euclidean T-SNE visualisations $(albert-xlarge-v2_wikitext_\langle Qx, Ky\rangle_trained-params_dense-off_24-layers)$ layer 0 layer 3 layer 1 layer 2 100 100 100 100 50 50 50 50 0 0 -50-50-50-50 -100-100-100-10050 0 50 100 50 100 0 100 0 50 100 -50 -100 -500 -50-50 -100 -100-100 layer 4 layer 6 layer 7 layer 5 100 100 100 100 50 50 50 50 0 0 0 -50-50 -50 **-**50 -100-100-100-50 0 50 100 -100-50 0 50 100 -100-50 0 50 100 -50 0 50 100 -100 -100layer 9 layer 10 layer 11 layer 8 100 100 100 100 50 50 50 50 0 0 0 0 -50-50-50-100 -100 -100-10050 50 100 100 0 50 100 0 50 100 -100-50 -100-50 0 -100-50 -100-50 layer 12 layer 13 layer 14 layer 15 100 100 100 100 50 50 50 50 0 0 0 -50 -50**-**50 -50-100-100 -100-100 0 50 -100 -50 0 50 50 -50 0 50 -100 -50100 100 -100-50 0 100 -100100 layer 16 layer 17 layer 18 layer 19 100 100 100 100 50 50 50 50 0 0 0 -50 **-**50 -50 --50 --100 | -100 -100 --100 | -100 -100 -Ó 50 -50 0 50 100 -100 -50 0 50 100 -50 100 -50 0 50 100 -100 layer 20 layer 22 layer 21 layer 23 100 100 100 100 50 50 50 50 0 0 0 0 -50-50 -50 -50 -100 --100 --100 --100 --100 **-**50 0 50 100 -100**-**50 50 100 -100-50 50 100 -100 -50 50 100