Verification of Deep Neural Networks in Control Systems

Changliu Liu

Assistant Professor

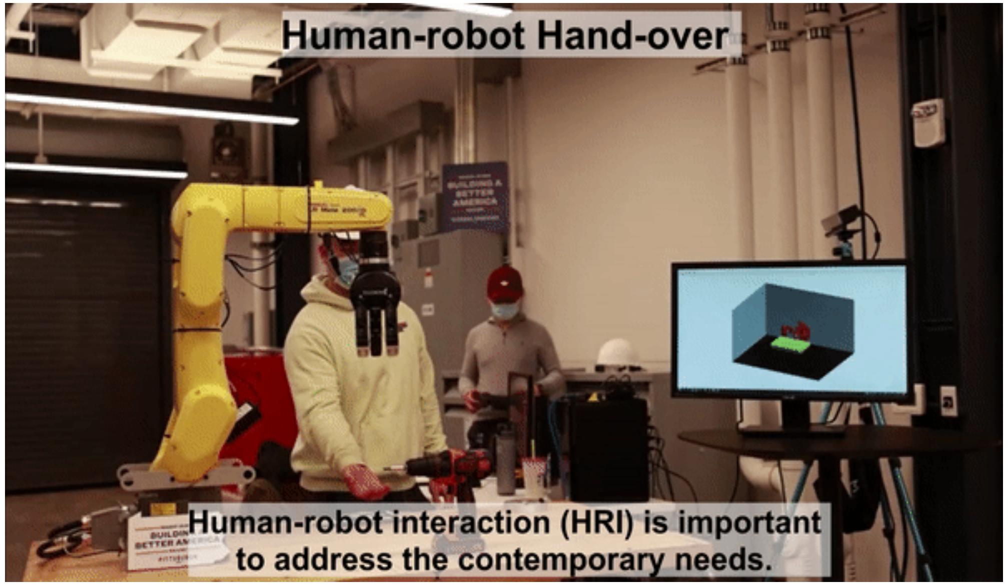
Robotics Institute

Carnegie Mellon University

Carnegie Mellon University

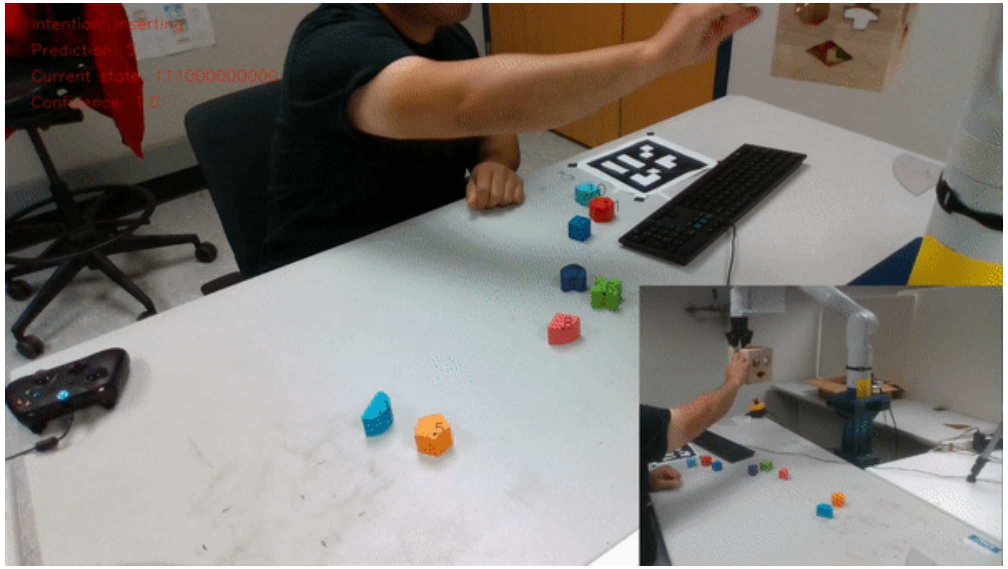
INTELLIGENT CONTROL LAB

Collaborative Handling

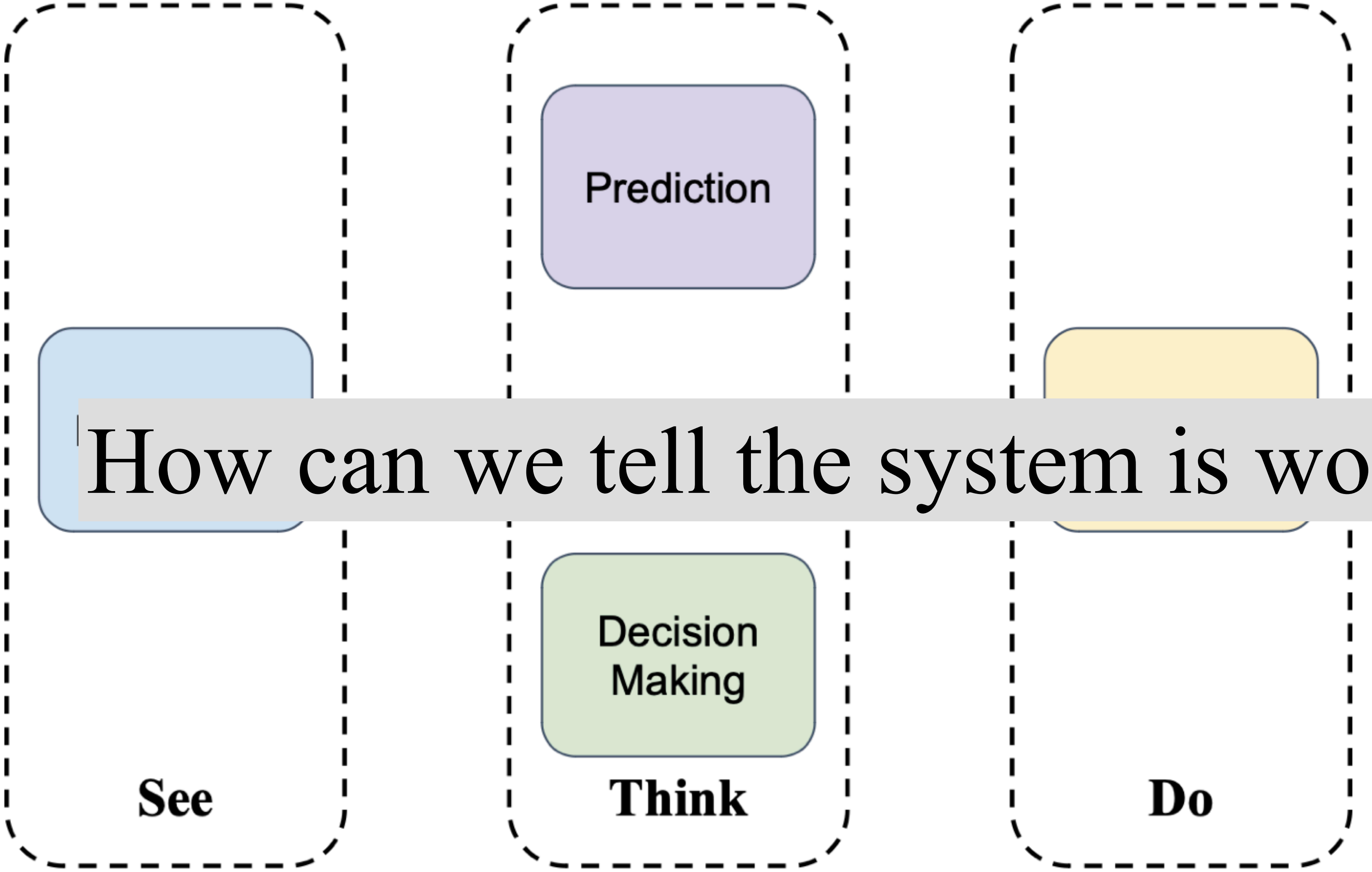

Human-robot Hand-over
Human-robot interaction (HRI) is important
to address the contemporary needs.

Handover

Learning from Demonstration

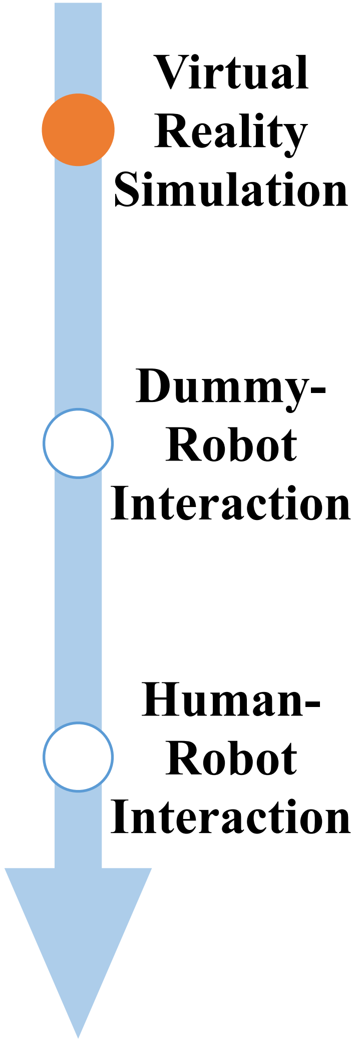


Co-assembly



How can we tell the system is working well?

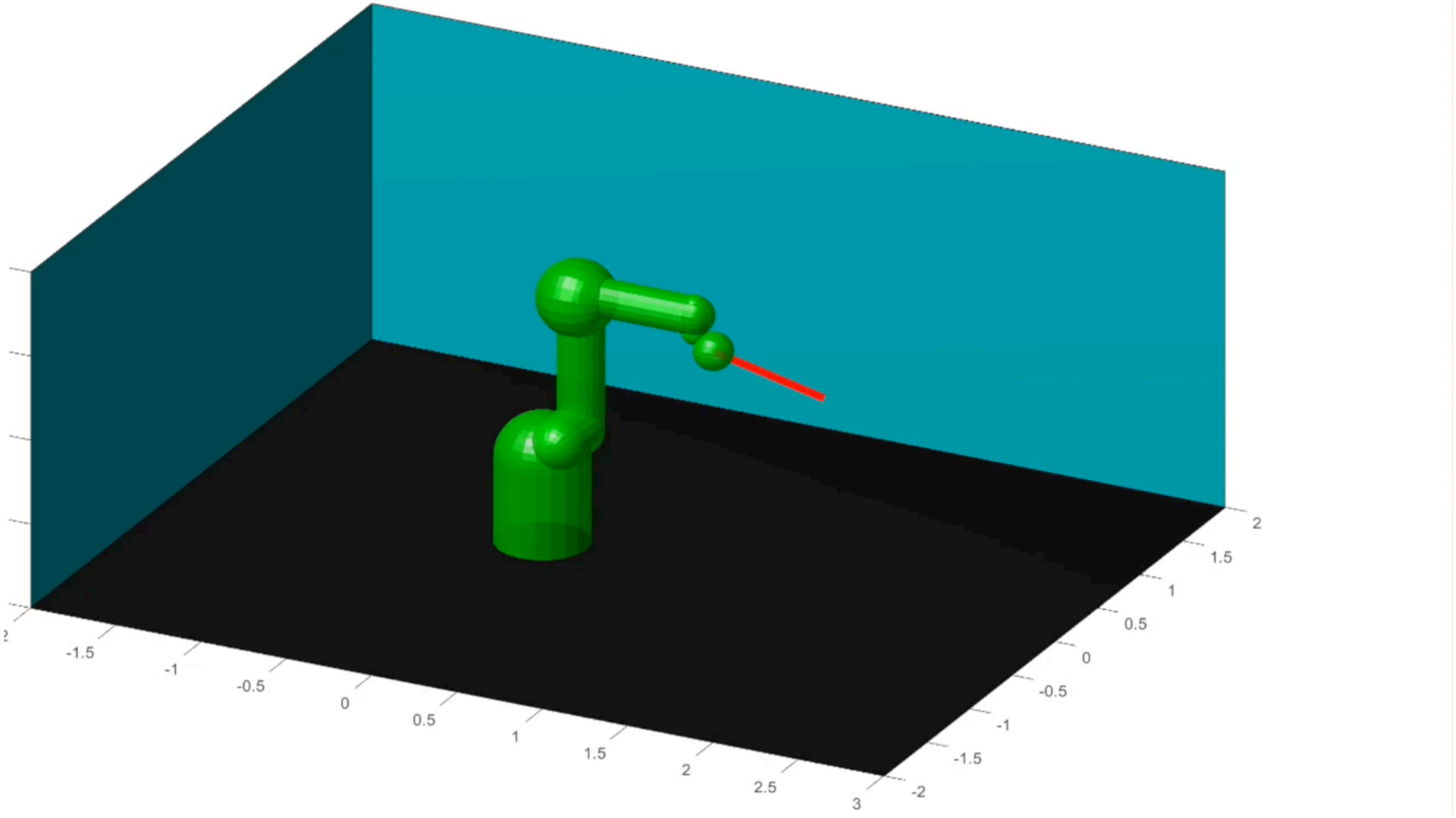
Evaluation of the Robot Control System



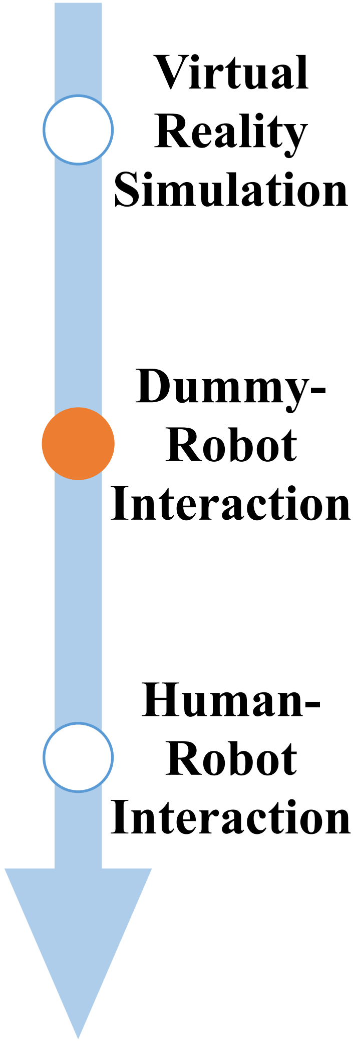
Virtual Reality Simulation

Dummy-Robot Interaction

Human-Robot Interaction



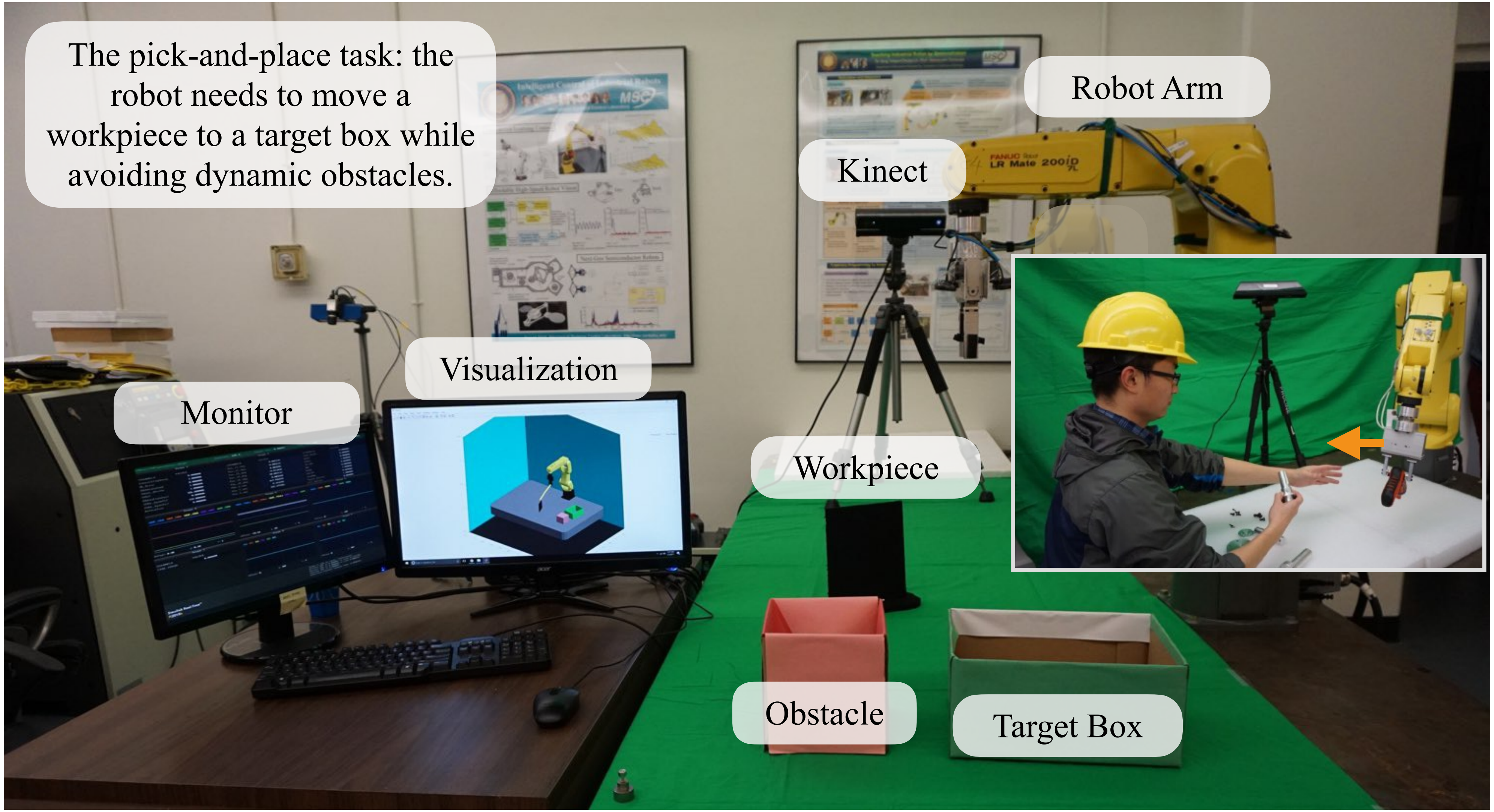
Evaluation of the Robot Control System



Virtual Reality Simulation

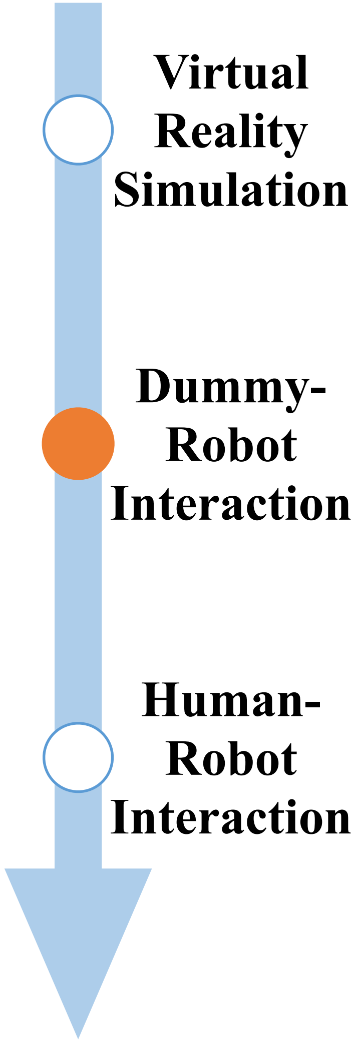
Dummy-Robot Interaction

Human-Robot Interaction



The pick-and-place task: the robot needs to move a workpiece to a target box while avoiding dynamic obstacles.

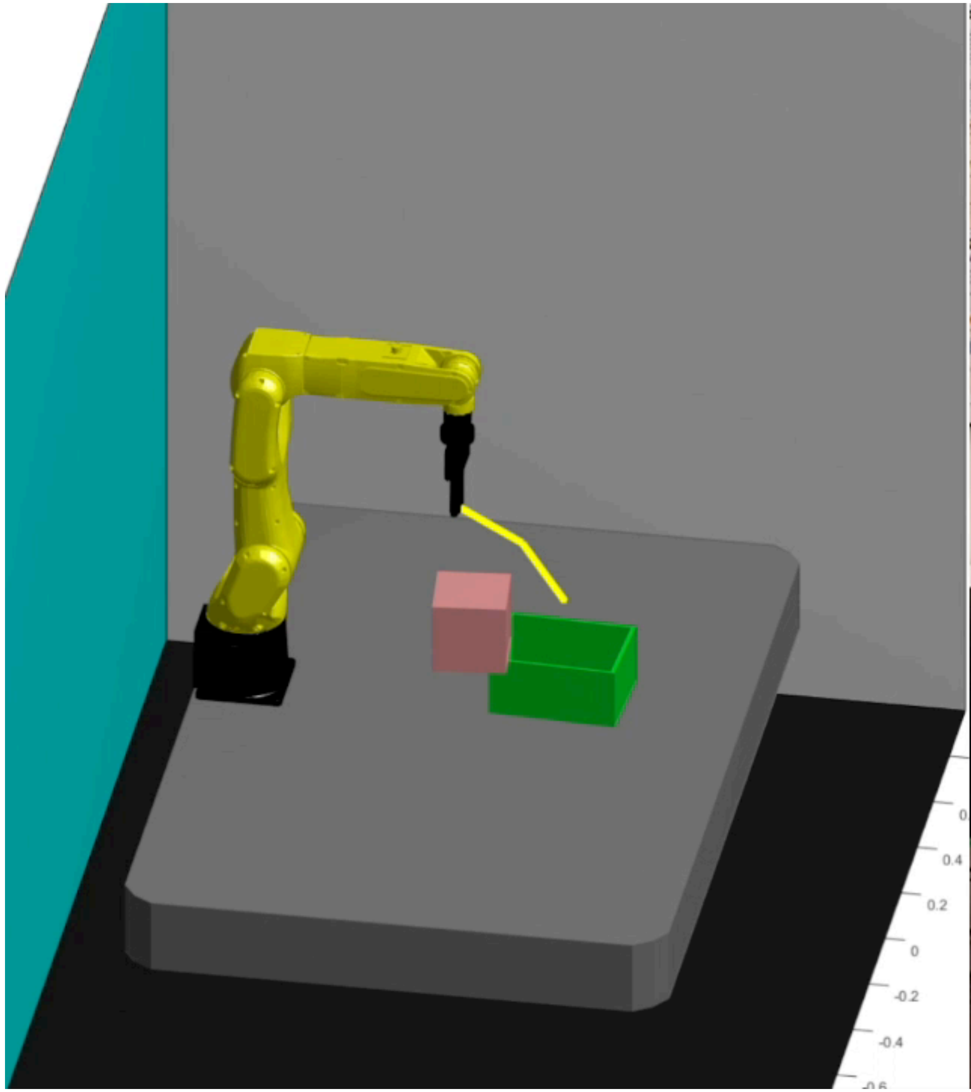
Evaluation of the Robot Control System

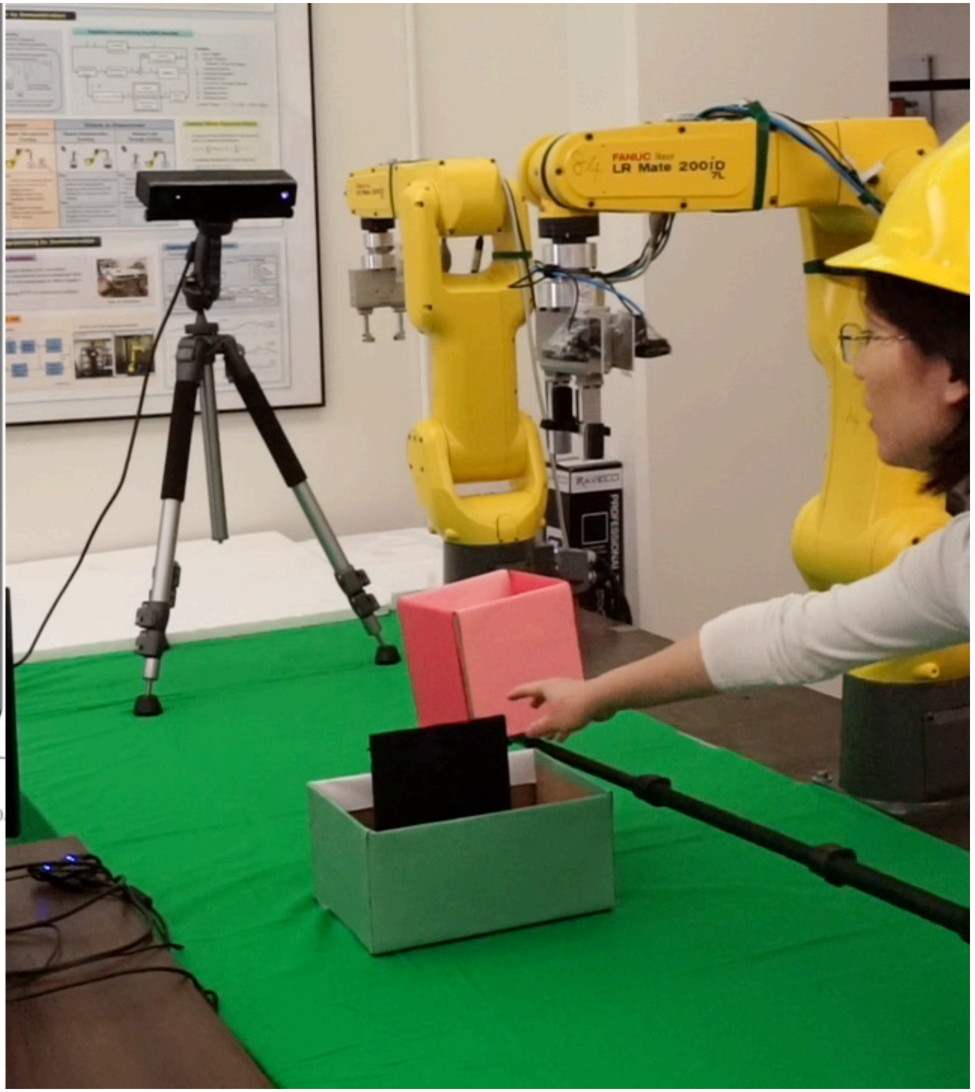


○ Virtual Reality Simulation

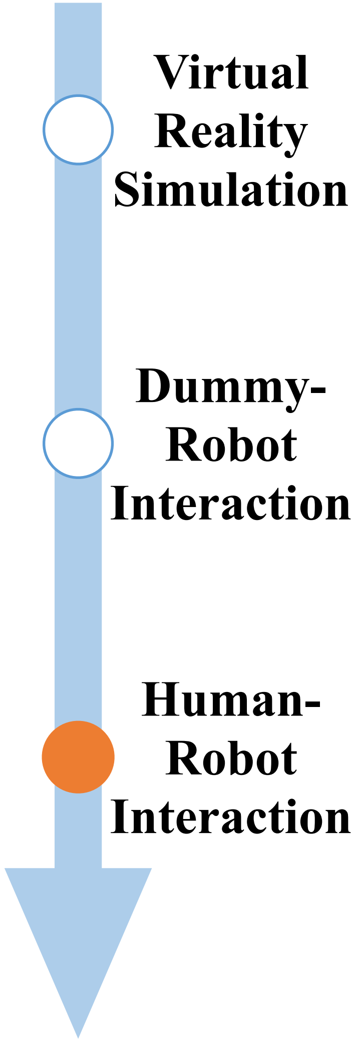
● Dummy-Robot Interaction

○ Human-Robot Interaction





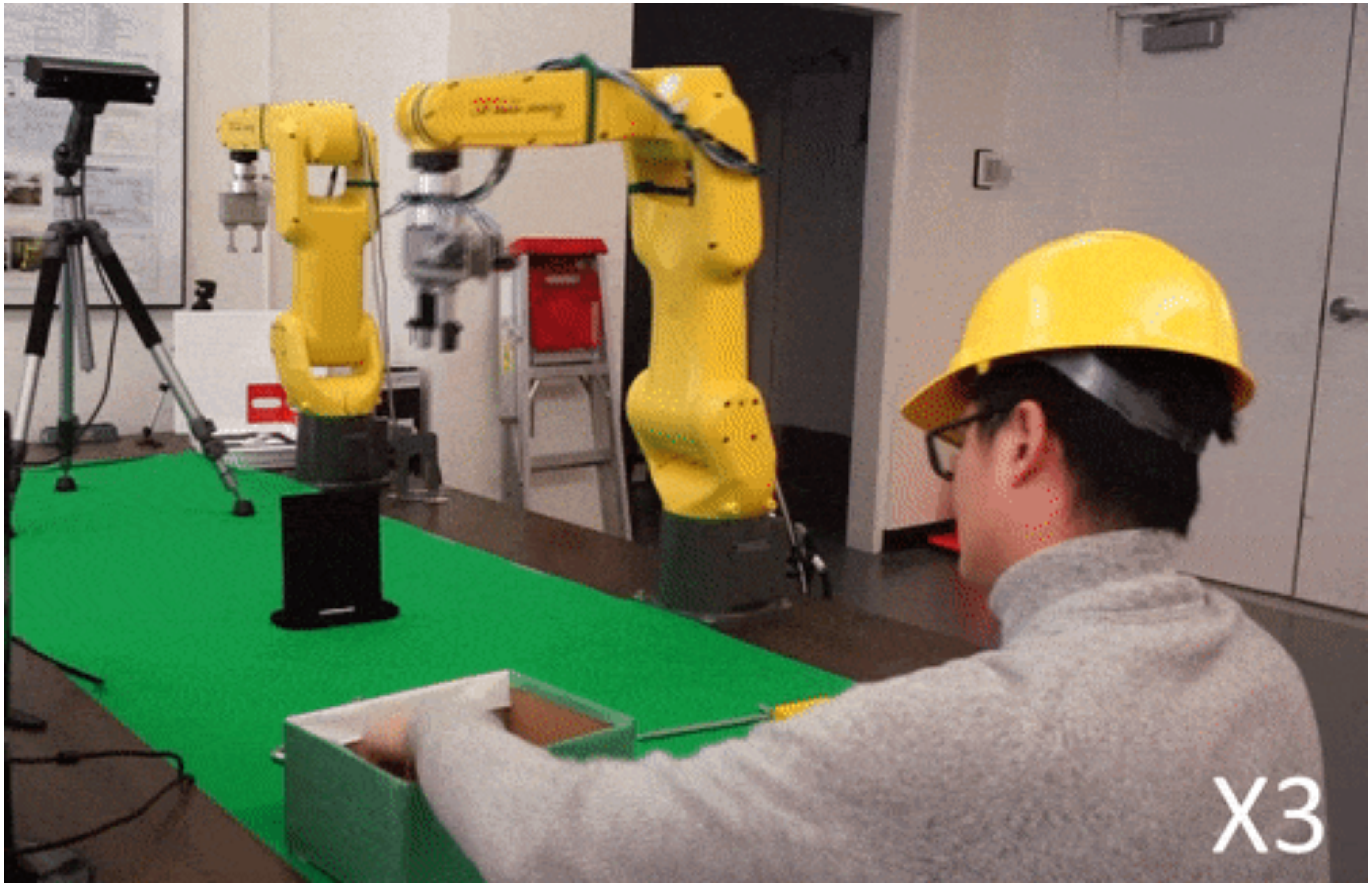
Evaluation of the Robot Control System



○ Virtual Reality Simulation

○ Dummy-Robot Interaction

● Human-Robot Interaction



Testing vs Verification

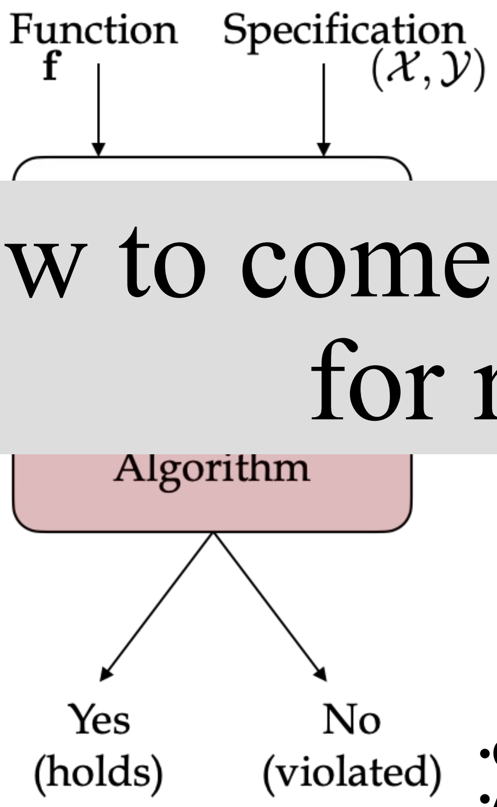
- Testing: sample based evaluation
  - Modular testing
  - System testing
  - Human study

Time consuming

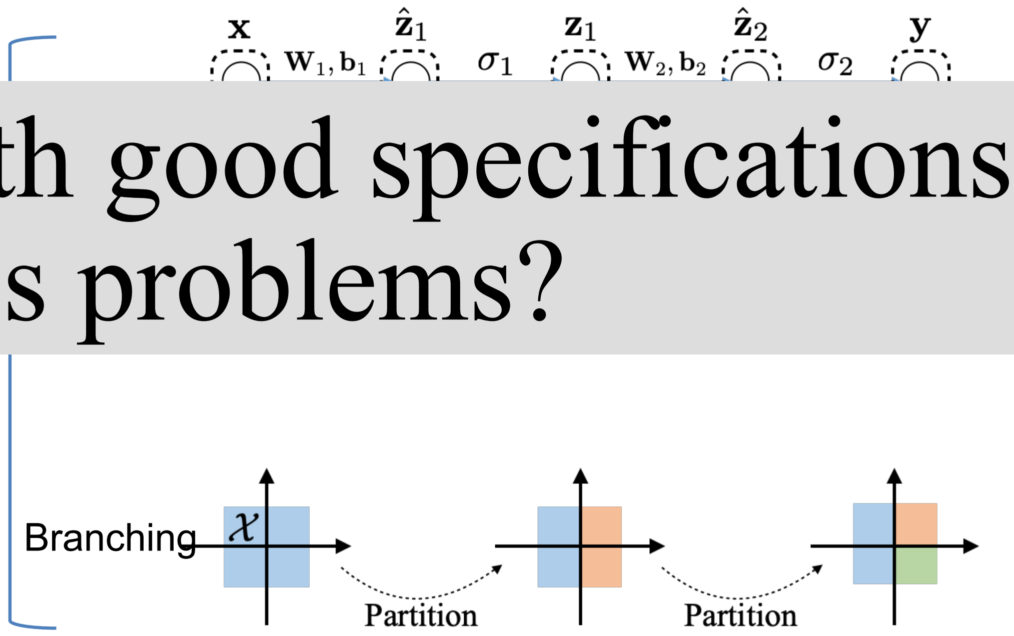
- Verification: mathematical proofs
  - System/Modular verification through Lyapunov analysis
  - Neural network verification

Neural Network Verification Tools

How to come up with good specifications for robotics problems?



- Counter example
- Adversarial input bound
- Output reachable set

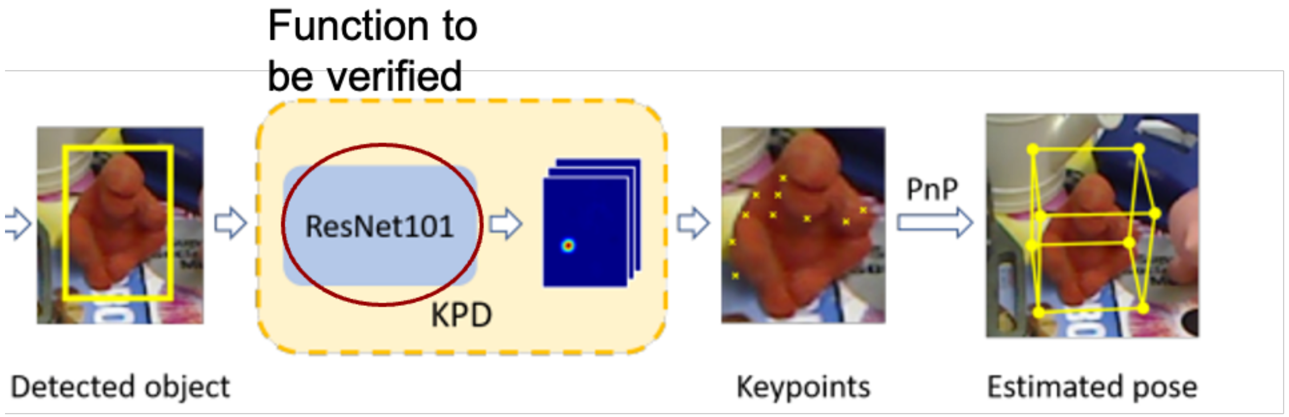


https://github.com/sisl/NeuralVerification.jl/

C. Liu, T. Arnon, C. Lazarus, C. Strong, C. Barrett, and M. Kochenderfer, "Algorithms for verifying deep neural networks," Foundations and Trends in Optimization, 2021.

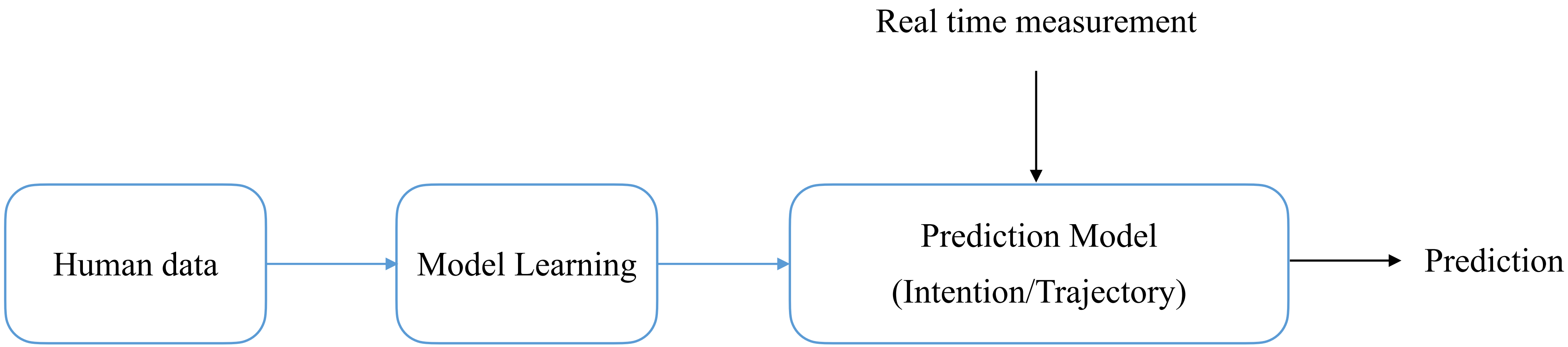
Case 1: Object Pose Estimation

- Existing approach: robustness against Lp disturbances on sampled images
- More practical specifications:
  - Whether the pose estimation error is bounded under 1) camera movement; 2) lighting changes, etc.

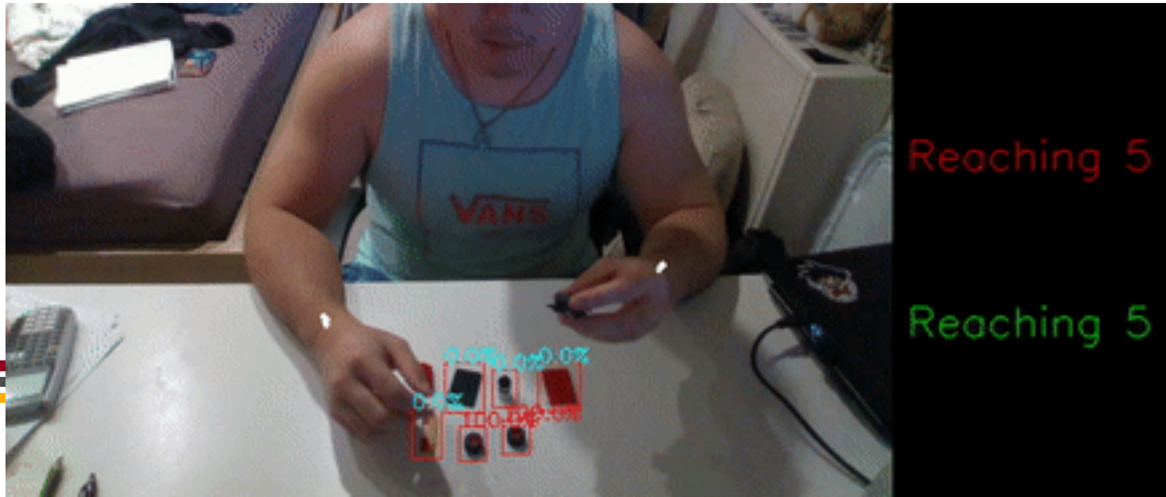


Hu, Hanjiang, Changliu Liu, and Ding Zhao. "Robustness Verification for Perception Models against Camera Motion Perturbations." *ICML Workshop on Formal Verification of Machine Learning (WFVML)*. 2023.

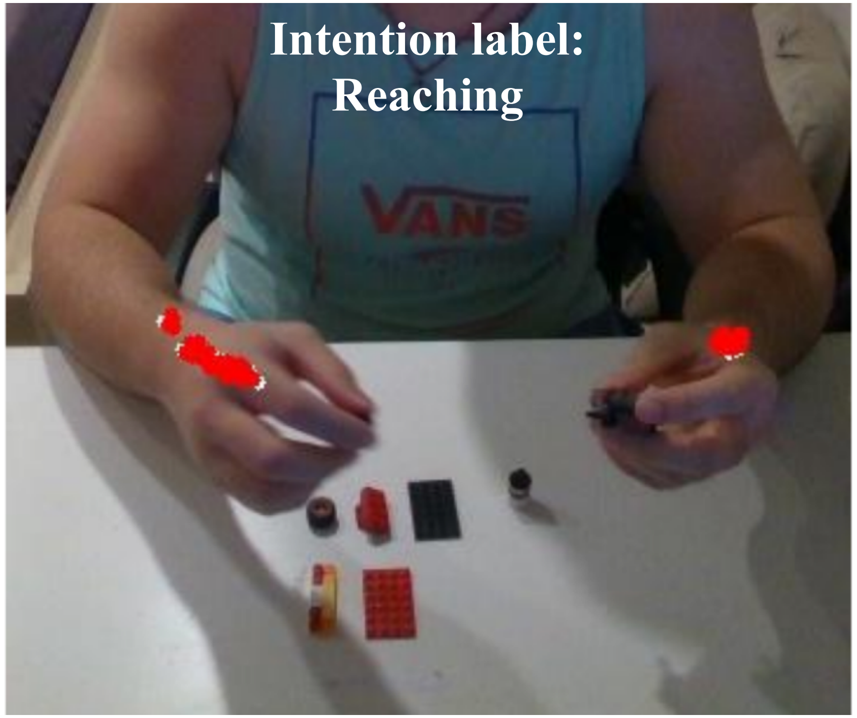
Case 2: Human Prediction Model



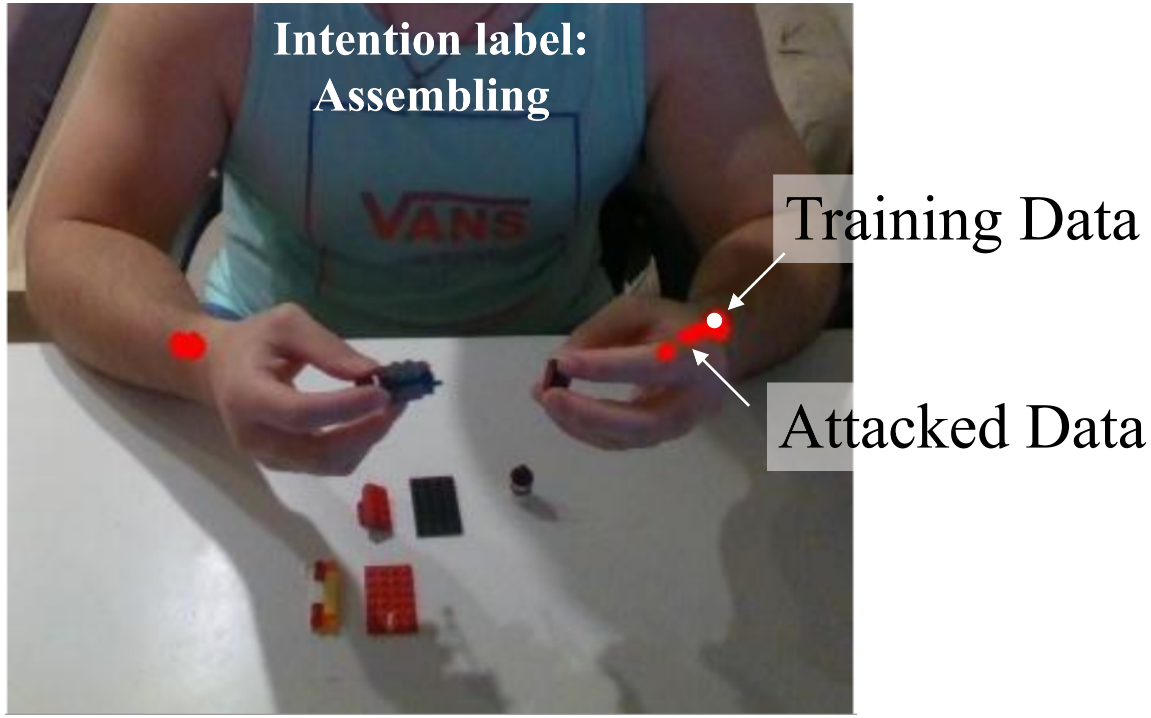
- Should the model be Lp robust to every human trajectory? (Returning the same intention prediction given small perturbations on the human trajectory)



Case 2: Human Prediction Model



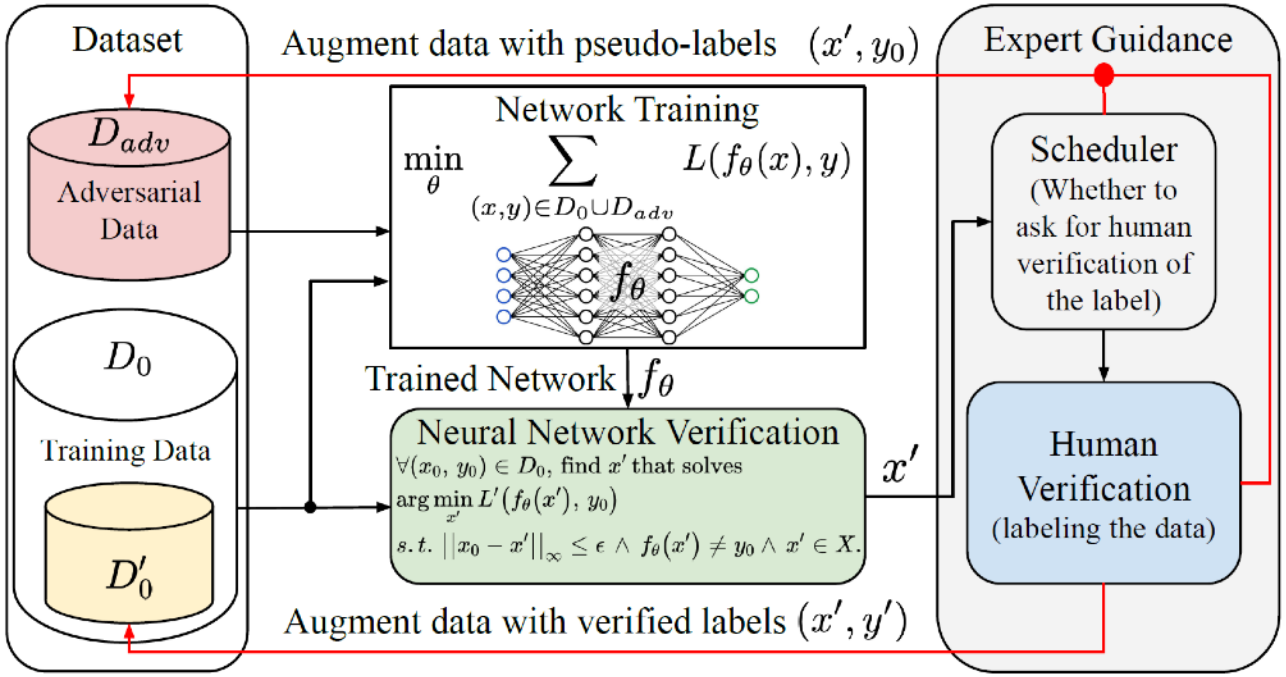
True Adversary



False Adversary

R. Liu, and C. Liu, "IADA: Iterative Adversarial Data Augmentation Using Formal Verification and Expert Guidance," ICML Workshop on Human in the loop learning, 2021.

Case 2: Human Prediction Model


Dataset
D_{adv}
Adversarial Data
D_0
Training Data
D_0'
Augment data with pseudo-labels (x', y_0)
Network Training
$\min_{\theta} \sum_{(x,y)\in D_0 \cup D_{adv}} L(f_\theta(x), y)$
f_θ
Trained Network f_θ
Neural Network Verification
$\forall (x_0, y_0) \in D_0$, find x' that solves
$\arg\min_{x'} L'(f_\theta(x'), y_0)$
$s.t. ||x_0 - x'||_\infty \le \epsilon \wedge f_\theta(x') \ne y_0 \wedge x' \in X.$
x'
Expert Guidance
Scheduler (Whether to ask for human verification of the label)
Human Verification (labeling the data)
Augment data with verified labels (x', y')

	Supervised Training	Adversarial Training	Iterative Adversarial Data Augmentation
Epochs: 500	81.99%	**82.53%**	**82.53%**
Epochs: 1k	**85.92%**	81.99%	85.48%
Epochs: 2k	85.26%	84.06%	**88.75%**
Epochs: 3k	82.97%	82.31%	**89.52%**
Epochs: 4k	82.86%	68.56%	**90.83%**
Epochs: 5k	81.55%	58.52%	**92.03%**

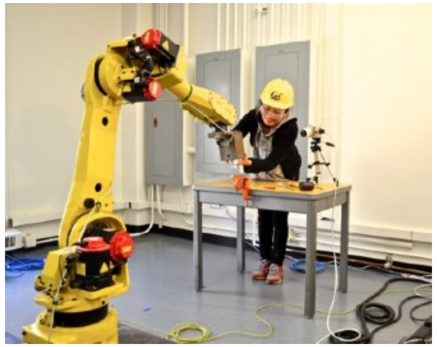
The learned model does not try to “robustify” every data point, but tries to fit the decision boundary well.

R. Liu, and C. Liu, “IADA: Iterative Adversarial Data Augmentation Using Formal Verification and Expert Guidance,” ICML Workshop on Human in the loop learning, 2021.

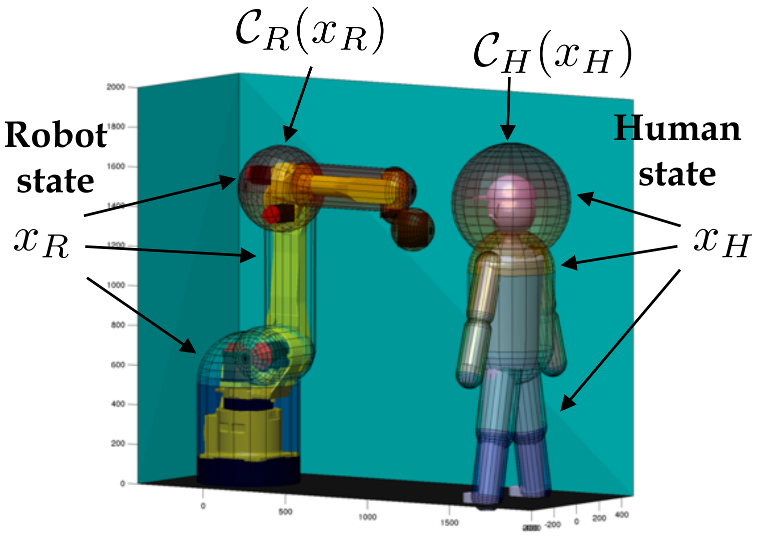
Case 3: Robot Policy

- For human-robot collision avoidance

The requirements on different states are different

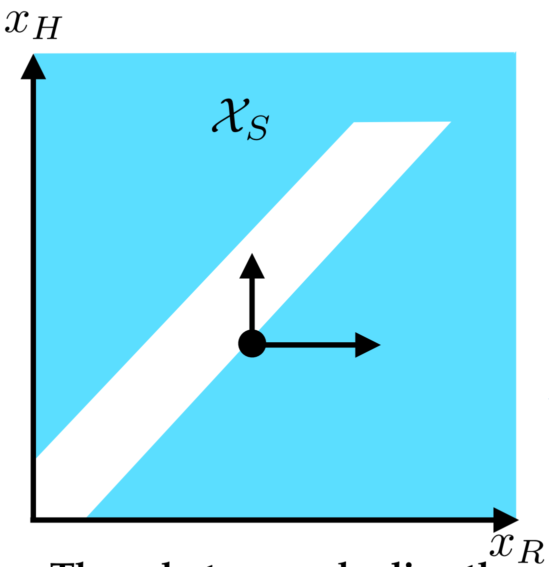


Real World Situation

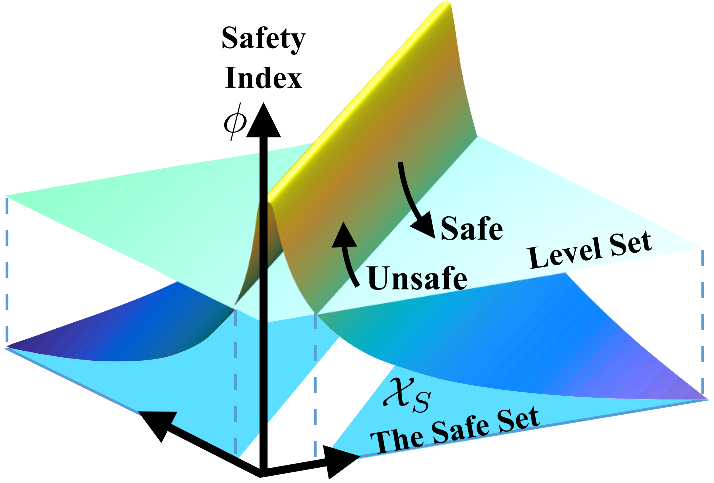


Computation model in Cartesian space

$$d(\mathcal{C}_R(x_R), \mathcal{C}_H(x_H)) \geq \gamma$$



The robot can only directly affect its own state.



C. Liu, and M. Tomizuka, "Algorithmic safety measures for intelligent industrial co-robots", in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 3095 – 3102.

Remarks

- What to verify highly depends on the system-under-test
- There exist gaps between the problems that verification algorithms can solve and the problems that need to be verified.
    - Example: (local) robustness to sampled panda images versus (global) robustness to all panda images
- Looking into real applications will offer more insights on what verification tools need to be developed

Thank You!

NSF

Ford

amazon

SIEMENS

BOEING

FANUC

LOCKHEED MARTIN

YASKAWA

ARM INSTITUTE ADVANCED ROBOTICS FOR MANUFACTURING

EFORT

INTELLIGENT CONTROL LAB

Students