

Step 1: Image-Instruction Pair Collection

Step 2: MLLM Response Collection

Step 3: MLLM Judge v.s. Human
Annotation