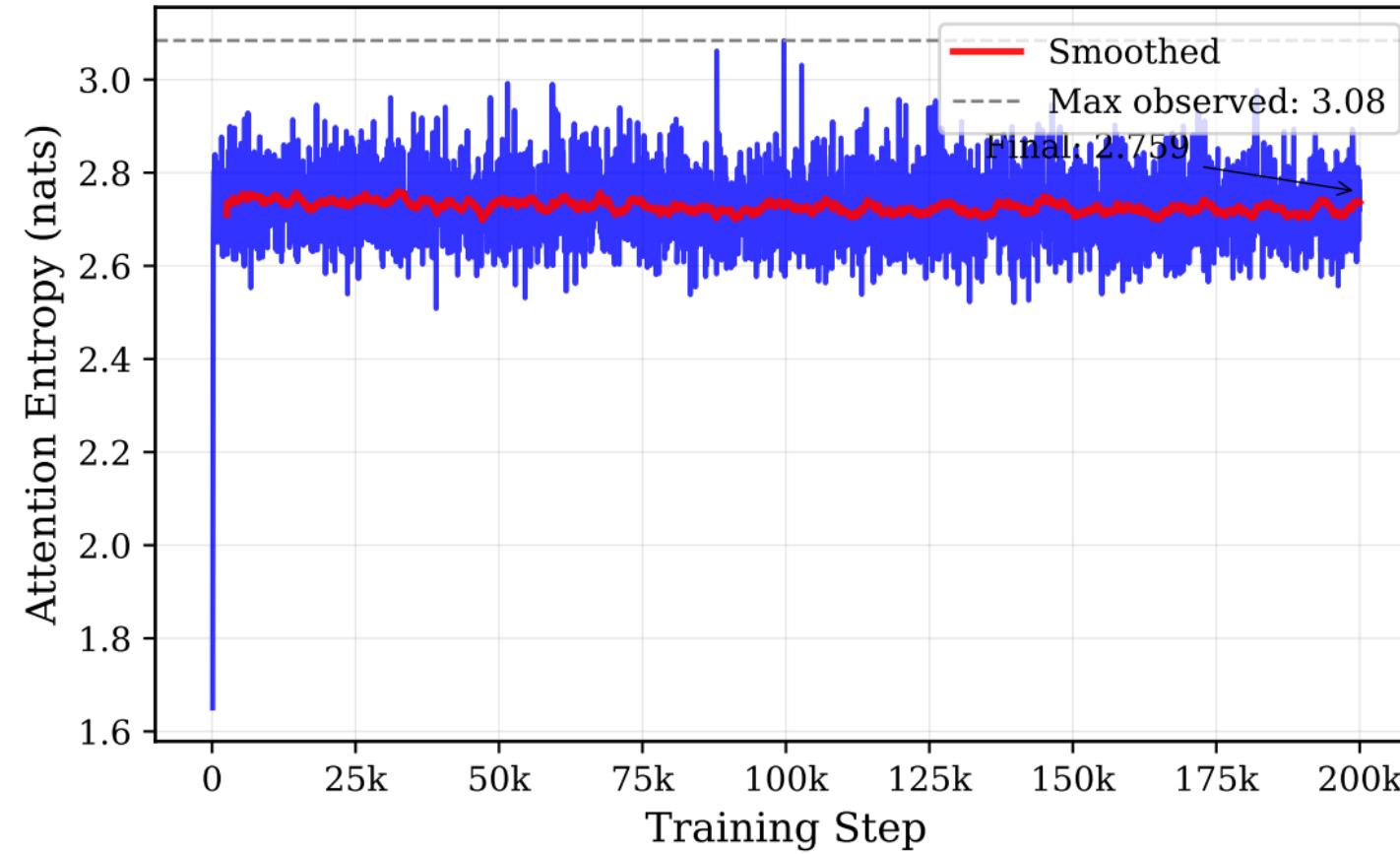
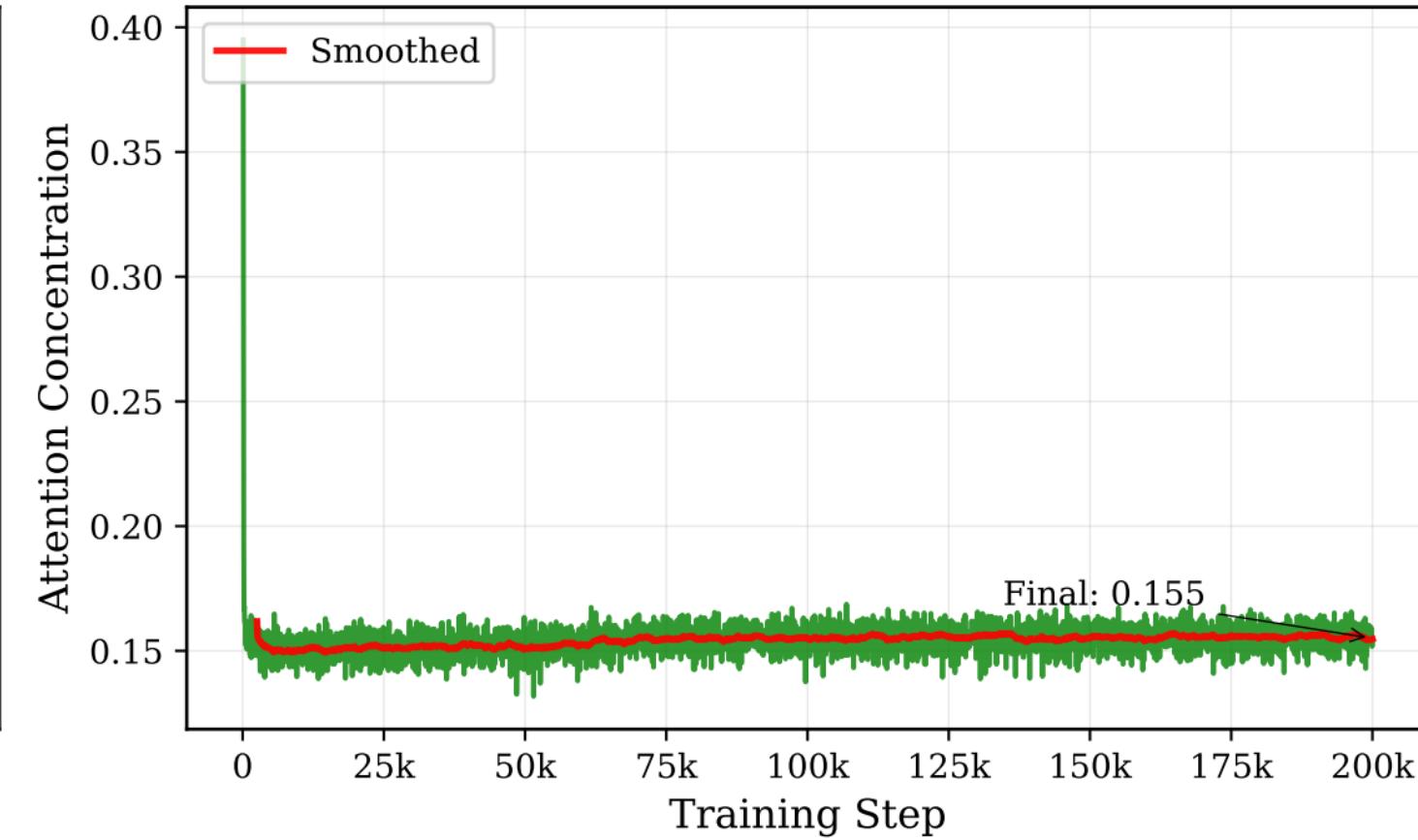


(a) Attention Entropy Over Training



(b) Attention Concentration Over Training



*Lower entropy = sharper attention (focused). Higher entropy = uniform attention (diffuse).*