**Handwritten Text Recognition Using a CRNN Model**



**DLP Project Report**

**Section: BCS – 8A**

21K-3838 Arqam Habib  
 **Course Instructor:** Ms. Summaiya Zahid

1. **Objective:**

The primary objective of this project is to develop an efficient Handwritten Text Recognition system capable of transcribing handwritten text from images. The system leverages a simplified Convolutional Recurrent Neural Network model with Connectionist Temporal loss to recognize text from the IAM Words dataset. By reducing the model complexity to four convolutional layers and a single LSTM layer, the project aims to achieve accurate transcription with lower computational requirements, making it suitable for resource-constrained environments.

1. **Problem Statement:**

Handwritten text recognition poses significant challenges due to variability in handwriting styles, image quality, and text alignment. Traditional HTR systems require complex models with deep convolutional and recurrent layers, which can be computationally ex pensive and prone to overfitting on smaller datasets like IAM Words. The problem is to design a simplified CRNN model that:

1. Accurately transcribes handwritten text from grayscale images.
2. Reduces model complexity to minimize computational resources.
3. Generalizes well to diverse handwriting styles in the IAM dataset.
4. **Methodology:**

The methodology encompasses data preparation, model design, training, and evaluation as implemented in the provided Jupyter Notebook.

**3.1 Dataset:**

The IAM Words dataset is used, containing grayscale images of handwritten words and corresponding transcriptions. Metadata in words.txt provides image paths and labels. A subset of 30,000 records is processed, split into 90% training (18,074 samples) and 10%

validation (2,007 samples).

* 1. **Data Preprocessing**

**Images:** Resized to (32, 128, 1), inverted, normalized to [0, 1], and padded if necessary.

**Labels:** Encoded as indices based on a 78-character set (punctuation, digits, letters).

**CTC Parameters:** Input length set to 32 timesteps, label length set to the word length.

* 1. **Model Architecture**

The simplified CRNN model is defined as follows:

**Input**: Grayscale images of shape (32, 128, 1).

**CNN Layers:**

1. Conv2D (64 filters, 3x3, ReLU, same padding) + MaxPool2D (2x2, stride 2).

2. Conv2D (128 filters, 3x3, ReLU, same padding) + MaxPool2D (2x2, stride 2).

3. Conv2D (256 filters, 3x3, ReLU, same padding) + Batch Normalization + MaxPool2D (2x1).

4. Conv2D (512 filters, 3x3, ReLU, same padding) + Batch Normalization.

**Output shape:** (1, 32, 512).

Squeeze: Lambda layer with tf.squeeze to produce (32, 512).

**RNN Layer:** Single LSTM (256 units, return sequences, 20% dropout).

**Output:** Dense layer with 79 units (78 characters + 1 blank) and softmax activation, producing (32, 79).

The model outputs a sequence of probabilities for CTC loss.

* 1. **Training**

The model is wrapped in a CTC training model using Keras, with inputs for images, labels, input length (32), and label length. It is compiled with the Adam optimizer and CTC loss.

1. **Results:**

The simplified CRNN model successfully transcribes simple words from the IAM Words dataset, as demonstrated in the notebook’s visualization.

Key observations:

**Accuracy**:

1. Training Accuracy: 95%
2. Validation Accuracy: 64%



