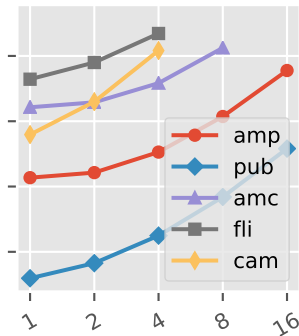
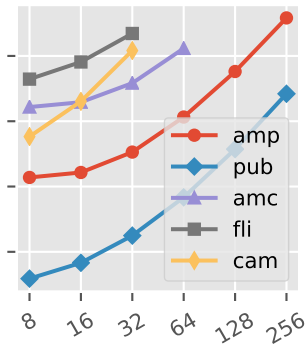
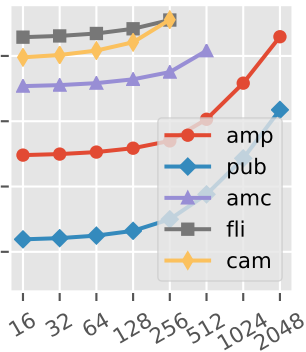


Training Memory Usage (MB)



#Head
 ($\dim(\mathbf{h}_x^1) = 64, d_a = d_v = d_m = 32$)