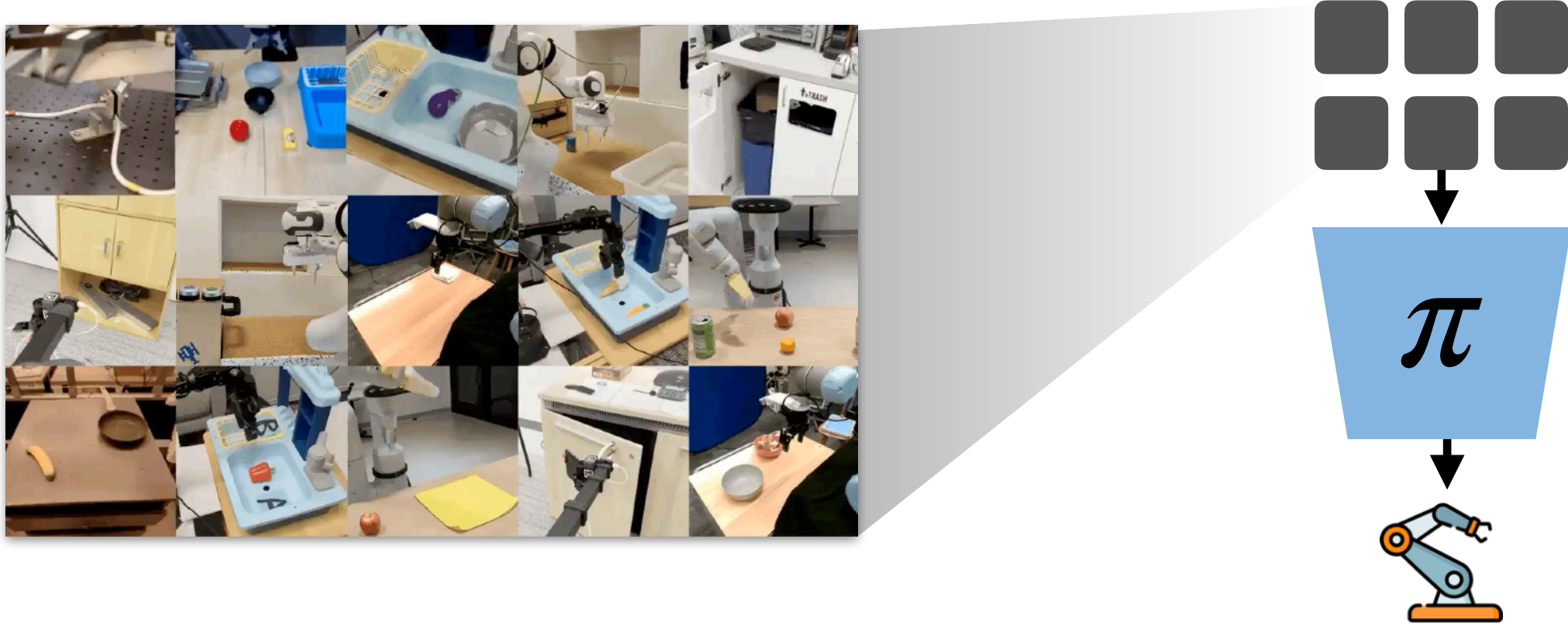# DataMIL🌾 : Selecting Data for Robot Imitation Learning using Datamodels

Shivin Dass*, Alaa Khaddaj*, Logan Engstrom, Aleksander Madry, Andrew Ilyas[†], Roberto Martín-Martín[†]

1. Motivation

Robotics is seeing a rise in the **size** and **diversity** of robot datasets

How can we **select data** from these datasets for training **task-specific policies**?

2. Problem Formulation

Given a dataset $\mathcal{D}$ and a learning algorithm $\mathcal{A}(.)$, we can learn a policy as,

$$\pi = \mathcal{A}(\mathcal{D})$$

The objective of data selection is to find the subset of data that maximizes performance of a policy trained on it on a target metric $\mathcal{M}$,

$$\underset{\mathcal{D}' \subset \mathcal{D}}{\mathrm{argmax}}\, \mathcal{M}(\mathcal{A}(\mathcal{D}')).$$

$\mathcal{M}$, for example, can be policy rollouts

How can we select data from large datasets to boost task performance?

3. Datamodels and Overview
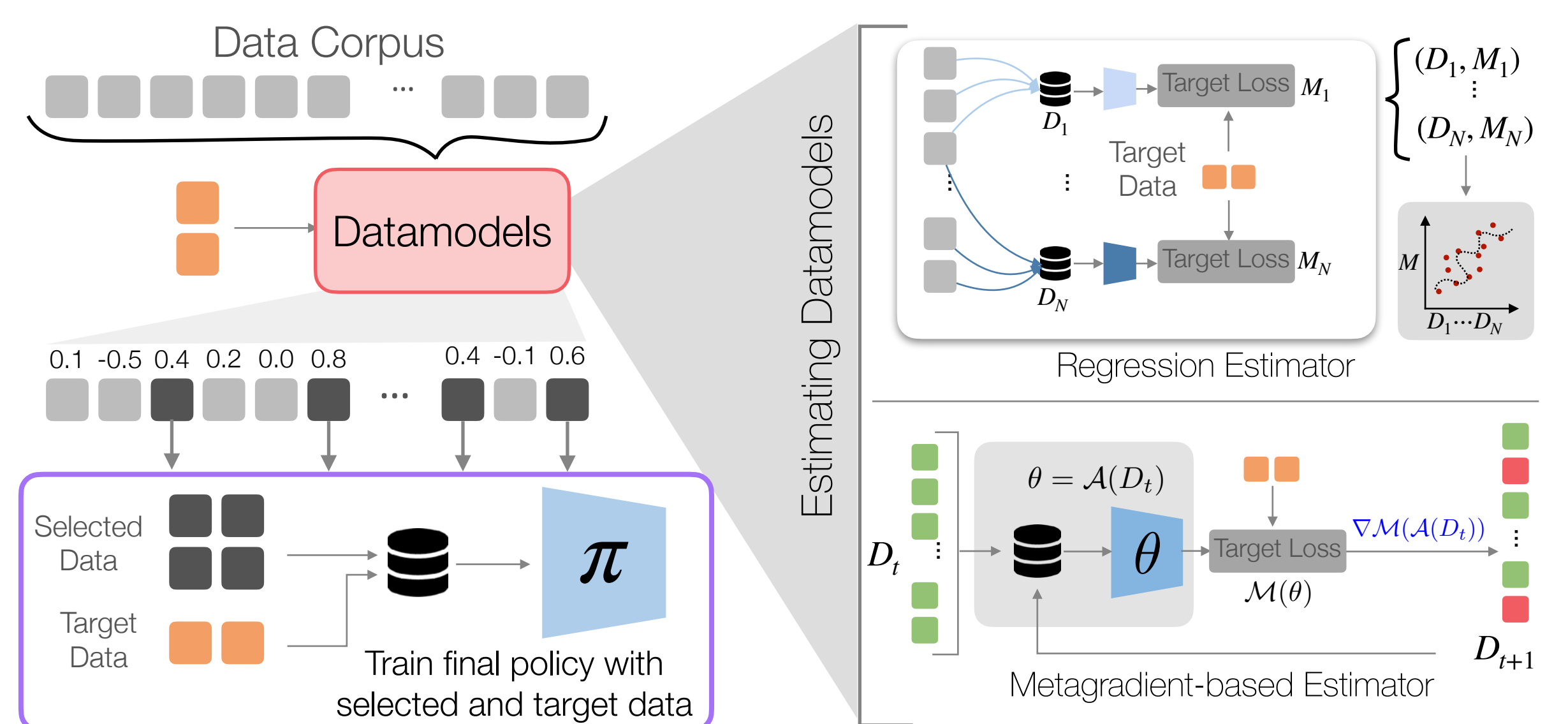
What are datamodels?

Datamodels aim to **cheaply predict** the behavior of a model if it were trained on a given subset of data.

$$\hat{f}(\mathcal{D}') \approx \mathcal{M}(\mathcal{A}(D')), \quad \forall \mathcal{D}' \subset \mathcal{D}$$

We explore methods that estimate the model outcome **linearly**,

$$\hat{f}(\mathcal{D}') = \sum_{z_i \in \mathcal{D}'} \tau(z_i)$$

More useful data points have a higher $\tau(z_i)$.

Data Corpus

Datamodels

0.1 -0.5 0.4 0.2 0.0 0.8 ... 0.4 -0.1 0.6

Selected Data / Target Data — Train final policy with selected and target data

Estimating Datamodels

(D_1, M_1) ... (D_N, M_N)

Target Data — Target Loss M_1 ... Target Loss M_N

Regression Estimator

$\theta = \mathcal{A}(D_t)$

D_t — θ — Target Loss $\mathcal{M}(\theta)$ — $\nabla \mathcal{M}(\mathcal{A}(D_t))$ — D_{t+1}

Metagradient-based Estimator

4. Adapting Datamodels to Robotics
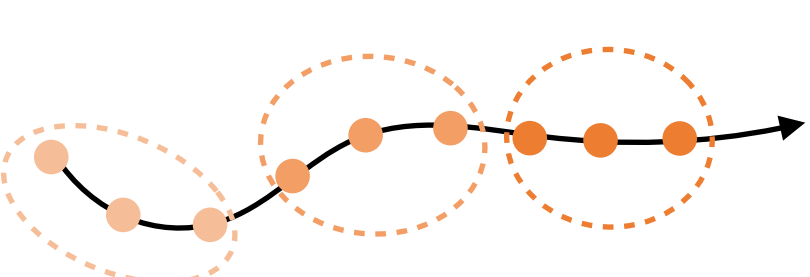
1. BC loss over a few target demos as a proxy for real world rollouts

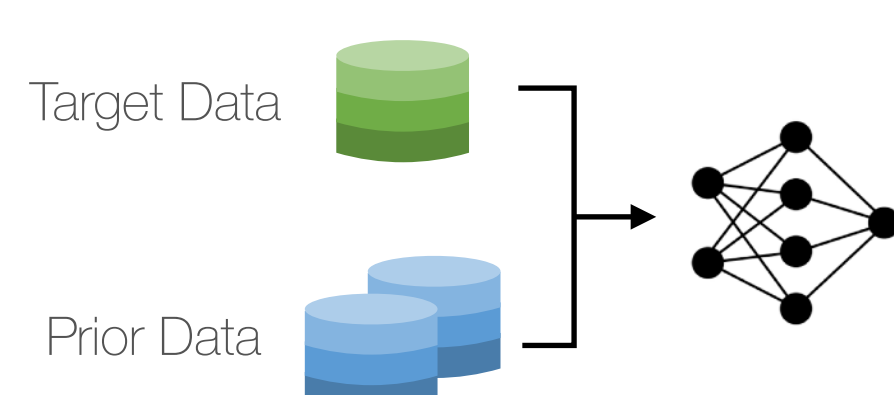$$\widehat{M} = -\mathbb{E}_{D_{target}}[L_{BC}(\pi(s), a)]$$

Empirically we show that the proxy objective can effectively stand in for expensive rollouts
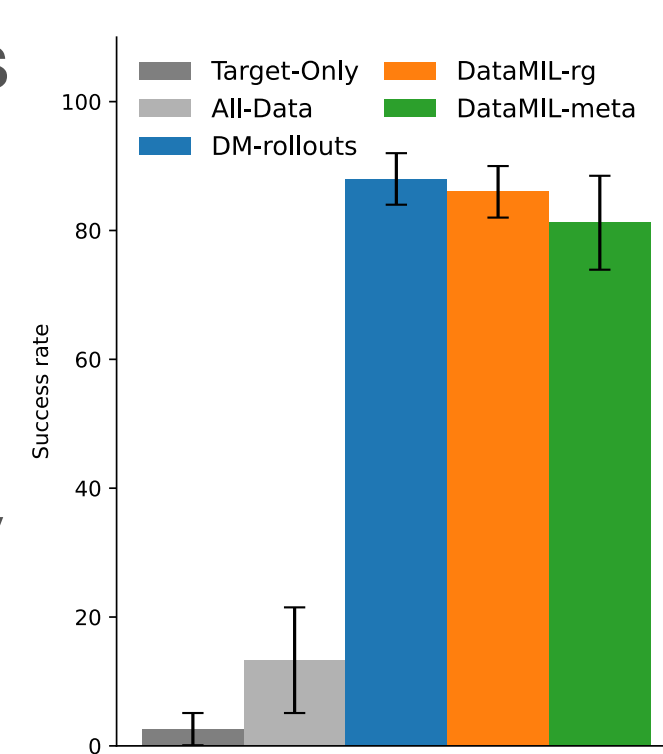
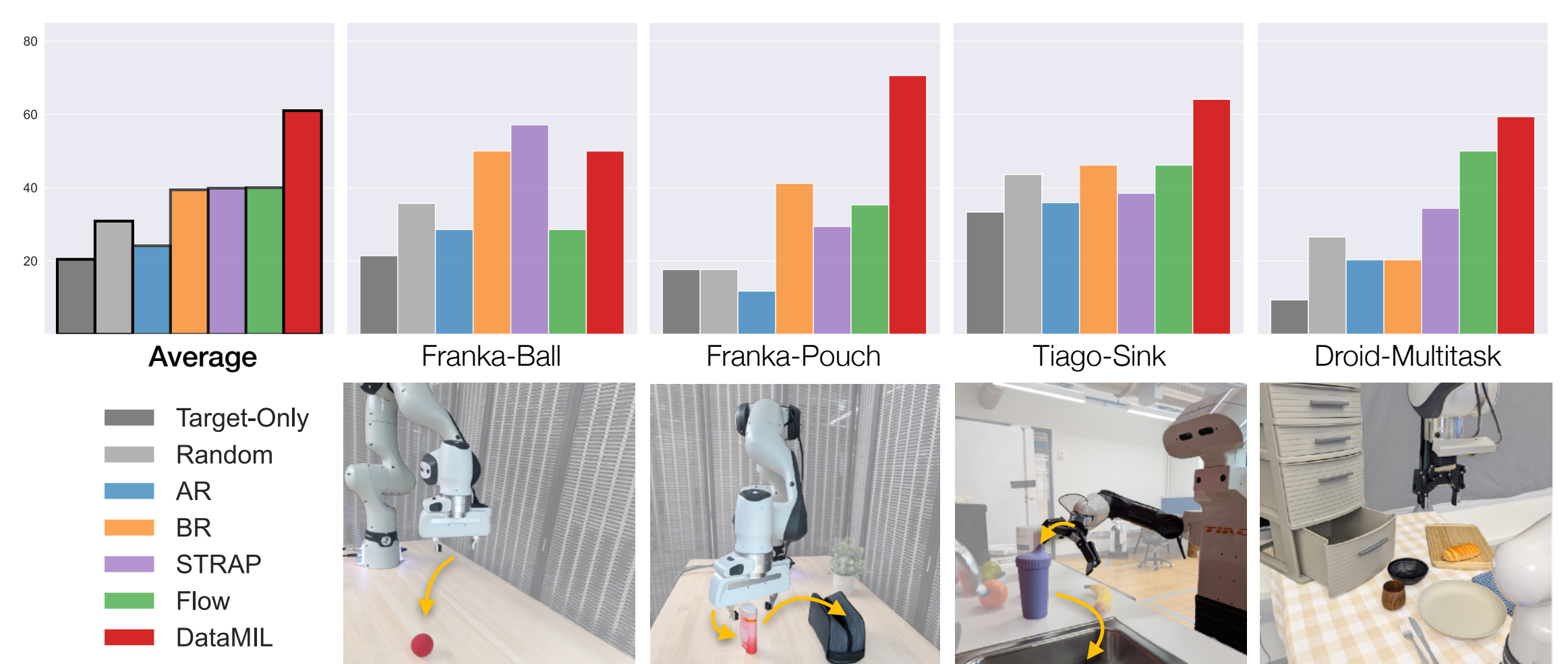2. Cluster data points to reduces datamodel estimation noise

3. Co-training with target data to minimize distribution shift

Target Data
Prior Data

5. Evaluation

We perform data selection on the **OXE dataset** and show successful selection on new **tasks** and **embodiments**

Average | Franka-Ball | Franka-Pouch | Tiago-Sink | Droid-Multitask

Target-Only
Random
AR
BR
STRAP
Flow
DataMIL

Interested to learn more?

Checkout our website   OR   Reach out to:
sdass@utexas.edu