3.4 Vision Language Model: Scene Understanding

Note:

1) **This section requires the configuration of the API key in "1.3.2 Vision Language Model Accessing" before proceeding. Additionally, ensure that the images to be used in this section are imported.**

2) **This experiment requires either an Ethernet cable or Wi-Fi connection to ensure the main control device can access the network properly.**

3) **In this course, we will use a program to send an image to the large model for recognition and generate a description of the content within the image.**

1. Experiment Steps

1) Execute the following command to navigate to the directory of Large Model.

```
cd large_models/
```

2) Run the program:

```
python3 openai_vllm_understand.py
```

2. Function Realization

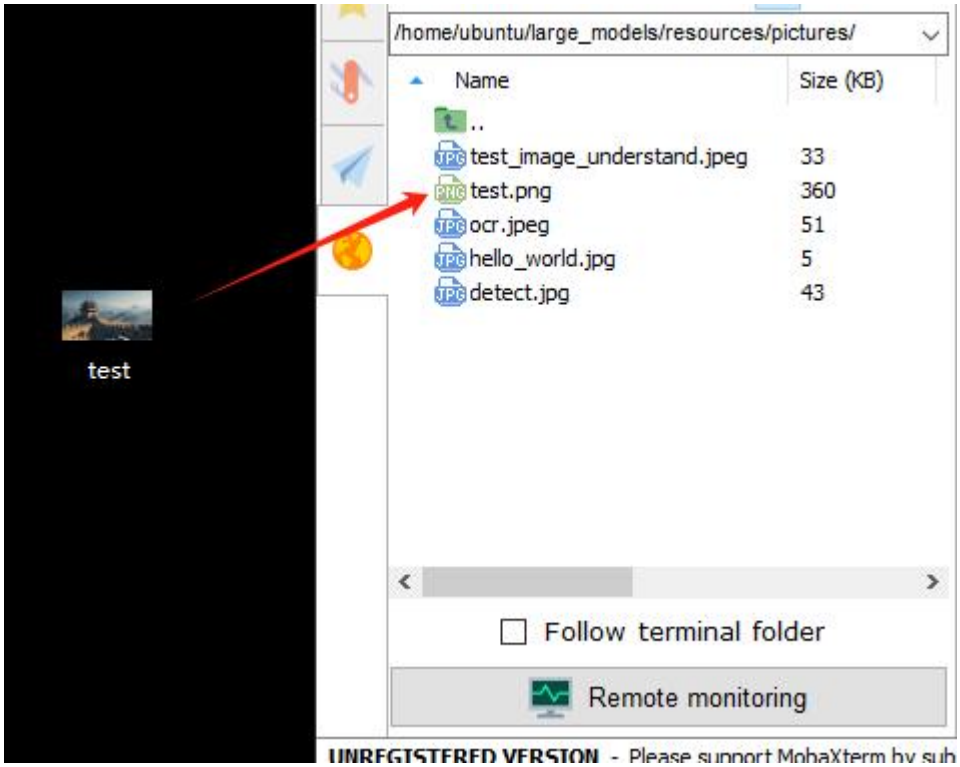
After running the program, the output printed matches our request of "Describe the image."

```
The image shows an assortment of fresh vegetables arranged in a collage. It includes:

1. Zucchini with its flower.
2. Tomatoes and eggplant.
3. A couple of tomatoes, one whole and one sliced.
4. An ear of corn partially husked.
5. A group of vegetables including a red bell pepper, cabbage, and garlic.
6. A few cucumbers.

The background is a transparent checkerboard pattern typical of image files with no background.
```

3. Function Expansion

If you need to recognize your own image, you should place the image in the corresponding path and modify the image path in the program.

1) First, drag your image directly into the ~/large_models/resources/pictures/ path using MobaXterm, and rename the image to test.png.



2) Then, open the scene understanding script by entering the following command in the terminal:

```
vim ~/large_models/vllm_understand.py
```

3) Change the image path in the code to reflect the name of your image (e.g., test.png).

```
client = speech.OpenAIAPI(api_key, base_url)

image = cv2.imread('./resources/pictures/test.png')
# 此处以qwen-vl-max-latest例，可按需更换模型名称。模型列表：https://help.aliyun.com/zh/model-studio/getting-started/models
print(client.vllm('图片里内容是什么?', image, prompt='', model='qwen-vl-max-latest'))
```

4) Run the program:

python3 ~/large_models/openai_vllm_understand.py

```
> python3 ~/large_models/openai_vllm_understand.py
```

```
www.hiwonder.com
The image showcases a vibrant assortment of fresh vegetables arranged in two rows. In the top row, there are leafy green plants alongside various colored tomatoes (yellow and orange), a dark purple eggplant, and green stems. In the bottom row, the display includes plump red tomatoes, an ear of corn, a green cabbage, and a mixture of radishes. Additionally, there is a red bell pepper, a cauliflower, and several green cucumbers. The overall composition highlights the diversity and freshness of these vegetables, emphasizing their bright colors and appealing textures.
```