3.5 Vision Language Model: Optical Character Recognition

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 use a program to transmit an image to the large model for recognition, extracting and identifying the text 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_ocr.py
```

2. Function Realization

After running the program, the output printed will be consistent with the content of the image sent.

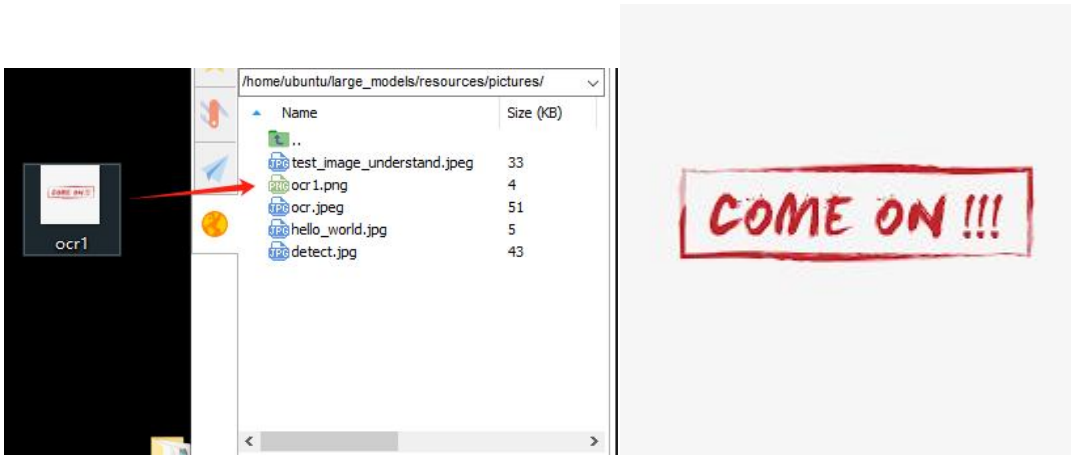
```
HELLO
WORLD.
```

3. Function Expansion

We can switch the image and change the large model to experience different functionalities of various models.

4. Change Pictures

1) Drag the image directly into the '~/large_models/resources/pictures/' path using MobaXterm. Here, we can drag in the image named 'ocr1.png' as an example, and let the program recognize the text 'COME ON'.



2) Then, input the command:

vim ~/large_models/openai_vllm_ocr.py

```
> vim ~/large_models/openai_vllm_ocr.py
```

3) Press the "i" key on your keyboard, which will display "INSERT" at the bottom.

i

```
INSERT large_models/openai_vllm_ocr.py
-- INSERT --
```

4) Change the image recognition path from: ./resources/pictures/ocr.jpeg

To: image = cv2.imread('./resources/pictures/ocr1.png')

image = cv2.imread('./resources/pictures/ocr1.png)

```
image = cv2.imread('./resources/pictures/ocr1.png')
```

5) Run the program:

python3 ~/large_models/openai_vllm_ocr.py

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

```
www.hiwonder.com
COME ON !!!
```