Rules :

	Scalar :	Vector
① $z = Wx$	$\boxed{\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z} \cdot W \qquad \frac{\partial L}{\partial W} = x \cdot \frac{\partial L}{\partial z}}$	$\begin{bmatrix} z: N\times 1 \\ x: M\times 1 \\ W: N\times M \end{bmatrix}$ $\boxed{\nabla_x L = (\nabla_z L) W \qquad \nabla_W L = x(\nabla_z L)}$
② $z = x \circ y$ $\circ$: component-wise multiplication		$[x, y, z: N\times 1]$ $\boxed{\nabla_x L = (\nabla_z L) \circ y^T \qquad \nabla_y L = (\nabla_z L) \circ x^T}$
③ $z = x + y$	$\boxed{\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z} \qquad \frac{\partial L}{\partial y} = \frac{\partial L}{\partial x}}$	$\begin{bmatrix} x, y, z: N\times 1 \\ \nabla_z L: 1\times N \end{bmatrix}$ $\boxed{\nabla_x L = \nabla_z L \qquad \nabla_y L = \nabla_z L}$
④ $z = g(x)$	$\boxed{\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z} g'(x)}$	$\begin{bmatrix} x, z: N\times 1 \\ \nabla_z L: 1\times N \\ J_x g: \text{Jacobian} \end{bmatrix}$ $\boxed{\nabla_x L = \nabla_z L \, J_x g = \nabla_z L \circ g'(x)^T}$
⑤ $z = g(x)$ $y = h(x)$	$\boxed{\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z} g'(x) + \frac{\partial L}{\partial y} h'(x)}$	

$$\text{delta} = \frac{\partial L}{\partial x(l+1, t)} + \frac{\partial L}{\partial h(l, t+1)}$$

① $z_t = \sigma(W_{zh} h_{t-1} + W_{zx} x_t)$

② $r_t = \sigma(W_{rh} h_{t-1} + W_{rx} x_t)$

③ $\tilde{h}_t = \tanh(W_h (r_t \otimes h_{t-1}) + W_x x_t)$

④ $h_t = (1 - z_t) \otimes h_{t-1} + z_t \otimes \tilde{h}_t \quad \Rightarrow \quad h_t = h_{t-1} - z_t \circ h_{t-1} + z_t \circ \tilde{h}_t$

$\Rightarrow$ ① : $z_1 = W_{zh} h_{t-1}$.
$z_2 = W_{zx} x_t$
$z_3 = z_1 + z_2$.
$z_t = \sigma(z_3)$

② : $z_4 = W_{rh} h_{t-1}$
$z_5 = W_{rx} x_t$
$z_6 = z_4 + z_5$
$r_t = \sigma(z_6)$.

③ $z_7 = r_t \circ h_{t-1}$
$z_8 = W_h \cdot z_7$
$z_9 = W_x x_t$
$z_{10} = z_8 + z_9$
$\tilde{h}_t = \tanh(z_{10})$

④ $z_{11} = 1 - z_t$
$z_{12} = z_{11} \circ h_{t-1}$
$z_{13} = z_t \circ \tilde{h}_t$
$h_t = z_{12} + z_{13}$.

$\nabla_{z_{12}} L = \nabla_{h_t} L$.
$\nabla_{z_{13}} L = \nabla_{h_t} L$
$\nabla_{z_t} L = \nabla_{z_{13}} L \circ \tilde{h}_t^T$
$\nabla_{\tilde{h}_t} L = \nabla_{z_{13}} L \circ z_t^T$
$\nabla_{z_{11}} L = \nabla_{z_{12}} L \circ h_{t-1}^T$
$\nabla_{h_{t-1}} L = \nabla_{z_{12}} L \circ z_{11}^T$
$\nabla_{z_t} L \mathrel{+}= -\nabla_{z_{11}} L$.

$\nabla_{z_{10}} L = \nabla_{\tilde{h}_t} L \circ (1 - \tanh^2(z_{10}))^T$
$\nabla_{z_8} L = \nabla_{z_{10}} L$.
$\nabla_{z_9} L = \nabla_{z_{10}} L$.
$\nabla_{W_x} L = x_t (\nabla_{z_9} L)$
$\nabla_{x_t} L = (\nabla_{z_9} L) W_x$
$\nabla_{W_h} L = z_7 (\nabla_{z_8} L)$
$\nabla_{z_7} L = (\nabla_{z_8} L) W_h$.
$\nabla_{r_t} L = (\nabla_{z_7} L) \circ h_{t-1}^T$
$\nabla_{h_{t-1}} L \mathrel{+}= (\nabla_{z_7} L) \circ r_t^T$

$\nabla_{z_6} L = \nabla_{r_t} L \circ \sigma(z_6)^T \circ (1 - \sigma(z_6))^T$
$\nabla_{z_4} L = \nabla_{z_6} L$
$\nabla_{z_5} L = \nabla_{z_6} L$.
$\nabla_{W_{rx}} L = x_t (\nabla_{z_5} L)$.
$\nabla_{x_t} L \mathrel{+}= (\nabla_{z_5} L) W_{rx}$
$\nabla_{W_{rh}} L = h_{t-1} (\nabla_{z_4} L)$.
$\nabla_{h_{t-1}} L \mathrel{+}= (\nabla_{z_4} L) W_{rh}$

$$\nabla_{z_3} L = \nabla_{z_t} L \circ \sigma(z_3 L)^T \circ (1-\sigma(z_3))^T$$

$$\nabla_{z_2} L = \nabla_{z_3} L .$$

$$\nabla_{z_1} L = \nabla_{z_3} L$$

$$\nabla_{W_{zx}} L = X_t \, (\nabla_{z_2} L)$$

$$\nabla_{X_t} L \mathrel{+}= (\nabla_{z_2} L) \, W_{zx}$$

$$\nabla_{W_{zh}} L = h_{t-1} \, (\nabla_{z_1} L)$$

$$\nabla_{h_{t-1}} L \mathrel{+}= (\nabla_{z_1} L) \, W_{zh} .$$