

CS 4391 Introduction to Computer Vision

Project Proposal Description

Instructor: Yu Xiang

February 10, 2026

1 Introduction

For the computer vision course project, students will use the following hardware to do a vision-related project:

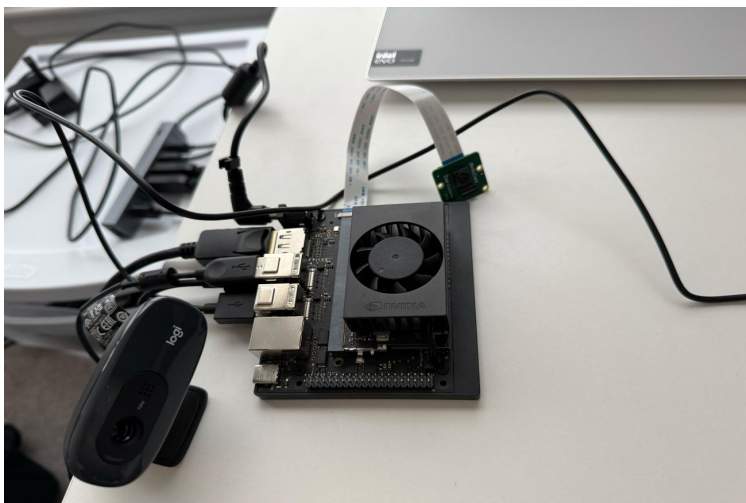


Figure 1: Hardware used for the course project

- **NVIDIA Jetson Orin Nano**
<https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-orin/nano-super-developer-kit/>
- **Logitech C270 HD Webcam**
https://www.amazon.com/dp/B004FH05Y6?ref=ppx_yo2ov_dt_b_fed_asin_title&th=1
- **Arducam 8MP IMX219 Camera**
https://www.amazon.com/dp/B082NVPC1N?ref=ppx_yo2ov_dt_b_fed_asin_title&th=1

Each project needs to perform the following stages:

- Data collection using the real camera

- Data labeling if supervision is needed
- Model training with the labeled data
- Model optimization using tensorRT
- Model deployment to Jetson Nano Orin
- Model evaluation on datasets
- Demonstration in classroom

2 Topics

2.1 Topic 1: Custom Object Detection or Tracking

This topic focuses on detecting or tracking specific object instances or categories.

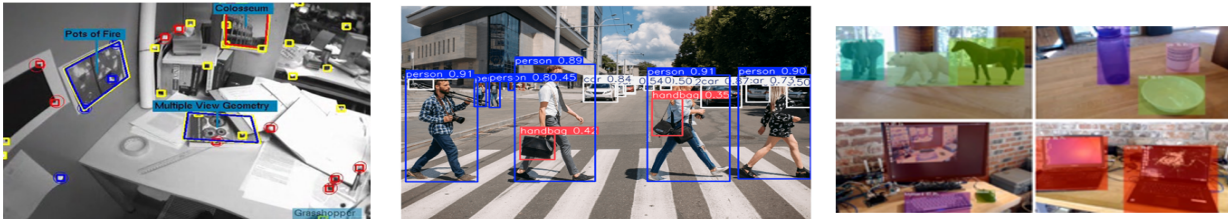


Figure 2: Object detection

Suggested model:

- Ultralytics YOLO <https://github.com/ultralytics/ultralytics>

2.2 Topic 2: Semantic Segmentation of Indoor Scenes

This topic focuses on labeling pixels into semantic classes.



Figure 3: Semantic segmentation

Suggested models:

- DeepLabV3+ <https://github.com/VainF/DeepLabV3Plus-Pytorch>
- U-Net <https://github.com/milesial/Pytorch-UNet>

2.3 Topic 3: Hand or Human Pose Estimation

This topic focuses on detecting or tracking human hands or human bodies.



Figure 4: Hand or body pose estimation

Suggested models:

- YOLO26 Pose Estimation <https://docs.ultralytics.com/tasks/pose/>
- MoveNet <https://github.com/fire717/movenet.pytorch>

2.4 Topic 4: Human-Object Interaction Recognition

This topic focuses on recognizing human actions from images or video clips.

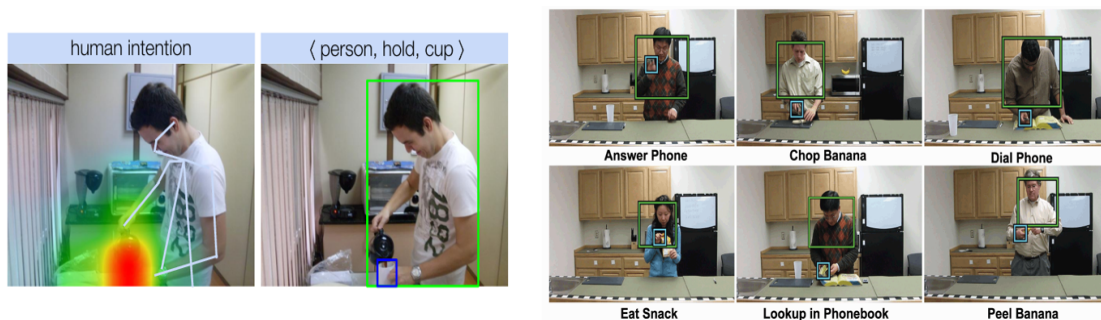


Figure 5: Human-Object interaction recognition

Suggested models:

- Ultralytics YOLO to detect both hands and objects <https://github.com/ultralytics/ultralytics>
- Hand object detector https://github.com/ddshan/hand_object_detector

2.5 Topic 5: Visual Grounding

This topic focuses on localizing a target object in a given image according to a language description.



Figure 6: Visual grounding

Suggested models:

- GLIP <https://github.com/microsoft/GLIP>
- Open-GroundingDino <https://github.com/longzw1997/Open-GroundingDino>

2.6 Topic 6: Monocular Depth Estimation

This topic focuses on estimating depth values from images.



Figure 7: Monocular depth estimation

Suggested models:

- Depth-Anything-V2 <https://github.com/DepthAnything/Depth-Anything-V2>
- ZoeDepth <https://github.com/isl-org/ZoeDepth>
- Monodepth2 <https://github.com/nianticlabs/monodepth2>

2.7 Topic 7: Proposal your own idea

You can propose your own idea of the project, but **it needs to be approved by the instructor.**

3 Proposal Submission (8%)

The project proposal should be prepared using the CVPR latex template. A useful online LaTeX tool is Overleaf <https://www.overleaf.com/>. We have the CVPR latex template accessible here via overleaf: <https://www.overleaf.com/read/gpjssbtrrpqm>. You can download a copy of the template or make a copy in overleaf for your own project, and then edit it.

The project proposal should be a **1-page PDF** using the latex template with the following items:

- **Title.** Let's give a name to your project.
- **Team Members.** List the names of the team members as the authors in the proposal. We expect you to work in **groups of 3-4 students** for the course projects.
- **Problem Statement.** Describe what is the problem you are trying to solve in this project. **Select one of the topics above.**
- **Approach.** Describe what is your plan to solve the problem. It is fine if some details have not been figured out in the project proposal. But students should have rough ideas on how to proceed.
- **Data.** Describe what data you will collect for training and evaluation.
- **Evaluation.** Describe how to evaluate the success of the project. For example, what evaluation metrics will be used to evaluate the performance of the method?
- **References.** Cite related works in the proposal.

Please submit your pdf file to eLearning.

4 Proposal Video Submission (2%)

Please record a video of the NVIDIA Jetson Orin Nano kit connecting to the Logitech C270 USB camera. Follow Lecture 5 to set it up and run the python script:

- `python test_usb_camera.py`

The video needs to show the camera is correctly launched and getting video frames in the Python window. **Please upload a .mp4 video to eLearning.**

5 Deep Learning Resources

Most recent vision methods leverage deep learning to train neural networks to tackle various problems in computer vision. If your project requires training of deep neural networks, you may

need to have GPUs for training. **Google Colab** is a great free resources for small amounts of GPU resources: <https://colab.research.google.com/>. Two widely-used deep learning frameworks:

- PyTorch <https://pytorch.org/>
- TensorFlow <https://www.tensorflow.org/>