

Do Video MLLMs use Motion in Visual Grounding?

Location



Please segment, except for the two bears in front, all other bears **in the distance** walking.



Sure, it is [SEG]

Heading



Please segment, the plane going left.



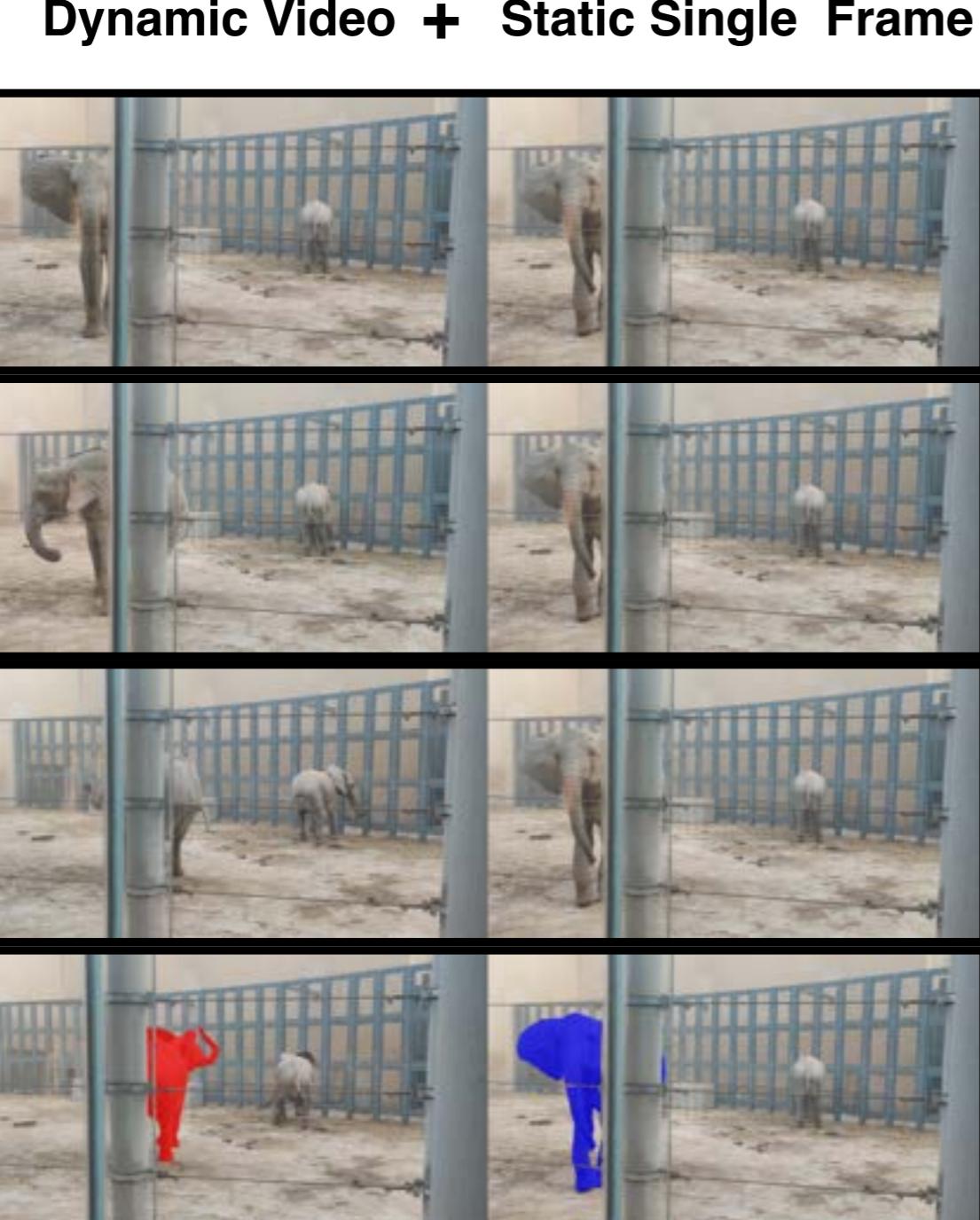
Sure, it is [SEG]

Motion Understanding is not Required

Motion Centric Benchmark (MoCentric-Bench)

1

Motion Existence

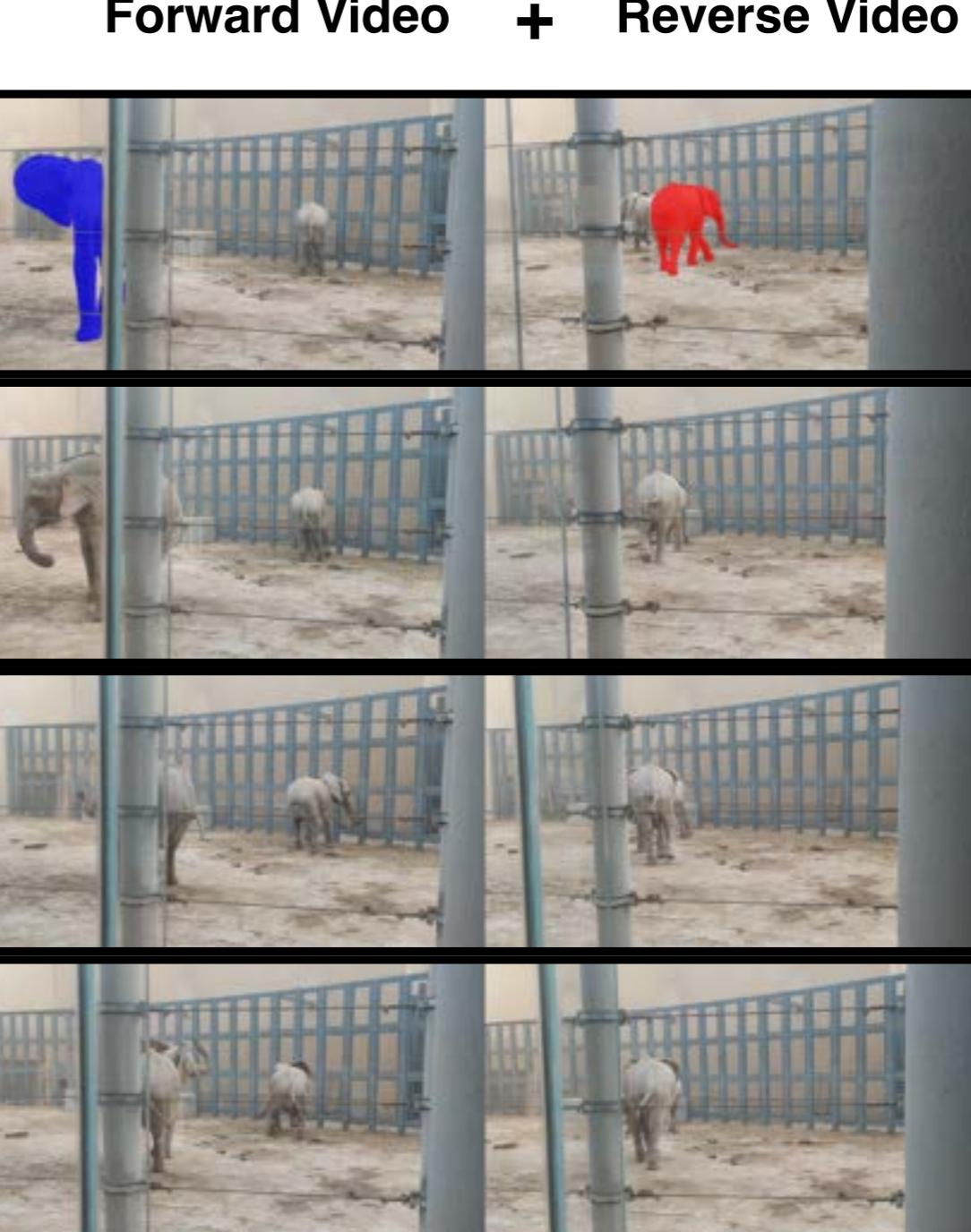


✓

✗

2

Motion Ordering



✗

✓



Please segment, front elephant walking to backwards.

Please segment, backwards elephant walking to front.

Motion Understanding is Required
Multi-Video Layout for Visual Grounding