

$$x_{pr} \in \mathbb{R}^{H \times 1}$$

$$H_{conv_2} 1 \times 3$$

$$H_{conv_3} 1 \times 3$$

$$F_K$$

$$RTB_K$$

$$F_1$$

$$RTB_1$$

$$H_{conv} 1 \times 3$$

$$WTL_L$$

$$\dots F_{0,1}$$

$$WTL_1$$

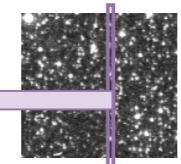
$$F_{0,0}$$

$$APE$$

$$H_{conv_1} 1 \times 3$$

$$\text{Normalization}$$

$$x_{lq} \in \mathbb{R}^{H \times 1}$$



WTL

$$+$$

$$MLP$$

$$\text{LayerNorm}$$

$$+$$

$$+$$

$$MLP$$

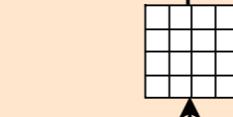
$$\text{LayerNorm}$$

$$+$$

Window based MSA

Head₁

$$Att = QK^T + RPE \rightarrow \text{Sigmoid}$$



$$Q, K, V \in \mathbb{R}^{wsize \times \frac{c}{nhead}}$$

$$\text{Linear}, c, 3c$$

$$RPE$$

$$\in \mathbb{R}^{wsize \times c}$$

Head_{nhead}

MSA

MSA

$$Window_1 \quad Window_2 \quad \dots \quad Window_n$$

$$\text{LayerNorm}$$

$$F_{j,i-1} \in \mathbb{R}^{N \times c} \quad c$$

$$N = \frac{H}{p}$$