

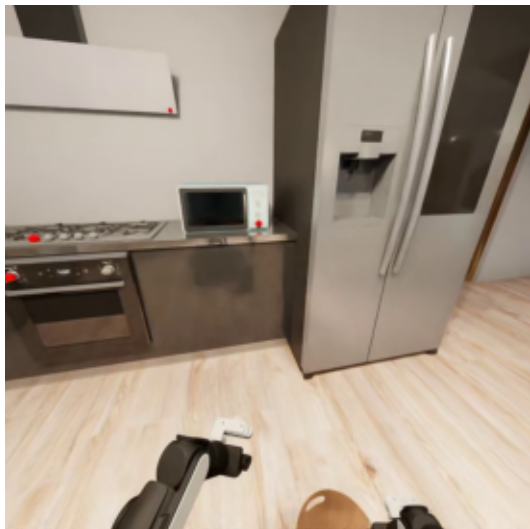
BEHAVIOR focus-mask via GPT selection

shipped seg_instance_id video -> clean instance candidates ->
GPT picks the subgoal target (no id->color map, no replay)

6 subgoal examples

task-0001 episode_00010690 skill 1: pick up from | target = 'trash can' (trash_can_116) | frame 656

RGB frame



16 instance candidates (lettered)



GPT pick 'O' = focus mask

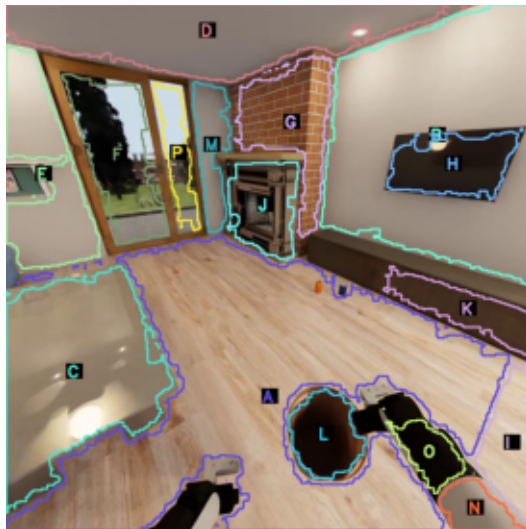


task-0001 episode_00010690 skill 3: pick up from | target = 'can of soda' (can_of_soda_114) | frame 1981

RGB frame



16 instance candidates (lettered)



GPT pick 'N' = focus mask



task-0001 episode_00010690 skill 6: pick up from | target = 'can of soda' (can_of_soda_115) | frame 4335

RGB frame



16 instance candidates (lettered)



GPT pick 'H' = focus mask

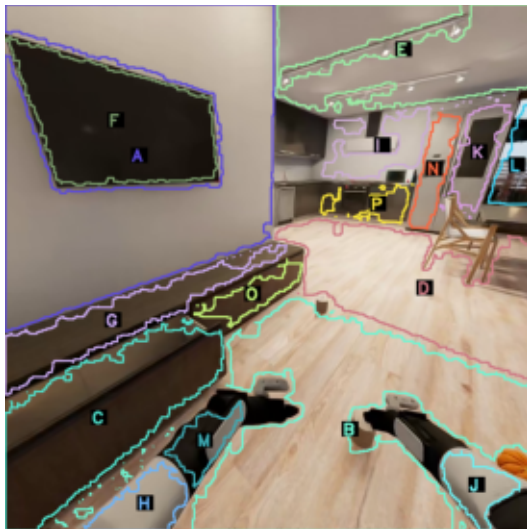


task-0002 episode_00020010 skill 3: pick up from | target = 'pillar candle' (pillar_candle_89) | frame 2040

RGB frame



16 instance candidates (lettered)

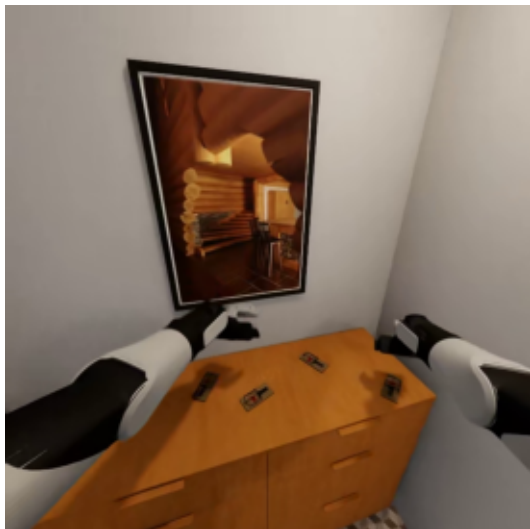


GPT pick 'B' = focus mask

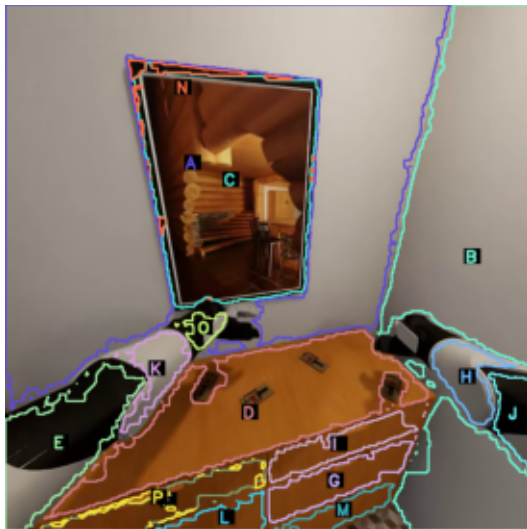


task-0005 episode_00050020 skill 1: pick up from | target = 'mousetrap' (mousetrap_47) | frame 888

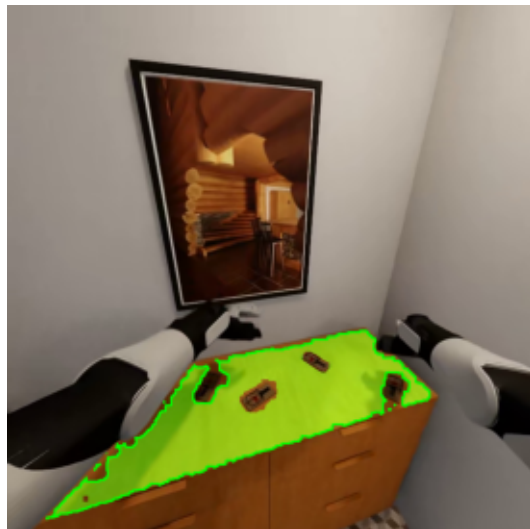
RGB frame



16 instance candidates (lettered)



GPT pick 'D' = focus mask

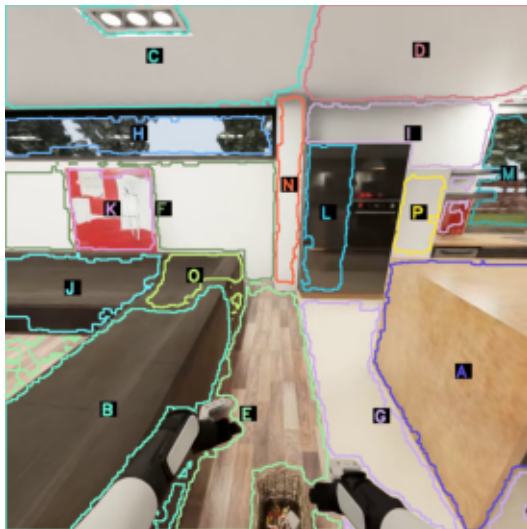


task-0009 episode_00090010 skill 1: pick up from | target = 'wicker basket' (wicker_basket_218) | frame 1243

RGB frame



16 instance candidates (lettered)



GPT pick 'E' = focus mask

