自适应融合层级特征的混合退化图像复原算法

白亮¹⁾, 刘辉^{1,2)}, 尚振宏^{1,3)*}

¹⁾ (昆明理工大学信息工程与自动化学院 昆明 650500)

²⁾ (中国科学院云南天文台 昆明 650216)

³⁾ (昆明理工大学云南省人工智能重点实验室 昆明 650500)
(shangzhenhong@126.com)

摘要: 多种退化类型混合的图像比单一类型的退化图像降质更严重, 很难建立精确模型对其复原, 研究端到端的神经网络算法是复原的关键. 现有的基于操作选择注意力网络的算法(operation-wise attention network, OWAN)虽然有一定的性能提升, 但是其网络过于复杂, 运行较慢, 复原图像缺乏高频细节, 整体效果也有提升的空间. 针对这些问题, 提出一种基于层级特征融合的自适应复原算法. 该算法直接融合不同感受野分支的特征, 增强复原图像的结构; 用注意力机制对不同层级的特征进行动态融合, 增加模型的自适应性, 降低了模型冗余; 另外, 结合 L_1 损失和感知损失, 增强了复原图像的视觉感知效果. 在 DIV2K, BSD500 等数据集上的实验结果表明, 该算法无论是在峰值信噪比和结构相似性上的定量分析, 还是在主观视觉质量方面, 均优于 OWAN 算法, 充分证明了该算法的有效性.

关键词: 自适应复原; 混合退化; 层级特征融合; 感知损失

中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2021.18482

Mixed Degraded Image Restoration Algorithm Based on Adaptive Fusion of Hierarchical Features

Bai Liang¹⁾, Liu Hui^{1,2)}, and Shang Zhenhong^{1,3)*}

¹⁾ (Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500)

²⁾ (Yunnan Observatories, Chinese Academy of Sciences, Kunming 650216)

³⁾ (Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500)

Abstract: The degradation of mixed degraded images is more serious than that of single degradation types, and it is difficult to restore them by precise modeling. The key to restore mixed degraded images is to study the end-to-end neural network algorithm. Although the existing operation-wise attention network (OWAN) algorithm has a certain performance improvement, its network is too complex, it runs slowly, the restored image lacks high-frequency details, and the overall effect also has room for improvement. To solve these problems, an adaptive restoration algorithm based on hierarchical feature fusion is proposed. The algorithm directly fuses the features of different receptive field branches to enhance the structure of the restored image. The attention mechanism is used to dynamically fuse the features of different hierarchies to increase the adaptability and reduce the redundancy of the model. In addition, combining the L_1 loss and perception loss, the visual perception effect of the restored image is enhanced. Experimental results on DIV2K, BSD500 and other data sets show that the proposed algorithm is better than the OWAN algorithm in terms of quantitative analysis of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), as well as subjective visual quality.

收稿日期: 2020-07-09; 修回日期: 2020-08-12. **基金项目:** 国家自然科学基金(12063002, 11873027, 61462052); 云南省重大科技专项计划(202002AD080001). 白亮(1994—), 男, 硕士研究生, 主要研究方向为计算机视觉、模式识别; 刘辉(1969—), 男, 博士, 研究员, 硕士生导师, 主要研究方向为机器学习、数据挖掘; 尚振宏(1975—), 男, 博士, 副教授, 硕士生导师, CCF 会员, 论文通讯作者, 主要研究方向为计算机视觉、图像处理.

Key words: adaptive restoration; mixed degradation; hierarchical feature fusion; perceptual loss

图像复原主要研究如何从退化图像中复原出高质量的清晰图像,是数字图像处理领域的研究热点之一,也是目标识别、检测跟踪、图像语义分析及转换等高层任务的基石。由于图像的退化过程不可逆,图像复原是一个典型的欠定问题。目前,图像复原技术主要包括基于退化理论模型的传统算法和基于统计拟合的深度学习算法。传统算法依赖对先验知识的了解程度和精确性,稳定性有局限,而且算法效率较低,很难适用于大数据背景下的海量数据高速复原。近几年,由于出色的性能和强大的学习能力,基于卷积神经网络的深度学习算法越来越受到图像复原研究者的欢迎,在各种复原任务中取得了令人瞩目的成绩。

针对噪声、模糊等单一退化类型影响的图像,研究者提出了很多不错的复原算法^[1-4]。然而,真实场景下往往出现多种不同退化类型混合的情况,如自动驾驶汽车、监控摄像机等得到的图像数据,受到多种退化类型未知强度的混合。与单一类型的退化图像相比,混合退化图像的降质更严重,退化情况组合非常多,很难建立精确模型对其复原,常见的效果较好的单一退化类型的图像复原算法应用于混合退化图像复原时效果也并不理想,因此研究端到端的神经网络图像复原算法十分重要。高斯噪声、高斯模糊和 JPEG 压缩 3 种退化的混合是一种比较常见的混合退化场景,针对这 3 种退化类型的随机混合,基于强化学习的复原算法 RL-Restore^[5]构造一个包含特定于不同复原任务的小规模卷积网络工具箱,通过强化学习方法学习复原策略,能根据不同的待复原图像选择适当的工具组成动态工具链逐步恢复损坏图像。基于操作选择注意力网络的复原算法 OWAN^[6]在其核心的选择模块中并行执行 8 种由不同卷积和池化等组成的基础操作进行特征提取,并用注意力机制对提取的特征进行加权组合。

作为目前已知效果最好的混合退化图像复原算法,OWAN 虽然比 RL-Restore 有一定的性能提升,但是其本质上仍是一种通过增加网络的复杂度来提升算法效果的方法,用并行执行多种不同操作的方式提高算法中每一层学习到的特征的多样性,并没有对不同层级特征之间的互补性和可

重用性加以利用优化,导致网络模型过于复杂。虽然使用大量的分离卷积操作降低模型的参数规模,但是也加大了训练和测试阶段的时间速率损失。文献[7-12]已经验证,充分利用层级特征的相关特性对降低网络冗余、提升算法效果具有较大帮助,已经成为一种比较通用的做法。另外,OWAN 算法复原的图像纹理过于平滑,缺乏高频细节,整体复原质量也有进一步的提升空间。为了解决这些问题,本文提出一种自适应融合层级特征的混合退化图像复原算法(hierarchical-feature fusion adaptive networks, HFANet),其代码已经发布到开源网站^①,本文的主要工作如下。

(1) 将不同感受野的空洞卷积分支提取的结构特征进行直接融合,形成结构特征融合块,增强复原图像的结构信息。

(2) 用多个顺序连接的小尺寸卷积层提取层级特征,并用注意力机制对不同层级的特征进行动态加权融合,形成层级特征动态融合块,充分利用层级特征的可重用性和互补性,降低模型冗余,增加模型的自适应性,同时加快模型的运行效率。

(3) 将 L_1 损失和感知损失组合,作为算法的优化目标,提高复原图像的视觉感知效果。

1 相关工作

1.1 注意力机制

人类的视觉系统倾向关注图像中感兴趣的部分而忽略不相关的信息,在计算机视觉中,注意力机制可被视为一种视觉指导,能对图像中信息量较大的部分施加更多的关注,分配更多的计算资源。常见的形式是对输入的数据进行特征提取、非线性映射形成动态权重,对网络中不同通道层的特征进行加权,从而形成对特定通道层的注意力关注。注意力机制可以增强网络模型的动态性和提高算法的性能,已被广泛应用于超分辨率处理、图像翻译等多种计算机视觉任务中^[13-15],并取得了显著的效果。这些基于注意力机制的神经网络算法的共同特点是,可以根据输入内容动态调整不同通道或区域特征的重要程度,从而提高模型的自适应性。

① <https://github.com/beiyan1911/HFANet>

1.2 基于层级特征融合的图片复原

深层卷积网络算法的显著特点是通常由多层卷积、残差块或其他基础模块堆叠而成, 可以对较复杂的非线性变化进行拟合, 这样就会在多个基础模块间或每个基础模块内形成不同层级的特征: 低层级的特征包含细节信息较多, 语义信息较少; 高层级的特征包含的语义信息更强, 细节信息较少. 如何对层级特征的特点加以充分地利用, 是提高算法效果的一个重要因素. Huang 等^[16]最早提出一种基于层级特征密集连接的卷积网络架构, 用密集连接机制将不同层级的特征相结合, 加强了特征的融合与传播, 在图像分类任务中取得明显的精度提升. 之后, 各种基于层级特征结合方式的改进很快被研究者引入图像复原中. Zhang 等^[7]提出一种基于层级特征密连接的深度残差网络(residual dense networks, RDN), 将残差块(residual block, RB)与密集连接机制相结合, 明显提升了图像超分辨率的效果. Chen 等^[8]提出一种将 Boosting 算法与层级特征稠密连接相结合的图像去噪算法(deep Boosting framework, DBF), 取得较好的图像去噪效果. 这 2 种算法以及文献[9-12]都充分证实, 不同层级特征之间的整合利用已经成为提升卷积网络复原算法的一种通用做法. 考虑现有的融合算法中层级特征连接过于密集、模型运行时资源占用过多, 本文在局部模块内用增强的注意力机制将不同层级特征动态加权融合, 既保证层级特征之间的复用和模块的轻量化, 又能提高混合退化图

像复原算法的自适应性.

1.3 感知损失

感知损失最早由 Johnson 等^[17]在图像风格迁移算法中提出, 用于增加风格迁移结果的视觉感知效果. 感知损失从预训练的分类网络中提取高维的图像特征, 用高维特征之间的距离作为损失函数对算法进行优化, 与 L_1 等逐像素比较的损失函数相比, 得到的结果图像具有更多的细节, 整体视觉效果更好. Kupyn 等^[18]将感知损失引入图像去模糊算法中, Ledig 等^[19]将感知损失引入图像超分辨率处理中, 都得到了具有更多细节和更高视觉感知质量的结果图像. 由于混合退化图像细节信息丢失更严重, 复原难度更大, 因此本文将 L_1 损失与感知损失进行加权结合, 用于对图像轮廓等高级语义信息进行增强, 以提升整体的复原效果.

2 本文算法

2.1 混合退化图像复原网络

如图 1 所示, 本文算法采用端到端的网络模型框架, 分为特征提取块、15 个顺序连接的动态适应块(dynamic adaptive block, DAB)和输出块 3 部分. 特征提取块用于提取多维度浅层图像特征, 多个动态适应块对浅层特征进行自适应提取融合与增强, 输出块组合输出经过增强的特征. 算法输入混合退化图像 I_{lr} , 输出复原之后的清晰图像 I_{hr} .

图 1 网络架构图

为了使算法最优化, 本文的目标函数由 L_1 损失和感知损失 2 部分构成. 最小化 L_1 损失就是最小化输出图像 I_{hr} 和高清标签图像 I_{label} 之间的像素距离, 会使复原的结果具有较好的峰值信噪比(peak signal-to-noise ratio, PSNR)和结构相似性(structural similarity, SSIM). 本文使用的 L_1 损失表示为

$$L_1 = \frac{1}{N} \sum_{n=1}^N \| \text{HFANet}(I_{lr}) - I_{label} \|_1 \quad (1)$$

其中, HFANet 表示本文算法; N 表示训练样本的个数.

最小化感知损失可以使输出图像与标签图像的高级语义信息更接近, 缓解只使用 L_1 损失带来的复原结果过于平滑的问题. 本文使用 VGG19 模型^[20]的第 1 层卷积和第 6 层卷积作为特征提取器, 用 ϕ_1 和 ϕ_6 表示, 分别提取浅层和深层的图像语义信息. 构造的感知损失为

$$L_{\text{perceptual}} = \frac{1}{N} \sum_{n=1}^N \| \phi_1(I_{hr}) - \phi_1(I_{label}) \|_1 + \frac{1}{N} \sum_{n=1}^N \| \phi_6(I_{hr}) - \phi_6(I_{label}) \|_1 \quad (2)$$

本文算法的总体优化目标为

$$\hat{\theta} = \arg \min_{\theta} (\alpha L_1 + \beta L_{\text{perceptual}}) \quad (3)$$

其中, θ 表示算法的参数集合; α 和 β 为 L_1 损失和感知损失的权重. 为了综合 L_1 损失和感知损失对复原效果的提升, 本文实验中取 $\alpha=1$, $\beta=0.0006$.

2.2 动态适应块

由于高斯噪声、高斯模糊和 JPEG 压缩进行未知强度随机混合的组合情况非常多, 常规的复原算法很难学习拟合该复杂情况, 而且极端情况下图像的高频细节和纹理可能严重丢失, 给复原带来巨大挑战. 为此, 本文将不同大小感受野的空洞卷积分支特征进行直接融合, 组成结构特征融合块(structure-feature fusion block, SFB)对复原图像结构进行增强. 同时, 为了增加算法自适应性和泛化能力, 使其能对尽可能多的混合退化组合进行统计学习, 设计了用注意力机制对层级特征进行自适应融合的层级特征动态融合块(hierarchical-feature dynamic fusion block, HDFB). SFB 和 HDFB 组合形成 DAB.

2.2.1 结构特征融合块

Anwar 等^[21]最早在图像去噪算法中提出合并-运行单元, 即本文中引入的 SFB. 如图 2 所示, Dilated Conv $\times p$ 表示空洞大小为 p 的空洞卷积操作, SFB 的主体部分包含小感受野分支和大感受野分支, 2 个分支分别由不同大小的空洞卷积和 ReLU 组成, 提取的特征经过拼接与合并操作, 再与输入的特征相加后输出. 由于采用直接融合不同感受野分支特征的方式, 模块可以学习到丰富的结构特征. 用 f_{k1} 和 f_{k2} 分别表示第 k 个 DAB 中 SFB 的输入和输出特征, 用 T_1 和 T_2 分别表示小感受野分支和大感受野分支, 用 C_1 表示拼接与合并操作, 则 SFB 可表示为

$$f_{k2} = C_1([T_1(f_{k1}), T_2(f_{k1})]) + f_{k1} \quad (4)$$

图 2 结构特征融合块

2.2.2 层级特征动态融合块

为了充分利用层级特征的特性, 本文通过注意力机制动态融合 HDFB 中不同层级的特征, 克服常规算法只增大每一层内不同操作的个数导致网络过于复杂的问题, 从而使算法更轻量化. 如图 3a 所示, HDFB 的左上角是注意力层(attention layer), 输入特征 f_{k2} , 输出注意力权重 f_{kw} . 受卷积网络中常用多个小尺寸卷积代替大尺寸卷积的启发, 本文使用由 3×3 卷积和 ReLU 组成的 5 个顺序连接的基本块(base block)进行特征提取, 将 5 个基本块提取到的特征组合形成不同层级的特征 f_{ksd} , f_{ksd} 与 f_{kw} 相乘, 得到动态加权的层级特征组合; 再经过拼接与合并操作, 与输入特征 f_{k2} 相加进行最后输出. 用 $B_1 \sim B_5$ 分别表示 5 个基本块, 用 A 表示注意力层, C_2 表示拼接与合并操作, 则 HDFB 可表示为

$$f_{k3} = C_2(A(f_{k2}) \times [B_1(f_{k2}), B_2(B_1(f_{k2})), \dots, B_5(B_4(\dots))] + f_{k2} \quad (5)$$

与常见的只使用全局平均池化和非线性映射生成注意力权重的方式不同, 本文结合全局最大值池化和全局平均值池化, 使生成的注意力权重具有更强的表示能力. 如图 3b 所示, 注意力层分为平均值池化分支 T_a 和最大值池化分支 T_m , 2 个分支的输出特征相加, 再经过 Softmax 处理形成注意力权重. 注意力层可表示为

$$f_{kw} = \text{Softmax}(T_a(f_{k2}) + T_m(f_{k2})) \quad (6)$$

a. 层级特征动态融合块

图 3 层级特征动态融合块结构图

3 实验分析和讨论

3.1 实验设置及数据集

基于 Windows 10 操作系统, 使用 Python 3.7 和 PyTorch 1.5 框架^[22]进行实验, 硬件采用 Intel Xeon E5-2678 v3@2.50 GHz 处理器, 32 GB 内存, NVIDIA RTX 2080Ti 显卡, 12 GB 显存, 使用 CUDA 10.1 和 CuDNN 7.0 模块实现 GPU 加速。

模型使用 Adam^[23]作为参数优化器, 训练集总迭代次数为 100, 学习率从初始值 10^{-3} 按照余弦退火方式(cosine annealing)逐渐衰减到 10^{-6} , 训练时 batch 为 48. 本文使用 PSNR 和 SSIM 作为评价指标, 使用与 RL-Restore^[5]相同的数据集 DIV2K^[24]进行训练和测试, 实验结果和目前已知效果最好的混合退化图像复原算法 OWAN^[6]进行比较。

DIV2K 数据集包含 800 幅高清图像, 涵盖从人、手工作品到环境、动植物等丰富的内容, 经常被作为基准数据集使用. 与 RL-Restore^[5]相同, 实验用前 750 幅图像制作训练集, 后 50 幅图像制作测试集, 每幅高清图像加入高斯模糊、高斯噪声和 JPEG 压缩 3 种退化类型的随机混合, 并根据退化程度划分为轻微退化(Mild)、中等退化(Moderate)和严重退化(Severe), 然后裁剪成多幅小块图像. 具体来说, 就是将高斯模糊标准差、高斯噪声标准差和 JPEG 压缩质量分别在 $[0, 5]$, $[0, 50]$ 和 $[10, 100]$ 均分为 11 档, 通过随机算法生成大量的 3 种退化档位组合, 去掉较小和较大的情况, 将档位和在 $[9, 20]$ 的退化组合分为 3 种程度: 档位和在 $[9, 11]$ 为 Mild; 在 $[12, 17]$ 为 Moderate; 在 $[18, 20]$ 为 Severe. 另外, 还对每种组合中的退化数值添加随机扰动, 使退化数值尽可能均匀覆盖整个范围. 最终制作的训练集包含 230139 幅 63×63 像素的 Moderate 图像, 测试集包含轻微、中等、严重 3 种退化程度, 每种包含 3584 幅 63×63 像素的图像. 模型在 Moderate 程度的训练集上训练, 在 Mild, Moderate 和 Severe 3 种退化程度的测试集上进行测试。

3.2 定量比较

3.2.1 在测试集上对比

为了验证本文算法在混合退化图像上的复原效果, 用训练好的算法模型在测试集上进行测试, 并将测试结果与 OWAN^[6]算法进行对比. 通过表 1 中的对比可以看出, 本文算法在 Mild, Moderate, Severe 3 种退化程度的混合退化图像复原任务上均有更高的 PSNR 和 SSIM.

表 1 2 种算法的 PSNR 和 SSIM 对比结果

算法	Mild		Moderate		Severe	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
OWAN ^[6]	28.33	0.7455	27.07	0.6787	25.88	0.6167
本文	28.40	0.7502	27.10	0.6804	25.91	0.6179

注: 表中加粗的数据为表现优异的指标。

3.2.2 在其他数据集上的泛化性实验对比

为了测试本文算法的泛化性, 在 BSD500^[25]和 Set14 数据集上分别随机添加高斯模糊、高斯噪声、JPEG 压缩 3 种单一退化以及同时包含 3 种的混合退化, 制作成 Blur, Noise, JPEG 和 Mix 共 4 个测试集. 3 种单一退化的测试集是由清晰图像随机添加 $[0, 5]$ 均匀分布的高斯模糊、 $[0, 50]$ 均匀分布的高斯噪声、 $[10, 100]$ 均匀分布的 JPEG 压缩形成, 混合退化测试集 Mix 的制作方法同第 3.1 节. 分别用本文算法和 OWAN 算法在 4 个测试集上进行测试, 对比结果如表 2 所示。

表 2 2 种算法在 2 个数据集上的实验结果对比

算法	测试集	BSD500		Set14	
		PSNR/dB	SSIM	PSNR/dB	SSIM
OWAN ^[6]	Mix	23.46	0.6719	22.04	0.6577
	Blur	20.36	0.6481	16.89	0.6105
	Noise	25.35	0.7379	24.58	0.7213
	JPEG	24.37	0.7851	21.66	0.7542
本文	Mix	25.81	0.7134	25.52	0.7209
	Blur	25.07	0.7099	24.95	0.7094
	Noise	24.65	0.6790	23.11	0.6519
	JPEG	28.94	0.8379	28.06	0.8305

注: 表中加粗的数据为表现优异的指标。

通过表 2 中定量对比可以看出, 本文算法在 Mix, Blur 和 JPEG 这 3 个测试集的复原任务上, PSNR 和 SSIM 指标均高于 OWAN 算法; 在 Noise 测试集复原任务上, PSNR 和 SSIM 指标略低于 OWAN 算法. 在实际的效果对比中也发现, OWAN 算法对于小块退化图像的复原结果较稳定, 当退化图像尺寸较大时, 复原结果中就会出现一些蓝色褶皱, 一定程度降低了复原效果. 综合来说, 本文算法比 OWAN 算法具有更好的泛化性。

3.2.3 时间效率对比

在相同的软硬件条件下, 对包含 Mild, Moderate 和 Severe 3 种退化程度各 3584 幅图像的测试集, 分别记录本文算法和 OWAN 算法进行复原所消耗的时间, 测试时数据读取批次大小为 20, 数据读取线程数为 8, 结果如表 3 所示. 可以看出,

与 OWAN 算法相比, 本文算法较大幅度地加快了复原速率. 速率加快的原因, 一方面是本文算法使用的常规卷积比 OWAN 算法使用的可分离卷积执行速度更快; 更重要的是, 本文使用小尺寸卷积提取特征, 层级特征融合方式比较高效, 使模块更轻量化、算法运行速率更快.

表 3 2 种算法时间性能对比

算法	总耗时/s
OWAN ^[6]	225.60
本文	141.23

注: 表中加粗的数据为表现优异的指标.

3.3 定性比较

为了更直观地将本文算法的复原结果与 OWAN 算法进行对比, 分别从 2 种算法的复原图像中列出几组进行对比. 从图 4 可以看出, 图 4a 第 5 列中, 本文算法复原的蜘蛛图像比 OWAN 算法的结果具有更清晰的结构; 图 4b 第 5 列中, 本文算法复原的图像在护栏处具有更多的细节; 图 4c 第 3 列中, OWAN 算法复原的结果在蓝色背景处明显有多余的褶皱. 总体来说, 本文算法复原的图像更加逼真, 整体视觉效果更好, 这也证明了本文算法的有效性.

3.4 消融实验

3.4.1 算法动态适应性消融实验对比

为了探究层级特征动态融合块中, 注意力机制的动态加权对最终复原效果的影响, 分别进行使用和不使用注意力机制的消融实验. 具体做法是将本文算法中的注意力权重设置为一个固定值 (即在实验中将式(6)得到的注意力权重 f_{kw} 取固定值 0.2), 在相同环境和数据集下重新训练网络模型并进行测试, 将使用注意力机制的复原图像与没有使用注意力机制的复原图像进行对比, 对比结果如表 4 所示. 通过对比可以看出, 在相同条件下, 使用注意力机制的算法比不使用注意力机制的算法在 Mild, Moderate 和 Severe 3 种退化程度的测试集上均取得了更好的 PSNR 和 SSIM 指标. 这也说明本文用注意力机制实现不同层级特征动态融合的方式可以增强模型的自适应性, 提高复原图像的质量.

3.4.2 感知损失消融实验对比

为了验证感知损失对本文算法复原结果的影响, 分别在相同的实验环境下做了使用感知损失和不使用感知损失的消融实验, 将只使用 L_1 损失函数训练的复原结果与同时使用 L_1 损失和感知损失训练的复原结果进行对比. 从图 5 可以看出, 图

5c 第 1 行同时使用 L_1 损失和感知损失的复原结果在果盘处明显比图 5b 第 1 行只使用 L_1 损失的复原结果结构更加清晰; 图 5c 第 2 行的复原图像比图 5b 第 2 行的复原图像具有更清晰的细节结构; 图 5c 第 3 行的复原图像在衣服褶皱处比图 5b 第 3 行更接近图 5d 第 3 行的真实标签图. 这些都验证感知损失确实提升了本文算法复原结果的视觉感知质量.

图 4 2 种算法复原结果的主观对比

表 4 使用和不使用注意力机制的实验结果对比

注意力 机制	Mild		Moderate		Severe	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
不使用	28.26	0.7441	27.02	0.6778	25.82	0.6144
使用	28.40	0.7502	27.10	0.6804	25.91	0.6179

注: 表中加粗的数据为表现优异的指标.

图5 使用和不使用感知损失下本文算法的复原结果对比

3.5 讨论

本文从层级特征融合的角度去提高混合退化图像复原的效果, 实验分析对比也验证了这种思路的正确性以及本文算法的有效性. 但是, 与 OWAN 算法相比, 本文算法的提高并不是非常显著, 设备和算力的局限是一方面因素, 更多还是因为混合退化图像丢失的信息较多, 神经网络算法对这种欠定问题的统计学习能力也有一定的瓶颈. 另外, 在对算法的分析中发现, 注意力机制得到的权重确实会根据输入数据进行动态变化, 但是这种变化是围绕某一固定值周围进行非常轻微的波动, 注意力机制对算法性能的提升效果也并不显著, 这与当下注意力机制在各个研究方向大受欢迎的现象有些不符, 值得后续进行更深入的研究.

可分离形式的卷积操作更多地应用在资源比较紧张的移动设备上, 用以减少神经网络模型的参数. OWAN 算法使用大量的可分离卷积, 较大程度地降低了模型参数, 最终的模型只有 1.8MB, 但是算法复杂度并没有降低, 使用分离卷积也更加降低了算法的运行效率. 本文算法保存的模型大小为 5.7MB, 运行时占用的显存与 OWAN 算法基本相同, 但是运行速度比 OWAN 算法具有较大提升, 综合对比下, 本文算法更具优势.

4 结语

本文提出一种基于层级特征融合的混合退化图像自适应复原算法, 通过注意力机制动态融合不同层级的特征, 引入结构特征融合块将不同感受野分支的特征进行结合, 充分利用层级特征之间的互补性和可重用性, 解决了现有算法因增加算法复杂度导致网络过于复杂、耗时较长、性能有局限的问题. 在此基础上, 本文加权结合 L_1 损失和感知损失, 提高了复原图像的视觉感知效果. 在 DIV2K, BSD500 等数据集上的实验结果表明, 本文算法获得的 PSNR, SSIM 指标和视觉感知质量均优于 OWAN 算法, 有效地提升了混合退化复原算法的效果, 大幅提高了算法的运行速率.

参考文献(References):

- [1] Zhang K, Zuo W M, Chen Y J, *et al.* Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising[J]. IEEE Transactions on Image Processing, 2017, 26(7): 3142-3155
- [2] Chakrabarti A. A neural approach to blind motion deblurring[C] //Proceedings of European Conference on Computer Vision. Heidelberg: Springer, 2016: 221-235
- [3] Lehtinen J, Munkberg J, Hasselgren J, *et al.* Noise2Noise: learning image restoration without clean data[C] //Proceedings

- of the 35th International Conference on Machine Learning. Stockholm: PMLR Press, 2018: 2965-2794
- [4] Soltanayev S, Chun S Y. Training deep learning based denoisers without ground truth data[C] //Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM Press, 2018: 3257-3267
 - [5] Yu K, Dong C, Lin L, *et al.* Crafting a toolchain for image restoration by deep reinforcement learning[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 2443-2452
 - [6] Suganuma M, Liu X, Okatani T. Attention-based adaptive selection of operations for image restoration in the presence of unknown combined distortions[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 9301-9040
 - [7] Zhang Y L, Tian Y P, Kong Y, *et al.* Residual dense network for image restoration[C] //Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence. Los Alamitos: IEEE Computer Society Press, 2020: 1-16
 - [8] Chen C, Xiong Z W, Tian X M, *et al.* Deep boosting for image denoising[C] //Proceedings of European Conference on Computer Vision. Heidelberg: Springer, 2018: 3-19
 - [9] Huang Y W, Qin M. Densely connected high order residual network for single frame image super resolution[OL]. [2020-07-09]. <https://arxiv.org/pdf/1804.05902.pdf>
 - [10] Li C Y, Guo J C, Porikli F, *et al.* A cascaded convolutional neural network for single image dehazing[J]. IEEE Access, 2018, 6: 24877-24887
 - [11] Song Y D, Zhu Y F, Du X. Dynamic residual dense network for image denoising[J]. Sensors, 2019, 19(17): 3809
 - [12] Tong T, Li G, Liu X J, *et al.* Image super-resolution using dense skip connections[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 4809-4817
 - [13] Lu Y, Zhou Y, Jiang Z Q, *et al.* Channel attention and multi-level features fusion for single image super-resolution[OL]. [2020-07-09]. <https://arxiv.org/pdf/1810.06935.pdf>
 - [14] Liu D, Wen B H, Fan Y C, *et al.* Non-local recurrent network for image restoration[C] //Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM Press, 2018: 1680-1689
 - [15] Tang H, Xu D, Sebe N, *et al.* Multi-channel attention selection GAN with cascaded semantic guidance for cross-view image translation[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 2417-2426
 - [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] Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution[OL]. [2020-07-09]. <https://arxiv.org/pdf/1603.08155.pdf>
 - [18] Kupyn O, Budzan V, Mykhailych M, *et al.* DeblurGAN: blind motion deblurring using conditional adversarial networks[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 8186-8192
 - [19] 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
 - [20] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[OL]. [2020-07-09]. <https://arxiv.org/pdf/1409.1556.pdf>
 - [21] Anwar S, Barnes N. Real image denoising with feature attention[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 3155-3164
 - [22] Paszke A, Gross S, Massa F, *et al.* PyTorch: an imperative style, high-performance deep learning library[OL]. [2020-07-09]. <https://arxiv.org/pdf/1912.01703.pdf>
 - [23] Kingma D P, Ba J. Adam: a method for stochastic optimization[OL]. [2020-07-09]. <https://arxiv.org/pdf/1412.6980.pdf>
 - [24] 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. Los Alamitos: IEEE Computer Society Press, 2017: 1122-1131
 - [25] 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