Research on Key Technologies in the Field of Computer Vision Based on Deep Learning

Miaofen Feng, *
1 College of Internet of Things, Jiangxi Teachers College, Yingtan, 335000, China

*Corresponding author e-mail: fengmiaofen@jxsfgz.com

Abstract. At present, introducing deep learning has opened up a new way for the computer vision neighborhood. It can perform feature extraction and analysis by simulating the human brain, which greatly improves the development and application of computer vision technology. This paper mainly analyzes the research status and development trends of deep learning in the three fields of image classification, object detection, and face recognition.

Keywords: Image Classification, Face Recognition, Deep Learning

1. Introduction

Computer vision uses electronic imaging systems to replace biological vision systems, then the program replaces the human brain for processing and interpretation, and finally the computer can understand the environment like a human. The short-term goal is to establish a visual system that can accomplish some simple intelligent tasks based on visual perception. For example, image classification, image matching, target detection, behavior recognition and image semantic segmentation, image question and answer, etc [1].

2. Deep learning

For the definition of deep learning, different people have different descriptions. The more appropriate is Bengo's definition: deep learning is a subfield of machine learning, which is an algorithm that attempts to abstract data through a series of multi-layer nonlinear transformations. The current deep learning is based on the neural network structure, and its deep term is relative to the previous shallow learning. Traditional machine learning methods (such as support vector machine, Boosting, nearest neighbor classifier, etc.) can be simulated with a neural network containing one or two hidden layers, which is a shallow learning structure. The model structure of deep learning usually has 3 or more hidden layers, and can use big data to learn features [2]. There are three main types of existing deep
learning architectures, namely deep convolutional neural network (CNN), deep confidence network (DBN), and stack autoencoder (SAE).

2.1. Deep Convolutional Neural Network (CNN)

Convolutional neural network is not a new algorithm. As early as the 1980s, Fukushima proposed a convolutional neural network calculation model. Later, LeCun et al. improved and trained the convolutional neural network on this basis and applied it to handwritten characters. Recognition tasks have achieved good results. However, limited by the training methods and computer hardware at the time, it was not until Hinton proposed the idea of deep learning that the once silent convolutional neural network was reproduced with a deep structure. Convolutional neural networks are generally composed of convolutional layers and sub-sampling layers. Each layer contains multiple two-dimensional planes, and each plane contains several independent neurons. A multilayer deep convolutional network structure is shown in Figure 1 [3].

![Figure 1. Convolutional neural network.](image)

Generally, the C layer is a convolutional layer, in which each neuron is connected to the local receptive field of the upper layer to extract local features. Neurons in the same layer form a feature map of the input, and different feature maps are obtained by convolving the input with different convolution kernels. The C1 layer in Figure 1 has 4 types of convolution kernels. The S layer is a downsampling layer, which samples the feature plane of the convolutional layer, reduces its spatial resolution, retains the main information, and reduces the weight parameters of the network. The convolutional layer and the down-sampling layer alternately form a multilayer convolutional neural network. Its groundbreaking part is: partial receptive field and weight sharing. The neuron of each feature mapping surface has the same weight, which reduces the complexity of the network model and reduces the weight parameters of the network [4]. This deep network structure is highly invariant to translation, scaling, tilt, or other forms of deformation, so it can extract the salient features of observation data that are invariant to translation, scaling, and rotation. It has been widely used in speech recognition and image recognition. And many other fields.

2.2. Deep Belief Network (DBN)

In 2006, the deep structure proposed by Hinton is the Deep Confidence Network (Dep Belief Networks), which is a deep neural network model with multiple hidden layers, which is composed of multiple restricted Boltzmann machines (RBM). RBM is an energy-based probability generation model. Its essence is to make the learned model produce the most qualified samples. Its structure is a
two-layer cyclic neural network, including a visible layer and a hidden layer. The neurons in each layer are not connected to each other, only the visible layer and the hidden layer have non-directional symmetry connections. All neuron nodes are random binary variable nodes (only 0 or 1 value), and the nodes in the same layer are independent of each other. These limitations make training RBM more efficient and make it a structural unit of Deep Belief Network (DBN). The structures of DBM and RBM are shown in Figure 2 and Figure 3, respectively [5].

![Figure 2. RBM](image)

![Figure 3. DBN](image)

Conducted from simple RBM into DBN with deep structure, Hinton et al. proposed a greedy layer-by-layer unsupervised training method, which solved the difficulty of optimizing weights in deep structure. That is, the output of the bottom layer of DBN is used as the input of the upper layer, and each layer is an RBM. After training a layer, keep the obtained parameters unchanged, and use its output as the input of the next layer to start training. And so on, until the entire network is trained. Finally, the supervised training method is used to fine-tune the parameters of each layer, so that DBN has the most optimized weight [6]. The whole training process is divided into two steps: 1) Train the network in different levels, training only one level at a time; 2) Parameter tuning, that is, the process of supervising and fine-tuning. DBN has good results in terms of time and algorithm efficiency, and it uses continuous layers of binary or true-valued variables to learn the distribution of high-level representations, so it has good flexibility. Currently, DBN has been successfully applied in different fields, such as text representation, audio event classification, and visual data analysis.

3. Application of deep learning algorithms to image classification

The convolutional neural network model based on deep learning has achieved breakthroughs in many fields, such as image classification and object detection.

Image classification is to classify images into certain categories by analyzing them, emphasizing the judgment of the overall semantics of the image. There are many labeled data sets that can be used
to judge such algorithms, and the more commonly used ones are ImageNet and CIFAR-10/100. Among them, ImageNet contains about 15 million tagged high-resolution image materials, and these images are divided into about 22,000 categories [7]. Looking back at ImageNet's various image classification competitions, the participating teams in the past 10 years and 11 years mostly used classic feature extraction algorithms with traditional classification algorithms to solve the problem. Classic feature extraction algorithms include HOG, SIFT, etc., with an error rate of about 30%.

In 2012, the AlexNet deep learning algorithm was applied to large-scale image classification for the first time, and it took the top spot with a 16.4% error rate. AlexNet is a classic convolutional neural network. The first five layers are convolutional, the latter three layers are fully connected, and the classic softmax algorithm is used for classification. The team also proved that Relu is more suitable as an activation function than sigmoid and tanh. At the same time, it proposed to use Dropout technology to reduce the overfitting problem caused by too many parameters and insufficient data. Since the introduction of AlexNet, models based on deep convolutional neural networks have gradually become prominent in image classification algorithms with their advantages, and eventually replaced traditional algorithms as the most common mainstream method in image classification competitions.

In 2013, the winning team proposed a more in-depth understanding of convolutional neural network method-convolutional neural network visualization, based on the continuous characteristics of the image itself, using deconvolutional network to visualize the convolutional layer, so as to analyze the learning of each layer Characteristics. This method undoubtedly answers a question: Why can convolutional neural networks achieve good results in image classification? The algorithm error rate is as low as 11.7% [8].

In 2014, the Google team won the championship with GoogleNet, with an error rate of 6.7%, which is half of the best record. The network has 22 layers and uses multi-scale processing to improve the convolutional neural network. The innovation of the network is that it proposes the Inception module. The structural innovation in the use of this module is to replace sparse components with dense components in order to find the most effective and time efficient structure.

In 2015, the MSRA team won the three competitions of classification, detection and positioning with a 152-layer residual network, and the error rate in the classification project was reduced to 3.46%. The training difficulty of the deep learning algorithm becomes more and more difficult as the number of layers deepens, so many algorithms are produced to solve this problem, including layer-by-layer initialization [9]. The central idea of the residual network is to optimize the objective function to approximate an identity map, not a 0 map, which makes it easier to transfer gradients, and it is not easy to cause gradients to disappear or explode, thereby reducing the difficulty of network training and improving accuracy rate.

ImageNet's large scale and many types of characteristics make it more suitable for transfer learning ideas. The general scenario of transfer learning is: a deep model trained in a common scenario can share its network structure and parameters to a task scenario with a small data set, so that small data tasks can be fine-tuned on the basis of the deep model. In addition, the field of image classification is the basic field of computer vision and the first field to use deep learning. In other fields, based on the experience of classification tasks, the classification model is transferred to other tasks, which is also the development of various fields of computer vision. trend.
4. Deep learning applied to face recognition

The use of face recognition is now becoming more and more extensive. All walks of life are trying to use face recognition to improve customer experience and optimize service levels. In addition to Baidu’s search for images application, there is also automatic recognition of people through pictures uploaded by users. Faces are classified into albums by people, or physical anti-theft through face recognition, find wanted criminals, and design security systems. Face recognition mainly includes face detection and positioning and face feature extraction and recognition. The former is to segment the face target from the background, after normalization processing, the latter requires the feature extraction algorithm to be invariant. The technical difficulty faced by the former lies in the diversity of face target patterns and the complexity of the background. Therefore, it is necessary to make reasonable assumptions about the context to simplify the problem. On this basis, the construction of a high-dimensional spatial face model is the most complicated part, and it is very difficult to establish an accurate estimation. Therefore, how to achieve accurate estimation is one of the research hotspots. The technical difficulty faced by the latter is that the face is an elastic model, and the modeling difficulty is higher than that of a rigid body. Therefore, any feature extraction method based on rigid body characteristics is difficult to achieve satisfactory results. In addition, the variability of human faces is related to psychological and physiological characteristics. Therefore, it is better to combine elastic modeling with human visual characteristics [10].

Traditional face feature extraction and recognition methods mainly include the following:

1. Geometric feature method.

2. Eigenface algorithm.

3. Elastic model method. Compared with traditional feature extraction algorithms, the CNN algorithm has the characteristics of high stability, strong robustness, and high anti-interference, so that the technology of face recognition can be applied to business.

5. Conclusion

There are countless successful commercial cases of computer vision technology, which are mostly reflected in security surveillance, face recognition, criminal body recognition, household registration management, and e-commerce.

Computer vision technology replaces human vision systems, and due to its advantages of long working hours, large storage space, and fast calculation speed, it has been widely used in industry, military and other fields. The development of image processing and detection technology has greatly improved the accuracy and efficiency of related tasks in the industry. In addition, computer vision is also used in agriculture. The task of accurate and rapid classification of fruits and vegetables is very difficult to automatically classify. The main reason is that natural products such as fruits and vegetables have large differences in appearance. Therefore, manual classification is often adopted in my country. The current computer vision technology has been able to achieve automatic classification, mainly to detect the shape, size, color, surface defects and damage of fruits and vegetables, and use this as an evaluation standard to achieve classification.
The range of academic research in the direction of deep learning is very wide, using convolutional neural networks for learning and object feature extraction has been widely accepted. There are many areas of deep learning application and it has very strong versatility. I believe that with the optimization of the feature technology extracted by the network, deep learning will make greater progress in various applications of artificial intelligence.

References

[1] Tang Yiping, Yang Zhao, Shi Xingmin, Zhong Yuyun. Research on key technologies of parrot behavior analysis based on computer vision [J]. Small and Micro Computer System, 2016: 203-208.

[2] Zhu Xuehai, Zhang Shuai, Zhang Dongxing, Zhang Aping, Luo Yunfei. Research and Application of Intelligent Identification Technology for Ship Water Scale Based on Machine Vision and Deep Learning [J]. Journal of Inspection and Quarantine, 2019: 113-116+122.

[3] Bai Linting, Wen Pengcheng, Li Yahui. Research on Visual Question Answering Technology Based on Deep Learning [J]. Aeronautical Computing Technology, 2018: 344-348.

[4] ZHOU Yiqi. Application Progress of Deep Learning in the Field of Computer Vision [J]. Electronic Manufacturing, 2018: 30-31.

[5] YANG Chunfu. Research on Depth Estimation Method Based on Computer Vision [J]. Journal of Yantai Nanshan University, 2017: 63.

[6] LI Yi-rong. Research on Computer Vision Technology Based on OpenCV [J]. Electronic World, 2014: 147-148.

[7] WU Sikun. Research on Computer Vision Technology Based on Surveillance Video [J]. Digital World, 2018.

[8] Gao Zhengchuang, Dai Yaowei, Ye Jiming, Tang Shuyin, Li Hongyu. Research on key technologies of unmanned driving based on deep learning [J]. Electronic World, 2019: 78-79.

[9] Jia Jingping, Qin Yihua. Review of Visual Tracking Algorithms Based on Deep Learning [J]. Computer Science, 2017: 29-33.

[10] Liu Haiyan, Lu Yingfeng. Depth estimation method based on computer vision [J]. Science and Technology Information, 2018: 43-44.