Input to the cell

H

key value query

dk_1 x A_1   dv_1 x A_1   dk_1 x A_1

Multi-head Attention

dv_1 x A_1

proj

H

Add & Layer norm

+

H

FFN_1

F_1

FFN_2

H

Add & Layer norm

+

H

Output of the cell