CSE472
Machine Learning Sessional

Handwritten Text Recognition

<u>Submitted By</u>
Souvik Das - 1705034
Md. Musharaf Hossain - 1705050

<u>Supervised By</u>
Md. Tareq Mahmood
Assistant Professor, BUET

Problem Definition:

Handwritten text recognition is the task of converting handwritten text images into machine-readable text. This project aims to develop a deep learning model using TensorFlow to accurately recognize and transcribe handwritten text from images. The project will use two different models for comparison: a ResNet and a simple CNN.

The ResNet is a deep neural network that uses residual blocks to enable deeper architectures while mitigating the vanishing gradient problem. The simple CNN, on the other hand, is a standard convolutional neural network with fewer layers. The goal of this project is to compare the performance of these two models for handwritten text recognition.

The main challenge in this project is to accurately recognize the text despite the variability in the input images. Handwritten text can have variations in writing style, slant, thickness, and spacing, which can make recognition difficult. Both models need to be trained on a large dataset of handwritten text images with various styles and variations to be able to generalize well and accurately transcribe unseen text.

Overall, the goal of this project is to compare the performance of ResNet and a simple CNN for handwritten text recognition using TensorFlow. The models should be evaluated on two metrics. They are CER (Character Error Rate) and WER (Word Error Rate).

Dataset:

From the **IAM Handwriting Database**, we are using handwritten **lines**.

The IAM Lines Dataset is a benchmark dataset for handwritten text recognition that is widely used in research. It consists of a collection of more than 13,000 text lines from more than 500 writers in English, taken from various historical and contemporary sources.

Each line is provided as a grayscale image with a resolution of 300dpi and a size of approximately 1024x128 pixels. The dataset also includes a transcription file for each line, which provides the corresponding ground truth text.

The IAM Lines Dataset is a challenging benchmark for handwritten text recognition due to the variability in writing styles, line spacing, and text layout. It is often used to evaluate the performance of deep learning models for handwritten text recognition, including models based on convolutional neural networks (CNNs) and recurrent neural networks (RNNs).

Many researchers use the IAM Lines Dataset as a standard benchmark for their handwritten text recognition models. The dataset is freely available for research purposes, and there are several open-source implementations of models that have achieved state-of-the-art performance on this dataset.

Proposed Solution:

For this handwritten text recognition project, we propose to develop and compare two deep learning models: a ResNet and a simple CNN. The ResNet is a deep neural network that uses residual blocks to enable deeper architectures while mitigating the vanishing gradient problem. The simple CNN, on the other hand, is a standard convolutional neural network with fewer layers.

The ResNet Model:

Residual Block:

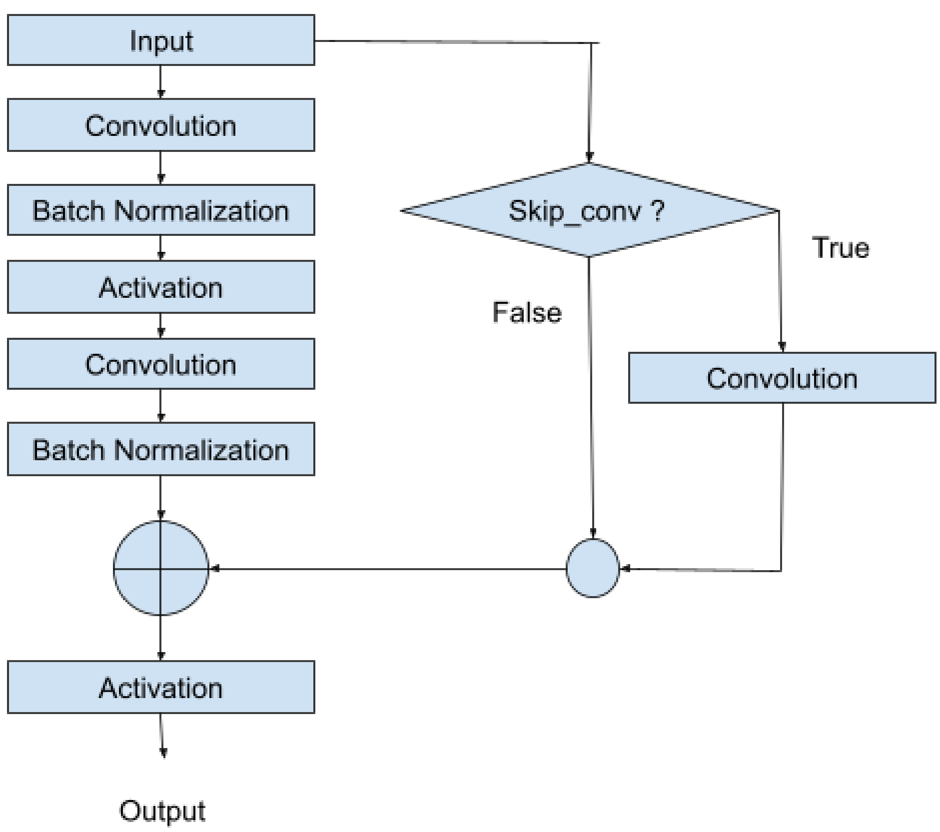

Figure-1: Residual Block

ResNet Model:

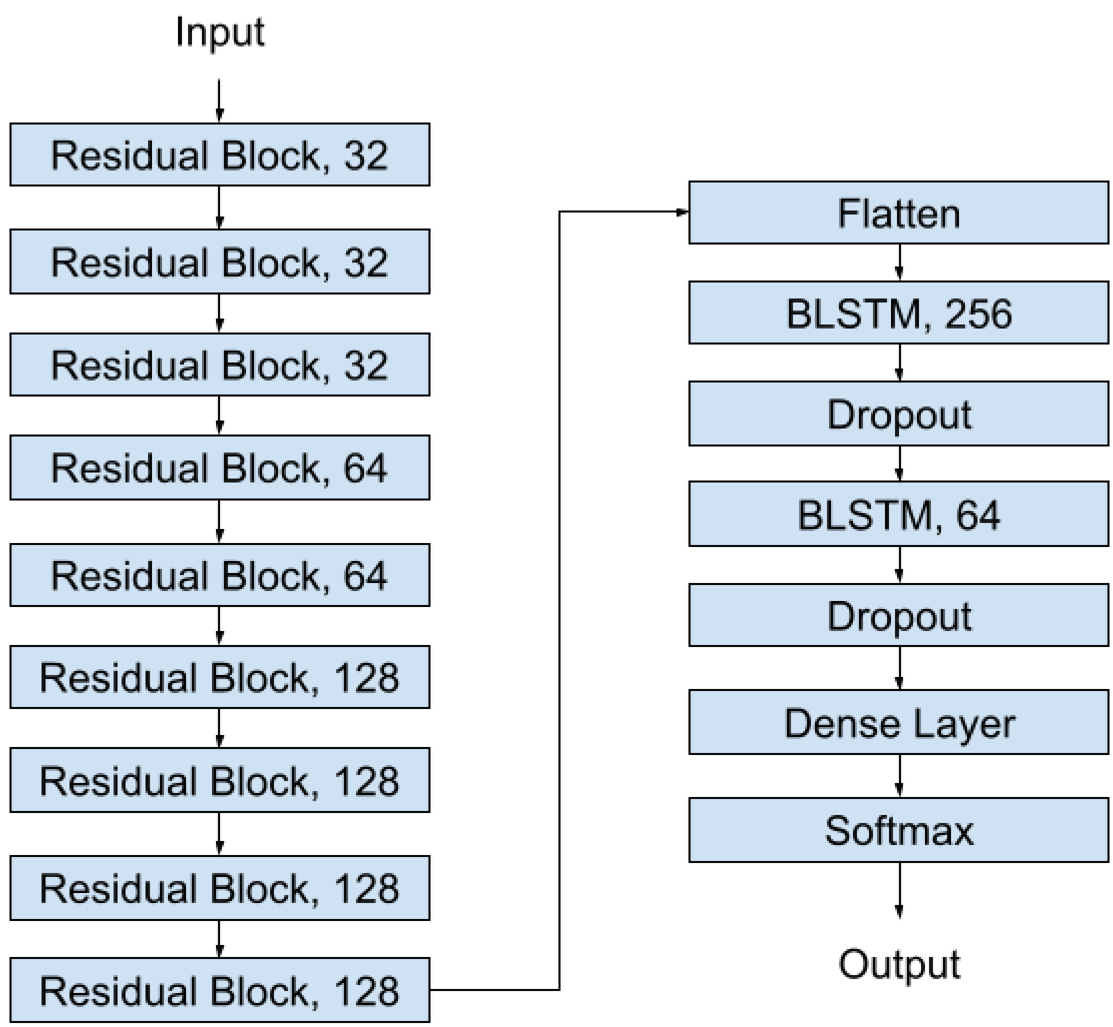

Figure-2: ResNet Model

CRNN Model:

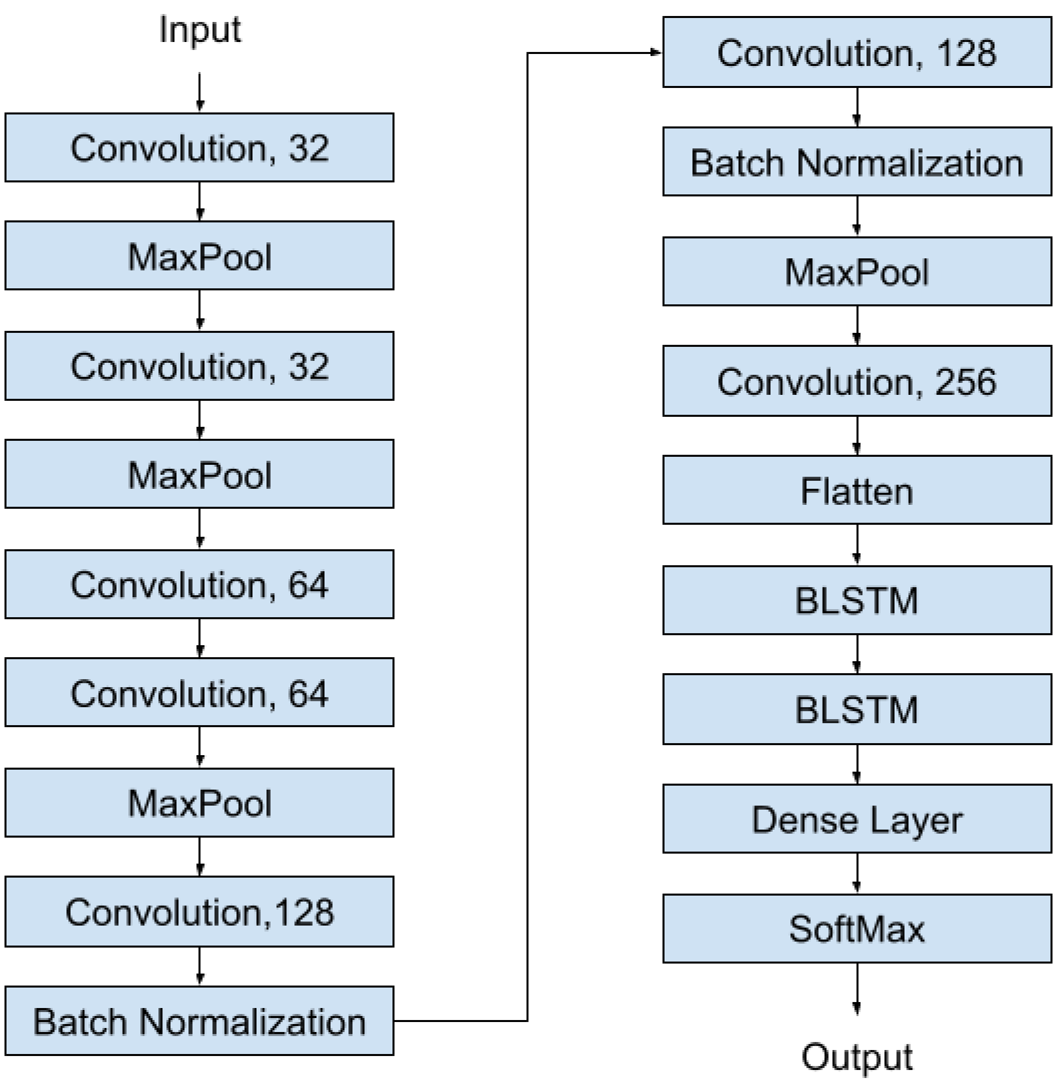

Figure-3: CRNN Model

Both models will be trained on a dataset of handwritten text images using TensorFlow. The dataset will be split into a training set, and a validation set. The training set will be used to train the models, the validation set will be used to tune the hyperparameters of the models and to evaluate the performance of the models.

The models will be evaluated on two metrics: Character Error Rate (CER) and Word Error Rate (WER). CER is the percentage of characters that are incorrectly recognized in the predicted text, while WER is the percentage of words that are incorrectly recognized. Both CER and WER are widely used metrics for evaluating the performance of handwriting recognition systems.

We will also perform a qualitative analysis of the models' predictions to identify common errors and areas for improvement. This will involve analyzing the predicted text and comparing it to the ground truth text to identify errors in character recognition and word segmentation.

Overall, the proposed solution is to develop and compare two deep learning models for handwritten text recognition, using TensorFlow and evaluation with CER and WER metrics.

Loss function:

In this project **CTC** Loss function is used.
CTC (Connectionist Temporal Classification) loss is a loss function commonly used in sequence-to-sequence tasks such as handwriting recognition. It works by calculating the probability of each possible alignment between the input sequence and the output sequence and then maximizing the probability of the correct output sequence over all possible alignments. CTC loss is designed to handle variable-length inputs and outputs, making it a powerful and effective tool for neural network-based handwriting recognition systems.

Since in a line, the number of characters is not fixed, and there are many possible alignment between the input sequence, CTC loss function is a good choice for the project.

Performance Report:

CER and WER values for both the models are given in the table below.

Model	Character Error Rate (CER)	Word Error Rate (WER)
ResNet Model (after 55 epoch)	2.09%	8.097%
CRNN Model (after 65 epoch)	4.62%	18.18%

As we train the models, loss, CER, WER values gradually decrease. Keeping these values along with y-axis and number of epochs in x-axis, some graphs are plotted. They are given below.

ResNet:

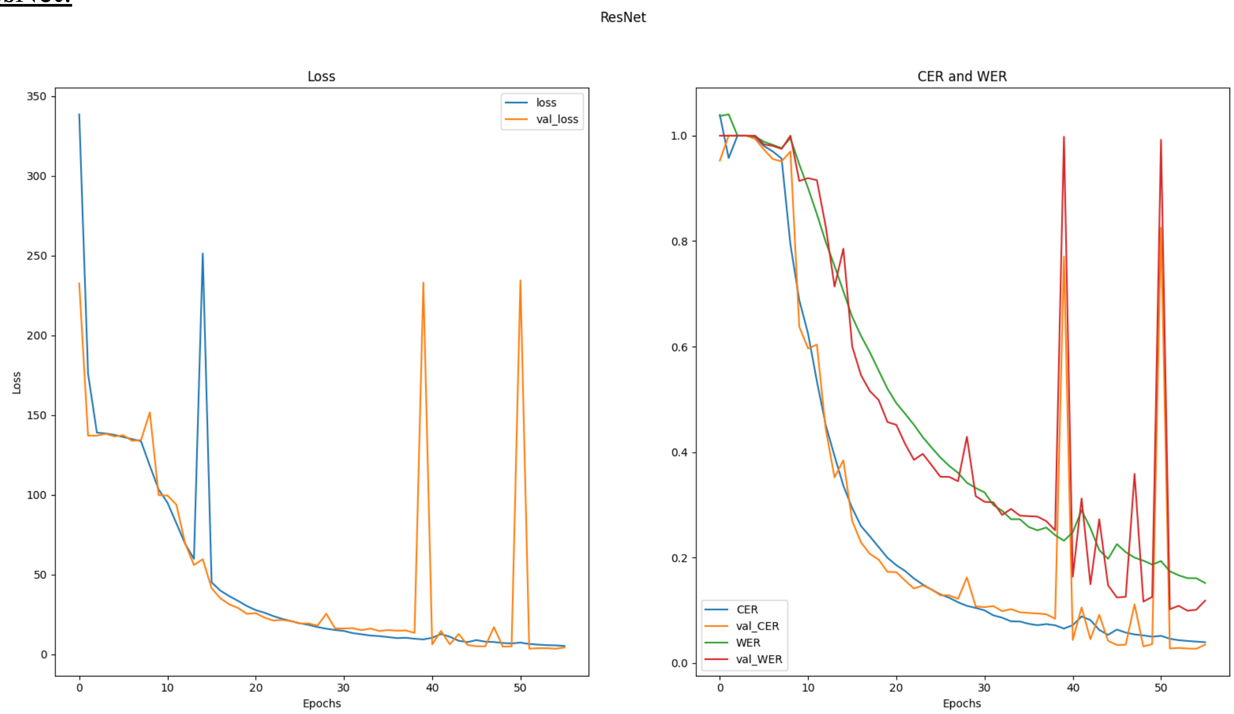

Figure-4: Loss, CER, WER of ResNet Model as number of epoch increases

CRNN:

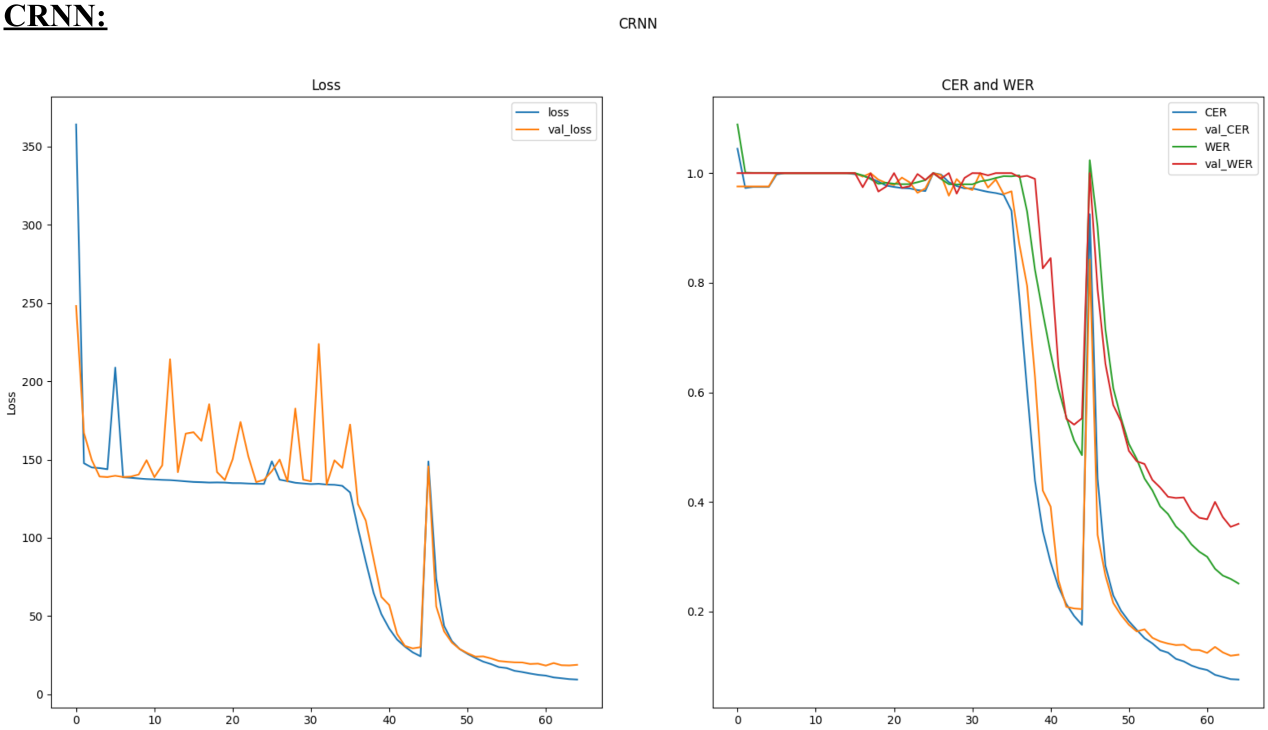

Figure-4: Loss, CER, WER of CRNN Model as number of epoch increases

Comparison:

Comparison between the two models is based on the performance measure Loss, CER and WER values are given below.

Loss: As we can see in the figure (Figure-5), loss is converging after 8 epochs for the ResNet Model, and after 35 epochs for the CRNN Model. Therefore, the ResNet model is more useful according to this measure.

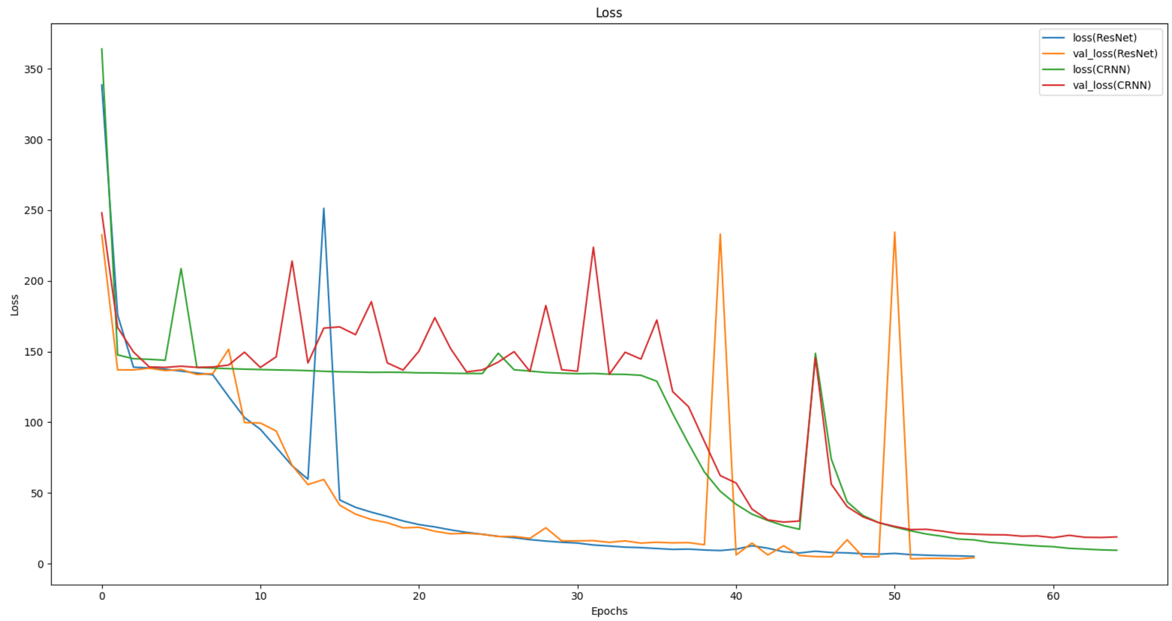

Figure-5: Loss comparison between the two models

CER: From the figure (Figure-6), CER is converging after 10 epochs for the ResNet Model, and after 38 epochs for the CRNN Model. Therefore, the ResNet model is more useful according to this measure.

WER: We can see from the figure (Figure-7), From the figure (Figure-4), WER starts decreasing after 10 epochs for the ResNet Model, and after 38 epochs for the CRNN Model. Therefore, the ResNet model is more useful according to this measure.

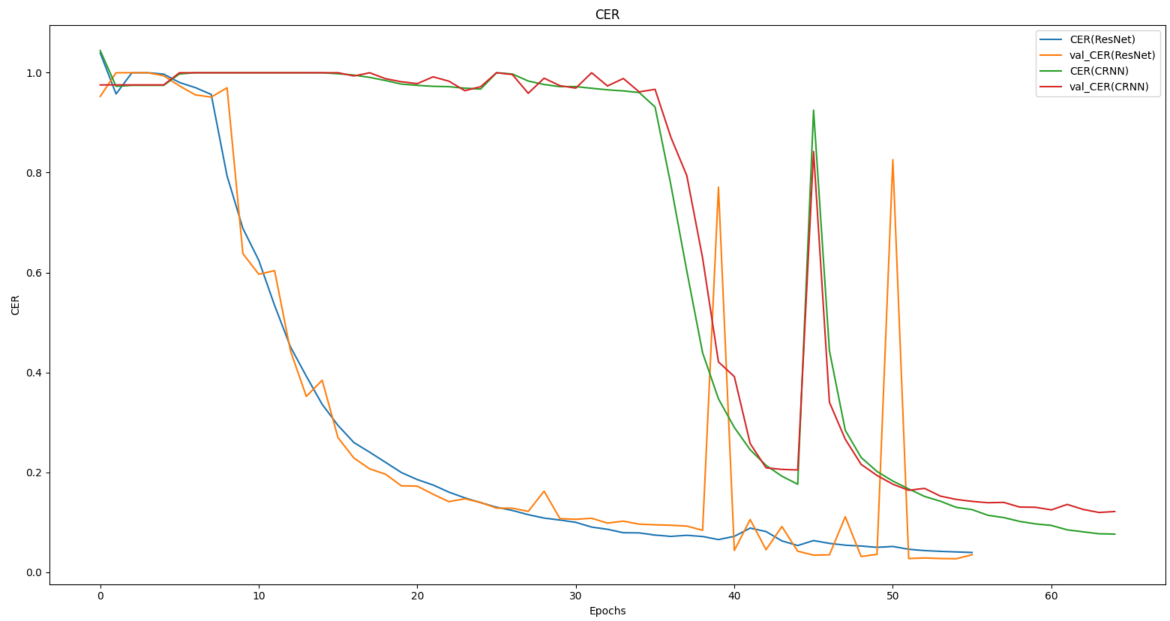

Figure-6: CER comparison between the two models

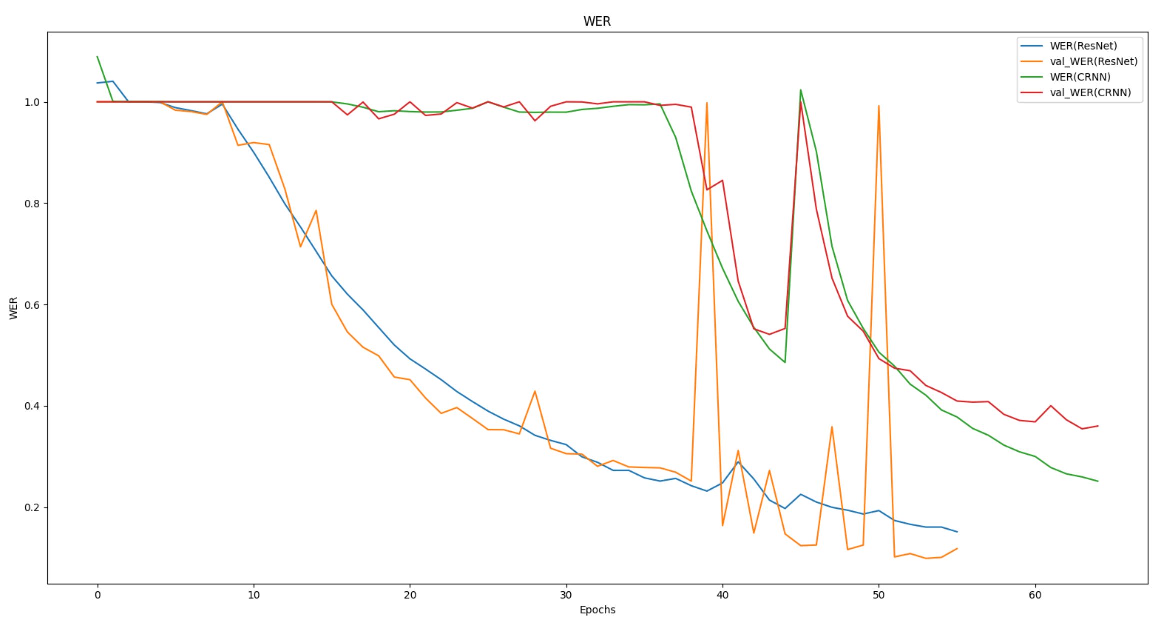

Figure-7: WER comparison between the two models

Though in every performance measure, ResNet is better, it is notable that CRNN is a less weighted model compared to ResNet. CRNN has almost half learning parameters than that of the ResNet model. But learning time for each epoch is almost the same. So, ResNet is more useful to recognize handwritten text.

Challenges:

- Time taken for training was the main challenge for us.
- Though we got a GPU, we couldn't use it to its max limit because of various dependencies among python, tensorflow, CUDA, cuDNN versions.
- To run the code in our local machine, we had to make a lot of changes in the original codebase.
- Python does not have backward compatibility. Some packages are for specific python versions. Packages have dependencies among them. But they depend on specific versions of other packages. Installing each package so that they are all compatible with each other was very much challenging.

Github Repository:

Link: https://github.com/musharaf-hossain-cs/ml-project-handwritten-text-recognition