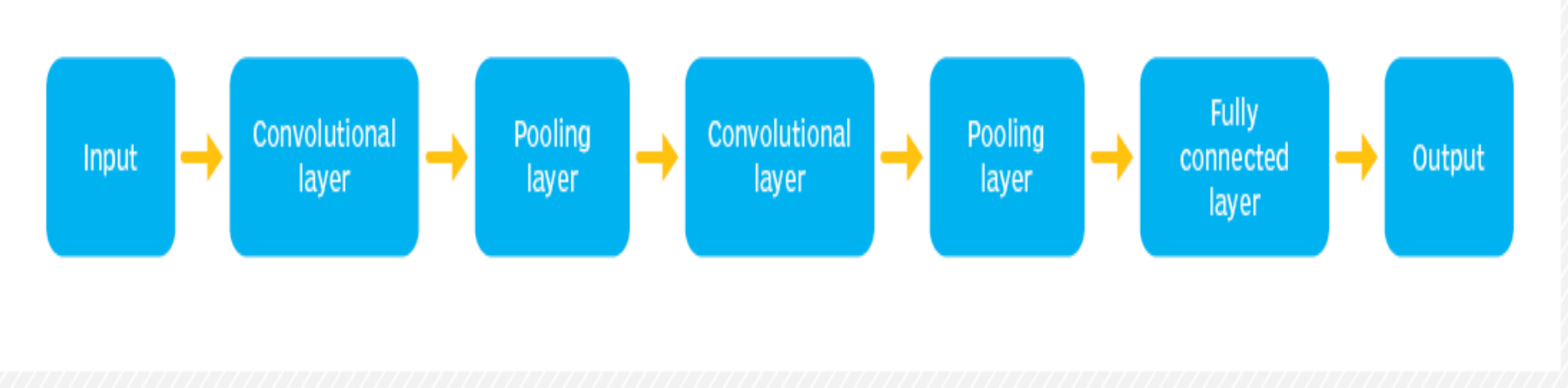
Convolutional Neural Networks

Architecture of a CNN



Convolutional layer

- This layer uses a filter or kernel -- a small matrix of weights -- to move across the receptive field of an input image to detect the presence of specific features.
- The process begins by sliding the kernel over the image's width and height, eventually sweeping across the entire image over multiple iterations

Convolutional layer Operations

- Let's say we have a 5x5 input image **X**
- And a 3x3 filter **W**:

```
1  2  3  4  5
6  7  8  9  10
11 12 13 14 15
16 17 18 19 20
21 22 23 24 25
```

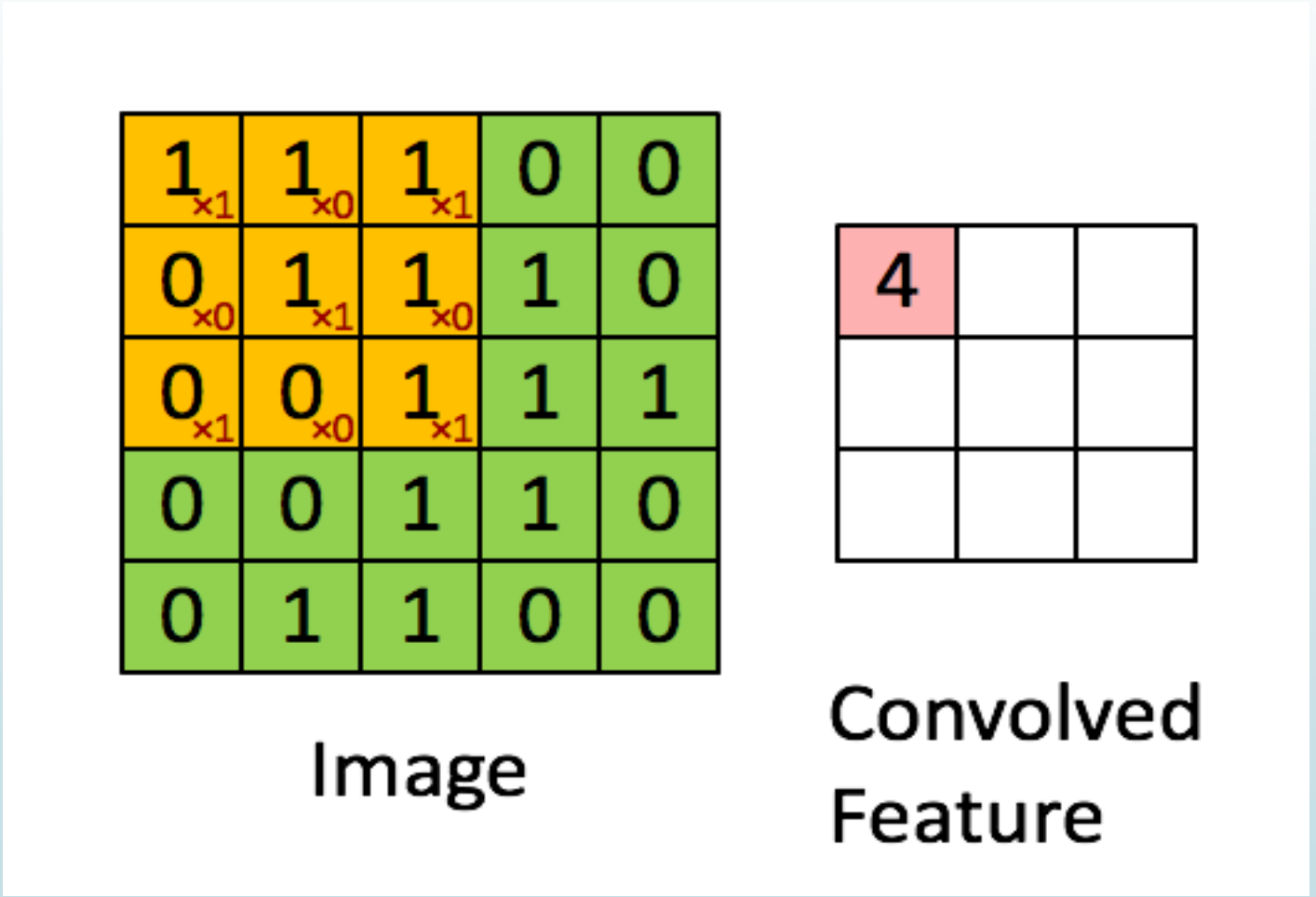
```
-1 0 1
-1 0 1
-1 0 1
```

Convolutional layer Operations

- With stride **s** = 1 and padding **p** = 0, the output feature map **Z** would be 3x3

```
-8 -10 -12
-13 -15 -17
-18 -20 -22
```

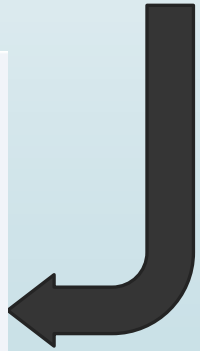
Convolutional layer Operations



Problem in Convolutional Operations

```
1  2  3  4  5
6  7  8  9  10
11 12 13 14 15
16 17 18 19 20
21 22 23 24 25
```

```
-1 0 1
-1 0 1
-1 0 1
```



```
-8  -10 -12
-13 -15 -17
-18 -20 -22
```

Padding

- Without padding, the output feature map of a convolutional layer will be smaller than the input image
- Padding adds extra pixels (usually with value 0) around the borders of the input image

image

	0	0	0	0	0	0	
	0	0	0	0	0	0	
	0	0	0	0	0	0	
	1	1	1	1	1	1	
	1	1	1	1	1	1	
	1	1	1	1	1	1	

*

kernel

-1	-1	-1
0	0	0
1	1	1

=

feature map

0	0	0	0	0	0
0	0	0	0	0	0
1	1	1	1	1	1
1	1	1	1	1	1
0	0	0	0	0	0
0	0	0	0	0	0

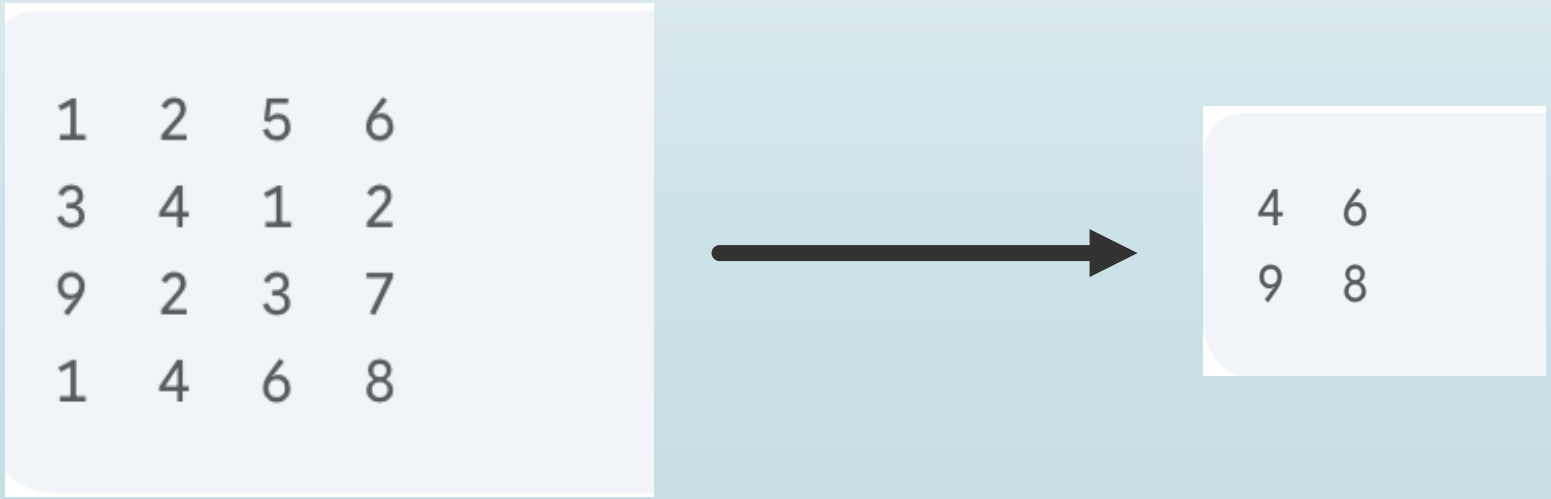
(6x6)

Max pooling

- Max pooling is a downsampling operation that reduces the size of the feature maps generated by convolutional layers

Let's say we have a 4x4 feature map

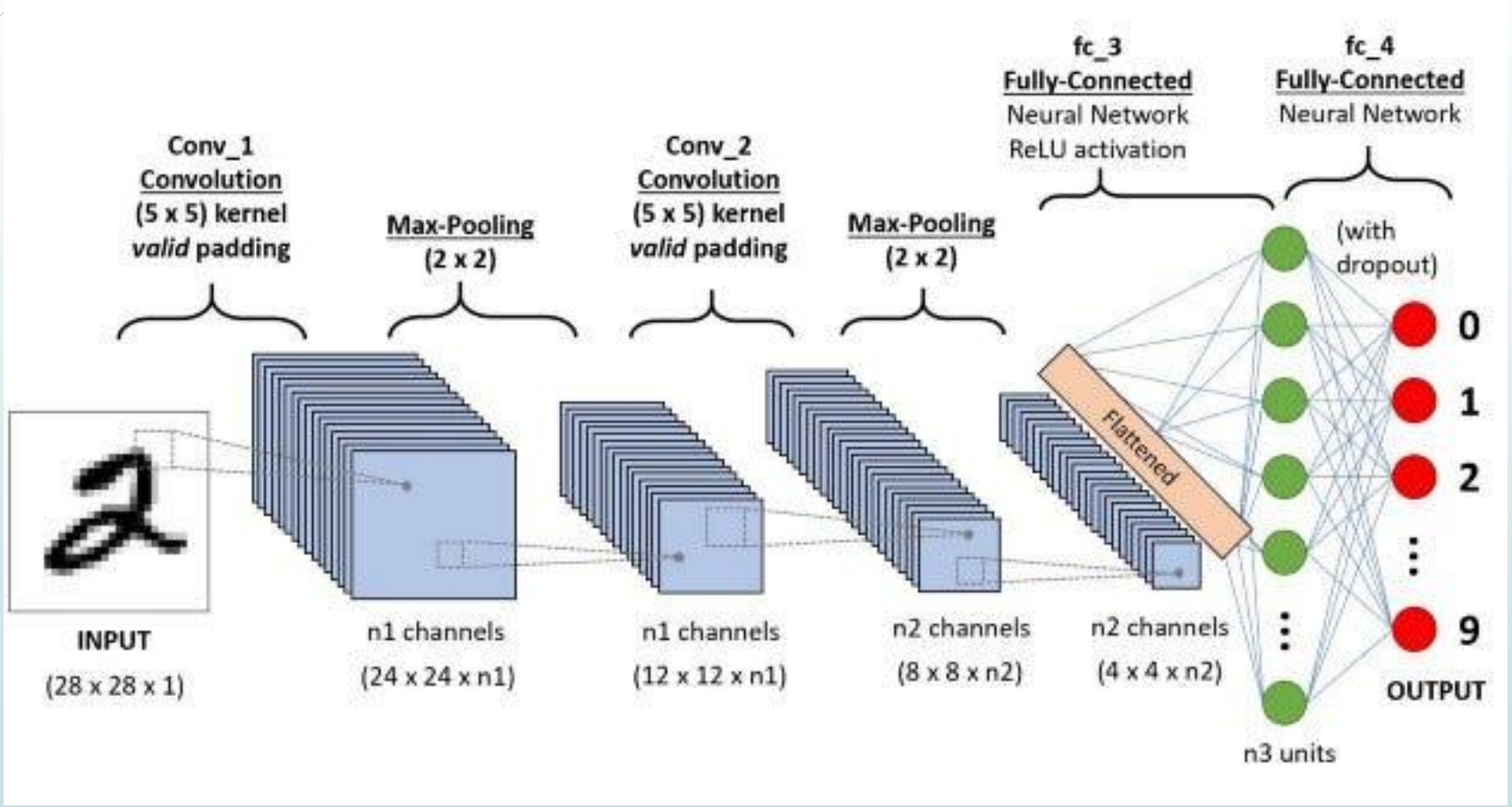
With a 2x2 pooling window and a stride of 2, the max pooling operation would produce the following 2x2 output



Why Max Pooling

1. Dimensionality Reduction
2. Spatial Invariance
3. Feature Emphasis

Fully Connected Layer



https://vijay-choubey.medium.com/understanding-convulutional-neural-networks-9b0cbd9b3055