Dehaze-RetinexGAN: Real-World Image Dehazing via Retinex-based Generative Adversarial Network

Xinran Wang[1*], Guang Yang[1*], Tian Ye[2], Yun Liu[1,3†]

[1]College of Artificial Intelligence, Southwest University, Chongqing, China
[2]The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
[3]College of Computing and Data Science, Nanyang Technological University, Singapore, Singapore
wowwowyahoo@gmail.com, learnforai@gmail.com, owentianye@hkust-gz.edu.cn, yunliu@swu.edu.cn

Abstract

Deep learning based dehazing networks trained on paired synthetic data have shown impressive performance, but they struggle with significant degradation in generalization ability on real-world hazy scenes. In this paper, we propose Dehaze-RetinexGAN, a lightweight Retinex-based Generative Adversarial Network for real-world image Dehazing using unpaired data. Our Dehaze-RetinexGAN consists of two stages: self-supervised pre-training and weakly-supervised fine-tuning. During the pre-training, we reduce the image dehazing task to an illumination-reflectance decomposition task based on the duality correlation between Retinex and dehazing. Specifically, a decomposition network named DecomNet is constructed to obtain an illumination and a reflectance, simultaneously. Moreover, a self-supervised learning strategy is developed to construct the connection between the preliminary dehazed result and the input hazy image, which constrains the solution space of DecomNet and accelerates training, leading to a more realistic dehazed result. In the fine-tuning stage, we develop a dual DTCWT-based attention module and embed it into the U-Net architecture to further improve the quality of preliminary result in the frequency domain. In addition, the adversarial learning is employed to constrain the relevance between the clean image and the final dehazed result in a weakly supervised manner, which can promote more natural performance. Extensive experiments on several real-world datasets demonstrate that our proposed framework performs favorably over state-of-the-art dehazing methods in visual quality and quantitative evaluation.

Introduction

The presence of haze leads to reduced visibility, decreased contrast, blurred details, and color distortion, severely affecting the performance of real-world vision-based systems such as autonomous driving and object detection. Therefore, image dehazing that aims to restore scene details is an important task in low-level computer vision.

Early dehazing techniques are mainly based on enhancement strategies, such as histogram equalization (Xu, Liu, and Chen 2009) and Retinex theory (Noori, Gholizadeh, and Rafsanjani 2024), to improve the perceptual visibility of hazy images. However, directly employing these enhancement methods on hazy images may lead to unnatural results due to the lack of physical mechanisms.

Figure 1: Dehazing results on a real-world hazy image.

Relying on the atmospheric scattering model, prior-based dehazing algorithms, such as dark channel prior (DCP) (He, Sun, and Tang 2011), region line prior (RLP) (Ju et al. 2021), and saturation line prior (SLP) (Ling et al. 2023), have been proposed to estimate the transmission and further recover the haze-free image. Unfortunately, these presented priors are not well-suited for handling diverse real-world scenarios. For instance, the well-known DCP fails in situations where the scene colors are similar to the airlight.

Leveraging the feature extraction capabilities of deep learning, MSCNN (Ren et al. 2016) and DehazeNet (Cai et al. 2016) were proposed to estimate the transmission. Subsequently, several dehazing networks, such as AODNet (Li et al. 2017), FFANet (Qin et al. 2020), MSBDN (Dong et al. 2020), Dehamer (Guo et al. 2022), and DEA-Net (Chen, He, and Lu 2024), have been developed to construct the mapping relationship between hazy images and clean data. However, due to the domain shift between synthetic and real data, these models trained solely on paired synthetic datasets suffer from limited generalization ability on real-world scenes.

To overcome the dependency on paired data, many

*These authors contributed equally.
†Corresponding Author
Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

unpaired learning-based methods (Engin, Genc, and Kemal Ekenel 2018; Shao et al. 2020; Chen et al. 2022) have been proposed to achieve image dehazing. The above approaches exploit CycleGAN (Zhu et al. 2017) framework to construct the dehazing and re-dehazing cycles, maintaining the content consistency. Nevertheless, they fail to consider the physical properties of real-world hazy conditions. Although RefineDNet (Zhao et al. 2021) and D4 (Yang et al. 2022) integrate the atmospheric scattering model and CycleGAN to boost the generalization ability, they overlook the intrinsic connection between dehazed results and hazy images, resulting in unnatural performance for real-world scenarios. Moreover, these methods generally require time-consuming training, which often causes efficiency issues in real applications. Recently, PSD (Chen et al. 2021), RIDCP (Wu et al. 2023), and KANet (Feng et al. 2024) have been proposed to focus on restoring real-world hazy scenes, whereas the dehazing performance of these algorithms needs further improvement.

In this paper, we propose a lightweight Retinex-based Generative Adversarial Network for real-world image dehazing, called Dehaze-RetinexGAN. Our Dehaze-RetinexGAN comprises two phases: self-supervised pre-training and weakly-supervised fine-tuning. In the pre-training, we propose a decomposition network named DecomNet to split the inverted intensity of an input hazy image into two components: illumination and reflectance. In this way, the illumination and the inverted reflectance correspond to the transmission and the preliminary dehazed result, respectively. Moreover, considering the connection between the input hazy image and the preliminary dehazed result, we develop a self-supervised learning strategy to make the dehazed result more realistic. In the fine-tuning stage, a dual Dual-Tree Complex Wavelet Transform (DTCWT)-based attention module, embedded into the U-Net architecture, is employed to improve the structures and textures of the preliminary result in the frequency domain. In addition, the adversarial learning is adopted to build the relationship between dehazed results and unpaired real-world clean data in a weakly supervised manner, which promotes the final results to conform more closely to the distribution of clean data, thereby achieving natural haze removal performance for real-world scenes. In Fig. 1, our Dehaze-RetinexGAN generates a more realistic dehazed result and outperforms the other two real-world image dehazing methods, namely DAD (Shao et al. 2020) and PSD (Chen et al. 2021).

The main contributions of this paper are as follows:

- We propose Dehaze-RetinexGAN, a novel lightweight two-stage framework for real-world image dehazing. By leveraging both self-supervised and weakly-supervised learning, our proposed Dehaze-RetinexGAN, trained on unpaired data, effectively improves generalization and natural performance on real-world hazy scenes.
- We propose a self-supervised learning strategy to consider the correlation between the dehazed results and the hazy images, which helps constrain the solution space of dehazing and improve the efficiency of the network.
- We propose a dual DTCWT-based attention module to enhance the structures and textures in the frequency domain, thus achieving visually pleasing dehazed results with richer details.
- Extensive experiments on several real-world datasets prove that our Dehaze-RetinexGAN achieves competitive performance compared to existing state-of-the-art dehazing methods both qualitatively and quantitatively.

Related Work

Traditional Dehazing Methods

In the early stages, enhancement-based methods are employed to achieve single image dehazing by improving the contrast and visibility of hazy images, such as CLAHE (Xu, Liu, and Chen 2009), Retinex theory (Land 1977), and wavelet transform (Zhang et al. 2014). However, these methods overlook the physical principles of haze imaging, limiting their dehazing performance and leading to unnatural results. Therefore, researchers focus on the atmospheric scattering model (Narasimhan and Nayar 2002) to describe the imaging procedure of a hazy image and explore effective priors (DCP (He, Sun, and Tang 2011), RLP (Ju et al. 2021), and SLP (Ling et al. 2023)) to recover haze-free images. In addition, several variational models (Liu et al. 2023b; Li et al. 2024) have been presented for haze removal. However, these priors or constraints may be valid only in specific situations and are not well-suited for complex real-world haze conditions.

Learning-Based Dehazing Methods

Unlike traditional methods that rely on enhancement techniques or hand-crafted priors, learning-based dehazing methods aim to design network architecture and train it on large-scale paired data to learn the transmissions and haze-free results from hazy images. For example, DehazeNet (Cai et al. 2016) and MSCNN (Ren et al. 2016) have been proposed to estimate the transmission maps by extracting the haze-relevant features. Subsequently, many end-to-end neural networks (Li et al. 2017; Qin et al. 2020; Dong et al. 2020; Guo et al. 2022; Chen, He, and Lu 2024; Liu et al. 2023a; Ye et al. 2023; Chen et al. 2024; Jin, Yang, and Tan 2022) construct the mapping relationships from degraded images to clean results. Although these approaches achieve excellent performance on synthetic data, they usually struggle to generalize effectively to real-world hazy scenes.

Since the paired data of real-world hazy and clean images is difficult to obtain, unpaired learning-based dehazing networks gradually become the focus. RefineDNet (Zhao et al. 2021) combines the CycleGAN architecture and the atmospheric scattering model to restore visibility and improve realism in a weakly supervised fashion. D4 (Yang et al. 2022) decomposes the transmission map into the density and depth for unpaired haze synthesis and removal. To improve the real-world dehazing performance, PSD (Chen et al. 2021) employs real-world hazy images to fine-tune the network in an unsupervised manner. RIDCP (Wu et al. 2023) exploits high-quality codebook priors to achieve real-world image

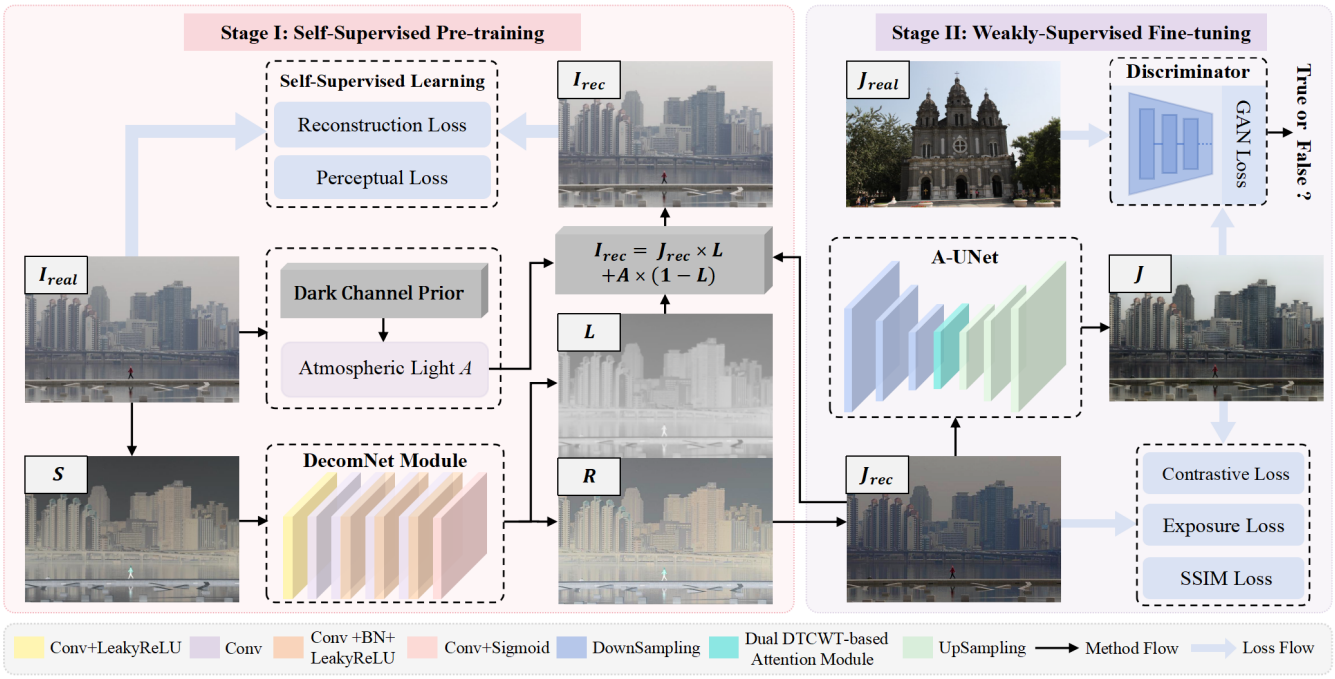


Figure 2: Overall architecture of our Dehaze-RetinexGAN. Our proposed Dehaze-RetinexGAN mainly consists of two stages: self-supervised pre-training and weakly-supervised fine-tuning. In the pre-training stage, an input real-world hazy image I_{real} is first inverted and a lightweight DecomNet module is developed to decompose the inverted hazy image S into an illumination L and a reflectance R, simultaneously. Then, the classic atmospheric scattering model is employed to reconstruct the hazy image I_{rec}. Subsequently, a self-supervised learning paradigm is introduced to construct the correlation between I_{real} and J_{rec}, which further constrains the solution space of DecomNet to obtain a more realistic clean image J_{rec}. For fine-tuning, a dual DTCWT-based attention block is devised and embedded into U-Net (Ronneberger, Fischer, and Brox 2015) architecture to enhance the structures and textures of J_{rec} in the frequency domain. The adversarial learning is imposed between the clean image J_{real} and dehazed result J in a weakly supervised fashion, allowing the knowledge from the clean image to be transferred to the dehazed result without requiring paired data.

dehazing. KANet (Feng et al. 2024) follows "localization-and-removal" pipeline to help the network capture haze-relevant features and reduce the dependency between clean background features and haze. Unfortunately, the above algorithms fail to account for the intrinsic connections between the dehazed results and the input hazy images, leading to unnatural performance on real-world hazy images. Therefore, there is still room for further improvement in real-world image dehazing.

Methodology

Framework Overview

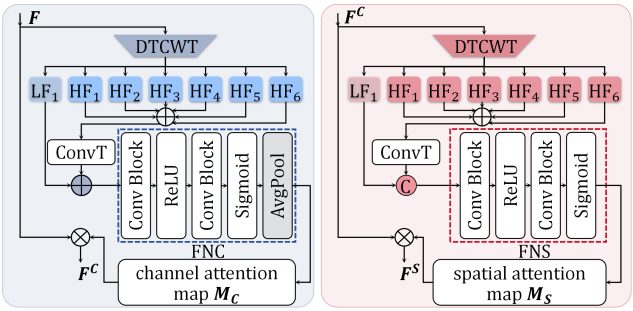
As shown in Fig. 2, our Dehaze-RetinexGAN consists of two phases: self-supervised pre-training and weakly-supervised fine-tuning.

Self-Supervised Pre-training. Based on the relationship between dehazing problem and Retinex theory (Galdran et al. 2018), the image dehazing task can be transformed into an illumination-reflectance decomposition problem. Thus, a lightweight DecomNet module is devised to separate an inverted hazy image into an illumination L and a reflectance R. The obtained illumination and inverted reflectance are considered as the transmission L and preliminary dehazed result J_{rec}, respectively. The classic atmospheric scattering model, which combines the DCP-based atmospheric light A, the transmission L, and the preliminary result J_{rec}, is adopted to acquire the reconstructed hazy image I_{rec}. Furthermore, a self-supervised learning paradigm is developed to construct the relationship between the input hazy image I_{real} and preliminary dehazed result J_{rec}, instructing the DecomNet to generate a more realistic dehazed result.

Weakly-Supervised Fine-tuning. To further improve the visual quality of preliminary dehazed result J_{rec}, we propose A-UNet, embedded with a novel dual DTCWT-based attention block, to enhance the structures and details in the frequency domain. Moreover, adversarial learning is employed to constrain the correlation between the unpaired real-world clean image J_{real} and the restored result J in a weakly supervised manner, making the final dehazed result J of our Dehaze-RetinexGAN closer to the clean data.

Motivation

According to the duality relationship between Retinex and dehazing (Galdran et al. 2018), the image dehazing task can be reduced to the Retinex decomposition on inverted inten-



(a) DTCWT-based channel attention (b) DTCWT-based spatial attention

Figure 3: Overall architecture of our designed dual DTCWT-based attention module, consisting of (a) DTCWT-based channel attention and (b) DTCWT-based spatial attention.

sity of the input hazy image I:

$$\text{Dehazing}\left(I\left(x\right)\right)=1-\text{Retinex}\left(1-I\left(x\right)\right) \tag{1}$$

For simplicity, we assume $S\left(x\right)=1-I\left(x\right)$ and $S(x)$ can be expressed as the pixel-wise multiplication of a reflectance $R(x)$ and an illumination $L(x)$: $S(x)=R(x)L(x)$.

On the other hand, assuming $A=1$, the classic atmospheric scattering model can be rewritten as:

$$1-I\left(x\right)=\left(1-J\left(x\right)\right)T\left(x\right) \tag{2}$$

where $T(x)$ and $J(x)$ respectively denote the medium transmission and reconstructed dehazed result. From the above derivation, the decomposed illumination $L(x)$ and inverted reflectance $1-R\left(x\right)$ produced by DecomNet can be considered as the transmission and the preliminary dehazed result.

Network Architecture

DecomNet Module. We propose the DecomNet module as an illumination-reflectance decomposition network to separate the inverted hazy image S into the desired transmission L and the preliminary dehazed result J_{rec}, simultaneously. Specifically, our constructed DecomNet module first utilizes a 3×3 convolutional layer followed by a LeakyReLU activation function for initial feature extraction, and then applies a 1×1 convolutional layer for channel expansion. Next, the DecomNet includes three repeated blocks, each consisting of a 3×3 convolutional layer, a 1×1 convolutional layer, batch normalization (BN), and a LeakyReLU activation function for further feature extraction and non-linearity. Finally, we adopt a 3×3 convolutional layer for final feature mapping, followed by a sigmoid activation function to produce the output. Due to the difficulty of acquiring paired data for real scenes, we employ the traditional illumination-reflectance decomposition method (Li et al. 2018) to compute $\hat{L}$ and $\hat{R}=S/\hat{L}$, and then the estimated ($\hat{L}$, $\hat{R}$) and the inverted hazy image S are used to form the paired data for training our DecomNet module in a supervised fashion.

Self-Supervised Learning. The preliminary dehazed result J_{rec} produced by our DecomNet appears to lack realism and naturalness due to inaccuracies in the training data. Therefore, we propose a self-supervised learning paradigm to account for the relevance between the dehazed result and the hazy input, thereby constraining the solution of the DecomNet and providing more realistic dehazed result.

Specifically, the classic DCP (He, Sun, and Tang 2011) is utilized to estimate the global atmospheric light A. Then, based on the atmospheric scattering model, we combine A, the transmission L and the preliminary dehazed result J_{rec} to produce the reconstructed hazy image I_{rec}:

$$I_{rec}(x)=J_{rec}(x)L(x)+A(1-L(x)) \tag{3}$$

In order to construct the relationship between the dehazed result and the input hazy image, we first adopt the reconstruction loss L_{rec} to regularize the reconstructed hazy image I_{rec}, the illumination L, and the reflectance R, simultaneously. L_{rec} is defined as:

$$L_{rec}=\|I_{rec}-I_{real}\|_1+\lambda_L\|L-\hat{L}\|_1+\lambda_R\|R-\hat{R}\|_1 \tag{4}$$

where λ_R and λ_L are weights to control different terms.

Furthermore, the perceptual loss L_{per} is introduced to improve the visual quality of preliminary dehazed result:

$$L_{per}=\|\phi(I_{rec})-\phi(I_{real})\|_1 \tag{5}$$

where ϕ refers to the feature maps obtained from specific layers of VGG-19 (Simonyan and Zisserman 2014).

In summary, the overall loss function in the self-supervised pre-training stage is formulated as:

$$L_{pre}=L_{per}+L_{rec}. \tag{6}$$

A-UNet. In order to further improve the perceptual quality of preliminary dehazed result J_{rec}, we design a novel dual DTCWT-based attention module embedded with U-Net architecture as a generator for fine-tuning and leverage the adversarial learning to discriminate the naturalness of the fine-tuned result in a weakly supervised fashion. In Fig. 3, our proposed dual DTCWT-based attention module includes two parts: DTCWT-based channel attention and DTCWT-based spatial attention. For the input features F from the downsampling operations of U-Net, the output features F^C and F^S can be expressed as:

$$F^C=M_C\otimes F,\quad F^S=M_S\otimes F^C \tag{7}$$

where $\otimes$ represents element-wise multiplication. M_C and M_S respectively represent the attention maps of channel and spatial attention. F^C and F^S are the output features of DTCWT-based channel and spatial attention module.

Firstly, F is fed into the DTCWT to obtain six high-frequency components $\{HF_1,\ldots,HF_6\}$ and a low-frequency component LF_1. Then, we sum the six high-frequency parts and upsample them using a transposed convolution operator to match the size of the low-frequency part:

$$HF_s=\text{ConvTranspose}\left(\textstyle\sum_{i=1}^{6}HF_i\right) \tag{8}$$

Then, we further design FNC and FNS blocks to generate the wavelet channel attention map M_C and spatial attention map M_S:

$$M_C=\text{FNC}(LF_1+HF_s), \tag{9}$$

$$M_S=\text{FNS}(Cat(LF_1,HF_s)) \tag{10}$$

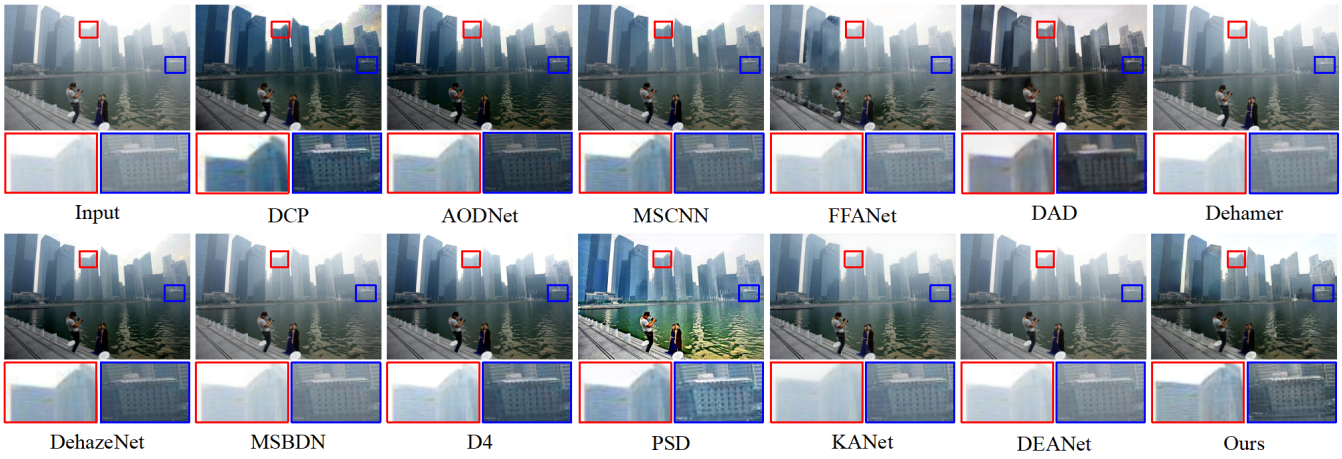


Figure 4: Visual comparisons on a real-world hazy image from RTTS dataset (Li et al. 2019).

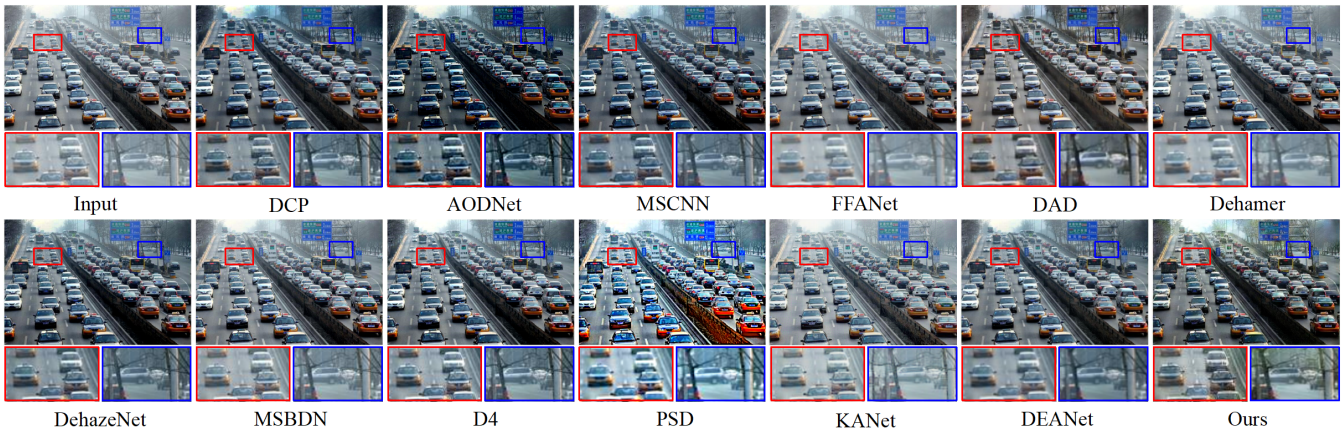


Figure 5: Visual comparisons on a real-world hazy image from URHI dataset (Li et al. 2019).

where Cat denotes concatenation operation. FNS consists of two convolutional layers, with a ReLU activation and a Sigmoid function, while FNC performs an additional average pooling operation.

Compared to previous DWT-based attention (Yang et al. 2023), our proposed dual DTCWT-based attention module possesses strong directional sensitivity ability through six distinct wavelet filters (typically at ±15°, ±45°, and ±75°), effectively capturing structures across different directions and improving details in the frequency domain.

Weakly-Supervised Learning. To fine-tune the A-UNet, we design three losses to improve the quality of the output dehazed result J. Concretely, we first utilize the SSIM loss to constrain the structural similarity between J and J_{rec}:

$$L_{SSIM} = 1 - SSIM(J, J_{rec}), \tag{11}$$

Then, the contrastive regularization loss (Wu et al. 2021) L_{con} is introduced to pull J to the reconstructed dehazed result J_{rec} and push J to its hazy input I_{real}:

$$L_{con} = \rho\left(\phi(I_{real}), \phi(J_{rec}), \phi(J)\right) \tag{12}$$

The exposure loss (Guo et al. 2020) L_{exp} is adopted to eliminate the overexposure of the output dehazed result J.

In addition, the adversarial learning is also introduced to establish the relevance between the output dehazed result J and unpaired real-world clean data J_{real}, which guarantee that the obtained J conforms more closely to the feature distribution of clean data. To achieve this goal, the classic GAN loss (Goodfellow et al. 2014) L_{GAN} is utilized to update the A-UNet and the discriminator D in a weakly supervised manner:

$$\begin{aligned} L_{GAN}(AUNet, D) = &\mathbb{E}_{J_{real} \sim P_{J_{real}}} \left[\log D(J_{real})\right] \\ &+ \mathbb{E}_{J_{rec} \sim P_{J_{rec}}} \left[\log(1 - D(AUNet(J_{rec})))\right] \end{aligned} \tag{13}$$

where $P_{J_{real}}$ and $P_{J_{rec}}$ respectively stand for the set of all possible J_{real} and J_{rec}.

Overall, the joint loss function in the weakly supervised fine-tuning stage is formulated as:

$$\begin{aligned} L_{\text{fine}} = &\lambda_{SSIM} L_{SSIM} + \lambda_{con} L_{con} + \lambda_{exp} L_{exp} \\ &+ \arg \min_{AUNet} \max_{D} \lambda_{GAN} L_{GAN} \end{aligned} \tag{14}$$

where λ_{SSIM}, λ_{con}, λ_{exp}, and λ_{GAN} are trade-off weights.

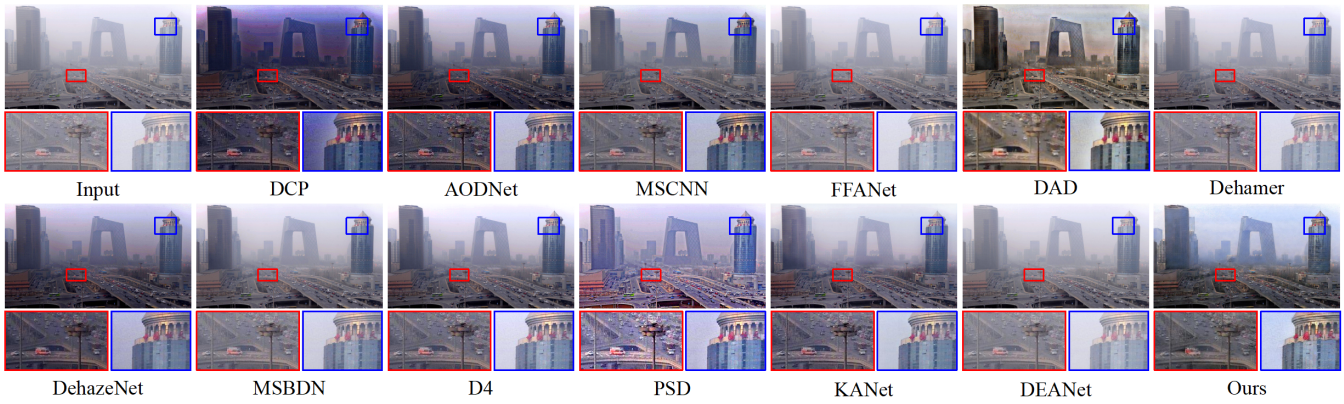


Figure 6: Visual comparisons on a real-world hazy image from HSTS dataset (Li et al. 2019).

Method	Publication	RTTS			URHI		Fattal		HSTS (real-world)	
		BRISQUE↓	NIMA↑	SSEQ↓	BRISQUE↓	NIMA↑	NIMA↑	SSEQ↓	NIMA↑	BRISQUE↓
Hazy image	-	36.6422	3.6389	38.5069	31.6256	3.6237	4.6155	15.3299	2.9921	27.6035
DCP (He, Sun, and Tang 2011)	TPAMI'11	38.7186	3.1589	40.3328	35.6648	3.1950	3.8408	25.9414	2.6676	36.5323
MSCNN (Ren et al. 2016)	ECCV'16	33.0982	3.7045	39.0049	33.3838	3.6505	4.6739	14.9321	3.1166	32.9732
DehazeNet (Cai et al. 2016)	TIP'16	35.0086	3.7163	41.0284	33.2972	3.6364	4.7219	15.0489	3.2102	33.4765
AODNet (Li et al. 2017)	ICCV'17	32.0727	3.8782	35.8573	29.6924	3.9057	4.7676	14.7494	3.1952	31.1289
MSBDN (Dong et al. 2020)	CVPR'20	27.9514	3.7364	35.0615	23.1545	3.7418	4.5042	16.3705	2.5447	40.0361
DAD (Shao et al. 2020)	CVPR'20	32.4602	3.2524	35.9945	29.8898	3.0770	3.7113	30.9603	2.7378	33.6936
FFANet (Qin et al. 2020)	AAAI'20	33.2453	3.7183	37.3002	28.4541	3.6938	4.6693	15.3466	2.9883	28.4745
PSD (Chen et al. 2021)	CVPR'21	27.5818	3.9922	36.0744	32.2545	3.8416	4.5400	15.9418	3.3282	33.2621
RefineDNet (Zhao et al. 2021)	TIP'21	18.9052	3.8594	28.9494	16.8559	3.8420	4.6736	14.2020	3.2044	23.3456
D4 (Yang et al. 2022)	CVPR'22	33.2184	3.7232	39.6694	29.7474	3.7474	4.6958	14.2753	3.1189	30.8643
Dehamer (Guo et al. 2022)	CVPR'22	33.8739	3.7330	38.1866	28.9035	3.7089	4.6249	16.4067	2.9923	28.3641
RIDCP (Wu et al. 2023)	CVPR'23	17.4362	4.4930	21.1056	16.9952	4.4621	4.6113	15.3256	3.2145	35.7810
MB-TaylorFormer (Qiu et al. 2023)	CVPR'23	33.1165	3.7163	37.6693	29.0139	3.6828	4.6597	15.0501	2.5054	40.0805
C2PNet (Zheng et al. 2023)	CVPR'23	34.2688	3.7146	38.0314	31.5900	3.6542	4.6977	15.0369	3.0059	28.6815
Dehazeformer (Song et al. 2023)	TIP'23	32.6476	3.7061	37.9226	29.5789	3.6785	4.6583	15.2638	3.0043	29.0388
DEANet (Chen, He, and Lu 2024)	TIP'24	30.4837	3.7856	37.2266	25.7643	3.7585	4.6826	14.6516	2.9779	28.0282
KANet (Feng et al. 2024)	TPAMI'24	20.2791	3.5711	29.3497	19.1414	3.5511	4.3233	19.2364	3.0486	22.8854
Dehaze-RetinexGAN (Ours)	-	18.2351	4.5313	25.0780	16.9776	4.4534	4.9475	12.3564	3.6763	22.3243

Table 1: Quantitative comparisons of seventeen state-of-the-art methods on four real-world datasets. The top two are marked in red and blue, respectively.

Experiments

Experiment Setup

Datasets. In the first stage of our Dehaze-RetinexGAN, we randomly select 4000 images from the URHI (Unannotated Real-world Hazy Images) dataset (Li et al. 2019) for training. For the fine-tuning stage, we use the above 4000 images and the overall SOTS (Synthetic Objective Testing Set) dataset (Li et al. 2019) for unpaired training. To maintain the performance of our Dehaze-RetinexGAN, the training for the two stages is conducted separately. For testing, we employ four real-world datasets, namely RTTS (Real-world Task-driven Testing Set) dataset (4322 real-world hazy images) (Li et al. 2019), Fattal dataset (31 classic real-world hazy images) (Fattal 2014), HSTS (Hybrid Subjective Testing Set) (10 real-world hazy images) (Li et al. 2019), and URHI test dataset (the remaining 809 real-world hazy images), for quantitative and qualitative comparisons.

Evaluation Metrics. For real-world datasets (RTTS, Fattal, HSTS, and URHI), three non-reference metrics are employed to evaluate the visual quality of the dehazed results, including BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator) (Mittal, Moorthy, and Bovik 2011), NIMA (Neural Image Assessment) (Talebi and Milanfar 2018), and SSEQ (Spatial and Spectral Entropy-based Quality) (Liu et al. 2014).

Implementation Details. Our Dehaze-RetinexGAN is implemented using PyTorch (Paszke et al. 2019) and trained on a single NVIDIA RTX 3090 GPU (24GB) with a batch size of 36. All the input images are resized to 256×256. We utilize the Adam optimizer (Kingma and Ba 2014) with initial momentum $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate is 1×10^{-4}. The cosine annealing algorithm is employed to progressively reduce the learning rate. The hyperparameters λ_{SSIM}, λ_{con}, λ_{exp}, λ_{GAN}, λ_L and λ_R are set to 4.0, 2.0, 0.003, 0.5, 0.2 and 0.2, respectively.

Comparisons with State-of-the-art Methods

We conduct quantitative and qualitative comparisons with seventeen state-of-the-art dehazing methods to demonstrate the superiority of our proposed Dehaze-RetinexGAN.

Quantitative Comparisons. Table 1 reveals the quantita-

Method	Param (M)	Platform	Time (s)
MSCNN (Ren et al. 2016)	-	Matlab	0.3176
DehazeNet (Cai et al. 2016)	-	Matlab	0.1413
MSBDN (Dong et al. 2020)	31.35	Pytorch	0.0182
PSD (Chen et al. 2021)	6.20	Pytorch	0.0327
RefineDNet (Zhao et al. 2021)	65.79	Pytorch	0.0178
D4 (Yang et al. 2022)	10.73	Pytorch	0.0118
Dehamer (Guo et al. 2022)	132.45	Pytorch	0.0594
RIDCP (Wu et al. 2023)	29.48	Pytorch	0.0540
KANet (Feng et al. 2024)	55.66	Pytorch	0.0328
Dehaze-RetinexGAN (Ours)	10.00	Pytorch	0.0093

Table 2: Comparisons with state-of-the-art methods.

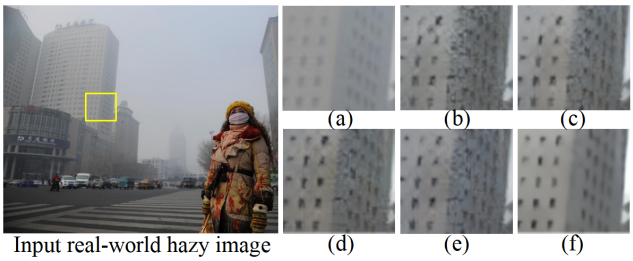


Figure 7: Visual comparisons of ablation study on loss function. (a) an input real-world hazy image. (b) w/o L_{SSIM}. (c) w/o L_{con}. (d) w/o L_{exp}. (e) w/o L_{GAN}. (f) Ours.

tive comparisons with seventeen dehazing methods on four real-world datasets. In Table 1, our Dehaze-RetinexGAN achieves the best NIMA score on the RTTS dataset, with an improvement of 0.85%. For the Fattal dataset, our approach demonstrates an improvement of 3.77% and 13.00% in NIMA and SSEQ metrics, respectively. NIMA and BRISQUE scores increase by 10.46% and 2.45% respectively on the HSTS dataset. In summary, our Dehaze-RetinexGAN shows competitive dehazing performance on four real-world datasets.

Qualitative Comparisons. Figs. 4-6 show the visual comparisons on real-world hazy images. Prior-based dehazing method, namely DCP, tends to produce overly dark dehazed results. Learning-based algorithms trained on synthetic data suffer from limited generalization ability on real-world scenes. Real-world dehazing approaches, such as PSD and KANet, may struggle with over-dehazing and under-dehazing, resulting in unnatural dehazing performance. In comparisons, our Dehaze-RetinexGAN produces the high-quality and natural dehazed results.

Efficiency Comparisons. In Table 2, we show the comparisons of average running times and model parameters. For fairness, all the test images in the RTTS dataset are resized to 256×256 and conducted on the same device to calculate the average running times. Our Dehaze-RetinexGAN provides the optimal execution time with fewer parameters, which achieves a well-balanced trade-off between dehazing performance and computational complexity.

Ablation Study

To verify the effectiveness of our devised each component, the ablation studies are performed on the RTTS dataset.

Variant	BRISQUE↓	NIMA↑
Hazy image	36.6422	3.6389
w/o Self-Supervised	19.7146	4.0684
w/o Spatial Attention	23.1729	4.2347
w/o Channel Attention	25.0785	4.1480
w/o Dual Attention	24.6284	4.0757
w/o SSIM Loss (L_{SSIM})	22.8872	4.4395
w/o Contrastive Loss (L_{con})	21.7813	4.4050
w/o GAN Loss (L_{GAN})	23.9156	4.2514
w/o Exposure Loss (L_{exp})	18.7454	4.4576
Dehaze-RetinexGAN (Ours)	**18.2351**	**4.5313**

Table 3: Ablation studies on RTTS dataset (Li et al. 2019).

Ablation on Self-Supervised Learning. We remove the self-supervised learning paradigm from the first pre-training stage to demonstrate its contribution. Table 3 illustrates the effectiveness of our self-supervised learning strategy.

Ablation on Dual DTCWT-based Attention Module. Our proposed attention module can effectively improve the structures and details of dehazed result. To prove its contribution, we design three settings to conduct the ablation study: (i) without channel attention in the dual DTCWT-based attention module; (ii) without spatial attention; (iii) without dual DTCWT-based attention module. Table 3 demonstrates the highest BRISQUE and NIMA scores of our devised dual attention module.

Ablation on Loss Function. To demonstrate the effectiveness of each loss, the visual results and quantitative comparisons are revealed in Fig. 7 and Table 3, respectively. From Fig. 7, the absence of L_{SSIM} may be prone to poor structural preservation, while the lack of L_{con} or L_{GAN} can result in unnatural dehazing performance. The result without using L_{exp} exhibits unbalanced brightness and contrast. Table 3 also verifies the contribution of each loss function.

Conclusion

In this paper, we propose Dehaze-RetinexGAN, consisting of self-supervised pre-training and weakly-supervised fine-tuning, for real-world image dehazing. First, we reformulate the image dehazing task as an illumination-reflectance decomposition problem and then design the DecomNet to achieve Retinex decomposition. In addition, we develop a self-supervised learning paradigm to construct the connection between the preliminary dehazed result and the input hazy image, improving the generalization performance of the DecomNet. In the fine-tuning stage, a dual DTCWT-based attention module is presented to further improve the structures and details of the dehazed result obtained from the first stage. Furthermore, we employ adversarial learning to enforce the relevance between the fine-tuned result and the clean data in a weakly supervised manner. Experiments on multiple real-world datasets prove the effectiveness of our Dehaze-RetinexGAN.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant No. 62301453) and in part by the scholarship from China Scholarship Council (CSC) (No. 202206995010).

References

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(11): 5187–5198.

Chen, E.; Chen, S.; Ye, T.; and Liu, Y. 2024. Degradation-adaptive neural network for jointly single image dehazing and desnowing. *Frontiers of Computer Science*, 18(2).

Chen, X.; Li, Y.; Kong, C.; and Dai, L. 2022. Unpaired Image Dehazing With Physical-Guided Restoration and Depth-Guided Refinement. *IEEE Signal Processing Letters*, 29: 587–591.

Chen, Z.; He, Z.; and Lu, Z.-M. 2024. DEA-Net: Single Image Dehazing Based on Detail-Enhanced Convolution and Content-Guided Attention. *IEEE Transactions on Image Processing*, 33: 1002–1015.

Chen, Z.; Wang, Y.; Yang, Y.; and Liu, D. 2021. PSD: Principled synthetic-to-real dehazing guided by physical priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7180–7189.

Dong, H.; Pan, J.; Xiang, L.; Hu, Z.; Zhang, X.; Wang, F.; and Yang, M.-H. 2020. Multi-Scale Boosted Dehazing Network With Dense Feature Fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2157–2167.

Engin, D.; Genc, A.; and Kemal Ekenel, H. 2018. Cycle-Dehaze: Enhanced CycleGAN for Single Image Dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 825–833.

Fattal, R. 2014. Dehazing Using Color-Lines. *ACM Transactions on Graphics (TOG)*, 34(1): 1–14.

Feng, Y.; Ma, L.; Meng, X.; Zhou, F.; Liu, R.; and Su, Z. 2024. Advancing Real-World Image Dehazing: Perspective, Modules, and Training. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12): 9303–9320.

Galdran, A.; Alvarez-Gila, A.; Bria, A.; Vazquez-Corral, J.; and Bertalmío, M. 2018. On the Duality Between Retinex and Image Dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 8212–8221.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Nets. In Ghahramani, Z.; Welling, M.; Cortes, C.; Lawrence, N.; and Weinberger, K., eds., *Advances in Neural Information Processing Systems*, volume 27.

Guo, C.; Li, C.; Guo, J.; Loy, C. C.; Hou, J.; Kwong, S.; and Cong, R. 2020. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1780–1789.

Guo, C.-L.; Yan, Q.; Anwar, S.; Cong, R.; Ren, W.; and Li, C. 2022. Image Dehazing Transformer With Transmission-Aware 3D Position Embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5812–5820.

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(12): 2341–2353.

Jin, Y.; Yang, W.; and Tan, R. T. 2022. Unsupervised night image enhancement: When layer decomposition meets light-effects suppression. In *European Conference on Computer Vision*, 404–421.

Ju, M.; Ding, C.; Guo, C. A.; Ren, W.; and Tao, D. 2021. IDRLP: Image Dehazing Using Region Line Prior. *IEEE Transactions on Image Processing*, 30: 9043–9057.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Land, E. H. 1977. The retinex theory of color vision. *Scientific american*, 237(6): 108–129.

Li, B.; Peng, X.; Wang, Z.; Xu, J.; and Feng, D. 2017. AOD-Net: All-In-One Dehazing Network. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 4770–4778.

Li, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; and Wang, Z. 2019. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1): 492–505.

Li, C.; Hu, E.; Zhang, X.; Zhou, H.; Xiong, H.; and Liu, Y. 2024. Visibility restoration for real-world hazy images via improved physical model and Gaussian total variation. *Frontiers of Computer Science*, 18(1): 181708.

Li, M.; Liu, J.; Yang, W.; Sun, X.; and Guo, Z. 2018. Structure-Revealing Low-Light Image Enhancement Via Robust Retinex Model. *IEEE Transactions on Image Processing*, 27(6): 2828–2841.

Ling, P.; Chen, H.; Tan, X.; Jin, Y.; and Chen, E. 2023. Single Image Dehazing Using Saturation Line Prior. *IEEE Transactions on Image Processing*, 32: 3238–3253.

Liu, L.; Liu, B.; Huang, H.; and Bovik, A. C. 2014. No-reference image quality assessment based on spatial and spectral entropies. *Signal Processing: Image Communication*, 29(8): 856–863.

Liu, Y.; Yan, Z.; Chen, S.; Ye, T.; Ren, W.; and Chen, E. 2023a. Nighthazeformer: Single nighttime haze removal using prior query transformer. In *Proceedings of the 31st ACM International Conference on Multimedia*, 4119–4128.

Liu, Y.; Yan, Z.; Tan, J.; and Li, Y. 2023b. Multi-Purpose Oriented Single Nighttime Image Haze Removal Based on Unified Variational Retinex Model. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(4): 1643–1657.

Mittal, A.; Moorthy, A. K.; and Bovik, A. C. 2011. Blind/Referenceless Image Spatial Quality Evaluator. In *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 723–727.

Narasimhan, S. G.; and Nayar, S. K. 2002. Vision and the atmosphere. *International journal of computer vision*, 48: 233–254.

Noori, H.; Gholizadeh, M. H.; and Rafsanjani, H. K. 2024. Digital image defogging using joint Retinex theory and independent component analysis. *Computer Vision and Image Understanding*, 245: 104033.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32.

Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; and Jia, H. 2020. FFA-Net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 11908–11915.

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. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 12802–12813.

Ren, W.; Liu, S.; Zhang, H.; Pan, J.; Cao, X.; and Yang, M.-H. 2016. Single Image Dehazing via Multi-scale Convolutional Neural Networks. In *Computer Vision–ECCV 2016*, 154–169.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015*, 234–241.

Shao, Y.; Li, L.; Ren, W.; Gao, C.; and Sang, N. 2020. Domain Adaptation for Image Dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2808–2817.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Song, Y.; He, Z.; Qian, H.; and Du, X. 2023. Vision Transformers for Single Image Dehazing. *IEEE Transactions on Image Processing*, 32: 1927–1941.

Talebi, H.; and Milanfar, P. 2018. NIMA: Neural Image Assessment. *IEEE Transactions on Image Processing*, 27(8): 3998–4011.

Wu, H.; Qu, Y.; Lin, S.; Zhou, J.; Qiao, R.; Zhang, Z.; Xie, Y.; and Ma, L. 2021. Contrastive Learning for Compact Single Image Dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10551–10560.

Wu, R.-Q.; Duan, Z.-P.; Guo, C.-L.; Chai, Z.; and Li, C. 2023. RIDCP: Revitalizing Real Image Dehazing via High-Quality Codebook Priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 22282–22291.

Xu, Z.; Liu, X.; and Chen, X. 2009. Fog Removal from Video Sequences Using Contrast Limited Adaptive Histogram Equalization. In *2009 International Conference on Computational Intelligence and Software Engineering*, 1–4.

Yang, Y.; Jiao, L.; Liu, X.; Liu, F.; Yang, S.; Li, L.; Chen, P.; Li, X.; and Huang, Z. 2023. Dual wavelet attention networks for image classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(4): 1899–1910.

Yang, Y.; Wang, C.; Liu, R.; Zhang, L.; Guo, X.; and Tao, D. 2022. Self-Augmented Unpaired Image Dehazing via Density and Depth Decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2037–2046.

Ye, T.; Chen, S.; Bai, J.; Shi, J.; Xue, C.; Jiang, J.; Yin, J.; Chen, E.; and Liu, Y. 2023. Adverse weather removal with codebook priors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12653–12664.

Zhang, H.; Liu, X.; Huang, Z.; and Ji, Y. 2014. Single image dehazing based on fast wavelet transform with weighted image fusion. In *2014 IEEE International Conference on Image Processing (ICIP)*, 4542–4546.

Zhao, S.; Zhang, L.; Shen, Y.; and Zhou, Y. 2021. RefineDNet: A Weakly Supervised Refinement Framework for Single Image Dehazing. *IEEE Transactions on Image Processing*, 30: 3391–3404.

Zheng, Y.; Zhan, J.; He, S.; Dong, J.; and Du, Y. 2023. Curricular Contrastive Regularization for Physics-Aware Single Image Dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5785–5794.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.