

L(Sentence Length)
Min_L = 2
Max_L = 256
D_model = 512
N_heads = 8
D_model/n_heads = 64

