



Object Category

horse

Bbox Coordinates

[223.07, 267.79, 101.47, 59.39]

Image Caption

The image depicts a group of people on horseback in a forested area. The person riding the horse in the red box is wearing a blue shirt and dark pants, and appears to be leading the group. The horse is brown and is walking on a dirt path. The riders are wearing backpacks, suggesting they might be on a trail ride or a guided tour. The forest is lush with green trees, and the sky is visible through the canopy, indicating it's a sunny day. The overall atmosphere of the image is one of outdoor adventure and exploration.



Can you specify if the raw expression that **{brown horse walking on sand}** is referring to the target object? If not, please modify it to a correct expression.



### Planning

**Reasoning:** The expression describes a brown horse that is walking on sand, but the image caption mentions the horse is walking on a dirt path, and there is no mention of sand in the caption. Based on this information, I reason the state is 4 where I am uncertain of object relation with surroundings. So I plan for Action 4 where I can crop a local object region in this image for VLM to further re-perceive. The description of VLM supplements the current information for me to conduct a more thorough expression examination.

**Action:** [{"Tool\_name": "VLM", "Tool\_parameters": {"Prompt": "Questions: 1. Is the horse walking on sand or on a dirt path?", "Image editing": "extended object crop."}}]

**Response:** I will consult with a VLM to acquire supplement information that will enable me to evaluate the current expression more thoroughly.



### Tool use

**VLM output:** 1. The horse is walking on a dirt path, not on sand.



### Reflection

**Feedback:** The expression **{brown horse walking on sand}** describes a wrong relation (on sand) of the horse.



### Planning

**Reasoning:** Based on the feedback, I reason the state is 2 where I am certain the expression **{brown horse walking on sand}** is wrong. So I plan for Action 2 where I use an LLM to rewrite the expression, and then check the correctness of the new expression.

**Actions:** [{"Tool\_name": "LLM", "Tool\_parameters": {"Prompt": "Image Caption, ..., Response, ..."}}]

**Response:** The expression **{brown horse walking on sand}** contains flaws, now I will first leverage the LLM to create an alternative expression for you.



### Tool use

**LLM output:** New Description: **{brown horse walking on dirt path with rider in blue shirt and hat, leading the way}**



### Reflection

**Feedback:** The expression **{brown horse walking on dirt path with rider in blue shirt and hat, leading the way}** describes the object category 'horse' and provides similar context to the caption.



### Planning

**Reasoning:** Based on the feedback, I reason the state is 1 where I am certain the new expression **{brown horse walking on dirt path with rider in blue shirt and hat, leading the way}** is correct. So I plan for Action 1 to stop.

**Actions:** []

**Response:** The expression **{brown horse walking on sand}** is a wrong one, and a correct one could be **{brown horse walking on dirt path with rider in blue shirt and hat, leading the way}**.

The expression **{brown horse walking on sand}** is a wrong one, and a correct expression could be **{brown horse walking on dirt path with rider in blue shirt and hat, leading the way}**.



Real-LOD