

Bing

Yes, you are correct. The 2x2x2 cube of [0.1, 0.2], [0.3, 0.4]],

[[0.5, 0.6],
[0.7, 0.8]] represents a spatial feature of the first channel of the input tensor x. The 1x2 row of [0.6, 0.7] is the attention that corresponds