

Received April 3, 2020, accepted April 22, 2020, date of publication May 11, 2020, date of current version June 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2993887

# Research on Medical Image Classification Based on Machine Learning

HAI TANG<sup>1</sup> AND ZHIHUI HU

School of Electrical and Information Engineering, Hubei University of Automotive Technology, Shiyan 442002, China

Corresponding author: Hai Tang (smile\_tangh@126.com)

This work was supported in part by the Research Project of Hubei Provincial Department of Education (Research on Active Safety Technology for Anti-Slip Rolling of Vehicles in Curved Roads) under Grant B2018070, in part by the Key Laboratory of Automotive Power Train and Electronics, Hubei University of Automotive Technology (Research on Active Safety Technology System for Vehicle Curved Driving), under Grant ZDK1201603, and in part by the Doctoral Research Startup Fund, Hubei University of Automotive Technology (Research on Key Technologies of Structured Road Lane Detection), under Grant BK201202.

**ABSTRACT** In this paper, we propose a new method for CT pathological image analysis of brain and chest to extract image features and classify images. Because the deep neural network needs a large number of labeled samples to complete the training, and the cost of medical image labeling is very high, the training samples needed to train the deep neural network are insufficient. In this paper, a semi supervised learning based image classification method is proposed, which uses a small amount of labeled pathological image data to train the network model, and then integrates the features extracted by the network to classify the image. The results show that the classification effect of the neural network is better than convolution neural network and other traditional image classification models. To some extent, it can reduce the dependence of neural network on a large number of training samples, and effectively reduce the over fitting phenomenon of the network. Through the analysis of the overall classification accuracy and kappa coefficient of different classification methods under different sample numbers, it is found that the overall classification accuracy and kappa coefficient are increasing with the increasing number of training samples. Especially in the case of a small number of training samples, compared with other deep neural networks and traditional classification methods, the classification accuracy of the counter neural network is about 10% higher than that of other neural networks and traditional classification methods, and the advantages are more obvious.

**INDEX TERMS** Generative adversarial network, deep learning, feature extraction, image classification.

## I. INTRODUCTION

With the increasing demand for faster and more accurate treatment, medical imaging plays an increasingly important role in the early detection, diagnosis and treatment of diseases. Thanks to the development of physics, electronic engineering and computer science and technology, the resolution of medical image is higher and higher, and the image mode is more and more abundant. At the same time, the number of medical image is increasing rapidly. At present, X-ray imaging, computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET / CT), ultrasound imaging are widely used in clinic.

The key to achieve accurate diagnosis and treatment is the accurate interpretation of medical images, but the interpretation of images highly depends on the subjective judgment of doctors, so doctors at different levels have great deviation on the results of image interpretation. In recent years,

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei<sup>1</sup>.

with the emergence of a large number of labeled natural image data sets and the breakthrough of deep learning in computer vision, image classification, target detection and image segmentation have been significantly improved. Many researches have been carried out on early disease detection and diagnosis based on supervised learning, Cireşan [1] applied deep neural network to medical image analysis, which played an important role in skin cancer classification [2], breast cancer diagnosis [3], brain tumor segmentation [4]. Hinton improved the deep convolution neural network and applied it to medical image analysis [5], [6]. The experimental results of Hafemann *et al.* [7] show that the feature extraction ability of convolutional neural network is better than that of traditional texture descriptors, and the accuracy of image recognition is improved by maximizing pooling, multiplying, summing and other operations. Because the convolution neural network can only segment the general contour of the object, and the result of edge processing is poor, Jonathan proposed the full convolution neural network based on the convolution neural network, which extends the

classification of image level to pixel level [8]. Olaf improved the full convolution neural network, proposed the u-net network, and applied it to the segmentation of neuron structure in microscope image [9]. Rania used u-net image for breast ultrasound image segmentation, and removed noise area from segmentation results [10].

In addition to the field of medical image classification, deep learning has also made significant progress in the fields of medical image detection [11]–[13], image segmentation [14], [15], image registration [16], [17], image fusion [18], image retrieval [19], [20], image annotation [21], image generation and enhancement [22], [23], computer-aided diagnosis and prognosis [24]–[26]. Deep learning has made significant progress in medical imaging, because of the superior performance of deep learning in data hierarchical feature extraction and feature learning, on the other hand, the large number of labeled data sets. However, in practical application, medical image annotation is significantly influenced by medical professional knowledge, medical industry standards, medical system, etc. The data scale of labeled data set is far from the scale needed, and it needs huge human and material resources for medical image annotation. Therefore, the classification and detection of medical images have been widely concerned by scholars when there are few labeled samples.

In order to improve the accuracy of image detection, saliency detection is widely used in target detection, scene classification, fixation prediction and other visual tasks. In the aspect of feature significance detection, Achanta proposed the local contrast method [27], Hou and Zhang [28] proposed the spectrum residual method, and Harel *et al.* [29] proposed the significance detection based on the graph method. With the development of new features and deep learning, especially the emergence of convolution neural network, cyclic neural network and other deep neural network models, it breaks through the limitations of manual feature extraction in the past, and learns the features from the image independently. Convolution neural network, cyclic neural network and other deep neural networks need a large number of learning samples. Because the labeled samples need certain medical knowledge and clinical experience, the medical image label samples are relatively lack, which is difficult to meet the needs of deep learning.

As a new machine learning model, adversarial neural network effectively combines the generation model and the discrimination model, optimizes the neural network training process through the dynamic game between them, and finally makes the whole network reach the optimal state of Nash equilibrium. Based on this, this paper proposes a semi supervised image classification method based on the generative adversarial network (GAN). In the case of limited labeled samples, a large number of unlabeled sample data is introduced to improve the accuracy of image classification. Combining labeled sample and unlabeled sample data effectively, the feature extraction and image of medical image are

realized in the case of sufficient labeled samples and limited labeled samples Classification.

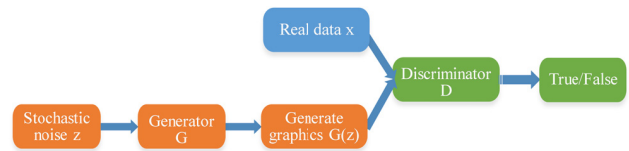


FIGURE 1. Generative adversarial network model [33].

## II. BASIC PRINCIPLE OF ANTAGONISM NEURAL NETWORK

GAN is a hot research field in machine learning in recent two years. It has been widely used in image super-resolution, image segmentation, text to image generation, image restoration and other fields [8], [30]–[32]. GAN is an unsupervised generation model [33] Based on the idea of zero-sum game in game theory put forward by Professor Goodfellow in 2014. Its process is shown in Figure 1. The generative adversarial network (GAN) originates from the two person zero sum game in game theory. It consists of the generating network and the judging network. It generates the network learning sample data for reconstruction, and the judging network estimates the probability that a sample comes from the real data. The model consists of two parts: a generator  $G$  and a discriminator  $D$ . The generator  $G$  maps the random noise  $z$  according to the specific distribution  $P_z$  to the target domain, learns the probability distribution of the real samples  $P_{data}$ , and generates the samples  $G(z)$  according to the distribution  $P_{data}(x)$  of the real data as much as possible. The discriminator judges whether the input samples come from the real data  $x$  or the generated data  $G(z)$ , so as to output a probability value belonging to the real data. The essence of counter neural network training is to make the discriminator recognize the samples generated by the generator as much as possible, and make the samples generated by the generator cheat the discriminator as much as possible. The two networks are optimized alternately in the training process, so as to form a competitive confrontation until both sides reach a dynamic balance.

The feed-forward propagation in GAN network is to transfer the generated samples from  $G$  network to  $D$  network. On the other hand, the back propagation process is that  $G$  network receives the gradient feedback from  $D$  network and updates it, thus completing the information flow between the whole networks. In the training process, in order to let the generator learn the probability distribution  $P_{data}$  of the real sample, we assume the  $P_x$  of the real sample, take the input  $z$  of the noise variable as the input, and get the mapping  $G(z)$  of the data space. Finally, the  $D(x)$  discriminators judge that the sample  $x$  is the probability from the real sample. When the input sample is a real sample, the goal of discriminator  $D$  is to make the output probability  $D(x)$  close to 1; otherwise, when the input sample is a false sample generated by noise, the goal of discriminator  $D$  is to make  $D(G(z))$  close to 0, and the goal of generator  $g$  is to make it as close to 1 as possible. Therefore, the training of antagonistic neural

network is a binary minimax game problem [33], and the objective function is:

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

### III. MEDICAL IMAGE CLASSIFICATION BASED ON ANTAGONISTIC NEURAL NETWORK

#### A. DATA AND PREPROCESSING

The images used in this study are from a data center of a hospital. The database includes 12000 CT images of brain, chest and cervical spine. In order to improve the generalization performance of the model and eliminate the image imbalance, the image was amplified and normalized. Taking brain CT as an example, the CT dataset of the brain includes 5 common brain diseases with high incidence rate in the brain CT images. Taking into account the actual application, 8 types of data are added to normal brain images without brain diseases and other brain diseases with the exception of the 5 high incidence rate. The treatment of chest CT and cervical CT is the same as above. In order to ensure the same size of image input, the image size is unified to 128pixel \* 128pixel. Finally, the selected data set is divided into training set, verification set and test set according to 7:1:2. The training data is divided into labeled sample data and unlabeled sample data, and the labeled sample data is far less than unlabeled sample data. After the network training, the new medical images are divided into normal and sick by using the network, and the classification results of the network are counted.

#### B. NETWORK STRUCTURE

Based on the above network structure model, aiming at the problem of the scarcity of labeled samples in medical image classification, this paper uses GAN network to classify medical images. At present, as a widely used model and branch of deep neural network, convolution neural network has strong processing and representation ability for complex data, which can effectively extract robust and invariant features from big data, and then is conducive to the subsequent image classification. Therefore, this paper uses the combination of convolution neural network and antagonism neural network to generate antagonism network Deep Convolutional Generative Adversarial Networks (DCGAN) [34]. DCGAN is to replace generator and discriminator in antagonism neural network with convolution neural network. In order to improve the feature extraction ability of the network and the stability of model training, the generator  $g$  and the discriminator  $D$  are adjusted as follows:

(1) Cancel all pooling layers. The generator network uses the transposed convolutional layer for up sampling, and the discriminator  $D$  network adds the string convolution to replace the pooling.

(2) Using batch normalization in distinguishing network  $D$  and generating network  $G$ .

(3) Remove FC layer and make the network become full convolution network.

(4) The generated network uses Relu as the activation function, and the last layer uses tanh activation function

(5) The LeakyRelu activation function is used in the discrimination network. The network diagram of generator  $g$  in DCGAN is shown in Figure 2 below:

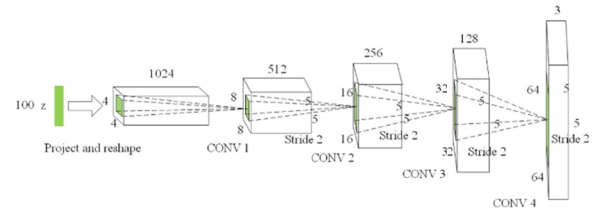


FIGURE 2. Generator network structure [34].

The discriminator network is mainly completed by the full connection layer, which mainly uses the output of the network of feature extraction as the input, and then performs multiple nonlinear mapping on the features. Multiple convolutional neural networks are used to further extract the previous features, and finally the global equalization layer gives the image classification results. In this stage, we should try our best to ensure that the classification model has a high classification accuracy, so that the features extracted from the convolution layer are more discriminative, and then integrate the output features of multiple network layers, and finally give the results of image suspicious area marking.

#### C. NETWORK MODEL TRAINING

The classification network mainly includes the training based on the generation of the network and the training based on the convolution neural network. The network input generated against network training is the result of real pathological effect of brain CT and even sampling. Firstly, the 100 dimensional samples are mapped through the full connection layer to obtain 1024 feature graphs of 4 \* 4 in size. Then, the feature graphs are upsampled through the upsampled network with strides 2, followed by the nonlinear transformation of the network output using the activation layer. The resulting image is 64 \* 64 \* 3. Mark the generated image as 0, the original input image as 1, and then use these marked data to supervise the training of the discriminator. The training process adopts the algorithm of random gradient descent.

In the actual training process, because the analytical solution can't be obtained directly, the iterative numerical method is used to train the model. Because the limited data set may lead to the over fitting of the discrimination network, we choose k-training discriminator  $D$ , 1-training generator  $g$ , and the selection of  $K$  depends on the complexity of discriminator  $D$  and the size of data volume. In the actual experiment,  $K$  is taken as 4. As long as  $G$  transformation is slow enough,  $D$  remains near the optimal solution after every complete training. Finally, the parameters of GAN network model are trained through multiple training rounds.

The noise  $z$  conforming to the Gaussian distribution is sent into the generator network  $G$  to generate the generated

image  $G(z)$  conforming to the real data as much as possible; the generated image sample  $G(z)$  and the database image sample are input into the discriminator network  $D$  together, the database sample data includes a small number of labeled sample data and a large number of unlabeled data. The discriminator  $D$  is composed of multiple convolution layers and full connection layers. Through continuous cross iterative training, the generator  $g$  gradually fits the distribution of database image samples to generate realistic image samples. Meanwhile, the discriminator  $D$ 's classification and discrimination performance for input is constantly improved.

Through the training in the above stages, a stable feature extraction network is obtained. Then the fixed feature extraction network is trained separately. The input image size is  $64 * 64 * 3$ . After the feature extraction network, the feature size is  $3 * 3 * 64$ . Finally, the feature is mapped to 64 dimensions through the global level pooling layer, and then the convolution neural network and the global average pooling layer are used to complete the final classification task.

## IV. TEST RESULTS AND ANALYSIS

### A. TRADITIONAL CLASSIFICATION METHOD

In order to compare the accuracy of traditional classification method and deep learning classification method for image classification, the traditional classification method adopts support vector machine algorithm, and deep learning algorithm selects neural network for confrontation, and then compares the image classification effect of SVM and GAN. The traditional pattern recognition based on classification method is divided into two parts: feature extraction and model selection. The ability of feature selection to express the image itself directly affects the effect of the final image classification. In order to ensure the accuracy of classification, here we select scale invariant feature transform (SIFT) feature as the description operator, and support vector machine as the statistical classifier.

In the process of image classification, the most important thing is to choose a suitable way to describe the characteristics of the image itself. Through the feature engineering of the image, the image is represented as a series of vectors. If the feature vector of an image has a similar pattern to the feature vector of the corresponding category, then the image can be determined as the category, and vice versa. Therefore, the ability of image feature vector to express image itself directly affects the performance of the final image classification system.

Excellent image feature engineering should satisfy the uniqueness of feature description, better anti-interference and high efficiency of fast matching in a short time. Good uniqueness ensures that the feature description can capture the tiny changes of image structure sensitively. When the organizational structure of image data changes, even the structural changes of local small areas can also reflect the feature description sensitively. The better anti-interference can make the feature descriptors recognize the target category image

data accurately when the image data changes in scale, rotation and noise. The efficiency of fast matching has certain requirements on the complexity of algorithm calculation and time complexity. If the dimension of description operator eigenvector is too high, the complexity of similarity matching between eigenvectors is high. Because STFT features have strong advantages in these three aspects, this paper uses SIFT features to describe the image.

SIFT feature, also known as scale invariant feature transformation, is an algorithm used to detect and describe local features of images. It has rotation invariance and scale invariance by finding extreme points as feature points in scale space. SIFT was proposed by Lowe [35], and is widely used in object recognition, robot map perception and navigation, 3D model building, gesture recognition, image tracking and action comparison. The basic idea of SIFT is to introduce a parameter which is regarded as scale into the image information processing model. Through finding the key points in different scale space, calculating the size, direction and scale information of the key points, using these information to constitute the key points to describe the feature points, to realize edge and corner detection and feature extraction in different resolutions. SIFT feature detection mainly includes the following four basic steps:

- 1) The construction of the scale space of DOG
- 2) Key search and location
- 3) Direction assignment
- 4) Key description. SIFT matching is the process of matching key points (feature points). The process is shown in Figure 3 below:

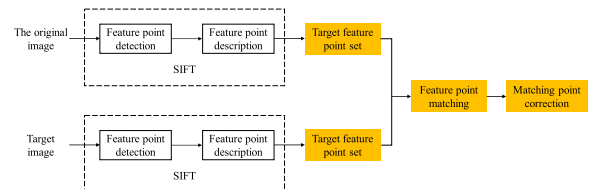


FIGURE 3. SIFT feature point matching process.

Medical image classification is essentially a supervised learning problem in data mining. Firstly, the image data with known category is divided into training set and test set, and the parameters of statistical learning discriminator are adjusted through multiple iterations to achieve good classification effect in the training data set, then the classification accuracy of the discriminator is tested on the test data and test data. Support vector machine (SVM) is a typical statistical classifier, which can not only deal with linearly separable data samples, but also map low-dimensional non separable linear features to high-dimensional features through kernel function, so that they can be separable in higher dimensions. SVM is widely used in various supervised learning scenarios because of its powerful classification ability [36]–[38]. Therefore, SVM is used as the classifier of traditional classification methods.

Support vector machines (SVM) is a kind of two classification model. Its purpose is to find a hyperplane to segment

samples. The principle of segmentation is to maximize the interval, which is finally transformed into a convex quadratic programming problem to solve. In linear separable scene, we assume that the optimal hyperplane is  $\vec{w}\vec{x} + b = 0$ . We use  $|\vec{w}\vec{x} + b|$  to represent the distance from sample point  $x$  to the hyperplane, and compare the symbols of  $\vec{w}\vec{x} + b$  and  $y$  to determine whether the samples are correctly classified. If  $\vec{w}\vec{x} + b = 0$  is consistent with the  $Y$  symbol of label, sample  $x$  is classified correctly, otherwise, the classification is wrong. Therefore, when the samples are classified correctly,  $y(\vec{w}\vec{x} + b = 0) > 0$ , and vice versa,  $y(\vec{w}\vec{x} + b = 0) < 0$ , the function distance can be defined based on this. The function interval is as follows (1).

$$\hat{y} = y(\vec{w}^T \vec{x} + b) = yf(x) \tag{1}$$

$X$  is used to represent the eigenvector of the training sample,  $y$  is used to represent the label of the sample,  $(x_i, y_i)$  is used to represent the eigenvector of the  $i$ th training sample and the label of the  $i$ th sample. Therefore, the function interval is the minimum value of the function interval of the distance classification hyperplane of all training samples. The optimal hyperplane for linear separability is shown in Figure 4.

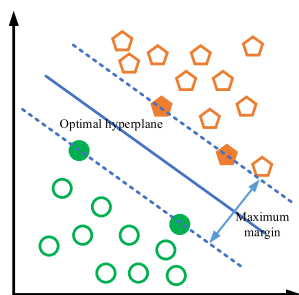


FIGURE 4. Optimal hyperplane for linear separability [37].

When the classification hyperplane is far away from the classification data, the confidence of the obtained classification hyperplane is higher. We should try our best to maximize the distance between the hyperplane and all kinds of data samples. The function interval will increase with the increase of the proportion of  $\vec{w}$  and  $b$ . While the geometric interval eliminates the influence of  $\vec{w}$ , which is only related to the hyperplane itself. Therefore, the function interval should not be used to maximize the interval, but the geometric interval should be used. According to the principle of maximizing the geometric interval, the optimal hyperplane is obtained. The hyperplane represented by the solid line which separates the two kinds of samples is the optimal hyperplane. The distance between the two dashed lines parallel to it is the maximum geometric distance  $\max \hat{y}$ . The classification sample point on the dotted line is the support vector, which satisfies  $y_i(\vec{w}\vec{x}_i + b) = 1$ , and for other correctly classified sample points satisfies  $y_i(\vec{w}\vec{x}_i + b) > 1$ .

For the linear non separable problem, it is necessary to map data to a high-dimensional space so that it can be transformed into a linear separable problem. The mathematical model of

SVM is as follows:

$$\begin{aligned} \min_{w,b} & \frac{1}{2} \|w\|^2 \\ \text{s.t.} & y_i(w^T \Phi(x_i) + b) \geq 1, \quad i = 1, 2, \dots, m \end{aligned}$$

where,  $\Phi(x_i)$  represents the sample feature after  $x_i$  is mapped.

The Lagrange multiplier method can be used to obtain the dual problem of the problem and further solve the mathematical model. The formula is as follows:

$$\begin{aligned} \max_{\alpha} & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \Phi(x_i)^T \Phi(x_j) \\ \text{s.t.} & \sum_{i=1}^m \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad i = 1, 2, \dots, m \end{aligned}$$

where  $\alpha_i$  is the Lagrange multiplier. In the above formula, we need to calculate the inner product  $\Phi(x_i)^T \Phi(x_j)$  after the sample is mapped to a high-dimensional space. Because of its high dimension, we need factor kernel function  $\kappa(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$ . That is, by directly bringing the low-dimensional features into the kernel function, the high-dimensional features are transformed, and then the inner product of the transformed features is obtained. Different kernel functions correspond to different feature mapping methods. Common kernel functions include linear kernel, polynomial kernel, Gaussian kernel, sigmoid kernel and Laplace kernel. Generally, linear kernel is used for text data, and Gauss kernel is used for other cases.

In practical application, there will be some noise points in the data, which can not be used to divide all samples correctly, so the concept of soft interval is introduced. The mathematical model is modified as follows:

$$\begin{aligned} \min_{w,b} & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \\ \text{s.t.} & y_i(w^T \Phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, m \end{aligned}$$

Among them,  $\xi_i$  is the relaxation variable, indicating the degree of incorrect sample classification.

### B. RESULTS AND ANALYSIS

In this paper, neural network and support vector machine are used to classify medical images. In addition, this paper explores the combination of neural network automatic feature extraction and traditional classification methods, which saves the work of artificial feature engineering, such as CNN-SVM algorithm, RBF-SVM algorithm. Then, the image classification effects of GAN, convolutional neural network, RBF-SVM algorithm, PCA-CNN, CNN-SVM algorithm and SVM algorithm are compared. The comparison of different classification methods is shown in Figure 5.

Convolutional neural network (CNN) is one of the most representative neural networks in the field of deep learning technology. One of the advantages of convolutional neural network compared with the traditional image processing algorithm is that it avoids the complex pre-processing

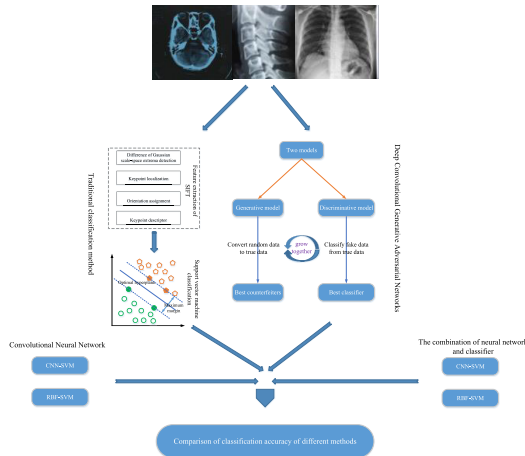


FIGURE 5. Comparison of different classification methods.

process of image. Especially in the process of image pre-processing, convolutional neural network can directly input the original image for a series of work, which has been widely used in various image related applications. The basic idea of RBF neural network is to use RBF as the “base” of the hidden unit to form the hidden layer space, so that the input vector can be directly mapped to the hidden space without weight connection. Principal component analysis (PCA) is one of the most widely used data dimensionality reduction algorithms. The main idea of PCA is to map the n-dimensional feature to the k-dimensional feature, which is a new orthogonal feature, also known as the main component. It is a k-dimensional feature reconstructed on the basis of the original n-dimensional feature. By comparing the classification accuracy of DCGAN, CNN, RBF-SVM, PCA-CNN, CNN-SVM and SVM in different situations, this paper analyzes the advantages and disadvantages of DCGAN compared with other algorithms in different situations.

Generally, we test the generalization performance of machine learning model through experiments. Specifically, the samples are divided into training set and test set. In the training set, the model is trained. In the test set, the discrimination ability of the test model to the new samples is used as the approximation of the model generalization performance. At present, there are three ways to divide training set and test set: hold out, cross validation and bootstrapping.

Set aside method is to directly divide data set into mutually exclusive training set and test set. About 2/3~4/5 of the partition is used for training, and the remaining data is used for testing. Although the proportion of the division is roughly determined, there are still two points to be noted in the division. First, the proportion of different categories in training set and test set should be consistent to ensure the consistency of data distribution in training set and test set. In addition, even if the above two conditions are met, there will still be many ways of division that may lead to the deviation of results. Therefore, in the experiment, it can be divided many times, then trained on the corresponding data, and finally average the results of all test sets as the generalization error estimation of the final draft of the model.

The cross validation method divides the data set into n mutually exclusive subsets, which should also be divided hierarchically to ensure the same proportion of different categories in each subset. Then N-1 subset is used for training each time, and the remaining 1 subset is used as the test. In this way, the subset of test can be changed, so that N training and test division can be carried out, and N models can be trained. The final result of the model is the average of the results on these N test sets. Generally, n is taken as 10, which is called 10 fold cross validation.

The self-service method is to sample the original data again, that is, the samples collected before may still be collected in the subsequent sampling process, but the samples not collected are used as tests. Assuming that the samples put back to M, we can get the data sets of M samples. It is estimated that about 36.8% of the samples have never been sampled.

Taking all aspects into consideration, this paper uses the method of reserve to divide training set and test set. In addition to data set partitioning, the problem of class imbalance is often encountered in machine learning. The category imbalance refers to the situation that the sample quantity gap is too large in different categories. In general, there are less positive samples and more negative samples. In this paper, the contribution of different classes in loss function is modified to deal with unbalanced data. First, we modify the contribution of different categories in the loss function, that is, the contribution value of a positive example error to the loss function is higher than that of a negative example error. If the number of positive cases is M and the number of negative cases is N, then the contribution ratio of positive cases and negative cases to the loss function is N/M: 1.

There are many indicators to represent the accuracy of classification, among which the most commonly used are confusion matrix, overall classification accuracy and Kappa coefficient. The confusion matrix can clearly see the number of correct classification of each figure, as well as the wrong classification and number. However, the confusion matrix can not see the classification accuracy at a glance. Therefore, a variety of classification accuracy indicators are derived from the confusion matrix, among which the overall classification accuracy (OA) and kappa coefficient (Kappa) are most widely used [39], [40].

Overall accuracy (OA) refers to the ratio of the number of correctly classified class pixels to the total number of class pixels. Although OA value can well represent the classification accuracy, it has a great influence on the extremely unbalanced multi class terrain, and can not well represent each class terrain. Kappa coefficient refers to the proportion of error reduction between classification and completely random classification, and its calculation formula is as follows:

$$Kappa = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \times x_{+i})}$$

**TABLE 1. Comparison of classification effect of brain CT image data set.**

Number	Method	DCGAN	CNN	RBF-SVM	PCA-CNN	CNN-SVM	SVM
10000	OA(%)	99.57	98.78	85.76	98.12	88.35	97.53
	Kappa	99.52	98.43	89.74	98.01	86.78	97.47
8000	OA(%)	95.43	95.33	85.96	93.67	88.54	94.50
	Kappa	95.27	95.06	84.37	93.28	87.47	94.06
6000	OA(%)	88.87	78.76	83.97	86.26	85.78	82.02
	Kappa	89.77	82.87	81.65	87.58	84.35	81.07
4000	OA(%)	87.90	82.75	76.75	88.02	79.96	79.77
	Kappa	85.32	80.78	73.28	87.34	77.92	79.56
2000	OA(%)	69.53	58.53	46.65	54.67	53.48	56.21
	Kappa	64.42	57.89	49.46	56.32	52.64	56.79

where,  $N$ ——1 total number of pixels.

- $x_{i+}$ ——Sum of each line of confusion matrix.
- $x_{+i}$ ——Sum of each column of confusion matrix.
- $x_{ii}$ ——confusion matrix diagonal elements.

In this experiment, in order to compare the classification accuracy and results of different methods more truly and comprehensively, the overall classification accuracy and kappa coefficient are used as evaluation indexes to evaluate the classification accuracy of six different feature extraction and classification methods. In order to compare and analyze the influence of different models and methods on the classification results more comprehensively, the proposed method is also combined with convolution neural network (CNN), Gaussian kernel support vector machine (RBF-SVM), convolution neural network (PCA-CNN), convolution neural network and support vector machine (CNN-SVM) after principal component analysis and dimensionality reduction, and the experimental results of SVM algorithm Ratio [41], [42].

The training and testing samples in this experiment are randomly selected from the whole data set. In order to reflect the performance of DCGAN model structure more comprehensively, we have adopted multiple groups of experiments with different training samples to verify. In addition, in order to eliminate the interference of accidental factors and make the test results more stable, we have carried out five experiments to obtain the overall average result Fruit. In the experiment, five different sizes of annotation data sets were defined, and the number of each kind of annotation samples was 2000, 4000, 6000, 8000, 10000 respectively. Table 1, table 2 and table 3 respectively show the average accuracy of different algorithms for classification of brain CT images, chest CT images and cervical CT images in different quantitative labeling situations.

As can be seen from table 1, for brain CT image data set, when 1000 training numbers are selected, the OA of DCGAN is 99.57%, and its classification accuracy is higher than that of CNN, PCA-CNN and SVM. In addition, the proposed method

**TABLE 2. Classification effect comparison of chest CT image data set.**

Number	Method	DCGAN	CNN	RBF-SVM	PCA-CNN	CNN-SVM	SVM
10000	OA(%)	94.06	90.82	83.38	91.65	89.40	92.17
	Kappa	92.43	88.68	81.98	89.47	88.32	90.65
8000	OA(%)	92.65	89.37	81.32	90.45	81.12	92.35
	Kappa	90.78	89.66	81.78	89.23	81.38	90.46
6000	OA(%)	86.84	85.74	76.56	86.34	76.45	87.22
	Kappa	84.34	84.66	71.45	83.67	72.43	85.29
4000	OA(%)	75.43	77.45	68.43	78.86	66.46	79.21
	Kappa	74.38	73.86	66.75	75.43	58.98	75.53
2000	OA(%)	63.43	61.53	54.35	58.89	43.67	59.66
	Kappa	63.32	62.89	54.78	46.78	51.63	47.54

**TABLE 3. Comparison of classification effect of cervical CT image data set.**

Number	Method	DCGAN	CNN	RBF-SVM	PCA-CNN	CNN-SVM	SVM
10000	OA(%)	98.51	97.64	88.37	98.32	97.38	97.35
	Kappa	98.42	97.53	87.21	98.15	96.84	98.32
8000	OA(%)	96.99	95.80	80.19	96.41	90.57	96.31
	Kappa	96.23	94.97	78.32	95.69	89.36	96.12
6000	OA(%)	88.96	85.76	76.54	86.59	76.34	88.78
	Kappa	87.53	84.38	74.38	84.87	74.51	87.32
4000	OA(%)	82.83	77.78	68.46	78.85	68.76	82.45
	Kappa	81.22	68.24	64.33	74.66	64.83	82.74
2000	OA(%)	79.58	65.57	67.68	69.78	62.44	76.59
	Kappa	74.10	68.32	66.59	66.24	61.59	74.33

is superior to the traditional RBF-SVM and CNN-SVM in OA and kappa by 13.81% and 11.22%, respectively. When the number of selected samples is reduced to one-fifth of the original number, for example, 2000 training samples are selected, the convolutional neural network will have a reduction in classification accuracy due to the lack of training samples. In this case, GAN shows better performance. It can be seen from the table that in the case of less samples, the method proposed in this paper has achieved better classification results compared with other classification methods. Therefore, GAN can reduce the dependence of neural network on a large number of training samples to a certain extent. In other words, GAN can effectively reduce the over fitting phenomenon of neural network.

Table 2 describes the classification effect of different classification methods on chest CT images. In the same way as brain CT image processing, 2000, 4000, 6000, 8000 and

10000 samples were randomly selected as the training set of five experiments. It can be seen from the results that when only 2000 training samples are selected, the average classification accuracy of GAN is 63.43%, which is 1.9% higher than that of CNN's OA61.53%, and also higher than other classification methods. When the training samples are 4000, 6000, 8000 and 10000, the classification accuracy of GAN is not very different from other classification methods. On the other hand, this paper expounds that GAN can play a better role in the case of limited training samples.

From table 3, it can be seen that DCGAN can achieve the classification performance with CNN only with less annotation data. For example, GAN can achieve 74.10% with only 2000 samples, while CNN needs 4000 samples to achieve the corresponding classification accuracy. In the case of less labeling, GAN improves the classification performance obviously, but with the increase of the number of labeled samples, the improvement ability of classification performance gradually decreases, but it is better than CNN and other traditional classification methods in general.

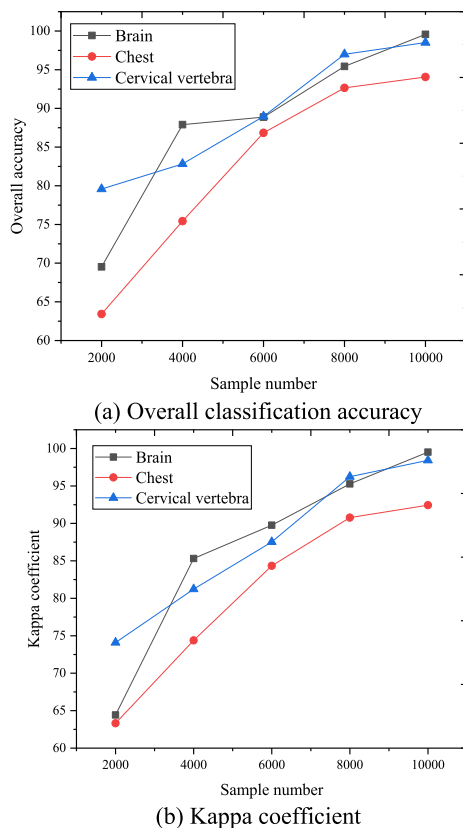


FIGURE 6. Variation of overall classification accuracy and Kappa coefficient with sample number.

Figure 6 shows the trend of the overall classification accuracy and Kappa coefficient with the number of samples. As can be seen from the figure, with the increasing number of samples, the classification accuracy and Kappa coefficient are increasing, and the growth trend of classification accuracy and kappa coefficient is the same. When the number of training samples is increased to 10000, the classification accuracy

is improved by 30%, 35.1% compared with 2000 training samples. In the data set of chest CT images, the classification accuracy was improved by 30.63% and 29.11%, respectively. On CT images of cervical spine, the accuracy of classification increased by 18.93% and 24.32%, respectively. From the above results, we can see that in the deep learning model, the number of labeled training samples has a great influence on the optimization of network parameters. With the increase of the number of labeled samples, the classification performance of the network will be further improved.

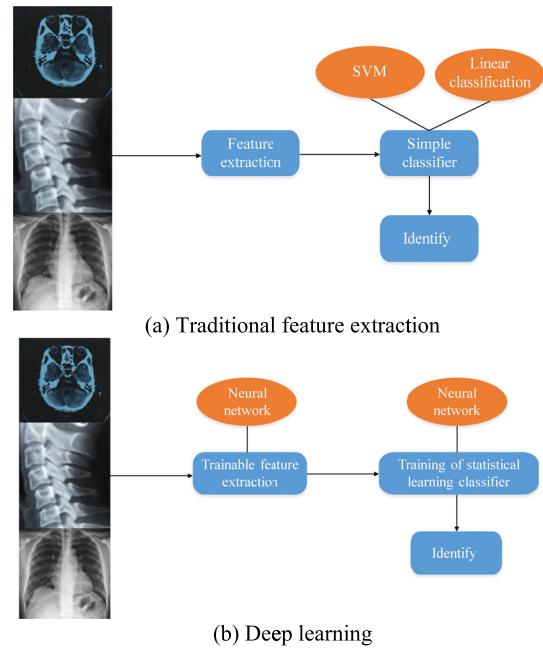


FIGURE 7. Image classification process based on traditional feature extraction and deep learning.

The medical image classification method based on deep learning firstly inputs the image data into the neural network, optimizes the error loss function by using the forward propagation and back propagation error algorithms, and obtains the model with the best classification effect by constantly updating the weights, then applies this model to the medical image recognition analysis, and the specific process is shown in figure 7. The traditional classification method based on statistical machine learning optimizes the feature extraction and classifier separately, while the image classification method based on deep learning can automatically learn the feature of training data, and quickly learn the feature representation of training data according to the training data, which saves the part of feature extraction. Therefore, the strength of deep learning lies in self feature learning, but there are problems that the training samples are not large enough. The dynamic game can automatically generate images against neural network, which just makes up for the shortage of neural network samples leading to the low accuracy of training results.

In order to compare the classification effect of different methods more intuitively, figure 8 further shows the comparison effect of different overall classification accuracy obtained



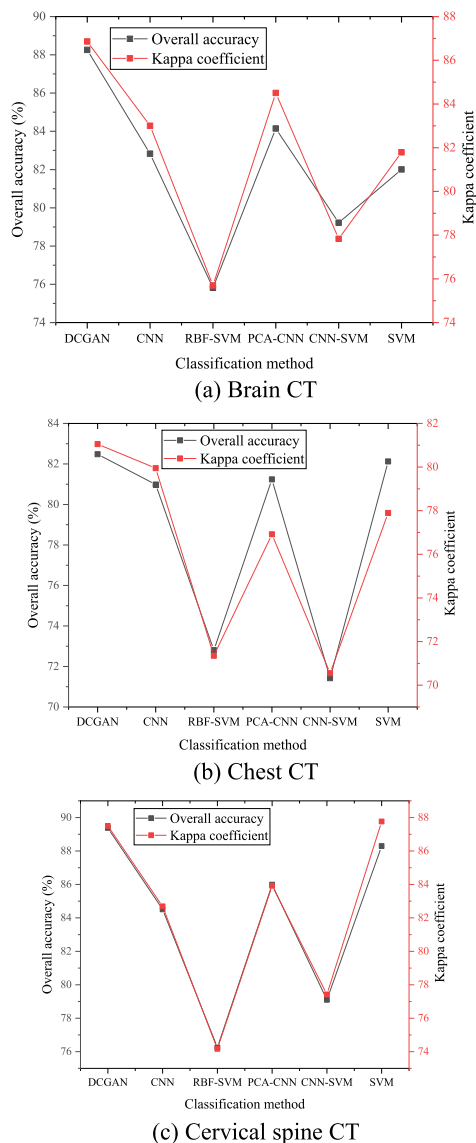


FIGURE 8. Comparison of overall classification accuracy and kappa coefficient of different classification methods.

by selecting different classification methods under the three medical image data sets of brain, chest and cervical spine. The overall classification accuracy and kappa coefficient of different sample numbers are averaged. It can be seen from the figure that the classification effect of deep convolution antagonism neural network is slightly better than that of convolution neural network (CNN), Gaussian kernel support vector machine (RBF-SVM), convolution neural network (PCA-CNN) after principal component analysis and dimensionality reduction, convolution neural network and support vector machine combined method (CNN-SVM), SVM algorithm. The deep convolution antagonism neural network effectively combines the dynamic game idea of antagonism neural network to extract the features of the samples, reduces the dependence of deep learning on the labeled samples, and ensures the robustness of the model in the case of small samples. In the experiment, by learning the distribution of real data, the generator generates an image similar to the

real data. By comparing the real data with the generated image, it is found that the generated sample outline is clear and the structure is reasonable.

V. CONCLUSION

Based on the weak feature extraction ability of traditional image classification methods and the lack of training samples in practical application, this paper proposes a semi supervised image classification method based on the generated antagonism network. In the case of limited labeled samples, this method effectively combines labeled samples and unlabeled sample data, and improves the result of less labeled samples. The decision plane alleviates the problem of the scarcity of labeled samples. The image classification of brain, thorax and cervical spine CT images was carried out by using deep convolution antagonism neural network, and the overall classification accuracy and Kappa coefficient of different classification methods under different sample numbers were analyzed and found. The results show that with the increasing number of training samples, the overall classification accuracy and Kappa coefficient are increasing. Especially in the case of a small number of training samples, compared with other depth neural networks and traditional classification methods, the classification accuracy of counter neural networks is about 10% higher than other neural networks and traditional classification methods, and the advantages are more obvious. The rationality and validity of the classification model are verified by comparative experiments, which provides a new idea for medical image classification.

REFERENCES

- [1] D. Ciresan, A. Giusti, and L. M. Gambardella, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 2843–2851.
- [2] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.
- [3] R. Platania, S. Shams, S. Yang, J. Zhang, K. Lee, and S.-J. Park, "Automated breast cancer diagnosis using deep learning and region of interest detection (BC-DROID)," in *Proc. 8th ACM Int. Conf. Bioinf., Comput. Biology, Health Informat.*, Aug. 2017, pp. 536–543.
- [4] K.-L. Tseng, Y.-L. Lin, W. Hsu, and C.-Y. Huang, "Joint sequence learning and cross-modality convolution for 3D biomedical segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6393–6400.
- [5] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
- [6] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
- [7] L. G. Hafemann, L. S. Oliveira, and P. Cavalin, "Forest species recognition using deep convolutional neural networks," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 1103–1107.
- [8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Dec. 2014.
- [9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, 2015, pp. 234–241.
- [10] R. Almajalid, J. Shan, Y. Du, and M. Zhang, "Development of a deep-learning-based method for breast ultrasound image segmentation," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Orlando, FL, USA, Dec. 2018, pp. 1103–1108.

- [11] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1240–1251, May 2016.
- [12] G. van Tulder and M. de Bruijne, "Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted Boltzmann machines," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1262–1272, May 2016.
- [13] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. C. Mok, L. Shi, and P.-A. Heng, "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1182–1195, May 2016.
- [14] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality iso-intense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, Mar. 2015.
- [15] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, and A. Biller, "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping," *NeuroImage*, vol. 129, pp. 460–469, Apr. 2016.
- [16] G. Wu, M. Kim, Q. Wang, B. C. Munsell, and D. Shen, "Scalable high-performance image registration framework by unsupervised deep feature representations learning," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1505–1516, Jul. 2016.
- [17] M. Simonovsky, B. Gutiérrez-Becker, and D. Mateus, "A deep metric for multimodal registration," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 10–18.
- [18] H.-I. Suk, S.-W. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage*, vol. 101, pp. 569–582, Nov. 2014.
- [19] Y. Anavi, I. Kogan, E. Gelbart, O. Geva, and H. Greenspan, "Visualizing and enhancing a deep learning framework using patients age and gender for chest X-ray image retrieval," *Proc. SPIE Med. Imag. Comput.-Aided Diagnosis*, vol. 9785, Jul. 2016, Art. no. 978510.
- [20] X. Liu, H. R. Tizhoosh, and J. Kofman, "Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 2872–2878.
- [21] H.-C. Shin, K. Roberts, L. Lu, D. Demner-Fushman, J. Yao, and R. M. Summers, "Learning to read chest X-rays: Recurrent neural cascade model for automated image annotation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2497–2506.
- [22] A. Janowczyk, A. Basavanthally, and A. Madabhushi, "Stain normalization using sparse AutoEncoders (StaNoSA): Application to digital pathology," *Comput. Med. Imag. Graph.*, vol. 57, pp. 50–61, Apr. 2017.
- [23] A. Benou, R. Veksler, and A. Friedman, "De-noising of contrast-enhanced MRI sequences by an ensemble of expert deep neural networks," *Deep Learning and Data Labeling for Medical Applications*. Cham, Switzerland: Springer, 2016, doi: 10.1007/978-3-319-46976-8\_11.
- [24] H.-I. Suk, T. Alzheimer's Disease Neuroimaging Initiative, S.-W. Lee, and D. Shen, "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis," *Brain Struct. Function*, vol. 220, no. 2, pp. 841–859, Mar. 2015.
- [25] H. I. Suk and D. Shen, "Deep learning in diagnosis of brain disorders," in *Recent Progress in Brain and Cognitive Engineering*, vol. 5, 2015, pp. 203–213.
- [26] H.-I. Suk, C.-Y. Wee, S.-W. Lee, and D. Shen, "State-space model with deep learning for functional dynamics estimation in resting-state fMRI," *NeuroImage*, vol. 129, pp. 292–307, Apr. 2016.
- [27] R. Achanta, F. Estrada, and P. Wils, "Salient region detection and segmentation," in *Proc. Int. Conf. Comput. Vis. Syst.* Berlin, Germany: Springer, 2008, pp. 66–75.
- [28] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
- [29] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 545–552.
- [30] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.
- [31] S. Reed, Z. Akata, and X. Yan, "Generative adversarial text to image synthesis," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1–10.
- [32] Z. Cao, S. Niu, and J. Zhang, "Occlusion image restoration algorithm based on generating confrontation net-work," *J. Beijing Univ. Posts Telecommun.*, vol. 41, no. 3, pp. 81–86, 2018.
- [33] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
- [34] J. Li, J. Jia, and D. Xu, "Unsupervised representation learning of image-based plant disease with deep convolutional generative adversarial networks," in *Proc. 37th Chin. Control Conf. (CCC)*, Jul. 2018, pp. 9159–9163.
- [35] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [36] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *J. Mach. Learn. Res.*, vol. 2, pp. 45–66, Nov. 2001.
- [37] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2560–2574, Aug. 2007.
- [38] S. Osowski, L. T. Hoai, and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 582–589, Apr. 2004.
- [39] W. Li, M. Jia, J. Wang, J. Lu, J. Deng, J. Tang, and C. Liu, "Association of MMP9-1562C/T and MMP13-77A/G polymorphisms with non-small cell lung cancer in southern Chinese population," *Biomolecules*, vol. 9, no. 3, p. 107, 2019.
- [40] Y. Lou, J. Shi, D. Guo, A. K. Qureshi, and L. Song, "Function of PD-L1 in antitumor immunity of glioma cells," *Saudi J. Biol. Sci.*, vol. 24, no. 4, pp. 803–807, May 2017, doi: 10.1016/j.sjbs.2015.06.025.
- [41] Y. Lou, J. Yang, L. Wang, X. Chen, X. Xin, and Y. Liu, "The clinical efficacy study of treatment to chiari malformation type i with syringomyelia under the minimally invasive surgery of resection of submeningeal cerebellar tonsillar herniation and reconstruction of cisterna magna," *Saudi J. Biol. Sci.*, vol. 26, no. 8, pp. 1927–1931, Dec. 2019, doi: 10.1016/j.sjbs.2019.07.012.
- [42] Z. Xiong, Y. Wu, C. Ye, X. Zhang, and F. Xu, "Color image chaos encryption algorithm combining CRC and nine palace map," *Multimedia Tools Appl.*, vol. 78, no. 22, pp. 31035–31055, Nov. 2019.



**HAI TANG** was born in October 1975. He received the bachelor's degree in engineering from the School of Power and Mechanical Engineering, Wuhan University, in July 1998, and the master's and Ph.D. degrees in engineering from the School of Computer Science, Wuhan University, in 2006 and 2010, respectively. After graduation, he worked as a Computer Teacher with the School of Electrical and Information Engineering, Hubei University of Automotive Technology (HUAT).

He is also a Lecturer and a full-time University Teacher. Since then, he has been actively engaged in scientific research and published more than ten articles as the first author, including six EI retrieval articles (two EI journal articles) and one Chinese core journal (Journal of Central China Normal University Natural Science Edition). He presided over eight scientific research projects. He has twice won the title of Excellent Teacher of HUAT.



**ZHIHUI HU** was born in March 1980. He received the bachelor's degree from the Hubei University of Automotive Technology (HUAT), in 2001, and the master's degree from Wuhan University, in 2007. He is currently a Lecturer with the School of Electrical and Information Engineering, HUAT. Since then, he has been engaged in the teaching and research work of computer science. He has undertaken teaching reform projects of school-level courses such as computer networks and C language programming. In addition, he has compiled and published textbooks such as *University Computer Basics* and *Research on Computer Network Technology*. He has participated in many city-level research projects and published more than ten articles.

...