AmazonML Challenge

Submitted by

Team Leader:

Jaimin Gulale

Dept. of Electronics and Communication Engineering

Team Members:

Het Patel

Dept. of Electronics and Communication Engineering

Bhavya Goradia

Dept. of Electronics and Communication Engineering

Siddharth Dalia

Dept. of Electronics and Communication Engineering

Institute of Technology,
Nirma University, Ahmedabad

Introduction

The goal of this project is to accurately extract text-based entity values from product images using advanced image processing techniques. The entities extracted include various dimensions (e.g., height, width), weight, voltage, and other product-specific metrics. This is achieved by processing images obtained via URLs, enhancing them with super-resolution techniques, and utilizing Optical Character Recognition (OCR) to extract text values. These values are then classified based on predefined entity-unit mappings.

Objectives

- Text Extraction: Extract entity-specific values from product images.
- Super-Resolution: Enhance image quality using ESRGAN for better OCR accuracy.
- OCR: Utilize EasyOCR for robust text detection and recognition.
- Entity Classification: Map extracted values to the correct entity using SI units and predefined entity-unit mappings.
- Automation: Automate the extraction process for large datasets of product images.

Technology used

- OpenCV: Image processing (preprocessing, reading images).
- EasyOCR: Optical Character Recognition (OCR) for text extraction.
- ESRGAN: Super-resolution model to enhance low-quality images before OCR.
- CRAFT Text Detector: For detecting text regions in the images.
- Pandas: Data handling and exporting results to CSV files.
- Python: The overall project is coded in Python, making use of libraries like OpenCV, EasyOCR, and TensorFlow.

System Requirements

- Python Libraries: OpenCV, EasyOCR, CRAFT Text Detector, Pandas, NumPy, TensorFlow, Pytesseract, etc.
- GPU Support: For faster execution of ESRGAN and OCR, it's recommended to run the project on a GPU-powered environment (e.g., Google Colab with GPU enabled).
- Dataset: CSV file containing image URLs and the corresponding entity name for each image.

Workflow

Step 1: Image Preprocessing

- Before performing OCR, the images are preprocessed to improve accuracy. This involves:
- Grayscale Conversion: Converting the image to grayscale for simplified processing.
- Binarization: Thresholding the image to differentiate text from the background.
- Noise Removal: Applying median blurring to reduce noise that may interfere with text recognition.
- Scaling: Resizing the image to enhance small text regions.

Step 2: Image Enhancement Using Super-Resolution

- Images that are too small, blurry, or of low resolution are enhanced using the ESRGAN (Enhanced Super-Resolution Generative Adversarial Network). ESRGAN improves the image resolution, making text easier to recognize.

Step 3: Text Detection and Extraction

Once the image has been preprocessed and enhanced, the next step is to detect and extract the text. This is achieved by using the CRAFT Text Detector to identify text regions in the image, followed by EasyOCR to recognize the text within those regions.

3.1. Text Detection Using CRAFT

The CRAFT (Character Region Awareness for Text detection) model is used to detect the regions of the image where text is located. This helps in identifying the areas where OCR should be applied.

3.2. Text Recognition Using EasyOCR

Once text regions are detected using CRAFT, EasyOCR is used to extract the text from those regions. EasyOCR is a simple and effective OCR tool that supports multiple languages and provides good accuracy for both printed and handwritten text.

SI Unit Classification

After extracting the text from the images, the next step is to classify the extracted text values into their corresponding SI units (e.g., volts, watts, kilograms). This step is crucial because the extracted values need to be mapped to the correct entity (e.g., voltage, weight).

Conclusion

In this project, we developed a robust and scalable system for feature extraction from product images using a combination of image preprocessing, enhancement, and advanced text detection and recognition techniques. The key steps included:

1. Image Preprocessing and Enhancement: We improved the quality of the images using techniques like grayscale conversion, resizing, and advanced upscaling using ESRGAN to handle low-quality or distorted images.
2. Text Detection: The CRAFT Text Detector efficiently identified regions containing text in the images, focusing the OCR process on relevant areas and improving overall accuracy.
3. Text Extraction: Using EasyOCR, we successfully extracted the text from the detected regions, handling a wide variety of image types and text formats.
4. Unit Classification: After extracting text, we used a combination of regular expressions and predefined SI units to classify and map extracted values to their corresponding entities (e.g., volts, weight, power).
5. Output Formatting: Finally, the results were saved in the required CSV format, ensuring accurate predictions of entity values and units for each image.

Through these steps, we addressed the challenge of extracting text and numeric data from images, handling various complications such as blurry, rotated, or distorted text. This system lays a solid foundation for applying machine learning models to classify and predict entity values with high precision.

Overall, this workflow demonstrates how combining image processing techniques and OCR tools can result in a powerful and adaptable solution for large-scale image-based data extraction tasks.

Code link :
https://colab.research.google.com/drive/1GUvq1u83P654hBD6HoRtkMv0LMxtoXh3

CODE LINK