ML HW_04 Report

1. Make a brief introduction about a variant of Transformer. (2 pts)

Ans: Big Bird is a variant of a transformer which focuses on improving efficiency of the attention matrix. Its attention matrix is the combination of random attention, window attention and global attention. This improvement makes Big Bird focus on finding relationships between random frame, neighbor frame and globally information; therefore, Big Bird typically yields a better result than vanilla transformer.

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

Convolutional layers could work greatly when data has strong locality. For example, frames in audio usually have a tight relationship with its neighbor frames. Neighbor frames could form a word or phoneme when considered as a whole, which gives the network a lot more information than considering a single frame. Thus, adding convolutional layers to the transformer is to exploit this trait in the dataset; and in most cases, this adjustment could boost its performance.