

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

Conformer is a Convolution-augmented Transformer.

Transformer models are good at capturing **content-based global** interactions, while **CNNs** exploit **local features** effectively.

Conformer combine convolution neural networks and transformers to model **both local and global dependencies** in a parameter-efficient way.

Ref:

<https://arxiv.org/abs/2005.08100>

<https://github.com/sooftware/conformer>

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

Transformer models are good at capturing **content-based global** interactions, while **CNNs** exploit **local features** effectively.

如果只有 **Transformer** 的話，無法有效 exploit local features，因此加入 CNN 可以補足 **Transformer** 不足的地方。