Report for analog devices and cnn ,nn

Detailed Report on Image-to-Speech Converter using CNN, OCR, and TTS

Our project is designed to extract text from images using convolutional neural networks based optical character recognition (OCR) system ans convert into speech using TTS engine.

Steps:

1. Convert to Grayscale: Removes color information to improve text detection.

2. Thresholding: Converts the image to pure black and white for better OCR accuracy.

3. Resizing: Ensures uniform size (128×32) for CNN input.

The entire system is integrated into a single script image\_to\_speech.py, which performs the following steps:

1. Takes an image as input.

2. Preprocesses the image.

3. Extracts text using CNN OCR (or Tesseract as fallback).

4. Processes the extracted text.

5. Converts text to speech and plays it through the speaker.

1. Components required :

1.ESP32 cam

2.Max98357A amp

3.speaker

4.FTDI module

1. Circuit connections
2. Connecting ESP32 cam to FTDI module
3. FTDI -ESP32 cam
4. VCC – 3.3 V
5. GND to GND
6. TX to RX
7. RX to Tx
8. GND – GPIO

* connecting to MAX98357A to ESP32 cam

1. Vin – 5v
2. GND – GND
3. LRC – GPIO 25
4. BCLK – GPIO 26
5. Din – GPIO 22
6. Connecting MAX98357A to speaker
7. Speaker+ to MAX98357A+
8. Speaker- to MAX98357A-

The working of components are as follows :

1. ESP32 cam captures the image and sends it to the model
2. The cnn model with OCR converts it into greyscale and extract text from the image .
3. The extracted text is further converted into speech by using a TTS engine by converting text into phenomes and further converting them to audio signals
4. The audio signals are then relayed to the MAX98357A module which converts the digital signal into an analog signal
5. The analog signal is further amplified and sent through the speaker for the output