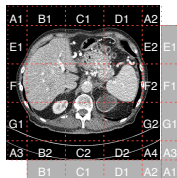
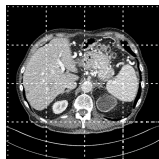
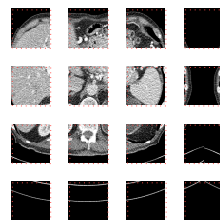
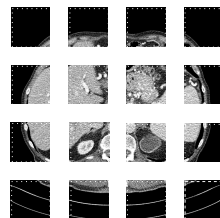




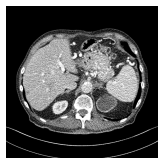
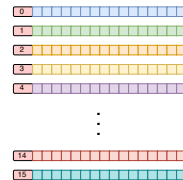
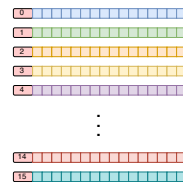
$shift = 0$ and $shift = \frac{win_size}{2}$



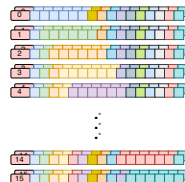
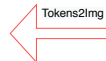
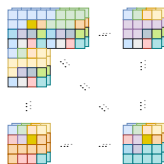
Patches extraction



Img2Tokens + Position Embedding



Pointwise Convolution+ GELU



Transformer Block+ Concatenate

