

Homework 4

Notes you want the TAs to consider when grading.

First, since I am on windows, I have install the packages through visual studio.

Second, my computer is a bit old now, it takes a relatively longer time to do the pybullet simulation. The videos are therefore a bit longer. I use google meet to record them.

Third, some of my image doesn't fit into the file, I will concatenate them and place them. Thank you!

Problem 2

Problem 2-a

Why do we use gaussian scoremap to generate the affordance target instead of a one-hot pixel image?

The main reason is that it allows us to represent the uncertainty in the grasp point location. The scoremap will have probability distribution around the actual grasp point, the higher the score, the closer it is to the grasp point. On the other hand, one-hot pixel is less practical to assume that the model can always detect the location of grasp point.

Problem 2-b

The aug pipeline randomly transform the input image during training. It consists of a random translation and a random rotation. The random translation is controlled within -20% to 20% of image size while the random rotation is controlled within (angle delta/2, angle delta/2)

Problem 2-d

My best checkpoint has validation (test) loss 0.0009 at epoch 100. At the end of 101 epochs, the train loss is 0.0012 and test loss is 0.0010

See Figure 1

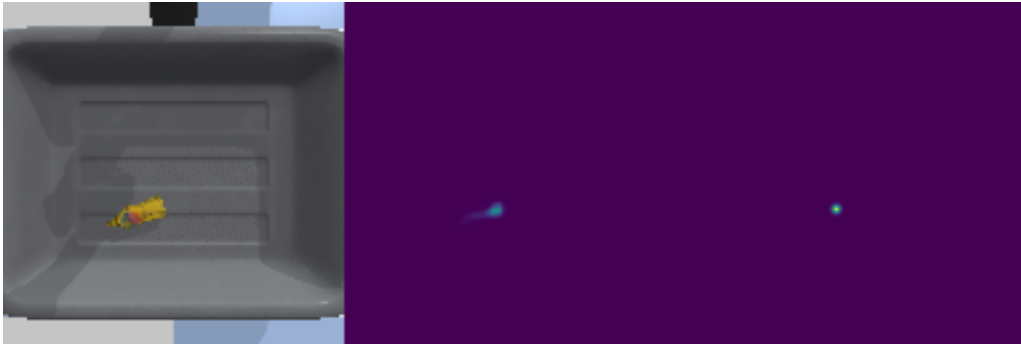


Figure 1: affordance training vis 003

Problem 2-f

Recorded Video: <https://youtu.be/rZyPn2GnnC4>

Success Rate: 73%

See Figure 2

Note: I have concatenate the original image of the Mustard Bottle to fit into the file

Problem 2-g

Recorded Video: <https://youtu.be/gcqLxu4V8SY>

Success Rate: 63%

See Figure 3

Visualization comparison: Since Banana is not in the training set, the model has never seen it before. However, it is still able to capture a general shape.

Problem 2-h

Recorded Video: https://youtu.be/1uk_MNigL9U

Success Rate: 10 picked up, 5 left

Problem 2-i

One of the reason of why this method is sample efficient is that we use augmentation to increase the dataset size. Another reason is that visual affordance model use neural network to learn the patterns that may be generalizable even to unseen objects.

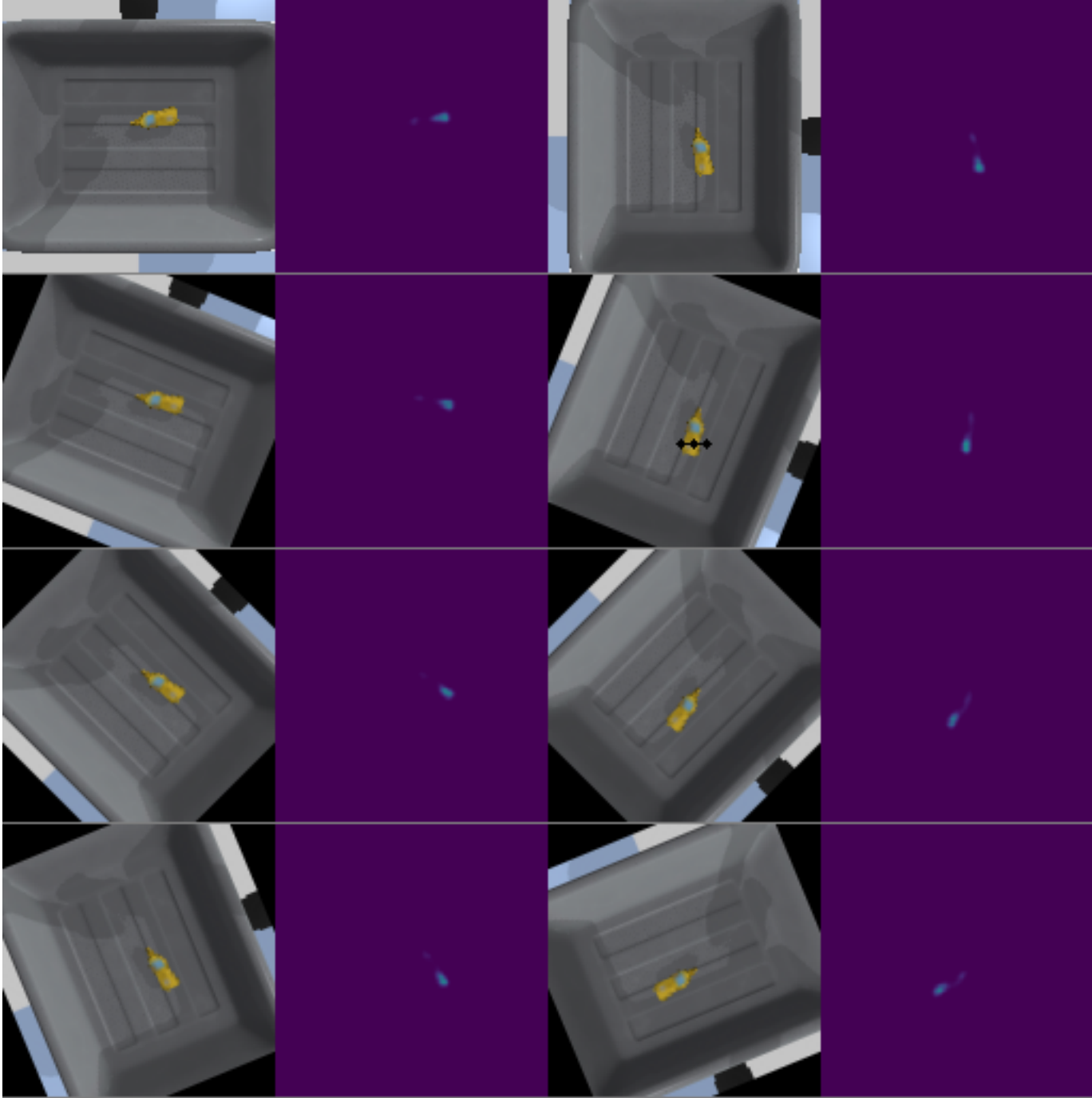


Figure 2: affordance eval training YcvMustardBottle

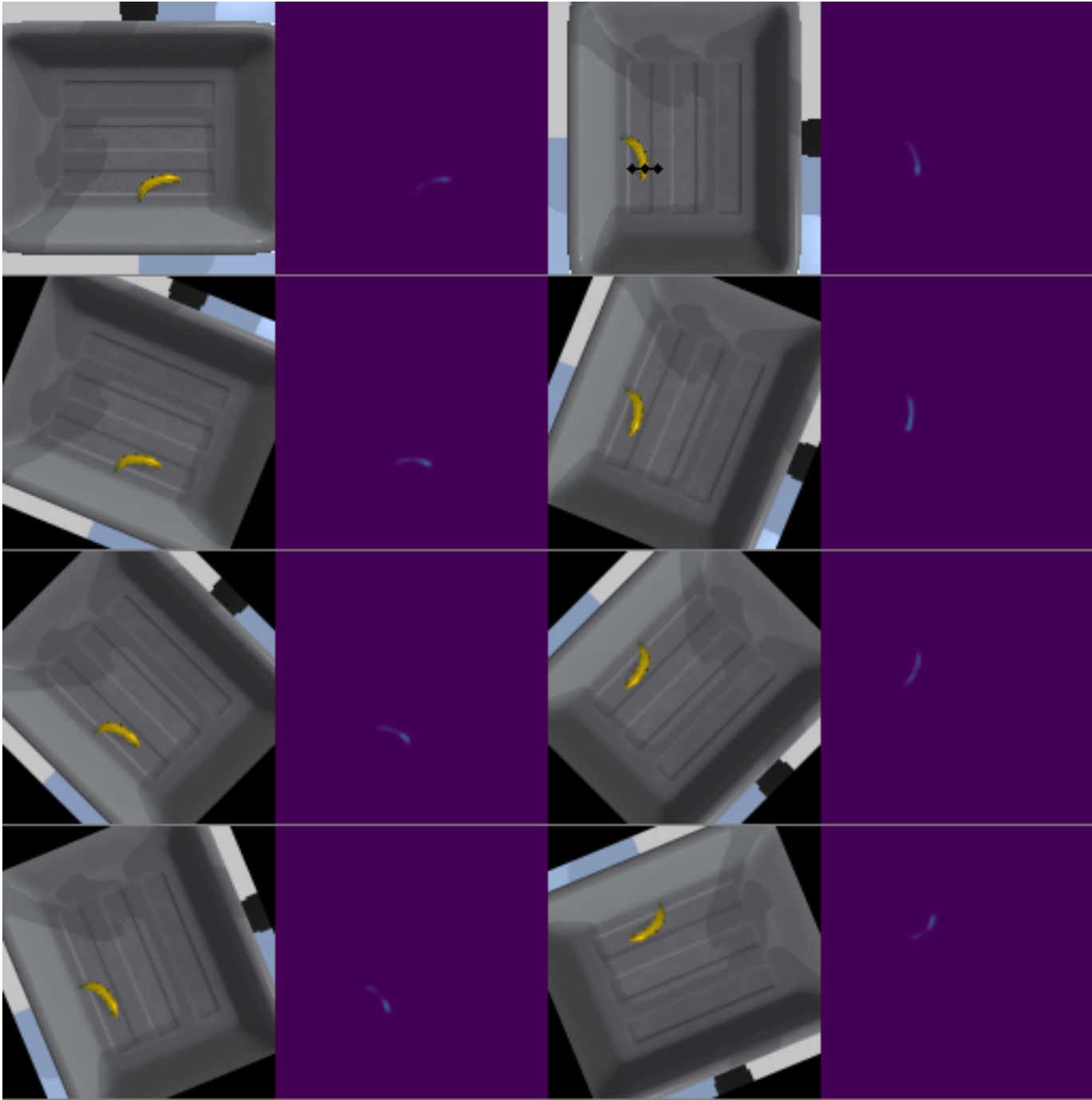


Figure 3: affordance eval testing Banana

Problem 3

Problem 3-a

Recorded Video: <https://youtu.be/dbCLBx1Rs2Q>

Success Rate: 73%

Problem 3-b

Recorded Video: <https://youtu.be/vOfQR5Pm8L4>

Success Rate: 84%

Problem 3-c

Recorded Video: <https://youtu.be/u5G6q3sRTj8>

Success Rate: 14 picked up, 1 left

Problem 3-d

Time-wise, evaluation for problem 3 takes longer because it allows for multiple trials. For 3-a, it's the same. For 3-b, we can see an improvement since sometimes the next best actions might be more suitable for grasping. However, there are still undetectable objects which all 8 attempts failed. For 3-c, we also see an improvement trying out different angle and location to pick up the items.

Problem 4

Problem 4-a

Best model: saved at epoch 60, with test loss of 0.0551 and training loss of 0.034

See Figure 4

Problem 4-b

Recorded Video: <https://youtu.be/7I13UN5wXIM>

Success Rate: 20%, See Figure 5 and 6



Figure 4: action regression 003



Figure 5: MustardBottle 2

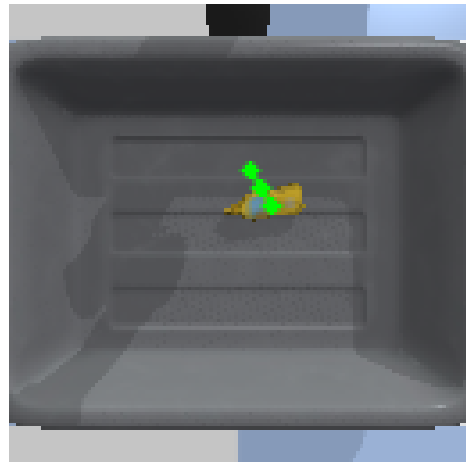


Figure 6: MustardBottle 1

Problem 4-c

Recorded Video: <https://youtu.be/xIatHEUUnLk>

Success Rate: 2 picked up, 13 left

The reason why Action Regression is worse than Visual Affordance is because the affordance map has much more learning signal than a 3-dimension action vector. Therefore, it is easier to learn in the low-data environment we have.