

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

Conformer.

Conformer is used to replace transformer encoder in our HW4. The main difference of conformer is the "conformer block", which has a "convolution module" in-between, but transformer block hasn't. To be more specific, "convolution module" is try to imitate the CNN structure.

Through the paper, the motivation to combine "transformer" and "CNN" is that:
"transformer models are good at capturing content-based global interactions, while CNNs exploit local features effectively."

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

This is because, the convolutional layer can obtain the local features or local content effectively. This helps a lot for the transformer to make up its disadvantage, since transformer is not good at extract fine-grained local feature patterns.

As a result, we can obtain "both local and global dependencies of an audio sequence in a parameter-efficient way".



HW4

● GRADED

STUDENT

梁峻璋

TOTAL POINTS

4 / 4 pts

QUESTION 1

Make a brief introduction about a variant of Transformer.

2 / 2 pts

QUESTION 2

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

2 / 2 pts