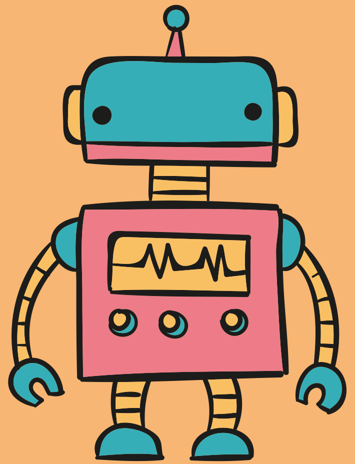
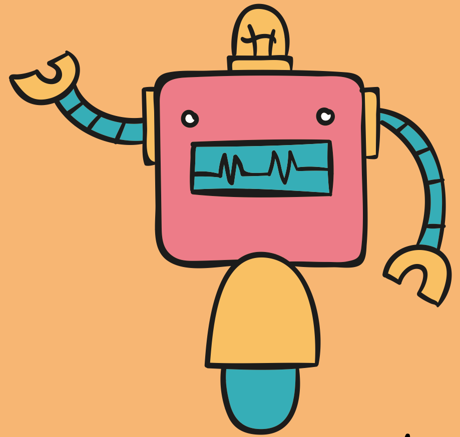
Cozmo Robot
Depth Map

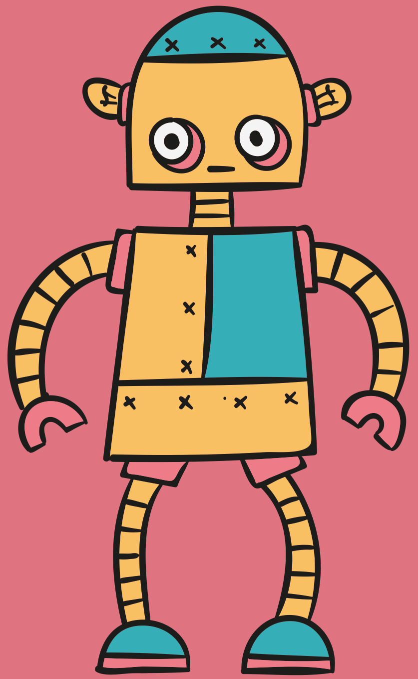




Maxwell Jones & Akshath Burra

Problem

We look to take a cozmo view and then return a mapping of how far away each pixel in the image is from the robot. This is formally known as a Depth Map.



The Approach

Relative Depth Computation

Use SOTA relative depth model to predict depth from a single camera image

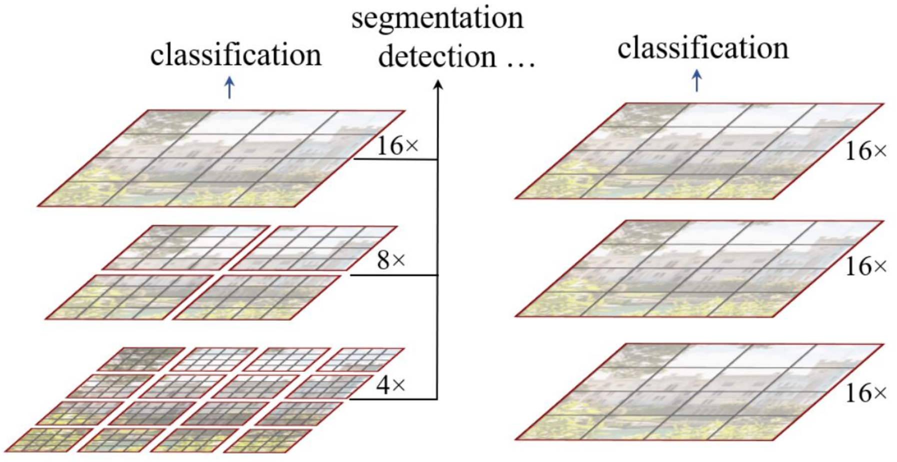
Cube Location Determination

Calculate pixel-wise location of cubes as well as cube distance from camera

Depth Map Calibration

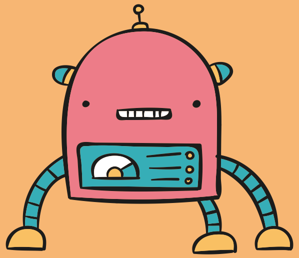
Find mathematically optimal scaling factor for depth map given cube locations and apply it

Swin Transformers

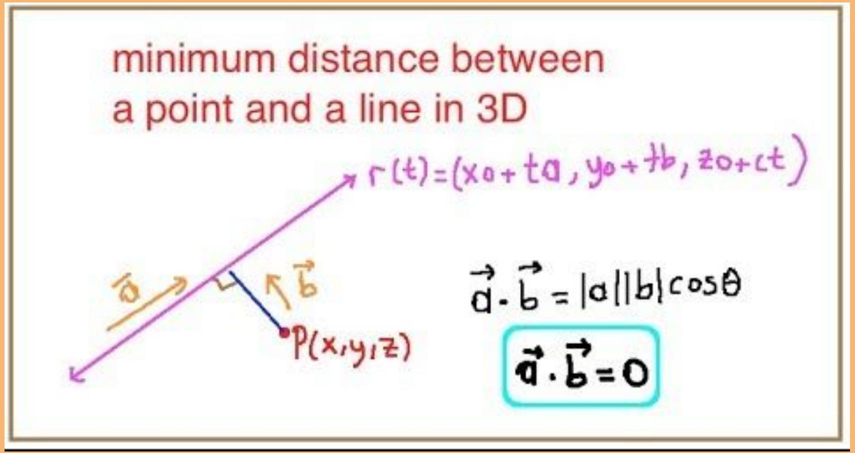


(a) Swin Transformer (ours)
(b) ViT

Swin Transformers allow for smaller image patches for self attention without increased time complexity. They accomplish this by reducing number of self-attention heads and restricting the self-attention connection to certain regions at lower layers. See here for more details

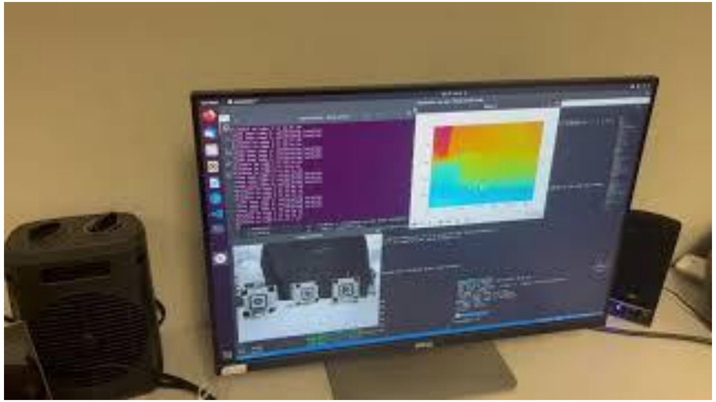


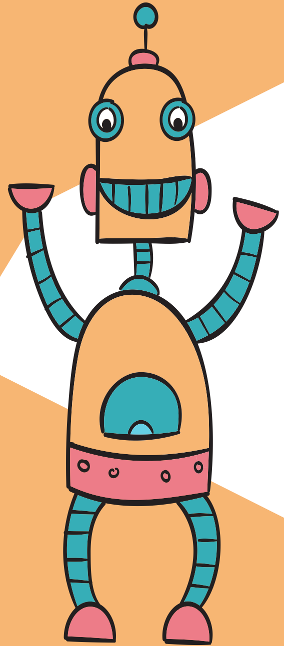
Optimal Depth Scaling Calculation



Let (x, y, z) be the distances of the cubes. The relative depth map has values (a, b, c) for the cubes. We want to find some t such that (ta, tb, tc) is close to (x,y,z). This boils down to finding the point on the line (tb, tb, tc) closest to (x,y,z). See here for a full proof in n dimensions of the optimal constant c

Final Results





Demo video link here

Result Analysis

Successes

Able to calibrate depth based on cube locations, query depth at any pixel in the image, and produce more accurate scaling factors as cube number increased

Weaknesses

Scaling factor unable to completely account for inaccuracies in depth prediction, poor image quality yielded pixel-wise depth maps that were slightly inaccurate

Future Extensions

Smarter Interpolation

Instead of using a scaling factor to adjust depth, find some smarter method to change the depth map that respects all cube positions better

Super Resolution

Use some super resolution mechanism to improve the quality of the camera images produced by cozmo

Movement Allocation

Use cozmo's movement across the world to have better pixel-wise depth estimates (structure from motion)

Thank You!

Any Questions?