


Analysis for Advancements of Optical Character Recognition in Handwriting Recognition

Yitao Yao ^a

School of International Education, Shanghai Normal University Tianhua College, Shanghai, China

Keywords: Handwriting Recognition, Hidden Markov Models, Convolutional Neural Networks, Recurrent Neural Networks.

Abstract: In the digital age, despite the widespread use of digital documents, many handwritten documents still need to be converted into digital formats. Optical Character Recognition (OCR) technology-based handwriting recognition addresses this need by converting printed or handwritten text into machine-readable form, improving work efficiency. This paper examines key OCR technologies, including Hidden Markov Models (HMM), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) with Long Short-Term Memory (LSTM). The methodology section discusses how HMMs use probabilistic models to recognize text in noisy environments, while CNNs automatically extract features from images. RNNs and LSTMs capture temporal dependencies and context in sequential data, making them effective for recognizing continuous characters and complex text structures. The paper also explores the combination of CNNs with LSTMs for end-to-end text recognition, further enhancing OCR capabilities. The discussion highlights the strengths and limitations of these technologies: HMMs are efficient but limited in expressive power, CNNs excel in feature extraction but require large datasets, and LSTMs handle long-term dependencies well but are computationally intensive. Despite advancements, OCR still faces challenges. This paper offers a comprehensive overview of key models in OCR technology, guiding future research in selecting suitable models for specific tasks and improving accuracy and efficiency.


1 INTRODUCTION

In the digital age, although digital documents are widely used, there are still many handwritten documents that need to be converted into digital formats. These documents include a large number of historical records, handwritten notes, contracts, and invoices. Optical Character Recognition (OCR) is a technology that converts printed or handwritten documents into machine-readable text and is considered an effective tool for improving work efficiency and achieving systematic document management.

The early conceptualization of OCR began with patents submitted by Tausheck and Handel (Mori, 1992). These patents laid the foundation for template/mask matching methods. The earliest OCR technology emerged in the 1950s and was designed to recognize printed text. These early systems used template matching, which involved comparing

characters with predefined templates. In the 1960s and 1970s, OCR technology began to rely on manually designed feature extraction methods, such as analyzing the distribution of horizontal and vertical lines and the contours of character shapes (Mori, 1992). While these methods were effective for recognizing specific fonts, they lacked adaptability to diverse fonts and handwritten text.

From the 1980s to 2000, OCR systems shifted from hardware-based implementations to software-based solutions capable of handling both printed and handwritten text. During the 1980s, Hidden Markov Models (HMM) were introduced to OCR, particularly for handwriting recognition. HMM allowed for the modeling of sequential data, enabling the processing of continuous handwritten text. In the 1990s, with increased computational power, neural networks began to be used in OCR. Early neural networks, such as Multilayer Perceptrons (MLP), could learn more features, improving recognition accuracy. After 2010,

^a <https://orcid.org/0009-0007-8685-0186>

the rise of Convolutional Neural Networks (CNN), introduced by Yann LeCun in the late 1980s, revolutionized the OCR field. CNNs use hierarchical feature extraction to automatically learn rich character features from data, significantly enhancing text recognition accuracy (Memon, 2020).

Later, the introduction of Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) further advanced OCR. RNNs are primarily used for processing and recognizing sequential data. Unlike traditional OCR methods, RNNs capture the temporal relationships and contextual information within input data, making them particularly effective for text recognition, especially handwriting (Memon, 2020). LSTM, a special type of RNN used for processing and predicting sequential data, gained widespread attention for its ability to address the gradient vanishing and exploding problems in traditional RNNs (Namysl, 2019). In the OCR field, LSTM is used to improve text recognition accuracy, especially for variable-length sequences or when strong contextual dependencies exist (Namysl, 2019). The application of LSTM in OCR began around 2015 when many research teams started combining it with CNNs to form end-to-end text recognition systems. Google's Tesseract 4.0 integrated LSTM, further enhancing its performance in complex text and multilingual environments. Overall, the introduction of LSTM has made OCR systems more proficient at handling complex text structures, contextual dependencies, and variable-length inputs, marking a significant shift from traditional handcrafted feature extraction methods to more intelligent and automated approaches. The evolution of OCR technology reflects a transition from traditional feature extraction to deep learning, and with ongoing improvements in computing power and algorithms, future OCR technology will be able to handle increasingly complex and diverse text scenarios. Due to the rapid development and significance of OCR technologies, it is necessary to conduct a comprehensive review in this domain.

The remainder of the paper is organized as follows. First, in Section 2, this article will introduce three key technologies in the history of OCR development: HMM, CNN, and RNN LSTM. In Section 3, the advantages and disadvantages of these technologies as well as the challenges and future prospects in this field will be analyzed. Finally, this paper will summarize the whole paper.

2 METHOD

2.1 The Introduction of OCR Recognition Workflow

The process of extracting text from documents or images into machine-readable text using OCR technology involves several key steps (Karthick, 2019). Figure 1 illustrates the OCR recognition workflow. The process starts with obtaining the image containing the handwritten text. This could range from scanned documents, photographs, or any image containing textual information. And then the image is adjusted through normalization, skew correction, noise reduction and binarization during preprocessing step. These steps ensure the image is clear and properly formatted for recognition. After that, the image is segmented into lines, words and characters. Old OCR systems may still use pattern matching to compare the extracted features against known character templates to recognize the characters. Modern OCR systems use deep learning models, like CNN or RNN for accurate character recognition. Then recognized characters are assembled into words and sentences, with contextual analysis and error correction applied to improve accuracy. Finally, the recognized text is output in the desired format for further use.

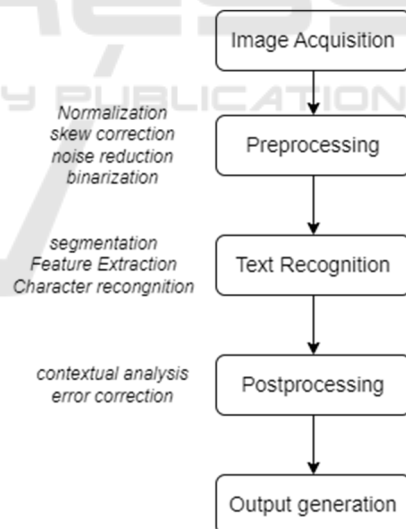


Figure 1: OCR recognition workflow (Photo/Picture credit: Original).

2.2 HMM

Lawrence et al. pioneered the use of HMMs for tasks like OCR, especially for recognizing handwritten text (Rabiner, 1989). HMMs model sequential data by representing a character sequence as hidden states,

with each state corresponding to a character and transitions between states indicating the probability of moving from one character to the next. HMMs rely on two key principles: state transitions and observation probabilities. This approach allows HMMs to handle noisy or unclear text effectively, making them adaptable to various languages, fonts, and styles.

Cao et al. (Cao, 2011) demonstrated the use of HMMs in recognizing mixed handwritten and typewritten text. The system uses OCR to identify word boundaries and applies HMMs to classify text types. By leveraging context and features like image intensity, HMMs achieve a lower error rate (4.75%) than older methods, enhancing OCR accuracy and reliability for mixed-type documents.

2.3 CNN

CNNs are deep learning models specifically designed for processing and analyzing image data, widely used in OCR for recognizing both printed and handwritten text. CNNs enhance recognition accuracy by automatically extracting features such as edges, textures, and shapes from images, making them effective in handling variations like deformations, rotations, and scaling. The key components of CNNs include convolutional layers that capture local features, pooling layers that reduce data size while preserving essential information, activation functions like Rectified Linear Unit (ReLU) to learn complex patterns, and fully connected layers that map these features to output classes for final text recognition. This structure allows CNNs to minimize the need for manual feature engineering and remain robust against image variations.

In a study by Alrehali et al. (Alrehali, 2020), CNNs were applied to recognize characters in historical Arabic manuscripts. The architecture used multiple convolutional and pooling layers to learn character shapes and reduce feature sizes, achieving an accuracy of 88.20% by increasing training samples. Despite slightly lower performance than traditional methods, CNNs demonstrated unique strengths in complex manuscript recognition, showing potential for future advancements.

2.4 RNN and LSTM

RNNs are designed for processing sequential data, making them ideal for OCR, particularly in recognizing handwritten text. Unlike traditional feedforward networks, RNNs utilize contextual information from sequences, allowing them to handle continuous characters and lengthy texts more

effectively. They achieve this by maintaining a "recurrent" structure, where the current hidden state is updated at each step based on both the current input and the previous hidden state. RNNs use the same weights to handle sequences of varying lengths and learn temporal dependencies through the Backpropagation Through Time (BPTT) algorithm. However, traditional RNNs struggle with long sequences due to the vanishing gradient problem, which often necessitates the use of LSTM networks or Gated Recurrent Units (GRU) to improve their capacity for modeling long-term dependencies.

LSTM networks are a more advanced type of RNN specifically designed to address the vanishing gradient problem encountered in long data sequences. LSTMs are particularly effective in OCR tasks that require long-term dependency recognition, such as handwritten and complex printed text. They employ memory cells with gating mechanisms to control information storage, updating, and forgetting, which helps retain relevant context while filtering out noise. This adaptability to various OCR scenarios, including multilingual recognition and document analysis, enhances both text recognition accuracy and reliability.

In a study by Jun (Ma, 2024), RNNs and LSTMs were applied to sequence modeling for handwritten text recognition. While RNNs could handle sequential data, they struggled with long sequences. LSTMs overcame this limitation with gates that capture long-term dependencies. The study used a Bidirectional LSTM, which processes data in both forward and backward directions, improving context capture for letters and words. By combining CNNs for feature extraction with LSTMs for sequence modeling, the hybrid approach reduced the Word Error Rate (WER) to 12.04% and the Character Error Rate (CER) to 5.13%, outperforming standalone CNN models.

Similarly, Su et al. (Su, 2015) employed RNNs and LSTMs to enhance scene text recognition without requiring character segmentation. Their method converted word images into sequential HOG feature vectors, which were classified using a multi-layer RNN. By integrating LSTMs with input, forget, and output gates, the model captured long-term dependencies effectively. Using Connectionist Temporal Classification (CTC), the approach aligned unsegmented input with correct word outputs, achieving high accuracy rates (up to 92%) on datasets like ICDAR 2003, ICDAR 2011, and SVT. This combined RNN-LSTM method proved effective in recognizing text in complex, real-world conditions, setting a strong benchmark for future research.

Table 1: advantages and disadvantages of different models.

Model	Advantages	Disadvantages
HMM	<ul style="list-style-type: none"> - Simple and interpretable - High training efficiency - Suitable for sequential data 	<ul style="list-style-type: none"> - Limited expressive power - Fixed state transitions
CNN	<ul style="list-style-type: none"> - Automatic feature extraction - Spatial invariance - Excellent in image and video processing 	<ul style="list-style-type: none"> - Requires large datasets - High computational cost
RNN	<ul style="list-style-type: none"> - Handles sequential data - Context memory 	<ul style="list-style-type: none"> - Difficulty in learning long-term dependencies - High computational complexity
LSTM	<ul style="list-style-type: none"> - Effectively handles long-term dependencies - Versatile - Selective memory 	<ul style="list-style-type: none"> - Complex structure - Long training time

3 DISCUSSIONS

For the mutual benefit and protection of Authors and Publishers, it is necessary that Authors provide formal written Consent to Publish and Transfer of Copyright before publication of the Book. The signed Consent ensures that the publisher has the Author's authorization to publish the Contribution.

The development of OCR technology has evolved over nearly 80 years, shifting from rule-based methods to machine learning, particularly deep learning. This paper explores the applications of HMM, CNN, RNN, and LSTM in recognizing handwritten text and compares their strengths and limitations.

HMM uses probabilistic models for sequential data, like step-by-step handwritten character recognition, but struggles with noise, complex backgrounds, and long-term dependencies. CNNs automatically learn image features, such as edges and shapes, and are robust to variations like rotation and scaling, making them effective for visual data. However, they handle sequential information poorly and require large datasets and high training costs. RNNs are suitable for processing continuous text and variable-length sequences but suffer from gradient issues in long sequences, impacting performance. LSTMs overcome these limitations with gating mechanisms to capture long-term dependencies, improving recognition accuracy, but they have a complex structure and high computational costs.

This Table 1 summarizes the main advantages, disadvantages of HMM, CNN, RNN, and LSTM in OCR.

Although OCR technology has significantly increased efficiency and convenience, it still faces

several challenges and limitations at this stage. First, current OCR technology struggles with accurately recognizing handwritten text and complex fonts, resulting in relatively low accuracy. Additionally, the output of OCR models can sometimes be difficult to interpret and verify. For example, when recognition errors occur, it can be challenging to determine whether the problem stems from font characteristics, image quality, or limitations of the model itself. This lack of sufficient explainability and transparency restricts the use of OCR in fields that require high accuracy and strict scrutiny, such as medicine or finance.

To address these issues, researchers are actively promoting technological innovations. Deep learning methods are increasingly being applied in the field of OCR, particularly in recognizing complex text patterns. For instance, new techniques like self-supervised learning and Generative Adversarial Networks (GANs) are being introduced. Self-supervised learning enables OCR systems to leverage unlabeled data for pre-training, enhancing the ability to recognize handwritten and irregular fonts, while GANs are used to generate diverse handwriting samples, improving the model's generalization capabilities. Additionally, attention mechanisms (such as Transformer architecture) have shown great potential in handling complex text structures by more effectively focusing on key content in images and reducing interference from background noise.

There are also many new research directions aimed at improving the explainability and transparency of OCR technology. For example, feature visualization techniques, such as Gradient-weighted Class Activation Mapping (Grad-CAM), can help analyze how a model responds to different text regions during the recognition process, thereby providing a more intuitive understanding of the model's decision-making process. Tools like Local

Interpretable Model-agnostic Explanations (LIME) and SHapley Additive Explanations (SHAP) are also being used to analyze the impact of various features on recognition results, enhancing the model's interpretability and reliability. These advancements are expected to make OCR systems more reliable and widely applicable in practice.

4 CONCLUSIONS

This paper has reviewed the evolution of OCR technology, from early template matching methods to advanced machine learning models such as HMM, CNN, and RNN with LSTM. HMMs provided an early framework for handling sequential data, particularly in noisy environments. CNNs revolutionized the field by enabling automatic feature extraction from images, improving the recognition of printed and handwritten text. The introduction of RNNs and LSTMs allowed for better handling of sequential dependencies and context, significantly enhancing text recognition accuracy. Together, these technologies have broadened the scope of OCR applications, making it more versatile and effective in handling complex and diverse text scenarios.

This paper provides valuable overview of key models used in OCR technology. The analysis can guide future studies in selecting appropriate models based on specific OCR tasks and inspire further exploration to enhance model accuracy and efficiency. Additionally, the discussion on the latest technological advancements offers insights into potential directions for improving OCR systems.

Future research should focus on enhancing model transparency and interpretability while reducing computational costs. Continued innovation, particularly in deep learning and model explainability, will be essential for advancing OCR's effectiveness in diverse and complex text environments.

REFERENCES

- Alkawaz, M. H., Seong, C. C., & Razalli, H. 2020, February. Handwriting detection and recognition improvements based on hidden markov model and deep learning. In 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA) (pp. 106-110). IEEE.
- Alrehali, B., Alsaedi, N., Alahmadi, H., & Abid, N. 2020, March. Historical Arabic manuscripts text recognition using convolutional neural network. In 2020 6th conference on data science and machine learning applications (CDMA) (pp. 37-42). IEEE.
- Cao, H., Prasad, R., & Natarajan, P. 2011, September. Handwritten and typewritten text identification and recognition using hidden Markov models. In 2011 International Conference on Document Analysis and Recognition (pp. 744-748). IEEE.
- Karthick, K., Ravindrakumar, K. B., Francis, R., & Ilankannan, S. 2019. Steps involved in text recognition and recent research in OCR; a study. International Journal of Recent Technology and Engineering, 8(1), 2277-3878.
- Ma, J. 2024. A Study on a Hybrid CNN-RNN Model for Handwritten Recognition Based on Deep Learning. scitepress.org
- Memon, J., Sami, M., Khan, R. A., & Uddin, M. 2020. Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR). IEEE access, 8, 142642-142668.
- Mori, S., Suen, C. Y., & Yamamoto, K. 1992. Historical review of OCR research and development. Proceedings of the IEEE, 80(7), 1029-1058.
- Namysl, M., & Konya, I. 2019, September. Efficient, lexicon-free OCR using deep learning. In 2019 international conference on document analysis and recognition (ICDAR) (pp. 295-301). IEEE.
- Nikitha, A., Geetha, J., & JayaLakshmi, D. S. 2020, November. Handwritten text recognition using deep learning. In 2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT) (pp. 388-392). IEEE.
- Rabiner, L. R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257-286.
- Su, B., & Lu, S. 2015. Accurate scene text recognition based on recurrent neural network. In Computer Vision-ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part I 12 (pp. 35-48). Springer International Publishing.