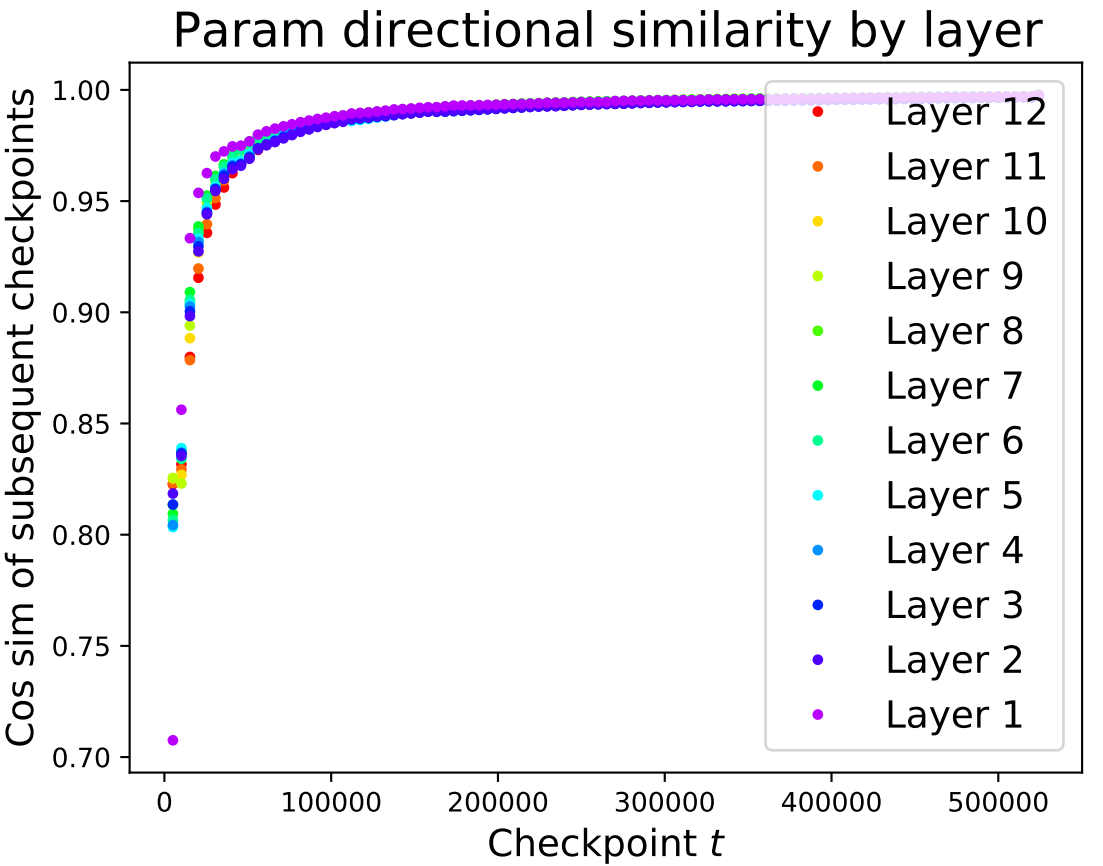
Param directional similarity by layer
Cos sim of subsequent checkpoints
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0
100000
200000
300000
400000
500000
Checkpoint t
Layer 12
Layer 11
Layer 10
Layer 9
Layer 8
Layer 7
Layer 6
Layer 5
Layer 4
Layer 3
Layer 2
Layer 1