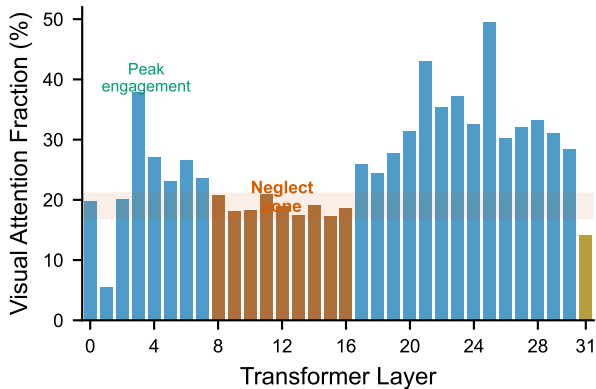


(a) Fraction of attention on visual tokens



(b) Visual attention entropy ratio

