Egocentric VR Hand Tracking

Hershed Tilak and Royce Cheng-Yue

Department of Electrical Engineering, Stanford University

Motivation

Having the ability to interact with a virtual environment is one of the most challenging tasks in VR. Glasses-bound VR companies have looked into controllers to interact with an environment, but this method is not natural.

Through this project, we aim to create a real-time egocentric hand tracking solution that could be integrated into a head mounted display.

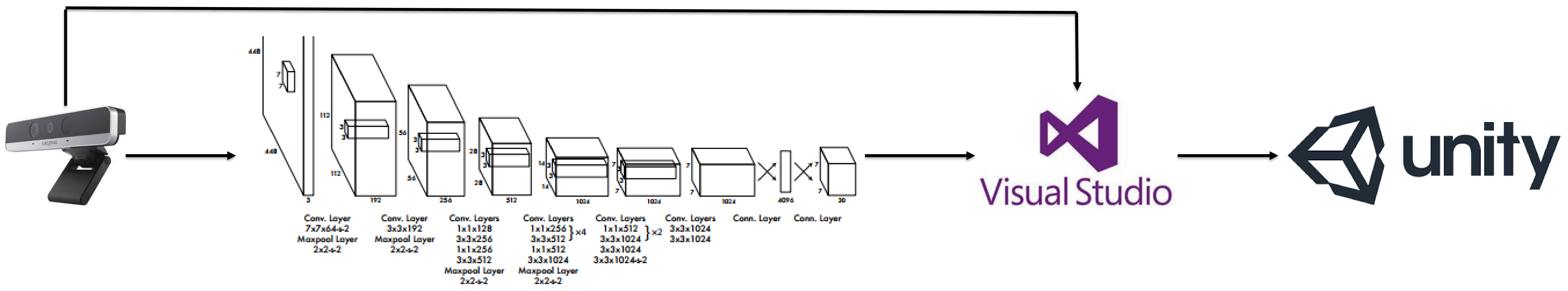
Related Work

We generated our dataset from the EgoHands and General-HANDS datasets. In total, we have 2,829 images for training, 403 images for validation, and 807 images for test.

After 2013, CNNs have become popular in object detection. Specifically, Faster R-CNN is a popular approach that involves generating region proposals and regressing within each proposal. Another approach is YOLO, which treats the overall image as a regression problem.

We plan to use YOLO, which achieves 63.4% mAP at 22 ms latency on the VOC 2012 test set.

Hand Tracking Pipeline



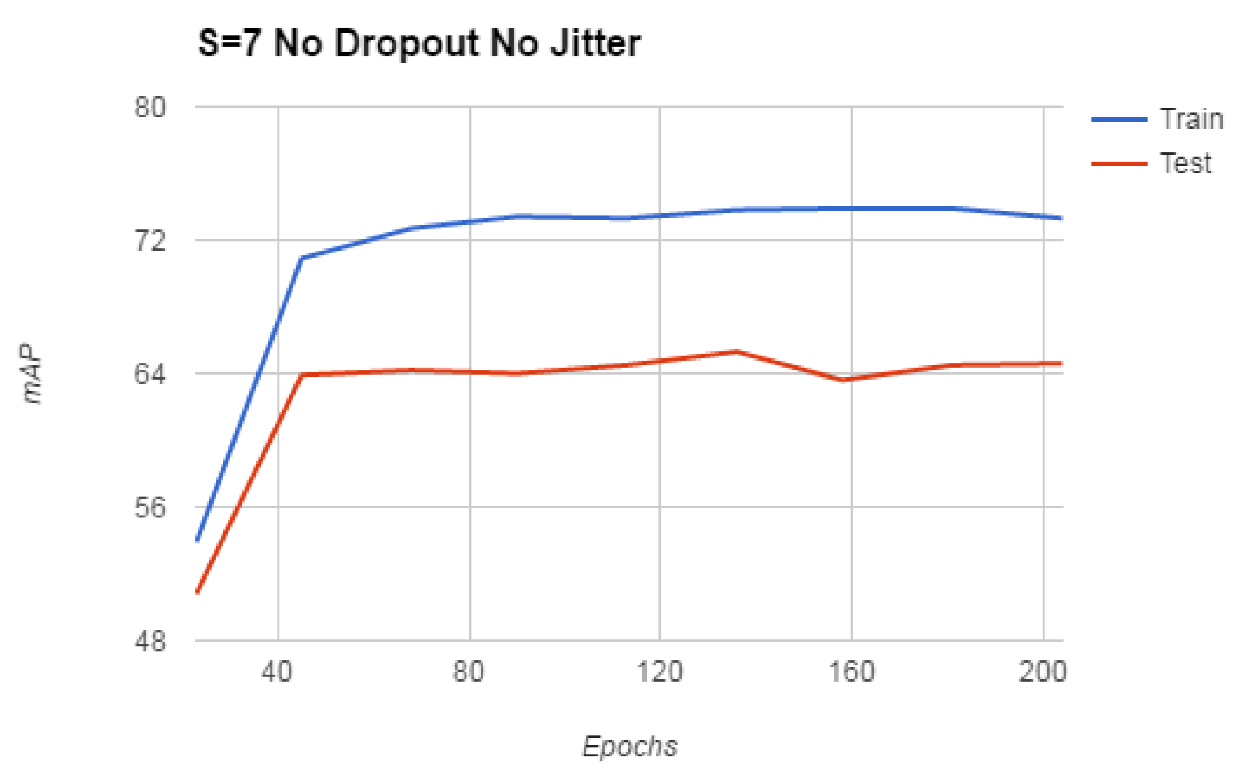
- Acquire RGB and depth images from the Intel RealSense
- Send RGB images to YOLO

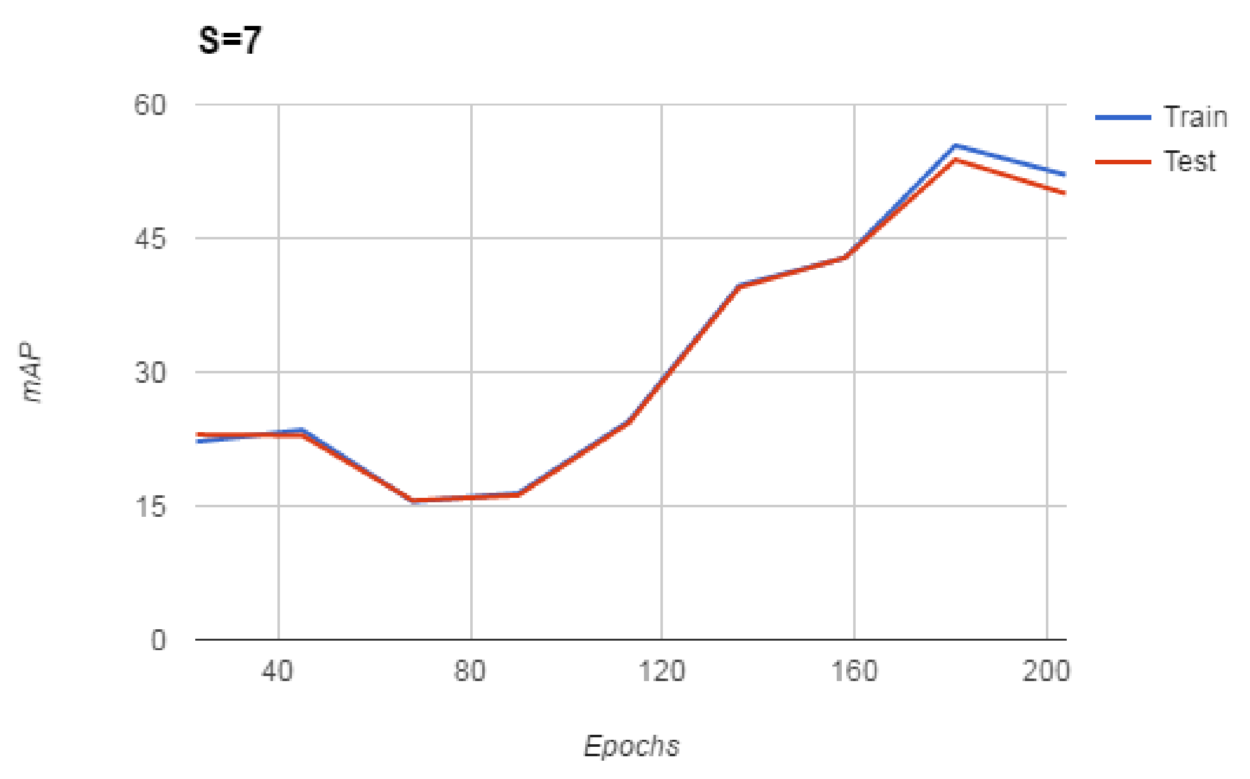
- Resize 640 x 480 to 448 x 448
- Add downsampling layer to fit model into GPU
- Replace last fully connected layer to be compatible with dataset
- Divide input image into an S x S grid
- Each cell regresses on 2 bounding box coordinates and confidences, as well as hand probability
- Looks at global features to determine bounding boxes of hands

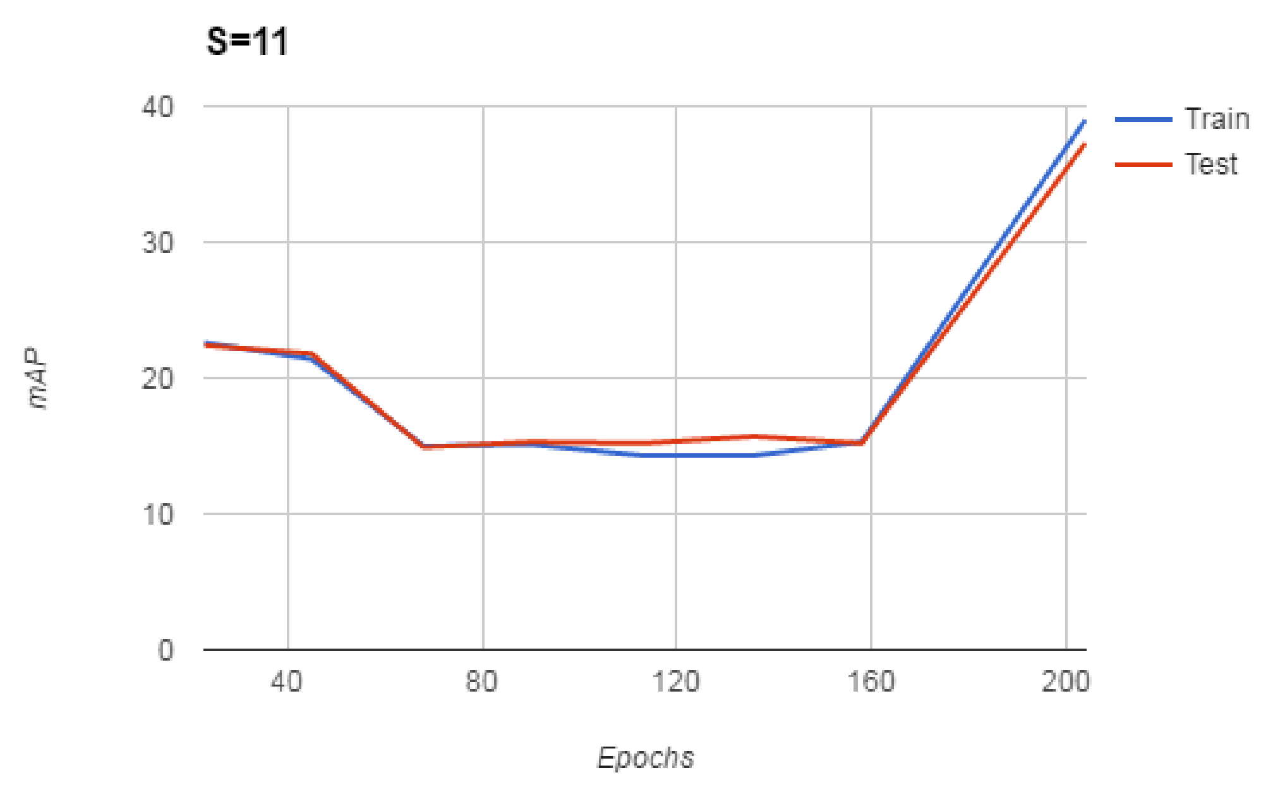
- Use YOLO output as x and y positions
- Extract z position from depth image
- Send data to Unity

- Update rendered hands in virtual environment
- Determine object collisions in VR game

Evaluation

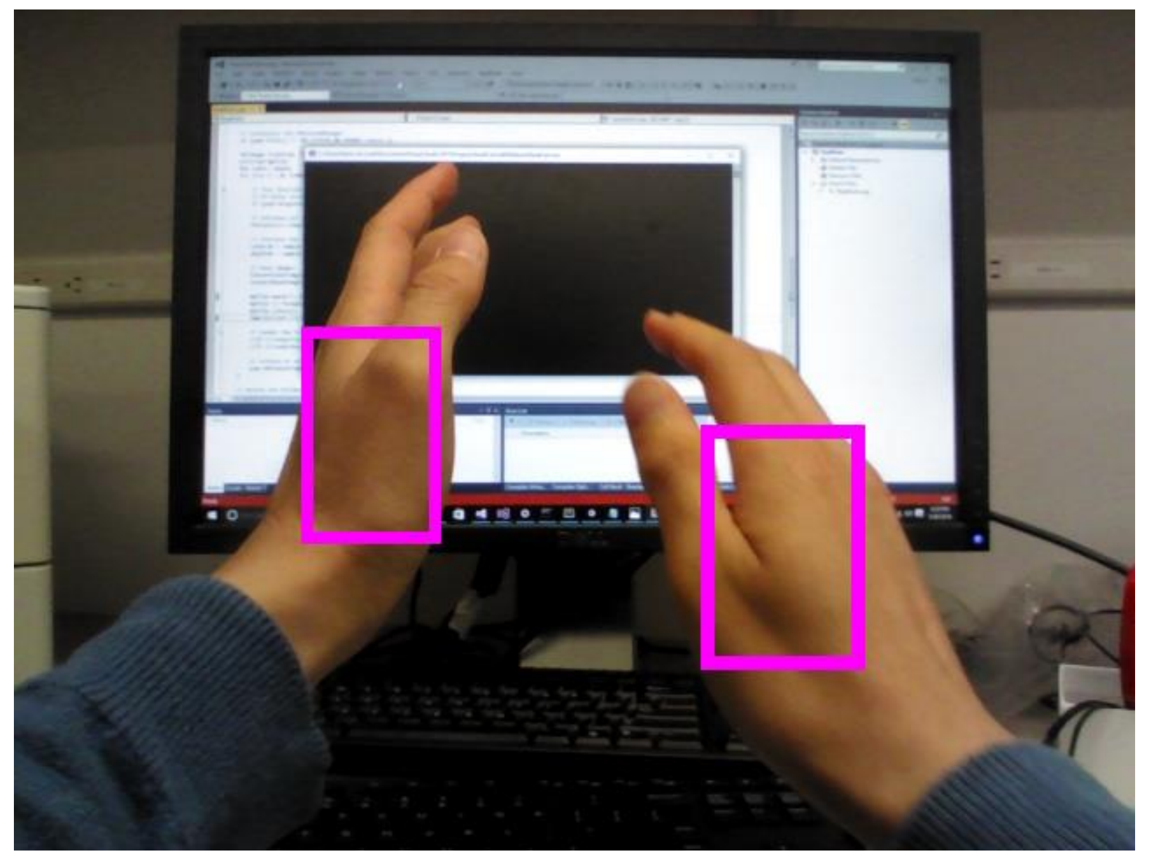




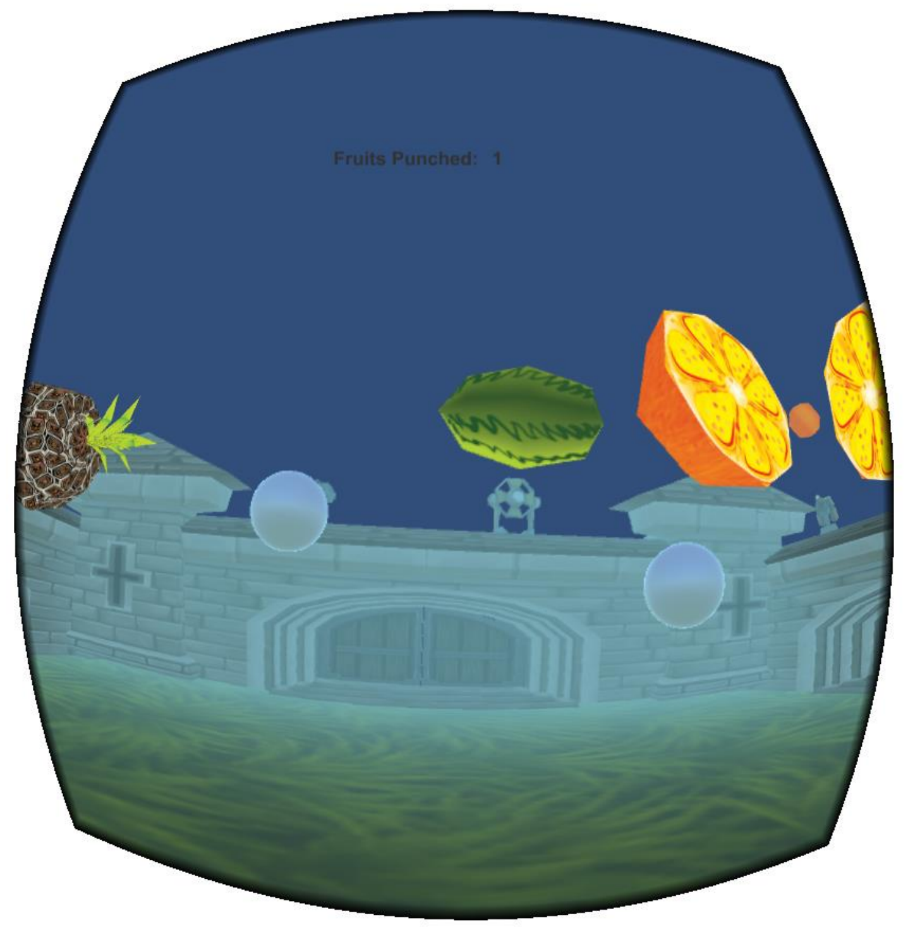


- Trained three YOLO models: S=7 with no dropout or jitter, S=7 with 0.5 dropout and jitter, S=11 with 0.5 dropout and jitter
- S=7 No Dropout No Jitter achieves best mAP but slightly overfits
- S=7 and S=11 models should outperform the S=7 No Dropout No Jitter model with more training

Experimental Results



Output of YOLO Neural Network



Rendered Hands in VR Environment