Unsupervised Haze Removal for Aerial Imagery Based on Asymmetric Contrastive CycleGAN

XIN HE¹, WANFENG JI¹, AND JINPENG XIE¹
¹School of Basic Sciences for Aviation, Naval Aviation University, Yantai 264001, China
Corresponding author: Xin He (e-mail: 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 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...
where $\beta$ and $I_0$ denote the haziness and the clear image, respectively. Note that, there are two critical parameters: $A$ refers to the atmospheric light and $t(x)$ is termed as the medium transmission map. Under the homogeneous haze, the map of transmitting process $t(x)$ is defined as:

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

where $\beta$ and $d(x)$ denote the atmospheric 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 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], H2RLNet [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:

$$I(x) = F(J(x)),$$

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 laborable to collect paired synthetic 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 example in which the hazy 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 succeeded 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 CycleGAN 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.
The features of each example 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 \( F_{\text{pair}}: D_{\mathbb{D}}^{w \times h} \rightarrow D_{\mathbb{D}}^{W \times H} \) based on the paired dataset \( \mathcal{D}_{\text{pair}} = \{(I_h, I_n)\}_{i=1}^{N} \), so that the dehazed image

\[
I_n = F_{\text{pair}}(I_h) \in \mathbb{D}_n^{W \times H},
\]

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

\[
I_n = F_{\text{unpair}}(I_h) \in \mathbb{D}_n^{W \times H},
\]

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 cycle-consistency branch \( h \rightarrow n \rightarrow h' \), 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 \( n \rightarrow h_n \rightarrow n' \), 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.
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 |
|            | 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 |
|             | Upampling-1   | DeConv+SN+ReLU | 256,128,128 | 128,256,256 |
|             | Upampling-2   | DeConv+SN+ReLU | 128,256,256 | 64,512,512 |
| Decoder    | Conv          | RP+Conv+SN+Tanh | 64,512,512 | 3,512,512 |
| Feature    | Residual Block-10| Conv+IN+ReLU+Conv+IN+RC | 256,128,128 | 256,128,128 |
| Transfer   | Residual Block-11| Conv+IN+ReLU+Conv+IN+RC | 256,128,128 | 256,128,128 |
| Network    | Residual Block-12| Conv+IN+ReLU+Conv+IN+RC | 256,128,128 | 256,128,128 |

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 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 away.

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:

\[
\text{sim} \left( u, v \right) = \frac{\mathbf{u}^\top \mathbf{v}}{\| \mathbf{u} \| \| \mathbf{v} \|}.
\]

(7)

In formula (6), \( \tau \) 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. After having a stack of features \( \{z_l\}_L = \left( G_{enc}^l \left( x \right) \right)_L \), we select a query each time, refer the corresponding positive as \( z_1^+ \), and other negatives as \( z_i^{\setminus 1\, \setminus s} \). 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

\[
\mathcal{L}_{\text{CONTR}} (G, H) = \mathbb{E}_{h \sim H} \sum_{i=1}^{L} \sum_{s=1}^{S_i} \ell \left( z_1^+, z_i^+, z_i^{\setminus 1\, \setminus s} \right).
\]

(8)

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

\[
\mathcal{L}_{\text{CONTR}} (F, N) = \mathbb{E}_{n \sim N} \sum_{i=1}^{L} \sum_{s=1}^{S_i} \ell \left( z_1^+, z_i^+, z_i^{\setminus 1\, \setminus s} \right).
\]

(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...
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:

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

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

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

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:

$$L_{cycle} = \mathbb{E}_{n \sim \mathcal{N}} [\|n - n^*\|_1] + \mathbb{E}_{h \sim \mathcal{H}} [\|h - h^*\|_1] . \quad (13)$$

3) Contrastive Loss
As mentioned above, the overall contrastive loss can be defined as:

$$L_{cont} = L_{cont_h}(G, H) + L_{cont_n}(F, N). \quad (14)$$

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:

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

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:

$$L_{per} = \|\varphi_1(h) - \varphi_1(F(G(h)))\|_2^2 + \|\varphi_2(r) - \varphi_2(G(F(r)))\|_2^2 , \quad (16)$$

where $\varphi_1(\cdot)$ denotes the features extracted from $l$-th layer of the pretrained CNN.

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

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. For the better performance, we train our model for total 200 epochs. The proposed network is trained from the scratch for 100

| Datasets | RESIDE | SateHaze1k | Overwater | RealHaze |
|----------|--------|------------|-----------|----------|
| Train-Set | Hazy Images | 2903 | Thin | 355 | Moderate | 355 | Thick | 355 | 2090 | 0 |
| Clear Images | 3577 | 355 | 355 | 355 | 2441 | 0 |
| Test-Set | Hazy Images | 100 | 45 | 45 | 45 | 127 | 30 |
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 $286 \times 286$ resolution and randomly crop them into $256 \times 256$ patches during training. For the selection of negative samples, we sample 256 negatives from within the same 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 perform 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.

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.
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      |

FIGURE 8. Qualitative evaluations on the Overwater real-world dataset. Please zoom in the figures to better see the comparison.

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

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|
When training on small-scale datasets, our proposed method external negatives, which also proves that internal negatives can observe that the performance goes down sharply with truthful structures. This benefits 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 forward 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 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. CONCLUSIONS

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, pp. 3926, 2021.
[2] R. Richter, “Atmospheric correction of satellite data with haze removal including a haze/clear transition region,” Comput. Geosci., vol. 22, pp. 675-681, 1996.
[3] C. Feng, J. Ma, Q. Dai, and X. Chen, “An improved method for cloud removal in ASTER data change detection,” in Proceedings of the 2004 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), September, 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, pp. 210-217, 2002.
[5] J. Jang, S. Kim, and J. Ra, “Enhancement of optical remote sensing images by subband-decomposed multiscale retinex with hybrid intensity transfer function,” IEEE Geosci. Remote Sens. Lett., vol. 8, pp. 983-987, 2011.
B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, “DehazeNet: An End-to-End Networks,” IEEE Trans. Geosci. Remote Sens., vol. 52, pp. 5895-5902, 2014.

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, pp. 173-187, 2002.

D. Haboudane, J. Miller, E. Pattee, P. Zarco-Tejada, and I. 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, pp. 337-352, 2004.

E. Cloutis, “Review Article Hyperspectral geologic remote sensing: evaluation of analytical techniques,” Int. J. Remote Sens., vol. 17, pp. 2215-2242, 1996.

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, New John Wiley: New York, 1999, vol. 3, pp. 251-307.

Y. Li, C. Kong, L. Dai, X. Chen, “Single-Stage Detector With Dual Feature Alignment for Remote Sensing Object Detection,” IEEE Geosci. Remote S., vol. 19, pp. 1-5, 2022.

M. Mandal, M. Shah, P. Meena, and S. Vipparthi, “SSSDET: Simple Short and Shallow Network for Resource Efficient Vehicle Detection in Aerial Scenes,” in 2019 IEEE International Conference on Image Processing (ICIP), Sept., IEEE: Taipei, Taiwan, 2019, pp. 3098-3102.

P. Garg, A. 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, pp. 1-18, 2021.

F. Yang, J. Guo, H. Tan, and J. Wang, “Automated Extraction of Urban Water Bodies from ZY-3 Multi-Spectral Imagery,” Water, vol. 9, pp. 144, 2017.

J. Guo, J. Yang, H. Yue, H. Tan, and K. Li, “RSDHeazenet: Dehazing Network With Channel Refinement for Multispectral Remote Sensing Images,” IEEE Trans. Geosci. Remote S., vol. 59, pp. 2532-2549, 2020.

S. Nayar, and S. Narasimhan, “Vision in bad weather,” in Proceedings of the Seventh IEEE International Conference on Computer Vision, Sept., IEEE: Krykya, Greece, 1999, pp 820-827.

S. Narasimhan, and S. Nayar, “Vision and the Atmosphere,” Int. J. Comput. Vis., vol. 48, pp. 233-254, 2002.

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, pp. 5197-5198, 2016.

W. Ren, L. Si, Z. Hua, J. Pan, and M. Yang, “Single Image Dehazing via Multi-scale Convolutional Neural Networks,” in European Conference on Computer Vision, Berlin, Heidelberg, 17 Sept., 2016; Springer, Cham: Berlin, Heidelberg, pp. 243-257, 2016.

B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, “AOD-Net: All-in-One Dehazing Network,” in 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22-29 Oct., IEEE: Venice, Italy, 2017, pp 4780-4788.

W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M. Yang, “Gated Fusion Network for Single Image Dehazing,” in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18-23 June, IEEE: Salt Lake City, UT, USA, 2018, pp 3253-3261.

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 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikola, HI, USA, 7-11 Jan., IEEE: Waikola, HI, USA, 2019, pp 1375-1383.

Liu, Y. Ma, Z. Shi, and J. Chen, “GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing,” in 2019 IEEE Conference Internation Conference on Computer Vision (ICCV), Seoul, Korea (South), 27 Oct.-2 Nov., IEEE: Seoul, Korea (South), 2019, pp 7313-7322.

H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M. Yang, “Multi-scale boosted dehazing network with dense feature fusion,” in 2020 IEEE/CVF conference on computer vision and pattern recognition, Seattle, WA, USA, June 13-19, IEEE: Seattle, WA, USA, 2020, pp 2154-2161.

D. Engin, A. Genc, and H. Ekenel, “Cycle-Dehaze: Enhanced CycleGAN for Single Image Dehazing,” in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18-22 June, IEEE: Salt Lake City, UT, USA, 2018, pp 938-9388.
IEEE Conference on Computer Vision and Pattern Recognition (CVPR),
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
Proceedings of the Asian Conference on Computer Vision (ACCV), 2020.

[48] K. Jeong, Y. Jin, Y. Li, D. Yoon, D. Kim, and H. Ko. “Adverse Weather
Image Translation with Asymmetric and Uncertainty-aware GAN,” in
arXiv preprint arXiv:2112.04283 2021.

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 are remote sensing image processing and computer vision.

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.

JINPENG XIE received the M.S. degree in transportation engineering from 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.

***