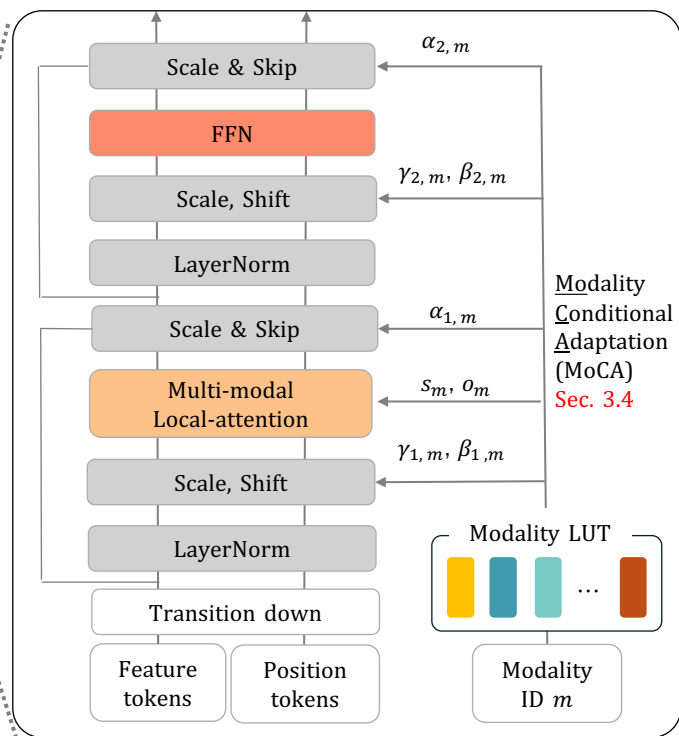
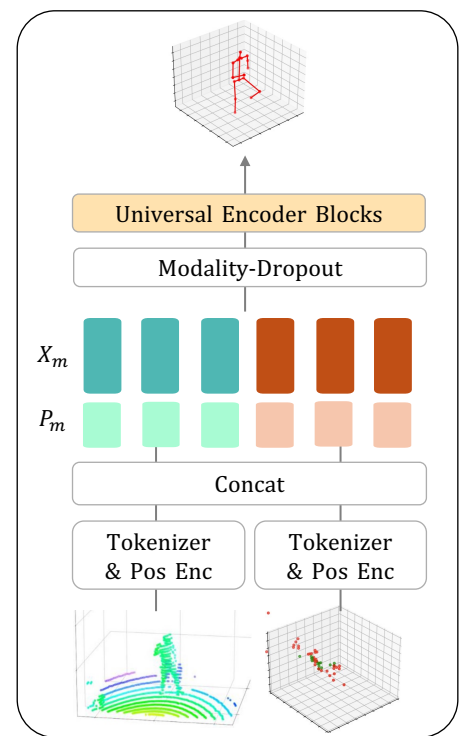


(a) Universal Transformer



(b) Universal Encoder Block



(c) Multi-modal fusion (Sec 3.5)