

Research on Image Recognition Based on Deep Learning Technology

Hao Zhai

School of Computer and Communication Engineering, Tianjin University of Technology, Tianjin
300384, China;

824982245@qq.com

Keywords: Image Recognition; Deep Learning;

Abstract. Nowadays image recognition technology is widely used, and plays a very important in various fields. Deep learning technology uses multilayer structure to analyze and deal with image features, which can improve the performance of image recognition. The popular models of deep learning contain AutoEncoder, Restricted Boltzmann Machine (RBM), Deep Belief Network (DBN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and other improved methods. The applications of image recognition based on deep learning technology including image classification, facial recognition, image search, object detection, pedestrian detection, video analysis. We believe that in the future deep learning will develop rapidly in theory, algorithm, and application and they will make our lives more intelligent.

1. Introduction

In 2016, in the deep learning field, the most sensational event is that in March AlphaGo defeats Li Shishi, who is a professional nine player and the world chess champion, with the total score 4:1. This is AlphaGo's second victory. In October 2015, it became the first Computer Go program to defeat the European Go champion Fan Hui, a 2-dan (out of 9 dan possible) professional, five to zero. The AlphaGo system was developed by British computer company DeepMind which was bought by Google in 2014. AlphaGo's algorithm uses a Monte Carlo tree search to find its moves based on knowledge previously "learned" by deep learning, both from human and computer play [2]. Deep learning simulates biological neural brain to process precise and complex calculating, like people recognizing objects.

Image recognition is an important application and requirement in life and industry. With the need of social practice, there are more and more kinds of things to be classified and identified, and the content of the object is more and more complex. For example, in the traffic management system, the car license automatic recognition based on image recognition technology can record the illegal behaviors of vehicles. So image recognition technology plays a very important role in the fields of industry and our daily life. Therefore, it is of great significance to improve the performance of image recognition technology by using advanced deep learning algorithms.

2. Definition of Deep Learning

Deep learning is a new field in the research of machine learning, which is a kind of unsupervised learning. The motivation of deep learning is to simulate the human brain to establish the neural network, which can interpret the data, such as images, sound and text, like human brain. In the neural network, attribute categories or features are represented by combining low-level features to form more abstract high-level features, in order to discover the distributed feature representation of data. One of the promises of deep learning is replacing handcrafted features with efficient algorithms for unsupervised or semi-supervised feature learning and hierarchical feature extraction [3].

In 2006, Hinton proposed the concept of Deep Learning and Restricted Boltzmann Machine (RBM) model, which used artificial neural networks with multiple hidden layers to improve the performance of visualization and classification [4]. After that, Hinton presented Deep Belief Network (DBN) model based on RBM [5]. On the other hand, deep Convolutional Neural Network

(CNN) was used in image recognition and achieved good performance [6]. Then in 2012, Hinton team won the ImageNet Challenge competition, which is about image recognition. After that, deep learning methods attract a lot of people to study and use. Now these methods are widely used in image and video recognition, recommender systems and natural language processing.

Compared with traditional machine learning methods, deep learning methods are particularly suitable for dealing with large data. The deep learning methods can reduce model bias to improve statistical estimation accuracy with more complex models. In addition, deep learning is almost the only end to end machine learning system, which abandons the intermediate step of artificial rules and applies the prior knowledge of the data structure to the new model structure. These advantages make deep learning methods very suitable for image recognition.

3. Technology of Deep Learning

Deep learning is a branch of machine learning based on a set of algorithms that attempts to model high-level abstractions in data by using a deep graph with multiple processing layers, composed of multiple linear and non-linear transformations [7] [8][9].

Deep learning algorithms are based on distributed representations. The underlying assumption behind distributed representations is that observed data are generated by the interactions of factors organized in layers. Deep learning adds the assumption that these layers of factors correspond to levels of abstraction or composition. Varying numbers of layers and layer sizes can be used to provide different amounts of abstraction. These architectures are often constructed with a greedy layer-by-layer method. Deep learning helps to disentangle these abstractions and pick out which features are useful for learning [10].

There is huge number of variants of deep learning models, such as AutoEncoder, RBM, DBN, CNN, RNN and other improved methods. In this part, we briefly introduce these models.

(1) AutoEncoder

AutoEncoder method is the simplest method of deep learning methods to learn a representation (encoding) for a set of data, typically for the purpose of dimensionality reduction [10]. AutoEncoder assumes that the output data is the same with the input data, then the parameters of AutoEncoder are adjusted in the training process to get the weight of each layer. The target of Autocoder is to reappear the input data. So AutoEncoder must capture the most important factors, which can the most important factors, like PCA. Recently, Autoencoder has become more widely used for learning generative models of data.

(2) Restricted Boltzmann Machine

A Restricted Boltzmann Machine (RBM) is a generative stochastic artificial neural network that can learn a probability distribution over its set of inputs. In the model of RBM, there is a bipartite graph, and there is no link between the nodes at each layer. One layer is a visible layer, namely the input data layer. The other layer is the hidden layer. The values of all nodes are random binary values (only 0 or 1). Meanwhile, the total probability distribution satisfies Boltzmann distribution [4].

If the number of the hidden layer increases, RBM transforms to DBM [11]. On the other hand, if a Bayesian Belief Network is used near the visual layer, and RBM is used in the most remote part of the visual layer, RBM transforms to DBN [12].

RBM has found applications in dimensionality reduction, classification, collaborative filtering, feature learning and topic modelling. They can be trained in either supervised or unsupervised ways, depending on the task.

(3) Deep Belief Networks

DBN is a probabilistic generative model in order to establish a joint distribution model between the observation data and the tags. DBN is composed of a plurality of RBN layers [8]. DBN is very flexible, which makes it easier to expand. One extension is the convolution of DBN (CDBN) [13]. At present, and the related research of DBN includes stacking automatic encoder, which replaces the traditional RBM inside the DBN.

(4) Convolutional Neural Networks

CNN is the first learning model to really succeed in training the multilayer network structure. It uses spatial relation to reduce the number of parameters in order to improve the training performance of the general forward BP algorithm. CNN, as a depth learning architecture, is proposed to minimize the process of data dealing [14].

CNN is more effective in image recognition. In CNN, a small part of the image is treated the input data of the lowest layer in the hierarchical structure. Then information of the image is transmitted to different layers in sequence. Each layer can obtain the most significant characteristics of the observed data by a digital filter. Even if the images are translated, scaled and rotated, CNN can deal with the observation data of significant features. Therefore, CNN is an important and effective method in image recognition [6].

(5) Recurrent Neural Network

The purpose of Recurrent Neural Network (RNN) is used to process the sequence data. In the traditional neural network model, the data is from the input layer to the hidden layer, then to the output layer, where the layers are fully connected, and the nodes between each layer are not connected. But this common neural network is incapable of many problems. However, in RNN, the data sequence of the current output is related to the front output. In the process of RNN, the network can remember the front information and apply them to the current output of the computation. This means that the nodes in the hidden layer are connected. On the other hand, the input data of the hidden layer not only includes the output data of the input layer, but also include the output data of the last hidden layer [15].

RNN can be applied to many tasks, such as generating image descriptions, unsegmented connected handwriting recognition, machine translation and speech recognition.

4. Application of Image Recognition Based on Deep Learning Technology

Deep learning attracts a lot of people to research and apply. World famous universities have found Research Center, such as Toronto, and MIT Technology Review. And science and technology companies also set up special research groups, such as Google's Google brain project, Baidu depth research institute. Deep learning is widely used in image classification, facial recognition, image search, object detection, pedestrian detection, video analysis.

(1) Image Classification

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is a benchmark in object category classification, and detection on hundreds of object categories and millions of images [16]. The challenge has been run annually from 2010 to present, attracting participation from more than fifty institutions [17]. The goal of ILSVRC is to estimate the content of photographs for the purpose of retrieval and automatic annotation. Over the past 5 years, ILSVRC has made many achievements and promoted the development of image classification. There are many potential directions of improvement and growth for ILSVRC and other large-scale image datasets.

(2) Facial Recognition

Face recognition is another important challenge in the field of computer vision and deep learning. In academia, Labeled Faces in the Wild (LFW) is one of the most famous face recognition test sets, founded in 2007. There are 6000 pairs of face images in the testing set, containing 3000 positive samples and 3000 negative samples. At present, recognition rate of this testing set can reach 99.47% [18].

In industry, as a more efficient way of authentication and identification, face recognition technology has gradually entered the stage of practical application. Face recognition technology can be used in government, military, banking, social security, electronic commerce, security defense and other fields.

(3) Image Search

In the 6 months after Hinton's team won the ImageNet competition, Google and Baidu made a

new search engine based on image content. They follow the Hinton in the ImageNet contest with the depth of the learning model, apply to their data, and find that the accuracy of image search has been greatly improved. At present, a number of web sites launch image search service, such as Google, Baidu. They search similar images based on the content of the input image. The research results can be close to the user's need.

(4) Object Detection

Deep learning also brings a great improvement in the objects detection, which is more difficult than object recognition. An image may contain multiple objects belonging to different categories, and object detection needs to determine the location and category of each object. In 2013, ImageNet ILSVRC organizers added the task of object detection, which was to need to detect 200 kinds of objects, about forty thousand internet pictures. At that time, the accuracy was 22.581%, then in 2014 the accuracy was improved to 43.933%, with the deep learning methods. In recent years, more influential works include RCNN, Overfeat, GoogLeNet, deepid-Net, VGG, and spatial pyramid pooling in deep CNN.

(5) Video Analysis

How to analyze video with deep learning method? The most direct approach is to make the video as a three-dimensional image, and CNN is applied to analyze video data. In the latest research work, the length of memory network (LSTM) is being widely concerned, which can capture the long-term dependence on the video of the complex dynamic modeling [19].

With the development of deep learning, Google, Facebook and other enterprises have made a significant breakthrough in the field of graphics recognition. Meanwhile, a start-up company called Clarifai provides a new service, which understands the video content with deep learning technology [20].

5. Future of Deep Learning technology

The development of image recognition is in the ascendant, and there is a huge space in the future. In object recognition and object detection, image recognition tends to use more and deeper network structure. Meanwhile, the size of the training data is also rapidly increasing, and there is an urgent need to study new algorithms and develop new parallel computing systems to train big data.

After inferring the latest studies, we conclude there are several future development trends in image recognition. (1) The layers of deep learning will be deepened, in order to deal with more data. (2) The original models will be improved and their structure will be more complex, in order to get higher recognition accuracy. (3) The recognition will be more precise, after the big training image samples are pre-treated in various ways. Comparing with image recognition, the application of deep learning in video classification is far from mature. To achieve this goal, it not only needs to set up a large-scale training data set, but also needs to study new deep learning model, which is suitable for video analysis.

6. Conclusion

From 2012, deep learning has greatly promoted the research progress of image recognition. The representative results are ImageNet ILSVRC, Google Brain and Baidu Institute of Deep Learning. They pay a lot of attention to research and apply deep learning technology. The essence of deep learning is that automatic learning features from the big data through the multilayer nonlinear transformation, which can replace the feature of the manual design. In one image, there are various implicit factors related together in nonlinear way. But deep learning can separate these factors so that classification becomes simple. We believe that in the future deep learning will develop rapidly in theory, algorithm, and application, and they will make our lives more intelligent.

References

- [1] Artificial intelligence: Google's AlphaGo beats Go master Lee Se-dol. <http://www.bbc.com/news/technology-35785875>. BBC News. Retrieved 17 March 2016.
- [2] Lamb E, Levy A, Geddes L, et al. News Peculiar pattern found in 'random' prime numbers[J]. *Nature*, 2016, 531: 296-286.
- [3] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.
- [4] Hinton G, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets [J]. *Neural computation*, 2006, 18(7): 1527-1554.
- [5] Hinton G E. Deep belief networks[J]. *Scholarpedia*, 2009, 4(5): 5947.
- [6] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//*Advances in neural information processing systems*. 2012: 1097-1105.
- [7] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.
- [8] Bengio Y. Learning deep architectures for AI[J]. *Foundations and trends® in Machine Learning*, 2009, 2(1): 1-127.
- [9] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2013, 35(8): 1798-1828.
- [10] Deng L, Seltzer M L, Yu D, et al. Binary coding of speech spectrograms using a deep auto-encoder[C]//*Interspeech*. 2010: 1692-1695.
- [11] Salakhutdinov R, Hinton G E. Deep Boltzmann Machines[C]//*AISTATS*. 2009, 1: 3.
- [12] Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines[C]//*Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 2010: 807-814.
- [13] Lee H, Grosse R, Ranganath R, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations[C]//*Proceedings of the 26th annual international conference on machine learning. ACM*, 2009: 609-616.
- [14] LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series[J]. *The handbook of brain theory and neural networks*, 1995, 3361(10): 1995.
- [15] Mikolov T, Karafiát M, Burget L, et al. Recurrent neural network based language model[C]//*Interspeech*. 2010, 2: 3.
- [16] ImageNet. <http://www.image-net.org/>
- [17] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge[J]. *International Journal of Computer Vision*, 2015, 115(3): 211-252.
- [18] Sun Y, Wang X, Tang X. Deeply learned face representations are sparse, selective, and robust[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015: 2892-2900.
- [19] Donahue J, Anne Hendricks L, Guadarrama S, et al. Long-term recurrent convolutional networks for visual recognition and description[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015: 2625-2634.
- [20] Clarifai. <https://www.clarifai.com/>