Journal Pre-proof
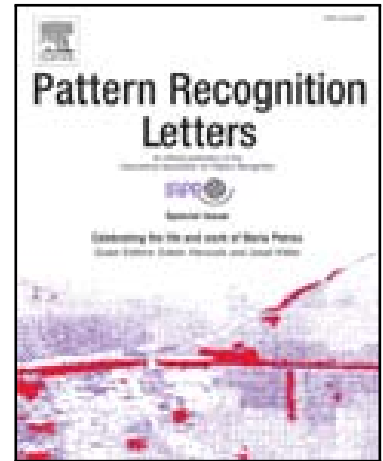
Comparative Analysis of Image Classification Algorithms Based on
Traditional Machine Learning and Deep Learning

Pin Wang , En Fan , Peng Wang

Please cite this article as: Pin Wang , En Fan , Peng Wang , Comparative Analysis of Image Classification Algorithms Based on Traditional Machine Learning and Deep Learning, *Pattern Recognition Letters* (2020), doi: https://doi.org/10.1016/j.patrec.2020.07.042

Highlights

- Representative SVM and CNN algorithms in traditional machine learning and deep learning for research.
- Under other conditions being the same, the data sets are different. The impact of the results varies.
- This article compares and analyzes the accuracy and running time.

Comparative Analysis of Image Classification Algorithms Based on Traditional Machine Learning and Deep Learning

Pin Wang1[a], En Fan2[b*], Peng Wang3[c]

Affiliation 1, School of Mechanical and Electrical Engineering, Shenzhen Polytechnic, Shenzhen 518055, Guangdong, China
Affiliation 2, Department of Computer Science and Engineering, Shaoxing University, Shaoxing 312000, Zhejiang, China
Affiliation 3, Garden Center, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650，Guangdong ,China

ABSTRACT

Image classification is a hot research topic in today's society and an important direction in the field of image processing research. SVM is a very powerful classification model in machine learning. CNN is a type of feedforward neural network that includes convolution calculation and has a deep structure. It is one of the representative algorithms of deep learning. Taking SVM and CNN as examples, this paper compares and analyzes the traditional machine learning and deep learning image classification algorithms. This study found that when using a large sample mnist dataset, the accuracy of SVM is 0.88 and the accuracy of CNN is 0.98; when using a small sample COREL1000 dataset, the accuracy of SVM is 0.86 and the accuracy of CNN is 0.83. The experimental results in this paper show that traditional machine learning has a better solution effect on small sample data sets, and deep learning framework has higher recognition accuracy on large sample data sets.

Key words: Traditional Machine Learning, Deep Learning, Support Vector Machines, Convolutional Neural Networks

1. Introduction

In the information age, pictures carry a lot of information and play an indispensable role. For massive images, it is very important to find useful image information within the effective time. Therefore, the excellent performance of the image classification algorithm has a certain influence on the image classification results. There are many image classification algorithms. The more common ones are machine learning and deep learning. Different models have different effects in different problems. Traditional machines Learning image classification algorithms and deep learning image classification algorithms have their own advantages. Comparing and analyzing image classification algorithms based on traditional machine learning and deep learning is of great significance for selecting algorithms to classify pictures.

Juan W proposed a hyperspectral remote sensing image classification method based on the improved optimal exponential factor (OIF) and support vector machine (SVM) algorithm, using the "one to one" classification strategy of support vector machine sort. The experimental results of Juan W show that the support vector machine (SVM) algorithm can effectively obtain the optimal classification band combination with high classification accuracy [1-2]. Zhao L draws on the idea of cropping KNN and uses the improved KNN classification algorithm to apply it to the object-oriented classification of high-resolution remote sensing images. Zhao L compares the classification results. Zhao L's experiment shows that under the same training set and test set, the improved KNN algorithm can achieve higher classification accuracy in high-resolution remote sensing image classification [3-4].

Image classification has always been a research hotspot, and machine learning algorithm has always been a commonly used image classification algorithm. As a branch of machine learning, deep learning has powerful functions and flexibility. In this paper, the representative SVM and CNN algorithm are selected, and the accuracy and time are compared, so as to provide a reference value for selecting the appropriate image classification algorithm. In the selection of data sets, in order to compare the algorithms from different perspectives, four types of data sets are used. It is proved that, under the same other conditions, different data sets have different effects on the experimental results of image classification. This paper is not only to verify the performance of the algorithm from a certain aspect, but also to compare the accuracy and running time, which is more convincing than the single confirmation accuracy or time.

2.Proposed Method

2.1 Machine Learning and Deep Learning

(1) Basic concepts

Machine learning is the core of the field of artificial intelligence. It is a multidisciplinary interdisciplinary subject, involving probability theory, statistics, approximation theory, convex analysis, algorithm complexity theory and other disciplines. A discipline that specializes in how computers simulate or implement human learning behavior to acquire new knowledge or skills and reorganize existing knowledge structures to improve their own performance. Generally speaking, machine learning is to learn laws from a large number of historical data through related algorithms, and make predictions or judgments on new sample data, and then learn like human beings. Deep learning is a new field in machine learning. Its motivation is to build and simulate the neural network of human brain for analysis and learning. It imitates the mechanism of human brain to interpret data, such as text, sound and image. It is a kind of unsupervised learning. Its concept comes from the research of artificial neural network, so it is also called deep neural network. Multi layer perceptron with multiple hidden layers is a kind of deep learning structure. The features obtained by deep learning are expressed level by level, and more abstract high-level semantic features are formed by combining low-level features to represent attribute categories or features, so as to discover the distributed features of data.

Most of the machine learning methods deal with data in shallow structure. These structural models have only one or two layers of nonlinear feature transformation at most, which can be

---

regarded as a structure with one layer of hidden layer or no hidden layer at all.

The reason why deep learning is called depth is relative to the above shallow learning. Different from traditional shallow learning:

1) Deep learning emphasizes the depth of the model structure, usually there are five, six or more hidden layers;

2) Highlight the importance of feature learning. We know that in the field of image recognition, extracting image features is the most critical part of the pattern recognition system. The quality of feature extraction directly affects the final recognition rate of the system. Through layer by layer feature space conversion, in-depth learning can get the most excellent expression of features;

3) Deep learning can automatically learn features from data. The features acquired from shallow structure are designed by hand. It is difficult to make use of the advantages of big data by relying on the prior knowledge and parameter adjustment experience of designers.

(2) Deep learning model

There are many kinds of deep learning models, among which convolutional neural network, deep trust network model, self coding network model, restricted Boltzmann machine model are commonly used.

1）CNN

An important role of supervisory pre training is to accelerate the training process of deep neural network. Taking convolutional neural network as an example, it is developed from neurocognitive machine according to the principle of visual field in human brain. It has the characteristics of local connection and parameter sharing, and has the excellent properties of time-shift invariance when processing data. At present, many pattern recognition systems, such as handwritten character recognition system, face recognition system and speech recognition system, have used CNN, and achieved good results.

2）DBN

Deep reliability network (DBN) is essentially a multi-layer neural network structure composed of multiple RBM networks, which can be understood as a Bayesian probability generation model. The first layer is called the input layer, the last layer is called the output layer, and the middle layers are called the hidden layer. The training principle is: first, train the first RBM layer according to the input data; then take the output of this layer as the input of the next RBM layer to continue the training; finally complete the training of the whole DBN network by repeating the previous process. The training process of DBN network is also called unsupervised training process, which realizes the reconstruction of input data.

3）AE

The automatic encoder is a special ANN. The AE model assumes that the output and the input data are the same, and then adjusts its parameters through training, so that the encoded input data can be decoded by the automatic encoder and restored to the original input data to the greatest extent, and the encoded data can be regarded as a feature extraction of the original data after being abstracted, so as to achieve this goal, It needs the ability of automatic encoder to find and summarize the characteristics of the original input data. In the process of building automatic encoder model, it may also be applied to sparse coding, de-noising coding, prediction sparse decomposition and other algorithms. The main purpose of these algorithms is to prevent the mapping of the model itself as the characteristics of data from being learned by the model and so on. In these methods, the evaluation of the effect of automatic coding is mainly based on some regularization processing, in which sparse coding can correct the deviation of hidden unit and form the effective output of hidden unit.

4）RBM

RBM consists of visible and hidden elements, both of which are binary variables. The network structure of RBM is a bipartite graph. There is an edge between the visible unit and the hidden unit, but there is no edge connection between the visible unit and the visible unit, and between the hidden unit and the hidden unit. The total probability distribution of RBM satisfies Boltzmann distribution. In the visual layer, each neural node can be understood as a characteristic representation of input data. For the visible layer and the hidden layer, when one of the network layers is known, all the nodes in the other layer can be considered as conditionally independent. Therefore, the parameters of hidden nodes can be calculated through the input of the visible layer, and the parameters of neural nodes in the visible layer can be calculated after the hidden output results are obtained. The main purpose of RBM is to train the parameters of each layer, The input data of the visible layer can be transformed into the hidden output, and the input of the visible layer can be calculated reversely from the hidden output. The hidden output can be regarded as a feature expression of the input data of the visible layer, so RBM can extract the features of the data, and the hidden unit is RBM features.

2.2 Artificial Neural Network

The emergence of artificial neural network is based on the exploration of brain working mechanism by early neurobiologists. It is a mathematical model that abstracts and simulates the way of processing information of biological central nervous system according to the knowledge of network topology. The neural network has the ability of parallel and distributed data computing, adaptive feature learning, good fault tolerance and robust robustness.

(1) Neural network model

Artificial neural network is formed by a large number of information processing units connected with each other. These information processing units are called artificial neurons. The construction of artificial neuron model originates from the structure of biological neuron. The main components of biological neuron include cell body, axon and dendrite. Artificial neuron establishes the structure of artificial neuron by simulating biological neuron. The artificial neuron is composed of three parts: multiple connection weights, a summation term and a nonlinear activation function.

1) Sigmoid function

Sigmoid function is also called logistic function. Because of its easy derivation, it has become the most frequently used activation function of early deep neural network. Its value range is from O to 1, which is suitable for two classification problems. The application effect is better when the characteristics are different. The mathematical expression of sigmoid function is:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

Sigmoid function has three main disadvantages. The first point is that in the process of deep neural network training, it is easy to lose the gradient. Sigmoid function will cause the derivative to appear 0, which makes the network weight unable to be updated. The second point is that the output of sigmoid function is always larger than o, which results in slow convergence of the model. The third point is that power operation results in the increase of training time. The above problems make sigmoid function gradually abandoned later.

2) Tanh function

Tanh function is also called double tangent function. The value range is from - 1 to L. it has a good effect in the

application of large feature differences, and it will expand the differences between features in the process of repeated training. Because the output of tanh function is O-centered, the effect of using gam function as activation function is better than sigmoid function. But tanh function also has the problems of gradient vanishing and power operation taking too long. The mathematical expression of tanh function is:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2)$$

3) Relu function

Relu function is the most frequently used activation function in recent years. Its emergence solves the problem of gradient disappearance caused by the use of sigmoid function and tanh function with the increase of network layers. The convergence speed of random gradient decline is far faster than sigmoid function and tanh function. In addition, there is no power operation, which greatly improves the convergence speed of the model. The mathematical expression of the relu function is:

$$\text{Re}\,LU = \max(0, x) \quad (3)$$

However, the relu function also has disadvantages, and the output is not centered on 0. Some neurons in the network may not be activated all the time, resulting in their weights can not be updated.

Neural network system is a network topology structure formed by a large number of neurons interconnection, including interconnection type and hierarchical structure. At present, the hierarchical model is the mainstream application structure of neural network model, and it is also the focus of this section. Generally speaking, artificial neural network is a neuron structure including input layer, multiple hidden layers and output layer. The input of each layer network is the output of the previous layer network, and the mapping process from input to output is nonlinear. Through the information transmission of each layer network, the final result output is achieved. In this mode, to calculate the output of neural network, it is necessary to carry out forward propagation step by step, input the initial vector into the input layer, calculate all the activation values of the next layer one by one, and so on, until the output layer outputs the results. Because the topology of the network is not closed-loop or loop, the network model is a feedforward neural network. Depending on the ability of these large number of neurons to process and transmit information, the complex information feature extraction and prediction tasks can be realized by using the distributed parallel processing mode to map the signal nonlinear.

(2) Back propagation algorithm

Back propagation algorithm is a learning algorithm which is widely used in artificial neural network and has good effect. The main idea is to update the parameters of feedforward neural network through error transmission and realize network training. The artificial neural network based on the back propagation algorithm is a supervised feature expression model. First, the error between the input result of the forward network and the real value is calculated. The chain derivation method is used to transfer the error layer by layer, update the weight of each layer of the network, and through repeated iterative training process, the output value of the cost function is reduced, so that the output value of the network is as close to the ideal value as possible.

The specific steps of back propagation algorithm are as

follows:

In the first step, the activation value of each neuron is obtained by using the forward conduction formula. The formula is as follows:

$$y = (wv + b) \quad (4)$$

The second step is to calculate the residual between the output result of the output layer and the output expectation.

In the third step, the network residuals of each layer are calculated in turn, and the network parameters are optimized.

In the fourth step, the gradient descent algorithm is used repeatedly to reduce the residual value and realize the final convergence of the network.

Due to its solid theory, rigorous derivation process and strong portability, back propagation algorithm is currently used in many deep learning models to train network parameters. But its disadvantages are also obvious. Firstly, the convergence time of the algorithm is too long. Secondly, when the network encounters the local minimum value, it will have a high probability that it has reached the optimal result and cannot guarantee to obtain the global optimal solution. Finally, the number of neurons in the hidden layer of the network needs to be determined through a lot of experiments, repeated parameter adjustment and optimization, and cannot be determined through theoretical calculation.

2.3 Traditional Machine Learning-Support Vector Machine

Support vector machine (SVM) is a very powerful classification model in machine learning and a common discriminant method[5]. Compared with other classification models in data mining, it has better generalization ability [6-7]. And for nonlinear separable data, it has a set of advanced theoretical methods to deal with. For linearly separable cases, the specific method is to find the line with the largest sum of distances from neighboring points to separate; for linearly inseparable cases, the kernel function is needed [8-9]. SVM is mainly applicable to two situations. The first category is linearly separable data, and the second category is linearly inseparable data [10-11]. For linearly separable data, the kernel technique is used to map the data from low-dimensional space to high-dimensional space, and then the data becomes linearly separable through techniques such as relaxation variables.

The expression of the binary classification discriminant function is written as: $g(x) = w^T x + b$, Then $g(x) = 0$ stands for hyperplane H, which is used to separate the two types of samples, and the classification rules are as in formula (5):

$$y_i(w^T x_i + b) \geq 1 \quad (5)$$

The mathematical operations in the SVM classification problem are all expressed in the form of inner products. We replace the inner product operations with kernel functions to complete the corresponding feature mapping [12-13]. At present, there are three main types of kernel functions, the formula is as follows:

(1) Polynomial kernel function:

$$K_{ploy}(x, x_i) = [(x \cdot x_i) + 1]^q \quad (6)$$

The result is a polynomial classifier of order q;

(2) Radial basis kernel function (RBF):

$$K_{ebf}(x, x_i) = \exp\left(-\frac{|x - x_i|^2}{\sigma^2}\right) \quad (7)$$

(3) Sigmoid kernel function:

$$K(x, x_i) = \tanh(v(x \cdot x_i) + c) \quad (8)$$

According to different problems, different kernel functions need to be selected [14]. Radial basis kernel functions and polynomial kernel functions are widely used because of their powerful classification capabilities [15]. At the same time, the selection of each parameter value of SVM also determines the classification effect.

Kernel function is essentially a kind of mapping from one dimension to multi dimension. When the case of linear indivisibility is mapped to multi dimension by kernel function, it may become linear separable, that is, it can be separated by method one. At this point, the linear non separable data may become the linear separable data. The kernel function is used to calculate the inner product of two vectors in the low-dimensional space in the high-dimensional space. As long as the function satisfies Mercer condition, it can be used as the kernel function.

2.4Deep Learning-Convolutional Neural Network

(1) Convolutional neural network

Convolutional neural network has a strong ability to extract local features of image. The biggest difference between convolutional neural network and traditional neural network is that partial connection network is used, and the concept of local receptive field is proposed. The traditional neural network for image feature learning, first of all, the image of two-dimensional data into one-dimensional data, into the network input layer to start training. This process destroys the spatial structure information of image, which makes it difficult for neural network to learn the spatial characteristics between pixels. In addition, the parameters of the neural network are too many, which makes the training time of the network too long. The local receptive field is inspired by the research of brain on visual information processing in biology. The visual neurons in the cerebral cortex receive a two-dimensional image local signal. Convolutional neural network is based on this principle. Each neuron in the network is only connected with the local neuron in the upper layer, and at the same time, the parameters in the network are reduced.

Another key concept of convolution neural network is parameter sharing. Convolution neural network uses convolution kernel to extract image features. Each convolution kernel can extract a feature in a local image, but in fact, different local images may have the same statistical characteristics. Using a convolution kernel, we can learn the same features, In this way, the global image can be convoluted by using convolution check to extract the same features in different positions of the global image. Through the way of parameter sharing, the parameters in the network are greatly reduced and the network training time is reduced. At present, the commonly used convolution neural network structure includes input layer, convolution layer, activation layer, pooling layer, full connection layer and output layer.

1) Input layer. The input layer is the first layer of convolutional neural network. When using convolutional neural network to process images, the pixel value matrix of an image or a local image block is generally used as the input. Convolutional neural network can input image matrix of single channel or multiple channels.

2) Convolution layer. Convolution layer is the most important part of convolution neural network, whose core operation is convolution operation. Different from the traditional full connection layer, the input of each node in the convolution layer is only a small block of the network matrix in the previous layer, which is usually 3×3 or 5×5. The convolution layer is composed of several characteristic matrices which are obtained by convolution operation of multiple convolution check input matrices. The eigenmatrix is obtained by using multiple convolution kernels to move on the input matrix and inner

product it. Through the above process, we can get the feature description of the local area of the image.

3) Activate layer. The activation function is used to activate each element of the convolution layer, so that the input-output of the network is a nonlinear mapping process, which does not change the size of the matrix.

4) Pool layer. In the pool layer, the most important features in the local feature matrix are obtained by the down sampling method, which further abstracts the dimensionality reduction features of the feature matrix, further reduces the number of parameters in the final full connection layer network, so as to reduce the parameters of the whole network, reduce the complexity of the model, reduce the possibility of over fitting problems in training, and improve the robustness of the model, At the same time, the calculation speed of the model is accelerated. The operation of pool layer usually includes taking the mean value, maximum value and random value of area matrix. Mean pooling method can effectively reduce the impact of image noise, but it destroys the image structure information. Maximum pooling can reduce the error of convolution and keep the structure information of image effectively, so it is widely used.

5) Full connection layer. The fully connected layer is a feature vector which is composed of deep layer feature permutation after the feature extraction process of multiple convolutions and pooling layers.

6) Output layer. The classifier is used to map the high-level features in the full connection layer to the category probability of the input image and output the classification result.

The training process of convolutional neural network is supervised. The advanced line propagates forward to the network, and then optimizes the network parameters through the back propagation algorithm.

(2) Image classification based on convolutional neural network

Convolutional neural network (CNN) is a special type of neural network, which is generally widely used in the field of image recognition [16-17].The CNN network structure is composed of an input layer, 2 convolutional layers, 2 pooling layers, 2 fully connected layers and output layers, a total of 8 layers [18-19]. Let the mth input graph of the convolutional layer be X_m, $W_{n, m}$ represent the convolution kernel from the mth input graph to the nth feature graph of the current layer, then the nth feature of the convolution layer node in the current layer The graph output y_n can be expressed as follows:

$$y_n = f\left(\sum_m X_m * W_{n,m} + b_n\right) \quad (9)$$

Where bn is the offset parameter of the nth feature map of the current layer, and * is the discrete convolution operator. f is the activation function of the neuron, which is generally a nonlinear mapping [20-21].

Pooling operations are divided into max pooling and average pooling, that is, maximum pooling and average pooling [22-23]. The key role of the pooling layer is to compress the image, less occurrence of overfitting, and facilitate optimization [24].After the features are obtained by convolution, all the extracted features can be used to train the classifier, such as softmax classifier, but it faces the challenge of computational complexity. For example: for a 96x96 pixel image, suppose we have learned 400 features defined on the 8×8 input, each feature and image convolution will get a (968 + 1) (968 + 1) = 7921 dimensional convolution feature, because there are 400 features, each example will get a $892 \times 400 = 3168400$ dimensional convolution feature vector.

Assuming that the image input to the pooling layer is $x^{(l-1)}$, the image after pooling is $x^{(l)}$, and down means pooling operation, then the pooling part can be defined as:

$$x^{(l)} = dowm\left(x^{(l-1)}\right) \quad (10)$$

The fully connected layer is at the tail of the convolutional neural network. It converts the two-dimensional feature map of the convolution output into a one-dimensional vector, that is, connects all the features, and finally sends the output value to the classifier, such as Softmax classifier [25]. After Softmax, the output can be expressed as:

$$S(y)_i = \frac{e^{yi}}{\sum_{j=1}^{n} e^{yi}} \quad (11)$$

3.Experiments

3.1 Experimental Environment

This experiment is written on the pycharm platform in the Windows environment, using Python language. Computer hardware environment configuration: the system is 64-bit windows10, the processor is intercorei5-8250UCPU@1.6GHz, the memory is 4.0GBRAM, and the graphics card is InterUHDGraphics620. The deep learning framework has no GPU version of tensorflow, and its version is 1.8.0.

3.2 Experimental Design

In order to compare the performance of the two algorithms in all aspects, this paper tests from the sample data size and different picture types. The image classification process is divided into training and testing. The sample set is divided into a training set and a test set. The training set is used for model training; the test set is used to detect the performance of this model and whether it can complete the image classification task more accurately. The first is to train a large number of data sets to generate a model and save it, then use the trained model to test the picture, and finally get the classification results.

(1) Large sample data set test

The large sample data set used in this paper is the mnist handwritten digital picture data set. There are 10,000 images in 28×28 format, which are 0-9 handwritten digits, as shown in Figure 1 (a). The experiment selected 2500 sheets as the test set and 7,500 sheets as the training set.

(2) Small sample data set test

The small sample data set used in this article is the COREL1000 picture set. In the experiment, the pictures in the picture set are adjusted to 96×96. The example is shown in Figure 1 (b). The picture set is divided into ten categories. Marked, in turn labeled 0-9, in the experiment selected 200 as the test set, 800 as the training set.

(a)MNIST handwritten digital picture data set

(b)Corel1000 picture collection
Figure 1. Sample picture data set

(3) Testing of different picture classifications

1) Test of different sizes of pictures

The data set used is the COREL1000 picture set. The performance of the two algorithms is evaluated from different picture sizes. The data of the sample set is normalized, and the picture size is processed as: 64×64, 128×128, and 256×256. carry out testing.

2) Different types of pictures

The data set used is the COREL1000 picture set, the size of the picture is set to 128×128 size, and then 2 types, 4 types and 6 types are randomly selected from the sample for testing.

4.Discussion

4.1 Test Analysis of Data Set Size

(1) Test analysis of large sample data set

When using the large sample data set mnist for image classification, the classification effect of the two models is shown in Figure 2.
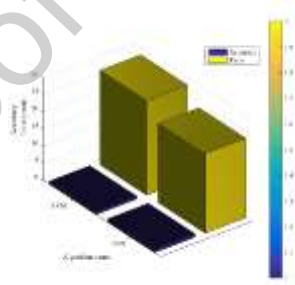

Figure 2. MNIST classification results analysis

As can be seen from Figure 2, when using a large sample mnist data set, the accuracy of SVM is 0.88, the accuracy of CNN is 0.98, the time required for SVM is 27.6min, and the time required for CNN is 23.2min.

(2) Test analysis of small sample data set

When using the small sample COREL1000 data set for image classification, the classification effect of the two models is shown in Figure 3.
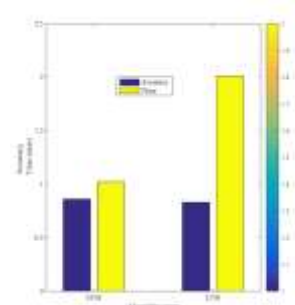

Figure 3. Analysis of corel1000 classification results

As can be seen from Figure 3, when using the small sample COREL1000 data set, the accuracy of SVM is 0.86, the accuracy of CNN is 0.83, the time required for SVM is 1.02min, and the time required for CNN is 2.01min.

4.2 Analysis of Test Results of Different Picture Classifications

(1) Test analysis of different sizes of pictures

In the case of different picture sizes, the performance of the two algorithms is compared and analyzed. The results are shown in Table 1 and Figure 4.

Table 1. Accuracy comparison of two algorithms under different picture sizes

Picture size	Number of categories	SVM classification Accuracy	CNN classification Accuracy

64×64	10	0.62	0.71
128×128	10	0.64	0.74
256×256	10	0.61	0.95

As can be seen from Table 1 and Figure 4, under the same conditions, the size of the picture has an impact on the accuracy of the algorithm. Among them, when the size of the picture is 64 × 64, the accuracy of the SVM algorithm is 0.62, and the accuracy of the CNN algorithm is 0.71 When the picture size is 128 × 128, the SVM algorithm accuracy is 0.64, and the CNN algorithm accuracy is 0.74; when the picture size is 256 × 256, the SVM algorithm accuracy is 0.61, and the CNN algorithm accuracy is 0.95.

(2) Test analysis of different types of pictures

Set the size of the picture to 128 × 128, and then randomly select 2 types, 4 types, and 6 types from the sample, as shown in Table 2, and the accuracy and running time are shown in Figure 5.
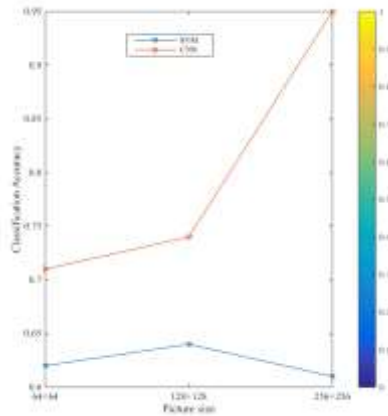

Figure 4. Comparison and analysis of the accuracy of two algorithms under different image sizes

Table 2. Test situation

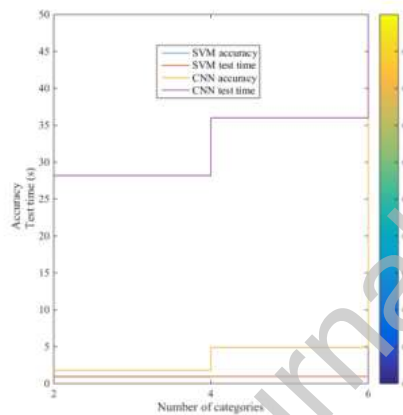| Picture size	Number of categories	Training set	Test set
128×128	2	160	40
128×128	4	320	80
128×128	6	480	120

have better capabilities.

Figure 5. Classification performance analysis of two algorithms for different types of pictures

It can be seen from Figure 5 that under other picture types with the same conditions, the accuracy of the two algorithms is almost the same, and the test time of SVM is shorter and the test time of CNN is longer.

5.Conclusions

The traditional machine learning image recognition model has various advantages, but it still has many deficiencies. In order to improve the accuracy of image recognition, the structure of deep learning models is proposed. Compare SVM and CNN algorithms.

In this paper, four different data sets are used for research and comparison. In terms of sample size and picture type, the accuracy and time of the two algorithms are compared and analyzed.

In this paper, through comparative analysis of experimental data, we can see that on small-scale data sets, traditional machine learning models have more classification advantages, and on large-scale data and recognition accuracy, deep learning models

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Juan W , Xian-Xiang W U , Yan-Ling C , et al. An image classification algorithm based on BBO-MLP and texture features[J]. Journal of Optoelectronics·laser, 2016, 27(11):1214-1219.

Luong Anh, Thi-Ngoc-Thanh Nguyen. Traffic Image Classification using Horizontal Slice Algorithm[J]. International Journal of Computer Applications, 2016, 148(11):30-34.

Zhao L , Zhang W , Sun Z , et al. Brake pad image classification algorithm based on color segmentation and information entropy weighted feature matching[J]. Qinghua Daxue Xuebao/Journal of Tsinghua University, 2018, 58(6):547-552.

Wang Y J . Image Classification Algorithm Based on Optimal Feature Weighting[J]. Zhongbei Daxue Xuebao, 2017, 38(2):196-201.

Jaesung Choi, Eungyeol Song, Sangyoun Lee. L-Tree: A Local-Area-Learning-Based Tree Induction Algorithm for Image Classification[J]. Sensors, 2018, 18(1):306.

Zhang, XB, Wang, JZ, Zhang, KQ. Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm[J]. Electric Power Systems Research, 2017, 146(2):270-285.

Noi P T , Kappas M . Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery[J]. Sensors, 2017, 18(1):18.

Saranjam Khan, Rahat Ullah, Asifullah Khan. Analysis of dengue infection based on Raman spectroscopy and support vector machine (SVM)[J]. Biomedical Optics Express, 2016, 7(6):2249-2256.

Chu D , He Q , Mao X . 1880. Rolling bearing fault diagnosis by a novel fruit fly optimization algorithm optimized support vector machine[J]. Journal of Vibroengineering, 2016, 18(1):151-164.

Omid Naghash Almasi, Modjtaba Rouhani. Fast and de-noise support vector machine training method based on fuzzy clustering method for large real world datasets[J]. Turkish Journal of Electrical Engineering and Computer Sciences, 2016, 24(1):219-233.

Nadejda Lupolova, Timothy J Dallman, Louise Matthews. Support vector machine applied to predict the zoonotic potential of E. coli O157 cattle isolates[J]. Proceedings of the National Academy of Sciences, 2016, 113(40):11312-11317.

Liu, Chuan, Wang, Wenyong, Wang, Meng. An efficient instance selection algorithm to reconstruct training set for support vector machine[J]. Knowledge Based Systems, 2017, 116(1):58-73.

Huiru Wang, Zhijian Zhou, Yitian Xu. An improved v-twin bounded support vector machine[J]. Applied Intelligence, 2017, 48(3):1-13.

Ali Anaissi, Madhu Goyal, Daniel R. Catchpoole. Ensemble Feature Learning of Genomic Data Using Support Vector Machine[J]. Plos One, 2016, 11(6):e0157330.

Wei-Chun Hsu, Li-Fong Lin, Chun-Wei Chou. EEG Classification of Imaginary Lower Limb Stepping Movements Based on Fuzzy Support Vector Machine with Kernel-Induced Membership Function[J]. International Journal of Fuzzy Systems, 2016, 19(2):1-14.

Ding, Changxing, Tao, Dacheng. Trunk-Branch Ensemble Convolutional Neural Networks for Video-based Face Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, PP(99):1-1.

Saito, Shunta, Yamashita, Takayoshi, Aoki, Yoshimitsu. Multiple Object Extraction from Aerial Imagery with Convolutional Neural Networks[J]. Electronic Imaging, 2016, 60(1):10402-1/10402-9.

Shen, Wei, Zhou, Mu, Yang, Feng. Multi-crop Convolutional Neural Networks for lung nodule malignancy suspiciousness classification[J]. Pattern Recognition, 2017, 61(61):663-673.

Yonghong Hou, Zhaoyang Li, Pichao Wang. Skeleton Optical Spectra-Based Action Recognition Using Convolutional Neural Networks[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2018, 28(3):807-811.

Nogueira R F , Lotufo R D A , Machado R C . Fingerprint Liveness Detection Using Convolutional Neural Networks[J]. IEEE Transactions on Information Forensics & Security, 2017, 11(6):1206-1213.

Fan Zhang, Bo Du, Liangpei Zhang. Weakly Supervised Learning Based on Coupled Convolutional Neural Networks for Aircraft Detection[J]. IEEE Transactions on Geoence and Remote Sensing, 2016, 54(9):1-11.

Ioannou Y , Robertson D , Shotton J , et al. Training Convolutional Neural Networks with Low-rank Filters for Efficient Image Classification[J]. Journal of Bacteriology, 2016, 167(3):774-783.

Poudel R P K , Lamata P , Montana G . Recurrent Fully Convolutional Neural Networks for Multi-slice MRI Cardiac Segmentation[J]. lecture notes in computer science, 2016, 3824(1):164-173.

Zhou W , Newsam S , Li C , et al. Learning Low Dimensional Convolutional Neural Networks for High-Resolution Remote Sensing Image Retrieval[J]. Remote Sensing, 2016, 9(5):489.

Haesol Park, Kyoung Mu Lee. Look Wider to Match Image Patches with Convolutional Neural Networks[J]. IEEE Signal Processing Letters, 2016, PP(99):1-1.

Pin Wang was born in Dingtao County, Shandong Province, P.R. China, in 1983, she received the Ph.D. from Shenzhen University, P.R. China, in 2012. Now she is a teacher of Shenzhen Polytechnic. Her research interests include cloud data fusion, signal processing, and multi-target tracking.
E-mail: wangpin@vip.qq.com

En Fan was born in Wuhan City, Hubei Province, P.R. China, in 1982. He received the Ph.D. degree from Xidian University, P.R. China, in 2015. Now he is a staff of Shaoxing University.
E-mail: efan@szpt.edu.cn

Peng Wang was Born in Dingtao County, Shandong Province, P.R. China, in 1987, He received the Master`s degree from Zhongnan University of Economics and Law, P.R. China, in 2018. Now he is a staff of Garden Center, South China Botanical Garden, Chinese Academy of Sciences.
E-mail: sdhzdtwp@126.com

Graphical abstract