



多头注意力机制



用assert判断，词向量的维度需要能够整除头数（多头的头的个数）



