

Figure 1: **Illustration of  $\mathcal{GNC}$  and train accuracy across different network architectures on CIFAR100 (top), Tiny-ImageNet (bottom) datasets.** We train the networks on CIFAR100 with  $d = 10$ ,  $K = 100$  and Tiny-ImageNet with  $d = 10$ ,  $K = 200$ .

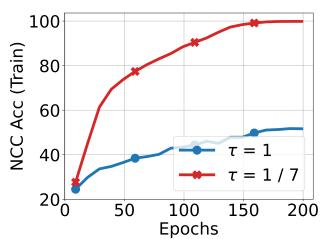


Figure 2: **Illustration of NCC training accuracy with different temperatures.**

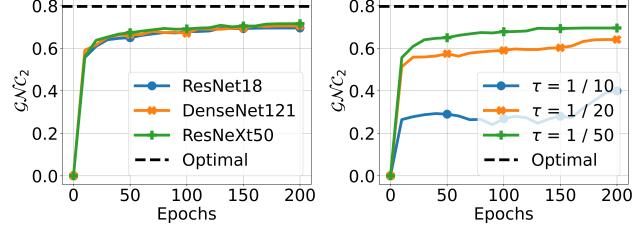


Figure 3: **Illustration of  $\mathcal{GNC}_2$  on BUPT-CBFace-50 dataset across different network architectures and temperatures.** We train the networks on BUPT-CBFace-50 with  $d = 512$ ,  $K = 10,000$ .

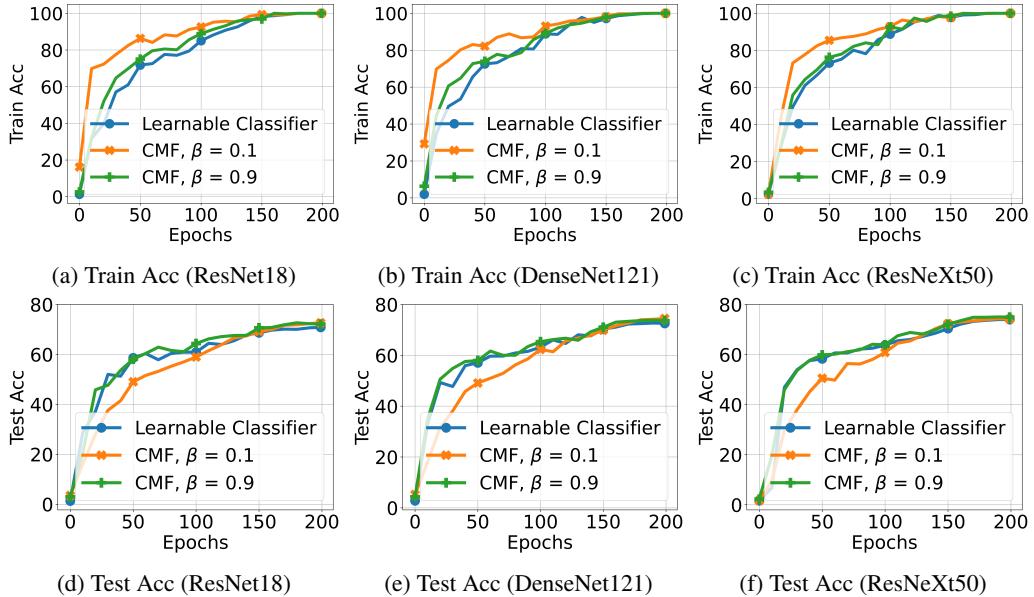


Figure 4: **Comparison of the learning curve with learned classifiers vs. CMF classifiers across ResNet18 (left), DenseNet121 (middle), and ResNeXt50 (right) on CIFAR100 dataset.** We plot the evolutions of training accuracy (top) and test accuracy (bottom) under  $K = 100$  and  $d = 10$ .