Research on the image recognition technology

Yida Wang

University of Science and Technology Beijing, China

Keywords: Image Recognition, Artificial Intelligence, Convolution Neural Network (CNN), Nonlinear Dimension Reduction.

Abstract: From 3D picture construction to video processing, computer vision is playing an important role in many fields and will continue to enlarge its influence in the future. Image recognition is a technology within the range of computer vision, performing the function of processing and identifying the images to achieve the automatic tasks. With decades of development, the approach of image recognition has changed a lot and now most of it is based on deep learning. In this article, the background of image recognition technology and deep learning in image recognition technology will be introduced. And this article will focus on the development of the CNN model and its structure. Moreover, image recognition technology is successful in many aspects and this article also provides research on application of image processing and identification technology, the state of image recognition technology and the popular research area.

1. Introduction

Image recognition, a subcategory of Computer Vision and Artificial Intelligence, represents a set of methods for detecting and analyzing images to enable the automation of a specific task. It is a technology that is capable of identifying places, people, objects and many other types of elements within an image, and drawing conclusions from them by analyzing them. So that Image Recognition can be applied to many other fields such as transportation and medical treatment. Photo or video recognition can be performed at different degrees of accuracy, depending on the type of information or concept required. So there are different tasks that Image recognition can perform: Classification, Tagging, Detection and Segmentation. The approach to realize image recognition involves the extraction of image features and the classification of those featured images, so that the differences between extraction methods can affect the efficiency of recognition. A more detailed recognition procedure begins with image acquisition, to transform the natural light signal into digital image data for the following steps. In order to meet the demand of the methods and to ensure a better extraction performance, it should be pre-processed with techniques like denoising and blurring before the digital image is handled by feature extraction methods. When the image is finishing processing, the extraction methods can get various features of the image. Finally, those features can enter the classifier that has been adjusted by the training set and the objects in the image can be identified.

2. Image recognition technology

2.1 Background of image recognition technology

The Field of image processing is continually evolving. During the past five years, there has been a significant increase in the level of interest in image morphology, neural networks, full-color image processing, image data compression, image recognition, and knowledge-based image analysis systems. Image is better than any other information form for our human being to perceive. Vision allows humans to perceive and understand the world surrounding us. Image understanding, image analysis, and computer vision aim to duplicate the effect of human vision by electronically perceiving and understanding images. In digital image processing system, theoretically, image recognition is mostly based on Deep Learning. Deep Learning, a subcategory of Machine Learning, refers to a set of

automatic learning techniques and technologies based on Artificial Neural Networks. Image recognition technology has developed a lot with Deep Learning.

2.2 Deep learning in image recognition technology

Deep learning models currently applied to Image Recognition and analysis research mainly include Convolution Neural Network (CNN).

In the 1950s and 1960s, David Hubel and Torsten Wiesel discovered "simple cell" and "complex cell" that can respond to image patterns, and the "receptive fields" mechanism in the signal transfer between them.[1] In 1980, Fukushima designed the "Neocognitron" model using the concept of simple and complex cells.[2] In 1989, Yann LeCun et al. used back-propagation to improve a previous zip code recognition system that designed the parameters manually. Then in 1998, they developed the model "LeNet-5" to recognize 32x32 pixel images, and it became a classic CNN model.[3] [4] In 2012, Alex Krizhevsky developed the model "AlexNet" with new features such as Multiple GPUs and ReLU. This model won the ILSVRC, resulting in the popularity of CNN networks.[5]

The basic CNN structure consists of three types of layers[6][7]: Convolution Layer, Pooling Layer and Fully Connected Layer. Many Convolution Layers are connected with each other which is followed by a Pooling Layer. Fully Connected Layers are at the end of the CNN network. In the Convolution Layer, image data was inputted and convoluted by a filter. In other words, the numbers in the area, which has the same width and height as the filter, were multiplied by those corresponding numbers in the filter and then were summed up. The area moved from left to right and from up to down so that the whole image can be convoluted. The size of the results depended on the size of image, the size of filter and the area selection stride. There can be different filters executing the same procedure to get multiple depth data results. After that, the results from the convolution layer were sent to the pooling layer. In the pooling layer, the results data was divided into several little squares. For each square of data, the maximum or average number was chosen to substitute the square. By this procedure, useful information can be sorted out and the total amount of data can be reduced in order to improve the efficiency of the network[8][9].

Usually, a CNN has more than 1 set of convolution layer and pooling layer. Every time the data passed through a new layer set, a higher-level feature of the image came out. Finally, the highest-level feature which was close to human's normal recognition of the objective was sent to the fully connected layer and was determined whether the image described the objective. The fully connected layer acted the same way as the hidden layer in normal neural network, using weighted parameters in decision.
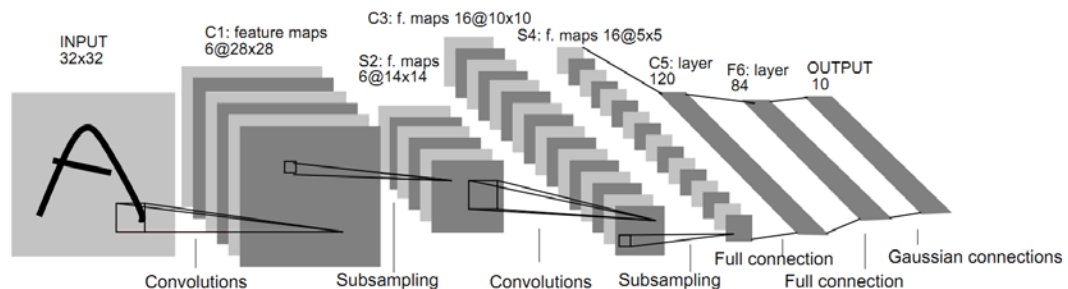

Figure 1 Architecture of Convolution Neural Network

2.3 Nonlinear image recognition technology

Computer image recognition technology is a high-dimensional technology. Regardless of the resolution of the image, it is always to be multi-dimensional which brings a lot of difficulties to computer recognition. To improve the ability to identify efficiently, the most effective way is to reduce dimensionality. Dimension reduction is divided into linear dimension reduction and nonlinear dimension reduction. Principal components analysis and linear discriminant analysis are two main

algorithms in linear dimension reduction. Principal components analysis can help identify features which are not necessary in model and make features mutually independent. The method is to choose vectors to project on which can maximize the data variance after the projection and finally transform high dimensional data into lower dimensional data. Those basis vectors are perpendicular to each other and are called principal components.

It has been verified that the linear dimension reduction has high computational complexity and takes up relatively more time and storage. Therefore, an image recognition technology based on nonlinear dimension reduction has been improved, which is an extremely effective method for nonlinear feature extraction.

3. The state of the image recognition technology

There are too many advantages in the development of image recognition technology recently. For example, the accuracy of recognition has reached a high level, benefited from the data pre-processing and various models[10][11]. The operations can include many different processing methods both linear and nonlinear with more choice. The information can be extracted from various sources as long as it can be described in two-dimensional array form, such as camera photo and digital image in painting software. And the information has a good potential to be compressed due to the relevance between nearby pixels[12]. Along with these advantages, however, there are also some problems restricting the development of Image Recognition[13]. One of them is the computer capability. Because the digital image usually contains a large amount of information and the training set is expected to contain many images, it requires a large storage capacity in the computer to store the information. Besides, it will take a long time to process all images if the computer processing speed is not fast enough, which requires the recognition model to be efficient[14]. Another problem is the recognition of 3D objects since it cannot be fully projected to 2D images without information loss so that it is hard to process and understand 3D objects. Considering these restrictions, attempts were made both in the computer aspect to improve the computing capability and in the recognition model aspect to optimize the network.

4. Research focus of image processing and recognition technologies

4.1 Artificial Intelligence Visualization

Visualization is a focus of current research of Image Recognition. Facilitated by a new neural network, fuzzy recognition and nonlinear dimensionality reduction, the recognition and processing technology has greatly improved[15][16]. And the connection between Image Recognition and Artificial Intelligence has produced new applications like ball tracking in sport games and generating the combination of two images with different styles. These applications provide the intuitive understanding of data and the exploration of art.

4.2 Virtual Realities

Virtual Realities have been the fastest growing and most promising applications in computer vision today. Virtual Realities applications based on the image processing and image recognition technology are numerous, real-time three-dimensional face recognition and tracking recognition of human movements in human-computer interaction[17][18]. Therefore, all require the deep integration of image processing and recognition technology. It is conceivable that this popular field will become more technical achievements and more closely related to daily life application.

5. Application of image processing and identification technology

5.1 Intelligent transportation

Image recognition technology applied to many applications, self-driving cars are a famous one. Many companies have researched this topic such as Google and Tesla. By capturing and identifying

objects on the roads such as cars, people and other moving objects, the self-driving car is able to avoid crashes and obey rules while driving[19][20]. And deep-learning based image recognition has improved its accuracy and speed. But self-driving cars are not fully achieved now and there are challenges to deal with.

5.2 Medical treatments

Medical imaging is another application field of image recognition. Medical imaging systems can acquire the patient data in different ways such as X-ray imaging system, computed tomography and nuclear medicine system. Then through image processing and recognizing, medical institutions can identify the abnormal organs and tissues in the patient body[21][22]. Among those image acquiring methods, computed tomography itself also uses image recognition technology when every scanned layer is put together to construct a 3D image[23].

6. Conclusion

In general, the image recognition technology mainly involves feature extraction and classification theoretically, and uses deep learning networks in practice. A classic deep learning model Convolution Neural Network sets up with three types of layers: Convolution Layer, Pooling Layer and Fully Connected Layer, each layer performs a specific function. To increase the speed and accuracy of recognition, images can be pre-processed with the Nonlinear Dimension Reduction technology which includes principal components analysis method. The application and development stage of image recognition technology has been mentioned above, however, there are still many restrictions in the development of Image Recognition like computer capability. The applications are abundant and creative with the connection of other research fields including transportation and medical treatment and more. From the point of view, I believe that the image recognition technology can be better applied to human life, but also the direction and its development.

References

[1] Hubel, D. H.; Wiesel, T. N. "Receptive fields and functional architecture of monkey striate cortex"

[2] Fukushima, Kunihiko "Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position"

[3] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition"

[4] LeCun, Yann; Léon Bottou; Yoshua Bengio; Patrick Haffner "Gradient-based learning applied to document recognition"

[5] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton "ImageNet Classification with Deep Convolutional Neural Networks"

[6] Zeiler, M. D., & Fergus, R. "Visualizing and understanding convolutional networks"

[7] Vincent Dumoulin, et al, "A guide to convolution arithmetic for deep learning"

[8] Venkatesan, Ragav; Li, Baoxin "Convolutional Neural Networks in Visual Computing: A Concise Guide"

[9] Ciresan, Dan, Ueli Meier; Jonathan Masci, Luca M. Gambardella, Jurgen Schmidhuber "Flexible, High Performance Convolutional Neural Networks for Image Classification"

[10] Zhenan Sun, Yunhong Wang, Tieniu Tan, Jiali Cui, "Improving iris recognition accuracy via cascaded classifiers"

[11] Jenkins, R. and Burton, A. M. "100\% Accuracy in Automatic Face Recognition"

[12] R. A. DeVore, B. Jawerth and B. J. Lucier, "Image compression through wavelet transform coding"

[13] M. Pak and S. Kim, "A review of deep learning in image recognition"

[14] K. Yanai and Y. Kawano, "Food image recognition using deep convolutional network with pre-training and fine-tuning"

[15] Seokwon Yeom, Inkyu Moon and B. Javidi, "Real-Time 3-D Sensing, Visualization and Recognition of Dynamic Biological Microorganisms"

[16] Potiyaraj, P., Subhakalin, C., Sawangharsub, B. and Udomkichdecha, W. "Recognition and re-visualization of woven fabric structures"

[17] J. Weissmann and R. Salomon, "Gesture recognition for virtual reality applications using data gloves and neural networks"

[18] Deyou Xu, "A Neural Network Approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG"

[19] C. N. E. Anagnostopoulos, I. E. Anagnostopoulos, V. Loumos and E. Kayafas, "A License Plate-Recognition Algorithm for Intelligent Transportation System Applications"

[20] A. Mammeri, E. Khiari and A. Boukerche, "Road-Sign Text Recognition Architecture for Intelligent Transportation Systems"

[21] F. Milletari, N. Navab and S. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation"

[22] M. Naf, O. Kubler, R. Kikinis, M. E. Shenton and G. Szekely, "Characterization and recognition of 3D organ shape in medical image analysis using skeletonization"

[23] Lu, Le, et al. "Deep learning and convolutional neural networks for medical image computing"