Automation of Time and Motion Data Collection Processes

Daelyn Bergsman, Will Locatelli, Audrey Slater, Sara Stavaski, Payam Vafadari, Bill Zhao

**DUBBUCKS**

STARBUCKS CAPSTONE — ISE @ W

PROBLEM

Current time study methods at Starbucks are: **inefficient**, **resource intensive**, and **error prone**.

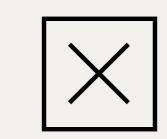They are done infrequently due to high costs, and lack accuracy.

Problem Statement

How can we develop a preliminary system that will **minimize resources** and **maximize accuracy** in time and motion study data collection?
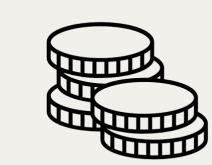
CURRENT SYSTEM
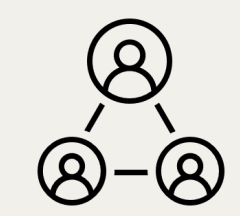
Manual: All time studies taken by hand.

Error Prone: Timestamps are estimated from multiple human recorded timestamps to account for human inaccuracy.

Expensive: Starbucks spends ~**$100K** each time they conduct time/motion studies. ~**70 labor hrs/wk** are required for dedicated time/motion study managers, engineers, and baristas.

METHODS

1 Data Collection:
Recorded baristas with two Microsoft Kinect DKs at a mock store set up in the Tryer center.

4 Baristas × 20 Lid to Handoff Action = 80 Videos + 20 **additional actions**
- Pump syrup
- Add whipped cream
- Pour milk

2 Labeling
Labeled videos using Microsoft VoTT.
2 actions:
"Lid to handoff" (key activity is occurring)
"Unknown" (key activity not occurring)

3 DeepHAR (Deep Human Activity Recognition):

Existing CNN model trained to classify **60 activities**. We re-trained the model to classify our **2 actions**.

Input
Video data collected at Tryer Center
>>> DeepHAR >>>
Output
Probability distribution of actions for video

4 Time Stamping:
Developed a sliding window technique.

Breaking the video into sections allows us to iterate through the entire video, processing individual sections at a time while classifying them based on the model's probability output.

Stride | Window
Full Length of the Video
Action Probability: 85%
Action Probability: 35%
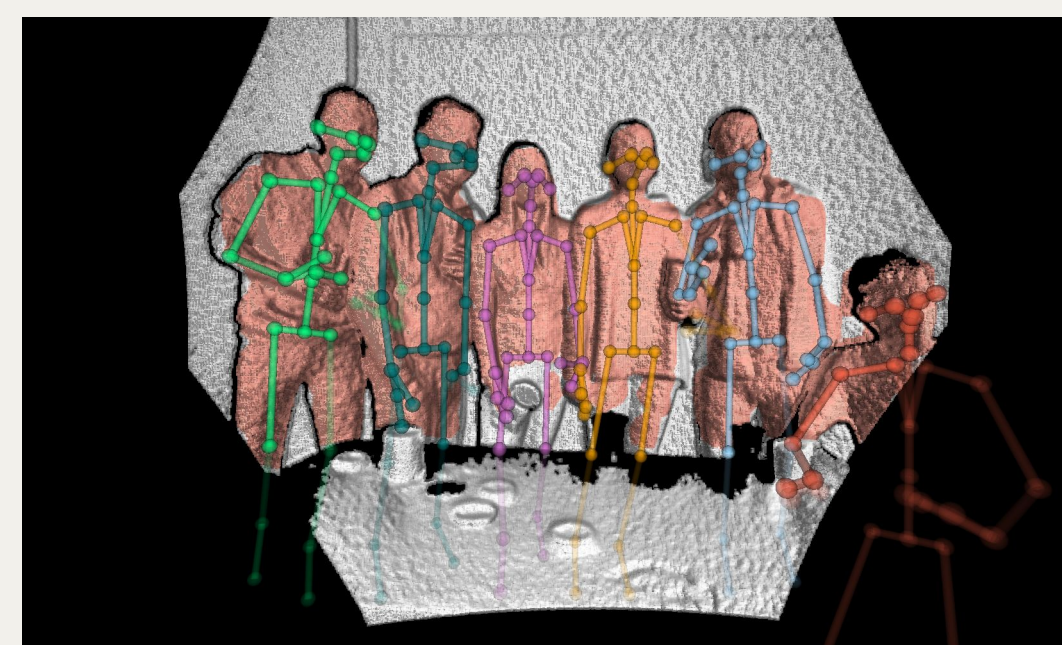
PROPOSED SOLUTION

Objective
To prove the **feasibility** of a system that could accurately recognize one key object manipulation activity in the Starbucks store and time how long it took to complete.
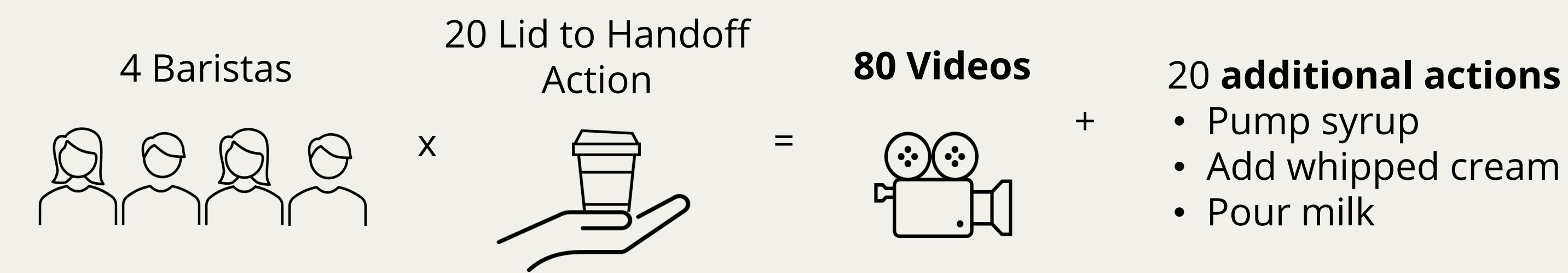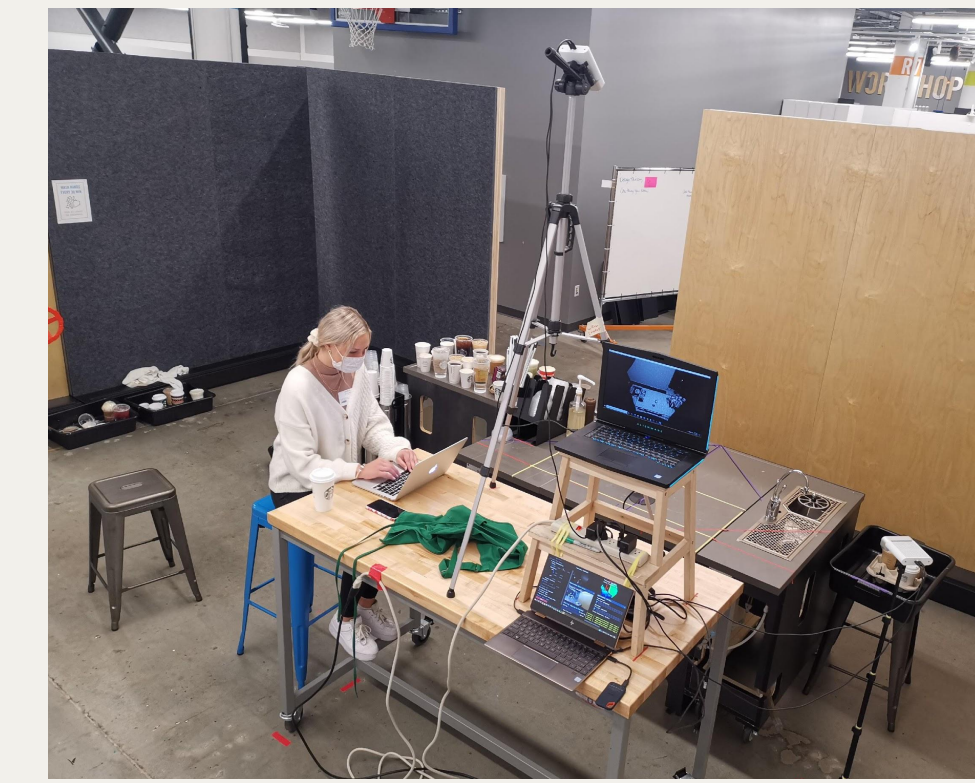
Key Activity
Lid to Handoff motion

Solution
Utilize transfer learning with machine learning models used to classify human actions with video data.
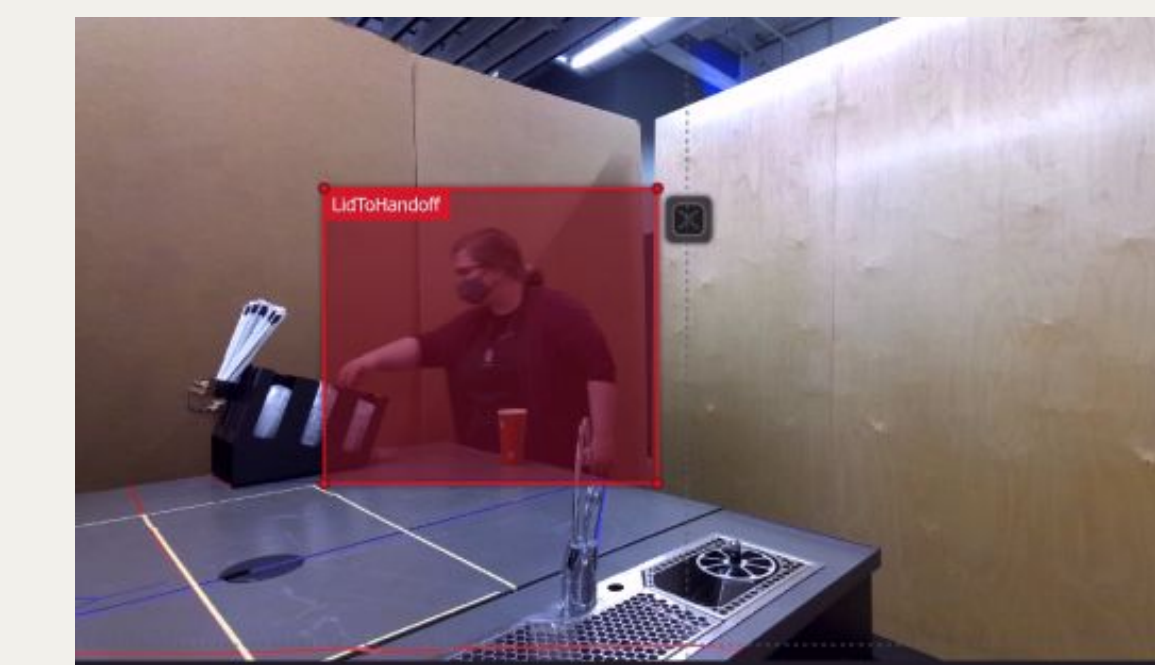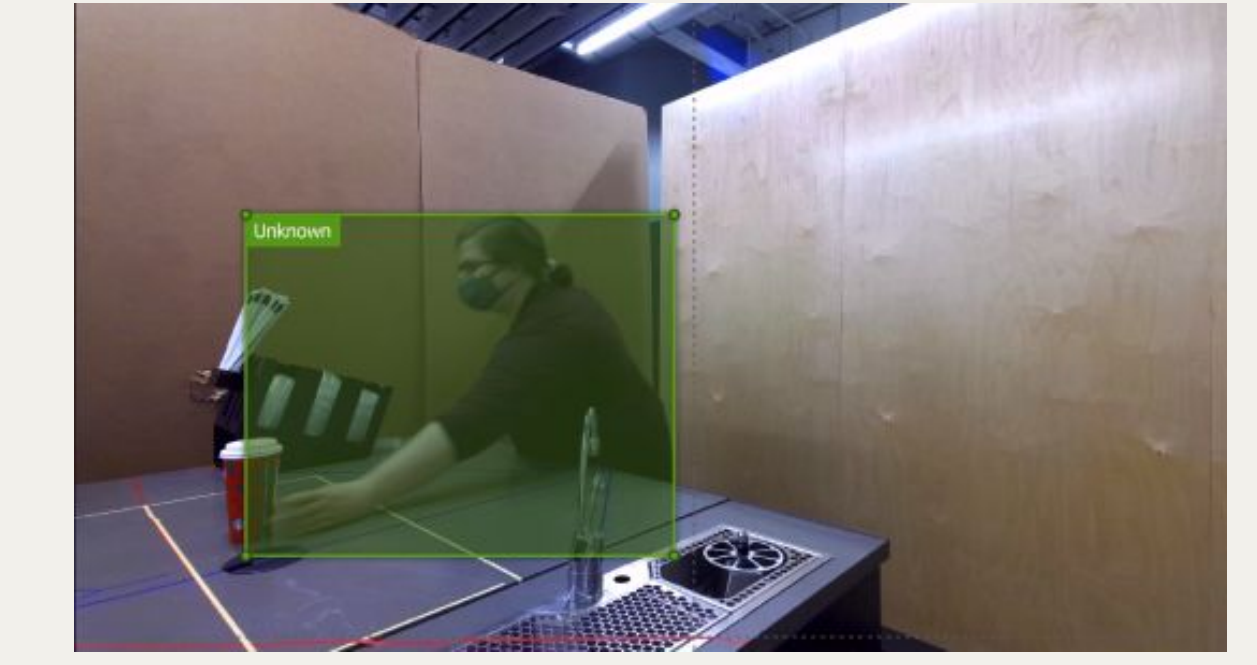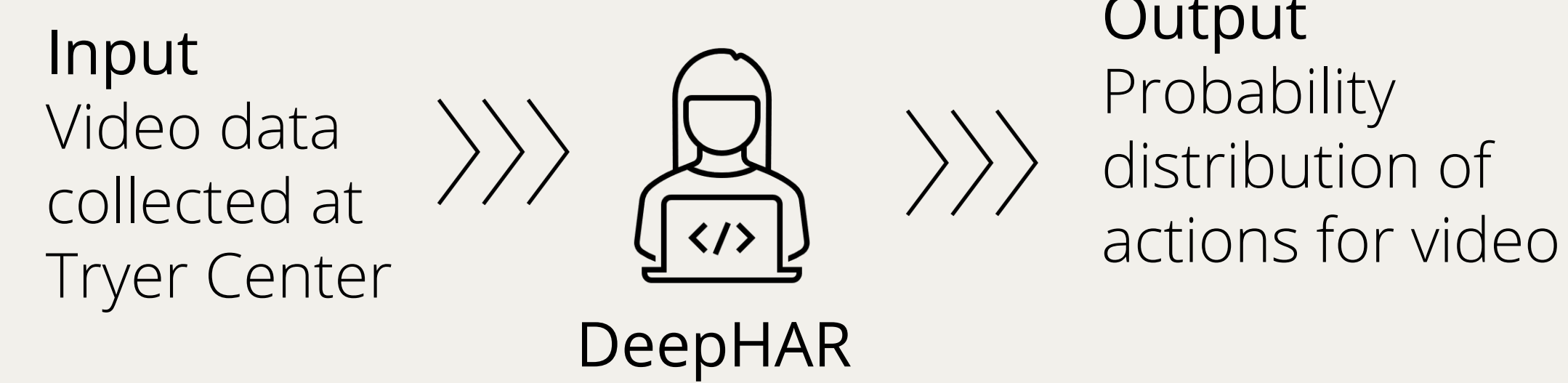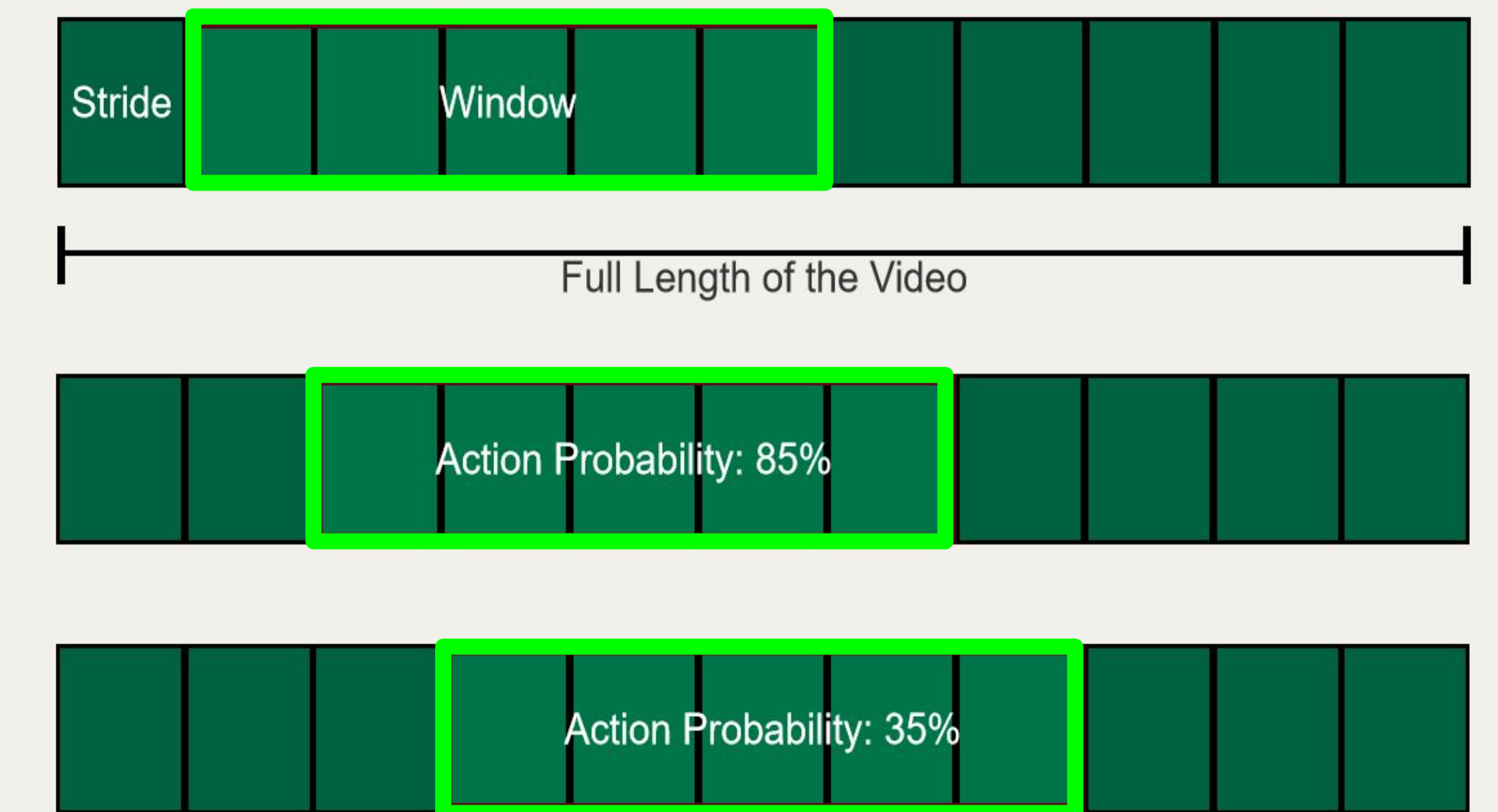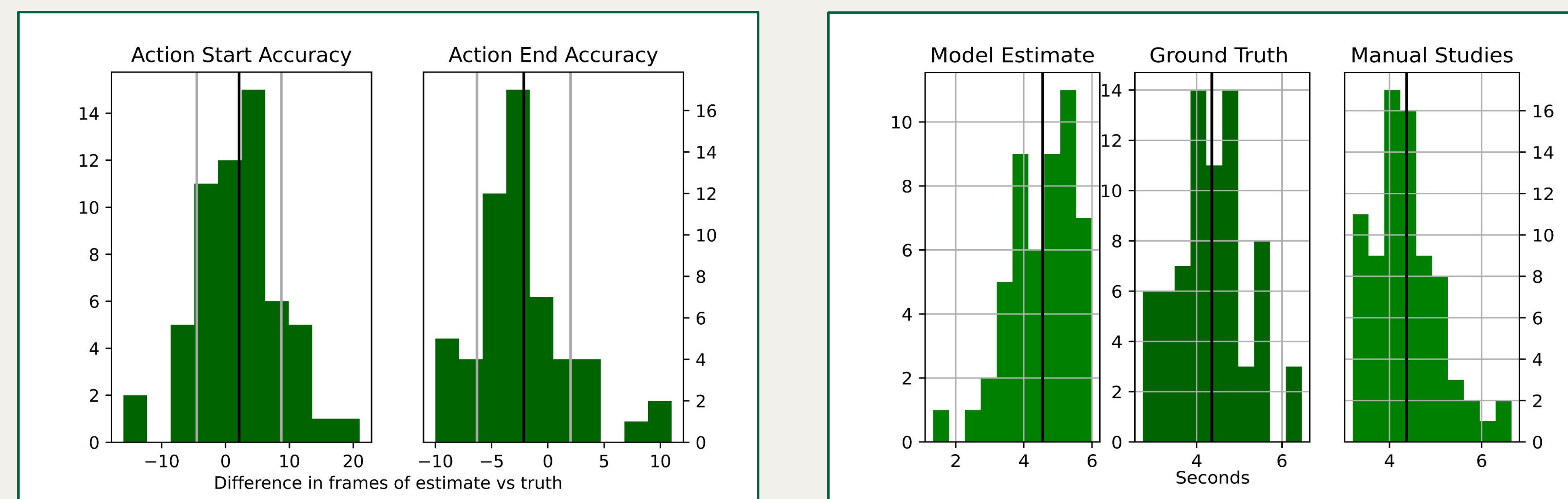
RESULTS

95% accuracy recognizing Lid to Handoff | **71%** accuracy timestamping
0.2 second difference in estimated duration

Action Start Accuracy | Action End Accuracy
Difference in frames of estimate vs truth

Model Estimate | Ground Truth | Manual Studies
Seconds

Next phases include continued research by a UW PHD student and further investigation into the capabilities of DeepHAR!

RECOMMENDATIONS

1 Expand the DeepHAR human activity recognition to include all motions of the time study process.

2 Take videos of each micromotion and label video frames using VoTT to train the model.

3 Implement privacy features into the video script.

Scan here to learn more and watch our video!