Hybrid Image Processing with OCR and CNN for Entity Value Prediction

Overview

This solution integrates Optical Character Recognition (OCR) and a Convolutional Neural Network (CNN) model to predict entity values from images, particularly where traditional OCR methods may fail. The hybrid approach leverages OCR to extract text-based information and uses CNN for cases where OCR does not provide accurate predictions.

System Components

1. **OCR with Tesseract**:
    - The system applies OCR using the Tesseract engine to extract text from images.
    - Extracted text is cleaned and processed to identify numerical values associated with measurement units.
2. **CNN for Image Processing**:
    - A CNN model is used for cases where OCR fails to extract a valid entity value.
    - The CNN predicts continuous values based on preprocessed images.

Packages and Libraries:

1. **General Libraries**: os, random, pandas, numpy, matplotlib, tqdm, re, requests
2. **OCR**: pytesseract (OCR engine), Pillow (PIL) (Image handling), cv2 (OpenCV) (Image preprocessing)
3. **CNN**: tensorflow (Deep learning framework), keras (High-level API for CNN)
4. **Preprocessing and Utilities**: sklearn (Data splitting and scaling), BytesIO, itertools

Workflow

1. **Image Loading**:
    - Images are loaded from a dataset, resized to 224x224 dimensions, and preprocessed to ensure compatibility with the CNN input.
2. **OCR Processing**:
    - OCR is performed using Tesseract, extracting text from the image. This text is cleaned, and potential entity values are identified.
3. **Hybrid Prediction**:
    - If OCR provides valid results, those are returned as the predicted entity value.
    - If OCR fails, the image is passed to the CNN for prediction. The CNN outputs a numerical value with the corresponding measurement unit.
4. **Model Training**:
    - A CNN model is built with layers for feature extraction and dense layers for regression.
    - The model is trained using images from the dataset, where the target is the actual entity value.
    - Training is done over 10 epochs with a batch size of 32.

Diagram of Solution Architecture

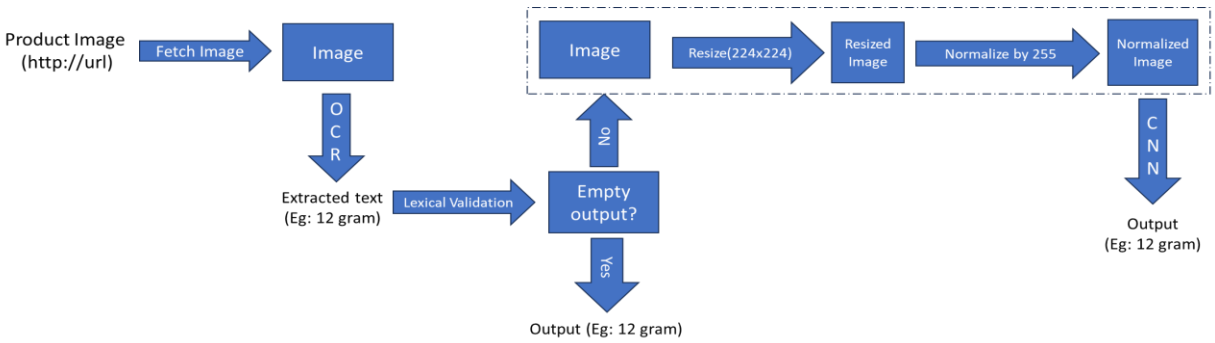

Figure 1: Structure of the Project and Dataflow during prediction

Table of Results

Original Image	OCR Detection	Preprocessed for CNN	Output
	Failed in detecting 	Number with units detected 	“item_volume”:”30 millilitre”
	No number with unit detected 	Numbers without unit detected 	“”
	Numbers with unit detected 	No need	“Width”: “20 centimetre”

Conclusion

This hybrid approach provides a robust solution for entity value prediction from images, combining the strengths of OCR and CNN to handle both text-based and visual data effectively. The system is trained on a dataset of images, and the model can predict continuous values with high accuracy, providing fallback mechanisms where OCR is insufficient.