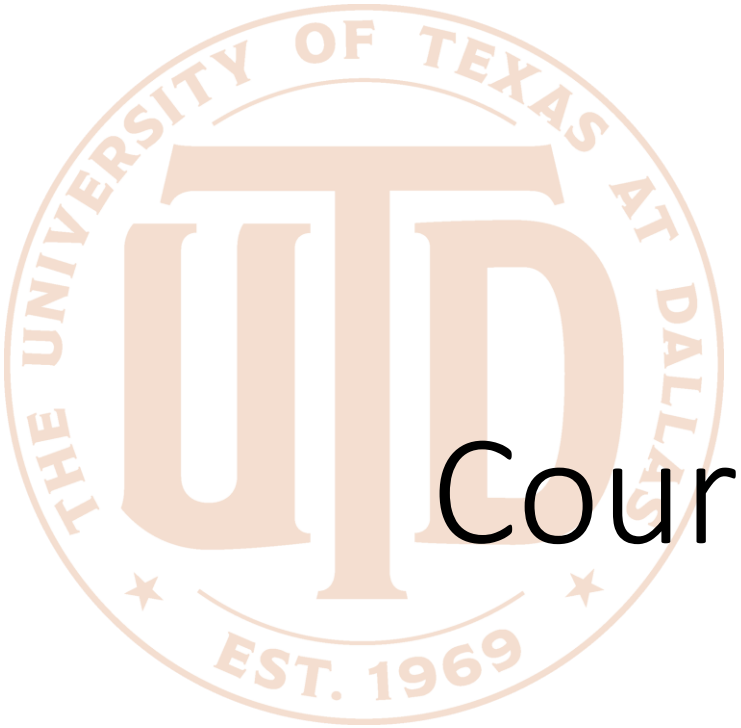


Course Project Description

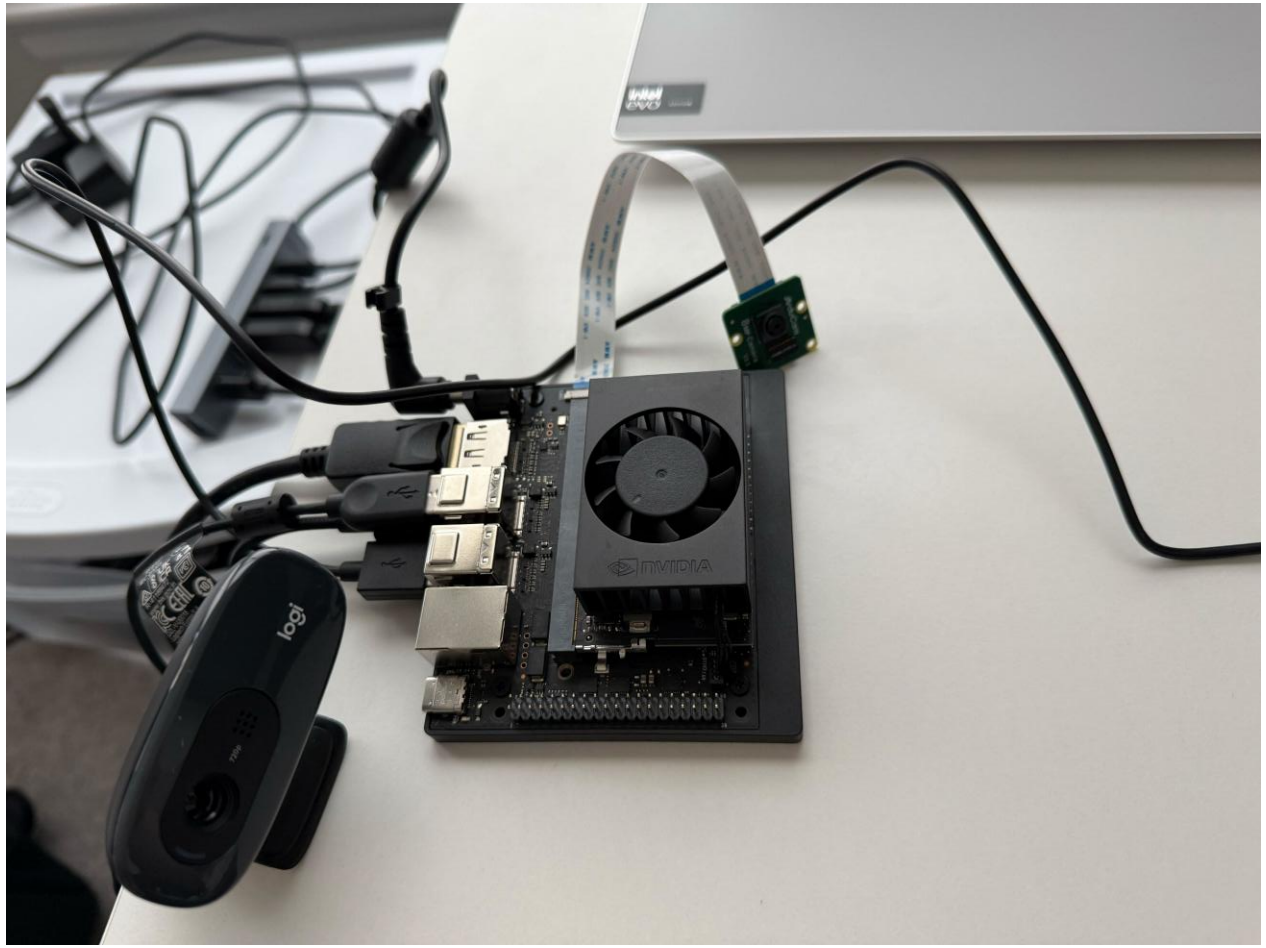
CS 4391 Introduction Computer Vision

Instructor: Yu Xiang

The University of Texas at Dallas

Hardware

- NVIDIA Jetson Orin Nano
- IMX219 Camera or
- Logitech C270 HD Webcam



Course Project

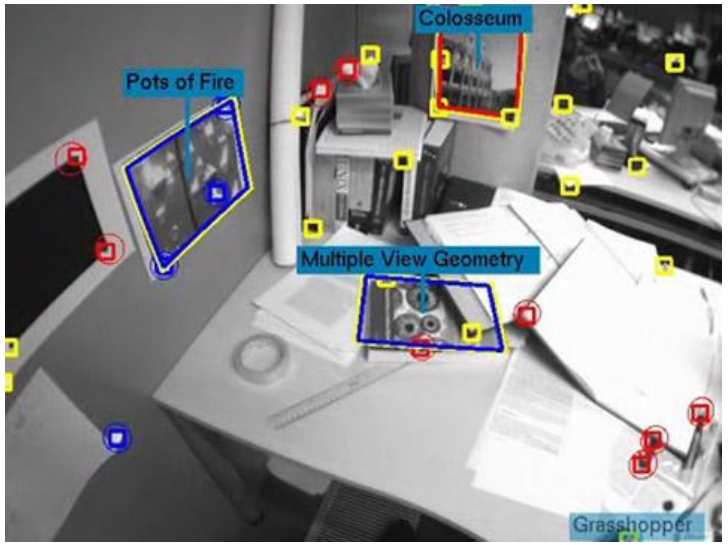
- Team Project (45%)
  - 4 students for a project: **sign up in eLearning**
  - Project proposal (10%)
  - Project mid-term report (10%)
  - Project presentation (10%): **in-class demo with Jetson and camera**
  - Project final report (15%)

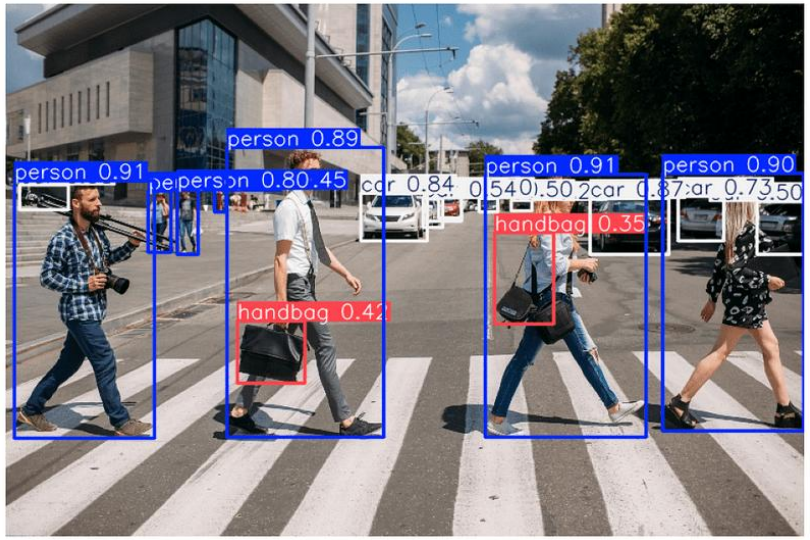
Course Project

- Each project needs to perform the following stages
  1. Data collection using the real camera
  2. Data labeling if supervision is needed
  3. Model training with the labeled data
  4. Model optimization using tensorRT
  5. Model deployment to Jetson Nano Orin
  6. Model evaluation on datasets
  7. Demonstration in classroom

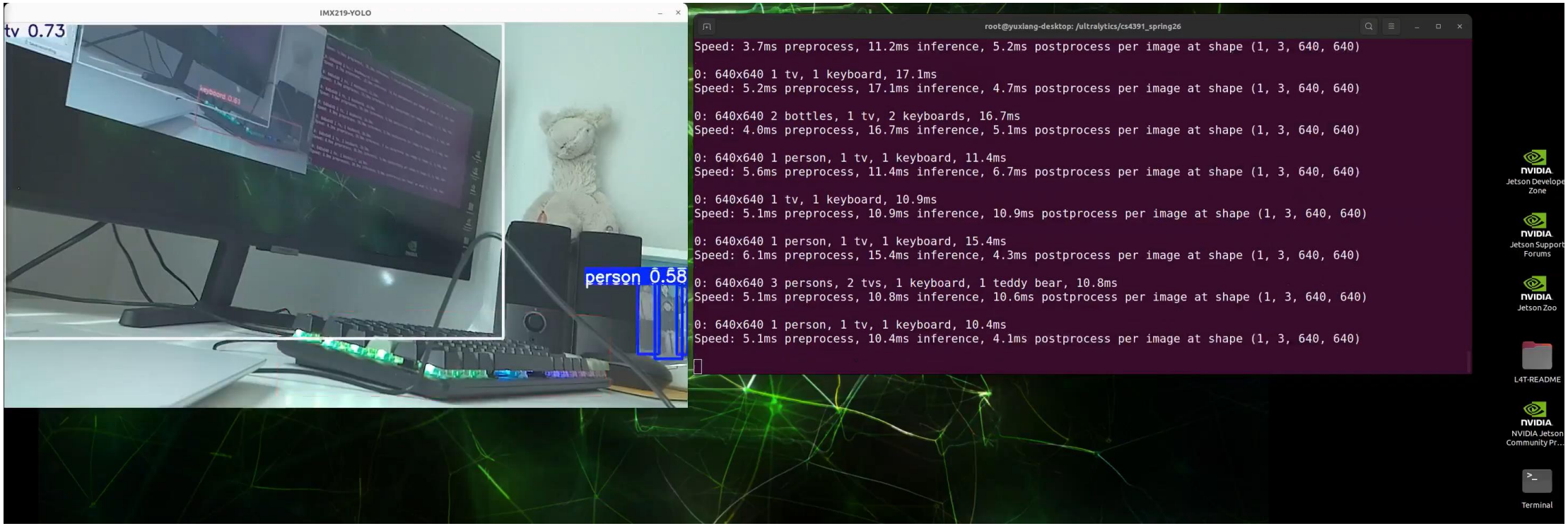
Topic 1: Custom Object Detection + Tracking

- Detecting or tracking specific object instances or categories



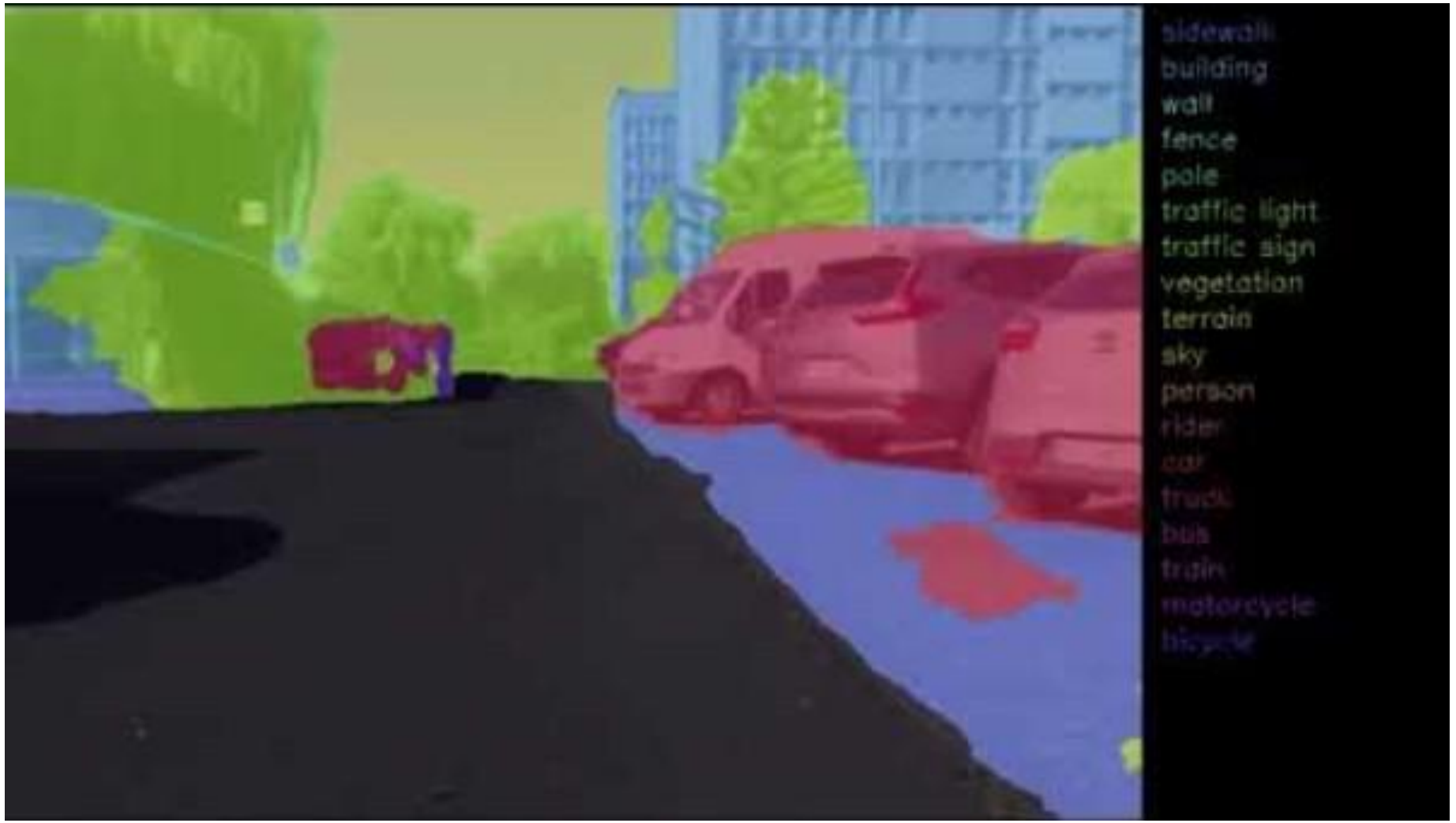


Demo


IMX219-YOLO
tv 0.73
person 0.58
root@yuxiang-desktop: /ultralytics/cs4391_spring26
Speed: 3.7ms preprocess, 11.2ms inference, 5.2ms postprocess per image at shape (1, 3, 640, 640)
0: 640x640 1 tv, 1 keyboard, 17.1ms
Speed: 5.2ms preprocess, 17.1ms inference, 4.7ms postprocess per image at shape (1, 3, 640, 640)
0: 640x640 2 bottles, 1 tv, 2 keyboards, 16.7ms
Speed: 4.0ms preprocess, 16.7ms inference, 5.1ms postprocess per image at shape (1, 3, 640, 640)
0: 640x640 1 person, 1 tv, 1 keyboard, 11.4ms
Speed: 5.6ms preprocess, 11.4ms inference, 6.7ms postprocess per image at shape (1, 3, 640, 640)
0: 640x640 1 tv, 1 keyboard, 10.9ms
Speed: 5.1ms preprocess, 10.9ms inference, 10.9ms postprocess per image at shape (1, 3, 640, 640)
0: 640x640 1 person, 1 tv, 1 keyboard, 15.4ms
Speed: 6.1ms preprocess, 15.4ms inference, 4.3ms postprocess per image at shape (1, 3, 640, 640)
0: 640x640 3 persons, 2 tvs, 1 keyboard, 1 teddy bear, 10.8ms
Speed: 5.1ms preprocess, 10.8ms inference, 10.6ms postprocess per image at shape (1, 3, 640, 640)
0: 640x640 1 person, 1 tv, 1 keyboard, 10.4ms
Speed: 5.1ms preprocess, 10.4ms inference, 4.1ms postprocess per image at shape (1, 3, 640, 640)
NVIDIA Jetson Developer Zone
NVIDIA Jetson Support Forums
NVIDIA Jetson Zoo
L4T-README
NVIDIA NVIDIA Jetson Community Pr...
Terminal

Topic 2: Semantic Segmentation of Indoor Scenes

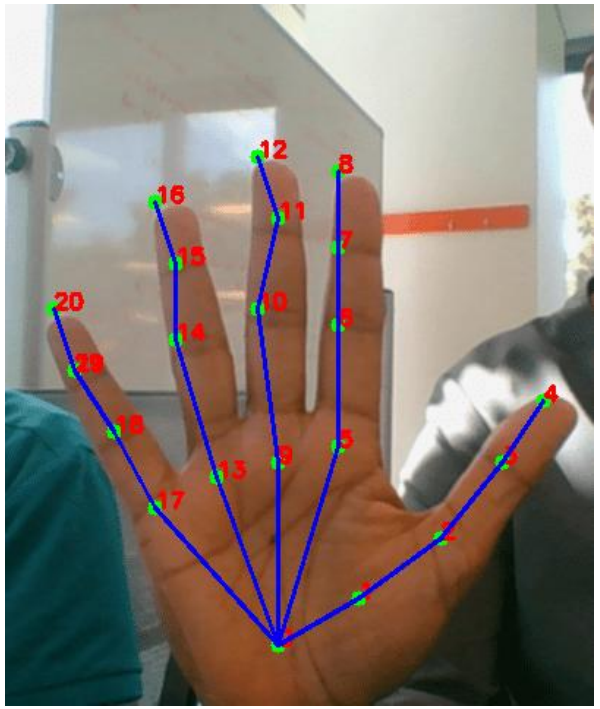
- Labeling pixels into semantic classes



https://www.youtube.com/watch?v=ILzbgNE0fPk

Topic 3: Hand or Human Pose Estimation

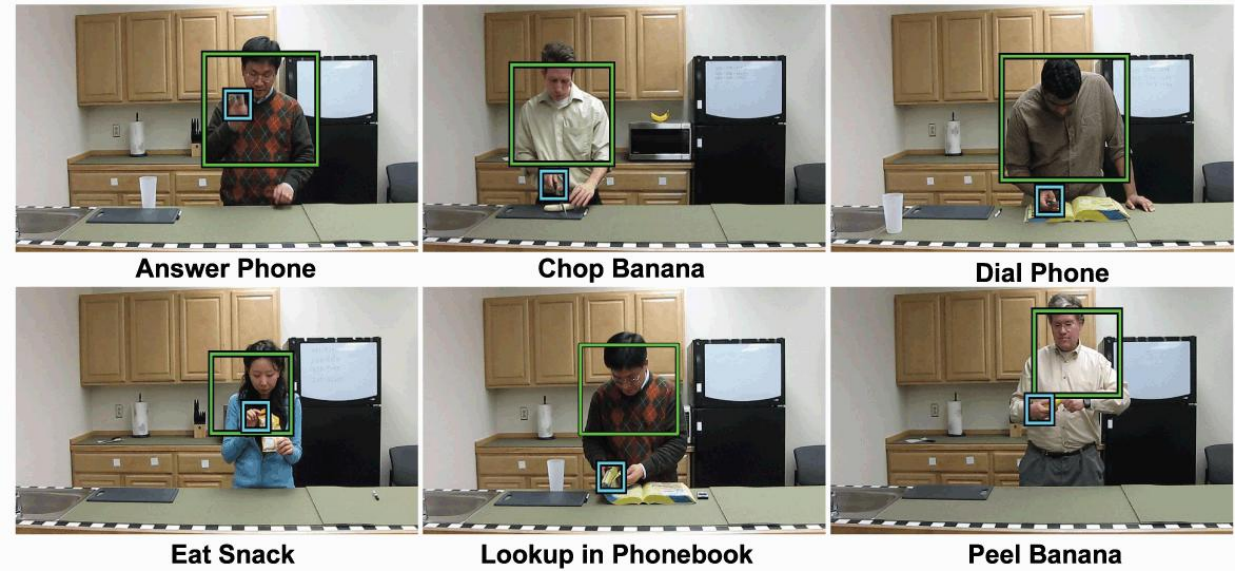
- Detecting or tracking human hands or human bodies

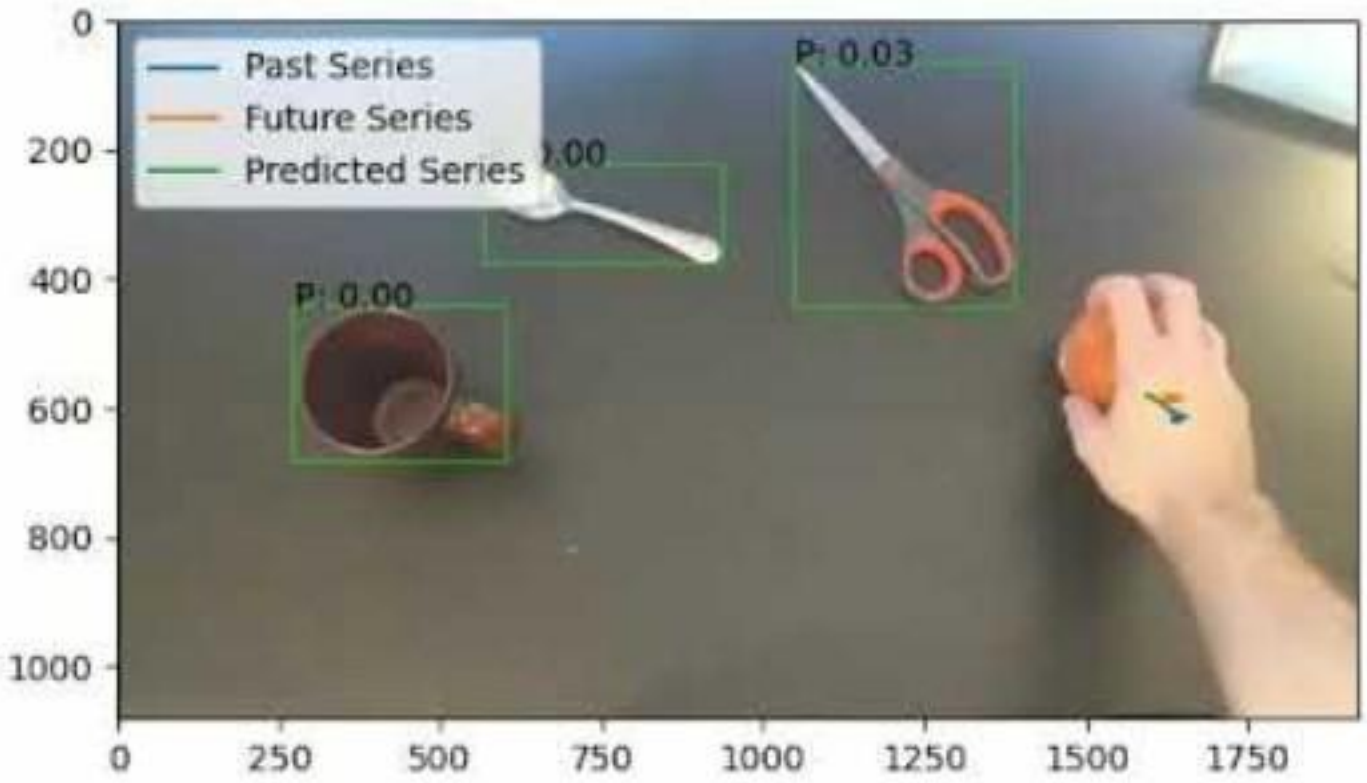


https://youtu.be/RMgrAxds3DU

Topic 4: Human-Object Interaction Recognition

- Recognizing the human actions from images or video clips



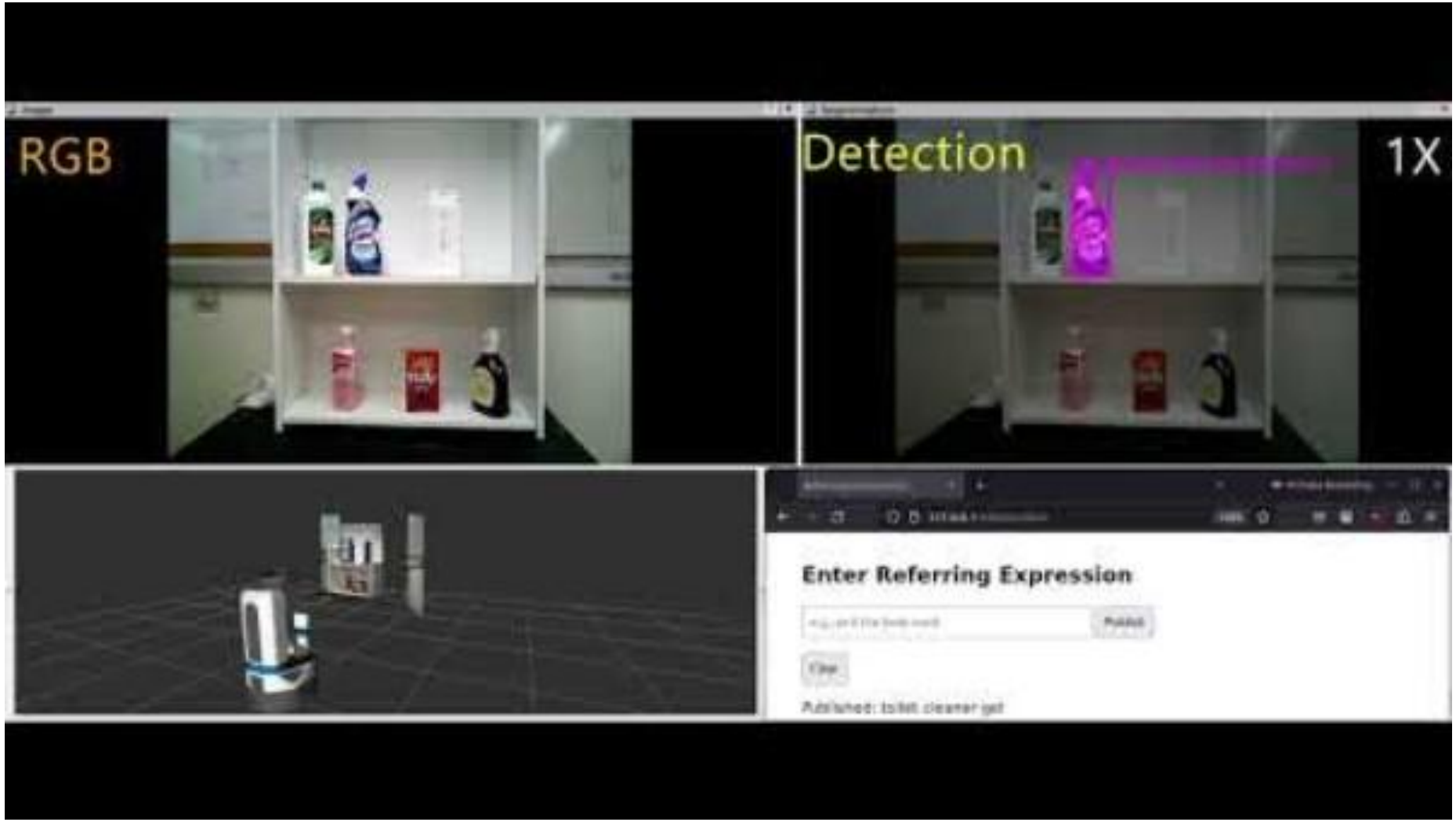


https://youtu.be/nKqXu4bZbFY

Topic 5: Visual Grounding

- Localize a target object in a given image according to a language description

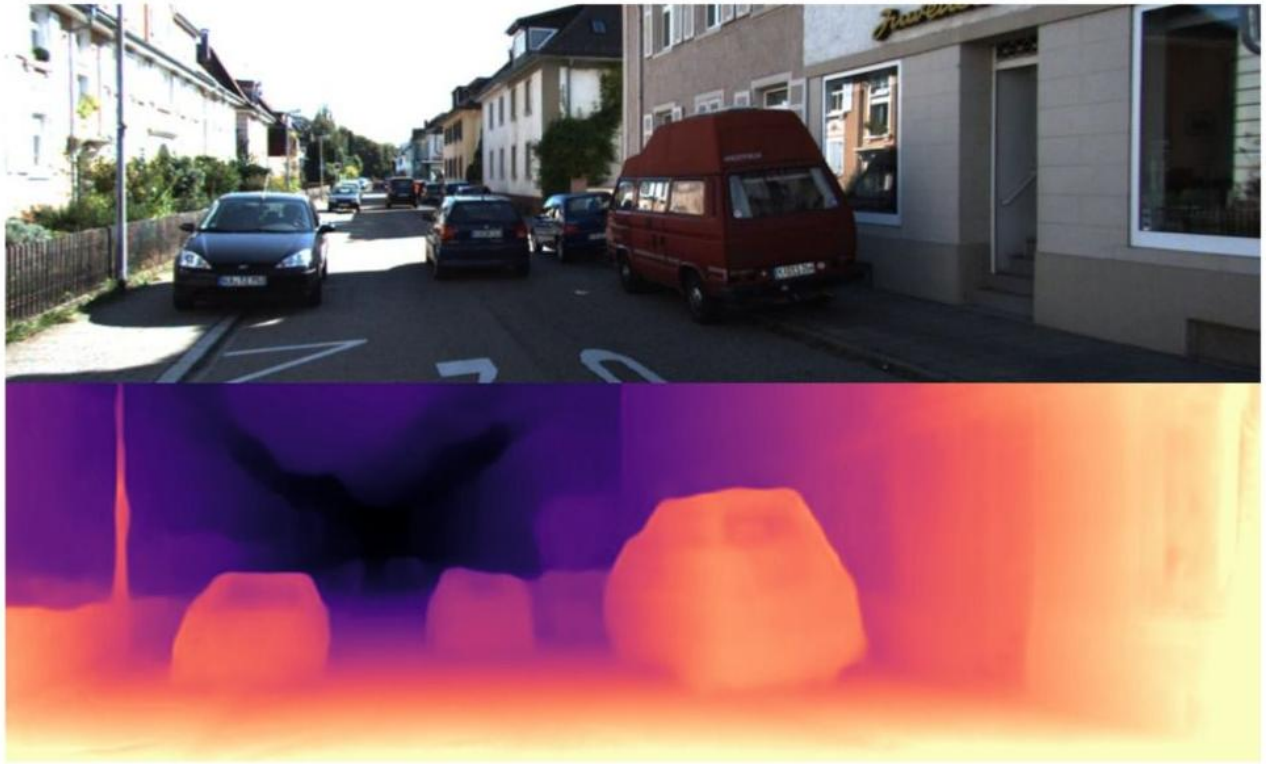
The shampoo bottle with a 2-in-1 label

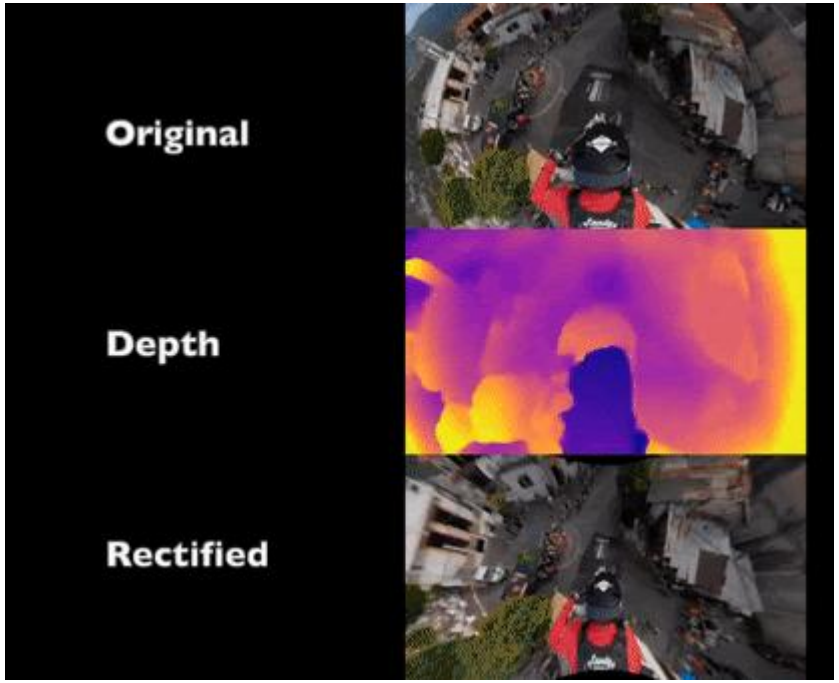


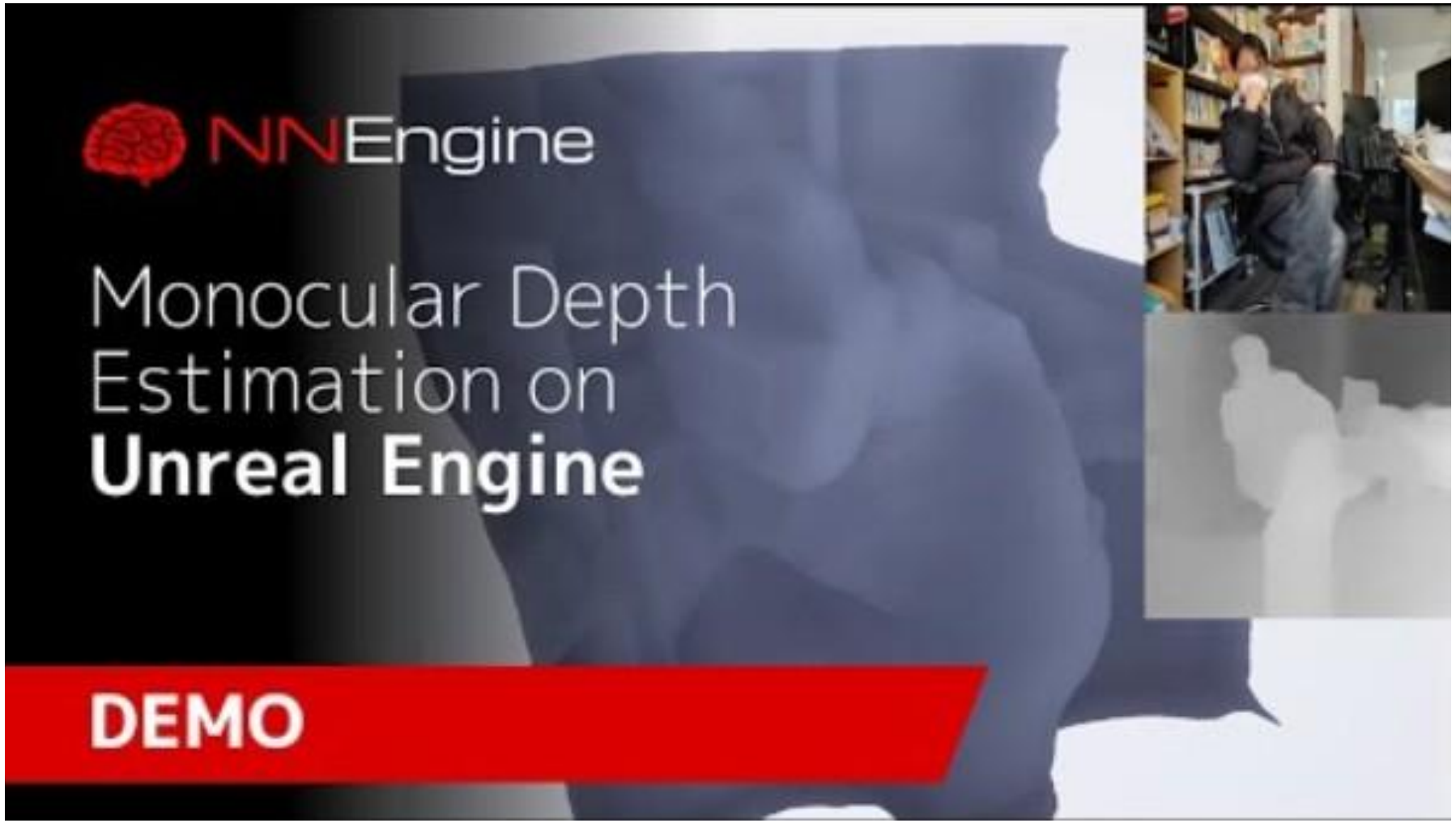
https://irvlutd.github.io/MultiGrounding/

Topic 6: Monocular Depth Estimation

- Estimating depth from images







https://youtu.be/glq34Wdi_3w

Topic 7: Proposal your own idea

- Needs to be approved by the instructor

- Needs to include these steps
  1. Data collection using the real camera
  2. Data labeling if supervision is needed
  3. Model training with the labeled data
  4. Model optimization using tensorRT
  5. Model deployment to Jetson Nano Orina
  6. Model evaluation on datasets
  7. Demonstration in classroom

Sign-up for Topics

- Topic 1 to topic 6, at most 4 groups for each topic
- Topic 7 (propose your own idea) needs to be approved
- Form a team and start thinking about what topic to work on
- We will have sign-up for topics soon

Questions?