SAR-to-Optical Image Translation via Neural Partial Differential Equations

Mingjin Zhang[1,3], **Chengyu He**[1*], **Jing Zhang**[2], **Yuxiang Yang**[4*], **Xiaoqi Peng**[1] and **Jie Guo**[1]
[1]State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xi'an 710071, China
[2]The University of Sydney, NSW 2006, Australia
[3]JD Explore Academy, China
[4]Hangzhou Dianzi University, Hangzhou 310018, China
{mjinzhang, 20011210556}@xidian.edu.cn, jing.zhang1@sydney.edu.au, yyx@hdu.edu.cn, xqpengxidian@163.com, jguo@mail.xidian.edu.cn

Abstract

Synthetic Aperture Radar (SAR) becomes prevailing in remote sensing while SAR images are challenging to interpret by human visual perception due to the active imaging mechanism and speckle noise. Recent researches on SAR-to-optical image translation provide a promising solution and have attracted increasing attention, though still suffering from low optical image quality with geometric distortion due to the large domain gap. In this paper, we mitigate this issue from a novel perspective, i.e., neural partial differential equations (PDE). First, based on the efficient numerical scheme for solving PDE, i.e., *Taylor Central Difference (TCD)*, we devise a basic TCD residual block to build the backbone network, which promotes the extraction of useful information in SAR images by aggregating and enhancing features from different levels. Furthermore, inspired by the *Perona-Malik Diffusion (PMD)*, we devise a PMD neural module to implement feature diffusion through layers, aiming at removing the noise in smooth regions while preserving the geometric structures. Assembling them together, we get a new SAR-to-Optical image translation network named S2O-NPDE, which delivers optical images with finer structures and less noise. Experiments on the popular SEN1-2 dataset show that S2O-NPDE outperforms state-of-the-art methods in both objective metrics and visual quality.

1 Introduction

With all-day and all-weather acquisition capability, synthetic aperture radar (SAR) has been widely used in many applications including change detection, resource exploration, and agricultural planning [Zhang and Tao, 2020; Yang *et al.*, 2022]. Unlike optical remote sensing images, which are friendly to interpret by human visual perception, the interpretation of SAR images is a difficult task for non-experts due to its unique active imaging mechanism and the high-frequency speckle noise. SAR-to-optical (S2O) image translation, as a promising solution, aims to generate a corresponding optical image from the SAR image. Benefiting from the development in deep learning, S2O image translation via deep neural networks has gained increasing attention in recent years [Fu *et al.*, 2021; Grohnfeldt *et al.*, 2018], most of which are based on generative adversarial networks (GAN), such as conditional GAN (cGAN) [Isola *et al.*, 2017] and cycle-consistent GAN (CycleGAN) [Zhu *et al.*, 2017]. However, these GAN-based methods suffer from generating geometric distortions in the optical image with low quality due to the large domain gap between SAR images and optical images, which makes the S2O image translation task very challenging.

One of the challenges originates from the unique active imaging mechanism of SAR instead of the passive one of optical sensors, which results in different image domains. SAR images record the amount of energy reflected back after interacting with the target where the energy is produced by SAR sensors. Hence, SAR images emphasize the physical properties of the surface, while optical images provide more visible structural details [Fuentes Reyes *et al.*, 2019]. For example, the paddy fields and land are easy to distinguish in optical images while being similar in SAR images. Furthermore, the electromagnetic wave signals used to generate SAR images may be reflected several times before being received, which produces geometric distortions in SAR images, often not matching the physical structures in the real environment [Auer *et al.*, 2009]. The difference makes the network trained only with the content loss, e.g., L1-norm loss and L2-norm loss, generate blurry optical images. For example, the cGAN-based method can characterize the land cover information well but produces ambiguous results with less structural information when applied to complex scenes [Wang *et al.*, 2019]. Using the cycle consistency loss, the CycleGAN-based method can keep the structure information in SAR image, while tending to produce hallucinations and mismatched structures [Fuentes Reyes *et al.*, 2019]. Other methods try to address this issue through effective network designs. For example, Yang *et al.* [2022] proposed a parallel feature fusion generator to improve the feature extraction of SAR image while Guo *et al.* [2021a] utilized edge information to help generate optical images with better texture, which however

*Corresponding Author

are heuristic and lack of an explicit design principle.

Another challenge originates from the common speckle noise in SAR images. Speckle noise is generated by the coherent interference of signal responses from individual scatters in resolution cells, which obscures part of the effective information in SAR images [Simard *et al.*, 1998], making it difficult to extract effective features from SAR images. To handle the speckle noise, Wang *et al.* [2018] developed the SAR-GAN consisting of two sub-networks to perform the despeckling task and coloring task, respectively. However, the two-step design idea ignores the geometric distortion of SAR images and can only be used to process partially distortion-free SAR images. Zhang *et al.* [2020] proposed a feature-guided method that uses a discrete cosine transform (DCT)-based loss to suppress the influence of speckle noise, while it may also affect the structures in the generated image.

In this paper, we make the first attempt to address these issues from a novel perspective, i.e., neural partial differential equations (PDE), and propose a SAR-to-Optical image translation method named S2O-NPDE. Different from previous methods, our S2O-NPDE is devised following an explicit and explainable design principle derived from PDE. First, based on the efficient numerical scheme for solving PDE, i.e., *Taylor Central Difference (TCD)*, we devise a basic TCD residual block to build the backbone network, which promotes the extraction of useful information in SAR images by aggregating and enhancing features from different levels. Furthermore, inspired by the *Perona-Malik Diffusion (PMD)*, we devise a differentiable PMD neural module to implement feature diffusion through layers, aiming at removing the noise in smooth regions while preserving the geometric structures. It is achieved by constraining the output of the PMD neural module to be consistent with the high-frequency information of optical image via a PDE-based loss. Assembling them together, we obtain the S2O-NPDE model, which is skilled at better feature extraction, speckle noise removal, and geometric structure preservation by design. Experiments on the popular SEN1-2 dataset show that our S2O-NPDE outperforms state-of-the-art (SOTA) methods in terms of both objective metrics and visual quality, i.e., generating better optical images with finer structures and less noise.

The main contribution of this paper is three-fold:

- We address the S2O image translation task from the perspective of neural partial differential equations and accordingly propose a novel S2O-NPDE model following an explicit and explainable design principle. To the best of our knowledge, this is the first attempt and offers a new insight into the S2O image translation task.
- We propose two essential components and assemble them together as the S2O-NPDE model. One is derived from the efficient numerical scheme for solving PDE, i.e., the TCD residual block to build the backbone network, while the other is derived from the well-known Perona-Malik Diffusion model, i.e., the PMD neural module. They help S2O-NPDE extract better features for S2O image translation while removing noise and preserving geometric structures by design respectively.
- Our S2O-NPDE delivers optical images with finer structures and less noise and outperforms SOTA methods in terms of objective metrics and visual quality on the popular SEN1-2 dataset containing diverse scenes.

2 Related Work

2.1 SAR-to-Optical Image Translation

Most of S2O image translation methods are based on generative adversarial networks [Goodfellow *et al.*, 2014; Qiao *et al.*, 2019] that implicitly learn a conditional data distribution of optical images matching the real distribution and complete the task by sampling an optical image from the distribution given a SAR image [Yang *et al.*, 2022]. Many research efforts have been made in the design of loss functions and network structures. For example, an L1-norm content loss is used in the cGAN-based method [Grohnfeldt *et al.*, 2018] while the cycle consistency loss is used in the CycleGAN-based method [Fuentes Reyes *et al.*, 2019]. However, these methods suffer from blurry artifacts or geometric distortions due to the large domain gap between SAR images and optical images. Other methods try to address the issue by exploring the characteristics of SAR images and devising effective network structures accordingly, e.g., leveraging the residual network [He and Yokoya, 2018], utilizing the edge information [Guo *et al.*, 2021a], designing chromatic aberration loss [Yang *et al.*, 2022]. Although these methods could generate promising results, they are designed in a heuristic manner and lack of an explicit design principle. Our model is also based on the CycleGAN framework, but different from them, we propose to solve the S2O image translation task from the perspective of neural PDE for the first time. Specifically, we devise a novel S2O-NPDE model consisting of two neural PDE modules by following an explicit and explainable design principle derived from PDE.

2.2 Neural Partial Differential Equations

Many deep networks have been designed inspired by the PDE [Guo *et al.*, 2021b], e.g., learning discriminative and rotation-invariant features for image classification [Fang *et al.*, 2017] and learning the optimal nonlinear anisotropic diffusion for image denoising [Chen *et al.*, 2015]. Recently, instead of designing networks based on the complex derivation of PDE, researchers have discovered the interesting connection between neural networks and the ordinary differential equation (ODE) [Lu *et al.*, 2018; He *et al.*, 2019], which is a simplified form of the PDE. Most existing ODE-inspired networks are designed by solving the ODE with the Euler method due to its simplicity and ease of implementation, although the Taylor method can deliver better accuracy [Anderson and Wendt, 1995]. In this paper, we investigate the network design based on the Taylor Central Difference method and propose a novel residual block to promote the extraction of useful information in SAR images by aggregating and enhancing features from different levels.

Besides, as a well-known PDE, Perona-Malik Diffusion has been widely adopted in the image processing community, especially for image denoising [Perona and Malik, 1990], by designing new diffusion kernels [Keeling and Stollberger, 2002] to efficiently remove noise while preserving structures.

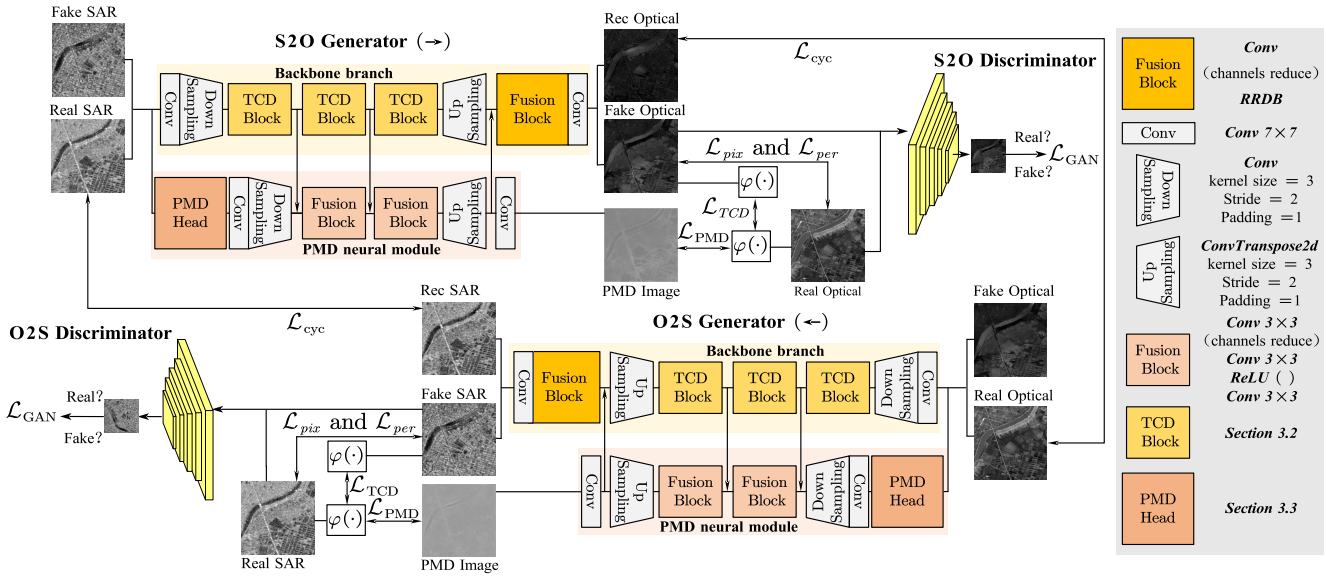

Figure 1: Overview of the proposed S2O-NPDE model, which is based on the CycleGAN framework. The generator includes a backbone branch of three TCD residual blocks (Section 3.2) and a parallel PMD neural module (Section 3.3), where feature interactions are encouraged between them. $\varphi(\cdot)$ denotes the mapping function learned by the PMD head. The adversarial loss $\mathcal{L}_{GAN}$, cycle consistency loss $\mathcal{L}_{cyc}$, pixel loss $\mathcal{L}_{pix}$, perceptual loss $\mathcal{L}_{per}$, and the proposed PDE losses, i.e., $\mathcal{L}_{TCD}$ and $\mathcal{L}_{PMD}$, are used to train S2O-NPDE (Section 3.4).

In this paper, we try to implement PMD in neural networks by devising a differentiable PMD neural module to specifically mitigate the influence of speckle noise in SAR images.

3 Proposed Method

3.1 Overall Architecture

We devise our S2O-NPDE model based on the CycleGAN framework [Zhu *et al.*, 2017]. As shown in Figure 1, the generator contains a backbone network of three cascaded TCD residual blocks (Section 3.2) and a parallel PMD neural module (Section 3.3). In the backbone, the SAR image first passes through a convolution layer and a downsampling layer, then goes through the three TCD residual blocks to aggregate the multi-level features. Besides, we devise a parallel PMD neural module to promote the generation of high-frequency information in the optical image. Specifically, the SAR image is first fed into the PMD head to eliminate the speckle noise while preserving the structure details. The output features are fused with the feature maps from TCD residual blocks in the fusion blocks for further refinement. After an upsampling layer and a convolutional prediction layer, a side output called PMD image is generated. The fused features are also fed into the fusion block in the backbone to promote feature interactions and guide the generator to produce a final optical image with fine details and less noise. Specific losses (Section 3.4) are used to supervise the generation of both the side output and the final optical image.

3.2 TCD Residual Block

Due to the large appearance discrepancy between SAR images and optical images, some detail informations such as lines will be lost during S2O image translation, making the generated optical image lack texture. To address this issue, we propose a specific residual block to promote multi-level feature fusion. In this way, fine detail features can be encoded and used for optical image generation. To this end, we implement the residual block as a solver of ODE based on the efficient numerical scheme TCD, i.e., we devise a basic TCD residual block to build the backbone network. We choose TCD instead of the Euler method due to its better accuracy in solving ODE [Anderson and Wendt, 1995] .

Specifically, we discretize the ODE using the Taylor central finite difference equations with third-order accuracy, i.e.,

$$\left(\frac{\partial u}{\partial x}\right)_j = \frac{2u_{j+1} - 3u_j + 2u_{j-1} - u_{j-2}}{2\Delta x}, \tag{1}$$

where u is a target depending on the input variable x. The above formula can be rewritten as:

$$\left(\frac{\partial u}{\partial x}\right)_j \Delta x = u_{j+1} - \frac{3}{2}u_j + u_{j-1} - \frac{1}{2}u_{j-2}. \tag{2}$$

We define the left side of the equation as the change of u from step $j-2$ to step $j+1$, which can be approximated with a neural module, denoted as $\delta f(x)$. Therefore, the above equation can be rewritten as:

$$u_{j+1} = u_j + \frac{1}{2}(u_j - u_{j-1}) - \frac{1}{2}(u_{j-1} - u_{j-2}) + \delta f(x). \tag{3}$$

Then, we use Δu_j and Δu_{j-1} to denote the residual between u_j and u_{j-1} and the residual between u_{j-1} and u_{j-2}, respectively, i.e., $\Delta u_j = u_j - u_{j-1}$ and $\Delta u_{j-1} = u_{j-1} - u_{j-2}$. Then, Eq. (3) can be rewritten as:

$$u_{j+1} = u_j + \frac{1}{2}\Delta u_j - \frac{1}{2}\Delta u_{j-1} + \delta f(x). \tag{4}$$

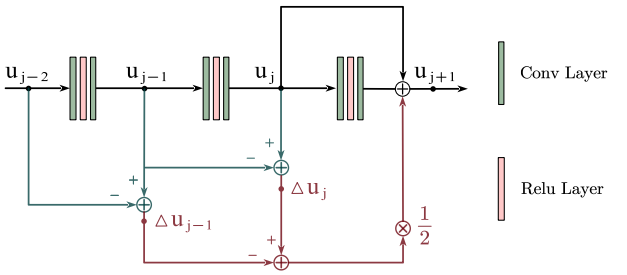

Figure 2: The structure of the TCD residual block, which embodies the TCD solver of ODE by learning the feature change using cascaded convolutional layers. See Section 3.2 for details.

In this paper, we devise the TCD residual block to implement Eq. (4), where $\delta f(x)$ is implemented as a series of two convolutional layers with a ReLU layer inserted. The detailed structure of TCD residual block is shown in Figure 2.

3.3 PMD Neural Module

PMD has been widely studied in image processing, especially for denoising while retaining the structural information in the image. Since there is speckle noise in SAR images, it is attractive to explore PMD in the S2O image translation task to generate noise-free optical images. To this end, we devise a PMD neural module that explicitly functions like PMD.

Given an image or feature u, the PMD equation is:

$$\begin{cases} \frac{\partial u}{\partial t} = div(g(|\nabla u|)\nabla u) \\ g(|\nabla u|) = \frac{1}{1+\frac{|\nabla u|^2}{k^2}} \end{cases}, \tag{5}$$

where $g(|\nabla u|)$ is the diffusion coefficient, t is the diffusion step and k is the shape constant [Guo *et al.*, 2011]. It is easy to find that in flat or smooth regions with small gradients ($|\nabla u| \to 0$), the diffusion coefficient g is near one, while in areas with rich texture or structural details ($|\nabla u| \to \infty$), the coefficient is near zero. With the spatially varying diffusion coefficient, PMD can keep details while removing noise.

After simple derivation based on second-order approximation and $u_{xx} + u_{yy} = u_{\alpha\alpha} + u_{\beta\beta}$, Eq. (5) becomes:

$$\frac{\partial u}{\partial t} = \frac{k^2}{k^2+|\nabla u|^2}\frac{k^2-|\nabla u|^2}{k^2+|\nabla u|^2}u_{\alpha\alpha} + \frac{k^2}{k^2+|\nabla u|^2}u_{\beta\beta}, \tag{6}$$

where u_{xx} and u_{yy} are the second-order partial derivatives of u along the x and y directions, respectively. $u_{\alpha\alpha}$ and $u_{\beta\beta}$ are the second-order partial derivatives of u along the normal direction (image gradient direction) and the tangential direction (edge direction), as shown in Figure 3 .

To better preserve the structural details such as edges, it needs to be smoothed as little as possible in the image gradient direction [Guo *et al.*, 2011]. To this end, we set the coefficient of $u_{\alpha\alpha}$ to zero, i.e., $\frac{k^2}{k^2+|\nabla u|^2}\frac{k^2-|\nabla u|^2}{k^2+|\nabla u|^2} = 0$, implying $k^2 = |\nabla u|^2$. If the diffusion step size Δt is set to one,

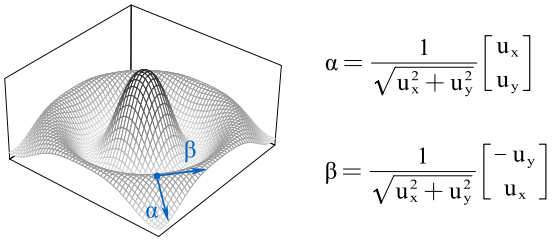

Figure 3: The image gradient direction α and the edge direction β.

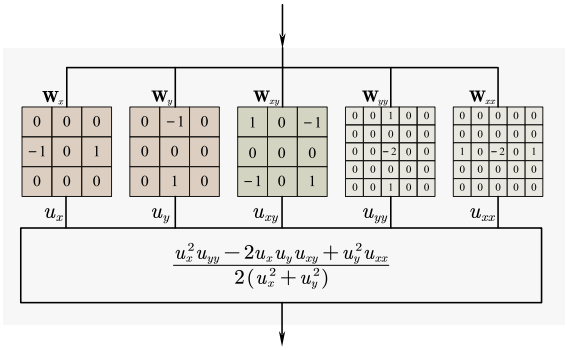

Figure 4: The structure of the PMD head module.

we have the following equation from Eq. (6):

$$\begin{aligned} u_{i+1} - u_i &= \frac{k^2}{k^2+|\nabla u|^2}u_{\beta\beta} = \frac{1}{2}u_{\beta\beta} \\ &= \frac{u_x^2 u_{yy} - 2u_x u_y u_{xy} + u_y^2 u_{xx}}{2(u_x^2+u_y^2)}, \end{aligned} \tag{7}$$

where u_x and u_y are the first-order partial derivatives of u along the x and y directions, respectively. Eq. (7) can be formulated as a residual neural block, where the right side of the equation is the learned residual Δu. Indeed, the residual neural block can be implemented as some convolution layers with fixed kernels. As shown in Figure 4, using convolution kernels like $\mathbf{W}_x = [0,1,0;\mathbf{0};0,-1,0]$, $\mathbf{W}_y = [\mathbf{0};-1,0,1;\mathbf{0}]$, $\mathbf{W}_{xx} = [0,0,-1,0,0;\mathbf{0};0,0,-2,0,0;\mathbf{0};0,0,1,0,0]$, $\mathbf{W}_{yy} = [\mathbf{0};\mathbf{0};1,0,-2,0,1;\mathbf{0};\mathbf{0}]$, and $\mathbf{W}_{xy} = [1,0,-1;\mathbf{0};-1,0,1]$, we can obtain u_x, u_{xx}, u_y and u_{yy} as well as Δu after a network forward pass, e.g., $u_x = \mathbf{W}_x * u$, where $*$ represents the convolution operation. We call these convolution layers with fixed kernels as PMD head, where the mapping function is denoted as $\varphi(\cdot)$.

In addition to learning (calculating) the residual using fixed convolution kernels, we also devise some fusion blocks to promote feature interactions between the backbone network and the PMD head and further refine the feature. Specifically, as shown in Figure 1, the feature from the PMD head module first goes through a convolution layer and a downsampling layer before concatenated and fused with the feature maps from TCD residual blocks in two sequential fusion blocks, each of which consists of three convolution layers. After an upsampling layer and a convolutional prediction layer, we get a side output called PMD image.

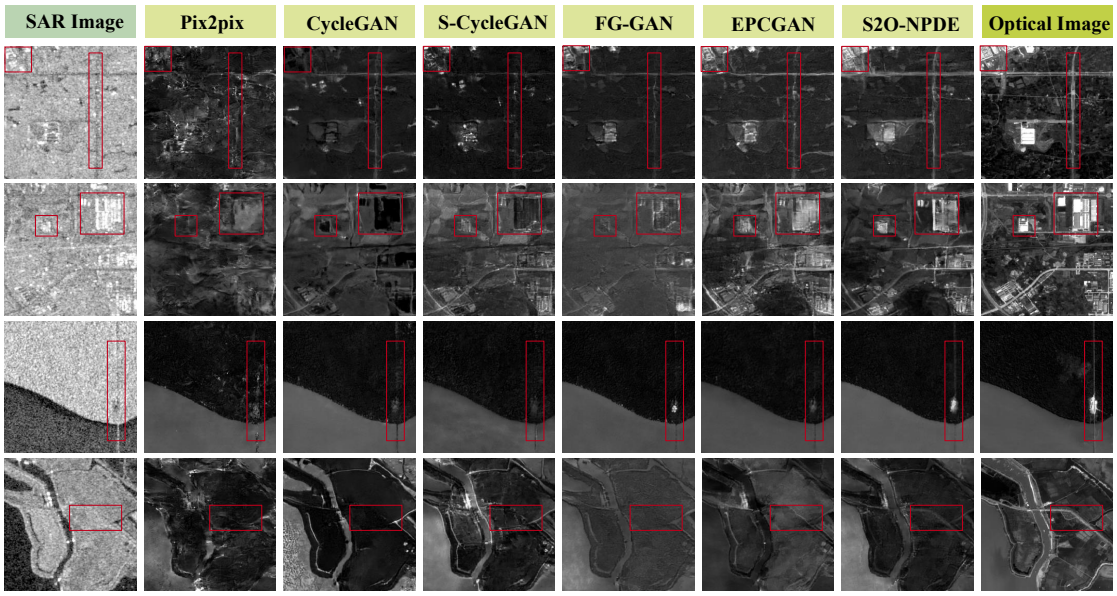

Figure 5: Visual results of some representative SAR-to-optical methods and the proposed S2O-NPDE.

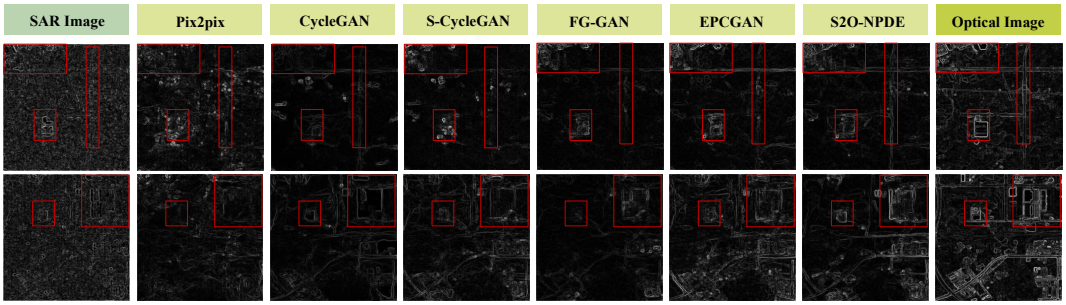

Figure 6: Comparison of structural information achieved by different SAR-to-optical methods

3.4 Loss Function

To train the proposed S2O-NPDE to generate target optical images, we leverage several typical loss functions widely used in previous works [Isola *et al.*, 2017; Zhu *et al.*, 2017; Yang *et al.*, 2022], including the adversarial loss $\mathcal{L}_{GAN}$, cycle consistency loss $\mathcal{L}_{cyc}$, pixel loss $\mathcal{L}_{pix}$, perceptual loss $\mathcal{L}_{per}$. In addition, we also devise a PDE loss $\mathcal{L}_{PDE}$ to emphasize the high-frequency details. Specifically, the final loss function can be defined as follows:

$$\mathcal{L} = \mathcal{L}_{GAN} + \lambda_1 \mathcal{L}_{pix} + \lambda_2 \mathcal{L}_{per} + \lambda_3 \mathcal{L}_{cyc} + \lambda_4 \mathcal{L}_{PDE}, \quad (8)$$

where $\lambda_1 \sim \lambda_4$ are loss weights and set to 10 empirically. Due to the limitation of space, we only introduce the $\mathcal{L}_{PDE}$ while details of other losses can be found in [Isola *et al.*, 2017; Zhu *et al.*, 2017; Yang *et al.*, 2022].

$$\begin{aligned} \mathcal{L}_{PDE} &= \mathcal{L}_{TCD} + \mathcal{L}_{PMD} \\ &= \mathbb{E}_{x,y\sim p_{data}(x,y)} \left\| \varphi(y) - \varphi\left(G_{TCD}(x)\right) \right\|_1 + \\ &+ \mathbb{E}_{x,y\sim p_{data}(x,y)} \left\| \varphi(y) - G_{PMD}(x) \right\|_1, \end{aligned} \quad (9)$$

where $\mathcal{L}_{PMD}$ and $\mathcal{L}_{TCD}$ represent the losses used to supervise the side output from the PMD neural module and the final prediction from the backbone network, respectively. $\varphi(\cdot)$ denotes the function of the PMD head as described in Section 3.3. x and y are the real SAR image and optical image, respectively. The loss function constrains that the high-frequency feature $\varphi(y)$ extracted from real optical image should be consistent with both the high-frequency feature $\varphi\left(G_{TCD}(x)\right)$ of the generated optical image, and the generated PMD image $G_{PMD}(x)$ from the PMD neural module. In this way, $\mathcal{L}_{PMD}$ indeed promotes to learn useful high-frequency features inside the PMD neural module.

4 Experiment

4.1 Dataset and Implementation Details

Following [Guo *et al.*, 2021a], we select 1,600 high-quality SAR-optical pairs as the training dataset and 300 pairs as the test set (denoting as Test1) from the SEN1-2 datatset [Schmitt *et al.*, 2018], which are in the size of 256×256 cropped from the original 258 high-resolution SAR-optical pairs. The Test1 and training dataset contain a wide range of terrain types including forests, lakes, mountains, rivers, buildings, farmland,

Method	PSNR			SSIM		
	Test1	Test2	Test3	Test1	Test2	Test3
Pix2pix	17.17	15.94	16.47	0.3451	0.2664	0.2714
CycleGAN	16.52	15.12	15.92	0.3423	0.2955	0.2895
S-CycleGAN	18.06	15.66	16.54	0.4081	0.2898	0.2840
FGGAN	18.55	16.81	17.15	0.4439	0.3626	0.3301
EPCGAN	18.98	16.54	17.46	0.4491	0.3615	0.3454
S2O-NPDE	**19.46**	**17.54**	**18.25**	**0.4931**	**0.3950**	**0.3748**

Table 1: Quantitative results of different methods.

Method	PSNR			SSIM		
	Test1	Test2	Test3	Test1	Test2	Test3
Base	18.43	16.50	17.41	0.4554	0.3481	0.3424
+TCD	18.74	17.01	17.88	0.4617	0.3682	0.3514
+PMD	18.68	17.36	18.06	0.4756	0.3878	0.3661
+TCD+Grad	18.95	17.41	17.55	0.4606	0.3792	0.3518
+RK2+PMD	18.54	17.15	17.97	0.4713	0.3747	0.3557
+TCD+PMD	**19.46**	**17.54**	**18.25**	**0.4931**	**0.3950**	**0.3748**

Table 2: Ablation study results.

roads, and others. In addition, we select 52 pairs in complex scenes of mountainous areas and 68 pairs in complex scenes of suburban areas as another two test sets, denoting as Test2 and Test3, respectively.

We adopt ADAM optimizer with $\beta_1 = 0.5$, $\beta_2 = 0.999$ to train the model for 200 epochs with a batch size of 1. The learning rate is set to 2×10^{-4} and linearly reduced to zero from the 100th epoch. We select Pix2pix [Isola *et al.*, 2017], CycleGAN [Wang *et al.*, 2019], S-CycleGAN [Zhu *et al.*, 2017], FGGAN [Zhang *et al.*, 2020], EPCGAN [Guo *et al.*, 2021a] as the representative S2O methods for comparison. All the experiments are implemented in PyTorch and on NVIDIA GTX 2080Ti GPUs.

4.2 Quantitative Results

We report the PSNR and SSIM [Wang *et al.*, 2004] scores of different methods on the three test sets in Table 1. As can be seen, the proposed S2O-NPDE achieves the best performance. For example, it outperforms the second best one EPCGAN by average 0.8 dB and 0.0368 SSIM score. Typical image translation methods such as CycleGAN and Pix2pix fail to achieve good results as they have not explicitly modeled the structures and reduce the side effect of the speckle noise. By contrast, our S2O-NPDE can effectively extract multi-level features via residual learning in the TCD residual blocks as well as preserve high-frequency structural details and remove speckle noise in the PMD neural module. These PDE-based modules have explicit designs to function like those PDEs and promote generating target optical images.

4.3 Visual Results

In Figure 5, we present some visual results generated by different methods. Pix2pix produces blurry results since it has not explicit preserved structures and addressed speckle noise. CycleGAN can generate images with clear structure owing to the cycle consistency loss, but these structures may have geometric distortions as indicated by the red boxes. FGGAN and S-CycleGAN have similar phenomena. Although EPCGAN can enhance the structural information, but some details have still not been recovered. For complex SAR images with speckle noise, the proposed S2O-NPDE can generate optical images with more accurate structural details and less noise than other methods. Since the gradient information of an image can reflect its texture and structure information, we also compare the gradient maps of generated optical images by different methods to demonstrate the effectiveness of our method in preserving texture and structure information. It can be seen from Figure 6 that the speckle noise in the SAR image significantly pollutes texture information, thereby degenerating previous methods and resulting in geometric distortions in the translated optical images. By contrast, our S2O-NPDE benefits from the TCD residual block and PMD neural module and produces better results closer to the ground truth.

4.4 Ablation Study

To investigate the effectiveness of each components in S2O-NPDE, we perform an ablation study and present the results in Table 2. Note that the Base model has not used both TCD residual blocks as well as the parallel PMD neural module, i.e., without any feature interactions and the auxiliary PDE loss. Then, we try different variants by adding them on Base. As can be seen, both modules are useful and help the model achieve better performance. Moreover, they are complementary to each other, where using them together generates the best results. We also try alternative designs of TCD and PMD by using RK2 [He *et al.*, 2019] and the gradient branch [Guo *et al.*, 2021a], respectively. However, they show inferior performance to our designs. We argue that although the gradient branch can also extract structural information, it is sensitive to the speckle noise. Besides, RK2-block is designed based on the Runge-Kutta solver of ODE, but it only extracts features from adjacent layers with limited representation ability unlike our TCD block that extracts features from multiple layers.

5 Conclusion

In this paper, we propose a novel S2O-NPDE model for SAR-to-Optical image translation. It is explicitly designed from the perspective of neural partial differential equations, including a backbone of TCD residual blocks and a parallel PMD neural module. The results on the SEN1-2 dataset show that S2O-NPDE can generate clear optical images with finer structures and less noise, owing to the effective designs that function like PDEs in extracting useful features from multiple layers, preserving structural details, and removing noise.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grants 61902293, 62036007, the Equipment Advance Research Field Fund Project under Grant 80913010601, the Shaanxi Province Key Research and Development Program Project under Grant 2021GY-034, the Youth Talent Promotion Project of China Association for Science and Technology 2021QNRC001, the Youth Talent Promotion Project of Shaanxi University Science and Technology Association under Grant 20200103.

References

[Anderson and Wendt, 1995] John David Anderson and J Wendt. *Computational Fluid Dynamics*. Springer, 1995.

[Auer *et al.*, 2009] Stefan Auer, Stefan Hinz, and Richard Bamler. Ray-tracing simulation techniques for understanding high-resolution sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 48(3):1445–1456, 2009.

[Chen *et al.*, 2015] Yunjin Chen, Wei Yu, and Thomas Pock. On learning optimized reaction diffusion processes for effective image restoration. In *CVPR*, 2015.

[Fang *et al.*, 2017] Cong Fang, Zhenyu Zhao, Pan Zhou, and Zhouchen Lin. Feature learning via partial differential equation with applications to face recognition. *Pattern Recognition*, 69:14–25, 2017.

[Fu *et al.*, 2021] Shilei Fu, Feng Xu, and Ya-Qiu Jin. Reciprocal translation between sar and optical remote sensing images with cascaded-residual adversarial networks. *Science China Information Sciences*, 64(2):1–15, 2021.

[Fuentes Reyes *et al.*, 2019] Mario Fuentes Reyes, Stefan Auer, Nina Merkle, Corentin Henry, and Michael Schmitt. Sar-to-optical image translation based on conditional generative adversarial networks—optimization, opportunities and limits. *Remote Sensing*, 11(17):2067, 2019.

[Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[Grohnfeldt *et al.*, 2018] Claas Grohnfeldt, Michael Schmitt, and Xiaoxiang Zhu. A conditional generative adversarial network to fuse sar and multispectral optical data for cloud removal from sentinel-2 images. In *IGARSS*, 2018.

[Guo *et al.*, 2011] Zhichang Guo, Jiebao Sun, Dazhi Zhang, and Boying Wu. Adaptive perona–malik model based on the variable exponent for image denoising. *IEEE Transactions on Image Processing*, 21(3):958–967, 2011.

[Guo *et al.*, 2021a] Jie Guo, Chengyu He, Mingjin Zhang, Yunsong Li, Xinbo Gao, and Bangyu Song. Edge-preserving convolutional generative adversarial networks for sar-to-optical image translation. *Remote Sensing*, 13(18):3575, 2021.

[Guo *et al.*, 2021b] Ping Guo, Kaizhu Huang, and Zenglin Xu. Partial differential equations is all you need for generating neural architectures–a theory for physical artificial intelligence systems. *CoRR abs/2103.08313*, 2021.

[He and Yokoya, 2018] Wei He and Naoto Yokoya. Multi-temporal sentinel-1 and-2 data fusion for optical image simulation. *ISPRS International Journal of Geo-Information*, 7(10):389, 2018.

[He *et al.*, 2019] Xiangyu He, Zitao Mo, Peisong Wang, Yang Liu, Mingyuan Yang, and Jian Cheng. Ode-inspired network design for single image super-resolution. In *CVPR*, pages 1732–1741, 2019.

[Isola *et al.*, 2017] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, 2017.

[Keeling and Stollberger, 2002] Stephen L Keeling and Rudolf Stollberger. Nonlinear anisotropic diffusion filtering for multiscale edge enhancement. *Inverse Problems*, 18(1):175, 2002.

[Lu *et al.*, 2018] Yiping Lu, Aoxiao Zhong, Quanzheng Li, and Bin Dong. Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations. In *ICML*, pages 3276–3285, 2018.

[Perona and Malik, 1990] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.

[Qiao *et al.*, 2019] Tingting Qiao, Jing Zhang, Duanqing Xu, and Dacheng Tao. Mirrorgan: Learning text-to-image generation by redescription. In *CVPR*, 2019.

[Schmitt *et al.*, 2018] M Schmitt, LH Hughes, and XX Zhu. The sen1-2 dataset for deep learning in sar-optical data fusion. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4(1), 2018.

[Simard *et al.*, 1998] Marc Simard, Gianfranco DeGrandi, Keith PB Thomson, and Goze B Benie. Analysis of speckle noise contribution on wavelet decomposition of sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 36(6):1953–1962, 1998.

[Wang and Patel, 2018] Puyang Wang and Vishal M Patel. Generating high quality visible images from sar images using cnns. In *Proceedings of the 2018 IEEE Radar Conference*, pages 0570–0575, 2018.

[Wang *et al.*, 2004] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[Wang *et al.*, 2019] Lei Wang, Xin Xu, Yue Yu, Rui Yang, Rong Gui, Zhaozhuo Xu, and Fangling Pu. Sar-to-optical image translation using supervised cycle-consistent adversarial networks. *IEEE Access*, 7:129136–129149, 2019.

[Yang *et al.*, 2022] Xi Yang, Jingyi Zhao, Ziyu Wei, Nannan Wang, and Xinbo Gao. Sar-to-optical image translation based on improved cgan. *Pattern Recognition*, 2022.

[Zhang and Tao, 2020] Jing Zhang and Dacheng Tao. Empowering things with intelligence: a survey of the progress, challenges, and opportunities in artificial intelligence of things. *IEEE Internet of Things Journal*, 8(10):7789–7817, 2020.

[Zhang *et al.*, 2020] Jiexin Zhang, Jianjiang Zhou, and Xiwen Lu. Feature-guided sar-to-optical image translation. *IEEE Access*, 8:70925–70937, 2020.

[Zhu *et al.*, 2017] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, pages 2223–2232, 2017.