

Context Frames + Text Conditioning

Predicted

t_{-3}

t_{-2}

t_{-1}

t_0

t_{+1}



"action: dx=+1.338, dy=-0.659, dyaw=-0.707, rel_t=0.055"

