IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 11, Issue 6, June 2021, ||Series -II|| PP 07-11

Analysis of manuscript recognition based on Deep Learning

Tian Jipeng
Computer Science College, Zhongyuan University of Technology, China

Hou Lingmei
Dos in Computer Science, University of Mysore, Mysore,India

G. Hemantha Kumar
Dos in Computer Science, University of Mysore, Mysore,India
Received 10 June 2021; Accepted 24 June 2021

ABSTRACT
Recognition the library's ancient books, transform them into text coding information, can reduce the workload of the ancient book workers .Deep Learning ODR methods divided some cumbersome processes into two main steps, one is the location of text detection for positioning text, and the other is text identification for identifying text. The Common text identification frames mainly have two, namely CNN + RNN + CTC and CNN + SEQ2SEQ + Attention. The Structure is similar, and the previous encoder CNN is mainly used for encoding extraction features, and the subsequent decoder RNN + CTC and SEQ2Seq + Attention are different. The text identification method has direct row text, curved text, character scale, and predetermined part. In deep learning, optimization and generalization techniques are particularly important, so a gradient decrease algorithm has occurred, and new technologies such as regular gradient cuttings used during training. Convolutional neural networks are one of the important algorithms that include convolutional calculations and have depth structures, is one of the important algorithms of deep learning.
KEYWORDS: CNN, Machine Learning, Manuscript, RNN, Recognition

I. INTRODUCTION

We take the signboard information of the roadside in our daily, extract text logo, integrated the current GPS positioning, you can directly match the evaluation page of the store, so it will be used to enter keyword search;It is an important technology-Optical Character Recognition (OCR)after the above scenes .The Chinese character recognition was earliered by IBM's engineer Casey and Nagy, and they published the first Chinese character identification related articles in 1966, using template matching. 1000 print body characters can be identified. After that, OCR technology has been studied, and it is now widely used in various fields. Before the depth learning method has not occupied the dominant position, text detection is mainly based on manual extraction characteristics, and the classic method has previously mentioned SWT, MSER, HOG, etc. The traditional way is to first set the image setting feature pyramid, then use the sliding window to scan, then enter the manual extraction feature phase (SWT, MSER, HOG, etc.), then extract the classification by the sliding window, and finally summarize the text area. Deep learning OCR methods divide some cumbersome processes into two main steps, one is text detection, mainly used to locate the location of the text. The other is text identification, mainly used to identify the specific content of the text. The object of text recognition is to identify the positioned text area, mainly solved the problem of transcribing a string image as a corresponding character. Deep study and text recognition have a deep origin. Text identification technology has been using convolutional neural networks and circulating neural networks before depth learning has not become a hot research object. For example, the Lenet5 network that occupies an important position in the deep learning system is applied to the OCR, serializing model LSTM, BLSTM, and CTC LOSS have also been applied in handwriting English.

Common text identification frames mainly have two, namely CNN[1] + RNN + CTC[2,3,4] and CNN + Seq2Seq + Attention. The structure is similar, the previous encoder CNN is mainly used for encoding extraction features, subsequent decoder RNN + CTC and SEQ2seq + Attention is different. The structure is similar, and the previous encoder CNN is mainly used for encoding extraction characteristics, and subsequent decoder RNN + CTC and SEQ2seq + Attention are different. The text identification method has direct row text, curved text, character scale, and predetermined part.

II. DEEP LEARNING ON MANUSCRIPT RECOGNITION

The development of deep learning, especially in the field of object detection, has greatly promoted the development of text detection, but text detection has its particularity compared to general object detection. The specific description is as follows.

- Diversified background. In a natural scene, the background of the text line can be any background, and it will be affected by some backgrounds with similar texture structures.
- Variety of text line shape and direction. Such as horizontal, vertical, inclined, curved, etc.
- Variety of text line colors, fonts, and scales (the length of the text varies, and the aspect ratio can reach 1:100 or even higher).
- Different degrees of perspective transformation .Poor lighting conditions and varying degrees of occlusion.

2.1 Preprocessing

In text detection and text recognition, the quality of the image is directly related to the detection rate and recognition rate. Therefore, preprocessing the image is an important link that cannot be ignored. Commonly used algorithms in image preprocessing include binarization, denoising and tilt angle detection and correction.

2.1.1 Binarization

Image binarization refers to setting the gray value of the pixel to 0 or 255 to make the image appear obvious black and white effect. On the one hand, binarization reduces the data dimension, on the other hand, by eliminating the interference caused by the noise in the original image, it can highlight the contour structure that is effectively removed. The OCR effect depends largely on this step, and high-quality binary images can significantly improve the accuracy of recognition. Currently, Binarization of Angfa is mainly divided into global threshold method (Global Binarization), local threshold method (Local Binarization), deep learning-based methods and other methods.

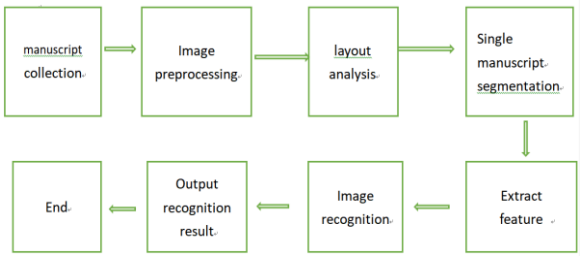

Fig. 1 The flow chart of mancuscript recognition

With the rapid development of deep learning technology, more and more workers are trying to build neural networks to binarize images and reach the level of state-of-the-art.

The Multi-Scale Fully Convolutional Neural Network was proposed by Chris Tensmeyer and others in 2017 [5], using a multi-scale fully convolutional neural network to binarize the image of the document image, and use it on the two public data sets of DIBCO and PLM both achieved good results.

In 2016, Rupinder Kaur et al. proposed a document image binarization method based on morphology and threshold[6]. The realization of this method can be divided into the following four steps.

- Convert RGM image into Grayscale image.
- Image filtering processing .
- Mathematical Morphology Operation .
- Threshold Calculation.

Among them, filtering is divided into Wiener Filter (Wiener Filter) and Gaussian Filter (Gaussian Filter). Wiener Filter, also known as Least Square Filter, uses the correlation and spectral characteristics of a stationary random process to filter signals mixed with noise ; Gaussian filter is a linear smoothing filter, adapted to eliminate Gaussian noise.

Mathematical morphology is one of the widely used technologies in image processing. By extracting the image components that are meaningful for expressing and describing the shape of the region from the image, the subsequent work can capture the most distinguishing shape of the target object feature. Among them, the basic operations of the binary image include: corrosion, expansion, opening operation and closing operation.

2.1.2 Smooth and dry

Image noise refers to unnecessary or redundant interference messages in the image data, which are generated in the image acquisition, quantification or transmission process, and will have a great impact on the

post-processing and analysis of the image. Therefore, it is a good solution. The noise method needs to maintain the boundaries and details of the image while removing the noise. Early denoising methods were mostly Spatial Filter. Spatial filtering consists of a field and predefined operations performed on pixels in the field. After the center of the filter traverses each pixel of the input image, the processed image is obtained. After each pixel point, the pixel value of the center coordinate of the field is replaced with the calculation result of the predefined operation. If a linear operation is performed on the image pixels, the filter is called a linear spatial filter, otherwise it is called a nonlinear spatial filter. Kin Gwn Lore et al. proposed the concept of The Low-Light Net (LLNet) [5]in 2017. LLNet realizes the adaptive enhancement (brightening and denoising) of low-noise images by introducing the idea of the sequential similarity detection algorithm (Stacked Sparse Denoising Autoencoder, SSDA).

2.1.3 Tilt angle detection and correction

During the scanning process, document rotation and displacement are prone to occur. Therefore, subsequent text line extraction and text recognition are inseparable from the Skew Detection and Correction link. Common methods include Hough Transform, Randon Transform, and PCA (Principal Component Analysis)-based methods.

2.2 Feature Extraction

Please use a 9-point Times Roman font, or other Roman font with serifs, as close as possible in appearance to Times Roman in which these guidelines have been set. The goal is to have a 9-point text, as you see here. Please use sans-serif or non-proportional fonts only for special purposes, such as distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

Feature extraction

Convolutional neural network is a kind of feed forward neural network that contains convolution calculation and has a deep structure. It is one of the important algorithms of deep learning. LeNet5 [7]was proposed by LeCun et al. in 1998 and consists of a convolutional layer, a pooling layer and a fully connected layer. LeNet5 and its subsequent variants define the basic structure of modern convolutional neural networks. By effectively extracting the translation-invariant features of the image, CNN has been widely used in computer vision and other fields. In traditional detection, Hear can express the edge information of the object due to its fast extraction speed, and can use

The integral graph is quickly calculated, so it is widely used. In terms of texture information expression, the LBP algorithm has better adaptability to uniformly changing illumination. The HOG algorithm uses histogram statistics to encode the edges of objects, and obtains a better feature expression, and has a wide range of applications in object detection, tracking, and recognition.

The encoder provides the decoder with features that can be learned, and the quality of the features is determined by whether the features can be distinguished.

- RRAM[8]
- FAN(Focusing Attention Network)
- Spatial Attention
- STN-OCR[9]
- ASTER[10]

2.3 classification

In a deep learning system, after detecting the target, it is often necessary to use a classifier to identify the detection area. In the real world, there are many types of text data, thousands of fonts, hundreds of languages, and some language families have many categories (for example, Chinese characters are divided into simplified and traditional Chinese characters, more than 6000 commonly used simplified characters, etc.), plus various Deformation, tilt, transparency, distortion, perspective changes, artistic characters, etc., determine the need to rely on a large number of simulations to achieve the purpose of generating training data. This requires us to generate the corresponding text according to the specific scene.

The first step is to collect the target fonts according to the specific scene. Generally, there may be more than a dozen fonts. Of course, some common fonts (such as Kaiti and Songti) can be added by default to enhance the generalization ability of the model; the second step is to collect relevant background pictures , Collect, crawl, and clean the corpus, and thus form a rich font library, background library, and corpus; the third step is to randomly select and combine various perspective transformations to generate a wider range of data that is as close to the actual scene as possible. Reduce the difference between the generated data and the actual scene data to get the best results.

Generative Adversarial Network (GAN)[11], which was proposed by Ian Goodfeiiow. Up to now, there are more than 400 variants of GAN, and many of the latest articles are related to GAN. The application scenarios of GAN basically cover all areas of AI, such as image and audio generation, image style transfer, image restoration (denoising and demosaicing, etc.), image super-resolution reconstruction, and text generation in NLP.

Generative Model (Generative Model) has always occupied a pivotal position in the history of machine learning. It is mainly used to describe the distribution of data and can directly model the data. When we have a large amount of data (for example, images, text, voice, etc.), if the generative model can help to fit these high-dimensional data, it is convenient to perform operations such as prediction and inference on the data. And for some scenarios where data is relatively scarce, the generative model can generate a large amount of data, improve the quality of the data, and use semi-supervised methods to improve the efficiency of learning. Generating a confrontational network, as the name suggests, is to introduce the idea of confrontational games on the basis of a generative model. We can use GAN to transfer images of one style to another style; we can also use GAN to generate artistic characters, and we can also use GAN to generate license plate images.

If the entire game process is described in mathematical language, it can be expressed as follows: Suppose the generative model is g(z), where z is a random noise, and the generative model g(z) can convert random noise into data type x. Still taking the picture problem as an example, the output of the generated model here is a picture. Assuming that D is a discriminant model, for any input x, the output of D(x) is a real number in the range (0, 1), which is used to determine the probability that the picture output by the generated model is a real picture.

Table 1 Traditional test methods

Methods	Year	Advantages	Disadvantages
Li et al.	2000	Supports detection and tracking of text in videos	Only supports horizontal text detection
Chen et al.	2004	Support text image detection in complex scenes, fast	Only supports horizontal text detection
Liu et al.	2006	Support natural scene text image detection	Only supports horizontal text detection
Neumann et al.	2010	Support text image detection in complex scenes, multi-language, faster	Only Support horizontal text detection
Yi et al.	2011	Support multi-directional text image detection, multi-language	Limited to simple scenarios, relying on manual design rules
Yao et al.	2012	multi-language, faster	relying on manual design rules
Huang et al.	2014	Support text image detection in complex scenes with superior performance	Only supports horizontal text detection

GAN has produced many variants in the continuous development (for example: conditional confrontation network, cyclic confrontation network, dual generation confrontation network, etc.). Through practice, it can be found that many adversarial networks are not suitable for the style transfer of text, because the text itself has small characteristics, and the domain-to-domain approach is generally not suitable for text generation. However, pair-to-pair (pairs) Is more suitable. For example, we use the improved pix2pix method to binarize text. In addition, pix2pix can also be used to generate data in fixed natural scenarios such as ID number generation and bank card number generation.

III. EXPERIMENTAL ANALYSIS

The most basic definition of manuscript recognition is the process of converting a text image contained in a picture into a computer text format. Because the source of the image and the scene cannot be restricted, standard data sets under different scenes are produced. The standard data set is a recognized benchmark for testing different performance and solving different difficulties in tasks. At present, according to the shape of the characters in the recognition task, the standard data set is mainly divided into arbitrary shape, multi-orientation, horizontal orientation data set, etc.; and according to the source of the image, it can be divided into network synthetic image data set and natural scene image data set ; According to different fonts, it can also be divided into handwritten and printed data sets.

IV. CONCLUSON

In recent years, due to the popularization of deep learning in various fields, handwritten document recognition has also made considerable development and progress. Researchers use deep learning to improve character recognition algorithms. Both recognition tasks and detection tasks have been significantly improved, and even in scenes such as license plate recognition, the trend of algorithm commercialization is shown. This is not only the technological innovation brought about by deep learning, but also the result of everyone's unremitting efforts day and night. Although in some scenarios, handwritten document recognition algorithms are already very mature, in more complex scenarios, the existing algorithms still cannot meet the demand. Therefore, there is still a lot of research and innovation space in the field of manuscript recognition waiting for our region to explore.

REFERENCES

[1]. Sermanet P, Chintalas, LeCun Y, Convolutional Neural Networks Applied to house numbers digit Classfication [J],Pattern Recognition. International Conferernce on 2012:3288-3291.
[2]. He P, et al. Reading Scene Text in Deep Convolutional Sequences[J]. Thirtieth AAAI Conference on Artificial Intelligence, 2016.
[3]. Shi B, Bai X, Yao C. an End-to –end Trainable Neural Network for Image –based Sequence Recognition and Its Application to Scene Text Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016.
[4]. Wang K, Babenko B, Belongie S. End-to –end Scene text recognition [C]//2011 Intenational Conference on Computer Vision IEEE,2012:1457-1464
[5]. Tensmeyer C, Martinez T. Document Image Binarization with Fully Convolutional Neural Networks[C].2017 14th IAPR International Conference on Document Analysis and Recognition(ICDAR). IEEE,2017.1:99-104
[6]. yoti D,Raj B,Sharma A,ET AL. Document Image Binarization Technique for Degraded Document Images by Using Morphological Operators[J]. Int Adv Res Ideas Innov,2016,2(3):1-7.
[7]. Burger HC,Schuler C J,Harmeling S. Image Denoising with Multi-layer Perceptrons,Comparision with Existing Alogrithms and with Bounds, Computer Science,2012
[8]. Lee C Y,Osindero S. Recurrent Nets with Attention Modeling for Ocr in the Wild[C]. THE Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2016:2231-2239.
[9]. Bartz C,Yang H,Meinel C.stn-ocr:a Single Neural Network for Text Detection and Text Recognition[J]. Arxiv Preprint Arxiv,2017:1707-8831
[10]. Shi B, Yang M, Wang X,et al. Aster: an Attentional Scene Text Recognizer with Flexible Rectification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.
[11]. Goodfellow I,Pouget-Abadie J, Mirza M, et al. Generative Adversarial Nets[J].Advances in Neural Information Processing Systems,2014:2672-2680
[12]. DAFENG ZHANG, Research on Character Recognition Algorithm Based on Deep Convolutional Neural Network [D]. Guizhou University,2019
[13]. Xue Gao, Lianwen JIN, Junxun YI, A handwritten Chinese character recognition method based on SVM.[J]Electronicnewspaper,2002,30(5):651-654

Tian Jipeng, et. al. "Analysis of manuscript recognition based on Deep Learning." *IOSR Journal of Engineering (IOSRJEN)*, 11(06), 2021, pp. 07-11.