

## ML2021Spring HW5 Report

<BME> <LI-WEI YANG>

<b07611001>

Public Score	Private Score
32.17	30.65

The methods I used to pass the strong baselines include:

1. Change the model structure to big in [1706.03762] Attention Is All You Need, with head number = 8, max token = 8192/2, to prevent GPU from out of memory situation.
2. Average checkpoint\_bests, not checkpoint\_last. Also, try and error on the number of checkpoints averaged, from 2 - 5, pick the best one out of it.