



Review of Generative Adversarial Networks in Object Detection

Chenyang Zhou^(✉), Siman Kong, and Jianzhi Sun^(✉)

College of Computer Science, Beijing Technology and Business University, Beijing 100048,
China

563588355@qq.com, sunjz@th.btbu.edu.cn

Abstract. Generative Adversarial Network (GAN) has become a research focus in the field of deep learning, and its research output has grown exponentially. This brand-new technology provides new ideas and methods for object detection, and has achieved remarkable success. Firstly, this paper introduces the basic GAN model and its derivative models in the field of object detection. Then analyzes the application status of GAN from object detection fields, such as industrial defect detection, medical image detection, remote sensing image detection, and face detection. Finally, summarize and prospect the technology development of generative adversarial networks.

Keywords: deep learning · generative adversarial network · adversarial learning · object detection

1 Introduction

In recent years, with the surge of data volume, the research of neural network algorithm and the improvement of computing power, the field of artificial intelligence has developed rapidly. Unsupervised learning has received more and more attention due to its unique methodology, becoming a hot direction of deep learning in recent years. Variational Auto-Encoder, Deep Belief Network, Flow-based model have emerged, however, the generalization ability of these models is insufficient. The proposal of Generative Adversarial Network (GAN) brings a new breakthrough in the field of artificial intelligence in the direction of generation. At present, generative adversarial network has become a research hotspot in deep learning, which has been applied in computer vision, language processing, anomaly detection and localization, information security, object detection and other fields.

2 GAN and Its Derivative Models in the Field of Object Detection

In 2014, Goodfellow et al. [1] proposed a new framework, Generative Adversarial Network (GAN), to estimate generative models through adversarial training.

The generative adversarial network consists of a generator network and a discriminator network. The generator G captures the original data distribution and transforms the random noise into pseudo samples that are close to the real samples, while the discriminator D is used to determine whether the input data comes from the original data distribution or from the pseudo data generated by the generator. The output of the discriminator D is fed back to the generator G , which is constantly trained to make the generated data closer to the real data. The core idea of generative adversarial network is derived from the two-person zero-sum game [2]. When training GAN, the generator and the discriminator play a minimax game. After training, iteratively optimizes and finally reaches a Nash equilibrium [3]. The model structure of GAN is shown in Fig. 1.

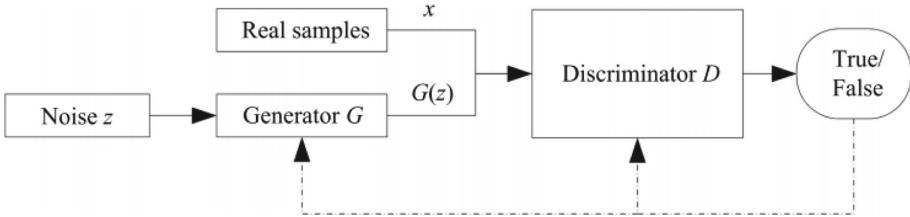


Fig. 1. Model structure of GAN

Compared with traditional generative models, GAN has many advantages: 1) It does not need to use Markov chain, only need to use back propagation to obtain gradient. 2) No inference is required in the learning process, and a large number of training tricks and loss functions that have been proposed can be used. 3) The generator of GAN is trained indirectly by discriminator, which means that the input source data is not directly copied into the parameters of the generator. 4) GAN can represent very sharp or even degenerate distributions.

GAN solves many problems of generative models, but it also has some limitations: 1) GAN training needs to achieve Nash equilibrium, but how to find the Nash equilibrium point is a difficulty and challenge [4, 5]. 2) The training process needs to ensure that the discriminator and the generator are trained synchronously, otherwise there will be a mode collapse problem. 3) Poor interpretability, unable to use mathematical formulas or parameters to represent the sample distribution generated. In addition, there are also problems such as vanishing gradients, too free models, difficulty in evaluating model training, and unsuitability for processing discrete data.

2.1 Perpetual GAN

Jianan Li et al. [6] proposed the Perceptual GAN in 2017, which is mainly used for small object detection. By mining structural correlation between objects of different scales, the feature representation of small objects is improved to make them similar to large objects. Perceptual GAN includes a generator network and a perceptual discriminator network. The generator network is a deep residual feature generation model, which converts the original poor features into high-score deformed features by introducing low-level

fine-grained features. On the one hand, the discriminator network distinguishes the high-resolution features generated by small objects from the real large objects, and on the other hand, it could use the perceptual loss to improve the detection rate. Experiments on Traffic-sign Detection Datasets and Pedestrian Detection Datasets demonstrated the effectiveness of Perceptual GAN for small object detection.

2.2 MTGAN

Yancheng Bai et al. [7] proposed an end-to-end multi-task generative adversarial network (MTGAN). The model consists of a generator network and a discriminator network. In the generator, a super-resolution network (SRN) is introduced, which can up-sample smaller target images to a larger scale, and SRN can generate higher quality images. The discriminator is a multi-task network that simultaneously distinguishes real and generated super-resolution images, predicts object categories, and refines predicted bounding boxes. Furthermore, in order for the generator to recover more details for detection, the classification and regression losses in the discriminator are back-propagated into the generator during training. Extensive experiments on the COCO dataset demonstrate the effectiveness of the method in recovering sharp super-resolved images from small blurred images, and show improved detection performance over the new technique. Based on this model, reference [8] combines the reconstruction error with the discriminator output to improve the performance of anomaly detection.

2.3 CGAN

Mirza et al. [9] proposed Conditional Generative Adversarial Networks (CGAN) in 2014. The main contribution is to add extra information y to the input of the generator and discriminator of the GAN. In the generator, prior input noise $P_z(z)$ and conditional information y are combined as joint hidden layer representation. In the discriminator, the real data x and extra information y are used as input to the discriminant function. Reference [10] proposed an image-to-image conversion framework based on CGAN, using CGAN loss L_{CGAN} and reconstruction loss $LL1$ to learn and observe the normal internal characteristics of crowd moving scenes. A large number of training G and D are performed on the normal frame of the moving scene and its corresponding optical flow images, and anomalies are detected by calculating the local difference between the generated content and the real frame.

3 GAN'S Application in the Field of Object Detection

Object detection is one of the classic tasks in the field of computer vision. With the large-scale application of deep learning in the field of object, the accuracy and speed of detection technology have been greatly improved, so it has been widely used in industrial defect detection, medical image detection, remote sensing image detection, face detection and other fields. Although the current object detection algorithm has achieved good results compared with traditional methods, it still cannot meet the needs of some special detection problems. The proposal of generative adversarial network provides a certain solution to the challenges in the field of object detection. Table 1 summarizes the models applied in different fields of object detection.

Table 1. GAN's application in the field of target detection

Field	Model
Industrial Defect Detection	AnoGAN [11], GANomaly [12], MVAE-GAN [13], Two-stage material surface defect detection based on CycleGAN [14], Defect simulation algorithm based on GLS-GAN [15]
Medical Image Detection	Sparse-GAN [16], MADGAN [17], Detection of lesion area based on GAN [18]
Remote Sensing Image Detection	Attention-GAN-Mask R-CNN [19], Ship identification based on GAN [20], FPN-GAN [21], EESRGAN [24]
Face Detection	SRGAN [29], GAN-based end-to-end convolutional neural network [30], C-GAN [31], GAN-based de-occlusion architecture [32]

3.1 Industrial Defect Detection

With the in-depth integration of the new generation of information technology and the manufacturing industry, people have higher and higher requirements for product quality. Product surface defects not only destroy the appearance quality of the product, but also may cause serious damage to the performance of the product. Surface defect detection is very important in order to detect problems in time, so as to effectively control product quality. The challenge of surface defects lies in the lack of sufficient training samples, especially defective samples. Insufficient training samples are prone to the problem of overfitting of deep learning models, so the research process of unsupervised learning methods is greatly accelerated. The manifestations of deep learning methods based on unsupervised learning are defect-free sample training methods and simulated defect sample training methods.

The defect-free sample training method obtains the defect detection result by learning the sample distribution, reconstructing the defect-free sample, and comparing the difference between the reconstructed sample and the input sample. Schlegl T et al. [11] proposed AnoGAN, the first method to introduce generative adversarial networks into defect detection. In this method, non-defective samples are used as training samples for unsupervised training. The idea is to learn the distribution of normal samples through GAN, and then map defective samples to the latent variables, and then reconstruct the samples from the latent variables. The reconstructed image will eliminate the defective part on the basis of retaining the original image characteristics, so the defect could be located by the residual between the reconstructed image and the input image. Akcay et al. [12] proposed a new anomaly detection algorithm, GANomaly, which utilizes conditional generative adversarial networks to jointly learn the generation and latent reasoning of high-dimensional image space and uses encoder-decoder-encoder in the generator network. The network, by comparing the latent variables obtained by coding and the latent variables obtained by reconstructing coding, could judge whether it is an abnormal sample. Through experiments on datasets from different fields, the validity of the model is verified. Li [13] proposed an image reconstruction model MVAE-GAN

based on the generative adversarial network and the variational autoencoder. By training non-defective samples, it could learn the latent feature information of non-defective samples and make it have the reconstruction ability of normal samples. Experiments show that the model performs better in various indicators such as structural similarity and peak signal-to-noise ratio.

The simulation defect sample training method solves the problem of insufficient defect samples in practical applications by generating annotated simulation defect samples and training the defect detection model. Tsai et al. [14] proposed a two-stage CycleGAN to automatically synthesize and annotate local defective pixels. The first stage uses two CycleGAN models to automatically synthesize and annotate defective pixels in images. Then, the defect images synthesized in CycleGAN model and their corresponding annotation results are used as input-output pairs to train U-NET network. Experiments show that the scheme has sufficient generality for industrial detection applications. Liu [15] proposed a defect simulation algorithm based on GLS-GAN, which fused the network structure of U-shaped network and residual network characteristics. The region training strategy for local defect generation enables the generator network to create simulation defects based on the real image. Using simulation samples to train defect identification model and defect segmentation model can greatly reduce the number of necessary real defect samples.

3.2 Medical Image Detection

Medical images reflect the internal structure or internal function of anatomical area, and are one of the main bases for modern medical diagnosis. There are many types of images in the medical field, and they are greatly affected by the environment of the equipment, which will affect the doctor's diagnosis to a certain extent. Introducing deep learning into medical image detection, training the network based on imaging data and theoretical guidance, and improving the accuracy of diagnosis. Traditional segmentation and classification methods are mainly based on supervised learning and good matching of images or voxel labels, relying on large-scale unlabeled images of healthy subjects. 2D/3D single medical image reconstruction to detect outliers in the learned feature space or from high reconstruction loss.

Deep learning has been successful in retinal disease detection, but usually relies on large-scale labeled data. To break this limitation, Kang et al. [16] proposed a sparse constrained generative adversarial network (Sparse-GAN) for image anomaly detection using only health data. Sparse-GAN maps the reconstructed image into the latent space and attaches an encoder to reduce the effect of image noise, is able to predict anomalies in the latent space rather than image-level anomalies, and is also constrained by a novel sparse regularization network. The feasibility of OCT image anomaly detection and the effectiveness of the method are verified by public datasets, and the abnormal activation map of lesions is displayed, which makes the results more interpretable. Han et al. [17] proposed MADGAN, a two-step unsupervised medical anomaly detection method based on GAN-based multi-slice reconstruction. Combined with the WGAN-GP gradient penalty term and the L1 loss, train on three healthy brain MRI axial slices and reconstruct the next three slices, the L1 loss only generalizes well to unseen images

with similar distribution to the training images, and the WGAN-GP loss captures recognizable structures. Since squared error is sensitive to outliers, L2 loss is used to clearly distinguish healthy samples from abnormal samples. Using 1133 healthy T1 MRI scans for training, the AUC was 0.727 when AD was detected in early MCI and 0.894 when AD was detected late. Based on a GAN model, Chen et al. [18] realized the detection of diseased regions in an unsupervised manner by learning the brain MRI data distribution of healthy subjects. The model is trained using T2-weighted health MRI images extracted from the Human Connectome Project dataset. The generator uses an adversarial autoencoder (AAE) and a variational autoencoder (VAE) to generate the health data distribution. The discriminator detects the lesion area by the pixel-wise intensity difference between the original image and the reconstructed image. The results showed that the AUC of the AAE model reached 0.923.

3.3 Remote Sensing Image Detection

Remote sensing image object detection has a wide application prospect in environmental supervision, military, transportation, civil industry and other fields. With the development of remote sensing platforms and high-performance sensors, the detailed information of ground objects obtained is more abundant. However, the traditional object detection algorithm is not ideal in the case of variable environment, complex background, object aggregation, too many small objects and so on, and could not extract valuable information.

Li et al. [19] proposed a remote sensing image object detection model Attention-GAN-Mask R-CNN based on attention mechanism and generative adversarial network. The model introduced a generative adversarial network in the Mask branch. The generators in the adversarial network are defined the same, so use a separate generative adversarial network to pre-train the Mask generation network of the Mask branch, thereby improving the accuracy of the generator in the original Mask branch. Lin et al. [20] proposed a SAR image ship object recognition method based on GAN pre-training CNN. Under the condition of limited training data, GAN was used to generate samples of corresponding categories, and then real samples with category annotations were used for fine-tuning to achieve higher feature extraction capability. The MSTAR dataset proves that the algorithm has good classification and recognition performance for multi-class objects.

To solve the problem of low detection performance of small objects in remote sensing images, Ahmad et al. [21] constructed a novel end-to-end FPN-GAN network architecture to solve the problem of small object detection. In the generator network, the feature pyramid is combined with the convolution layer, and the least squares loss is used for both global and local images in the discriminator network [22]. In order to improve the quality and efficiency of the model, Resnet-50 is used as the backbone network architecture. Through the experiments on the large-scale benchmark dataset DIOR [23] of optical remote sensing image object detection, the performance of the model in terms of accuracy, precision, recall, and validation loss is analyzed, and the superiority of the method is verified. Rabbi et al. [24], inspired by Edge Enhancement GAN (EEGAN) and ESRRGAN, studied a novel Edge Enhancement Super-Resolution GAN (EESRRGAN) to improve the quality of remote sensing images and train the network in an end-to-end

manner. The whole architecture consists of EESRGAN network and detector network. For generator and edge enhancement network, residual dense block (RRDB) [25] is used. These blocks contain multi-level residual networks with dense connections and perform well in image enhancement. And the Charbonnier loss [26] is used in the edge-enhanced network, and finally different detectors are used to detect small objects from SR images. The method is applied to the created oil and gas storage tank (OGST) dataset [27] and COWC dataset [28], and the detection performance of different use cases is compared. The results show that the method is superior to the latest research results.

3.4 Face Detection

In recent years, face detection has been applied to people's daily life. With the rapid development of deep learning, a large number of face detection algorithms have emerged. However, due to the gradual expansion of the application scope and the complex use scenarios, the current technology have problems of misjudgment in the case of low resolution, angle, occlusion, different face image styles, and face forgery. The emergence of generative adversarial networks has played an important role in solving the above problems.

Generative adversarial network improves the face detection effect in different environments and scenes by using context information, super-resolution reconstruction, image enhancement and other methods. SRGAN [29] is the first deep learning algorithm to apply generative adversarial network to the field of super-resolution reconstruction. However, in low-resolution face images, the obtained images are blurry and lack details. Bai et al. [30] proposed based on GAN end-to-end convolutional neural network. In the generator, using a super-resolution network (SRN) to upsample small faces, and make use of the surrounding region information of the face cropped by enlarged window to train GAN, but there are still problems of ambiguity and lack of detailed information. Then further introduce an improved sub-network (RN) to restore the missing information and generate high-resolution images. In the discriminator, designed a new loss function to complete the discrimination of real/fake face and face/non-face. Zhang et al. [31] proposed a Contextual based Generative Adversarial Network(C-GAN), which added a regression branch to improve the border detection of difficult faces. Through ablation experiments, the effectiveness of C-GAN on blurring small faces is verified. Gu [32] proposed a de-occlusion architecture with generative adversarial network as the main body. By improving U-net network, convolution with padding was used, and the edge information was fully utilized to improve the network performance. The improved SU-net network is used as generator network, which effectively improves the face detection accuracy.

4 Conclusion

This paper reviews the basic theory of GAN and its research progress, focusing on a systematic review of its application in object detection fields, such as industrial defect detection, medical image detection, remote sensing image detection, and face detection. As a generative model, GAN provides a good solution to the problems of insufficient

samples, low image resolution, and difficulty in feature extraction, and it could make the network more robust to occlusion and deformation problems, improve the accuracy of detection.

Object detection has very important value for the current information society. The application of GAN in object detection could be deeply explored. It could innovate the algorithm, select an appropriate loss function, optimize the network structure, and then combine with specific application scenarios to improve the real-time and accuracy of detection. GAN could be used for sample generation to expand the datasets to address the lack of training data for many object detection scenarios. Improving the interpretability and the evaluation criteria of the model also have very important research value.

References

1. Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al.: Generative adversarial nets. In: Neural Information Processing Systems. MIT Press (2014)
2. Mertens, J.F., Zamir, S.: The value of two-person zero-sum repeated games with lack of information on both sides. *Int. J. Game Theory* **1**(1), 39–64 (1971)
3. Ratliff, L.J., Burden, S.A., Sastry, S.S.: Characterization and computation of local Nash equilibria in continuous games. In: Proceedings of the 51st Communication, Control, and Computing (Allerton), Monticello, IL, USA, pp. 917–924. IEEE (2013)
4. Arjovsky, M., Bottou, L.: Towards principled methods for training generative adversarial networks. [arXiv:1701.04862](https://arxiv.org/abs/1701.04862). <https://arxiv.org/abs/1701.04862>
5. Wang, B., Liu, K., Zhao, J.: Conditional generative adversarial networks for common-sense machine comprehension. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence, 19–25 August 2017, Melbourne, Australia, Melbourne: IJCAI, pp. 4123–4129 (2017)
6. Li, J., Liang, X., Wei, Y., Xu, T., Feng, J., Yan, S.: Perceptual generative adversarial networks for small object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1951–1959 (2017). <https://doi.org/10.1109/CVPR.2017.211>
7. Bai, Y., Zhang, Y., Ding, M., Ghanem, B.: SOD-MTGAN: small object detection via multi-task generative adversarial network. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Computer Vision – ECCV 2018. LNCS, vol. 11217, pp. 210–226. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01261-8_13
8. Song, K.: Research and implementation of abnormal detection in aircraft agent based on machine learning. University of Electronic Science and Technology of China (2022). <https://doi.org/10.27005/d.cnki.gdzku.2022.002510>
9. Mirza, M., Osindero, S.: Conditional generative adversarial nets. [arXiv:1411.1784](https://arxiv.org/abs/1411.1784) (2014)
10. Yarlagadda, S.K., Güera, D., Bestagini, P., Maggie Zhu, F., Tubaro, S., Delp, E.J.: Satellite image forgery detection and localization using GAN and one-class classifier. *Electron. Imaging* **2018**(7), 214-1–214-9 (2018)
11. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: Niethammer, M., Styner, M., Aylward, S., Zhu, H., Oguz, I., Yap, P.-T., Shen, D. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 146–157. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_12
12. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: Ganomaly: semi-supervised anomaly detection via adversarial training. In: Jawahar, C.V., Li, H., Mori, G., Schindler, K. (eds.) ACCV 2018. LNCS, vol. 11363, pp. 622–637. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20893-6_39

13. Li, J.: Research on defect inspection of wooden floor based on unsupervised learning. Beijing Jiaotong University (2021). <https://doi.org/10.26944/d.cnki.gbfnj.2021.000622>
14. Tsai, D.M., Fan, S.K.S., Chou, Y.H.: Auto-annotated deep segmentation for surface defect detection. *IEEE Trans. Instrum. Meas.* **70**, 1–10 (2021)
15. Liu, L.: Research on surface defect detection algorithm based on deep learning. Huazhong University of Science & Technology (2019). <https://doi.org/10.27157/d.cnki.ghzku.2019.001438>
16. Zhou, K., et al.: Sparse-Gan: sparsity-constrained generative adversarial network for anomaly detection in retinal OCT image. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 1227–1231 (2020). 10.1109/ISBI45749.2020.9098374
17. Han, C., Rundo, L., Murao, K., et al.: MADGAN: unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction. *BMC Bioinform.* **22**, 31 (2021). <https://doi.org/10.1186/s12859-020-03936-1>
18. Chen, X., Konukoglu, E.: Unsupervised detection of lesions in brain MRI using constrained adversarial auto-encoders (2020). <https://arxiv.org/abs/1806.04972>
19. Li, J., Deng, Y., Wu, X., et al.: Object detection in remote sensing image based on attention mechanism and GAN. *Comput. Syst. Appl.* **31**(6), 182–191 (2022). <http://www.c-s-a.org.cn/1003-3254/8490.html>
20. Lin, Z.: Ship detection and recognition in remote sensing images based on deep learning. National University of Defense Technology (2018). <https://doi.org/10.27052/d.cnki.gzjgu.2018.000972>
21. Ahmad, T., Chen, X., Saqlain, A.S., Ma, Y.: FPN-GAN: multi-class small object detection in remote sensing images. In: 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), pp. 478–482 (2021). <https://doi.org/10.1109/ICCCBDA51879.2021.9442506>
22. Jolicoeur-Martineau, A.: The relativistic discriminator: a key element missing from standard GAN (2018)
23. Li, K., Wan, G., Cheng, G., Meng, L., Han, J.: Object detection in optical remote sensing images: a survey and a new benchmark. *ISPRS J. Photogram. Remote Sens.* **159**, 296–307 (2020)
24. Rabbi, J., Ray, N., Schubert, M., Chowdhury, S., Chao, D.: Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network. *Remote Sens.* **12**(9), 1432 (2020). <https://doi.org/10.3390/rs12091432>
25. Wang, X., et al.: ESRGAN: enhanced super-resolution generative adversarial networks. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018. LNCS, vol. 11133, pp. 63–79. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11021-5_5
26. Charbonnier, P., Blanc-Féraud, L., Aubert, G., Barlaud, M.: Two deterministic half-quadratic regularization algorithms for computed imaging. In: Proceedings of International Conference on Image Processing, vol. 2, pp. 168–172 (1994)
27. Rabbi, J., Chowdhury, S., Chao, D.: Oil and Gas Tank Dataset. In Mendeley Data, V3 (2020). <https://data.mendeley.com/datasets/bkxj8z84m9/3>
28. Mundhenk, T.N., Konjevod, G., Sakla, W.A., Boakye, K.: A large contextual dataset for classification, detection and counting of cars with deep learning. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 785–800. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46487-9_48
29. Ledig, C., Theis, L., Huszár, F., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4681–4690 (2017)
30. Bai, Y., Zhang, Y., Ding, M., et al.: Finding tiny faces in the wild with generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 21–30 (2018)

31. Zhang, Y., Ding, M., Bai, Y., et al.: Detecting small faces in the wild based on generative adversarial network and contextual information. *Pattern Recogn.* **94**, 74–86 (2019)
32. Gu, J.: Design and Implementation of Yujiao Robot Face Recognition System. Chongqing University (2019). <https://doi.org/10.27670/d.cnki.gcqdu.2019.001616>