Adaptable Human Intention and Trajectory Prediction for Human-Robot Collaboration

Abulikemu Abuduweili,
Siyan Li,
Changliu Liu

AI-HRI 2019

Carnegie Mellon University

INTELLIGENT CONTROL LAB

Introduction

INTENTION AND TRAJECTORY PREDICTION

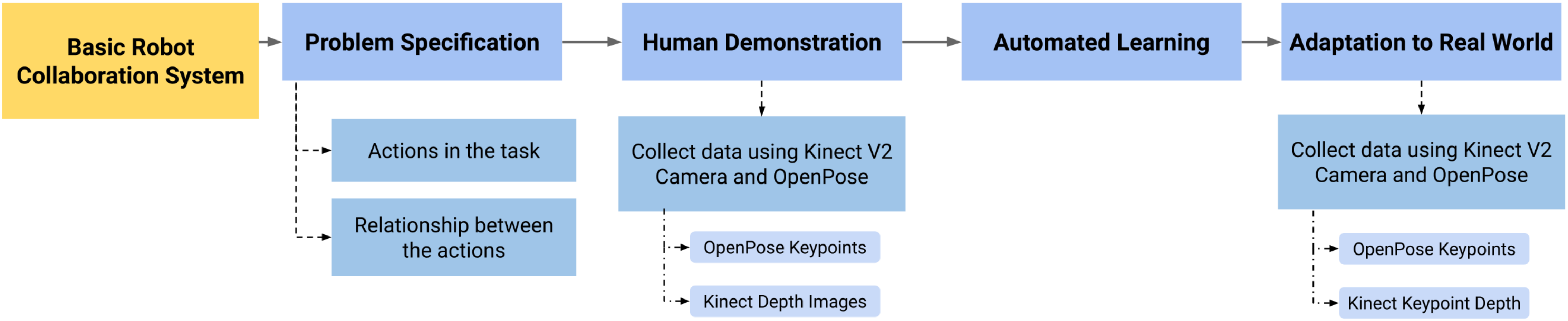
- Usually separated, would like to combine
- Often require wearable devices (Pistohl et al. 2008, Wang et al. 2018)
- Computer-vision-based methods

ONLINE ADAPTATION

- RLS-PAA to adapt the last linear layer of a fully connected network (Si, Wei, and Liu 2019)
- Adapting non-linear layers in more complex networks

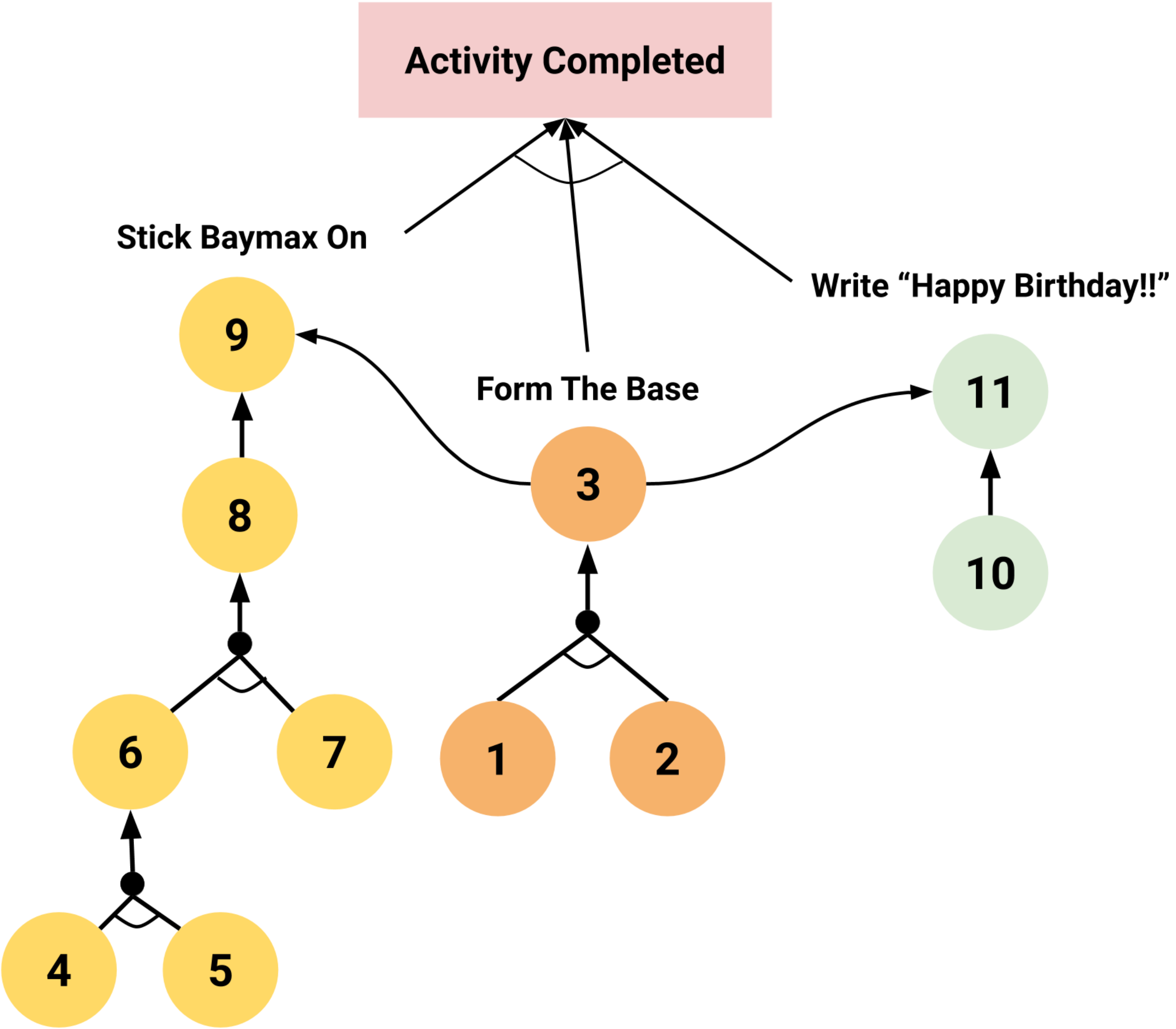
1.
PIPELINE STRUCTURE

A general overview of our framework


Basic Robot Collaboration System
Problem Specification
Human Demonstration
Automated Learning
Adaptation to Real World
Actions in the task
Relationship between the actions
Collect data using Kinect V2 Camera and OpenPose
OpenPose Keypoints
Kinect Depth Images
Collect data using Kinect V2 Camera and OpenPose
OpenPose Keypoints
Kinect Keypoint Depth

Problem Specification

- A task consisting of atomic individual actions
- Must be able to be represented using an and-or graph
- This step must be performed manually at this stage



Form The Base
1. Take the Card
2. Take the Red Sharpie
3. Draw Lines

Stick Baymax On

4. Take Baymax
5. Take Scissors
6. Cut Out Baymax
7. Take Glue Stick
8. Put Glue on Baymax
9. Stick Baymax on the Card

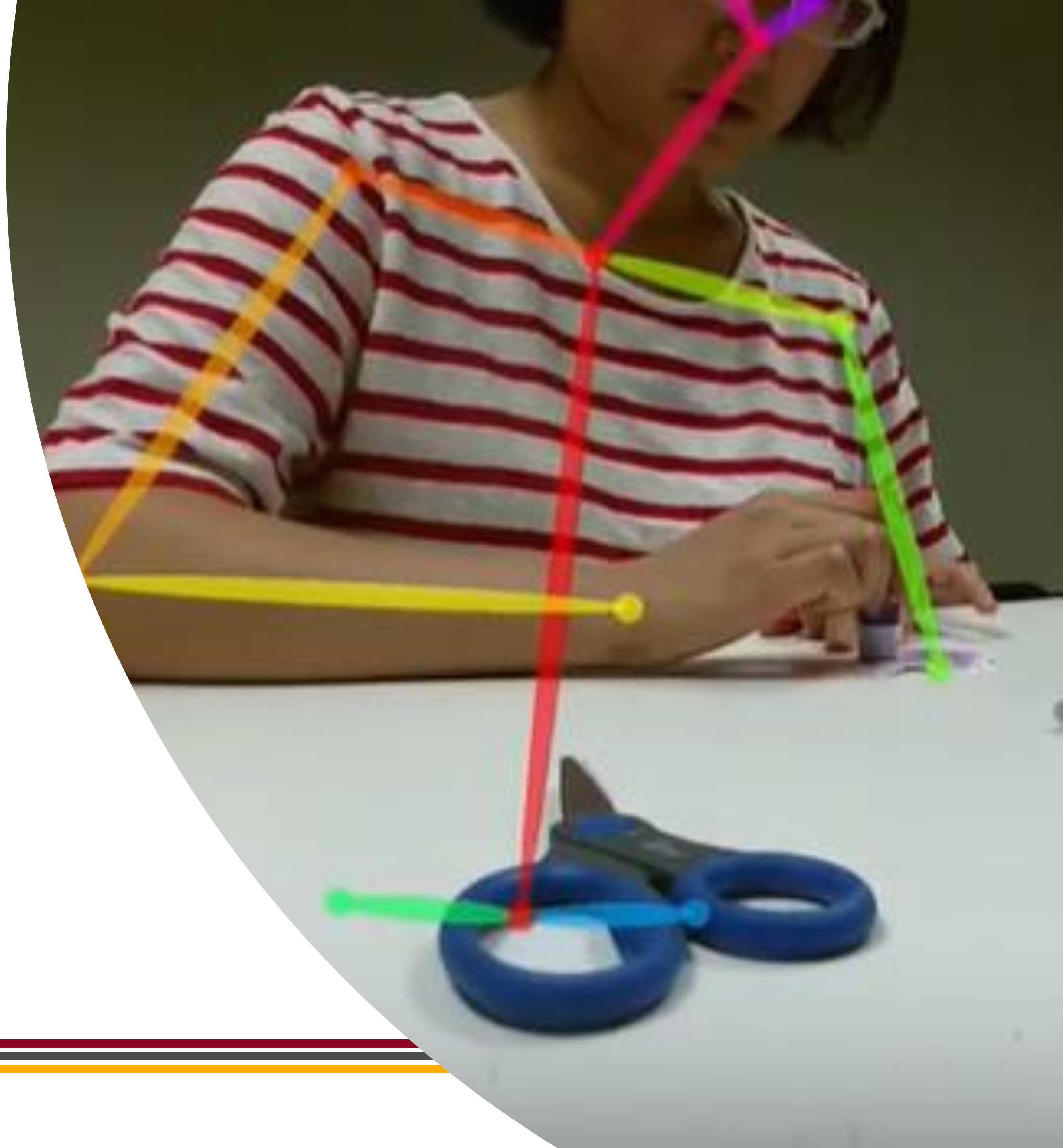
Write "Happy Birthday!"

10. Take the Black Sharpie
11. Write Words

12. Take Back

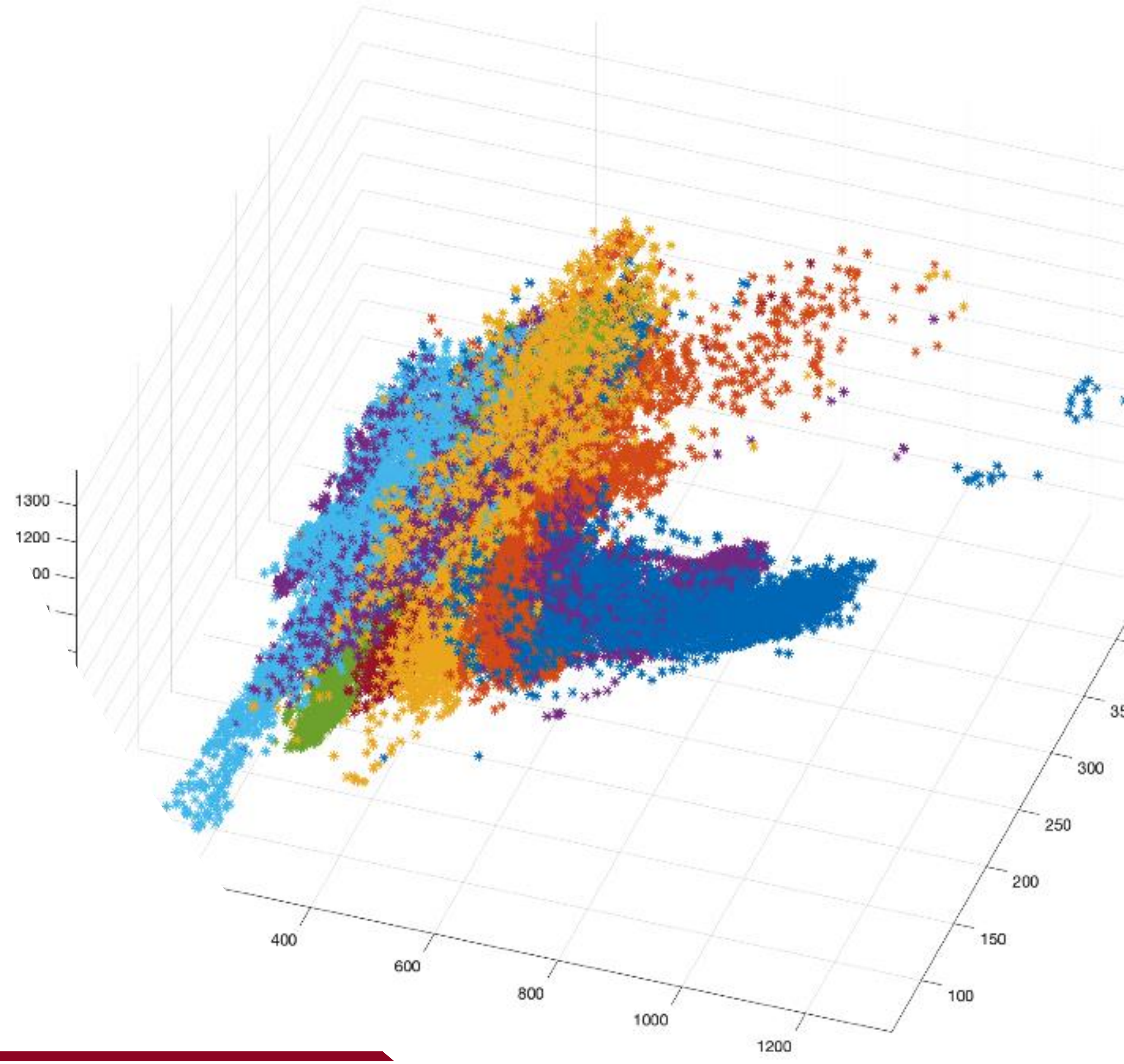
Human Demonstration

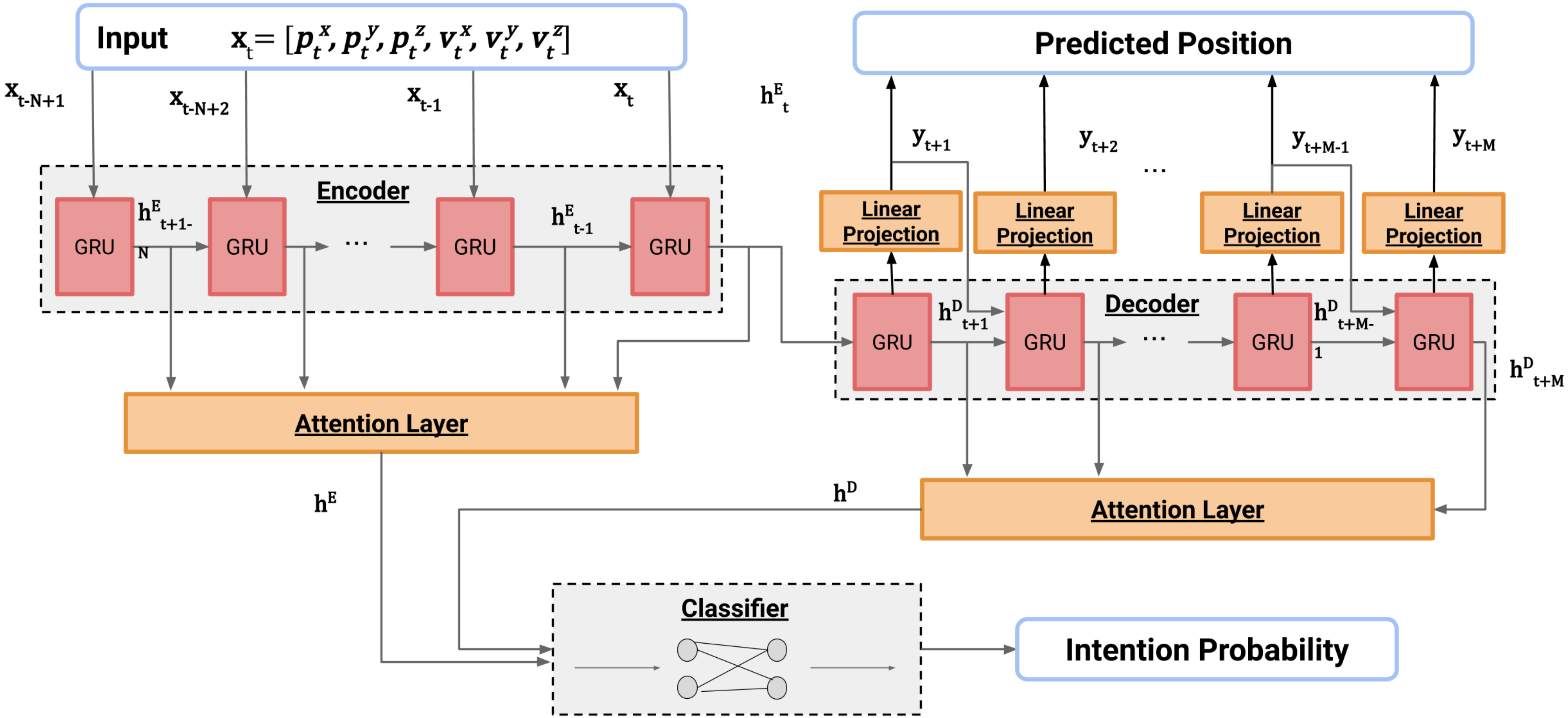
- Atomic actions must be repeated to the system
- Gathering data for the neural network in the pipeline to learn features of the actions



Learning Trajectories and Intentions

- **Multi-task model**
  - Trajectory Prediction = **Encoder-Decoder Seq2Seq**
  - Intention Prediction = **Encoder-Decoder-Attention-Classifier**
- <u>Input:</u> x, y, z positions and velocities in x, y, z directions for the past N time steps
- <u>Output:</u> trajectory and intention prediction for the next M time steps

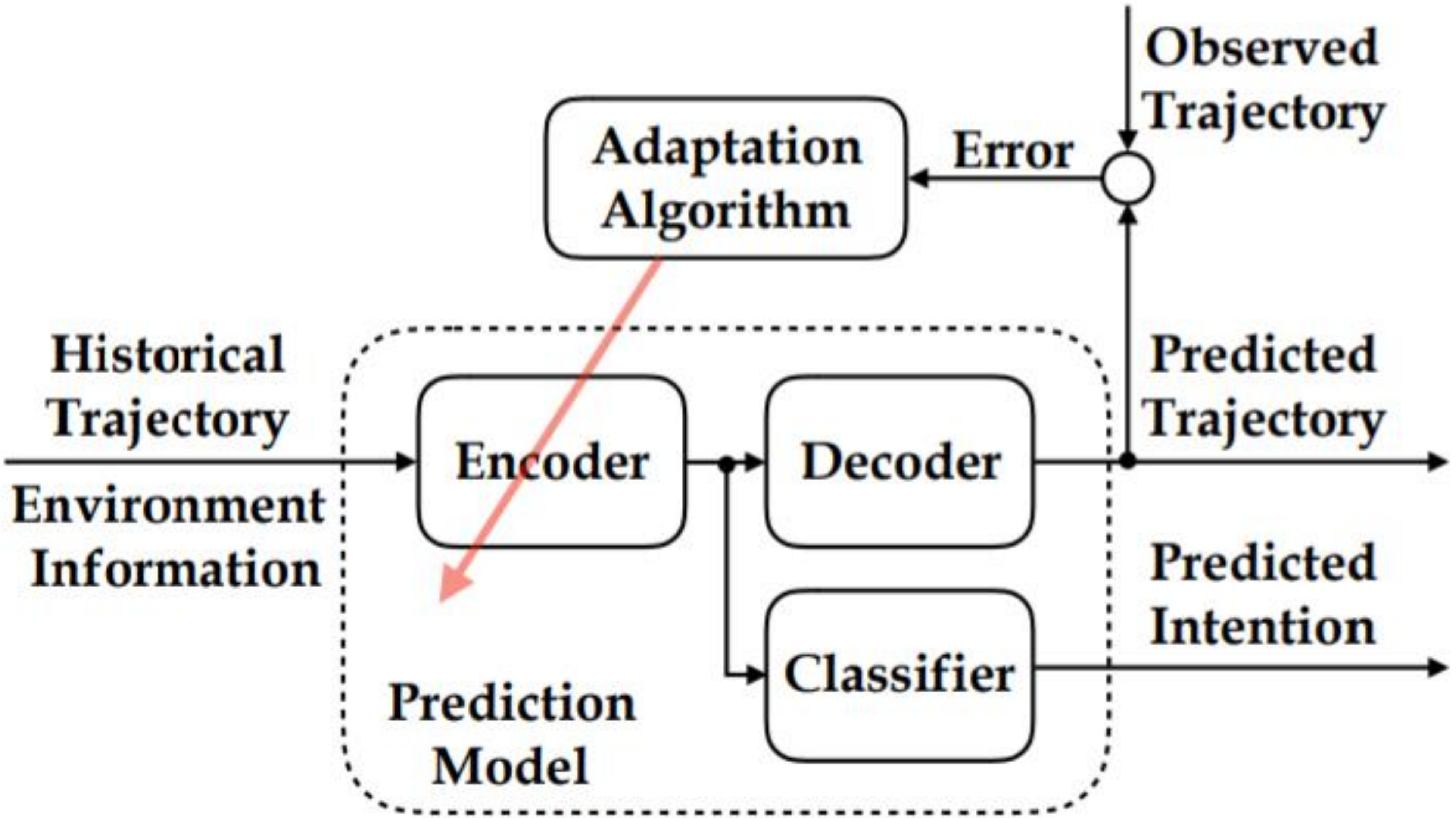


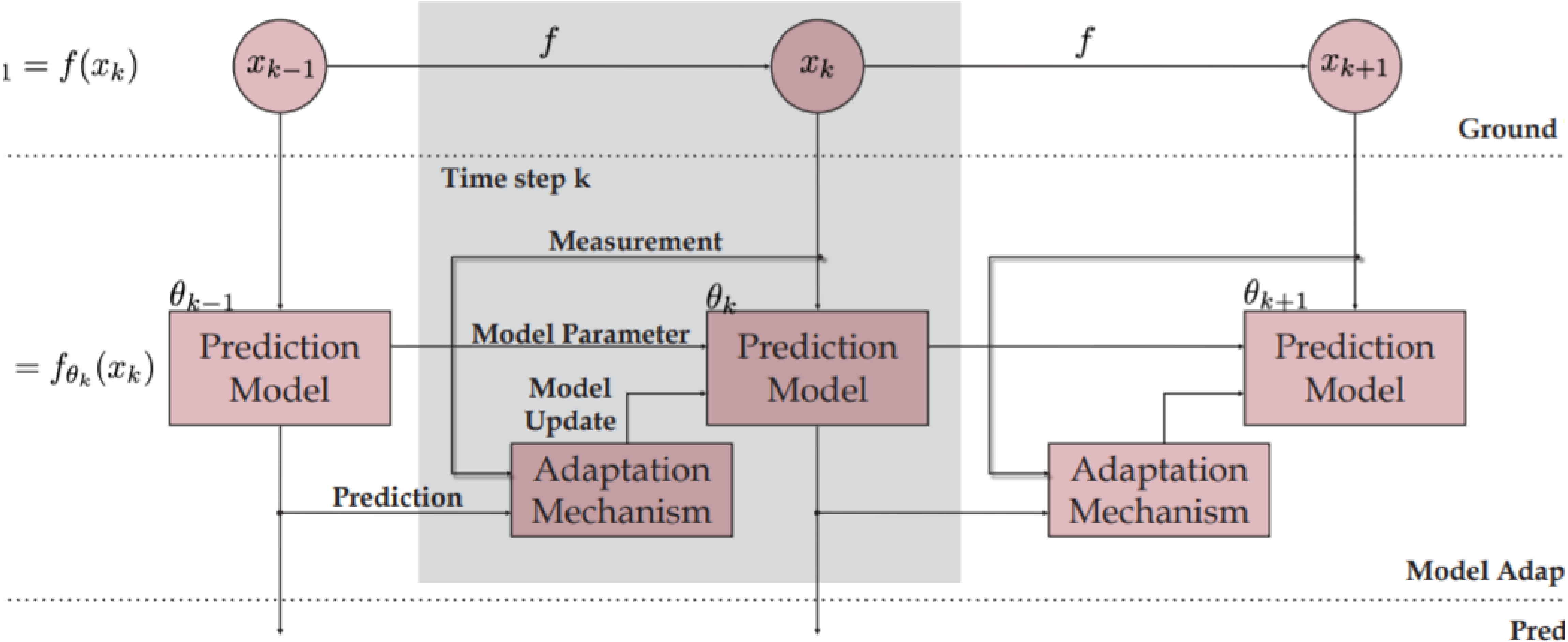


Input
$x_t = [p_t^x, p_t^y, p_t^z, v_t^x, v_t^y, v_t^z]$
x_{t-N+1}
x_{t-N+2}
x_{t-1}
x_t
h^E_t
Encoder
GRU
h^E_{t+1-N}
h^E_{t-1}
Attention Layer
h^E
Predicted Position
y_{t+1}
y_{t+2}
y_{t+M-1}
y_{t+M}
Linear Projection
Decoder
h^D_{t+1}
h^D_{t+M-1}
h^D_{t+M}
h^D
Classifier
Intention Probability

Adapting to Real-world Tasks

- **Non-linear Recursive Least Square Parameter Adaptation Algorithm (NRLS-PAA)**
- Model updated every time ground-truth is received
- Need to wait for the new ground truth


Observed Trajectory
Adaptation Algorithm
Error
Historical Trajectory
Environment Information
Encoder
Decoder
Classifier
Predicted Trajectory
Predicted Intention
Prediction Model



$_1 = f(x_k)$
x_{k-1}
f
x_k
f
x_{k+1}
Ground
Time step k
Measurement
θ_{k-1}
θ_k
θ_{k+1}
$= f_{\theta_k}(x_k)$
Prediction Model
Model Parameter
Prediction Model
Prediction Model
Model Update
Adaptation Mechanism
Adaptation Mechanism
Prediction
Model Adap
Pred

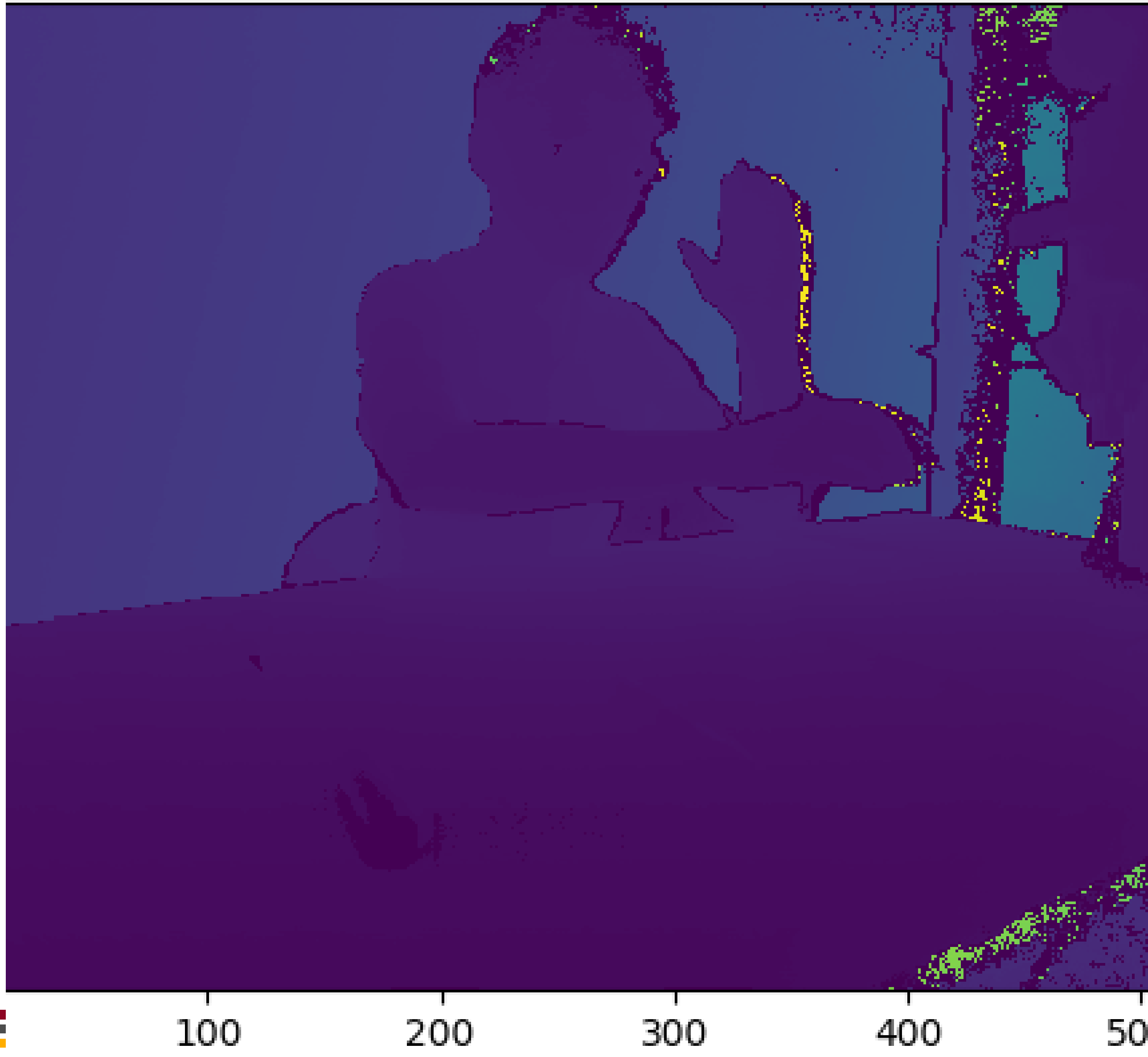
2.

EXPERIMENTS

Experiments we conducted to evaluate both the multi-task model and our adaptation

Collected Data

- Collected from two Actors, A and B
- Actor A performed each of the 12 actions 50 times (80% offline training, 20% offline validation)
- Actor B performed each action 10 times (100% online testing)



Using Multiple Sets of Adaptation Steps

- Implemented 1-step, 2-step, and 5-step adaptations
- As adaptation steps increase, the time it takes to perform the adaptation increases as well

Using Multiple Sets of Adaptation Steps

	Accuracy	MSE (cm^2)
Without Adaptation	0.930	5.508
1-step Adaptation	0.938	4.919
2-step Adaptation	0.938	4.488
5-step Adaptation	**0.946**	**3.964**

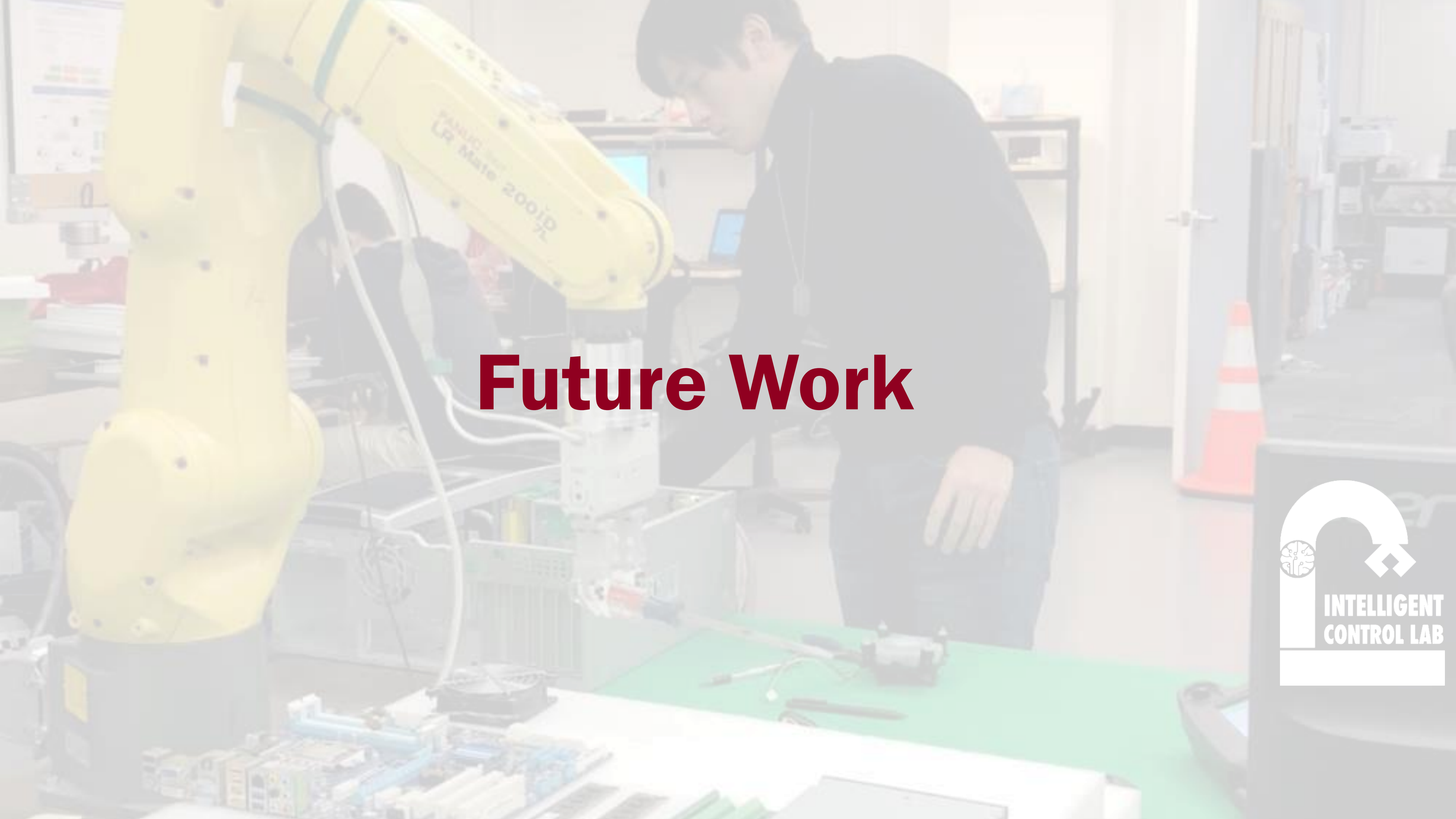
Using Single-task and Multi-task Models

- Single-task models for intention and trajectory predictions
- Sharing encoder weights between single and multi-task models

Using Single-task and Multi-task Models

	Accuracy	MSE (cm^2)
Single-task Intention Prediction	0.899	-
Single-task Trajectory Prediction	-	5.909
Multitask Simultaneous Prediction	**0.930**	**5.508**

Future Work



Prompts:

- To which extent is the human intention and trajectory predictable?
- How fast will the adaptation be considered fast enough?

Questions?

References

- [Pistohl et al. 2008] Pistohl, T.; Ball, T.; Schulze-Bonhage, A.; Aertsen, A.; and Mehring, C. 2008. Prediction of arm movement trajectories from ecog-recordings in humans. Journal of neuroscience methods 167(1):105–114.
- [Si, Wei, and Liu 2019] Si, W.; Wei, T.; and Liu, C. 2019. Agen: Adaptable generative prediction networks for autonomous driving. In IEEE Intelligent Vehicle Symposium, 2019.
- [Wang et al. 2018] Wang, W.; Li, R.; Chen, Y.; and Jia, Y. 2018. Human intention prediction in human-robot collaborative tasks. In Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, HRI ’18, 279–280. New York, NY, USA: ACM.