Robust Single Image Dehazing Based on Consistent and Contrast-Assisted Reconstruction

De Cheng[1†], Yan Li[2†], Dingwen Zhang[3,4], Nannan Wang[1*], Xinbo Gao[5], Jiande Sun[2*]
[1]Xidian University
[2]Shandong Normal University
[3]Northwestern Polytechnical University
[4]Institute of Artificial Intelligence, Hefei Comprehensive National Science Center
[5]Chongqing University of Posts and Telecommunications
{dcheng, nnwang}@xidian.edu.cn, {liyan.ly, jiandesun}@hotmail.com,
zhangdingwen2006yyy@gmail.com, gaoxb@cqupt.edu.cn

Abstract

Single image dehazing as a fundamental low-level vision task, is essential for the development of robust intelligent surveillance system. In this paper, we make an early effort to consider dehazing robustness under variational haze density, which is a realistic while under-studied problem in the research filed of singe image dehazing. To properly address this problem, we propose a novel density-variational learning framework to improve the robustness of the image dehzing model assisted by a variety of negative hazy images, to better deal with various complex hazy scenarios. Specifically, the dehazing network is optimized under the consistency-regularized framework with the proposed Contrast-Assisted Reconstruction Loss (CARL). The CARL can fully exploit the negative information to facilitate the traditional positive-orient dehazing objective function, by squeezing the dehazed image to its clean target from different directions. Meanwhile, the consistency regularization keeps consistent outputs given multi-level hazy images, thus improving the model robustness. Extensive experimental results on two synthetic and three real-world datasets demonstrate that our method significantly surpasses the state-of-the-art approaches.

1 Introduction

Image dehazing aims to recover the clean image from a hazy input, which is essential for the development of robust computer vision systems. It helps to mitigate the side-effect of image distortion induced by the environmental conditions, on many visual analysis tasks, such as object detection [Li *et al.*, 2018a; Zhang *et al.*, 2021a] and scene understanding [Sakaridis *et al.*, 2018; Zhang *et al.*, 2021b]. Therefore, single image dehazing has attracted more and more attention, and many dehazing methods have been proposed recently [Qin *et al.*, 2020; Wu *et al.*, 2021].

Figure 1: The results of two dehazing methods on the same image with different hazy densities. The first row shows the input hazy images with three different densities. The second and third rows show the corresponding dehazed images by current two representative methods (denoted as "Dehazing Model 1" and "Dehazing Model 2"), respectively. It shows that such methods are lack of robustness to cope with hazy images in different scenarios.

Great efforts have been made in the past few years and end-to-end deep learning based dehazing methods has achieved great success in dealing with complex scenes [Dong *et al.*, 2020a; Li *et al.*, 2017; Hong *et al.*, 2020; Shao *et al.*, 2020]. However, when performing on scenarios with different haze density, these methods still cannot always obtain desirable dehzing performance witnessed by the inconsistency results as shown in Figure 1. We can clear see that the same image with different hazy densities usually generate dehazed images with different qualities, by some current designed dehazing models. This phenomenon illustrates that such image dehazing models are not robust to some complex hazy scenarios, which is not what we expected for a good dehazing model. Unfortunately, this situation may usually happen in real world. Consequently, how to improve the robustness of dehazing model

†Equal contribution.
*Corresponding author.

becomes an important yet under-studied issue.

The inconsistency results shown in Figure 1 also inspire us to regularize the learning process by utilizing these multi-level hazy images with different densities, to improve the model robustness. To achieve this goal, we creatively propose a Contrast-Assisted Reconstruction Loss (CARL) under the consistency-regularized framework for single image dehazing. It aims to train a more robust dehazing model to better deal with various hazy scenarios, including but not limited to multi-level hazy images with different densities.

In the proposed CARL, we fully exploit the negative information to better facilitate the traditional positive-orient dehazing objective function. Specifically, we denote the ground-truth clear image and the restored dehazed image as the anchor and positive example, respectively. The negative examples can be constructed not only from the original hazy image and its variants with different hazy densities, but also from the restored images generated by some other dehazing models. The CARL enables the network prediction to be close to the clear image, while push it far away from the hazy images in the feature space. Specifically, pushing anchor point far away from various negative images seems to squeeze the anchor point to its positive example from different directions. Elaborately selecting the negative examples can help the CARL to improve the lower bound for approximating to its clear image under the regularization of various negative hazy images. Besides, more negative examples in the contrastive loss can usually contribute more performance improvement, which has been demonstrated in metric learning research field [Khosla *et al.*, 2020]. Therefore, we also try to adopt more negative hazy images to improve the model capability to cope with various hazy scenarios.

To further improve the model robustness, we propose to train the dehazing network under the consistency-regularized framework on top of the CARL. The success of consistency regularization lies in the assumption that the dehazing model should output very similar or even same dehazed images when fed the same hazy image with different densities. Such constraint meets the requirement of a good dehazing model to deal with hazy images in different hazy scenarios. Specifically, we implement this consistency constraint by the mean-teacher framework [Tarvainen and Valpola, 2017]. For each input hazy image, we also resort to previously constructed images with different hazy densities or some other informative negative examples. Then, L1 loss is used to minimize the discrepancy among all the dehazed images which corresponds to one clear target. The consistency regularization significantly improves the model robustness and performance of the dehazing network, and it can be easily extended to any other dehazing network architectures. The main contributions are summarized as follows:

- We make an early effort to consider dehazing robustness under variational haze density, which is a realistic while under-studied problem in the research filed of singe image dehazing.
- We propose a contrast-assisted reconstruction loss under the consistency-regularized framework for single image dehazing. This method can fully exploit various negative hazy images, to improve the dehzing model capability of dealing with various complex hazy scenarios.
- Extensive experimental results on two synthetic and three real-world image dehazing datasets demonstrate that the proposed method significantly surpasses the state-of-the-art algorithms.

2 Related Work

We can roughly divide existing dehazing methods into categories: the physical-scattering-model dependent methods and the model-free methods.

The physical-scattering-model dependent methods try to recover the clear image through estimating the atmospheric light and transmission map by some specially designed priors or network architectures. For the prior-based image dehazing methods, they usually remove the haze using different statistic image prior from empirical observations, such as the traditional dark channel prior (DCP) [He *et al.*, 2010], non-local prior [Berman *et al.*, 2016] and contrast maximization [Tan, 2008]. Although these methods have achieved great successes, the priors can not handle all the cases in the unconstraint wild environment.

Recently, as the prevailing success of deep learning in image processing tasks, many deep dehazing methods depending on the atmosphere scattering model have been proposed. [Zhang and Patel, 2018] directly embedded the physical model into the dehazing network by a densely connected encoder-decoder structure. [Ren *et al.*, 2016] proposed a coarse-to-fine multi-scale CNN model to estimate the transmission map. [Li *et al.*, 2017] reformulated the original atmospheric scattering model and jointly estimated the global atmospheric light and the transmission map. However, such physical-scattering-model based learning methods may produce accumulative error and degrade the dehazing results, since some relatively small inaccurate or biased estimation of the transmission map and global atmospheric light could lead to larger reconstruction error between the dehazed and clear images.

Model-free deep dehazing methods try to directly learn the mapping between the hazy input image and clear result without using atmospheric scattering model. Most of such methods focus on strengthening the dehazing network. For instance, [Liu *et al.*, 2019b] designed a dual residual neural network architecture to explore the potential of paired operations for image restoration tasks. [Qu *et al.*, 2019] proposed a pixel-to-pixel dehazing network to obtain perceptual pleasing results. [Qin *et al.*, 2020] proposed a fused attention mechanism to strengthen the flexibility of the network to deal with various information. These methods only minimize the reconstruction loss between the dehazed image and its clear target, without any regularization on images or features.

Recently, there also appeared some methods, which adopted distance metric regularization to further improve the reconstruction loss. [Wu *et al.*, 2021] proposed the divided-contrast regularization for single image dehazing. Our method also falls into this category, but it is very different from them. We specifically propose the CARL under the consistency regularization framework for single image dehazing. It can not only fully exploit existing negative hazy exam-

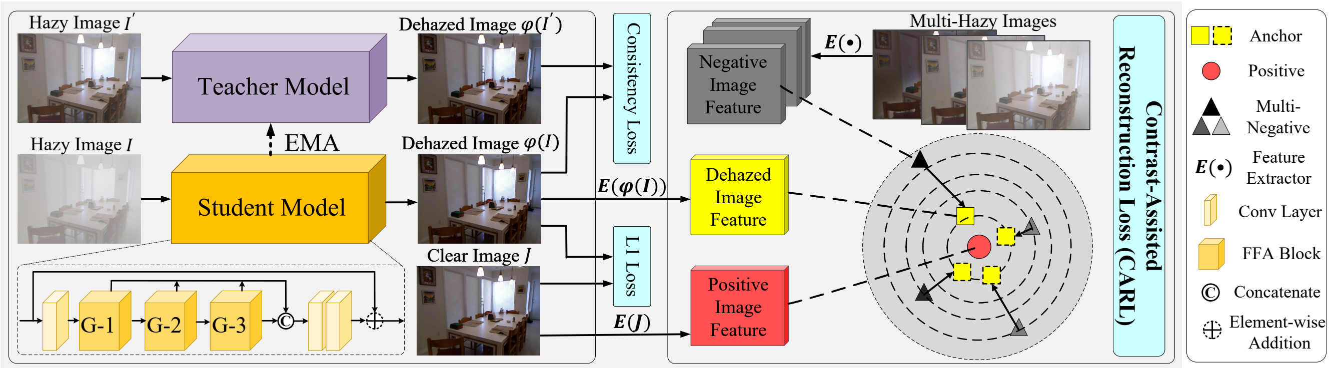

Figure 2: The framework of the proposed method for robust single image dehazing. It consists of the same dehazing network architecture for the student and teacher network, the Contrast-Assisted Reconstruction Loss (CARL) to fully exploit various negative hazy images to squeeze the dehazed image to its clean target from different directions, the consistency regularization framework to explicitly deal with multi-level hazy scenarios to further improve its robustness, and the traditional L1 reconstruction loss.

ples to assist the dehazing model to produce more natural and sharper dehazed images, but also can improve the model robustness to deal with various complex hazy scenarios.

3 Proposed Method

The proposed method makes great efforts to deal with various hazy scenarios from the following two aspects: 1) The CARL improves the traditional positive-orient dehazing objective function, by fully exploiting the various negative information to squeeze the dehazed image to its clean target from different directions; 2) The consistency regularization aims to further improve the model robustness by explicitly utilizing the constraint outputs of multi-level hazy images. In the following, we describe the algorithm in detail shown in Figure 2.

3.1 Dehazing Network Architecture

In this paper, we adopt the previously proposed FFA-Net [Qin *et al.*, 2020] as our backbone network architecture. As shown in Figure 2, the student and teacher network share the same network architecture (FFA-Net), which includes the following components: the shallow feature extraction part, several group attention architecture (Denoted as G-x), the feature concatenation module, the reconstruction part and global residual skip connection.

3.2 Contrast-Assisted Reconstruction Loss

The contrastive learning method has achieved a series of successes in representation learning, it aims to learn discriminative representations by pulling "positive" pairs close, while pushing "negative" pairs far apart. Inspired by this, we propose the "Contrast-Assisted" Reconstruction Loss (CARL) to improve the traditional positive-orient dehazing methods, by fully exploiting various negative information to squeeze the dehazed image to its clean target from different directions.

To imitate the traditional contrastive learning, there are two aspects we need to consider: one is how to obtain useful positive and negative training examples, the other is how to apply the CARL in the dehazing framework [Wu *et al.*, 2021]. As we known, elaborately constructing efficient positive and negative training examples is very crucial to better optimize the CARL. For the image dehazing task, obviously the positive pair is the dehazed image and its corresponding clear one, which can be denoted as anchor point and positive example. Our final goal is just to minimize the discrepancy between them. Meanwhile, pushing anchor point far away from several negative examples is to squeeze the anchor point to positive example from different directions, as illustrated in Figure 2. Therefore, we generate negative examples from several aspects, which includes the original hazy image, multi-level hazy images with different densities, some relatively low-quality dehazed images by previous model, and some other variants of the input hazy images. For the latent space to apply the CARL, we adopt the the fixed pre-trained model (e.g. VGG-19 [Simonyan and Zisserman, 2014]) as the feature extractor "E" and then extract the image features from different intermediate network layers, which was also used in [Wu *et al.*, 2021; Johnson *et al.*, 2016].

Denote the input hazy image as I, its corresponding dehazed image as $\phi(I)$ which is generated by the dehazing network ϕ, and the ground-truth hazy-free image as J. The selected negative images corresponding to $\phi(I)$ denote as $I_i^-, i \in \{1, \cdots, K\}$, K is the number of negative examples used in the CARL. We define the features extracted by the fixed pre-trained VGG model as $E(J)$, $E(\phi(I))$ and $E(I_i^-)$ for the positive, anchor and negative examples, respectively. Then, the m-th CARL function $\mathcal{L}^m_{CARL}$ can be formulated as:

$$-\log \frac{\mathrm{e}^{-|E_m(\phi(I)) - E_m(J)|/\tau}}{\mathrm{e}^{-|E_m(\phi(I)) - E_m(J)|/\tau} + \sum_{i=1}^{K} \mathrm{e}^{-|E_m(\phi(I)) - E_m(I_i^-)|/\tau}}. \tag{1}$$

In Eq. 1, "e" denotes the exponential operation, $E_m, m = \{1, 2, \cdots, M\}$, extracts the image features from the m-th intermediate network layer of the fixed pre-trained model VGG [Simonyan and Zisserman, 2014]. $|\cdot|$ demotes the $L1$ distance, which usually achieves better performance compared to $L2$ distance for image dehazing task. $\tau > 0$ is the temperature parameter that controls the sharpness of the output.

Therefore, the final CARL can be expressed as follows,

$$\mathcal{L}_{CARL} = \sum_{m=1}^{M} \omega_m \mathcal{L}_{CARL}^{m}, \tag{2}$$

where ω_m is the weight coefficient for the m-th CARL using the intermediate feature generated by the fixed VGG model.

Note that, in Eq. 1, the positive point $E_m(J)$ and all the negative points $E_m(I_i^-)$ are constant values, minimizing $\mathcal{L}_{CARL}$ can optimize the parameters of the dehazing network ϕ through the dehazed image features $E_m(\phi(I))$. Related to our CARL, perceptual loss [Johnson *et al.*, 2016] minimizes the reconstruction error between the restored and the ground-truth images by using multiple intermediate features extracted from the fixed pre-trained deep model. On top of this, one divided-contrastive learning method [Wu *et al.*, 2021] adopted the original hazy image as negative image to regularize the solution space. Different from above methods, the proposed CARL method aims to minimize the reconstruction error between the restored image and its corresponding clear ground-truth, as well as pushing the restored dehazed image far away from various negative hazy examples, which acts as a way to squeeze prediction to its ground truth from different directions in the constraint learning space. Thus, it enables the dehazing model to deal with various complex hazy scenarios.

The main difference between the traditional contrastive learning and our proposed CARL is that: The traditional contrastive learning aims to learn discriminative feature representations to distinguish instances from different classes or identities, which cares about both the inter-class discrepancy and intra-class compactness. However, for such image dehazing task, we just consider the reconstruction error between the dehazed image and its corresponding clean target. Therefore, our final goal of CARL is to better optimize the reconstruction loss with the help of various negative hazy examples in the contrastive manner. This is also the reason why we call the proposed method "Contrast-Assisted Reconstruction loss".

3.3 The Consistency-Regularized Framework

We propose the consistency regularization based on the assumption that a good dehazing model should output very similar or even same dehazed images when fed the same hazy image with different densities. Such explicit constraint further improves the model robustness to deal with multi-level hazy images. This learning paradigm is implemented by training a student neural network $\phi_s(\cdot)$ and a teacher neural network $\phi_t(\cdot)$, which share the same network architecture, but are parameterized by θ_s and θ_t respectively.

Specifically, we first construct hazy images with different densities using the physical-scattering model. Usually, the synthetic dehazing datasets themselves contain several hazy images with different densities corresponding to one clean image. For the real-world dehazing dataset, we generate different hazy images with the help of the transmission map and atmospheric light given for other dataset, and some relatively poor-quality dehazed image obtained by previous dehazing model, such as DCP [He *et al.*, 2010].

Given two different hazy images denoted as I and I', which correspond to the same clear image J, they are fed into the student and teacher network respectively. These two hazy images are processed by dehazing network $\phi_s(\cdot)$ and $\phi_t(\cdot)$ to obtain the dehazed images $\phi_s(I)$ and $\phi_t(I')$. The consistency regularization $\mathcal{L}_{CR}$ can be expressed as follows,

$$\mathcal{L}_{CR} = |\phi_s(I) - \phi_t(I')|. \tag{3}$$

In Eq. 3, we use the L_1 loss to implement the consistency regularization. Minimizing the loss function $\mathcal{L}_{CR}$ can directly optimize parameters of the student network $\phi_s(\cdot)$, while the parameters of the teacher network is updated by the exponential moving average (EMA) techniques, which is based on the previous teacher network and current student network parameters. Unlike previous teacher network for image dehazing [Hong *et al.*, 2020], we do not have a predefined high quality model as the fixed teacher model, we build it from past iteration of the student network $\phi_s(\cdot)$. The updating rule of "EMA" is $\theta_t \leftarrow \lambda\theta_t + (1-\lambda)\theta_s$, and λ is a smoothing hyper-parameter to control the model updating strategy [Tarvainen and Valpola, 2017].

Therefore, such a consistency regularization keeps hazy image with different densities or under different scenarios, have the same haze-free output, which greatly improves the robustness of the image dehazing model.

3.4 The Overall Loss Function

Apart from above introduced CARL and the consistency regularization, we also adopt the traditional reconstruction loss $\mathcal{L}_1$ between the prediction $\phi(I)$ and its corresponding ground truth J in the data field. It can be implemented by the $L1$ loss as follows,

$$\mathcal{L}_1 = |\phi_s(I) - J|. \tag{4}$$

Therefore, the overall objective function $\mathcal{L}$ can be expressed as,

$$\mathcal{L} = \mathcal{L}_1 + \lambda_1\mathcal{L}_{CR} + \lambda_2\mathcal{L}_{CARL}, \tag{5}$$

where λ_1 and λ_2 are two hyper-parameters to balance above three terms in the overall loss function.

4 Experiments

4.1 Experiment Setup

Datasets and Metrics. To comprehensively evaluate our proposed method, we conduct extensive experiments on two representative synthetic datasets and three challenging real-world datasets. The widely used synthetic dataset: RESIDE[Li *et al.*, 2018b], contain two subsets, i.e., Indoor Training Set (ITS), and Outdoor Training Set (OTS). ITS and OTS are used for training, and they have corresponding testing dataset, namely, Synthetic Objective Testing Set (SOTS), which consists of 500 indoor hazy images (SOTS-Indoor) and 500 outdoor hazy images (SOTS-Outdoor). We also evaluate the proposed model on four popular real-world datasets: NTIRE 2018 image dehazing indoor dataset (referred to as I-Haze) [Ancuti *et al.*, 2018b], NTIRE 2018 image dehazing outdoor dataset (O-Haze) [Ancuti *et al.*, 2018a], and NTIRE 2019 dense image dehazing dataset (Dense-Haze) [Ancuti *et al.*, 2019]. We adopt two commonly used evaluation metrics, i.e., the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity index (SSIM), for the objective measurement.

Method	SOTS-Indoor		SOTS-Outdoor	
	PSNR	SSIM	PSNR	SSIM
DCP [He *et al.*, 2010]	16.62	0.8179	19.13	0.8148
MSCNN [Ren *et al.*, 2016]	17.57	0.8102	20.73	0.8187
DehazeNet [Cai *et al.*, 2016]	21.14	0.8472	22.46	0.8514
AOD-Net [Li *et al.*, 2017]	19.06	0.8504	20.29	0.8765
DCPDN [Zhang and Patel, 2018]	19.00	0.8400	19.71	0.8300
GFN [Ren *et al.*, 2018]	22.30	0.8800	21.55	0.8444
EPDN [Qu *et al.*, 2019]	25.06	0.9232	22.57	0.8630
DuRN-US [Liu *et al.*, 2019b]	32.12	0.9800	19.41	0.8100
Grid-Net [Liu *et al.*, 2019a]	32.16	0.9836	30.86	0.9819
KDDN [Hong *et al.*, 2020]	34.72	0.9845	–	–
DA [Shao *et al.*, 2020]	25.30	0.9420	26.44	0.9597
MSBDN [Dong *et al.*, 2020a]	32.00	0.9860	30.77	0.9550
FFA-Net [Qin *et al.*, 2020]	36.39	0.9886	<u>32.09</u>	0.9801
FD-GAN [Dong *et al.*, 2020b]	23.15	0.9207	–	–
DRN [Li *et al.*, 2020]	32.41	0.9850	31.17	<u>0.9830</u>
IDRLP [Ju *et al.*, 2021]	23.56	0.9383	–	–
AECR-Net [Wu *et al.*, 2021]	37.17	<u>0.9901</u>	–	–
DIDH [Yoon *et al.*, 2021]	<u>38.91</u>	0.9800	30.40	0.9400
Ours	**41.92**	**0.9954**	**33.26**	**0.9849**

Table 1: Quantitative comparisons with SOTA methods on SOTS-Indoor and SOTS-Outdoor synthetic datasets.

Implementation Details. We implement our method based on PyTorch with NVIDIA RTX 2080Ti GPUs. In the training process, we randomly crop 240×240 image patches as input and adopt Adam optimizer for optimization. The learning rate is initially set to 1×10^{-4} and is adjusted using the cosine annealing strategy [He *et al.*, 2019]. We follow [Wu *et al.*, 2021] to select the features of the 1st, 3rd, 5th, 9th and 13th layers from the fixed pre-trained VGG-19 [Simonyan and Zisserman, 2014] to calculate the L1 distance in Eq.(2), and their corresponding weight factors ω_m are set as $\frac{1}{32}$, $\frac{1}{16}$, $\frac{1}{8}$, $\frac{1}{4}$ and 1, respectively. The number of negative examples used in Eq. 1 is set to $K = 5$, and parameter $\tau = 0.5$. The hyper-parameters λ_1 and λ_2 in Eq. 5 is set to 1.0 and 10.0, respectively.

4.2 Comparisons with State-of-the-art Methods

Results on Synthetic Datasets. We follow the settings of [Qin *et al.*, 2020] to evaluate our proposed method on two representative synthetic datasets, and compare it with seventeen state-of-the-art (SOTA) methods. The results are shown in Table 1. We can see that our proposed method achieves the best performance on both SOTS-Indoor and SOTS-Outdoor datasets. On SOTS-Indoor, our proposed method achieves 41.92dB PSNR and 0.9954 SSIM, surpassing the second-best 3.01dB PSNR and 0.0053 SSIM, respectively. On the SOTS-Outdoor, our proposed method achieves the gain with 1.17dB PSNR and 0.0019 SSIM, compared with the second-best method. In Figure 3, we show the visual effects of representative methods of dehazed images for subjective comparison. We can observe that DCP suffers from the color distortion, where the images are darker and unrealistic. MSCNN, DehazeNet and AOD-Net cannot remove haze completely, and there are still a lot of haze residues. Although Grid-Net and FFA-Net can achieve better dehazed effect, there is still a gap between their results and the ground truth. In contrast, the images generated by our proposed method are closer to ground truth and more natural.

Method	I-Haze		O-Haze		Dense-Haze	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DCP	14.43	0.7520	16.78	0.6530	10.06	0.3856
MSCNN	15.22	0.7550	17.56	0.6500	11.57	0.3959
DehazeNet	14.31	0.7220	16.29	0.6400	13.84	0.4252
AOD-Net	13.98	0.7320	15.03	0.5390	13.14	0.4144
Grid-Net	16.62	0.7870	18.92	0.6720	13.31	0.3681
KDDN	–	–	<u>25.46</u>	<u>0.7800</u>	14.28	0.4074
FFA-Net	–	–	–	–	14.39	0.4524
MSBDN	<u>23.93</u>	**0.8910**	24.36	0.7490	15.37	<u>0.4858</u>
IDRLP	17.36	0.7896	16.95	0.6990	–	–
AECR-Net	–	–	–	–	**15.80**	0.4660
Ours	**25.43**	<u>0.8807</u>	**25.83**	**0.8078**	<u>15.47</u>	**0.5482**

Table 2: Quantitative comparisons with SOTA methods on I-Haze, O-Haze and Dense-Haze real-world datasets.

Method	SOTS-Indoor		Dense-Haze	
	PSNR	SSIM	PSNR	SSIM
$\mathcal{L}_1$	36.39	0.9886	14.39	0.4524
$\mathcal{L}_1$+$\mathcal{L}_{DivC}$	37.54	0.9915	14.82	0.5354
$\mathcal{L}_1$+$\mathcal{L}_{CARL}$	39.56	0.9939	15.30	0.5438
$\mathcal{L}_1$+$\mathcal{L}_{CARL}$ + $\mathcal{L}_{CR}$	**41.92**	**0.9954**	**15.47**	**0.5482**

Table 3: Ablation study on SOTS-Indoor and Dense-Haze datasets.

Results on Real-world Datasets. We also evaluate our proposed method on three challenging real-world datasets. As shown in Table 2, our proposed method outperforms most state-of-the-art methods, and we have obtained 25.43dB PSNR and 0.8807 SSIM on I-Haze dataset, 25.83dB PSNR and 0.8078 SSIM on O-Haze dataset, 14.47db PSNR and 54.82 SSIM on Dense-Haze dataset. On I-Haze and Dense-Haze datasets, although our proposed method is slightly lower than the state-of-the-art methods in SSIM and PSNR, but higher in terms of PSNR and SSIM, respectively. In general, our proposed method can still maintain advanced results on real-world datasets, and some visual comparison presented in Figure 3 also verifies it.

4.3 Ablation Study

To reveal the effectiveness of each ingredient, ablation study are conducted to analyze different items in the framework, including the $\mathcal{L}_1$, $\mathcal{L}_{CR}$ and $\mathcal{L}_{CARL}$, on both synthetic and real-world hazy image datasets.

We conduct all the experiments based on the same dehazing network architecture (FFA-Net [Qin *et al.*, 2020]). We implement the following four variants of the proposed method: 1) $\mathcal{L}_1$: Training the network by the traditional L_1 loss function, which works as the baseline method; 2) $\mathcal{L}_1$+$\mathcal{L}_{CARL}$: Training the network jointly with the L_1 loss and our proposed CARL; 3) $\mathcal{L}_1$+$\mathcal{L}_{DivC}$ [Wu *et al.*, 2021]: Training the network jointly with the L_1 loss and the divided-contrast loss [Wu *et al.*, 2021], which is to make a comparison

Metrics	$\lambda_2 = 1$	$\lambda_2 = 5$	$\lambda_2 = 10$	$\lambda_2 = 15$
PSNR	41.48	40.28	**41.92**	39.91
SSIM	0.9952	0.9935	**0.9954**	0.9932

Table 4: Parameter sensitivity analysis on SOTS-Indoor dataset.

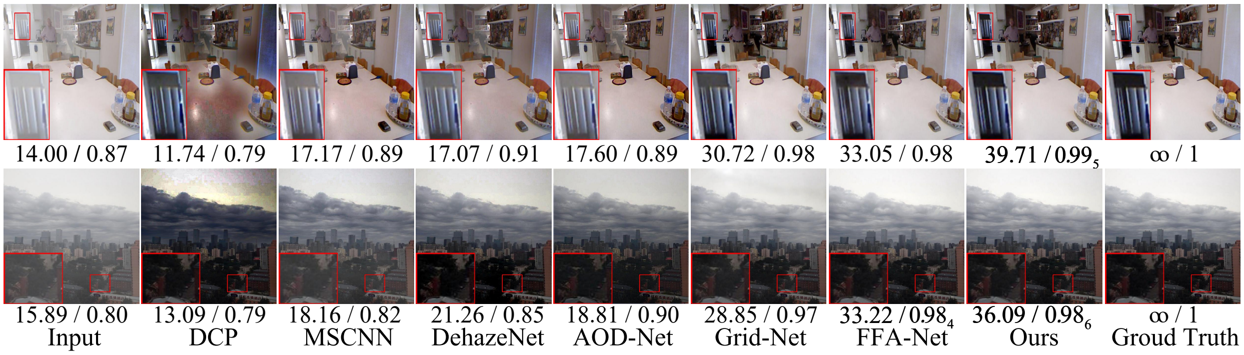

Figure 3: Visual comparisons on two hazy images with different methods. The first row: the dehazed results from SOTS-Indoor dataset. The second row: the dehazed results from SOTS-Outdoor dataset. The numbers under each image represent the PSNR and SSIM values.

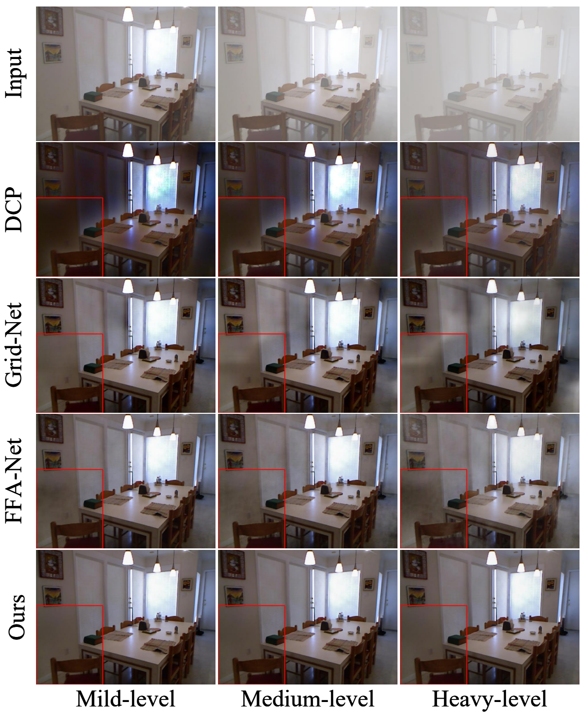

Figure 4: The dehazed results of different methods on the image with different hazy densities.

with ours; 4) $\mathcal{L}_1$+$\mathcal{L}_{CARL}$ + $\mathcal{L}_{CR}$: Training the network jointly with the L_1 loss, CARL and the consistency regularization, which is our final algorithm as illustrated in Eq. 5.

The performance of above-mentioned methods are summarized in Table 3, we can see that by adding our proposed CARL into the traditional L_1 loss, we can improve the baseline performance 3.17db and 0.91db PSNR on the SOTS-Indoor and Dense-Haze datasets, respectively. Compared with the relevant method $\mathcal{L}_{DivC}$ [Wu *et al.*, 2021], our proposed method outperforms $\mathcal{L}_{DivC}$ by a margin of 2.02db and 0.48db PSNR on these two datasets, respectively. When further adding the proposed consistency regularization $\mathcal{L}_{CR}$, we can improve the performance by 2.36db and 0.17db PSNR on these two datasets, respectively. As shown in Figure 4, we can see that our method generates more consistent dehazed images, and can deal with multi-level hazy images well.

Parameter sensitivity analysis is shown in Table 4. As defined in Eq. 5, our final loss function contains three terms: $\mathcal{L}_1$, $\mathcal{L}_{CR}$ and $\mathcal{L}_{CARL}$. To investigate the effect of hyper-parameters on the performance, we conduct comprehensive experiments with various values of these parameters. Here, we just list the performance with various λ_2 when $\lambda_1 = 1.0$ in Table 4, for the limitation of paper length. We can clearly see that our method yields best performance when $\lambda_2 = 10.0$.

5 Conclusion

In this paper, we propose a contrast-assisted reconstruction loss for single image dehazing. The proposed method can fully exploit the negative information to better facilitate the traditional positive-orient dehazing objective function. Besides, we also propose the consistency regularization to further improve the model robustness and consistency. The proposed method can work as a universal learning framework to further improve the image dehazing performance on top of various cutting edge dehazing network architectures, without bringing in additional computation cost or parameters in the testing phase. In the future, we will extend our method to many other relevant tasks, such as image deraining, image super-resolution, etc.

Acknowledgements

This work was supported in part by the National Key Research and Development Program of China under Grant 2018AAA0103202, in part by the National Natural Science Foundation of China under Grant 62176198, 61922066, 61876142, 62036007 and 62106184, in part by the Joint Project for Smart Computing of Shandong Natural Science Foundation (ZR2020LZH015), in part by the Technology Innovation Leading Program of Shaanxi under Grant 2022QFY01-15 and in part by Open Research Projects of Zhejiang Lab under Grant 2021KG0AB01.

References

[Ancuti *et al.*, 2018a] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *CVPRW*, 2018.

[Ancuti *et al.*, 2018b] Cosmin Ancuti, Codruta O Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-haze: a dehazing benchmark with real hazy and haze-free indoor images. In *ACIVS*, 2018.

[Ancuti *et al.*, 2019] Codruta O Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *ICIP*, 2019.

[Berman *et al.*, 2016] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *CVPR*, 2016.

[Cai *et al.*, 2016] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE TIP*, 2016.

[Dong *et al.*, 2020a] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *CVPR*, 2020.

[Dong *et al.*, 2020b] Yu Dong, Yihao Liu, He Zhang, Shifeng Chen, and Yu Qiao. Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing. In *AAAI*, 2020.

[He *et al.*, 2010] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *TPAMI*, 2010.

[He *et al.*, 2019] Tong He, Zhi Zhang, Hang Zhang, Zhongyue Zhang, Junyuan Xie, and Mu Li. Bag of tricks for image classification with convolutional neural networks. In *CVPR*, 2019.

[Hong *et al.*, 2020] Ming Hong, Yuan Xie, Cuihua Li, and Yanyun Qu. Distilling image dehazing with heterogeneous task imitation. In *CVPR*, 2020.

[Johnson *et al.*, 2016] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, 2016.

[Ju *et al.*, 2021] Mingye Ju, Can Ding, Charles A Guo, Wenqi Ren, and Dacheng Tao. Idrlp: Image dehazing using region line prior. *IEEE TIP*, 2021.

[Khosla *et al.*, 2020] Prannay Khosla, Piotr Teterwak, and etal Wang. Supervised contrastive learning. *arXiv:2004.11362*, 2020.

[Li *et al.*, 2017] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *ICCV*, 2017.

[Li *et al.*, 2018a] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. End-to-end united video dehazing and detection. In *AAAI*, 2018.

[Li *et al.*, 2018b] Boyi Li, Wenqi Ren, Dacheng Tao, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE TIP*, 2018.

[Li *et al.*, 2020] Pengyue Li, Jiandong Tian, Yandong Tang, Guolin Wang, and Chengdong Wu. Deep retinex network for single image dehazing. *IEEE TIP*, 2020.

[Liu *et al.*, 2019a] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *ICCV*, 2019.

[Liu *et al.*, 2019b] Xing Liu, Masanori Suganuma, Zhun Sun, and Takayuki Okatani. Dual residual networks leveraging the potential of paired operations for image restoration. In *CVPR*, 2019.

[Qin *et al.*, 2020] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *AAAI*, 2020.

[Qu *et al.*, 2019] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *CVPR*, 2019.

[Ren *et al.*, 2016] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *ECCV*, 2016.

[Ren *et al.*, 2018] Wenqi Ren, Lin Ma, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *CVPR*, 2018.

[Sakaridis *et al.*, 2018] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *IJCV*, 2018.

[Shao *et al.*, 2020] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In *CVPR*, 2020.

[Simonyan and Zisserman, 2014] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*, 2014.

[Tan, 2008] Robby T Tan. Visibility in bad weather from a single image. In *CVPR*, 2008.

[Tarvainen and Valpola, 2017] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *NeurIPS*, 2017.

[Wu *et al.*, 2021] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *CVPR*, 2021.

[Yoon *et al.*, 2021] Kuk-Jin Yoon, Pranjay Shyam, and Kyung-Soo Kim. Towards domain invariant single image dehazing. In *AAAI*, 2021.

[Zhang and Patel, 2018] He Zhang and Vishal Patel. Densely connected pyramid dehazing network. In *CVPR*, 2018.

[Zhang *et al.*, 2021a] Dingwen Zhang, Junwei Han, Gong Cheng, and Ming-Hsuan Yang. Weakly supervised object localization and detection: A survey. *IEEE TPAMI*, 2021.

[Zhang *et al.*, 2021b] Dingwen Zhang, Wenyuan Zeng, Jieru Yao, and Junwei Han. Weakly supervised object detection using proposal-and semantic-level relationships. *IEEE TPAMI*, 2021.