

$x_{l+1}: [N, L, D]$

Layer Norm

addition

FFN

Layer Norm

addition

MHA

$x_l: [N, L, D]$

