



ActionNet

<https://action-net.csail.mit.edu>

A Multimodal Dataset for Human Activities Using Wearable Sensors

Joseph DelPreto*, Chao Liu*, Yiyue Luo, Michael Foshey, Yunzhu Li
Antonio Torralba, Wojciech Matusik, and Daniela Rus

Neural Information Processing Systems (NeurIPS)
Track on Datasets and Benchmarks
2022



MIT Computer Science
and Artificial Intelligence Lab

Funded in part by the
Gwangju Institute of
Science and Technology (GIST)



Overview and Goals

Rich multimodal sensing focused on wearable devices

Motion



Tactile Sensing



Muscle Activity



Body and Finger Tracking



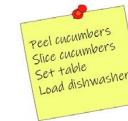
Eye Tracking



RGB+D Video



Audio

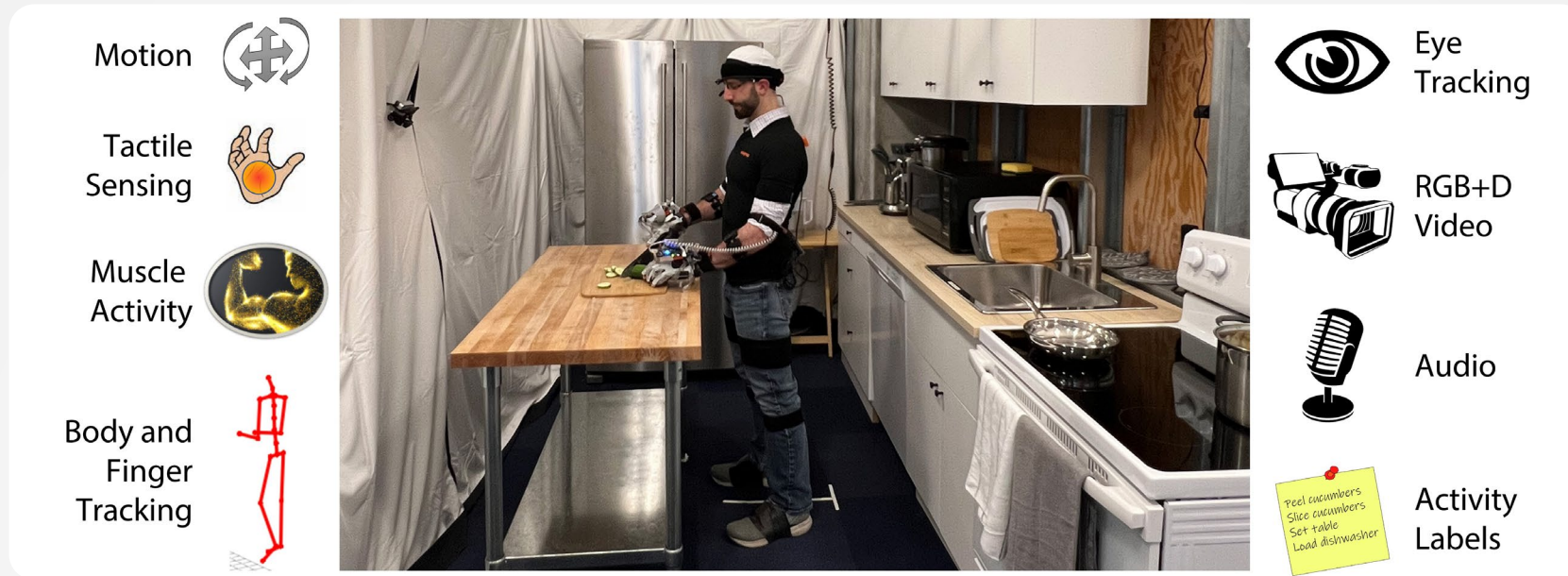


Activity Labels

Peel cucumbers
Slice cucumbers
Set table
Load dishwasher

Overview and Goals

Rich multimodal sensing focused on wearable devices



Insights about how humans interact with physical or virtual worlds

Train robot assistants

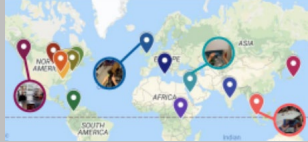
Train people

Interactive simulations

Smart textiles

Related Datasets

Large-Scale Image/Video



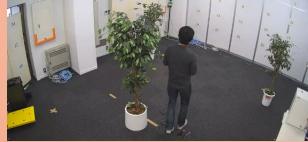
Ego4D: Grauman et al., 2022

EPIC-KITCHENS-100: Damen et al., 2022

Action Genome: Ji et al., 2020

ActivityNet: Heilbron et al., 2015

Multimodal with Wearables



MMAct: Kong et al., 2019

Stanford ECM: Nakamura et al., 2017

UTD-MHAD: Chen et al., 2015

CMU-MMAC: de la Torre et al., 2009

Data-Driven Robotics Applications



Levine et al., 2018

Pinto et al., 2016

Finn et al., 2016

Agrawal et al., 2016

Simulation



BEHAVIOR: Srivastava et al., 2021

Meta-World: Yu et al., 2020

VirtualHome: Puig et al., 2018

Columbia: Goldfeder et al., 2009

Demonstrations



MIME: Sharma et al., 2018

RoboMimic: Mandlekar et al., 2021

Related Datasets

Large-Scale Image/Video



Ego4D: Grauman et al., 2022

EPIC-KITCHENS-100: Damen et al., 2022

Action Genome: Ji et al., 2020

ActivityNet: Heilbron et al., 2015

Multimodal with Wearables



MMAct: Kong et al., 2019

Stanford ECM: Nakamura et al., 2017

UTD-MHAD: Chen et al., 2015

CMU-MMAC: de la Torre et al., 2009

Data-Driven Robotics Applications



Levine et al., 2018

Pinto et al., 2016

Finn et al., 2016

Agrawal et al., 2016

Simulation



BEHAVIOR: Srivastava et al., 2021

Meta-World: Yu et al., 2020

VirtualHome: Puig et al., 2018

Columbia: Goldfeder et al., 2009

Demonstrations



MIME: Sharma et al., 2018

RoboMimic: Mandlekar et al., 2021

- Highly multimodal
- Wearable and global
- Activity labels
- Structured tasks
- Parsing and recording frameworks



Sensors: Wearable Devices

Motion Tracking

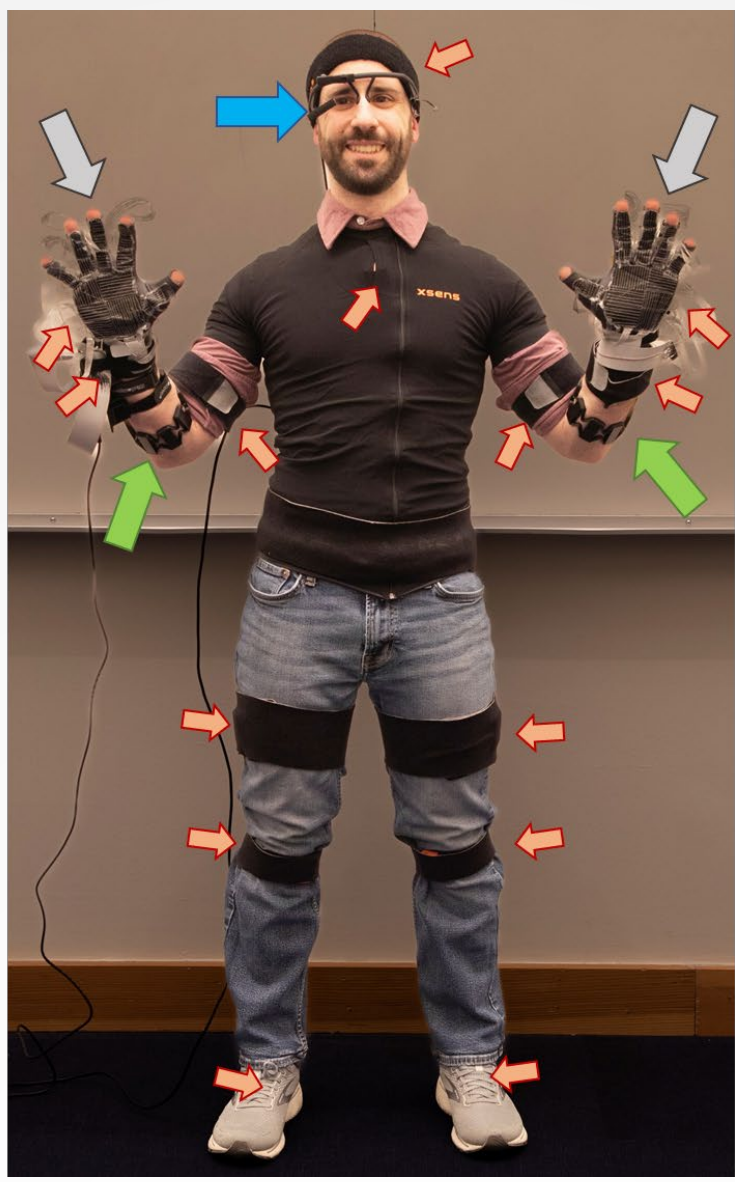


Xsens Awinda body tracking

Manus Gloves finger tracking

Custom tactile sensors

Interaction Forces



Myo Armband muscle sensor

Pupil Labs eye tracking and first-person video



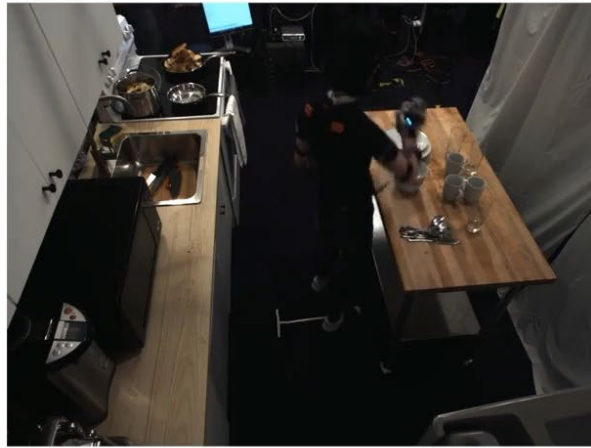
Muscle Activity



Attention



Sensors: Global Perspectives

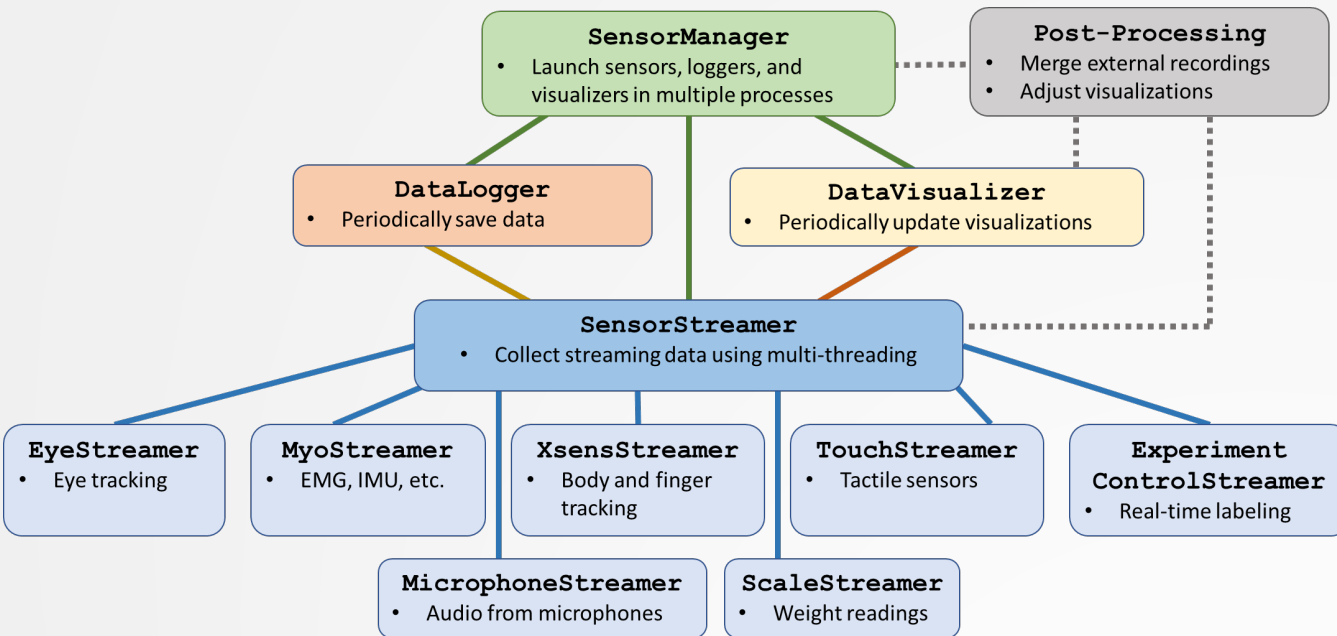


5x

-  ➤ 5 RGB cameras
-  ➤ 1 RGBD camera
-  ➤ 2 microphones

Software Infrastructure: Recording

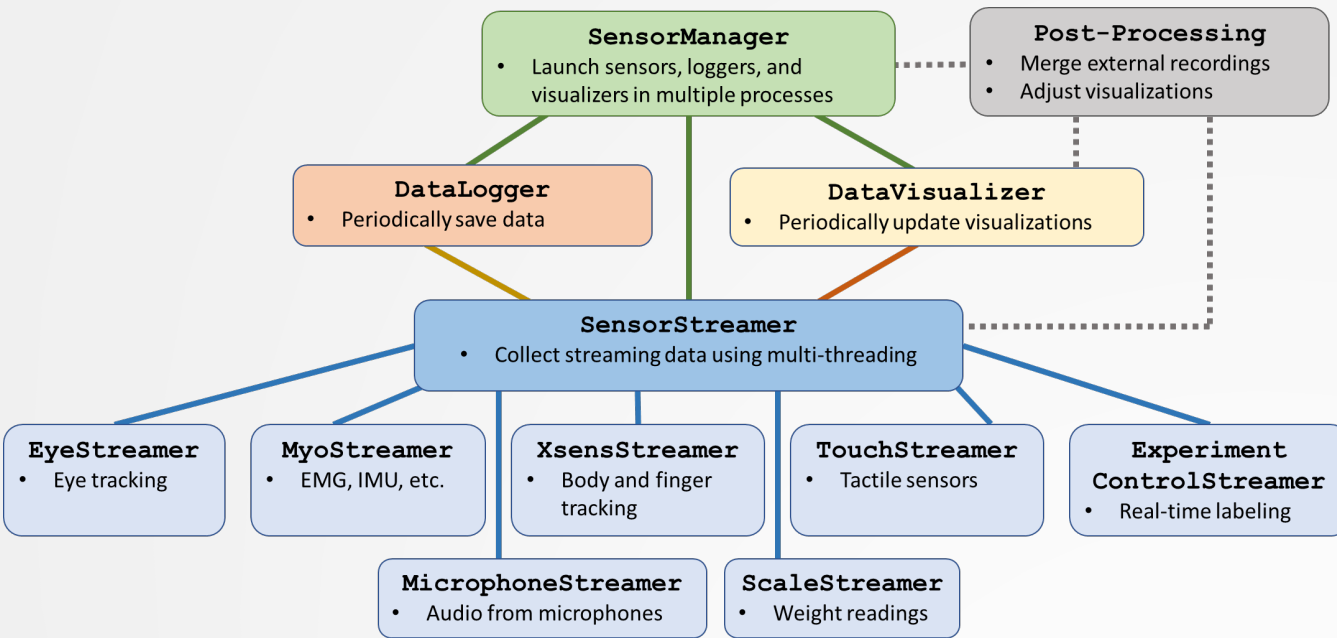
Extensible Recording Framework



- Stream, synchronize, save, and visualize data from all sensors
- Post-processing for higher-quality data
- Easily add or replace sensors

Software Infrastructure: Labeling

Extensible Recording Framework



- Stream, synchronize, save, and visualize data from all sensors
- Post-processing for higher-quality data
- Easily add new sensors

Experiment Control and Real-Time Labeling

Calibration

Body Arms Hands Gaze Tactile Third-Party

Left hand pose/object Flat hand on scale

Right hand pose/object

Time since last calibrations:

- body
- arms
- hand_poses
- gaze
- tactile_gloves
- third_party

Start Calibration Mark Good Mark Maybe Mark Bad

Activities

Slice a cucumber

Start Activity Mark Good Mark Maybe Mark Bad

Number of each activity performed [good/maybe/bad]:

- 2 / 0 / 0: Get/replace items from refrigerator/cabinets/drawers
- 3 / 1 / 0: Clear cutting board
- 3 / 0 / 1: Peel a cucumber
- 1 / 0 / 0: Slice a cucumber
- 0 / 0 / 0: Peel a potato
- 0 / 0 / 0: Slice a potato
- 0 / 0 / 0: Slice bread
- 0 / 0 / 0: Spread almond butter on a bread slice
- 0 / 0 / 0: Spread jelly on a bread slice
- 0 / 0 / 0: Open/close a jar of almond butter
- 0 / 0 / 0: Pour water from a pitcher into a glass
- 0 / 0 / 0: Clean a plate with a sponge
- 0 / 0 / 0: Clean a plate with a towel
- 0 / 0 / 0: Clean a pan with a sponge
- 0 / 0 / 0: Clean a pan with a towel
- 0 / 0 / 0: Get items from cabinets: 3 each large/small plates, bowls, n
- 0 / 0 / 0: Set table: 3 each large/small plates, bowls, mugs, glasses, s
- 0 / 0 / 0: Stack on table: 3 each large/small plates, bowls
- 0 / 0 / 0: Load dishwasher: 3 each large/small plates, bowls, mugs, gl
- 0 / 0 / 0: Unload dishwasher: 3 each large/small plates, bowls, mugs,

General notes

Submit Notes

- Interactive GUI for synchronized ground truth
- Label, rate, and annotate activities as they occur
- Keep track of experimental flow

Dataset Activities: Kitchen Tasks

Food Preparation



Peel a Cucumber



Peel a Potato



Clear Cutting Board



Spread Almond Butter



Slice a Cucumber



Slice a Potato



Slice Bread



Spread Jelly

Cleaning Tableware



Clean a Plate with a Sponge or Towel



Clean a Pan with a Sponge or Towel



Manipulations



Open/Close a Jar



Pour Water

High-Level Tableware Tasks



Fetch Tableware



Fetch/Return Items



Stack Tableware



Set Table



Load Dishwasher



Unload Dishwasher

Visualizations: Multimodal Wearable Data



Tactile



Muscles

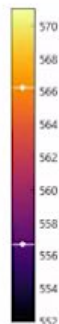
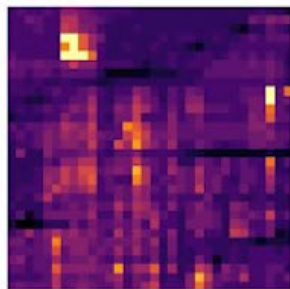


Motion

2.5x

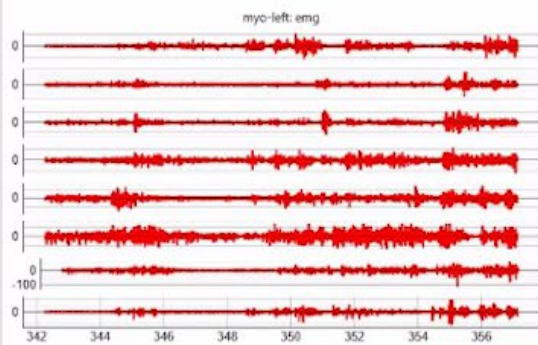
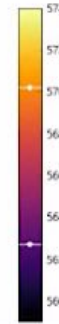
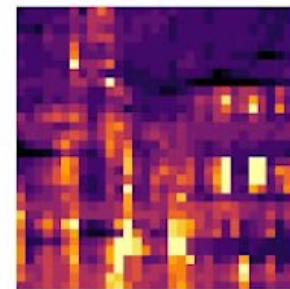
Left Arm

tactile-glove-left: tactile_data

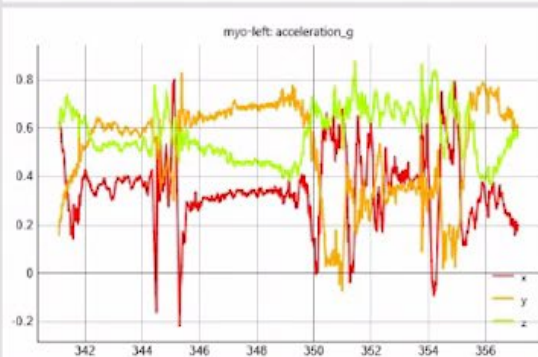


Right Arm

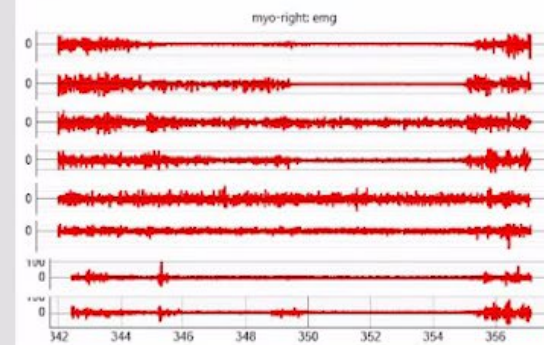
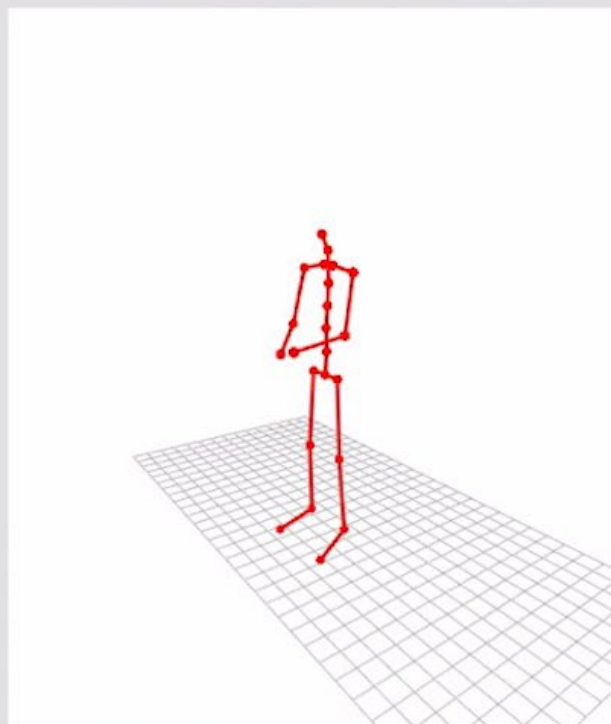
tactile-glove-right: tactile_data



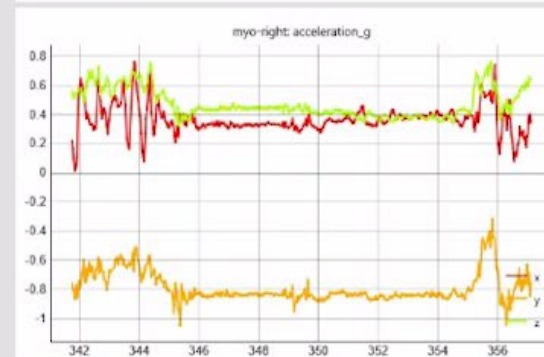
myo-left: emg



myo-left: acceleration_g



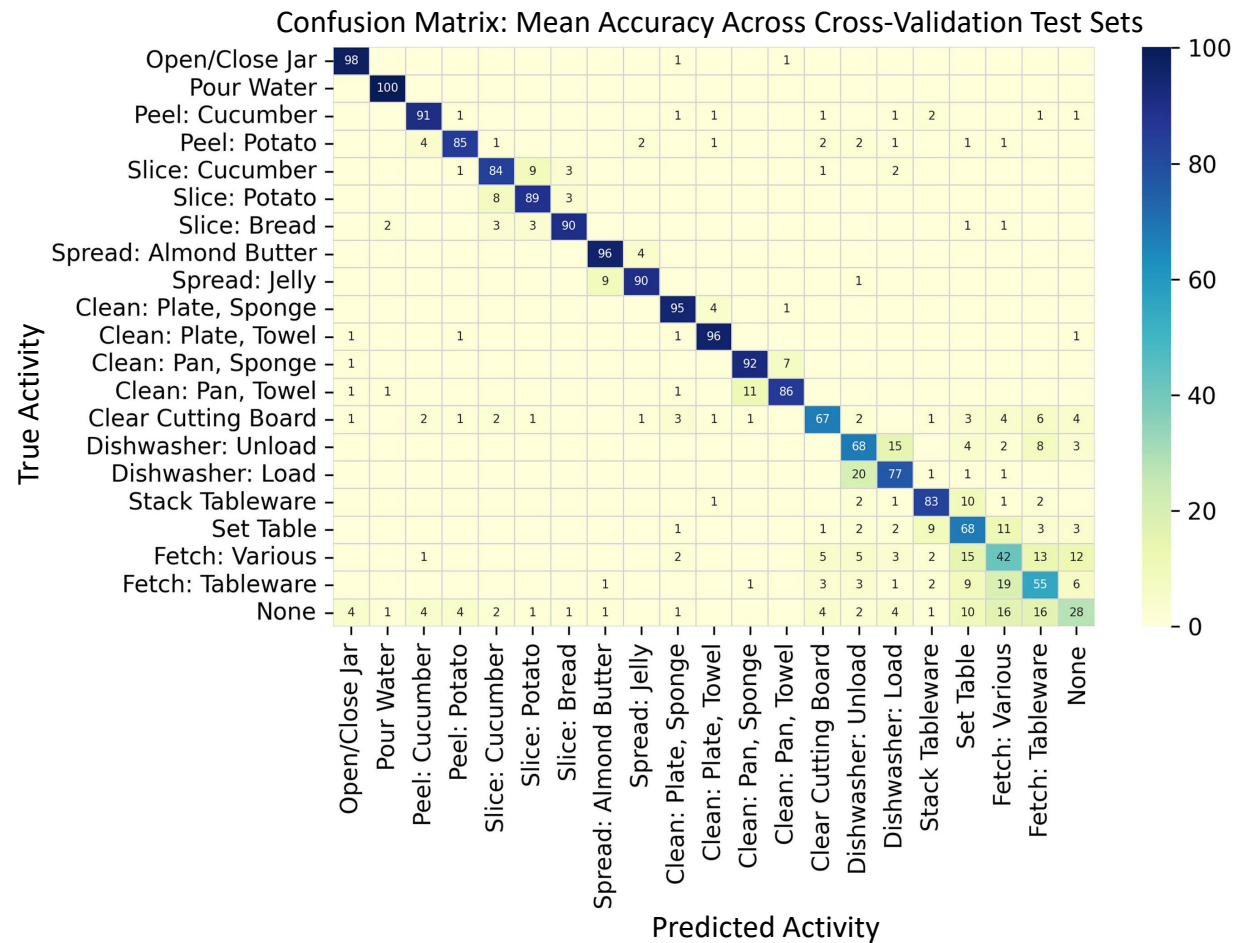
myo-right: emg



myo-right: acceleration_g

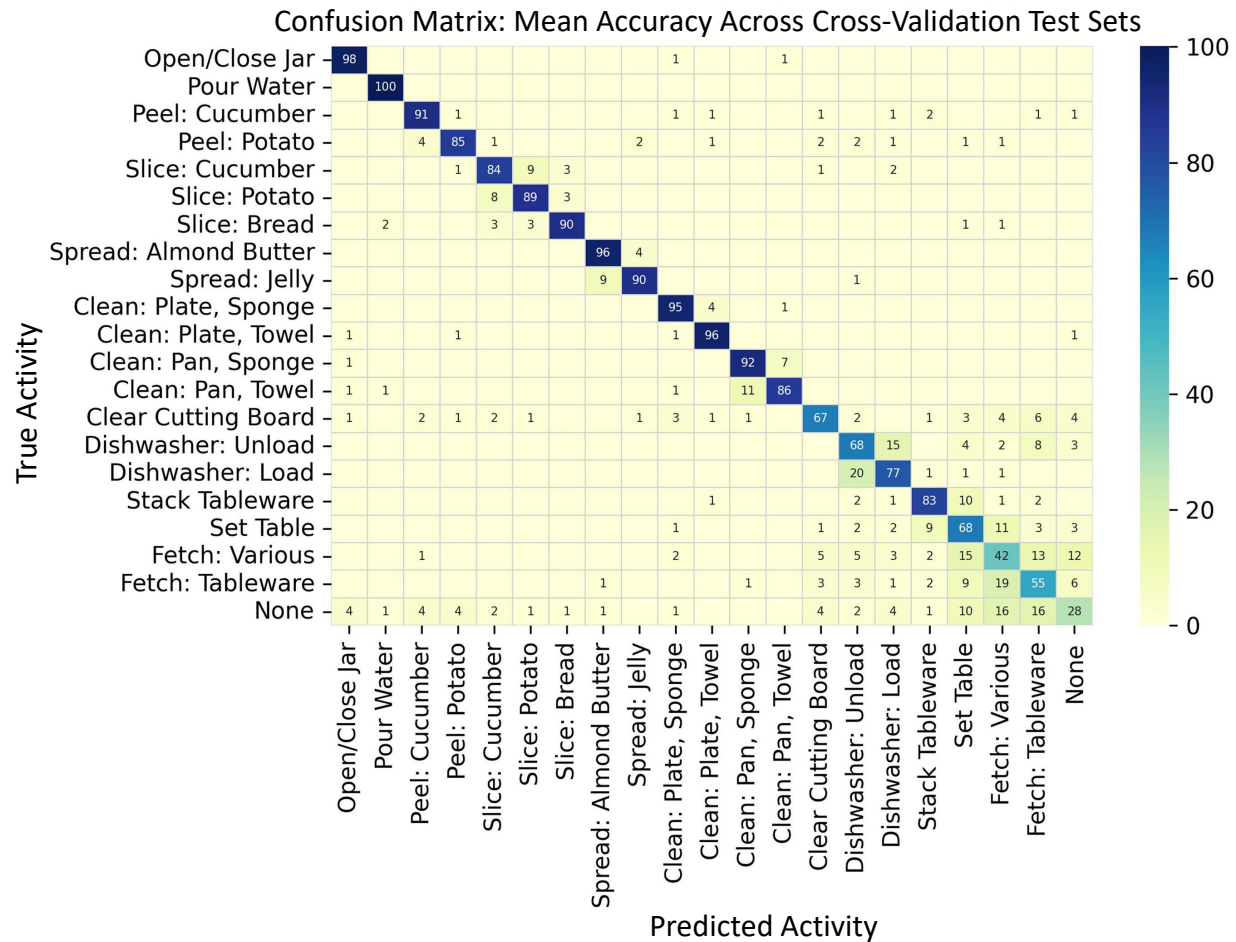
Applications: Classification and Modality Studies

Activity Classification and Modality Ablations Using LSTMs

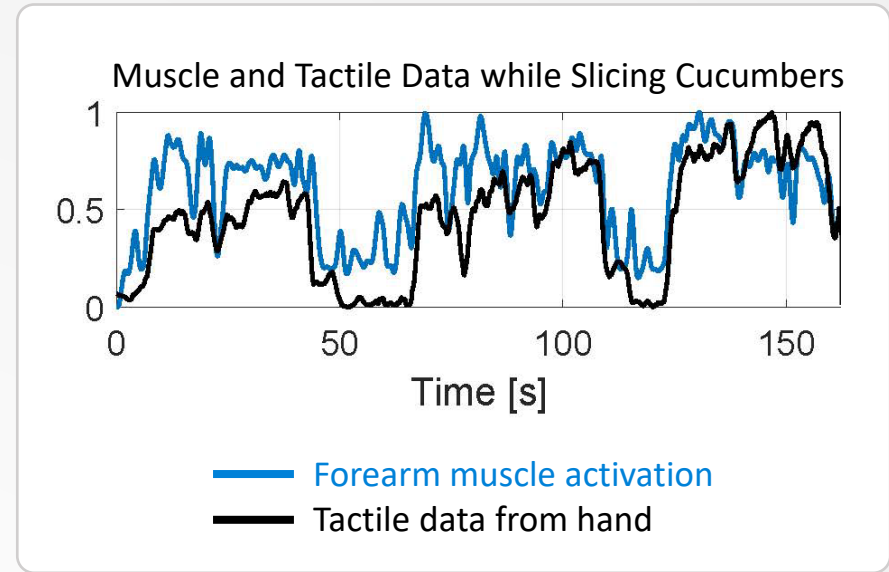


Applications: Classification and Modality Studies

Activity Classification and Modality Ablations Using LSTMs



Cross-Modal Analysis



Conclusion and Future Work

- Multimodal dataset focused on wearable devices and common household tasks
- Synchronized global data and ground-truth labels
- Extensible recording and processing framework

Models and analysis

- Cross-modal predictions
- Sequence reasoning

Teaching from demonstrations

- Robotic assistants
- Personalized training

Extensions and integrations

- Immersive simulations
- New task domains

<https://action-net.csail.mit.edu>

