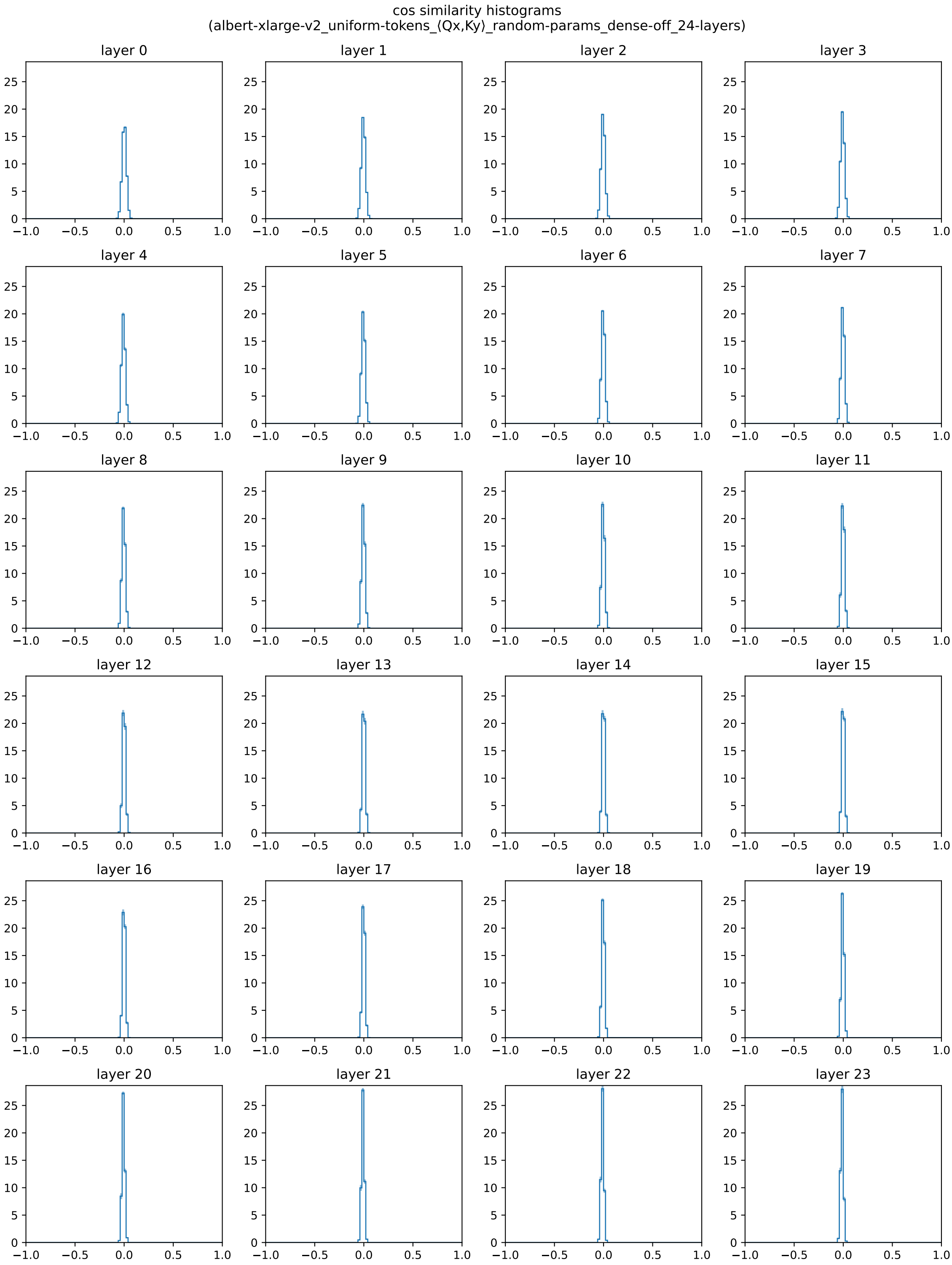
cos similarity histograms
(albert-xlarge-v2_uniform-tokens_⟨Qx,Ky⟩_random-params_dense-off_24-layers)
layer 0
layer 1
layer 2
layer 3
layer 4
layer 5
layer 6
layer 7
layer 8
layer 9
layer 10
layer 11
layer 12
layer 13
layer 14
layer 15
layer 16
layer 17
layer 18
layer 19
layer 20
layer 21
layer 22
layer 23
0
5
10
15
20
25
−1.0
−0.5
0.0
0.5
1.0