Object Detection with GPT Vision Model

Shruthi Rengarajan *June 2024*

Project Summary

Objective

The objective of this project is to develop an object detection system using OpenAI's GPT-4 Vision model. The system is designed to analyze an image, detect objects within it, provide detailed descriptions of these objects along with their coordinates, and identify the most important object based on contextual information. This information will be used to draw bounding boxes around the detected objects and highlight the most important one.

API usage

OpenAI GPT-4 Turbo:

- Utilized for image analysis and object detection.
- Generates detailed descriptions of objects in the image and gives their coordinates. - Identifies the most important object based on the context of the image.

OpenAI API Key: - Secured via environment variables to ensure safe usage and prevent unauthorized access.

Challenges encountered - Coordinate Extraction: - Extracting accurate coordinates for each object from the GPT-4 response.

- Integration:
- Integrating the GPT-4 model with the SSD model for object detection.
- Ensuring the bounding boxes drawn by the SSD model align with the coordinates provided by GPT-4.

Methodology

-> Detection of objects in a Image

- Utilizing GPT-4 to analyse the image and generate a detailed description of objects present.

-> Getting details of the objects for Ranking

- Extracting object coordinates and contextual information from the GPT-4 output. - Parsing the description to identify the most important object

-> Ranking all the objects

- Ranking objects based on the importance identified by GPT-4. - Highlighting the most important object using visual cues.

Implementation

The implementation involves two main components: a JavaScript module (generic.js) and a Python script (draw_bounding_box.py).

generic.js:

- OpenAI GPT-4 Turbo model is used to analyze the image.
- We send a request to the GPT-4 API with the image URL.
- The solution entails a detailed description of the objects in the image, including their coordinates. - Then, we save the description and image URL to a JSON file (image_description.json).

draw_bounding_box.py:

- The saved JSON file is read to get the image description and URL.
- We then load the image from the URL and processes it using the SSD object detection model.
- The program draws bounding boxes around all detected objects.
- It highlights the most important object based on the coordinates extracted from the description.

Prompt Given:

Analyse the provided image and deliver a detailed description of all identifiable objects. For each object, provide the coordinates in the format (x, y) with the origin (0, 0) located at the top-left corner of the image. Additionally, suggest which object appears to be the most important based on the context within the image along with its coordinate.

Time Taken:

generic.js : 11.769194400072097 seconds

draw_bounding_box.py : 14.07 seconds

Output:

This is the input image given:

generic.js Output:
The image depicts a vibrant outdoor market setting, primarily focused on a single individual. Here are the identified objects and their approximate coordinates:

1. **Man in the Center with a Crate of Vegetables**
   - **Description**: A smiling man, dressed in a brown apron over a dark shirt, holding a wooden crate filled with various fresh vegetables. - **Coordinates**: (220, 180) for the center of the man.

2. **Wooden Crate of Vegetables**
   - **Description**: This is a wooden crate filled with a vibrant array of vegetables including bell peppers, tomatoes, and greens. - **Coordinates**: (215, 230) for the center of the crate.

3. **Market Stalls**
- **Description**: Several market stalls with canopies in shades of white and blue covering them.
- **Coordinates**: (100, 100) and (500, 100) for the approximate centers of these stalls visible behind the man.

4. **Additional Market Attendees**
- **Description**: Other individuals can be seen in the background, shopping and interacting at the market.
- **Coordinates**: (520, 250) for the center of an individual partially visible on the right side of the image.

5. **Vegetable and Fruit Displays**

- **Description**: Assorted fruit and vegetable produce can be seen on tables under the market stalls. - **Coordinates**: (530, 160) for the center of a vegetable display partially visible.

Based on the context within the image, the most important object appears to be the **Wooden Crate of Vegetables** that the man is holding. This object, combined with his engaging smile, forms the focal point of the image, symbolizing freshness and the appeal of local produce. Its central coordinate is approximately (215, 230). This object underlines the primary theme of the image – the vibrant and wholesome nature of the market and its products.

draw_bounding_box.py Output: