

频率引导的双稀疏自注意力单图像去雨算法

文渊博,高 涛,陈 婷,张千禧

(长安大学信息工程学院,陕西西安 710064)

摘 要: 现有基于自注意力网络 Transformer 的单图像去雨算法尽管在合成雨图上在取得良好效果,但却造成巨大的计算负担,且无法有效处理真实雨图. 对此,本文提出一种频率引导的双稀疏自注意力单图像去雨算法(Frequency-guided Dual Sparse self-Attention TransFormer, FDSATFormer). 首先,该算法利用空间稀疏因子和通道降维因子在提取准确全局信息的同时减少计算量,进而提出双稀疏自注意力特征学习网络(Dual Sparse self-attention Feature Learning, DSFL)以解决 Transformer 在高分辨率雨图上难以表征自注意力的问题. 其次,该算法通过探究图像去雨前后的频谱变化,提出频率引导的特征增强模块(Frequency-guided Feature Enhancer, FFE),其利用频域的全局信息指导特征编码阶段对空域特征的学习. 此外,现有去雨网络的编解码结构采用相近的设计,这导致网络的整体计算负担倍增. 对此,本文提出层级特征解码重建网络(Hierarchical Feature Decoding and Reconstructing network, HFDR),其利用无参的空间邻域移位操作(Spatial Neighborhood Shift, SNS)构建特征解码网络,在取得良好效果的同时进一步减少整体的计算负担. 实验表明,相比表现优秀的 Uformer 和 Restormer,本文算法所得结果的平均峰值信噪比分别提升 3.13 dB 和 0.12 dB,平均结构相似性分别提升 1.39% 和 1.07%.

关键词: 计算机视觉;图像去雨;自注意力网络;稀疏注意力;空间移位;频率引导学习

基金项目: 国家自然科学基金(No.52172379, No.62001058);长安大学中央高校基本科研业务费专项资金资助项目(No.300102242901)

中图分类号: TP391.4

文献标识码: A

文章编号: 0372-2112(2023)10-2812-09

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20221420

Frequency-guided Dual Sparse Self-Attention Algorithm for Single Image Deraining

WEN Yuan-bo, GAO Tao, CHEN Ting, ZHANG Qian-xi

(School of Information Engineering, Chang'an University, Xi'an, Shaanxi 710064, China)

Abstract: Existing Transformer-based algorithms for single image deraining achieve state-of-the-art performance but leading to reasonable computational loads while failing to effectively process real-world rainy images. To this end, we propose a frequency-guided dual sparse self-attention Transformer for single image deraining (FDSATFormer). Initially, our proposed method utilizes the spatial sparse factor and the channel reduction factor to extract accurate global information and significantly decreases the amount of computation. Furthermore, we present dual sparse self-attention feature learning network (DSFL) to deal with the problem that Transformer is difficult to represent self-attention on high-resolution feature maps. Meanwhile, by exploring the spectral changes of rainy image before and after removing rain streaks, we develop a frequency-guided feature enhancement module (FFE), which exploits the global information from the frequency domain to guide the accurate learning of spatial features in network encoders. In addition, the encoder and decoder of most existing methods follow the similar principles, resulting in almost double computational burden. To handle with this issue, we propose a hierarchical feature decoding and reconstructing network (HFDR), which uses non-parametric spatial neighborhood shift (SNS) to construct the feature decoding network and achieves fine results while further reducing the overall computing burden. Experimental results show that, our method improves the average peak signal noise ratio by 3.13 dB and 0.12 dB, and achieves performance gains of 1.39% and 1.07% in average structure similarity over the state-of-the-art Uformer and Restormer.

Key words: computer vision; image deraining; Transformer; sparse attention; spatial shift; frequency-guided learning

Foundation Item(s): National Natural Science Foundation of China (No.52172379, No.62001058); Fundamental Research Funds for the Central Universities (No.300102242901)

1 引言

图像去雨旨在去除雨图中的雨痕,进而提升后续计算机视觉算法的表现^[1-3].传统工作利用非局部均值滤波^[4]和引导滤波^[5,6]等方法去除雨痕,无法适应雨痕在密度,大小和分布等方面的多样性.在基于卷积神经网络(Convolution Neural Network, CNN)的单图像去雨算法中:Wei等^[7]同时利用有监督的合成雨图和无监督的自然雨图来训练网络,从而提升算法的泛化能力;Zhang等^[8]引入雨痕的密度信息进而提出一种多流密集单图像去雨网络;Yasarla等^[9]利用不确定性信息先检测雨痕再去雨;Li等^[10]引入压缩激励机制和空洞卷积;Ren等^[11]提出一种多阶段渐进的循环去雨网络,以极少的参数量取得良好的去雨性能;Jiang等^[12]利用雨痕的多尺度信息来提升算法的泛化能力;Zamir等^[13]采用多阶段网络将图像去雨分为多个子任务.然而,CNN的局部感受野特点缺乏对全局信息的关注,而权重共享机制无法自适应输入特征.

近来, Vision Transformer (ViT)^[14]利用多头自注意力(Multi-Head Self-Attention, MHSA)以某像素周围所有像素值的加权和作为其响应,从而有效地弥补了CNN的不足.然而, MHSA的计算量与输入特征的分辨率呈平方关系.对此,窗口自注意力网络Swin Transformer^[15]将MHSA的计算限制在不重叠的分割窗口内.基于Swin Transformer, Gao等人^[16]引入密集连接和残差学习来去除雨痕. Wang等人^[17]构建U形图像恢复网络,并将深度可分离卷积引入多层感知机(Multi-Layer Perceptron, MLP).然而,窗口自注意力违背MHSA捕获全局信息的初衷.不同的是, Zamir等人^[18]转置查询向量和键向量进而在通道维度表征MHSA.尽管该操作相比在空间维度表征MHSA能够降低计算负担,但却忽视了雨痕的稀疏性造成的特征内部空间冗余性,而这些冗余的信息会干扰雨图的特征学习和去雨图像的重建.

此外,雨痕主要存在于雨图的高频分量,去雨的过程中频谱变化是明显的,因而引入频率信息来指导网络的特征学习可以令网络学得更加准确的特征.在引入频域信息的去雨算法中, Fu等人^[2]首先将雨图分离为高频和低频两个分量,再将去雨后的高频分量与增强后的低频分量相加作为去雨图像.类似地,文献[19, 20]也是先分离高低频分量再分别去除雨痕.但是,这类算法的去雨表现严重依赖于频率分量分离方法的优劣.近来, Yang等人^[21]在网络中直接引入频率变换,从而学习雨图的频域特征.但是,该方法将频谱变换和频

域特征学习作为网络的核心组件,导致在后续阶段必须进行逆变换.同时,文献[22~24]引入离散小波变换(Discrete Wavelet Transform, DWT)来获得细化的频率特征.但是, DWT的结果实际并非频谱信息.另外,现有的方法在特征解码阶段采用与编码阶段近似的网络设计,这导致网络的整体计算负担翻倍.然而,文献[25~27]表明这一架构的近似设计不是必须的.

综上所述,现有方法存在计算量大、冗余信息多,未有效利用频域信息和特征解码网络复杂等不足.对此,本文提出一种频率引导的双稀疏自注意力单图像去雨算法(Frequency-guided Dual Sparse self-Attention TransFormer, FDSATFormer).该算法包括三部分:双稀疏自注意力特征学习网络(Dual Sparse self-attention Feature Learning, DSFL),频率引导的特征增强模块(Frequency-guided Feature Enhancer, FFE)和层级特征解码重建网络(Hierarchical Feature Decoding and Reconstructing network, HFDR).

2 频率引导的双稀疏自注意力算法

图1是本文算法的主要结构.图中, PatchEmbedding表示切分映射;双稀疏自注意力特征学习网络DSFL包括频率引导的特征增强模块FFE和双稀疏自注意力网络块(Dual Sparse Self-Attention Transformer Block, DSSA Transformer Block);DownSample为下采样操作;UpSample为上采样操作;层次特征解码重建网络HFDR由4个解码器Decoder块构成,其包括UpSample和空间邻域移位残差块(Spatial Neighborhood Shift Residual Block, SNSRB);SNSRB由空间邻域移位(Spatial Neighborhood Shift, SNS)操作和 1×1 卷积构成;Depthwise MLP是DSSA Transformer Block的前馈网络.输出层由输出头(Output Head, OH)和输出尾(Output Tail, OT)组成.

2.1 双稀疏自注意力特征学习网络

如图2所示,本文基于空间稀疏映射(Spatial Sparse Mapping, SSM)和通道降维变换(Channel Reducing Transform, CRT)提出双稀疏自注意力(Dual Sparse Self-Attention, DSSA).DSSA一方面显著将对计算负担,另一方面令网络获得更准确丰富的特征.ViT中对MHSA的表征可以公式化为

$$\text{MHSA} = \left(\mathbf{Q}_{l_0 \times c_0} \mathbf{K}_{l_0 \times c_0}^T / \sqrt{d_k} \right) \mathbf{V}_{l_0 \times c_0} \quad (1)$$

式(1)中, \mathbf{Q} , \mathbf{K} 和 \mathbf{V} 分别代表查询向量,键向量和值向量. \mathbf{T} 表示矩阵转置操作, $\sqrt{d_k}$ 是维度缩放因子, c_0 和 l_0 分别为输

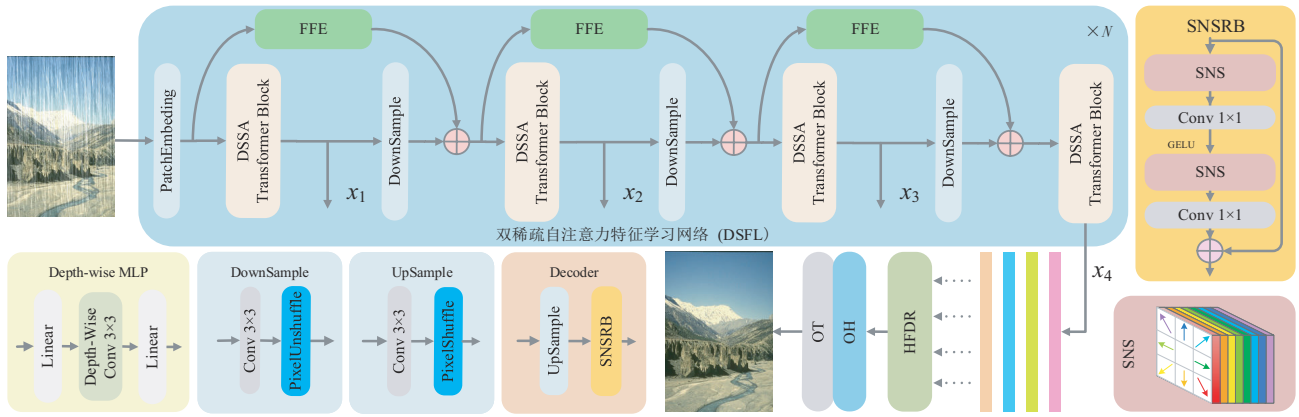


图1 频率引导的双稀疏自注意力单图像去雨算法的示意图

入特征的通道数和一维化后的向量长度. 本文根据MHSA的内积计算特点提出双稀疏自注意力DSSA,其包括稀疏

空间自注意力(Sparse Spatial Self Attention, SSSA)和稀疏通道自注意力(Sparse Channel Self Attention, SCSA).

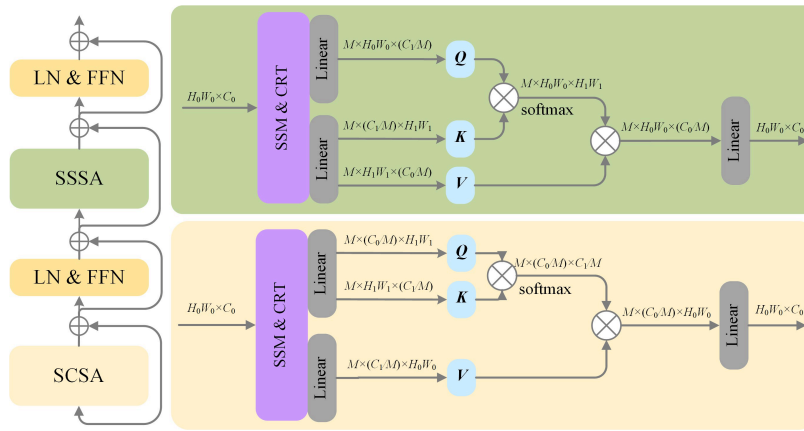


图2 双稀疏自注意力示意图

SSSA 在计算 K 和 V 前先对输入特征进行稀疏映射,进而经过全连接层 Linear 得到向量长度减小的键向量和值向量. 稀疏后的向量长度可以表示为

$$l_1 = l_0 \times r_s \quad (2)$$

式中, r_s 为空间稀疏因子. 降维后的通道数量可以表示为

$$c_1 = c_0 \times r_c \quad (3)$$

式中, r_c 为通道降维因子. SSSA 可以用公式表示为

$$SSSA = \left(Q_{l_0 \times c_1} K_{l_1 \times c_1}^T / \sqrt{d_k} \right) V_{l_1 \times c_0} \quad (4)$$

式中: l_1 代表经稀疏映射后的向量长度. 与 SSSA 不同的是, SCSA 在表征 Q 和 K 前先对输入特征进行稀疏映射,进而通过全连接层得到向量长度减小的查询向量和键向量. SCSA 可以用公式表达为

$$SCSA = \left(Q_{l_1 \times c_0}^T K_{l_1 \times c_1} / \sqrt{d_k} \right) V_{l_0 \times c_1} \quad (5)$$

稀疏映射的本质是通过一个核大小与步长一致的卷积实现,其所得特征不存在大于输入特征的冗余信息. 举例而言,对于一个结构为 $(l_0, c_0) = (32 \times 32, 64)$ 的

输入特征, DSSA 在表征自注意力时设置其空间稀疏因子 r_s 和通道降维因子 r_c 分别为 0.5 和 0.5, 则 SSSA 中查询向量, 键向量和值向量的结构分别为 $(32 \times 32, 32)$, $(32, 16 \times 16)$, $(16 \times 16, 64)$, 而 SCSA 中查询向量, 键向量和值向量的结构分别为 $(64, 16 \times 16)$, $(16 \times 16, 32)$, $(32, 32 \times 32)$. 本文提出的 DSSA 与不进行空间稀疏和通道降维的双自注意力(Dual Self-Attention, DSA)的计算量对比如表 1 所示, 其中, DSA 由空间自注意力(Spatial Self-Attention, SSA)和通道自注意力(Channel Self-Attention, CSA)构成. 可以发现, 相比双自注意力 DSA, 双稀疏自注意力 DSSA 的乘法计算量下降 80.56%, 加法计算量下降 78.89%.

表1 不同自注意力的计算量对比

自注意力		乘法计算量	加法计算量
DSA	SSA	1.43×10^8	1.41×10^8
	CSA		
DSSA	SSSA	2.78×10^7	2.74×10^7
	SCSA		

另外, DSSA 使用 SSM 和 CRT 前后计算量的变化情况如表 2 所示. 可以发现, 单独采用 SSM 和 CRT 相比原自注意力的计算量分别降低 72.87% 和 27.97%. 因此, SSM 和 CRT 均可以降低计算负担.

表 2 采用 SSM 和 CRT 前后自注意力的计算量对比

自注意力	SSM	CRT	乘法计算量
i	×	×	1.43×10^8
ii	√	×	3.88×10^7
iii	×	√	1.03×10^8
DSSA	√	√	2.78×10^7

2.2 频率引导的特征增强模块

在常用的频谱变换方法中, 离散傅里叶变换 (Discrete Fourier Transform, DFT) 的结果为复数, 但本文算法为实数网络, 因而无法对其进行无损计算. 尽管 DWT 可以得到不同频率水平的图像, 但其本质上并非频谱. 不同的是, 离散余弦变换 (Discrete Cosine Transform, DCT) 的输入输出均为实数. 基于 DCT, 本文提出频率引导的特征增强模块 FFE, 如图 3 所示. 其中, FSM 为频域稀疏映射 (Frequency-domain Sparse Mapping, FSM). FFE 位于双稀疏自注意力特征学习网络 DSFL 中, 其将前一层级特征进行频率变换, 再利用频域全局信息来指导空域特征学习.

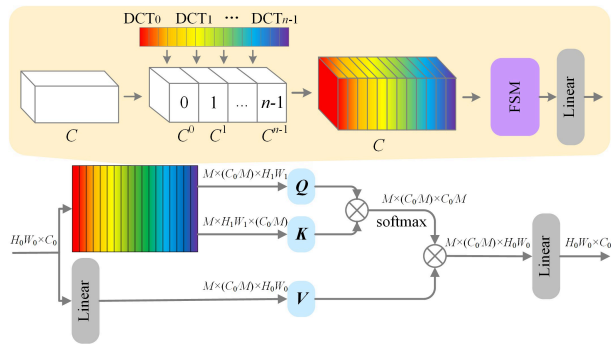


图 3 频率引导的特征增强模块的示意图

在目前引入频率信息进行图像恢复的算法中, 往往对频域信息进行处理后会再通过反变换回到空域, 这种设计给网络造成不必要的计算负担. 对此, 本文算法直接利用频域的全局信息来指导空域的特征学习. FFE 首先利用 DCT 得到频谱映射图, 2 维 DCT 的公式为

$$F(u, v) = c(u)c(v) \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} f(i, j)g(i, j; u, v) \quad (6)$$

式(6)中, H, W 为输入特征的空间分辨率, $f(i, j)$ 为空间上坐标为 i, j 的像素值. $g(i, j; u, v)$ 的数学形式为

$$g(i, j; u, v) = \cos\left[\frac{(i+0.5)\pi}{H}u\right] \cos\left[\frac{(j+0.5)\pi}{W}v\right] \quad (7)$$

系数 $c(u)$ 的数学形式为

$$c(u) = \begin{cases} 1/\sqrt{H}, & u=0 \\ 2/\sqrt{H}, & 1 \leq u \leq H-1 \end{cases} \quad (8)$$

式(8)中, $c(v)$ 的数学形式与 $c(u)$ 同理. 在得到频谱图之后, FFE 利用频域的稀疏映射 FSM 将分辨率降低, 进而通过全连接层得到查询向量和键向量. 式(6)表明, DCT 结果本身已包含一定的空间全局信息, 因此 FFE 利用稀疏频域自注意力 (Sparse Frequency-domain Self-Attention, SFSA) 在通道维度计算自注意力, 从而实现维度信息的互补, 即

$$\text{SFSA} = \left(\mathbf{Q}'_{l_1 \times c_0} \mathbf{K}'_{l_1 \times c_0} / \sqrt{d_k} \right) \mathbf{V}'_{l_0 \times c_0} \quad (9)$$

具体而言, FFE 通过转置频域的 \mathbf{Q}' 和 \mathbf{K}' 来表征通道自注意力, 再利用其对空域的值向量进行增强, 最终将结果嵌入到下一层级特征编码网络以引导其特征学习.

2.3 层级特征解码重建网络

基于文献[25]和文献[27], 本文认为同时在空间与通道两维度进行信息交互是解码器网络的基本功能. 本文提出空间邻域移位残差块 SNSRB, 其由无参的空间邻域移位操作 SNS 和 1×1 卷积构成. SNSRB 可以表示为

$$\begin{aligned} X' &= \text{GELU}(\text{Conv}_{1 \times 1}(\text{SNS}(X_{\text{in}}))) \\ X_{\text{out}} &= \text{Conv}_{1 \times 1}(\text{SNS}(X')) + X' \end{aligned} \quad (10)$$

式(10)中, $\text{GELU}(\cdot)$ 为激活函数, $\text{Conv}_{1 \times 1}$ 为核大小为 1×1 的卷积, X_{in} 和 X_{out} 分别为解码器的输入和输出. SNSRB 首先将输入特征分为 8 组, 每组分别向 8 邻域中的一个方向进行一个像素的移动, 移位后的特征通过一个 1×1 卷积进行特征融合. 在数字图像中, 邻域移位操作可以简单分为图 4 的三种情况. 本文同时对三种不同的移位操作进行实验发现选择 8 邻域移位操作时 FDSATFormer 取得最好的去雨表现, 因此本文实验均采用 8 邻域移位操作. 其次, 根据文献[28]的研究, 本文在残差结构中使用 GELU 激活函数取代 ReLU 激活函数来进一步提升算法的性能. 此外, 以往工作在去雨图输出部分仅采用一层卷积实现, 导致输出层所能表征的解空间有限. 本文将其分为输出头和输出尾两部分, 其中 OH 是基于未稀疏化的转置自注意力的特征优化网络, OT 是 3×3 卷积.

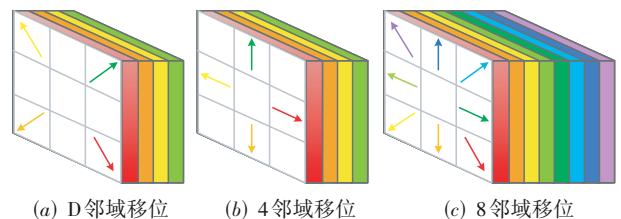


图 4 不同的邻域移位操作示意图

3 实验与分析

3.1 数据集设置

为验证本文算法在合成雨图上的表现,实验数据集选用来自 MPRNet^[13]的合成雨图数据集 Rain13K,其由 Rain100H^[29], Rain100L^[29], Rain800^[30], Rain12000^[8], Rain14000^[2]和 Rain12^[31]组成. Rain13K的训练集应包括 13 712 对雨图和无雨图,然而本文实际从数据集链接下载得到的训练集为 13 711 对图像. 经对比发现 MPRNet^[13]和 Restormer^[18]所采用的数据集均缺少第 12 300 对图像. 为保持对比实验的公平性,本文并未将其补充进训练集中,即实际训练集的图像对仍为 13 711. 同时,为验证本文算法在真实雨图上的泛化能力,本文选用来自文献[32]的 300 张真实雨图. 由于该数据集中部分图像的分辨率过大,在现有推理环境下无法测试对比方法 Uformer 和 Restormer,因此实验对其分辨率进行了一定调整,这与本文之前的工作^[16]存在区别.

3.2 实验环境和训练设置

实验在 NVIDIA Tesla A100 显卡上训练,在 NVIDIA GeForce RTX 3090 显卡上测试. DSFL 共包含 4 个 DSSA 和 3 个 FFE. 在 DSFL 特征编码器的四个阶段, MHSA 的注意力头数分别为 1、2、4、8,空间稀疏因子分别为 0.25、0.5、0.5、1,通道降维因子 1、0.5、0.5、0.25. 在 FFE 的三个阶段, DCT 的分块大小为 8、4、2. 在累计 500 次训练中,梯度优化算法 Adam^[33]的初始学习率为 2×10^{-4} ,其经过余弦退火学习率调整策略逐步衰减到 1×10^{-7} . 同时,本文采用混合率为 0.5 的 MixUp^[34]以及随机旋转,镜像变

换来增强训练数据. 训练时每批输入的图像对为 32,补丁大小为 256×256 . 损失函数为负结构相似性损失与平均绝对误差损失的和.

3.3 合成雨图实验对比

该实验将本文算法 FDSATFormer 与 10 个现有的单图像去雨算法在峰值信噪比 (Peak Signal Noise Ratio, PSNR) 和结构相似性 (Structure SIMilarity, SSIM) 上进行定量对比,对比算法分别为 DerainNet^[2], SEMI^[7], DIDMDN^[8], UMRL^[9], RESCAN^[10], PReNet^[11], MSPFN^[12], MPRNet^[13], Uformer^[17]和 Restormer^[18],结果如表 3 所示. 表中,加粗和加下划线的数据分别表示当前算法在对应测试集上的指标最好和排名第二. 实验表明,本文算法相比前 8 个算法在所有测试集上的平均表现最优,具体表现为平均 PSNR 提升 4.12%~51.60%,平均 SSIM 提升 2.61%~27.00%. 在 Test1200 测试集上, Uformer 的 SSIM 指标最高,本文算法指标排名第二. 在 Rain100L 和 Test2800 两个测试集上,本文的 SSIM 指标超越且 PSNR 指标接近目前表现优秀的 Restormer.

此外,本文还将 PReNet, MSPFN, MPRNet, Uformer, Restormer 五种单图像去雨算法与本文算法的去雨图像进行对比. 如图 5 所示,本文算法在雨痕分布密集的第 1 幅雨图上能彻底去除雨痕,而其他算法的处理结果中存在明显的雨痕. 在第 2 幅,第 3 幅,第 4 幅雨图上, PReNet 和 MSPFN 处理后的雨图中存在大量雨痕, MPRNet 和 Uformer 得到的去雨图像中仍旧残留少量雨痕,而本文算法与 Restormer 在完全去除雨痕的同时保持丰富的细节.

表 3 不同算法在合成雨图测试集上的定量对比

算法	Test100		Rain100H		Rain100L		Test2800		Test1200		平均	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DerainNet ^[2]	22.77	0.810	14.92	0.592	27.03	0.884	24.31	0.861	23.38	0.835	22.48	0.796
SEMI ^[7]	22.35	0.788	16.56	0.486	25.03	0.842	24.43	0.782	26.05	0.822	22.88	0.744
DIDMDN ^[8]	22.56	0.818	17.35	0.524	25.23	0.741	28.13	0.867	29.65	0.901	24.58	0.770
UMRL ^[9]	24.41	0.829	26.01	0.832	29.18	0.923	29.97	0.905	30.55	0.910	28.02	0.880
RESCAN ^[10]	25.00	0.835	26.36	0.786	29.80	0.881	31.29	0.904	30.51	0.882	28.59	0.857
PReNet ^[11]	24.81	0.851	26.77	0.858	32.44	0.950	31.75	0.916	31.36	0.911	29.42	0.897
MSPFN ^[12]	27.50	0.876	28.66	0.860	32.40	0.933	32.82	0.930	32.39	0.916	30.75	0.903
MPRNet ^[13]	30.27	0.897	30.41	0.890	36.40	0.965	33.64	0.938	32.91	0.916	32.73	0.921
Uformer ^[17]	29.05	0.921	27.01	0.880	35.91	0.970	30.91	<u>0.950</u>	31.87	0.938	30.95	0.932
Restormer ^[18]	<u>32.00</u>	<u>0.923</u>	<u>31.46</u>	<u>0.904</u>	38.99	<u>0.978</u>	34.18	0.944	<u>33.19</u>	0.926	<u>33.96</u>	<u>0.935</u>
本文算法	32.15	0.939	31.55	0.911	<u>38.80</u>	0.982	<u>33.14</u>	0.957	34.76	<u>0.934</u>	34.08	0.945

3.4 消融实验

为验证本文所提算法的有效性,实验对不同模块进行消融分析,如表 4 所示. 对于双稀疏自注意力特征学习网络 DSFL,实验分别对比移位窗口自注意力

(Shifted Window Self-Attention, SWSA)^[12]、稀疏空间自注意力 SSSA,稀疏通道自注意力 SCSA 和双稀疏自注意力 DSSA;对于特征增强模块 (Feature Enhancer, FE),实验分别对比空域特征增强模块 (Spatial Fea-

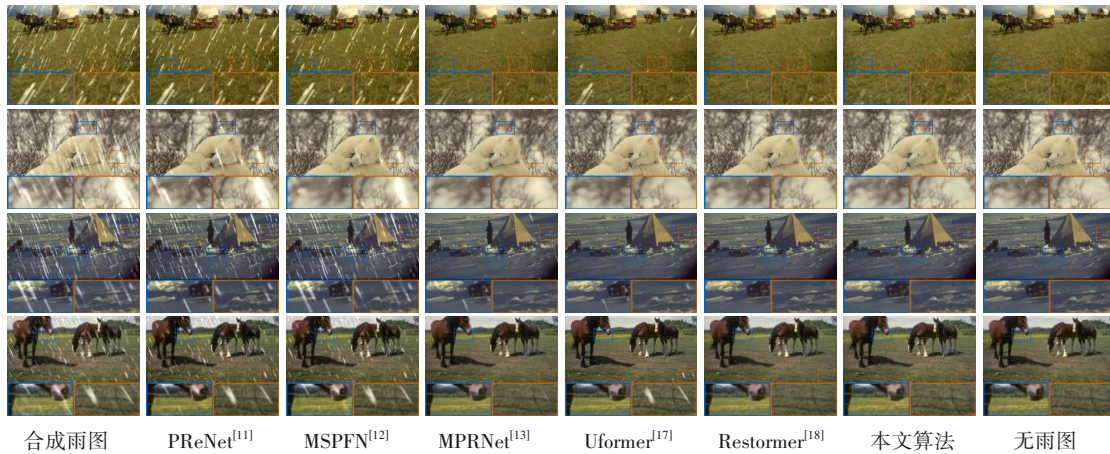


图5 不同算法在合成雨图上的视觉对比

ture Enhancer, SFE)和频域引导的特征增强模块 FFE; 对于特征解码器 Decoder,实验分别对比卷积层 Conv 和空间邻域移位残差块 SNSRB;对于输出层,实验分

别对比单独使用输出尾 OT 和结合输出头 OH. 本文参与对比的基线网络的解码器和编码器均为残差块.

表 4 不同网络组成在 Test100 测试集上的消融实验

基线网络		√	×	×	×	×	×	×	×	×
DSFL	SWSA	×	√	×	×	×	×	×	×	×
	SSSA	×	×	√	×	×	×	×	×	×
	SCSA	×	×	×	√	×	×	×	×	×
	DSSA	×	×	×	×	√	√	√	√	√
FE	SFE	×	×	×	×	×	√	×	×	×
	FFE	×	×	×	×	×	×	√	√	√
Decoder	Conv	×	√	√	√	√	√	√	×	×
	SNSRB	×	×	×	×	×	×	×	√	√
OL	OT	√	√	√	√	√	√	√	√	√
	OH	×	×	×	×	×	×	×	×	√
PSNR		23.41	25.17	27.33	25.91	27.58	28.13	29.43	<u>31.28</u>	32.15
SSIM		0.815	0.829	0.854	0.867	0.871	0.887	0.898	<u>0.924</u>	0.939

由表 4 可知,网络使用 SWSA 后相比基线网络在 PSNR 上提升 1.92 dB. SSSA 相比 SWSA 提升 2.16 dB. 同时,SCSA 相比 SSSA 在 PSNR 上下降 1.42 dB,但却在 SSIM 上提升 1.52%. DSSA 的表现超越 SSSA 和 SCSA. 在 FE 中,采用 FFE 相比 SFE 取得 1.30 dB 的提升. 在 Decoder 中,SNSRB 相比单独使用卷积提升 1.85 dB. 最后,在 OL 中引入 OH 相比单独使用 OT 提升 0.87 dB. 实验证明,本文提出网络模块均对整体的去雨表现有提升.

为进一步证明 FFE 的有效性,实验分别对比将 DFT 和 DWT 引入 FE 与本文提出的 FFE. 由表 5 可知,相比使用 DFT 和 DWT 引入频率信息的方法,FFE 最终的去雨效果分别提升 3.85 dB 和 2.69 dB. 因此,FFE 能有效地将频率信息引入去雨网络.

3.5 真实雨图实验对比

为验证本文算法在真实雨图上的表现,实验将

表 5 不同频率引导方法在 Test100 测试集上的去雨表现对比

频率利用方法	PSNR	SSIM
DFT	28.30	0.872
DWT	29.46	0.890
FFE	32.15	0.939

PReNet^[11], MSPFN^[12], MPRNet^[13], Uformer^[17], Restormer^[18]算法和本文算法进行对比,如表 6 所示. 真实去雨评价指标采用自然度图像质量评估器(Naturalness Image Quality Evaluator, NIQE)和空间-光谱熵质量(Spatial-Spectral Entropy-based Quality, SSQE),其值越小表示图像的质量越高.

由表 6 可知,本文算法的真实雨图去雨效果在 NIQE 上为 4.891,为所有方法中最小,优于排名第二的算法 0.08%. 在 SSEQ 指标上,MSPFN 所得结果为 21.50,其为所有算法

表6 不同算法在真实雨图测试集上的定量对比

算法	NIQE	SSEQ
PReNet ^[11]	4.930	27.28
MSPFN ^[12]	5.377	21.50
MPRNet ^[13]	5.030	27.35
Uformer ^[17]	4.985	27.81
Restormer ^[18]	<u>4.897</u>	26.67
本文算法	4.891	<u>26.35</u>

中最小. 本文算法在该指标上排名第二, 但均高于在合成雨图上表现优秀的 Uformer 和 Restormer. 不同算法的自然去雨图像的视觉效果如图6所示. 其中, 对于雨痕分布稀疏的第2幅和第4幅雨图, 本文算法有效去除雨痕, 而其他算法的结果中均残留雨痕. 在其余真实雨图上, 本文

算法同样去除大部分雨痕, 而其他算法几乎对其没有作用.

3.6 算法效率分析

为验证本文算法的效率, 实验将 PReNet^[11], MSPFN^[12], MPRNet^[13], Uformer^[17], Restormer^[18] 和本文算法的参数量, 计算量和推理用时三个指标进行对比, 其中推理时长为处理 100 张分辨率 256×256 的雨图所需时间的平均值. 由表7可知, 在计算量上本文算法小于 PReNet, 是 MPRNet 计算量的 29.1%, Uformer 计算量的 50.4%, Restormer 的 30.7%. 在推理时长上, PReNet 所需时间最少, MPRNet 次之. 尽管本文算法并非用时最少, 但相比目前表现优秀的 Uformer 和 Restormer 分别减少 21.5% 和 67.9%. 综上所述, 本文算法在图像去雨表现和推理用时上取得目前最好的平衡.

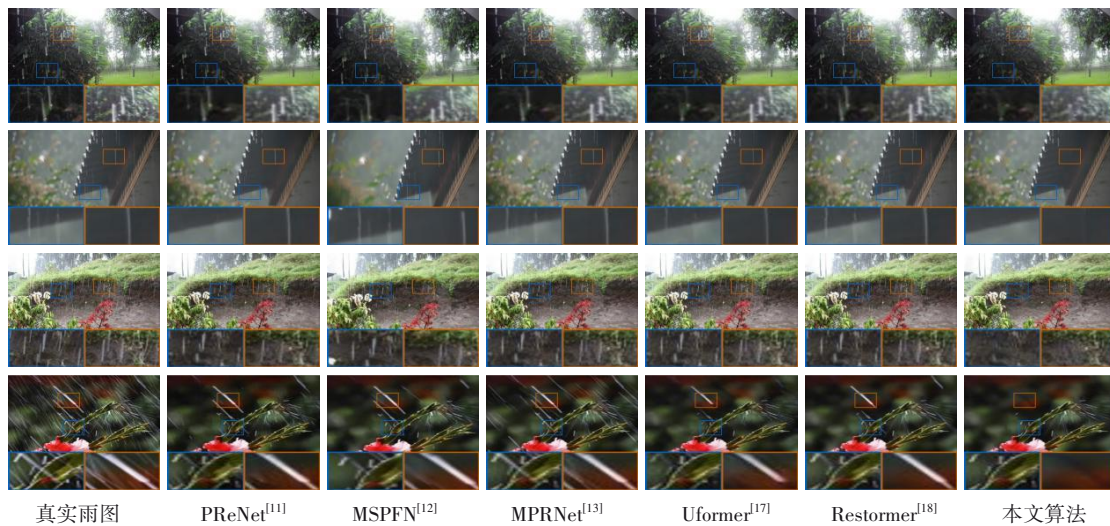


图6 不同算法在真实雨图上的视觉对比

表7 不同算法处理效率的定量分析

算法	参数量/M	计算量/G	推理用时/s
PReNet ^[11]	0.169	66.250	0.044
MSPFN ^[12]	15.820	604.700	0.073
MPRNet ^[13]	<u>3.637</u>	148.600	<u>0.072</u>
Uformer ^[17]	50.390	<u>85.770</u>	0.121
Restormer ^[18]	26.130	141.000	0.296
本文算法	52.820	43.220	0.095

4 结论

为解决现有单图像去雨算法的不足, 本文提出一种频率引导的双稀疏自注意力算法, 其利用双稀疏自注意力, 频率引导学习策略和高效的层级特征解码重建网络在合成雨图和真实雨图上的去雨表现均优于当下表现优秀的算法, 同时网络的计算量和推理用时也显著降低, 因此其有望成为单图像去雨研究新的范式.

参考文献

- [1] 盖杉, 王俊生. 基于深度学习的非局部注意力增强网络图像去雨算法研究[J]. 电子学报, 2020, 48(10): 1899-1908.
GAI S, WANG J S. Image raindrop algorithm research using nonlocal attention enhanced network based on deep learning[J]. Acta Electronica Sinica, 2020, 48(10): 1899-1908. (in Chinese)
- [2] FU X Y, HUANG J B, DING X H, et al. Clearing the skies: A deep network architecture for single-image rain removal[J]. IEEE Transactions on Image Processing, 2017, 26(6): 2944-2956.
- [3] 秦菁, 文渊博, 高涛, 等. 面向多天气退化图像恢复的自注意力扩散模型[J/OL]. 上海交通大学学报, 2023. <https://doi.org/10.16183/j.cnki.jsjtu.2023.043>.
QIN J, WEN Y B, GAO T, et al. Transformer-based diffusion

- model for all-in-one weather-degraded image restoration[J/OL]. Journal of Shanghai Jiaotong University, 2023. <https://doi.org/10.16183/j.cnki.jsjtu.2023.043>. (in Chinese)
- [4] KIM J H, LEE C, SIM J Y, et al. Single-image deraining using an adaptive nonlocal means filter[C]//2013 IEEE International Conference on Image Processing. Piscataway: IEEE, 2014: 914-917.
- [5] XU J, ZHAO W, LIU P, et al. Removing rain and snow in a single image using guided filter[C]//2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE). Piscataway: IEEE, 2012: 304-307.
- [6] ZHENG X H, LIAO Y H, GUO W, et al. Single-image-based rain and snow removal using multi-guided filter[C]//Neural Information Processing: 20th International Conference, ICONIP 2013. Cham: Springer, 2013: 258-265.
- [7] WEI W, MENG D Y, ZHAO Q A, et al. Semi-supervised transfer learning for image rain removal[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 3872-3881.
- [8] ZHANG H, PATEL V M. Density-aware single image deraining using a multi-stream dense network[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2018: 695-704.
- [9] YASARLA R, PATEL V M. Uncertainty guided multi-scale residual learning-using a cycle spinning CNN for single image de-raining[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 8397-8406.
- [10] LI X A, WU J L, LIN Z C, et al. Recurrent squeeze-and-excitation context aggregation net for single image deraining[C]//Computer Vision – ECCV 2018. Cham: Springer International Publishing, 2018: 262-277.
- [11] REN D, ZUO W, HU Q, et al. Progressive image deraining networks: A better and simpler baseline[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 3932-3941.
- [12] JIANG K, WANG Z Y, YI P, et al. Multi-scale progressive fusion network for single image deraining[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 8343-8352.
- [13] ZAMIR S W, ARORA A, KHAN S, et al. Multi-stage progressive image restoration[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 14816-14826.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 1-11.
- [15] LIU Z, LIN Y T, CAO Y E, et al. Swin Transformer: Hierarchical vision Transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2022: 9992-10002.
- [16] 高涛, 文渊博, 陈婷, 等. 基于窗口自注意力网络的单图像去雨算法[J]. 上海交通大学学报, 2023, 57(5): 613-623.
GAO T, WEN Y B, CHEN T, et al. A single image deraining algorithm based on swin Transformer[J]. Journal of Shanghai Jiao Tong University, 2023, 57(5): 613-623. (in Chinese)
- [17] WANG Z D, CUN X D, BAO J M, et al. Uformer: A general U-shaped Transformer for image restoration[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 17662-17672.
- [18] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: Efficient Transformer for high-resolution image restoration [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 5718-5729.
- [19] ZHANG Y, XIANG Y J, CAI L, et al. Single image deraining with high-low frequency guidance[C]//ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2022: 2330-2334.
- [20] 柳长源, 王琪, 毕晓君. 基于多通道多尺度卷积神经网络的单幅图像去雨方法[J]. 电子与信息学报, 2020, 42(9): 2285-2292.
LIU C Y, WANG Q, BI X J. Research on rain removal method for single image based on multi-channel and multi-scale CNN[J]. Journal of Electronics & Information Technology, 2020, 42(9): 2285-2292. (in Chinese)
- [21] YANG H, ZHOU D M, CAO J D, et al. DpNet: Detail-preserving image deraining via learning frequency domain knowledge[J]. Digital Signal Processing, 2022, 130: 103740.
- [22] HUANG H B, YU A J, CHAI Z H, et al. Selective wavelet attention learning for single image deraining[J]. International Journal of Computer Vision, 2021, 129(4): 1282-1300.
- [23] YANG H H, YANG C H H, WANG Y C F. Wavelet channel attention module with a fusion network for single

- image deraining[C]//2020 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2020: 883-887.
- [24] YANG W H, LIU J Y, YANG S A, et al. Scale-free single image deraining via visibility-enhanced recurrent wavelet learning[J]. IEEE Transactions on Image Processing, 2019, 28(6): 2948-2961.
- [25] WU X A, HUANG T Z, DENG L J, et al. A decoder-free Transformer-like architecture for high-efficiency single image deraining[C]//Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2022: 1474-1480.
- [26] CHEN P X, ZHANG M D, SHEN Y H, et al. Efficient decoder-free object detection with Transformers[C]//Lecture Notes in Computer Science. Cham: Springer, 2022: 70-86.
- [27] JOSE VALANARASU J M, YASARLA R, PATEL V M. TransWeather: Transformer-based restoration of images degraded by adverse weather conditions[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 2343-2353.
- [28] LIU Z, MAO H Z, WU C Y, et al. A ConvNet for the 2020s[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 11966-11976.
- [29] YANG W H, TAN R T, FENG J S, et al. Deep joint rain detection and removal from a single image[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 1685-1694.
- [30] ZHANG H, SINDAGI V, PATEL V M. Image de-raining using a conditional generative adversarial network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(11): 3943-3956.
- [31] LI Y, TAN R T, GUO X J, et al. Rain streak removal using layer priors[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 2736-2744.
- [32] FU X Y, LIANG B R, HUANG Y, et al. Lightweight pyramid networks for image deraining[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(6): 1794-1807.
- [33] KINGMA D P, BA J. Adam: A method for stochastic optimization[EB/OL]. (2017-01-30) [2023-09-13]. <https://arxiv.org/abs/1412.6980>.
- [34] ZHANG H, CISSE M, DAUPHIN Y N, et al. MixUp: Beyond empirical risk minimization[EB/OL]. (2018-04-27) [2022-12-19]. <https://arxiv.org/abs/1710.09412>.

作者简介



文渊博 男, 1999年2月出生于陕西省咸阳市. 现为长安大学在读博士研究生, 主要从事计算机视觉和智能交通系统方面的研究工作.

E-mail: wyb@chd.edu.cn



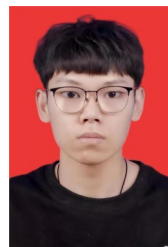
高涛(通讯作者) 男, 1980年9月出生于陕西省富平县. 2010年博士毕业于西北工业大学电子信息学院. 现为长安大学教授, 从事人工智能和图像处理方面的研究工作.

E-mail: tgaochd@126.com



陈婷 女, 1982年4月出生于陕西省西安市. 2011年博士毕业于西安电子科技大学通信工程学院. 现为长安大学信息工程学院教授, 主要从事模式识别和智能交通系统方面的研究工作.

E-mail: tchenchd@126.com



张千禧 男, 2000年5月6日出生于河南省濮阳市. 现为长安大学在读硕士研究生, 主要从事计算机视觉方面的研究工作.

E-mail: zhangqx1216@gmail.com