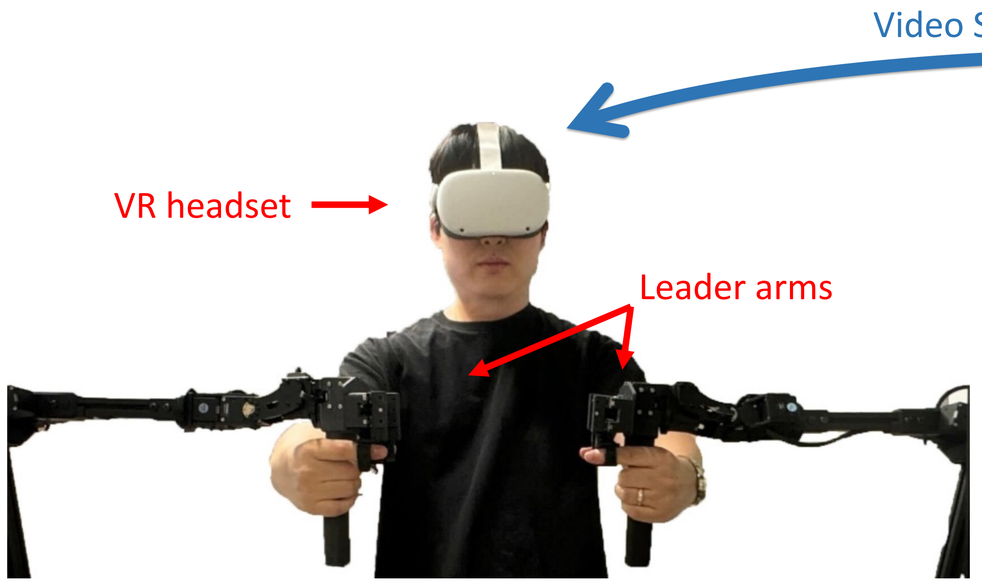
Active Vision Might Be All You Need: Exploring Active Vision in Bimanual Robotic Manipulation

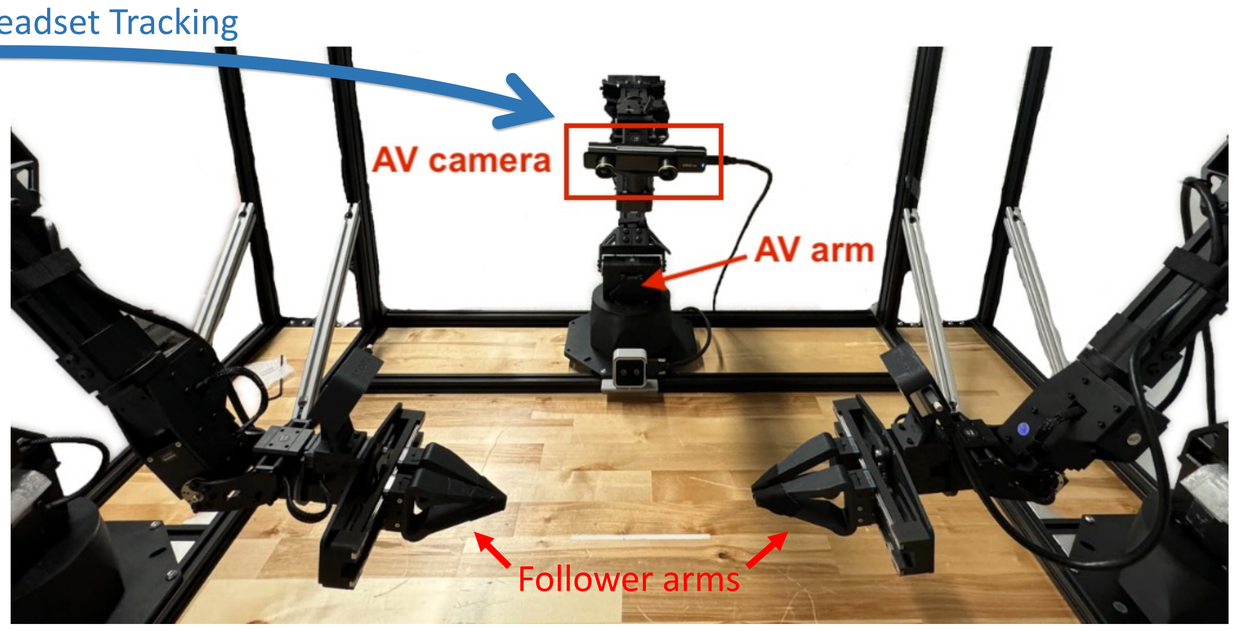
Ian Chuang*[1,2], Andrew Lee*[1], Dechen Gao[1], Mahdi Naddaf[1], Iman Soltani[1]

[1]University of California Davis, [2]University of California Berkeley *Equal Contribution

How to overcome **occlusions** and **limited field of view** in **imitation learning?**

Introducing… **AV-ALOHA!** 👀



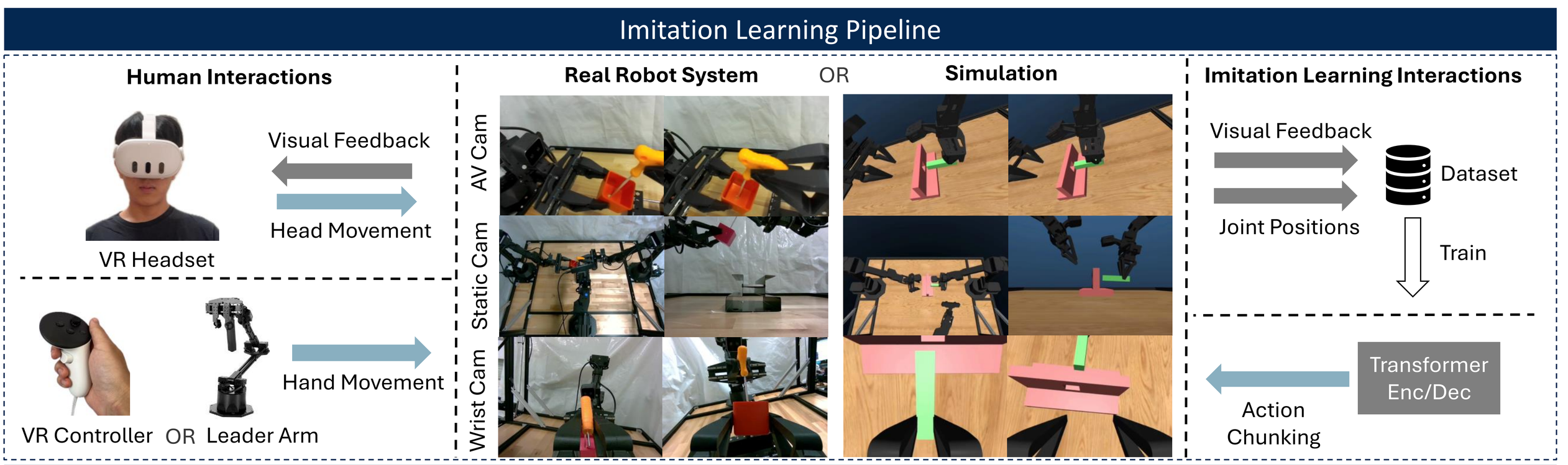


AV-ALOHA:

- A **bimanual robot system** with **7-DoF active vision (AV)** and **1st person VR control!**
- Overcome limitations of fixed or eye-in-hand cameras with **dynamic viewpoint control**.
- Platform to explore learning **human-guided active vision.**
- Open-source code, hardware, **simulation environment**, and datasets.

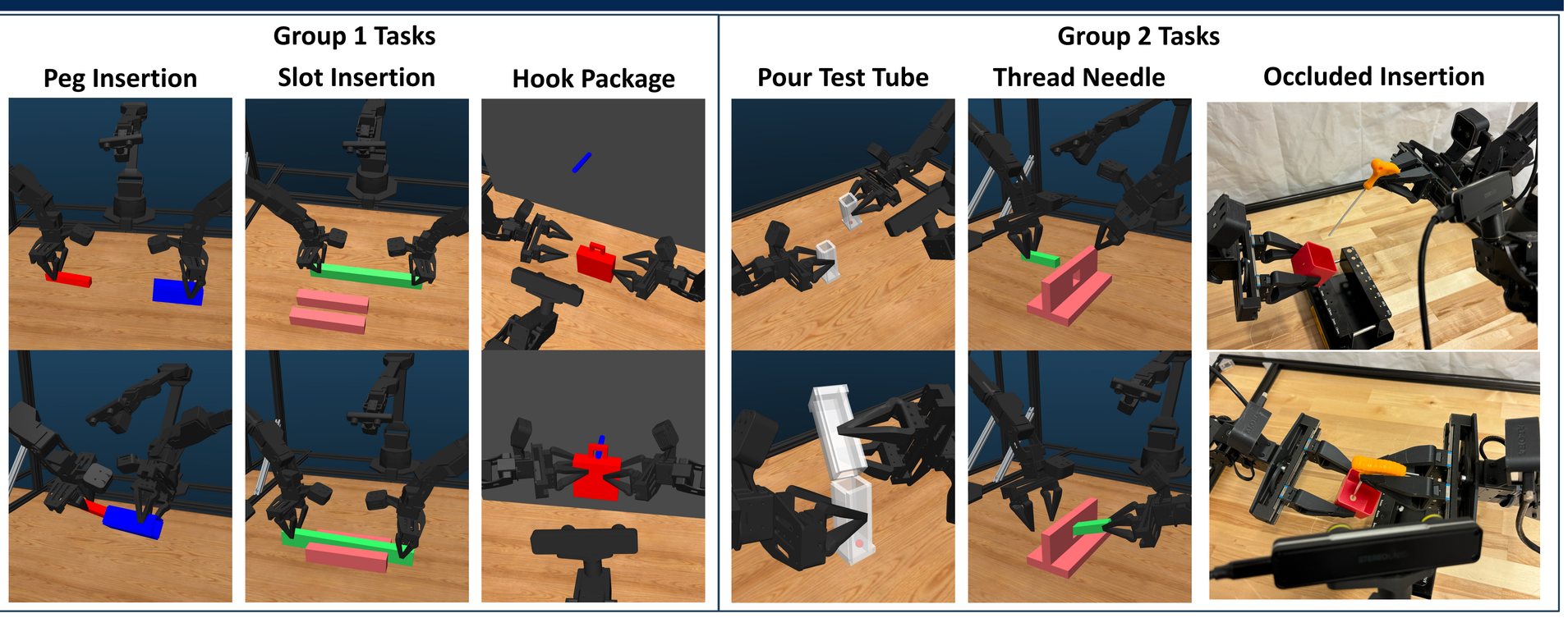
Control simulation with VR!

Imitation Learning Pipeline



Results

- We evaluate on 6 tasks. Some tasks are designed to not necessarily require active vision **(Group 1)** and other tasks have occlusions that might benefit from active vision **(Group 2)**.
- We collect data and evaluate **ACT Policy** across different camera configurations:
  - **Active Vision (AV)**
  - **Static** cameras
  - **Wrist** eye-in-hand cams
- **Active vision improves performance** when there are occlusions or limited visibility!



Success Rates (%) of ACT Policy on Different Tasks and Camera Configurations

	Group 1						Group 2					
	Peg Insertion		Slot Insertion		Hook Package		Pour Test Tube		Thread Needle		Occluded Insertion	
	Grasp	Insert	Grasp	Insert	Grasp	Hook	Grasp	Pour	Grasp	Thread	Grasp	Insert
AV	74	42	88	50	100	22	66	**14**	98	**52**	60	20
AV + Static	84	46	100	62	100	34	50	10	98	26	20	0
AV + Wrist	82	34	96	44	100	22	70	**14**	92	**52**	95	**30**
AV + Static + Wrist	78	36	100	36	100	24	36	8	90	40	40	5
Static	84	**48**	98	66	100	**44**	44	8	88	30	85	20
Static + Wrist	88	40	100	**78**	100	30	46	6	38	22	100	15
Wrist	84	42	98	44	92	8	44	10	94	44	60	15

Project Page