2023 Fall Deep Learning Report

109651066林柏佑

**Screenshot of task-1**

* + Number of layers
  + Parameter size
  + Accuracy

**Screenshot of task-2**

* + Number of layers
  + Parameter size
  + Accuracy

**In task-2**

* + Which kind of transformer-like model do you choose？
  + The reason why you choose this model.
  + The advantage of chosen model.

**Anything you do to improve the performance.**

其實有學者提出，先做Layer Norm再做Residual，效果上會有提升！

參考論文：[On Layer Normalization in the Transformer Architecture](https://arxiv.org/abs/2002.04745)

**Screenshot of your transformer code for both encoder layer and encoder**

* + **Plagiarism is forbidden !!!**
  + **There should be comment in your code**
  + **For both Task-1 and Task-2**