

### Part i

Test accuracy is 0.2525

Compare the accuracies we got in HW7 (in which the lowest accuracy is 0.3925), our test accuracy is relatively low. I think it's due to the small learning rate (0.0001) considering we only have 5 epochs. Smaller learning rate means that it will take longer for the model to adapt to the problem. During each update, the changes to the weights are small, so it needs more epochs, more than 5 in our case.

### Part ii

1) Test accuracy: 0.9850

The accuracy has improved a lot and is high. This may be due to that we transfer the layers that were pre-trained on large amounts of data.

2) Test accuracy: 0.9350

Compare with previous part i model (the number of layers from Alexnet: 0) and the part ii 1) model (the number of layers from Alexnet: some), and this model (the number of layers from Alexnet: many), we can see that transferring pre-trained layers from Alexnet can improve the performance, but it doesn't mean more transferring the better.

### Part iii

Case 1. when the number of training images is 3 per class

1) 'RandRotation' = [0 360], 'RandScale' = [0.5 1]

Test accuracy: 0.9940

2) 'RandXTranslation' = [-5 5], 'RandScale' = [0.5 1]

Test accuracy: 0.9711

3) 'RandXTranslation' = [-5 5], 'RandXReflection' = 1

Test accuracy: 0.9541

The combination 1) has the highest test accuracy. Compare with Part ii 1) where we have no data augmentation (which has test accuracy 0.9850), Part iii 1) improves the test accuracy while the other combinations have lower results.

Case 2. when the number of training images is 10 per class

1) 'RandRotation' = [0 360], 'RandScale' = [0.5 1]

Test accuracy: 0.9527

2) 'RandXTranslation' = [-5 5], 'RandScale' = [0.5 1]

Test accuracy: 0.9893

3) 'RandXTranslation' = [-5 5], 'RandXReflection' = 1

Test accuracy: 0.9786

First, compare with the non-augmentation model in part ii (0.9850), only combination 2) in this case is higher while combinations 1) and 3) both have lower test accuracies. Then, compare with case 1 where the number of training images is 3 per class, the combination that has the highest accuracy in case 1 has the lowest test accuracy in case 2. And other two combinations also have different behaviors in case 2. So we can see that the number of training images per class has a great impact on the results. Even for the same combination, different number of training images can lead to totally different results.

Part iv

Test accuracy: 0.9884

Compare with the accuracy when we don't freeze the transferred layers (0.9850), freezing transferred layers give a better result but very similar. I think the small improvement may be that when we freeze the layers, their weights will not be updated during backpropagation. Since we have other layers after the transferred layers and their weights are being trained, freezing the transferred layers prevent overpowering the weights they already have.