Journal of Physics:
Conference Series

PAPER • OPEN ACCESS

A Comparison of Traditional Machine Learning and Deep Learning in Image Recognition

To cite this article: Yunfei Lai 2019 *J. Phys.: Conf. Ser.* **1314** 012148

A Comparison of Traditional Machine Learning and Deep Learning in Image Recognition

Yunfei Lai[1]

[1]College of mechanical engineering, Wuhan University of Technology, NO. 1178, Yangyuan Road, Wuchang District, Wuhan, P. R. China

Email: 1246605084@qq.com

Abstract. The growth of the mobile Internet, smartphones and social networks has brought in huge amounts of picture information, and traditional manual identification is not able to meet the demand well enough. Therefore, the automatical image recognition [1] has been proposed which can help us recognize the image efficiently and get the corresponding information. Although traditional machine learning methods [2] have already been widely used in the field of image recognition, most of these methods are designed to handle one-dimensional vector information. Thus, we should first stretch image matrix to one-dimensional vector or extract features from images to employ traditional image recognition methods, which would lose the adjacent information in images and miss some important features. With the development of computer technology, deep learning [3] is gradually applied to the field of image recognition. It can deal with two-dimensional image data naturally and extract features automatically. Compared with the traditional machine learning methods, deep learning is popular for its good learning ability and low generalization error. In this paper, we compare the differences between SVM [4] and deep learning on image recognition, with an application to handwritten digital images recognition. The results show that the deep learning method is more accurate and more stable in image recognition.

1. Introduction

Image recognition technology has been widely used in our life. For example: The scene recognition through deep learning algorithm which can automatically recognize some common scenes in images, such as sky, grass, people, and so on. Based on that function, the client's application can easily realize the automatic management, grouping and search images, complete the intelligent management of large image library, and save a lot of time. The significance of image recognition technology is that it frees people from heavy and mechanical repetitive work and gives them more time to deal with other more meaningful things, thus greatly increasing the efficiency of working.

The traditional machine learning algorithmic system has been very mature, and a lot of them (such as SVM) can be applied to image recognition. For the present computer processing and computing ability, it is easy to use traditional machine learning method to solve some problems. However, when these methods are applied in the field of image recognition, there are some obstacles to overcome. Traditional machine learning method can only apply to deal with one-dimensional data while the images showed up as a matrix form.

There are mainly two methods proposed to handle this problem. One is to stretch a matrix line by line or column by column into one-dimensional vectors. This method is relatively easy to implement, but by doing so, the position relation of rows and rows in the original matrix will be lost, which will cause some problems that the important information can't embody. For example, the color features among several neighboring pixels in the image are similar, but they are no longer adjacent after being stretched into vectors, which makes it difficult for the machine to recognize the image with the similar features lost. The second method is to extract the feature information of the image step by step, such as using histogram form to count the area of a specific color in the picture, or the number of lines with certain shape in the picture to identify the image content. This method is similar to the human image recognition method. But the disadvantage of this approach is also clear: after a picture is entered, the machine has no way of knowing which features are useful and which are redundant. Then in order to recognize the image better, the machine needs to count and calculate a large number of features, consuming a lot of unnecessary time and computing resources. Generally speaking, although traditional machine learning algorithms have their own advantages, they are not suitable for the task of image recognition.

Recently, deep learning is widely used for image recognition. In deep learning, the image can be directly used as the input of image recognition network. Different from the traditional algorithm, where the original image is processed by splicing or extracting features, the depth learning algorithm automatically extracts the feature from the image by convolution of the images. After convolution is completed, the algorithm processes the convolution image data at different resolution scales by pooling[5] or upsampling[6]. This method can extract the feature information in different resolutions and make the image information more complete. Such a method of information extraction makes the deep learning have strong learning ability that the traditional machine learning method does not have, which makes it more accurate when dealing with the problem of image recognition. But the strong learning ability also brings some problems. In the training concentration, the deep learning mode will exist the problem of over-fitting[7], which makes the actual recognition effect worse and limits its ability. In order to make the depth learning achieve higher recognition accuracy, weight sharing of convolution kernel[8] can be used to reduce the overall computation, so as to avoid over-fitting in the algorithm. In conclusion, the deep learning method has stronger learning ability and it is more suitable to solve the problem of image recognition.

In this paper, we test and compare the advantages and disadvantages between traditional ML methods (we choose SVM as the representative) and deep learning. After the preparatory work, we use SVM to build a recognition model. Since the image data we use are linear and inseparable, we use radial basis function to make the data linear separable. Then we stretch the image data into a vector and transfer the transformed vector data to the SVM model in parts. The accuracy of the model in training sets is between [90.7, 98.0], and the accuracy in testing sets is between [90.4, 98.8].

Then we reconstructed a deep learning network, which has two convolution layers and two full-connection layers, to categorize numbers written by hand. Finally, the accuracy rate of the model on the training set can reach 99%, while the accuracy rate on the test set was stable above 99.5%.

The rest part of this article will be organized as follows. In section 2, there will have a brief introduction about the background of the data samples that were used in this article. In section 3, some details about the SVM model will be contained. In section 4, the passage will write about the deep learning network that used in our experiment. Finally, we compared the traditional method with the deep learning, and summarize the article.

2. Data description

The image data used in this experiment is Mnist[9] dataset, which is an open handwritten digital dataset. The digit images in Mnist dataset were naturally selected and experimented with by Chris. Burges and Corinna Cortes. The digits have been size-normalized and centered in a fixed-size image (28 X 28).

This data set can be found at http://yann.lecun.com/exdb/mnist.com, where the original image data downloaded are saved as idx3-ubyte format and we transformed them into Jpeg format. Three sample images are shown in Figure 1.

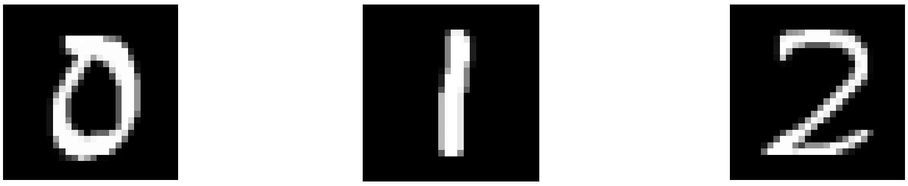

Figure 1: Three sample images of Mnist dataset

Of the 70,000 images in the entire training set, 60,000 images were used as a training set, and the remaining 10, 000 images were used as test sets. To explore the effectiveness of both models in identifying all handwritten digits, the containing status of numbers in training streets and test sets are summarized in Table 1.

Table 1: The Containing Status of Numbers in Training Set and Test Set

Number / Data	0	1	2	3	4	5	6	7	8	9
Train	6000	6000	6000	6000	6000	6000	6000	6000	6000	6000
Test	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

In the above table, we can see that the number of each number in the training set and the testing set is the same, and each number in the training set is 6:1 compared with the number in the testing set. This will ensure the model can be fully trained in different data before testing.

3. SVM Algorithm for Mnist Data Recognition

SVM Algorithm (Support vector machine) as a classical algorithm, it was proposed by Vladimir N. Vapnik[10] and Alexey Ya. Chervonenkis[11] in 1963, but it can only deal with linear classification problems. later on, Bernhard E. Boser[12], Isabelle M. Guyon[13] and Vladimir N. Vapnik[14] improved it with a method called 'kernel trick'[15], which they proposed in 1993. After that, SVM algorithm can solve the problem of nonlinear classification. The principle of the SVM algorithm is similar to the linear regression problem in mathematics. It can distribute the learning samples into two different classes in the high-dimensional space by linear classification and make use of these data to create a non-probabilistic binary linear classifier. Once a new sample is analyzed, the SVM model will compare the new model with the data from the learning sample and judge the new model category by the position of the new sample in the original space.[16] (Basic principle of SVM is showed in Figure 2)

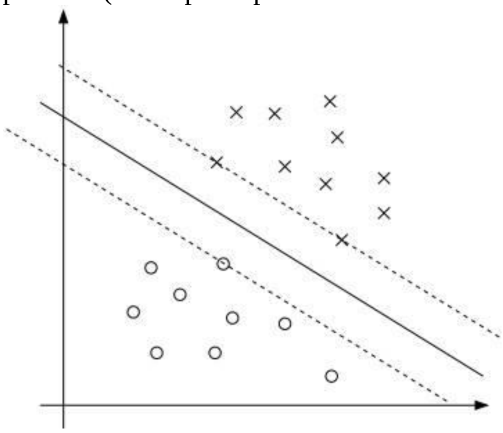

Figure 2: The Principle of SVM Algorithm

The quality of the model depends on whether the linear classifier correctly classifies points in the high dimensional space. The idea of SVM is to find a hyperplane to separate the sample data, so that the total distance from the hyperplane to the two nearest sample point in each side can be maximized.

In practical applications, we are faced with nonlinear separable problems in many cases. In order to deal with these nonlinear problems, James Mercer[17] proved the Mercer theorem[18] in the early 20th century and laid the foundation for the emergence of the kernal method. In brief, the kernal method maps the linear inseparability problem in the low-dimensional space into the high-dimensional space, which makes the problem linear separable in high-dimensional space. Kernal can be combined with many arithmetic operations, and the SVM used in this experiment is one of them. Kernal has many branches, and in this experiment, we used RBF (radial basis function kernal)[19] as the basic method.

The SVM model is designed for binary classification, but we have 10 numbers here which is a multi-classification problem. There are two kinds of methods to solve this problem in SVM: OvO scheme[20] and OvR scheme[21]. OvO (one vs one) scheme is that we compare all the categories in pairs. For this experiment, the scheme will generate 10*(10-1)/2 binary tasks based on data. The final predictions are decided according to the votes of these binary tasks. The OvR (one vs rest) scheme takes one category as positive sample and other categories as negative samples. In this experiment, the model automatically generates 10 classifiers after receiving new images. Similarly, the final prediction is made by voting method. In this experiment, in order to increase the efficiency of identification, we adopt the OvR method.

To make the calculation more convenient, we first resize the resulting image to a 32*32 image format. Then we reshape the image matrix to a vector of 1024 length, and use the data in this form as input to the SVM model. We take the advantage of RBF and use OvR multi-classification to train our SVM model. In our experiment, the SVM model totally spend 11.63min running on the training set and the confusion matrix of training set is presented in Table 1.

Table 2：Confusion Matrix of Training Set

Predict / True	0	1	2	3	4	5	6	7	8	9
0	5803	1	13	11	13	24	22	7	26	3
1	2	6603	41	19	11	15	5	7	29	10
2	38	25	5554	45	73	22	55	50	83	13
3	19	29	**100**	5604	9	**173**	21	58	86	32
4	11	20	26	0	5557	4	25	16	7	**176**
5	31	51	29	**134**	48	4981	79	5	42	21
6	37	13	31	0	33	71	5719	0	14	0
7	11	36	55	14	70	12	2	5922	12	**131**
8	21	**105**	52	84	37	**136**	32	16	5319	49
9	26	21	24	86	**184**	28	1	**147**	39	5393

Generally speaking, the performance of SVM is pretty well in training set for it owns 94.09% total accuracy rate, and the accuracy of model is stable between 90.6% and 98% for each class. However, due to the data in the obfuscation matrix, we found that the SVM model is not satisfactory for some special numbers, which have been coarsely identified in Table 1. For example, the distinction between the number 4 and the number 9 is deficient in SVM, where 184 examples of 9 are projected to be 4 and 176 examples of 4 are predicted to be 9. Through the analysis, we find that numbers 4 and 9 have similar writing characteristics and contain similar circular structure, which is easy to confuse the SVM model. The same thing happens between 3 and 5, 7 and 9. This also shows that the SVM model is not suitable to handle image recognition tasks with similar features in the pictures.

Table 3: Confusion Matrix of Testing Set

True \ Predict	0	1	2	3	4	5	6	7	8	9
0	968	0	1	1	0	6	1	1	1	1
1	0	1120	2	3	0	4	3	0	3	0
2	9	2	954	8	13	1	12	11	**19**	3
3	0	1	16	942	0	**21**	3	11	12	4
4	1	1	4	1	930	0	9	3	2	**31**
5	7	6	4	**35**	6	806	10	3	12	3
6	7	3	3	1	6	11	926	0	1	0
7	2	**15**	**22**	6	9	0	0	950	3	**21**
8	3	5	8	**18**	10	**26**	10	6	882	6
9	7	5	1	13	**37**	8	1	**16**	7	914

The confusion matrix of SVM on the testing set is listed in Table 3. Similarly, SVM performs well in digital number recognition and get 93.92% total accuracy. The accuracy rate of model is stable between 90.6% and 98.8% for each class. The accuracy of our SVM model is comparable, which implies that it does not over-fit the training set. The confusion matrix shows a similar pattern for training set and testing set. It is more difficult to distinguish between 4 and 9,3 and 5,7 and 9 for SVM.

4. Deep Learning for MNIST Data Recognition

Deep learning is a kind of machine learning method based on neural network. It refers to the thinking mode of the animals' brain, and can recognize the features automatically without any supervision. It can process big data accurately, thus saving a lot of human resources. The term deep learning was first introduced in 1986 by Rina Dechter[22] in the field of machine learning while the first complete in-depth learning network algorithm was published by Alexey Ivakhnenko[23] and Lapa[24] in 1965. Because of its high accuracy and efficiency, deep learning algorithms are now widely used in various areas of life, including medicine, unmanned, machine translation, image restoration and even armament research and development.

Neural network is a machine learning framework that processes data by connecting point to point between different layers. It can extract features accurately. Deep learning adds convolution, pooling and other means to this framework, which can effectively extract the features of the pictures.

First, to make computing and processing easier, we resize the images to 20*20 and convert the images into 0-1 images. Then, we construct a three-layer convolution neural network, and the network structure is shown in Figure 3.

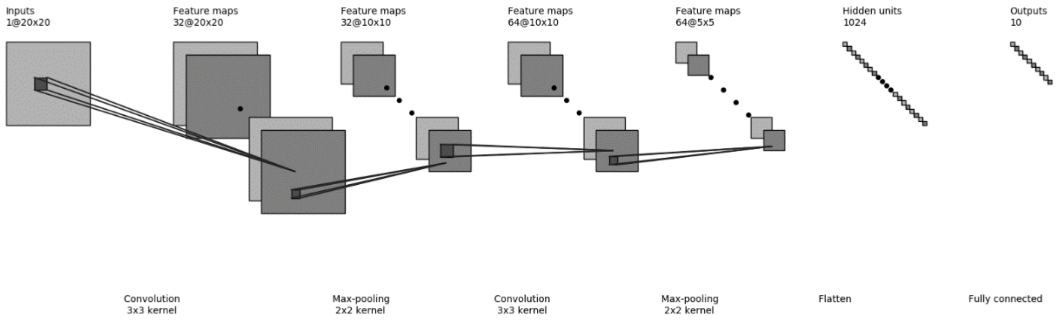

Figure 3: The Diagram of Neural Network

In the Layer 1 network, we deploy 32 3*3 convolution kernels to collect the feature information of the image. In order to keep the size of the image invariably before and after convolution, we use the method called padding to add 0 to the image after convolution to get 32 20*20 pixels images after convolution. To reduce the possibility of over-fitting and reduce the size of the output data, we employ max pooling with kernel size 2*2 on these images and finally we get 32 10*10 images. In the second-level convolution network, we set up two convolution layers that convert the first-layer output image data into 64 10 x 10 images. Then the max pooling with size 2*2 converts these images to 5*5 images.

We then processed the data on the former floor, in the Flatten layer, we integrate the 64 5*5 images from the second layer into a 5*5*64 vector and reshape it into a 1600-length vector. Then we use a full connected layer to convert this vector into a vector of 1024 length. In the full connected layer, we convert the vector length to a vector length of 10, and finally we use Softmax function[25] activation as the final output. To ensure the correct rate, we performed 5000 iterations on the data. During this experiment, the deep learning model ran a total of 11h50m41.2s on the training set, and the confusion matrix on the training set was displayed in the Table 4.

Table 4: The Confusion Matrix of Deep Learning on Training Set

Predict / True	0	1	2	3	4	5	6	7	8	9
0	5923	0	0	0	0	0	0	0	0	0
1	0	6742	0	0	0	0	0	0	0	0
2	0	0	5958	0	0	0	0	0	0	0
3	0	0	0	6131	0	0	0	0	0	0
4	0	0	0	0	5842	0	0	0	0	0
5	0	0	0	0	0	5421	0	0	0	0
6	0	0	0	0	0	0	5918	0	0	0
7	0	0	0	0	0	0	0	6265	0	0
8	0	0	0	0	0	0	0	0	5851	0
9	0	0	0	0	0	0	0	0	0	5949

It can be seen that the accuracy rate of the deep learning model on the training set reaches 100%, which is far higher than the result of the SVM model. That result means that all the handwritten image data are correctly classified by the model, which shows that the learning and fitting ability of the deep learning algorithm is very strong. Compared with the SVM model, in the training set, the deep learning model does not classify the image in the wrong selection because of the similarity of the image features, which indicates that the depth learning is better than the traditional machine learning method in doing image recognition work.

Table 5: The Confusion Matrix of Deep Learning on Testing Set

Predict / True	0	1	2	3	4	5	6	7	8	9
0	974	1	1	0	0	0	2	1	1	0
1	0	1133	0	0	0	0	1	1	0	0
2	1	1	1021	0	1	0	0	4	4	0
3	0	0	1	999	0	**7**	0	0	3	0
4	0	1	0	0	970	0	1	1	0	**9**
5	1	0	0	**7**	0	883	1	0	0	0
6	3	3	0	0	2	1	945	0	4	0
7	0	4	5	1	2	0	0	1012	0	4
8	3	0	1	3	0	0	0	2	962	3
9	0	3	0	0	**10**	2	0	**6**	2	986

The confusion matrix of the deep learning algorithm on the test set is shown in Table 5. On the test set, we see that the accuracy for each number is between 97.7% and 99.4% the overall correct rate was 98.85%. In Table 5, compared with SVM, the error rate of the deep learning model is relatively low. Besides, the digital patterns of 9-4,3-5 that SVM has some difficulties to distinguish seems have not troubled the deep learning method. Although there still exist some mistakes, the error rate is very low. This shows that deep learning in the recognition of handwritten numerals is better than SVM algorithm, and can learn the characteristics that are not easy to be distinguished in the picture. At the same time, the high accuracy on the test set also shows that the depth learning model has strong generalization ability.

5. Conclusion

Image recognition has become an indispensable technology in people's life, from the input of handwritten text to the application of military, it can be said to be ubiquitous. Therefore, the research on image recognition technology is very promising and meaningful. This technique can be implemented either by traditional machine learning methods or by deep learning algorithms. In this paper, we first used a classical machine learning algorithm, SVM, as a representative of traditional machine learning methods to analyze MNIST handwritten digital image data. We also built a three-layer CNN to recognize these handwritten digits, and compared the results of CNN to that of SVM. The details for these comparisons are summarized in Table 6.

Table 6: Comparisons of SVM and Deep Learning

Methods / Comparison	SVM	Deep Learning
Operating Time	46m54s	11h50m41.2s
Accuracy in Training Set	94.09%	100%
Accuracy in Testing Set	93.92%	98.85%
Means for Extracting Features	Manually and Subjective	Automatically and Objective
Means for Processing Data	Turn Images into Vector	Directly Using Images

Although deep learning needs more training time and more computing resources, it can achieve better prediction accuracy and generalization ability, which indicates that deep learning has better learning ability. Compared with the traditional machine learning, deep learning can extract features from the image automatically and efficiently. The deep learning method can clearly distinguish the image with the similar characteristics which is difficult for traditional machine learning methods to recognize. What's more, deep learning can extract features objectively by itself and can directly process two-dimensional image data, unlike traditional machine learning methods, which need subjective feature extraction to convert binary vectors into one-dimensional vectors. In a word, deep learning is more suitable for the modeling of image data. Its strong feature extraction ability and learning ability are not acquired by traditional machine learning methods.

References

[1] Olague G, Cagnoni S, Lutton E. Introduction to the special issue on evolutionary computer vision and image understanding[J]. Pattern Recognition Letters, 2006, 27(11):1161-1163.
[2] Liu H, Cocea M. Traditional Machine Learning[J]. 2018.
[3] Deng L, Yu D. Deep learning: methods and applications[J]. Foundations & Trends in Signal Processing, 2014, 7(3):197-387.
[4] Rakotomamonjy A. Variable Selection Using SVM-based Criteria[J]. Journal of Machine Learning Research, 2003, 3(7-8):1357-1370.

[5] Wischik D, Handley M, Braun M B. The resource pooling principle[J]. ACM SIGCOMM Computer Communication Review, 2008, 38(5):47.
[6] Shan Q, Li Z, Jia J, et al. Fast image/video upsampling[J]. ACM Transactions on Graphics, 2008, 27(5):1.
[7] Schaffer C. Overfitting avoidance as bias[J]. Machine Learning, 1993, 10(2):153-178.
[8] Gao L, Chen P Y, Yu S. Demonstration of Convolution Kernel Operation on Resistive Cross-Point Array[J]. IEEE Electron Device Letters, 2016, 37(7):1-1.
[9] Chervonenkis A Y. Early History of Support Vector Machines[M]// Empirical Inference. 2013.
[10] Boser B E, Guyon I M, Vapnik V N. Algorithm for Optimal Margin Classifiers[C]// Computational Learning Theory. 1992.
[11] Hai Y. An Incremental Learning Algorithm for SVM based on Voting Principle[C]// International Conference on Advanced Information Management & Service. 2010.
[12] Mercer J. Functions of positive and negative type and their connection with the theory of integral equations[J]. Philosophical Transactions of the Royal Society of London, 1909, 209:415-446.
[13] Park J. Approximation and radial-basis-function networks[M]. MIT Press, 1993.
[14] Pan L, Zhang D, Jiang N, et al. Improving robustness against electrode shift of high-density EMG for myoelectric control through common spatial patterns. [J]. Journal of Neuroengineering & Rehabilitation, 2015, 12(1):1-16.
[15] Rina Dechter (1986). Learning while searching in constraint-satisfaction problems. University of California, Computer Science Department, Cognitive Systems Laboratory.
[16] Chen K, Kvasnicka V, Kanen P C, et al. Multi-Valued and Universal Binary Neurons: Theory, Learning, and Applications [Book Review][J]. IEEE Transactions on Neural Networks, 2001, 12(3):647-647.
[17] Jörg Lücke, Sahani M. Generalized Softmax Networks for Non-linear Component Extraction[C]// Artificial Neural Networks-icann, International Conference, Porto, Portugal, September. DBLP, 2007.