

IFOR: Iterative Flow Minimization for Robotic Object Rearrangement

Ankit Goyal^{1,2} Arsalan Mousavian¹ Chris Paxton¹ Yu-Wei Chao¹

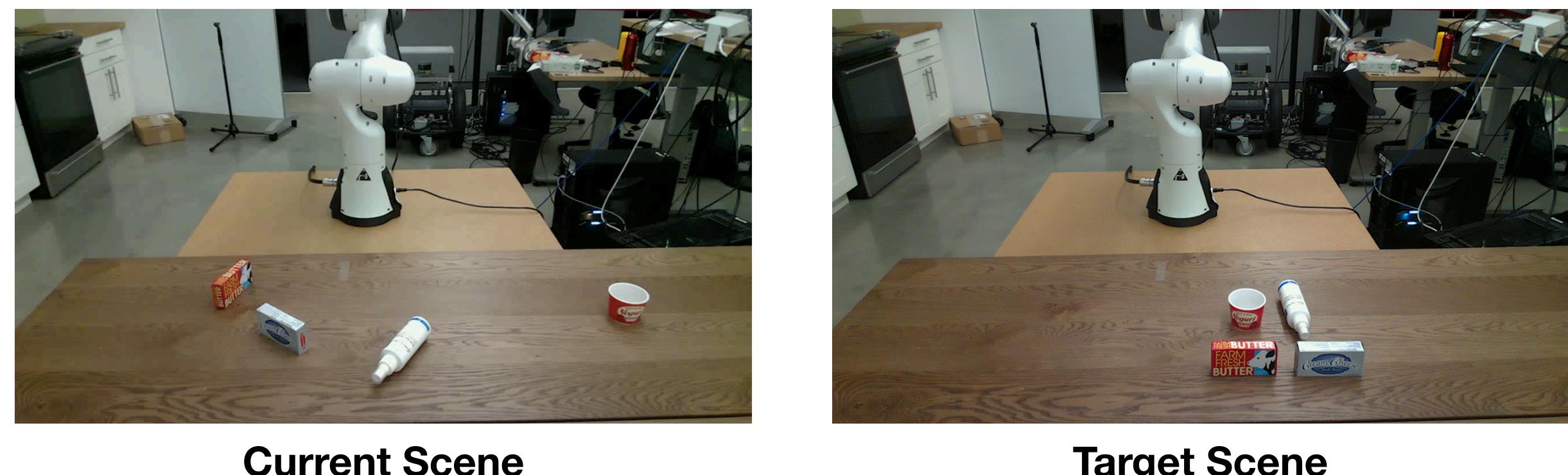
Brian Okorn^{1,3} Jia Deng² Dieter Fox¹

¹NVIDIA ²Princeton University ³CMU

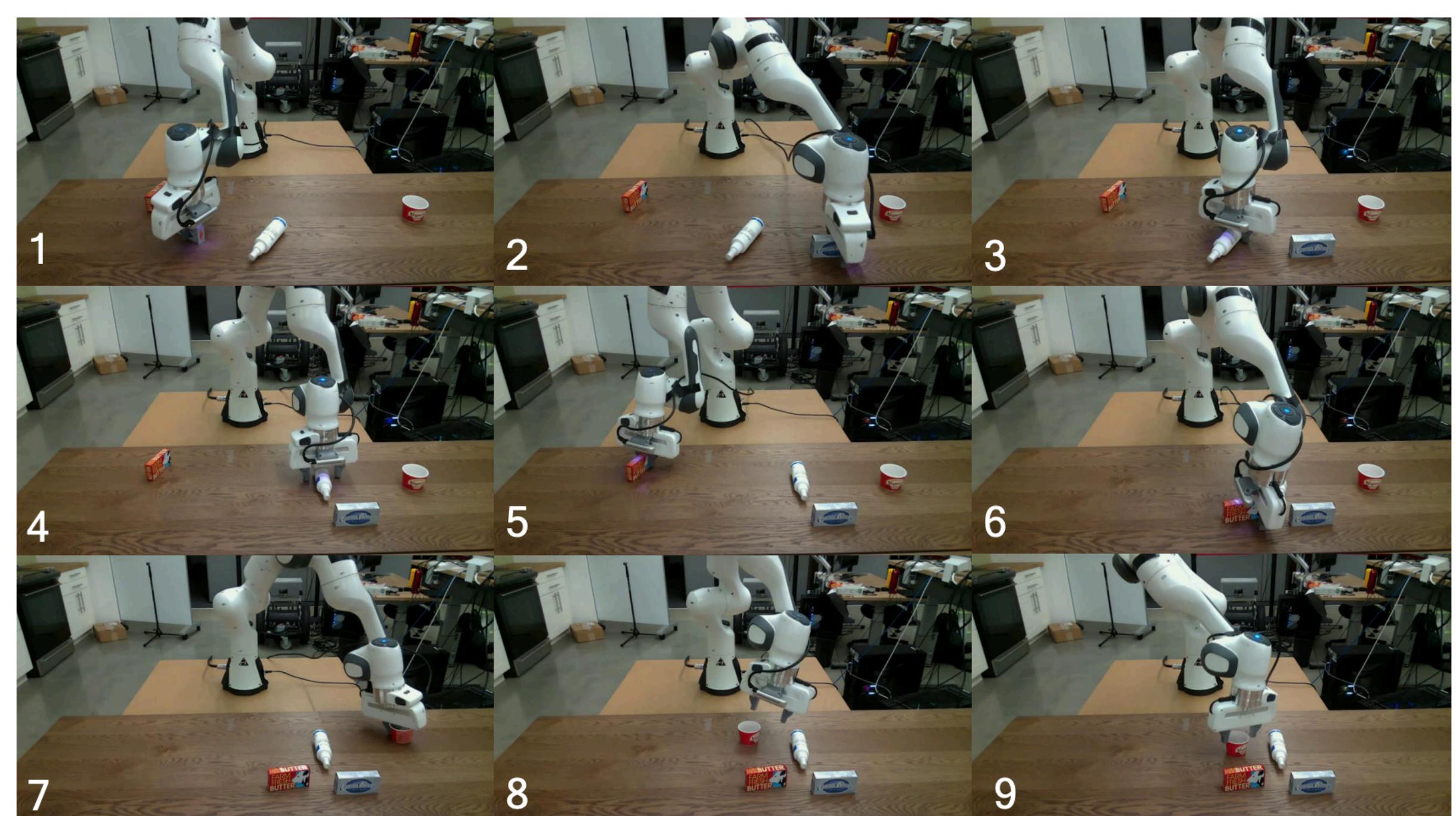


Rearrangement

Input: RGBD of the current and target scene

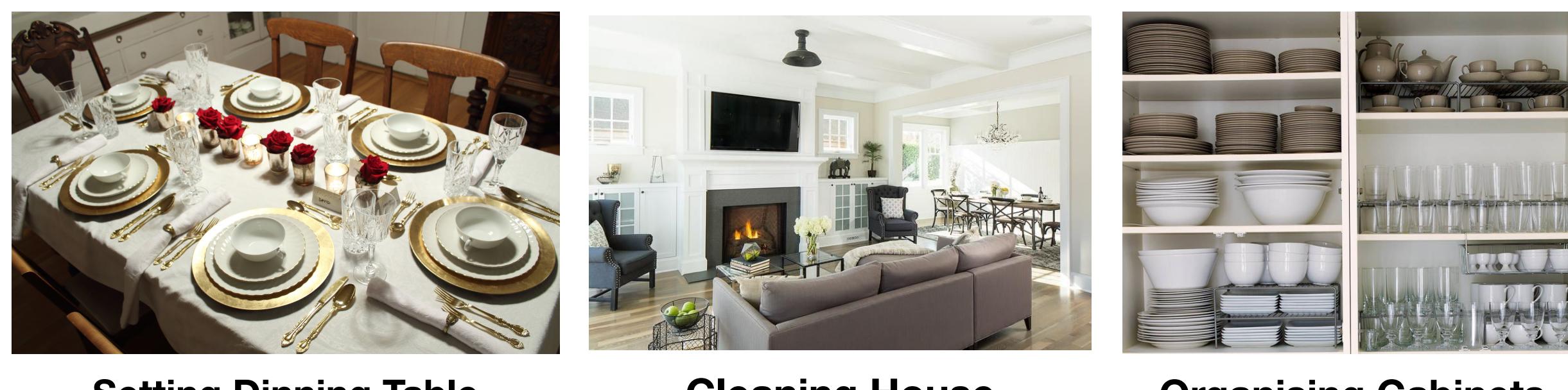


Objective: Rearrange to the target configuration

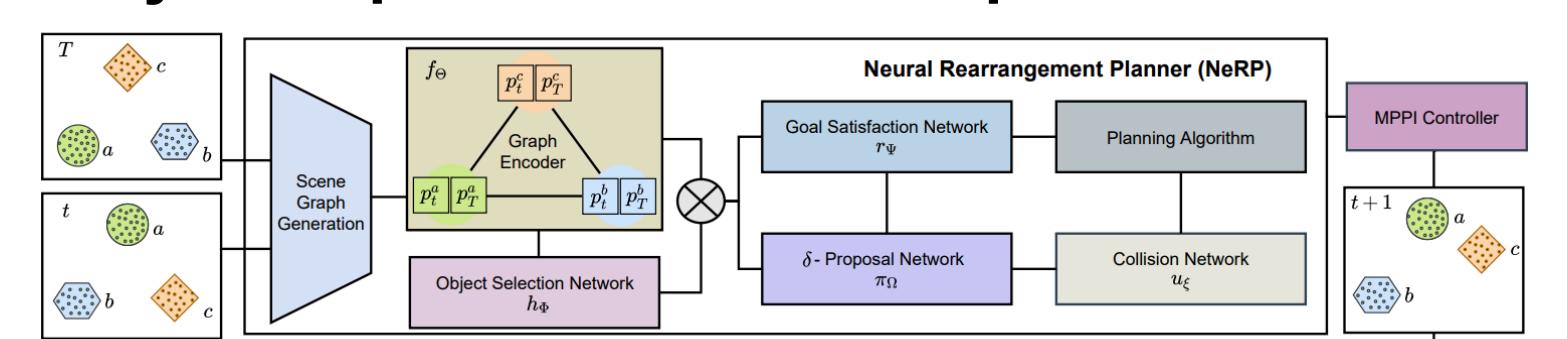


Why Rearrange?

Canonical Embodied AI task and appears in many scenarios



SOTA many components => Complicated + Prone to Failures



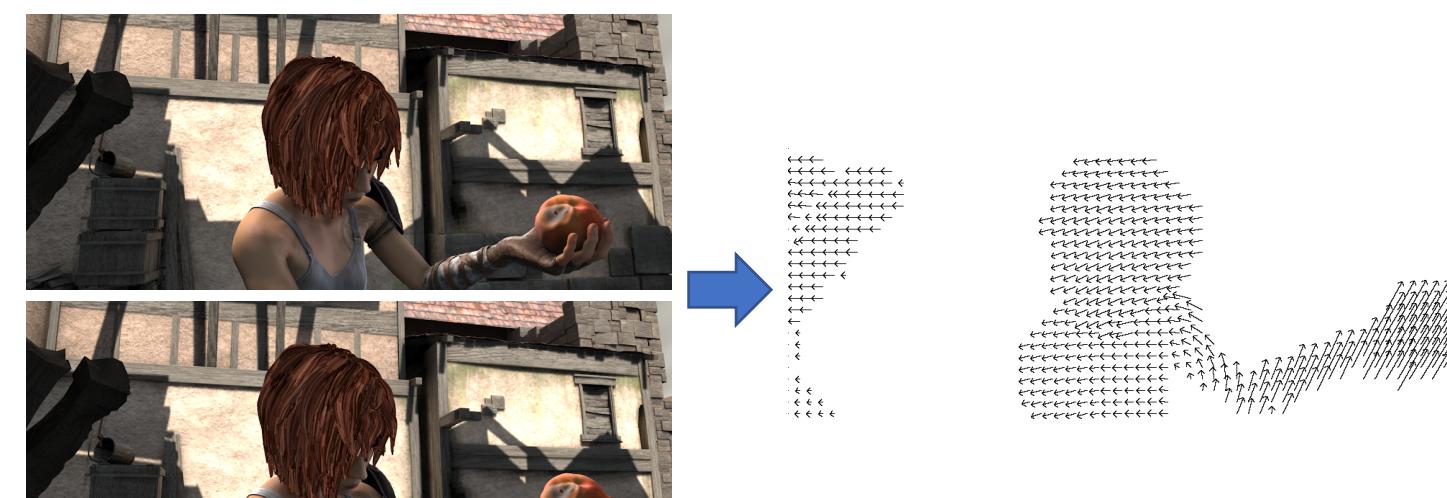
How to Rearrange?

Need to recognize the change in pose of objects.



Challenges

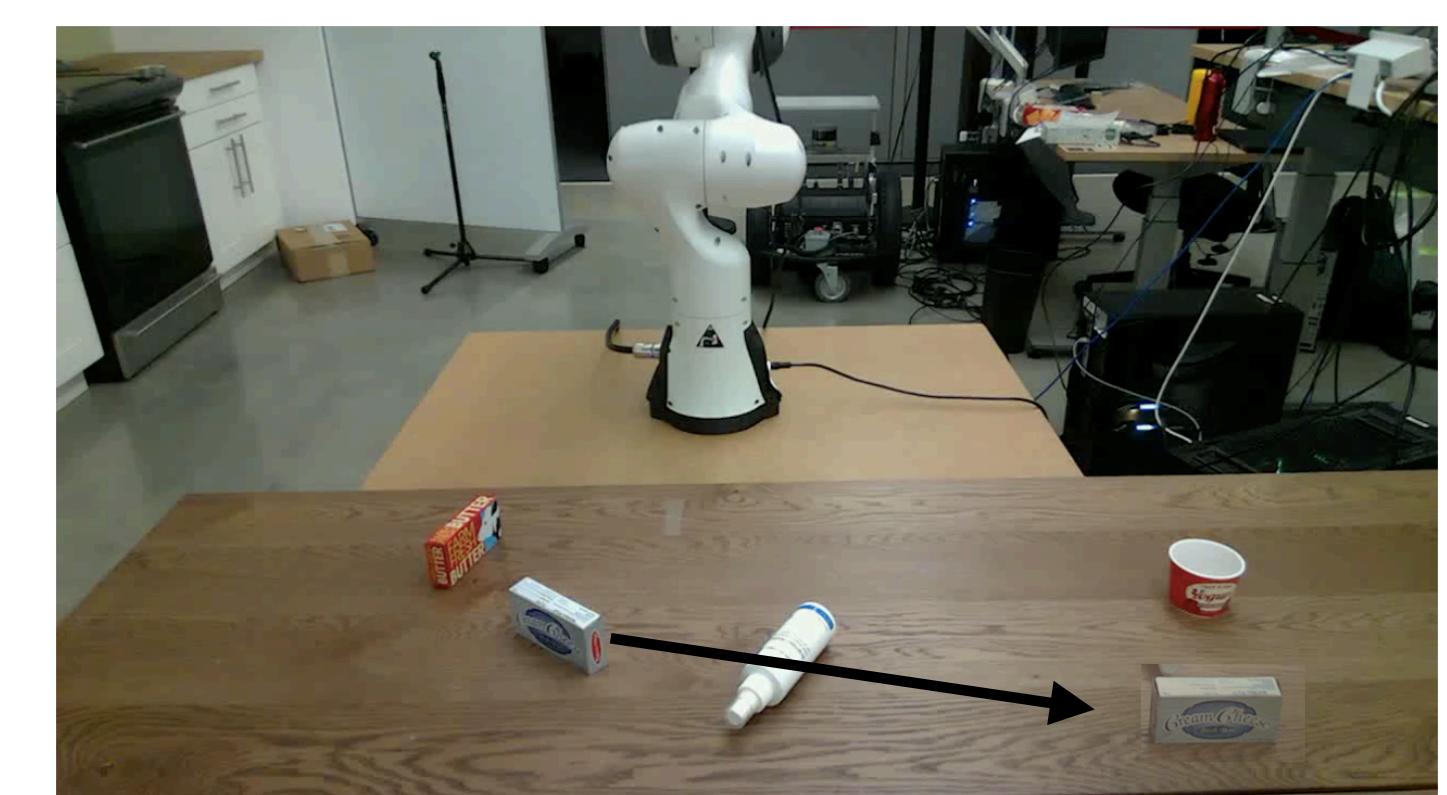
- Challenge: Raw pixel; no privileged info.; unseen object



- Traditional pose estimators won't work 😞

- Object-invariant intermediate representation like flow?

- Solve rigid-body transform from flow (+ segmentation) 😊



- Challenge: Flow values large

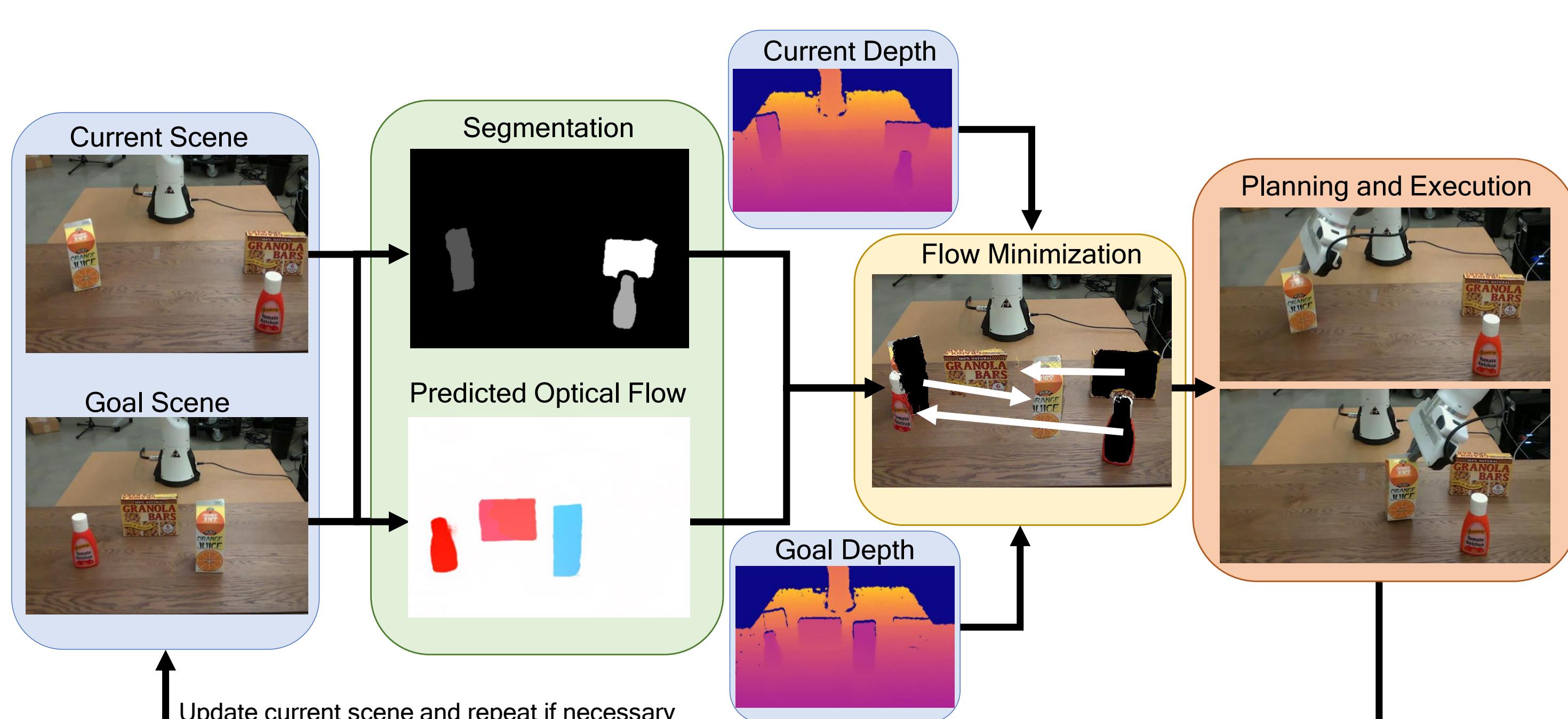
- Traditional flow estimators won't work 😞

- A suitable neural flow estimator with trained correct data?

- Works very well! Transfers from sim-to-real in zero shot 😊

Method: IFOR

IFOR: Iterative Flow Minimization for Robotic Object Rearrangement



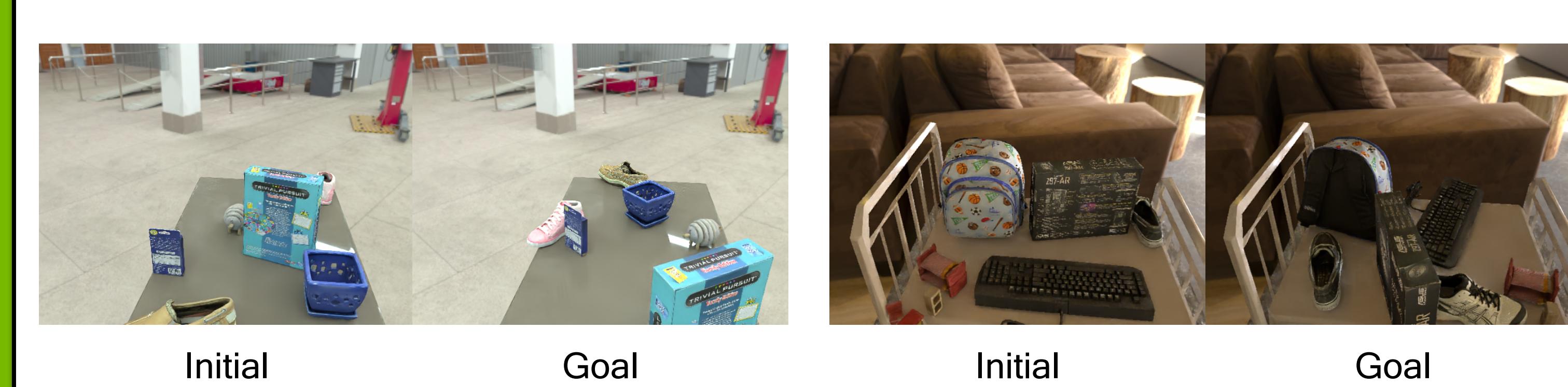
Segmentation: Off-the-shelf Unseen Object Segmentation

Flow Estimation: RAFT + Synthetic Dataset

Solve for rigid-body transformation: Multi-view geometry + RANSAC

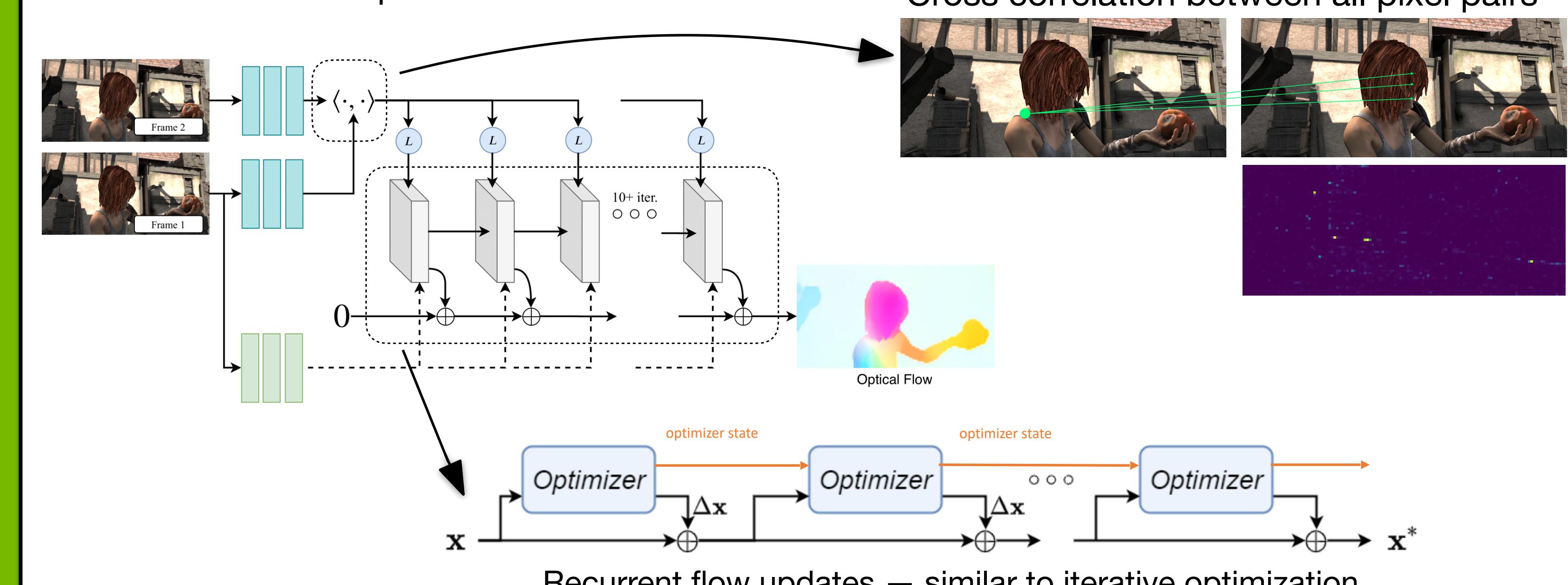
Planning: Prioritizes large transforms and avoids collision

Synthetic Dataset: NVISII Renderer; 50K Samples for Training



Method: IFOR (cont.)

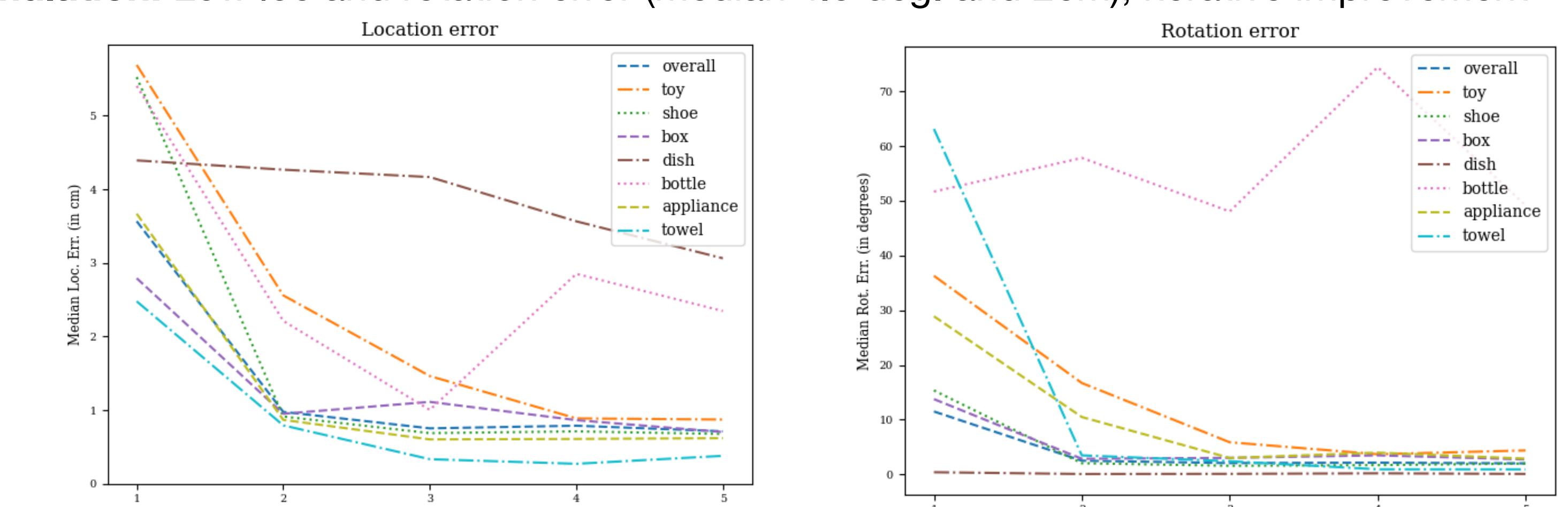
RAFT: Recurrent All-pairs Feature Transforms



Key Observation: Compares each pixel to all other pixels => In theory, learn large flows

Results

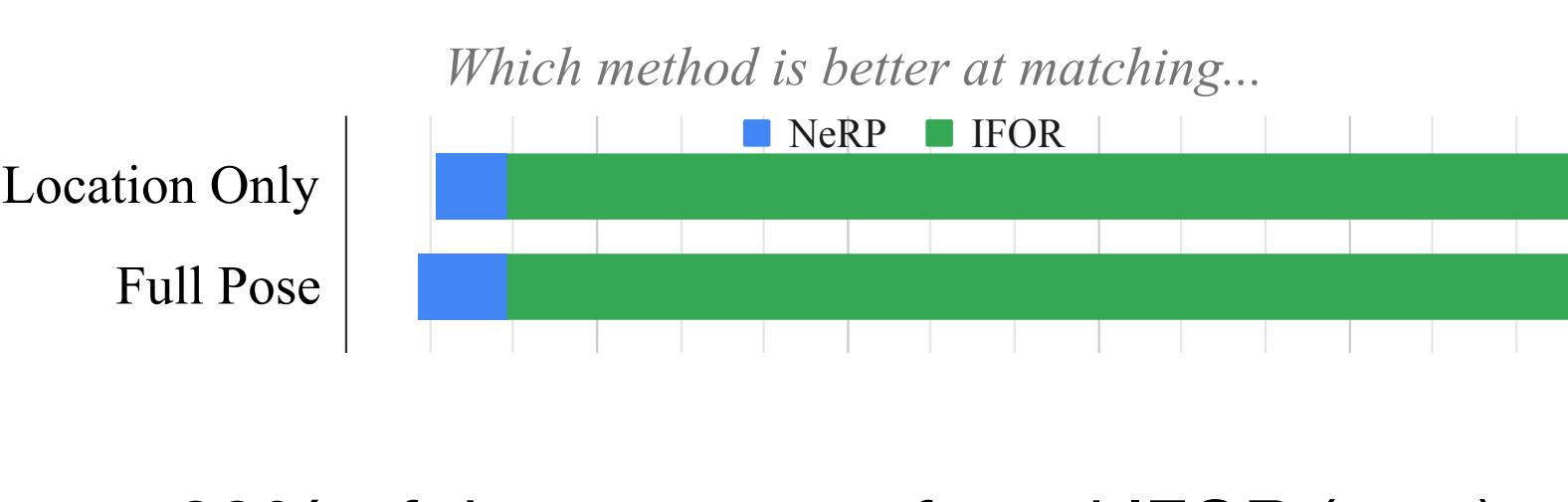
Simulation: Low loc and rotation error (median 4.5 deg. and 2cm); iterative improvement



Real World: Outperforms prior work



IFOR (ours) was consistently rated to perform good!



>92% of time users preferred IFOR (ours) over prior-art

Summary

- Robotic system for vision based object rearrangement; iteratively minimizes flow.
- Input is RGBD images of the current and target scene; no privileged information used
- Trained entirely in simulation; zero-shot transfer to real world.
- Tested on objects not seen during training

