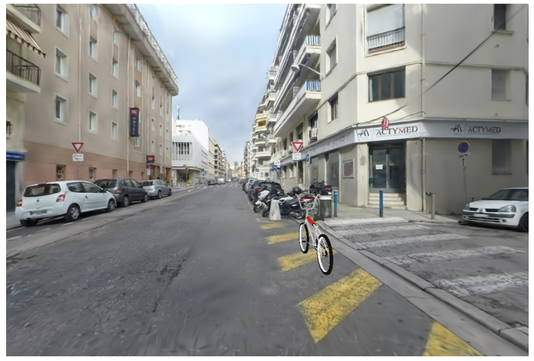
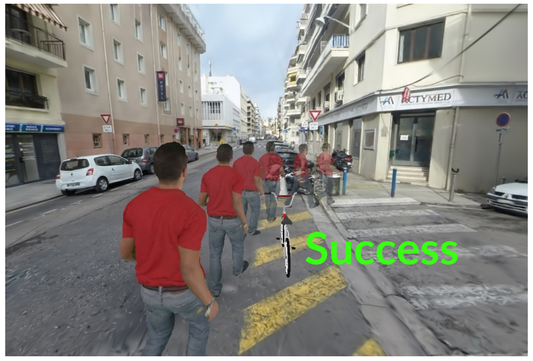
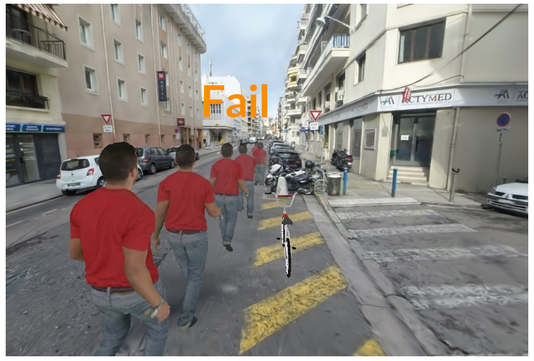
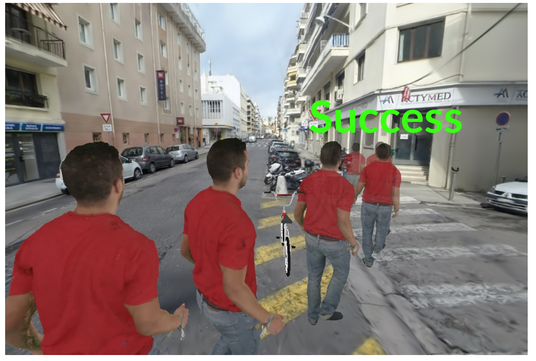
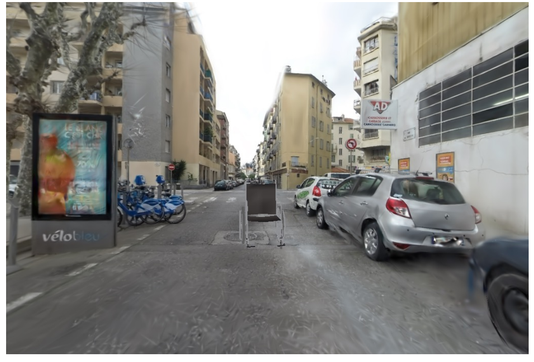
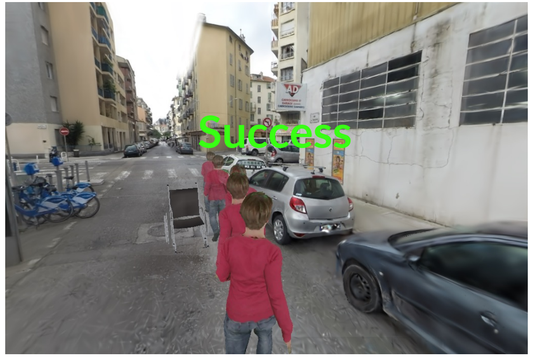
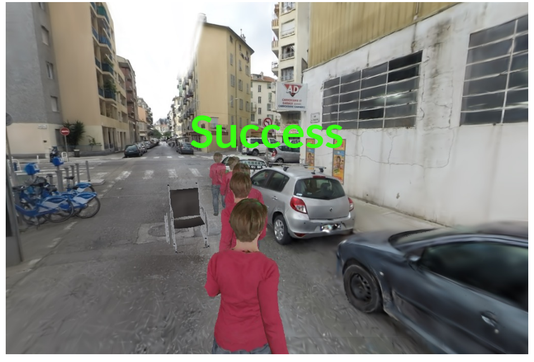
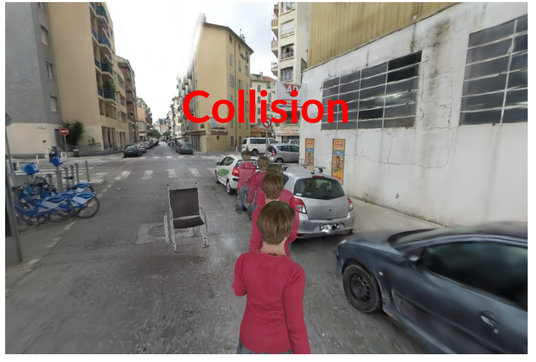
Goal	Initial Ego-centric View	Ours	w/o Visual Prompting	w/o Iterative Reasoning
"Green cylindrical trash bin"				
"White service car with 'GARNERO' branding"				
"Light blue sedan car"