Understanding the Convolution Operation in CNNs

1. The Convolution Formula

The mathematical operation used in Convolutional Neural Networks (CNNs) is called **convolution**. It combines two functions or matrices:

- The **input image** (I), and
- The **filter or kernel** (K).

The general formula is:

$$(I * K)(x, y) = \sum_m \sum_n I(x + m, y + n) \cdot K(m, n)$$

This formula tells us how to calculate the new value of a single pixel in the output (called a **feature map**) at position (x, y).

2. What Each Term Means

- $I(x, y)$ = pixel value of the input image at position (x, y)
- $K(m, n)$ = value inside the kernel or filter at position (m, n)
- $(I * K)(x, y)$ = output value (feature map) after applying the kernel
- The two summations $\sum_m \sum_n$ mean that we add up the multiplication of image and kernel values for all pixels in the kernel area.

In simple words: We place the small filter (kernel) on top of the image, multiply each overlapping pixel, add them together, and that becomes one pixel in the new output image.

3. Step-by-Step Example

Let us take a small 3×3 input image and a 2×2 kernel to see how convolution works.

Input Image (I):

$$I = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Kernel (K):

$$K = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

This kernel will help detect changes or edges in the image.

4. Calculating Step by Step

The kernel slides over the image from top-left to bottom-right. At each position, multiply the overlapping values and sum them.

Step 1: Place K on the top-left corner of I:

$$\begin{bmatrix} \boxed{1} & \boxed{2} & 3 \\ \boxed{4} & \boxed{5} & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Multiply element-wise:

$$(1 \times 1) + (2 \times 0) + (4 \times 0) + (5 \times (-1)) = 1 + 0 + 0 - 5 = -4$$

So, output at $(1, 1)$ is -4.

Step 2: Move the kernel one step right:

$$(2 \times 1) + (3 \times 0) + (5 \times 0) + (6 \times (-1)) = 2 + 0 + 0 - 6 = -4$$

Output at $(1, 2)$ is -4.

Step 3: Move kernel down to next row:

$$(4 \times 1) + (5 \times 0) + (7 \times 0) + (8 \times (-1)) = 4 + 0 + 0 - 8 = -4$$

Output at $(2, 1)$ is -4.

Step 4: Bottom-right position:

$$(5 \times 1) + (6 \times 0) + (8 \times 0) + (9 \times (-1)) = 5 + 0 + 0 - 9 = -4$$

Output at $(2, 2)$ is -4.

Final Output Feature Map:

$$(I * K) = \begin{bmatrix} -4 & -4 \\ -4 & -4 \end{bmatrix}$$

5. What Happened Here?

- The convolution captured how the pixel values change in the image.
- Since this kernel is a simple edge detector, it gives similar negative values where the intensity changes.
- If we used a different kernel (for example, a vertical or blur filter), we would get very different outputs.

6. Real-Life Analogy

Imagine you are looking at a photo through a small window (the kernel). You move the window around and write down what you see inside each small region. By doing this everywhere, you create a smaller, summarized version of the whole image — that is exactly what convolution does.

7. Summary

- Convolution is the process of applying a filter (kernel) over an image.
- Each filter extracts a specific feature such as edges, color transitions, or patterns.
- The mathematical operation combines multiplication and addition for each region.
- The output is called a **feature map**, which shows where those features occur in the image.

  In simple terms, convolution allows CNNs to "look" at small parts of an image, detect patterns, and build an understanding of the whole picture step by step.