

DOI:10.3969/j.issn.1003-5060.2025.03.005

多流选择性核去雾网络

侯 策, 杨依忠, 刘雪晴

(合肥工业大学 微电子学院, 安徽 合肥 230601)

摘要:目前图像去雾算法虽已取得较高的指标,但其恢复出的图像仍存在着颜色失真、雾霾残留、图像伪影等问题。文章提出一种可端到端训练的多流选择性核(multi-stream selective kernel, MSSK)去雾网络,借助图像的低频信息和高频信息优化去雾效果,实现在低频信息的指导下提高图像的颜色保真度,在高频信息的指导下优化图像的边缘和细节。该网络为多分支结构,通过分支指导块实现分支间的信息交互;采用选择残差块依据雾霾质量浓度自适应地调整感受野大小。在多个基准数据集上对网络进行评估的实验结果表明,文章所提出的网络具有较高的结构相似度(structural similarity, SSIM)和峰值信噪比(peak signal-to-noise ratio, PSNR)指标,能较好地解决颜色失真、雾霾残留、图像伪影等问题。

关键词:深度学习;图像去雾;多流结构;自适应融合;端到端训练

中图分类号:TP391.41

文献标志码:A

文章编号:1003-5060(2025)03-0320-07

Multi-stream selective kernel dehazing network

HOU Ce, YANG Yizhong, LIU Xueqing

(School of Microelectronics, Hefei University of Technology, Hefei 230601, China)

Abstract: Although current image dehazing algorithms have achieved high indicators, the restored images still have problems such as color distortion, haze residue, and image artifacts. Therefore, this paper proposes an end-to-end multi-stream selective kernel(MSSK) dehazing network, which can optimize the dehazing effect by using the low-frequency and high-frequency information of the image. The color fidelity of the image is improved under the guidance of low-frequency information, and the edges details of the image are optimized under the guidance of high-frequency information. The network is a multi-branch structure, and the information interaction between branches is realized by branch guidance blocks. It uses the selective residual block to adaptively adjust the size of the receptive field according to the haze concentration. A comprehensive evaluation was carried out on multiple benchmark datasets. The experimental results show that the proposed network has high structural similarity (SSIM) and peak signal-to-noise ratio(PSNR), which can solve the problems of color distortion, haze residue and image artifacts.

Key words: deep learning; image dehazing; multi-stream structure; adaptive fusion; end-to-end training

0 引 言

在雾霾天气下拍摄的图像,由于其在成像过程中受到大气中烟雾、灰尘和其他大气颗粒的吸

收和散射影响,会出现低对比度、模糊等问题,进而显著影响其他高级视觉任务的性能^[1-2],因此图像去雾算法得到了广泛的关注。

早期多数去雾研究工作针对图像本身进行增

收稿日期:2023-03-29;修回日期:2023-04-26

基金项目:安徽省自然科学基金资助项目(2208085MF177);中央高校基本科研业务费专项资金资助项目(JZ2021HGQA0262)

作者简介:侯 策(1998—),男,湖南岳阳人,合肥工业大学硕士生;

杨依忠(1975—),男,安徽合肥人,合肥工业大学副教授,硕士生导师,通信作者, E-mail: yangyizhong@hfut.edu.cn.

强^[3],即通过调整对比度、亮度等因素来改善雾霾图像视觉质量。随后,文献[4]提出了大气散射模型:

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

其中: $I(x)$ 为有雾图像; $J(x)$ 为无雾图像; x 为像素点位置; A 为全球大气光强; $t(x)$ 为场景透射率。在该理论的支撑下,基于模型的方法迅速发展,展现出良好的性能。

随着深度学习的迅速发展,卷积神经网络(convolutional neural networks, CNN)在各项计算机视觉任务中取得重大突破。在去雾领域,学者们尝试借助 CNN 学习 $t(x)$ 以结合大气散射模型恢复图像,或使用 CNN 学习有雾图像到无雾图像的直接映射或残差映射。

一幅图像的低频通道通常蕴含物体的基本结构、色彩信息等,高频通道则主要蕴含物体的边缘、细节等。对于有雾图像,雾霾主要分布于低频通道,而由于雾霾的影响高频通道的信息较少,表现为边缘不明显、细节少。

有雾图像、无雾图像及其各自的低通、高通滤波图像,如图1所示。

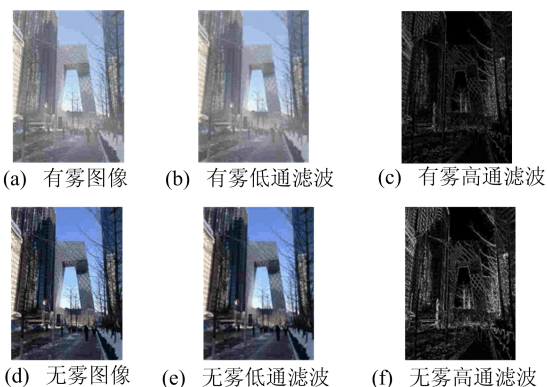


图1 有雾图像、无雾图像及其各自的低通、高通滤波图像

不同于目前大多数 CNN 算法仅将有雾图像作为输入,本文将有雾图像的低通、高通信息有效利用,提出一个多流选择性核(multi-stream selective kernel, MSSK)去雾网络。MSSK 由低通分支、去雾分支和高通分支组成,各分支均采用编解码器结构。低通分支用以指导去雾分支的色彩、对比度等信息的恢复,高通分支用以指导去雾分支的细节、边缘等信息的恢复。

为了解决现有 CNN 算法对雾霾质量浓度变化的感知不敏感问题,受文献[5]思想的启发,本文提出了一个选择残差块,凭借多扩张卷积分支结构,使模型能捕捉到雾霾质量浓度的细微变化。

实验结果表明,本文所提出的算法具有优秀的去雾性能。

1 相关工作

目前图像去雾的算法可以分为两类:基于物理模型的方法和基于深度学习的方法。

文献[6]发现无雾图像中的非天空区域像素至少有一个低强度的通道;文献[7]发现,无雾图像像素的饱和度和亮度近乎相同,当受到雾霾的干扰时,饱和度会下降、亮度会提升,雾质量浓度越高现象越明显;文献[8]提出了一种非局部先验,发现无雾图像中的颜色信息在色彩空间会呈现紧密的簇,并且会在雾的影响下变为线。以上基于物理模型的方法由于其对先验假设的依赖性,在某些恶劣环境下,去雾的效果会降低甚至失效。

文献[9]使用 CNN 进行图像去雾的早期尝试,虽结构简单,但相比于传统的算法能更准确地估计 $t(x)$;文献[10]对大气散射模型进行重构,将需要分别估计的 A 和 $t(x)$ 进行联立,减少了整体误差;文献[11]通过将有雾图像分解,从基本分量和细节分量 2 个角度对有雾图像进行恢复;文献[12]提出了一个网格化的网络,减轻了不同尺度间潜在的精度干扰;文献[13]通过结合通道注意力和像素注意力,从通道角度和像素位置角度突出重要信息、抑制不重要的信息。以上基于深度学习的方法在去雾效果、适用范围等方面比传统方法更具优势,但现有方法恢复出的无雾图像存在着雾霾残留、颜色失真、细节缺失等问题。

2 MSSK 去雾网络结构

2.1 整体结构

MSSK 去雾网络具有低通分支、去雾分支和高通分支。低通分支将有雾图像的低通滤波图像作为输入,去雾分支将有雾图像作为输入。因有雾图像高通滤波图像难以提取到足够的信息去指导去雾分支,故高通分支也采用有雾图像作为输入,如图2所示。

对于低通滤波器和高通滤波器的选择,本文分别采用高斯滤波器和拉普拉斯滤波器。除了单一的卷积层和反卷积层外,每个分支还包含编码器块、选择残差块和解码器块 3 种模块,数量分别为 3、6、3。去雾分支在此基础上还额外具有 5 个分支指导块进行信息交互。为了防止浅层特征消失,各分支均通过跳跃连接将其与深层特征相

结合。

图像进入各分支后经过 2 次下采样和 2 次上采样。为了减轻下采样带来的信息损失,将下采

样后的特征通道数增加 1 倍,各尺度上的通道数分别为 16、32、64。相应地,上采样后的特征通道数变为原来的 1/2。

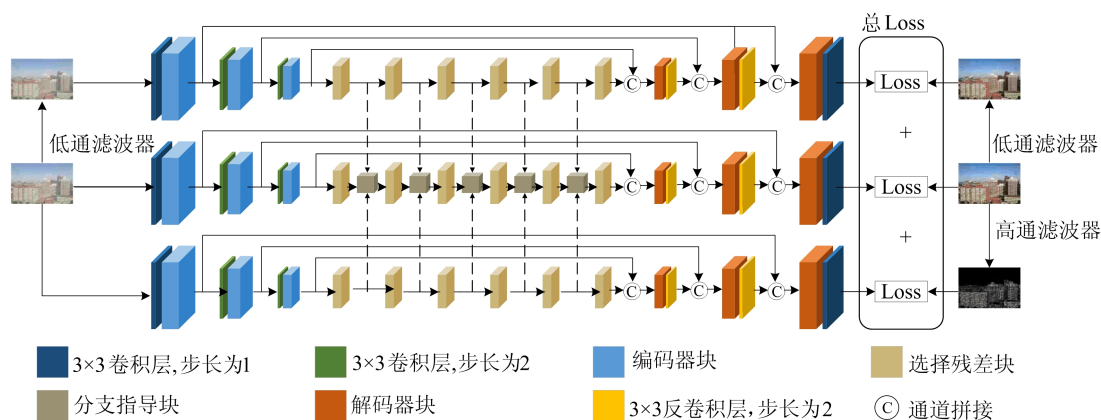


图 2 多流选择性核去雾网络结构

2.2 编码器块、解码器块与分支指导块

本文基于残差块与注意力机制^[14]构建出有效的编码器块和解码器块。两者具有相同的结构,共包括 2 个 3×3 的卷积层、1 个 ReLU 函数和 1 个通道注意力机制,并通过跳跃分支构成残差学习,如图 3a 所示。

注意力机制首先对输入进行全局平均池化操作以得到通道描述符;然后通过 1×1 卷积层对其压缩与还原后,使用 Sigmoid 函数得到代表通道重要程度的注意力权重;最后以元素相乘进行权重赋值,如图 3b 所示。

此外,本文通过分支指导块将低通分支和高通分支的信息提供给去雾分支,以减少颜色失真、加强边缘检测的能力。为了更好地信息交互,各分支提取特征层次需一致,即 3 个部分特征需要有相似的深度,因此在构建分支指导模块时以简单有效为导向。分支指导块为 3 输入、1 输出的结构,首先接受低通分支、高通分支的特征以使用卷积和 Sigmoid 函数生成各分支注意力图;其次将各分支注意力图与去雾分支特征用元素相乘操作结合;最后使用通道拼接操作得到去雾分支的最终特征,如图 3c 所示。

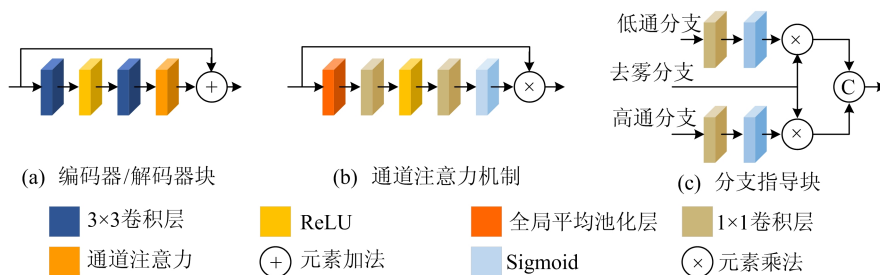


图 3 编码器块、解码器块与分支指导块结构

2.3 选择残差块

在神经科学领域,视觉皮层神经元的感受野大小能依据所受的刺激进行调节^[15-16],文献[5]提出了一种动态选择机制使 CNN 中各层神经元具有感受野自调节的能力。

基于此工作,本文提出了选择残差块,在面对大块的雾霾时,较大扩张率的卷积具有更强的响应程度,而在检测雾霾细微的变化时,较小扩张率的卷积具有更强的响应程度,以此达到感受野自

适应调整的效果,如图 4 所示。

首先,选择残差块将输入特征分 4 个支路进行学习,各支路分别使用扩张率不同的 3×3 扩张卷积,具有不同大小的感受野。具体表示为:

$$y_i^c = \delta(\text{Conv}_{3 \times 3, d=i}^c(x^c)) \quad (2)$$

其中: x^c 为输入特征; y_i^c 为各分支特征; $\text{Conv}_{3 \times 3, d=i}^c(\ast)$ 为扩张率为 i 的 3×3 扩张卷积, $i \in \{1, 3, 5, 7\}$; $\delta(\ast)$ 为 ReLU 激活函数;上标 c 为特征宽度或卷积核数。

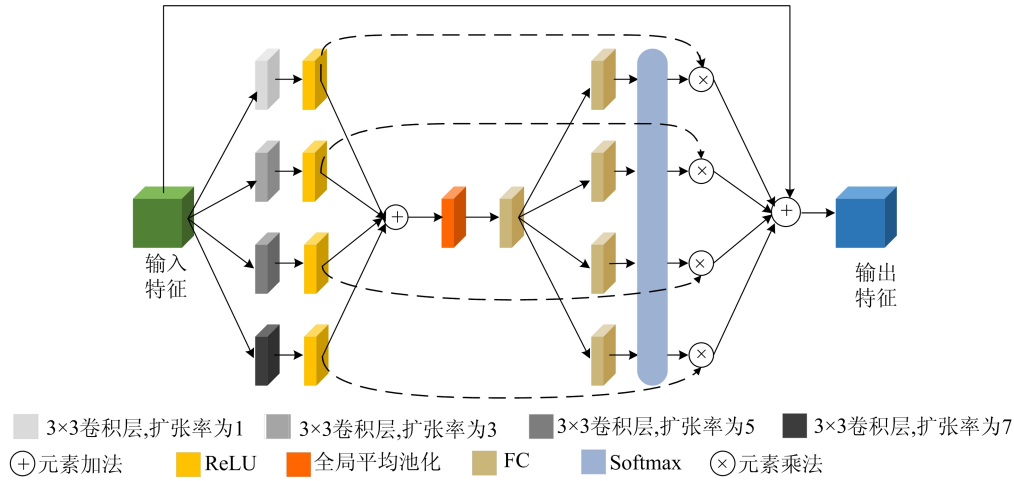


图 4 选择残差块结构

其次,通过元素加法将各支路特征聚合,计算全局信息后再通过 2 层全连接层。其中:第 1 层全连接层对信息进行压缩;第 2 层将信息还原并分为 4 个支路与前面相对应。具体表示为:

$$m_i^c = F_i^c(F^{c/8}(G(\sum_j y_j^c))) \quad (3)$$

其中: m_i^c 为各个分支的特征描述符; $G(\ast)$ 为全局平均池化; $F(\ast)$ 为全连接层。

最后,将各支路的特征堆叠后使用 Softmax 函数生成各支路的权重并通过元素乘法进行赋值。使用元素加法将各支路最终的特征聚合,并添加跳跃分支构成残差学习以防止梯度消失、加快模型收敛。具体过程表示为:

$$z^c = x^c \oplus \sum_i [\varphi(S(m_1^c, m_3^c, m_5^c, m_7^c)) \otimes y_i^c] \quad (4)$$

其中: $S(\ast)$ 为特征堆叠操作; $\varphi(\ast)$ 为 Softmax 函数; z^c 为最终输出特征。

2.4 损失函数

除了大部分文献使用的 L_1 损失外,本文使用结构相似度损失 L_s 来进一步帮助模型训练。 L_1 和 L_s 损失定义如下:

$$L_1 = \frac{1}{n} \sum_{i=1}^n |\hat{Y}_i - Y_i|, \\ L_s = \frac{1}{n} \sum_{i=1}^n [1 - \text{SSIM}(\hat{Y}_i, Y_i)] \quad (5)$$

其中: \hat{Y}_i, Y_i 分别为第 i 幅由模型恢复的去雾图像、对应场景下的真实标签;SSIM(\ast)为结构相似度函数。

相比于 L_1 损失, L_s 损失更关注两幅图像间结构内容上的差异,通过将两者相结合构建出模型最终的损失 L_t 。结合方式定义为:

$$L_{t-1} = c_1 L_1(\hat{Y}_1, Y_1) + c_d L_1(\hat{Y}_d, Y_d) + \\ c_h L_1(\hat{Y}_h, Y_h),$$

$$L_{s-t} = c_1 L_s(\hat{Y}_1, Y_1) + c_d L_s(\hat{Y}_d, Y_d) + \\ c_h L_s(\hat{Y}_h, Y_h),$$

$$L_t = \omega_1 L_{t-1} + \omega_2 L_{s-t} \quad (6)$$

其中: $\hat{Y}_1, \hat{Y}_d, \hat{Y}_h$ 分别为由模型恢复的低通图像、去雾图像和高通图像; Y_1, Y_d, Y_h 分别为前者的对应真实标签; $c_1, c_d, c_h, \omega_1, \omega_2$ 分别为各部分的权重。考虑到高通图像的通道数为 1,设置 $c_1 = 0.4, c_d = 0.4, c_h = 0.2$ 。本文更注重恢复图像与有雾图像结构上的差异,因此设置 $\omega_1 = 0.3, \omega_2 = 0.7$ 。

3 实验与结果分析

3.1 数据集

本文使用 RESIDE(realistic single image dehazing)^[17]、I-Haze(indoor haze)^[18] 和 O-Haze(outdoor haze)^[19] 数据集进行相关实验。对于 RESIDE 数据集,选取其中的 ITS(indoor training set)数据集和 OTS(outdoor training set)数据集进行训练,并在 SOTS(synthetic objective testing set)数据集上进行评估测试。ITS 包括 1 399 幅室内无雾图像和 13 990 幅室内有雾图像,OTS 包括 8 970 幅室外无雾图像和 313 950 幅室外有雾图像。SOTS 分为室内、室外 2 个部分,每个部分均包含 500 幅图像用于测试。

I-Haze 数据集由 35 对有雾和无雾的图像对构成,官方将其划分为 3 个子集,包括训练集 25 对、验证集 5 对、测试集 5 对。类似地,O-Haze 数据集包括训练集 35 对、验证集 5 对、测试集 5 对。由于 2 个数据集图像数量较少,此处将各自的验证集和测试集组合为具有 10 对图像的新数据集进行评估验证。

3.2 实验设置

对于 RESIDE 数据集,本文从原始图像中随机裁剪大小为 256×256 的图像块作为模型的输入。而对于 I-Haze 和 O-Haze 数据集,由于训练集图像数量较少、图像尺寸较大,为了防止模型过拟合并考虑到显卡内存限制,本文从原始图像中随机裁剪出 512×512 、 $1\ 024 \times 1\ 024$ 、 $1\ 536 \times 1\ 536$ 、 $2\ 048 \times 2\ 048$ 共 4 种不同大小的图像块,并将除 512×512 之外的图像块进行双线性插值操作调整至 512×512 大小作为模型输入。此外,为了突破 GPU 内存限制,本文将其测试图像也调整至 512×512 。

在训练中,使用 Adam(adaptive moment estimation)优化器,其参数 β_1 和 β_2 的值分别设置为 0.9 和 0.999,以 2×10^{-4} 为初始学习率,采用余弦退火策略^[20]逐步衰减学习率。RESIDE 数据集批量大小设置为 8,训练 2×10^5 步。对于 I-Haze 和 O-Haze 数据集批量大小设置为 2,训练 1×10^4 步。基于 Pytorch 框架使用 1 个 rtx 2080 ti GPU 实现模型。

3.3 实验结果

为评估所提算法的性能,选用结构相似度(structural similarity,SSIM)和峰值信噪比(peak signal-to-noise ratio,PSNR)作为定量评价指标,选取以下 5 种具有代表性的算法进行对比,即 DCP(dark channel prior)^[6]、AOD(all-in-one dehazing network)^[10]、GDN(GridDehazeNet)^[12]、GCA(gated context aggregation network)^[21]、FFA(feature fusion attention network)^[13]。

所选算法在各数据集上的 SSIM 和 PSNR 指标表现见表 1 所列。表 1 中部分数据采用相应文献汇报的结果,对于各文献尚未涉及的数据集,本文采用相同实验设置进行实验以确保公平有效。从表 1 可以看出,本文所提的方法在 SOTS 室外、I-Haze、O-Haze 等数据集上取得了最优指标,在 SOTS 室内数据集上表现次优,与最优的 FFA 表现相差不大。各算法在 RESIDE 数据集上的视觉表现如图 5 所示。图 5 中:前 2 行为室内图像;后 2 行为室外图像。

表 1 各数据集的 SSIM 和 PSNR 指标

数据集	指标	DCP	AOD	GDN	GCA	FFA	本文
SOTS 室内	SSIM	0.819 4	0.827 9	0.983 6	0.980 0	0.988 6	0.988 2
	PSNR	21.73	20.62	32.16	30.23	36.39	34.40
SOTS 室外	SSIM	0.814 2	0.845 5	0.981 9	0.943 9	0.984 0	0.985 4
	PSNR	17.25	18.74	30.86	27.99	33.57	33.71
I-Haze	SSIM	0.681 6	0.754 1	0.782 9	0.810 9		0.875 1
	PSNR	13.09	15.68	16.31	21.26		21.42
O-Haze	SSIM	0.758 9	0.742 8	0.843 1	0.876 0		0.898 7
	PSNR	17.06	19.06	21.66	23.84		25.35

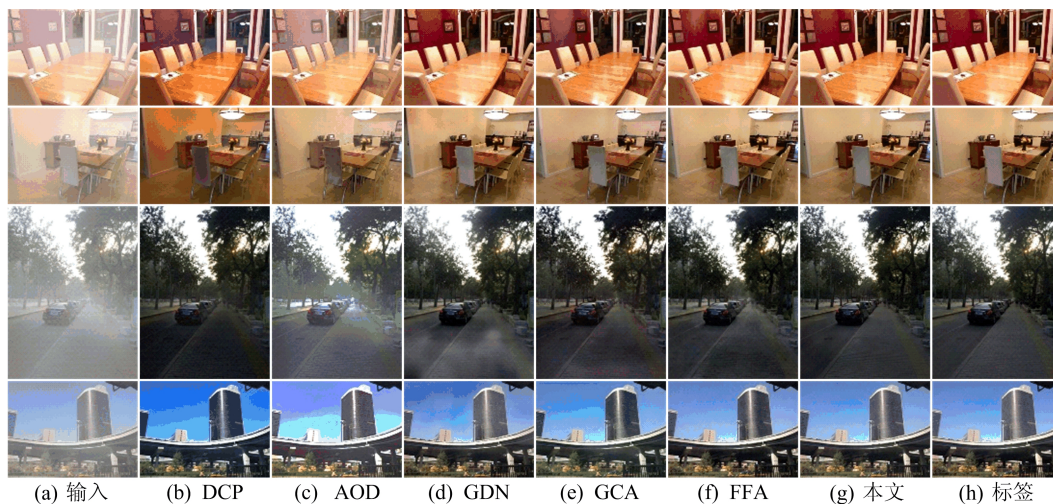


图 5 不同算法在 SOTS 数据集上的定性比较

由图 5 可知:DCP 恢复出的图像颜色失真问题较明显,无法良好地处理天空区域,如第 4 幅的天空区域;AOD 恢复出的图像存在颜色失真和较大面积的雾霾残留;GDN 和 GCA 相比于前两者,虽颜色质量都有明显的提升,但依旧存在薄雾残留、存在伪影等问题;FFA 在第 1 幅图像中的左上角存在着小面积的雾霾残留,第 3 幅图像中的街道部分存在伪影;相比于其他算法,本文提出的方法具有雾霾残留少,颜色保真度高、无伪影等

方面优势。

各算法在 I-Haze 数据集和 O-Haze 数据集上的视觉表现如图 6 所示,其中前 2 行图像来自 I-Haze 数据集。从图 6 可以看出:DCP 恢复出的图像依旧存在着较明显的颜色失真问题;AOD 和 GDN 恢复出的图像存在着明显的雾霾残留;GCA 和本文提出的方法对雾霾去除得更为彻底,相比于 GCA,本文方法恢复出的图像色彩更接近于真实情况。

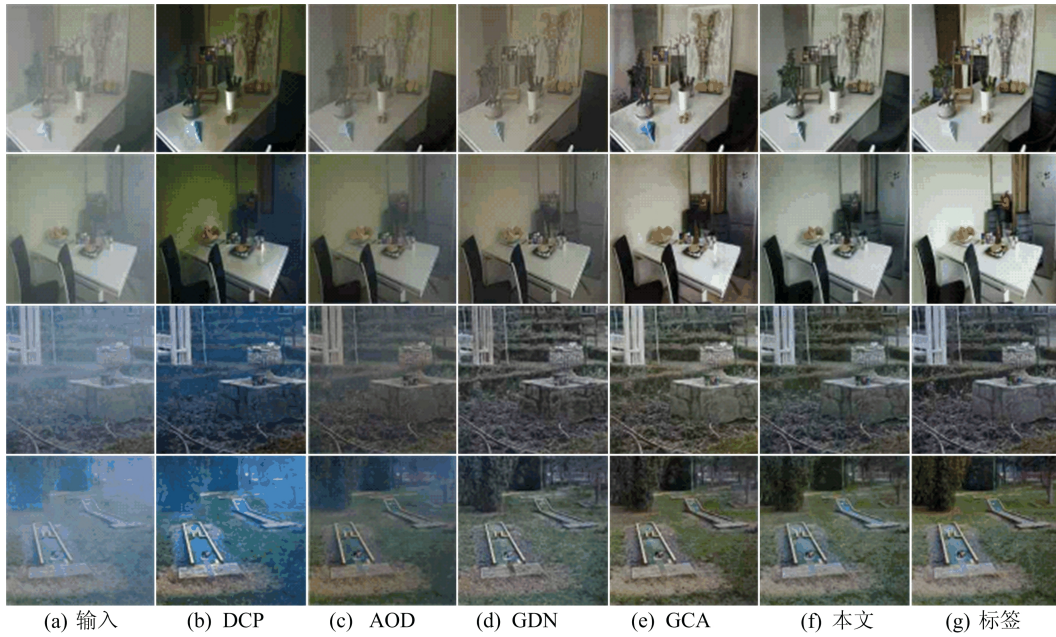


图 6 不同算法在 I-Haze 和 O-Haze 数据集上的定性比较

3.4 消融实验

为了验证方法的有效性,本文研究了 MSSK 网络中低通分支和高通分支的贡献,分别采用将两者从网络中去除的方法设计实验。各实验均使用 ITS 进行训练,在 SOTS 室内部分进行评估,实验结果见表 2 所列。

表 2 分支消融实验结果

实验号	低通分支	高通分支	SSIM	PSNR
1			0.975 3	31.46
2	✓		0.985 9	33.14
3		✓(高通)	0.983 8	32.32
4		✓(雾图)	0.986 5	33.51
5	✓	✓	0.988 2	34.40

通过对比各项实验结果可知,相比于不使用或仅使用 1 个高通分支或低通分支,同时采用高通分支和低通分支能达到更高的结构相似度和峰值信噪比。此外,若仅使用 1 个低通分支或高通

分支也比不使用任何分支效果更优。

实验 3 和实验 4 探究了高通分支的不同输入对结果的影响。从表 2 可以看出,将输入从有雾图像的高通滤波图像替换为有雾图像后,算法性能有了明显的提升。这说明有雾图像的高通滤波图像受雾霾影响时,因为边缘、细节等信息缺失较多,所以难以提取到足够的特征指导去雾,而通过将有雾图像作为高通分支输入可解决该问题。

4 结 论

本文提出了一种新的可端到端训练的图像去雾网络,其多分支结构能将图像低频和高频信息纳入模型训练过程。通过分支指导模块有效地进行分支间信息交互,利用残差选择块所具有的感受野调节能力使网络能更好地感知雾霾质量浓度变化。实验结果表明,本文所提方法能有效解决雾霾残留、颜色失真、伪影等问题,具有优良的性能。

[参 考 文 献]

- [1] 万隆. 基于颜色描述的快速视频去雾算法[J]. 电光与控制, 2021, 28(5): 36-41.
- [2] TU Z, CHEN X, YUILLE A L, et al. Image parsing: unifying segmentation, detection, and recognition[J]. *International Journal of Computer Vision*, 2005, 63: 113-140.
- [3] 王万国, 王滨海, 张晶晶, 等. 基于直方图规定化的图像去雾算法[J]. 计算机技术与发展, 2014, 24(9): 241-244.
- [4] NARASIMHAN S G, NAYAR S K. Vision and the atmosphere[J]. *International Journal of Computer Vision*, 2002, 48(3): 233-254.
- [5] LI X, WANG W, HU X, et al. Selective kernel networks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2019: 510-519.
- [6] HE K, SUN J, TANG X. Single image haze removal using dark channel prior[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 33(12): 2341-2353.
- [7] ZHU Q, MAI J, SHAO L. A fast single image haze removal algorithm using color attenuation prior[J]. *IEEE Transactions on Image Processing*, 2015, 24(11): 3522-3533.
- [8] BERMAN D, AVIDAN S. Non-local image dehazing[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2016: 1674-1682.
- [9] 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.
- [10] LI B, PENG X, WANG Z, et al. Aod-net: all-in-one dehazing network[C]//Proceedings of the IEEE International Conference on Computer Vision. [S. l.]: IEEE, 2017: 4770-4778.
- [11] YEH C H, HUANG C H, KANG L W. Multi-scale deep residual learning-based single image haze removal via image decomposition[J]. *IEEE Transactions on Image Processing*, 2019, 29: 3153-3167.
- [12] LIU X, MA Y, SHI Z, et al. GridDehazeNet: attention-based multi-scale network for image dehazing[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S. l.]: IEEE, 2019: 7314-7323.
- [13] 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. [S. l.]: AAAI, 2020: 11908-11915.
- [14] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2018: 7132-7141.
- [15] PETTET M W, GILBERT C D. Dynamic changes in receptive-field size in cat primary visual cortex[J]. *Proceedings of the National Academy of Sciences*, 1992, 89(17): 8366-8370.
- [16] SPILLMANN L, DRESP-LANGLEY B, TSENG C H. Beyond the classical receptive field: the effect of contextual stimuli[J]. *Journal of Vision*, 2015, 15(9): 1-23.
- [17] LI B, REN W, FU D, et al. Benchmarking single-image dehazing and beyond[J]. *IEEE Transactions on Image Processing*, 2018, 28(1): 492-505.
- [18] ANCUTI C, ANCUTI C O, TIMOFTE R, et al. I-HAZE: a dehazing benchmark with real hazy and haze-free indoor images[C]//Proceedings of International Conference on Advanced Concepts for Intelligent Vision Systems. [S. l.]: Springer, 2018: 620-631.
- [19] ANCUTI C O, ANCUTI C, TIMOFTE R, et al. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. [S. l.]: IEEE, 2018: 754-762.
- [20] HE T, ZHANG Z, ZHANG H, et al. Bag of tricks for image classification with convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2019: 558-567.
- [21] CHEN D, HE M, FAN Q, et al. Gated context aggregation network for image dehazing and deraining[C]//2019 IEEE Winter Conference on Applications of Computer Vision. [S. l.]: IEEE, 2019: 1375-1383.

(责任编辑 李 凯)