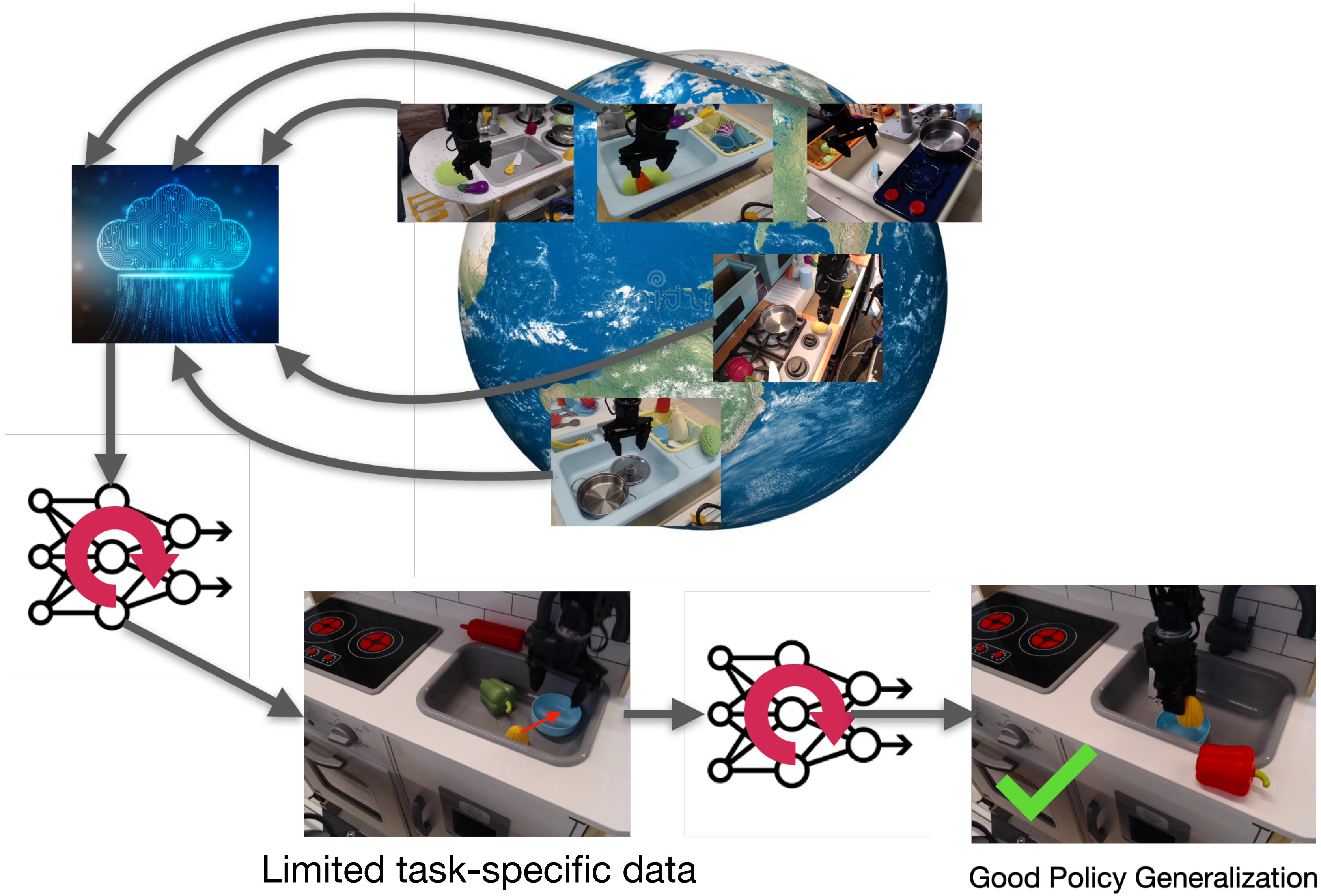
Pre-Training for Robots:

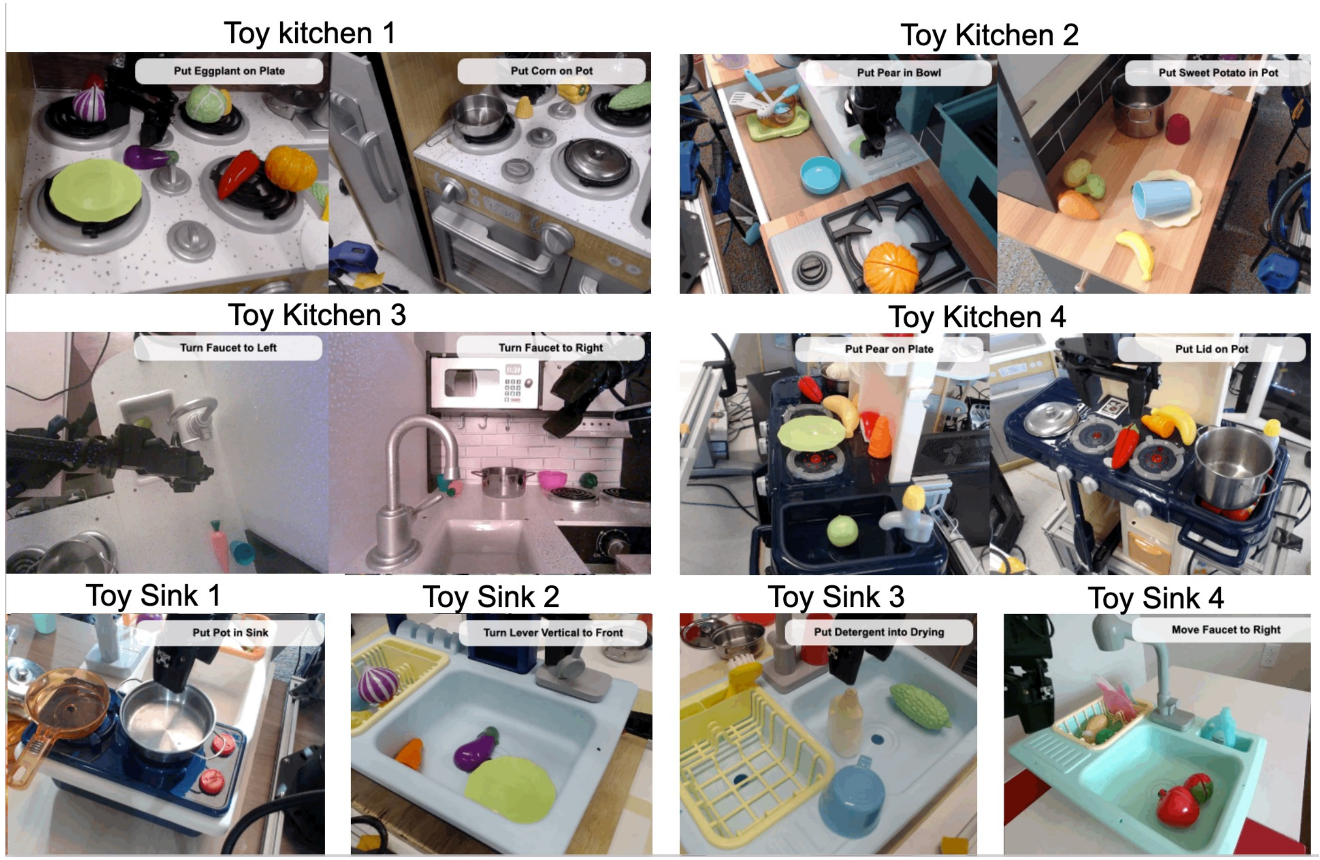
How Offline RL Enables Learning New Tasks from a Handful of Trials

Aviral Kumar*, Anikait Singh*, Frederik Ebert*,
Yanlai Yang, Chelsea Finn, Sergey Levine

Our Vision: Incorporate Large Robotic Datasets



How To Learn From Large Robot Datasets



Imitation learning

Ebert et al. 2021
Young et al. 2021
and many more

Pre-trained representation

Imitation

Nair
al. 2022
y more….

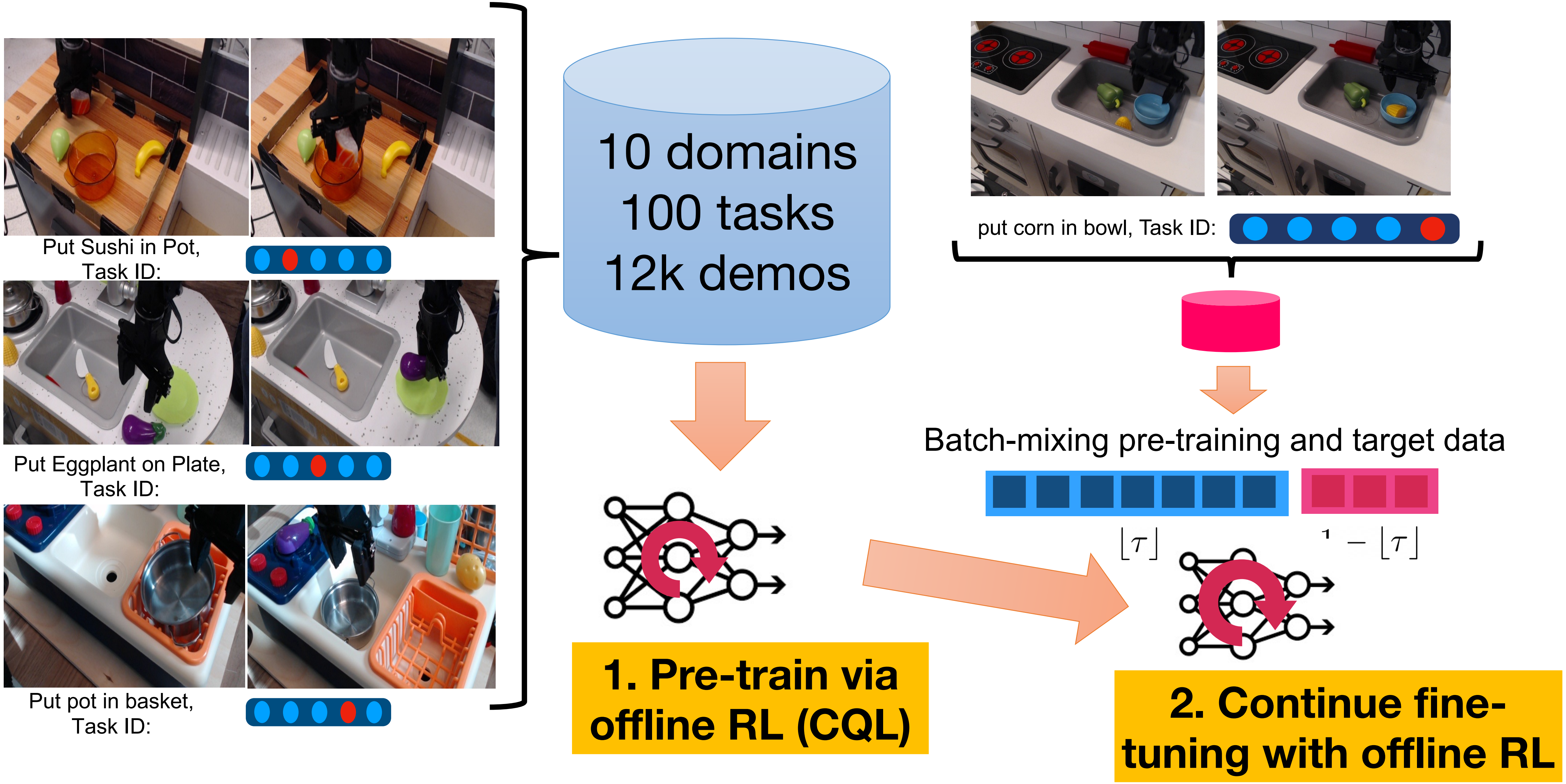
Can we instead use offline RL for <u>both</u> pre-training & fine-tuning?

Why?

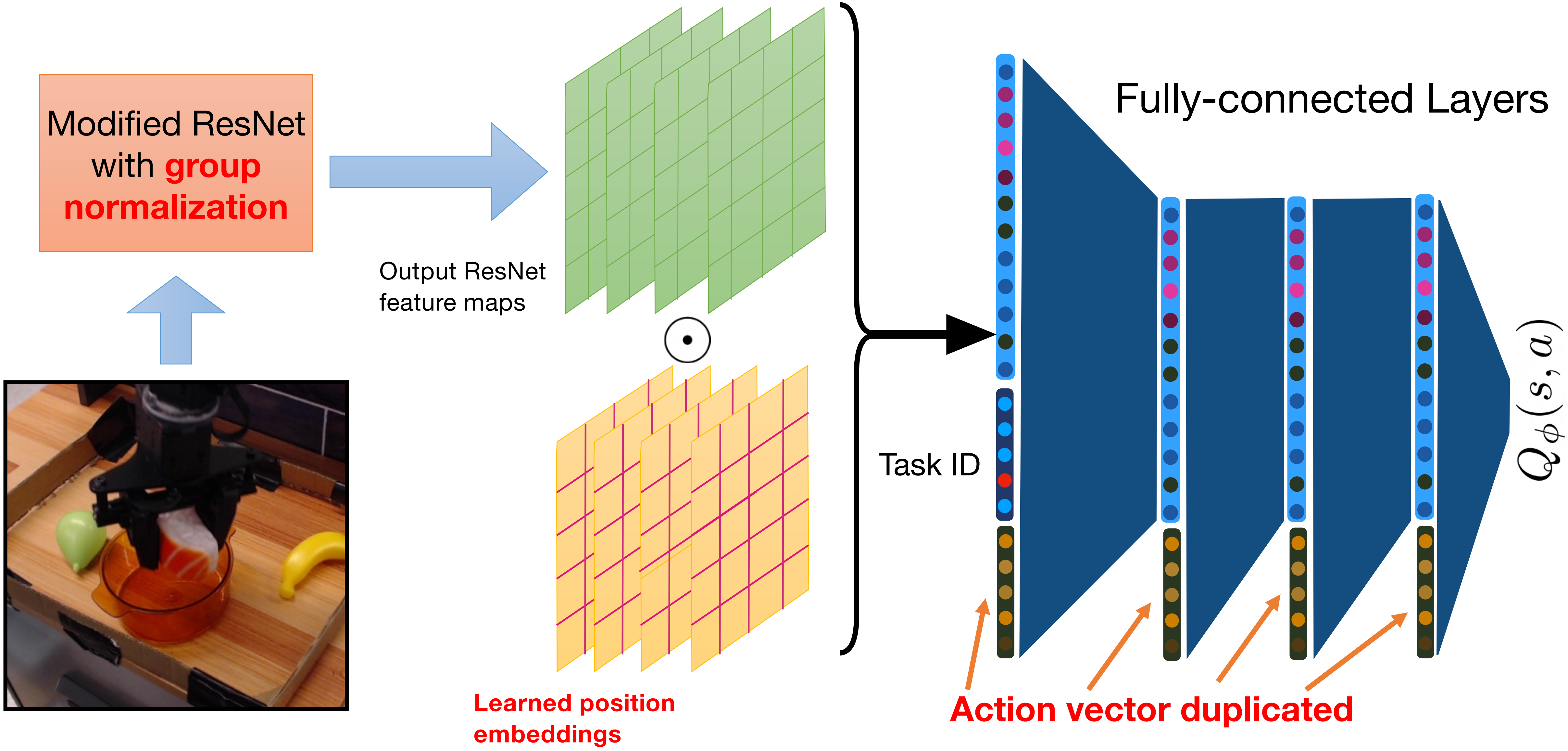
Pre-training on broad data (e.g., representation learning)

Adaptation on <u>limited,</u> task-specific data

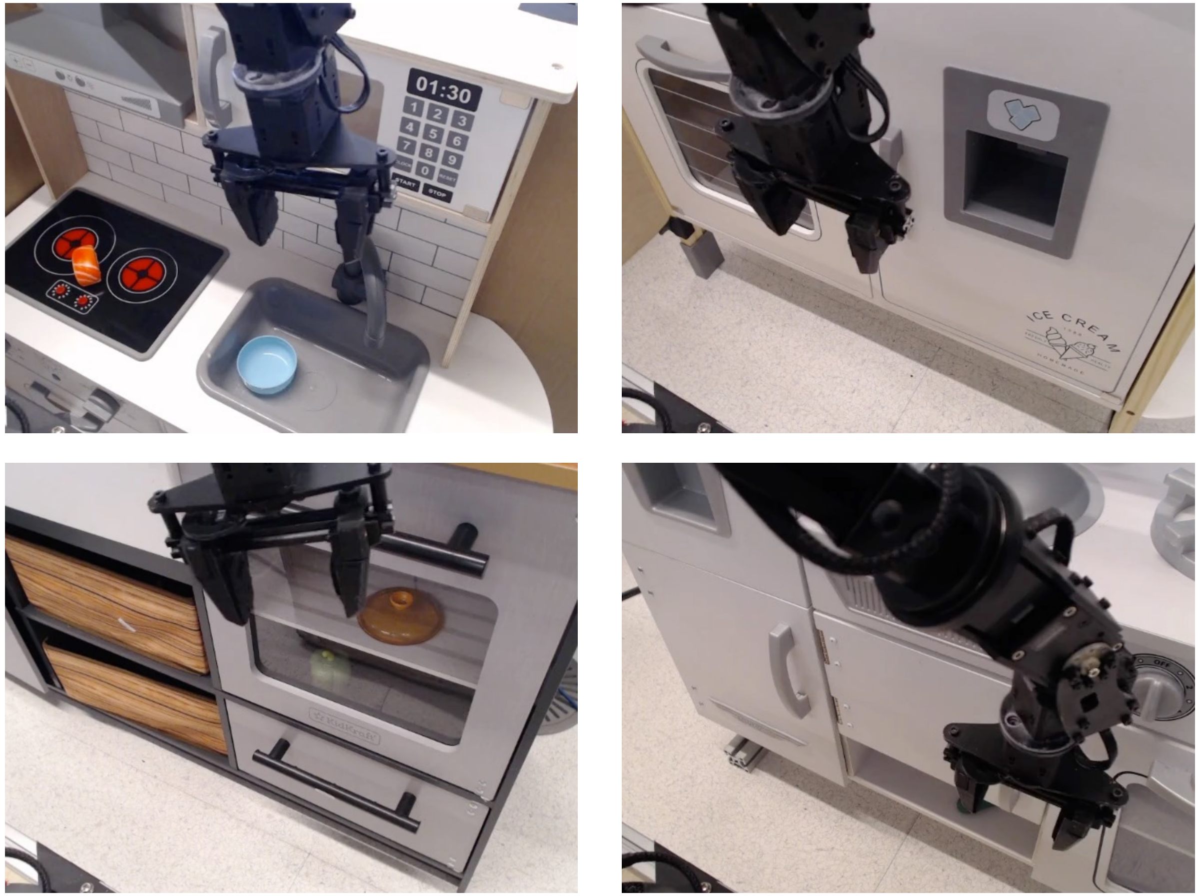
Pre-Training for Robots Using Offline RL



Main Innovation: Architecture



Task: Solving A Task in A New Domain



1. Pre-Train on Bridge Data, 12 doors 800 demonstrations

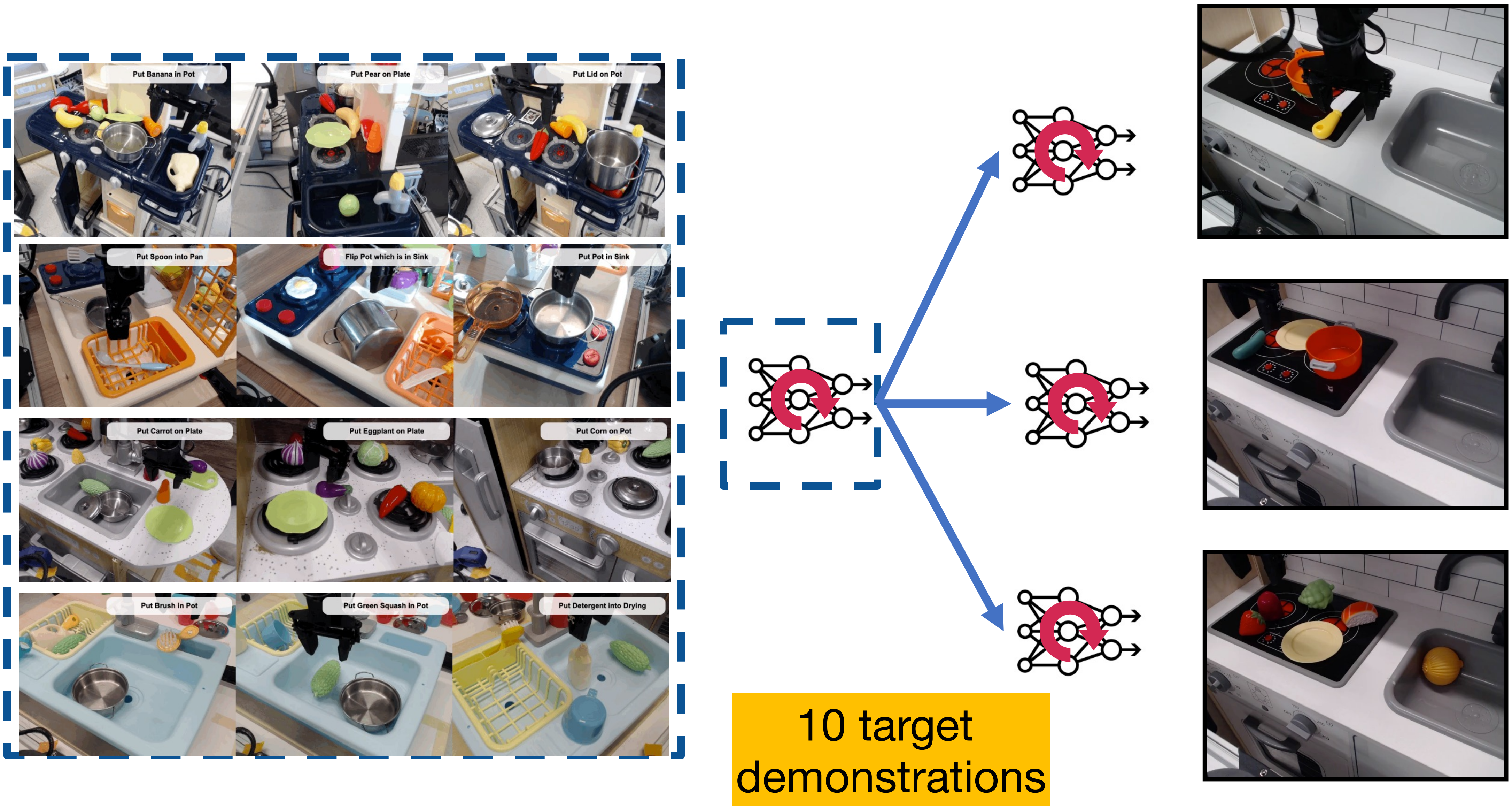
2. Fine-Tune on Target Domain Data:
1 door, 10 demonstrations

Results: Solving A Task in A New Domain

Method:
Imitation (Best prior method)

Method:
PTR (Ours)

Task: Solving New Tasks in New Domains



Results: Solving New Tasks in New Domains

Best Prior method

PTR (Ours)

Some Quantitative Results

Task	PTR (Ours)	BC finetuning: BC (fine.)	BC finetuning: Autoreg. BC	BC finetuning: BeT	Joint training: COG	Joint training: BC	Target data only: CQL	Target data only: BC	Pre-train. rep. + BC finetune: R3M	Pre-train. rep. + BC finetune: MAE
Take croissant from metal bowl	**7/10**	3/10	5/10	1/10	4/10	4/10	0/10	1/10	1/10	3/10
Put sweet potato on plate	**7/20**	1/20	1/20	0/20	0/20	0/20	0/20	0/20	0/20	1/20
Place knife in pot	**4/10**	2/10	2/10	0/10	1/10	3/10	3/10	0/10	0/10	0/10
Put cucumber in pot	**5/10**	0/10	1/10	0/10	2/10	1/10	0/10	0/10	0/10	0/10

Imitation (using transformers, auto-regressive)

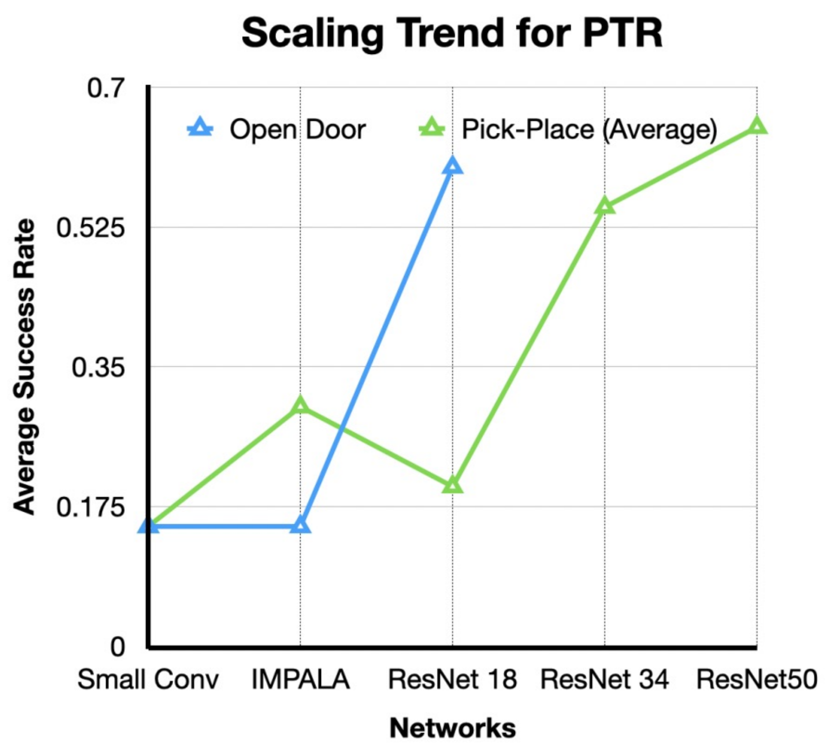
Better fine-tuning!

Self-supervised pre-training from internet data / bridge data

Representation learning

Takeaway: Offline RL learn useful representations + better fine-tuning

Scaling And Analysis

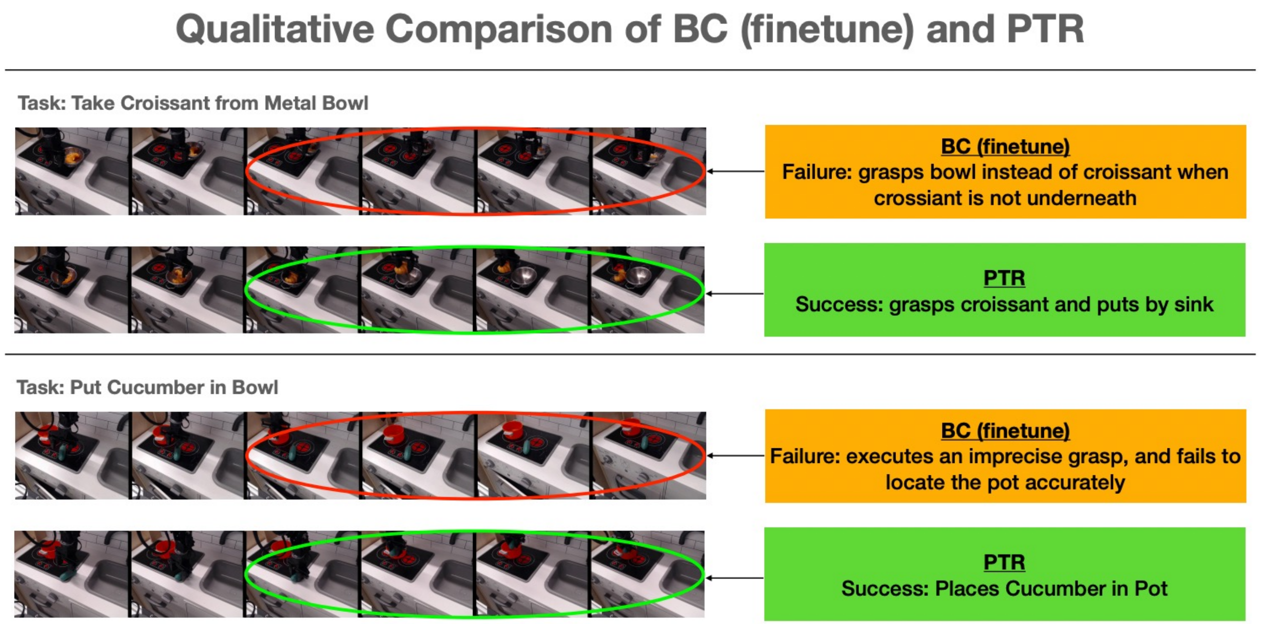


The larger the network, the better!

Why would RL enable better performance…

….when the data is collected via human teleoperation?

Spoiler: Value-functions can learn which decisions are more critical than others!



Takeaways and Future Work

- Offline RL can be good for both **representation learning** and **control, even with human demonstration data**

- Future Directions:
  - **Goal specification:** language? goals? reward learning?
  - **Multi-modal data:** videos?

Paper & Code:

Thank You!