HW4

Graded

Student

黃資芸

Total Points

4 / 4 pts

Question 1

Make a brief introduction about a variant of Transformer. **2** / 2 pts

✔ **– 0 pts** Correct

– 0.5 pts slightly wrong

– 1 pt Not detailed enough.

– 1 pt incorrect

– 2 pts Wrong

– 2 pts Directly copy from the internet is not allowed.

– 2 pts Not a variant of Transformer.

Question 2

Briefly explain why adding convolutional layers to Transformer can boost performance. **2** / 2 pts

✔ **– 0 pts** Correct

– 0.5 pts Slightly incorrect.

– 0.5 pts Not detailed enough

– 1 pt This is not the main reason.

– 1 pt Should be more detailed.

– 1 pt There're some mistakes

– 2 pts Wrong

Questions assigned to the following page: 1 and 2

R10945015.

1. Make a brief introduction about a variant of Transformer.

ex. sandwich transformer: reorder sublayer module, more self-attention toward bottom and more freeforward sublayer toward the top.

ex. universal transformer: 一般transformer 輸入經過Attention後，會進入fully connected layer，而 universal transformer 會一直進入共享權重的 transition function 繼續計算，且用 Adaptive Computation Time (ACT) 控制循環次數.

2. Briefly explain why adding convolutional layers to Transformer can boost performance.

Transformer 因為self-attention的設計，針對大範圍前後有相關特徵資訊，但會損失局部特徵，加上convolutional layer 可提取局部細微特徵，綜合各自的優缺點.