

# Pre-Training for Robots:

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

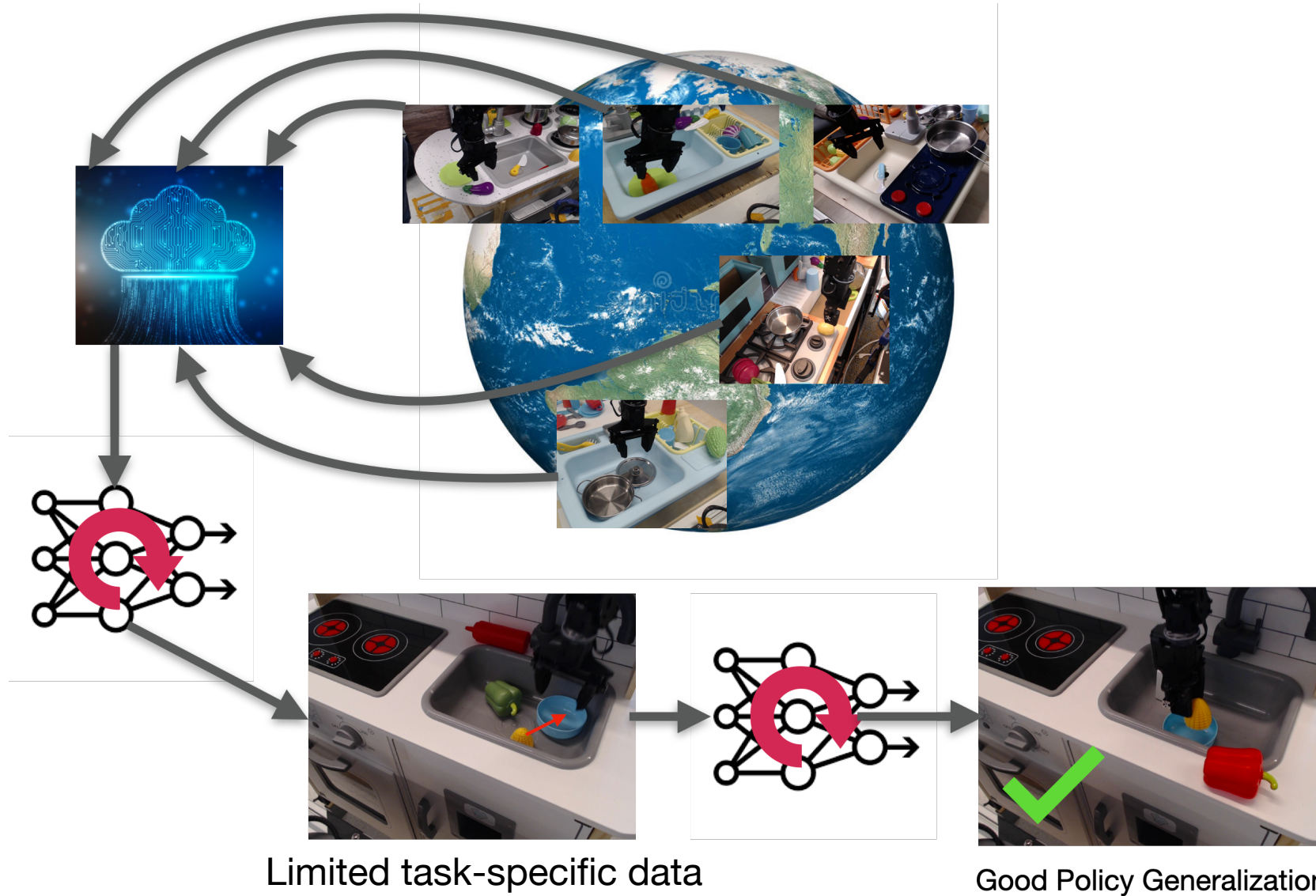
**Aviral Kumar\***, Anikait Singh\*, Frederik Ebert\*,  
Yanlai Yang, Chelsea Finn, Sergey Levine



Berkeley  
UNIVERSITY OF CALIFORNIA



# Our Vision: Incorporate Large Robotic Datasets



# How To Learn From Large Robot Datasets



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



Adaptation on  
limited, **task-**  
**specific data**

Imitation learning

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

Pre-trained representation



Imitation

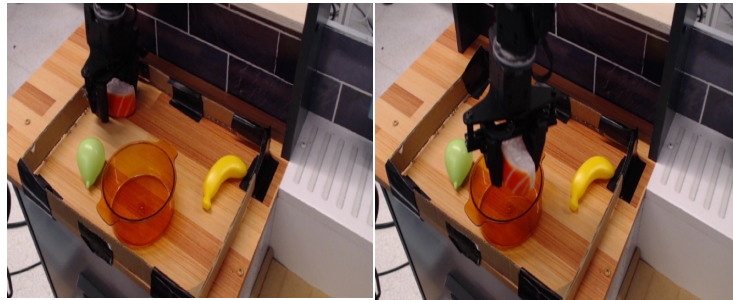
Nair et al. 2022  
and many more....

**Can we instead use offline RL for  
both pre-training & fine-tuning?**

**Why?**



# Pre-Training for Robots Using Offline RL



Put Sushi in Pot,  
Task ID:



Put Eggplant on Plate,  
Task ID:



Put pot in basket,  
Task ID:



10 domains  
100 tasks  
12k demos



**1. Pre-train via  
offline RL (CQL)**



put corn in bowl, Task ID:



Batch-mixing pre-training and target data

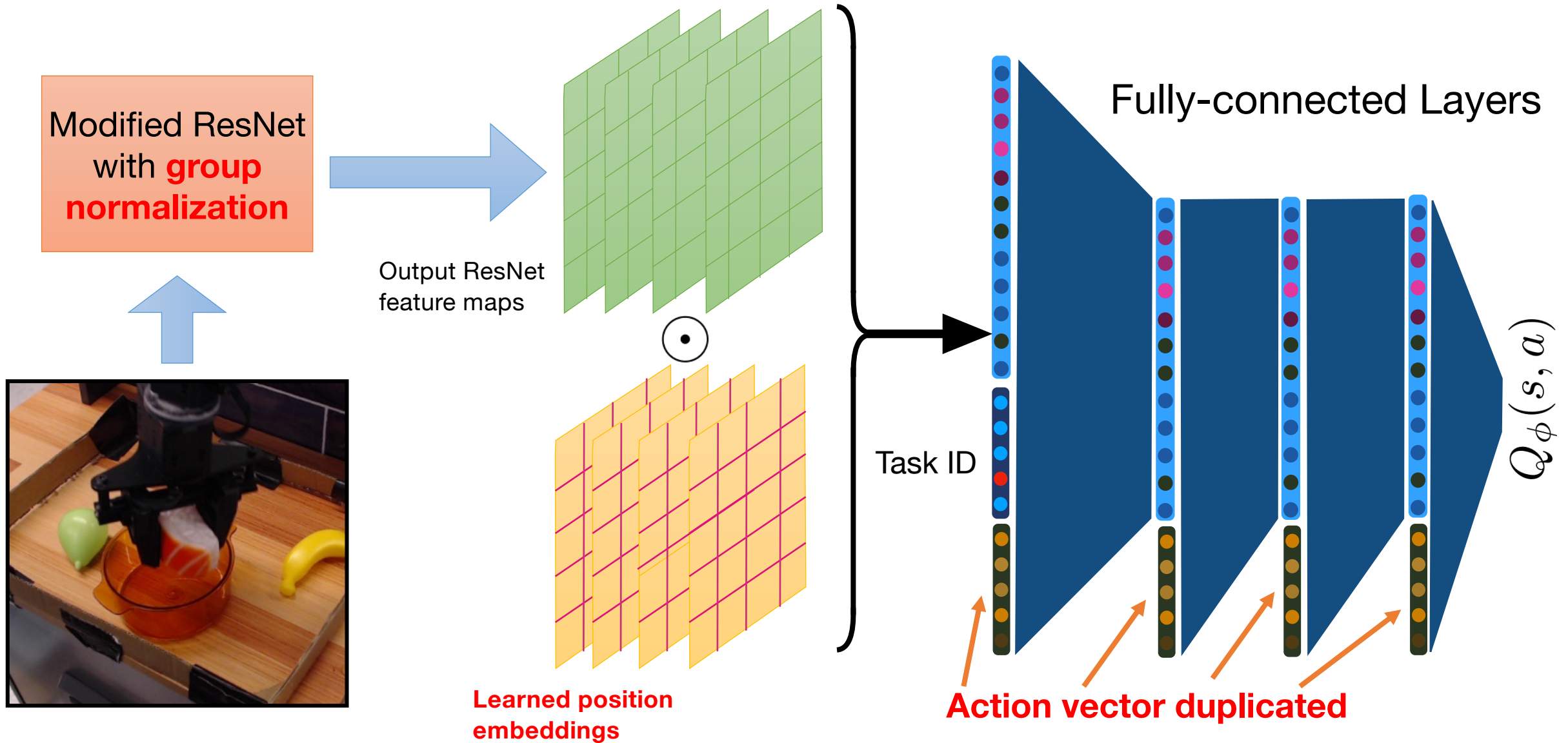


$\lceil \tau \rceil$   $1 - \lceil \tau \rceil$

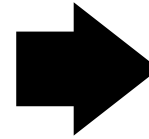


**2. Continue fine-tuning  
with offline RL**

# Main Innovation: Architecture



# Task: Solving A Task in A New Domain



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

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



# Results: Solving A Task in A New Domain

## Method:

Imitation (Best prior method)

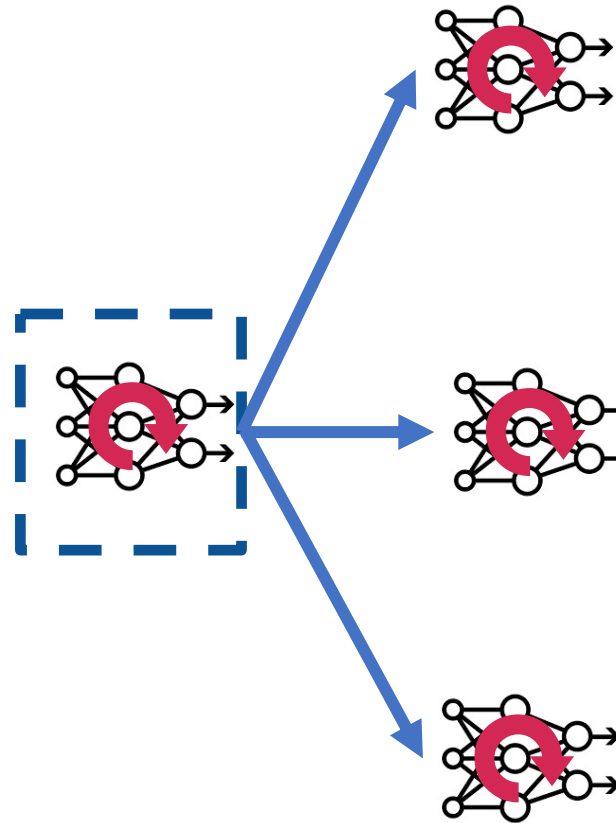


## Method:

**PTR (Ours)**



# Task: Solving **New** Tasks in New Domains



10 target  
demonstrations





# Results: Solving **New** Tasks in New Domains

**Best Prior  
method**



**PTR (Ours)**



# Some Quantitative Results

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

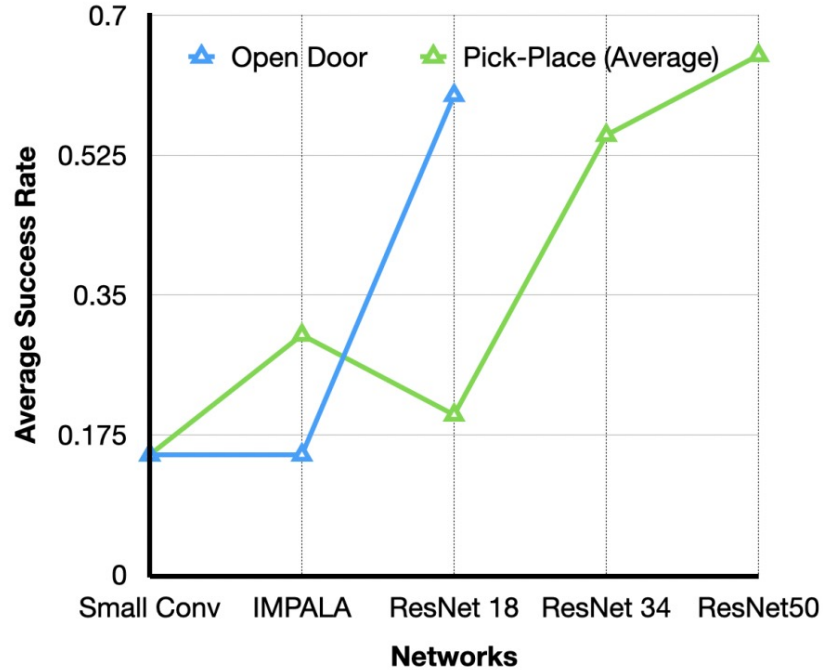


Representation learning

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

# Scaling And Analysis

Scaling Trend for PTR



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!

Qualitative Comparison of BC (finetune) and PTR

Task: Take Croissant from Metal Bowl



**BC (finetune)**  
Failure: grasps bowl instead of croissant when  
croissant is not underneath



**PTR**  
Success: grasps croissant and puts by sink

Task: Put Cucumber in Bowl



**BC (finetune)**  
Failure: executes an imprecise grasp, and fails to  
locate the pot accurately



**PTR**  
Success: Places Cucumber in Pot



# 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!