Evolution of loss during training
accuracy=0.7712
lr=0.005, betas=(0.9, 0.999), weight_decay=0.0005, hessian_power=1