

Mix_case1 — step: Open the wooden cabinet top region [manipulate]

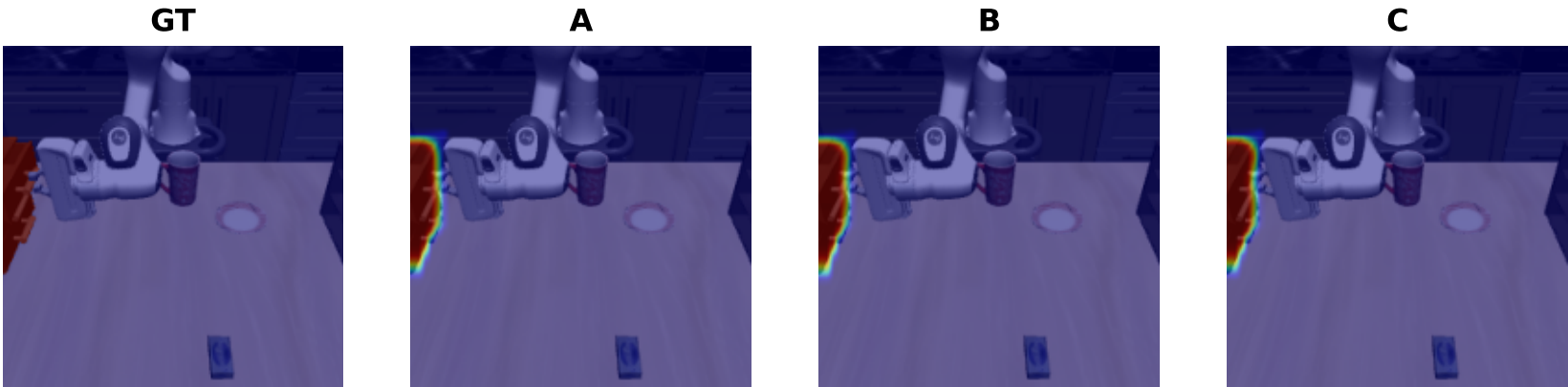
goal: Find and pick the butter from the wooden cabinet, pour the cream cheese from the kitchen table



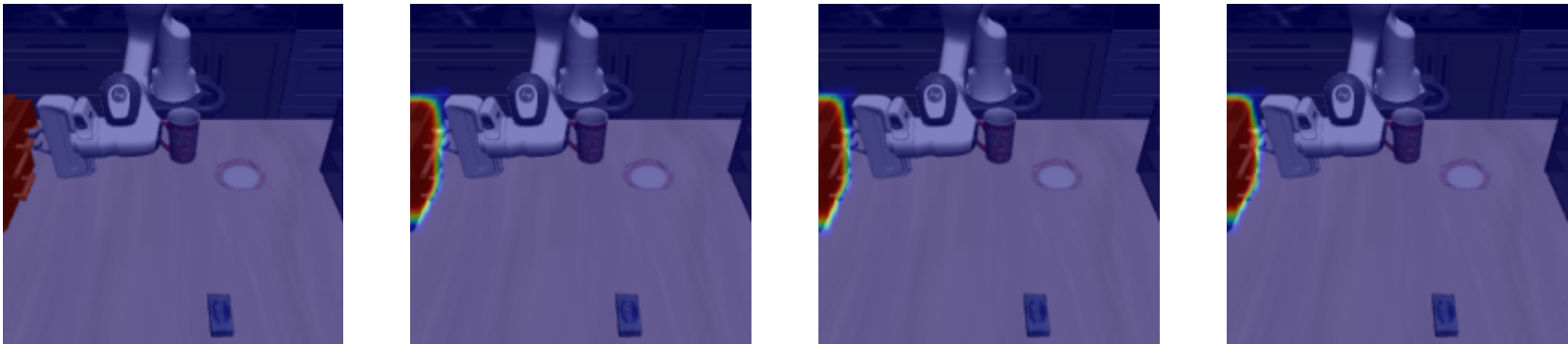
Mix_case1 — step: Close the wooden cabinet top region [manipulate]

goal: Find and pick the butter from the wooden cabinet, pour the cream cheese from the kitchen table

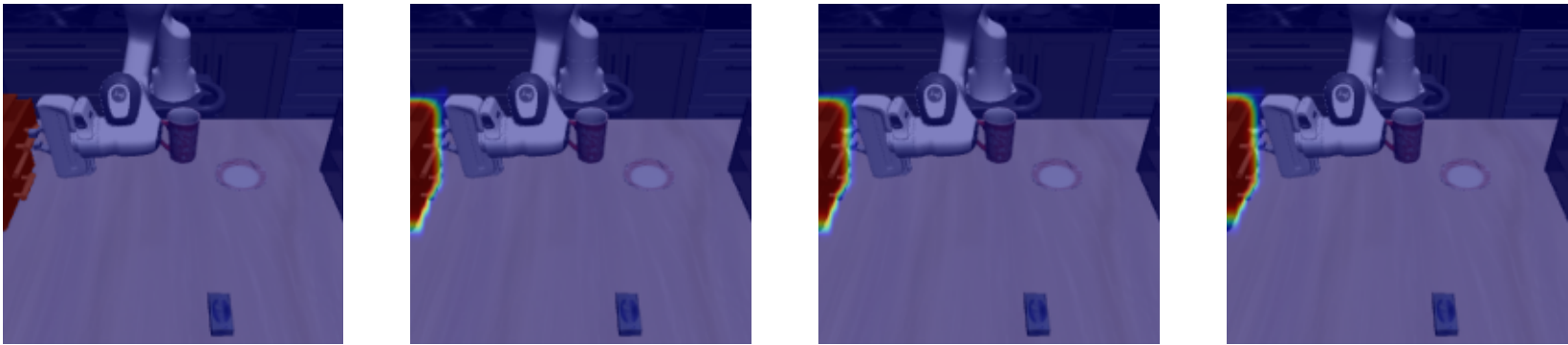
DINO-only



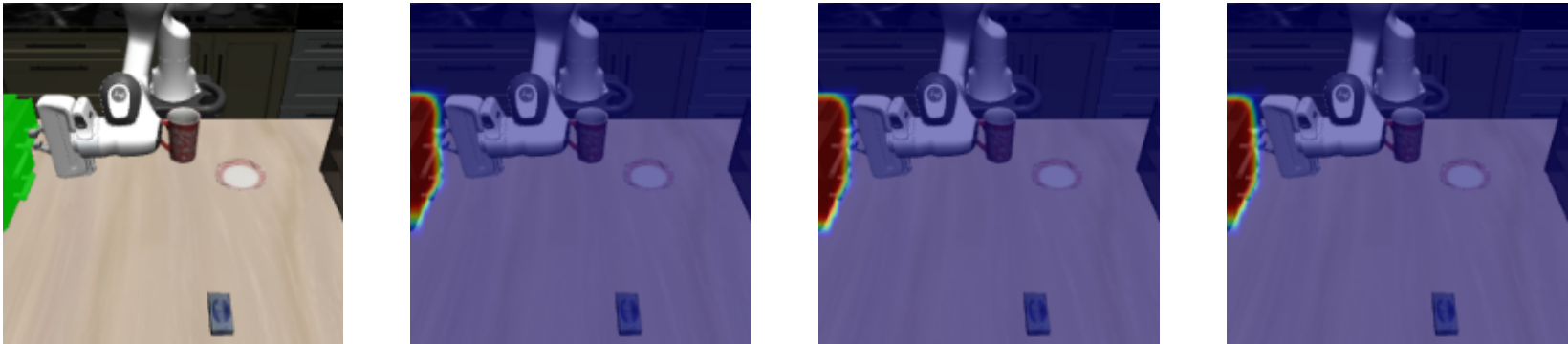
CLIP-only



DINO+CLIP



role-aware



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

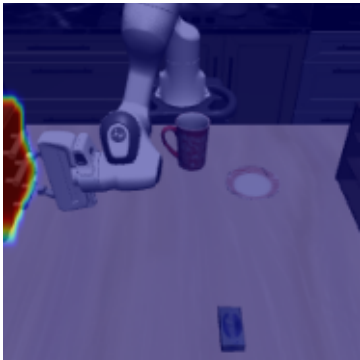
Mix_case1 — step: Open the wooden cabinet middle region [manipulate]

goal: Find and pick the butter from the wooden cabinet, pour the cream cheese from the kitchen table

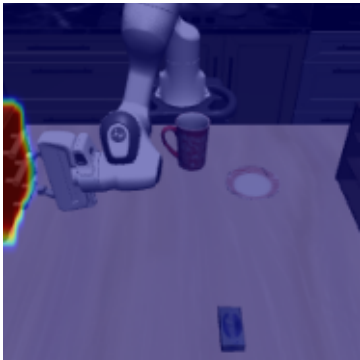
DINO-only



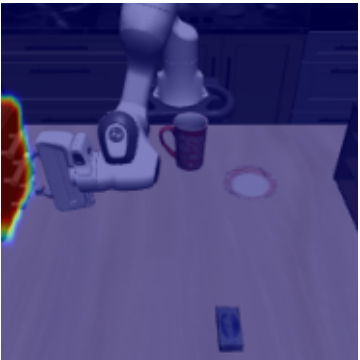
A



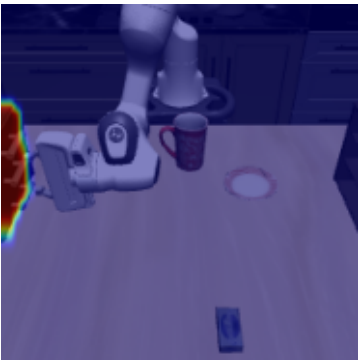
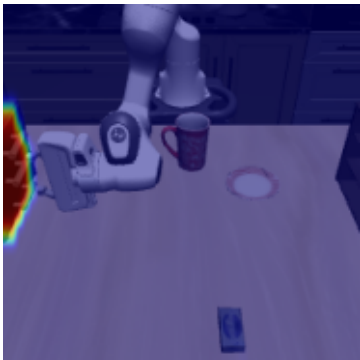
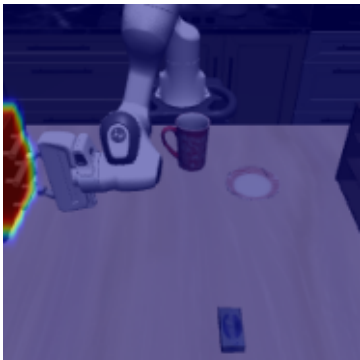
B



C

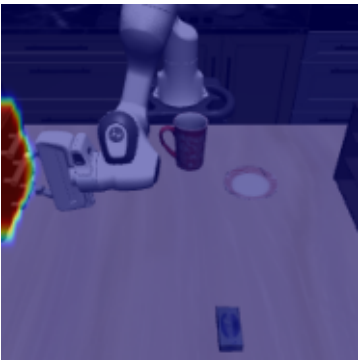
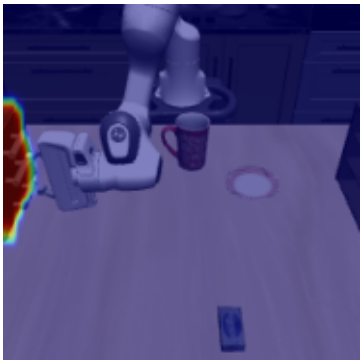
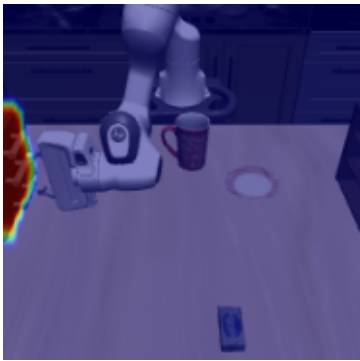


CLIP-only

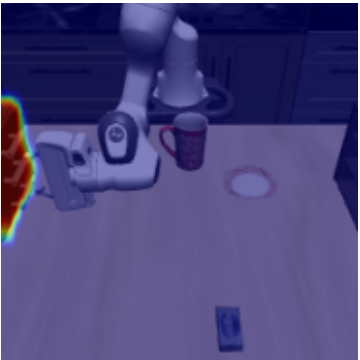
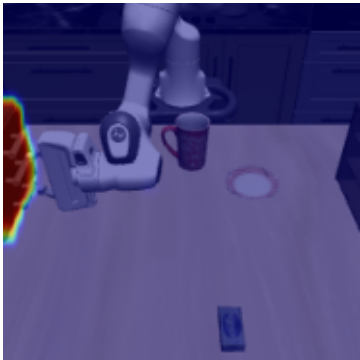
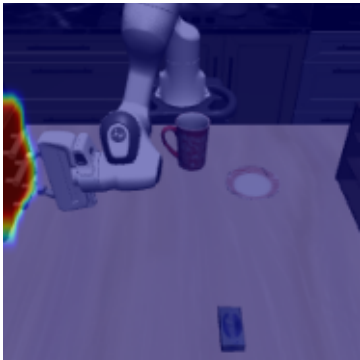


single-object GT (old) ↑ dual-object role-aware GT ↓

DINO+CLIP



role-aware



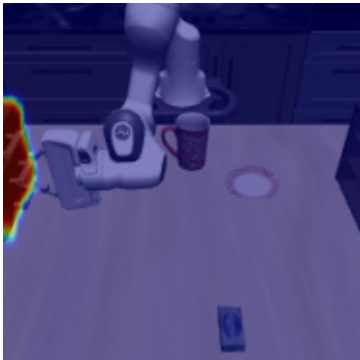
Mix_case1 — step: Close the wooden cabinet middle region [manipulate]

goal: Find and pick the butter from the wooden cabinet, pour the cream cheese from the kitchen table

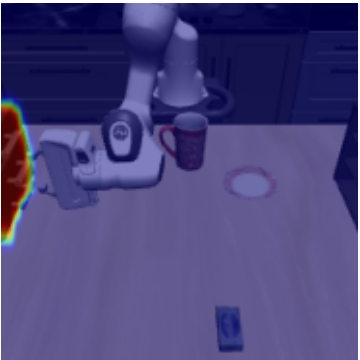
DINO-only



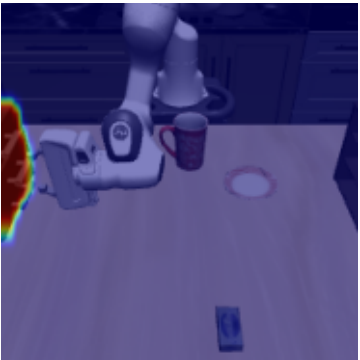
A



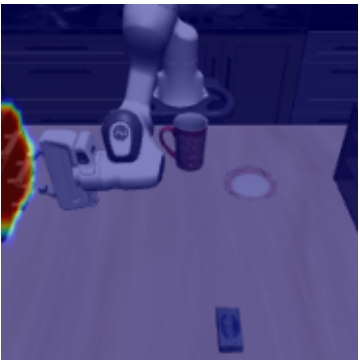
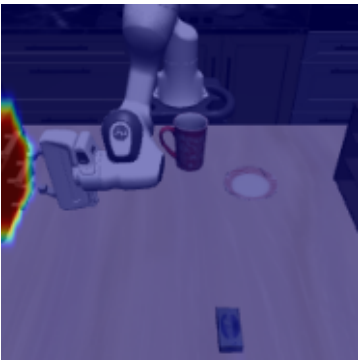
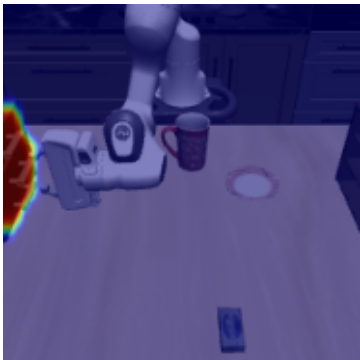
B



C

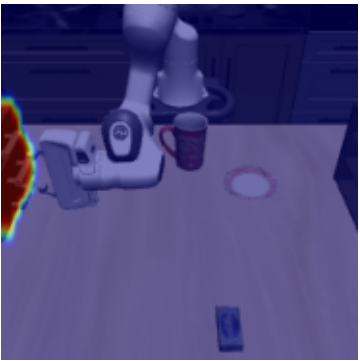
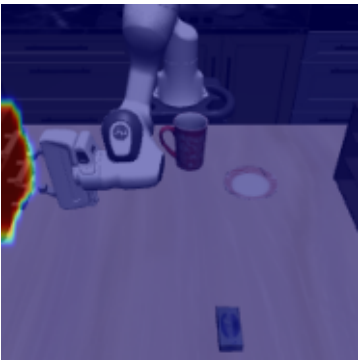
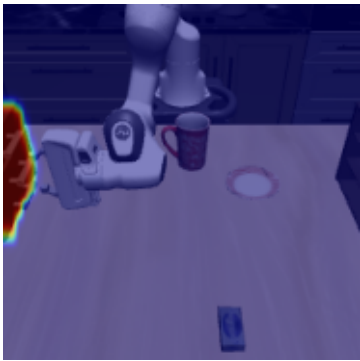


CLIP-only

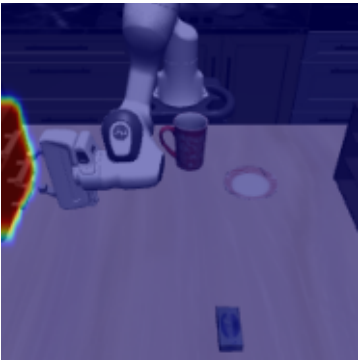
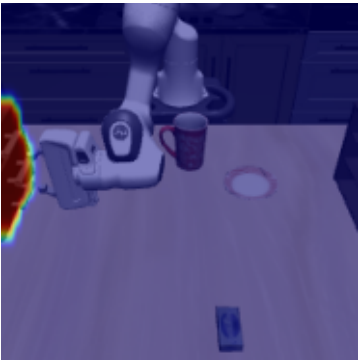
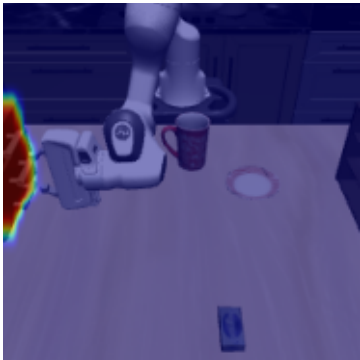


single-object GT (old) ↑ dual-object role-aware GT ↓

DINO+CLIP

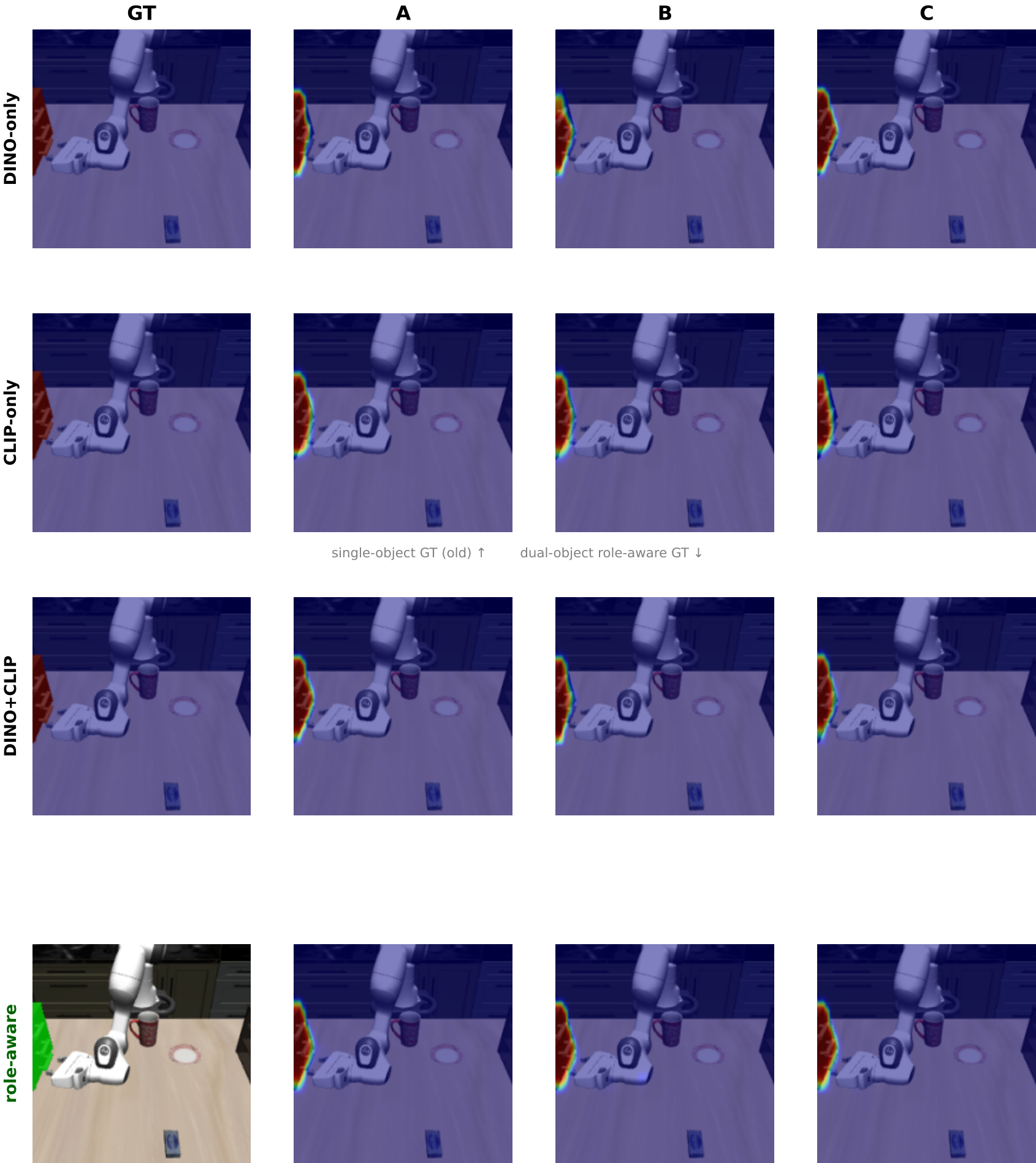


role-aware



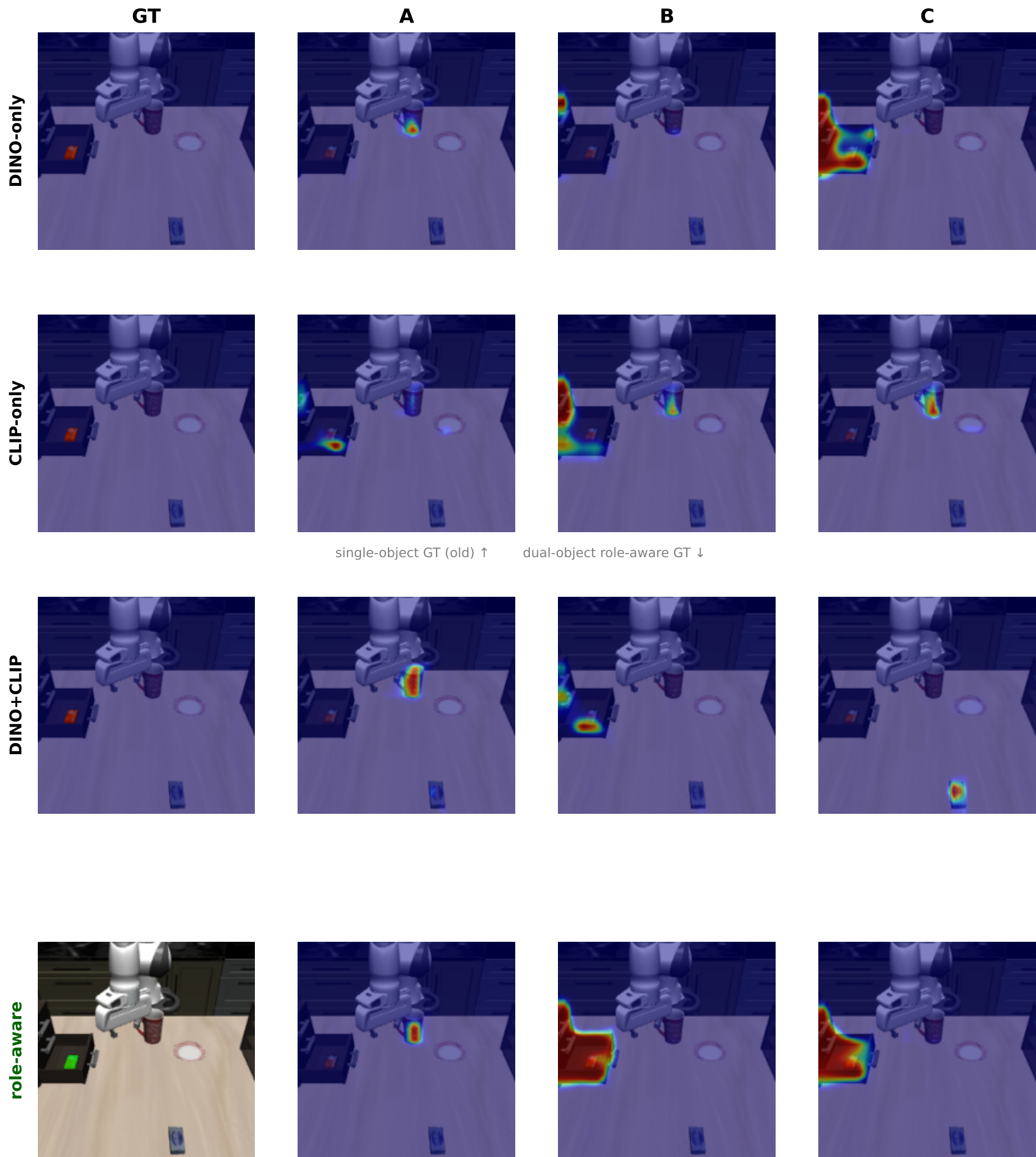
Mix_case1 — step: Open the wooden cabinet bottom region [manipulate]

goal: Find and pick the butter from the wooden cabinet, pour the cream cheese from the kitchen table



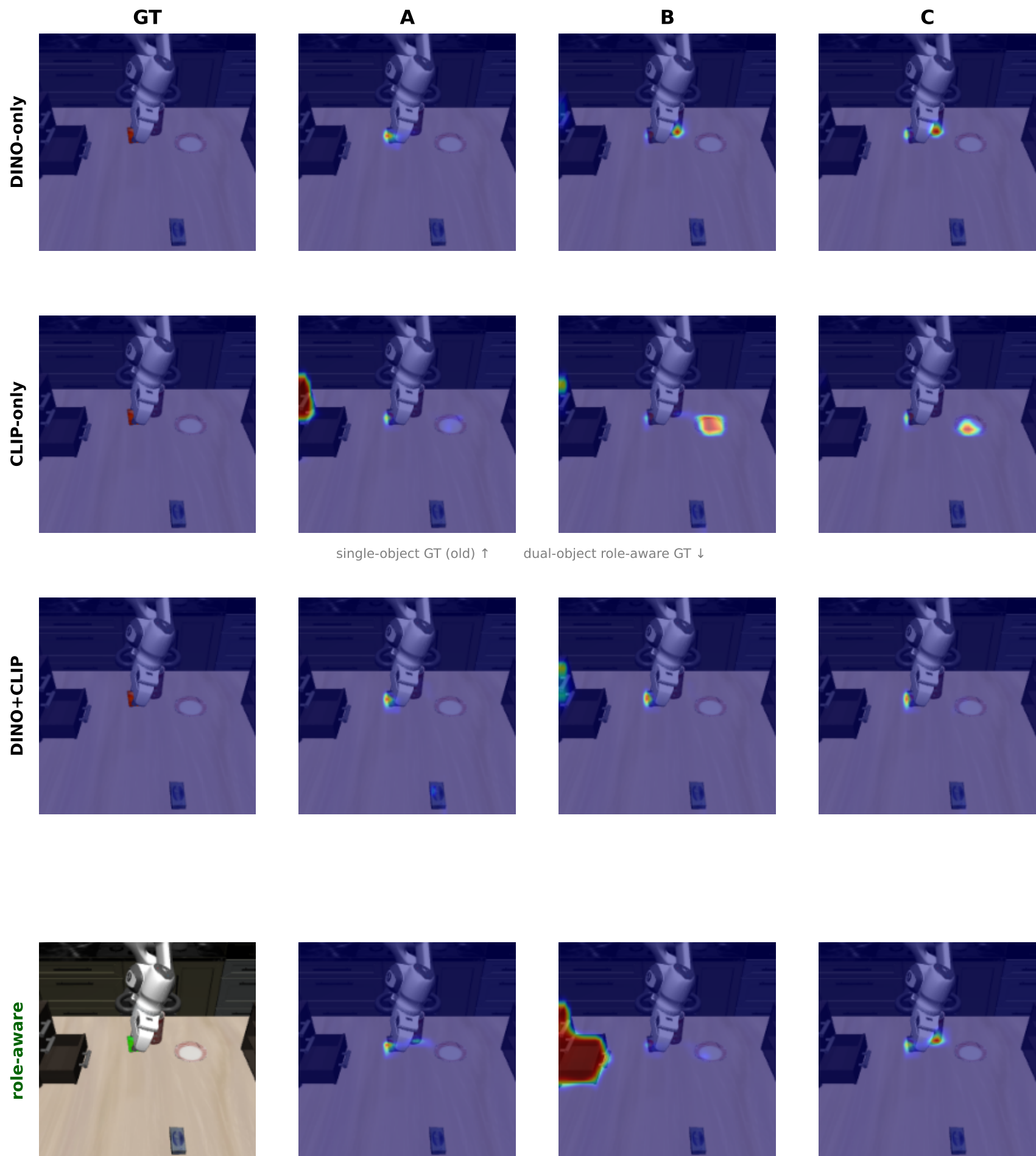
Mix_case1 — step: Pick up the butter from the wooden cabinet bottom region [grasp]

goal: Find and pick the butter from the wooden cabinet, pour the cream cheese from the kitchen table



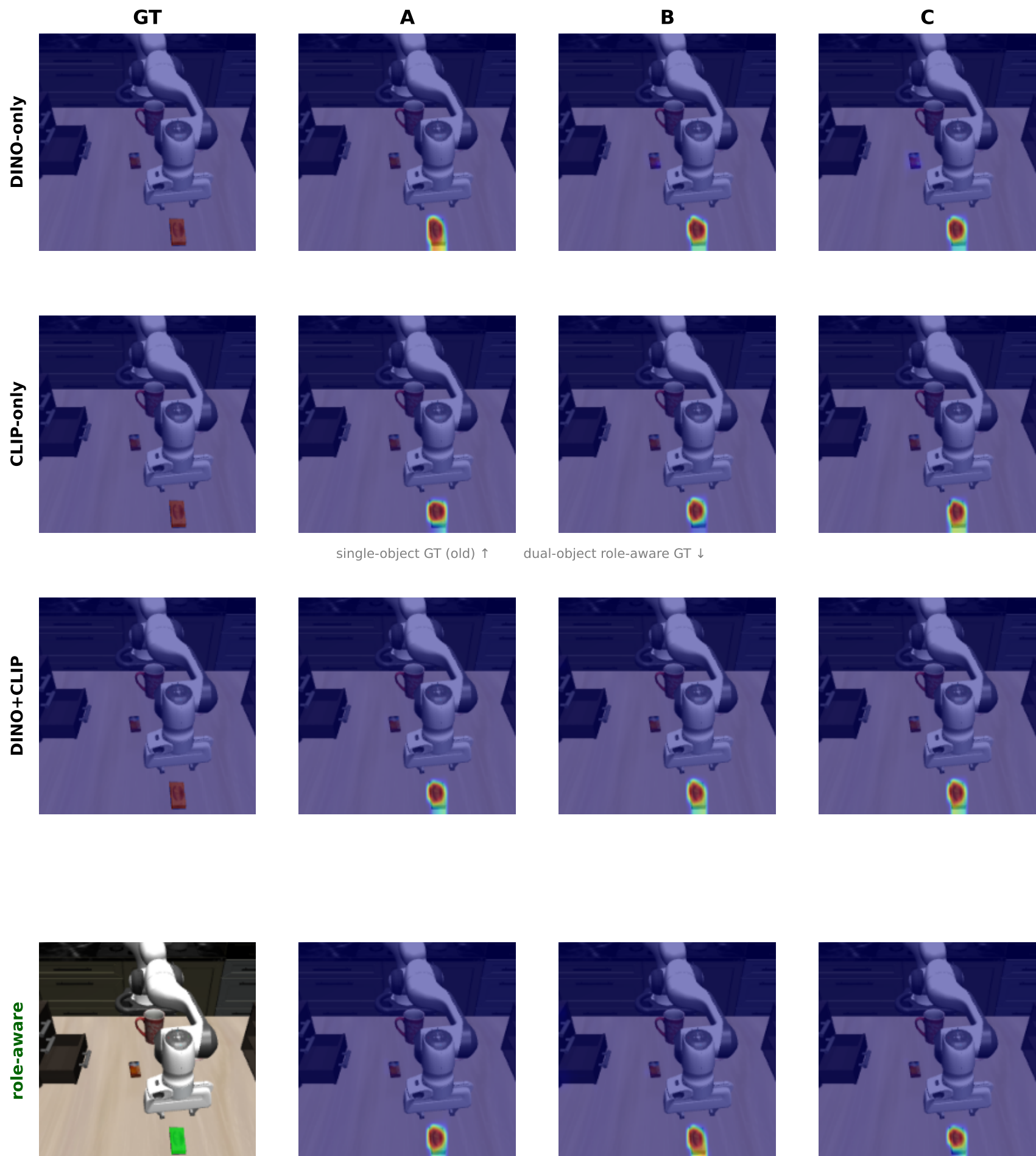
Mix_case1 — step: Place the butter on the kitchen table [release]

goal: Find and pick the butter from the wooden cabinet, pour the cream cheese from the kitchen table



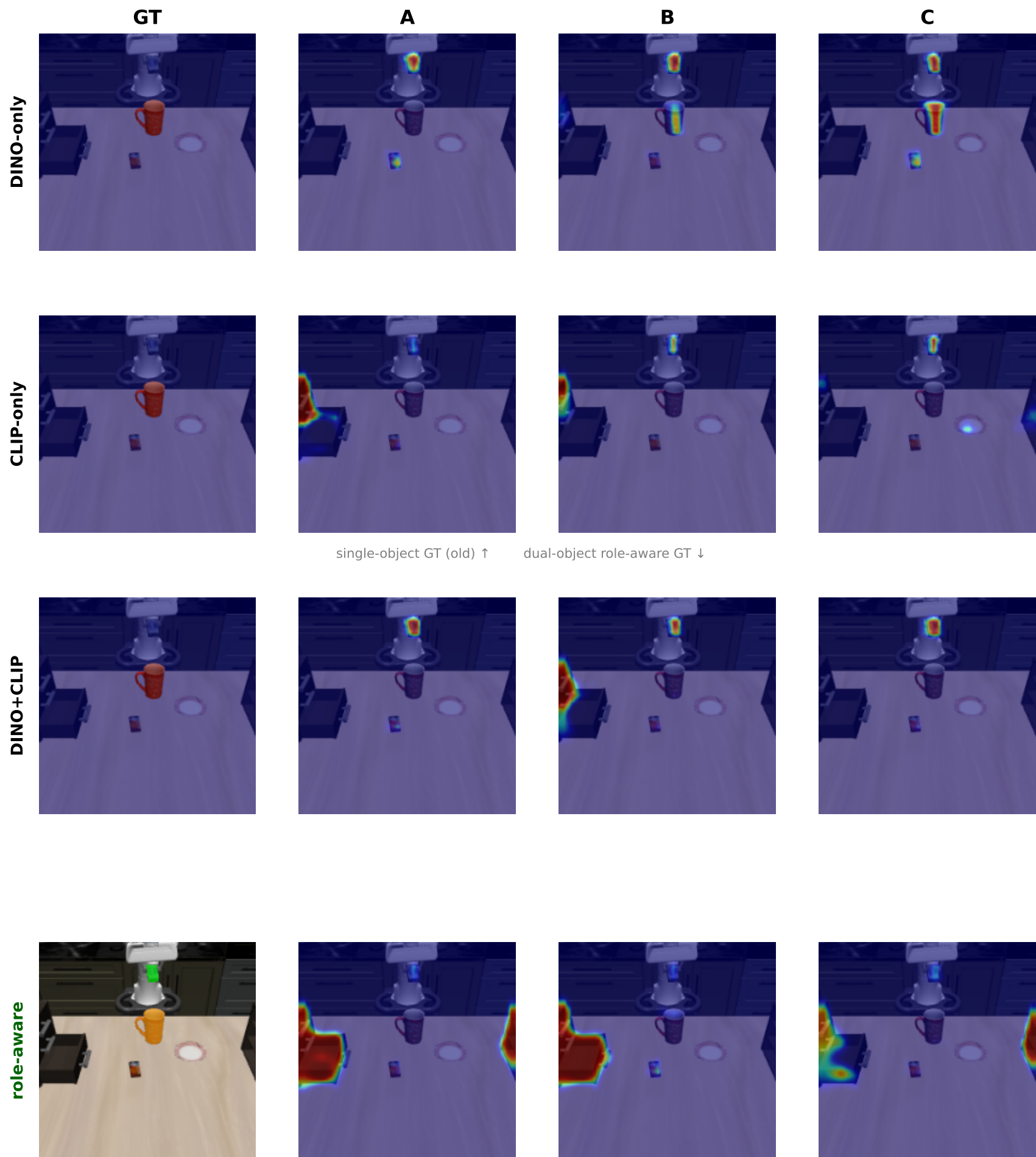
Mix_case1 — step: Pick up the cream cheese from the kitchen table [grasp]

goal: Find and pick the butter from the wooden cabinet, pour the cream cheese from the kitchen table



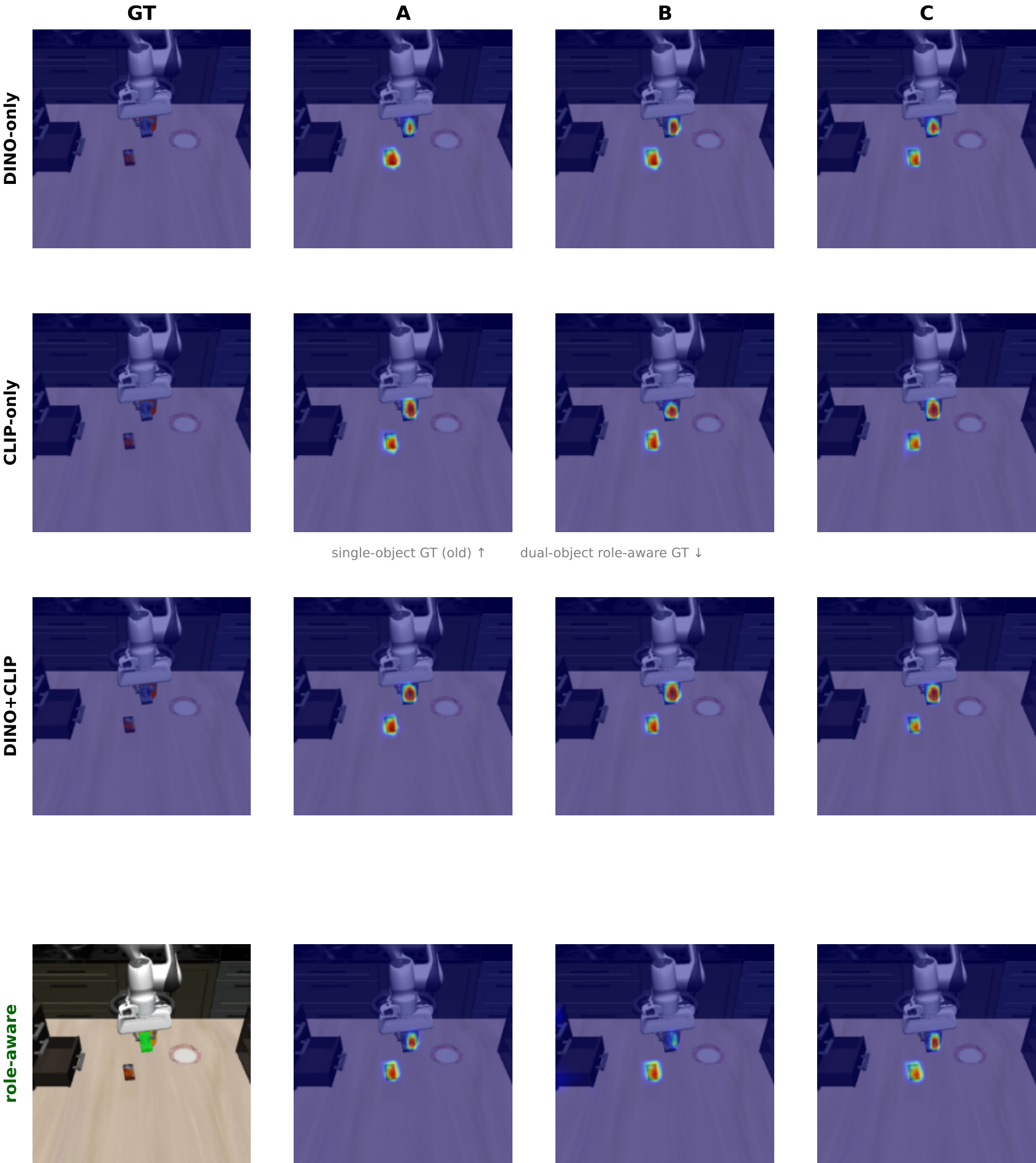
Mix_case1 — step: Pour the cream cheese into the red coffee mug [manipulate]

goal: Find and pick the butter from the wooden cabinet, pour the cream cheese from the kitchen table



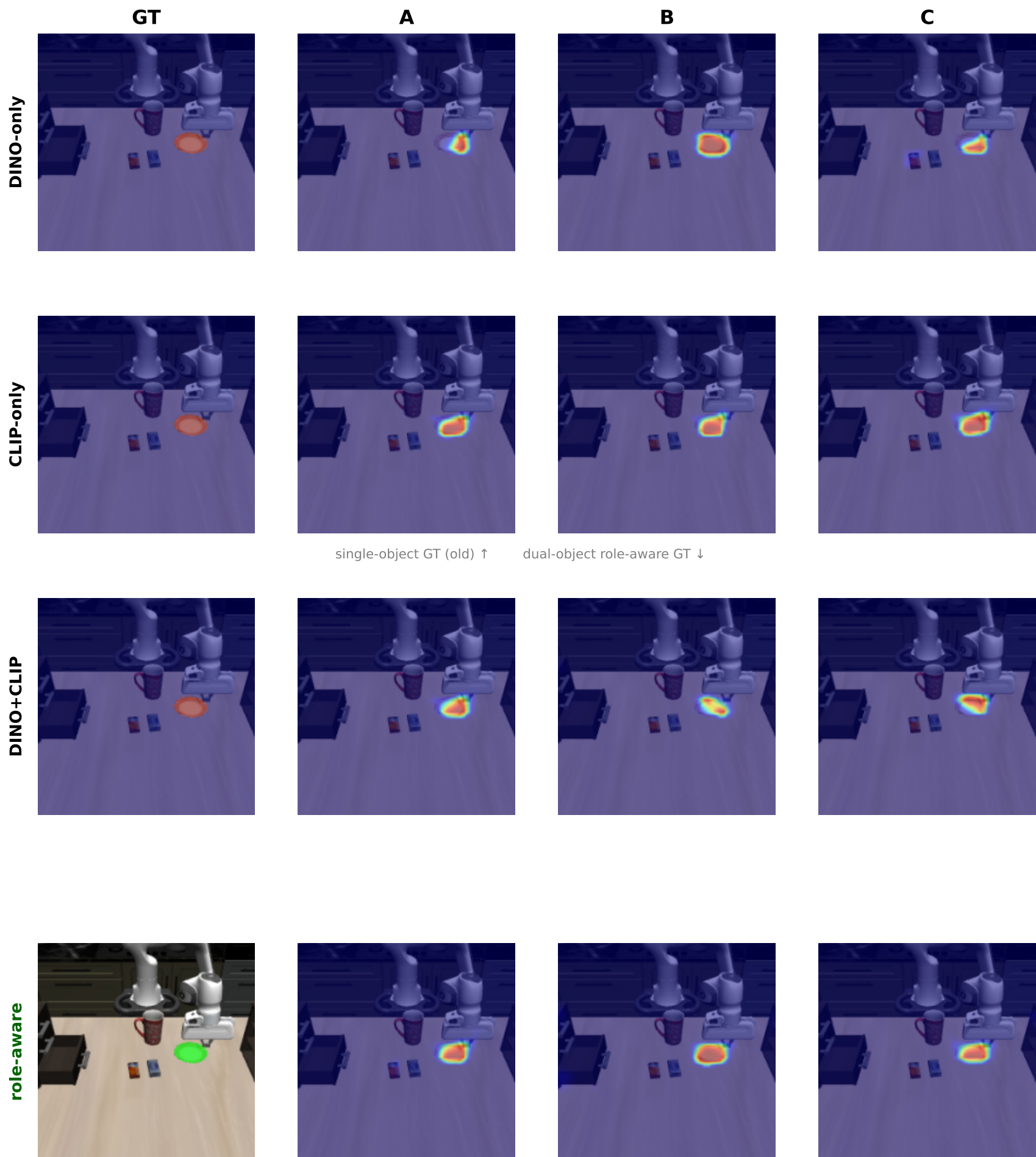
Mix_case1 — step: Place cream cheese on the kitchen table [release]

goal: Find and pick the butter from the wooden cabinet, pour the cream cheese from the kitchen table



Mix_case1 — step: Pick up the plate from the kitchen table [grasp]

goal: Find and pick the butter from the wooden cabinet, pour the cream cheese from the kitchen table



Mix_case1 — step: Place the plate into the wooden shelf top side [release]

goal: Find and pick the butter from the wooden cabinet, pour the cream cheese from the kitchen table

