



Multiple heads  
(num= $h$ )

## Self Attention

### Scaled Dot-Product Attention

