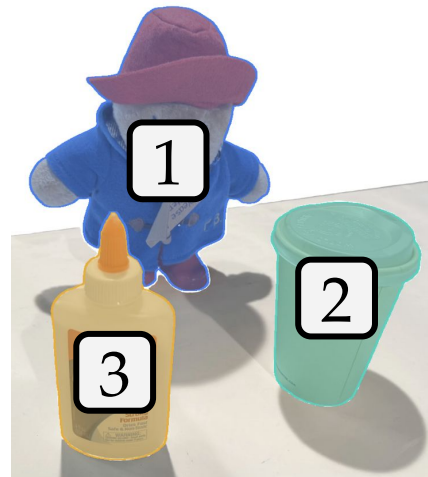


Original Scene *I*



Annotated Image I'

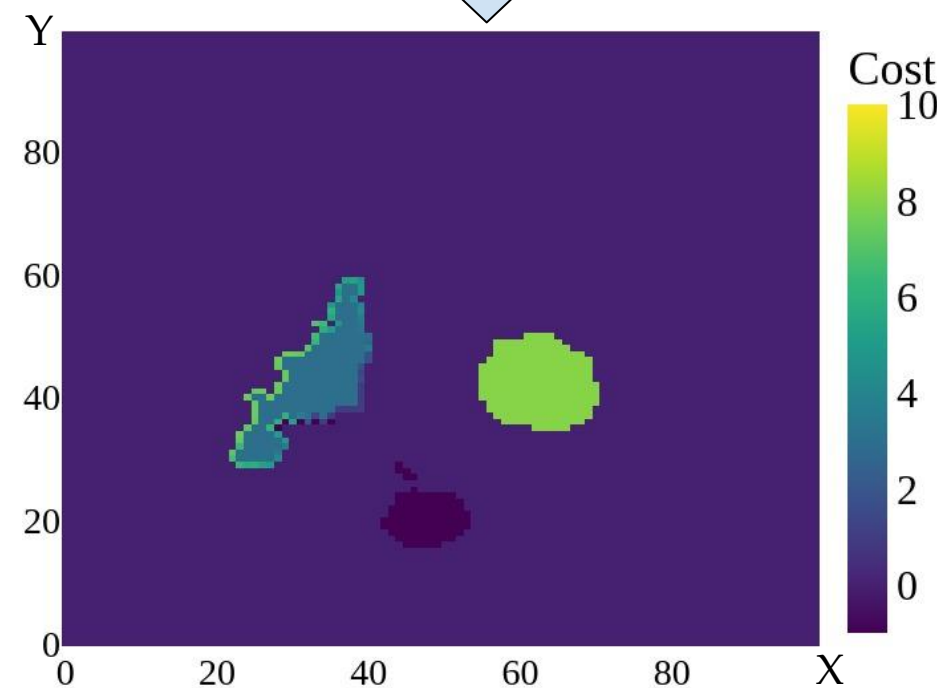
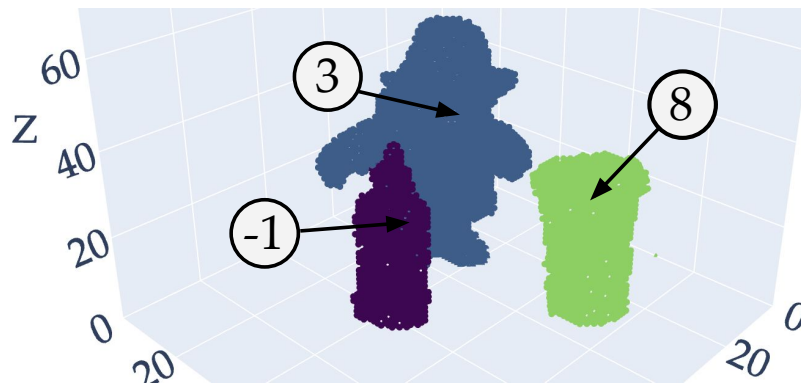


There is {(1) a toy bear, (2) a coffee cup and (3) a glue}.
Please assign safety scores.

Language Instruction ℓ

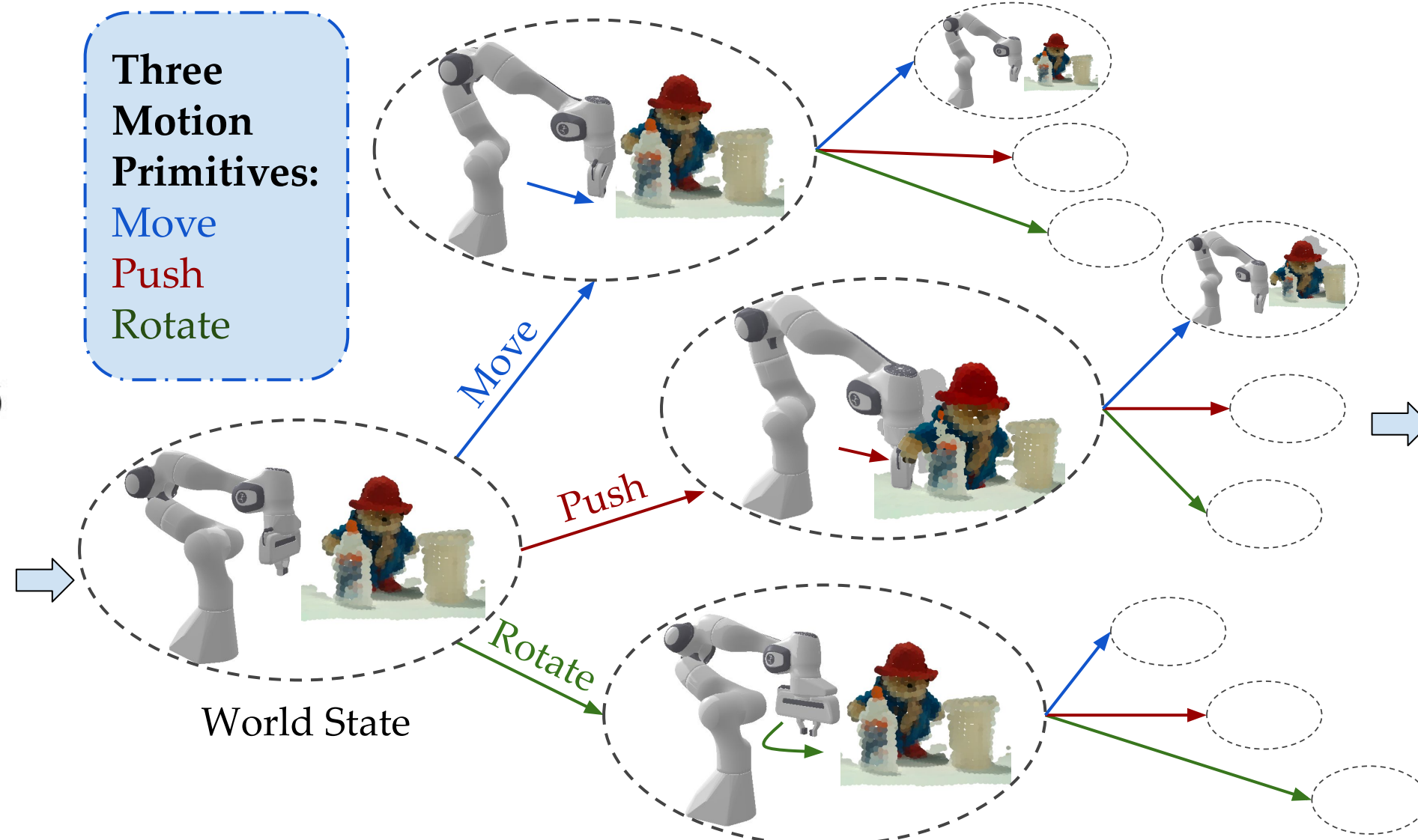
Vision Language Model

Directional Cost Map



Contact-Aware Motion Planning

Three Motion Primitives: Move Push Rotate



Execution

