ISSN: 2664-9640

DAGAN based technology and application for generating small sample infrared images of power equipment

Zhankang Liu

School of NCEPU, North China Electric Power University, Heibei 10054, China;

*Corresponding author Email: 1224510309@qq.com

Abstract

Because infrared image has the advantages of strong anti-interference ability, good spatial adaptability and high target contrast, complex tasks such as power equipment classification and detection based on infrared image have been unprecedentedly developed. Accurately identifying the type and state of power equipment is of great significance for the maintenance of the whole power system and the safety of people's life and property. However, the disadvantage is that the structure of deep learning target detection network is generally complex, the number of parameters is large, and its requirements for the quantity and quality of data are higher than those of traditional machine learning algorithms. When there are few sample data, the parameters of the network become uncertain, the trained network model is easy to over fit, and the generalization ability of the network is also very poor. Due to its high sensitivity and confidentiality, the infrared image for power equipment is lack of public data set, which seriously restricts the development of deep learning target detection network for power infrared image. To solve this problem, this paper uses the improved data enhanced generation countermeasure network (Dagan) to expand the data. Firstly, the generator structure is changed and a new generator structure mul-u-resnet3 + is proposed. The structure optimizes the network structure and residual module. The u-net structure is optimized into mul-u-net3 + structure, and the residual block is changed into sandglass module. In addition, the gradient penalty of its loss function is changed into random gradient penalty, so as to reduce the difficulty of network training. In this paper, the algorithm is tested with the data of a 220kV and 500kV substation, and the expanded image is used for re detection. The experimental results show that the expanded samples have a better effect in target detection, which proves the feasibility and effectiveness of this technology.

Keywords

Data Augmentation,Generative Adversarial Networks,Infrared image of power equipment, Image Generation.

1. Introduction

At present, deep learning methods represented by convolutional neural networks have gradually been widely applied in various complex problems. The idea of applying deep learning object detection networks to the field of infrared object detection is receiving increasing attention from researchers. The training of infrared target detection networks requires a large amount of data to obtain a learning model with strong generalization ability, which highly relies on the feature information provided by a large amount of infrared image data. Infrared image data is an image generated by infrared sensors by detecting target thermal radiation [1]. Due to the need for special imaging equipment to capture, it is limited by natural factors, data recording, and human and material resources. Obtaining massive training data is very expensive and requires a large amount of manpower and material resources. The lack of infrared image data will bring negative effects such as overfitting to the training of infrared target detection network, which limits the further application of depth learning in the field of infrared target detection. The size and quality of the dataset directly affect the final performance of deep learning algorithms, including object detection networks. Therefore, expanding image data to increase the amount of data is an important way to improve the performance of object detection networks.

Simply put, data augmentation is about expanding the sample. Common data augmentation operations include Flip, Rotation, Scale, Crop, Translation, Brightness adjustment, and Gaussian noise addition. As long as minor changes are made to the existing dataset, the network will perceive it as a different image. But these operations are not universal. An effective generation method is to correctly mine common patterns in a small number of known samples and apply the learned patterns to create new samples. The Generative Adversarial Network (GAN) has received widespread attention since its inception. GAN mainly uses the idea of zero sum game in game theory to make GAN reach Nash equilibrium through the continuous game of generating network G (Generator) and discriminator network D (Discriminator). At this time, G learns the distribution of data and uses the learned information to generate new samples. GAN has been widely applied in fields such as style conversion, image generation, and speech recognition, but its training is prone to problems such as gradient vanishing and pattern collapse. So people have proposed various improvements to GAN and achieved very good results.

This article has made some improvements on the basis of DAGAN, proposing a new network and naming it MUL-DAGAN3+. In order to fully utilize the potential information of the image, MUL-DAGAN3+changed the original ResNet [3] U-Net [4] network model to MUL-U-Net3+[5] with multiple feature inputs; Ordinary residual blocks can easily lead to information loss and gradient confusion. This article uses the SandGlass module to solve this problem. This article tests the network using infrared datasets of 220Kv and 500Kv substation equipment, and designs comparative experiments to compare the accuracy of infrared target detection before and after expansion. The experimental results indicate that the expanded data can effectively improve the recognition rate of the device.

2. Related Work

The goal of data augmentation is to expand the training set with new samples. Traditional data augmentation techniques such as cropping, rotation, and color dithering can only generate limited image diversity and cannot produce realistic images. In contrast, the depth generative model can use the distribution of training data to generate more diverse and real samples for feature enhancement and image enhancement.

Variational self encoder [6] and generation countermeasure network and other generative model have made significant breakthroughs in generation. To achieve the goal of generating samples using a few images. CGAN [7] introduces a condition variable y and adds additional information y to the model when generating and discriminating images, which can guide the data generation process. However, the network layers of its generator and discriminator both use full connections, which results in a slight lack of effectiveness in extracting image feature values. The edges of the generated images are also not continuous enough, and the authenticity is relatively poor. ACGAN [8] is used for the generation of specific category data and classification. When ACGAN is used for the generation of small sample category data, there are two outputs to judge whether it is true or false and to distinguish classification, and the corresponding two loss function. When a small sample category is generated, its two losses will conflict, and it is unable to generate images with small sample size. FIGR [9] combines meta learning with Reptile to learn the data distribution of similar images. But the generated image quality is poor. ProGAN [10] adopts a dynamically growing network structure, gradually learning image features from low resolution to high resolution, and ultimately generating high-resolution simulation images. Reference [11] successfully applied GAN to the field of image style transfer, solving the problem of one-to-one image mapping from visible light image style to infrared image style. However, this method requires a large amount of data for training and cannot achieve high-quality image data conversion in small sample situations.

3. Model Framework

3.1. Network Model.

The goal of this model is to learn the features of a small number of image samples and generate new images similar to them through this feature. In this framework, the generator encodes the incoming image through an automatic encoder, adds noise, and decodes it to obtain new images of the same class.

The principle of the model in this article is the same as GAN, and the objective function of the original GAN is as follows:

$$\min_G \max_D V(G,D) = \mathbb{E}_{x\sim P_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim P_z(z)}\left[\log\left(1 - D\left(G(z)\right)\right)\right] \tag{1}$$

Among them, x represents the real image, z represents the noise of the input generator network, and G (z) represents the image generated by the generator network.

D (x) represents the probability that the discriminator network determines whether a real image is real. And D (G (z)) is the probability that the discriminator network determines whether the image generated by the generator is true.

The generator network should make the generated images as close to real images as possible. In other words, we need to make D (G (z)) as large as possible, and then V (D, G) will become smaller. The discriminator aims to distinguish between real images and generated images as much as possible, with D (x) being larger and D (G (x)) being smaller. At this point, V (D, G) will become larger.

The model is also composed of a generator and a discriminator, as shown in Figure 1. The left generator receives the image xi of class c and obtains the feature vector ri through the generator's encoder. The other input is linearly processed Gaussian noise, which is concatenated with the image's feature vectors and input together into the decoder to generate the image xg.

The discriminator has three inputs. Two images of class C and the generated image xg, the discriminator should try to identify the real image and the generated image as much as possible.

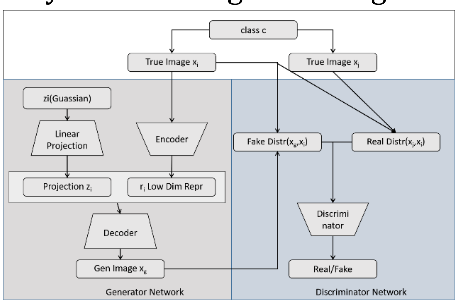

Fig. 1 DAGAN Network Architecture

3.2. Generator Model

Net lacks the ability to explore sufficient information from the full scale. In order to make up for this deficiency, MUL-UNet 3+uses the multi-scale feature tensor of the input image as input on the basis of U-net3+, in order to better extract local features. Each decoder layer incorporates small-scale and co scale feature maps from the encoder, as well as large-scale feature maps from the decoder, which learn fine-grained and coarse-grained semantics at full scale. Combining high-level and low-level semantics from feature maps at different scales; Learn hierarchical representations from multi-scale aggregated feature maps. The network structure is shown in Figure 2.

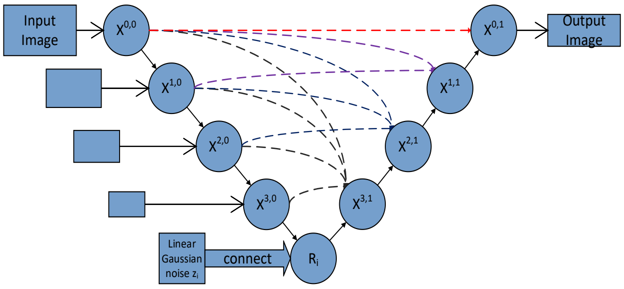

Fig. 2 Generator Architecture

3.3. DiscriminatorModel.

The discriminator has two data inputs, one is the generator image xg with input xi, and the other is the data xi. The goal of the discriminator is to minimize the Wasserstein distance between the generated image and the real image.

This article uses a discriminator for DenseNet network structure. DenseNet consists of 1 Dense Blocks and 3 Transition Layers. Each dense block consists of four layers of convolution and one layer of transition layer, as shown in Figure 3. In the last convolutional layer of each dense block, a weight pruning of 0.5 was used to set the training growth rate k to 64, the training cycle epoch to 200, and the learning rate to 0.00015. The parameters of the Adam optimizer are β1= 0.9, β2= 0.9.

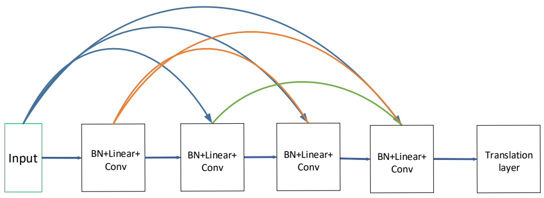

Fig. 3 Dense Blocks Architecture

4. Model Framework

4.1. Dataset

The dataset used in this article is various infrared image samples from a 220Kv and a 500Kv substation, totaling 3200 images. This includes eight major categories of equipment: transformers, circuit breakers, current transformers (CT), voltage transformers (PT), GIS, capacitors, isolation switches, and lightning arresters. Image size is 448 × 336 pixels, divide the data into training, validation, and testing sets in an 8:1:1 ratio.

4.2. Comparative Experiments and Evaluation Indicators

In order to compare the enhancement effect of the model in this article, three experiments were designed for comparison. One group did not use data enhancement algorithms to process images, and the dataset was recorded as B1; Expand the training and validation sets using traditional geometric transformation methods such as translation, cropping, and flipping, and record the enhanced dataset as B2; The dataset expanded by the image enhancement algorithm in this article is denoted as B3. The sample size of the three test sets remains unchanged, and the other control variables are also the same. The data was trained using Faster R-CNN, SSD, and YOLOv4 object detection algorithms. Three sets of experiments were recorded as A1, A2, and A3, respectively. The network was trained to convergence and tested on the original dataset. The accuracy of the object detection algorithm was evaluated using AP (Average Precision) values.

4.3. Analysis of experimental results

The experiment was run on Windows 64 and Nvidia GTX2060 operating systems, and network construction was carried out based on the TensorFlow framework. After 100 iterations, the loss function basically converges, and the experimental results are shown in Figure 4.

From the experimental results, it can be seen that the method of relative and geometric transformation generates infrared images through the algorithm in this paper, while ensuring device features, increasing the diversity of target features, theoretically improving the accuracy of device detection and classification.

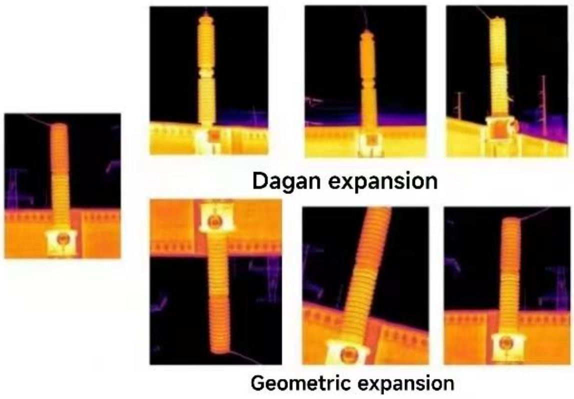

Fig. 4 Comparison of Image Amplification Effects of Voltage Transformers

Three target detection algorithms were used to train three sets of experiments, and the results are shown in Table 1. It can be seen that the AP value obtained from target detection of group A2 and A3 after data expansion is higher than that of group A1 without data expansion, which indicates that the algorithm in this paper and the traditional data enhancement methods can improve the problem of overfitting of target detection algorithm caused by too few targets. By comparing the A2 and A3 groups, the improvement effect of the A3 group is more significant. This indicates that although traditional data augmentation methods increase the number of targets in the image, they do not increase the diversity of target features, which can improve the accuracy of target detection to a certain extent. Data augmentation methods based on generative adversarial networks can not only increase the number of targets but also enrich the diversity of target features, while also increasing the applicability and generalization of the model.

Table 1 Number of images in each group of the dataset

group	B1	B2	B3
A1	75.8%	80.2%	82.3%
A2	81.6%	83.4%	85.9%
A3	83.7%	86.1%	90.5%

5. Summary

This article combines multi-scale fusion of MUL-U-Net3+and generative adversarial neural networks to propose a new data enhanced generative network. Conduct tests on infrared equipment images of a 220Kv and a 500Kv substation, and quantitatively analyze the experimental results using three object detection algorithms. The results show that compared with existing methods, the generation algorithm proposed in this paper can generate more diverse and realistic images, effectively improving the accuracy of object detection in cases of data imbalance and data scarcity. f you follow the “checklist” your paper will conform to the requirements of the publisher and facilitate a problem-free publication process.

References

[1] Shi Fukai Research on infrared image data expansion method based on generative adversarial networks [D]. Xi'an University of Electronic Science and Technology,2020.DOI:10.27389/d.cnki.gxadu.2020.000269.

[2] Goodfellow I, Pouget-Abadie J, Mirza M, et al.Generative adversarial nets//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2014:2672-2680.

[3] He K, Zhang X, Ren S, et al. Deep residual learning forimage recognition // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas,USA, 2016:770-778.

[4] Ronneberger O., Fischer P., Brox T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham.

[5] Huimin H., Lanfen L., Ruofeng T., Hongjie H., Qiaowei Z. et al.A Full-Scale Connected UNet for Medical Image Segmentation.ariXiv preprint arXiv:2004.08790.

[6] Zhai Zhengli, Liang Zhenming, Zhou Wei, Sun Xia. Overview of Variational Autoencoder Models [J]. Computer Engineering and Applications,2019,55(03):1-9.

[7] lsola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks// Proceedings of theIEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017:1125-1134

[8] Liu Qun, Chen Forgingsheng. Classification of hyperspectral remote sensing images using ACGAN and multi feature fusion [J]. Journal of Huaqiao University (Natural Science Edition), 2019,40(01):113-120.

[9] Alex Nichol, Joshua Achiam, and John Schulman. 2018. On first-order metalearning algorithms. arXiv preprint arXiv:1803.02999 (2018).

[10] Karras T, Aila T, Laine S, et al．Progressive growing of GANs for improved quality, stability and variation［J］．Arxiv Preprint Arxiy, 2017: 1710． 10196．

[11] ZHU J Y, PA R K T, ISOLA P, et al.Unpaired image-to-image translation using cycle-consistent adversarial networks［C］/ /IEEE International Conference on Computer Vision (ICCV), 2017:2342-2351．