

$$\begin{aligned}
Z_0 &= [X_{class}; X_p^1 \mathbf{E}; X_p^2 \mathbf{E}; \dots X_p^N \mathbf{E}] + \mathbf{E}_{pos}, \mathbf{E} \in \mathbb{R}^{(p^2, c)D}, \mathbf{E} \in \mathbb{R}^{(N+1)D} \\
Z'_\ell &= MSA(LN(Z_{\ell-1})) + Z_{\ell-1} \ell = 1 \dots L \\
Z_\ell &= MLP(LN(Z'_\ell)) + Z'_\ell \ell = 1 \dots L \\
Y &= LN(Z_L^0)
\end{aligned} \tag{1}$$