Recognition of Telugu Ancient Characters And Information Retrieval From Temple Epigraphy Using Deep Learning

VEENA SAI NIGAMA
STUDENT ID: 1129215
SEPTEMBER 2024

Agenda

- Introduction
- Aims & Objectives
- Literature Review
- Problem Statement
- Methodology
- Results & Discussion
- Conclusion & Future works

Introduction

- Inscriptions on temple walls hold significant historical, cultural, and religious value. These inscriptions, often carved in stone, provide insights into the period during which the temple was constructed, the rulers or patrons who commissioned the temple, and the religious practices and social life of the time

Aims & Objectives

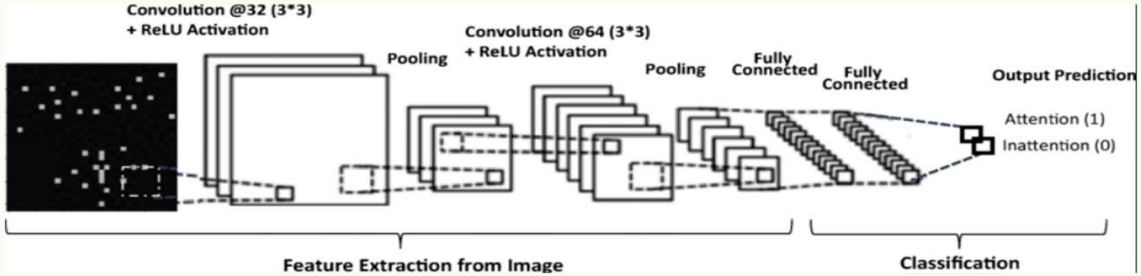
Understand an Image

Using CNNs to learn the image features

Transliterate into user friendly language

Train the model to understand the text and transliterate the test to user friendly language.

Literature Review



Computer vision has witnessed significant advancements in recent years, transforming how machines perceive and interpret visual data. Here are some key developments driving the progress in this field:

1.Deep Learning and Convolutional Neural Networks (CNNs): The advent of deep learning, especially CNNs, has revolutionized image recognition and processing. CNNs excel at recognizing patterns in images, allowing for breakthroughs in tasks like object detection, classification, and segmentation.

Literature Review

1.Vision Transformers (ViT): Transformers, initially developed for natural language processing, have been adapted for vision tasks through the Vision Transformer model. ViT models have demonstrated impressive results, especially in tasks like image classification and understanding complex visual patterns.

1.Pretrained Models and Transfer Learning: Large-scale pretrained models, such as ResNet, VGG, and EfficientNet, allow researchers to leverage vast amounts of data to fine-tune models for specific tasks with smaller datasets. Transfer learning has reduced the need for extensive labeled data, making AI accessible for a broader range of applications.

Literature Review

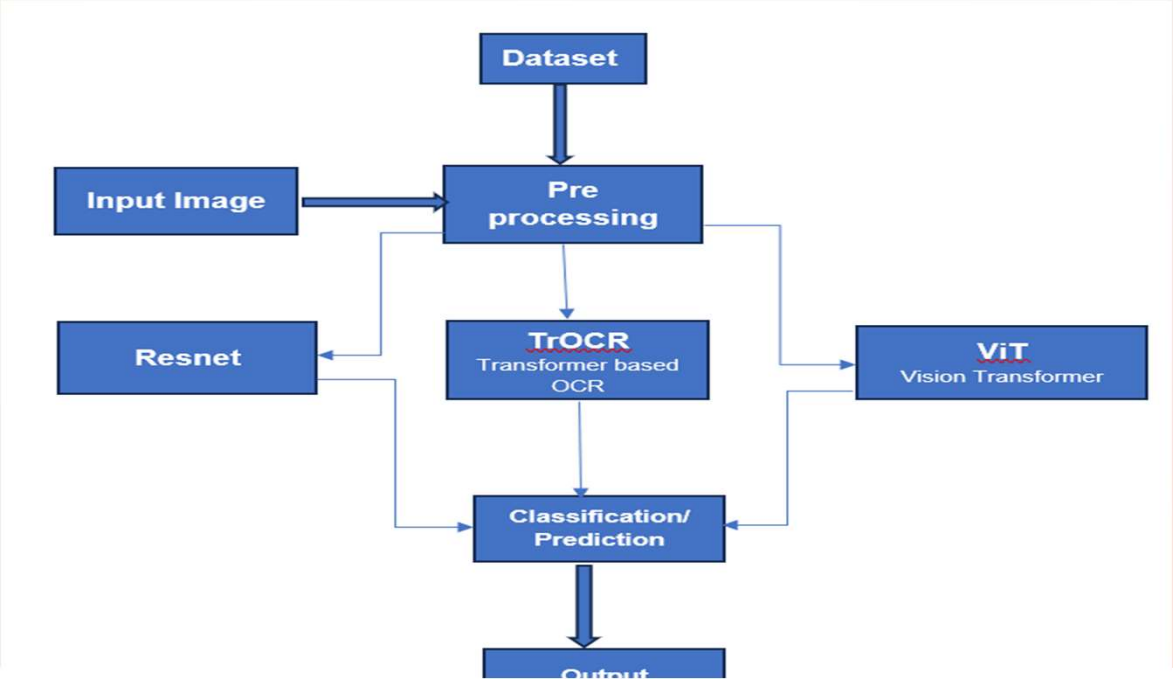
After studying various papers and knowing about the various techniques used for optical character recognition, I decided the workflow for the project. The image is taken as input by reading or scanning and is converted as a image of digital type. Enhancement techniques like binarization can be helpful for reducing noise in the image. Segmentation techniques like lie, word or character segmentation are used to extract the main components of the image. With the help of feature extraction important features are extracted leaving behind the undesired. Then a convolution neural network can be used for training, classification, and recognition of old Telugu letters. For implementing various methods of recognizing letters, multiple neural networks are analyzed and compared to fetch the best performing model

Problem Statement

The main aim of this research is to develop the robust system for extraction and recognizing telugu handwritten characters from images of ancient temple inscriptions. Using advanced neural network techniques like CNNs, TrOCR(Transformer based OCR), ViTs(Vision transformers), the goal is to accurately digitize and archive these inscriptions, preserving valuable historical, cultural, and linguistic information.

Methodology

The project will involve acquiring high-resolution images of temple inscriptions and enhancing their quality through preprocessing techniques such as noise reduction and binarization. Characters will be isolated using segmentation methods. Convolutional neural networks (CNNs), and Transformer based OCR models will be employed to recognize the handwritten characters



Dataset

Database is simply a collection of similar information or data that is modified, stored and transmitted. For this model set of images are fetched from dataset mentioned below. I organize digital images into central location for speed sharing. The image data usually comes in different forms, like video sequences, multi-dimensional data extracted from a medical scanner, or view from various cameras at various different angles. Hand-written images of 10 telugu characters each with different variations in size, angle, background, etc. I have taken 10 different styles for each character and created a database of 100 samples from the dataset

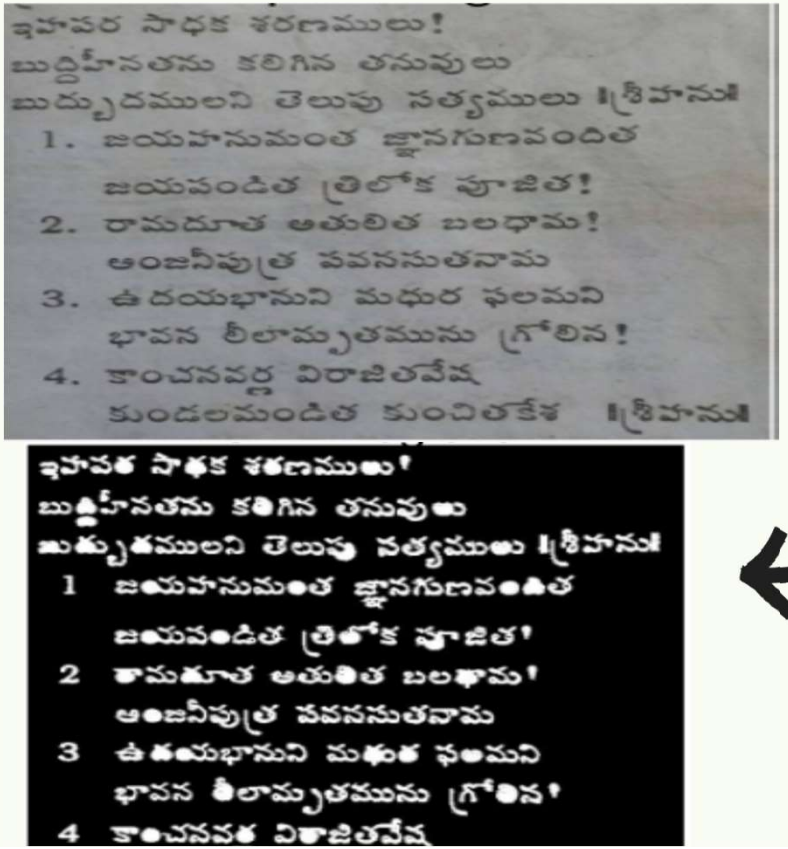
Dataset consists of all Telugu handwritten characters taken from below IEEE site.

https://ieee-dataport.org/open-access/telugu-handwritten-character-dataset

File name of each character is embedded with its 'English' transliterated form.

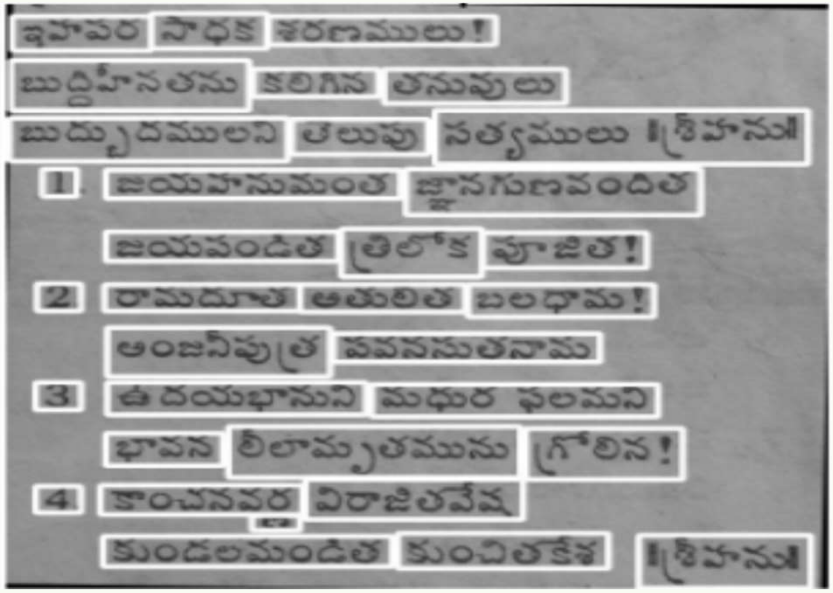
Data Preprocessing

Segmentation- Thresholding



In this stage we basically try to reduce noise or color in the image. Since historical inscriptions on walls of ancient temples are of poor quality often, the scanned images might have some disturbance or noise that needs to be improvised in order to make the recognition of Telugu characters very easy and more efficient. The given image is initially changed into grayscale by following threshing method.

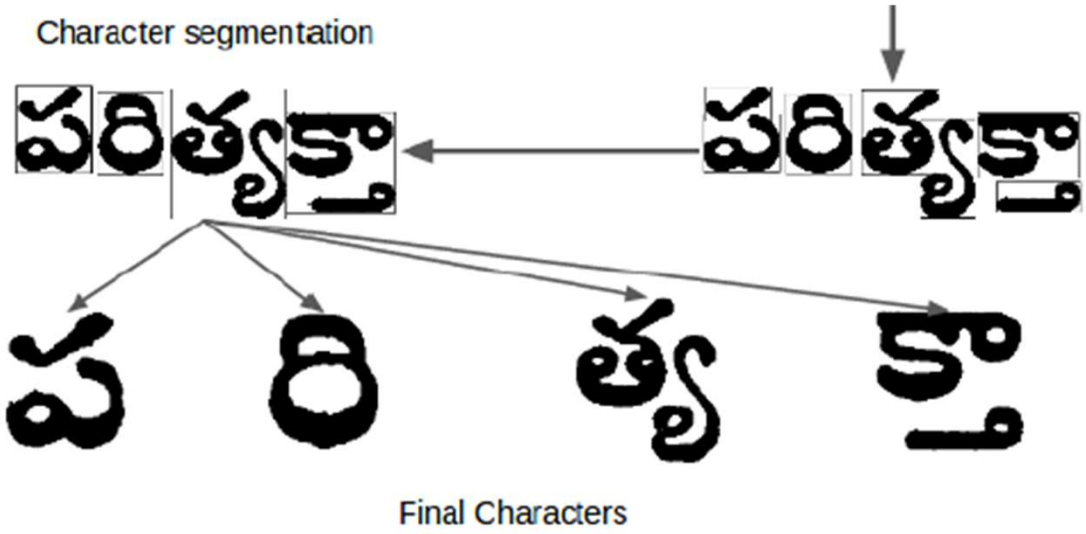
Segmentation



Word segmentation from the contour regions

After image binarization a segmentation technique is applied. Segmentation is the technique for isolating text in an image. Basically, it helps to separate graphics like the image of tree, or any other symbols from remaining text in an image. It also helps to contrast one text from other, based on the criteria that processing a single word or single character at a time is usually faster than the processing of the entire image at once

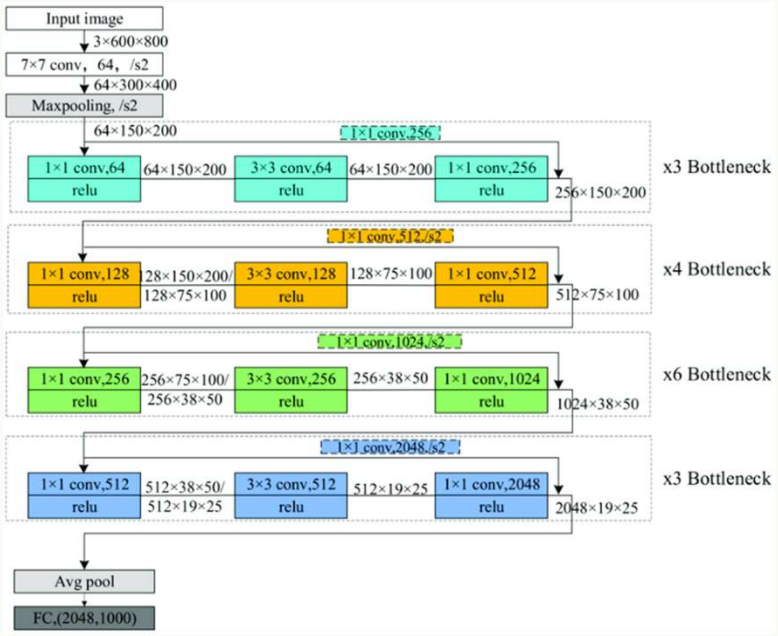
Segmentation



Character segmentation

Character Segmentation is the method to extract characters only from a word. Character Segmentation is quite difficult for this project as the characters of Inscriptions are difficult to be recognised and to obtain a meaningful text from image. This stel decompose a line of text into each character.

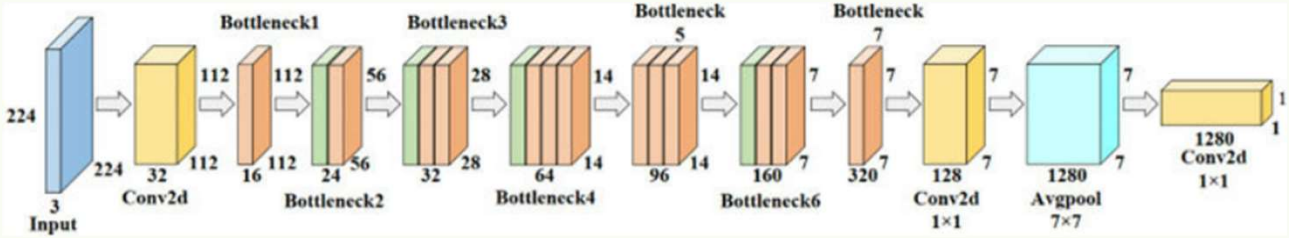
Classification-Resnetv2 50



Architecture

ResNet-50 is a deep CNN architecture which is known for introducing residual learning through residual blocks, which will utilize skip connections to mitigate the vanishing gradient problem, allowing for the training of deeper networks. Comprising 50 layers, including convolutional layers and batch normalization, ResNet-50 is organized into multiple groups of residual blocks that progressively learn abstract features.

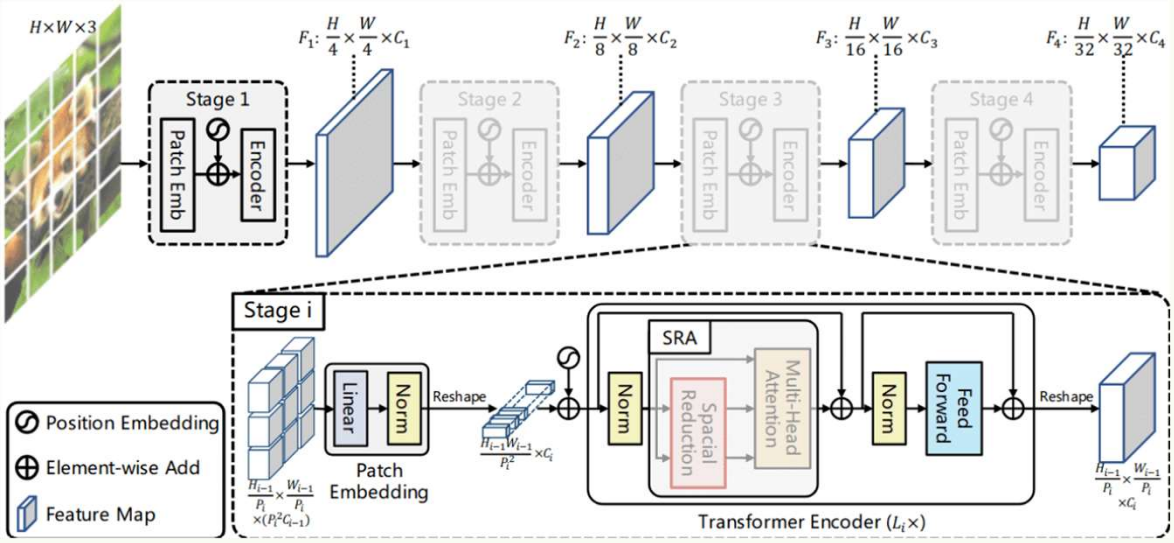
Classification- Mobilenet v2



Architecture

An effective deep neural network architecture called MobileNetV2 was created for contexts with limited resources and mobility. Two major advances are introduced: linear bottlenecks and inverted residual blocks. In contrast to conventional residual blocks, inverted residual blocks raise the number of channels again after first compressing the input with a pointwise convolution, then applying a depthwise convolution.

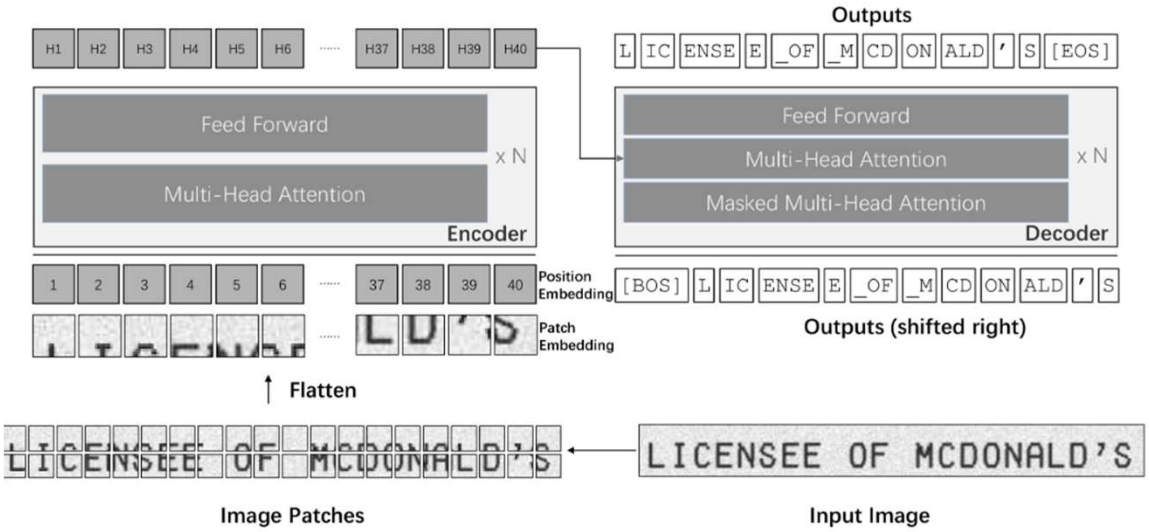
Classification- ViT



Vision Tranformer(ViT)

ViT employs patch-based input representation, where an image is divided into fixed-size patches (e.g., 16x16 pixels) instead processing entire images using convolutional layers detect features hierarchically. Each patch is converted into a vector and handled like "token," much like how NLP transformers handle the words. The transformer model is fed this series of patch tokens.

Classification-TrOCR



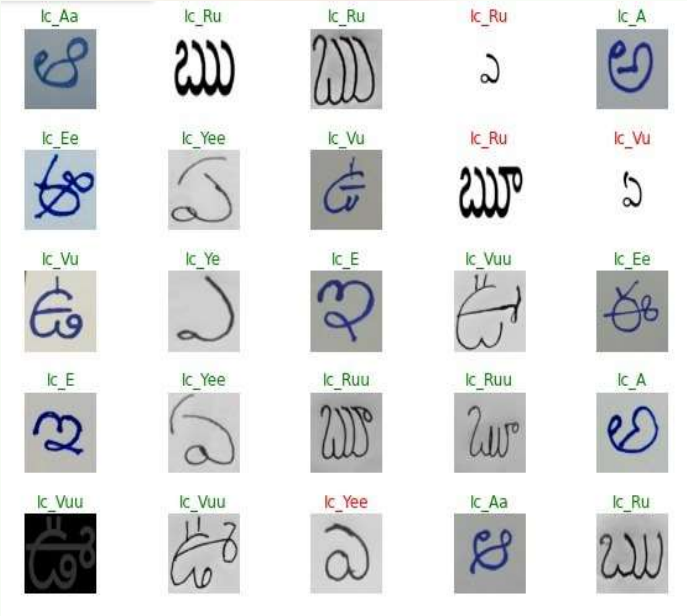
Tranformer Based OCR(TrOCR)

Text recognition is long-standing problem in research for document digitalization. This is existing approach for text recognition, are built based CNN for an image understanding and RNN for char-level text generation.

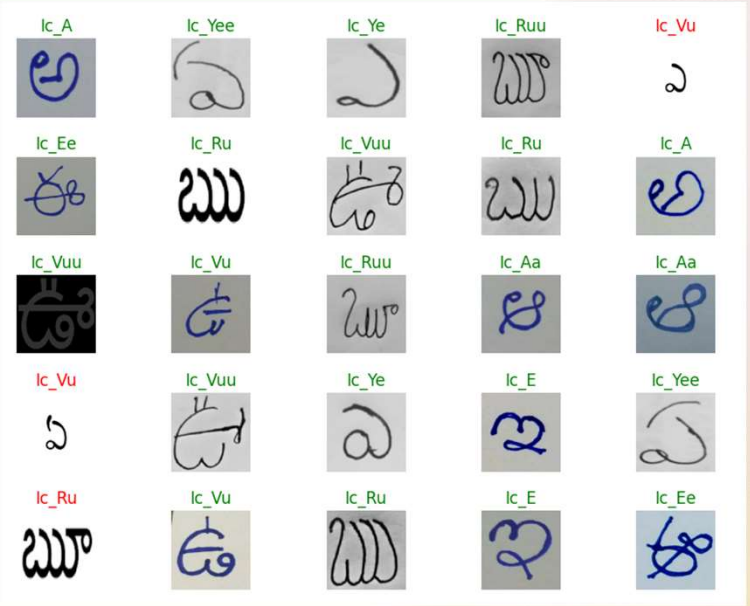
In addition, other language model is needed in improving the overall performance and accuracy as post-processing step. This is an end-to-end text recognition approach with pre-trained image Transformer and text Transformer models, namely TrOCR

Results & Discussion

Predictions on Resnet and Mobile V2



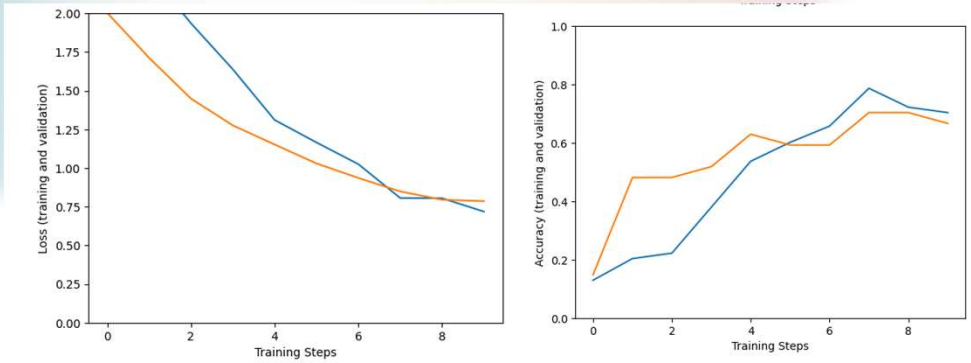
Test data on Mobilenet



Test data on Resnet 50

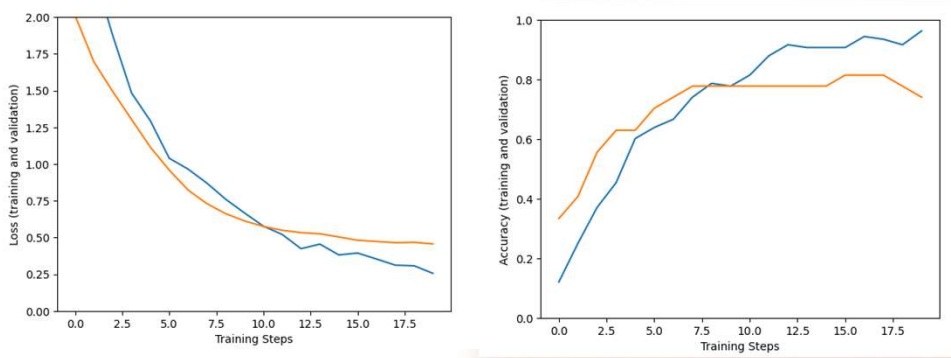
Results & Discussion

Metrics of Resnet, Mobilenet and Vision Transformer



Metrics of Mobilenet model

Mobilenet model has been inconsistent with respect to accuracy



Metrics of Resnet 50 Model

Resnet model has been consistent and giving appropriate results.

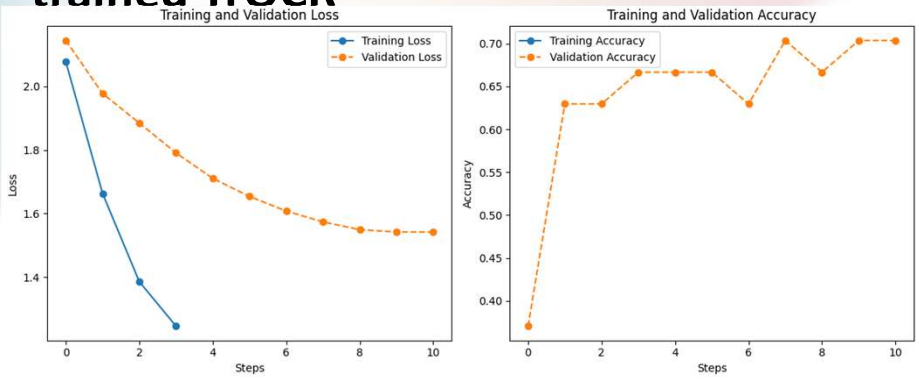
Results & Discussion

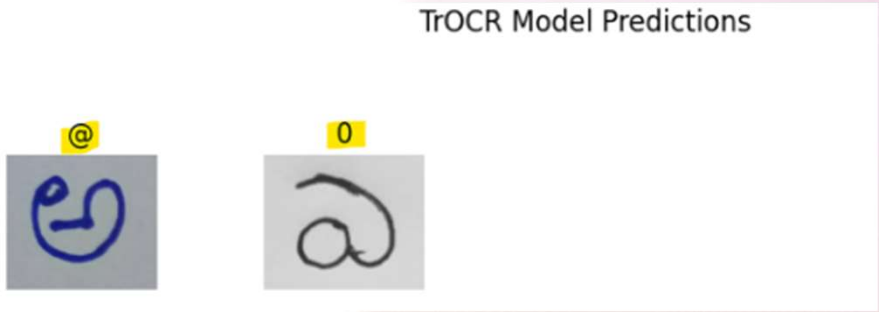
Summary of Resnet, Mobilenet and Vision Transformer

Mobilenet v2 has given 74% of accuracy which is allowed. Overall loss of Validation and accuracy drops down over the training steps, which signifies the the model is trained well. Where as the accuracy increased eventually and gave 74% of training accuracy after 5 epoches. While training the model 5 epoches are used as the data is less and using high epoch count will lead to overfitting the model. As part of the experiment, I have introduced epoch count 5, 8,10. But anything after 5 lead to overfitting. Now let's see how Resnet 50 works. Resnet has gradual increase in accuracy unlike Mobilenet v2.

Results & Discussion

Metrics of Vision Transformer and Predictions on pre trained TrOCR





Metrics of ViT model

The training of ViT model has been inconsistent and it didn't yield appropriate training accuracy. I have tried with hyper parameter tuning by changing the epochs count, updating the configuration, logging steps to 10, and by adding metrics callback method to fetch the training accuracy results

Predictions on TrOCR pretrained model

The model has predicted wrong values and didn't perform as expected. This can be avoided by training a new model with required dataset

Conclusion & Future Works

A Database of images of Telugu characters is extracted from a public dataset. This Database consists of 100 samples of 10 different Telugu characters, which is trained and tested using Convolutional Neural Network Classifier(CNN)- Resnet 50, Mobilenet v2, Transformer Based OCR and Vision Transformers. The performance accuracy of this model is achieved as **74% with Resnet 50**. By taking more samples of each character, the accuracy of our model can be improvised. This proposed work can be extended further to various other languages in India, with the scope of having a common OCR system for all the languages.

Though the advanced Transformer based are used for the research problem, they couldn't perform well due to below Drawbacks

Conclusion & Future Works

A. Data Requirements:

- High-Quality Datasets: TrOCR, ViT requires large, high-quality datasets of image-text pairs for effective training. Poor or limited data can lead to suboptimal performance. The dataset we considered is less in this research problem, could be one of the limitations of not getting the appropriate results.

A. Computational Resources:

- Training Complexity: Training TrOCR models is computationally intensive, requiring significant GPU/TPU resources. This can be a barrier for those with limited hardware capabilities.

A. Generalization:

- Adaptation to New Languages or Scripts: TrOCR models trained on specific languages or scripts may need retraining or fine-tuning to handle new languages or scripts effectively. In this research, I have used Telugu language, but ViT and TrOCR are trained on English Launguage. So this can be a cause of their underperformance.

In the future research, I would recommend to use further hyper parameter tuning techniques in the recent advanced techniques and I would recommend to use high volume datasets to overcome the limitations that we have got in this thesis.

Thank you!