

## Traditional T2I Benchmarks

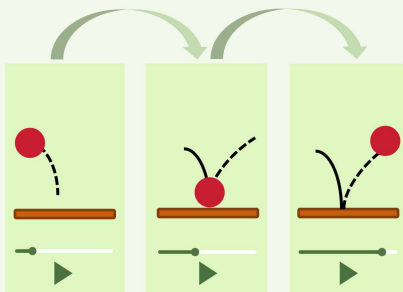
Single Image



General-domain Visual Fidelity

## Video Generation Benchmarks

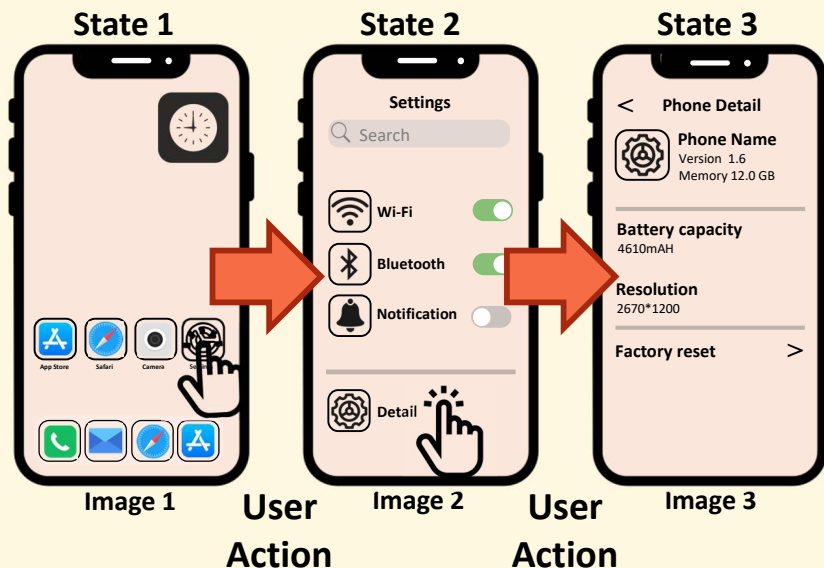
Video Frames



Continuous State Transitions

## GUI Environment Benchmark (Ours)

Multi Images



Discrete State Transitions

## Main Challenges

Semantic Interpretation of Icons

Refining text rendering

Improving grounding point localization

