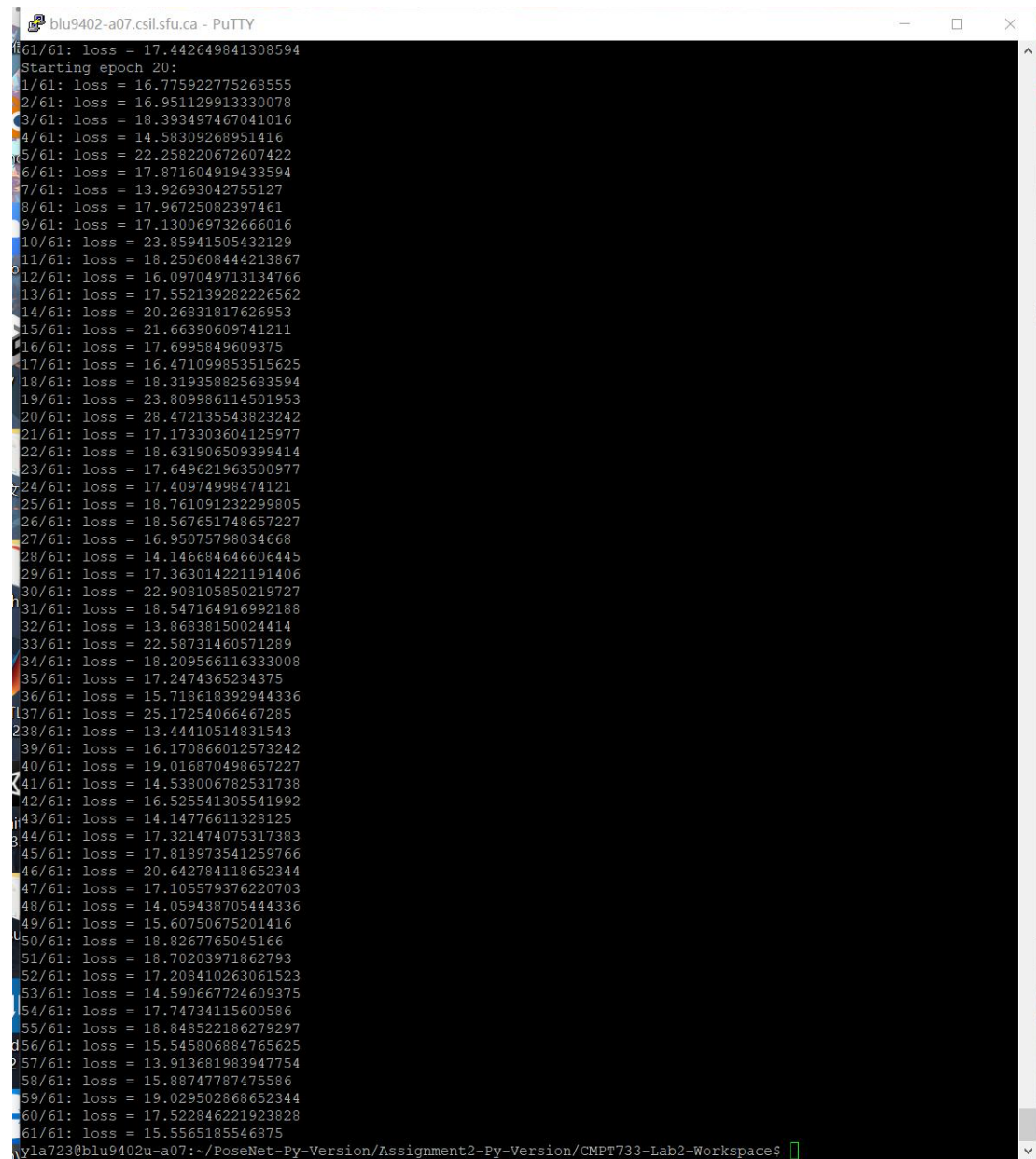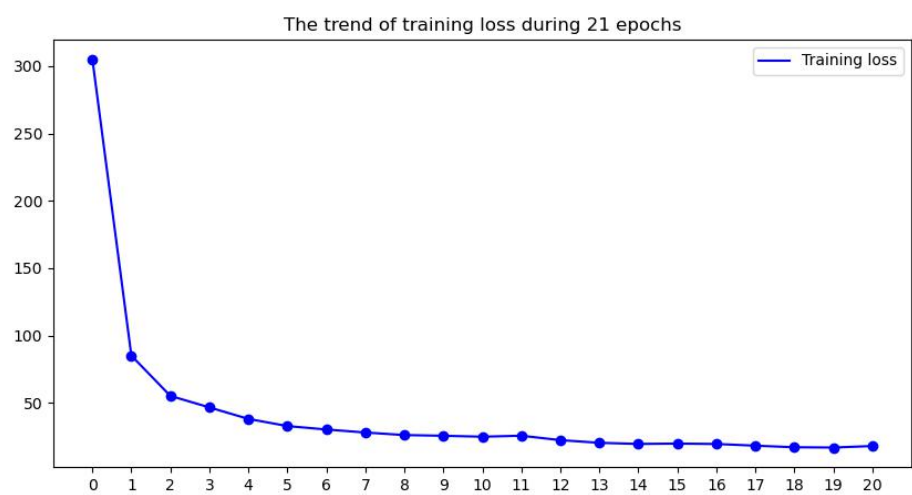I use learning rate=0.0001, batch size=20 and Adam optimizer to train the model 21 epochs, the following graph is the training loss of the last epoch and the graph of the trend of training loss in the 21 epochs. In the graph, we can see trend of the training loss is keeping decrease during the training process, and getting stable after epoch 13.

To calculate the loss function, I use the mse instead of the norms in the formula, because this can reduce both the training error and test error.


```
blu9402-a07.csil.sfu.ca - PuTTY                                          —   □   ×
61/61: loss = 17.442649841308594
Starting epoch 20:
1/61: loss = 16.775922775268555
2/61: loss = 16.951129913330078
3/61: loss = 18.393497467041016
4/61: loss = 14.58309268951416
5/61: loss = 22.258220672607422
6/61: loss = 17.871604919433594
7/61: loss = 13.92693042755127
8/61: loss = 17.96725082397461
9/61: loss = 17.130069732666016
10/61: loss = 23.85941505432129
11/61: loss = 18.250608444213867
12/61: loss = 16.097049713134766
13/61: loss = 17.552139282226562
14/61: loss = 20.26831817626953
15/61: loss = 21.66390609741211
16/61: loss = 17.6995849609375
17/61: loss = 16.471099853515625
18/61: loss = 18.319358825683594
19/61: loss = 23.809986114501953
20/61: loss = 28.472135543823242
21/61: loss = 17.173303604125977
22/61: loss = 18.631906509399414
23/61: loss = 17.649621963500977
24/61: loss = 17.40974998474121
25/61: loss = 18.761091232299805
26/61: loss = 18.567651748657227
27/61: loss = 16.95075798034668
28/61: loss = 14.146684646606445
29/61: loss = 17.363014221191406
30/61: loss = 22.908105850219727
31/61: loss = 18.547164916992188
32/61: loss = 13.86838150024414
33/61: loss = 22.58731460571289
34/61: loss = 18.209566116333008
35/61: loss = 17.2474365234375
36/61: loss = 15.718618392944336
37/61: loss = 25.17254066467285
38/61: loss = 13.44410514831543
39/61: loss = 16.170866012573242
40/61: loss = 19.016870498657227
41/61: loss = 14.538006782531738
42/61: loss = 16.525541305541992
43/61: loss = 14.14776611328125
44/61: loss = 17.321474075317383
45/61: loss = 17.818973541259766
46/61: loss = 20.642784118652344
47/61: loss = 17.105579376220703
48/61: loss = 14.059438705444336
49/61: loss = 15.60750675201416
50/61: loss = 18.8267765045166
51/61: loss = 18.70203971862793
52/61: loss = 17.208410263061523
53/61: loss = 14.590667724609375
54/61: loss = 17.74734115600586
55/61: loss = 18.848522186279297
56/61: loss = 15.545806884765625
57/61: loss = 13.913681983947754
58/61: loss = 15.88747787475586
59/61: loss = 19.029502868652344
60/61: loss = 17.522846221923828
61/61: loss = 15.5565185546875
yla723@blu9402u-a07:~/PoseNet-Py-Version/Assignment2-Py-Version/CMPT733-Lab2-Workspace$ █
```

The trend of training loss during 21 epochs

I test the model with epoch_21.pth, the following graph shows the median position error=7.95 and median orientation error=4.67.

Before calculate the test error, I divided the GT coordinates by 100 to make them in the same order of magnitude as the PRED coordinates(because we normalized the PRED). After calculate the test loss, we multiple the loss value by 100 to change it back to the original order of magnitude. This method helps the learning process and makes the test error smaller.