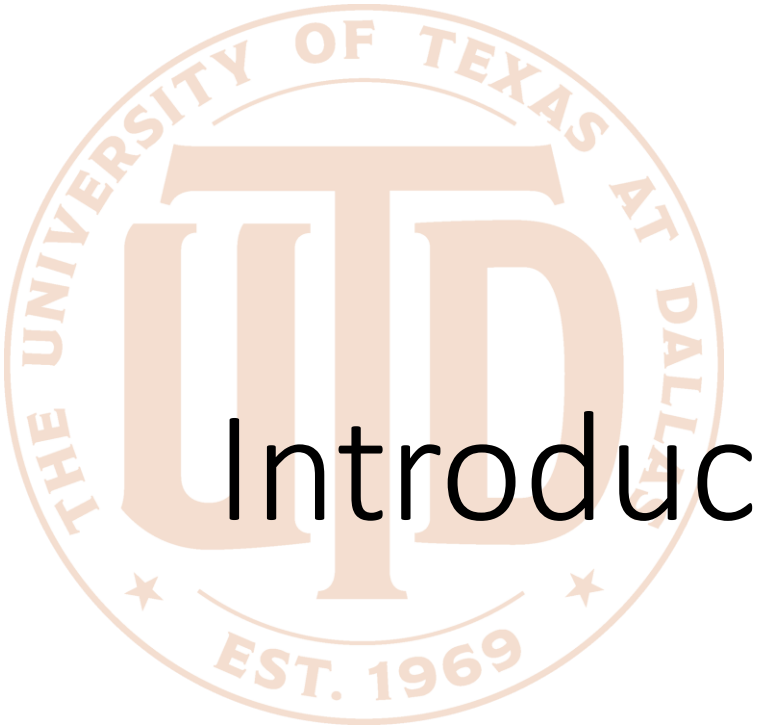


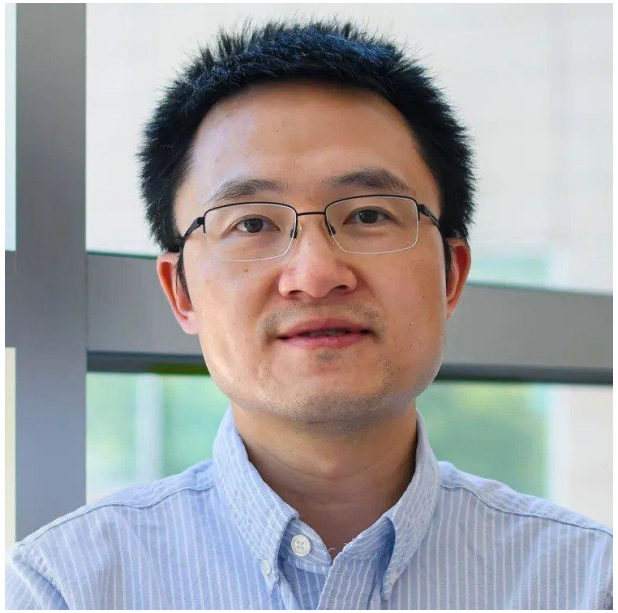
Introduction to Computer Vision

CS 4391 Introduction Computer Vision

Instructor: Yu Xiang

The University of Texas at Dallas

Who am I?



- Assistant Professor in CS at UTD (joined Fall 2021)
- Intelligent Robotics and Vision Lab (IRVL) at UTD

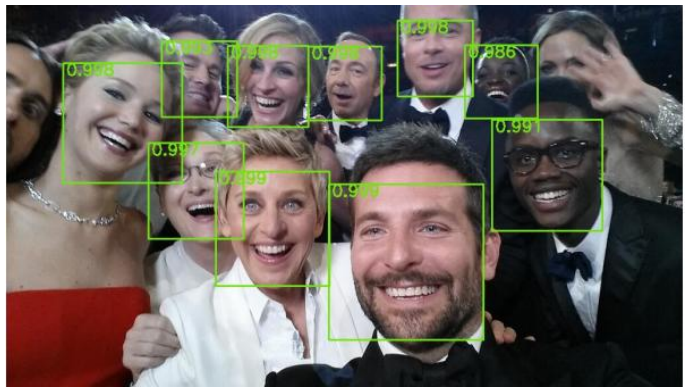
https://labs.utdallas.edu/irvl/

- Senior Research Scientist at NVIDIA (2018 – 2021) Robotics
- Ph.D. University of Michigan at Ann Arbor 2016

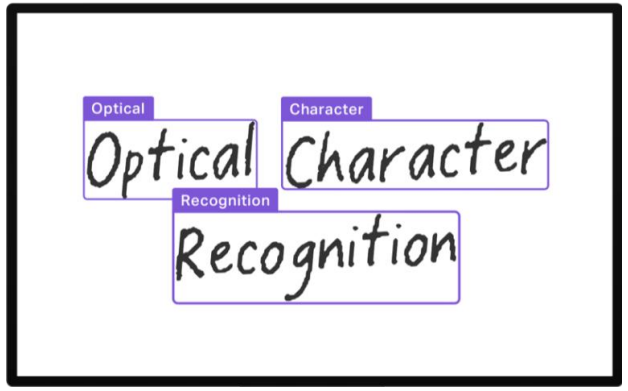
Introduce yourself

- Name
- Major program
- Which year in the program?
- What experience you have with computer vision?

What is Computer Vision?



Face Detection



Optical Character Recognition (OCR)

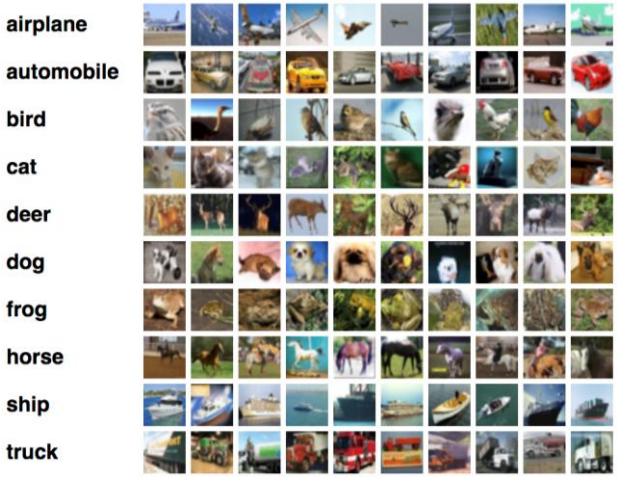
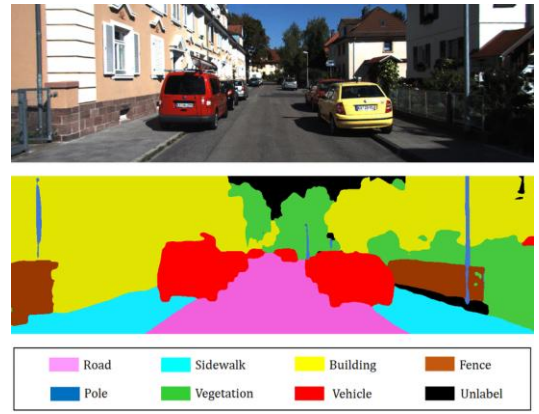


Image Classification

Panorama Stitching

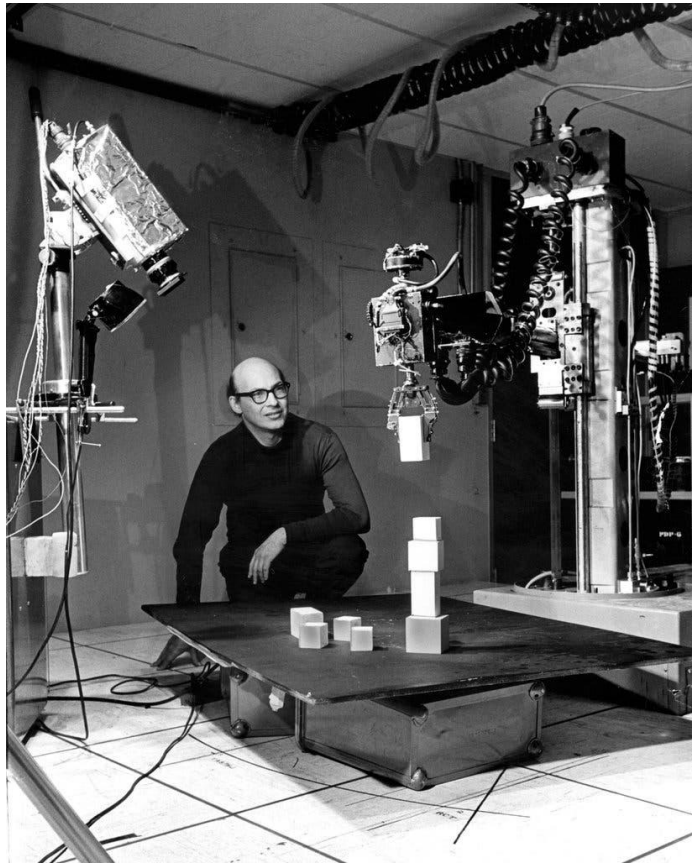
Surveillance



Semantic Segmentation

Computer vision is much more beyond image classification and processing

The Origin of Computer Vision



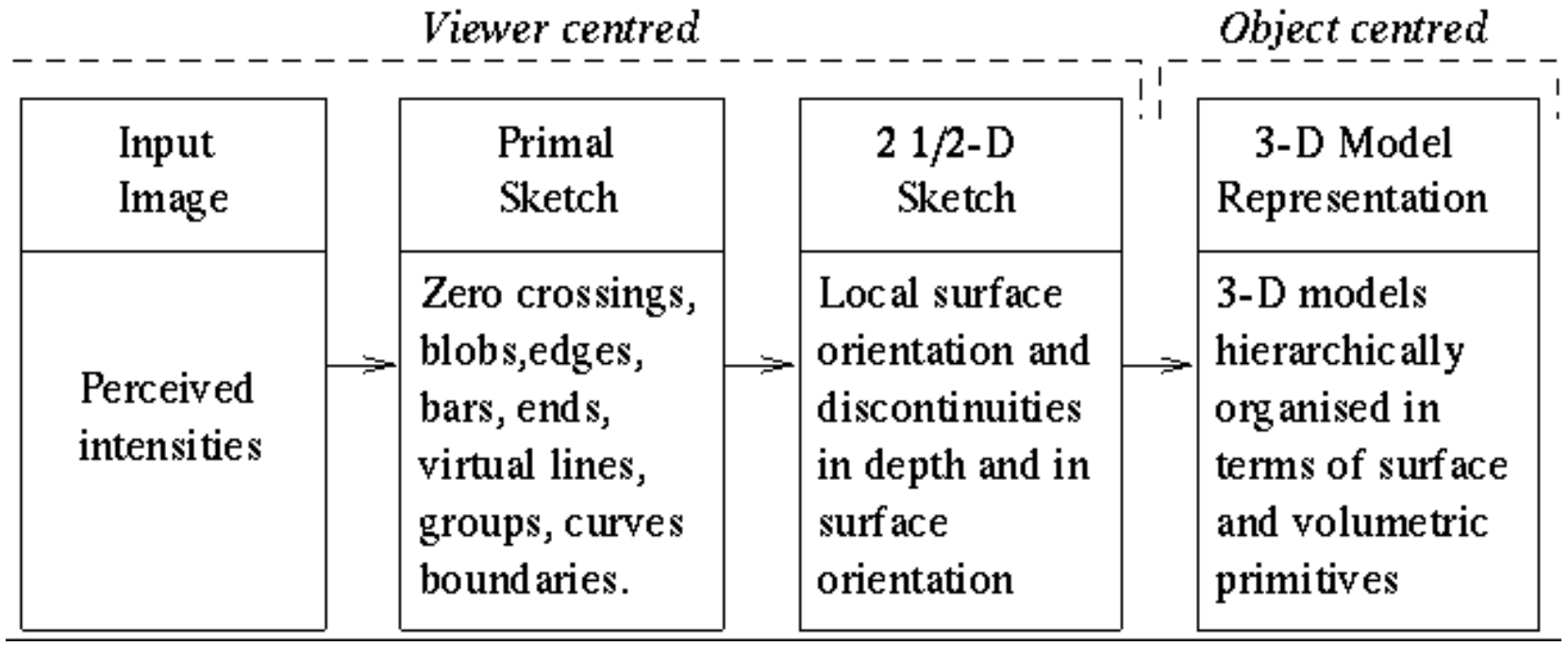
Marvin Minsky in a lab at MIT in 1968

An undergraduate project assigned by Marvin Minsky in 1966

“spend the summer linking a camera to a computer and getting the computer to describe what it saw”

Understand the 3D world from 2D images like humans

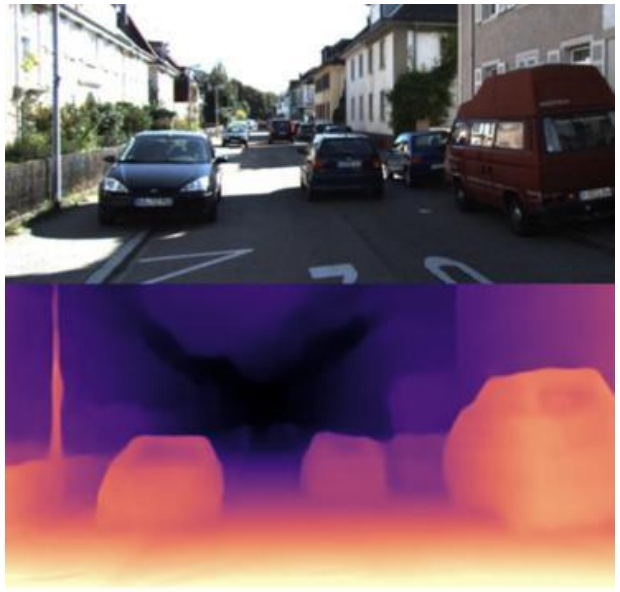
David Marr's Theory of Vision (Neuroscientist)



https://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/GOMES1/marr.html

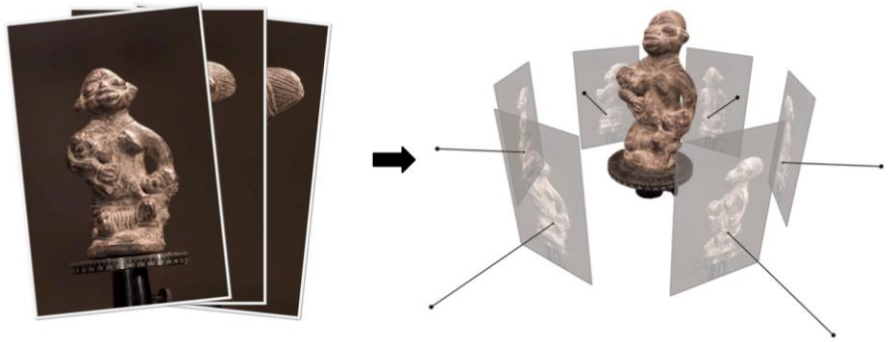
D. Marr. Vision. W. H. Freeman and Co., 1982.

What is Computer Vision?

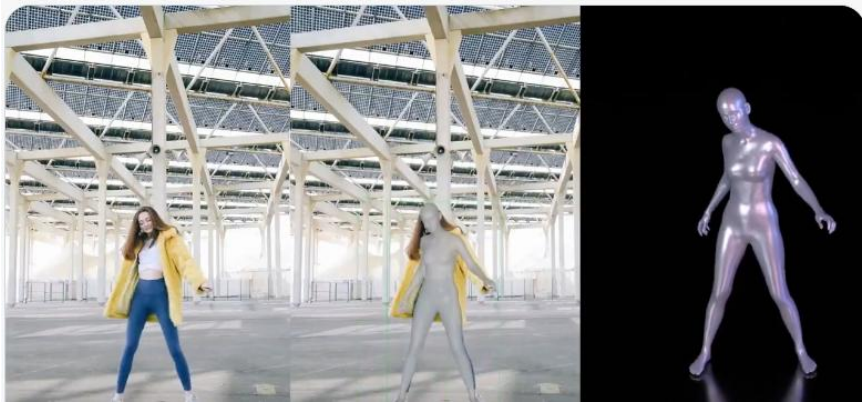


Depth Estimation

Structure from Motion



3D Reconstruction



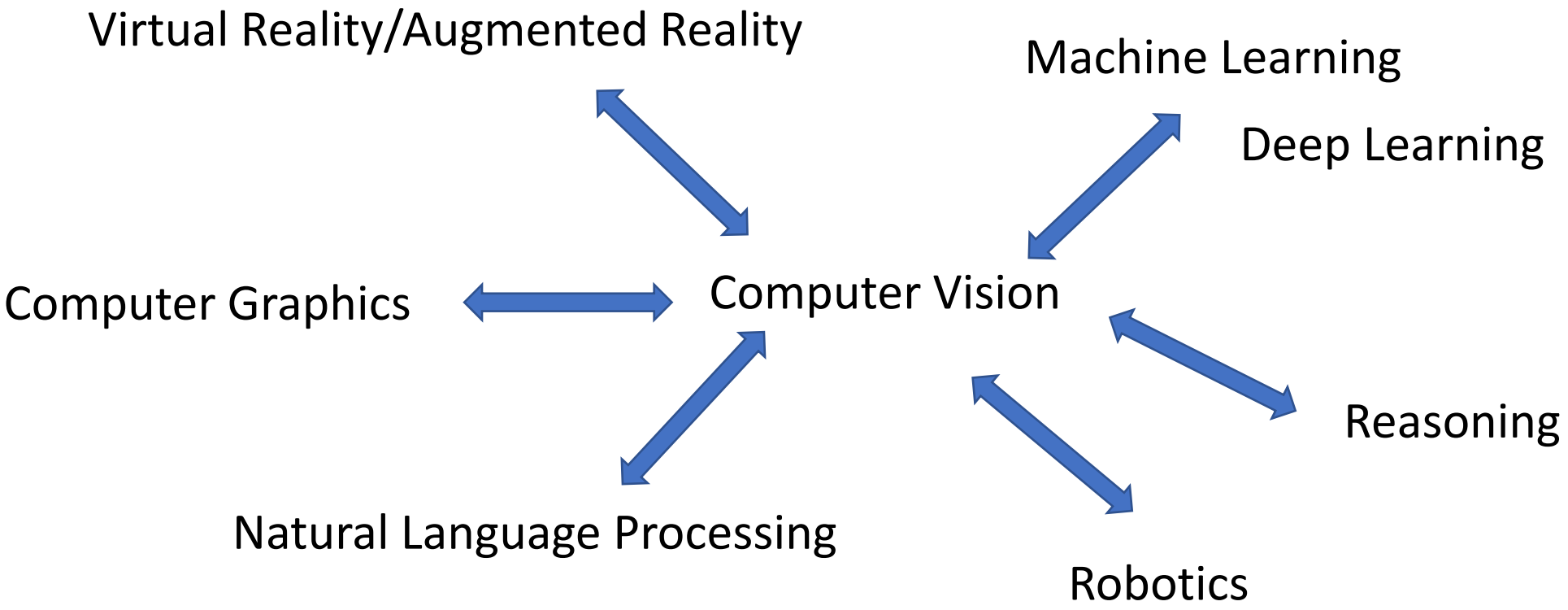
3D Human Pose and Shape Estimation

Understand the 3D world from 2D images

Vision-based Autonomous Driving

https://youtu.be/BBfjgfPOx58

Computer Vision in AI



Computer Vision in AI

Datasets $\neq$ Real World

Test your algorithms in the real world, e.g., with a camera

What will you learn in this course?

- Geometry in computer vision
  - Camera model, stereo geometry, multi-view geometry, etc.
- Deep learning in computer vision
  - Convolutional neural networks, recurrent neural networks, transformers, etc.
- Visual recognition
  - Object detection, semantic segmentation, etc.
- Hand-on programming with NVIDIA Jetson and cameras

Grading Policy

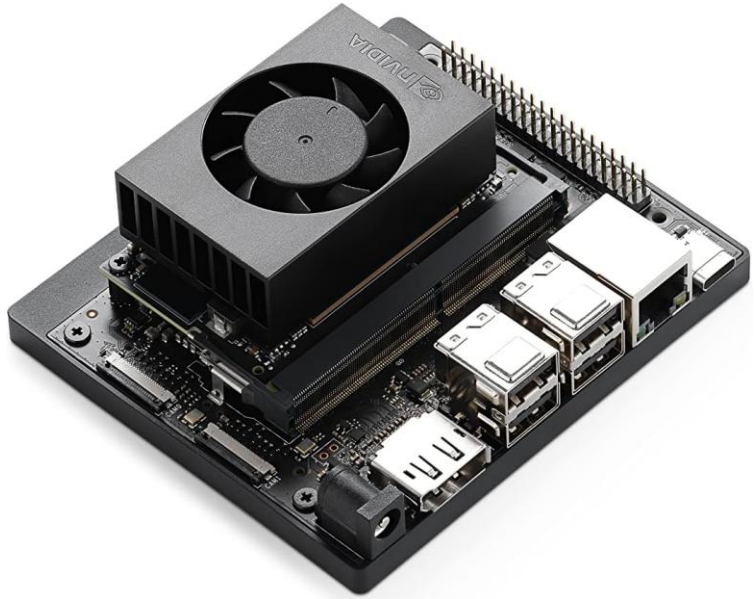
- Homework (50%): Individual submission
  - (10%) Homework #1
  - (10%) Homework #2
  - (10%) Homework #3
  - (10%) Homework #4
  - (10%) Homework #5
- Team Project (45%)
  - (10%) Project proposal
  - (10%) Project mid-term report
  - (10%) Project presentation
  - (15%) Project final report
- In-Class Activity (5%)

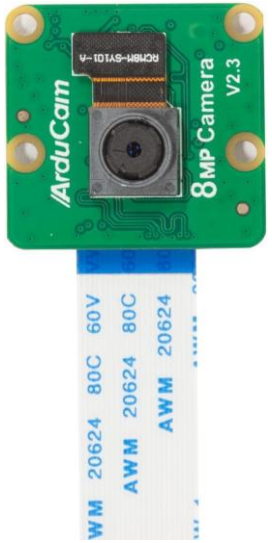
Team Projects

- 4 students for each team

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

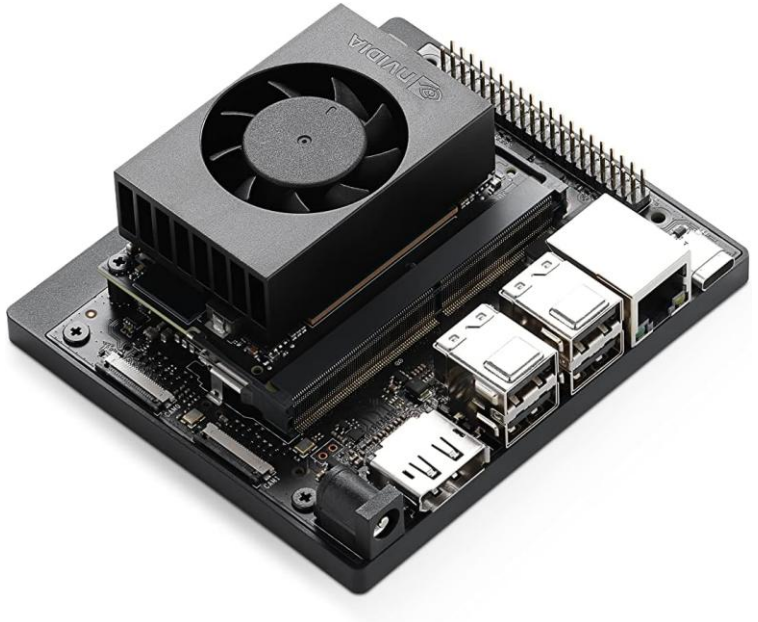
- Propose your own idea to work on
  - In-class demonstration in the final presentation





NVIDIA Jetson Orin Nano

- A **small, low-power GPU computer** for running AI models on the edge $249



- NVIDIA Ampere GPU, Up to **40 TOPS** AI performance
  - Desktop RTX GPU, 100 ~ 1000+ TOPS

- Linux-based (Ubuntu + CUDA)

JetPack ↔ CUDA mapping (typical)

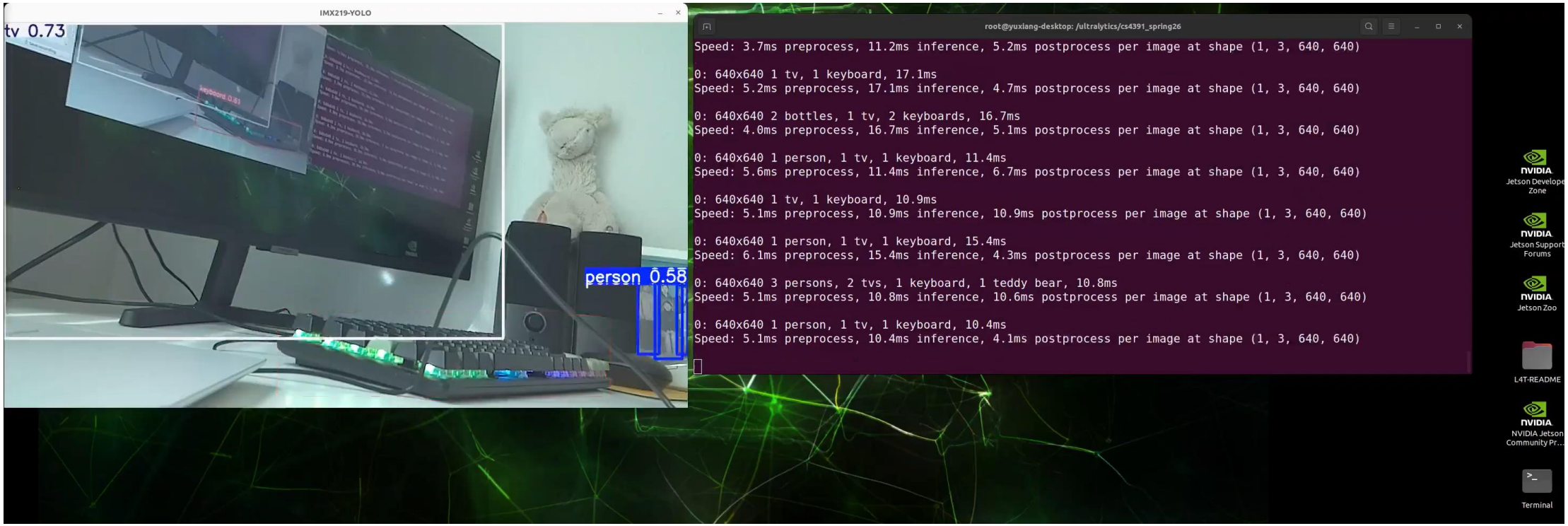
JetPack	Ubuntu	CUDA
JetPack 5.x	20.04	CUDA 11.x
JetPack 6.x	22.04	CUDA 12.x

Course Details

- Textbook
  - Richard Szeliski. **Computer Vision: Algorithms and Applications**. 2nd Edition. Springer. Available online https://szeliski.org/Book/
  - David Forsyth, Jean Ponce. **Computer Vision: A Modern Approach**, 2nd Edition. Pearson, 2011. (Optional)
  - Richard Hartley. **Multiple View Geometry in Computer Vision**, 2nd Edition. Cambridge University Press, 2004. (Optional)
- My office hour

  Tuesday & Thursday 3:00PM – 4:00PM
  ECSS 4.702
- TA office hour: TBD
- Course access and navigation: eLearning

Demo Time


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

Questions?