IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 15, 2019, accepted August 30, 2019, date of publication September 5, 2019,
date of current version September 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939649

SAR-to-Optical Image Translation Using Supervised Cycle-Consistent Adversarial Networks

LEI WANG[1], XIN XU[1], YUE YU[1], RUI YANG[1], RONG GUI[1], ZHAOZHUO XU[2], (Student Member, IEEE), AND FANGLING PU[1,3], (Member, IEEE)

[1]Collaborative Sensing Laboratory, School of Electronic Information, Wuhan University, Wuhan 430079, China
[2]Electrical Engineering Department, Stanford University, Stanford, CA 94305, USA
[3]Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China

Corresponding author: Xin Xu (xinxu@whu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0502600, in part by the Thirteen-Five Civil Aerospace Planning Project — Integration of Communication, Navigation and Remote Sensing Comprehensive Application Technology, and in part by the Chinese Technology Research and Development of the Major Project of High-Resolution Earth Observation System under Grant 03-Y20A10-9001-15/16.

ABSTRACT Optical remote sensing (RS) data suffer from the limitation of bad weather and cloud contamination, whereas synthetic aperture radar (SAR) can work under all weather conditions and overcome this disadvantage of optical RS data. However, due to the imaging mechanism of SAR and the speckle noise, untrained people are difficult to recognize the land cover types visually from SAR images. Inspired by the excellent image-to-image translation performance of Generative Adversarial Networks (GANs), a supervised Cycle-Consistent Adversarial Network (S-CycleGAN) was proposed to generate large optical images from the SAR images. When the optical RS data are unavailable or partly unavailable, the generated optical images can be alternative data that aid in land cover visual recognition for untrained people. The main steps of SAR-to-optical image translation were as follows. First, the large SAR image was split to small patches. Then S-CycleGAN was used to translate the SAR patches to optical image patches. Finally, the optical image patches were stitched to generate the large optical image. A paired SAR-optical image dataset which covered 32 Chinese cities was published to evaluate the proposed method. The dataset was generated from Sentinel-1 (SEN-1) SAR images and Sentinel-2 (SEN-2) multi-spectral images. S-CycleGAN was applied to two experiments, which were SAR-to-optical image translation and cloud removal, and the results showed that S-CycleGAN could keep both the land cover and structure information well, and its performance was superior to some famous image-to-image translation models.

INDEX TERMS SAR-to-optical image translation, visualization, GAN, Sentinel, cloud removal.

I. INTRODUCTION

Synthetic aperture radar (SAR) and optical remote sensing (RS) sensors have been widely used in land use planning, disaster prevention, target detection and so on [1]–[3]. The bad weather and clouds commonly cause severe contamination in optical RS imagery, seriously affecting the Earth observation. SAR can overcome the disadvantages of optical RS data and has the advantages of working all day and night, working under all weather conditions, large scope, and certain penetration capacity. However, SAR images are different from optical RS images and untrained people are difficult to recognize the land cover types from SAR images visually mainly due to the following three reasons. First of all, the color information of land cover types in SAR images is very different from optical RS images. Especially, single polarization (single-pol) SAR images do not contain any color information. Second, SAR images are contaminated by speckle, which never appears in optical RS images. Third, the special side-looking active imaging mechanism of SAR leads to the geometry distortion and shadows. If a SAR image can be translated to the image which looks like an

The associate editor coordinating the review of this manuscript and approving it for publication was Fan Zhang.

optical RS image, then the generated optical image can be a good alternative that aid in land cover visual recognition for untrained people. Therefore, exploring the method of SAR-to-optical image translation is of great significance.

In the past decades, many methods were proposed to make the SAR images more suitable for visual recognition [4]–[6], and they can be mainly divided into two categories: image enhancement algorithms and pseudo-color encoding algorithms. The image enhancement algorithms aim to show the objects more clearly in SAR images [5], [7]. To effectively obtain the image suitable for human observation, Jiao *et al.* [8] introduced a new SAR image visualization algorithm to map the high dynamic range SAR amplitude values to low dynamic range displays via reflectivity distortion preserved entropy maximization. In [9], an adaptive two-scale enhancement method was proposed to visualize and further recognize ocean vessels and aircrafts. The pseudo-color encoding algorithms encode the pixels with different colors [10]–[12], which are especially proper for single-pol SAR images. Then the pseudo-color encoded SAR images can be used in other SAR image applications. The SAR images, which were processed by above-mentioned methods, became more suitable for visual recognition by professionals, but they were still very different from optical RS images and untrained people were still hard to recognize the land cover types.

Deep learning [13] is a branch of machine learning and it has developed rapidly in recent years. More and more deep learning methods were applied to SAR image processing [14]–[16]. Deep learning can deal with the image-to-image translation tasks [17], which are often formulated as per-pixel classification or regression. Generative Adversarial Networks (GANs) [18] are recently regarded as a breakthrough in deep learning. As typical probability generation models, GANs can find out the internal data distribution through learning plenty of data. GANs generate data in an unsupervised manner but they cannot control the data generation process [19]. Then the Conditional Generative Adversarial Networks (cGANs) [19] were proposed to solve this problem. By conditioning the models on additional information [20], [21], the data generation process of cGANs can be directed. Many GAN based methods achieved great results for image-to-image translation tasks, such as pix2pix [20], Cycle-Consistent Adversarial Networks (CycleGAN) [22] and DualGAN [23]. GANs can be used to translate SAR images to optical images, which are more suitable for untrained people to recognize the land cover types visually. It is a new way to solve the visual observation problem of SAR images.

In SAR image processing field, some image-to-image translation methods were already proposed. Song *et al.* [24] proposed a method to convert single-pol gray-scale SAR image to polarimetric SAR (PolSAR) image using Convolution Neural Networks (CNNs). Ao *et al.* [25] used a dialectical GAN to translate Sentinel-1 image to TerraSAR-X SAR image. Recently, more and more researchers focus on

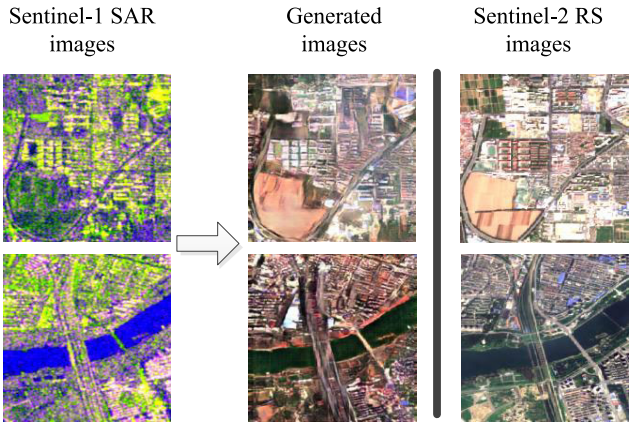


FIGURE 1. Examples of SAR-to-optical image translation in patch level. The images from left to right are Sentinel-1 pseudo-color images, generated images and Sentinel-2 RS images used for comparison.

SAR-to-optical image translation based on GANs [26]–[31]. Schmitt *et al.* [32] published the SEN1-2 dataset to foster deep learning research in SAR-optical data fusion and used pix2pix to translate SAR to optical images. However, these methods mainly had two problems. First, the SAR images were translated in patch level and there was no large image translation result. Large SAR image translation method is still worth studying. Second, some methods used CycleGAN and pix2pix for SAR-to-optical image translation, but both CycleGAN and pix2pix have disadvantages. CycleGAN can well keep the structure information but some land cover information is lost. In contrast, pix2pix can well characterize the land cover information, but it produces blurry results and the structure information of some objects is missed. We considered combining the advantages of CycleGAN and pix2pix, so a supervised Cycle-Consistent Adversarial Network (S-CycleGAN) was proposed to keep both the land cover and structure information in SAR-to-optical image translation. Fig. 1 shows the examples of SAR-to-optical image translation in patch level. Our method is designed to translate large SAR images to optical images and can be used in practical applications. The images that generate from SAR images are helpful for SAR image visualization. When the optical RS data are unavailable, the generated optical images can aid in land cover visual recognition for untrained people.

One of the applications of the generated optical images is the thick cloud removal. Optical RS images are commonly affected by the cloud, fog and rain. If an area is covered by large thick clouds, the land cover types under the thick clouds can be hardly observed through the optical RS images. Existing methods for thick cloud removal can be categorized into interpolation and substitution techniques [34]. The interpolation techniques reconstruct the areas obscured by clouds approximately within an arrangement of cloud-free pixels [35], [36]. These methods require a high temporal resolution and a sufficient length of the image acquisition period to cover seasonal aspects [34]. The substitution techniques simply replace the pixels of clouds with the pixels at the

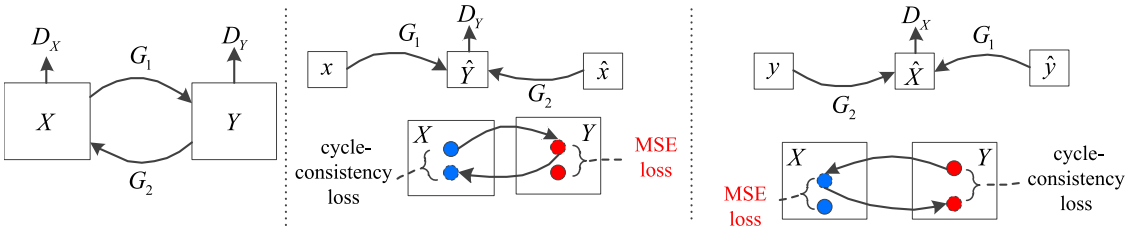


FIGURE 2. Our model is improved from CycleGAN model. CycleGAN introduces two cycle consistency losses and translates the images from domain X to domain Y and back again. In this paper, we add the MSE loss to the two mappings between domain X and Y for better performance. Left: the two mapping functions $G_1 : X \rightarrow Y$ and $G_2 : Y \rightarrow X$ and associated adversarial discriminators D_Y and D_X. Middle: cycle-consistency loss $x \rightarrow G_1(x) \rightarrow G_2(G_1(x)) \approx x$ and the MSE loss $G_1(x) \approx y$. Right: cycle-consistency loss $y \rightarrow G_2(y) \rightarrow G_1(G_2(y)) \approx y$ and MSE loss $G_2(y) \approx x$.

same location of another cloud-free images [37]. However, the temporal differences may exist and are often neglected by these methods. The SAR and cloudy optical RS images which are imaged at the same day can be used to solve the problem of temporal differences, but the visual differences between SAR and optical RS images affect the image quality [34], [38]. With the development of deep learning, researchers attempted to use GANs to remove filmy clouds [39]–[42]. In this paper, the substitution technique was used to for cloud removal. The generated optical images, which were translated from the SAR images by S-CycleGAN, were used to replace the thick cloud areas of the cloudy optical RS images.

The rest of this paper is organized as follows. In section II, the proposed S-CycleGAN is described and a paired image dataset for SAR-to-optical image translation is introduced. In section III, experiment results of SAR-to-optical image translation and cloud removal are illustrated. The seam elimination method used in the two experiments and the potential applications of the generated images are discussed in Section IV. Finally, the conclusions and future work are illustrated in Section V.

II. Proposed Method

A. GENERATIVE ADVERSARIAL NETWORK

Traditionally, GANs consist of two modules, one is the generator G and the other is the discriminator D. The generator learns the data distribution and can produce realistic data to fool the discriminator. The discriminator tries to distinguish whether its input is from the training data or the generator. This framework corresponds to a minimax two-player game. GANs are unconditioned generative models that learn the mapping of input random noisy z to output y.

$$G : z \rightarrow y \tag{1}$$

In contrast, cGANs learn a mapping from the extra input x and random noisy z to output y.

$$G : \{x, z\} \rightarrow y \tag{2}$$

Both the generator and discriminator of cGANs are conditioned on extra information x, which can be text [21], labels [43] and other forms of the images [20].

Some image-to-image translation models are trained in supervised manner and the training images are strictly paired, such as the famous supervised model pix2pix. Pix2pix as a widely used image-to-image translation framework has proven stable and powerful for image-to-image translation. The generator of pix2pix is a "U-Net" [44] architecture and the skip connections between the encoder layers and decoder layers are added to share the low-level information. The architecture of the discriminator is PatchGAN [20] that can model the image as a Markov random field. Pix2pix achieves very good results for many image-to-image translation tasks. For some image-to-image translation tasks, obtaining paired training data can be difficult and expensive. CycleGAN was proposed to solve the problem. The generators of CycleGAN are from [45] and the discriminators are PatchGAN, too. Although CycleGAN is trained without supervision, qualitative results on several tasks demonstrate that the performance of CycleGAN is close to pix2pix [22].

Based on the architectures of pix2pix and CycleGAN, we proposed S-CycleGAN for SAR-to-optical image translation.

B. SUPERVISED CYCLEGAN

Proposed S-CycleGAN is improved from CycleGAN, so we first introduce the framework of CycleGAN. CycleGAN learns the translation between two image domains X and Y with training samples $\{x_i\}_{i=1}^{N} \in X$ and $\{y_j\}_{j=1}^{M} \in Y$. It contains two generators and two discriminators, and introduces two mappings $G_1 : X \rightarrow Y$ and $G_2 : Y \rightarrow X$, as shown in Fig. 2.

The generator G_1 tries to generate images $G_1(x)$ that look similar to images from domain Y and the discriminator D_Y aims to discriminate between y and $G_1(x)$. In the same way, G_2 tries to generate images $G_2(y)$ that look similar to images from domain X and the discriminator D_X aims to discriminate between x and $G_2(y)$. The adversarial losses for the two mappings are shown as below:

$$\begin{aligned}\mathcal{L}_{GAN}(G_1, D_Y, X, Y) = {} & \mathbb{E}_{y \sim p_{data}(y)}[logD_Y(y)] \\ & + \mathbb{E}_{x \sim p_{data}(x)}[1 - D_Y(G_1(x))] \end{aligned} \tag{3}$$

$$\begin{aligned}\mathcal{L}_{GAN}(G_2, D_X, Y, X) = {} & \mathbb{E}_{x \sim p_{data}(x)}[logD_X(x)] \\ & + \mathbb{E}_{y \sim p_{data}(y)}[1 - D_X(G_2(y))] \end{aligned} \tag{4}$$

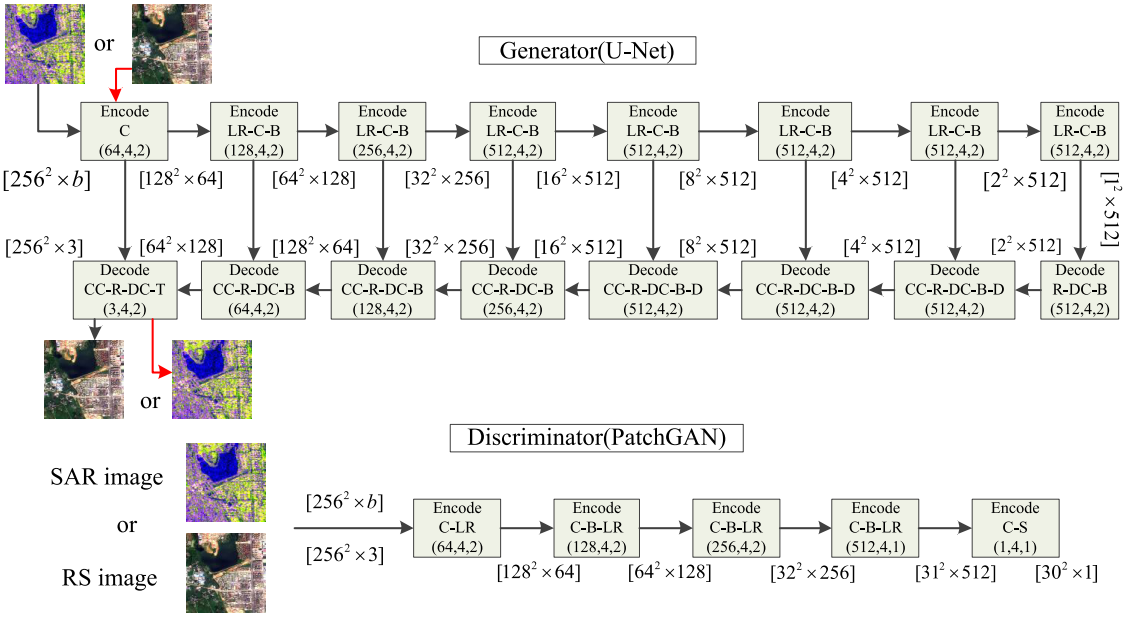


FIGURE 3. Architectures of the generator (top) and discriminator (bottom) networks. Acronyms in the encoder and decoder units are as follows: b: The number of bands, C: Convolutional, R: ReLU, LR: Leaky ReLU, B: Batch Normalization, D: Dropout, DC: Deconvolutional, CC: Concatenation, T: Tanh, S: Sigmoid. The three numbers in round brackets shown in all encoding and decoding layers indicate the number of filters, filter size and stride, respectively. The numbers in square brackets indicate the size of feature maps.

To train an unsupervised image-to-image translation model, CycleGAN introduces the cycle consistency loss. For each input image x, the image translation cycle should be able to bring x back to the original image, which can be illustrated as bellow:

$$x \to G_1(x) \to G_2(G_1(x)) \approx x \tag{5}$$

Similarly, for each image y from domain Y, G_1 and G_2 should also satisfy backward cycle consistency.

$$y \to G_2(y) \to G_1(G_2(y)) \approx y \tag{6}$$

Therefore, the cycle consistency loss can be illustrated as below:

$$\mathcal{L}_{cyc}(G_1, G_2) = \mathbb{E}_{x\sim p_{data}(x)}[||G_2(G_1(x)) - x||_1] + \mathbb{E}_{y\sim p_{data}(y)}[||G_1(G_2(y)) - y||_1] \tag{7}$$

The loss function of CycleGAN has the following form:

$$\mathcal{L}_{CycleGAN}(G_1, G_2, D_X, D_Y) = \mathcal{L}_{GAN}(G_1, D_Y, X, Y) + \mathcal{L}_{GAN}(G_2, D_X, Y, X) + \lambda\mathcal{L}_{cyc}(G_1, G_2) \tag{8}$$

where $\mathcal{L}_{GAN}$ denotes the adversarial loss, $\mathcal{L}_{cyc}$ denotes the cycle consistency loss and λ controls the importance of $\mathcal{L}_{cyc}$.

In SAR image processing field, the land cover information is very important. In order to keep the land cover information unchanged between the generated images and the input SAR images, we borrow the idea of supervised model pix2pix. The pixel-level MSE loss named L_p is added to the two mappings of CycleGAN. Then CycleGAN is changed from unsupervised model to supervised model and we name it S-CycleGAN, as shown in Fig. 2. The L_p represents for the difference between the generated image and the target real image at pixel level. Therefore, the training samples of S-CycleGAN should be paired and then S-CycleGAN can learn the SAR-to-optical image translation in supervised manner. The loss function has the following form:

$$\mathcal{L}_p(G_1, G_2) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[||G_1(x) - y||_2] + \mathbb{E}_{y,x\sim p_{data}(y,x)}[||G_2(y) - x||_2] \tag{9}$$

In summary, the loss function of S-CycleGAN is shown as bellow:

$$\mathcal{L}_{S-CycleGAN}(G_1, G_2, D_X, D_Y) = \mathcal{L}_{CycleGAN}(G_1, G_2, D_X, D_Y) + \beta\mathcal{L}_p(G_1, G_2) \tag{10}$$

where $\mathcal{L}_p$ denotes the pixel-wise MSE loss and β controls the importance of $\mathcal{L}_p$. Finally the objective of S-CycleGAN is

$$G_1^*, G_2^* = \arg\min_{G_1,G_2}\max_{D_X,D_Y} \mathcal{L}_{S-CycleGAN}(G_1, G_2, D_X, D_Y) \tag{11}$$

The network architectures of the generators and the discriminators of S-CycleGAN are the same as pix2pix, which are the U-Net and PatchGAN, respectively. The detailed information is shown in Fig. 3. The inputs of the generators and discriminators can be SAR images or optical images. If the inputs of the generators are SAR image, then the outputs will be the generated optical images, and vice versa.

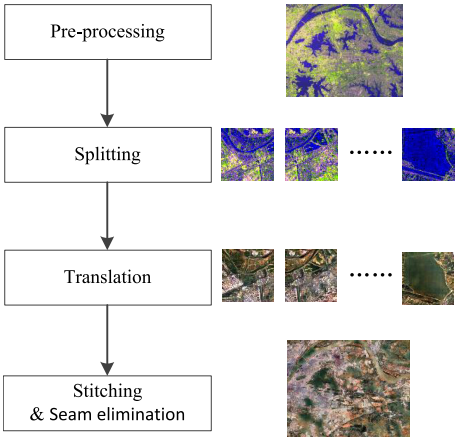


FIGURE 4. The steps of SAR-to-optical image translation.

C. SAR-TO-OPTICAL IMAGE TRANSLATION

Based on S-CycleGAN, a SAR-to-optical image translation framework is discussed in this section. Because S-CycleGAN still translates images in patch level, large SAR images cannot be input to S-CycleGAN directly. We first split large SAR images to small patches and then input the patches to S-CycleGAN model to generate corresponding optical image patches. Finally, we stitch the generated images patches to get large optical images. The detailed steps are shown as Fig. 4.

(a) Pre-processing: correct the geometry and terrain distortion of the SAR image.
(b) Splitting: split the SAR image to small patches with size of 256×256.
(c) Translation: input the small SAR patches to the trained S-CycleGAN to generate optical image patches.
(d) Stitching and seam elimination: stitch the generated optical image patches to get the large optical image. Meanwhile, the traditional weighted averaging method is used for seam elimination.

Because the brightness and tone may be different between the generated optical image patches, obvious seam lines will exist in the generated large optical image if no seam elimination method is used. The traditional weighted averaging method is chosen for seam elimination, which is fast and good enough for proposed framework. The weighted averaging method replaces the overlapped region of two images with the averaged new pixels. The new pixels are calculated based on the distance between the pixels and the two boundaries of the overlapped region, which has the following form:

$$I(x,y) = \begin{cases} I_1(x,y) & (x,y) \in I_1 \\ \omega_1 I_1(x,y) + \omega_2 I_2(x,y) & (x,y) \in I_1 \cap I_2 \\ I_2(x,y) & (x,y) \in I_2 \end{cases} \quad (12)$$

where $I_1(x,y)$ and $I_2(x,y)$ are the pixel values of two images in the coordinates (x,y), and ω_1 and ω_2 are the weights, which are determined by the distances between the pixel (x,y) and the two boundaries of the overlapped region. $\omega_1 + \omega_2 = 1$, $0 < \omega_1 < 1$, and $0 < \omega_1 < 1$.

The pre-condition of using weighted averaging method is that two images should overlap with each other. Therefore, when we split the SAR image to patches in the step (b), each SAR image patch should overlap with its neighborhood patches. Then each generated optical image patch is overlapped with its neighborhood patches and the weighted averaging method can used to eliminate the seam lines in four directions.

D. DATASET

To evaluate the performance of S-CycleGAN, a paired SAR-optical image dataset was generated from the Sentinel-1 (SEN-1) and Sentinel-2 (SEN-2) images. SEN-1 was conducted by the European Space Agency (ESA) and contains two satellites, Sentinel-1A and Sentinel-1B. Both satellites carry a C-band SAR sensor and can provide dual-polarization SAR images in all-weather, day or night. SEN-2, which also contains two satellites (Sentinel-2A and Sentinel-2B), was developed by ESA and can provide multi-spectral RS images.

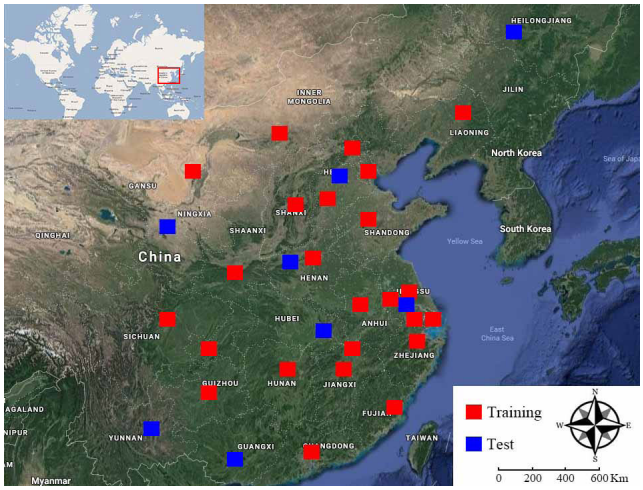


FIGURE 5. The rough locations of scenes in WHU-SEN-City dataset. Red rectangles: The scenes for training. Blue rectangles: The scenes for test.

The paired SAR-optical dataset named WHU-SEN-City[1] covers 32 Chinese cities. The rough locations of the scenes are shown in Fig. 5 and the detailed information is shown in Table 1. WHU-SEN-City dataset contains many kinds of land covers, such as mountains, forests, lakes, rivers, bare lands, buildings, farmlands, vegetation, roads, bridges, and so on. It is a challenge to translate SAR images to optical images on the dataset. The detailed steps of generating WHU-SEN-City dataset are as follows. First, SEN-1 and SEN-2 images were downloaded from the website https://scihub.copernicus.eu/dhus/#/home. As modern cities often develop quickly, the image times of the corresponding SEN-1 and SEN-2 images should be close to each other. Second, the SEN-1 and SEN-2 images were preprocessed, and the steps were shown in Fig. 6. The SEN-2 images

[1] Available at https://github.com/whu-csl/WHU-SEN-City

TABLE 1. The information of SEN-1 and SEN-2 images of WHU-SEN-City dataset. NLA: North latitude bound. WLO: West longitude bound. SLA: South latitude bound. ELO: East longitude bound.

Use	City	Imaging time (SEN-1)	Imaging time (SEN-2)	NLA	WLO	SLA	ELO	Image Size
Training	Beijing	2-7-2018	29-6-2018	40.083	116.304	39.665	117.113	8925×4611
	Changsha	19-4-2018	18-4-2018	28.368	112.886	28.058	113.124	2604×3391
	Chengdu	15-4-2018	17-4-2018	30.70	104.003	30.452	104.287	3133×2737
	Chongqing	20-7-2018	23-7-2018	29.773	106.349	29.493	106.716	4013×3061
	Fuzhou	21-4-2018	19-4-2018	26.162	119.114	25.856	119.495	4143×3327
	Guangzhou	7-3-2018	11-3-2018	23.181	113.053	22.809	113.634	6330×4053
	Guiyang	3-4-2018	9-4-2018	26.774	106.543	26.473	106.855	3415×3295
	Hangzhou	10-5-2018	4-5-2018	30.454	119.846	30.237	120.211	3957×2352
	Hefei	7-8-2018	31-7-2018	31.905	117.185	31.673	117.493	3286×3440
	Huhhot	26-3-2018	25-3-2018	40.905	111.418	40.722	111.88	5091×2016
	Jinan	27-5-2018	30-5-2018	36.85	116.771	36.513	117.062	3211×3719
	Jiujiang	24-8-2018	23-8-2018	29.817	115.713	29.591	116.059	3786×2473
	Nanchang	10-4-2018	14-4-2018	28.834	115.695	28.50	116.064	4040×3656
	Nanjing1	6-6-2018	8-6-2018	32.218	118.376	31.863	118.771	4314×3878
	Nanjing2	6-6-2018	4-4-2018	32.271	118.775	31.908	119.101	3561×3965
	Shanghai	23-4-2018	19-4-2018	31.244	121.242	30.957	121.637	4320×3139
	Shenyang	22-7-2018	23-7-2018	41.807	123.232	41.593	123.59	3943×2359
	Shijiazhuang	13-6-2018	12-6-2018	38.215	114.268	38.02	114.588	3450×2102
	Suzhou	11-4-2018	19-4-2018	31.421	120.431	31.142	120.882	4890×3025
	Taiyuan	5-6-2018	6-6-2018	37.931	112.416	37.734	112.651	2561×2147
	Tianjin	4-6-2018	3-6-2018	39.297	117.181	38.884	117.594	4552×4551
	Xian	5-4-2018	9-4-2018	34.316	108.842	34.10	109.174	3626×2359
	Yangzhou	4-4-2018	9-4-2018	32.414	119.327	32.08	119.619	3162×3616
	Yinchuan	14-6-2018	14-6-2018	38.651	105.985	38.214	106.25	2921×4816
	Zhengzhou1	6-6-2018	7-6-2018	34.994	113.501	34.791	113.755	2744×2193
	Zhengzhou2	6-6-2018	7-8-2018	34.746	113.608	34.506	113.921	3382×2593
Test	Baoding	8-6-2018	12-6-2018	39.05	115.32	38.763	115.637	3456×3128
	Changzhou	4-4-2018	9-4-2018	31.917	119.811	31.614	120.106	3194×3279
	Haerbin	18-5-2018	31-5-2018	45.913	126.458	45.554	126.844	4193×3901
	Kunming	14-2-2018	16-2-2018	25.139	102.565	24.767	102.965	4326×4024
	Lanzhou	9-5-2018	15-5-2018	36.129	103.585	36.02	103.961	4105×1189
	Luoyang	18-6-2018	12-6-2018	34.851	112.297	34.616	112.561	2879×2564
	Nanning	20-8-2018	14-8-2018	22.822	108.232	22.663	108.405	1885×1733
	Wuhan	23-2-2017	15-4-2017	30.687	114.285	30.371	114.681	4296×3429

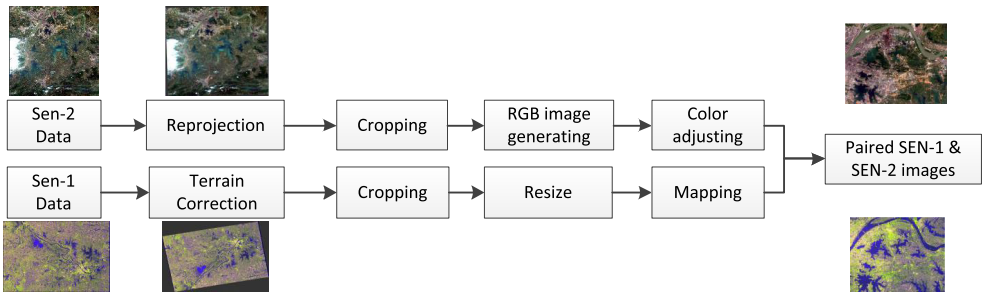


FIGURE 6. The preprocessing of SEN-1 and SEN-2 images for WHU-SEN-City dataset.

were projected to the World Geodetic System 1984 (WSG84) and then cropped according to the longitudes and latitudes in Table 1. The red, green, and blue bands of SEN-2 images were used to generate RGB images and color adjusting was needed to make sure that all SEN-2 images had similar color distribution. For SEN-1 images, the terrain distortion should be corrected and the cropping was also needed. Then the SEN-1 images were resized to the size of corresponding SEN-2 images. For each band of SEN-1 images, the pixel values were converted to dB first, and the minimum and maximum of each band were calculated. Then a linear transformation was used to map the dB values to the range of [0, 255]. After the above-mentioned preprocessing, we got the paired large SEN-1 and SEN-2 images. A free software named SNAP[2] was used to do the reprojection, terrain correction, cropping, RGB image generating, and color adjusting. SNAP contains the Sentinel 1, 2 and 3 Toolboxes, which are free open source toolboxes for the scientific exploitation of earth observation missions. Finally, the paired large SEN-1 and SEN-2 images of different cities were split to training and test sets. Because China is a big country with a vast territory, the land cover types from different regions have different

[2]http://step.esa.int/main/

FIGURE 7. Some paired training samples. Top: Pseudo-color SEN-1 images. Bottom: SEN-2 images.

visual appearances. Both the training and test sets contain the cities from the east, south, west, north, and central regions of China. For the training set, a sliding window was used to generate 18542 paired training samples. The window size was 256×256 and the stride was 128. Fig. 7 shows some of the paired training samples. The test SEN-1 and SEN-2 images were only used to validate the performance of S-CycleGAN and did not used in the training. The test SEN-1 images were used for the SAR-to-optical image translation, which was introduced in Section II-C. The test SEN-2 images were only used to calculate the image quality of the generated optical images quantitatively.

The dual-polarization SEN-1 data are Ground Range Detected (GRD) products and they have two bands, which are VH and VV with a spatial resolution of 20m in range direction and 22m in azimuth direction. We used VH and VV bands to generate a new band VH/VV. Then VH, VV and VH/VV were used as the input data of S-CycleGAN. Only the red, green and blue bands of the SEN-2 images were used. The spatial resolution of the three bands is 10m.

III. EXPERIMENTS

S-CycleGAN was applied to two experiments, one was the SAR-to-optical image translation and the other was the cloud removal. In this section, the training strategy of S-CycleGAN is introduced first, and then the results of SAR-to-optical image translation and cloud removal are illustrated.

A. TRAINING

The paired training set of WHU-SEN-City was used to train the models. We first trained CycleGAN to generate optical images with object structure well reserved. Then S-CycleGAN was fine tuned based on the trained CycleGAN model. The batch size was 24, λ was 100, and β was 100. S-CycleGAN was implemented based on the Pytorch implementation of CycleGAN.[3] Except the batch size, λ, and β, the other training parameters were the same as CycleGAN. For example, the input data standardization, normal distribution initialization, batch normalization, and least square GAN loss were all used by default. Random cropping and flipping were used to avoid overfitting.

B. SAR-TO-OPTICAL IMAGE TRANSLATION

1) RESULTS

The large SAR image translation results are shown in this section. Due to the limitation of space, Fig.8 only shows the results of two test cities, which are Wuhan and Kunming. The quantitative translation results of other test cities are illustrated in Section III-B.2. Some local areas of the generated images are shown in Fig. 9, which can more clearly show the detailed translation results.

The experiment results show that proposed method is capable of SAR-to-optical image translation. The geometry information and color appearance of different land cover types in the generated large optical images are almost consistent with the real SEN-2 optical RS image, such as the buildings, lakes, rivers, forests, bare lands and so on, as shown in Fig. 9. The man-made structures, such as the very high density residential areas of Wuhan and Kunming in Fig. 8, are also well translated. Untrained people can more easily recognize the land cover types from the generated images than from the original SAR images.

The generated image can be used to validate the correctness of SAR image visual recognition results if optical RS data are not available. Some land cover types that have similar backscattering properties are hard to be distinguished in SAR images. It is difficult and time-consuming to compare the differences of the land cover types and sometimes we are still not sure whether the visual recognition results are right or not. With the help of generated image, we can validate the correctness of SAR image visual recognition. For example, the land cover types in the two red boxes in Fig. 9(a) look very similar to each other in the SEN-1 image, but they can be recognized as the vegetation and bare land easily in the generated image. In Fig. 9(b), the land cover types in the red boxes are factory buildings and residences. From the SEN-1 image, it is hard to recognize these two kinds of land cover types. In the generated optical image, the colors of the two land types are different and the visual appearance is consistent with the SEN-2 image.

Generally speaking, S-CycleGAN is a useful method for large SAR image visualization. When there is no available optical RS data, the generated images can provide useful land cover information for untrained people.

2) COMPARISON TO DIFFERENT METHODS

Three image quality assessment (IQA) methods, which are peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [46] index, and chrominance feature-similarity ($FSIM_c$) [47] index, are introduced to evaluate the performance of S-CycleGAN for SAR-to-optical image translation quantitatively. PSNR is a traditional IQA index and a higher PSNR generally indicates that the image is of higher quality. SSIM and $FSIM_C$ calculate the similarity of target images to reference images. If the target image is the same with the reference image, SSIM and $FSIM_C$ will equate to 1. In this paper, the target images are the generated large optical images and the reference images are the SEN-2 images. CycleGAN, DualGAN and pix2pix are used to compare the performance with S-CycleGAN. The image translation results of different methods of Wuhan are shown in Fig. 10 and some local results

[3] https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix

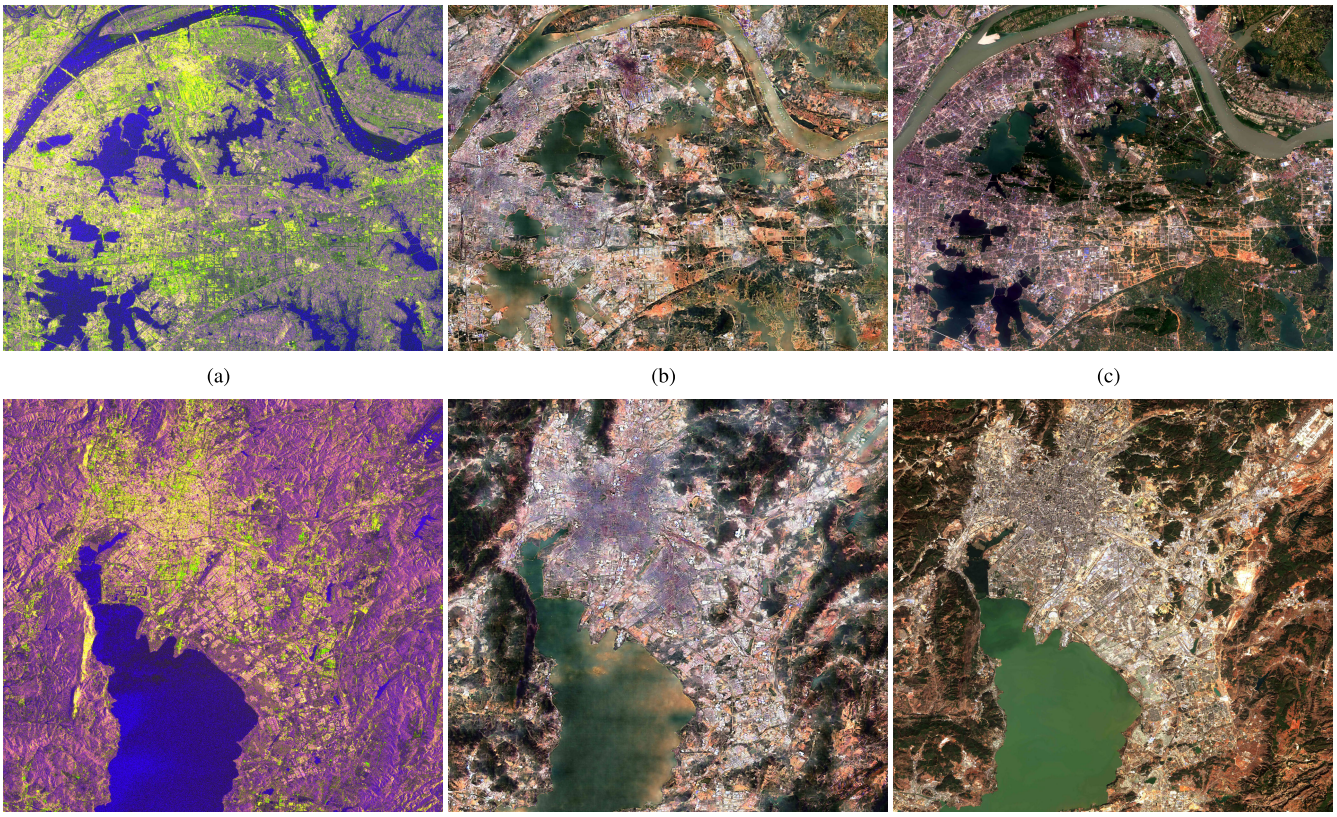


FIGURE 8. Translation results of Wuhan and Kunming cities. (a)(d) SEN-1 pseudo-color images of Wuhan and Kunming. (b)(e) Generated optical images. (c)(f) SEN-2 images.

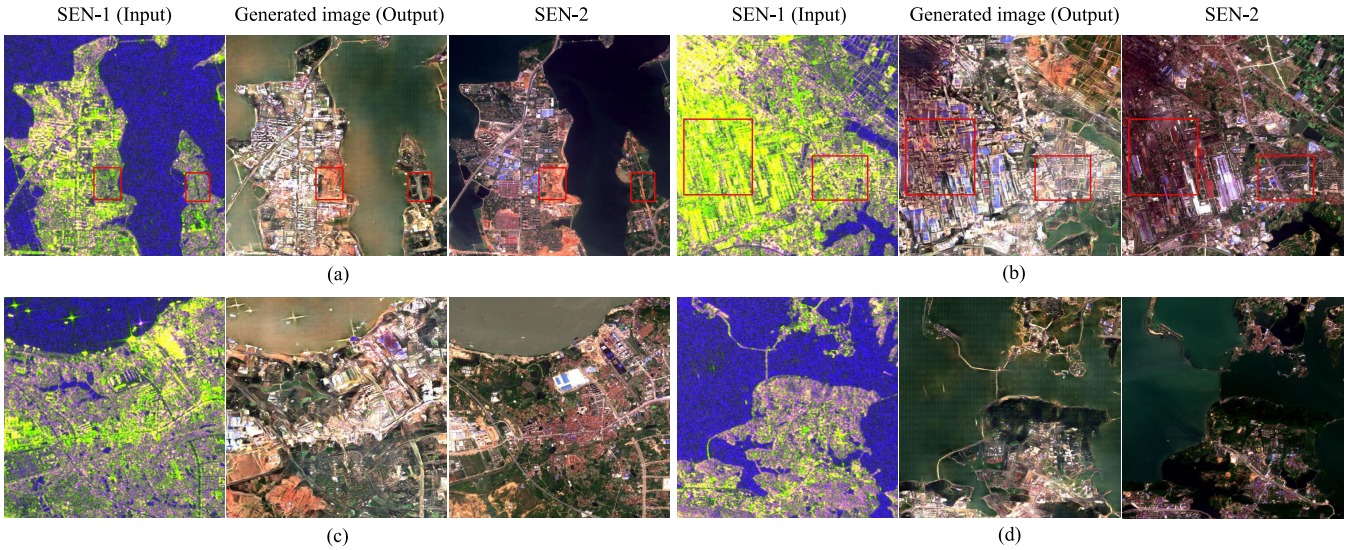


FIGURE 9. Local translation results. The size of all images is 500 × 500.

are shown in Fig. 11. The IQA results of 8 test cities are illustrated in Table 2.

First, we discuss the translation results in Fig. 10 and Fig. 11. CycleGAN and DualGAN cannot well translate some land cover information but pix2pix and S-CycleGAN are better. The land cover information is very important in remote sensing image processing field, so it is a severe disadvantage of CycleGAN and DualGAN. For example, the land cover

FIGURE 10. Different methods for SAR-to-Optical image translation results. (a) SEN-2 optical RS image. (b) CycleGAN. (c) DualGAN. (d) Pix2pix. (e) S-CycleGAN.

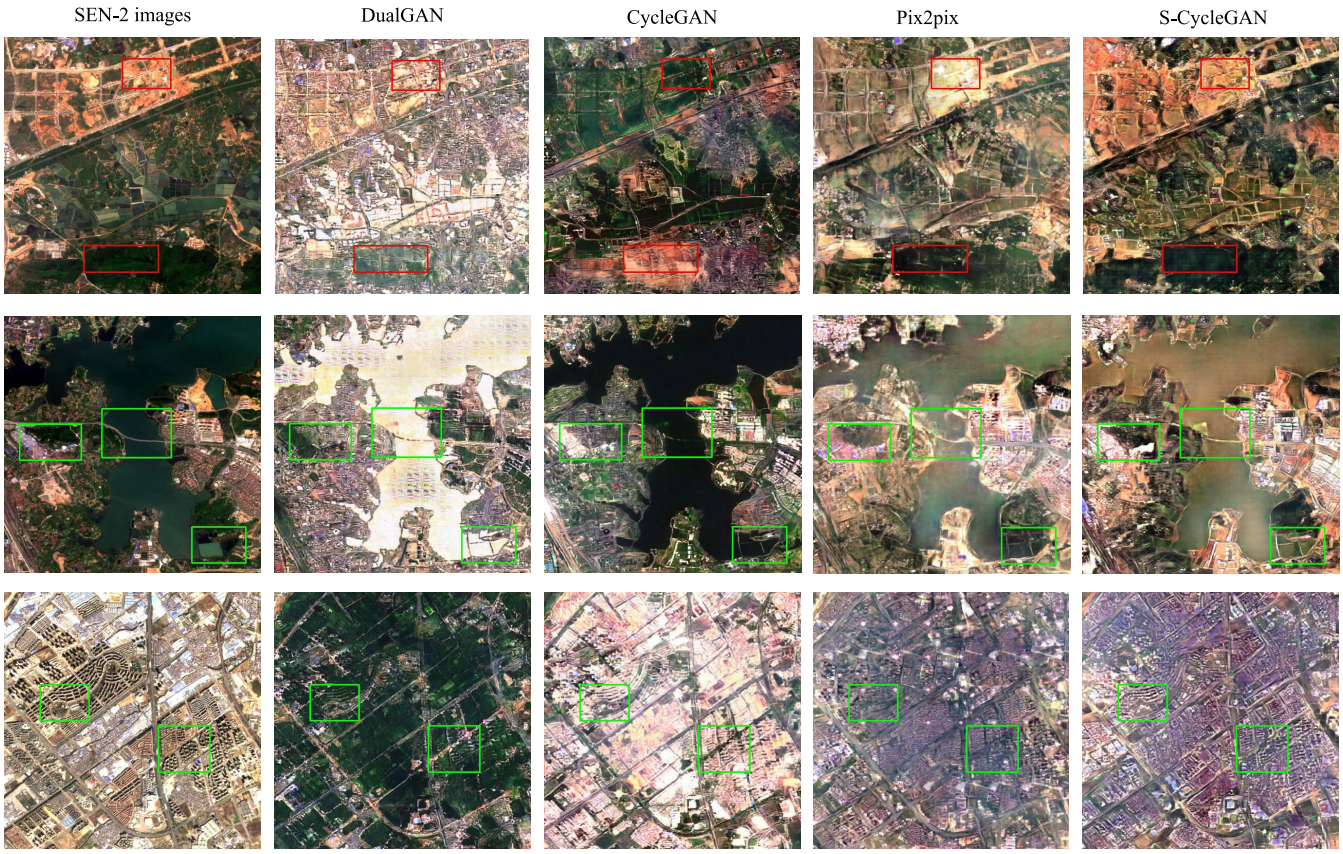


FIGURE 11. The local translation results of different methods. The size of all images is 500 × 500. The images from left to right are SEN-2 optical RS images, DualGAN, CycleGAN, pix2pix, and S-CycleGAN.

TABLE 2. IQA results of different methods.

IQA	Method	Baoding	Changzhou	Haerbin	Kunming	Lanzhou	Luoyang	Nanning	Wuhan	Average
PSNR	DualGAN	8.521	8.368	7.980	7.663	5.691	6.131	10.717	6.696	7.721
	CycleGAN	10.076	10.541	9.991	9.671	8.165	**10.759**	7.723	10.797	9.715
	pix2pix	11.272	11.202	11.551	12.678	8.426	8.959	11.857	11.017	10.870
	S-CycleGAN	**11.745**	**11.478**	**11.713**	**12.930**	**9.360**	9.620	**11.410**	**11.517**	**11.203**
SSIM	DualGAN	-0.095	-0.162	-0.177	0.013	-0.102	-0.021	0.129	-0.077	-0.062
	CycleGAN	0.217	0.291	0.299	0.181	0.192	0.165	-0.004	0.245	0.198
	pix2pix	0.379	0.376	0.361	0.467	0.189	0.190	0.261	0.333	0.319
	S-CycleGAN	**0.472**	**0.421**	**0.410**	**0.525**	**0.294**	**0.227**	**0.267**	**0.399**	**0.377**
$FSIM_c$	DualGAN	0.706	0.708	0.725	0.682	0.643	0.570	0.635	0.688	0.669
	CycleGAN	0.707	0.710	0.706	0.680	0.639	0.661	0.620	0.708	0.679
	pix2pix	0.741	0.731	0.718	0.759	0.670	**0.690**	0.691	0.729	0.716
	S-CycleGAN	**0.764**	**0.740**	**0.727**	**0.777**	**0.678**	0.678	**0.707**	**0.744**	**0.727**

types in the red boxes in Fig. 11 are the bare land and forest. We cannot figure out these two land cover types in the translation results of CycleGAN and DualGAN but we can clearly recognize them in the results of pix2pix and S-CycleGAN. One of the advantages of CycleGAN and DualGAN is that the structure information of some objects can be well kept,

such as the objects in the green boxes in Fig. 11. The objects in the results of pix2pix are blurry and twisty. S-CycleGAN not only can translate the land cover information but also can keep the structure information of the objects. Compare with CycleGAN, DualGAN and pix2pix, S-CycleGAN achieves the best translation results.

Second, we analyze the IQA results of the four methods in Table 2. DualGAN achieves the worst IQA results. The PSNR, SSIM and $FSIM_C$ results of the 8 cities are almost the lowest and the average PSNR and SSIM are much lower than the other three methods. The IQA results of CycleGAN are better than DuanGAN but also worse than pix2pix and S-CycleGAN. Except the Luoyang city, the IQA results of S-CycleGAN are all better than pix2pix. The Luoyang scene contains many bare mountains, bare farmlands and a few forests. The land cover types in the Luoyang SEN-2 image all look yellow. The overall tone of the S-CycleGAN result does not match the Luoyang SEN-2 image, so the IQA results of S-CycleGAN are a little lower. In a word, S-CycleGAN achieves the best average IQA results. The IQA results again illustrate that the performance of S-CycleGAN is the best.

Finally, we have the following conclusions. In SAR-to-optical image translation task, the performance of CycleGAN and DualGAN is lower than pix2pix and S-CycleGAN. CycleGAN and DualGAN are unsupervised image-to-image translation models and may need more special tuning for SAR-to-optical image translation. By combining the advantages of CycleGAN and pix2pix, the performance of S-CycleGAN is the best. S-CycleGAN can well keep both the structure and land cover information and the visual appearances of the generated images are close to real SEN-2 images.

3) FAILURE CASES

The generated optical images also have some problems. First, some special land cover types, such as roads, playgrounds, airport runways, shadows of mountains, rivers and lakes with different colors, are not translated properly. Because of the imaging mechanism of SAR sensor, some of the special land cover types have similar backscattering properties and are hard to be distinguished in the SEN-1 SAR images, so they are not translated well in the generated images. For example, the backscattering property of airport runways is similar to water. The training samples of water are much more than airport runways, so the airport runway in the green box in Fig. 12(a) is translated to water. In Fig. 12(b), the color of the river and the shadows of the mountains are not translated well, too. Second, the texture information of some land cover types is not well learned. From the red boxes in Fig. 12(a) we can see that, the forest in the generated image is just green but has no texture information. The reason could be that the texture information of SAR and optical RS images is different and not matched. The mismatched textures lead to the smoothness in the generated images and the texture information is lost.

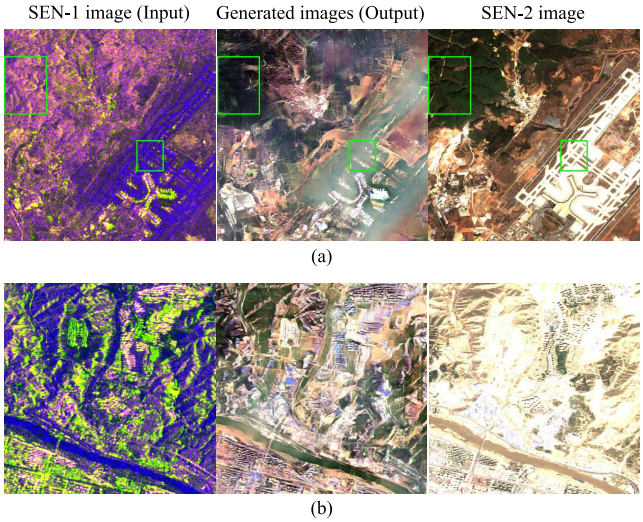


FIGURE 12. Examples of failed translation results. The size of all images is 500 × 500.

C. CLOUD REMOVAL

The SEN-1 and SEN-2 images used for cloud removal are shown in Fig. 13(a) and Fig. 13(b), which cover the scene of northwest Wuhan. The north latitude is 30.687, the west longitude is 113.387, the south latitude is 30.371, and the east longitude is 114.283. The imaging time of the SEN-1 and SEN-2 images is March 26, 2018 and the image size is 4224 × 3328. The cloud areas in red color in Fig. 13(c) were labeled manually. Fig. 13(d) is the generated large optical image, which is translated from Fig. 13(a). It can be used to reconstruct the thick cloud areas of the cloudy SEN-2 RS image.

1) EXPERIMENT STEPS

The labeled cloud areas of the SEN-2 image are replaced with the same areas of the generated RS image. To minimize the differences of brightness and tone between the cloudy RS image and the generated image, the uncontaminated areas from the cloudy RS image are used to fine tune the trained S-CycleGAN. The detailed steps are as follows:

(a) Label the pixels of clouds and shadows in the cloudy SEN-2 image manually.
(b) Generate paired samples from the SEN-1 image and the uncontaminated areas of SEN-2 image and fine tune the trained S-CycleGAN.
(c) Use the fine tuned S-CycleGAN to generate large optical image from the SEN-1 image, which is illustrate in Section II-C.
(d) Replace the labeled pixels with the pixels in the generated image. Meanwhile, the weighted averaging method is also used to eliminate the seam lines.

Although S-CycleGAN is fine tuned, the brightness and tone are still a little different between the generated optical image and the cloudy SEN-2 image and the seam lines also

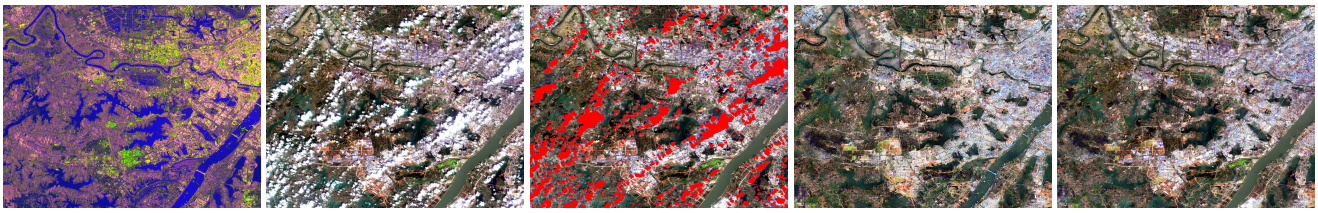


FIGURE 13. Cloud removal. (a) Pseudo-color image of SEN-1. (b) Cloudy SEN-2 image. (c) Ground truth of clouds. (d) Generated optical image. (e) Cloud removal result.

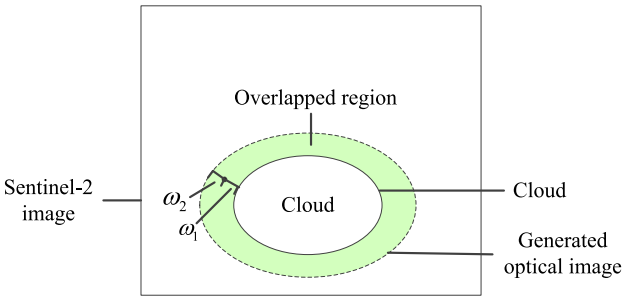


FIGURE 14. Weighted averaging method for seam elimination. The rectangle denotes the real optical RS image. The ellipse with the solid line denotes the cloud. The ellipse with the dotted line denotes the generated optical image, which is used to replace the cloud. The area with green color denotes the overlapped region of the SEN-2 image and the generated image.

exist in the result image of cloud removal. We improve the weighted averaging method to eliminate arbitrary curved seam lines. As shown in Fig. 14, the rectangle denotes the SEN-2 image. The ellipse with the solid line denotes the cloud and the ellipse with the dotted line denotes the generated image. The area between the two ellipses is the overlapped region. The pixels in the overlapped region are also calculated by (12), where I_1 and I_2 denote the SEN-2 image and the generated image, respectively. The weights ω_1 and ω_2 are calculated based on the distance between each pixel and the boundaries of the overlapped region.

2) CLOUD REMOVAL RESULTS

Fig. 13(e) shows the cloud removal result and the local results are shown in Fig. 15. All kinds of clouds, such as big thick clouds and lines of small clouds, can be removed. It is feasible to use the generated image to reconstruct the pixels of cloud areas. The SEN-1 and SEN-2 images were imaged at the same day, so no temporal difference between the generated optical image and the SEN-2 image.

In cloudy days, the optical RS data are partly unavailable and it is impossible to observe the land cover types under thick clouds visually. SAR has certain penetration capacity and can work under all weather conditions. The images that translated from SAR images can reconstruct the contaminated pixels of an optical RS image and then untrained people can observe the land cover types of the cloud areas. Therefore, the proposed method shows its ability of assisting in land cover visual recognition in cloudy days.

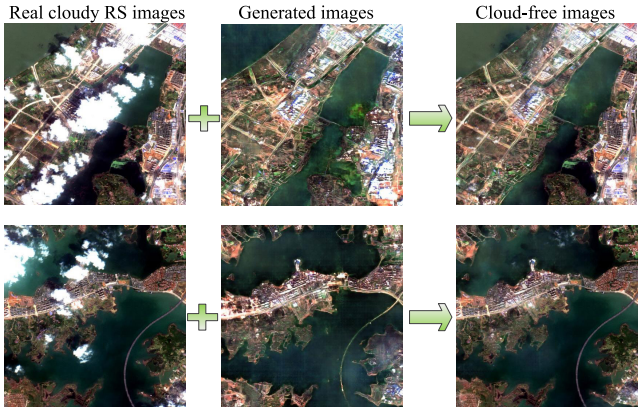


FIGURE 15. Local results of cloud removal. The size of all images is 500 × 500. The images from left to right are SEN-2 cloudy RS images, generated images and cloud-free images.

IV. DISCUSSION

A. ASSISTANCE OF VISUAL RECOGNITION

Optical RS data are usually unavailable in bad weather days or partly unavailable in cloudy days. SAR can provide important land cover information in all weather conditions, but untrained people are difficult to recognize the land cover types from SAR images. S-CycleGAN can be used to translate SAR images to optical images and then untrained people can recognize the land cover types from the generated images instead of SAR or unavailable optical RS images.

Two examples illustrate that the generated images can assist in land cover type visual recognition when lack of useful optical RS data. The cloud removal introduced in Section III-C could be a typical example. The land cover information under thick clouds is unrecognizable in optical RS images. The image which generated from a SAR image can provide useful land cover information and reconstruct the cloud areas of the cloudy RS image. Another example is the rainforest monitoring and it is an important Earth observation task [48]–[50]. The Amazon rainforest is always covered by clouds and the rainforest monitoring through optical RS images is difficult. With the help of generated images, the visual observation of Amazon rainforest becomes much easier. Untrained people can visually monitor the land cover changes in the Amazon rainforest, such as the deforestation and destruction.

B. POTENTIAL APPLICATIONS

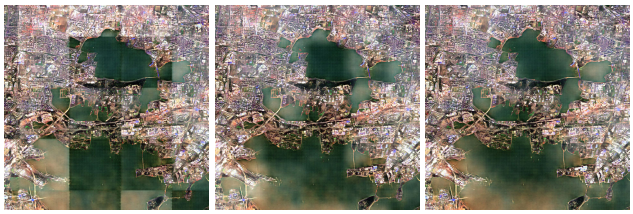
The trained S-CycleGAN learns rich optical RS data features, so the generated optical images can be used in some RS image processing. Here we discuss two potential applications of the generated images.

The SAR and optical RS data fusion has been reported to be useful as SAR and optical data are able to provide complementary information for each other [51]. Zhang *et al*. [51] studied the feature normalization methods on SAR and optical data fusion for land cover classification. Yousif *et al*. [52] investigated the fusion of SAR and optical images for change detection application. Due to clouds or bad weather, the optical RS data are not always available, so the generated images could be an alternative. By using both of SAR and generated optical images, the performance of SAR image processing could be improved.

The generated images can be used to calculate the visible-band vegetation indices (VIs) when optical RS data are unavailable. VIs can indicate whether the target being observed contains live green vegetation or not, and are widely used in agricultural remote sensing monitoring and so on [53]–[55]. In [56], the selected VIs derived from Sentinel-2 data and three machine learning methods were used to estimating the defoliation of Scots pine stands. Visible-band VIs are usually calculated from one or more visible bands, such as excess green index (ExG) [57] and normalized green-red difference index (NGRDI) [53]. The generated images also contain three visible bands and could be alternative data for visible-band VI calculation. Moreover, the other bands of SEN-2 data could be added to our dataset and S-CycleGAN can learn to generate the near-infrared band or other bands. Then some VIs, which need near-infrared data, e.g. the normalized difference vegetation index [58], can also be calculated from the generated data.

C. DISCUSSION ON SEAM ELIMINATION

The weighted averaging method is used to eliminate the seam lines for SAR-to-optical image translation and cloud removal. The size of the overlapped regions affects the seam elimination results. We adjust the size of overlapped regions according to experiment results and finally get an appropriate value.



(a) (b) (c)

FIGURE 16. Results of image stitching with different sizes of overlapped regions. The size of all images is 1000 × 1000. (a) The size of the overlapped regions is 16 pixels. (b) The size of the overlapped regions is 56 pixels. (c) The size of the overlapped regions is 128 pixels.

Fig. 16 shows the SAR-to-optical image translation results with different sizes of the overlapped regions. When the size of the overlapped regions is 16 pixels or 56 pixels, obvious seam lines still exist in the mosaicked images. When the size of the overlapped regions is 128 pixels, i. e. one half of an image patch is overlapped with its neighborhood image patches, the seam lines are eliminated without loss of too many details.

(a) (b) (c)

FIGURE 17. Results of different sizes of overlapped regions for cloud removal. The size of all images is 500 × 500. (a) The SEN-2 cloudy image. (b) The size of the overlapped region is 10 pixels. (c) The size of the overlapped regions is 50 pixels.

Fig. 17 shows the cloud removal results with different sizes of the overlapped regions. When the size of the overlapped regions is 10 pixels, seam lines still can be seen. When the size of the overlapped regions is 50 pixels, the seam lines are almost disappeared. No need to increase the size of the overlapped regions over 50 pixels.

V. CONCLUSION

In this paper, S-CycleGAN is proposed to translate large SAR images to optical images and can be used in practical applications. The land cover types in the generated optical images can be recognized visually for untrained people. S-CycleGAN takes the advantages of CycleGAN and pix2pix, so it can characterize both the land cover information and the object structure information. Experiment results on WHU-SEN-City dataset show that the performance of S-CycleGAN is better than the famous DualGAN, CycleGAN and pix2pix in SAR-to-optical image translation task. In addition, the generated optical image is used for cloud removal and the cloud remove results are good. Based on the superior performance of S-CycleGAN, the generated images can aid in land cover visual recognition for untrained people when optical RS data are unavailable. The potential applications of the generated images in SAR and optical image fusion and VI calculation are also discussed.

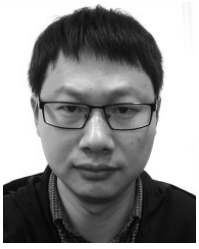
Although S-CycleGAN is powerful for SAR-to-optical image translation, some special land covers and the texture information are not translated well. In the future, we will consider introducing some more data from other sources and further improve the translation results. We will also investigate the SAR and generated image fusion for SAR image classification, change detection and so on.

REFERENCES

[1] J.-S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*. Boca Raton, FL, USA: CRC Press, 2009.

[2] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.
[3] C. O. Dumitru, S. Cui, D. Faur, and M. Datcu, "Data analytics for rapid mapping: Case study of a flooding event in Germany and the tsunami in Japan using very high resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 1, pp. 114–129, Jan. 2015.
[4] A. C. van den Broek, A. Smith, and A. Toet, "Visual interpretation of polarimetric SAR imagery," *Proc. SPIE*, vol. 4541, pp. 169–179, Jan. 2002.
[5] Y. Chambenoit, N. Classeau, E. Trouve, and J.-P. Rudant, "Performance assessment of multitemporal SAR images' visual interpretation," in *Proc. IEEE Geosci. Remote Sens. Symp.*, vol. 6, Jul. 2003, pp. 3911–3913.
[6] J. Tao, G. Palubinskas, P. Reinartz, and S. Auer, "Interpretation of SAR images in urban areas using simulated optical and radar images," in *Proc. Urban Remote Sens. Event*, Apr. 2011, pp. 41–44.
[7] S. G. Dellepiane and E. Angiati, "A new method for cross-normalization and multitemporal visualization of SAR images for the detection of flooded areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2765–2779, Jul. 2012.
[8] Y. Jiao, Y. Niu, L. Liu, G. Zhao, G. Shi, and F. Li, "Dynamic range reduction of SAR image via global optimum entropy maximization with reflectivity-distortion constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2526–2538, May 2018.
[9] B. Zhang, C. Wang, H. Zhang, and F. Wu, "An adaptive two-scale enhancement method to visualize man-made objects in very high resolution SAR images," *Remote Sens. Lett.*, vol. 6, no. 9, pp. 725–734, 2015.
[10] X. Zhou, C. Zhang, and S. Li, "A perceptive uniform pseudo-color coding method of SAR images," in *Proc. CIE Int. Conf. Radar*, Oct. 2006, pp. 1–4.
[11] Z. Li, J. Liu, and J. Huang, "Dynamic range compression and pseudo-color presentation based on Retinex for SAR images," in *Proc. IEEE Int. Conf. Comput. Sci. Softw. Eng.*, Dec. 2008, pp. 257–260.
[12] Q. Deng, Y. Chen, W. Zhang, and J. Yang, "Colorization for polarimetric SAR image based on scattering mechanisms," in *Proc. Congr. Image Signal Process.*, vol. 4, May 2008, pp. 697–701. doi: 10.1109/CISP.2008.366.
[13] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[14] Z. Zhang, H. Wang, F. Xu, and Y.-Q. Jin, "Complex-valued convolutional neural network and its application in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7177–7188, Dec. 2017.
[15] R. Shang, J. Wang, L. Jiao, R. Stolkin, B. Hou, and Y. Li, "SAR targets classification based on deep memory convolution neural networks and transfer parameters," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 8, pp. 2834–2846, Aug. 2018.
[16] S. De, L. Bruzzone, A. Bhattacharya, F. Bovolo, and S. Chaudhuri, "A novel technique based on deep learning and a synthetic target database for classification of urban areas in PolSAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 154–170, Jan. 2018.
[17] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. CVPR*, Jun. 2015, pp. 3431–3440.
[18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 3, 2014, pp. 2672–2680.
[19] M. Mirza and S. Osindero, "Conditional generative adversarial nets," Nov. 2014, *arXiv:1411.1784*. [Online]. Available: https://arxiv.org/abs/1411.1784
[20] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2016, pp. 5967–5976.
[21] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," in *Proc. 33rd Int. Conf. Mach. Learn.*, 2016, pp. 1060–1069.
[22] 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. 2242–2251.
[23] 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. 2868–2876.
[24] Q. Song, F. Xu, and Y. Q. Jin, "Radar image colorization: Converting single-polarization to fully polarimetric using deep neural networks," *IEEE Access*, vol. 6, pp. 1647–1661, 2017.
[25] D. Ao, C. O. Dumitru, G. Schwarz, and M. Datcu, "Dialectical GAN for SAR image translation: From Sentinel-1 to TerraSAR-X," Jul. 2018, *arXiv:1807.07778*. [Online]. Available: https://arxiv.org/abs/1807.07778
[26] P. Wang and V. M. Patel, "Generating high quality visible images from SAR images using CNNs," in *Proc. IEEE Radar Conf.*, Apr. 2018, pp. 0570–0575.
[27] N. Merkle, S. Auer, R. Muller, and P. Reinartz, "Exploring the potential of conditional adversarial networks for optical and SAR image matching," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 6, pp. 1811–1820, Jun. 2018.
[28] L. Liu and B. Lei, "Can SAR images and optical images transfer with each other?" in *Proc. IGARSS*, Jul. 2018, pp. 7019–7022.
[29] A. Ley, O. Dhondt, S. Valade, R. Haensch, and O. Hellwich, "Exploiting GAN-based SAR to optical image transcoding for improved classification via deep learning," in *Proc. 12th Eur. Conf. Synth. Aperture Radar (EUSAR)*, Jun. 2018, pp. 1–6.
[30] K. Enomoto, K. Sakurada, W. Wang, N. Kawaguchi, M. Matsuoka, and R. Nakamura, "Image translation between SAR and optical imagery with generative adversarial nets," in *Proc. IGARSS*, Jul. 2018, pp. 1752–1755.
[31] M. Schmitt, L. Hughes, K. Körner, and X. X. Zhu, "Colorizing Sentinel-1 SAR images using a variational autoencoder conditioned on Sentinel-2 imagery," *Int. Arch. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 42, no. 2, pp. 1045–1051, 2018.
[32] M. Schmitt, L. H. Hughes, and X. X. Zhu, "The SEN1-2 dataset for deep learning in SAR-Optical data fusion," *ISPRS Ann. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 4, pp. 141–146, Sep. 2018.
[33] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. CVPR*, Jun. 2016, pp. 2536–2544.
[34] R. Eckardt, C. Berger, C. Thiel, and C. Schmullius, "Removal of optically thick clouds from multi-spectral satellite images using multi-frequency SAR data," *Remote Sens.*, vol. 5, no. 6, pp. 2973–3006, 2013.
[35] C. Zhang, W. Li, and D. J. Travis, "Restoration of clouded pixels in multispectral remotely sensed imagery with cokriging," *Int. J. Remote Sens.*, vol. 30, no. 9, pp. 2173–2195, 2009.
[36] Q. Cheng, H. Shen, L. Zhang, and P. Li, "Inpainting for remotely sensed images with a multichannel nonlocal total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 175–187, Jan. 2014.
[37] D. Lu, "Detection and substitution of clouds/hazes and their cast shadows on IKONOS images," *Int. J. Remote Sens.*, vol. 28, no. 18, pp. 4027–4035, 2007.
[38] B. Huang, Y. Li, X. Han, Y. Cui, W. Li, and R. Li, "Cloud removal from optical satellite imagery with SAR imagery using sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1046–1050, May 2015.
[39] K. Enomoto, K. Sakurada, W. Wang, H. Fukui, M. Matsuoka, R. Nakamura, and N. Kawaguchi, "Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets," *Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1533–1541.
[40] C. Grohnfeldt, M. Schmitt, and X. Zhu, "A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from Sentinel-2 images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 1726–1729.
[41] P. Singh and N. Komodakis, "Cloud-Gan: Cloud removal for Sentinel-2 imagery using a cyclic consistent generative adversarial networks," in *Proc. IGARSS*, Jul. 2018, pp. 1772–1775.
[42] J. D. Bermudez, P. N. Happ, D. A. B. Oliveira, and R. Q. Feitosa, "SAR to optical image synthesis for cloud removal with generative adversarial networks," *ISPRS Ann. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 4, no. 1, pp. 5–11, 2018.
[43] D. Eigen and R. Fergus, "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 2650–2658.
[44] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*. Cham, Switzerland: Springer, 2015, pp. 234–241.
[45] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 694–711.
[46] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[47] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[48] M. J. Delgado-Aguilar, F. E. Fassnacht, M. Peralvo, C. P. Gross, and C. B. Schmitt, "Potential of TerraSAR-X and Sentinel 1 imagery to map deforested areas and derive degradation status in complex rain forests of Ecuador," *Int. Forestry Rev.*, vol. 19, no. 1, pp. 102–118, 2010.
[49] J. R. Santos, C. C. Freitas, L. S. Araujo, L. V. Dutra, J. C. Mura, F. F. Gama, L. S. Soler, and S. J. S. Sant'Anna, "Airborne p-band SAR applied to the aboveground biomass studies in the Brazilian tropical rainforest," *Remote Sens. Environ.*, vol. 87, no. 4, pp. 482–493, 2003.
[50] T. Kraus, M. Schmidt, S. W. Dech, and C. Samimi, "The potential of optical high resolution data for the assessment of leaf area index in East African rainforest ecosystems," *Int. J. Remote Sens.*, vol. 30, pp. 5039–5059, Sep. 2009.
[51] H. Zhang, H. Lin, and Y. Li, "Impacts of feature normalization on optical and SAR data fusion for land use/land cover classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1061–1065, May 2015.
[52] O. Yousif and Y. Ban, "Fusion of SAR and optical data for unsupervised change detection: A case study in Beijing," in *Proc. IEEE Joint Urban Remote Sens. Event (JURSE)*, Mar. 2017, pp. 1–4.
[53] A. J. Pérez, F. López, J. V. Benlloch, and S. Christensen, "Colour and shape analysis techniques for weed detection in cereal fields," *Comput. Electron. Agricult.*, vol. 25, no. 3, pp. 197–212, Feb. 2000.
[54] T. Hague, N. D. Tillett, and H. Wheeler, "Automated crop and weed monitoring in widely spaced cereals," *Precis. Agricult.*, vol. 7, no. 1, pp. 21–32, 2006.
[55] T. Jingguo, W. Shudong, Z. Lifu, W. Taixia, S. Xiaojun, and J. Hailing, "Evaluating different vegetation index for estimating lai of winter wheat using hyperspectral remote sensing data," in *Proc. IEEE 7th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2015, pp. 1–4.
[56] P. Hawryło, B. Bednarz, P. Wężyk, and M. Szostak, "Estimating defoliation of Scots pine stands using machine learning methods and vegetation indices of Sentinel-2," *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 194–204, 2018.
[57] D. M. Woebbecke, G. E. Meyer, K. Von Bargen, and D. A. Mortensen, "Color indices for weed identification under various soil, residue, and lighting conditions," *Trans. ASAE*, vol. 38, no. 1, pp. 259–269, 1995.
[58] J. Verrelst, M. E. Schaepman, B. Koetz, and M. Kneubühler, "Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data," *Remote Sens. Environ.*, vol. 112, no. 5, pp. 2341–2353, May 2008.



LEI WANG received the B.S. degree in circuit and system and the M.S. degree in communication engineering from Wuhan University, Wuhan, China, in 2009 and 2011, respectively, where he is currently pursuing the Ph.D. degree with the School of Electronic Information.

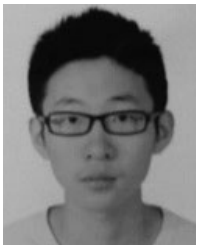
His research interests include deep learning and polarimetric SAR image processing.

XIN XU received the B.S. degree in electronic engineering, the M.S. degree in signal and information processing, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1989, 1996, and 2003, respectively, where he is currently a Professor with the School of Electronic Information.

His research interests include image understanding and statistical signal processing with applications to remote sensing.

YUE YU received the B.Eng. degree in electronic engineering from Hubei University, Wuhan, China, in 2017.

He is currently a Graduate Researcher with Wuhan University. His research focuses on remote sensing image processing.



RUI YANG received the B.Eng. degree in electronic engineering from Wuhan University, Wuhan, China, in 2017, where he is currently a Graduate Researcher.

His research interests include machine learning and computer vision, especially applications on remote sensing imagery processing.

RONG GUI received the M.E. degree from Wuhan University, Wuhan, China, in 2016, where she is currently pursuing the Ph.D. degree in signal and information processing.

Her research interests include laser measurement technology, pattern recognition, and machine learning in remote sensing.

ZHAOZHUO XU (S'13) received the B.Eng. degree in electronic engineering from Wuhan University, Wuhan, China, in 2017.

He is currently a Graduate Researcher with Stanford University. His research interests include machine learning and computer vision, especially applications on object detection and time series modeling.

FANGLING PU received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, in 2005. She was a Research Assistant with the Institute for Theoretical Information Technology, RWTH Aachen University, in 2007. Since 2000, she has been an Associate Professor with the School of Electronic Information, Wuhan University.

Her research interests include system and networking, signal and information processing, and data mining.