



## Image Dehazing Network Based on Dual-Domain Feature Fusion

---

Weijia Wang, Fei Chen, Wanling Liu, Hang Cheng and  
Meiqing Wang

EasyChair preprints are intended for rapid  
dissemination of research results and are  
integrated with the rest of EasyChair.

March 31, 2024

# 基于双域特征融合的图像去雾网络

王炜嘉<sup>1)</sup>, 陈飞<sup>1)\*</sup>, 刘莞玲<sup>1,2)</sup>, 程航<sup>3)</sup>, 王美清<sup>3)</sup>

<sup>1)</sup> (福州大学计算机与大数据学院 福州 350108)

<sup>2)</sup> (天津大学智能与计算学部 天津 300350)

<sup>3)</sup> (福州大学数学与统计学院 福州 350108)  
(chenfei314@fzu.edu.cn)

**摘要:** 图像去雾的目标是从有雾图像中恢复潜在的无雾图像。现有方法利用清晰/退化图像对在空间域和频率域的差异进行去雾并取得一定的效果, 但是仍存在双域特征融合中空间域特征提取与融合不够充分、频率域特征融合效果不佳等问题。为此, 提出一种新颖的双域特征融合网络 DFFNet。首先, 设计更适合图像软重建的空间域特征融合模块 SFFM, 采用 Transformer 风格架构, 并引入大核注意力与像素注意力, 通过不同感受野分别对全局特征和局部特征进行建模, 并通过卷积前馈网络进行映射与融合。同时, 设计频率域特征融合模块 FFFM, 使用卷积层来放大并丰富高频特征, 并采取一种新颖的特征融合策略, 利用通道交互来强调与融合多种高频特征。结合这两种关键设计提出的 DFFNet 在两个基准数据集上展现出与最先进方法相当甚至更好的性能。DFFNet-L 是第一个在 SOTS-Indoor 数据集上峰值信噪比(peak signal-to-noise ratio, PSNR)超过 43dB 的去雾方法, PSNR 为 43.83dB。代码发布于 <https://github.com/WWJ0720/DFFNet>。

**关键词:** 图像去雾; 空间域特征融合; 频率域特征融合; 注意力机制; 深度学习

**中图法分类号:** DOI:

## Image Dehazing Network Based on Dual-domain Feature Fusion

Wang Weijia<sup>1)</sup>, Chen Fei<sup>1)\*</sup>, Liu Wanling<sup>1,2)</sup>, Cheng Hang<sup>3)</sup> and Wang Meiqing<sup>3)</sup>

<sup>1)</sup> (College of Computer and Data Science, Fuzhou University, Fuzhou 350108)

<sup>2)</sup> (College of Intelligence and Computing, Tianjin University, Tianjin 300350)

<sup>3)</sup> (School of Mathematics and Statistics, Fuzhou University, Fuzhou 350108)

**Abstract:** The goal of image dehazing is to restore the latent haze-free image from a hazy image. Existing methods utilize the differences between clear/degraded image pairs in both spatial-domain and frequency-domain for dehazing and achieve some achievement. However, there are still issues with insufficient spatial-domain feature extraction and fusion, as well as unsatisfactory frequency-domain feature fusion in dual-domain feature fusion. To address this, a novel dual-domain feature fusion network called DFFNet is proposed. Firstly, a spatial-domain feature fusion module (SFFM) is designed, which is more suitable for image soft reconstruction. It adopts a Transformer-style architecture and incorporates large kernel attention and pixel attention to model global and local features separately using different receptive fields, and performs mapping and fusion through a convolutional feed-forward network. Simultaneously, a frequency-domain feature fusion module (FFFM) is designed, which uses convolutional layers to amplify and enrich

收稿日期: 20\*\*-\*\*-\*\*; 修回日期: 20\*\*-\*\*-\*\*。基金项目: 国家自然科学基金项目(61771141, 62172098); 福建省自然科学基金项目(2021J01620); 福建省中青年教育科研项目(科技类 JAT231010)。王炜嘉(2000—), 男, 硕士研究生, 主要研究方向为图像处理与计算机视觉; 陈飞(1980—), 男, 博士, 副教授, 硕士生导师, 论文通讯作者, 主要研究方向为图像处理、计算机视觉与机器学习; 刘莞玲(1991—), 女, 博士研究生, 实验师, CCF 会员, 主要研究方向为智能决策与计算机视觉; 程航(1979—), 男, 博士, 教授, 博士生导师, 主要研究方向为机器学习与图像处理; 王美清(1967—), 女, 博士, 教授, 博士生导师, 主要研究方向为图像处理;

high-frequency features. It employs a novel feature fusion strategy that emphasizes and fuses multiple high-frequency features through channel interaction. Combining these two key designs, the proposed DFFNet demonstrates comparable or even better performance than state-of-the-art methods on two benchmark datasets. DFFNet-L is the first dehazing method to achieve a peak signal-to-noise ratio (PSNR) exceeding 43dB on the SOTS-Indoor dataset, with a PSNR of 43.83dB. The code has been released on <https://github.com/WWJ0720/DFFNet>.

**Key words:** image dehazing; spatial-domain feature fusion; frequency-domain feature fusion; attention mechanism; deep learning

在有雾场景下捕获的图像通常会出现对比度降低和颜色失真等严重退化, 这些退化会显著影响高级计算机视觉任务的性能, 例如目标检测和场景理解. 因此, 图像去雾任务作为图像恢复领域的研究重点引起了学术界和工业界的广泛关注, 并在监控、自动驾驶技术和遥感等领域发挥着重要的作用.

图像去雾的目标是从有雾图像中恢复潜在的无雾图像. 在图像恢复领域, 通常采用大气散射模型(atmospheric scattering models, ASM)<sup>[1,2,3]</sup>来对有雾图像的退化过程进行建模:

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

其中  $I$ 、 $J$ 、 $t$ 、 $A$  分别表示有雾图像、无雾图像、介质透射率和全局大气光.  $t$  可以进一步表示为  $t(x) = e^{-\beta d(x)}$ , 其中  $\beta$  是大气的散射系数,  $d$  是场景深度. 早期的图像去雾方法基于经验统计<sup>[4,5,6]</sup>, 利用先验来估计  $A$  和  $t$ . 然而, 这些方法通常只在满足其先验条件的特定场景中表现良好.

近年来, 用于图像去雾的深度学习方法经历了快速的发展, 并在性能上显著超越了传统方法. 根据网络架构的不同, 现有的图像去雾模型可大致分为两类: 基于卷积神经网络(convolutional neural network, CNN)的模型<sup>[7,8,9]</sup>主要关注增加网络的深度和宽度来提高模型的代表能力和泛化能力, 同时扩大网络的感受野, 以更好地理解图像中的纹理、形状等结构化信息. 基于 Transformer 的模型<sup>[10,11]</sup>具有出色的长距离建模能力和全局感受野, 可以捕捉到不同区域雾的分布和特征, 在恢复大规模有雾区域时表现出色, 但需要巨大的参数规模和昂贵的训练成本. 基于 CNN 的模型虽然感受野大小不如基于 Transformer 的模型, 但可以充分利用卷积和池化操作的局部性, 从而大大减少计算资源的开销. 所以, 一些现有的工作<sup>[12,13,14]</sup>尝

试结合 Transformer 和 CNN 的优势来设计图像去雾模型, 并取得了卓越的成果.

同时, 除了利用清晰/退化图像对在空间域的差异进行去雾以外, 清晰/退化图像对在频率域的差异对去雾也有着显著的作用, 可以为空间域特征融合结果提供补充信息. 如图 1 所示, (a)、(b)是 RESIDE-Indoor 数据集中的无雾图像与对应的重度、轻度有雾图像相减所得残差图像的离散傅里叶变换结果. 可以看出无雾/有雾图像对在频域的差异大部分在低频, 少部分在高频. 高频特征代表图像的边缘和细节, 低频特征代表图像的整体结构和大致形状. 现有的工作利用变换工具, 如小波变换<sup>[15]</sup>、傅里叶变换<sup>[16]</sup>、池化技术<sup>[17]</sup>和滤波器<sup>[18]</sup>, 将输入特征分解成不同频率子带, 然后分别处理这些子带以重建相应的输出特征.

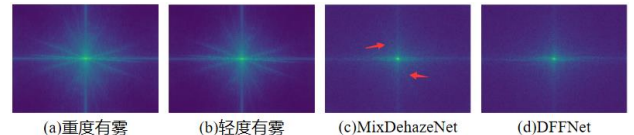


图 1 不同图像与 RESIDE-Indoor 中的无雾图像相减所得残差图像的离散傅里叶变换结果

然而, 双域特征融合存在两个主要的问题: (1) 空间域特征提取与融合不够充分. 传统空间域特征提取模块的小感受野难以应对大范围有雾的情况. 同时, 上下文信息的不充分利用导致模型对图像整体结构和细节的理解不够准确. 并且没有考虑多语义特征的融合, 不能很好地适应基于 ASM 的软重建(soft reconstruction, SR)<sup>[13]</sup>对于特征的要求. (2) 频率域特征融合效果不佳. 现有的频率域特征融合模块广泛采用选择性核融合(selective kernel fusion, SKFusion)<sup>[13]</sup>用于特征融合, 但是 SKFusion 存在两个明显的问题: SKFusion 的初始特征融合方式采取简单的加法以及 SKFusion 只能

融合两种频率特征. 初始特征融合的质量会直接影响到融合后的重建特征的质量, 同时融合频率特征数量的限制也会对重建特征的质量产生显著影响.

针对上述问题, 本文提出一种新颖的双域特征融合网络(dual-domain feature fusion network, DFFNet). DFFNet 基于软重建<sup>[13]</sup>来恢复潜在的无雾图像, 并结合两个关键设计: 空间域特征融合模块(spatial-domain feature fusion module, SFFM)和频率域特征融合模块(frequency-domain feature fusion module, FFFM).

SFFM 是一个 Transformer 风格的模块, 对视觉注意力网络(visual attention network, VAN)<sup>[19]</sup>进行重新设计使其更适合图像去雾任务, 并引入了像素注意力(pixel attention, PA)和大核注意力(large kernel attention, LKA), 以不同感受野分别对软重建所需的局部特征与全局特征进行建模, 并将特征通过卷积前馈网络(convolutional feed-forward network, CFFN)<sup>[20]</sup>进行映射与融合.

如图 1 所示, (c)、(d)分别是目前最先进方法 MixDehazeNet 和提出的 DFFNet 的去雾结果与无雾图像相减所得残差图像的离散傅里叶变换结果. MixDehazeNet 在高频和低频的恢复都展现出不错的效果, 但仍与无雾图像存在一定的差异, 高频的差异如(c)中箭头所指. 由于低频的恢复主要基于空间域, 所以 DFFNet 中的 FFFM 主要针对高频进行恢复. FFFM 受 Cui 等<sup>[18]</sup>和 Dai 等<sup>[21]</sup>的启发, 采用多个卷积层作为高通滤波器来放大高频特征的方差与幅度, 增加高频特征的多样性, 并采取一种新颖的特征融合策略来强调与融合多种高频特征, 以更好地恢复图像中的高频子带.

DFFNet 在 SOTS-Indoor 和 Haze4K 两个基准数据集上展现出与最先进方法相当甚至更好的性能. DFFNet-L 与其他图像去雾方法在 SOTS-Indoor 数据集上的比较如图 2 所示, 圆的大小代表参数规模. DFFNet-L 以 46%的参数量以及 67%的每秒浮点运算次数(floating point operations per second, FLOPs)在峰值信噪比(peak signal-to-noise ratio, PSNR)上超越当前图像去雾领域的最先进方法 MixDehazeNet-L<sup>[14]</sup> 1.21dB, 并且 DFFNet-L 是第一个在 SOTS-Indoor 数据集上超过 43dB 的方法.

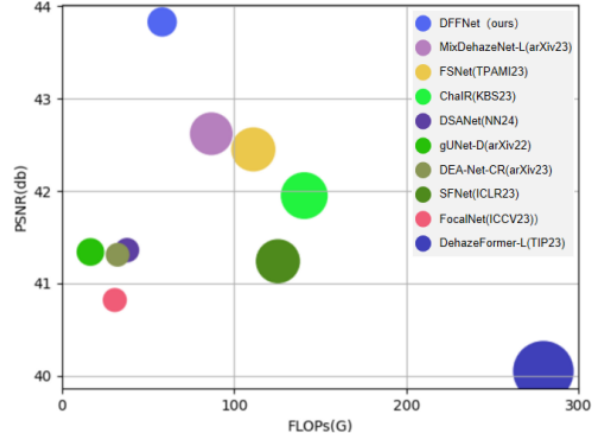


图 2 DFFNet 与其他图像去雾方法在 SOTS-Indoor 数据集上的比较

## 1 相关工作

### 1.1 基于空间域的图像去雾

图像去雾方法可分为基于先验的方法和基于学习的方法. 基于先验的方法依赖 ASM<sup>[1,2,3]</sup>, 通常使用手工设计的先验来去雾, 例如暗通道先验(dark channel prior, DCP)<sup>[4]</sup>、颜色衰减先验(color attenuation prior, CAP)<sup>[5]</sup>和非局部先验(non-local prior, NLP)<sup>[6]</sup>. 然而, 由于这些先验是基于经验统计的, 因此当场景不满足这些先验的假设时, 这些方法往往表现不佳.

基于学习的方法依赖数据集来恢复潜在的无雾图像. 早期基于学习的方法<sup>[22,23]</sup>使用神经网络来预测 ASM 中的  $A$  和  $t$ , 以恢复潜在的无雾图像. 然而, 预测的不准确性可能会累积, 从而影响无雾图像的恢复质量.

随着 RESIDE 等大规模数据集的出现, 近几年的方法<sup>[7,8,9]</sup>倾向于直接恢复无雾图像. GridDehazeNet<sup>[7]</sup>通过基于注意力的多尺度估计来解决瓶颈问题, 通过后处理模块减少伪影, 并首次指出了利用 ASM 提供的降维来进行图像去雾的问题并给出可能的解释. FFA-Net<sup>[8]</sup>的特征注意力模块分别使用通道注意力和像素注意力来处理不同信息, 并提出基于注意力的不同层级特征融合结构给予重要特征更高的权重. PMNet<sup>[9]</sup>通过可分离混合注意力模块和密度图来建模雾的不均匀分布, 从而有效地捕捉特征级别的不均匀退化.

考虑到 Transformer 具有出色的长距离建模能力和全局感受野, 一些现有的方法<sup>[10,11]</sup>尝试设计基于 Transformer 的去雾模型. Valanarasu 等<sup>[10]</sup>使用

内部补丁 Transformer 编码器和具有可学习的天气类型嵌入的 Transformer 解码器来适应天气退化. Wang 等<sup>[11]</sup>引入了局部增强窗口 Transformer 块和多尺度恢复调制器来减少额外参数和计算成本.

但是基于 Transformer 的模型仍然难以训练, 一些现有的工作<sup>[12,13,14]</sup>尝试结合 Transformer 和 CNN 的优点来设计去雾模型. Guo 等<sup>[12]</sup>提出通过学习调制矩阵在 Transformer 特征的条件下调制 CNN 特征, 拥有 Transformer 的全局上下文建模能力和 CNN 的局部表示能力. Song 等<sup>[13]</sup>修改了 Swin Transformer 中不适合图像去雾的关键设计并提出改进后的 DehazeFormer. Lu 等<sup>[14]</sup>提出了一个 Transformer 式的混合结构块, 通过并行多尺度并行大卷积核来引入大感受野和多尺度特征, 通过增强型并行注意模块处理不均匀雾气分布.

## 1.2 基于频率域的图像去雾

最近的工作尝试利用清晰/退化图像对在频率域的差异进行去雾. 无雾/有雾图像对在频域上的差异大部分在低频, 少部分在高频. 由于低频主要基于空间域进行恢复, 基于频率域的图像去雾的研究重点在于如何提取输入特征中的高频特征以及如何融合高频特征以更好地恢复图像中的高频子带. FocalNet<sup>[17]</sup>提出了一种频率选择模块, 通过使用全局平均池化获得输入特征中的低频特征, 用输入特征减去低频特征来获得高频特征, 并使用高频特征与输入特征的逐元素乘法以及残差连接得到输出特征. FSNet<sup>[24]</sup>提出了多分支动态选择频率模块和多分支紧凑型选择频率模块, 分别使用多尺度卷积和池化将特征解耦为不同的频率特

征, 并使用 SKFusion<sup>[13]</sup>和加法来重建输出特征. ChaIR<sup>[18]</sup>引入了隐式频域通道注意模块, 该模块使用卷积来放大高频特征并使用 SKFusion<sup>[13]</sup>融合高频特征以重建输出特征.

## 2 本文方法

### 2.1 整体架构

DFFNet 的总体架构如图 3 所示. DFFNet(a) 的架构采用与 UNet 相似的编码器-解码器架构, 由三个尺度的编码器和解码器组成, 每个尺度堆叠  $N$  个 SFFM(b). SFFM 主要由 LKA(c)、PA 和 CFFN(d) 组成. 为了减少计算成本, FFFM(e) 仅使用于第三尺度的编码器与解码器之间. 第一、第二尺度编码器的输出特征分别与经上采样的第二、第三尺度解码器的输出特征通过 SKFusion<sup>[13]</sup>动态融合, 第一尺度解码器的输出特征经过块还原后与输入图像通过 SR 来恢复潜在的无雾图像.

SR 的设计受<sup>[13]</sup>的启发, 将式(1)重写为:

$$J(x) = K(x)I(x) + B(x) + I(x) \quad (2)$$

其中,  $K(x) = 1/t(x) - 1$ ,  $B(x) = -(1/t(x) - 1)A$ . SR 约束了  $K(x)$  和  $B(x)$  的关系, 将第一尺度解码器的输出特征  $D_1 \in \mathbb{R}^{4 \times H \times W}$  分解为  $K \in \mathbb{R}^{1 \times H \times W}$  和  $B \in \mathbb{R}^{3 \times H \times W}$ , 并通过式(2)重建潜在的无雾图像. 当  $K(x) = 0$  时, SR 通过预测全局残差来恢复无雾图像. SR 通过基于 ASM 的先验更好地约束图像的结构和特征, 从而提高无雾图像的恢复效果.

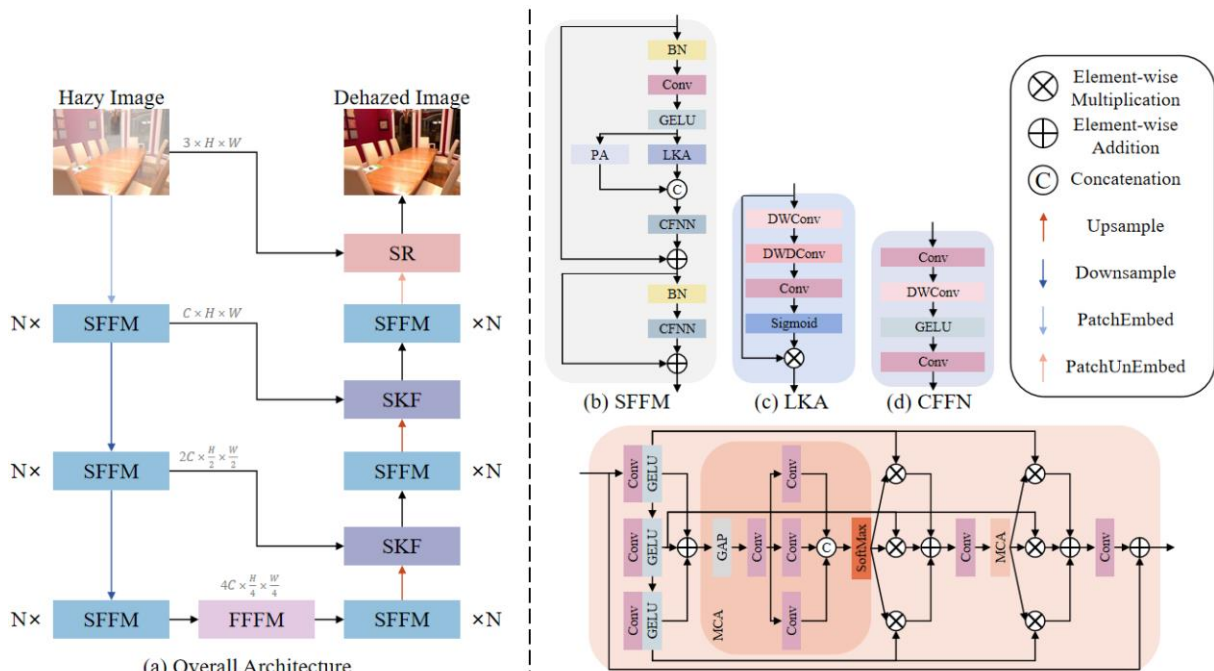


图 3 DFFNet 的总体架构

## 2.2 SFFM

SFFM 受 VAN<sup>[19]</sup> 的启发, 采用 Transformer 风格架构, 并由两个关键组件构成: 结合 LKA 和 PA 来模拟的多头自注意力和使用 CFFN<sup>[20]</sup> 作为多层感知机的前馈神经网络. SFFM 在拥有大感受野与更适合图像去雾软重建<sup>[13]</sup> 的设计的同时, 保持着较小的参数规模和计算复杂度.

具体来说, 给定输入特征  $X \in \mathbb{R}^{C \times H \times W}$ , 首先通过对  $X$  进行批归一化(batch normalization, BN), 采用  $1 \times 1$  卷积进行特征提取, 并使用高斯误差线性单元(gaussian error linear unit, GELU)激活后得到中间特征  $X'$ .

接着, 分别使用 LKA 和 PA 来选择性地关注  $X'$  中的全局特征和局部特征. LKA 通过将大核卷积分解成深度卷积、深度扩张卷积和  $1 \times 1$  卷积的方式, 在引入大感受野的同时极大地减少了计算成本. 这样的大感受野可以更充分地利用上下文信息来建模  $X'$  中的全局特征  $A$ , 并关注图像中雾的分布, 更好地恢复图像中的大片有雾区域. 而 PA 通过对  $X'$  中的每个像素位置应用注意力机制, 更好地对  $X'$  中位置相关的局部特征  $t$  进行建模, 更好地恢复图像的边缘和细节. LKA 和 PA 所建模的特征可以互相补充, 以满足软重建对于全局特征和局部特征的需求.

$$X'_{LKA} = \sigma \left( C_{1 \times 1} \left( DWDC_{7 \times 7} \left( DWC_{5 \times 5} (X') \right) \right) \right) \otimes X' \quad (3)$$

$$X'_{PA} = \sigma \left( C_{1 \times 1} \left( GELU \left( C_{1 \times 1} (X') \right) \right) \right) \otimes X' \quad (4)$$

其中,  $DWC_{5 \times 5}$  代表  $5 \times 5$  深度卷积,  $DWDC_{7 \times 7}$  代表  $7 \times 7$  深度扩张卷积,  $C_{1 \times 1}$  代表  $1 \times 1$  的卷积,  $\sigma$  代表 Sigmoid 函数. 将经 LKA 和 PA 强调后的输出特征  $X'_{LKA}$  和  $X'_{PA}$  连接起来并将其输入 CFFN, 因为每个通道的退化模式不一样, 而 CFFN 中的深度卷积可以进行通道分离变换, 在 SFFM 的多头自注意力中对特征使用 CFFN 既可以对特征中不同通道进行融合与映射以拟合软重建中的特征, 也可以增强 SFFM 的表征能力.

$$X_{CFFN} = C_{1 \times 1} \left( GELU \left( DWDC_{3 \times 3} \left( C_{1 \times 1} \left( [X'_{LKA}, X'_{PA}] \right) \right) \right) \right) \quad (5)$$

其中  $[, \cdot]$  表示连接. 将 CFFN 的输出  $X_{CFFN}$  与输入特征  $X$  进行残差连接得到多头自注意力部分的输出  $X_{MHSA}$ .

SFFM 中的前馈神经网络部分对  $X_{MHSA}$  进行映射与非线性变换以学习更丰富的表示. 前馈神

经网络部分首先对  $X_{MHSA}$  进行 BN, 然后使用 CFFN 进行映射并将输出特征与  $X_{MHSA}$  进行残差连接, 得到 SFFM 的输出  $X_{SFFM}$ .

$$X_{SFFM} = C_{1 \times 1} \left( GELU \left( DWDC_{3 \times 3} \left( C_{1 \times 1} \left( BN \left( X_{MHSA} \right) \right) \right) \right) \right) + X_{MHSA} \quad (6)$$

## 2.3 FFFM

FFFM 的设计受 ChaIR<sup>[18]</sup> 的启发, 堆叠使用三个卷积层作为高通滤波器来放大输入特征中的高频特征并丰富高频特征的多样性. 更多的堆叠层数可以提高输出特征的质量, 但出于参数规模和计算效率的考虑, DFFNet 堆叠三个卷积层来提取三种高频特征. 提取效果如图 4 所示, (a) 中的高频特征被卷积层放大后得到 (b)、(c)、(d), 并且经过更多卷积层放大后的高频特征更突出, 所以可以通过不同的卷积层获得丰富的高频特征表示.

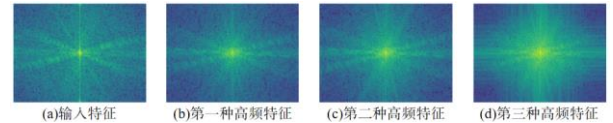


图 4 不同卷积层的特征提取效果

具体来说, 假设输入特征为  $X \in \mathbb{R}^{C \times H \times W}$ , FFFM 强调输入特征中的高频特征的过程如下:

$$x^1 = GELU \left( C_{3 \times 3}^1 (X) \right) \quad (7)$$

$$x^i = GELU \left( C_{3 \times 3}^i (x^{i-1}) \right), i \in \{2, 3\} \quad (8)$$

其中,  $x^i$  和  $C_{3 \times 3}^i$  分别表示经卷积层放大后得到的第  $i$  种高频特征和对应的卷积层. 对于放大后得到的三种高频特征, 不同于常见的 SKFusion<sup>[13]</sup> 使用相加作为初始特征融合, FFFM 通过多分支通道注意力(multi-branch channel attention, MCA)来计算每种高频特征对应的注意力权重, 并对它们进行加权融合作为初始特征融合的结果.

$$m^i = C_{1 \times 1}^i \left( C_{1 \times 1} \left( GAP \left( \sum_{j=1}^3 x^j \right) \right) \right), i \in \{1, 2, 3\} \quad (9)$$

其中,  $m^i$  分别表示  $x^i$  对应的中间注意力权重, GAP 表示全局平均池化. 接着, 将三个中间注意力权重连接, 并用 Softmax 激活, 得到  $x^i$  对应的三个通道注意力权重  $w^i$ . MCA 采用  $1 \times 1$  卷积以避免破坏频率特征中的结构信息. 将每种高频特征通过对应的通道注意力权重加权后融合, 并通过  $1 \times 1$

卷积进行特征细化后得到初始特征融合  $X_{IFF}$ . 这种初始特征融合方式利用通道交互动态融合了多种高频特征, 并强调高频特征中的重要部分, 更好地恢复图像的边缘和细节等高频特征, 有效地提高了 FFFM 的表征能力和恢复的无雾图像的质量.

$$X_{IFF} = C_{1 \times 1} \left( \sum_{i=1}^3 w^i x^i \right) \quad (10)$$

为了方便起见, 后续的频率特征融合部分采用和初始特征融合相似的设计, 使用 MCA 来融合多种高频特征.

$$M^i = C_{1 \times 1}^i \left( C_{1 \times 1} \left( \text{GAP} \left( X_{IFF} \right) \right) \right), i \in \{1, 2, 3\} \quad (11)$$

其中,  $M^i$  指的是  $x^i$  对应的后续频率特征融合部分的中间注意力权重. 将中间注意力权重连接, 并通过 Softmax 激活后得到  $x^i$  对应的通道注意力权重  $W_i$ . 与初始特征融合类似, 将每种高频特征通过通道注意力权重动态加权融合, 并通过  $1 \times 1$  卷积进行特征细化后与输入特征  $X$  进行残差连接得到 FFFM 输出特征  $X_{FFFM}$ .

$$X_{FFFM} = C_{1 \times 1} \left( \sum_{i=1}^3 W^i x^i \right) + X \quad (12)$$

## 2.4 损失函数

为了促进双域的无雾图像恢复过程, DFFNet 在空间域与频率域都使用了  $L_1$  损失, 并且在空间域引入了使用 ResNet-152 作为固定预训练模型<sup>[14]</sup>的对比正则化<sup>[25]</sup>. 最终的损失函数由空间域损失  $L_s$  和频域损失  $L_f$  组合而成.

$$L_s = \|J - \hat{J}\|_1 + \alpha \sum_{i=1}^n \omega_i \cdot \frac{D(R_i(J), R_i(\hat{J}))}{D(R_i(I), R_i(\hat{J}))} \quad (13)$$

$$L_f = \|F(J) - F(\hat{J})\|_1 \quad (14)$$

$$L = L_s + \beta L_f \quad (15)$$

其中  $F$  代表快速傅里叶变换,  $\hat{J}$  代表恢复的无雾图像,  $R_i, i=1, 2, \dots, n$  是从 ResNet-152 中提取第  $i$  个隐藏特征,  $D(x, y)$  是  $x$  和  $y$  之间的  $L_1$  距离,  $\omega_i$  是权重, 超参数  $\alpha$  和  $\beta$  都设置为 0.1.

## 3 实验

### 3.1 实验设置

实现细节. DFFNet 在 4-Card NVIDIA GeForce RTX 3090 GPU 上使用 Pytorch 2.1.1 实现. 在训练期间, 图像被随机裁剪为  $256 \times 256$  的块. DFFNet 有两种规模的版本, -S、-L 分别表示小模型和大模型, 分别对应 2 层、8 层的 SFFM 的堆叠层数. DFFNet-S 足够轻量级, 可以应用于实际应用中的实时去雾场景, 而 DFFNet-L 可以应用于对去雾结果要求严格的任务. DFFNet 从固定预训练的 Resnet-152 中提取第 11、35、143、152 层的特征用于进行对比正则化, 它们对应的权重  $\omega_i$  依次为  $\left\{ \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, 1 \right\}$ . DFFNet 使用 AdamW 优化器 ( $\beta_1 = 0.9, \beta_2 = 0.999$ ) 进行训练, 初始学习率为  $2 \times 10^{-4}$ , 通过余弦退火策略逐渐降低至  $2 \times 10^{-6}$ . FLOPs 在  $256 \times 256$  的块上进行测量.

数据集. 实验设置遵循 MDNet<sup>[14]</sup>, 分别在 RESIDE-Indoor 训练集(13990 个图像对)和 Haze4K 训练集(3000 个图像对)上进行训练, 并在对应的 SOTS-Indoor 测试集(500 个图像对)和 Haze4K 测试集(1000 个图像对)上进行测试.

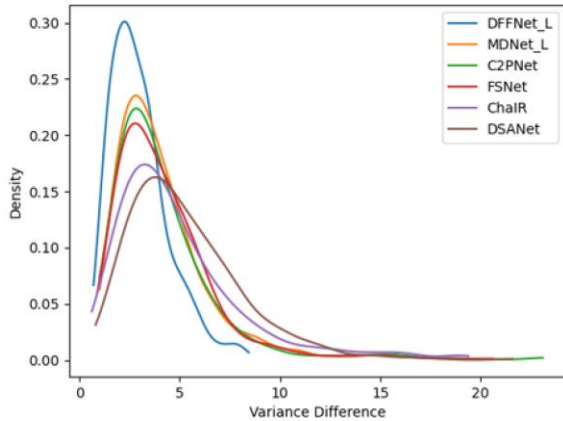
### 3.2 与最先进方法的对比

定量比较. 在定量比较中使用 PSNR 和结构相似性(structural similarity, SSIM)来评估去雾模型的性能, 使用参数量(parameters, Params)和 FLOPs 衡量模型的计算效率. 定量比较的结果如表 1 所示, 在 SOTS-Indoor 测试集上, DFFNet-L 领先于最先进的方法, 并且是第一个在 SOTS-Indoor 测试集上 PSNR 超过 43dB 的去雾模型, 超越最先进的方法 MDNet<sup>[14]</sup>1.21dB, SSIM 与 MDNet 持平. 同时, DFFNet 更加轻量级, Params 和 FLOPs 更小. 在 Haze4K 数据集上, DFFNet 以 MDNet 46% 的 Params 和 67% 的 FLOPs 取得了次优的结果, PSNR 和 SSIM 分别为 35.29dB 和 0.9916, 稍微落后于 MDNet 0.35dB 和 0.0004. 此外, 在两个数据集上分别测量了 DFFNet-L 以及最先进去雾模型恢复的无雾图像与数据集中的真实图像之间的方差差异, 结果如图 5 所示. 结果表明 DFFNet-L 在 SOTS-Indoor 数据集(a)上恢复的无雾图像最接近真实情况, 而在 Haze4K 数据集(b)上恢复效果仅次于最先进方法 MDNet-L. 定量比较的结果表明, DFFNet 在室内室外场景都具有出色的去雾性能,

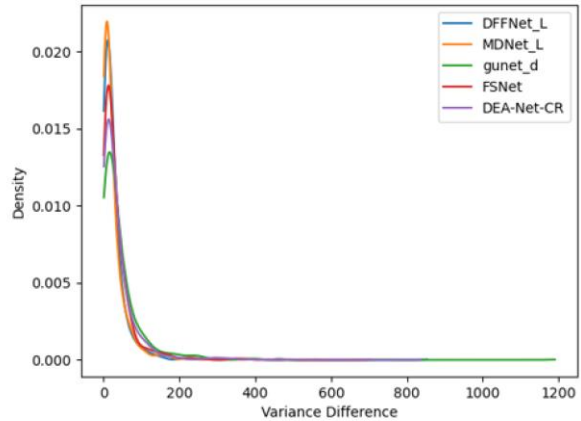
表 1 在 SOTS-Indoor 测试集和 Haze4K 测试集与最先进的方法进行定量比较

方法	刊物&年份	SOTS-Indoor		Haze4K		Params/M	FLOPs/G
		PSNR/dB	SSIM	PSNR/dB	SSIM		
DCP <sup>[4]</sup>	TPAMI10	16.62	0.818	14.01	0.760	-	-
DehazeNet <sup>[22]</sup>	TIP16	19.82	0.821	19.12	0.840	0.0009	0.581
AOD-Net <sup>[23]</sup>	ICCV17	19.06	0.850	17.15	0.83	0.002	0.115
GridDehazeNet <sup>[7]</sup>	ICCV19	32.16	0.984	23.29	0.93	0.956	21.49
MSBDN <sup>[26]</sup>	CVPR20	33.67	0.985	22.99	0.850	31.35	41.54
FFA-Net <sup>[8]</sup>	AAAI20	36.39	0.989	26.96	0.950	4.456	287.8
AECR-Net <sup>[25]</sup>	CVPR21	37.17	0.990	-	-	2.611	52.20
PMNet <sup>[9]</sup>	ECCV22	38.41	0.985	33.49	0.98	18.90	81.13
DehazeFormer-L <sup>[13]</sup>	TIP23	40.05	0.996	-	-	25.44	279.7
SANet <sup>[27]</sup>	IJCAI23	40.40	0.996	-	-	3.81	37.26
FocalNet <sup>[17]</sup>	ICCV23	40.82	0.996	-	-	3.74	30.63
IRNeXt <sup>[28]</sup>	ICML23	41.21	0.996	-	-	5.46	41.95
SFNet <sup>[29]</sup>	ICLR23	41.24	0.996	-	-	13.27	125.43
DEA-Net-CR <sup>[30]</sup>	TIP24	41.31	0.995	34.25	0.99	3.653	32.23
gUNet-D <sup>[31]</sup>	arXiv22	41.34	0.996	33.52	0.988	5.025	16.48
DSANet <sup>[32]</sup>	NN24	41.36	0.997	-	-	3.86	37.72
ChaIR <sup>[18]</sup>	KBS23	41.95	0.997	-	-	15.02	140.75
FSNet <sup>[24]</sup>	TPAMI23	42.45	0.997	34.12	0.99	13.28	111.14
C2PNet <sup>[33]</sup>	CVPR23	42.56	0.995	-	-	7.17	460.95
MixDehazeNet-L <sup>[14]</sup>	arXiv23	<u>42.62</u>	<u>0.997</u>	<b>35.64</b>	<b>0.992</b>	12.42	86.7
DFNet-S	Ours	41.59	0.996	32.84	0.9879	1.72	15.86
DFNet-L	Ours	<b>43.83</b>	<b>0.997</b>	<u>35.29</u>	<u>0.9916</u>	5.71	58.15

注. 加粗数值为最优结果, 下划线为次优结果.



(a) SOTS-Indoor 测试集



(b) Haze4K 测试集

图 5 SOTS-Indoor 测试集(a)和 Haze4K 测试集(b)上不同去雾模型预测的真实图像和数据集中真实图像之间的方差差异

并且具有优秀的参数量和计算效率.

定性比较. DFFNet 与 SOTS-Indoor 数据集、Haze4K 数据集上最先进的模型相比的视觉效果如图 6、图 7 所示. 第一行与第二行分别是重度有雾和轻度有雾的室内和室外图像. 由图可知, DFFNet 恢复的图像是最清晰、最接近真实情况的, 因为 DFFNet 具有大感受野, 可以更好地利用上下文信息进行去雾, 所以恢复的图像中的雾残留更少、颜色更均匀、视觉效果更好. 同时 DFFNet 利用清晰/退化图像对之间的频率差异来更好地恢复图像的高频信息, 使得恢复图像的边缘与细节更清晰. 图 6 和图 7 中的矩形框用于突出展示不同模型恢复的

无雾图像的差异. 同时, 图 8 展示了 DFFNet 与最先进的模型的去雾结果与无雾图像相减所得残差图像的离散傅里叶变换结果, 由图可知 DFFNet 恢复的无雾图像在频率域上最接近无雾图像.

### 3.3 消融实验

通过在 RESIDE-Indoor 训练集上训练 DFFNet-T 模型及其变体, 并在 SOTS-Indoor 数据集上进行测试来进行消融实验, 以此分析 DFFNet 中主要模块 SFFM 和 FFFM 的有效性. DFFNet-T 中每个尺度的 SFFM 堆叠层数  $N$  均为 1. DFFNet-T 训练轮次为 400, 初始学习率为  $4 \times 10^{-4}$ , 批量大小为 32. 基线(Baseline)通过将 MDNet<sup>[14]</sup>中





图6 SOTS-Indoor 数据集上的室内图像去雾结果

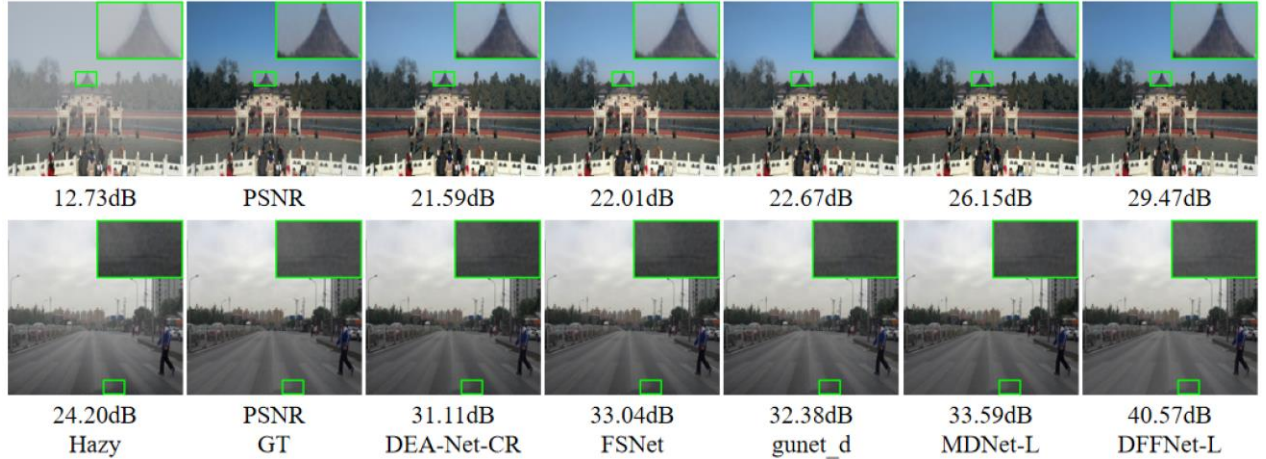


图7 Haze4K 数据集上的室外图像去雾结果

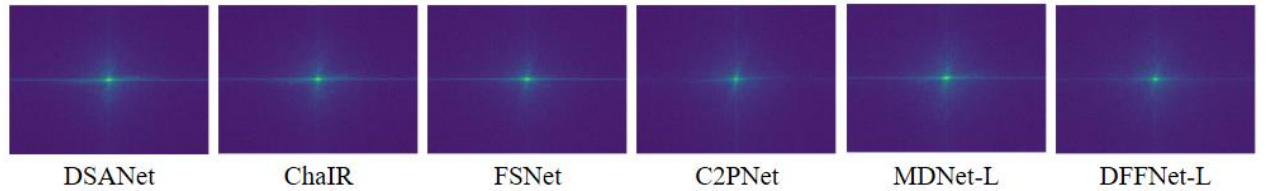


图8 不同模型的去雾结果与无雾图像相减所得残差图像的离散傅里叶变换结果

的 MSBlock 换成 VAN<sup>[19]</sup>获得. 不同组件的影响如表 2 所示. Baseline 的 PSNR 为 37.68dB. Baseline+SFFM(将 Baseline 中的 VAN 替换为 SFFM)和 Baseline+FFFM(向 Baseline 中加入 FFFM)的 PSNR 分别比 Baseline 提高了 0.64dB 和 1.37dB. 通过结合 SFFM 和 FFFM, DFFNet 相比 Baseline PSNR 提高了 2.31dB. 定量验证了 SFFM 和 FFFM 的有效性.

表2 SOTS-Indoor 数据集上 DFFNet 不同组件消融研究

方法	PSNR/dB	SSIM	Params/M	FLOPs/G
Baseline	37.68	0.9932	0.426	4.403
Baseline+SFFM	38.32	0.9938	0.747	7.716
Baseline+FFFM	39.05	0.9943	0.732	5.499
DFFNet	39.99	0.9951	1.052	8.812

DFFNet 中不同组件处理后特征的视觉效果如图 9 所示. 第一行与第二行分别是重度有雾和轻度有雾的图像, 第三列、第四列和第五列分别表示初始输入特征、经 SFFM 处理后的输入特征和经 SFFM 和 FFFM 处理后的输入特征. 观察图 9 可得, SFFM 主要使模型更加关注退化区域, 而 FFFM 通过放大高频特征并进行强调与融合以获得更明确的边缘和更精细的细节. 图 9 中的矩形框用于突出展示不同组件对特征的影响.

表 3 展示了 SFFM 与其他空间域特征融合模块的比较. 通过改进 Baseline, 将 SFFM 与三种流行的空间域特征融合模块进行对比, 来进一步证明 SFFM 的优越性. 如表 3 所示, 与 Resblock<sup>[18]</sup>和 VAN<sup>[19]</sup>相比, SFFM 的 PSNR 分别高出 2.50dB 和 0.64dB. 同时, 虽然 SFFM 的 PSNR 略低于目前最

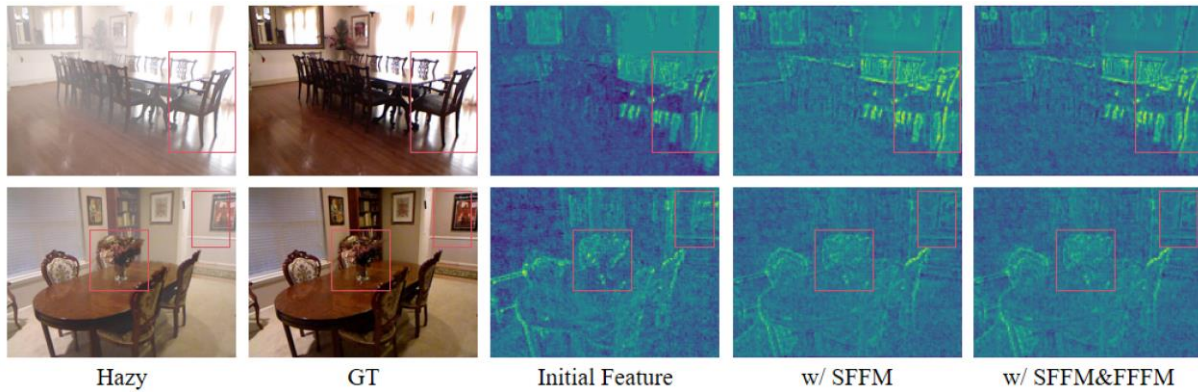


图9 SFFM 和 FFFM 处理特征后的视觉效果

先进方法 MDNet<sup>[14]</sup>中的 MSBlock, 但是 SFFM 的参数规模仅为 MSBlock 的 46%, FLOPs 仅为 53%, 具有更高的计算效率和更小的计算开销, SFFM 可以简单方便地替代其他去雾网络的空间域特征融合模块来提高去雾性能.

表3 SFFM 与其他空间域特征提取模块的比较

方法	PSNR/dB	SSIM	Params/M	FLOPs/G
ResBlock <sup>[18]</sup>	35.82	0.9913	0.766	7.421
VAN <sup>[19]</sup>	37.68	0.9932	0.426	4.403
MSBlock <sup>[14]</sup>	38.35	0.9941	1.623	14.621
SFFM	38.32	0.9938	0.747	7.716

表4展示了 FFFM 与其他频率域特征融合模块的比较. 通过改进 Baseline, 将 FFFM 与三种最新的频率域特征融合模块进行对比, 来进一步证明 FFFM 的优越性. 如表4所示, 与 FSNet<sup>[24]</sup>中的 MDSF+MCSF、FNet<sup>[17]</sup>中的 FSM 和 ChaIR<sup>[18]</sup>中的 FCA 相比, FFFM 的 PSNR 高出 1.05dB、0.35dB 和 0.28dB. 比较结果在展示频率域特征融合在去雾领域的有效性的同时, 也展示了 FFFM 的优越性.

表4 FFFM 与其他频率域特征融合模块的比较

方法	PSNR/dB	SSIM	Params/M	FLOPs/G
MDSF+MCSF <sup>[24]</sup>	38.00	0.9937	0.454	4.423
FSM <sup>[17]</sup>	38.70	0.9939	0.426	4.403
FCA <sup>[18]</sup>	38.77	0.9941	0.704	5.461
FFFM	39.05	0.9943	0.732	5.499

## 4 结 语

本文针对双域特征融合中存在的问题, 提出了一种用于图像去雾的双域特征融合网络 DFFNet.

设计 SFFM 以增强模型对特征的建模能力, 并使模型更好地关注图像中雾的分布. 设计 FFFM 以更好地恢复图像中的边缘和细节等高频特征. 在两个图像去雾基准数据集上对 DFFNet 的性能进行评估, 结果表明 DFFNet 与最先进的方法相当甚至更好. 同时, 通过消融实验验证了双域特征融合机制的有效性和通用性. DFFNet 的架构足够简单和轻量级, 可以方便地扩展至其他视觉任务.

不同的双域特征融合方式会显著影响恢复的无雾图像的质量, 未来将继续对空间域和频率域的特征融合方式进行详细探讨, 充分利用清晰/退化图像对在空间域和频率域的差异, 发挥提升去雾性能的潜力. 其次, 针对复杂多样的雾霾天气, 将考虑进一步设计针对场景的先验, 利用场景信息来应对更复杂的去雾问题. 最后, 尝试将这项工作扩展至视频去雾上, 结合多帧的信息来辅助模型更好地去雾.

## 参考文献(References):

- [1] McCartney E J. Optics of the atmosphere: scattering by molecules and particles[J]. New York, 1976
- [2] Nayar S K, Narasimhan S G. Vision in bad weather[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 1999, 2: 820-827
- [3] Narasimhan S G, Nayar S K. Vision and the atmosphere[J]. International Journal of Computer Vision, 2002, 48: 233-254
- [4] He K M, Sun J, Tang X O. Single image haze removal using dark channel prior[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(12): 2341-2353
- [5] Zhu Q S, Mai J M, Shao L. A fast single image haze removal algorithm using color attenuation prior[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3522-3533
- [6] Berman D, Avidan S. Non-local image dehazing[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 1674-1682
- [7] Liu X H, Ma Y R, Shi Z H, et al. Griddehazenet: Attention-based multi-scale network for image dehazing[C] //Proceedings of the IEEE/CVF International Conference on

- Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 7314-7323
- [8] Qin X, Wang Z L, Bai Y C, *et al.* FFA-Net: Feature fusion attention network for single image dehazing[C] //Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020, 34(07): 11908-11915
- [9] Ye T, Jiang M C, Zhang Y C, *et al.* Perceiving and modeling density is all you need for image dehazing[OL]. [2021-11-18]. <https://arxiv.org/abs/2111.09733>
- [10] Valanarasu J M J, Yasarla R, Patel V M. Transweather: Transformer-based restoration of images degraded by adverse weather conditions[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 2353-2363
- [11] Wang Z D, Cun X D, Bao J M, *et al.* Uformer: A general u-shaped transformer for image restoration[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 17683-17693
- [12] Guo C L, Yan Q X, Anwar S, *et al.* Image dehazing transformer with transmission-aware 3d position embedding[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 5812-5820
- [13] Song Y D, He Z Q, Qian H, *et al.* Vision transformers for single image dehazing[J]. IEEE Transactions on Image Processing, 2023, 32: 1927-1941
- [14] Lu L P, Xiong Q, Chu D F, *et al.* MixDehazeNet: Mix Structure Block For Image Dehazing Network[OL]. [2023-05-18]. <https://arxiv.org/abs/2305.17654>
- [15] 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. Los Alamitos: IEEE Computer Society Press, 2021: 4196-4205
- [16] Guo S, Yong H W, Zhang X D, *et al.* Spatial-Frequency Attention for Image Denoising[OL]. [2023-02-27]. <https://arxiv.org/abs/2302.13598>
- [17] Cui Y N, Ren W Q, Cao X C, *et al.* Focal network for image restoration[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2023: 13001-13011
- [18] Cui Y N, Knoll A. Exploring the potential of channel interactions for image restoration[J]. Knowledge-Based Systems, 2023, 282: 111156
- [19] Guo M H, Lu C Z, Liu Z N, *et al.* Visual attention network[J]. Computational Visual Media, 2023, 9(4): 733-752
- [20] Wang W H, Xie E Z, Li X, *et al.* Pvt v2: Improved baselines with pyramid vision transformer[J]. Computational Visual Media, 2022, 8(3): 415-424
- [21] Dai Y M, Gieseke F, Oehmcke S, *et al.* Attentional feature fusion[C] //Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 3560-3569
- [22] Cai B L, Xu X M, Jia K, *et al.* Dehazenet: An end-to-end system for single image haze removal[J]. IEEE Transactions on Image Processing, 2016, 25(11): 5187-5198
- [23] Li B Y, Peng X L, Wang Z Y, *et al.* Aod-net: All-in-one dehazing network[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 4770-4778
- [24] Cui Y N, Ren W Q, Cao X C, *et al.* Image Restoration via Frequency Selection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023
- [25] Wu H Y, Qu Y Y, Lin S H, *et al.* Contrastive learning for compact single image dehazing[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 10551-10560
- [26] Dong H, Pan J S, Xiang L, *et al.* Multi-scale boosted dehazing network with dense feature fusion[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 2157-2167
- [27] Cui Y N, Tao Y, Jing L X, *et al.* Strip attention for image restoration[C] //Proceedings of the 32th International Joint Conference on Artificial Intelligence. Freiburg: IJCAI, 2023, 2
- [28] Cui Y N, Ren W Q, Yang S N, *et al.* IRNeXt: Rethinking Convolutional Network Design for Image Restoration[C] //Proceedings of the 40th International Conference on Machine Learning. New York: ICML, 2023
- [29] Cui Y N, Tao Y, Bing Z S, *et al.* Selective Frequency Network for Image Restoration[C] //Proceedings of the 11th International Conference on Learning Representations. Washington DC: ICLR, 2022
- [30] Chen Z X, He Z W, Lu Z M. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention[J]. IEEE Transactions on Image Processing, 2024
- [31] Song Y D, Zhou Y, Qian H, *et al.* Rethinking performance gains in image dehazing networks[OL]. [2022-09-23]. <https://arxiv.org/abs/2209.11448>
- [32] Cui Y N, Knoll A. Dual-domain strip attention for image restoration[J]. Neural Networks, 2024, 171: 429-439
- [33] Zheng Y, Zhan J H, He S F, *et al.* Curricular contrastive regularization for physics-aware single image dehazing[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 5785-5794