

基于改进CycleGAN模型的图像去雾算法

王 悦, 姜超谦

青岛大学计算机科学技术学院, 山东 青岛

收稿日期: 2024年10月25日; 录用日期: 2024年11月21日; 发布日期: 2024年11月29日

摘 要

图像去雾一直以来都是一个极具挑战的课题。本文基于循环生成对抗网络(cycle-consistent generative adversarial network, CycleGAN)进行改进。CycleGAN的生成器模块将引入基于泰勒展开的线性Transformer网络(MB-TaylorFormer), 有效解决注意力机制的二次计算复杂度, 同时使用不同大小卷积以及深度可分离卷积和可变形卷积结合进行多尺度标记, 提高去雾能力。实验表明, 算法改进后图像去雾任务性能有效提高。

关键词

图像去雾, 循环生成对抗网络, MB-TaylorFormer, 多尺度

Image Dehazing Algorithm Based on Improved CycleGAN Model

Yue Wang, Chaoqian Jiang

School of Computer Science and Technology, Qingdao University, Qingdao Shandong

Received: Oct. 25th, 2024; accepted: Nov. 21st, 2024; published: Nov. 29th, 2024

Abstract

Image dehazing has always been a highly challenging subject. This paper presents improvements to the cycle-consistent generative adversarial network (CycleGAN). The generator module of CycleGAN incorporates an MB-Taylor Former network, which is a linear Transformer based on Taylor expansion, effectively addressing the quadratic computational complexity of the attention mechanism. Additionally, it uses convolutions of varying sizes in combination with depthwise separable convolutions and deformable convolutions to perform multi-scale marking, thereby enhancing dehazing capabilities. Experiments demonstrate that the performance of image dehazing tasks is effectively improved following the enhancement of the algorithm.

Keywords

Image Dehazing, CycleGan, Mb-TaylorFormer, Multi-Scale

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

在生活中我们可以发现,雾的存在会严重影响图像的质量,遮盖图像的信息。图像与信息模糊、失真会影响视觉应用的效果。因此,图像去雾算法的研究对于自动驾驶、视频监控、目标检测等视觉系统具有重要意义。近年来深度学习技术不断发展,准确、高效去除图像中的雾也备受关注。传统的去雾算法根据是否使用物理模型来简单地分成两类。第一类是基于图像增强的算法,包含直方图均衡化[1]、Retinex 理论[2]、小波变换算法等,它未考虑雾的物理特性,仅通过调整图像的对比度和饱和度来实现去雾,因此在复杂场景的去雾表现不佳。第二类是基于大气散射模型[3]的图像去雾算法,利用先验知识或假设图像退化的物理模型实现去雾。HE 等[4]将暗通道先验(dark channel prior, DCP)算法映射到大气散射模型,设计了一种有效的去雾算法。在此基础上多人进行了方法的改进得到了广泛应用。但这类方法对参数估计的精度要求高,当区域不满足先验算法会失效,出现颜色失真和光晕现象。近年来深度学习方法广泛应用于各类图像处理的任务中,图像去雾也有了更有效的方法。基于深度学习的去雾网络根据输入输出的不同也可以简单分为两类。第一类是非端到端的去雾网络,它通过训练来获得输入图像的无雾透射图。在非端到端去雾网络中,CAI 等[5]基于大气模型提出了深度学习的去雾算法 DehazeNet,算法将边缘细节恢复较好,但仍然存在残留;Ren 等[6]设计了一种由粗尺度卷积网络和细尺度卷积网络组成多尺度卷积神经网络模型 MSCNN 用于图像去雾。另一类是端到端的去雾网络,它可以通过学习直接将有雾图像转换成清晰无雾图像,不需估计介质传输图及大气光图像,减少误差,得到更好的去雾效果。在端到端去雾网络中,LI 等[7]提出了 AOD-Net 通过一体化网络架构实现了高效的图像去雾。Wang [8]等提出的 FFA-Net 通过特征融合注意力机制显著提升了去雾效果。Li 等[9]提出的 MB-TaylorFormer 通过多分支高效变压器结合泰勒公式显著提升了去雾效果。传统的图像增强和基于大气散射模型的去雾算法相比,基于深度学习的算法效果更理想,但目前图像去雾依然存在很多问题等待进一步优化。例如,大多数算法基于成对数据来训练,同时收集有雾图像与无雾图像难度较大,在合成数据集上训练的网络对真实图像去雾的效果较差;注意力机制的二次计算复杂度使得 Transformer [10]应用受到了限制等。为解决这些问题,本文选用 CycleGAN [11],使用非成对的数据进行训练并将生成器进行改进,替换为改进后的 MB-TaylorFormer 网络,增强 CycleGAN 的去雾能力。改进后的生成器为 Transformer 变体,采用泰勒扩展来近似注意力机制,实现了线性轻量级的复杂度,增强了去雾效果。

2. CycleGAN 网络结构

CycleGAN 是在非配对数据集上完成端到端去雾网络的训练,无需数据对齐和标注,用生成对抗的方式学习映射关系。模型有两个生成器和两个判别器,通过两个生成器完成循环过程通过增加训练成本换取训练的灵活性,网络结构如图 1 所示,有雾图像经生成器 G_{AB} 转换为无雾图像,无雾图像通过 G_{BA} 转换成有雾图像,通过原始输入图像与循环后转换图像计算差异,判别器网络对生成器网络处理后的图像进行评估和反馈来调整生成器网络的模型参数,使生成器网络模型可以更好地生成去雾的图像。本文主

要通过改进该模型的生成器部分来增强模型对信息的利用，提高模型生成图像的质量。

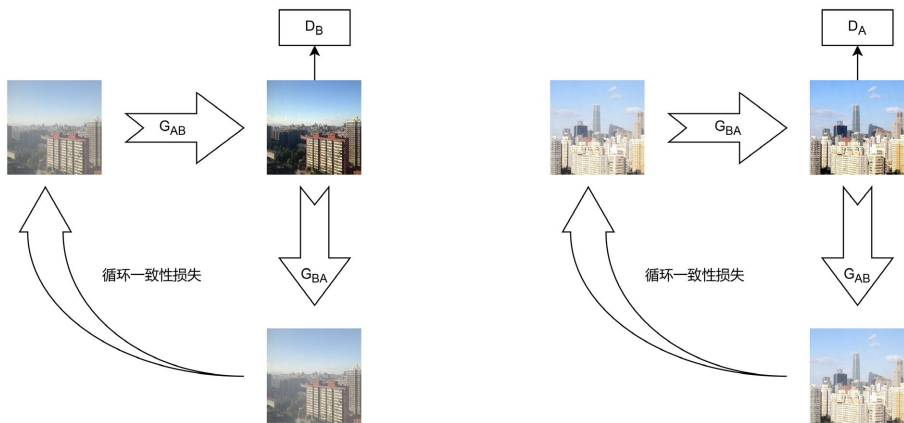


Figure 1. Network architecture of CycleGAN
图 1. CycleGAN 网络结构

3. 改进 CycleGAN 网络架构

3.1. 生成器

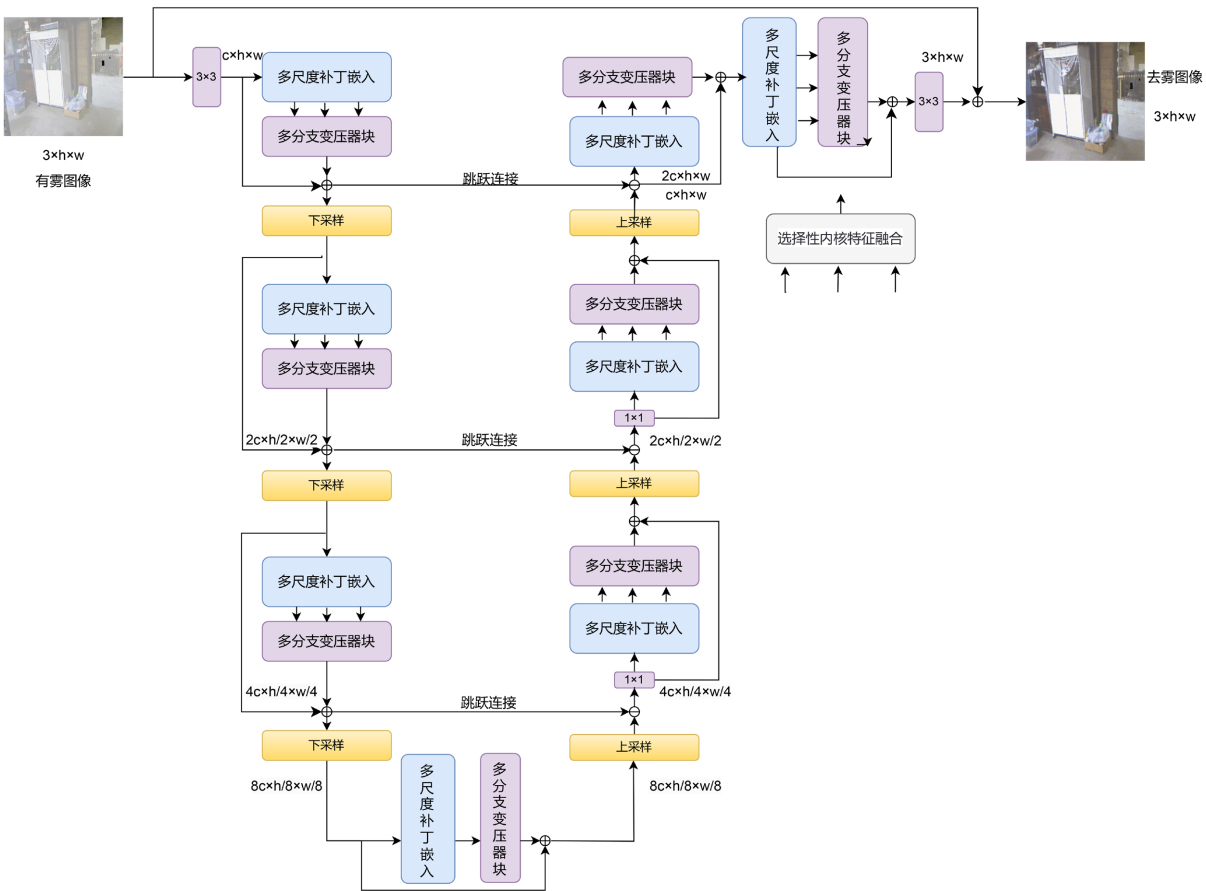


Figure 2. Overall architecture of generator
图 2. 生成器整体结构

原有生成器结构去雾效果较差, 因此本文将生成器进行了改进, 由于选用 CycleGAN 本身增加了训练成本, 因此在生成器部分本文使用 Transformer 变体, 降低计算成本。正如图 2 网络可见, 该网络采用 U-net [12] 结构, 首先使用 3×3 卷积操作进行浅层特征提取, 随后利用“四级编码器—解码器”网络对深层特征进行提取。在每级编码器中引入了由多尺度补丁嵌入和多分支 Transformer 组成的残差块, 其中多尺度补丁嵌入模块通过不同大小的卷积核及可变形卷积来捕捉图像的局部与全局特征, 并将其输入到多个 Transformer 分支当中。每个 Transformer 分支由多个 Transformer 编码器组成, 用于处理来自多尺度补丁嵌入模块的特征。最后使用选择性核融合(SKFF) [13] 模块融合不同分支生成的特征。每个阶段使用像素取消洗牌和像素洗牌操作[14]对特征进行下采样和上采样, 以保留更多局部信息并避免信息丢失。为了聚合编码器和解码器的信息, 我们采用了跳跃连接[15]机制, 并通过 1×1 卷积层进行降维。为进一步恢复图像的精细结构和纹理细节, 在“编码器—解码器”结构之后, 我们加入了残差块。最终, 通过 3×3 卷积层减少通道数并输出残差图像。这一改进后的生成器结构有效地提升了去雾性能, 同时保持了较高的计算效率。

3.1.1. 多尺度补丁嵌入模块

采用堆叠扩展感受野的操作, 本文选用堆叠多个具有不同小核的可变形卷积层增加网络的深度, 提供多层次的语义信息, 帮助减少参数和计算负担。并在所有可变形卷积层后面使用 Hardswish [16] 激活函数来激活可变形卷积层的输出。这一模块可以更好地捕捉不同尺度和层次的视觉特征, 提高对视觉元素多样性的建模能力, 多尺度补丁嵌入模块结构如图 3 所示。

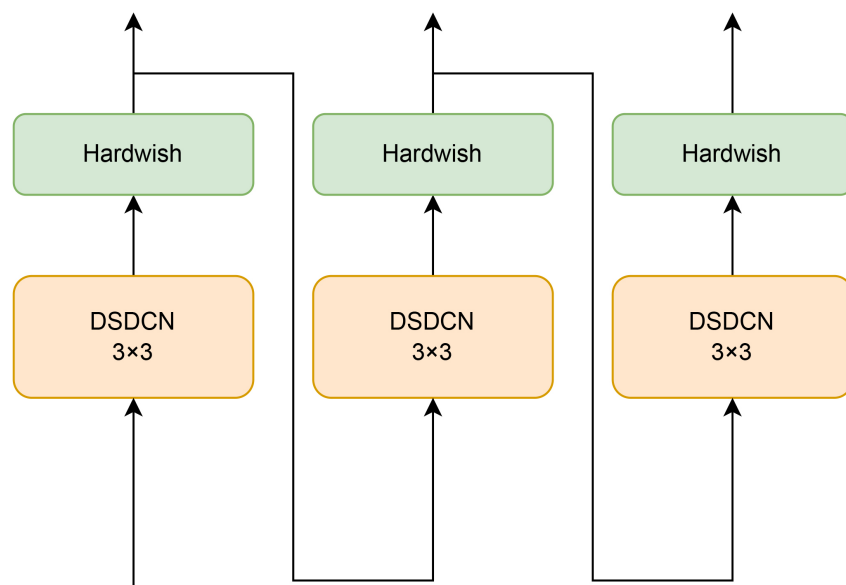


Figure 3. Architecture of multi-scale patch embedding
图 3. 多尺度补丁嵌入模块结构

3.1.2. 泰勒扩展自注意力机制

Softmax 自我关注的计算复杂度为 O , 计算成本昂贵。因此本文学习引入泰勒扩展自注意力机制, 降低计算成本, Q , K , V 采用深度卷积生成的方法, 强调局部上下文。使用多头结构, 从层级的顶部到底部, 头部的数量逐渐增加。因一阶泰勒展开忽略余数产生误差, 引入 MSAR 模块通过学习查询和关键字的局部信息进行修正, 多泰勒扩展自注意力机制结构如图 4 所示。

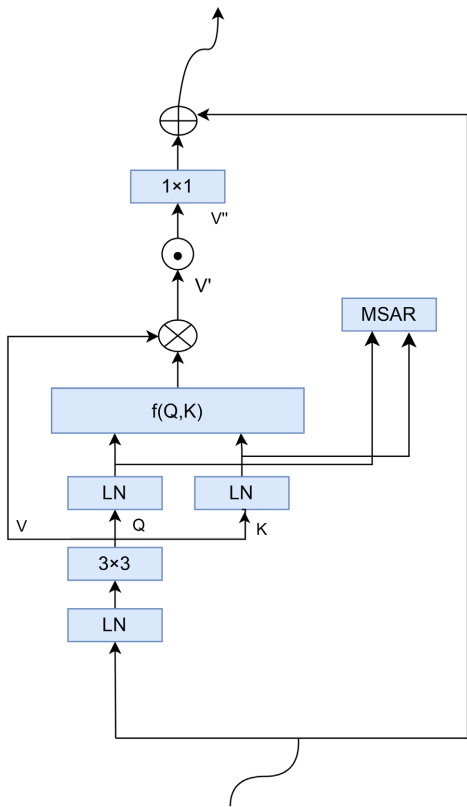


Figure 4. Architecture of Taylor expanded multi-head self-attention
图 4. 泰勒扩展自注意力机制结构

3.2. 判别器

本文判别器依然保留原始 CycleGAN 模型当中的判别器为 DA 和 DB 确保模型在图像去雾任务中的有效性和鲁棒性，DA 和 DB 分别用于判别源域(有雾图)和目标域(去雾图)中的图像真实性。这两个判别器遵循 PatchGAN 结构，旨在捕捉图像的局部特征，并提高模型的学习能力。每个判别器由多个卷积层组成，逐步降低输入图像的空间分辨率，同时增加特征图的深度。每个判别器的整体结构如图 5 所示。

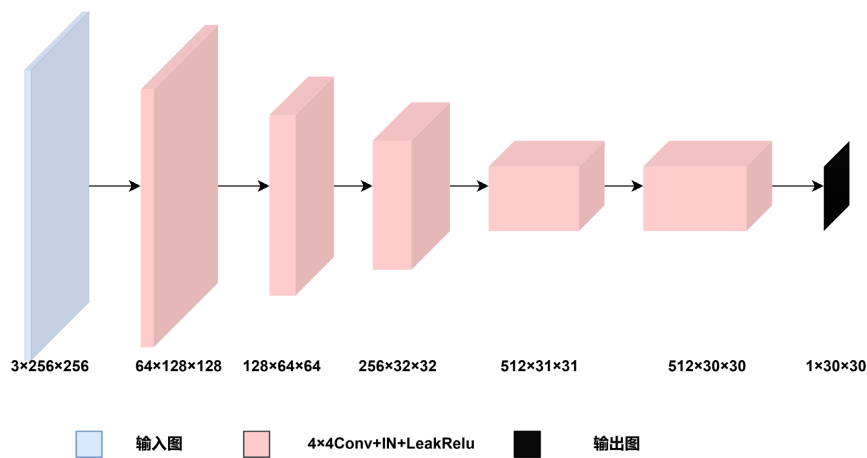


Figure 5. Architecture of discriminator
图 5. 判别器结构

4. 损失函数

生成对抗损失用于训练生成器和判别器之间的对抗过程。生成对抗损失用于训练生成器, 使其生成的去雾图像可以欺骗判别器。通过最小化生成对抗损失, 学会生成更加逼真的图像去雾图像。我们采用二分类交叉熵损失(Binary Cross Entropy, BCE)来计算生成对抗损失, 对于生成器和判别器, 可以分别表示为

$$L_g = -E_{x \sim P_{noise}(x)} [\log D(G(x))] \quad (1)$$

$$L_d = -E_{y \sim P_{data}(y)} [\log D(y)] - E_{x \sim P_{data}(x)} [\log (1 - D(G(x)))] \quad (2)$$

其中, L_g 为生成器生成对抗损失函数, L_d 为判别器损失函数, $y \sim P_{data}(y)$ 是真实无雾图像的分布, $x \sim P_{data}(x)$ 是有雾图像的分布。

循环一致性损失用于确保从源域到目标域再回到原域的变换是一致的, 通过计算两个张量元素之间绝对值差异的平均值帮助模型学习双向映射关系, 我们采用 L1 范数损失(L1 Norm Loss)函数来计算循环一致性损失:

$$L_{cyc} = E_{x \sim P_{data}(x)} [\|G_{BA}(G_{AB}(x)) - x\|_1] + E_{y \sim P_{data}(y)} [\|G_{AB}(G_{BA}(y)) - y\|_1] \quad (3)$$

其中, $E_{x \sim P_{data}(x)}$ 和 $E_{y \sim P_{data}(y)}$ 分别表示对来自数据分布 $P_{data}(x)$ 的样本 x 和 y 的期望值, G_{AB} 和 G_{BA} 是两个生成器网络, G_{AB} 用于将域 A 中的样本 x 转换成域 B 中的样本, G_{BA} 用于将域 B 中的样本 y 转换成域 A 中的样本。 $\|\cdot\|_1$ 表示 L1 范数, 即绝对值之和。

本文生成器和判别器的损失分别计算, 并且在每次迭代中分别更新。各模块综合几个损失函数来训练整体网络, 生成器综合损失函数 $GTotalloss$ 为

$$GTotalloss = \lambda_1 L_{cyc} + \lambda_2 L_g \quad (4)$$

其中, λ_1 和 λ_2 分别为两个子损失函数的权重参数。经过实验验证, 当参数分别设置为 8 和 0.8 时去雾的效果最好。判别器综合损失函数 $DTotalloss$ 为

$$DTotalloss = L_d \quad (5)$$

5. 实验结果与分析

5.1. 实验配置

实验在型号为 GeForce GTX 1080 Ti 的 GPU 上使用 pytorch 框架完成。Adam optimizer [17] 用来训练网络, 其中衰减率 β_1 和 β_2 采用默认值 0.5 和 0.999。网络的 batchsize 为 1, 初始学习率为 0.001, 训练迭代次数为 200 次。实验直接采用真实世界非配对图片, 训练时, 选取 800 张有雾图像和无雾图像作为训练集。在对训练集预处理时, 选择将图像的尺寸裁剪为 256×256 , 并使用水平和垂直翻转图像来增强数据集的质量。

5.2. 评价指标

本文采用峰值信噪比 PSNR (Peak Signal to Noise Ratio) 与结构相似性指数 SSIM (Structural Similarity Index) 作为评价指标对实验结果进行评价。PSNR 通过均方差进行定义, 其数值越高图像质量越好、处理后图像与清晰图像之间误差越小、失真越少。SSIM 用于衡量两幅图像相似程度, 其数值越接近于 1 代表两幅图结构越相似。

5.3. 实验结果和分析

为验证算法的性能,将本文算法与基于先验知识的 DCP 去雾算法;基于大气模型的深度学习的 AOD-Net 去雾算法;基于端到端的深度学习的 GCANet、FFA-Net、MB-TaylorFormer 去雾算法进行比较,实验结果如图 6 所示,SSIM 和 PSNR 指标定量比较不同算法的去雾结果,见表 1。DCP、AOD-Net 去雾算法的色彩饱和度比较好,但部分位置出现颜色失真的问题,图像不真实;GCANet 和 FFANet 细节恢复较好,但处理后的图片中仍有残留的雾,天空颜色过亮;MB-TaylorFormer 可以将雾基本去除,但整体的亮度偏暗。与上面接种去雾算法比较,本文的算法可以更好的去雾并保留下更清晰的细节信息特征,在图像的色彩饱和度、对比度及亮度等方面的恢复最接近真实图像。

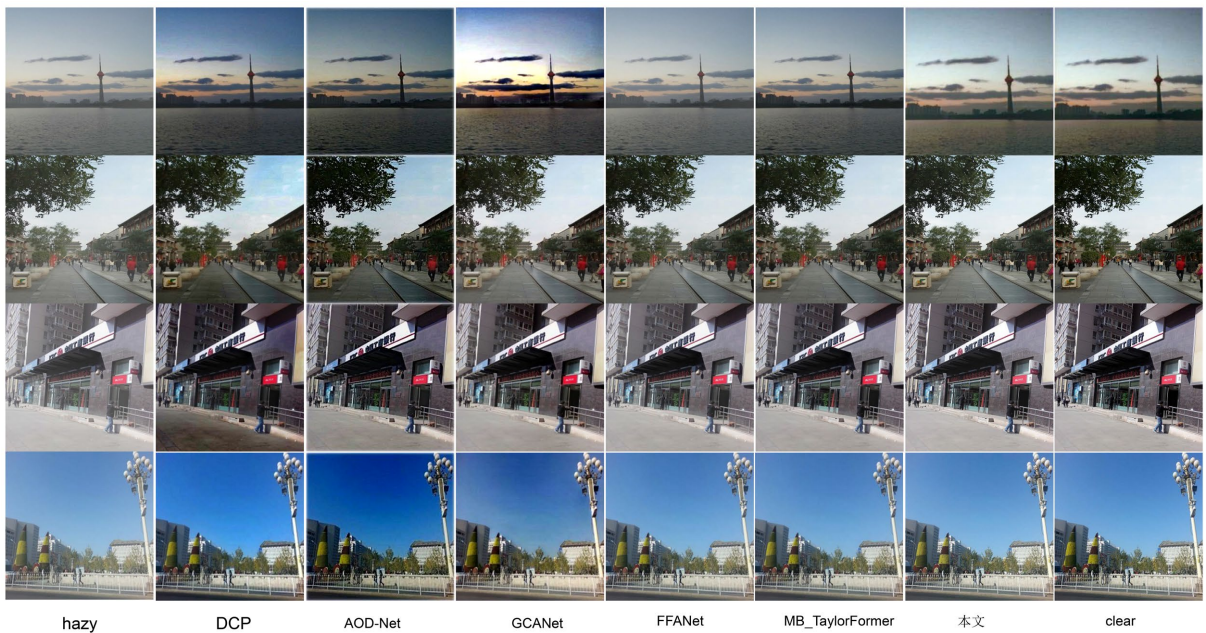


Figure 6. Results of comparison experiment
图 6. 对比实验结果图

Table 1. Analysis of experimental results
表 1. 不同算法的实验结果分析

评价指标	DCP	AOD-Net	GCANet	FFA-Net	MB_TaylorFormer	本文算法
SSIM	0.81	0.85	0.89	0.90	0.92	0.93
PSNR	18.49	20.06	24.23	27.31	27.35	28.75

6. 结论

本文提出了一种基于 CycleGAN 模型改进的去雾算法,使用非配对数据集避免了同时收集有雾图像与无雾图像难度较大、不对齐的问题,由于 CycleGAN 模型计算成本大,在生成器中本文选择改进后的 MB-TaylorFormer 网络,在更好捕捉图像多尺度特征的同时有效解决注意力机制的二次计算复杂度。另外本文算法属于端到端去雾,通过直接学习输入图像的特征,不依赖传统的大气散射模型,这样避免了因估计不准确的透射率和大气光值而导致的去雾不彻底和颜色偏差等问题,从而实现了更高质量的去雾效果。

参考文献

- [1] 纪平, 胡学友, 张瑞琦. 基于直方图均衡算法的图像增强技术研究[J]. 蚌埠学院学报, 2021, 10(2): 40-43.
- [2] Land, E.H. and McCann, J.J. (1971) Lightness and Retinex Theory. *Journal of the Optical Society of America*, **61**, 1-11. <https://doi.org/10.1364/josa.61.000001>
- [3] Gao, T., Li, K., Chen, T., Liu, M., Mei, S., Xing, K., *et al.* (2020) A Novel UAV Sensing Image Defogging Method. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **13**, 2610-2625. <https://doi.org/10.1109/jstars.2020.2998517>
- [4] He, K., Sun, J. and Tang, X. (2011) Single Image Haze Removal Using Dark Channel Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **33**, 2341-2353. <https://doi.org/10.1109/tpami.2010.168>
- [5] Cai, B., Xu, X., Jia, K., Qing, C. and Tao, D. (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>
- [6] Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X. and Yang, M. (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
- [7] Li, B., Peng, X., Wang, Z., Xu, J. and Feng, D. (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>
- [8] Qin, X., Wang, Z., Bai, Y., Xie, X. and Jia, H. (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>
- [9] Qiu, Y., Zhang, K., Wang, C., Luo, W., Li, H. and Jin, Z. (2023) MB-TaylorFormer: Multi-Branch Efficient Transformer Expanded by Taylor Formula for Image Dehazing. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 12756-12767. <https://doi.org/10.1109/iccv51070.2023.01176>
- [10] 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.
- [11] Engin, D., Genc, A. and Ekenel, H.K. (2018) Cycle-Dehaze: Enhanced CycleGAN for Single Image De-Hazing. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, 18-22 June 2018, 825-833. <https://doi.org/10.1109/CVPRW.2018.00127>
- [12] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S. and Yang, M. (2022) Restormer: Efficient Transformer for High-Resolution Image Restoration. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 5718-5729. <https://doi.org/10.1109/cvpr52688.2022.00564>
- [13] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M., *et al.* (2020) Learning Enriched Features for Real Image Restoration and Enhancement. *Computer Vision—ECCV 2020*, Glasgow, 23-28 August 2020, 492-511. https://doi.org/10.1007/978-3-030-58595-2_30
- [14] Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A.P., Bishop, R., *et al.* (2016) Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 1874-1883. <https://doi.org/10.1109/cvpr.2016.207>
- [15] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Munich, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [16] Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L., Tan, M., *et al.* (2019) Searching for MobileNetV3. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 1314-1324. <https://doi.org/10.1109/iccv.2019.00140>
- [17] Kingma, D.P. and Ba, J. (2014) Adam: A Method for Stochastic Optimization. arXiv:1412.6980.