Hindawi
Scientific Programming
Volume 2020, Article ID 7607612, 14 pages
https://doi.org/10.1155/2020/7607612

*Research Article*

# Image Classification Algorithm Based on Deep Learning-Kernel Function

**Jun-e Liu**[1] **and Feng-Ping An**[2,3]

[1]*School of Information, Beijing Wuzi University, Beijing 100081, China*
[2]*School of Physics and Electronic Electrical Engineering, Huaiyin Normal of University, Huaian, Jiangsu 223300, China*
[3]*School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China*

Correspondence should be addressed to Feng-Ping An; anfengping@163.com

Received 30 June 2019; Revised 30 November 2019; Accepted 10 December 2019; Published 31 January 2020

Academic Editor: Cristian Mateos

Although the existing traditional image classification methods have been widely applied in practical problems, there are some problems in the application process, such as unsatisfactory effects, low classification accuracy, and weak adaptive ability. This method separates image feature extraction and classification into two steps for classification operation. The deep learning model has a powerful learning ability, which integrates the feature extraction and classification process into a whole to complete the image classification test, which can effectively improve the image classification accuracy. However, this method has the following problems in the application process: first, it is impossible to effectively approximate the complex functions in the deep learning model. Second, the deep learning model comes with a low classifier with low accuracy. So, this paper introduces the idea of sparse representation into the architecture of the deep learning network and comprehensively utilizes the sparse representation of well multidimensional data linear decomposition ability and the deep structural advantages of multilayer nonlinear mapping to complete the complex function approximation in the deep learning model. And a sparse representation classification method based on the optimized kernel function is proposed to replace the classifier in the deep learning model, thereby improving the image classification effect. Therefore, this paper proposes an image classification algorithm based on the stacked sparse coding depth learning model-optimized kernel function nonnegative sparse representation. The experimental results show that the proposed method not only has a higher average accuracy than other mainstream methods but also can be good adapted to various image databases. Compared with other deep learning methods, it can better solve the problems of complex function approximation and poor classifier effect, thus further improving image classification accuracy.

## 1. Introduction

According to the Internet Center (IDC), the total amount of global data will reach 42ZB in 2020. And more than 70% of the information is transmitted by image or video. To extract useful information from these images and video data, computer vision emerged as the times require. At present, computer vision technology has developed rapidly in the field of image classification [1, 2], face recognition [3, 4], object detection [5–7], motion recognition [8, 9], medicine [10, 11], and target tracking [12, 13]. As an important research component of computer vision analysis and machine learning, image classification is an important theoretical basis and technical support to promote the development of artificial intelligence. Image classification began in the late 1950s and has been widely used in various engineering fields, human-car tracking, fingerprints, geology, resources, climate detection, disaster monitoring, medical testing, agricultural automation, communications, military, and other fields [14–19]. A large number of image classification methods have also been proposed in these applications, which are generally divided into the following four categories. (1) Image classification methods based on statistics: it is a method based on the least error, and it is also a popular image statistical model with the Bayesian model [20] and Markov model [21, 22]. (2) Image classification methods based on traditional colors, textures, and local features: the typical feature of local features is scale-invariant feature transform (SIFT). This method was first proposed by David in 1999, and it was perfected in 2005 [23, 24]. SIFT looks for the position, scale, and rotation invariants of extreme points on different spatial scales. It is widely used in object recognition [25], panoramic image stitching [26], and modeling and recognition of 3D scenes and tracking [27]. However, this type of method has problems such as dimensionality disaster and low computational efficiency. (3) Image classification method based on shallow learning: in 1986, Smolensky [28] proposed the Restricted Boltzmann Machine (RBM), which is widely used in feature extraction [29], feature selection [30], and image classification [31]. In 2017, Sankaran et al. [32] proposed a Sparse Restricted Boltzmann Machine (SRBM) method. Its sparse coefficient is determined by the normalized input data mean. It defines a data set whose sparse coefficient exceeds the threshold as a dense data set. It achieves good results on the MNIST data set. However, the characteristics of shallow learning are not satisfactory in some application scenarios. (4) Image classification method based on deep learning: in view of the shortcomings of shallow learning, in 2006, Hinton proposed deep learning technology [33]. For the first time in the journal science, he put forward the concept of deep learning and also unveiled the curtain of feature learning. In view of this, many scholars have introduced it into image classification. Krizhevsky et al. presented the AlexNet model at the 2012 ILSVRC conference, which was optimized over the traditional Convolutional Neural Networks (CNN) [34]. It mainly includes building a deeper model structure, sampling under overlap, ReLU activation function, and adopting the Dropout method. It is applied to image classification, which reduces the image classification Top-5 error rate from 25.8% to 16.4%. Therefore, this method became the champion of image classification in the conference, and it also laid the foundation for deep learning technology in the field of image classification. Since then, in 2014, the Visual Geometry Group of Oxford University proposed the VGG model [35] and achieved the second place in the ILSVRC image classification competition. It reduces the Top-5 error rate for image classification to 7.3%. Its structure is similar to the AlexNet model, but uses more convolutional layers. In 2015, Girshick proposed the Fast Region-based Convolutional Network (Fast R-CNN) [36] for image classification and achieved good results. Compared with the previous work, it uses a number of new ideas to improve training and testing speed, while improving classification accuracy. In 2017, Lee and Kwon proposed a new deep convolutional neural network that is deeper and wider than other existing deep networks for hyperspectral image classification [37]. In 2018, Zhang et al. proposed an image classification method combining a convolutional neural network and a multilayer perceptron of pixels. It consistently outperforms pixel-based MLP, spectral and texture-based MLP, and context-based CNN in terms of classification accuracy. This study provides an idea for effectively solving VFSR image classification [38]. Some scholars have proposed image classification methods based on sparse coding. For example, Zhang et al. [39] embedded label consistency into sparse coding and dictionary learning methods and proposed a classification framework based on sparse coding automatic extraction. Jing et al. [40] applied label consistency to image multilabel annotation tasks to achieve image classification. Zhang et al. [41] proposed a valid implicit label consistency dictionary learning model to classify mechanical faults. However, this type of method still cannot perform adaptive classification based on information features.

Although the deep learning theory has achieved good application results in image classification, it has problems such as excessive gradient propagation path and over-fitting. In view of this, this paper introduces the idea of sparse representation into the architecture of the deep learning network and comprehensively utilizes the sparse representation of good multidimensional data linear decomposition ability and the deep structural advantages of multilayer nonlinear mapping. It will complete the approximation of complex functions and build a deep learning model with adaptive approximation capabilities. It solves the problem of function approximation in the deep learning model. At the same time, a sparse representation classification method using the optimized kernel function is proposed to replace the classifier in the deep learning model. It will improve the image classification effect. So, this paper proposes an image classification algorithm based on the stacked sparse coding depth learning model-optimized kernel function nonnegative sparse representation. The novelty of this paper is to construct a deep learning model with adaptive approximation ability. At the same time, this paper proposes a new sparse representation classification method for optimizing kernel functions to replace the classifier in the deep learning model.

Section 2 of this paper will mainly explain the deep learning model based on stack sparse coding proposed in this paper. Section 3 systematically describes the classifier design method proposed in this paper to optimize the nonnegative sparse representation of kernel functions. Section 4 constructs the basic steps of the image classification algorithm based on the stacked sparse coding depth learning model-optimized kernel function nonnegative sparse representation. Section 5 analyzes the image classification algorithm proposed in this paper and compares it with the mainstream image classification algorithm. Finally, the full text is summarized and discussed.

## 2. Deep Learning Model Based on Stacked Sparse Coding

*2.1. Stacked Sparse Autoencoder.* The Automatic Encoder Deep Learning Network (AEDLN) is composed of multiple automatic encoders. If multiple sparse autoencoders form a deep network, it is called a deep network model based on Sparse Stack Autoencoder (SSAE).

The sparse autoencoder [42, 43] adds a sparse constraint to the autoencoder, which is typically a sigmoid function. During learning, if a neuron is activated, the output value is approximately 1. If the output is approximately zero, then the neuron is suppressed. The network structure of the automatic encoder is shown in Figure 1. The basic principle of forming a sparse autoencoder after the automatic encoder is added to the sparse constraint as follows.

It is assumed that the training sample set of the image classification is $\{x^{(1)}, x^{(2)}, \ldots, x^{(\mathrm{m})}\}$, and $x^{(\mathrm{m})}$ is the image to be trained. Training is performed using a convolutional neural network algorithm with the output target $y^{(i)}$ set to the input value, $y^{(i)} = x^{(i)}$.

In Figure 1, the autoencoder network uses a three-layer network structure: input layer $L_1$, hidden layer $L_2$, and output layer $L_3$. Its training goal is to make the output signal $\widehat{x}$ approximate the input signal $x$, that is, the error value between the output signal and the input signal is the smallest.

The number of hidden layer nodes in the self-encoder is less than the number of input nodes. If the number of hidden nodes is more than the number of input nodes, it can also be automatically coded. At this point, it only needs to add sparse constraints to the hidden layer nodes. In general, high-dimensional and sparse signal expression is considered to be an effective expression, and in the algorithm, it is generally not specified which nodes in the hidden layer expression are suppressed, that is, artificially specified sparsity, and the suppression node is the sigmoid unit output is 0. Specifying $\rho$ sparsity parameter in the algorithm represents the average activation value of the hidden neurons, i.e., averaging over the training set. In node $j$ in the activated layer $l$, its automatic encoding can be expressed as $a_j^{(l)}$:

$$a_j^{(l)} = f\left(\sum_{i}^{s_{l-1}} W_{ji}^{(t-1)} * \alpha_i^{(t-1)} + b^{(l-1)}\right), \tag{1}$$

where $f(x)$ is the sigmoid function, the number of nodes in the $L$th layer can be expressed as $s_l$ the weight of the $i, j$th unit can be expressed as $W_{ji}$, and the offset of the $L$th layer can be expressed as $b^{(l)}$. Therefore, $a_j^{(2)}(x)$ can be used to represent the activation value of the input vector $x$ for the first hidden layer unit $j$, then the average activation value of $j$ is

$$\widehat{\rho}_j = \frac{1}{m}\sum_{i=1}^{m}\left[\alpha_j^{(2)}(x, y)\right]. \tag{2}$$

The above formula indicates that for each input sample, $j$ will output an activation value. Therefore, the activation values of all the samples corresponding to the node $j$ are averaged, and then the constraints are

$$\widehat{\rho}_j = \rho, \tag{3}$$

where $\rho$ is the sparse parameter of the hidden layer unit. To achieve the goal of constraining each neuron, usually $\rho$ is a value close to 0, such as $\rho = 0.05$, i.e., only 5% chance is activated. The goal of e-learning is to make $\widehat{\rho}$ as close as possible to $\rho$. That is to say, to obtain a sparse network structure, the activation values of the hidden layer unit nodes must be mostly close to zero. In order to achieve the purpose of sparseness, when optimizing the objective function, those $\widehat{\rho}$ which deviate greatly from the sparse parameter $\rho$ are punished. This paper chooses to use KL scatter (Kullback Leibler, KL) as the penalty constraint:

$$\gamma\left(\rho_j\right) = \sum_{j=1}^{s_2}\rho\log\left(\frac{\rho}{\widehat{\rho}_j}\right) + (1-\rho)\log\frac{1-\rho}{1-\widehat{\rho}_j}, \tag{4}$$

where $s_2$ is the number of hidden layer neurons in the sparse autoencoder network, such as the method using KL divergence constraint, then formula (4) can also be expressed as follows:

$$\gamma\left(\rho_j\right) = \sum_{j=1}^{s_2}\mathrm{KL}\left(\rho \,\|\, \widehat{\rho}_j\right). \tag{5}$$

Among them,

$$\mathrm{KL}\left(\rho \,\|\, \widehat{\rho}_j\right) = \rho\log\left(\frac{\rho}{\widehat{\rho}_j}\right) + (1-\rho)\log\frac{1-\rho}{1-\widehat{\rho}_j}. \tag{6}$$

When $\widehat{\rho}_j = \rho$, $\mathrm{KL}(\rho \,\|\, \widehat{\rho}_j) = 0$, if the value of $\widehat{\rho}_j$ differs greatly from the value of $\rho$, then the $\mathrm{KL}(\rho \,\|\, \widehat{\rho}_j)$ term will also become larger. The overall cost function can be expressed as follows:

$$H_{\text{sparse}}(W, b) = H(W, b) + \beta\sum_{j=1}^{s_2}\mathrm{KL}\left(\rho \,\|\, \widehat{\rho}_j\right). \tag{7}$$

Among them, the coefficient $\beta$ is a sparse penalty term, the value of $\widehat{\rho}_j$ related to $W$, $b$, and $H(W, b)$ is a loss function, which can be expressed as follows:

$$\begin{aligned} H(W, b) &= \left[\frac{1}{m}\sum_{i=1}^{m} H\left(W, b; x^{(i)}\right)\right] + 0.5\lambda\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{i=1}^{s_{l+1}}\left(W_{ji}^{(l)}\right)^2 \\ &= \left[\frac{1}{m}\sum_{i=1}^{m}\frac{1}{2}\left\|h_{W,b}\left(x^{(i)}\right)\right\|^2\right] + 0.5\lambda\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{i=1}^{s_{l+1}}\left(W_{ji}^{(l)}\right)^2. \end{aligned} \tag{8}$$

The abovementioned formula gives the overall cost function, and the residual or loss of each hidden layer node is the most critical to construct a deep learning model based on stacked sparse coding. To this end, the residuals of the hidden layer are described in detail below, and the corresponding relationship is given. The residual for layer $l$ node $i$ is defined as $\delta^{(l)}$. It is used to measure the effect of the node on the total residual of the output. The sparse penalty item only needs the first layer parameter to participate in the calculation, and the residual of the second hidden layer can be expressed as follows:

$$\delta^{(2)} = \left(\sum_{j=1}^{s_2} W_{ji}^{(2)}\delta^{(3)}\right) f'\left(z_i^{(2)}\right). \tag{9}$$

After adding a sparse constraint, it can be transformed into

$$\delta^{(2)} = \left(\sum_{j=1}^{s_2} W_{ji}^{(2)}\delta^{(3)}\right) + \beta\left(-\frac{\rho}{\widehat{\rho}_j} + \frac{1-\rho}{1-\widehat{\rho}_j}\right) f'\left(z_i^{(2)}\right), \tag{10}$$

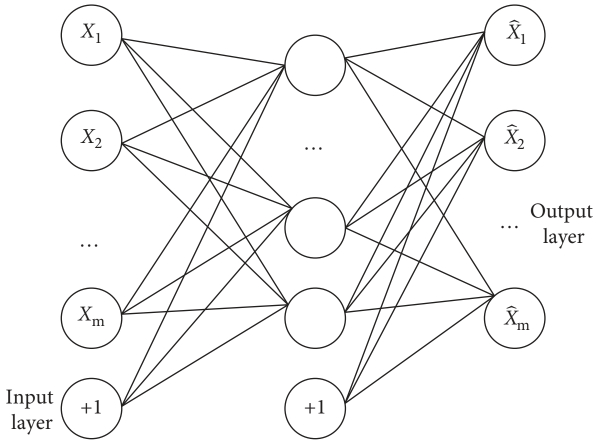

Figure 1: Basic structure of a sparse autoencoder.

where $z_j^{(l)}$ is the input of the activation amount of the $L$th node $j$, $f'(z_j^{(l)}) = \alpha_j^{(l)}$.

*2.2. Stack Sparse Autoencoder Model and Training Ideas.* In general, the dimensionality of the image signal after deep learning analysis increases sharply and many parameters need to be optimized in deep learning. Therefore, sparse constraints need to be added in the process of deep learning. It can effectively control and reduce the computational complexity of the image signal to be classified for deep learning. It can efficiently learn more meaningful expressions.

The stack sparse autoencoder is a constraint that adds sparse penalty terms to the cost function of AE. Therefore, it can automatically adjust the number of hidden layer nodes according to the dimension of the data during the training process. It avoids the disadvantages of hidden layer nodes relying on experience. Based on the study of the deep learning model, combined with the practical problems of image classification, this paper, sparse autoencoders are stacked and a deep learning model based on Sparse Stack Autoencoder (SSAE) is proposed. In the process of deep learning, the more layers of sparse self-encoding and the feature expressions obtained through network learning are more in line with the characteristics of data structures, and it can also obtain more abstract features of data expression.

*2.2.1. SSAE Model Structure.* The SSAE is implemented by the superposition of multiple sparse autoencoders, and the SSAE is the same as the deep learning model. It is also a generation model. The SSAE is characterized by layer-by-layer training sparse autoencoder based on the input data and finally completes the training of the entire network. Then, by comparing the difference between the input value and the output value, the validity of the SSAE feature learning is analyzed. The basic structure of SSAE is as shown in Figure 2. The SSAEs are stacked by an M-layer sparse autoencoder, where each adjacent two layers form a sparse autoencoder. During the training process, the output reconstruction signal of each layer is used to compare with the input signal to minimize the error.

*2.2.2. SSAE Model Training Ideas.* The SSAE depth model directly models the hidden layer response of the network by adding sparse constraints to the deep network. Since each hidden layer unit is sparsely constrained in the sparse autoencoder. Therefore, it can get a hidden layer sparse response, and its training objective function is

$$H_{\min}(W,c,b) = -\sum_{l=1}^{m} \log \sum_{h} \rho(x(l),h(l)) + \lambda \sum_{j=1}^{K} \left| \rho - \frac{1}{m} \sum_{j}^{K} E\left[ h_j(l) \mid x(l) \right] \right|^2 . \tag{11}$$

In the formula, the response value of the hidden layer is between $[0,1]$. $\rho \in (0,1)$ represents the response expectation of the hidden layer unit. The smaller the value of $\rho$, the more sparse the response of its network structure hidden layer unit. $\sum_h \rho(x(t),h(t))$ represents the probability of occurrence of the $l$th sample $x(l)$. $E[h(l) \mid x(l)]$ represents the expected value of the $j$th hidden layer unit response. $h(l)$ represents the response of the hidden layer. $m$ represents the number of training samples.

SSAE training is based on layer-by-layer training from the ground up. The idea of SSAE training is to train one layer in the network each time, that is, to train a network with only one hidden layer. In training, the first SAE is trained first, and the goal of training is to minimize the error between the input signal and the signal reconstructed after sparse decomposition. Then, the output value of the M-1 hidden layer training of the SAE is used as the input value of the Mth hidden layer. Repeat in this way until all SAE training is completed. SSAE itself does not have the function of classification, but it only has the function of feature extraction. Therefore, if you want to achieve data classification, you must also add a classifier to the last layer of the network. The classifier for optimizing the nonnegative sparse representation of the kernel function proposed in this paper is added here.

In this paper, the output of the last layer of SAE is used as the input of the classifier proposed in this paper, which keeps the parameters of the layers that have been trained unchanged. The weights obtained by each layer individually training are used as the weight initialization values of the entire deep network. Then, fine tune the network parameters. The basic flow chart of the constructed SSAE model is shown in Figure 3.

*2.3. Advantages of SSAE Deep Learning Model in Image Classification.* Compared with the deep belief network model, the SSAE model is simpler and easier to implement. SSAE's model generalization ability and classification accuracy are better than other models. This is due to the inclusion of sparse representations in the basic network model that makes up the SSAE. Sparse autoencoders are often used to learn the effective sparse coding of original images, that is, to acquire the main features in the image data. The SSAE model is an unsupervised learning model that can extract high autocorrelation features in image data during training,

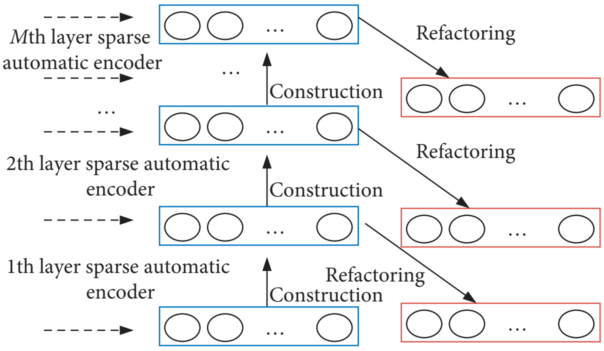
Mth layer sparse automatic encoder
Refactoring
Construction
...
2th layer sparse automatic encoder
Refactoring
Construction
1th layer sparse automatic encoder
Refactoring
Construction

Figure 2: Basic schematic diagram of the stacked sparse autoencoder.

and it can also alleviate the optimization difficulties of convolutional networks. Since the learning data sample of the SSAE model is not only the input data, but also used as the target comparison image of the output image, the SSAE weight parameter is adjusted by comparing the input and output, and finally the training of the entire network is completed.

The SSAE depth model is widely used for feature learning and data dimension reduction. Due to the constraints of sparse conditions in the model, the model has achieved good results in large-scale unlabeled training. Moreover, the weight of its training is more in line with the characteristics of the data itself than the traditional random initialization method, and the training speed is faster than the traditional method.

The image classification algorithm studied in this paper involves a large number of complex images. These large numbers of complex images require a lot of data training to dig into the deep essential image feature information. Since the calculation of processing large amounts of data is inevitably at the expense of a large amount of computation, selecting the SSAE depth model can effectively solve this problem. The SSAE deep learning network is composed of sparse autoencoders. In the process of training object images, the most sparse features of image information are extracted. It can reduce dimension information. Then, through the deep learning method, the intrinsic characteristics of the data are learned layer by layer, and the efficiency of the algorithm is improved. Applying SSAE to image classification has the following advantages:

(1) The essence of deep learning is the transformation of data representation and the dimensionality reduction of data. In DNN, the choice of the number of hidden layer nodes has not been well solved. However, the sparse characteristics of image data are considered in SSAE. It is calculated by sparse representation to obtain the eigendimension of high-dimensional image information. The sparsity constraint provides the basis for the design of hidden layer nodes. In summary, the structure of the deep network is designed by sparse constrained optimization.

(2) Because deep learning uses automatic learning to obtain the feature information of the object measured by the image, but as the amount of calculated data increases, the required training accuracy is higher, and then its training speed will be slower. Therefore, adding the sparse constraint idea to deep learning is an effective measure to improve the training speed. The SSAE model proposed in this paper is a new network model architecture under the deep learning framework.

Therefore, the SSAE-based deep learning model is suitable for image classification problems. The model can effectively extract the sparse explanatory factor of high-dimensional image information, which can better preserve the feature information of the original image. It can reduce the size of the image signal with large structure and complex structure and then layer the feature extraction. The features thus extracted can express signals more comprehensively and accurately. It is an excellent choice for solving complex image feature analysis.

## 3. Classifier Design for Optimizing Nonnegative Sparse Representation of Kernel Functions

*3.1. Basic Principle of Nonnegative Sparse Coding.* Assuming that images are a matrix of $w \times h$, the autoencoder will map each image into a column vector $v \in R^{\mathrm{d}}$, $d = w \times h$, then $n$ training images form a dictionary matrix, that is, $D = [v_1, v_2, \ldots, v_n] \in R^{d \times n}$. Let $D_1 \in R^{d \times k}$ denote the target dictionary and $D_2 \in R^{d \times (n \times k)}$ denote the background dictionary, then $D = [D_1, D_2]$.

Under the sparse representation framework, the pure target column vector $y \in R^d$ can be obtained by a linear combination of the atom $v$ in the dictionary and the sparse coefficient vector $C$. The details are as follows:

$$y = \sum_{i=1:\,n} v_i \cdot c_i. \tag{12}$$

Among them, the sparse coefficient $C = [0, \ldots, 0, c_1^t, 0, \ldots, 0] \in R^n$. In the ideal case, only one coefficient in the coefficient vector is not 0. In the real world, because of the noise signal pollution in the target column vector, the target column vector is difficult to recover perfectly. So, add a slack variable to formula (12):

$$y = \sum_{i=1:\,n} v_i \cdot c_i + r, \tag{13}$$

where $y$ is the actual column vector and $r \in R^d$ is the reconstructed residual. In formula (13), $v_i$ and $y$ are known, and it is necessary to find the coefficient vector corresponding to the test image in the dictionary. To this end, it must combine nonnegative matrix decomposition and then propose nonnegative sparse coding. Therefore, its objective function becomes the following:

$$\min_{C} H(C) = \left\| y - \sum_{i=1:\,n} v_i \cdot c_i \right\|_2^2 + \lambda \|C\|_1, \tag{14}$$

where $\lambda$ is a compromise weight. When $\lambda$ increases, the sparsity of the coefficient increases. $\|C\|_1 = \sum_{i=1:\,n} \|c_i\|$, $c_i \geq 0$, $v_{i,j} \geq 0$.

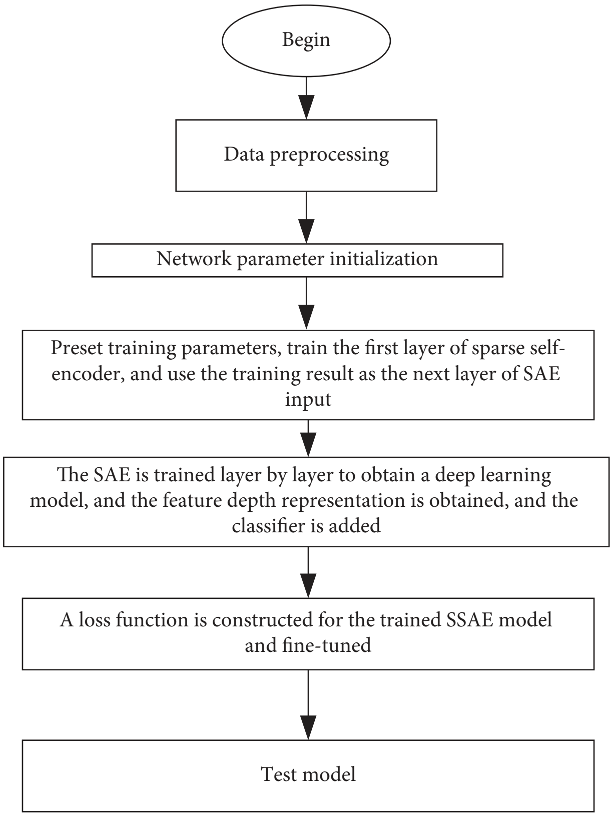
Begin
Data preprocessing
Network parameter initialization
Preset training parameters, train the first layer of sparse self-encoder, and use the training result as the next layer of SAE input
The SAE is trained layer by layer to obtain a deep learning model, and the feature depth representation is obtained, and the classifier is added
A loss function is constructed for the trained SSAE model and fine-tuned
Test model

Figure 3: Basic flow chart of SSAE model training.

*3.2. The Basic Principle of Classification of Nonnegative Sparse Representation of Kernel Function.* The premise that the nonnegative sparse classification achieves a higher classification correct rate is that the column vectors of $D = [v_1, v_2, \ldots, v_n]$ are not correlated. But in some visual tasks, sometimes there are more similar features between different classes in the dictionary. If the two types of problems are considered, the correlation of the column vectors of $D_1$ and $D_2$ is high, and the nonzero solutions of the convex optimization may be concentrated on the wrong category. A kernel function is a dimensional transformation function that projects a feature vector from a low-dimensional space into a high-dimensional space. It can increase the geometric distance between categories, making the linear indivisible into linear separable. This method has many successful applications in classic classifiers such as Support Vector Machine. Inspired by [44], the kernel function technique can also be applied to the sparse representation problem, reducing the classification difficulty and reducing the reconstruction error. Its basic idea is as follows.

Let function $\phi$ project the feature from dimensional space $d$ to dimensional space $h$: $R^d \rightarrow R^h$, $(d < h)$. The class to be classified is projected as $y \rightarrow \phi(y)$, and the dictionary is projected as $D \rightarrow D = (\phi(v_1), \phi(v_2), \ldots, \phi(v_n))$. Let $K(x, x) = \phi(x)^T \phi(x) = 1$. Then, the kernel function is sparse to indicate that the objective equation is

$$\begin{aligned} \min_{C} \quad & H(C) = \left\| \sum_{i=1:\, n} \phi(v_i) \cdot c_i - \phi(y) \right\|_2^2 + \lambda \|C\|_1 \\ \text{s.t.} \quad & c_i \geq 0, v_{i,j} \geq 0. \end{aligned} \tag{15}$$

It can be known that the convergence rate of the random coordinate descent method (RCD) is faster than the classical coordinate descent method (CDM) and the feature mark search FSS method. However, because the RCD method searches for the optimal solution in the entire real space, its solution may be negative. It does not conform to the nonnegative constraint $c_i \geq 0$ in equation (15). Therefore, this paper proposes a kernel nonnegative Random Coordinate Descent (KNNRCD) method to solve formula (15). The condition for solving nonnegative coefficients using KNNRCD is that the gradient of the objective function $R(C)$ conforms to the Coordinate-wise Lipschitz Continuity, that is,

$$\left| \nabla_c H\left(C + \mathrm{he}_j\right) - \nabla_c H(C) \right| \leq L_j |h|. \tag{16}$$

When $c_i \neq 0$, the partial derivative of $J(C)$ can be obtained:

$$\begin{aligned} k\left(v_j, v_i\right) &= \varphi\left(v_j\right)^T \cdot \varphi(v_i), \\ k\left(v_j, y\right) &= \varphi\left(v_j\right)^T \cdot \varphi(y), \\ \nabla_c H(C) &= \sum_{i=1:n} c_i \cdot K\left(v_j, v_i\right) - K\left(v_j, y\right) + \lambda. \end{aligned} \tag{17}$$

Calculated by the above mentioned formula,

$$L_j = \max\left( \frac{\partial\left(\nabla_c J(C)\right)}{\partial c} \right) = k\left(v_j, v_i\right), \tag{18}$$

where $k\,(v_j, v_i) = 1$. It can be seen that the gradient $\nabla_c H(C)$ of the objective function is divisible and its first derivative is bounded. So, the gradient of the objective function $H(C)$ is consistent with Lipschitz's continuum. According to [44], the update method of RCD is

$$c_i^{k+1} = c_i^k - L_i^{-1} \cdot \nabla H(C) \cdot e_i, \tag{19}$$

where $i$ is a random integer between $[0, n]$. But the calculated coefficient result may be $c_i^{k+1} < 0$. So, it needs to improve it to

$$c_i^{k+1} = \max\left(0, c_i^k - L_i^{-1} \cdot \nabla H(C) \cdot e_i\right). \tag{20}$$

For the coefficient selection problem, the probability that all coefficients in the RCD are selected is equal. This strategy leads to repeated optimization of the zero coefficients. In order to improve the efficiency of the algorithm, KNNRCD's strategy is to optimize only the coefficient $c_i$ greater than zero. Specifically, the computational complexity of the method is $O((n/\varepsilon)\log(1/\rho))$, where $\varepsilon$ is the convergence precision and $\rho$ is the probability. Therefore, for any kernel function $K(\cdot,\cdot)$, the KNNRCD algorithm can iteratively optimize the sparse coefficient $C$ by the abovementioned formula.

This paper proposes the Kernel Nonnegative Sparse Representation Classification (KNNSRC) method for classifying and calculating the loss value of particles. If $r_s$ is the residual corresponding to class $s$, then

$$\begin{aligned} r_s &= \left\| \sum_{i=1}^{n} D \cdot C_s - \varphi(y) \right\|_2^2 \\ &= \left( C_S^T K\left(v_i, v_j\right)_{n\times n} C_s - 2K\left(v_i, y\right)_{n\times 1}^T C_s \right), \end{aligned} \tag{21}$$

where $C_s$ is the corresponding coefficient of the S-class. For the two classification problem available,

$$\begin{cases} l_y = 1, & \text{if } r_1 < r_2, \\ l_y = 0, & \text{if } r_1 > r_2, \end{cases} \tag{22}$$

where $l_y$ is the category corresponding to the image $y$. For a multiclass classification problem, the classification result is the category corresponding to the minimum residual $r_s$. The particle loss value required by the NH algorithm is $l_{i,t} = r_1$.

The KNNRCD method can combine multiple forms of kernel functions such as Gaussian Kernel and Laplace Kernel. The final classification accuracy corresponding to different kinds of kernel functions is different.

## 4. Image Classification Algorithm Based on Stacked Sparse Coding Deep Learning Model-Optimized Kernel Function Nonnegative Sparse Representation

Firstly, the sparse representation of good multidimensional data linear decomposition ability and the deep structural advantages of multilayer nonlinear mapping are used to complete the approximation of the complex function of the deep learning model training process. Then, a deep learning model based on stacked sparse coding with adaptive approximation ability is constructed. The classifier of the nonnegative sparse representation of the optimized kernel function is added to the deep learning model. Finally, an image classification algorithm based on stacked sparse coding depth learning model-optimized kernel function nonnegative sparse representation is established. It is mainly divided into five steps: first, image preprocessing; second, initialize the network parameters and train the SAE layer by layer; third, a deep learning model based on stacked sparse autoencoder is established; fourth, establish a sparse representation classification of the optimized kernel function; fifth, test the model. The basic flow chart of the proposed image classification algorithm is shown in Figure 4. Its basic steps are as follows:

(1) First preprocess the image data.

(2) Initialize the network parameters and give the number of network layers, the number of neural units in each layer, the weight of sparse penalty items, and so on.

(3) The approximation of complex functions is accomplished by the sparse representation of multidimensional data linear decomposition and the deep structural advantages of multilayer nonlinear mapping. It will build a deep learning model with adaptive approximation capabilities. At the same time, combined with the practical problem of image classification, this paper proposes a deep learning model based on the stacked sparse autoencoder. In deep learning, the more sparse self-encoding layers, the more characteristic expressions it learns through network learning and are more in line with the data structure characteristics. It is also capable of capturing more abstract features of image data representation.

(4) In order to improve the classification effect of the deep learning model with the classifier, this paper proposes to use the sparse representation classification method of the optimized kernel function to replace the classifier in the deep learning model. It can improve the image classification effect.

(5) Based on steps (1)–(4), an image classification algorithm based on stacked sparse coding depth learning model-optimized kernel function nonnegative sparse representation is established. The algorithm is used to classify the actual images.

## 5. Example Analysis

*5.1. Daily Database Example Analysis.* This section uses Caltech 256 [45], 15-scene identification data set [45, 46], and Stanford behavioral identification data set [46] for testing experiments. All the pictures are processed into a gray scale image of $128 \times 128$ pixels, as shown in Figure 5. The images covered by the above databases contain enough categories. This is the main reason for choosing this type of database for this experiment. The method in this paper identifies on the above three data sets. The maximum block size is taken as $l = 2$ and the rotation expansion factor is 20. Randomly select 20%, 30%, 40%, and 70% of the original data set as the training set and the rest as the test set. Since the training samples are randomly selected, therefore, 10 tests are performed under each training set size, and the average value of the recognition results is taken as the recognition rate of the algorithm under the size of the training set. In order to reflect the performance of the proposed algorithm, this algorithm is compared with other mainstream image classification algorithms. The experimental results are shown in Table 1.

It can be seen from Table 1 that the recognition rates of the HUSVM and ScSPM methods are significantly lower than the other three methods. This is because the linear combination of the training test set does not effectively represent the robustness of the test image and the method to the rotational deformation of the image portion. The reason that the recognition accuracy of AlexNet and VGG + FCNet methods is better than HUSVM and ScSPM methods is that these two methods can effectively extract the feature information implied by the original training set. It facilitates the classification of late images, thereby improving the image classification effect. The accuracy of the method proposed in this paper is significantly higher than that of AlexNet and

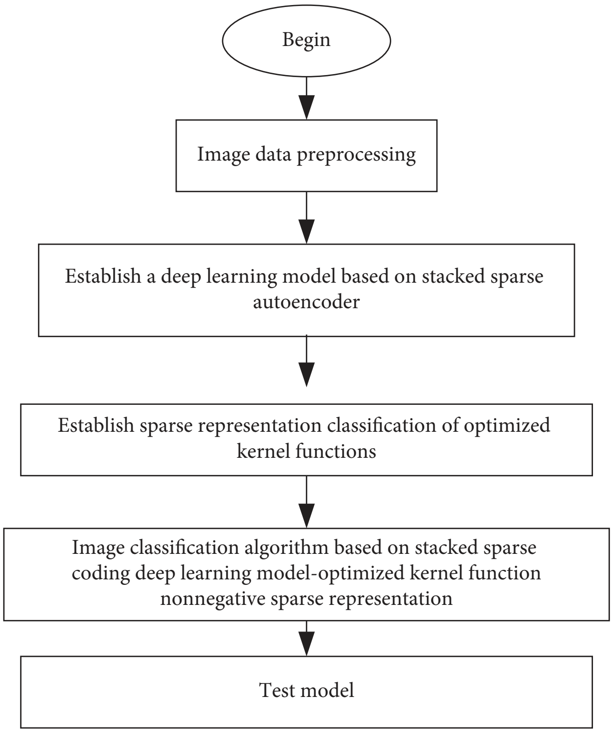
Begin
Image data preprocessing
Establish a deep learning model based on stacked sparse autoencoder
Establish sparse representation classification of optimized kernel functions
Image classification algorithm based on stacked sparse coding deep learning model-optimized kernel function nonnegative sparse representation
Test model

Figure 4: Basic flow chart of image classification algorithm based on stack sparse coding depth learning-optimized kernel function nonnegative sparse representation.

(a) (b) (c)

(d) (e) (f)

Figure 5: Sample image of the data set: (a) cannon, (b) coin, (c) duck, (d) horse, (e) microwave, and (f) mouse.

VGG + FCNet. Because although this method is also a variant of the deep learning model, the deep learning model proposed in this paper has solved the problems of model parameter initialization and classifier optimization. It can train the optimal classification model with the least amount of data according to the characteristics of the image to be tested. This is also the main reason why the method can achieve better recognition accuracy under the condition that the training set is low. In the case where the proportion of images selected in the training set is different, there are

Table 1: Different methods identify accuracy at various training set sizes (unit:%).

| Method type | The proportion of training sets. | | | |
|---|---|---|---|---|
| | 20% | 30% | 40% | 70% |
| HUSVM [47] | 82.2 | 86.8 | 91.6 | 93.8 |
| ScSPM [48] | 88.1 | 90.1 | 92.4 | 94.5 |
| AlexNet [49] | 90.5 | 92.3 | 94.5 | 97.2 |
| VGG + FCNet [50] | 91.4 | 94.7 | 95.2 | 98.1 |
| Ours | 94.5 | 96.1 | 97.8 | 99.2 |

certain step differences between AlexNet and VGG + FCNet, which also reflects the high requirements of the two models for the training set.

To further verify the universality of the proposed method. In this section, the experimental analysis is carried out to verify the effect of the multiple of the block rotation expansion on the algorithm speed and recognition accuracy, and the effect of the algorithm on each data set. The block size and rotation expansion factor required by the algorithm for reconstructing different types of images are not fixed. For example, in the coin image, although the texture is similar, the texture combination and the grain direction of each image are different. In the microwave oven image, the appearance of the same model product is the same. Although there are angle differences when taking photos, the block rotation angles on different scales are consistent. Therefore, when identifying images with a large number of detail rotation differences or partial random combinations, it must rotate the small-scale blocks to ensure a high recognition rate. For any type of image, there is no guarantee that all test images will rotate and align in size and size. Therefore, the recognition rate of the proposed method under various rotation expansion multiples and various training set sizes is shown in Table 2. When calculating the residual, the selection principle of the block dictionary of different scales is adopted from the coarse to the fine adaptive principle.

It can be seen from Table 2 that the recognition rate of the proposed algorithm is high under various rotation expansion multiples and various training set sizes. For different training set ratios, it is not the rotation expansion factor, the higher the recognition rate is, because the rotation expansion of the block increases the completeness of the dictionary within the class. On the other hand, it has the potential to reduce the sparsity of classes. So, if the rotation expansion factor is too large, the algorithm proposed in this paper is not a true sparse representation, and its recognition is not accurate. At the same time, as shown in Table 2, when the training set ratio is very low (such as 20%), the recognition rate can be increased by increasing the rotation expansion factor. When the training set ratio is high, increasing the rotation expansion factor reduces the recognition rate. This is because the completeness of the dictionary is relatively high when the training set is high. However, while increasing the rotation expansion factor while increasing the in-class completeness of the class, it greatly reduces the sparsity between classes. It will cause the algorithm recognition rate to drop.

*5.2. Medical Database Example Analysis.* In order to further verify the classification effect of the proposed algorithm on medical images. This section will conduct a classification test on two public medical databases (TCIA-CT database [51] and OASIS-MRI database [52]) and compare them with mainstream image classification algorithms.

*5.2.1. Database Introduction and Test Process Description.* In 2013, the National Cancer Institute and the University of Washington jointly formed the Cancer Impact Archive (TCIA) database [51]. The TCIA-CT database is an open source database for scientific research and educational research purposes. According to the setting in [53], this paper also obtains the same TCIA-CT database of this DICOM image type, which is used for the experimental test in this section. Some examples of images are shown in Figure 6. This paper also selected 604 colon image images from database sequence number 1.3.6.1.4.1.9328.50.4.2. The TCIA-CT database contains eight types of colon images, each of which is 52, 45, 52, 86, 120, 98, 74, and 85. The size of each image is 512 $*$ 512 pixels. For this database, the main reason is that the generation and collection of these images is a discovery of a dynamic continuous state change process. Even within the same class, its difference is still very large. Therefore, if the model is not adequately trained and learned, it will result in a very large classification error. Finally, this paper uses the data enhancement strategy to complete the database, and obtains a training data set of 988 images and a test data set of 218 images.

The OASIS-MRI database is a nuclear magnetic resonance biomedical image database [52] established by OASIS, which is used only for scientific research. The database contains a total of 416 individuals from the age of 18 to 96. The database brain images look very similar and the changes between classes are very small. To this end, this paper uses the setting and classification of the database in the literature [26, 27], which is divided into four categories, each of which contains 152, 121, 88, and 68 images. An example picture is shown in Figure 7. Figure 7 shows representative maps of four categories representing brain images of different patient information. From left to right, the images of the differences in pathological information of the patient's brain image. From left to right, they represent different degrees of pathological information of the patient. It can be seen from Figure 7, it is derived from an example in each category of the database. The classification of images in these four

Table 2: Identification accuracy of the proposed method under various rotation expansion multiples and various training set sizes (unit: %).

| Rotational expansion factor | Proportion of training set | | | |
|---|---|---|---|---|
| | 20% | 30% | 40% | 70% |
| 5 | 90.1 | 91.3 | 94.3 | 94.2 |
| 10 | 91.5 | 92.6 | **98.8** | **99.2** |
| 20 | 92.4 | **98.1** | 97.2 | 98.9 |
| 40 | 93.3 | 95.9 | 95.2 | 98.0 |
| 80 | **97.5** | 94.6 | 94.7 | 97.4 |

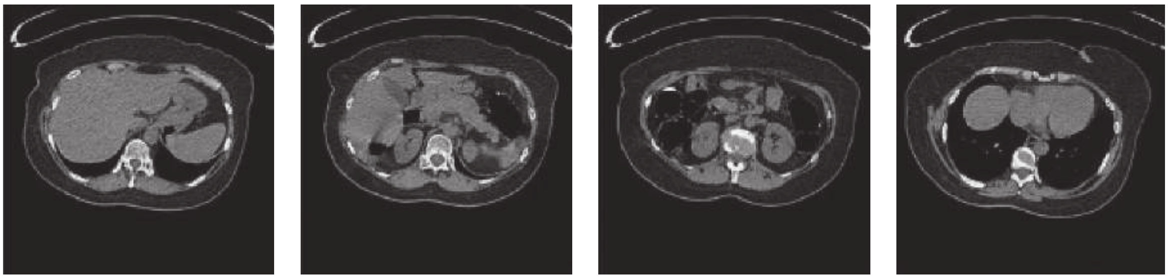
(a) (b) (c) (d)

Figure 6: Example picture of the TCIA-CT database.

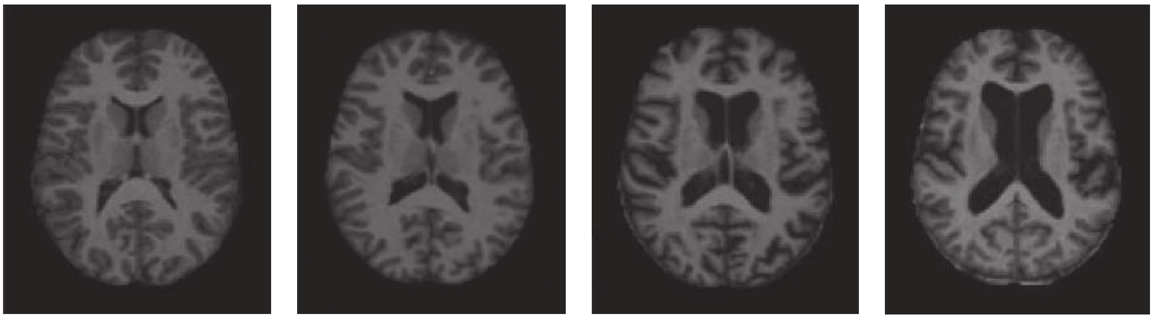
(a) (b) (c) (d)

Figure 7: Example picture of the OASIS-MRI database.

categories is difficult; even if it is difficult for human eyes to observe, let alone use a computer to classify this database. Based on the same data selection and data enhancement methods, the original data set is extended to a training set of 498 images and a test set of 86 images.

*5.2.2. Classification Results and Analysis.* The classification algorithm proposed in this paper and other mainstream image classification algorithms are, respectively, analyzed on the abovementioned two medical image databases. According to the experimental operation method in [53], the classification results are counted. The statistical results are shown in Table 3.

It can be seen from Table 3 that the image classification algorithm based on the stacked sparse coding depth learning model-optimized kernel function nonnegative sparse representation is compared with the traditional classification algorithm and other depth algorithms. The classification accuracy obtained by the method has obvious advantages. At the same time, the performance of this method is stable in both medical image databases, and the classification accuracy is also the highest. Specifically, the first three corresponding traditional classification algorithms in the table are mainly to separate the image feature extraction and classification into two steps, and then combine them for classification of medical images. The latter three corresponding deep learning algorithms can unify the feature extraction and classification process into one whole to complete the corresponding test. In general, the integrated classification algorithm achieves better robustness and accuracy than the combined traditional method. This is also the main reason why the deep learning image classification algorithm is higher than the traditional image classification method.

For the performance in the TCIA-CT database, only the algorithm proposed in this paper obtains the best classification results. The classification accuracy of the three algorithms corresponding to other features is significantly lower. The results of the other two comparison depth models

Table 3: Comparison table of classification accuracy of different classification algorithms on two medical image databases (unit: %).

| Method type | Medical database type | |
|---|---|---|
| | TCIA-CT | OASIS-MRI |
| LBP + SVM | 71.8 | 57.5 |
| HOG + KNN | 85.1 | 67.6 |
| HOG + SVM | 87.3 | 81.6 |
| DeepNet1 | 98.7 | 89.2 |
| DeepNet3 | 99.2 | 92.1 |
| Ours | 100 | 95.2 |

DeepNet1 and DeepNet3 are still very good. Although 100% classification results are not available, they still have a larger advantage than traditional methods. For the most difficult to classify OASIS-MRI database, all depth model algorithms are significantly better than traditional types of algorithms. This also shows that the accuracy of the automatic learning depth feature applied to medical image classification tasks is higher than that of artificially designed image features. The HOG + KNN, HOG + SVM, and LBP + SVM algorithms that performed well in the TCIA-CT database classification have poor classification results in the OASIS-MRI database classification. In particular, the LBP + SVM algorithm has a classification accuracy of only 57%.

In short, the traditional classification algorithm has the disadvantages of low classification accuracy and poor stability in medical image classification tasks. It shows that this combined traditional classification method is less effective for medical image classification. However, the classification accuracy of the depth classification algorithm in the overall two medical image databases is significantly better than the traditional classification algorithm. This also proves the advantages of the deep learning model from the side. In addition, the medical image classification algorithm of the deep learning model is still very stable. Among them, the image classification algorithm based on the stacked sparse coding depth learning model-optimized kernel function nonnegative sparse representation is compared with DeepNet1 and DeepNet3. It achieved the best classification performance. This is because the deep learning model proposed in this paper not only solves the approximation problem of complex functions, but also solves the problem in which the deep learning model has poor classification effect.

*5.3. ImageNet Database Example Analysis.* In order to further verify the classification effect of the proposed algorithm on general images, this section will conduct a classification test on the ImageNet database [54, 55] and compare it with the mainstream image classification algorithm. The ImageNet data set is currently the most widely used large-scale image data set for deep learning imagery. It is also the most commonly used data set for image classification tasks to be validated and model generalization performance. Due to the uneven distribution of the sample size of each category, the ImageNet data set used as an experimental test is a subcollection after screening. There are a total of 1000 categories, each of which contains about 1000 images. An example of an image data set is shown in Figure 8. In this paper, the image in the ImageNet data set is preprocessed before the start of the experimental process, with a uniform size of 256 × 256. At the same time, the mean value of each pixel on the training data set is calculated, and the mean value is processed for each pixel. The specific experimental results are shown in Table 4.

It can be seen from Table 4 that the image classification algorithm proposed in this paper has certain advantages over other mainstream image classification algorithms. At the same time, the performance of this method in both medical image databases is relatively stable, and the classification results are also very accurate. Specifically, this method has obvious advantages over the OverFeat [56] method. Both the Top-1 test accuracy rate and the Top-5 test accuracy rate are more than 10% higher than the OverFeat method. This also shows that the effect of different deep learning methods in the classification of ImageNet database is still quite different. Because the dictionary matrix $D$ involved in this method has good independence in this experiment, it can adaptively update the dictionary matrix $D$. Furthermore, the method of this paper has good classification ability and self-adaptive ability. Compared with the VGG [44] and GoogleNet [57–59] methods, the method improves the accuracy of Top-1 test by nearly 10%, which indicates that the deep learning method proposed in this paper can better identify the sample better. The Top-5 test accuracy rate has increased by more than 3% because this method has a good test result in Top-1 test accuracy. The VGG and GoogleNet methods do not have better test results on Top-1 test accuracy. These two methods can only have certain advantages in the Top-5 test accuracy. This is because the deep learning model constructed by these two methods is less intelligent than the method proposed in this paper. This method is better than ResNet, whether it is Top-1 test accuracy or Top-5 test accuracy. It only has a small advantage.

In short, the early deep learning algorithms such as OverFeat, VGG, and GoogleNet have certain advantages in image classification. In Top-1 test accuracy, GoogleNet can reach up to 78%. GoogleNet can reach more than 93% in Top-5 test accuracy. The deep learning algorithm proposed in this paper not only solves the problem of deep learning model construction, but also uses sparse representation to solve the optimization problem of classifier in deep learning algorithm. Therefore, the proposed algorithm has greater advantages than other deep learning algorithms in both Top-1 test accuracy and Top-5 test accuracy.

Figure 8: ImageNet database example diagram.

Table 4: Comparison table of classification results of different classification algorithms on ImageNet database (unit: %).

| Method type | Top-1 test accuracy | Top-5 test accuracy |
|---|---|---|
| OverFeat | 68.31 | 85.82 |
| VGG | 76.30 | 93.20 |
| GoogleNet | 78.91 | 93.33 |
| ResNet | 85.72 | 96.43 |
| Ours | 87.18 | 97.15 |

## 6. Conclusion

In this paper, a deep learning model based on stack sparse coding is proposed, which introduces the idea of sparse representation into the architecture of the deep learning network and comprehensive utilization of sparse representation of good multidimensional data linear decomposition ability and deep structural advantages of multilayer nonlinear mapping. It solves the approximation problem of complex functions and constructs a deep learning model with adaptive approximation ability. Then, a sparse representation classifier for optimizing kernel functions is proposed to solve the problem of poor classifier performance in deep learning models. This sparse representation classifier can improve the accuracy of image classification. On this basis, this paper proposes an image classification algorithm based on stacked sparse coding depth learning model-optimized kernel function nonnegative sparse representation.

The basic idea of the image classification method proposed in this paper is to first preprocess the image data. Reuse sparseness to represent good multidimensional data linear decomposition capabilities and deep structural advantages of multilayer nonlinear mapping. It solves the approximation problem of complex functions and constructs a deep learning model with adaptive approximation ability. At the same time, combined with the basic problem of image classification, this paper proposes a deep learning model based on the stacked sparse autoencoder. Then, in order to improve the classification effect of the deep learning model with the classifier, this paper proposes to use the sparse representation classification method of the optimized kernel function to replace the classifier in the deep learning model. It enhances the image classification effect. Finally, an image classification algorithm based on stacked sparse coding depth learning model-optimized kernel function nonnegative sparse representation is established. The image classification algorithm is used to conduct experiments and analysis on related examples.

This paper verifies the algorithm through daily database, medical database, and ImageNet database and compares it with other existing mainstream image classification algorithms. The experimental results show that the proposed method not only has a higher average accuracy than other mainstream methods but also can be well adapted to various image databases.

## Data Availability

The data used to support the findings of this study are included within the paper.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This paper was supported by the National Natural Science Foundation of China (no. 61701188), China Postdoctoral Science Foundation funded project (no. 2019M650512), and Scientific and Technological Innovation Service Capacity Building-High-Level Discipline Construction (city level).

## References

[1] Y. Wei, W. Xia, M. Lin et al., "Hcp: a flexible cnn framework for multi-label image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 9, pp. 1901–1907, 2016.

[2] T. Xiao, Y. Xu, and K. Yang, "The application of two-level attention models in deep convolutional neural network for fine-grained image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 842–850, Boston, MA, USA, June 2015.

[3] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: a unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823, Boston, MA, USA, June 2015.

[4] C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 2049–2058, 2015.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[6] T. Y. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125, Honolulu, HI, USA, July 2017.

[7] T. Y. Lin, P. Goyal, and R. Girshick, "Focal loss for dense object detection," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, October 2018.

[8] G. Chéron, I. Laptev, and C. Schmid, "P-CNN: pose-based CNN features for action recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3218–3226, Santiago, Chile, December 2015.

[9] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1933–1941, Las Vegas, NV, USA, June 2016.

[10] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4293–4302, Las Vegas, NV, USA, June 2016.

[11] L. Wang, W. Ouyang, and X. Wang, "STCT: sequentially training convolutional networks for visual tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1373–1381, Las Vegas, NV, USA, June 2016.

[12] R. Sanchez-Matilla, F. Poiesi, and A. Cavallaro, "Online multi-target tracking with strong and weak detections," *Lecture Notes in Computer Science*, vol. 9914, pp. 84–99, 2016.

[13] K. Kang, H. Li, J. Yan et al., "T-CNN: tubelets with convolutional neural networks for object detection from videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 10, pp. 2896–2907, 2018.

[14] L. Yang, P. Luo, and C. Change Loy, "A large-scale car dataset for fine-grained categorization and verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3973–3981, Boston, MA, USA, June 2015.

[15] R. F. Nogueira, R. de Alencar Lotufo, and R. Campos Machado, "Fingerprint liveness detection using convolutional neural networks," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 6, pp. 1206–1213, 2016.

[16] C. Yuan, X. Li, and Q. M. J. Wu, "Fingerprint liveness detection from different fingerprint materials using convolutional neural network and principal component analysis," *Computers, Materials & Continua*, vol. 53, no. 3, pp. 357–371, 2017.

[17] J. Ding, B. Chen, and H. Liu, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 364–368, 2016.

[18] A. Esteva, B. Kuprel, R. A. Novoa et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[19] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, 2016.

[20] S. Sanjay-Gopal and T. J. Hebert, "Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm," *IEEE Transactions on Image Processing*, vol. 7, no. 7, pp. 1014–1028, 1998.

[21] L. Sun, Z. Wu, J. Liu, L. Xiao, and Z. Wei, "Supervised spectral-spatial hyperspectral image classification with weighted Markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1490–1503, 2015.

[22] G. Moser and S. B. Serpico, "Combining support vector machines and Markov random fields in an integrated framework for contextual image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2734–2752, 2013.

[23] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, pp. 1150–1157, Kerkyra, Greece, September 1999.

[24] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[25] P. Loncomilla, J. Ruiz-del-Solar, and L. Martínez, "Object recognition using local invariant features for robotic applications: a survey," *Pattern Recognition*, no. 60, pp. 499–514, 2016.

[26] F.-B. Wang, P. Tu, C. Wu, L. Chen, and D. Feng, "Multi-image mosaic with SIFT and vision measurement for microscale structures processed by femtosecond laser," *Optics and Lasers in Engineering*, no. 100, pp. 124–130, 2018.

[27] J. Tran, A. Ufkes, and M. Fiala, "Low-cost 3D scene reconstruction for response robots in real-time," in *Proceedings of the IEEE International Symposium on Safety, Security, and Rescue Robotics*, pp. 161–166, Kyoto, Japan, November 2011.

[28] P. Smolensky, *Information Processing in Dynamical Systems:*

*Foundations of Harmony Theory*, University of Colorado Boulder Dept of Computer Science, Boulder, CO, USA, 1986.

[29] A. Coates, A. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *Proceedings of the fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 215–223, Lauderdale, FL, USA, 2011.

[30] J. VanderPlas and A. Connolly, "Reducing the dimensionality of data: locally linear embedding of sloan galaxy spectra," *The Astronomical Journal*, vol. 138, no. 5, pp. 1365–1379, 2009.

[31] H. Larochelle and Y. Bengio, "Classification using discriminative restricted Boltzmann machines," in *Proceedings of the 25th International ACM Conference on Machine Learning*, pp. 536–543, Helsinki, Finland, July 2008.

[32] A. Sankaran, G. Goswami, M. Vatsa, R. Singh, and A. Majumdar, "Class sparsity signature based restricted Boltzmann machine," *Pattern Recognition*, no. 61, pp. 674–685, 2017.

[33] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 21, pp. 1097–1105, 2012.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.

[36] R. Girshick, "Fast r-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, Santiago, Chile, December 2015.

[37] H. Lee and H. Kwon, "Going deeper with contextual CNN for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4843–4855, 2017.

[38] C. Zhang, X. Pan, H. Li et al., "A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, no. 140, pp. 133–144, 2018.

[39] Z. Zhang, F. Li, T. W. S. Chow, L. Zhang, and S. Yan, "Sparse codes auto-extractor for classification: a joint embedding and dictionary learning framework for representation," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3790–3805, 2016.

[40] X.-Y. Jing, F. Wu, Z. Li, R. Hu, and D. Zhang, "Multi-label dictionary learning for image annotation," *IEEE Transactions on Image Processing*, vol. 25, no. 6, pp. 2712–2725, 2016.

[41] Z. Zhang, W. Jiang, F. Li, M. Zhao, B. Li, and L. Zhang, "Structured latent label consistent dictionary learning for salient machine faults representation-based robust classification," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 644–656, 2017.

[42] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, no. 89, pp. 171–178, 2016.

[43] X. Han, Y. Zhong, B. Zhao, and L. Zhang, "Scene classification based on a hierarchical convolutional sparse auto-encoder for high spatial resolution imagery," *International Journal of Remote Sensing*, vol. 38, no. 2, pp. 514–536, 2017.

[44] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3128–3137, Boston, MA, USA, June 2015.

[45] G. Griffin, A. Holub, and P. Perona, *Caltech-256 Object Category Dataset*, California Institute of Technology, Pasadena, CA, USA, 2007.

[46] T. Xiao, H. Li, and W. Ouyang, "Learning deep feature representations with domain guided dropout for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1249–1258, Las Vegas, NV, USA, June 2016.

[47] F. Yan, W. Mei, and Z. Chunqin, "SAR image target recognition based on Hu invariant moments and SVM," in *Proceedings of the IEEE Conference on Information Assurance and Security*, pp. 585–588, Xi'an, China, August 2009.

[48] Y. Nesterov, "Efficiency of coordinate descent methods on huge-scale optimization problems," *SIAM Journal on Optimization*, vol. 22, no. 2, pp. 341–362, 2012.

[49] M. Z. Alom, T. M. Taha, and C. Yakopcic, "The history began from AlexNet: a comprehensive survey on deep learning approaches," 2018, https://arxiv.org/abs/1803.01164.

[50] R. Cheng, J. Zhang, and P. Yang, "CNet: context-aware network for semantic segmentation," in *Proceedings of the IEEE Conference on 4th IAPR Asian Conference on Pattern Recognition*, pp. 67–72, Nanjing, China, November 2017.

[51] K. Clark, B. Vendt, K. Smith et al., "The cancer imaging archive (TCIA): maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[52] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *Journal of Cognitive Neuroscience*, vol. 19, no. 9, pp. 1498–1507, 2007.

[53] S. R. Dubey, S. K. Singh, and R. K. Singh, "Local wavelet pattern: a new feature descriptor for image retrieval in medical CT databases," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5892–5903, 2015.

[54] http://www.image-net.org/.

[55] J. Deng, W. Dong, and R. Socher, "Imagenet: a large-scale hierarchical image database," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 248–255, Miami, FL, USA, June 2009.

[56] P. Sermanet, D. Eigen, and X. Zhang, "Overfeat: integrated recognition, localization and detection using convolutional networks," 2013, https://arxiv.org/abs/1312.6229.

[57] P. Tang, H. Wang, and S. Kwong, "G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition," *Neurocomputing*, no. 225, pp. 188–197, 2017.

[58] F.-P. An, "Medical image classification algorithm based on weight initialization-sliding window fusion convolutional neural network," *Complexity*, vol. 2019, Article ID 9151670, 15 pages, 2019.

[59] C. Zhang, J. Liu, and Q. Tian, "Image classification by non-negative sparse coding, low-rank and sparse decomposition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1673–1680, Springs, CO, USA, 2011.