

# Dehaze-AGGAN: Unpaired Remote Sensing Image Dehazing Using Enhanced Attention-Guide Generative Adversarial Networks

Yitong Zheng, Jia Su<sup>✉</sup>, Member, IEEE, Shun Zhang<sup>✉</sup>, Member, IEEE,  
Mingliang Tao<sup>✉</sup>, Senior Member, IEEE, and Ling Wang

**Abstract**—Remote sensing image dehazing is of great scientific interest and application value in both military and civil fields. In this article, we propose an enhanced attention-guide generative adversarial network (GAN) network, Dehaze-AGGAN, to solve the remote sensing images dehazing problem, which does not require paired training data. Since haze images have a great influence on remote sensing object detection, the dehazing of remote sensing images has become significantly important. Typical image dehazing methods require a hazy input image and its ground truth in a paired manner, while paired training data are usually not available in the field of remote sensing. To solve this problem, we propose the Dehaze-AGGAN network and train it by feeding unpaired clean and hazy images into the model. We present a novel total variation loss combined with the cycle consistency loss to eliminate wave noise and improve the target edge quality in the test dataset. Moreover, we present a new dehazing dataset called remote sensing dehazing dataset (RSD), which contains 7000 simulate and real hazy images including 3500 warship images and 3500 civilian ship images, and evaluate our method in the dataset. We conduct experiments on RSD. Extensive experiments demonstrate that the proposed Dehaze-AGGAN is effective and has strong robustness and adaptability in different settings.

**Index Terms**—Attention guided, dehaze, generative adversarial networks (GANs), total variation loss.

## I. INTRODUCTION

### A. Background

SATELLITE imagery has great importance in various applications during the development of remote sensing technology, such as Earth observation, climate change, and environmental monitoring. However, optical remote sensing is usually affected by fog. Fog in the air will reduce the information entropy of the image and limit the use of these optical remote sensing images in other applications. Since most object detection models and methods are trained on datasets under sunny and fog-free conditions, to obtain better results in object detection, they require a certain degree of enhanced visualization [1].

Manuscript received 14 June 2022; revised 22 August 2022; accepted 1 September 2022. Date of publication 7 September 2022; date of current version 27 September 2022. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62171379 and Grant 62271409 and in part by the Innovation Capability Support Program of Shaanxi under Grant 2021KJXX-99. (Corresponding authors: Jia Su; Ling Wang.)

The authors are with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: jiasu1011@126.com; lingwang@nwpu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2022.3204890

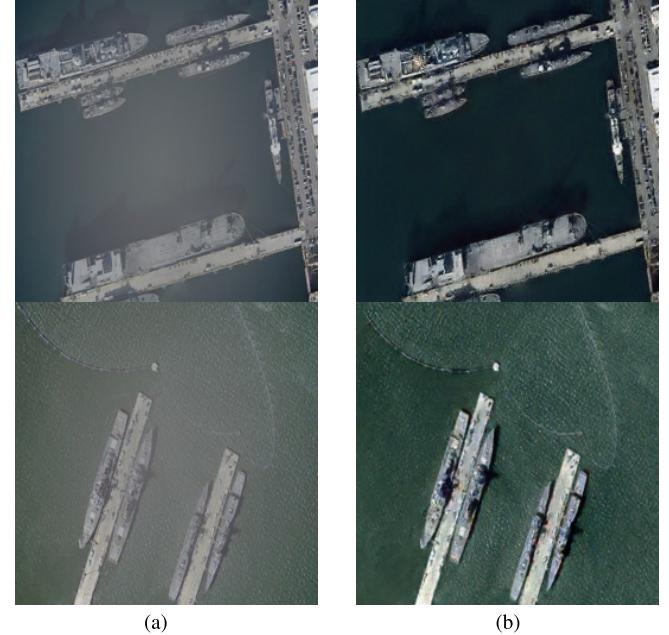


Fig. 1. Dehazing results on remote sensing hazy images. (a) Hazy images. (b) Dehazed images.

This pretreatment is often referred to as image dehazing. Image dehazing technology can further provide fog-free images for remote sensing ship detection, which is significantly important (as shown in Fig. 1).

### B. Related Works

The early research on single-image dehazing typically can be classified into prior information-based methods and learning-based methods. Prior information-based methods are mainly based on estimating the transmission map, such as dark channel priors [2], color attenuation prior [2], and haze-line prior [3], [4], [5]. Unfortunately, incorrect transmission estimations caused by incorrect prior information are easy to occur. Therefore, restoring the image quality is not an ideal method.

To solve this problem, neural networks are proposed to estimate the mapping parameters or directly generate the defogging images. Years of research in computer vision, image processing, computational photography, and graphics have produced powerful dehazing systems in the supervised setting,

where example image pairs are available, e.g., [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20].

Relying on a large number of paired real hazy images and haze-free images has become an important factor of restricting the development of deep learning in dehazing. However, paired data are usually difficult or expensive to acquire.

Years of research in computer vision, image processing, computational photography, and graphics have produced powerful translation systems in the supervised setting, where example image pairs. Fortunately, CycleGAN [21] estimates the transmission map with unpaired data. DualGAN [22], DiscoGAN [23], and SMAPGAN [24] use CycleGAN architecture and fine-tuning some details to achieve better results. However, in the remote sensing scenario, CycleGAN [21] is often influenced by background changes and lightly influenced by the structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) values.

To overcome the unwanted changes and the influences caused by background changes, AttentionGAN [25] was proposed for unpaired image translation tasks. However, in the image dehazing task, AttentionGAN [25] cannot converge (we will show the result in Section III). We summarize our concerns in the following four areas to describe the problems we encountered and the corresponding solutions.

1) *Single Image Dehazing*: As we have described above, the early research on single-image dehazing can be typically divided into two parts: prior information-based methods and learning-based methods. Prior information-based methods are mainly based on predicting the transfer mapping. Physical scattering models are usually presented as follows:

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

where  $I(x)$  represents a foggy photograph. In the meantime,  $J(x)$  represents the haze-free image,  $A$  denotes global atmospheric light, and  $t(x)$  denotes transmission mapping

$$t(x) = \exp^{-\beta d(x)} \quad (2)$$

where  $\beta$  is the scattering coefficient and  $d(x)$  is the distance between the observer and the camera. In most dehazing methods, these methods attempt to predict the transmission mapping and the atmospheric light.

Due to being inconsistent with the actual scene, these image priors can lead to inaccurate transmission approximations. To solve this problem, learning-based methods [CNNs and generative adversarial networks (GANs)] are proposed to estimate the parameters or directly generate the defogging image. The models based on convolutional neural networks mostly concentrate on predicting transmission mapping [3] to recover clean images. On the contrary, GANs mainly focus on generating clean images, and these models try to predict the parameters of the scattering model [26].

2) *Generative Adversarial Networks*: They are enlightened by a zero-sum theory [27]. GAN's confrontational evolution methods can generate more realistic samples. For the reasons mentioned above, GAN has a more forceful imitating ability than traditional models. However, we need more object features to train a model based on GANs such as paired images. Especially in image dehazing task, GAN models need paired

input-output samples. However, the paired data in the real world are usually not available.

3) *Unpaired Image Dehazing*: To solve the problem of less-paired data, some recent works propose the unpaired dehazing models. The models learn the mapping function without the requirement of paired training data. For example, CycleGAN [21] learns the mappings between two hazed images and dehazed images. For single-image dehazing, Cycle-Dehaze [28], CDNet [29], and Dd-cyclegan [30] adopt the CycleGAN framework, which can be used for unpaired training.

However, in the remote sensing scenario, CycleGAN [21] often generates noise in the dehazed images, and the SSIM and PSNR are lightly influenced. In the image dehazing area, CycleGAN [21] cannot focus on the foggy area of the input information, which causes the noise pollution and feature loss.

4) *Attention-Guided Image-to-Image Translation*: To tackle this issue, a few methods add attention modules to enhance image dehazing. Most of these methods focus on the masks to get the key features from the input images. Through the masks, the model can get information from the key areas, which needs to be concentrated on. For example, Liang *et al.* [31] used the object mask annotations to train the model. Chen *et al.* [32] used an additional network to get attention masks. Tang *et al.* [25] used a CycleGAN structure and attention mechanism for image translation task.

Through our experiments, AttentionGAN [25] performs excellently in the unpaired image-to-image translation task. However, in a single-image dehazing task, AttentionGAN [25] cannot converge. To fix the limitation, we propose Dehaze-AGGAN, which not only generates attention masks by the generators but also estimates mapping between hazed images and dehazed images. We also combine cycle-consistency and total variation losses to avoid noise from the sea and target edges.

### C. Contributions

In this article, we summarize three problems (less-paired data problem, noise pollution problem, and feature spotlight problem) of the remote sensing dehazing and propose novel methods to solve the problems. The contributions of our work are summarized as follows.

- 1) To overcome the feature spotlight problem, we proposed enhanced attention-guided GANs for remote sensing images dehazing task and test on our dataset remote sensing dehazing dataset (RSD). We use a CycleGAN architecture to solve the unpaired problem of image translation. Our model contains two enhanced attention-guide generators and two enhanced attention-guide discriminators to focus on the most informative locations in the image.
- 2) As for the noise pollution problem, we present a novel loss function by combining cycle-consistency and total variation losses to restrain the noise caused by sea waves and improve the edge quality of the targets in remote sensing images. The total variation loss is often used in signal processing and signal denoising. Our innovation is

to combine cycle-consistency and total variation losses so that the noise caused by sea waves is able to be denoising.

- 3) We proposed a RSD to solve the less-paired data problem. RSD is aimed at single-image dehazing and ship detection of military and civil ships. The dataset includes 7000 annotated simulate and real hazy images of ships, which contains 3500 paired hazed and haze-free images.

The reminder of this article is structured as follows. Section II introduces our proposed methods and explains the principle clearly. We show comparison, object detection, and ablation study result in Section III. Finally, conclusions are drawn in Section IV.

## II. PROPOSED METHOD

We present three main problems in Section I (less-paired data problem, noise pollution problem, and feature spotlight problem). In our consideration, we use a CycleGAN architecture to empower the model to extract unpaired information and enhanced attention-guided architecture to focus on the most informative locations in the image. As for the noise pollution problem, we combine the total variation loss and cycle-consistency loss. Finally, RSD is proposed to solve the less-paired data problem.

### A. Enhanced Attention-Guided Generation

Dehaze-AGGAN is an improved method of AttentionGAN [25] for unpaired image dehazing. As can be demonstrated in Fig. 2, Dehaze-AGGAN architecture is composed of two generators  $G$  and  $F$  and two discriminators  $D_x$  and  $D_y$ . We enhanced the architecture of the generators and make it possible for this architecture to do image dehazing. To clean/add the haze and remove the noise, our method combines the cycle-consistency and total variation losses and gets better performance.

Like CycleGAN [21], we used two mappings between hazed images domain  $X$  and dehazed images domain  $Y$ , i.e.,  $x \rightarrow [A_x^e, C_x^e] \rightarrow G(x)$  and  $y \rightarrow [A_y^e, C_y^e] \rightarrow F(y)$ .  $A_x^e$  and  $A_y^e$  are the attention masks.  $C_x^e$  and  $C_y^e$  are the content masks.

Enhanced attention-guided generator generate  $n$  attention masks  $\{A_y^e\}_{e=1}^n$ ,  $\{A_x^e\}_{e=1}^n$  and  $n$  content masks  $\{C_y^e\}_{e=1}^n$ ,  $\{C_x^e\}_{e=1}^n$ . The attention masks  $A_x^e$  and  $A_y^e$  represent the ratio of each pixel in the content masks  $C_x^e$  and  $C_y^e$ . After that, we fuse the attention masks  $A_x^e$  and  $A_y^e$  and the content masks  $C_x^e$  and  $C_y^e$  to get the final images  $G(x)$  and  $F(G(x))$ .

$G(x)$  can be calculated as follows:

$$G(x) = \sum_{e=1}^n (C_y^e * A_y^e). \quad (3)$$

In the same way,  $F(G(x))$  can be calculated as

$$F(G(x)) = \sum_{e=1}^n (C_x^e * A_x^e). \quad (4)$$

As we can see in Fig. 2, generator's input image is  $x \in \mathbb{R}^{H \times W \times 3}$ . The outputs are attention masks ( $A_x^e$  and  $A_y^e$ ) and content masks ( $C_x^e$  and  $C_y^e$ ). Specifically, the attention mask

is  $A_x^e, A_y^e \in \{0, \dots, 1\}^{H \times W}$  and the content mask is  $C_x^e, C_y^e \in \mathbb{R}^{H \times W \times 1}$ .

Different from AttentionGAN [25], the model we proposed pays more attention on the attention masks and the content masks, which leads to a better performance in remote sensing dehazing. We directly use input images to generate attention masks and contents masks without encoding, which reduces the amount of calculation significantly. Thus, we can fuse the attention masks and content masks directly without real image information.

### B. Total Variation Losses

Total variational regularization is widely used in signal processing, which has important applications in noise pollution removal. The principle is that this noise cancellation technique is superior to simple techniques such as linear smoothing or median filtering, which can reduce noise as well as remove edges more or less. In contrast, even at low signal-to-noise ratio (SNR), total variation denoising is very effective at preserving the edges while smoothing out of the noise in flat regions.

In image processing, the total variation loss can make the image smooth. Thus, we use total variation loss to get rid of the noise caused by sea waves and improve the edge quality of the targets in remote sensing images, and we express the objective as

$$\mathcal{L}_{TV}(\hat{x}) = \frac{1}{hwc} \sum_{i,j,k} \sqrt{(\hat{x}_{i,j+1,k} - \hat{x}_{i,j,k})^2 + (\hat{x}_{i+1,j,k} - \hat{x}_{i,j,k})^2} \quad (5)$$

where  $h$ ,  $w$ , and  $c$  are the height, width, and channel of  $x$ , respectively.

### C. Adversarial Loss and Cycle-Consistency Loss

We use adversarial losses in our model. For the mapping function  $G : X \rightarrow Y$  and its discriminator  $D_Y$ , we express the objective as

$$\mathcal{L}_{GAN}(G, D_Y) = \mathbb{E}_{x \sim P_{\text{data}}(x)} [\log(1 - D_Y(G(x))] + \mathbb{E}_{y \sim P_{\text{data}}(y)} [\log D_Y(y)]. \quad (6)$$

For each image  $x$  from domain  $X$ , the dehazing cycle should be able to bring  $x$  back to the original image, i.e.,  $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ . We express the cycle-consistency loss as

$$\mathcal{L}_{\text{cycle}}(G, F) = \mathbb{E}_{x \sim P_{\text{data}}(x)} [\| F(G(x)) - x \|_1] + \mathbb{E}_{y \sim P_{\text{data}}(y)} [\| G(F(y)) - y \|_1]. \quad (7)$$

### D. Full Objective

Our full objective is

$$\mathcal{L} = \mathcal{L}_{GAN} + \lambda_{\text{cycle}} * \mathcal{L}_{\text{cycle}} + \lambda_{TV} * \mathcal{L}_{TV} \quad (8)$$

where  $\mathcal{L}_{GAN}$ ,  $\mathcal{L}_{\text{cycle}}$ , and  $\mathcal{L}_{TV}$  are adversarial loss, cycle-consistency loss, and total variation loss, respectively. The parameters  $\lambda_{\text{cycle}}$  and  $\lambda_{TV}$  are to control the relative importance of the two objectives.

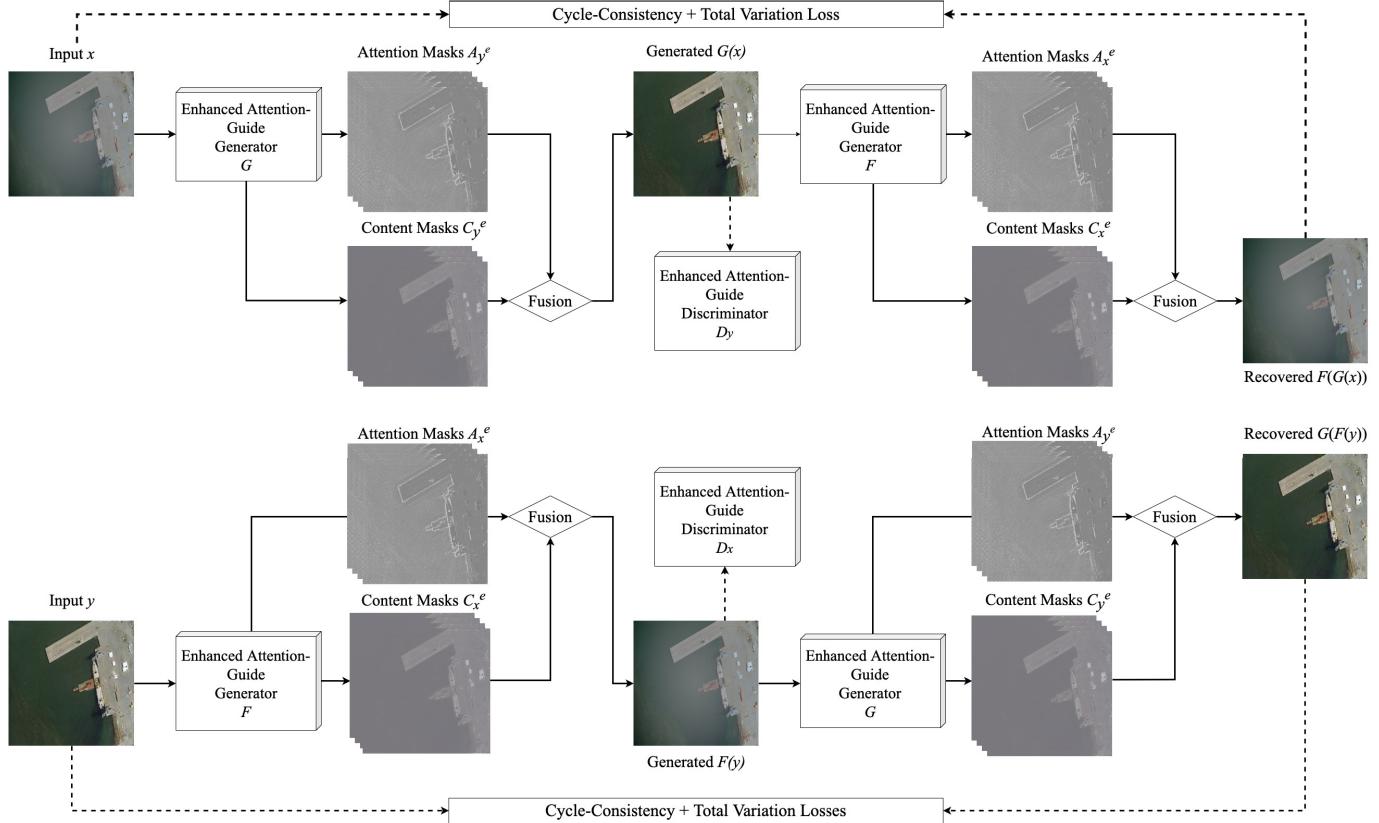


Fig. 2. Framework of the proposed Dehazed-AGGAN. Our model includes two generators  $G$  and  $F$ . We show two mappings in this figure, i.e.,  $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$  and  $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$ . The attention-guided generators can generate  $n - 1$  different content masks and attention masks, which helps us to estimate the dehazed images. We fuse the content masks and the attention masks to synthesize the final result.

### E. RSD: Remote Sensing Dehazing Dataset

We also propose a new remote sensing dehazing dataset, which is aimed at single-image dehazing and ship detection called RSD.

1) *Source of Images*: The RSD dataset includes 7000 annotated images of ships belonging to warships or civilian ships. We use the remote sensing images in Mcship [33], which contains six subcategories: aircraft carrier, submarine, landing ship, auxiliary ship, destroyer, and missile boat, while the civilian ships contain seven subcategories: sailboat, speedboat, fishing boat, passenger ship, container ship, tugboat, and support ship. The images are selected from major search engines, portals, forums, ship videos, and surveillance videos, and all kinds of ship pictures have been carefully selected and annotated at least five times. Table I summarizes the statistics of the dataset, including the images, objects, and their percentage of each ship category. We also add real hazy images to the test set to avoid data bias when training the network.

2) *Methods to Generate Paired Images*: We use the same method (atmospheric scattering model) to get paired images as in the dataset RESIDE [34], which uses a new large-scale benchmark consisting of both synthetic and real-world hazy images. In (1), we randomly set the values of  $A$  from 0.4 to 0.6 to ensure the simulation of various conditions of global atmospheric light. Also, in (2), we randomly set the values of  $\beta$  from  $5e^{-6}$  to  $7e^{-6}$  to ensure the simulation of various

TABLE I  
DATASET INFORMATION

Ship Type	Num	Percent	Annotated
Warship	3500	50%	PASCAL VOC
Civilian ship	3500	50%	PASCAL VOC
All	7000	100%	PASCAL VOC

conditions of foggy days. We set  $d(x)$  as the average distance of the remote sensing satellites from the ground. Through the settings above, we can get the haze image  $I(x)$  by input haze-free scene radiance  $J(x)$ . We use the add weight function of the CV2 Library to add  $A$  and  $\beta$  weights to the input haze-free scene radiance  $J(x)$ . Finally, we save  $I(x)$  as the output images as .png files.

### III. EXPERIMENTS AND RESULTS

To evaluate the single-image dehazing ability of our model, we show several experimental results in this section. We compare our result with the state-of-the-art single-image dehazing approaches first. Then, we show the object detection performance of ship on the RSD dataset and the convergence of the proposed algorithm. Finally, we conduct the ablation study as follows: 1) without total variation loss and 2) generator with the connection between fuse feature original image.

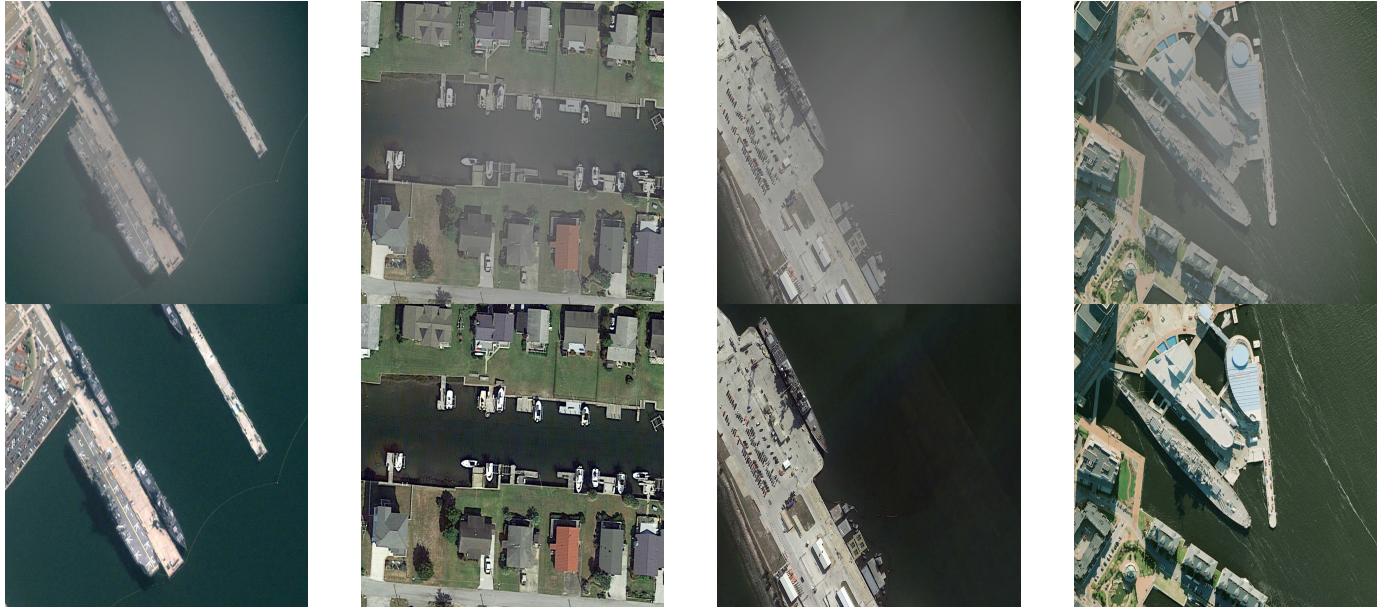


Fig. 3. Example images of our RSD datasets. RSD is aimed at single-image dehazing and ship detection of military and civil ships. The dataset includes 7000 annotated images of ships, which contains 3500 paired hazy and haze-free images.

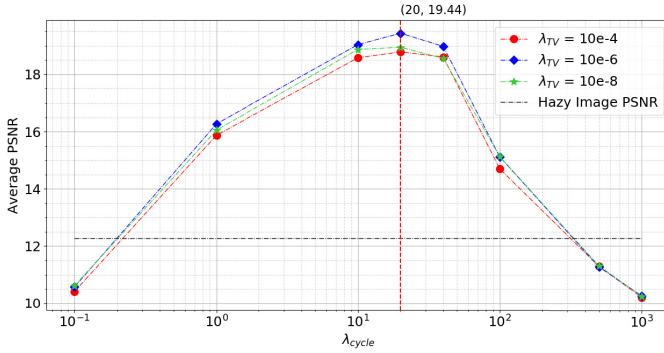


Fig. 4. Average PSNR of dehazed images that are generated by models with different  $\lambda_{cycle}$  and  $\lambda_{TV}$  values.

### A. Datasets

Because of lacking of remote sensing dehazing dataset, we propose a new dehazing dataset named RSD. RSD can be used for evaluating the tasks of single-image dehazing and detection of military and civil ships. The dataset includes 7000 simulate and hazy images of ships, which contains 3500 paired hazy and haze-free images synthesized by python (as shown in Fig. 3). We use our dataset to get qualitative and quantitative results by comparing with state-of-the-art results.

### B. Experiment Details

1) *Training Strategy*: We use the Pytorch framework to build our model. Around 20 epochs are performed on each dataset. Following the standard optimization method by GANs, we optimize the proposed network where we switch steps between a gradient decent one on generators and another one on discriminators in the field of training methods. Specifically, we set  $n = 10$  in our experiments. Moreover, we use a least-square loss to stabilize our model during the training

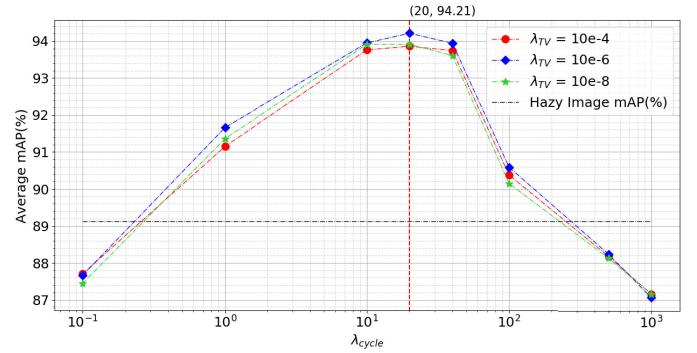


Fig. 5. Average YOLOv5 detection mAPs of dehazed images that are generated by models with different  $\lambda_{cycle}$  and  $\lambda_{TV}$  values.

procedure. We also use a history of generated images to update discriminators similar to CycleGAN.

2) *Parameter Setting*: For all datasets, images are rescaled to  $256 \times 256$ . We do left-right flip and random crop for data augmentation. We set the number of image buffer to 50. We use the Adam optimizer with the momentum terms  $\beta_1 = 0.5$  and  $\beta_2 = 0.999$ . We set  $\lambda_{cycle} = 20$  and  $\lambda_{TV} = 1e^{-6}$ . As shown in Figs. 4 and 5, we can get the best performance of the proposed method when  $\lambda_{cycle} = 20$  and  $\lambda_{TV} = 1e^{-6}$  (with PSNR = 19.44 and mAP = 94.1). It is obvious that the connection between PSNR and mAP is strong. With the increase of PSNR, the quality of the dehazed images is getting better, which causes the increase of mAP. Finally, we set  $\lambda_{cycle} = 20$  and  $\lambda_{TV} = 1e^{-6}$  for the experiments.

3) *Competing Models*: We have conducted our experiments on the RSD dataset to illustrate the performance of our method compared to the other state-of-the-art methods in unpaired single-image dehazing area. We choose AttentionGAN [25], CycleGAN [28], GCANet [35], and DehazeNet [11] for the experiments.

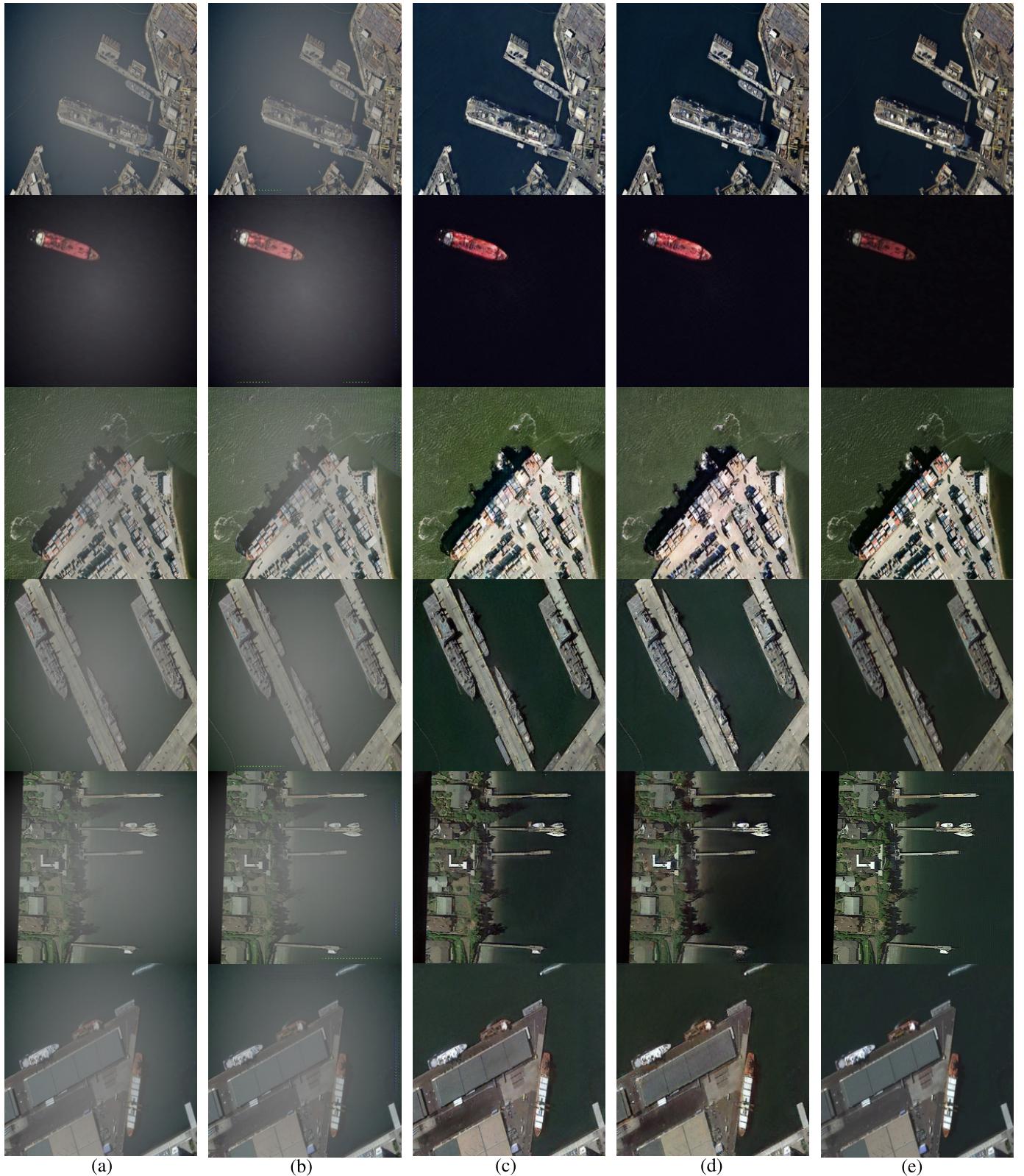


Fig. 6. Qualitative results on RSD datasets. We test our dataset on three baseline models, including our method. Order of results for each row is (a) hazy image, (b) AttentionGAN result, (c) CycleGAN result, (d) our result (proposed result), and (e) ground truth.

**4) Evaluation Metrics:** Whether we are doing experiments or engaging in other needs, some processing of the image may damage the quality of the image or change the content information. The change can be measured by the image quality

evaluation index. PSNR/SSIM, analyzed by A Hore [36], can meet the demand. In this article, we use PSNR/SSIM to measure the image quality. We also test the mAPs values on YOLOv5.

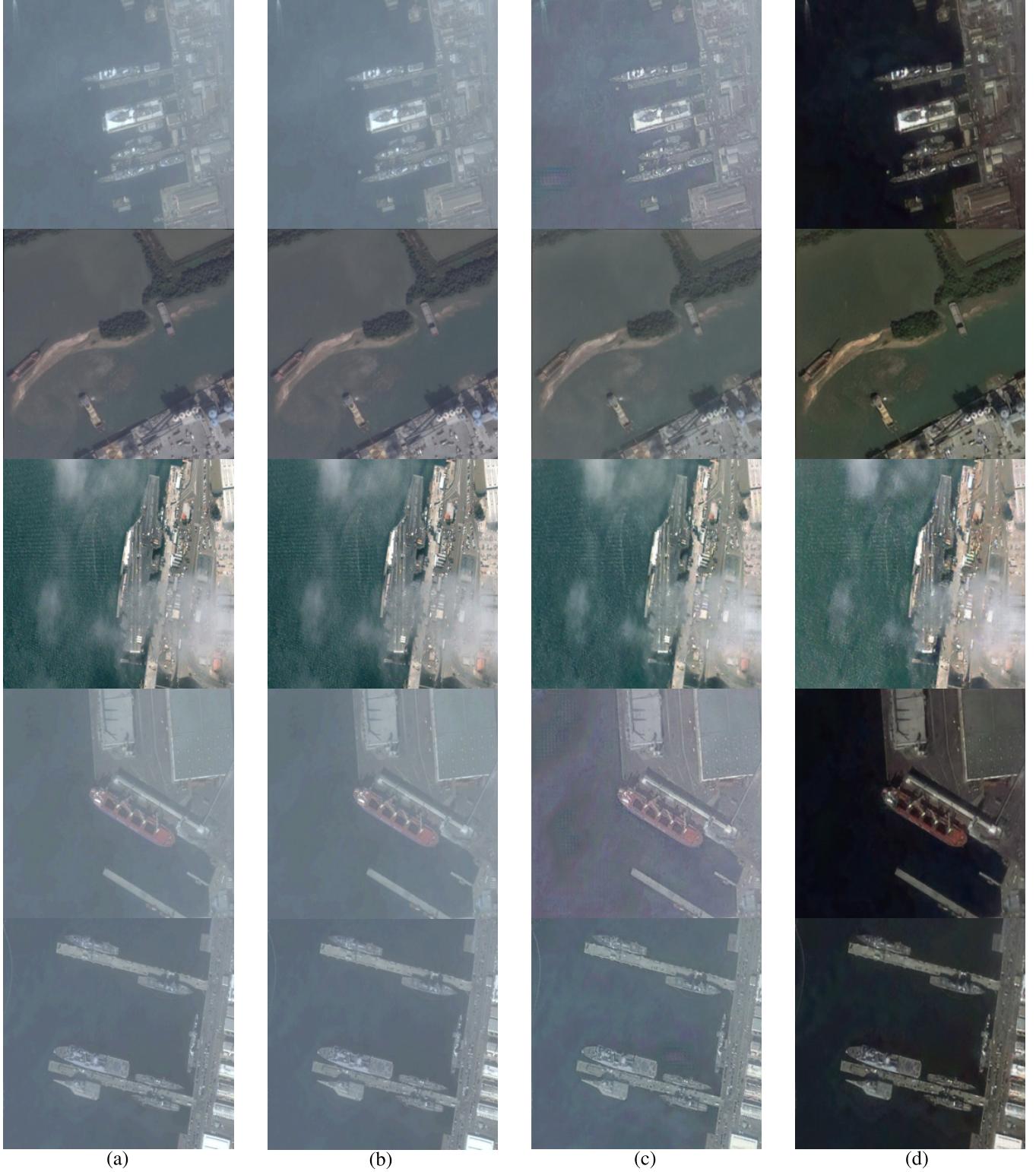


Fig. 7. Qualitative results on RSD datasets. We test real hazy images on three baseline models, including our method. The order of results for each row is (a) hazy image, (b) AttentionGAN result, (c) CycleGAN result, (d) our result (proposed result).

### C. Comparison With State-of-the-Art Approach

We have conducted our experiments on the RSD dataset to illustrate the performance of our method compared to the other state-of-the-art methods in unpaired

single-image dehazing areas. We choose AttentionGAN [25], CycleGAN [21], GCA-Net [35], and DehazeNet [11] for the experiments. The outcomes are presented in Figs. 6–8, and Tables II and III present the average PSNR/SSIM values. We also calculate the object detection results by a YOLOv5x

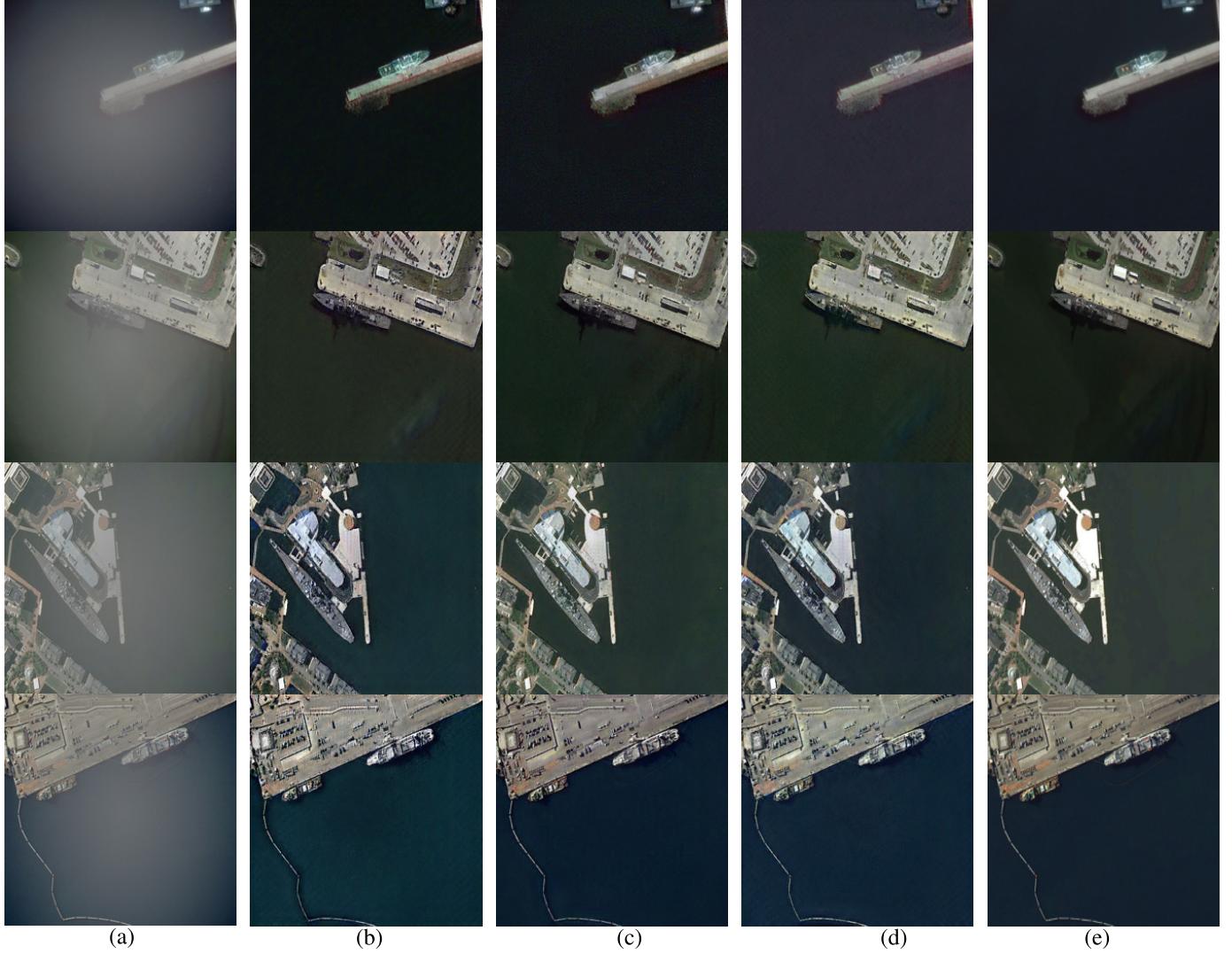


Fig. 8. Qualitative results on RSD datasets. We also test our dataset on some supervised models, including our method. Order of results for each row is (a) hazy image, (b) DehazeNet [11] result, (c) GCANet [35] result, (d) our result, and (e) ground truth.

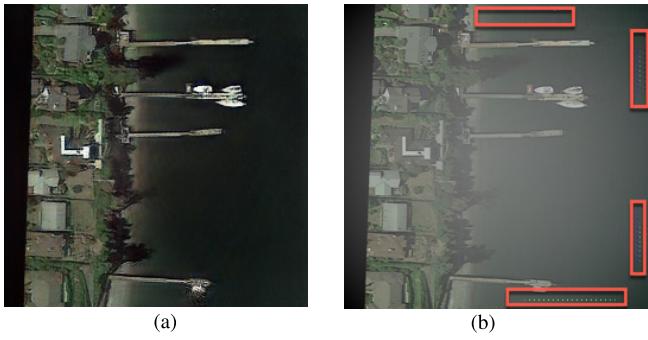


Fig. 9. Dehazing results on remote sensing hazy images. AttentionGAN result cannot remove the fog but generate small spots at the edge of the images (in red rectangle). (a) Proposed result. (b) AttentionGAN result.

detector after dehazing and the results are shown in Fig. 10 and Table V.

1) Qualitative Results: Fig. 6 shows the qualitative difference between CycleGAN [21], AttentionGAN [25], and Dehaze-AGGAN. It is obvious that AttentionGAN [25] cannot

remove the fog but generate small spots at the edge of the images, which is the reason why AttentionGAN [25] gets much higher PSNR. We can easily see that Dehaze-AGGAN has less noise and more details than CycleGAN [21], where total variation loss reduced the noise and enhanced structure leads to better detail collection ability. Specifically, Dehaze-AGGAN retains the original color toning after dehazing. Fig. 7 shows the qualitative results tested on real hazy images. It is obvious that AttentionGAN [25] and CycleGAN [21] cannot remove the fog. It can be easily seen that Dehaze-AGGAN could remove the haze well with the model trained on simulation images, which means that our model is applicable to most real-time remote sensing tasks. Fig. 8 shows the qualitative difference between GCANet [35], DehazeNet [11], and Dehaze-AGGAN. It is obvious that DehazeNet [11] restores the image with intact details while avoiding image oversaturation. At the same time, DehazeNet [11] does not restore well at depth-of-field changes, and information is easily lost. DehazeNet [11] has more parameters and is inefficient. It can be seen that GCANet [35] has less gridding artifacts from the dilated

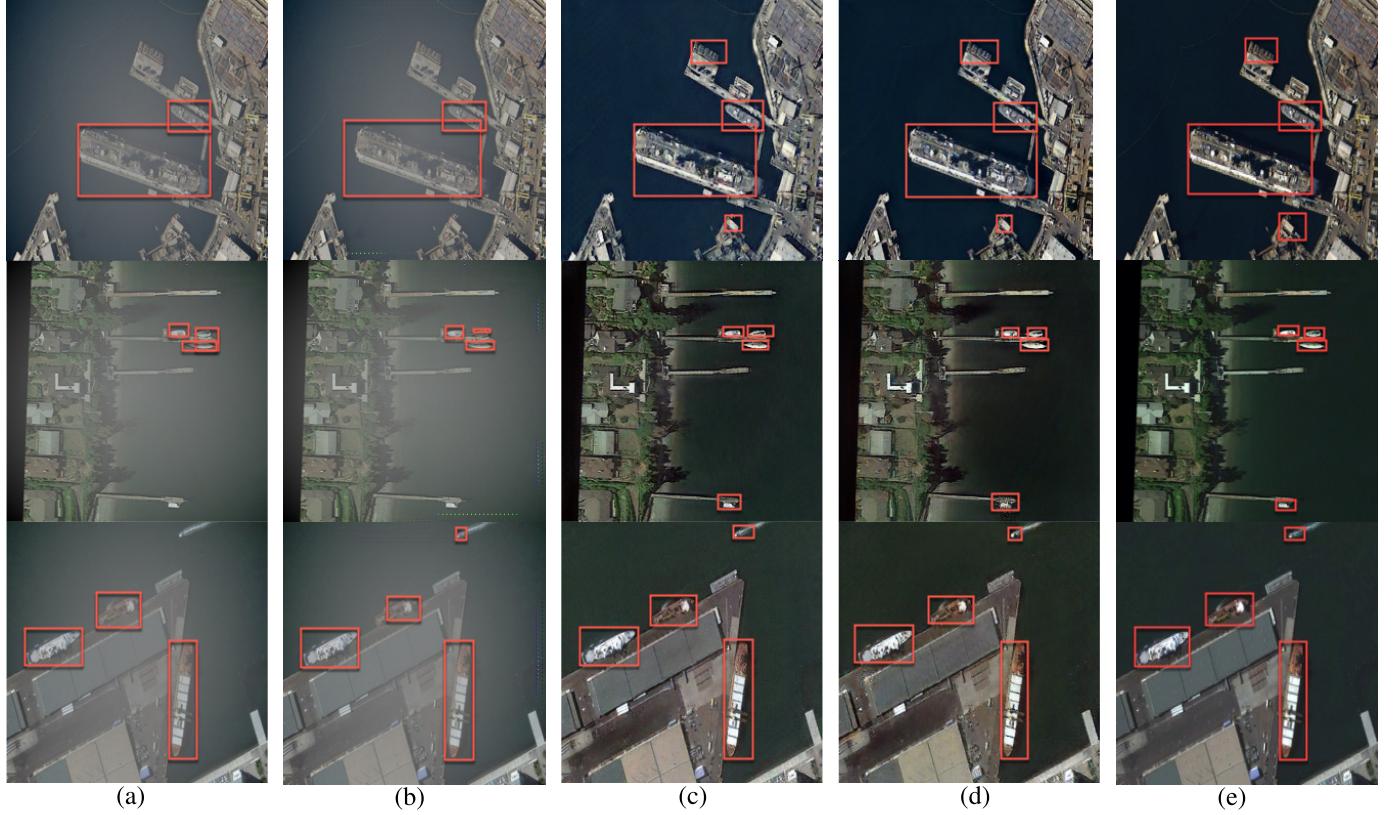


Fig. 10. Object detection results on dehazing results. We test the dehazing results with the baseline detector YOLOv5x. Order of results for each row is (a) hazy image, (b) AttentionGAN result, (c) CycleGAN result, (d) our result, and (e) ground truth.

TABLE II  
AVERAGE PSNR AND SSIM RESULTS ON RSD DATASETS  
(UNPAIRED SINGLE IMAGE DEHAZING METHODS)

Methods	Metrics	Validation	Test
None	PSNR/SSIM	13.22/0.65	12.27/0.59
CycleGAN [21]	PSNR/SSIM	19.56/0.88	18.61/0.87
AttentionGAN [25]	PSNR/SSIM	45.19/0.63	47.19/0.56
AttentionGAN(ws)	PSNR/SSIM	13.43/0.66	12.34/0.61
<b>Proposed</b>	PSNR/SSIM	<b>20.11/0.92</b>	<b>19.44/0.89</b>

convolution, where the latest smoothed dilated technique is used.

2) *Quantitative Results*: Table II presents the average PSNR and SSIM results on RSD datasets. The first row of the results shows the values that are average PSNR and SSIM results calculated directly between each hazy and its ground-truth image. (*ws* means without spots. We manually removed the light spots in the image and calculate the PSNR/SSIM again.) As we can see in Table II, our method performs better than CycleGAN [21] in terms of PSNR and SSIM values. This shows that our new structure can focus on the details of the objects and recover foggy images better, which leads to better performance on SSIM. The total variation loss also reduces the noise of the dehazed images, which leads to better performance on PSNR. However, AttentionGAN [25] gets much higher PSNR (45.19 on test set and 47.19 on validation set) values than the other methods. AttentionGAN [25] fusion

input image information into generated images, which leads attention masks and content masks, cannot converge and get better dehazing results, where AttentionGAN [25] generates some small spots at the edge of the images, which increases the maximum brightness.

We have marked the location of the bright spots that caused the PSNR to increase significantly in Fig. 9. Thus, we manually removed the light spots in images and calculate the PSNR/SSIM again (AttentionGAN(ws) in Table II), and the result has nearly the same PSNR/SSIM as hazy image, which indicates that AttentionGAN [25] is not available for image dehazing. In general, our model has a better performance in unpaired single-image dehazing.

As for paired remote sensing single-image dehazing methods, it is shown that the supervised methods, such as DehazeNet [11] and GCANet [35], achieve better PSNR and SSIM performances than the proposed unsupervised method in Table III. These two supervised learning methods build a knowledge base from the preclassified patterns that support classification and generate new patterns. In contrast with supervised learning, there are no explicit target outputs or environmental evaluations associated with each input for the unsupervised method. The unsupervised method brings to bear prior biases as to what aspects of the structure of the input should be captured in the output [37]. This explains why these supervised methods have better performance. However, in practical applications such as remote sensing image dehazing, the labeled image pairs are often unavailable, which

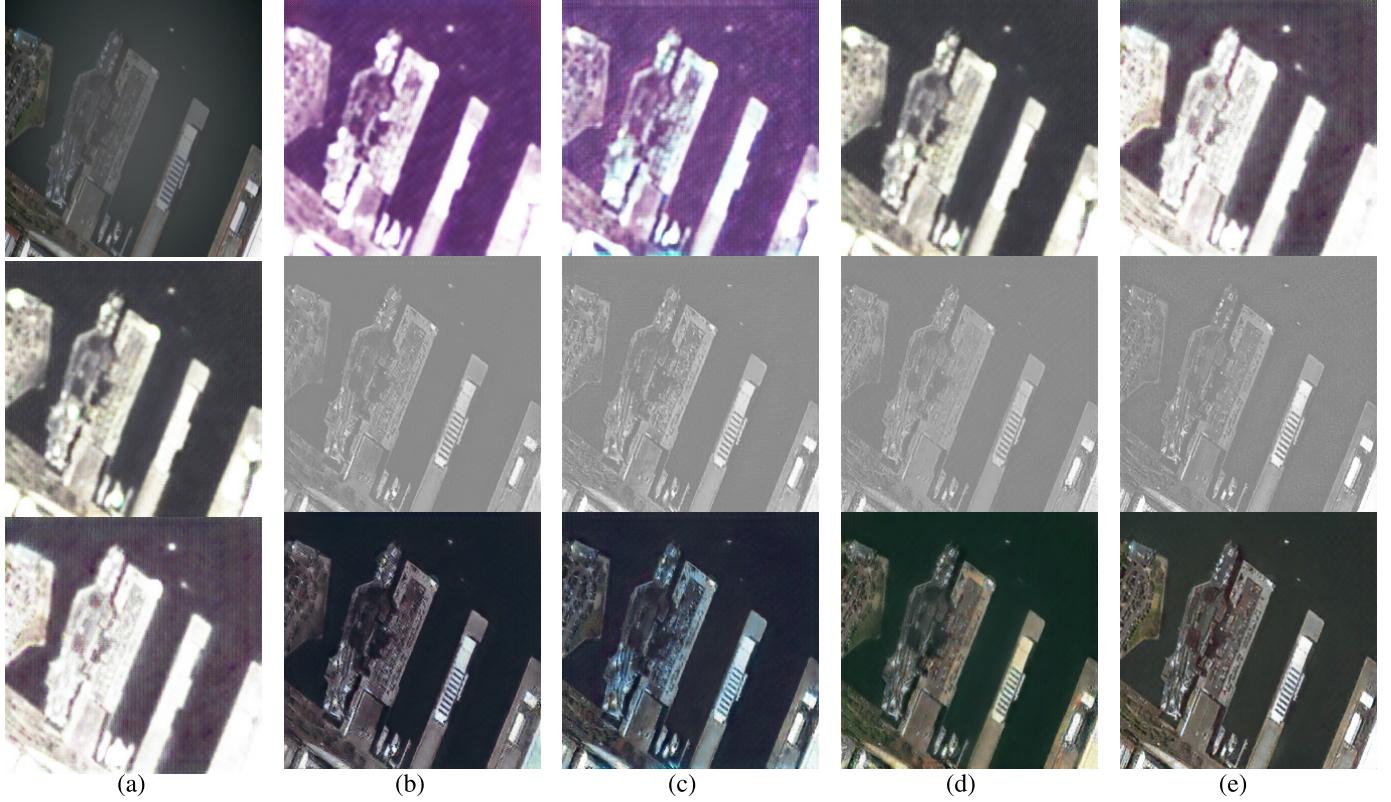


Fig. 11. Evolution of attention masks and content masks ( $n = 10$  and  $e = 1$ ). (From Top to Bottom) Content masks, attention masks, and output images. (a) Epoch 0. (b) Epoch 5. (c) Epoch 20. (d) Epoch 50. (e) Epoch 150.

TABLE III

AVERAGE PSNR AND SSIM RESULTS ON RSD DATASETS  
(PAIRED SINGLE IMAGE DEHAZING METHODS)

Methods	Metrics	Validation	Test
None	PSNR/SSIM	13.22/0.65	12.27/0.59
DehazeNet [11]	PSNR/SSIM	20.18/0.93	19.78/0.91
GCANet [35]	PSNR/SSIM	21.40/0.93	21.12/0.92
<b>Proposed</b>	PSNR/SSIM	<b>20.11/0.92</b>	<b>19.44/0.89</b>

limits the applicability of the supervised approaches, while the proposed unsupervised method shows promising results. Therefore, the impact of the slight performance gap could be alleviated and is considered acceptable.

#### D. Object Detection Results

Due to the fine-grained categorization of the RSD dataset, we also compare the object detection results of the images after dehazing by these models on the RSD dataset. We use YOLOv5 as a baseline detector and calculate the mAP of ship.

1) *YOLOv5 Training Details*: We trained the YOLOv5 detector on both hazy images and hazy-free images to compare with data augmentation in detector training. We follow the standard optimization method from the official YOLOv5 model and test the YOLOv5 detectors on both hazy images and hazy-free test set. Table IV shows the cross datasets' experiments with the YOLOv5 detector. It is obvious that the detector trained on hazy images can get better mAP

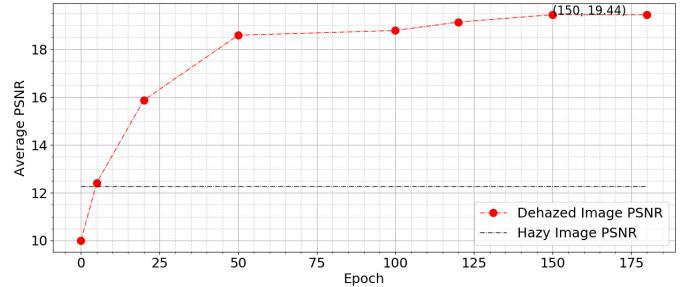


Fig. 12. Evolution of average PSNR with epochs.

TABLE IV  
YOLOv5 RESULTS WITH DIFFERENT TRAIN SETS AND TEST SETS

Train Set - Test Set	Precision	Recall	mAP(0.5:0.95)
Clear - Hazy	0.974	0.993	0.952
Clear - Clear	0.972	0.995	0.975
Hazy - Hazy	0.977	0.997	<b>0.979</b>
Hazy - Clear	0.953	0.981	0.919

performance on the hazy test set than the detector trained on clear images. Also, the mAP of YOLOv5 trained on hazy images is 0.979, which denotes that our proposed methods can enhance the detection performance and nearly eliminate the effect of haze weather on remote sensing object detection. Finally, we use the model train on hazy images to test the object detection results of the images dehazed by the state-of-the-art methods.

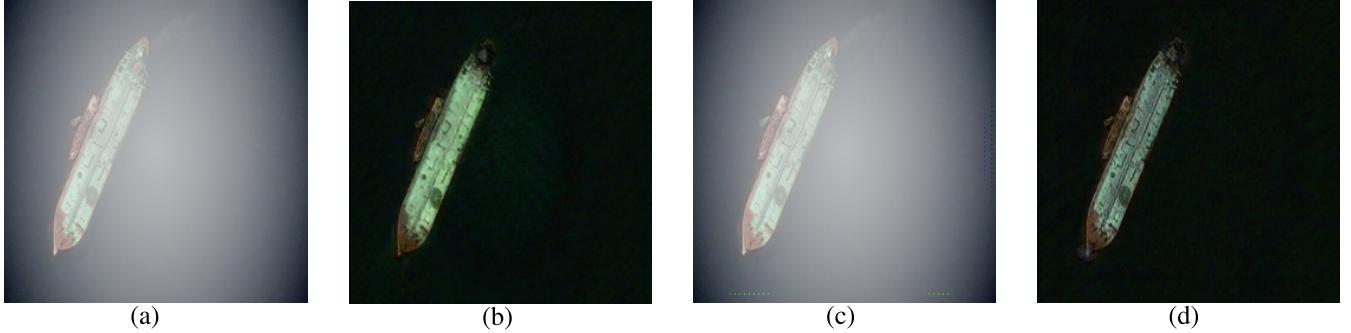


Fig. 13. Results of the ablation study. (a) Hazy image. (b) Without TVloss. (c) With image connection. (d) Proposed result.

TABLE V

OBJECT DETECTION PERFORMANCE (mAP AND THE IMPROVEMENT OF HAZY IMAGES) OF SHIP ON OUR RSD DATASET

	Hazy	AttentionGAN	CycleGAN	Proposed	GT
mAP(%)	89.12	89.78	92.38	<b>94.87</b>	95.09
Gain	-	+0.66	+3.26	<b>+5.75</b>	+5.97

2) *Qualitative Results:* Fig. 10 shows the qualitative difference between the object detection results of the models on the RSD dataset. Detection results of hazy image and AttentionGAN generate few false and missing detections. The dehazing results generated by CycleGAN and the proposed methods enable the YOLOv5 model to detect smaller targets and improve the accuracy of the detection frame. The result shows that the proposed methods can enhance the detection performance and nearly eliminate the effect of haze weather on remote sensing object detection.

3) *Quantitative Results:* Table V shows the experimental results in the field of mAP and the mAP improvements compared to hazy images. Our work performs best in all three methods and takes 5.75 higher than hazy images in mAP. The result shows that our method restores more image features and has a positive effect on remote sensing image recognition.

#### E. Convergence of the Proposed Algorithm

We evaluate the convergence of the proposed method by comparing the masks and PSNR performances of the proposed method with epochs.

1) *Qualitative Results:* Fig. 11 shows the evolution of the masks (attention and content) where mask numbers  $n = 10$  and  $e = 1$  represent the first mask in all masks. Fig. 11(a) shows the input hazy image in the RSD dataset and the original attention mask and content mask. In Fig. 11, the attention mask and content mask of epochs 5, 20, 50, and 150 are shown. We can see that with the training epoch rise, the attention masks become more clear and more detailed. The results show that our model is effective and evolvable.

2) *Qualitative Results:* Fig. 12 shows the evolution of the average PSNR of output dehazed images while training. We can see that with the training epoch rise, the average PSNR increases from 10 to 19.44. When the training epoch is 150, the average PSNR is 19.44. After epoch 150, with the epoch

TABLE VI

AVERAGE PSNR AND SSIM RESULTS ON RSD DATASETS IN ABLATION STUDY. (ws MEANS WITHOUT SPOTS. WE MANUALLY REMOVED THE LIGHT SPOTS IN THE IMAGE AND CALCULATE THE PSNR/SSIM AGAIN)

Methods	Metrics	Validation	Test
Hazy image	PSNR/SSIM	13.22/0.65	12.27/0.59
Without TVloss	PSNR/SSIM	15.41/0.87	16.12/0.83
With Connection	PSNR/SSIM	43.61/0.63	45.21/0.57
With Connection(ws)	PSNR/SSIM	13.43/0.66	12.34/0.61
<b>Ours</b>	<b>PSNR/SSIM</b>	<b>20.11/0.92</b>	<b>19.44/0.89</b>

rising, the average PSNR nearly remains unchanged. The result shows that our model converges after 150 epochs.

#### F. Results on Ablation Study

In order to indicate the utility of our loss committee, we implement an ablation study with two parts: 1) without total variation loss and 2) generator with the connection between fuse feature and original image.

1) *Qualitative Results:* We show the qualitative results in Fig. 13. Fig. 13(a) shows a hazy image in the RSD dataset. The outline of the ship is covered by fog, which produces a lot of noise and blurring. In Fig. 13(b), the methods without total variation loss result in poor light uniformity of the restored image and more noise than Fig. 13(d). Fig. 13(c) has the same spots as the result generated by AttentionGAN [25]. Also, the function with connection cannot remove the fog in this image from the RSD dataset.

It can be seen that Fig. 13(d) has less noise and more details than Fig. 13(b) and (c), where total variation loss reduced the noise and enhanced structure leads to better detail collection ability. The results demonstrate that our model is effective and evolvable.

2) *Quantitative Results:* As we can see in Table VI, our method gets better PSNR/SSIM values than hazy images and the methods without TVloss. However, the method with image connection gets a much higher PSNR (43.61 on test set and 45.21 on validation set) value than the other methods. The AttentionGAN [25] structure will lead to this result as we analyzed in Section III-C. We manually removed the light spots in images and calculate the PSNR/SSIM again. Finally, we can observe that the aggregate with all the two tricks

generates the **clearest result**, which indicates that our proposed method works.

#### IV. CONCLUSION

In this article, we propose an unpaired remote sensing dehazing network called Dehaze-AGGAN, which generate masks from the generators and discriminator and fuse the output image. Our contributions are listed in the following.

- 1) Experiments on the datasets show that our Dehaze-AGGAN achieves **remarkable performance** compared to other state-of-the-art methods. Extensive tests demonstrate that our Dehaze-AGGAN is **effective** and has strong robustness and adaptability in **different environments**.
- 2) We **combine cycle-consistency loss with total variation loss** to improve the output image quality. Results present that our **model generates better quality images**. Our methods not only get better **PSNR** and **SSIM** than CycleGAN [21] and AttentionGAN [25] architecture but also are **applicable to most** real-time remote sensing tasks.
- 3) Our **RSD dataset** is the first dataset in the **remote sensing dehazing area**. The dataset shows great potential for future research.
- 4) Although the dehazing performance shown by our method is the best in **unpaired single-image dehazing models**, there is still room to improve our level compared to other paired single-image dehazing model. We still believe that the convenience on **collecting data** and the adaption to more tasks of our methods can bridge these performance gaps, and we hope to achieve better results in the future.

#### REFERENCES

- [1] 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.
- [2] 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, Dec. 2011.
- [3] D. Berman *et al.*, "Non-local image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1674–1682.
- [4] D. Berman, T. Treibitz, and S. Avidan, "Air-light estimation using haze-lines," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, Oct. 2017, pp. 1–9.
- [5] 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, no. 9, pp. 5895–5905, Sep. 2014.
- [6] G. Xu, B. Zhang, J. Chen, F. Wu, J. Sheng, and W. Hong, "Sparse inverse synthetic aperture radar imaging using structured low-rank method," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2021.
- [7] 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. 2535–2549, Mar. 2021.
- [8] J. Guo, J. Yang, H. Yue, C. Hou, and K. Li, "Landsat-8 OLI multispectral image dehazing based on optimized atmospheric scattering model," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10255–10265, Dec. 2020.
- [9] D. Shen, J. Liu, Z. Wu, J. Yang, and L. Xiao, "ADMM-HFNet: A matrix decomposition-based deep approach for hyperspectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2021.
- [10] J. Yuan, Z. Cai, and W. Cao, "TEBCF: Real-world underwater image texture enhancement model based on blurriness and color fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2021.
- [11] B. L. Cai, X. M. Xu, K. Jia, C. M. Qing, and D. C. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Aug. 2016.
- [12] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 154–169.
- [13] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3194–3203.
- [14] J. Zheng, X.-Y. Liu, and X. Wang, "Single image cloud removal using U-Net and generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6371–6385, Oct. 2020.
- [15] 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.*, Oct. 2017, pp. 4770–4778.
- [16] R. Li, J. Pan, Z. Li, and J. Tang, "Single image dehazing via conditional generative adversarial network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8202–8211.
- [17] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced Pix2pix dehazing network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8160–8168.
- [18] W. Ren *et al.*, "Gated fusion network for single image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3253–3261.
- [19] X. Zhang, "Research on remote sensing image de-haze based on GAN," *J. Signal Process. Syst.*, vol. 94, no. 3, pp. 1–9, 2021.
- [20] X. Tian, K. Li, Z. Wang, and J. Ma, "VP-Net: An interpretable deep network for variational pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022.
- [21] 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.*, Oct. 2017, pp. 2223–2232.
- [22] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2849–2857.
- [23] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1857–1865.
- [24] X. Chen *et al.*, "SMAPGAN: Generative adversarial network-based semisupervised styled map tile generation method," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4388–4406, Sep. 2020.
- [25] H. Tang, H. Liu, D. Xu, P. H. S. Torr, and N. Sebe, "AttentionGAN: Unpaired image-to-image translation using attention-guided generative adversarial networks," 2019, *arXiv:1911.11897*.
- [26] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.
- [27] A. Rapoport, *Two-Person Game Theory*. Chelmsford, MA, USA: Courier Corporation, 2013.
- [28] D. Engin, A. Genç, and H. K. Ekenel, "Cycle-Dehaze: Enhanced CycleGAN for single image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 825–833.
- [29] A. Dudhane and S. Murala, "CDNet: Single image De-hazing using unpaired adversarial training," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 1147–1155.
- [30] J. Zhao *et al.*, "DD-CycleGAN: Unpaired image dehazing via double-discriminator cycle-consistent generative adversarial network," *Eng. Appl. Artif. Intell.*, vol. 82, pp. 263–271, Jun. 2019.
- [31] X. Liang, H. Zhang, and E. P. Xing, "Generative semantic manipulation with contrasting GAN," 2017, *arXiv:1708.00315*.
- [32] X. Chen, C. Xu, X. Yang, and D. Tao, "Attention-GAN for object transfiguration in wild images," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 164–180.
- [33] Y. Zheng and S. Zhang, "Mcships: A large-scale ship dataset for detection and fine-grained categorization in the wild," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2020, pp. 1–6.
- [34] B. Li *et al.*, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.
- [35] D. Chen *et al.*, "Gated context aggregation network for image dehazing and deraining," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 1375–1383.
- [36] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.
- [37] Y. Reddy, P. Viswanath, and B. E. Reddy, "Semi-supervised learning: A brief review," *Int. J. Eng. Technol.*, vol. 7, nos. 1–8, p. 81, 2018.



**Yitong Zheng** received the B.Eng. degree in electronic and information engineering from Northwestern Polytechnical University, Xi'an, China, in 2016, where he is currently pursuing the Ph.D. degree in information and communication engineering.

His research interests include machine learning and computer vision, with a focus on object detection, image classification, and image dehazing.



**Jia Su** (Member, IEEE) received the B.Eng. degree in communication engineering and the M.Eng. degrees in optical communication from the Guilin University of Electronic Technology, Guilin, China, in 2008 and 2011, respectively, and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2015.

He is currently an Associate Professor with Northwestern Polytechnical University (NPU), Xi'an. His research interests include radar signal processing and time-frequency analysis.



**Shun Zhang** (Member, IEEE) received the B.S. and Ph.D. degrees in electronic engineering from Xi'an Jiaotong University, Xi'an, China, in 2009 and 2016, respectively.

He is currently an Associate Professor with the School of Electronic and Information, Northwestern Polytechnical University, Xi'an. His research interests include machine learning, computer vision, and human-computer interaction, with a focus on object detection, visual tracking, person reidentification, image classification, feature extraction, and sparse representation.



**Mingliang Tao** (Senior Member, IEEE) received the B.Eng. degree in intelligent science and technology from Xidian University, Xi'an, China, in 2011, and the Ph.D. degree in signal processing from the National Laboratory of Radar Signal Processing, Xidian University, in 2016.

In July 2016, he joined Northwestern Polytechnical University (NPU), Xi'an, as an Associate Professor. He has authored more than 30 journals and conference papers in the last five years. His research interests are radio frequency interference and radar

signal processing.

Dr. Tao was a recipient of the Doctoral Dissertation Award by the China Education Society of Electronics and the Shaanxi Province of China, the Young Scientist Award from the International Union of Radio Science (URSI) in 2017 and 2020, and among the Finalists for Postdoctoral Innovation Talent Support Program of China in 2017.



**Ling Wang** received the B.Sc., M.Sc., and Ph.D. degrees in electronic engineering from Xidian University, Xi'an, China, in 1999, 2002, and 2004, respectively.

From 2004 to 2007, he worked at Siemens and Nokia Siemens Networks, Beijing. Since 2007, he has been with the School of Electronic and Information, Northwestern Polytechnical University, Xi'an, where he was promoted to Professor in 2012. His research interests include array processing and smart antennas, wideband communications, cognitive radio, adaptive antijamming for satellite communications, satellite navigation, and date link systems.