remote sensing

MDPI

Article

Edge-Preserving Convolutional Generative Adversarial Networks for SAR-to-Optical Image Translation

Jie Guo [1], Chengyu He [1], Mingjin Zhang [1,*], Yunsong Li [1], Xinbo Gao [1,2] and Bangyu Song [3]

[1] Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China; jguo@mail.xidian.edu.cn (J.G.); 20011210556@stu.xidian.edu.cn (C.H.); ysli@mail.xidian.edu.cn (Y.L.); xbgao@mail.xidian.edu.cn (X.G.)

[2] Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

[3] School of Integrated Circuit Science and Engineering, Beihang University, Beijing 100191, China; 18373216@buaa.edu.cn

* Correspondence: mjinzhang@xidian.edu.cn; Tel.: +86-188-2955-2213

Abstract: With the ability for all-day, all-weather acquisition, synthetic aperture radar (SAR) remote sensing is an important technique in modern Earth observation. However, the interpretation of SAR images is a highly challenging task, even for well-trained experts, due to the imaging principle of SAR images and the high-frequency speckle noise. Some image-to-image translation methods are used to convert SAR images into optical images that are closer to what we perceive through our eyes. There exist two weaknesses in these methods: (1) these methods are not designed for an SAR-to-optical translation task, thereby losing sight of the complexity of SAR images and the speckle noise. (2) The same convolution filters in a standard convolution layer are utilized for the whole feature maps, which ignore the details of SAR images in each window and generate images with unsatisfactory quality. In this paper, we propose an edge-preserving convolutional generative adversarial network (EPCGAN) to enhance the structure and aesthetics of the output image by leveraging the edge information of the SAR image and implementing content-adaptive convolution. The proposed edge-preserving convolution (EPC) decomposes the content of the convolution input into texture components and content components and then generates a content-adaptive kernel to modify standard convolutional filter weights for the content components. Based on the EPC, the EPCGAN is presented for SAR-to-optical image translation. It uses a gradient branch to assist in the recovery of structural image information. Experiments on the SEN1-2 dataset demonstrated that the proposed method can outperform other SAR-to-optical methods by recovering more structures and yielding a superior evaluation index.

Keywords: SAR-to-optical image translation; deep learning; generative adversarial networks; edge-preserving convolution

check for updates

Citation: 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. https://doi.org/10.3390/rs13183575

Academic Editor: Dusan Gleich

Received: 31 July 2021
Accepted: 1 September 2021
Published: 8 September 2021

1. Introduction

With the continuous development of remote sensing technology, optical remote sensing data and synthetic aperture radar (SAR) remote sensing data have been widely leveraged in disaster monitoring, environmental monitoring, resource exploration, and agricultural planning, etc. [1–4]. Optical remote sensing image data are more representative of what we can observe with the naked eye, which means that these data contain rich spectral information, but capture depends heavily on the clarity of the environment. Heavy clouds and bad weather seriously reduce the quality of optical remote sensing images, and light conditions limit observation times, resulting in limited use of optical remote sensing data [5]. By relying on the microwave band electromagnetic waves, SAR can work in all weather and all light conditions to obtain SAR remote sensing data. However, the interpretation of SAR images is a difficult task for people without professional training, which is not the case with optical remote sensing images. Firstly, there is usually a lot of speckle noise in SAR images, generated by coherent interference of radar echoes from target scatters, which makes effective information in SAR images difficult to obtain [6,7]. Secondly, while the SAR signal wavelengths (mm to cm) do not belong to the visible part of the electromagnetic spectrum, which is familiar to human eyes, the easily distinguishable features in optical images may appear similar in SAR images. Overall, the features in SAR images are difficult to distinguish. In Figure 1, the water, building and paddy fields, which can be well differentiated in optical images, are partly visible in the SAR image. The water can be distinguished by the shape, but the building and paddy fields cannot be identified according to the SAR image. Thirdly, SAR images inevitably involve geometric distortion due to their special imaging mechanism, as SAR images are acquired through the reflected electromagnetic wave signals that are received by moving sensors [8]. Furthermore, electromagnetic wave signals may be reflected several times before being received, which makes the features in an SAR image sparse and distorted, often not matching the physical structures in the real environment [9]. Therefore, the interpretation of SAR image features is still a difficult task, despite the continuous improvement of remote sensing technology.

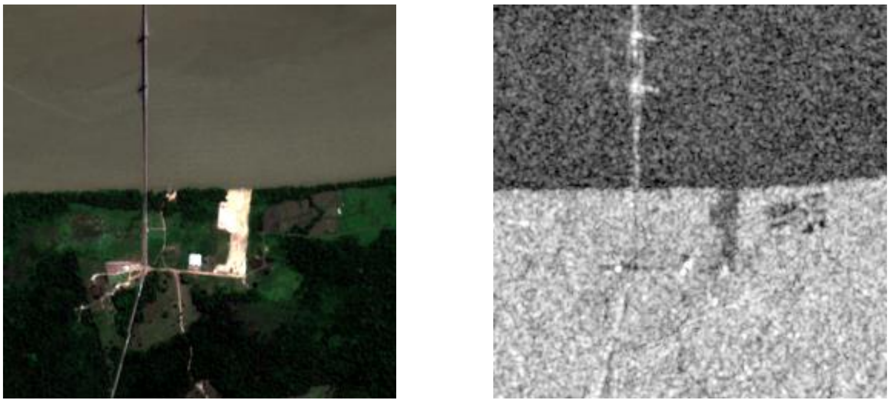


Figure 1. A comparison between SAR (**right**) and optical (**left**) images.

Due to the above points, it is necessary to leverage some technical means to increase the readability of SAR images. In the past few decades, some methods have been proposed to enhance the readability of SAR images based on the ideas of image enhancement and image colorization. SAR image enhancement aims to make the target in an SAR image more obvious through processing [10–12]. Odegard et al. [13] presented a method to reduce the speckle noise in SAR images based on wavelet transform, but it may cause an increase in the amount of natural clutter. An adaptive processing method was developed in [14], which combines with filtering, histogram truncation and equalization steps. An example application, the generation of a flood image, proved the validity of the method. SAR image colorization tries to make SAR images resemble optical imagery by encoding the pixels in the SAR images [15–17]. These methods are mainly for single-pol SAR images, as single-pol SAR images are single-channel images that are visually close to grayscale images. Image colorization is a process of entropy increase, which strongly depends on the establishment of the model; therefore, performance degradation may occur in actual use. The SAR images processed by the above method have improved visual features and perform better in feature detection and recognition. However, differently from optical remote sensing images, these processed images are only suitable for expert recognition and untrained people still cannot recognize the features in the images [18].

Deep learning is the field of machine learning; of handling complex tasks by building neural network models, which have developed rapidly with the improvements of computing ability in recent years [19]. Deep learning can be used to achieve image-to-image translation tasks, which are regression tasks [20–23]. Some methods for the SAR-to-optical image translation task have been presented. These convert the more readily available SAR images into optical images that are more compatible with human visual perception [5,18,24–29]. They are mainly based on generative adversarial networks (GAN), as GAN have the ability to produce images in line with real data distributions when there is a big difference between the SAR image and the optical image. These methods can generate grayscale or RGB optical images through SAR images by slightly adjusting the network for the image-to-image translation task. As these methods often do not take into account the special nature of image conversion and the network structure is not specifically made for SAR images, the optical images obtained often lose the structural information in SAR images and may contain conversion errors. Some work has been performed to improve them. A feature-guided method combined with a loss function based on discrete cosine transform (DCT) was developed in [30]. Zhang et al. [31] focused on the influences of edge information and polarization on the recovery process of SAR-to-optical image translation. However, the network structure is not specially designed for SAR-to-optical image translation. In addition, the whole feature maps are convolved with the same convolution filter, which is designed to reduce the parameters and complexity of neural networks in a standard convolution layer. However, the details and structural information of SAR images would be ignored while the content of each window is different but the filter is the same. It can also be understood that the parameters of the convolution kernel are globally optimal in an ideal situation, but are only sub-optimal for the contents in each window. This can degrade the quality of the generated image, especially in a difficult task such as SAR-to-optical image translation. Some methods try to predict convolutional filter weights at each pixel with a separate sub-network [32,33], but they increase the number of parameters, leading to more memory usage, longer training time and the corresponding marked dataset.

In this paper, we propose an edge-preserving convolutional generative adversarial network (EPCGAN) to enhance the structural information and visual clarity in the generated optical image. Inspired by decomposition theory utilized in traditional image enhancement methods and the pixel-adaptive convolution (PAC) [34], edge-preserving convolution (EPC) is proposed to perform content-adaptive convolution on feature maps while preserving the structural information. We first decompose the content of the feature map based on structural information extraction, and then perform content-adaptive convolution on the obtained content components, which combines the decomposition theory of traditional reinforcement methods with deep learning theory. The filter weight in the content-adaptive convolution is obtained by multiplying the weights of standard convolution kernel and the weights of edge-preserving kernel generated from the content component in each sliding window. Combined with the proposed EPC, EPCGAN, which has a gradient branch to assist the recovery of structural information, is proposed for the SAR-to-optical translation task. The gradient branch continuously receives the content information from the backbone network to simulate the gradient of a real optical image and finally feeds back the gradient information to the backbone network to assist in the image generation, which aims to make full use of the structural information in the SAR image for the SAR-to-optical image translation. In order to verify the effectiveness of our proposed edge-preserving convolution and edge-preserving convolutional generative adversarial networks, we conducted comparative experiments and ablation studies on the SEN1-2 dataset. Experimental results prove that our proposed method can obtain better visual properties with more defined texture and better evaluation indexes than other methods for SAR-to-optical image translation.

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

1. Edge-preserving convolution (EPC) is proposed for SAR-to-optical image translation. It performs content-adaptive convolution on a feature graph while preserving structural information according to decomposition theory, leading to good structure in the generated optical images.
2. For the situations in which SAR image interpretation is difficult, a novel edge-preserving convolutional generative adversarial network (EPCGAN) for SAR-to-

optical image translation is proposed, which can improve the quality of the structural information in the generated optical image by utilizing the gradient information of the SAR image and the optical image as a constraint.

3. The experiments on the training set selected from the SEN1-2 dataset [35] containing multi-modal data (forests, rivers, waters, plains, mountains, etc.) prove the superiority of the proposed algorithm. Meanwhile, ablation studies are given.

The organization of the remainder of the paper is as follows. Section 2 gives a comprehensive review of related methods. The proposed edge-preserving convolutional generative adversarial network for SAR-to-optical image translation is introduced in Section 3. We present the experiment results on the SEN1-2 dataset in Section 4 and comprehensive analyses in Section 5. Finally, the conclusions are illustrated in Section 6.

2. Related Works

2.1. Image-to-Image Translation

Image-to-image translation refers to the conversion of an image into another type of image, which has become one of top research topics in deep learning. Examples of translation include converting sketches to real pictures and realistic images to anime images [36–40]. Calculating the loss only through the content loss function, such as the L1-norm loss function or L2-norm loss function, will lead to the output having poor visual quality, which limit the results of the image-to-image translation task in the early stage. Generative adversarial networks have been widely applied in image-to-image translation, since the generator in GAN can generate images with excellent visual properties. The conditional generative adversarial network (cGAN) is a widely used framework for image-to-image translation tasks due to its ability to generate images based on not only content but also style [41]. Isola et al. presented a novel network named Pix2pix for image-to-image translation based on the cGAN framework, where the generator is based on U-Net [20]. Then, a high-resolution network Pix2pixHD was developed in line with Pix2pix, which can realize high-resolution image-to-image translation and semantic editing [22]. Pix2pix and Pix2pixHD has shown excellent conversion capabilities in sketch-to-real image conversion and style transfer experiments, but a large amount of paired data from different domains is needed, which is sometimes hard to acquire. Based on the ideas of symmetry and circulation, the networks named CycleGAN and DualGAN were proposed, which can utilize unpaired datasets for training [21,23]. Both Pix2pix and CycleGAN aim for one-to-one conversion, that is, the conversion from one domain to another domain. When there are multi-domain images that need to be converted, it takes a long time to retrain a model for each domain translation. Choi et al. presented a network named StarGAN, which can realize multi-domain image translation and only requires one training period [42]. Some methods also try to control some features in the output image through encoded variables [43]. A lightweight network for image-to-image translation was also proposed [44,45]. SAR-to-optical translation is also a part of image-to-image translation. However, there are huge differences between SAR images and optical images due to the datasets and speckle noise. Therefore, this particular case is indeed different than most image-to-image translation tasks. Unfortunately, when a network designed for “ordinary” image-to-image translation tasks is applied to SAR-to-optical translation, the outcome is poor. Therefore, our method for SAR-to-optical translation is meaningful.

2.2. Deep Learning-Based Methods for SAR Data

Deep learning has been used in SAR image optimization for different reasons. Based on the boundary equilibrium generative adversarial network (BEGAN) proposed in [46], a generative adversarial network for SAR image generation was developed, and it was demonstrated that synthetic data generated by the proposed network could improve the accuracy of classification [47]. Chierchia et al. [48] presented a deep learning-based method to remove the speckle noise in SAR images, and the network is based on the residual network, which is presented in [49]. The results came close to those of some state-of-the-art denoising methods for SAR images, which proves the potential of deep learning-based methods for SAR images. In order to enhance the quality of SAR images, the dialectical generative adversarial network (Dialectical GAN) was proposed to generate TerraSAR-X data with a ground-range resolution of 2.9 m and Sentinel-1 data based on a ground-range resolution of 20 m, which is similar to the effect of super-resolution in computer vision [50]. In addition, researchers also discussed the possibility of SAR-to-optical image translation to enhance the utilization of SAR images. Most solutions are based on the cGAN framework. Merkle et al. [25] proposed a method for optical and SAR image matching by converting single-pol SAR images to optical images with a U-net architecture and cGAN. Wang et al. [26] developed the SAR-GAN network consisting of two sub-networks to perform the despeckling task and coloring task, respectively; however, the two-step design idea ignores the different imaging principles of SAR images and optical images. Multi-temporal SAR data have also been considered, He et al. [51] developed a method that can generate optical images based on a meticulously designed residual network and cGAN. Some methods first convert SAR images into optical images and then fuse the SAR-to-image images with cloud images and SAR images to obtain cloud-free images, which contain RGB information [29,52] or hyperspectral information [28]. Schmitt et al. [35] published the SEN1-2 dataset, containing 282,384 pairs of corresponding image patches, which provides sufficient training data for the SAR-to-optical image translation task. cGAN requires strictly corresponding datasets, and the quality of datasets seriously affects the training results. Mario et al. [5] leveraged an unsupervised learning network CycleGAN [21] for SAR-to-optical image translation and discussed the fundamental limitations affecting SAR-to-optical image translation. Wang et al. [18] presented the supervised cycle-consistent adversarial network (S-CycleGAN) based on Pix2pix and CycleGAN to keep both the land cover and structural information. Furthermore, some methods that consider SAR image characteristics have been proposed. Zhang et al. [30] developed a feature-guided method with DCT loss, and Zhang et al. [31] utilized edge information to assist with SAR-to-optical image translation. However, these methods are usually simply modified versions of networks for general image-to-image translation that were not designed for SAR-to-optical image translation. In SAR-to-optical translation, we hope to recover an optical image with good lines. However, SAR images contain strong speckle noise, the edges of the image may be ignored in the standard convolution and the weight of the convolution kernel is content-independent, resulting in the output image having poor definition and blurred structural edges. Differently from the previously described methods, the proposed EPC and EPCGAN were designed for SAR-to-optical translation based on the characteristics of optical images and SAR images.

3. Methods

In this section, we first introduce the edge-preserving convolution. Then we present the details of edge-preserving convolutional generative adversarial networks and loss functions, accordingly.

3.1. Edge-Preserving Convolution

The convolutional neural network, a pioneering achievement, is described in [53]. It is one of the most widely used network structures in deep learning. The feature maps are convolved with a convolution kernel of specified size in a standard convolution layer. The standard convolution layer has far fewer parameters and far less of a computational load during training than fully connected layers, which effectively increases the depth of the neural network and decreases the difficulty of training. The weights of the convolutional layer are *spatially shared* but also *content insensitive*. Formally, the standard convolution from image features $\mathbf{X}$ with c channels to image features $\mathbf{X}'$ with c' channels can be written as:

$$\mathbf{X}'_{c'}(p) = b_{c'} + \sum_{p' \in \Phi(p)} \mathbf{W}\langle p' - p \rangle \times \mathbf{X}_c(p'), \tag{1}$$

where $\mathbf{W} \in \mathbb{R}^{c' \times c \times k \times k}$ are the weights of the convolution kernel, p are pixel coordinates in the image features, $\Phi(\cdot)$ is the range of $k \times k$ around the pixel coordinate of input and $b_{c'}$ denotes biases. With a slight abuse of notation, we use $\langle p' - p \rangle$ to denote the indexing of the spatial dimensions of an array with 2D spatial offsets. It can be seen from Equation (1) that the weight of the pixel multiplication in the convolutional layer is only related to the position. Once a convolutional neural network is trained, the same convolutional filter bank is applied to all images and all pixels, regardless of their content. Therefore, the structural information and details of the image are ignored, which limits the quality of the output image from the network.

To solve this limitation, we draw lessons from the traditional edge-sensing decomposition method for improving the convolution operation. Image decomposition techniques are widely used in traditional edge-aware image operators to achieve image enhancement [54–56], which is also used for the processing of SAR images [57,58]. Traditional decomposition methods can be summarized as:

$$\widehat{\mathbf{X}} = \mathcal{E}(\mathbf{X}), \tag{2}$$

$$\widetilde{\mathbf{X}} = \mathbf{X} - \widehat{\mathbf{X}}, \tag{3}$$

$$\mathbf{X}' = g\left(\widehat{\mathbf{X}}\right) + f\left(\widetilde{\mathbf{X}}\right), \tag{4}$$

where $\widehat{\mathbf{X}} = \mathcal{E}(\mathbf{X})$ is the content component; $\mathcal{E}(\cdot)$ is the operation of extracting content from an image, which is usually an edge-aware filter; $\widetilde{\mathbf{X}} = \mathbf{X} - \widehat{\mathbf{X}}$ is the texture component, which is the difference between image and content components; $g(\cdot)$ and $f(\cdot)$ are different processes for the content component and texture component, which can be referred to as a non-linear function. These traditional edge-aware decomposition methods leverage edge-aware filters to obtain the content component, which is usually considered to consist of the low-frequency components of the image, and the texture component, which is usually considered to consist of the high-frequency components of the image. Applying different modifications to content components will result in the changes in contrast and tone adjustments of image, and the image can be sharpened by enhancing the texture component.

While the gradient of the image is considered to contain the texture information of the image, we first extract the gradient of the image as the texture component in the image and keep the texture component unchanged, and then perform a convolution operation on the content component. Since the goal of the module we designed is not to change the number of channels in the feature map, the subsequent channels are unified to c. The standard convolution of a content component and of the processing of content components can be defined as:

$$\widehat{\mathbf{X}}'_c(p) = b_c + \sum_{p' \in \Phi(p)} \mathbf{W}\langle p' - p \rangle \times \left(\mathbf{X}_c(p) - \widetilde{\mathbf{X}}_c(p)\right) \tag{5}$$

Inspired by PAC algorithm in [34], an edge-preserving kernel $k\langle p' - p \rangle$ is proposed to modify the standard convolutional filter weights adaptively, according to the features in the content component. The edge-preserving kernel $k\langle p' - p \rangle$ is generated by the difference between the value of the $\mathbf{X}_c(p)$ and the surrounding pixel value, which provides the amplitude of the edge and small-scale detail. The edge-preserving kernel $k\langle p' - p \rangle$ can be written as:

$$k\langle p' - p \rangle = e^{-\frac{1}{2\sigma^2}\left(\frac{\mathbf{X}_c(p') - \mathbf{X}_c(p)}{v(p)}\right)^2} \tag{6}$$

Equation (6) is actually a modified Gaussian function. The σ is the standard deviation, which can control the degree of edge retention in the convolution. The $\alpha(p)$ is a regularization parameter added to limit the range of differences. The regularization $v(p)$ can be defined as:

$$v(p) = \max_{p' \in \langle p' - p \rangle}\left(\left\|\mathbf{X}_c(p') - \mathbf{X}_c(p)\right\|\right) \tag{7}$$

Combined with Equation (6), the processing of content components in edge-preserving convolution can be defined as:

$$\widehat{\mathbf{X}}'_c(p) = b_c + \sum_{p' \in \Phi(p)} k\langle p' - p \rangle \times \mathbf{W}\langle p' - p \rangle \times \left(\mathbf{X}_c(p) - \widetilde{\mathbf{X}}_c(p)\right) \tag{8}$$

The kernel $\mathbf{W}\langle p' - p \rangle$ is the same as the standard convolution kernel, whose purpose is to learn the corresponding relationship between the SAR image and the optical image through training; the edge-preserving kernel $k\langle p' - p \rangle$, which is generated from the content of each convolution window, can keep the edges in the content component by decreasing the influence of pixels with amplitudes that differ from that of the center pixel in Equation (8).

After the content component undergoes the operation in Equation (8), we merged the texture component, and the content component to obtain new features. Finally, the edge-preserving convolution (Figure 2) can be written as:

$$\mathbf{X}'_c(p) = \widetilde{\mathbf{X}_c}(p) + \left(b_c + \sum_{p' \in \Phi(p)} k\langle p' - p \rangle \times \mathbf{W}\langle p' - p \rangle \times \left(\mathbf{X}_c(p) - \widetilde{\mathbf{X}_c}(p)\right)\right) \tag{9}$$

This operation can be used in image restoration or translation, which preserves edges in each convolution and implements content-adaptive enhancement.

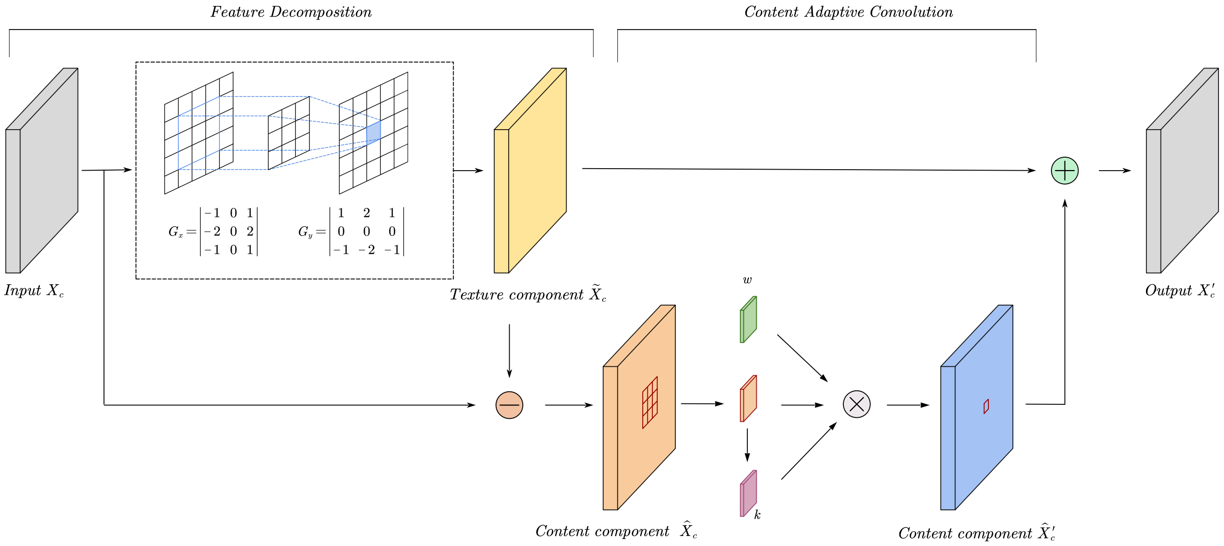

Feature Decomposition
Content Adaptive Convolution
$G_x = \begin{vmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{vmatrix}$
$G_y = \begin{vmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{vmatrix}$
Input X_c
Texture component $\widetilde{X}_c$
Output X'_c
w
k
Content component $\hat{X}_c$
Content component $\hat{X}'_c$

Figure 2. An overview of edge-preserving convolution. The gradient information is obtained via convolution, wherein the kernel weight is fixed as a Sobel operator and is regarded as the texture component. In fact, we can also extract other information from the image, such as the texture features and the curvature. EPC decomposes the input X_c into a content component and texture component based on the obtained texture component, and then convolution is performed on the content component $\widehat{X}_c$. The convolution kernel w is modified by the edge-preserving kernel k that is generated from each window in the content component $\widehat{X}_c$; ⊗ denotes the convolution operation.

3.2. *Edge-Preserving Convolutional Generative Adversarial Networks*

3.2.1. Network Framework

In addition to speckle noise, there always exist great differences between optical images and SAR images of the same scene, which are mainly due to their different imaging concepts [5]. The physical properties of the objects' surfaces will be highlighted in the SAR image, but the optical image provides more structural details; hence, the design of network is a problem that needs careful consideration. cGAN is an effective choice that can enhance the visual likeness of output image by GAN and the intensity constraint of the conversion process with the pixel loss between the output image and the target image. However, while some obvious features in the SAR image or optical image may not be obvious at all in the other, the loss of strong constraints would make the network unstable and produce blurry results with missing structural information of some objects. CycleGAN is another choice, which does not rely on the strong constraint loss function between the output image and the target image. CycleGAN can preserve the structural information well, but some land cover information is lost and translation errors may occur without a strong constraint loss function.

The edge-preserving convolutional generative adversarial network was designed based on the CycleGAN framework, but strong constraint loss between the input image and output image is added. In order to reduce the negative effect of strong constraint loss, we add some other losses to reduce the impact of strong constraint loss. In addition, our network also has a branch structure to make better use of the structural information in the input image. The overall framework of EPCGAN is shown in Figure 3.

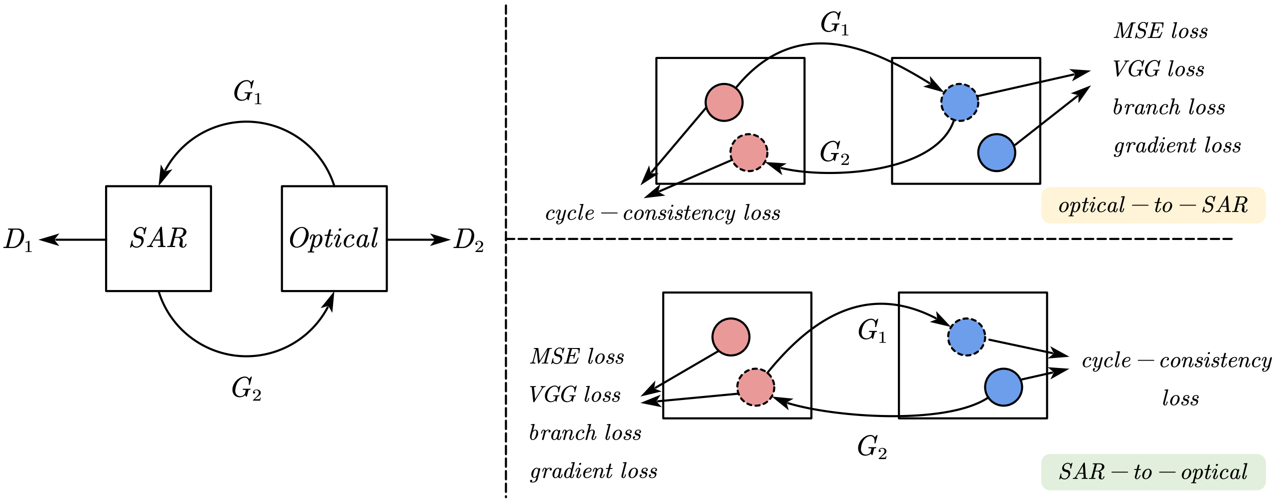


Figure 3. Our framework is based on the CycleGAN model. We added the loss of mean square error and other losses in the mapping between the SAR image and optical image to achieve better results.

3.2.2. Generator

Based on the proposed EPC, we designed a generator that contains a gradient branch. The backbone network utilizes the proposed EPC to extract features and merges the information provided by the gradient branch to output the converted image. The gradient part takes the gradient of the input image as the input, continuously integrates the auxiliary information provided by the backbone network, and finally feeds back to the backbone network for the final image reconstruction. Detailed information is shown in Figure 4.

The backbone network first leverages a 7 × 7 convolution and the proposed EPC, which can produce effective feature extraction of an image. After that, the size of the feature map is reduced through the convolution layer to reduce the network parameters, which has been proven to be effective in image-to-image translation [18]. We incorporate the feature maps from the 3th, 6th and 9th blocks into the gradient branch as auxiliary information and introduce the residual in the residual dense block (RRDB) proposed in [59] to fuse the feature map of the backbone network with the output of the gradient branch.

The goal of the gradient branch is to estimate the conversion of the gradient map between the SAR image and the optical image. The gradient branch first obtains the gradient map from the input image, just as the proposed EPC does. The gradient map can be obtained by calculating the differences between pixels, which can be realized by a convolutional layer with a fixed kernel. The acquisition of the gradient map can be expressed as:

$$\begin{aligned} F_x(z) &= F(x+1,y) - F(x-1,y); \\ F_y(z) &= F(x,y+1) - F(x,y-1); \\ \alpha(z) &= \left\| \left(F_x(z), F_y(z) \right) \right\|_2 \end{aligned} \tag{10}$$

where $\alpha(\cdot)$ stands for the operation to extract the gradient map, and $z = (x; y)$ are the coordinates in image F. The gradient branch will continuously combine the feature maps in the backbone network in order to restore the gradient map, an implicit reflection of whether the recovered regions should be sharp or smooth. In the generator, we provide the feature map generated by the penultimate layer of the gradient branch to the backbone network. At the same time, we apply these feature maps as input to generate the output gradient map through a 1 × 1 convolution layer.

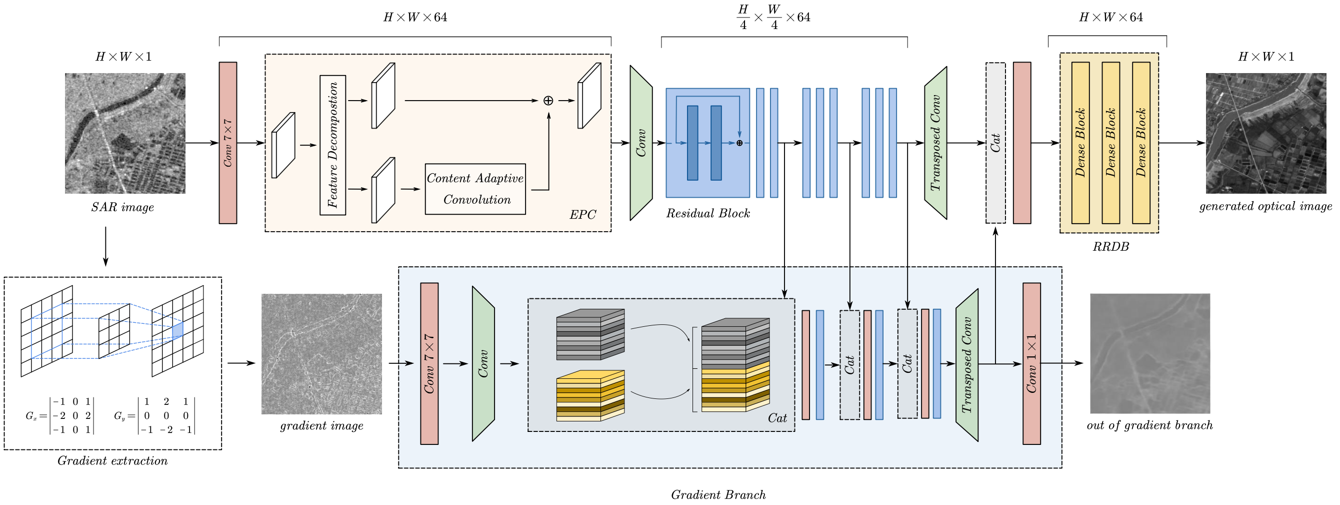


Figure 4. The generator of EPCGAN. The generator shown is for SAR-to-optical translation. The gradient branch aims to resolve the SAR image gradient map with its the optical image counterpart. It combines multi-level auxiliary information from the backbone network to reduce parameters and outputs gradient information to assist the generation of the final optical image.

3.2.3. Discriminator

The discriminator adopted the PatchGAN architecture, consisting of five convolutional layers, which has effective discrimination ability with fewer parameters [20]. Each convolutional layer is followed by a leaky ReLU, and a sigmoid output layer is set in the end for classification. The advantage of this method is that only the local image is discriminated, not the whole image, so that the image can be better judged better.

3.3. Loss Function

There exist two generators and two discriminators in our proposed EPCGAN to learn the translation between SAR image domain X and optical image domain Y with paired data samples $\{x_i\}_{i=1}^N \in X$ and $\{y_i\}_{i=1}^N \in Y$. The generator G_1 attempts to generate an image $G_1(x)$ that looks similar to the optical image based on the input SAR image x, and the discriminator D_y aims to distinguish real optical image y and generated optical image $G_1(x)$. In the same way, generator G_2 generates an image $G_2(y)$ that looks similar to the SAR image from the input optical image y, and the discriminator D_x is designed to distinguish real SAR image x from generated SAR image $G_2(y)$. The adversarial losses are shown as below:

$$\mathcal{L}_{GAN}(G_1, D_Y) = \mathbb{E}_{y\sim p_{data}(y)}[\log D_Y(y)] + \mathbb{E}_{x\sim p_{data}(x)}[1 - D_Y(G_1(x))] \tag{11}$$

$$\mathcal{L}_{GAN}(G_2, D_X) \;=\; \mathbb{E}_{x\sim p_{data}(x)}[\log D_X(x)] + \mathbb{E}_{y\sim p_{data}(y)}[1 - D_X(G_2(y))] \tag{12}$$

Cycle consistency loss was proposed in CycleGAN [21]. For each input SAR image x, it is converted to $G_1(x)$ by the generator G_1, and then converted to $G_2(G_1(x))$ by the generator G_2. The input x is expected to be consistent with $G_2(G_1(x))$, and the cycle-consistency loss is as follows:

$$\mathcal{L}_{cyc}(G_1, G_2) \;=\; \mathbb{E}_{x\sim p_{data}(x)}\|G_2(G_1(x)) - x\|_1 + \mathbb{E}_{y\sim p_{data}(y)}\|G_1(G_2(y)) - y\|_1 \tag{13}$$

Most SAR-to-optical image translation methods optimize well-designed networks through common pixel loss, which can reduce the average pixel difference between generated optical images and real optical images, but may lead to fuzzy results with loss of structural information. We also leverage the loss function to accelerate convergence and improve SAR-to-optical performance. Since there are two generators in our network, the pixel loss can be expressed as:

$$\mathcal{L}_{pix}(G_1, G_2) \;=\; \mathbb{E}_{x,y\sim p_{data}(x,y)}\|G_1(x) - y\|_2 + \mathbb{E}_{y,x\sim p_{data}(y,x)}\|G_1(y) - x\|_2 \tag{14}$$

In order to improve the perceptual quality of the generated image, the concept of perceptual loss was proposed in [60]. The features containing semantic information are extracted from the pretrained VGG network. The Euclidean distances between the features of input images and generated ones are minimized in perceptual loss:

$$\mathcal{L}_{per}(G_1, G_2) \;=\; \mathbb{E}_{x,y\sim p_{data}(x,y)}\|\varphi_i(G_1(x)) - \varphi_i(y)\|_1 + \mathbb{E}_{y,x\sim p_{data}(y,x)}\|\varphi_i(G_1(y)) - \varphi_i(x)\|_1 \tag{15}$$

where $\varphi_i(\cdot)$ denotes the ith layer output of the pretrained VGG model.

If the model is only optimized by L1 loss or MSE loss in the image space, we usually obtain images with blurry edges given an input test sequence where the ground truth has sharp edges. In order to enhance the structural information of the generated optical image as much as possible, we propose a gradient loss that is calculated by the gradient of the generated image and the gradient of the target image as follows:

$$\mathcal{L}_{grad}(G_1, G_2) \;=\; \mathbb{E}_{x,y\sim p_{data}(x,y)}\|\alpha(G_1(x)) - \alpha(y)\|_1 + \mathbb{E}_{y,x\sim p_{data}(y,x)}\|\alpha(G_1(y)) - \alpha(x)\|_1 \tag{16}$$

where $\alpha(\cdot)$ denotes the operation of gradient extraction. In the proposed EPCGAN, the output of the generator includes the output $G_1(x)$ of the backbone network and output $G_{1branch}(x)$ of the gradient branch. The function of the gradient branch in the generator is to extract effective structural information according to the input image to assist with image translation. In order to restrict the function of the gradient branch, we utilize the distance between the output of the gradient branch and the gradient graph of the target image to constrain the updating of the gradient branch parameters:

$$\mathcal{L}_{branch}(G_1, G_2) = \mathbb{E}_{x,y\sim p_{data}(x,y)}\|G_{1branch}(x) - \alpha(y)\|_1 + \mathbb{E}_{y,x\sim p_{data}(y,x)}\|G_{2branch}(y) - \alpha(x)\|_1 \tag{17}$$

In summary, we have two discriminators D_X and D_Y, which are optimized with $\mathcal{L}_{GAN}(G_1, D_Y)$ and $\mathcal{L}_{GAN}(G_2, D_X)$. For the generator, $\mathcal{L}_{GAN}$ and $\mathcal{L}_{cyc}$ are used to improve the visual realism of the output image while maintaining the structures. The $\mathcal{L}_{pix}$ and $\mathcal{L}_{per}$ are to provide corresponding constraints based on the pixel distance between the generated image and the target image. Gradient loss and branch loss cooperate with each other to improve the structural information of the output image according to the pixel distance between the generated image and the target image. The overall objectives are defined as follows:

$$\begin{aligned} G_1, G_2, D_X, D_Y = \arg\min_{G_1,G_2}\max_{D_X,D_Y} \big(&\mathcal{L}_{GAN}(G_1, D_Y) + \mathcal{L}_{GAN}(G_2, D_X) + \lambda_{cyc}\mathcal{L}_{cyc}(G_1, G_2) \\ &+ \lambda_{pix}\mathcal{L}_{pix}(G_1, G_2) + \lambda_{per}\mathcal{L}_{per}(G_1, G_2) \\ &+\lambda_{grad}\mathcal{L}_{grad}(G_1, G_2) + \lambda_{branch}\mathcal{L}_{branch}(G_1, G_2)\big) \end{aligned} \tag{18}$$

where λ_{cyc}, λ_{pix}, λ_{per}, λ_{grad} and λ_{branch} denote the weight parameters of different losses.

4. Experiments

In order to prove the effectiveness of the proposed method, we conducted experiments with some methods for SAR-to-optical translation based on the same training set and test set, which were selected from the SEN1-2 dataset [35].

4.1. Implementation Details

4.1.1. Dataset

The selection of the experimental dataset is a very important issue when proving the robustness of any method. SEN1-2, containing 282,384 paired image blocks collected from across the globe and throughout all meteorological seasons, has been proven to be usable for SAR-to-optical translation tasks. These image blocks were obtained by medium-range clipping from multiple paired SAR and optical images, and the size of each image block is 256 × 256 pixels. A common method of selecting datasets is to take some image blocks from a picture as the training set and some other image blocks as the test set, under the condition of ensuring that there are no overlapping pixels in the two kinds of image blocks. When the paired data resources for the task are difficult to obtain, this method is indeed reasonable. However, there is always a high degree of similarity between image blocks that come from the same large picture. When the network is trained with image blocks from the same picture as the test set, the model will perform better on the test set than it should, and the robustness of the model cannot be reflected in such experimental results.

We selected 1551 pictures from SEN1-2 as the training set, which contained multiple terrain types, including forests, lakes, mountains, rivers, buildings, farmlands, roads, and bridges, etc. At the same time, we selected pictures to form four test sets, *Test*_1,*Test*_2, *Test*_3 and *Test*_4, to evaluate the model. *Test*_1 contained 289 image blocks with various terrains, which were used to evaluate the performance of the model. Some image blocks in *Test*_1 and some image blocks in the training set came from the same large pictures, which were collected by us and named *Test*_2. In addition, we also added *Test*_3 and *Test*_4, which contained 62 image blocks and 111 image blocks, respectively. Those two datasets show mountains and urban suburbs with complex layouts, and the image blocks in the two datasets were from the large pictures that did not participate in the training of the model; therefore, *Test*_3 and *Test*_4 were completely unseen datasets. They were used to prove the robustness of our method. Details of each dataset are tabulated in Table 1.

4.1.2. Training Details

We trained and tested EPCGAN and the other SAR-to-optical methods on the same dataset. For each model, we used the same preprocessing method, and random rotating and flipping were utilized to avoid overfitting. ADAM optimizer [61] with $\beta_1 = 0.5$, $\beta_2 = 0.999$ was used for the optimization of EPCGAN. In particular, the two generators in EPCGAN shared the Adam optimizer, and the two discriminators also shared another Adam optimizer. The EPCGAN was trained for 200 epochs at a batch size of 1 in the experiments. We set the learning rates to 2×10^{-4} for both generator and discriminator, and linearly reduced them to zero starting from epoch 100. As for the weight parameters of losses, the λ_{cyc} was set to 10 following the settings in [21], and λ_{pix}, λ_{per}, λ_{grad} and λ_{branch} were set to 10 to balance the impressions of different losses. All the experiments were implemented on PyTorch and trained on NVIDIA GTX 2080Ti GPUs.

Table 1. The information of images involved in the SAR-to-optical translation.

Dataset	Number	Scene Content
train	1551	bridge rivers road mountain forests town farmland
Test_1	289	bridge rivers road mountain forests town farmland
Test_2	45	bridge rivers road
Test_3	62	mountain road
Test_4	111	farmland town rivers road

4.2. Results and Analysis

To evaluate the proposed EPCGAN quantitatively, we applied peak signal-to-noise ratio (PSNR), mean square error (MSE) and structural similarity (SSIM) for comparison. MSE represents the average gap between corresponding pixels. In order to make the results easy to observe, we first reduced the image pixel value of (0–255) to (0–1) and then calculated the MSE. The smaller the MSE, the smaller the distortion. The PSNR was based on the MSE between corresponding pixels in the reconstructed optical image and the real optical image. The higher the PSNR, the smaller the distortion. While the PSNR treated each pixel equally, the score of PSNR often deviated from the visual quality acquired by human eyes. Considering the human visual system, we also used SSIM to evaluate the similarities in brightness, contrast and structure. The higher the SSIM, the smaller the distortion. It is worth noting that our framework has the ability to convert optical images into SAR images and convert SAR images into optical images. We only discuss the translation from SAR images to optical images here.

We compare the proposed method quantitatively with some methods for SAR-to-optical translation, including Pix2pix [20], CycleGAN [21] and S-CycleGAN [18]. Pix2pix and CycleGAN are well-known methods for image-to-image translation that have been proven to be feasible in SAR-to-optical translation in some previous works [5,24]. S-CycleGAN was proposed in [18] for SAR-to-optical translation, which combines pixel loss and cycle-consistency loss. The results of PSNR and SSIM values are presented in Table 2. In each row, the best results are highlighted in bold. We can see in all the testing datasets that the proposed EPCGAN achieved the best PSNR and SSIM performance. Pix2pix could obtain good performance in PSNR compared with other methods and achieved the second highest PSNR values on *Test*_3—second only to EPCGAN; however, the SSIM values acquired by Pix2pix were the lowest on all the testing datasets due to the L_1 loss used in training. The L_1 loss was calculated according to the difference between the pixels of the generated picture and the target image, which is similar to the calculation principle of MSE. Thus, Pix2pix is more like a PSNR-oriented SAR-to-optical method, with which it is easy to produce relatively fuzzy results with high PSNR values.

We also visually compare these SAR-to-optical methods. From Figure 5, we see that they have better structural information and visual effects than other methods. For the first image, EPCGAN successfully restored the road, which is vaguely reflected in the SAR image based on the input SAR image, indicating that our method is capable of capturing structural characteristics in SAR images. At the same time, the edges of the recovered port are more standardized than other methods, which proves that EPCGAN can effectively constrain the edges of the graphics in the generated image through the gradient branch. Making full use of the features and structural information in the input SAR image, the EPCGAN generate results with better texture in the second and fourth image and more natural and realistic results in the third image.

CycleGAN can generate images with good structural information, but unsatisfactory partial translation results usually appear in the images (such as the port in the first image, the building in the fourth image and the additional artifacts of the third image) due to the lack of pixel loss calculated based on the generated image and the target image in the training process. Pix2pix only uses L1 loss to update the network during the training process, which leads to disappointing visual effects when Pix2pix is applied for SAR-to-optical translation. We cannot distinguish the river in the result generated by Pix2pix in the fourth image as it includes a number of undesirable artifacts. The first image and fourth image were from *Test*_3 and *Test*_4. The image blocks were not from the large picture from which some blocks were chosen for the training set, which proves that Pix2pix may have insufficient robustness when applied to SAR-to-optical translation. S-CycleGAN can generate better results than Pix2pix and CycleGAN, but the structural information and edges in the generated pictures often do not respond well. The visual comparison proves that our proposed method can better utilize and maintain the structural information in the SAR image based on the gradient branch and the proposed EPC, which helps to generate optical images that are easier to detect and recognize.

Table 2. Image quality assessment (IQA) results of different methods. The best values for each quality index are shown in bold.

IQA	Dataset	Pix2pix	CycleGAN	S-CycleGAN	EPCGAN
PSNR	Test_1	17.0482	16.3082	17.9046	**19.3627**
	Test_2	22.1012	22.4319	23.2056	**23.8345**
	Test_3	16.2285	15.7547	16.1178	**17.4944**
	Test_4	15.9798	15.4854	16.0738	**17.0195**
MSE	Test_1	0.0318	0.0322	0.0222	**0.0151**
	Test_2	0.0069	0.0068	0.0057	**0.0047**
	Test_3	0.0240	0.0285	0.0268	**0.0197**
	Test_4	0.0296	0.0351	0.0272	**0.0228**
SSIM	Test_1	0.3481	0.3424	0.4107	**0.4771**
	Test_2	0.4840	0.5331	0.5547	**0.5799**
	Test_3	0.2833	0.3140	0.2998	**0.3827**
	Test_4	0.2658	0.2944	0.2799	**0.3399**

4.3. A comparison of Textural and Structural Information

The gradient information of the image can well reflect the texture and structure of the image. In order to demonstrate the effectiveness of our method in image texture and structure restoration, we obtained the corresponding gradient map through the last images generated by different methods in Figure 5, and the results are shown in Figure 6. We can see that there are great differences in textural information between SAR image and optical images, and that speckle noise in the SAR image seriously pollutes texture information. Pix2pix had poor visual results on the unseen dataset. CycleGAN and S-CycleGAN can reduce the influence of speckle noise, but the structures of roads and buildings cannot be restored well. The proposed EPCGAN created the image with the best textural and structural information. It is worth noting that there was also a gap between the textural information of images generated by EPCGAN and optical images, which was due to the lack of information contained in SAR images, and the reasons were discussed in our introduction. Higher resolution SAR data are expected to reduce the impacts of these factors.

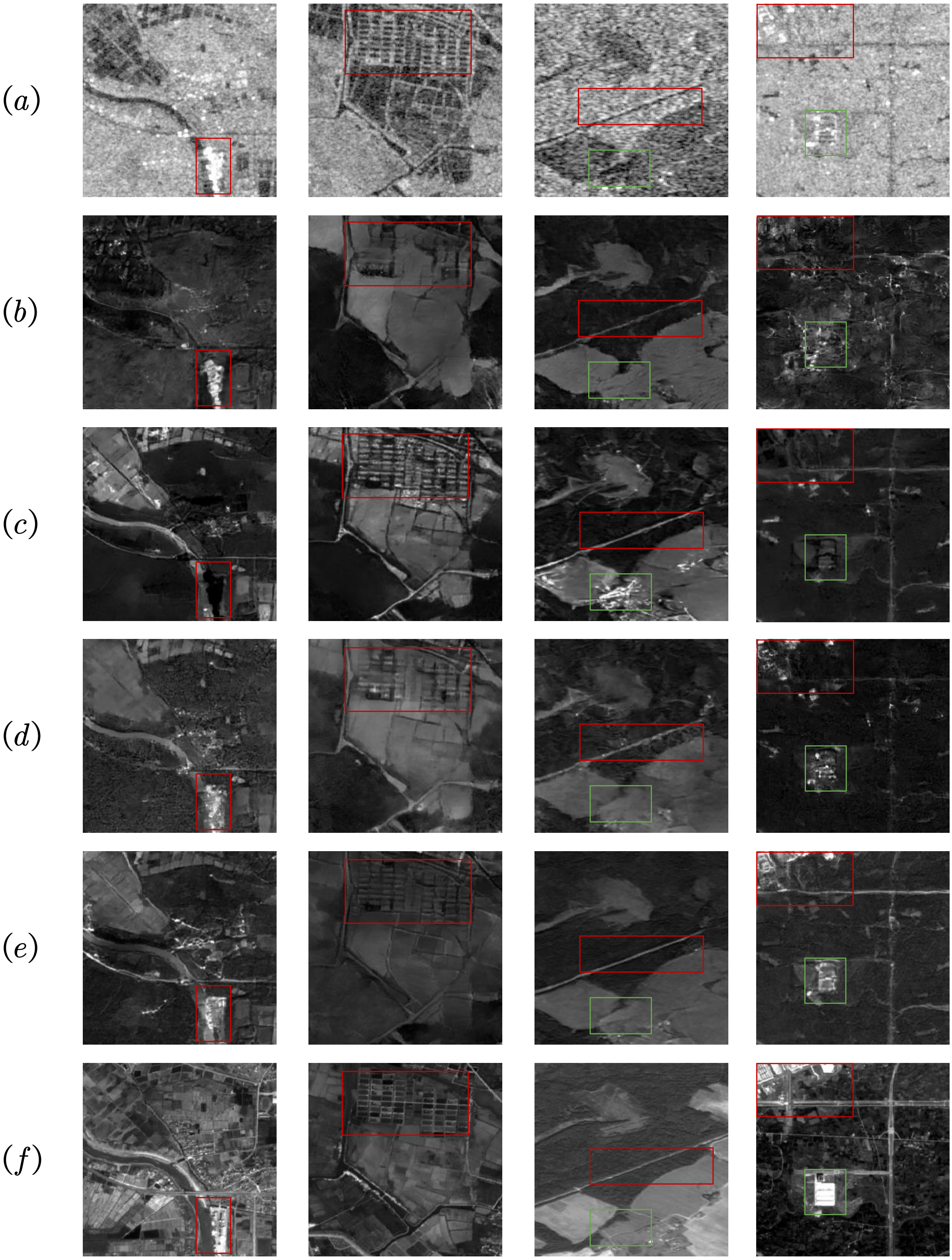
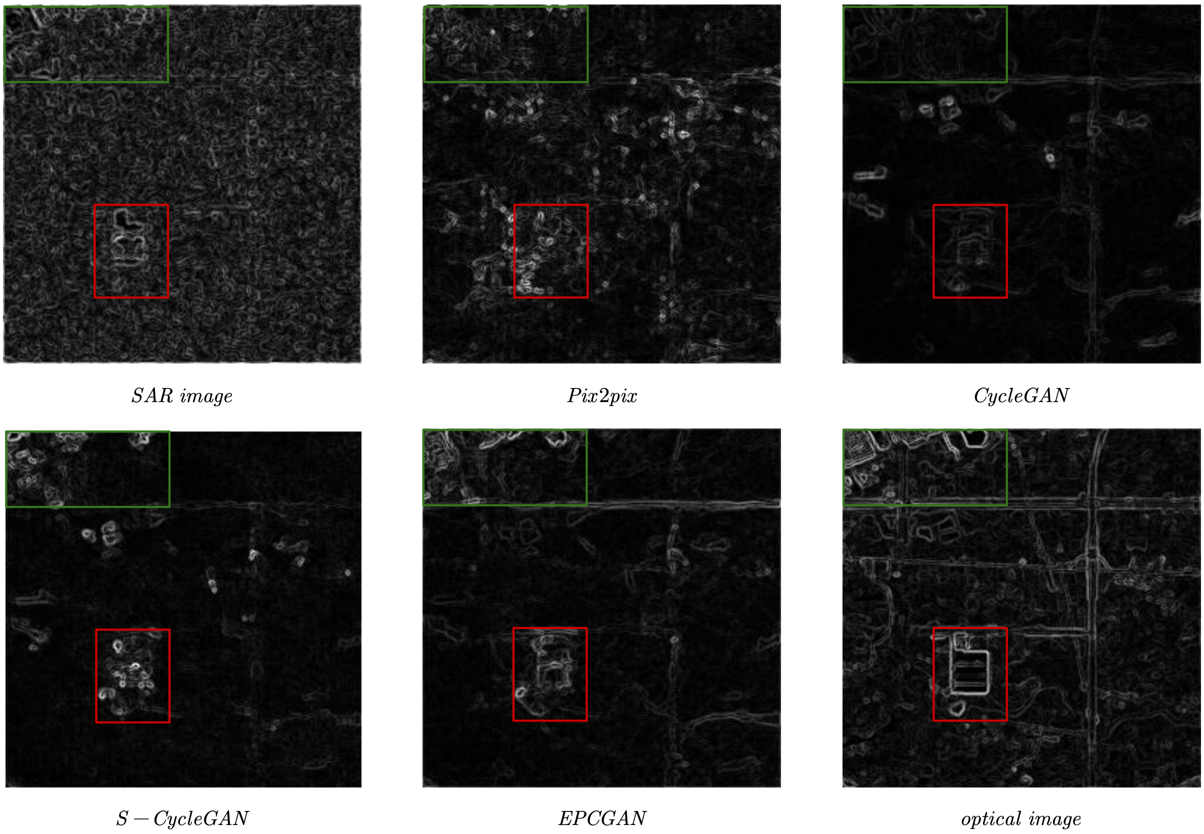


Figure 5. A visual comparison of different SAR-to-optical methods. The size of all images was 256 × 256. (**a**) SEN-1 SAR image. (**b**) Pix2pix. (**c**) CycleGAN. (**d**) S-CycleGAN. (**e**) EPCGAN. (**f**) SEN-2 optical image.



SAR image *Pix2pix* *CycleGAN*

$S - CycleGAN$ *EPCGAN* *optical image*

Figure 6. A gradient map of different SAR-to-optical methods. The size of all images is 256 × 256.

4.4. Model Complexity Analysis

In this section, the influence of the proposed EPCGAN on model complexity is studied. We summarize the parameters and floating-point operations (FLOPs) of the proposed EPCGAN and other methods for the SAR-to-optical translation compared in Section 4.2. Model parameter numbers refer to the numbers of parameters in the network that needed to be updated during training, which determined the neural network's demand on video memory. Generally, researchers hope to obtain better performance indicators with fewer parameters, whereas fewer parameters representing the model can be more easily deployed in industrial scenarios. FLOPs is the index that is used to measure the complexity of the model. Since the SAR-to-optical translation was realized by the generator after the network was trained, we only calculated the parameters and FLOPs of a single generator.

Figure 7 illustrates the PSNR values, SSIM values and parameter numbers of different methods on *Test*_1. Compared with the other methods, the proposed EPCGAN had a smaller model and better performance than them. It should be noted that the proposed EPC achieved edge-preserving and content-adaptive convolution without introducing extra parameters, whose parameters were equal to the convolutional layer with the same specifications.

Table 3 illustrates the training time and FLOPs of different methods. While achieving the best results with good structure and texture information, the FLOPs of the EPCGAN were higher, and more training time was required due to the use of RRDB and gradient branches. It is worth noting that EPC will introduce the calculation of multiple variables during back propagation, resulting in the extension of the network training time. We are considering optimizing this part in future work.

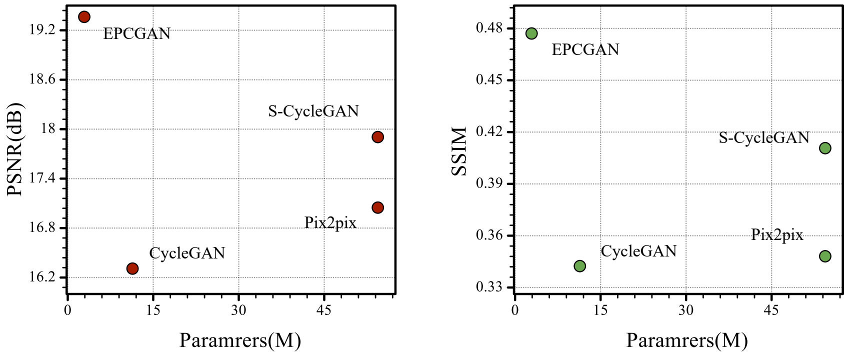


Figure 7. A comparison of the parameters of different SAR-to-optical methods.

Table 3. Training times and FLOPs of different methods.

	Pix2pix	**CycleGAN**	**S-CycleGAN**	**EPCGAN**
Training time (h)	3	9	12	31
FLOPs (G)	17.8	56.0	17.8	64.4

4.5. Ablation Experiment

In our method, the proposed EPC and gradient branch are used, both of which play unique roles. In order to prove the effectiveness of gradient branch and EPC, we did an ablation study to show the effects of the gradient branch and EPC. It should be noted that in EPCGAN (without EPC), we only delete the EPC in the network structure, and all loss functions are reserved for training. For EPCGAN (without gradient branching) we delete the gradient branch, and the gradient loss is removed during training. We trained on the same training set and tested on four test sets.

EPCGAN achieved the highest SSIM values on all test sets in Table 4, indicating that the complete method had better results in terms of structure and vision. Both EPC and gradient branch could effectively improve the quality of translated images, but due to the difficulty of the task, EPCGAN with EPC and gradient branch could only achieve less improvement than EPCGAN (without EPC) and EPCGAN (without gradient branch). We also performed a visual comparison in the ablation experiment. For the second image in Figure 8, without the EPC and gradient branch, the bridge in the generated optical image was translated into having small irregular bends, which is inconsistent with the real scene. The bridge in the image that was generated by the model without gradient branching is less obvious due to the lack of the overall gradient auxiliary information provided by the gradient branch. Additionally, without content-adaptive EPC, buildings that are not obvious enough in the gradient map will also not be obvious enough in the generated optical image. The complete EPCGAN generated the optical image with the best visual effects and good structure.

Table 4. Image quality assessment (IQA) results of ablation experiments. The best values for each quality index are shown in bold.

IQA	**Dataset**	**Ours (w/o EPC and Gradient Branch)**	**Ours (w/o EPC)**	**Ours (w/o Gradient Branch)**	**Ours**
SSIM	Test_1	0.4199	0.4647	0.4602	**0.4771**
	Test_2	0.4335	0.5195	0.5152	**0.5799**
	Test_3	0.3375	0.3783	0.3650	**0.3827**
	Test_4	0.3041	0.3362	0.3102	**0.3399**

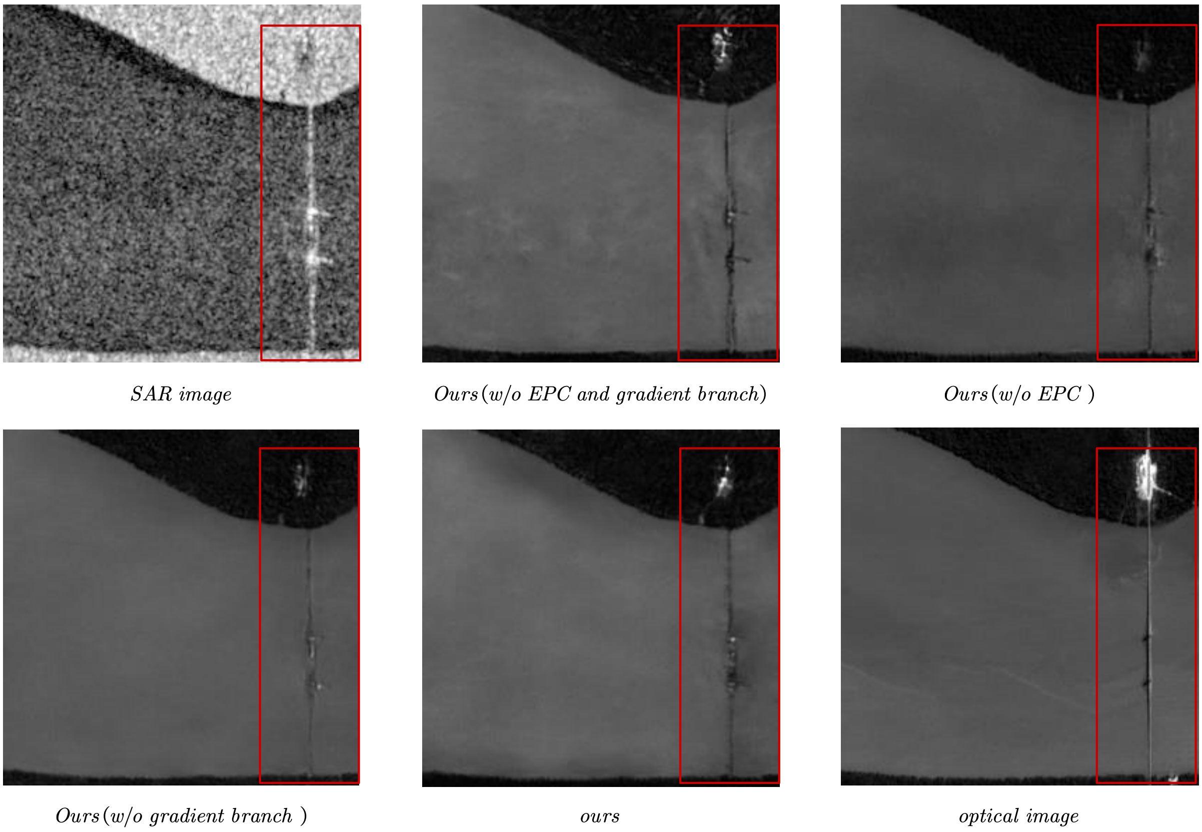

SAR image
Ours(w/o EPC and gradient branch)
Ours(w/o EPC)
Ours(w/o gradient branch)
ours
optical image

Figure 8. A visual comparison of ablation experiments. The size of all images is 256 × 256.

5. Discussion

5.1. Goals and Difficulties for SAR-to-Optical Translation

SAR-to-optical translation is a difficult task due to the huge differences between SAR images and optical images. In most image-to-image translation tasks that transform images belonging to one domain to another domain, the images between the two domains are different but often have a strong connection. For example, converting a character photo into an anime photo is a task of image-to-image translation, the contours of the characters provide a reliable basis for the generation of animation photos, and then CGANs can be utilized to generate visually realistic corresponding pictures. However, the SAR-to-optical translation is different for multiple reasons.

First of all, as we mentioned in Section 1, there exists a big gap between an SAR image and its optical image due to the imaging principle. Some features in the optical image will not be reflected in the SAR image. Accordingly, some obvious targets in the optical image may be consistent with the surrounding environment in the SAR image. The SAR image can provide detailed surface characteristics of the object, so the different coverage information will be obvious in the SAR image. However, the same kind of coverage information may have many different appearances in an optical image; for instance, deep water and shallow water often appear almost the same in an optical image, and optical images obtained in different weather conditions and lighting of the same place are very different. All these factors create great difficulties for SAR-to-optical translation.

Secondly, there is severe speckle noise and possible geometric distortion in SAR images. Speckle noise masks the real effective information to affect the feature extraction, and distortion seriously affects the translation, usually resulting in a distorted generated optical image.

At last, differently from other image-to-image translation tasks whose goal is to produce an overall visually realistic effect, we hope to obtain a sufficiently realistic optical image through SAR-to-optical translation. However, due to the different resolutions of remote sensing and the reasons mentioned above, it is difficult to recover optical images with excellent visual effects.

Based on the points we discussed above, we can understand that SAR-to-optical translation is a unique and difficult task. This leads to serious performance degradation when many image-to-image translation methods are directly applied to SAR-to-optical translation, and it is very necessary to design the network structure, loss function and preprocessing method according to the characteristics of the SAR image. In addition, because the optical image does not match the information in the SAR image, the goal of this conversion should be to use as much information in the SAR image as possible to output an optical image with a better structure. Our method was designed based on this.

5.2. Comparative Analysis of PAC and EPC

PAC is pixel adaptive convolution, proposed in [34], with excellent performance, which modifies the weight of the filter by the content in each window. However, the weight modification is not effectively restricted in PAC and the content may have an excessive influence on the weight of the filter. Therefore, the feature map used to influence the weight of the filter usually has a very small coefficient, which should be obtained through multiple adjustments, to avoid excessive influence in the experiment in [34]. At the same time, the coefficient may not be optimal for each image due to differences between images. We effectively restrict the weight modification through regularization in EPC. In addition, the structural information in SAR images may be lost or blurred when PAC modifies the filter weight, and our method can effectively retain the texture information of the SAR image. In order to prove the effectiveness of EPC, we conducted a comparative experiment in which the only difference between the EPCGAN and EPCGAN(PAC) was which model was chosen out of PAC and EPC.

Our method achieved the highest SSIM and PSNR values on all test sets (see Table 5), indicating that the EPC method had better results for the mean square error, structure and vision. We also provided a visual comparison in Figure 9. EPC can achieve clearer texture edges and better visual effects.

Table 5. Image quality assessment (IQA) results of EPCGAN(PAC) and EPCGAN. The best values for each quality index are shown in bold.

IQA	Dataset	EPCGAN (PAC)	EPCGAN
PSNR	Test_1	18.9468	**19.3627**
	Test_2	22.7652	**23.8345**
	Test_3	17.3758	**17.4944**
	Test_4	16.9336	**17.0195**
SSIM	Test_1	0.4575	**0.4771**
	Test_2	0.5272	**0.5799**
	Test_3	0.3631	**0.3827**
	Test_4	0.3389	**0.3399**

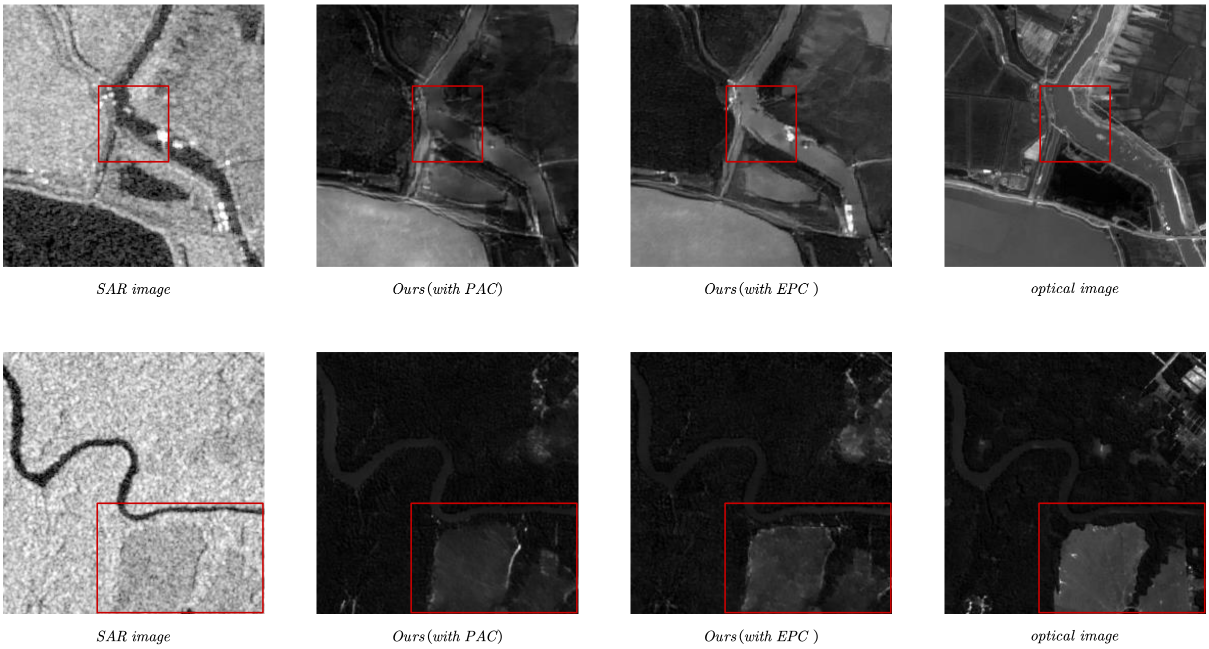


Figure 9. A visual comparison of EPCGAN(PAC) and EPCGAN.

5.3. Network Structure and Loss Function for SAR-to-Optical Translation

For the overall network framework, cGAN is the current optimal solution for SAR-to-optical translation, which leverages GAN to enhance the visual quality of the generated images. Nevertheless, the loss function needs to be designed according to the characteristics of SAR-to-optical translation. The generated image will be blurry with poor structure if only the pixel loss calculated based on the generated image and the target image is used in training. Perceived loss, DCT loss and some other losses will be effective options that first convert the output image and the target image before calculating the loss. In our method, multiple loss functions are used, which have different effects. In order to prove the correctness of our choice, we conducted an ablation study to show the effects of different loss functions. For the combination of multiple loss functions, when one loss was removed, its influence could clearly be reflected in the translation process.

An image quality assessment and a visual comparison are given in Table 6 and Figure 10. When the MSE loss is not used, we obtained poor PSNR value, and the translation error in CycleGAN would also appear, which proves that MSE loss can effectively constrain the translation process. It is worth noting that the MSE loss can be replaced with L1 loss; both of them are calculated based on the error between pixels. When the VGG loss is not used, the generated image is blurred, and the key target has unsatisfactory visual quality. It is worth noting that the images generated with our method had good structural edges due to the use of gradient loss. When the gradient loss was not used, we found that the edge of the port was very irregular and blurred, which proves that our gradient loss can help the recovery of the image edge. The rationality and effectiveness of our loss function can be proved by this phenomenon.

It is effective to extract additional information from SAR images to assist with generating optical images. The edge information of SAR images contains a lot of information, which is helpful for the generation of optical images. The proposed method provides auxiliary information for the reconstruction of the optical image according to the gradient map of the input image through the gradient branch, which is proved to be effective in the experiment. In addition, owing to the limited information possessed by single-channel SAR images, the use of multi-pol SAR images for image restoration is also a direction worth exploring.

Table 6. Image quality assessment (IQA) results of different loss functions. The best values for each quality index are shown in bold.

IQA	Dataset	EPCGAN (w/o MSE Loss)	EPCGAN (w/o VGG Loss)	EPCGAN (w/o Grad Loss)	EPCGAN
PSNR	Test_1	18.4377	18.6029	19.3625	**19.3627**
	Test_2	22.2648	20.7552	22.5868	**23.8345**
	Test_3	16.4162	17.4345	16.9904	**17.4944**
	Test_4	15.6180	16.9060	16.9023	**17.0195**
SSIM	Test_1	0.4454	0.4369	0.4594	**0.4771**
	Test_2	0.5174	0.4550	0.5587	**0.5799**
	Test_3	0.3427	0.3722	0.3669	**0.3827**
	Test_4	0.3357	0.3044	0.3362	**0.3399**

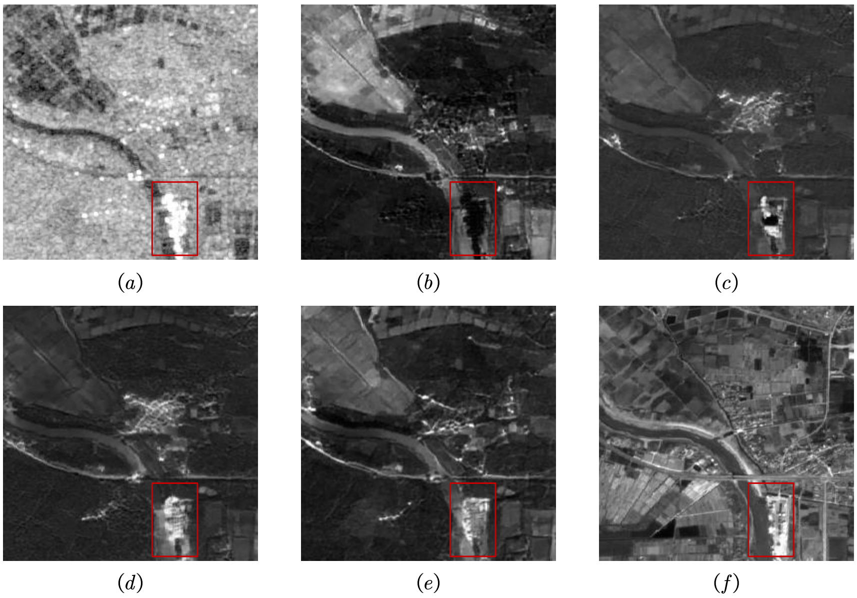


Figure 10. A visual comparison of different loss functions. (**a**) SEN-1 SAR image. (**b**) EPCGAN (without MSE loss). (**c**) EPCGAN (without VGG loss). (**d**) EPCGAN (without grad loss). (**e**) EPCGAN. (**f**) SEN-2 optical image.

6. Conclusions

In this paper, we summarized the difficulties and goals in SAR-to-optical translation based on the discussion of the characteristics of optical images and SAR images. After that, we proposed the EPCGAN and EPC for SAR-to-optical translation and conducted comparative experiments and ablation studies that demonstrated excellent performance of WDCNN against the other methods for SAR-to-optical translation. The trained standard convolution is content agnostic, which will cause the model to ignore some of the content features that we hope to be reflected in the generated optical image when facing complex SAR images. By combining traditional decomposition methods, we developed a novel EPC to perform content-adaptive convolution on SAR images while maintaining the texture characteristics in the SAR image. The EPC decomposes the content of convolution windows based on the texture component extracted by the edge extraction operator and achieves content-adaptive convolution by multiplying convolutional filter weights with an edge-preserving kernel generated from the content component in each window. Based on the proposed EPC, a new model EPCGAN was introduced for SAR-to-optical translation tasks. EPCGAN has two generators and two discriminators based on the CycleGAN framework, which can learn SAR-to-optical translation and optical-to-SAR translation at the same time. Since an SAR image contains very rich structural information, we designed the gradient branch in the generator of EPCGAN to leverage the edge information in an SAR image, which contains abundant useful information and basic features of the target structure. The introduction of edge information through the gradient branch and the proposed EPC effectively improve the structural quality of the generated image. The graphics in the generated image have clearer edges, and the generated image is more realistic and natural to our vision. At the same time, EPCGAN has excellent robustness that can handle SAR images with complex terrain, since EPC is content-adaptive. In addition, we discussed the impact of the loss function and the specific network structure on the SAR-to-optical translation. These findings provide an important reference for the design of a network structure and loss function in SAR-to-optical translation tasks. In addition, our scheme provides ideas for how to improve the structural information and visual quality of optical images and how to make full use of the complex information in SAR images. Since the design of EPCGAN is based on network structure, the proposed EPCGAN has the potential to be used in other tasks and become a general method for GAN. We will consider conducting experiments to explore the possibility of creating a general method of GAN and construct datasets to conduct detection experiments to prove the utility of our method in practical applications in the future.

Author Contributions: Conceptualization, J.G.; methodology, M.Z.; project administration, X.G.; software, C.H.; data curation, C.H. and B.S.; writing—review and editing, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 61902293; in part by the Key Research and Development Program of Shaanxi under Grant 2021GY-034; in part by the Young Talent Fund of University Association for Science and Technology, Shaanxi, China, under Grant 20200103; in part by the Fundamental Research Funds for the Central Universities under Grant XJS200112.

Data Availability Statement: The SEN1-2 dataset is downloaded free of charge from the library of the Technical University of Munich according to the link in [35].

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

References

1. Bazzi, H.; Baghdadi, N.; Amin, G.; Fayad, I.; Zribi, M.; Demarez, V.; Belhouchette, H. An Operational Framework for Mapping Irrigated Areas at Plot Scale Using Sentinel-1 and Sentinel-2 Data. *Remote Sens.* **2021**, *13*, 2584. [CrossRef]
2. Huang, L.; Yang, J.; Meng, J.; Zhang, J. Underwater Topography Detection and Analysis of the Qilianyu Islands in the South China Sea Based on GF-3 SAR Images. *Remote Sens.* **2021**, *13*, 76. [CrossRef]
3. Bayramov, E.; Buchroithner, M.; Kada, M.; Zhuniskenov, Y. Quantitative Assessment of Vertical and Horizontal Deformations Derived by 3D and 2D Decompositions of InSAR Line-of-Sight Measurements to Supplement Industry Surveillance Programs in the Tengiz Oilfield (Kazakhstan). *Remote Sens.* **2021**, *13*, 2579. [CrossRef]
4. Rajaneesh, A.; Logesh, N.; Vishnu, C.L.; Bouali, E.H.; Oommen, T.; Midhuna, V.; Sajinkumar, K.S. Monitoring and Mapping of Shallow Landslides in a Tropical Environment Using Persistent Scatterer Interferometry: A Case Study from the Western Ghats, India. *Geomatics* **2021**, *1*, 3–17. [CrossRef]
5. Fuentes Reyes, M.; Auer, S.; Merkle, N.; Henry, C.; Schmitt, M. SAR-to-Optical Image Translation Based on Conditional Generative Adversarial Networks—Optimization, Opportunities and Limits. *Remote Sens.* **2019**, *11*, 2067. [CrossRef]
6. Lee, J.S. Speckle suppression and analysis for synthetic aperture radar image. *Opt. Eng.* **1986**, *25*, 255636. [CrossRef]
7. Simard, M.; DeGrandi, G.; Thomson, K.P.B.; Benie, G.B. Analysis of speckle noise contribution on wavelet decomposition of SAR images. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1953–1962. [CrossRef]
8. Argenti, F.; Lapini, A.; Bianchi, T.; Alparone, L. A tutorial on speckle reduction in synthetic aperture radar images. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–35. [CrossRef]
9. Auer, S.; Hinz, S.; Bamler, R. Ray-Tracing Simulation Techniques for Understanding High-Resolution SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 1445–1456. [CrossRef]
10. Chambenoit, Y.; Classeau, N.; Trouvé, E. Performance assessment of multitemporal SAR images' visual interpretation. In Proceedings of the 2003 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Toulouse, France, 21–25 July 2003; pp. 3911–3913.
11. Zhang, B.; Wang, C.; Zhang, H.; Wu, F. An adaptive two-scale enhancement method to visualize man-made objects in very high resolution SAR images. *Remote Sens. Lett.* **2015**, *6*, 725–734. [CrossRef]
12. Li, Y.; Gong, H.; Feng, D.; Zhang, Y. An adaptive method of speckle reduction and feature enhancement for SAR images based on curvelet transform and particle swarm optimization. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3105–3116. [CrossRef]

13. Odegard, J.E.; Guo, H.; Lang, M.; Burrus, C.S.; Hiett, M. Wavelet Based SAR Speckle Reduction and Image Compression. In Proceedings of the SPIE—The International Society for Optical Engineering, San Diego, CA, USA, 9 July 1995; p. 2487.
14. Dellepiane, S.G.; Angiati, E. A new method for cross-normalization and multitemporal visualization of SAR images for the detection of flooded areas. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 2765–2779. [CrossRef]
15. Zhou, X.; Zhang, C.; Li, S. A perceptive uniform pseudo-color coding method of SAR images. In Proceedings of the 2006 CIE International Conference on Radar, Shanghai, China, 16–19 October 2006; pp. 1–4.
16. Uhlmann, S.; Kiranyaz, S.; Gabbouj, M. Polarimetric SAR classification using visual color features extracted over pseudo color images. In Proceedings of the 2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Melbourne, Australia, 21–26 July 2013; pp. 1999–2002.
17. Chen, Y.X.; Wu, W.B. Pseudo-color Coding of SAR Images Based on Roberts Gradient and HIS Color Space. *Geomat. Spat. Inf. Technol.* **2017**, *4*, 85–88.
18. 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]
19. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [CrossRef] [PubMed]
20. 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 2017(CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
21. 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 2017(ICCV), Venice, Italy, 22–29 October 2017; pp. 2223–2232.
22. 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 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 8798–8807.
23. Yi, Z.; Zhang, H.; Tan, P.; Gong, M. Dualgan: Unsupervised dual learning for image-to-image translation. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2849–2857.
24. Merkle, N.; Fischer, P.; Auer, S.; Müller, R. On the possibility of conditional adversarial networks for multi-sensor image matching. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Forth Worth, TX, USA, 23–28 July 2017; pp. 2633–2636.
25. Merkle, N.; Auer, S.; Muller, R.; Reinartz, P.; Pu, F. 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]
26. Wang, P.; Patel, V.M. Generating high quality visible images from SAR images using CNNs. In Proceedings of the IEEE Radar Conference 2018(RadarConf18), Oklahoma City, OK, USA, 23–27 April 2018; pp. 0570–0575.
27. Ley, A.; Dhondt, O.; Valade, S.; Haensch, R.; Hellwich, O. Exploiting GAN-based SAR to optical image transcoding for improved classification via deep learning. In Proceedings of the European Conference on Synthetic Aperture Radar 2018 (EUSAR), Aachen, Germany, 4–7 June 2018; pp. 1–6.
28. Grohnfeldt, C.; Schmitt, M.; Zhu, X. A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from Sentinel-2 images. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spian, 22–27 July 2018; pp. 1726–1729.
29. 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]
30. Zhang, J.; Zhou, J.; Lu, X. Feature-Guided SAR-to-Optical Image T ranslation. *IEEE Access* **2020**, *8*, 70925–70937. [CrossRef]
31. 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]
32. Xue, T.; Wu, J.; Bouman, K.L.; Freeman, W.T. Visual dynamics: Probabilistic future frame synthesis via cross convolutional networks. In Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS), Barcelona, Spain, 5–10 December 2016; pp. 91–99.
33. Bako, S.; Vogels, T.; McWilliams, B.; Meyer, M.; Novák, J.; Harvill, A.; Rousselle, F. Kernel-predicting convolutional networks for denoising Monte Carlo renderings. *ACM Trans. Graph.* **2017**, *36*, 97:1–97:14. [CrossRef]
34. Su, H.; Jampani, V.; Sun, D.; Gallo, O.; Learned-Miller, E.; Kautz, J. Pixel-adaptive convolutional neural networks. In Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 11166–11175.
35. Schmitt, M.; Hughes, L.H.; Zhu, X.X. The SEN1-2 dataset for deep learning in SAR-optical data fusion. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *4*, 141–146. [CrossRef]
36. Zhang, M.; Li, J.; Wang, W.; Gao, X. Compositional Model-Based Sketch Generator in Facial Entertainment. *IEEE Trans. Cybern.* **2018**, *48*, 904–915. [CrossRef] [PubMed]
37. Zhang, M.; Wang, N.; Li, Y.; Gao, X. Deep Latent Low-Rank Representation for Face Sketch Synthesis. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3109–3123. [CrossRef] [PubMed]
38. Zhang, M.; Wang, N.; Li, Y.; Gao, X. Neural Probabilistic Graphical Model for Face Sketch Synthesis. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 2623–2637. [CrossRef]
39. Zhang, M.; Wang, N.; Li, Y.; Gao, X. Bionic Face Sketch Generator. *IEEE Trans. Cybern.* **2020**, *50*, 2701–2714. [CrossRef]

40. Zhang, M.; Wang, R.; Li, J.; Gao, X.; Tao, D. Dual-transfer Face Sketch-Photo Synthesis. *IEEE Trans. Image Process.* **2019**, *28*, 642–657. [CrossRef] [PubMed]
41. Zhang, J.; Zhou, J.; Li, M.; Zhou, H.; Yu, T. Quality Assessment of SAR-to-Optical Image Translation. *Remote Sens.* **2020**, *12*, 3472. [CrossRef]
42. 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 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 8789–8797.
43. Li, X.; Zhang, S.; Hu, J.; Cao, L.; Hong, X.; Mao, X.; Ji, R. Image-to-image Translation via Hierarchical Style Disentanglement. In Proceedings of the 2021 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Virtually, 19–25 June 2021; pp. 8639–8648.
44. Chen, R.; Huang, W.; Huang, B.; Sun, F.; Fang, B. Reusing discriminators for encoding: Towards unsupervised image-to-image translation. In Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 8168–8177.
45. Liu, M.Y.; Huang, X.; Mallya, A.; Karras, T.; Aila, T.; Lehtinen, J.; Kautz, J. Few-shot unsupervised image-to-image translation. In Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 10551–10560.
46. Berthelot, D.; Schumm, T.; Metz, L. Began: Boundary equilibrium generative adversarial networks. *arXiv* **2017**, arXiv:1703.10717.
47. Marmanis, D.; Yao, W.; Adam, F.; Datcu, M.; Reinartz, P.; Schindler, K.; Stilla, U. Artificial generation of big data for improving image classification: A generative adversarial network approach on SAR data. In Proceedings of the 2017 conference on Big Data from Space (BiDS), Toulouse, France, 28–30 November 2017; pp. 293–296.
48. Chierchia, G.; Cozzolino, D.; Poggi, G.; Verdoliva, L. SAR Image Despeckling Through Convolutional Neural Networks. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 5438–5441.
49. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
50. Ao, D.; Dumitru, C.O.; Schwarz, G.; Datcu, M. Dialectical GAN for SAR Image Translation: From Sentinel-1 to TerraSAR-X. *Remote Sens.* **2018**, *10*, 1597. [CrossRef]
51. He, W.; Yokoya, N. Multi-Temporal Sentinel-1 and -2 Data Fusion for Optical Image Simulation. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 389. [CrossRef]
52. Bermudez, J.; Happ, P.; Oliveira, D.; Feitosa, R. SAR to optical image synthesis for cloud removal with generative adversarial networks. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *4*, 1. [CrossRef]
53. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
54. Durand, F.; Dorsey, J. Fast bilateral filtering for the display of high-dynamic-range images. In Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, San Antonio, TX, USA, 23–26 July 2002; pp. 257–266.
55. Paris, S.; Hasinoff, S.W.; Kautz, J. Local Laplacian filters: Edge-aware image processing with a Laplacian pyramid. *ACM Trans. Graph.* **2011**, *20*, 68.
56. Bovik, A.C. Nonlinear filtering for image analysis and enhancement. In *The Essential Guide to Image Processing*; Academic Press: New York, NY, USA, 2000; pp. 263–291.
57. Jing, W.; Jin, T.; Xiang, D. Edge-Aware superpixel generation for SAR imagery with one iteration merging. *IEEE Geosci. Remote Sens. Lett.* **2020**, *99*, 1–5. [CrossRef]
58. Choi, H.; Jeong, J. Speckle Noise Reduction Technique for SAR Images Using Statistical Characteristics of Speckle Noise and Discrete Wavelet Transform. *Remote Sens.* **2019**, *11*, 1184. [CrossRef]
59. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*; Stefan, R., Laura, L., Eds.; Springer: Cham, Switzerland, 2018; pp. 63–79.
60. Johnson, J.; Alahi, A.; Li, F. Perceptual losses for real-time style transfer and super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*; Bastian, L., Jiri, M., Nicu, S., Max, W., Eds.; Springer: Cham, Switzerland, 2016; pp. 694–711.
61. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.