文章编号:1671-6833(2026)03-0100-08

基于多尺度动态滤波的图像增强模型

尹 毅,吕 培,李凯江,郑昊坤,徐 豪,陈梦婕

(郑州大学 计算机与人工智能学院,河南 郑州 450001)

摘 要:为了解决传统图像增强方法中难以同时兼顾全局平滑与局部纹理细节的问题,提出了一个基于多尺度动态滤波分解的 MDFD 图像增强模型。首先,利用可学习的低通滤波器和高通滤波器来分别提取图像的低频与高频图像分量;其次,结合这两种频域图像分量,提出了跨低频通道注意力融合模块(LFCA)和跨高频空间注意力融合模块(HFSA),以实现图像全局与局部的协同增强;最后,通过引入多尺度融合策略,综合利用不同尺度下的高频和低频信息进行特征融合。多尺度融合的优点在于能够通过有效整合不同尺度上的细节和全局特征,在多个层面显著提升图像的增强效果。实验结果表明:MDFD 模型在 FiveK 和 PPR10K 数据集上的验证中表现出色,其中峰值信噪比(*PSNR*)分别达到 25.90 和 27.35,结构相似性指数(*SSIM*)分别为 0.964 和 0.945,ΔE_{ab} 分别为 7.38 和 6.50。这表明 MDFD 模型在复杂环境和颜色丰富等场景下具有优越的图像增强性能。

关键词:图像增强; 低通滤波; 高通滤波; 多尺度融合; 频域变换

中图分类号:TP37;TP391.9 **文献标志码**:A **doi**:10.13705/j.issn.1671-6833.2025.03.016

图像增强作为计算机视觉的核心任务之一,旨在提升图像的对比度、细节清晰度和色彩表现,从而使其在视觉上更加清晰,进而改善后续的视觉处理效果。目前,图像增强方法主要分为传统方法和基于神经网络的方法两大类。传统的图像增强方法依赖于手工设计的规则与图像统计特性,典型的技术包括直方图均衡[1]和伽马校正[2]等。但是,这些方法无法充分捕捉图像的高层次语义信息,在处理复杂场景时表现不佳。与之相比,基于神经网络的图像增强方法利用大量数据训练模型来学习图像的复杂特征,使得这类方法在处理局部细节和全局结构时更加智能和高效,尤其在复杂多变的图像场景中展现了更优的特性。

Chen 等[3]提出了一种轻量化的快速重参数残差网络(FRR-NET)用于低光图像增强。通过设计轻量化残差块和基于 Transformer 的亮度增强模块解决亮度增强后图像特征退化的问题。Zhou 等[4]提出了一种高效的全引导信息流网络(UGIF-Net),通过多色彩空间引导的颜色估计模块,结合两个色彩空间的信息,精确恢复颜色,旨在解决颜色恢复准确性和抗干扰能力的不足。Shen 等[5]提出了一种基于卷积神经网络和 Retinex 理论的低光图像增强模型,将低光图像增强视为一个机器学习问题,直接学习暗光与亮光图像之间的端到端映射。基于可学习的 3D LookUpTable 的图像增强方法[6-7],通过可学习的权重比例动态计算对应的 3D LookUpTable,能够实时将退化的图像进行逐像素增强。RSFNet[8]采用一个可以并行区域特定滤镜进行照片修饰的方法。该模型同时生成每个过滤器的参数(例如饱和度、对比度、色调),从而可以更轻松地训练更多样化的滤波器类别。虽然基于白盒的方式可以针对特定属性针对性的增强,但缺乏属性间的交互增强,导致局部细节关联性减弱。

尽管上述基于神经网络的图像增强方法在不同场景下取得了较好的效果,但它们普遍存在一个共同的缺陷,即忽视了图像整体图像的平滑性与局部纹理细节之间的关系处理。在图像增强过程中,如何有效协同处理全局结构与局部细节,以实现图像

收稿日期:2025-02-21;**修订日期**:2025-07-21

基金项目:国家自然科学基金资助项目(62372415);装备预研教育部联合基金项目(8091B032257);河南省杰出青年科学基金项目(242300421050)

通信作者:吕培(1986—)男,河南孟州人,郑州大学教授,博士,博士生导师,主要从事人工智能、虚拟现实、外骨骼机器人研究,E-mail:ielvpei@zzu.edu.cn。

引用本文:尹毅,吕培,李凯江,等. 基于多尺度动态滤波的图像增强模型[J]. 郑州大学学报(工学版),2026, 47(3):100-107. (YIN Y, LYU P, LI K J, et al. Image enhancement model based on multi-scale dynamic filtering[J]. Journal of Zhengzhou University (Engineering Science),2026, 47(3):100-107.)

整体平滑自然,同时展现丰富的图像细节,是一项重要挑战。当前图像增强领域亟须解决如何协调处理这两者关系的关键问题。

为了解决上述问题,本文提出了一种基于多尺度动态滤波分解的图像增强方法,其算法流程如图 1 所示。本文的贡献如下:①本文模型采用可学习的低通和高通滤波器,分别提取图像的低频和高频成分以得到全局平滑特征和局部纹理信息;②本文设计了 LFCA(low-frequency channel attention fusion)模块与 HFSA(high-frequency spatial attention fusion)模块,从而将提取得到的低频成分和高频成分与原图像融合,协同增强图像的全局和局部信息,进而在提升图像整体的平滑性的同时确保图像局部细节的视觉质量;③本文模型引入多尺度融合策略,综合利用不同尺度下的低频与高频特征,从而实现全局结构与局部细节的协调增强。

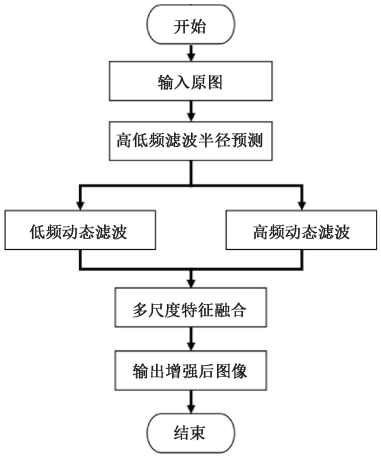

图 1　基于多尺度动态滤波的图像增强流程

Figure 1　Image enhancement flowchart based on multi-scale dynamic filtering

1　相关工作

1.1　图像增强技术

图像增强技术旨在通过提升图像的对比度、亮度、细节清晰度等视觉特性,以改善图像质量,进而提高后续视觉任务的性能。传统的图像增强方法主要基于图像的低层次特征,常见的技术包括直方图均衡[1]和伽马校正[2]等。这些方法虽然计算简单且易于实现,但由于无法有效捕捉图像的高层次语义信息,处理复杂场景时效果受限。

近年来,随着深度学习的兴起,基于神经网络的图像增强方法得到了快速发展。通过大规模数据集的训练,深度学习模型可以自动学习复杂图像特征,并进行端到端的增强处理。卷积神经网络 CNN 在图像增强任务中广泛应用,如 SRCNN[9](super resolution convolutional network)用于超分辨率重建。HDRNet[10]利用输入/输出图像对训练卷积神经网络,以预测双边空间中局部映射模型的系数,并学习如何实现局部、全局和内容相关的映射,以逼近所需的图像转换。此外,自注意力机制在图像增强中也取得了显著进展,能够关注图像中重要的区域,实现全局与局部信息的高效结合。

1.2　频域图像处理

在频域中,图像的特性可以通过其频率成分进行分析。根据傅里叶变换理论,图像可以被分解为不同频率的正弦波,这使得在频域中进行处理变得更加高效[11]。

高频和低频滤波在图像处理领域的应用广泛。高频滤波常用于边缘检测和细节增强,Luo 等[12]设计了多尺度高频特征提取模块,通过提取多个尺度的高频噪声进行人脸伪造检测。Bai 等[13]通过对抗训练直接增强高频分量来弥补 ViT(vision transformer)模型捕获图像高频分量的能力。相对而言,低频滤波则在去噪和图像平滑中发挥重要作用。噪声在不同频率层中表现出不同程度的对比度,在低频层中比在高频层中更容易检测到噪声。Xu 等[14]利用这一特性有效地处理了低光图像的噪声。常青等[15]利用小波变换通过分解超声 C 图像的高频与低频部分,以增强图像的对比度和细节信息。

2　MDFD 模型

本文提出了一种基于多尺度动态滤波分解的 MDFD(multi-scale dynamic filtering decomposition)图像增强模型,如图 2 所示。首先,通过动态频域分频对原始图像进行处理,将其分解为低频和高频成分。通过频域特征分离,分别得到整体平滑的低频图像与具有纹理清晰的高频图像;其次,通过跨低频通道注意力融合模块 LFCA 增强原图的全局整体特征,同时通过跨高频空间注意力融合模块 HFSA 增强原图的边缘细节特征;最后,联合 LFCA 与 HFSA 构建了基础特征融合块,通过多尺度的特征融合得到最终的增强后的图像。

2.1　动态频域分频

为了全面优化图像特征,本文针对全局和局部角度对输入图像进行了频域空间特征分离。如图 3 所示,为了增强频域空间中滤波的多样性,采用了 ResNet-18[16]对原始图像进行特征提取。通过提取的图像特征,自适应地预测了高低频滤波的半径。通过动态可学习的滤波半径,可以增强网络中的低频图像和高频图像的表征空间。具体计算过程如式

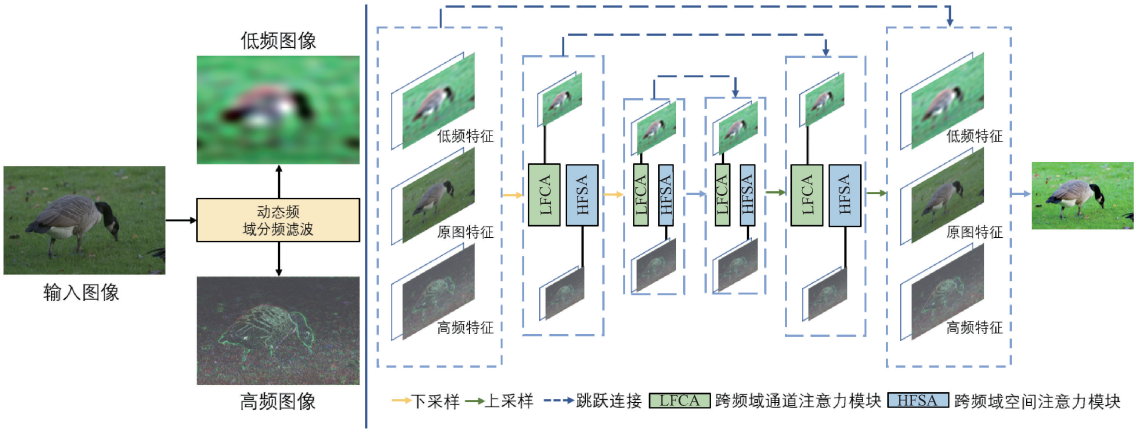

图 2 MDFD 网络架构

Figure 2 MDFD network structure

(1)所示。

$$\begin{cases}\alpha = \mathrm{Head}(\mathrm{ResNet}_{18}(I_{\mathrm{img}}^{\alpha}));\\ \beta = \mathrm{Head}(\mathrm{ResNet}_{18}(I_{\mathrm{img}}^{\beta}))。\end{cases} \tag{1}$$

式中:α、β 分别代表由 Head 推断出的高频滤波半径比例和低频滤波半径比例。

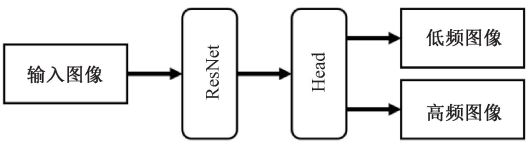

图 3 动态频域分频

Figure 3 Dynamic frequency domain divider

进一步利用可学习的滤波半径进行高频与低频滤波,具体表现如式(2)所示。

$$\begin{cases}I_{\mathrm{low}} = \mathrm{IFFT}(\mathrm{Mask}(\mathrm{FFT}(I_{\mathrm{img}}^{\alpha}),\alpha));\\ I_{\mathrm{high}} = \mathrm{IFFT}(\mathrm{Mask}(\mathrm{FFT}(I_{\mathrm{img}}^{\beta}),\beta))。\end{cases} \tag{2}$$

式中:I_{high}、I_{low} 分别代表分离出的高频与低频图像;FFT 表示快速傅里叶变换[17];Mask(*,α) 表示对输入二维频域图像在半径为 α 比例内的像素进行掩码操作;IFFT 表示快速傅里叶逆变换[17]。

2.2 跨低频通道注意力融合(LFCA)

通道注意力能够有效增强关键特征,抑制噪声并提高特征的表征能力,从而能显著提升图像整体效果。然而,仅依靠通道注意力缺乏特征引导,往往难以达到理想效果。本文利用低频图像特征来增强图像的全局优化。低频图像特征通常包含图像的整体信息和结构。通过跨低频通道注意力融合(LFCA)减少图像的噪声和伪影,进而确保图像整体的平滑性。LFCA 结构如图 4 所示。首先,图像特征经过卷积得到查询特征 $\boldsymbol{Q}_{\mathrm{c}}$,低频特征经过卷积将通道变为 2 倍。其次,在通道维度将低频特征分解为特征 $\boldsymbol{K}_{\mathrm{c}}$、$\boldsymbol{V}_{\mathrm{c}}$,如式(3)所示。

$$\begin{cases}\boldsymbol{Q}_{\mathrm{c}} = \mathrm{Conv}(\boldsymbol{F}_{\mathrm{img}});\\ (\boldsymbol{K}_{\mathrm{c}},\boldsymbol{V}_{\mathrm{c}}) = \mathrm{Chunk}^{2}(\mathrm{Conv}(\boldsymbol{F}_{\mathrm{low}}))。\end{cases} \tag{3}$$

式中:Conv 代表卷积核尺寸为 3×3 的卷积操作;Chunk^{2} 表示按第 2 个维度拆分成两个子特征。

最后,特征融合过程中,采用在通道维度进行注意力的融合,如式(4)所示。

$$\overline{\boldsymbol{F}} = \mathrm{Softmax}(\boldsymbol{Q}_{\mathrm{c}}\boldsymbol{K}_{\mathrm{c}}^{\mathrm{T}}/\delta)\boldsymbol{V}_{\mathrm{c}}。 \tag{4}$$

式中:δ 为可学习的参数;$\overline{\boldsymbol{F}}$ 为低频特征与输入图像特征的融合特征。

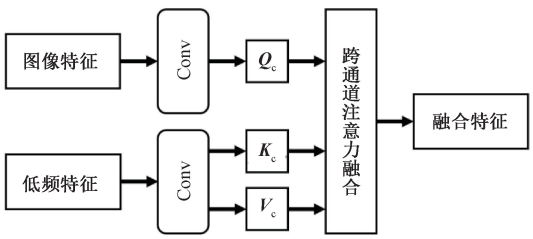

图 4 LFCA 结构图

Figure 4 LFCA structure diagram

2.3 跨高频空间注意力融合(HFSA)

空间注意力能够通过关注图像不同区域的显著特征来提升局部信息,有效增强边缘和纹理细节。然而,泛化的对所有特征进行空间维度的卷积难以精确聚焦在特定位置,限制了其效果。本文引入了高频图像特征,专注于提取图像边缘纹理信息,以引导图像局部细节的优化和微调。通过综合运用空间注意力和高频特征,使用跨高频空间注意力融合(HFSA),有针对性地调整图像的空间细节,从而显著提升图像的局部增强效果。HFSA 结构如图 5 所示。首先,分别将融合特征 $\overline{\boldsymbol{F}}$ 经过卷积操作后将通道扩展为原来的 3 倍,高频特征 $\boldsymbol{F}_{\mathrm{high}}$ 经过卷积操作后通道扩展为原来的 2 倍;其次,将融合特征在通道维度拆分为特征 $\boldsymbol{Q}'_{\mathrm{s}}$、$\boldsymbol{K}'_{\mathrm{s}}$、$\boldsymbol{V}'_{\mathrm{s}}$,高频特征在通道维度

拆分为特征 $\boldsymbol{K}''_s$、$\boldsymbol{V}''_s$,如式(5)所示。

$$\begin{cases}(\boldsymbol{Q}'_s,\boldsymbol{K}'_s,\boldsymbol{V}'_s)=\mathrm{Chunk}^3(\mathrm{Conv}(\boldsymbol{F}));\\(\boldsymbol{K}''_s,\boldsymbol{V}''_s)=\mathrm{Chunk}^2(\mathrm{Conv}(\boldsymbol{F}_{\mathrm{high}}))。\end{cases}\tag{5}$$

式中:Chunk^3 表示按照第 2 个维度分别拆分 3 个子特征。

最后,使用空间自注意力分别对融合特征进行特征提取,作为跨高频空间注意力融合的查询 $\boldsymbol{Q}''_s$。如式(6)所示。

$$\begin{cases}\boldsymbol{Q}''_s=((\boldsymbol{Q}'^{\mathrm{T}}_s\boldsymbol{K}'_s/\gamma)\boldsymbol{V}'^{\mathrm{T}}_s)^{\mathrm{T}};\\\overline{\boldsymbol{F}}=((\boldsymbol{Q}''^{\mathrm{T}}_s\boldsymbol{K}''_s/\eta)\boldsymbol{V}''^{\mathrm{T}}_s)^{\mathrm{T}}。\end{cases}\tag{6}$$

式中:γ、η 均为可学习的参数;$\overline{\boldsymbol{F}}$ 代表经过空间细节优化后的特征。

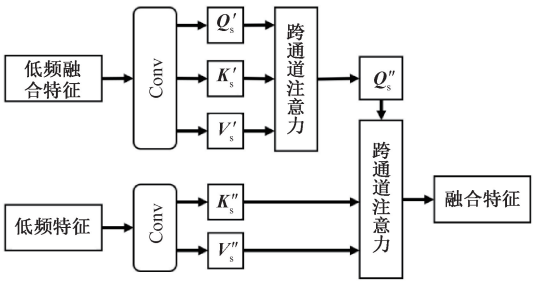

图 5　HFSA 结构图

Figure 5　HFSA structure diagram

2.4　损失函数

MDFD 模型可以通过端到端的训练方式来实现。参考常规图像增强的设计方法,该模型采用均方差损失 L_{mse} 来优化网络中的可学习参数,如式(7)所示。

$$L_{\mathrm{mse}}=\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i-y_i)^2。\tag{7}$$

式中:N 代表图像中的像素数量;y_i 为目标图像 y 中第 i 个像素的值;$\hat{y}_i$ 为模型生成的增强图像 y 中的第 i 个像素的值。

3　模型训练

3.1　数据集介绍

文本采用 MIT-Adobe FiveK[18] 和 PPR10K[19] 数据集进行实验评估。FiveK 数据集包含 5 000 张 RAW 图像,每张图像均经过 5 种不同风格的手动修饰(A/B/C/D/E),实验中使用了版本 C。数据集中 4 500 对用于训练,500 对用于测试,训练和测试分别采用 480p 及 4K 分辨率图像。PPR10K 数据集由 11 161 张 RAW 肖像图像组成,每张图像有 3 个不同的真实修饰版本(a/b/c),其中 8 875 对用于训练,2 286 对用于测试,实验在 360p 分辨率下进行。在该数据集上执行了照片修饰任务,数据增强策略遵循现有的常用设置。

3.2　环境与超参数设置

本文的多尺度动态滤波的图像增强的实验环境:Intel CPU i9-9980XE;NVIDIA Tesla V100,训练环境基于 Python 3.7,CUDA 11.1,PyTorch 1.8.1。实验中采用标准的 Adam 优化器来最小化方程(7)中的损失函数。在 FiveK 和 PPR10K 数据集上,批量大小分别设定为 1 和 16。所有模型均使用固定学习率 1×10^{-4} 进行了 40 轮次的训练。为使 MDFD 学习过程更加稳定,在前 5 个训练轮次中冻结了参数,并将学习率降低至原来的 0.1。

4　实验结果

4.1　评价指标

实验采用 3 种度量标准,包括峰值信噪比 *PSNR*、结构相似性指数 *SSIM*[20] 和色差 ΔE_{ab}[19] 作为评估指标,部分评估指标如下。

$$PSNR=10\lg\left(\frac{a_{\max}}{b_{\mathrm{mse}}}\right)。\tag{8}$$

式中:$a_{\max}$ 表示图像像素的最大值;b_{mse} 表示均方方差。

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)}。\tag{9}$$

式中:μ_x 表示图像 x 的均值;μ_y 表示图像 y 的均值;σ_x^2 表示图像 x 的方差;σ_y^2 表示图像 y 的方差;σ_{xy} 表示图像 x 和 y 之间的协方差;C_1 和 C_2 表示常数,本例中分别为 0.01 和 0.03。

ΔE_{ab}[19] 是在 CIELAB 颜色空间中定义的颜色差异度量标准,已被证明与人类感知一致。与 *PNSR* 和 *SSIM* 相反,较小的 ΔE_{ab} 代表更好的性能。

$$\Delta E_{\mathrm{ab}}=\sqrt{(L_2^*-L_1^*)^2+(a_2^*-a_1^*)^2+(b_2^*-b_1^*)^2}。\tag{10}$$

式中:L_1^* 表示第 1 张图像的亮度值;L_2^* 表示第 2 张图像的亮度值;a_1^* 和 a_2^* 分别表示第 1 张和第 2 张图像在 CIELAB 色彩空间中 a 轴的值;b_1^* 和 b_2^* 分别表示第 1 张和第 2 张图像在 CIELAB 色彩空间中 b 轴的值。

4.2　低频特征融合实验

低频特征融合旨在实现全局性能优化。为了验证 LFCA 的优势,本文将其分别与 CNN、空间注意力 Space-att 融合方式进行了对比。由表 1 可知,与其他的特征融合方式相比,LFCA 的 *PSNR*、*SSIM*、ΔE_{ab} 均取得了最优效果。LFCA 采用的通道注意力机制可以更好地融合低频图像中的全局特征,表现出更好的特征提取能力。相比之下,计算消耗更高空间

注意力融合的 Space-att 由于忽略了通道间的相关性,仅通过空间维度加权不足以捕捉到所有有价值的信息,进而在融合低频图像全局特征中效果不佳。同时值得注意的是,在特征融合时,常用 CNN 进行特征融合,卷积核的感受野较小,对于大尺寸物体或者跨区域特征的关系信息捕捉能力有限,CNN 在图像全局特征提取时效果不明显,这也导致了通用的 CNN 采用 Concat 增强效果不佳。实验结果表明,与其他低频特征融合模块相比,LFCA 在应对图像整体优化上可以具备更强的全局特征提取能力。

表 1 不同低频特征融合方式模型对比

Table 1 Model comparison of different low-frequency feature fusion methods

模块	*PSNR*	*SSIM*	ΔE_{ab}
CNN	25.39	0.936	7.73
Space-att	25.50	0.951	7.51
LFCA	25.90	0.964	7.38

4.3 高频特征融合实验

高频图像可以凸显图像中的边缘纹理,主要在于优化图像的局部细节。与低频特征融合不同,高频特征融合既要考虑图像本身周围像素级关联关系,又要考虑在边缘细节上进行的局部特征的加强。因此,HFSA 首先进行了空间自注意力的像素级优化。然后,通过跨高频空间注意力融合对边缘细节进行具有针对性的优化。为此,本文在 FiveK 数据集通过照片修饰任务对 HFSA 与 CNN、空间注意力 Space-att、通道注意力 Channel-att 融合方式进行了对比分析。如表 2 所示,通过对比常规的特征融合方法,HFSA 在 *PSNR*、*SSIM*、ΔE_{ab} 评估指标上均得到了最优的效果,在照片修饰任务中表现出最佳性能。

表 2 不同高频特征融合方式模型对比

Table 2 Model comparison of different high-frequency feature fusion methods

模块	*PSNR*	*SSIM*	ΔE_{ab}
CNN	25.63	0.959	7.49
Channel-att	25.39	0.955	7.55
Space-att	25.75	0.961	7.41
HFSA	25.90	0.964	7.38

4.4 消融实验

为了研究高频与低频图像特征对模型的影响,在场景数据集 FiveK 上对不同特征进行了消融分析。比较在 MDFD 模型中高频特征和低频特征的不同组合融合。如表 3 所示,当在 MDFD 模型中不使用频域特征时,*PSNR* 指标仅为 24.59;分别添加高频特征和低频特征时,*PSNR* 分别为 25.65 和 25.85;联合高频特征和低频特征时,MDFD 在全局与细节方面取到最佳结果,*PSNR* 指标达到 25.90。实验同时表明,低频特征在一定的程度上改善图像的效果。与低频特征相比,高频特征能够显著地优化图像像素级的增强效果,因此在各项指标中,其增强效果优于低频特征的增益幅度。

表 3 消融实验结果

Table 3 Ablation experiment results

频域特征	高频特征	*PSNR*	*SSIM*	ΔE_{ab}
		24.59	0.954	8.05
√		25.65	0.960	7.51
	√	25.85	0.961	7.46
√	√	25.90	0.964	7.38

注:√表示使用该特征。

4.5 实验结果分析

为了进一步展示不同图像增强方法的增强效果,将 MDFD 和当前主流模型在 FiveK 场景数据集上的图像增强效果进行对比,如图 6 所示。在图 6 中右上角为增强效果与真值的像素差异。从图 6 可以看出,在光线变化和低照度等场景下 MDFD 拥有更优的增强效果。通过观察右上角的差异图像,可以看出 MDFD 对图像中的细节纹理、光线不足问题展现出更好的可视效果。MDFD 在图像细节部分通过增强高频成分使得像素差异分布相对于现在主流模型更加集中,对图像细节部分内容的增强与真值基本吻合。此外,MDFD 通过跨低频注意力融合使得图像整体平滑度明显高于当前主流模型的结果,在图像整体轮廓上展现出明显的增强效果。

为验证模型有效性,在 FiveK 数据集上进行了照片修饰任务和色调映射任务,并与主流图像增强方法对比,结果见表 4 和表 5。从表 4 可知,在 FiveK 数据集上,通过照片修饰任务验证 MDFD 模型在图像的整体外观和质量上均可取得最佳结果。而色调映射任务主要通过验证模型在细粒度像素级别的性能。此外,不同模型的运行时间对比表明,MDFD 模型在保持优越的图像增强性能的同时,也具备一定的运行效率优势。经典的 U-Net 架构模型[21]包含大量的卷积操作与上下采样两条通路,导致计算复杂度和内存开销较大,运行时间相对较长。尽管 3D-LUT 系列模型[6-7]运行时间较短,但其依赖查找表直接进行颜色映射,虽然计算复杂度低,但难以应对复杂场景和全局信息的协同处理。

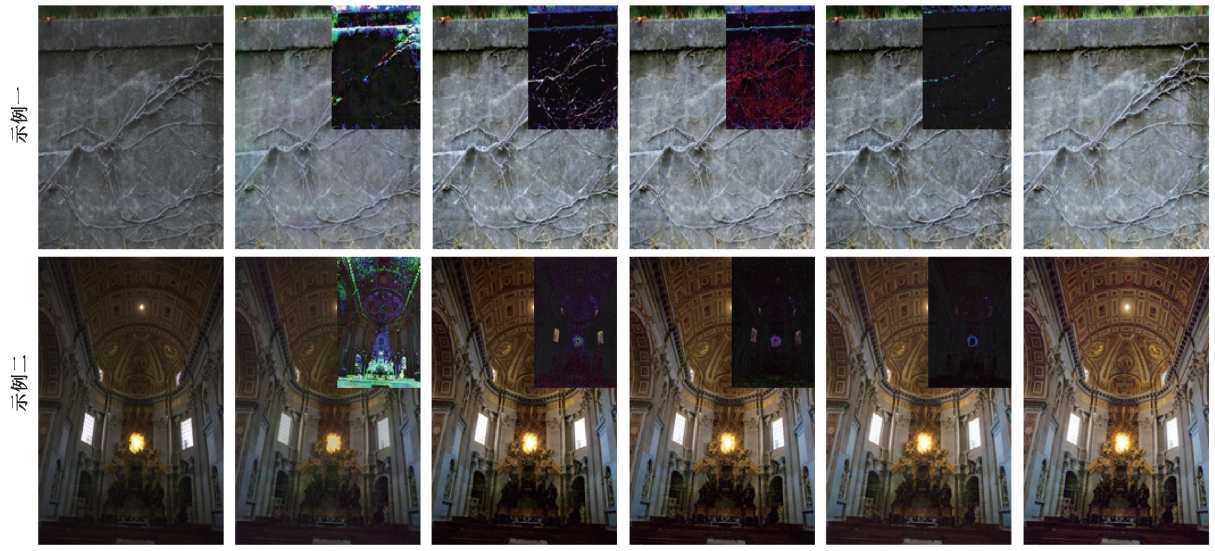

图 6　在 FiveK 数据集的增强效果比较

Figure 6　Comparison of enhancement results on the FiveK dataset

表 4　在 FiveK 数据集上照片修饰性能比较

Table 4　Comparison of photo retouching performance on the FiveK dataset

方法	*PSNR*	*SSIM*	ΔE_{ab}	运行时间/s
Harmonizer[21]	24.11	0.904	8.23	16.74
HDRNet[10]	24.66	0.915	8.06	—
CSRNet[22]	25.17	0.924	7.75	3.49
3D-LUT[6]	25.29	0.923	7.55	—
3D-LUT+AdaInt[7]	25.49	0.926	7.47	1.59
RSFNet-map[8]	25.49	0.924	7.23	9.98
MDFD	25.90	0.964	7.38	9.42

表 5　在 FiveK 数据集上色调映射应用性能比较

Table 5　Comparison of tone mapping application performance on the FiveK dataset

方法	*PSNR*	*SSIM*	ΔE_{ab}
HDRNet[10]	24.52	0.915	8.14
3D-LUT[6]	25.07	0.920	7.55
CSRNet[22]	25.19	0.921	7.63
3D-LUT+AdaInt[7]	25.28	0.925	7.48
MDFD	25.32	0.937	7.39

如表 5 所示,本文在 FiveK 数据集上色调映射任务上及在 *PSNR*、*SSIM*、ΔE_{ab} 指标上均获得了最佳的效果。这充分表明 MDFD 模型通过对高频域信息和低频域信息结合,能更好地提取照片整体轮廓和局部细节的内容特征,从而展现出优秀的照片修饰和色调映射应用性能。相比之下,对图像进行多尺度复杂处理的 HDRNet 计算开销大,内存消耗高,表现不佳。Harmonizer 无法区分在不同光照下的前景与背景,导致对全局颜色感知不足。CSRNet 虽然通过全局特征来对局部特征进行优化,但是过分精简网络,导致采用的 MLPs 层数过少,感受野受限导致全局上下文颜色失真。3D-LUT 由于是基于全局查找表进行色彩变化,缺乏对局部图像细粒度的控制,因此在处理局部风格差异明显的照片时性能较差。此外,为了验证 MDFD 在标准照片修饰任务上的性能,本文在 PPR10K 的摄影数据集上进行了模型验证。如表 6 所示,在大规模数据集上,MDFD 的性能得到了充分挖掘,在目前最优的 3D-LUT+AdaInt 效果上实现了显著提升。

表 6　PPR10K 数据集上照片修饰性能比较

Table 6　Comparison of photo enhancement performance on the PPR10K dataset

方法	PPR10K-a			PPR10K-b			PPR10K-c		
	PSNR	*SSIM*	ΔE_{ab}	*PSNR*	*SSIM*	ΔE_{ab}	*PSNR*	*SSIM*	ΔE_{ab}
Harmonizer[21]	24.76	0.929	—	22.79	0.885	—	24.56	0.902	—
HDRNet[10]	23.93	—	8.70	23.96	—	8.84	24.08	—	8.87
CSRNet[22]	22.72	—	9.75	23.76	—	8.77	23.17	—	9.45
3D-LUT[6]	25.64	—	6.97	24.70	—	7.71	25.18	—	7.58
3D-LUT+AdaInt[7]	26.33	0.947	6.56	25.40	0.936	7.33	25.68	0.919	7.31
RSFNet-map[8]	25.58	0.949	—	24.88	0.945	—	25.52	0.939	—
MDFD	27.54	0.945	6.50	27.11	0.944	6.71	27.40	0.946	6.52

5 结论

本文提出了一种基于多尺度动态滤波分解的MDFD图像增强模型,旨在解决传统图像增强技术中局部细节与全局纹理之间的冲突及无法协同增强的问题,通过可学习的高通滤波器与低通滤波器分别提取图像的高频与低频成分。本文提出了LFCA模块与HFSA模块协同的全局与局部优化的图像增强模型,其中LFCA模块确保了图像整体的平滑性,而HFSA模块则提升了图像的局部细节特征。通过使用多尺度融合方法综合高频与低频的信息,实现了全局平滑与局部纹理的结合,确保图像增强的效果更加自然和平衡。在数据集FiveK和PPR10K上的实验结果表明,MDFD模型相较于多种图像增强模型,均取得了更优的图像增强效果,证明MDFD模型在复杂环境和颜色丰富等场景下具有优越的图像增强性能。在耗时上,虽然与经典的U-Net架构模型[21]相比,MDFD模型在性能提升的同时具备较低的计算耗时。然而,与3D-LUT系列模型[6-7]相比,其耗时仍略显偏高。为进一步提升效率,计划在后续研究中对MDFD模型进行轻量化优化。

参考文献:

[1] 刘华军, 张瑞珏, 刘建锋, 等. 基于FPGA的高分辨率视频图像实时增强去雾系统[J]. 郑州大学学报(工学版), 2020, 41(2): 19-24.
LIU H J, ZHANG R J, LIU J F, et al. High resolution video image real-time enhancement system based on FPGA[J]. Journal of Zhengzhou University (Engineering Science), 2020, 41(2): 19-24.

[2] GAUTAM C, TIWARI N. Efficient color image contrast enhancement using range limited Bi-histogram equalization with adaptive gamma correction[C]//2015 International Conference on Industrial Instrumentation and Control (ICIC). Piscataway: IEEE, 2015: 175-180.

[3] CHEN Y H, ZHU G, WANG X Q, et al. FRR-NET: a fast reparameterized residual network for low-light image enhancement[J]. Signal, Image and Video Processing, 2024, 18(5): 4925-4934.

[4] ZHOU J C, LI B S, ZHANG D H, et al. UGIF-net: an efficient fully guided information flow network for underwater image enhancement[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 4206117.

[5] SHEN L, YUE Z H, FENG F, et al. MSR-net: low-light image enhancement using deep convolutional network[EB/OL]. (2017-11-07)[2024-08-11]. https://arxiv.org/abs/1711.02488.

[6] ZENG H, CAI J R, LI L D, et al. Learning image-adaptive 3D lookup tables for high performance photo enhancement in real-time[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4): 2058-2073.

[7] YANG C Q, JIN M G, JIA X, et al. AdaInt: learning adaptive intervals for 3D lookup tables on real-time image enhancement[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 17501-17510.

[8] OUYANG W Q, DONG Y, KANG X Y, et al. RSFNet: a white-box image retouching approach using region-specific color filters[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2023: 12126-12135.

[9] YAHIAOUI M L, KHARFI F, BOUKERDJA L. Resolution enhancement of neutron radiography image using combined SRCNN-POCS method[J]. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2023, 1050: 168123.

[10] GHARBI M, CHEN J W, BARRON J T, et al. Deep bilateral learning for real-time image enhancement[J]. ACM Transactions on Graphics, 2017, 36(4): 1-12.

[11] HALIDOU A, MOHAMADOU Y, ARI A A A, et al. Review of wavelet denoising algorithms[J]. Multimedia Tools and Applications, 2023, 82(27): 41539-41569.

[12] LUO Y C, ZHANG Y, YAN J C, et al. Generalizing face forgery detection with high-frequency features[J]. (2021-03-23)[2024-08-11]. https://arxiv.org/abs/2103.12376.

[13] BAI J W, YUAN L, XIA S T, et al. Improving vision transformers by revisiting high-frequency components[C]// Computer Vision-ECCV 2022. Cham: Springer, 2022: 1-18.

[14] XU K, YANG X, YIN B C, et al. Learning to restore low-light images via decomposition-and-enhancement[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 2278-2287.

[15] 常青, 杨程伟, 罗彬杰, 等. 基于小波变换的扩散焊超声C图像融合算法[J]. 郑州大学学报(工学版), 2023, 44(4): 54-59, 87.
CHANG Q, YANG C W, LUO B J, et al. Ultrasonic C image fusion algorithm for diffusion welding based on wavelet transform[J]. Journal of Zhengzhou University (Engineering Science), 2023, 44(4): 54-59, 87.

[16] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 770-778.

[17] DUHAMEL P, VETTERLI M. Fast Fourier transforms: a tutorial review and a state of the art[J]. Signal Processing, 1990, 19(4): 259-299.

[18] BYCHKOVSKY V, PARIS S, CHAN E, et al. Learning photographic global tonal adjustment with a database of input/output image pairs[C]//CVPR 2011. Piscataway: IEEE, 2011: 97-104.

[19] LIANG J, ZENG H, CUI M M, et al. PPR10K: a large-scale portrait photo retouching dataset with human-region mask and group-level consistency[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway:IEEE, 2021:00071.

[20] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[EB/OL]. (2014-01-01)[2024-07-16]. https://ieeexplore.ieee.org/document/1284395.

[21] KE Z H, SUN C Y, ZHU L, et al. Harmonizer: learning toPerform white-box image andVideo harmonization[C]// Computer Vision-ECCV 2022. Cham: Springer, 2022: 690-706.

[22] HE J W, LIU Y H, QIAO Y, et al. Conditional sequential modulation for efficient global image retouching[C]// Computer Vision-ECCV 2020. Cham: Springer, 2020: 679-695.

Image Enhancement Model Based on Multi-scale Dynamic Filtering

YIN Yi, LYU Pei, LI Kaijiang, ZHENG Haokun, XU Hao, CHEN Mengjie

(School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China)

Abstract: To address the issue of collaborative enhancement between global smoothness and local textures in traditional image enhancement techniques, in this study an MDFD image enhancement model based on multi-scale dynamic filtering decomposition was proposed. Initially, learnable low-pass and high-pass filters were utilized to extract the low-frequency and high-frequency image components, respectively. Subsequently, by combining these two frequency-domain image components, the cross low-frequency channel attention fusion module (LFCA) and cross high-frequency spatial attention fusion module (HFSA) were introduced to achieve collaborative enhancement of image global and local features. Finally, a multi-scale fusion strategy was introduced to comprehensively utilize high-frequency and low-frequency information at different scales for feature fusion. The advantage of multi-scale fusion lay in its ability to effectively integrate details and global features at different scales, significantly enhancing the image at multiple levels. Experimental results showed that the MDFD model performed excellently in the validation on the FiveK and PPR10K datasets, with peak signal-to-noise ratio (*PSNR*) reaching 25.90 and 27.35, structural similarity index (*SSIM*) being 0.964 and 0.945, and ΔE_{ab} being 7.38 and 6.50, respectively. These results indicated that the MDFD model could offer superior image enhancement performance in complex environments and color-rich scenes.

Keywords: image enhancement; low-pass filtering; high-pass filtering; multi-scale fusion; frequency domain transformation