

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-16

论文引用格式: Lin Xiao, Dong Dateng, Li Yan, Hong Xuan, Yang Qizhe. Spatial-frequency multi-scale sparse attention for single image raindrop removal[J/OL]. Journal of Image and Graphics, XXXX:1-16. DOI: 10.11834/jig.260323. (林晓, 董大腾, 李岩, 洪璇, 杨启哲. 空频联合多尺度稀疏注意力的单幅图像去雨[J/OL]. 中国图象图形学报, XXXX:1-16. DOI: 10.11834/jig.260323.) [DOI:10.11834/jig.260323]

空频联合多尺度稀疏注意力的单幅图像去雨

林晓, 董大腾, 李岩, 洪璇, 杨启哲*

上海师范大学信息与机电工程学院, 上海市 201418

摘要: 目的 针对图像修复中空间异质性退化(同时包含中心区域均匀模糊与边缘区域复杂折射)的难题, 现有方法因缺乏退化属性先验与频域建模, 难以兼顾两类退化属性的修复需求。为此, 提出一种基于空间和频域联合处理与不确定性引导的多尺度稀疏变换器。**方法** 首先设计多尺度物理解耦模块, 利用不同膨胀率与核大小的并行卷积分支分别侧重提取全局模糊、过渡细节与高频折射特征; 进而引入空间-频域联合处理模块, 通过空域深度可分离卷积与频域幅度相位调制互补增强特征表示; 在此基础上构建不确定性估计模块, 生成空间不确定性图以量化各区域修复难度, 并据此设计不确定性引导的稀疏注意力机制, 使模型在高不确定性区域保持精细纹理, 在低不确定性区域抑制冗余计算。同时, 采用融合重建损失与不确定性图正则化的复合损失函数, 约束不确定性图聚焦于退化区域。**结果** 在 SPAD 雨纹数据集和 AGAN-Data 雨滴数据集上, 所提方法的峰值信噪比(peak signal-to-noise ratio, PSNR)和结构相似性(structural similarity index, SSIM)均优于现有基线模型; 在 Dense-Haze 真实雾霾数据集上, PSNR 达到最优。在雨滴去除任务中, 相比对比方法中的最优方法, PSNR 指标提升了 0.30 dB。消融实验验证了多尺度解耦、空间-频域联合处理及不确定性引导机制对性能的贡献。**结论** 本文提出的空频联合与不确定性引导的多尺度稀疏变换器, 能够有效应对空间异质性退化, 在雨滴、雨纹和雾霾去除任务上均取得优越的修复效果。实验代码与数据已开源发布(<https://doi.org/10.57760/sciencedb.j00240.00269>)。

关键词: 图像修复; 空间和频域联合处理; 不确定性引导; 稀疏注意力; 多尺度特征

Spatial-frequency multi-scale sparse attention for single image raindrop removal

Lin Xiao, Dong Dateng, Li Yan, Hong Xuan, Yang Qizhe*

School Of Information, Mechanical And Electrical Engineering, Shanghai Normal University, Shanghai 201418, China

Abstract: Objective Single image deraining is a critical task in image restoration, aiming to remove raindrop degradations that exhibit spatially heterogeneous characteristics: central regions of raindrops produce uniform blur due to semi-transparent occlusion, while edge regions introduce complex refraction distortions. These two physically distinct degradation patterns coexist and impose conflicting requirements on restoration models – effective suppression of blur in central areas yet preservation of fine textures and structural details at edges. Existing methods, including conventional model-based approaches and deep learning frameworks, lack explicit physical priors on degradation properties and effective

收稿日期: 2026-06-06; 修回日期: 2026-08-16

* 通信作者: 杨启哲 qzyang@shnu.edu.cn

基金项目: 国家自然科学基金(62302167, 62477013); 上海市自然科学基金(24ZR1456100); 上海市科委“科技创新行动计划”科普专项“青少年科创教育培养”项目(24DZ2305900)

Supported by: National Natural Science Foundation of China (Grant Nos. 62302167, 62477013); Natural Science Foundation of Shanghai (Grant No. 24ZR1456100); Science and Technology Innovation Action Plan of Shanghai Science and Technology Commission — Youth Science and Technology Innovation Education and Training Program (Grant No. 24DZ2305900)

frequency-domain modeling, making it difficult to address both types of degradation simultaneously. Convolutional neural networks (CNNs) are limited by local receptive fields, hindering accurate modeling of complex light-path distortions at raindrop edges. Transformer-based architectures leverage global self-attention but rely on feature similarity or computational efficiency heuristics without explicit modeling of degradation physics, and their sparse attention decisions are often globally homogeneous. To overcome these limitations, this paper proposes a novel single image deraining method based on spatial-frequency joint processing and uncertainty-guided multi-scale sparse attention. **Method** The proposed method adopts a symmetric encoder-decoder (U-Net) architecture. The core innovations are four modules. First, a multi-scale physical decoupling block employs three parallel depthwise separable convolution branches with different kernel sizes and dilation rates: a large-kernel low-dilation branch (7×7 , dilation 1) captures global blur; a medium-kernel moderate-dilation branch (5×5 , dilation 2) preserves transitional details; and a small-kernel high-dilation branch (3×3 , dilation 3) focuses on high-frequency refraction features at raindrop edges. A cross-scale interaction unit adaptively fuses these multi-scale features via spatial attention weights, providing physics-guided multi-scale representations. Second, a spatial-frequency joint processing module enhances feature representations complementarily. The spatial branch employs depthwise separable convolution with channel attention to capture local texture information. The frequency branch transforms features into the Fourier domain via real FFT, applies a lightweight complex modulation network to jointly model real and imaginary parts, then transforms back. A dynamic gating unit adaptively fuses the two branches based on global image context. Third, an uncertainty guidance module estimates a spatial uncertainty map that quantifies restoration difficulty at each pixel position. This map is generated through three stacked 3×3 convolutions followed by a 1×1 convolution and Sigmoid activation, then fused with global context from average pooling. The uncertainty map dynamically modulates attention scores – enhancing responses in high-uncertainty regions (e.g., raindrop edges) while suppressing redundant computation in low-uncertainty regions (simple backgrounds). Fourth, an uncertainty-guided sparse attention mechanism incorporates the modulated uncertainty signals into Transformer attention, enabling physics-aware adaptive resource allocation. The overall loss function is a composite of Charbonnier reconstruction loss, sparsity loss to encourage the uncertainty map to focus on degraded regions, and total variation loss to smooth the uncertainty map. Experiments are conducted on three benchmark datasets: the AGAN-Data raindrop dataset, the SPAD rain streak dataset, and the Dense-Haze real-world haze dataset. All input images are cropped to 256×256 patches. Training uses the AdamW optimizer with an initial learning rate of 2×10^{-4} and cosine annealing decay to 1×10^{-6} on an NVIDIA A100 GPU. **Result** Extensive experiments demonstrate that the proposed method achieves state-of-the-art performance. On the AGAN-Data raindrop removal test-a subset, the method attains a PSNR of 32.97 dB and an SSIM of 0.942, outperforming all compared methods including AttnGAN, Quan's model, Uformer, IDT, MAXIM-2S, AST-B, RQ-D2UNet, and WeatherNet. It surpasses the best competing method (WeatherNet, 32.67 dB) by 0.30 dB in PSNR and matches the best SSIM. On the SPAD rain streak removal dataset, the method achieves a PSNR of 47.65 dB and an SSIM of 0.9931, improving PSNR by 0.31 dB over IDT (47.34 dB) and achieving the second-best SSIM. On the Dense-Haze real-world haze dataset, the method attains a PSNR of 17.51 dB, which is 0.56 dB higher than MB-TaylorFormer-B (16.95 dB), demonstrating strong generalization across different degradation types. Qualitative visual comparisons show that the proposed method effectively removes raindrops and rain streaks while preserving sharp edges and fine textures, whereas competing methods either leave residual artifacts or over-smooth details. Ablation studies on the AGAN-Data dataset systematically evaluate the three key components: multi-scale physical decoupling, spatial-frequency joint processing, and uncertainty guidance. The full model with all three modules achieves the best performance (PSNR 32.97 dB, SSIM 0.942). The combination of multi-scale decoupling and spatial-frequency joint processing yields 32.86 dB, while adding uncertainty guidance further improves to 32.97 dB. Each individual module, when introduced alone, slightly degrades performance relative to the baseline (32.76 dB) due to disruption of feature distribution, but combinations exhibit clear synergistic effects. Among the three, spatial-frequency joint processing contributes slightly more than uncertainty guidance (32.86 vs. 32.80 when paired with multi-scale decoupling), highlighting the unique advantage of frequency-domain modeling for edge refraction recovery. **Conclusion** In this paper, we have proposed a single image deraining method based on spatial-frequency joint multi-scale sparse attention. By integrating multi-scale physical decoupling, spatial-frequency complementary enhancement, and uncertainty-guided sparse attention, the pro-

posed method explicitly incorporates degradation physics into the attention mechanism and dynamically allocates computational resources according to regional restoration difficulty. Extensive experiments on raindrop removal, rain streak removal, and real-world haze removal benchmarks demonstrate that the method consistently outperforms state-of-the-art approaches in both quantitative metrics (PSNR and SSIM) and visual quality. Ablation studies confirm the synergistic contributions of multi-scale decoupling, spatial-frequency joint processing, and uncertainty guidance mechanisms. The proposed framework shows strong generalization across heterogeneous degradations and holds promise for future extensions to video restoration and real-time deployment. The experimental code and data have been made publicly available (<https://doi.org/10.57760/sciencedb.j00240.00269>).

Key words: image restoration; spatial-frequency joint processing; uncertainty guidance; sparse attention; multi-scale feature

论文引用格式: Lin Xiao, Dong Dateng, Li Yan, Hong Xuan, Yang Qizhe. Spatial-frequency multi-scale sparse attention for single image raindrop removal [J/OL]. Journal of Image and Graphics. DOI: 10.11834/jig.260323. (林晓, 董大腾, 李岩, 洪璇, 杨启哲. 空频联合多尺度稀疏注意力的单幅图像去雨 [J/OL]. 中国图象图形学报. DOI: 10.11834/jig.260323.)

0 引言

图像修复旨在恢复高质量清晰图像,是计算机视觉领域的基础性问题之一(Wang等,2020;Zhang等,2021)。在实际拍摄场景中,雨水附着于镜头或物体表面所形成的雨滴退化具有独特的空间异质性,雨滴中心区域因半透明遮挡产生均匀模糊,而边缘区域则因光折射引发复杂的纹理畸变(Eigen等,2013)。这两类物理属性截然不同的退化模式同时存在,对修复模型提出了矛盾的需求,既要为中心强遮挡区域进行过滤以消除伪影,又需在边缘折射区保留细微的纹理变化。然而,现有方法在处理此类退化时仍面临显著挑战(Quan等,2019)。

传统基于模型的方法利用雨滴的光度或几何先验进行显式建模,虽无需配对数据,但受限于先验的表达能力,难以适应真实场景中雨滴形态的多样性(He等,2010)。近年来,基于深度学习的卷积神经网络(convolutional neural networks, CNN)因其强大的特征学习能力在图像修复领域取得进展(Wang等,2020;Qian等,2018),但卷积操作的局部感受野限制了其对雨滴边缘复杂光路畸变的建模。变换器(Transformer)架构凭借全局自注意力机制在该任务中展现出潜力,如 Uformer (Wang 等, 2022) 和

Restormer(Zamir等,2022)通过窗口或通道注意力平衡计算效率与全局建模,DRSformer(Chen等,2023)引入 Top-K 通道选择实现特征稀疏化。然而, Uformer (Wang 等, 2022) 和 Restormer (Zamir 等, 2022)分别采用窗口注意力和通道注意力,其权重完全由特征相似度决定,无法感知退化区域的属性差异;DRSformer(Chen等,2023)的 Top-K 通道选择是全局固定的,无法适配异质性需求;AST(Zhou等,2024)虽融合稠密与稀疏分支,但其稀疏决策基于特征能量而非退化模式;USTN(Chen等,2023)利用不确定性采样但未将其反馈至注意力调制。上述方法均难以同时满足中心模糊与边缘折射的矛盾处理需求。本文显式引入多尺度退化解耦分支,将不同退化模式分离后再进行注意力计算。

针对上述问题,本文提出一种基于空频联合与不确定性引导的多尺度稀疏变换器(spatial-frequency multi-scale sparse transformer, SFFormer)。所提方法的核心思想是将区域退化属性先验引入注意力机制的设计中,有机融合多尺度退化解耦、空频互补增强与不确定性驱动的注意力调制,形成一个面向异质性退化的统一修复框架,而非独立模块的简单堆叠。具体地,通过多尺度退化解耦模块,利用不同膨胀率的深度可分离卷积分支分别侧重提取全局模糊、过渡细节与高频折射特征,实现退化模式的隐式分离;进而引入空间-频域联合处理模块,在空域深度可分离卷积与频域幅度相位调制之间形成互补增强,全面捕捉退化图像的局部纹理与全局结构信息。在此基础上,构建不确定性估计模块生成空间不确定性图以量化各区域修复难度,并设计不确定性引导的稀疏注意力机制。与仅将不确定性用于通道注意力加权的文献(Shao等,2021)不同,本文的不确定性图同时参与多尺度特征的跨尺度交互权

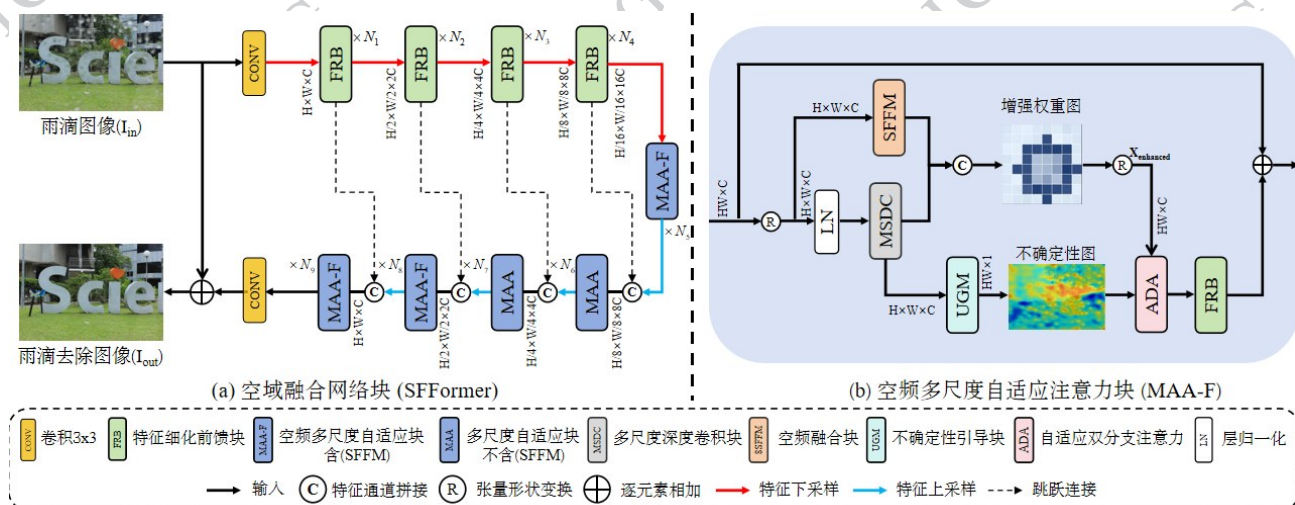
重、注意力分数的动态调制、残差更新的软掩码以及损失函数的正则化约束,实现了从特征提取到注意力再到监督信号的全流程物理感知,使模型能够根据区域退化特性动态分配资源。同时,采用融合重建损失与不确定性图正则化的复合损失函数,约束不确定性图聚焦于退化区域。最后在雨滴去除、雨纹去除和真实雾霾去除任务上的实验验证了所提框架的有效性与泛化能力。

1 相关工作

1.1 图像雨滴去除与多尺度特征学习

图像雨滴去除旨在从附着雨滴的退化图像中恢复清晰背景,对自动驾驶、目标检测等下游任务具有重要意义。早期方法基于物理建模,利用雨滴的光学特性或几何形状进行显式表示,如 Roser 和 Geiger

(2009)通过对雨滴光度特性建模实现检测与去除。此类方法无需成对数据,但受限于先验表达能力,难以适应真实场景中雨滴形态的多样性。深度学习的引入推动了该领域的发展,Eigen等(2013)最早使用卷积网络处理镜头附着物去除任务,Qian等(2018)设计具有注意力机制的生成对抗网络,通过生成器与判别器的对抗学习提升去雨滴效果。Quan等(2019)提出编码器-解码器架构,引入边缘注意力和通道注意力以增强雨滴区域检测与特征鲁棒性。Liu等(2019)研究深度残差网络,提出对偶残差连接以提升网络表现。Zhang和Patel(2018)采用双分支架构分别关注大尺度与小尺度雨滴区域,并将分支结果融合得到细化输出。Shao等(2021)使用连续掩码表示雨滴模糊程度以处理不同透明度的退化。此外,国内学者也在该方向开



((a) Spatial Fusion Network Block; (b) Spatial-Frequency Multi-Scale Adaptive Attention Block (MAA-F))

图1 基于空频联合与不确定性引导的多尺度稀疏变换器(SFFormer)的整体架构

Fig. 1 Overall Architecture of the Spatial-Frequency Joint and Uncertainty-Guided Multi-Scale Sparse Transformer (SFFormer)

展了深入研究,如汤红忠等(2024)提出基于双阶段特征解耦网络的单幅图像去雨方法,采用渐进式学习方式从粗到细分阶段进行去雨,实现了背景图像与有雨分量的有效分离。胡明娣等(2022)对单幅图像去雨数据集与深度学习算法进行了系统的评估;王美华等(2022)将Transformer与3D注意力机制相结合,提出了融合3D注意力和Transformer的图像去雨网络,有效改善了单幅图像的去雨效果。

多尺度特征提取能帮助模型捕捉不同尺度的退化模式。Ronneberger等(2015)提出的U型网络(U-

Net)编码器-解码器架构通过下采样与上采样的组合实现多尺度特征提取,成为图像修复的经典范式。在雨滴去除任务中,Zhang和Patel(2018)设计双分支架构分别关注大尺度与小尺度雨滴区域,Shao等(2021)提出多尺度注意力网络以适应雨滴尺寸变化。然而,上述方法虽在雨滴去除上取得进展,但对雨滴退化的异质性特征本质建模仍显不足。雨滴图像中既存在中心区域的均匀模糊,又存在边缘区域的复杂折射畸变,这两类物理属性截然不同的退化模式需要差异化的特征处理,现有方法难以同时满

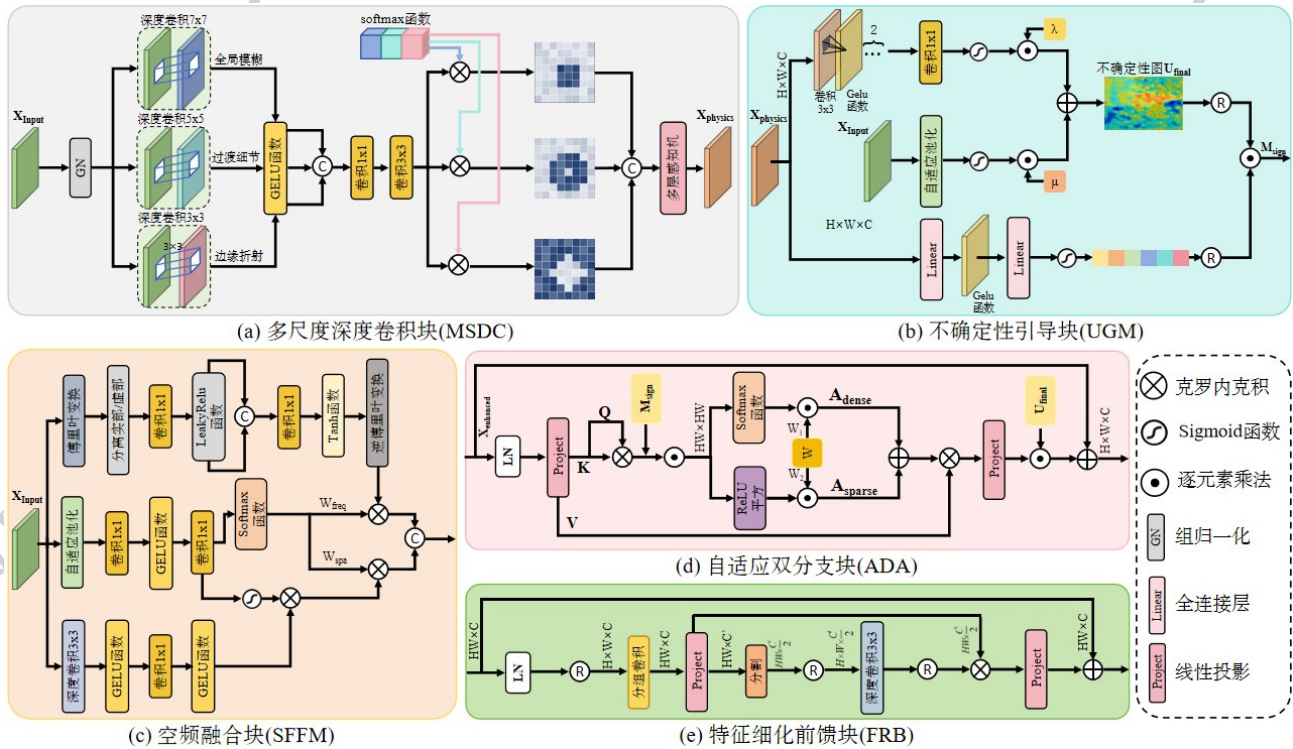
足这一需求。

1.2 注意力机制与稀疏 Transformer

注意力机制在视觉感知中具有重要作用,能够引导模型聚焦于关键信息区域。Hu 等(2018)提出的 Squeeze-and-Excitation 网络是通道注意力的经典代表,Woo 等(2018)将通道与空间注意力串联提出卷积块注意力模块。在图像修复领域,Zhang 等(2018)将通道注意力引入图像超分辨率任务,Yang 等(2020)利用通道注意力与分层残差网络实现图像修复。在雨滴去除任务中,Qian 等(2018)在生成对抗网络中引入视觉注意力模块,Quan 等(2019)将边缘注意力与通道注意力结合,Zhang 和 Patel(2018)利用双分支注意力关注不同尺度雨滴区域,Shao 等(2021)提出多尺度注意力网络并引入不确定性引导。林晓等(2022)提出利用多尺度特征联合注意力模型。上述方法虽验证了注意力的有效性,但多使用单一类型或简单组合,尚未充分探索多重注意力的协同作用,且注意力权重学习主要基于特征统计

特性,缺乏与雨滴退化模式的显式关联。

Transformer 凭借全局自注意力在计算机视觉中广泛应用。IPT(Chen 等,2021)首次将 Transformer 用于图像修复,Uformer(Wang 等,2022)设计窗口注意力平衡全局建模与效率,Restormer(Zamir 等,2022)采用通道注意力降低开销,SwinIR(Liang 等,2021)利用移位窗口策略促进跨窗口交互。稀疏性研究方面,Sparse Transformer(Correia 等,2019)提出因子化自注意力,DRSformer(Chen 等,2023)引入 Top-K 通道选择实现稀疏化,CODE(Ye 等,2023)将特征投影至超像素空间,AST(Zhou 等,2024)融合稠密与稀疏自注意力。然而,DRSformer(Chen 等,2023)的 Top-K 选择全局固定;AST(Zhou 等,2024)的决策基于特征能量而非退化属性;USTN(Chen 等,2023)的不确定性引导仅用于降低计算量,未反馈至注意力权重。因此,它们均无法根据中心模糊与边缘折射的异质性需求实现差异化处理。



((a) Multi-Scale Depthwise Convolutional block; (b) Uncertainty Guidance Module; (c) Spatial-Frequency Fusion Module; (d) Adaptive Dual-branch Attention; (e) Feature Refinement Feed-Forward Block)

图2 核心模块内部结构示意图

Fig. 2 Schematic Diagram of Internal Structure of Core Modules

1.3 频域分析与空域频域联合处理

频域分析在图像处理中具有独特优势,傅里叶

变换将图像分解为幅度与相位分量,分别对应能量分布与结构信息(Oppenheim 和 Lim, 2005)。近年

来,部分研究探索将频域信息引入深度学习模型,Wavelet-CNN(Liu等,2018;Fujieda等,2017)将小波变换与卷积网络结合以增强多尺度表示,FFT-based方法(温博等,2011)利用快速傅里叶变换(fast Fourier transform,FFT)加速大尺寸图像处理。在图像修复任务中,频域处理有助于捕捉周期性退化模式与边缘方向信息(Xu等,2020)。然而,现有方法多将频域作为空域的补充工具,尚未形成系统的空域与频域联合建模框架(Chen等,2023)。雨滴边缘的折射畸变在空域表现为复杂的纹理变化,在频域则对应高频分量的局部增强,若能实现空域与频域的联合建模与互补增强,有望更全面地捕捉雨滴退化的多维度信息,但目前鲜有研究将频域信息显式引入雨滴去除任务。

针对上述不足,本文提出一种基于空域与频域联合处理及不确定性引导的多尺度稀疏变换器,通过多尺度退化解耦、空域频域互补增强及不确定性引导的注意力分配,实现雨滴图像中异质性退化的有效修复。

2 本文方法

本节首先介绍本文提出的基于空频联合与不确定性引导的多尺度稀疏变换器(SFFormer)的整体架构,然后分别详细阐述所设计的多尺度深度卷积块、空频融合块、不确定性引导块、自适应双分支块、特征细化前馈块以及空频多尺度自适应注意力块,最后介绍模型训练所使用的损失函数。

2.1 网络模型的整体结构

如图1(a)所示,本文提出的空频联合与不确定性引导的多尺度稀疏变换器(SFFormer)采用对称的编码器-解码器(U-Net)架构,包含4层编码器和4层解码器:输入雨滴图像首先经 3×3 卷积映射为初始特征,编码器每层堆叠多个特征细化前馈块(FRB)后跟下采样以扩大感受野,瓶颈层由空频多尺度自适应注意力块(Multi-scale Adaptive Attention with Frequency, MAA-F)构成,且根据预设仅在瓶颈层和解码器最后两层启用空频融合块(SFFM)以强化深层频域建模;解码器每层先上采样恢复分辨率,再与对应编码器层输出跳跃连接融合后送入MAA-F细化处理,最后解码器输出经 3×3 卷积生成残差图像并与输入相加得到去雨结果,从而实现高质量图像

修复。

最终,解码器最后一层输出通过一个 3×3 卷积生成残差图像 R ,去雨结果表示为 $I_{out} = I_{in} + R$ 。整个网络通过分层递进的特征提取与融合,实现高质量的图像雨滴去除。

2.2 多尺度深度卷积块

针对雨滴图像中同时存在中心均匀模糊与边缘复杂折射的异质退化模式,本文设计了多尺度深度卷积块(Multi-Scale Depthwise Convolutional block, MSDC),MSDC的核心假设是:不同膨胀率与核大小的深度卷积具有不同的频域响应特性,因此可以引导网络倾向性提取不同频率的退化成分,而非严格意义上的信号分离。

如图2(a)所示,给定输入特征 $X \in \mathbf{R}^{C \times H \times W}$,MSDC首先采用三条具有不同核大小与膨胀率的深度可分离卷积分支进行特征提取:

$$F_s = \phi(\text{DWConv}_{k_s, d_s}(X)) \quad (1)$$

式中 $s \in \{1, 2, 3\}$,并且 DWConv_{k_s, d_s} 表示核大小为 k_s 、膨胀率为 d_s 的深度卷积, ϕ 为高斯误差线性单元(Gaussian error linear unit, GELU)激活函数。分支 $s=1$ 采用大核低膨胀率($k_1=7, d_1=1$)具有低通滤波特性,以捕获广域均匀模糊区域的低频信息,对应深度卷积核 7×7 ;分支 $s=2$ 采用中等核与适度膨胀($k_2=5, d_2=2$)可覆盖中频过渡带,以保持模糊与清晰区域间的中频过渡结构,对应深度卷积核 5×5 ;分支 $s=3$ 采用小核高膨胀率($k_3=3, d_3=3$)且小核本身具有高通滤波特性,以聚焦雨滴边缘高频折射畸变的局部细节,对应深度卷积核 3×3 。各分支的有效感受野在保持参数效率的同时实现了差异化的空间覆盖能力。

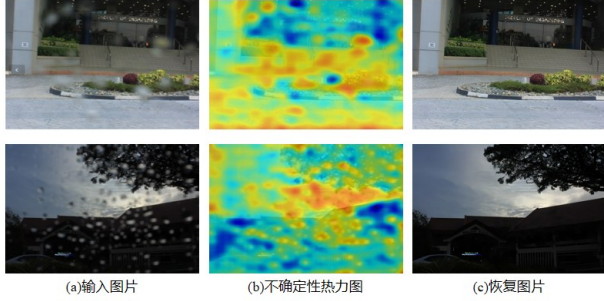
为融合三个分支的多尺度特征,MSDC设计了跨尺度交互单元。首先将分支输出拼接为 $Z = [F_1, F_2, F_3] \in \mathbf{R}^{3C \times H \times W}$,然后通过一个小型卷积网络生成空间注意力权重:

$$[w_1, w_2, w_3] = \text{Softmax}(f_{3 \times 3}(f_{1 \times 1}(Z))) \quad (2)$$

式中 $f_{1 \times 1}$ 和 $f_{3 \times 3}$ 分别为 1×1 卷积和 3×3 卷积,SoftMax沿通道维度归一化使得每个空间位置的三分支权重和为1。加权融合时,三个分支的特征先与对应权重逐元素相乘,再引入轻量的残差连接以稳定训练;随后将更新后的三个分支特征拼接,通过一个由两个 1×1 卷积和GELU激活构成的轻量多层感知机

(multi-layer perceptron, MLP)进行通道重组与非线性变换,最后与输入 X 相加得到输出特征 $X_{physics} \in \mathbb{R}^{C \times H \times W}$ 。

MSDC使模型能根据像素位置动态选择合适感受野,为后续处理提供富含区域退化模式信息的多尺度特征表示。



((a) Input image; (b) Uncertainty heatmap; (b) Restored image)

图3 不确定性引导块可视化示例

Fig. 3 Visualization Example of Uncertainty Guidance Block

2.3 不确定性引导块

针对图像区域修复难度差异导致计算资源分配均质化的问题,本文设计了不确定性引导块(Uncertainty Guidance Module, UGM)。

如图2(b)所示,给定由MSDC输出的退化解耦特征 $X_{physics} \in \mathbb{R}^{C \times H \times W}$,UGM首先通过不确定性图生成网络估计每个像素位置的修复难度。该网络由三个连续的“3×3卷积+GELU”操作和一个1×1卷积构成,最后经Sigmoid激活输出不确定性图 $U \in \mathbb{R}^{1 \times H \times W}$:

$$U = \sigma \left(f_{1 \times 1} \left(\phi_{3 \times 3}^3 (X_{physics}) \right) \right) \quad (3)$$

式中 $\phi_{3 \times 3}^3(\cdot)$ 表示三层堆叠, σ 为Sigmoid函数。 U 中每个元素 u_{ij} 量化了位置 (i,j) 的修复难度,值越接近1表示该区域越难修复(如雨滴边缘折射区),越接近0则表示简单背景。

图3展示了由所提不确定性引导块生成的不确定性图示例。从图中可以看出,确定性图在雨滴的边缘折射区域和中心半透明遮挡区域呈现出较高的响应(图中红色高亮部分),而在背景区域响应较低,表明模型能够准确感知不同区域的修复难度。结合后续的注意力调制,该模块使模型在高不确定性区域保留精细纹理,在低不确定性区域抑制冗余计算,从而实现计算资源的自适应分配。

为避免单一特征源带来的估计偏差,将原始输入特征的全局上下文融入不确定性图。首先对 X 进行全局平均池化,再经Sigmoid激活得到全局上下文图,与 U 加权融合:

$$U_{final} = \lambda \cdot U + \mu \cdot \sigma(AAP(X)) \quad (4)$$

式中 λ, μ 为可学习的平衡系数, $AAP(\cdot)$ 表示自适应平均池化(Adaptive Average Pooling, APP),后经广播操作恢复至空间尺寸。融合后的 U_{final} 兼具局部特性和全局语义信息。

与此同时,门控分支将输入特征 X 重塑为序列后依次通过两个线性层和Sigmoid激活,得到门控向量 g ,用以衡量各位置在重建任务中的全局语义重要性。

最终,将不确定性图序列化为 U_{seq} 并与门控向量逐元素相乘得到调制信号 $M_{sign} = g \cdot U_{seq}$ 。该信号用于动态调整后续注意力得分,使模型在高不确定性区域增强响应,在低不确定性区域抑制冗余计算。

2.4 空频融合块

针对单一域处理难以全面捕捉雨滴图像多维退化特征的问题,本文提出空频融合块(Spatial-Frequency Fusion Module, SFFM),通过并行空域与频域分支互补增强特征表示,并引入动态门控机制实现自适应融合,如图2(c)所示。

给定输入特征 $X \in \mathbb{R}^{C \times H \times W}$,SFFM首先分别通过空域分支与频域分支进行特征提取。

空域分支采用轻量级空域处理块,其由深度可分离卷积与通道注意力组成,可高效捕获局部纹理信息:

$$X_{spa} = X + \text{Attn}(f_{dconv}(X)) \odot f_{dconv}(X) \quad (5)$$

式中 f_{dconv} 为3×3深度可分离卷积, $\text{Attn}(\cdot)$ 为基于全局平均池化的通道注意力, $\odot$ 表示逐元素相乘。

频域分支将输入通过实部快速傅里叶变换转换到频域,得到复数张量 $\tilde{X} \in \mathbb{C}^{C \times H \times (W/2 + 1)}$ 。随后采用轻量级复数调制网络对其实部与虚部进行联合建模。该网络分别对实部 $\tilde{X}_{real}$ 和虚部 $\tilde{X}_{imag}$ 应用1×1卷积,并将输出拼接后通过Tanh激活生成调制系数 M :

$$M = \tanh(f_{fuse}([f_{real}(\tilde{X}_{real}), f_{imag}(\tilde{X}_{imag})])) \quad (6)$$

式中 f_{real}, f_{imag} 为1×1卷积, f_{fuse} 为融合卷积。调制后的频域特征为:

$$\tilde{X}_{out} = \tilde{X} \odot (1 + \alpha \cdot M) + 0.1\tilde{X} \quad (7)$$

式中 α 为可学习的全局缩放因子(初始化为0.01),残差连接保证了原始信息的保留。最后通过逆傅里叶变换得到空域特征 $X_{\text{freq}} \in \mathbf{R}^{C \times H \times W}$ 。

为自适应融合两个分支的输出,SFFM设计了动态门控单元。首先对输入 X 进行全局平均池化,得到全局上下文向量;然后通过两个 1×1 卷积层和一个Softmax函数预测两个分支的融合权重:

$$[w_{\text{spa}}, w_{\text{freq}}] = \text{Softmax}(f_{\text{gate}}(\text{GAP}(X))) \quad (8)$$

式中 $\text{GAP}(\cdot)$ 为全局平均池化, f_{gate} 由两个 1×1 卷积(中间含GELU激活)组成,输出维度为2。权重 $w_{\text{spa}}, w_{\text{freq}} \in \mathbf{R}^{1 \times 1 \times 1}$ 在通道和空间维度上广播,实现逐样本的加权融合。为增强信息流动性,将融合特征与原始输入拼接后送入特征融合层。

SFFM使模型能够根据图像内容自适应结合局部纹理与全局频域信息,为后续注意力机制提供更全面的特征表示。

2.5 自适应双分支注意力

是为将退化属性引导的不确定性信息融入注意力决策,本文设计自适应双分支块(Adaptive Dual-branch Attention, ADA)。该模块接收由MSDC和SFFM融合后的增强特征 $X_{\text{enhanced}} \in \mathbf{R}^{H \times W \times C}$,以及来自UGM的调制信号 $\mathbf{M}_{\text{sign}}$ 和最终不确定性图 $\mathbf{U}_{\text{final}}$,通过不确定性调制、双分支融合与残差加权,实现不确定性引导的自适应注意力计算。

首先,将输入特征 X 通过线性投影生成查询 Q 、键 K 、值 V 。初步注意力分数矩阵 A 由 Q 与 K 的点积除以 $\sqrt{d}$ 并加上相对位置偏置 B 得到。为引导注意力聚焦于高难度区域,利用调制信号 $\mathbf{M}_{\text{sign}}$ 对 A 进行动态调制:

$$\mathbf{A}_{\text{mod}} = A \odot (1 + \gamma \cdot \tanh(\mathbf{M}_{\text{sign}})) \quad (9)$$

式中 γ 为可学习缩放系数。调制后的分数 $\mathbf{A}_{\text{mod}}$ 同时蕴含了特征相似度与退化属性不确定性信息。

在双分支融合中,ADA采用Softmax密集分支与平方修正线性单元(rectified linear unit, ReLU)稀疏分支分别保持稳定性和主动稀疏化,并通过可学习参数加权求和得到注意力输出;最后利用不确定性图 $\mathbf{U}_{\text{final}}$ 作为软掩码调节残差更新幅度,使模型根据区域修复难度自适应调整特征更新量。

2.6 特征细化前馈块

为在Transformer层中增强局部细节建模能力,本文设计特征细化前馈块(Feature Refinement Feed-

Forward Block, FRB)。如图2(e)所示,FRB通过引入深度卷积与门控机制,在保持计算效率的同时实现特征的精细化表达。

给定输入序列 $X \in \mathbf{R}^{C \times H \times W}$,FRB首先通过一个全连接层将通道数扩展为原来的4倍,并应用GELU激活:

$$\mathbf{X}_m = \text{GELU}(XW_1 + b_1) \quad (10)$$

式中 $W_1 \in \mathbf{R}^{C \times 4C}$ 。为引入局部空间交互,将 $\mathbf{X}_m$ 重塑为二维特征图,并应用 3×3 深度可分离卷积进行空间邻域特征提取。随后将卷积后的特征展平回序列形式,并与原始 $\mathbf{X}_m$ 逐元素相乘实现门控调制。最后通过第二个全连接层将通道数恢复为原始维度 C ,得到输出 $\mathbf{X}_{\text{out}}$ 。整个过程可表示为:

$$\mathbf{X}_{\text{out}} = (\mathbf{X}_m \odot \text{DWConv}(\mathbf{X}_m))W_2 + b_2 \quad (11)$$

式中 $\text{DWConv}(\cdot)$ 包含必要的二维重塑与展平操作。

FRB通过升维—空间卷积—门控降维的架构,有效增强了特征的局部细节表达能力,同时保持了Transformer的全局建模能力。

2.7 损失函数

本文采用不确定性感知的复合损失函数监督网络训练,在保证重建精度的同时引导不确定性图学习具有明确指示意义的分布。总损失由Charbonnier重建损失、稀疏性损失和平滑性损失三部分构成:

$$L_{\text{total}} = \lambda_{\text{char}} L_{\text{char}} + \lambda_{\text{sparse}} L_{\text{sparse}} + \lambda_{\text{tv}} L_{\text{tv}} \quad (12)$$

Charbonnier损失(Charbonnier等,1994)用于约束重建图像 $\mathbf{I}_{\text{out}}$ 与真实图像 $\mathbf{I}_{\text{gt}}$ 的像素级差异,其平滑特性可增强训练的稳定性:

$$L_{\text{char}} = \sqrt{\|\mathbf{I}_{\text{out}} - \mathbf{I}_{\text{gt}}\|^2 + \varepsilon^2} \quad (13)$$

式中 $\varepsilon = 10^{-3}$ 为常数。

稀疏性损失鼓励不确定性图 $\mathbf{U}$ 仅在雨滴遮挡等难以修复区域具有较大响应,而在简单背景区域趋近于零:

$$L_{\text{sparse}} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_{ij} \quad (14)$$

全变分损失(Rudin等,1992)用于平滑不确定性图,避免因噪声干扰产生不连续的突变:

$$L_{\text{tv}} = \frac{1}{H \times W} \left(\sum_{i=1}^{H-1} \sum_{j=1}^W (U_{i+1,j} - U_{ij})^2 + \sum_{i=1}^H \sum_{j=1}^{W-1} (U_{i,j+1} - U_{ij})^2 \right) \quad (15)$$

该复合损失使网络在精确重建的同时,学习到具有明确区域修复难度指示意义的不确定性分布,为后续的注意力调制提供有效指导。

3 实验

3.1 数据集和实验细节

为全面评估所提方法的性能,本文在三个具有代表性的图像修复数据集上开展实验:雨滴去除采用广泛使用的 Raindrop 数据集 (Qian 等, 2018), 该数据集通过在相机镜头前放置附有水滴的玻璃拍摄真实雨滴图像, 训练集包含 861 对图像, 测试集

仅使用严格对齐的 test-a 子集 (58 对图像); 雨纹去除在 SPAD 数据集 (Wang 等, 2019) 上进行评估, 该数据集包含 14 000 对训练图像 (有雨/清晰) 和 1 000 对测试图像, 涵盖多种雨纹模式和强度; 真实雾霾去除采用 Dense-Haze 数据集 (Ancuti 等,

2019) 进行去雾性能验证, 该数据集包含真实世界的密集雾霾图像及其对应的清晰图像, 为去雾任务提供了具有挑战性的真实场景。所有图像在输入网络前均被裁剪为 256×256 的图像块, 每个数据集的训练集和测试集划分均严格遵循其原始设定, 以确保与现有方法的公平比较。所有实验均在相同的评估设置下在相应的测试集上完成。

在实验细节上本文方法基于 PyTorch 深度学习框架实现, 使用 NVIDIA A100 GPU 进行训练和测试。模型中的特征基础通道数设置为 $C = 32$, 编码器各层通道维度逐层增至 $\{32, 64, 128, 256, 512\}$ 。编码器包含四个阶段, 分别堆叠 $[1, 2, 8, 8]$ 个处理块 (FRB); 瓶颈层包含 2 个块; 解码器对应阶段堆叠 $[8, 8, 2, 1]$ 个处理块。所有自注意力模块采用 8×8 的窗口大小, 多头注意力头数设置为 $[1, 2, 4, 8, 8, 4, 2, 1]$, 并遵循 Zamir 等 (2022) 的方法使每个头的特征维度保持一致。

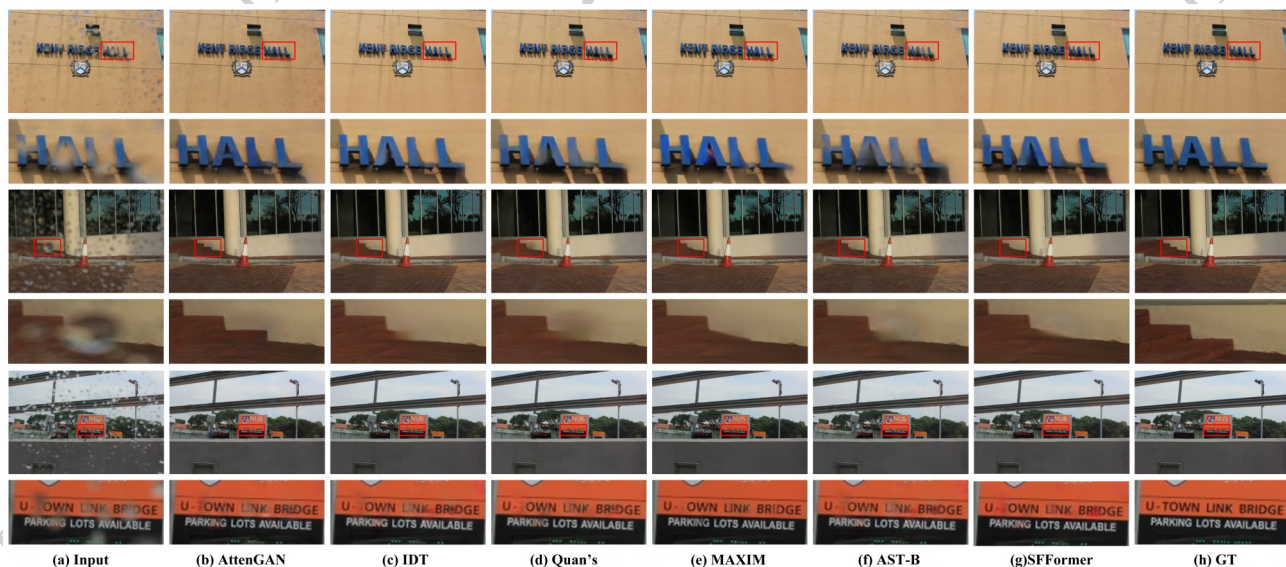


图4 在 AGAN 数据集上不同方法的雨滴去除的视觉对比

Fig. 4 Visual Comparison of Different Methods for Raindrop Removal on the AGAN Dataset

训练采用 AdamW (Loshchilov 和 Hutter, 2019) 优化器, 初始学习率设为 2×10^{-4} , 通过余弦衰减 (Loshchilov 和 Hutter, 2017) 策略降至 1×10^{-6} 。损失函数采用第 2.7 节所述的复合损失函数, 其中稀疏性损失权重 $\lambda_{\text{sparse}} = 0.1$, 全变分损失权重 $\lambda_{\text{tv}} = 0.01$ 。训练过程中, 输入图像随机裁剪为 128×128 像素块以节省训练时间, 而在计算复杂度时使用 256×256 输入尺寸。应用随机翻转和旋转增强,

渐进式学习策略用于节省时间, 类似于 Ren 等 (2019) 和 Zamir 等 (2021)。

3.2 实验结果

为验证本文方法的有效性, 我们在三个公开数据集上进行了定量对比实验: AGAN 雨滴去除数据集、Dense Haze 真实雾霾数据集以及 SPAD 雨纹去除数据集。评价指标采用峰值信噪比 (PSNR) 和结构相似性 (SSIM)。

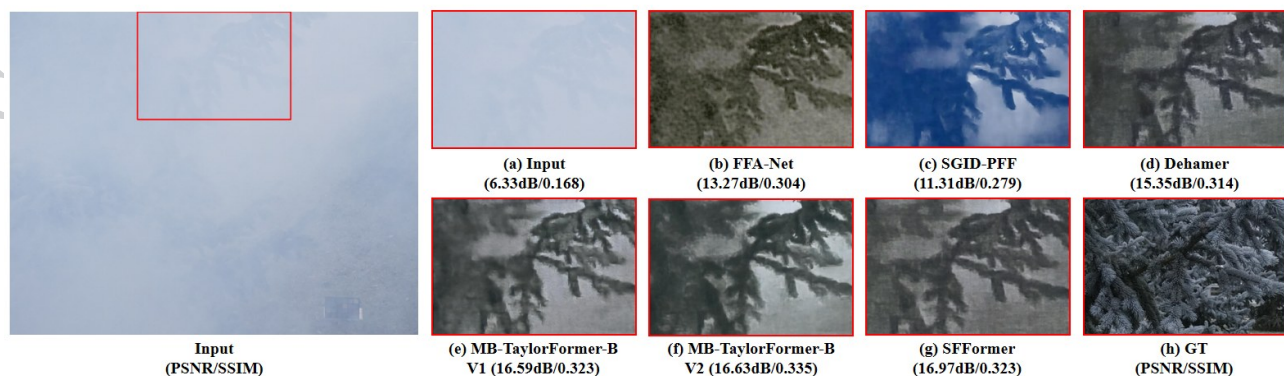


图5 在Dense-haze数据集上不同方法的密集雾霾去除的视觉对比

Fig. 5 Visual Comparison of Different Methods for Dense Haze Removal on the Dense-Haze Dataset

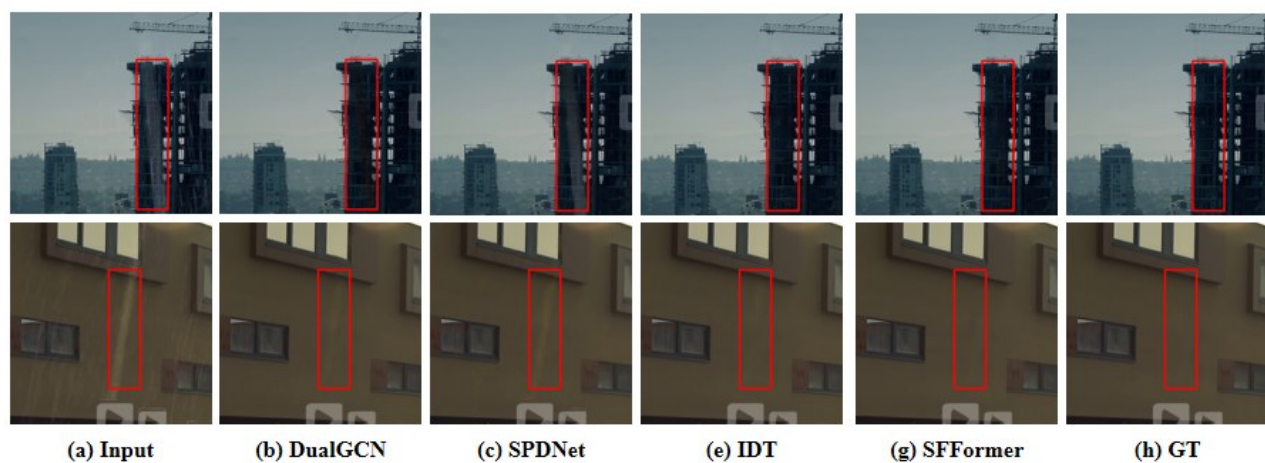


图6 在SPAD数据集上不同方法的雨纹去除的视觉对比

Fig. 6 Visual Comparison of Different Methods for Rain Streak Removal on the SPAD Dataset

AGAN数据集(Qian等,2018)结果如表1所示。本文提出的SFFormer取得了32.97dB的PSNR和0.942的SSIM,均优于所有对比方法。与次优方法WeatherNet(Xu等,2025)(32.67dB)相比,PSNR提升了0.30dB,SSIM达到与最优方法持平的水平,验证了所提模块在异质性雨滴退化修复中的优势。

Dense-Haze数据集(Ancuti等,2019)结果如表2所示。SFFormer的PSNR达到17.51dB,高于所有对比方法,比次优的MB-TaylorFormer-B(Jin等,2025)(16.95dB)提升0.56dB。然而,本文方法的SSIM为0.54,低于ConvIR-B(0.61)和MB-TaylorFormer-B(0.62)。分析其原因,Dense-Haze为均匀密集雾霾场景,退化缺乏空间异质性,本文专为异质性退化设计的解耦与稀疏注意力模块在此产生一定冗余,可能过度平滑了局部结构;同时,以Charbonnier损失为主的优化目标对结构相似性的约束弱于SSIM损

失。尽管如此,PSNR的领先仍表明本文方法在像素级重建精度上具有优势。

SPAD数据集(Wang等,2019)结果如表3所示。在雨纹去除任务中均优于对比方法。与次优方法IDT(Xiao等,2022)(47.34dB)相比,PSNR提升了0.31dB,SSIM指标也达到次优水平,验证了所提框架在不同图像修复任务中的良好泛化能力。

我们在图4、图5和图6中分别展示了AGAN-Data(Qian等,2018)雨滴去除、Dense-Haze(Ancuti等,2019)去雾和SPAD(Wang等,2019)去雨纹的视觉对比结果。雨滴图像存在中心遮挡与边缘折射的异质性退化,雾霾图像全局对比度下降,雨纹图像则呈现复杂的线条状退化模式。现有方法要么残留退化伪影,要么过度平滑纹理细节。本文SFFormer通过空域-频域联合处理模块,在空域增强局部纹理的同时在频域恢复全局结构;结合多尺度退化解耦模

表 1 不同方法在 AGAN 测试集 test-a 上的对比结果
Table 1 Comparison of Different Methods on the AGAN Test-a Subset

方法	会议/期刊	PSNR/ dB	SSIM
AttenGAN (Qian 等, 2018)	CVPR' 2018	31.43	0.926
Quan's (Quan 等, 2019)	ICCV' 2019	31.37	0.918
CCN (Quan 等, 2021)	CVPR' 2021	31.34	0.929
Uformer (Wang 等, 2022)	CVPR' 2022	29.42	0.929
IDT (Xiao 等, 2022)	TPAMI' 2022	31.87	0.931
MAXIM-2S (Tu 等, 2022)	CVPR' 2022	31.87	<u>0.935</u>
USTN (Chen 等, 2023)	TIM' 2024	31.62	0.924
AST-B (Zhou 等, 2024)	CVPR' 2024	32.10	<u>0.935</u>
RQ-D ² UNet (Gai 和 Ni, 2025)	TIP' 2025	32.42	0.942
WeatherNet (Xu 等, 2025)	TCSVT' 2025	<u>32.67</u>	0.942
SFFormer	—	32.97	0.942

注:粗体表示最优值,下划线表示次优值。

块分离不同退化模式,并利用不确定性引导的稀疏注意力聚焦修复难点,从而在有效去除各类退化的同时保持图像细节和边缘锐利度,验证了我们的方法的有效性和泛化能力。

3.4 消融实验

3.4.1 模块有效性分析

为了验证所提各个模块的有效性,本文在

AGAN-Data 雨滴数据集 (Qian 等, 2018) 的 test-a 子集上进行了消融实验。分别对多尺度深度卷积块 (MSDC)、不确定性引导块 (UGM) 和空频融合块 (SFFM) 进行组合测试,实验结果如表 4 所示。基线模型 (不包含任何所提模块) 采用简单的两层卷积残差模块替代上述模块,其余网络结构保持不变。所有实验均在相同训练设置下进行,以确保公平对比。

从表 4 可以看出,单独引入任一模块时,PSNR 均低于基线,这可能是由于单个模块的引入破坏了原有残差结构的特征分布,导致训练初期难以收敛。然而,当模块组合使用时,性能显著提升,表明本文所提模块之间存在着良好的协同效应。

具体而言,当同时引入 MSDC 与 SFFM 时,PSNR 达到 32.86 dB,较基线提升 0.10 dB,说明多尺度感受野与空频联合处理的互补建模能够有效协同,共同提升特征表达能力。在此基础上进一步加入 UGM,完整模型取得最佳性能,PSNR 达 32.97 dB,SSIM 达 0.942,较“MSDC+SFFM”组合提升 0.11 dB,较基线提升 0.21 dB。这表明不确定性引导能够有效聚焦于高难度区域,与多尺度解耦和空频增强形成互补,进一步提升修复质量。

此外,“MSDC+SFFM”与“MSDC+UGM”组合可见,前者 (32.86 dB) 优于后者 (32.80 dB),说明空频融合块对性能的贡献略大于不确定性引导块,反映出频域信息建模在雨滴边缘折射恢复中的独特优

表 2 不同方法在 Dense-Haze 数据集上的对比结果
Table 2 Comparison of Different Methods on the Dense-Haze Dataset

方法	会议/期刊	PSNR/dB	SSIM
AOD-Net (Li 等, 2017)	ICCT' 2017	13.14	0.41
GridDehazeNet (Liu 等, 2019)	ICCV' 2019	13.31	0.37
FFA-Net (Qin 等, 2020)	AAAI' 2020	14.39	0.54
SGID (Bai 等, 2022)	TIP' 2022	13.09	0.52
Uformer (Wang 等, 2022)	CVPR' 2022	15.22	0.43
Restormer (Zamir 等, 2022)	CVPR' 2022	15.78	0.55
DeHamer (Guo 等, 2022)	CVPR' 2022	16.62	0.56
MB-TaylorFormer-s (Qiu 等, 2023)	TIP' 2023	16.29	0.51
ConvIR-B (Cui 等, 2024)	TPAMI' 2024	16.86	<u>0.61</u>
MB-TaylorFormer-B (Jin 等, 2025)	TPAMI' 2025	<u>16.95</u>	0.62
SFFormer	—	17.51	0.54

注:粗体表示最优值,下划线表示次优值。

表3 不同方法在SPAD数据集上的对比结果

Table 3 Comparison of Different Methods on the SPAD

Dataset		
方法	PSNR/dB	SSIM
SPDNet(Yi等,2021)	43.20	0.9871
DualGCN(Fu等,2021)	44.18	0.9902
IDT(Xiao等,2022)	47.34	0.9929
Fu等(2023)	45.03	0.9907
SCD-Former(Guo等,2023)	46.86	0.9941
SFFormer	47.65	<u>0.9931</u>

注:粗体表示最优值,下划线表示次优值。

表4 SFFormer模块的消融实验

Table 4 Comparison of Different Methods on the SPAD

Dataset				
MSDC	UGM	SFFM	PSNR/dB	SSIM
×	×	×	32.76	0.940
×	×	√	32.35	0.934
×	√	×	32.49	0.938
×	√	√	32.69	0.940
√	×	×	32.66	0.937
√	×	√	32.86	0.941
√	√	×	32.80	0.941
√	√	√	32.97	0.942

注:粗体表示最优值

3.3 视觉效果对比

势。而所有包含UGM的组合均优于对应不含UGM的组合,证明不确定性估计模块在不同特征基础上均能带来稳定增益。

综上所述,消融实验结果充分验证了所提三个模块各自的有效性及其协同作用,为模型性能提供了可靠保障。

3.4.2 损失函数超参数分析

如第2.7节所述,本文采用复合损失函数,包含Charbonnier重建损失、不确定性图的稀疏性损失(权重 λ_{sparse})和全变分平滑损失(权重 λ_{tv})。为分析这两个正则项对模型性能的影响,我们在AGAN-Data测试集上进行了超参数消融实验,结果如表5所示。

从表5可以看出,当同时移除稀疏性和平滑性正则化(实验②)时,PSNR相比默认设置下降0.43 dB,SSIM下降0.005,证明两项正则项对提升修复

表5 损失函数超参数消融实验结果

Table 5 Loss function hyperparameter ablation study

方法	λ_{sparse}	λ_{tv}	PSNR/dB	SSIM
①(默认)	0.1	0.01	32.97	0.942
②	0	0	32.54	0.937
③	0.05	0.01	32.85	0.940
④	0.2	0.01	32.61	0.934
⑤	0.1	0	32.23	0.934
⑥	0.1	0.05	32.31	0.936

注:粗体表示最优值

质量均具有积极作用。

对比实验①、③、④可以看出,固定 $\lambda_{\text{tv}}=0.01$ 时, λ_{sparse} 从默认值0.1减小到0.05(实验③)或增大到0.2(实验④),PSNR分别下降0.12 dB和0.36 dB,SSIM也均有下降。这表明过小的稀疏性惩罚无法有效鼓励不确定性图聚焦于退化区域,而过大的稀疏性惩罚则可能过度抑制不确定性响应,导致模型对难修复区域的关注不足。

进一步分析TV损失的影响(实验①、⑤、⑥)。当 $\lambda_{\text{tv}}=0$ 时(实验⑤),PSNR降至32.23 dB,SSIM降至0.934,性能下降最为显著,且观察不确定性图发现大量离散噪声点(此处可辅以可视化说明),说明TV损失对抑制不确定性图的随机波动至关重要。当 λ_{tv} 增大至0.05(实验⑥)时,性能同样低于默认值,原因是过强的平滑性约束会使不确定性图过于均匀,无法准确反映区域修复难度的差异。

综上所述,默认超参数 $\lambda_{\text{sparse}}=0.1$ 、 $\lambda_{\text{tv}}=0.01$ 在稀疏性聚焦与平滑性之间取得了最佳平衡,后续实验均采用此设置。

3.4.3 模块有效性分析

为了进一步量化各模块的“参数-性能”权衡,本文基于组合消融实验对三个核心模块的独立参数

量进行了减法拆解,结果如表6所示。SFFM、UGM和MSDC的参数量分别为8.22M、9.09M和15.78M,三者合计为33.09M。其中MSDC作为多尺度退化解耦的核心,参数量占比最大(15.78M),这是实现异质性退化建模的必要代价;而SFFM和UGM合计仅占三者总和的52.3%(17.31M),表明空频融合与不确定性引导以较小的参数开销实现了关键的功能增强。

表6 损失函数超参数消融实验结果

Table 6 Loss function hyperparameter ablation study

模型	参数量 (M)	占完整 模型比 例	PSNR 提升 (dB/M)
SFFM(空频融合)	8.22	7.4%	+0.10 dB
UGM(不确定性引导)	9.09	8.2%	+0.04 dB
MSDC(多尺度解耦)	15.78	14.2%	-0.10 dB

值得注意的是,UGM 仅以 9.09M 的参数,便在“MSDC+SFFM”组合(32.86dB)的基础上进一步将 PSNR 提升至 32.97dB,贡献了 0.11dB 的纯增益;SFFM 以 8.22M 的参数,与 MSDC 协同带来了 0.10dB 的提升(32.76dB 到 32.86dB)。这表明,虽然 MSDC 占用了较多参数用于物理先验建模,但新增的 UGM 和 SFFM 均以较低参数代价换取了可观的性能收益,验证了本文设计的参数效率。

此外,本文所设计的不确定性引导稀疏注意力机制,其核心价值在于动态抑制平坦背景区域的冗余计算,而非单纯压缩模型体积。因此,模型参数规模与推理时的实际计算量(FLOPs)并非严格正相关,这为后续通过剪枝与蒸馏实现轻量化部署预留了充足空间。

4 结 论

本文针对图像修复中空间异质性退化(如雨滴同时呈现中心模糊与边缘折射)的恢复难题,提出了一种空频联合与不确定性引导的多尺度稀疏变换器(SFFormer)。该方法通过空频联合增强实现局部纹理与全局结构的互补建模,引入不确定性估计量化区域修复难度,并设计稀疏注意力机制实现计算资源的动态分配,有效解决了异质性退化对模型提出的差异化处理需求。实验结果表明,SFFormer 在雨滴与雨纹去除任务中取得了优于现有主流方法的 PSNR 与 SSIM;在真实雾霾去除任务中,PSNR 领先但 SSIM 略低于部分对比方法,表明模型对均匀雾霾退化的结构保真度有待进一步提升,视觉对比进一步验证了其在退化去除与细节保持方面的优越性。未来工作将探索该框架在视频修复、移动端实时部署等场景下的扩展,并结合可学习的退化结构先验提升模型的可解释性与适应性。

数据可用性声明

5 数据可用性声明

支撑本研究的科学数据与代码已在中国科学院科学数据银行(Science Data Bank)ScienceDB 平台公开发布,访问地址为 <https://www.doi.org/10.57760/sciencedb.j00240.00269> 或 <https://cstr.cn/31253.11.sciencedb.j00240.00269>。数据采用 CC BY-NC-ND 4.0 许可协议开放共享。

参考文献(References)

- Wang Z H, Chen J and Hoi S C H. 2021. Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43 (10) : 3365-3387 [DOI: 10.1109/TPAMI.2020.2982166]
- Zhang K, Li Y, Zuo W, Zhang L and Timofte R. 2020. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44 (10) : 6360-6376 [DOI: 10.1109/TPAMI.2021.3088914]
- Eigen D, Krishnan D and Fergus R. 2013. Restoring an image taken through a window covered with dirt or rain//*Proceedings of the IEEE International Conference on Computer Vision*. Sydney: IEEE: 633-640 [DOI: 10.1109/ICCV.2013.84]
- Quan Y H, Deng S, Chen Y X and Ji H. 2019. Deep learning for seeing through window with raindrops//*Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul: IEEE: 2463-2471 [DOI: 10.1109/ICCV.2019.00255]
- Wang Z D, Cun X D, Bao J M and Shi W Z. 2022. Uformer: A general u-shaped transformer for image restoration//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans: IEEE: 17683-17693 [DOI: 10.1109/CVPR52688.2022.01718]
- Zamir S W, Arora A, Khan S, Hayat M, Khan F S and Yang M H. 2022. Restormer: Efficient transformer for high-resolution image restoration//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans: IEEE: 5728-5739 [DOI: 10.1109/CVPR52688.2022.00564]
- He K M, Sun J and Tang X O. 2011. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33 (12) : 2341-2353 [DOI: 10.1109/TPAMI.2010.168]
- Qian R, Tan R T, Yang W H, Su J J and Liu J Y. 2018. Attentive generative adversarial network for raindrop removal from a single image//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE: 2482-2491 [DOI: 10.1109/CVPR.2018.00255]

- 10.1109/CVPR.2018.00263]
- Chen X, Li H R, Li M and Pan J S. 2023. Learning a sparse transformer network for effective image deraining//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE: 5896-5905 [DOI: 10.1109/CVPR52729.2023.00571]
- Xiao J, Fu X Y, Liu A D, Wu F and Zha Z J. 2022. Image de-raining transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11): 12978-12995 [DOI: 10.1109/TPAMI.2022.3183612]
- Lin X, Zhou Y X, Li D Z, Huang W and Sheng B. 2022. Image restoration using multi-scale feature joint attention model. *Journal of Computer-Aided Design & Computer Graphics*, 34(8): 1260-1271 (林晓,周云翔,李大志,黄伟,盛斌. 2022. 利用多尺度特征联合注意力模型的图像修复. *计算机辅助设计与图形学学报*, 34(8): 1260-1271) [DOI: 10.3724/SP.J.1089.2022.19172]
- Roser M and Geiger A. 2009. Video-based raindrop detection for improved image registration//Proceedings of the 12th IEEE International Conference on Computer Vision Workshops (ICCV Workshops). Kyoto: IEEE: 570-577 [DOI: 10.1109/ICCVW. 2009.5457650]
- Liu X H, Ma Y R, Shi Z H and Chen J. 2019. GridDehazeNet: Attention-based multi-scale network for image dehazing//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE: 7314-7323 [DOI: 10.1109/ICCV. 2019.00741]
- Zhang H and Patel V M. 2018. Density-aware single image de-raining using a multi-stream dense network//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 695-704 [DOI: 10.1109/CVPR.2018.00079]
- Shao Y J, Li L R H, Ren W Q, Gao C X and Sang N. 2020. Domain adaptation for image dehazing//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE: 2808-2817 [DOI: 10.1109/CVPR42600.2020.00288]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: Convolutional networks for biomedical image segmentation//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Cham: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8): 2011-2023 [DOI: 10.1109/TPAMI.2019.2913372]
- Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: Convolutional block attention module//Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]
- Loshchilov I and Hutter F. 2019. Decoupled weight decay regularization//Proceedings of the 7th International Conference on Learning Representations (ICLR). New Orleans: OpenReview. net [DOI: 10.48550/arXiv.1711.05101]
- Loshchilov I and Hutter F. 2017. SGDR: Stochastic gradient descent with warm restarts//Proceedings of the 5th International Conference on Learning Representations (ICLR). Toulon: OpenReview. net [DOI: 10.48550/arXiv.1608.03983]
- Zhang Y L, Li K, Li K, Wang L C, Zhong B N and Fu Y. 2018. Image super-resolution using very deep residual channel attention networks//Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer: 286-301 [DOI: 10.1007/978-3-030-01234-2_18]
- Yang F Z, Yang H, Fu J L, Lu H and Guo B N. 2020. Learning texture transformer network for image super-resolution//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE: 5791-5800 [DOI: 10.1109/CVPR42600.2020.00583]
- Ancuti C O, Ancuti C, Sbert M and Timofte R. 2019. Dense-Haze: A benchmark for image dehazing with dense-haze and haze-free images//Proceedings of the IEEE International Conference on Image Processing (ICIP). Taipei: IEEE: 1014-1018 [DOI: 10.1109/ICIP.2019.8803046]
- Chen H T, Wang Y H, Guo T Y, Xu C, Deng Y P, Liu Z L, et al. 2021. Pre-trained image processing transformer//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual: IEEE: 12299-12310 [DOI: 10.1109/CVPR46437. 2021.01212]
- Liang J Y, Cao J Z, Sun G L, Zhang K and Timofte R. 2021. SwinIR: Image restoration using Swin Transformer//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal: IEEE: 1833-1844 [DOI: 10.1109/ICCVW54120.2021.00210]
- Correia G M, Niculae V and Martins A F T. 2019. Adaptively sparse transformers//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics: 2174-2184 [DOI: 10.18653/v1/D19-1223]
- Ye T, Chen S H, Bai J H, Shi B W and Xue J H. 2023. Adverse weather removal with codebook priors//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE: 12619-12630 [DOI: 10.1109/ICCV51070.2023.01163]
- Zhou S H, Chen D D, Pan J S, Shi J L and Yang J H. 2024. Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE: 2952-2963 [DOI: 10.1109/CVPR52733.2024.00285]
- Oppenheim A V and Lim J S. 1981. The importance of phase in signals. *Proceedings of the IEEE*, 69(5): 529-541 [DOI: 10.1109/PROC. 1981.12022]
- Liu P J, Zhang H, Zhang K, Lin L and Zuo W M. 2018. Multi-level

- wavelet-CNN for image restoration//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City: IEEE: 773-782 [DOI: 10.1109/CVPRW.2018.00121]
- Fujieda S, Takayama K and Hachisuka T. 2017. Wavelet convolutional neural networks for texture classification[EB/OL]. [2026-06-05]. <https://arxiv.org/abs/1707.07394>
- Xu K, Qin M, Sun F, Wang Y, Chen Y K and Ren F. 2020. Learning in the frequency domain//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE: 1740-1749 [DOI: 10.1109/CVPR42600.2020.00181]
- Chen S H, Ye T, Bai J H, Chen E J, Shi J and Zhu L. 2023. Sparse sampling transformer with uncertainty-driven ranking for unified removal of raindrops and rain streaks//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris: IEEE: 13060-13071 [DOI: 10.1109/ICCV51070.2023.01205]
- Charbonnier P, Blanc-Féraud L, Aubert G and Barlaud M. 1994. Two deterministic half-quadratic regularization algorithms for computed imaging//Proceedings of the 1st IEEE International Conference on Image Processing. Austin: IEEE: 168-172 [DOI: 10.1109/ICIP.1994.413553]
- Rudin L I, Osher S and Fatemi E. 1992. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4): 259-268 [DOI: 10.1016/0167-2789(92)90242-F]
- Tang H Z, Xiong P Q, Wang W, Wang S Y and Chen L. 2024. Two-stage feature disentanglement network for single image rain removal. *Journal of Computer-Aided Design & Computer Graphics*, 36(2): 273-282 (汤红忠, 熊珮全, 王蔚, 王晒雅, 陈磊. 2024. 基于双阶段特征解耦网络的单幅图像去雨方法. *计算机辅助设计与图形学学报*, 36(2): 273-282) [DOI: 10.3724/SP.J.1089.2024.19828]
- Ren D W, Zuo W M, Hu Q H, Zhu P F and Meng D Y. 2019. Progressive image deraining networks: A better and simpler baseline//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 3937-3946 [DOI: 10.1109/CVPR.2019.00406]
- Zamir S W, Arora A, Khan S, Hayat M, Khan F S and Yang M H. 2021. Multi-stage progressive image restoration//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual: IEEE: 14821-14831 [DOI: 10.1109/CVPR46437.2021.01458]
- Shao M W, Li L, Meng D Y and Zuo W M. 2021. Uncertainty guided multi-scale attention network for raindrop removal from a single image. *IEEE Transactions on Image Processing*, 30: 4828-4839 [DOI: 10.1109/TIP.2021.3076283]
- Wang T Y, Yang X, Xu K, Chen S Z, Zhang Q and Lau R W H. 2019. Spatial attentive single-image deraining with a high quality real rain dataset//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 12270-12279 [DOI: 10.1109/CVPR.2019.01255]
- Wen B, Zhang Q H and Zhang J L. 2011. Real-time processing method of 2D-FFT/IFFT for high-resolution image and hardware implementation. *Application Research of Computers*, 28(11): 4376-4379 (温博, 张启衡, 张建林. 2011. 高分辨图像二维FFT正/反变换实时处理方法及硬件实现. *计算机应用研究*, 28(11): 4376-4379) [DOI: 10.3969/j.issn.1001-3695.2011.11.102]
- Quan R J, Yu X, Liang Y Z and Yang Y. 2021. Removing raindrops and rain streaks in one go//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual: IEEE: 9147-9156 [DOI: 10.1109/CVPR46437.2021.00903]
- Tu Z Z, Talebi H, Zhang H, Yang F, Milanfar P and Bovik A C. 2022. MAXIM: Multi-axis MLP for image processing//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE: 5759-5770 [DOI: 10.1109/CVPR52688.2022.00568]
- Gai S and Ni Y. 2025. Reduced biquaternion dual-branch deraining U-network via multi-attention mechanism. *IEEE Transactions on Image Processing*, 34: 6286-6301 [DOI: 10.1109/TIP.2025.3612841]
- Xu J, Hu X, Zhu L and Heng P A. 2025. Unifying physically-informed weather priors in a single model for image restoration across multiple adverse weather conditions. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(10): 9575-9591 [DOI: 10.1109/TCSVT.2025.3561470]
- Li B Y, Peng X L, Wang Z Y, Xu J Z and Feng D. 2017. AOD-Net: All-in-one dehazing network//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE: 4770-4778 [DOI: 10.1109/ICCV.2017.511]
- Qin X, Wang Z L, Bai Y C, Xie X D and Jia H Z. 2020. FFA-Net: Feature fusion attention network for single image dehazing//Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI Press, 34(7): 11908-11915 [DOI: 10.1609/aaai.v34i07.6865]
- Bai H, Pan J S, Xiang X G and Tang J H. 2022. Self-guided image dehazing using progressive feature fusion. *IEEE Transactions on Image Processing*, 31: 1217-1229 [DOI: 10.1109/TIP.2022.3140609]
- Guo C L, Yan Q X, Anwar S, Cong R M, Ren W Q and Li C Y. 2022. Image dehazing transformer with transmission-aware 3D position embedding//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE: 5812-5820 [DOI: 10.1109/CVPR52688.2022.00572]
- Qiu Y, Zhang K, Wang C, Luo W, Li H and Jin Z. 2023. MB-TaylorFormer: Multi-branch efficient transformer expanded by Taylor formula for image dehazing//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris: IEEE: 12756-12767 [DOI: 10.1109/ICCV51070.2023.01176]
- Cui Y N, Ren W Q, Cao X C and Knoll A. 2024. Revitalizing convolu-

- tional network for image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12): 9423-9438 [DOI: 10.1109/TPAMI.2024.3419007].
- Jin Z, Qiu Y, Zhang K, Li H and Luo W. 2025. MB-TaylorFormer V2: Improved multi-branch linear transformer expanded by Taylor formula for image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence [DOI: 10.1109/TPAMI.2025.3559891].
- Yi Q, Li J, Dai Q, Fang F, Zhang G and Zeng T. 2021. Structure-preserving deraining with residue channel prior guidance//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE: 4218-4227 [DOI: 10.1109/ICCV48922.2021.00420].
- Fu X Y, Qi Q, Zha Z J, Zhu Y and Ding X H. 2021. Rain streak removal via dual graph convolutional network//Proceedings of the AAAI Conference on Artificial Intelligence. Virtual: AAAI Press, 35(2): 1352-1360 [DOI: 10.1609/aaai.v35i2.16224].
- Fu X Y, Xiao J, Zhu Y R, Liu A P, Wu F and Zha Z J. 2023. Continual image deraining with hypergraph convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(8): 9534-9551 [DOI: 10.1109/TPAMI.2023.3241756].
- Guo Y, Xiao X Y, Chang Y, Deng S M and Yan L X. 2023. From sky to the ground: A large-scale benchmark and simple baseline towards real rain removal//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris: IEEE: 12097-12107 [DOI: 10.1109/ICCV51070.2023.01111].
- Hu M D, Wu Y, Song Y, Yang J B, Zhang R F, Wang H, et al. 2022. Joint evaluation and prospect of single image rain removal datasets

and deep learning algorithms. Journal of Image and Graphics, 27(5):1359-1391 (胡明娣, 吴怡, 宋尧, 杨静冰, 张瑞芳, 王红, 等. 2022. 单幅图像去雨数据集和深度学习算法的联合评估与展望. 中国图象图形学报, 27(5):1359-1391) [DOI: 10.11834/jig.210735].

Wang M H, Ke F H, Liang Y, Fan Z, Liao L. 2022. Image deraining network combining 3D attention and Transformer. Journal of Image and Graphics, 27(5):1509-1521 (王美华, 柯凡晖, 梁云, 范衡, 廖磊. 2022. 融合 3D 注意力和 Transformer 的图像去雨网络. 中国图象图形学报, 27(5):1509-1521) [DOI: 10.11834/jig.210473].

作者简介

林晓, 女, 教授, 主要研究方向为图像视频编辑处理、人工智能+教育、人工智能+医疗。E-mail: lin6008@shnu.edu.cn。

董大腾, 男, 硕士研究生, 主要研究方向为图像修复与增强中的多尺度空频联合建模及不确定性引导的稀疏注意力机制。E-mail: 1000569311@smail.shnu.edu.cn。

李岩, 男, 讲师, 主要研究方向为可信人工智能、大语言模型、计算机视觉、智能医疗、智能教育。E-mail: yanli@shnu.edu.cn。

洪璇, 女, 副教授, 主要研究方向为数字签名算法、密码学安全模型、区块链技术。E-mail: hong@shnu.edu.cn。

杨启哲, 通信作者, 男, 讲师, 主要研究方向为计算机视觉与机器学习、人工智能+医疗、人工智能+教育。E-mail: qzyang@shnu.edu.cn。