Attacking Layer model.layer4.0.conv2 Max Success rate (%), 10 trials, Normal Init 0.003 - 0.0 6.2 12.5 12.5 6.2 12.5 6.2 12.5 6.2 6.2 12.5

- 100

80

0.006 - 0.0 18.8 12.5 12.5 18.8 18.8 18.8 12.5 18.8 12.5 18.8 6.2 43.8 31.2 43.8 31.2 31.2 18.8 25.0 18.8 25.0 31.2 0.013 - 6.2 43.8 43.8 37.5 31.2 37.5 43.8 31.2 31.2 25.0 25.0 0.016 - 6.2 62.5 62.5 56.2 50.0 50.0 50.0 43.8 37.5 43.8 25.0 0.019 - 6.2 81.2 81.2 81.2 75.0 68.8 56.2 62.5 43.8 50.0 43.8 0.022 - 6.2 81.2 81.2 93.8 93.8 87.5 81.2 75.0 62.5 68.8 56.2

60 40 0.025 - 6.2 93.8100.087.5 93.8 81.2 81.2 81.2 75.0 75.0 75.0 20 6.2 93.8100.093.8100.093.8 87.5 93.8 81.2 87.5 75.0 0.031 - 12.5100.0100.093.8100.093.8100.093.887.593.887.50.009 0.013