

FE-DARFormer:基于频域增强与退化感知路由Transformer的图像去雪模型

秦溢, 战鹏祥, 鲜峰, 柳晨龙, 王明辉

引用本文

秦溢, 战鹏祥, 鲜峰, 柳晨龙, 王明辉. FE-DARFormer:基于频域增强与退化感知路由Transformer的图像去雪模型[J]. 计算机科学, 2025, 52(11): 113-122.

QIN Yi, ZHAN Pengxiang, XIAN Feng, LIU Chenlong, WANG Minghui. FE-DARFormer:Image Desnowing Model Based on Frequency Enhancement and Degradation-aware Routing Transformer [J]. Computer Science, 2025, 52(11): 113-122.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[DCDAD:考虑上下文依赖差异化的时间序列异常检测模型](#)

DCDAD:Differentiated Context Dependency for Time Series Anomaly Detection Method

计算机科学, 2025, 52(6): 106-117. <https://doi.org/10.11896/jsjcx.240600001>

[多尺度下幅度谱与相位谱相融合的视觉注意建模](#)

Visual Attention Modeling Based on Multi-scale Fusion of Amplitude Spectrum and Phase Spectrum

计算机科学, 2017, 44(7): 293-298. <https://doi.org/10.11896/j.issn.1002-137X.2017.07.053>

[融入频域信息聚焦特征的显著性目标检测算法](#)

Saliency Object Detection Algorithm Integrating Focusness Feature of Frequency Domain Information

计算机科学, 2018, 45(10): 261-266. <https://doi.org/10.11896/j.issn.1002-137X.2018.10.048>

FE-DARFormer: 基于频域增强与退化感知路由 Transformer 的图像去雪模型

秦溢 战鹏祥 鲜峰 柳晨龙 王明辉

四川大学计算机学院 成都 610065

(2022223045231@stu.scu.edu.cn)

摘要 图像去雪的目标是从包含复杂雪景退化的图像中恢复清晰的场景信息。与雨的规律性和半透明性不同,雪具有各种退化形态和尺度,严重退化的区域会严重遮挡场景信息。近年来,许多方法通过自注意力机制来恢复不同的退化现象。然而,对图像所有区域进行全局自注意力计算成本较高。为了降低计算成本,这些方法通常将自注意力计算限制在有限的窗口内。但是由于严重退化区域的遮挡效应,这些退化区域的恢复只能依赖捕捉周围区域的信息,对图像进行恢复时严重退化区域受到感受野瓶颈的限制,难以聚合更多信息。因此,这些方法难以有效恢复大面积退化的区域。为了进一步提升去雪性能,提出了一种新颖有效的去雪方法。从退化感知路由与频域增强的角度出发,提出了退化感知路由 Transformer 和双频域增强 Transformer,并将两者结合,提出了新的网络架构——FE-DARFormer。FE-DARFormer 能够针对严重退化区域进行动态路由和全局自注意力计算,从而获得全局感受野,有效恢复大面积退化区域,并降低计算成本。此外,该方法能通过离散小波分解出高低频信息,从而有效恢复多尺度的雪景退化并识别多样化的雪花形态与纹理特征。

关键词: 图像去雪;退化感知路由;双频域增强;全局感受野

中图分类号 TP391

FE-DARFormer: Image Desnowing Model Based on Frequency Enhancement and Degradation-aware Routing Transformer

QIN Yi,ZHAN Pengxiang,XIAN Feng,LIU Chenlong and WANG Minghui

College of Computer Science,Sichuan University,Chengdu 610065,China

Abstract The goal of image desnowing is to restore clear scene information from images degraded by complex snowy scenes. Unlike the regularity and semi-transparency of rain, snow exhibits various forms and scales of degradation, with severely degraded regions often obstructing important scene details. Recent methods have employed self-attention mechanisms to address different degradation phenomena. However, global self-attention computation across all image regions is computationally expensive, leading these methods to restrict attention to smaller windows. Yet, due to the occlusion effects in severely degraded areas, the recovery of these regions relies heavily on capturing information from surrounding areas, which results in a receptive field bottleneck, limiting the ability to aggregate sufficient information. As a result, these methods struggle to effectively restore large-scale degraded regions. To improve desnowing performance, this paper proposes a novel approach, introducing a new network architecture called FE-DARFormer, which combines a Degradation-Aware Routing Transformer and a Dual-Frequency Enhancement Transformer. FE-DARFormer dynamically routes and applies global self-attention to severely degraded regions, enabling a global receptive field for effective restoration of large degraded areas while reducing computational cost. Additionally, it uses discrete wavelet decomposition to handle multi-scale snow degradation, enhancing the recovery of diverse snowflake shapes and textures.

Keywords Image desnowing, Degradation-aware routing, Dual-frequency enhancement, Global receptive field

1 引言

图像去雪旨在去除雪天拍摄图像中的各种退化现象,并恢复清晰的图像。单图像的雪花去除在许多下游任务,如目标检测^[1]、目标跟踪^[2]和视频监控^[3-4]等相关视觉应用中,对

于获取高质量图像信息至关重要。然而与其他恶劣天气场景相比,被雪花遮挡的场景具有更加复杂多样的退化特征。在雪花退化场景中,雪花通常呈现出不同的形状、大小、反射度以及透明度,使得对雪花进行去除和恢复干净图像具有挑战性。与以往的研究^[5]一样,本文将雪景的图像退化模型表示为:

到稿日期:2024-12-24 返修日期:2025-03-09

基金项目:国家重点研发计划(2022YFC2407604)

This work was supported by the National Key Research and Development Program of China(2022YFC2407604).

通信作者:王明辉(wangminghui@scu.edu.cn)

$$I(x) = K(x)t(x) + A(1-t(x)) \quad (1)$$

其中, $t(x)$ 和 A 分别表示透射图和大气光; 雪退化的图像表示为 $I(x)$; $K(x)$ 可写为 $K(x) = J(x)(1 - Z(x)R(x)) + C(x)Z(x)R(x)$, $J(x)$ 表示无雪退化的干净图像, $R(x)$ 表示雪的位置信息, 通过提供一个二值掩码来定义, $C(x)$ 和 $Z(x)$ 分别表示色差图像和雪掩码。

可以看出, 雪的退化模型是一个病态问题, 需要同时关注图像中雪的透明度、形状和反射特性。为了解决雪花引起的退化问题, 已有研究者提出了多种方法^[6-8], 主要集中于使用图像的先验特征和通过信号分离的方法进行算法设计, 取得了一定成效。然而, 由于依赖手工先验, 这些方法难以准确区分雪花与场景, 尤其是雪花形态多样且部分纹理与场景相似时, 可能导致去雪不彻底或细节丢失。

近年来, 深度学习神经网络发展迅速, 并广泛应用于各种视觉任务中^[9-11]。针对图像去雪问题, 已有许多基于深度学习的方法^[12-13]被提出, 且表现出了良好的性能。Liu 等^[12]提出了首个大规模合成雪景数据集 Snow-100K。为了应对各种复杂的雪退化问题, 许多架构设计被提出。HDCW-Net^[14]以分层结构来分别处理不同频域的特征; JSTASR^[15]采用多尺度卷积操作来应对多样的雪退化场景。此外, Restormer^[15]提出了通道级自注意力机制, 通过在特征通道层面进行全局自注意力计算来增强网络的非局部能力, 以解决复杂的雪退化问题。

然而, 目前图像去雪领域仍存在两个主要问题。

1) 雪退化引起严重遮挡效应。特别是在大面积退化区域, 场景信息难以重建。恢复被遮挡的信息需依赖周围区域, 因此捕捉更多周围信息有助于图像重建。而要恢复图像中被完全遮挡部分的信息, 只能捕捉遮挡区域周围信息进行重建。考虑到在雪退化的场景图像中, 大尺度近景雪花占比相对较少, 动态分配计算资源进行全局自注意力计算是一个理想的方案。ViT^[16]首次将自注意力算法引入计算机视觉任务, 通过在输入图像的所有嵌入令牌(Token)上执行自注意力计算来建立全局依赖。然而, 这种方法计算成本过高。而基于窗口的方法^[17-18]虽降低了成本, 却限制了感受野, 影响了严重退化区域的信息聚合, 削弱了自注意力的全局特性。Focal-Net^[19]虽然考虑了雪景图像中退化程度不均匀的问题, 并引入了双域选择的机制针对难以恢复的部分进行增强, 但该方法仍然没有有效地扩大感受野。因此, 本文认为动态地针对这些严重的退化区域进行全局自注意力计算的算法, 更适合图像去雪任务。

2) 不同尺度和形态的雪退化难以消除。值得注意的是, 不同尺度的雪退化会对图像场景产生不同的影响: 大尺度的雪花往往退化严重, 所带来的遮挡效应主要导致大面积的图像信息损失, 从而导致难以被有效重建; 小尺度雪花则会造成与其纹理相似的场景物体被一起去除和残留。Desnow-Net^[12]采用多尺度的设计来模拟雪花的多样性, 但是面对形态多样化的雪花仍然缺乏针对纹理细节的有效识别。Restormer^[15]通过在通道之间进行自注意力计算的方式来捕捉长距离的依赖关系, 从而实现多种退化的有效去除。然而, 通道之间进行自注意力运算难以关注空间域上的信息变化。

另外, 自注意力通过归一化打分的方式来聚合信息, 难以关注到高频细节纹理信息。鉴于此, 设计出一种既能有效识别和恢复图像高频细节纹理信息, 又能有效重建图像低频信息的方法, 是一个理想的解决方案。

本文提出频域增强路由 Transformer 模块(Frequency Enhanced Routing Transformer Block, FERTB)以解决上述问题。该模块包含退化感知路由 Transformer (Degradation-Aware Routing Transformer, DART) 和双频域增强 Transformer (Dual-Frequency Enhancement Transformer, DFET)。DART 包含一个退化感知路由注意力 (Degradation-Aware Routing Attention, DARA) 和 MF2N^[20], DART 负责识别并路由严重退化区域至窗口区域, 并对该区域执行全局注意力计算以及对其余区域进行卷积以捕捉局部细节, 最终融合两部分特征。以往的自注意力算法^[16-18]中, ViT 直接对全图像进行自注意力运算, 其由于计算成本高昂, 因此难以运用到图像生成类的密集型预测任务中。Swin Transformer 降低了计算复杂度, 但是感受野受到窗口大小的瓶颈限制。而雪花的成像受观察视角的影响, 呈现出以下规律: 近景中雪花稀疏、尺度较大, 主要分布在图像的低频分量; 远景中雪花则稠密、尺度较小, 主要分布在图像的高频分量。因此, 受雪花成像特性以及随机分布特点的启发, 本文方法的 DART 首次提出对卷积滤波后的退化感知特征进行窗口划分与投射路由, 实现感受野最大化与全局建模, 有效提升重建效果并降低计算成本。其机制对比如图 1 所示。

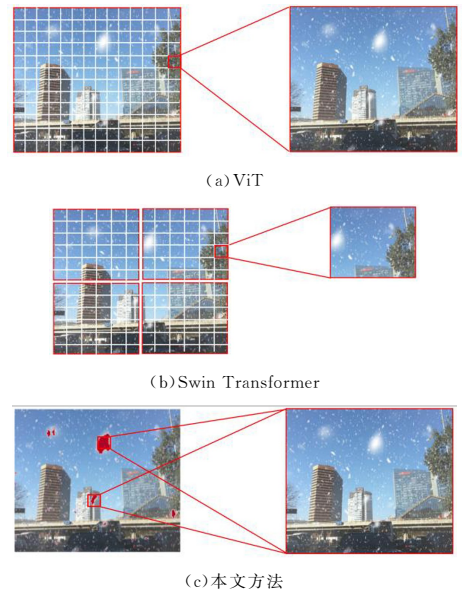


图 1 本文方法与现有方法机制对比的示意图

Fig. 1 Schematic comparison of the proposed method and existing methods

虽然 DART 能够有效重建退化严重区域, 但由于路由信息总量固定, 难以覆盖所有退化区域。对此, 本文提出 DFET, 以有效消除不同尺度雪花的退化效应并捕捉细节纹理, 缓解细节丢失的问题。具体而言, DFET 首先利用离散小波变换^[21] (Discrete Wavelet Transform, DWT) 对输入特征进行分解, 获得低频分量, 然后采用基于窗口的自注意力机制^[17-18]对低频分量进行重建, 扩大感受野并减少计算量。

同时,对输入特征应用中心差分卷积^[22] (Central Difference Convolution, CDC)捕获丰富的纹理细节特征,得到高频增强特征,之后通过 DWT 分解获得高频分量。最后,使用离散小波逆变换(IDWT)融合重建后的低频分量与高频分量。由此,DFET 充分结合了 CDC 与自注意力机制在不同频域的优势,能够有效去除多种退化现象。

本文的主要贡献总结如下:

1)提出了 DART,利用退化感知路由模块(Degradation-Aware Routing Module, DARM),通过融合预训练掩码生成器生成的雪退化掩码图与输入特征融合生成引导图,并通过图像窗口划分投射路由的机制,对与严重退化区域相关的部分进行路由,然后经过自注意力机制对这些路由部分进行全局聚合。与现有流行的基于窗口的自注意力机制^[17-18]相比,本文首次提出利用图像窗口划分投射路由的机制,针对受严重退化影响的图像区域进行真正的全局建模,最大化地扩大了对严重退化区域的感受野和提高了全局信息聚合的能力,同时保持高效的计算性能。值得注意的是,本文方法理论上还可以迁移到其他图像恢复类领域中。

2)提出了 DFET,结合中心差分卷积(CDC)与自注意力机制,利用不同频域的优势应对多尺度雪退化问题。考虑到图像的主要信息集中于低频分量,本文将计算复杂度较高的自注意力机制应用于低频成分,以扩大感受野并降低计算开销。同时,利用轻量化的差分卷积增强高频细节,并结合小波变换的先验约束,使 CDC 与自注意力机制分别聚焦于不同频域,有效弥补了自注意力机制天然的平滑性导致的在高频信息处理上的不足。此外,多尺度设计提升了对不同尺度雪退化的适应性。相比传统基于窗口的自注意力机制或单纯卷积方法,本文方法在恢复质量上具备显著优势,并可泛化至其他图像恢复任务。

3)在多个合成和真实场景数据集上进行实验,验证了该模型的有效性。定性和定量的结果验证了本文模型具有显著效果,并在多个数据集达到了当前先进的水平。

2 相关工作

2.1 单图像去雪

2.1.1 基于物理的图像去雪方法

传统的雪图像去除方法^[6-8,23]主要基于雪特征的物理先验知识来处理退化问题。这些方法通常通过对雪的几何特征、光学特性,以及其与场景背景的关系进行建模来实现雪的检测与去除。例如,Bossu 等^[23]提出了一种基于混合高斯模型(MoG)的方法,将前景和背景分离以检测和去除雪粒子,同时结合方向梯度直方图(HOG)特征来恢复清晰的图像。此外,Pei 等^[6]则利用了雨滴和雪花的物理先验知识,包括饱和度和亮度和可见性等属性,构建了一种用于检测和去除雪雨的模型。该方法通过分析图像中雪粒子与背景的差异,能够有效地区分透明或半透明的雪粒子,从而减少对背景细节的干扰。虽然这些传统先验方法在一定程度上能够应对雪图像的退化问题,但由于依赖物理先验知识,它们在处理复杂场景时存在显著的局限性,难以适应多样化的雪粒子分布和光照条件变化。

2.1.2 基于深度学习的图像去雪方法

深度学习在许多视觉任务中取得了显著的进展,并展现出卓越的性能^[24-31]。在图像去雪问题上,Liu 等^[12]提出了 DesnowNet,通过结合透明度处理模块和残差生成模块,有效恢复了清晰的无雪图像,展现了优异的效果。此外,JSTASR^[5]结合尺寸感知和透明度感知的方法,成功应对了多种复杂的雪景挑战,大幅提升了去雪算法的性能。

针对雪景的复杂性,Chen 等^[14]开发了 HDCW-Net,利用双树复小波变换对雪景图像进行分解,更加注重处理雪花在场景中的复杂分布特性,但是该方法仅在图像高频分量进行了多尺度特征提取,忽略了图像主要成分集中于低频分量,缺乏对低频成分的多尺度特征关注。此外,该方法缺乏有效的局部与非局部处理的方法,导致难以有效应对雪花的复杂退化问题。与此同时,许多无监督方法也被提出^[31-32],这些方法通过结合生成对抗网络(GAN)框架以及域自适应学习的方式,实现了针对真实雪景图像的高效去雪。此外,DDMS-Net^[33]深入探索了退化图像中的语义和深度信息,将这些特性融入去雪过程。ConvIR^[34]通过引入自适应卷积机制,让卷积核根据输入的图像内容进行调整,同时通过重新加权高通滤波器的权重,关注了图像中具有信息量的频率成分,从而更好地优化了图像修复性能。这些基于深度学习的方法显著推动了去雪技术的发展,进一步提升了复杂场景下的算法性能,但它们仍然缺乏对不同尺度雪花在图像频域中所造成不同影响的关注,难以有效地从不同频域中捕获重建信息。

2.2 面向图像恢复的视觉 Transformer

源于语言模型的视觉 Transformer^[16] (Vision Transformer, ViT)在众多视觉任务中展现出卓越的性能,并成功应用于多个领域^[35-36]。然而,由于 Transformer 中的自注意力计算复杂度随着图像分辨率的增长呈现二次方的增加,导致其难以应用到低层视觉任务中。Swin^[17]通过将自注意力计算限制在局部区域内,有效降低了计算复杂度,在生成式视觉任务中被广泛应用。例如,Uformer^[36]将 Transformer 模块融入 U 型结构用于图像恢复,并引入了一种轻量化的可学习的多尺度恢复调节器,但是该方法仍然将自注意力计算限制在很小的窗口内,在应对严重退化的大尺度雪花时难以捕捉更多周围信息以更好地重建场景信息;Restormer 通过引入通道级自注意力机制来考虑通道信息,从而实现全局特征提取,突破了局部区域的限制。然而,通道级自注意力机制的方法忽视了特征中不同空间域之间的信息差异,而这些分布于空间域中的信息对于从复杂多样的雪退化中恢复高质量图像信息至关重要。CPLFormer^[37]通过跨尺度原型学习的 Transformer 架构从不同尺度的特征中提取信息,但是该方法仍然难以实现针对退化区域的全局上下文学习。

2.3 差分卷积

差分卷积的起源可追溯至局部二值模式方法通过编码局部像素差异来实现纹理分类。随着卷积神经网络的发展,Xu 等^[38]提出了局部二值卷积(Local Binary Convolution, LBC),该方法结合非线性激活和卷积层对像素差异进行处理。随后,Yu 等^[22]提出了中心差分卷积(Central Difference Convolution, CDC),通过可学习的权重编码像素差异。鉴于差分卷

积能够有效捕捉图像纹理细节信息, 本文将其引入图像去雪中任务中, 用于重建高频细节纹理特征, 以更好地消除不同形态的小尺度雪花带来的影响。

3 本文方法

本文提出的频域增强与退化感知路由 Transformer 的图像去雪网络模型 (Frequency Enhancement and Degradation-Aware Routing Transformer, FE-DARFormer) 的整体框架如图 2 所示。为了更好地消除多尺度的雪花退化, 其设计遵循流行的多层级的编码器-解码器结构风格。除了第一层级采用串联 DFET 的结构外, 其余层级均统一使用多个 FERTB 模块串联, 以学习鲁棒的去雪特征。对于输入的雪退化图像

$I \in \mathbb{R}^{3 \times H \times W}$, FE-DARFormer 首先使用一个预训练好的普通 UNet^[39] 生成雪花掩码 $M_1 \in \mathbb{R}^{3 \times H \times W}$, 再对输入图像进行一个卷积操作生成特征嵌入 $F_0 \in \mathbb{R}^{C \times H \times W}$, 其中 H 和 W 分别表示图像的高度和宽度, C 表示通道数。接着, 浅层特征 F_0 被送入一个多层级的编码器-解码器组成的网络, 其中每一个层级采用跳跃连接^[19] 融合编码和解码特征。从第二层级开始, 本文对雪退化掩码图像 M_1 使用平均池化下采样的操作生成尺度上与当前层级特征图对齐的掩码图 $M_i \in \mathbb{R}^{3 \times H_i \times W_i}$ (i 表示当前层级), 随后将其送入 DART 的 DARM 中来路由受退化影响的相关区域。在解码器的最后通过一个卷积操作对输出特征进行精调, 并生成残差图像 $R \in \mathbb{R}^{3 \times H \times W}$ 。最后, 将该残差图像叠加到退化图像上, 得到复原后的图像 $J = I + R$ 。

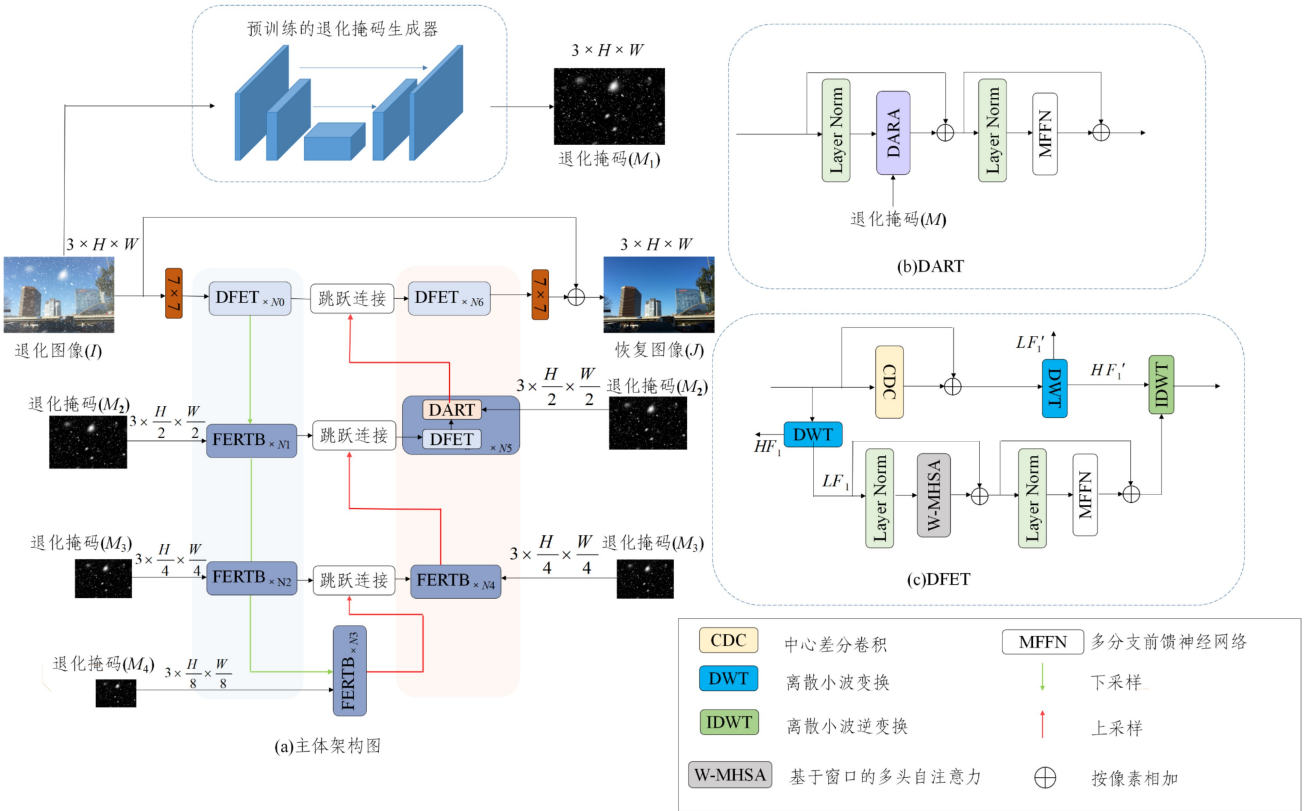


图 2 FE-DARFormer 整体框架

Fig. 2 Overall framework of FE-DARFormer

3.1 退化感知路由 Transformer

在传统的 Transformer 架构^[16] 中, 计算成本随输入分辨率二次增长, 限制了自注意力在低层视觉任务中的应用。Swin 通过局部自注意力降低计算复杂度, 但雪退化的非均匀性和严重遮挡效应要求更大感受野以聚合更多周围信息, 使其不适用于图像去雪任务。如图 2 所示, 本文提出的 DART 由 DARA 和 MFN 组成。其中 DARA 如图 3(a) 所示, 通过路由的设计仅对部分受严重退化影响的关键区域进行全局建模。相对于 ViT 对每个像素点进行全局计算的方式, DARA 显著降低了计算复杂度, 同时能保证退化区域的恢复效果。其输出可以表示为:

$$F' = \text{DARA}(\text{LN}(F_{\text{in}})) + F_{\text{in}} \quad (2)$$

$$F_{\text{out}} = \text{MFN}(\text{LN}(F')) + F' \quad (3)$$

如图 3(b) 所示, 本文在 DARA 中引入了退化感知路由

模块 (DARM), 用于将输入特征中受严重退化影响的相关区域信息路由到特定的窗口区域, 并生成映射权重。对于第 i 层级的输入特征 $F_{\text{in}} \in \mathbb{R}^{C_i \times H_i \times W_i}$, 首先通过一个线性层生成查询特征 $F_Q \in \mathbb{R}^{C_i \times H_i \times W_i}$, 其计算式为 $F_Q = W_p^Q F_{\text{in}}$, 其中 $W_p^{(\cdot)}$ 表示逐点卷积。为了动态路由退化区域, 一个路由权重生成网络首先通过线性层 W_p^M 将输入的退化掩码图 $M \in \mathbb{R}^{3 \times H_i \times W_i}$ 映射为 $W_p \in \mathbb{R}^{C_i \times H_i \times W_i}$, 然后与当前特征 $F_Q \in \mathbb{R}^{C_i \times H_i \times W_i}$ 拼接融合, 再通过由卷积组成的路由引导图生成器生成路由引导图 $F \in \mathbb{R}^{1 \times H_i \times W_i}$ 。接下来, 将路由引导图划分为多个不重叠的窗口区域 $P \in \mathbb{R}^{1 \times H_p \times W_p \times N}$, 其中 N 表示窗口的数量, H_p 和 W_p 分别为每个窗口的高度和宽度。本文通过 $\text{argmax}(\cdot)$ 函数确定每个窗口内像素的路由索引, 其计算式如下:

$$M_{u,v} = \text{argmax}(P_{u,v}^0, P_{u,v}^1, \dots, P_{u,v}^{N-1}) \quad (4)$$

其中, $M \in \mathbb{R}^{C \times H_p \times W_p}$ 表示路由索引图, 其值对应每个窗口中

最大值的索引; (u, v) 表示映射图中的像素位置, $P_{u,v}^i$ 表示第 i 个窗口 (u, v) 位置处的像素值。通过这种方式,可以为每个像素点分配一个路由索引,从而精确定位和聚焦退化相关的关键区域。接着,将 F_Q 划分为不重叠的窗口,记为 $\mathbf{X}^j \in \mathbb{R}^{C_i \times H_p \times W_p}$, 其中 j 表示具体的窗口索引。在窗口维度上对关键区域的路由过程可以表示为:

$$\hat{\mathbf{X}}_{u,v}^j = \mathbf{X}_{u,v}^{M_{u,v}^j}, M_{u,v}^j \in [0, N-1] \quad (5)$$

其中, $\hat{\mathbf{X}} \in \mathbb{R}^{C_i \times H_p \times W_p}$ 表示从 F_Q 路由出的与退化相关的路由区域。然而, $\text{argmax}(\cdot)$ 操作会导致梯度在这一部分被截断,无法进行反向传播。因此,本文重新设计了该函数,以确保有效的梯度流动。如图 3 所示,本文用 $\text{softmax}(\cdot)$ 替代 $\text{argmax}(\cdot)$ 来生成路由权重图。由此,路由过程中路由区域 $\hat{\mathbf{X}} \in \mathbb{R}^{C_i \times H_p \times W_p}$ 在 (u, v) 位置的计算式可以表示为:

$$\hat{M}_{u,v}^j = \frac{e^{P_{u,v}^j}}{\sum_{m=0}^{N-1} e^{P_{u,v}^m}}, \hat{\mathbf{X}}_{u,v} = \hat{\mathbf{M}}_{u,v}^T \mathbf{X}_{u,v} \quad (6)$$

其中, $\hat{\mathbf{M}} \in \mathbb{R}^{C_i \times H_p \times W_p \times N}$ 表示路由权重图。通过这种方法, $M_{u,v}$ 被近似为 $\hat{\mathbf{M}}_{u,v}$, 其中 $\hat{\mathbf{M}}_{u,v}$ 表示为 $[\hat{M}_{u,v}^0, \dots, \hat{M}_{u,v}^{N-1}]$ 。通过上述设计,图像特征中的退化关键区域被路由到一个路由窗口区域 $\hat{\mathbf{X}} \in \mathbb{R}^{C_i \times H_p \times W_p}$ 中。为了加速 $\hat{\mathbf{M}}_{u,v}$ 收敛到近似 one-hot 的形式,本文引入了平均绝对偏差 (Mean Absolute Deviation, MAD) 作为正则化项,其定义如下:

$$\text{MAD}(\hat{\mathbf{M}}) = -\sum_i |\hat{\mathbf{M}} - E_{\text{last}}(\hat{\mathbf{M}})| \quad (7)$$

其中, E_{last} 表示对 $\hat{\mathbf{M}}$ 的最后一个维度进行平均操作。通过这种正则化,模型能够更快地调整路由权重,从而精确定位退化相关的关键区域并实现更有效的特征路由。

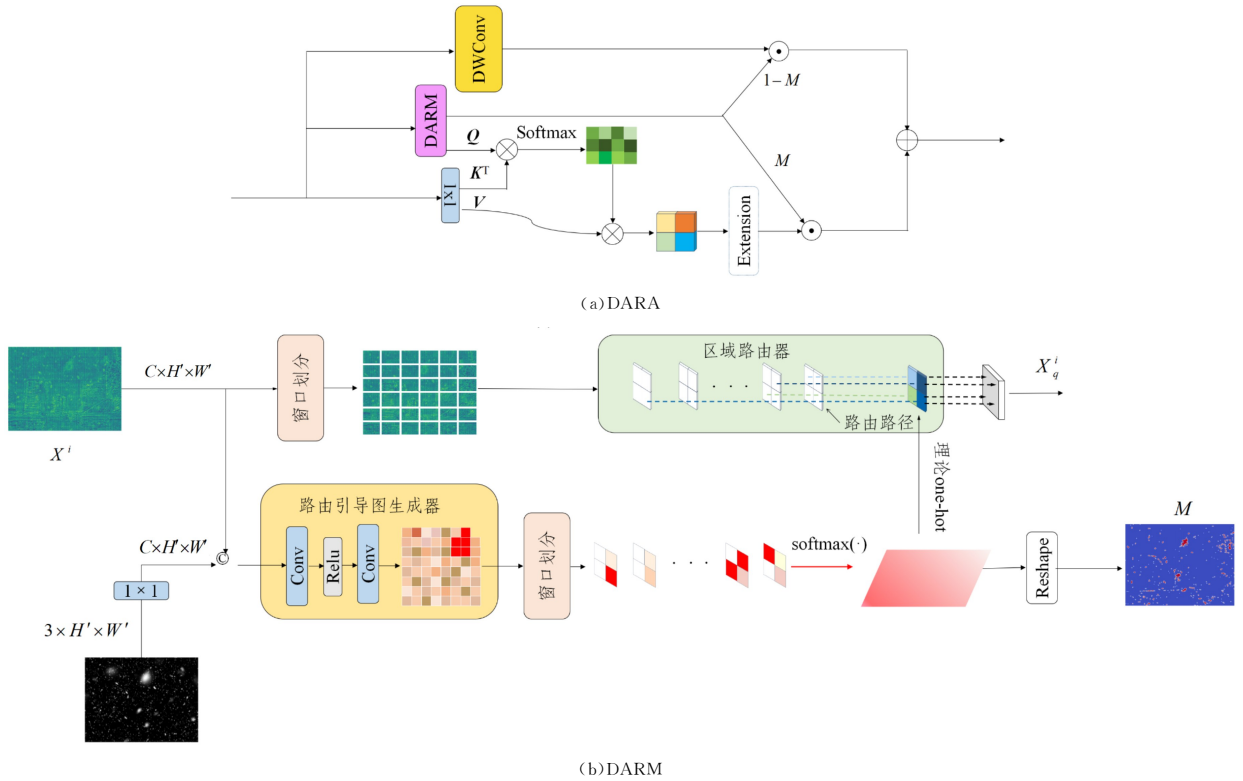


图 3 退化感知路由注意力示意图

Fig. 3 Schematic diagram of degradation-aware routing attention

如前文所述,通过区域路由器,将通常包含严重退化的关键区域路由到窗口 $\hat{\mathbf{X}} \in \mathbb{R}^{C_i \times H_p \times W_p}$ 。为了有效地为窗口内的路由区域进行全局依赖建立,同时保持计算效率,本文聚焦退化相关的区域。具体而言,本文首先将窗口内的信息作为查询特征,然后利用线性层从输入特征中生成键 $\mathbf{F}_K \in \mathbb{R}^{N_i \times C_i}$ 和值 $\mathbf{F}_V \in \mathbb{R}^{N_i \times C_i}$ ($N_i = H_i \times W_i$), 随后针对退化相关区域进行全局建模,可表示为:

$$\hat{\mathbf{X}}_{\text{out}} = \text{Softmax}\left(\frac{\hat{\mathbf{X}} \mathbf{F}_K^T}{\sqrt{C_i}}\right) \mathbf{F}_V \quad (8)$$

其中, C_i 表示维度的大小。为了有效融合经过全局建模后的路由区域 $\hat{\mathbf{X}}_{\text{out}} \in \mathbb{R}^{C_i \times H_p \times W_p}$ 与输入特征,首先以堆叠的方式扩展 $\hat{\mathbf{X}}_{\text{out}}$, 并重塑形状为 $\mathbf{X}_{\text{out}} \in \mathbb{R}^{C_i \times H_i \times W_i}$ 。接着,将之前获得的

路由权重图 $\hat{\mathbf{M}}$ 重塑形状为 $\mathbf{W} \in \mathbb{R}^{C_i \times H_i \times W_i}$, 用于特征融合。上述操作虽然对输入特征中部分与退化相关的区域进行了全局依赖的计算,但为了保证图像的恢复效果和计算效率,本文对其余非路由区域采取深度可分离卷积的操作。其处理过程定义如下:

$$\mathbf{F}' = \mathbf{W} \odot \mathbf{X}_{\text{out}} + (1 - \mathbf{W}) \odot \text{DWConv}(\mathbf{F}_{\text{in}}) \quad (9)$$

其中, \odot 表示逐元素乘法, $\text{DWConv}(\cdot)$ 表示深度可分离卷积。为了进一步增强特征表达能力,采用 MFFN^[20] 作为前馈网络。总体而言, DART 通过将更多的计算资源用于退化相关的关键区域进行全局聚合,既能在恢复退化区域图像信息时获得全局感受野,又保证了计算效率。

3.2 双频域增强 Transformer

雪退化图像的信息受大面积退化的遮挡和小尺度雪花对

高频细节的影响,导致难以恢复图像细节和纹理丢失。大尺度雪花主要引起低频信息损失。文献[15]通过特征通道建立全局性处理多样化退化,但难以关注空间域的信息变化和低频细节。自注意力计算中 $\text{softmax}(\cdot)$ 打分的设计使得其对图像低频成分的重建更为敏感,同时图像信息的主要成分集中在低频部分。因此,本文引入离散小波变换(DWT)分解图像低频信息,并基于窗口的自注意力机制进行重建,扩大感受野并降低计算开销。同时,针对图像高频信号处理,引入中心差分卷积(CDC)以融合差分卷积算子捕捉纹理信息,并利用DWT滤除低频分量。具体而言,DWT分解输入特征提取低频分量,鉴于自注意力的动态加权和非局部建模机制被视为低通滤波[40]的有效工具,同时在理论上自注意力机制的归一化操作使其对低频信息更加敏感,本文应用基于窗口的自注意力方法[18]重建低频分量,采用CDC增强高频特征,最终通过DWT逆变换融合两者。该方法缩小了低频分量的空间分辨率,使非局部建模在更大感受野下进行,从而更有效地去除不同尺度的雪退化问题。

如图2所示,输入特征为 $\mathbf{X} \in \mathbb{R}^{C_i \times H_i \times W_i}$, 首先将其分解为高频子带和低频子带。该过程可表示为:

$$(\mathbf{LF}_1, \mathbf{HF}_1^1, \mathbf{HF}_1^2, \mathbf{HF}_1^3) = \text{DWT}(\mathbf{X}) \quad (10)$$

其中, $\text{DWT}(\cdot)$ 表示离散小波变换操作, \mathbf{LF}_i 表示第 i 层的低频子带, \mathbf{HF}_i^k 表示第 i 层的第 k 个高频子带。随后,在低频子带 $\mathbf{LF}_1 \in \mathbb{R}^{C_i \times \frac{H_i}{2} \times \frac{W_i}{2}}$ 上通过 W-MHSA[18] 进行自注意力计算,其过程可表示为:

$$\mathbf{LF}_1' = \mathbf{LF}_1 + \text{W-MHSA}(\text{LN}(\mathbf{LF}_1)) \quad (11)$$

$$\mathbf{LF}_1'' = \mathbf{LF}_1' + \text{MFFN}(\text{LN}(\mathbf{LF}_1')) \quad (12)$$

其中, LN 表示层归一化(LayerNorm)。W-MHSA的具体形式如下:

$$\text{Softmax}\left(\frac{\mathbf{QK}^T}{\sqrt{C}} + p\right)\mathbf{V} + \text{DWConv}(\mathbf{V}) \quad (13)$$

其中, \mathbf{Q}, \mathbf{K} 和 \mathbf{V} 是从 \mathbf{LF}_1 线性投影得到的特征,被划分为窗口; p 表示对应的位置编码。

在另一分支中,本文首先引入CDC从输入特征 $\mathbf{X} \in \mathbb{R}^{C_i \times H_i \times W_i}$ 中提取经过高频增强的特征 $\mathbf{X}' \in \mathbb{R}^{C_i \times H_i \times W_i}$, 以提升纹理细节的处理能力; 然后利用DWT提取 \mathbf{X}' 的高频分量 $\mathbf{HF}_1' \in \mathbb{R}^{C_i \times \frac{H_i}{2} \times \frac{W_i}{2}}$; 最后使用IDWT将高频分量 \mathbf{HF}_1' 与通过自注意力重建的低频子带 $\mathbf{LF}_1'' \in \mathbb{R}^{C_i \times \frac{H_i}{2} \times \frac{W_i}{2}}$ 结合。该过程可表示为:

$$\mathbf{X}' = \text{CDC}(\mathbf{X}) + \mathbf{X} \quad (14)$$

$$(\mathbf{LF}_1', \mathbf{HF}_1^1, \mathbf{HF}_1^2, \mathbf{HF}_1^3) = \text{DWT}(\mathbf{X}') \quad (15)$$

$$\mathbf{X}_{\text{out}} = \text{IDWT}(\mathbf{LF}_1', \mathbf{HF}_1^1, \mathbf{HF}_1^2, \mathbf{HF}_1^3) \quad (16)$$

其中, $\text{CDC}(\cdot)$ 表示中心差分卷积操作, $\text{IDWT}(\cdot)$ 表示离散小波逆变换。

3.3 损失函数

FE-DARFormer通过以下3种损失函数进行优化:

1) \mathcal{L}_1 损失函数,即平均绝对误差,用于最小化恢复图像 J 与对应的真实值 GT 之间的距离。其定义为:

$$\mathcal{L}_1 = \|J - GT\|_1 \quad (17)$$

2) 聚焦频域损失函数 \mathcal{L}_{FFL} (Focal Frequency Loss)[41], 用

于通过自适应加权动态聚焦不同频率分量。由于雪退化图像的退化程度和尺度是非均匀的,在图像频域上表现为不同频率分量学习难易程度不同,因此,使用FFL损失函数可以动态地为图像中更难学习的频率分量分配更大的权重,从而优化网络性能,提升图像恢复质量。其定义为:

$$\omega(u, v) = |F_J(u, v) - F_{GT}(u, v)|^\alpha \quad (18)$$

$$\mathcal{L}_{\text{FFL}} = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \omega(u, v) |F_J(u, v) - F_{GT}(u, v)|^2 \quad (19)$$

其中, $F_J(u, v)$ 和 $F_{GT}(u, v)$ 表示分别对恢复图像 J 与真实图像 GT 进行傅里叶变换之后得到的频谱图像; α 是缩放因子,本文采用默认值1。

3) $\mathcal{L}_{\text{MADR}}$, 即平均绝对偏差正则化,用于约束路由权重掩码的形式接近 one-hot 形式。其定义为:

$$\mathcal{L}_{\text{MADR}} = \sum_i \text{MAD}(M_i) \quad (20)$$

其中, M_i 表示第 i 个 FERTB 中预测的路由权重掩码。因此,本文设计的网络总损失函数定义如下。

$$\mathcal{L} = \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_{\text{FFL}} + \lambda_3 \mathcal{L}_{\text{MADR}} \quad (21)$$

其中, λ_1, λ_2 和 λ_3 在实验中分别按经验设定为 1, 0.2 和 0.1。

4 实验

4.1 数据集

本文在合成数据集和真实拍摄数据集上验证了所提方法的有效性。数据集具体描述如下:

综合雪数据集(Comprehensive Snow Dataset, CSD)[14]: 一个涵盖了雪线、雪花等各种形态的综合合成数据集,一共有8000张训练集和2000张测试集图片。

Snow100K[12]: 包含根据雪的特征合成的大约10万张成对数据集以及真实场景中的1300多张雪退化图像。

真实场景中的雪去除(Snow Removal in Realistic Scenario, SRRS)[5]: 通过基于当前流行的雾基准数据集 RESIDE[13]来合成具有似雾遮蔽效应的图像,然后用Photoshop合成雪景图像并标注了雪的透明度、大小和位置。

从每个数据集中选择2000张图像进行评估。需要注意的是, Snow100K数据集被划分为3个子集(即S, M和L)。因此,本文分别从S, M和L子集中选择了1200张、400张和400张图像进行评估。同时,为了评估方法在真实场景中的泛化能力,使用了Snow100K中的真实雪样本,并通过视觉主观对比的方法与现有先进的方法进行了对比。

4.2 实验设置

本文提出的FE-DARFormer基于PyTorch深度学习平台实现。在整体网络设计中,本文采用了多尺度金字塔的架构,各层的特征维度分别设为 $\{32, 64, 128, 256, 128, 64, 32\}$, 各阶段的模块数量设置为 $[2, 2, 3, 3, 3, 2, 2]$, DART的路由窗口大小是根据输入特征的尺度进行动态调节的,输入特征大小 $H \times W$ 设置为 $\frac{H}{8} \times \frac{W}{8}$ 。掩码生成器采用UNet对雪掩码进行预训练。优化器采用Adam[42], 其参数 $\beta_1, \beta_2, \epsilon$ 均设置为默认值,即0.9, 0.999和 1×10^{-8} 。此外,初始学习率和批量大小分别设为 2×10^{-4} 和 8。学习率调整策略采用余弦退火策略,将学习率从初始值调整至 2×10^{-6} 。在数据增强

方面,本文随机裁剪大小为 256×256 的图像块,并对其进行固定角度旋转(即 $0^\circ, 90^\circ, 180^\circ$ 或 270°),以及垂直或水平翻转。在整个训练阶段,模型共训练 1 200 个周期。在 Snow100K 数据集上,本文随机选择了 20 000 张图片进行训练。

4.3 评价指标

与前人工作保持一致,本文采用了计算机视觉领域中常用的峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)和结构相似性指数(Structural Similarity Index Measurement, SSIM)评估去雪性能。其中,PSNR 能够反映恢复图像与原始图像之间像素值的差异;SSIM 能够体现恢复图像与原始图像的亮度、对比度和结构信息之间的相似性。

4.4 对比基线

为对比不同模型的性能,本文采用的基线方法包括:面向图像恢复任务的方法 FocalNet^[19], ConvIR^[34], Restormer^[15], NAFNet^[43], Uformer^[36], DGUNet^[44];面向恶劣天气任务的方法 TKL^[45], TransWeather^[11];面向图像去雪任务的方法 JSTASR^[5], DesnowNet^[12], HDCW-Net^[14], DDMSNet^[33], SMGARN^[46], CPLFormer^[37]。

4.5 实验结果与分析

4.5.1 定量评估

将 FE-DARFormer 与多个现有的先进方法在上述 3 个去雪基准数据集上进行了对比实验。定量评估结果如表 1 所列,粗体表示最优结果,下划线表示次优结果。在 CSD 数据集上,本文方法的 PSNR 指标比最优方法 ConvIR-B 低 0.33 dB。主要原因在于 CSD 数据集由真实场景结合人工合成的雪和雾构成,而本文方法主要侧重于雪花部分的恢复。然而,图片中的雾区受大气散射影响,导致光线衰减和对比度下降,从而限制了本文方法在该数据集上的表现。因此,未来将考虑引入针对雪雾的建模方法,以提升对复杂天气条件下图像的恢复效果。在 Snow100K 数据集上,本文方法的 PSNR 较 ConvIR-B 高出 0.88 dB。此外,本文方法的 FLOPs 仅为 ConvIR-B 的约 30%,复杂度显著降低。在其他数据集上,本文方法在 PSNR 与 SSIM 指标上全面超越了其他方法,主要原因在于:首先本文 DFET 针对雪花在图像不同频域的特性,分别利用全局和局部处理的方法对不同频域进行更为有效的重建;其次,本文 DART 可以使得退化严重区域获得更全局的信息以达到更好的恢复效果。

表 1 FE-DARFormer 与几种最先进方法的定量对比实验结果

Table 1 Quantitative comparisons between FE-DARFormer and several state-of-the-art methods

Methods	CSD		SRRS		Snow100K		Overhead	
	PSNR/dB \uparrow	SSIM \uparrow	PSNR/dB \uparrow	SSIM \uparrow	PSNR/dB \uparrow	SSIM \uparrow	# Param	# FLOPs
DesnowNet ^[12]	20.13	0.81	20.38	0.84	30.50	0.94	1.560×10^7	1.70 KG
JSTASR ^[5]	27.96	0.88	25.82	0.89	23.12	0.86	6.500×10^7	—
DDMSNet ^[33]	28.79	0.90	27.03	0.91	30.76	0.91	2.2945×10^9	—
HDCW-Net ^[14]	29.06	0.91	27.78	0.92	31.54	0.95	6.990×10^6	9.78 G
SMGARN ^[46]	31.93	0.95	29.14	0.94	31.92	0.93	6.860×10^6	450.30 G
TransWeather ^[11]	31.76	0.93	28.29	0.92	31.82	0.93	2.190×10^7	5.64 G
TKL ^[45]	33.89	0.96	30.82	0.96	34.37	0.95	3.135×10^7	41.58 G
DGUNet ^[44]	34.74	0.97	31.28	0.96	34.21	0.95	1.218×10^7	199.74 G
Uformer ^[36]	33.80	0.96	30.12	0.96	33.81	0.94	9.030×10^6	19.82 G
Restormer ^[15]	35.43	0.98	32.19	0.95	34.67	0.95	2.612×10^7	140.99 G
NAFNet ^[43]	35.13	0.97	32.13	0.97	34.49	0.95	2.240×10^7	12.12 G
FocalNet ^[19]	37.18	0.99	31.34	0.98	33.53	0.95	3.740×10^6	30.63 G
CPLFormer ^[37]	37.83	0.98	<u>33.49</u>	<u>0.98</u>	<u>35.85</u>	<u>0.96</u>	8.780×10^6	11.12 G
ConvIR-S ^[34]	38.43	0.99	32.25	0.98	35.46	0.96	5.530×10^6	42.10 G
ConvIR-B ^[34]	39.10	<u>0.99</u>	32.39	0.98	35.59	0.96	8.630×10^6	71.22 G
DFE-DARFormer	<u>38.77</u>	0.99	34.47	0.98	36.47	0.96	1.006×10^7	21.86 G

4.5.2 定性评估

图 4 对比了本文算法与现有先进方法在 CSD 数据集上的视觉效果。从图中可以观察到:HDCW-Net 仍然存在明显的雪痕或者严重的伪影;Restormer 通过在通道层面上进行全局建模,一定程度上提升了恢复大尺度雪花遮挡区域的效果,但是其忽略了在空间域上信息的差异,导致部分场景细节丢失;NAFNet 和 FocalNet 能保留更多的场景细节,但是在大面积退化区域恢复的表现仍然不如本文方法。虽然 FocalNet 通过设计双域选择的机制增强了对重要区域的处理,但却未能有效地扩大此部分区域的感受野来聚合更多的信息,使得对严重退化区域的恢复效果仍然不如本文方法;虽然 ConvIR 通过多尺度的设计以及动态地对信息丰富的频率成分进行关注,使

其在细节纹理与大尺度退化区域重建均有一定的效果,但其对大尺度退化区域仍然缺乏足够的感受野,相比本文方法,其重建后的图像在大尺度退化区域仍残留较多伪影。得益于退化路由机制,本文方法 DFE-DARFormer 能够针对退化相关区域聚合全局信息,使得大面积图像信息损失的区域得到更有效的恢复。同时,通过结合双频域增强设计,本文方法在高频和低频信息的恢复上均表现出色。从图 4 中可以观察到,本文方法不仅在去除雪残留方面表现更优,保留了更多的场景细节,更为重要的是,其在恢复被大雪严重遮挡的场景区域上也表现出显著的效果提升。相比之下,其他方法或是由于去除不彻底,在图像中留下明显的雪痕,或是去除后引入伪影或导致场景细节丢失。

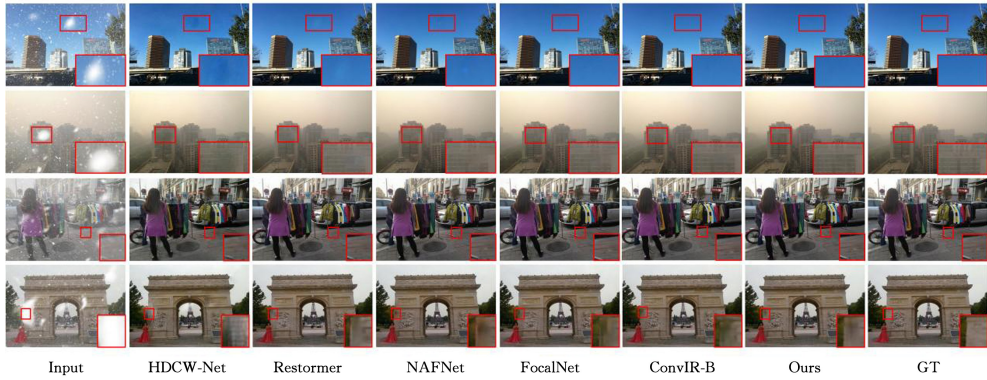


图 4 不同方法在合成雪景图像上的去雪效果视觉比较

Fig. 4 Visual comparison of snow removal performance on synthetic snowy images using different methods

为了更好地评估本文方法的性能,还在 Snow100K 的真实场景数据集上进行了测试。

如图 5 所示,本文方法生成了在视觉效果上更加令人

满意的去雪效果。其他方法的结果或是场景信息被过度去除,或是大面积退化区域恢复效果不理想,而本文方法表现出更优的效果。

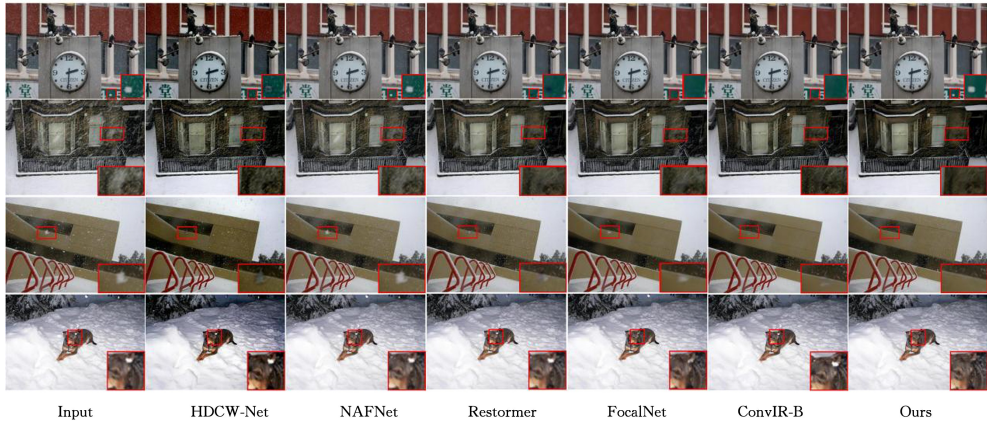


图 5 不同方法在 Snow100K 真实雪景数据集上的去雪效果视觉比较

Fig. 5 Visual comparison of snow removal performance using different methods on the Snow100K real-world snowy dataset

4.5.3 消融实验

为了验证 DFE-DARFormer 的有效性,本文在 CSD 数据集上进行消融实验,以评估各个模块的作用。本文设计了以下两组消融实验。

1)退化感知路由 Transformer 的消融实验。其中 MSA 表示将退化感知路由 Transformer 替换为基于窗口的自注意力机制;Base 表示将整个退化感知路由 Transformer 移除,而双频增强 Transformer(DFET)保持不变;Model-A 表示在只将退化感知路由注意力(DARA)中的退化感知路由注意力机制(DARM)移除。

如表 2 所列,本文方法取得了最佳性能。为了验证本文设计的退化感知路由注意力机制的确如上述理论作用于严重退化区域,还设计了 Model-B 表示直接在本文方法的预训练模型中关闭所有 DARM 的路由作用,并进行了主观视觉对比,如图 6 所示。

可以观察到,直接在预训练模型中关闭退化感知路由机制的 Model-B 的输出结果只在严重退化区域出现大面积伪影以及雪痕,即只作用于严重退化区域。同时可以观察到,Model-A 在大面积退化区域仍然残留较多雪痕,且场景结构轮廓的恢复效果明显不如本文提出的方法。这表明本文中的

退化感知路由 Transformer 在严重雪退化区域具有显著的恢复效果。

表 2 在 CSD 数据集上对 DART 模块进行消融实验的定量比较
Table 2 Quantitative comparison of DART in ablation experiments on the CSD dataset

对比项	Base	MSA	Model-A	FERTB
PSNR/dB	36.73	38.25	38.17	38.77
SSIM	0.97	0.98	0.98	0.99
# Param	3.970×10^6	7.460×10^6	6.820×10^6	1.006×10^7

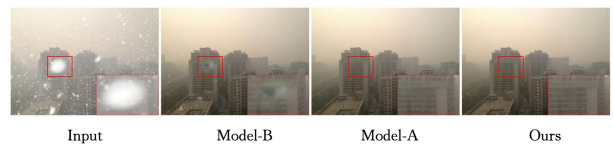


图 6 DARA 消融实验的去雪效果视觉比较

Fig. 6 Visual comparison of snow removal performance in DARA ablation study

2)双频域增强 Transformer 的消融实验。为了验证本文双频域增强 Transformer(DFET)的有效性,从以下几个方面进行了消融实验;Base 表示移除 DFET 模块;CDC 表示仅用中心差分卷积替代 DFET 模块;MSA 表示仅用基于窗

口模块;HFEB表示用残差卷积块替代DFET模块的低频分支所使用的自注意力机制;LFEB表示用残差卷积块替代DFET中的中心差分卷积块。不同模型的PSNR和SSIM性能指标对比结果如表3所列,可以看出,无论是离

散小波对特征图进行鲁棒分频的设计,还是中心差分卷积与自注意力机制分别在高低频优势的结合,均有助于提升模型的性能指标,缺少其中任何一个模块都会导致模型性能下降。

表3 在CSD数据集上对DFET模块进行消融实验的定量比较

Table 3 Quantitative comparison of the ablation study of the DFET module on the CSD dataset

	Base	CDC	MSA	HFEB	LFEB	Ours
CDC	×	√	×	√	×	√
MSA	×	×	√	×	√	√
RB	×	×	×	√	√	×
DWT	×	×	×	√	√	√
PSNR/dB	36.41	37.14	37.96	37.41	38.47	38.77
SSIM	0.97	0.97	0.98	0.98	0.99	0.99
# Param	5.860×10^6	6.040×10^6	8.060×10^6	6.120×10^6	8.140×10^6	1.006×10^7

结束语 图像去雪在许多下游任务中具有重要的应用。恢复大面积雪退化遮挡区域的场景信息仍是当前的主要挑战。自注意力机制因能够建立长距离依赖,被广泛用于图像领域。然而,当前的大多数研究仍然是在限制的窗口感受野内进行信息聚合,导致缺乏进一步地扩大感受野以聚合更多图像信息,提高对严重退化区域的恢复能力。

针对以上问题,本文提出了FE-DRAT模型,通过退化感知路由注意力重点关注退化区域的全局性,通过路由机制对退化区域进行全局依赖建立,可在节约计算成本的前提下获得全局感受野。此外,双频域增强Transformer通过小波分频设计结合差分卷积与自注意力机制,在各频域放大优势,有效解决了小尺度雪花难识别、大尺度雪花难恢复的问题。

未来将探索如何将模型扩展到其他图像恢复领域,并改进模型以更好地适应真实雪景图像。其次,将探索如何设计轻量高效的模型和算法,以实现实时的图像去雪。

参考文献

- [1] YANG R, LI W, SHANG X, et al. KPE-YOLOv5: an improved small target detection algorithm based on YOLOv5 [J]. *Electronics*, 2023, 12(4): 817.
- [2] FAN J, YANG X, LU R, et al. Long-term visual tracking algorithm for UAVs based on kernel correlation filtering and SURF features [J]. *The Visual Computer*, 2023, 39(1): 319-333.
- [3] CHANG Y, TU Z, XIE W, et al. Video anomaly detection with spatio-temporal dissociation [J]. *Pattern Recognition*, 2022, 122: 108213.
- [4] LIU Y, YANG D, WANG Y, et al. Generalized video anomaly event detection: Systematic taxonomy and comparison of deep models [J]. *ACM Computing Surveys*, 2024, 56(7): 1-38.
- [5] CHEN W T, FANG H Y, DING J J, et al. JSTASR: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal [C] // *Computer Vision-ECCV 2020; 16th European Conference*. 2020: 754-770.
- [6] PEI S C, TSAI Y T, LEE C Y. Removing rain and snow in a single image using saturation and visibility features [C] // *2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. IEEE, 2014: 1-6.
- [7] ZHENG X, LIAO Y, GUO W, et al. Single-image-based rain and snow removal using multi-guided filter [C] // *Neural Information Processing; 20th International Conference, ICONIP 2013*. Berlin: Springer, 2013: 258-265.
- [8] VORONIN V, SEMENISHCHEV E, ZHDANOVA M, et al. Rain and snow removal using multi-guided filter and anisotropic gradient in the quaternion framework [C] // *Artificial Intelligence and Machine Learning in Defense Applications*. SPIE, 2019: 227-233.
- [9] HUANG J, LIU Y, FU X, et al. Exposure normalization and compensation for multiple-exposure correction [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022: 6043-6052.
- [10] HUANG J, ZHAO F, ZHOU M, et al. Learning sample relationship for exposure correction [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023: 9904-9913.
- [11] VALANARASU J M J, YASARLA R, PATEL V M. Trans-weather: Transformer-based restoration of images degraded by adverse weather conditions [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022: 2353-2363.
- [12] LIU Y F, JAW D W, HUANG S C, et al. Desnownet: Context-aware deep network for snow removal [J]. *IEEE Transactions on Image Processing*, 2018, 27(6): 3064-3073.
- [13] LI B, REN W, FU D, et al. Benchmarking single-image dehazing and beyond [J]. *IEEE Transactions on Image Processing*, 2018, 28(1): 492-505.
- [14] CHEN W T, FANG H Y, HSIEH C L, et al. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss [C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021: 4196-4205.
- [15] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: Efficient transformer for high-resolution image restoration [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022: 5728-5739.
- [16] ALEXEY D. An image is worth 16x16 words: Transformers for image recognition at scale [J]. *arXiv*; 2010. 11929, 2020.
- [17] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021: 10012-10022.
- [18] SONG Y, HE Z, QIAN H, et al. Vision transformers for single image dehazing [J]. *IEEE Transactions on Image Processing*, 2023, 32: 1927-1941.

- [19] CUI Y, REN W, CAO X, et al. Focal network for image restoration[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023:13001-13011.
- [20] CHEN S, YE T, LIU Y, et al. Dual-former: Hybrid self-attention transformer for efficient image restoration[J]. Digital Signal Processing, 2024, 149:104485.
- [21] MALLAT S. A wavelet tour of signal processing[M]. Academic Press, 1999.
- [22] YU Z, ZHAO C, WANG Z, et al. Searching central difference convolutional networks for face anti-spoofing[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020:5295-5305.
- [23] BOSSU J, HAUTIERE N, TAREL J P. Rain or snow detection in image sequences through use of a histogram of orientation of streaks[J]. International Journal of Computer Vision, 2011, 93: 348-367.
- [24] HUANG L W, CUI W C, SHAO H. Research on Pedestrian Detection Method Based on Multi-Layer Feature Fusion[J]. Computer Science, 2024, 51(S2):489-495.
- [25] LIU Z K, YIN J B. Text-Driven Emotional Diversification in Facial Animation Generation[J]. Computer Science, 2024, 51(S2): 323-330.
- [26] LI W X, ZHANG J, ZHUO L, et al. Advances in Vision Segmentation Techniques Based on Transformer[J]. Journal of Computer Science and Technology, 2024, 47(12):2760-2782.
- [27] LI T, ZHAO E D, YANG J. Road Obstacle Detection Method Based on Self-Attention and Bidirectional Feature Fusion[J]. Computer Science, 2024, 51(S2):287-291.
- [28] JIANG C J, HE X Y, XIANG J. LOL-YOLO: Low-Light Object Detection with Multi-Attention Mechanism Fusion[J]. Computer Engineering and Applications, 2024, 60(24):177-187.
- [29] TANG S, ZENG W L, YANG S L, et al. Single Image Super-Resolution Reconstruction Network Based on Transformer with Block-Intra and Block-Inter Dual Aggregation[J]. Journal of Computer Science and Technology, 2024, 47(12):2783-2802.
- [30] YANG Y D, GE H B, XIN S A, et al. Lightweight Remote Sensing Image Small Object Detection Fusing Super-Resolution and Feature Enhancement[J]. Computer Engineering, 2024, 50(11): 284-296.
- [31] REN J, ZHOU G, ZHU Y, et al. Unsupervised Domain Adaptive Learning for Image Desnowing with Real-World Data[C] // 2023 IEEE International Conference on Image Processing (ICIP). IEEE, 2023:3050-3054.
- [32] YASARLA R, SINDAGI V A, PATEL V M. Unsupervised restoration of weather-affected images using deep gaussian process-based cyclegan[C] // 2022 26th International Conference on Pattern Recognition (ICPR). IEEE, 2022:1967-1974.
- [33] ZHANG K, LI R, YU Y, et al. Deep dense multi-scale network for snow removal using semantic and depth priors[J]. IEEE Transactions on Image Processing, 2021, 30:7419-7431.
- [34] CUI Y, REN W, CAO X, et al. Revitalizing convolutional network for image restoration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12):9423-9438.
- [35] LIANG J, CAO J, SUN G, et al. Swinir: Image restoration using swin transformer[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021:1833-1844.
- [36] WANG Z, CUN X, BAO J, et al. Uformer: A general u-shaped transformer for image restoration[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:17683-17693.
- [37] CHEN S, YE T, LIU Y, et al. CPLFormer: Cross-scale Prototype Learning Transformer for Image Snow Removal[C] // Proceedings of the 31st ACM International Conference on Multimedia. 2023:4228-4239.
- [38] XU F, NARESH BODDATI V, SAVVIDES M. Local binary convolutional neural networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 19-28.
- [39] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C] // Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. 2015:234-241.
- [40] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:7794-7803.
- [41] JIANG L, DAI B, WU W, et al. Focal frequency loss for image reconstruction and synthesis[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 13919-13929.
- [42] KINGMA D P. Adam: A method for stochastic optimization[J]. arXiv:1412.6980, 2014.
- [43] CHEN L, CHU X, ZHANG X, et al. Simple baselines for image restoration[C] // European Conference on Computer Vision. Cham: Springer, 2022:17-33.
- [44] MOU C, WANG Q, ZHANG J. Deep generalized unfolding networks for image restoration[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 17399-17410.
- [45] CHEN W T, HUANG Z K, TSAI C C, et al. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:17653-17662.
- [46] CHENG B, LI J, CHEN Y, et al. Snow mask guided adaptive residual network for image snow removal[J]. Computer Vision and Image Understanding, 2023, 236:103819.



QIN Yi, born in 1996, master. His main research interests include image processing and computer vision.



WANG Minghui, born in 1971, Ph. D., professor. His main research interests include intelligent medical image processing and computer vision.