

Classroom Activity: One CNN Step by Hand

Question with worked solution

Question

A grayscale image is represented by the following input matrix. A CNN uses the filter shown below.

Input image I

1	2	0
0	1	3
2	1	1

Filter / kernel K

1	0
0	1

Settings: padding $p = 1$; convolution stride $s = 1$; max pooling window 2×2 ; pooling stride $= 2$.

Tasks: add padding, determine convolution output size, calculate the feature map, apply max pooling, and state the final output.

Worked Solution

Step 1 - Add padding

Because $p = 1$, add one layer of zeros around the original 3×3 image:

0	0	0	0	0
0	1	2	0	0
0	0	1	3	0
0	2	1	1	0
0	0	0	0	0

Step 2 - Calculate convolution output size

Use: $(N - F + 2P) / S + 1$. Here $N = 3$, $F = 2$, $P = 1$ and $S = 1$.

$$(3 - 2 + 2(1)) / 1 + 1 = 4$$

Therefore, the convolution feature map is 4×4 .

Step 3 - Perform convolution

At each position, multiply the 2×2 image patch element-by-element by the kernel, then add the four products.

Output (1,1)

$$(0 \times 1) + (0 \times 0) + (0 \times 0) + (1 \times 1) = 1$$

0	0
0	1

Output (1,2)

$$(0 \times 1) + (0 \times 0) + (1 \times 0) + (2 \times 1) = 2$$

0	0
1	2

Output (2,2)

$$(1 \times 1) + (2 \times 0) + (0 \times 0) + (1 \times 1) = 2$$

1	2
0	1

Output (2,3)

$$(2 \times 1) + (0 \times 0) + (1 \times 0) + (3 \times 1) = 5$$

2	0
1	3

Repeating the same process at every valid position gives the full feature map:

1	2	0	0
0	2	5	0
1	1	2	3
2	1	1	1

Step 4 - Apply 2×2 max pooling

1	2
0	2

max = 2

1	1
2	1

max = 2

0	0
5	0

max = 5

2	3
1	1

max = 3

Final pooled output

2	5
2	3

Final answer: Feature map = 4 x 4; pooled output = 2 x 2.