

Qwen2.5VL

Qwen2.5VL

- 1. Model Size
- 2. Performance
- 3. Using Qwen2.5VL
 - 3.1 Running Qwen2.5VL
 - 3.2 Conducting a Dialogue
 - 3.3 Visual Functions
 - 3.4 Ending the Dialogue
 - 3.5 Chinese Dialogue
- References

Demo Environment

Development Boards: Jetson Orin, RDK X5, Raspberry Pi 5 series motherboard

Note: Jetson Orin Nano 4GB, Raspberry Pi 5 2GB, and RDK X5 4GB cannot run this model due to performance limitations. The Raspberry Pi 5 4GB version will run very slowly.**

Qwen2.5-VL is Qwen's new flagship visual language model and a significant leap forward compared to the previous Qwen2-VL.

1. Model Size

Model	Size
qwen2.5vl:3b	3.2GB
qwen2.5vl:7b	6.0GB

2. Performance

		Qwen2.5-VL-3B	InternVL2.5-4B	Qwen2-VL-7B
College-level Problems	MMMU	53.1	52.3	54.1
	MMMU Pro	31.6	32.7	30.5
Document and Diagrams Reading	DocVQA	93.9	91.6	94.5
	InfoVQA	77.1	72.1	76.5
	CC-OCR	74.5	-	61.6
	OCRBenchV2	54.3	-	47.8
General Visual Question Answering	MegaBench	28.9	-	34.4
	MMStar	55.8	58.3	60.7
	MMBench1.1	77.6	79.3	80.7
Math	MathVista	62.3	60.5	58.2
	MathVision	21.2	20.9	16.3
Video Understanding	VideoMME	61.5	62.3	63.3
	MMBench-Video	1.6	1.7	1.4
	LVBench	43.3	-	-
	CharadesSTA	38.8	-	-
Visual Agent	AITZ	76.9	-	-
	Android Control	63.7	-	-
	ScreenSpot	55.5	-	55.3
	ScreenSpot Pro	23.9	-	1.6

		Qwen2.5-VL 7B	Qwen2-VL 7B	GPT-4o Mini	Other Best <small>Open LLM With Similar Size</small>
College-level Problems	MMMU	58.6	54.1	60.0	56.0
	MMMU Pro	38.3	30.5	37.6	34.3
Document and Diagrams Reading	DocVQA	95.7	94.5	-	93.0
	InfoVQA	82.6	76.5	-	77.6
	CC-OCR	77.8	61.6	-	61.6
	OCRBenchV2	56.3	47.8	43	47.8
General Visual Question Answering	MegaBench	36.8	34.4	43.1	34.4
	MMStar	63.9	60.7	54.8	62.8
	MMBench1.1	82.7	80.7	76.0	79.4
Math	MathVista	68.2	58.2	52.5	67.2
	MathVision	25.1	16.3	-	19.7
Video Understanding	VideoMME	65.1	63.3	64.8	63.3
	MMBench-Video	1.8	1.4	-	1.7
	LVBench	45.3	-	-	38.4
	CharadesSTA	43.6	-	-	48.4
Visual Agent	AITZ	81.9	-	-	53.3
	Android Control	60.1	-	-	61.5
	ScreenSpot	84.7	55.3	-	89.5
	ScreenSpot Pro	29.0	1.6	-	35.7

3. Using Qwen2.5VL

3.1 Running Qwen2.5VL

Use the `run` command to start running the model. If this model has not been downloaded before, it will automatically pull the model from the Ollam model library:

```
ollama run qwen2.5vl:3b
```

```

:~$ ollama run qwen2.5vl:3b
pulling manifest
pulling e9758e589d44: 100% 3.2 GB
pulling a242d8dfdc8f: 100% 487 B
pulling 75357d685f23: 100% 28 B
pulling 832dd9e00a68: 100% 11 KB
pulling 52d2a7aa3a38: 100% 23 B
pulling 97a23b280c2e: 100% 567 B
verifying sha256 digest
writing manifest
success
>>> Send a message (/? for help)

```

3.2 Conducting a Dialogue

```
Please tell me how many hours there are in a day
```

The response time depends on hardware configuration; please be patient!

```

:~$ ollama run qwen2.5vl:3b
>>> Please tell me how many hours there are in a day
There are 24 hours in a day.
>>> Send a message (/? for help)

```

3.3 Visual Functions



what do you see in this picture? ../test_pic.png
#Using ":" + image path" in the dialogue allows the model to use its visual functions to parse information from the image.

```
~$ ollama run qwen2.5vl:3b
>>> What do you see in this picture?../test_pic.png
Added image '../test_pic.png'
The picture shows a person sitting on a rooftop or elevated platform, looking up towards the sky. The person is wearing a white tank top and appears to be in a relaxed or contemplative pose. In the foreground, there is a clear plastic bottle. The background features a bright, sunny sky with some clouds and a cityscape with buildings visible in the distance. The overall atmosphere of the image is calm and serene.
```

3.4 Ending the Dialogue

Use the shortcut `Ctrl+d` or `/bye` to end the dialogue!

3.5 Chinese Dialogue

For those without a Chinese input method, please refer to the Chinese input method switching tutorial.

Chinese Dialogue:

```
>>> 一个小时是多少分钟？请用中文回答我
一个小时是60分钟。
>>> Send a message (/? for help)
```

References

Ollama

Official Website: <https://ollama.com/>

GitHub: <https://github.com/ollama/ollama>

Qwen2.5VL

GitHub: <https://github.com/QwenLM/Qwen2.5-VL>

Ollama Corresponding Model: <https://ollama.com/library/qwen2.5vl>