Training set loss 0.5 loss decode.loss_ce 0.4 0.3 Loss 0.2 0.1 0 5000 10000 15000 20000 25000 30000 step