```
I [[ 6.,  1.,  0.,  3.,  4.,  0.],
        [-1.,  2.,  0.,  6.,  0.,  1.],
        [ 2., -3.,  4.,  0.,  1.,  2.]]
W_K [[-1.,  0.,  3.,  2.],
        [ 2.,  1., -4.,  0.],
        [ 3.,  0.,  2.,  1.],
        [ 4.,  1.,  6.,  0.],
        [-5.,  0., -4.,  6.],
        [ 6.,  1.,  1.,  1.]]
W_Q [[ 2.,  0.,  6., -4.],
        [ 1., -6.,  5.,  2.],
        [ 2., -3.,  1.,  6.],
        [ 1., -1.,  0.,  5.],
        [ 2.,  6.,  2., -1.],
        [ 1.,  1.,  0.,  0.]]
W_V [[-3.,  4.,  1.,  0.],
        [ 2., -6.,  2.,  2.],
        [ 1.,  0., -1.,  0.],
        [ 1.,  1.,  1.,  1.],
        [ 0.,  1.,  0.,  0.],
        [ 3.,  2.,  2., -2.]]

Keys [[-12.,   4.,  16.,  36.],
        [ 35.,   9.,  26.,  -1.],
        [ 11.,  -1.,  24.,  16.]]
Queries [[ 24.,  15.,  49., -11.],
        [  7., -17.,   4.,  38.],
        [ 13.,  14.,   3.,   9.]]
Values [[-13.,  25.,  11.,   5.],
        [ 16.,  -8.,  11.,   8.],
        [ -2.,  31.,  -4., -10.]]

attention before softmax [[ 160., 2260., 1249.],
        [1280.,  158.,  798.],
        [ 272.,  650.,  345.]]

attention after softmax [[0., 1., 0.],
        [1., 0., 0.],
        [0., 1., 0.]]

Attention applied to V (the output [[ 16.,  -8.,  11.,   8.],
        [-13.,  25.,  11.,   5.],
        [ 16.,  -8.,  11.,   8.]]
```