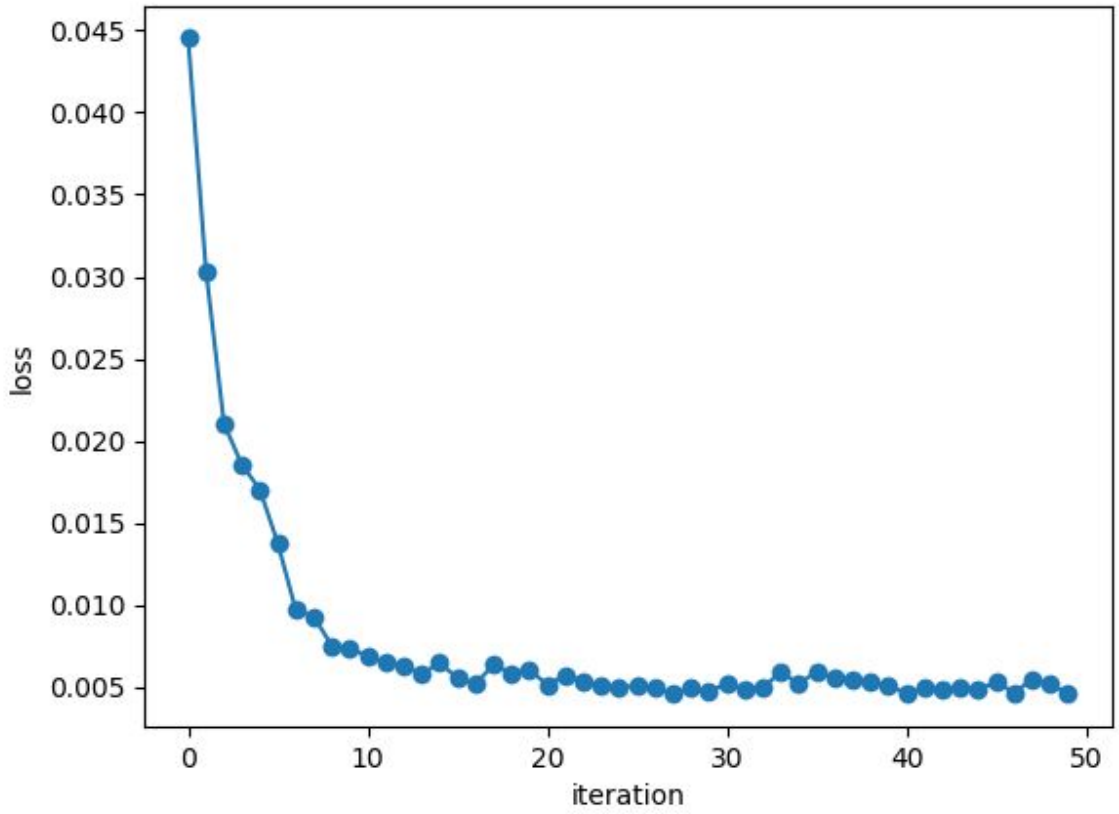
Section 2

Setting: Convolutional neural network trained on the Humanoid environment



Section 3.1:

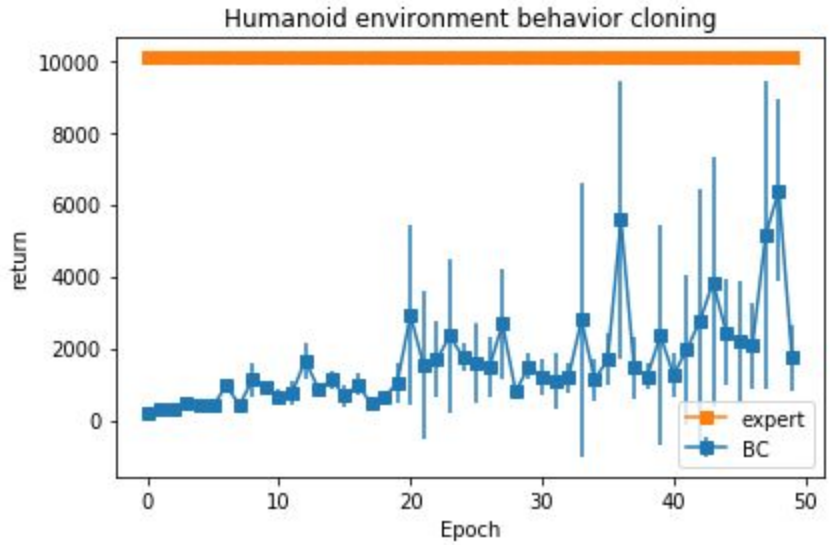
Architecture: (Basically using the same model as Watcher & Zijie did. I tried convolutional neural network once for behavior cloning in Humanoid environment. The result was not good at all. The loss never went down below 0.001, and the mean returns from the trained convolutional neural network never reached above 1,000.)

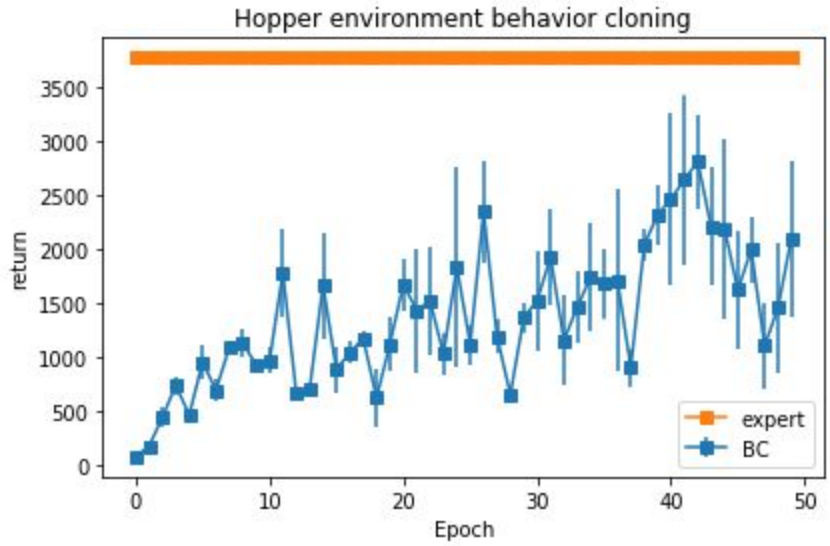
```
class CNN(nn.Module):
    def __init__(self, INPUTDIM, OUTPUTDIM):
        super(CNN, self).__init__()
        self.fc1 = nn.Linear(INPUTDIM, 512)
        self.bn1 = nn.BatchNorm1d(512)
        self.fc2 = nn.Linear(512, 512)
        self.bn2 = nn.BatchNorm1d(512)
        self.fc3 = nn.Linear(512, 512)
        self.bn3 = nn.BatchNorm1d(512)
        self.fc4 = nn.Linear(512, 512)
        self.bn4 = nn.BatchNorm1d(512)
        self.fc5=nn.Linear(512,OUTPUTDIM)

    def forward(self,x):
        x=self.bn1(self.fc1(x))
        x = F.relu(x)
        x=self.bn2(self.fc2(x))
        x = F.relu(x)
        x=self.bn3(self.fc3(x))
        x = F.relu(x)
        x=self.bn4(self.fc4(x))
        x = F.relu(x)
        x=self.fc5(x)
        return x
```

Expert policy rollout numbers=20
Epoch=50

Results of behavior cloning:



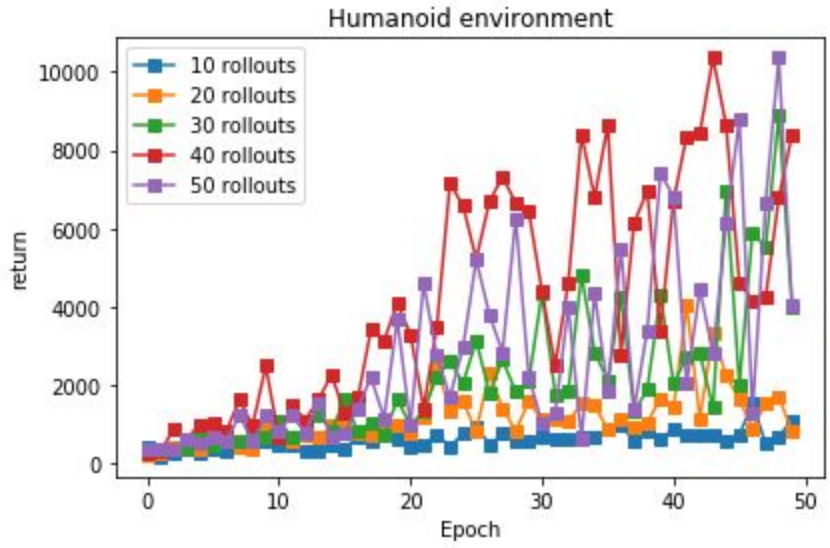


While behavior cloning achieves comparable performance in the Hopper environment, it fails to do so in the Humanoid environment. There are two factors that might contribute to the difference here. First, in the Humanoid environment the dimension of the observation and action spaces are 376 & 17 respectively, while in the Hopper environment the dimensions are only 11 & 3. In other words, the Humanoid environment is much more complex than the Hopper environment.

Second, in the Humanoid environment, the expert policy might not be so “expert” after all. While generating rollout data, I ran into situations where the expert terminated early and gained little reward(see appendix).

Section 3.2:
Hyperparameter: number of demonstrations(expert policy rollout numbers)
Environment: Humanoid
Architecture: same as above

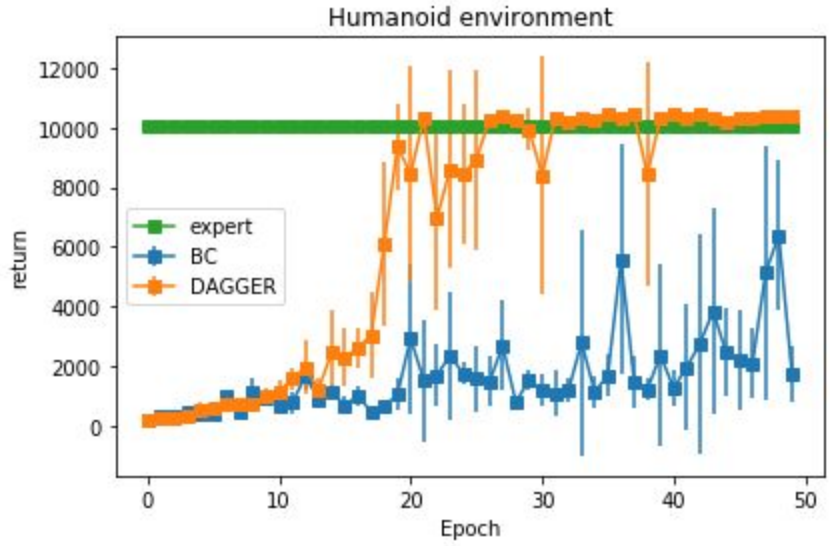
Results:

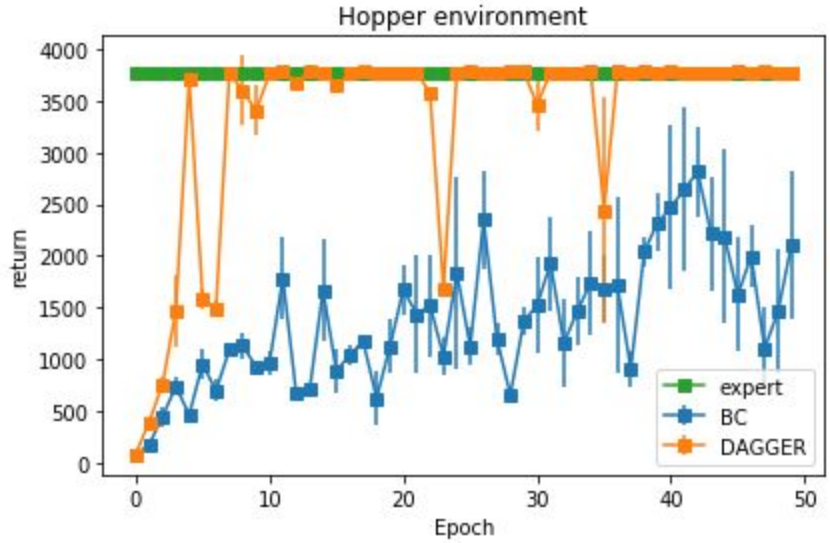


Rationale:

As the number of demonstration increases, we have a larger training set better approximating the real world distribution. Hence, we can expect network trained on this dataset yields a better performance. This is indeed the case when the rollout number increases from 10 to 40. However, when the rollout number increases from 40 to 50, there is no more improvement gained. One possible explanation is that the marginal benefit of approximation as we increase the rollout number is decreasing. The (observation, action) sequence generated by the expert policy typically will not include those “dangerous” observations, and we need DAGGER to further improve the performance of our network.

Section 4:


Humanoid environment
return
Epoch
expert
BC
DAGGER
0
2000
4000
6000
8000
10000
12000
0
10
20
30
40
50


Hopper environment
return
Epoch
expert
BC
DAGGER
0
500
1000
1500
2000
2500
3000
3500
4000
0
10
20
30
40
50

Appendix

#Humanoid environment

humanoid_expert_mean_of_all_rollouts=[
10448.195872253944, 10481.7435684638, 10397.913210861252, 10384.384788014513,
10402.170481802876, 10141.75646949923, 10390.418363061635, 10433.168370310226,
10351.581482619016, 10427.743799942833, 10422.676598963639, 10436.77727547405,
10409.031334560947, 10461.674997138833, 10356.660778171954,
4035.5473932482714,
10344.81489536892, 10390.202293611559, 10399.123974525617, 10364.837145147623]

humanoid_expert_mean=10074.021154652035

humanoid_expert_std=1386.9620644584668

humanoid_bc_history_of_mean=[198.74934639, 302.52814483, 306.34311862,
465.90731334, 419.40772583,
404.62749195, 963.70166043, 428.20162684, 1134.15195957, 915.19929965,
676.51570316, 763.26356822, 1668.30396635, 878.39590502, 1124.98711335,
680.69403097, 1004.51927894, 456.36368959, 630.07180768, 1047.36149923,
2928.17515015, 1525.03190116, 1694.81873183, 2349.38382791, 1746.85380643,
1571.87819229, 1482.19395389, 2678.0006785, 797.11124427, 1509.87301989,
1208.41587278, 1092.82487057, 1177.23091743, 2791.30560508, 1123.77438302,
1697.38115054, 5572.83038513, 1462.57338181, 1226.73860153, 2358.23221008,
1261.33537639, 1965.22515665, 2754.02118757, 3833.5919805, 2451.74439261,
2184.98437304, 2081.91471044, 5146.90531397, 6381.81306457, 1739.97525937]

humanoid_bc_history_of_std=[45.83623212, 29.76137656, 16.88434883, 184.82437839,
63.51145525,
53.71137286, 176.36303151, 43.03416106, 481.68503985, 77.47506309,
220.65912737, 355.49562141, 497.15491819, 61.16478727, 260.14746176,
293.16377805, 322.5398694, 81.32718838, 166.5406618, 547.98575404,
2524.1018258, 2042.77111937, 1056.13006608, 2128.55662276, 379.71620631,
1105.52014345, 855.26269913, 1514.17802109, 179.29869462, 333.96195849,
521.08809117, 793.42250212, 407.11808226, 3799.30398839, 561.19846322,
712.46403596, 3845.94856797, 845.99404779, 348.3510279, 3063.85852221,
584.44642703, 2093.08162632, 3684.18104143, 3471.88368315, 1463.64084918,
1683.77806484, 1186.86462092, 4258.09757061, 2530.64081616, 916.37140946]

humanoid_dagger_history_of_mean=[166.66742887, 248.8699933, 282.40357268,
323.11090458,
548.65626431, 578.53973195, 741.06613821, 701.75856566,
709.08414352, 1004.51701204, 1141.06157366, 1575.4999752,
1946.01444465, 1219.59475144, 2467.15432537, 2295.6560252,
2618.68159081, 3031.37679247, 6101.26555856, 9366.30542689,
8442.47202737, 10330.32383737, 6967.19617835, 8591.98927776,

8450.01997426, 8925.9144808, 10271.89622127, 10379.75183767,
10280.37632883, 9950.57091231, 8398.77401877, 10319.70315391,
10171.51640376, 10342.75940189, 10266.77950211, 10480.20163195,
10306.77670544, 10432.80670426, 8412.46348243, 10333.87848357,
10453.96518447, 10301.51538306, 10436.7767714, 10340.54379608,
10163.69643079, 10345.24516707, 10316.69123817, 10405.47789784,
10386.03090612, 10413.34912549]

humanoid_dagger_history_of_std=[3.28584602e+00, 4.60867751e+00, 7.47420106e+01,
7.11120796e+01,
2.26809470e+02, 1.73373296e+02, 1.66364235e+02, 2.15276179e+02,
1.44071487e+02, 2.77182225e+02, 4.11846935e+02, 3.56141056e+02,
9.14061958e+02, 4.03010324e+02, 1.37986320e+03, 9.75123732e+02,
6.60626787e+02, 1.42002914e+03, 2.75887649e+03, 1.43019824e+03,
3.63456190e+03, 4.98599163e+01, 3.10919483e+03, 3.31991289e+03,
2.33881917e+03, 3.02379347e+03, 1.04132860e+02, 5.33078129e+01,
4.78411253e+01, 6.75425877e+02, 3.97843166e+03, 8.75206099e+01,
9.84895032e+01, 2.44686688e+01, 4.22190855e+01, 7.69836642e+01,
6.49710946e+01, 5.84588909e+01, 3.76002721e+03, 1.35765161e+02,
4.00903323e+01, 9.50387573e+01, 5.34759982e+01, 4.24985024e+01,
2.79431869e+01, 2.78899376e+01, 5.52430289e+01, 5.63756027e+01,
3.80576850e+01, 3.16145575e+01]

#Hopper environment
hopper_expert_mean_of_all_rollouts=[
3776.258622118129, 3781.1758918982596, 3774.9131759824318, 3780.697286920128,
3778.015266814751,
3780.398301396628, 3780.159025960202, 3779.23286603148, 3780.3554876264006,
3772.179704576205,
3780.863795447612, 3779.6210141672736, 3780.0712395515247, 3784.1917665582832,
3782.912734535562,
3780.554797600858, 3784.4771139131567, 3779.3136667971144, 3780.8298738615235,
3770.778163925877]
hopper_expert_mean=3779.3499897841707
hopper_expert_std=3.411681908561538

hopper_bc_history_of_mean=[77.5749774, 171.99408051, 441.19279213, 737.6911625,
462.34557888,
953.25167442, 693.72475143, 1090.64674805, 1129.40765254, 927.78279493,
966.33340295, 1782.07457105, 665.20262153, 702.93763794, 1666.22692634,
883.80145523, 1042.36891828, 1172.92529044, 623.91193873, 1121.19286636,
1674.54207613, 1436.78709428, 1519.81192529, 1028.87528032, 1838.66032777,

1118.556477, 2352.84925591, 1193.16681312, 652.51431751, 1367.76096979,
1523.30888001, 1929.00812289, 1158.11473003, 1463.02124274, 1742.99441471,
1683.37803877, 1711.58611772, 901.89924849, 2039.35621377, 2321.86323384,
2468.73579463, 2646.87355012, 2814.85149666, 2212.21810204, 2192.41836001,
1624.40431995, 1997.40609769, 1103.80090045, 1460.90048944, 2103.71450705]

hopper_bc_history_of_std=[1.95949087, 1.46339657, 93.3402985, 80.66277763,
8.44252752,
151.0097446, 105.57994663, 12.09145082, 132.49685146, 39.32360171,
120.46552067, 400.97481541, 6.48864476, 11.435424, 492.30294586,
209.56854416, 99.53509409, 67.99224007, 265.0034293, 258.48915153,
243.40591314, 575.97843791, 493.56059155, 189.74591272, 926.13606862,
184.62455251, 473.23479632, 153.40180777, 9.60263639, 137.66403281,
466.44720256, 453.00816886, 421.12656138, 327.05721336, 500.13933208,
326.12263409, 849.58960994, 172.19807034, 149.91033643, 277.46391404,
797.8943807, 790.08512065, 433.64586753, 543.88804961, 836.84702634,
550.22540776, 302.79162519, 399.35490665, 602.3109092, 722.70670832]

hopper_dagger_history_of_mean=[78.7909229, 390.06690026, 753.24269139,
1467.48149215, 3713.9392316,
1589.69703598, 1489.83960845, 3767.75776687, 3602.25520156, 3404.59445452,
3762.28949963, 3785.72467596, 3677.67906309, 3788.60370856, 3775.78404034,
3644.72184697, 3763.15441125, 3782.85805361, 3767.71071975, 3775.18976201,
3764.547246, 3775.19221769, 3575.40613712, 1683.5058464, 3770.87186674,
3790.1633175, 3774.39503398, 3775.26676248, 3784.64168673, 3780.72434625,
3466.71789388, 3776.12333878, 3776.59759881, 3771.27365797, 3784.77699232,
2443.6345721, 3782.59090385, 3777.24014412, 3780.79144262, 3773.89753015,
3782.28764808, 3778.15698236, 3778.82573831, 3775.4038308, 3778.16083582,
3781.37125161, 3779.84258366, 3781.59192968, 3777.58634821, 3779.06070733]

hopper_dagger_history_of_std=[2.04820997, 4.75656887, 79.6880605, 350.497520,
12.7683424, 99.0026359, 80.9408023, 6.56486116,
335.356434, 242.707975, 4.55423401, 2.98274356,
51.0307524, 4.95203532, 4.18194156, 48.9840333,
1.66268257, 2.18029345, 2.60286553, 3.37625638,
3.80912956, 0.538623888, 15.2579304, 68.1600339,
2.60382783, 6.00651899, 3.20672387, 2.67311983,
4.53040408, 2.12900530, 264.223719, 1.87383876,
3.48836126, 4.29545198, 3.65105411, 1085.13649,
2.65460146, 3.04287439, 1.77162960, 1.76576520,
2.19586539, 5.81418685, 5.93445220, 4.14281793,
1.34092909, 2.52873230, 2.45648546, 3.77128771,
1.49637285, 1.76093482]