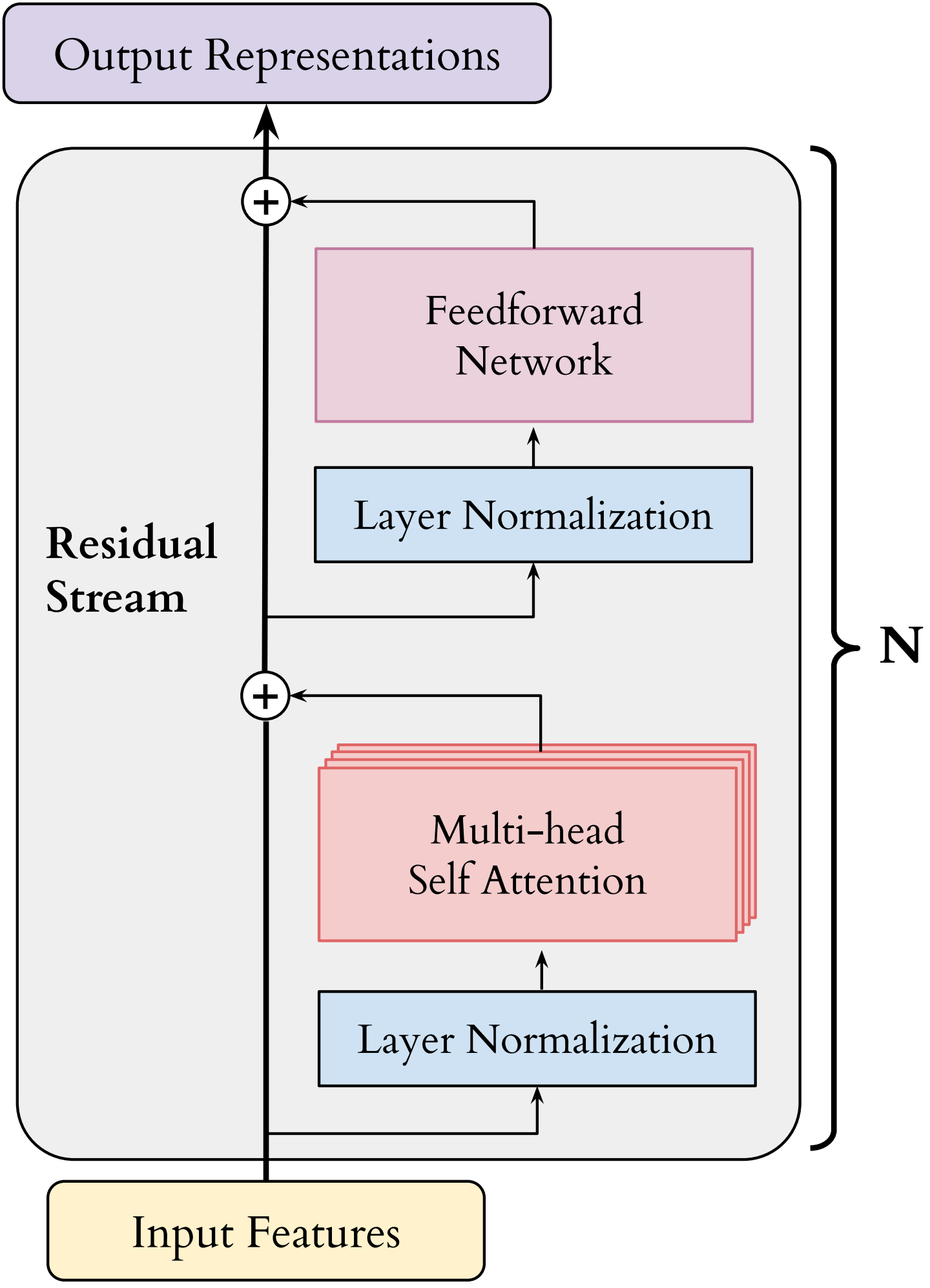
Output Representations
Feedforward Network
Layer Normalization
Residual Stream
Multi-head Self Attention
Layer Normalization
N
Input Features