Survey Paper on AI-Based Mathematical Handwritten Formula Recognition Using DL

Deshpande Rashmi Wamanrao[1], Dr. Abhishek Kumar[2*]

[1]PhD scholar, YBN University, Ranchi, India.
[2*]Associate Professor, EE Department, School of Engineering, YBN University, Ranchi, India

Abstract
Handwritten formula recognition is a rapidly developing area of computer technology, with people having the greatest ability to identify images and items. However, selecting suitable component sets for handwriting recognition is challenging due to the vast variety of writer's handwriting. Many fields, such as medical imaging, science, and archaeology, rely heavily on handwriting, such as forensic medicine. The rapid growth of handheld devices, virtual textbooks, and specialized communication devices has prompted increased attention to handwriting recognition. To train a deep learning network, a dataset of handwriting is constructed, containing 2200 illustrations of each alphabet and a viewable database. Multidimensional Long Short-Term Memory networks are used in emerging strategies for fine Handwritten Text Classification, but they have limitations, such as removing feature vectors close to those extracted by convolution layers, which are less computationally complex.

Keywords: Computer science, handwritten text recognition, word segmentation, machine learning, deep learning

Introduction

Handwriting recognition efficiency is influenced by the software suite used to represent input data. This leads to a variety of evaluation metrics, each complex in terms of characteristics. Image processing involves the processing of objects in software perception. With advancements in technology, various programs for image engineering have emerged. Text recognition is crucial in various fields, but computer-based methods are challenging. Offline handwritten text identification is essential for releasing millions of ancient documents in archives. Most of these records have been digitized, but exposure to their information is limited. Handwritten Text Transcription (HTR) programs aim to transcribe these records for better understanding and preservation. Handwriting recognition (HTR) has significantly advanced over the past two centuries due to comprehensive learning and identification methods in Automated Speech Recognition (ASR) and the growing availability of viewable databases for training and certification HTR programs. HTR systems can be classified into various groups, with the remembered text period being the most essential factor. Offline or online recognition is possible, and multiple source strings distinguish issue forms. Character collection, writing style, and restrictions significantly impact recognition systems, allowing for the identification of characters, words, or entire sentences in text.

A. Off Line Recognition

Software significantly impacts handwriting recognition, with text recognition technologies being the first to be implemented. These technologies digitize paper sheets into two-dimensional objects, improving and removing attributes for identification. Offline recognition, where text is remembered after the writing task, is another form of acknowledgement. Optical character recognition (OCR) is a subset of offline handwritten text. Although not ideal for real-time interaction, they are useful for converting paper files into data that can be read or processed by machines.

B. On Line Recognition

Handwriting is captured and recognized in real-time through internet-based systems. Writing devices, such as notebooks or monitors, gather data on the position and movement of a pen-like visual object on a printed sheet. This data is sent to an identification system, which uses a medical grade flip on the pen tip to indicate pen position and stroke differentiation. On-line handwriting recognition systems offer interactive web applications for those who require converter-based mechanisms but cannot use hand-held devices. Most online character recognition programs are used for filling out paperwork, allowing minimal interaction. In more sophisticated technologies, a pointed pen and syntax highlighting device may replace a keyboard and mouse.

Man-machine interaction primarily involves keyboards and aiming tools, which can be uncomfortable when the unit is thicker or similar to a human palm. Normal keyboards are difficult to fit into portable screens, and the mechanism's width is typically determined by it.

Online handwriting recognition is a creative way to communicate with machines using a pen-like stylus, particularly in languages with large entry groups like Chinese and Japanese. Traditional targeting tools like race rollers or pencils are inadequate for sequential feedback programs. New input techniques, such as expression and handwriting classification models, have been established to address these issues. Research on Latin, Chinese,

and Japanese characters has been published, but little has been done on Arabic letters due to the difficulty of Arabic text and lack of Arabic
repositories. The Handwritten Text Recognition (HTR) group has adopted recurrent neural networks, particularly Multidimensional Long Short-Term Memory (MDLSTM) networks, which currently holds the state-of-the-art score in the field.

Literature Review

The feature extraction step is crucial for handwritten character recognition due to variations in written text caused by factors like size, society, standard of knowledge, and source. MDLSTM networks differ from standard LSTM networks in that they implement relapse across a single axis, while MDLSTM implements repeat across two dimensions. This allows for long-term 2D-LSTM systems are a popular line-level HTR strategy that uses a character-level representation of the input line image by translating 2D data into an ID sequence. These systems are computationally intensive compared to convolutional operations. They are physically similar to two-dimensional clustering algorithms, allowing for the extraction of meaningful attributes in lower layers of the stack without the need for recurrent layers' wide context.
Recognizing text from handwritten documents for digitization is crucial, especially when dealing with complex documents like Fig. 3a (Bangla language) and Fig. 3b (Bangla language). Detecting text regions in Fig. 2a is easy due to the absence of figures or plots, but identifying text regions in Fig. 3b is more challenging due to the presence of non-text devices within visual elements. To differentiate between these objects, the first step is to exclude non-text objects. The identification of text regions in Fig
Handwritten text recognition systems come in two types: online and offline. Off-line recognition records writing visually, while online recognition uses two-dimensional dimensions and author sequences. Online methods are considered more efficient for character recognition. Handwriting style is limited in automated programs to improve efficiency. Characters are often written to simplify differentiation or have a leading line for more accurate writing. These restrictions are intended to increase the recognizer's efficiency and can impact a normative decision guideline. If too stringent, the device may be refused by customers. Restrictions affect writing style, pace, and classification accuracy. For example, graffiti character is shown in Fig.

1.HTR Technology

The most popular methods for HTR (HMM-GMM) include N-gram based methods (LM) and optical character modelling using HMMs with Gaussian mixture filled. Multilayer perceptron's (HMM-MLP) are used to approach emissions chances. Both discriminative and exclusionary learning styles improve perceptual modelling. Recently, RNNs for optical simulation have gained traction in HTR precision. Optical models are trained using lines images and transcripts, which are considered "diplomatic" if they match the paper material

2.Language Modeling

CRNNs can capture lexical and linguistic meaning independently, unlike classic LM approaches that focus on modeling contextual patterns and constraints. These systems are based on plain text and can be used quickly. The SRILM Toolkit calculates statistical character N-grams, which can enhance CRNNs' HTR results. CRNN performance character probabilities can be subjected to trained N-gram contextual constraints using a stochastic finite-state transducer. The edge chances of this transmitter are computed by averaging approximate N-gram percentages.

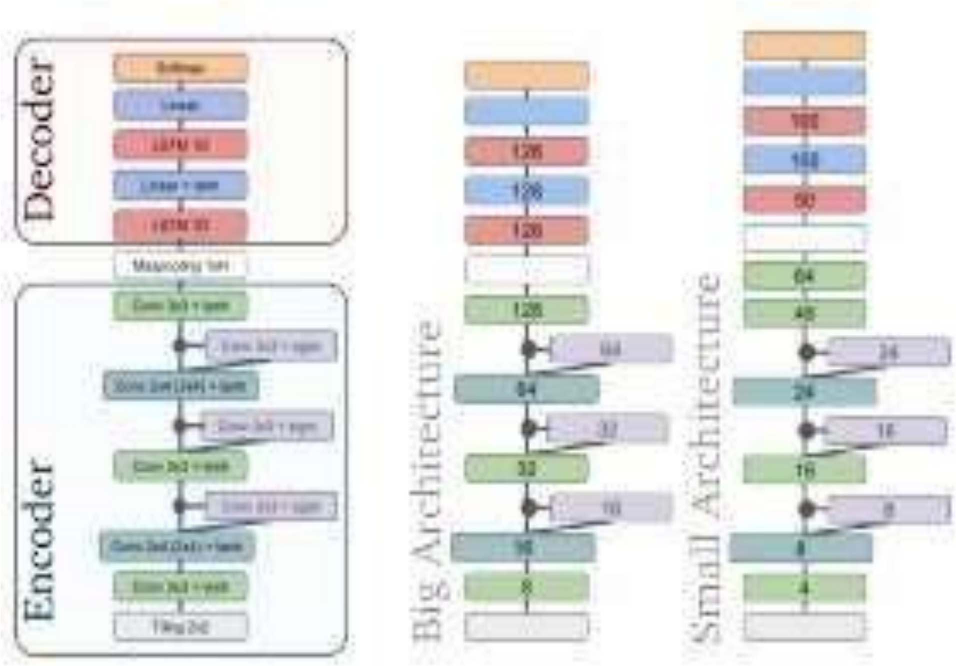
Methodology

Handwritten character recognition systems have been developed by various researchers, including a face detection system using fuzzy logic, a VLSI architecture system that doesn't impact shift and distortion, and a revolutionary method for recognizing handwritten Tamil characters using the Kohonen Self-Organizing Map (SOM). The Kohonen Self-Organizing Map (SOM) is an unsupervised neural network that produces near-perfect results, but errors may occur if the handwritten symbols aren't divided correctly. A multi-layer feed forward neural network has also been proposed for verifying a person's identity using handwriting. These advancements have significant implications in the field of handwritten character recognition.
The study compares physical and digital writing methods using a system that considers the advantages of both.

Users are advised to write on regular paper as if using an off-line platform. An USB Imaging System is used to capture image attributes as they are written on paper, rather than waiting for an entire page to be written and inserted into a device. Character strokes are deleted from the video sequence using various video processing methods. The system does not require specialized tools, as many laptops come equipped with a small tape recorder for net meetings applications. An USB CCD camera captures image sequences with 320 x 240 pixels, 6 frames per second, and adequate illumination. The writing area shown in the video images is approximately 20 x 12 cm in size.
Offline and online leaming are both anxious and sleepy processes, and both mechanisms must be incorporated in the framework for gradual learning based on user feedback and offline leaming on permanent information. RPROP is the preferred algorithm for offline leaming ue to its benefits over pure backpropagation. It outperforms backpropagation in terms of the number of epochs needed to reach a minimum distance of decision variables. However, RPROP is not a digitally-based teaching and learning method.The study suggests that using RPROP for online education is equivalent to using backpropagation for training data, scale image, performing two step histogram projections to detect text lines and individual characters, and extracting any visible noise in the scanning final picture. This approach ensures the same quality of training data for online education.

1.Preprocessing:
The grey-scale image was preprocessed to remove non-homogeneous contrast in the foreground, resulting in a screened image. The image's binarization is expressed as Imin, Imout, and th, with the default value of 0.5. This process can achieve both effects by adjusting the input and output images.
The system for splitting hand-printed Arabic text involves using binary trees and a simultaneous smoothing algorithm. The system involves tracing the thinned picture from right to left using a 33 window and recording the arrangement of traced bits. A binary tree with many nodes is constructed using specified rules, with each vertex defining the form component of a linked entity. Thresholding is used to reduce the number of branches, Freeman code sequence, and distortion in the dried picture. The tree structure is broken into several semis, each representing a different individual.Arabic uses a unique set of dots, with up to three dots per letter, positioned above or below the main body. These dots can cause misrepresentation of characters, so effective pre-processing methods are necessary to avoid ignoring them and altering the character's identity. Figure 9 shows three Arabic phrases with one, two, or three sub-words.
Writing on a page is typically divided into lines and sentences, with gaps between the lines and between the sentences.
This study focuses on removing handwriting sign frames from a short grayscale image of an odd 1D sequence of symbols, demonstrating that the template is not limited to Latin characters, with no gaps among characters and no need for anticipating room.This work focuses on the use of symbols among characters in a text recognition system. The application's core is the FCN pattern prediction, which uses symbols as grouping labels and is based on an FCN model. The Block Length CNN moves generate word blocks of 32-16 N in size, and the number of output predictions is proportional to the predicted block length N
when sampled blocks are fed into the Symbol Prediction FCN. The system uses a Deep Neural Network (NN) to search for handwritten text, with the input data often made smaller. The minimum prerequisite for HTR introduction is TF, and the CPU can handle Ote (Ote) or a GPU (GPU).
The Android app for our framework allows consumers to digitize handwritten documents by snapping pictures using the Android software and device screen, as illustrated in figure 12.
An Android app has been developed to allow customers to take pictures of written characters on their phones. The app processes the clicking picture and uses a database to recognize handwritten text. The backend of the system hosts a pre-trained neural network architecture for predictions, while the handwritten text image undergoes image analysis. The TensorFlow-trained neural network architecture and a Python script using the OpenCV library are used. The Deep Convolution System model is used for image analysis.
The convolutional neural network (CNN), a cutting-edge technology, is widely used in fields like image and video recognition, natural language processing, and recommendation systems.

2.Architecture Details
The study developed two machine learning frameworks based on principles for accuracy, speed, and size. Both frameworks use a convolution layers codec, limit around the vertical axis, and a recurrent demodulator. The major system has 3x3 convolution layers with 8 functions, 2x4 convolution layers with 16 functions, a 3x3 convolution layers gate, and a 3x3 fully connected with 32 functions. The tiny channel's codec includes 3x3 convolution layers with 64 functions. The video codecs consist of two symmetric video codecs.
The study uses LS'I'M layers with 128 neurons and a neural network of 128 units for two networks: a slow system for actual pictures and a quick system for photos transcoded to approximately '/0 million of the current length.



The study identifies two major implementations, Fast (F) for large ones, and FastSmall (S) and FasterSmall (X) for minor ones. Big frameworks have 750k criteria, while tiny implementations only have management accounting. Mass encoding occupies 1.1MB and 500kB intervertebral discs, respectively.

3.Neural Network Overview

The deep algorithm is divided into three main parts, based on the use of neural networks in machine learning and symmetric LS'IM for word implementations. The compression codec, which converts two-dimensional images into two-dimensional feature maps, is the slowest part of the model, accounting for 80% of simulation time. The goal is to make it universally applicable, reducing processing time. The GUI converts 2D photo descriptions into the intended ID interpretation, predicting character sequences. The lstm Model RNN decoder uses attribute series to anticipate character sequences, retaining most of the channel's power and loading quickly, reducing computing effort.

Results

A USB CCD Camera is used to identify pencil tip locations in photos, calculating distal part and utilizing dimension and contextual awareness for face detection. The roles are transferred to the LaTeX data transformation executable, which supports range creation, reference, segment, implant system, formula, direct quotes, and regular text conditions. The acceptance rate is 90.21 percent, the incorrect rate is 7.02 percent, and the rejected rate is 3.77 percent. The main causes of the theoretical problem are abnormal writing and uncertainty between identical formed characters, with word pairs like "e" and "c," "5" and "S," and "u" and "v" causing the most uncertainty.

A USB CCD Camera is used to identify pencil tip locations in photos, calculating distal part. This technique aids in situational awareness and resolution of uncertainty. Table II shows LaTeX code used to create standard text, line, citation, array, item list, equation, verbatim text, and mathematical expressions, including equations and mathematical expressions.

Figure 15 shows the testing of the training algorithm and RPROP architectures on various datasets, resulting in a teaching and learning curve. The Lemming curve highlights error as a feature of capabilities, indicating high bias or courage and bravery.

The study reveals no noticeable issues with clients or autocorrelation problems in the learning curves. After 100 epochs, the RPROP algorithm converges to a better maximum than back-propagation due to its advantages over pure back-propagation. Table III shows no evidence of overfitting or underfitting in the learning curves. The RPROP algorithm converges to a better maximum after 100 epochs than back-propagation due to its advantages. Table III shows the comparison of learning algorithms used for practice. False alarm values are computed around the training curve's 'x' axis, ignoring values when m 40. The method was applied to paper images, observing 71180 objects across 52 document images, with 70411 successfully noted. The detection rate was 98.9%, with a false positive rate of 1.1%.

This paper presents the findings of the largest technology on all available data and translations, referred to as Generic models. The four channels (Accurate, Quick, FastSmall, and FasterSmall) are equated to A2iA TextReaderl M's previous edition. Each word is represented by two neural nets in A2iA'léxtReaderTM, each using an MDLS'I'M-RNN design. The concise description works with actual (300DPI) photos, ensuring accurate and fast sequencing.The easy version uses transcoded pictures to 70% of their length, with initial weights for near-zero language originating from a camera mounted on another word, each based on dialect information.

A. Multilingual Education

This section discusses the impact of the largest technology on all existing evidence and translations, referred to as Generic models. The four networks (Accurate, Quick, FastSmall, and FasterSmall) are compared to A2iA's previous version, with each phrase defined by two machine learning in A2iA'léxtReaderTM. Each channel is an MDLS'I'M-RNN, following the structure described in [30]. The concise description works with initial images, while the quick version works with pictures that have been transcoded to 70% of their current length. These systems are based on dialect data, with initial weights for reduced language originating from a camera mounted on a foreign language.

Conclusion

Online handwriting recognition is a creative way to interact with computers using a pen-like stylus, combining the benefits of online and offline systems. The pen-tip's roles are tracked using an Usb Imaging System, and directional data is calculated. Text reclassification uses spatial data from the pen motion points, and fine identification is achieved using spatial data. The Android app uses touch, picture, and video input for text classification. A Java package containing class enforces the neural network leaning model, which can be used in other implementations. Several enhancements to the application or learning model can be proposed, such as restricting the neural network's extracted features to operate on more thoroughly pre-processed data and merging multiple classifiers based on different attributes. A fuzzy classification learning model, such as an ART system, can be used on raw input. A handwriting recognition neural network model achieves state-of-the-art performance, using a convolution layers algorithm to extract common characteristics from input documents and an LSTM encoding predictions different characteristics in each word. The computer can now recognize symbols and words, enhancing the interaction between humans and machines

References

1. T.-N. Truong, C. T. Nguyen, R. Zanibbi, H. Mouchère, and M. Nakagawa, "A Survey on Handwritten Mathematical Expression Recognition: The Rise of Encoder-Decoder and GNN Models," Pattern Recognition, vol. 153, Art. no. 110531, Sept. 2024, doi: 10.1016/j.patcog.2024.110531. (ScienceDirect)
2. Y. Zhang and G. Li, "Improving Handwritten Mathematical Expression Recognition via Integrating Convolutional Neural Network With Transformer and Diffusion-Based Data Augmentation," IEEE Access, vol. 12, pp. 67945–67956, 2024, doi: 10.1109/ACCESS.2024.3399919. (Directory of Open Access Journals)
3. P. Fu, G. Xiao, and H. Yang, "SATD: Syntax-Aware Handwritten Mathematical Expression Recognition Based on Tree-Structured Transformer Decoder," The Visual Computer, vol. 41, pp. 883–900, 2025. (DOI)
4. J. Tang, H. Guo, J. Wu, F. Yin, and L. Huang, "Offline Handwritten Mathematical Expression Recognition with Graph Encoder and Transformer Decoder," Pattern Recognition, vol. 148, Art. no. 110155, 2024. (Mindat)
5. N. V. Bay and T. H. Vinh, "A Short Review for Handwritten Math Expression Recognition Techniques," Procedia Computer Science, vol. 235, pp. 231–239, 2024. (ScienceDirect)
6. J. Zhu, W. Zhao, Y. Li, X. Hu, and L. Gao, "TAMER: Tree-Aware Transformer for Handwritten Mathematical Expression Recognition," arXiv:2408.08578, 2024. (arXiv)
7. T. Guan, C. Lin, W. Shen, and X. Yang, "PosFormer: Recognizing Complex Handwritten Mathematical Expression with Position Forest Transformer," arXiv:2407.07764, 2024. (arXiv)
8. C. Liu et al., "NAMER: Non-Autoregressive Modeling for Handwritten Mathematical Expression Recognition," arXiv:2407.11380, 2024. (arXiv)
9. Y. Liu, W. Ke, and J. Wei, "Attention Guidance Mechanism for Handwritten Mathematical Expression Recognition," arXiv:2403.01756, 2024. (arXiv)
10. A. Vaswani et al., "Attention Is All You Need," Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017. (Still one of the most cited foundational references for Transformer-based HMER methods.)