

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

A variant of Transformer is Conformer. Conformer is a model which combines models with self-attention and models with convolutional layers to achieve better performance. The convolutional layers are added in the encoder of Transformer.

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

Convolutional layers can extract local features and patterns greater than models with self-attention such as transformer. In speech recognition, audio context and sequences have some important local features to be captured.