remote sensing

MDPI

Technical Note

High-Resolution SAR-to-Multispectral Image Translation Based on S2MS-GAN

Yang Liu [1,*,†], Qingcen Han [2,†], Hong Yang [2,3] and Huizhu Hu [1]

[1] College of Optical Science and Engineering, Zhejiang University, Hangzhou 310058, China; huhuizhu2000@zju.edu.cn
[2] School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China; hanqingcen@mail.nwpu.edu.cn (Q.H.); yanghong@aircas.ac.cn (H.Y.)
[3] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China
* Correspondence: yang047@zju.edu.cn
† These authors contributed equally to this work.

Abstract: Synthetic aperture radar (SAR) has been extensively applied in remote sensing applications. Nevertheless, it is a challenge to process and interpret SAR images. The key to interpreting SAR images lies in transforming them into other forms of remote sensing images to extract valuable hidden remote sensing information. Currently, the conversion of SAR images to optical images produces low-quality results and incomplete spectral information. To address these problems, an end-to-end network model, S2MS-GAN, is proposed for converting SAR images into multispectral images. In this process, to tackle the issues of noise and image generation quality, a TV-BM3D module is introduced into the generator model. Through TV regularization, block-matching, and 3D filtering, these two modules can preserve the edges and reduce the speckle noise in SAR images. In addition, spectral attention is added to improve the spectral features of the generated MS images. Furthermore, we construct a very high-resolution SAR-to-MS image dataset, S2MS-HR, with a spatial resolution of 0.3 m, which is currently the most comprehensive dataset available for high-resolution SAR-to-MS image interpretation. Finally, a series of experiments are conducted on the relevant dataset. Both quantitative and qualitative evaluations demonstrate that our method outperforms several state-of-the-art models in translation performance. The solution effectively facilitates high-quality transitions of SAR images across different types.

Keywords: SAR image; SAR interpretation; multispectral image; high-quality image translation; remote sensing dataset

check for updates

Citation: Liu, Y.; Han, Q.; Yang, H.; Hu, H. High-Resolution SAR-to-Multispectral Image Translation Based on S2MS-GAN. *Remote Sens.* **2024**, *16*, 4045. https://doi.org/10.3390/rs16214045

Academic Editors: Mengdao Xing, Jianlai Chen, Yu Li and Yongkang Li

Received: 14 September 2024
Revised: 18 October 2024
Accepted: 24 October 2024
Published: 30 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the last decades, remote sensing technology for detection and imaging has developed rapidly. Among these different kinds of imaging means, multispectral (MS) imaging and SAR imaging are extensively used. Images in multiple spectral bands are captured simultaneously to obtain rich feature information. These bands typically include visible and invisible light (e.g., near-infrared), which means MS images have more spectral channels than normal optical RS images. Also, MS images have comparatively high spectral resolution, and they are easier to interpret than hyperspectral images. MS images are mostly used in forest prediction [1], agricultural monitoring [2], disaster response [3], land cover monitoring [4], military [5], etc. But multispectral devices cannot be used at night and are limited by adverse weather conditions such as haze or clouds, which may lead to missing information in these particular cases. For SAR, radars emit penetrating microwaves, and the sensors receive energy reflected to the image. Because of the imaging methodology of SAR images, they are easily prone to noise issues and cannot be directly interpreted. SAR also captures more variations, details, and higher-frequency noise [1–6].

SAR can be used under all-day and all-weather conditions alongside its disadvantages. In some poor atmospheric conditions, only SAR images can be obtained; converting them to MS images can compensate for the shortcomings of MS and SAR images. So recently, some researchers are willing to complete SAR-to-MS image translation (S2MST) in order to avoid their drawbacks [7,8]. S2MST through manual labeling by humans requires substantial manpower and time investment. At the same time, image-to-image translation (I2I) by artificial intelligence [9] is continuously growing with the development of machine learning (ML) and deep learning (DL) [10–13]. Most of the I2I studies are based on a generative adversarial network (GAN) [14] and its variants [15–17]. pix2pix [14] is based on conditional generative adversarial networks [17], which is the first method to achieve widely recognized good results in I2IT. And then a cycle-consistent adversarial network (CycleGAN) [16] was raised for the well-known unsupervised I2IT. Nevertheless, those methods are designed to solve natural images' translation problems. So presently, the majority of S2OT works are using CycleGAN as their backbone network for image translation [18–20].

The fundamental concept behind the researchers is to use paired data as prior knowledge, enabling the computer to learn the corresponding feature distributions through deep learning techniques. We aim to translate very high-resolution SAR images into MS images. However, due to the lack of high-quality paired data and the model for SAR-to-MS image translation, most of the SAR images' translation studies focus on SAR-to-optical translation (S2OT) [21–27].

Currently, the dataset of MS images and SAR images available for training and testing purposes is primarily derived from open-access sources, such as Sentinel-2 [28], Landsat-8 [29], Quickbird [30], WorldView-3 and Sentinel-1 [31], ALOS-2 [32], TerraSAR-X [33], etc. But the resolution of these open-access remote sensing images is relatively low; most of them are above 1m or even lower. While these data are directly accessible, the relatively low resolution of these images (especially the MS images) is not paired and is already preprocessed for immediate use. The lack of high-quality data is a significant impediment to conducting S2MST. Simultaneously, the existing methods of S2OT are not fully suitable for S2MST because they do not consider the differences in multispectral channels. Additionally, the current S2OT methodologies fail to resolve the emergent noise issue of very high-resolution SAR images, which may cause a reduction in image generation quality and the emergence of artifacts. Therefore, transitioning from S2OT to S2MST requires both data enhancements and model modifications. SAR images and optical images exhibit distinct display characteristics due to their respective sensor technologies and imaging principles [34], making them examples of heterogeneous remote sensing data [35]. Theoretically, there is no inherent correspondence between these two types of images. However, the SAR and multispectral images used in this study were captured over the same geographical region, allowing them to share consistent viewing angles and coverage areas. By applying preprocessing techniques such as polarimetric pseudo-color generation and image registration [36], we ensure that the resulting images depict the same ground features, albeit presented through different visual modalities of the same observed environment. The SAR images employed in this study are pseudo-color RGB images generated via polarization processing, while the multispectral images are spectral images derived from ortho-rectification and color adjustment. This shared foundation enables the use of rich prior knowledge to explore the task of SAR-to-MS image translation [37].

In order to translate SAR images to MS images, we set a dataset containing paired SAR and MS images with very high resolution [38]. And we propose an end-to-end SAR-to-multispectral generative adversarial network (S2MS-GAN). To solve the issues mentioned above, we only kept the basic framework of CycleGAN, and we redesigned the structure of the generator and the discriminator. Given the ultra-high resolution of our SAR images, the associated high frequency inevitably introduces increased noise levels, particularly speckle noise, which can adversely impact the quality of the generated images. So, in the generator part, we designed a module named TV-BM3D, which leverages BM3D (block-

batching and 3D filtering) [39] and Total Variation Regularization in turns to remove speckle noise from the SAR images. So that TV-BM3D can extract critical features and enhance the performance of SAR-to-MS image generation. For the discriminator, we aim to direct the discriminator's attention to focus on spectral channel information by incorporating the classical attention module. Our proposed framework is capable of performing both S2MST and multispectral-to-SAR translation (MS2ST), with a primary focus on S2MST, which is shown in Figure 1.

In summary, our key contribution can be summarized as follows:

1. A challenging task of high-resolution image translation from SAR-to-multispectral images is proposed. In contrast to current S2OT solutions, the S2MS-GAN in this paper is able to provide more remote sensing spectral information during the SAR image translation process.
2. An end-to-end network model is designed, and a TV-BM3D module is introduced into the generator. The speckle noise in high-resolution SAR images can be effectively reduced by TV regularization and the BM3D module. Meanwhile, spectral attention is added to improve the spectral features of the generated multispectral images. In a series of evaluation experiments, the images generated by this method have higher accuracy, and the visualization results are better.
3. A very high-resolution SAR-MS image dataset named S2MS-HR is constructed by performing paired preprocessing on each image with a spatial resolution of 0.3 m. S2MS-HR can provide the model with better learning generalization ability and offer a more extensive data basis for the interpretation and generation of SAR images.

The rest-outline is structured as follows: In Section 2, we provide a comprehensive review of the related work. Section 3 describes our model and the method utilized. In Section 4, we emphasize the data processing techniques and results of the evaluation across different baseline models. Section 5 offers the conclusion.

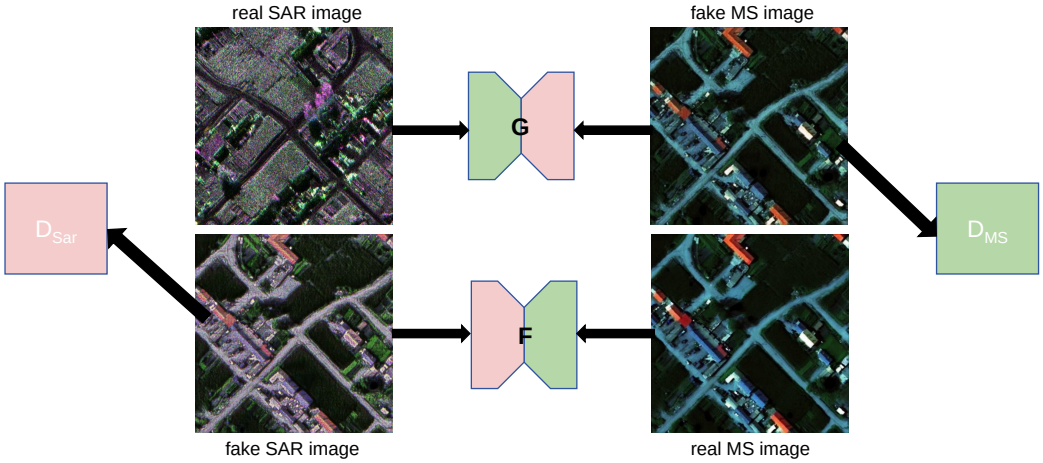


Figure 1. The S2MST and MS2ST cycle–generate framework. We can conduct bi-directional generation of SAR and MS images (the main focus of this work is S2MST).

2. Related Work

2.1. Generative Adversarial Network

In recent years, the field of deep learning has experienced rapid advancements, leading to a significant expansion in its applications and notable achievements across various domains [40], including generative adversarial networks (GAN) which was first raised in 2014 by Goodfellow et al. [14]. In GAN, the generator is used as a key component to generate realistic images from random noise. Most of the early generators were based on simple fully connected layers and convolution layers, such as the original cGAN [23], which generated clearer low-resolution images by stacking convolutional and transposed convolutional layers. As for the discriminator, it is designed to distinguish the real images

from the generated images from the generator. Most of the early discriminators used simple convolutional neural network (CNN) [41] architectures to achieve the discriminator's task through multiple convolutional and fully connected layers. With the development of GAN, researchers have proposed a variety of improved discriminator architectures to enhance the quality of the generated images and the performance of the discriminators. For example, PatchGAN [15] enhances attention to local details by determining whether the local patches of an image are realistic or not and is widely used in image translation tasks. On the other hand, the ProGAN [42] discriminator adopts a step-by-step growth network structure to increase the resolution layer by layer, making the generated high-resolution images more realistic. Through these various studies on GAN, some researchers apply GAN to image-to-image translation. But existing studies primarily apply methods designed for natural image-to-image translation or develop network structures from the perspective of natural images. These methodologies almost overlook the inherent complexity and unique characteristics of SAR images. Moreover, the generators and discriminators implemented in contemporary SAR-to-optical image translation techniques are inadequate in noise suppression, consequently resulting in poor quality of the generated images.

2.2. Remote Sensing Image Translation

For instance, there are no former studies on SAR-to-MS image translation. Since the MS images and optical images are highly relevant, and both S2MST and S2OT are branches of the image-to-image translation task in remote sensing image translation, our related work primarily focuses on an analysis of SAR-to-optical image translation tasks. For S2OT, in 2018, Merkle et al. conducted experiments on translating SAR and visible light images using conditional GANs (cGANs) [22]. Enomoto et al. also explored similar experiments using cGANs [26] in that same year. Based on the generative adversarial framework, Isola et al. [15] introduced the pix2pix network early, and Toriya et al. were the first to apply pix2pix in S2OT. And Zhu et al. created CycleGAN [16] by using two generators and two discriminators to achieve unsupervised image translation with cycle-consistency loss, which can share the information between images in two different domains. Based on CycleGAN, Guo et al. proposed EP-GAN [18], focusing on maintaining the edge information of the image to improve the quality and detail fidelity of SAR-translated images. In the year 2022, in a similar approach, Wang et al. [43] developed a supervised CycleGAN designed to transform SAR images into large-scale optical images. Also, Zhang et al. [20] in 2023 have proposed a thermodynamic theory-based network (S2O-TDN) in which the interpretative network as well as the use of traditional methods based on CycleGAN for the translation of SAR-to-optical images.

2.3. BM3D, TV, and Attention

Block-matching and 3D filtering (BM3D) was first proposed by Dabov et al. in 2007 [39]. BM3D cuts the image into different patches based on similar noise intensity and groups the similar patches together to form a 3D matrix. After, BM3D inverse log transformation using the denoised image obtained in the first stage, which can generate the denoised image. In 2015, Subramanyam, M. V., and Prasad, G. first applied BM3D into the field of remote sensing [44]. They discovered that the InSAR image can be denoised by using BM3D and have a great result. Also in 2018, Francescopaolo et al. extend BM3D to a more efficient model by using a cosine similarity based on phase difference. In 2019, Devapal et al. further study the denoising algorithm on the SAR image [45], in which they use the Curvelet transform to extract the transform coefficients. The Curvelet better captures the curved edge features in the image. In 2019, Lavreniuk et al. [27] introduced a new BM3D method for denoising with CNNs, which contains the deep learning method and the traditional method. In 2022, Malik et al. designed a new BM3D method that can detect edge standards based on the Curvelet transform and the Canny edge detection operator [46]. With the edge detection, using BM3D can both denoise the image and keep the most information of the original image.

Total Variation Regularization was first raised in 1992 by Rudin et al. to solve the image denoising problem [47]. TV regularization suppresses high-frequency noise in images by minimizing the sum of the image gradients. Unnecessary noise is eliminated while maintaining image structure. This contrasts with other common regularization methods (e.g., L2 regularization), where TV regularization does not tend to over-smooth the image, allowing significant jumps at the edges of the image to be preserved. Later TV regularization has shown remarkable results in applications such as image denoising [48], image deblurring [49], and super-resolution reconstruction [50]. With the technological development of deep learning, TV regularization was applied to deep learning frameworks [51]. By introducing more prior information, the performance of TV regularization in complex scenarios is further improved.

For attention, Vaswani, A. set attention to a new stage in 2017, and a lot of attention works are based on his study [52]. In 2019, SE-Net was raised by Hu et al. and had a great success through the different weights on images' channels [53]. Also, Hu designed a network named GAE-Net [54]. Through experiments, the authors show that the GE framework can significantly improve the performance of architectures such as ResNet. Later in 2020, Qilong Wang et al. use ECA-Net to extract the feature and achieve a balance between computational efficiency and performance [55].

3. Methods

In this section, we proposed S2MS-GAN with the framework of CycleGAN, which exhibits robust image translation capabilities, significantly enhancing the quality of translated multispectral images and generating more realistic multispectral imagery. Additionally, S2MS-GAN effectively elevates the fidelity of the translated multispectral images, producing outputs that closely align with real multispectral data. S2MS-GAN consists of two integral components: a generator designed to eliminate speckle noise of SAR images and a discriminator engineered to enhance spectral discrimination within multispectral images as shown in Figure 2.

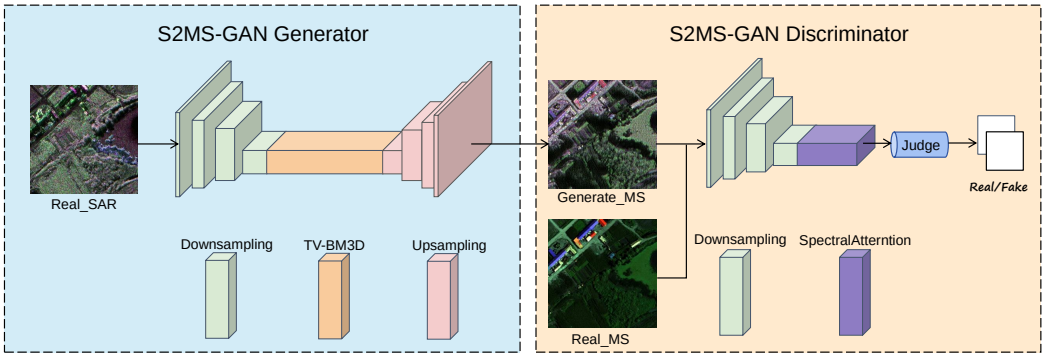


Figure 2. The overview framework of S2MS-GAN. The architecture consists of two primary components: the generator and the discriminator. The generator takes real SAR images as input, applies downsampling followed by noise removal and edge enhancement using a TV-BM3D module, and finally upscales the output to produce the generated MS images. These generated MS images, alongside the real MS images, are fed into the discriminator. In the discriminator, the input undergoes downsampling and spectral attention before passing through the final classification stage to determine authenticity.

3.1. S2MS-GAN Generator

The generator is built on the CycleGAN, and the S2MS generator contains a resnet-based generator that can extract images' features through the skip connection structure. Although using the skip connection from resnet, we cannot directly denoise from the origin SAR image. The TV-BM3D is aimed at removing the speckle noise, which uses Total

Variation Regularization and adaptive BM3D, which are more suitable for our work. TV regularization operates by minimizing the image gradient during the denoising process.

$$\min_u \left\{ \frac{1}{2} \|u - u_0\|^2 + \lambda TV(u) \right\} \tag{1}$$

where u_0 is the image with noise. u is the denoised image. λ is the regularization parameter and controls balance. Larger values of λ imply more regularization, leading to more smoothing, while smaller values of λ allow more detail to be retained in the image.

As shown in Figure 3. in the first stage, similar block matching is firstly required, among which we use SSIM to replace the traditional L2 distance to determine block similarity. SSIM not only considers pixel-level similarity but also incorporates luminance, contrast, and structural information, allowing for a more comprehensive capture of the relationships between image blocks. During the second stage, leveraging the denoising output from the preceding stage, a Wiener filter is constructed to further refine the image, enhancing its denoising ability. This approach ensures that residual noise is substantially reduced, improving the overall image quality by addressing localized noise artifacts. The Wiener filter, specifically tailored to the local signal-to-noise ratio (SNR) of each block, plays a critical role in optimizing the denoising process. By adaptively adjusting to the unique noise characteristics of individual blocks, it ensures maximal preservation of image details while effectively minimizing noise, resulting in superior denoising performance.

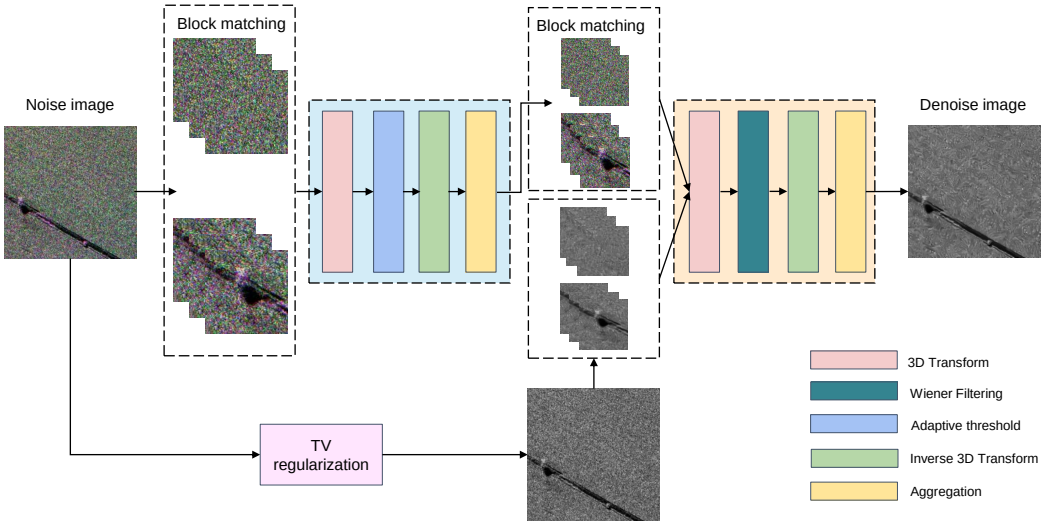


Figure 3. The TV-BM3D framework incorporates several modular components, each represented by a different color in the diagram, with each module fulfilling a specific function. This architecture utilizes two BM3D modules, where the second module processes the output generated by the TVR as its input. Both the output from the TVR and the output to the first BM3D module undergo block matching to ensure optimal performance in noise reduction and detail preservation.

BM3D typically has two stages. We design to feed the denoised outputs from the first stage along with the TV-regularized results into the 3D transform of the second stage, then going Wiener filtering. This approach offers the advantage of pre-smoothing noise while preserving image details through TV regularization and further suppressing noise that the initial BM3D stage might not fully achieve. This method effectively enhances the denoising performance by retaining local details and edges in remote sensing images.

Additionally, an adaptive thresholding mechanism is applied to the transform coefficients of each block group, grounded in the results of local noise estimation. In regions with high noise levels, a larger threshold is employed to remove a greater amount of noise. In regions with lower noise levels or high detail complexity, a smaller threshold is applied to preserve more intricate detail information. After group-matching, former

adaptive BM3D can be performed by setting the corresponding filtering thresholds for different groups, e.g., 0.5 for the heavy noising group and 0.3 for the weak noising one. This content-driven adaptation enables the algorithm to effectively handle diverse image regions, leading to enhanced performance in both smoothness and highly textured areas, thus ensuring consistent denoising across the entire image.

3.2. S2MS-GAN Discriminator

PatchGAN [12] is used in CycleGAN to divide the input image into a number of small patches. PatchGAN can effectively detect texture changes in an image, ensuring that the generated image remains real to the local details while the style is transformed. Therefore, we continue to use PatchGAN as our discriminator's base model. Since generating MS images is our goal, we decide to add an attention module to our discriminator. We choose to add Channel Attention Mechanism (CAM) into our discriminator as our spectral attention module. CAM was raised with Spatial Attention Module (SAM) in and combined as CBAM. We choose to use the CAM part for our task. Initially, the spectral attention module extracts the global information of individual channels via both max-pooling and average-pooling processes. Then a shared multilayer perceptron (MLP) computes the channel-specific weights based on this extracted information. These pooling outputs are then combined, followed by normalization using a Sigmoid activation function. Finally, the resulting attention weights, derived from the Sigmoid normalization, are applied to re-weight the input features, selectively amplifying the significant channels.

For our task, different channels equal different spectral. We will re-weight the spectrum as we re-weight the channels, which means we can focus on the more important spectral channel and generate an image with more information. Combining the ability to focus on detail of PatchGAN and the ability to focus on different spectral channels of the spectral attention module will make the discriminator more powerful.

3.3. Loss Fuction

Our proposed S2MS-GAN uses the former loss functions, so the total loss function consists of adversarial loss and cycle-consistency loss.

3.3.1. Adversarial Loss

Using adversarial loss aims to make the image generated by the generator as real as possible to the real MS image.

$$\begin{aligned}\mathcal{L}_{\mathrm{GAN}}(G, D_{MS}, SAR, MS) = \mathbb{E}_{ms\sim p_{\mathrm{data}}(ms)}[\log D_M S(ms)] \\ +\mathbb{E}_{sar\sim p_{\mathrm{data}}(sar)}[\log(1 - D_{MS}(G(sar)))]\end{aligned} \tag{2}$$

where $sar \sim p_{\mathrm{data}}(sar)$ represent the real SAR image sampled from the SAR distribution. $ms \sim p_{\mathrm{data}}(ms)$ is the real MS image sampled from the MS distribution. $\|\cdot\|_1$ shows the L1 paradigm, which indicates the difference between the two distributions.

3.3.2. Cycle Consistency Loss

Cycle-consistency loss is used so that the samples generated by the two generators do not conflict with each other. The previous adversarial loss only ensures that the generator-generated samples are identically distributed with the real samples, but we want the images between the corresponding domains to be one-to-one, which means SAR-MS-SAR can also transfer back again.

$$\begin{aligned}\mathcal{L}_{\mathrm{cyc}}(G, F) = \mathbb{E}_{sar\sim p_{\mathrm{data}}(sar)}[\|F(G(sar)) - sar\|_1] \\ + \mathbb{E}_{ms\sim p_{\mathrm{data}}(ms)}[\|G(F(ms)) - ms\|_1].\end{aligned} \tag{3}$$

The total loss can be written as follows:

$$\begin{aligned}\mathcal{L}(G, F, D_{SAR}, D_{MS}) = & \mathcal{L}_{\text{GAN}}(G, D_{MS}, SAR, MS) \\ & + \mathcal{L}_{\text{GAN}}(F, D_{SAR}, MS, SAR) \\ & + \lambda \mathcal{L}_{\text{cyc}}(G, F). \end{aligned} \tag{4}$$

Generator G implements transfer from SAR to MS as much as possible, and generator F implements transfer from MS to SAR as much as possible. And at the same time, it is hoped that the generators can iterate back to each other to return to their own. In the total loss function, the parameters are continuously adjusted as the model undergoes training. The ultimate goal is an ongoing process of optimizing the overall loss function.

4. Experimental Section

In this section, we present a dataset containing paired SAR images and MS images, which is named S2MS-HR. Based on the dataset, we compared our method with several baselines using different evaluation methods. And quantitative and qualitative evaluations are presented below. Additionally, we conducted an ablation study to measure the validity of our method. Furthermore, we present a comparative analysis of the ablation study conducted in this experiment. Details are provided in the following sections.

4.1. Datasets and Preprocess

Obtaining very high-resolution SAR and MS images presents significant challenges due to data confidentiality, spatial and temporal inconsistencies, and limited availability. As a result, acquiring precisely aligned SAR and MS images of the same region remains difficult. To address this, we aim to construct a paired dataset for our experiments, featuring images with a very high resolution of 0.3 m for both SAR and MS data. Our SAR images, characterized by a shorter wavelength and a higher signal-to-noise ratio compared with standard SAR data, deliver enhanced resolution. This improvement in spatial resolution offers the potential for more detailed and accurate image generation, surpassing the capabilities of existing open-source remote sensing datasets. Table 1 provides a comparative overview of the resolution of various remote sensing image datasets.

Table 1. Available multispectral, SAR image resolution. Resolution refers to the spatial resolution to show the resolving power of the data collected by the individual satellites.

Multispectral Satellite	MS Image Resolution	SAR Satellite	SAR Image Resolution
LANDSET-8	15 m	GAOFEN-3	1 m
Sentinel-2	10 m	SeaSet	25 m
Worldview-1	0.5 m	TerraSAR-X	6 m
Worldview-2	1.8 m	Alos	10 m
Quickbird	2.44 m	Sentinel-1	2 m

In this step, the initial images are preprocessed by geometric correction and atmospheric correction in order to enhance the clarity of the images. In the meantime, the SAR images and multispectral images are paired correctly. All the paired images are cropped into 512 $\times$ 512 for training and testing. Our S2MS-HR dataset has 1136 paired images, which contain various topographic categories, such as forests, lakes, residential buildings, factory buildings, farmlands, playgrounds, and roads. Using different classes of images can enhance the robustness of trained models and efficiently evaluate our model's generation capability in general remote sensing image scenes. The training and testing datasets are partitioned according to a 9:1 ratio.

Figure 4 showcases a variety of scenes, including areas within the forest and along the lake's edge. These regions are captured using multispectral imaging, which goes beyond the standard RGB spectrum to provide a richer array of information. This capability highlights the advantages of multispectral imagery over conventional optical images. Additionally,

the two zoomed-in images demonstrate the exceptionally high resolution, allowing us to discern various types of buildings and vegetation. The detailed layout of the houses and even the shapes of the rooftops are clearly visible.

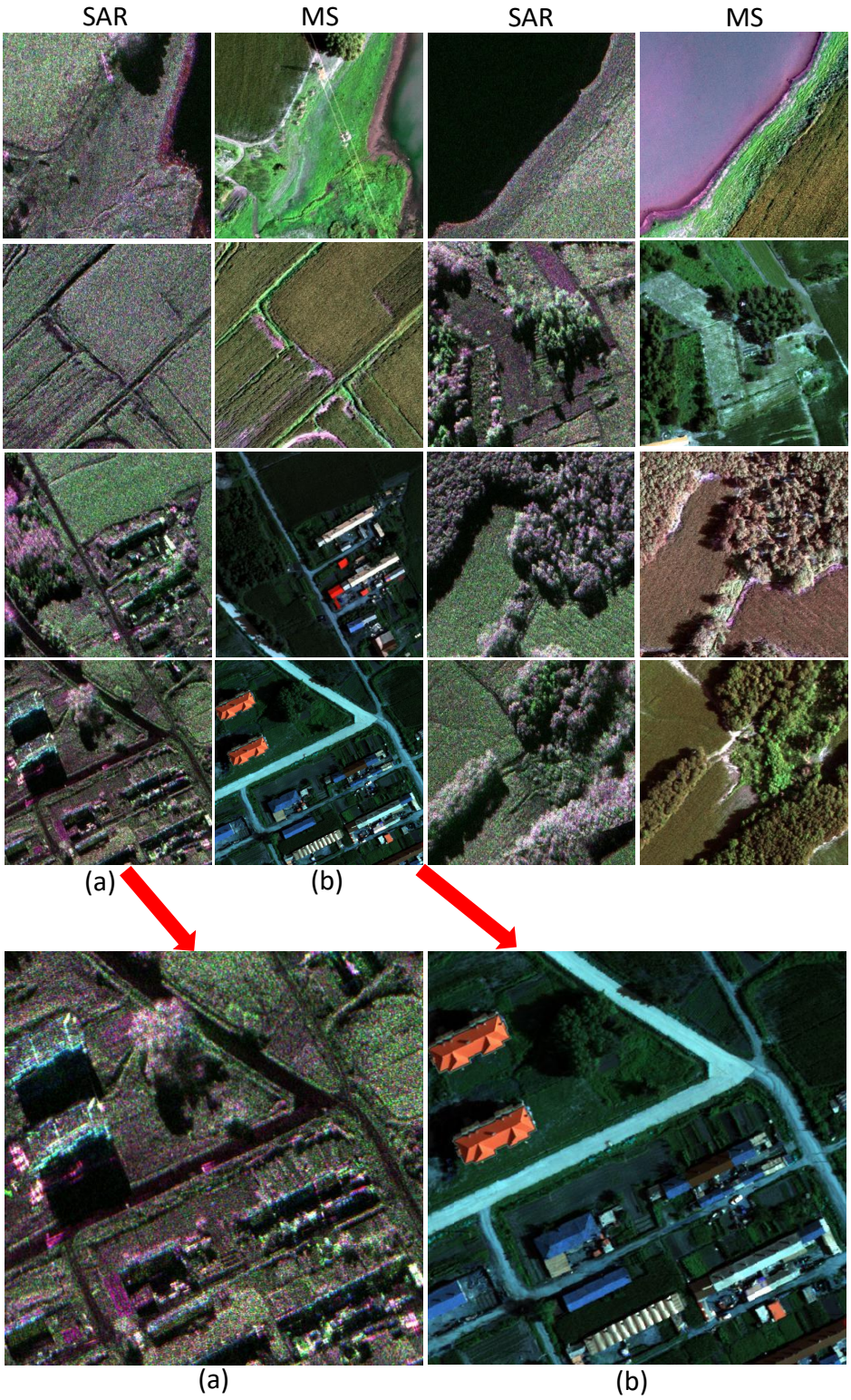


Figure 4. Sample images from S2MS-HR dataset. This is sorted from the first to the eighth group in a left-to-right, top-to-bottom direction. The seventh group of images was enlarged, shown by SAR image (**a**) and MS image (**b**).

4.2. Implementation of S2MS-GAN

In the training stage, we use the PyTorch and a single NVIDIA RTX Titian X with 24 GB of GPU memory. And all environments are set on the server system of Windows. Also, we set batchsize at 16, λ_{TV} at 0.8, and we use Adam Optimizer. We begin to set the learning rate at 0.0002 for the first 250 epochs and continue to decay linearly for the remaining 250 epochs. Since our main task is image translation from SAR image to MS image, we modify the ratio of the weights of the cycle-consistency loss function, i.e., $\lambda_S 2MS$: $\lambda_M S2SAR$ = 2:1, which can put the most attention on the process of S2MST.

4.3. Evaluation Metrics

A series of evaluations are conducted through three mainstream metrics, which are peak signal-to-noise ratio (PSNR) [56], structural similarity index (SSIM) [57], learned perceptual image patch similarity (LPIPS) [58], Multiscale Structural Similarity Index (MSSSIM) [59], and Spectral Angle Mapper (SAM) [60].

PSNR is an objective evaluation metric widely used to measure the quality of an image and is particularly suitable for evaluating image quality in tasks such as image denoising, generation, and so on. The primary objective of PSNR is to precisely quantify the extent of distortion introduced during compression or generation processes by evaluating the pixel-wise differences between two images.

Based on the perceptual model of the human visual system, the SSIM is designed to measure three aspects of two images: luminance, contrast, and structure. The three measures are integrated into one formula, and the values of the formula are from -1 to 1. A value of 1 indicates that the two images are identical, 0 indicates no correlation, and a negative value indicates inverse correlation.

LPIPS is based on a deep neural network to model human perception of image similarity, which utilizes features in a pre-trained CNN to measure the differences between two images. The core idea of LPIPS is that the effectiveness of image generation should rely on high-level semantic information rather than just a simple comparison between pixels. The LPIPS formula can be shown as follows:

$$\mathrm{LPIPS}(x, y) = \sum_{l} w_l \cdot \|f_l(x) - f_l(y)\|_2 \tag{5}$$

where $f_l(x)$ and $f_l(y)$ are the features of images x and y in the NO. l layer of the pre-trained model. w_l is a weighting parameter for each layer, and $\|\cdot\|_2$ is the L2 paradigm for calculating the distance between features.

MSSSIM, an extension of SSIM, provides a more comprehensive evaluation of remote sensing images by considering multiple scales. At higher resolutions, it emphasizes detailed features, while at lower resolutions, it captures the overall structure of the image. This multi-scale approach enables MSSSIM to more accurately reflect the image's characteristics across various levels of detail, offering improved precision and stability in assessing perceived image quality compared with the single-scale SSIM. The computation of MSSSIM involves the following steps: Starting with the original image, downsampling is performed to generate images at different scales. For each scale, the SSIM value between images is computed. These SSIM values are then combined using a weighted average to produce the final MSSSIM score. Spectral Angle Mapper (SAM) evaluates the similarity between two spectral vectors by calculating the angle between them; the smaller the angle, the more similar they are. SAM is not affected by changes in illumination because it only considers the direction of the spectral vectors and not their magnitude and is therefore particularly suitable for remote sensing image processing. Here is the formula of SAM:

$$\theta = \arccos\left(\frac{\sum_{i=1}^{n} T_i R_i}{\sqrt{\sum_{i=1}^{n} T_i^2} \times \sqrt{\sum_{i=1}^{n} R_i^2}}\right) \tag{6}$$

T_i is the target spectral vector of the i-band value of the target spectral vector. R_i is the i-th band value of the reference spectral vector. n is the total number of spectral bands. θ is the angle between two spectral vectors indicating their spectral similarity.

4.4. Results

4.4.1. Quantitative Evaluation

We validate the effectiveness of the S2MS-GAN with different comparative experiments such as pix2pix [11], CycleGAN [13], and Attention-GAN. We use CycleGAN as our baseline. And each baseline for comparing uses the parameters from their original paper, and every training dataset we use is the same. We use two comparison metrics to evaluate the efficiency of our method: quantitative evaluation and qualitative evaluation. The quantitative evaluation is set with three evaluation metrics and shown in Table 2. S2MS-GAN achieves the best results for the evaluation metrics, which are SSIM, PSNR, LPIPS, MSSSIM, and SAM in our S2MS-HR dataset. The above three metrics are raised to 2.2%, 0.8%, 1.43%, 1.28%, and 0.88° to the best model before. In contrast to other perceptual models that focus on pixel-level discrimination, LPIPS emphasizes the structural, textural, and semantic information of an image. LPIPS is trained by pre-training the network with VGG, which makes it possible to focus more on the salient features of the image. However, due to the differences between natural images and remote sensing images, we place more emphasis on metrics at the pixel level. Based on all the above metrics, our model is also the best for S2MST.

Table 2. Quantitative evaluation results with different models on the S2MS-HR dataset.The upward arrow (↑) signifies that a higher value corresponds to better performance, while the downward arrow (↓) indicates that a lower value is preferred for optimal performance. **Bolded values** highlight the best results.

Models	SSIM↑	PSNR↑	LPIPS↓	MSSSIM↑	SAM°↓
pix2pix	0.1369	9.611	0.6122	0.1637	14.74
BicycleGAN	0.1146	9.746	0.6560	0.1410	14.76
CycleGAN	0.1760	11.35	0.4990	0.2241	13.42
AttentionGAN	0.2092	11.62	0.4224	0.2593	11.37
U-GAT-IT	0.1956	10.89	0.4659	0.2581	11.29
S2MS-GAN(Ours)	**0.2319**	**12.47**	**0.4081**	**0.2721**	**10.41**

4.4.2. Qualitative Evaluation

Similarly, we demonstrate the effectiveness of our model through qualitative evaluation by presenting visual results, allowing for a more intuitive comparison of the generated outputs with those of other models and highlighting our model's superior performance.

In Figure 5, the first image shows that pix2pix introduces noticeable yellowing on the right side, while BicycleGAN renders the entire image with a reddish hue. Both CycleGAN and U-GAT-IT underrepresent the green areas. In contrast, AttentionGAN produces a result that closely aligns with the ground truth. Compared with these methods, our model generates colors that are more faithful to the real image. In the second image, which depicts a countryside scene, only our model and AttentionGAN generate outputs that resemble the true colors of the scene. However, due to its learning process involving excessive vegetation details, AttentionGAN mistakenly converts some of the houses into forested areas, deviating from reality. Our model, on the other hand, accurately translates the image into a village landscape, preserving the correct positions of various buildings. The third image contains a significant portion of forested areas, and most models perform reasonably well in this case. In the fourth image, both pix2pix and BicycleGAN perform not that well, converting the lake area into dense vegetation. CycleGAN fails to adequately capture the terrain along the shore. Although AttentionGAN achieves better results, it struggles with the accurate translation of the lake. U-GAT-IT introduces severe blurring during style transfer, leading to the loss of critical details. Our model, by contrast, preserves the

most information, accurately representing roads and shoreline woodlands, making it the best-performing model in this context.

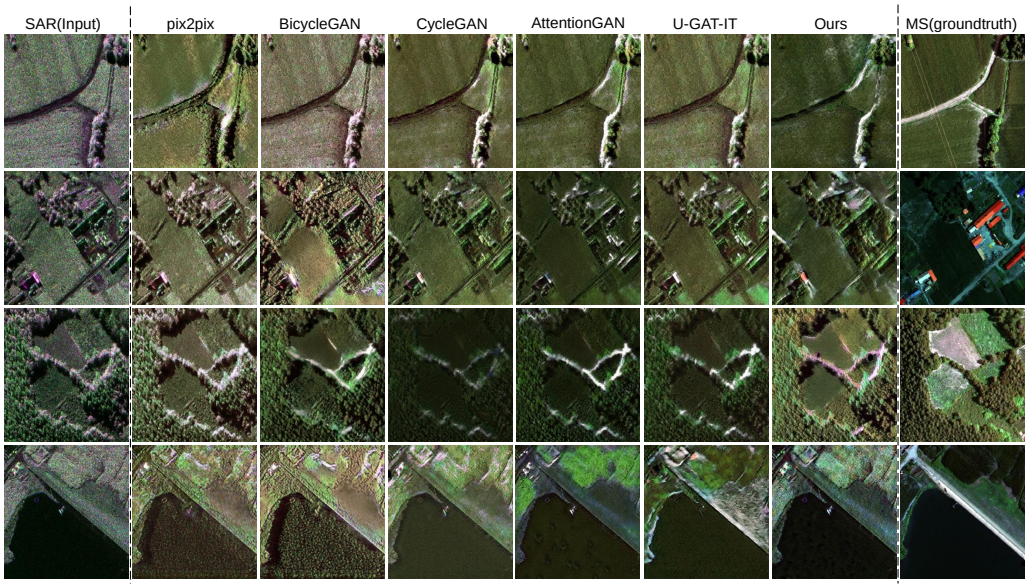


Figure 5. Example of S2MST with different baseline models. From left to right, the SAR (input) results from pix2pix, BicycleGAN, CycleGAN, AttentionGAN, U-GAT-IT, our proposed model, and MS (ground truth) are presented.

The conclusions drawn from Figure 6 are largely consistent with those from Figure 5. In the first image, CycleGAN exaggerates the green areas and reduces brightness excessively, whereas the other models achieve relatively better translation results. On the right side of the forest clearing, the SAR image contains significant noise, which our model successfully smooths, improving the image's readability. The second image presents the challenge of translating the road, which proves difficult. Compared with other models, which produce blurry translations in the bottom–right corner, our model closely aligns with the original image and maintains a consistent overall color tone. The third image is particularly notable, as only our model successfully captures the road adjacent to the vegetation. Other models either blur the road or fail to interpret the road information entirely. AttentionGAN, in particular, over-focuses on irrelevant details, generating artifacts in the center of the image. The fourth image vividly illustrates the importance of the denoising module in our model. Other translation models not only fail to remove the noise from the vegetation on the left side of the SAR image but even amplify it. In contrast, our model significantly reduces the noise while preserving the original structure of the vegetation.

Similar trends are observed in the other generated images. Our model produces results that align more closely with human visual perception, both in terms of overall spectral consistency and localized denoising capabilities, outperforming other models in these respects.

As shown in Figure 7, in the detailed view of the first set, the focus is on the magnified edge of the image. It is clearly obvious that our model effectively captures the forest edge, while other models do not perform well. Although U-GAT-IT also shows the edge, it has missing information on both sides and appears rather blurred. In the second set, the lower left corner where there is a house is magnified. Only our model successfully converts the red house, providing more abundant overall spectral information and generating a much smoother corresponding surface image.

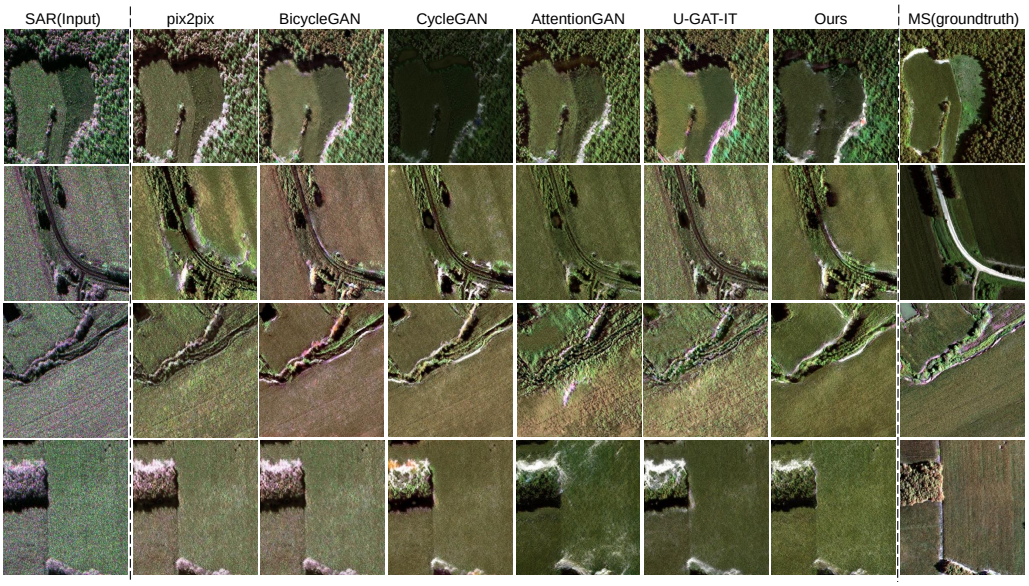


Figure 6. The other example of S2MST with different baseline models.From left to right, the SAR (input) results from pix2pix, BicycleGAN, CycleGAN, AttentionGAN, U-GAT-IT, our proposed model, and MS (ground truth) images are presented.

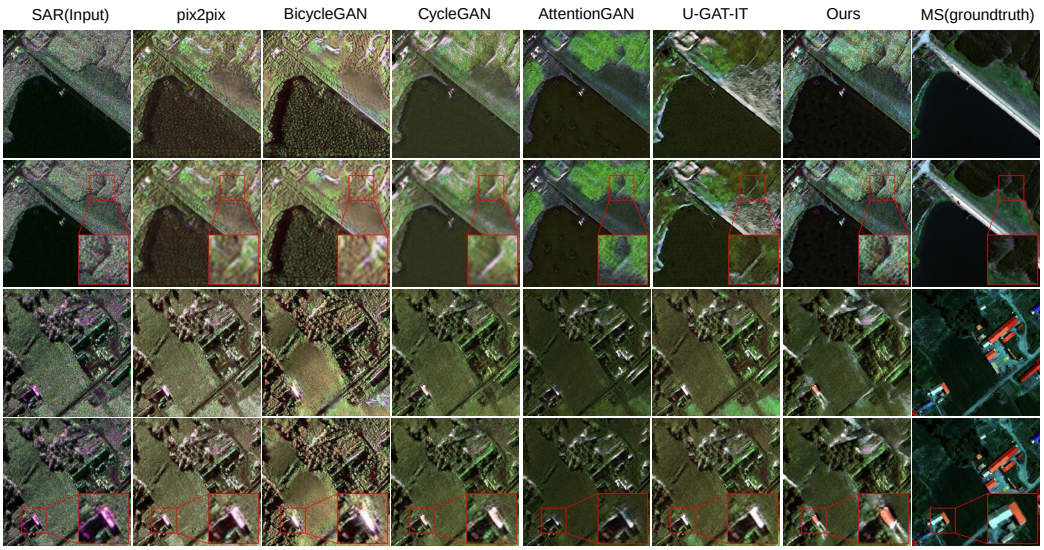


Figure 7. Enlarged views of the image details after translation by different models are presented. This figure provides a more detailed depiction of the generated results. The first and third rows represent the original images, while the second and fourth rows display the enlarged detail views.

4.5. Ablation Study

The experimental results verify the effectiveness of integrating the TV-BM3D module and spectral attention module in our task. To validate the contributions of these two components, we provide a comparison of various performance metrics based on the baseline model (CycleGAN) both with and without the individual components, as well as their combined implementation (S2MS-GAN).

In this part, +TV-BM3D and +spectral attention refer to the components separately, while +all represents S2MS-GAN. All consistent hyperparameter settings are maintained the same in the ablation experiments. As shown in Table 3, incorporating TV-BM3D and spectral attention modules improved both PSNR and SSIM values, with TV-BM3D having a more pronounced impact. This suggests that the TV-BM3D module enhances image generation by effectively removing speckle noise while preserving details. The spectral

attention module indirectly boosts the network's learning capacity by reinforcing the discriminator's ability to distinguish image features. Regarding the MSSSIM metric, both TV-BM3D and spectral attention contribute positively to image quality, with TV-BM3D slightly outperforming spectral attention. This indicates that TV-BM3D enhances the perceptual similarity of the images across multiple scales, which is critical for remote sensing imagery. TV-BM3D's superior performance in MSSSIM suggests that its denoising capability preserves structural and textural information better across scales, which is key in maintaining image quality in remote sensing applications. For the SAM metric, the introduction of spectral attention demonstrated the most significant improvement, with a substantial reduction in spectral angle compared with the baseline. This indicates that by focusing on spectral characteristics, spectral attention helps preserve the spectral fidelity of the generated images more effectively. In contrast, TV-BM3D, which primarily targets noise reduction, contributes less to the improvement of SAM, as its focus is not on spectral accuracy but rather on denoising the image while maintaining structural features. For the LPIPS metric, the addition of TV-BM3D does not lead to a significant improvement, while the spectral attention module achieves enhancement. We attribute this to the fact that TV-BM3D primarily denoises the image while retaining most of its original features, thus having minimal impact on LPIPS, which works by measuring perceptual similarity between images. In contrast, spectral attention optimizes the generated images by focusing on brightness, saturation, etc., bringing them resemblance to real images.

Table 3. Translation performance with different modules. The upward arrow (↑) signifies that a higher value corresponds to better performance, while the downward arrow (↓) indicates that a lower value is preferred for optimal performance. **Bolded values** highlight the best results.

Models	SSIM↑	PSNR↑	LPIPS↓	MSSSIM↑	SAM°↓
Baseline	0.1760	11.35	0.4990	0.2241	13.42
+TV-BM3D	0.1902	11.93	0.4817	0.2577	12.93
+Spectral Attention	0.1874	11.41	0.4385	0.2558	11.50
+all(S2MS-GAN)	**0.2319**	**12.47**	**0.4081**	**0.2721**	**10.41**

In all evaluation metrics, +all achieves the best performance. It optimizes structural details, enhances spectral channel information, and strengthens perceptual quality, while maintaining stability in the adversarial training process by refining both the generator and discriminator.

5. Conclusions

In this paper, we propose a novel high-quality image translation model from SAR images to MS images, which is named S2MS-GAN. In order to reduce the speckle noise in SAR images, we introduced a TV-BM3D module, allowing for the smoothing of noisy areas while preserving image edges and crucial details. This module integrates Total Variation (TV) regularization with the BM3D algorithm. By restricting local variations, it prevents blurring caused by excessive smoothing and minimizes the total variation of the image gradient. At the same time, the BM3D algorithm uses block matching to group similar blocks in the image and performs transformation and filtering in the three-dimensional domain. The block matching mechanism ensures the utilization of non-local similarity so as to realize the global recognition of similar structures and further enhance the noise reduction ability. In addition, we added an attention mechanism for spectra in the discriminator, which enhanced the attention to the spectral characteristics of the generated images and ensured the integrity of the spectral channel information in the MS images.

On the other hand, to address the challenge of lacking high-resolution data, we constructed a paired high-resolution SAR-to-MS image dataset, which is named S2MS-HR. Each image in this dataset has a resolution of 0.3 m and contains corresponding SAR and MS images, making it directly applicable for SAR-to-MS image translation tasks. Based on

the S2MS-HR dataset, our model has also achieved higher-quality image translation results, providing more inspiration for SAR image applications.

Author Contributions: Conceptualization, Y.L. and Q.H.; methodology, Q.H.; software, Q.H.; validation, Q.H., H.H., and Y.L.; formal analysis, Y.L.; investigation, Q.H.; resources, H.Y.; data curation, H.Y.; writing—original draft preparation, Q.H.; writing—review and editing, Y.L.; visualization, Q.H.; supervision, Y.L.; project administration, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (Grant No. 2021YFB390050*).

Data Availability Statement: The S2MS-HR dataset in this article is not readily available because the data are part of an ongoing study and research. Relevant data will be opened up progressively in the future. Requests to access the datasets should be directed to https://github.com/TimmHa/S2MS-HR (accessed on 13 September 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SAR	Synthetic Aperture Radar
MS	Multispectral
GAN	Generative Adversarial Network
S2OT	SAR-to-Optical Image Translation
S2MST	SAR-to-MS Image Translation
BM3D	Block-Matching and 3D Filtering

References

1. Naik, P.; Dalponte, M.; Bruzzone, L. Prediction of forest aboveground biomass using multitemporal multispectral remote sensing data. *Remote Sens.* **2021**, *13*, 1282. [CrossRef]
2. Berni, J.A.; Zarco-Tejada, P.J.; Suárez, L.; Fereres, E. Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 722–738. [CrossRef]
3. Quan, Y.; Zhong, X.; Feng, W.; Dauphin, G.; Gao, L.; Xing, M. A novel feature extension method for the forest disaster monitoring using multispectral data. *Remote Sens.* **2020**, *12*, 2261. [CrossRef]
4. Thakur, S.; Mondal, I.; Ghosh, P.B.; Das, P.; De, T.K. A review of the application of multispectral remote sensing in the study of mangrove ecosystems with special emphasis on image processing techniques. *Spat. Inf. Res.* **2020**, *28*, 39–51. [CrossRef]
5. Farlik, J.; Kratky, M.; Casar, J.; Stary, V. Multispectral detection of commercial unmanned aerial vehicles. *Sensors* **2019**, *19*, 1517. [CrossRef]
6. Li, Y.; Chen, J.; Zhu, J. A new ground accelerating target imaging method for airborne CSSAR. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 4013305. [CrossRef]
7. Bermudez, J.D.; Happ, P.N.; Feitosa, R.Q.; Costa, G.A.O.P.; da Silva, S.V.; Soares, M.S. Synthesis of multispectral optical images from SAR/optical multitemporal data using conditional generative adversarial networks. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1220–1224. [CrossRef]
8. Abady, L.; Barni, M.; Garzelli, A.; Basarab, A.; Pascal, C.; Frandon, J.; Dimiccoli, M. GAN generation of synthetic multispectral satellite images. In Proceedings of the Image and Signal Processing for Remote Sensing XXVI, Online, 21–25 September 2020; Volume 11533, pp. 122–133.
9. Pang, Y.; Lin, J.; Qin, T.; Chen, Z. Image-to-image translation: Methods and applications. *IEEE Trans. Multimed.* **2021**, *24*, 3859–3881. [CrossRef]
10. Alotaibi, A. Deep generative adversarial networks for image-to-image translation: A review. *Symmetry* **2020**, *12*, 1705. [CrossRef]
11. Kaji, S.; Kida, S. Overview of image-to-image translation by use of deep neural networks: Denoising, super-resolution, modality conversion, and reconstruction in medical imaging. *Radiol. Phys. Technol.* **2019**, *12*, 235–248. [CrossRef]
12. El Mahdi, B.M.; Abdelkrim, N.; Abdenour, A.; Zohir, I.; Wassim, B.; Fethi, D. A novel multispectral maritime target classification based on ThermalGAN (RGB-to-thermal image translation). *J. Exp. Theor. Artif. Intell.* **2023**, *1*, 1–21.
13. El Mahdi, B.M. A Novel Multispectral Vessel Recognition Based on RGB-to-Thermal Image Translation. *Unmanned Syst.* **2024**, *12*, 627–640. [CrossRef]
14. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 139–144.
15. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.

16. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
17. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
18. Guo, Z.; Zhang, Z.; Cai, Q.; Liu, J.; Fan, Y.; Mei, S. MS-GAN: Learn to Memorize Scene for Unpaired SAR-to-Optical Image Translation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 11467–11484. [CrossRef]
19. Guo, J.; He, C.; Zhang, M.; Li, Y.; Gao, X.; Song, B. Edge-preserving convolutional generative adversarial networks for SAR-to-optical image translation. *Remote Sens.* **2021**, *13*, 3575. [CrossRef]
20. Zhang, M.; Xu, J.; He, C.; Shang, W.; Li, Y.; Gao, X. SAR-to-Optical Image Translation via Thermodynamics-inspired Network. *arXiv* **2023**, arXiv:2305.13839.
21. Tasar, O.; Happy, S.L.; Tarabalka, Y.; Alliez, P. SemI2I: Semantically consistent image-to-image translation for domain adaptation of remote sensing data. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Honolulu, HI, USA, 26 September–2 October 2020; pp. 1837–1840.
22. Merkle, N.; Auer, S.; Mueller, R.; Reinartz, P. Exploring the potential of conditional adversarial networks for optical and SAR image matching. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1811–1820. [CrossRef]
23. Yang, X.; Zhao, J.; Wei, Z.; Wang, N.; Gao, X. SAR-to-optical image translation based on improved CGAN. *Pattern Recognit.* **2022**, *121*, 108208. [CrossRef]
24. Li, Y.; Fu, R.; Meng, X.; Jin, W.; Shao, F. A SAR-to-optical image translation method based on conditional generation adversarial network (cGAN). *IEEE Access* **2020**, *8*, 60338–60343. [CrossRef]
25. Wei, J.; Zou, H.; Sun, L.; Cao, X.; He, S.; Liu, S.; Zhang, Y. CFRWD-GAN for SAR-to-optical image translation. *Remote Sens.* **2023**, *15*, 2547. [CrossRef]
26. Enomoto, K.; Sakurada, K.; Wang, W.; Kawaguchi, N.; Matsuoka, M.; Nakamura, R. Image translation between SAR and optical imagery with generative adversarial nets. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 1752–1755.
27. Katkovnik, V.; Ponomarenko, M.; Egiazarian, K. Complex-Valued Image Denoising Based on Group-Wise Complex-Domain Sparsity. *arXiv* **2017**, arXiv:1711.00362.
28. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. SENTINEL 2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]
29. Roy, D.P.; Wulder, M.A.; Loveland, T.R.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Zhu, Z. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.* **2014**, *145*, 154–172. [CrossRef]
30. Contreras, D.; Blaschke, T.; Tiede, D.; Jilge, M. Monitoring recovery after earthquakes through the integration of remote sensing, GIS, and ground observations: The case of L'Aquila (Italy). *Cartogr. Geogr. Inf. Sci.* **2016**, *43*, 115–133. [CrossRef]
31. Mazzanti, P.; Scancella, S.; Virelli, M.; Frittelli, S.; Nocente, V.; Lombardo, F. Assessing the Performance of Multi-Resolution Satellite SAR Images for Post-Earthquake Damage Detection and Mapping Aimed at Emergency Response Management. *Remote Sens.* **2022**, *14*, 2210. [CrossRef]
32. Aoki, Y.; Furuya, M.; De Zan, F.; L-band Synthetic Aperture Radar: Current and future applications to Earth sciences. *Earth Planets Space* **2021**, *73*, 56. [CrossRef]
33. Jiang, B.; Dong, X.; Deng, M.; Wan, F.; Wang, T.; Li, X.; Zhang, G.; Cheng, Q.; Lv, S. Geolocation Accuracy Validation of High-Resolution SAR Satellite Images Based on the Xianning Validation Field. *Remote Sens.* **2023**, *15*, 1794. [CrossRef]
34. Freeman, A.; Durden, S.L. A Three-Component Scattering Model for Polarimetric SAR Data. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 963–973. [CrossRef]
35. Pohl, C.; Van Genderen, J.L. Multisensor Image Fusion in Remote Sensing: Concepts, Methods and Applications. *Int. J. Remote Sens.* **1998**, *19*, 823–854. [CrossRef]
36. Ulaby, F.T.; Moore, R.K.; Fung, A.K. *Microwave Remote Sensing: Active and Passive. Volume 1—Microwave Remote Sensing Fundamentals and Radiometry*; Artech House: London, UK, 1981.
37. Toriya, H.; Dewan, A.; Kitahara, I. SAR2OPT: Image alignment between multi-modal images using generative adversarial networks. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 923–926.
38. Qin, R.; Liu, T. A Review of Landcover Classification with Very-High Resolution Remotely Sensed Optical Images—Analysis Unit, Model Scalability and Transferability. *Remote Sens.* **2022**, *14*, 646. [CrossRef]
39. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095. [CrossRef] [PubMed]
40. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
41. O'Shea, K.; Nash, R. An introduction to convolutional neural networks. *arXiv* **2015**, arXiv:1511.08458.
42. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv* **2017**, arXiv:1710.10196.
43. Wang, L.; Xu, X.; Yu, Y.; Yang, R.; Gui, R.; Xu, Z.; Pu, F. SAR-to-optical image translation using supervised cycle-consistent adversarial networks. *IEEE Access* **2019**, *7*, 129136–129149. [CrossRef]
44. Subramanyam, M.V.; Prasad, G. A new approach for SAR image denoising. *Int. J. Electr. Comput. Eng.* **2015**, *5*, 5.

45. Devapal, D.; Kumar, S.S.; Sethunadh, R. Discontinuity adaptive SAR image despeckling using curvelet-based BM3D technique. *Int. J. Wavelets Multiresolut. Inf. Process.* **2019**, *17*, 1950016. [CrossRef]
46. Malik, M.; Azim, I.; Dar, A.H.; Asghar, S. An Adaptive SAR Despeckling Method Using Cuckoo Search Algorithm. *Intell. Autom. Soft Comput.* **2021**, *29*, 1. [CrossRef]
47. Rudin, L.I.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Phys. D Nonlinear Phenom.* **1992**, *60*, 259–268. [CrossRef]
48. Wang, Y.; Yang, J.; Yin, W.; Zhang, Y. A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. Imaging Sci.* **2008**, *1*, 248–272. [CrossRef]
49. Cai, J.F.; Chan, R.H.; Shen, Z. A framelet-based image inpainting algorithm. *Appl. Comput. Harmon. Anal.* **2009**, *24*, 131–149. [CrossRef]
50. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 184–199.
51. Zhang, T.; Wiliem, A.; Yang, S.; Lovell, B. TV-GAN: Generative Adversarial Network Based Thermal to Visible Face Recognition. In Proceedings of the 2018 International Conference on Biometrics (ICB), Gold Coast, QLD, Australia, 20–23 February 2018; pp. 174–181.
52. Vaswani, A. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762.
53. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
54. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Vedaldi, A. Gather-excite: Exploiting feature context in convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 1–11.
55. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11534–11542.
56. Tanchenko, A. Visual-PSNR measure of image quality. *J. Vis. Commun. Image Represent.* **2014**, *25*, 874–878. [CrossRef]
57. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
58. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 586–595.
59. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale Structural Similarity for Image Quality Assessment. In Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003; Volume 2, pp. 1398–1402.
60. Kruse, F.A.; Lefkoff, A.B.; Boardman, J.W.; Heidebrecht, K.B.; Shapiro, A.T.; Barloon, P.J.; Goetz, A.F.H. The Spectral Image Processing System (SIPS)—Interactive Visualization and Analysis of Imaging Spectrometer Data. *Remote Sens. Environ.* **1993**, *44*, 145–163. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.