cos similarity histogram - layer 23 (albert-xlarge-v2 uniform-tokens (Qx,Ky) trained-params dense-on 24-layer 6 5 4 3 2 1 -0.500.00 0.25 0.50 -1.00-0.75-0.250.75 1.00