Team members

Nitya Tyagi, Akash Singh Yadav. We both are currenting pursuing Integrated Masters (4th year) at IIT Roorkee.

ABOUT THE PROJECT

An OCR (Optical Character Recognition) model refers to a neural network-based system designed to recognize and extract text information from images. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are commonly used architectures in OCR models.

Process

1.Getting Data

File was unzipped and stored, since the images were in jpg, they were converted to png for better quality image. Seeing how many images and unique characters we have.

2.Preprocessing

Mapping characters to integers-> define custom train split function-> convert to grey scale to reduce processing-> create dataset object.

3.Visualization of the data

4.Model

Creating CTC layer, defining loss function. We reshaped the data to be compatible with the loss function.

Model Architecture: The OCR model architecture is defined using the Keras functional API. The model consists of two convolutional blocks followed by max-pooling layers. Bidirectional LSTM layers are used for capturing sequential information. The output layer is a dense layer with a softmax activation function to predict character probabilities. The build_model function creates an instance of the OCR model. It compiles the model using the Adam optimizer.

5.Training

early_stopping_patience: This variable defines the number of epochs with no improvement in the monitored metric after which training will be stopped.

Why this model was chosen. OCR often involves dealing with variability in fonts, text sizes, orientations, and other visual characteristics. OCR benefits from bidirectional RNNs, which process sequences in both forward and backward directions.