Robotic Telekinesis:

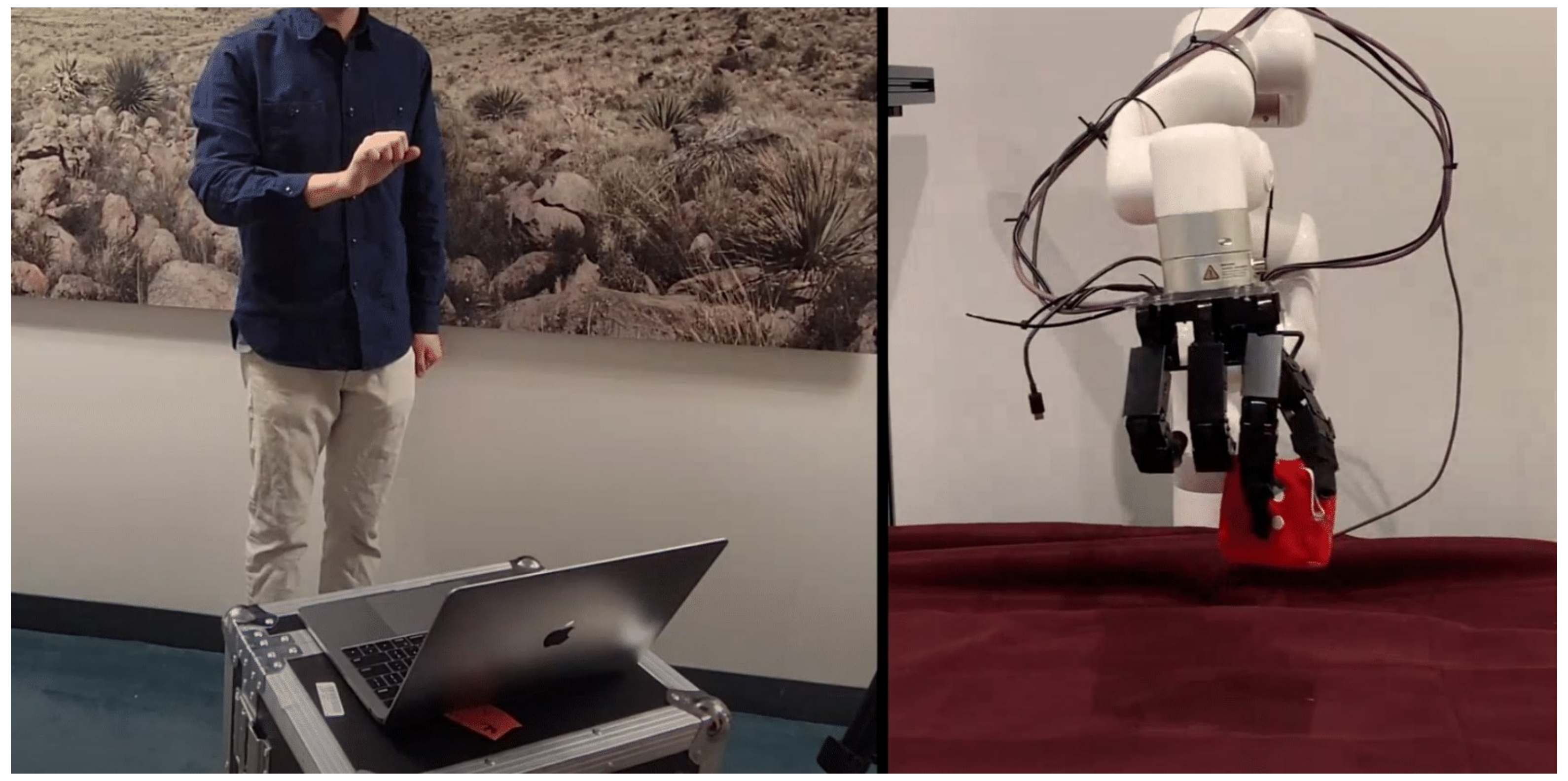
Learning a Robotic Hand Imitator by Watching Humans on YouTube

Aravind Sivakumar* Kenneth Shaw* Deepak Pathak

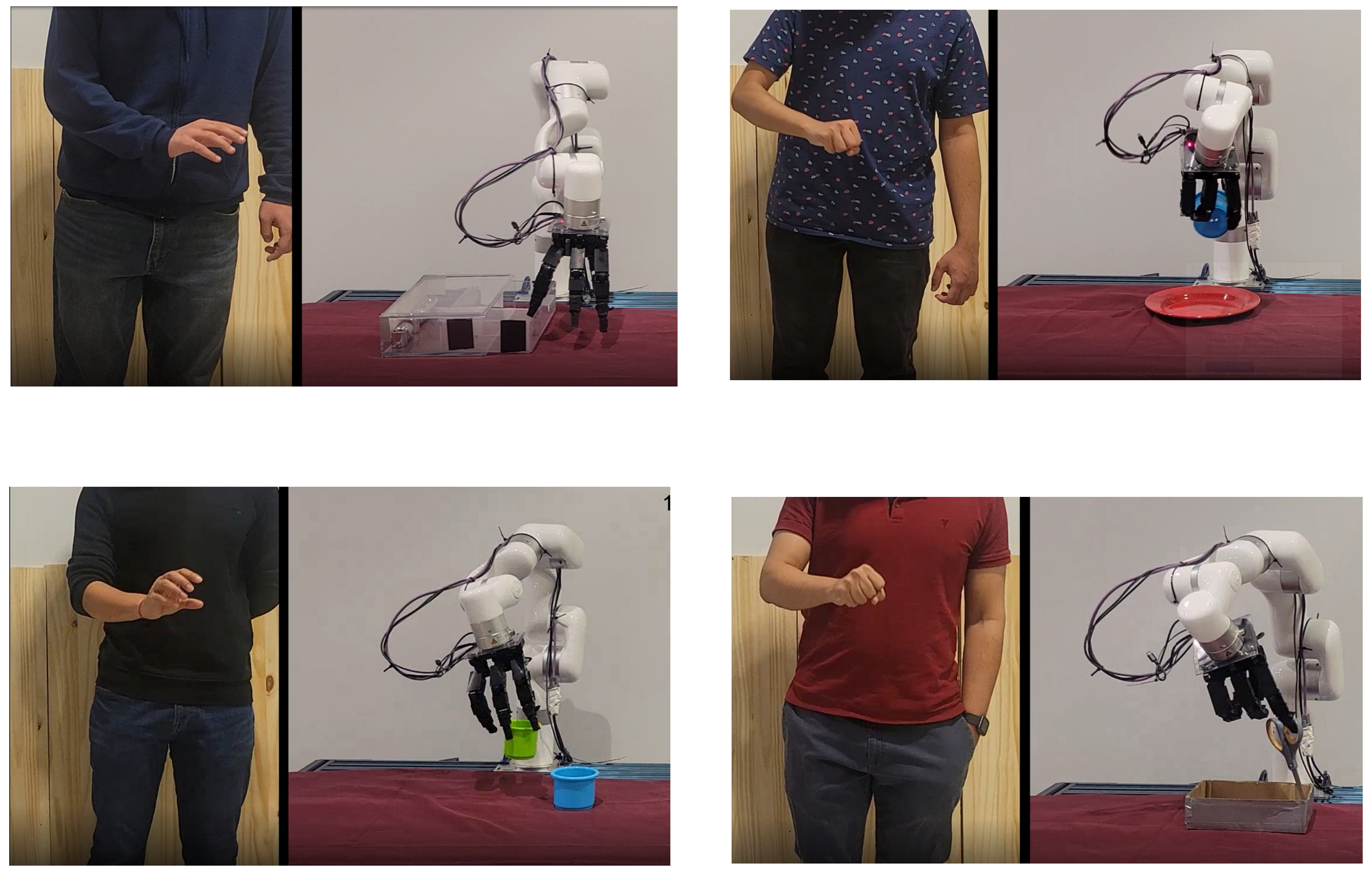
Carnegie Mellon University

See Demo videos at https://robotic-telekinesis.github.io/

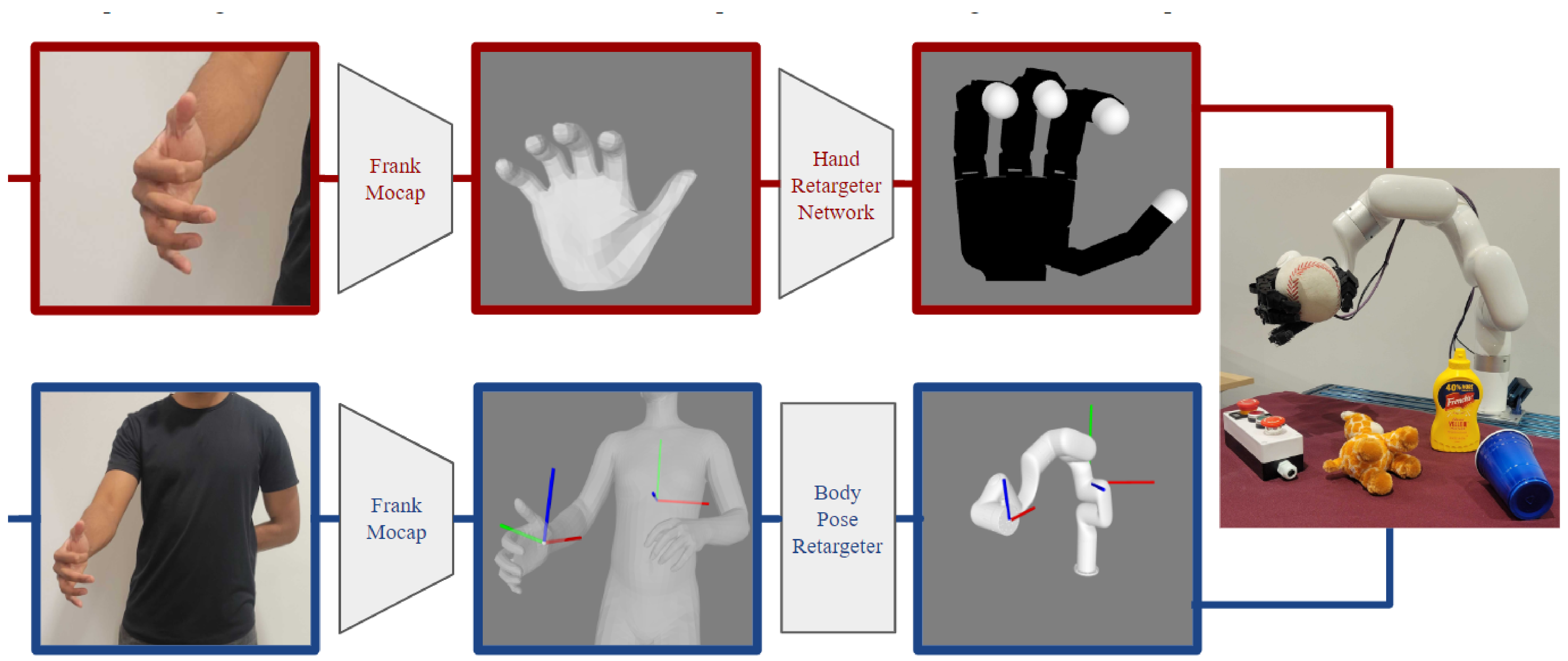
We use passive data from the internet to enable robotic real-time imitation in-the-wild!



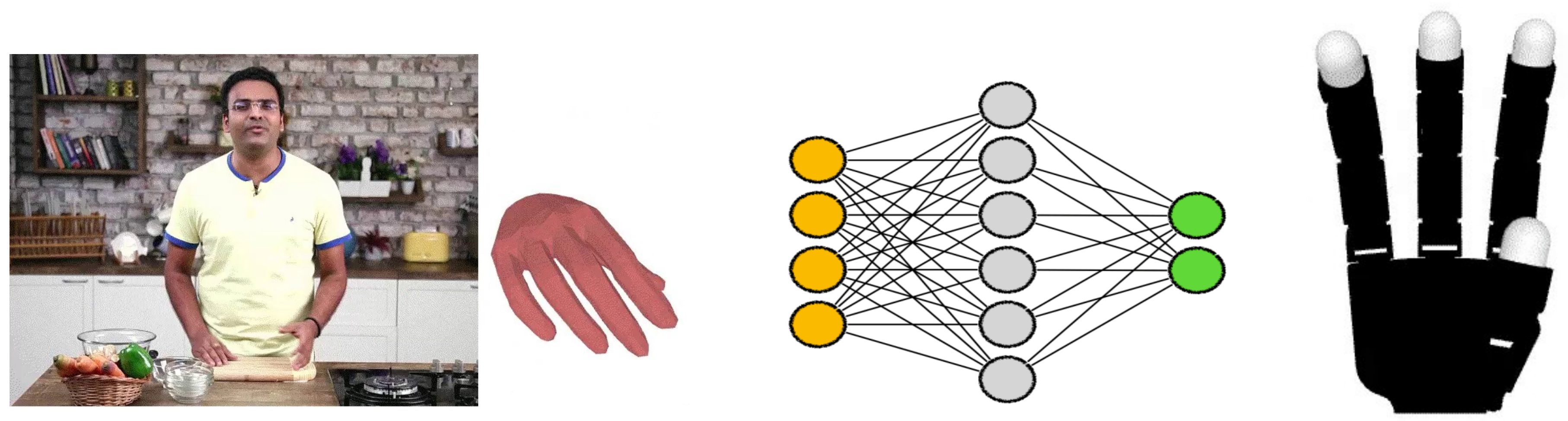
The operator can use any camera, including webcams!



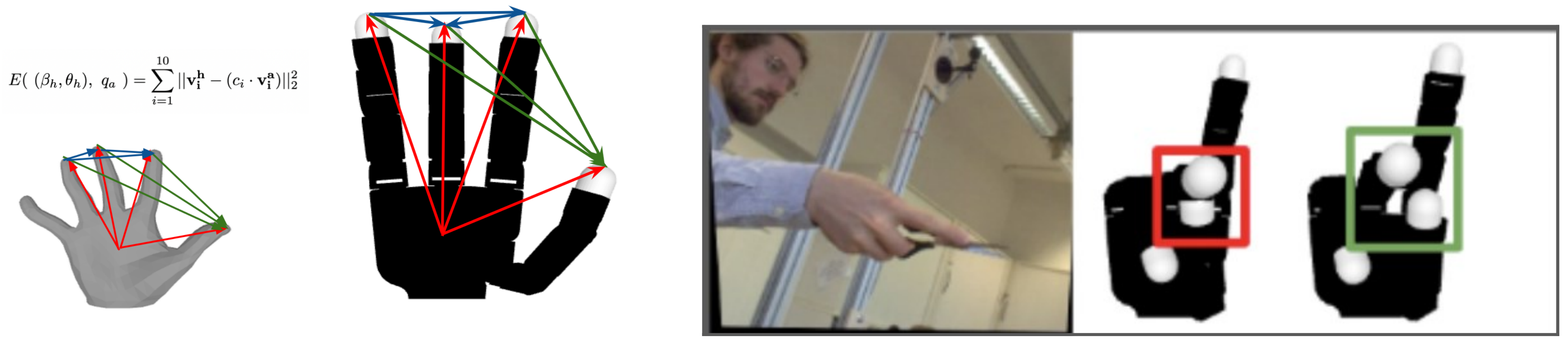
Any operator can complete any type of dexterous task!



Top: A hand pose estimator, and the hand retargeting network maps the estimated human hand to robot hand pose. **Bottom:** A body pose estimator and cross-morphology correspondences determine the desired arm pose.

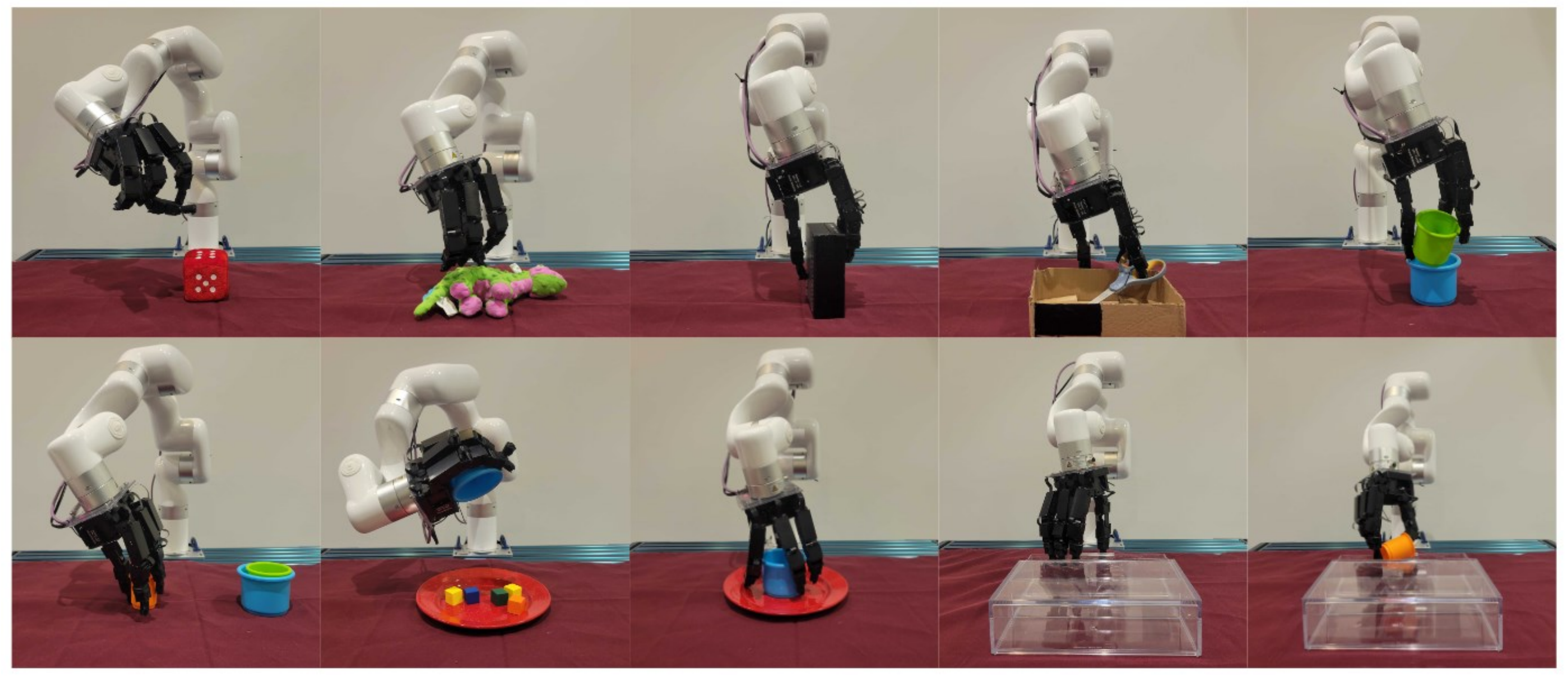


We train a human-to-robot retargeting network to map human hands to robot poses by watching thousands of hours of YouTube videos of people using their hands



$$E(\ (\beta_h,\theta_h),\ q_a\ ) = \sum_{i=1}^{10} ||\mathbf{v_i^h} - (c_i \cdot \mathbf{v_i^a})||_2^2$$

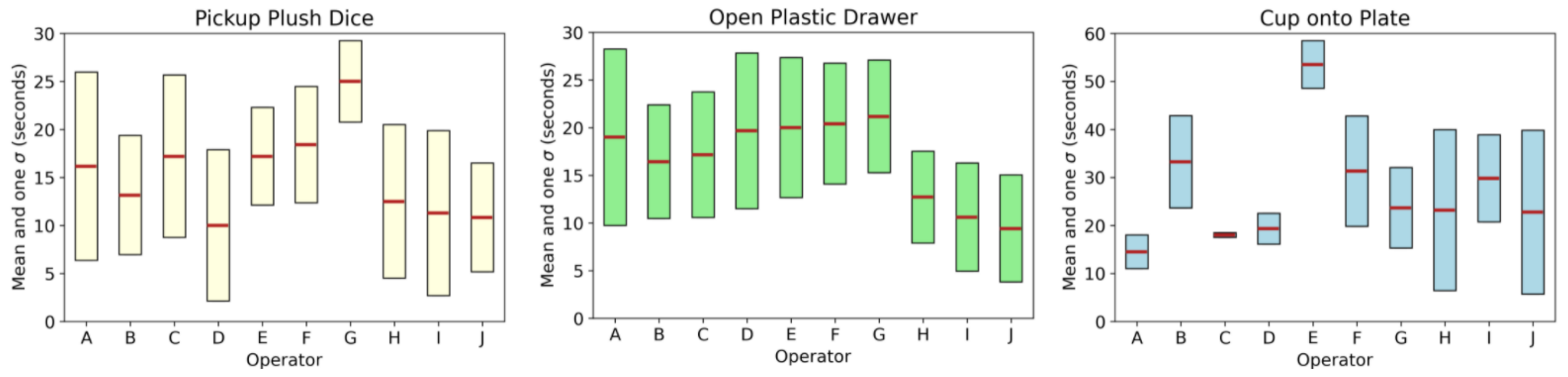
An advantage of NN retargeters is we can use a trained self-collision classifier that is an adversary and penalizes self-colliding poses.



Ten Tasks Completed on Telekinesis

	Success (rate)		Completion Time (sec)	
Task	Ours	DexPilot-Mono*	Ours	DexPilot-Mono*
Pickup Dice Toy	**0.9**	0.7	**8.6 (2.65)**	13.5 (5.47)
Pickup Dinosaur Doll	**0.9**	0.6	**8.2 (3.49)**	11.00 (3.95)
Box Rotation	**0.6**	0.3	37.2 (12.6)	**16.33 (10.69)**
Scissor Pickup	**0.7**	0.5	28.6 (9.4)	**27.66 (11.09)**
Cup Stack	0.6	**0.7**	**21.5 (7.6)**	22.85 (16.57)
Two Cup Stacking	**0.3**	0.1	**27.3 (11.0)**	45.00 (0.0)
Pouring Cubes onto Plate	**0.7**	0.5	36.80 (17.7)	**13.8 (4.02)**
Cup Into Plate	**0.8**	0.7	**10.6 (4.4)**	13.71 (5.44)
Open Drawer	**0.9**	**0.9**	23.6 (12.3)	**14.88 (4.40)**
Open Drawer and Pickup Cup	0.6	**0.7**	33.7 (8.1)	**28.14 (11.48)**

DexPilot-Mono* is nearly identical to ours, but uses online gradient descent for hand pose retargeting inspired by DexPilot. Ours uses a neural network retargeter and out-performs the baseline.



Inexperienced Operators can complete tasks as well!