

Research on character recognition technology based on deep learning

Daodong Xiang

Wenzhou Polytechnic, Wenzhou, 325035, China

Abstract: In the security field, it is necessary to extract the information of each camera in real time. However, due to various reasons, each camera may have wrong shooting time, location and other information during the use process. If it cannot be found in time, it will bring great hidden danger to the security. This research based on CRNN depth learning algorithm to detect and recognize the text information on these pictures, and the results can effectively improve the accuracy of recognition.

Keywords: Deep learning; Character recognition technology; Complex background.

1. Introduction

OCR technology has a wide range of application scenarios. In many scenarios, there are requirements for text detection or character recognition, and there are more and more application scenarios for character recognition in natural scenarios. Due to the diversity, complexity and many interference factors of text in natural scenes, it is still difficult to accurately detect and recognize text, which is also a current research hotspot.

In the field of security, it is necessary to extract the information of each camera in real time. However, due to various reasons, the camera may have wrong shooting time, location and other information during the use process. If it cannot be found in time, it will bring great hidden dangers to the security.

Traditional OCR still stays on image processing and artificial feature extraction methods. In recent years, with the rapid development of image processing technology, especially the emergence of deep learning and convolutional neural networks, more and more traditional methods have been replaced by methods based on deep learning. This paper uses CRNN depth learning algorithm to detect and recognize characters on complex background images in specific application scenarios to improve character recognition rate.

2. Text features in complex background

(1) Diversity of texts

Different from text with regular font and neat arrangement in document images, text in complex backgrounds may have completely different fonts, colors, proportions and directions even in the same scene.

(2) Complexity of the background

In natural scenes, the background in images and videos can be very complex. Elements such as logos and tags are actually indistinguishable from real text, so it is easy to cause confusion and errors.

(3) Many interferences factor

Various interference factors, such as noise, blurring, distortion, low resolution, uneven illumination and partial occlusion, may lead to the failure of text detection and recognition.

3. Deep learning method

OCR identification process:

OCR recognition preprocessing: graying (if it is a color image), noise reduction, binarization, character segmentation and normalization. After binarization, there are only two colors left in the image, namely black and white. One is the image background, and the other is the text to be recognized; Noise reduction is very important at this stage, and the quality of noise reduction algorithm has a great impact on feature extraction. Character segmentation is to divide the characters in the image into single characters - one character at a time. If the text line is slanted, it is often necessary to correct the slant. Normalization is to normalize the text image of a single medium to the same ruler.

(2) Feature extraction and dimension reduction: features are the key information used to identify characters, and each different character can be distinguished from other characters through features. For numbers and English letters, this feature extraction is relatively easy, because there are only 10 numbers and 52 English letters, all of which are small character sets. For Chinese characters, feature extraction is difficult, because first of all, Chinese characters are large character sets. In the national standard, there are 3755 first level Chinese characters that are most commonly used; The second Chinese character has a complex structure and many similar characters. After determining which feature to use, depending on the situation, it is also possible to reduce the dimension of the feature. This situation is that if the dimension of the feature is too high, the efficiency of the classifier will be greatly affected. In order to improve the recognition rate, it is often necessary to reduce the dimension. This process is also very important. It is necessary to reduce the dimension and make the feature vector after reducing the dimension retain enough information.

(3) Classifier design, training and actual recognition: the classifier is used for recognition, that is, for the second step, extract features from a text image to the classifier, and the classifier will classify it to tell you which text this feature should be recognized.

(4) OCR recognition post-processing: post-processing is used to optimize the classification results. First, sometimes the classification of the classifier is not completely correct. For example, for the recognition of Chinese characters, it is easy to recognize a character as a similar character because of

the existence of similar characters in Chinese characters. This problem can be solved in post-processing. For example, the language model is used for correction. If the classifier identifies "where" as "where to store", it will find that "where to store" is wrong through the language model, and then correct it. Second, OCR recognizes images that often have a large number of characters, and these characters are stored in complex situations such as typesetting and font size. In post-processing, you can try to format the recognition results, such as arranging them according to the typesetting in the image.

The system flow of traditional OCR technology is basically shown in the figure 1.

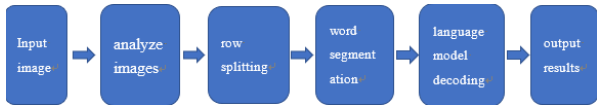


Figure 1. Traditional OCR technology

With the continuous development of deep learning technology, the traditional OCR technology framework is ignored. In the field of computer vision, convolutional neural networks are used to do research on OCR technology, and the recognition rate is very high, the effect is very good. You don't need to spend a lot of time designing character features.

In OCR system based on deep learning, neural network model is used as feature extractor and classifier. The input is character image, and the output is recognition result, so the whole workflow becomes very simple.

The typical neural network structure is shown in Figure 2.

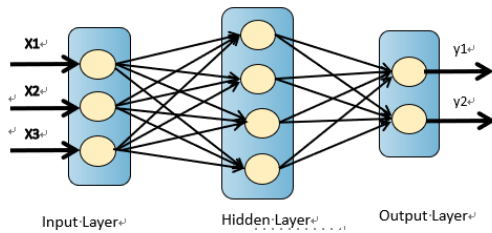


Figure 2. Typical neural network structure

Neural network (NNs) is an algorithm mathematical model that imitates the behavior characteristics of animal neural networks and conducts distributed parallel information processing. This kind of network relies on the complexity of the system, and achieves the purpose of processing information by adjusting the interconnection between a large number of internal nodes. The purpose of neural network design is to try to write a general network model, and then train against the data, so as to continuously improve the parameters in the model, and finally achieve the output results meet the expectations.

Common neural networks are as follows:

Feedforward neural network

A unidirectional multilayer structure is adopted. Each layer contains several neurons. The neurons in the same layer are not connected with each other, and the information between layers is transmitted only in one direction. The first layer is called the input layer. The last layer is the output layer. The middle layer is the hidden layer, referred to as the hidden layer. The hidden layer can be one layer. It can also be multi-layer.

Convolution neural network:

The convolutional neural network is composed of one or more convolution layers and the top full connection layer (corresponding to the classical neural network), as well as the associated weight and pooling layer. This structure enables the convolutional neural network to use the two-dimensional structure of input data. Compared with other deep learning

structures, convolutional neural network can give better results in image and speech recognition.

Cyclic neural network

RNN is a neural network used to process sequence data. It introduces structural units with "memory" property. Besides this input, the calculation also includes the last calculation result.

4. CRNN model structure

CRNN is a convolutional recurrent neural network structure, which is used to solve image-based sequence recognition problems, especially scene character recognition problems. CRNN network realizes indefinite length verification, combines CNN and RNN network structures, uses bidirectional LSTM cyclic network for timing training, and finally introduces CTC loss function to realize end-to-end indefinite length sequence recognition

The network structure consists of three parts, from bottom to top as Figure 3.

(1) Convolutional layer. The function is to extract feature sequences from input images.

(2) Cycle layer. The function is to predict the label (real value) distribution of the feature sequence obtained from the convolution layer.

(3) Transcription layer. The function is to convert the label distribution obtained from the loop layer into the final recognition result through de integration and other operations.

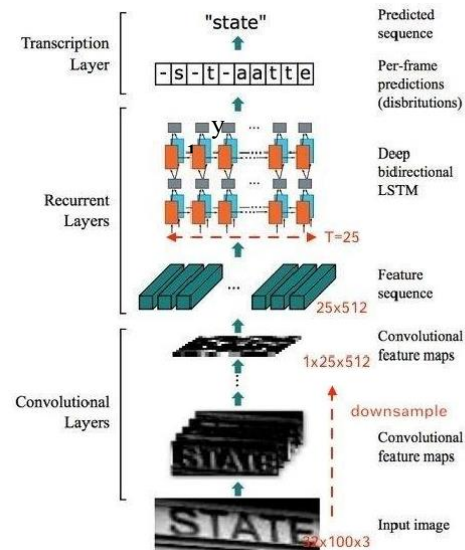


Figure 3. CRNN network structure

From: <https://zhuanlan.zhihu.com/p/43534801>

5. Experimental process

1)Collection of data sets

The pictures in this study are from real-time photos taken by cameras installed at different points. Through sorting out the collected data, we found that the most prone to errors were the irregular arrangement of characters and the lack of characters.

2)Dataset Extension

Preprocess the real background texture material. For image data, expand the training data by rotating, converting, and distorting the image, and train the model on the extended data set.

3)Model construction

Build a basic CNN network. In the training process, first

use the standard CNN network to extract the features of the text image, then use BLSTM to fuse the feature vectors to extract the context features of the character sequence, then obtain the probability distribution of each column of features, and finally predict the text sequence through the transcription layer (CTC).

4) Model training

The specific model training process is as follows:

- a) Zoom the input image to $32 * W * 3$
- b) Use CNN to extract the image convolution feature, and the size is $1 * W/4 * 512$.
- c) Extract sequence features from LSTM through the above input to obtain $W/4 * n$ posterior probability matrix
- d) Utilize CTC loss to realize one-to-one correspondence between labels and outputs for training.

6. Conclusion

Limited by traditional computer vision algorithms, traditional OCR only performs well on regular printed documents, such as high-quality scanned documents. Because traditional OCR often relies on complex process optimization and manual design to adapt to scenarios, the scenarios have poor universality, and in different business scenarios, a lot of manual fine-tuning is often required to adapt to differences; Under complex scenes, the performance and accuracy of character recognition are not ideal.

Text recognition for images in natural scenes is much more difficult than text recognition in conventional scanned document images, because the form of text presentation in natural scenes is very complex and irregular. CRNN model is

trained to solve the problem that some Chinese cannot be recognized, reduce the recognition error rate of some characters caused by background interference.

References

- [1] Zhang Tingting. Based on Tesseract_ Research on OCR character recognition system [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2020.
- [2] Wang Yiwen. Research on the Application of Deep Convolution Neural Network in OCR [D]. Chengdu: University of Electronic Science and Technology of China, 2018.
- [3] Wang Yang, Li Zhendong, Yang Guanci. Research on the application of OCR character recognition based on deep learning in the banking industry [J]. Computer Application Research, 2020.
- [4] Zhao Shanshan. Design and implementation of image form data recognition system based on OCR technology[D]. Nanjing: Southeast University, 2020.
- [5] Zeng Yue, Ma Mingdong. Research on character recognition based on Tesseract_OCR [J]. Computer Technology and Development, 2021.
- [6] Xiao Jian. Learning Based OCR Character Recognition [J]. The Computer Age, 2018.
- [7] Han Ping, Liu Zexu. An Improved Method of X-ray Baggage Image Enhancement Based on Gray Level Grouping [J]. Journal of Civil Aviation University of China, 2011, 29 (4): 23-26.