Hand Gesture Interface

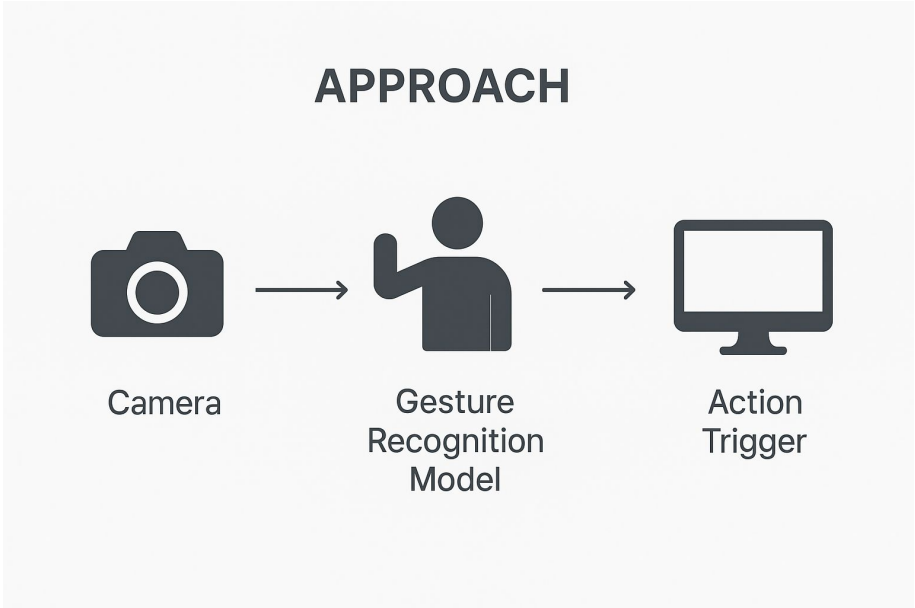
Aniket Patel, Matthew Kasper, Matej Popovski

Why is this problem important?

- A gesture-based computer interface can provide a more personalized and comfortable method of interacting with computers for those suffering from arthritis, carpal tunnel syndrome, or other conditions that limit hand mobility.
- This system can also make Interaction with some types of computers, such as VR/AR systems and smart TVs, more convenient for a general user when compared to traditional control methods.

Approach

- YOLO
- Training Data
- Real Time Detection and Computer Actions



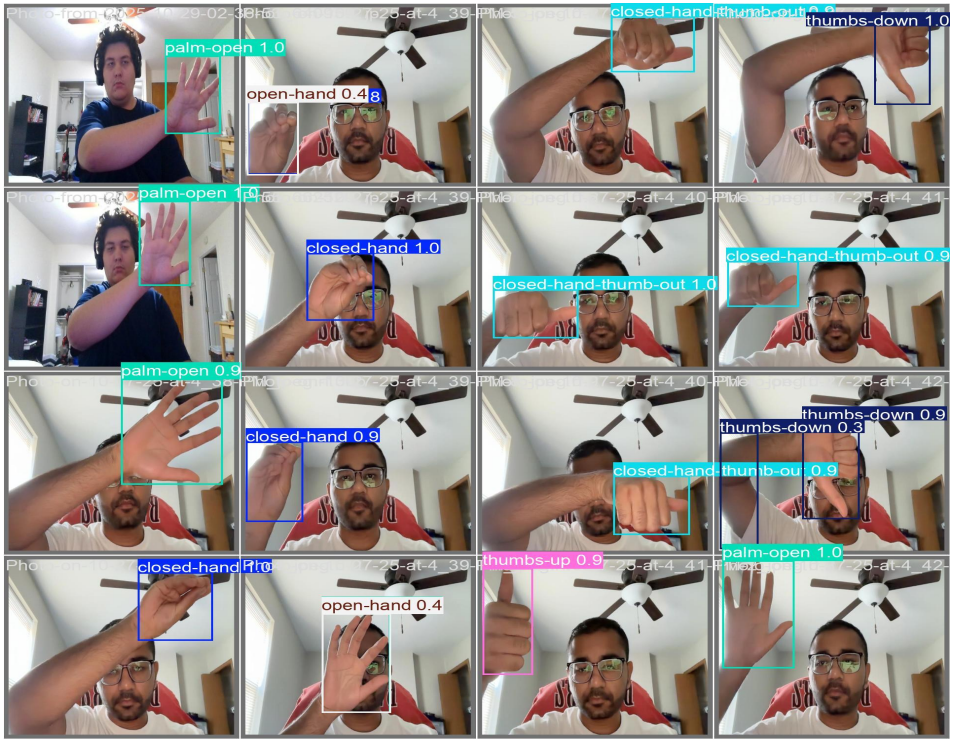
YOLO (You Only Look Once)

- YOLO is one of the leading object detection models out there.
- Its key strength is speed
- It’s widely used in applications such as autonomous systems, video analytics, and robotics.

https://docs.ultralytics.com/tasks/detect/

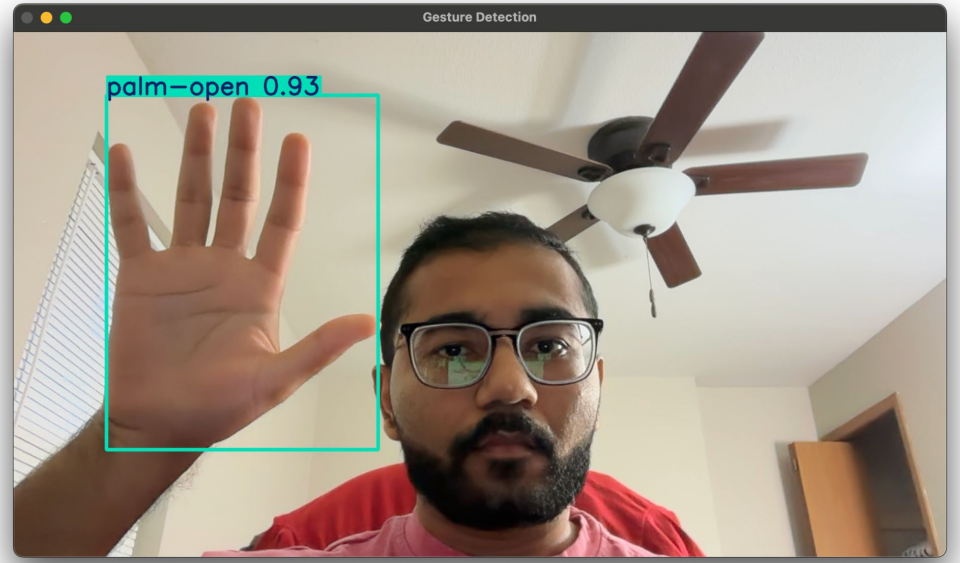
Training Data

- Custom labeled dataset of 445 images
- Augmented the custom dataset by randomly changing hue, brightness, rotation, etc. to scale the dataset to 1071 total images.
- Fine-tuned the YOLOv12n model with our custom dataset.



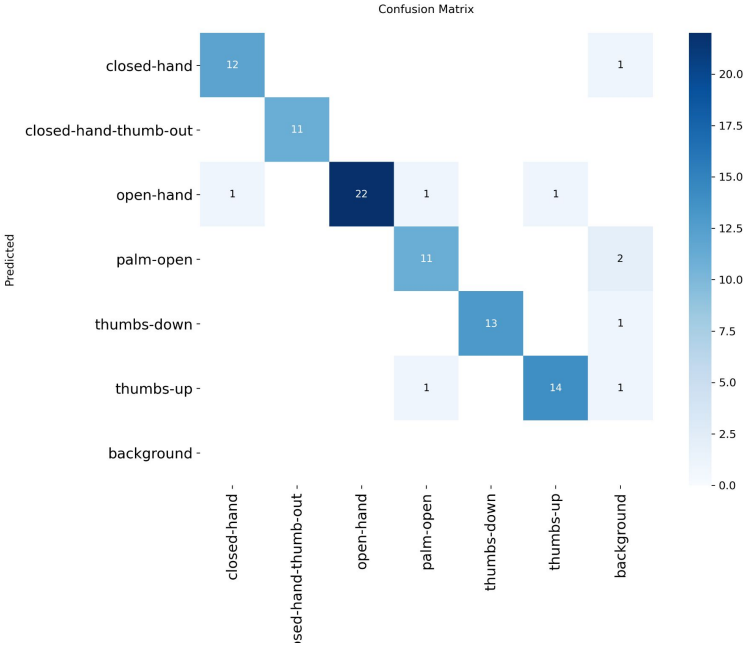
Real Time Detection and Computer Actions

- Created a live feed with real time gesture annotations.
- Mapped gestures to actions using the pynput library
- Actions include: volume up/down, pause/play, skip back, and mouse movement/click.



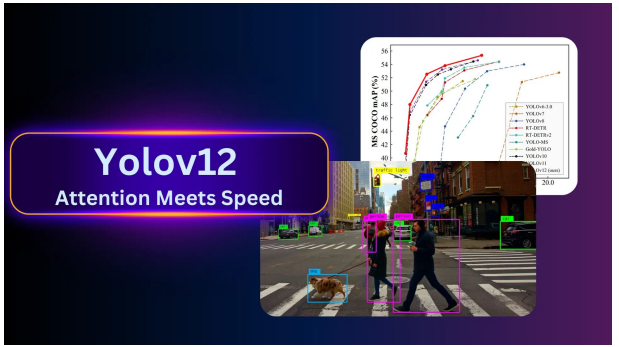
Training Results

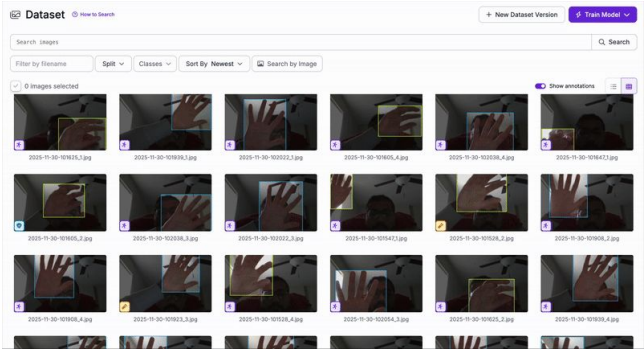
- Fine-tuning YOLO from an existing, robust model allowed us to achieve good results for our special gesture cases with a relatively small dataset.
- In addition, YOLO has very high performance during inference, allowing our program to run in real time on standard personal computers.



What We Learned

- From this project, we learned how crucial data quality is for real-world computer vision. Building our own dataset showed us that imbalance, inconsistent gesture execution, and limited variety can significantly hurt model performance.
- Finetuning YOLOv12 also taught us how powerful transfer learning is—allowing us to get strong results quickly—but it also made clear that real-time systems require robustness, not just accuracy on a test set.
- Creating our own dataset taught us how hard good labeled data is to obtain—manual annotation is slow, small datasets cause confusion between similar gestures, and simply balancing class counts isn't enough—lessons we wouldn't have learned using an existing dataset.





Where This Could Lead in the Future

This work can lead to a fully hands-free computer interface that’s accessible, low-cost, and flexible. With more diverse data, better augmentation, and potentially combining YOLO with landmark-based methods, the system could be used for accessibility tools, AR/VR interaction, smart TV control, or general gesture-based computing using only a standard webcam.