

Concatenate



Positional Encoding



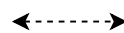
Layer Normalization



DropOut



Summation



Weights sharing



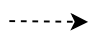
Loop start

D

Latent dimension

MHA

Multi-Head Attention



Teacher-forcing

