

CNN-Based Handwritten Text Recognition Using Deep Learning Techniques

¹Mohith V K, ²Rajesh B R, ³Thanush, ⁴Manoj Y M, ⁵Mrs.Madevamma

dept. of Information Science & Enginnering Rajeev Institute of Technology Hassan 573201

Abstract - This paper presents a handwritten text recognition system using Convolutional Neural Networks (CNN), BiLSTM, and CTC loss. The proposed model performs image preprocessing, feature extraction, and sequence prediction to recognize handwritten text efficiently. The system is trained using the IAM Handwriting Dataset and implemented using TensorFlow. Experimental results demonstrate improved recognition accuracy and reduced error rates for offline handwritten text recognition applications.

Keywords— Handwritten Text Recognition, CNN, BiLSTM, CTC Loss, OCR, Deep Learning, TensorFlow.

Introduction

Handwritten Text Recognition (HTR) is the process of converting handwritten text present in images into machine-readable digital text. Unlike printed text, handwritten text contains large variations in writing style, spacing, orientation, and character structure, making recognition a difficult problem. Traditional Optical Character Recognition (OCR) systems relied on handcrafted feature extraction methods and statistical classifiers, which produced limited accuracy when dealing with complex handwriting styles.

Recent developments in deep learning have significantly improved the performance of handwriting recognition systems. Convolutional Neural Networks (CNNs) provide efficient spatial feature extraction, while Recurrent Neural Networks (RNNs), particularly Bidirectional Long Short-Term Memory (BiLSTM) networks, capture contextual dependencies in sequential data. The integration of Connectionist Temporal Classification (CTC) loss eliminates the need for explicit character segmentation. This work proposes a CNN-BiLSTM-CTC based handwritten text recognition system capable of recognizing handwritten words from images. The system is trained on the IAM Handwriting Dataset and deployed using ONNX Runtime for efficient inference.

Literature Survey Several researchers have contributed to the advancement of handwritten text recognition systems using deep learning architectures.

Baoguang Shi et al. proposed the CRNN architecture combining CNN and RNN layers with CTC loss for sequence recognition tasks. Their work demonstrated the effectiveness of end-to-end trainable neural networks for OCR applications. Hongjian Zhan et al. introduced a combination of Residual Networks and RNN-CTC architecture for handwritten digit string recognition. Residual CNN blocks improved feature extraction and overall recognition accuracy. Ahmed AL-Saffar proposed a dynamically configurable CRNN architecture optimized using metaheuristic algorithms, improving handwriting recognition performance on IAM datasets. Zahangir Alom demonstrated the effectiveness of CNN architectures for handwritten character recognition and showed that dropout improves generalization. Recent studies also explored transformer-based OCR systems and multilingual handwriting recognition models, indicating the growing importance of deep learning in OCR research.

Problem Statement

Despite advancements in OCR technologies, accurate handwritten text recognition remains challenging due to:

- Variability in handwriting styles
- Irregular spacing and alignment
- Image noise and distortion
- Lack of clear character segmentation
- Variable-length text sequences

Traditional OCR systems fail to handle these challenges effectively. Therefore, there is a need for an intelligent system capable of recognizing handwritten text without explicit character segmentation.

Proposed Methodology

The proposed system uses a hybrid deep learning architecture consisting of CNN, BiLSTM, and CTC loss.

A. 4.1 System Workflow

The overall workflow of the system includes:

1. Dataset Preparation
2. Image Preprocessing
3. CNN Feature Extraction
4. Sequence Conversion
5. BiLSTM Sequence Modelling
6. CTC-Based Prediction

- ONNX Inference Engine
- Evaluation Module

The input image is first preprocessed and passed through CNN layers for feature extraction. The sequential features are processed using BiLSTM layers, followed by CTC decoding to generate the predicted text.

7. ONNX Deployment

B. 4.2 Image Preprocessing

Input handwritten images are resized to fixed dimensions and normalized to improve model performance. The preprocessing steps include:

- Image resizing
- Pixel normalization
- Label encoding
- Data augmentation

C. 4.3 CNN Feature Extraction

The CNN layers extract spatial features from handwritten images. Residual convolutional blocks improve gradient flow and feature learning.

D. 4.4 Sequence Modelling Using BiLSTM

The extracted feature maps are reshaped into sequential vectors and processed using BiLSTM layers. Bidirectional learning captures dependencies in both forward and backward directions.

E. 4.5 CTC Loss Function

CTC loss enables alignment-free prediction and removes the requirement for explicit character segmentation.

$$L_{CTC} = -\ln P(y|x)$$
The CTC decoder converts probability sequences into readable text outputs.

System Architecture

The architecture of the proposed system consists of the following modules:

- Input Image Module
- Preprocessing Module
- CNN Feature Extractor
- BiLSTM Sequence Model
- CTC Decoder

Implementation

6.1 Development Environment

The proposed system is implemented using Python and TensorFlow.

- Software Requirements
- Python 3.x
- TensorFlow
- Keras
- OpenCV
- NumPy
- Pandas
- ONNX Runtime
- tf2onnx
- Hardware Requirements
- Minimum 8 GB RAM
- Multi-core processor
- Optional NVIDIA GPU with CUDA support

6.2 Dataset Preparation

The IAM Handwriting Dataset is used for training and evaluation. The dataset contains handwritten word images and corresponding text labels.

Dataset preparation steps include:

- Dataset extraction
- Annotation parsing
- Invalid data filtering
- Vocabulary construction
- Label padding

6.3 Model Training

The model is trained using supervised learning.

Training Configuration

- Optimizer: Adam
- Learning Rate: 0.001
- Batch Size: 128

- Epochs: 100
- Training Callbacks
- Early Stopping
- Model Checkpoint
- Tensor Board
- ReduceLROnPlateau

- High recognition accuracy
- No explicit character segmentation required
- Robust feature extraction using CNN
- Efficient sequence modelling using BiLSTM
- Alignment-free prediction using CTC
- Cross-platform deployment using ONNX
- Scalable architecture for real-world applications

II. RESULTS AND DISCUSSION

The proposed handwritten text recognition system achieved good recognition performance on the IAM Handwriting Dataset.

A. 7.1 Evaluation Metrics

The model performance is evaluated using:

- Character Error Rate (CER)
- Word Error Rate (WER)

$$\text{CER} = \frac{S+D+I}{N}$$

Where:

- S = Substitutions
- D = Deletions
- I = Insertions
- N = Total characters

Lower CER and WER values indicate better recognition performance.

B. 7.2 Experimental Results

The experimental analysis indicates:

- Improved recognition accuracy
- Efficient sequence prediction
- Faster inference using ONNX Runtime
- Better generalization through data augmentation

The model successfully recognizes handwritten text with good consistency and reduced error rates.

III. ADVANTAGES OF PROPOSED SYSTEM

The proposed system offers several advantages:

IV. LIMITATIONS

Some limitations observed during testing include:

- Reduced accuracy for unclear handwriting
- Dependency on image quality
- Fixed input size limitations
- Limited multilingual support

V. FUTURE ENHANCEMENTS

The system can be improved further by:

- Implementing transformer-based architectures
- Adding multilingual handwriting support
- Integrating attention mechanisms
- Improving real-time deployment
- Using beam search decoding
- Extending recognition to sentence and paragraph level

Conclusion

This paper presented a deep learning-based handwritten text recognition system using CNN, BiLSTM, and CTC loss. The system effectively recognizes handwritten words without requiring explicit character segmentation. The integration of CNN for feature extraction and BiLSTM for sequence modelling significantly improves recognition performance. Experimental results demonstrate the effectiveness of the proposed approach in handling variable handwriting styles and complex image patterns. The deployment using ONNX Runtime provides efficient and portable inference capabilities. The proposed system can be applied in document digitization, banking, healthcare, and educational applications.

VI. References



[1] B. Shi, X. Bai, and C. Yao, "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*, vol. 39, no. 11, pp. 2298–2304, 2017.

[2] H. Zhan, Q. Wang, and Y. Lu, "Handwritten Digit String Recognition by Combination of Residual Network and RNN-CTC," *IEEE Access*, vol. 7, pp. 10254–10261, 2019.

[3] A. AL-Saffar, "A Sequential Handwriting Recognition Model Based on a Dynamically configurable CRNN," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 6, pp. 202–210, 2022.

[4] M. Z. Alom et al., "Handwritten Bangla Digit Recognition Using Deep Learning," *Neural Computing and Applications*, vol. 33, no. 4, pp. 1235