Vision : CNN [Convolution, padding, stride]

- 2D Convolution

An image is represented as a matrix of numbers.
A filter (kernel) is a small matrix that slides over the image.

At each position, convolution computes:

$$\text{output}(i, j) = \sum_{m} \sum_{n} \text{image}(i+m, j+n) \cdot \text{Kernel}(m, n)$$

This means:
- multiply corresponding values
- add them together
- store the result in the feature map.

- Padding

Padding adds extra pixels (usually zeros) around the image.

If:
- input size = N
- filter size = F
- padding = P
- stride = S

Then output size is:

$$\boxed{\text{Output size} = \frac{N - F + 2P}{S} + 1}$$

Padding helps control output size and keeps edge information.

- Stride

stride defines how far the filter moves each step.
- stride = 1 → detailed feature map
- stride > 1 → downsampled output

mathematically, we skip pixels by stepping S units each time.

- Key Idea

Convolution transforms an image into feature maps by applying learned filters that detect patterns like edges and textures.