Evolution of loss during training: renormalized, adam
accuracy=0.37
lr=0.001, betas=(0.9, 0.999), weight_decay=0.0005, hessian_power=1