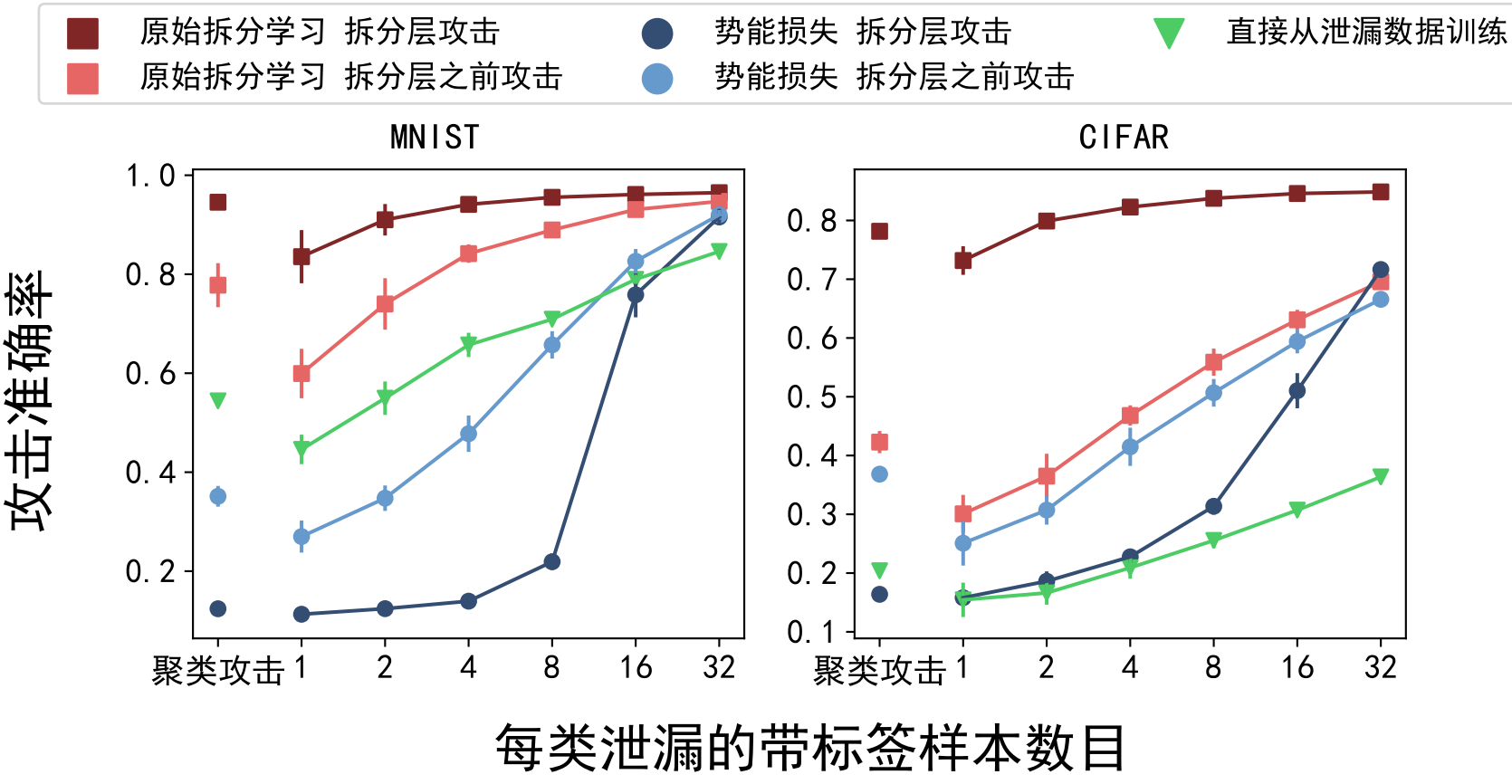
原始拆分学习 拆分层攻击
原始拆分学习 拆分层之前攻击
势能损失 拆分层攻击
势能损失 拆分层之前攻击
直接从泄漏数据训练
MNIST
CIFAR
攻击准确率
1.0
0.8
0.6
0.4
0.2
0.7
0.5
0.3
0.1
聚类攻击
1
2
4
8
16
32
每类泄漏的带标签样本数目