Vision to Text tasks (input=vision, output=text)

Support examples

A cat wearing sunglasses.

Elephants walking in the savanna.

Query

<BOS><image>Output: A cat wearing sunglasses.<EOC><image>Output: Elephants walking in the savanna.<EOC><image>Output:

Processed prompt

Visual Question Answering Task (input=vision+text, output=text)

Support examples

What's the cat wearing? sunglasses

How many animals? 3

Query

What is on the water?

<BOS><image>Question: What's the cat wearing? Answer: sunglasses<EOC><image>Question: How many animals? Answer: 3<EOC><image>Question: What is on the water? Answer:

Processed prompt