International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM03136
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Handwritten Text Recognition: A Deep Learning Based Approach to Digitize Handwritten Text

P. Nancy Celine [1], Gugulothu Krishnaveni[2], Potta Pramodita[3]
[12345]*Vasavi College of Engineering, Hyderabad, Telangana, India*
nancy@staff.vce.ac.in, krishnavenig9959@gmail, pramodita832@gmail.com

Abstract

Our project employs deep learning methods to digitize handwritten pages into text. The process has a number of steps, ranging from identifying individual words, and identifying individual characters. Through the use of convolutional neural networks (CNNs) and other sophisticated machine learning models, the project seeks to perform accurate and efficient recognition of handwritten text. The model is trained on the IAM Handwriting Dataset, which contains a diverse collection of handwritten text samples, allowing the system to generalize well across different handwriting styles. Key components of the project include preprocessing the input images, extracting features using CNN layers, and optimizing the model over multiple training epochs. The result is a robust handwriting recognition system capable of converting handwritten words into editable digital text, with applications in document digitization, historical manuscript preservation, and automated data entry systems.

Keywords: Handwritten Text Recognition, Deep Learning Methods, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs)

1. Introduction

Handwritten Text Recognition (HTR) is a domain-specific subfield of Optical Character Recognition (OCR) that deals with the recognition of handwritten text into digital modes. Legacy HTR systems have traditionally used rule-based approaches and statistical models such as Hidden Markov Models (HMMs), which were mostly limited in handling varying handwriting styles and were mostly done manually through feature extraction. The rise of deep learning brought a major shift to handwritten text recognition (HTR), making Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) key players in advancing the field. CNNs are particularly good at extracting spatial features from handwritten images, whereas RNNs, especially Long Short-Term Memory (LSTM) networks, are able to capture the sequential nature of text, allowing for the recognition of whole words and sentences. The existence of big data, including the IAM Handwriting Dataset, has played a critical role in pushing the frontiers of HTR, with varying handwriting samples for training powerful models. Data augmentation techniques, including rotations and noise addition, further enhance model generalization. Despite significant progress, challenges persist due to variations in handwriting, cursive text, and overlapping characters. Modern HTR systems incorporate language models and dictionaries in post-processing to correct recognition errors and improve output accuracy.

2. Related Work

Early text recognition methods, particularly for natural scene images, were mostly targeted at recognizing either isolated characters or entire words. Character-based systems generally used sliding window methods, connected component analysis, part-based hierarchical models, or stroke width transformations to recognize individual characters. After being detected, classifiers — most commonly aided by feature extraction algorithms like Convolutional Neural Networks (CNNs) and Histograms of Oriented Gradients (HOG) with random ferns or multi-scale features with random forests — were used to identify the characters. After identification, fixed lexicon-based grouping algorithms were utilized to restore words from identified characters.

Within the larger context of Handwritten Text Recognition (HTR), the process started with rule-based techniques and template matching, wherein handwritten characters were matched against a repository of pre-defined templates. While useful for rigid and simple scripts, these techniques faltered in the face of natural

ISSN: 2583-6129
DOI: 10.55041/ISJEM03136

variability present in human handwriting. To bridge these limitations, statistical models such as Hidden Markov Models (HMMs) came into being, providing enhanced ability to capture sequential dependencies. However, even HMM-based systems were requiring a huge amount of hand-engineered feature design and remained handwriting variability sensitive, which hampered their application in real-life scenarios.

3. Proposed Methodology

3.1 Dataset Characteristics

Our project utilizes words.tgz and words.txt datasets of IAM handwriting dataset to train and test the model to recognize the hand written words. The IAM Words Dataset is a subset of the IAM Handwriting Database, specifically designed for isolated word recognition tasks. It contains thousands of handwritten word images along with their corresponding transcriptions, provided in a words.txt file. In this project, the dataset is preprocessed to extract individual word images and organize them with clear labels, making it suitable for training a CRNN (Convolutional Recurrent Neural Network) with CTC Loss for handwritten text recognition. This setup enables the model to learn how to accurately recognize handwritten words and forms the foundation for building a full handwriting recognition pipeline.

Attribute	Value
Dataset Name	IAM Handwriting Dataset
Number of Samples	967
Number of Samples	232
Image Size	256x256 pixels
Number of Classes	10 or based on data
Language	English

Fig 1. Dataset Description

3.2 Architecture

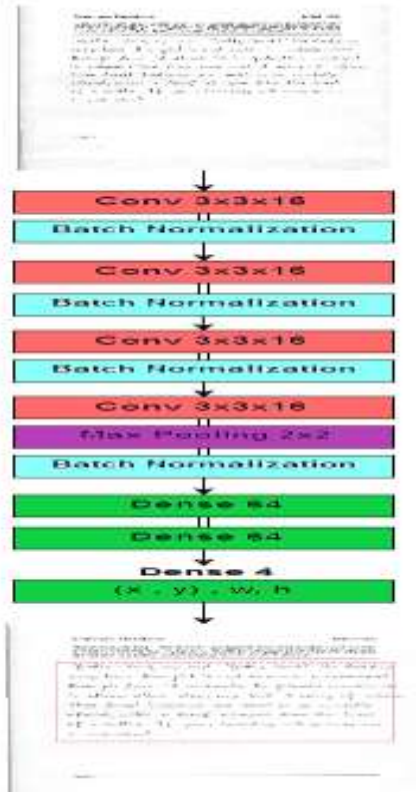

Fig 2. Model Architecture

1. Convolutional Layers (Conv 3x3x16)

- Conv 3x3x16 refers to a convolutional layer with a kernel size of 3×3 and 16 filters.
- Conv layers extract low-level features like edges, textures and patterns from the image.
- Repeated application of Conv layers (stacked multiple times) allows the model to learn more complex, hierarchical features.
- Key Purpose: Identify patterns in handwriting like strokes, curves, or letters.

2. Batch Normalization

- After each convolutional layer, Batch Normalization is applied.
- This normalizes the intermediate feature maps to stabilize and speed up training.
- Key Purpose: Prevent overfitting and improve training efficiency.

3. Max Pooling (2x2)

- Following a few convolutional and batch normalization layers, a Max Pooling layer with a kernel of 2×2 is used.
- This approach decreases the spatial size of the feature maps while preserving meaningful information.

4. Fully Connected Layers (Dense Layers)

- The convolutional layers' output is flattened and fed into two Dense (fully connected) layers:
- Dense 64: Has 64 neurons to learn higher-level features.

Layer Type	No.of Units/Neurons	Activation Function	Output Shape
Input Layer	-	-	(256,256,3)
Convolutional Layer 1	32	ReLu	(256,256,3)
Maxpooling Layer 1	-	-	(128,128,32)
Convolutional Layer 2	64	ReLu	(128,128,64)
Maxpooling Layer 2	-	-	(64,64,64)
Dense Layer	64	ReLu	64
Output Layer	1	Softmax	10

ISSN: 2583-6129
DOI: 10.55041/ISJEM03136

4. Implementation

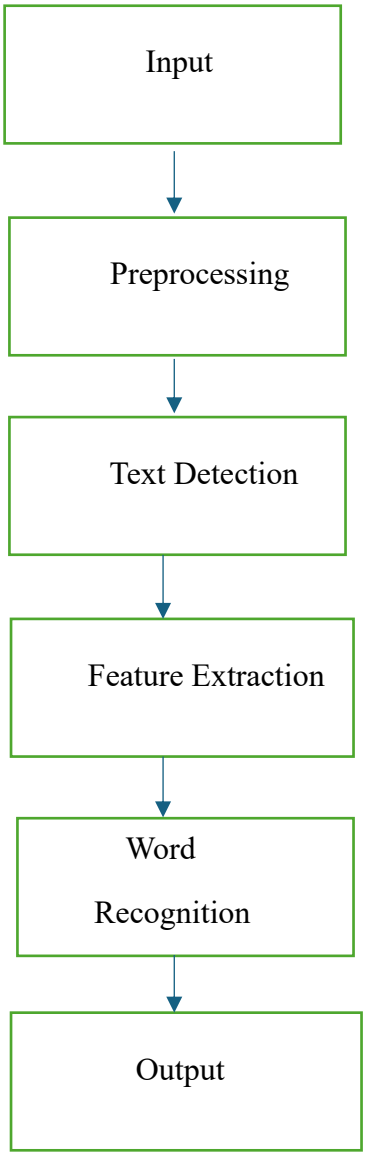

Fig.4.1 Flowchart

Pre-Processing

Pre-processing includes a set of steps on the input image to make it of higher quality and ready for segmentation accurately. The objective is to differentiate clearly the meaningful patterns from the background. Techniques such as data augmentation (the input image is copied and minute changes such as minimal rotation of the image are performed and both these images are passed through to the model in order to grow it's data examples), grey-scaling (images are converted to black and white so the model can perfectly identify if the input has handwriting on it or not.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM03136
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Feature Extraction

The process of identification of useful features in a machine learning problem may be time consuming and highly subjected to human prejudice but with Convolution Neural Networks, we are capable of finding features by itself by comparing related patterns in the images. In getting features from the DCNN model, to begin with we need to train the CNN network using the final sigmoid/logistic dense layer. The training network is responsible for determining the correct weight of the network through backward and forward iteration repeated as many times as it can until ultimately mean square error is minimized. We utilize MXNet to solve the issue for a particular problem and choose MSE (Mean Square Error) as the measure of evaluation. We shall be optimizing the model by trying to minimize MSE value with every new epoch.

Handwriting Recognition

The final model that is the handwritten text recognition model that takes in a line and converts the line into digital text. The model is built using a CNN-biLSTM architecture.

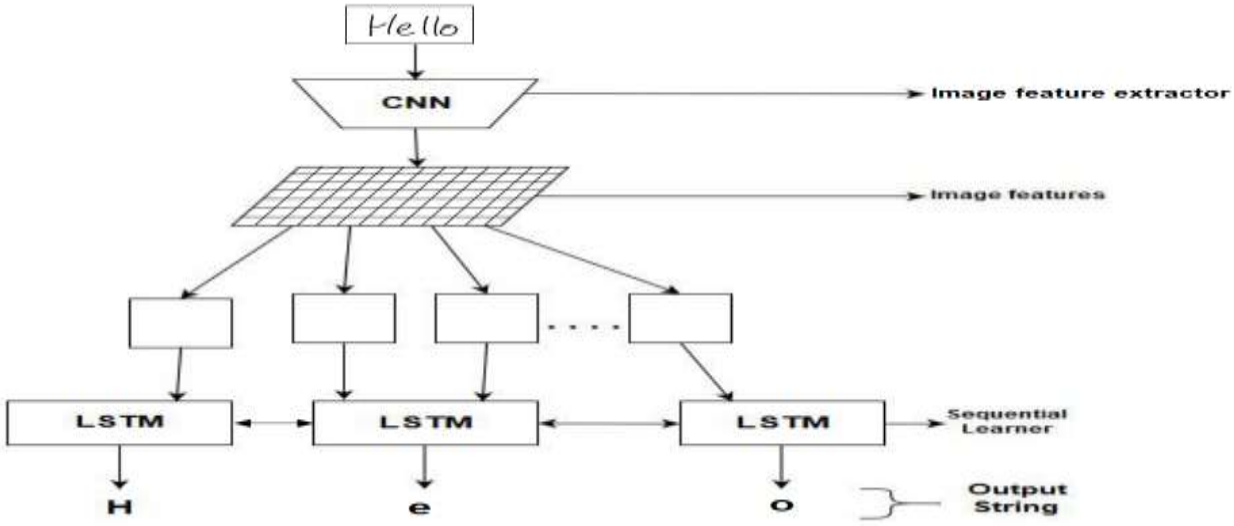

Fig.4.2 Handwritten Word Recognition

5. Results

Fig.5.1 Ground Truth Results

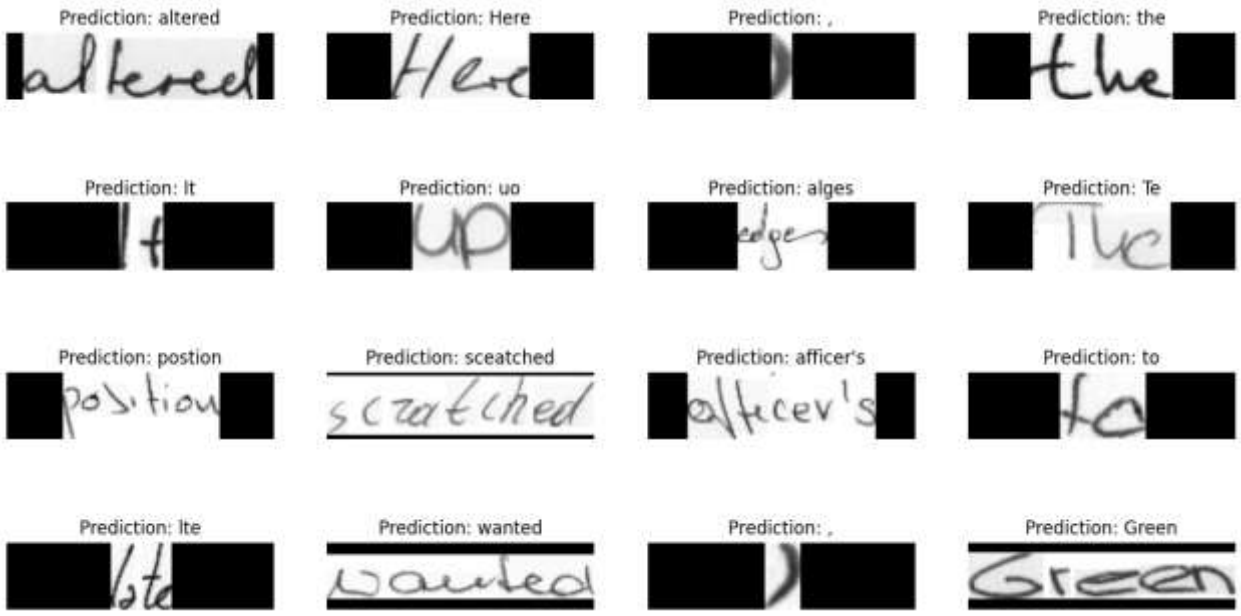

Fig.5.2 Predicted Output

6. Conclusion

This project demonstrates the effective use of deep learning techniques for converting handwritten text into digital format. By leveraging advanced models like Convolutional Neural Networks (CNNs) and datasets such as the IAM dataset, the system successfully performs tasks like word detection and recognition. The model achieves robust performance in handling diverse handwriting styles and layouts, showing its potential to streamline processes in document digitization, archival, and automation.

References

[1] ASGHAR ALI CHANDIO 1,2, MD. ASIKUZZAMAN 2,MARK R.PICKERING 2, (Member, IEEE), AND MEHWISH LEGHARI1

Cursive Text Recognition in Natural SceneImages Using Deep ConvolutionalRecurrent Neural Network

https://ieeexplore.ieee.org/document/9315679

[2] Handwritten Text Recognition and Conversion Using Convolutional Neural Network (CNN) Based Deep Learning Model

[3] A. Jain and U. Bhattacharya, "End-to-End Handwritten Text Recognition with Vision Transformers," presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022. Link: https://arxiv.org/abs/2202.12356

[4] R.-C. Chen, "Automatic License Plate Recognition Using Sliding-Window Darknet-YOLO Deep Learning Models," published in Image and Vision Computing, Vol. 87, July 2019, pp. 47–56.