EgoUse

Object use from egocentric video

Visual Perception for AR Glasses Project

Daniel Heller

Advisors: Zihao Ding, Prof. Yao liu

WINLAB
WIRELESS INFORMATION NETWORK LABORATORY

RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Where did I put my keys?

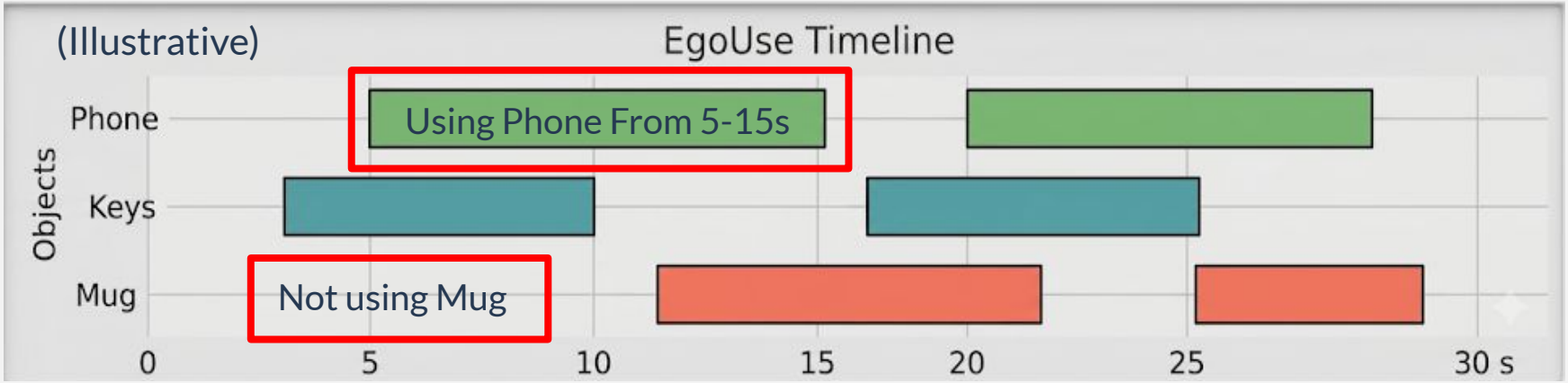
- **Smart glasses** are more of a trend than ever
- They record **egocentric** video (first-person view)
- How might we be able to use them to help us find our keys we lost earlier?
- **Difficulty:** This software must **run live** and and be **lightweight** to not drain battery

Goal: Determine when the wearer is using an object and identify which object

- Later use this information to find your keys

What we want EgoUse to output

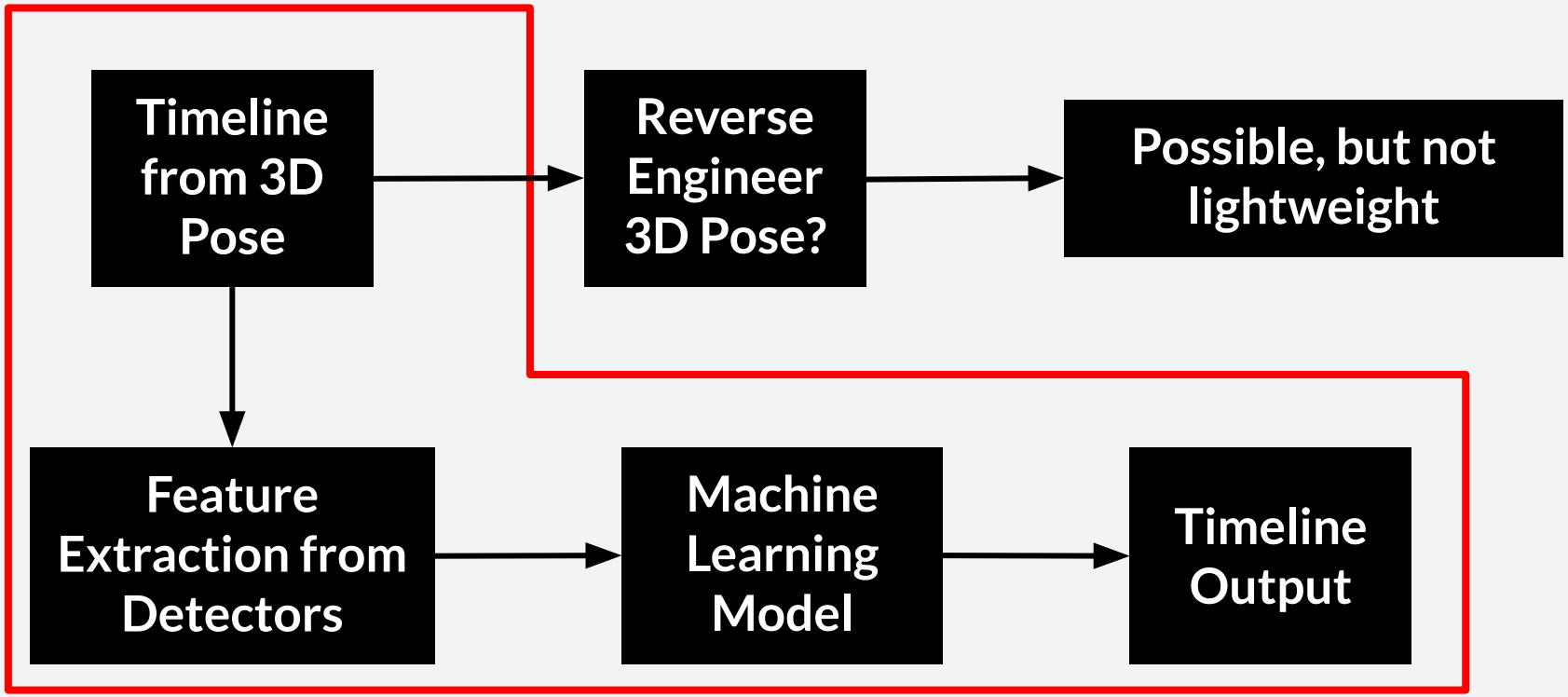
- **Output:** A timeline showing which objects the wearer is using
  - Colored blocks: Object use



Building a timeline with 3D positions

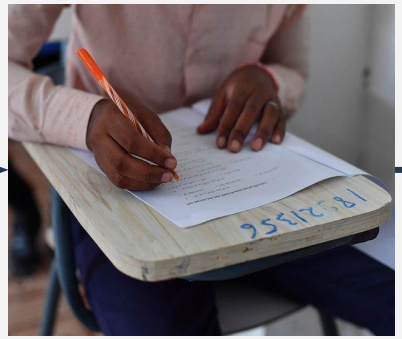
- **Dataset:** HOT3D (Meta) – Egocentric dataset of video from **Project Aria** smart glasses.
  - Accurate 3D positions of hands and objects (132 videos)

Interaction based on **Two Rules** based on 3D positional data


Timeline from 3D Pose
Reverse Engineer 3D Pose?
Possible, but not lightweight
Feature Extraction from Detectors
Machine Learning Model
Timeline Output

Teaching the student model

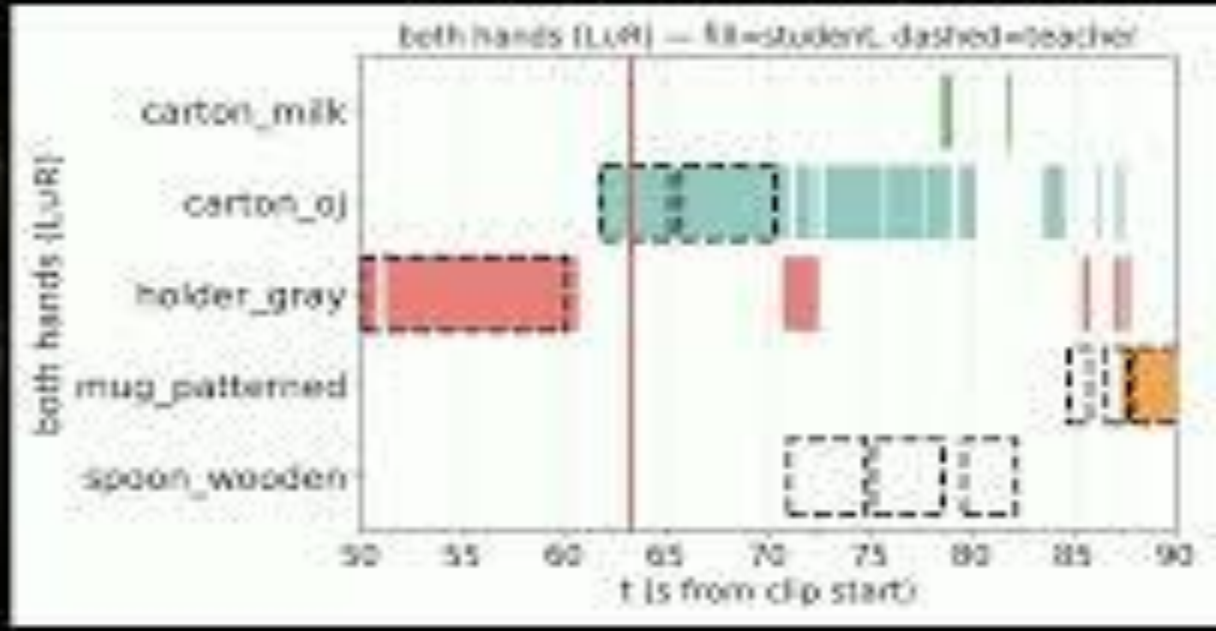
- **Teacher-Student model learning:** We train a student model to reproduce the teachers results with limited information
  - **Teacher:** The timelines from 3D positional data
  - **Student:** Only receives features from detectors, not video itself



Student Produces Timeline

Student Submits

Teacher gives feedback off the Answer Key



Video with Detectors

Student = Filled (color), Teacher = Dashed Boxes

Findings

	Recall (Frame)
Training (122)	93.9%
Validation (10)	18.0%

How many of the interactions did it get?

	2 minute clip
Model	<1 second
Entire Pipeline	~5 minutes

Detector speed takes up ~90% of time

Limitations

Pre-recorded footage, not live

1 Dataset, 1 Device

EgoUse

Object use from egocentric video

- Strong Timelines from 3D research data
- Weaker timelines from egocentric video
- Tools to evaluate and plot timelines

Scan for Wiki Page

WINLAB
WIRELESS INFORMATION NETWORK LABORATORY

RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY