Neural Computing and Applications (2021) 33:8191–8200
https://doi.org/10.1007/s00521-020-04933-4

S. I : INTELLIGENT COMPUTING METHODOLOGIES IN MACHINE LEARNING FOR IOT APPLICATIONS

CT image classification based on convolutional neural network

Yuezhong Zhang[1] · Shi Wang[2] · Honghua Zhao[3] · Zhenhua Guo[4] · Dianmin Sun[5]

Received: 1 February 2020 / Accepted: 6 April 2020 / Published online: 29 April 2020

Abstract

With the rapid development of the Internet, image information is explosively growing. Traditional image classification methods are difficult to deal with huge image data and cannot meet people's requirements on the accuracy and speed of image classification. In recent years, the convolutional neural network (CNN) has been developing rapidly, and it has performed extremely well. The image classification method based on CNN breaks through the bottleneck of traditional image classification methods and becomes the mainstream image classification algorithm at present. CT image classification algorithm is one of the research hot spots in the field of medical image. The purpose of this paper is to apply convolutional neural network to CT image classification, so as to speed up CT image classification and improve the accuracy of CT image classification and so as to reduce the workload of doctors and improve work efficiency. In this paper, CT images are classified by CDBN model. Vector machine SVM is used as the feature classifier of CDBN model to enhance feature transfer and reuse so as to enrich the features. It also suppresses features that are not very useful for current tasks and improves the performance of the model. Using CDBN to classify CT images, several commonly used gray images are compared. Comparing the results of the ordinary gradient algorithm with Adam algorithm, we can get the CDBN model using Adam optimization algorithm. In CT image classification, both accuracy and speed have a good effect. The experimental results show that the training speed of CDBN model of Adam optimization algorithm in CT image classification is 3% faster than that of general gradient algorithm.

Keywords Convolution neural network · Convolution layer · CT image classification · CDBN model

✉ Dianmin Sun
sundianmin@sdu.edu.cn

Yuezhong Zhang
zhangyuezhong168@163.com

Shi Wang
wangshi@ict.ac.cn

Honghua Zhao
me_zhaohh@ujn.edu.cn

Zhenhua Guo
guozhenhua@inspur.com

[1] Department of Ultrasound, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan 250117, Shandong, China

[2] Key Laboratory of Intelligent Information Processing, Institute of Computer Technology, Chinese Academy of Sciences, Beijing, 100190, China

[3] School of Mechanical Engineering, University of Jinan, Jinan 250022, Shandong, China

[4] State Key Laboratory of High-End and Storage Technology, Inspur Electronic Information Industry Co., Ltd, Jinan 250013, Shandong, China

[5] Department of Thoracic Surgery, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan 250117, Shandong, China

1 Introduction

The most common health check in physical examination is computed tomography (CT), which is widely used in medical diagnostics and is one of the most commonly used medical images [1]. With the development of technology, medical image imaging technology has also been developed and widely used. Medical imaging systems and medical image processing have greatly helped doctors diagnose diseases. Medical imaging devices such as computed tomography, cardiovascular imaging, magnetic resonance imaging, positron emission tomography, and medical ultrasound examination are widely used [2–5]. These medical imaging devices can present various tissue structures of the human body and then combined with medical image processing technology for three-dimensional reconstruction, and the doctor can intuitively observe the pathological changes of the organ tissues in the patient and better assist the doctor's diagnosis [6–8]. Medical imaging diagnosis has become an indispensable reference for clinical medical diagnosis. However, due to the wide application of medical imaging technology, the burden of reading by clinicians has increased, which has affected the correctness of doctors' diagnosis to some extent [9]. In general, 10 mm CT scans usually produce 16–23 pictures. The hospital receives a large number of patients every day, and the amount of data generated is very large. The diagnosis of the disease depends on the collection of medical images and the interpretation of medical images [10]. If the traditional method is used, depending on the image data output from the imaging device to the display screen or film for observation and diagnosis, the doctor will check a large number of images every day. On the one hand, the workload is huge and it is difficult to bear. On the other hand, today, data informatization is also necessary to store large amounts of image data in different categories.

Convolutional neural networks have a large number of research results in the field of medical imaging and are the most widely used models, which have become the mainstream application of medical image analysis. In the field of medical imaging, convolutional neural networks are mainly used for image segmentation and target detection. CT image classification is an important branch of the application field of convolutional neural networks, and especially in the application of digital handwriting, the error rate has almost been reduced to the extreme [11–13]. Due to the complex diversity of CT images, the requirements for classification algorithms are getting higher and higher. Convolutional neural networks have good characteristics such as image rotation invariance, and it is very important to improve their performance in practical applications. Applying the convolutional neural network to the field of CT image classification can relieve the doctor's work pressure and improve the efficiency of diagnosis. On the other hand, using computer to classify medical image pictures is also beneficial to the similar pathological images that appear in the future diagnosis process. Doing scientific classification and collation also has a certain role in clinical research [14–16].

The Moeskops team proposed a method for automatically segmenting MR brain images into many organizational categories using convolutional neural networks. To ensure that the method achieves accurate segmentation detail and spatial consistency, the network uses multiple patch sizes and multiple convolution kernel sizes to obtain multi-scale information about each voxel. The method does not rely on explicit features, but learns to identify information that is important to the classification based on the training data. This method requires only a single anatomical MR image. The segmentation method is applicable to five different datasets: coronal T2-weighted images of preterm infants obtained at 30 weeks post-menstrual (PMA) and 40 weeks of PMA, axial T2-weighted images of preterm infants obtained at PMA 40 weeks, axis A-weighted image of an elderly adult with a T1-average age of 70 years, a T1-weighted image of a young adult with an average age of 23 years. For each dataset, the method obtains the following average Dice coefficients for all segmented organizational categories: 0.87, 0.82, 0.84, 0.86 and 0.91, and accurate segmentation is obtained in all five groups, thus proving its robustness of age and acquisition protocol differences [17]. The Kruthiventi team proposed DeepFix, a fully convolutional neural network that models the bottom-up mechanism of visual attention through significant predictions. Unlike classic works, classic works use a variety of hand-crafted features to depict saliency maps, and their models automatically learn features in a layered manner and predict saliency maps in an end-to-end manner. DeepFix captures semantics of multiple sizes while considering the global context by using a network layer with a large receiver domain. Full convolutional networks are spatially invariant, which prevents them from modeling location-dependent patterns (center bias). Their network solves this problem by incorporating a novel position-biased convolutional layer. They evaluated their models on a number of challenging and significant datasets and showed that the model reached the latest results [18]. The Anthimopoulos team proposed and evaluated a convolutional neural network (CNN) for ILD pattern classification. The proposed network consists of five convolutional layers with 2×2 cores and Leaky ReLU activation, followed by an average pool with a size equal to the size of the final feature map and three dense layers. The last dense layer has seven outputs, which is equivalent to the category considered: health, frosted glass opacity (GGO), micronodules, consolidation, mesh, honeycomb and GGO/mesh combination. To train and evaluate CNN, they used a dataset of 14,696 image patches, which were derived from 120 CT scans from different scanners and hospitals. To the best of their knowledge, this is the first deep CNN designed for a specific problem. Comparative analysis demonstrates the effectiveness of the proposed CNN over previous methods in challenging datasets. Classification performance (85.5%) demonstrates the potential of CNN for the analysis of lung type. Future work includes extending CNN to the 3D data provided by CT volume scans and integrating the proposed method into a CAD system designed to provide differential diagnosis for ILD as an aid to radiologists [19]. Salehi's team proposed a technique based on automated convolutional neural network (CNN), in which the local and global image features are learned through 2D patches of different window sizes. In this architecture, three parallel 2D convolution paths for three different directions (axial, coronal and sagittal) can implicitly learn 3D image information without the need to calculate expensive 3D convolutions. The posterior probability map generated by the network is used iteratively together with the original image block as context information to understand the local shape and connectivity of the brain for extraction from non-brain tissue. The brain extraction results obtained from their algorithm are superior to those obtained recently in the literature on two publicly available benchmark datasets, LPBA40 and OASIS, in which they obtained Dice overlap coefficients of 97.42% and 95.40%. In addition, they evaluated the performance of the algorithm in a challenging problem of extracting an arbitrarily oriented fetal brain from a reconstructed fetal brain magnetic resonance imaging (MRI) dataset. In this application, our algorithm is much better than other methods (scorpion coefficient: 95.98%). In other methods, other methods are not effective due to the irregular orientation and geometry of the fetal brain in MRI. Their CNN-based approach provides accurate, geometrically independent brain extraction in challenging applications [9].

The algorithm designed in this paper has certain pertinence for CT image classification. The vector machine algorithm is used to classify CT images. Support vector machine (SVM) is used as the feature classifier of CDBN model to make features richer, enhance feature transfer and reuse and adaptively learn feature weights by feature recalibration. Improve the network's ability to learn useful features, suppress features that are of little use to the current task and improve the performance of the model and thus feature extraction of CT images. The CT images are classified by the model CDBN, and the parallel implementation of CDBN model training is proposed to realize the parallelization of training and detection. The experimental method in this paper shows that the CDBN model of the Adam optimization algorithm has a 3% higher training speed in CT image classification than the ordinary gradient algorithm.

2 Proposed method

2.1 Convolutional neural network

Convolutional neural networks are a very important model in deep neural networks. Originally designed for use in image recognition classification, it can now be applied to time series signals such as text data. Convolutional neural networks have certain translation invariance and rotational scaling invariance, which determines that it is suitable for image classification processing. Due to its excellent features, it is now used not only for images and video, but also for other computer vision tasks such as natural language processing. The traditional image classification method generally needs to preprocess the image, and the convolutional neural network can directly operate on the original image data without this step. Convolutional neural networks are complex neural network models consisting of multilayer feature extraction and superposition. The convolutional neural network consists of two parts: the feature extraction layer and the full-link layer. The feature extraction layer of the convolutional neural network consists of a convolutional layer, a nonlinear layer, a pooling layer and an overlay layer. Convolutional neural networks, as the name suggests, differ from other network architectures in that they have a convolutional network structure. One of the main structural advantages of convolutional neural networks is the property of weight sharing. This method of weight sharing can greatly reduce the memory usage of the network and greatly reduce the number of parameters in the network model. Over-fitting has been somewhat relieved. The local receptive field, shared weight and downsampling of convolutional neural network design ensure the invariance of scale, translation and distortion in the calculation process.

The name of the convolutional neural network comes from the existence of a convolution operation, which just gives the model the properties of this convolution operation. The convolutional layer, also known as the feature extraction layer, is the core of the convolutional neural network. It is also one of the different CNN elements from traditional neural networks. Most of CNN's calculations are performed at the convolutional level. In the feature extraction process, a fixed convolution kernel is used to convolve the image and obtain a plurality of feature maps, each of which is a composite structure composed of a plurality of neurons. The purpose of the convolution operation is to extract key information from the image. The convolution operation can extract useful feature information from a small block of the input image without affecting the spatial relationship between the pixels. The convolution operation only transmits a small number of pixels of the upper layer network, which is quite different from the characteristics of the fully connected layer. Convolutional patches usually have different sizes and are typically set to (3×3) or (5×5) size. The convolutional layer can obtain deeper abstract feature information contained in a small block of pixels by convolution operations. In general, the depth of the convolutional node matrix becomes deeper and deeper. As shown in Fig. 1, this block is called a convolution kernel or filter. The convolution kernel can convolve the matrix of child nodes on the current neural network and transform it into the next layer of the neural network. In the convolution operation of the convolutional layer, the size (length and width) of the square of the target area of the convolutional kernel convolution is manually set, and the target area depth of the convolution kernel processing is consistent with the current depth of the neural network it processes. The size of the convolution kernel refers to the size (length and width) of the input node matrix, and the depth refers to the depth of the output unit node. When the size of the convolution kernel is not 1×1, the boundary of the current layer matrix can be filled. TensorFlow provides two identical and valid choices, where SAME means that all zeros are added and the convolution kernel has a certain number of dimensions. When the size is even, only one side of the dimension (bottom, right) is filled, and VALID means no padding is added. At the same time, you can adjust the size of the matrix by setting the step size of the convolution kernel move.

In a convolutional neural network, the convolution kernel used by each convolutional layer has the same parameter size, and the convolutional layer is also a very important convolutional neural network. Shared convolution kernels can make full use of valid information between pixels during image processing without being affected by location. Through the sharing of weights, the number of parameters in the neural network training process can be greatly reduced. Due to the existence of weight parameter sharing, the parameters of the convolution kernel are the same in each convolution process. This is also an important feature of convolutional neural networks that distinguish them from other networks. Due to the existence of such weight parameter sharing, it becomes possible to acquire image pixel information without being affected by the position. Sharing the parameters of the filters in each convolutional layer can significantly reduce the parameters of the neural network.

In a convolutional neural network, the pooling layer is also referred to as a downsampling layer. After the convolutional layer, the resulting features are usually higher in dimension. If these high-dimensional features are directly input into the classification, the classification of the device not only makes the network over-fitting, but also has a large amount of calculation and a poor classification effect. Pooling is a statistical analysis of the characteristics of a certain area of an image, and statistical results are used to represent the overall characteristics of the entire area. This statistical analysis operation is a pooling operation. The main purpose of pooling is to achieve high-dimensional to low-dimensional hierarchical transformation of image features through feature mapping. In addition, by introducing a pooling operation, over-fitting can be avoided, and the amount of calculation can be reduced, thereby speeding up the training of the network. The pooling process mainly has the following forms: the largest pooling process of space, the space average pooling process and the spatial addition pooling process. Maximum pooling is one of the commonly used pooling methods in convolutional neural networks. It uses the maximum value in the pool domain as the pooled output value during the pooling process. This has the advantage of reducing the error caused by the mean shift of the convolutional layer parameter error estimate and keeping the texture information of the image as much as possible. A common feature is to scan the largest element 2×2 in the feature image with a 2×2 window of a spatial domain and record it, as shown in Fig. 2. Average pooling is another common pooling method in convolutional neural networks. It differs from the maximum pool direct pool domain maximum as the pool's output value. Average pooling is to find all the values in the pool domain. After summing, using the average as the pool output can reduce the average pool. Due to the limitation of the size of the neighborhood, the

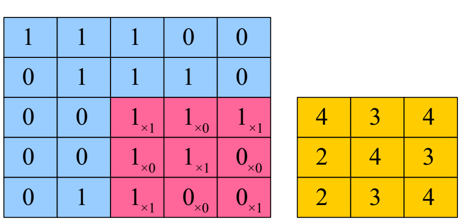

Fig. 1 Image convolution process

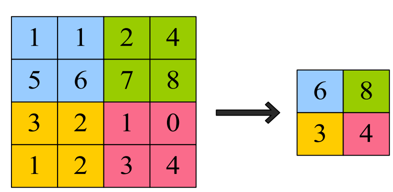

Fig. 2 Maximum pooling process

error caused by the increase in the estimated value is increased, thereby preserving the background information of the image as much as possible. Since the average pool processing principle is similar to the maximum pool processing, usually 2×2 windows are used to scan 4 of the 2×2 windows in the feature map, 4 in each window and averaged, and the average is recorded.

The pooling process can reduce the spatial dimension of the input feature image, which not only reduces the dimension of the original feature space, but also greatly reduces the network space and computational complexity of a large number of parameters, and can prevent overfitting to a certain extent. This makes the output after the maximum pool processing impossible to be greatly affected, even if small changes occur in the image because the combination of the maximum pool or the average pool retains valid information and local weak redundancy. This change is not sensitive, and we can use this feature of the pool to eliminate the influence of location information and effectively detect the target in the image.

In a convolutional neural network, the role of the fully connected layer is to integrate the distributed feature representations learned by the network and map them to the sample tag space. The fully connected layer is characterized in that the neurons in the layer are fully connected to all neurons in the upper layer. All connection operations, numerous parameters, data redundancy, occupy a large amount of computing resources, and model efficiency needs to be improved. In practical applications, the researchers proposed to replace the fully connected layer in CNN with convolution operation or global average pooling method to speed up the model training and avoid the problem of over-fitting of the model due to too many parameters. The classification accuracy of the model has also been improved. After the connection layer is fully connected, there are no other layers or convolutional layers or hidden layers, because after the full connection operation, the dimensional information of the image has been reduced from two dimensions to one dimension, and the convolution operation is performed in the two-dimensional convolution. Therefore, the convolution operation is no longer possible. The main function of the full connection is to map the learned features to the sample mark space, which essentially converts one feature space linearly into another. In convolutional neural networks, full connections often appear in the last layers of the network to measure the weight of the previously designed features.

The output layer is the last layer of the convolutional neural network, shouldering the task of return. Usually the classifier acts as the last layer of the output layer of the network model. The Softmax classifier is a good classification logistic regression model that predicts the likelihood of multiple classifications as long as the label of the sample is unique. If the sample label is not unique, the Softmax regression model cannot be used. Note that the input feature is $x^{(i)}$, the label of each sample is y_i, the training set $S = \{(x^{(1)}, y^{(1)}), \ldots, (x^{(m)}, y^{(m)})\}$, and then, assume that the function and the logistic regression cost function form are as follows:

$$ho(x) = \frac{1}{\sum_{j=1}^{k} \mathrm{e}^{o_j^t x^{(i)}}} \begin{bmatrix} e^{o_1^{\mathrm{T}} x^{(i)}} \\ e^{o_2^{\mathrm{T}} x^{(i)}} \\ e^{o_3^{\mathrm{T}} x^{(i)}} \\ \vdots \\ e^{o_k^{\mathrm{T}} x^{(i)}} \end{bmatrix} \tag{1}$$

where o_1, o_2, o_3 are the learnable parameters of the model, and 1 is the normalized term.

$$J(o) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\left\{y^{(i)} = j\right\} \log \frac{\mathrm{e}^{o_j^T x^{(i)}}}{\sum_{I=1}^{k} \mathrm{e}^{o_I^{\mathrm{T}} x^{(i)}}}\right] \tag{2}$$

The result of this function is 1 only if the value in parentheses in the above formula is true, otherwise 0. The optimal solution for solving the Softmax cost function usually uses an iterative algorithm such as stochastic gradient descent method (SGD), Newton method and quasi-Newton method (L-BFGS). For the Softmax classifier and multiple binary classifiers, when the number of categories is 2, the Softmax regression model degenerates into a logistic regression model. This shows that the Softmax regression model is a general form of the logistic regression model. When the objects to be classified are mutually exclusive, the Softmax regression classifier should be selected. When the classification objects are not mutually exclusive, it is best to set up several independent logistic regression classifiers.

2.2 Image classification

Image classification is an image processing method that distinguishes different kinds of objects according to different characteristics reflected in image information. The classification of image classification is very diverse, and the classification results are also different. According to different image semantics, images can be divided into object classification, scene classification, event classification and emotion classification. The main processes of image classification include image preprocessing, image feature description and extraction and classifier design. The preprocessing includes image filtering (such as median filtering, mean filtering, Gaussian filtering) and dimensional normalization operations, the purpose of which is to facilitate subsequent processing of the target image; the image feature is a description of the prominent features or attributes, each image has its own characteristics, feature extraction, that is, according to the characteristics of the image itself, according to a certain image classification method, the appropriate features are selected and effectively extracted; the classifier is an algorithm for classifying the target image according to the selected features. The role of images in people's lives is increasingly important. People need a reasonable, efficient and fast way to analyze and process large amounts of image data, which can greatly improve the efficiency of people looking for useful images. Life provides great convenience. However, if you use manual methods to classify images, you need a lot of human resources. These workers will spend a lot of time repeating work and the efficiency will be low. In addition, everyone's knowledge and energy is limited. This means that there is an upper limit to the accuracy of the classification results. The focus of image classification is how to extract useful information from various images and classify them by content. A complete and practical image classification system mainly consists of original image information collection and digitization, image preprocessing, feature extraction and classification. In addition to acquisition and digitization, the image classification algorithm is also reflected in other parts. Image classification is an important area of pattern recognition. How to extract effective information is an urgent problem to be solved. Feature extraction, as the core of the image classification algorithm, is directly related to the classification task. Different classification tasks, different feature requirements, different feature extraction methods should be different.

2.3 CT image feature classification

The support vector machine is used as the feature classifier of the CDBN model. Support vector machine is a basic linear two-class model. By solving a convex quadratic optimization problem, the separated hyperplane $w_x + b$ and the decision function which maximize the positive and negative sample intervals are finally obtained as classifiers.

$$f(x) = \mathrm{sign}(w_x + b) \tag{3}$$

SVM combined with nuclear techniques can classify nonlinear data. Choosing the appropriate kernel function for different data has a great influence on the classification effect. The choice of kernel function is usually determined by many training experiments. We use the MATLAB toolbox to assist in training the support vector machine. The training process of the SVM is as follows:

The structured image data file is first input into the CDBN model to obtain $N \times N \times d \times N$, that is, the tag feature map group. Each feature map of size $N \times N \times d$ is then expanded into a row vector as a sample input of the support vector machine, and the support vector machine is trained using the SMO method. The classifier is selected by the cross-validation method: The sample set is randomly divided into ten sub-sample sets before each training, each sub-sample set contains 110 feature data, nine of which are used as training data, and one subset is used as test set. Then, through several trainings and tests, the support vector machine with the smallest classification loss is found. In each training process, the classification effect of the support vector machine is adjusted by adjusting the box constraint parameters of the SVM and the range of the kernel function, and finally, the minimum error classification is adopted.

3 Experiments

3.1 Experimental environment

The environment provided by this experiment is: CPU is Intel i7-8700, 00 GHz; GPU is NVIDIA GTX 1060 with 6G memory, including 16G memory configuration, operating system is Windows 10. Under the operating system, a convolutional neural network framework MatConvNet based on C language and MATLAB is used. Use MATLAB simulation software, the version is MATLAB 2016a.

3.2 Experimental data set

The CT image dataset used in this paper mainly comes from CT images of lung, liver and brain provided by Lung Image Database Consortium, Polytechnique Montréal & CHUM Research Center, Tel Aviv University & Sheba Medical Center, IRCAD, etc. These data come from different medical institutions, and the data acquisition equipment is different. Each patient's image has different layer thicknesses, but they are all between 0.5 and 5 mm, and the resolution is 512×512. Due to personal subjective reasons, each expert's annotation of the same slice is not the same. Therefore, three to four experts are selected as the final reference standard, and then the sample image with poor imaging quality (for example, the image contains serious artifacts) is removed. Finally, 1500 CT scans are selected as the experimental data.The CT image in the dataset is shown in Fig. 3.

3.3 CT image whitening process steps

The specific steps are: normalizing the CT image data and obtaining a dataset $\{x_1, x_2, x_3, \ldots, x_n\}$, containing 1500 CT images. Each image X has a size of 100×100, and each CT image is expanded into a row vector of dimension 10,000. The 1500 image samples are then combined in columns into a sample matrix X of size $1500 \times 10{,}000$.

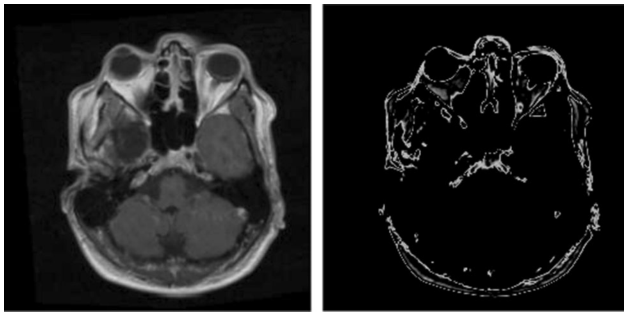

Fig. 3 Example of a CT image in a dataset

Use the formula to find the covariance matrix of the sample matrix $\sum$, and the formula is:

$$\sum = \frac{1}{m} X X^{\mathrm{T}} \tag{4}$$

The SVD decomposition is performed on the covariance matrix to obtain a matrix U composed of eigenvectors. Use the formula to rotate the sample dataset matrix X to the main direction.

$$x_{\mathrm{rotate}} = U^{\mathrm{T}} x = \begin{bmatrix} u1 \\ u2 \\ \cdots \\ un \end{bmatrix} x \tag{5}$$

Normalize the variance of the sample dataset matrix.

$$x_{\mathrm{PCAWrite}} = \begin{pmatrix} \frac{1}{\sqrt{\lambda 1}} \\ \frac{1}{\sqrt{\lambda 2}} \\ \vdots \\ \frac{1}{\sqrt{\lambda n}} \end{pmatrix} x_{\mathrm{rotate}} \tag{6}$$

After the PCA whitening feature ^P is turned over, the whitened image is finally obtained by multiplying the feature vector u. Finally, the ZCA whitened sample matrix is restored to the original image size to obtain the image group of the whitened sample. After completing the ZCA whitening of the image group, in order to facilitate the training and processing of the program, we store the image group and label in a two-dimensional MATLAB cell array structure, the first row of the array stores the image matrix, and the second row stores the image label. The picture sample is a normal picture, and the digital label of the set image is 1, otherwise 0, and the image data are normalized to the [0, 1] interval before each picture is stored in the matrix file. In order to maintain the stability of the parameter update of the training process, we store the preprocessed image groups in a random order into the.mat file and prepare to input the network for training.

4 Discussion

4.1 CDBN analysis of CT classification

The model parameters are first updated using a stochastic gradient descent method. The maximum iteration batch numbers of the first layer, the second layer and the third layer are epoch = 100, epoch = 60 and epoch = 40, respectively. Within each batch, we randomly selected 10% of the CT images of the training set as training samples to train the CDBN model. In order to observe the training situation of the model, the reconstruction error is used to evaluate the model. The reconstruction error refers to the squared difference between the original sample and the original sample after Gibbs sampling by the distribution of the RBM model with the given training data as the standard, as the evaluation index of the model. The reconstruction error can reflect the degree of fitting of the RBM model to the sample data to some extent.

$$\text{Reconstruction error} = \frac{\left\| v_{\mathrm{origin}} - v_{\mathrm{reconstructed}} \right\|^2}{M \times M} \tag{7}$$

The reconstruction error curve of the first layer CDBN and the second layer CDBN of the model after each batch training iteration is completed using the batch gradient descent, as shown in Fig. 4.

In the first 25 batches of training shown in Fig. 5, the CDBN model decreases in the first layer reconstruction error CRBM faster and then gradually decreases. After the 80th batch of training, the model reconstruction error is gradually stable between 0.03 and 0.025. The reconstruction error of the second layer CDBN increased slightly in the previous iterations, and it stabilized between 0.035 and 0.03 after ten iterations. During the experiment, we found that the possible reason for the slight increase in the reconstruction error is that the model is affected by sparse

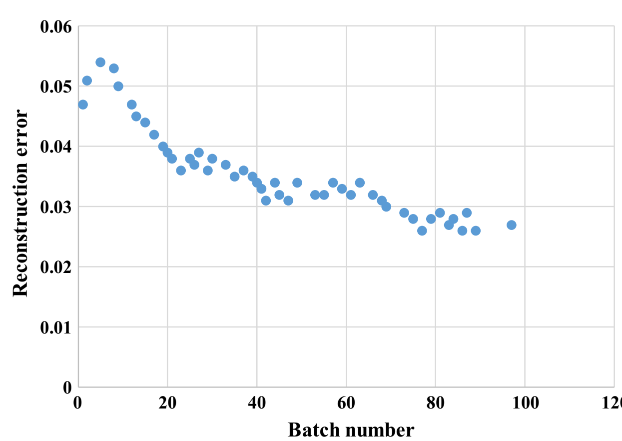

Fig. 4 Variation curve of reconstruction error for each iteration of CDBN layer—first

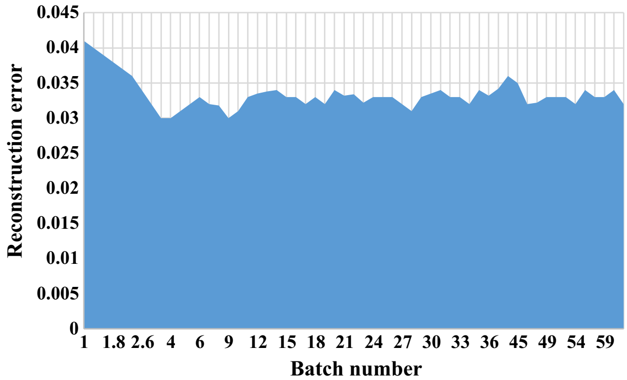

Fig. 5 Variation curve of reconstruction error for each iteration of CDBN layer—second

constraints when updating the model parameters, so that the parameters of the model are not completely updated and moved in the optimal direction.

4.2 CDBN analysis of CT classification accuracy

With the increase in the number of network layers, the CT image feature expression obtained by CDBN has a better classification effect. We can find that the training time of each batch of Adam.CDBN is slightly longer than the CDBN based on impulse. We can find that the reconstruction error of Adam.CDBN can be stabilized in about 15 batches, and the number of iteration batches is greatly reduced compared with the momentum-based model. The CDBN using the Adam algorithm has a much faster convergence rate than the CDBN based on the impulse algorithm. We then calculated the classification accuracy of the Adam.CDBN feature under the RBF.SVM classifier as shown in Table 1 and Fig. 6.

The classification effect is calculated using CDBN features and several other commonly used grayscale image features, including gray-level co-occurrence matrix and Gabor feature. The classification method uses a support vector machine with a radial basis function (RBF). The comparison effect is shown in Table 2 and Fig. 7.

We found that the training time per batch for each layer of Adam.CDBN was slightly longer than the CDBN based

Table 2 Comparison of other feature classification effects

Feture + classification	Positive accuracy (%)	Negative accuracy
Glcm + svm	80	74
Gabor + svm	82	63
1layerCDBN + svm	87	76
2layerCDBN + svm	91	79
3layerCDBN + svm	93	83

Table 1 Feature dimensions and classification accuracy of Adam–CDBN Layers

	First layer	Second layer	Third layer
Feature dimension	51,421	17,880	5175
Positive classification accuracy	86%	92%	93%
Negative classification accuracy	76%	77%	81%

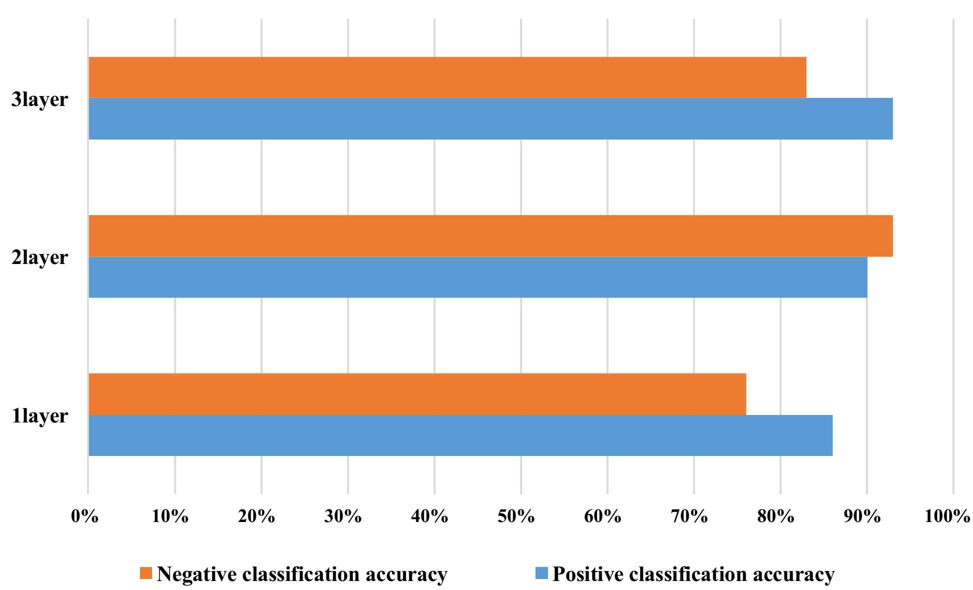

Fig. 6 Feature dimensions and classification accuracy bar charts for Adam–CDBN layers

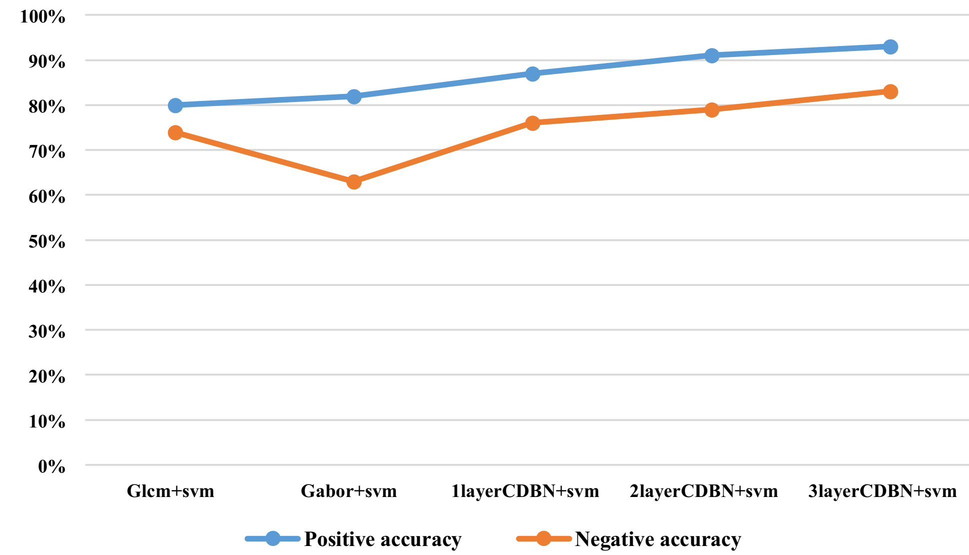

Fig. 7 Comparison of other feature classification effects

on impulse. However, the reconstruction error of Adam.CDBN can be stabilized within about 15 batches. It can be concluded that the CDBN using the Adam algorithm has a much faster convergence rate than the CDBN based on the impulse algorithm. Adam optimizes the CDBN training process, which can effectively improve the training speed while maintaining classification accuracy.

5 Conclusions

Since convolutional neural networks have unique advantages in the field of image classification, images can be directly used as input data. This paper discusses the CT image classification algorithm based on convolutional neural network. By introducing the superiority of convolutional neural networks in image classification processing, several basic image data preprocessing techniques are introduced. By sampling the image data of different medical institutions, the CT images are classified by CDBN, and good results are obtained. Finally, several commonly used grayscale images are compared and compared with the effects of the ordinary gradient algorithm and the Adam algorithm. It can be concluded that the CDBN model using the Adam optimization algorithm has a good effect on the classification accuracy and speed of CT images.

First of all, the basic structure of the neural network is introduced. The special functions of the convolution neural network are the pooling layer and the convolution layer. The advantages of the convolution neural network in the task of image processing, classification and classification are shown. Then, using support vector machine SVM as the feature classifier of CDBN model, the feature extraction of CT images is performed by using support vector machine algorithm to classify CT images.

Secondly, through the analysis of CT classification effect of cdbn and CT classification accuracy of cdbn, it can be concluded that using Adam to optimize the training process of cdbn, the classification accuracy of the obtained features is not much different from the accuracy of CT image classification based on the momentum based cdbn model algorithm. The adaptive moment estimation method can effectively accelerate the training process of the model. However, the use of non-parallel training and detection methods will take a lot of time, so this paper achieves parallelization of CDBN training and detection.

Although the method based on convolutional neural network has achieved good results for some simple image classification tasks, the performance of some complex image classifications needs to be improved. The algorithm designed in this paper shows good characteristics for CT image classification. However, there are also deficiencies. Due to the number of samples, the generalization ability of the model still needs to be verified, and the sparse algorithm also has room for improvement. Therefore, in terms of how to better classify CT images, these aspects require further exploration.

Acknowledgements This study was supported by Grant No. 2018GSF118221 from the Key Research and Development Program of Shandong Province.

Compliance with ethical standards

Conflict of interest There are no potential competing interests in our paper. And all authors have seen the manuscript and approved to submit to your journal. We confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

References

1. Filho PPR, Rebouças EDS, Marinho LB et al (2017) Analysis of human tissue densities: a new approach to extract features from medical images. Pattern Recogn Lett 94:211–218
2. Dandıl E (2018) A computer-aided pipeline for automatic lung cancer classification on computed tomography scans. J Healthc Eng 2018(4):1–12
3. Sethi G, Rawal K, Saini BS (2018) Computer aided diagnosis from CT images using wavelet features with progressive classification. In: International conference on intelligent circuits and systems (ICICS). pp 58–62
4. Filho PPR, Barros ACdS, Ramalho GLB et al (2019) Automated recognition of lung diseases in CT images based on the optimum-path forest classifier. Neural Comput Appl 31:901–914
5. Sebatubun MM, Haryawan C, Windarta B (2017) Classification of ground glass opacity lesion characteristic based on texture feature using lung CT image. J Exp Theor Artif Intell 30(2):1–13
6. Fang Y, Chi Z, Yang W et al (2018) Blind visual quality assessment for image super-resolution by convolutional neural network. Multimed Tools Appl 77(10):1–18
7. Yoon Y, Jeon HG, Yoo D et al (2017) Light-field image super-resolution using convolutional neural network. IEEE Signal Process Lett 24(6):848–852
8. Ševo I, Avramović A (2017) Convolutional neural network based automatic object detection on aerial images. IEEE Geosci Remote Sens Lett 13(5):740–744
9. Salehi SSM, Erdogmus D, Gholipour A (2017) Auto-context convolutional neural network (auto-net) for brain extraction in magnetic resonance imaging. IEEE Trans Med Imaging 36:2319–2330
10. Yin X, Liu X (2017) Multi-task convolutional neural network for pose-invariant face recognition. IEEE Trans Image Process 27:964–975
11. Valverde S, Cabezas M, Roura E et al (2017) Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. Neuroimage 155:159–168
12. Acharya UR, Oh SL, Hagiwara Y et al (2017) A deep convolutional neural network model to classify heartbeats. Comput Biol Med 89:389–396
13. Rocco I, Arandjelovic R, Sivic J (2017) Convolutional neural network architecture for geometric matching. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol 99. p 1
14. Ma X, Dai Z, He Z et al (2017) Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. Sensors 17(4):818
15. Wang Z, Li J, Enoh M (2019) Removing ring artifacts in CBCT images via generative adversarial networks with unidirectional relative total variation loss. Neural Comput Appl 31:5147–5158
16. Chen H, Zhang Y, Kalra MK et al (2017) Low-dose CT with a residual encoder-decoder convolutional neural network. IEEE Trans Med Imaging 36:2524–2535
17. Moeskops P, Viergever MA, Mendrik AM et al (2016) Automatic segmentation of MR brain images with a convolutional neural network. IEEE Trans Med Imaging 35:1252–1261
18. Kruthiventi SSS, Ayush K, Babu RV (2017) DeepFix: a fully convolutional neural network for predicting human eye fixations. IEEE Trans Image Process 26(9):4446–4456
19. Anthimopoulos M, Christodoulidis S, Ebner L et al (2016) Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans Med Imaging 35:1207–1216