

$$\text{FFN}_t(\mathbf{x}'_t)$$

Feed-Forward Network

$$\text{FNN}_t^{(n)}$$

$$\mathbf{h}_1^{(1)}$$

...

$$\mathbf{h}_1^{(d_{ff})}$$

$$\mathbf{x}$$

$$\mathbf{x}'_t$$

Output Layer
($1 \times d_{model}$)

$$\mathbf{W}_2^{d_{ff} \times d_{model}}$$

Hidden Layer
($d_{ff} \times d_{model}$)

$$\mathbf{W}_1^{d_{model} \times d_{ff}}$$

Input Layer
($1 \times d_{model}$)