SKILLCONNECT
Innovating Engineering!

Document Analysis

Problem Statement:

Extracting Handwritten and Printed Text from well formatted document images.

Description:

This problem has two phases. Initially, the team will be working on Extracting Information from printed documents which have a fixed format. For eg. Aadhar Cards and Pan Cards. After successfully implementing this project, the next phase would involve reading Handwritten Forms and digitising the data.

Requirements:

Develop an end to end application for both iOS and Android that can accurately do the above. Also, the size of the application must be as small as possible. Use of APIs is not allowed. Use of Automated ML tools is allowed, but only in accordance with the license.

Expected Outcome:

Given an Image of Document, understand the type of document, correct the orientation of the document and perform Key information Extraction. For eg. Extracting Name, Father's name, DOB etc from Aadhaar Cards.
Develop an Android and Web Interface for users to upload the Documents and get the results.

Final work would be being able to read Handwritten Forms and making a digital copy of it.

Tech-stack:

Deep Learning, Android, Web, iOS (preferably React), Flask server (Better if on device processing can be done)

Some Existing Work:

https://github.com/shubham99bisht/Aadhaar_Pan_Card_OCR