Translation between High- and Low-frequency SAR Images using Cycle-Consistent Conditional Adversarial Network

Daquan He, Tian Jin*, Yongkun Song, Chen Wu

College of Electronic Science and Technology, National University of Defense Technology, No.109, Deya Road, Changsha 410003, China

*tianjin@nudt.edu.cn

Abstract. The bi-frequency (high- and low) synthetic aperture radar (SAR) images cannot be directly compared due to their distinct statistical properties. To diminish their statistical difference, we manage to translate the bi-frequency SAR images into one another. Therefore, we propose a cycle-consistent conditional adversarial network to achieve the goal. The cycle-consistency criteria in the Cycle GAN and the conditional generation adversarial networks in the Pix2Pix are integrated to construct the cycle-consistent conditional adversarial network. Experiments on Ku-band and P-band SAR images validate that our method outperforms Cycle GAN and Pix2Pix.

Keywords: SAR, Image Translation, Cycle GAN, Pix2Pix.

1. Introduction

Due to the ability of actively imaging at high resolution in all-day and all-weather, synthetic aperture radar (SAR) is widely applied in the military and surveillance fields. Bi-frequency SAR system can simultaneously obtain paired images in two wavelengths. However, the challenge remains in the interpretation of bi-frequency SAR images due to their distinct statistical properties and physical backscattering mechanism.

Recently, deep learning technologies have been widely applied in remote sensing image processing, such as terrain surface classification [1], target detection [2], parameter inversion [3], despeckling [4] et al. Especially, inspired by the excellent data fitting performance of the generative adversarial networks (GAN) [5], GAN had been adopted to learn the mapping between source images and target images and generate the desired images with high similarity. [6] and [7] employed conditional GAN (cGAN) to fuse SAR image and optical image to synthesize optical images. Niu et al. [8] utilized cGAN to translate the SAR image and optical image into a homogeneous domain for change detection. Ao et al. [9] applied GAN in C-band and X-band SAR images for super-resolution. Marmanis et al. [10] augment SAR datasets by using the boundary equilibrium generative adversarial networks for improving the classification accuracy. Fu et al. [11] translate SAR image into optical image for SAR image interpretation by using cascaded-residual adversarial networks.

In this paper, Pix2Pix [12] and Cycle GAN [13] are combined to achieve the translation between bi-frequency SAR images. Experiments on P-band and Ku-band SAR images show that our method could keep the structural details while diminishing their distribution difference. The rest of this paper is
organized as follows. Section 2 presents the proposed method, where the architecture and the loss function are described. Section 3 and 4 show the experiments and conclusions respectively.

2. Methodology

2.1. Cycle-Consistent Conditional Adversarial Network
As shown in Fig 1, the cycle-consistent conditional network is designed to achieve the translation between low-frequency SAR image (LF SAR) and high-frequency SAR image (HF SAR). The network consists of two branches of HF SAR to translated LF SAR and LF SAR to translated HF SAR, which are highlighted by the red path and the blue path respectively. In the branch of HF SAR to translated LF SAR, the generator G1 is trained to translate the HF SAR into translated LF SAR, and the discriminator D1 is trained to distinguish the real LF SAR and the translated LF SAR. The translated LF SAR will be translated into reconstructed HF SAR by the generator G2. The $L_1$ distance between the reconstructed HF SAR and the real HF SAR is employed as the cyclic loss 1. A conventional binary classification loss is employed to train the discriminator D1, while the combination of cGAN loss 1 and cyclic loss 1 is applied to train the generator G1. Conversely, the other branch is trained to translate the LF SAR into translated HF SAR and then generate reconstructed LF SAR. The $L_1$ distance between the reconstructed LF SAR and the real LF SAR is employed as the cyclic loss 2. The combination of cGAN loss 2 and cyclic loss 2 is utilized to train the generator G2. The generators of G1 and G2 are trained jointly, while the discriminators of D1 and D2 are trained separately.

Our image translation network is based on two well-known image-to-image translation networks of Pix2Pix [12] and Cycle GAN [13]. The major structure of our networks follows the cycle-consistent architecture of the Cycle GAN. The Patch-GAN and the U-Net [12] in the Pix2Pix are used as discriminator and generator respectively. The generator and discriminator are depicted in Fig. 2 (a) and (b), respectively. The input and output of the generator are the real SAR image and the translated SAR image respectively, and their sizes are $128 \times 128 \times 1$. The input of the discriminator is the real SAR image or the translated image, and the output of the discriminator is a probability map. The input size and the output size of the discriminator are $128 \times 128 \times 1$ and $16 \times 16 \times 1$ respectively. Both of the generator and discriminator are assembled by a set of units of convolution filter, deconvolution filter, activation function, and concatenation. The arrows indicate the flow of the data, and the symbol ‘+’ denotes concatenation. ‘Leaky ReLU’, ‘Tanh’, and ‘Sigmoid’ are three activation functions. ‘Conv’ and ‘DeConv’ represent the convolution filter and deconvolution filter, respectively. ‘IN’ and ‘Dropout’ mean the operations of Instance Normalization and Dropout, respectively. The alphanumeric characters combined by ‘Conv’, ‘DeConv’, ‘n’, ‘k’, ‘s’ and numbers are used to illustrate the filters of convolution

![Fig. 1 Framework of the proposed method.](image-url)
and deconvolution, where ‘Conv’ means the convolution filter, ‘DeConv’ denotes the deconvolution filter, the numbers followed by the letter ‘n’, ‘k’, ‘s’ count the channel number, the kernel size, and the stride, respectively. The receive field of the discriminator is $70 \times 70$.

\[ \text{Fig. 2} \ (a) \ \text{Generator (U-Net)}, \ (b) \ \text{Discriminator (Patch-GAN)} \]

### 2.2. Loss Function

The training of the generators of $G_1$ and $G_2$ and the discriminators of $D_1$ and $D_2$ in the image translation network are adversarial. The loss of the generators $G_1$ and $G_2$ can be defined as

\[
\begin{align*}
L(G_1, G_2) &= E_{x \sim P_{\text{data}}(x)}[\log D_1(x, G_1(x)) + E_{y \sim P_{\text{data}}(y)}[\log D_2(y, G_2(y))] + \lambda \cdot L_{\text{cyclic}}(G_1, G_2) \\
&= E_{x \sim P_{\text{data}}(x)}[\log D_1(x, y)] + E_{y \sim P_{\text{data}}(y)}[\log D_2(x, y)] + \lambda \cdot L_{\text{cyclic}}(G_1, G_2)
\end{align*}
\]

(1)

where $x$ and $y$ represent the input images of HF-SAR and LF-SAR, $P_{\text{data}}(x)$ and $Q_{\text{data}}(y)$ are the underlying data distributions. $G_1(x)$ and $G_2(y)$ represent the translated images of generated LF-SAR and generated HF-SAR, respectively. The losses of the discriminators $D_1$ and $D_2$ can be defined as

\[
\begin{align*}
L(D_1) &= E_{x \sim P_{\text{data}}(x), y \sim P_{\text{data}}(y)}[\log D_1(x, y)] + E_{x \sim P_{\text{data}}(x), y \sim P_{\text{data}}(y)}[\log D_1(x, y)] + \beta_1 \cdot L_{\text{SSIM}}(y, G_1(x)) \\
&= E_{x \sim P_{\text{data}}(x), y \sim P_{\text{data}}(y)}[\log D_1(x, y)] + E_{x \sim P_{\text{data}}(x), y \sim P_{\text{data}}(y)}[\log D_1(x, y)] + \beta_1 \cdot L_{\text{SSIM}}(x, y) \\
L(D_2) &= E_{x \sim P_{\text{data}}(x), y \sim P_{\text{data}}(y)}[\log D_2(x, y)] + E_{x \sim P_{\text{data}}(x), y \sim P_{\text{data}}(y)}[\log D_2(x, y)] + \beta_2 \cdot L_{\text{SSIM}}(x, y) \\
&= E_{x \sim P_{\text{data}}(x), y \sim P_{\text{data}}(y)}[\log D_2(x, y)] + E_{x \sim P_{\text{data}}(x), y \sim P_{\text{data}}(y)}[\log D_2(x, y)] + \beta_2 \cdot L_{\text{SSIM}}(x, y)
\end{align*}
\]

(2)

(3)

where $L(D_1)$ and $L(D_2)$ are parts of cGAN loss 1 and cGAN loss 2 in the two image translation branches respectively. They are augmented by the structural similarity (SSIM) loss [10]. $\| \|$ denotes L1 distance. $L_{\text{SSIM}}(P_1, P_2)$ is the SSIM loss between the paired images of $P_1$ and $P_2$.

\[
L_{\text{SSIM}}(P_1, P_2) = \frac{1}{N} \sum_{p \in P_1} 1 - \frac{2m_m^2 + C_1}{m_m^2 + C_1} \cdot \frac{2m_n^2 + C_2}{m_n^2 + C_2}
\]

(4)

where $m(p)$ and $n(p)$ are the pixels located at $p$ in the patch $P_1$ and $P_2$; $m_m$, $m_n$, $\sigma_m$, $\sigma_n$, and $\sigma_{mn}$ denote the mean, standard deviation, and covariance of $P_1$ and $P_2$ respectively; $N$ is the total number of the pixels in the image $P_1$ or $P_2$; The image pairs $P_1 - P_2$ can be regarded as $y - G_1(x)$ or $x - G_2(y)$. $L_{\text{cyclic}}(G_1, G_2)$ is the cyclic consistent loss in the whole network

\[
L_{\text{cyclic}}(G_1, G_2) = E_{x \sim P_{\text{data}}(x)}[\| x - G_2(G_1(x)) \|_1] + E_{y \sim P_{\text{data}}(y)}[\| y - G_1(G_2(y)) \|_1]
\]

(5)
where the two parts represent cyclic loss 1 and cyclic loss 2 respectively. $\lambda$ controls the weight of the cyclic loss in function (1). The purpose of the training is to solve:

$$G_1^*, G_2^* = \arg \min_{G_1, G_2} \max_{D_1, D_2} [L(G_1, G_2), L(D_1, D_2)]$$

(6)

3. Experiments and results

3.1. Datasets and setting
Experiments on P-band and Ku-band SAR images are conducted to evaluate the translation performance. The P-band and Ku-band SAR data come from an experimental airborne bi-frequency SAR system, where the images are obtained simultaneously. The co-registered images are cut into 128x128 patches for training the network. The datasets contain 1760 pairs of images and features four scenes of farmland, road, river, and forest with each scene containing 400 pairs for training, 20 pairs for testing, and 20 pairs for validating.

We compare the proposed method with Cycle GAN and Pix2Pix. Four indexes of SSIM, peak signal to noise ratio (PSNR), histogram-based Kullback-Leibler divergence (KLD), and histogram-based mean squared error (MSE) are employed to measure the similarity between paired images, for instance, the pairwise real P-band SAR image (P) and translated P-band SAR image (T-P) are evaluated in this study. The higher the PSNR or SSIM, or the less the KLD or the MSE, the better the translation.

In our method, the model is trained on 200 epochs with batch size 1 and learning rate 0.0002, and the dropout rate is set as 0.5. $\lambda$ in the objective function (1) is set to 10, and both of $\beta_1$ and $\beta_2$ in (2) and (3) are set to 10. Adam optimizer is used to update the generators and the discriminators. In the calculation of the histogram-based KLD and MSE, the number of the bin is set as 64.

3.2. Results and analysis
In this paper, four representative pairs of different scenes are selected to show how the methods perform. Correspondingly, the real P-band and Ku-band SAR images are shown in Fig. 3 (a), and the translated images of Pix2Pix, Cycle GAN, and our method are shown in Fig. 3 (b), (c), and (d), respectively. Comparisons between Fig. 3 (a) and (b) show that the translated images produced by the Pix2Pix are blurred. Comparisons between Fig. 3 (a) and (c) show that the translated images produced by the Cycle GAN lose many structural details. By contrast, the translated images produced by the proposed method retain more structural details.

The four metrics of PSNR, SSIM, KLD, and MSE of the translated images and paired real images are shown in Table 1. Compared with Pix2Pix and Cycle GAN, our method performs better in the two indicators of PSNR and SSIM, especially the SSIM improved remarkably over 0.5262. All the three methods can comparatively diminish the distribution difference between the real Ku-band SAR image and the real P-band SAR image based on KLD and MSE.

Our method draws on the advantages of Pix2Pix and Cycle GAN and avoids the shortcomings in visual inspection and quantitative analysis because the Pix2Pix is limited by one-to-one mapping [12], the Cycle GAN is based on unpaired training datasets [13].
Fig. 3 (a) Real P-band and Ku-band SAR images,  
(b) Translated P-band and Ku-band SAR images of the Pix2Pix,  
(c) Translated P-band and Ku-band SAR images of the CycleGAN,  
(d) Translated P-band and Ku-band SAR images of the proposed method

|                | P and Ku | Pix2Pix | CycleGAN | Proposed |
|----------------|----------|---------|----------|----------|
|                | P and T-P| Ku and T-Ku | P and T-P| Ku and T-Ku | P and T-P| Ku and T-Ku |
| PSNR           | 17.2179  | 18.8730 | 15.8704  | 21.5135  | 18.4763 | 21.5377 | 25.7575  |
| SSIM           | 0.1776   | 0.2548  | 0.0763   | 0.3269   | 0.2243  | 0.3331  | 0.5262   |
| KLD            | 1.1080   | 0.9289  | 0.5786   | 1.9892   | 0.8465  | 0.4882  | 0.9037   |
| MSE            | 6.7295×10\(^{-3}\) | 4.4710×10\(^{-4}\) | 4.7834×10\(^{-4}\) | 0.0016   | 7.7479×10\(^{-4}\) | 4.1413×10\(^{-4}\) | 5.2992×10\(^{-4}\) |

Tab. 1 Quantitative evaluation of translated images with different methods

4. Conclusions
To achieve the translation between the high- and low-frequency SAR images, this paper presents a cycle-consistent conditional adversarial network. The proposed network takes advantages of Pix2Pix and Cycle GAN, where the cycle-consistent architecture of the Cycle GAN is employed as the major structure of our network, and the Patch-GAN and the U-Net in the Pix2Pix are used as discriminator and generator respectively. Experiments on Ku- and P-band SAR images show that our method outperforms Cycle GAN and Pix2Pix with higher PSNR and SSIM and comparative KLD and MSE.

References
[1] Parikh H, Patel S and Patel V 2020 J. Classification of SAR and PolSAR images using deep learning: a review 11 pp1-32
[2] Chen S and Wang H 2014 SAR target recognition based on deep learning 2014 International Conference on Data Science and Advanced Analytics (DSAA) pp 541-547
[3] Wang L, Scott K A and Xu L 2014 J. Ice concentration estimation from dual-polarized SAR images using deep convolutional neural networks IEEE (Transactions on Geoscience and Remote Sensing)
[4] Wang P, Zhang H and Patel V M 2017 SAR Image Despeckling Using a Convolutional Neural Network IEEE Signal Processing Letters 24 pp1763-1767
[5] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 J. Generative adversarial nets Advances in neural information processing systems pp 2672-2680
[6] Grohfeldt C, Schmitt M and Zhu X 2018 A conditional generative adversarial network to fuse sar and multispectral optical data for cloud removal from sentinel-2 images IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium pp1726-1729
[7] He W and Yokoya N 2018 Multi-temporal sentinel-1 and-2 data fusion for optical image simulation ISPRS International Journal of Geo-Information 7 p389
[8] Niu X, Gong M, Zhan T and Yang Y 2018 A conditional adversarial network for change detection in heterogeneous images IEEE Geoscience and Remote Sensing Letters 16 pp45-49

[9] Ao D, Dumitru C O, Schwarz G and Datcu M 2018 J. Dialectical GAN for SAR image translation: From Sentinel-1 to TerraSAR-X Remote Sensing 10 p1597

[10] Marmanis D, Yao W, Adam F, Datcu M, Reinartz P, Schindler K, Wegner J D and Stilla U 2017 J. Artificial generation of big data for improving image classification: A generative adversarial network approach on SAR data arXiv preprint arXiv:1711.02010

[11] Fu S, Xu F and Jin Y Q 2019 J. Reciprocal translation between SAR and optical remote sensing images with cascaded-residual adversarial networks arXiv preprint arXiv:1901.08236

[12] Isola P, Zhu J Y, Zhou T and Efros A A 2017 Image-to-image translation with conditional adversarial networks Proceedings of the IEEE international conference on computer vision pp1125-1134

[13] Zhu J Y, Park T, Isola P and Efros A A 2017 Unpaired image-to-image translation using cycle-consistent adversarial networks Proceedings of the IEEE international conference on computer vision pp 2223-2232