

base : 16 n : 1024 emb : 128 depth : 2 heads : 4

epochs: 400 l2: 0.0001 lr: 0.0010 dropout: 0.5000 params: 398080