

# 基于跨层注意力特征交互和多尺度通道注意力的 单幅图像去雾网络

孙 航<sup>1,2</sup>, 付秋月<sup>1</sup>, 李勃辉<sup>1</sup>, 但志平<sup>1,2\*</sup>, 余 梅<sup>1,2</sup>, 万 俊<sup>3</sup>

(1. 三峡大学计算机与信息学院, 湖北宜昌 443000; 2. 三峡大学湖北省水电工程智能视觉监测重点实验室, 湖北宜昌 443000;  
3. 中南财经政法大学信息与安全工程学院, 湖北武汉 430073)

**摘 要:** 近年来, 基于 U 型结构的卷积神经网络在去雾领域取得了显著的成果. 然而, 大多数基于 U 型结构的去雾网络将编码层特征直接传递到对应尺度的解码层, 忽略了不同层次特征信息的有效利用. 此外, 去雾网络中广泛使用的通道注意力受感受野的限制, 没有充分地利用上下文信息, 从而对通道权重的学习起负面作用, 使得重构的清晰图像不够理想. 为了解决上述问题, 本文提出了一种跨层注意力特征交互和多尺度通道注意力的去雾算法. 具体来说, 跨层注意力特征交互模块利用编码层的多尺度跨层特征学习层级权重, 然后将这些跨层特征聚合传递到对应解码层, 从而减少了去雾网络重构清晰图像过程中的特征稀释. 此外, 为了挖掘对于去雾网络非常重要的特征通道信息, 本文设计了多尺度通道注意力机制, 利用不同空洞率的空洞卷积提取多尺度特征信息, 形成一个多尺度上下文并行学习的通道注意力机制, 可以更有效地为去雾网络的特征分配权重. 实验结果表明, 本文提出的去雾算法在 4 个公开的数据集上相比现有的 12 种去雾方法取得了较好的客观评价指标和视觉效果. 本文的代码已上传至 <https://github.com/bohuisir/AAFMAN>.

**关键词:** 图像去雾; 跨层注意力特征交互; 特征稀释; 多尺度通道注意力; 空洞卷积

**基金项目:** 国家自然科学基金(No.62002233); 湖北省自然科学基金(No.2021CFB004)

**中图分类号:** TP391.4

**文献标识码:** A

**文章编号:** 0372-2112(2024)11-3711-16

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

**DOI:** 10.12263/DZXB.20230821

## Cross-Layer Attention Feature Interaction and Multi-Scale Channel Attention Network for Single Image Dehazing

SUN Hang<sup>1,2</sup>, FU Qiu-yue<sup>1</sup>, LI Bo-hui<sup>1</sup>, DAN Zhi-ping<sup>1,2\*</sup>, YU Mei<sup>1,2</sup>, WAN Jun<sup>3</sup>

(1. College of Computer and Information Technology, China Three Gorges University, Yichang, Hubei 443000, China;

2. Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering, China Three Gorges University, Yichang, Hubei 443000, China;

3. School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan, Hubei 430073, China)

**Abstract:** In recent years, U-shaped convolutional neural networks (CNNs) have achieved remarkable progress in image dehazing. However, most U-shaped dehazing networks directly pass encoder features to the decoder at the corresponding scale, ignoring effective utilization of multi-scale features. In addition, channel attention widely used in dehazing networks is restricted by receptive fields, failing to sufficiently leverage contextual information, which adversely affects learning of channel weights. To address the above issues, this paper proposes a novel dehazing algorithm with cross-layer attentive feature interaction and multi-scale channel attention. Specifically, the cross-layer attentive feature interaction module learns hierarchical weights for multi-scale encoder features, and aggregates these cross-layer features for transfer to the decoder, thereby reducing feature dilution during the dehazing network's reconstruction of clear images. Moreover, to uncover channel information that is critical for dehazing networks, we devise a multi-scale channel attention mechanism that extracts multi-scale features by dilated convolutions with different dilation rates, forming a parallel learning scheme of channel attention with multi-scale contexts for more effective weight allocation for dehazing network features. Experimental results demonstrate that the

proposed dehazing algorithm achieves better objective metrics and visual performance compared to 12 existing methods on 4 public datasets. The code for this paper has been uploaded to <http://github.com/bohuisir/AAFMAF>.

**Key words:** image dehazing; cross-layer attention feature interaction; feature dilution; multi-scale channel attention; dilated convolution

**Foundation Item(s):** National Natural Science Foundation of China (No.62002233); Natural Science Foundation of Hubei Province (No.2021CFB004)

## 1 引言

过去几十年,随着工业化的迅速发展,全国各地雾霾天气频发,使得成像设备拍摄到的图像内容存在颜色失真、对比度降低以及细节丢失等问题,从而对目标检测<sup>[1, 2]</sup>、目标跟踪<sup>[3, 4]</sup>、人脸检测<sup>[5, 6]</sup>等高级视觉算法的应用造成严重的负面影响. 因此,图像去雾领域成为众多研究人员关注的热点. 早期的图像去雾算法通常利用先验信息估计大气散射模型<sup>[7]</sup>的参数实现去雾. 虽然基于先验的去雾方法<sup>[8, 9]</sup>已经取得了一定的进展,但这些去雾方法在无约束条件下,通常恢复出的清晰图像质量不够理想. 近年来,随着深度学习的不断发展,研究人员提出了许多基于卷积神经网络的去雾方法,主要分为参数估计<sup>[10, 11]</sup>去雾和端到端去雾<sup>[12, 13]</sup>方法两大类.

参数估计的去雾方法容易产生不准确的结果,导致去雾图像存在颜色失真和雾残留等问题. 例如, Cai 等人<sup>[10]</sup>提出了一种 DehazeNet 去雾网络,该方法通过卷积神经网络估计传输图,从而实现去雾效果. Zhang 等人<sup>[11]</sup>提出了一种密集连接金字塔估计传输图和大气光的方法,再通过大气散射模型实现去雾. 但这些方法所得估计值依旧与真实值存在一定差异,导致最终去雾图像不够理想.

目前,大多数去雾算法使用 U 型网络结构作为基础实现端到端<sup>[14-16]</sup>的去雾. 为了实现更好的去雾性能,许多研究人员在 U 型网络中设计不同的信息融合方式提升特征的表达能. 例如 Dong 等人<sup>[17]</sup>设计了一种密集特征融合的多尺度增强去雾网络 MSBDN (Multi-Scale Boosted Dehazing Network with dense feature fusion),该方法基于 DenseNet<sup>[18]</sup>的思想,在编码和解码层上分别进行密集跳用于多尺度的特征交互. Wu 等人<sup>[19]</sup>基于对比学习的思路提出了一种去雾网络 AECD-Net (AutoEncoder and Contrastive Regularization Network),该模型采用有雾图像作为负样本来约束网络的学习. 此外,在网络模型的设计中该算法采用 Mixup 在对应编解码层之间进行信息融合,并获得了较好的去雾性能. 虽然这些端到端的去雾方法取得了一定进展,但是忽略了网络中非对应解码和编码层之间的跨层信息交互,导致特征信息稀释,使得重构的清晰图像不够理想.

近年来,通道注意力 SE (Squeeze-and-Excitation net-

works)<sup>[20]</sup>被广泛应用于目标检测<sup>[21, 22]</sup>、语义分割<sup>[23, 24]</sup>和图像分割<sup>[25]</sup>等视觉领域, SE 的核心思想是通过评估各特征通道的重要程度,针对不同的任务学习各通道权重,达到增强或抑制不同特征通道的目的. Qin 等人<sup>[26]</sup>首次将通道注意力机制引入去雾网络模型,提出了一种称为 FFA-Net (Feature Fusion Attention Network for single image dehazing) 的去雾算法,达到良好的去雾效果. Wu 等人<sup>[19]</sup>基于 FFA-Net 中的特征注意力模块构建了一种紧凑的去雾网络,并实现了更好的去雾性能. 然而, SE 通道注意力通过直接在特征图上进行全局平均池化来提取通道特征,存在感受野受限,上下文信息没有充分考虑的问题,无法实现准确的通道特征加权. 因此,如何设计一个有效的通道注意力对图像去雾模型的性能至关重要.

为了解决上述问题,本文提出了一种跨层注意力特征交互和多尺度通道注意力的去雾网络 CLMSA-Net (Cross-Layer attention feature interaction and Multi-Scale channel Attention dehazing Network). 具体而言,为了有效地利用不同层级的信息,跨层注意力特征交互模块 CLA (Cross-Layer Attention feature interaction module) 将聚合后的不同编码层信息进行融合,从而保留浅层和深层信息. 再将这些跨层聚合特征在分配相应权重后传递到对应解码层,减少了编码器提取特征过程中的特征稀释. 此外,为了获取编解码不同层级中的多尺度特征表达,本文设计了一个多尺度注意力机制 MSCA (Multi-Scale Channel Attention module),利用不同空洞率的空洞卷积获得有效的多尺度特征,并通过并行注意力机制分配通道权重,以获得有效的多尺度特征通道信息. 该注意力实现了从不同感受野的分支中汇集多尺度信息,有效地丰富了特征表示,提高了网络的去雾效果. 综上所述,本文的主要贡献如下:

(1) 为了缓解网络模型中的特征稀释并增强网络的表达能力,本文提出了一种跨层注意力特征交互模块 CLA. 通过获取编码器内不同的层级特征,并对这些跨层特征进行聚合. 经过权重分配后,再将包含边缘纹理和语义信息的聚合特征传递到相应解码层,有效缓解了重构清晰图像过程中的特征稀释问题.

(2) 设计了一种多尺度注意力机制 MSCA. 该注意力机制提取网络中同一特征的不同感受野信息后分别

进行通道权重的学习,可以有效地将这些包含丰富上下文信息的特征进行聚合,形成一个更合理的特征权重分配机制.

(3)在自然图像数据集 Reside-Outdoor<sup>[27]</sup>、Dense-Haze<sup>[28]</sup>、NH-Haze<sup>[29]</sup>,以及遥感图像数据集 StateHaze-1k<sup>[30]</sup>上的实验证明,本文提出的方法在客观指标和主观视觉效果上优于最近提出的 12 种优秀去雾算法.

## 2 相关工作

传统的去雾算法大多是基于物理模型实现图像去雾.其中,最具代表性的是 He 等人<sup>[9]</sup>提出的 DCP(Dark Channel Prior)单幅图像去雾算法,根据图像的暗通道先验知识估计传输图,再通过大气散射模型得到去雾图像. Tang 等人<sup>[31]</sup>提出了一种随机森林的方法学习回归模型,估计有雾图像的传输图并通过大气散射模型进行图像去雾.近年来,基于深度学习的图像去雾方法受到广泛关注,主要包括:基于参数估计、端到端的模型和注意力机制.下面将从这 3 类方法对图像去雾的发展进行回顾.

### 2.1 基于参数估计的模型

随着深度学习的发展,研究人员利用卷积神经网络来估计传输图和大气光值,再根据大气散射模型得到清晰的图像,取得了比传统的参数估计更好的去雾效果. Cai 等人<sup>[10]</sup>设计了一种 DehazeNet 算法,首次利用卷积神经网络对传输图进行估计,建立了有雾图像和传输图之间的映射关系并通过大气散射模型来进行图像去雾. Ren 等人<sup>[32]</sup>通过学习有雾图像与传输图之间的映射,提出了一种用于单图像去雾的多尺度卷积神经网络,该算法由粗尺度和细尺度网络组成,分别用于预测整体传输图和局部结果的优化,进一步提升了去雾效果. Zhang 等人<sup>[11]</sup>针对 DehazeNet 没有对大气光值进行有效利用的问题,提出了一种密集连接金字塔去雾网络,对传输图和大气光值均进行了估计,再通过大气散射模型得到清晰图像. Li 等人<sup>[33]</sup>提出了一种面向单图去雾的感知渐进网络 LAP-Net(Level-Aware Progressive Network).不同于之前的多阶段算法, LAP-Net 在不同雾霾级别的监督下逐级学习传输图.同时,通过密集残差池化块学习大气光值,使得网络能够处理雾浓度较厚的图像. Ren 等人<sup>[34]</sup>提出了一种基于整体边缘的多尺度卷积神经网络的图像去雾算法,利用粗细尺度网络来学习有雾图像与传输图之间的映射,实现单幅图像的去雾. Li 等人<sup>[35]</sup>提出了一种多尺度的方法来对传输图和大气光值进行估计,以实现图像去雾的效果.虽然基于参数估计的去雾算法实现了图像去雾,但对大气光值和透射图的估计难以真正拟合自然界中复杂的大气分布,所得到的去雾图像往往伴随着伪影、

颜色失真、色差等图像退化现象.

### 2.2 基于端到端的模型

目前,基于端到端图像去雾方法逐渐成为研究的主流方向.通过卷积神经网络直接学习有雾图像与清晰图像之间的映射关系,从而实现图像去雾.根据是否使用配对数据集进行训练,这些方法可大致分为无监督学习和监督学习两大类别.

研究人员开始探索无监督学习在去雾方面的应用.例如, Zhu 等人<sup>[36]</sup>提出的 CycleGAN(Cycle Generative Adversarial Network),该模型能够利用非配对数据集,实现不同域之间图像的转换. Engin 等人<sup>[37]</sup>在 CycleGAN 的基础上提出了 Cycle-dehaze 的无监督去雾网络,该方法在原有的 CycleGAN 框架之上引入感知一致性损失,以提高纹理信息的质量,生成视觉效果更好的去雾图像. Anvari 等人<sup>[38]</sup>提出一种 Dehaze-GLCGAN(Dehazing Global-Local cycle-Consistent Generative Adversarial Network)网络,通过在 CycleGAN 框架上添加全局和局部判别器,从而实现较好的去雾效果. Dan 等人<sup>[16]</sup>提出一种双判别器异构对抗的端到端去雾网络,引入双判别器异构循环框架,通过异构同一任务生成器和局部感受野约束生成器,增强模型对局部区域的学习能力,提升算法收敛性,有效适用真实场景的去雾.

虽然无监督去雾算法取得了一定的去雾效果,但无监督去雾算法在去雾效果以及图像细节的重构等方面与监督学习相比存在一定的差距.例如, Li 等人<sup>[39]</sup>提出一种基于卷积神经网络的一体化去雾网络 AOD-Net(All-in-One Dehazing Network),更有效地实现图像去雾. Qu 等人<sup>[40]</sup>设计了一种增强型 pix2pix 去雾网络,通过多分辨率生成器和多尺度判别器,并结合增强模块,实现了很好的去雾性能. Qin 等人<sup>[26]</sup>提出了一种特征融合注意力网络(FFA-Net),该算法设计了一种通道注意力和像素注意力结合的模块 FA(Feature Attention),可自适应地学习特征权重的分配,取得了良好的去雾效果.此外, Wu 等人<sup>[19]</sup>提出了一种基于对比学习的去雾算法 AEER-Net,利用清晰图像和有雾图像分别作为正负样本构建了对比正则化损失,并在编解码对应层之间采用 Mixup 进行自适应的特征融合,实现了 SOTA(State Of The Art)的去雾性能. Dong 等人<sup>[17]</sup>通过在编码层和解码层中设计密集的跳连接,提出了一种具有密集特征融合的多尺度去雾算法 MSBDN,取得了较好的去雾效果. Guo 等人<sup>[41]</sup>在 MSBDN 的基础上提出了 MARG-UNet(a single image dehazing Network based on Multimodal Attention Residual Group).该网络通过在残差块中嵌入通道注意模块和像素注意模块,提高了非均匀特征提取能力和模型的去雾性能.

### 2.3 注意力机制

近年来,注意力机制被广泛应用于图像处理领域并取得了显著的成绩. Hu 等人<sup>[20]</sup>通过设计挤压和激励操作,实现了特征图中通道权重的分配,称为 SE 注意力机制. Woo 等人<sup>[42]</sup>通过结合空间注意和通道注意力提出了 CBAM (Convolutional Block Attention Module) 注意力模型,进一步提升了分类效果. Wang 等人<sup>[43]</sup>设计了一种高效的通道注意力机制 ECA-Net (Efficient Channel Attention for deep convolutional neural Networks),避免通道维度降维可能带来的负面影响的同时降低了参数量.与此同时,许多研究者针对 U-Net 模型存在特征之间的长距离依赖关系以及编解码层之间语义差异等问题进行了研究,如 Cai 等人<sup>[44]</sup>通过在 U-Net 基础上引入多尺度和注意力机制设计了 MA-Unet (an improved version of Unet based on Multi-scale and Attention mechanism for medical image segmentation) 模型,取得较好的分割效果. Ates 等人<sup>[45]</sup>提出了一种双交叉注意力模块 DCA (Dual Cross-Attention),有效提取多尺度编码器特征之间的通道和空间依赖,以解决语义差距问题.最近几年,注意力机制广泛应用于图像去雾的网络模型设计, Qin 等人<sup>[26]</sup>通过将通道注意力和像素注意力相结合,设计了一种特征融合注意力网络 (FFA-Net),提高网络的表达能力. Shyam 等人<sup>[46]</sup>将编码层信息通过空间感知通道注意力机制 (SACA) 融合到对应分辨率的解码层特征中,获得了较好的去雾效果. Lin 等人<sup>[47]</sup>提出了一种基于多尺度注意力特征融合网络 MSAFF-Net (MultiScale

Attention Feature Fusion Networks),由特征提取模块、通道注意力模块和并行的多尺度空间注意力模块组成,具有良好的去雾性能. Dan 等人<sup>[16]</sup>通过引入多阶特征统计量提出了多阶通道注意力模块,从而进一步挖掘对于图像去雾重要的特征通道信息增强网络去雾性能.

虽然这些方法都取得了较好的去雾效果,但是大多数去雾网络将编码层特征信息直接传递到对应尺度的解码层,忽略了不同层次特征信息的有效利用,生成的图像仍然存在较多伪影和色彩失真等现象.此外,通道注意力受其感受野限制无法有效地利用上下文信息,并且通道注意力中间过程的降维操作会对通道权重的预测产生负面影响,影响去雾网络性能.为此,本文提出了基于跨层注意力特征交互和多尺度通道注意力的去雾网络.

## 3 本文方法

### 3.1 基于跨层注意力特征交互和多尺度通道注意力的去雾网络

本文提出的去雾网络 CLMSA-Net 的整体框架如图 1 所示. 本文所设计的网络采用编码-解码的 U 型结构,主要包含编码器、解码器、跨层注意力特征交互模块和多尺度通道注意力模块. 其中,定义输入的有雾图像为  $I \in \mathbb{R}^{H \times W \times C}$ , 图像的高度、宽度和通道数分别由  $H, W, C$  来代表. 在 4 倍下采样的过程中,通过 1 个步幅为 1 的标准卷积和 2 个步幅为 2 的卷积来提取图像特征,从而得到不同层次的特征  $X_L (L=1, 2, 3)$ .

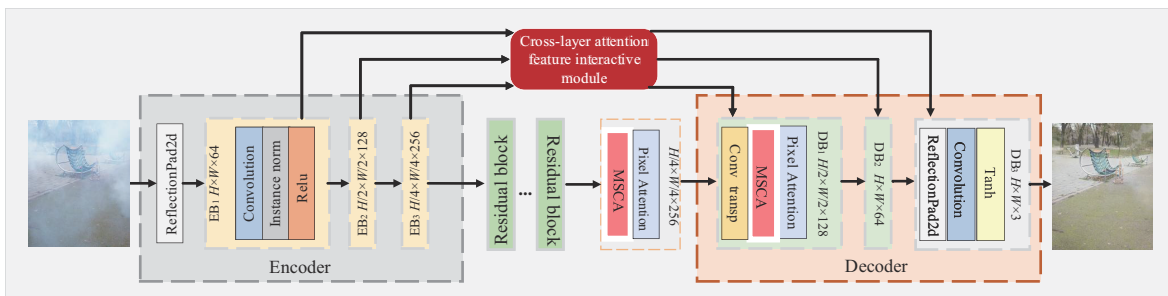


图 1 CLMSA-Net 整体框架

### 3.2 跨层注意力特征交互模块

大多数基于 U 型结构的去雾网络将编码层特征直接传递到对应尺度的解码层,忽略了不同层次特征信息的有效利用. 因此,为了更有效地利用编码阶段不同层次的信息,减少特征稀释的同时提升解码过程中重构去雾图像的效果,本文提出了跨层注意力特征交互模块如图 2 所示.

首先,对于编码器不同层次的特征  $X_L$  采用全局平均池化操作  $H_{\text{gap}}$  (Global Average Pooling) 进行压缩,得

到每一层特征图的通道描述符  $G_L \in [G_1, G_2, G_3]$ ,从而保留各层特征图最显著的特征. 其中,第  $L$  层得到的通道描述符表述为:  $(X_L)_n(i, j)$  表示第  $L$  个编码层中第  $n$  个单通道特征图在位置  $(i, j)$  处的值,特征图  $X_L$  的形状从  $H_L \times W_L \times C_L$  压缩到  $1 \times 1 \times C_L$ ,所得到的通道描述符对应的通道数为  $C_L \in [64, 128, 256]$ .

$$(G_L)_n = H_{\text{gap}}((X_L)_n) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (X_L)_n(i, j) \quad (1)$$

再将通道描述符在通道维度上进行 Concat 拼接:

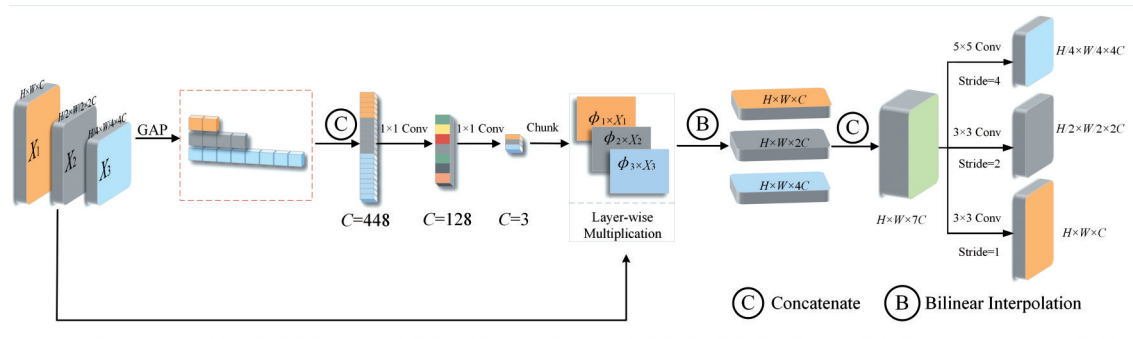


图2 跨层注意力特征交互模块

$$D_n = \text{Concat}(G_1, G_2, G_3) \quad (2)$$

其中,  $D_n$  代表拼接后的通道描述符. 再根据各编码层对聚合特征的重要性对各编码进行特征权值学习, 其中聚合的特征将高级语义信息以及低级边缘纹理信息相结合. 然后, 采用卷积-激活-卷积的方式来学习编码层对应的权值, 即

$$\varphi_L = \text{Chunk}(\text{Conv}(\partial(\text{Conv}(D_n)))) \quad (3)$$

其中,  $\text{Conv}$  表示卷积核为  $1 \times 1$  大小的 2D 卷积,  $\partial$  表示激活函数,  $\text{Chunk}$  表示将得到的  $1 \times 1 \times 3$  大小的一维向量拆分为 3 个权重  $\varphi_L$ . 再将得到的各层权值  $\varphi_L$  乘以原始层的特征  $(X_L)_n$ , 实现给各编码层分配权重的作用, 计算式如下:

$$Z_L = (X_L)_n \times \varphi_L \quad (4)$$

其中,  $Z_L$  表示各编码层分配了权重后的特征图. 为了能够将不同尺度大小的编码层融合到解码层中, 本文采用双线性插值函数  $B$  将特征图放大到相同尺寸并进行  $\text{Concat}$  操作得到聚合特征  $F$ . 其具体表示如下:

$$F = \text{Concat}(B(Z_1), B(Z_2), B(Z_3)) \quad (5)$$

接下来采用不同的卷积实现下采样操作, 将聚合特征下采样到对应解码层的尺寸使聚合特征融合到解码层的各层, 其中  $\text{Conv}_L(F)$  表示聚合特征  $F$  经过不同步长的卷积核的过程,  $Y_L$  表示解码器的第  $L$  层. 跨层注意力特征交互模块, 通过将编码阶段不同层次的特征信息进行聚合, 使其包含更丰富的细节纹理以及结构语义信息, 再将聚合特征下采样到对应解码层, 使聚合特征融合到解码层的各层以引导网络生成更高质量的去雾图像, 有效缓解了  $U$  型网络编解码过程中特征稀释的问题.

$$Y_L = \text{Concat}(\text{Conv}_L(F), Y_{L+1}) \quad (6)$$

跨层注意力特征交互模块能够整合编码阶段所提取的浅层纹理信息和深层语义信息, 用于无雾图像的重构. 然而, 现有的通道注意力机制如 SE 和 ECA 等, 由于感受野的限制, 使得各个解码层无法有效学习跨层注意力特征交互模块所聚合的边缘纹理、语义信息. 因此需要采用本文 3.3 节提出的多尺度通道注意力模块,

来聚合跨层注意力特征交互模块传递过来的多种特征信息, 从而进一步提升网络重构图像的性能.

### 3.3 多尺度通道注意力模块

传统的通道注意力模块只具有单一尺度的感受野<sup>[20, 26, 43]</sup>, 未能充分利用上下文信息, 这在一定程度上限制了注意力权重的有效分配. 为了改善这一情况, 本文设计了一种多尺度通道注意力模块, 利用不同空洞率的空洞卷积提取多尺度特征信息, 形成一个多尺度上下文并行学习的通道注意力机制. 其工作机制如图 3 所示, 主要包括分散-聚合两个阶段. 在分散的过程中, 使用不同尺度的空洞卷积获取多尺度特征表达. 在聚合的过程中, 将所获取的权值分别乘以对应的特征图, 并将拥有不同感受野的特征进行融合.

通过引入不同空洞率的空洞卷积, 并行地学习由跨层注意力特征交互模块传递的特征信息, 形成了一个并行学习的通道注意力机制, 能够更加有效地为去雾网络的特征分配权重. 首先, 对输入的特征图  $X$ , 利用不同空洞率  $r \in \{1, 2, 3\}$  的卷积核  $\text{Conv}_r$  进行卷积, 提取不同感受野的特征  $\text{SF}_r \in \{\text{SF}_1, \text{SF}_2, \text{SF}_3\}$ , 数学表达式为

$$\text{SF}_r = \text{Conv}_r(X), \quad r \in \{1, 2, 3\} \quad (7)$$

与现有的卷积相比, 空洞卷积在每个卷积核元素之间插入  $r-1$  行和列个 0, 不增大参数数量的同时扩大卷积核的感受野, 通过设置不同的空洞率, 获取不同大小的感受野, 从而拥有多尺度信息. 这种方法相较于单一尺度的通道注意力<sup>[20]</sup>可以捕获到更广阔、更丰富以及更复杂的上下文信息. 紧接着对拥有不同感受域的特征分别进行全局平均池化操作, 具体为

$$(\text{SG}_r)_n = H_{\text{gap}}((\text{SF}_r)_n) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (\text{SF}_r)_n(i, j) \quad (8)$$

其中,  $(\text{SF}_r)_n(i, j)$  表示经过空洞率为  $r$  的卷积后的第  $n$  个单通道特征图在位置  $(i, j)$  处的值;  $\text{SG}_r$  表示不同感受野特征对应的通道描述符,  $\text{SG}_r \in \{\text{SG}_1, \text{SG}_2, \text{SG}_3\}$ . SE 通道注意力采用全连接层学习各通道的权值, 但通道降维可能会导致一些重要特征丢失, 本文采用一维卷积进行通道权重的学习, 减少参数数量的同时避免了通道降

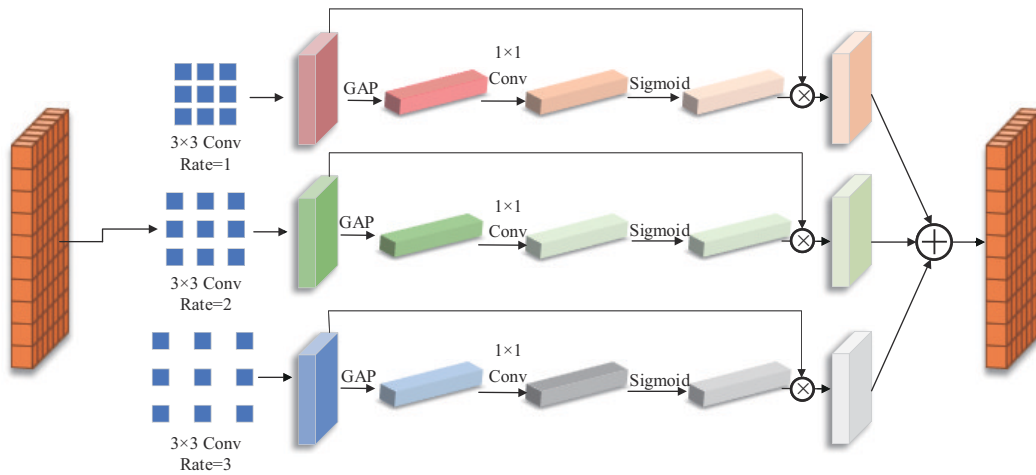


图3 多尺度注意力机制

维所带来的负面影响. 接着采用激活函数对各层权重进行归一化操作得到  $(M_r)_n$ , 表示为经过空洞率为  $r$  的卷积和通道注意力后得到的第  $n$  个通道权值. 其表达式如下:

$$(M_r)_n = \delta(\text{Conv1d}((\text{SG}_r)_n)) \quad (9)$$

其中, Conv1d 代表一维卷积,  $\delta$  表示 Relu 激活函数. 最后, 将所获取的权值分别乘以对应的特征, 并且将拥有不同感受野的特征进行融合, 以得到出不同感受域的通道特征  $\hat{X}$ .

$$\hat{X} = \sum_{n=1}^N ((\text{SF}_r)_n \times (M_r)_n) \quad (10)$$

因此, 相较于仅使用单一尺度的通道注意力机制, 本文提出的多尺度通道注意力模块能够获取不同大小感受野的特征图用于通道权重的学习, 使得各个解码层更有效地学习跨层特征交互模块所聚合的纹理、语义信息. 此外, 在 4.4 节中, 消融实验进一步验证了本文提出的多尺度通道注意力在图像去雾性能上优于现有的单尺度通道注意力.

### 3.4 损失函数

在神经网络学习过程中, 损失函数被引入以量化期望输出与模型预测输出之间的差异. 本文采用了 4 种不同的损失函数来指导网络训练, 包括 Smooth- $l_1$  损失、感知损失<sup>[36]</sup>、多尺度结构相似度损失以及对抗损失<sup>[48]</sup>. 本节将介绍所采用的损失函数.

#### 3.4.1 Smooth- $l_1$ 损失函数

本文采用 Smooth- $l_1$  损失来约束图像重构过程, 以确保生成图像与对应地面真实图像 (Ground Truth, GT) 之间具有较高的相似度. 表达式如下:

$$L_{\text{smooth-}l_1} = \begin{cases} 0.5(y - G(x))^2, & |y - G(x)| < 1 \\ |y - G(x)| - 0.5, & |y - G(x)| \geq 1 \end{cases} \quad (11)$$

其中,  $x$  和  $y$  分别代表有雾图像和 GT 图像,  $G$  表示去雾网络 CLMSA-Net,  $G(x)$  表示去雾后的图像. 当真实值

和预测值之间的差异较小时, 损失函数将采用均方误差, 而预测值和真实值相差较大时, 损失函数采用  $L_1$  范数形式进行计算.

#### 3.4.2 感知损失函数

感知损失广泛应用于图像分割、图像去噪和图像去雾等领域. 具体而言, 在各个特征层级中, 通过将去雾网络生成的图像和 GT 间的欧氏距离 ( $L_2$  范数) 计算作为损失, 实现网络生成的去雾图像在从低级纹理到高级语义的各个层次的有力约束. 使得生成的图像更好地匹配 GT 图像的纹理、细节与颜色. 表达式定义如下:

$$L_{\text{perc}} = \frac{1}{N} \sum_{i=1}^N \frac{1}{C_i H_i W_i} \|\varphi_i(y) - \varphi_i(x)\|_2^2 \quad (12)$$

其中,  $x$  和  $y$  分别代表有雾图像和 GT 图像,  $\varphi_i(\cdot)$  代表获取 VGG<sub>16</sub><sup>[49]</sup> 预训练模型第  $i$  层大小为  $C \times H \times W$  的特征图,  $\|\cdot\|_2^2$  表示  $L_2$  范数,  $N$  表示在感知损失中使用 VGG<sub>16</sub> 预训练模型特征层的数量.

#### 3.4.3 多尺度结构相似度损失函数

为了使生成器恢复的图像和 GT 图像具有更好的结构相似性, 本文引入了多尺度结构相似度损失 MS-SSIM 可定义为

$$L_{\text{ms-ssim}} = 1 - \prod_{m=1}^M \left( \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \right)^{\beta_m} \left( \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \right)^{\gamma_m} \quad (13)$$

其中,  $x$  代表生成图像,  $y$  代表 GT 图像,  $\mu_x$  和  $\mu_y$  分别代表生成图像和 GT 图像的均值,  $\sigma_x$  和  $\sigma_y$  分别代表生成图像和 GT 图像的标准差,  $\sigma_{xy}$  代表生成图像和 GT 图像之间的协方差,  $\beta_m$  和  $\gamma_m$  表示两项相对重要性,  $C_1$  和  $C_2$  为常数项.

#### 3.4.4 对抗损失函数

对抗损失在各去雾网络中被证明有效, 使用 GAN 的对抗性损失<sup>[48]</sup>, 其目的是区分 GT 图像和生成器生成的去雾图像. 本文对抗损失函数表达式如下:

$$L_{\text{adv}} = \sum_{n=1}^N -\log D(G(x)) \quad (14)$$

其中,  $G$  表示生成器,  $D$  表示判别器,  $D(G(x))$  表示去雾后的图像,  $G(x)$  判断为清晰图像的概率.

### 3.4.5 总损失函数

为了更好地训练本文提出的模型, 提出的总损失函数如式(15)所示:

$$L_{\text{total}} = \lambda_1 L_{\text{adv}} + \lambda_2 L_{\text{smooth-}l_1} + \lambda_3 L_{\text{perc}} + \lambda_4 L_{\text{ms-ssim}} \quad (15)$$

其中,  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$  为每个损失函数的超参数.

## 4 实验结果与分析

本节内容安排如下: 首先, 在 4.1 节中介绍了实验参数配置、所用数据集以及评估指标; 其次, 在 4.2 节中将展示本文算法和目前优秀的 12 种去雾算法在合成和真实数据集上的定量结果评析; 再次, 在 4.3 节中将从定性方面展示本文算法与其他 11 种当前较优去雾算法的比较结果; 最后, 在 4.4 节中通过消融实验验证了本文算法中的各个模块的作用和贡献.

### 4.1 实验配置

#### 4.1.1 参数设置

实验使用显存为 24 GB 的 RTX 3090 显卡, 代码基于 Pytorch 框架实现, CUDA (Compute Unified Device Architecture) 版本为 11.0. 选择 Adam 优化策略来调整网络, 其中动量衰减参数  $\beta_1=0.9$  和  $\beta_2=0.999$ . 而损失函数的超参数  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$  分别设置为 1.0、0.5、0.01、0.5%. 在训练过程中, 输入到网络的图像尺寸为  $256 \times 256$ . 为了扩充训练集, 防止模型过拟合, 还引入了  $90^\circ$ 、 $180^\circ$  和  $270^\circ$  旋转数据增强.

#### 4.1.2 数据集

本文选择了 RESIDE 合成有雾数据集<sup>[27]</sup>、Dense-Haze 真实浓雾数据集<sup>[28]</sup>、NH-Haze21 真实非均匀雾数据集<sup>[29]</sup>以及遥感数据集 StateHaze-1k<sup>[30]</sup>来评估本文的模型. 其中, 在 RESIDE 数据集中, 采用室外训练集 OTS (Outdoor Training Set) 对网络进行训练, SOTS (Synthetic Objective Testing Set) 的 Outdoor 作为测试集. Dense-Haze 包含 45 张浓雾数据集. NH-Haze21 包含 25 张非均匀有雾图像, 由于其验证集和测试集 GT 图像还未公布, 选取前 20 张作为训练集, 剩余 5 张作为测试集用于评估. 公开遥感数据集 StateHaze-1k 包含 3 个子数据集 (Thin fog, Moderate fog, Thick fog), 不同雾的浓度的每个子训练集包含 320 张图像, 验证集 35 张, 测试集 45 张. 由于真实有雾数据集和遥感数据集较少, 将数据集中的图片随机裁剪到 1 024、512、256 的大小, 并且在利用双线性插值的方式将图片 resize 到 256 的尺度扩充数据集.

### 4.1.3 评价指标

为评估模型性能, 使用峰值信噪比 PSNR (Peak Signal to Noise Ratio) 和结构相似性指数 SSIM (Structural Similarity Index) 为定量评估评价指标. 这 2 项指标常用于去雾图像的质量评价. 图像质量指数 UIQI (Universal Image Quality Index) 是一种考虑图像失真和丢失信息的综合图像质量评价指标. 它结合了图像的相关性、亮度失真和对比度失真三个方面, 值域在  $[-1, 1]$  之间, 值越大表示图像质量越好. 均方根误差 RMSE (Root Mean Squared Error) 反映重构图像与原始图像之间的偏差程度, 它计算图像所有像素点灰度误差的平方根, 值越小表示重构图像与原图越接近, 图像质量越好.

### 4.2 定量分析

本节将展示本文算法与其他 12 种优秀去雾算法之间的比较. 其中包括 DCP<sup>[9]</sup>、AOD-Net、EPDN (Enhanced Pix2pix Dehazing Network)<sup>[40]</sup>、GCA-Net (Gated Context Aggregation Network)<sup>[50]</sup>、MSBDN<sup>[17]</sup>、FFA<sup>[26]</sup>、AECR<sup>[19]</sup>、TBN (a Two-Branch Neural network for non-homogeneous dehazing via ensemble learning)<sup>[51]</sup>、SGID (Self-Guided Image Dehazing using progressive feature fusion)<sup>[52]</sup>、Huang.SAR (single satellite optical imagery dehazing using SAR image prior based on conditional generative adversarial networks)<sup>[30]</sup>、TMS-GAN (a Two-fold Multi-Scale Generative Adversarial Network for single image dehazing)<sup>[53]</sup> 和 ITB<sup>[54]</sup>. 表 1 和表 2 展示了各去雾算法在合成雾数据集 SOTS-Outdoor、浓雾数据集 Dense-Haze、非均匀雾数据集 NH-Haze 以及遥感数据集 StateHaze-1k 的测试结果. 表 1 列出了 PSNR 和 SSIM 指标上的性能, 表 2 展示了 UIQI 和 RMSE 指标上的表现. 随后, 表 3 和表 4 继续展示了相同数据集上的测试结果. 表 3 提供了 PSNR 和 SSIM 指标的性能, 表 4 则涵盖了 UIQI 和 RMSE 指标的结果.

在自然有雾数据集上 PSNR 和 SSIM 的结果分析如表 1 所示, 在真实浓雾数据集 Dense-Haze 和真实非均匀雾数据集 NH-Haze 中, 本文所提算法在 PSNR 和 SSIM 上分别达到了 17.42 dB、0.603 和 22.78 dB、0.845, 并取得了这 2 个数据集上最好的性能结果. 在 RESIDE 室外合成测试集 SOTS 中, 本文算法相较于其他算法取得了最高的 PSNR 指标和第二的 SSIM 指标, 相较于 FFA 和 MSBDN, 本文提出的去雾算法在 PSNR 上分别提高了 0.13 dB 和 1.54 dB. 由于 FFA 没有下采样的过程, 保留了去雾图像的结构信息, 所以本文去雾算法在 SSIM 上略低于 FFA 算法. 相较于 ITB 和 TMS-GAN 算法, 本文算法在 SSIM 和 PSNR 上均有所提升, 例如在 Dense-Haze 数据集上, PSNR 和 SSIM 分别提升了 1.11 dB、0.042 和 1.27 dB、0.077. 因此, 本文算法综合自然合成有雾数据集、真实

表 1 在自然场景数据集上 PSNR 和 SSIM 的对比

Method	SOTS-Outdoor		Dense-Haze		NH-Haze	
	PSNR /dB	SSIM	PSNR /dB	SSIM	PSNR /dB	SSIM
(TPAMI 10) DCP	15.55	0.815	10.85	0.404	11.30	0.605
(ICCV 17) AOD-Net	19.63	0.861	13.30	0.469	13.22	0.613
(CVPR 19) EPDN	22.57	0.863	16.24	0.536	18.35	0.784
(WACV19)GCANet	21.71	0.891	13.16	0.483	14.84	0.446
(CVPR 20)MSBDN	32.16	0.976	14.84	0.494	19.63	0.804
(AAAI 20) FFA	33.57	0.984	16.26	0.545	20.40	0.806
(CVPR 21) AECR-Net	30.90	0.968	15.10	0.451	18.64	0.750
(CVPRW 21) TBN	30.56	0.967	16.36	0.582	21.66	0.843
(Tip 22)SGID	30.20	0.975	13.09	0.519	16.42	0.682
(TCSVT 22)TMS-GAN	32.58	0.964	16.15	0.526	20.94	0.811
(CVPRW 23) ITB	32.66	0.970	16.31	0.561	21.67	0.838
Ours	33.70	0.977	17.42	0.603	22.78	0.845

浓雾场景数据集以及真实非均匀雾场景数据集, 本文的算法在综合性能上达到了 SOTA 性能。

表 2 在自然场景数据集上 UIQI 和 RMSE 的对比

Method	SOTS-Outdoor		Dense-Haze		NH-Haze	
	UIQI	RMSE	UIQI	RMSE	UIQI	RMSE
(TPAMI 10) DCP	0.844	9.66	0.234	10.34	0.354	10.23
(ICCV 17) AOD-Net	0.951	9.34	0.317	10.25	0.531	10.01
(CVPR 19) EPDN	0.946	9.76	0.678	10.09	0.705	9.91
(WACV19)GCANet	0.798	8.34	0.497	10.35	0.530	10.20
(CVPR 20)MSBDN	0.980	4.99	0.537	10.13	0.906	9.57
(AAAI 20) FFA	0.989	5.01	0.669	10.01	0.898	9.59
(CVPR 21) AECR-Net	0.900	6.29	0.443	10.11	0.840	10.01
(CVPRW 21) TBN	0.901	6.09	0.688	9.94	0.935	9.21
(Tip 22)SGID	0.899	6.23	0.467	10.15	0.820	9.57
(TCSVT 22)TMS-GAN	0.974	5.88	0.679	10.11	0.924	9.47
(CVPRW 23) ITB	0.991	5.13	0.683	9.90	0.932	9.33
Ours	0.996	4.97	0.727	9.85	0.937	8.82

表 2 展示了在自然场景数据集上, UIQI 和 RMSE 等 2 种评估指标的比较结果. 在 RESIDE 室外合成测试集 SOTS 中, 本文提出的算法在 UIQI 和 RMSE 指标上均获得了最优的表现, 达到了 0.996 和 4.97, 相比于 ITB 分别提升和降低了 0.005 和 0.16. 同样, 在自然数据集中的真实浓雾数据集 Dense-Haze 和真实非均匀雾数据集 NH-Haze 上, 本文算法在 UIQI 和 RMSE 指标上都取得了最优的结果, 分别达到了 0.727、9.85 和 0.937、8.82. 在真实浓雾数据集 Dense-Haze 的 UIQI 指标上相比于 ITB 和 TMS-GAN 分别提升了 0.044 和 0.048. 其中在真实非均匀雾数据集 NH-Haze 上, 相较于 ITB 和 TMS-GAN 算法, 本文算法在 UIQI 指标上分别提高了 0.005 和 0.013, 在 RMSE 上分别降低了 0.51 和 0.65. 证明了

CLMSA-Net 算法在去雾任务上的有效性.

在真实数据集上的结果分析: 表 3 展示了各去雾算法在遥感数据集上的 PSNR 和 SSIM 性能指标, 并且在对比算法中添加了 Huang. SAR 遥感去雾算法, 由于该算法并未开源代码, 本文只对其进行了定量分析的比较. 此外, Huang 等人提出的算法在去雾网络中引入多孔径雷达图像作为辅助信息, 该数据包含许多图像细节信息, 可以有效地帮助去雾算法恢复去雾图像, 但是作者提供的多孔径雷达图像有误(多孔径雷达图像和 RGB (Red, Green, Blue) 图像没有对应关系). 本文算法在未引入多孔径雷达数据的情况下, 在薄雾和中浓雾上 PSNR 和 SSIM 分别达到了 25.67 dB、0.921 和 27.55 dB、0.946, 取得了这 2 项数据子集的最优性能. 在遥感薄雾数据集和中浓雾数据集上相较于 Huang. SAR、TBN, 本文算法在 PSNR 和 SSIM 分别提升了 1.51 dB、0.015、2.24 dB、0.02 和 1.15 dB、0.024、0.13 dB、0.002. 本文在遥感薄雾数据集上与 TMS-GAN 算法相比, PSNR 提升了 2.22 dB, 在 SSIM 上提升了 0.012. 与 ITB 相比, 在遥感薄雾和中等雾数据集上, 本文算法在 PSNR 和 SSIM 分别提升了 1.06 dB、0.009 和 0.07 dB、0.002. 由于浓雾会引起图像信息的严重退化, 从而导致去雾难度更大. Huang 等人提出的算法利用了多孔径雷达数据, 该数据包含图像更多的细节纹理以及结构信息, 有利于图像细节的构建. 本文的算法在遥感浓雾数据子集上 PSNR 和 SSIM 达到了第二的性能, 依然优于其他算法.

表 4 表示各去雾算法在遥感数据集不同雾浓度下的 UIQI 和 RMSE 指标比较结果显示. 本文算法在薄雾、表 3 在 StateHaze-1k 遥感数据集不同雾浓度数据集上 PSNR 和 SSIM 的对比

Method	Thin fog		Moderate fog		Thick fog	
	PSNR /dB	SSIM	PSNR /dB	SSIM	PSNR /dB	SSIM
(TPAMI 10) DCP	13.45	0.701	9.78	0.591	10.89	0.572
(ICCV 17) AOD-Net	18.62	0.851	17.91	0.882	15.21	0.739
(CVPR 19) EPDN	21.74	0.882	24.05	0.910	19.73	0.786
(WACV19)GCANet	18.71	0.791	20.58	0.799	17.73	0.724
(CVPR 20) MSBDN	20.09	0.832	22.94	0.873	18.67	0.733
(AAAI 20) FFA	23.75	0.903	26.50	0.941	22.03	0.840
(CVPR 21) AECR-Net	22.90	0.879	24.03	0.879	20.61	0.812
(CVPR 21) TBN	24.52	0.897	27.42	0.944	21.56	0.837
(Tip 22) SGID	23.34	0.907	23.95	0.935	19.20	0.823
Huang.w/o SAR	21.74	0.816	22.09	0.827	22.12	0.784
(WACV 20)Huang.SAR	24.16	0.906	25.31	0.926	25.07	0.864
(TCSVT 22)TMS-GAN	23.45	0.909	26.13	0.931	21.85	0.831
(CVPRW 23) ITB	24.61	0.912	27.48	0.944	22.10	0.839
Ours	25.67	0.921	27.55	0.946	23.37	0.863

中雾和浓雾 3 个不同浓度遥感数据集上的 UIQI 分别达到 0.976、0.979 和 0.957, 相比于 TMS-GAN 分别提升了

表 4 在 StateHaze-1k 遥感数据集不同雾浓度数据集上 UIQI 和 RMSE 的对比

Method	Thin fog		Moderate fog		Thick fog	
	UIQI	RMSE	UIQI	RMSE	UIQI	RMSE
(TPAMI 10) DCP	0.762	10.01	0.594	10.19	0.608	10.09
(ICCV 17) AOD-Net	0.859	9.64	0.822	9.96	0.724	9.98
(CVPR 19) EPDN	0.936	9.31	0.944	8.63	0.887	9.68
(WACV19)GCANet	0.948	9.26	0.932	9.41	0.922	9.44
(CVPR 20) MSBDN	0.927	9.53	0.948	9.14	0.887	9.68
(AAAI 20) FFA	0.968	8.37	0.973	7.96	0.946	9.06
(CVPR 21) AECR-Net	0.953	9.06	0.947	8.67	0.915	9.60
(CVPR 21) TBN	0.969	8.46	0.977	7.72	0.938	9.18
(Tip 22) SGID	0.963	8.86	0.972	8.97	0.939	9.62
(TCSVT 22) TMS-GAN	0.968	9.02	0.967	8.24	0.939	9.23
(CVPRW 23) ITB	0.974	8.41	0.978	7.42	0.945	9.10
Ours	0.976	8.30	0.979	7.37	0.957	8.80

0.008、0.012 和 0.018, 相比于 TBN 分别提升了 0.007、0.002 和 0.019. 本文算法的 RMSE 分别为 8.30、7.37 和 8.80. 相比于 ITB, 本文算法在薄雾、中雾和浓雾 3 个不同浓度上 UIQI 分别提升了 0.002、0.001 和 0.012, RMSE 分别降低了 0.11、0.05 和 0.3. Huang 等人在其算法中引入了多孔径雷达图像作为辅助信息, 可以提供额外的图像细节帮助去雾. 但本文算法直接在遥感数据集上进行训练和测试, 在没有辅助数据的前提下, 本文算法也取得了较好的去雾效果, 其 UIQI 和 RMSE 指标均优于对比算法.

### 4.3 定性分析

本节展示了本文算法与其他 11 种优秀去雾算法的定性分析. 由于 Huang.SAR 遥感去雾算法未开源代码, 不能获取该算法的结果, 因此只对该算法进行了定量分析.

图 4 对各算法在 SOTS 室外合成测试集 Reside-Outdoor 上的效果进行了可视化展示. DCP 算法在处理天空区域时容易产生曝光过度 and 颜色失真的问题, 这一现象可以归因于基于物理先验的方法在无约束条件下估计大气光或传输图参数时存在较大的误差, 从而对生成的去雾图像产生负面影响.

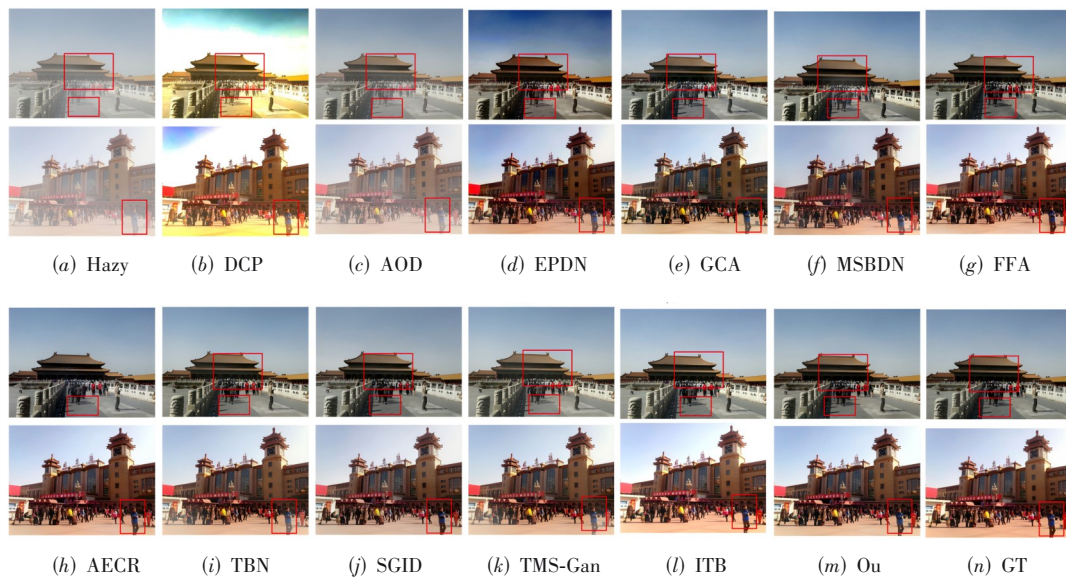


图 4 Reside-Outdoor 合成数据集可视化结果

同样受大气散射模型的限制, AOD 去雾算法没有达到良好的去雾效果, 在去雾图片上仍存在大量雾霾 (图 4(c)). EPDN 相较于 AOD 拥有良好的去雾效果, 但是去雾后生成的图像颜色过饱和, 例如图 4(d) 中的天空区域和钟楼部分, 这是因为, EPDN 算法采用 U 型网络结构, 缺乏浅层与深层的特征交互. MSBDN 和 FFA 算法在合成雾图数据集上都能够生成高质量的去雾图

像, 但仍有一些区域存在问题. TMS-Gan 去雾算法拥有良好的去雾效果, 但在去雾图像上依旧存在雾残留的问题. ITB 算法的生成结果与 TBN 算法类似, 存在颜色偏浅的问题. 与上述算法相比, 本文算法在效果和图像细节、颜色上更接近 GT 图像. 由于本文算法对各层次特征充分交互, 并对通道注意力机制进行优化, 从而筛选出更有助于去雾的特征.

在 Dense-Haze 浓雾数据集上的可视化结果如图 5 所示,在浓雾场景下,图像存在严重的退化现象,因

此将浓雾图像恢复成清晰图像在去雾领域仍是一大挑战.

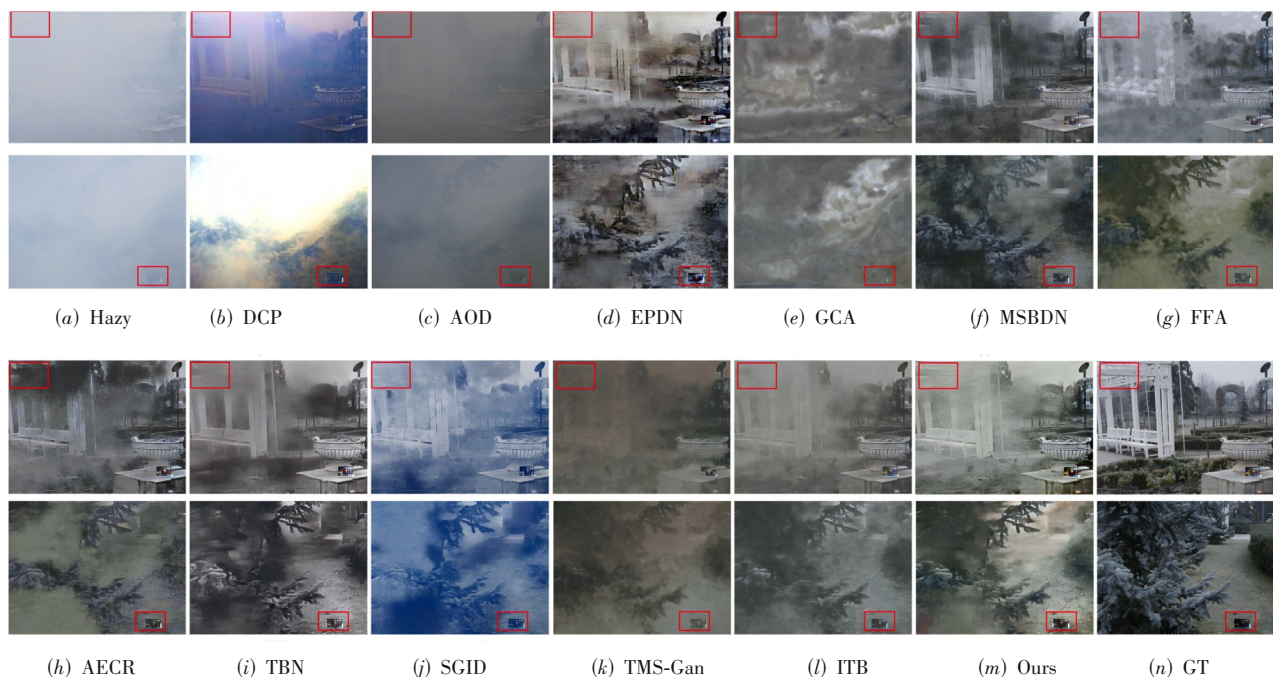


图 5 Dense-haze 浓雾数据集可视化结果

例如,DCP算法主要依赖暗通道估计参数,但其在无约束条件下的假设往往并不准确,导致去雾效果较差.如图5(b)所示,从DCP恢复的图像可见存在显著的曝光偏差和颜色失真.AOD-Net在浓雾数据集上的去雾效果也相对有限,如在图5(c)中可以观察到,它未能完全恢复图像的原始信息且去雾图像场景信息存在大量雾霾.EPDN去雾算法利用多尺度生成器与多尺度判别器指导网络在不同分辨率上进行去雾,达到一定的去雾效果,但其输出的图像在颜色饱和度上仍有不足.MSBDN去雾算法相较于GCA-Net恢复图像包含更多纹理细节信息,MSBDN融合了编码层和解码层各自的多尺度特征信息,但是其生成结果仍存在一些伪影,例如图5(f)左上角存在部分伪影.图5(g)中,FFA算法采用了没有下采样过程的直筒去雾结构,并通过叠加通道注意力与像素注意力达到良好的去雾效果,使生成的图像保留了良好的结构信息,但FFA的直筒型结构可能在捕获图像特征和颜色信息方面存在局限性,导致恢复图像的颜色与GT图像有所偏差.TBN算法基于集成学习的思想,在网络中加载预训练模型取得了良好的去雾效果,但其去雾图像与GT图像相比整体色调偏白(如图5(i)所示).如图5(k)所示,TMS-Gan算法恢复的图像存在色调偏红的问题.ITB算法在细节的恢复上依旧存在不足,去雾效果不

佳.本文算法在细节、纹理和颜色方面使恢复图像更接近于GT图像,但对浓雾图像,还存在进一步优化空间.

图6展示了本文算法在非均匀雾数据集NH-Haze21上的去雾可视化结果.DCP算法生成的去雾图像存在失真,并在某些区域出现过度曝光,如图6(b)所示.AOD算法恢复图像部分区域存在大量雾霾,如图6(c)所示.与基于先验预估的网络相比,EPDN表现更出色,但显示仍存在问题,如图6(d)所示,其去雾图像在某些区域仍有雾霾残留和伪影.GCA-Net算法可以恢复出图像的大致轮廓,但去雾图像仍包含大量伪影.

MSBDN、FFA、TBN算法能够达到良好的去雾效果,但是去雾图像仍存在部分问题,例如图6(f)中MSBDN去雾图像和图6(g)中FFA去雾图像在门框旁仍存在部分雾霾.另外,在图6(i)TBN恢复图像中,其颜色趋向于暗绿,这是由于TBN同样采用U型架构作为网络的一部分,未融合多层次特征.SGID存在颜色失真问题,并且去雾图像颜色偏蓝.本文提出网络恢复图像包含更多的图像细节、纹理以及颜色光照信息.图6(k)所示,TMS-Gan算法大部分还是存在雾霾残留.ITB也存在少量的雾残留,如图6(l)房屋门区域和草坪区域.相较于以上算法,本文算法在细节、颜色以及光照上更接近GT图像.

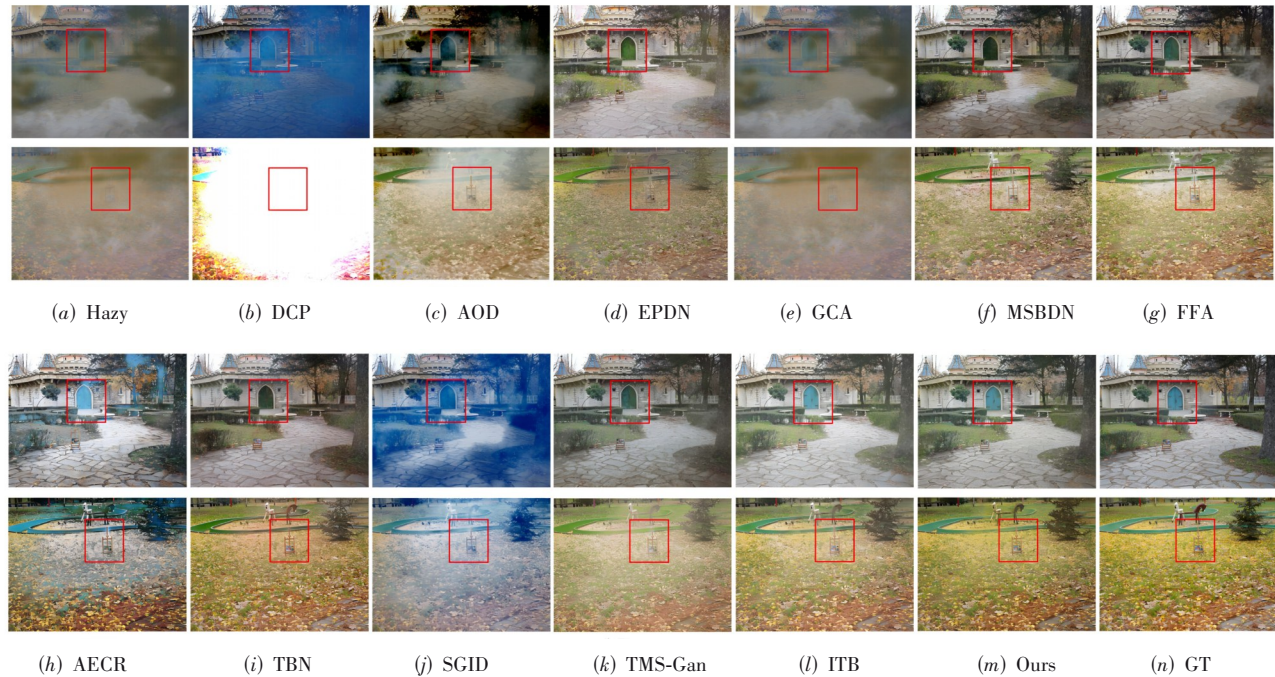


图6 NH-haze非均匀雾数据集可视化结果

图7展示了在不同雾霾浓度下的遥感数据集的去雾效果。在薄雾环境下(图7(b)),DCP存在色彩失真的问题,AOD相较于DCP具有较好的去雾效果,但图像中仍存在大量的雾霾.EPDN能有效地去除大部分雾霾,但其去雾图像与真实图像(GT)仍有偏差.MSBDN和FFA能取得较好的去雾效果,但生成的图像会出现少量的伪影.TBN算法在薄雾环境下表现出良好的去雾效果,但其生成的图像存在少量雾霾,并且颜色略显偏白.SGID算法去雾后的图像存在颜色失真的问题.如图7(k)所示,TMS-Gan算法所生成的图像存在雾残留的现象,同时色调与GT图像对比也存在偏浅的问题.ITB也存在少量雾残留的现象.

在中等雾环境下(图7(b)),由于过度依赖先验知识,DCP生成的图像出现曝光现象.AOD去雾后的图像仍有大量雾霾,颜色偏暗.MSBDN通过在编码和解码层利用多尺度信息,使去雾图像细节更加丰富,但部分区域仍会出现伪影.SGID算法存在图像失真以及颜色偏绿的问题,这是由于在编码过程中,SGID算法利用了有雾图像的特征,其中可能包含部分有雾信息,且该算法受限于大气散射模型.如图7(k)和图7(l)所示,框选的TMS-Gan算法的图片中存在明显的雾残留现象,而ITB存在少量雾残留的现象.

遥感图像通常在高空拍摄,浓雾会导致图像丢失大量细节信息,从而造成严重的图像退化.因此,在浓雾环境下,由于对先验知识预测错误(图7(b)),DCP的效果较差,图像出现严重失真.AOD恢复的图像中仍然

残留有大量雾霾.相较于AOD,EPDN和GCA在去雾效果上表现得更好,但部分区域仍有雾霾存在,且EPDN去雾算法恢复的图像颜色过饱和,与GT图像相差较大.MSBDN算法在浓雾测试集上的去雾效果不佳,SGID同样出现了图像颜色失真且颜色偏绿的问题.如图7(k)和图7(l)所示,TMS-Gan和ITB算法存在少量的雾残留现象和颜色失真的问题.相较于上述算法,本文算法在不同浓度的雾霾下都取得了较好的视觉效果,细节纹理和颜色恢复更接近GT图像,但还需要进一步改进.

#### 4.4 消融实验

为了评估本文模型各模块的有效性,本节根据框架和创新点设计消融实验,其中包括5个实验.(1) Base: U-Net体系架构,主要由编码层、解码层和6个残差块组成.值得注意的是,编码层和解码层通过跳跃连接直接连接.(2) Base+CLA:在U-Net基础框架上将中间的跳跃连接改进,添加跨层注意力特征交互模块(CLA).(3) Base+CLA+CA:在实验(2)的基础上添加通道注意力.(4)在实验(2)的基础上添加ECA通道注意力.(5)在实验(2)的基础上添加MSCA注意力机制.通过添加不同的注意力,以验证本文提出的注意力模型优于其他通道注意力模型.

由表5可知,Base框架在PSNR和SSIM上为29.55 dB、0.963,通过添加跨层注意力特征交互模块,PSNR可以达到31.42 dB,SSIM达到0.964.相较于基准模型PSNR提升了1.87 dB,SSIM提升了0.001.模型

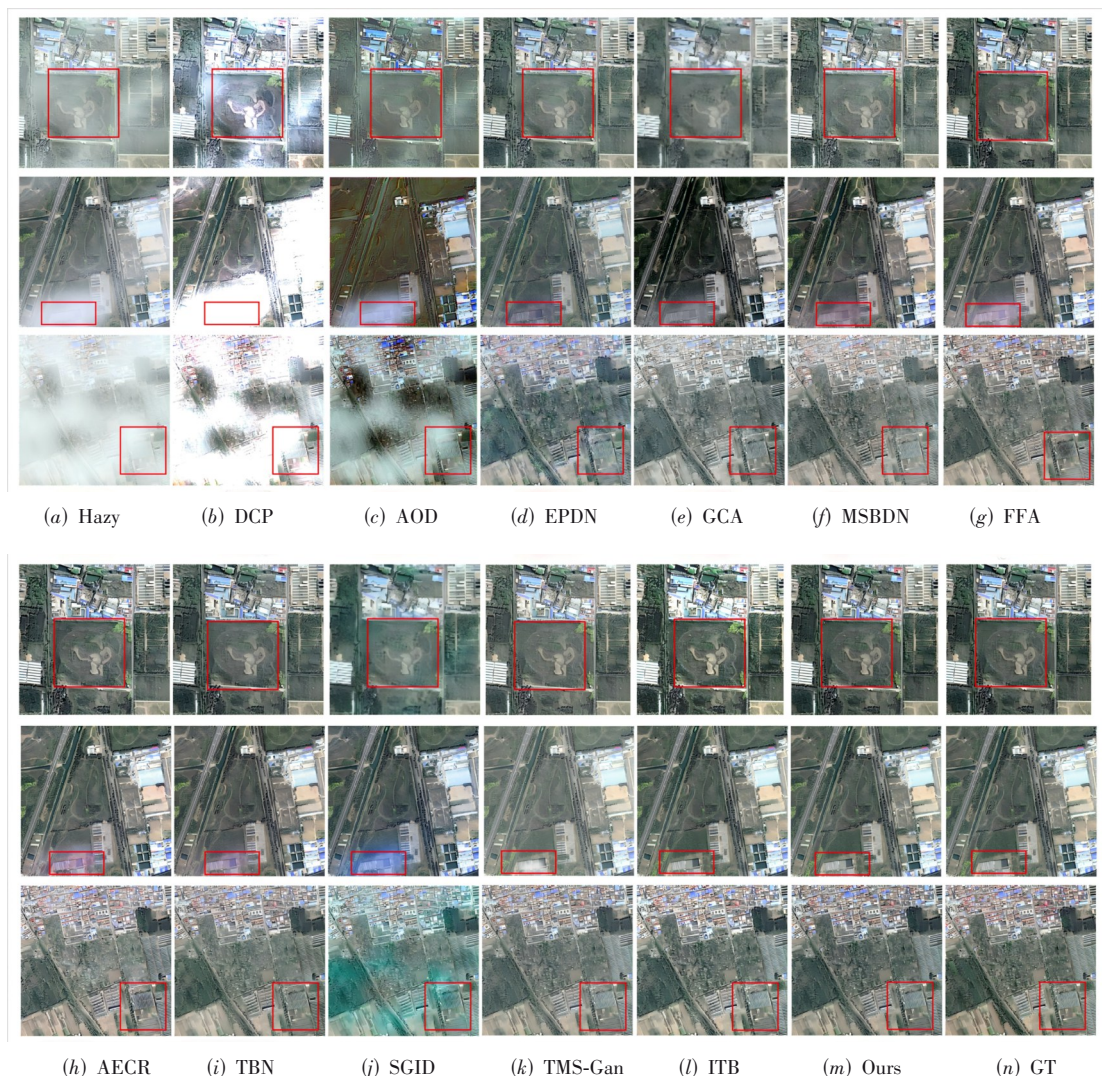


图7 StateHaze-1k 遥感数据集薄-中-厚雾可视化结果

在添加了通道注意力之后,性能得到进一步提升. 本文的注意力模块考虑到注意力机制感受野受限的问题,同时通过ECA通道注意力的启发采用一维卷积作为中间过程以减少模型的参数量,设计了多尺度通道注意力MSCA. 与添加CA相比,本节的注意力模块MSCA分别将PSNR和SSIM提高了0.18 dB、0.002. 与ECA相比,本文的注意力模块在PSNR和SSIM上提升了0.17 dB、0.002,验证了所提注意力模块的有效性.

本文以网络模型中解码层的第一层为例,将该层的256个特征通道划分为4行64列作为样例进行可视化,并通过色条展示了3种机制对每个通道权重的分配和分布范围. 如图8所示,CA、ECA的权值谱图中突出显示和抑制的特征通道数量基本保持一致,这是因为CA和ECA存在感受野受限问题,无法获取有效的上下

表5 在SOTS-Outdoor数据集上PSNR和SSIM结果

Model	PSNR/dB	SSIM
Base	29.55	0.963
Base+CLA	31.42	0.964
Base+CLA+CA	33.52	0.975
Base+CLA+ECA	33.53	0.975
Base+CLA+MSCA	33.70	0.977

文信息用于通道权重的学习. 相比而言,MSCA通过不同空洞率的空洞卷积来捕捉多尺度上下文信息用于学习更有效的通道权重分配. 如图8所示,本文提出的MSCA注意机制在权值谱图中只有少数通道被分配了较大的权值,与CA和ECA相比更符合稀疏原理<sup>[55,56]</sup>.

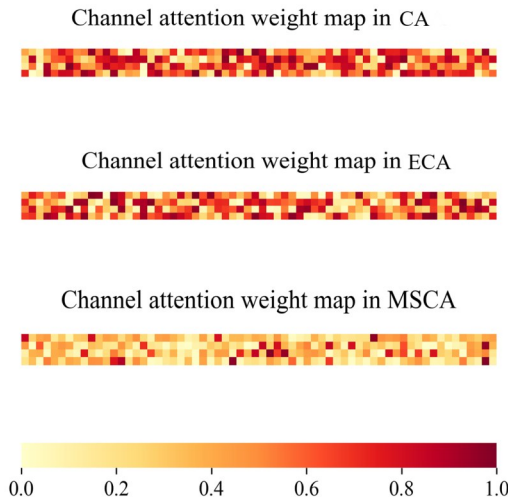


图 8 MSCA通道注意力权重图

## 5 结论

由于雾霾对视觉内容的遮挡,图像去雾在计算机视觉领域仍是一个具有挑战性的研究话题.本文提出了一种基于跨层注意力特征交互和多尺度通道注意力的去雾网络 CLMSA-Net,网络结构主要包括跨层注意力特征交互模块和多尺度通道注意力模块.跨层注意力特征交互模块根据各编码层对聚合特征的重要性自适应地学习各层次权重,以取得良好的聚合特征,并且将聚合特征分别融合进对应解码层以缓解U型网络特征稀释的问题.多尺度通道注意力模块结合不同空洞率的空洞卷积来获取拥有不同感受野的特征通道,充分利用上下文信息以探索对于图像去雾至关重要的特征通道信息.通过与当前主流去雾算法的对比实验,证明了所提出的 CLMSA-Net 具有出色的有效性和鲁棒性.

### 参考文献

- [1] KALWAR S, PATEL D, AANEGOLA A, et al. GDIP: Gated differentiable image processing for object detection in adverse conditions[C]//2023 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE, 2023: 7083-7089.
- [2] 张智, 易华挥, 郑锦. 聚焦小目标的航拍图像目标检测算法[J]. 电子学报, 2023, 51(4): 944-955.  
ZHANG Z, YI H H, ZHENG J. Focusing on small objects detector in aerial images[J]. Acta Electronica Sinica, 2023, 51(4): 944-955. (in Chinese)
- [3] LI S Y, FISCHER T, KE L, et al. OVTrack: Open-vocabulary multiple object tracking[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 5567-5577.
- [4] CHEN X, PENG H W, WANG D, et al. SeqTrack: Sequence to sequence learning for visual object tracking[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 14572-14581.
- [5] WAN J, XI H, ZHOU J, et al. Robust and precise facial landmark detection by self-calibrated pose attention network[J]. IEEE Transactions on Cybernetics, 2023, 53(6): 3546-3560.
- [6] YANG X, LIU C, XU L L, et al. Towards effective adversarial textured 3D meshes on physical face recognition[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 4119-4128.
- [7] CANTOR A. Optics of the atmosphere: Scattering by molecules and particles[J]. IEEE Journal of Quantum Electronics, 1978, 14(9): 698-699.
- [8] BERMAN D, TREIBITZ T, AVIDAN S. Non-local image dehazing[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 1674-1682.
- [9] 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, 2011, 33(12): 2341-2353.
- [10] 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.
- [11] ZHANG H, PATEL V M. Densely connected pyramid dehazing network[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 3194-3203.
- [12] LI R D, PAN J S, HE M, et al. Task-oriented network for image dehazing[J]. IEEE Transactions on Image Processing, 2020, 29: 6523-6534.
- [13] 高继蕊, 李华锋, 张亚飞, 等. 双注意力引导的细节和结构信息融合图像去雾网络[J]. 电子学报, 2023, 51(1): 160-171.  
GAO J R, LI H F, ZHANG Y F, et al. Dual attention-guided detail and structure information fusion network for image dehazing[J]. Acta Electronica Sinica, 2023, 51(1): 160-171. (in Chinese)
- [14] SUN H, LI B H, DAN Z P, et al. Multi-level feature inter-

- action and efficient non-local information enhanced channel attention for image dehazing[J]. *Neural Networks*, 2023, 163: 10-27.
- [15] ZUNAIR H, BEN HAMZA A. Sharp U-Net: Depthwise convolutional network for biomedical image segmentation[J]. *Computers in Biology and Medicine*, 2021, 136: 104699.
- [16] 但志平, 方帅领, 孙航, 等. 基于双判别器异构 CycleGAN 框架下多阶通道注意力校准的室外图像去雾[J]. *电子学报*, 2023, 51(9): 2558-2571.
- DAN Z P, FANG S L, SUN H, et al. Outdoor image dehazing based on multi-order channel attention calibration using a dual-discriminator heterogeneous CycleGAN framework[J]. *Acta Electronica Sinica*, 2023, 51(9): 2558-2571. (in Chinese)
- [17] DONG H, PAN J S, XIANG L, et al. Multi-scale boosted dehazing network with dense feature fusion[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 2154-2164.
- [18] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 2261-2269.
- [19] WU H Y, QU Y Y, LIN S H, et al. Contrastive learning for compact single image dehazing[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 10546-10555.
- [20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
- [21] WEI X C, ZHANG Y, ZHENG Y H. BSFCDet: Bidirectional spatial-semantic fusion network coupled with channel attention for object detection in satellite images[J]. *Remote Sensing*, 2023, 15(13): 3213.
- [22] YANG X, LV Z Y, ATLI BENEDIKTSSON J, et al. Novel spatial-spectral channel attention neural network for land cover change detection with remote sensed images[J]. *Remote Sensing*, 2022, 15(1): 87.
- [23] LI H F, QIU K J, CHEN L, et al. SCAAttNet: Semantic segmentation network with spatial and channel attention mechanism for high-resolution remote sensing images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 18(5): 905-909.
- [24] ZHONG Z L, LIN Z Q, BIDART R, et al. Squeeze-and-attention networks for semantic segmentation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 13062-13071.
- [25] LI J, LIU K Y, HU Y T, et al. Eres-UNet++: Liver CT image segmentation based on high-efficiency channel attention and Res-UNet++[J]. *Computers in Biology and Medicine*, 2023, 158: 106501.
- [26] QIN X, WANG Z L, BAI Y C, et al. FFA-net: Feature fusion attention network for single image dehazing[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 11908-11915.
- [27] LI B Y, REN W Q, FU D P, et al. Benchmarking single image dehazing and beyond[J]. *IEEE Transactions on Image Processing*, 2018, 28(1): 492-505.
- [28] ANCUTI C O, ANCUTI C, SBERT M, et al. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images[C]//2019 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2019: 1014-1018.
- [29] ANCUTI C O, ANCUTI C, TIMOFTE R. NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2020: 1798-1805.
- [30] HUANG B H, LI Z, YANG C, et al. Single satellite optical imagery dehazing using sar image prior based on conditional generative adversarial networks[C]//2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2020: 1795-1802.
- [31] TANG K T, YANG J C, WANG J. Investigating haze-relevant features in a learning framework for image dehazing[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 2995-3002.
- [32] REN W Q, LIU S, ZHANG H, et al. Single image dehazing via multi-scale convolutional neural networks[[M]]// *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2016: 154-169.
- [33] LI Y N, MIAO Q G, OUYANG W L, et al. LAP-net: Level-aware progressive network for image dehazing[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 3275-3284.

- [34] REN W Q, PAN J S, ZHANG H, et al. Single image dehazing via multi-scale convolutional neural networks with holistic edges[J]. *International Journal of Computer Vision*, 2020, 128(1): 240-259.
- [35] LI B Y, GOU Y B, GU S H, et al. You only look yourself: Unsupervised and untrained single image dehazing neural network[J]. *International Journal of Computer Vision*, 2021, 129(5): 1754-1767.
- [36] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 2242-2251.
- [37] ENGIN D, GENÇ A, EKENEL H K. Cycle-dehaze: Enhanced CycleGAN for single image dehazing[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2018: 938-946.
- [38] ANVARI Z, ATHITSOS V. Dehaze-GLCGAN: Unpaired single image de-hazing via adversarial training[EB/OL]. (2020-08-15) [2023-03-29]. <http://arxiv.org/abs/2008.06632v1>.
- [39] LI B Y, PENG X L, WANG Z Y, et al. AOD-net: All-in-one dehazing network[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 4780-4788.
- [40] QU Y Y, CHEN Y Z, HUANG J Y, et al. Enhanced Pix2pix dehazing network[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 8152-8160.
- [41] GUO H F, PIAO J C. MARG-UNet: A single image dehazing network based on multimodal attention residual group[C]//2022 IEEE 2nd International Conference on Information Communication and Software Engineering (ICICSE). Piscataway: IEEE, 2022: 105-109.
- [42] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018: 3-19.
- [43] WANG Q L, WU B G, ZHU P F, et al. ECA-net: Efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 11531-11539.
- [44] CAI Y T, WANG Y. MA-Unet: An improved version of Unet based on multi-scale and attention mechanism for medical image segmentation[C]//Third International Conference on Electronics and Communication; Network and Computer Technology (ECNCT 2021). SPIE, 2022: 1-13.
- [45] ATES G C, MOHAN P, ÇELİK E. Dual cross-attention for medical image segmentation[J]. *Engineering Applications of Artificial Intelligence*, 2023, 126: 107139.
- [46] SHYAM P, YOON K J, KIM K S. Towards domain invariant single image dehazing[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(11): 9657-9665.
- [47] LIN C Y, RONG X W, YU X Y. MSAFF-net: Multiscale attention feature fusion networks for single image dehazing and beyond[J]. *IEEE Transactions on Multimedia*, 2022, 25: 3089-3100.
- [48] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 105-114.
- [49] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10) [2023-03-29]. <http://arxiv.org/abs/1409.1556v6>.
- [50] CHEN D D, HE M M, FAN Q N, et al. Gated context aggregation network for image dehazing and deraining[C]//2019 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2019: 1375-1383.
- [51] YU Y K, LIU H, FU M H, et al. A two-branch neural network for non-homogeneous dehazing via ensemble learning[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2021: 193-202.
- [52] BAI H R, PAN J S, XIANG X G, et al. Self-guided image dehazing using progressive feature fusion[J]. *IEEE Transactions on Image Processing*, 2022, 31: 1217-1229.
- [53] WANG P Y, ZHU H Q, HUANG H, et al. TMS-GAN: A twofold multi-scale generative adversarial network for single image dehazing[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(5): 2760-2772.
- [54] LIU Y Y, LIU H, LI L Y, et al. A data-centric solution to

- NonHomogeneous dehazing via vision transformer[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2023: 1406-1415.
- [55] MEI Y Q, FAN Y C, ZHOU Y Q. Image super-resolution with non-local sparse attention[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 3516-3525.
- [56] AINSLIE J, ONTAÑÓN S, ALBERTI C, et al. ETC: Encoding long and structured inputs in transformers[EB/OL]. (2020-10-27)[2023-03-29]. <http://arxiv.org/abs/2004.08483v5>.

### 作者简介

**孙 航** 男,1986年4月出生于湖北省武汉市. 武汉大学博士,现为三峡大学副教授,硕士生导师. 主要研究方向为计算机视觉、图像去雾和 underwater 图像修复.

E-mail: sunhang0418@whu.edu.cn

**付秋月** 女,1998年8月出生于广东省佛山市. 现为三峡大学硕士研究生. 主要研究方向为计算机视觉、图像去雾.

E-mail: 15272874521@163.com

**李勃辉** 男,1998年10月出生于河南省信阳市. 现为三峡大学硕士研究生. 主要研究方向为计算机视觉、图像去雾.

E-mail: snowwhite@uestc.edu.cn

**但志平** 男,1976年3月出生于湖北省宜昌市. 现为三峡大学教授. 主要研究方向为模式识别、计算机视觉.

E-mail: zp\_dan@ctgu.edu.cn

**余 梅** 女,1980年1月,湖北省枣阳市. 武汉大学博士,现为三峡大学讲师. 主要研究方向为计算机视觉、新型电力系统.

E-mail: yumei@ctgu.edu.cn