

Input (5×5)

Output (3×3)

Kernel (3×3)

Stride = 1, Padding = 0

0	2	0	1	3
4	-4	2	0	1
0	0	1	3	2
2	0	1	0	
1	2	2	0	1

Note: Shown as cross-correlation (kernel not flipped), which is how most deep learning libraries implement "convolution." Flip K if you need the strict signal-processing definition.

$$y_{1,1} = \sum_{i=1}^3 \sum_{j=1}^3 A_{i,j} K_{i,j} = 5$$

(element-wise multiply over the blue region and sum)

1,2,2,0,1