

## 基于多补丁和多尺度层级聚合网络的快速非均匀图像去雾



杨坤 张娟 方志军

上海工程技术大学电子电气工程学院 上海 201620

(1249959224@qq.com)

**摘要** 尽管基于卷积神经网络的去雾算法在合成的均匀雾气数据集上已经取得了巨大进展,但在真实的非均匀有雾图像上仍然表现不佳。为了快速有效地去除图像中的非均匀雾气,文中首先提出了一种多补丁和多尺度层级聚合网络结构(Multi-patch and Multi-scale Hierarchical Aggregation Network, MPSHAN),融合了多补丁局部化信息和多尺度全局化信息。其次,提出了层级融合模块(Hierarchical Fusion Module, HFM),既解耦了残差融合以实现更丰富的非线性特征表达,又通过通道注意力机制提升了关键位置的特征融合质量。同时,对层级结构使用扩张卷积获得多尺度信息,增强特征图以优化融合效果。此外,在损失函数中加入频域损失以恢复更好的边缘质量。实验结果表明,所提算法在非均匀雾气图像上具有很好的鲁棒性,1200×1600 高分辨率图像的平均处理时间仅有0.044 s,相比其他去雾算法,其在图像去雾效果和运行时间之间实现了更好的平衡。

**关键词:** 多补丁;多尺度;层级融合模块;注意力机制;扩张卷积;图像去雾

**中图分类号** TP391

## Multi-patch and Multi-scale Hierarchical Aggregation Network for Fast Nonhomogeneous Image Dehazing

YANG Kun, ZHANG Juan and FANG Zhi-jun

School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

**Abstract** Despite dehazing algorithms based on convolutional neural networks have made tremendous progress in synthetic uniform hazy datasets, they still perform poorly on real nonhomogeneous hazy images. In order to achieve fast and effective nonhomogeneous image dehazing, we propose a multi-patch and multi-scale hierarchical aggregation network (MPSHAN), which fuses multi-patch local information and multi-scale global information. Secondly, we propose a hierarchical fusion module (HFM), which not only decouples residual fusion to achieve richer non-linear feature expression, but also improves the feature fusion quality at key locations through the channel attention mechanism. At the same time, dilated convolution is used on hierarchies to obtain multi-scale information, which enhances feature maps to optimize the fusion effect. In addition, in the loss function, we add frequency domain loss to restore better edge quality. The experimental results show that the proposed algorithm has good robustness on nonhomogeneous hazy images, and the average processing time of 1200×1600 high-resolution images is only 0.044 s. Compared with other dehazing algorithms, it achieves a better balance between image dehazing effect and running time.

**Keywords** Multi-patch, Multi-scale, Hierarchical fusion module, Attention mechanism, Dilated convolution, Image dehazing

### 1 引言

在恶劣的雾霾天气中,所拍摄图像的颜色、纹理、对比度的衰减直接导致了图像场景的可见性下降。这样的模糊图像会直接加大许多高级视觉任务的挑战难度,如目标检测<sup>[1]</sup>、语义分割<sup>[2]</sup>、姿态估计<sup>[3]</sup>等。因此,从有雾图像中恢复清晰图像成为了重要的视觉任务之一。

尽管基于深度学习的去雾算法<sup>[4-10]</sup>取得了很好的效果,

但是驱动网络的数据大多是合成的,往往仅在合成的有雾图像上取得很好的性能。并且,由于实际生活中雾气通常并不均匀,因此在真实的有雾图像上去雾效果并不理想。在2020年CVPR的工作室NTIRE(New Trends in Image Restoration and Enhancement)比赛中,主办单位首次提供了真实的非均匀雾气数据集NH-HAZE<sup>[11]</sup>,本文主要基于该数据集展开研究。

在NTIRE2020的非均匀去雾比赛中,Das等<sup>[12]</sup>将深度

收稿日期:2020-09-07 返修日期:2021-05-26 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61772328)

This work was supported by the National Natural Science Foundation of China(61772328).

通信作者:张娟(zhang-j@foxmail.com)

多补丁层级网络(Deep Multi-patch Hierarchical Network, DMPHN)用于去雾,以最快的运行速度取得了优秀的去雾效果。同时,他们也探索了深度多尺度层级网络(Deep Multi-scale Hierarchical Network, DMSHN)在去雾中的应用。实验结果证明,深度多补丁层级网络(DMPHN)比深度多尺度层级网络(DMSHN)恢复的无雾图像效果更好。

受他们工作的启发,我们产生了一种创新的想法:聚合 DMPHN 和 DMSHN,综合多补丁网络获取的局部化信息和多尺度网络获取的全局化信息,在保持快速运行的情况下,提升图像的去雾效果。为了进一步提升层级特征融合质量,本文提出了层级融合模块(HFM)。一方面,原网络中残差式的信息融合限制了网络的表达空间,HFM 模块使用拼接式的融合方法以解耦特征融合过程中潜在的组合方式。另一方面,在模块中,本文使用通道注意力机制优化关键融合处的特征图质量。对于多尺度网络部分,本文对层级结构使用扩张卷积以获取更密集的特征图,为层级融合部分提供优质的初始融合特征。此外,在损失函数部分,本文基于 Laplace 离散值核函数,加入了二阶微分频域信息,以强化边缘信息的恢复。在最终的结果中可以看到,本文算法在快速去雾模型中具有很好的表现。

本文的贡献如下:

- (1)提出了一种新颖的多补丁和多尺度层级聚合网络结构(MPSHAN),用于快速去除图像中的非均匀雾气。
- (2)设计了层级融合模块 HFM,基于拼接式特征融合和通道注意力机制优化了层级特征融合效果。
- (3)对于多尺度部分的层级结构采用扩张卷积,以更密集的特征图提升多尺度部分的特征质量。
- (4)在混合损失函数中加入了频域损失,以恢复更好的边缘信息。

## 2 相关工作

### 2.1 传统去雾算法和深度学习去雾算法

去雾方法主要可以分为基于统计先验的传统算法和基于数据驱动的深度学习算法。其中很大一部分算法是基于大气散射模型<sup>[13]</sup>的,定义为:

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

其中, $I$ 表示有雾图像, $J$ 表示清晰无雾图像, $t$ 表示透射图, $A$ 表示全局大气光值, $x$ 表示像素点的空间位置。由式(1)可知,在 $t$ 和 $A$ 同时未知的情况下,依靠大气散射模型的一个约束条件求解 $J$ 是一个不适应问题。

早期的基于传统算法的工作,一般通过统计的先验信息添加约束条件,估算出 $t$ 和 $A$ ,从而恢复出 $J$ 。其中,He等<sup>[14]</sup>提出的暗通道先验(Dark Channel Prior,DCP)是一种代表性方法。他们通过先验的统计信息发现,无雾图像在RGB 3个颜色通道中通常存在一个低像素值。DCP假设其值为零获得约束条件,并结合大气光值的近似估算,从而恢复无雾图像。Zhu等<sup>[15]</sup>将红外图像信息加权融合到可见光图像中,以有效地恢复图像对比度。Chen等<sup>[16]</sup>通过逆通道和改进的滤

波方法,进一步提升了去雾效果。然而,基于统计先验的传统算法通常存在一些个例,如DCP在天空等亮度较高的地方并不成立,并且由于多次近似估算,得到的透射图并不精确,因此恢复的无雾图像质量有待提升。

而基于深度学习的算法,通过数据驱动直接学习特征到透射图或无雾图像之间的映射关系,打破了基于统计先验的传统算法的局限性,甚至可以不依赖于大气散射模型。随着深度学习的迅猛发展,许多相关去雾算法已经被提出。Cai等<sup>[4]</sup>提出的网络结构DehazeNet,端到端学习有雾图像到透射图之间的映射关系。Ren等<sup>[5]</sup>将多尺度卷积神经网络(Multi-scale Convolutional Neural Networks, MSCNN)用于去雾,通过多尺度卷积核获取更加有效的特征。Li等<sup>[6]</sup>提出的多合一去雾网络(All-in-One Dehazing Network, AOD-Net),将大气散射模型中的透射图和大气光值整合成一个参数,网络直接回归无雾图像。Zhao等<sup>[7]</sup>将生成对抗网络结构用于去雾,这种方法并不需要成对训练数据。Chen等<sup>[8]</sup>提出的门控上下文聚合网络(Gated Context Aggregation Network, GCANet)基于残差的方式,采用平滑扩张卷积和融合了级联不同级别信息的门控子网络实现了端到端去雾。Liu等<sup>[9]</sup>提出的GridDehazeNet使用基于注意力机制的多尺度模型,有效缓解了去雾网络中多尺度结构的瓶颈问题。Liu等<sup>[17]</sup>提出的三叉戟除雾网络(Trident Dehazing Network, TDN)取得了最佳的非均匀去雾效果,但网络模型复杂,每秒一到两帧的处理速度难以满足多数视觉任务的实时性要求。Qin等<sup>[10]</sup>提出的特征融合注意力网络(Feature Fusion Attention Network, FFA-Net)在网络中保持输入图像的分辨率,基于注意力机制的级联融合和残差思想,所恢复的无雾图像质量优秀但模型结构有待精简。

### 2.2 深度多补丁层级网络和深度多尺度层级网络

Zhang等<sup>[18]</sup>在图片去模糊工作中受多尺度网络启发,提出了深度多补丁层级网络DMPHN。单个层级由简易的基于残差块的编码器-解码器结构组成,层级之间通过编码特征以及解码特征,分别与上层的编码特征以及输入图像信息残差连接实现层级之间的信息融合。该网络通过将精细网格内的图像输入到自编码器结构中来获得局部感受野的信息,将底层得到的细化特征辅助上层更大补丁图像的特征生成,由细到粗实现了优秀的图像去模糊效果。

Das等<sup>[12]</sup>把去模糊领域中的多补丁层级网络用于非均匀去雾工作。网络结构部分使用了深度多补丁层级网络变体结构DMPHN1-2-4,直接学习有雾图像到无雾图像之间的映射关系。在他们的去雾工作中,对深度多尺度层级网络DMSHN也进行了广泛研究。该网络与深度多补丁层级网络在结构上相似,细微的区别之处在于,多尺度网络基于插值的方式输入不同下采样比例的有雾图像,从而获得不同尺度的信息。多尺度网络将下采样比例最大化,将得到的最大感受野的信息输入到更上层的网络中,用底层得到的更全局的信息来辅助更小感受野信息的生成,由粗到细实现了图像去雾。

### 2.3 扩张卷积

Chen等<sup>[19]</sup>在语义分割任务中,通过使用扩张卷积替换

了网络级联部分的下采样模块。这种方法在保证网络感受野的同时获得了更加密集的特征图,从而降低了网络最后上采样恢复原图尺寸信息的难度,实现了更好的语义分割效果。扩张卷积是在正常卷积核中插入一定比例的零值,在增大卷积核尺寸的同时并不会增加浮点计算,因此可以在扩大感受野的同时不增大计算开销。

## 2.4 注意力机制

随着注意力机制在自然语言领域的兴起,近年来在计算机视觉领域中也进行了一系列注意力机制相关工作的探索。Hu等<sup>[20]</sup>提出的通道注意力机制(Squeeze-and-Excitation, SE)在把空间域信息全局池化后留下通道维度,基于通道间的相互依赖关系重新校正各通道的特征响应,激励有用的特征,抑制无用的冗余特征。在此基础上,Woo等<sup>[21]</sup>在池化层中加入了最大池化,进一步提升了通道注意力的性能,并提出了空间域注意力,从而对特征图赋予更细化的注意力权重。相比SE,Zhong等<sup>[22]</sup>提出的挤压和注意(Squeeze-and-Attention, SA)模块不再将特征图空间域信息全部压缩,而是通过局部池化捕获空间像素组的注意力权重。Hou等<sup>[23]</sup>采用条形池化建模长距离上下文信息,并通过叠加纵横条形信息构建局部信息,这种结构获得的注意力权重在空间域具有全局感受野。

考虑到计算开销和运行时间,本文将采用高缩小率的通道注意力机制SE。

## 3 本文方法

本文提出了一种快速的基于多补丁和多尺度的层级聚合网络结构(MPSHAN),用于去除图像中的非均匀雾气。此外,本文设计了层级融合模块,以及对多尺度层级结构进行扩张卷积,优化了特征融合质量,增强了网络的表达力。

### 3.1 多补丁和多尺度层级聚合网络结构(MPSHAN)

为了充分利用局部和全局信息,本文聚合了多补丁和多尺度结构,提出了一种新颖的网络架构MPSHAN。它融合了两种网络结构的特性:一种是深度多补丁层级结构DMPHN1-2-4变体,另一种是深度多尺度层级结构(DM-SHN)。

MPSHAN是一种多层次结构,如图1所示(未使用层级融合模块和扩张卷积),分为多补丁部分(Multi-patch Part, MPP)、主体部分(Main Part, MP)和多尺度部分(Multi-scale Part, MSP)。在最初的聚合结构MPSHAN中,多补丁部分和主体部分的聚合,就是深度多补丁层级网络DMPHN1-2-4的变体结构,而多尺度部分和主体的聚合就是深度多尺度层级结构。每一层级由独立的编码器-解码器结构组成,层级之间进行信息融合。编码器结构包括3个卷积层和6个简单残差块,每个简单残差块包含两个卷积层、一个线性整流函数(Rectified Linear Unit, ReLU)层以及一个残差连接。而解码器将对称位置的两个步长为2的下采样卷积替换成转置卷积,从而恢复图像尺度。图中编码器和解码器上的参数由上至下依次为输入通道数、输出通道数、卷积核大小以及步长。

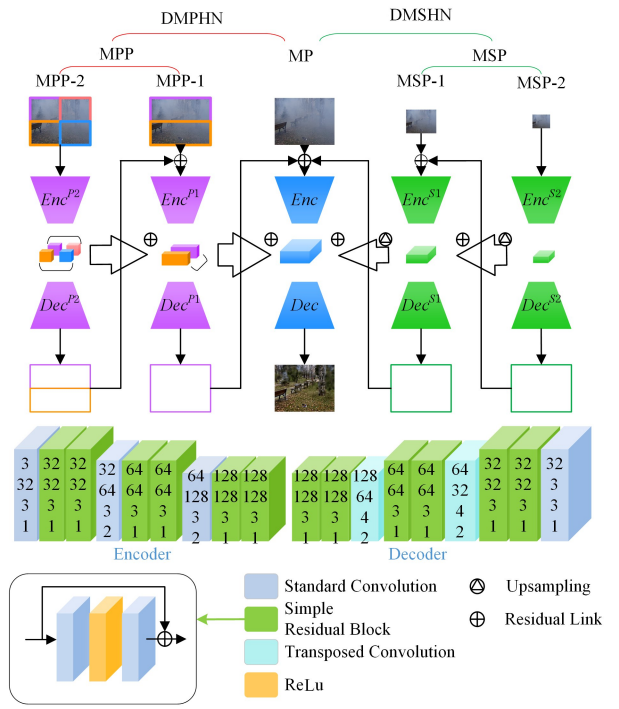


图1 多补丁和多尺度的层级聚合网络结构示意图

Fig. 1 Structure diagram of multi-patch and multi-scale hierarchical aggregation network

具体而言,假设输入有雾图像为 $I$ ,在MPP中,第一个层级结构MPP-1中输入图像将被水平分为 $I_{11}^p$ 和 $I_{12}^p$ 上下两层,第二个层级结构MPP-2中输入图像又在垂直方向被分割,上半部分由左至右表示为 $I_{21}^p$ 和 $I_{22}^p$ ,下半部分由左至右表示为 $I_{23}^p$ 和 $I_{24}^p$ 。在初始的多尺度部分,MSP中有雾图像通过下采样插值后输入网络中,第一层MSP-1双线性插值比率为 $1/2$ ,表示为 $I_{12}^s$ ,第二层MSP-2双线性插值比率为 $1/4$ ,表示为 $I_{14}^s$ 。编码器-解码器结构在MPP- $i$ ( $i$ 表示层级)中将被分解为编码器 $Enc^p_i$ 和解码器 $Dec^p_i$ ,在MP中表示为 $Enc$ 和 $Dec$ ,在MSP- $i$ 中表示为 $Enc^s_i$ 和 $Dec^s_i$ 。

多补丁部分通过局部感受野的细粒度信息辅助更粗粒度信息的生成。首先在MPP-2中,最细粒度的4个补丁块 $I_{2j}^p$ 将被输入编码器 $Enc^{p2}$ 中,按照对应空间位置生成4个编码特征图 $M_{1j}^{p2}$ ,表示为:

$$M_{1j}^{p2} = Enc^{p2}(I_{2j}^p), \forall i \in \{1, 2\}, \forall j \in \{1, 2\} \quad (2)$$

我们注意到,在编码器的每一个层级结构内同等结构将共享参数。在输入到解码器之前,将特征图 $M_{1j}^{p2}$ 水平串联生成新的特征图 $N_1^{p2}$ 。

$$N_1^{p2} = Cat_H(M_{11}^{p2}, M_{12}^{p2}), \forall i \in \{1, 2\} \quad (3)$$

其中, $Cat_H(\dots, \dots)$ 表示将特征图水平方向拼接,拼接后的特征图 $N_1^{p2}$ 将被输入到解码器 $Dec^{p2}$ 中生成类残差部分 $Q_1^{p2}$ :

$$Q_1^{p2} = Dec^{p2}(N_1^{p2}), \forall i \in \{1, 2\} \quad (4)$$

所获得的类残差部分 $Q_1^{p2}$ 将会把信息传播到下一层,补充输入图像信息 $I_{11}^p$ ,之后输入到MPP-1中的编码器 $Enc^{p1}$ ,得到特征图 $M_i^{p1}$ :

$$M_i^{p1} = Enc^{p1}(I_{11}^p + Q_1^{p2}), \forall i \in \{1, 2\} \quad (5)$$

类似地,式(6)中 $Cat_V(\dots, \dots)$ 表示将特征图垂直串联,将特征图 $M_i^{p1}$ 垂直串联后,在新的特征图中融合之前层级更细粒

度的特征  $N_i^{P2}$ , 经解码器  $Dec^{P1}$  解码后获得类残差信息  $Q^{P1}$ :

$$N^{P1} = Cat_V(M_1^{P1}, M_2^{P1}) + Cat_V(N_1^{P2}, N_2^{P2}) \quad (6)$$

$$Q^{P1} = Dec^{P1}(N^{P1}) \quad (7)$$

多尺度部分与多补丁部分的结构类似,用更全局的信息辅助图像恢复。在最初的工作中,多尺度部分通过插值的方法获得不同尺度的信息。在 MSP-2 中将下采样四倍的图像  $I_{1/4}^{S2}$  输入编码器  $Enc^{S2}$ , 获得编码特征  $N^{S2}$ , 送入解码器  $Dec^{S2}$  后, 获得类残差信息  $Q^{S2}$ , 表示为:

$$N^{S2} = Enc^{S2}(I_{1/4}^{S2}) \quad (8)$$

$$Q^{S2} = Dec^{S2}(N^{S2}) \quad (9)$$

类似多补丁结构, 将会在网络下一层级, 即 MSP-1 的编码器-解码器结构中融合 MSP-2 中的更大感受野的特征信息。由于融合特征图尺寸不一, 融合过程中将对特征图进行上采样, 表示为:

$$M^{S1} = Enc^{S1}(I_{1/2}^{S1} + Up(Q^{S2})) \quad (10)$$

$$N^{S1} = M^{S1} + Up(N^{S2}) \quad (11)$$

$$Q^{S1} = Dec^{S1}(N^{S1}) \quad (12)$$

其中,  $M^{S1}$  是 MSP-1 中的编码特征,  $Up(\dots)$  表示使用双线性插值上采样两倍,  $N^{S1}$  表示融合后新的特征图,  $Q^{S1}$  是解码后的类残差信息。

为了同时利用多补丁结构和多尺度结构的优点, 需将这二者进行聚合。即在网络主干部分将 MP 的编码器-解码器结构同时融合多补丁部分和多尺度部分的局部和全局的信息, 表示为:

$$M = Enc(I + Q^{P1} + Up(Q^{S1})) \quad (13)$$

$$N = M + Cat_V(M_1^{P1}, M_2^{P1}) + Up(M^{S1}) \quad (14)$$

$$J = Dec(N) \quad (15)$$

输入图像  $I$  融合多补丁部分残差信息  $Q^{P1}$  和上采样后的多尺度部分残差信息  $Up(Q^{S1})$  后, 输入主干网络的编码器  $Enc$ , 得到特征图  $M$ , 融合多补丁和上采样后的多尺度的编码信息  $M_i^{P1}$  和  $Up(M^{S1})$  后得到新的特征图  $N$ , 通过解码器  $Dec$  后得到恢复的无雾图像  $J$ 。

### 3.2 层级融合模块(HFM)

初始网络中所有的跨层级特征融合部分都是采用残差连接的。但是残差式融合是一种比较直接的先验表达, 精简的同时也会限制网络的表达力, 尤其是当网络为轻量级水平时, 这种限制尤为明显。因此, 本文提出了 HFM, 如图 2 所示。该模块把来自跨层级的融合特征拼接起来输入下一层, 以解耦更多潜在的非线性表达, 而不是简单使用加法。在生成下一层特征图时, 残差连接是拼接连接中潜在的一种方式。

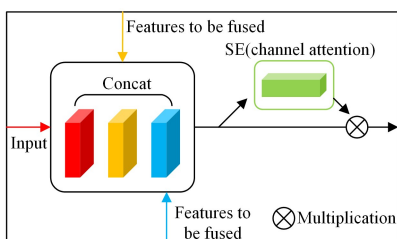


图 2 层级融合模块结构

Fig. 2 Hierarchical fusion module structure

由于当前特征图中的每一个通道是分别基于前层特征图

卷积变换后产生的, 一般的网络结构无法充分利用通道域特征之间的相互依赖关系, 导致获得的特征质量在通道方向参差不齐或冗余, 这将限制网络学习的能力。因此, 本文在 HFM 模块中引入通道注意力机制  $SE^{[20]}$ , 通过动态的通道域特征响应的再校核, 激励优质的特征, 抑制较差或冗余的特征, 从而强化网络表达力。

如图 3 所示, 在网络结构中本文将使用 HFM 模块取代层级间所有的残差连接, 以提升信息融合质量。

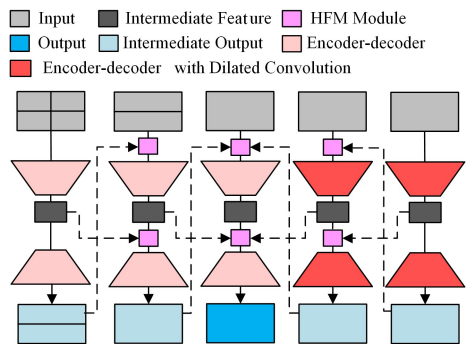


图 3 MPSHAN 的体系结构

Fig. 3 Architecture of MPSHAN

### 3.3 基于扩张卷积的多尺度部分

由于多尺度部分中上采样插值得到的特征图过于粗糙, 而网络要处理的非均匀图像去雾问题中, 雾气分布较为复杂, 因此我们认为粗糙的特征图是网络融合效果的瓶颈之一。受 Chen 等的相关工作<sup>[19]</sup>的启发, 本文将扩张卷积和多尺度部分相结合, 以获得更密集的特征图, 增强层级特征的融合效果。简而言之, 如图 3 所示, 该结构使用层级融合模块和扩张卷积, 在多尺度部分, 本文将不再对初始图像  $I$  进行下采样, 而是直接把初始分辨率图像送入网络。本文将不再使用标准卷积, 而是对多尺度部分的两个层级结构使用不同扩张比率的扩张卷积(文中设置为 2 和 3), 以获得更大的感受野, 且不会降低特征图的分辨率。

具体来说, 多尺度部分的第二层 MSP-2 将直接把有雾图像  $I$  输入编码器  $Enc^{S2}$ , 获得编码特征图  $N^{S2}$ , 继而输入本层解码器  $Dec^{S2}$ , 获得类残差部分  $Q^{S2}$ , 表示为:

$$N^{S2} = Enc^{S2}(I) \quad (16)$$

$$Q^{S2} = Dec^{S2}(N^{S2}) \quad (17)$$

与基于插值方式的融合不同, 由于使用了扩张卷积, MSP 部分的特征尺度将保持一致, 在与下一层 MSP-1 融合时将不再需要把特征上采样, 表示为:

$$M^{S1} = Enc^{S1}(I + Q^{S2}) \quad (18)$$

$$N^{S1} = M^{S1} + N^{S2} \quad (19)$$

$$Q^{S1} = Dec^{S1}(N^{S1}) \quad (20)$$

同样, 在 MSP-1 和主网络部分融合时也不需要再进行上采样过程。

## 4 实验结果及分析

### 4.1 数据集

为了对模型的有效性进行验证, 本文基于真实的非均匀雾气数据集 NH-HAZE<sup>[11]</sup> 展开了广泛的消融和对比实验。

该高分辨率数据集的训练集中包含 45 张非均匀有雾图,并提供了真实清晰的图像;验证集只提供了 5 张有雾图像,并不提供对应的真实清晰图像,需要上传算法去雾后的图像至官方服务器,由 NH-HAZE 数据集官方返回评估结果;测试集也包括 5 张有雾图像,同样不提供真实的清晰图像,但用来测试的服务器已经关闭。为了弥补测试集的缺失,并做到广泛准确的定量分析,我们重新划分该数据集,在原训练集中取出 5 对高分辨率图像作为验证集,原官方服务器的验证评估将用于测试。此外,本文除了使用真实的非均匀雾气数据集 NH-HAZE 进行评估外,还在真实的密集雾数据集 Dense-HAZE<sup>[24]</sup>上与各种方法进行了对比。在 Dense-HAZE 数据集上,与 NH-HAZE 数据集类似,本文选取 45 对图像进行训练,5 对图像作为验证集以及 5 对图像用于测试,以评估算法的性能。

为了充分利用  $1600 \times 1200$  高分辨率图片,在网络训练阶段,本文把训练图片裁剪为原有尺度的  $1/10$ ,并以  $1/20$  大小的采样步长进行间隔采样。在非均匀雾气数据集 NH-HAZE 上将获得的 16245 张图片,用于最终的训练测试。由于卷积在应用尺度上具有一致性,基于不同尺度训练得到的模型将可以直接用于高分辨率非均匀雾图的恢复。这样,在验证和测试阶段,相当于分别评测了 500 张训练图片。单张高分辨率图像中用于评估的内容丰富,因此在评测模型性能上具有很好的泛化能力。

## 4.2 实验设置

### 4.2.1 训练细节

本文基于 Pytorch 深度学习框架,在英伟达 RTX2080 型号显卡上进行网络训练。我们使用 Adam 优化器<sup>[25]</sup>的默认参数,学习率从 0.0001 开始逐渐衰减,表示为:

$$lr = 0.0001 \times \cos\left(\frac{\pi}{2} \times \frac{epoch}{EPOCH}\right) \quad (21)$$

其中,  $lr$  表示学习率,  $epoch$  表示当前的训练批次,  $EPOCH$  表示总训练批次。

### 4.2.2 损失函数

本文使用混合损失  $L_H$ <sup>[12]</sup>,它包含了用于重构的平均绝对误差  $L_1$  和均方误差  $L_2$ ,用于感知衡量的感知损失  $L_p$ <sup>[26]</sup>以及用于平滑图像的总变差(Total Variation, TV)损失  $L_{tv}$ ,表示为:

$$\begin{cases} L_H = \lambda_1 L_1 + \lambda_2 L_2 + \lambda_p L_p + \lambda_{tv} L_{tv} \\ L_1 = \| J_{Net} - J_{GT} \|_1 \\ L_2 = \| J_{Net} - J_{GT} \|_2 \\ L_p = \| \varphi(J_{Net}) - \varphi(J_{GT}) \|_2 \\ L_{tv} = \| \nabla_x(J_{Net}) \|_2 + \| \nabla_y(J_{Net}) \|_2 \end{cases} \quad (22)$$

其中,  $J_{Net}$  表示网络恢复的无雾图像,  $J_{GT}$  表示真实清晰图像,各项系数  $\lambda_1 = 0.6$ ,  $\lambda_2 = 0.4$ ,  $\lambda_p = 6 \times 10^{-3}$ ,  $\lambda_{tv} = 2 \times 10^{-8}$ ,  $\varphi(\dots)$  表示将图像输入预训练好的 VGG16<sup>[27]</sup>网络中以抽取感知特征,  $\nabla_x(\dots)$  和  $\nabla_y(\dots)$  分别表示求取图像在水平和垂直方向上的梯度矩阵。除此之外,在混合损失  $L_H$  中加入频域损失  $L_F$ ,以恢复更好的边缘质量,损失函数  $L$  表示为:

$$\begin{cases} L = L_H + \lambda_F L_F \\ L_F = \lambda_1 \| conv_G(J_{Net}) - conv_G(J_{GT}) \|_1 + \\ \lambda_2 \| conv_G(J_{Net}) - conv_G(J_{GT}) \|_2 \end{cases} \quad (23)$$

其中,  $\lambda_F$  表示频域损失系数,  $conv_G(\dots)$  表示使用 Laplace 二阶微分算子离散核函数处理图像矩阵获得频域信息。 $\lambda_F$  的具体选用将在实验中进行比较。

### 4.2.3 评估方法

本文基于峰值信噪比 PSNR 和结构性损失 SSIM<sup>[28]</sup> 对各实验结果进行评估。

## 4.3 消融实验

### 4.3.1 模型结构消融实验

本文对模型结构的选取进行了广泛的消融实验,表 1 中的  $D$  表示使用扩张卷积,运行时间为处理单张图片的平均时间。在测试集评估中,相比深层多补丁结构 DMPHN,本文提出了初始多补丁和多尺度层级聚合结构 MP SHAN, PSNR 增长了 0.27 dB, SSIM 增长了 0.0086,运行时间延长了 48% (0.012s),由于共享了主干网络部分,运行时间延长短于深度多尺度网络结构 DMSHN。为了更好地利用层级融合信息,本文用层级融合模块 HFM 替换层级残差连接。相比未使用该模块, PSNR 延长了 0.31 dB, SSIM 增长了 0.0031,运行时间延长了 14% (0.005s),说明 HFM 相比初始聚合方法在提升 PSNR 和抑制运行时间延长方面更占优势。在聚合结构 MP SHAN 上,单独使用层级扩张卷积获得密集特征图时, PSNR 和 SSIM 几乎没有变化。而搭配使用 HFM 时,相比单独添加 HFM 模块, PSNR 增加了 0.037 dB, SSIM 增加了 0.0111,运行时间仅延长了 5% (0.002s),模型性能显著提升。我们合理推测,模型最初的层级残差连接是限制网络表达力的瓶颈,即使有更好的特征融合信息,如扩张卷积获得的密集特征,也难以提升聚合结构的网络性能。在使用 HFM 后,解耦了更多潜在的非线性表达,并优化了融合特征质量,从而打破了限制网络表达力的瓶颈,提升了去雾效果。整体来看,本文最终的聚合网络结构 MP SHAN 模型参数仅有 42.8 M,相比多补丁结构 DMPHN, PSNR 提升了 0.95, SSIM 提升了 0.0228,运行时间延长了 76% (0.19s)。

表 1 模型结构消融研究结果

Table 1 Ablation study results of model structure

Algorithm	MPP	MSP	HFM	D	PSNR/dB		SSIM		Runtime/s
					Val	Test	Val	Test	
DMPHN <sup>[12]</sup>	✓	—	—	—	18.18	16.94	0.6749	0.6177	0.025
DMSHN <sup>[12]</sup>	—	✓	—	—	17.71	16.42	0.6615	0.5991	<b>0.015</b>
MP SHAN	✓	✓	—	—	18.43	17.21	0.6911	0.6263	0.037
MP SHAN+HFM	✓	✓	✓	—	18.87	17.52	0.6989	0.6294	0.042
MP SHAN+D	✓	✓	—	✓	18.52	17.16	0.6941	0.6314	0.039
MP SHAN+HFM+D(ours)	✓	✓	✓	✓	<b>19.46</b>	<b>17.89</b>	<b>0.7044</b>	<b>0.6405</b>	<b>0.044</b>

而之前的工作,从多尺度网络 DMSHN 到多补丁网络 DMPHN,PSNR 提升了 0.52dB,SSIM 增长了 0.0186,运行时间延长 67%(0.01s)。在运行时间增加比例接近的情况下,本文最终的聚合网络结构 MPSHAN 的 PSNR 增长几乎是之前工作的两倍,SSIM 增长也略有提升。进一步,根据验证集对消融实验进行辅助验证,本文最终的结构相比多补丁网络 DMPHN,PSNR 提升了 1.28dB,SSIM 提升了 0.0295,而多补丁网络 DMPHN 相比 DMSHN,PSNR 提升了 0.47dB,SSIM 提升了 0.0134,所提出的聚合网络结构 MPSHAN 在 PSNR 和 SSIM 上的增长都是之前工作的两倍以上。

在之后的实验中,所有聚合结构 MPSHAN 都默认使用了层级融合模块和层级扩张卷积。

#### 4.3.2 损失函数消融实验

如表 2 所列,分别设置  $\lambda_F = 0.1, \lambda_F = 0.05, \lambda_F = 0.02$  以及  $\lambda_F = 0$  进行了频域损失函数消融实验。

表 2 频域损失函数的消融研究结果

Table 2 Ablation study results of frequency domain loss function

$\lambda_F$	PSNR/dB		SSIM	
	Val	Test	Val	Test
0	19.46	17.89	0.7044	0.6405
0.02	<b>19.63</b>	<b>17.97</b>	<b>0.7122</b>	<b>0.6424</b>
0.05	19.43	17.88	0.7106	0.6412
0.1	19.33	17.79	0.7110	0.6417

在测试集评估中,相比不加入频域损失,当  $\lambda_F = 0.1$  时 PSNR 降低了 5.6%,SSIM 提升了 1.9%,整体恢复效果略微降低;当  $\lambda_F = 0.05$  时,PSNR 近似相等,SSIM 提升了

1.1%,图像质量略微提升;当  $\lambda_F = 0.02$  时,PSNR 提升了 4.5%,SSIM 提升了 3%,图像恢复质量提升明显。在验证集评估中,同样在  $\lambda_F = 0.02$  时,图像恢复质量提升最多,PSNR 提升了 8%,SSIM 提升了 11%。实验结果表明,添加适当比例的频域损失,有利于恢复更好的图像质量。

#### 4.4 方法对比

##### 4.4.1 NH-HAZE 上的方法对比

在 NH-HAZE<sup>[11]</sup>数据集上与各种方法进行了对比,在验证和测试结果中,本文的 PSNR 和 SSIM 都取得了最好的成绩。定量的结果如表 3 所列,其中 DCP<sup>[14]</sup>方法在 CPU 上进行测试,其余均在 GPU 上进行。

表 3 不同方法在 NH-HAZE 数据集上的定量对比

Table 3 Quantitative comparison of different methods on NH-HAZE

Algorithm	PSNR/dB		SSIM		Runtime/s
	Val	Test	Val	Test	
DCP <sup>[14]</sup>	10.66	11.32	0.4426	0.4184	>1
AOD-Net <sup>[6]</sup>	14.21	13.04	0.5274	0.4429	<b>0.004</b>
GCANet <sup>[8]</sup>	16.46	14.63	0.6272	0.5421	0.205
GridDehazeNet <sup>[9]</sup>	17.87	16.55	0.6561	0.6036	0.027
FFA <sup>[10]</sup>	17.07	15.14	0.6457	0.5691	>1
DMPHN <sup>[12]</sup>	18.18	16.94	0.6749	0.6177	0.025
MPSHAN (ours)	<b>19.63</b>	<b>17.97</b>	<b>0.7122</b>	<b>0.6423</b>	0.044

可视化结果的对比如图 4 所示,相比其他方法,可以看到本文方法所恢复的无雾图像在局部纹理和对比度方面更好。在快速去雾模型中,本文提出的多补丁和多尺度层级聚合网络不管是定量还是定性的评估对比都取得了更好的效果。

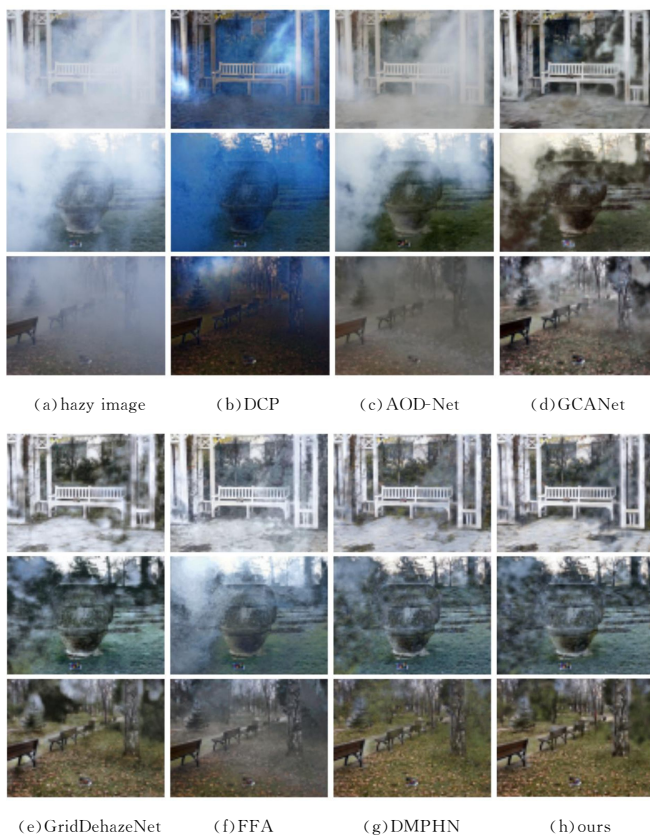


图 4 不同去雾方法在 NH-HAZE 数据集上的定性对比

Fig. 4 Qualitative comparison of different dehazing methods on NH-HAZE

## 4.4.2 Dense-HAZE 上方法的对比

在均匀的密集雾气数据集 Dense-HAZE<sup>[24]</sup>上,对本文算法的有效性进行了进一步的评估。表 4 列出了本文方法与各种方法的对比结果。相比深度多补丁层级结构 DMPHN,本文方法在验证集上 PSNR 提高了 0.62 dB, SSIM 提高了 0.0163,在测试集上 PSNR 提高了 0.56 dB, SSIM 提高了 0.0257,并且仍然保持了快速的运行速度。

表 4 不同方法在 Dense-HAZE 数据集上的定量对比

Table 4 Quantitative comparison of different methods on

Algorithm	Dense-HAZE				Runtime/s
	PSNR/dB		SSIM		
	Val	Test	Val	Test	
DCP <sup>[14]</sup>	11.76	10.86	0.3623	0.4	>1
AOD-Net <sup>[6]</sup>	13.38	13.65	0.4155	0.5002	<b>0.004</b>
GCANet <sup>[8]</sup>	11.40	13.37	0.3398	0.4476	0.339
GridDehazeNet <sup>[9]</sup>	13.96	13.60	0.4425	0.4650	0.027
FFA <sup>[10]</sup>	14.08	14.25	0.4557	0.5019	>1
DMPHN <sup>[12]</sup>	13.96	14.02	0.4759	0.4936	0.025
MPSHAN (ours)	<b>14.58</b>	<b>14.58</b>	<b>0.4922</b>	<b>0.5193</b>	0.042

图 5 给出了去雾后的视觉对比,相比其他方法,本文恢复的无雾图像的纹理质量和色彩饱和度更好。

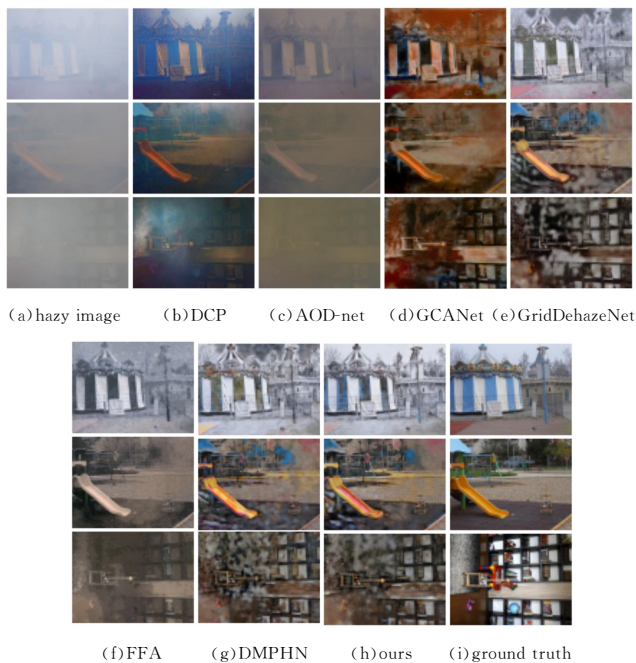


图 5 不同去雾方法在 Dense-HAZE 数据集上的定性对比  
Fig. 5 Qualitative comparison of different dehazing methods on Dense-HAZE

**结束语** 为了快速有效地去除真实图像中的非均匀雾气,本文提出了一种多补丁和多尺度层级聚合网络结构(MP-SHAN),并通过层级融合模块、层级结构扩张卷积以及频域损失对算法进行了优化。在模型保持快速运行的情况下,图像去雾效果得到了明显提升。在定量和定性的实验中证明了本文算法的有效性和鲁棒性。但是,更多层级的引入不可避免地提高了模型复杂度,增加了单张图片的处理时间。在未来的工作中,我们将进一步优化网络的拓扑结构,利用神经网络中模型压缩和特征融合技术实现更高效的图像去雾算法。

## 参考文献

- [1] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions Pattern Analysis Machine Intelligence, 2017, 39(6): 1137-1149.
- [2] SHELHAMER E, LONG J, DARRELL T. Fully Convolutional Networks for Semantic Segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [3] RANJAN R, PATEL V M, CHELLAPPA R. HyperFace: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(1): 121-135.
- [4] CAI B, XU X, 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.
- [5] REN W, LIU S, ZHANG H, et al. Single Image Dehazing via Multi-scale Convolutional Neural Networks [C] // Proceedings of the European Conference on Computer Vision (ECCV). Springer, 2016: 154-169.
- [6] LI B, PENG X, WANG Z, et al. AOD-Net: All-in-One Dehazing Network [C] // Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017: 4780-4788.
- [7] ZHAO J, ZHANG J, LI Z, et al. DD-CycleGAN: Unpaired image dehazing via Double-Discriminator Cycle-Consistent Generative Adversarial Network [J]. Engineering Applications of Artificial Intelligence, 2019, 82: 263-271.
- [8] CHEN D, HE M, FAN Q, et al. Gated Context Aggregation Network for Image Dehazing and Deraining [C] // Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2019: 1375-1383.
- [9] LIU X, MA Y, SHI Z, et al. GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing [C] // Proceedings of the 2019 IEEE International Conference on Computer Vision (ICCV). IEEE, 2019: 7313-7322.
- [10] QIN X, WANG Z, BAI Y, et al. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing [C] // Proceedings of the AAAI Conference on Artificial Intelligence. AAAI, 2020: 11908-11915.
- [11] ANCUTI C O, ANCUTI C, TIMOFTE R. NH-HAZE: An Image Dehazing Benchmark with Non-Homogeneous Hazy and Haze-Free Images [C] // Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2020: 1798-1805.
- [12] DAS S D, DUTTA S. Fast Deep Multi-patch Hierarchical Network for Nonhomogeneous Image Dehazing [C] // Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2020: 1994-2001.
- [13] HIDE R. Optics of the Atmosphere: Scattering by Molecules and Particles [J]. Physics Bulletin, 1977, 28(11): 521.
- [14] HE K M, SUN J, TANG X O. Single image haze removal using

- dark channel prior[C]//Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE,2009:1956-1963.
- [15] ZHU Z,HUANG R,ZANG T G,et al. Single Image Defogging Method Based on Weighted Near-Infrared Image Fusion[J]. Computer Science,2020,47(8):241-244.
- [16] CHEN Z Y, LONG D Y, WANG X, et al. Dark Channel Defogging Algorithm Based on Inverse Channel and Improved Guided Filtering[J]. Computer Engineering,2021,47(6):245-252.
- [17] LIU J, WU H, XIE Y, et al. Trident Dehazing Network[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2020:1732-1741.
- [18] ZHANG H, DAI Y, LI H, et al. Deep Stacked Hierarchical Multi-Patch Network for Image Deblurring[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE,2019:5971-5979.
- [19] CHEN L, PAPANDEIOU G, KOKKINOS I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018,40(4):834-848.
- [20] HU J, SHEN L, SUN G. Squeeze-and-Excitation Networks [C]//Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018: 7132-7141.
- [21] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block Attention Module[C]//Proceedings of the European Conference on Computer Vision(ECCV). Springer,2018:3-19.
- [22] ZHONG Z, LIN Z Q, BIDART R, et al. Squeeze-and-Attention Networks for Semantic Segmentation[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE,2020:13062-13071.
- [23] HOU Q, ZHANG L, CHENG M M, et al. Strip Pooling: Rethinking Spatial Pooling for Scene Parsing[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE,2020:4002-4011.
- [24] ANCUTI C O, ANCUTI C, SBERT M, et al. Dense-Haze: A Benchmark for Image Dehazing with Dense-Haze and Haze-Free Images[C]//Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP). IEEE,2019:1014-1018.
- [25] KINGMA D , BA J. Adam: A Method for Stochastic Optimization [C] // Proceedings of the International Conference for Learning Representations (ICLR). 2015:1-15.
- [26] JOHNSON J, ALAHI A, LI F F. Perceptual Losses for Real-Time Style Transfer and Super-Resolution[C]//Proceedings of the European Conference on Computer Vision (ECCV). Springer,2016:694-711.
- [27] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. arXiv:1409.1556, 2014.
- [28] ZHOU W, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing,2004,13(4):600-612.



**YANG Kun**, born in 1996, postgraduate. His main research interests include computer vision and deep learning.



**ZHANG Juan**, born in 1975, Ph.D, associate professor, is a member of China Computer Federation. Her main research interests include computer vision and so on.