

**REVIEW**

Review of Optical Character Recognition for Power System Image Based on Artificial Intelligence Algorithm

Xun Zhang¹, Wanrong Bai¹ and Haoyang Cui^{2,*}

¹Gansu Electric Power Research Institute of State Grid Corporation of China, Lanzhou, China

²College of Electronic and Information Engineering, Shanghai University of Electric Power, Shanghai, China

*Corresponding Author: Haoyang Cui. Email: cuihy@shiep.edu.cn

Received: 17 November 2021 Accepted: 17 August 2022

ABSTRACT

Optical Character Recognition (OCR) refers to a technology that uses image processing technology and character recognition algorithms to identify characters on an image. This paper is a deep study on the recognition effect of OCR based on Artificial Intelligence (AI) algorithms, in which the different AI algorithms for OCR analysis are classified and reviewed. Firstly, the mechanisms and characteristics of artificial neural network-based OCR are summarized. Secondly, this paper explores machine learning-based OCR, and draws the conclusion that the algorithms available for this form of OCR are still in their infancy, with low generalization and fixed recognition errors, albeit with better recognition effect and higher recognition accuracy. Finally, this paper explores several of the latest algorithms such as deep learning and pattern recognition algorithms. This paper concludes that OCR requires algorithms with higher recognition accuracy.

KEYWORDS

Optical character recognition; artificial intelligence; power system image; artificial neural network; machine learning; deep learning

1 Introduction

Optical Character Recognition (OCR) [1,2] refers to a technology that uses image processing technology and character recognition algorithms to identify characters on an image file input in a computer, before storing the obtained information in a computer file. OCR research shares close connections with research in fields such as pattern recognition, Artificial Intelligence (AI), Machine Learning (ML), and biomedicine.

Unlike computers, human beings acquire image information in an intuitive way. To date, researchers have been grappling with the challenge of designing a computer which can acquire information from images and convert it into text in an accurate and intuitive manner. The learning logic of AI [3] does not realistically mimic the visualization of objects by human beings; instead, AI must follow a learning curve process of learning features, from simple to complicated. To date, research and development of AI in the field of image recognition have achieved good results. As a powerful technology used for implementing AI, deep learning is related to existing learning methods in ML. The advances in AI in recent years has led to a substantial improvement in the quality of life



for many people. Nowadays, deep learning is widely used, and considerable success has been achieved in fields such as character recognition, image reconstruction [4], image restoration [5], and machine translation.

Handwritten Text Recognition (HTR) [6,7] is one of the key research directions in OCR, and researchers have been constantly researching and exploring new directions for its development. In recent years, it has been possible for data in handwritten documents to be extracted by AI systems, and the development of HTR technology has gained new impetus, with higher potential value. Before the spread of electronic information technology, most documents were stored in paper form. To make full use of the information processing power of the computer today, document-based information must be entered into the computer before being processed. If manual entry is adopted, objective conditions such as human emotions and physical strength will have an impact on work efficiency. In addition, there may be errors in the storage of information. Computer recognition of handwritten characters is easier and less error-prone than manual recognition. First, paper documents or certificates are transformed into image information that can be used by computer via various techniques, and then OCR is used to convert them into text information that can be recognized by computer. The recognition of handwritten characters has a high practical value as it is able to reduce the workload of workers in many industries, for example, automated processing of tax forms and check in banks. Therefore, the use and popularity of this technology has increased the efficiency of staffs in handling their work [7].

There are two means by which handwritten characters can be recognized: offline recognition, and online recognition. In this paper, offline recognition of handwritten characters, which is more challenging, is selected as the study object. In the process of recognizing characters, attributes such as stroke information and trajectories can provide more clues for online recognition. The process of offline recognition can be completed only by extracting feature information from the image [3].

Several scholars have investigated and contributed significantly to developments in optical recognition technology (The current research status of the optical recognition technology field at home and abroad will be introduced in the next section). Character recognition usually refers to the recognition of characters that have been optically processed. OCR refers to the conversion of any handwritten or printed image into a data file that can be read and edited by a computer. The image is converted into text using computer-related technologies such as scanning paper articles or books, so as to improve work efficiency and text storage capabilities [3–5].

The first research patent for OCR was first published in Germany around 1930, while handwritten letter recognition originated in the 1950s [8]. Since then, there has been extensive research into online recognition of English character and number symbols, but progress has generally been limited. The continued popularity of newspapers and other media has spurred OCR research by scholars, especially in Europe and the United States. Since 1990, many text recognition technologies developed in these regions have been applied, and English text recognition technologies are now widely adopted in commercial and official settings in European and American countries [9]. Innovations relating to the performance of platform-based scanning equipment have also guided developments in technology related to character recognition.

The 1980s witnessed rapid advances in artificial neural network algorithm design. Jia et al. [10] proposed using backpropagation to improve weights for training of multi-layer neural networks, and then using Parallel Distributed Processing Theory; their method is referred to as the BP algorithm. The BP algorithm has been widely used in the field of pattern recognition. This algorithm feeds back the learning results to each layer of neurons in the network by learning errors via backpropagation,

and modifies the weight matrix according to the results, so as to achieve the purpose of learning. The BP algorithm solves the problem of difficult modification of the hidden layer weights of the multi-layer neural network, enabling the neural network to exhibit excellent computing capabilities. In addition, the successful application of the BP algorithm theoretically proved the feasibility of using neural networks to perform complex multi-classification of the feature space. During this period, a large number of scholars conducted in-depth studies on artificial neural network design. For example, Bishop [11] studied and analyzed artificial neural networks from the perspective of statistical pattern recognition.

In the 1990s, the emergence of Support Vector Machines (SVMs) provided a new research concept for the field of pattern recognition [12]. Unlike neural networks, an SVM transforms image features into high-dimension feature space, applies linear functions for feature learning, and optimizes the weight matrix by minimizing structural risks.

In recent years, the AI industry has developed rapidly, with a rapid growth in the number of applications of cutting-edge technologies in the field of AI to various fields of computer vision. As a representative of cutting-edge AI technologies, deep learning is widely used in the fields of computer vision and pattern recognition, speech and natural language processing, and has achieved far better performance than traditional methods. One of the advantages of deep learning is that it can automatically learn the optimal features for a specific task from a large amount of data and does not need to manually design features for a specific task [9]. Since the turn of the 21st century, the ongoing development of deep learning has promoted a new round of OCR development. In 2008, Torralba et al. [13] investigated Deep Belief Network (DBN) technologies in order to improve the effect of image classification. In addition, there exist many network models that can recognize characters, for example, deep perceptron, the deep Boltzmann machine [14], Recurrent Neural Network (RNN) [15], and deep Convolutional Neural Networks (CNN) [16,17]. Yang et al. obtained the LVRC of ImageNet with their AlexNet model [18], which achieved an accuracy rate of 84.7%. Since then, deep CNN has been favored by researchers and has been widely used and exerted its power in the field of image recognition. In August 2014, Microsoft Research Asia announced the results of its ICDAR-2013 test set in natural scenes at the International Conference on Pattern Recognition; the final recognition result reached an accuracy rate of 92.1% and a recall rate of 92.3%.

At present, the mainstream OCR algorithms mainly include artificial neural networks, ML, SVMs, deep learning, etc. However, each different algorithm has its own problems. Therefore, it is critical and necessary to analyse the role of AI algorithms with their different underlying principles in OCR applications.

The remainder of this paper is structured as follows:

[Section 2](#) describes the design of the artificial neural network-based OCR algorithm, and the principle of OCR; [Section 3](#) examines ML-based OCR algorithms; [Section 4](#) examines deep learning-based OCR algorithms; and, [Section 5](#) summarizes the findings of this review paper.

2 OCR Principle and Artificial Neural Network

2.1 OCR Principle

This paper aims to review the core algorithms required by OCR systems (characters include Arabic numerals and English uppercase and lowercase letters) suitable for most machine vision fields; it analyzes the feasibility and practicality of a series of algorithms by taking actual scientific research

needs as an example; and, it compares processing algorithms with strong anti-noise ability and good effect [19].

Printed or handwritten characters are converted into discrete digital images through the image acquisition system, that is, through the optical imaging system and digital acquisition system, and the resulting digital images are transferred to the computer for pattern recognition processing [20]. The optical imaging system includes optical lenses (conventional optical lens and self-focusing lens array lens) and image sensor chip (CMOS, CCD); and, the digital acquisition system mainly includes the digital image acquisition card [21].

The key image processing algorithm of OCR is shown in the Fig. 1 below.

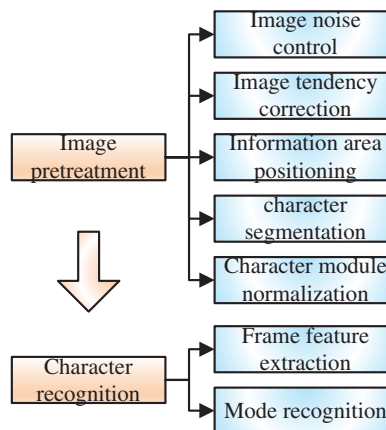


Figure 1: OCR diagram

The recognition process undertaken by the recognition algorithm can be roughly divided into the following three steps [22]:

- 1) The *preprocessing step*, which includes enhancing the quality of the input image, positioning the digital image through the image processing algorithm and intercepting the area of interest, and continuing segmenting and extracting the relevant structure on the region of interest according to the specific requirements of tasks;
- 2) The *feature extraction step*, that is, extracting the feature vector of the discretized character and digital image. The aim is to extract the feature vectors with a high degree of difference between characters; and,
- 3) The *pattern recognition step*, that is, inputting the extracted feature vector, and performing recognition and description using the pattern matching algorithm to correctly distinguish the characters, before finally completing the image processing task.

2.2 Neural Network Based OCR

With the rapid development of neural network technology, Convolutional Neural Networks (CNN) has unique advantages in image processing. This paper takes license plate character recognition as an example to discuss the effectiveness and limitations of OCR based on artificial neural network. Many studies have begun to use target detection networks such as SSD [13], Faster-RCNN [23], YOLO [24] to complete the license plate detection task. However, the prediction boxes of the above three kinds of multi-target detection neural networks are head-up rectangles. In actual situations, when the evaluated license plate image is rotated to a certain degree because the scene is manually

captured by traffic police or there is deviation of the camera, more background information may be selected. Therefore, Xie et al. [25] proposed a CNN-based MD-YOLO framework, which detects multi-directional rotating license plates by predicting accurate rotation angles and using a fast cross-evaluation strategy. The difference from YOLO is that this method not only uses coordinates and classification as the output of inference, but also adds tilt angle information for each Region of Interest (ROI). The method they proposed can cope well with the license plate rotation problem in real-time scene. The WPOD-NET framework proposed by Silva et al. [26] uses the common sense that a license plate corresponds to a specific vehicle. The framework first uses YOLOv2 to detect the specific vehicle, and then searches for the position of the license plate to be detected on the specific vehicle. During the process, the interference of background noise is avoided. In addition, the output of the license plate detection network is a transformation matrix from a square to a specific license plate shape, which effectively solves the problem of perspective and the tilt of the shooting angle and improves the accuracy of detection [27]. License plate pictures taken from surveillance cameras are often small and blurred due to the movement of cars and low resolution of cameras. Maglad [28] proposed a new hybrid cascade structure for quickly detection of small and fuzzy license plates in large and complex surveillance scenes. In order to quickly extract candidate license plates, they proposed two cascade detectors, including a pixel cascaded color space conversion detector and a contrast color Haar type detector. These two cascade detectors can perform detection from coarse to fine ones at the front and middle of the hybrid cascade. Finally, they proposed a cascaded convolutional network structure, which includes two Detection-ConvNets and a Calibration-ConvNet, aiming to carry out fine inspection.

The second step of the LPDR task is License Plate Recognition (LPR). The previous license plate recognition tasks are mostly to first segment the characters in the license plate, and then use Optical Character Recognition (OCR) technology to recognize the segmented characters. Because the color of the license plate image is monotonous and uniform, it can usually be binarized. The characters in the license plate can be segmented in various ways. For example, Hsu et al. [29] used Connected Component Analysis (CCA) method in segmenting the characters, in which the characters are segmented based on different connected domains. Gou et al. [30] used the specific area of specific characters to distinguish different license plate characters before completing the task of character segmentation. However, character segmentation becomes a challenging task due to the effect of brightness, shadows, and image noise of license plate images. Because once segmentation errors occur, single character recognition will be affected, and the final result will become inaccurate. With the development of neural networks, the LPR algorithm without character segmentation is proposed. In the framework proposed by Špaňhel et al. [31], the license plate recognition is based on the CNN structure, which processes the entire image as a whole, thereby avoiding the segmentation of license plate characters. Li et al. [32] believed that license plate recognition is a sequence task. Recurrent Neural Network (RNN) and Connectionist Temporal Classification (CTC) can be used to label character sequences. After extracting image features, CNN does not perform character segmentation. As a similar task, the conventional method of text recognition in natural scene images based on deep learning have been developed rapidly. Text recognition algorithms can provide inspiration and reference for license plate recognition. Although the results of applying RNN and CTC method to recognize license plate are good, the computing speed of RNN on embedded devices is slow. Zherzdev et al. [33] replaced the RNN network with CNN network and further reduced the number of parameters of the network to achieve real-time operation on the embedded license plate recognition task without reducing the recognition accuracy.

However, artificial neural network-based OCR is affected by the following problems:

- 1) Insufficient generalization of the neural network, which makes the recognition rate low and the recall rate of recognition high;
- 2) The instability and low robustness of the neural network, which leads to fixed recognition errors; and,
- 3) Extreme low adaptability and recognition rate of the artificial neural network, due to the number of layers of neural network being too small to enable complex correlations.

Therefore, an artificial neural network-based OCR algorithm cannot deliver a good OCR effect.

3 OCR Algorithm Based on Machine Learning

3.1 Recognition Principle Based on Machine Learning

Machine learning is a discipline that studies how to make computer programs simulate human brain for learning. It requires the computer to analyze the environmental data, summarize it, and build the corresponding knowledge base on its own like the human brains, so as to improve the performance of the algorithm based on the learning results. Machine learning can not only self-learn to improve the execution ability of a algorithm, but also summarize a large amount of data, which gives machine learning a strong advantage in solving large-scale classification and recognition problems with unclear internal characteristics of the image. The figure is the machine learning model proposed by Simon, a master in the field of artificial intelligence [34].

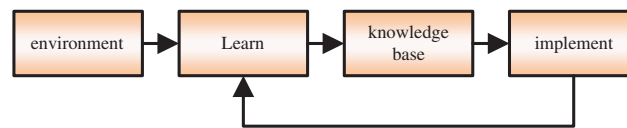


Figure 2: Machine learning flow chart

The process of machine learning is follows: the environment provides input data to the learner, and then the learning algorithm inside the learner receives the data, analyzes the data and converts it into knowledge; the knowledge base receives and stores the knowledge, and the data in the knowledge base has an impact on the subsequent data input; after the executor completes the task, it feeds back the result to the learner, and then the learner analyzes the result and reacts; the process continues until the learning task of the given goal is completed, that is, a learner completes its training, which is shown in Fig. 2.

In the learning system, in order to articulate knowledge the following principles need to be adhered to:

- (1) *Small computational complexity.* During the process of learning, the learner has to convert a large amount of environmental data into knowledge. If the conversion method is too complicated and requires a large number of calculations, it will be time-consuming and the overall performance of the algorithm will be affected. For example, when it is necessary to compare the two expressions to see if they are equivalent, if they are feature vectors, the judgment can be made via a quick and simple calculation of distance between the two vectors; whereas, if they are in first-order logic language, it is necessary to convert two relational expressions to the same expression form before comparing them, and this process will increase the number of calculations needed.

- (2) *Easy retrieval and modification.* The learning system needs to continuously update and modify the knowledge content in the knowledge base in the process of learning, and so the knowledge in the learning system should have a clear structure and organization, such as storage rules and feature vectors. In the process of updating the knowledge base, if the new knowledge required is identical to or contradicts the old knowledge, the old knowledge should be deleted. Therefore, in the process of expanding the knowledge base, the existing knowledge content should be checked and adjusted.
- (3) *Extensible knowledge base.* The basic requirement for a knowledge base is that it should express various kinds of knowledge; and, a further requirement is for the base to construct new knowledge on the basis of existing knowledge, so that it can expand the knowledge field and include more representative and more complex data, so as to improve the execution ability of the learning system. The extensibility of the knowledge base is a very important factor affecting the performance of large scale recognition systems.

Current mature ML algorithms include the Naive Bayesian algorithm, logistic regression algorithm, analogical learning, linear regression, neural network learning algorithm, SVM, nearest neighbor algorithm, etc. [35]. These algorithms can summarize knowledge and expand the knowledge base by finding rules between the sample data input and output for a given size of sample data. This kind of structural model, which can learn and generate certain knowledge content and classify it, is known as a classifier.

With ML models, researchers no longer need to spend a lot of time analyzing data in and determining rules for huge databases. Instead, they can determine how to extract more representative features, and develop classifiers and knowledge models with better performance, and enable the classifier to converge quickly. It is precisely because of this advantage that ML is able to replace traditional pattern matching and thus play a significant role both in industry and society.

3.2 OCR Algorithm Based on Machine Learning

Character recognition methods based on statistical classifiers mainly include SVM-based character recognition methods [36,37]. SVM is a machine learning method. Thanks to its excellent learning performance, the technology has quickly become a research hotspot in the machine learning industry. SVM has great advantages in solving small sample problems. The character recognition method based on SVM is divided into a training stage and a testing stage. The training stage is to extract features from the samples of training set, set SVM parameters and train the SVM classifier; the test stage is to extract the features of the samples in the test set, substitute the features into the trained SVM classifier, calculate the decision value, and judge the category of the sample based on the decision value.

Reference [37] proposed a character recognition algorithm combining KNN (K-Nearest Neighbor) algorithm and improved SVM. When this algorithm is used to test 1000-character samples, the recognition rate is 96.1%. In addition, the reference compared the recognition rates of several commonly used character recognition methods for the same 1000 character samples. Among them, the recognition rate of the Hausdorff distance-based template matching method was 93.5% and that of the BP neural network-based method was 93.2%. The experimental results showed that the recognition rate of character recognition algorithm combining KNN and the improved SVM was higher than both that of Hausdorff distance-based template matching method and that of the BP neural network-based method.

In the nonlinear SVM model, the common kernel functions are as follows:

(1) Linear kernel function

$$K(x_1, x_2) = (x_1, x_2) \quad (1)$$

where x_1 and x_2 is the arguments.

(2) Polynomial kernel function

$$K(x_1, x_2) = ((x_1, x_2) + R)^d \quad (2)$$

The dimension of this space is C_{p+d}^d , wherein, p is the original spatial dimension.

(3) Gaussian kernel function

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \quad (3)$$

When σ is too large, the weight of high-order feature attributes decays rapidly, and the value is still equivalent to a low-dimensional space; when σ is too small, any given data can be linear separable and is prone to overfitting.

(4) Sigmoid Kernel function

$$K(x_1, x_2) = \tanh(\gamma(x_1, x_2) + c) \quad (4)$$

where γ and c is the constants.

The advantages of the SVM algorithm are as follows:

- 1) It can be used for linear/non-linear classification and regression, with low generalization error rate, which means that it has good learning ability, and the learned results can be will generalized.
- 2) It can solve the machine leaning problem in the case of small samples and high-dimensional problems, and avoid the problems of neural network structure selection and local minima.
- 3) SVM is the best ready-made classifier. "Ready-made" means that it can be used directly without modification. In addition, there is a lower error rate, and SVM can make good classification decisions for data points outside the training set.

As ML algorithms, SVMs have the following disadvantages:

- 1) An SVM performs very well when processing small-scale samples as its generalization ability is outstanding among classification algorithms. However, when faced with large-scale high-dimension data, its performance is restricted.
- 2) It is sensitive to missing data.
- 3) The kernel support vector machine is a very powerful model that performs well on various data sets. An SVM allows a complex decision boundary, even if there are only a few data features. It performs well on both low-dimensional data and high-dimensional data (i.e., few features and many features), but it does not perform well when scaling the number of samples. It may perform well on data comprising up to 10,000 samples, but if the sample number approaches or exceeds 100,000, it may incur runtime and memory usage problems.
- 4) An SVM requires careful preprocessing of data and parameters adjustment. This is why, in many applications today, algorithms are using tree-based models such as random forests and the gradient-boosted tree (requiring little or no preprocessing). In addition, its processes and

prediction methods are so elaborate that any descriptions thereof cannot be conveyed to a non-expert audience.

4 OCR Algorithm Based on Deep Learning

4.1 Recognition Principle Based on Deep Learning

In 1980, inspired by the learning process of human brain neurons, Fukushima proposed a hierarchical neural network model. The model can recognize patterns by understanding the shape of objects. In the field of detection and recognition, Redmon et al. [24] applied a backpropagation algorithm and CNN to digital and handwriting recognition, and achieved good recognition results. With the rapid developments of hardware along with improvements in deep learning theory, deep neural networks have shown powerful feature expression capabilities and nonlinear mapping capabilities. There are many cases of successful deep learning-based OCR applications in various fields.

The basic elements that make up a neural network are weight ω , bias b , and structural feature of neural network. In order to make the neural network output the desired results, the network needs to be trained, so that the expected output value can be obtained in the process of continuously optimizing the parameters. During the training process, parameter optimization algorithms are used. Among them, the gradient descent method is the currently the most popular optimization algorithm and the most effective method to optimize the neural network. The core of the gradient descent algorithm is to minimize the objective loss function $f(\omega)$, wherein, ω is the weights to be optimized in the neural network. The initial learning rate is defined as α , and the gradient value g_t and momentum of the loss function are calculated. The calculation method is as follows:

$$g_t = \nabla f(w_t) \quad (5)$$

$$m_t = \phi(g_1, g_2, \dots, g_t) \quad (6)$$

$$V_t = \psi(g_1, g_2, \dots, g_t) \quad (7)$$

where m_t and V_t represent the first-order and second-order momentum of the gradient, respectively. Then, according to the gradient and momentum, the gradient loss η_t at the current time and the gradient ω_{t+1} at the next time are calculated:

$$\eta_t = \alpha \times \frac{m_t}{\sqrt{V_t}} \quad (8)$$

$$w_{t+1} = w_t - \eta_t \quad (9)$$

There are many improved algorithms based on the gradient descent method that enables the network to converge more quickly.

4.1.1 CNN

According to Fig. 3, Convolutional Neural Network (CNN) is one of the most widely used and most powerful network structures in deep learning. By adjusting the structure of the network, CNN can effectively solve all kinds of complex problems. All neurons between two adjacent layers of fully connected network are connected to each other. Through the nonlinear activation function, the fully connected network has a very strong fitting ability. However, when the fully connected network processes image data, the amount of parameters will increase by square as the size of the picture increases, which will increase the amount of calculation of the network, and prone to over-fitting. CNN is very similar to a fully connected layer network. Both of them are composed of neurons that

can learn weights and biases, but CNN uses the weight sharing of the convolution layer to make the number of parameters only related to the size of the convolution kernel, and uses the pooling layer to down sample the feature graph of each layer, so as to reduce the feature dimension and enable the model to have better generalization ability and faster inference speed.

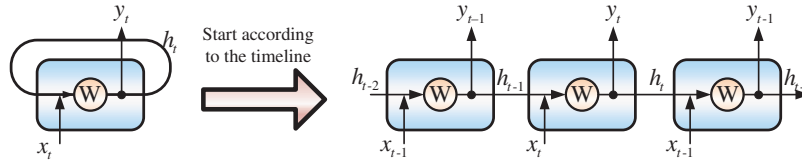


Figure 3: Convolution network diagram

4.1.2 RNN

According to Fig. 4, Recurrent Neural Network (RNN) originated in the 1980s. Hopfield first used a cyclically connected network structure to simulate human memory units. In 1990, Elman proposed a Simple Recurrent Network (SRN), which uses a backpropagation algorithm to train SRN. After that, Schmidhuber and Hochreiter proposed a Long and Short-Term Memory Network [26] (LSTM), which allows information to be stably transmitted in the network through a carefully constructed gating control unit, effectively prevents gradients elimination through back propagation, and effectively captures long-term dependencies. Therefore, LSTM is widely used in fields such as language models, speech recognition, translation conversion, and together with CNN has become the foundation stone of deep learning.

The basic structure of RNN has the characteristics of cyclic connection of hidden layer neurons, and that the output of the current moment is associated with the current input and the hidden state of the previous moment. Assuming the input signal is X, the cyclic dynamic system can be expressed as:

$$s_t = f(s_{t-1}, x_t) \tag{10}$$

The hidden state s at t moment needs to depend on the input at t moment and the hidden state at $t - 1$ moment. According to the characteristics of the circulatory system, the results at different moments can be expressed by the state at previous moment. For example, the state at $t = 3$ moment can be expressed as:

$$s_3 = f(s_2, x_3) = f(f(s_1, x_2), x_3) \tag{11}$$

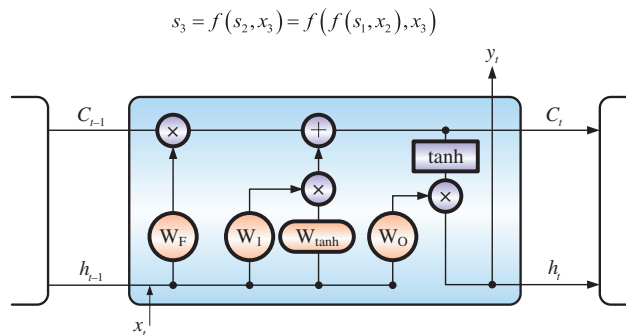


Figure 4: RNN diagram

4.2 OCR Principle Based on Deep Learning

The process of traditional text recognition methods includes image preprocessing-character segmentation-character recognition. Due to the complex background and different text forms, the character segmentation process is difficult, which will seriously restrict the accuracy of text recognition. In order to prevent character segmentation from causing error transmission, the text recognition methods based on deep learning are used to avoid the character segmentation process, which mainly includes the method based on the CTC [38,39] and the method based on the attention mechanism [40,41].

The CTC-based method assumes that each character is related to each other and not independent of each other. It calculates the conditional probability under the prediction result of the previous frame. The input during training and test output are both character label sequences, which requires no character alignment and character labeling. When observing an image, it will focus on and pay more attention to a certain part of the content of interest, which is the principle of the attention mechanism. Inspired by it, the method based on the attention mechanism determines the information that is beneficial to text recognition to reduce the amount of information need to be processed, simplify the task and improve the performance of text recognition.

Recurrent Neural Network (RNN) is often used to process sequence data. For example, Shi et al. [38] proposed an end-to-end CRNN model to jointly train CNN and RNN. It used convolutional layers to extract features, used recurrent layers to predict the label distribution of each frame, and used the interpretation layer (CTC layer) to output the label sequence. The CRNN model can recognize text images of different lengths, and does not require a pre-defined dictionary. It performs outstandingly in the text recognition in actual scenes. However, the RNN+CTC method requires a large amount of computation quality and have difficulties in training. Therefore, Gao et al. [39] used a stack network to obtain the contextual relevance of the input sequence, which is simpler and easier to achieve parallelism than the RNN network. After the introduction of the residual attention module, the recognition accuracy can be increased by about 2%, with a speed 9 times faster than the BLSTM model.

Lee et al. [40] proposed an unconstrained scene text recognition model integrating RNN and attention mechanism. It used RNN model to achieve fewer model parameters and extract image features more efficiently, and used soft attention mechanism to select and coordinate the image features extracted in the previous step. The experimental results show that the accuracy of the model is improved by 9% on the SVT data set, and it is improved by 8.2% on ICDAR2013. Liu et al. [41] proposed a binary encoding-decoding neural network (B-CED NeT) to convert text images into categories. The feature map reduces memory requirements and speeds up operation while preserving character category, space, and morphological information. The experimental results show that the running time of recognizing scene text and images with size a is only 4.59 ms on average.

However, existing deep learning-based OCR methods have two shortcomings:

- 1) *Optimization is difficult*. The design of existing methods regards OCR as two independent tasks-text detection and recognition-which therefore are independently optimized.
- 2) *Slow speed*. Existing methods cannot easily meet real-time requirements as the complicated process restricts the recognition speed to a certain extent.

In the field of deep learning, large-scale data sets are very important. However, due to the high cost of collecting data or the limited number of research objects, data sets often have problems such as insufficient data and poor diversity. For example, the collection of defective product image data sets in

the industrial field requires experienced professionals to set up machine vision system, with a high labor cost; In addition, due to the wide variety of defects, different shapes, and few numbers, it is impossible to obtain thousands or even millions of data. It is possible to obtain all kinds of defects. Insufficient data can easily lead to overfitting of the CNN model, and poor diversity reduces the generalization ability of the model. In order to solve the problems of insufficient data and poor diversity, scholars have proposed many data enhancement methods.

Therefore, a high-performance character recognition system is conducive to high-precision, real-time recognition.

5 Conclusion

It can reach the following conclusions:

- 1) In order to conduct an in-depth study on the recognition effect of OCR based on artificial intelligence algorithms, this paper classifies and analyzes OCR using different artificial intelligence algorithms. First of all, the mechanism and characteristics of OCR using artificial neural network are summarized, and the results show that the OCR based on the artificial neural network obviously has problems such as low recognition rate and unsatisfying accuracy. Then, this paper introduces the OCR based on machine learning, and concludes that the machine learning algorithms-based OCR is still in its infancy, with low generalization, and is prone to fixed recognition errors, though it has better recognition effect and higher recognition accuracy. Finally, this paper introduces a large number of the latest algorithms such as Deep learning and pattern recognition algorithms. This paper concludes that OCR requires algorithms with higher recognition accuracy.
- 2) This paper analyzes the application status of three types of artificial intelligence algorithms in OCR, including artificial neural network, machine learning, and deep learning. Among them, artificial neural network algorithm has the characteristics of rapid recognition, but has low generalization. The insufficient generalization of neural network results in low recognition rate and high recall rate of recognition. The SVM algorithm in the machine learning algorithm can avoid the structure selection of neural network and local minima. Its generalization ability is outstanding among the classification algorithms. However, when faced large-scale high-dimension data, its effect is not satisfying. The deep learning-based OCR method has strong adaptability and accurate recognition capabilities, but it has two shortcomings as follows: (1) Difficult optimization. The existing methods regard OCR as two independent tasks—text detection and recognition, to performed independent optimization; (2) Slow speed, which is difficult to meet the real-time requirements as the complex process restricts the recognition speed to a certain extent.
- 3) In the future, there is still a need for an OCR algorithm with accurate recognition capabilities and high robustness for high-precision and real-time recognition.

Funding Statement: This work was jointly supported by science and technology projects of Gansu State Grid Corporation of China (52272220002U).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
2. Nair, V., Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th International Conference on Machine Learning*, Haifa, Israel.
3. Yang, B., Zhong, L., Zhang, X., Shu, H., Yu, T. et al. (2019). Novel bio-inspired memetic salp swarm algorithm and application to MPPT for PV systems considering partial shading condition. *Journal of Cleaner Production*, 215(3), 1203–1222. DOI 10.1016/j.jclepro.2019.01.150.
4. Yang, B., Wang, J., Zhang, X., Yu, T., Yao, W. et al. (2020). Comprehensive overview of meta-heuristic algorithm applications on PV cell parameter identification. *Energy Conversion and Management*, 208(5), 112595. DOI 10.1016/j.enconman.2020.112595.
5. Yang, B., Yu, T., Zhang, X., Li, H., Shu, H. et al. (2019). Dynamic leader based collective intelligence for maximum power point tracking of PV systems affected by partial shading condition. *Energy Conversion and Management*, 179(18), 286–303. DOI 10.1016/j.enconman.2018.10.074.
6. Yang, B., Yu, T., Shu, H., Dong, J., Jiang, L. (2018). Robust sliding-mode control of wind energy conversion systems for optimal power extraction via nonlinear perturbation observers. *Applied Energy*, 210(1), 711–723. DOI 10.1016/j.apenergy.2017.08.027.
7. Hanmandlu, M., Murthy, O. V. R. (2007). Fuzzy model based recognition of handwritten numerals. *Pattern Recognition*, 40(6), 1840–1854. DOI 10.1016/j.patcog.2006.08.014.
8. Hinton, G. E., Osindero, S., Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7), 1527–1554. DOI 10.1162/neco.2006.18.7.1527.
9. Govindan, V. K., Shivaprasad, A. P. (1990). Character recognition—A review. *Pattern Recognition*, 23(7), 671–683. DOI 10.1016/0031-3203(90)90091-X.
10. Jia, Z. B., Tian, L. Y. (2006). Recognition algorithm using BP neural network for ID. *Journal of Suzhou Vocational University*, 17(3), 92–94.
11. Bishop, C. M. (1995). *Neural network for pattern recognition*. England: Oxford University Press.
12. Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2), 121–167. DOI 10.1023/A:1009715923555.
13. Torralba, A., Fergus, R., Weiss, Y. (2008). Small codes and large image databases for recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, USA.
14. Salakhutdinov, R., Hinton, G. (2009). Deep boltzmann machines. *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, Bellevue, USA.
15. Socher, R., Lin, C. C. Y., Ng, A. Y., Manning, C. D. (2011). Parsing natural scenes and natural language with recursive neural networks. *Proceedings of the 28th International Conference on Machine Learning*, Bellevue, WA, USA.
16. Huang, G., Liu, Z., van Der Maaten, L., Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, AK, USA.
17. Hu, G., Peng, X., Yang, Y., Hospedales, T. M., Verbeek, J. (2017). Frankenstein: Learning deep face representations using small data. *IEEE Transactions on Image Processing*, 27(1), 293–303. DOI 10.1109/TIP.2017.2756450.
18. Yang, B., Swe, T., Chen, Y., Zeng, C., Shu, H. et al. (2021). Energy cooperation between Myanmar and China under one belt one road: Current state, challenges and perspectives. *Energy*, 215(1), 119130. DOI 10.1016/j.energy.2020.119130.
19. Yang, B., Wang, J., Zhang, X., Yu, T., Yao, W. et al. (2020). Comprehensive overview of meta-heuristic algorithm applications on PV cell parameter identification. *Energy Conversion and Management*, 208(5), 112595. DOI 10.1016/j.enconman.2020.112595.

20. Ren, S., He, K., Girshick, R., Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 39(6), 1137–1149. DOI 10.1109/TPAMI.2016.2577031.
21. Wang, L., Ouyang, W., Wang, X., Lu, H. (2015). Visual tracking with fully convolutional networks. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Bellevue, USA.
22. Nam, H., Han, B. (2016). Learning multi-domain convolutional neural networks for visual tracking. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, AK, USA.
23. Ren, S., He, K., Girshick, R., Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28, 91–99.
24. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Bellevue, USA.
25. Xie, L., Ahmad, T., Jin, L., Liu, Y., Zhang, S. (2018). A new CNN-based method for multi-directional car license plate detection. *IEEE Transactions on Intelligent Transportation Systems*, 19(2), 507–517. DOI 10.1109/TITS.2017.2784093.
26. Silva, S. M., Jung, C. R. (2018). License plate detection and recognition in unconstrained scenarios. In: *Lecture notes in computer science*, vol. 8690. Cham, Springer.
27. Liu, C., Chang, F. (2018). Hybrid cascade structure for license plate detection in large visual surveillance scenes. *IEEE Transactions on Intelligent Transportation Systems*, 20(6), 2122–2135. DOI 10.1109/TITS.2018.2859348.
28. Maglad, K. W. (2012). A vehicle license plate detection and recognition system. *Journal of Computer Science*, 8(3), 310–315. DOI 10.3844/jcssp.2012.310.315.
29. Hsu, G. S., Chen, J. C., Chung, Y. Z. (2012). Application-oriented license plate recognition. *IEEE Transactions on Vehicular Technology*, 62(2), 552–561. DOI 10.1109/TVT.2012.2226218.
30. Gou, C., Wang, K., Yao, Y., Li, Z. (2015). Vehicle license plate recognition based on extremal regions and restricted Boltzmann machines. *IEEE Transactions on Intelligent Transportation Systems*, 17(4), 1096–1107. DOI 10.1109/TITS.2015.2496545.
31. Špaňhel, J., Sochor, J., Juránek, R., Herout, A., Maršík, L. et al. (2017). Holistic recognition of low quality license plates by CNN using track annotated data. *14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Lecce, Italy.
32. Li, H., Shen, C. (2016). Reading car license plates using deep convolutional neural networks and LSTMs. arXiv preprint arXiv:1601.05610.
33. Zherzdev, S., Gruzdev, A. (2018). LPRNet: License plate recognition via deep neural networks. arXiv preprint arXiv:1806.10447.
34. Cui, H., Huo, S., Ma, H., Guo, W., Ge, C. et al. (2018). Effects of view angle and measurement distance on electrical equipment UV corona discharge detection. *Optik*, 171, 672–677. DOI 10.1016/j.ijleo.2018.06.080.
35. Luger, G. F. (2005). *Artificial intelligence: structures and strategies for complex problem solving*. 4th ed. London: Pearson Education.
36. Wang, X., Wang, X. (2005). Improved approach based on SVM for license plate character recognition. *Journal of Beijing Institute of Technology*, 14(4), 378–381.
37. Capar, A., Gokmen, M. (2006). Concurrent segmentation and recognition with shape-driven fast marching methods. *18th International Conference on Pattern Recognition (ICPR'06)*, Hong Kong, China.
38. Shi, B., Bai, X., Yao, C. (2016). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(11), 2298–2304. DOI 10.1109/TPAMI.2016.2646371.

39. Gao, Y., Chen, Y., Wang, J., Lu, H. (2017). Reading scene text with attention convolutional sequence modeling. arXiv preprint arXiv: 1709.04303.
40. Lee, C. Y., Osindero, S. (2016). Recursive recurrent nets with attention modeling for OCR in the wild. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Bellevue, USA.
41. Liu, Z., Li, Y., Ren, F., Goh, W. L., Yu, H. (2018). Squeezedtext: A real-time scene text recognition by binary convolutional encoder-decoder network. *AAAI Conference on Artificial Intelligence*, New Orleans, USA.