ITM Web of Conferences **70**, 03010 (2025) https://doi.org/10.1051/itmconf/20257003010
DAI 2024

Advancements and Challenges in Character Recognition: A Comparative Analysis of CNN and Deep Learning Approaches

Ximin Yang

Graduate School of Arts & Science, Shanghai New York University, 200124 Shanghai, China

Abstract. This paper provides a comprehensive review of character recognition technologies, focusing on the application of Convolutional Neural Networks (CNN) and deep learning methodologies. Through an analysis of three key studies, the research highlights the strengths and limitations of current approaches. Study by Zib emphasizes the challenges in segmenting and recognizing English characters using CNN, revealing the need for supplementary techniques to mitigate errors. Research by Nikitha explores the impact of increasing the dimensionality of analysis, demonstrating that higher dimensions improve accuracy but also extend training times. Similarly, work conducted by Pradeep shows that larger vector sizes enhance recognition accuracy but at the cost of greater computational resources. The collective findings suggest that while CNN and deep learning models have significantly advanced character recognition, there remains a need for enhanced segmentation techniques and a balanced approach to optimizing training efficiency and accuracy. Future research should focus on integrating supportive methods to improve segmentation and finding an optimal trade-off between variable complexity and computational efficiency, thereby advancing the practical application of character recognition systems across various domains.

1 Introduction

Written character recognition technology is a widely used technology of transferring written words into digital words. Such technology can be used for different languages including Chinese, English and other languages. There are two main types of recognition, one is “Online Character Recognition” [1], the other one is offline, or can call it “Optical Character Recognition” [2], since it is reading the text through a camera. This technology can be used in various situation. For example, when taking action to the money stored in a bank, a signature is needed; notes written by professors in classes need to be stored and read; when responding to messages, words were written on screens; when changing files from paper based to electronic based, or to create an electronic record of the historical discoveries, Character recognition is needed. It is important especially for Chinese since it is difficult to quickly find the right word through Pinyin because there are lots of words with the same

Corresponding author: xy2456@nyu.edu

Pinyin, but character recognition can directly transfer the input to the result writer want [3]. Through reviewing the development of character recognition, this article can provide an insight to freshman of this field.

According to “Off-line Recognition of Chinese Handwriting: From Isolated Character to Realistic Text” by Su Tonghua, optical character recognition began its development in 1950s, at first those systems were only able to recognize specific character such as numbers, further developments enable them to read letters but only in specific printed form, such as the product named International Business Machines Corporation (IBM) 1418 is only capable of identifying words printed in IBM407 format [4]. After 1960 there came an advancing period of the system, Chinese characters and English words became recognizable objects of the system, handwriting was also included. Aspet-Oriented Programming (ASPET) developed in 1971 and IBM1975 in 1975 were also enable to recognize printed papers in low quality. The 1980s is the golden age for Chinese character recognition, multiple databases for Chinese characters were developed by different organizations. From 1990s until now is the time which recognition systems become partially mature. Multiple data sets for handwriting were built in this period, helping with recognition in different languages such as IRONOFF for French and Hierarchical Consistency Learning (HCL) 2000 for Chinese. For English, recognition to a entire sentence in different fields began to emerge and different ways of recognizing sentences were developed [2]. There are several existing methods of character recognition and classification such as the vector machine, Hidden Markov Model (HMM), and backup propagation (BP) network according to Jiao Weiwei [5]. Special methods for recognizing characters that had been partially covered by pollution such as dirt or something else had also developed. The data used to test such method is with a little noise, and it produced a result of 94% recognition [6].

The primary objective of this study is to review existing research on character recognition and summarize the key concepts and background information. This study traces the development of character recognition technology, highlighting several algorithms as examples. The study delves into the names, structures, solutions, processes, and significance of these algorithms. It also compares various approaches, examining their impact on different datasets and identifying the fields that could benefit from these methods to drive advancements. Additionally, the study discusses the advantages and disadvantages of each algorithm, with a focus on their potential for future development. This study aims to provide newcomers to the field of character recognition with a comprehensive overview of its evolution, offering insights into both its historical context and future potential. By understanding the strengths and limitations of existing algorithms, readers can better appreciate how this technology has been applied globally and how it might continue to evolve. This knowledge will be valuable for those looking to contribute to the ongoing development of character recognition technologies.

2 Methodology

2.1 Dataset description and preprocessing

The most widely used datasets in character recognition research include the Identity and Access Management (IAM) online and offline handwriting databases, among others. These datasets are frequently utilized due to their comprehensive collection of handwritten characters from various languages and contributors. The IAM handwriting databases, for instance, consist of scanned documents containing handwritten text in multiple languages, contributed by a diverse group of individuals. These databases are highly regarded and widely used in studies focusing on handwriting recognition. In addition to these well-known datasets, researchers often compile their own datasets to suit specific research needs. For example, in a study conducted by Zin TT [7], the author utilized a self-collected dataset comprising handwritten samples from students ranging from kindergarten to primary school in the targeted countries. This approach allowed for a more tailored dataset that met the specific requirements of the research. In another study by Nikitha [8], the IAM online and offline handwriting databases were employed, with a focus on scanned documents featuring Latin scripts. This study specifically leveraged the diversity of handwriting styles captured in these datasets to enhance the robustness of the recognition models. Study conducted by Pradeep, on the other hand, included fifty datasets, each containing an English alphabet written by a different individual [9]. This study primarily focused on feature extraction, utilizing the diverse handwriting styles to test the ability of the model to accurately identify and classify characters. These examples highlight the diversity of datasets available for character recognition research, as well as the varying approaches researchers take to develop and test their models. By utilizing both established datasets like IAM and custom-collected samples, researchers can address specific challenges and advance the field of character recognition.

2.2 Methods

When preforming character recognition, the characters are scanned at the beginning. The raw image is not suitable for recognition, it will go through a serious of preprocessing steps, including noise reduction, normalization, compression and feature extraction. Each step reduces irrelevant information with characters and make the image clearer than before until it is suitable for recognition. After preprocessing, there are multiple measures available for assisting recognition, including deep learning, Hidden Markov Model, neuron networks [10]. Fig. 1 shows how all the steps line up.

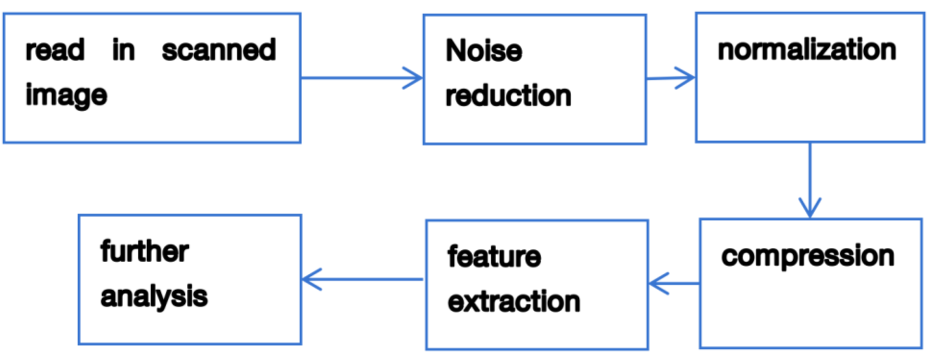

Fig. 1. How character recognition works (Picture credit: Original).

2.2.1 Technologies used by studies Convolutional Neural Networks (CNN)

CNN are similar to traditional Artificial Neuron Networks (ANNS) since both of them are made up with neuron nodes that are able to improve themselves through learning. Each node in the network will receive an input, preform a specific operation according to the setting by the author, and give an output. After several layers of nodes is the output. CNN are experts when preforming pattern recognition within images, while character recognition is a task that focuses on images. The main structure of a CNN is made up with three types of layers: convolutional layers, pooling layers and fully-connected layers. After the input layer read the pixel values of the image, they are feed into the convolutional ones. The convolutional layer will determine if the output of a neuron is connected to local region through a serious of calculation of its weight and the region connected to input value. Then the pooling layer will perform downsampling to reduce the number of parameters with in it. The fully-connected layer will produce class scores that are used for classification. CNN is used by Zin in feature extraction and classification. Deep learning is also used in the study conducted by Nikitha, while Pradeep chose to use neuron networks but not convolutional ones [11].

2.2.2 How written character recognition works

Handwriting of everyone is different. The paper with characters on it might be contaminated, making the scanning image full of noise. Before different algorithms can be applied to recognize the character, the data have to first go through several preprocessing measures to make it suitable for further action. There are three main objectives of preprocessing according to Arica: noise reduction, normalization of the data and compression in amount of information to be retained. Each objective requires different measures [10].

First the programme needs to get rid of noises. For noise reduction there are three different measures. When scanning the paper, things like dirt or ink which is dropped purposeless, will cause line segments, dots or loops other than the character needed to recognize. They will cause errors if get recognized as characters, so noises have to be removed before processing. There are hundreds of solutions to this that can be stored into three main categories. The first category is filtering. Measures in this category aims to remove noise and diminish spurious points introduced by uneven surface or by data collecting device. There exists a lot of different filters for this, all carries the basic idea of convolute a predefined mask to alter the image and eliminating the noises. The second category is morphological operations. They targets noised due to low quality of paper and/or ink and different hand movements that could cause the noise. These measures will make up the mistakes through different measures to make the image look like written with good handwriting using better material compared to raw data. The last category is noise modelling. If a model for a specific kind of noise is available, this kind of noise could be removed through using the model. However, it can do little with noises caused by sampling such as a blur image taken by a shaking hand, it is advised to take a clear image instead [12,13].

The second step is normalization. After dealing with noises, the data need to be normalized to remove the variations of the writing and obtain standardized data. Skew normalization and baseline extraction will help correct the curve caused by handwriting or the scanning process, making the characters line up. Baseline extraction can help find weather the two characters are in the same line or not, helping with skew normalization. The width, height and form of the characters are various considering different writer had different writing style and the requirements for the size of characters are also different to different task, so they need to be normalized into the same size and form. Slant normalization is used to deal with form while size normalization is used to deal with width and height. After these two steps, the characters will be in similar form with similar size, improving the efficiency of recognition.

Last step is compression. The data is in standard size and shape, but the images are in the format of space domains, they need to be translated into shape information for analysis. That is the job of compression. Two approaches are called thresholding and thinning, the former changes the image to binary image reducing the storage required, the latter extracts the shape of the characters [10].

3 Discussion of Results

3.1 Results analysis

According to the study conducted by Zin, his system manages to achieve an accuracy of more than 80% with all tasks, but there are some issues that the system needs further improvements. After comparing two-dimensional Long Short-Term Memory (2DLSTM) with 1DLSTM, Nikitha declared that the 2DLSTM method works better. In the first approach conducted by Pradeep, the feature vector size is chosen as 54, without row wise and column wise features. After 744 epochs the performance goal was reached. After included the row wise and column wise features, the feature vector size increased to 69 and it took 923 epochs to reach the performance goal same as the first approach. The accuracy of the first approach is 96.52% while the second one is 97.84%. The results are listed in Table 1.

Table 1. Result of different studies.

	Measure 1	Result 1	Measure 2	Result 2
Zin	Task 1 : Segmentation	accuracy : 98.5%	Task 2 : Recognition	Character accuracy :98.76% Word accuracy : 95.6%
Nikitha	2DLSTM	Character error rate : 8.2 Word error rate : 27.5	CNN-1DLSTM-CTC	Character error rate : 6.2 Word error rate : 20.5
Pradeep	Without row wise and column wise	vector: 54 epochs : 744 accuracy :96.52%	With row wise and column wise	vector : 69 epoches : 923 Accuracy : 97.84%

3.2 Discussions

It is evident that while these models have shown considerable promise, they also have certain limitations. However, by combining the strengths of different models, that can potentially achieve more robust and accurate results. For instance, the CNN model developed by Zin still requires further refinement to address issues related to misinterpretation of characters. Once the dissemination of letters is corrected, this model could serve as a highly effective feature extraction and segmentation tool, significantly enhancing the preprocessing stage. Study by Nikitha demonstrated that incorporating more vectors during the recognition process can lead to substantial improvements in the overall accuracy of the outcomes. This finding is echoed in study by Pradeep, which also observed that while introducing additional variables increases the training time, the resulting boost in accuracy justifies the extended duration. Furthermore, integrating these various approaches—such as combining the enhanced CNN model by Zin with the vector-rich methodology proposed by Nikitha—could lead to even greater accuracy and efficiency. This approach underscores the potential for hybrid models that leverage the strengths of multiple techniques, ultimately advancing the field of character recognition and addressing the current limitations in a more comprehensive manner. By continuing to explore these combinations, future research can focus on optimizing both performance and efficiency in character recognition systems.

4 Conclusion

This study reviews several previous research efforts in the field of character recognition, focusing on the various methods and techniques employed. Among these, CNNs and deep learning approaches were prominently featured across the three studies examined by the author. The study conducted by Zin targeted the recognition of English characters, highlighting the need for additional measures to improve the segmentation and recognition accuracy of CNN due to existing errors. Research by Nikitha explored the impact of analysis dimensions on the outcome, finding that increasing the dimensionality positively influences accuracy. Similarly, study by Pradeep investigated the effect of vector size on training results, revealing that larger vectors enhance accuracy, though at the cost of increased training time. The findings indicate that while CNN are powerful tools, they require supplementary techniques to address segmentation challenges. Additionally, while increasing dimensionality and vector size can boost accuracy, it also lengthens the training process. Future research should focus on developing supportive methods to enhance segmentation, as well as finding an optimal balance between the number of variables, training time, and accuracy. This balance is crucial to maximizing the efficiency and effectiveness of character recognition algorithms, paving the way for more robust and practical applications in the field.

References

1. X.J. Liu, Research on character root extraction algorithm in on-line handwritten Chinese character recognition, Ph.D. thesis, Huazhong University of Science and Technology (2008)
2. T.H. Su, Offline Chinese handwriting recognition - from isolated Chinese characters to real text, Ph.D. thesis, Harbin Institute of Technology (2008)
3. S. Wang, Market status and prospect analysis of handwriting technology, 9(2), 124-128 (2009)
4. E.C. Greanias, Some important factors in the practical utilization of optical character readers. Optical Character Recognition, 129-146 (1962)
5. W.W. Jiao, Research on off-line handwriting character recognition technology, Ph.D. thesis, Xinjiang University (2014)
6. W.Y. Sun, Recognition of defaced Chinese characters by finite sets, Ph.D. thesis, Nanjing University of Aeronautics and Astronautics (2008)
7. T.T. Zin, S. Thant, M.Z. Pwint, T. Ogino, Handwritten Character Recognition on Android for Basic Education Using Convolutional Neural Network. Electronics. 10(8), 904 (2021)
8. A. Nikitha, J. Geetha and D.S. JayaLakshmi, Handwritten Text Recognition using Deep Learning. International Conference on Recent Trends on Electronics, Information, Communication & Technology, 388-392 (2020)
9. J. Pradeep, E. Srinivasan and S. Himavathi, Diagonal based feature extraction for handwritten character recognition system using neural network, International Conference on Electronics Computer Technology, 364-368 (2011)
10. N. Arica and F.T. Yarman-Vural, An overview of character recognition focused on off-line handwriting, IEEE Transactions on Systems, Man, and Cybernetics, 31(2), 216-233 (2001)
11. K. O'Shea, R.A. Nash, Introduction to Convolutional Neural Networks, National Key Lab for Novel Software Technology (2015)

12. J. Serra, Morphological filtering: An overview, Signal Process, 38(1), 3-11 (1994)
13. M. Sonka, V. Hlavac and R. Boyle, Image Processing Analysis and Machine Vision, CA (1999)