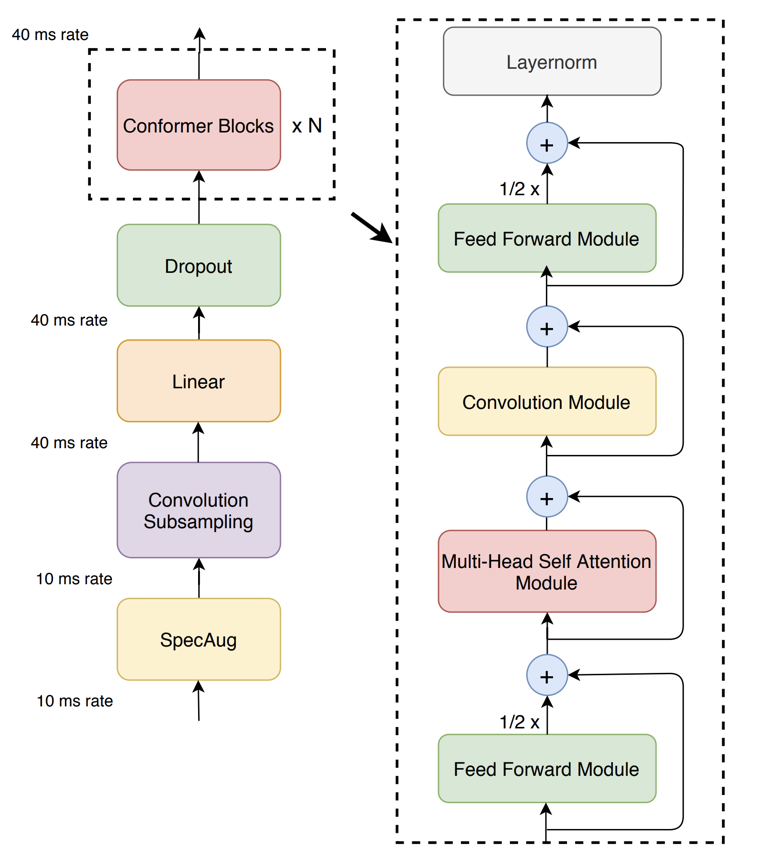
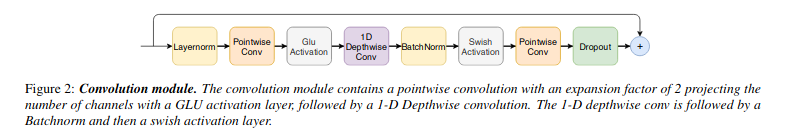
1. Make a brief introduction about a variant of Transformer.

Ans:

Transformer是Google 提出的一種不使用 RNN和CNN，而使用self-attention的架構，在處理NLP這種有前後順序、上下文關係很有效。而在一般的圖像處理中，大多使用的網路是有卷積層的CNN架構，利用卷積來提取局部特徵。Conformer就是結合了原本的Transformer與卷積這兩種特性，應用在NLP上。



Conformer 主要是由4 個block組成，也就是上圖右邊的feed forward module、multi-head self attention module、convolution module最後再接一個feed forward module。



convolution module中先對layer做normaliztion，使用 pointwise convolution、GLU activation layer，一個近年發現能收斂比Relu快的Activation，然後是一維的depthwise convolution、Batchnorm，然後是 Swish activation layer。

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

performance.

Ans:

上述有提到Transformers很適合用於長形的context中，但是對於較局部的local feature pattern就很難提取出。而將convolutional layers加入到Transformers中便能解決這個缺點，因為CNN的convolutional layers能有效提取local information，且使用position-based kernels，可以維持 translation equivariance，加強整體performance。