Single Satellite Optical Imagery Dehazing using SAR Image Prior Based on conditional Generative Adversarial Networks

Binghui Huang[1*], Zhi Li[2*], Chao Yang[1], Fuchun Sun[1], Yixu Song[1]
[1] Beijing National Research Center for Information Science and Technology (BNRist),
State Key Lab on Intelligent Technology and Systems,
Department of Computer Science and Technology, Tsinghua University
[2] Shanghai University Of Engineering Science
hbh18@mails.tsinghua.edu.cn
lizhi3158@foxmail.com, yangchao18@mails.tsinghua.edu.cn
fcsun@tsinghua.edu.cn, songyx@mail.tsinghua.edu.cn

Abstract

Satellite image dehazing aims at precisely retrieving the real situations of the obscured parts from the hazy remote sensing (RS) images, which is a challenging task since the hazy regions contain both ground features and haze components. Many approaches of removing haze focus on processing multi-spectral or RGB images, whereas few of them utilize multi-sensor data. The multi-sensor data fusion is significant to provide auxiliary information since RGB images are sensitive to atmospheric conditions. In this paper, a dataset called ***SateHaze1k*** *is established and composed of 1200 pairs clear Synthetic Aperture Radar (SAR), hazy RGB, and corresponding ground truth images, which are divided into three degrees of the haze, i.e. thin, moderate, and thick fog. Moreover, we propose a novel fusion dehazing method to directly restore the haze-free RS images by using an end-to-end conditional generative adversarial network(cGAN). The proposed network combines the information of both RGB and SAR images to eliminate the image blurring. Besides, the dilated residual blocks of the generator can also sufficiently improve the dehazing effects. Our experiments demonstrate that the proposed method, which fuses the information of different sensors applied to the cloudy conditions, can achieve more precise results than other baseline models.*

1. Introduction

With the development of the remote sensing technology in recent years, the quantity and quality of the satellite images have dramatically improved, which can be applied to many fields, *e.g.* building extraction [28], earthquake damage assessment [2], and image decomposition [6]. In many cases, however, optical remote sensing images are sensitive and vulnerable to the weather and light conditions. Cloud, fog or snow often affects the visibility and accuracy of the optical images, as a result of which, the applications of the optical images maybe invalid in some scenarios *e.g.* Due to the obscuration by the cloud, the real situation cannot be obtained in time for the applications of the disaster evaluation and visual interpretation. Dehazing can not only improve the quality of the visible images, but also enhance the

Figure 1. Satellite image haze removal example. From left to right: satellite hazy images, haze removal results of our method.

*Denotes equal contributions. Corresponding author: Fuchun Sun.

usability of the practical frameworks. Therefore, it is significant to investigate the methods of removing the haze in the remote sensing images.

Single image haze removal is challenging because there is no time series image information to refer to. Traditional researchers often solve this problem by utilizing various kinds of prior knowledge, such as He *et al*. presented the dark channel prior [16], Tan *et al*. proposed Maximum Contrast (MC) method [34], and Zhu *et al*. raised Color Attenuation Prior (CAP) method [29]. However, the human visual system does not rely on these explicit feature transformations to estimate the concentration of the fog and the depth of the scene.

Since the deep neural network has made great progress in tasks such as object detection [12] and semantic segmentation [32], researchers begin to use deep learning based methods. These methods can be mainly divided into two types. On the one hand, based on the physical corruption model, the parameters of the dehazing model can be estimated by the neural network in a few methods. On the other hand, the input hazy images are directly utilized to obtain the dehazed image. The latest dehazing methods are similar to the latter framework. These works usually employ machine learning methods to learn transmission maps, such as random forest regression in [35], Convolutional Neural Networks (CNNs) in [30] and Generative Adversarial Network (GAN) in [25, 38, 7]. They usually rely on the quality of the training data, thus declining the efficiency and achievements of learning-based approaches.

However, fog or haze produces information loss. As a result, it is difficult for dehazing to enhance the missing covered information in essence. Moreover, not all of the abovementioned models and strategies are specifically designed for the satellite images with complex backgrounds, abundant objects, large area of space. Particularly, the satellite images are frequently blinded by mist and haze. In addition, many methods generate so simple haze mask that the synthetic images lack sufficient expressiveness for the complicate scenes. Therefore, these solutions are actually not appropriate for the remote sensing images dehazing. Furthermore, besides optical RGB three-channel images, satellite images also include near-infrared, multispectral, synthetic aperture radar (SAR) *etc*. which reflect different wavebands information of the landscapes and objects features. As for the SAR, mathematical techniques are used to combine reflected signal phase and amplitude information. In addition, a high-resolution image can be built up according to several adjacent-in-time RADAR pulses, which are seen as a function of time. SAR is often used because of its all-weather and day-or-night capability. Even though facing the weather with extremely low visibility, SAR images are weather resistant and can maintain a good quality of satellite image. Therefore, when the interpretation and availability of the optical remote sensing images are confined, the SAR radar images can be an alternative approach to retrieve the real situation of the target area under bad weather *e.g*. cloudy or snowy, and its practical application is promising. Overall, as a supplement to the optical images, SAR images are beneficial for the remote sensing image applications. Besides, remote sensing applications such as building change detection, segmentation or earthquake damage assessment put forward higher requirements for the recovering quality and image details. Therefore, the SAR information is introduced due to the unique texture details. These are vital to detecting buildings after the satellite haze removal process for recovering cloudless high-resolution images. However, SAR and RGB usually have different modalities, resolutions, reflection angles, and different time zones. Therefore, it is very difficult to fuse the SAR information in the defogging tasks. At present, the only authoritative SAR-Optical-paired public dataset [31] was released in July, 2018, and [13] introduced Gaussian noise into the raw RGB images to generate fogging masks, afterwards explored defogging algorithms using 11-band concatenation of SAR and multispectual images as input to train a GAN.

To handle the obstacles of the missing information caused by fog occlusion, and normal photos dehazing approaches inapplicable to satellite images, we train a conditional generative adversarial networks (cGAN) by adding a SAR constraint, which can fit the real ground truth image distribution for the purpose of obtaining more natural effects. In our experiments, we show that RGB and SAR information fusion not only enhances the visual effects, but also promotes the performance of subsequent application tasks, such as building footprints detection and segmentation. Our method achieves better results contrasting with other methods which only utilize RGB remote sensing images for dehazing and SAR-Opt-cGAN designed for SAR and RGB information fusion proposed by [13], exposing that multi-sensor fusion is capable of promoting both vision and segmentation performance after remote sensing images haze removal.

Our contributions can be summarized as follows:

- For the sake of lacking optical and SAR paired data, we construct a new dataset ***SateHaze1k*** which includes 1200 the same region pairs of cloudy RGB images, SAR images and corresponding RGB haze-free ground truth, on which researchers can fuse SAR with RGB to conduct haze removal tasks.

- We propose a generative approach, specifically, use deep neural network (DNN) to construct an end-to-end network, applying SAR images as condition, RGB images as input and dehazed images as output to enhance dehazing performance. We also put forward a dilated residual block eliminating image indistinctness.

- Considerable empirical results on challenging benchmarks indicate that our method attains consistent improvements over other SAR-used GAN method and many baseline models.

The rest of the paper is organized as follows. Firstly the summary of related work is in Section 2, the ***SateHaze1k*** dataset is described in Section 3. Our proposed image dehazing method is demonstrated in Section 4. Finally, the comprehensive experiments results and ablation studies are discussed in Section 5 and conclusion in Section 6.

2. Related Works

In this section, we will introduce the concerning low-level vision algorithms, the applications of attention mechanism and the development of GANs.

2.1. Low Level Vision

In recent years, research on low-level vision [16, 4, 20, 21, 39, 11, 26] tasks, including haze removal, rain removal and snow removal, is mainly divided into two classes. The first class is from the perspective of the model [16, 20, 21], which focus on using the traditional model to describe the transmission map and background separately. He *et al*. [16] proposed an effective haze removal method using the dark channel prior to fit physical corruption model. A bilateral filter [20] is used to decompose the rainy images into the high-frequency and low-frequency components. Kim *et al*. [21] first used kernel regression to detect rain streaks and then used a non-local means filter to remove the detected rain streaks. The second class is based on deep learning to construct a network framework to accomplish the image enhancement tasks [4, 39, 11, 26]. Zhang et al. [39] advanced a convolutional neural network (CNN)-based algorithm which first estimates the rain density and then removes the corresponding rain-streaks. Fu *et al*. [39] introduced a method for removing rain streaks from an image based on CNN; it directly learns the transmission map between the rainy and clean images. Cai *et al*. [4] proposed an end-to-end system called DehazeNet, it firstly estimates the transmission map and finally used a classical algorithm to recover the clear image. Liu *et al*. [26] proposed a multi-stage network, DesnowNet, to manage the snow removal problem. In these image enhancement tasks, haze removal is more complicated because the density of the haze is an additional factor that makes the hazy images more difficult to process.

2.2. Generative Adversarial Networks (GANs)

Over the last few years, with the development of deep learning, the GANs gradually become a research hotspot. GANs has achieved amazing results in fitting data distribution and generating images. Usually, GANs consists of generative network and discriminative network. The GANs model attracted much attention because they do not require prior knowledge, however, the free training methods also makes GANs tremendously sensitive to initial parameters, resulting in the training process is extremely unstability and serious mode collapse problems. Arjovsky *et al*. [1] discussed the difficulties in GANs training caused by J-S divergence approximation and put forward the use of earth-mover distance $w(q, p)$. In addition, Mao *et al*. [27] proposed a Least Squares Generative Adversarial Networks (LSGANs), which applies the least square loss function to the discriminative network to overcome the problem of mode collapse and disappearing gradient in the training stage.

At present, GANs has has been applied to various low-level vision tasks, such as style transfer [5, 9], photo enhancement [8], super-resolution [24], imitation learning [37] and others [3, 14, 10].

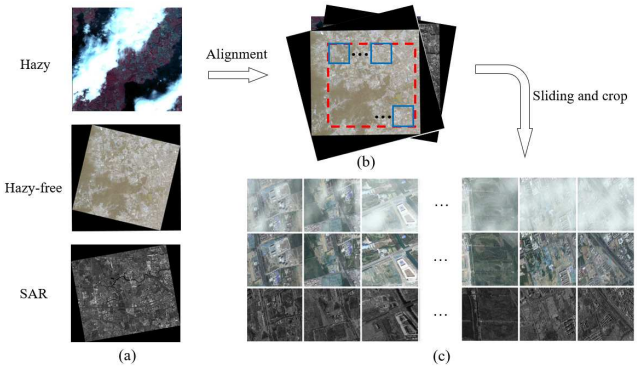

Figure 2. Illustration of the dataset generation pipeline. (a) Raw data of hazy RGB, haze-free RGB and SAR images that are captured by the earth remote sensing satellites (SSR). (b) Alignment of multi-channel remote sensing images according to geographic location. (c) Sliding on the overlap of the remote sensing images and cropping into 512×512 images. Each row (from top to bottom) represents the hazy images, hazy-free (ground truth) images, and SAR images respectively. Note that the raw remote sensing images have a very large resolution.

3. Haze Satellite Dataset

The new haze satellite dataset on which we evaluate our approach contains 1200 individual pairs of hazy images, corresponding hazy-free images and SAR images. The dataset generation process is shown in Figure 2. The raw multi-channel remote sensing images may be misaligned due to the satellite synchronization where hazy and haze-free images were captured using the GF-2 satellite and the SAR image was collected by GF-3 satellite. So before sliding and cropping the images, we have to align all the remote sensing images to ensure that the same image area represents the same geographic area.

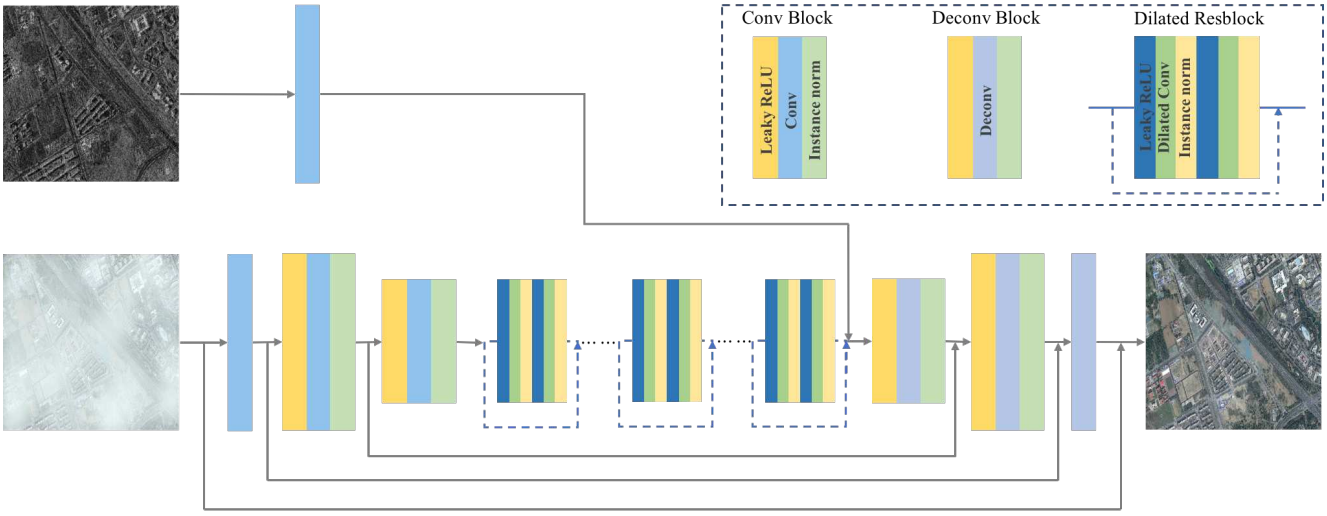

Figure 3. The overall network structure of the proposed method is a basic auto-encoder structure. It consists of one common convolution layer, and two convolution blocks, followed by six dilated resblocks as the feature extraction part, two deconvolution blocks and one deconvolution layer. The SAR prior is added to the feature maps after all of the dilated resblocks, that highway extracts the texture details information. The dilated resblocks are inserted in the middle network to aggregate context information. To fuse the features from different levels, the skip-layers connection is leveraged. During the output image and the hazy input image in an end-to-end way. Note that we adopt the patchGAN[19] as our discriminator network, so we omitted it here.

In order to guarantee the facticity, abundance, and diversity of haze masks in our dataset, we use *Photoshop* Software to extract real haze masks of the easily accessible original hazy remote sensing images to generate transmission maps for synthetic images. The dataset consists of 3 levels of fog, called *Thin fog*, *Moderate fog*, *Thick fog*. In the synthetic images covered by thin fog, the haze mask will be only mist which picks up from the original real cloudy image. For the moderate fog image, samples overlap with mist and medium fog. But for the thick fog, the transmission maps are selected from the dense haze.

Training, validation and test folds. Our training, validation and test folds were approximately 80%, 10%, 10% of the total data respectively. More details of the dataset can be found in the persistent link[1]. We split every 400 images to train, valid, and test set, and artificially label 45 of thick fog images for segmentation purposes.

4. Methodology

In this section, firstly we present the structure of our GANs including the generative and discriminative network. As shown in Figure 3, for the generator we utilize the encode-decoder network including convolutional blocks, dilated resblocks and deconvolutional blocks. For the discriminator, we adopt 70×70 PatchGAN in [19] and change the network input to the concatenation of the haze-free image and original hazy image. Then, we introduce our loss function in detail for generator and discriminator separately.

4.1. Generative Networks

The generative network structure is shown in Figure 3. It is composed of two modules, *i.e.*, the encoder-decoder network and the SAR information highway. The encoder blocks aim to obtain both global and local features of the input haze image, the decoder deconvolutional blocks restore the feature maps to original size and produce an output image. The SAR information highway extracts the features of the SAR prior.

In the training step, we first use a convolution layer to obtain the features of SAR image. Then we feed the input haze image into the encoder-decoder network to get the detailed features. Finally, we perform an element-wise sum operation on the above results of two modules to obtain the fusing feature to generate a clean image. We also utilize skip connections to fuse multi-scale information. Furthermore, we adopt the handmade designed dilated residual blocks which expands the receptive fields significantly and avoid gradient vanishing effectively.

Different from previous GANs which rely on adversarial loss, we propose to use LSGANs loss to both generator and discriminator, which can be expressed as:

$$\begin{aligned} L_{ls} =& E_{x,y\sim P_{data}(x,y)}\left[(D(x,y)-0)^2\right] + \\ & E_{x\sim P_{data}(x)}\left[(D(x,G(x))-1)^2\right] \end{aligned} \tag{4.1}$$

where G represents the generative network, and D represents the discriminator network, x is an input hazy image,

[1] https://www.dropbox.com/s/k2i3p7puuwl2g59/Haze1k.zip?dl=0

y is a ground truth clean image. In order to stable training stage, we introduce the least square loss function to regularize the two modules, the loss functions are defined as:

$$L_1 = \|G(x) - y\|_1 \tag{4.2}$$

where y represents the corresponding clean image of the input hazy image.

The overall loss function of the generator network can be expressed as:

$$L_G = \alpha L_{ls} + \beta L_1 \tag{4.3}$$

The setting of α, β will be described in Section 5.1.

Encoder-decoder network: The proposed encoder-decoder network comprises a stack of strided convolutional blocks, nine residual blocks [17], and four transposed convolutional blocks. Each residual block consists of a convolutional layer, an instance normalization layer [36], and a ReLU activation layer [23]. After the first convolutional layer in each ResBlock [17], the Dropout [33] with probability 0.5 is added to the generative network.

4.2. Discriminator Network

The structure of the discriminator network is like 70×70 PatchGAN [19]. The discriminator network, comprising a stack of convolution layers, each of which is followed by a batch normalization layer [18] and Leaky ReLU activation function. The final layer of the discriminator network is the sigmoid activation function whose probability of the output and input image pairs is true or false.

The whole loss function of the discriminator network can be expressed as:

$$L_D = E_{x \sim P_{\text{data}}(x)} \left[(D(x, G(x)) - 0)^2 \right] \tag{4.4}$$

The samples of various loss functions are illustrated in Figure 5.

5. Experiments

In this section, we will demonstrate the effectiveness of the training details and evaluation metrics in our tasks. Furthermore, we analyze and discuss the effects of the proposed algorithm, including loss functions, SAR image prior and the dilated blocks.

5.1. Training Details

Each convolution block of the proposed encoder-decoder network in the generator network consists of a convolution, instance normalization and LeakyReLU. The rates of six dilated resblock are set as 2, 2, 2, 4, 4, 4. Generator input and the output image size are set as $512 \times 512 \times 3$. The size of the input images in the discriminator is set as $512 \times 512 \times 3$, and the size of the output vector is $30 \times 30 \times 1$. During training process, we set $\alpha = 1, \beta = 1$. The learning rate is set at 0.0002. We use Adam [22] optimization to train our network. Our method is implemented in Python 3.5 with a Nvidia RTX 2080Ti.

5.2. Metrics

We evaluate the performance of our method from two aspects: visual performance and task-oriented performance. For visual performance, we use peak signal to noise ratio(PSNR) and structural similarity index(SSIM) to assess the quality of recovered images. For the latter, The segmentation algorithms are used in the generated images to confirm whether the segmentation accuracy is sufficiently promoted. We refine the MASK-RCNN[15] to detect the building footprints of the remote sensing RGB images. In this respect, we use general metrics in segmentation task: Intersection over Union(IoU), F1-Score and pixel accuracy(PA). The F1-Score metric is based on proposed buildings whose IoU with the ground truth is larger than 0.5.

Training: Inspired by the training process on COCO dataset in [15], we train the network through the following three phases as follows: the first phase for training the head with 40 epochs, the second phase for training 4th and more rear stage of ResNet-101 with 20 epochs, and the last phase for training all layers with 20 epochs. We train a 3 channel input mask-rcnn network performing better than top-3 winners (Nofto, Wleite, and XD_XD) of the SpaceNet Challenge, while XD_XD uses 11 channel input image to build up building footprints.

5.3. Comparative Evaluations

In this section, the proposed method is quantitatively evaluated and compared with the state-of-the-art ones on synthetic images and real-world hazy images.

Quantitative Analysis: Several kinds of previous dehazing methods are realized as contrast experiments, containing traditional approach DCP [16], DNN-based model Dehazenet [4], and the method which applies SAR information SAR-Opt-cGAN.

Corresponding quantitative results of different methods are presented in Table 2. Apparently, our proposed method can achieve excellent performance in the three cases of fogs in Figure 4. Moreover, distinct from other methods, our models exhibit generalization ability to remove different levels of fog.

Qualitative Analysis: Figure 4 illustrates three real-world hazy images and corresponding output images generated by our advanced methods. The method proposed by He *et al*. DCP is a traditional approach to make use of dark channel prior cannot produce a clear image, as shown in Figure 4(c). The previous method [4] used a convolutional neural network to estimate the transmission map and a physical corruption model to recover clear images. However, the haze removal results still contain some artifacts and residu-

Figure 4. The visual effects of different methods implemented in our dataset. From left to right: (a)Input images, (b)Ground truth, (c)DCP, (d)DehazeNet, (e)SAR-Opt-cGAN, (f)Ours. The generative methods *e.g*. SAR-Opt-cGAN and ours generate more clear haze-free images than the methods based on atmospheric scattering model *e.g*. DCP and DehazeNet. Our method conditions the GAN model on additional SAR information and performs better than [13] whose name is called SAR-Opt-cGAN but actually use the normal GAN).

Method	Thin fog		Moderate fog		Thick fog	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Original	12.7712	0.7241	12.5867	0.7399	8.5893	0.4215
DCP[16]	13.1517	0.7246	9.7830	0.5735	10.2513	0.5850
Dehazenet[20]	19.7529	0.8950	18.1250	0.8552	14.3321	0.7064
SAR-Opt-cGAN[13]	20.1950	0.8419	21.6616	0.7941	19.6553	0.7573
Our without SAR	21.4744	0.8168	22.0945	0.8274	22.1206	0.7842
Our with SAR	**24.1638**	**0.9061**	**25.3111**	**0.9264**	**25.0731**	**0.8640**

Table 1. Dehazing performance comparison by two evaluation metrics, PSNR and SSIM, in three levels of haze. Our method without SAR prior exclusively uses RGB images as input. Our method with SAR prior generate clear images as the Figure 3.

Method	Iou	F1-Score	PA
Original	0.208	0.042	**0.962**
SAR-Opt-cGAN	0.297	0.107	0.957
Ours without SAR	0.276	0.069	0.956
Ours with SAR	**0.319**	**0.113**	0.961

Table 2. The results of our method using SAR image prior or not. The SAR-based methods perform better in the segmentation tasks.

als of the haze especially the dense fog. Moreover, the covered objects are confusing and complex to be recognized. In contrast, our algorithm has a more strong generalization ability to handle different fog and achieves better results.

5.4. Ablation Study

Effect of Loss Function: In order to regularize the output of the generator network and get sufficiently similar images to ground truth, we adapt L_1 in Eq. (4.2) and L_{ls} in Eq. (4.1) to compose the loss function. In order to demonstrate the validity of the loss function, we train the same network with different loss functions. As shown in Figure. 5 and Table 3, the loss function we use achieves better performance than others, *e.g*.L_{ad} loss. L_1 loss can greatly restrict the distribution of the inputs and targets of the generative networks. The least square loss is kernel-based square loss and gains better results than traditional adversary loss due to its second-order derivative regularization property for gradient

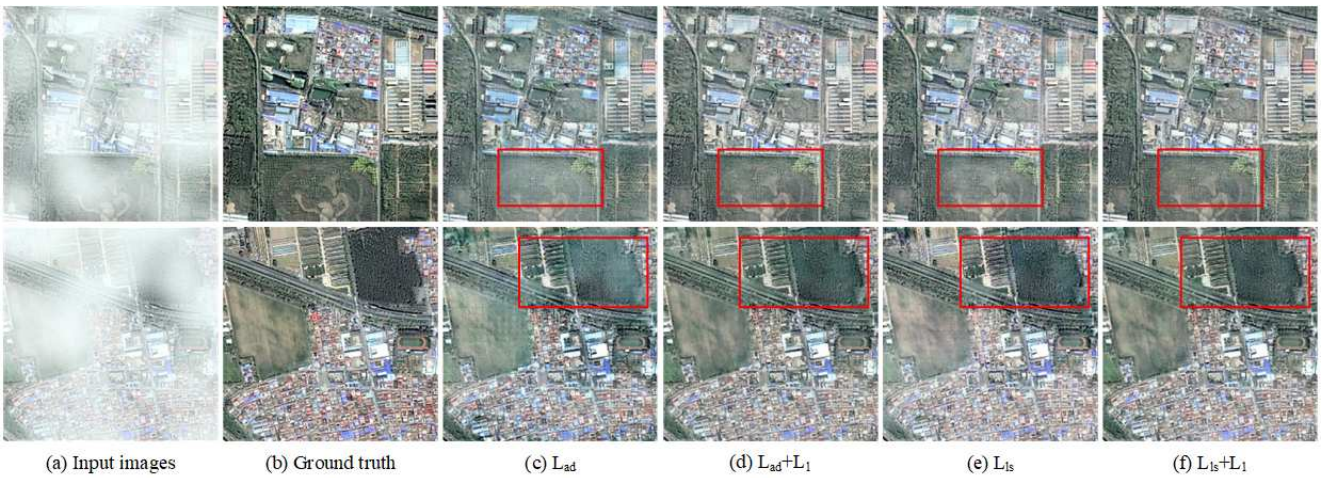

Figure 5. Visualization results of different loss function settings. From left to right: Input hazy image, ground truth, the generating results of the four loss functions $L_{ad}, L_{ad} + L_1, L_{ls}, L_{ls} + L_1$. It can easily be seen that the $L1$ loss plays an important role in generating the clear images, either L_{ad} or L_{ls} can get a increase after we add L_1 loss to the overall loss function.

backward optimization process.

Loss	L_{ad}	$L_{ad} + L_1$	L_{ls}	$L_{ls} + L_1$
Avg. PSNR	16.721	23.824	18.874	**24.945**
Avg. SSIM	0.707	0.938	0.809	**0.946**

Table 3. The performance of our method when using different loss functions.

Effect of Sar Image Prior and Dilated Residual Block: We also intend to investigate the influence of the SAR image prior, so we abandon the SAR image feature extraction path in our model to test whether the SAR image prior plays an important role in the haze removal task. The results are presented in the Table 2 and Table 4, the SAR prior feature attains performance improvement both in the visual effect and task-oriented Iou, F1 score and pixel accuracy both in SAR related methods than others without SAR inputs, SAR-Opt-cGAN [13] and our method with SAR inputs. Our method performs better in the Iou and pixel accuracy, while is quite close to SAR-Opt-cGAN in F1-Score. In conclusion, the SAR images introduce additional detailed prior information which can be of much significance in dehazing tasks.

w/SAR Prior	×	×	√	√
w/Dilated Block	×	√	×	√
Avg.PSNR	22.099	22.851	23.039	**24.945**
Avg.SSIM	0.827	0.851	0.842	**0.946**

Table 4. Dilation and SAR image prior ablation study analysis for each component with different training settings, which shows the combination of our SAR prior and designed block achieving best performance.

6. Conclusion

In this paper, due to the lack of wonderful SAR-Optical pair cloud removal dataset so far, we built ***SateHaze1k*** dataset including 1200 pairs of images, where the optical images in each 400 pairs are covered by thin, moderate and thick fog. For the last moderate test 45 images, we also label out the building footprints for the segmentation tasks use. Moreover, we adopt the conditinal GANs in the satellite haze removal tasks since the conventional methods focus on the estimation of the transmission map from the input hazy image. In order to produce better dehazed results, we proposed a novel architecture introducing SAR prior as limits to reconstruct original details, which can be trained end-to-end to generate natural haze removal images. We propose a practical dehazing convolution dilated resblock to smooth the output and generate distinct images. The proposed method compares favorably with the most advanced methods in synthetic datasets.

References

[1] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein gan. 2017.

[2] D. Brunner, G. Lemoine, and L. Bruzzone. Earthquake damage assessment of buildings using vhr optical and sar imagery. *IEEE Transactions on Geoscience & Remote Sensing*, 48(5):2403–2420, 2010.

[3] A. Bulat and G. Tzimiropoulos. Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans. 2017.

[4] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016.

[5] H. Chang, J. Lu, F. Yu, and A. Finkelstein. PairedCycleGAN: Asymmetric style transfer for applying and removing makeup. June 2018.

[6] Chen and Xiaoli. An edge-preserving variational method for image decomposition. *Chinese Journal of Electronics*, 22(1):109–113, 2013.

[7] J. Chen and L. Zhang. Joint multi-image saliency analysis for region of interest detection in optical multispectral remote sensing images. *Remote Sensing*, 8(6), 2016.

[8] Y. Chen, Y. Wang, M. Kao, and Y. Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. June 2018.

[9] Y. Choi, M. Choi, M. Kim, J. W. Ha, S. Kim, and J. Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. 2017.

[10] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, and J. Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. 2017.

[11] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, 26(6):2944–2956, 2017.

[12] R. Girshick. Fast r-cnn. *Computer Science*, 2015.

[13] C. Grohnfeldt, M. Schmitt, and X. Zhu. A conditional generative adversarial network to fuse sar and multispectral optical data for cloud removal from sentinel-2 images. 2018.

[14] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. 2018.

[15] K. He, G. Gkioxari, P. Dollar, and R. Girshick. Mask r-cnn. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, PP(99):1–1, 2017.

[16] K. He, S. Jian, and X. Tang. Single image haze removal using dark channel prior. 2009.

[17] K. He, X. Zhang, S. Ren, and S. Jian. Deep residual learning for image recognition. 2016.

[18] S. Ioffe and C. Szegedy. Batch normalization: accelerating deep network training by reducing internal covariate shift. 2015.

[19] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. 2016.

[20] L. W. Kang, C. W. Lin, and Y. H. Fu. *Automatic Single-Image-Based Rain Streaks Removal via Image Decomposition*. 2012.

[21] J. H. Kim, C. Lee, J. Y. Sim, and C. S. Kim. Single-image deraining using an adaptive nonlocal means filter. In *IEEE International Conference on Image Processing*, 2014.

[22] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *Computer Science*, 2014.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. 2012.

[24] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. 2016.

[25] R. Li, J. Pan, Z. Li, and J. Tang. Single image dehazing via conditional generative adversarial network. 2018.

[26] Y. F. Liu, D. W. Jaw, S. C. Huang, and J. N. Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, PP(99):1–1, 2017.

[27] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, W. Zhen, and S. P. Smolley. Least squares generative adversarial networks. 2016.

[28] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone. Morphological attribute profiles for the analysis of very high resolution images. *IEEE Transactions on Geoscience & Remote Sensing*, 48(10):3747–3762, 2010.

[29] Z. Qingsong, M. Jiaming, and S. Ling. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11):3522–3533, 2015.

[30] W. Ren, L. Si, Z. Hua, J. Pan, X. Cao, and M. H. Yang. Single image dehazing via multi-scale convolutional neural networks. 2016.

[31] M. Schmitt, L. H. Hughes, and X. X. Zhu. The sen1-2 dataset for deep learning in sar-optical data fusion. *arXiv preprint arXiv:1807.01569*, 2018.

[32] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 39(4):640–651, 2014.

[33] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[34] R. T. Tan. Visibility in bad weather from a single image. 2008.

[35] K. Tang, J. Yang, and J. Wang. Investigating haze-relevant features in a learning framework for image dehazing. 2014.

[36] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Instance normalization: The missing ingredient for fast stylization. 2016.

[37] C. Yang, X. Ma, W. Huang, F. Sun, H. Liu, J. Huang, and C. Gan. Imitation learning from observations by minimizing inverse dynamics disagreement. In *Advances in Neural Information Processing Systems*, pages 239–249, 2019.

[38] H. Zhang and V. M. Patel. Densely connected pyramid dehazing network. 2018.

[39] H. Zhang and V. M. Patel. Density-aware single image deraining using a multi-stream dense network. 2018.