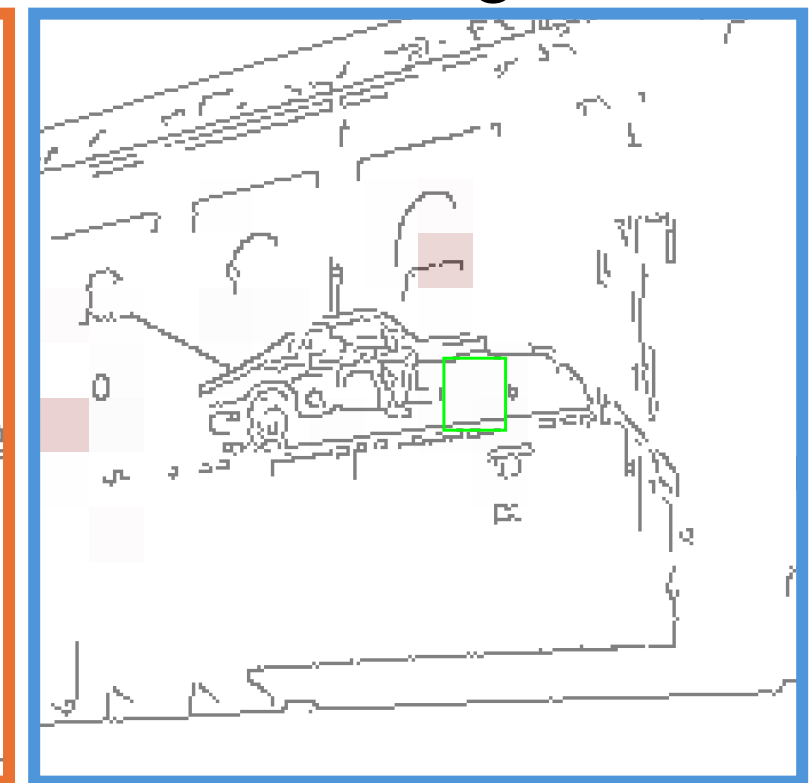
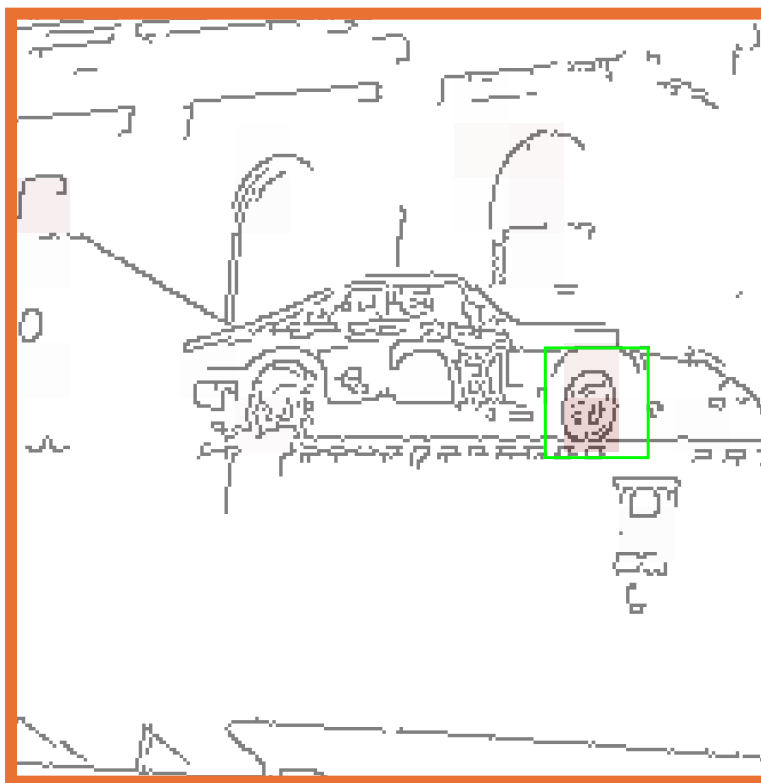


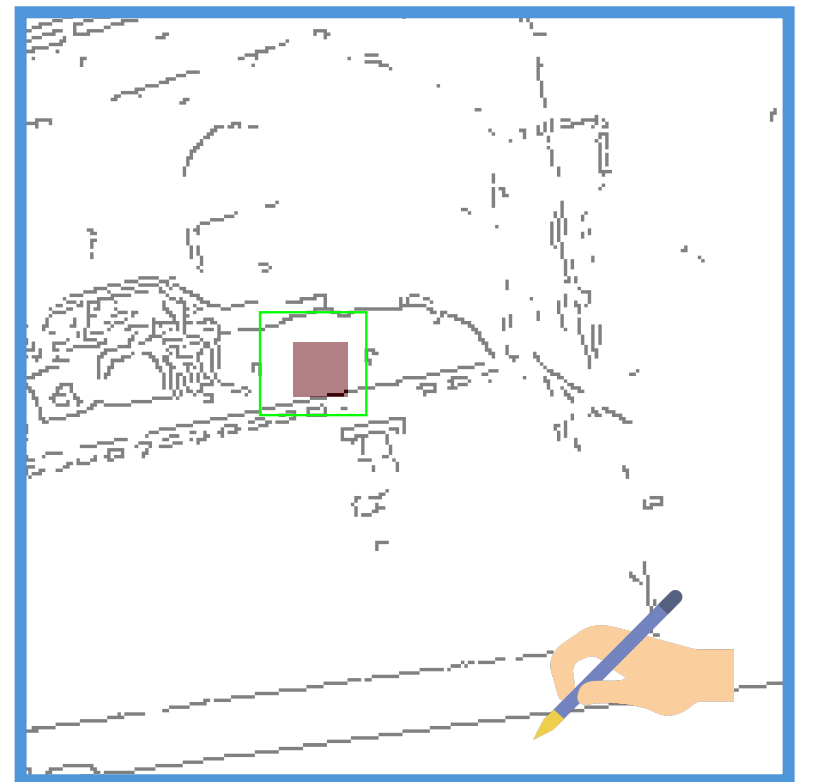
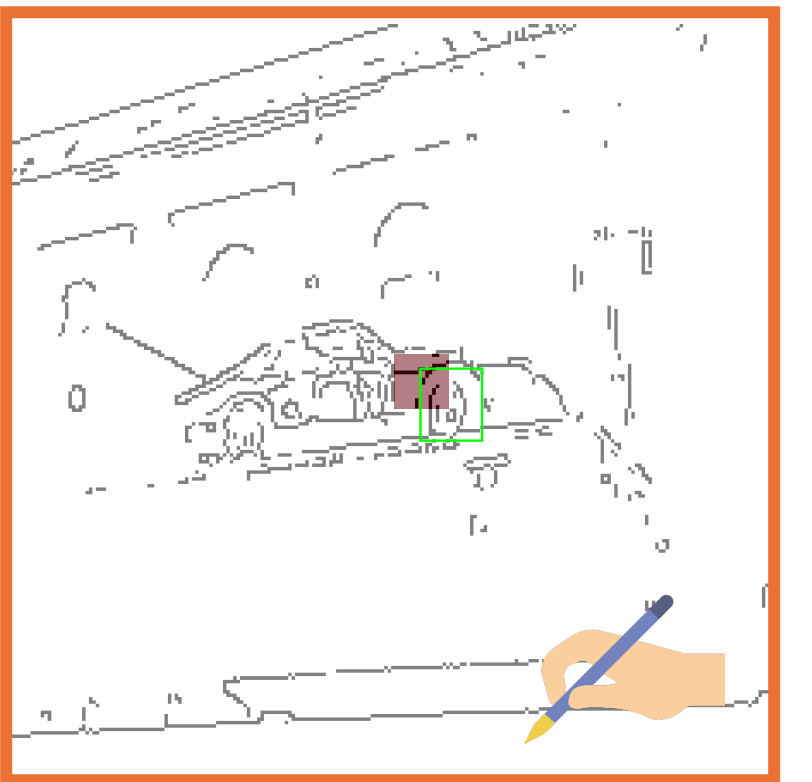
(b) MHPA attention maps  
CLIP4IDC: the window is missing ❌



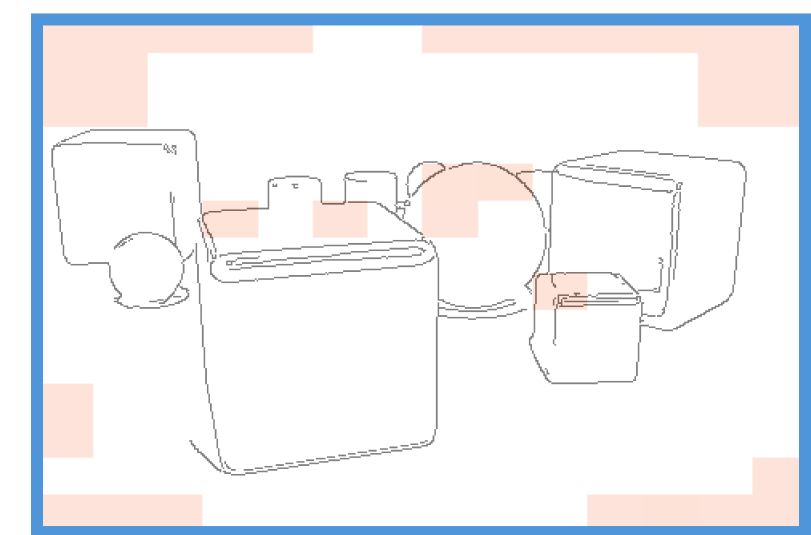
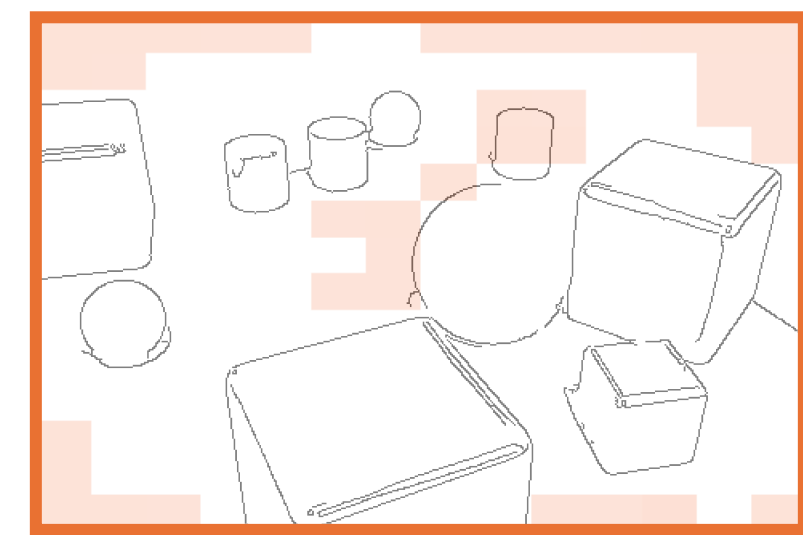
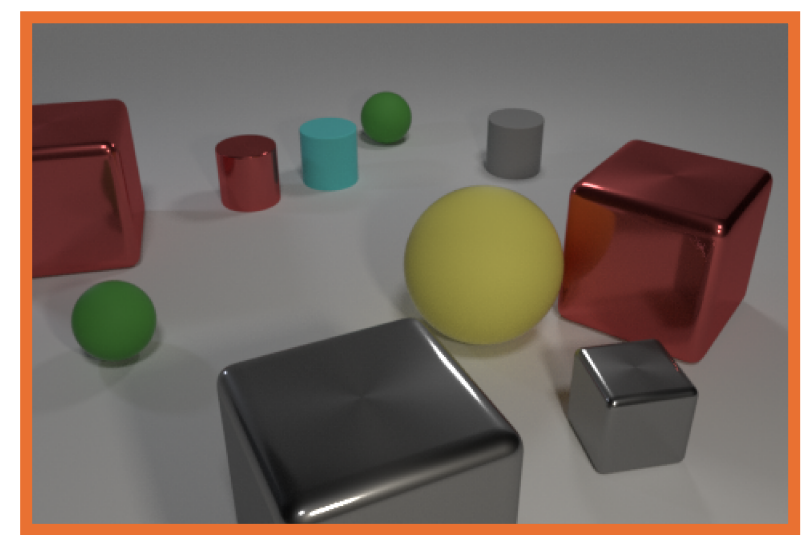
(a) Change images  
GT: the wheel is missing

(c) TAB attention maps  
TAB4IDC: the window is missing ❌

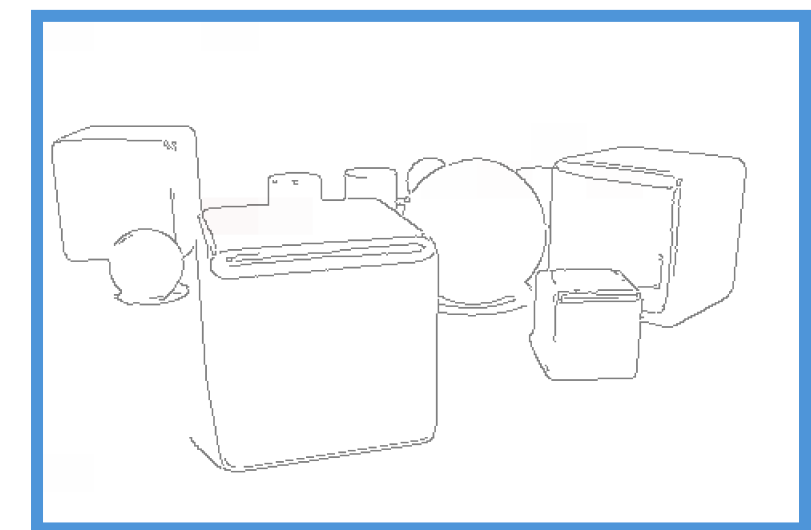
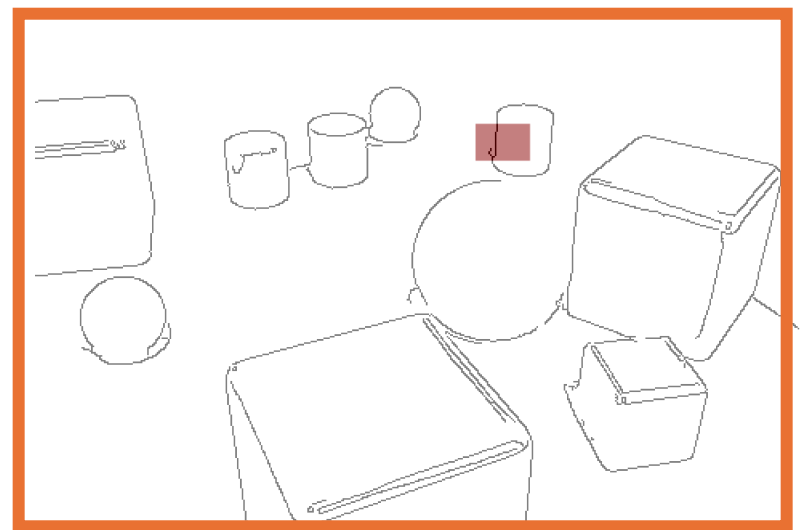
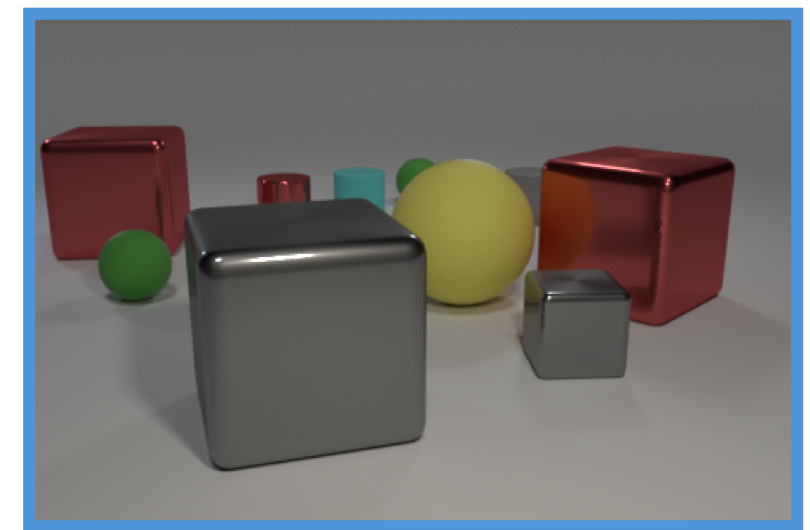
(d) Corrected MHPA attention maps  
CLIP4IDC: the window is missing ❌



(e) Corrected TAB attention maps  
TAB4IDC: the wheel is missing ✔️



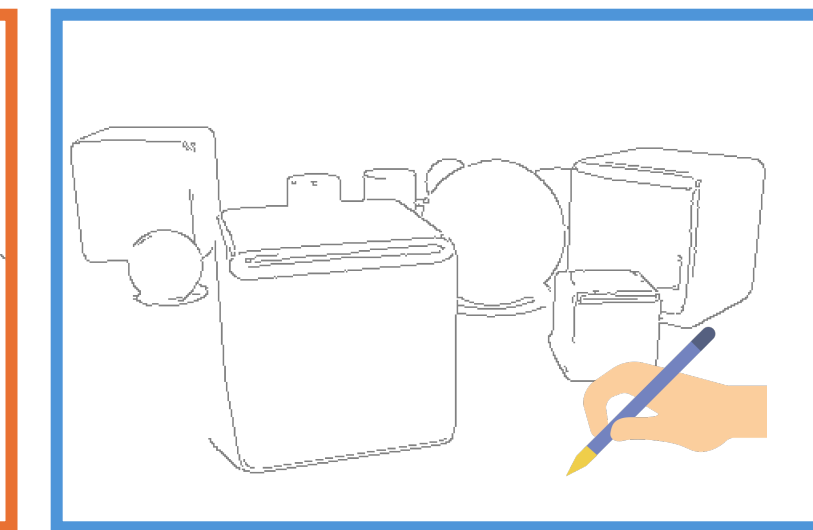
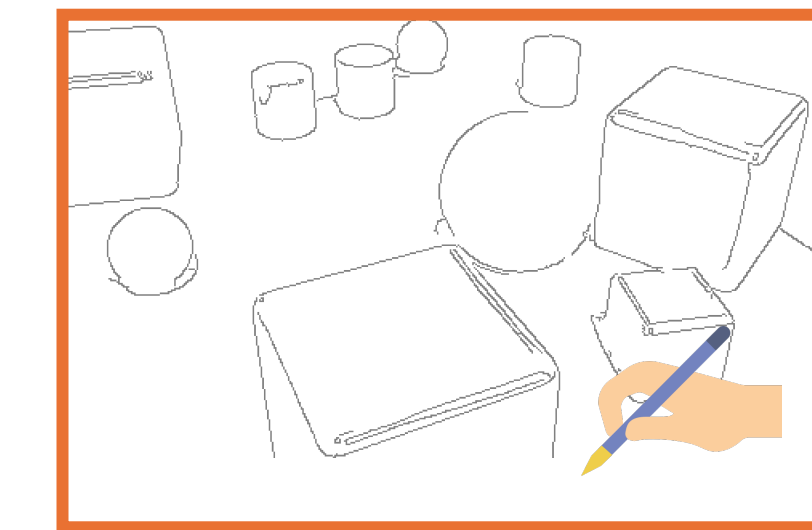
(g) MHPA attention maps  
CLIP4IDC: the small gray matte cylinder that is behind the small red matte object changed its location ❌



(f) No-Change images  
GT: there is no change

(h) TAB attention maps  
TAB4IDC: the tiny gray matte cylinder that is behind the big red metal object is gone ❌

(i) Zeroed MHPA attention maps  
CLIP4IDC: the small gray matte cylinder that is behind the small red matte object changed its location ❌



(j) Zeroed TAB attention maps  
TAB4IDC: there is no change ✔️