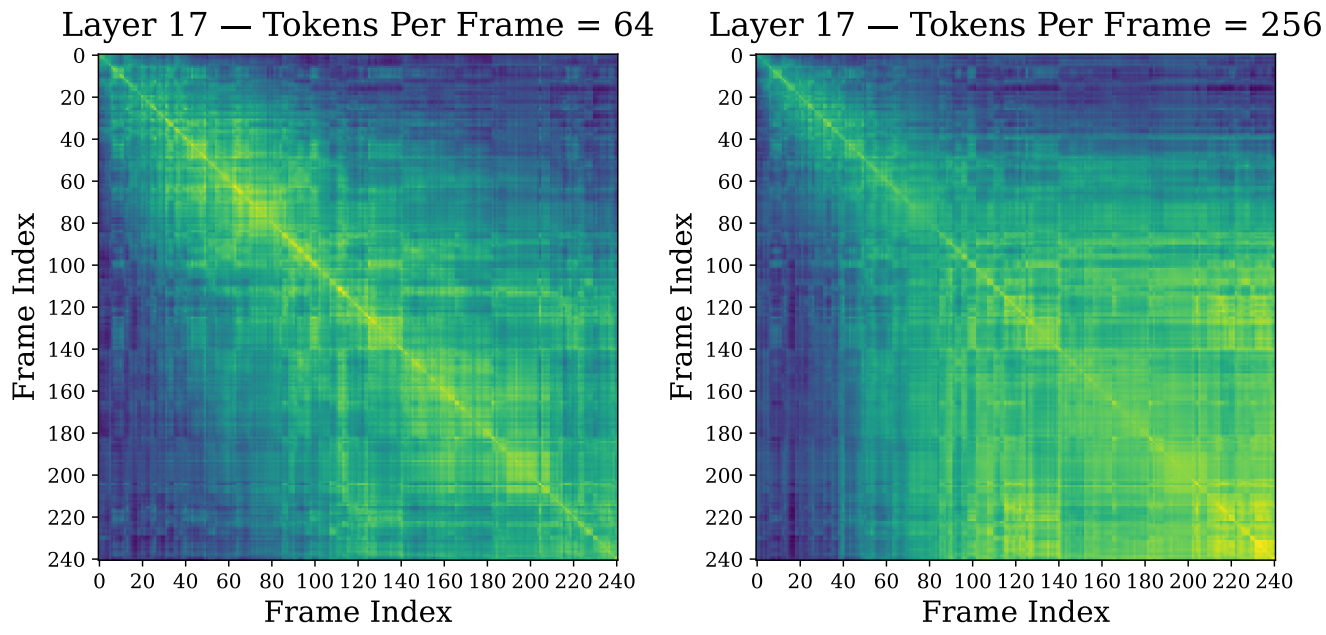
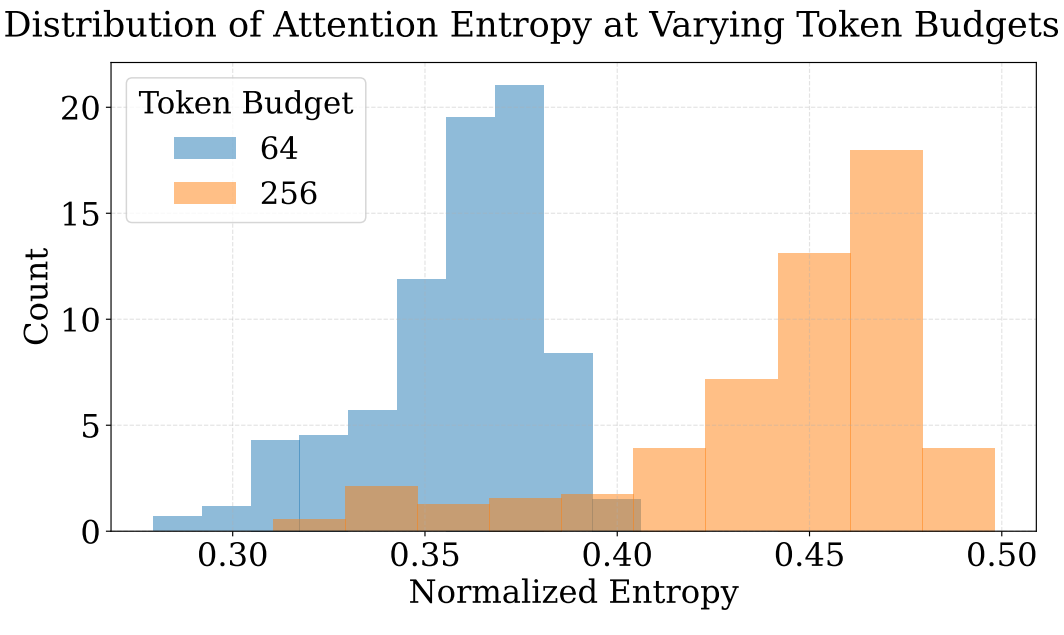


(a) Self-Similarity Matrices



(b) Attention Entropy Distribution