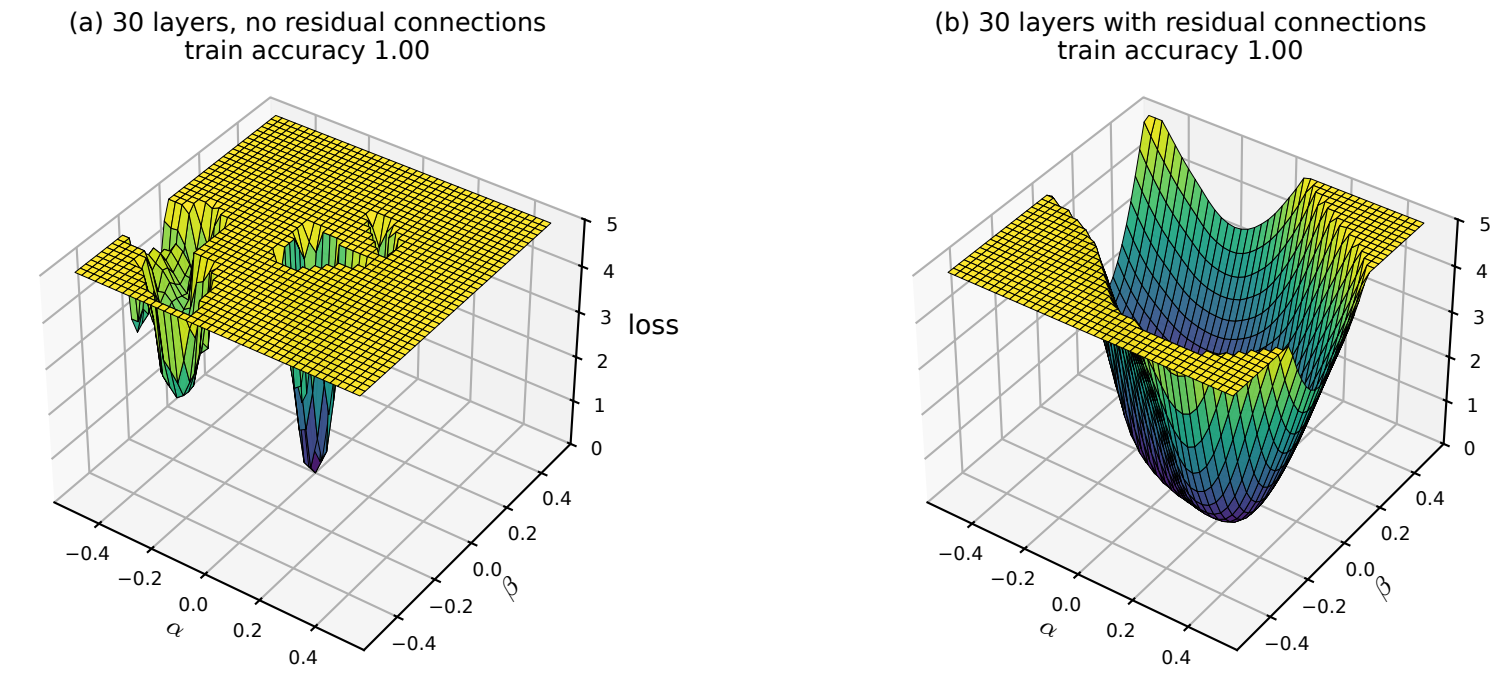

(a) 30 layers, no residual connections
train accuracy 1.00
(b) 30 layers with residual connections
train accuracy 1.00
loss
0
1
2
3
4
5
−0.4
−0.2
0.0
0.2
0.4
α
β