Remote Sensing Images Data Augmentation Based on Style Transfer under the Condition of Few Samples

Yuchen Jiang\textsuperscript{a}, Bin Zhu\textsuperscript{b}, Bo Xie\textsuperscript{c}

State Key Laboratory of Pulsed Power Laser Technology, College of Electronic, Hefei, China
\textsuperscript{a}E-mail: jyc2647118@126.com \textsuperscript{b}E-mail: zhubinee@163.com
\textsuperscript{c}E-mail: xiebo17@nudt.edu.cn,

Abstract—To solve the problem that the detection accuracy of remote sensing image is affected by convolution neural network overfitting under the condition of small samples, a data augmentation method based on style transfer is proposed, in which new data is generated from texture of external domain to source domain by using cycle-consistent adversarial networks (CycleGAN). The experiment results show that the accuracy of detection and recognition is improved after adding the generated data to the original data.

1. Introduction

With the rapid development of remote sensing technology, remote sensing image object detection and recognition is widely used in various fields. Convolutional neural network is often used as an important research method. The excellent performance and strong feature extraction ability of convolutional neural network are based on rich datasets. When the amount of data is insufficient, it will lead to overfitting and affect the algorithm performance. However, in the real world, due to the limitation of data recording conditions, it is usually unrealistic to get a large number of labeled remote sensing image data, and there are only a few labeled data samples. In practical application, the number of samples in some datasets is too small, which limits the performance of existing detection and recognition algorithms.

Aiming at how to reduce the influence of overfitting under the condition of few samples, the current commonly used data augmentation methods mainly include image affine change, information deletion, image fusion, generative model\textsuperscript{[1-4]}. Image affine change mainly refers to the image translation, scale transformation, contrast transformation, noise disturbance and other methods. In reference \textsuperscript{[5]}, discrete optimization of the operation in image affine change is applied to image classification. Information deletion mainly refers to the deletion of a part of the image, among which the more typical methods such as cutout, gridmask\textsuperscript{[6-7]} and so on. Image fusion mainly refers to the fusion of two or more images into one image, and the typical methods are mixup and augmix\textsuperscript{[8-9]}. The method of generative model mainly uses the method of generating data. The commonly used generation model mainly includes variational auto encoder (VAE), generative adversarial networks (GAN)\textsuperscript{[10-11]}, etc.

In this paper, cycle-consistent adversarial networks (CycleGAN)\textsuperscript{[12]} is improved and applied to the remote sensing image data augmentation under the condition of few samples. The yolov3 object detection algorithm\textsuperscript{[13]} is used to detect the augmented dataset, and the detection accuracy is improved. The main work includes three aspects.
• This paper proposes a remote sensing image data augmentation process based on style transfer;
• In order to improve the quality of image generation and integrate the local and global information of the image, the network is improved by using multi-scale convolution and attention mechanism;
• the effect of data augmentation is evaluated by the detection accuracy of the detection model.

2. Cycle-consistent Adversarial Networks

A generative model G and a discriminative model D are set up in the generative adversarial networks. The noise signal is taken as the input variable in the generative model G, which is mapped to the data space of the generative model to obtain the probability distribution of the real data. The input of discriminative model D is real data and fake data from generation model G, and the output is a scalar to reflect the probability that the input comes from real data rather than generated data.

The advantage of GAN is that it is a new generative model using uncertain probability density function. Compared with other generative models, it is more convenient to calculate. It does not need to carry out complex derivation and learning, and does not need to use Markov chain\(^{(14)}\). It only needs to use backpropagation algorithm to get the gradient. The generative model of GAN does not use real data to update and iterate directly, but uses the gradient returned by discriminative model to update, which means that the components of real data do not directly affect the parameters of the generative model, which simplifies the calculation process. As far as input is concerned, there are few restrictions and no requirements except differentiability. Theoretically, it can be any distribution, that is, the goal of GAN is to have infinite modeling ability.

The disadvantage of GAN is that it is difficult to guarantee the stability and convergence in the training process. Since there is no fixed expression for the probability distribution of generation model G, in most cases, in order to ensure that the learning rate of generative model G is consistent with that of discriminative model D, we need to deliberately stop the learning process of one model, or reduce the learning rate, so that another model can catch up and train. Different from the optimization process of other generative models, the goal of GAN is to achieve Nash equilibrium, which may lead to oscillation without convergence. In the case of non-convergence of GAN, the problem of mode collapse is more prominent, that is, the generative of model G may generate identical images repeatedly in some cases. The model collapse is originated from minimax game. The gradient descent has no clear priority for minmax and maxmin. When we want it to behave as minmax, we often get the result of maxmin.

CycleGAN aims to solve the problem of unpaired image translation, that is, the image is mapped from source domain A to source domain B. In view of the lack of structural constraints of the original generative adversarial networks, CycleGAN not only learns the mapping from source domain A to source domain B, but also from source domain B to source domain A to strengthen the constraints. The structure of this network has two generative models and two discriminative models, and the network learns the mapping from source domain A to source domain B and from source domain B to source domain A at the same time. The architecture diagram is shown in the Fig. 1:

![Figure 1 Basic structure frame diagram of CycleGAN.](attachment:image.png)

In the figure, A and B represent the source domain A and source domain B respectively; G and F represent the generative model from A to B and the generative model from B to A respectively; \(D_A\) and
DB represent the discriminative models used to distinguish the authenticity of domain A and domain B respectively. The loss function of CycleGAN is divided into two parts, and the first part is the adversarial loss, which is the sum of the two domains:

\[
L_{adv} = E_{x\sim p_A}[\log D_A(B)] + E_{x\sim p_B}[\log(1 - D_B(G(A)))] + E_{x\sim p_A}[\log D_A(A)] + E_{x\sim p_B}[\log(1 - D_B(G(B)))]
\]

The second part is the loss of cycle consistency, that is, source domain B is obtained from source domain A through generative model G, and then reconstructed from source domain B through generative model F to source domain A. The reconstructed source domain A should be consistent with the original source domain A; vice versa, source domain A is obtained from source domain B through generative model F, and then reconstructed back to source domain B from source domain A through generative model G. The reconstructed source domain B should be the same as the original source domain B, and the loss of cycle consistency is as follows:

\[
L_{cyc} = E_{x\sim p_A}[\|F(G(A)) - A\|] + E_{x\sim p_B}[\|G(F(B)) - B\|]
\]

The total loss function is:

\[
L = L_{adv} + \lambda \cdot L_{cyc}
\]

Among them, \(\lambda\) controls the weight of the two losses. In CycleGAN, both the adversarial loss of and the loss of cycle consistency have a positive effect on the effect of image generation. Removing any loss will cause the quality of the generated image to decline. In addition, the use of two-way cycle consistent loss makes the training of the model more stable than that of the single direction.

In terms of network structure, the discriminant model in CycleGAN uses 70 \times 70 PatchGANs\(^{[15]}\) to judge the authenticity of 70 \times 70 image blocks. Such a discriminative model based on image blocks has fewer parameters than the discriminant model based on full image, and can process images of any size in full convolution mode. In the generative model, different number of residual blocks are used for different resolution images, six residual blocks are used for 128 \times 128 images, and nine residual blocks are used for 256 \times 256 and above images.

3. Process and algorithm of remote sensing images data augmentation

3.1. Remote Sensing Images Data Augmentation Process

Different from image affine transformation, information deletion, image fusion and other data augmentation methods, this paper uses training generative adversarial networks to transfer image style to get new samples. The method is as follows: the original dataset is used as the source domain A, and the external dataset is used as the source domain B to train CycleGAN. After the training, the network is used to map the source domain A to the source domain B, and the label of the newly generated data does not change.

3.2. Improvement of CycleGAN

In order to improve the quality of the image generated after the style transfer of CycleGAN model, two improvements have been made. One is to adjust the single scale convolution module of the original generative model to multi-scale convolution, and the other is to introduce the attention mechanism into the generative model and the discriminative models at the same time.

The multi-scale convolution transforms the single scale convolution kernel in the original generation model into multi-scale convolution kernel, and integrates the information of each scale. The specific structure diagram is shown in the following figure:
Attention mechanism usually calculates the response of a certain position in a sequence by paying attention to all positions in the same sequence. In this paper, attention mechanism is applied to generative adversarial networks to utilize global information. As shown in the Fig. 3, f(x), g(x) and h(x) are 1x1 convolutions with different number of output channels, where the number of f(x) and g(x) is 1/8 of the original feature maps, and the number of h(x) is consistent with that of the original feature maps.

4. Experiment and analysis

4.1. Making dataset
In this paper, DIOR dataset[^16] is selected as the research object. The DIOR dataset contains a total of 23463 remote sensing images, including 11725 images in the training set and validation set, and 11738 images in the test set. The image resolution is 800 × 800, and there are 20 categories. In order to study the influence of data augmentation algorithm on target detection algorithm in small sample dataset, this paper reduces the number of training sets, randomly selects 7035 images from training set and validation set as training set, and the test set remains unchanged. The target domain dataset of style transfer selects the crosshatched category in the descriptive textures dataset (DTD).

4.2. Experimental Results and Analysis
The improved CycleGAN is used to transfer the style of the original training set, and the new sample is consistent with the label of the original training set. The example of the original and generated sample is shown in the Fig.4 and Fig.5:
In this paper, the original CycleGAN, CycleGAN with multi-scale convolution, and CycleGAN with multi-scale convolution and attention mechanism are used for data augmentation on small sample datasets. The number of generated data is equal to the number of training sets. Then, yolov3 detection and recognition model is used to detect the data, and the experimental results are shown in the table:
TABLE I. COMPARISON OF DETECTION ACCURACY OF DIFFERENT DATA AUGMENTATION METHODS

| Data Augmentation                  | Precision(mAP) |
|------------------------------------|----------------|
| None                               | 46.06%         |
| CycleGAN                           | 46.74%         |
| CycleGAN+ multi-scale convolution   | 48.42%         |
| CycleGAN+ multi-scale convolution+attention | 49.15%         |

The experimental results show that the detection accuracy of the original training set is 46.06%, the detection accuracy of data augmentation using the original CycleGAN is 46.74%, the detection accuracy of CycleGAN with multi-scale convolution is 48.42%, and CycleGAN with multi-scale convolution and attention mechanism achieves the highest detection accuracy of 49.15%, which shows the effectiveness of this method.

5. Conclusion
Deep neural network is prone to have overfitting problems in small dataset. In this paper, a data augmentation method based on style transfer is proposed for remote sensing images with few samples. The effectiveness of the proposed method is verified by experiments. This method is suitable for remote sensing image data augmentation and helps to improve the detection and recognition accuracy of detection model.

REFERENCES
[1] Esben J B.(2017)“Enumeration as Data Augmentation for Neural Netwo rk Modeling of Molecules,” http://arxiv.org/abs/1703.07 076? context =cs.LG. html
[2] Zhun Z, Liang Z, Guo L K, Shaozi Li, Yi Yang.(2017) “Random erasing data augmentation , ” http://arxiv.org/abs/1708.04896? context=cs.LG. html.
[3] Lemley J, Bazrafkan S, Corcoran P.(2017)“Smart Augmentation - Learning an Optimal Data Augmentation Strategy,” [J].IEEE Access 5,2017: 5858-5869.
[4] Dinh L, Krueger D, Bengio Y.(2014) “Nice: non-linear independent components estimation,"http://arxiv.org/abs/1410.8516? context=cs. LG. html
[5] Elkin D.C, Barret Z, Dandelion M, Vijay V, Quoc V. Le.(2019) “AutoAugment: Learning Augmentation Strategies From Data,” [C]//Proceedings of the IEEE Conference on Computer Vision and PatternRecognition. United States,2019: 113-123
[6] Terrance D, Graham W T.(2017) “Improved regularization of convolutional neural networks with cutout,” http://arxiv.org/abs/1708.04552? con -text=cs. LG.html
[7] Chen P, Liu S, Zhao H, Jial J.(2020) “GridMask Data Augmentation,” http://arxiv.org/abs/2001.04086?context=cs.LG.html
[8] Hongyi Z, Moustapha C, Yann N, David L.(2017) “mixup: Beyond Empirical Risk Minimization,” http://arxiv.org/abs/1710.09412?context= cs. LG.html.
[9] Hendrycks D , Mu N , Cubuk E D , Barret Z, Justin G, et al.(2019) “AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty,” http://arxiv.org/abs/1912.02781?context=cs. LG.html
[10] Kingma D P, Welling M.(2013) “Auto-encoding variational bayes,” http://arxiv.org/abs/1312.6114?context=cs. LG.html.
[11] Goodfellow I, Pouget J, Mirza M, Bing X, David W F, et al.(2014) “Generative adversarial nets,” [C]. Annual Conference on Neural Information Processing Systems 2014, Montreal, Quebec, Canada. 2014: 2672-2680.
[12] Zhu J Y , Taesung P, Phillip I, Alexei A.(2017) “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,” http:// arxiv.org/abs/1703.10593?context=cs.LG.html
[13] Redmon J, Farhadi A. (2018) “YOLOv3: An Incremental Improvement -t,” http://arxiv.org/abs/1804.02767?context=cs.LG.html.
[14] Ackley D H, Hinton G E, Sejnowski T J. (1985) “A learning algorithm for Boltzmann machines,” [J]. Cognitive science 1985, 9(1): 147-169.

[15] Isola P, Zhu J Y, Zhou T, Alexei A, Efros A.. (2017) “Image-to-Image Translation with Conditional Adversarial Networks,” [C]/// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, 5967-5976.

[16] Li K , Wan G , Cheng G , Meng L, Han J,et al. (2019)“Object Detection in Optical Remote Sensing Images: A Survey and A New Benchmark,” http://arxiv.org/abs/1909.00133?context=cs.LG.html