International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 7 July 2023, pp: 654-663 www.ijaem.net ISSN: 2395-5252

Scene recognition classification technology of controllable diversity generative adversarial networks

Zhaotao[1], Pengfeng[2], Zhou Fachao[2], Gao Mingya[2], Liu Chao[2], Luo Nan[2], Zhu Shijiang[1]

[1]*Hubei Provincial Key Laboratory of Intelligent Construction and Management of Water Engineering, China Three Gorges University, Yichang 443002;*
[2]*Hubei Provincial Office of River and Lake Chiefs 430064;*

Date of Submission: 15-07-2023 Date of Acceptance: 25-07-2023

ABSTRACT:
Using deep learning for high-resolution remote sensing image scene classification can achieve higher accuracy. Deep models usually require a large amount of high-quality training samples; however, the number of samples for some applications is inherently small and lacks diversity, and at the same time, due to the cost, it is difficult to collect a large number of samples; this situation will lead to monotonous of spatial information and small of sample set size; final cause deep models easy fit to specific feature and obtain lower classification ability. To address the above problems, this research proposes a scene recognition classification technology of controllable diversity generative adversarial networks(CD-GAN). The key character of CD-GAN is the ontrollable generation of description tensors with spatial diversity, and according to description tensors near-real remote sensing scene samples can be generated; the generated samples can greatly improve the diversity of spatial features and structures of the original sample set, and further facilitates the CNN to discover the key spatial features of the scene during the training process and improve the classification accuracy. In the experiment, five different methods were introduced for comparison, and CD-GAN obtained the highest classification accuracy; results show that CD-GAN can generate samples with more spatial information diversity, and these samples can significantly improve the accuracy of scene classification.

KEYWORDS: Convolutional neural networks; Generative adversarial network; Scene classification; Remote sensing image; Sample set augment

I. INTRODUCTION

Remote sensing image scene classification technology can automatically divide remote sensing image blocks into different categories, which has great application value in urban planning, environmental monitoring, land resource management and so on (Zhao et al., 2014). With the advancement of artificial intelligence technology, the classification technology of deep models represented by Convolutional Neural Networks (CNN) has become a standard method for remote sensing scene classification methods (Gu et al., 2019; Zhu et al., 2017). CNN can extract higher-level attribute information in remote sensing scenes and obtain classification accuracy far exceeding that of traditional shallow models (Xia et al., 2017). At the same time, based on the CNN structure, the attention mechanism and memory mechanism can be added to the scene classification to continuously improve the accuracy of scene classification (Wang et al., 2018; wang et al., 2022; Li et al., 2021).

Deep models such as CNN require a large number of samples for training. However, in practical work, it is often encountered that the scope of the study area is small or the number of objects to be classified is itself small. At this time, difficulties will be encountered: either provide very few samples, the classification model cannot be fully trained, and thus cannot be effectively classified; either almost all the objects to be distinguished are added to the sample set, and automatically classification is no longer required (Tao et al., 2020). Therefore, CNN needs to deal with the problem of small samples in practical scene classification applications (Akodad et al., 2020).

Different from the traditional model, the number of CNN parameters is large enough to ' remember ' all the details in a small number of samples, so CNN is easy to fit with the specific differences in the scene when the sample size is small (Nogueira et al., 2017; Lv et al., 2019). In order to improve the performance of CNN in the case of small samples, the main methods currently used in the field of remote sensing and image recognition are: using geometric transformation to transform the sample image in a certain direction and scale to increase the number of samples (Yu et al., 2017); this method performs well on large general data sets, but it may cause the feature of category confusion to be amplified in the case of fewer samples (Andresini et al., 2021).

Generative Adversarial Networks (GAN) provide a new way to solve the problem of small samples. The generative adversarial network contains generator G and discriminator D, and the GAN can identify the key features in the sample through the adversarial training of the two models (Goodfellow et al., 2014); GAN can effectively overcome the influence of overfitting, noise and sample overlap, and effectively generate image data that can be used to augment the sample set (Shin et al., 2018; Yang et al., 2019). In the field of remote sensing scene classification, GAN has received more attention. In terms of identifying key features, GAN can effectively deal with the relationship between high dimension (such as using hyperspectral images) and low sample size and find the key features of the category (Zhu et al., 2018; Lin et al., 2017). In terms of sample set amplification: Xu et al.constructed a linear exponential structure to enhance the ability of traditional GAN to generate high-resolution scene images (Xu et al., 2018); Ma et al.realized the generation of online high-diversity scene samples based on GAN, and the accuracy of the obtained classification model is higher than that of the traditional geometric transformation enhanced model (Ma et al., 2018). The key to using GAN to generate remote sensing images and improve classification accuracy is to improve the diversity of spatial information. GAN can improve the space to form more spatial content combinations and avoid the amplification of confusing attributes that often occur in geometric transformation (Pan et al., 2020). The spatial diversity of remote sensing samples manufactured by generative adversarial networks comes from the input random string z. At this time, two problems may be encountered for z: 1) Randomness reduction: some randomness may be treated as noise in the first few layers of G neural network and lose part of the random content (Odena et al., 2017); 2) Uncontrollability: from random string to random spatial structure is completely controlled by weights, and the training of weights is random, so it is difficult to determine which structures are controlled by the specific position of the string (Pan et al., 2021). This problem directly leads to a dilemma when using GAN to generate remote sensing samples. In order to ensure the quality of a single sample, a more hierarchical structure may be introduced, which will significantly reduce the spatial diversity. However, reducing the number of network layers in order to improve spatial diversity will reduce the quality of a single sample. This makes the sample space generated by the GAN-based scene classification method limited in diversity and very difficult to control and experiment with.

In order to solve the above problems, this paper proposes a scene recognition classification technology of controllable diversity generative adversarial networks(CD-GAN). The key feature of CD-GAN is the introduction of tensor generator T, which actively constructs a two-dimensional description tensor with diversity in spatial structure and details. The generator G converts the two-dimensional description tensor into a remote sensing scene image; the controllability of spatial diversity is achieved by controlling the output CD-GAN of T. In the experiment, two remote sensing scene classification datasets, UC-Merced and AID, were introduced for comparison. Compared with the other four traditional methods, CD-GAN achieves higher classification accuracy; especially in the case of fewer samples, the advantages of CD-GAN are more obvious. This shows that CD-GAN can be adapted to small sample remote sensing classification scenarios and can play a greater role in practical applications.

II. METHOD DESCRIPTION

Overall process of the method

In view of the problems existing in the existing research, the main goal of CD-GAN is to generate samples with sufficient spatial diversity to promote the training process of CNN. The overall process of the CD-GAN method is as follows :

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 7 July 2023, pp: 654-663 www.ijaem.net ISSN: 2395-5252

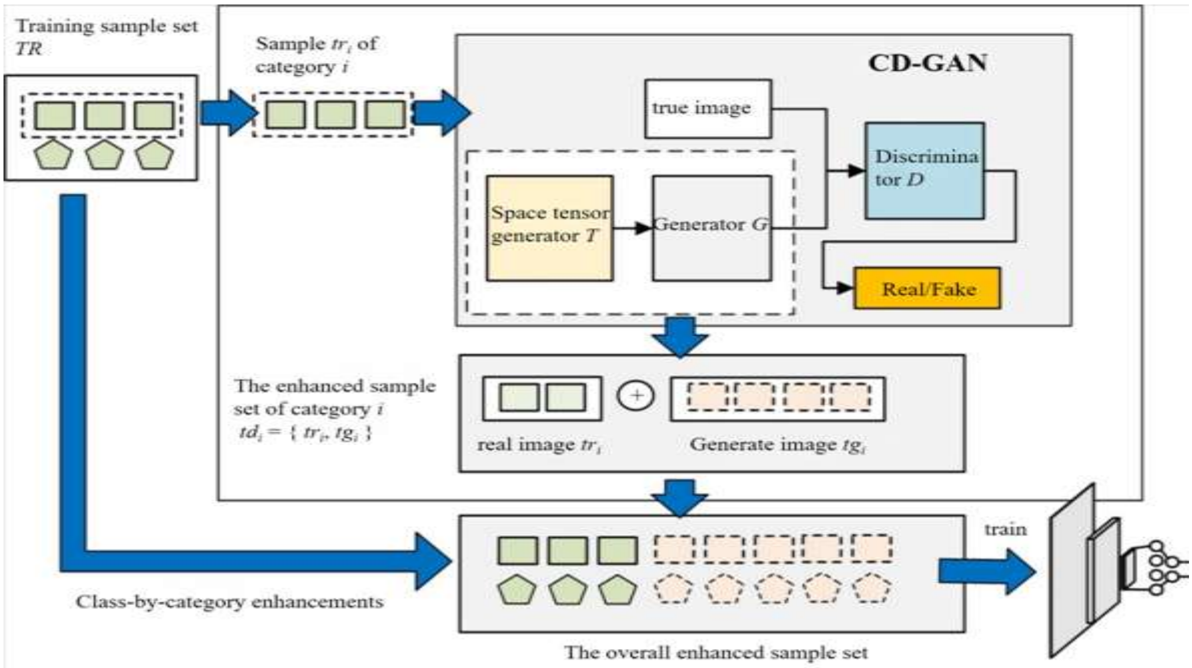


Fig.1 The overall flow chart of CD-GAN method

As shown in Figure 1: For a remote sensing scene classification task, the input is a real training sample set TR = { tr_1, tr_2..., $tr_{Ncategory}$} containing Ncategory categories; among them, tr_i is a collection of remote sensing image samples of category i. CD-GAN consists of three components: Spatial Tensor Generator, T, Generator. G and Discriminator, D. T are used to randomly generate a spatial description z, and G receives a z to generate a pseudo-scene image G(z). For the input image x, the decision result of D is D(x) (the image x is real from tr_i or forged G(z)). D and G have an antagonistic relationship. The target formula of CD-GAN is:

$$\min_G \max_D V(D,G) = E_{x\sim p(x)}[\log(D(x))] + E_{z\sim p(z)}[\log(1-D(G(z)))] \quad (1)$$

Where E is the expected operator, the distribution of the real image is p(x), and the corresponding distribution of the z generated by the generator is p(z). By gradually optimizing Formula (1), the image generated by G is more and more close to the real image. At this time, for a scene image category i adversarial model CD-GAN can be expressed as a generation process:

$$CD-GAN(\mathrm{tr}_i) \rightarrow tg_i \quad (2)$$

At this time, CD-GAN has the ability to discover key spatial features in tr_i and generate a series of images td_i in a random manner. The iterative application of CD-GAN 's training and generation process can enhance the spatial information of the original sample set TR. The overall process of the method described in this article is as follows:

CD-GAN Overall Process Algorithm CG-GAN overall process（GAN-OP）
Input: TR
Output:M_{cnn}
Begin
TD=ø;
M_{CD-GAN}=Initialize a CD-GAN model;
for i=1: $N_{category}$
Using Formula 1 and using the sample set tr_i to train M_{CD-GAN};
The sample tg_i is generated by using T and G of M_{CD-GAN}.;
td_i={tr_i, tg_i};
TD←td_i;
M_{cnn}=Construct a CNN model M, and use TD to train M;
return M_{cnn};
End

The input of GAN-OP is to train the remote sensing scene sample set TD; for each category tr_i, CD-GAN trains and discovers a combination of high-level spatial features in the sample. T generates random tensors and uses G to convert these tensors into corresponding generated samples tg_i. At this time, a set of enhanced samples td_i ={tr_i, tg_i} corresponding to category i can be constructed. After processing all categories, the diversity-enhanced training sample set TD = {td_1, td_2..., $td_{Ncategory}$} can be obtained. TD can be used to

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 7 July 2023, pp: 654-663 www.ijaem.net ISSN: 2395-5252

train a CNN classification model Mcnm; at this time, Mcnm has a more stable spatial feature recognition ability.

Model structure of CD-GAN

The model structure with the traditional CD-GAN consists of three components: tensor generator T, generator G and discriminator D. The structure of the tensor generator T is shown in Figure 2:

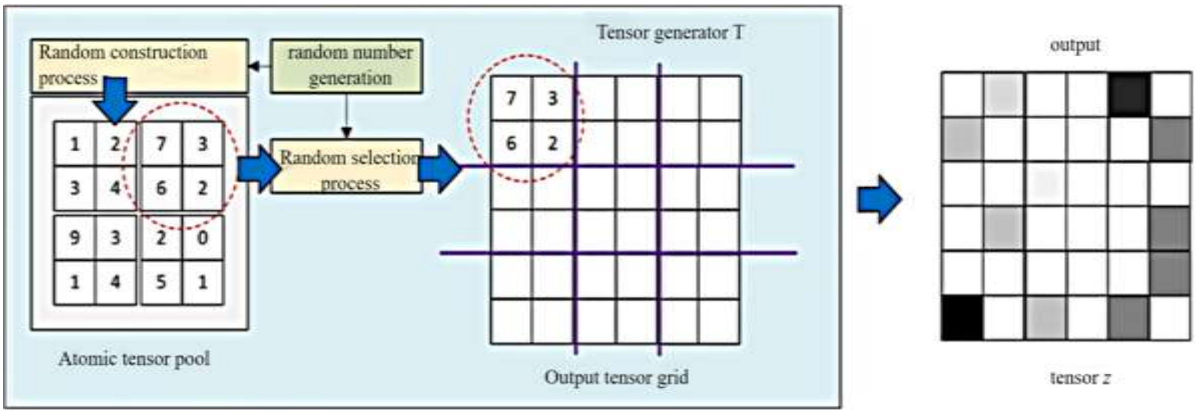


Fig. 2The structure of tensor generator

As shown in Figure 2, the tensor generator is mainly responsible for generating a two-dimensional tensor z. The tensor generator first constructs a set of ' atomic tensors ' to form an atomic tensor pool T_{atom}. The process is as follows:

Atorm Tensor Pool Construction Algorithm (ATPCA)
Input: $Size_{atom}$, $Rand_{Max}$, $Rand_{Max-atom}$
Output: T_{atom}
Begin
$T_{atom}= \varnothing$;
for i=1: $Rand_{Max-Tatom}$
M_{atom}=The maximum value for generating $Size_{atom}$ × $Size_{atom}$ is the $Rand_{Max}$ random number matrix;
$T_{atom} \leftarrow M_{atom}$;
returnT_{atom};
End

After obtaining the T_{atom}, T constructs the tensor output network. The elements in T_{atom} are randomly selected and added to the output network. Finally, the output tensor z is obtained by normalizing the whole output network. The algorithm of this process is described as follows:

Tensor Generator Output Algorithm Tensor generator output algorithm(TGOA)
Input: $Size_{grid}$, $Rand_{Max-atom}$
Output:z
Begin
Grid=Create a 2D array of $Size_{grid}$ × $Size_{grid}$;
for i in 1: $Size_{grid}$
for j in 1: $Size_{grid}$
Grid[i, j]=T_{atom} randomly selects an element between 1 and $Rand_{Max-atom}$;
z=Join all the elements in the Grid to build a two-dimensional matrix;
z=For all elements in z, normalized to between 0 and 1;
return z;
End

Each time T calls TGOA, a two-dimensional random tensor z will be generated, which will drive the generator G to generate the corresponding remote sensing scene image. The value of using tensor generator T in CD-GAN is mainly reflected in two aspects:
(1) Introducing sufficient spatial structure randomness.

In the existing GAN-based scene classification research, it can be seen that if the traditional random string-based method is used to drive G to generate images, this randomness may be absorbed in the first few layers of the G neural network, so that the images generated by G are concentrated in a limited number or even a spatial structure. The z of CD-GAN is equivalent to directly introducing the randomness of the structure into G, which directly drives it to generate diverse spatial structures.
(2) Reduce the randomness of element content and output noise

For a specific category of remote sensing image scene, although the spatial information is

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 7 July 2023, pp: 654-663 www.ijaem.net ISSN: 2395-5252

diverse, the content elements are usually limited. If too much random content is introduced, on the one hand, too much input data change range will make G difficult to train. On the other hand, G may encounter random content that is not encountered in training during the generation process. This random content will produce obviously an unreasonable color/texture combination or noise through the neural network. The z of CD-GAN is composed of a fixed range of atoms, which makes it possible to avoid excessive random elements and noise elements to a certain extent.

The G of CD-GAN is responsible for generating remote-sensing scene images, and its structure is shown in Figure 3:

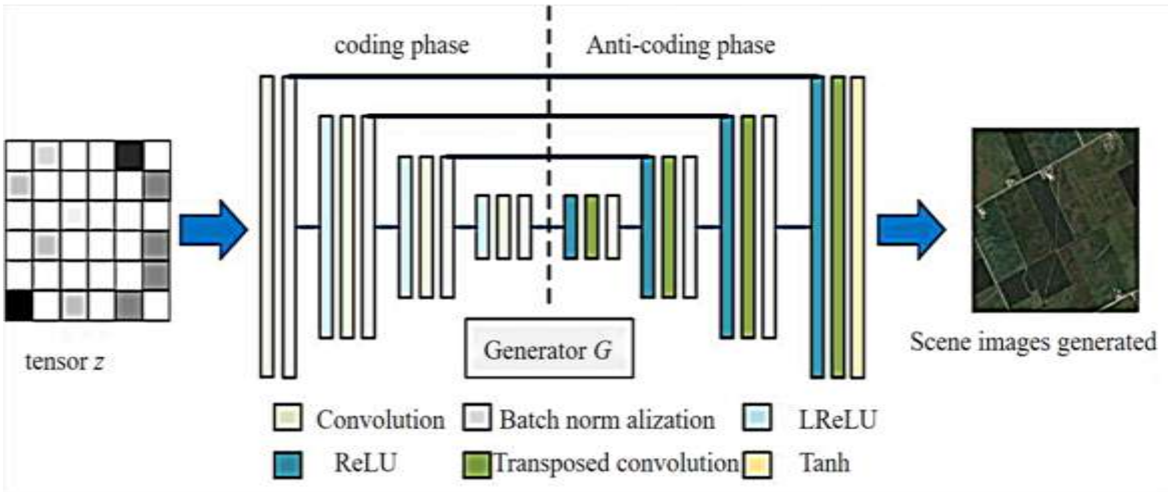


Fig 3The structure of generator G

The generator G generates the corresponding scene remote sensing image after receiving the tensor z of the T output, as shown in Figure 3: The structure of G adopts two stages of coding and anti-coding close to the U-Net neural network :

(1) Coding stage

In the coding stage, G uses four groups of layers, each group contains a convolutional layer (the size of the convolution kernel is 5 × 5, stride is 2 × 2) and a Batch normalization layer; the LReLU layer is used for processing between groups and groups. Due to the introduction of the convolution layer with stride, each group of Feature maps will be reduced to half of the original size. Through this parity is G negative layer by layer extraction to express the spatial relationship in z.

(2) Anti-coding stage

In the anti-coding stage, four groups of layers are also used: each group of layers contains ReLU, Transposed convolution layer and Batch normalization layer, and finally the results are generated through the Tanh layer output. Each group in the coding stage will also directly connect the output to the corresponding anti-coding stage, which makes it possible to take into account the spatial relationship information and specific detail information in the generation process.

The D of CD-GAN is responsible for testing whether the remote sensing scene image is yes, and its structure is shown in Figure 4:

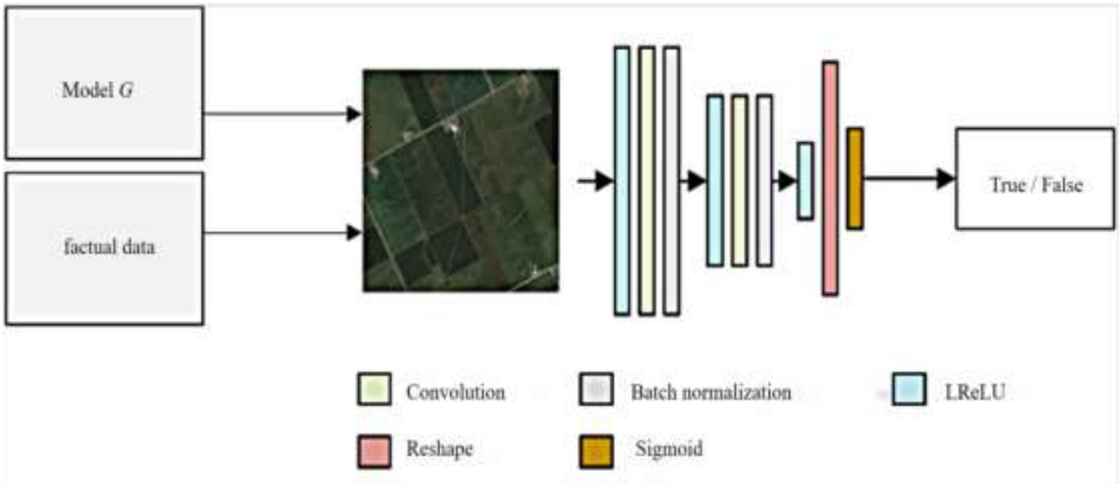


Fig.4 The structure of the discriminator D

As shown in Figure 4, the discriminator D is responsible for distinguishing whether the input image comes from the real data (true) or the generated data of the model G. By describing the confrontation process of T, G and D through formula (1), the image generated by G is gradually close to the real remote sensing scene. Using G, the required remote sensing image samples can be obtained. The method of CD-GAN from training to output is as follows :

CD-GAN Generate samples algorithm CD-GAN Generate samples algorithm (GSA)

Input: tr_i, $N_{initsamples}$
Output: tg_i

Begin

Initialize the T, G, and D models

//1.Pre-training phase

The ATPCA algorithm is used to establish the atomic tensor pool.;

trainx=The $N_{initsamples}$ tensor is established by using the TGOA algorithm.;

trainy=Randomly select $N_{initsamples}$ from tri for different images.;

Trainx and trainy are trained as the input and output of G.;

//2.stage of training

for i in 4:1

temp=Resample tr_i with 1 / i.;

The output of T is used as the input of G and the CD-GAN is trained with a temp real image.;

//3.output phase

testx=The TGOA algorithm is used to establish the tensor with the same number of tr_i samples.;

tg_i =Testx is used as the input of G to generate remote sensing images.;

return tg_i;

End

The GSA algorithm is divided into three stages: 1) Pre-training stage. In CD-GAN, due to the large number of layers of G and the complexity of remote sensing scene data, it is far from being able to generate images in the case of neural network initialization; the goal of this stage is to use a certain number of tensor z to the output image sample so that G can initially achieve the function of generating images; able to continue with D against the overall training. 2) In the training phase, the complete tr_i content is not used at the beginning of the training phase, but the resampled samples of tr_i are used for training, so that the entire training process can begin to focus on the overall spatial structure (easier to train), and then gradually focus on the details of the training on the basis of the previous round, until CD-GAN can adapt to the full resolution of the tr_i content. 3) In the output stage, the TGOA algorithm is used to generate the tensor and the G is converted to the corresponding generated image set tg_i. The GSA algorithm can be used to realize the sample generation based on CD-GAN. At this time, due to the introduction of T, more abundant spatial diversity can be achieved.

III. EXPERIMENTS

Algorithm implementation and experimental data set

This study uses Python 3.8 to implement all CD-GAN algorithms; the deep learning part is implemented by Keras, and the image access and

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 7 July 2023, pp: 654-663 www.ijaem.net ISSN: 2395-5252

processing use the scikit-image development kit. In order to deeply compare the remote sensing scene classification ability of the algorithm, this paper compares the following methods:
(1) CNN: simply use CNN to learn and classify samples.
(2) CNN + GT: For samples, geometric transformation (GT) is used to perform 5 transformations of 90,180,270 degrees rotation, left-right and up-down flipping to increase the number of samples, and CNN is used for learning and classification.
(3) Diversity-GAN: A self-growing GAN neural network is used for remote sensing classification. The feature of the GAN is to use a random string to drive the GAN to generate remote sensing images and enhance the classification ability of CNN.
(4) MARTA-GAN: GAN is used to extract the key features of remote sensing images, and based on this feature, the shallow model is trained to achieve the classification goal.
(5) CD-GAN : The method proposed in this paper.

For the test data set, this paper uses two famous data sets, UC-Merced and AID, to train. All algorithms are running and tested on an i7-10700F 32G / GTX 1070 8G computer. Due to the limited memory space, the images of the two data sets are scaled to 256 × 256 for testing. In order to test the classification ability of each model in the case of insufficient samples, 10,20,30,40 and 50 samples were selected as training samples in the two data sets, and the remaining samples were used as test samples to test the accuracy. Five tests were conducted to evaluate the accuracy, and the overall accuracy (OA) was used as the evaluation index.

Remote sensing scene content generated by CD-GAN method

In the case of inputting 50 samples, the typical remote sensing image samples generated by the GSA method of CD-GAN are shown in Figure 5.

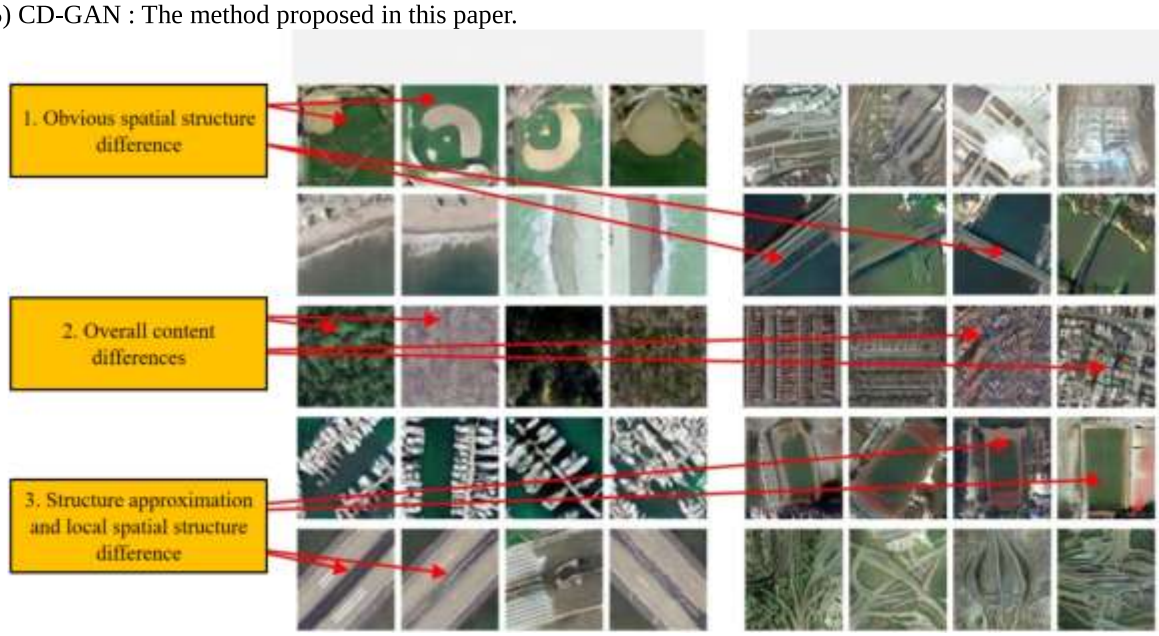


Fig.5Scene image generated by CD-GAN

It can be seen from Fig.5 that the scene image generated by CD-GAN is not perfect, and it does not reach the point where human recognition can be deceived. The intuitive feature is that the boundaries of some artificial buildings are not straight (CD-GAN does not introduce an evaluation function to measure whether the boundaries of buildings are straight). However, the main goal of CD-GAN is to produce samples that can improve classification accuracy. These samples have the following characteristics:

(1) Obvious spatial structure difference.

Whether it is the UC-Merced and AID datasets, the direction of the generated samples and the location of the key objects can be changed in a large range.

(2) Differences in the overall content

Some of the generated samples can also show significant changes in the overall content, such as color and composition.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 7 July 2023, pp: 654-663 www.ijaem.net ISSN: 2395-5252

(3) Changes in details

Even if some samples are similar in spatial structure and direction, their details, such as road width, width and composition, are different.

Thanks to the introduction of T, a wide range of changes can be input into G. The above results show a key feature of CD-GAN, which is the diversity of spatial structure and spatial content. This diversity can further enrich the content of the sample set, and then achieve the purpose of improving the classification accuracy of CNN.

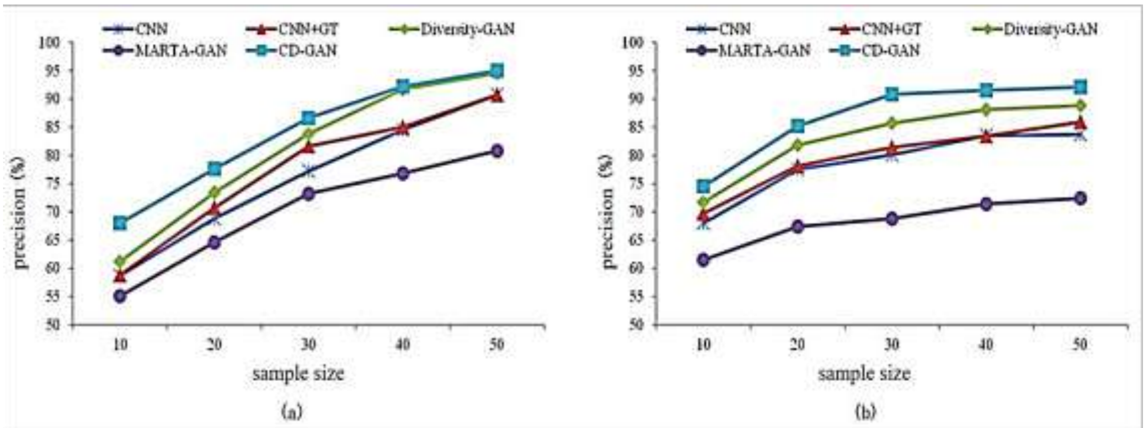
Comparison of the accuracy of the five methods

The comparison of the accuracy of the five methods is shown in Table 1:

Tab.1 Accuracy comparison of five methods

data set	sample fraction(%)	Overall Accuracy OA (%)				
		CNN	CNN+GT	Diversity-GAN	MARTA-GAN	CD-GAN
UC-Merced	10	58.7±2.6	58.8±1.6	61.2±1.6	55.1±1.9	68.0±2.2
	20	68.8±1.0	70.7±1.1	73.4±0.9	64.6±0.4	77.6±0.8
	30	77.2±0.5	81.6±0.2	83.8±0.5	73.2±0.6	86.6±0.9
	40	84.5±0.5	85.0±0.5	91.7±0.4	76.8±0.3	92.2±0.6
	50	90.7±0.3	90.7±0.9	94.5±0.7	80.8±0.3	95.0±0.4
AID	10	68.0±1.7	69.7±1.5	71.7±0.7	61.5±2.9	74.5±2.8
	20	77.5±1.1	78.1±0.6	81.8±1.1	67.4±0.4	85.2±0.7
	30	80.1±1.1	81.4±1.2	85.7±0.6	68.8±0.9	90.8±0.8
	40	83.5±0.5	83.4±0.7	88.1±0.7	71.4±1.1	91.5±0.4
	50	83.7±0.7	85.9±0.4	88.8±0.2	72.4±0.8	92.1±0.6

It can be seen from Table 1 that CD-GAN has achieved the highest classification accuracy among the five methods, no matter with fewer 10 samples or more 50 samples. For the UC-Merced data set, the classification accuracy of 95.0 ± 0.4 was achieved at 50 samples, and the classification accuracy of 92.1 ± 0.6 was achieved for the AID data set. Based on the sample size and accuracy, the comparison of the five methods is shown in Figure 6 :



(a) Comparisons based on UC-Merced datasets(b) Comparison based on AID datasets
Fig.6Accuracy comparison of five methods

It can be seen from Figure 6 that the trends of the five methods for the two data sets are basically consistent; for CNN, the classification accuracy is low in the case of a small sample size, and the accuracy is rapidly improved with the increase of the number of samples. Since the number of AID data sets is much larger than that of UC-Merced (spatial diversity is also higher and more difficult to classify), the classification accuracy of AID data sets is higher than that of UC-Merced data sets at 50 samples. For CNN + GT, although the geometric transformation can increase the number of samples, the key difference between this model and the GAN series method is that the simple geometric transformation may lead to the existence of confused samples after the same change confusion still exists, even increased, so CNN + GT compared to CNN accuracy improvement is not obvious; for Diversity-GAN, it introduces a random string to drive the diversity of samples, and the accuracy is improved obviously. MARTA-GAN focuses on discovering the spatial feature information of samples, so the accuracy will be lower than that of ordinary CNN. The goal of CD-GAN is consistent with that of Diversity-GAN, that is, to generate samples with more spatial diversity. CD-GAN introduces a tensor generator T to control the output of diverse spatial structures, which makes CD-GAN significantly better than Diversity-GAN in spatial diversity, especially in the case of small sample size.

IV. CONCLUSION

Using deep models such as CNN for remote sensing scene classification can obtain higher classification quality, but massive and high-quality sample sets are key to training deep models.

It can't meet the needs of training CNN. In practice, it is often difficult to collect enough samples, so it is very important to use small samples to train CNN model. This paper proposes a scene recognition classification technology of controllable diversity generative adversarial networks. This method has the following characteristics:

(1) Using generated samples to improve the diversity of spatial information: CD-GAN adds samples generated by *G* to the original sample set; from the experiment, it can be seen that in the structure, content and details of spatial information, the generated samples can appear a random combination of multiple contents. These combinations can enrich the spatial content of the original sample set, so that the classification model can more easily summarize the spatial characteristics of specific categories, and then improve the CNN classification ability.

(2)Different from the traditional GAN method that introduces random strings to drive the generation behavior of *G*, CD-GAN uses the tensor generator *T* to generate a two-dimensional tensor and then converts the two-dimensional tensor into an image through a neural network. This generation process can directly introduce diversity into the spatial structure, rather than relying on the complex training process to identify the spatial randomness in the string; this makes the samples generated by CD-GAN more directly generate spatial diversity, and the experimental results also prove that this model has more advantages than Diversity-GAN using random strings directly.

Based on the above characteristics, CD-GAN can significantly enhance the spatial information content of the original sample set and obtain higher classification accuracy in typical application scenarios such as small sample size, huge sample collection cost, and small research area. The application environment suitable for a small sample size makes CD-GAN have practical application value.

REFERENCES

[1]. Akodad, S., Bombrun, L., Xia, J., Berthoumieu, Y., Germain, C., 2020. Ensemble learning approaches based on covariance pooling of CNN features for high resolution remote sensing scene classification. *Remote Sensing* 12, 3292.

[2]. Andresini, G., Appice, A., De Rose, L., Malerba, D., 2021. GAN augmentation to deal with imbalance in imaging-based intrusion detection. *Future Generation Computer Systems* 123, 108-127.

[3]. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. *Advances in neural information processing systems* 27.

[4]. Gu, Y., Wang, Y., Li, Y., 2019. A survey on deep learning-driven remote sensing image scene understanding: Scene classification, scene retrieval and scene-guided object detection. *Applied Sciences* 9, 2110.

[5]. Li, M., Lei, L., Tang, Y., Sun, Y., Kuang, G., 2021. An attention-guided multilayer feature aggregation network for remote sensing image scene classification. *Remote Sensing* 13, 3113.

[6]. Lin, D., Fu, K., Wang, Y., Xu, G., Sun, X., 2017. MARTA GANs: Unsupervised representation learning for remote sensing image classification. *IEEE Geoscience and Remote Sensing Letters* 14, 2092-2096.

[7]. Lv, J., Liu, Y., 2019. Data augmentation based on attributed scattering centers to train robust CNN for SAR ATR. *IEEE Access* 7, 25459-25473.

[8]. Ma, D., Tang, P., Zhao, L., 2019. SiftingGAN: Generating and sifting labeled samples to improve the remote sensing image scene classification baseline in vitro. *IEEE Geoscience and Remote Sensing Letters* 16, 1046-1050.

[9]. Nogueira, K., Penatti, O.A., Dos Santos, J.A., 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition* 61, 539-556.

[10]. Odena, A., Olah, C., Shlens, J., 2017. Conditional image synthesis with auxiliary classifier gans, International conference on machine learning. *PMLR, pp*. 2642-2651.

[11]. Pan, X., Zhang, C., Xu, J., Zhao, J., 2021. Simplified object-based deep neural network for very high resolution remote sensing image classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 181, 218-237.

[12]. Pan, X., Zhao, J., Xu, J., 2019. A scene images diversity improvement generative adversarial network for remote sensing image scene classification. *IEEE Geoscience and Remote Sensing Letters* 17, 1692-1696.

[13]. Shin, H.-C., Tenenholtz, N.A., Rogers, J.K., Schwarz, C.G., Senjem, M.L., Gunter, J.L., Andriole, K.P., Michalski, M., 2018. Medical image synthesis for data augmentation and anonymization using generative adversarial networks, International workshop on simulation and synthesis in medical imaging. *Springer, pp*. 1-11.

[14]. Tao, C., Qi, J., Lu, W., Wang, H., Li, H., 2020. Remote sensing image scene classification with self-supervised paradigm under limited labeled samples. *IEEE Geoscience and Remote Sensing Letters*.

[15]. Wang, J., Wang, X., Xing, L., Liu, B.-D., Li, Z., 2022. Class-Shared SparsePCA for Few-Shot Remote Sensing Scene Classification. *Remote Sensing* 14, 2304.

[16]. Wang, Q., Liu, S., Chanussot, J., Li, X., 2018. Scene classification with recurrent attention of VHR remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 57, 1155-1167.

[17]. Xia, G.-S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., Zhang, L., Lu, X., 2017. AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing* 55, 3965-3981.

[18]. Xu, S., Mu, X., Chai, D., Zhang, X., 2018. Remote sensing image scene classification based on generative adversarial networks. *Remote sensing letters* 9, 617-626.

[19]. Yan, Y., Tan, Z., Su, N., 2019. A data augmentation strategy based on simulated samples for ship detection in RGB remote sensing images. *ISPRS International Journal of Geo-Information* 8, 276.

[20]. Yu, X., Wu, X., Luo, C., Ren, P., 2017. Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. *GIScience & Remote Sensing* 54, 741-758.

[21]. Zhao, L.-J., Tang, P., Huo, L.-Z., 2014. Land-use scene classification using a concentric circle-structured multiscale bag-of-visual-words model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7, 4620-4631.

[22]. Zhu, L., Chen, Y., Ghamisi, P., Benediktsson, J.A., 2018. Generative adversarial networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* 56, 5046-5063.

[23]. [23] Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine* 5, 8-36.