

Training with:
base_lr = 0.000500; stepsize = 50000; lr_mul = 0.000100
Finetuning: trial_5; Iter = 50000

