

























features }10^{−1} 10^{-6} sep -10^{-11} non-trained layer -10^{-16} - 10⁻²¹ -10^{-26} - 10⁻³¹ trained layer - 10⁻³⁶ 0 1 2 3 4 5 6 7 8 91011 head