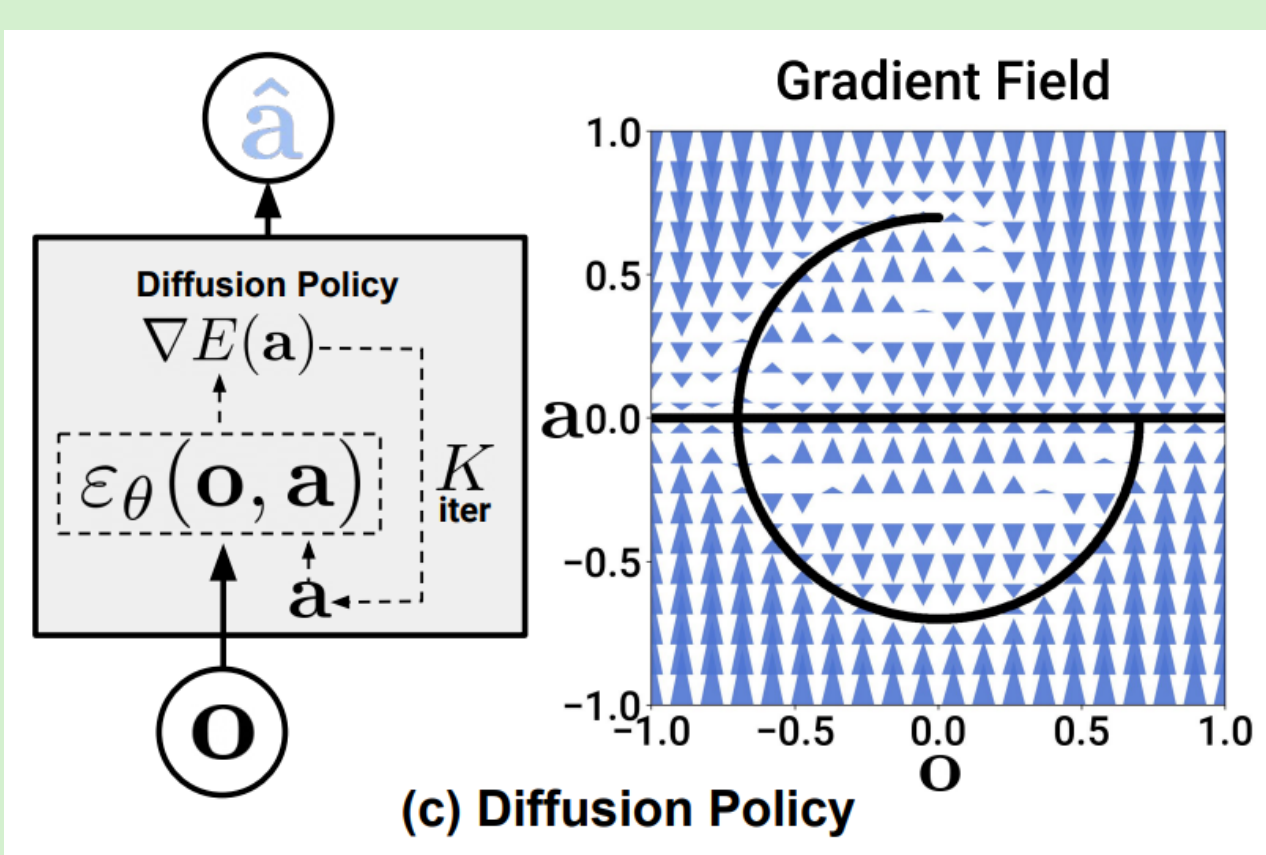
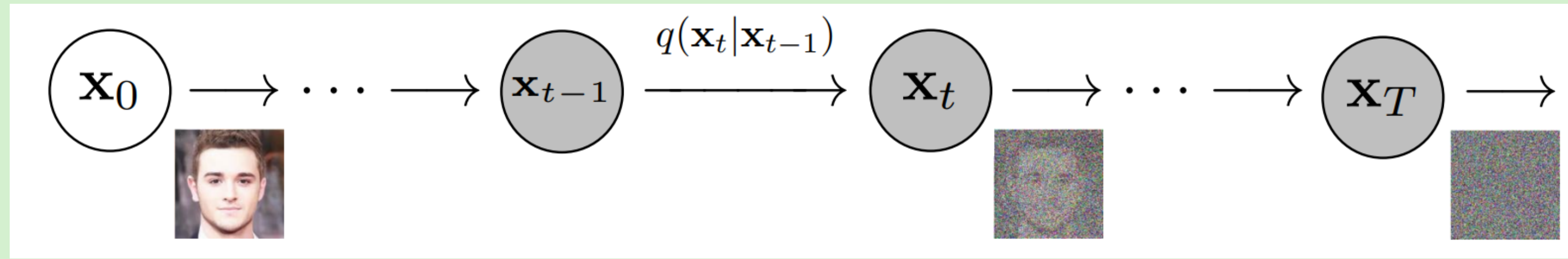


Experiment

Diffusion Policy [1]

Diffusion Model:

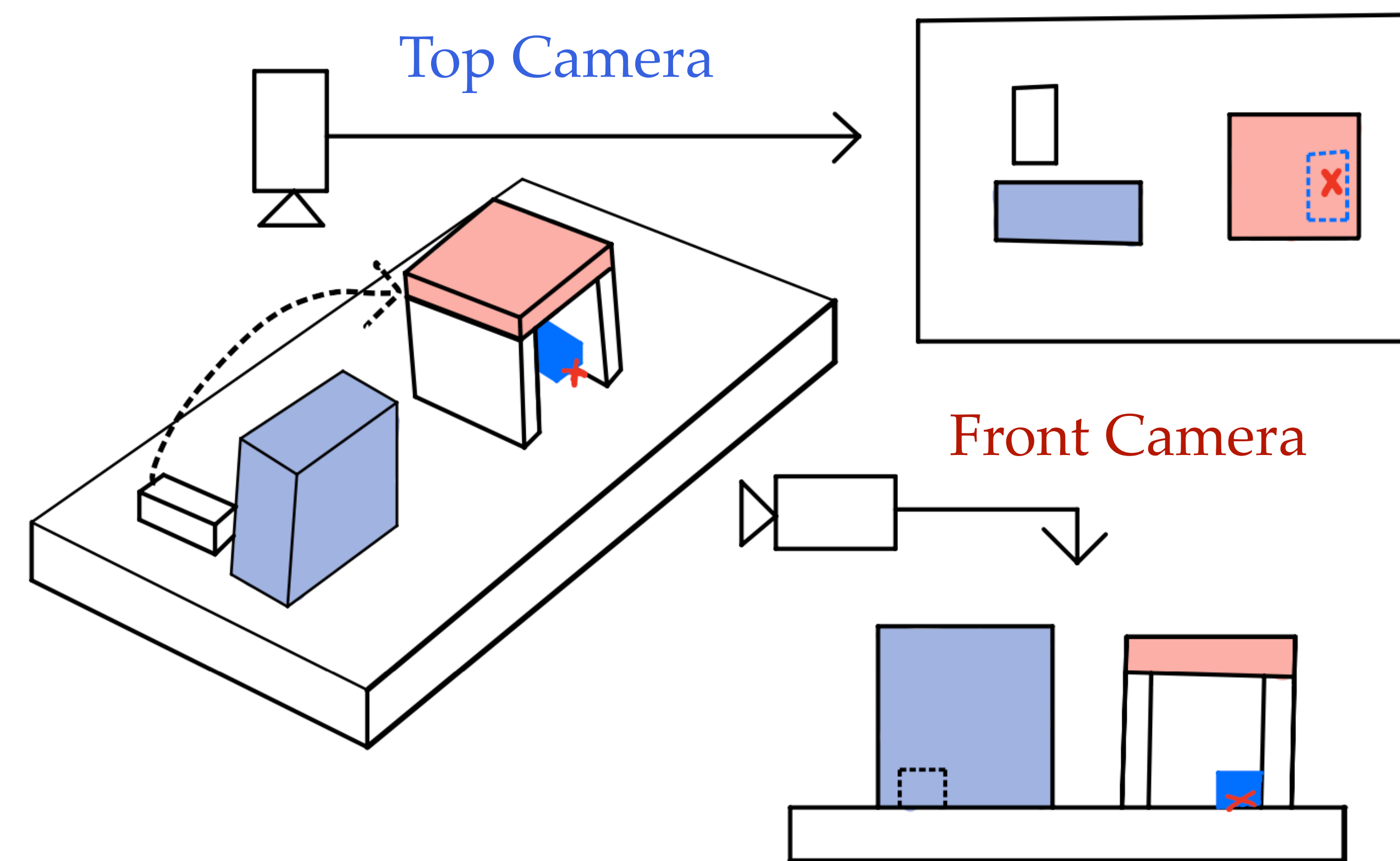


Diffusion Policy:

Diffusion model for visuomotor policy learning on robots

Pros: Multimodality, data efficiency

Task: pick and place object to target



Top camera can see pick, front camera can see place

We train 3 models conditioned on 1) **top camera feed** 2) **front camera feed** and 3) **both cameras**

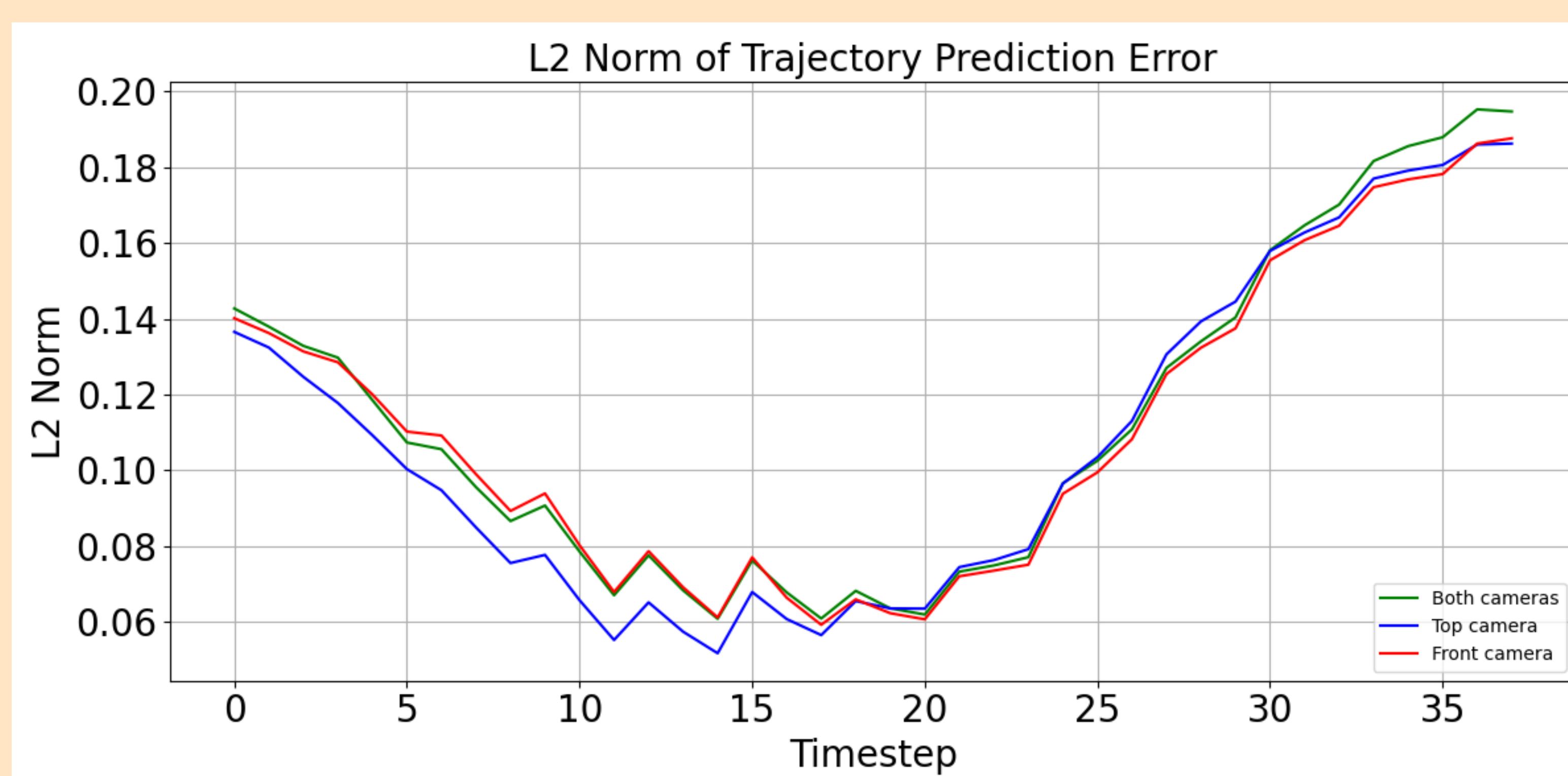
Evaluation

Quantitative Analysis

Q1) Did the models converge?



Q2) How much error across a trajectory?



Errors across different models were similar across trajectory: Inconclusive

Qualitative Analysis

Q1) Real World Evaluation

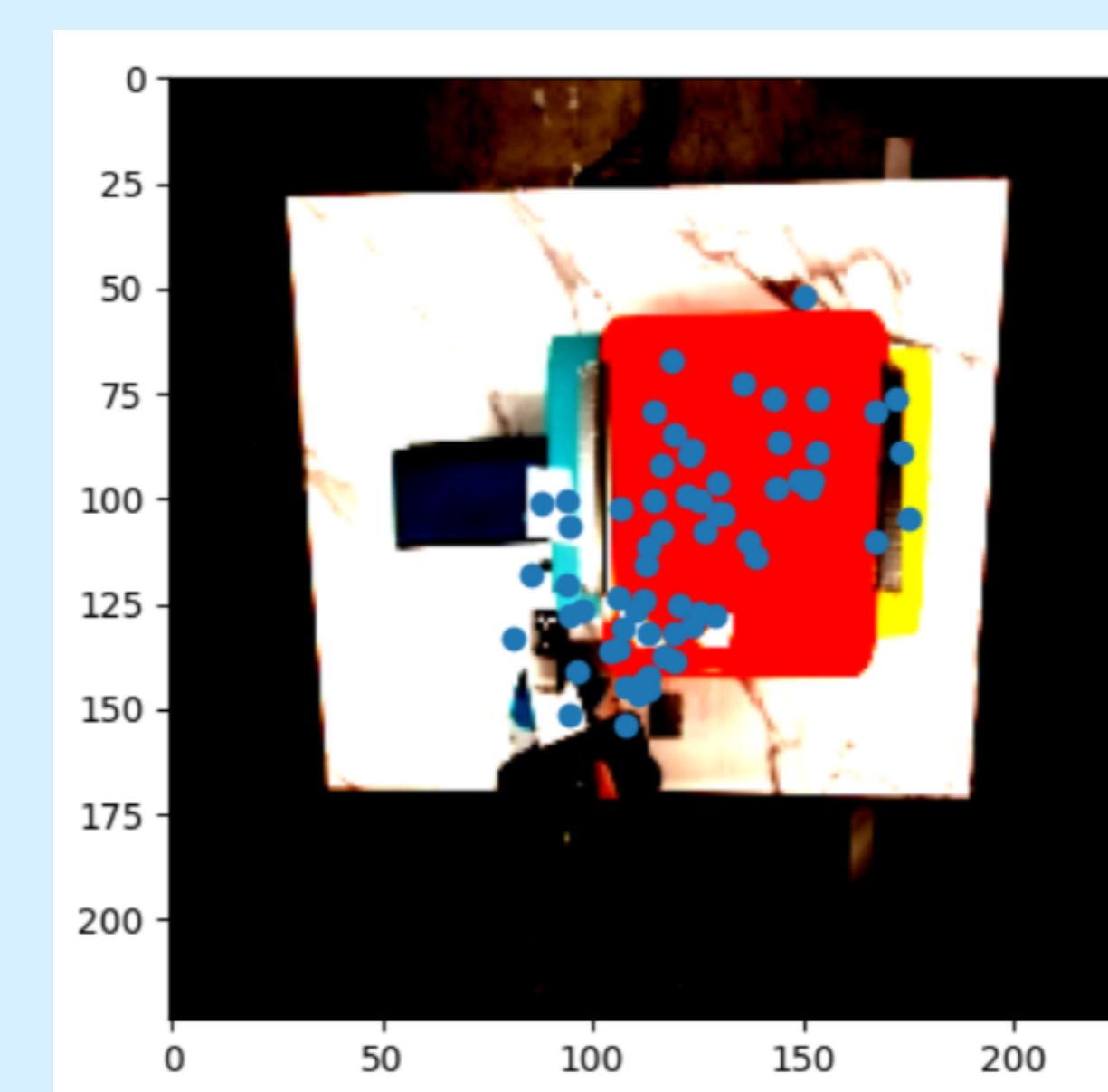
Top camera view:
pick location visible
Robot approaches the target

Front camera view:
pick location not visible
Robot does not move!

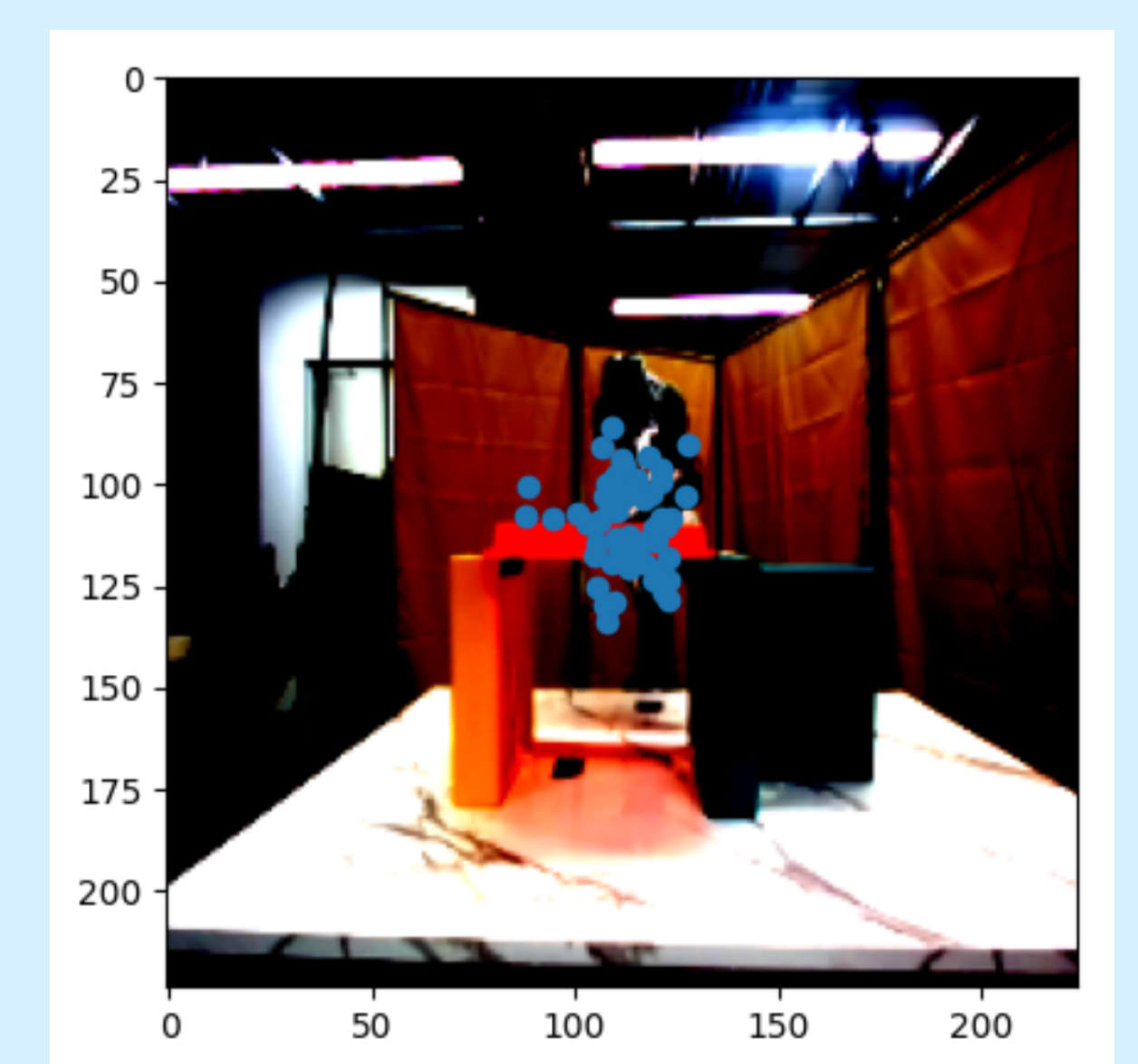


The two models show significant difference between success rates of task

Q2) Where does my model pay attention to? [2]



Top camera policy



Front camera policy

Future Work

- Address the issue of quantitative evaluation not matching qualitative evaluation
- Explore other modes of overcoming occlusion from camera frame such as image inpainting and 3D representation
- Test more complex tasks requiring multiple camera views due to more occlusions

