pacity during training process for different data sizes, numbers of parameters, and numbers of layers Experiment 3: 100000 model size, number of layers 90000 16, 1 32, 1 80000 **---** 64, 1 128, 1 70000 16, 2 32, 2 60000 -**-** 64, 2 Capacity 128, 2 50000 data size 1000 40000 10000 50000 30000

