OCR for handwritten text

Vinay Sutar (vinaysutar336@gmail.com)
Aditya Anand (21d070007@iitb.ac.in)

Applications and Motivation

- OCR projects help convert printed or handwritten text into digital format, making it easy to search, edit, and store. This technology is used to automate tasks like data entry, scanning invoices, and reading vehicle number plates.
- It also makes documents accessible to visually impaired people and helps preserve old books and records. Working on OCR is valuable because it improves skills in recognizing text from images and has practical uses in many fields like banks, healthcare, and law.
- It’s a project with real-world benefits and opens doors to new innovations in text recognition.

Implementation

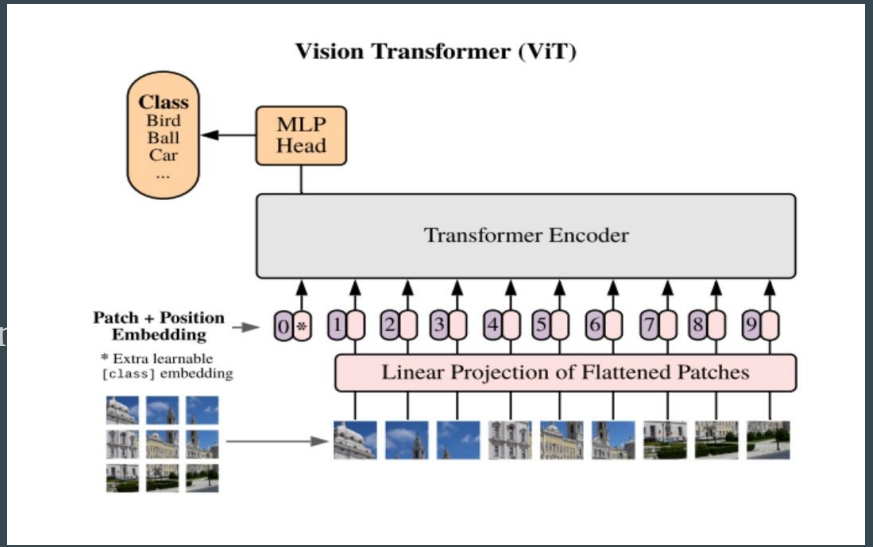
- Past works in this field utilised CNN's, Multidimensional LSTM's equipped with attention mechanisms. In this project we used a Transfor.er based approach for OCR.
- The dataset contains images of handwritten words that is used for training and testing the model. These images are performing various preprocessing steps are used as inputs to the Transformer. A RESNET50 backbone is used for this transformer model.
- The Encoder of the transformer receives these images after the 2-D image is flattened into 1-D sensors.
- These feature maps are then fed to the Multi-Layer Peceptron which gives out probability scores corresponding to each alphanumeric class.

Vision Transformer

This is the original architecture of a vision transformer by Google.

The words are flattened and fes to as inputs along with the positional encoding scheme, which helps keep track of the way the letters are arra

This feature gives Transformer a upper hand compared to CNN and LSTM, which has only a limited access to past memory.



Testing

- The probabilities are used to predict the corresponding and the positional encoding mechanisms utilised to arrange these letters to output the correct word.
- Our model has shown a test accuracy of 78 percent. Some illustrations of this is shown in the image attached
- Future work in this project can be utilizing text segmenting techniques to obtain text segments of a complete handwritten page and then feed the text segments to the model.
- Human writing vary a lot, higher precision can be obtainedby using more training data and more complex architectures.