Readme

The Experiments For 《Generalization Bound and New Algorithm for Clean-Label Backdoor Attack》

What we offer here is the code that find backdoor poison by algorithm 1.

1, 'Backdoor_attack_method.py' is code of the algorithm 1 in paper on dataset CIFAR-10, and the effect of poisoning was also tested in this file.

1.1, In its line 29, 'bud' is the budget of poison, you can change it from 8/255 to 32/255, or some others you like; in line 30, 'num' is the number of poisoned samples in training set, you can change it from 300 to 500, not more than 5000; in line 31, network=1 means use ResNet, network=2 means use VGG; in line 32, lp is target label, not more than 9.

1.2, This file can be run directly.

1.3, For the purpose of saving time, we load a trained network F_1 'vgg-9l.pt'(which was trained by 'small_vgg_9layers.py') instead of training it from scratch, if you want to train F_1 but not load 'vgg-9l.pt', use 'Small_vgg_9layers.py' to get a new trained F_1.

2, 'Small_vgg_9layers.py' is code to train F_1.

2.1, In its line 188, it is the path to save F_1, once you change it, in order to load the model correctly in 'Backdoor_attack_method.py', you need to modify line 124 of 'Backdoor_attack_method.py' .

2.2, This file can be run directly.

3, See 'Requiresments' for experiment environment.