Survey of Text Recognition Processes in Natural Scenes

Minghao Diao [a], Yanjun Yin [b]

College of computer science and technology, Inner Mongolia normal university, Hohhot 010022, China.

[a]nsddiaominghao@163.com, [b]ciecyyj@163.com

Keywords: OCR; detection algorithm; recognition technology.

Abstract: With the rapid development of Internet and artificial intelligence technology and the maturity of OCR technology, more software applications with OCR technology have entered people's daily lives. With the continuous development of deep learning technology in the field of vision, OCR technology has also broken through the bottleneck of the traditional technology framework. In recent years, a variety of detection algorithms such as anchor-based, pixel-based and pixel-anchor have emerged, and the use of machine recognition instead of manual entry has become a research hotspot. In this paper, the general research process of OCR in natural scenes, namely image preprocessing, text line location detection and text recognition, is reviewed. The problems of OCR text recognition and future research directions are expounded.

1. Introduction

With the rapid development of the Internet and artificial intelligence, more and more traditional work and repetitive work have been replaced by automated systems. OCR (Optical Character Recognition), one of the classic technologies in the field of computer vision, is commonly used for print recognition, document recognition, license plate recognition, license/business card/ticket recognition, and handwriting recognition. OCR technology can not only improve people's recognition efficiency, but also recognize many languages. It has high accuracy, high stability and strong applicability. It has been successfully applied to Microsoft, Nuance, OPENTEXT, Industrial and Commercial Bank of China, Huaxin Airlines and other well-known domestic and foreign companies, and has been tested by a large number of customers and various complex scenarios.

With the popularization of information automation and office automation in China and the rapid growth of deep learning in recent years, the further development of OCR text recognition technology has been greatly promoted. This paper first gives a brief overview of the concept and process of text recognition. Then, the domestic and international research status and future research directions of image preprocessing, text line location detection algorithm and text recognition technology are elaborated and analyzed. Image preprocessing is mainly to process the image. Some processing methods make the image recognition more effective. It is indispensable to locate the text lines in the image to identify the text in the image. Therefore, the image preprocessing and text line location detection algorithms play an important role in the overall recognition process.

2. OCR Text Recognition Overview

OCR text recognition refers to the process of checking printed characters on an electronic device (such as a digital camera or scanner) and translating the shape into computer text using a character recognition method. That is the process of scanning the document, analyzing and processing the document image, and obtaining the text and layout information, which is suitable for multi-scene. The concept of optical character recognition was first proposed in 1929 by Tausheck, a German scientist, and applied for a patent. But this dream did not come true until the birth of computers. Up to now, it has become the use of optical technology to scan and recognize characters and characters. The development of OCR can be divided into three stages: the first stage, in the early 1960s, several famous companies in the world, such as IBM and NCR, launched their own OCR identification

software. It can be described as the first generation of OCR identification products. At that time, the technology could only identify the numbers, English letters and some symbols of the printed body, and must be the specified font. In the second stage, OCR identification products are limited to the recognition of handwritten digits in the early stage; in the third stage, OCR develops to identify poor quality documents and large character sets. Text detection usually uses convolutional neural networks to extract scene image features and then process them by target regression [1, 2, 3], with semantic segmentation [4, 5] or by adding attention mechanisms [6, 7]. With the popularization of information automation and office automation in China and the rapid growth of deep learning in recent years, the further development of OCR text recognition technology has been greatly promoted.

OCR is essentially image recognition, which includes three key technologies: image preprocessing, text detection, and text recognition. Firstly, the image is denoised, binarized and so on, to enhance the image feature effect; then the image feature is extracted and the target area is detected; finally, the characters in the target area are segmented and classified. In the research of traditional OCR text recognition, the classification process includes the following steps: first, image preprocessing; secondly, layout analysis, text line location; finally, character segmentation, recognition and post-processing. Until nearly five years ago, the traditional OCR recognition technology framework was still the most widely used in the industry. The key point of OCR is the rapid development of deep learning. With the rise of deep learning, the OCR recognition framework based on this technology and another new idea have quickly broken through the original technical bottlenecks, and has been widely used in industry.

For OCR, images of different definitions have a greater impact on text recognition. For a long time, complex backgrounds, exposures, cluttered typographic styles, handwriting recognition, multi-language mixing, formula recognition and other issues are still difficult to overcome in computer academia and industry. Text detection in natural scenes has been widely used in artificial intelligence, image retrieval [46, 47], letter extraction, computer vision, video subtitles [48] and other fields.

3. Text Recognition Process

3.1 Image Preprocessing

Image preprocessing is mainly to process the image. Some processing methods make it better to recognize the image when it is recognized. Therefore, image preprocessing plays an important role in the overall recognition process. Image preprocessing mainly includes noise reduction, enhancement, tilt correction and binarization.

Noise is an important cause of image interference. There are many ways to reduce noise. The common methods are adding salt and pepper noise (salt = white, pepper = black) to the original image or using median filtering, minimum filtering and maximum filtering algorithms to reduce noise.

Image enhancement is to enhance the useful information in the image, so that the contrast of the image is enlarged, the image becomes clear, and the features are more obvious. Grayscale stretching is one type of image enhancement, and grayscale stretching is also called contrast stretching. The commonly used image enhancement method also has a maximum entropy method, which distributes all the pixels in the image to different gray levels as much as possible, obtains the maximum entropy and performs histogram equalization processing. The image contrast of the histogram equalization is improved, and the image quality is improved.

The image obtained by the input device is inevitably tilted into the recognition system, which will cause difficulties in subsequent image segmentation, character recognition, etc. Therefore, tilt correction is also an important part of image preprocessing. Hough transform, as a commonly used tilt correction algorithm, is one of the basic methods for identifying geometric shapes from images. Since Hough was proposed in 1962, this method is robust to noisy images and can be implemented

in parallel [12,13]. As early as 2002, Liang Dong, Fu Qizhong et al. [14] proposed using Hough transform and inverse Hough transform to locate and reconstruct the license plate area. After Yang Xining, Duan Jianmin et al. [53] also proposed a lane detection method based on improved Hough transform, using Hough transform based on dynamic ROI for lane tracking; narrowing the voting space scale of Hough transform space and improving real-time detection Sex and stability. Zhu Guiying and Zhang Ruilin et al. [15] proposed to improve Hough by replacing the classical cyclic process with multidimensional arrays. By contrast, the Hough transform detection speed and detection speed of multidimensional arrays are improved.

In the threshold segmentation, the Otsu algorithm (Otsu method or the largest inter-class variance method) [49, 50] is widely used because of its advantages of simplicity, stability, and strong adaptability. It is an efficient algorithm for binarizing images. Compared with the traditional Otsu, Zhao Mengchao et al. [8] proposed an Otsu algorithm based on second-order oscillatory particle swarm optimization. For the segmentation results, the morphological analysis and the hole filling process are performed to obtain accurate segmentation results. The experimental results show that the proposed algorithm is about 88% higher than the traditional Otsu algorithm, and the global convergence is better than the traditional Otsu algorithm. Although the Otsu algorithm has the characteristics of simple operation and high efficiency, it has poor anti-interference ability. Lu Qiuju [10] proposed to eliminate the two-dimensional algorithm in the search for defects in the image, you can use two one-dimensional Otsu algorithms to find the threshold required for two-dimensional Otsu. For the problem of high complexity of 3D Otsu algorithm, the improved ion algorithm can be applied to the 3D Otsu algorithm. The experimental results show that adding the ion algorithm not only improves the program operation speed but also improves the anti-interference ability of the algorithm [11].

3.2 Text Line Location and Detection Algorithms

To recognize the text in the image, the text lines in the image should be located first. Detection algorithm plays an important role in image recognition. Traditional image text detection uses photoelectric scanning equipment to scan printed or handwritten documents into images, and then get the text [31] in document images. However, in the natural scene image, the background is complicated, the shape of the text, the shape and size of the font are not uniform, and the illumination is uneven, which is difficult to detect. Under the rapid development of deep learning, a variety of new detection algorithms have emerged, and anchor-based, pixel-based and pixel-anchor have shown good results in natural scenes.

The traditional method runs fast and has low hardware configuration requirements, but it is only suitable for images with simple background. In the natural scene image, it is difficult to perform text detection due to complex background, text shape and uneven illumination. For traditional detection, it can be roughly divided into two categories: natural scene text detection method based on sliding window and natural scene text detection method based on connected domain.

The natural scene text detection method based on the sliding window mainly uses a multi-scale sliding window to scan the entire image to search for the position of the text information in the image. Kim et al. [32] according to the consistency of text color and multi-resolution wavelet transform method, finally send the feature into the support vector machine, determine whether the candidate region is text, and finally get the text region. Chen et al. [33] proposed edge extraction of images using the Candy edge detection operator. Babenko et al. [34] used a sliding window in conjunction with a Histogram of Gradient (HOG) [35] and then filtered out the non-text regions in the graph using a Random Ferns classifier [36] to obtain a text region. Chen et al. [37] combined the multi-scale sliding window with the AdaBoost algorithm [38] to combine multiple weak text classifiers into a strong text classifier to filter out non-text regions in the graph. The detection speed of this method is significantly faster than other algorithms, but the accuracy of text detection is not high. Based on the connected-field natural scene text detection method, Jain et al. [39] proposed to first decompose the image into connected domains that do not overlap each other by color clustering, and then connect them into text lines according to the size and shape of the connected domains and

the distance relationship. Finally, the non-text lines are filtered according to the geometric rules of the text area. However, this method requires a lot of artificial definition of parameters, and the effect is not ideal in complex scenes. Epshtein et al. [40] proposed the Stroke Width Transform (SWT). The method utilizes the characteristics that the width of the character strokes is substantially the same, forms a stroke width map corresponding to the original image, and combines geometric reasoning to recover the original shape of the text, and extracts text of different scales in the complex background image, but does not apply to low resolution. Rate or occluded image.

At present, the mainstream target detection algorithms use the anchor mechanism. The anchor-based methods are representative of CTPN, SegLink, and TextBoxes++.

Zhi Tian, Weilin Huang et al. [16] proposed the link-based text proposal network CTPN, which can accurately locate the text lines in natural images, and innovatively developed a vertical anchoring mechanism, which greatly improved the accuracy of positioning. The advantage of CTPN is that the detected borders are accurate at 4 points above and below, and can detect long text. However, CTPN is only suitable for detecting horizontal text, not for oblique text and vertical text, and CTPN involves the merger of anchors. When to merge when it is not easy to master, post processing is complicated.

SegLink is also a detection algorithm proposed by Baoguang Shi, Serge Belongie et al. [18] in the past two years. It is a detection algorithm that can detect text at any angle and long text. For better retrieval of multi-angle texts, Baoguang Shi, Serge Belongie et al. proposed to learn a rotation parameter based on the original.

TextBoxes++ is based on the improvement of TextBoxes. TextBoxes can directly predict different scales of text, and use non-maximum suppression algorithm (NMS) [20] to process candidate text boxes to get the final text detection results, but the processing results are not ideal. In this regard, Yu Wei, Lu Yue et al. [21] proposed an improved TextBoxes network and improved non-maximum suppression algorithm. For the incomplete detection of non-maximum suppression algorithms, a text box fusion algorithm (Text-BBF) was proposed. The algorithm utilizes the positional relationship of the neighborhood candidate frames to fuse the candidate frames, remove redundant text boxes, and obtain more accurate text position information to improve text detection performance. But the TextBoxes algorithm can't detect multi-angle text in complex scenes, and TextBoxes++ is built on an end-to-end full convolutional network that can detect text in any direction. Li Weichong et al. [22] proposed an end-to-end trainable multi-directional scene image text recognition method (ie EX-TextBoxes++). Compared to two-stage text recognition, the proposed improvement shares convolution features between detection and recognition, and improves image text detection through multitasking.

The representative methods of pixel-based are EAST, FOTS and PSENNet.

There are multiple stages in text detection. For the usual detection algorithms [24, 25, 26], there are usually multiple stages such as candidate frame extraction, candidate frame filtering, bounding box regression, and candidate frame merging, and the text detection is divided into multiple stages. This increases the loss of text detection accuracy and running time to a certain extent, and thus Xinyu Zhou, Cong Yao et al. [23] proposed a new method to detect multi-angle texts in a simple, effective and accurate manner. The algorithm is called EAST. In order to realize end-to-end text detection, it is only divided into FCN to generate text line parameter stage and local perceptual non-maximum value suppression NMS stage. The simplicity of network model makes the speed and accuracy of text detection greatly improved. For a similar end-to-end trainable fast directional network, Liu Xuebo et al. [28] proposed a FOTS network for simultaneous text detection and recognition, sharing computational and visual information. FOTS has high performance, but it is not easy to detect long texts like EAST.

Since text detection based on bounding box is difficult to find text of any shape in an image, most pixel-based segmentation detectors cannot separate text instances that are very close to each other. Li Xiang, Wang Wenhai et al. [29] proposed a novel progressive extended network (PSENet). PSENet can detect long text and curved text, but post processing is more complicated. Li Xiang, Wang

Wenhai and others also proposed a progressive scale expansion algorithm, the idea comes from the breadth-first search (BFS) algorithm.

The combination of Pixel and anchor detection algorithm combines the advantages of pixel-based method and anchor-based method, and adopts FPN+ASSP mechanism to improve the network's receptive field. The combined model absorption advantages can solve texts of various scales and angles to a certain extent, but the anchor-based branch relies on strong manual setting anchors, and the anchor-based disadvantages may not be able to utilize the pixel-based method. Make up for the text that is back and cannot detect the curved arrangement.

Different detection methods have their own characteristics and deficiencies. Based on different detection algorithms and methods, as shown in Table 1, it is important to select detection methods according to different data sets.

Table 1. Detection algorithm comparison table

Category	**Algorithm**	**Advantage**	**Disadvantage**
anchor-based	CTPN	Detectable long text	Only detect horizontal text
	SegLink	Detects long text, multi-angle text	Post-detection is more complicated
	TextBoxes++	Can handle multi-angle text	Can only detect medium and small text
pixel-based	EAST	Can handle multi-angle text, simple structure	Poor detection of long text is not effective
	FOTS	Combined with detection and identification, high performance	Poor detection of long text is not effective
	PSENet	Detect long text, curve arrange text	Post-detection is more complicated
pixel-anchor	pixel-anchor	Text that can detect different scales and angles	Unable to detect curved text
Traditional method	Natural scene text detection method based on sliding window and natural scene text detection method based on connected domain	Fast running speed, suitable for fixed simple scenarios, low hardware configuration requirements	Cannot be applied to complex scenes

3.3 Text Recognition Technology

After the image is pre-processed and text-marked or detected, the input image can be input into the module for text recognition, thereby obtaining text information into the image. Tesseract-OCR, CRNN+CTC, and CTC+ Attention are all popular text recognition technologies.

Tesseract-OCR is a recognition engine developed by Google. It is an open source optical character recognition engine with high performance. This engine not only recognizes printed characters, but also recognizes standard handwritten characters, and has a certain ability of self-learning [51,52]. After training through the standard training set, the trained characters can be recognized more accurately. The Tesseract algorithm [41] is mainly divided into the following parts:

contour analysis, text line cutting, text block single character cutting, text feature extraction and first recognition, character recognition for misrecognition, language analysis, and so on. At present, Tesseract can support English, French, Italian, German, Spanish, Portuguese, Dutch recognition, and support ASCII, UTF-8 and other coding methods. Tesseract has many shortcomings. Tesseract requires that the input pictures have no complex background, horizontal text lines, space character spacing are basically consistent, and the computation speed is slow. It does not support multiple threads, nor does it support multiple instances in the same thread.

CRNN network [42] + CTC model, no need to cut the characters, you can directly identify the sequence characters, get the recognition results. The two-dimensional LSTM [45] is cascaded to form a deep structure. CTC is characterized by the introduction of the blank character, which solves the problem of no characters in some locations.

Attention-CTC model is a multi-task learning decoder [19] based on Attention Mechanism (AM) [54-55] and Connectionist Temporal Classification (CTC) [56] trained by Attention-CTC. It is a natural scene text algorithm proposed by Wenjie, Liu Jingbiao and others [17]. The new model is robust to text images with complex background and blurred fonts. It solves the problem that the long text sequence cannot be effectively predicted, is sensitive to noise, and is misaligned when decoding.

4. Summary

Natural scene detection has a wide range of applications in the fields of artificial intelligence, image and video processing. OCR was originally limited by traditional models and could not be applied to the detection and identification of complex scenes. With the rapid development of deep learning, the further development of OCR has been promoted, breaking through the limitations of traditional models. This paper mainly summarizes the overall process of OCR text recognition, and explains and analyzes image preprocessing, text line location detection algorithm and text recognition technology in detail.

From the current stage, there are still many issues worthy of further investigation for OCR text recognition. There are several issues that need to be studied in depth in the future:

Rich typographic style, complex background, image shooting angle, exposure, occlusion, text distortion and distortion will greatly affect the accuracy of image recognition. Currently, image preprocessing plays a vital role. Image preprocessing is the beginning of all recognition.

There are many types of text line location detection algorithms, but the inadequacies of each algorithm are also obvious. It is impossible to simultaneously recognize multi-angle long text, curve text and long text in complex natural scenes.

Therefore, we should strengthen the research of pretreatment technology to improve the restoring degree of image, to better carry out subsequent recognition. Although the combination of Pixel-anchor algorithm consumes longer time and cannot detect bending text, it improves the range requirements of text size, so we should try to improve and combine the detection algorithm to optimize the algorithm.

Acknowledgments

Our thanks to the Inner Mongolia Natural Science Foundation: 2014MS0614 for our support.

References

[1] M. Liao, B. Shi, X. Bai, X. Wang, W. Liu. Textboxes++: A Single-Shot Oriented Scene Text Detector. In IEEE Transactions on Image Processing 27(2018)3676-3690.

[2] Zhou X., Yao C., Wen H., Wang Y., Zhou S., He W., Liang J. East: An Efficient and Accurate Scene Text Detector. arXiv Preprint:1704.03155,2017.

[3] He W., Zhang, X. Y., Yin F., Liu C.L. Deep Direct Regression for Multi-Oriented Scene Text Detection. arXiv Preprint:1703.08289, 2017b.

[4] Shi B., Bai X., Belongie S. 2017. Detecting oriented text in natural images by linking segments. arXiv preprint arXiv:1703.06520.

[5] Qiangpeng Yang, Mengli Cheng, Wenmeng Zhou. IncepText: A New Inception-Text Module with Deformable PSROI Pooling for Multi-Oriented Scene Text Detection. In arXiv: 1805. 01167 [cs.CV].

[6] P. He, W. Huang, T. He, Q. Zhu, Y. Qiao, X. Li. Single Shot Text Detector with Regional Attention. arXiv Preprint:1709.00138,2017.6, 7.

[7] M. Jaderberg, K. Simonyan, A. Vedaldi, A. Zisserman. Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition. arXiv preprint:1406.2227,2014.2,3,6,7.

[8] Zhao Mengchao, Kong Lingcheng, Tan Zhiying. Segmentation of Coated Metal Strip Defects Based on Improved Otsu Method[J]. Computer Engineering and Design,2018,39(9):2811-2816.

[9] Yuan Jian, Cheng Guotao. Fast Otsu Image Segmentation Algorithm Based on Double Slope Division of Two-Dimensional Histogram[J]. Computer Applied Research,2017,34(6):1905-1908.

[10] Lu Qiuju. An Improved Two-Dimensional Otsu Image Segmentation Algorithm[J]. Journal of Natural Science of Xiangtan University,2018,40(6):82-87.

[11] Peng Wei. Research on Image Segmentation Based on Improved Particle Swarm Optimization Algorithm and 3D Otsu[D]. Wuhan University of Technology,2015.

[12] Gao Jun, Zhang Weiyong, Han Jianghong. Hough Transform Based on Neural Network and Its Optical Implementation [J]. Journal of Electronics,1999,27(2):37- 39.

[13] Kesidis A.L., Papamarkos N. On the Inverse Hough Transform[J]. IEEE Transactions on Pattern Analysis andMachine Intelligence, 1999,21(12):1329- 1343.

[14] Liang Dong, Gao Wei, Fu Qizhong et al. Vehicle license plate location and reconstruction based on shape characteristics and inverse Hough transform[J]. Computer Applications, 2002, 22 (5): 43-44+47.

[15] Zhu Guiying, Zhang Ruilin. Circle Detection Method Based on Hough Transform [J]. Computer Engineering and Design, 2008,29(6): 1462-1464.

[16] Zhi Tian, Weilin Huang, Tong He, Pan He. Detecting Text in Natural Image with Connectionist Text Proposal Network.arXiv:1609.03605v1 [cs.CV],2016.

[17] He Wenjie, Liu Jingbiao, Pan Mian, Lv Shuaishuai. Text Recognition Algorithms for Natural Scene Based on Attention-CTC [J/OL]. Electronic Technology, 2019(12):1-5[2019-06-26]. http:// kns.cnki.net/kcms/detail/61.1291.TN.20190315.1003.018.html.

[18] Baoguang Shi, Xiang Bai, Serge Belongie. Detecting Oriented Text in Natural Images by Linking Segments.arXiv:1703.06520 [cs.CV],2017.

[19] Xu K, Li D, Cassimatis N, et al. LCANet: end-to-end lipreading with cascaded attention-CTC[C]. Xi'an: China Automatic Face & Gesture Recognition,2018.

[20] Neubeck A, Gool L V. Efficient Non-Maximum Suppression[C]. International Conference on Pattern Recognition, 2006:850-855.

[21] Yu Zheng, Lu Yue. Text Detection Algorithm for Natural Scenes Based on Improved TextBoxes [D]. East China Normal University,2018.

[22] Li Weichong. Research on Multi-Directional Scene Character Recognition Algorithms Based on Improved TextBoxes+[J]. Graphics and Images,2018.

[23] Xinyu Zhou, Cong Yao, He Wen. EAST: An Efficient and Accurate Scene Text Detector. arXiv: 1704.03155 [cs.CV].

[24] M. Busta, L. Neumann, and J. Matas. Fastext: Efficient unconstrained scene text detector. In Proc. of ICCV, 2015.

[25] S. Tian, Y. Pan, C. Huang, S. Lu, K. Yu, and C. L. Tan. Textflow: A unified text detection system in natural scene images. In Proc. of ICCV, 2015.

[26] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman. Reading Text in the Wild with Convolutional Neural Networks. International Journal of Computer Vision, 116(1):1–20, jan 2016.

[27] T. Novikova, O. Barinova, P. Kohli, and V. Lempitsky. Large-lexicon attribute-consistent text recognition in natural images. In Proc. of ECCV, 2012.

[28] Xuebo Liu, Ding Liang, Shi Yan. FOTS: Fast Oriented Text Spotting with a Unified Network.arXiv:1801.01671 [cs.CV],2018.

[29] Xiang Li, Wenhai Wang, Wenbo Hou. Shape Robust Text Detection with Progressive Scale Expansion Network.arXiv:1806.02559 [cs.CV],2018.

[30] Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Be-longie. Feature pyramid networks for object detection. In CVPR, 2017.

[31] Dai Jin. Research on Text Detection Method Based on MSER[D]. Tianjin Normal University, 2014.

[32] Kim K C, Byun H R, Song Y J, et al. Scene text extraction in natural scene images using hierarchical feature combining and verification [C]. International Conference on Pattern Recognition. 2004: 679-682.

[33] Chen D, Odobez J M, Bourlard H. Text detection and recognition in images and video frames[J]. Pattern Recognition, 2004, 37(3): 595-608.

[34] Babenko B, Belongie S. End-to-end scene text recognition[C]. IEEE International Conference on Computer Vision. 2012: 1457-1464.

[35] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection [C] IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005: 886-893.

[36] Ozuysal M, Fua P, Lepetit V. Fast Keypoint Recognition in Ten Lines of Code [C] IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2007: 1-8.

[37] Chen X, Yuille A L. Detecting and reading text in natural scenes [C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2004: 366-373.

[38] Viola P, Jones M. Fast and Robust Classification using Asymmetric AdaBoost and a Detector Cascade[J]. Advances in Neural Information Processing Systems. 2001, 14: 1311-1318.

[39] Jain A K, Yu B. Automatic text location in images and video frames[J]. Pattern Recognition, 1998, 31(12): 2055-2076.

[40] Epshtein B, Ofek E, Wexler Y. Detecting text in natural scenes with stroke width transform [C]. IEEE Conference on Computer Vision and Pattern Recognition. 2010: 2963 -2970.

[41] Yin Lei, Zhang Honggang. Design and Implementation of Character Recognition Application Based on Android Platform [D]. Beijing University of Posts and Telecommunications, 2015.

[42] B. Shi, X. Bai, and C. Yao, A end-to-end trainable neural network for image-based recognition and its application to scene text recognition. CoRR, 2015.

[43] S. L. Phung and A. Bouzerdoum, “A pyramidal neural network for visual pattern recognition,” IEEE Transactions on Neural Networks, vol. 18, no. 2, pp. 329–343, 2007.

[44] S. W. Lee, H. H. Song A new recurrent neural network architecture for visual pattern recognition IEEE Transactions 1997 8(2):331-340.

[45] A. Graves, A. Mohamed, and G. E. Hinton. Speech recognition with deep recurrent neural networks. In ICASSP, 2013.

[46] Wang Qi, Chen Linqiang, Liang Xu. Caption Extraction in Video. Computer Engineering and Application, 2012, 48(5): 177-178.

[47] Hu Erlei, Feng Rui. Image Retrieval System Based on Deep Learning. Computer System Application, 2017, 26(3): 8-19.

[48] Chen Li. Design and Implementation of License Plate Recognition System. Modern Electronic Technology, 2012, 35(15): 142-144.

[49] Application of Yan Hongwen, Deng Xuefeng. OTSU Algorithm in Image Segmentation [J]. Agricultural Development and Equipment,2018.

[50] Huang Dongmei, Sun Yiqi, He Yuwen. Nomially Optimized Otsu Segmentation Algorithm for Remote Sensing Information of Different Sizes[J]. Remote Sensing Information, 2014, 34 (1):7-14.

[51] Zhang Yang, Shen Peiyi. Design and Implementation of Tesseract Optical Character Recognition Application [D]. Xi'an University of Electronic Science and Technology, 2013.

[52] Wan Song, Zhu Juan. Research and Implementation of Business Card Recognition System Based on Tesseract-OCR [D]. South China University of Technology,2014.

[53] Yang Xining, Duan Jianmin, Gao Dezhi et al. Lane Detection Technology Based on Improved Hough Transform [J]. Computer measurement and Control,2010,18(2):292-294+298.

[54] Bahdanau D, Chorowski J, Serdyuk D, et al. End-to-end attention-based large vocabulary speech recognition[C]. Shanghai: The 41st IEEE International Conference on Acoustics, Speech and Signal Processing,2016.

[55] Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation[C]. Lisbon: Empirical Methods in Natural Language Processing, 2015.

[56] Graves A, Gomez F. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks[C]. Hong Kong: International Conference on Machine Learning,2006.