Image Analysis, Segmentation and text Extraction

Anirudh Hegde

Bengaluru, India

---

1. Introduction

Text Extraction from image is concerned with extracting the relevant text data from a collection of images. Rapid development of digital technology has resulted in digitization of all categories of materials. Various methods are used for the extraction of text from colored journal images, camera captured images, video images, printed document, degraded document images, handwritten historical document, graphical and color document images, low resolution images, book cover and web pages. Image augmentation is used in deep learning and computer vision tasks to increase the quality of trained models. The purpose of image augmentation is to create new training samples from the existing data. Image segmentation is thus inevitable. Segmentation used for text-based images aim in retrieval of specific information from the entire image. This information can be a line or a word or even a character

2. Approach

- Utilized Google Cloud API for text extraction due to its robust OCR capabilities.
- Employed OpenCV for basic image segmentation to isolate visual elements.
- Structured the extracted content into a simple HTML format to differentiate text and visual elements.

3. Technologies Used

- Google-Cloud-Vision API: Integrates vision detection features within applications, including image labeling, face and landmark detection, optical character recognition (OCR), and tagging of explicit content.

- OpenCV-Python: For image segmentation.
- Pillow: For handling image files.

4. Implementation Details

- Set up Google Cloud Vision API for authentication and initialise the client
- Read the image file, performed text detection, and and segmented the image to identify and visualize individual elements.
- HTML content embedding these images along the extracted text.

5. Challenges Encountered

- Ensuring that the segment images which are converted to binary data which are stored in buffer as temporary storage are converted into text string using base64. So that Base64 image string is embedded directly into an HTML tag
- Ensuring that the extracted text and visual elements are correctly embedded in the HTML format.