0.7510	28.98/0.8816
	CARN	1592	90.9	121	32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837	30.47/0.9084
	OISR-RK2	5500	412.2	241	32.32/0.8965	28.72/0.7843	27.66/0.7390	26.37/0.7953	30.75/0.9082
	OISR-LF	5500	412.2	239	32.33/0.8968	28.73/0.7845	27.66/0.7389	26.38/0.7953	30.76/0.9080
	MSRN	6078	365.1	352	32.07/0.8903	28.60/0.7751	27.52/0.7273	26.04/0.7896	30.17/0.9034
	SeaNet	7397	3065.6	704	32.33/0.8981	28.72/0.7855	27.65/0.7388	26.32/0.7942	30.74/0.9129
	TSAN	4137	415.1	452	32.40/0.8975	28.73/0.7847	27.67/0.7398	26.39/0.7955	—/—
	PRFFN	3131	200.1	316	**32.43/0.8983**	**28.75/0.7857**	**27.70/0.7398**	**26.55/0.7974**	**30.93/0.9133**

集上, 本文方法均优于 TSAN. 特别是在 Urban100 数据集上, 平均 PSNR 最大改进分别为 0.1 dB 和 0.16 dB.

2) 主观效果比较. 各种方法 SR 的视觉效果比较见图 8 ~ 10. 可以看出, 本文方法恢复的 SR 图像视觉质量优于其他方法. 例如, 对于 Urban100 数据

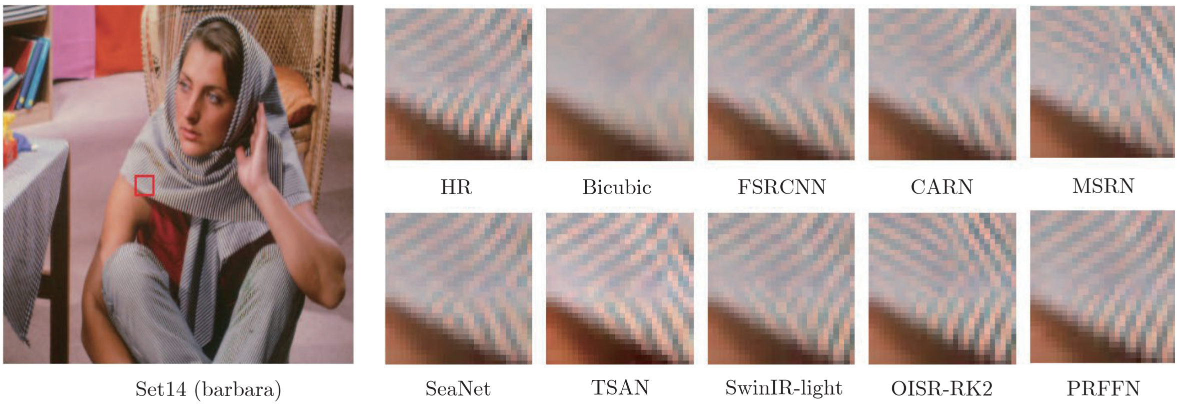

图 8 2 倍 SR 的视觉效果比较

Fig. 8 Visual comparison for 2 times SR

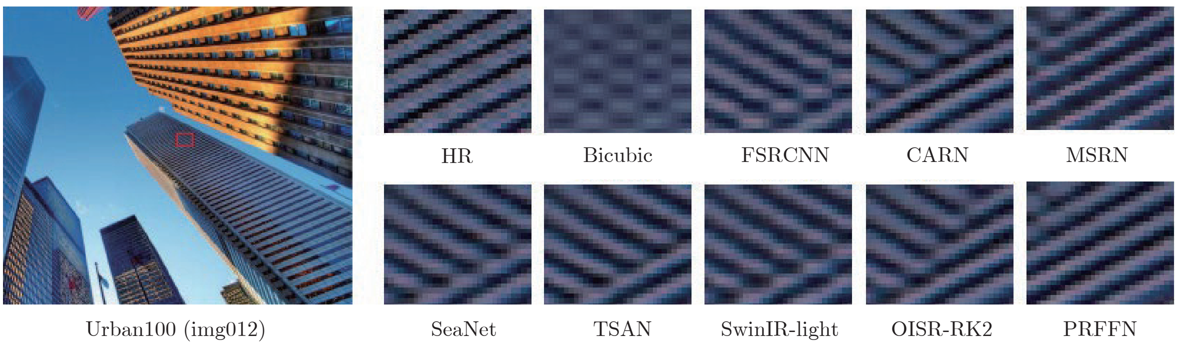

图 9 3 倍 SR 的视觉效果比较

Fig. 9 Visual comparison for 3 times SR

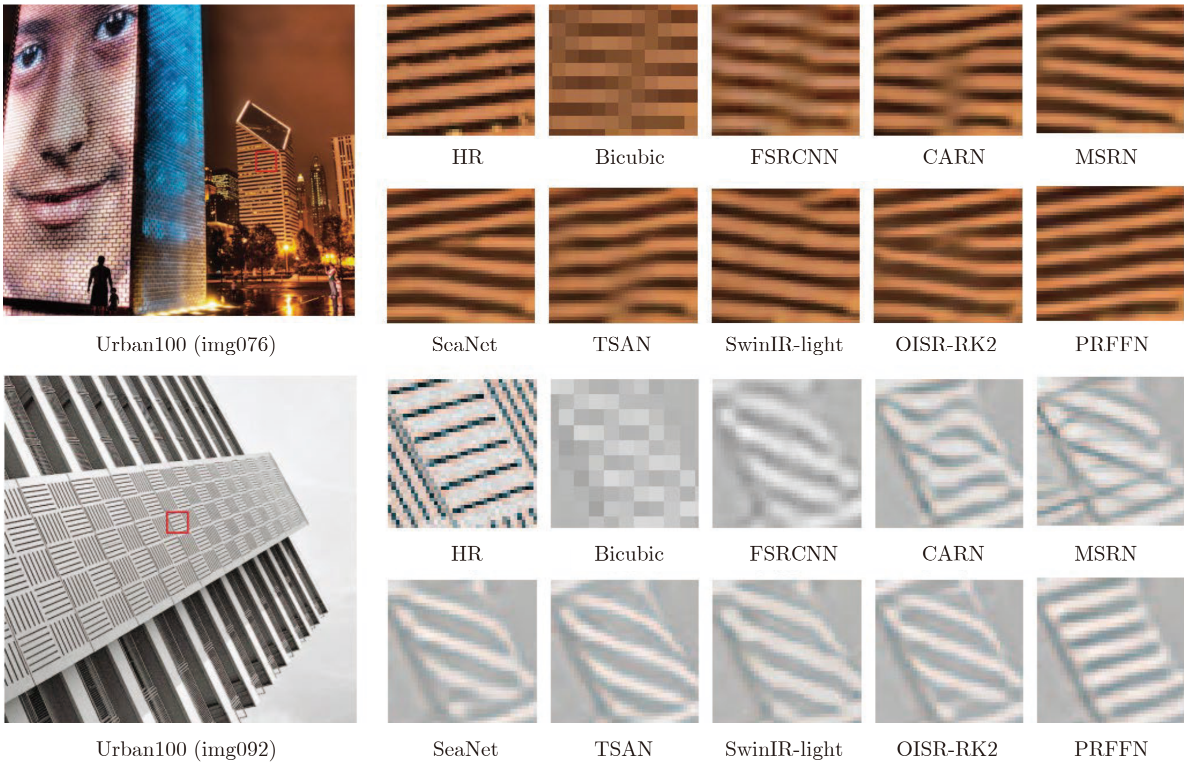

图 10　4 倍 SR 的视觉效果比较

Fig. 10　Visual comparison for 4 times SR

集中的图像 img076, 其他方法恢复的 SR 图像存在过度模糊或条纹方向错误, 本文方法恢复出的 SR 图像接近于原 HR 图像; 其他图像的恢复结果也类似.

5　结束语

本文提出一个通用的基于对比学习的逐像素损失函数 L_{cntr}, 以恢复出的 SR 图像像素作为锚样本, 原 HR 图像对应的像素作为正样本, 其他像素作为负样本. L_{cntr} 使锚样本尽可能接近正样本, 并远离负样本, 可以显著改进 SR 图像的保真度和视觉质量. 为了验证 L_{cntr} 的性能, 本文提出一个 SR 的网络架构 PRFFN. 实验结果表明, 本文的 PRFFN 组合 L_{cntr} 取得了很有竞争力的 SR 性能.

References

1　Freeman W T, Pasztor E C, Carmichael O T. Learning low-level vision. *International Journal of Computer Vision*, 2000, **40**(1): 25−47

2　Thornton M W, Atkinson P M, Holland D A. Subpixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super-resolution pixel-swapping. *International Journal of Remote Sensing*, 2006, **27**(3): 473−491

3　Zou W, Yuen P C. Very low resolution face recognition problem. *IEEE Transactions on Image Processing*, 2011, **21**(1): 327−340

4　Shi W, Caballero J, Ledig C, Zhuang X, Bai W, Bhatia K, et al. Cardiac image super-resolution with global correspondence using multi-atlas patch-match. In: Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention. Nagoya, Japan: 2013. 9–16

5　Chang H, Yeung D Y, Xiong Y M. Super-resolution through neighbor embedding. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC, USA: IEEE, 2004. 1−8

6　Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations. In: Proceedings of the International Conference on Curves and Surfaces. Avignon, France: 2010. 711–730

7　Timofte R, De Smet V, Van Gool L. Anchored neighborhood regression for fast example-based super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia: 2013. 1920–1927

8　Dong C, Loy C C, He K M, Tang X O. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **38**(2): 295–307

9　Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: 2016. 1646–1654

10　Lim B, Son S, Kim H, Nah S, Lee K M. Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Hawaii, USA: 2017. 136–144

11　Zhang Y L, Li K P, Li K, Wang L C, Zhong B N, Fu Y. Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision. Munich, Germany: 2018. 286–301

12　Tong T, Li G, Liu X J, Gao Q Q. Image super-resolution using dense skip connections. In: Proceedings of the IEEE International-al Conference on Computer Vision. Venice, Italy: 2017. 4799−4807

13 Zhang Y L, Li K P, Li K, Zhong B N, Fu Y. Residual non-local attention networks for image restoration. In: Proceedings of the International Conference on Learning Representations. New Orleans, USA: 2019. 1−18

14 Fan Y C, Shi H H, Yu J H, Liu D, Han W, Yu H C, et al. Balanced two-stage residual networks for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Hawaii, USA: 2017. 161−168

15 Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution. In: Proceedings of the European Conference on Computer Vision. Amsterdam, Netherlands: 2016. 694−711

16 Karen S, Andrew Z. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the 3rd International Conference on Learning Representations, Conference Track Proceedings. San Diego, USA: 2015. 7−9

17 Wang Y, Lin S, Qu Y, Wu H, Zhang Z, Xie Y, et al. Towards compact single image super-resolution via contrastive self-distillation. In: Proceedings of the 30th International Joint Conference on Artificial Intelligence. Montreal, Canada: 2021. 1122−1128

18 Wang Z, Bovik A C, Sheikh H R, Simoncelli E P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004, **13**(4): 600−612

19 Sajjadi M S, Scholkopf B, Hirsch M. EnhanceNet: Single image super-resolution through automated texture synthesis. In: Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: 2017. 4491−4500

20 Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA: 2017. 4681−4690

21 Fisher Y, Vladlen K. Multi-scale context aggregation by dilated convolutions. In: Proceedings of the International Conference on Learning Representations. San Juan, Puerto Rico: 2016. 1−13

22 Woo S, Park J, Lee J Y, Kweon I S. CBAM: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision. Munich, Germany: 2018. 3−19

23 Zhang Y L, Tian Y P, Kong Y, Zhong B N, Fu Y. Residual dense network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: 2018. 2472−2481

24 Niu B, Wen W L, Ren W Q, Zhang X D, Yang L P, Wang S Z, et al. Single image super-resolution via a holistic attention network. In: Proceedings of the European Conference on Computer Vision. Glasgow, UK: 2020. 191−207

25 Li J C, Fang F M, Mei K F, Zhang G X. Multi-scale residual network for image super-resolution. In: Proceedings of the European Conference on Computer Vision. Munich, Germany: 2018. 517−532

26 Liang J Y, Cao J Z, Sun G L, Zhang K, Van Gool L, Timofte R. SwinIR: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: 2021. 1833−1844

27 Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, et al. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: 2021. 10012−10022

28 Gao D D, Zhou D W. A very lightweight and efficient image super-resolution network. *Expert Systems With Applications*, 2023, **213**: Article No. 118898

29 He K M, Fan H Q, Wu Y X, Xie S N, Girshick R. Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: 2020. 9729−9738

30 Sermanet P, Lynch C, Chebotar Y, Hsu J, Jang E, Schaal S, et al. Time-contrastive networks: Self-supervised learning from video. In: Proceedings of the IEEE International Conference on Robotics and Automation. Queensland, Australia: IEEE, 2018. 1134−1141

31 Wu G, Jiang J J, Liu X M, Ma J Y. A practical contrastive learning framework for single image super-resolution. arXiv preprint arXiv: 2111.13924, 2021.

32 Wu H Y, Qu Y Y, Lin S H, Zhou J, Qiao R Z, Zhang Z Z, et al. Contrastive learning for compact single image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual Event: 2021. 10551−10560

33 Wang L G, Wang Y Q, Dong X Y, Xu Q Y, Yang J G, An W, et al. Unsupervised degradation representation learning for blind super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual Event: 2021. 10581−10590

34 Shi W Z, Caballero J, Huszár F, Totz J, Aitken A P, Bishop B, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: 2016. 1874−1883

35 Liu J, Zhang W J, Tang Y, Tang J, Wu G S. Residual feature aggregation network for image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: 2020. 2359−2368

36 Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: Proceedings of the International Conference on Machine Learning. Virtual Event: 2020. 1597−1607

37 Xie Z D, Lin Y T, Zhang Z, Cao Y, Lin S, Hu H. Propagate yourself: Exploring pixel-level consistency for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual Event: 2021. 16684−16693

38 Oord A V D, Li Y Z, Vinyals O. Representation learning with contrastive predictive coding. arXiv preprint arXiv: 1807.03748, 2018.

39 Timofte R, Agustsson E, Van Gool L, Yang M H, Zhang L. Ntire 2017 challenge on single image super-resolution: Methods and results. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Hawaii, USA: 2017. 114−125

40 Bevilacqua M, Roumy A, Guillemot C, Alberi-Morel M. Low-complexity single-image super-resolution based on non-negative neighbor embedding. In: Proceedings of the British Machine Vision Conference. Surrey, UK: 2012. 1−10

41 Martin D, Fowlkes C, Tal D, Malik J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of the 8th IEEE International Conference on Computer Vision. Vancouver, Canada: IEEE, 2001. 416−423

42 Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: 2015. 5197−5206

43 Yusuke M, Ito K, Aramaki Y, Fujimoto A, Ogawa T, Yamasaki T, et al. Sketch-based manga retrieval using Manga109 dataset. *Multimedia Tools and Applications*, 2017, **76**(20): 21811−21838

44 Zhang R, Isola P, Efros A A, Shechtman E, Wang O. The un-

reasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: 2018. 586−595

45 Dong C, Loy C C, Tang X O. Accelerating the super-resolution convolutional neural network. In: Proceedings of the European Conference on Computer Vision. Amsterdam, Netherlands: 2016. 391–407

46 Wang L G, Dong X Y, Wang Y Q, Ying X Y, Lin Z P, An W, et al. Exploring sparsity in image super-resolution for efficient inference. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual Event: 2021. 4917–4926

47 Zhou D W, Chen Y M, Li W B, Li J X. Image super-resolution based on adaptive cascading attention network. *Expert Systems With Applications*, 2021, **186**: Article No. 115815

48 Li Z, Wang C F, Wang J, Ying S H, Shi J. Lightweight adaptive weighted network for single image super-resolution. *Computer Vision and Image Understanding*, 2021, **211**: 103−254

49 Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: 2016. 1637–1645

50 Ahn N, Kang B, Sohn K A. Fast, accurate, lightweight super-resolution with cascading residual network. In: Proceedings of the European Conference on Computer Vision. Munich, Germany: 2018. 252–268

51 He X Y, Mo Z T, Wang P S, Liu Y, Yang M Y, Cheng J. Ode-inspired network design for single image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: 2019. 1732–1741

52 Fang F M, Li J C, Zeng T Y. Soft-edge assisted network for single image super-resolution. *IEEE Transactions on Image Processing*, 2020, **29**: 4656−4668

53 Zhang J Q, Long C J, Wang Y X, Piao H Y, Mei H Y, Yang X, et al. A two-stage attentive network for single image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, **32**(3): 1020−1033

周登文 华北电力大学控制与计算机工程学院教授. 主要研究方向为图像去噪, 图像去马赛克, 图像插值和图像超分辨率. 本文通信作者.
E-mail: zdw@ncepu.edu.cn
(**ZHOU Deng-Wen** Professor at the School of Control and Computer Engineering, North China Electric Power University. His research interest covers image denoising, image demosaic, image interpolation, and image super-resolution. Corresponding author of this paper.)

刘子涵 华北电力大学控制与计算机工程学院硕士研究生. 主要研究方向为计算机视觉, 深度学习.
E-mail: 120212227102@ncepu.edu.cn
(**LIU Zi-Han** Master student at the School of Control and Computer Engineering, North China Electric Power University. His research interest covers computer vision and deep learning.)

刘玉铠 华北电力大学控制与计算机工程学院硕士研究生. 主要研究方向为计算机视觉, 深度学习.
E-mail: liuyk@ncepu.edu.cn
(**LIU Yu-Kai** Master student at the School of Control and Computer Engineering, North China Electric Power University. His research interest covers computer vision and deep learning.)