Handwritten Text Recognition and Generation using Neural Networks

Project Title: NOTES AI

Sai Kumar R (22BLC1031)
Naveen M (22BLC1055)
Shaneel Reddy L (22BLC1164)

Abstract

Objective:
The primary goal is to enhance handwritten text recognition accuracy and automate handwritten text generation by using neural networks with **CNN** (Convolutional Neural Network) and **RNN** (Recurrent Neural Network) layers, especially with **LSTM** (Long Short-Term Memory) units for sequence handling.

Challenges:
Handwritten text comes with high variability in writing styles (e.g., cursive, printed), low-quality images (blurry or noisy), and background noise, which makes recognition difficult.

Key Results:
The model improves accuracy, especially in recognizing complex handwriting styles, compared to traditional methods that struggle with diverse handwriting.

Keywords

- **CNN (Convolutional Neural Network)**:
  CNN is a deep learning architecture primarily used for processing visual data. In this case, it extracts crucial character features from handwritten text images.
- **RNN (Recurrent Neural Network)**:
  RNN is designed for sequence-based data. It processes data sequentially (e.g., character-by-character or word-by-word) to understand the flow of handwriting.
- **LSTM (Long Short-Term Memory)**:
  An advanced type of RNN, LSTM is used to retain important sequence information over time, which is essential in recognizing handwriting that spans multiple characters or words.
- **Image Preprocessing**:
  Techniques such as resizing, noise removal, and binarization are used to enhance the quality of input images, making them more suitable for neural network processing.
- **IAM Dataset**:
  The IAM (Interactive Annotation Machine) dataset contains thousands of handwriting samples. It's commonly used to train handwriting recognition models.

Introduction

Handwriting Recognition Need:
Many documents, especially historical or personal records, are handwritten and need to be converted into a digital format for better accessibility, archiving, and analysis.

Traditional OCR Limitation:
OCR systems are effective for recognizing printed text but perform poorly with handwritten text, especially due to the large variation in handwriting styles.

Objective:
This research focuses on leveraging CNN, RNN, and LSTM networks to overcome the challenges posed by handwriting recognition, offering a more robust solution.

Problem Definition

- **Research Aim**:
  The aim is to create a model that can accurately interpret a wide variety of handwriting styles, making it capable of handling real-world handwritten documents.
- **Challenges Identified**:
  Handwriting variability, poor-quality images, and difficulties with real-time recognition are the key issues.
- **Goal**:
  Achieve high accuracy in recognizing diverse handwriting styles, making the model suitable for a wide range of real-world applications.

Related Works

Existing Methods:
Traditional OCR methods work well with printed texts (e.g., scanned books) but fail when dealing with handwritten texts due to differences in character shapes, spacing, and flow.

Previous Studies:
Prior research has attempted using CNNs and RNNs for handwriting recognition. However, many have struggled with variable handwriting styles and less-than-ideal data quality.

Our Approach:
The research builds on previous work by combining CNNs for feature extraction with RNNs and LSTMs for sequence prediction, creating a model that adapts better to different handwriting styles.

Model Architecture

Architecture Overview:
The model combines **CNN** for extracting image features (like character shapes) and **RNN** with **LSTM** for interpreting the sequential flow of characters to generate text.

Data Flow:
This diagram would typically show how the input image (handwritten text) is first processed by the CNN to extract features, then passed through the RNN layers (which capture the context of the text), and finally converted to textual output.

Layers' Functionality:

- **CNN**: Extracts local features like edges, curves, and strokes in the characters.
- **RNN**: Uses its recurrent structure to process the sequence of characters or words, understanding their relationships.

Data Collection and Preprocessing

Dataset:
The IAM dataset is a collection of handwriting samples annotated with the corresponding ground truth text. It is widely used for training handwriting recognition systems.

Preprocessing Steps:

- **Standardize Size**: Resize images to ensure consistency.
- **Enhance Clarity**: Techniques like contrast adjustment to make text clearer.
- **Binarization**: Convert the image to black and white, making text more distinct from the background.

Importance:
Preprocessing is crucial as it reduces noise and ensures that the neural network receives clean, uniform input data for better accuracy.

Model Training Process

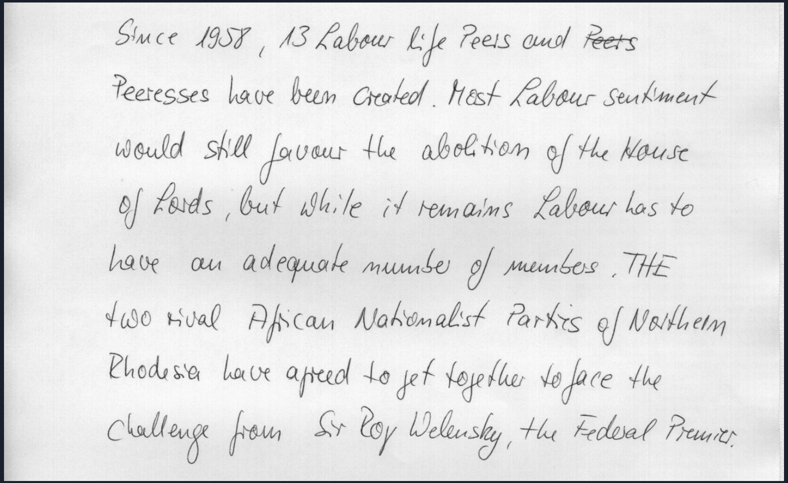
CNN and RNN Layer Training:
Each layer of the model is trained using the dataset. The CNN layers learn to extract relevant features from the images, while the RNN layers focus on the sequence of the characters.

Optimization Technique:
Gradient Descent is used to minimize the loss function, helping the model learn the correct weights to accurately predict the text.

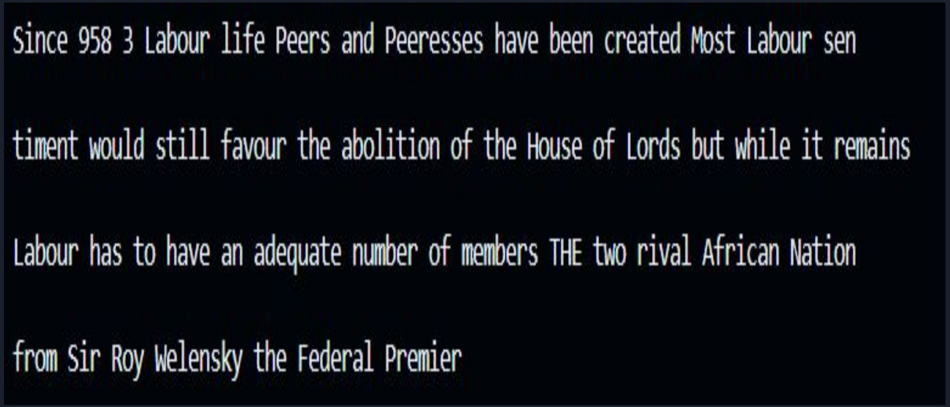
CTC Loss Function:
Connectionist Temporal Classification (CTC) loss function is used in sequence-to-sequence tasks where alignment between input and output is unknown, making it ideal for handwriting recognition where character segmentation isn't explicitly provided.

Results and Discussion

INPUT IMAGE


Since 1958, 13 Labour life Peers and ~~Peers~~
Peeresses have been created. Most Labour sentiment
would still favour the abolition of the House
of Lords, but while it remains Labour has to
have an adequate numbe of members. THE
two rival African Nationalist Parties of Northern
Rhodesia have agreed to get together to face the
challenge from Sir Roy Welensky, the Federal Premier.

GENERATED OUTPUT


Since 958 3 Labour life Peers and Peeresses have been created Most Labour sen
timent would still favour the abolition of the House of Lords but while it remains
Labour has to have an adequate number of members THE two rival African Nation
from Sir Roy Welensky the Federal Premier

Results and Discussion

```
Epoch: 1 | train_loss: 0.130 | val_loss: 0.013
Epoch: 2 | train_loss: 0.010 | val_loss: 0.009
Epoch: 3 | train_loss: 0.009 | val_loss: 0.009
Epoch: 4 | train_loss: 0.009 | val_loss: 0.009
Epoch: 5 | train_loss: 0.009 | val_loss: 0.009
Epoch: 6 | train_loss: 0.009 | val_loss: 0.009
Epoch: 7 | train_loss: 0.009 | val_loss: 0.009
Epoch: 8 | train_loss: 0.009 | val_loss: 0.009
Epoch: 9 | train_loss: 0.009 | val_loss: 0.009
Epoch: 10 | train_loss: 0.009 | val_loss: 0.009
Epoch: 11 | train_loss: 0.009 | val_loss: 0.009
Epoch: 12 | train_loss: 0.009 | val_loss: 0.009
Epoch: 13 | train_loss: 0.009 | val_loss: 0.009
Epoch: 14 | train_loss: 0.009 | val_loss: 0.009
Epoch: 15 | train_loss: 0.009 | val_loss: 0.009
Epoch: 16 | train_loss: 0.009 | val_loss: 0.009
Epoch: 17 | train_loss: 0.009 | val_loss: 0.009
Epoch: 18 | train_loss: 0.009 | val_loss: 0.009
Epoch: 19 | train_loss: 0.009 | val_loss: 0.009
Epoch: 20 | train_loss: 0.009 | val_loss: 0.009
Total Training Time : 660.581seconds
```

- The training logs indicate that the NotesAI model achieved rapid convergence in both training and validation losses. Initial epochs show a significant drop in training loss, starting from 0.130 in Epoch 1 and stabilizing at around 0.009 from Epoch 4 onward. The validation loss follows a similar trend, quickly stabilizing at 0.009 by the second epoch and remaining consistent throughout the training process.

- The low and stable validation loss suggests that the model generalizes well to unseen data, indicating minimal overfitting. This implies the CNN-LSTM architecture effectively learned to recognize patterns in the handwritten text dataset, likely due to the strength of CNNs in feature extraction and the ability of LSTMs to capture sequential dependencies.

Results and Discussion

Strengths:
The model performs well even on complex and diverse handwriting, demonstrating its adaptability.

Limitations:
It may struggle with very low-quality images or handwriting with extreme stylization (e.g., artistic scripts).

Future Enhancements:
Future work could focus on improving the model's flexibility by training it on a more diverse set of handwriting samples and improving preprocessing techniques for degraded images.

Conclusion

Summary:
The research successfully developed a neural network model that can accurately recognize handwritten text, addressing the challenges of handwriting variability.

Real-World Impact:
This model could be used to digitize handwritten archives, making historical records more accessible and improving accessibility for people with disabilities.

Next Steps:
Focus on improving model flexibility, enhancing real-time recognition capabilities, and exploring practical applications in fields like document digitization and education.

References

1. Khalkar, Rohini & Dikhit, Adarsh & Goel, Anirudh. (2021). Handwritten Text Recognition using Deep Learning (CNN & RNN). IARJSET. 8. 870-881. 10.17148/IARJSET.2021.86148.
2. Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. Physica D Nonlinear Phenomena, 404, 132306. 10.1016/j.physd.2019.132306
3. Yuming He (2020) Research on text detection and Recognition based on OCR recognition technology. (n.d.). 10.1109/ICISCAE51034.2020.9236870
4. Hemanth G R, Jayasree M, Keerthi Venii S, Akshaya P, Saranya R (2021). ICTACT Journals - view articles. (n.d.). 10.21917/ijsc.2021.0351

Thank You !