

lr: 0.0010

12 : 0.3000

epochs: 1000

heads: 4

params : 398208

dropout : 0.5000