基于多尺度特征融合注意力的半监督图像去雾

闫在爽 [1]
YAN Zaishuang

摘 要 在图像去雾任务中，有监督方法由于依赖大量合成配对图像，常面临泛化能力差和过拟合的问题；而无监督方法由于缺乏有效约束，往往去雾效果不理想，且容易导致图像纹理信息的丢失。为解决以上问题，文章提出一种基于多尺度特征融合注意力的半监督图像去雾网络。首先，通过半监督学习策略，将合成图像与真实图像一同用于网络训练，从而提高模型的泛化能力并增强去雾效果。其次，引入多尺度特征融合注意力模块，通过融合不同尺度的注意力，整合多层次特征信息，从而增强模型捕捉图像细节与全局信息的能力，改善恢复图像的细节和纹理质量。在公开的合成和真实有雾数据集上的相关实验表明，该算法能够显著提升去雾性能。

关键词 图像去雾；多尺度；注意力机制；特征融合；半监督学习

doi：10.3969/j.issn.1672-9528.2024.12.009

0 引言

雾霾会显著降低图像质量，导致细节模糊和颜色偏移，对自动驾驶、监控系统、遥感成像等实际应用构成挑战。因此，图像去雾是计算机视觉中的重要任务，旨在从雾化图像中恢复出清晰图像。目前，图像去雾方法主要分为两类：基于先验的方法和基于深度学习的方法。

基于先验的方法通过估计大气光值和传输图等参数，并结合大气散射模型来恢复无雾图像。常见的先验去雾方法包括暗通道先验（dark channel prior，DCP）[1] 和颜色衰减先验（color attenuation prior，CAP）[2] 等。虽然这些算法在部分场景下表现良好，但也存在局限性。参数估计困难，导致去雾效果欠佳，影响图像清晰度和细节恢复。此外，这些算法在不同场景中的鲁棒性较差，去雾效果不稳定。

基于深度学习的图像去雾方法分为有监督学习和无监督学习两种。有监督方法依赖于大量的合成配对图像，通过学习有雾图像与清晰图像之间的映射关系来直接恢复清晰图像。Cai 等人 [3] 提出的 DehazeNet 是首个基于深度学习的去雾模型，利用卷积神经网络（convolutional neural network，CNN）估计传输图，并根据大气散射模型恢复图像；Li 等人 [4] 提出的 AOD-Net 将传输图与大气光值合并为一个中间变量 K，再通过 CNN 生成清晰图像；Xu 等人 [5] 提出的 FFA-Net 结合了通道注意力（channel attention，CA）和像素注意力（pixel attention，PA）用于图像去雾，并在合成数据集上表现出色。然而，CA 将特征信息压缩到通道描述符中，对各通道进行加权处理，可能导致部分通道特征被削弱或忽略，继而影响 PA 的像素级处理效果，造成细节信息逐层丢失。此外，CA 和 PA 的串行顺序处理方式将全局特征增强与局部特征处理分开，未能有效整合全局与局部信息。

有监督的图像去雾方法依赖于大量合成配对图像，因此常面临泛化能力差和过拟合的问题。相比之下，无监督方法可通过非配对图像进行训练，避免了这些问题。Zhu 等人 [6] 提出了循环生成对抗网络（cycle-consistent generative adversarial network，CycleGAN），该模型能够利用非配对图像实现图像转译；Engin 等人 [7] 将 CycleGAN 迁移应用到图像去雾任务中，提出了 Cycle-Dehaze 端到端去雾网络，将图像去雾转换为图像转译问题。然而，由于无监督去雾算法缺乏有效约束，通常去雾效果欠佳，且容易导致图像细节丢失与色彩失真。

针对以上问题，本文提出一种基于多尺度特征融合注意力的半监督图像去雾网络，结合有监督与无监督学习优势，充分利用合成和真实图像进行训练，提升模型的泛化能力和去雾性能。同时，在网络的部分层级之间引入了多尺度特征融合注意力模块（multi-scale feature fusion attention，MSFFA），以融合不同尺度的特征信息，从而更全面地恢复图像的细节和颜色信息。

1 本文方法

1.1 网络框架

本文提出的基于多尺度特征融合注意力的半监督图像去雾网络整体框架如图 1 所示。该网络主要包含两个分支：有

1. 三峡大学计算机与信息学院 湖北宜昌 443002

监督分支和无监督分支。有监督分支利用成对的合成图像进行训练，以学习从有雾到无雾图像的直接映射关系，并逐步优化去雾效果。该分支通过均方误差（mean squared error，MSE）损失最小化去雾图像与真实清晰图像之间的误差，从而有效捕捉图像中的细节信息，确保去雾结果的准确性和高质量。无监督分支则利用真实有雾图像进行训练，无需清晰图像作为参考，通过总变分损失[8]和暗通道损失等来优化生成图像，从而增强模型在不同场景下的适应性和鲁棒性。

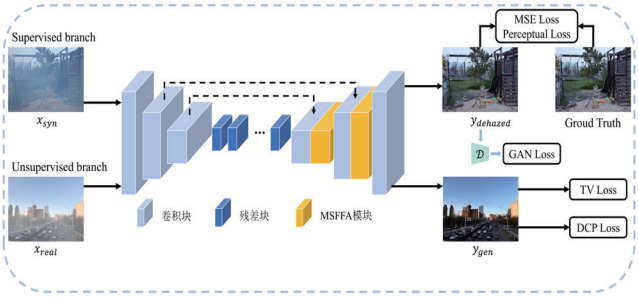

图 1 网络整体架构

去雾主网络由编码器、解码器和 MSFFA 模块组成。在编码器部分，有雾图像首先通过一个步长为 1、卷积核大小为 7 的 3×3 卷积块，将输入图像的通道数从 3 维提升至 64 维，以实现初步特征提取。随后通过 3 个连续的下采样卷积块，逐渐将图像特征图的尺寸缩小至原来的 1/8。在下采样过程中，编码器的输出会通过跳跃连接传递给解码器的相应层进行特征融合，以保留高分辨率的细节信息。在网络的中间层，采用 9 个残差块进行特征转换，帮助学习雾化与清晰图像的特征映射，并有效防止梯度消失，增强网络的表达能力。解码器通过逐步上采样并融合编码器传递的高分辨率特征，以恢复图像的细节和清晰度。同时，解码器中还集成了 MSFFA 模块，该模块通过多尺度卷积和充分利用图像的通道及空间信息，能够更精确地处理图像中的关键细节，从而进一步改善去雾效果。

1.2 多尺度特征融合注意力

通道信息在去雾任务中的重要性已经被 He 等人证实。Cai 等人通过分析雾在图像中的不均匀分布，并采用通道注意力和像素注意力来提升去雾效果。然而，传统的通道注意力机制主要关注通道间的全局信息，通过生成注意力权重对输入通道进行加权。这种方法虽然可以突出关键特征，但在处理复杂场景和细微差异的图像时，仍可能忽略一些重要细节。同时，CA 和 PA 的串行顺序处理方式将全局特征增强与局部特征处理分开，未能有效整合全局与局部信息。

针对以上问题，本文设计了多尺度特征融合注意力模块，其结构如图 2 所示。MSFFA 模块由混合通道注意力（hybrid channel attention，HCA）、空间注意力（spatial attention，SA）以及多尺度卷积模块组成。

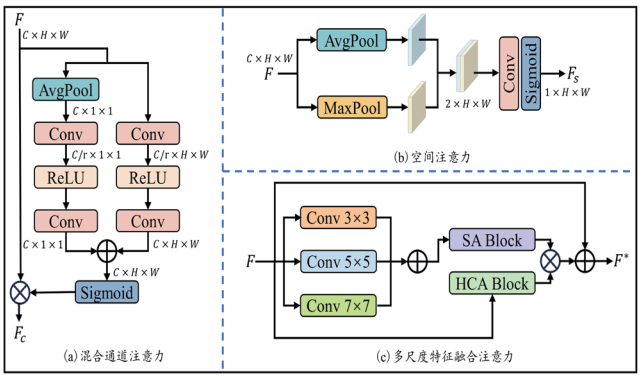

图 2 MSFFA 模块各部分结构

HCA 模块的结构包括两部分：局部通道注意力分支和全局通道注意力分支。对于输入特征 $F \in \mathbb{R}^{C\times H\times W}$，全局通道注意力分支首先通过全局平均池化（global average pooling，GAP）将全局信息转化为通道描述符，然后经过两个卷积层实现通道间的特征变换得到全局通道信息，公式为：

$$F_{\text{global}} = \text{Conv}(\delta(\text{Conv}(\text{GAP}(F)))) \tag{1}$$

式中：δ 表示 ReLU 激活函数。

在局部通道注意力分支中，为保证局部信息的完整性和准确性，未使用 GAP 来处理输入特征，而是直接通过一系列的卷积操作来提取局部通道信息，公式为：

$$F_{\text{local}} = \text{Conv}(\delta(\text{Conv}(F))) \tag{2}$$

随后将局部和全局通道特征信息按主元素相加进行融合，并通过 Sigmoid 激活函数生成混合通道注意力权重公式为：

$$\text{HCA} = \sigma(F_{\text{global}} \oplus F_{\text{local}}) \tag{3}$$

式中：σ 表示 Sigmoid 激活函数。

最终，将输入特征与 HCA 进行逐元素相乘，得到通道增强后的特征公式为：

$$F_c = F \otimes \text{HCA} \tag{4}$$

尽管 HCA 模块有效增强了通道特征的表达能力，但图像中的空间信息同样关键。在去雾任务中，由于雾的空间分布不均匀，单纯依赖通道信息可能导致局部细节的丢失，从而影响整体去雾效果。针对此问题，本文在 MSFFA 模块中引入了多尺度卷积模块和空间注意力机制，以进一步提升图像细节的捕捉能力和去雾效果。

多尺度卷积模块通过采用 3×3、5×5 和 7×7 三种不同尺度的卷积核对输入特征进行并行卷积操作，提取出不同感受野下的特征表示。并将这些特征图进行融合，形成包含多尺度信息的融合特征公式为：

$$F_{\text{fused}} = \text{Conv}_3(F) \oplus \text{Conv}_5(F) \oplus \text{Conv}_7(F) \tag{5}$$

空间注意力机制则进一步强化图像中的空间特征。为计算空间注意力，首先沿着通道轴对输入特征图进行平均池化

和最大池化操作，然后将这两个结果沿通道维度拼接，生成综合特征公式为：

$$F_r = [\text{AvgPool}(F_{\text{fused}}), \text{MaxPool}(F_{\text{fused}})] \tag{6}$$

接下来，F_r 通过卷积核大小为 7 的卷积层，进一步提取空间维度上的特征信息，并使用 Sigmoid 激活函数对输出进行非线性变换，生成单通道的空间注意力图公式为：

$$M_s = \sigma(\text{Conv}(F_r)) \tag{7}$$

最后，将 HCA 模块的输出结果与空间注意力图 M_s 相结合，得到 MSFFA 模块的最终输出结果为：

$$F^* = F \oplus (F_c \otimes M_s) \tag{8}$$

1.3 损失函数

在有监督去雾分支中，采用了均方误差损失、感知损失以及对抗损失。均方误差损失用于衡量去雾结果与清晰图像之间的像素差异，其公式为：

$$L_{\text{MSE}} = \frac{1}{n}\sum_{i=1}^{n}\left\| y^i_{\text{dehazed}} - y^i \right\|_2^2 \tag{9}$$

式中：y_{dehazed} 为去雾图像；y 为对应清晰图像。

感知损失用于提高生成图像的纹理质量。通过预训练的 VGG19 网络提取清晰图像和去雾图像的多层次特征，计算它们的差异以保证在不同特征层面的一致性，其公式为：

$$L_p = \left\| \phi(y_{\text{dehazed}}) - \phi(y) \right\|_2^2 \tag{10}$$

式中：ϕ 代表 VGG19 网络的特征提取层。

为了使生成的清晰图像更加接近真实图像的分布，使用对抗损失对图像生成进行约束，其定义为：

$$L_{\text{adv}} = \mathbb{E}_{y\sim \text{Pdata(y)}}[\log D(y)] + \mathbb{E}_{x\sim \text{Px(x)}}[\log(1 - D(G(x_{\text{syn}})))] \tag{11}$$

式中：D 为判别器；G 为生成器；x_{syn} 为输入的合成有雾图像。

在无监督去雾分支中，采用了总变分损失和暗通道损失。总变分损失用于减少图像的噪声和伪影，增强去雾图像的平滑性，公式为：

$$L_{tv} = \frac{1}{n}\sum_{i=1}^{n}\left(\left\| \partial_h(y^i_{\text{gen}}) \right\|_1 + \left\| \partial_v(y^i_{\text{gen}}) \right\|_1 \right) \tag{12}$$

式中：y_{gen} 为生成的去雾图像；∂_h 和 ∂_v 分别代表水平和垂直方向上的微分运算。

He 等人提出的 DCP 去雾算法基于以下假设：在无雾图像的任意局部窗口内，至少有一个通道的亮度值接近零。其数学公式为：

$$\text{DC}(y_{\text{gen}}) = \min_{c\in\{r,g,b\}}\left(\min_{z\in\Omega(x)} y^c_{\text{gen}}(z)\right) \to 0 \tag{13}$$

式中：$\Omega(x)$ 是用于计算每个像素位置最小值的局部窗口；$y^c_{\text{gen}}(z)$ 为像素 z 在通道 c 上的亮度值。根据这一假设，引入暗通道损失，使去雾图像的暗通道与清晰图像的暗通道一致，以确保生成的去雾图像更接近真实的清晰图像，其定义为：

$$L_{\text{DCP}} = \left\| \text{DC}(y_{\text{gen}}) \right\|_1 \tag{14}$$

综上所述，最终的损失函数可以表示为：

$$L_{\text{total}} = \lambda_1 L_{\text{MSE}} + \lambda_2 L_P + \lambda_3 L_{\text{adv}} + \lambda_4 L_{\text{tv}} + \lambda_5 L_{\text{DCP}} \tag{15}$$

式中：λ_1、λ_2、λ_3、λ_4 和 λ_5 分别为相应损失的权重。

2 实验结果与分析

2.1 数据准备

为了取得更好的训练效果，本文从常用的图像去雾数据集 RESIDE 中随机抽取了合成有雾图像和真实场景的有雾图像，用于去雾网络的训练。具体来说，本文从 RESIDE 数据集中的室外训练集（outdoor training set，OTS）和室内训练集（indoor training set，ITS）中各随机抽取 2000 对配对图像用于有监督分支训练。此外，还从未标注真实有雾图像（unannotated realistic hazy images，URHI）数据集中随机抽取 2000 张真实有雾图像用于无监督分支训练，以构建本文的训练数据集。

在算法测试中，本文采用了 RESIDE 中的合成客观测试集（synthetic objective testing set，SOTS）和真实图像去雾数据集 DENSE-HAZE 进行评估。SOTS 测试集包括 500 张室外和 500 张室内的有雾图像及其对应的无雾清晰图像；而 DENSE-HAZE 数据集则由 55 张真实的有雾图像及其对应的清晰图像组成。

2.2 实验设置

本文基于 PyTorch 框架，在 Ubuntu 环境下利用 Nvidia RTX3090 GPU 进行网络模型训练。采用 Adam 优化器进行优化，动量衰减指数 β_1=0.5，β_2=0.999，初始学习率为 0.000 2。针对不同分辨率的原始图像，将图像随机裁剪成像素大小均为 256 px×256 px 的图像块作为网络的输入，batch size 大小设置为 4。

2.3 对比实验

为评估不同去雾算法的性能，本文使用了图像去雾领域常用的性能指标，包括结构相似性指数（structural similarity index，SSIM）和峰值信噪比（peak signal-to-noise ratio，PSNR）对实验结果进行评价。

本文将所提出的去雾算法与其他五种先进的去雾算法，包括 Cycle-Dehaze、DA Dehaze、RefineDNet[9]、D4[10] 以及 UCL-Dehaze[11]，在 Outdoor、Indoor 和 DENSE-HAZE 数据集上进行实验对比。表 1 展示了各算法的定量实验结果，包括每个数据集上的 SSIM 和 PSNR 指标。其中，第一名用粗体标注；第二名画线标注；第三名斜体标注。实验结果表明：本文提出的算法在各个测试数据集上均表现优异，尤其是在 Outdoor 数据集上，其 SSIM 和 PSNR 分别达到了 0.942 和 26.72 dB，领先于其他方法。在 Indoor 和 DENSE-HAZE 数据集上，本算法同样展现了更好地去雾效果，其中在 DENSE-HAZE 数据集上的表现突出，SSIM 为 0.512，PSNR 为 14.89 dB，证明了本算法在处理浓雾场景时的鲁棒性。

表 1 各算法在 Outdoor、Indoor 和 DENSE-HAZE 数据集上的定量对比

Method	Outdoor		Indoor		DENSE-HAZE		Average	
	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB
Cycle-Dehaze	0.899	21.31	0.854	20.11	0.357	13.30	0.703	18.24
DA Dehaze	0.930	26.38	**0.939**	**25.72**	0.436	13.98	0.768	22.03
RefineDNet	0.886	20.67	0.866	20.48	0.391	*13.79*	0.714	18.31
D4	*0.927*	*25.53*	*0.917*	*25.19*	0.384	13.44	*0.743*	*21.39*
UCL-Dehaze	0.895	24.87	0.903	24.26	*0.423*	13.66	0.740	20.93
Ours	**0.942**	**26.72**	0.926	25.37	**0.512**	**14.89**	**0.793**	**22.33**

总体而言，本文算法在多个数据集上的平均 SSIM 和 PSNR 均优于现有主流方法，验证了其在图像去雾任务中的有效性。各个去雾算法在 Outdoor 数据集上的可视化结果如图 3 所示，相比于其他五种算法，本文提出的去雾算法在饱和度、清晰度和亮度方面最接近目标图像，尽管存在微小差异，但总体去雾效果更为出色，能更好地保留色彩和细节，使图像更自然和真实。

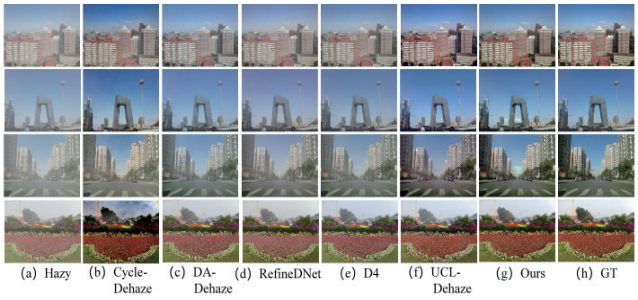

图 3 各算法在 Outdoor 数据集上的定性对比

3 结论

本文提出的基于多尺度特征融合注意力的半监督图像去雾算法，通过结合半监督学习策略，有效利用合成图像与真实图像进行网络训练，从而提高模型的泛化能力和去雾效果。此外，引入的多尺度特征融合注意力整合了不同尺度的特征信息，增强对图像细节和全局信息的捕捉，改善恢复图像的细节和纹理质量。实验结果表明，本文算法在合成和真实有雾数据集上的性能和视觉效果均优于现有的五种主流去雾算法。展望未来，研究工作将聚焦于去雾模型的轻量化设计以及实时性的优化提升，旨在进一步拓展该算法于实际应用场景中的适用性与高效性。

参考文献：

[1] 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. Piscataway:IEEE, 2009: 1956-1963.

[2] 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, 2015, 24(11): 3522-3533.

[3] 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.

[4] 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: 4770-4778.

[5] XU Q, WANG Z L, BAI Y C, et al. FFA-Net: feature fusion attention network for single image dehazing[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2020, 34(7): 11908-11915.

[6] ZHU J Y, PARK T, ISOLAI P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway:IEEE,2017:2223-2232.

[7] ENGIN D, GENC A, EKENEL H K. Cycle-dehaze: enhanced cycleGAN for single image dehazing[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.Piscataway:IEEE, 2018: 825-833.

[8] SHAO Y J, REN W Q, GAO C X, et al. Domain adaptation for image dehazing[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 2808-2817.

[9] ZHAO S Y, ZHANG L, SHEN Y, et al. RefineDNet: a weakly supervised refinement framework for single image dehazing[J]. IEEE transactions on image processing, 2021, 30: 3391-3404.

[10] YANG Y, WANG C Y, LIU R S, et al. Self-augmented unpaired image dehazing via density and depth decomposition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2022: 2037-2046.

[11] WANG Y Z, YAN X F, WANG F L, et al. UCL-Dehaze: Towards real-world image dehazing via unsupervised contrastive learning[J]. IEEE transactions on image processing, 2024, 33: 1361-1374.

【作者简介】

闫在爽（1997—），男，湖北襄阳人，硕士研究生，研究方向：图像处理、深度学习。

（收稿日期：2024-09-04）