cos similarity heatmaps
(albert-xlarge-v2_wikitext_(x,y)_trained-params_dense-on_24-layers)