International Journal of Applied Mathematics in Control Engineering 5 (2022) 47-52

Contents lists available at YXpublications

International Journal of Applied Mathematics in Control Engineering

Journal homepage: http://www.ijamce.com

Data Enhancement Based on Attention Mechanism and Generative Adversarial Networks

Hongyu Shi[a,*], Jiawei Tan[a]

[a] *School of Mathematics and Statistics, Changchun University of Technology, Jilin, China*

ARTICLE INFO

Article history:
Received 21 January 2022
Accepted 28 March 2022
Available online 30 March 2022

Keywords：
Generative adversarial networks
Attention mechanism
Image generation
Data enhancement

ABSTRACT

In recent years, the communication between human and computer has reached the inseparable degree. At present, generative adversarial network is widely used in the field of image generation with excellent performance. To solve the problem of low recognition rate of deep learning defect detection algorithm under small sample conditions, a data enhancement model (AttGAN) based on the combination of attention mechanism and generative adversation network was proposed. Firstly, the discriminator adopts CNN model incorporating attention mechanism as feature vector to enhance feature extraction ability. Secondly, the features extracted by the discriminator are sent to the generator as step-by-step guidance signals to guide the generator to generate images, making the generated images more inclined to the reference image. Finally, the generator completes the image generation and passes it to the discriminator to judge whether the image is true or false, in order to confirm whether the image conforms to the standard. Experimental results show that when attention mechanism is incorporated into the generative adversace network model, the image contains global information, and the feature extraction performance of the model is significantly improved.

1. Introduction

As the upstream task of image recognition, object detection and semantic segmentation, image processing has important research significance. The current data enhancement is introduced from the classification task, but the enhancement of the classification task is not necessarily applicable to the detection task. As the number of transformations becomes larger and larger, it becomes difficult to combine them efficiently manually. Machine learning is used to search for a combination of strategies that are more suitable for target detection tasks. Because the labeling cost of target detection data is higher than that of image classification, data enhancement strategy can not only improve model performance but also save data cost in the case of limited data.

Traditional data enhancement methods include inversion, rotation, scaling, clipping, rotation distortion transformation and polar coordinate transformation. Although the above method can realize data enhancement, it cannot effectively improve the performance of the detection network because of the repeated deep features and a lot of redundant information generated in the adversarial network. Generative adversarial network (GAN) was proposed by Goodfellow et al in 2014(Goodfellow et al.,2014). It has remarkable effect in data enhancement and is regarded as the most imaginative deep learning network model in the recent decade. Conditional generative adjudgment network (CGAN) with conditional constraints can annotate the generated image directly and generate the image with the specified label, but the training is difficult and the stability of image generation is poor (Mirza and Osindero 2014). Conditional generative adversarial networks have only one output for determining the truth and falsity of discriminator D but a semi-supervised learning primitive adversarial network based on conditional generative adversarial networks (Semi-GAN) can increase the number of data categories output by the discriminator to K+1(Odena A et al.,2016).However, as research has found that GAN has structural flaws, such as the difficulty of Nash equilibrium, the ease of gradient satiation, the training instability, unclear features of generated samples, etc., the DCGAN(the Radford, Alec et al.,2016) is a new approach to the original GAN structure and introduces a convolutional neural network (CNN), which combines supervised and unsupervised learning, effectively improving the GAN training stability. In order to stabilize the training process, the 2016 year, a loss function using least squares was proposed for LSGAN model, which is based on the principle of improving the generated data by penalizing the generated data that deviate from the discriminator network judgment boundaries and thus improving the generated data. The principle of LSGAN model is to improve

* Corresponding author.
E-mail addresses: 2227678478@qq.com (H. Shi)
doi:

the generated data quality (no gradient contribution) by penalizing the generated data that deviate from the discriminator network judgment boundary (Mao X et al.,2017).In 2017,Antreas Antoniou proposed by DAGAN model (Antoniou A et al.,2017), where the image features are extracted and used as input to a generator whose input is projected to a lower level projection, and then the random vector z are transformed, linked and fused with the bottleneck vectors, and then up-dimensioned by the network to produce an enhanced image. Picon A et al. in a study of Res Net50 improved the disease identification of wheat samples taken in the field with an accuracy of 96% (The Picon A,2019).

Chen (Chen et al.,2018) and Zhao (Zhao et al.,2017) Although the proposed model has superior segmentation performance, its feature extraction module extracts contextual information for the whole feature region through a non-adaptive approach, ignoring the differences in context dependence of different regions and kinds of objects, and is vulnerable to the limitations of complex scenes. In order to integrate the global pixel-level up contextual information together, scholars have proposed the self-attentive mechanism (Veličković et al.,2018) method by estimating the full image context information representation to enhance the traditional convolutional features.

Wasserstein GAN (WGAN) with gradient punishment successfully solved the problem of GAN training instability, ensuring the diversity of generated samples, but resulting in blurred details of generated images. Generated against network can generate diverse images, good results have been achieved in the field of data enhancement, can effectively enhance the performance of deep learning in affective computing, based on this, this paper use generated against network and attention mechanism for Fashion to enhance MNIST and MNIST data, in order to improve classification ability and the accuracy of the model.

2. Construction of data enhancement model

2.1 Generative adversarial networks

Generative adversarial networks mainly contain two structures: Generator (G) and Discriminator (D). To put it simply, the main function of generator G is to generate different samples according to different inputs, while the discriminator D is to judge whether the generated samples are "similar" to the real samples. In other words, determine whether the input data is real data. The training of both is in a state of game

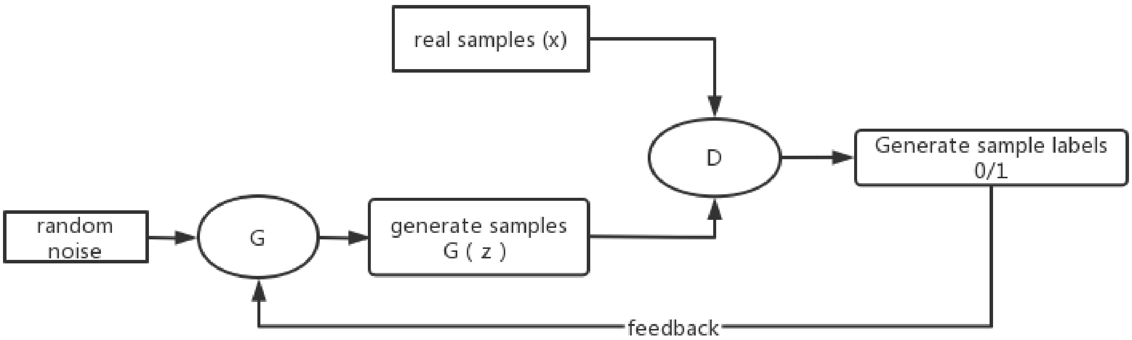


Fig. 1 The structure of Generative adversarial networks

In essence, they are two forms of mapping functions, which are constructed by neural network structure to realize the mapping from hidden variable space to sample space. The connection between G and D is mainly established through the loss functions D-loss and G-Loss. By adjusting the parameters of G and D according to the obtained discriminator probability, the pseudo sample distribution generated by G in the process is getting closer and closer $x \sim P_{data}(x)$, and accordingly, D is also improving the discriminant ability of sample authenticity. They are constantly optimized in the game and finally reach Nash equilibrium. At this point, G can reasonably fit the real sample distribution and generate G(z) that D cannot distinguish from X.

Assume pseudo samples $G(z) = x'$,and obey distribution $x' \sim P_g(x)$. The goal of GAN is $P_g(x)$ and $P_{data}(x)$ to be as similar as possible. During model training, the input sample outputs the corresponding probability through the mapping function D(x) of D, so as to determine whether the input sample is x or x'. D tries to classify x' into the pseudo sample class, while G aims to classify x into the real sample class. The logarithm of the discriminant result is taken and the mathematical expectation is expressed as follows:

$$V(G,D) = E_{x\sim P_{data}(x)} \log D(x) + E_{x\sim P_g(x)} \log(1 - D(x)) \tag{1}$$

where, E represents mathematical expectation, that is, the objective function $V(G,D)$ is expressed in the form of mathematical expectation. The first term represents the discriminant expectation of the discriminant on the real sample, and correspondingly, the second term represents the discriminant expectation of the discriminant on the pseudo sample. Since the objectives of D and G are opposite, in order to adapt to this situation, the optimal parameters of the model are solved and the minimum-maximum method is adopted to optimize the objective function. Therefore, the objective function of the training of the two is:

$$\min_{G}\max_{D} V\left(G,D\right) = E_{x\sim P_{data}(x)}\log D(x) + E_{x\sim P_g(x)}\log(1-D(x)) \quad (2)$$

The goal of D is to get D(x) to output as much as possible 1 for the sample from $P_{data}(x)$, 0 for the sample from $P_g(x)$, and the goal of G is just the opposite. A larger value of $V(G,D)$ indicates better performance of the discriminator and worse performance of the generator. On the contrary, when D is difficult to distinguish authenticity, the value of $V(G,D)$ will decrease rapidly. When optimizing the parameters of G, fixed the maximum D parameter last time and minimized the objective function value at this time, namely:

$$G' = \arg\min_{G}[\max_{D} V(G,D)] \quad (3)$$

where, G' is the optimal generative synthesizer, and the principle of solving the optimal discriminator D' is similar.

2.2 Wasserstein Generative adversarial network

Wasserstein GAN improved the traditional GAN from the perspective of measuring the distance function between distributions. As the traditional GAN uses KL divergence and JS divergence to calculate the distance between two distributions, the training process of the traditional GAN is very unstable and prone to mode collapse, which directly leads to the limitation of the application of the traditional GAN. Based on the above questions, Wasserstein distance measures the difference between the generated data distribution and the real data. Wasserstein distance is also known as EM distance, which is defined as the lower bound of the expected distance between samples within all possible value ranges under the joint distribution Y, as shown in Formula (4).

$$W(P_{real},P_g) = \inf_{y\sim\prod(P_{real},P_g)} E_{(x,y)\sim y}[\|x-y\|] \quad (4)$$

Compared with KL divergence and JS divergence, Wasserstein distance is more sensitive to the variation of difference, which can provide a more robust gradient guidance for the model, and the optimization target does not tend to be constant when the distribution overlap surface is small, thus avoiding the gradient disappearance problem. However, since Wasserstein distance is difficult to be solved directly, variation is adopted as the distance measurement between distributions, as shown in Formula (5).

$$W(P_{real},P_g) = \sup_{\|f_L\le 1\|}(E_{x\sim P_{data}}\|f(x)\| - E_{x\sim P_x}\|f(\hat{x})\|) \quad (5)$$

In the above formula $\|f_L \le 1\|$, function f needs to satisfy the 1-Lipschitz constraint. At this point, f can be seen as a discriminator module composed of neural network, and the above formula is established through constant range constraint on the parameters of the discriminator θ_d. The loss function of WGAN is shown in formula (6).

$$\min_{G}\max_{D} L(G,D) = E_{x\sim P_{data}}[D(x)] - E_{\hat{x}\sim P_x}[D(\hat{x})] \quad (6)$$

2.3 Attention Mechanism

According to its different scope, the attention mechanism can focus on the feature information of the image in the channel dimension and space dimension respectively by the channel attention mechanism and spatial attention mechanism.

2.3.1 Channel attention mechanism

The branching of channel attention produces feature information $M_c(F)\in R^c$ by acquiring features between different channels. Since each channel contains specific feature responses, in order to obtain feature maps in different channels, Global Average Pooling (GAP) is adopted to obtain channel vectors $F_c \in R^{c\times1\times1}$, which encode Global feature information in channels. To calculate the channel injection force between channel vectors。This paper uses Multi-Layer Perception (MLP) with a hidden Layer. To save parameter overhead, set the hidden activation size to $R^{c/r\times1\times1}$, where r is the reduction ratio. After MLP, a Batch Normalization (BN) layer was added to leverage spatial branching output for scaling.

2.3.2 Spatial attention mechanism

The branch of spatial attention produces feature information $M_s(F)\in R^{H\times W}$ by focusing on the contextual feature information in different spatial locations, measuring the importance of different feature information according to the weight, and establishing dynamic weight parameters by making relevant and irrelevant choices of feature information, so as to strengthen key information and weaken useless information. Firstly, the input features are projected onto the reduced size of 1×1 convolution to combine $F_c \in R^{c\times H\times W}$ and compress the feature graphs $R^{c/r\times H\times W}$ in the whole channel dimension. Then, two Dilated Convolution (DConv) of 3×3 cavities with a void rate of 4 is used to effectively utilize the context information. The void Convolution can expand the size of the receptive field in the spatial attention module and aggregate the feature information of the image. Finally, 1×1 convolution is used to simplify the features and improve the operation efficiency of the network.

2.4 AttGAN model

2.4.1 Network Framework

The network structure of AttGAN is shown in Figure 1, where AttGAN has a generator, a discriminator, and data enhancement.

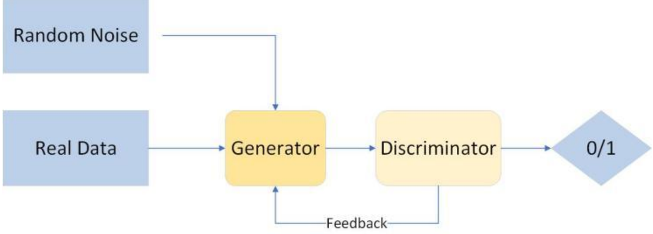


Fig. 2 Network structure of AttGAN

The input Gaussian noise of the generator is connected to the feature layer through the input layer, and then the feature layer can be obtained through the transpose convolution module. Finally, the TANH activation function is used to activate it and output the image. The transpose convolution module consists of transpose convolution layer, batch standardization layer and ReLU activation layer. The size of transpose convolution kernel is 3×3 and the step size is 2.

The function of discriminator is to retain more image content and restore details in the process of image generation. Its network structure is shown in Figure 2. The Network structure is input into the image, and the feature extraction layer adopts the improved Visual Geometry Group Network 16 (VGG-16) feature extraction layer. A convolution layer is added after the input layer. The size of the convolution kernel is 3×3 and the step size is 2. The receptive field of the network is increased to extract features of higher resolution images. The pooled layer is followed by batch standardization and Leaky ReLU activation functions, and the final layer uses the full connection layer to convert features into vectors. Then activated by Sigmoid layer, the authenticity of sample images and generated images can be identified. The purpose of the three repetitions is to increase the depth of the network and thereby improve feature extraction capabilities, providing additional flexibility in processing different content.

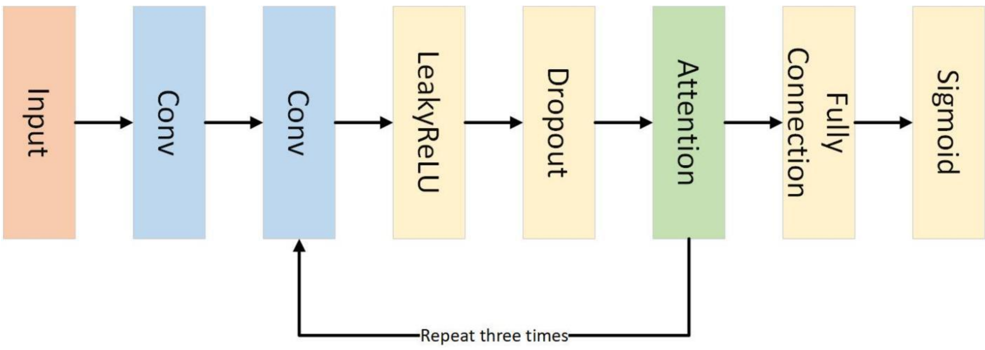


Fig. 3 The network structure of the Discriminator

AttGAN LOOP：

/iterations is the total number of iterations/

while number less than iterations do

/k is the number of times each iteration of the discriminator is executed/

for t=1,...,k do

/m is the batch size/

for i=1,...,m do

/ z_i is the noise , $\hat{y}_i$ is the result of the discriminator/

$$\hat{y}_i \leftarrow G(z_i)$$

/ y_i is the real results, λ is the weight loss/

$$L^{(i)} \leftarrow \log(y_i) + \log(1 - D(\hat{y}_i)) + \lambda \min(\|y_i - \hat{y}_i\|_2)$$

/ α、 β_1、 β_2 is hyperparameter of optimization function, w is hyperparameter of discrimination/

$$w \leftarrow Optim(\nabla_w \frac{1}{m}\sum_{i=1}^{m} L^{(i)}, w, \alpha, \beta_1, \beta_2)$$

end for

/ θ is the parameter of generator/

$$\theta \leftarrow Optim(\nabla_\theta \frac{1}{m}\sum_{i=1}^{m} \log(1 - D(\hat{y}_i)), \theta, \alpha, \beta_1, \beta_2)$$

end while

3. Experiment and Analysis

3.1 Datasets and Experimental design

Deep learning is adopted as the framework and python3.7 (Anaconda) is used for comparative experimental analysis. The overall working flow of the experiment is shown in the figure, and the processing steps are as follows:

1. Data preparation: MNIST data set was selected as the experimental data set;

2.AttGAN model training: network structure is designed and two data sets are used for model training and image generation;

3. Performance evaluation: Use the generated quality evaluation indicators to evaluate the quality of the samples generated by AttGAN.

In the experimental part, linear mapping to 128 dimensions is used in generators and discriminators to facilitate the linking of subsequent convolution layers. The experimental results of AttGAN and WGAN are compared with the original scheme GAN for image generation. The GAN implementation of AttGAN is trained using Adam optimizer, and the discriminator and generator use different learning rates, where the discriminator learning rate is set to 0.0004, the generator learning rate is set to 0.0001, and the hyperparameters beta1 and beta2 are set to 0.0 and 0.9, respectively. AttGAN's WGAN implementation uses RMSprop optimizer training, and the discriminator and generator learning rates are set to 0.0002 to achieve better results.

3.2 Experimental results and Analysis

1. MNIST

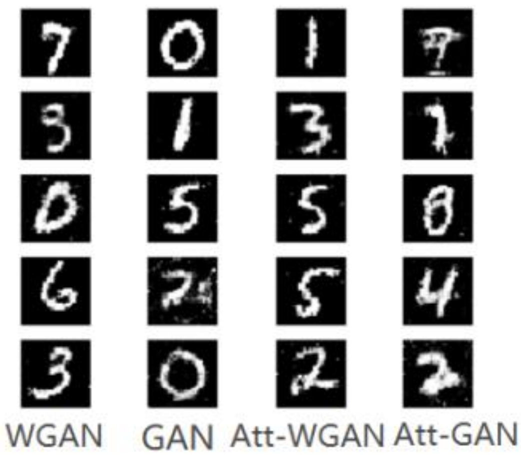


Fig 4 MNIST data generation result comparison

Table 1 The accuracy of data generated by different models on MNIST dataset

Model	Accuracy(%)
WGAN	84.52
GAN	83.96
Att-GAN	84.59
Att-WGAN	85.54

It can be concluded from Figure 3 and Table 1 that the ATT-WGAN model performs better than other comparison models in this experiment. As the number of iterations increased, the data generated by each model became clearer gradually. In the process of unified iteration 30001 times, it was found that the generation time of WGAN was shorter in the process of data generation. And they showed better results. Att-gan and ATT-WGAN, the ACC evaluation indexes used in this paper, have better results.

2. Fashion MNIST

The generated data set is shown in Figure 4, from which it can be seen that: The images generated by GAN and WGAN methods have problems such as insufficient processing of high-frequency information such as texture and pixel, and unclear details. In contrast, AttGAN method with attentional mechanism can effectively process different features and regions, improve the response ability of features, and make the generated images clearer.

It can be concluded from Figure 4 and Table 2 that the PERFORMANCE of Att-WGAN model in this experiment is superior to other comparison models. As the number of iterations increased, the data generated by each model gradually became clear. In the process of unified iteration 30001, it was found that the generation time of WGAN was shorter and the effect was better in the process of data generation, but the application effect on Fashion MNIST data set was not as good as that on MNIST data set. Therefore, generative adversarial network with attention mechanism is more suitable for simple data distribution.

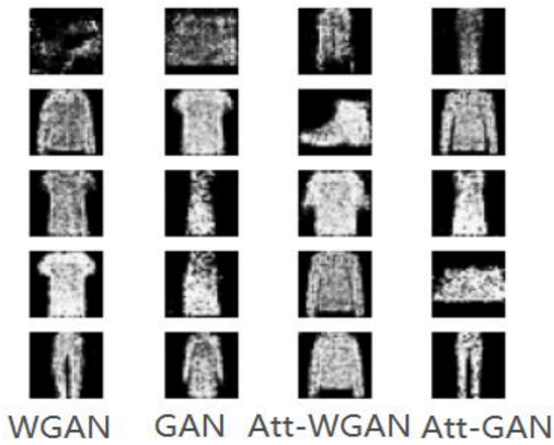


Fig 5 The result comparison of Fashion MNIST data generation

Table 2 The accuracy of data generated by different models on Fashion MNIST dataset

Model	Accuracy(%)
WGAN	83.98
GAN	83.56
Att-GAN	84.03
Att-WGAN	85.19

In the experiment, WGAN, GAN, ATT-gan and ATT-WGAN were used to generate pictures of different numbers and clothes and carry out relevant experiments, and the experimental results proved the effectiveness. Firstly, image generation and data enhancement experiments were carried out on Fashion MNIST and MNIST data sets to verify the effect of model generated images. The quality and diversity of generated images are measured by comparing feature vectors between different images. Experimental results show that ATT-WGAN generates better images. Compared with the other three methods, this method can effectively improve the quality of the generated image.

4. Conclusion

In this paper, a generative adversarial network AttGAN based on attention mechanism is implemented to improve the clarity of generated images. At the same time, two kinds of generative adversarial networks, namely GAN and WGAN, are used to implement the generative adversarial networks of attention mechanism, namely Att-GAN and Att-WGAN, and the GAN and WGAN schemes are applied to Fashion MNIST and MNIST data sets and their effects are compared. Among them, the generated against the attention mechanism in the network in the middle of the original convolution layer are calculated by different local pixel area, the relationship between the main solved the convolution operation can handle only partial correlation between pixels, and cannot be calculated over long distances, multi-layer correlation between local defects, and with the aid of attention mechanism to generate higher image quality. In addition, the generative adversarial network framework composed of multiple generators and discriminators is the current research trend. Therefore, the future research will focus on the introduction of multiple generators and discriminators generative adversarial network to improve the quality and effect of generated images.

References

[1] Ma Jiayi, Yu Wei, Liang Pengwei, et al. Fusion GAN: A generative adversarial network for infrared and visible image fusion [J]. Information Fusion, 2019, 48 (4): 11–26.

[2] GOODFELLOW I J, POUGET-ABADIE J, MIRZAM, et al. Generative adversarial nets [C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada: NIPS, 2014: 2672-2680.

[3] Mechrez R, Talmi I, Zelnik-Manor L. The contextual loss for image transformation with non-aligned data[C]. Proceedings of the European conference on computer vision (ECCV), 2018: 768-783.

[4] Antoniou A, Storkey A, Edwards H. Data augmentation generative adversarial networks[J]. ar Xiv preprint ar Xiv:1711.04340, 2017.

[5] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. ar Xiv preprint ar Xiv:1511.06434, 2015.

[6] Picon A, Alvarez-Gila A, Seitz M, Ortiz-Barredo A, Echazarra J, Johannes A. Deep convolutional neural networks for mobile capture device-based crodisease classification in the wild[J]. Computers and Electronics in Agriculture, 2019, 161: 280-290.

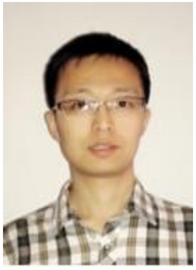
[7] Odena A. Semi-supervised learning with generative adversarial networks. ar Xiv: 1606.01583, 2016.

[8] Chen L C, Papandreou G, Kokkinos I, et al. Deep Lab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. //IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4):834-848.

[9] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]// Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Hawaii: IEEE, 2017: 2881-2890.

[10] Hung W. C., Tsai Y. H., Liou Y. T., et al. Adversarial learning for semi-supervised semantic segmentation[J]. //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1-17.

Hongyu Shi is currently studying for her master's degree at the School of Mathematics and Statistics, Changchun University of Technology. Her research interest covers machine learning and generative adversarial networks, optimization, and big data analysis.



Jiawei Tan currently studying for her master's degree at the School of Mathematics and Statistics, Changchun University of Technology. His research interest covers Numerical solution of partial differential equations.