MLP Formula with Updated Notation

The MLP contains two layers with a GELU non-linearity:

$$z_0 = \left[x_{\text{class}}, x_p^1 \mathbf{E}, x_p^2 \mathbf{E}, \dots, x_p^T \mathbf{E} \right] + \mathbf{E}_{pos},$$
$$z'_\ell = \text{MSA}(\text{LN}(z_{\ell-1})) + z_{\ell-1}, \quad \ell = 1 \dots L,$$
$$z_\ell = \text{MLP}(\text{LN}(z'_\ell)) + z'_\ell, \quad \ell = 1 \dots L,$$
$$y = \text{LN}(z_0^L).$$

where:

$$\mathbf{E} \in \mathbb{R}^{(P^2 \cdot \text{Col}) \times C},$$
$$\mathbf{E}_{pos} \in \mathbb{R}^{(T+1) \times C}.$$