Article

Sar2color: Learning Imaging Characteristics of SAR Images for SAR-to-Optical Transformation

Zhe Guo [ID], Haojie Guo *[ID], Xuewen Liu, Weijie Zhou, Yi Wang and Yangyu Fan

School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China
* Correspondence: ghj_xin@mail.nwpu.edu.cn

Abstract: Optical images are rich in spectral information, but difficult to acquire under all-weather conditions, while SAR images can overcome adverse meteorological conditions, but geometric distortion and speckle noise will reduce the quality of SAR images and thus make image interpretation more challenging. Therefore, transforming SAR images to optical images to assist SAR image interpretation will bring opportunities for SAR image application. With the advancement of deep learning technology, the ability of SAR-to-optical transformation has been greatly improved. However, most of the current mainstream transformation methods do not consider the imaging characteristics of SAR images, and there will be failures such as noisy color spots and regional landform deformation in the generated optical images. Moreover, since the SAR image itself does not contain color information, there also exist many color errors in these results. Aiming at the above problems, Sar2color, an end-to-end general SAR-to-optical transformation model, is proposed based on a conditional generative adversarial network (CGAN). The model uses DCT residual block to reduce the effect of coherent speckle noise on the generated optical images, and constructs the Light atrous spatial pyramid pooling (Light-ASPP) module to mitigate the negative effect of geometric distortion on the generation of optical images. These two designs ensure the precision of texture details when the SAR image is transformed into an optical image, and use the correct color memory block (CCMB) to improve the color accuracy of transformation results. Towards the Sar2color model, we have carried out evaluations on the homologous heterogeneous SAR image and optical image pairing dataset SEN1-2. The experimental results show that, compared with other mainstream transformation models, Sar2color achieves the state-of-the-art effect on all three objective and one subjective evaluation metrics. Furthermore, we have carried out various ablation experiments, and the results show the effectiveness of each designed module of Sar2color.

Keywords: SAR image; optical image; SAR-to-optical transformation; conditional generative adversarial network (CGAN); deep learning

1. Introduction

With the continuous development of space remote sensing detection technology, remote sensing images have a wide range of application needs in land planning, environmental monitoring, resource prospection, military reconnaissance and other fields [1–4]. In the practical application environment, the optical remote sensing images contain high spectral resolution, which is more in line with human visual perception. However, the severe weather conditions such as clouds and fog will cause serious pollution to the visible light band of the optical images [5], thus reducing the quality of the optical remote sensing images, and then affecting the earth observation task. At the same time, the light conditions also restrict the long time observation, resulting in the limited use of optical images. In contrast, the microwave band of synthetic aperture radar (SAR) [6] can penetrate adverse meteorological conditions and work under all-day and all-weather conditions [7] to obtain high-resolution SAR image data. However, because SAR images are limited by the inherent side-view property of radar during imaging, there will be geometric distortion [8] in some

areas of the image, which makes SAR images mismatched with the physical structures in the real environment. At the same time, SAR is a coherent imaging method, and the complex electromagnetic wave scattering process will make SAR images appear granular appearance, and the resulting coherent speckle noise [9] will affect the acquisition of valid information in SAR images. These imaging characteristics of SAR images increase the difficulty of SAR interpretation. The imaging principles of SAR and optical images are fundamentally different, in terms of measurement methods, wavelengths, detection instruments and viewing angles [10,11]. SAR images mainly represent the structure and dielectric properties of the observed target, and the spectral information is insufficient, while optical images are rich in spectral information easily understood by human eyes, which facilitates the visual interpretation of the landscape. As shown in Figure 1, horizontal and vertical roads, as well as residential areas and adjacent green spaces, can be easily distinguished in the right optical image. However, for the left SAR image, only horizontal roads can be distinguished, and other geographic targets are difficult to interpret, which makes the interpretation of SAR images very difficult for those people who are not specialized in studying the background of SAR images.
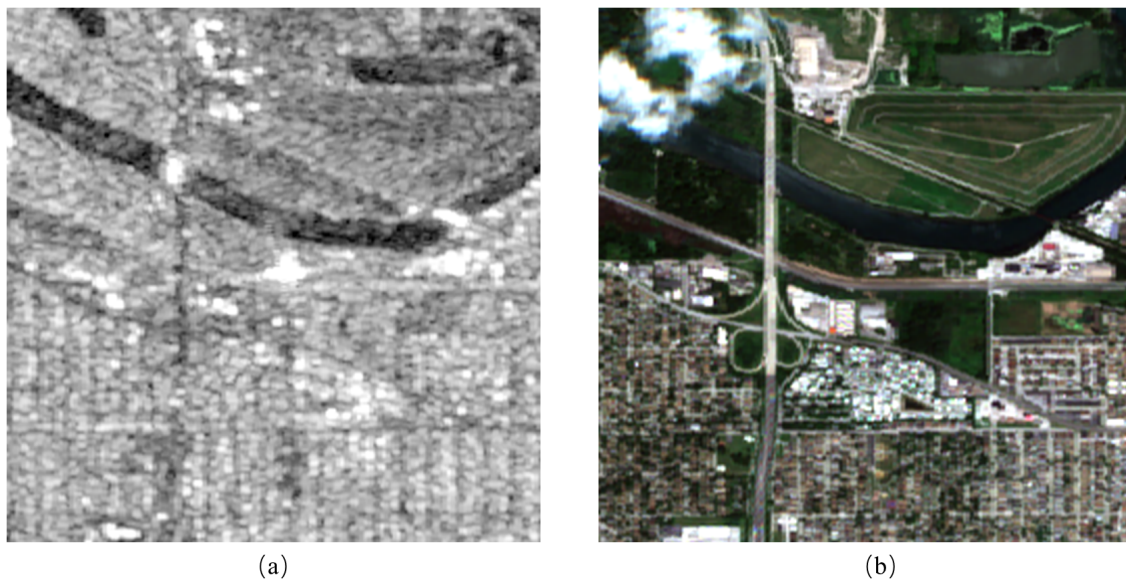

(a)   (b)

Figure 1. The comparison between (**a**) SAR image and (**b**) optical image. Best viewed in color.

When both of the above remote sensing images can be acquired, reliable registration techniques of SAR and optical images [12,13] can be used to combine the image information, which gives full play to their respective advantages and maximizes the acquisition of terrain characteristics of the target scene. However, for some specific application scenarios, we cannot acquire both SAR and optical images, so the above image registration techniques cannot be effectively applied. In some cases, only SAR images can be obtained, while optical images are difficult to be obtained [10]. At this time, under the condition of reasonable use of amplitude information of SAR image, converting SAR image to an image similar to optical representation is conducive to the interpretation of SAR image and better helps people to understand different geographical target features and scene information in SAR image. The process of inputting SAR images to generate optical images is called the transformation from SAR image to optical image, which is an image generation transformation task.

In recent years, due to the rapid development of deep learning technology, generative adversarial network (GAN) [14], as an excellent deep generative model, uses the generator and the discriminator for generative adversarial learning to learn the internal distribution characteristics of data. In addition, conditional generative adversarial network (CGAN) [15] uses the condition information corresponding to the image data on the basis of GAN for training, which can generate images distributed by specific data according to the given con-

dition information, and has been widely used in the field of image transformation. In 2017, Isola et al. proposed Pix2pix [16] based on CGAN, which serves as a general image transformation framework and provides a broad reference for subsequent image transformation work. Meanwhile, Zhu et al. proposed CycleGAN [17], which consists of two generators and discriminators, and can use unpaired images for image transformation tasks. In 2018, Wang et al. proposed the high-resolution image transformation method Pix2pix-HD [18], which uses a multi-level structure in the generator part to continuously improve the resolution and also utilizes a multi-scale discriminator to improve the generation quality. In recent years, there are also many studies on image transformation using other conditional information such as input image [16,19], text [20,21], color information [22,23], domain information [24,25], etc. The above mentioned image transformation works are focused on natural images and sketch data applications, and the work on SAR-to-optical image transformation using deep learning techniques started late. There are large differences between SAR images and optical images, CGAN can generate images that match the real data distribution, and can generate optical images based on SAR images by adjusting the network of the image-to-image transformation task. Merkle et al. [26] used the generated images to achieve image registration, while they also tried to transform SAR images to optical images, showing the great potential of deep learning in the field of remote sensing image transformation. Kento et al. [27] introduced regional information into the image transformation network based on the CGAN structure. Specifically, feature vectors from the pre-trained classification network are fed to the generator and discriminator networks, which alleviates the problem of colorization errors in generated images due to the lack of color information in SAR images. Yu et al. [28] addressed the problem of speckle noise in SAR images by introducing attention mechanism into the CGAN structure, which allows the network to emphasize useful features and ignore unimportant noise features, and thus improves the quality of optical image generation. Faramarz et al. [29] used two CGAN structures in SAR-to-optical image transformation and cloud removal tasks, replaced vanilla U-net with dilated residual inception blocks (DRIBs) in the generator, and further reduced the number of downsampling and extended the received field of view by expanding the convolution, which improves the quality of the generated images. Zuo et al. [30] designed a histogram of orientated phase congruency (HOPC) as a loss function in the SAR image transformation in order to better preserve the structural information after image transformation. Javier et al. [31] proposed an Atrous CGAN structure that enhances the fine details in the generated optical images by using multi-scale spatial context. Guo et al. [11] proposed an edge-information-preserving CGAN (EPCGAN) for SAR-to-optical image transformation, which enhances the structural information and sharpness of the generated optical images. In 2021, Tan et al. [32] proposed a feature-preserving heterogeneous remote sensing image transformation model called Serial GAN for the SAR-to-optical transformation problem, which uses two sub-networks, Serial Despeckling GAN and Colorization GAN, to complete the transformation of SAR image to optical grayscale image and optical grayscale image to optical color image, reducing the semantic distortion and spectral distortion in the direct SAR-to-optical transformation process.

For the SAR-to-optical image transformation task, SAR images and optical images belong to homogeneous and heterogeneous data, and the imaging characteristics of the two images are different significantly. The above mentioned Merkle et al., Kento et al., Faramarz et al., Zuo et al., and Javier et al. all improve the transformation quality by improving the network model and loss function, without considering the imaging characteristics of SAR images. Yu et al. consider the speckle noise, but the attention mechanism greatly increases the network parameters, resulting training difficulties. Guo et al. consider the structural information in SAR images to recover more structural information in the generated optical images, but the color information in their generated optical images is scarce. Tan et al. propose an alternative idea of SAR image transformation by decoupling SAR-to-optical into two subtasks to reduce the transformation difficulty, but its network does not consider the geometric distortion characteristics in SAR images, and some regional features in the

generated optical images do not match the real physical features, and the network is not an end-to-end network.

In order to implement SAR-to-optical image transformation more efficiently, we propose an end-to-end SAR-to-optical image transformation model called Sar2color in full consideration of the imaging characteristics of SAR images. Our model is based on the CGAN structure, and the DCT residual block (DCTRB) is designed in the encoding part of the generator to reduce the influence of coherent speckle noise in SAR image on the generated optical images. The Light atrous spatial pyramid pooling (Light-ASPP) module is designed in the middle part of the generator to improve the utilization of spatial multi-scale information and alleviate the negative effect of geometric distortion on the generated optical images. To improve the color quality of generated optical images, a correct color memory block (CCMB) is designed in the decoding part of the generator. The above design reduces the difficulty of transforming SAR images into optical images, and also makes the generated optical images closer to the real optical images.

The main contributions of this paper are as follows:

1. We propose an end-to-end SAR-to-optical image transformation model called Sar2color, which takes into account the imaging characteristics of SAR images, reduces the adverse effects of SAR images brought to the generated optical images during the transformation process, improves the color quality of the generated optical images, and is conducive to assisting the interpretation of SAR images;
2. In this paper, DCTRB and Light-ASPP modules are designed to reduce the negative effects of coherent speckle noise and geometric distortion characteristics in SAR images on the generation of optical images, and thus reduce the difficulty of SAR-to-optical image transformation task;
3. A CCMB module is proposed to alleviate the problem of color deviation that occurs in generating optical images;
4. This paper evaluates the results of the proposed method on the remote sensing image and optical image paired dataset SEN1-2 [33], while achieving the state-of-the-art effect on four different mainstream evaluation metrics, such as peak signal to noise ratio (PSNR) [34], structural similarity index metric (SSIM) [35], mean square error (MSE) [36], and learned perceptual image patch similarity (LPIPS) [37].

2. Methods

We propose the Sar2color to solve SAR-to-optical transformation task, which has the ability to transform SAR images into images that are closer to real optical images, and the overall model structure is shown in Figure 2. Sar2color model is based on CGAN structure and consists of generator and discriminator. The generator is composed of DCTRB, Light-ASPP, CCMB, and other modules (contains convolution and deconvolution operations, bacth normalization and some activation functions). The discriminator has the dual-scale patch discriminator structure, which is a multi-layer network composed of multiple convolutions, batch normalization, activation functions and Light-ASPP blocks. The generator is responsible for training the input SAR image to obtain the resulting image, and then the discriminator will discriminate the generated image from the real optical image, and the generator and discriminator are trained against each other to improve model performance. The generator attempts to produce a result closer to the real optical image, while the discriminator tries to distinguish the generated result from the real image as much as possible.
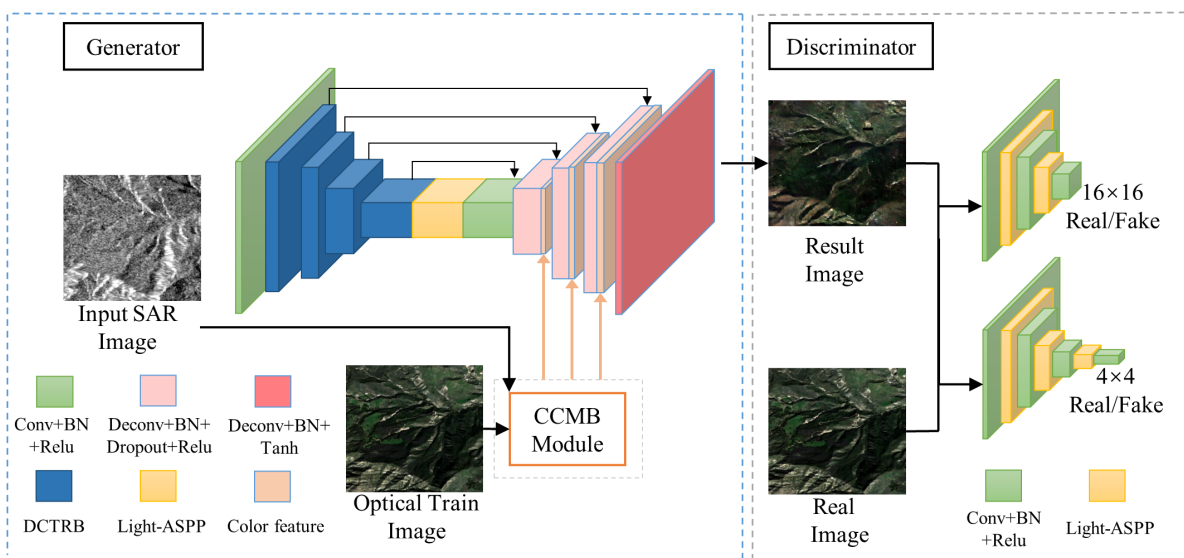
Figure 2. The overall architecture of Sar2color. Best viewed in color.

2.1. The Structure of Sar2color Model

2.1.1. Generator

As shown in Figure 2, our generator part here uses a U-net [38] structure with a downsampled image encoder section and an upsampled image decoder section.

We design DCTRB with different sub-sampling resolutions in the encoder of the generator, when the encoder extracts the main features of the SAR image, it can reduce the interference of invalid speckle noise features in the SAR image to the subsequent generation of optical image. The specific structure of the DCTRB block is shown in Figure 3.
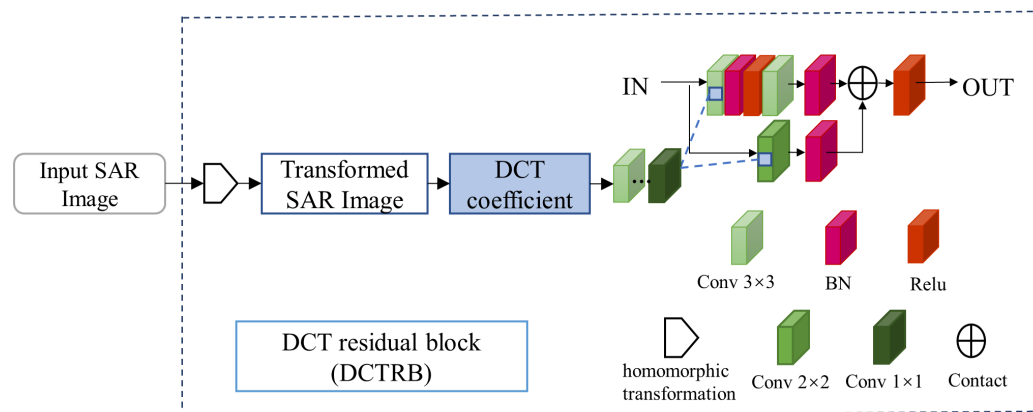

Figure 3. The architecture of DCTRB. DCT coefficient is a 256×256 size matrix, conv 3×3, conv 2×2 and conv 1×1 are the convolution operations of different convolution kernel sizes. The blue dotted line represents the concatenation on the feature channel of the DCT coefficient feature map and the SAR image feature map. BN represents batch normalization, Relu represents an activation function, and Contact represents addition operation. Best viewed in color.

Here, DCT coefficients are introduced into the residual shrinkage block [39], the DCT [40] is a real domain transformation. For an $M \times N$ image matrix, its DCT is defined as Formula (1).

$$F(m,n) = \alpha(m)\alpha(n) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \times \cos\left[\frac{(2x+1)\pi m}{2M}\right] \cos\left[\frac{(2y+1)\pi n}{2N}\right],$$

$$\alpha(m) = \begin{cases} \frac{1}{\sqrt{M}}, & m = 0 \\ \sqrt{\frac{2}{M}}, & 1 \leqslant m \leqslant M-1 \end{cases}, \alpha(n) = \begin{cases} \frac{1}{\sqrt{N}}, & n = 0 \\ \sqrt{\frac{2}{N}}, & 1 \leqslant n \leqslant N-1 \end{cases} \tag{1}$$

where $M = N = 256$, the coefficients $\alpha(m)$ and $\alpha(n)$ make the DCT transform matrix an orthogonal matrix. x and y represent the position of the image pixel in the horizontal and vertical directions, $f(x,y)$ is the pixel value matrix of the image in the spatial domain (x,y) coordinates. m and n are the frequencies of the image in the horizontal and vertical directions after calculation, $F(m,n)$ is the transformation matrix in the frequency domain after calculation.

DCT has the characteristics of low computational complexity and concentrated energy. It can remove the redundant information of image and retain the information that can best express the features of the image. After DCT, most of the energy of the image is concentrated in a small range of the frequency domain, including the main visual information in the image, which can be selected and extracted according to the practical applications. As shown in Figure 4, after DCT, most of the SAR image energy is concentrated in the low-frequency DCT coefficient, so the low-frequency coefficient is more important than the high part. As the speckle noise in SAR image belongs to the multiplicative noise and high-frequency noise, it is difficult to deal with the multiplicative relationship between the multiplicative noise and the image signal. We first do the homomorphic transformation process on the SAR image to transform the multiplicative noise in the image into additive noise, and then calculate the DCT coefficients of the transformed SAR image to remove the high frequency noise information in the image. Nevertheless, discarding high-frequency information completely will lose more image details.
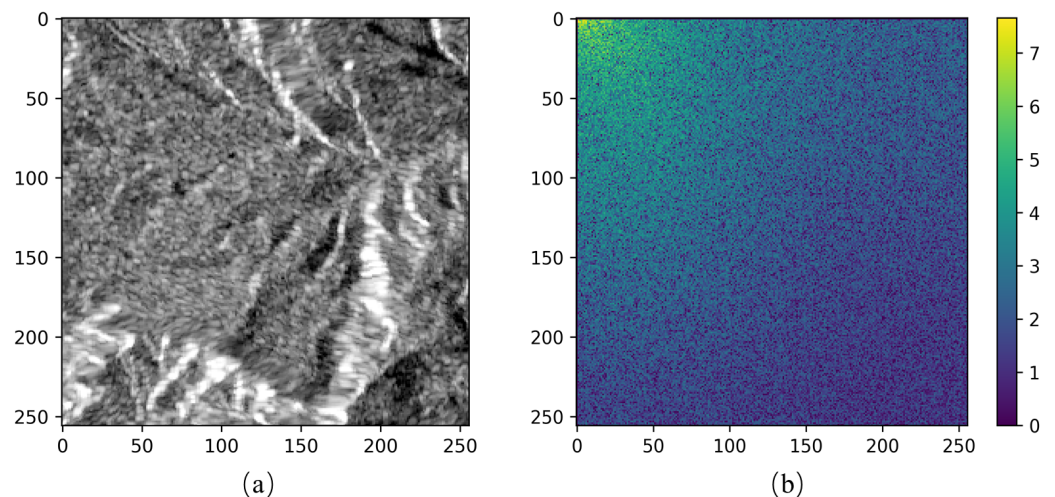

Figure 4. The characteristic of the DCT. (**a**) The SAR image, and (**b**) the corresponding DCT image. Best viewed in color.

In order to remove the high-frequency noise as much as possible and also retain the coefficients that can best express the image features, we utilize the coefficient discriminant method to select and retain the low-frequency coefficients and the medium and high-frequency coefficients with higher energy. The specific execution process is as follows:

- Firstly, the SAR image is processed by homomorphic transformation, and the transformed SAR image is obtained by the following formula: $f(x,y) = \ln(g(x,y))$, $f(x,y)$ and $g(x,y)$ represent the transformed SAR image and SAR image under the spatial coordinates (x,y), respectively;

- Then, the transformed SAR image is divided into 8 blocks, DCT is applied in each sub-block, and then ZigZag [41] scanning is used to retain 10 low-frequency coefficients in the upper left corner of the coefficient matrix after DCT;
- For other medium and high-frequency coefficients, the average coefficient of each sub-block is set as the screening threshold, and the remaining 54 coefficients are compared with this threshold of each sub-block. If the coefficient is larger than the threshold, it is retained, and if less than 0, set it to 0. In this way, the medium and high-frequency coefficients with higher energy can be retained. Finally, the filtered coefficients of each sub-block are integrated.

We calculate the two-dimensional pixel matrix of the input transformed SAR image to get the corresponding DCT coefficient. After distinguishing the coefficient, we retain the DCT coefficient of the effective image information, and get a coefficient matrix of 256×256 size. After the matrix is operated by conv 3×3 and conv 1×1, we get a 64×64 size DCT coefficient characteristic graph. The input 256×256 size SAR image is convoluted by the first layer in the encoder to get the 64×64 size feature image, which is spliced with the DCT coefficient feature map on the feature channel, and then continues to be trained in the network for DCTRB with different downsampling resolutions in the encoder. The DCT coefficient matrix performs a number of conv 3×3 and conv 1×1 operations to obtain the DCT coefficient feature map that matches the size of the SAR image feature map.

To reduce the negative impact of geometric distortion characteristics in SAR image on the generated optical image, we design the Light-ASPP structure to improve the utilization of spatial multi-scale information. The Light-ASPP structure originates from the DeepLabv3+ paper [42]. ASPP uses the spatial pyramid pooling module to further extract multi-scale information. To reduce the computational complexity without compromising accuracy, we use more conv 1×1 operations and a lower rate of Atrous conv with a lower rate of (1,6,12) on the original ASPP structure to preserve the extraction of multi-scale features, while adding a global pooling layer to retain global context information. Consequently, we get a lightweight structure called Light-ASPP, we insert the Light-ASPP after DCTRB of the encoder. Multi-scale feature extraction of the encoded SAR image can reduce the adverse effects of local geometric distortion and provide better conditions for subsequent decoding and restoration of optical image. The output of Light-ASPP is connected to the back convolution module through a conv 1×1 operation, which ensures the matching of feature sizes.

In the decoder of the generator, we use deconv operation to upsample and restore the image resolution. As SAR does not contain color information, if directly upsampling, there will be a larger coloring error in the resulting image. To improve the color accuracy of the generated optical image, inspired by the paper [23,42,43], we design CCMB to implement color enhancement processing on the upsampled image.

The structure of CCMB is shown in Figure 5. For the optical remote sensing images in the training set, we take the pooling layer information of the corresponding grayscale images after conv 3_2 through the pre-trained ResNet18 network [39,44] as spatial information s and then use colorthief [45] to extract the Top-5 color values of different optical remote sensing images in the training set. After normalization, we get the color information c, so the image matching information M_{Info} can be described as:

$$M_{\text{Info}} = (s_1, c_1), (s_2, c_2), \ldots, (s_m, c_m) \tag{2}$$

where m represents the number of the training image. During the training step, the KNN method [46] is used to match the pooling layer information of the input SAR image after the conv 3_2 of the pre-trained ResNet18 network with the spatial information of the training set, and the color information corresponding to the most similar spatial information is obtained. The color information is then output at different characteristic scales through a full convolutional encoder network. The feature map of the corresponding scale of the

decoder is connected with Concat [38] to enhance the color information of the restored optical image.
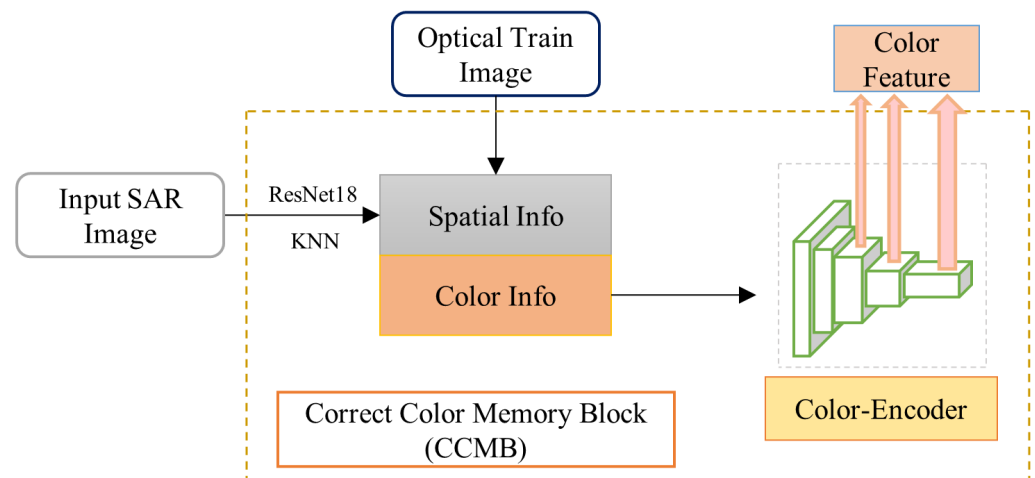

Figure 5. The architecture of CCMB. The spatial information of the input SAR image is calculated by the ResNet18 network, and the color information corresponding to the most similar spatial information in the optical training set image is returned by the KNN algorithm. The color information is input into the full convolutional encoder network, and the color features of different feature scales are cascaded with the decoder. Best viewed in color.

2.1.2. Discriminator

The goal of the discriminator is to distinguish between the generated image and the real image obtained by the generator. In the network training, the generator continues to improve the effect of the generated image to be closer to the real image. The discriminator continues to improve the discriminant ability to distinguish the generated image from the real image as much as possible, and the two carry out confrontation training to get a higher quality generated image.

As shown in Figure 2, we use a dual-scale patch discriminator, which consists of two sub-discriminators [18,47] with different output scales. To better distinguish the resulting image and the real image, the Light-ASPP is connected after the convolution layer of the sub-discriminator to improve the discrimination effect of the context information. The output discriminant matrix of the last layer of the two sub-discriminators is 16×16 and 4×4, respectively, from top to bottom, indicating that the discriminant results are applied on different scales of the image. Using the sigmoid function as the last layer output for the conventional discriminator may lead to the extreme discrimination in a single case. Therefore the sub-discriminators here are all patch discriminators, while the patch discriminator discriminates on different image blocks, and the final discriminant matrix is the mean value of the discriminant result of each block. Because there may be great differences in the size of different objects in the remote sensing image, the dual-scale discriminator can be utilized to get different receptive field sizes on the image to better constrain the targets of different sizes. In the training stage, the loss of the discriminator is obtained by the weighted sum of two sub-discriminators.

2.2. Loss Functions

There is a great difference between SAR image and optical image, and the design of loss function has a significant influence on the quality of the generated optical image. In the training of the Sar2color model, we use adversarial loss [15], smooth L1 distance loss [18] and style loss.

Adversarial Loss: Sar2color is based on the CGAN structure, so we use the adversarial loss for the entire generator and discriminator, defined as follows:

$$L_c = E_{x,y \sim p_{\text{data}}(x,y)}[\log(D(x,y))] + E_{x,z \sim p_{\text{data}}(x,z)}[\log(1 - D(x, G(x,z)))] \tag{3}$$

where G and D denote generator and discriminator, x is the SAR image, y is the real optical image, and z is the noise. The generator aims at minimizing the target loss, while on the contrary, the discriminator works to maximize the target loss, and the two play games with each other to continuously improve the generation effect.

Smooth L1 Loss: To reduce the pixel difference between the real optical image and the generated image obtained by the generator, we use smooth L1 loss, defined as follows:

$$L_{sL1}(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & \text{for } |y - \hat{y}| \leq \delta \\ \delta|y - \hat{y}| - \frac{1}{2}\delta^2 & \text{otherwise.} \end{cases} \tag{4}$$

where y represents the generated image, and $\hat{y}$ represents the real optical image.

Style Loss: For the generated image, we also want to include more texture details. As the above loss only pays attention to the overall structure of the image content, and does not take into account the slight texture difference between the two images, so we use style loss here to enhance the performance of texture details. In this paper, we only experiment on the SAR image and the L channel from generated image. We adopt the idea of Gram matrix matching (representing feature correlation) [48] extracted from some layers of the pre-trained classification network VGG19 [44]. The Gram matrix is defined as follows:

$$G_{ij}^l = \sum_k C_{ik}^l C_{jk}^l \tag{5}$$

where $G_{ij}^l \in R^{N_l \times N_l}$, the N_l represents the number of feature maps in the lth layer of the network, and the C_{ik}^l represents the output of the ith filter at the kth channel of the lth layer. The C_{jk}^l represents the transpose vector matrix of C_{ik}^l. We use the third layer (relu2_2, relu3_2, and relu4_2) of the VGG19 network to form the loss L_s.

The final completed total loss function of the generator is defined as:

$$L_G = \lambda_1 * L_c + \lambda_2 * L_{sL1} + \lambda_3 * L_s \tag{6}$$

In our experiments, the weight is set as: $\lambda_1 = 1$, $\lambda_2 = 10$, $\lambda_3 = 0.1$.

3. Results

In this section, we first describe the details of the experiment, then introduce the used dataset and evaluation metrics, and then carry on the qualitative analysis and quantitative comparison of the experimental results. Finally, we carry out the ablation experiment of each part of our proposed Sar2color model.

3.1. Implementation Details

In the experiments, we use the PyTorch1.7 framework and an NVIDIA RTX2080TI GPU for development. In the training stage, the epoch is set to 500 and the initial learning rate is set to 10^{-3}, which attenuates to 0.5 of the original learning rate every 10 epochs at the beginning of 400 epoch. The generator is developed with an Adam optimizer of $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and the discriminator is developed with the SGD optimizer. Here, we train so that the size of all images is 256×256.

3.2. Dataset Setting

In this paper, the SEN1-2 dataset [33] proposed by Schmitt et al. is used for SAR-to-optical transformation. At present, this dataset is the unique large remote sensing dataset used for SAR-optical data fusion, which is obtained from the Sentinel-1 and Sentinel-2 satellites all over the world and in various seasons. Sentinel-1 consists of two polarorbiting

satellites, equipped with C-band SAR sensors, which enables them to acquire imagery regardless of the weather. For the Sentinel-1 images in this dataset, so-called ground-range-detected (GRD) products acquired in the most frequently available interferometric wide swath (IW) mode were used. These images contain the σ^0 backscatter coefficient in dB scale for every pixel at a pixel spacing of 5 m in azimuth and 20 m in range. For sake of simplicity, the dataset pays more attention to vertically polarized (VV) data, ignoring other potentially available polarizations. Sentinel-2 comprises twin polarorbiting satellites in the same orbit, phased at 180° to each other. The mission is meant to provide continuity for multi-spectral image data of the SPOT and LANDSAT kind, which have provided information about the land surfaces of our Earth for many decades. For the Sentinel-2 part of the dataset, it only used the red, green, and blue channels (i.e., bands 4, 3, and 2) in order to generate realistically looking optical RGB images.

The image block is cut to 256 × 256. As shown in Figure 6, SEN-1 and SEN-2 represent SAR images and optical color images, respectively, which contain 282384 pairs of SAR and optical color images with four seasons. Due to the detection limitations of the sensor, the SEN1-2 dataset has a large amount of overlapping data. Then, the same data processing method as Serial GAN is used to sample the original SEN1-2 dataset. Five typical geographical features are selected for this model: river valley, mountains and hills, urban residential area, coastal city and desert. Because the same geographical features are different in the four seasons, we extract image pairs evenly in different seasons to ensure that the number of images in different seasons is the same as possible, and the ratio of the train to test is set as 4:1. The selected dataset in our experiment is shown in Table 1.

In the training step, we aggregate the data of all the different terrains of the four seasons together to obtain an overall dataset, and use this dataset to train from scratch to obtain a single model. Then, we use this model to make predictions and obtain the resulting images.

Table 1. The number of different season types in our selected dataset.

Season	Train	Test
Spring	2000	500
Summer	1600	400
Fall	2000	500
Winter	1600	400
Total	7200	1800

Figure 6. The examples of the SEN1-2 dataset. The first row is SEN-1 data, representing SAR images, and the second row is SEN-2 data, representing optical RGB images. Best viewed in color.

3.3. Metrics

For the evaluation metric of image results in the field of remote sensing image transformation, we select four different mainstream evaluation metrics to measure the actual test

results, which are peak signal to noise ratio (PSNR) [34], structural similarity index metric (SSIM) [35], mean square error (MSE) [36], and learned perceptual image patch similarity (LPIPS) [37].

PSNR and MSE here objectively reflect the pixel distance between the generated image and the real image, and SSIM reflects the image structure similarity between the generated image and the real image. Here, x is used to represent the generated image, y represents the real image, $x_{c,h,w}$ and $y_{c,h,w}$ correspond to their pixel values, respectively. PSNR is defined as follows:

$$\text{PSNR}(x,y) = 20\log_{10}\text{Max}(y_{c,h,w}) / \left\|x_{c,h,w} - y_{c,h,w}\right\|_2 \tag{7}$$

SSIM is defined as follows:

$$\text{SSIM}(x,y) = \frac{(2x_{c,h,w}y_{c,h,w} + c_1)(2\sigma_{xy} + c_2)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{8}$$

where μ_x and μ_y represent the mean image pixel of x and y. σ_x and σ_y denote the standard deviation of image x and y, respectively, σ_{xy} represents the covariance of image x and y, c_1 and c_2 are constants.

MSE is defined as follows:

$$MSE = \frac{1}{C \cdot H \cdot W} \sum_{c=h=w=1}^{C \cdot H \cdot W} (x_{c,h,w} - y_{c,h,w})^2 \tag{9}$$

where dimension $C = 3$, $H = W = 256$.

At the same time, to avoid the limitation of objective evaluation metrics, we also use the LPIPS metric, which is a subjective evaluation metric. By calculating the distance between two images after extracting the network through perceptual features, we can better measure the subjective perception distance between the generated image and the real image.

3.4. Qualitative Evaluation

To verify the performance of Sar2color in the SAR-to-optical transformation task, we compare this model with Pix2pix-HD, Faramarz, and Serial GAN models. As shown in Figures 7 and 8, the first column and the second column represent the SAR image and the corresponding optical color image, respectively, the third to the fifth column show the experimental results of Pix2pix-HD, Faramarz, and Serial GAN, and the last column shows the experimental results of our model.

From the overall visual effect, Sar2color is closer to the real optical color image in color accuracy, color saturation, and texture details, indicating that the DCTRB and Light-ASPP modules in Sar2color can reduce the negative impact of coherent speckle noise and geometric distortion characteristics in SAR images as much as possible and thus restore the texture details of the real optical color image. CCMB module enhances the color accuracy of the resulting image. The added style loss in the loss function further improves the color saturation of the resulting image. In contrast, the retention effect of Serial GAN on texture details is similar to that of Sar2color, but there is a certain gap with Sar2color in terms of color accuracy and saturation for the urban residential area and coastal city scenes. In particular, there is a color loss in the middle part and the upper left corner of the resulting images in the above two scenes. Visually, there are some color mottling phenomena and regional color errors in the transformation results of Pix2pix-HD and Faramarz algorithms in each scene. For example, in the river valley scene, these two algorithms mistakenly regard the water area as a plain, resulting in incorrect green areas in the image, and noisy color spots in scenes such as mountains and hills, coastal city, and desert. The information interpretation of the resulting images is seriously affected.
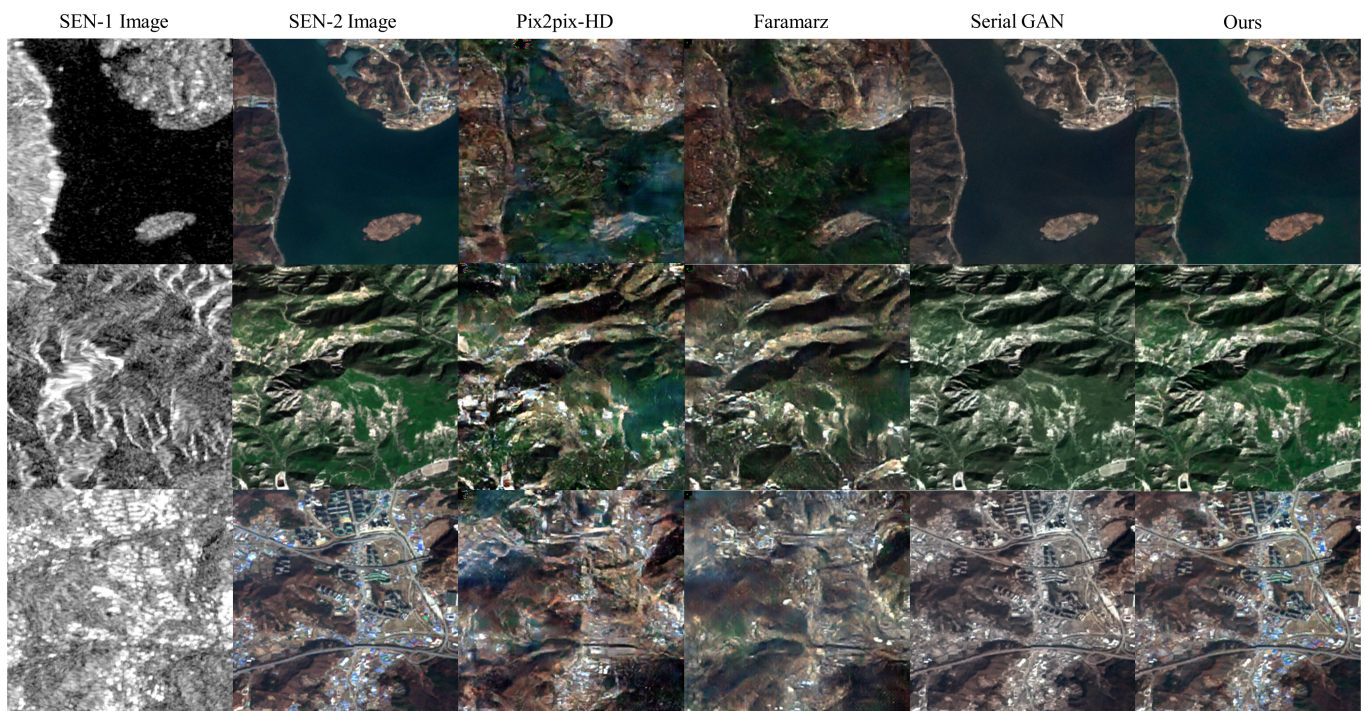
Figure 7. The results of different SAR-to-optical transformation models. The columns from left to right show SAR images, optical images, results of Pix2pix-HD, results of Faramarz et al. model, results of Serial GAN, and our results, respectively. The topography shown in the figure from top to bottom is: river valley, mountains and hills, and urban residential area. Best viewed in color.
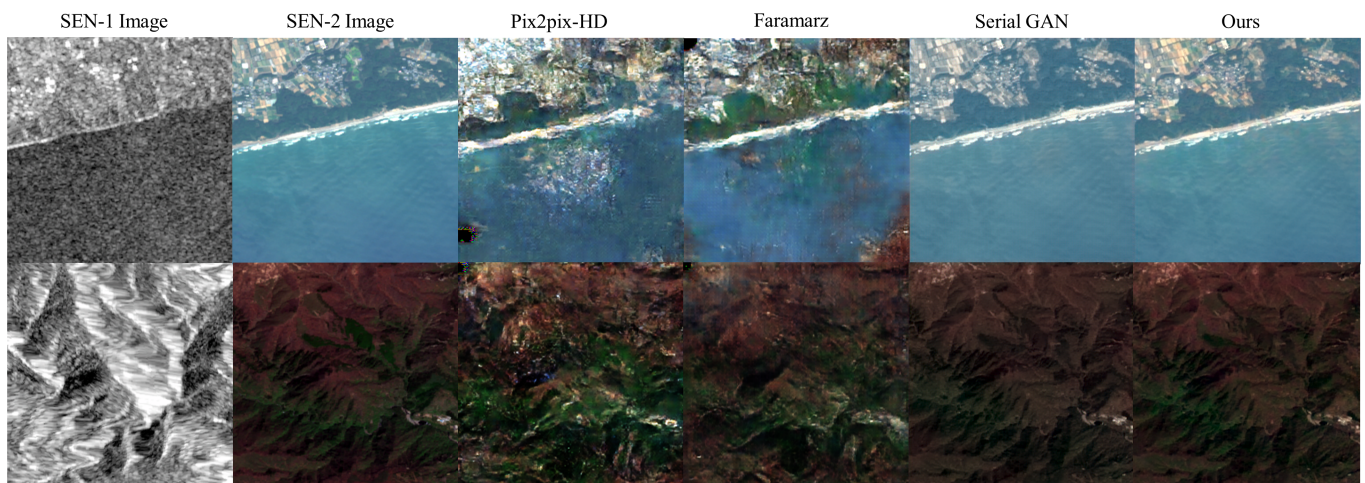

Figure 8. The results of different SAR-to-optical transformation models. The topography shown in the figure from top to bottom is: coastal city, and desert. Best viewed in color.

To further verify the performance differences between Sar2color and other models in the local details of the transformation results, as shown in Figure 9, the first column represents the optical color image corresponding to the SAR image, and the second to fifth columns show the experimental results of Pix2pix-HD, Faramarz, Serial GAN, and our Sar2color, respectively. Here, some noteworthy local areas in three different scenarios are circled in red boxes.

In the red box area of the first River valley scene, Sar2color and Serial GAN accurately translated the water area and the land boundary. The results of Pix2pix-HD and Faramarz overlap at the boundary, and Pix2pix-HD is more serious. In the red box area of the second line urban residential area scene, both Sar2color and Serial GAN show the road shape, but

Serial GAN has a color transformation error on the blue residential area of the roadside, while Sar2color gives the accurate transformation. Furthermore, Pix2pix-HD and Faramarz have the wrong transformation of the road shape and living area. In the red box area of the third line mountains and hills scene, both Sar2color and Serial GAN accurately translate the complex road shapes, while Pix2pix-HD and Faramarz mistranslate them into similar surrounding mountain landforms.

Based on the above overall visual effect and local detail comparison performance, our Sar2color model shows the best effect, and has certain performance advantages compared with other models.
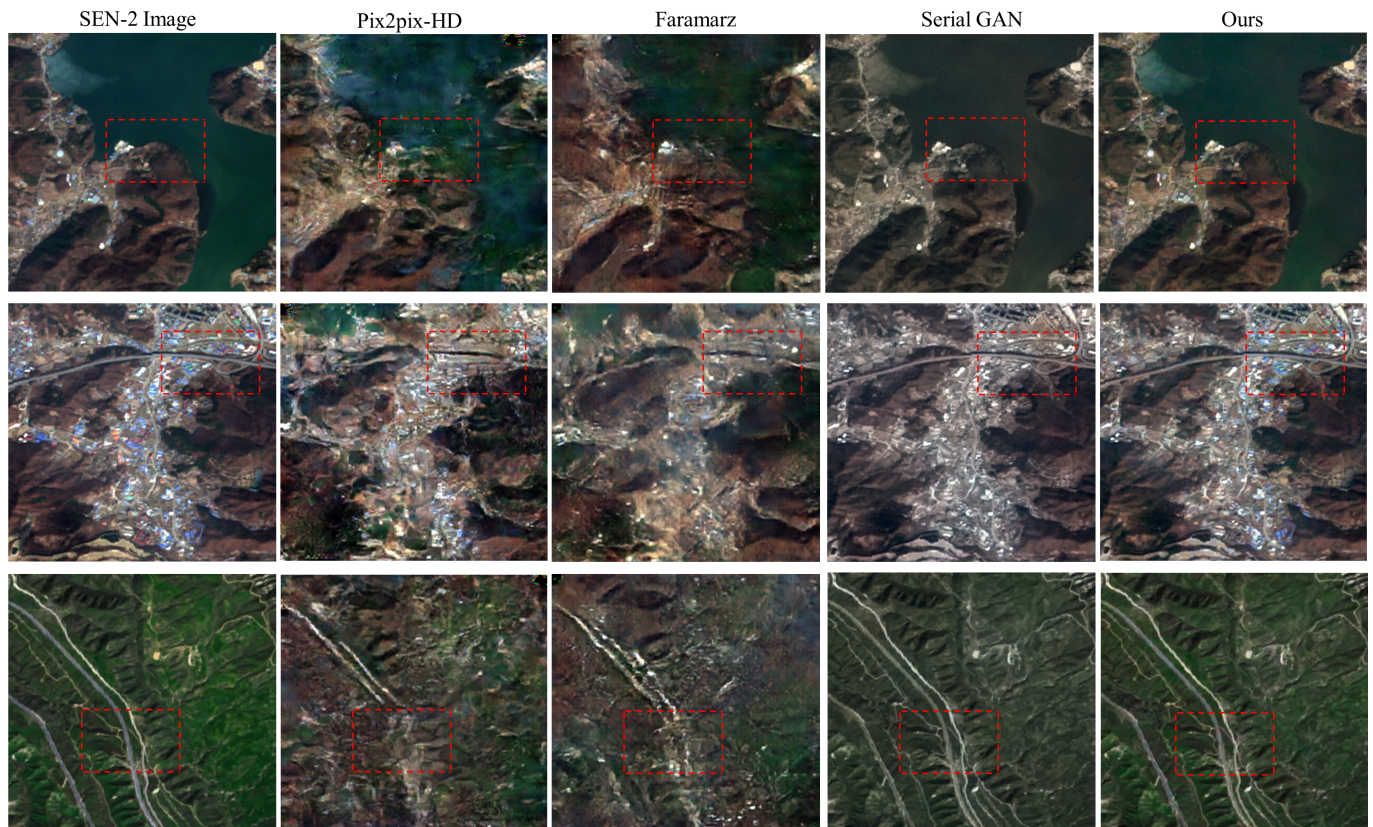

Figure 9. The local detail comparison of different SAR-to-optical transformation models, some noteworthy local areas in three different scenarios are circled in red boxes. The columns from left to right show Optical images, results of Pix2pix-HD, results of Faramarz et al. model, results of Serial GAN, and our results, respectively. Best viewed in color.

3.5. Quantitative Evaluation

To quantitatively compare the performance of different evaluation metrics of Sar2color and other models, we carried out experiments on four different metrics such as PSNR, SSIM, MSE, and LPIPS. From Table 2, we can see that Sar2color is much better than other models in terms of PSNR and MSE metrics, indicating that the transformation result of our model is closer to the real optical color image in pixel distance, and Sar2color is also better than other models in SSIM metric, indicating that the overall structure of the generated image obtained by our model is more similar to that of the real optical image. On the perceptual distance metric of LPIPS, the result of Sar2color is higher than that of other models, which reflects that our model has the best effect on human subjective visual perception, and provides convenience for subsequent image information interpretation.

Table 2. Comparison of results with quantitative evaluation of PSNR, SSIM, MSE, LPIPS by different methods on SEN1-2 dataset. The up-arrow ↑ means larger is better, while the down-arrow ↓ means smaller is better.

	PSNR↑	SSIM↑	MSE↓	LPIPS↓
Pix2pix-HD [18]	13.4253	0.2782	0.0487	0.3734
Faramarz [29]	15.5617	0.3451	0.0374	0.2511
Serial GAN [32]	16.3529	0.3876	0.0328	0.2649
Ours	**18.8145**	**0.4162**	**0.0315**	**0.1769**

3.6. Ablation Study

To prove the effectiveness of each part of Sar2color, we carried out the ablation experiments of each part of Sar2color on the PSNR, SSIM, and LPIPS metrics. In Table 3, the base model indicates that normal residual block is used in the DCTRB part of the original encoder, and the CCMB module of Light-ASPP and decoder is removed at the same time, and the others add different components based on the base model. Compared with the base model, it can be seen that after adding each module separately, the added DCTRB module has the greatest effect on the improvement of each metric, which shows that in SAR-to-optical transformation, the speckle noise in the SAR image has the greatest influence on the transformation ability of the model. At the same time, we can also see that the addition of the other two modules can also improve the base model, and the transformation effect of the model is the best after adding all the modules, which gives the best performance in all metrics.

Table 3. Ablation study of our Sar2color structure with three evaluation metrics on SEN1-2 dataset. The up-arrow ↑ means larger is better, while the down-arrow ↓ means smaller is better.

	PSNR↑	SSIM↑	LPIPS↓
Base	12.1417	0.2544	0.3659
Base + DCTRB	16.1325	0.3715	0.2434
Base + Light-ASPP	13.4263	0.2963	0.3187
Base + CCMB	12.8766	0.2731	0.3428
Base + All of Above	**18.8145**	**0.4162**	**0.1769**

For the effectiveness of the style loss L_s, we carried out the ablation experiment on the loss functions. In Table 4, Oral means that we only use adversarial loss L_c and smooth L1 loss L_{sL1}, the second line indicates that we have increased the perceptual loss L_{p_HD} used in Pix2pix-HD [18] based on Oral loss, the loss is to improve the natural vividness of the generated image, and the third line is to add the style loss based on Oral loss. From the experimental results of the three metrics in the table, adding style loss L_s to the Oral loss achieves the best results, which proves that the style loss can better eliminate the difference in texture details between the generated image and the real image.

Table 4. Ablation study of the loss functions with three evaluation metrics on SEN1-2 dataset. The up-arrow ↑ means larger is better, while the down-arrow ↓ means smaller is better.

	PSNR↑	SSIM↑	LPIPS↓
Oral	18.3427	0.4017	0.1936
Oral + L_{p_HD}	18.5641	0.4058	0.1842
Oral + L_s	**18.8145**	**0.4162**	**0.1769**

4. Discussion

The Sar2color model is proposed in this paper for the SAR-to-optical transformation task. Sar2color is based on the CGAN structure and consists of a generator and a discriminator. In the generator, we design the DCTRB, Light-ASPP and CCMB modules to

generate more realistic optical images, where the DCTRB and Light-ASPP modules can reduce the negative impact of coherent speckle noise and geometric distortion characteristics in SAR images to generate optical images, while the CCMB module has the ability to alleviate the impact of the lack of color information on the transformation effect of the SAR image. It makes the resulting image more consistent with visual cognition in terms of color accuracy and saturation. At present, the research on speckle noise and geometric distortion suppression in SAR images is almost carried out independently. Most of the research uses spatial filtering, non-local mean, and variational methods to suppress speckle noise directly, and uses feature fusion methods for reducing geometric distortion in image registration. As the network components of Sar2color, DCTRB and Light-ASPP designed by us can simultaneously handle speckle noise and geometric distortion of SAR images, which is the first time to regard SAR as a priori information, and start with these two imaging characteristics of the SAR image in the deep learning network structure design in SAR-to-optical task.

Furthermore, we analyze the influence of the DCTRB, Light-ASPP and CCMB modules, and also the loss function in our Sar2color model on the quality of the generated optical images. It can be seen from the results in Table 3 that the DCTRB module has the greatest improvement in the quality of the generated optical images. DCTRB module directly selects and retains the low-frequency coefficients and the medium and high-frequency coefficients with higher energy by a simple and straightforward strategy named coefficient discriminant, which shows a significant effect in reducing the influence of coherent speckle noise on the generated optical images. Comparing the improvement of Light-ASPP and CCMB modules for the quality of transformed images, it can be seen that the adding Light-ASPP improves the evaluation metrics more. This is because Light-ASPP is designed to alleviate the influence of geometric distortion of SAR image on the generated optical images, and the impact of geometric distortion processing is higher than that of CCMB on color enhancement. When designing CCMB, the main color values of Top5 in the training image are selected, because there are fewer color types in optical remote sensing image, and the Top5 color values in the image can represent its main color. This strategy also reduces the amount of calculation parameters. Compared with other methods, this paper uses style loss in the design of the loss function, which can reduce the texture difference between the generated optical image and the real optical image. From the local details of different methods marked by the red box in Figure 9, the restoration effect of local details in our model results is the best, while other methods have obvious texture differences compared with real optical images.

5. Conclusions

In this paper, an end-to-end SAR-to-optical transformation model called Sar2color is proposed. According to the imaging characteristics of SAR images, DCTRB and Light-ASPP module are designed in the generator to reduce the negative impact of coherent speckle noise and geometric distortion characteristics in SAR images to generate optical images. CCMB module is then introduced in the decoder to enhance the color accuracy of the transformed images. At the same time, we also add the style loss in the training stage, which improves the color saturation of the generated optical images. Through qualitative and quantitative comparison, compared with other mainstream models, Sar2color shows obvious metric advantages and visual effects on the SEN1-2 dataset. This model can provide a new general solution framework for the SAR-to-optical transformation task.

However, our method still has some shortcomings. The geometric distortion of SAR images varies greatly under different meteorological conditions, so it is difficult for a single Light-ASPP module to alleviate images with complex geometric distortions. In future research, we will consider the problem of SAR images transformation under complex geometry distortion by designing local spatial attention mechanism to extract effective information. Furthermore, we found that the current SAR-to-optical advanced methods use paired datasets, the transformation performance of SAR images in other fields will

be degraded. Cross-domain learning under multiple types of unpaired datasets will be a promising solution to get rid of the dependence of our Sar2color on paired data and further improve the application universality of the model.

Author Contributions: Conceptualization, Z.G. and H.G.; methodology, H.G.; software, H.G., X.L. and W.Z.; validation, H.G., X.L. and W.Z.; data curation, H.G.; visualization, X.L. and W.Z.; funding acquisition, Z.G. and Y.F.; writing—original draft preparation, H.G.; writing—review and editing, Z.G. and H.G.; supervision, Z.G., Y.W. and Y.F. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: We used publicly available datasets: SEN1-2, which is accessible from https://mediatum.ub.tum.de/1436631 (accessed on 1 November 2021). SEN1-2 is a dataset consisting of 282,384 pairs of corresponding synthetic aperture radar and optical image patches acquired by the Sentinel-1 and Sentinel-2 remote sensing satellites, respectively.

Conflicts of Interest: The each authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GAN	Generative adversarial network
CGAN	Conditional generative adversarial network
SAR	Synthetic aperture radar
DCTRB	DCT residual block
Light-ASPP	Light atrous spatial pyramid pooling
CCMB	Correct color memory block
PSNR	Peak signal-to-noise ratio
SSIM	Structural similarity index metric
MSE	Mean square error
LPIPS	Learned perceptual image patch similarity

References

1. Scarpa, G.; Gargiulo, M.; Mazza, A.; Gaetano, R. A CNN-based fusion method for feature extraction from sentinel data. *Remote Sens.* **2018**, *10*, 236. [CrossRef]
2. Lyu, H.; Lu, H.; Mou, L. Learning a transferable change rule from a recurrent neural network for land cover change detection. *Remote Sens.* **2016**, *8*, 506. [CrossRef]
3. Balz, T.; Liao, M. Building-damage detection using post-seismic high-resolution SAR satellite data. *Int. J. Remote Sens.* **2010**, *31*, 3369–3391. [CrossRef]
4. Singhroy, V.; Mattar, K.; Gray, A. Landslide characterisation in Canada using interferometric SAR and combined SAR and TM images. *Adv. Space Res.* **1998**, *21*, 465–476. [CrossRef]
5. Santangelo, M.; Cardinali, M.; Bucci, F.; Fiorucci, F.; Mondini, A.C. Exploring event landslide mapping using Sentinel-1 SAR backscatter products. *Geomorphology* **2022**, *397*, 108021. [CrossRef]
6. Zhang, T.; Zhang, X.; Shi, J.; Wei, S.; Wang, J.; Li, J.; Su, H.; Zhou, Y. Balance scene learning mechanism for offshore and inshore ship detection in SAR images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 1–5. [CrossRef]
7. Gao, J.; Yuan, Q.; Li, J.; Zhang, H.; Su, X. Cloud removal with fusion of high resolution optical and SAR images using generative adversarial networks. *Remote Sens.* **2020**, *12*, 191. [CrossRef]
8. Cigna, F.; Bateson, L.B.; Jordan, C.J.; Dashwood, C. Simulating SAR geometric distortions and predicting Persistent Scatterer densities for ERS-1/2 and ENVISAT C-band SAR and InSAR applications: Nationwide feasibility assessment to monitor the landmass of Great Britain with SAR imagery. *Remote Sens. Environ.* **2014**, *152*, 441–466. [CrossRef]
9. Maity, A.; Pattanaik, A.; Sagnika, S.; Pani, S. A comparative study on approaches to speckle noise reduction in images. In Proceedings of the 2015 International Conference on Computational Intelligence and Networks, Odisha, India, 12–13 January 2015; pp. 148–155.

10. Zhang, Q.; Liu, X.; Liu, M.; Zou, X.; Zhu, L.; Ruan, X. Comparative analysis of edge information and polarization on sar-to-optical translation based on conditional generative adversarial networks. *Remote Sens.* **2021**, *13*, 128. [CrossRef]

11. 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]

12. Kong, Y.; Hong, F.; Leung, H.; Peng, X. A Fusion Method of Optical Image and SAR Image Based on Dense-UGAN and Gram–Schmidt Transformation. *Remote Sens.* **2021**, *13*, 4274. [CrossRef]

13. Chen, Y.; Bruzzone, L. Self-supervised sar-optical data fusion of sentinel-1/-2 images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–11. [CrossRef]

14. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *arXiv* **2014**, arXiv:1406.2661.

15. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.

16. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.

17. 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.

18. Wang, T.C.; Liu, M.Y.; Zhu, J.Y.; Tao, A.; Kautz, J.; Catanzaro, B. High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8798–8807.

19. Cho, W.; Choi, S.; Park, D.K.; Shin, I.; Choo, J. Image-to-image translation via group-wise deep whitening-and-coloring transformation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 10639–10647.

20. Reed, S.; Akata, Z.; Yan, X.; Logeswaran, L.; Schiele, B.; Lee, H. Generative adversarial text to image synthesis. In Proceedings of the International Conference on Machine Learning (PMLR), New York, NY, USA, 19–24 June 2016; pp. 1060–1069.

21. Zhang, H.; Xu, T.; Li, H.; Zhang, S.; Wang, X.; Huang, X.; Metaxas, D.N. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5907–5915.

22. Bahng, H.; Yoo, S.; Cho, W.; Park, D.K.; Wu, Z.; Ma, X.; Choo, J. Coloring with words: Guiding image colorization through text-based palette generation. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 431–447.

23. Yoo, S.; Bahng, H.; Chung, S.; Lee, J.; Chang, J.; Choo, J. Coloring with limited data: Few-shot colorization via memory augmented networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 11283–11292.

24. Choi, Y.; Choi, M.; Kim, M.; Ha, J.W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8789–8797.

25. Pumarola, A.; Agudo, A.; Martinez, A.M.; Sanfeliu, A.; Moreno-Noguer, F. Ganimation: Anatomically-aware facial animation from a single image. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 818–833.

26. Merkle, N.; Auer, S.; Müller, 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]

27. Doi, K.; Sakurada, K.; Onishi, M.; Iwasaki, A. GAN-Based SAR-to-Optical Image Translation with Region Information. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 2069–2072.

28. Yu, T.; Zhang, J.; Zhou, J. Conditional GAN with Effective Attention for SAR-to-Optical Image Translation. In Proceedings of the 2021 3rd International Conference on Advances in Computer Technology, Information Science and Communication (CTISC), Shanghai, China, 23–25 April 2021; pp. 7–11.

29. Darbaghshahi, F.N.; Mohammadi, M.R.; Soryani, M. Cloud removal in remote sensing images using generative adversarial networks and SAR-to-optical image translation. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–9. [CrossRef]

30. Zuo, Z.; Li, Y. A SAR-to-Optical Image Translation Method Based on PIX2PIX. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 3026–3029.

31. Turnes, J.N.; Castro, J.D.B.; Torres, D.L.; Vega, P.J.S.; Feitosa, R.Q.; Happ, P.N. Atrous cgan for sar to optical image translation. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 1–5. [CrossRef]

32. Tan, D.; Liu, Y.; Li, G.; Yao, L.; Sun, S.; He, Y. Serial GANs: A Feature-Preserving Heterogeneous Remote Sensing Image Transformation Model. *Remote Sens.* **2021**, *13*, 3968. [CrossRef]

33. Schmitt, M.; Hughes, L.H.; Zhu, X.X. The SEN1-2 dataset for deep learning in SAR-optical data fusion. *arXiv* **2018**, arXiv:1807.01569.

34. Huynh-Thu, Q.; Ghanbari, M. Scope of validity of PSNR in image/video quality assessment. *Electron. Lett.* **2008**, *44*, 800–801. [CrossRef]

35. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.

36. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [CrossRef]

37. 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.

38. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.

39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

40. Rubel, O.S.; Lukin, V.V.; De Medeiros, F.S. Prediction of Despeckling Efficiency of DCT-based filters Applied to SAR Images. In Proceedings of the 2015 International Conference on Distributed Computing in Sensor Systems, Fortaleza, Brazil, 10–12 June 2015; pp. 159–168.

41. Meenakshi, K.; Swaraja, K.; Kora, P. A robust DCT-SVD based video watermarking using zigzag scanning. In *Soft Computing and Signal Processing*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 477–485.

42. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.

43. Guo, H.; Guo, Z.; Pan, Z.; Liu, X. Bilateral Res-Unet for Image Colorization with Limited Data via GANs. In Proceedings of the 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI), Washington, DC, USA, 1–3 November 2021; pp. 729–735.

44. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]

45. Peters, A.F.; Peters, P. *The Color Thief*; Albert Whitman and Company: Park Ridge, IL, USA, 2015.

46. Guo, G.; Wang, H.; Bell, D.; Bi, Y.; Greer, K. KNN model-based approach in classification. In *Proceedings of the OTM Confederated International Conferences On the Move to Meaningful Internet Systems*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 986–996.

47. Li, Y.; Chen, X.; Wu, F.; Zha, Z.J. Linestofacephoto: Face photo generation from lines with conditional self-attention generative adversarial networks. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 2323–2331.

48. Xian, W.; Sangkloy, P.; Agrawal, V.; Raj, A.; Lu, J.; Fang, C.; Yu, F.; Hays, J. Texturegan: Controlling deep image synthesis with texture patches. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8456–8465.