Deep Learning Homework 4

Ashwin Disa, Devesh Bhangale

Q1. a)

Epoch [1/25], Loss: 0.5456
Epoch [2/25], Loss: 0.4210
Epoch [3/25], Loss: 0.3817
Epoch [4/25], Loss: 0.3604
Epoch [5/25], Loss: 0.3423
Epoch [6/25], Loss: 0.3302
Epoch [7/25], Loss: 0.3168
Epoch [8/25], Loss: 0.3100
Epoch [9/25], Loss: 0.3009
Epoch [10/25], Loss: 0.2924
Epoch [11/25], Loss: 0.2817
Epoch [12/25], Loss: 0.2773
Epoch [13/25], Loss: 0.2734
Epoch [14/25], Loss: 0.2649
Epoch [15/25], Loss: 0.2631
Epoch [16/25], Loss: 0.2583
Epoch [17/25], Loss: 0.2493
Epoch [18/25], Loss: 0.2462
Epoch [19/25], Loss: 0.2452
Epoch [20/25], Loss: 0.2339
Epoch [21/25], Loss: 0.2326
Epoch [22/25], Loss: 0.2329
Epoch [23/25], Loss: 0.2276
Epoch [24/25], Loss: 0.2263
Epoch [25/25], Loss: 0.2190

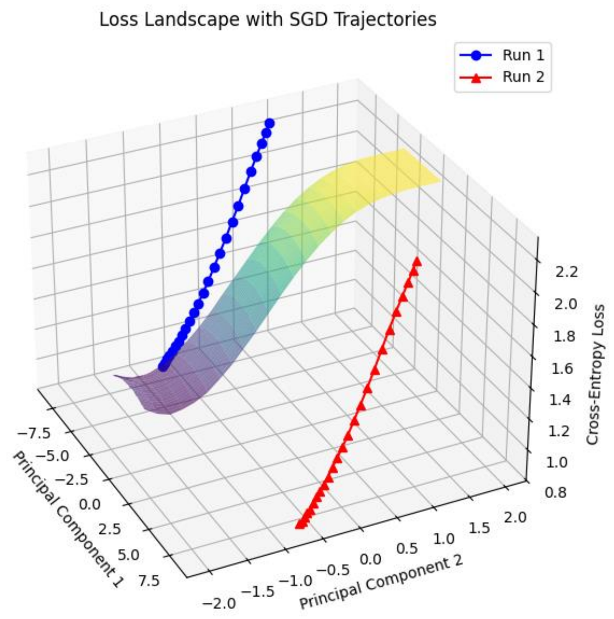
Accuracy on test data: 88.27%

c)

The discrepancy arises because PCA reduces the high-dimensional parameter space to just two dimensions, losing some information. The surface plot represents an approximation of the loss landscape based on these two principal components, while the SGD trajectories operate in full parameter space. Therefore, the projected points from PCA do not perfectly reconstruct the original parameters, leading to differences in the computed loss values. Creating a surface plot that exactly matches the "real" fCE values would be impractical due to the high dimensionality and complexity of the full parameter space.

d) i and iii are True.

Both PCA and inverse PCA are linear transformations.



Q2.

[Epoch 1, Mini-batch 200] loss: 1.145
[Epoch 1, Mini-batch 400] loss: 0.693
[Epoch 1, Mini-batch 600] loss: 0.602
[Epoch 1, Mini-batch 800] loss: 0.565
[Epoch 2, Mini-batch 200] loss: 0.483
[Epoch 2, Mini-batch 400] loss: 0.479
[Epoch 2, Mini-batch 600] loss: 0.447
[Epoch 2, Mini-batch 800] loss: 0.433
[Epoch 3, Mini-batch 200] loss: 0.404
[Epoch 3, Mini-batch 400] loss: 0.408
[Epoch 3, Mini-batch 600] loss: 0.382
[Epoch 3, Mini-batch 800] loss: 0.380
[Epoch 4, Mini-batch 200] loss: 0.351
[Epoch 4, Mini-batch 400] loss: 0.359
[Epoch 4, Mini-batch 600] loss: 0.346
[Epoch 4, Mini-batch 800] loss: 0.360
[Epoch 5, Mini-batch 200] loss: 0.331
[Epoch 5, Mini-batch 400] loss: 0.321
[Epoch 5, Mini-batch 600] loss: 0.331
[Epoch 5, Mini-batch 800] loss: 0.326
[Epoch 6, Mini-batch 200] loss: 0.315
[Epoch 6, Mini-batch 400] loss: 0.325
[Epoch 6, Mini-batch 600] loss: 0.310
[Epoch 6, Mini-batch 800] loss: 0.306
[Epoch 7, Mini-batch 200] loss: 0.294

[Epoch 7, Mini-batch 400] loss: 0.303
[Epoch 7, Mini-batch 600] loss: 0.295
[Epoch 7, Mini-batch 800] loss: 0.298
[Epoch 8, Mini-batch 200] loss: 0.299
[Epoch 8, Mini-batch 400] loss: 0.278
[Epoch 8, Mini-batch 600] loss: 0.282
[Epoch 8, Mini-batch 800] loss: 0.283
[Epoch 9, Mini-batch 200] loss: 0.277
[Epoch 9, Mini-batch 400] loss: 0.283
[Epoch 9, Mini-batch 600] loss: 0.273
[Epoch 9, Mini-batch 800] loss: 0.277
[Epoch 10, Mini-batch 200] loss: 0.265
[Epoch 10, Mini-batch 400] loss: 0.276
[Epoch 10, Mini-batch 600] loss: 0.281
[Epoch 10, Mini-batch 800] loss: 0.275
[Epoch 11, Mini-batch 200] loss: 0.251
[Epoch 11, Mini-batch 400] loss: 0.231
[Epoch 11, Mini-batch 600] loss: 0.229
[Epoch 11, Mini-batch 800] loss: 0.236
[Epoch 12, Mini-batch 200] loss: 0.217
[Epoch 12, Mini-batch 400] loss: 0.221
[Epoch 12, Mini-batch 600] loss: 0.226
[Epoch 12, Mini-batch 800] loss: 0.224
[Epoch 13, Mini-batch 200] loss: 0.222
[Epoch 13, Mini-batch 400] loss: 0.222
[Epoch 13, Mini-batch 600] loss: 0.219
[Epoch 13, Mini-batch 800] loss: 0.223
[Epoch 14, Mini-batch 200] loss: 0.219
[Epoch 14, Mini-batch 400] loss: 0.213
[Epoch 14, Mini-batch 600] loss: 0.217
[Epoch 14, Mini-batch 800] loss: 0.213
[Epoch 15, Mini-batch 200] loss: 0.210
[Epoch 15, Mini-batch 400] loss: 0.211
[Epoch 15, Mini-batch 600] loss: 0.216
[Epoch 15, Mini-batch 800] loss: 0.208
[Epoch 16, Mini-batch 200] loss: 0.213
[Epoch 16, Mini-batch 400] loss: 0.216
[Epoch 16, Mini-batch 600] loss: 0.211
[Epoch 16, Mini-batch 800] loss: 0.206
[Epoch 17, Mini-batch 200] loss: 0.208
[Epoch 17, Mini-batch 400] loss: 0.219
[Epoch 17, Mini-batch 600] loss: 0.207
[Epoch 17, Mini-batch 800] loss: 0.207
[Epoch 18, Mini-batch 200] loss: 0.206
[Epoch 18, Mini-batch 400] loss: 0.207
[Epoch 18, Mini-batch 600] loss: 0.213
[Epoch 18, Mini-batch 800] loss: 0.197
[Epoch 19, Mini-batch 200] loss: 0.213

[Epoch 19, Mini-batch 400] loss: 0.205
[Epoch 19, Mini-batch 600] loss: 0.211
[Epoch 19, Mini-batch 800] loss: 0.200
[Epoch 20, Mini-batch 200] loss: 0.203
[Epoch 20, Mini-batch 400] loss: 0.201
[Epoch 20, Mini-batch 600] loss: 0.203
[Epoch 20, Mini-batch 800] loss: 0.202
[Epoch 21, Mini-batch 200] loss: 0.202
[Epoch 21, Mini-batch 400] loss: 0.198
[Epoch 21, Mini-batch 600] loss: 0.205
[Epoch 21, Mini-batch 800] loss: 0.195
[Epoch 22, Mini-batch 200] loss: 0.195
[Epoch 22, Mini-batch 400] loss: 0.196
[Epoch 22, Mini-batch 600] loss: 0.201
[Epoch 22, Mini-batch 800] loss: 0.200
[Epoch 23, Mini-batch 200] loss: 0.191
[Epoch 23, Mini-batch 400] loss: 0.195
[Epoch 23, Mini-batch 600] loss: 0.202
[Epoch 23, Mini-batch 800] loss: 0.198
[Epoch 24, Mini-batch 200] loss: 0.188
[Epoch 24, Mini-batch 400] loss: 0.205
[Epoch 24, Mini-batch 600] loss: 0.196
[Epoch 24, Mini-batch 800] loss: 0.201
[Epoch 25, Mini-batch 200] loss: 0.199
[Epoch 25, Mini-batch 400] loss: 0.201
[Epoch 25, Mini-batch 600] loss: 0.193
[Epoch 25, Mini-batch 800] loss: 0.191
[Epoch 26, Mini-batch 200] loss: 0.197
[Epoch 26, Mini-batch 400] loss: 0.198
[Epoch 26, Mini-batch 600] loss: 0.201
[Epoch 26, Mini-batch 800] loss: 0.202
[Epoch 27, Mini-batch 200] loss: 0.198
[Epoch 27, Mini-batch 400] loss: 0.197
[Epoch 27, Mini-batch 600] loss: 0.193
[Epoch 27, Mini-batch 800] loss: 0.195
[Epoch 28, Mini-batch 200] loss: 0.188
[Epoch 28, Mini-batch 400] loss: 0.188
[Epoch 28, Mini-batch 600] loss: 0.198
[Epoch 28, Mini-batch 800] loss: 0.199
[Epoch 29, Mini-batch 200] loss: 0.199
[Epoch 29, Mini-batch 400] loss: 0.187
[Epoch 29, Mini-batch 600] loss: 0.185
[Epoch 29, Mini-batch 800] loss: 0.202
[Epoch 30, Mini-batch 200] loss: 0.201
[Epoch 30, Mini-batch 400] loss: 0.185
[Epoch 30, Mini-batch 600] loss: 0.196
[Epoch 30, Mini-batch 800] loss: 0.198
Finished Training

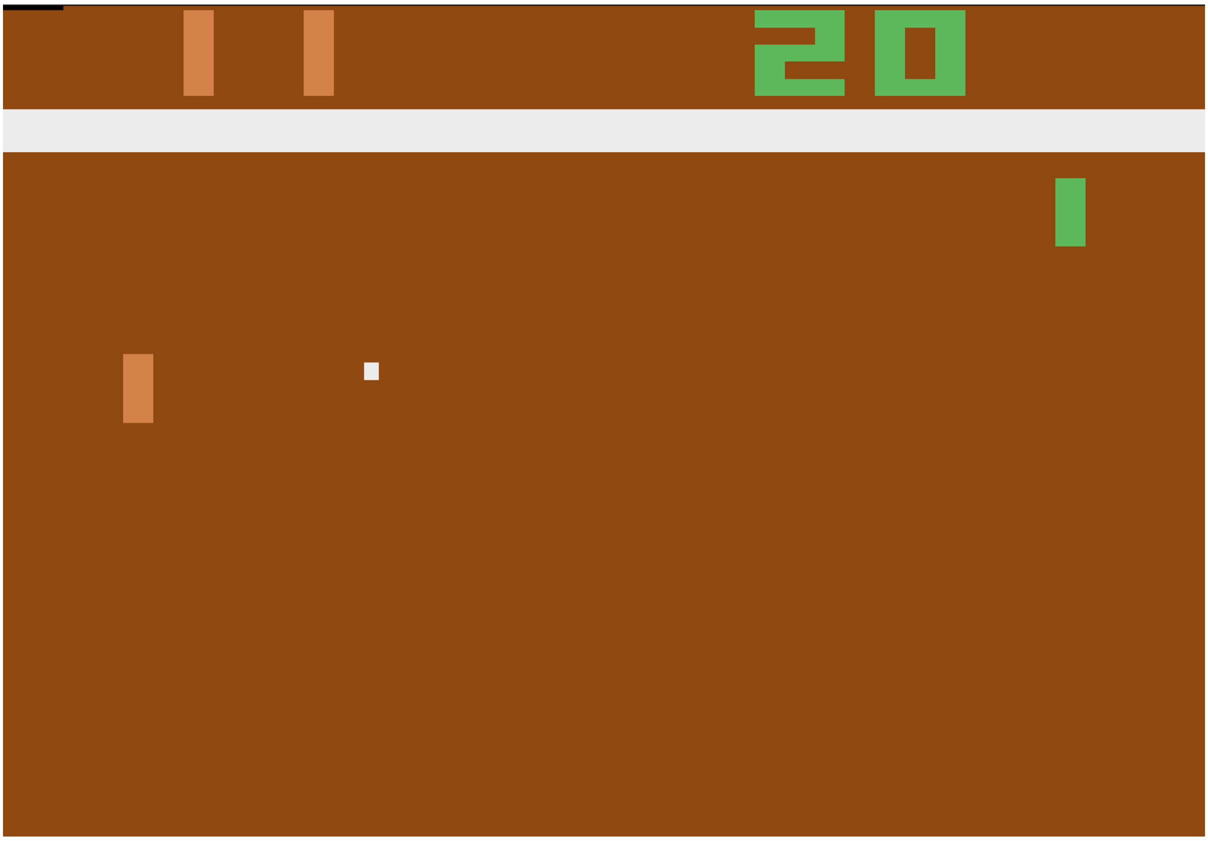
Accuracy of the network on the 10000 test images: 92.22%

Q3.

We chose to fine-tune the whole model including the weights of Resnet

train Loss: 0.6693 Acc: 0.7656
test Loss: 0.4568 Acc: 0.8277
train Loss: 0.4107 Acc: 0.8543
test Loss: 0.3557 Acc: 0.8719
train Loss: 0.3452 Acc: 0.8745
test Loss: 0.3390 Acc: 0.8749
train Loss: 0.3145 Acc: 0.8868
test Loss: 0.3329 Acc: 0.8795
train Loss: 0.2831 Acc: 0.8991
test Loss: 0.2864 Acc: 0.8981
train Loss: 0.2600 Acc: 0.9052
test Loss: 0.2921 Acc: 0.8911
train Loss: 0.2426 Acc: 0.9116
test Loss: 0.2821 Acc: 0.8970
train Loss: 0.1896 Acc: 0.9298
test Loss: 0.2364 Acc: 0.9134
train Loss: 0.1756 Acc: 0.9354
test Loss: 0.2347 Acc: 0.9134
train Loss: 0.1678 Acc: 0.9384
test Loss: 0.2353 Acc: 0.9138
train Loss: 0.1605 Acc: 0.9410
test Loss: 0.2311 Acc: 0.9166
train Loss: 0.1551 Acc: 0.9427
test Loss: 0.2341 Acc: 0.9141
train Loss: 0.1515 Acc: 0.9446
test Loss: 0.2312 Acc: 0.9157
train Loss: 0.1454 Acc: 0.9464
test Loss: 0.2275 Acc: 0.9209
train Loss: 0.1355 Acc: 0.9501
test Loss: 0.2268 Acc: 0.9192
Training complete in 22m 9s
Best val Acc: 0.920900

Q4.
The class CNNPolicyNet() with its methods is in the train.py file attached along with the play_atari.py
Also attached the model.cpt file in submission to replicate the following results.


11
20