

Output

Multi-Head Self-Attention



Add & Norm



Feed-Forward Network



Add & Norm



Input (Q , K , V)

