remote sensing

MDPI

Article

A Dual-Generator Translation Network Fusing Texture and Structure Features for SAR and Optical Image Matching

Han Nie [1], Zhitao Fu [1,*], Bo-Hui Tang [1,2], Ziqian Li [1], Sijing Chen [1] and Leiguang Wang [3]

[1] Faculty of Land and Resources Engineering, Kunming University of Science and Technology, Kunming 650031, China; nie_han@stu.kust.edu.cn (H.N.); tangbh@kust.edu.cn (B.-H.T.); liziqian1012@stu.kust.edu.cn (Z.L.); chensijing@stu.kust.edu.cn (S.C.)
[2] State Key Laboratory of Resources and Environment Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[3] Institute of Big Data and Artificial Intelligence, Southwest Forestry University, Kunming 650024, China; leiguangwang@swfu.edu.cn
* Correspondence: zhitaofu@kust.edu.cn

Abstract: The matching problem for heterologous remote sensing images can be simplified to the matching problem for pseudo homologous remote sensing images via image translation to improve the matching performance. Among such applications, the translation of synthetic aperture radar (SAR) and optical images is the current focus of research. However, the existing methods for SAR-to-optical translation have two main drawbacks. First, single generators usually sacrifice either structure or texture features to balance the model performance and complexity, which often results in textural or structural distortion; second, due to large nonlinear radiation distortions (NRDs) in SAR images, there are still visual differences between the pseudo-optical images generated by current generative adversarial networks (GANs) and real optical images. Therefore, we propose a dual-generator translation network for fusing structure and texture features. On the one hand, the proposed network has dual generators, a texture generator, and a structure generator, with good cross-coupling to obtain high-accuracy structure and texture features; on the other hand, frequency-domain and spatial-domain loss functions are introduced to reduce the differences between pseudo-optical images and real optical images. Extensive quantitative and qualitative experiments show that our method achieves state-of-the-art performance on publicly available optical and SAR datasets. Our method improves the peak signal-to-noise ratio (PSNR) by 21.0%, the chromatic feature similarity (FSIMc) by 6.9%, and the structural similarity (SSIM) by 161.7% in terms of the average metric values on all test images compared with the next best results. In addition, we present a before-and-after translation comparison experiment to show that our method improves the average keypoint repeatability by approximately 111.7% and the matching accuracy by approximately 5.25%.

Keywords: SAR-to-optical image translation; dual-generator; texture and structure fusing; SAR and optical image matching

Citation: Nie, H.; Fu, Z.; Tang, B.-H.; Li, Z.; Chen, S.; Wang, L. A Dual-Generator Translation Network Fusing Texture and Structure Features for SAR and Optical Image Matching. *Remote Sens.* **2022**, *14*, 2946. https://doi.org/10.3390/rs14122946

Academic Editors: Olga Sykioti, Gangyao Kuang and Xin Su

Received: 10 May 2022
Accepted: 17 June 2022
Published: 20 June 2022

1. Introduction

Different sensors can capture different features. In particular, synthetic aperture radar (SAR) images and optical images are widely used in map production [1]. Optical images conform to human vision, but are susceptible to objective factors such as cloud interference [2], whereas SAR images are immune to the imaging defects of optical images and have the advantages of all-weather acquisition, a long line of sight, and some level of penetration capability. Therefore, the fusion of optical and SAR images is widely used in pattern recognition [3], change detection [4], and landslide recognition [5]. However, a prerequisite for the fusion of SAR and optical images is high-accuracy matching. In recent decades, many feature matching methods for homologous images have been proposed, e.g., SIFT [6], SURF [7], ORB [8], and LoFTR [9]. The LoFTR method mainly focuses on dense

matching of weakly textured regions of homologous images. However, these methods are applicable to homologous image matching, but not to SAR and optical image matching because NRDs are not considered. Recently, to address severe NRDs between SAR and optical images, Cui et al. [10] implemented MAP-Net by introducing spatial pyramid aggregated pooling (SPAP) and an attention mechanism to improve the matching precision of optical and SAR images. Li et al. [11] proposed the radiation-variation insensitive feature transform (RIFT) for different types of images. Cui et al. [12] extended scale invariance based on RIFT, but their method was more sensitive to noise, and Li et al. [13] proposed the locally normalized image feature transform (LNIFT) using a local normalization filter to convert images of different modalities into the same intermediate modality, turning the multimodal image matching problem into a homogenous matching problem, and making different modalities similar to improve the matching accuracy. In recent years, deep learning has been successfully introduced into the field of remote sensing image processing for applications such as image matching [14], image fusion [15], and image translation [16]. It is noteworthy that generative adversarial networks (GANs) can better convert multimodal image matching problems into homologous matching problems. Many researchers have implemented matching between SAR and optical images based on SAR-to-optical translation. Quan, D. [17] proposed a generative matching network (GMN) to generate a corresponding simulated optical image for a real SAR image or a pseudo-SAR image for a single optical image, and then input these matched pairs into a matching network to infer whether they matched, achieving improved performance in SAR–optical image matching. Merkle et al. [18] jointly implemented the translation of single-polarization SAR images into optical images by means of a conditional generative adversarial network (CGAN) and verified the possibility of using the transformed pseudo-optical images for image matching. A k-means clustering-guided generative adversarial network (KCG-GAN) [19] has also been proposed for use in SAR and optical image matching, and the results showed that the quality of SAR-to-optical translation limits the matching accuracy between SAR and optical images. Therefore, the key question that needs to be urgently answered is how to design a high-precision SAR-to-optical translation method to enhance the SAR–optical matching performance.

In recent decades, many researchers have proposed methods, which are mainly based on image enhancement algorithms and pseudo-colour encoding algorithms, for SAR–optical translation. In the field of image enhancement, a wavelet transform-based method was used for SAR image denoising to achieve SAR image enhancement from the perspective of noise suppression, but it was found that there was a possibility of increasing the amount of other types of clutter [20]. By introducing visualization algorithms to map high-dynamic-range SAR amplitude values to low-dynamic-range displays via reflectivity distortion, entropy maximization can be preserved to improve the visual quality of SAR images by maximizing the display information [21], and an adaptive two-scale enhancement method can be used to visualize all greyscale information and enhance local target peaks [22]. However, the previous approaches enhance SAR images by means of visualization methods that cannot effectively resolve differences caused by nonlinear radiation distortion (NRDs). In the field of pseudo-colour coding, the pixels of SAR images are mainly encoded to make them as similar as possible to those of optical images [23–25]; however, a greyscale image is obtained instead of a three-channel image, and because the results are highly dependent on the specifics of the model, the performance may decline in practical use. The images processed by image enhancement algorithms and pseudo colour encoding algorithms are enhanced in terms of visual features, but both types of algorithms ignore the NRDs differences between SAR images and optical images; consequently, large differences in structure and texture remain in the resulting pseudo-optical images compared to the real optical images. For the task of automatic image colorization, a deep learning model can be used to predict the pixel-by-pixel colour histogram suitable for the colouring task without structurally transformed image pairs [26]. In the field of SAR image processing, a convolutional neural network (CNN)-based approach has been used to convert

a single-polarization greyscale SAR image into a full-polarization image [27]. Moreover, generative adversarial networks (GANS) [28] are widely used for image translation. A dialectical GAN using conditional Wasserstein generative adversarial network–gradient penalty (WGAN-GP) loss functions has been applied to translate Sentinel-1 images into TerraSAR-X images [29]. Based on the proposal of a boundary equilibrium generative adversarial network (BEGAN) [30], an adversarial network was designed for SAR image generation, and it was demonstrated that the proposed method could improve the classification accuracy [31]. Many GAN-based methods have also been used in SAR-to-optical transformation, such as Pix2pix [32], CycleGAN [33], S-CycleGAN [34], and EPCGAN [16]. Pix2pix and CycleGAN can both be used for SAR–optical translation, but they have certain drawbacks. With Pix2pix, the structure is vague, and some objects have missing structural information, whereas CycleGAN can retain structural information, but ignores land cover information; accordingly, S-CycleGAN combines the advantages of CycleGAN, preserving both land cover information and structural information. He, W. [35] proposed a model combining residual networks and CGANs that can simulate optical images from multitemporal SAR images. However, there is a major problem with such methods; they usually rely on a network structure designed for optical image transformation, with only simple modifications, which is not applicable for SAR–optical translation because of the differences between the imaging principles of SAR images and optical images. Based on this understanding, a feature-guided method based on a discrete cosine transform (DCT) loss has been proposed [36], and edge information has been used to guide SAR–optical translation to obtain pseudo-optical images with better edge information [37]. Similarly, EPCGAN considers the edge blurring problem for pseudo-optical images, and uses gradient information to preserve the edge information in generated pseudo-optical images. The pseudo-optical images obtained in this way contain better structural information, but a situation may arise in which structure and texture features cannot be effectively fed back, resulting in poor and unrealistic imaging effects due to the inability to achieve deep fusion of the structure and texture features. Inspired by [38], in which more natural image inpainting results were obtained by means of a two-branch network, we also treat SAR-to-optical image translation as consisting of two complementary subtasks, namely, texture translation and structure translation, considering the NRDs of SAR images. We reduce the gap between pseudo-optical and real optical images by introducing a spatial-domain loss function and frequency-domain loss function, and thus, obtain pseudo-optical translation results with high accuracy (see Figure 1).

Figure 1. *Cont.*

Figure 1. High-quality image translation results obtained with our method. The pseudo-optical image is the image generated from the SAR image through our method.

In this paper, we propose a dual-generator translation network that fuses texture and structure features to obtain enhanced pseudo-optical images for SAR–optical matching. The proposed network consists of dual generators, bidirectional gated feature fusion (Bi-GFF) [38], and contextual feature aggregation (CFA) [39] modules and discriminators. First, the input SAR image is decomposed into structure and texture features based on the Canny edge detection algorithm [40]. Then, the structure features and greyscale map are input into the structure encoder, the SAR image is input into the texture encoder, and the feature maps of different dimensions from the texture encoder and structure encoder are stitched together to join the structure and texture decoder to obtain texture features and structure features, which are then fused and refined by the Bi-GFF module and CFA module. Finally, a frequency-domain loss function (focal frequency loss [41]) and a spatial-domain loss function (mean square error) are introduced to reduce the gap between the pseudo-optical and real optical images during the learning process. We present comparative experiments and ablation experiments conducted on the same dataset. The experimental results show that the proposed method yields images with clearer textures and structures that are used to achieve better evaluation results that exhibit better visual properties than the results of Pix2pix [32], CycleGAN [33], S-CycleGAN [34], and EPCGAN [16].

Specifically, the major contributions of this paper are as follows:

1. We propose a dual-generator translation network that fuses texture and structure features to improve the matching of SAR images with optical images. The proposed network includes both structure and texture generators, and the structure and texture features are coupled with each other by these dual generators to obtain high-quality pseudo-optical images.
2. We introduce spatial-domain and frequency-domain loss functions to reduce the gap between pseudo-optical images and real optical images, and present ablation experiments to prove the superiority of our approach.
3. To demonstrate the superiority of the proposed algorithm, we select training and test data from public datasets, and we present keypoint detection and matching experiments for comparisons between pseudo-optical images and real optical images and between real optical images and SAR images before and after translation.

The remainder of this paper is organized as follows. The proposed dual-generator translation network fusing texture and structure features for SAR–optical image translation is introduced in Section 2. We present the experimental results and matching applications in Section 3. A discussion is provided in Section 4. Finally, the conclusions are summarized in Section 5.

2. Methods

In this section, we introduce the proposed dual-generator translation network fusing texture and structure features for SAR–optical image matching. As illustrated in Figure 2, the dual generators provide feedback to each other to obtain the structure and texture features, which are fused by the Bi-GFF and CFA modules. In the following subsections, we present the details of the generators, discriminator, and loss functions.

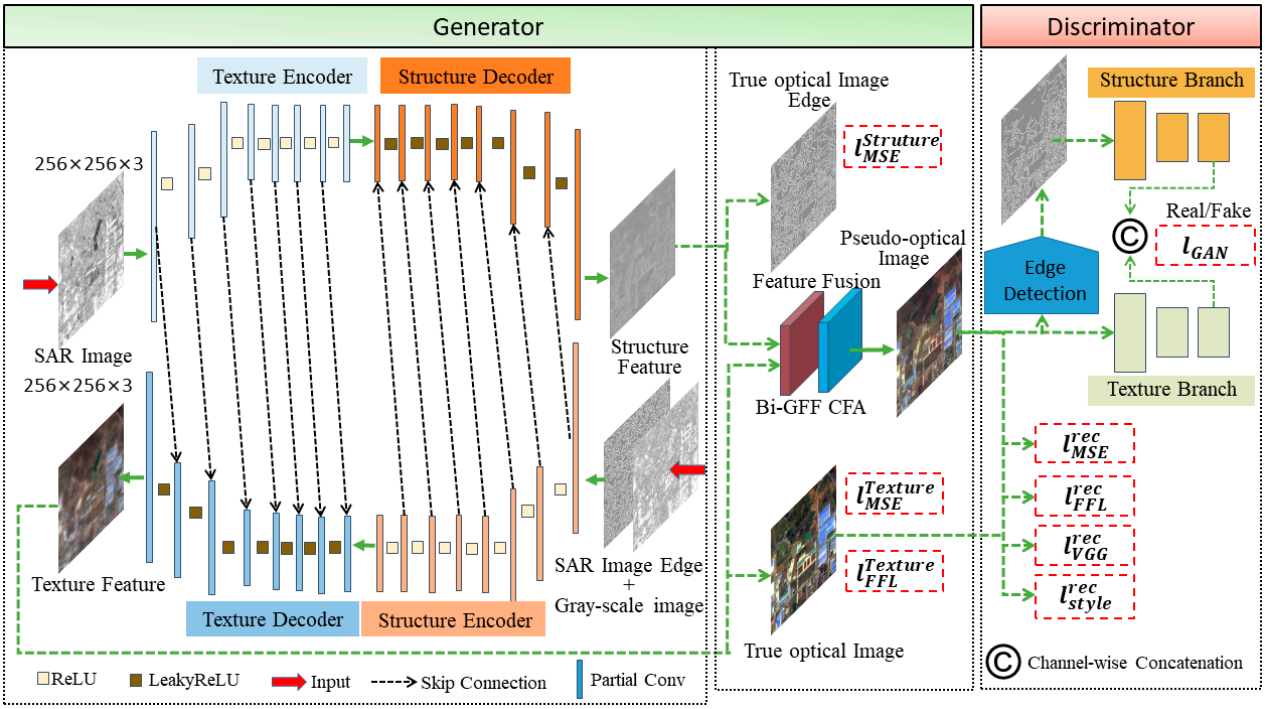

Figure 2. The generators and discriminator of our network. **Generators**: The SAR-to-optical translation process is divided between two generators, i.e., a structure generator and a texture generator, which borrow each other's depth features, and the Bi-GFF and CFA modules are used to refine and fuse the features from these structure and texture reconstruction branches to form the final pseudo-optical image. **Discriminator**: The texture branch guides texture generation, and the structure branch guides structure generation.

2.1. Generators

As shown in Figure 2, the generator part of the SAR-to-optical translation network is divided into two generators, namely, a structure generator and a texture generator, which are based on U-Net variants [42], where final features from the structure encoder and multilevel features from the texture encoder are added to the texture decoder via a skip connection, and final features from the texture encoder and multilevel features from the structure encoder are added to the structure decoder via a skip connection. We also show the structural details of the texture and structure generators in Table 1. In the encoder stage, the SAR image to be translated is passed to the texture encoder, and the greyscale image and edge structure image of the SAR image to be translated are passed to the structure encoder. In the decoder stage, the structure features from the structure encoder are used as constraints in the texture decoder, and the texture features from the texture encoder are used as constraints in the structure decoder. This coupled dual-generator structure ensures good complementarity between the structure and texture features. Compared with normal convolutional layers, partial convolutional layers can better capture the information of irregular boundaries [42]; accordingly, considering the severe scattering noise, NRD, and large irradiance differences between optical and SAR images, we also use partial convolutional layers instead of normal convolutional layers. In addition, we add skip