

文章编号 1004-924X(2023)18-2687-13

全局和局部特征融合图像去雾

姜鑫*, 聂海涛, 朱明

(中国科学院长春光学精密机械与物理研究所, 吉林 长春 130033)

摘要: 具有参数共享特性的卷积操作主要关注于图像局部特征的提取, 而无法对超出感受野范围的特征进行建模, 同时整幅图像共享同一个卷积核参数也忽略了不同区域的特性不同。为了克服现有方法的表达不足, 提出了全局和局部特征融合去雾网络, 分别利用 Transformer 和卷积操作提取图像全局和局部特征信息, 并将两者融合后输出, 充分发挥了 Transformer 建模长距离依赖关系和卷积操作局部感知特性的优势, 实现了特征的高效表达。在最终输出复原图像前, 设计了包含多尺度图像块的增强模块, 利用 Transformer 进一步聚合全局特征信息, 丰富复原图像细节。同时, 提出了一个全局位置编码生成器, 可自适应地根据全局图像内容信息生成位置编码, 进而实现对像素点间依赖关系的二维空间位置建模。实验结果表明, 所提出的去雾网络在合成和真实图像数据集上均展现出了较好的去雾性能, 复原图像更加真实, 细节还原度高。

关键词: 图像去雾; 生成式对抗网络; Transformer; 位置编码生成器

中图分类号: TP394.41 **文献标识码:** A **doi:** 10.37188/OPE.20233118.2687

Global and local feature fusion image dehazing

JIANG Xin*, NIE Haitao, ZHU Ming

(Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences,
Changchun 130033, China)

* Corresponding author, E-mail: xinjiang@zju.edu.cn

Abstract: Convolution operations with parameter sharing features primarily focus on the extraction of local features of images but fail to model the features beyond the range of the receptive field. Moreover, when the parameters of an entire image share the same convolution kernel, the characteristics of different regions are ignored. To address this limitation in existing methods, a global and local feature fusion dehazing network is proposed. We utilize transformer and convolution operations to extract global and local feature information from images, respectively. Subsequently, we merge and output these features, effectively employing the advantages of transformers in modeling long-distance dependencies and the local perception of convolution operations, thus achieving efficient feature expression. Before the final output of restored images, we incorporate an enhancement module that includes multi-scale patches to further aggregate global feature information and enhance the details of the restored images using a transformer. Simultaneously, we introduce a global positional encoding generator, which can adaptively generate positional encodings based on the global content information of images, thereby enabling 2D spatial location modeling of the dependency relationship between pixels. Experimental results demonstrate the superior performance of the

收稿日期: 2022-02-13; 修订日期: 2022-03-11.

基金项目: 广东省重点领域研发计划资助项目 (No. 2019B010155003)

proposed dehazing network on both synthetic and real image datasets, producing more realistic restored images and significantly reducing detail loss.

Key words: image dehazing; generative adversarial network; transformer; positional encoding generator

1 引言

近些年,随着工业社会的快速发展,社会经济不断增长的同时对自然环境的破坏也日趋严重,导致雾霾天气出现的越来越频繁,严重地影响了人们日常的生产与生活^[1]。雾霾是大气中的粒子吸收散射光而产生的一种自然现象,受其影响,光学设备无法获取有效的场景信息,所采集的图像质量较差,易出现颜色失真、饱和度降低、图像模糊等退化现象。雾霾天气下图像质量的下降不仅降低了图像的视觉观赏性,而且严重地干扰了视频监控、自动驾驶、无人机跟踪等各类视觉处理任务的有效运行^[2]。针对上述现象,如何在不会丢失图像细节的前提下,在雾霾场景中有效地去除图像中的雾霾,最大程度上复原出图像的色彩信息,提高图像质量,具有重要的研究意义和应用价值。

图像去雾是一个极具挑战性的问题,近些年来引起了学术界和工业界的广泛关注。目前有关图像去雾的研究主要分为两大类:一类是基于先验信息的去雾方法研究,另一类是基于神经网络学习的去雾方法研究。基于先验信息的去雾方法主要利用大气散射模型和手工设计的先验知识对雾天图像进行去雾处理;基于神经网络学习的去雾方法主要利用神经网络的特征提取能力和包含大量雾天图像的数据集来实现高效去雾。

基于大气散射模型,Tan等^[3]通过提高图像局部对比度的方式实现了图像去雾,但该方法容易出现复原后图像色差较大的问题。He等^[4]提出了基于暗通道先验的去雾方法并取得了很好的去雾效果,但易受到天空等高亮区域的影响,并且存在时间和空间复杂度高的问题。Ancuti等^[5]基于原始图像及其半逆图像之间的色差,提出了一种可以快速检测并去除图像雾霾的去雾方法。Zhu等^[6]在对大量图像分析的基础上建立了景物深度和景物亮度与饱和度差的线性模型,通过景物透射图有效地去除了图像中的雾霾。

尽管上述算法在某些场景中实现了一定的去雾效果,但受限于理想条件下的大气散射模型和各种先验假设,其在复杂雾气图像中的应用仍存在着较大的局限性。

随着深度学习技术的快速发展,大量卷积神经网络应用在了图像去雾领域中。Cai等^[7]第一次将卷积神经网络引入到图像去雾任务中,提出了一种端到端的可训练去雾网络,利用多尺度卷积操作提取雾霾特征,大幅提升了图像去雾性能。Li等^[8]提出了一个轻量级去雾网络 AOD-Net,该网络将大气散射模型中的多个中间变量集成为一个可训练参数,有效地降低了复原公式的重建误差,提高了去雾图像的质量。Chen等^[9]提出了一种结合残差学习和导向滤波的去雾算法,采用多尺度卷积提取雾霾特征,利用导向滤波保持图像边缘特性,解决了复原图像对比度和清晰度下降的问题。Feng等^[10]提出了一个双视觉注意网络的联合图像去雾和透射率估计算法,充分利用了透射率估计和去雾过程中捕获信息的相关性,实现了图像中雾气浓度信息的预测。Yang等^[11]提出了一种分离特征和协同网络下的去雾模型,利用神经网络提取不同深度的空间信息及细节特征,使得复原图像颜色自然,细节保持良好。大部分基于神经网络学习的去雾方法均利用了卷积操作来提取图像特征,但是,具有参数共享特性的卷积操作在应用时有两点弊端:一是卷积操作更加关注于局部特征信息的提取,不能对超出感受野范围的特征进行建模,因此无法很好地感知图像全局特征信息;二是卷积核与图像之间的交互并不能根据图像内容而自适应地调整,使用相同的卷积核来复原不同区域的图像可能并不是最好的选择。

最近,随着 Transformer^[12]的出现,基于自注意力机制的网络框架在很多机器视觉处理任务上取得了与卷积神经网络相当甚至更好的性能和效果^[13-15]。Transformer最初是用来解决自然语言处理任务的,通过使用自注意力机制使得模型可以并行化训练。与卷积神经网络相比,

Transformer 最大的优势在于每个特征学习层中的单元都具有全局的感受野,其更擅长于建立远距离依赖关系的模型,可以实现全局特征信息的有效聚合。

基于此,为了克服现有方法表达的不足,提升网络的特征表达能力,提出了全局和局部特征融合去雾网络。网络分别利用 Transformer 和卷积操作提取图像全局和局部特征信息,并将两者融合后输出,充分发挥了 Transformer 建模长距离依赖关系和卷积操作局部感知特性的优势,实现了特征的高效表达。实验结果表明,所提出的去雾网络图像复原效果更加真实,细节还原度高,去雾性能优异。

2 全局和局部特征融合去雾网络

2.1 去雾网络整体架构

条件式生成对抗网络^[16]在图像复原领域展现出了良好的性能和复原效果,已广泛应用于图像去噪^[17-18]、图像去雨^[19-20]、图像去雾^[21-22]等各类图像复原任务中。鉴于此,本文基于此框架设计了全局和局部特征融合去雾网络。图 1 为条件式生成对抗去雾网络整体架构图^[23],它主要由一个生成器 G 和一个判别器 D 组成。其中,“ x ”表示输入的原始有雾图像,“ $G(x)$ ”表示经过生成器复原后的无雾图像,“ y ”表示与之对应的真实清晰无雾图像。

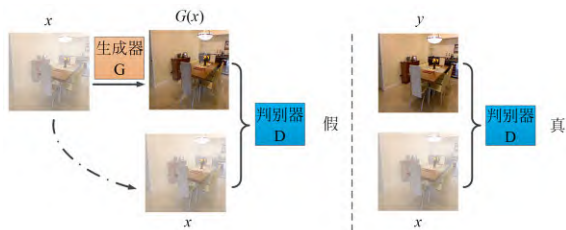


图 1 条件式生成对抗网络架构图

Fig. 1 Diagram of conditional generative adversarial network

生成器 G 的目标是根据输入的有雾图像复原出清晰无雾图像;判别器 D 的目标是准确分辨出生成器复原出的无雾图像和真实的无雾图像。生成器和判别器在样本训练的过程中进行博弈并不断提升各自性能,判别器 D 的目的是尽可能地将真实样本和生成样本区分开来,即判断生成

样本为假,真实样本为真;生成器 G 的目的是逐步学习真实清晰无雾图像的数据分布,并尽可能地生成与无雾图像类似的样本数据来欺骗判别器 D,使得判别器无法对其生成的样本做出准确判别。最终,在较为理想的情况下,生成器准确学习到了真实清晰图像的数据分布,能够生成“以假乱真”的无雾图像;同时,判别器无法准确分辨出生成样本和真实样本,从而达到纳什均衡的状态^[16]。在传统的生成式对抗网络框架中,生成器往往通过输入的随机噪声来复原出目标图像;而在条件式生成对抗网络框架中,生成器和判别器均可读取输入的原始有雾图像,有助于网络实现更好的复原效果^[23]。网络训练后,利用生成器 G,即可根据有雾图像生成与之对应的无雾图像,实现图像去雾效果,有关生成器 G 和判别器 D 的网络架构设计详见下文。

2.2 生成器

生成器采用 U-Net 型^[24]网络架构,如图 2 所示,其中“Conv”表示卷积操作,“BN”表示批标准化操作,“GELU”表示高斯误差线性单元激活函数,“MaxPool”表示最大池化操作,“GLFFM”表示全局和局部特征融合模块,“Enhancer”表示增强模块,“Tanh”表示双曲正切激活函数,“skip connection”表示跳跃连接操作。

生成器主要由四部分组成:编码器、全局和局部特征融合模块、解码器和增强模块。编码器主要由多个 Conv-BN-GELU 序列和最大池化操作组成,在逐步降低特征图分辨率的同时,不断地提取图像特征,并扩展特征图维度。在此基础上,全局和局部特征融合模块分别提取图像全局和局部特征信息,并将两者融合后输出,该模块的详细操作参见 2.3 节。接着,解码器采用多个 pixel shuffle 层^[25]对特征图进行上采样操作,在减少图像伪影的同时将特征图分辨率逐步向原始图像分辨率靠近。同时,为了实现高效的网络特征信息共享,编码器端每一级特征层都通过跳跃连接的方式与解码器端特征层进行直连,这有助于将编码器端浅层网络中提取到的边缘、纹理、轮廓等特征信息传递到解码器端的深层网络中,进而在解码的同时充分保留原始图像的细节、纹理等信息。之后,通过增强模块进一步融合图像全局特征信息,增强网络的特征表示能力,从而

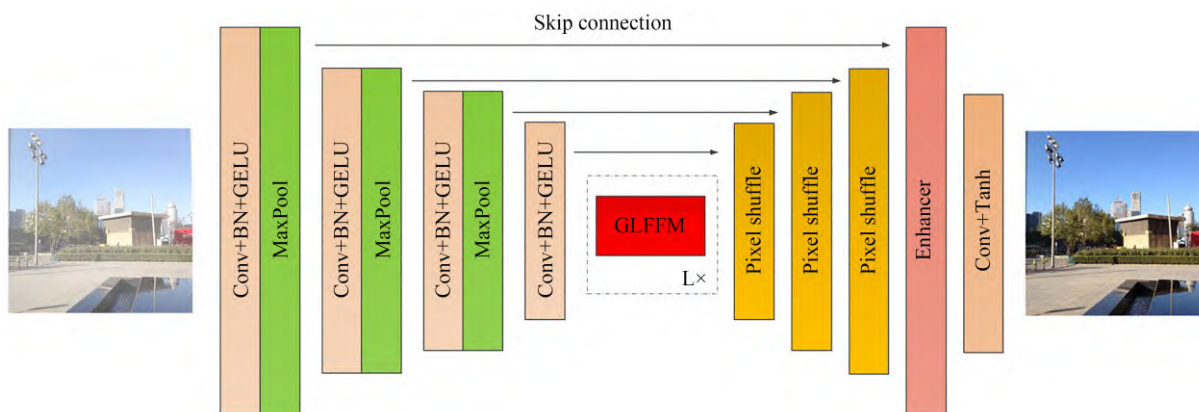


图 2 生成器示意图

Fig. 2 Framework of the generator

精细化复原结果,该模块的详细操作参见 2.4 节。最后,通过 Conv-Tanh 序列降低特征图维度并输出最终复原出的清晰无雾图像。生成器的详细参数信息如表 1 所示。

2.3 全局和局部特征融合模块

针对卷积操作特征表达的局限性,为了更好地提升去雾网络的整体特征表达能力,设计了一个全局和局部特征融合模块,分别利用 Transformer 和卷积操作提取图像全局和局部特征信息,发挥 Transformer 建模长距离依赖关系和卷积操作局部感知特性的优势,并将两者获取的特征信息进行融合后输出,进而实现特征的高效表达。全局和局部特征融合模块的架构示意图如图 3 所示,其中“PE”表示图像块压缩操作,“PEG”表示位置编码生成器^[26],”LN”表示层标准化操作,“MSA”表示多头自注意力机制,“MLP”表示多层感知机。

如图 3 所示,全局和局部特征融合模块主要包含局部特征提取分支、全局特征提取分支、特征融合单元和长距离跳跃连接。局部特征提取分支由 Conv-BN-GELU-Conv-BN 序列组成,通过卷积操作提取图像局部特征信息,此模块中所有的卷积核尺寸均为 3×3。全局特征提取分支由图像块压缩单元、位置编码生成器和多个 Transformer 编码器组成。由于 Transformer 需要的是类似于单词序列的一维输入信号,而图像本身为二维信号,因此需要先将整幅图像拆分为多个图像块,并将图像块展平压缩后拉伸成一维序列,再输入到 Transformer 编码器中。但是,这种

表 1 生成器的参数信息

Tab. 1 Parameter information of the generator

网络层	输出分辨率	输出维度	操作
输入图像	256×256	3	—
Conv+BN+GELU	256×256	64	3×3 conv
MaxPool	128×128	64	2×2 maxpool
Conv+BN+GELU	128×128	128	3×3 conv
MaxPool	64×64	128	2×2 maxpool
Conv+BN+GELU	64×64	256	3×3 conv
MaxPool	32×32	256	2×2 maxpool
Conv+BN+GELU	32×32	512	3×3 conv
GLFFM	32×32	512	详见 2.3 节
Pixel Shuffle	64×64	256	2 倍上采样
Pixel Shuffle	128×128	128	2 倍上采样
Pixel Shuffle	256×256	64	2 倍上采样
Enhancer	256×256	128	详见 2.4 节
Conv+Tanh	256×256	3	3×3 conv

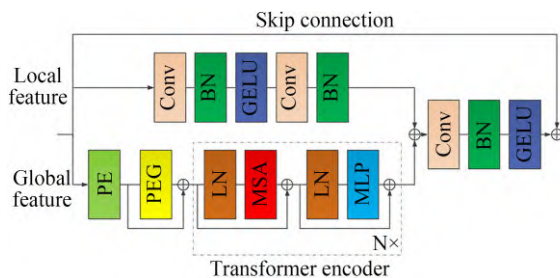


图 3 全局和局部特征融合模块示意图

Fig. 3 Diagram of global and local feature fusion module

方法并不能完美地建模图像,因为其缺少了二维图像中所包含的重要信息,即像素点间的位置信息。考虑到图像复原任务对位置信息是高度敏感的,因此在图像块序列输入到 Transformer 编码器前,设计了位置编码生成器,以实现针对不同位置图像块的有效编码,进而保留像素点间的二维空间位置关系。位置编码生成器的有关设计详见 2.5 小节。在去雾网络整体架构中,生成器中的编码器提取出了分辨率大小为 32×32 的特征图,输入到全局和局部特征融合模块。在全局特征提取分支中,Transformer 将此特征图拆分为 1 024 个图像块,每个图像块的分辨率大小为 1×1 。Transformer 编码器的结构与原始 Transformer 编码器^[12]保持一致,经其特征提取后,将输出的一维序列信号恢复成二维图像。之后,将局部特征信息和全局特征信息通过由 Conv-BN-GELU 序列构建的特征融合单元进行融合,在保证整个模块输出特征图通道数目与输入特征图通道数目一致的同时,进一步增强网络的特征表达能力。最后,通过长距离跳跃连接将输入特征图与融合特征图叠加后输出。长距离跳跃连接一方面可以避免反向传播过程中的梯度消失问题,加速训练过程中网络模型的收敛;另一方面可以使主干网络更加专注于有价值特征信息的提取和学习,而将价值量较低或不太重要的特征信息通过旁路向后传递。值得指出的是,全局和局部特征融合模块不改变输入特征图的宽度、高度和通道数目,因此可以在不改变其他网络主体结构的前提下,灵活嵌入移植到其他网络模型中,进而提高网络的特征提取和表达能力。

2.4 增强模块

为了进一步聚合图像的全局特征信息,在最终输出复原图像前,设计了包含多尺度图像块的增强模块,利用 Transformer 丰富复原图像的细节信息。增强模块的整体架构如下图 4 所示,其中“concat”表示特征图连接操作。

如图 4 所示,首先将输入的特征图依次拆分成分辨率大小为 4×4 , 8×8 和 16×16 的图像块,由于特征图的分辨率大小为 256×256 ,因此拆分后的图像块数目分别为 4 096, 1 024 和 256。多尺度图像块包含了多尺度图像特征信息,有助于网络在不同尺度上恢复图像的细节、纹理等信息。接着,利用卷积操作对图像块进行压缩,依次将图像块通道数目压缩成 16, 64 和 256,并将图像块展平成一维序列后,分别输入到 Transformer 中进行全局特征提取,此处的 Transformer 与上一小节的全局特征提取分支一致。特征提取后,将一维序列信号重新恢复成二维特征图,特征图的通道数目分别为 16, 64 和 256,分辨率大小分别为 64×64 , 32×32 和 16×16 。然后,利用 pixel shuffle 层^[25]分别对特征图进行 4 倍、8 倍和 16 倍上采样操作,依次得到三个通道数目为 1, 分辨率为 256×256 的特征图。雾气在图像中往往是非均匀分布的,此特征图可以有效地实现图像的空间注意力机制^[27],即允许网络选择性地学习价值量较大的像素区域,如浓雾区域、边缘区域等,而减少对价值较低像素区域的学习,如无雾区域、平坦区域等,进而高效地利用网络资源,提高特征学习和表达能力。之后,将此特征图自适应学习到的权重值与输入特征图进行逐像素点乘,并进行特征层连接。最后,通过 3×3 卷积

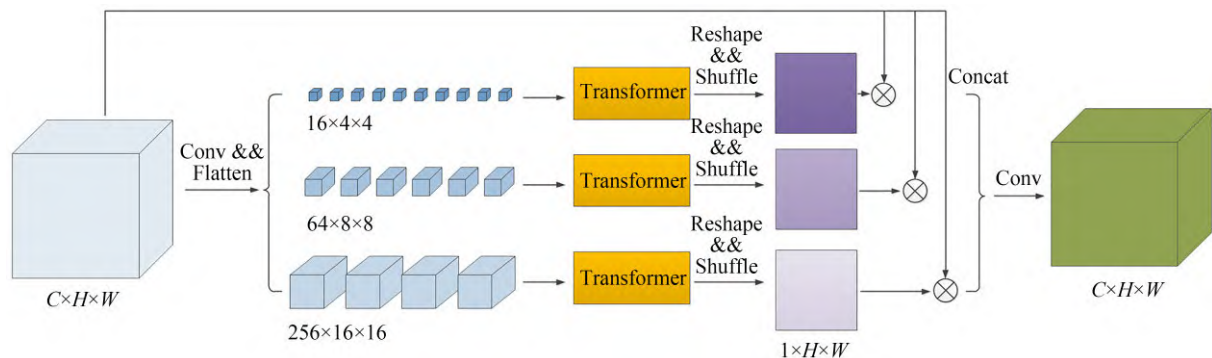


图 4 增强模块示意图

Fig. 4 Framework of enhancer

操作进行特征融合和通道数目降维,以保持输出特征图通道数目与输入特征图通道数目一致。

2.5 位置编码生成器

Transformer的自注意力机制有一个天然的缺陷,即它无法捕获输入图像块序列的位置信息。但对于图像复原任务来说,保留图像的二维空间位置信息对复原结果至关重要。因此,在图像块序列输入到Transformer编码器前,需要将每个图像块的位置信息进行编码,并与其对应的特征向量叠加后一起输入到编码器中。

传统的位置编码方式,主要包括固定位置编码^[12]、相对位置编码^[15]、可学习位置编码^[14]等。这些编码方法往往都是事先定义好的,且编码方式基本与图像内容无关。Chu等^[26]提出了一种

基于图像内容的动态位置编码生成器,可根据图像局部区域内的内容信息自适应地生成位置编码,如图5(a)所示。具体实现过程如下:首先将一维序列恢复成二维特征图,再利用 3×3 卷积操作生成局部位置编码特征图,接着将此特征图拆分成一维序列与原始输入序列进行叠加后输入到Transformer编码器中,其中卷积操作时采用边缘零填充的方式,一方面保证了输出特征图与输入特征图分辨率大小一致,另一方面提供了特征图内每个像素点的绝对位置信息。该编码方式同时提供了像素间的相对位置信息和绝对位置信息,并有效地结合图像局部区域内的内容信息生成位置编码,显著地提升了Transformer在各领域中不同任务的性能^[26]。

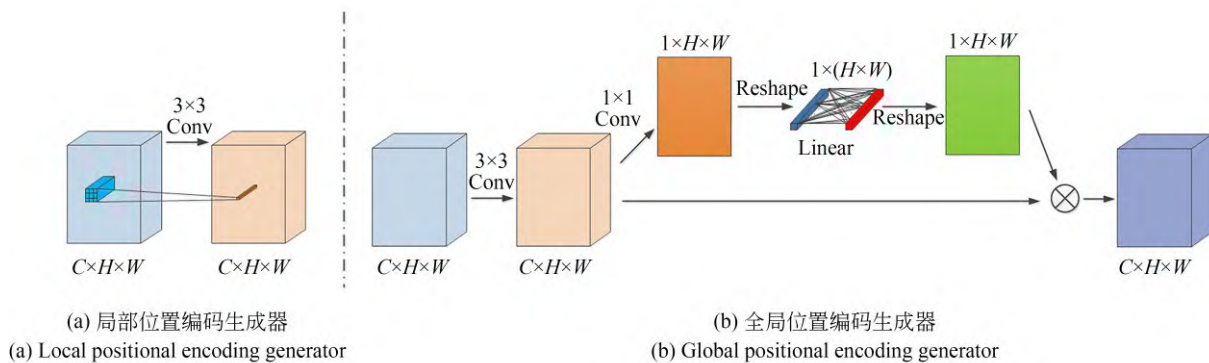


图5 位置编码生成器示意图

Fig. 5 Framework of the positional encoding generator

上述位置编码生成器采用 3×3 卷积操作生成位置编码特征图,特征图中每个像素点的所对应的感受野较小,且编码方式无法对长距离像素点间的依赖关系进行建模。在实际应用场景中,雾气在图像中往往是成片出现的,因此更大的感受野对图像复原过程可以提供更多的参考和指导。基于此,提出了全局位置编码生成器,如图5(b)所示,其中“Linear”表示线性操作。针对局部编码生成器生成的特征图,先通过 1×1 卷积操作对特征图进行降维,生成通道数为1的特征图。接着,将特征图展平成一维序列,并对其进行全连接操作,以使得特征图中所有的像素点均具有全局的感受野。然后,将一维序列恢复成二维特征图,并将该特征图与局部编码生成器生成的特征图进行逐像素点乘后输出。该编码方式提供

了特征图中任意两像素点间的相对位置关系,可自适应地根据全局图像内容信息生成位置编码,相较而言提供了更大的感受野,有助于提升图像整体复原效果。

2.6 判别器

判别器的目标是准确分辨出生成器复原出的样本图像和真实样本图像。本文采用了全卷积神经网络PatchGAN^[23],网络输出一个判别矩阵。矩阵中每个像素点代表了原始图像中某个图像块是复原样本或真实样本的概率,因此其可以充分考虑图像中不同区域对判别结果的影响,使得生成器在训练的过程更加专注于复原图像的细节、纹理等信息,并且有助于减少复原图像伪影。最终,计算矩阵中所有元素的均值进行输出。

2.7 损失函数

条件式生成对抗网络的损失函数在数学上可以表示为^[16]:

$$L_{cGAN}(G, D) = E_{x,y}[\log D(x, y)] + E_{x,z}[\log(1 - D(x, G(x, z)))] \quad (1)$$

其中: x 表示输入有雾图像, y 表示与之对应的清晰无雾图像, z 表示噪声信号。生成器 G 的目标是最小化该损失函数,而判别器 D 的目标是最大化该损失函数。在网络训练和推理时,都以 dropout 的形式替代噪声信号^[23]。

相关研究表明将上述对抗损失函数与 L1 损失函数混合使用有助于减少复原图像伪影及颜色失真现象^[23]。L1 损失函数可以从像素级层面保证复原图像与真实图像的一致性。L1 损失函数在数学上可以表示为:

$$L_{L1}(G) = E_{x,y,z}[\|y - G(x, z)\|_1] \quad (2)$$

同时,将感知损失考虑在内,在特征空间中对复原图像和真实图像进行比较,与像素空间形成互补。为了度量特征空间中的感知相似度,提取出了预训练 VGG16 模型^[28]中的第 2 特征层和第 5 特征层中的特征信息,并将感知损失表示为:

$$L_P(G) = \|\xi(y) - \zeta(G(x, z))\|_2^2 \quad (3)$$

其中, ξ 表示预训练 VGG16 模型的特征提取器。

将三者进行结合,总的损失函数可以表示为:

$$L(G, D) = L_{cGAN}(G, D) + \lambda_1 L_{L1}(G) + \lambda_2 L_P(G) \quad (4)$$

其中, λ_1 和 λ_2 均为权衡参数。

3 仿真实验

本节中,分别在公开的合成图像数据集和真实图像数据集上验证所提出全局和局部特征融合去雾网络的去雾效果。将所提去雾网络与下述经典的去雾算法进行对比,包括:DCP(Dark Channel Prior)^[4],CAP(Color Attenuation Prior)^[6],AODNet(All-in-One Dehazing Network)^[8],EPDN(Enhanced Pix2pix Dehazing Network)^[29],pix2pix^[23],FFA-Net(Feature Fusion Attention Network)^[30]和LD-Net(Light-DehazeNet)^[31]。同时,设计了消融实验来验证所提出增强模块和位置编码生成器的有效性。

3.1 数据集

合成图像数据集选择 RESIDE(Realistic Single Image Dehazing)数据集^[32],该数据集主要基于大气散射模型,随机设定大气光参数和环境散射系数,针对每张清晰无雾图像合成多张有雾图像。模型参数训练时,从OTS(Outdoor Training Set)集合中随机选取 4 000 张成对的有雾与无雾图像用于训练。测试时,从SOTS(Synthetic Objective Testing Set)集合中随机选取 600 张成对的有雾与无雾图像用于分析和比较。

真实图像数据集选择近些年 CVPR NTIRE 去雾挑战赛^[33-35]中公开的数据集 O-HAZE^[36],DENSE-HAZE^[37]和 NH-HAZE^[38]。三个数据集分别包含了 45 对户外有雾场景图像、55 对浓雾场景图像和 55 对非均匀分布有雾场景图像。这些图像都是通过专业的烟雾生成器,在户外模拟不同程度的真实有雾场景抓拍生成的。在这 155 对图像中,随机选择 140 对图像用于训练,剩下 15 对用于测试。

3.2 实施细节

基于 PyTorch 框架,去雾网络的训练和测试阶段均在 NVIDIA GEFORCE RTX 3090 TI GPU 上运行。所有图像在输入进网络前均被调整为 256×256 大小的分辨率,同时训练过程中使用了 ADAM 优化器。网络的整体训练过程共计 400 次迭代,其中前 200 次迭代时学习率固定为 0.000 1,后 200 次迭代学习率逐渐从 0.000 1 线性衰减为 0。参照 pix2pix 模型^[23],权衡参数 λ_1 和 λ_2 分别取 100 和 50。每张图像在 RTX 3090 TI GPU 上的平均运行时间为 0.096 s。

3.3 合成图像数据集上的效果

客观评价指标采用峰值信噪比(Peak Signal to Noise Ratio, PSNR)和结构相似度(Structural Similarity Index Measurement, SSIM)。不同去雾算法的客观评价结果如表 2 所示,视觉对比效果如图 6 所示。从表 2 中可以看出所提出的全局和局部特征融合去雾网络实现了最优的 PSNR 和 SSIM 指标,相较于 FFA-Net 分别提升了 2.063 7 dB 的 PSNR 和 0.018 7 的 SSIM。从图 6 中可以发现,DCP 和 CAP 算法在天空区域处理效果不佳,易出现颜色畸变;AODNet 网络去雾效果不明显,复原图像中仍存在着较多雾气区

表 2 合成图像数据集的客观评价结果

Tab. 2 Objective evaluation results on synthetic image datasets

去雾算法	PSNR	SSIM
DCP	17.692 1	0.877 8
CAP	18.475 8	0.820 4
AODNet	19.784 2	0.869 8
EPDN	21.503 6	0.879 2
Pix2pix	27.153 4	0.929 3
FFA-Net	31.126 5	0.958 3
LD-Net	24.765 3	0.917 2
Ours	33.190 2	0.977 0

域;EPDN网络生成的图像有时与真实清晰图像存在一定的色差;pix2pix算法复原的图像有时存在一定的伪影现象;FFA-Net重建的图像有时存在部分细节不清晰;LD-Net复原出的图像有时会出现颜色失真问题。与上述算法相比,所提去雾网络复原出了视觉效果更好的清晰无雾图像,

并且有效地保留了原始图像中的色彩和纹理信息。

3.4 真实图像数据集上的效果

不同去雾算法在真实图像数据集上的客观评价结果如表 3 所示,视觉对比效果如图 7 所示。从中可以看出 CAP 和 AODNet 算法的去雾效果并不明显,复原图像中仍存在着大量雾气;DCP

表 3 真实图像数据集的客观评价结果

Tab. 3 Objective evaluation results on real image datasets

去雾算法	PSNR	SSIM
DCP	17.982 3	0.698 2
CAP	17.104 6	0.656 4
AODNet	17.119 3	0.613 2
EPDN	17.216 8	0.708 1
Pix2pix	18.109 2	0.710 2
FFA-Net	17.658 2	0.691 0
LD-Net	17.132 8	0.632 5
Ours	19.311 5	0.747 8

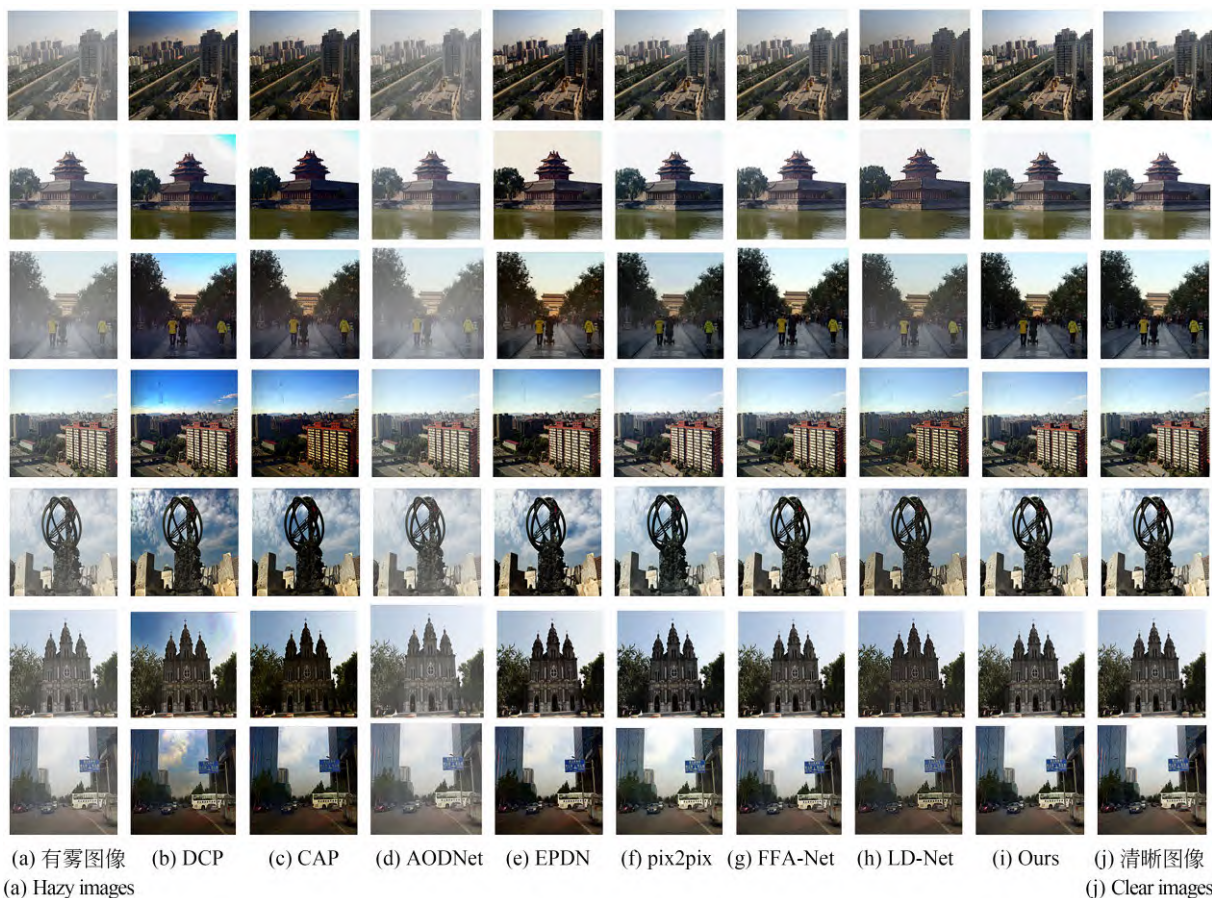


图 6 合成图像数据集的视觉对比效果

Fig. 6 Visual contrast effect on synthetic image datasets



图 7 真实图像数据集的视觉对比效果

Fig. 7 Visual contrast effect on real image datasets

取得了一定程度上的去雾效果,但复原图像有时颜色相对较暗;EPDN有效地去除了图像中的雾霾,但也造成了颜色失真;pix2pix和FFA-Net保留了图像的细节和纹理等信息,但去雾效果不够彻底;LD-Net在此类图像上的去雾效果表现不佳。与上述算法相比,所提出的全局和局部特征融合去雾网络展示出了较好的去雾效果,图像色彩还原度高,纹理也更加清晰,同时也取得了最优的PSNR和SSIM指标。虽然部分图像的还原细节有些模糊,但网络也展现出了较优的性能和较大的潜力。

为了进一步比较不同去雾算法的视觉效果,本文将去雾网络在真实户外有雾图像数据中进行了测试,视觉对比效果如图8所示。从图中可以看出,DCP和CAP在去雾的同时,会给复原图像带来一定的颜色失真问题。AODNet的去雾效果并不明显,复原图像中仍存在着较大雾气。EPDN恢复的图像在某些区域中较为模糊。pix2pix实现了一定的去雾效果,但去雾后图像的细节和纹理不够清晰。FFA-Net的去雾效果仍然不够理想,LD-Net有时也无法完全去除图像

中的雾霾。相较而言,所提去雾网络复原的图像更加真实和自然,色彩失真较少,图像细节和纹理也更加清晰。

3.5 消融实验

为了验证所提出位置编码生成器和增强模块的有效性,设计了两类消融实验进行测试。针对位置编码方法,对比了6种不同的编码方式,分别为:(1)全局位置编码生成器;(2)局部位置编码生成器^[26];(3)固定位置编码^[12];(4)相对位置编码^[15];(5)可学习位置编码^[14];(6)无位置编码。6种编码方式在合成图像数据集和真实图像数据集的客观评价结果如表4所示。从中可以看出,通过位置编码保留图像的二维空间位置信息,对图像复原结果至关重要。同时,仅提供相对位置编码信息是不够的,只有包含了绝对位置编码信息,才能实现较好的复原效果。所提出的位置编码生成器同时提供了像素点间的相对位置信息和绝对位置信息,并有效地结合了图像全局内容信息生成位置编码,实现了最优的PSNR和SSIM指标。

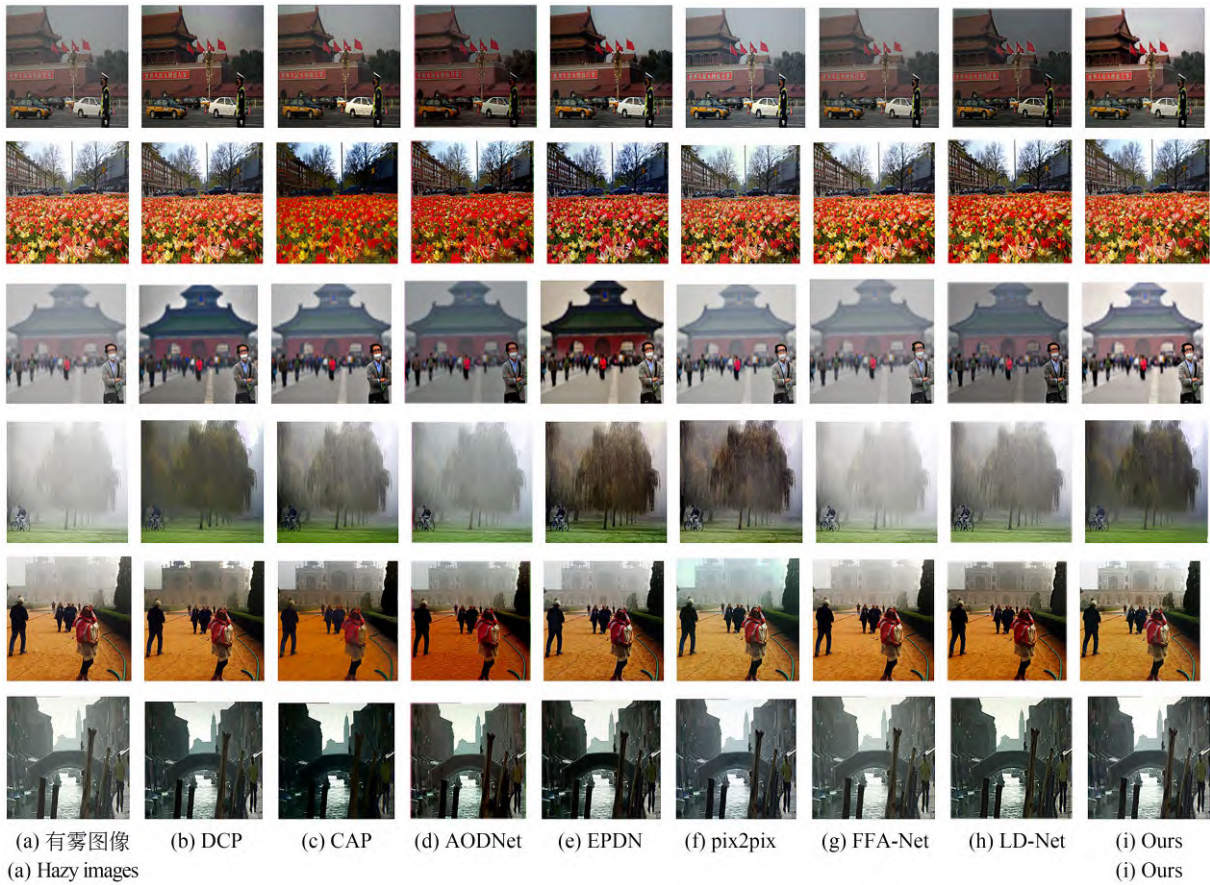


图 8 真实户外有雾图像的视觉对比效果
Fig. 8 Visual contrast effect on real outdoor hazy images

表 4 不同位置编码方式的客观评价结果

Tab. 4 Objective evaluation results of different positional encoding methods

编码方式	合成图像数据集		真实图像数据集	
	PSNR	SSIM	PSNR	SSIM
全局位置编码生成器	33.190 2	0.977 0	19.315 5	0.747 8
局部位置编码生成器	32.677 6	0.971 5	19.073 9	0.727 9
固定位置编码	32.698 2	0.972 5	18.977 3	0.724 7
相对位置编码	31.814 8	0.966 8	18.705 8	0.711 7
可学习位置编码	32.679 2	0.970 7	18.714 6	0.720 3
无位置编码	31.498 5	0.958 6	18.668 5	0.708 5

为了验证增强模块的有效性,分别对包含增强模块和不包含增强模块的情况进行了测试,对比结果如表 5 所示。从中可以看出,增强模块利用包含多尺度图像块的 Transformer 进一步融合了图像全局特征信息,通过较大的感受野,有效地提升了复原图像的质量。

表 5 增强模块的客观评价结果

Tab. 5 Objective evaluation results of the enhancer

是否包含增强模块	合成图像数据集		真实图像数据集	
	PSNR	SSIM	PSNR	SSIM
是	33.190 2	0.977 0	19.311 5	0.747 8
否	32.486 9	0.970 1	17.534 2	0.662 7

4 结 论

本文克服了现有卷积操作表达方法的不足,提出了全局和局部特征融合去雾网络。分别利用 Transformer 和卷积操作提取图像全局和局部特征信息,发挥各自建模长距离依赖关系和局部感知特性的优势,实现了特征的高效表达。同时,设计了包含多尺度图像块的增强模块,利用 Transformer 进一步聚合全局特征信息,丰富复

原图像细节。最后,提出了一个全局位置编码生成器,自适应地根据全局图像内容信息生成位置编码。实验结果表明:所提去雾网络展现出了较好的去雾性能,在合成图像数据集上可达到 33.190 2 dB 的 PSNR 和 0.977 0 的 SSIM 指标,在真实图像数据集上可达到 19.315 5 dB 的 PSNR 和 0.747 8 的 SSIM 指标,复原图像更加真实,细节还原度高。

参考文献:

- [1] 王卫星,赵恒. 结合改进 Retinex 及自适应分数阶微分的雾霾公路交通图像增强[J]. 光学精密工程, 2020, 28(8): 1820-1834.
WANG W X, ZHAO H. Haze traffic image enhancement based on improved retinex and adaptive fractional differential [J]. *Opt. Precision Eng.*, 2020, 28(8):1820-1834. (in Chinese)
- [2] JIANG X, ZHAO C L, ZHU M, *et al.* Residual spatial and channel attention networks for single image dehazing [J]. *Sensors (Basel, Switzerland)*, 2021, 21(23): 7922.
- [3] TAN R T. Visibility in Bad Weather from a Single Image[C]. 2008 *IEEE Conference on Computer Vision and Pattern Recognition*. 23-28, 2008, *Anchorage, AK, USA*. IEEE, 2008: 1-8.
- [4] HE K M, SUN J, TANG X O. Single image haze removal using dark channel prior[C]. 2009 *IEEE Conference on Computer Vision and Pattern Recognition*. 20-25, 2009, *Miami, FL*. IEEE, 2009: 1956-1963.
- [5] ANCUTI C O, ANCUTI C, HERMANS C, *et al.* A Fast Semi-Inverse Approach to Detect and Remove the Haze from a Single Image[M]. *Computer Vision-ACCV 2010*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011: 501-514.
- [6] 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: a Publication of the IEEE Signal Processing Society*, 2015, 24(11): 3522-3533.
- [7] 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: a Publication of the IEEE Signal Processing Society*, 2016, 25(11): 5187-5198.
- [8] 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)*. 22-29, 2017, *Venice, Italy*. IEEE, 2017: 4780-4788.
- [9] 陈清江,张雪. 混合残差学习与导向滤波算法在图像去雾中的应用[J]. 光学精密工程, 2019, 27(12):2702-2712.
CHEN Q J, ZHANG X. Application of hybrid residual learning and guided filtering algorithm in image defogging [J]. *Opt. Precision Eng.*, 2019, 27(12):2702-2712. (in Chinese)
- [10] 冯燕茹. 双视觉注意网络的联合图像去雾和透射率估计[J]. 光学精密工程, 2021, 29(4): 854-863.
FENG Y R. Joint transmission map estimation and image dehazing using dual vision attention network [J]. *Opt. Precision Eng.*, 2021, 29(4): 854-863. (in Chinese)
- [11] 杨燕,梁小珍,张金龙. 分离特征和协同网络下的端到端图像去雾[J]. 光学精密工程, 2021, 29(8):1931-1941.
YANG Y, LIANG X Z, ZHANG J L. End-to-end image dehazing under separated features and collaborative network [J]. *Opt. Precision Eng.*, 2021, 29(8):1931-1941. (in Chinese)
- [12] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need [C]. *Proceedings of the 31st International Conference on Neural Information Processing Systems. December 4 - 9, 2017, Long Beach, California, USA*. New York: ACM, 2017: 6000-6010.
- [13] LIANG J Y, CAO J Z, SUN G L, *et al.* Swin-IR: image restoration using swin transformer [C].

- 2021 *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. 11-17, 2021, Montreal, BC, Canada. IEEE, 2021: 1833-1844.
- [14] CHEN H T, WANG Y H, GUO T Y, *et al.* Pre-trained image processing transformer [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 12294-12305.
- [15] QIN Q, YAN J K, WANG Q, *et al.* ETDNet: an efficient transformer deraining model[J]. *IEEE Access*, 2021, 9: 119881-119893.
- [16] MIRZA M, OSINDERO S. Conditional Generative Adversarial Nets [EB/OL]. 2014: *arXiv*: 1411.1784. <https://arxiv.org/abs/1411.1784>
- [17] ZHANG H B, SAKURAI K. Conditional generative adversarial network-based image denoising for defending against adversarial attack[J]. *IEEE Access*, 2021, 9: 169031-169043.
- [18] WANG P. Image denoising using deep cgan with bi-skip connections[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 16-17, 2019, Long Beach, CA, USA. IEEE, 2020: 724-729.
- [19] ZHANG H, SINDAGI V, PATEL V M. Image de-raining using a conditional generative adversarial network [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30 (11) : 3943-3956.
- [20] LI R T, CHEONG L F, TAN R T. Heavy rain image restoration: integrating physics model and conditional adversarial learning [C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 15-20, 2019, Long Beach, CA, USA. IEEE, 2020: 1633-1642.
- [21] PARK J, HAN D K, KO H. Fusion of heterogeneous adversarial networks for single image dehazing [J]. *IEEE Transactions on Image Processing*, 2020, 29: 4721-4732.
- [22] LI R D, PAN J S, LI Z C, *et al.* Single image dehazing via conditional generative adversarial network [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 8202-8211.
- [23] ISOLA P, ZHU J Y, ZHOU T H, *et al.* Image-to-image translation with conditional adversarial networks [C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 5967-5976.
- [24] RONNEBERGER O, FISCHER P, BROX T. *U-Net: Convolutional Networks for Biomedical Image Segmentation* [M]. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234-241.
- [25] SHI W Z, CABALLERO J, HUSZÁR F, *et al.* Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 1874-1883.
- [26] CHU X, TIAN Z, ZHANG B, *et al.* Conditional Positional Encodings for Vision Transformers [EB/OL]. 2021: *arXiv*: 2102.10882. <https://arxiv.org/abs/2102.10882.pdf>
- [27] JIANG X, LU L, ZHU M, *et al.* Haze relevant feature attention network for single image dehazing [J]. *IEEE Access*, 2021, 9: 106476-106488.
- [28] 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)*. 18-22, 2018, Salt Lake City, UT, USA. IEEE, 2018: 938-9388.
- [29] 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)*. 15-20, 2019, Long Beach, CA, USA. IEEE, 2020: 8152-8160.
- [30] 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.
- [31] ULLAH H, MUHAMMAD K, IRFAN M, *et al.* Light-DehazeNet: a novel lightweight CNN architecture for single image dehazing [J]. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, 2021, 30: 8968-8982.
- [32] LI B Y, REN W Q, FU D P, *et al.* Benchmarking single-image dehazing and beyond [J]. *IEEE*

- Transactions on Image Processing*, 2019, 28(1): 492-505.
- [33] ANCUTI C, ANCUTI C O, TIMOFTE R, *et al.* NTIRE 2018 challenge on image dehazing: methods and results[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 18-22, 2018, Salt Lake City, UT, USA. IEEE, 2018: 1004-100410.
- [34] ANCUTI C, ANCUTI C, TIMOFTE R, *et al.* NTIRE 2019 image dehazing challenge report[C]. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA: CVPRW*, 2019: 2241-2253.
- [35] ANCUTI C O, ANCUTI C, VASLUIANU F A, *et al.* NTIRE 2020 challenge on nonhomogeneous dehazing[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 14-19, 2020, Seattle, WA, USA. IEEE, 2020: 2029-2044.
- [36] ANCUTI C O, ANCUTI C, TIMOFTE R, *et al.* O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 18-22, 2018, Salt Lake City, UT, USA. IEEE, 2018: 867-8678.
- [37] 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)*. 22-25, 2019, Taipei, China. IEEE, 2019: 1014-1018.
- [38] 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)*. 14-19, 2020, Seattle, WA, USA. IEEE, 2020: 1798-1805.

作者简介:



姜 鑫(1992—),男,吉林长春人,硕士,研究实习员,2015年和2018年分别于浙江大学获得学士学位和硕士学位,主要从事图像处理及人工智能方面的研究。E-mail: xinjiang@zju.edu.cn



朱 明(1964—),男,江西南昌人,研究员,博士生导师,1985年于南京航空航天大学获得学士学位,1991年于中科院长春光机所获得硕士学位,现为中国图象图形学会常务理事,吉林省图象图形学会会长,主要从事图像处理与人工智能理论及应用方向的研究。E-mail: zhu_mingca@163.com