

PAPER • OPEN ACCESS

Application of deep learning in image recognition

To cite this article: Lingyun Li 2020 *J. Phys.: Conf. Ser.* **1693** 012128

View the [article online](#) for updates and enhancements.

You may also like

- [Research on Image Recognition Algorithm Based on Computer Deep Learning](#)
Shuqiang Du
- [Research on Small Sample Image Recognition Based on Transfer Learning](#)
Guhai Yang, Baihui Tang and Sanxing Cao
- [Research on the Application of Deep Learning in Computer Image Recognition](#)
Miaofen Feng

Application of deep learning in image recognition

Lingyun Li

Rose-Hulman Institute of Technology, Computer Science

5500 Wabash Ave, Terre Haute, IN 47803

robinlingyunli@gmail.com

vivian.wang@cas-harbour.org

Abstract. Deep learning is a new field in machine learning, which is based on the neural network model to extract features by simulating human brain cognition. Image recognition is an essential subject in the field of deep learning and has made significant progress in recent years. In the training process, only the computer needs to cooperate, and no personnel need to participate in the process to get a good image recognition effect. This paper makes an in-depth analysis on the application of deep learning in image recognition, discusses the necessary steps of image trial and the basic principle of the neural network, and puts forward some problems and popular deep learning models used in image recognition.

1. Introduction

With the development of image technology, every moment of our life contains a lot of image information, but the speed of the human visual system to process the image is very limited, far short than the speed of the image produced. Therefore, people started to use the computer to achieve more accurate identification of images, and obtain image data that cannot be seen by human eyes. In image classification, recognition, and other problems, a deep learning method can simulate how the human brain responds to images, and analyze the image data more thoroughly. It has an excellent performance in large-scale image processing and has been widely used in many fields. Compared with other neural network structures, the convolutional neural network requires relatively few parameters, which enables it to be widely used. This article will introduce deep learning techniques in the field of the image. In the second part, the principle and development of deep learning will be introduced. In the third part, the three steps of image recognition are introduced: preprocessing, feature extraction, and classification and recognition. In the fourth part, a standard deep learning model for image recognition will be introduced. Finally, two neural networks that have achieved good rankings in the Large Scale Visual Recognition Challenge will be introduced.

2. The Basic principle of Deep Learning and Development

2.1. Principle

Deep learning is a new field extended from machine learning, aiming at building a neural network that can analyze and learn data like the human brain, and a series of new algorithms generated with the improvement of hardware computational power.[1] Deep learning is a form of data-based learning. The specific learning process is generally understood as that the computer runs data iteratively and updates the parameters between the layers of the deep learning network to make the training result



closer to the real value. The ultimate goal of deep learning is to learn the internal rules or trends of sample data and extract the acquired information such as characters and images so that machines can have the ability of analysis and learning like the human brain and can recognize characters and images and other information. In recent years, many countries have made breakthroughs in the research field of deep learning, especially the application of deep learning in the field of image recognition, which can effectively solve many difficulties in recognizing intricate patterns. It can be said that deep learning has effectively promoted the rapid development and progress of image recognition technology.

2.2. Development

The development process of machine learning is divided into two periods: the first is the shallow learning period, and the second is the deep learning period. At present, most of the classification, regression, and other learning algorithms belong to the shallow structure. The shallow structure usually contains only two or less nonlinear feature layers, and the typical shallow structure includes logistic regression (LR) and support vector machine (SVM). SVM is one of the most successful classification models.

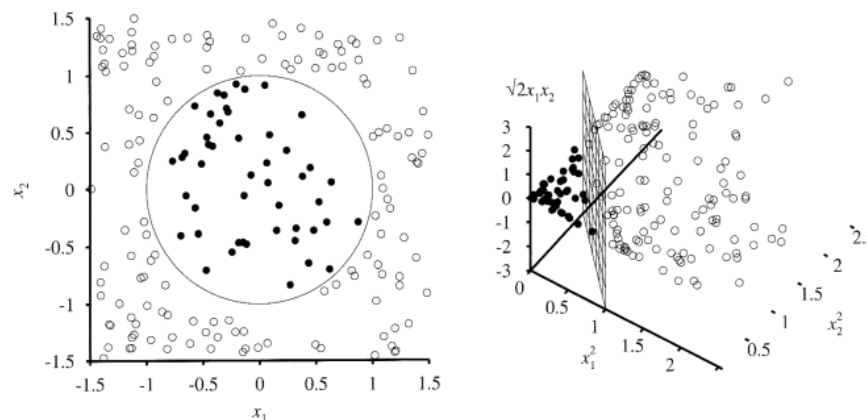


Fig.1 Feature Space Representation [2]

The SVM algorithm is used to achieve the optimal partition, aiming to maximize the interval between the training set and classification hyperplane. [3] When different categories of data vectors in a lower-dimensional space cannot be separated, it USES a specific function to map the data set to a higher-dimensional data space that allows for linear partitioning of existing data. The limitation of the shallow model lies in the representation ability of complex functions, and the analytical ability of complex classification problems will be significantly limited, which makes it difficult to solve some more complex problems, such as natural language processing and image recognition. On the contrary, deep learning can analyze some complex data by increasing the parameters of the learning model and making it have a deeply nonlinear network structure.

The shortcoming that shallow learning cannot analyze complex data effectively promotes the further research of deep learning algorithm. The rapid growth of deep learning has had a positive impact in most areas. For example, the application of the convolutional neural network plays a useful role in promoting the development of image recognition technology in the future. Besides, the recursive neural network model (RNN) has made an outstanding contribution to improving the accuracy of natural language processing. Also, deep learning has achieved good results in chess competitions, beating the world champion in chess several times in a row.

3. General Process of Image Recognition

Deep learning is widely used in image recognition due to its advantages, such as strong feature extraction ability and high recognition accuracy. With the help of computer technology, image recognition can be effectively expanded and preprocessed, features extracted, classification and

recognition. The image recognition framework is usually divided into three stages: preprocessing, feature extraction, classification and recognition of image results.

3.1. Preprocessing

In image recognition, preprocessing is the first step. Before adding the algorithm, some processing steps are done on input images to get inputs that are clearer or more consistent with the algorithm.

In the process of preprocessing, the relevant data of the image should be read first. The image data is then stored in the form of 0 and 1. The storage form of a color image in a computer is usually a two-dimensional matrix, and the two dimensions are the width and height of the image. The three-dimensional matrix of color images can be divided into three two-dimensional matrices, namely R, G and B. The elements in the matrix respectively represent the brightness of R, G and B at the corresponding position of the image, with the range from 0 to 255. One of the preprocessing methods for color images is normalization. To some extent, this method can be understood as changing the pixel value from 0 to 255 to 0-1, reducing its distribution distance and thus reducing data storage and computation.

3.2. Feature extraction

In image recognition, feature extraction is an important step. The original image, which is based on pixels, belongs to the signal data and has a large amount. The classifier has no way to judge these pixels. It can only recognize the high-level data information, which is extracted from the content part of the input image, namely the image features. This process is called feature extraction. For images, each of them has its features that can be distinguished from others. Some are natural features that can be intuitively felt, such as brightness and color. Others require further transformation or processing, such as principal components. Principal Component Analysis (PCA), a method for dimensionality reduction by a linear combination of features, is one of the attractive approaches to transformation these hidden features. The essence of PCA is to project the sample data in the high-dimensional space to the low-dimensional space through linear transformation on the premise of representing the original data as best as possible.

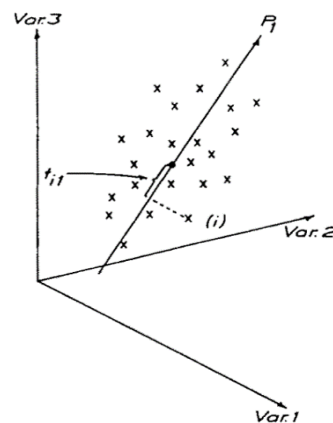


Fig. 2 A one-component PC model [4]

3.3. Classification and recognition

The training of classifier is particularly crucial for machine learning, increasing the number of iterations and continuously training the model parameters of the algorithm to make the classification performance better. After selecting feature examples with moderate computing amount and desired accuracy, traditional machine learning methods such as logistic regression, K-nearest neighbor, or Random Forest are used to train a classification model, which can be used to classify images. Furthermore, achieving optimal performance of the selected model by repeatedly adjusting the hyper-

parameters of the classifier. With the optimal parameters, the machine can classify the input data into a particular category accurately.

4. Deep Learning Model for Image Recognition

Deep learning is widely used in image recognition due to its advantages, such as strong feature extraction ability and high recognition accuracy. Comparing with some standard networks, such as RNN, Convolutional Neural Networks (CNN) has a noticeable effect on image recognition.

Conventional neural networks is a kind of deep learning model specially designed for image classification and recognition developed based on multi-layer Neural Networks. A typical deep learning model is demonstrated below:

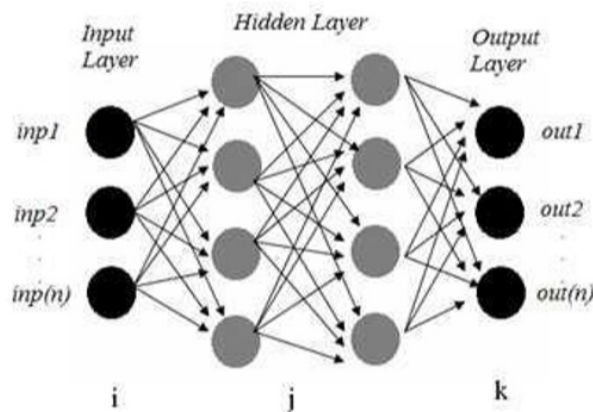


Fig. 3 Neural network architecture [5]

A typical deep learning model is a multi-layer neural network consists of an input layer and an output layer, with several hidden layers in the middle. There are several neurons in each layer, and each neuron in these layers are connected with a specific weight as the parameter. For each neuron in these hidden layers, a specific bias is added to make a further adjustment. In general recognition problems, the image data is passed through the input, and during some multiplications and additions, the final result is returned in the output layer.

In the image recognition problem, each neuron of a deep learning model in the input layer may represent a feature of the input image. However, there are several problems in this neural network model during image recognition. One is that the spatial structure of the image is not considered so that the recognition performance is limited. Second, the neurons in every two adjacent layers are fully connected, so that the training speed is slow due to an excess of parameters.

Unlike conventional deep learning models, convolutional neural networks can solve these problems. It uses a unique structure for data analysis and can be trained quickly than a conventional deep learning model. Because of the high speed, it is possible to have a large number of neural network layers, which has a significant advantage in recognition accuracy.

There are three basic concepts of a convolutional neural network: local receptive fields, shared weights, and pooling. Each idea is demonstrated below for the case of image recognition.

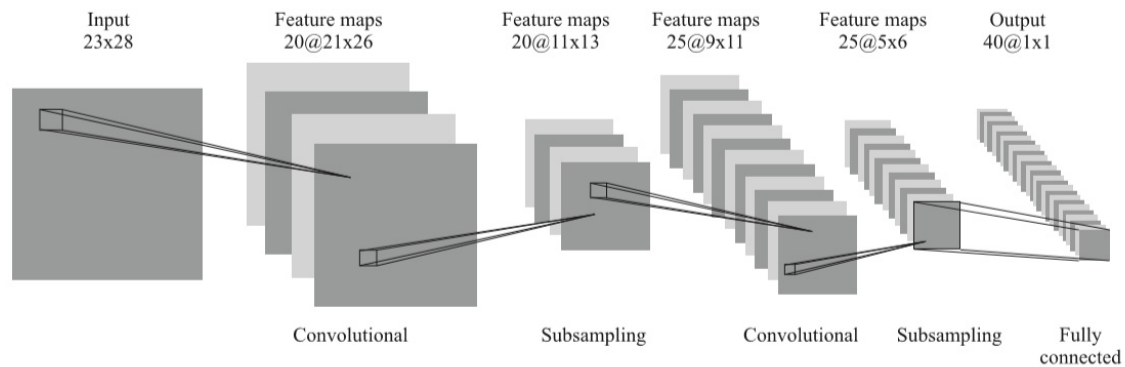


Fig. 4 A typical convolutional network [6]

4.1. Local receptive fields

As with conventional neural networks, neurons in the input layer need to connect with neurons in the hidden layer. However, instead of connecting each input neuron to each hidden neuron, connections are created only in a local area of an image. If every neuron has a full connection, too many parameters are needed, so the method of the local receptive field is adopted. Each neuron only needs to perceive the local information in the image, and then combine the information at a higher level to get the global information. In the diagram, a small part of the area has been reflected in the feature map in the convolutional stage. All the neurons in this local receptive field are connected with a corresponding neuron in the feature map. Each connection has a particular weight parameter, so the amount of weights in a receptive field is equal to the number of neurons in that field. By moving this receptive field up or down, all neurons in the input image will be transferred into a feature map. If the input data has a size of $n \times n$ and the local receptive field has a size of $m \times m$, then the size of the feature map obtained is $(n-m+1) \times (n-m+1)$.

4.2. Shared weights

In the input data, no matter where the local receptive field is, every neuron in the same local receptive field shares the same weight as neurons in other fields. It means that all neurons in the input detect the same feature at different locations in the image each time. This mapping from the input layer to the hidden layer is also called a feature map. In order to do image recognition, more than one feature map is usually needed; therefore, a complete convolutional layer usually contains several different feature maps. As can be seen from the figure, the first convolutional layer contains four feature maps. It can be seen that these feature maps reflect some unique spatial structures of the original image, so it learned some information related to spatial structures for identification.

4.3. Pooling

The pooling (subsampling) layer is after the convolution layer, aiming to simplify the output from the previous feature maps. For example, each neuron in the pooling layer might be the summation of neurons in a 2×2 region in the convolution layer. One commonly used pooling method is to obtain the square root of the sum of squares of all neurons in the 2×2 region of the convolutional layer as the output. Although different methods have different mathematical algorithms, they all aim to simplify the output information from the convolution layer.

5. Application of Deep learning in Image Recognition

The most critical development of deep learning in image recognition is the image classification task in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) challenge. The convolutional network is widely used in the classification of ImageNet. In many cases, the data obtained by traditional machine learning methods often have large errors, and the error rate on the test set is relatively high. However, deep learning can effectively solve these problems. Two popular convolutional network models that got good rankings in ILSVRC are AlexNet and VGG.

5.1. AlexNet

The most significant breakthrough of deep learning in the ILSVRC challenge occurred in 2012, when Hinton's research team used AlexNet, a new variant of convolution network, to win the champion of the ImageNet image classification tasks. Hinton's team used AlexNet to reduce the error rate to 15.315 percent, far below the second-place error rate. [7] Below is the architecture of AlexNet.

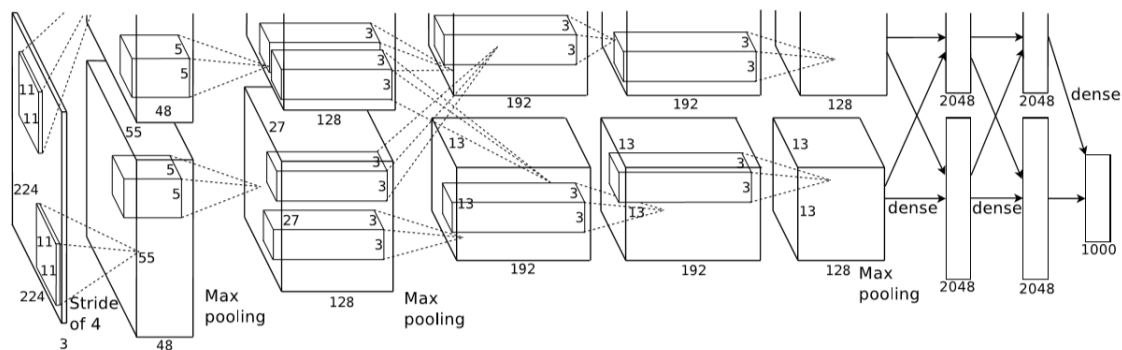


Fig.5 An illustration of the architecture of AlexNet [7]

Compared with the traditional convolutional network, AlexNet has three differences: First, AlexNet adopts the training strategy of Dropout, during which some neurons in the input layer and the middle layer are randomly set to zero which aims to simulate the situation of noise interferes from input data. Dropout makes the training speed faster and make the network model more flexible. Secondly, AlexNet adopts Rectified Linear Unit (ReLU) as the activation function. It helps significantly reduce the complexity of computation. Thirdly, AlexNet produces more training samples by mirroring training samples and therefore, reduces overfitting problems.

5.2. VGG

One improvement of VGG16 over AlexNet is to replace the larger convolution kernel in AlexNet (11x11, 7x7, 5x5) with successive 3x3 convolution kernel. VGG uses convolutional layers with the smaller kernel (3x3) instead of one with a larger kernel. On the one hand, it can reduce the number of parameters; on the other hand, it provides more nonlinear mapping, which helps reduce the overall computation workload.

A small convolution kernel is a vital characteristic of VGG. Although VGG imitates the network structure of AlexNet, it does not adopt the original convolution kernel size in AlexNet. Instead, it reduces the size of the convolution kernel and increases the number of convolution layers to achieve the same performance.

6. Conclusion

Deep learning mainly refers to learning the internal rules and representation level of samples, and its ultimate goal is to enable machines to have the ability of analysis and learning like humans, and to recognize data such as image and language accurately. Meanwhile, deep learning is also a relatively complex machine learning algorithm. Therefore, this paper discusses the application of deep learning in image recognition, mainly analyzes the convolutional network which is often used in image

recognition, with the ultimate purpose of promoting the rapid development of the field of image recognition.

Acknowledgement

This research work was based on the support of 2020 Summer ATEE-Plan Online Research Program and the guidance of Prof. Yiyun Shi.

References

- [1] Yann LeCun, Yoshua Bengio, Geoffrey Hinton, Deep learning, Review Insight.
- [2] Vikramaditya Jakkula, Tutorial on Support Vector Machine (SVM), School of EECS, Washington State University, Pullman 99164.
- [3] William S Noble, What is a support vector machine? 2006 Nature Publishing Group
- [4] Kim Esbensen, Paul Geladi, Principal Component Analysis, Research Group for Chemometrics, Institute of Chemistry, Umei University, S 901 87 Umed (Sweden)
- [5] Muhammad Imran Razzak, Saeeda Naz and Ahmad Zaib, Deep Learning for Medical Image Processing: Overview, Challenges and Future
- [6] Steve Lawrence, C. Lee Giles, Ah Chung Tsoi, Andrew D. Back, Face Recognition: A Convolutional Neural Network Approach,
- [7] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks