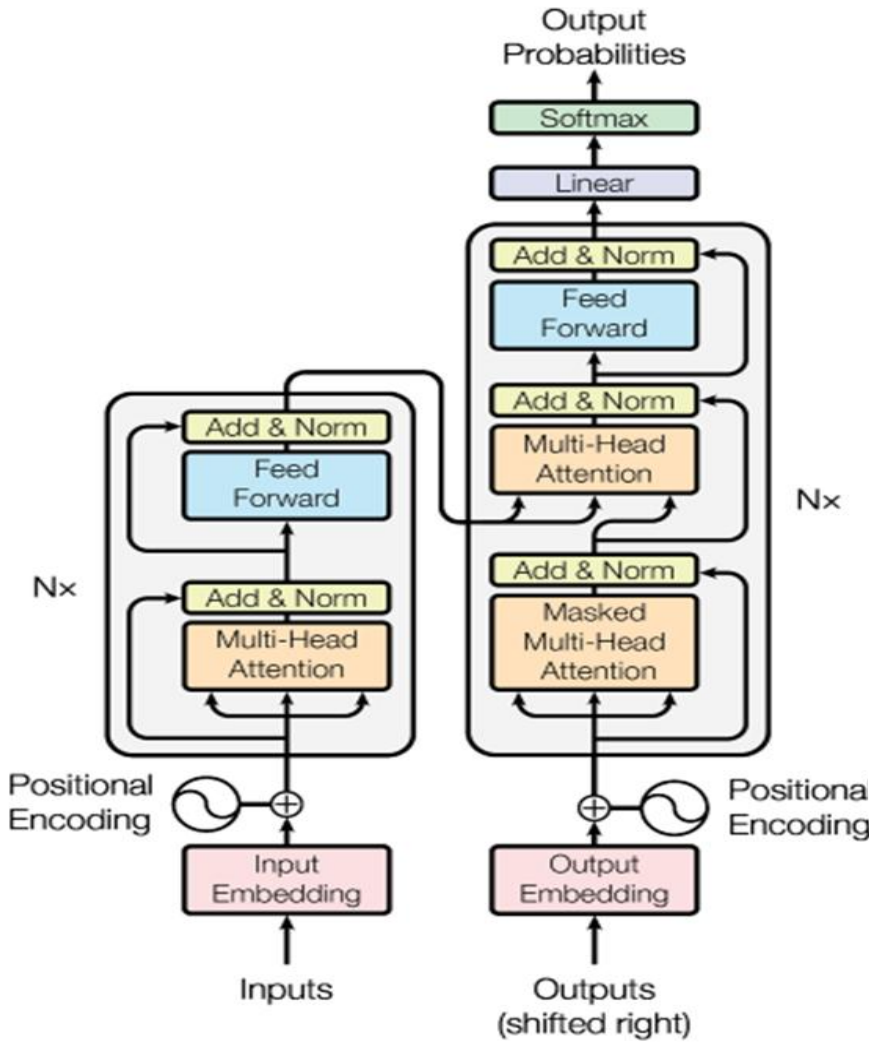
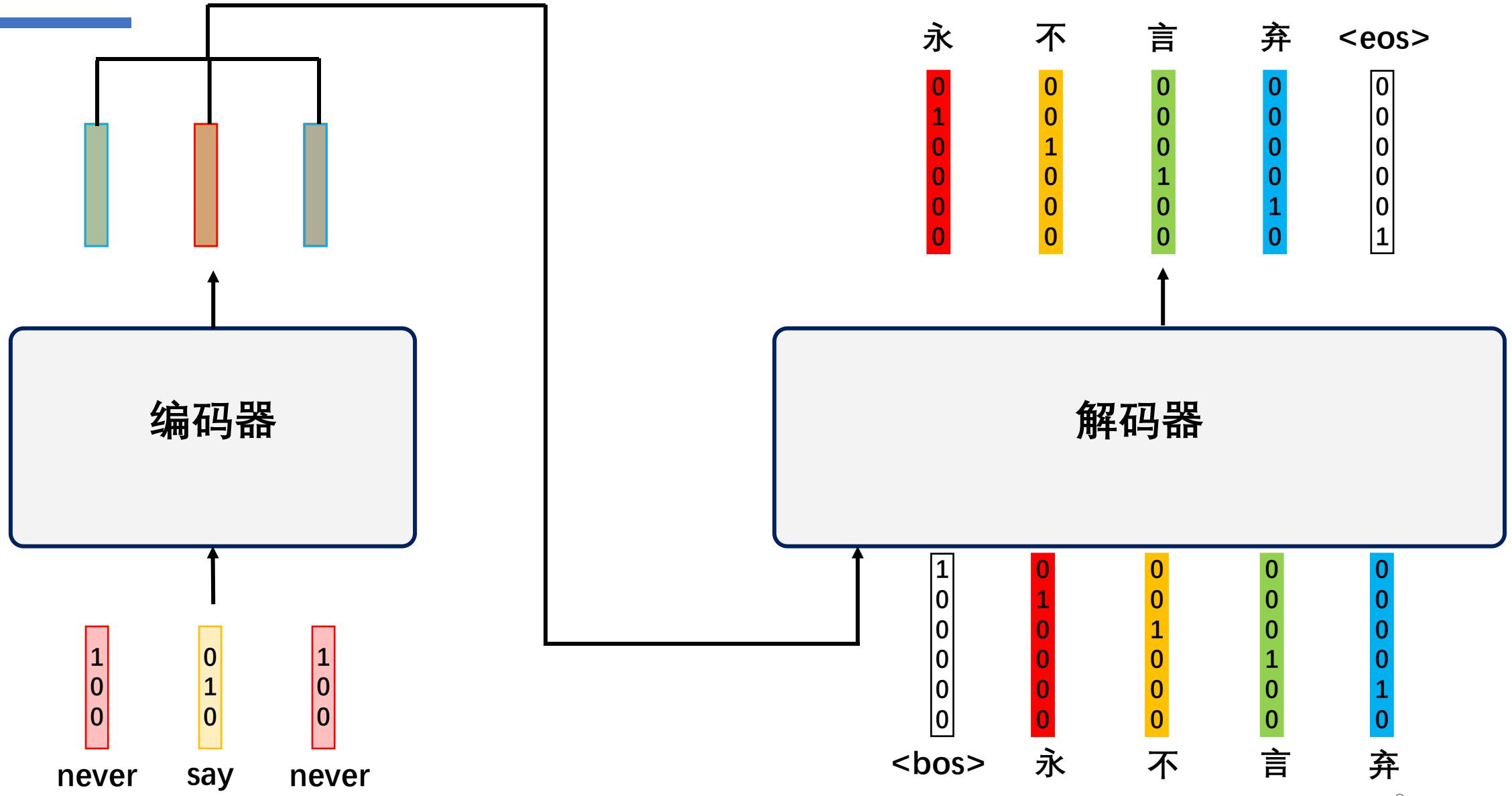
Transformer

Transformer

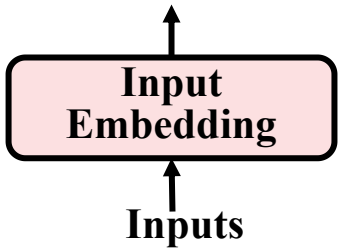


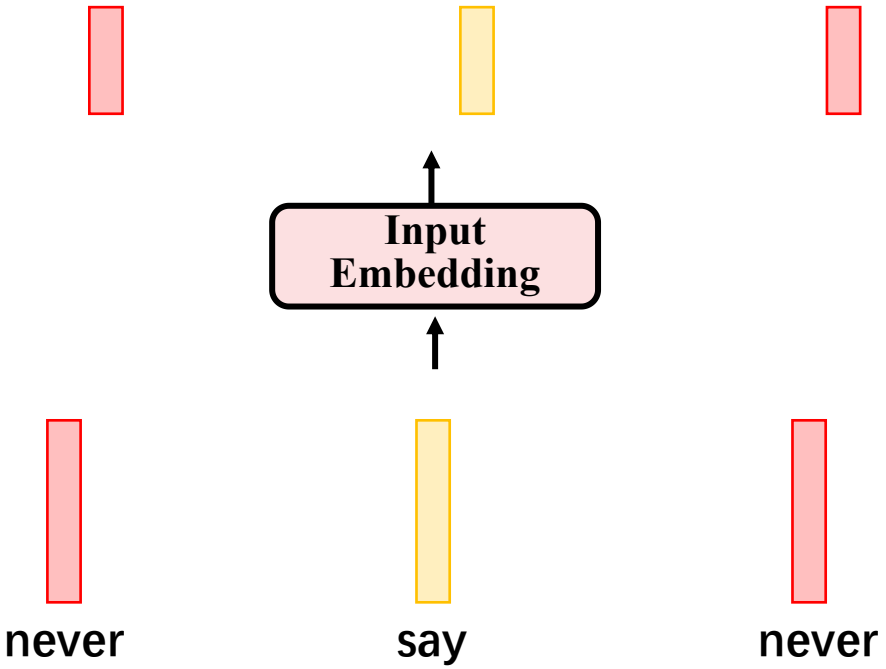
1.编码器-解码器
2.自回归

机器翻译

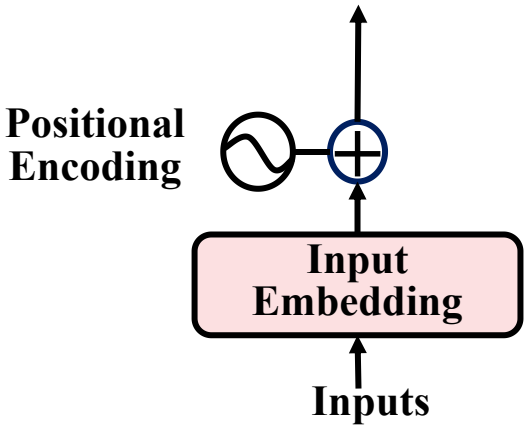


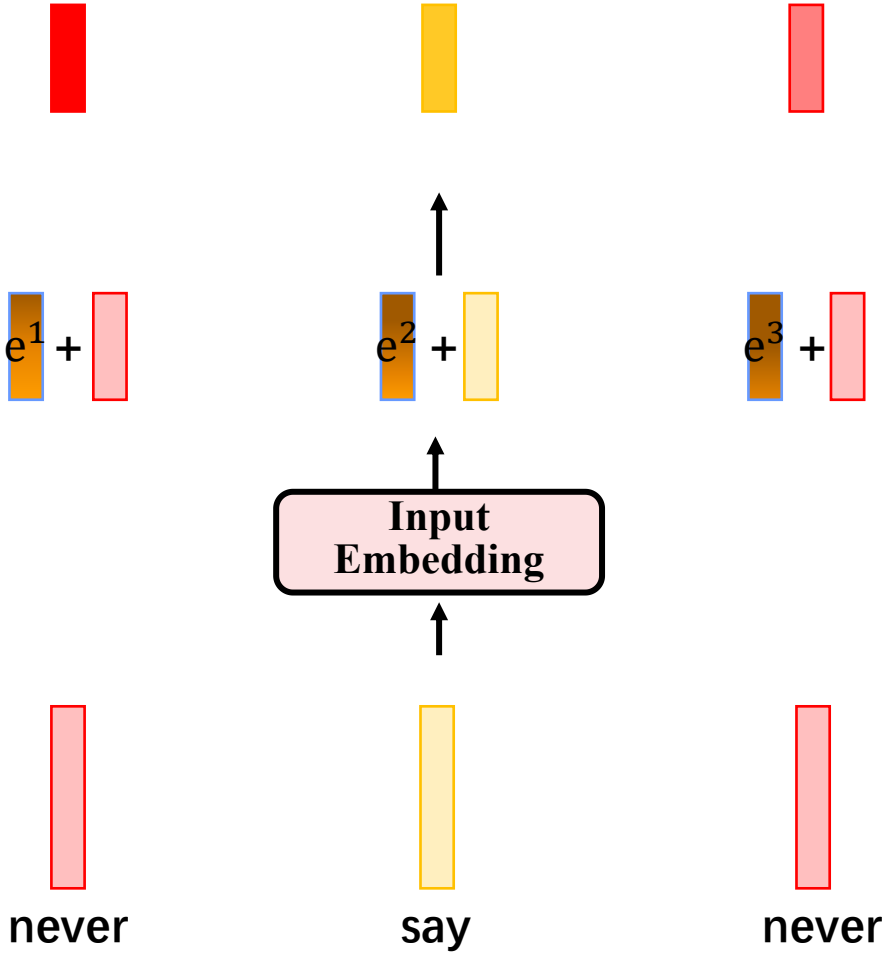
Transformer



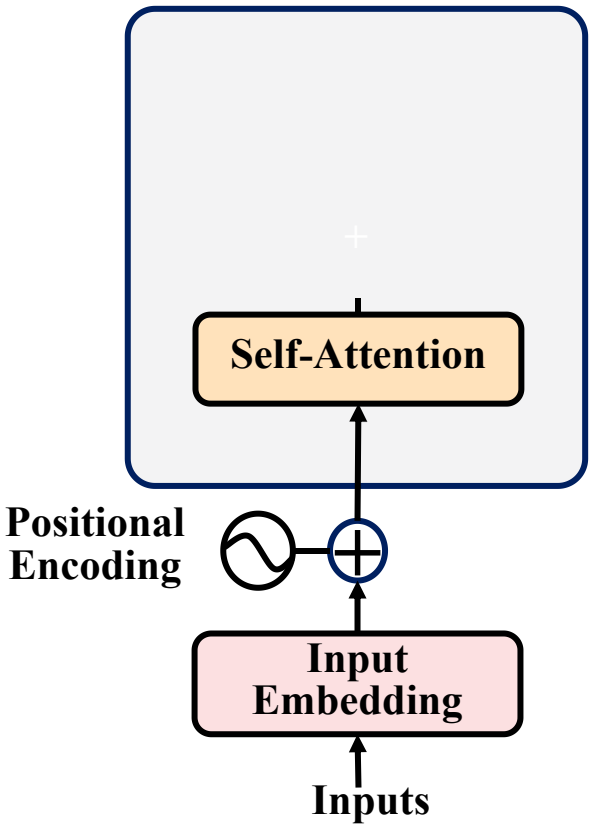


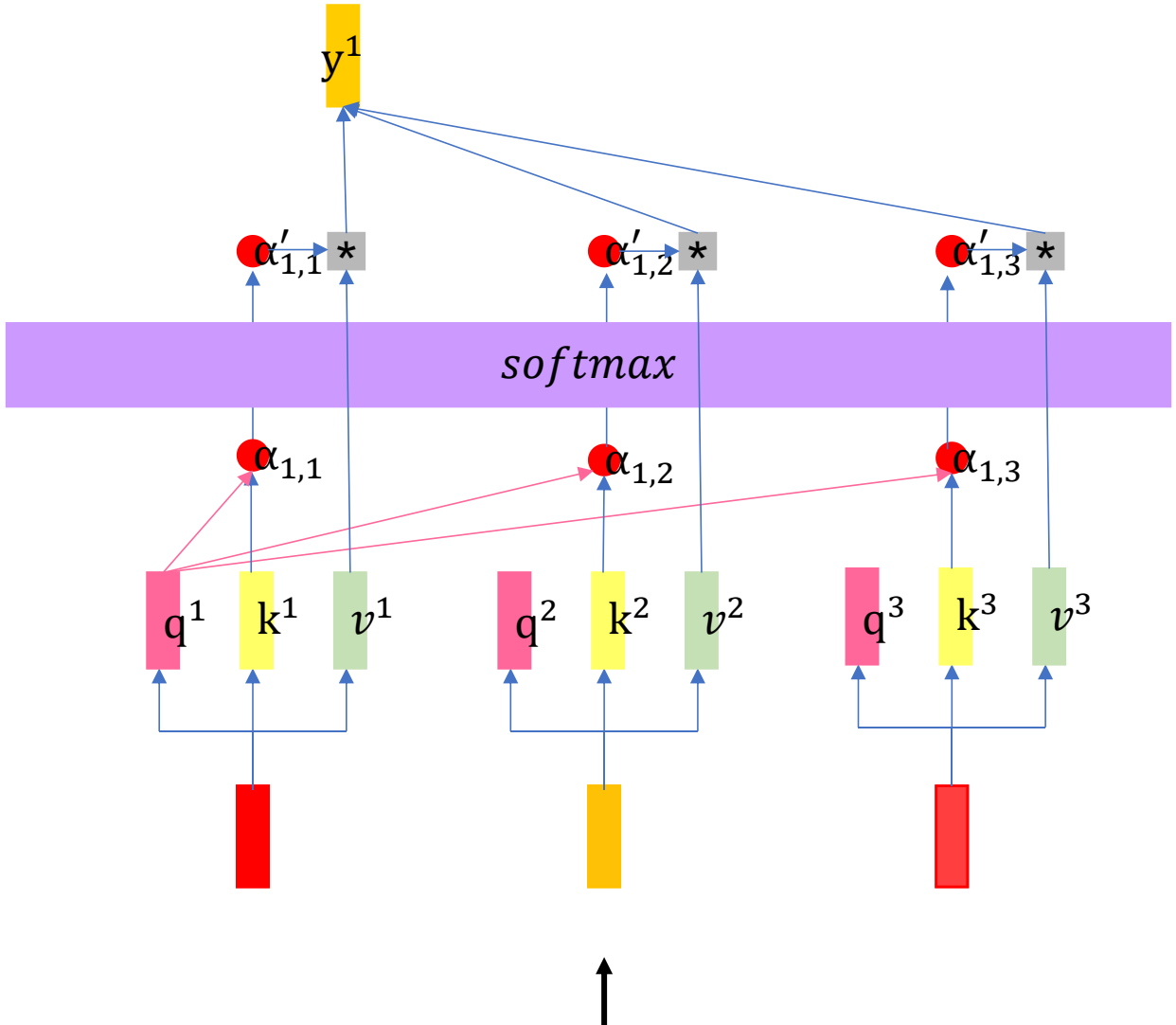
Transformer



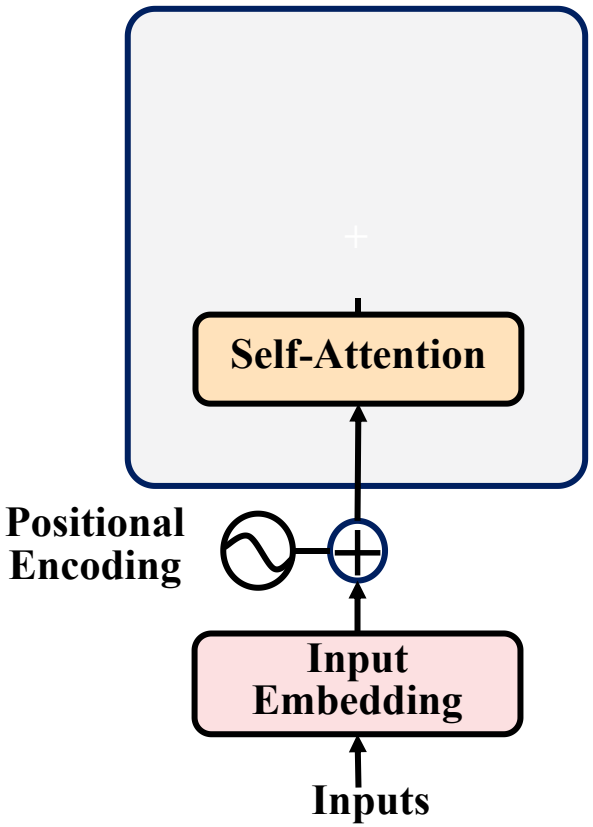


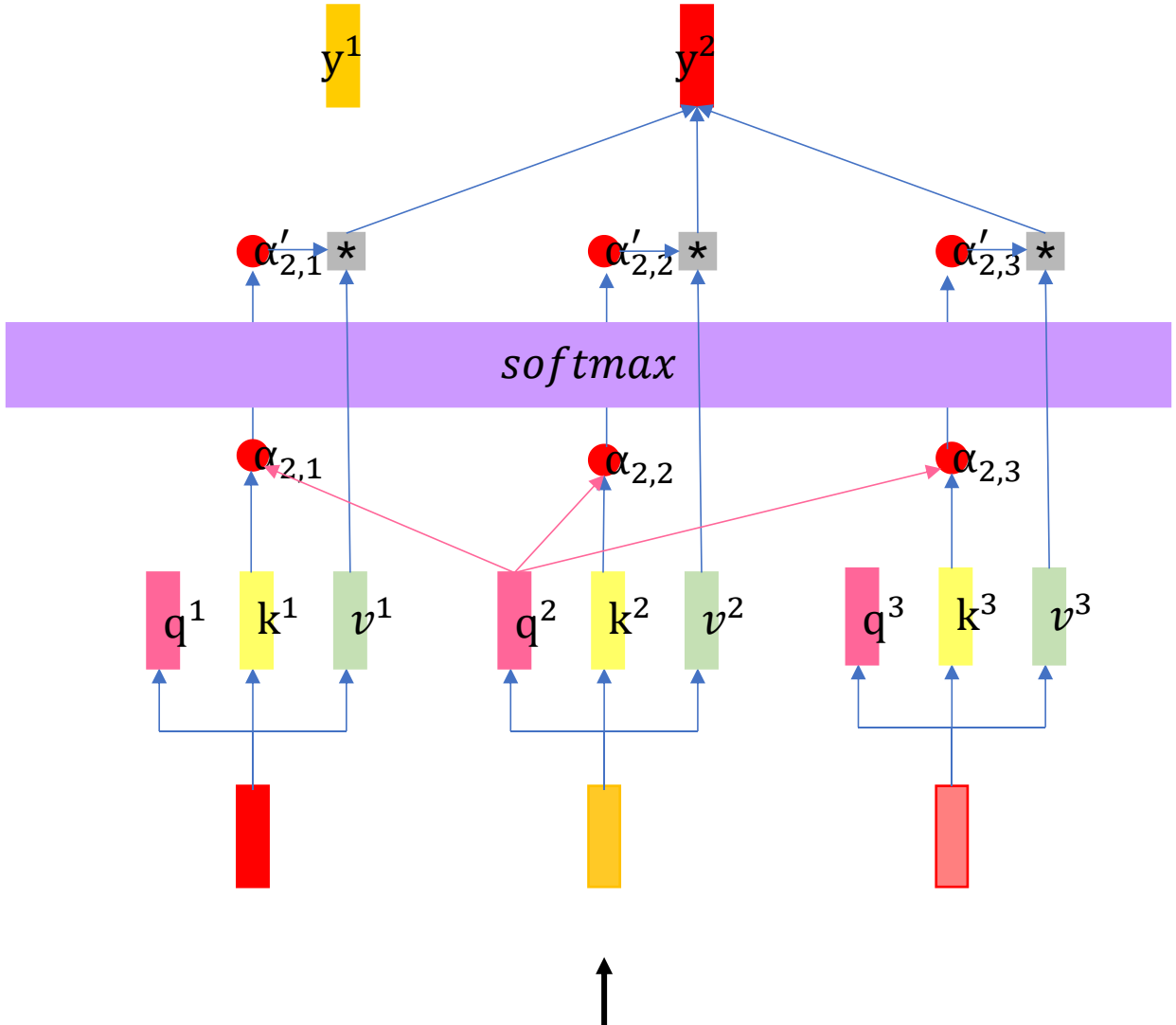
Transformer



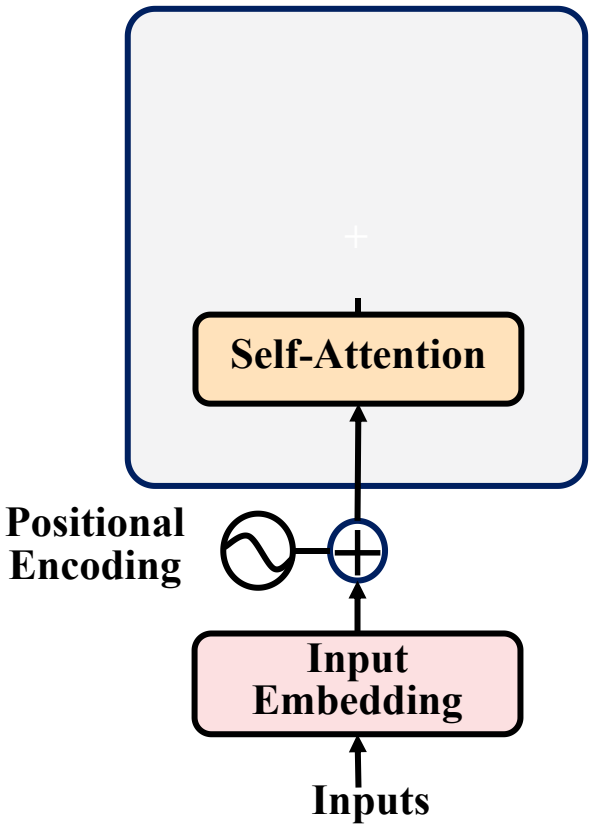


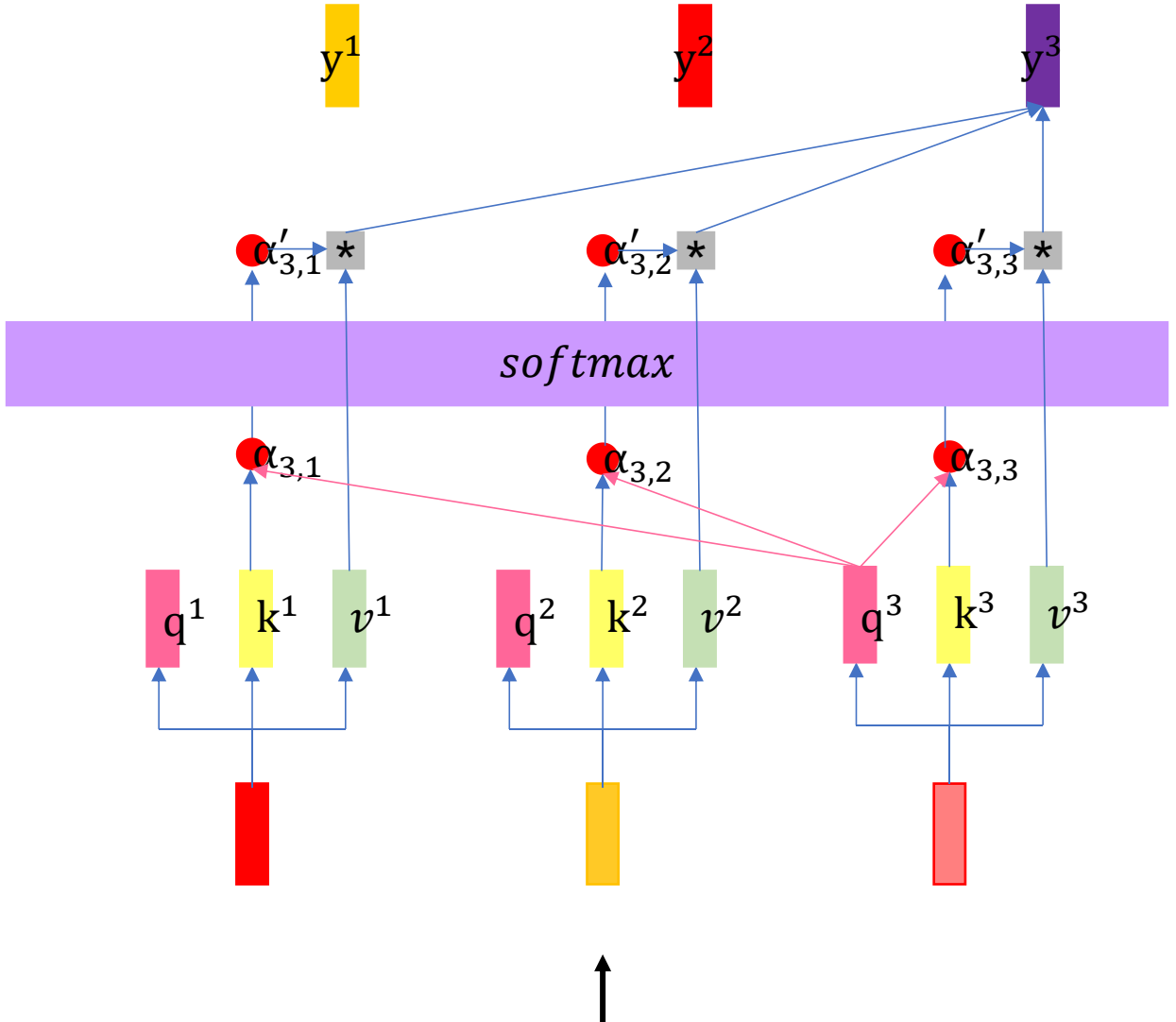
Transformer



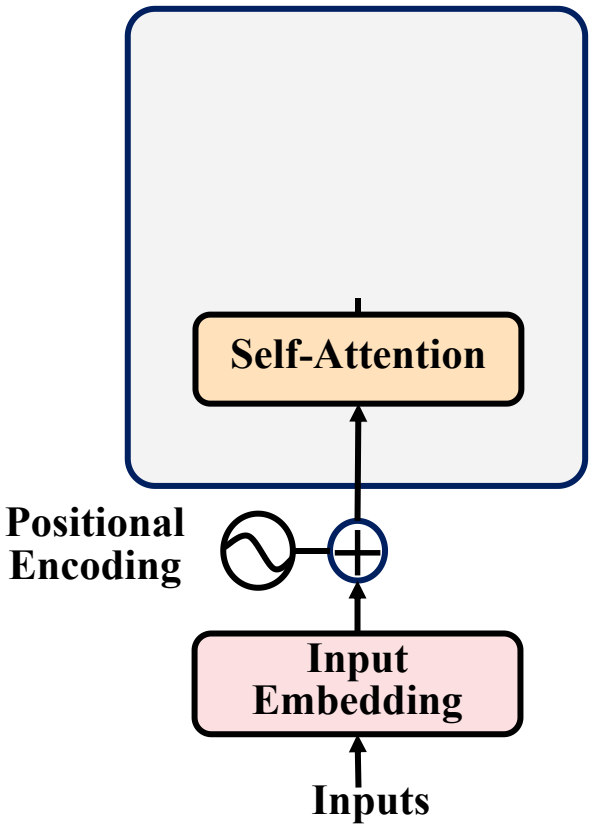


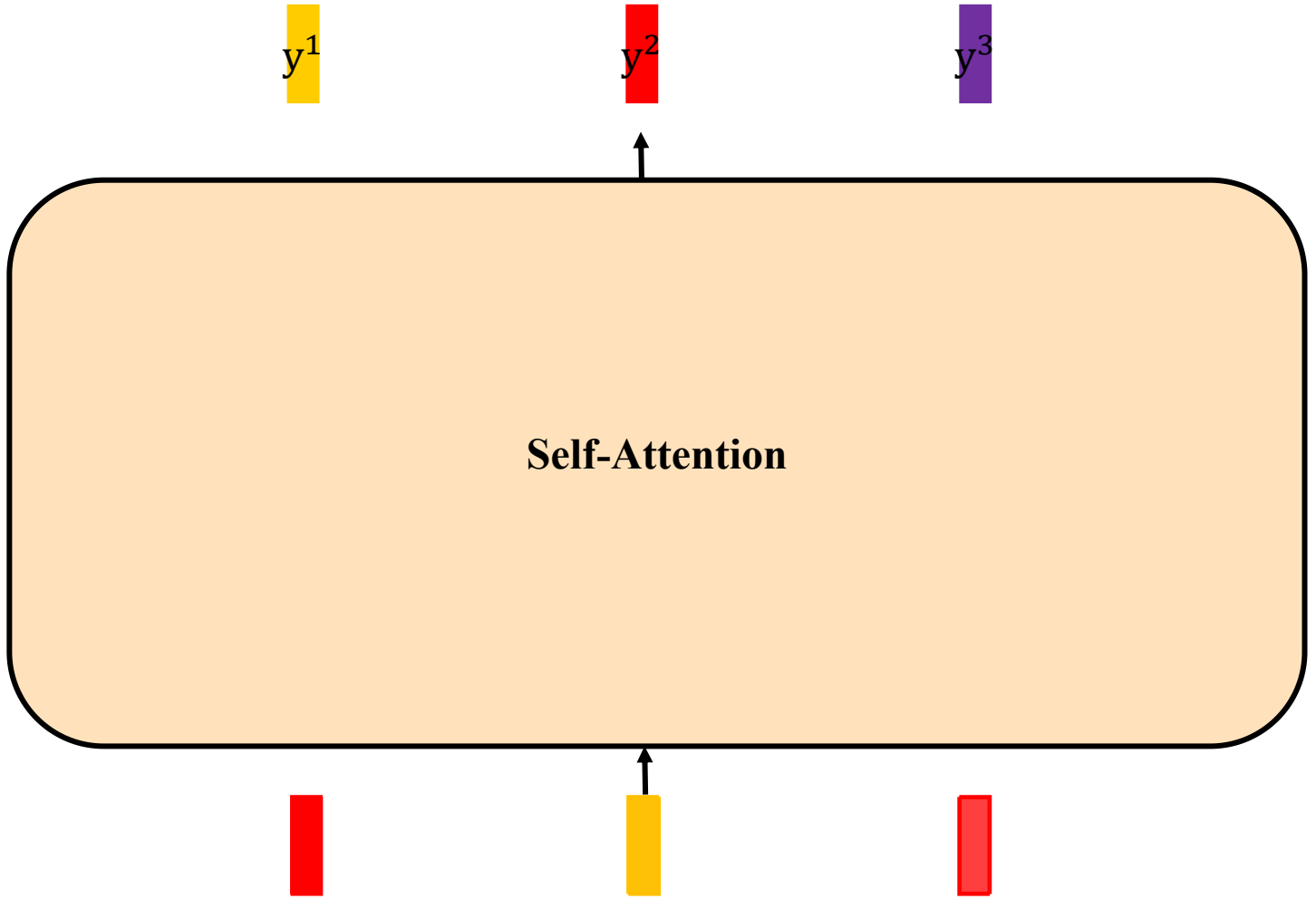
Transformer



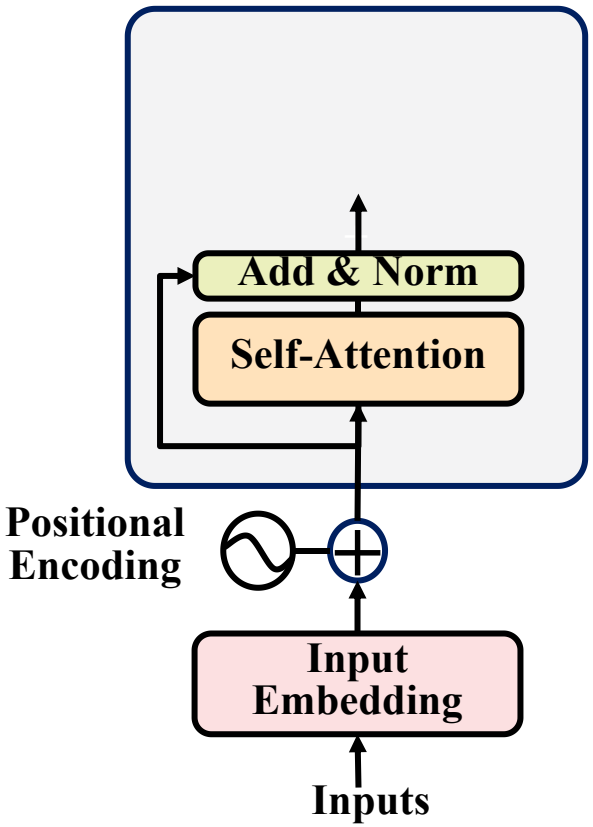


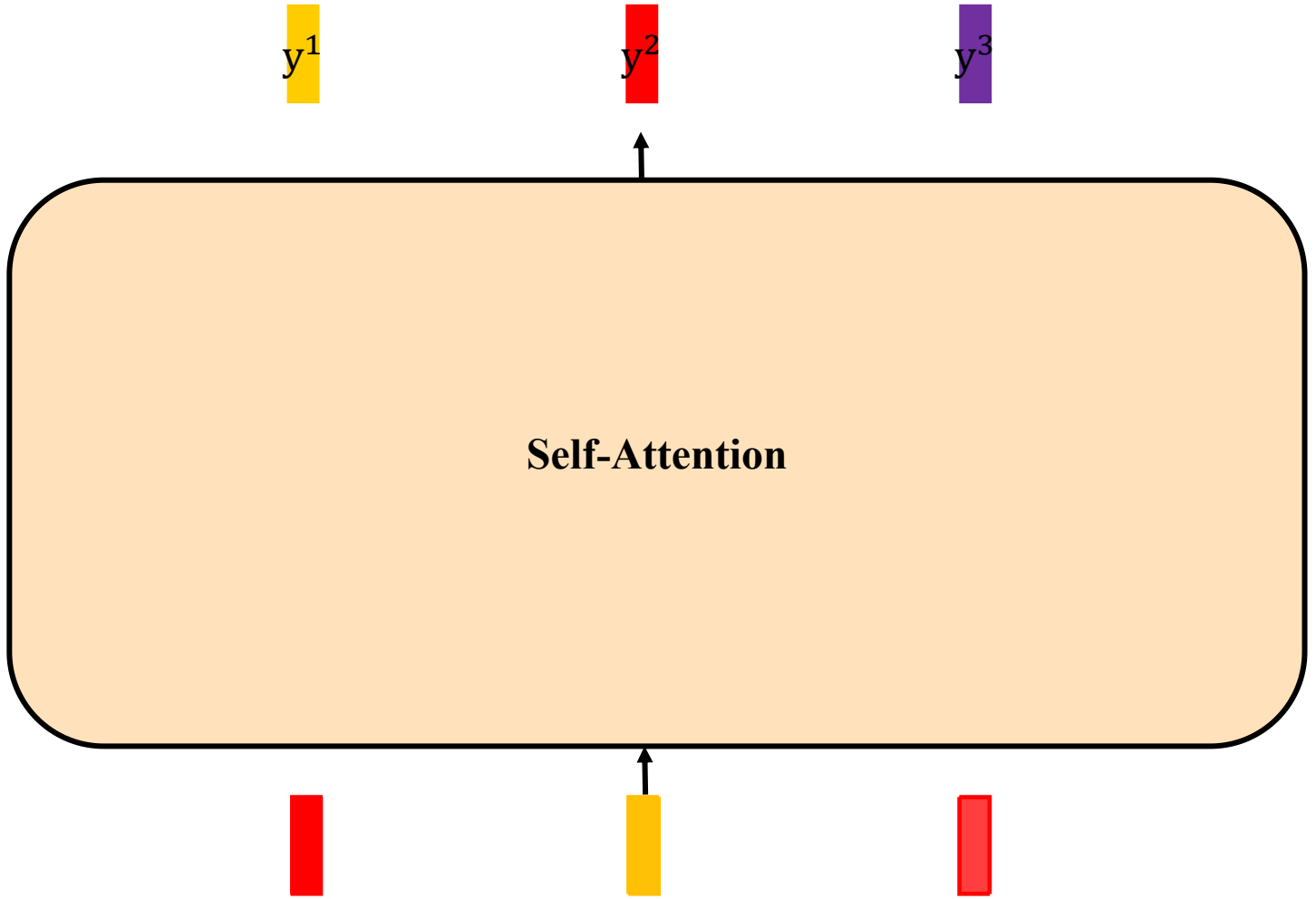
Transformer





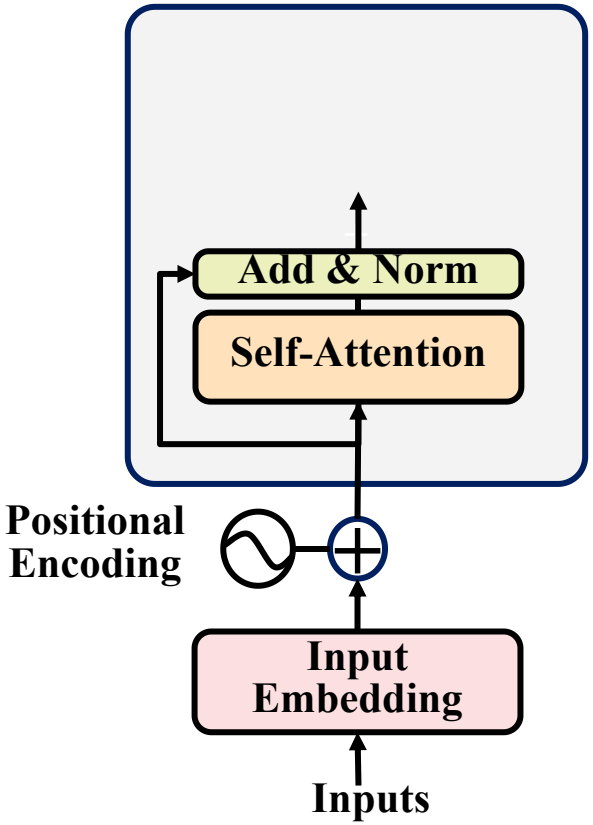
Transformer

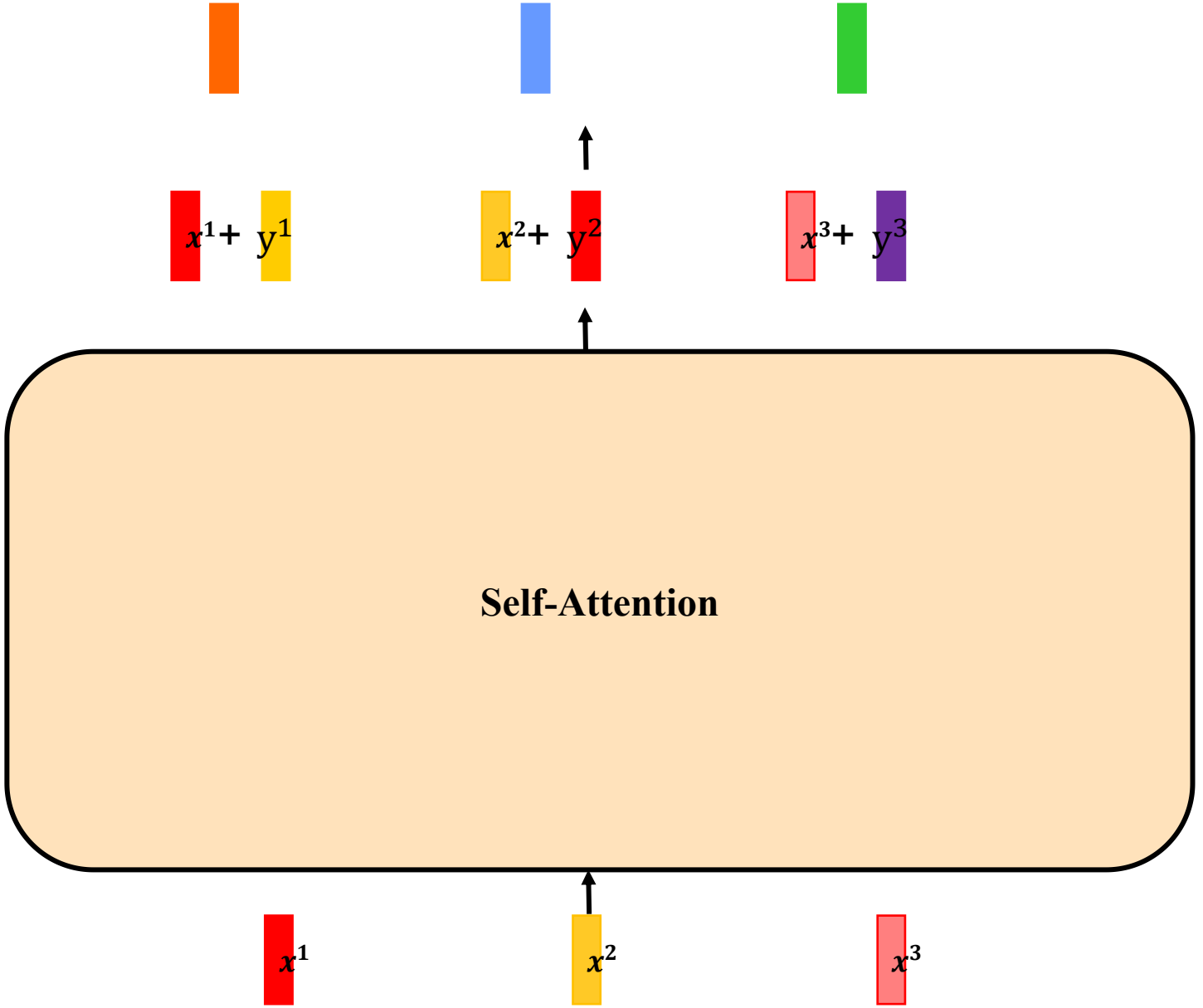




Transformer

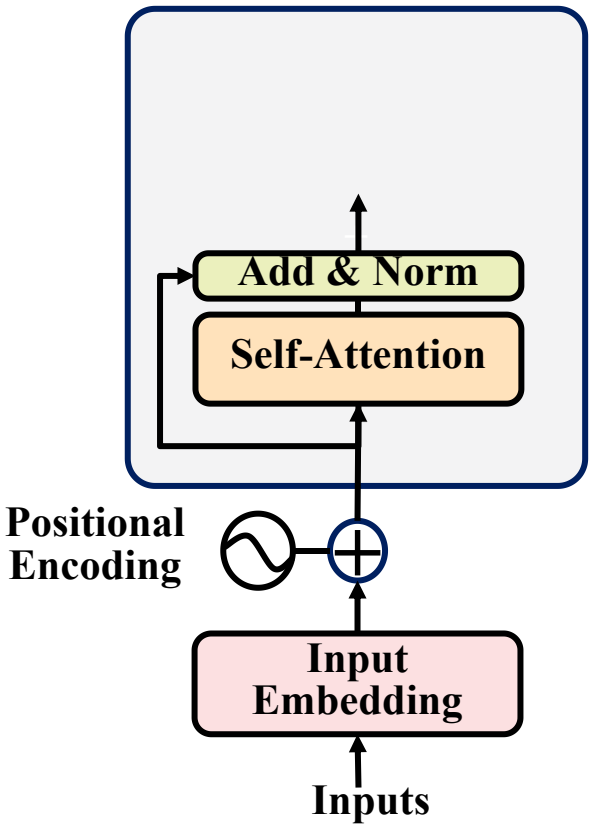
$$LayerNorm(x + Sublayer(x))$$





Transformer

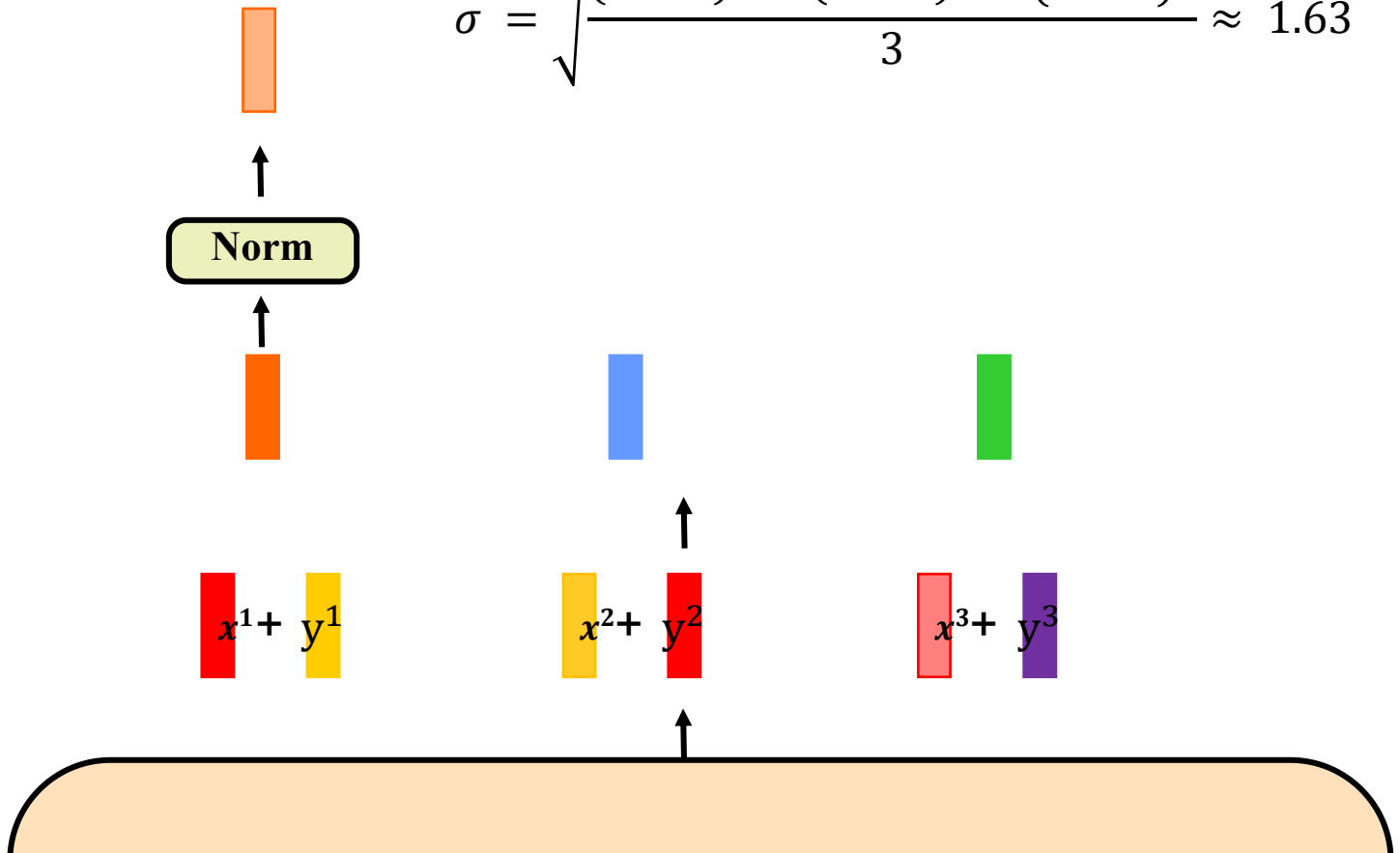
$$\boldsymbol{LayerNorm}(\boldsymbol{x} + \boldsymbol{Sublayer}(\boldsymbol{x}))$$



$$\begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}$$

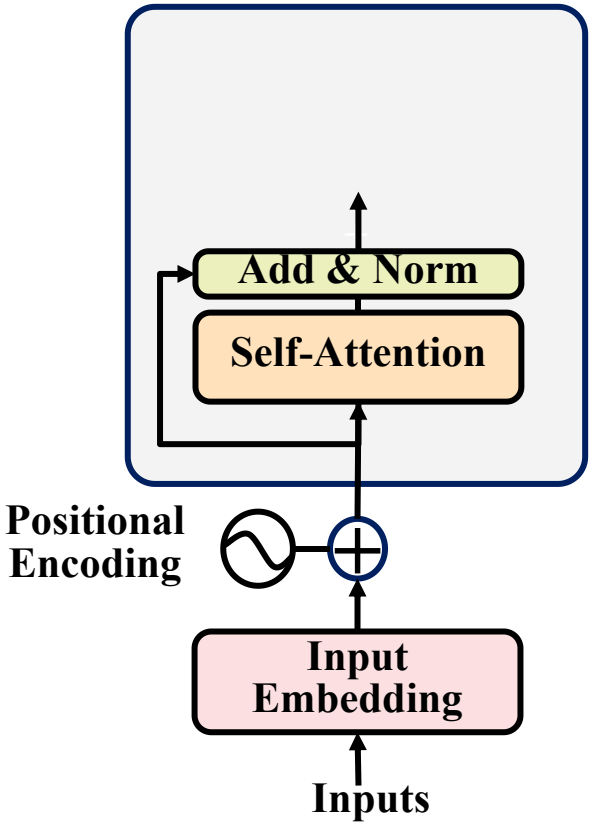
$$\mu = \frac{1+3+5}{3} = 3$$

$$\sigma = \sqrt{\frac{(1-3)^2 + (3-3)^2 + (5-3)^2}{3}} \approx 1.63$$



Transformer

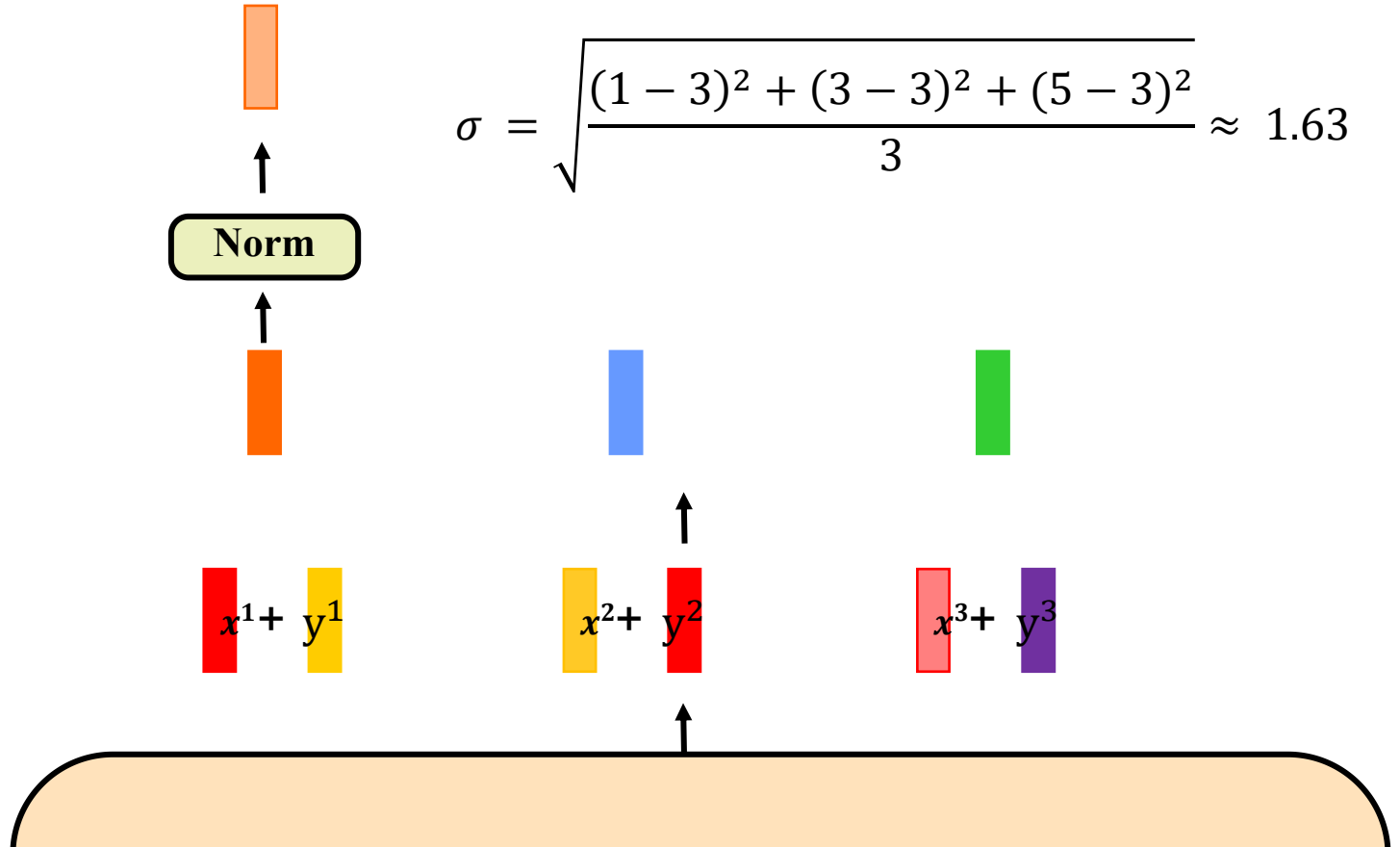
$$\boldsymbol{LayerNorm}(\boldsymbol{x} + \boldsymbol{Sublayer}(\boldsymbol{x}))$$



$$\begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} \qquad \begin{pmatrix} \frac{1-3}{1.63} \\ \frac{3-3}{1.63} \\ \frac{5-3}{1.63} \end{pmatrix} = \begin{pmatrix} -1.83 \\ 0 \\ 1.83 \end{pmatrix}$$

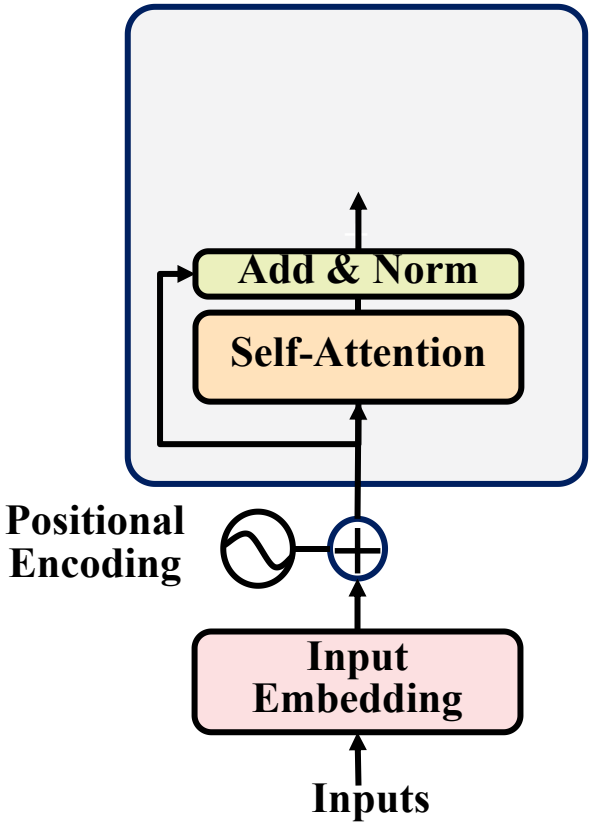
$$\mu = \frac{1+3+5}{3} = 3$$

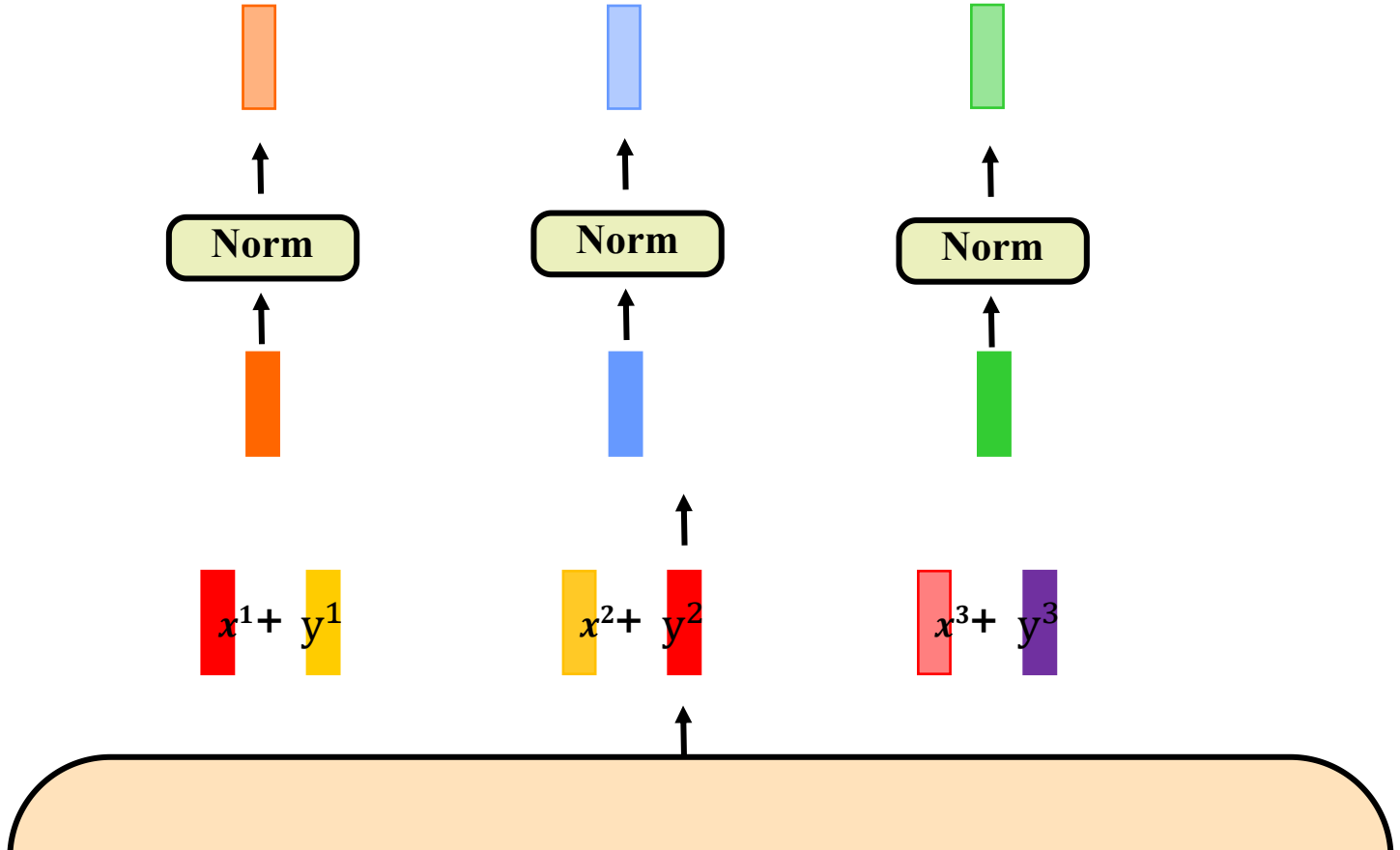
$$\sigma = \sqrt{\frac{(1-3)^2 + (3-3)^2 + (5-3)^2}{3}} \approx 1.63$$



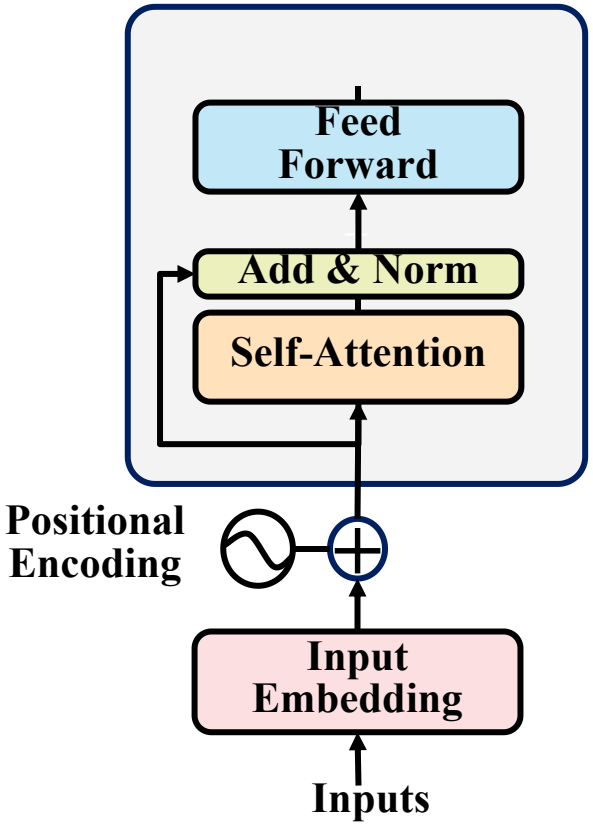
Transformer

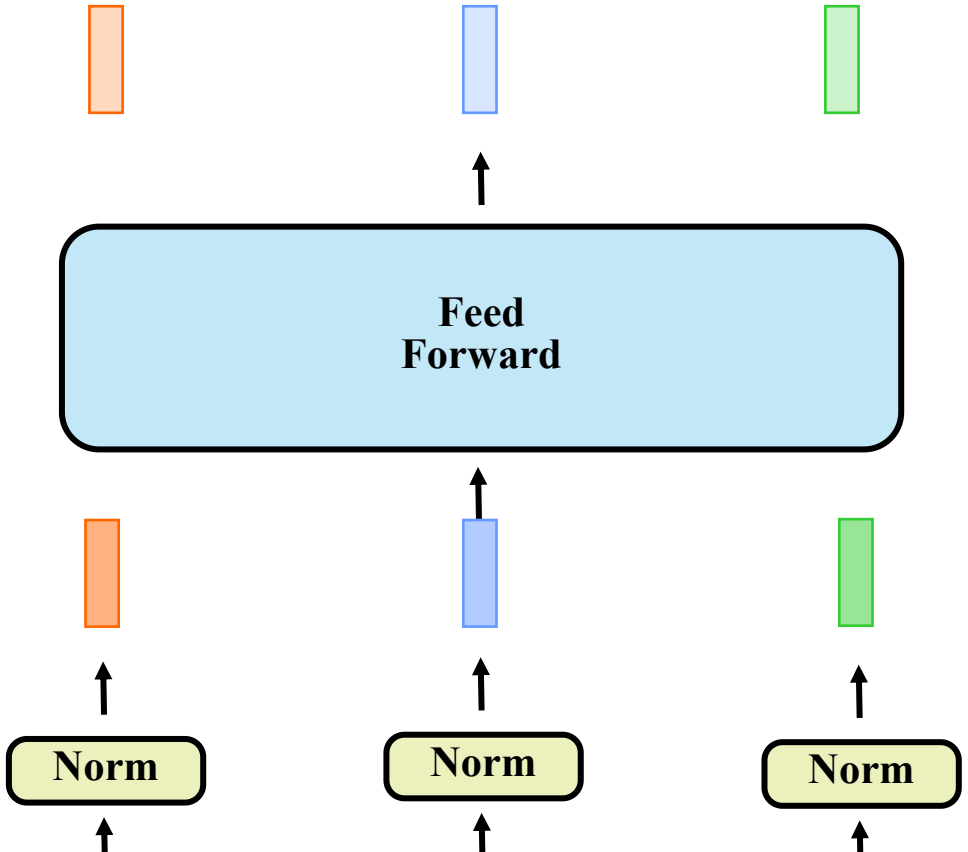
$$LayerNorm(x + Sublayer(x))$$



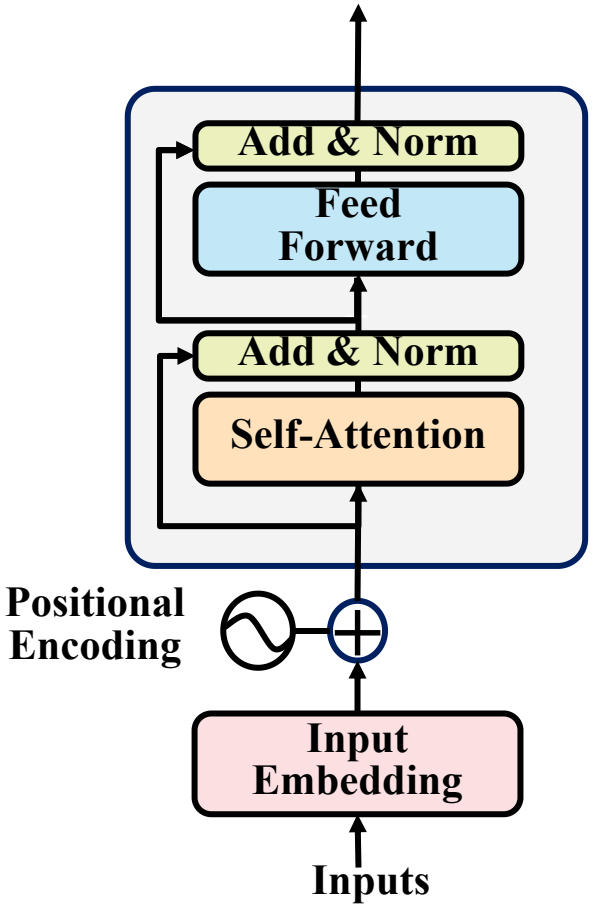


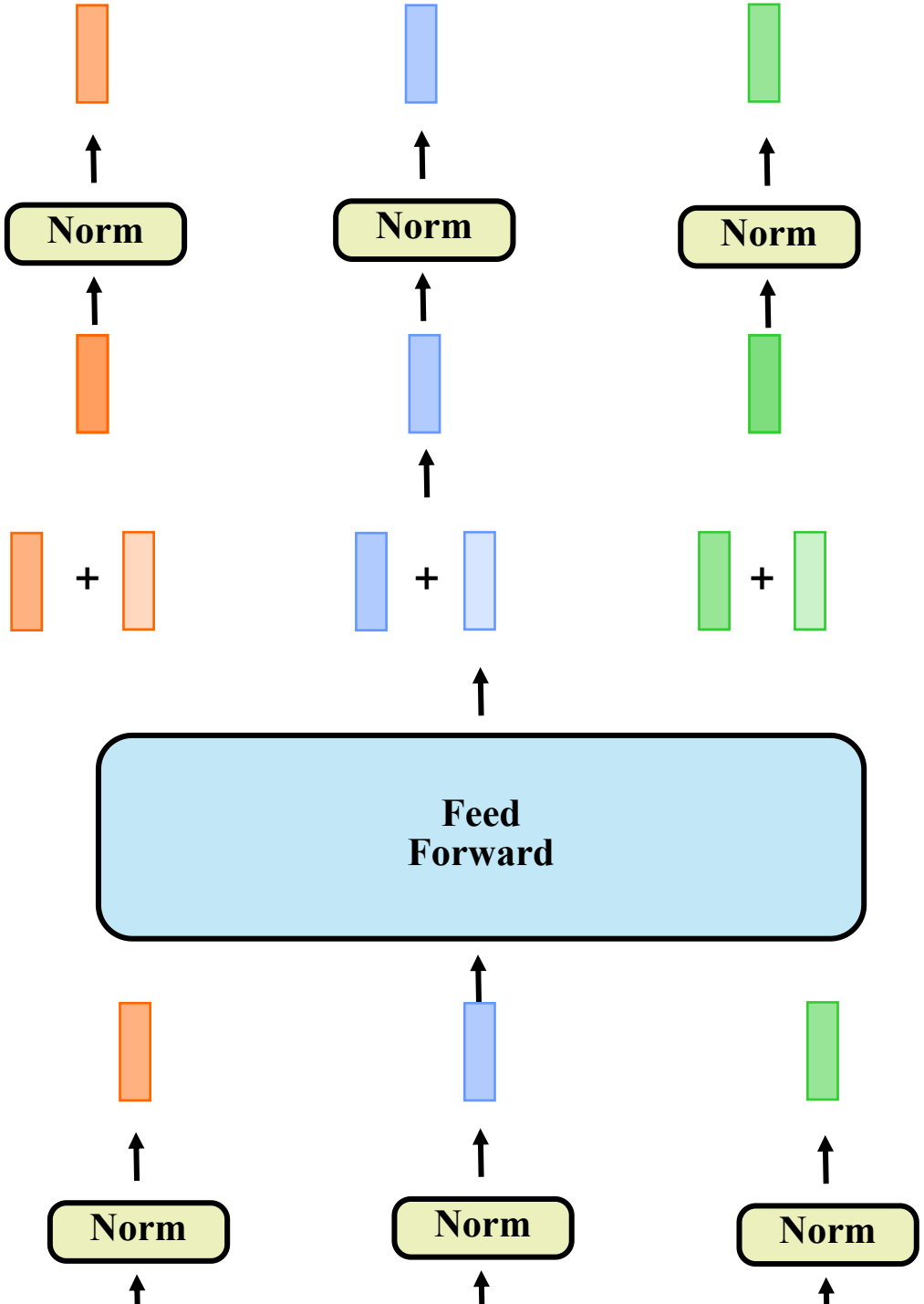
Transformer



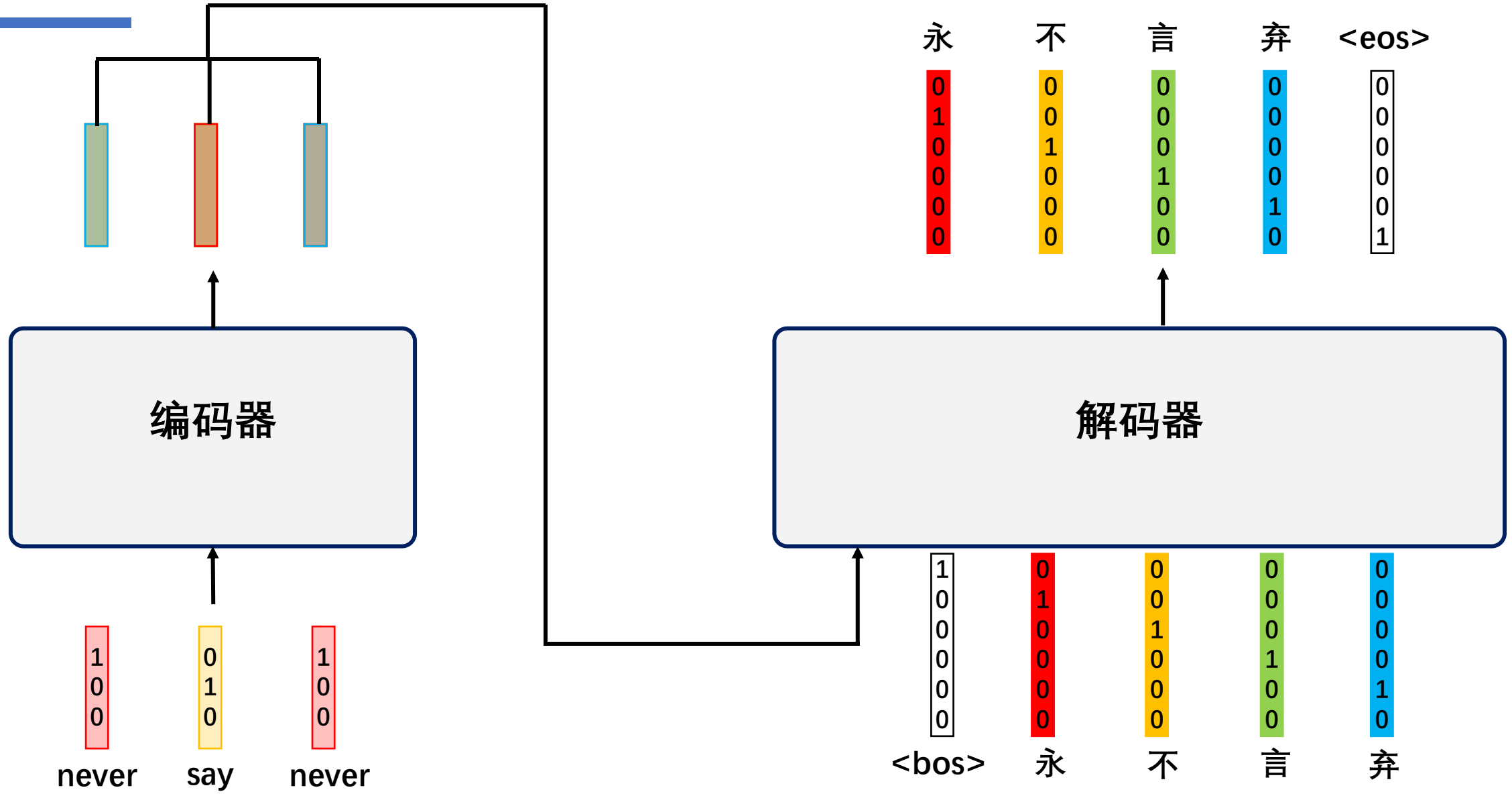


Transformer

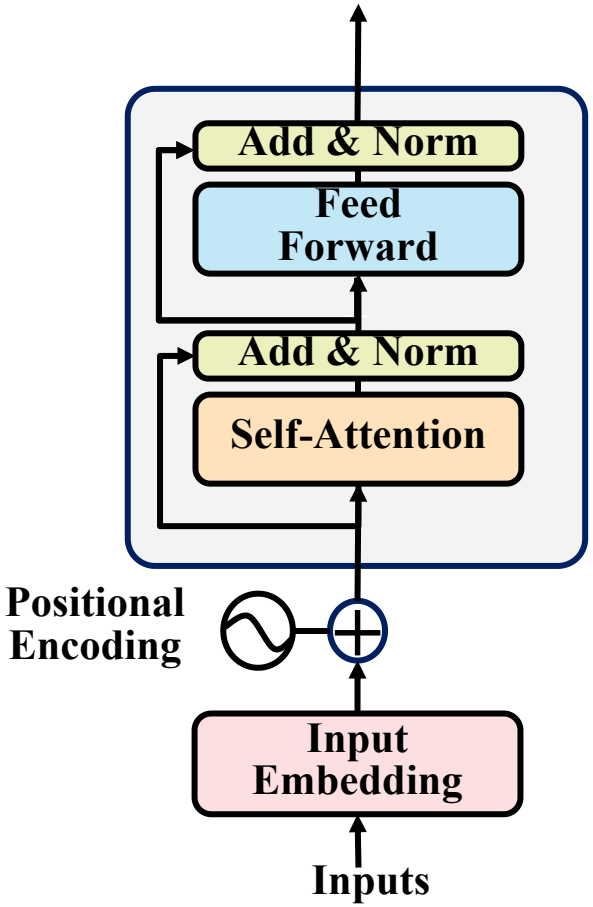


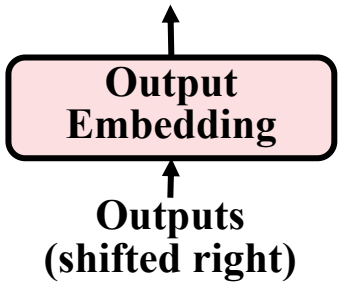


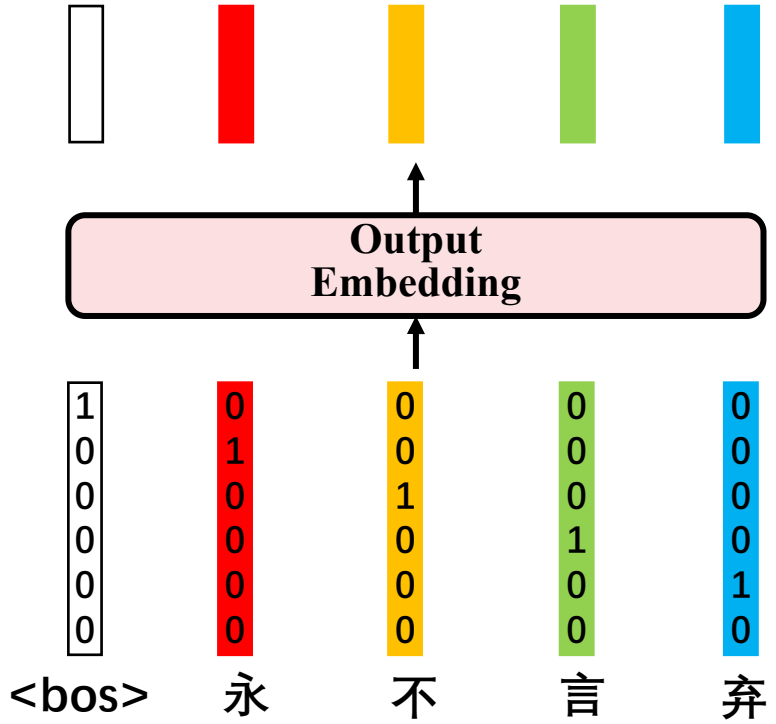
机器翻译



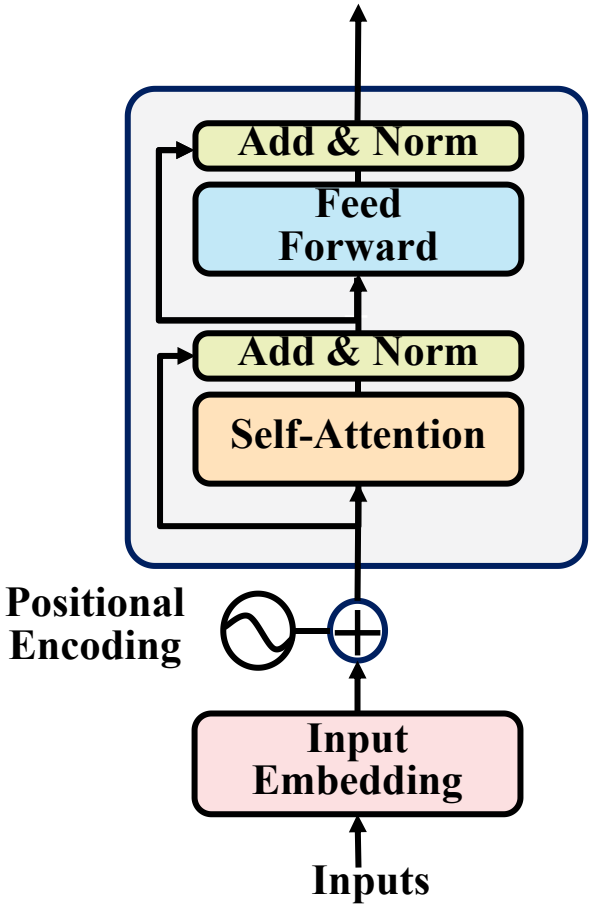
Transformer

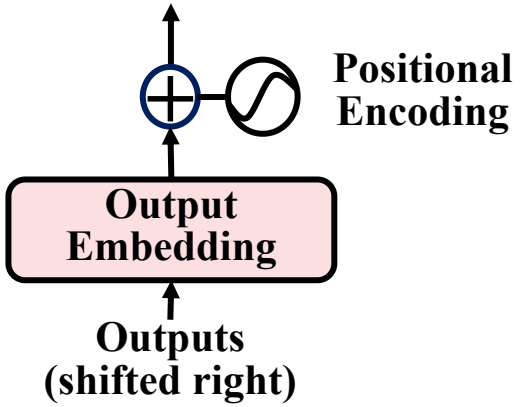


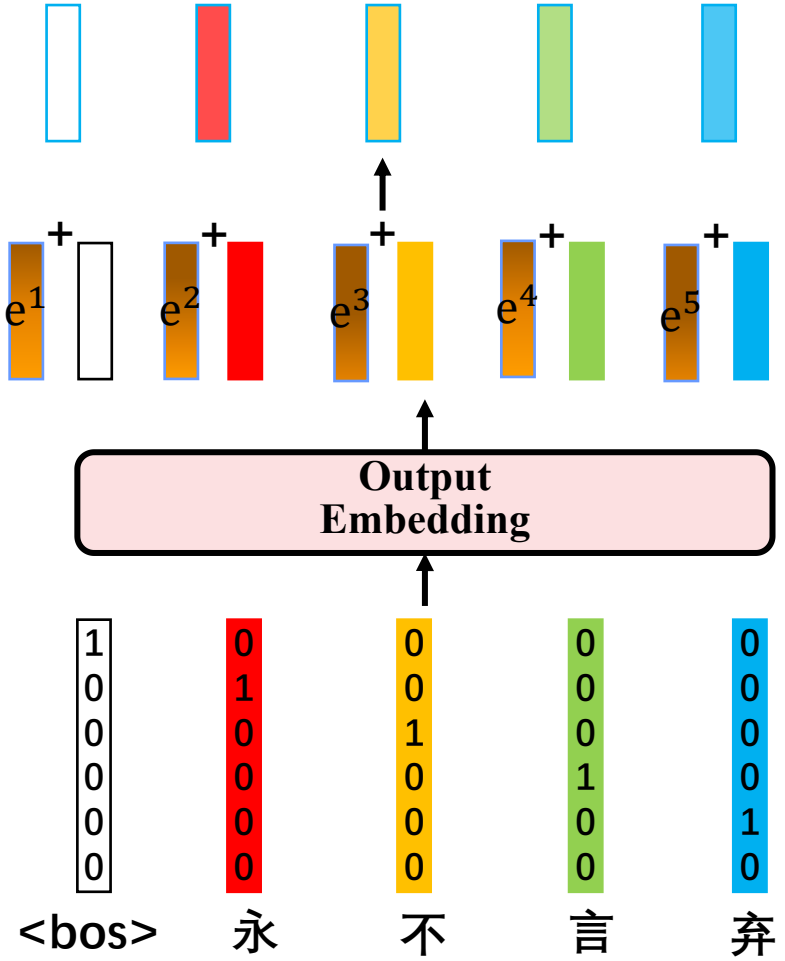




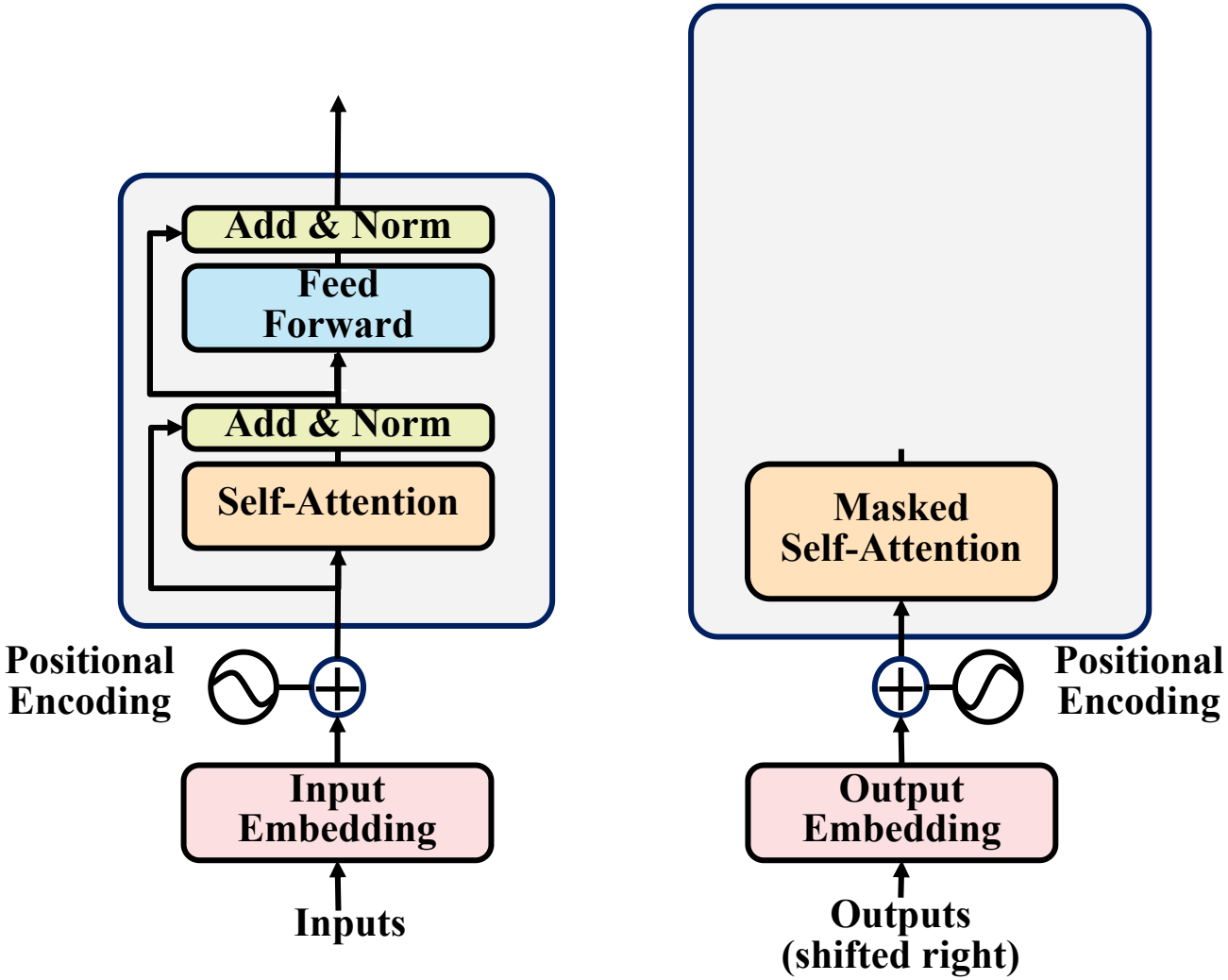
Transformer

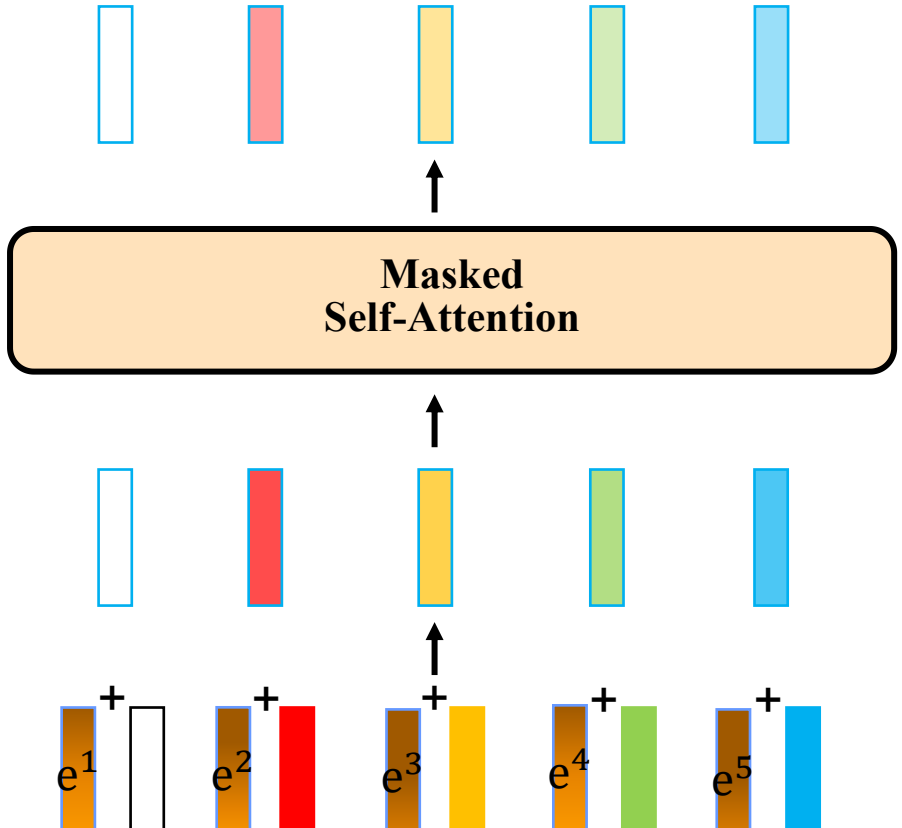




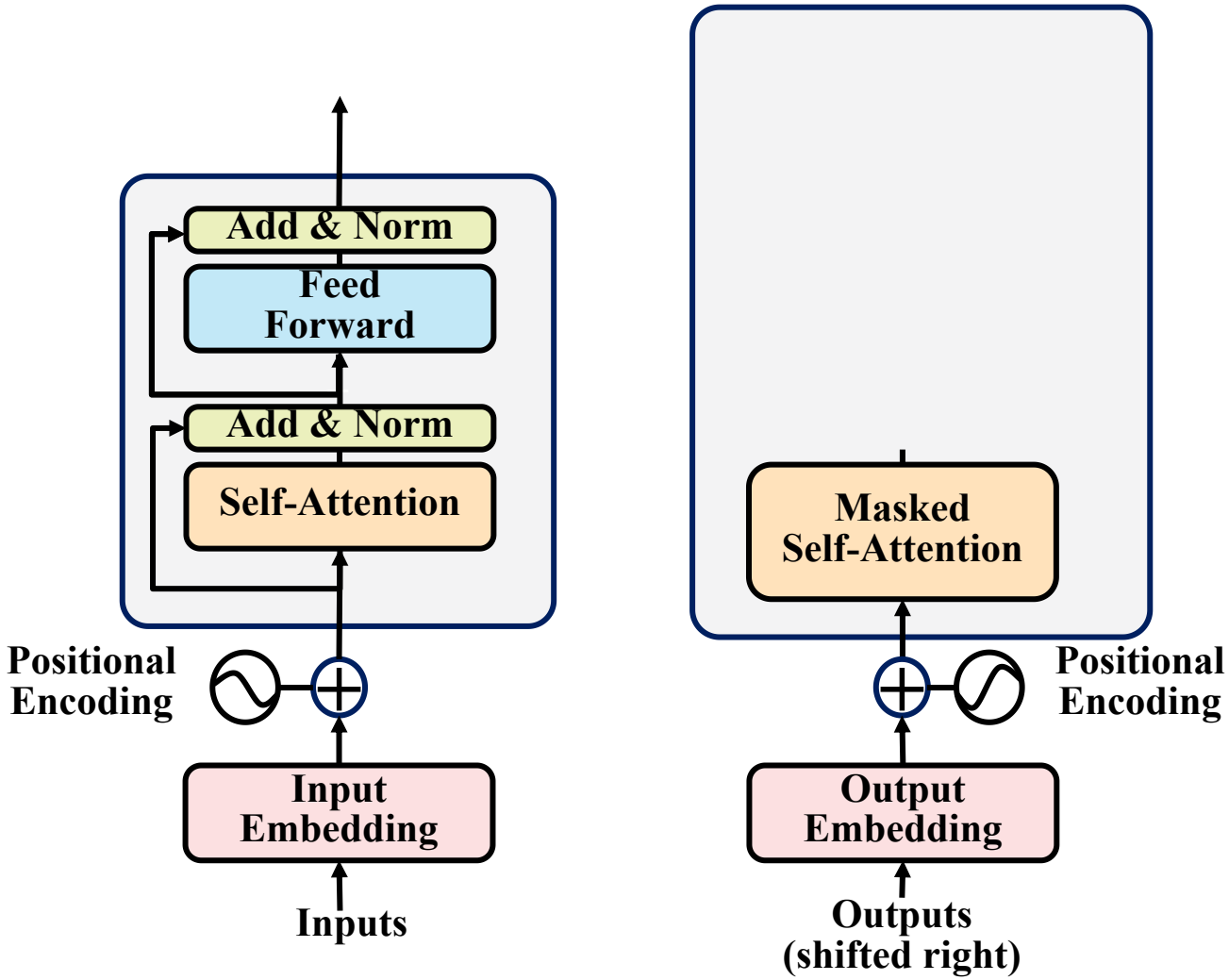


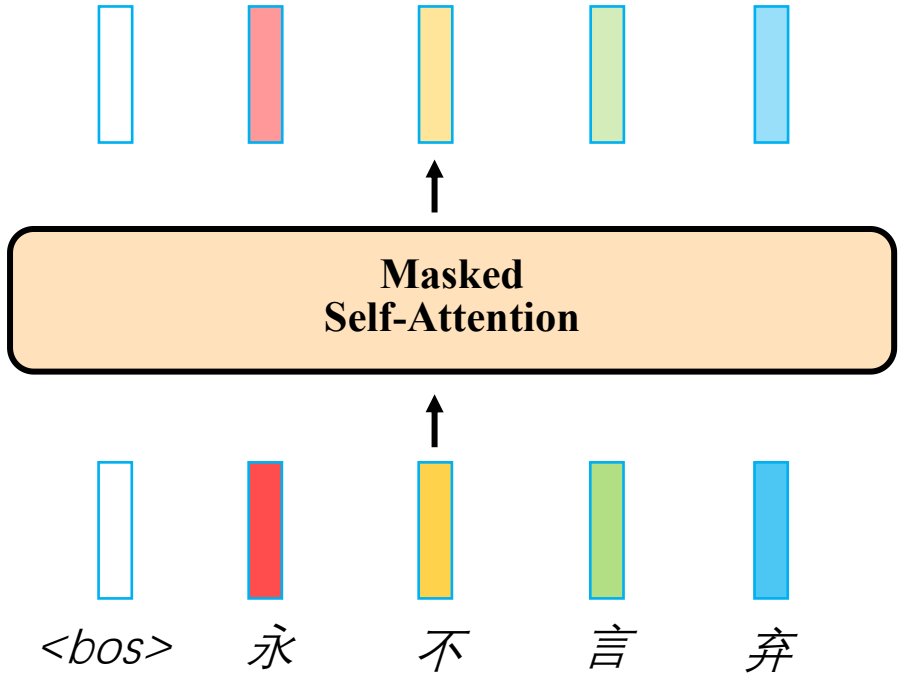
Transformer



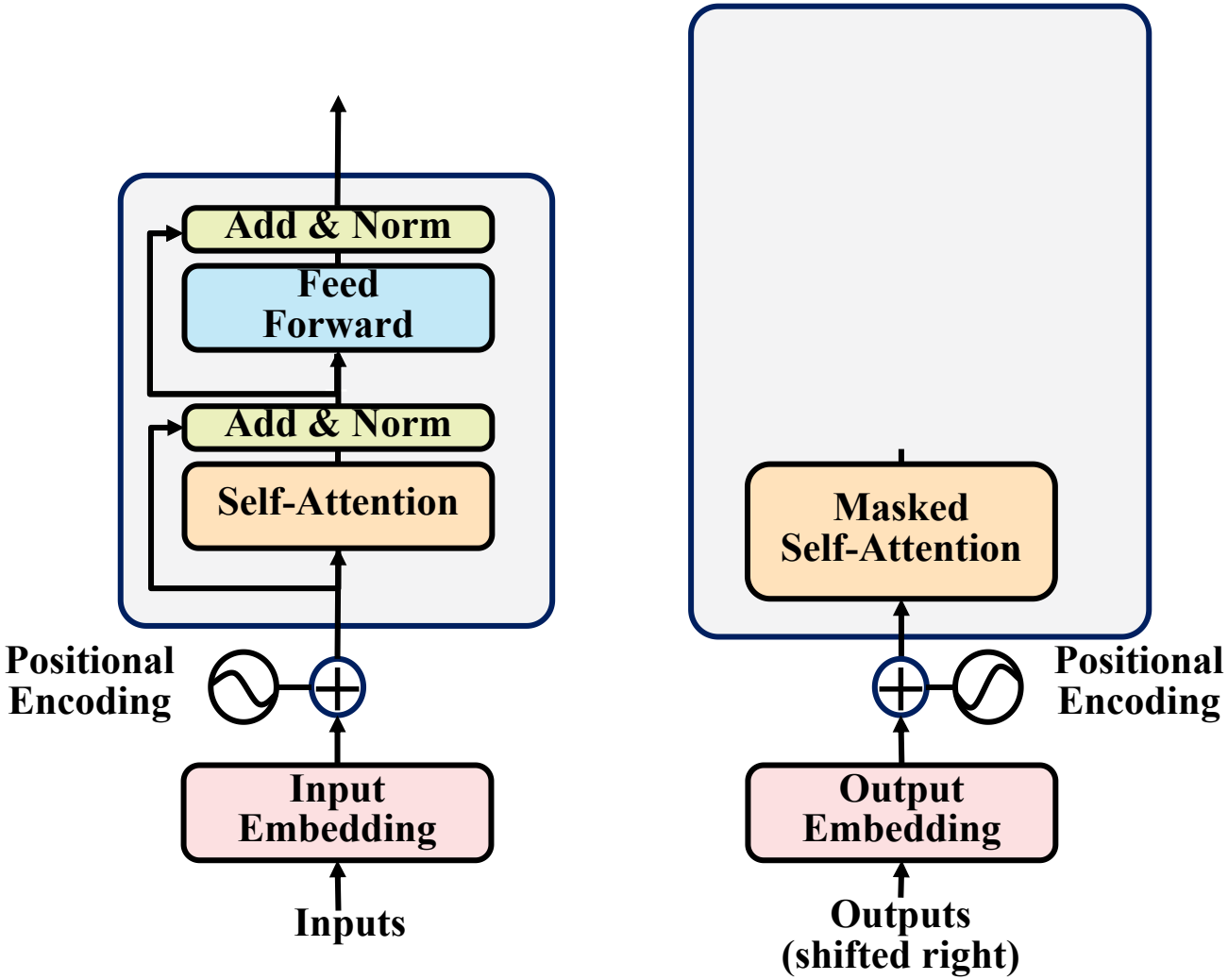


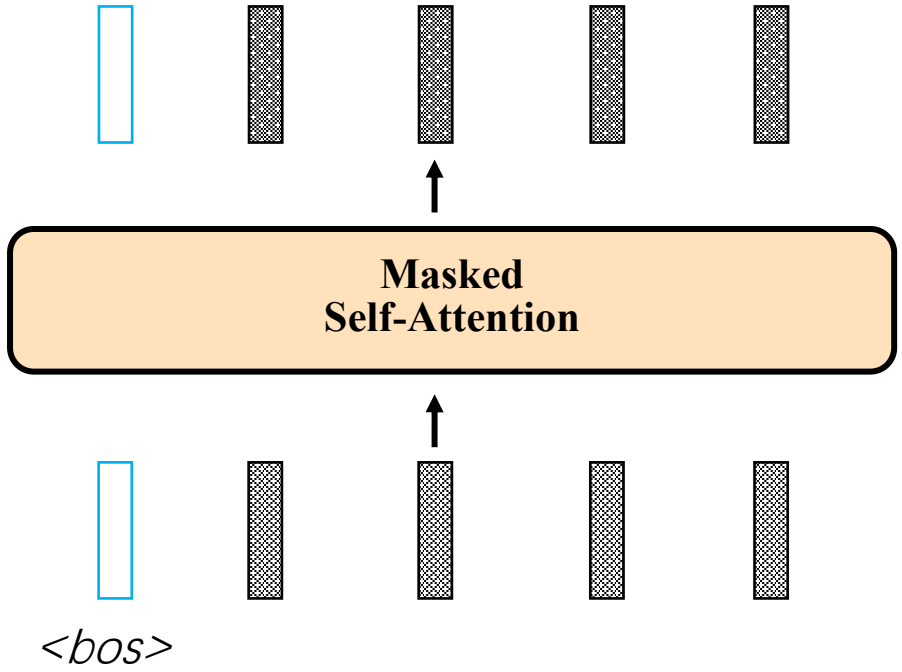
Transformer



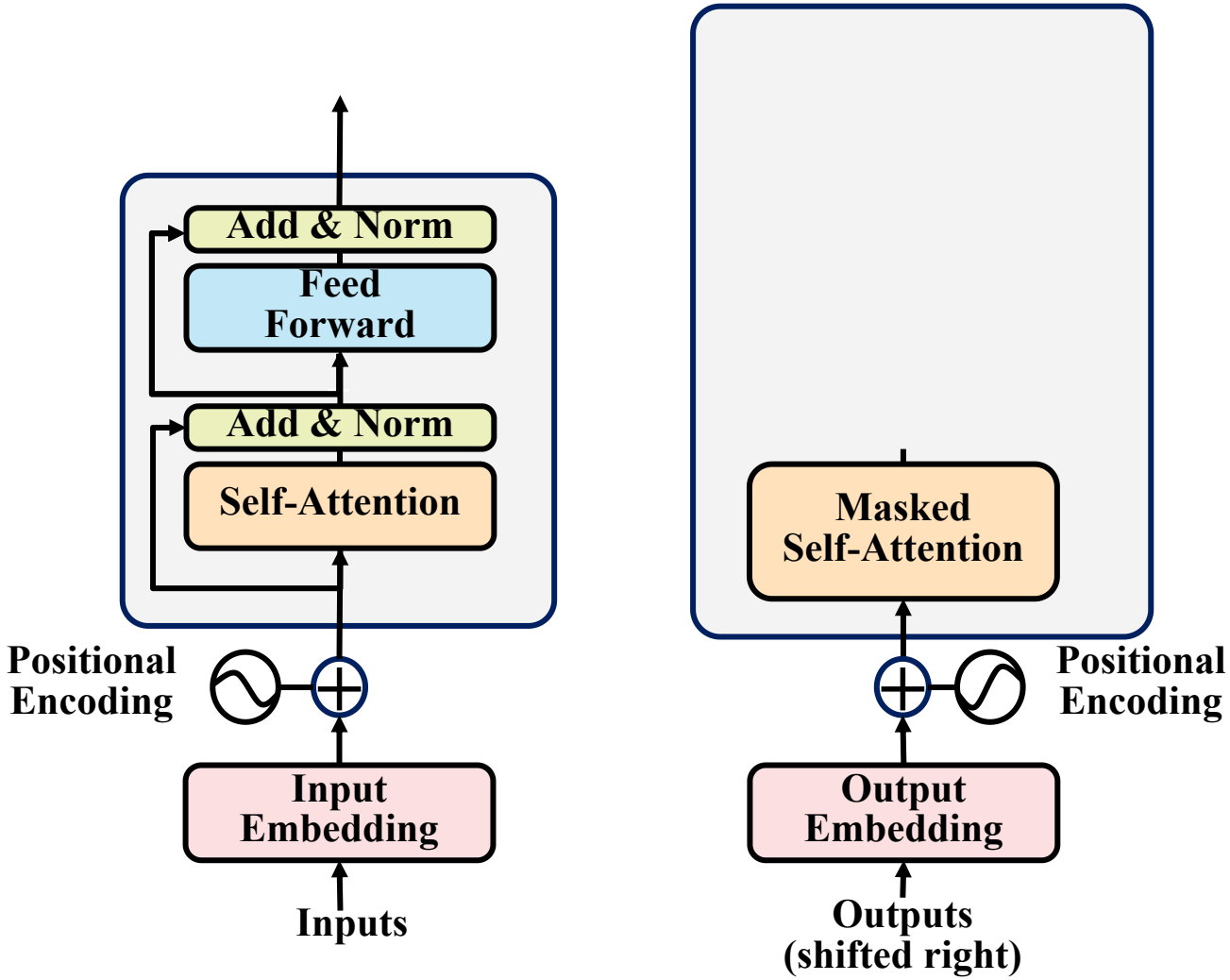


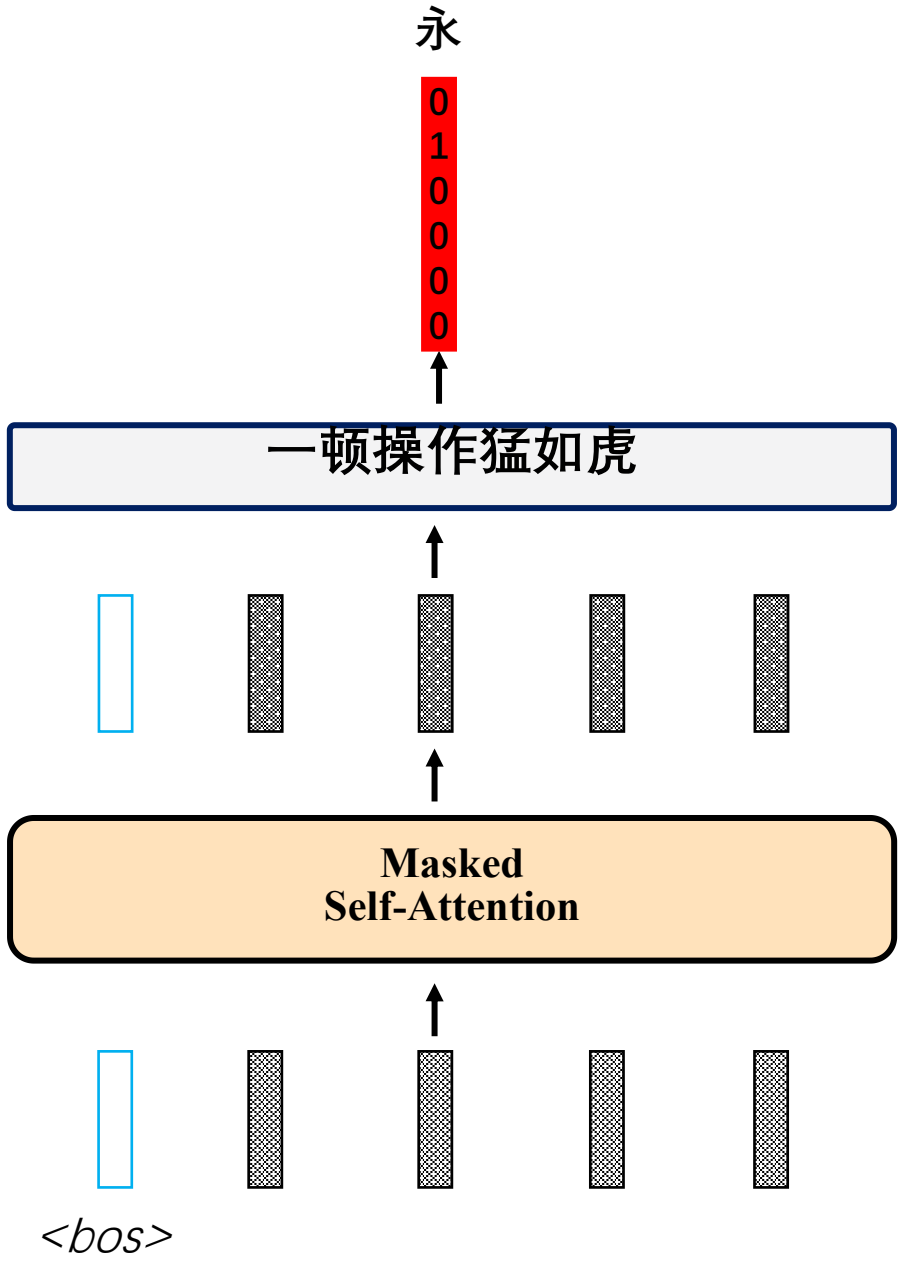
Transformer



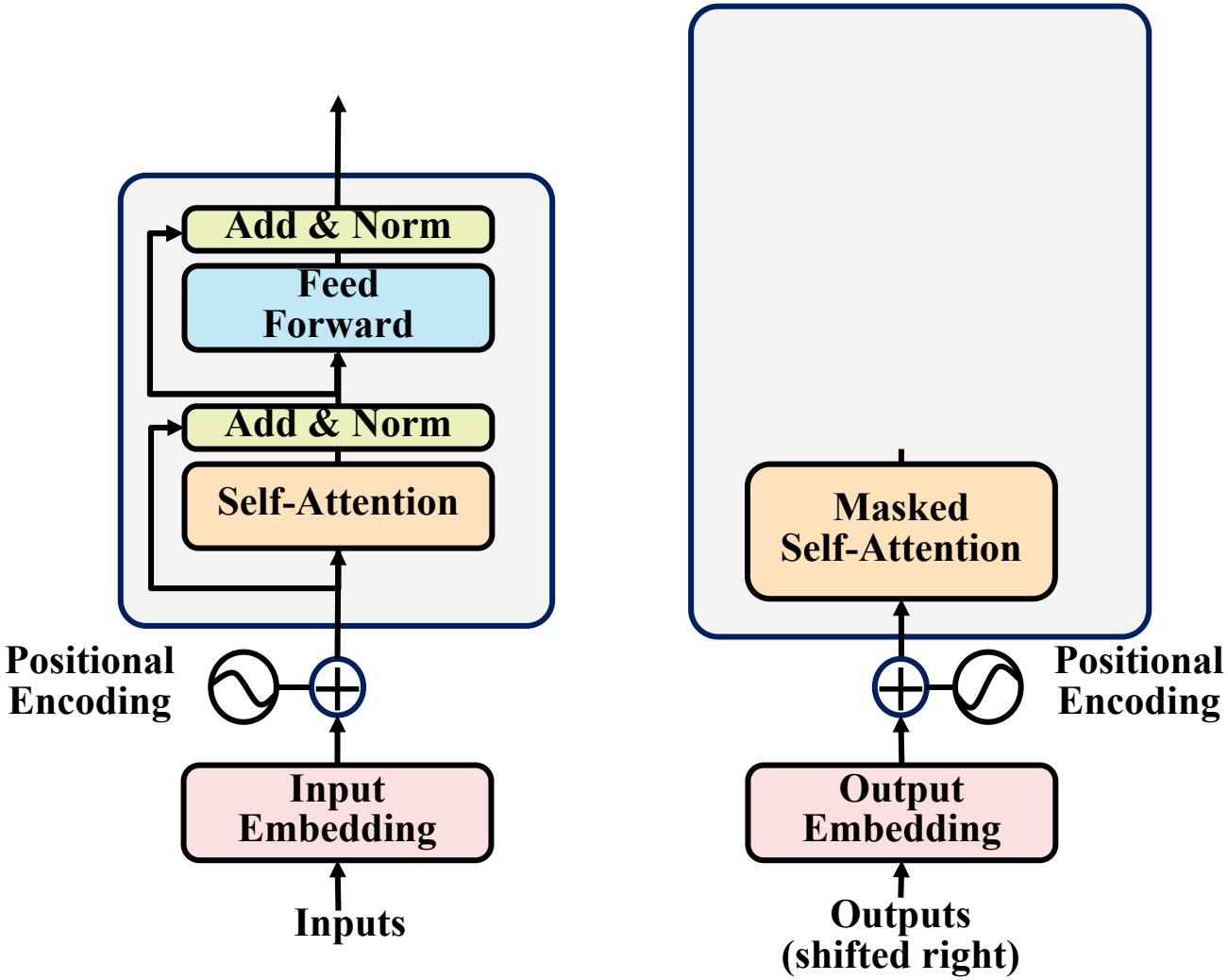


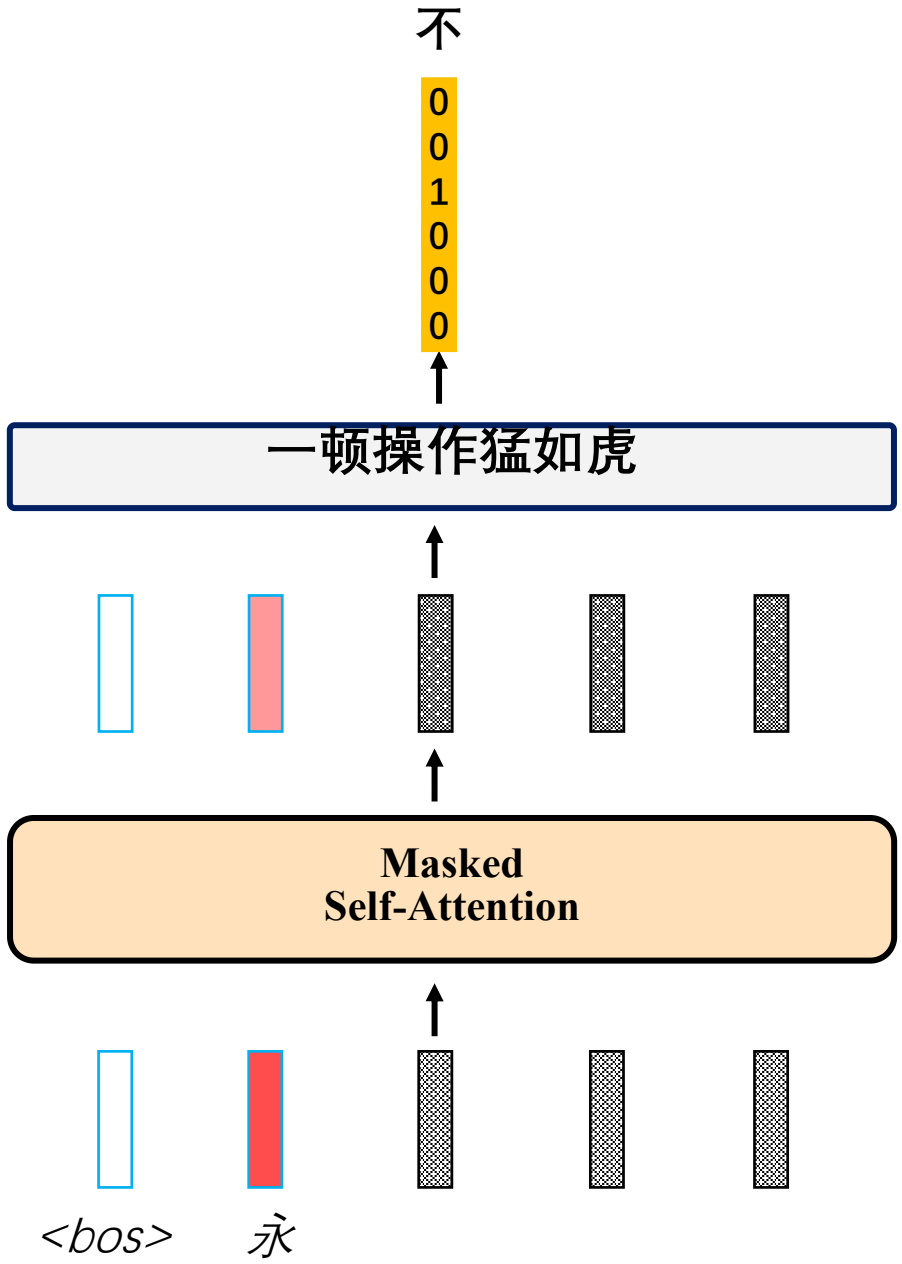
Transformer



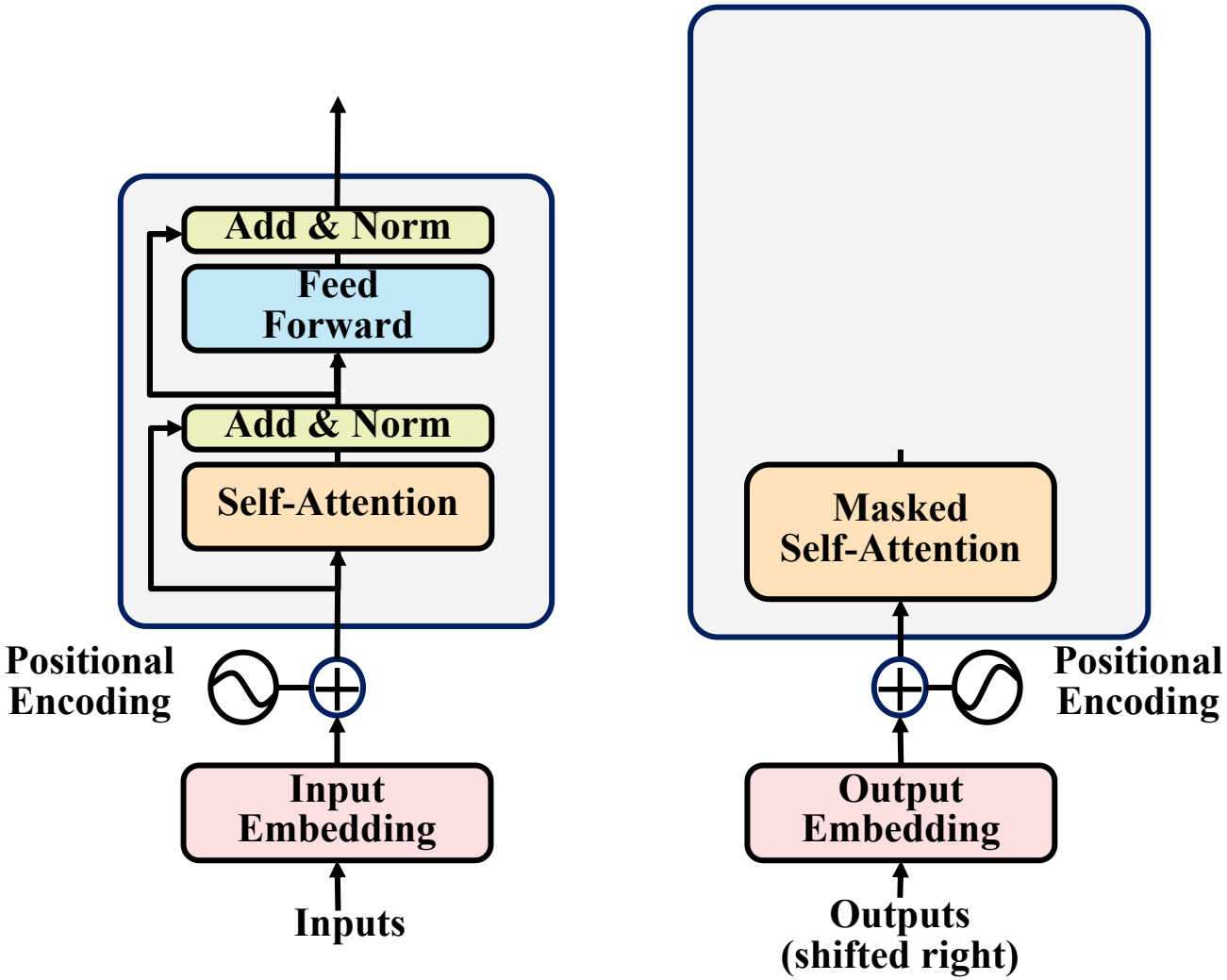


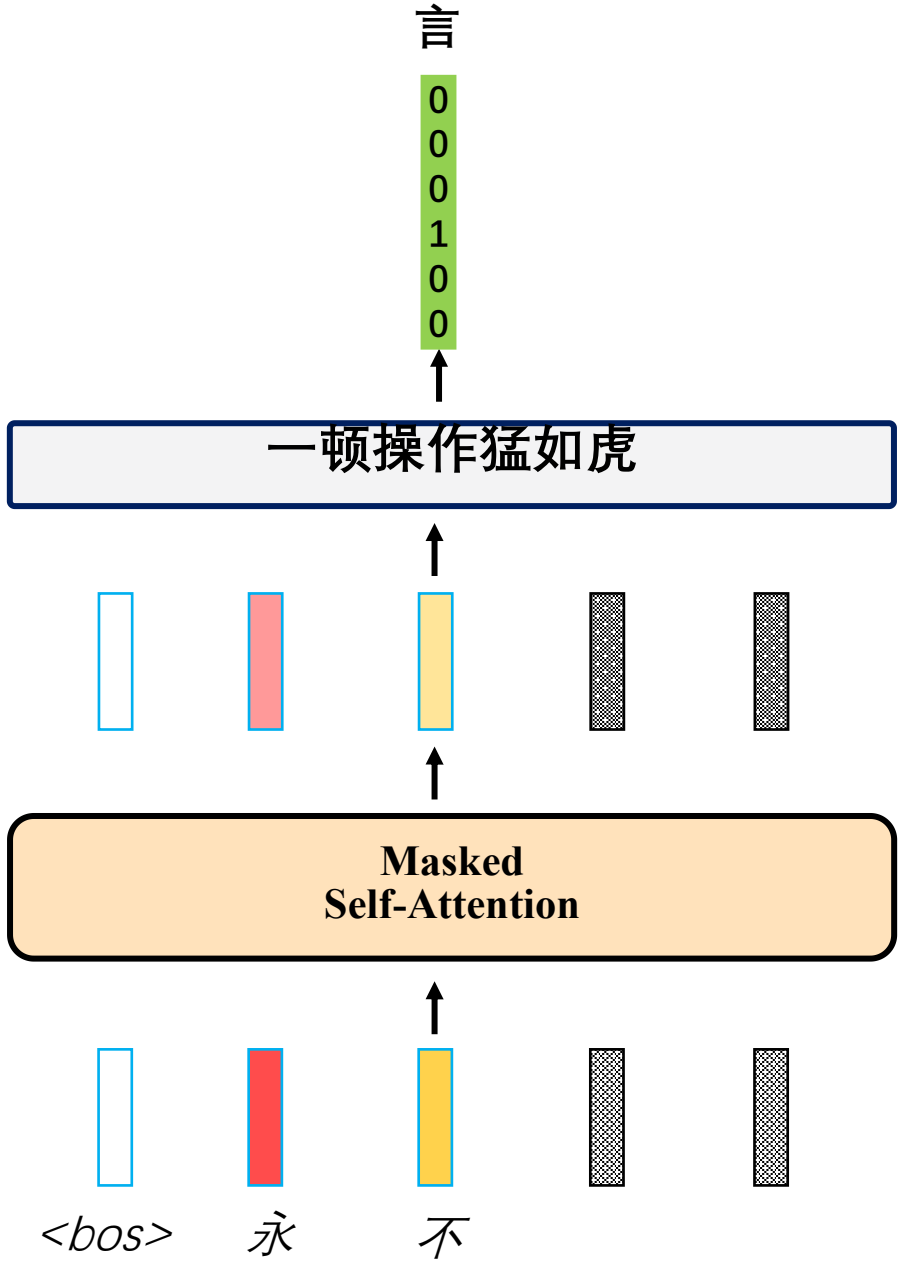
Transformer



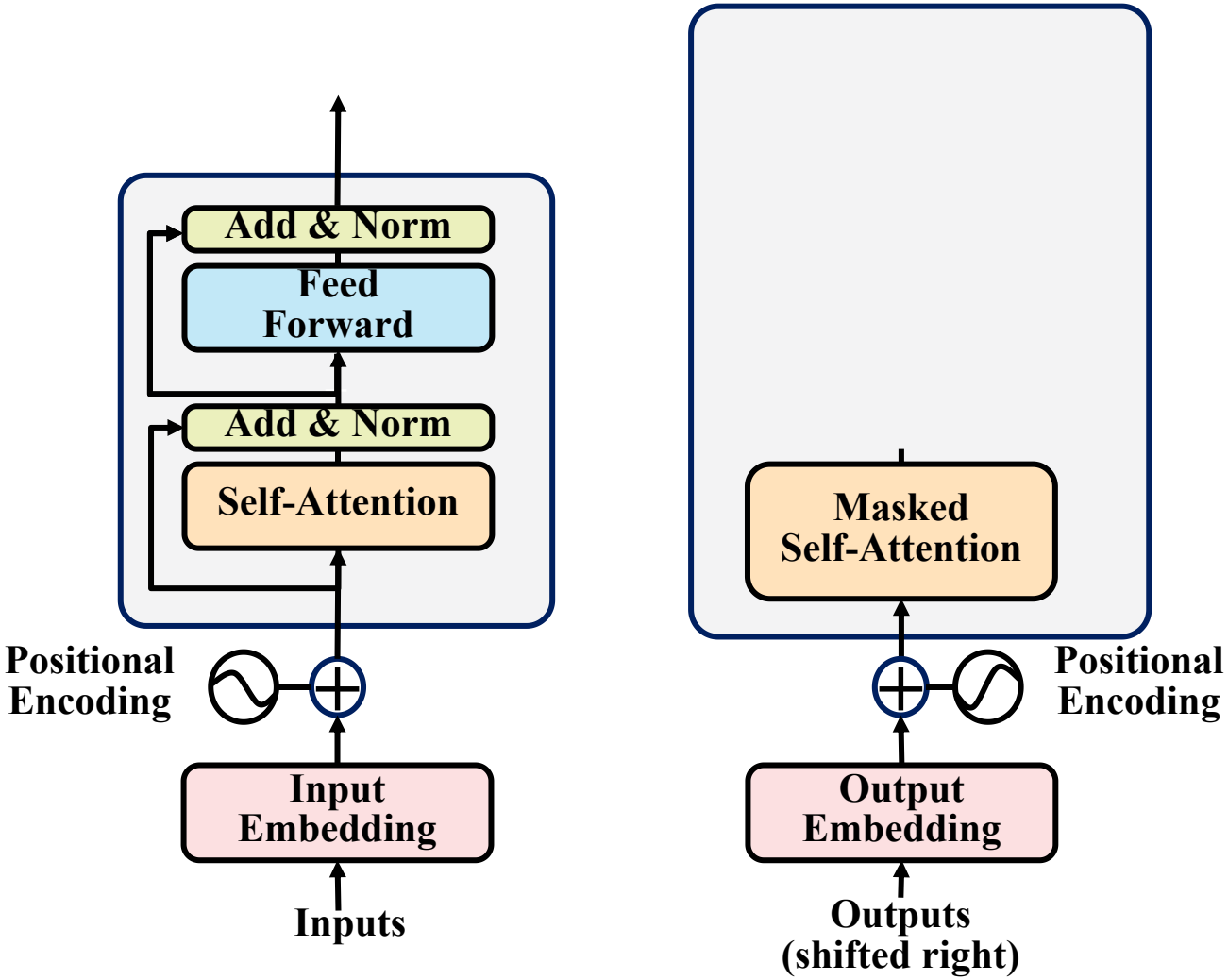


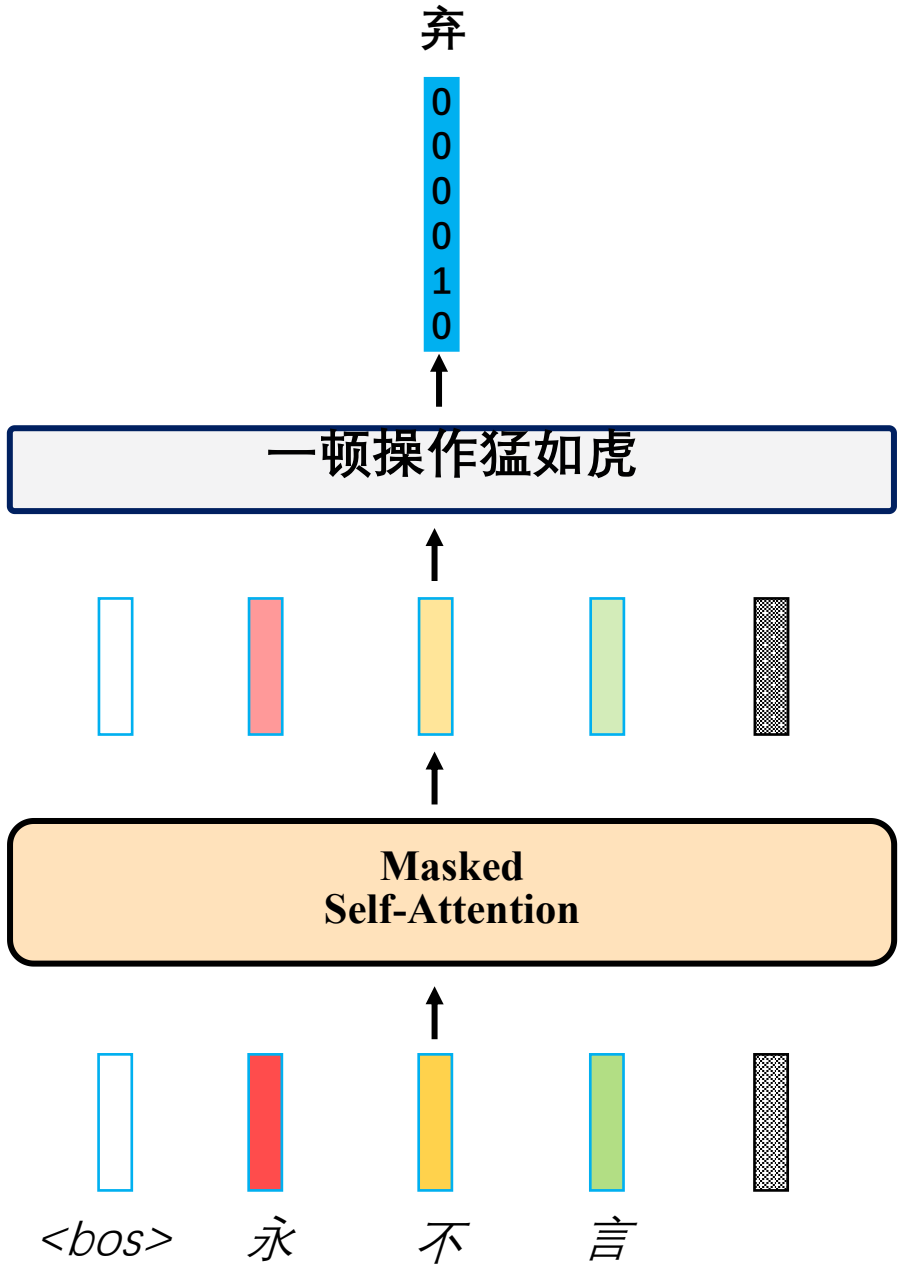
Transformer



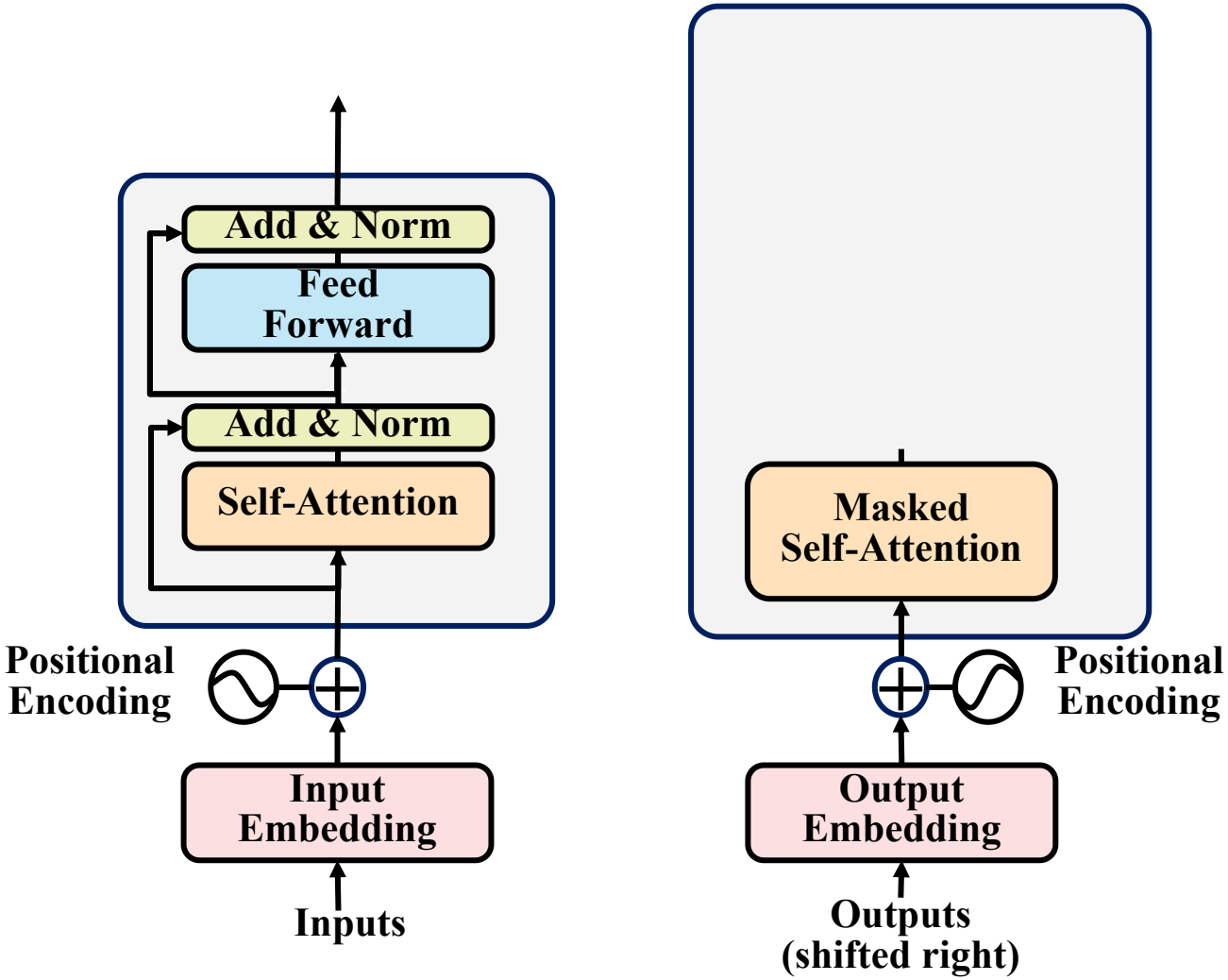


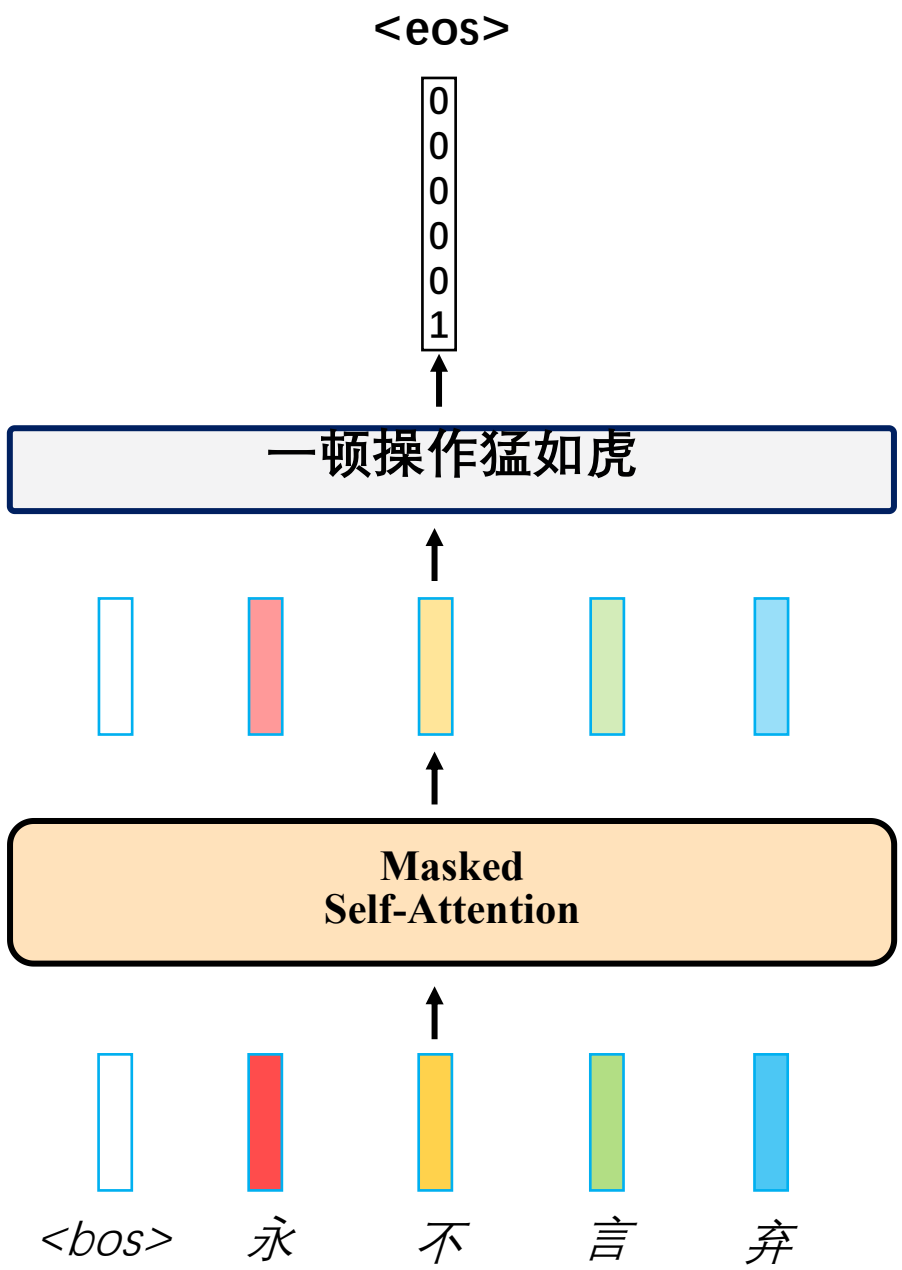
Transformer





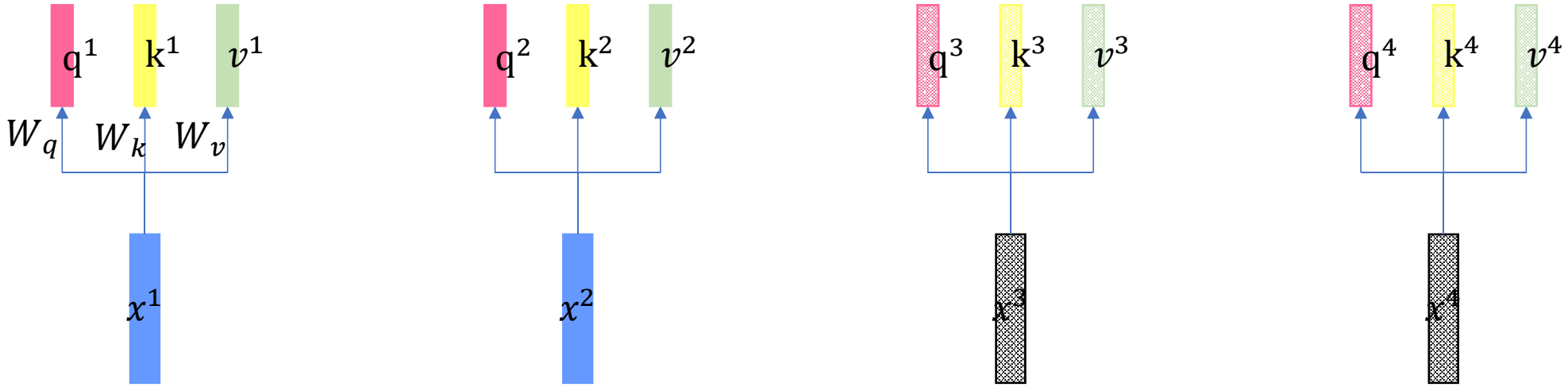
Transformer



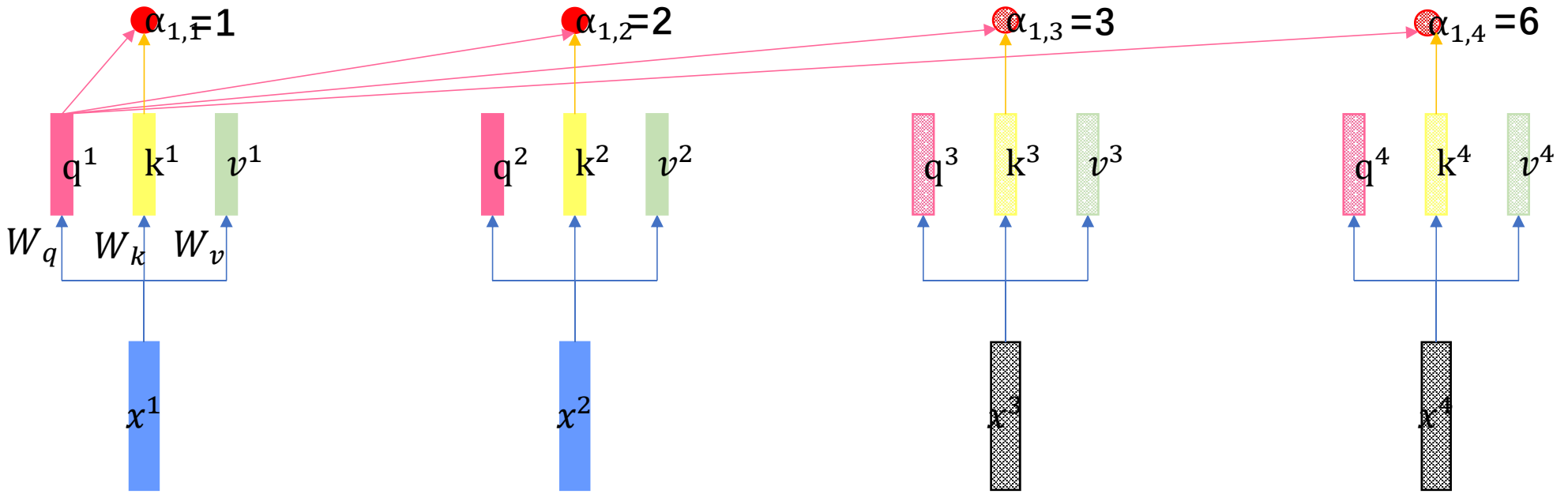


Masked-Attention

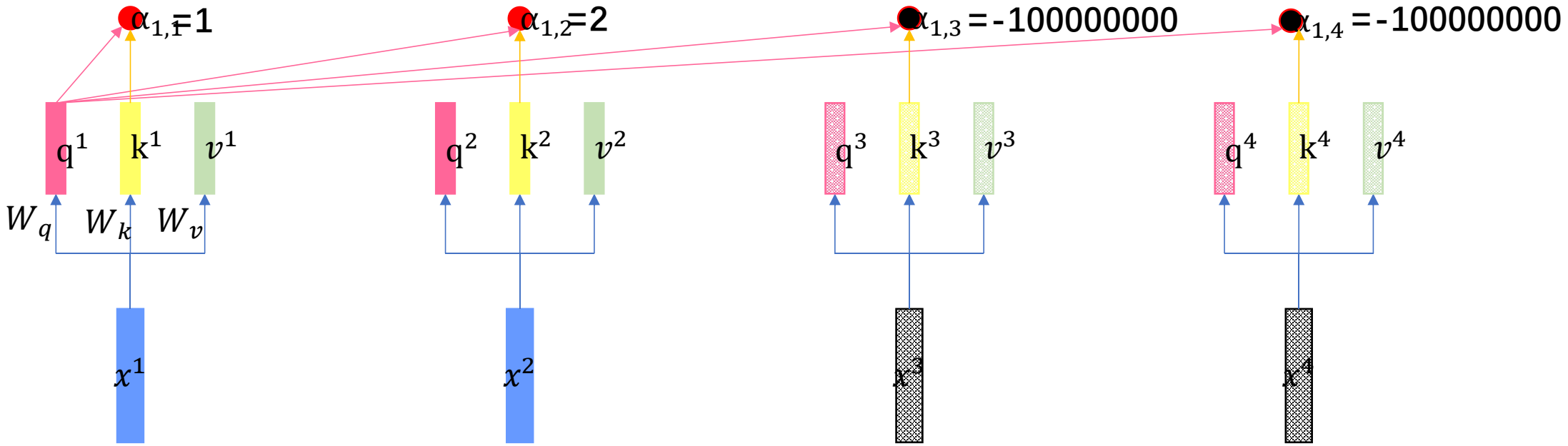
Masked-Attention



Masked-Attention

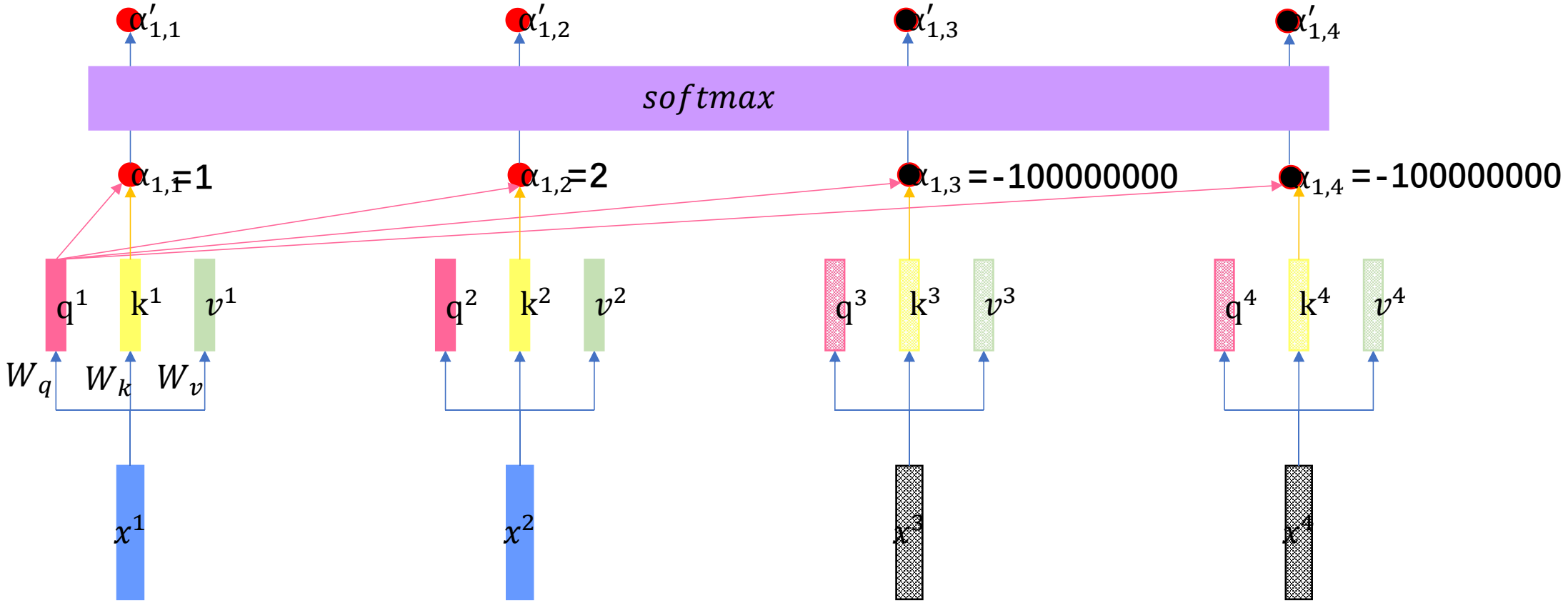


Masked-Attention

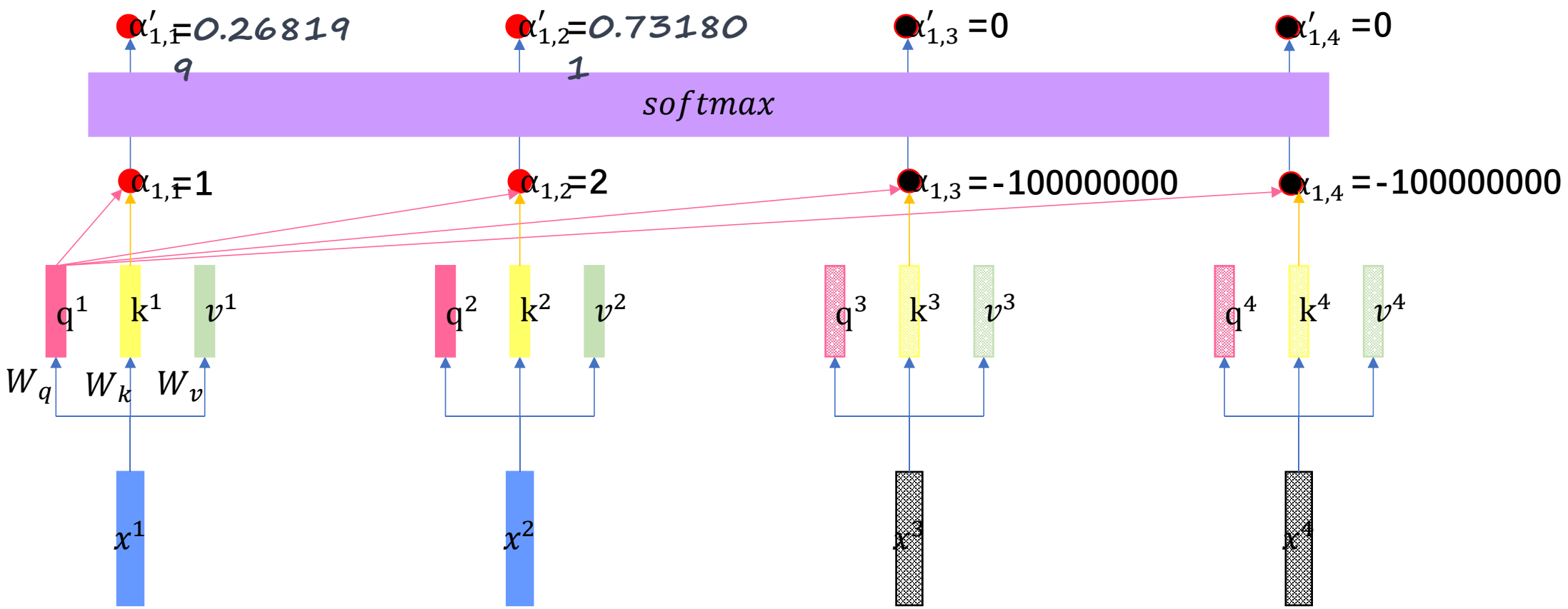


Masked-Attention

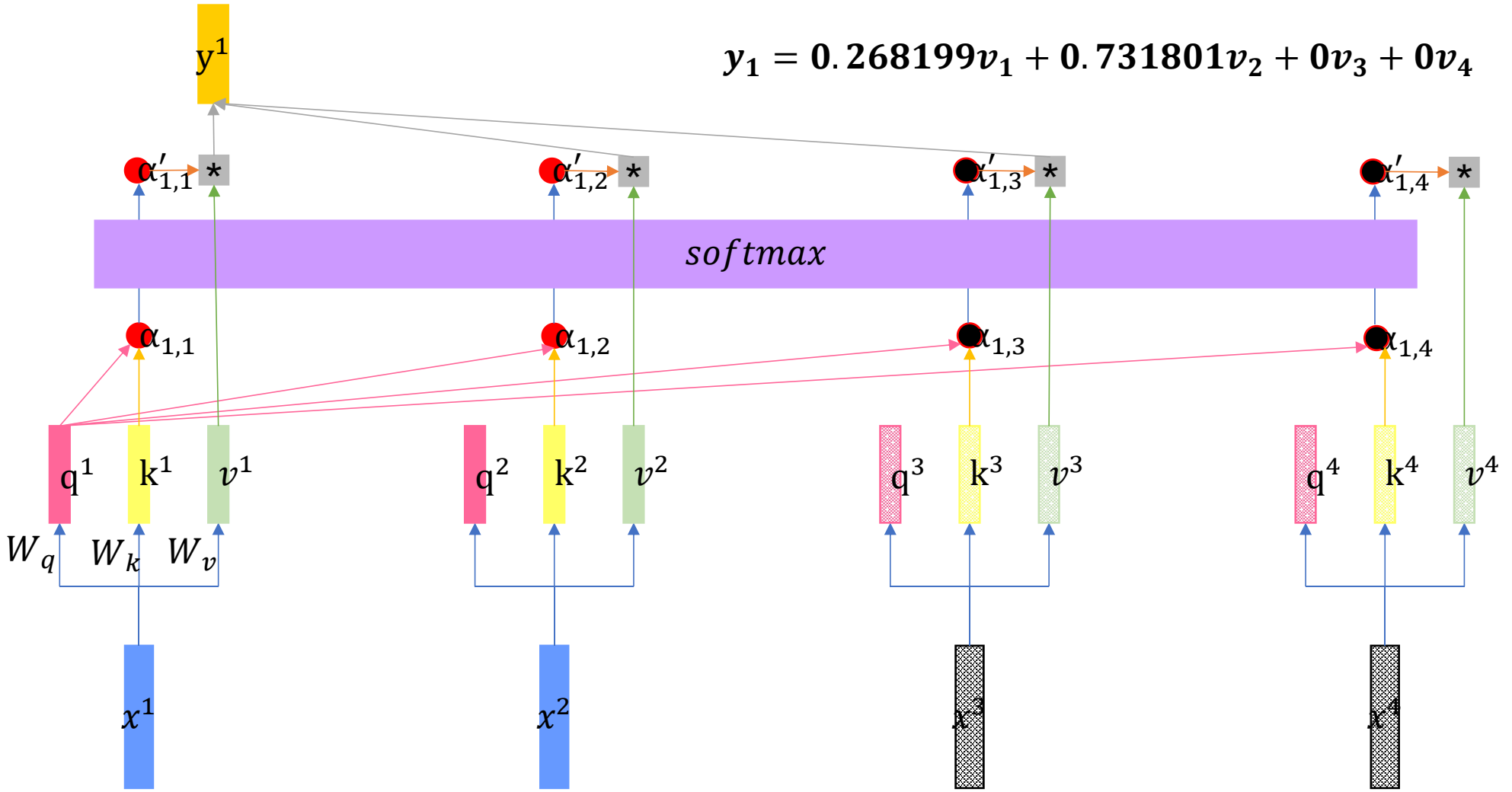
$$\alpha'_{1,1} = \frac{\exp(\alpha_{1,1})}{\exp(\alpha_{1,1})+\exp(\alpha_{1,2})+\exp(\alpha_{1,3})+\exp(\alpha_{1,4})} = \frac{2.71828}{2.71828+7.389056+0+0} = 0.268199$$



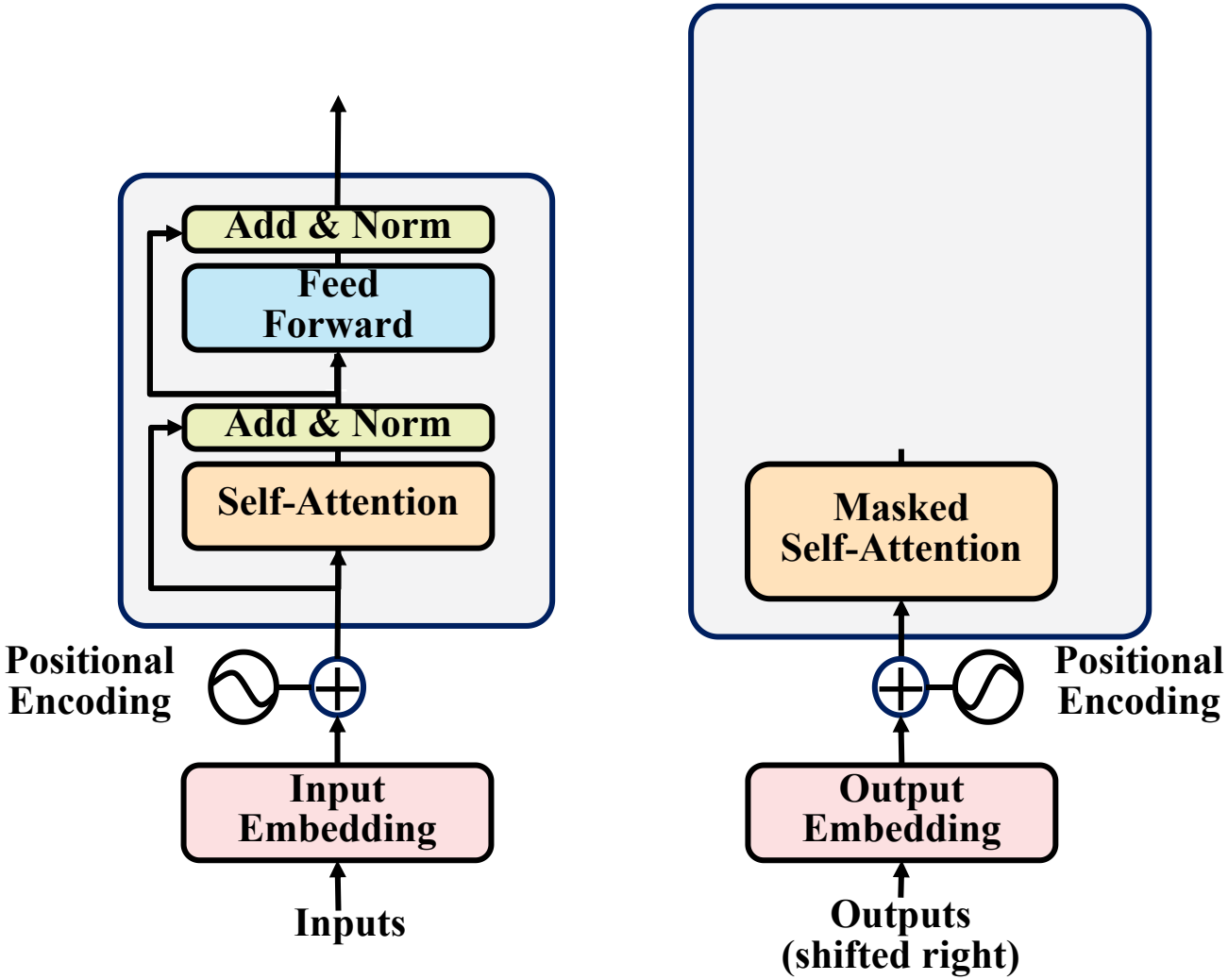
Masked-Attention

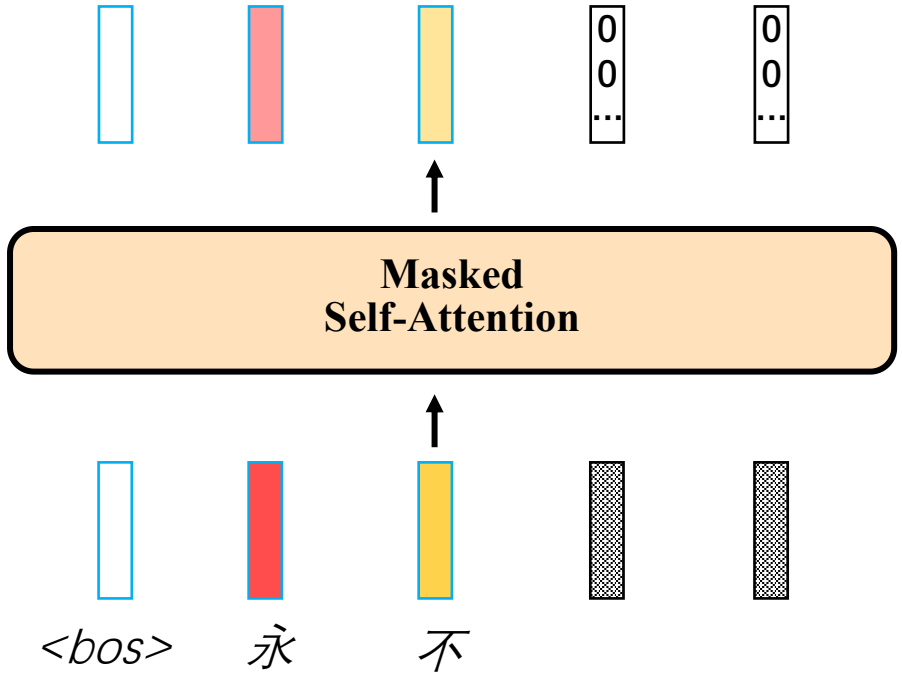


Masked-Attention

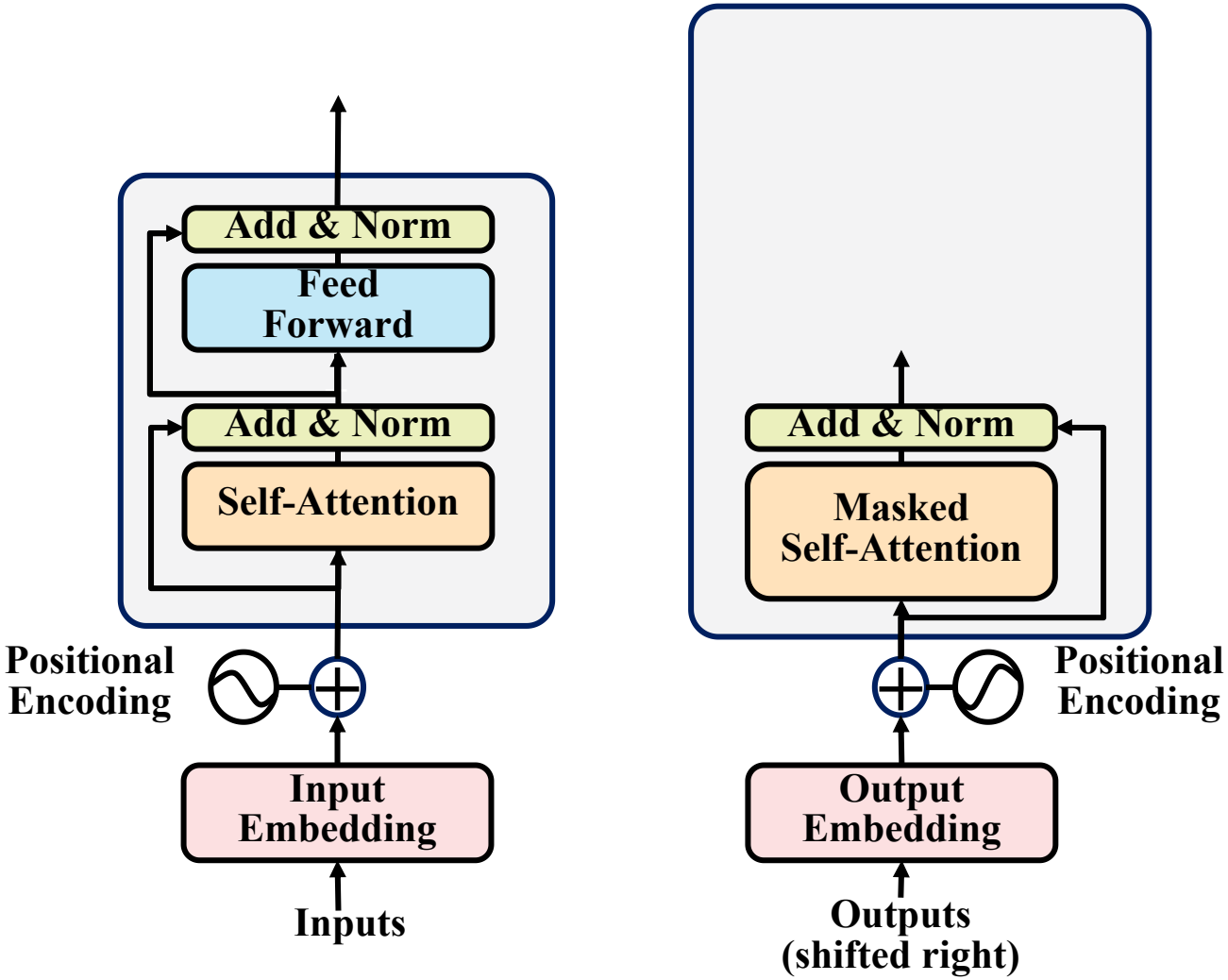


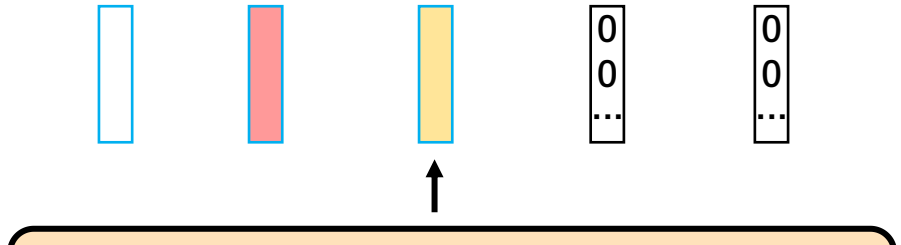
Transformer



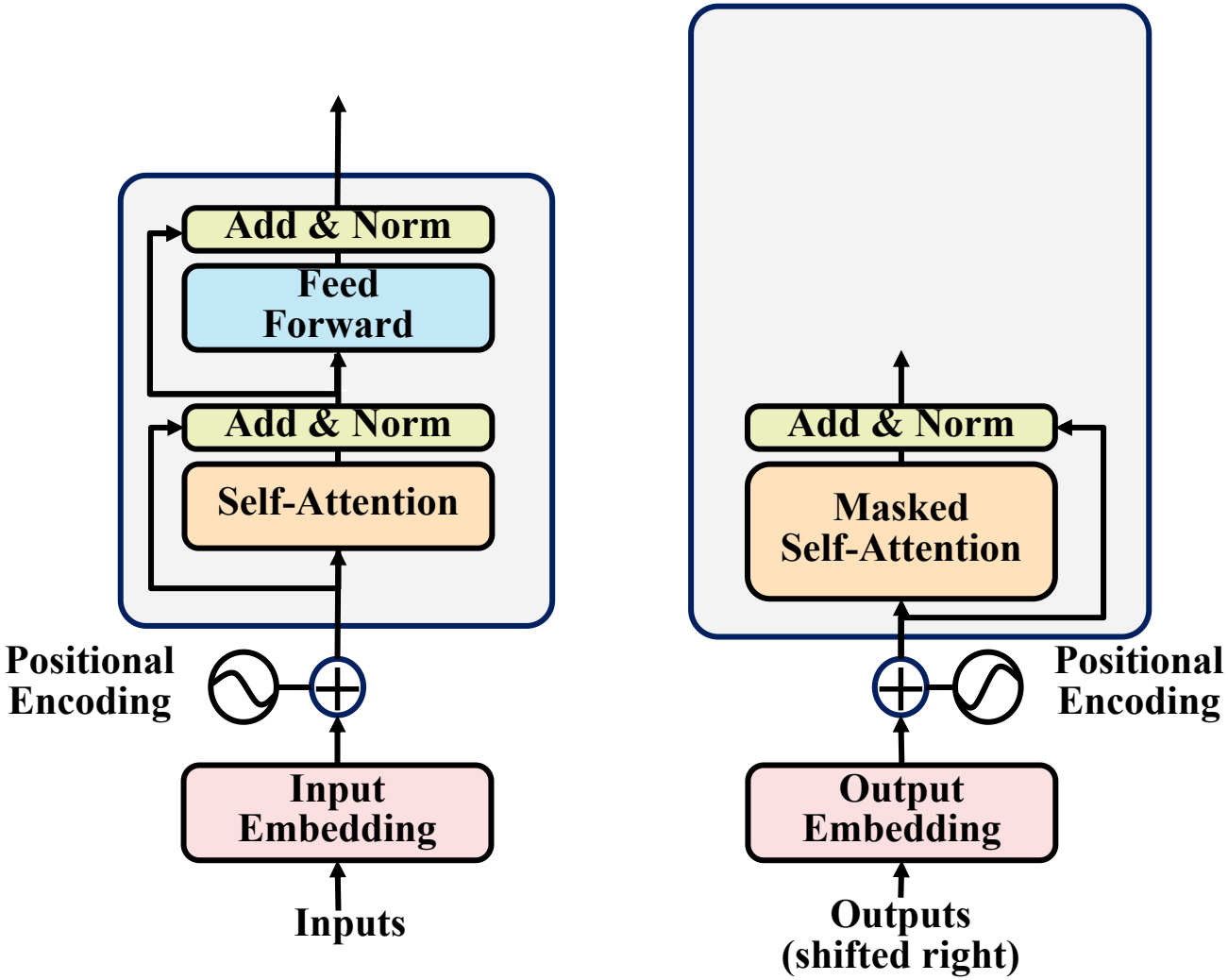


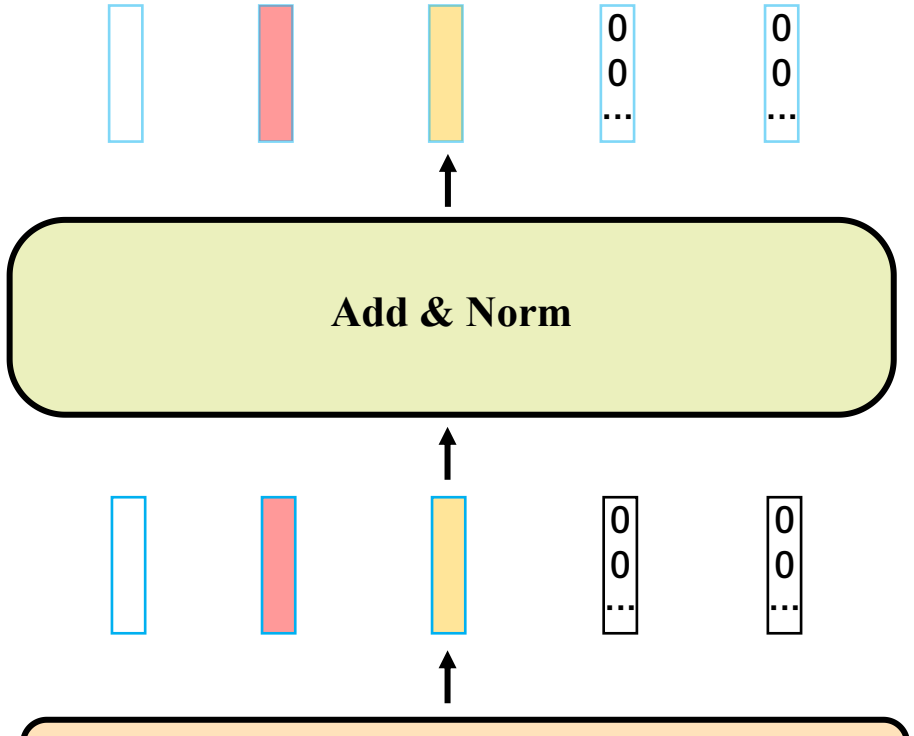
Transformer



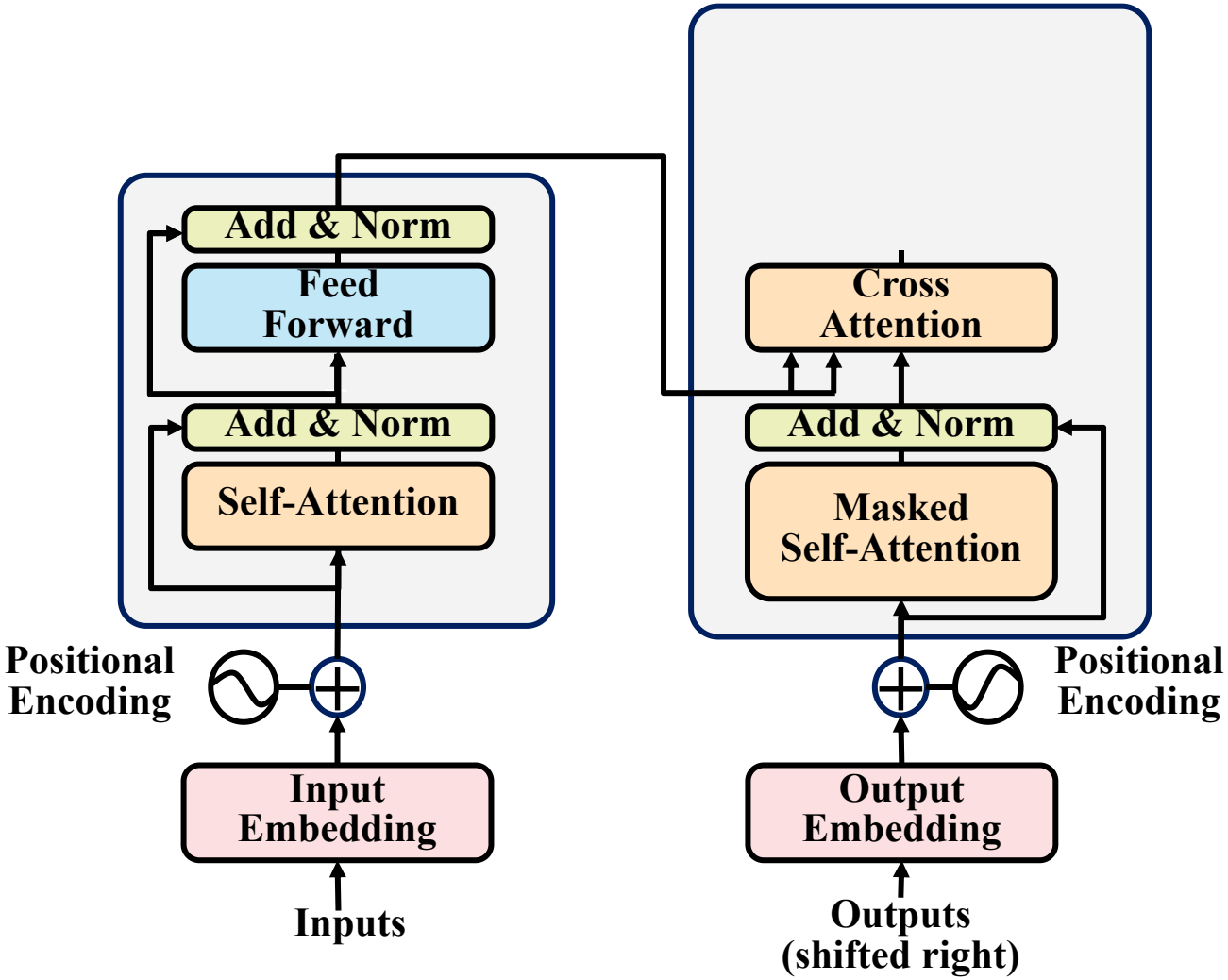


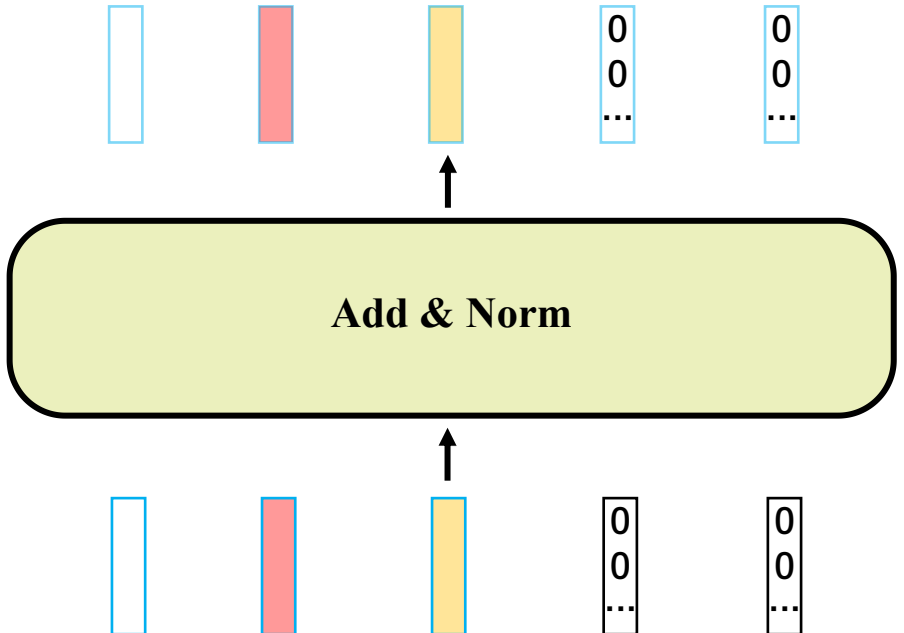
Transformer



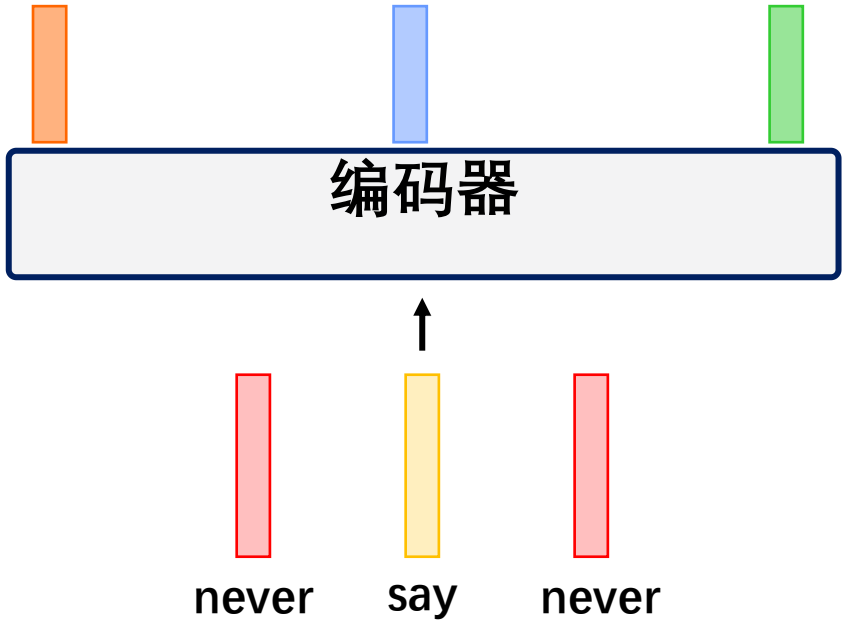


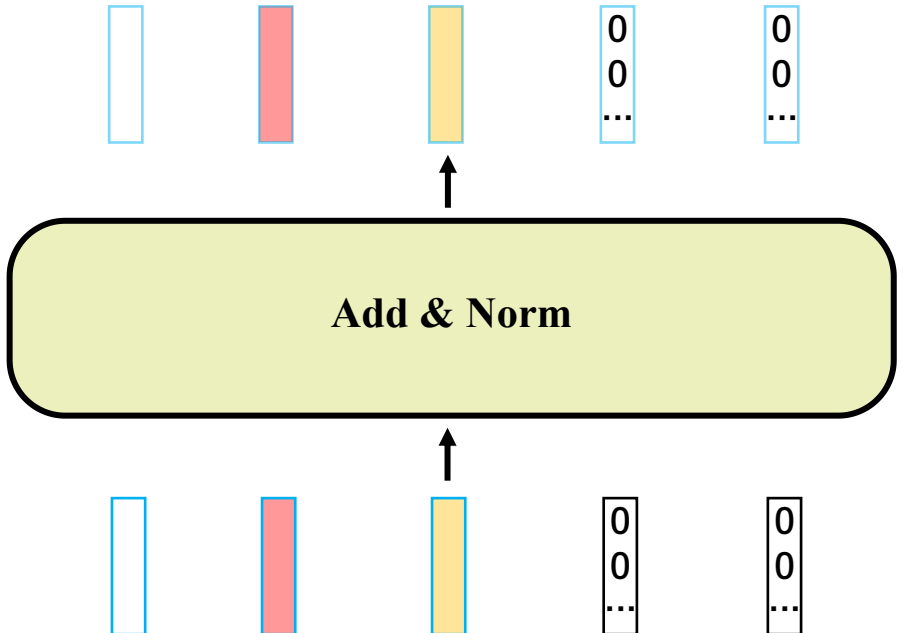
Transformer



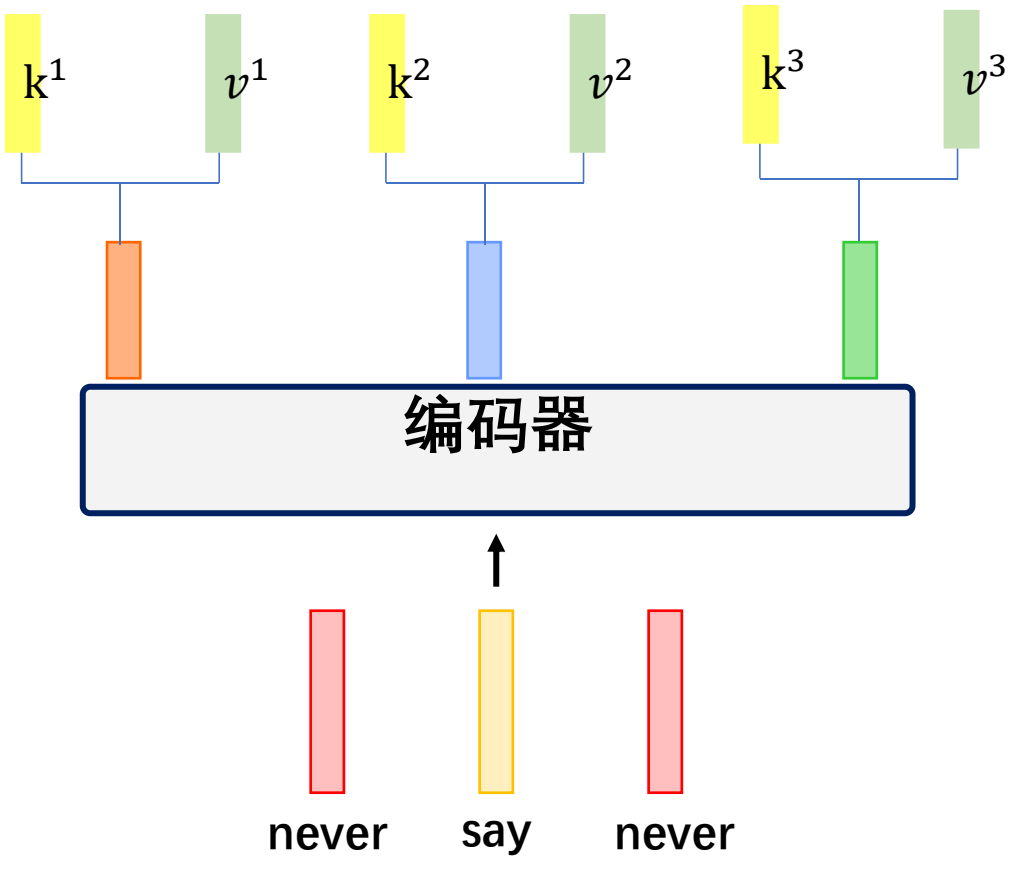


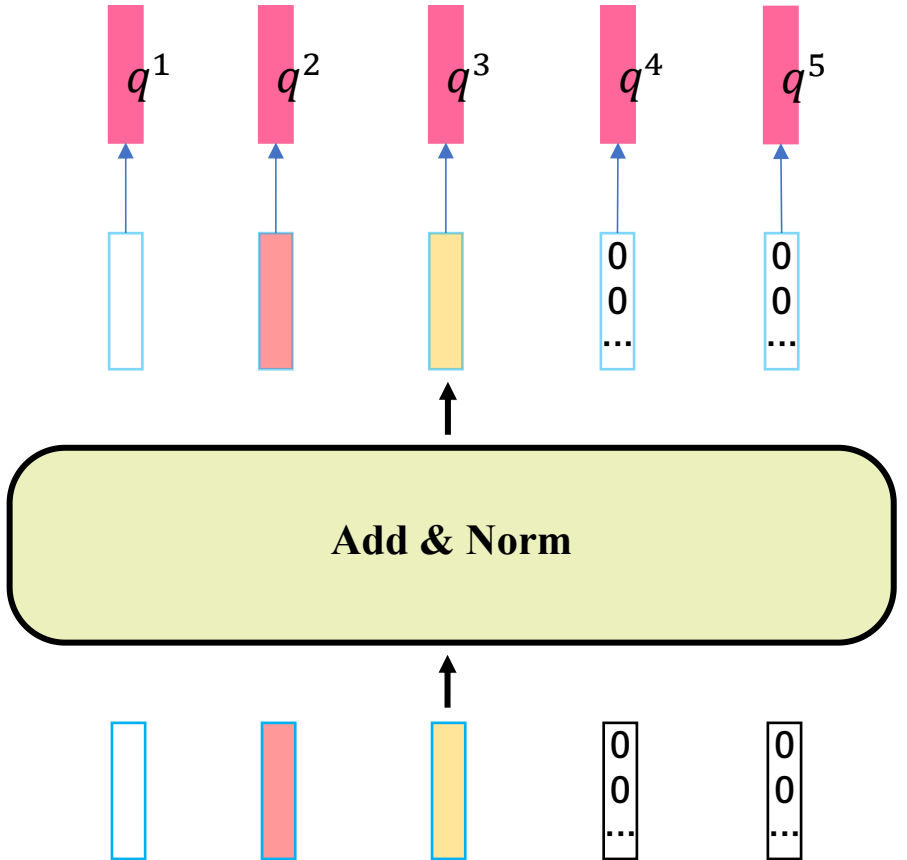
Transformer



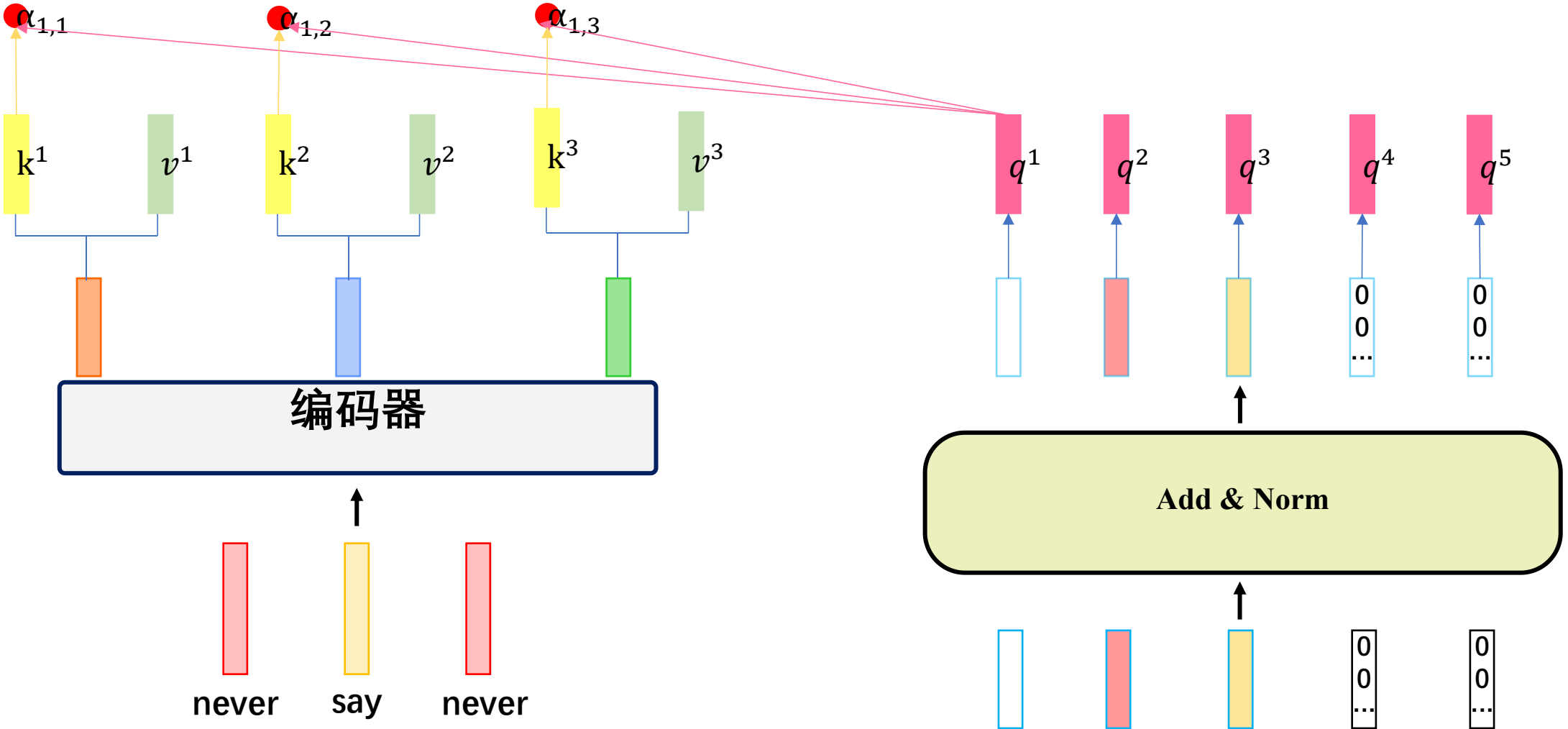


Transformer

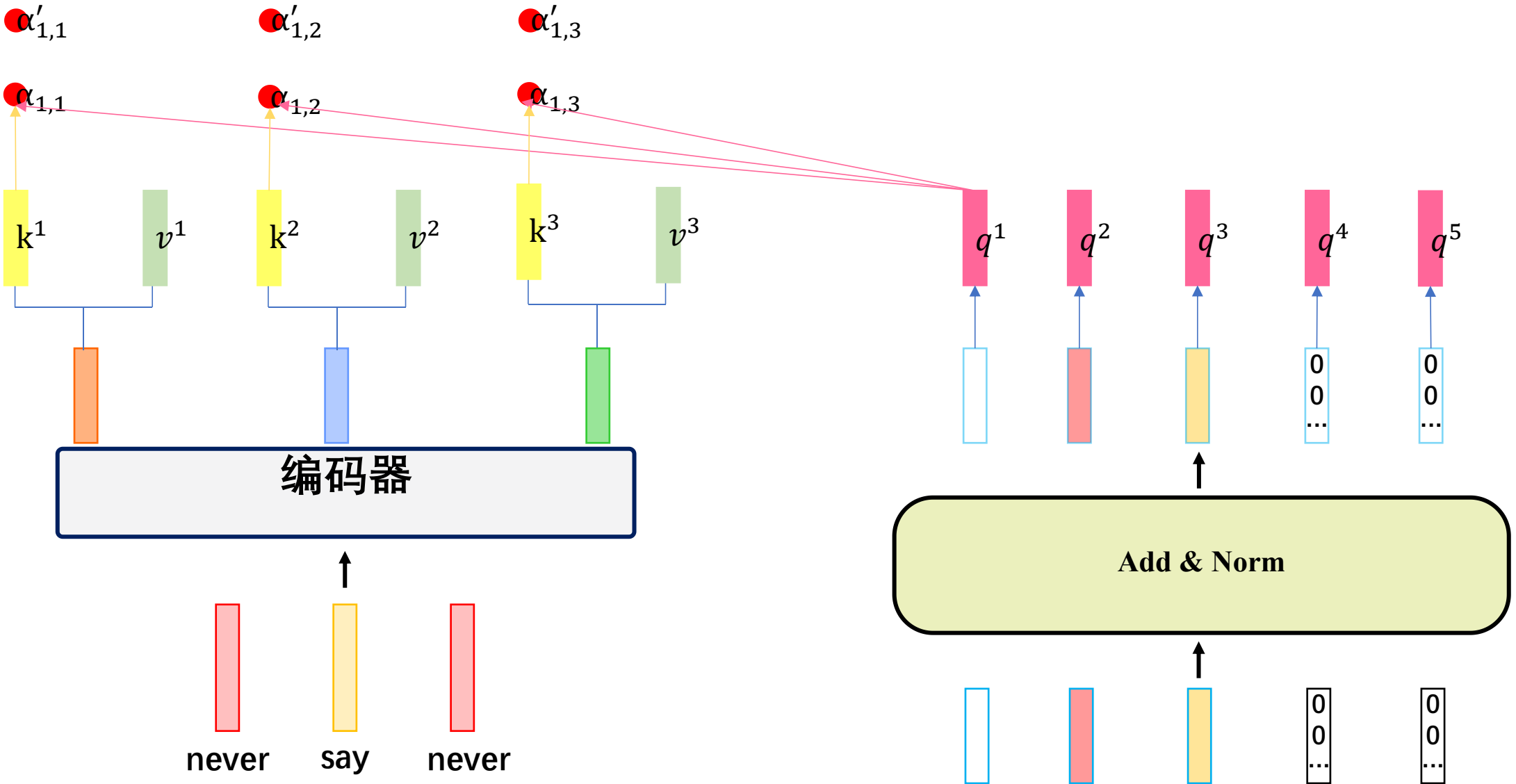




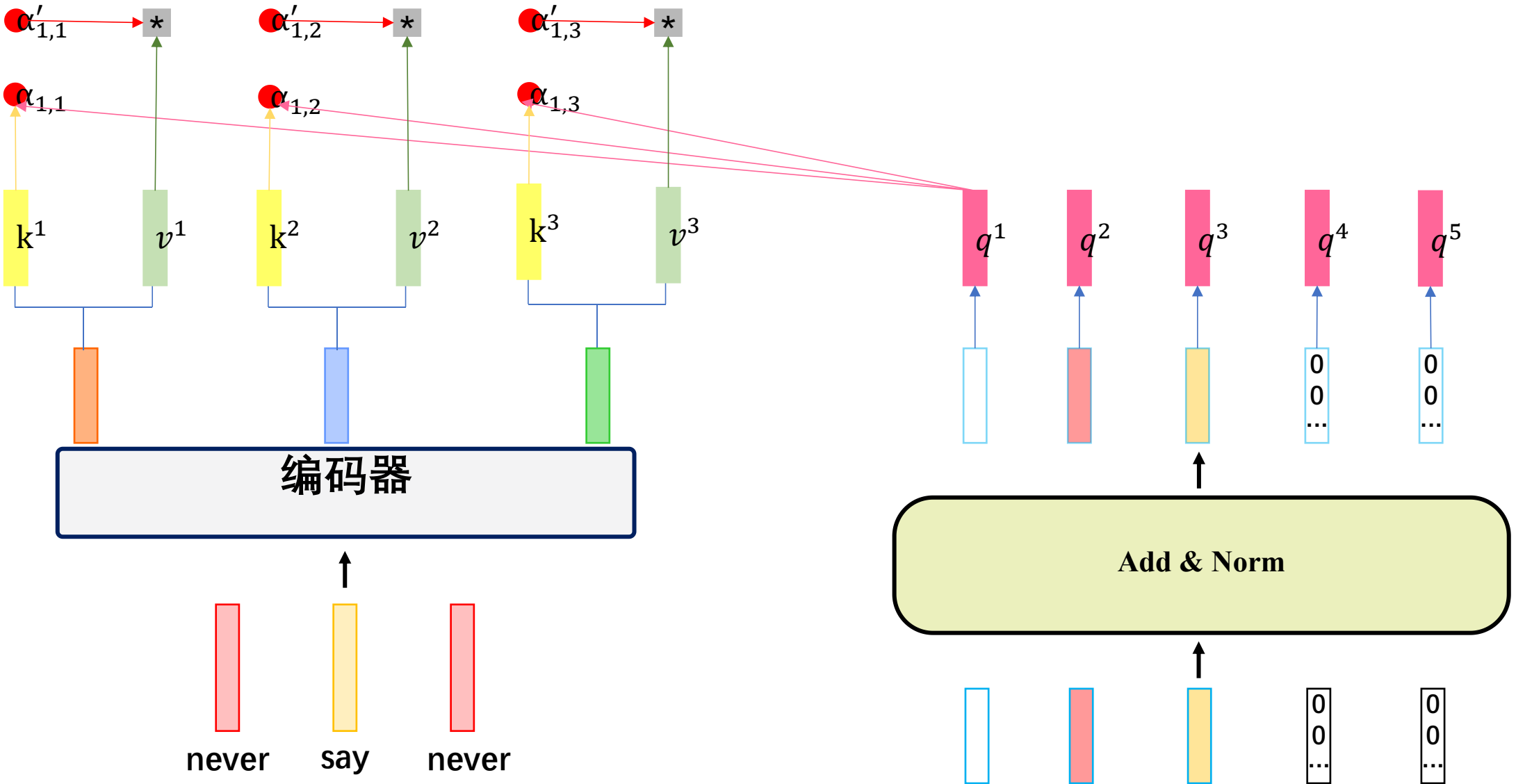
Transformer



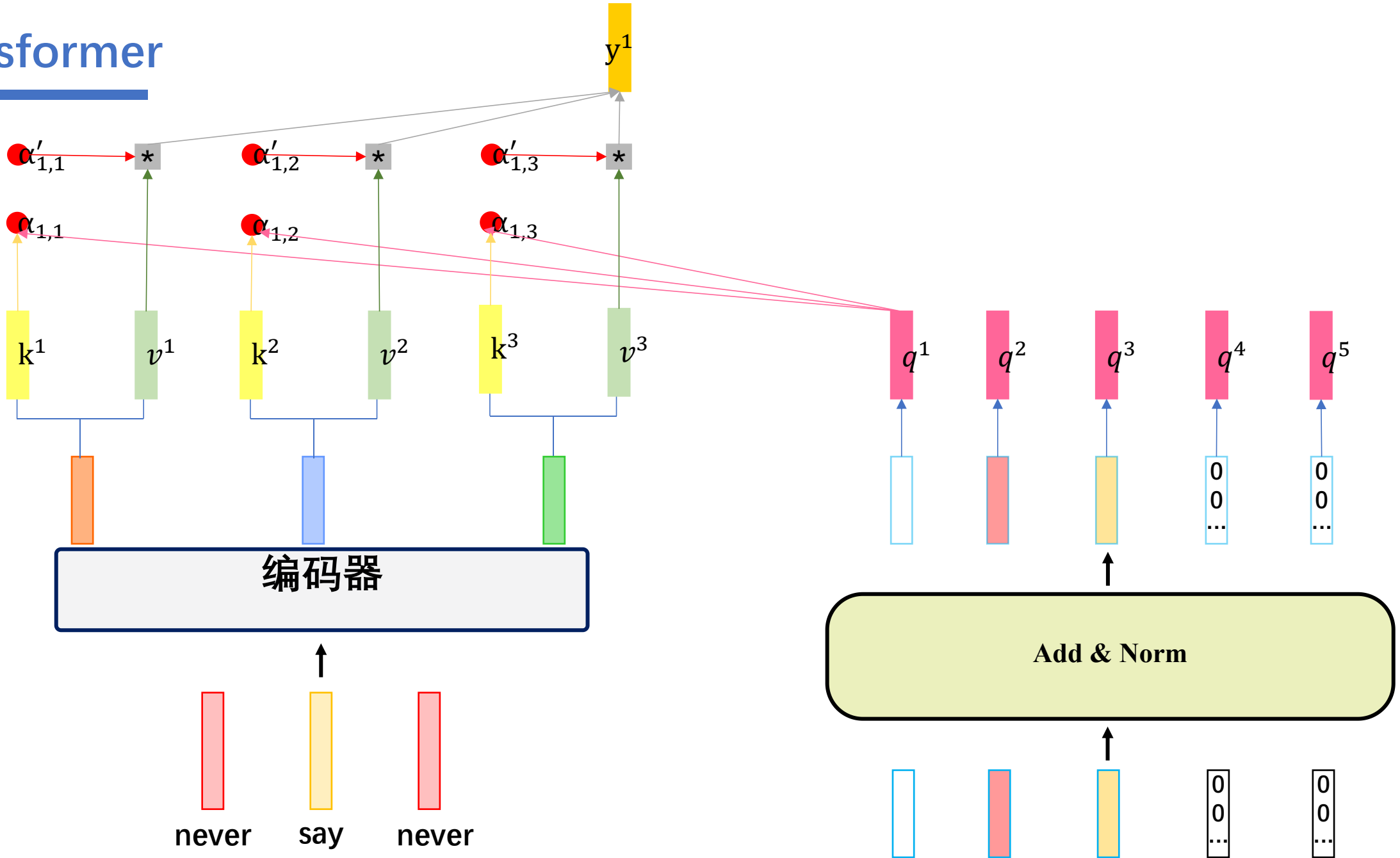
Transformer



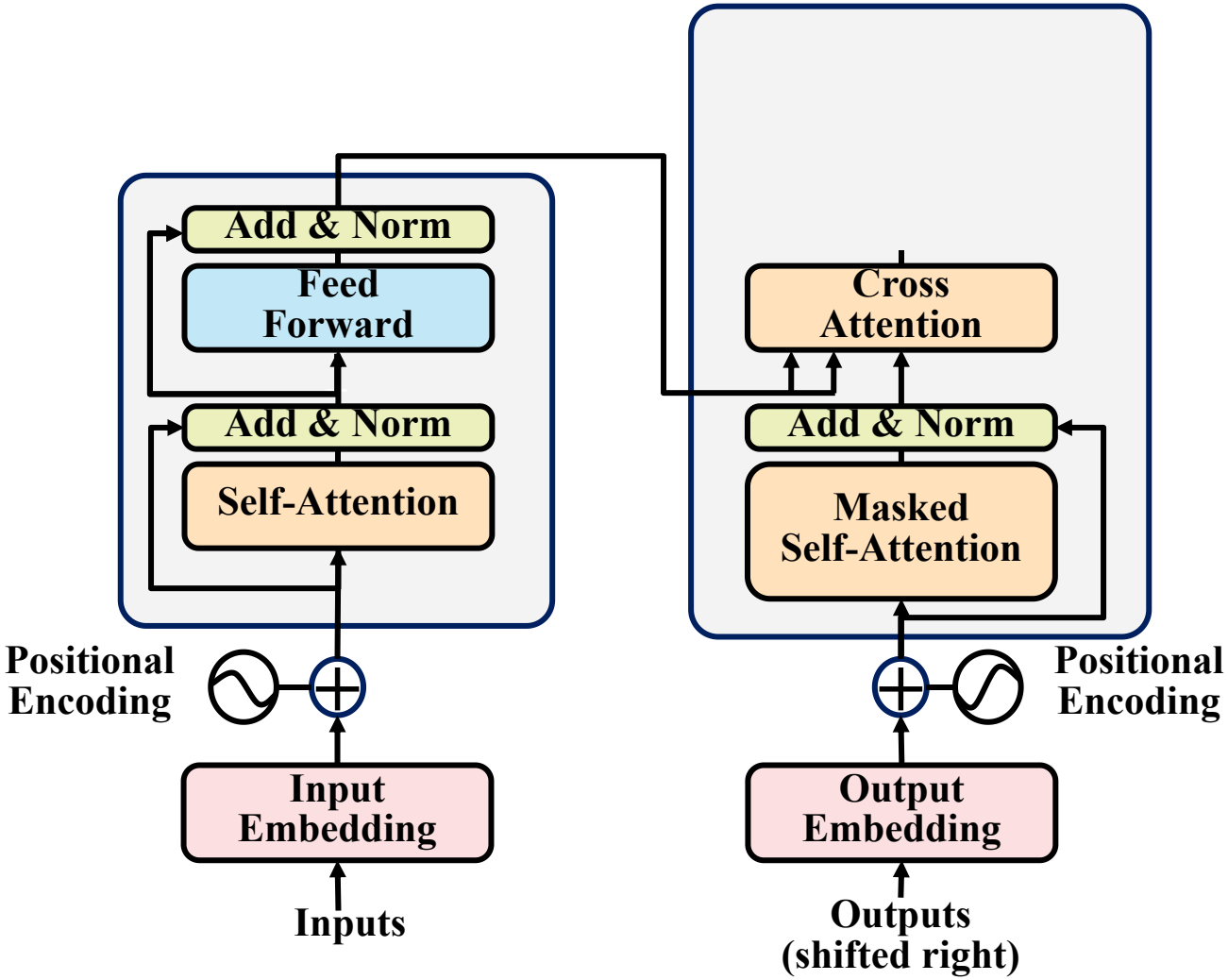
Transformer

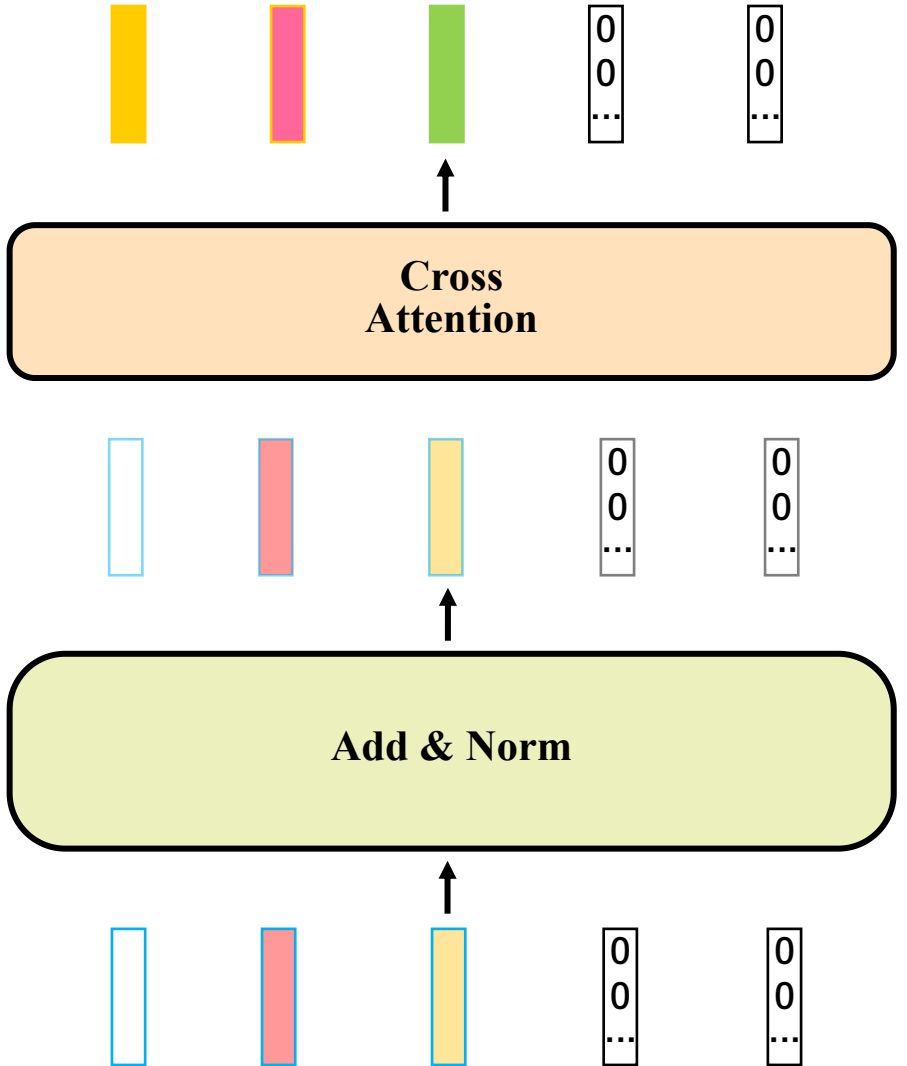


Transformer

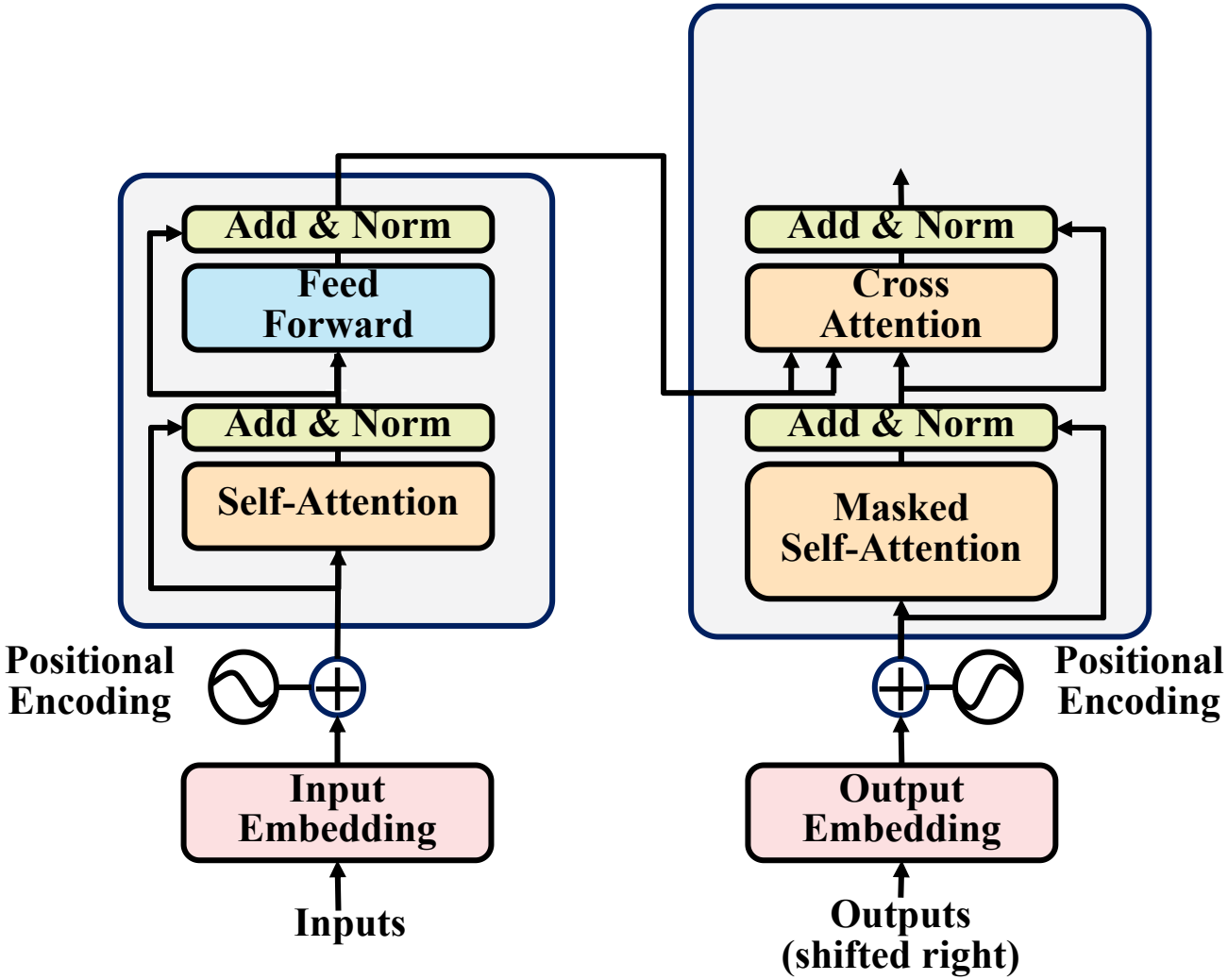


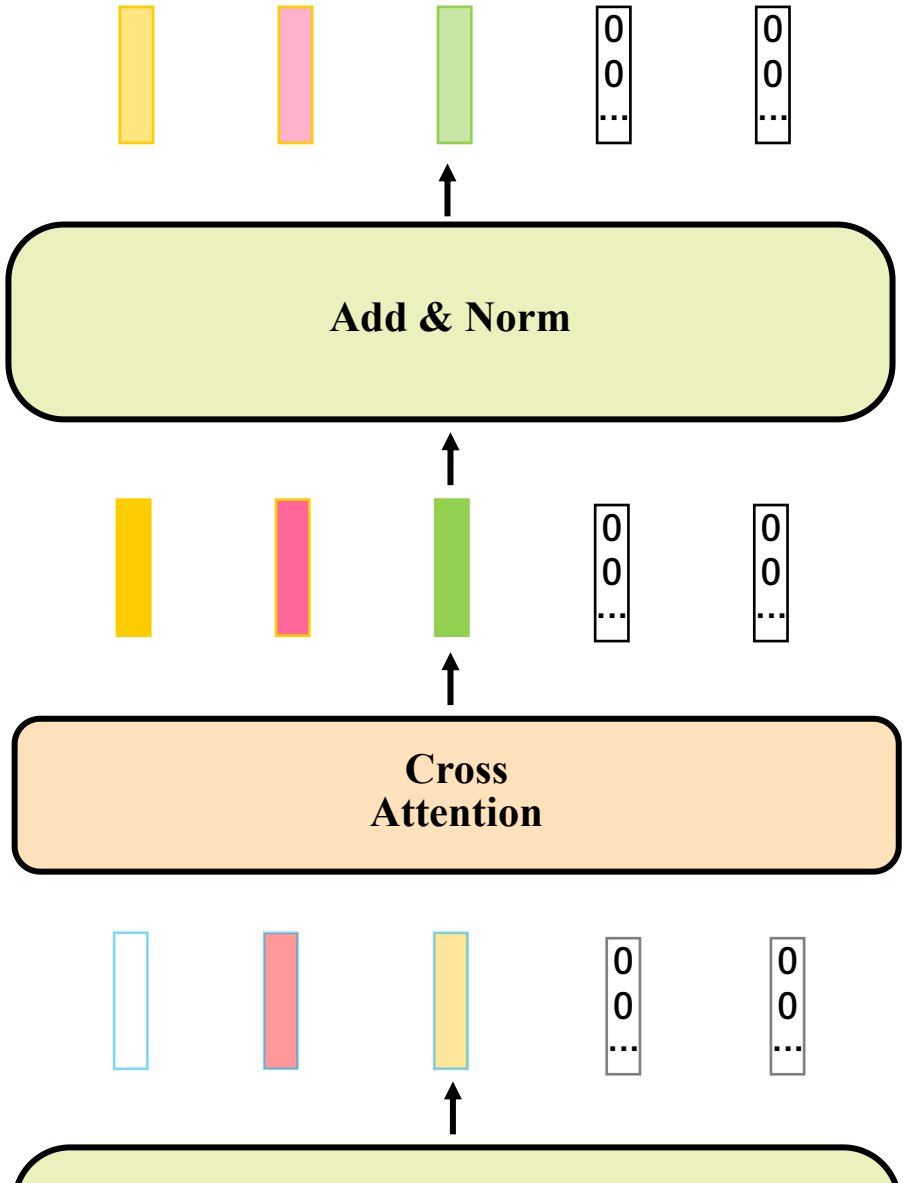
Transformer



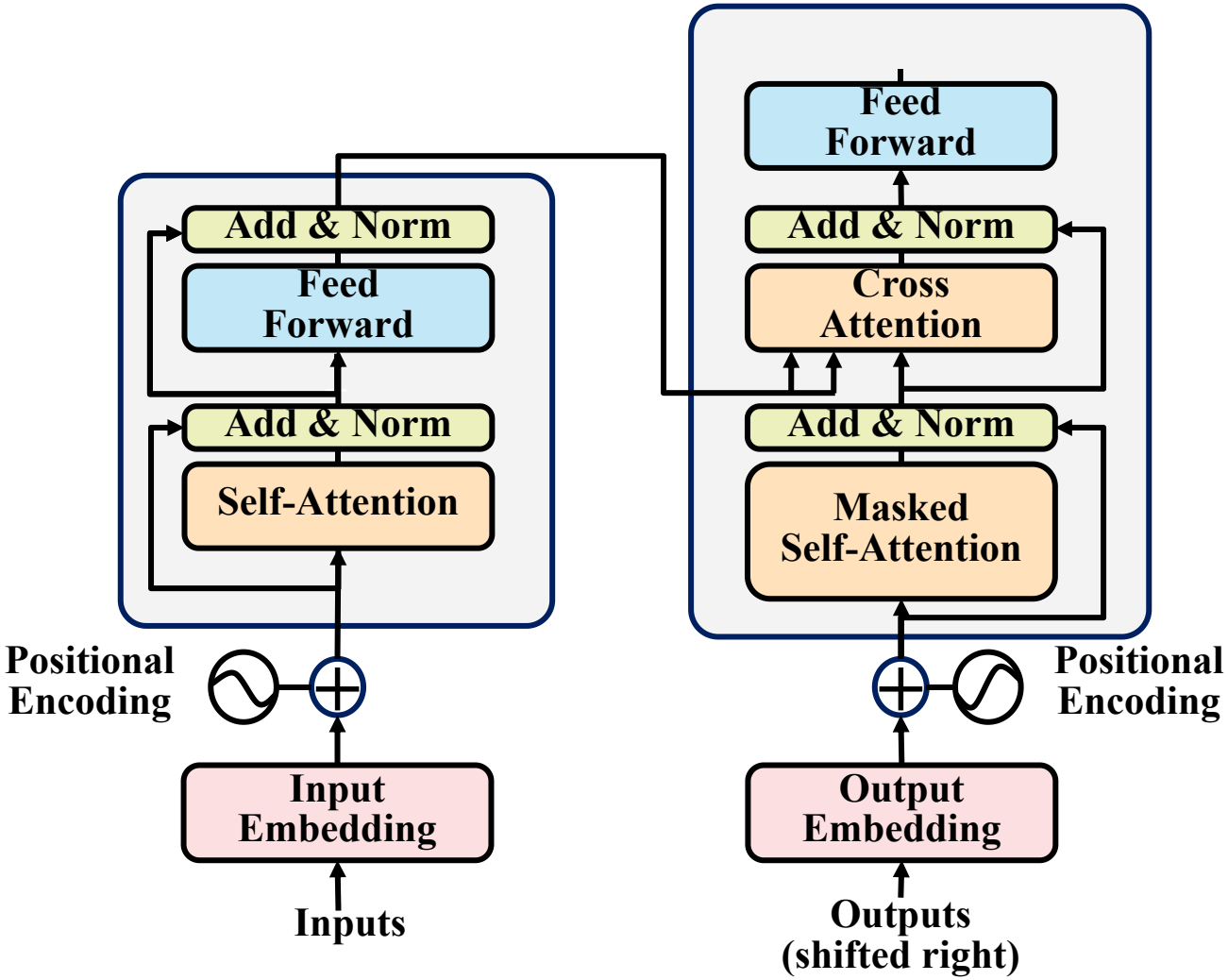


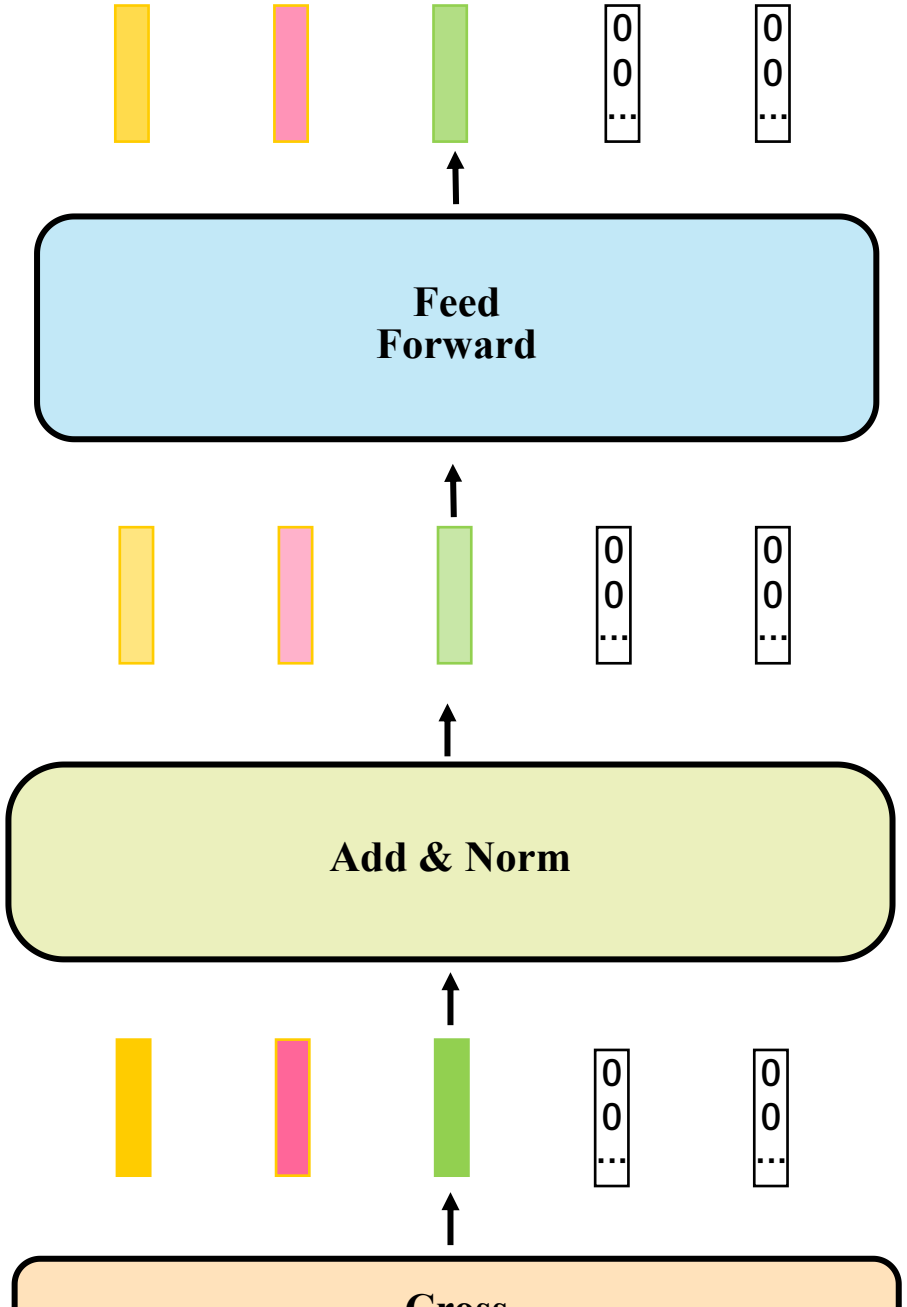
Transformer



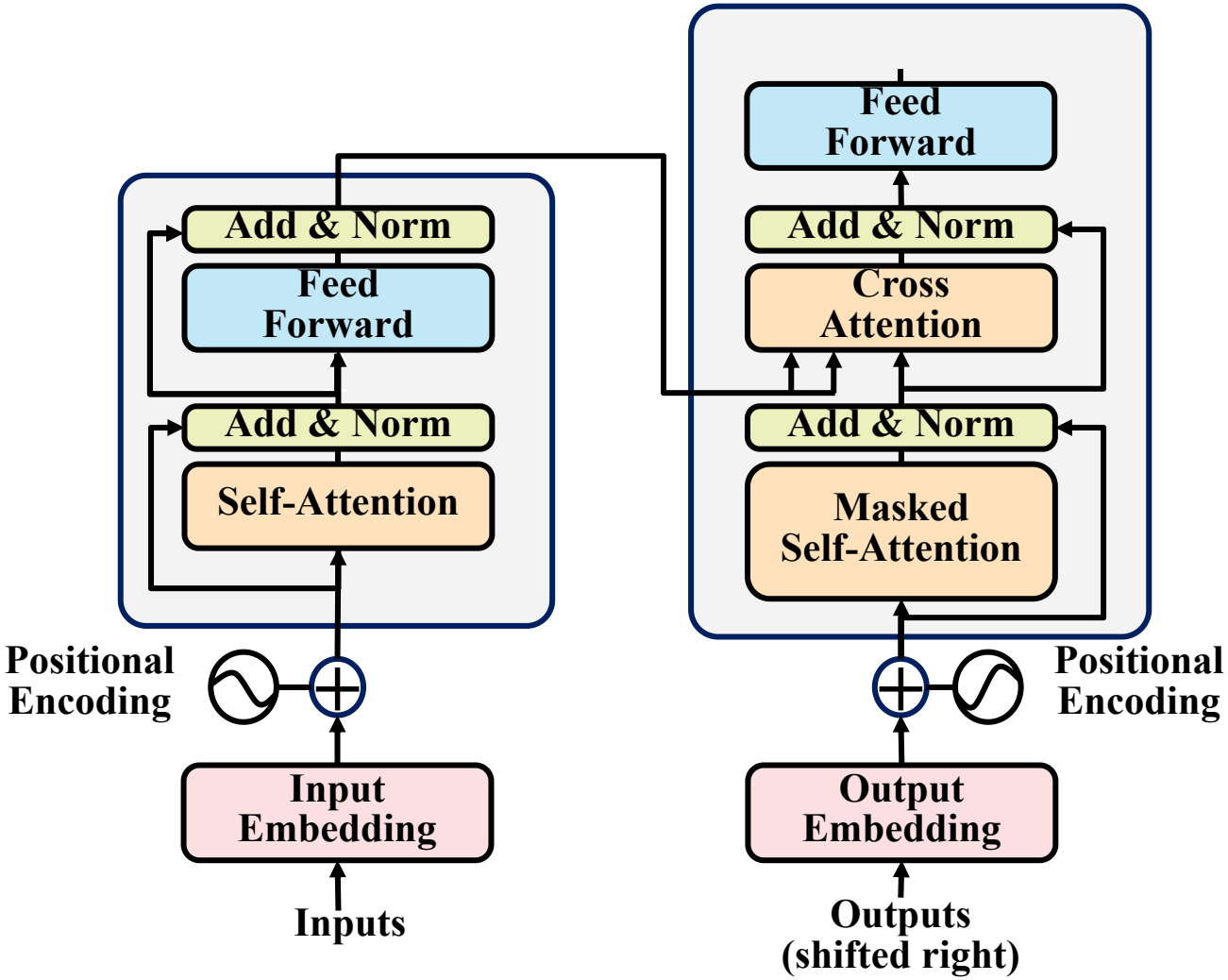


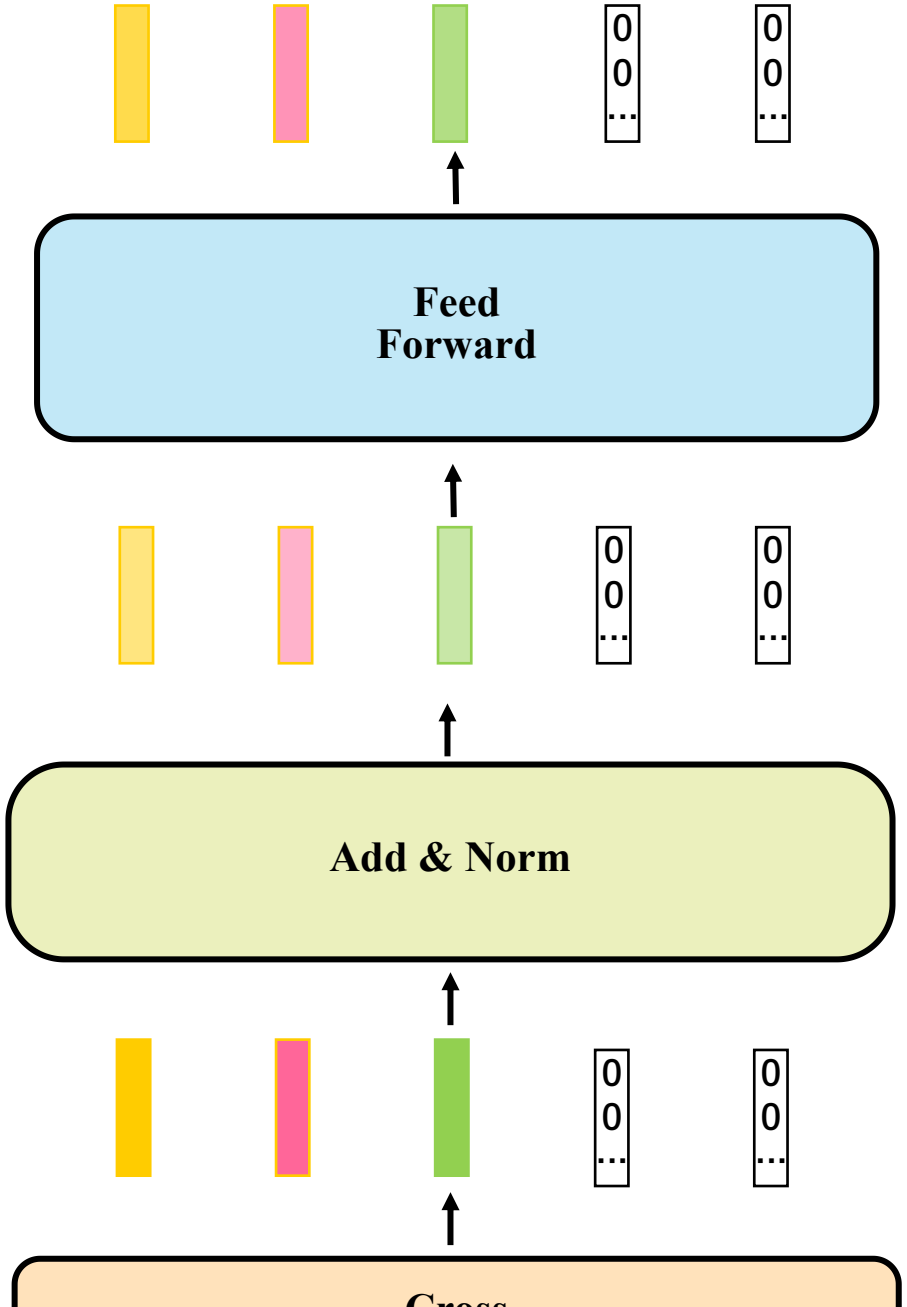
Transformer



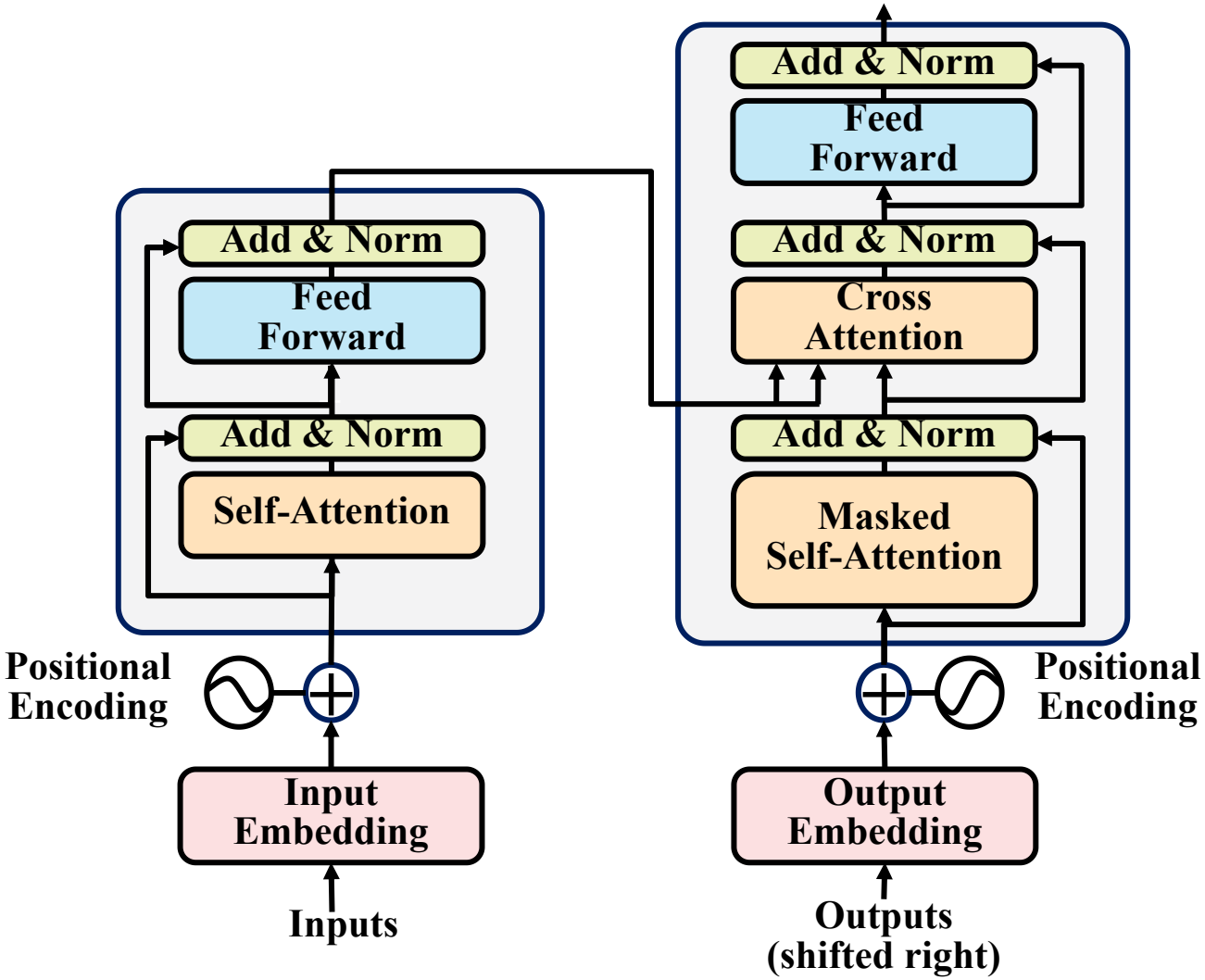


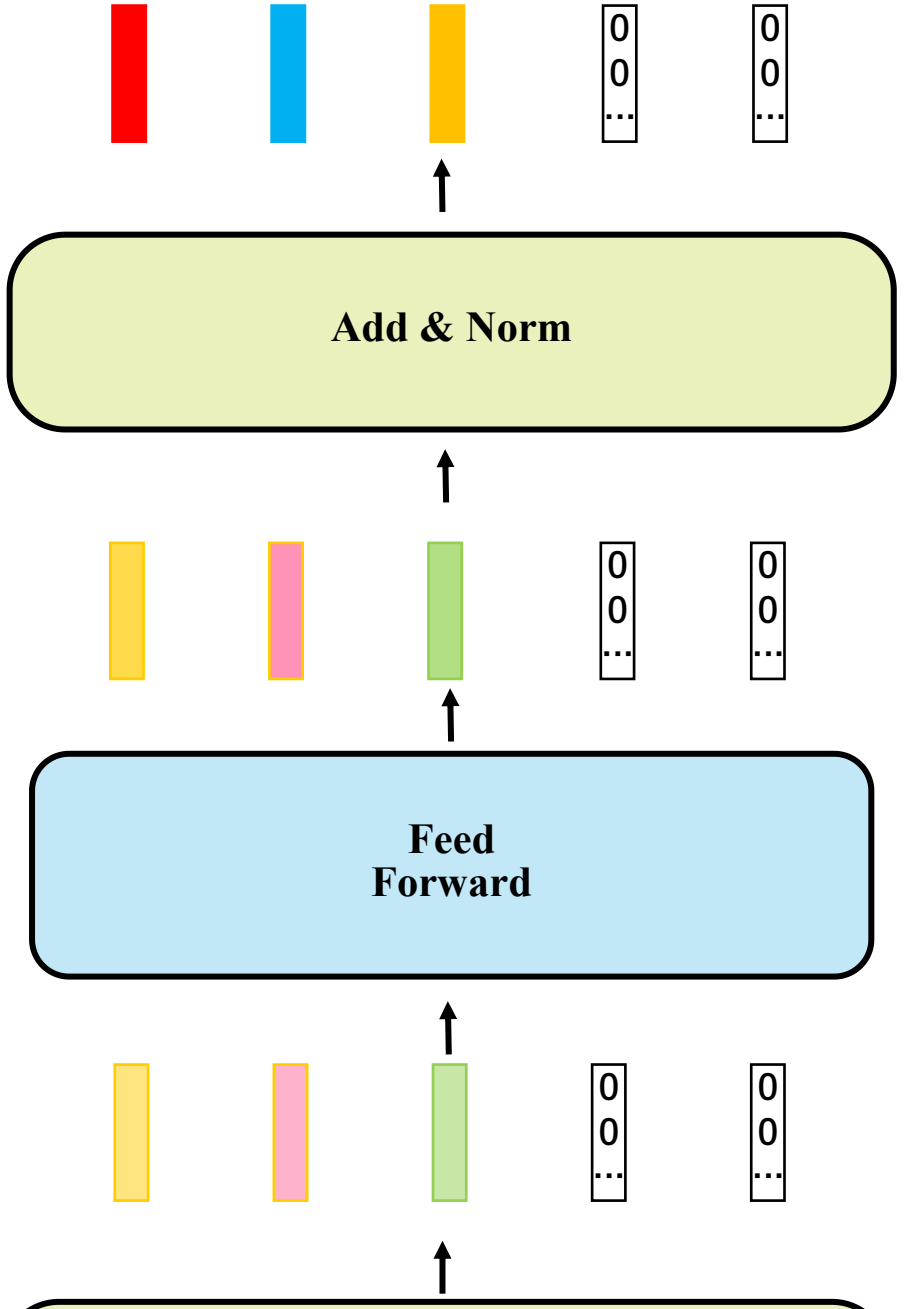
Transformer



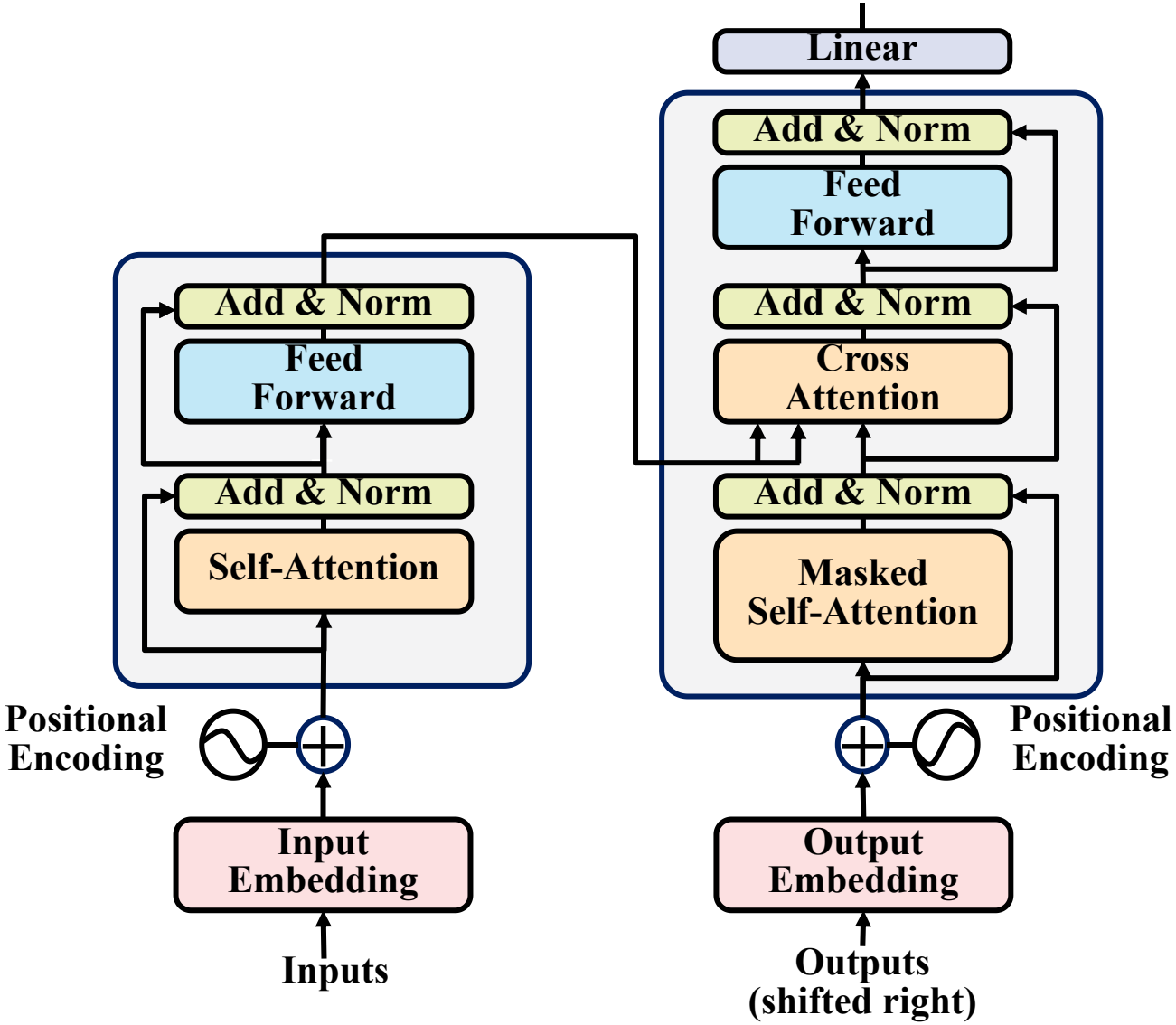


Transformer

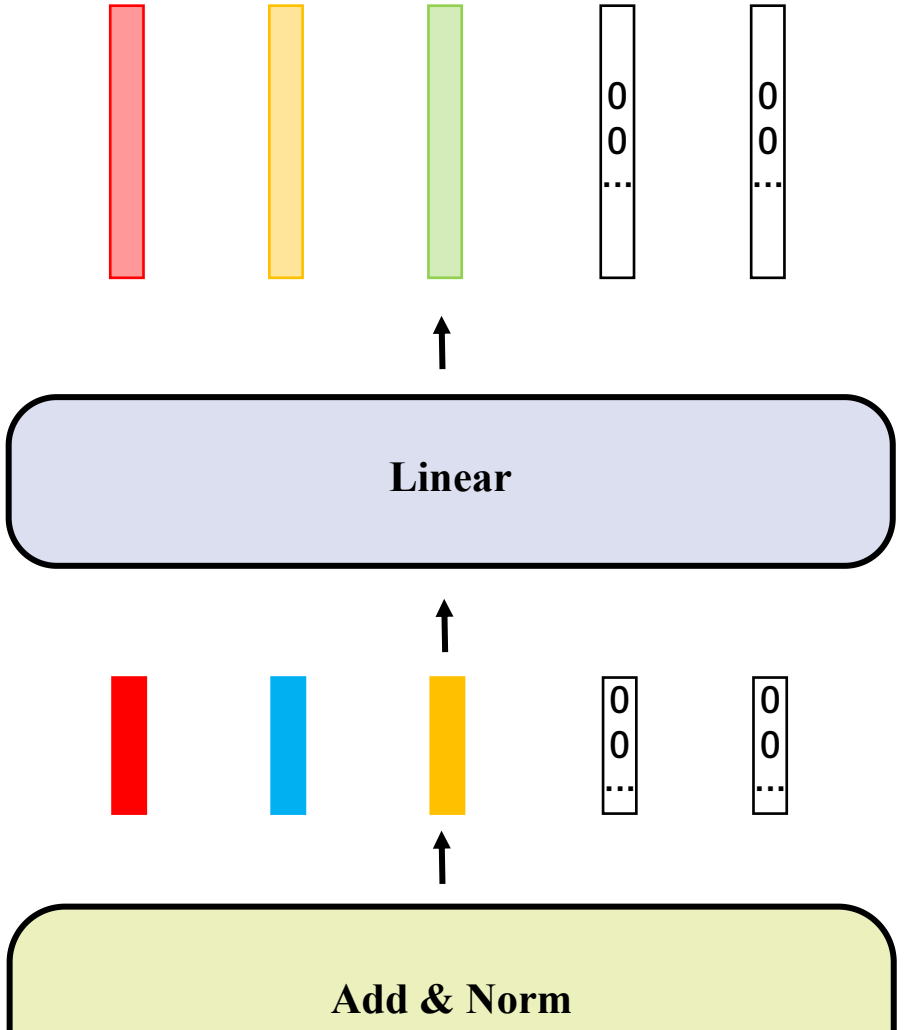




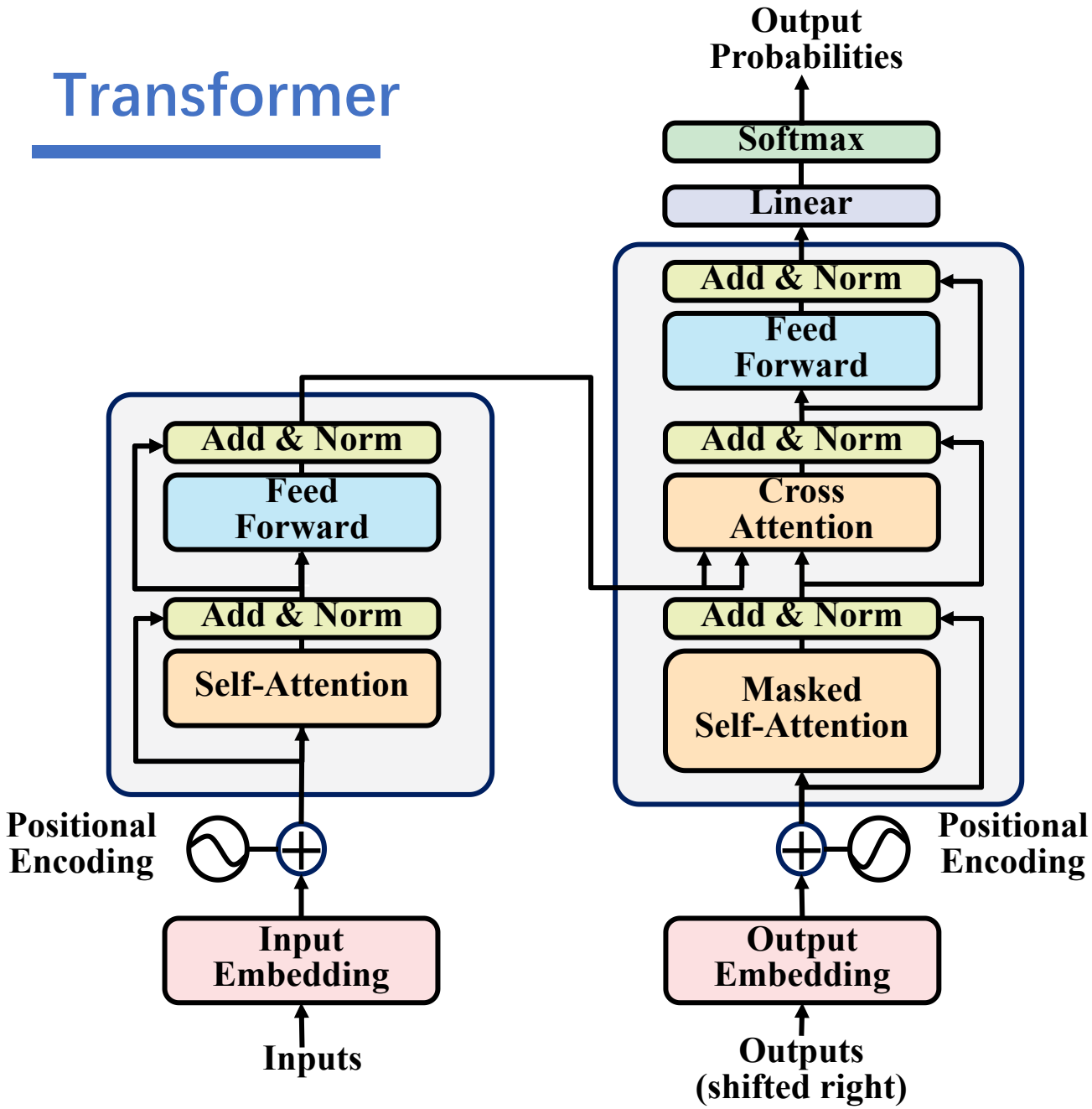
Transformer

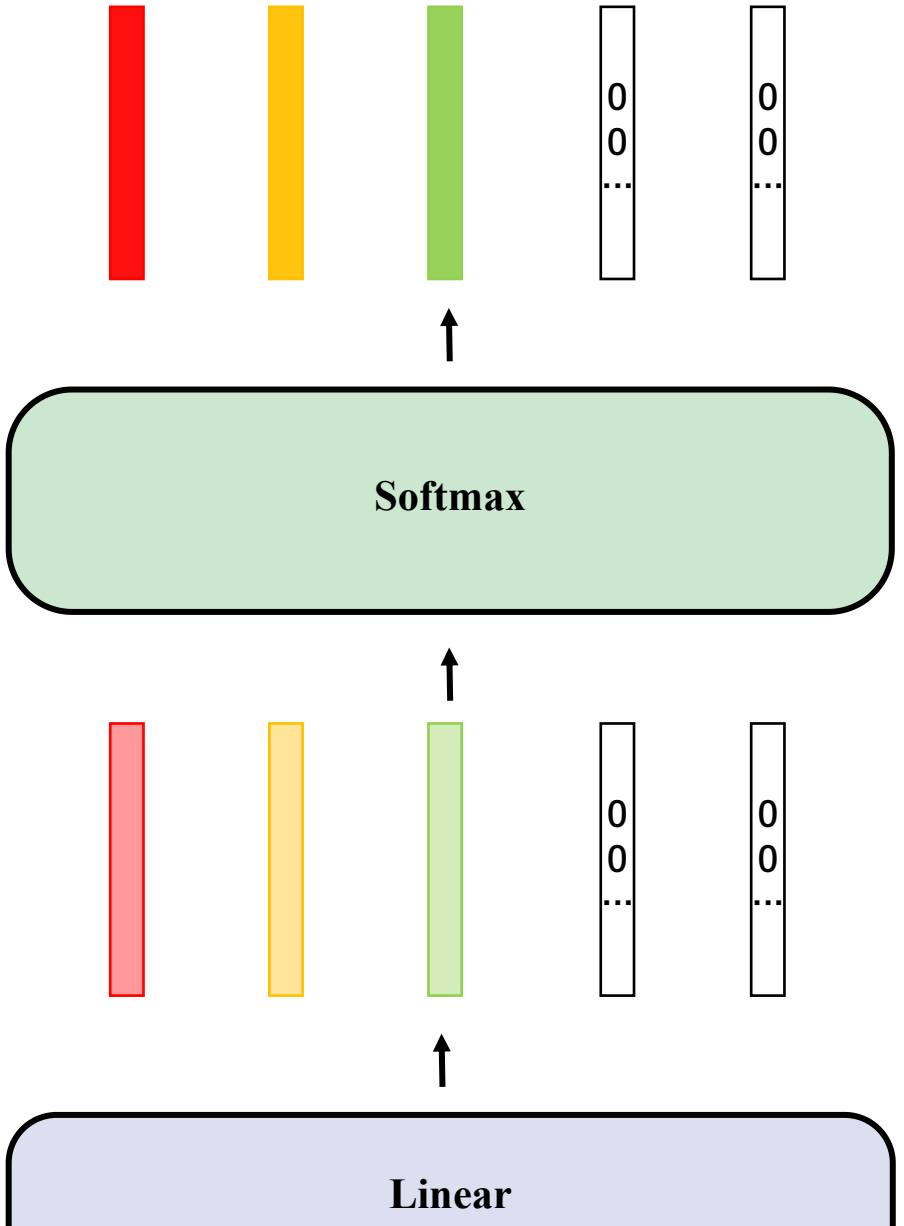


$$Linear(x) = Wx + b$$

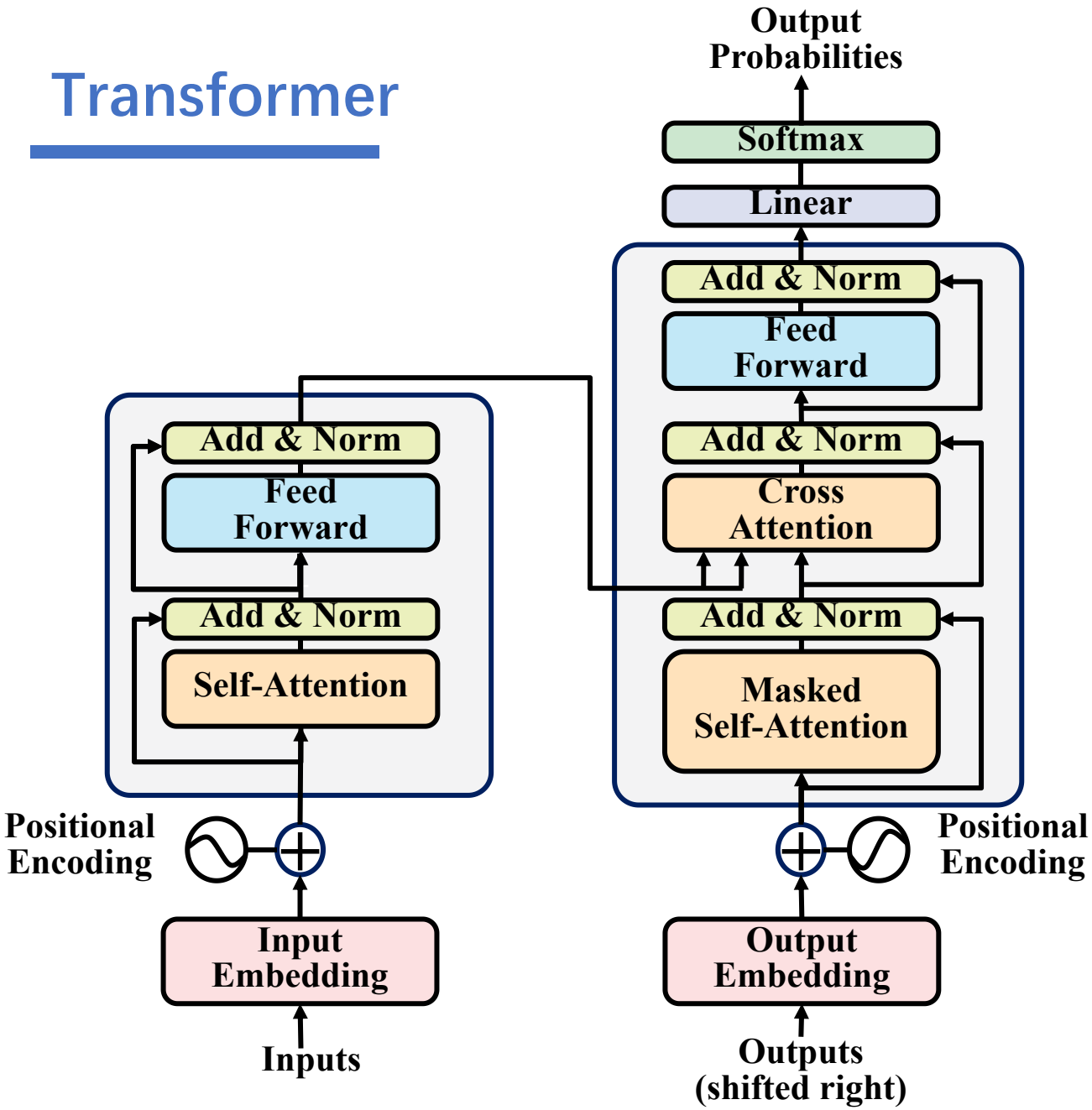


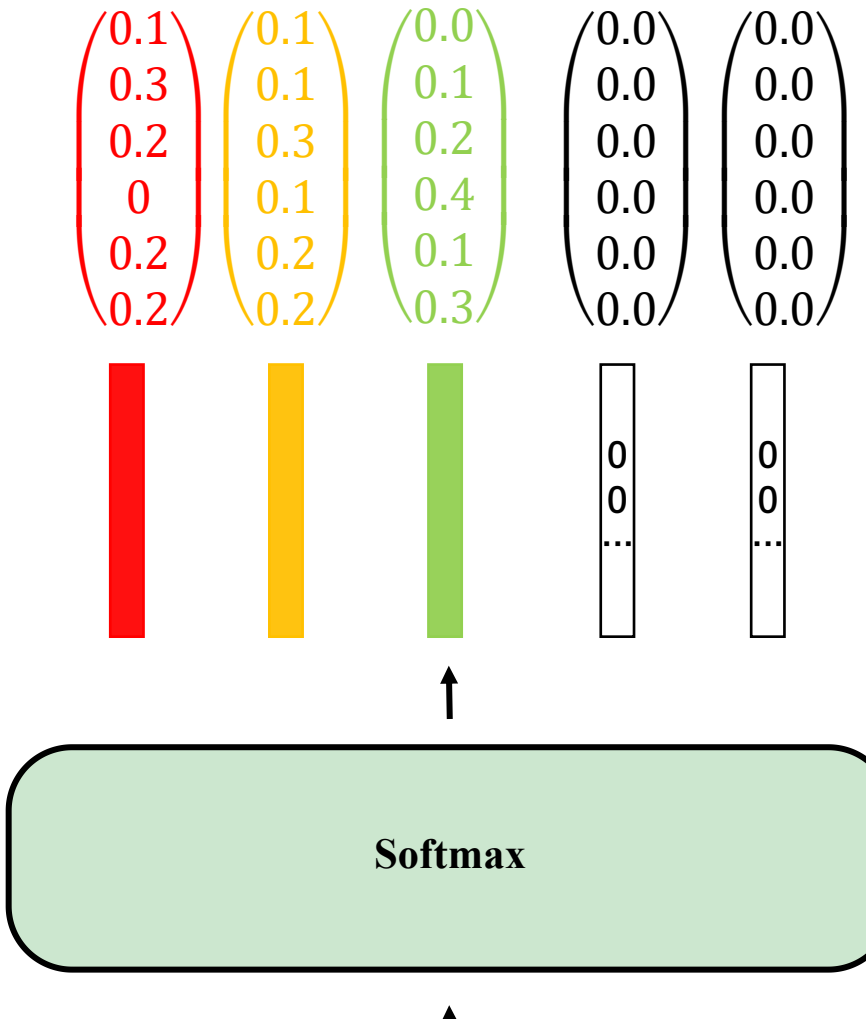
Transformer



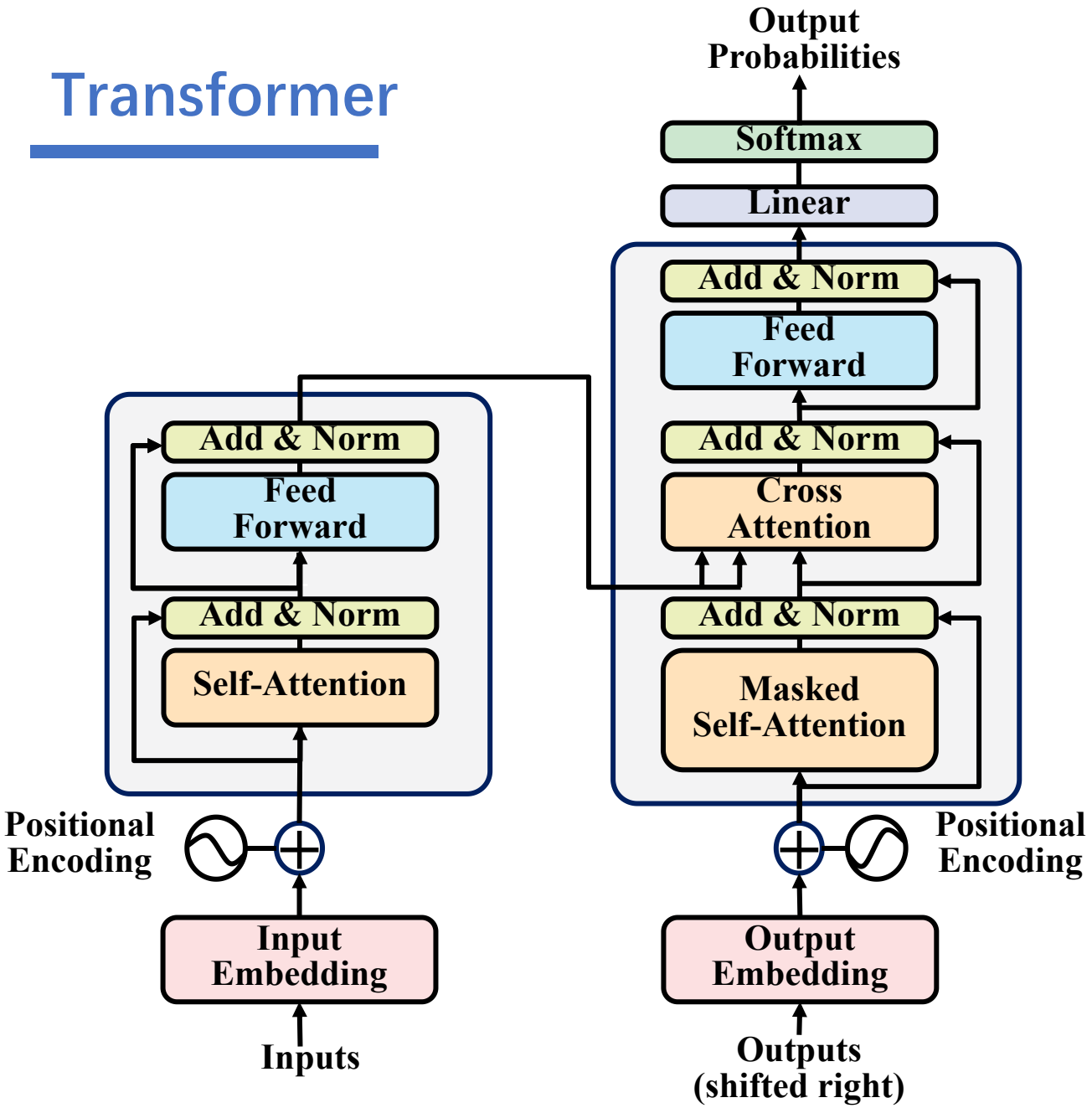


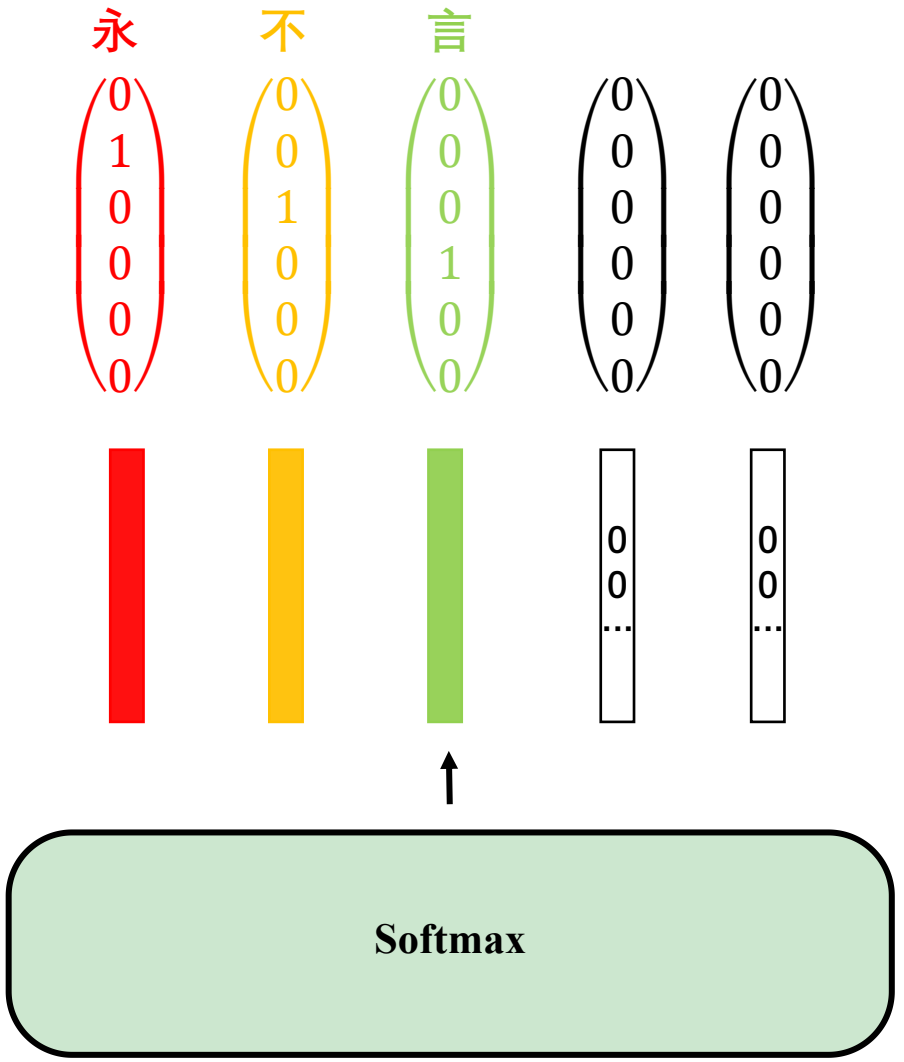
Transformer





Transformer





谢谢大家~