

SAR image to optical image translation technology based on conditional generative adversarial network

Haoxufei Liu, Zhixing Zhou, Cheng Lu and Huanxin Zou*

School of Electronic Science, National University of Defense Technology, Changsha, China

*Corresponding author e-mail: 237006237@qq.com

Abstract. To address the problems of poor readability and difficult interpretation caused by the special imaging mechanism of Synthetic Aperture Radar (SAR) images, this paper combines the latest advances in Generative Adversarial Network (GAN) technology in machine learning to overcome the problems of CycleGAN. In this paper, we combine the latest advances in GAN technology to overcome the problems of unstable training, failure to converge, and lack of diversity in generating a single image, and construct a supporting training dataset to design and optimize a multimodal image translation network model to explore a solution for translating SAR images into easily understood optical images. The research results of this paper are very important for realizing applications such as alignment, matching and change detection between multimodal images.

Keywords: synthetic aperture radar; generative adversarial networks; image conversion.

SAR is a microwave active imaging radar, capable of high-resolution imaging of the target detection area, with The advantages of all-day, all-weather operation, in the military, agriculture and urban and rural planning and other fields have a very wide range of applications. However, the special imaging mechanism of SAR images makes the images difficult to understand, which easily brings obstacles to the application of SAR images by non-specialists.

Image translation is a research hotspot in the field of computer vision and machine learning in recent years, which aims to transform the expression of the input image into another expression, and has wide applications in the fields of image creation, image stylization, image restoration, and domain adaptive learning.

However, CycleGAN and Star GAN proposed in recent years often suffer from problems such as unstable network training, failure to converge, or lack of diversity by generating only a single image.

Therefore, to address the above problems, the main purpose of this paper is to translate SAR images into optical images by designing relevant technical methods, which can help people understand SAR images, and the research results of this paper will also provide technical support for related applications in the field of multimodal image processing.

1. Research content of image translation

1.1 Construction of SAR image and optical image training data set

The open network acquires SAR images and optical images corresponding to the same region in similar time phases, realizes the pre-processing process of SAR images and optical images, completes the accurate alignment of SAR images and optical images with pixel-level accuracy, and provides the basic data set support for deep learning network model training.

1.2 Multimodal image translation network model based on conditional generation countermeasure network

The basic working principle of generative adversarial networks is investigated, the input constraints of generative adversarial networks are improved for multimodal image translation from SAR images to optical images, the depth network models corresponding to generators and discriminators in generative adversarial networks are selected, and the design of multimodal image translation network models is completed.

1.3 Design and optimization of loss function of network model

According to the structure of the multimodal image translation network model, the loss functions (objective functions) of the generator and discriminator networks are designed separately to form the total loss function of the network, and the optimisation strategy of the network model training process is proposed.

1.4 Experimental verification and evaluation of SAR image to optical image translation

Based on the constructed network training dataset, the training of the multimodal image translation network model is completed. At the same time, for the test dataset, the steps and process of the validation experiments for SAR image to optical image translation are designed, and the qualitative and quantitative analyses of the validation experiments are completed using the relevant evaluation metrics.

2. Comparative analysis of image translation methods

Generative adversarial networks have developed rapidly in the field of computer vision in recent years, especially in the field of image data generation, where image translation methods based on generative adversarial networks have also seen rapid development. pix2pix networks were proposed in 2017, using conditional generative adversarial networks with paired input image data, initially solving the problem of one-to-one paired image data translation. Later, to address the problem that the pix2pix network image data must be input in pairs, some scholars proposed the CycleGAN network, which introduced the idea of cyclic consistency in language translation to the image translation task, while using non-paired image data for training and learning, solving the image translation problem under non-paired data.

In 2018, the Star GAN network was proposed to add a mask vector to the target domain information, which was used to ensure that the network could focus on specific labels provided by certain datasets, solving the task of data translation between multiple domains and across datasets. 2019 Meng Xiangchao et al. disclosed a method for translating SAR remote sensing images to optical remote sensing images based on GAN and the feature class to which it belongs, obtaining better good results. However, generative adversarial translation model networks lack a quantitative evaluation system, training is prone to model collapse, and the networks are extremely sensitive to parameter tuning. The current Inception Score, Frechet Inception Distance and Sliced Wasserstein Distance cannot accurately evaluate whether the images are close to the real training data and the diversity of the generated images. GAN networks are unstable, do not converge, or only produce a single image, which lacks diversity. In addition, the migration of the network to different scenes is also a major difficulty.

Therefore, this project addresses the problem of SAR image to optical image translation, collects sensor-specific SAR images (e.g. Sentinel-1) and optical images (e.g. Sentinel-2) based on existing research results, constructs a network training dataset, designs a multimodal image translation model for SAR images to optical images based on conditional generative adversarial networks, designs a network loss The network training optimization strategy is designed, the procedure and steps of the SAR image to optical image translation validation experiment are designed, and the qualitative and quantitative analysis and evaluation of the SAR image to optical image translation validation experiment are completed based on multiple evaluation metrics[3].

3. Experiment and analysis

The research idea of this topic is: based on the SAR image (Sentinel-1) and optical image (Sentinel-2) datasets that can be collected by public networks or commercially purchased, the training dataset for matching SAR images and optical images is constructed using image alignment and other techniques; using the powerful domain learning capability of conditional generative adversarial

networks, the mapping between the two domains (SAR images The generator network and the discriminator network are trained alternately using images from two domains (SAR images, optical images) as training data to fit the mapping between the two data distributions to obtain the translation network model from SAR images to optical images; the SAR remote sensing images to be translated are pre-processed and input to the corresponding adversarial neural network prediction model, and the generator outputs the image translation results.

The research programme of this topic is shown in Figure 2, and the technical routes corresponding to each research element are described in detail as follows.

3.1 Construction of SAR image and optical image training date set

The datasets were derived from open-source SAR images and optical remote Pixel-level aligned SAR and optical image pairs for different scenarios were constructed and 70% of the image pairs under each set were randomly selected as the training sample set, while the rest of the image pairs were used as test samples. , the number of images is not less than not less than 5000 (resolution of 256×256).

3.2 Multimodal image translation network model based on conditional generation countermeasure network

GAN generative adversarial networks consist of two types of networks, generators and discriminators. The purpose of the generator is to try to learn the real data distribution, and the purpose of the discriminator is to try to correctly discriminate whether the input data is from the real data or from the generator; both need to be continuously optimized , each improving their generative and discriminative capabilities. Pix2Pix (Image-to-Image Translation) generates adversarial networks based on the condition that in the generator G In the generator G, an image x is input, an image y is output, and in the discriminator D, a pair of <x,y> is input and D determines whether the image pair is true. During the training iterations, G and D are continuously adjusted and optimised so that eventually D is unable to distinguish between <x,y>, at which point we obtain a GAN model that can implement Pix2Pix. The structure of the underlying conditional generative adversarial network model is shown in Table 1 below.

Table1 network architecture

Encoder	Decoder	Discriminator
CR (64,3,1)	CBRD (512,4,2)	CBR (64,4,2)
CBR (128,4,2)	CBRD (512,4,2)	CBR (128,4,2)
CBR (256,4,2)	CBRD (512,4,2)	CBR (256,4,2)
CBR (512,4,2)	CBR (512,4,2)	CBR (512,4,2)
CBR (512,4,2)	CBR (256,4,2)	C (1,3,1)
CBR (512,4,2)	CBR (128,4,2)	
CBR (512,4,2)	CBR (64,4,2)	
CBR (512,4,2)	C (4,3,1)	

In Table 1, C, B, R and D represent the convolution layer, batch regularization, ReLU activation function and Dropout layer, respectively. The numbers in brackets indicate, from left to right, the number of convolution filters, the size of the spatial filter and the step size, respectively.

3.3 Design and optimization of loss function of network model

The objective function of the network model consists of two loss functions: Confrontation losses L_{cGAN} and $L1$ losses L_{L1} . Let X be the input (grey-scale) amplitude SAR image and Y the output (colour) optical image. Contrast loss is defined as $L_{cGAN}(G, D) = E_{x,y}[\log(D(x, y))] + E_x[\log(1 - D(x, G(x)))]$

where G and D represent the outputs of the generator and discriminator, respectively. In addition, the $L1$ loss is defined as:

$$L_{L1}(G) = E_{x,y}[\|y - G(x)\|_1]$$

In summary, total network losses can be defined as: $G^* = \arg \min_G \max_D L_{cGAN}(G, D) + \lambda L_{L1}(G)$

where λ is $L1$ losses the proportion of losses to total losses.

3.4 Experimental verification and evaluation of SAR image to optical image translation

In order to qualitatively and quantitatively evaluate the SAR image to optical image translation results, this project trains a multimodal image translation network model based on the constructed network training dataset, and designs a test dataset for quantitative evaluation by peak signal-to-noise ratio (PSNR) and structural similarity ratio (SSIM) evaluation metrics. Among them, the lower PSNR is better and the higher SSIM is better. In addition, the results were evaluated using the Lee filter, the Kuan filter, mean filter, and median filter as pre-processing steps for speckle noise reduction are compared for experiments to compare whether the multimodal image translation network model is robust to

speckle noise. Specifically, 80% of the images from the entire training data are randomly selected as training data, and the remaining 20% of the images are used as test data. The number of minimum batches of the network model is set to 8, and the number of training cycles is set to 200. all quantitative evaluation results are set to the average of 5-fold cross-validation.

Table2 Evaluation results

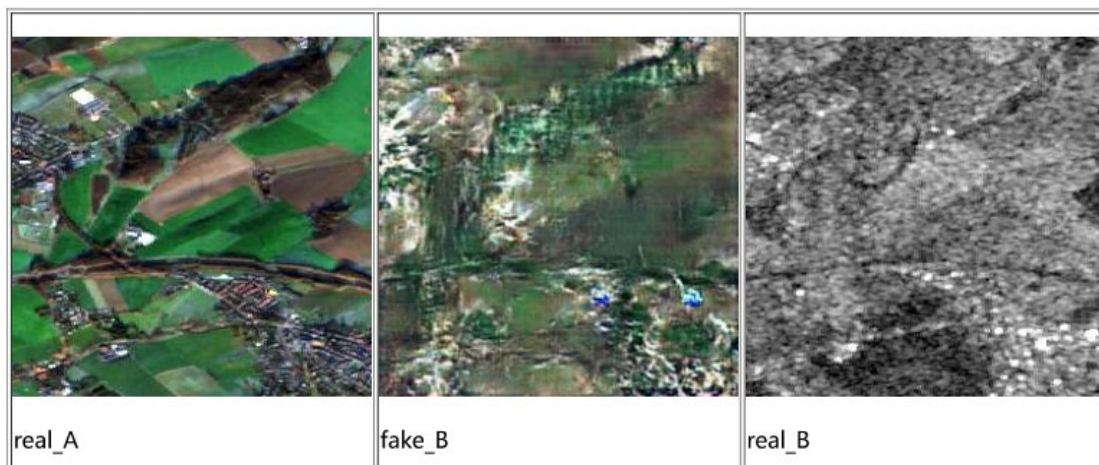
	example	CycleGAN	StarGAN
MSSIM	0.3	0.2	0.3
MSE	400	600	400
PSNR	22	20	22

The false images generated by the model are evaluated with the original optical images by SSIM, we get the average value of SSIM is 0.3, which is larger than the average value of CycleGAN 0.2. Also comparing with StarGAN, the model has better overall performance in different scenes such as desert, coastal, plain and mountainous areas. The experimental results are shown in Figure 2 below.

4. Conclusion

The multimodal image translation network model trained in this paper can better translate SAR images to optical images in different scenes such as desert, coastal, plain, mountain, etc., which can help people understand SAR images and improve the generation effect and adaptability to multiple scenes compared with other CycleGAN and StarGAN models. The research results of this paper will also provide technical support for related applications in the field of multimodal image processing such as alignment, matching and change detection between multimodal images, which is worth further research and exploration

ROIs2017_winter_s2_16_p917



ROIs2017_winter_s2_16_p936

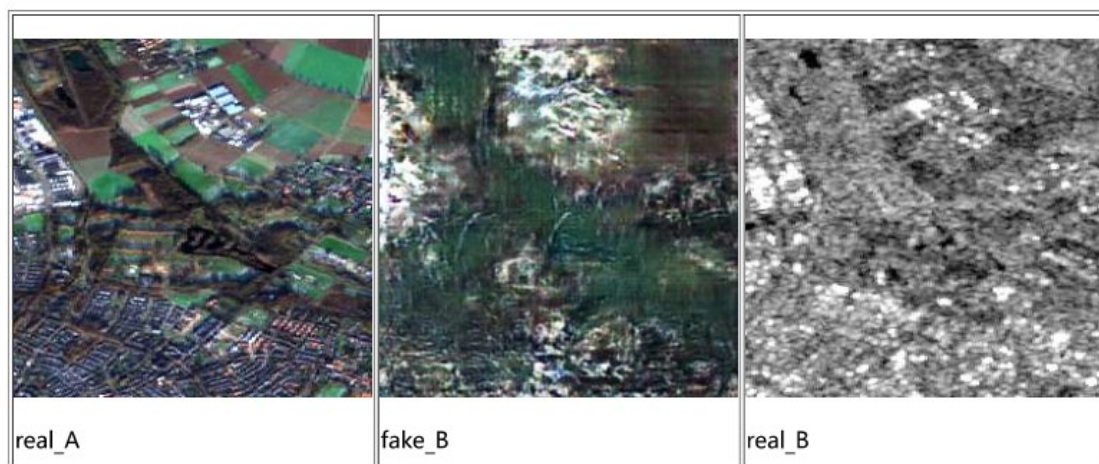


Fig.1 Example of model results.

References

- [1] ZHANG Wenyuan, TAN Guoxin, SUN Chuanming. An Approach to Translate SAR Image into Optical Image . Geomatics and Information Science of Wuhan University,2017, Vol. 42: 178-184, 192
- [2]] Qin Yong, Research on SAR and optical image translation theory and application. Wuhan University, 2017
- [3] Kong Y , Liu S , Peng X . Multi-Scale translation method from SAR to optical remote sensing images based on conditional generative adversarial network[J]. International Journal of Remote Sensing, 2022, 43(8):2837-2860.