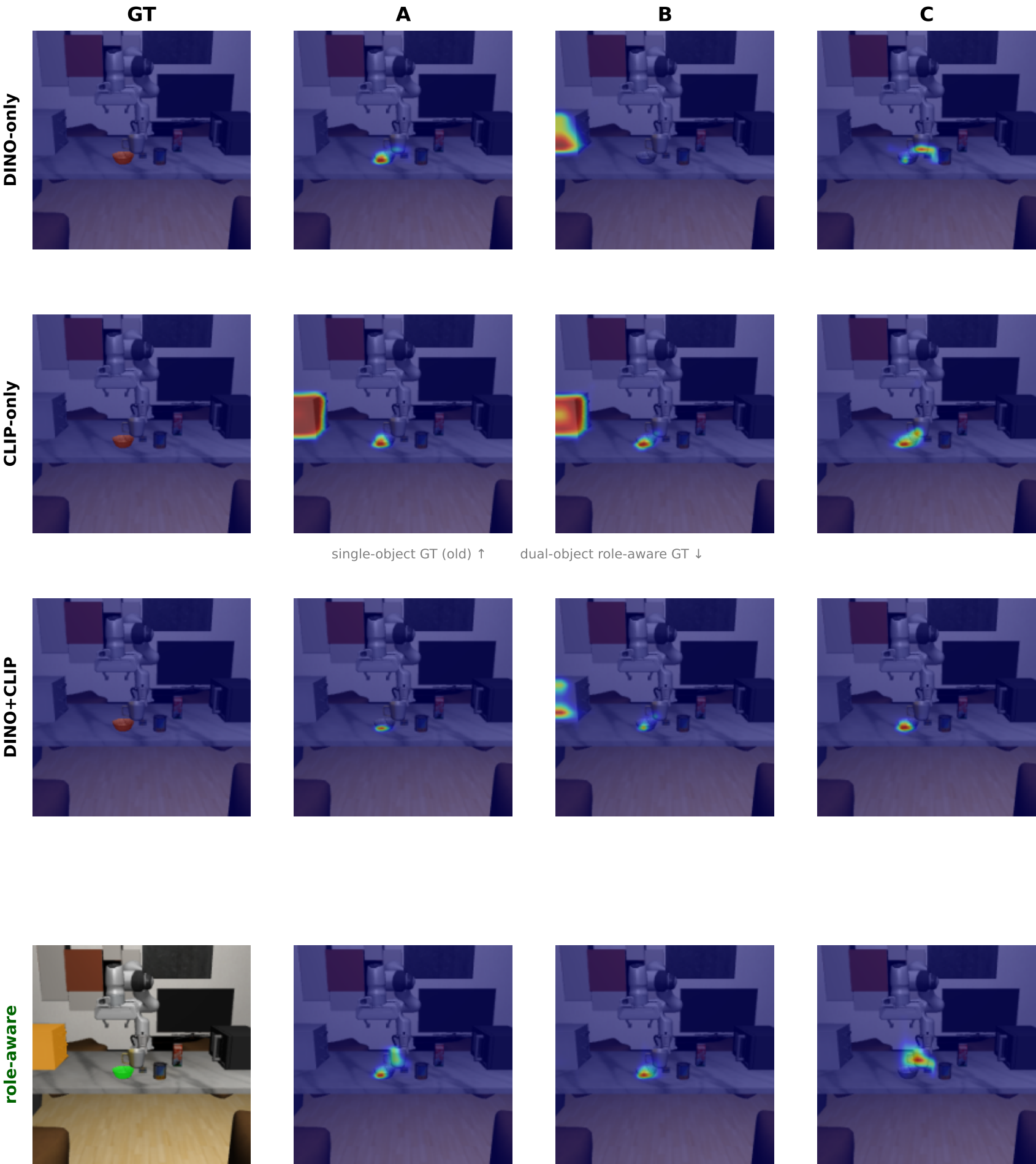


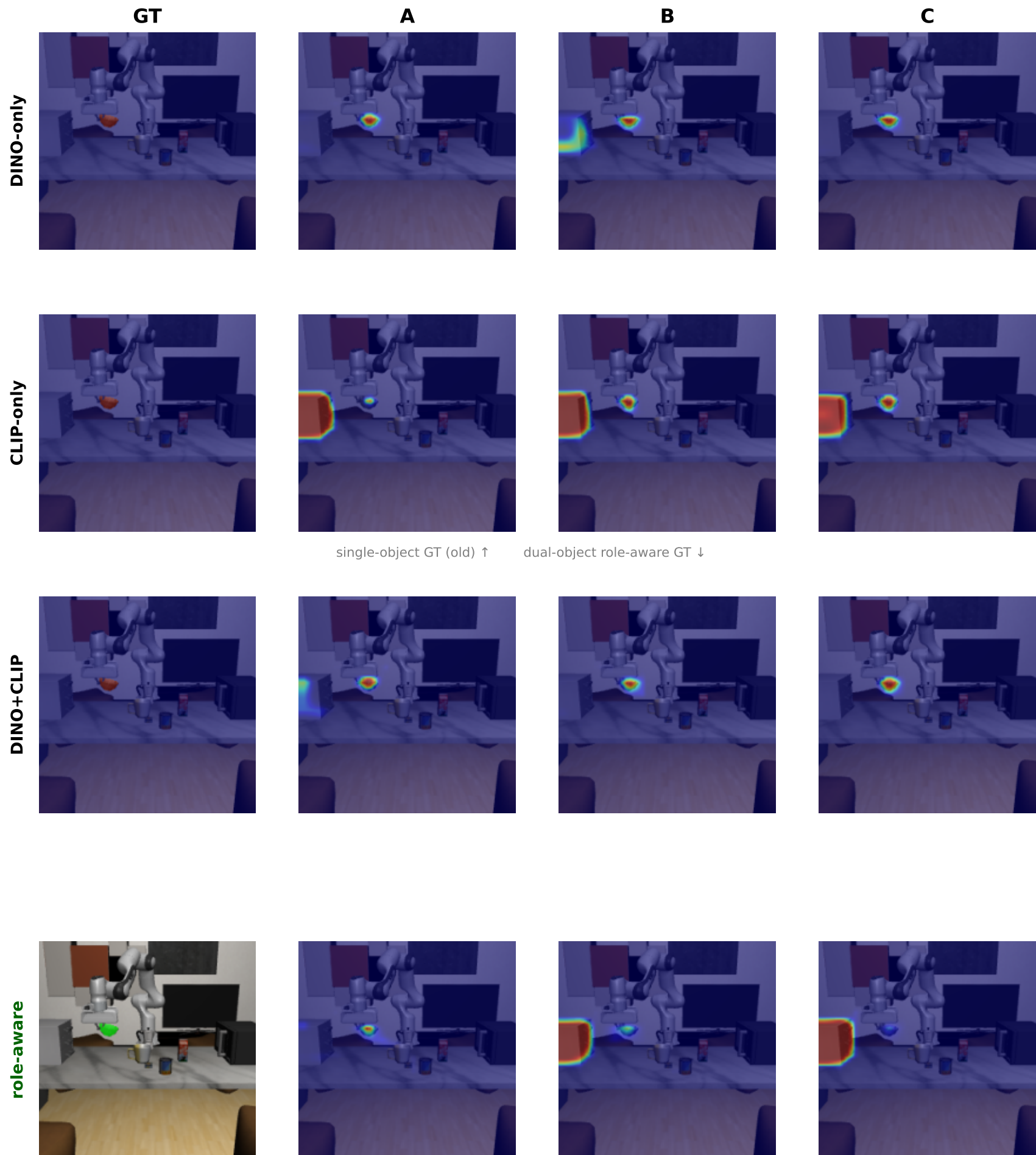
Mix_case5 — step: Pick up the akita black bowl from coffee table. [grasp]

goal: Heat the alphabet soup in a bowl using the microwave□pour a cup of milk into a mug, reposition t



Mix_case5 — step: Place the akita black bowl on table surface. [release]

goal: Heat the alphabet soup in a bowl using the microwave□pour a cup of milk into a mug, reposition t



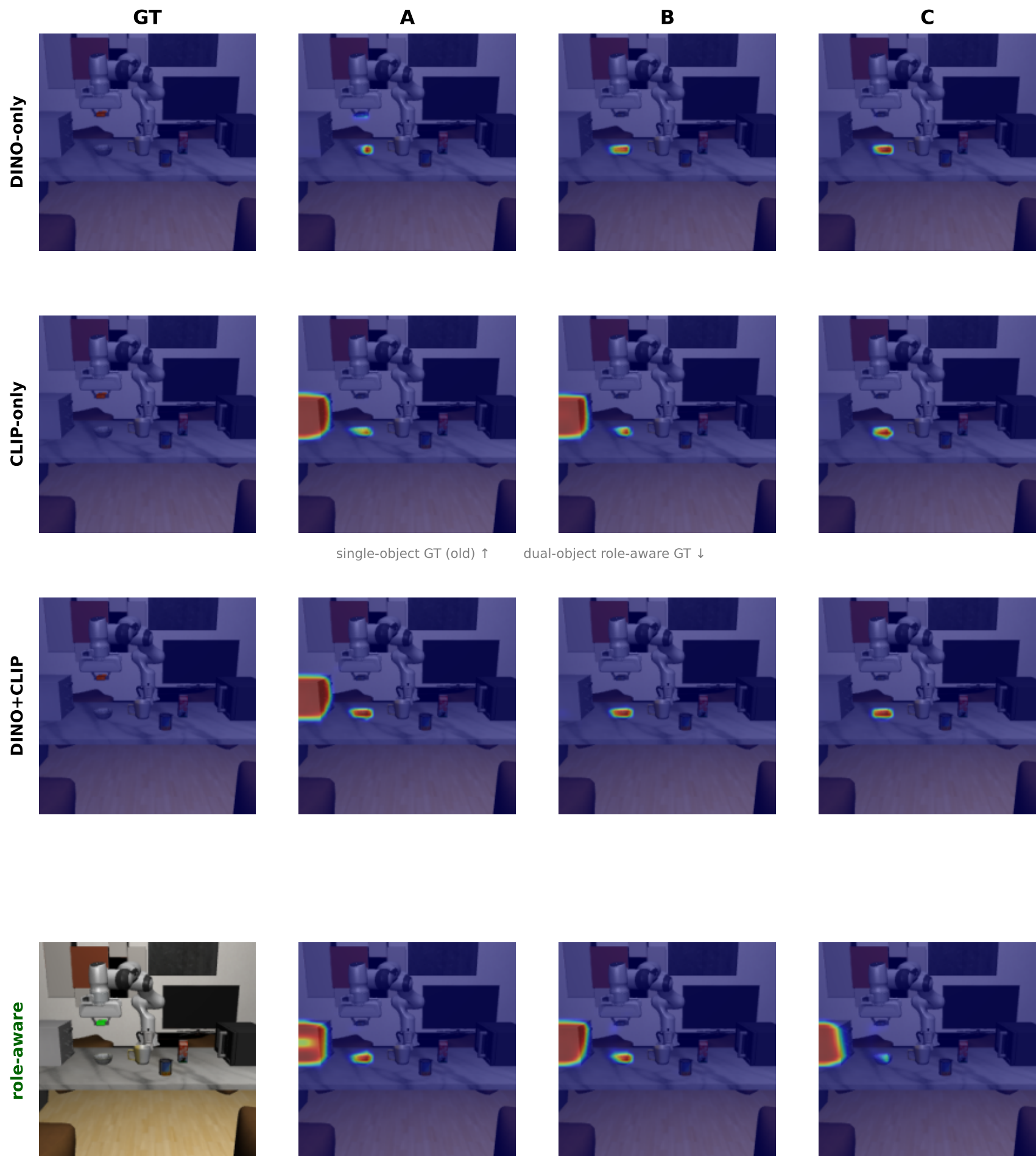
Mix_case5 — step: Pick up the popcorn from coffee table. [grasp]

goal: Heat the alphabet soup in a bowl using the microwave□pour a cup of milk into a mug,reposition t



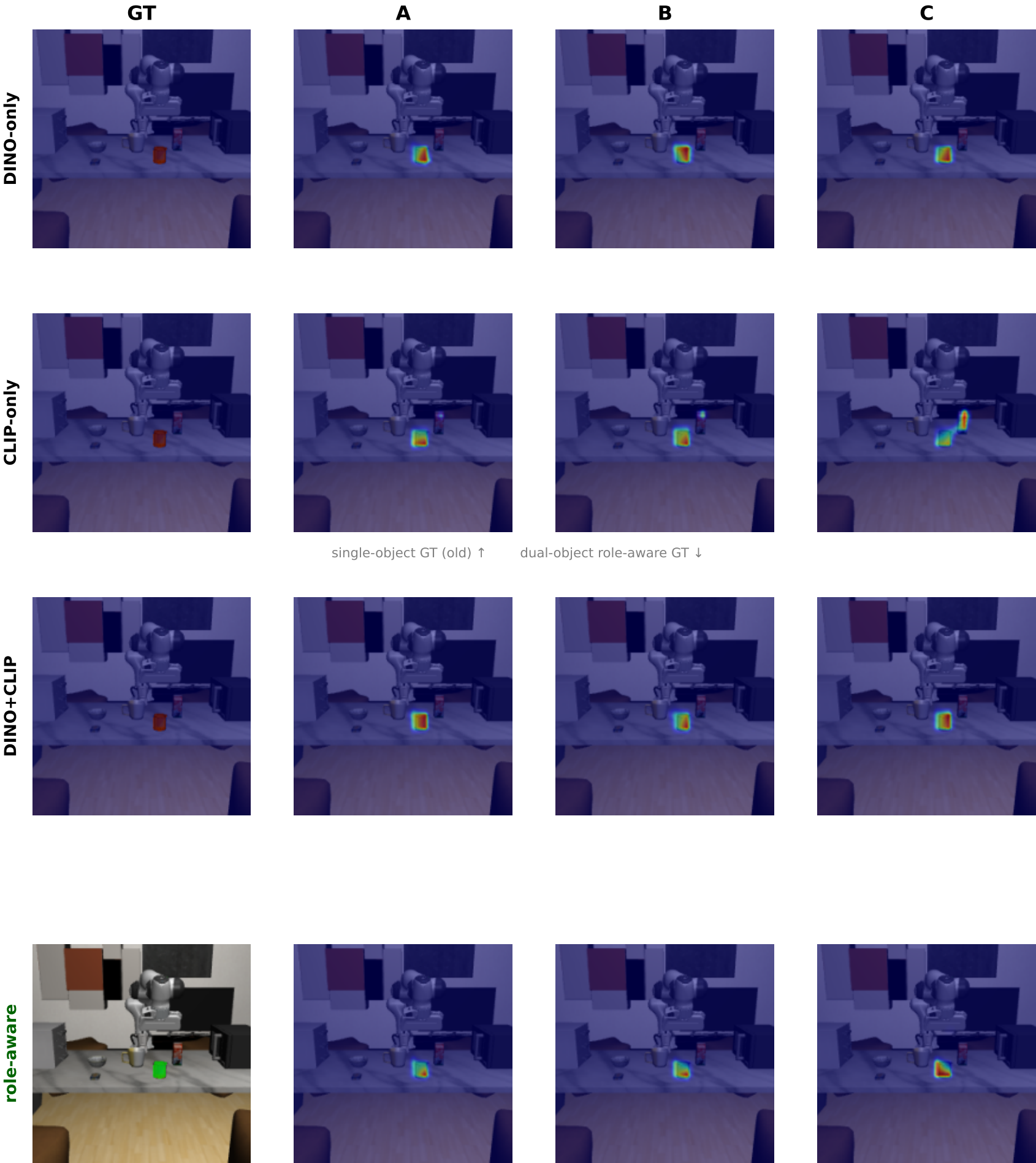
Mix_case5 — step: Place the popcorn on coffee table. [release]

goal: Heat the alphabet soup in a bowl using the microwave□pour a cup of milk into a mug, reposition t



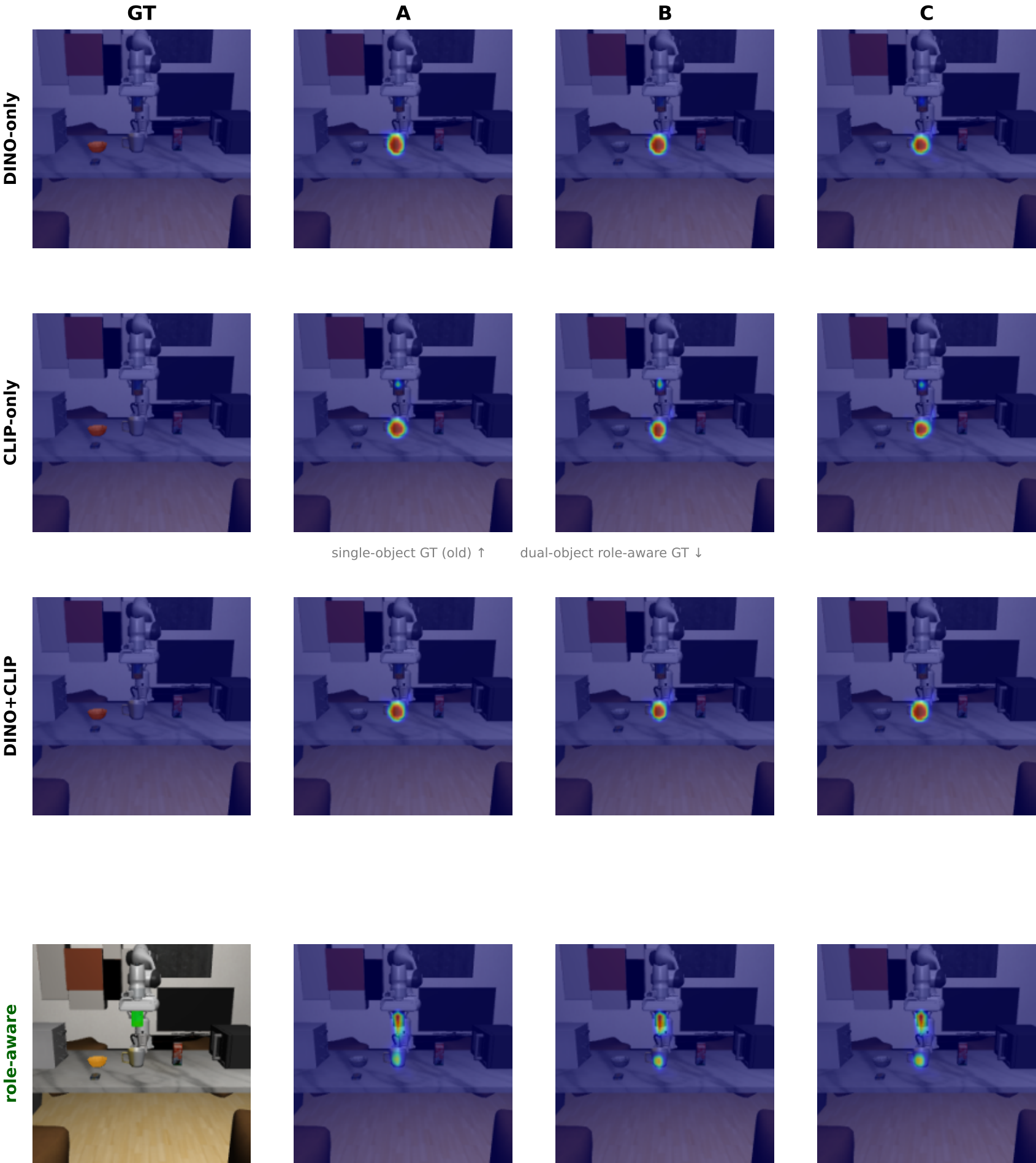
Mix_case5 — step: Pick up the alphabet soup from coffee table. [grasp]

goal: Heat the alphabet soup in a bowl using the microwave□pour a cup of milk into a mug, reposition t



Mix_case5 — step: Pour out the alphabet soup into the akita black bowl. [manipulate]

goal: Heat the alphabet soup in a bowl using the microwave□pour a cup of milk into a mug, reposition t



Mix_case5 — step: Place the empty alphabet soup on table surface. [release]

goal: Heat the alphabet soup in a bowl using the microwave□pour a cup of milk into a mug, reposition t

GT



A



B



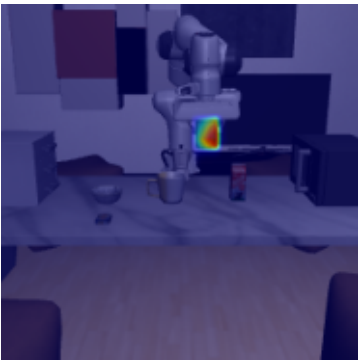
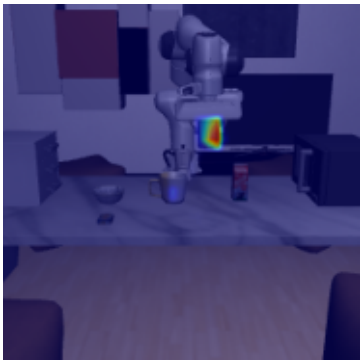
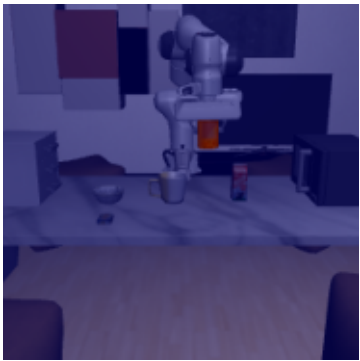
C



DINO-only



CLIP-only



DINO+CLIP

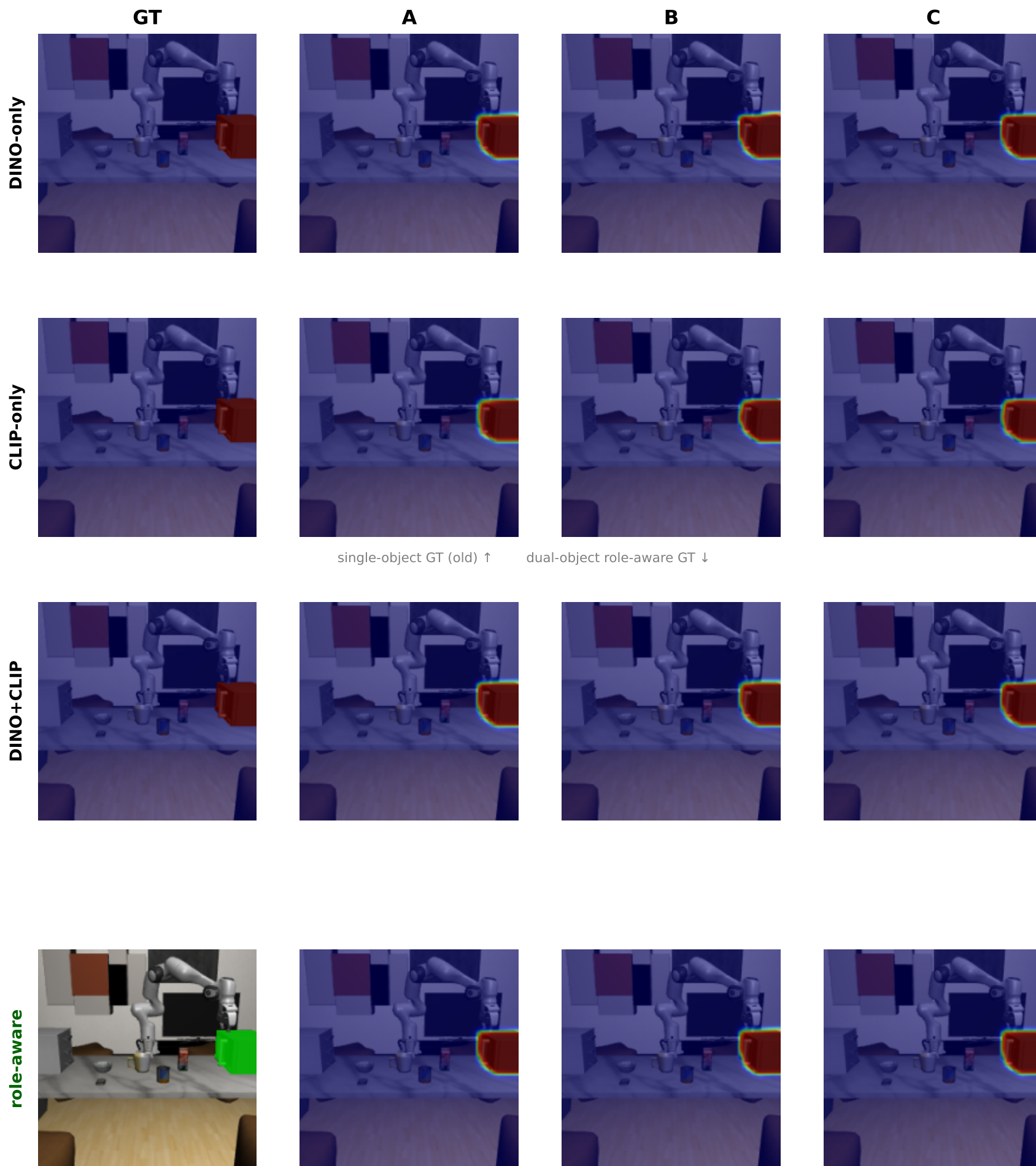


role-aware

single-object GT (old) ↑ dual-object role-aware GT ↓

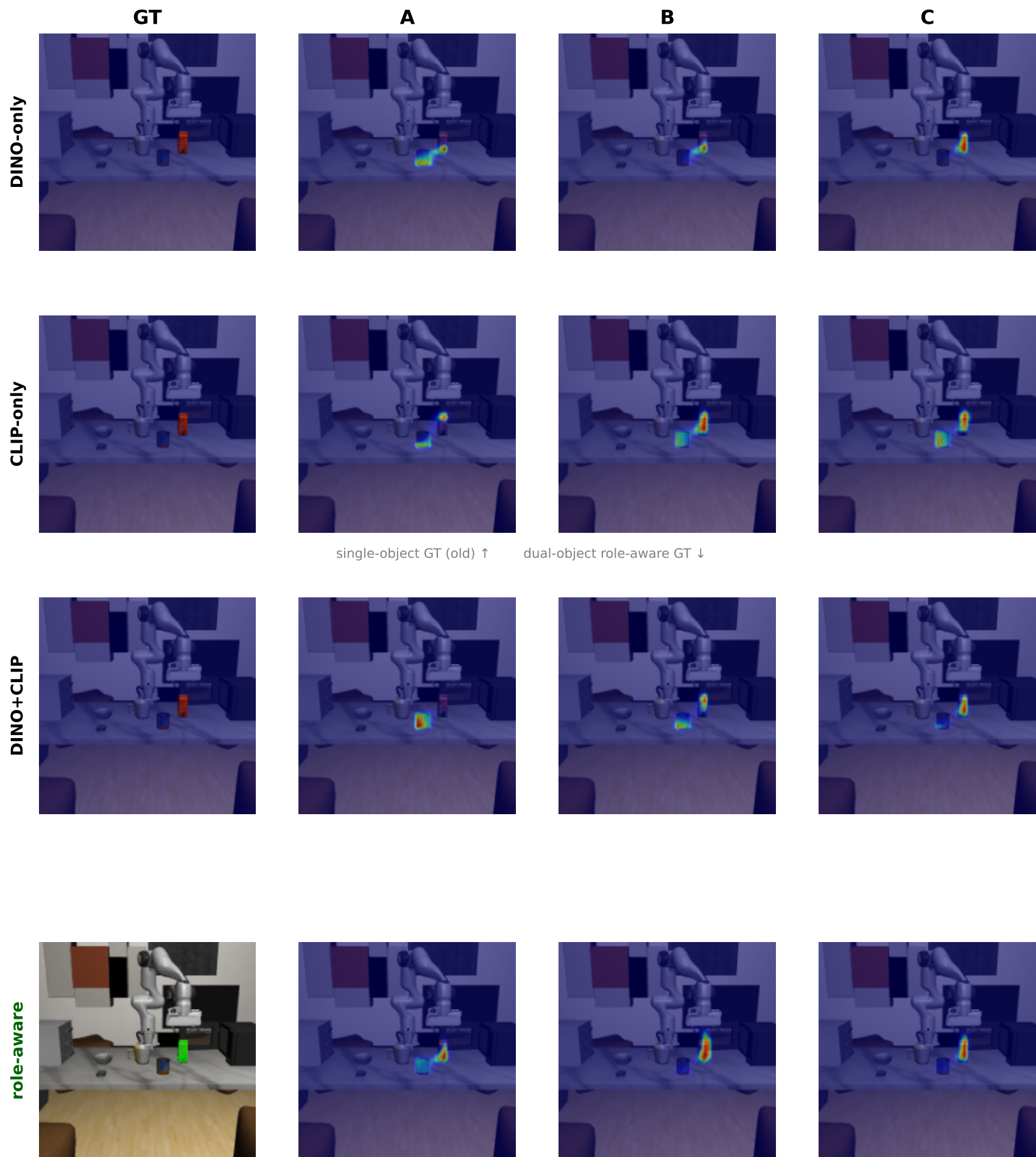
Mix_case5 — step: Open the microwave. [manipulate]

goal: Heat the alphabet soup in a bowl using the microwave□pour a cup of milk into a mug, reposition t



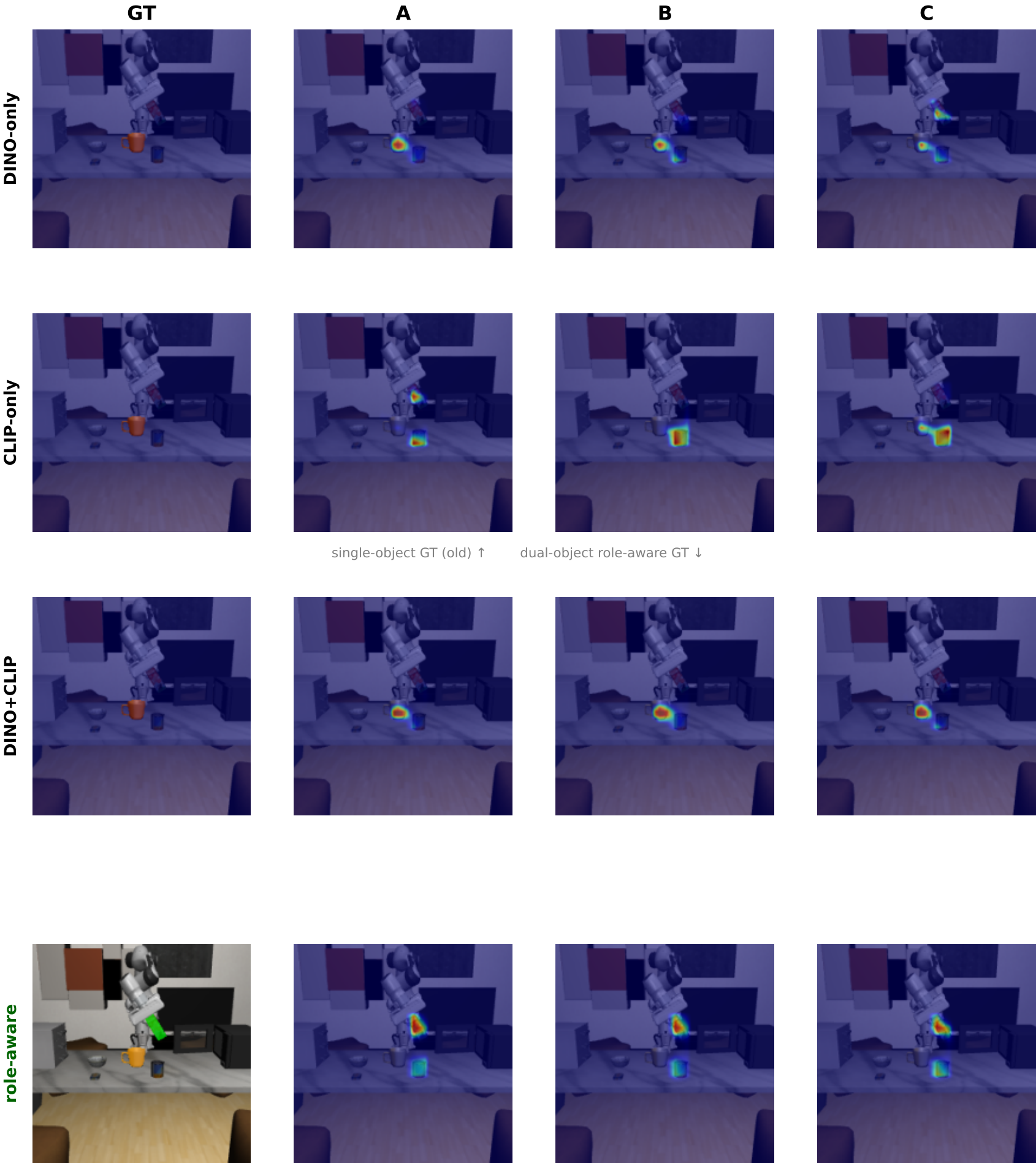
Mix_case5 — step: Pick up the milk from its storage. [grasp]

goal: Heat the alphabet soup in a bowl using the microwave□pour a cup of milk into a mug, reposition t



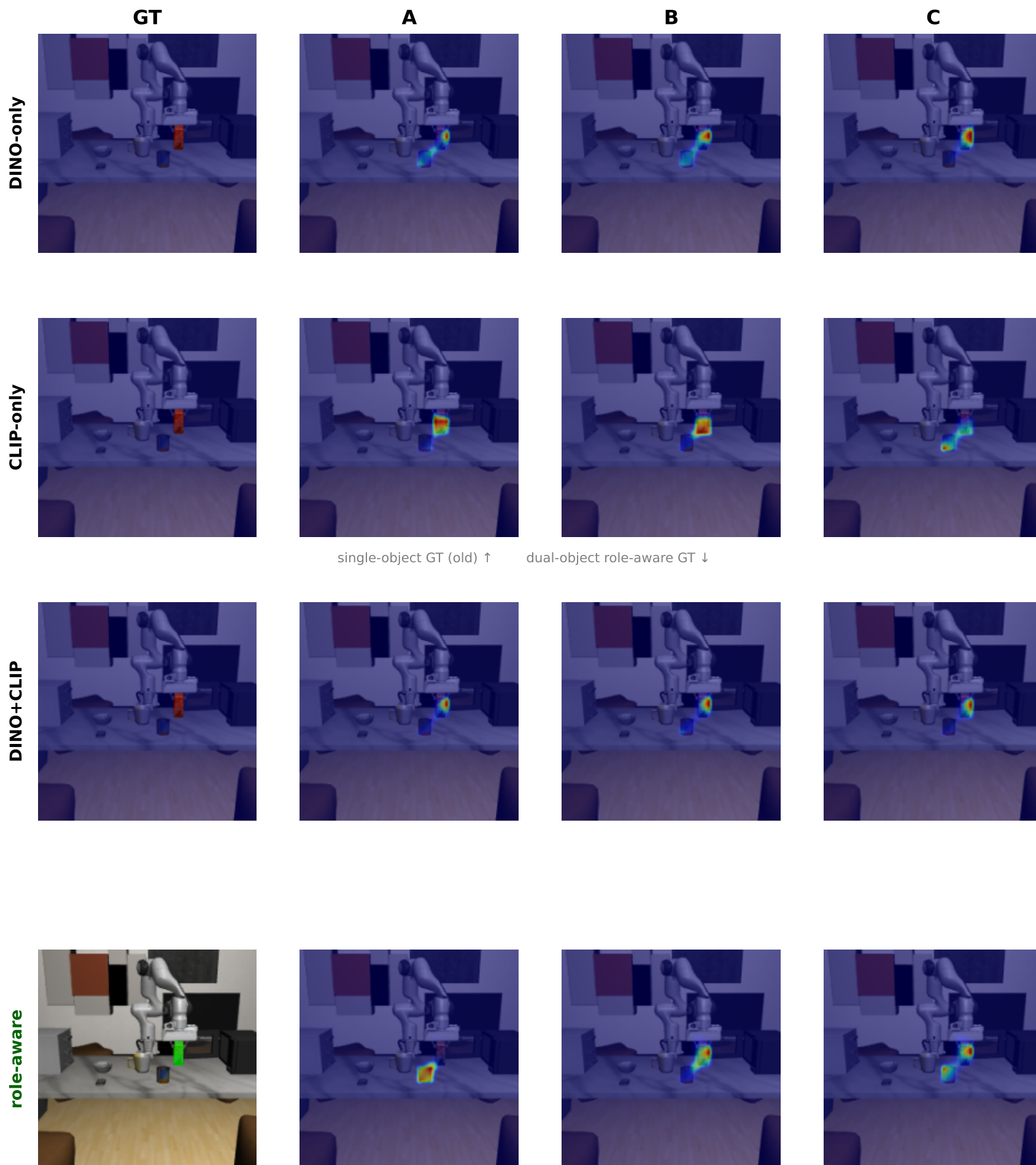
Mix_case5 — step: Pour out the milk into the white yellow mug. [manipulate]

goal: Heat the alphabet soup in a bowl using the microwave□pour a cup of milk into a mug, reposition t



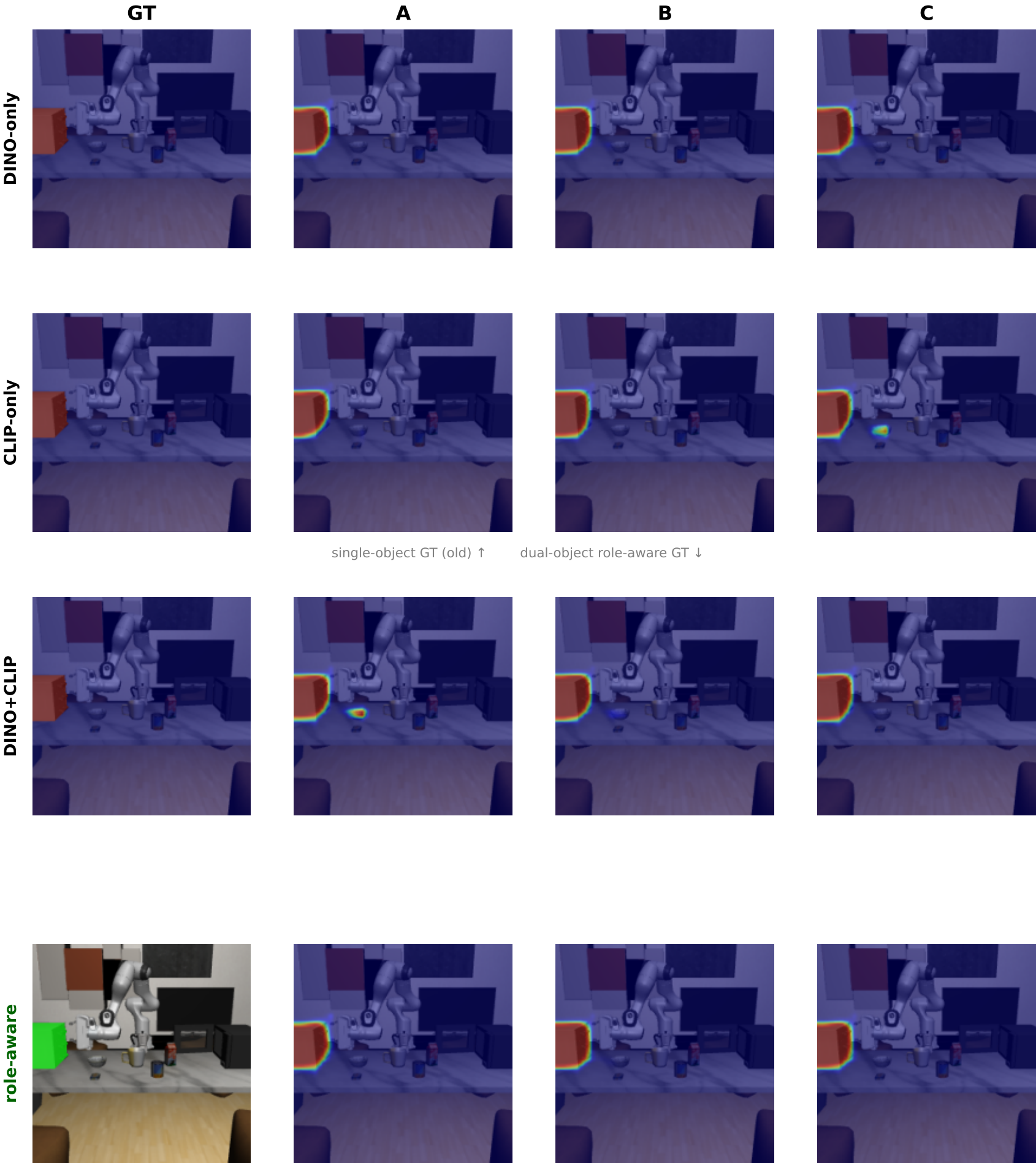
Mix_case5 — step: Place the milk on table surface. [release]

goal: Heat the alphabet soup in a bowl using the microwave□pour a cup of milk into a mug, reposition t



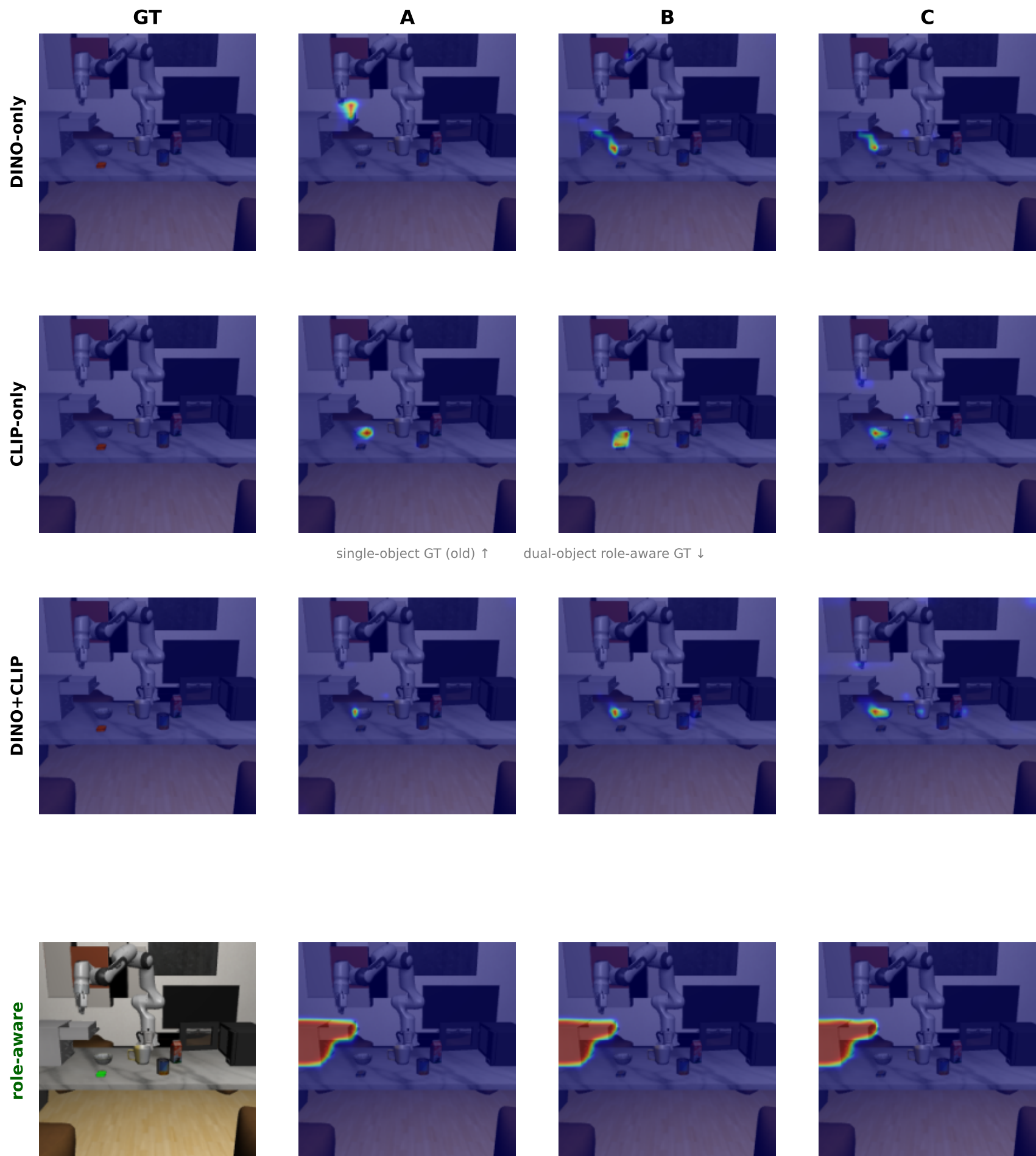
Mix_case5 — step: Open the top region of the white cabinet [manipulate]

goal: Heat the alphabet soup in a bowl using the microwave□pour a cup of milk into a mug, reposition t



Mix_case5 — step: Pick up the popcorn from coffee table. [grasp]

goal: Heat the alphabet soup in a bowl using the microwave□pour a cup of milk into a mug,reposition t



Mix_case5 — step: Place the popcorn on coffee table. [release]

goal: Heat the alphabet soup in a bowl using the microwave□pour a cup of milk into a mug,reposition t

