

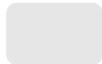
Autoregressive Transformer



Stage 1: Multiview World Model Post-training



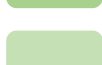
Stage 2: Action Fine-tuning



Language



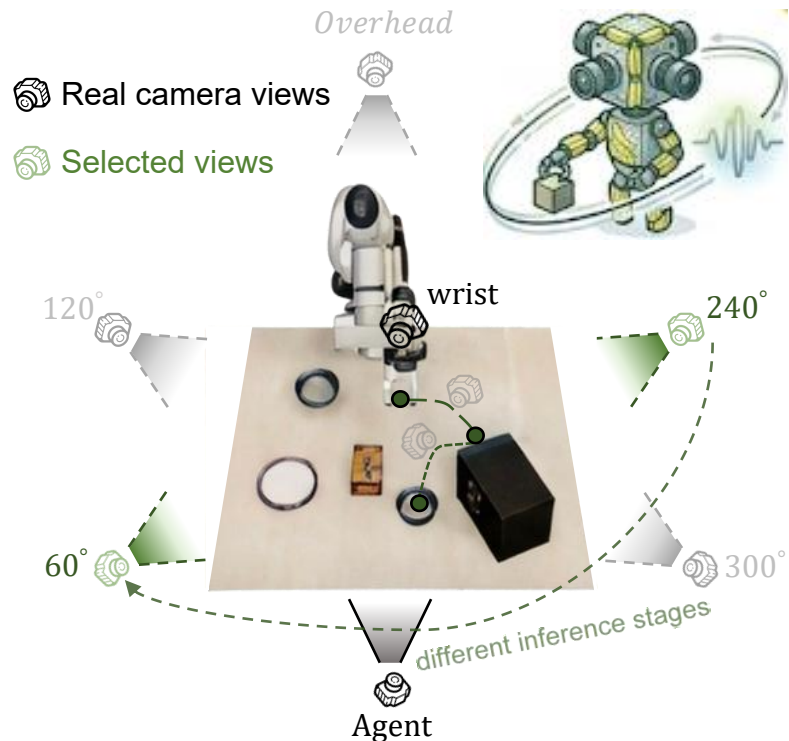
Dense aux-view



Compact aux-view



Action



Stage 3: Dynamic Multiview Selection Inference