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

Conformer:
與Transformer一樣也是一個seq2seq的模型，不同的是，它加入了卷積的運算進入Encoder的部分。
Conformer Encoder的Conformer Block之總體架構由三大部分組成：(1) 使用Swish的Activation Function和dropout的Feedforward Module (2) 使用相對位置編碼並加入pre-norm殘差與dropout的Multi-head self attention Module (3)使用pre-norm殘差的 Convolution Module。其中，每個Module都使用了Residual殘差計算。而Convolution和Attention透過串連可達到增強的效果。

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

基於Self attention layer設計的Transformer，在針對大範圍前後有相關的特徵互動的資訊，有較好的效果，但在提取局部細微的特徵時表現較為遜色。而CNN的Convolutional layer擅長提取局部細微的特徵，但缺點就是需要大量參數或是模型深度來理解整張圖片的全域特徵關係。

Conformer的架構就是將Self attention layer與Convolutional layer做結合，擷取各自的優點，並優化模型的表現。