attention_logit t-SNE embeddings (wikitext albert-large-v2 params-trained dense-off layers-48) after layer 25 after layer 28 after layer 26 after layer 27 100 100 100 100 50 50 50 50 0 -50-50-50-50-100-100-100-10050 50 100 0 50 0 50 100 -500 100 -100 0 -50100 -50-100 -50-100-100 after layer 29 after layer 31 after layer 32 after layer 30 100 100 100 100 50 50 50 50 0 0 0 0 -50-50 -50 -50-100-100-10050 100 -5050 100 -50 0 50 100 50 100 -100 -50-100-100-100 -50after layer 33 after layer 34 after layer 35 after layer 36 100 100 100 100 50 50 50 50 0 0 0 0 -50-50-50-50-100-100-100-1000 50 100 50 100 0 50 100 50 100 -100-50-100-50-100-50-100-50after layer 37 after layer 39 after layer 38 after layer 40 100 100 100 100 50 50 50 50 0 0 -50-50-50 -50-100-100-100 -100 50 50 0 -100-500 100 -100 -500 100 -500 50 -50 50 100 -100100 -100after layer 41 after layer 42 after layer 43 after layer 44 100 100 100 100 50 50 50 50 0 0 0 0 -50 -50 -50 -50 -100 ↓ -100 -100 | -100 -100 --100 -50 0 50 100 -100 -50 50 100 -50 0 50 100 -50 0 50 100 -100 after layer 45 after layer 48 after layer 46 after layer 47 100 100 100 100 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 -100-5050 100 -100 -50 0 50 100