(1)Vanilla Transformer

(2)cosFormer

(3)cosFormer (w/o re-weighting)

(4)*cos* re-weighting matrix