

SORNet: Spatial Object-Centric Representations for Sequential Manipulation

Wentao Yuan¹ Chris Paxton²
Karthik Desingh¹ Dieter Fox^{1,2}

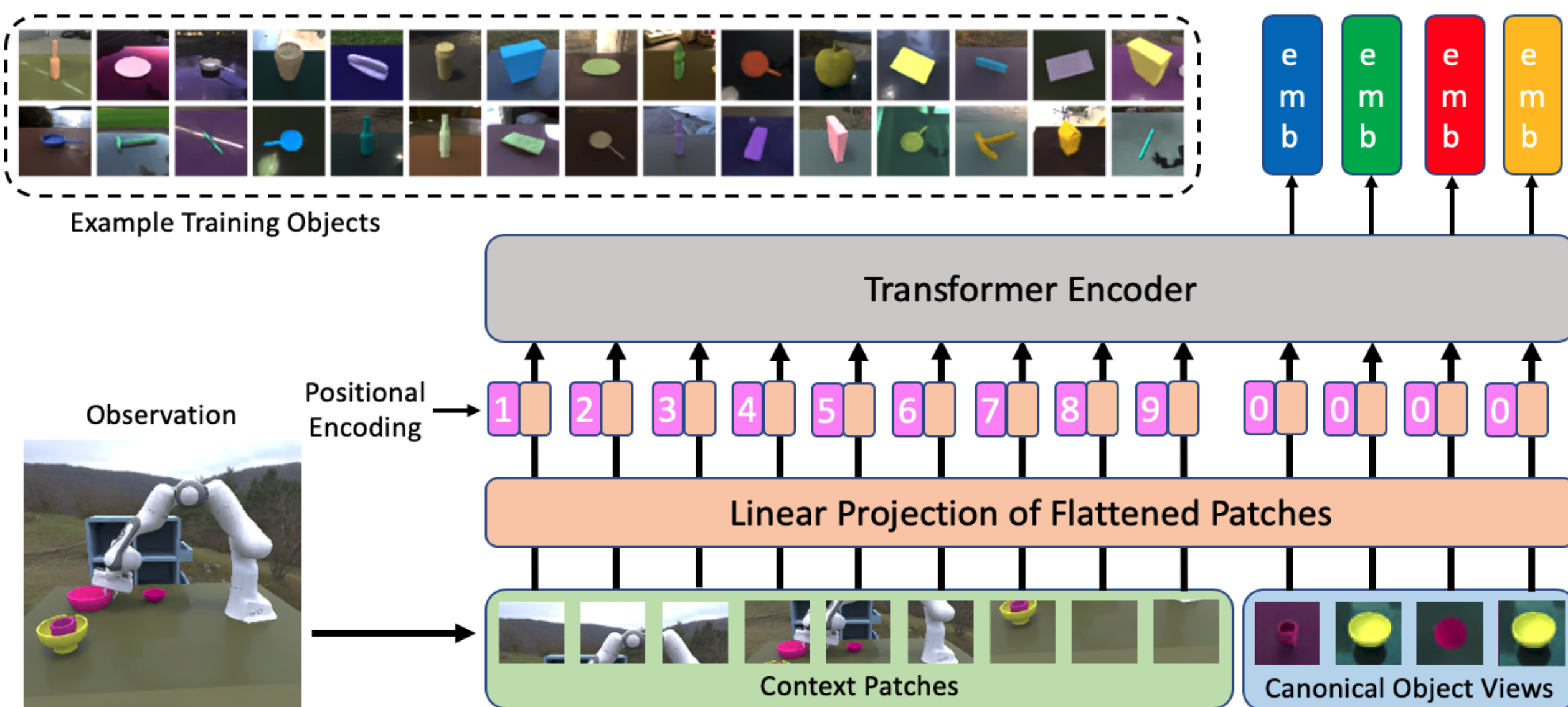
¹University of Washington ²NVIDIA

Contribution

We propose **SORNet**: **S**patial **O**bject-centric **R**epresentation **N**etwork to learn object-centric embeddings that encode spatial relations and enable **zero-shot** classification and regression of spatial relations among unseen objects.

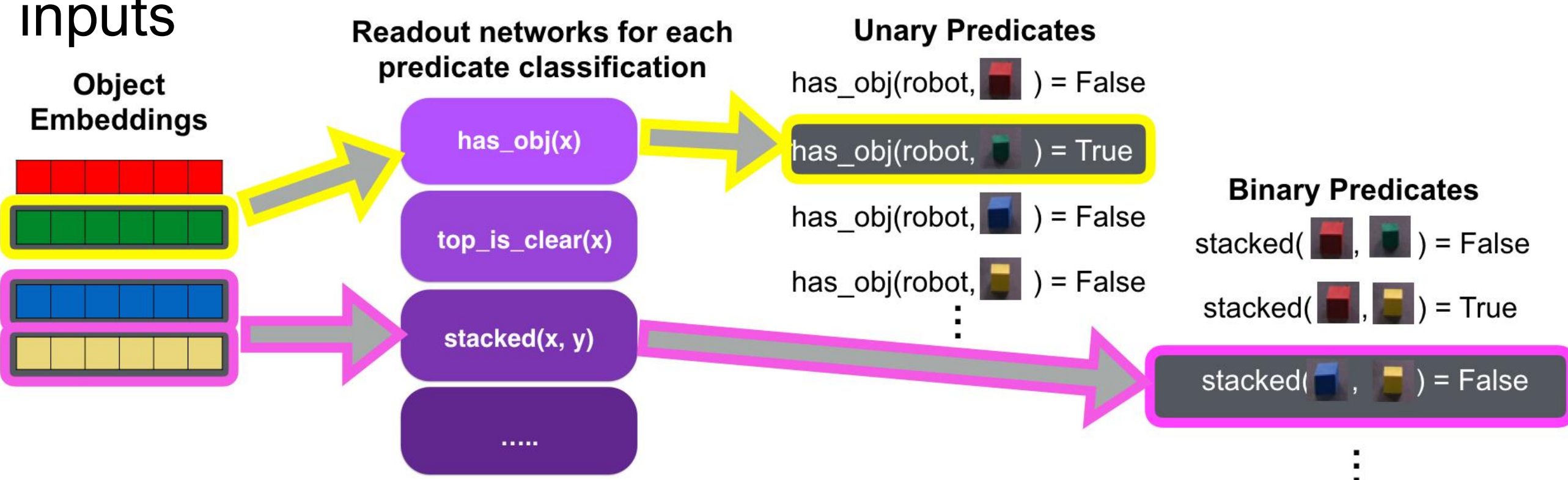
Embedding network

Permutation-invariant object embedding from the network

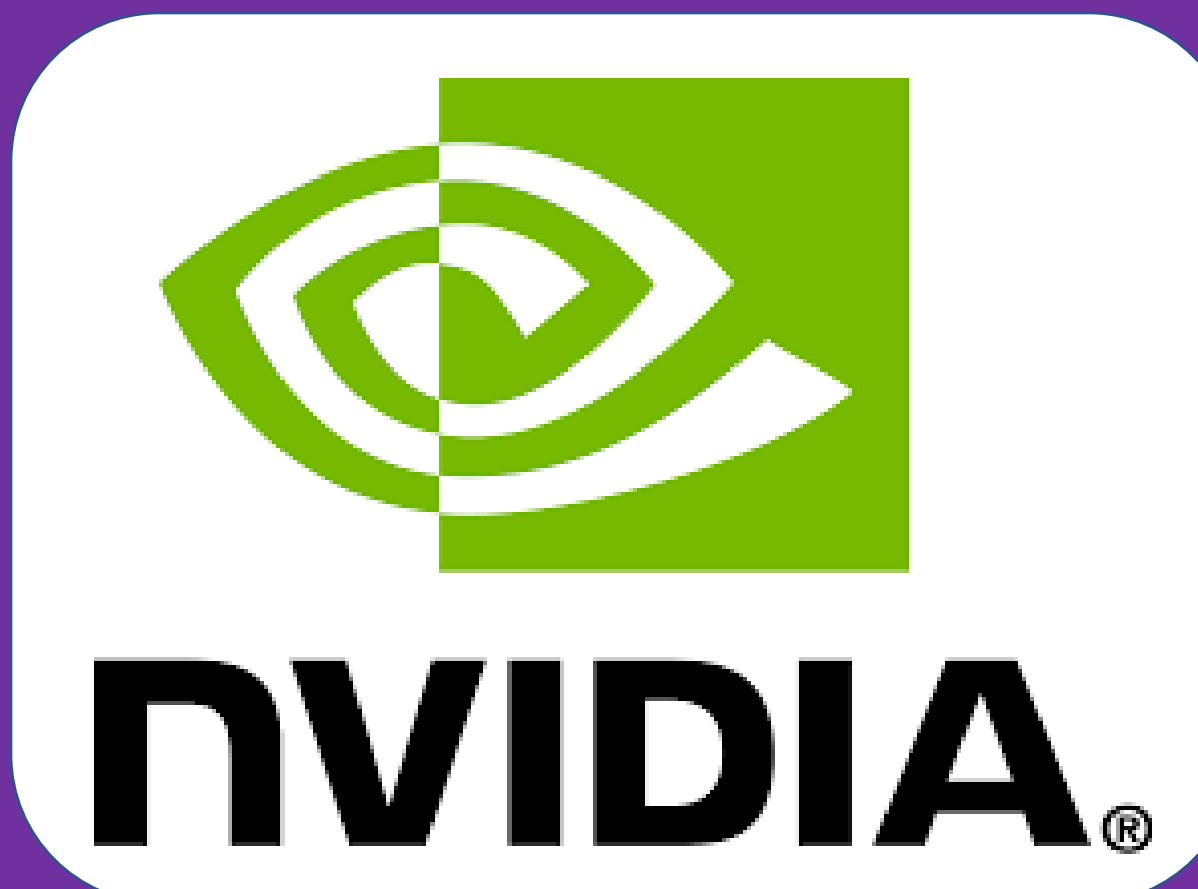
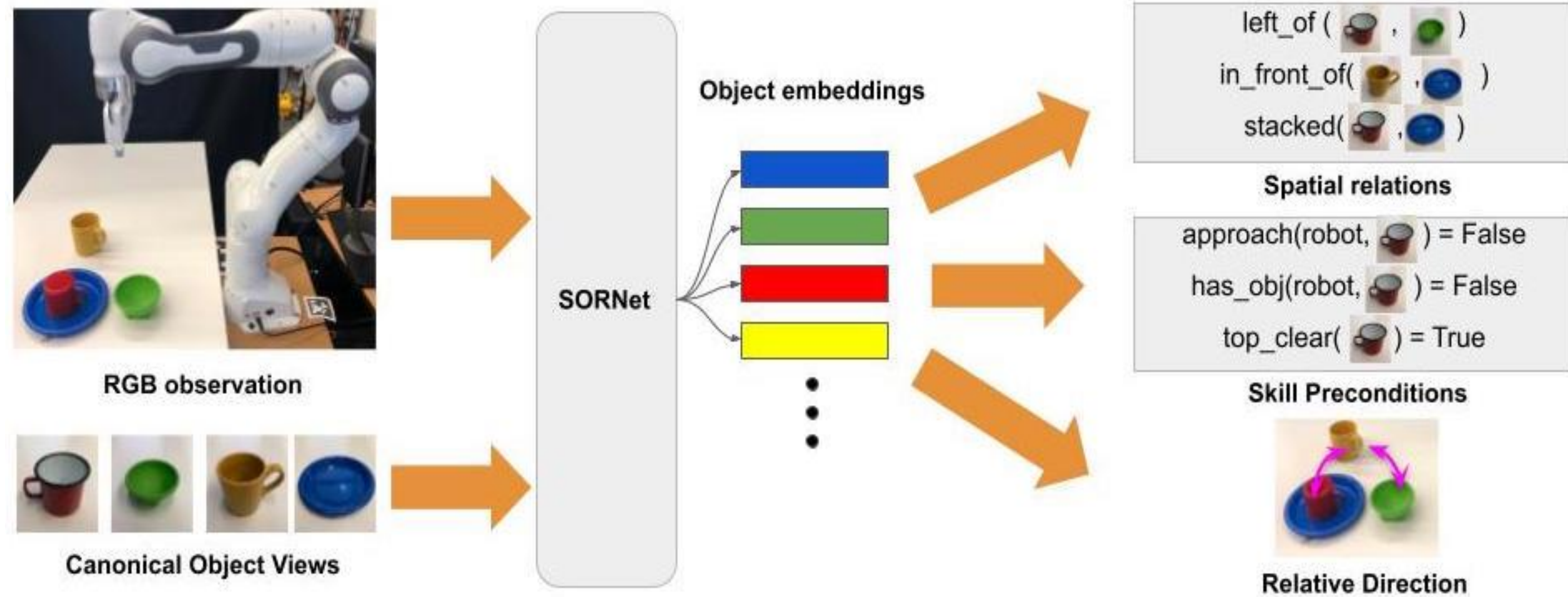


Readout networks

Number of outputs changes adaptively with number of inputs

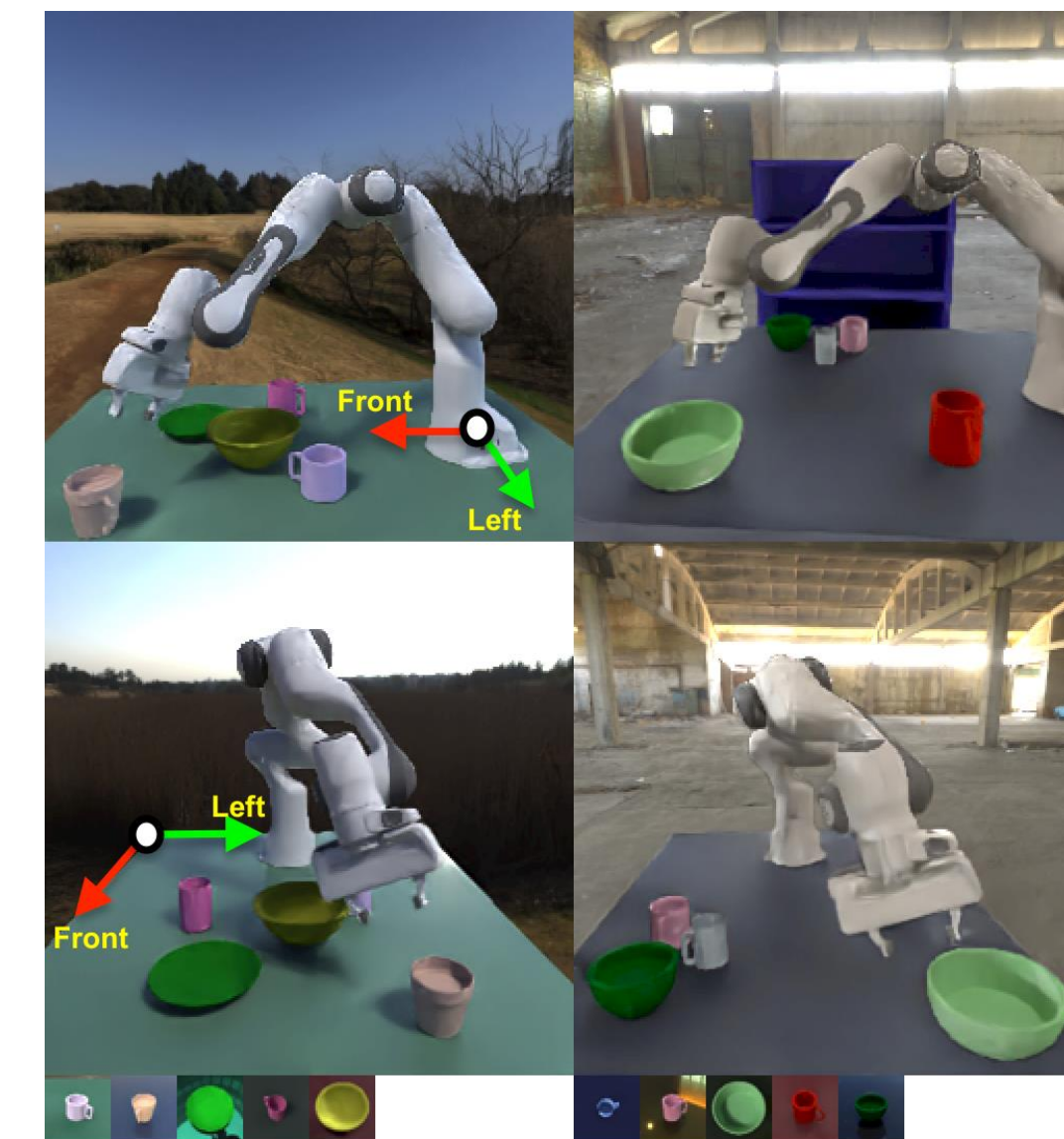


Can we learn implicit object-centric representations for spatial-reasoning and manipulation?

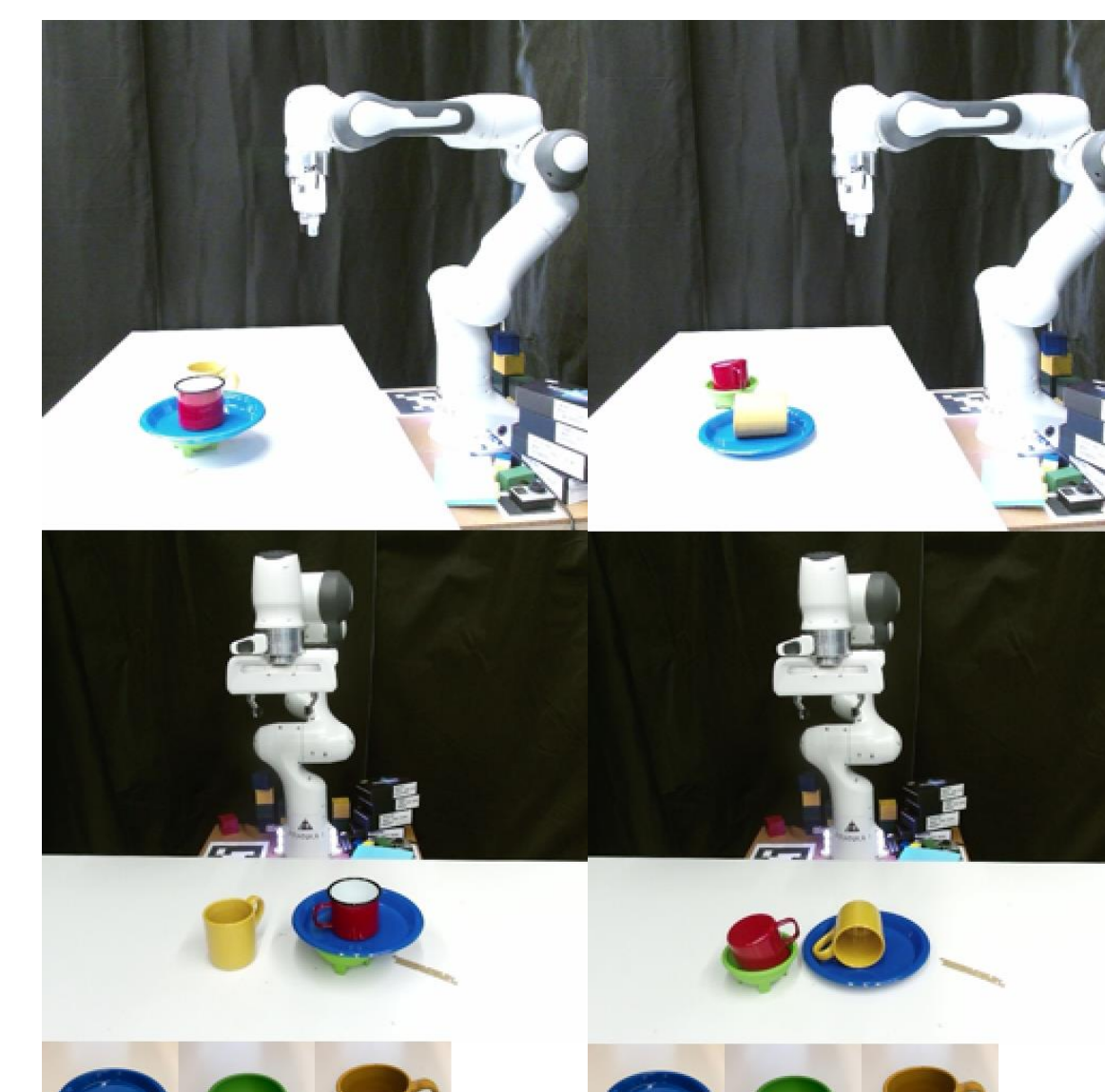


Spatial predicate classification

Simulation results



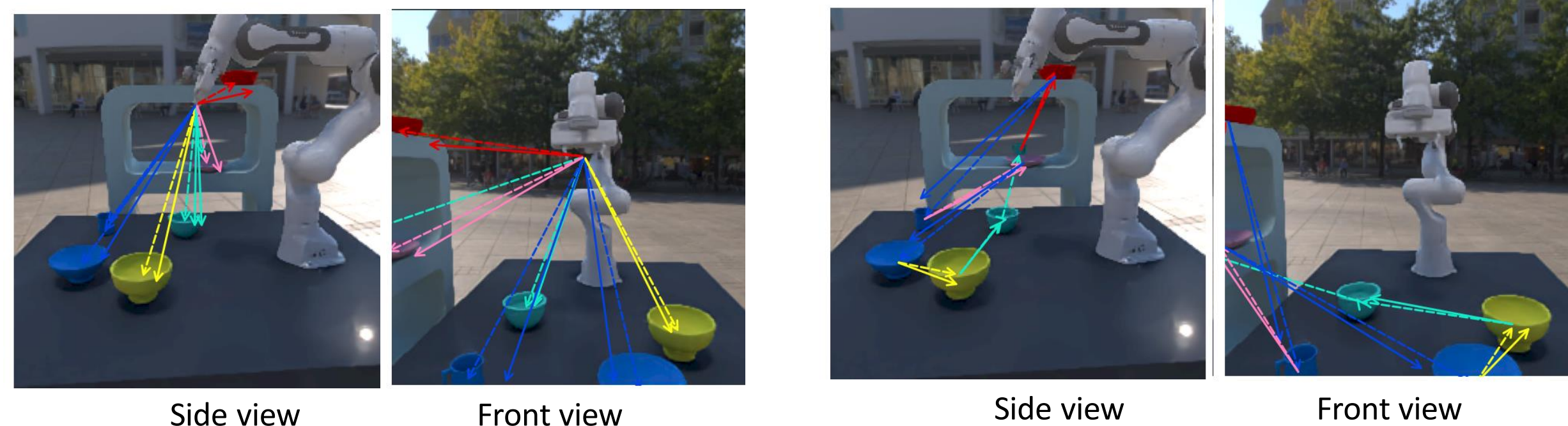
Real-world results



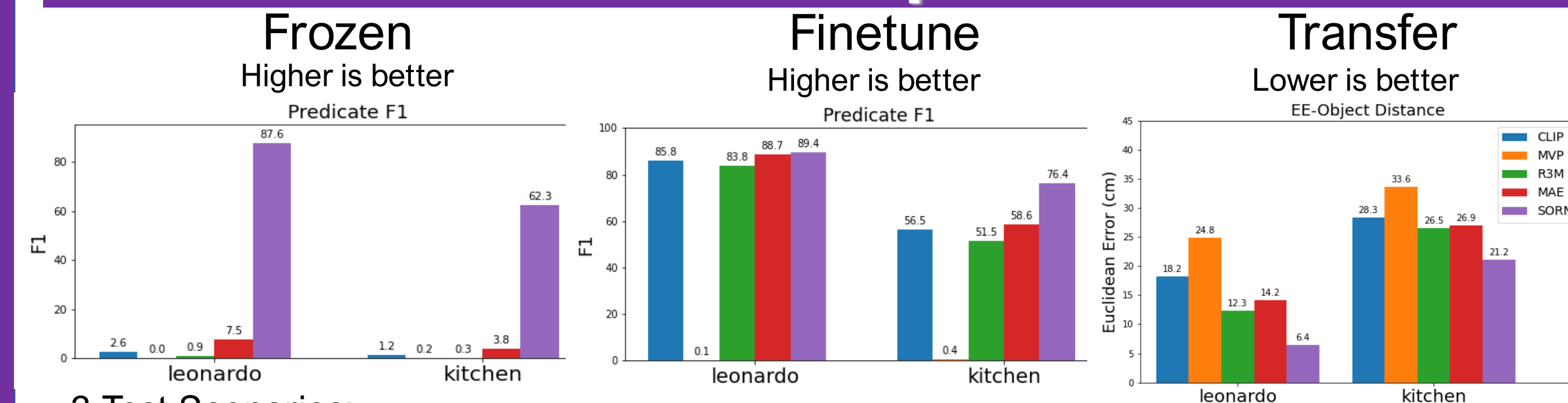
[black] True positive, [blue] False positive, [red] False negative

Spatial direction regression

End-effector to Object direction Object to Object direction



Baseline comparisons



3 Test Scenarios:

- Frozen: freeze embedding, train readout networks on predicate classification
- Finetune: finetune embedding together with readout networks on predicate classification
- Transfer: take embedding finetuned on classification, freeze and train readout networks on direction regression

Key Takeaway:

- SORNet's embedding contains much more spatial information compared to non-object-centric baselines, as shown by its advantage in all 3 test scenarios.
- This indicates that SORNet's embedding can be much more useful in manipulation, where spatial reasoning is critical.