Homework 2

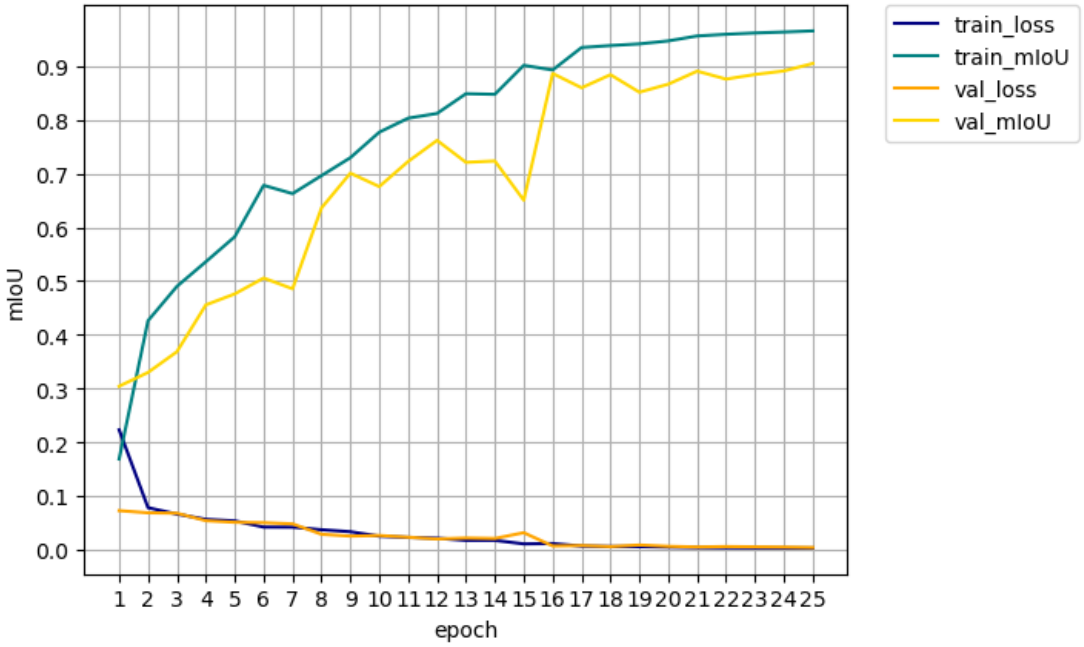
Problem 2

1. The instance segmentation problem consists of a training a model to segment certain classes of objects, meaning it was classify pixels to a certain color corresponding to the coded color of the object that the pixel belongs to. The model will take RGB-D images as input, corresponding ground truth masks as target labels, and will output the segmentation of the inputs with as a close accuracy to the target masks. In this way, this is a supervised learning problem given the training has ground truth masks as targets.

2.

3.

4. 1x1 kernel size is used for dimensionality reduction to reduce the number of channels. It captures the interaction of input channels in just one pixel of feature map. Therefore, it makes sense to use it right before the output. 3x3 kernel size for the preceding convolutional layers works well since it can extract useful information about each pixel using the local neighboring pixels.
   The 6 output channels corresponding to the classification classes for each pixels. In this case, there are 6 different classes that each pixel could be predicted as that correspond to the segmentation taskid name color: 0 background black, 1 sugar box red, 2 tomato soup can green, 3 tuna fish can blue, 4 banana yellow, 5 bowl purple. 1 channel ground truth masks store information about a certain class rather than multiple classes.



5. Learning curve for best model checkpoint shown above

Problem 3

1. When tuning the trimesh icp function, I initially thought that a higher number of max iterations and a lower threshold would intuitively yield better results. However, I found that beyond a limit, increasing max iterations made negligible improvements to the Average Closest Point Distance (ACPD) metric and instead caused the metric to worsen as well as wasted computational resources. Qualitatively, increasing max iterations also displayed larger variances between the predicted and gt masks in meshlab. I found that the optimal max iterations was 100. Similarly, the threshold produced similar results and I found that the optimal threshold hold was 1e-04. For the initial transformation I used the trimesh procrustes method and this greatly improved the ACPD metric and gave closer matches in meshlab.

2. One thing to note. For 1_gt, 1_pred, the ACDP could not be calculated reliably since my model failed to classify any pixels as belonging to the yellow banana class. Thus, the there was no metric for the banana class. This is also qualitively illustrated in meshlab, with no banana point cloud shown in 1_predmask_transformed.ply.