中图法分类号:TP391　　**文献标识码**:A　　**文章编号**:1006-8961(2025)11-3665-15

论文引用格式:Wang W J, Chen F, Liu W L, Cheng H and Wang M Q. 2025. Image dehazing network based on dual-domain feature fusion. Journal of Image and Graphics, 30(11):3665-3679(王炜嘉, 陈飞, 刘莞玲, 程航, 王美清. 2025. 基于双域特征融合的图像去雾网络. 中国图象图形学报, 30(11):3665-3679)[DOI:10.11834/jig.240655]

基于双域特征融合的图像去雾网络

王炜嘉[1],陈飞[1*],刘莞玲[1,2],程航[3],王美清[3]

1. 福州大学计算机与大数据学院,福州　350108; 2. 天津大学智能与计算学部,天津　300350;
3. 福州大学数学与统计学院,福州　350108

摘　要:**目的**　图像去雾旨在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3个主要问题:空间域特征提取与融合存在局限性、频率域特征融合效果不佳以及未能实现频空双域特征的高效融合。针对这些问题,提出专注于频空双域特征融合的双域特征融合网络(dual-domain feature fusion network,DFFNet)。**方法**　首先,设计更适合图像软重建的空间域特征融合模块(spatial-domain feature fusion module,SFFM),采用Transformer风格架构,通过大核注意力机制捕获全局特征并定位有雾区域,像素注意力机制建模局部特征并恢复边缘和细节,共同模拟多头自注意力机制,满足软重建需求。同时,提出频率域特征融合模块(frequency-domain feature fusion module,FFFM)。该模块采用隐式方法处理高频信息,通过多个卷积层增强高频分量、多分支通道注意力实现频率高效融合,并放置于网络瓶颈处实现频空双域特征高效融合。**结果**　结合这两种关键模块设计提出的DFFNet在两个基准数据集上展现出超越目前先进方法的性能表现。DFFNet-L是第1个在室内合成目标测试集(synthetic objective testing set-indoor,SOTS-Indoor)上峰值信噪比(peak signal-to-noise ratio,PSNR)超过43 dB以及第1个在Haze4K数据集上PSNR超过36 dB的去雾网络,PSNR分别为43.83 dB和36.39 dB,分别领先领域先进方法MixDehazeNet-L 1.21 dB和0.45 dB。并且DFFNet更加轻量级,参数量仅为MixDehazeNet-L的46.0%,浮点运算次数仅为其67.1%,同时,由于DFFNet的主要模块SFFM和FFFM具有良好的可迁移性和扩展性,这使得它们能够便捷地迁移到其他计算机视觉任务中,为提升模型性能提供新的解决方案。**结论**　本文所提出的双域特征融合网络,综合了卷积神经网络模型和Transformer模型的优点,有效解决了双域特征融合存在的问题,取得了卓越的去雾效果。代码发布于https://github.com/WWJ0720/DFFNet。

关键词:计算机视觉;图像去雾;双域特征融合;空间域特征融合;频率域特征融合;注意力机制;深度学习

Image dehazing network based on dual-domain feature fusion

Wang Weijia[1], Chen Fei[1*], Liu Wanling[1,2], Cheng Hang[3], Wang Meiqing[3]

1. College of Computer and Data Science, Fuzhou University, Fuzhou　350108, China; 2. College of Intelligence and Computing, Tianjin University, Tianjin　300350, China; 3. School of Mathematics and Statistics, Fuzhou University, Fuzhou　350108, China

Abstract: **Objective** Image dehazing is a crucial task in computer vision and image processing. This task aims to recover potential haze-free images from those degraded by atmospheric scattering phenomena such as haze and fog. Image dehazing research holds substantial practical importance and application value. It enhances the performance and reliability of visual systems under adverse weather conditions, which provides technical support for critical applications such as autonomous

收稿日期:2024-11-06;**修回日期**:2025-04-08;**预印本日期**:2025-04-15

*通信作者:陈飞　chenfei314@fzu.edu.cn

基金项目:国家自然科学基金项目(62471141,62172098);中国烟草总公司福建省公司科技计划项目(2022350000240056)

Supported by: National Natural Science Foundation of China (62471141, 62172098)

driving, intelligent surveillance, and remote sensing image analysis. It also serves as a preprocessing step to improve the performance of advanced computer vision tasks including object detection, image classification, and scene understanding. Simultaneously, image dehazing technology has driven the development of atmospheric scattering models and image restoration theory, which promotes innovative applications of deep learning in low-level computer vision tasks. It has great value for improving the effectiveness of public safety, environmental monitoring, and intelligent transportation systems, which serves as a crucial supporting technology for achieving all-weather, all-scenario visual perception capabilities. Existing methods utilize differential features of clear-degraded image pairs in spatial and frequency domains to achieve dehazing with certain effectiveness. However, three major problems remain: limitations in spatial domain feature extraction and fusion, poor performance in frequency domain feature fusion, and inefficient integration of features from both domains. To address these issues, we propose the dual-domain feature fusion network (DFFNet), which focuses on the fusion of features from spatial and frequency domains. **Method** DFFNet is based on soft reconstruction (SR) to recover potential haze-free images and consists of two key modules. First, we design a spatial-domain feature fusion module (SFFM) more suitable for image soft reconstruction. This module adopts a Transformer-style architecture that captures global features through large-kernel attention mechanisms to accurately locate hazy regions in images. Meanwhile, it models local features through pixel attention mechanisms to effectively restore image edges and details. The two attention mechanisms jointly satisfy the dual requirements for global and local features in the image soft reconstruction process. Specifically, these features are mapped and fused using a convolutional feed-forward network. The SFFM maintains a relatively small parameter scale and low computational complexity. Meanwhile, according to the spectral convolution theorem, visual feature processing in the spatial domain and frequency domain is essentially equivalent——convolution operations in the spatial domain are equivalent to multiplication operations in the frequency domain. Therefore, we choose to emphasize high-frequency components through convolution. We propose the frequency-domain feature fusion module (FFFM), which adopts an implicit method to process high-frequency information without requiring explicit Fourier transforms. By stacking multiple convolutions as high-pass filters, the module enhances high-frequency components in the input features and enriches their diversity. As a result, the capability of the model to restore high-frequency details in images is improved. Increasing the number of stacked convolutional layers can extract richer high-frequency features, which enhances the quality of the output features of the FFFM module. However, considering parameter scale and computational efficiency, we ultimately choose to stack three convolutional layers for extracting three different types of high-frequency features. To better capture contextual information while optimizing computational overhead, the FFFM module is only applied between the encoder and decoder at the third scale. The output features of the first and second scale encoders are adaptively fused with the upsampled output features of the second and third scale decoders through SKFusion, respectively. The output features of the first scale decoder, after patch unembed, are processed together with the input image through the SR module to recover the potential haze-free image. In DFFNet, the fusion of spatial and frequency domain features occurs in the FFFM module located at the network bottleneck. The features at the bottleneck have already extracted rich multi-scale spatial information through multiple SFFM layers, which possess strong semantic expression capabilities. Simultaneously, as the connection point between the encoder and decoder, the features fused at this position can directly influence the subsequent recovery process. These fused features allow high-frequency information to effectively guide the entire decoding process, which enhances restoration of image edges and details. **Result** The DFFNet, which combines the two key module designs, demonstrates performance exceeding those of current state-of-the-art methods on two benchmark datasets. DFFNet-L is the first dehazing network to achieve peak signal-to-noise ratio values exceeding 43 dB on the SOTS-Indoor test set and 36 dB on the Haze4K dataset. The specific values are 43.83 dB and 36.39 dB, which outperform those of the current state-of-the-art image dehazing method MixDehazeNet-L by 1.21 dB and 0.45 dB, respectively. Furthermore, DFFNet is more lightweight, with only 46.0% of the parameters and 67.1% of the floating point operations compared with MixDehazeNet-L. In the visualization experiments, DFFNet-L demonstrates the highest quality of haze-free image restoration. This superior performance is primarily attributed to its large receptive field design, which enables the model to fully capture and utilize global contextual information in images for precise dehazing. This advantage allows DFFNet-L to accurately identify the spatial distribution characteristics of haze in scenes, which achieves thorough haze removal, uniform and natural color restoration, and improved overall visual effects

in the recovered images. Simultaneously, DFFNet-L leverages the frequency domain differential features between clear and degraded image pairs, which significantly enhances its capability to restore high-frequency components of images. This improved capability results in sharper edge contours and clearer textural details in the recovered images, which enhance the accuracy of detail restoration and make the results highly closely approximate the haze-free ground truth images. Moreover, the main modules of DFFNet, SFFM, and FFFM possess good transferability and scalability, which allow them to be conveniently transferred to other computer vision tasks. As a result, new solutions for improving model performance are provided. **Conclusion** This study proposes a dual-domain feature fusion network that combines the advantages of convolutional neural network models and Transformer models, which effectively addresses the existing problems in dual-domain feature fusion and achieves excellent dehazing results. The code is available at https://github.com/WWJ0720/DFFNet.

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

0 引　言

在有雾场景中拍摄的图像通常会出现对比度降低和色彩失真等视觉质量退化，这些退化不仅影响人眼感知，而且严重制约了目标检测、图像分割及场景理解等高级计算机视觉任务的性能。因此，作为图像恢复领域的关键研究方向之一，图像去雾任务受到学术界和工业界的持续关注，可用于工业中的烟草烘烤图像去雾，并在监控系统、自动驾驶和遥感图像处理等应用领域具有重要价值，对提升恶劣天气条件下视觉系统的可靠性具有重要意义（邢晓敏和刘威，2016）。

图像去雾的目标是从有雾图像中恢复潜在的无雾图像。在图像去雾研究中，通常采用大气散射模型（atmospheric scattering model，ASM）（Narasimhan和Nayar，2002）对图像的退化过程进行建模，ASM的表达式为

$$\boldsymbol{I}(x)=\boldsymbol{J}(x)t(x)+\boldsymbol{A}\big(1-t(x)\big) \tag{1}$$

式中，$\boldsymbol{I}$、$\boldsymbol{J}$、t、$\boldsymbol{A}$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

$$t(x)=\mathrm{e}^{-\beta d(x)} \tag{2}$$

式中，β是大气的散射系数，d是场景深度。

图像去雾方法主要分为基于先验知识的方法和基于深度学习的方法。基于先验知识的方法依赖大气散射模型，通过手工设计的先验辅助去雾。He等人（2011）提出暗通道先验（dark channel prior，DCP），利用“无雾图像中至少一个颜色通道像素值极低，称为暗通道；而在有雾图像中暗通道由于大气散射的影响，值会增大”这一现象作为先验知识，通过估计大气光和透射率来恢复无雾图像。Zhu等人（2015）提出颜色衰减先验（color attenuation prior，CAP），利用“随着场景深度增加，物体的饱和度会降低而亮度会增加”这一现象作为先验知识，通过估计场景深度和大气光来恢复无雾图像。Berman等人（2016）提出非局部先验（non-local prior，NLP），利用“即使相隔较远的像素也可能属于同一颜色聚类，而这些相似颜色会因透射率差异在色彩空间中沿雾线延伸”这一现象作为先验知识来辅助去雾。然而，这些基于经验统计的先验在不符合假设的场景中表现欠佳，例如DCP在天空等无暗通道区域和纹理单一区域表现不佳，甚至会导致恢复的图像出现严重色彩失真。

基于深度学习的方法依赖数据集来恢复潜在的无雾图像。早期基于深度学习的方法通过深度神经网络估计ASM中的A和t或其变形公式中的参数，以恢复潜在的无雾图像。Cai等人（2016）提出DehazeNet，通过学习有雾图像与透射率之间的映射关系，精确拟合透射率并结合ASM来恢复无雾图像。Li等人（2017）提出多合一去雾网络（all-in-one dehazing network，AOD-Net），基于重新设计的ASM，在保留物理模型基础的同时，实现在单一网络中对变形公式中的唯一参数K进行估计。

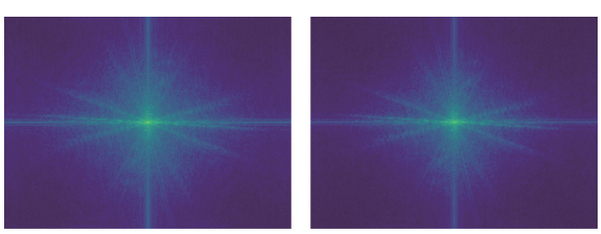
随着真实单图像去雾（realistic single image dehazing，RESIDE）等大规模有雾图像数据集的出现，近年来去雾方法倾向于直接恢复无雾图像。Liu等人（2019）提出Grid-DehazeNet，通过基于注意力机制驱动的多尺度估计来缓解瓶颈问题，并通过后处理模块减少伪影，以提高去雾质量。Qin等人（2020）提出特征融合注意力网络（feature fusion attention network，FFA-Net），该网络采用的特征注意

力模块分别使用通道注意力和像素注意力来处理不同类型信息，并通过注意力引导的多层级特征融合结构，强化关键特征权重。

考虑到Transformer具有优秀的长距离建模能力和全局感受野，研究者尝试设计基于Transformer的去雾模型。Valanarasu等人（2022）提出TransWeather，采用内部补丁Transformer编码器和具有可学习的天气类型嵌入的解码器来有效消除各类天气退化。Wang等人（2022b）提出Uformer，通过构建局部增强窗口Transformer块来优化局部特征捕获，并引入可学习的多尺度恢复调制器，在降低参数量和计算成本的同时，增强了模型捕获局部与全局依赖的能力。但是，基于纯Transformer的模型仍然难以训练，因此研究者开始尝试结合Transformer和卷积神经网络的优点来设计高效去雾模型。Guo等人（2022）提出DeHamer，通过学习调制矩阵在Transformer特征条件下调制卷积神经网络特征，成功融合了Transformer的全局上下文建模能力和卷积神经网络的局部表示能力。Song等人（2023）提出DehazeFormer，修改了Swin Transformer中不适合图像去雾的关键设计，提出轻量级且高效的改进架构，在保持性能的同时显著降低了计算复杂度。Lu等人（2024）提出MixDehazeNet，设计了一个Transformer式的混合结构块，通过多尺度并行大卷积核模块来引入大感受野和多尺度特征，并通过增强型并行注意模块处理不均匀雾气分布。

同时，在图像去雾领域，除了利用无雾图像与有雾图像在空间域的差异特征外，两类图像在频率域的特征差异同样具有重要价值，能够为空间域特征融合结果提供关键的补充信息。图1（a）（b）分别展示了SOTS-Indoor测试集（Li等，2019）中无雾图像与对应的重度有雾、轻度有雾图像相减所得残差图像经离散傅立叶变换后的结果。可以观察到，无雾/有雾图像对在频率域的差异主要分布在低频分量，少量分布在高频分量。高频特征反映了图像的边缘和细节信息，低频特征则体现了图像的整体结构和形态特征。

现有的图像恢复工作采用小波变换（Chen等，2021）、傅里叶变换（Guo等，2023b）、池化操作（Cui等，2023a）和卷积（Cui等，2024b）等变换工具，将输入特征分解成不同频率分量，并对各分量分别处理后重建输出特征。在图像去雾任务中，基于频率处理的网络普遍采用隐式方法处理高频信息，无需显式的傅里叶变换。Cui等人（2023a）提出FocalNet，设计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。Cui等人（2024b）提出频率选择网络（frequency selection network，FSNet），设计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，并使用选择性核融合（selective kernel fusion，SKFusion）和加法来重建输出特征。Cui和Knoll（2023）提出ChaIR（channel interaction for image restoration），引入了隐式频域通道注意模块，该模块使用卷积放大高频特征并使用SKFusion融合高频特征以重建输出特征。



(a) 重度有雾　　(b) 轻度有雾

图1　残差图像经离散傅里叶变换后的结果

Fig. 1　The results of the residual images after discrete Fourier transform ((a) heavy fog; (b) light fog)

然而，频空双域特征融合机制存在3个主要的问题：1）空间域特征的提取与融合存在局限性。传统空间域特征提取模块受限于有限的感受野，在处理大范围雾气分布时表现不佳。此外，上下文信息的不充分利用导致模型难以准确把握图像的整体结构和局部细节。更为关键的是，由于缺乏有效的多语义特征的融合机制，基于ASM的软重建（soft reconstruction，SR）（Song等，2023）的性能受到限制，从而降低了模型的去雾效果。2）频率域特征融合效果不佳。目前主流的隐式频率域特征融合模块广泛采用SKFusion（Song等，2023）用于特征融合，但该方法存在两个显著缺陷：其初始特征融合方式仅采用简单的加法运算以及仅能实现两种频率特征的融合。这些局限性直接影响了频率域特征融合模块的性能——初始特征融合的质量决定了重建特征的质量，而频率特征融合数量的限制也显著制约了重建特征的表现。3）未能实现频空双域特征的高效融

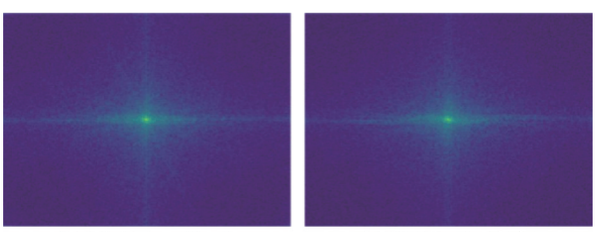
合。现有基于频率的去雾网络的频空融合效率有待提高，并且融合效果不够理想。

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）。该网络基于软重建来恢复潜在的无雾图像，并结合两个关键模块设计：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 模块采用 Transformer 风格设计，通过对视觉注意力网络（visual attention network，VAN）（Guo 等，2023a）进行重新设计使其更适用于图像去雾任务。该模块引入了像素注意力（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，并利用卷积前馈网络（convolutional feed-forward network，CFFN）（Wang 等，2022a）实现特征的映射与融合。

图 2（a）（b）分别展示了先进方法 MixDehazeNet-L 以及 DFFNet-L（L 表示大模型）在去雾处理后，得到的去雾图像与无雾图像相减所得残差图像经离散傅里叶变换后的结果。

MixDehazeNet 的模块主要基于空间域操作设计，在高频和低频信息的恢复方面展现出良好的效果，但仍与无雾图像存在一定的差异，特别表现在高频区域。根据相关研究（Bai 等，2022；Park 和 Kim，2022）表明，多头自注意力机制（multi-head self-attention mechanism，MHSA）在处理视觉数据时表现出独特的特征学习偏好：擅长捕获低频表征信息，但在提取高频特征方面存在局限性。与之形成鲜明对比的是，卷积操作恰好展现出互补的特性，即在高频特征的学习上具有天然优势，能够以隐式的方式放大频率域中的高频分量。所以 DFFNet 中的 FFFM 模块专注于针对高频信息进行恢复，通过卷积隐式恢复高频特征，与 Transformer 风格架构的 SFFM 可以形成有效的互补，充分发挥频空双域的优势来提高恢复的无雾图像的质量。FFFM 模块的设计受 ChaIR（Cui 和 Knoll，2023）和 AFF（attentional feature fusion）（Dai 等，2021）的启发，采用隐式方法处理高频信息，通过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，丰富高频特征的多样性，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，从而更有效地恢复图像的高频分量。并且，FFFM 仅放置于第 3 层编码器与解码器之间的网络瓶颈处，实现了频域与空域特征的高效融合。

结合这两种关键模块设计提出的 DFFNet 在两个基准数据集上展现出超越目前先进方法的性能表现。DFFNet-L 是第 1 个在 SOTS-Indoor 测试集上峰值信噪比（peak signal-to-noise ratio，PSNR）超过 43 dB 以及第 1 个在 Haze4K 数据集上 PSNR 超过 36 dB 的去雾网络，PSNR 分别为 43.83 dB 和 36.39 dB，分别领先领域先进方法 MixDehazeNet-L 1.21 dB 和 0.45 dB。并且 DFFNet 更加轻量级，参数量仅为 MixDehazeNet-L 的 46.0%，浮点运算次数仅为其 67.1%。同时，由于 DFFNet 的主要模块 SFFM 和 FFFM 具有良好的可迁移性和扩展性，使得它们能够便捷地迁移到其他计算机视觉任务中，为提升模型性能提供新的解决方案。



(a) MixDehazeNet-L　　(b) DFFNet-L (本文)

图 2　不同模型残差图像经离散傅里叶变换后的结果

Fig. 2　Results of different residual images after discrete Fourier transform ((a) MixDehazeNet-L; (b) DFFNet-L)

1　本文方法

1.1　DFFNet 整体网络架构

DFFNet 的整体网络架构如图 3（a）所示。图中浅蓝和深蓝箭头分别表示块嵌入和下采样，浅橙和深橙箭头分别表示块还原和上采样。DFFNet 采用与 UNet（Ronneberger 等，2015）类似的编码器—解码器架构，由 3 个尺度的编码器和解码器构成，每个尺度堆叠 N 个 SFFM 模块。为了在优化计算开销的同时更好地捕获上下文信息，FFFM 模块仅应用于第三尺度的编码器与解码器之间。SFFM 和 FFFM 的模块结构如图 3（b）（e）所示。空间域特征融合模块（SFFM）主要由像素注意力（PA）、大核注意力（LKA）和卷积前馈网络（CFFN）3 个关键组件组成。频率域特征融合模块（FFFM）使用卷积层来提取多种高频分量，并使用多分支通道注意力（multi-branch chan-

nel attention, MCA)来对高频分量进行自适应融合。根据谱卷积定理(Katznelson, 2004),在空间域和频率域中进行的视觉特征处理本质上是等价的——空间域的卷积运算等价于频率域的乘积运算。因此,FFFM采用多个卷积层来学习并融合高频特征,与Transformer风格架构的SFFM可以形成有效的互补。在DFFNet中,空间域特征和频率域特征的融合发生在网络的瓶颈位置。瓶颈处的特征已经通过多层SFFM提取了丰富的多尺度空间信息,具有较强的语义表达能力。同时,瓶颈处作为编码器和解码器的连接点,在此处融合后的特征可以直接影响后续的恢复过程,使得高频信息能够有效地指导整个解码过程,从而更好地恢复图像的边缘和细节。

在空间域特征融合方面,采用核大小为3、步长为1的标准卷积实现块嵌入,采用核大小为3、步长为2的步进卷积实现下采样。同时,采用由核大小为1、步长为1的卷积和像素重排组成的子像素卷积实现上采样,核大小为3、步长为1的标准卷积实现块还原。第一、第二尺度编码器的输出特征分别与上采样后的第二、第三尺度解码器的输出特征通过SKFusion自适应融合,第一尺度解码器的输出特征经过块还原后,与输入图像一起通过SR模块恢复潜在的无雾图像。SR模块的设计受DehazeFormer的启发,将式(1)重写为

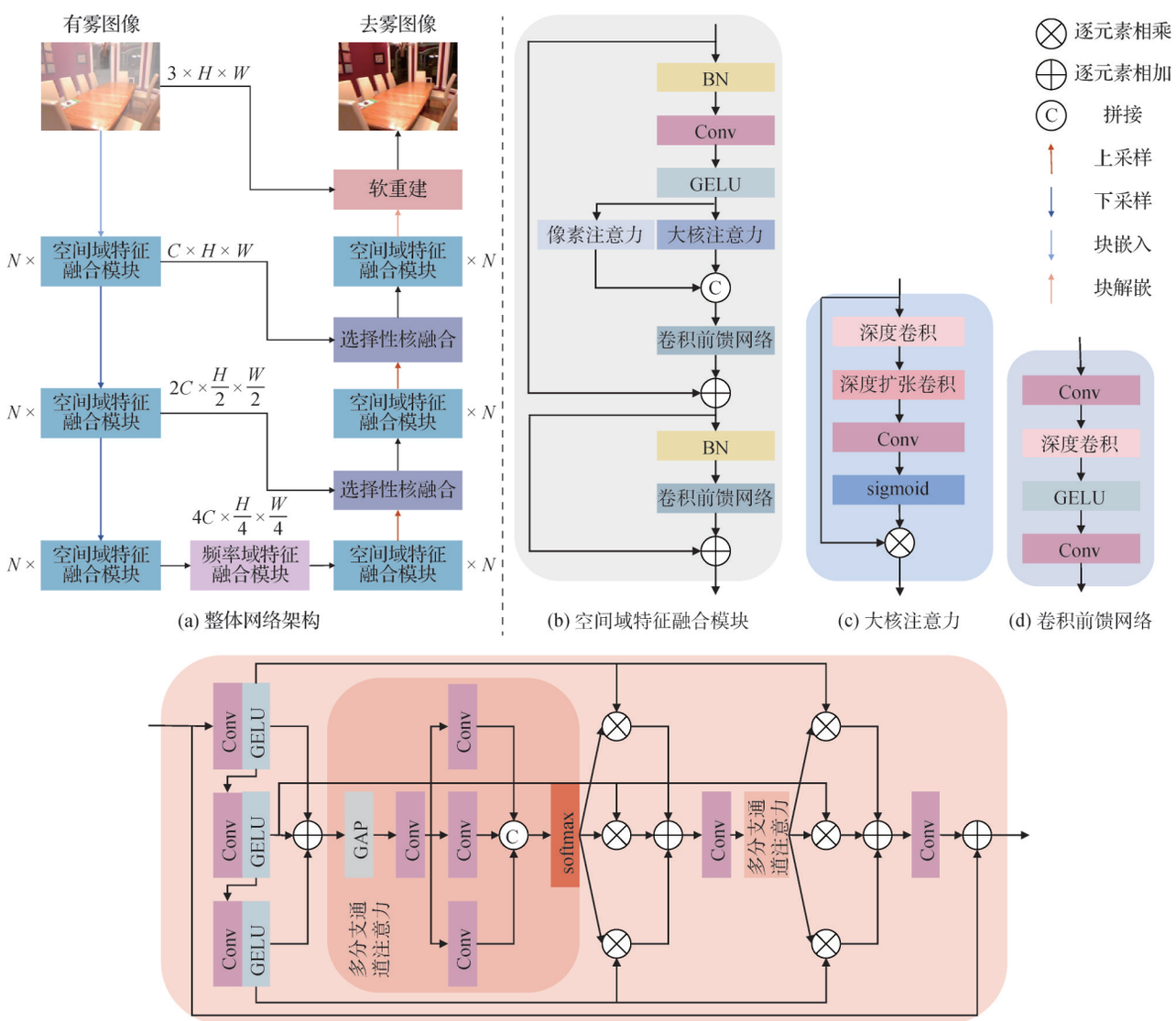


图3 DFFNet的整体网络框架与模块设计

Fig. 3 The overall network architecture and module design of DFFNet

((a) overall architecture; (b) SFFM; (c) LKA; (d) CFFN; (e) FFFM)

$$\boldsymbol{J}(x)=K(x)\boldsymbol{I}(x)+\boldsymbol{B}(x)+\boldsymbol{I}(x) \tag{3}$$

式中，$K(x)=1/t(x)-1$，$\boldsymbol{B}(x)=-\left(1/t(x)-1\right)\boldsymbol{A}$。SR软约束了$K(x)$和$\boldsymbol{B}(x)$的关系，将第一尺度解码器的输出特征$\boldsymbol{D}_1\in\mathbf{R}^{4\times H\times W}$分解为$\boldsymbol{K}\in\mathbf{R}^{1\times H\times W}$和$\boldsymbol{B}\in\mathbf{R}^{3\times H\times W}$。当$K(x)=0$时，SR通过预测全局残差来恢复无雾图像。SR借助基于ASM的弱先验知识，更有效地软约束了图像的结构和特征，从而提高无雾图像的恢复质量。

1.2　空间域特征融合

SFFM模块受VAN的启发，采用Transformer风格架构，并由两个核心组件构成：结合LKA和PA来模拟的多头自注意力机制以及使用CFFN作为多层感知机的前馈神经网络。SFFM模块在拥有大感受野与更适合图像去雾软重建的设计的同时，还保持着较小的参数规模和较低的计算复杂度。

1.2.1　多头自注意力机制

具体而言，假设输入特征为$\boldsymbol{X}\in\mathbf{R}^{C\times H\times W}$。为了确保训练的稳定性并增强模型的表征能力，在输入处理阶段，首先对$\boldsymbol{X}$进行批归一化（batch normalization，BN）处理，随后采用1×1卷积进行特征变换，并经过高斯误差线性单元（Gaussian error linear unit，GELU）激活后，最终得到中间特征表示$\boldsymbol{X}'$。接着，巧妙地结合了LKA和PA两种注意力机制，分别对$\boldsymbol{X}'$中的全局特征和局部特征进行选择性关注。LKA通过将大核卷积分解成深度卷积、深度扩张卷积和1×1卷积的方式，在引入大感受野的同时显著降低了计算成本。这样的大感受野设计使得模型能够更充分地利用上下文信息来建模$\boldsymbol{X}'$中的全局特征，并关注图像中的有雾区域分布，能够更好地恢复大片有雾区域。而PA通过在$\boldsymbol{X}'$中的每个像素位置上独立应用注意力机制，更好地对$\boldsymbol{X}'$中位置相关的局部特征进行建模，从而有效地恢复图像的边缘和细节。这两种注意力机制所提取的特征具有很强的互补性，共同满足了软重建过程中对全局特征和局部特征的双重需求，从而实现了更优质的图像恢复效果。LKA和PA可表示为

$$\boldsymbol{X}'_{\mathrm{LKA}}=\sigma\left(C_{1\times1}\left(DWDC_{7\times7}\left(DWC_{5\times5}(\boldsymbol{X}')\right)\right)\right)\otimes\boldsymbol{X}' \tag{4}$$

$$\boldsymbol{X}'_{\mathrm{PA}}=\sigma\left(C_{1\times1}\left(GELU\left(C_{1\times1}(\boldsymbol{X}')\right)\right)\right)\otimes\boldsymbol{X}' \tag{5}$$

式中，$DWC_{5\times5}$代表5×5深度卷积，$DWDC_{7\times7}$代表7×7深度扩张卷积，$C_{1\times1}$代表1×1的标准卷积，σ代表sigmoid函数。将经LKA和PA增强后的输出特征$\boldsymbol{X}'_{\mathrm{LKA}}$和$\boldsymbol{X}'_{\mathrm{PA}}$拼接起来，并将其输入CFFN进行处理。考虑到不同通道可能存在不同的退化模式，CFFN中的深度卷积能够实现有效的通道分离变换。在SFFM模块的多头自注意力机制中，使用CFFN不仅可以对不同通道的特征进行融合与映射以更好地拟合软重建所需的特征，同时还能增强SFFM模块的特征表征能力。CFFN可表示为

$$\boldsymbol{X}_{\mathrm{CFFN}}=C_{1\times1}\left(GELU\left(DWC_{3\times3}\left(C_{1\times1}\left(\left[\boldsymbol{X}'_{\mathrm{LKA}},\boldsymbol{X}'_{\mathrm{PA}}\right]\right)\right)\right)\right) \tag{6}$$

式中，$[\cdot,\cdot]$表示特征在通道维度上的拼接操作。将CFFN的输出特征$\boldsymbol{X}_{\mathrm{CFFN}}$与原始输入特征$\boldsymbol{X}$进行残差连接，得到多头自注意力机制部分的输出$\boldsymbol{X}_{\mathrm{MHSA}}$。这种残差连接的设计不仅有助于缓解深度网络的梯度消失问题，还能保持原始特征信息，使得网络能够更好地学习残差特征映射。

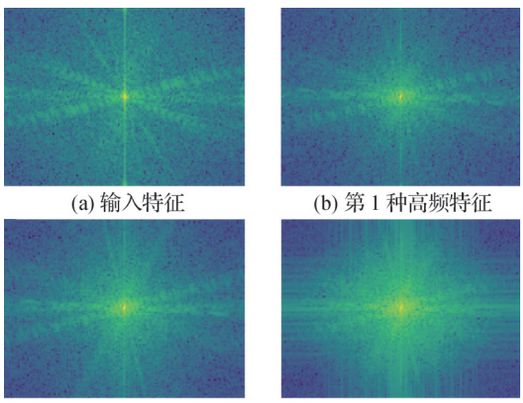
1.2.2　前馈神经网络

SFFM模块中的前馈神经网络部分对多头自注意力机制部分的输出特征$\boldsymbol{X}_{\mathrm{MHSA}}$进行映射与非线性变换，以学习更丰富的特征表示。首先对$\boldsymbol{X}_{\mathrm{MHSA}}$进行批归一化，然后使用CFFN进行非线性映射变换，最后将输出特征与$\boldsymbol{X}_{\mathrm{MHSA}}$进行残差连接，得到SFFM模块的最终输出$\boldsymbol{X}_{\mathrm{SFFM}}$。过程可表示为

$$\boldsymbol{X}_{\mathrm{SFFM}}=C_{1\times1}\left(GELU\left(DWC_{3\times3}\left(C_{1\times1}\left(BN\left(\boldsymbol{X}_{\mathrm{MHSA}}\right)\right)\right)\right)\right)+\boldsymbol{X}_{\mathrm{MHSA}} \tag{7}$$

1.3　频率域特征融合

FFFM模块的设计受ChaIR的启发，采用隐式方法融合高频信息，而无需显式的傅里叶变换。该模块通过堆叠使用多个卷积层作为高通滤波器来增强输入特征中的高频分量并丰富其多样性。增加卷积层的堆叠数量可以提取更丰富的高频特征，从而提升FFFM模块输出特征的质量。但出于参数规模和计算效率的考虑，DFFNet最终采用堆叠3个卷积层来提取3种不同的高频特征。各个卷积层的提取效果如图4所示。从图4(a)所示的原始输入特征出发，经过逐层卷积增强后分别得到图4(b)(c)(d)3种不同程度的高频特征表示。值得注意的是，随着卷积层数的增加，高频特征的表现愈发突出。这种多层次的高频特征提取机制使FFFM能够灵活地融合不同尺度的高频信息，从而更好地实现高频分



(a) 输入特征　(b) 第1种高频特征　(c) 第2种高频特征　(d) 第3种高频特征

图4　不同卷积层的特征提取效果

Fig. 4　Feature extraction effects of different convolutional layers ((a) input features; (b) the first type of high-frequency features; (c) the second type of high-frequency features; (d) the third type of high-frequency features)

量的恢复。

1.3.1　基于MCA的初始特征融合

假设输入特征为$\boldsymbol{X} \in \mathbf{R}^{C \times H \times W}$。以堆叠使用3个卷积层的情况为例，FFFM模块中的初始特征融合的过程表示如下。多个卷积层提取特征可表示为

$$\boldsymbol{x}^i = \begin{cases} GELU\left(C_{3\times3}^i(\boldsymbol{X})\right) & i = 1 \\ GELU\left(C_{3\times3}^i(\boldsymbol{x}^{i-1})\right) & i \in \{2,3\} \end{cases} \tag{8}$$

式中，$\boldsymbol{x}^i$和$C_{3\times3}^i$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。对于增强后得到的多种高频特征，FFFM采用了一种不同于常见SKFusion的初始特征融合(initial feature fusion, IFF)方式。SKFusion通常使用简单的加法运算作为初始特征融合的方式，而FFFM则引入了MCA来实现更有效的特征融合。具体而言，MCA会为每种高频特征计算相应的注意力权重，然后通过加权融合的方式得到初始特征融合结果。过程可表示为

$$\boldsymbol{m}^i = C_{1\times1}^i\left(C_{1\times1}\left(GAP\left(\sum_{j=1}^{3}\boldsymbol{x}^j\right)\right)\right),\ i \in \{1,2,3\} \tag{9}$$

式中，$\boldsymbol{m}^i$分别表示$\boldsymbol{x}^i$对应的中间注意力权重，GAP表示全局平均池化。接着，将3个中间注意力权重拼接，并通过softmax激活，得到$\boldsymbol{x}^i$对应的3个通道注意力权重$\boldsymbol{w}^i$。MCA采用1×1卷积以避免破坏特征中的结构信息。将每种高频特征通过对应的通道注意力权重进行加权融合，随后通过1×1卷积进行特征细化，最终得到初始特征融合结果$\boldsymbol{X}_{\mathrm{IFF}}$。这种初始特征融合方式通过通道间的动态交互实现了多种高频特征的有效融合，并强调高频特征中的重要部分，显著提升了图像边缘和细节等高频特征的恢复效果，有效地提高FFFM模块的特征表示能力和恢复的无雾图像的质量。初始特征融合过程表示为

$$\boldsymbol{X}_{\mathrm{IFF}} = C_{1\times1}\left(\sum_{i=1}^{3}\boldsymbol{w}^i\boldsymbol{x}^i\right) \tag{10}$$

1.3.2　后续频率特征融合

后续的频率特征融合部分采用了与初始特征融合相似的设计，同样利用MCA来融合多种高频特征。过程可表示为

$$\boldsymbol{M}^i = C_{1\times1}^i\left(C_{1\times1}\left(GAP\left(\boldsymbol{X}_{\mathrm{IFF}}\right)\right)\right) \tag{11}$$

式中，$\boldsymbol{M}^i$指的是$\boldsymbol{x}^i$对应的后续频率特征融合部分的中间注意力权重。将中间注意力权重拼接，并通过softmax激活后得到$\boldsymbol{x}^i$对应的通道注意力权重$\boldsymbol{W}^i$。与初始特征融合类似，将每种高频特征通过其对应的通道注意力权重动态加权融合，并通过1×1卷积进行特征细化，最后与输入特征$\boldsymbol{X}$进行残差连接得到FFFM模块的输出特征$\boldsymbol{X}_{\mathrm{FFFM}}$。这种设计不仅保持了特征融合的一致性，还通过残差连接有效地保留了原始特征信息。过程可表示为

$$\boldsymbol{X}_{\mathrm{FFFM}} = C_{1\times1}\left(\sum_{i=1}^{3}\boldsymbol{W}^i\boldsymbol{x}^i\right) + \boldsymbol{X} \tag{12}$$

1.4　损失函数

为了优化频空双域的无雾图像恢复效果，DFFNet采用了综合的损失函数设计。除了在空间域采用$\mathcal{L}_1$损失，还引入了频率域的$\mathcal{L}_1$损失(Cho等，2021)。同时，在空间域引入了使用ResNet-152作为固定预训练模型的对比正则化(Wu等，2021)。空间域损失$\mathcal{L}_{\mathrm{s}}$和频率域损失$\mathcal{L}_{\mathrm{f}}$表示为

$$\mathcal{L}_{\mathrm{s}} = \left\|\boldsymbol{J} - \hat{\boldsymbol{J}}\right\|_1 + \alpha\sum_{i=1}^{n}\omega_i \times \frac{\left\|\boldsymbol{R}_i(\boldsymbol{J}) - \boldsymbol{R}_i\left(\hat{\boldsymbol{J}}\right)\right\|_1}{\left\|\boldsymbol{R}_i(\boldsymbol{I}) - \boldsymbol{R}_i\left(\hat{\boldsymbol{J}}\right)\right\|_1} \tag{13}$$

$$\mathcal{L}_{\mathrm{f}} = \left\|\mathcal{F}(\boldsymbol{J}) - \mathcal{F}\left(\hat{\boldsymbol{J}}\right)\right\|_1 \tag{14}$$

式中，$\mathcal{F}$表示快速傅里叶变换，$\boldsymbol{J}$和$\hat{\boldsymbol{J}}$分别表示无雾真值图像和恢复的无雾图像，$\boldsymbol{I}$表示有雾图像，$\boldsymbol{R}_i$($i = 1, 2, \cdots, n$)表示从固定预训练模型ResNet-152中提取第i个隐藏特征，ω_i为对应的权重系数，超参数α设置为0.1。总体损失函数由$\mathcal{L}_{\mathrm{s}}$和$\mathcal{L}_{\mathrm{f}}$两部分构成，表示为

$$\mathcal{L} = \mathcal{L}_{\mathrm{s}} + \beta\mathcal{L}_{\mathrm{f}} \tag{15}$$

式中，超参数β设置为0.1。

2　实　验

2.1　实验设置

DFFNet 基于 PyTorch 2.1.1 框架在 NVIDIA GeForce RTX 3090 GPU 和 NVIDIA L40 GPU 平台上实现。在训练过程中，输入图像随机裁剪为 256 × 256 像素的块进行处理。DFFNet 有 3 种不同规模的变体，-T、-S、-L 分别表示消融模型、小型模型和大型模型，分别对应不同尺度编码器与解码器中 1、2、8 层的 SFFM 模块堆叠层数。DFFNet-T 模型专用于进行消融实验分析；DFFNet-S 模型具有轻量级的网络架构，适用于实际应用中的实时去雾场景。DFFNet-L 模型则适用于对恢复的无雾图像质量要求严格的高精度任务。为增强特征表达，DFFNet 从固定预训练的 Resnet-152 中提取第 11、35、143、152 层的特征用于对比正则化，对应的权重系数 ω_i 设置为 {1/16, 1/8, 1/4, 1}。数据集的实验参数配置遵循 MixDehazeNet 的设置。在 RESIDE-IN 训练集上，DFFNet-T 采用较大的初始学习率（4×10^{-4}）和批量大小（32）以加速收敛，便于快速完成消融实验；而 DFFNet-S、DFFNet-L 则使用较小的初始学习率（2×10^{-4}）和批量大小（16）以确保训练稳定性。同时，学习率通过余弦退火策略逐渐降低至初始学习率的 1%。DFFNet 使用 AdamW 优化器（Loshchilov 和 Hutter，2019）进行训练，其中 $\beta_1 = 0.9, \beta_2 = 0.999$。所有模型的参数量（parameters，Params）和浮点运算次数（floating point operations，FLOPs）基于 256 × 256 像素的图像块进行测量。

实验数据集：在 RESIDE-IN 训练集（13 990 个图像对）和 Haze4K 训练集（3 000 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（1 000 个图像对）上进行测试。

2.2　对比实验

2.2.1　定量实验

在定量实验中，采用 PSNR 和结构相似性指数（structural similarity index measure，SSIM）衡量去雾模型的性能，使用 Params 和 FLOPs 衡量去雾模型的计算效率。

在 SOTS-Indoor 测试集和 Haze4K 测试集上与较先进方法的定量对比如表 1 所示。在 SOTS-Indoor 测试集上，DFFNet-L 领先 MixDehazeNet-L 模型 1.21 dB，成为首个在该测试集上 PSNR 超过 43 dB 的去雾模型。同时，DFFNet-L 所需计算开销更小，Params 仅为其 46.0%，FLOPs 仅为 67.1%。同时，得益于 DFFNet 基于 SR 来恢复无雾图像，通过引入弱先验，DFFNet-S 能够以极低的计算资源开销实现优秀的图像去雾效果。这种高效的性能优势使其特别适合于如手术场景等对实时性和资源限制有严格要求的场景。

表 2 展示了 DFFNet 与其他先进的去雾方法对 SOTS-Indoor 测试集中的图像去雾所需时间的对比。表中运行时间为单幅 620 × 460 像素图像的处理时间。通过控制变量法确保运行时环境尽量一致。

从表 2 可以看出，虽然 DFFNet-L 在 Params 和 FLOPs 上具有优势，但是单幅图像的处理时间却长于其他方法。这种现象很可能是由于 DFFNet-L 中存在大量的深度卷积所导致。深度卷积虽然理论计算量比标准卷积更少，但是其并行度较低，无法充分利用 GPU 的并行计算能力，并且其不规则的内存访问模式导致内存带宽利用率较低，从而影响执行速度。但是在实际应用场景中，由于 DFFNet-L 的 Params 更小，内存占用更低，因此可以使用较大的批处理大小来提高 GPU 的利用率，弥补并行度不足的问题，从而在整体吞吐量上实现高效的去雾处理。同时，DFFNet-S 具有优秀的去雾性能和处理速度，可以用于实时去雾场景。

在 Haze4K 测试集上，DFFNet-L 领先 MixDehazeNet-L 模型 0.45 dB，成为首个在该测试集上 PSNR 超过 36 dB 的去雾模型。

2.2.2　定性实验

在定性实验中，通过比较模型恢复的无雾图像的视觉效果来衡量模型的去雾性能，并通过比较去雾结果与无雾图像的残差图像的离散傅里叶变换结果来比较模型在频率域的去雾性能。DFFNet-L 与 SOTS-Indoor 测试集上 MixDehazeNet-L 等优秀去雾模型恢复的无雾图像视觉效果如图 5 所示。通过详细比较可见，DFFNet-L 恢复的无雾图像是质量最优、最接近无雾真值图像的。这主要归功于其大感受野的设计，使模型能够充分捕捉并利用图像的全局上下文信息进行精确去雾。这一优势使 DFFNet-L 能够更准确地识别场景中雾气的空间分布特性，从而在恢复图像中实现更彻底的雾气消除、更均匀

表1　在SOTS-Indoor测试集和Haze4K测试集上的定量比较结果

Table 1　Quantitative comparison results on the SOTS-Indoor test set and Haze4K test set

方法	刊物和年份	SOTS-Indoor		Haze4K		Params/M	FLOPs/G
		PSNR/dB	SSIM	PSNR/dB	SSIM		
DCP	TPAMI 2010	16.62	0.818	14.01	0.760	–	–
DehazeNet	TIP 2016	19.82	0.821	19.12	0.840	0.000 9	0.581
AOD-Net	ICCV 2017	19.06	0.850	17.15	0.830	0.002	0.115
GridDehazeNet	ICCV 2019	32.16	0.984	23.29	0.930	0.956	21.49
MSBDN	CVPR 2020	33.67	0.985	22.99	0.850	31.35	41.54
FFA-Net	AAAI 2020	36.39	0.989	26.96	0.950	4.456	287.8
AECR-Net	CVPR 2021	37.17	0.990	–	–	2.611	52.20
PMDNet	ECCV 2022	38.41	0.985	33.49	0.980	18.90	81.13
DehazeFormer-L	TIP 2023	40.05	0.996	–	–	25.44	279.7
SANet	IJCAI 2023	40.40	0.996	–	–	3.81	37.26
OKNet	AAAI 2024	40.79	0.996	–	–	4.72	39.67
FocalNet	ICCV 2023	40.82	0.996	–	–	3.74	30.63
IRNeXt	ICML 2023	41.21	0.996	–	–	5.46	41.95
SFNet	ICLR 2023	41.24	0.996	–	–	13.27	125.43
DEA-Net-CR	TIP 2024	41.31	0.995	34.25	0.990	3.653	32.23
gUNet-D	arXiv 2022	41.34	0.996	33.52	0.988	5.025	16.48
DSANet	NN 2024	41.36	**0.997**	–	–	3.86	37.72
ChaIR	KBS 2023	41.95	**0.997**	–	–	15.02	140.75
FSNet	TPAMI 2023	42.45	**0.997**	34.12	0.990	13.28	111.14
C2PNet	CVPR 2023	42.56	0.995	–	–	7.17	460.95
MixDehazeNet-L	IJCNN 2024	42.62	**0.997**	35.94	0.992	12.42	86.7
DFFNet-S	本文	41.59	0.996	–	–	1.72	15.86
DFFNet-L	本文	**43.83**	**0.997**	**36.39**	**0.993**	5.71	58.15

注：加粗、下划线字体分别表示各列最优、次优结果。“–”表示无实验结果。

表2　不同方法计算效率和运行时间对比结果

Table 2　Comparison of computing efficiency and running time for different methods

方法	Params/M	FLOPs/G	运行时间/s
ChaIR	15.02	140.75	0.082 0
FSNet	13.28	111.14	0.123 0
MixDehazeNet-L	12.42	86.70	0.118 0
DFFNet-S（本文）	1.72	15.86	0.037 4
DFFNet-L（本文）	5.71	58.15	0.130 0

自然的色彩还原以及更优质的整体视觉效果。同时DFFNet-L利用清晰/退化图像对之间的频率域差异特征，显著增强了对图像高频分量的恢复能力，这使得恢复图像的边缘轮廓与细节纹理更加清晰锐利，提高图像的细节还原精度，使结果更加贴近无雾真值图像。图中的矩形框用于突出展示不同去雾模型恢复的无雾图像的视觉差异。

图6通过频率域分析提供了更客观的性能评估。图6展示了DFFNet和SOTS-Indoor数据集上MixDehazeNet-L等优秀去雾模型恢复的无雾图像与无雾真值图像相减所得残差图像在离散傅里叶变换后的频谱分布。分析结果表明，DFFNet恢复的图像在高频区域与无雾真值图像的差异最小，证实了其在频率域上具有最优的恢复效果。

图7展示了DFFNet-L与Haze4K测试集上最先进的去雾模型恢复的无雾图像视觉效果。从图中可

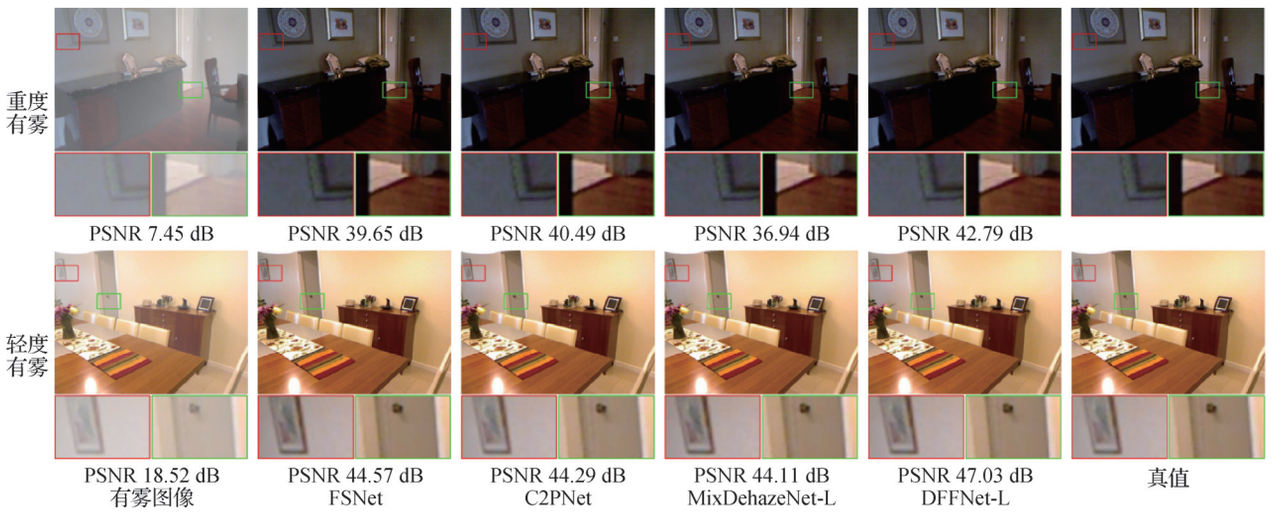


图 5　不同模型在 SOTS-Indoor 上恢复的室内无雾图像视觉效果

Fig. 5　Visual effects of haze-free indoor images restored by different models on SOTS-Indoor

以观察到，即使是在复杂多变的室外场景，DFFNet 仍对有雾图像具有良好的去雾能力。

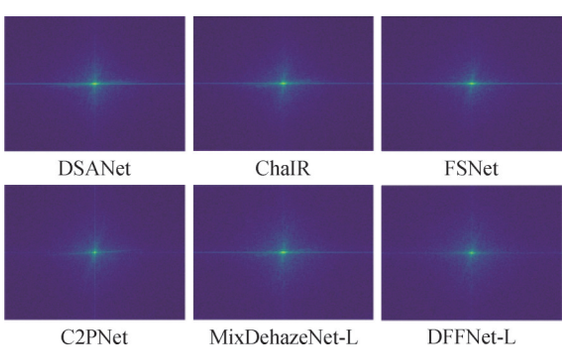


图 6　不同模型恢复的无雾图像与真实无雾图像残差结果的频谱分布

Fig. 6　Spectral distribution of residual results between haze- free images restored by different models and real haze-free images

2.3　消融实验

通过在 RESIDE-IN 训练集上训练 DFFNet-T 和基于基线的变体模型，并在 SOTS-Indoor 测试集上进行评估来进行消融实验。通过消融实验，能够精确分析 SFFM 和 FFFM 这两个核心组件对 DFFNet 去雾性能的实际贡献，验证它们在模型架构中的有效性与必要性。

不同组件的性能影响如表 3 所示。首先通过将 MixDehazeNet-T 中的 MSBlock 换成 VAN 来构建基线（baseline）模型，该基线模型在测试集上达到 37. 68 dB 的 PSNR。用 SFFM 模块替代基线中的 VAN 模块（基线 + SFFM）时，PSNR 提升 0. 64 dB；而当在基线模型中引入 FFFM 模块（基线 + FFFM）时，性能提升更为显著，达到 1. 37 dB 的增益。最终，通过同时整合这两个关键模块 SFFM 和 FFFM，完整的 DFFNet-T 架构相比基线模型实现了 2. 31 dB 的 PSNR 提升，有力地证明了这两个模块在图像去雾任务中的有效性与必要性。

图 8 直观展示了 DFFNet 中 SFFM 模块和 FFFM 模块对中间特征的处理效果。第 3—5 列则依次展示了初始输入特征、经 SFFM 处理后的特征以及经 SFFM 和 FFFM 共同处理后的特征。SFFM 模块主要发挥空间注意力机制的优势，使模型能够精准定位并聚焦于图像中的退化区域，从而在处理大面积雾气覆盖区域时表现出色。FFFM 模块则通过有效放大和强调高频信息，实现了边缘轮廓的锐化和细节纹理的增强，使最终恢复的无雾图像在视觉质量上更接近真实场景。这一可视化结果不仅验证了两个模块在不同层面上的互补作用，也解释了为何完整的 DFFNet 架构能够在定量评估中取得显著性能提升。SFFM 专注于大尺度雾气去除，而 FFFM 则着重于精细细节恢复，二者协同工作，互相补充，形成了高效的去雾处理流程。

表 4 展示了 SFFM 模块与其他空间域特征融合模块的性能对比。在基线模型的基础上进行改进，将 SFFM 与 3 种主流的空间域特征融合模块进行了全面对比。实验结果显示，SFFM 在性能上明显优于 ChaIR 中的残差块和 VAN 模块，PSNR 分别提高 2. 5 dB 和 0. 64 dB。值得注意的是，虽然 SFFM 的

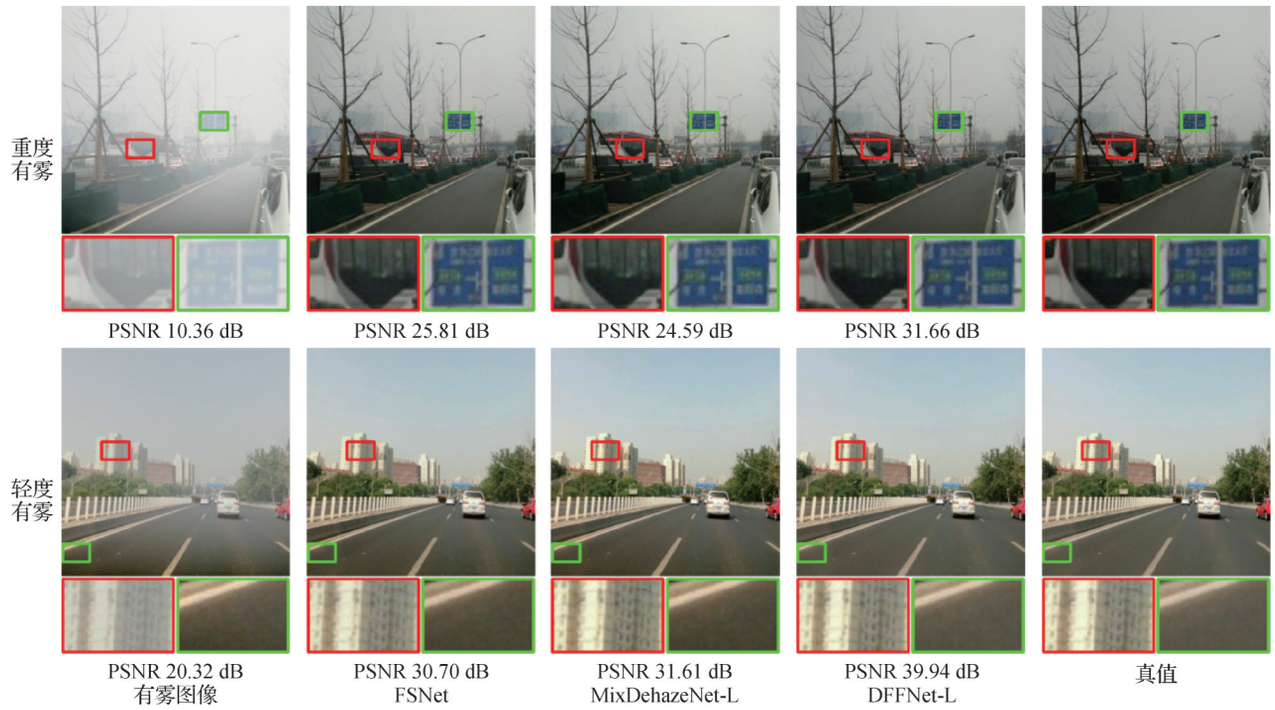


图7　不同模型在Haze4K数据集上恢复的室外无雾图像视觉效果

Fig. 7　Visual effects of haze-free outdoor images restored by different models on the Haze4K dataset

表3　SOTS-Indoor数据集上DFFNet不同组件消融研究

Table 3　Ablation study of different components of DFFNet on the SOTS-Indoor dataset

方法	PSNR/dB	SSIM	Params/M	FLOPs/G
基线	37.68	0.993 2	0.426	4.403
基线 + SFFM	38.32	0.993 8	0.747	7.716
基线 + FFFM	39.05	0.994 3	0.732	5.499
DFFNet-T	39.99	0.995 1	1.052	8.812

PSNR略低于MixDehazeNet中的混合结构块，但SFFM在计算资源利用方面展现出显著优势，其Params仅为混合结构块的46%，FLOPs仅为53%。这一特性使SFFM在保持竞争性能的同时，大幅降低了计算开销，提高运行效率。SFFM模块不仅具有出色的去雾性能，还兼具轻量化和高效率的特点，可作为一个通用且高效的组件，便捷地集成到各类图像恢复架构中，以提升整体恢复效果。

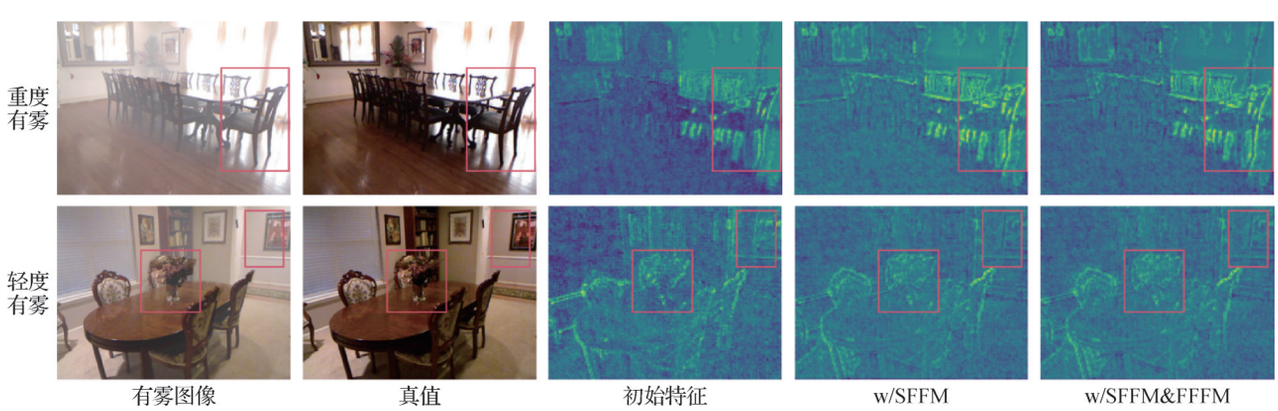


图8　经过SFFM和FFFM处理后特征的视觉效果

Fig. 8　Visual effects of features after processing by SFFM and FFFM

表5展示了FFFM模块与其他频率域特征融合模块的性能对比结果。为客观评估FFFM的优势，在基准模型的基础上进行改进，将FFFM与3种主流的频率域特征融合模块进行了全面对比。为确保对

比的公平性，所有对比模块均仅部署于网络的瓶颈位置。实验结果显示，相比FSNet中的多分支动态选择频率模块（multi-branch dynamic selective frequency，MDSF）+多分支紧凑选择频率模块（multi-branch compact selective frequency，MCSF）、FocalNet中的频率选择模块以及ChaIR中的频域通道注意力模块，FFFM的PSNR分别提高1.05 dB、0.35 dB和0.28 dB。对比结果不仅有力证明了频域特征融合技术在图像去雾任务中的重要价值，而且展示了FFFM的性能优势。并且，FFFM也可作为一个通用的组件，便捷地集成到各类图像恢复架构中，以提升整体恢复效果。

表6展示了FFFM模块中卷积层堆叠层数N对DFFNet-T整体性能的影响。实验结果表明，随着N的增加，FFFM能够融合更多样化的高频特征，从而提升重建特征的质量，使DFFNet的性能指标得到有效提升。然而，也观察到一个明显的边际效应：当卷积层深度持续增加时，性能提升的幅度逐渐减小，呈现出收益递减的趋势。出于对参数规模、计算效率以及性能增益的考虑，选择在FFFM堆叠3层卷积层来强调并融合3种高频特征。

表4　SFFM与其他空间域特征融合模块的比较

Table 4　Comparison of SFFM with other spatial domain feature fusion modules

方法	PSNR/dB	SSIM	Params/M	FLOPs/G
基线 + 残差块	35.82	0.991 3	0.766	7.421
基线	37.68	0.993 2	0.426	4.403
基线 + 混合结构块	38.35	0.994 1	1.623	14.621
基线 + SFFM	38.32	0.993 8	0.747	7.716

表5　FFFM与其他频率域特征融合模块的比较

Table 5　Comparison of FFFM with other frequency domain feature fusion modules

方法	PSNR/dB	SSIM	Params/M	FLOPs/G
基线 + MDSF + MCSF	38.00	0.993 7	0.454	4.423
基线 + 频率选择模块	38.70	0.993 9	0.426	4.403
基线 + 频域通道注意力模块	38.77	0.994 1	0.704	5.461
基线 + FFFM	39.05	0.994 3	0.732	5.499

表6　不同的卷积层堆叠层数N对DFFNet-T性能的影响

Table 6　Impact of different numbers of stacked convolutional layers N on DFFNet-T performance

N	PSNR/dB	SSIM	Params/M	FLOPs/G
2	39.92	0.995 0	0.959	8.472
3	39.99	0.995 1	1.052	8.812
4	40.05	0.995 0	1.144	9.152
5	40.08	0.995 1	1.236	9.492

3　结　论

本文针对双域特征融合中存在的问题，提出双域特征融合网络。针对在空域特征的提取与融合上存在局限性的问题，提出空间域特征融合模块。该模块采用Transformer架构，通过大核注意力机制捕获全局特征并准确定位图像的有雾区域分布，通过像素注意力机制来建模局部特征并有效恢复图像的边缘和细节，这两种注意力机制共同满足了图像软重建过程中对全局特征和局部特征的双重需求。针对频域特征恢复质量仍有提升空间以及未能实现频空双域特征的高效融合的问题，提出频率域特征融合模块。该模块通过多个卷积层来丰富并增强图像的高频分量，并通过多分支通道注意力对高频分量进行频率域特征融合，显著提升了模型恢复图像高频细节的能力。并且，该模块位于网络瓶颈处，实现了频域与空域特征的高效融合。实验结果表明，所提出网络在SOTS-Indoor测试集和Haze4K测试集上展现出先进的性能表现。同时，所提出的两个关键模块具有良好的可迁移性和扩展性，能够便捷地整合到其他计算机视觉任务中，为提升模型性能提供一种新颖且高效的解决方案。

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索，充分利用清晰/退化图像对在空间域和频率域的差异，发挥其提升模型去雾性能的潜力。其次，由于深度卷积的大量使用会降低网络的处理速率，未来将继续探索设计计算开销小且处理速率高的高效去雾网络。再者，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）

Bai J W, Yuan L, Xia S T, Yan S C, Li Z F and Liu W. 2022. Improving vision transformers by revisiting high-frequency components//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 1-18 [DOI: 10.1007/978-3-031-20053-3_1]

Berman D, Treibitz T and Avidan S. 2016. Non-local image dehazing//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 1674-1682 [DOI: 10.1109/CVPR.2016.185]

Cai B L, Xu X M, Jia K, Qing C M and Tao D C. 2016. DehazeNet: an end-to-end system for single image haze removal. IEEE Transactions on Image Processing, 25(11): 5187-5198 [DOI: 10.1109/TIP.2016.2598681]

Chen W T, Fang H Y, Hsieh C L, Tsai C C, Chen I H, Ding J J and Kuo S Y. 2021. ALL snow removed: single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 4176-4185 [DOI: 10.1109/ICCV48922.2021.00416]

Cho S J, Ji S W, Hong J P, Jung S W and Ko S J. 2021. Rethinking coarse-to-fine approach in single image deblurring//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 4621-4630 [DOI: 10.1109/ICCV48922.2021.00460]

Cui Y N and Knoll A. 2023. Exploring the potential of channel interactions for image restoration. Knowledge-Based Systems, 282: #111156 [DOI: 10.1016/j.knosys.2023.111156]

Cui Y N, Ren W Q, Cao X C and Knoll A. 2023a. Focal network for image restoration//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 12955-12965 [DOI: 10.1109/ICCV51070.2023.01195]

Cui Y N, Ren W Q, Cao X C and Knoll A. 2024b. Image restoration via frequency selection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(2): 1093-1108 [DOI: 10.1109/TPAMI.2023.3330416]

Dai Y M, Gieseke F, Oehmcke S, Wu Y Q and Barnard K. 2021. Attentional feature fusion//Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 3559-3568 [DOI: 10.1109/WACV48630.2021.00360]

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 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 5802-5810 [DOI: 10.1109/CVPR52688.2022.00572]

Guo M H, Lu C Z, Liu Z N, Cheng M M and Hu S M. 2023a. Visual attention network. Computational Visual Media, 9(4): 733-752 [DOI: 10.1007/s41095-023-0364-2]

Guo S, Yong H W, Zhang X D, Ma J Q and Zhang L. 2023b. Spatial-frequency attention for image denoising [EB/OL]. [2024-11-08]. https://arxiv.org/pdf/2302.13598.pdf

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]

Katznelson Y. 2004. An Introduction to Harmonic Analysis. 3rd ed. Cambridge: Cambridge University Press [DOI: 10.1017/CBO9781139165372]

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 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 4780-4788 [DOI: 10.1109/ICCV.2017.511]

Li B Y, Ren W Q, Fu D P, Tao D C, Feng D, Zeng W J and Wang Z Y. 2019. Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing, 28(1): 492-505 [DOI: 10.1109/TIP.2018.2867951]

Liu X H, Ma Y R, Shi Z H and Chen J. 2019. GridDehazeNet: attention-based multi-scale network for image dehazing//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 7313-73223 [DOI: 10.1109/ICCV.2019.00741]

Loshchilov I and Hutter F. 2019. Decoupled weight decay regularization//Proceedings of the 7th International Conference on Learning Representations. New Orleans, USA: OpenReview.net

Lu L P, Xiong Q, Xu B R and Chu D F. 2024. MixDehazeNet: mix structure block for image dehazing network//Proceedings of 2024 International Joint Conference on Neural Networks (IJCNN). Yokohama, Japan: IEEE: 1-10 [DOI: 10.1109/IJCNN60899.2024.10651326]

Narasimhan S G and Nayar S K. 2002. Vision and the atmosphere. International Journal of Computer Vision, 48(3): 233-254 [DOI: 10.1023/A:1016328200723]

Park N and Kim S. 2022. How do vision transformers work?//Proceedings of the 10th International Conference on Learning Representations. [s.l.]: OpenReview.net

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 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI: 11908-11915 [DOI: 10.1609/aaai.v34i07.6865]

Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//Proceedings of the 18th

International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]

Song Y D, He Z Q, Qian H and Du X. 2023. Vision transformers for single image dehazing. IEEE Transactions on Image Processing, 32: 1927-1941 [DOI: 10.1109/TIP.2023.3256763]

Valanarasu J M J, Yasarla R and Patel V M. 2022. TransWeather: transformer-based restoration of images degraded by adverse weather conditions//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 2343-2353 [DOI: 10.1109/CVPR52688.2022.00239]

Wang W H, Xie E Z, Li X, Fan D P, Song K T, Liang D, Lu T, Luo P and Shao L. 2022a. PVT v2: improved baselines with pyramid vision transformer. Computational Visual Media, 8(3): 415-424 [DOI: 10.1007/s41095-022-0274-8]

Wang Z D, Cun X D, Bao J M, Zhou W G, Liu J Z and Li H Q. 2022b. Uformer: a general u-shaped transformer for image restoration//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 17662-17672 [DOI: 10.1109/CVPR52688.2022.01716]

Wu H Y, Qu Y Y, Lin S H, Zhou J, Qiao R Z, Zhang Z Z, Xie Y and Ma L Z. 2021. Contrastive learning for compact single image dehazing//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 10546-10555 [DOI: 10.1109/CVPR46437.2021.01041]

Xing X M and Liu W. 2016. Haze removal for single traffic image. Journal of Image and Graphics, 21(11): 1440-1447 (邢晓敏, 刘威. 2016. 雾天交通场景中单幅图像去雾. 中国图象图形学报, 21(11): 1440-1447) [DOI: 10.11834/jig.20161103]

Zhu Q S, Mai J M and Shao L. 2015. A fast single image haze removal algorithm using color attenuation prior. IEEE Transactions on Image Processing, 24(11): 3522-3533 [DOI: 10.1109/TIP.2015.2446191]

作者简介

王炜嘉，男，硕士研究生，主要研究方向为图像处理与计算机视觉。E-mail: wwj20000720@163.com

陈飞，通信作者，男，教授，主要研究方向为图像处理、计算机视觉与机器学习。E-mail: chenfei314@fzu.edu.cn

刘莞玲，女，博士研究生，实验师，主要研究方向为智能决策与计算机视觉。E-mail: 1019218015@tju.edu.cn

程航，男，教授，主要研究方向为机器学习与图像处理。E-mail: hcheng@fzu.edu.cn

王美清，女，教授，主要研究方向为图像处理。E-mail: mqwang@fzu.edu.cn