Studies on the Application of Machine Learning and Deep Learning in Image Recognition

Zongjian Wu

The University of Queensland, Queensland, Australia

2482516799@qq.com

Abstract. In recent years, with the rapid development of the Internet and mobile technology, the image data in the network has shown explosive growth. It has been widely used in social security, military security, information security, identity authentication, and traffic supervision. Image data is simple and intuitive, contains rich information, and is widely used as a carrier of information exchange. Especially in recent years, the application of machine learning algorithms in the field of image recognition has provided new technical means for image classification, recognition, and feature extraction, making it a research hotspot in this field. Image recognition has received more and more attention due to its broad application value in agriculture, commerce, military affairs, and daily life. The technology serves as the basis of machine vision. For computers, there is no direct connection between the underlying features of an image and the high-level image semantics, so solving the "semantic gap" is the primary focus and difficulty of image recognition. The development of machine learning has gone through two stages shallow learning and deep learning. Experts and scholars have proposed many algorithm models and achieved many achievements in image recognition, speech recognition, and artificial intelligence. In recent years, with the successful promotion and application of deep learning technology in various application fields, image target recognition technology based on deep learning has gradually become the focus of many researchers. Different algorithms have also achieved varying degrees of research progress in image recognition. This paper analyzes and introduces the working principles of two typical machine learning algorithms (Support Vector Machine and Deep Belief Network). The conclusion of this paper confirms the effectiveness and vast application space of virtual samples in machine learning algorithms. Also, it provides powerful help for solving minor sample problems in machine learning.

Keywords: Machine Learning; Deep Learning; Support Vector Machine; Deep Belief Network; Image Recognition

1 Introduction

The development of human beings and the advancement of technology has brought unprecedented convenience to life. In this era full of opportunities and challenges, the

C. Chen et al. (eds.), *Proceedings of the 3rd International Conference on Digital Economy and Computer Application (DECA 2023)*, Atlantis Highlights in Computer Sciences 17,
https://doi.org/10.2991/978-94-6463-304-7_12

social progress brought about by technological change has amazed everyone, and it is a hymn to the current prosperity. The Internet and mobile Internet technologies allow people to access any desired network at any time, place, and in any way, breaking through the limitations of time and space and allowing people to quickly and effectively obtain the information they need. With the rapid development of science and technology, information technology significantly impacts all aspects of human life. Among them, the most intuitive information carrier represented by sound, text, and image carries a large amount of information. Compared with sound and text, image Information has the characteristics of a large amount of information and strong comprehension, so the image processing technology represented by image recognition shows a broader application prospect and research value.

Image recognition, as the most typical representative technology of image understanding, aims to enable the computer to extract the feature information in the image through learning, use it, and finally achieve the purpose of target classification and recognition (Zhang et al., 2020). The machine learning algorithm first needs to perform a series of preprocessing operations on the acquired data and then extract and select practical image features on the preprocessed data according to the artificially prescribed form based on expert opinions. Eventually, complex function processing completes the modeling process for image recognition. Feature representation is summarized as data acquisition, preprocessing, and feature extraction. The quality of feature representation directly impacts the model accuracy of machine learning algorithms and the performance of structured models.

With the rapid development of computer information technology, image recognition plays a vital role in various fields of social life. Image recognition has been deeply studied and developed rapidly because of its extensive and essential application value. Image recognition technology is mature and widely used to recognize faces, numbers, and other objects. In the field of public safety, public transportation safety, license plate location, target tracking, and moving object detection are also closely related to image recognition (Zheng et al., 2019). Face recognition is a relatively mature and precious image recognition technology. In addition, the application principles of the retina, fingerprint scanning, and other access control systems are similar (Liu et al., 2018). The hospital's clinical medical instruments also use image recognition to judge and analyze the condition (Zhang). All the above applications have use value and research significance.

However, after decades of machine learning development, many problems still need to be resolved, such as image recognition and positioning detection, image classification, graphic generation, speech recognition, and natural language processing (Bai, 2014). In deep learning research, research on image classification has always been the most basic, traditional, and urgent research direction.

2 Literature Review

As early as the 1950s, the concept of machine learning appeared along with artificial intelligence (Kim & Hong, 2009). It is envisaged that computers will be intelligentized and have the ability to autonomous learning and reasoning like humans, often referred to as artificial intelligence. However, due to the limitation of the environment and technical level at that time, the learning ability of computers was low, and machine learning and artificial intelligence gradually faded out of the vision of scientists. In the 1980s, the artificial neural network algorithm represented by the back-propagation algorithm was born, which overcomes the limitation of traditional algorithms that need to manually formulate rules and only need to input a large number of training samples to the machine. The machine can learn from the samples. After finding statistical laws, compared with traditional algorithms, neural network algorithms have advantages in algorithm complexity and target recognition accuracy (He et al., 2003)[1].

In an article on Deep Belief Networks (DBN) published in the journal Science by Professor Geoff Hinton of the University of Toronto in 2006, the D is a stack of many restricted Boltzmann machines (RBMs) trained by an unsupervised layer-by-layer greedy algorithm. This method solves the problems of local optimum and gradient disappearance and achieves a high recognition rate on the MN1ST handwriting database; the result is 98.8% (Bengio et al., 2007). Inspired by the DBN method, Benigo et al. replaced each layer of RBM in DBN with Auto-encoder in 2007 and proposed a Stacked Auto-encoder, the deep structure of SAE. In 2010, the U.S. Department of Defense jointly sponsored many universities to conduct deep learning and research on deep learning and achieved good results (Hinton et al., 2012). In 2011, researchers from Microsoft and Google applied deep neural networks to speech recognition, improved the accuracy of speech by 20 percent, and achieved breakthrough results, demonstrating the extraordinary performance of deep neural networks in speech recognition[2].

Compared with the foreign technical level, domestic research in China on virtual image samples started late, but in recent years, many researchers have begun to research virtual samples. In 2002, Wen Jinwei et al. used the linear category weighting of face images to construct virtual samples, expanded the number of samples, and verified the effect of virtual samples in the Bayesian network classifier after extracting features through PCA. At Yale University, the addition of virtual samples in the face database has made some improvements. In 2003, Zhu Changren and Wang Runsheng proposed to use feature points to represent the face, fit the feature points of the image through a high-order function, and use the least square method to solve the virtual samples generated by the image after different pose transformations. In the ORL face database, the virtual sample rate is increased from 52% to 88% by adding ten forward faces and 20 other poses. In 2006, Zhang Shengliang et al. introduced many geometric transformations to the image, such as rotation, scaling, and mirror transformation, to increase the number of virtual samples. By using two-dimensional principal component analysis (2DPCA) to extract features, Zhang added seven virtual samples and 11 virtual samples, experimented on the ORL face database, and got a better recognition rate improvement. Zhang Jianming used singular value decomposition to perform perturbation reconstruction and, at the same time, used traditional geometric transformation methods to construct virtual samples[4]. Experiments on the ORL face database show that adding virtual samples under single-sample conditions can increase the recognition rate by 0.61% on average[3].

However, deep learning still has bottlenecks in practical applications, the most prominent of which is poor interpretability. Since we cannot grasp the actual working mechanism of the human brain, artificial neural network models, including deep learning, are An abstract imitation of the workings of the human brain. China's "Brain Network Dome" project aims to uncover the secrets of human brain thinking, study its inner working mechanism and principles, and guide the development of artificial neural networks and the further optimization and development of deep learning. The "BRAIN" project in the United States is a necessary research set to study the operating mechanism of the human brain. It is an important scientific research project for mapping the complex circuits of the human brain. The U.S. Department of Defense hopes that the "BRAIN" project can simulate the working process and work of the human brain as highly as possible. Mechanisms to generate new information processing architectures or new computing methods.

3 The Development of Machine Learning

With the ultra-high-speed development of computer and Internet technology, the revolution brought by machine learning has brought a new dawn to the development of artificial intelligence. Machine learning has shown different styles in different historical periods in the long development process: shallow learning and deep learning. The revolution in machine learning comes from the research and development of biological neural networks. People have established relevant scientific models by simulating biological neural networks through computer neural networks and made breakthroughs in image recognition, speech recognition, and artificial intelligence. The progress has set off a wave of upsurge.

The shallow learning stage of machine learning can be traced back to 1986. At that time, the well-known neural network expert Hinton et al. published an article in "Nature" magazine on the method of the backpropagation algorithm. Training the neural network has strengthened the neural network. Experts and scholars have started to train the neural network through the backpropagation algorithm and achieved great success. Experts and scholars have trained many three-layer human neural networks through samples, statistics, and algorithms, and this period has become the shallow learning period of machine learning. With the continuous development of machine learning, experts and scholars have proposed many machine learning models. These models are used in different fields of machine learning according to different characteristics to solve different problems. Commonly used models include Gaussian models, logistic regression, and support vector machines, among which support vector machines are developing rapidly. Support vector machine training is simple and convenient. After much research and practice, it is widely used in speech and picture recognition[6].

The deep learning stage of machine learning can be traced back to 2006. At that time, Hinton et al. published a paper on deep learning and the Boltzmann machine, which proposes multiple hidden layers to strengthen the artificial neural network. Hinton puts forward two viewpoints in the article that contribute significantly to machine learning. First, more depth can enable the machine model to learn more features and better describe the essence of things, which can improve recognition accuracy. Accuracy; Second, multiple hidden layers can transmit information, each hidden layer can be trained independently, and each layer uses independent learning, which solves the embarrassing situation of manual training[5]. The deep machine learning model is established by simulating the human brain. When humans process external information, they go through multiple levels of abstraction and transmission. Deep learning also established a multi-level processing layer. High-level features can better represent things through the transfer and abstraction from low-level to high-level and the conversion from low-level features to high-level features. The deep learning model overcomes the difficulty of manually selecting features. In image processing, many complex features, such as HOG, LBP, and SIFT, can be automatically learned by the deep learning model through the underlying features, reflecting the advantages of deep learning models.

4 Applications in Image Recognition

Image recognition generally consists of image acquisition, preprocessing, feature extraction, dimensionality reduction, and classification recognition. The specific process is shown in Figure 1:

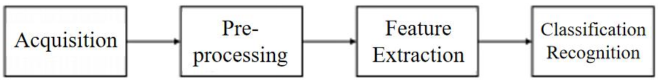

Fig. 1. Standard Process of Image Recognition

Image acquisition collects original images through electronic imaging equipment as a sample source for training and testing in the subsequent recognition process. However, in practice, the number of images collected is limited, the sample quality is uneven, the number of available samples is small, and it is difficult to obtain the required target.

Image preprocessing refers to the process of obtaining high-quality and sufficient image data that can be used for image recognition algorithms from the original image data acquired by the sensor. The original data obtained by image acquisition often contains a lot of noise and a large number of areas that are outside of classification and recognition, which significantly impacts the efficiency of image recognition and then restricts the performance of the entire image recognition system. Image preprocessing is to unify the size and format of the image, that is, to standardize it according to a particular method, and to perform related denoising processing on the image to reduce the interference information of the image. Therefore, it is necessary to complete the image data preprocessing process before image feature extraction.

Image preprocessing techniques often include adjusting brightness, contrast, and histogram equalization.

Feature extraction and dimensionality reduction refer to extracting feature information that can represent the primary information of the image that contributes to the target classification from the input data and kick out irrelevant, redundant information to realize feature extraction and dimensionality reduction of the data. Image features can be divided into global and local features according to different ranges[7]. The global features have global characteristics, which can well reflect the whole of the image, and the local features have local characteristics, which can highlight the critical points of the image. Image recognition focuses on the association between low-level features and high-level semantics[10]. The underlying features are features composed of pixels, which can be extracted using correlation methods. Commonly used underlying features are color, shape, and texture. The underlying features have good stability, so they are the best choice for image recognition. High-level semantic features are relatively abstract and need to be extracted from the underlying features and mapped to high-level semantic features. With the development of mathematical techniques, image processing techniques, morphology, wavelets, genetic algorithms, support vector machines, neural networks, and deep learning, images Feature extraction algorithms are getting more and more powerful[8].

Classification recognition refers to the data sample features obtained by the image recognition system based on the feature extraction and dimensionality reduction steps. According to the principle that different image samples have different feature distributions, a specific classification criterion is used to assign a category label to the input image sample. Complete The classification recognition step of the image recognition system. The classifier in the system will be based on specific criteria through the previous training on the feature vector of the target object. It will give a predicted classification label with the measured target object. The difficulty of classification depends on two points: first, the fluctuation of feature divisions of different samples of the same category; second, the difference in feature distribution of samples of different categories. Therefore, an appropriate classifier should be selected according to the target characteristics to match and classify the image target to be recognized (i.e., the test sample) and the trained target image (i.e., the training sample), and finally, output the recognition and classification results. The difficulty of the classification and recognition step is that there are differences in feature distribution between different types of image samples and differences in feature distribution between the same samples. To correctly complete the classification and recognition of image samples, it is necessary to increase the dispersion between samples as much as possible. Moreover, reduce the dispersion within the sample to complete the high-quality classification and recognition of image samples[9].

5 Support Vector Machines

In traditional image recognition algorithms, the index to measure the performance of the algorithm is generally the image recognition accuracy rate, which can better reflect the recognition performance of the image recognition system under the premise of sufficient image data. However, in practical application scenarios, obtaining a large number of sufficient sample sets in image acquisition is often challenging. In the case of low accuracy, overfitting is easy to occur. The machine learning model can only achieve a high recognition rate on the training sample set but cannot achieve a recognition rate comparable to that on the training set on the test set. To overcome the above problems, Corinna Cortes and Vapnik et al. proposed the Support Vector Machine (SVM) in 1995. The SVM model is based on statistical theory. It has advantages in the establishment of classification models. It can minimize structural and empirical risks and solve the problem of underfitting and overfitting in small sample spaces. It is essential for machine learning, and development plays a significant role in promoting it. SVM models have a similar functional form to neural networks and radial basis functions. However, SVM's ability to generalize and train on data far exceeds these traditional methods. SVM can be well-analyzed for complex, real-world problems such as text and image classification, handwriting recognition, and bioinformatics. Even with a few models to train, SVMs perform well on many different kinds of datasets. SVM seeks an optimal classification plane to separate data of different categories to the greatest extent (fig2).

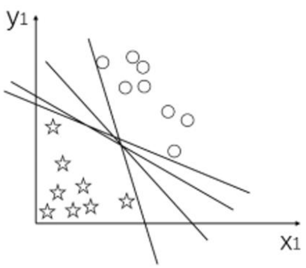

Fig. 2. Demonstration of SVM Separating Hyperplane

The optimal classification line sought needs to separate the two types of samples and keep the classification interval to the maximum. Usually, when dealing with complex data, the optimal classification hyperplane that separates the data can be found by mapping the data into a high-dimensional space. At the same time, the optimal plane sought by SVM is the global optimum, which has a significant advantage over the traditional neural network prone to local extremum problems. SVM needs to extract features artificially in image recognition, extract relevant feature vectors with the help of color, shape, and texture extraction methods, and then convert nonlinear feature vectors into linearly separable feature vectors through kernel functions to solve. The optimal classification surface can improve image recognition's accuracy and generalization effect.

Machine learning algorithms, including SVM and principal component analysis algorithms, have made ideal breakthroughs in image recognition and have been widely used in e-commerce, security systems, and automated driving technology. Compared with traditional statistics-based image recognition algorithms, machine learning algorithms can automatically classify or cluster the extracted feature vectors according to their distribution in the feature space to achieve the purpose of image recognition. Its most crucial identification process is the training of the classifier. The parameter setting of the classifier is not an artificial setting method. However, according to the different distributions of the input data, the model learns and sets better model parameters independently to achieve better performance. Therefore, the degree of human participation in the image recognition process is reduced to a certain extent, and the work efficiency and image recognition accuracy of the image recognition model are improved. However, on the one hand, machine learning algorithms rely on the manual selection of features. Machine learning algorithms must first complete the manual feature selection and determination, then the machine can complete the feature extraction and classification. Excessive manual participation affects not only the computational efficiency of the model but also the model.

On the other hand, the machine learning algorithm also depends on the selection of the classification model. Different image data types are suitable for classification models, such as linear regression, Logistic regression, and Softmax regression. The same data Different classification recognition performance can be obtained using different classification models. In summary, machine learning algorithms rely on manual feature selection and model selection, which limits the development of image recognition technology. Therefore, exploring a deeper network model is necessary, aiming to break through the limitations of machine learning and achieve better recognition performance.

6 Artificial Neutral Network

The artificial neural network is a computer neural network that simulates the biological neural network. It can extract the features in the image very well and then pass through the layers of the hidden layer to continuously abstract the higher-level features. It has a massive advantage in image recognition for its simple structure, usually connected by an input layer, one or more hidden layers, and an output layer. For the input layer, the processed data is used as the input data, and the number of neurons is the same as the input dimension. The relevant output is finally obtained for the output layer, which can be classified and judged, and then the image is recognized. The number of output neurons is the same as the output category.

Artificial neural network adopts the idea of bionics, which imitates the behavior of the advanced intelligent biological neural network, so it has many remarkable characteristics. The first is high parallelism. During the working process of the artificial neural network, each neuron node operates in parallel. This feature gives the artificial neural network model a more prominent advantage in processing high-dimensional and complex data. In the network, multiple neuron nodes interact with each other. Through the parameter interaction between neurons, the memory function of the processed information can be realized, and functions such as feature extraction, image recognition, and sequence prediction can be realized. Lastly, the artificial neural network has a profound self-learning and adaptive ability. It can continuously fit and learn the data distribution of the sample through continuous training under the condition of large samples.

7 Deep Belief Network

Hinton first proposed deep learning in an article published in Science magazine in 2006. In their article, Corinna Cortes and Vapnik et al. studied a layer-by-layer superposition method of Restricted Boltzmann Machines to form a neural network. In 1985, Ackley proposed the Boltzmann machine. The Boltzmann machine was built based on statistical mechanics. The Boltzmann machine was established according to the relevant model of statistical mechanics. The Boltzmann machine layer, connections between layers, and elements of a single layer are also connected, so the probability distribution is difficult to calculate. The Boltzmann machine is a distributed feature representation method with strong feature expression ability. At the same time, the Boltzmann machine is self-learning, which can automatically learn from low-level features and convert them to more abstract high-level It has strong abstract learning ability. Such a deep network has better feature learning ability, can effectively express data features, and can obtain ideal recognition results. In the article, Hinton used a method called "layer-by-layer initialization" to effectively solve the problems of deep networks in the past. Firstly, unsupervised learning is used to pre-train each deep neural network layer by layer. The results of the previous layer will be used as the input of the next layer during layer-by-layer training. Then, the specific parameters in the deep network are iteratively updated by employing supervised learning. Because after the network is pre-trained through unsupervised learning, the initial parameters in it will be set according to the input data instead of randomly selected in the past, so the results obtained are ideal. At the same time, the influence of gradient dispersion is essentially eliminated (fig3).

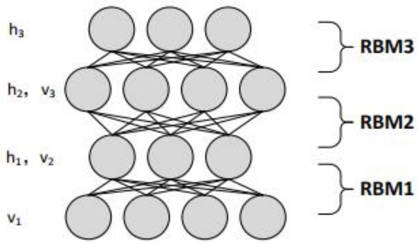

Fig. 3. Demonstration of a DBN Structural Model

As shown in Figure 3, DBN is formed by stacking multiple layers of RBM with an input layer, an output layer, and several hidden layers. Since the DBN model combines unsupervised and supervised methods, the network structure constructed by pre-training learning unlabeled training data will be easier to find the global optimum. Compared with the random distribution of weights in traditional neural networks, The classification recognition ability of DBN will be even better. DBN is a probabilistic generative model that has successfully solved complex problems such as computer vision, speech recognition, and natural language processing. Although DBN has many advantages in recognition, some things still need to be solved. Because the activation function in the traditional DBN network model uses the sigmoid function, the sigmoid function cannot guarantee the uniqueness of the network solution. In addition, the sigmoid function needs exponential magnitude to be calculated, its convergence speed is slow, and it is easy to saturate and causes the gradient to disappear. When the input is very large or very small, the neuron gradient is close to 0, which makes the parameters of the first few layers of DBN in the backpropagation algorithm. It needs to be effectively updated, which leads to the inability to recursively learn the input data, which is why the early DBN cannot be deepened.

8 Conclusion

With the rapid development of the Internet and mobile Internet, a large amount of picture data appears in the network. Massive picture data contains rich information, and pictures have the characteristics of simple form and intuitive content and are widely used as carriers of information exchange. With the emergence of mobile terminals such as smartphones, people can use the camera on the device to obtain pictures at any time and then use the massive pictures on the Internet to search for the information they want, realizing the effect of searching for pictures by picture. In order to realize such a wish, image recognition is significant as a fundamental work. Image recognition needs to extract features from the picture and then identify the picture through the mapping relationship between the underlying features and high-level semantics. However, there is a gap between the underlying features and high-level image semantics, which requires machine learning to achieve this goal.

This paper introduces the basic concepts of deep learning, and the specific workflow of image recognition, reviews the difficulties and difficulties of image target recognition technology based on deep learning and the current development status at home and abroad, and summarizes the principles of traditional machine learning. Deep-learning image recognition algorithms and algorithm derivation steps pointed out the deficiencies and limitations of traditional algorithms in image recognition tasks. Based on fully summarizing the research results in the field of image recognition technology based on deep learning, the research on the cause analysis, targeted improvement, and improvement effect test is carried out for the problems that often occur in practical applications.

This paper systematically introduces the basic principles and learning and training to the engineer of two commonly used machine learning algorithms, the SVM, and the DBN. It illustrates the mechanism and function of common feature extraction and dimensionality reduction methods through experiments. The basic theory of virtual samples is briefly introduced, and the evaluation criteria and corresponding evaluation basis for the validity of virtual samples are studied from the perspective of characteristics.

With the development of machine learning technology, the machine learning system used for identification and classification will inevitably show a more advanced development trend, which will put higher requirements for designing a more effective virtual sample algorithm. Image recognition is a widely used interdisciplinary topic. Applying the deep learning mechanism to the field of image target recognition can improve the recognition rate of image target recognition to a certain extent. However, in actual use scenarios, it will still be affected by many objective factors. This impact restricts the application of deep learning algorithms in image target recognition tasks in natural environments. The in-depth combination and breakthroughs of image recognition technology and deep learning methods are bound to appear in application scenarios closer to life. Due to time constraints and the lack of comprehensive knowledge involved in the subject, the virtual sample algorithm designed in this paper still needs to be improved and can be further studied.

References

1. Zhang S, Zhang R, Zhang Y. "A review of medical image recognition based on deep learning," J. China Health Statistics, vol.1, 2020.
2. Zheng Y, Li G, Li Y. "A Review of Deep Learning Application in Image Recognition," J. Computer Engineering and Application, vol.55, pp.20-36, 2019.
3. Liu J, Sun R, Qiao S. "Application Research of Deep Learning in Medical Image Recognition," J. Journal of Qingdao University (Natural Science Edition), vol.1, pp.69–74, 2018.
4. Zhang Q. Research and application of deep learning in medical image recognition, D.
5. Bai Y. Research and Improvement of Image Recognition System Based on Deep Neural Network, D. [Master Thesis]. Harbin: Harbin Institute of Technology, 2014.
6. Kim J. S., Hong K. S.. "Color texture segmentation using unsupervised graph cut," J, Pattern Recognition, vol.42, pp.735-750, 2009.
7. He X, King O, Mei W, et al. "Learning a Semantic Space from User's Relevance Feedback for Image Retrieval," J. IEEE Trans. on Circuit and Systems for Video Technology, vol.13, pp.39–48, 2003.
8. Bengio Y, Lamblin P, Popovici D, et al. "Greedy layer-wise training of deep networks," J. Advances In Neural Information Processing Systems, vol.19, p.153, 2007.
9. Hinton A et al. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580, 2012.
10. Sun Z, Xue J et al. "A review of deep learning research," J. Computer Application Research, vol.29, pp.2806-2810, 2018.