t-SNE visualisations $(albert-xlarge-v2_wikitext_\langle x,y\rangle_trained-params_dense-on_192-layers)$ after layer 8 after layer 16 after layer 24 after layer 32 100 100 100 100 50 50 50 50 0 0 -50-50 -50 -50 -100-100-100-10050 50 0 50 0 50 100 -500 0 100 -50100 -100 -50100 -50 -100 -100-100after layer 64 after layer 40 after layer 48 after layer 56 100 100 100 100 50 50 50 50 0 0 -50-50 **-**50 **-**50 -100-100-100-50 0 50 100 -100-50 0 50 100 -50 50 100 -50 0 50 100 -100-100-100after layer 80 after layer 96 after layer 72 after layer 88 100 100 100 100 50 50 50 50 0 0 0 -50-50**-**50 -100-100 -100-10050 100 -5050 100 -50 50 100 -50 0 50 100 -100-1000 -100-100-500 after layer 104 after layer 112 after layer 120 after layer 128 100 100 100 100 50 50 50 50 0 0 -50 -50 -50 -50-100 -100 -100-100 -50 0 50 50 -50 50 -50 0 50 -100 100 -100-500 100 -1000 100 -100100 after layer 136 after layer 144 after layer 152 after layer 160 100 100 100 100 50 50 50 50 0 -50 -50 --50 --50 -100 | -100 -100 -100 -100 -50 0 50 100 -100 -50 0 50 100 -50 0 50 100 -50 0 50 100 -100-100 after layer 176 after layer 168 after layer 184 after layer 192 100 100 100 100 50 50 50 50 0 0 0 -50-50 -50 -50

-100

-100

-50

100

50

-100

-100

-50

50

100

100

50

-100

-100

-50

0

50

-100

-100

-50

100