Proceedings of the 2023 International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/12/20230322

English text recognition technology and application analysis based on optical character recognition

Shangwei Liu

Computer science, Boston University, Boston, MA, USA, 02215

swliu@bu.edu

Abstract. With the rapid development of artificial intelligence, it is gradually applied to more and more fields. Ai's market and technological potential value are unpredictable, which involves interaction between Ai and humans. In addition, optical character recognition is one of the essential roles in the interaction between machines and humans. Recognition of human language from a machine can significantly increase the user's experience and potential efficiency. Hence, this paper will focus on analyzing today's Optical character recognition (OCR) techniques based on feature extraction and neural networks. This article first provides an overview of the important significance of English text recognition, and then selects two typical methods for overview. Then, comparative experiments are conducted on the above methods and their analysis is conducted. By comparing each technique, to decide the restriction and advantage, based on the analysis to conclude and make predictions in OCR optimization.

Keywords: word recognition and transfer, computer vision, deep learning, feature extraction.

1. Introduction

Optical character recognition (OCR) is a technology that converts text documents into digital format that can be edited and searched electronically. It also can be implemented beyond the basic level which allows the software to analyze an image of a handwritten document. OCR is widely used in document management systems, electronic archives, and various other applications where conversion of printed or handwritten text to digital format is required [1]. Until today, this technology has benefited each user's life with OCR, and users can significantly increase their working experience. Furthermore, OCRs are widely used in phone scanning and text transfer. With development, optical character recognition can be used in the Artificial intelligence field to help improve human interaction, which can much more efficiently decrease human labor resources [2].

There are many algorithms that can complete text recognition tasks [3]. A functional optical character recognition program can have a nearly 94% accuracy rate. In some cases, words font, angle, direction, and lines can directly affect the function's accuracy. Therefore, it involves algorithms to help the system analyze each character, different investments algorithms have different recognition effects in different application scenarios [4]. Therefore, this article will analyze and compare several typical OCR algorithms, and then analyze their usage scenarios and advantages and disadvantages [5].

In this article, the beginning will focus on the structure of Optical character recognition, mainly analyzing optical character recognition based on feature extraction and neural networks. The body will include its pre-process, recognition process, feature extraction stage, and post-processing stage. The last part of the article will conclude with optical character recognition, provide comparison results between each technique, and emphasize the advantage of each method. Moreover, a comparative analysis will be conducted between the predicted content and the experimental results.

2. OCR application analysis

OCR systems are widely used in many fields due to their excellent functionality [6,7]. Some of these applications include:

1) Document digitization: According to Apple's recent release on their IOS 15 cell phone system, a new method, Visual Look Up was introduced. This method increases image-driven interfaces and allows users to perceive information from the physical world, and able transfer language words from images into digital versions format. This powerful method involves the optical character recognition technique as one of roll when archiving.

2) License plate recognition: Optical Character Recognition (OCR) systems play a crucial role in traffic surveillance systems by accurately identifying and interpreting vehicle license plates. This capability proves invaluable in enhancing traffic enforcement measures and facilitating efficient vehicle tracking processes. An illustrative example of the effective utilization of OCR technology can be observed in China, where the License Plate Recognition (LRP) system significantly benefits both the transportation sector and law enforcement agencies across multiple domains.

3) Assistive technology: OCR systems can be incorporated into assistive technologies for individuals with visual impairments, converting text from images or printed materials into speech or braille. With assistive OCR technology, individuals with visual impairments can take a picture of printed materials using their smartphone or a specialized device, and the OCR system will convert the text into speech or Braille output that can be read by the user. This technology can be used to access a wide range of printed materials, including books, documents, product labels, and menus.

3. Methods

3.1. ORC based on template matching

Template matching is a technique that involves comparing an input image to a set of predefined templates [8]. The input image is analyzed to identify regions that match each test template, and template with the best matching template is identified. Each template will be transformed into different characters or symbols in computer before identifying. Figure 1 shows the English word character A converted into digital form.

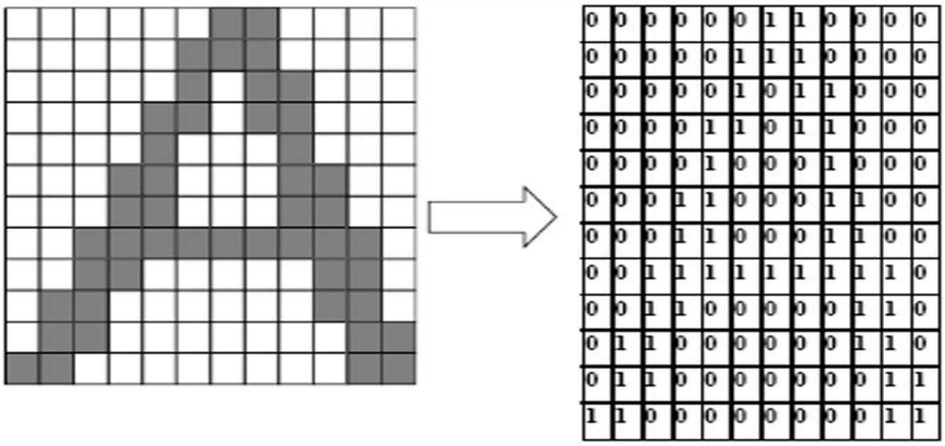

Figure 1. Image of character A in digital form [3].

As shown in Figure 1, the image in template matching extraction process is converted into 12*12-pixel points. The algorithm will find the pixel point if it's exact math to the template character's pixel points. In Optical character recognition by using template matching from Nadira Muda at university Malaysia Pahang researcher mentioned for recognition to occur, the input character set as $I(i, j)$ and

DOI: 10.54254/2755-2721/12/20230322

$Tn(i,j)$ as the template character image to recognize. By using corresponding equation can return a value that indicating the matching result as $S(I,Tn)$. The calculation equation used in City Block method is as shown in Eq. 1.

$$S(I,Tn) = \sum_{i=0}^{w} \sum_{j=0}^{k} I(i,j) - Tn(i,j) \tag{1}$$

There are many improved versions based on the principle of this formula, such as the Euclidean method and the Cross correlation method, whose calculation formulas are shown in Eq. 2 and Eq. 3.

$$S(I,Tn) = \sum_{i=0}^{w} \sum_{j=0}^{k} \left(I(i,j) - Tn(i,j)\right)^2 \tag{2}$$

$$S(I,Tn) = \sum_{i=0}^{w} \sum_{j=0}^{k} I(i,j)\, Tn(i,j) \tag{3}$$

Based on this technique character recognition is achieved by identifying. The function can provide the best matching value by only providing a set of high-quality templates, as well as the test templates. In addition, character's font, size, shape, and direction can affect function's accuracy on matching with template. Therefore, Template matching can be only used under certain conditions, and are constringed by the character's font etc.

3.2. ORC based on CNN

Feature extraction-based CNN for OCR involves using convolutional neural networks (CNNs) to extract relevant features from images of text and Once the features are extracted, they are fed into a classifier, which is trained to recognize specific patterns in the features. The classifier then uses this information to determine the most likely interpretation of the text in the input image. To train CNN, large datasets of labeled images are typically used. These datasets contain images of text that have been annotated with the correct labels, allowing the CNN to learn the features that are most useful for OCR [9,10].

Convolutional Neural Networks are a type of artificial neural network that is designed for image processing and computer vision tasks. The basic architecture of a CNN consists of three main layers. The convolutional layer applies a filter or kernel to the input image, which results in a feature map; The pooling layer reduces the dimensionality of the feature map., and the fully connected layer implement regression or classification tasks.

As previous part mentioned related to feature extraction, in addition, feature extraction is a crucial step in CNN which involves extracting meaningful features from an input image. The main goal of feature extraction is to capture the important characteristics of an image that are relevant to the task at hand. The quality of the features extracted from an image has a direct impact on the performance of network model. Figure 2 indicates the process about the CNN network.

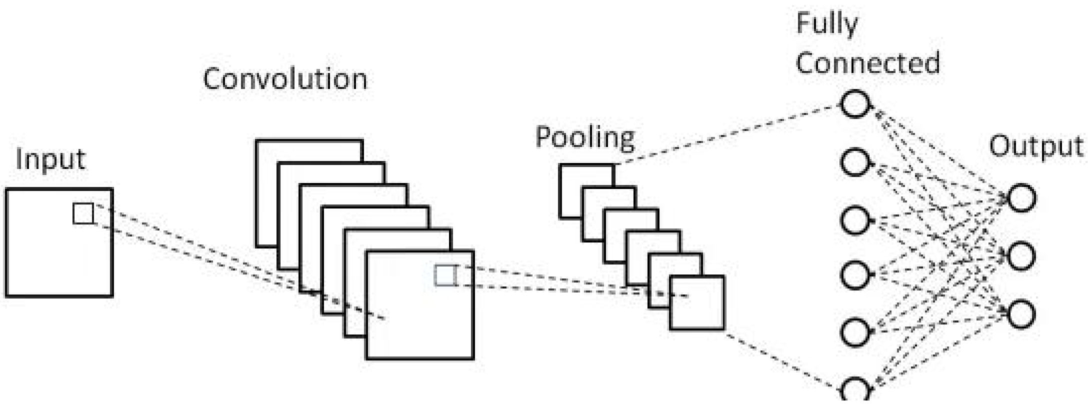

Figure 2. Flowchart of a convolutional neural network for recognition process.

DOI: 10.54254/2755-2721/12/20230322

3.2.1. Convolutional layer. The first layer structure for performing feature extraction in CNN is the convolutional layer. This layer extracts feature maps from the input image. The output feature map is then passed on to the next layer for further processing. Feature extraction-based optical character recognition techniques involve geometric morphometric characteristics. Geometric Morphometric Characteristics (GMC) refers to studying shape and size variations in biological and non-biological objects. In addition, GMC can be used to analyze the shape and size variations of English characters to improve the overall algorithm accuracy of character recognition. The features include Edge, Line, Direction, and component analysis. By feeding each feature vector to the convolutional neural network can establish an accurate prediction of the result [4]. Edge detection is one of the techniques in feature extraction. Edge detection can identify boundary features between different key regions in an image and distinguish characters from each other. In CNNS, edge detection is typically performed using a convolutional layer that applies a set of filter box to scan input image and scan for the corner of the character in image. However, only edge detection can affect the accuracy rate of the recognition, accuracy rate can be increased by combining several Feature detections.

3.2.2. Pooling layer. In a CNN network, the pooling layer serves to reduce the dimensionality of feature maps produced by the convolutional layer. By applying pooling operations, such as L2 pooling, which calculates the root mean square value of a set of values in the feature map, the overall size of the feature maps is decreased. This reduction helps to extract the most significant information while discarding unnecessary details, ultimately improving the efficiency and robustness of the network.

3.2.3. Activation function. The activation function plays a crucial role in introducing nonlinearity to the convolutional layer. Common activation functions like sigmoid and tanh are used, but the rectified linear unit (ReLU) is widely favored due to its simplicity and its ability to address the vanishing gradient problem. ReLU allows positive values to pass through unchanged, ensuring efficient gradient flow during backpropagation, thereby enabling the CNN to effectively learn complex patterns and relationships in the data.

3.2.4. Batch normalization. Batch normalization is a technique used to normalize the inputs of each layer in a CNN. By subtracting the mean and dividing by the standard deviation of the inputs within a batch, batch normalization mitigates internal covariate bias. This normalization process improves network stability, enables faster convergence, and enhances generalization performance by reducing dependencies on specific weight initializations.

3.2.5. Transfer learning. Transfer learning can use the pre-trained network model for feature extraction. The pre-trained model is trained on a large dataset, and can be used for a variety of computer vision tasks. Transfer learning helps to reduce the amount of training data required for a new task and can improve the performance of the CNN model. Overall feature extraction is a critical step in CNN that plays a vital role in many tasks. The quality of the features extracted from an input image has a direct impact on CNN model performance. This paper discussed the basic concept of CNN, the role of feature extraction in CNN, and the different methods for feature extraction in CNN.

4. Experiments

4.1. Experiment results based template matching

Template matching involves finding the location of a sub-image. Once the number of corresponding templates matches, these numbers can be used to determine the result. Hence, template matching requires very few similarities between a given set of templates and an input image. However, when matching the same fonts, the accuracy rate can reach a higher number in the range of 98% to 100%. Figure 3 indicates the accuracy rate of template matching in different fonts.

DOI: 10.54254/2755-2721/12/20230322

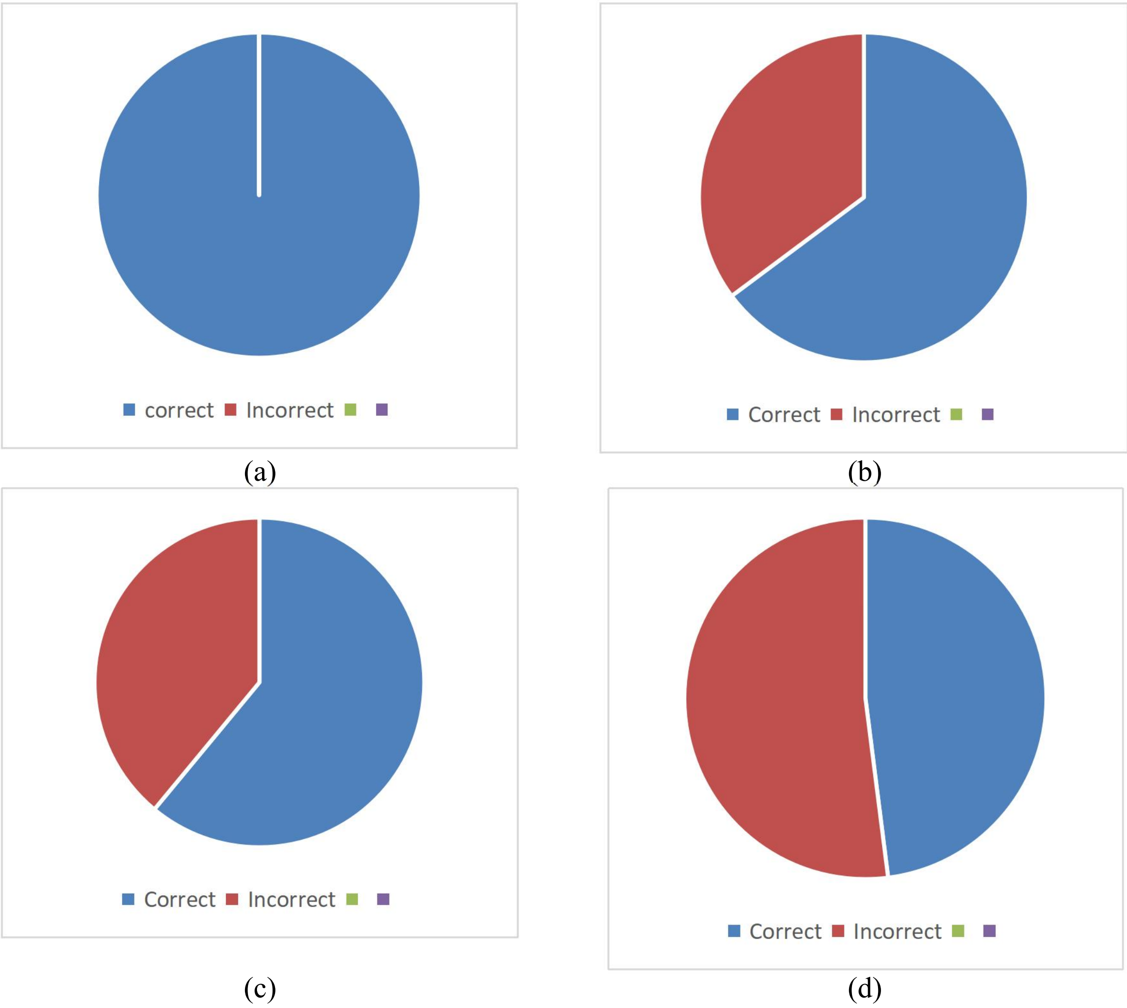

Figure 3. Comparison of accuracy in matching different font templates. (a) is the experimental result of Calibri font. (b) is the experimental result of Arial font. (c) is the experimental result of Times New Roman font. (d) is the experimental result of Cambria font.

Based on the results which created on experiment related to template matching OCR system. In the test experience, this part tested 100 words from each font. There are significant drawbacks due to variety fonts, which affect the template matching accuracy when the font of the template has low similarity to the input characters.

4.2. *Experiment result based feature extraction*

Feature extraction algorithms are to search for essential and unique components from a single character. Algorithms will determine based on the edge, corner, lines, and shape of the character. Combining these feature vectors increases the chance when classified each character. However, feature extraction requires large amounts of data set to study in a Convolutional neural network, with more data and templates the machine can study, the more accurate result will output in the final stage. Moreover, due to the methodology feature extraction applied, having different fonts or sizes will not affect the overall accuracy rate. The data set tested in feature extraction OCR program were 2000 handwritten English character, each image was converted in 8*8-pixel points. However, in the classify section which implement based on feature extraction seem to have relatively low accuracy in recognition with the highest in the program accuracy of 77.4%, Although with relatively poor accuracy result compared to Template matching. Throughout the experiment, it can be noticed that the poor accuracy were based on the amount of data set were fed to each layer. In other words, accuracy could highly increase if more data or template set were given. Therefore, Feature extraction is an accurate algorithm to implement when facing large sets of data or template sets when comparing with template matching.

5. Conclusion

Based on this research, it can be concluded that the most suitable algorithm for an optical character recognition program. Even Template matching has several benefits on the time and hardware consumption, however, template matching algorithms are more suitable for small organizations dealing with small or fewer amounts of data sets. When a program serves a large various number of clients, different fonts, shapes, or even signatures. Feature extraction can be done much more accurately than Template matching. In general, by studying more feature vectors from a character, the program can classify each character based on the set of feature vectors learned from previous procedures. In general, the accuracy rate based on feature extraction can be even further increased with more data sets that are inputted to the program.

Hence, Feature extraction is the most suitable program compared to template matching. The latter technique is also one of the simplest and most efficient approaches in the OCR field. Therefore, the clear disadvantage of utilizing this method is that the accuracy of recognition is heavily reliant on how much the input image resembles the stored templates.

References

[1] Fabien Lauer, Ching Y. Suen, Gérard Bloch. A trainable feature extractor for handwritten digit recognition. Pattern Recognition, 2007, 40 (6), pp.1816-1824. 10.1016/j.patcog.2006.

[2] J. Park, E. Lee, Y. Kim, I. Kang, H. I. Koo and N. I. Cho, "Multi-Lingual Optical Character Recognition System Using the Reinforcement Learning of Character Segmenter," in IEEE Access, vol. 8, pp. 174437-174448, 2020.

[3] Nadira Muda, Nik Kamariah Nik Ismail, Siti Azami Abu Bakar, Jasni Mohamad Zain Fakulti Sistem Komputer & Kejuruteraan Perisian,"Optical Character Recognition By Using Template Matching", University Malaysia Pahang Karung Berkunci 12, 25000 Kuantan, Pahang.

[4] Torres-Velázquez, Maribel & Chen, Wei-Jie & Li, Xue & McMillan, Alan. (2020). Application and Construction of Deep Learning Networks in Medical Imaging. IEEE Transactions on Radiation and Plasma Medical Sciences. 1-1.

[5] Greg CouncilMarch 5, 2018. (n.d.). Using OCR: How accurate is your data? Transforming Data with Intelligence. Retrieved April 27, 2023.

[6] Research, P. S. F. K. (2022, June 12). Apple introduces powerful suite of image-based tools for collaboration. PSFK. Retrieved April 29, 2023.

[7] Survision, License Plate Recognition Cameras. (n.d.). What is license plate recognition? Survision. Retrieved April 29, 2023.

[8] Optical Character Recognition Systems. The American Foundation for the Blind. (n.d.). Retrieved April 29, 2023.

[9] T. Wiatowski and H. Bölcskei, "A Mathematical Theory of Deep Convolutional Neural Networks for Feature Extraction," in IEEE Transactions on Information Theory, vol. 64, no. 3, pp. 1845-1866, March 2018.

[10] Y. Chen, H. Jiang, C. Li, X. Jia and P. Ghamisi, "Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks," in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 10, pp. 6232-6251, Oct. 2016.