RESEARCH ARTICLE

Unsupervised Haze Removal for Aerial Imagery Based on Asymmetric Contrastive CycleGAN

XIN HE, WANFENG JI, AND JINPENG XIE

School of Basic Sciences for Aviation, Naval Aviation University, Yantai 264001, China

Corresponding author: Xin He (hexin6770@163.com)

ABSTRACT Aerial image dehazing is an important preprocessing step, since haze extremely degrades the imaging quality and affects subsequent the applications of aerial imagery. Most current haze removal methods achieve encouraging performance by relying on paired synthetic data, while are limited to their generality and scalability in the practical tasks. To this end, this paper aims to learn an effective unsupervised dehazing model from an unpaired set of clear and hazy aerial images. Motivated by the great advantages of contrastive learning in unsupervised representation field, we first attempt to formulate a Asymmetric Contrastive CycleGAN dehazing framework (namely ACC-GAN) to maximize the mutual information between the hazy domain and the haze-free domain. In the latent representation space, the introduced contrastive constraint ensures that the restored image is pulled closer to the clear image and pushed away from the hazy image, so as to indirectly regularize the unsupervised dehazing process. Importantly, different from the standard CycleGAN, we develop an additional feature transfer network into the forward path to form the asymmetric structure of ACC-GAN, which can enhance encoded features from hazy domain to haze-free domain. During training, multi-dimension loss terms are jointly built into a loss committee for generating dehazed results with higher naturalness and better fidelity. Experimental results on synthesis and real-world datasets indicate that our method is superior to existing unsupervised dehazing approaches, and is also very competitive to other related supervised models.

INDEX TERMS Aerial imagery, haze removal, asymmetric CycleGAN, unsupervised learning, contrastive learning.
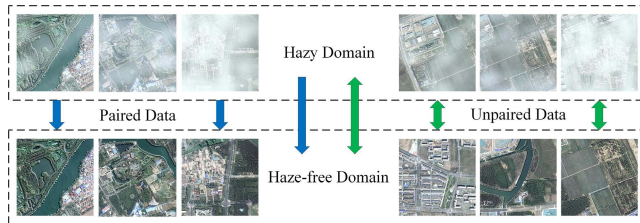
I. INTRODUCTION

Aerial images refer to photos taken from UAVs, helicopters and other aircraft, which have rich information content. Therefore, they are widely used in various fields, such as remote sensing [1], agriculture [9], geology [10] and earth science [11]. In addition, aerial images can also facilitate numerous subsequent high-level vision applications, such as target detection [12], aerial surveillance [13], scene understanding [14], land cover classification [15]. Since aerial images are perceived from considerable distances, these images often suffer from blurring, color shifts and low visibility due to variations in atmospheric paths. Such atmospheric effects, mainly in the form of clouds, fog and mist, can extremely degrade the imaging quality and aerial vision-based systems. Cloud is a giant cloud of ice crystals or water droplets moving through the atmosphere. Under cloudy conditions, the imaging path is severely obstructed, aerial

The associate editor coordinating the review of this manuscript and approving it for publication was Andrea F. Abate.

sensors cannot directly perceive the ground. Haze is the condensation of smoke, fine dust or light water vapor, resulting in the attenuation of light transmission through the air [16]. Aerial images taken in hazy conditions are typically characterized by low intensity, low contrast, and low visibility. Thus, it is of great interest to develop an effective aerial image dehazing algorithm to recover high-quality haze-free images.

The dehazing technology of aerial images has attracted increasing attention [1], and several classical approaches have been developed. Early methods employed image enhancement strategies, such as histogram equalization [2], homomorphic filter [3], wavelet transform [4] and Retinex [5], which can effectively recover clear images. However, they often only change the contrast of the image and do not really remove the haze. Later, some image restoration based dehazing algorithms have been designed to deal with haze removal, such as dark channel prior (DCP) [6], haze thickness map (HTM) [7] and haze optimized transformation (HOT) [8]. These algorithms start from the blur mechanism and the causes of haze image degradation. The mathematical

FIGURE 1. Schematic diagram of comparison between paired supervised based dehazing method (blue arrow) and unpaired without supervised based dehazing method (green arrow).

(a) Hazy Input   (b) DCP   (c) GFN   (d) CycleGAN   (e) Ours

FIGURE 2. Aerial image dehazing results by different methods, including DCP [6] (prior-based method), GFN [22] (paired learning-based method), and CycleGAN [35] (unpaired learning-based method). Zooming in the figures offers a better view at the dehazing capability.

model of imaging process is established to deal with the inverse process of image degradation, and then the clear image is restored. The commonly used atmospheric scattering model [17], [18] describes the simple approximation of haze effect, and its mathematical expression is

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

where $I(x)$ and $J(x)$ denote the hazy image under observation and the clear image, respectively. Note that, there are two critical parameters: A refers to the global atmospheric light and $t(x)$ is termed as the medium transmission map. Under the homogeneous haze, the map of transmitting process $t(x)$ is define as:

$$t(x) = \exp(-\beta d(x)), \qquad (2)$$

where β and $d(x)$ denote the atmosphere scattering coefficient and the scene depth. Although the above methods are generally simple and fast, their threshold parameters need to be fine-tuned to obtain satisfactory dehazing results, and their results are easily lead to serious color distortion (see Figure 2 (b)).

In recent years, numerous deep learning based methods have been employed convolutional neural network (CNNs) and generative adversarial networks (GANs) for image dehazing. The CNN based methods mainly utilize the atmospheric scattering model to regress the transmission map and clear image [19]–[22], or directly restore the results in end-to-end training manner without relying on the model [23]–[25]. The GAN based methods [26]–[28] adopt the game strategies to map the hazy image to a clear one with the help of generator and discriminator. However, due to the aviation scene is complex, the objects are diverse and the space is vast, the direct application of these previous ordinary dehazing models to aerial images often suffers from low-quality restored results. These models lack the ability to express complex scenes, resulting in less effective extraction of haze-related features. In particular, they do not consider non-uniform haze, which is the most common haze state in aerial images. To overcome such problems, a series of haze removal methods [1] for aerial image have started to emerge. On the one hand, the latest aerial image dehazing datasets have been established, such as SateHaze1k (three levels of fog, namely thin, moderate, and thick fog) [29], UN-HAZE (uniform haze-clear image pairs) [30], and NONUN-HAZE (nonuniform haze-clear image pairs) [30]. On the other hand,
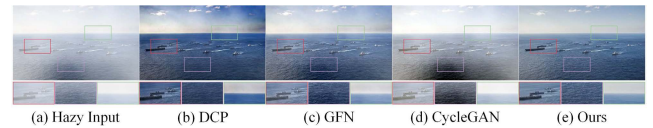
the recent aerial image dehazing algorithms exhibit remarkable performance, such as MRCNN [32], H2RL-Net [31], and RSDehazeNet [16]. However, it is worth noting that the above mentioned dehazing networks rely on paired synthetic data to train their models in a fully supervised manner, which directly learn a suitable representation from one-to-one mapping relationship, as shown in Figure 1. The expression can be defined as follows:

$$I(x) = F(J(x)), \qquad (3)$$

where $F(\cdot)$ represents the mapping function related to image degradation. For supervised dehazing models, there exist two major limitations: (1) Under complex and changeable hazy environments, it is time-consuming and laborious to collect paired training data in large scale. (2) Since the current synthetic hazy datasets are generally too simplified to represent the complex real-world hazy scenarios, the performance of those supervised methods in dealing with real aerial cases will significantly drop due to the intra-domain and inter-domain gap. Figure 2 (c) shows an real-world example in which the dehazing result is suboptimal.

Given that, some researchers [33], [34] start to shift to explore the unsupervised aerial image dehazing strategies. Most existing methods can be viewed as an image-to-image transfer learning problem, which tries to get the transferable and domain invariant feature through decreasing the distribution discrepancy of two domains. Following the advantage of CycleGAN [35] in unpaired setting, visual adversarial frameworks have performed feature-level and pixel-level adaptation jointly between the hazy domain and the haze-free domain, see Figure 1. Due lack of supervised constraints suitable for haze removal, directly applying limited CycleGAN-based adversarial frameworks for the dehazing task will inevitably suffer from the under-constrained problem. Figure 2 (d) gives a corresponding dehazing example. Fundamentally, existing CycleGAN-based unsupervised dehazing methods do not supply an effective mechanism that can better learn mutual information between the hazy domain and the haze-free domain and infer its latent-space representation, which could be helpful for unsupervised learning [46]. One more thing, since the domain gap between the hazy domain and the haze-free domain, previous unsupervised CycleGAN-based dehazing methods are not conducive to domain translation by using symmetric architecture [48].

Towards this end, a novel unpaired aerial image dehazing solution called Asymmetric Contrastive CycleGAN (ACC-GAN) is developed. CycleGAN is a common

framework for unsupervised low-level vision tasks. Motivated by the great advantages of contrastive learning in unsupervised representation field, we introduce recent contrastive learning into the CycleGAN framework to achieve end-to-end training without paired information. Contrastive learning was widely applied in unsupervised representation learning [36]–[38] and showed its great ability to capture useful visual features by leveraging both positive and negative samples. Considering the image dehazing task, the information of the hazy image and the clear image are taken as negative samples and positive samples. The proposed ACC-GAN aims to maximize mutual information, and the representations are learned in their embeddings by encouraging the positives closer while keeping the negatives further away. In other words, our introduced contrastive regularization indirectly constrains the latent space of corresponding patches to guide dehazing by acting as a "supervising teacher". To obtain best dehazing performance (see Figure 2 (e)), we combine multiple loss functions into a loss committee to further constrain the proposed ACC-GAN for the better jointly optimization.

In summary, we make the following contributions:

- We formulate an effective ACC-GAN to leverage contrastive learning to maximize the mutual information between the hazy domain and the haze-free domain, which helps the model trained in an unpaired fashion and improves aerial image dehazing generalization.
- To the best of our knowledge, this study is the first attempt to integrate contrastive learning into the architecture of asymmetric CycleGAN, which provides information supplement for one-way translation by utilizing the feature transfer network.
- We build an effective loss committee embedded in ACC-GAN for collaborative optimization and benefit to generate dehazed results with higher naturalness and better fidelity.
- Experimental results on multiple synthesis and real-world datasets indicate that our method is superior to existing unsupervised dehazing approaches, and is also very competitive to other related supervised dehazing models.

The remainder of this paper is organized as follows: Section II describes an overview of related work. In Section III, we provide the main technical details of our proposed approach. Then, Section IV presents the comprehensive experimental results with discussions. At last, the conclusions are given in Section V.

II. RELATED WORK

In this section, we briefly introduce several aerial image dehazing approaches based on priori, supervised learning and unsupervised learning types. In addition, recent successful contrastive learning mechanism is presented.

A. AERIAL IMAGE DEHAZING

Haze removal from aerial images has attracted increasing attention, and many classic methods have been developed.

Traditional researchers toward aerial image dehazing task by utilizing various kinds of prior knowledge, such as He *et al.* [6] proposed the dark channel prior (DCP), Zhang *et al.* [8] presented haze optimized transformation (HOT) method, and Makarau *et al.* [7] raised haze thickness map (HTM) method.

Recently, learning-based approaches start to employ CNNs and GANs for haze removal and achieve state-of-the-art performance. Based to the physical model, the CNN based methods devote to regress the transmission map or dehazed the image using multiple features. Cai *et al.* [19] first exploited DehazeNet to recover haze-free image by simple pixelwise operations. The lightweight AOD-Net [21] was proposed to be an end-to-end architecture design dealing with the dehazing problem. Liu *et al.* [24] presented a novel trainable Grid-Dehaze Network (GDN) that indicates how confident the network is about the multi-scale features are learned. Later, these models succussed inspired great efforts invested in the development of aerial image dehazing methods, such as MRCNN [32], H2RL-Net [31], RSDehazeNet [16], where more effective network architectures were mainly designed. Unlike CNN, the GAN based methods have adopted generation and discrimination networks to regularize the dehazed image to have reliable colors and structures. The work of [39] explored dehazing approach using SAR and multi-spectral images as input to train a GAN. Afterwards, Huang *et al.* [29] developed conditional GAN with SAR image prior to eliminate the image blurring. Despite the remarkable progress, the above data-driven learning methods depend on the paired training data that extremely difficult to acquire. In practice, however, for real hazy aerial images without ground-truth (i.e., unpaired), most existing supervised dehazing models may fail due to the irregularity and nonuniform haze.

To the best of our knowledge, there are only a few unsupervised efforts for aerial image dehazing. CycleGAN [35] has been employed to address the unpaired image-to-image translation problem, to achieve an unsupervised dehazing process. SkyGAN [34] was suggested by utilizing HSI guidance and multi-cue color input for dehazing. Hu *et al.* [33] used edge-sharpening cycle-consistent adversarial network without require prior information. However, these unsupervised dehazing performance is still insufficient, leaving room for further improvement. Note that, these unsupervised dehazing networks are still implemented on the standard CycleGAN, and few studies consider the impact of asymmetric Cycle-GAN on dehazing performance. Therefore, this paper focuses on learning an effective asymmetric CycleGAN-based dehazing model from an unpaired set of clear and hazy aerial images.

B. CONTRASTIVE LEARNING

In recent years, contrastive learning has been widely studied in self-supervised and unsupervised representation learning [40], [41]. Its working principle is to pull the positive sample near the anchor and push the negative sample away in the latent space, so as to increase the mutual information
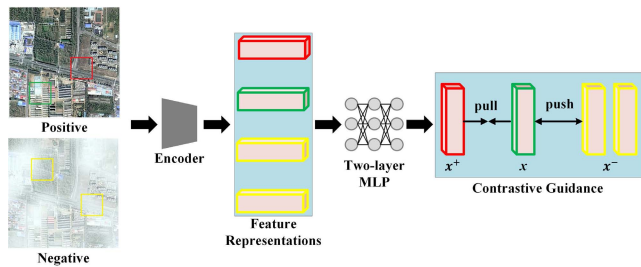
FIGURE 3. Schematic diagram of aerial image dehazing based on contrastive learning scheme. Specifically, the clear image and the hazy image are regarded as positive and negative exemples, respectively. First, the features of each exemples are extracted by the encoder, and then the stack feature representation is formed and finally sent to the two-layer MLP. The contrastive guidance aim to learns a representation to pull positive exemples close to the query and push negative exemples away.

in the learned representation. The selection of positive and negative samples may vary depending on the specific downstream tasks. Previous work has applied contrastive learning to a host of high-level vision tasks [42], [43], which has shown the ability to capture complex visual features and model the contrast between positive and negative features.

More recently, contrastive learning also made an important exploration in the low-level vision tasks. For exemple, CUT [36] first utilized noise contrast estimation to achieve image-to-image translation, which aims to learn the mutual relationship between the input image patches and the generated image patches. Based on the dual advantages of CUT [36] and CycleGAN [35], Han *et al.* [37] proposed DCLGAN to perform unsupervised image-to-image translation, and achieved new state-of-the-art performance. Wu *et al.* [38] developed a novel AECR-Net based on autoencoder network and contrastive regularization for single image dehazing. CWR [44] was a model of contrastive learning applied to underwater image restoration problem, which showes impressive restoration performance on the created underwater dataset HICRD. Chen *et al.* [46] designed an effective unsupervised single image deraining GAN to explore the mutual properties of the unpaired exemples by means of contrastive learning. Wang *et al.* [45] introduced a unsupervised super-resolution network, which uses contrastive learning to deal with various degradations based on the learned representations.

Inspired by the above related work, this paper introduces a contrastive learning scheme for aerial image dehazing task. Considering the properties of haze removal task, the information of the hazy image and the clear image are taken as negative samples and positive samples, respectively. The schematic diagram is shown in Figure 3. Specifically, the features of each exemples are first extracted by the encoder, and then the stack feature representation is formed and finally sent to the two-layer MLP. In the latent representation space, the introduced contrastive learning scheme ensures that the restored image is pulled closer to the clear image and pushed away from the hazy image.

III. PROPOSED APPROACH

In this section, we first illustrate the proposed approach, including the overall network design and technical implementation details. Then we introduce each member of the loss committee.

A. ASYMMETRIC ARCHITECTURE

Given a hazy domain $\mathbb{D}_h$ (a set of hazy images I_h) and a haze-free domain $\mathbb{D}_n$ (a set of haze-free images I_n), the goal of traditional paired image dehazing is to learn a direct mapping function $F_{pair}: D^{w \times h} \to D^{W \times H}$ based on the paired dataset $\mathcal{D}_{pair} = \left\{ \left(I_h^i, I_n^i \right) \right\}_{i=1}^N$, so that the dehazed image

$$I_n = F_{pair}\left(I_h \right) \in \mathbb{D}^{W \times H}, \tag{4}$$

can well estimate the corresponding ground-truth haze-free image. However, this way of paired learning often suffer from certain limitations. In this work, our goal is to learn an indirect mapping function $F_{unpair}: D^{w \times h} \to D^{W \times H}$ from unpaired training data, so that the dehazed image

$$I_n = F_{unpair}\left(I_h \right) \in \mathbb{D}^{W \times H}, \tag{5}$$

where we do not have any explicit association between individual images in $\mathbb{D}_h$ and $\mathbb{D}_n$. Since the ground truth labeled data is not available, unpaired image dehazing is much more challenging than paired image dehazing. In order to achieve this goal, it is very important to explore the internal relationship between the images of the two domains.

We now present our proposed ACC-GAN and its overall architecture is shown in Figure 4. In detail, it has two generators G, F generating the clean and hazy images respectively and two discriminators D_1, D_2 distinguishing between fake dehazed images and real clean images. Table 1 and Table 2 show the specific architecture details of the generator and discriminator. Specifically, the designed ACC-GAN covers two main branches below: (1) hazy to hazy cycle-consistency branch $h \to n_h \to h^*$, predicting clean images from hazy images and subsequently reconstructing them with the use of the generator; (2) haze-free to haze-free cycle-consistency branch $n \to h_n \to n^*$, generating hazy images from clean images and afterwards reconstructing them with the use of the generator. Most of CycleGAN-based [35] models adopt symmetrical translation structure for the optimization of stable equilibrium. Instead, motivated by the similar intuition in [48], we design a novel asymmetric CycleGAN for bidirectional translation process. The reason why we adopt asymmetric architecture is intuitive. If we treat the forward cycle and backward cycle equally, it is not enough to fully excavate the feature properties from hazy domain to haze-free domain. Therefore, we insert an additional feature transfer network into the forward path to achieve an enhanced and disentangled information for domain translation.

However, it hardly tackles with the heavy variations in haze removal task only by relying on cycle-consistency constraints, because this constraint is weak in complex image space [46]. Notably, additional knowledge from unpaired
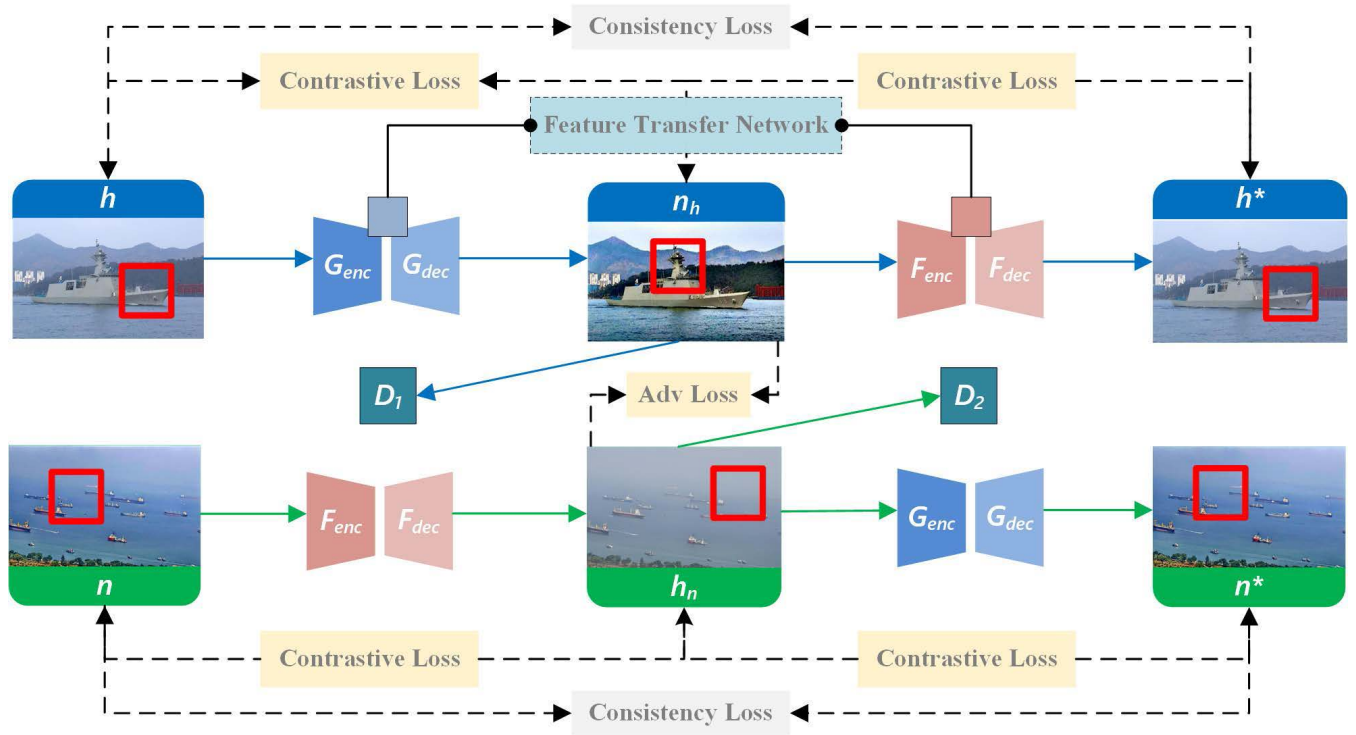
FIGURE 4. Overall architecture of the proposed Asymmetric Contrastive CycleGAN framework (ACC-GAN). We omit the identical preserving loss and perceptual loss here.

TABLE 1. The architecture of generator. RP means reflectionpad, IN denotes instance normalization, RC represents residual connection, and SN denotes spectral normalization.

Components		Layer Details	Input Shape	Output Shape
Encoder	Conv	RP+Conv+IN+ReLU	3,515,512	64,512,512
	Downsampling-1	Conv+IN+ReLU	64,512,512	128,256,256
	Downsampling-2	Conv+IN+ReLU	128,256,256	256,128,128
	Residual Block-1	Conv+IN+ReLU+Conv+IN+RC	256,128,128	256,128,128
	Residual Block-2	Conv+IN+ReLU+Conv+IN+RC	256,128,128	256,128,128
	Residual Block-3	Conv+IN+ReLU+Conv+IN+RC	256,128,128	256,128,128
	Residual Block-4	Conv+IN+ReLU+Conv+IN+RC	256,128,128	256,128,128
	Residual Block-5	Conv+IN+ReLU+Conv+IN+RC	256,128,128	256,128,128
Decoder	Residual Block-6	Conv+IN+ReLU+Conv+IN+RC	256,128,128	256,128,128
	Residual Block-7	Conv+IN+ReLU+Conv+IN+RC	256,128,128	256,128,128
	Residual Block-8	Conv+IN+ReLU+Conv+IN+RC	256,128,128	256,128,128
	Residual Block-9	Conv+IN+ReLU+Conv+IN+RC	256,128,128	256,128,128
	Upsampling-1	DeConv+SN+ReLU	256,128,128	128,256,256
	Upsampling-2	DeConv+SN+ReLU	128,256,256	64,512,512
	Conv	RP+Conv+SN+Tanh	64,512,512	3,512,512
Feature Transfer Network	Residual Block-10	Conv+IN+ReLU+Conv+IN+RC	256,128,128	256,128,128
	Residual Block-11	Conv+IN+ReLU+Conv+IN+RC	256,128,128	256,128,128
	Residual Block-12	Conv+IN+ReLU+Conv+IN+RC	256,128,128	256,128,128

information based regularization has been adopted, enabling the model to improve the quality of the restored images. Therefore, we further introduce contrastive regularization indirectly constrains the latent space of corresponding image patches to better guide dehazing by acting as a ''supervising teacher''. The component is elucidated below.

TABLE 2. The architecture of discriminator. SN denotes spectral normalization. For LeakyReLU, we set negative slope as 0.2.

Components		Layer Details	Input Shape	Output Shape
Discriminator	Downsampling-1	Conv+SN+LeakyReLU	3,512,512	64,256,256
	Downsampling-2	Conv+SN+LeakyReLU	64,256,256	128,128,128
	Downsampling-3	Conv+SN+LeakyReLU	128,128,128	256,64,64
	Downsampling-4	Conv+SN+LeakyReLU	256,64,64	512,64,64
	Downsampling-5	Conv	512,63,63	1,62,62

B. CONTRASTIVE REGULARIZATION

Inspired by contrastive learning, we seek to maximize the mutual information between the corresponding patches of input and output, which can supervise the dehazing process with feature disparity obtained by constructing patch-wise contrastive loss. For example, for the ground patch of the generated dehazed image, we should be able to associate it more closely with the same ground patch of the original input image rather than the rest of the image. In contrastive learning, there are two aspects to consider: one is how to construct the feature representation space for comparison, and the other is how to select positive and negative pairs. Considering the properties of haze removal task, the information of the hazy image, clear image and restored image are taken as negative samples, positive samples and anchors, respectively. The positive pair includes the restored image and the corresponding clear image, while the negative pair includes the restored image and the corresponding hazy image. The feature visualization process is shown in Figure 5. For a anchor point, contrastive learning aims to encourage the positives closer while keeping the negatives further awary.

To constrain features with the discovered pairs of patches in their embeddings, let us define the query, positive, and N negatives to v, v^+, and v_n^-. Motivated by [36], we also establish a $(N + 1)$ path classification problem and calculate the probability of choosing "positive" instead of "negative." Mathematically, this can be further regarded as the cross entropy loss, which is calculated by the following formula:

$$\ell\left(v, v^+, v^-\right)$$
$$= -\log\left(\frac{\exp\left(\text{sim}\left(v, v^+\right)/\tau\right)}{\exp\left(\text{sim}\left(v, v^+\right)/\tau\right) + \sum_{n=1}^{N}\exp\left(\text{sim}\left(v, v_n^-\right)/\tau\right)}\right)$$
(6)

where the cosine similarity between u and v can be expressed as:

$$\text{sim}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}^\top \boldsymbol{v} / \|\boldsymbol{u}\|\|\boldsymbol{v}\|.$$
(7)

In formula (6), τ indicates the temperature parameter to be used to scale the distance between the query and other examples. Here, we use 0.07 as the default value. The number of negatives is set to 255 by default. Technically, we extract features of images from the four encoding layers (the downsampling-1 and downsampling-2, the residual block-1 and residual block-5) of the generator and send them to a 2-layer multi-layer perceptron (MLP) projection head.
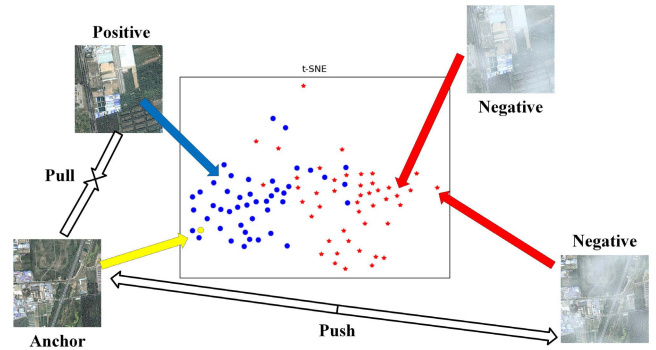
FIGURE 5. Visualization of latent features with t-SNE technology. For a given anchor point, contrastive loss aims to ush the anchor far away from negative samples and pull the anchor close to positive samples in the latent representation space.

After having a stack of features $\{z_l\}_L = \left(G_{\text{enc}}^l(\boldsymbol{x})\right)_L$, we select a query each time, refer the corresponding positive as z_l^s, and other negatives as $z_l^{S\backslash s}$. Such latent representations can be captured the relationship between the hazy domain $\mathbb{D}_h$ and the haze-free domain $\mathbb{D}_n$ by acting the role of a "supervising teacher". Thus, we introduce a contrastive regularization term to regularize the captured images' latent representations. In other word, this regularization term is employed in the form of contrastive loss (as a member of the loss committee). Mathematically, the contrastive loss is refined as

$$\mathcal{L}_{\text{CONT}_h}(G, H) = \mathbb{E}_{h\sim H} \sum_{l=1}^{L} \sum_{s=1}^{S_l} \ell\left(\hat{z}_l^s, z_l^s, z_l^{S\backslash s}\right).$$
(8)

Considering the reverse mapping on the other side, we also introduce a similar loss,

$$\mathcal{L}_{\text{CONT}_n}(F, N) = \mathbb{E}_{n\sim H} \sum_{l=1}^{L} \sum_{s=1}^{S_l} \ell\left(\hat{z}_l^s, z_l^s, z_l^{S\backslash s}\right).$$
(9)

Interestingly, this guidance from contrastive learning essentially can be regarded as an adversarial learning collaboration: the lower disparity the mutual feature is, the more similar the latent representation is, the cleaner the dehazed output is, and vice versa.

C. LOSS COMMITTEE

For aerial image dehazing task, the output of dehazing process should be close to the haze-free image domain in a certain level. In addition, another important goal is to ensure

TABLE 3. Descriptions of different experimental datasets, including the RESIDE dataset, SateHaze1k dataset, Overwater dataset, and RealHaze dataset.

Datasets		RESIDE	SateHaze1k			Overwater	RealHaze
			Thin	Moderate	Thick		
Train-Set	Hazy Images	2903	355	355	355	2090	0
	Clear Images	3577	355	355	355	2441	0
Test-Set	Hazy Images	100	45	45	45	127	30
Type			Synthetic			Real-world	

content consistency and avoid color-/structure- corruption after dehazing. To this end, we combine multi-dimension loss terms into a loss committee as the task-specific proxy guidance. These optimization losses are jointly embedded into the proposed ACC-GAN during the training stage. We describe the members of this committee in the following.

1) ADVERSARIAL LOSS

For promoting the generated images to be highly realistic, the adversarial loss is adopted in two domains. In terms of the mapping $G : \mathbb{D}_h \rightarrow \mathbb{D}_n$, the adversarial loss is formulated as:

$$L_{adv}(G, D_1) = \mathbb{E}_{n \sim N} \left[\log D_1(n) \right] \\ + \mathbb{E}_{h \sim H} \left[\log(1 - D_1(G(h))) \right]. \quad (10)$$

Similarly, we define same adversarial loss for the mapping $F : \mathbb{D}_n \rightarrow \mathbb{D}_h$:

$$L_{adv}(F, D_2) = \mathbb{E}_{h \sim H} \left[\log D_2(h) \right] \\ + \mathbb{E}_{n \sim N} \left[\log(1 - D_2(F(n))) \right]. \quad (11)$$

The overall adversarial loss is calculated by

$$L_{adv} = L_{adv}(G, D_1) + L_{adv}(F, D_2). \quad (12)$$

2) CYCLE CONSISTENCY LOSS

We also introduce a cycle consistency loss for solving the issue that an adversarial loss independent fails to ensure the matching of the target distribution and the output distribution. L_{cycle} is capable of limiting generated samples' space and preserving image content. The loss in image domain is expressed as:

$$L_{cycle} = E_{n \sim N} \left[\|n - n^*\|_1 \right] + E_{h \sim H} \left[\|h - h^*\|_1 \right]. \quad (13)$$

3) CONTRASTIVE LOSS

As mentioned above, the overall contrastive loss can be defined as:

$$L_{cont} = \mathcal{L}_{\text{CONT}_h}(G, H) + \mathcal{L}_{\text{CONT}_n}(F, N). \quad (14)$$

4) IDENTITY PRESERVING LOSS

To further preserve the identity information such as color invariance between the input and the output, we add an identity preserving loss. Considering training speed and time complexity, we do not use PatchNCE loss [36] as identity preserving loss. Instead, we define it as follows:

$$L_{idt} = \mathbb{E}_{n \sim N} \left[\|G(n) - n\|_1 \right] + \mathbb{E}_{h \sim H} \left[\|F(h) - h\|_1 \right]. \quad (15)$$

5) PERCEPTUAL LOSS

Other than the measurement of images' per-pixel difference, perceptual loss displays a relationship to the distinction of feature maps, covering a wide range of contents and perceptual quality. As such, we exploit the perceptual loss to encode the difference between the original hazy image and the corresponding dehazed image. The perceptual loss can be defined as:

$$L_{per} = \|\varphi_l(h) - \varphi_l(F(G(h)))\|_2^2 \\ + \|\varphi_l(r) - \varphi_l(G(F(r)))\|_2^2, \quad (16)$$

where $\varphi_l(\cdot)$ denotes the features extracted from l-th layer of the pretrained CNN.

6) TOTAL LOSS

$$L_{\text{total}} = \lambda_1 L_{adv} + +\lambda_2 L_{cycle} + \lambda_3 L_{cont} + \lambda_4 L_{idt} + \lambda_5 L_{per}, \quad (17)$$

where $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 2$, $\lambda_4 = 1$, and $\lambda_5 = 0.01$ in our experiments.

IV. EXPERIMENTS AND RESULTS

In this section, we describe the datasets and training details. Then, comprehensive dehazing experiments are employed to demonstrate the effectiveness of proposed ACC-GAN against the current comparing approaches. In addition, ablation studies are conducted to validate the efficiency of our designed framework.

A. DATASETS SETUP

We use two synthetic datasets including the RESIDE dataset [20] and SateHaze1k dataset [29] with different haze distribution states. Besides, real-world datasets are collected to evaluate the performance of dehazing and two datasets are involved: Overwater dataset [47] and RealHaze dataset [31]. The detailed descriptions are tabulated in Table 3, including the synthetic and real-world datasets. Note that we shuffle above-mentioned images randomly in order to achieve unpaired supervised learning.

B. TRAINING DETAILS

The detailed architecture and parameter settings of the proposed ACC-GAN are depicted in Table 1 and Table 2 using Pytorch framework. To accelerate the training process, the Adam optimization is applied with a batch size of 1.

FIGURE 6. Qualitative evaluations on the RESIDE synthetic dataset. Please zoom in the figures to better see the comparison.

(a) Hazy Input   (b) DCP   (c) DehazeNet   (d) AOD-Net

(e) GFN   (f) CycleGAN   (g) Ours   (h) Ground Truth

TABLE 4. Comparison of quantitative results on the RESIDE synthetic dataset.

Methods	DCP	DehazeNet	AOD-Net	GFN	CycleGAN	Ours
PSNR ↑	16.62	17.25	19.06	22.30	20.14	**24.35**
SSIM ↑	0.817	0.826	0.850	0.881	0.864	**0.912**
MSE ↓	523.608	347.652	237.416	185.463	472.220	**126.876**
Run Time ↓	0.135	0.208	**0.079**	0.431	0.162	0.167

For the better performance, we train our model for total 200 epochs. The proposed network is trained from the scratch for 100 epochs with the learning rate of 0.0001, followed by another 100 epochs with the learning rate linearly decayed to 0. The batch size is set to 1. We load all images in 286 × 286 resolution and randomly crop them into 256 × 256 patches during training. For the selection of negative samples, we sample 256 negatives from within the same
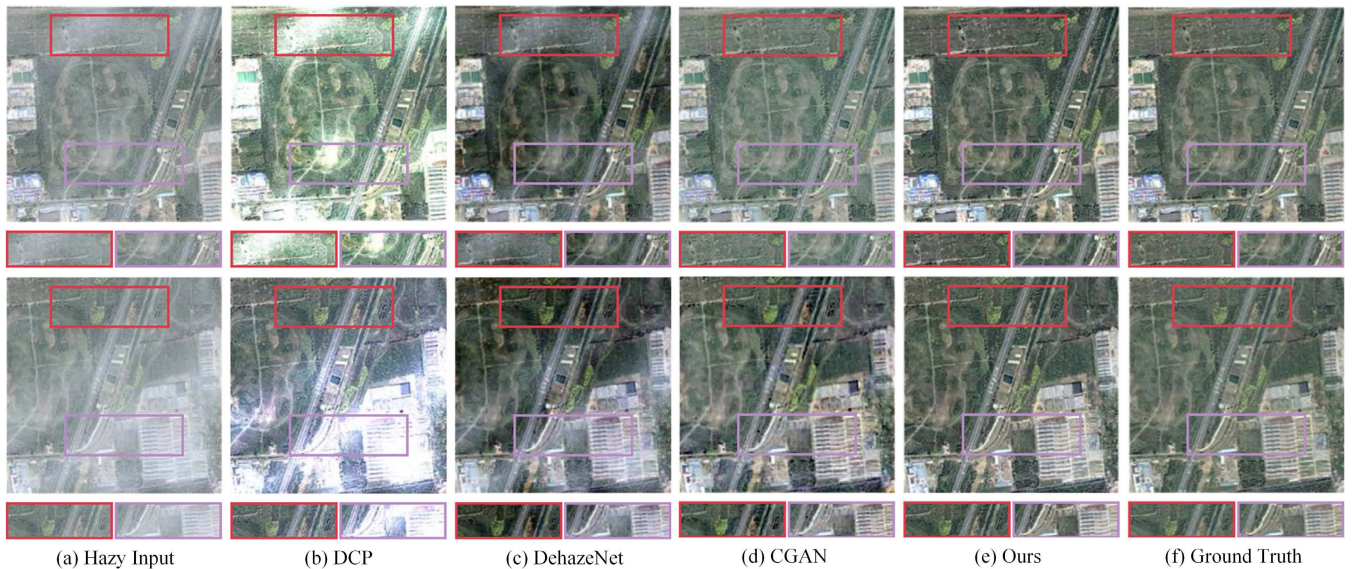
FIGURE 7. Qualitative evaluations on the SateHaze1k synthetic dataset. Please zoom in the figures to better see the comparison.

| (a) Hazy Input | (b) DCP | (c) DehazeNet | (d) CGAN | (e) Ours | (f) Ground Truth |

image (internal). To enhance the stability of adversarial learning, we keep an image buffer that stores the 50 previously created images. Due to the complexity of the inference phase of contrastive learning, the training time is determined by the number of samples. For the RESIDE and Overwater datasets, the training phase took about 5 days. For the SateHaze1k dataset, the training phase took about 3 days. All the comparing testing experiments performs with the same datasets and hardware environment is performed on a server with the NVIDIA Tesla P100-PCIE-16GB GPU. All parameters are defined via cross validation using the validation set, and the whole network is trained in the unsupervised manner.

C. RESULTS ON SYNTHETIC DATASETS
With the help of the ground truth in synthetic datasets, we perform the quantitative comparison using two commonly evaluation metrics: the Peak Signal to Noise Ratio (PSNR) and Structural SIMilarity index (SSIM).

1) EVALUATION ON RESIDE DATASET
We compare our method with other five comparing image dehazing methods, including DCP [6], DehazeNet [19], AOD-Net [21], GFN [22], and CycleGAN [35]. Meanwhile, the Mean Square Error (MSE) and computational cost are also compared. Through the Table 4, our proposed method gets the highest value both in PSNR/SSIM and the lowest value in MSE. The notable increasing scores reflects the excellent performance of ACC-GAN, which benefits from the improvement of contrastive regularization for unsupervised learning. Even though DCP and AOD-Net algorithms have advantages in running speed, their dehazing performance is far from enough. In contrast, our proposed method can

take into account both dehazing performance and running cost. Besides the quantitative results, we further present one example for visual observation comparisons in Figure 6. CycleGAN is an unsupervised-based learning algorithm. Unfortunately, its recovery performance is far from enough. In contrast, our network performs better than other models and deals with majority of fog. Since other comparing results contain more haze and color distortion, which keep consistent with the above PSNR results.

2) EVALUATION ON SATEHAZE1K DATASET
To objectively evaluate the aerial image dehazing performance, we compare our method with DCP [6], DehazeNet [19], FCTF [30] and CGAN [29]. Table 5 and Figure 7 show quantitative and qualitative results on the SateHaze1k dataset, respectively. The Satehaze1k dataset is more challenging because non-uniform haze is a common state in aerial images. Through Table 5, our proposed method gets the highest scores both in PSNR and SSIM except for the PSNR of thick fog. As displayed in Figure 7, DehazeNet leaves too much haze in the dehazed results, particularly in the thick fog condition. Zooming the color boxes, the main drawbacks of CGAN is that it tends to cause color distortion and blur the contents. Compared to the above approaches, our method can remove various densities of haze and preserve color and structural information.

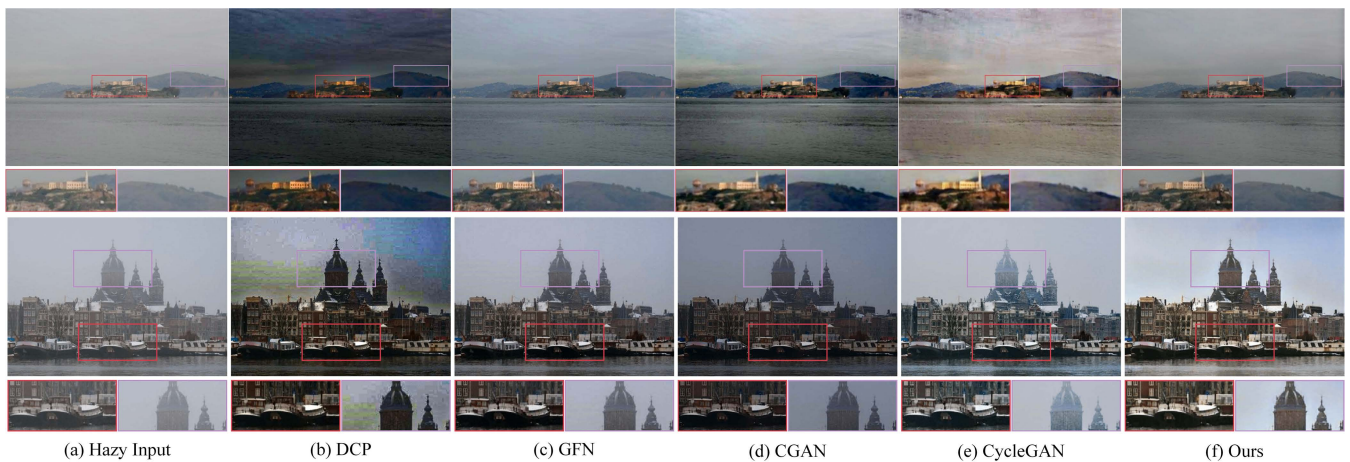### D. RESULTS ON REAL-WORLD DATASETS
For further general verification in practical use, we conduct experiments on two real-world hazy datasets. For the cases without ground truth, we only perform qualitative comparison.

TABLE 5. Comparison of quantitative results on the SateHaze1k synthetic dataset. Bold indicates the best dehazed results.

Methods	Thin Fog		Moderate Fog		Thick Fog		Average Value	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Input	12.77	0.724	12.58	0.739	8.58	0.421	11.31	0.628
DCP	13.15	0.725	9.78	0.573	10.25	0.585	11.06	0.627
DehazeNet	19.75	0.895	18.12	0.855	14.33	0.706	17.40	0.818
FCTF	20.41	0.885	23.59	0.931	18.37	0.802	20.79	0.873
CGAN	24.16	0.906	25.31	0.926	**25.07**	0.864	24.84	0.898
Ours	**24.83**	**0.912**	**25.56**	**0.928**	24.89	**0.886**	**25.09**	**0.908**

(a) Hazy Input   (b) DCP   (c) GFN   (d) CGAN   (e) CycleGAN   (f) Ours
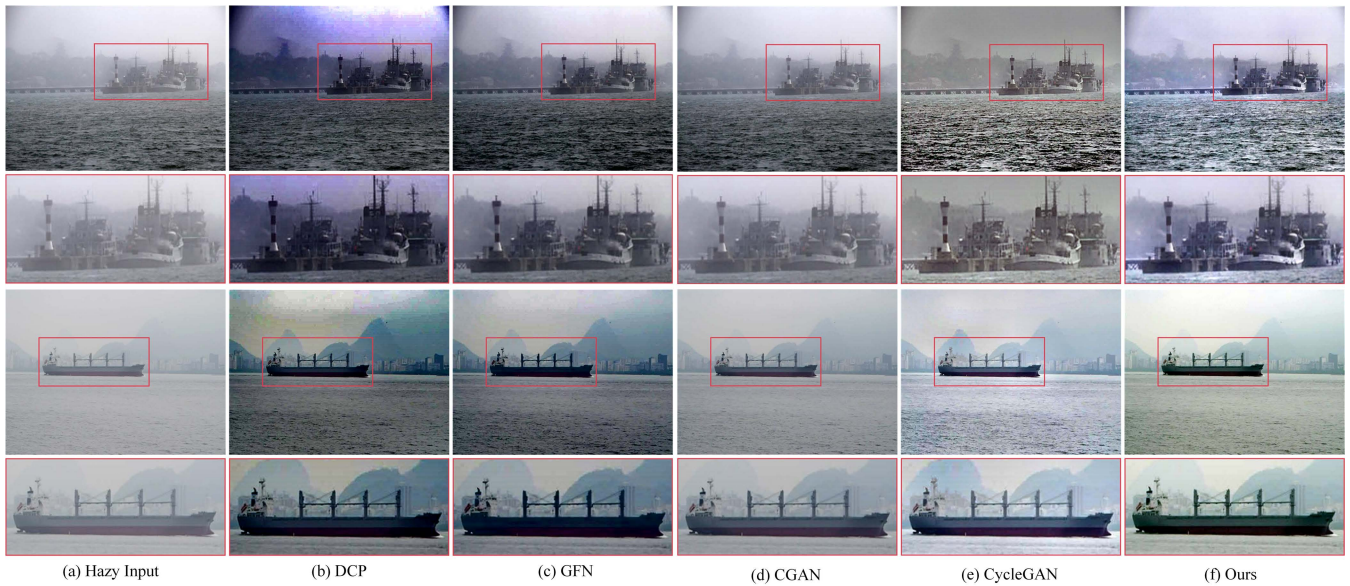
FIGURE 8. Qualitative evaluations on the Overwater real-world dataset. Please zoom in the figures to better see the comparison.

(a) Hazy Input   (b) DCP   (c) GFN   (d) CGAN   (e) CycleGAN   (f) Ours

FIGURE 9. Qualitative evaluations on the RealHaze real-world dataset. Please zoom in the figures to better see the comparison.

1) EVALUATION ON OVERWATER DATASET

Figure 8 compares the results on the Overwater dataset of all competing methods visually, including DCP [6], GFN [22], CGAN [35] and CycleGAN [35]. As hazy aerial images from ocean scenes are more complex, all the competing methods fail to recover high-quality clear images. It can be observed that the proposed method significantly competitive others in removing the majority of haze while preserving

TABLE 6. Quantitative analysis of different loss functions.

Models	w/o L_{adv}	w/o L_{cycle}	w/o L_{cont}	w/o L_{idt}	w/o L_{per}	Ours
PSNR	22.77	23.83	21.58	24.02	23.96	**24.35**

image structure even in the light effecting and dark surrounding. Obviously, developing unpaired learning-based dehazing method is more valuable for real-world generalized scenes.

2) EVALUATION ON REALHAZE DATASET

Similarly, we also further evaluate competing algorithms on RealHaze real-world dataset, including DCP [6], GFN [22], CGAN [35] and CycleGAN [35]. As shown in Figure 9, it can be observed that the proposed method still exhibits remarkable performance, and it can recover more clear images with truthful structures. This benefites from the contribution of the loss committee we established. In general, the above experiments reveal our method performs well on image dehazing in various real hazy scenarios, demonstrating both the generality and effectiveness of the proposed ACC-GAN.

E. ABLATION STUDIES

We study the impact of different loss components, asymmetric architecture, and negatives choices on the final dehazing performance. All the studies are performed in the same environment by using the RESIDE dataset.

1) STUDY ON LOSS COMMITTEE

To better demonstrate the effectiveness of our loss committee, we regularly remove one member to each configuration at one time. The values of PSNR are listed in Table 6. We observed that the best dehazing performance obtains 24.35 dB by using all the above loss terms, which indicates that each design strategy that has been considered has its own contribution to the final performance of ACC-GAN, especially the contrastive loss.

2) STUDY ON ASYMMETRIC ARCHITECTURE

We remove the feature transfer network that acted on the farward cycle path as the comparison model. As reported in Table 7, the proposed asymmetric CycleGAN architecture in ACC-GAN works better in PSNR than symmetrical model. Thanks to the further performance gains brought by the feature transfer network. The reason is that it helps to enhance encoded features from hazy domain to haze-free domain.

3) STUDY ON SELECTION OF NEGATIVES

The most important step in contrastive learning is how to select the negatives. In our work, we sample 256 negatives from within the same image (internal). Similar to [36], We further use external negatives instead of internal negatives to investigate the impact of different strategies. Figure 10 provides the visual dehazing results on a sample image. We can observe that the performance goes down sharply with external

(a) Hazy Input   (b) Internal Negatives   (c) External Negatives   (d) Ground Truth

FIGURE 10. Qualitative analysis of different negatives selection strategies.

TABLE 7. Quantitative analysis of different network architectures.

Models	w/o Feature Transfer Network	Ours
PSNR	23.66	**24.35**

negatives, which also proves that internal negatives are more effective than external.

F. LIMITATIONS

When training on small-scale datasets, our proposed method has limitations in dehazing performance, because contrastive learning often requires a large number of sample pairs to deliver satisfactory performance. In addition, another limitation is that the network training of GAN is unstable, and a large number of parameters need to be adjusted to converge.

V. CONCLUSION

In this paper, inspired by the significant advantage of contrastive learning, we are the first attempt to construct a novel Asymmetric Contrastive CycleGAN framework (called ACC-GAN) for aerial image dehazing. The proposed framework aims to learn an effective unsupervised dehazing model from an unpaired set of clear and hazy aerial images. Instead of only employing adversarial learning and cycle-consistency constraints, we introduce contrastive regularization to maximize the mutual information between the hazy domain and the haze-free domain. Furthermore, an additional feature transfer network is integrated into GAN to form asymmetric structure, which further enhances encoded features from hazy domain to haze-free domain. To avoid the dehazed results suffer from color-/structure- destroying effect, we combine multiple loss functions into a loss committee as the task-specific proxy guidance for the better jointly optimization. Experimental results on synthesis and real-world datasets considerably demonstrate that the effectiveness and generalization of our designed model. In future research work, we plan to explore the contribution of contrastive representation learning scheme in other low-level vision tasks, such as aerial image denoising and deblurring.

REFERENCES

[1] J. Liu, S. Wang, X. Wang, M. Ju, and D. Zhang, "A review of remote sensing image dehazing," *Sensors*, vol. 21, no. 11, p. 3926, Jun. 2021.

[2] R. Richter, "Atmospheric correction of satellite data with haze removal including a haze/clear transition region," *Comput. Geosci.*, vol. 22, no. 6, pp. 675–681, Jul. 1996.

[3] C. Feng, J.-W. Ma, Q. Dai, and X. Chen, "An improved method for cloud removal in ASTER data change detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2004, pp. 3387–3389.

[4] Y. Du, B. Guindon, and J. Cihlar, "Haze detection and removal in high resolution satellite image with wavelet analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 1, pp. 210–217, Jan. 2002.

[5] J. H. Jang, S. D. Kim, and J. B. Ra, "Enhancement of optical remote sensing images by subband-decomposed multiscale retinex with hybrid intensity transfer function," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 983–987, Sep. 2011.

[6] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Sep. 2010.

[7] A. Makarau, R. Richter, R. Müller, and P. Reinartz, "Haze detection and removal in remotely sensed multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, pp. 5895–5902, 2014.

[8] Y. Zhang, B. Guindon, and J. Cihlar, "An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images," *Remote Sens. Environ.*, vol. 82, nos. 2–3, pp. 173–187, 2002.

[9] D. Haboudane, J. R. Miller, E. Pattey, P. J. Zarco-Tejada, and I. B. Strachan, "Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture," *Remote Sens. Environ.*, vol. 90, no. 3, pp. 337–352, 2004.

[10] E. A. Cloutis, "Review article hyperspectral geological remote sensing: Evaluation of analytical techniques," *Int. J. Remote Sens.*, vol. 17, no. 12, pp. 2215–2242, Aug. 1996.

[11] J. Mustard and J. Sunshine, "Spectral analysis for earth science: Investigations using remote sensing data," in *Remote Sensing for the Earth Sciences: Manual of Remote Sensing*, vol. 3. New York, NY, USA: Wiley, 1999, pp. 251–307.

[12] Y. Li, C. Kong, L. Dai, and X. Chen, "Single-stage detector with dual feature alignment for remote sensing object detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[13] M. Mandal, M. Shah, P. Meena, and S. K. Vipparthi, "SSSDET: Simple short and shallow network for resource efficient vehicle detection in aerial scenes," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Taipei, Taiwan, Sep. 2019, pp. 3098–3102.

[14] P. Garg, A. S. Chakravarthy, M. Mandal, P. Narang, V. Chamola, and M. Guizani, "ISDNet: AI-enabled instance segmentation of aerial scenes for smart cities," *ACM Trans. Internet Technol.*, vol. 21, no. 3, pp. 1–18, Aug. 2021.

[15] F. Yang, J. Guo, H. Tan, and J. Wang, "Automated extraction of urban water bodies from ZY-3 multi-spectral imagery," *Water*, vol. 9, no. 2, p. 144, Feb. 2017.

[16] J. Guo, J. Yang, H. Yue, H. Tan, and K. Li, "RSDehazeNet: Dehazing network with channel refinement for multispectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2532–2549, Mar. 2021.

[17] S. K. Nayar and S. G. Narasimhan, "Vision in bad weather," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Kerkyra, Greece, Sep. 1999, pp. 820–827.

[18] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *Int. J. Comput. Vis.*, vol. 48, no. 3, pp. 233–254, 2002.

[19] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5197–5198, Nov. 2016.

[20] W. Ren, L. Si, Z. Hua, J. Pan, and M. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, Sep. 2016.

[21] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-One dehazing network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 4780–4788.

[22] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang, "Gated fusion network for single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 3253–3261.

[23] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, "Gated context aggregation network for image dehazing and deraining," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Waikoloa, HI, USA, Jan. 2019, pp. 1375–1383.

[24] X. Liu, Y. Ma, Z. Shi, and J. Chen, "GridDehazeNet: Attention-based multi-scale network for image dehazing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Nov. 2019, pp. 7313–7322.

[25] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M.-H. Yang, "Multi-scale boosted dehazing network with dense feature fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 2154–2164.

[26] D. Engin, A. Genc, and H. K. Ekenel, "Cycle-dehaze: Enhanced CycleGAN for single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Salt Lake City, UT, USA, Jun. 2018, pp. 825–833.

[27] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 3194–3203.

[28] Y. Huang, X. Chen, L. Xu, and K. Li, "Single image desmoking via attentive generative adversarial network for smoke detection process," *Fire Technol.*, vol. 57, no. 6, pp. 3021–3040, Nov. 2021.

[29] B. Huang, Z. Li, C. Yang, F. Sun, and Y. Song, "Single satellite optical imagery dehazing using SAR image prior based on conditional generative adversarial networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Snowmass, CO, USA, Mar. 2020, pp. 1795–1802.

[30] Y. Li and X. Chen, "A coarse-to-fine two-stage attentive network for haze removal of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 10, pp. 1751–1755, Oct. 2021.

[31] X. Chen, Y. Li, L. Dai, and C. Kong, "Hybrid high-resolution learning for single remote sensing satellite image dehazing," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[32] H. Jiang and N. Lu, "Multi-scale residual convolutional neural network for haze removal of remote sensing images," *Remote Sens.*, vol. 10, no. 6, p. 945, 2018.

[33] A. Hu, Z. Xie, Y. Xu, M. Xie, L. Wu, and Q. Qiu, "Unsupervised haze removal for high-resolution optical remote-sensing images based on improved generative adversarial networks," *Remote Sens.*, vol. 12, no. 24, p. 4162, Dec. 2020.

[34] A. Mehta, H. Sinha, M. Mandal, and P. Narang, "Domain-aware unsupervised hyperspectral reconstruction for aerial image dehazing," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Waikoloa, HI, USA, Jan. 2021, pp. 413–422.

[35] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2242–2251.

[36] T. Park, A. Efros, R. Zhang, and J. Zhu, "Contrastive learning for unpaired image-to-image translation," in *Proc. Eur. Conf. Comput. Vis.*, A. Vedaldi, H. Bischof, T. Brox, and J. Frahm, Eds. Berlin, Germany: Springer, Nov. 2020, pp. 319–345.

[37] J. Han, M. Shoeiby, L. Petersson, and M. A. Armin, "Dual contrastive learning for unsupervised image-to-image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Nashville, TN, USA, Jun. 2021, pp. 746–755.

[38] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma, "Contrastive learning for compact single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 10546–10555.

[39] C. Grohnfeldt, M. Schmitt, and X. Zhu, "A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from Sentinel-2 images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Valencia, Spain, Jul. 2018, pp. 1726–1729.

[40] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2020, pp. 1597–1607.

[41] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 9726–9735.

[42] O. Hénaff, A. Srinivas, J. Fauw, A. Razavi, C. Doersch, S. Eslami, and A. Oord, "Data-efficient image recognition with contrastive predictive coding," in *Proc. 37th Int. Conf. Mach. Learn.*, Jun. 2019, pp. 4182–4192.

[43] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018, *arXiv:1807.03748*.

[44] J. Han, M. Shoeiby, T. Malthus, E. Botha, J. Anstee, S. Anwar, R. Wei, L. Petersson, and M. A. Armin, "Single underwater image restoration by contrastive learning," 2021, arXiv:2103.09697.

[45] O. J. Hénaff, A. Srinivas, J. De Fauw, A. Razavi, C. Doersch, S. M. A. Eslami, and A. van den Oord, "Data-efficient image recognition with contrastive predictive coding," 2019, arXiv:1905.09272.

[46] X. Chen, J. Pan, K. Jiang, Y. Li, Y. Huang, C. Kong, L. Dai, and Z. Fan, "Unpaired deep image deraining using dual contrastive learning," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2022, pp. 2017–2026.

[47] S. Zheng, J. Sun, Q. Liu, Y. Qi, and S. Zhang, "Overwater image dehazing via cycle-consistent generative adversarial network," in Proc. Asian Conf. Comput. Vis. (ACCV), 2020, pp. 1–16.

[48] J.-G. Kwak, Y. Jin, Y. Li, D. Yoon, D. Kim, and H. Ko, "Adverse weather image translation with asymmetric and uncertainty-aware GAN," 2021, arXiv:2112.04283.

WANFENG JI received the Ph.D. degree in military operations research from the Academy of Military Sciences, China, in 2010. He is currently an Associate Professor with Naval Aviation University. His current research interests include vessel command technology and route planning.

XIN HE received the M.S. degree in transportation engineering from Dalian Maritime University, China, in 2017. He is currently a Senior Lecturer with Naval Aviation University. His current research interests include remote sensing image processing and computer vision.

JINPENG XIE received the M.S. degree in transportation engineering from the Wuhan University of Technology, China, in 2020. He is currently a Senior Lecturer with Naval Aviation University. His current research interests include vessel traffic risk visualization and autonomous collision avoidance.

• • •