

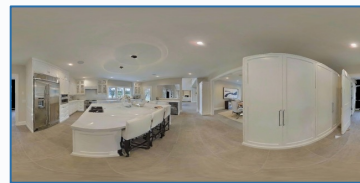
Panoramic inputs

Uniform history sampling



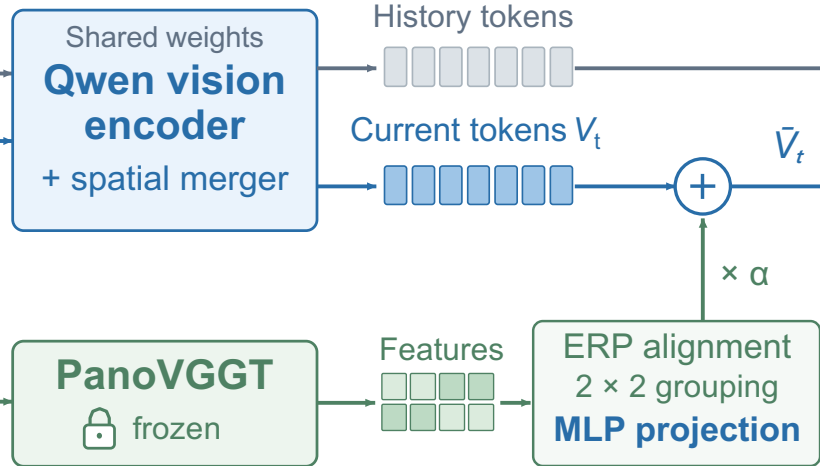
≤ 10 frames / last 100

Current RGB ERP P_t

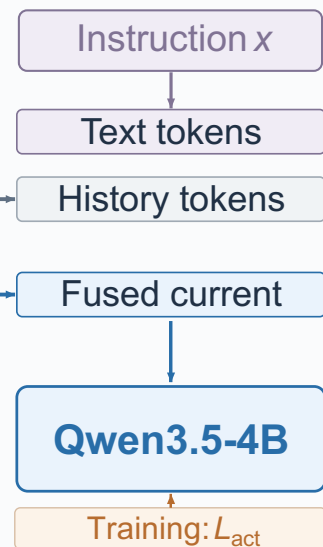


RGB-only observations

Semantic-geometric fusion



Action prediction



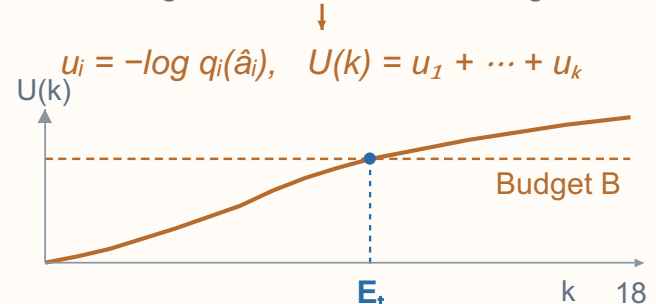
Confidence-guided execution

Predict $H = 18$ actions

F F R R R R R R F F F F F F F F S

Autoregressive action words + logits

$u_i = -\log q_i(\hat{a}_i), U(k) = u_1 + \dots + u_k$



Execute prefix

Fixed H, adaptive E_t

Reobserve