RH20T: A Robotic Dataset for Learning Diverse Skills in One-Shot

Hao-Shu Fang, Hongjie Fang, Zhenyu Tang, Jirong Liu, Junbo Wang, Haoyi Zhu, Cewu Lu
Shanghai Jiao Tong University
fhaoshu@gmail.com, {galaxies, tang_zhenyu, jirong, sjtuwjb3589635689, zhuhaoyi, lucewu}@sjtu.edu.cn

ROBOTICS SCIENCE AND SYSTEMS

rh20t.github.io

140+ diverse skills
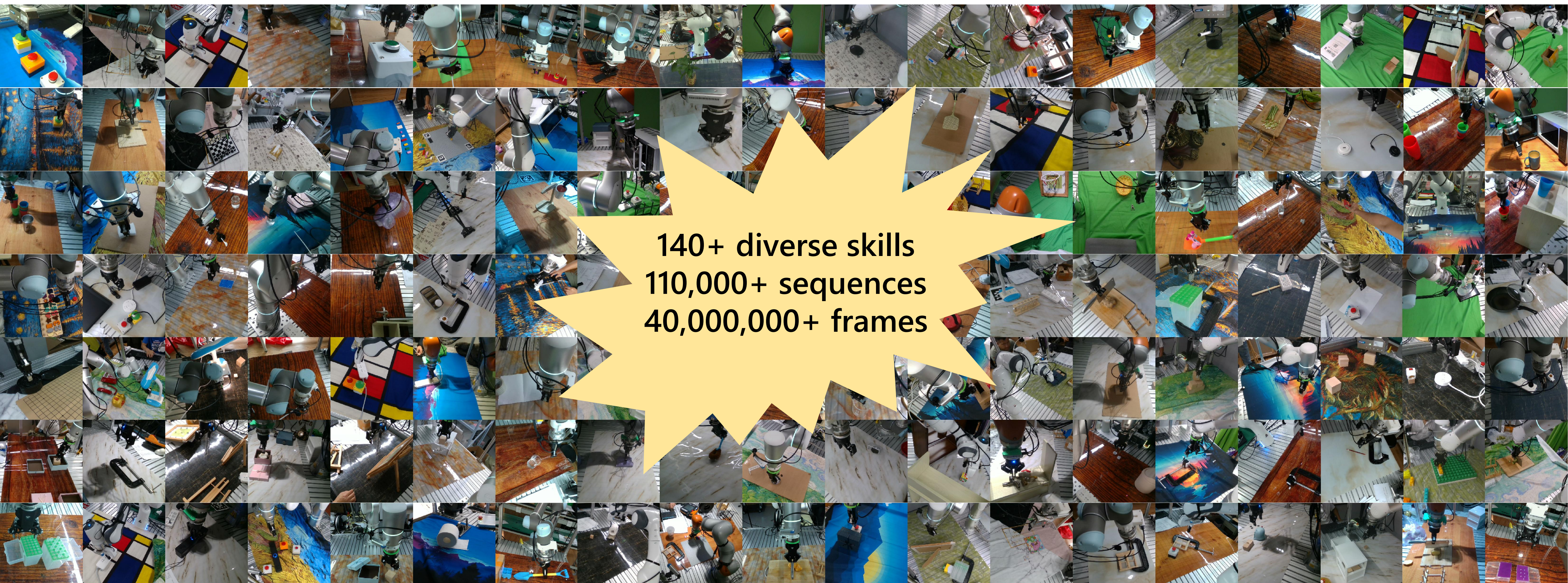110,000+ sequences
40,000,000+ frames

Fig.1: Sampled RBG images from RH20T. Our dataset contains diverse skills, robots, viewpoints, objects, backgrounds, etc. Note that these images are center-cropped for better visualization. A more detailed illustration of all the tasks is given in the paper.
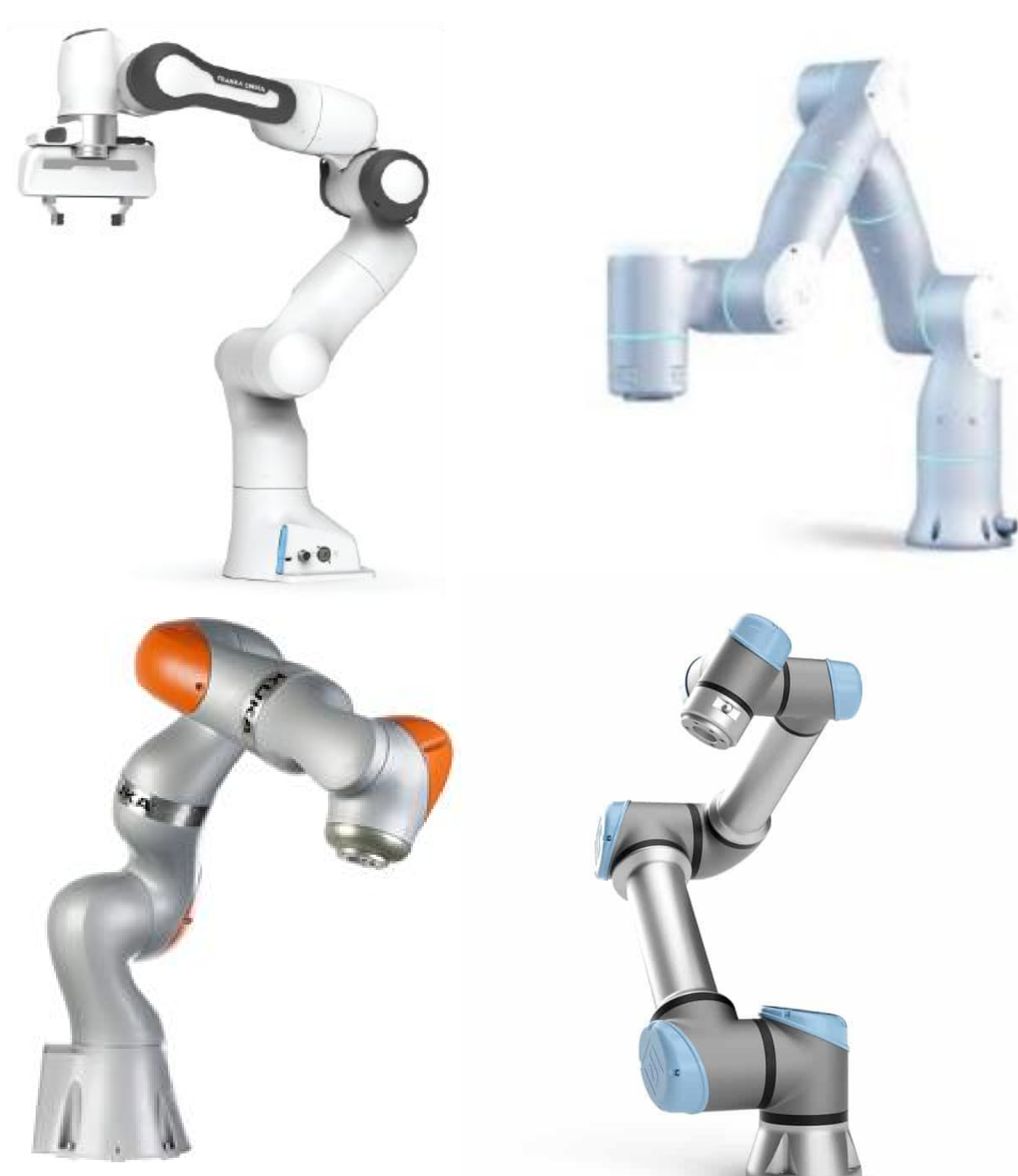

Fig.2: The dataset contains 4 robots, 4 grippers, 70 camera viewpoints, 3 force-torque sensors, 1 tactile sensor, etc.
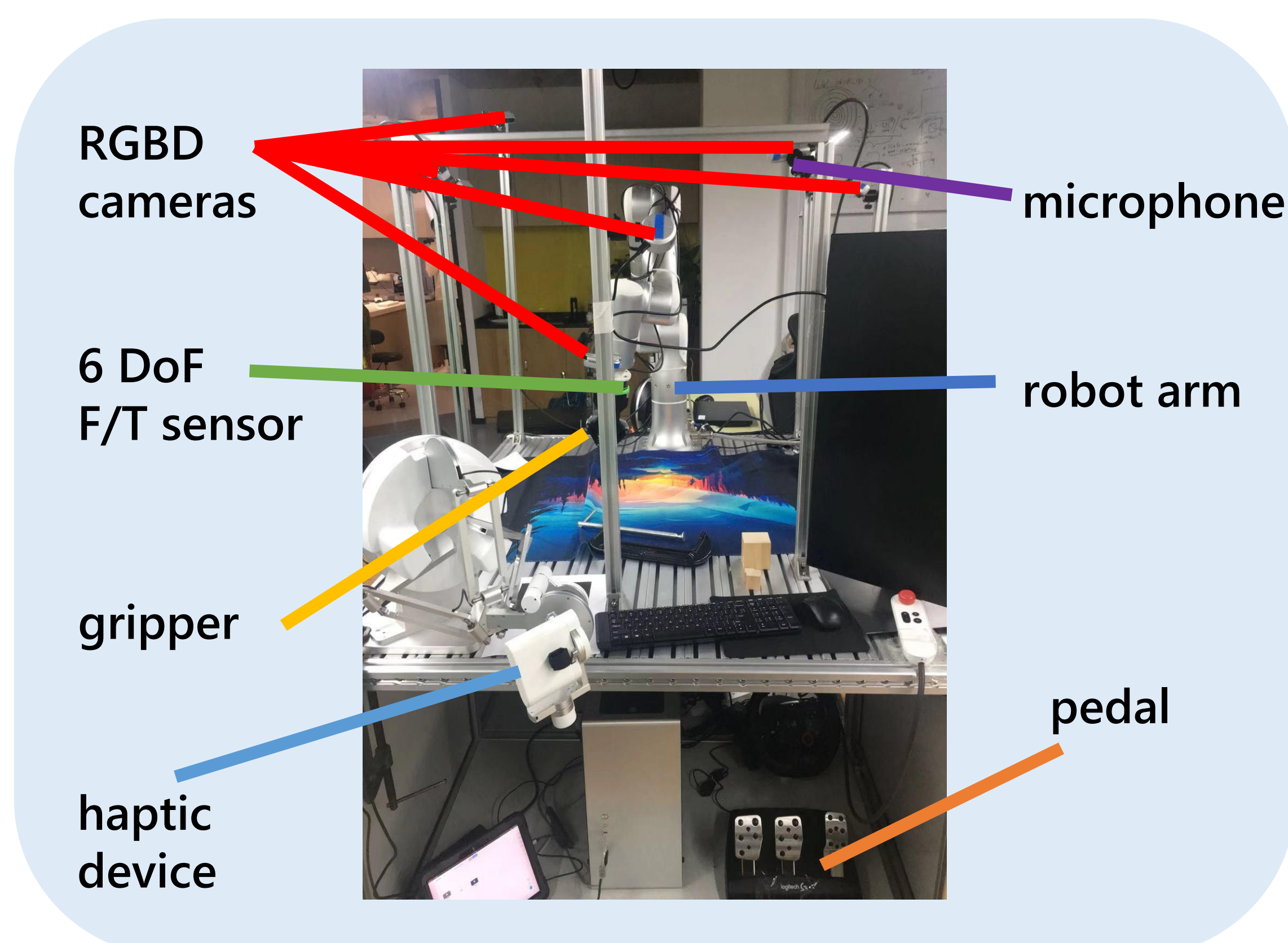

RGBD cameras
6 DoF F/T sensor
gripper
haptic device
microphone
robot arm
pedal

Fig.3: Each platform contains 8-10 RGBD cameras for data collection. A haptic device and a pedal are utilized to tele-operate the robot accurately.

Modal	Size	Freq.
RGB image	1280×720	10 Hz
Depth image	1280×720	10 Hz
Binocular IR images	1280×720	10 Hz
Robot joint angle	7	10 Hz
Robot joint torque	7	10 Hz
Gripper Cartesion pose	7	100 Hz
Gripper width	1	10 Hz
6DoF F/T	6	100 Hz
Audio	N/A	30 Hz
Fingertip tactile	2×16×3	200 Hz

Tab.1: Data modality in our dataset. The last data modality of fingertip tactile sensing is only available in robot Cfg 7.
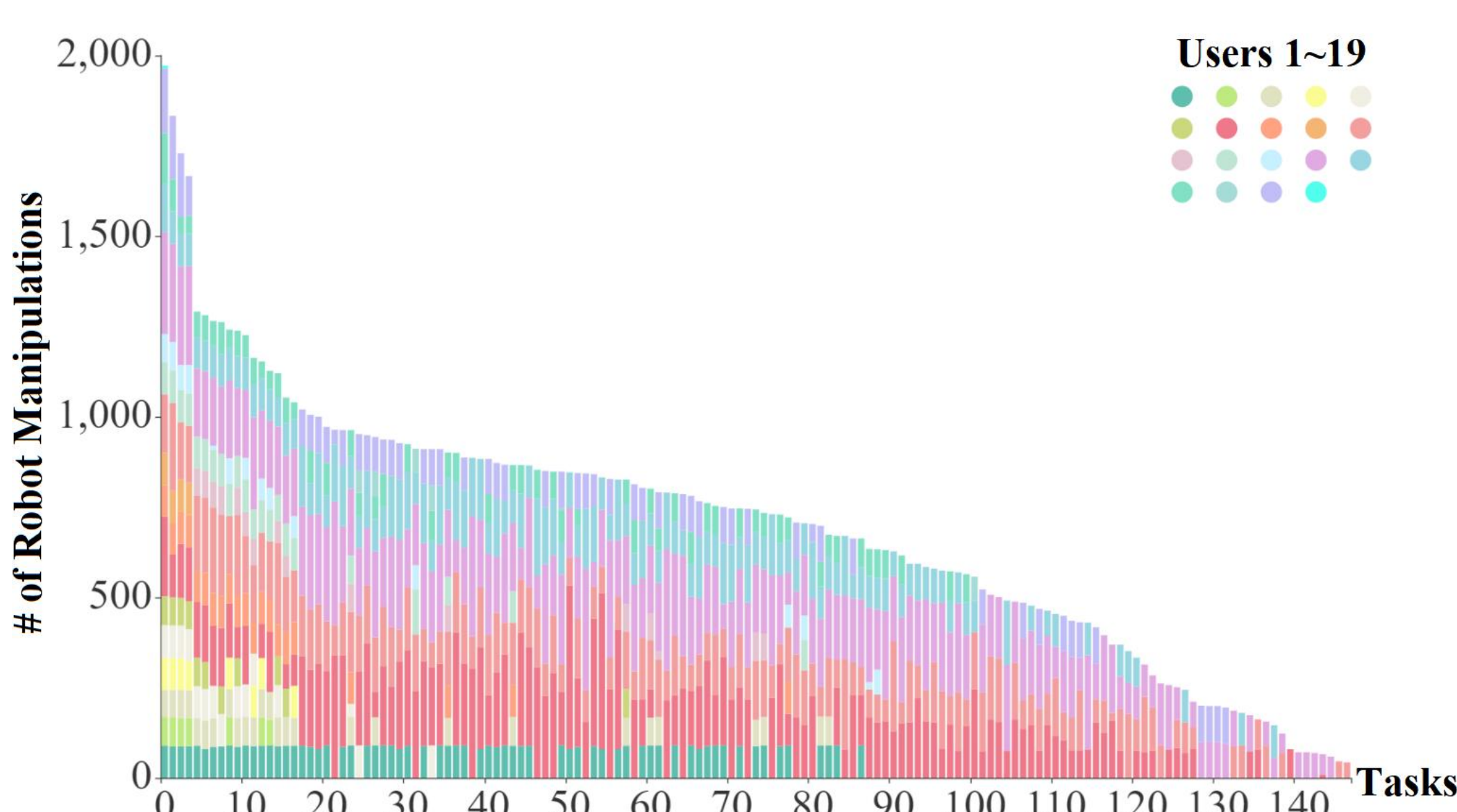

Fig.4: Statistics on the amount of robotic manipulation for different tasks.
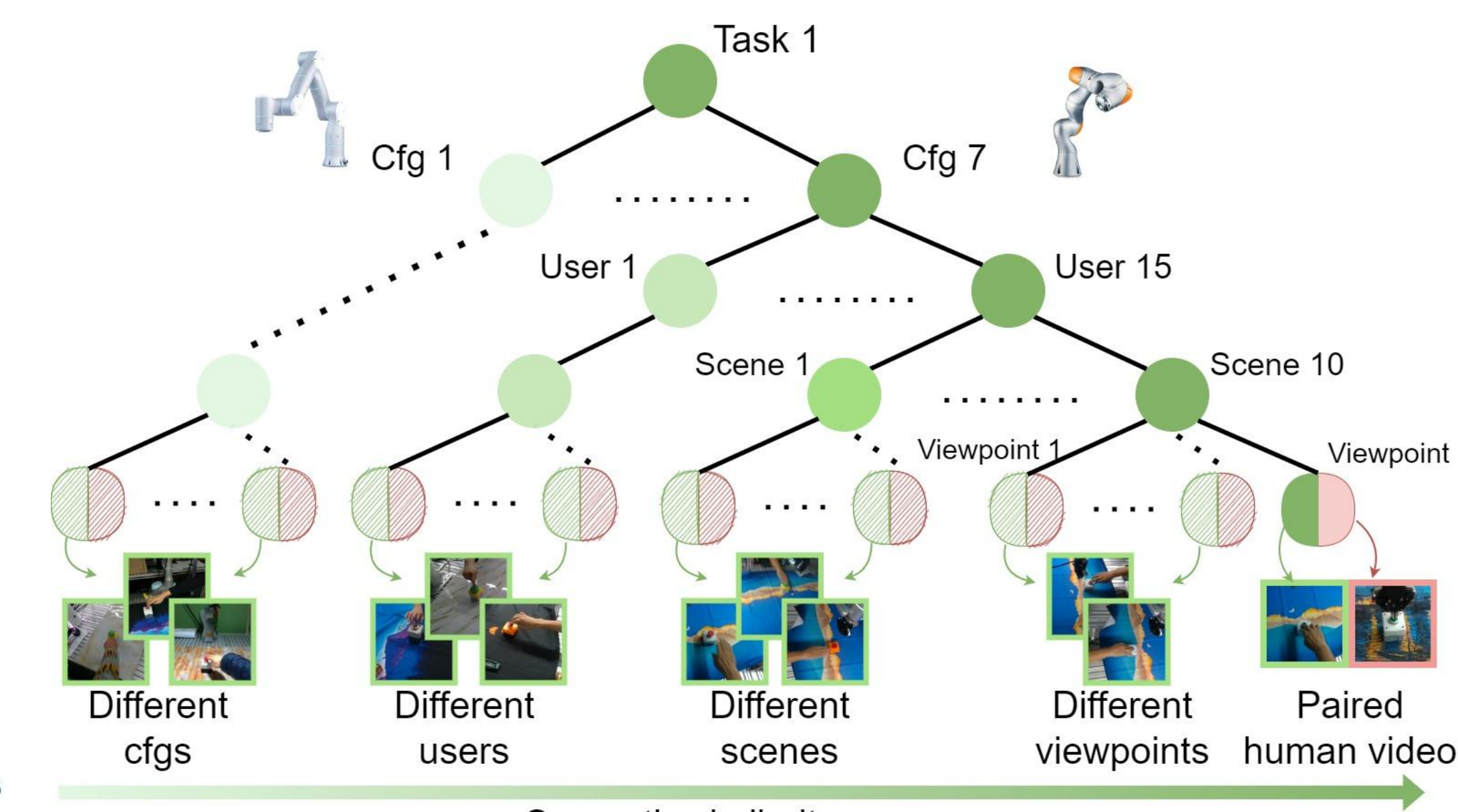

Fig.5: Example of data hierarchy. We can pair a robot manipulation sequence with different human demonstration videos.
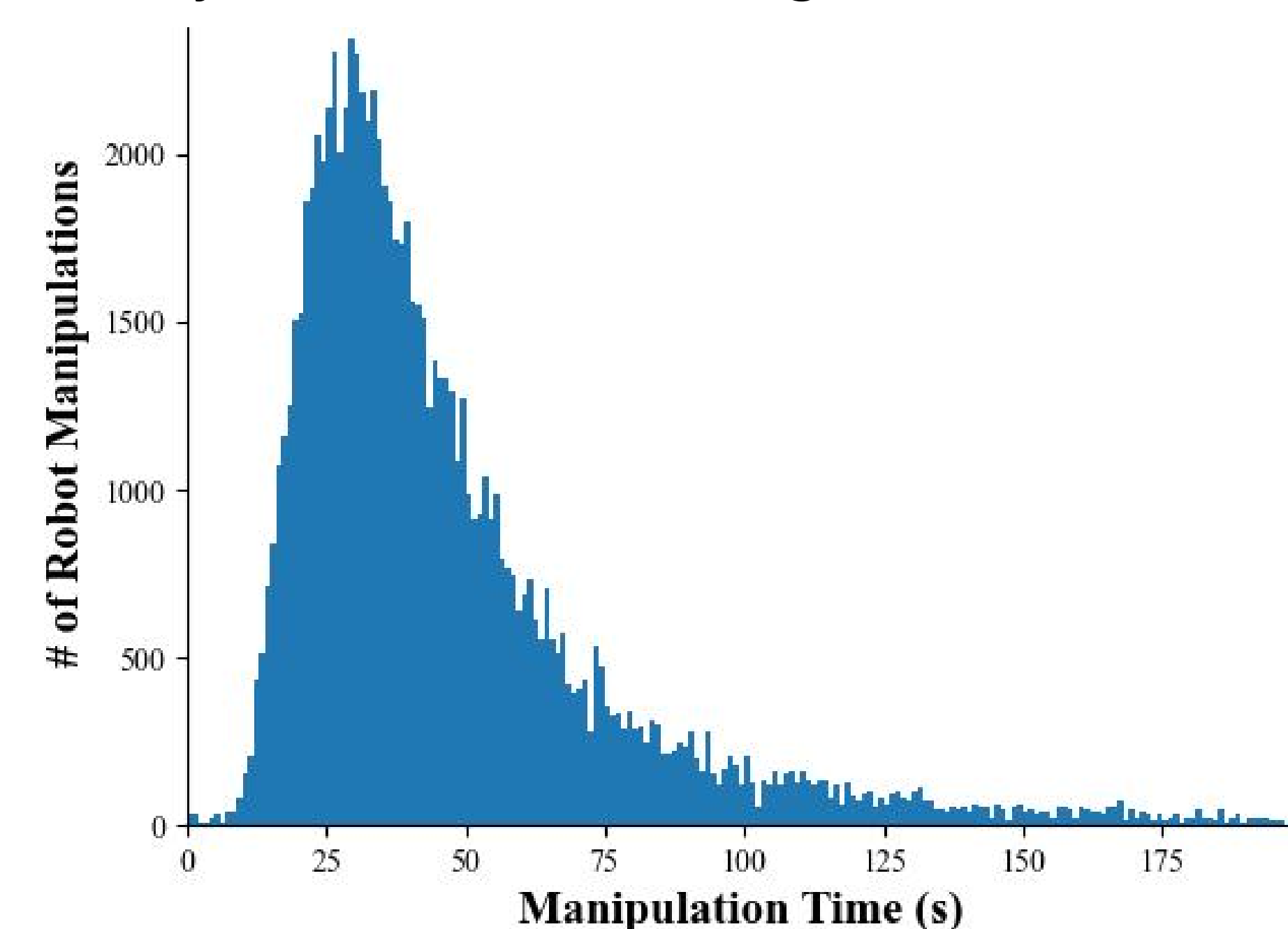

Fig.6: Statistics on the execution time of different robotic manipulations in our dataset.

Data sample

Task: *Insert plug into the socket*

Human Demonstration (multiple views, only one is demonstrated here)

Robot Manipulation (multiple views)

Fused point cloud and robot model