# *ScrewNet: Category-Independent Articulation Model Estimation From Depth Images Using Screw Theory*

Ajinkya Jain, Rudolf Lioutikov,
Caleb Chuck, Scott Niekum

## MOTIVATION

Robots in human environments will need to interact with a wide variety of articulated objects such as cabinets, drawers, and dishwashers while assisting humans in performing day-to-day tasks. Existing methods either require objects to be textured or need to know the articulation model category a priori for estimating the model parameters for an articulated object.

## ScrewNet

- estimates articulation model for an object directly from raw depth images
- does not require knowing articulation model category a priori for model estimation
- uses screw theory to unify the representation of different articulation types, enabling it to perform category-independent articulation model estimation
- uses a single network for estimating models for all common articulation model categories unlike prior methods [1, 2]
- can also represent an additional articulation model category, the helical model, without making any changes to the network or the training procedure

## Evaluation

We evaluate our approach on two benchmarking datasets[1, 3] and a real-world dataset.

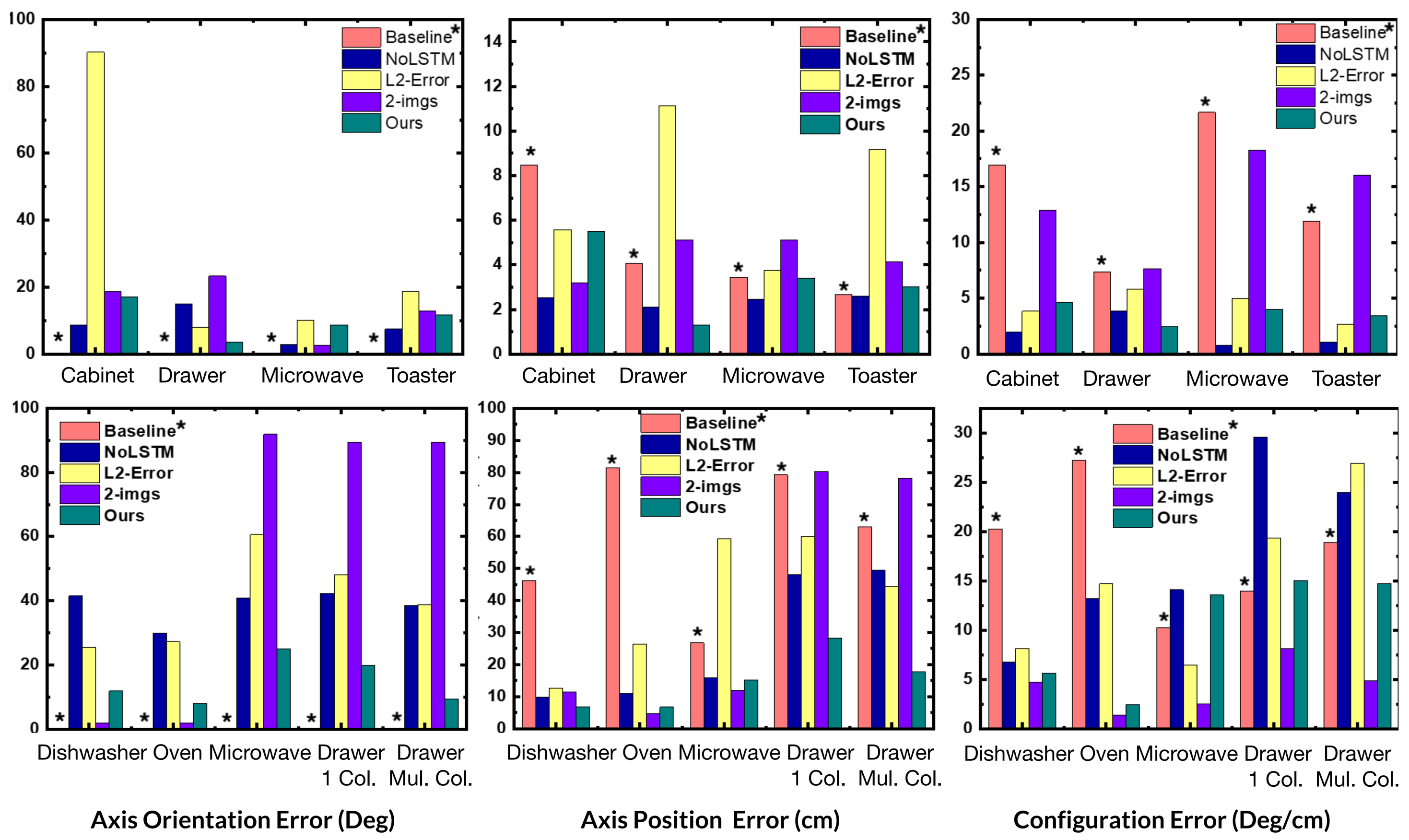



**Fig. 1. [Benchmarking datasets]** Mean error values for the joint axis orientations, positions, and joint configurations for 1000 test objects for each object class from **(top)** simulated articulated objects dataset[1] and **(bottom)** PartNet-Mobility Dataset[3]. Results demonstrate that ScrewNet (green) can estimate the articulation models and their parameters for novel objects across articulation model categories with better on average accuracy than the prior state-of-the-art method[1]. (baseline; pink) and its ablated versions (other colors). Lower errors denote better performances. Symbol (⋆) denote that the baseline[1] has a significant advantage over other methods as it uses a separate network for each object class.

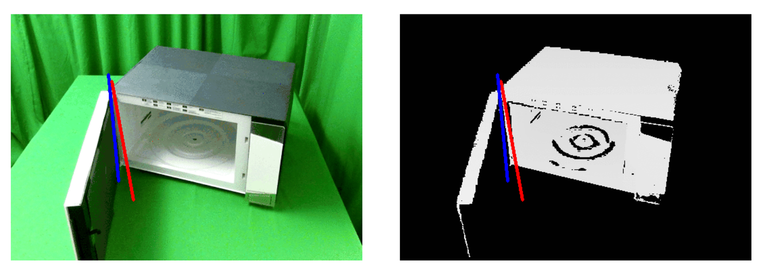
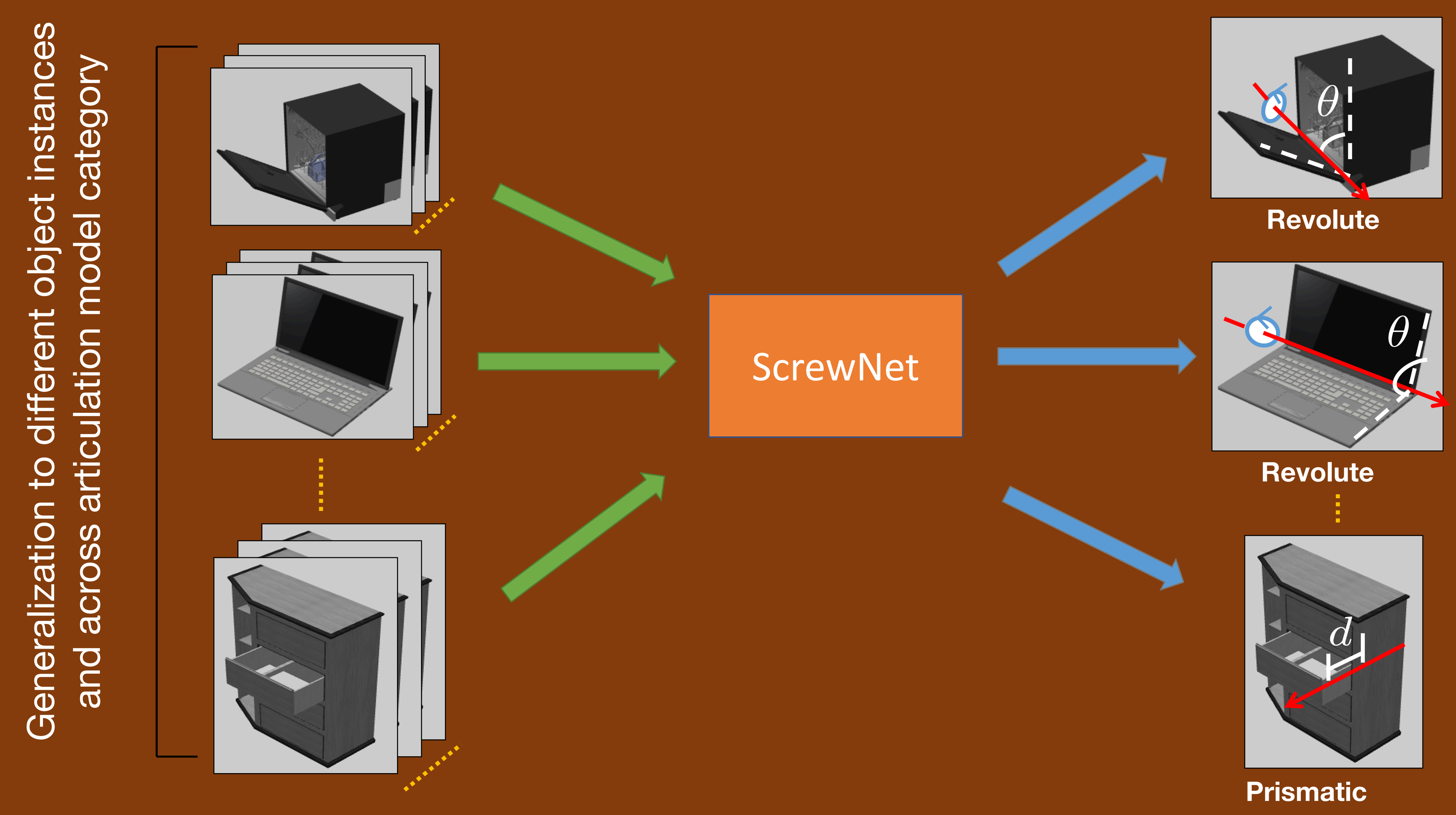


**Fig. 2. [Real-world Dataset]** Color and depth Images with overlayed ground-truth joint axis (blue) and ScrewNet's predicted axis (red) for a microwave where the network was trained solely using simulated data

# Our screw-theory based method learns a unified representation, that estimates **articulation properties** for various objects **directly from depth images**, without requiring a priori knowledge of the model category!



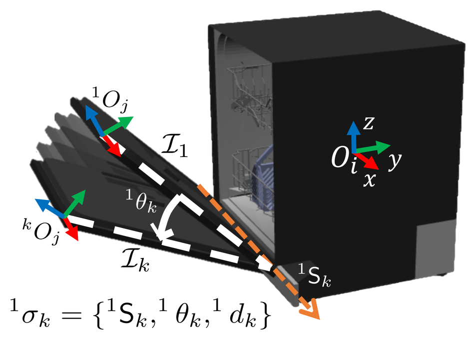

Input: Sequence of depth Images

Output:
Articulation Model Category
Joint Parameters
Joint Configurations

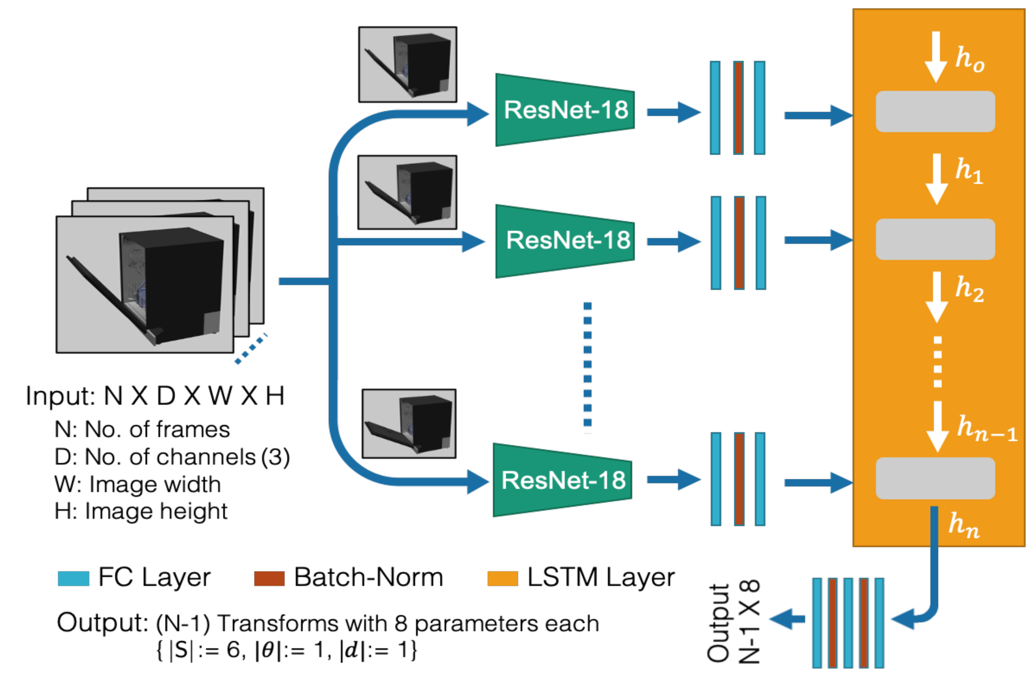
**Take a picture** for more details

## Background

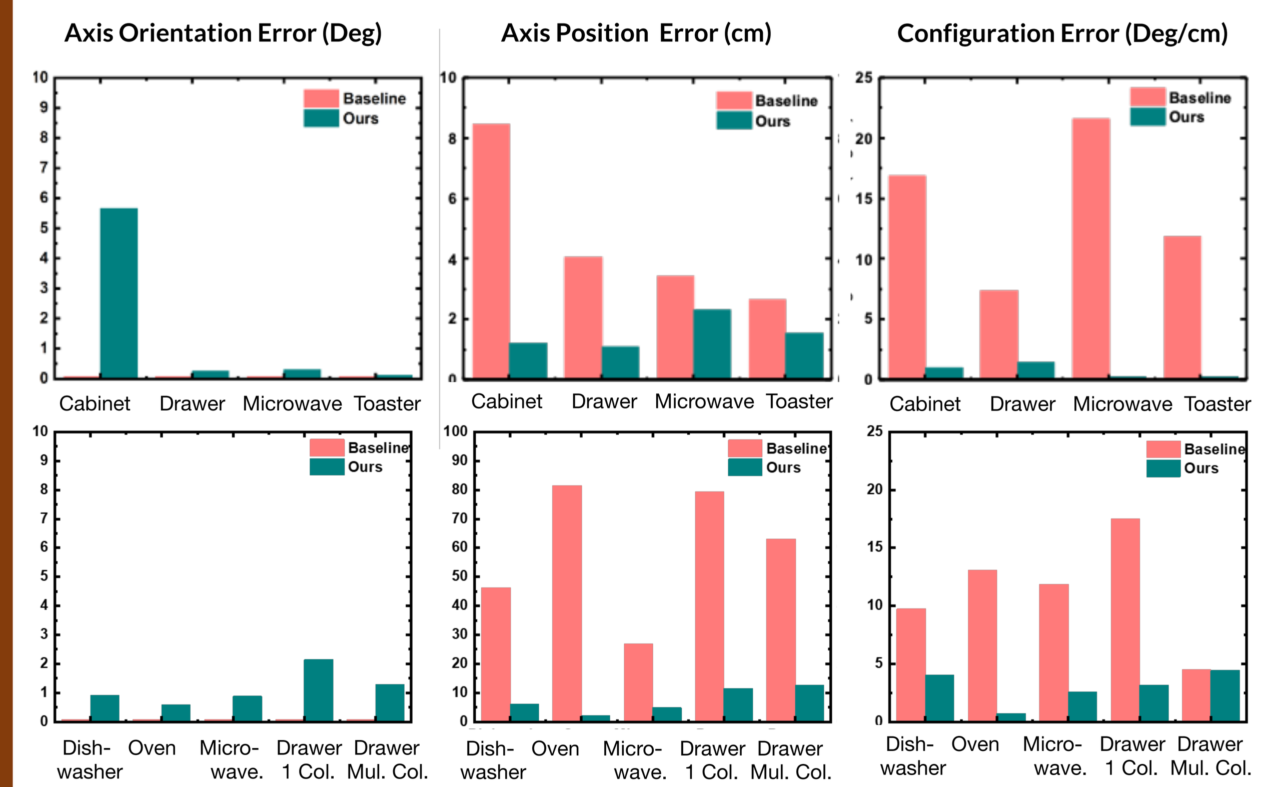
- Chasles' Theorem: "*Any displacement of a body in space can be accomplished by means of a rotation of the body about a unique line in space accompanied by a translation of the body parallel to that line*"
- Motion of an articulated body can be seen as a sequence of rigid body displacements along a shared screw axis



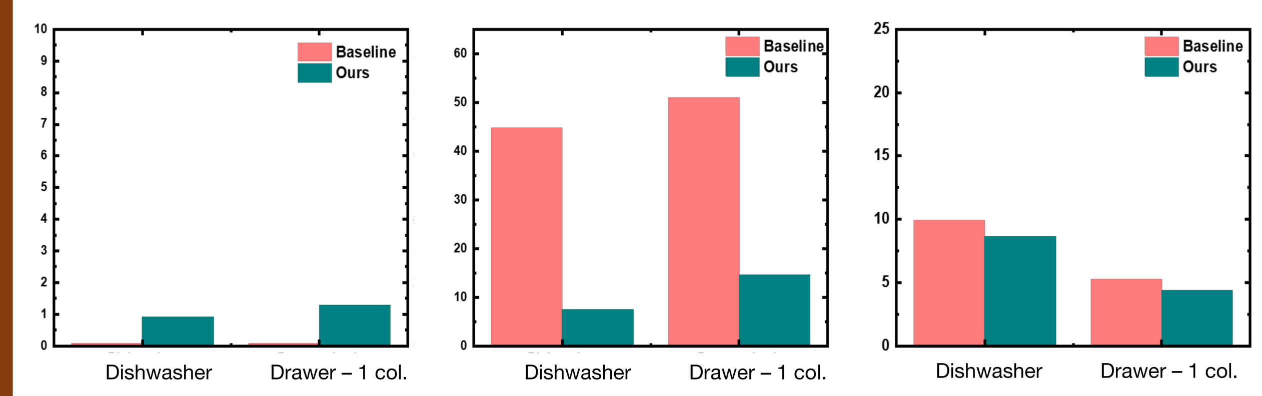

## ScrewNet Architecture




## Further Results




Unseen object instances belonging to known object classes. **(Top)** simulated articulated objects dataset[1] and **(Bottom)** PartNet-Mobility Dataset[3] for. A single ScrewNet and a baseline network were trained for individual object classes. Results show that ScrewNet (green) outperforms the baseline (pink)




Generalization performance to unseen object classes belonging to same articulation model category on the PartNet-Mobility Dataset

## References:

[1] Abbatematteo et al., "Learning to Generalize Kinematic Models to Novel Objects," CoRL'19
[2] Li et al. "Category-Level Articulated Object Pose Estimation". CVPR'20
[3] Xiang et al., "SAPIEN: A SimulAted Part-based Interactive ENvironment," CVPR'20

TEXAS The University of Texas at Austin
Personal Autonomous Robotics Lab
PeARL