Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230875

Image recognition system based on deep learning technology

Zetai Lin

School of Computer Science and Electronic Engineering, Essex University, Colchester, Britain, CO4 3SQ

zl19012@essex.ac.uk

Abstract. In recent years, people have been able to easily obtain a large amount of visual information through various devices, but not all of these visual information are useful, and there is a lot of useless information mixed with them, which makes it difficult for people to identify and track the target parent in the picture. Due to its various characteristics, deep learning technology can quickly obtain information from massive information and compare, analyze and utilize it, so it has become the key to solving the problem of image recognition and tracking. This paper will summarize the existing literature, focus on the structure of the image detection system based on deep learning technology, explain its basic principles, and briefly introduce the basis of deep learning technology - neural network. This paper finds that there are many technical routes for image recognition systems based on deep learning technology, each of which has its own advantages and disadvantages.

Keywords: Neural Networks, Deep Learning, Image Recognition.

1. Introduction

An important research area in the science of computer vision is image recognition, a technology that enables computers to process, analyze, and comprehend images to recognize targets and objects of diverse patterns[1]. It has huge practical value and application prospects. Nowadays, the application fields of image recognition and tracking technology include intelligent video surveillance, robot navigation, obstacle detection in vehicle camera images, etc. Many research results have been applied in life. Early image recognition systems mainly used feature extraction methods such as Scale-invariant feature transform, which were difficult to extract a large number of features and had low recognition accuracy[2].

Deep learning technology has excellent characteristics such as “unsupervised” in the realization of image recognition. It can extract image features by building a neural network, subject the original data to a series of nonlinear transformations to generate high-level abstract representations, and use a large amount of material to train itself to accurately identify targets in images without supervision. Because of this, deep learning is widely used in image recognition systems and has achieved quite good results. This paper will try to describe the overall structure and construction method of an image recognition system based on deep learning technology, and sort out its construction process.

DOI: 10.54254/2755-2721/6/20230875

2. The overall design of target recognition and tracking

The three stages of the image recognition process are loosely categorized as follows: region proposal, feature representation, and region classification. First, it proposes possible target positions in the image, that is, proposes some possibilities for the candidate area containing the target. Then the appropriate feature model is used to obtain the feature representation. Finally, the classifier is used to determine whether each area contains a specific type of target, and through some post-processing operations, such as non-maximum suppression, border position regression, etc., Get the final target border.

In this process, we do not need to input a large number of rules, but only need to build a model with a small number of rules, and the computer can create new data through learning, so as to analyze and judge new things. The deep learning method realizes the automatic search, identification, classification, marking and tracking of the target by the system. Deep learning combines low-level features to form more abstract high-level representation attribute categories or features to discover distributed feature representations of data, so that machines have self-learning capabilities, and can automatically detect, identify and track specific targets, such as vehicles, people, and specified items [3].

Because the target background environment is complex and the scale changes drastically, traditional recognition algorithms mostly use known specific features for target recognition. For changeable targets and background environments, traditional algorithms often have poor recognition results. Target detection based on deep learning neural networks Recognition does not depend on specific features of fixed types has high recognition ability for image targets, and can complete recognition tasks in complex background environments.

After the accurate image features are obtained, the target object is tracked. The three main forms are discussed below.

The target is locked manually and then tracked by the computer. Technically, the traditional technical path of sample template matching is mostly used, the robustness of the target tracking algorithm is weak, and the tracking effect is limited. In this system design, the multi-scale self-learning tracking algorithm of "sample-training-detection" is comprehensively used, and at the same time, the powerful capabilities of the central processing unit (CPU) and field-programmable gate array (FPGA) are fully utilized, and the integrated A variety of target detection algorithms can solve the problem of target tracking robustness in harsh environments such as scene motion, non-rigid motion of targets, occlusion and self-occlusion, weak and noisy environments, and make the tracking of various targets more reliable[4]. Support inertial tracking, the tracking target is temporarily lost due to occlusion, and the inertial tracking of the tracking target can be maintained within a set time. Supports dynamic switching of tracking targets, without exiting the tracking of the existing target, and can quickly switch to the tracking of the next selected target.

Target library matches. Public datasets for target visual detection in order to promote the research progress of target visual detection, the construction of large-scale public datasets is an inevitable requirement. At present, the commonly used public datasets for target visual detection research include ImageNet, PASCAL VOC, SUN, and MS COCO. Before executing the task, the unmanned intelligent task platform binds the target image of this task into the memory of the intelligent image processing system, and sets the matching degree threshold. When the task starts, the intelligent image processing system automatically detects the target in the field of view, and matches and compares it with the target library in real time. When the matching degree is higher than the preset threshold, the system can automatically lock the target according to the signal, or manually participate in making the next step. Step by step instructions.

Automatic target detection and recognition, manual decision-making. After the intelligent image processing system performs deep learning on the specific target before the task is executed, it can automatically detect and identify the specific target in the field of view and mark the target when the task is executed. At this time, the servo system of the computer does not participate in the target tracking. When the decision maker selects a specific target in the field of view, the servo system tracking loop starts to work, and the target is always kept in the center of the field of view for the decision-making system to make the next action instruction.

DOI: 10.54254/2755-2721/6/20230875

3. Technical route

VGG network is a multi-layer deep neural network based on a convolutional network (CNN), which is the basis of this deep learning technique. There are two typical VGG networks, VGG16 and VGG19, with 16 and 19 layers, respectively. CNN is mainly composed of convolution layer, subsampling layer (pooling layer), fully connected layer and classification layer, each layer is composed of multiple two-dimensional planes (feature maps), and each plane is composed of multiple independent neurons. Each plane represents a featured image, and each pixel represents a neuron. The first few layers are alternately composed of convolution layers and pooling layers, followed by 2~3 fully connected layers and a Softmax classification layer [5]. This kind of deep structure extracts the features such as edges and blocks in the shallow layer to the complex semantic features in the high-level sequentially from shallow to deep. When the amount of data is large enough, the high-level semantic features can describe objects in a fine manner, and with some parameter optimization strategies, they can achieve high-precision target type recognition ability [6].

In order to efficiently detect larger targets and express the essential features of small targets, the target area is first up-sampled and down-sampled by means of an image pyramid, and then each layer of the pyramid is input to the convolutional neural network to extract features. The features extracted by this method can not only be robust to the scale changes of the target but also enrich the feature expression of the target by using the intermediate layer features of the convolutional neural network.

4. Data preprocessing

The first step in training is to collect data and make samples. The source of the sample is generally collected pictures or videos with the target and then use the labeling tool to mark the target to obtain a series of sample data sets. These data should be as true as possible in real life. The augmented data also needs to pay attention to the simulation. What the data looks like in real life. It is best to collect samples according to the actual data distribution. After processing, the collected samples should be consistent with the samples used when the model is used. The samples should be kept diverse, and various situations should be considered as much as possible.

The second step is data augmentation. Neural network training often requires a large number of samples. In addition to some necessary strategies during training, the larger the sample size, the easier the network is to fit and converge during training, and the stronger the network generalization ability. When faced with a specific task of classification and recognition, there is often not a large amount of data, so it is often necessary to augment the sample data to increase the sample size. The data augmentation strategy can also improve the recognition accuracy of difficult-to-classify sample types, which is often very effective for the classification and recognition of targets in specific application scenarios. A series of affine transformations can be performed on each type of target based on the original samples to generate new data, thereby increasing the sample size [7].

The last step is normalization and normalization. There are often some singular samples that differ greatly from the average in the collected data. Their existence will cause a series of problems such as increasing the training time and failing to converge the processing results. Normalization makes the preprocessed data limited to a certain range to eliminate the adverse effects caused by singular sample data. The purpose of standardization is to make the scale of the samples uniform. If the scales of the samples are inconsistent, consistent input data can be obtained in this way. For example, if pictures are collected by different cameras, there will be pictures of different scales due to various problems such as light and shade, contrast and so on. Some values that are too large will affect the convergence results. At this time, the convergence speed can be accelerated after normalization. It can also, to a certain extent, increase convergence accuracy.

5. Network training

Network training is the main process of neural network target recognition, which mainly includes network initialization, network fine-tuning training, network parameter model selection and other processes, and finally determines the recognition model.

DOI: 10.54254/2755-2721/6/20230875

Network initialization: For a deep convolutional neural network, features are combined layer by layer, so the more basic the features at the bottom layer, the greater the commonality. For example, the bottom layer convolution kernels learn mostly edge, block and other features, such features are often referred to as texture features. High-level features are often more complex, and the features learned near the top layer can roughly describe an object. The features at this time are called semantic features.

Network fine-tuning training: Network training and fine-tuning can be roughly divided into the feature extraction process of forward propagation and the parameter adjustment process of back propagation. The two are often inseparable as a whole during training. The training process is generally performed on batch data. Iterative process of this process until all training data is used up. Deep convolutional networks require multiple iterations of training until the network converges.

Network parameter model selection: After training, the training network parameters of multiple iterations will be saved. According to the parameters with the highest training accuracy and verification accuracy and the convergence situation, according to the higher verification accuracy and the smaller difference between the verification accuracy and the training accuracy. The principle of choosing the best network parameters as the final network model parameters for the application.

6. Conclusion

At present, the feasibility and practicability of this system have been proved by many researchers. Although its accuracy is not as good as that of the human eye, it can recognize images at an extremely fast speed, and the cost is very low, so it has been widely used. There are many technical routes for image recognition systems based on deep learning technology, each of which has its own advantages and disadvantages. It can be seen that this system is very flexible, and the image recognition system can be used in different fields by changing the deep learning model and the training data set. However, this paper only discussed one of them and did not study the technical details. It is better to show the differences and similarities between different models by comparing experimental data and giving the specific steps of the experiment for the convenience of readers' reference or suggestions.

References

[1] Lu Hongtao, Zhang Qinchuan. A review of the application of depth convolution neural n etwork in computer vision [J]. Data acquisition and processing,2016,31(01):1-17.DOI:1 0.16337/j.1004-9037.2016.01.001.

[2] Lowe D G.Distinctive image features from scale-invariant keypoints[J].International Journa l of Computer Vision,2004,60(2):91-110.

[3] Dumitru Erhan, Aaron Courville, Yoshua Bengio, Pascal Vincent. Why Does Unsupervise d Pre-training Help Deep Learning?. Journal of Machine Learning Research, 2010, 11: 625-660

[4] Zhang Junyang, Wang Huili, Guo Yang, Hu Xiao. A review of research related to deep learning [J]. Computer application research, 2018,35(07):1921-1928+1936.

[5] Guan Q, Wang Y, Ping B, et al. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot s tudy[J]. Journal of Cancer, 2019, 10(20): 4876.

[6] Chang Liang, Deng Xiaoming, Zhou Mingquan, Wu Zhongke, Yuan Ye, Yang Shuo, Wa ng Hong'an. Convolution Neural Network in Image Understanding [J]. Journal of Aut omation, 2016,42(09):1300-1312.DOI:10.16383/j.aas.2016.c150800.

[7] Lin Chengchuang, Chunchun, Zhao Gansen, Yang Zhirong, Peng Jing, Chen Shaojie, Hu ang Runhua, Li Zhuangwei, Yi Xusheng, Du Jiahua, Li Shuangyin, Luo Haoyu, Fan Xiaomao, Chen Bingbing. Overview of image data enhancement in machine vision ap plications [J]. Computer Science and Exploration,2021,15(04):583-611.