A review of neural networks in handwritten character recognition

Ruoxin Li

Electronic and Electrical Engineering, University of Sheffield, Sheffield, S10 2TN, United Kingdom

liruoxin1209@gmail.com

Abstract. Handwritten character recognition has been a significant research focus in the fields of pattern recognition and artificial intelligence over the past decade. With the advent of neural networks, particularly deep learning models, the accuracy and efficiency of offline handwritten character recognition have dramatically improved. This paper presents a comprehensive review of recent developments in applying neural networks to handwritten character recognition. The literature review covers studies conducted between 2016 and 2023, providing insights into the methodologies, data processing techniques, and evaluation metrics used. The review spans various neural network architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid models. It categorizes and compares their performance across multiple benchmark datasets, highlighting specific improvements in recognition accuracy and efficiency. Furthermore, the review discusses the challenges faced in large-scale datasets, such as the diversity of handwriting styles and computational cost constraints. Notably, CNNs have shown outstanding performance, but the integration of advanced techniques like transfer learning and Generative Adversarial Networks (GANs) is explored as a potential avenue to enhance future recognition systems.

Keywords: Hand-written character recognition, convolutional neural network, recurrent neural network, hybrid model, deep learning.

1. Introduction

As a research focus in the field of pattern recognition for several decades, handwritten character recognition (HCR) has been applied to several areas including postal address reading, bank check processing, and historical document digitization [1]. It can be categorized as online HCR and offline HCR. The complexity of this problem arises from the diversity of individual writing styles, distortions, and noise in the data. Early studies state that the traditional method primarily involved stages of data preprocessing, feature extraction, and classification for character recognition. Despite these efforts, the recognition rates achieved by most renowned commercial brands were below 90% [1]. The advent of neural networks has brought groundbreaking progress and vitality to this field. A remarkable model called convolutional neural network (CNN), as proposed by research, is capable of automatically extracting rich and interconnected features from images. Furthermore, it can achieve significant recognition accuracy, as research has shown that it achieves an impressive 99.87% recognition accuracy on the MNIST dataset [2]. This paper reviews the development and application of neural network models

in handwritten character recognition, and explores various neural network architectures including CNNs, RNNs and hybrid models, discussing their methodologies and performance metrics. A suggested direction for future research is considered in this paper as well. This research enhances the efficiency and accuracy of operations in postal services, banking, and document preservation, reducing errors and saving resources. In the AI field, it advances the development of more robust recognition systems.

2. Early handwritten character recognition development

Handwriting recognition systems traditionally comprise three main components: data preprocessing, feature extraction, and classification.

Data preprocessing involves several steps to optimize input data, including sample normalization, noise removal, and geometric transformations such as rotation and scaling to correct distortions. Additionally, techniques like generating pseudo-samples and adding virtual strokes are employed.

Feature extraction is crucial in handwriting recognition, and it can be divided into structural and statistical features. Structural features analyze the character's structure, strokes, or components to extract shape and layout information. However, for handwritten characters, statistical features have proven to be more effective because handwriting is highly variable and inconsistent, with differences in stroke thickness, slant, and style across different writers. Statistical features, such as directional features like Gabor and gradient features, capture the distribution and directionality of pixel intensities, which makes them more robust in handling the variations inherent in handwritten text. These features are widely used for offline handwriting character recognition (HCCR) because they can effectively capture the essential characteristics of handwriting despite its variability.

Classification involves using models like the Modified Quadratic Discriminant Function (MQDF), Support Vector Machines (SVM), Hidden Markov Models (HMM), Discriminative Learning Quadratic Discriminant Function (DLQDF), and Learning Vector Quantization (LVQ). These models classify characters based on the extracted features. Text line recognition is another critical aspect, which can be approached using segmentation-based or segmentation-free strategies. Segmentation-based methods use projection and connected component analysis to segment text lines into individual characters, which are then recognized using character classifiers.

Segmentation-free methods employ sliding window techniques, where a window moves across the text line, and character recognition is performed within the window, often combined with statistical language models in a Bayesian framework to model the context and generate the final recognition result. Despite significant advancements, early offline handwriting recognition systems faced challenges in handling diverse handwriting styles and large-scale datasets. However, these traditional methods laid the foundation for subsequent research and inspired modern neural network-based approaches that have further improved recognition accuracy and robustness.

3. Convolutional Neural Networks (CNNs)

3.1. Convolutional Neural Networks (CNNs) model architecture and updates

CNNs have become the standard for image recognition tasks, including Handwritten Character Recognition (HCR).

LeNet-5, proposed by Yann LeCun et al. in 1998, is a classic CNN architecture primarily used for handwritten digit recognition, such as the MNIST dataset. It consists of two convolutional layers followed by subsampling layers, a fully connected layer, and an output layer. This architecture was revolutionary at its time and laid the groundwork for future CNN development.

AlexNet, introduced by Alex Krizhevsky et al. in 2012, is a deep CNN architecture that gained fame by winning the ImageNet Large Scale Visual Recognition Challenge. Although it was designed for complex image classification tasks, AlexNet can also be applied to handwritten character recognition. Its deeper structure, consisting of five convolutional layers and three fully connected layers, allows it to capture more complex features of the input data.

ResNet, developed by Kaiming He et al. in 2015, addresses the vanishing gradient problem in deep networks by introducing residual blocks. These blocks enable the training of much deeper networks by allowing the smooth flow of gradients through skip connections. ResNet's ability to train extremely deep models has proven to be advantageous in achieving high accuracy for various recognition tasks, including HCR. The field has been further advanced by improved models such as Relaxation CNN and ART CNN.

3.2. Benchmark performance evaluation

The performance of CNN architectures can be evaluated using benchmark datasets. For example, in the Arabic handwritten character dataset, the ResNet architecture achieved an accuracy of 0.72, a precision of 0.74, and a recall of 0.70. AlexNet, in contrast, achieved higher scores with an accuracy of 0.8107, a precision of 0.8270, and a recall of 0.8024 [3]. LeNet, while foundational, showed lower performance metrics with an accuracy of 0.6435, a precision of 0.8489, and a recall of 0.6381. These results highlight the advancements in CNN architectures over time. While LeNet provided a solid foundation for CNN applications in HCR, more recent architectures such as AlexNet and ResNet have significantly improved performance by capturing more complex features and enabling the use of deeper networks. Future research and model improvements, such as Relaxation CNN and ART CNN, continue to push the boundaries of what is achievable in handwritten character recognition, holding the promise of even greater accuracy and robustness in diverse and complex datasets.

4. Recurrent Neural Network (RNN)

4.1. General model

Recurrent Neural Networks (RNNs) have been widely used for sequence modeling in handwriting recognition. RNNs are particularly useful in recognizing cursive writing, where context understanding is crucial. In the standard RNN architecture, modifications are often made to the basic framework to improve its ability to capture dependencies across sequences. These modifications include altering the depth of the network, adjusting the number of hidden units, and fine-tuning the activation functions to better handle the nuances of cursive script. However, despite these improvements, basic RNNs still have several limitations. One of the most significant issues is the vanishing gradient problem, which hampers the network's ability to effectively learn long-range dependencies. Additionally, RNNs are prone to overfitting, especially when the training data is limited, and they struggle with high computational demands, which reduces their efficiency in processing very long sequences.

4.2. Improved model

Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) represent significant advancements over traditional Recurrent Neural Networks (RNNs), particularly in tasks such as handwriting recognition where context and long-term dependencies play a crucial role. LSTMs introduce memory cells equipped with three gating mechanisms—input, forget, and output gates—that regulate the flow of information. These gates ensure that relevant data is retained over long sequences, while irrelevant information is discarded, effectively solving the vanishing gradient problem that hampers traditional RNNs.

Conversely, GRUs simplify the LSTM architecture by combining the functions of the input and forget gates into a single update gate, thus eliminating the necessity for separate memory cells. GRUs also include a reset gate, which controls the influence of previous states on the current output. This simpler structure allows GRUs to achieve similar performance to LSTMs but with fewer parameters, resulting in lower computational costs.

The vanishing gradient problem in RNNs occurs when gradients used to update the network's weights diminish as they are propagated backward through time, leading to extremely slow or stalled learning in the earlier layers of the network. This problem is particularly pronounced in tasks requiring the learning of long-term dependencies, such as handwriting recognition. LSTM networks address this issue by

introducing memory cells that maintain a constant error, allowing gradients to remain effective across numerous time steps. Similarly, GRUs simplify this architecture while maintaining effectiveness, making them computationally more efficient [4]. These improved models have become essential tools in sequence modeling, particularly in applications like handwriting recognition, where understanding the sequential nature of input data is critical.

4.3. Benchmark performance evaluation
In comparing the performance of GRU and LSTM models across multiple handwriting recognition datasets, the GRU-based model (Att-BGRU-V) consistently outperforms LSTM-based models. The GRU model achieves lower RMSE and ED values while maintaining higher or comparable recognition rates. For instance, on the IRONOFF lower-case dataset, the GRU model records an RMSE of 2.9 and a recognition rate of 93.2%, outperforming LSTM models which have higher RMSEs and lower accuracy. Similarly, on the LMCA dataset, the GRU model not only matches the LSTM's recognition accuracy of 98.9% but also demonstrates lower RMSE. This indicates that GRU models offer superior performance in handwriting recognition tasks, likely due to their simplified architecture and lower computational requirements [5].

5. Hybrid models
Recent studies have investigated hybrid models that combine CNN and RNN architectures in order to capitalize on the strengths of both. These models are designed to enhance accuracy by capturing both the spatial and sequential dependencies of handwritten text.

5.1. Typical architecture
The specific architecture is based on a sequence-to-sequence (Seq2Seq) model with an attention mechanism. This hybrid model utilizes a CNN for feature extraction, which efficiently captures spatial hierarchies in the handwriting. Convolutional layers process the input images to extract detailed features such as edges, curves, and textures. These features are crucial for understanding the spatial aspects of the handwriting.

Following feature extraction, the model employs a bidirectional gated recurrent unit (BGRU) for encoding and decoding the extracted features. The BGRU processes the sequential data in both forward and backward directions, which allows it to capture the context from both past and future sequences. This bidirectional approach is particularly beneficial for recognizing and interpreting complex handwriting patterns where context is essential.

The attention mechanism within the Seq2Seq model further enhances the hybrid architecture by focusing on relevant parts of the input sequence during the decoding process. This mechanism dynamically weights the importance of different features, allowing the model to concentrate on critical aspects of the handwriting while generating the output sequence.

5.2. Performance evaluation and benchmarking
On the RIMES dataset, the H2TR (CNN-RNN) model achieved a character accuracy of 98.14%, demonstrating superior performance compared to standalone CNN and RNN models [6]. More specifically, the H2TR model demonstrated a 2% improvement over the CNN model and an 11% improvement over the RNN model.

The performance of the H2TR model can be attributed to its ability to effectively capture both spatial and sequential dependencies. By combining the strengths of CNNs and RNNs, the hybrid model can better understand and interpret the nuances of handwritten text. This leads to more accurate recognition and classification of characters, even in complex and varied handwriting styles.

In addition to character accuracy, the hybrid model also shows improvements in other metrics such as word error rate (WER) and sequence error rate (SER). These improvements demonstrate the robustness and versatility of the hybrid CNN-RNN architecture in effectively handling a wide range of handwriting recognition tasks.

6. Discussion

Despite significant strides in handwritten character recognition, several challenges persist across different neural network architectures. Convolutional Neural Networks (CNNs), while effective in image recognition tasks, are limited by their fixed receptive fields and reliance on large datasets for optimal performance. Recurrent Neural Networks (RNNs), on the contrary, face issues such as vanishing gradients and high computational complexity, particularly in long sequences.

Hybrid models that combine CNNs and RNNs have been proposed to leverage their respective strengths, but these models often suffer from complexity and tuning difficulties. An emerging approach that shows promise in overcoming these challenges involves the use of Generative Adversarial Networks (GANs) for data augmentation.

GANs consist of two neural networks: a generator and a discriminator. The generator synthesizes artificial data samples, while the discriminator learns to distinguish between real and fake data. This adversarial process leads to the generation of highly realistic synthetic data that can effectively expand the training dataset for recognition tasks.

By leveraging GANs for data augmentation, they offer a robust solution to improve recognition accuracy and efficiency. The synthetic data generated by GANs enriches the training dataset, and enhances the overall performance of recognition systems by reducing overfitting and improving generalization [7]. This approach addresses the inherent limitations of traditional CNN and RNN.

In parallel, federated learning emerges as a crucial technique for privacy-preserving handwriting recognition. In scenarios where sensitive handwriting data, such as personal signatures or medical notes, are involved, federated learning allows models to be trained across multiple decentralized devices without the need to share raw data. This ensures the maintenance of privacy while still benefiting from the diverse datasets across different sources, thereby improving the robustness and generalization of the models. The use of federated learning could have a significant impact in industries like banking or healthcare, where data privacy is paramount [8].

Meanwhile, Meta-learning with its ability to improve learning algorithms based on multiple tasks, offers several advantages that can address some challenges in HCR. One of the significant benefits of meta-learning is its ability to perform well with limited data, which is particularly useful in HCR where annotated data can be scarce [9]. Few-shot learning techniques within meta-learning allow models to learn new characters or styles with very few examples. This capability can be crucial for expanding the recognition system to new alphabets or styles without the need for extensive retraining. Meta-learning can also optimize the learning algorithms themselves, enhancing their efficiency and efficacy in handling HCR tasks. This includes learning the best initialization parameters, optimization strategies, and even the most suitable hyperparameters for the HCR models.

Future research should focus on data augmentation and standardization, model optimization and architecture improvement, and the integration of advanced techniques like meta-learning and reinforcement learning. It is crucial to enhance model robustness through multimodal fusion and testing on noisy data. Additionally, fostering cross-disciplinary collaboration and sharing models, data, and research outcomes. Additionally, fostering cross-disciplinary collaboration and sharing models, data, and research outcomes will drive progress. Key future directions include utilizing GANs for data augmentation, exploring Transformers and hybrid models, advancing few-shot and zero-shot learning, applying self-supervised and unsupervised learning, enhancing model interpretability and real-time processing capabilities, integrating multimodal information, and adopting federated learning for privacy protection. These efforts will significantly improve the accuracy, efficiency, and adaptability of HCR systems.

7. Conclusion

This comprehensive review explored various neural network architectures and their applications in offline HCR. CNNs have emerged as the mainstream approach due to their robust feature extraction capabilities. However, hybrid models and advanced techniques such as Meta-learning and GANs are gaining attention for their potential to address existing challenges in HCR. The review concentrated

heavily on well-established models such as CNNs, RNNs, LSTMs, and GRUs, potentially neglecting emerging techniques and less explored hybrid models that could offer alternative perspectives or innovative solutions. Looking forward, future research could address these limitations by expanding the scope of the literature review to include a broader range of studies, especially those that explore novel or underrepresented techniques. Additionally, incorporating a wider variety of datasets and evaluation metrics would provide a more holistic understanding of model performance. Furthermore, exploring the integration of advanced techniques such as meta-learning, transfer learning, and generative models could offer new avenues for enhancing HCR systems. Finally, fostering cross-disciplinary collaboration and focusing on the practical deployment of HCR technologies in real-world settings will be essential for translating research advancements into tangible societal benefits.

References
[1] Jin, L., Zhong, Z., & Yang, Z. (2016). Applications of Deep Learning for Handwritten Chinese Character Recognition: A Review. ACTA AUTOMATICA SINICA, 42(8), 1125–1141.
[2] Ahlawat, S., Choudhary, A., Nayyar, A., Singh, S., & Yoon, B. (2020). Improved handwritten digit recognition using convolutional neural networks (CNN). Sensors, 20(12), 3344.
[3] Nugraha, G. S., Darmawan, M. I., & Dwiyansaputra, R. (2023). Comparison of CNN's architecture GoogleNet, Alexnet, VGG-16, Lenet -5, resnet-50 in Arabic handwriting pattern recognition. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control.
[4] Graves, A., Liwicki, M., Fernandez, S., Bertolami, R., Bunke, H., & Schmidhuber, J. (2009). A novel connectionist system for unconstrained handwriting recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(5), 855–868.
[5] Rabhi, B., Elbaati, A., Boubaker, H., Hamdi, Y., Hussain, A., & Alimi, A. M. (2021). Multi-lingual character handwriting framework based on an integrated deep learning based sequence-to-sequence attention model. Memetic Computing, 13(4), 459–475.
[6] Geetha, R., Thilagam, T., & Padmavathy, T. (2021). Effective offline handwritten text recognition model based on a sequence-to-sequence approach with CNN–RNN networks. Neural Computing and Applications, 33(17), 10923–10934.
[7] Elaraby, N., Barakat, S., & Rezk, A. (2022). A conditional GAN-based approach for enhancing transfer learning performance in few-shot HCR tasks. Scientific Reports, 12(1).
[8] Zhuofan Mei, The Recognition of Tibetan Handwritten Numbers Based on Federated Learning. Journal of Artificial Intelligence Practice (2021) Vol. 4: 1-12. DOI:
[9] Hospedales, T., Antoniou, A., Micaelli, P., & Storkey, A. (2021). Meta-learning in Neural Networks: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1–1.