cosine T-SNE visualisation - layer 9 $(albert-xlarge-v2_wikitext_\langle Qx,Ky\rangle_random-params_dense-off_24-layers)\\ 100 + (Qx,Ky)_random-params_dense-off_24-layers)$ 75 50 25 0 -25-50-75-10050 -100-75-50-250 25 75 100