

β_2	0.99	0.999	0.9999	0.99999
$T = 10^3$	0.7183 ± 0.0169	0.6990 ± 0.0005	0.7313 ± 0.0129	0.7004 ± 0.0108
$T = 10^4$	0.0384 ± 0.0075	0.0398 ± 0.0002	0.0238 ± 0.0007	0.0217 ± 0.0002
$T = 10^5$	0.0071 ± 0.0022	0.0023 ± 0.0007	0.0027 ± 0.0022	0.0000 ± 0.0000

Table 1: Final training loss (average of two runs) with standard deviations. The ResNet-18 model was trained on the CIFAR-10 dataset with varying β_2 and total iteration number T . The other hyper-parameter setups follow the default ones. **Bold numbers** indicate the lowest training loss for each T , which roughly aligns with the result.