

一种基于无监督学习的两阶段解耦图像去雾算法研究

林 盛, 黎 敏, 胡 杰, 赵 丽*

温州大学计算机与人工智能学院, 浙江省安全应急智能信息技术重点实验室, 浙江 温州

收稿日期: 2024年3月15日; 录用日期: 2024年4月18日; 发布日期: 2024年4月19日

摘 要

利用卷积神经网络进行监督学习是解决图像去雾问题的一种常用的解决方法。然而, 现有的方法大多主要使用成对的合成雾霾数据集, 这可能不能准确地代表真实雾霾天气的场景。针对这一问题, 本文提出了一种基于无监督学习的两阶段解耦去雾网络。该网络由三个子网络组成, 它们分别将观测到的雾霾图像分解为无雾霾图像层、透射图层和大气光层。同时, 该网络分为恢复无雾图像和透射图两个阶段。在第一阶段, 利用嵌入的暗通道先验来获得无雾霾像和透射图的粗略估计。在第二阶段, 通过两个子网络对第一阶段的结果进行细化, 以产生更精确的无雾霾图像和透射图, 而大气光则由另一个子网络直接估计。此外, 本文还设计了一种新的多尺度注意力模块, 作为细化无雾图像的子网。多尺度注意力模块在自注意力中执行多尺度的标记聚合, 以捕获不同尺度的特征。实验结果表明, 本文提出的网络获得了有效的雾性能和令人满意的视觉效果, 且在PSNR、SSIM和主观视觉效果方面均优于现有的无监督去雾方法。

关键词

图像去雾, 注意力机制, Transformer, 无监督学习

A Two-Stage Disentanglement Image Dehazing Algorithm Based on Unsupervised Learning

Sheng Lin, Min Li, Jie Hu, Li Zhao*

Zhejiang Province Key Laboratory of Security Emergency Intelligent Information Technology, School of Computer and Artificial Intelligence, Wenzhou University, Wenzhou Zhejiang

Received: Mar. 15th, 2024; accepted: Apr. 18th, 2024; published: Apr. 19th, 2024

*通讯作者。

文章引用: 林盛, 黎敏, 胡杰, 赵丽. 一种基于无监督学习的两阶段解耦图像去雾算法研究[J]. 计算机科学与应用, 2024, 14(4): 163-176. DOI: 10.12677/csa.2024.144087

Abstract

The use of convolutional neural networks for supervised learning is a commonly used method to solve the problem of image dehazing. However, most existing methods mainly use paired synthetic haze datasets, which may not accurately represent real-world haze weather scenarios. This article proposes a two-stage decoupling dehazing network based on unsupervised learning to address this issue. This network consists of three sub networks, which respectively decompose the observed haze images into a haze free image layer, a transmission layer, and an atmospheric light layer. Meanwhile, the network is divided into two stages: restoring fog free images and transmitting images. In the first stage, rough estimates of haze free images and transmission images are obtained using embedded dark channel priors. In the second stage, the results of the first stage are refined through two sub networks to produce more accurate haze free images and transmission maps, while atmospheric light is directly estimated by another sub network. In addition, this article also designs a new multi-scale attention module as a subnet for refining haze free images. The multi-scale attention module performs multi-scale label aggregation in self attention to capture features of different scales. The experimental results show that the proposed network achieves effective defogging performance and satisfactory visual effects, and is superior to existing unsupervised defogging methods in terms of PSNR, SSIM, and subjective visual effects.

Keywords

Image Dehazing, Attention Mechanism, Transformer, Unsupervised Learning

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

雾霾是一种当大量颗粒物聚集并悬浮在空气中时出现的现象。当光穿过大气层时，它会被这些粒子吸收和散射，导致各种设备捕获的图像的清晰度和对比度降低。这种图像质量的下降给后续的图像处理操作带来了巨大的挑战。在严重的情况下，它甚至会扭曲颜色信息，使计算机视觉技术的应用具有挑战性。因此，图像去雾已成为计算机视觉领域的研究热点[1] [2]。在图像去雾任务中，雾霾退化过程采用大气散射模型[3] [4] [5]进行建模，可以表示为：

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

其中 I 为雾霾图像， J 为潜在的无雾图像， A 为全局大气光， t 为透射图。而透射图可以表示为：

$$t(x) = e^{-\beta d(x)} \quad (2)$$

其中 β 是大气的散射系数， d 是场景深度。可以看出，从一个雾霾图像中估计一个干净的图像是一个困难的不适定问题。目前，已经提出了许多图像去雾方法，大致可分为基于先验的方法和基于学习的方法两类。基于先验的方法通常依赖于图像中固有的先验知识来进行去雾。例如，He 等人[6]引入了暗通道先验(Dark Channel Prior, DCP)来识别图像中雾霾的分布；Zhu 等人[7]提出了颜色衰减先验(Color Attenuation Prior, CAP)来估计透射图。虽然这些基于先验的方法已经取得了一些结果，但所使用的先验知识往往容易受到实际干扰，而且在复杂的现实场景中鲁棒性较差。

近年来,随着深度学习在各种计算机视觉和图像处理任务中的成功应用,基于学习的方法在图像去雾处理领域得到了广泛的应用。这些基于学习的方法允许网络以数据驱动的方式学习透射图和大气光。例如,Cai 等人[8]提出了 DehazeNet,一个在大数据集上训练来估计等式(1)中的 $t(x)$ 的网络,同时借助真实雾霾参考图像的透射图。Zhang 等人[9]提出了密集连接的金字塔去雾网络(Densely Connected Pyramid Dehazing Network, DCPDN),它将透射图、大气光和去雾网络相结合,以进行总体学习。此外,其他一些工作[10]并不依靠网络来估计透射图,而是直接利用网络来恢复无雾霾图像。

许多研究表明,基于学习的方法可以通过利用大型数据集有效地学习复杂的输入输出之间的映射,而且它们通常具有优于基于先验的方法的性能。然而,这些基于学习的去雾方法有其局限性,需要进一步的细化和增强。一方面,目前大多数基于学习的去雾方法[11][12][13]都是数据驱动的,通常需要图像对以完全监督的方式训练网络,但很难获得大量的标记数据。另一方面,基于大气散射模型[8][9][14]学习了许多方法,其中输入的雾霾图像是三个因素的组合:透射图、全局大气图和潜在的无雾霾图像。这表明,网络从输入的雾霾图像中学习到的特征在大气散射物理模型中包含了这三个因素。然而,许多现有的方法只预测一个因素或平等地预测这三个因素,导致图像去雾性能的提升有限。

为了解决这些问题,本文提出了一种基于无监督学习的两阶段解耦去雾网络。与其他基于学习的方法不同,该网络只需要一个单一的雾霾图像进行训练,不需要任何配对或未配对的图像,这减轻了模型对成对数据的依赖,大大降低了训练成本。同时,本文提出的网络模型利用三个子网络,根据大气散射的物理模型,将观测到的雾霾图像分解为三个潜在层,然后同时预测这三个潜在层。值得注意的是,该网络专注于预测无雾霾图像和透射图。如等式(1)所示,雾霾图像的信息与底层的无雾霾图像和透射图相关,而全局大气光图不包含图像内容。因此,本文提出的网络包括两个阶段。在第一阶段,利用嵌入的暗通道先验来获得无雾霾图像和透射图的粗略估计。在第二阶段,通过两个子网络对第一阶段的结果进行细化,以产生更精确的无雾图像和透射图,而另一个子网络直接估算大气光。这种对网络中三个潜在层的两阶段估计集成了基于先验的方法和基于学习的方法的优势,使网络能够恢复更真实的图像。

为了获得更精确的无雾图像,本文提出了一种新的多尺度注意力模块,利用了 Transformer 在图像去雾[11]和其他领域[15]中的优异性能。在多尺度注意力模块的自注意力中,执行多尺度标记聚合来获取不同大小的键和值,使其能够捕获不同尺度的特征。

本文工作的贡献总结如下:

- 本文提出了一种基于无监督学习的两阶段解耦去雾网络,旨在准确估计大气散射物理模型中的无雾图像和透射图。
- 本文提出了一种新的多尺度注意力模块,并将其应用于子网络,以获得精细的无雾霾图像。多尺度注意力模块中的自注意力采用多尺度令牌聚合来捕获不同尺度的特征,增强多个尺度上图像雾霾的清除,从而细化无雾霾图像。
- 大量实验表明,本文提出的网络在各种雾霾数据集上取得了优异的去雾性能,超过了以往的无监督去雾方法。

2. 相关工作

2.1. 图像去雾

2.1.1. 基于先验的方法

基于先验的方法,也称为传统的去雾方法,依赖于大气散射模型,利用图像的先验知识去除雾霾。例如,He 等人[6]提出了 DCP 来获取图像的雾霾分布。该方法利用图像无雾区域内暗通道的统计规律性来估计透射图和大气光。同样,Zhu 等人[7]引入了 CAP 来估计透射图,它利用了图像中不同区域的雾霾

的颜色一致性。根据雾霾对图像有颜色相关的衰减效应, Fattal [16]提出了一种使用颜色线模型来估计透射图和大气光的技术。Berman 等人[17]提供了一种基于局部先验信息的图像去雾方法。基于先验的方法虽然取得了一定的进展, 但也有一定的局限性。例如, 它们所依赖的先验知识在实际场景中很容易被推翻, 这限制了它们在复杂的现实场景中的鲁棒性。

2.1.2. 基于学习的方法

随着深度学习在各种计算机视觉任务中的成功应用, 基于学习的去雾方法逐渐取代了基于先验的去雾方法, 并在图像去雾领域占据主导地位。早期的方法试图将卷积神经网络(Convolutional Neural Networks, CNNs)应用于图像去雾, 以直接从输入的雾霾图像中估计透射图和大气光。例如, Cai 等人[8]首次将卷积神经网络引入图像去雾任务。该方法使用卷积神经网络来估计透射图和大气光, 并且在图像去雾性能方面比传统的基于先验的方法有了显著的改进。Ren 等人[14]提出了一种多尺度卷积神经网络(Multi-scale Convolutional Neural Network, MSCNN), 该网络通过估计场景透射图来恢复潜在的干净图像。Li 等人[10]提出了一种深度卷积神经网络, 称为 AOD-Net, 它以雾霾图像为输入; AOD-Net 不是像其他方法那样估计透射图和大气光, 而是通过网络直接生成干净的图像。这些方法已经取得了令人印象深刻的结果, 并展示了深度学习在图像去雾方面的潜力。

后来, 这一领域的研究转向了从单一的雾霾图像中直接恢复潜在的干净图像, 而不考虑大气散射物理模型[12] [13] [18] [19] [20] [21] [22]。例如, Li 等人[19]提出了一种基于条件生成对抗网络框架的图像去雾霾网络, 他们集成了 VGG 特征和 L1 正则化梯度先验, 直接从输入的雾霾图像中生成无雾图像。Engin 等人[20]提出了 Cycle-Dehaze, 该网络基于循环生成对抗网络, 结合周期一致性损失和感知损失, 训练网络, 产生具有更好视觉效果的去雾图像。Chen 等人[18]开发了一种端到端门控上下文聚合网络, 该网络利用平滑扩张技术来消除由扩张卷积引起的网格伪影, 并利用不同级别特征的门控子网络融合。Qu 等人[21]提出了一种图像去雾霾网络, 该网络在条件生成对抗网络框架中加入了一个增强模块, 以增强在颜色和细节上的去雾霾效果。Qin 等人[12]提出了一种端到端特征融合注意网络, 该网络使用通道和像素注意机制来处理分布在不同图像像素上的非均匀雾霾特征。此外, Wu 等人[23]将对比学习引入图像去雾, 提出了基于对比学习的对比正则化, 并开发了一种基于类自编码器框架的紧凑去雾网络 AEER-Net。

近年来, 视觉 Transformer (Vision Transformer, ViT) [24]在高级计算机视觉任务中的显著性能挑战了 CNN 的主导地位, 并显示出了取代它们的巨大潜力。这促使了基于 ViT 的去雾网络的提出。例如, Zhao 等人[25]将一个混合的局部 - 全局视觉 Transformer (HyLoG-ViT)集成到一个去雾网络中, 以捕获全局和远程依赖关系。从 SwinTransformer [26]开始, Song 等人[11]提出了一种名为 DehazeFormer 的 ViT 架构, 并将其适应于去雾任务。Guo 等人[27]提出了 DeHamer, 该方法在 Transformer 中引入了一种新的传输感知三维位置嵌入模块, 并结合了与雾霾密度相关的先验。然而, 这些基于学习的方法依赖于图像对进行监督, 而本文提出的方法是一种无监督方法, 只需要单个输入图像。同时, 与另一种基于 Transformer 的归一化层、激活函数和空间信息聚合方案的[11]相比, 本文的设计主要围绕自注意机制。

2.2. 无监督学习

虽然深度学习方法在图像去雾霾方面取得了重大进展, 但这些方法通常需要大量的监督训练数据, 并且对真实世界的雾霾图像的泛化能力有限。因此, 人们对开发无监督方法[28] [29] [30]越来越感兴趣, 该方法可以在不依赖成对训练数据的情况下执行图像去雾。例如, 甘德尔斯曼等人[29]提出了一个基于无监督图像先验[28]的单图像无监督统一框架, 可用于执行包括图像去雾霾处理在内的计算机视觉任务。Li 等人[30]提出了一种专门设计的无监督方法 YOLY。Li 等人[31]提出的 ZID 是一种不需要合成雾霾图像的零样本学习方法, 用于图像去雾, 可以避免在处理真实图像时出现的域转移问题。同样, ZIR [32]

是一种基于科施米德模型的零样本图像恢复方法，适用于图像去雾霾、水下图像恢复等场景。然而，ZIR并不是专门为图像去雾任务而设计的。此外，虽然 ZID 可以通过零样本学习实现图像去雾，但它也存在一些缺陷，如能够平等地预测大气模型中的三个因素，从而限制了去雾性能的提高。

3. 方法

3.1. 两阶段去雾

如图 1 所示，输入到网络中的雾霾图像 x 将经过两个阶段的处理。第一阶段是粗略估计阶段，第二个阶段是细化阶段。本节详细介绍了网络的两个阶段及其具体操作。

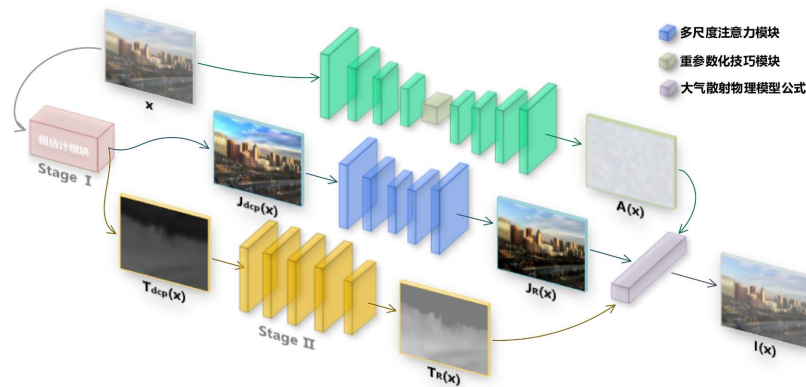


Figure 1. The network structure of the proposed dehazing algorithm

图 1. 所提出的去雾网络结构

3.1.1. 粗略估计阶段

在粗略估计阶段，利用 DCP 分别依次估计透射图和无雾霾图像的粗略估计图像 $T_{dcp}(x)$ 和 $J_{dcp}(x)$ 。给定一个输入的三通道彩色图像 $I(x)$ ，首先找到三个 RGB 通道的最小值，生成一个灰度图像。然后，将此灰度图像倒置，并进行最大池化操作。得到的图像记为 I^{dark} ，可以通过以下公式得到：

$$I^{dark}(x) = -\max\text{pool}\left(-\min_{c \in R, G, B} (I^c(x))\right) \quad (3)$$

其中 I^c 是指 I 的 R、G、B 通道之一。将 I^{dark} 代入大气散射物理模型得到：

$$I^{dark}(x) = J^{dark}(x)t(x) + A(1-t(x)) \quad (4)$$

其中 $I^{dark}(x)$ 和 $J^{dark}(x)$ 分别是图像 I 和 J 在像素 x 处的暗通道。由于天空的颜色通常类似于雾霾图像中呈现的大气光 A ，本文选择了 I^{dark} 中前 0.1% 的像素，并计算它们的平均值得到 A 。然而，这个对大气光 A 的估计仅用于 $T_{dcp}(x)$ 和 $J_{dcp}(x)$ 的推断。为了获得一个全面的大气光图，本文进一步利用了 A-Net，这是一个大气光估计网络。对于 J^{dark} ，基于 DCP 的假设，室外无雾图像 J 的非天空区域的暗通道通常接近于零，从而满足以下条件：

$$J^{dark} \rightarrow 0 \quad (5)$$

结合等式(4)，可以得到 T_{dcp} 的初步结果如下：

$$t(x) = 1 - \frac{I^{dark}(x)}{A} \quad (6)$$

为了解决暗通道图像中不平滑边缘过渡的问题，采用引导滤波来提高 $T_{dcp}(x)$ 边缘的平滑性。具体来

说, 内核大小为 19×19 、步幅为 1 的平均池化用于实现引导滤波。根据等式(1), 对于已知大气光 A 和透射图 $t(x)$ 、 $J_{dep}(x)$ 可表示为:

$$J(x) = \frac{I(x) - A}{t(x)} + A \quad (7)$$

3.1.2. 细化阶段

在细化阶段, 使用三个子网络, 包括无雾图像细化网络(JR-Net)、透射图细化网络(TR-Net)和大气光估计网络(A-Net), 对三个不同的潜在层进行预测。

- JR-Net 是对 U-Net 架构的改进, 它由五个层组成。在该网络中, 将卷积块替换为多尺度注意力模块, 并使用 SK 融合模块[11]来连接和融合不同的层。为了获得全局残差, 该网络使用了软重构层。在网络的头部和尾部有一个 3×3 卷积层, 多尺度注意力模块的每一层都进行降采样或上采样操作。具体来说, 多尺度注意力模块的前两个级别之后是降采样操作, 而第三个和第四个多尺度注意力模块之后是上采样操作。

- 虽然透射图是本文的雾霾去除算法的关键组成部分, 但本文并没有使用所提出的多尺度注意力模块来评估透射图。而是采用[19]的非退化结构来设计 TR-Net。具体来说, TR-Net 由四个层组成, 每一层都包含卷积层、批处理归一化层[33]和 LeakyReLU 激活函数[34]。最后一层包括卷积层和一个 sigmoid 函数, 它将输出归一化到[0,1]的范围。

- 采用对称编码解码器结构估计大气光, 产生的大气光图像用于雾霾图像重建。网络的编码器和解码器由四层组成, 其中编码器包括卷积层、ReLU 激活函数[35]和每一层的最大池化层, 解码器包括上采样层、卷积层、批归一化层和每一层的 ReLU 激活函数。此外, 在编解码器中间使用重参数化技巧模块[36]模块将编码器的输出转换为潜在的高斯分布, 然后重新采样后输入解码器。

3.2. 多尺度注意力模块

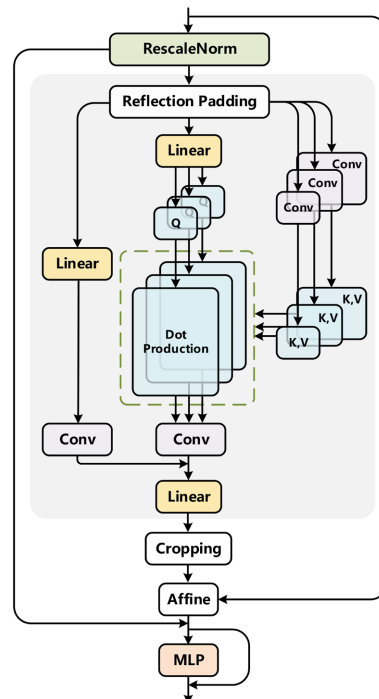


Figure 2. The structure of the multi scale attention block
图 2. 多尺度注意力块的结构

所提出的多尺度注意力模块基于 Transformer 编码器架构, 具体的结构如图 2 所示, 其中主要包括一个 RescaleNorm 层、一个多尺度自注意力(Multi Scale Self-attention, MSSA)和一个多层感知器层。与现有的使用 Transformer 的去雾方法[11]相比, 多尺度注意力模块通过自注意中的多尺度令牌聚合来捕获特征, 这种独特的设计使多尺度注意力模块能够捕获不同尺度的特性, 提高了其处理多尺度信息的能力。相比之下, 之前的方法[11]倾向于主要在单一尺度上提供令牌特征, 这导致捕获和处理多尺度特征的能力有限。下面详细介绍了 RescaleNorm 层和 MSSA。

3.2.1. RescaleNorm 层

归一化层是神经网络的一个重要组成部分, 因为它可以极大地提高神经网络的收敛速度、稳定性和泛化能力。为了提高网络的性能, 本文采用了更适合的 RescaleNorm 层[11], 它适用于图像去雾任务。RescaleNorm 的具体归一化过程可以表示为:

$$\hat{x} = F\left(\frac{x - \mu}{\sigma} \cdot \gamma + \beta\right) \cdot (\sigma W_\gamma + B_\gamma) + (\mu W_\beta + B_\beta) \quad (8)$$

其中 $\mu, \sigma \in \mathbb{R}^{b \times 1 \times 1}$ 分别表示均值和标准差, $\gamma, \beta \in \mathbb{R}^{b \times 1 \times c}$ 分别是学习到的比例因子和偏差, $W_\gamma, W_\beta \in \mathbb{R}^{1 \times c}$ 和 $B_\gamma, B_\beta \in \mathbb{R}^{1 \times 1 \times c}$ 分别为两个用于转换 μ 和 σ 的线性层权重和偏置项, 转换过程表示为 $\{\gamma, \beta\} = \{\sigma W_\gamma + B_\gamma, \mu W_\beta + B_\beta\}$ 。为了加速收敛, B_γ 和 B_β 初始化为 1 和 0。

3.2.2. 多尺度自注意力

与传统的注意力机制不同, 本文提出了多尺度自注意力机制来聚合窗口内的多尺度特征信息, 同时一定程度上保留邻域的信息。具体来说, 本文提出的多尺度自注意力具有两个分支。在主要分支中对 Q 、 K 、 V 进行不同的处理。对于 Q , 保留其原始的尺度, 对其进行普通的全连接操作。而对于键 K , 值 V , 在同一自注意力层中, 使用尺度变换(Scale Transformation, ST)来获得不同尺寸的 K 和 V , 具体表示为:

$$\begin{aligned} Q &= XW^Q \\ K_i, V_i &= ST_i(X)W_i^K, ST_i(X)W_i^V \\ V_i &= V_i + LE_i(V_i) \end{aligned} \quad (9)$$

其中, W^Q, W_i^K, W_i^V 是自注意力层的同一头部中第 i 个尺度层的线性投影的参数, $LE_i(\cdot)$ 为深度卷积的局部增强分量。而 $ST_i(\cdot)$ 是第 i 个尺度层的尺度变换操作, 它提供对 X 的降采样, 可具体表示为:

$$ST_i(X) = LN(Convi(X)) \quad (10)$$

其中, $LN(\cdot)$ 是指 Layer Normalization [37]。 $Convi(\cdot)$ 是指第 i 个尺度层所使用的卷积, 不同尺度层所用的内核大小和步幅不同, 从而造就了不同的尺度变换。因此, 在同一自我注意力层中, 键和值可以捕捉到不同尺度的特征。然后, 自注意力头可以计算为:

$$head = \text{Soft max}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (11)$$

其中 d 是维度。将不同的头连接, 通过以下方法计算主要分支中的多头自我注意:

$$MH(Q, K, V) = \text{Concat}(head_i, \dots, head_j)W^O \quad (12)$$

其中 $\text{Concat}(\cdot)$ 来源于工作[15]中的连接操作, W^O 是线性投影参数。而在另一个保留邻域信息的分支中, 仅对输入的特征依次进行线性投影和卷积操作。综上所述, 本文提出的多尺度自注意力可表示为:

$$MSSA = \text{Concat} \left(MH \{Q, K_i, V_i\}_{i=1, \dots, m} \right) + \text{Conv}_0 \left(XW^0 \right) \quad (13)$$

3.3. 损失函数

本文网络的制定是基于从大气散射模型解耦的三个子网预测组合成的雾霾图像和输入的雾霾图像之间的相似性的关系。因此，相应的采用了重建损失 L_{rec} 来测量输入图像与重建后图像的误差，以此来约束整个网络。 L_{rec} 定义为：

$$L_{rec} = \|I(x) - x\|_F \quad (14)$$

其中， $\|\cdot\|_F$ 表示给定矩阵的弗罗比尼乌斯范数范数。 x 是输入的雾霾图像， $I(x)$ 是由三个子网的输出重建成的雾霾图像。除了对网络顶端的输出进行监督外，本文对大气光估计网络和和无雾图像细化网络的输出也进行了监督，设计了相应的损失函数。对无雾图像细化网络，本文根据工作[7]对图像的观测结果，即无雾图像的深度与其亮度和饱和度之差呈正相关，将其损失定义为预测的无雾图像 $J(x)$ 的亮度和其饱和度之间的差，具体表示为：

$$L_J = \|V(J_R(x)) - S(J_R(x))\|_F \quad (15)$$

其中， $V(J_R(x))$ 表示亮度， $S(J_R(x))$ 表示饱和度。

对于大气光估计网络，其损失由三项损失构成：

$$L_A = L_H + L_{KL} + \lambda L_{Reg} \quad (16)$$

其中 L_H 是估计的大气光 $A(x)$ 和 $A_{init}(x)$ 之间的损失， $A(x)$ 是从输入的图像数据中自动估计的初始化大气光。 L_H 具体表示为：

$$L_H = \|A(x) - A_{init}(x)\|_F \quad (17)$$

L_{KL} 是大气估计光网络中重参数化模块的的变分推理项损失，它的目的是使网络中重采样后的潜码 $\hat{z} \in N(\mu_z, \sigma_z^2)$ 和采样前 $z \in N(0, I)$ 的差异最小化。 L_{KL} 可表示为：

$$L_{KL} = KL \left(N(\mu_z, \sigma_z^2) \| N(0, I) \right) \quad (18)$$

其中， $KL(\cdot)$ 表示两个分布之间的 Kullback-Leibler 散度， z_i 表示 z 的第 i 个维度。 L_{Reg} 是为了避免大气光估计网络过拟合设置的正则项，具体表示为：

$$L_{Reg} = \frac{1}{2n} \sum_{i=1}^n \left(A(x)_i - \frac{1}{|N(A(x)_i)|} \sum_{y_i \in N(A(x)_i)} y_i \right)^2 \quad (19)$$

式中， $N(\cdot)$ 为二阶邻域， $|N(\cdot)|$ 为邻域大小， n 表示像素 x 的数量。 L_{Reg} 前的参数 λ 是用来平衡正则化的一个非负参数，在实验中，它被设置为 0.1。综上所述，网络的总损失通过结合以下五种损失函数来定义：

$$L = L_{Rec} + L_J + L_H + L_{KL} + \lambda L_{Reg} \quad (20)$$

4. 实验结果与分析

在本节中，本文提出的方法将在两个合成数据集和两个真实数据集上实现，以证明网络的有效性。同时，与 10 种去雾方法进行了定量和定性比较。下面详细介绍了实验设置、在各种数据集上的结果以及消融实验。

Table 1. Comparative analysis of dehazing outcomes across the synthetic datasets SOTS-Indoor and HSTS
表 1. 对合成数据集 SOTS-Indoor 和 HSTS 的去雾结果的比较分析

Type	Method	SOTS-Indoor		HSTS		Average	
		PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow
Learning-based	DehazeNet [8]	21.14	0.8472	24.48	0.9153	22.81	0.8812
	MSCNN [12]	17.57	0.8102	18.64	0.8168	18.10	0.8135
	AOD-Net [10]	19.06	0.8504	20.55	0.8973	19.80	0.8738
	CAP [7]	19.05	0.8364	21.53	0.8726	20.29	0.8545
	DCP [6]	16.62	0.8179	14.84	0.7609	15.73	0.7894
Prior-based	FVR [38]	15.72	0.7483	14.48	0.7624	15.10	0.7553
	NLD [17]	17.29	0.7489	18.92	0.7411	18.10	0.745
	DDIP [29]	16.97	0.7147	20.91	0.8842	18.94	0.7994
Unsupervised Learning	YOLY [30]	19.41	0.8327	23.82	0.9125	21.61	0.8726
	ZID [31]	19.83	0.8353	22.65	0.9011	21.24	0.8682
	Ours	20.47	0.8544	26.15	0.9507	23.31	0.9025

4.1. 实验设置与结果

4.1.1. 数据集

每种图像去雾方法都在两个公开的基准数据集上进行评估,包括 SOTS-Indoor 和 HSTS。SOTS-Indoor 和 HSTS 数据集都来自于 RESIDE 标准数据集[39]。SOTS-Indoor 包含 500 张室内合成雾霾图像,而 HSTS 包含 10 张合成雾霾图像和 10 张真实世界的雾霾图像。为了评价本文提出的方法的在真实世界的雾霾图像的去雾性能,从互联网上收集了 8 张真实世界的雾霾图像进行测试。

4.1.2. 评价指标

本文提出的方法是全面的评估,与其他最先进的方法使用定量和定性的比较和评估。为了保证公平的比较,本文采用了两种广泛使用的指标,即峰值信噪比(Peak Signal-to-noise Ratio, PSNR)和结构相似度 (Structural Similarity, SSIM) [40]来进行定量分析。这两个指标的值越高,表明去雾性能越好。同时,采用人眼观察的方法对不同方法的有效性进行了定性评价。

4.1.3. 实验配置

本文的实验是在 NVIDIA Tesla V100 GPU 上进行的,并使用了 PyTorch 工具箱。为了优化本文的网络,本文使用了 ADAM 优化器[41],并将学习率设置为默认值 0.001。对于本文的比较中包含的基线方法,取他们在原始论文中报告的结果。在没有所需结果的情况下,使用相应论文作者提供的源代码进行实验。

4.2. 比较实验

4.2.1. 定量结果

表 1 显示了各种去雾方法在 SOTS-Indoor、HSTS 上的 PSNR 和 SSIM 得分。可以看出,DehazeNet 在 SOTS-Indoor 的 PSNR 得分最高,本文所提出的网络在 SOTS-Indoor 的 SSIM 得分最高。在 HSTS 上,本文提出的网络优于所有其他方法,获得了最高的 PSNR 和 SSIM 分数。总之,本文提出的网络在所有

两个数据集上的性能始终优于其他方法。值得注意的是，本文的方法在两个合成数据集上都优于以前的无监督方法。总的来说，在所有三个类别中都优于或接近于表现最好的方法：基于学习的方法、基于先验的方法和无监督的方法。

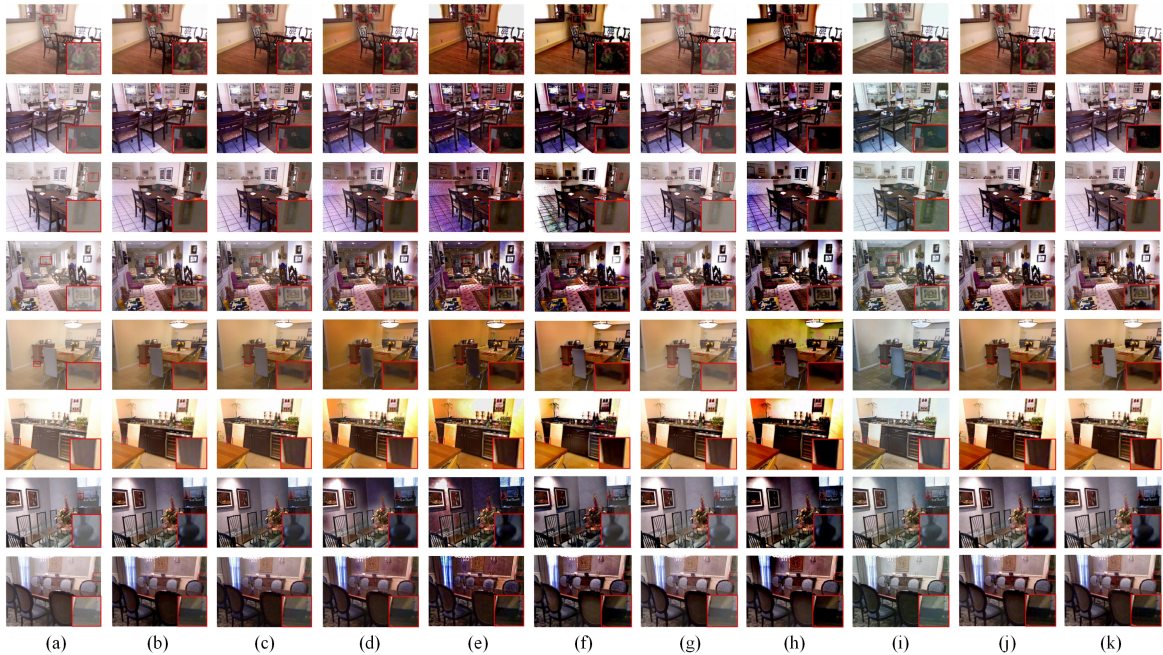


Figure 3. Qualitative comparison of methods for image dehazing on SOTS-Indoor
图 3. SOTS-Indoor 数据集中图像去雾处理方法的定性比较

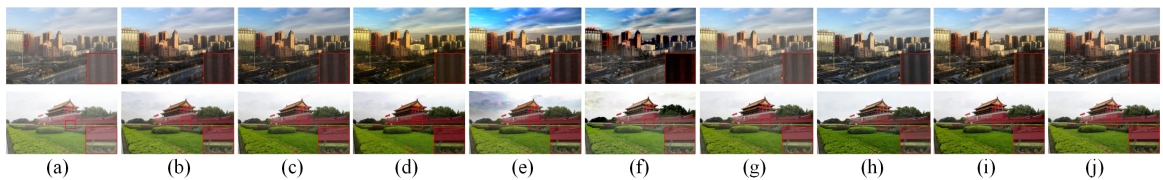


Figure 4. Qualitative comparison of methods for image dehazing on HSTS
图 4. 在 HSTS 数据集上进行图像去雾处理的方法的定性比较

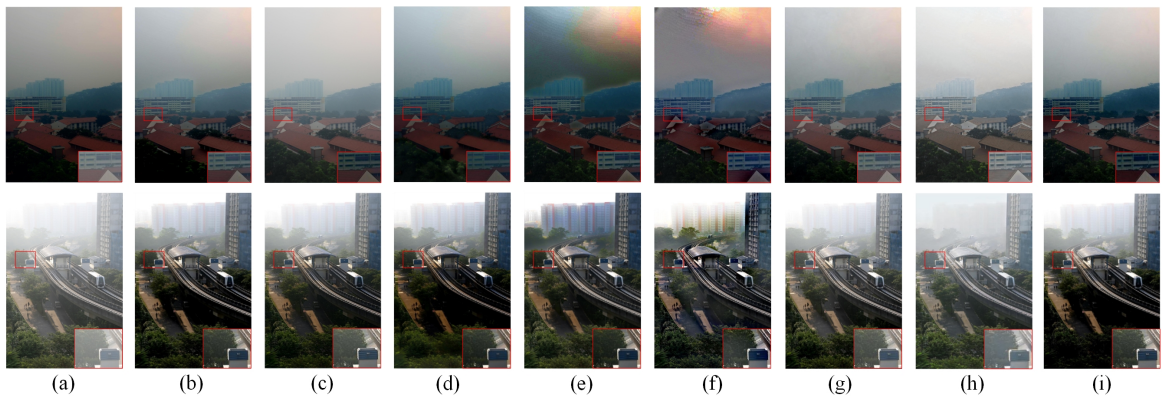


Figure 5. Qualitative comparison of methods for image dehazing on the real-world dataset
图 5. 在真实世界的数据集上的图像去雾处理方法的定性比较

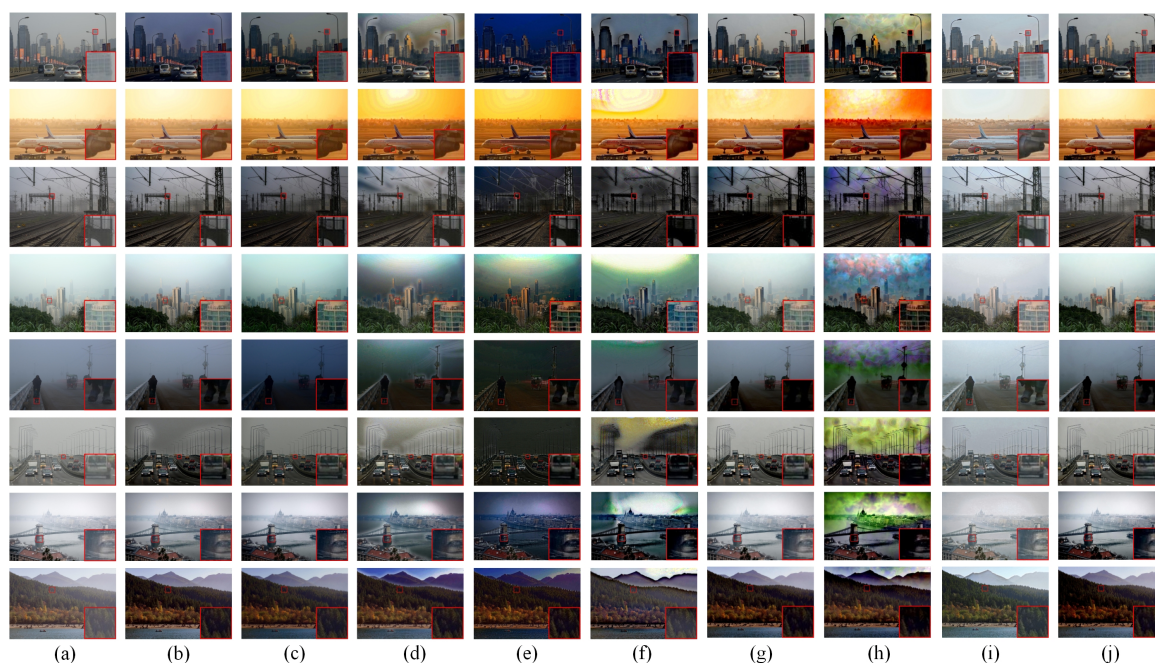


Figure 6. Qualitative comparison of methods for image dehazing on the real-world scenes
图 6. 对现实世界场景中的图像去雾处理方法的定性比较

4.2.2. 定性结果

图 3 和图 4 分别显示了不同方法在合成数据集 SOTS-Indoor 和 HSTS 上的去雾结果。图 3 中(a)~(k)列分别代表雾霾图像, DehazeNet、MSCNN、AOD-Net、DCP、FVR、DDIP、YOLY、ZID、本文提出网络的去雾结果和清晰无雾霾图像。图 4 中(a)~(j)列分别代表输入雾霾图像, DehazeNet、MSCNN、AOD-Net、CAP、DCP、DDIP、YOLY、本文提出网络的去雾结果和清晰无雾霾图像。如图 3 所示, 虽然 DehazeNet 等基于学习的方法在去雾性能方面优于基于 DCP 的方法, 但仍不能获得没有雾霾细节的去雾化图像(如图 3(b)~图 3(d)的第五和第六张图像所示)。同时, DCP 等基于先验的方法可以去除图像中的大部分雾霾, 但可能会发生图像失真(如图 3(e)~图 3(f)的第七、第八幅图像所示)。类似的视觉伪影也出现在以前的无监督方法中(图 3(g)~图 3(i))。相比之下, 尽管本文的网络没有获得最高的定量评估分数, 但它的图像在视觉上更吸引吸引力, 更接近参考的清晰图像。此外, 网络有效地消除了难以检测的雾霾细节, 而不会引起图像失真, 从而导致更高的视觉质量和更令人满意的去雾霾效果。

图 5 和图 6 分别显示了各种去雾方法对 HSTS 和本文从互联网上手动采集的真实世界雾霾图像的视觉去雾结果。如图 6 所示, ZID 在几乎所有 8 张图像中都在天空区域产生了明显的彩色斑块。FVR 倾向于使图像的色调变暗, 这在图 6(f)的第一和第三个图像中尤为明显。虽然 DehazeNet 和 MSCNN 去除了很大一部分的雾霾, 但它们经常会产生颜色失真(如图 6 中的第一张图片所示)。相比之下, 本文提出的方法有效地消除了雾霾, 并产生了视觉上令人满意的结果, 同时避免了这些问题。

Table 2. Evaluation of the effectiveness of the Multi Scale Attention block on different datasets
表 2. 评价多尺度注意力模块在不同数据集上的有效性

Setting	SOTS-Indoor		HSTS	
	PSNR	SSIM	PSNR	SSIM
gUNet	19.85	0.8360	22.74	0.9123

续表

Dehazeformer	18.77	0.7779	22.21	0.8747
TR-Net	20.07	0.8465	23.33	0.9194
Ours	20.47	0.8544	26.15	0.9507

4.3. 消融实验

4.3.1. 多尺度注意力模块的分析

为了验证在网络中提出的多尺度注意力模块的有效性, 本文替换了 JR-Net, 采用 gUNet 工作[13]中的 gConv 块、DehazeFormer [11]和 TR-Net 进行实验。其中, TR-Net 是网络的另一个子网络, 它由一个 5 级卷积层组成。实验结果见表 2。可以看出, 本文提出的多尺度注意力模块在两个数据集上都显著优于其他结构, 这表明多尺度注意力模块在从雾霾图像中恢复潜在的干净图像方面是非常有效的。

4.3.2. 两阶段去雾架构的分析

为了研究两阶段去雾结构在零网中的有效性, 本文通过构建 StageI和 StageII网络进行了消融研究。在 StageI网络中, 删除了 JR-Net 和 TR-Net, 而在 StageII中, 删除了粗估计模块。为了保证解耦网络的完整性, 实验将 StageI的 A-Net 保留, 粗估计模块只估计网络第一阶段的干净图像和透射图。假设去除网络顶部的大气光估计网络 A-Net, 那么缺失大气光估计后将无法重建基于物理模型的雾霾图像。表 3 为消融研究的结果。可以看出, 本文的两阶段去雾方法比仅使用 StageI和 StageII获得了更高的 PSNR 和 SSIM 分数, 这表明本文的两阶段去雾架构提高了网络的去雾性能。然而, StageII在两个数据集上的性能与两阶段去雾方法相比指标较低, 这表明粗估计模块对网络有一定的贡献。

Table 3. Evaluation of the effectiveness of the two-stage dehazing strategy on different datasets

表 3. 评价两阶段去雾在不同数据集上的有效性

Setting	SOTS-Indoor		HSTS	
	PSNR	SSIM	PSNR	SSIM
StageI	15.92	0.7599	15.20	0.6699
StageII	19.19	0.8220	22.13	0.8996
Two-stage (ours)	20.47	0.8544	26.15	0.9507

5. 结论

本文提出了一种基于无监督学习的两阶段解耦去雾网络, 利用三个子网络将观测到的雾霾图像分解为几个潜在层, 包括无雾霾图像层、透射图层和大气光层。本文网络的主要重点是恢复无雾霾图像和传输图。为了获得更精确的无雾图像, 引入了一种新的多尺度注意力模块, 并将其应用于无雾图像细化子网络中。多尺度注意力模块在自注意部分进行多尺度令牌聚合, 可以捕获不同尺度的特征, 有效地恢复雾霾图像中的潜在场景信息。实验结果表明, 该网络的网络具有良好的去雾霾性能和满意的视觉效果, 在 PSNR、SSIM 和主观视觉效果方面都优于现有的无监督去雾霾方法。

基金项目

国家自然科学基金资助项目(U2033210、62101387、62201400), 浙江省自然科学基金资助项目(LDT23F02024F02)。

参考文献

- [1] Huang, P., Zhao, L., Jiang, R., *et al.* (2021) Self-Filtering Image Dehazing with Self-Supporting Module. *Neurocomputing*, **432**, 57-69. <https://doi.org/10.1016/j.neucom.2020.11.039>
- [2] Li, X., Yu, H., Zhao, C., *et al.* (2023) DADRnet: Cross-Domain Image Dehazing via Domain Adaptation and Disentangled Representation. *Neurocomputing*, **544**, Article 126242. <https://doi.org/10.1016/j.neucom.2023.126242>
- [3] McCartney, E.J. and Hall, F.F. (1977) Optics of the Atmosphere: Scattering by Molecules and Particles. *Physics Today*, **30**, 76-77. <https://doi.org/10.1063/1.3037551>
- [4] Nayar, S. and Narasimhan, S. (1999) Vision in Bad Weather. *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Kerkyra, 20-27 September 1999, 820-827. <https://doi.org/10.1109/ICCV.1999.790306>
- [5] Narasimhan, S.G. and Nayar, S.K. (2002) Vision and the Atmosphere. *International Journal of Computer Vision*, **48**, 233-254. <https://doi.org/10.1023/A:1016328200723>
- [6] He, K., Sun, J. and Tang, X. (2009) Single Image Haze Removal Using Dark Channel Prior. 2009 *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, 20-25 June 2009, 1956-1963. <https://doi.org/10.1109/CVPR.2009.5206515>
- [7] Zhu, Q., Mai, J. and Shao, L. (2015) A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior. *IEEE Transactions on Image Processing*, **24**, 3522-3533. <https://doi.org/10.1109/TIP.2015.2446191>
- [8] Cai, B., Xu, X., Jia, K., *et al.* (2016) DehazeNet: An End-to-End System for Single Image Haze Removal. *IEEE Transactions on Image Processing*, **25**, 5187-5198. <https://doi.org/10.1109/TIP.2016.2598681>
- [9] Zhang, H. and Patel, V.M. (2018) Densely Connected Pyramid Dehazing Network. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 3194-3203. <https://doi.org/10.1109/CVPR.2018.00337>
- [10] Li, B., Peng, X., Wang, Z., *et al.* (2017) AOD-Net: All-in-one Dehazing Network. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 4780-4788. <https://doi.org/10.1109/ICCV.2017.511>
- [11] Song, Y., He, Z., Qian, H., *et al.* (2023) Vision Transformers for Single Image Dehazing. *IEEE Transactions on Image Processing*, **32**, 1927-1941. <https://doi.org/10.1109/TIP.2023.3256763>
- [12] Qin, X., Wang, Z., Bai, Y., *et al.* (2020) FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 11908-11915. <https://doi.org/10.1609/aaai.v34i07.6865>
- [13] Song, Y., Zhou, Y., Qian, H., *et al.* (2022) Rethinking Performance Gains in Image Dehazing Networks. arXiv 2209.11448.
- [14] Ren, W., Liu, S., Zhang, H., *et al.* (2016) Single Image Dehazing via Multi-Scale Convolutional Neural Networks. *Computer Vision—ECCV 2016*, Amsterdam, 11-14 October 2016, 154-169. https://doi.org/10.1007/978-3-319-46475-6_10
- [15] Wang, Z., Cun, X., Bao, J., *et al.* (2022) Uformer: A General U-Shaped Transformer for Image Restoration. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 17662-17672. <https://doi.org/10.1109/CVPR52688.2022.01716>
- [16] Fattal, R. (2015) Dehazing Using Color-Lines. *ACM Transaction on Graphics*, **34**, Article No. 13. <https://doi.org/10.1145/2651362>
- [17] Berman, D., Treibitz, T. and Avidan, S. (2016) Non-Local Image Dehazing. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 1674-1682. <https://doi.org/10.1109/CVPR.2016.185>
- [18] Chen, D., He, M., Fan, Q., *et al.* (2019) Gated Context Aggregation Network for Image Dehazing and Deraining. 2019 *IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 7-11 January 2019, 1375-1383. <https://doi.org/10.1109/WACV.2019.00151>
- [19] Li, R., Pan, J., Li, Z., *et al.* (2018) Single Image Dehazing via Conditional Generative Adversarial Network. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 8202-8211. <https://doi.org/10.1109/CVPR.2018.00856>
- [20] Engin, D., Genc, A. and Kemal Ekenel, H. (2018) Cycle-Dehaze: Enhanced CycleGAN for Single Image Dehazing. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, 18-22 June 2018, 938-9388. <https://doi.org/10.1109/CVPRW.2018.00127>
- [21] Qu, Y., Chen, Y., Huang, J., *et al.* (2019) Enhanced Pix2pix Dehazing Network. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 8152-8160. <https://doi.org/10.1109/CVPR.2019.00835>
- [22] Bai, H., Pan, J., Xiang, X., *et al.* (2022) Self-Guided Image Dehazing Using Progressive Feature Fusion. *IEEE Trans-*

- actions on Image Processing*, **31**, 1217-1229. <https://doi.org/10.1109/TIP.2022.3140609>
- [23] Wu, H., Qu, Y., Lin, S., *et al.* (2021) Contrastive Learning for Compact Single Image Dehazing. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 10546-10555. <https://doi.org/10.1109/CVPR46437.2021.01041>
- [24] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- [25] Zhao, D., Li, J., Li, H., *et al.* (2022) Complementary Feature Enhanced Network with Vision Transformer for Image Dehazing. arXiv 2109.07100.
- [26] Liu, Z., Lin, Y., Cao, Y., *et al.* (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 10012-10022. <https://doi.org/10.1109/ICCV48922.2021.00986>
- [27] Guo, C., Yan, Q., Anwar, S., *et al.* (2022) Image Dehazing Transformer with Transmission Aware 3D Position Embedding. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 5802-5810. <https://doi.org/10.1109/CVPR52688.2022.00572>
- [28] Lempitsky, V., Vedaldi, A., Ulyanov, D. (2018) Deep Image Prior. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 9446-9454. <https://doi.org/10.1109/CVPR.2018.00984>
- [29] Gandelsman, Y., Shocher, A., Irani, M. (2019) “Double-DIP”: Unsupervised Image Decomposition via Coupled Deep-Image-Priors. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 11018-11027. <https://doi.org/10.1109/CVPR.2019.01128>
- [30] Li, B., Gou, Y., Gu, S., *et al.* (2020) You Only Look Yourself: Unsupervised and Untrained Single Image Dehazing Neural Network. *International Journal of Computer Vision*, **129**, 1754-1767. <https://doi.org/10.1007/s11263-021-01431-5>
- [31] Li, B., Gou, Y., Liu, J.Z., *et al.* (2020) Zero-Shot Image Dehazing. *IEEE Transactions on Image Processing*, **29**, 8457-8466. <https://doi.org/10.1109/TIP.2020.3016134>
- [32] Kar, A., Dhara, S.K., Sen, D., *et al.* (2021) Zero-Shot Single Image Restoration through Controlled Perturbation of Koschmieder’s Model. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 16200-16210. <https://doi.org/10.1109/CVPR46437.2021.01594>
- [33] Ioffe, S. and Szegedy, C. (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, Lille, 6-11 July 2015, 448-456.
- [34] Maas, A.L. (2013) Rectifier Nonlinearities Improve Neural Network Acoustic Models. *Proceedings of the 30th International Conference on Machine Learning*,
- [35] Nair, V. and Hinton, G.E. (2010) Rectified Linear Units Improve Restricted Boltzmann Machines. *Proceedings of the 27th International Conference on International Conference on Machine Learning*, Haifa, 21-24 June 2010, 807-814.
- [36] Kingma, D. and Welling, M. (2013) Auto-Encoding Variational Bayes. *International Conference on Learning Representations*, Canada, 20 December 2013, 1-14.
- [37] Ba, J., Kiros, J.R. and Hinton, G.E. (2016) Layer Normalization. ArXiv abs/1607.06450.
- [38] Tarel, J.P., Hautière, N. (2009) Fast Visibility Restoration from a Single Color or Gray Level Image. 2009 *IEEE 12th International Conference on Computer Vision*, Kyoto, 29 September-2 October 2009, 2201-2208. <https://doi.org/10.1109/ICCV.2009.5459251>
- [39] Li, B., Ren, W., Fu, D., *et al.* (2019) Benchmarking Single-Image Dehazing and Beyond. *IEEE Transactions on Image Processing*, **28**, 492-505. <https://doi.org/10.1109/TIP.2018.2867951>
- [40] Wang, Z., Bovik, A., Sheikh, H., *et al.* (2004) Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, **13**, 600-612. <https://doi.org/10.1109/TIP.2003.819861>
- [41] Kingma, D.P. and Ba, J. (2015) Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*, San Diego, 7-9 May 2015, 1-15.