Visual Input → **Mode Selector** → **Discrete Interaction Mode** → **Action Predictor** → **Trajectory Inference**