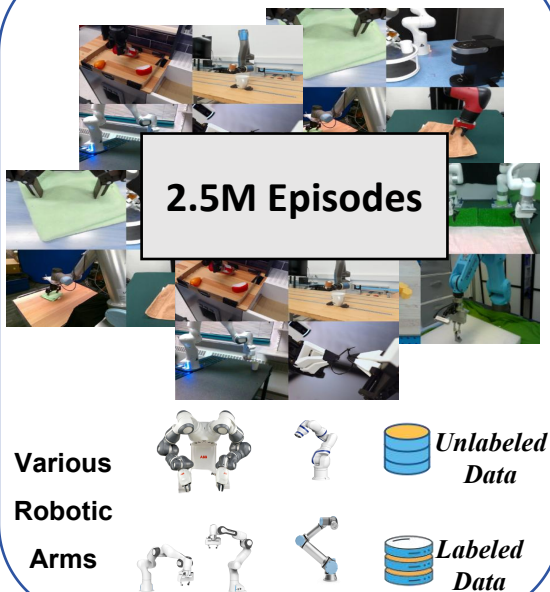
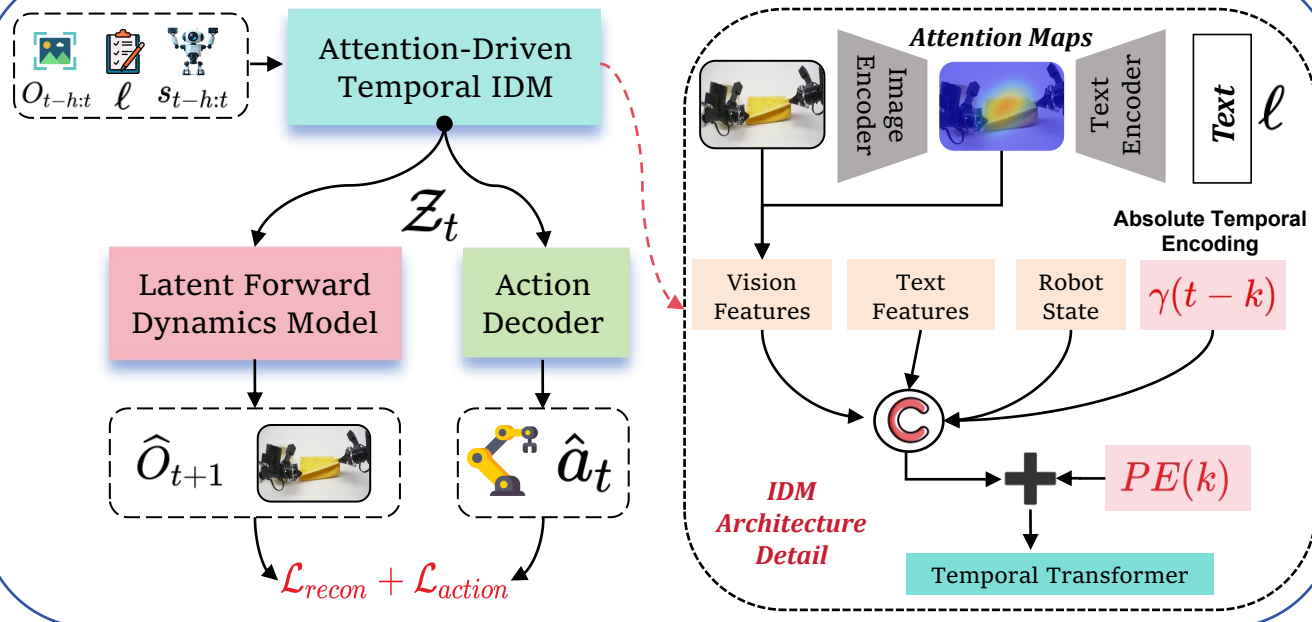


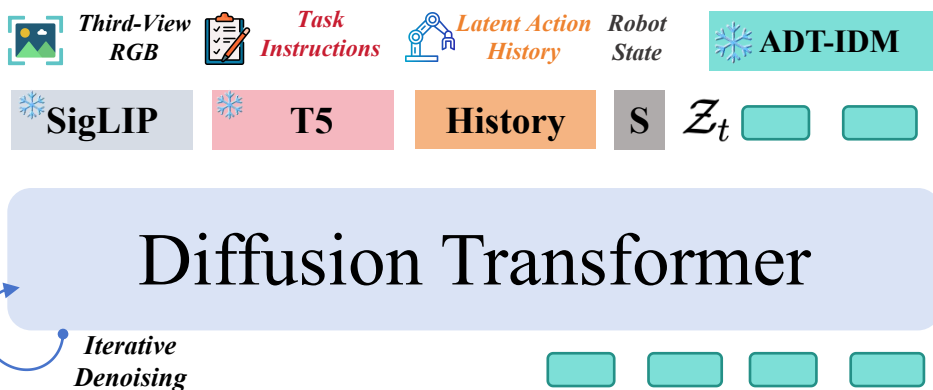
Pre-training Data



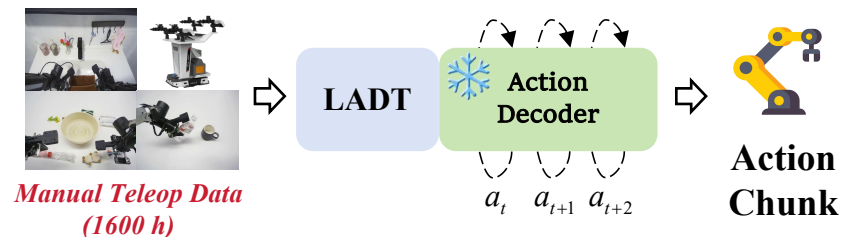
Stage 1: Temporal-Attentive Latent Action Model (TA-LAM)



Stage 2: Latent Action Diffusion Transformer (LADT)



Stage 3: Fine-Tuning



Zero-Shot Generalizability



Few-Shot Learning

