

Weight-Decay $\lambda=0.01$, Mini-Batch Size B=500

