📌 The Challenge of Images in Neural Networks

Consider a high-resolution grayscale image of size 1000 × 1000 pixels.

- The input layer of a simple Artificial Neural Network (ANN) would require 1 million neurons (one per pixel).
- If the next hidden layer has just 100 neurons, the number of parameters becomes:
  $1{,}000{,}000 \times 100 = 100{,}000{,}000$ (100 million)
- For larger images (say 2000 × 2000 pixels), the parameters quickly grow into billions, making the model computationally expensive, memory-hungry, and prone to overfitting.

👉 This is why using a simple ANN for images is impractical.

📌 Why Convolution is Needed

In an ANN, all pixels are treated as independent inputs. A pixel in the top-left corner and one in the bottom-right corner are just two numbers with no relation. The network does not know they are neighbors in the image grid.

But in reality, images have strong spatial structure:

- Nearby pixels are usually related.
- Patterns such as edges, corners, textures, and shapes emerge only when pixels are considered together in local neighborhoods.
- Higher-level understanding (like recognizing eyes, faces, or objects) comes from combining these local patterns in a spatial hierarchy.

👉 Convolutions solve this problem:

- A filter (kernel) slides across the image, always looking at small local patches of neighboring pixels.
- This ensures that the network learns spatial features - how pixels relate to each other in 2D space.
- By stacking multiple convolution layers, CNNs learn a hierarchy of spatial features:
  - Early layers → edges & textures
  - Middle layers → shapes & parts
  - Later layers → full objects

✅ This makes CNNs far more efficient and powerful than simple ANNs for image data.

📌 Convolution- Convolution is an operation where a filter (kernel) is applied to an input (such as an image) to extract useful features like edges, textures, or patterns.

✒️ CNN - A CNN is a deep learning model designed to process images efficiently. It is widely used in computer vision tasks such as image classification, object detection, facial recognition, and medical image analysis.

📌 Edge Detection Matrix

Edge detection is a fundamental technique used to identify boundaries and object structures within an image. Matrix dot products (convolutions) play a crucial role in detecting edges.

✒️ Vertical Edge Detection

-1	0	1	|	1	0	-1
-1	0	1	|	1	0	-1
-1	0	1	|	1	0	-1

✒️ Horizontal Edge Detection

-1	-1	-1	|	1	1	1
0	0	0	|	0	0	0
1	1	1	|	-1	-1	-1

✒️ Filter/Kernal - A filter (kernel) in a Convolutional Neural Network (CNN) is a small matrix used to extract important features such as edges, textures, and patterns from an image. Filters slide over the input image and perform convolution operations, transforming the input data into a more useful representation for classification, detection, or segmentation.

✒️ Stride - In Convolutional Neural Networks (CNNs), stride refers to the step size at which the filter (kernel) moves across the input image during the convolution operation. It determines how much the filter shifts at each step and influences the output size of the feature map.

📌 ANN vs CNN – Similarities and Differences

🔹 Similarities

- Both are made of neurons that take inputs, apply weights, add bias, and pass through an activation function.
- Both learn weights and biases during training.
- Both are trained using backpropagation and gradient descent.
- Both aim to extract features from data to make predictions.

🔹 Differences

- Connections:
    - ANN → every neuron connects to all inputs.
    - CNN → each neuron connects only to a local region (patch) of the input.

- Parameters:
    - ANN → has a very large number of parameters for images.
    - CNN → uses filters with shared weights, so parameters are far fewer.

- Weights:
    - ANN → each connection has its own unique weight.
    - CNN → the same filter (set of weights) is reused across the whole image.

- Output:
    - ANN → outputs a vector of activations.
    - CNN → outputs a feature map, showing where certain patterns (edges, textures, shapes) are found.

- Spatial Information:
    - ANN → ignores spatial structure, treats all pixels independently.
    - CNN → preserves spatial relationships and builds features hierarchically (edges → shapes → objects).

👉 Key Idea:
Both ANN and CNN work with neurons, weights, and activations, but CNNs are designed for images - they use filters to capture spatial features and output feature maps instead of just flat vectors.

✒️ Dimensionality Reduction - Dimensionality reduction can be understood as a process that reduces the resolution or complexity of an image while preserving important structural information. Blurring is one such technique that reduces dimensionality by smoothing high-frequency details, such as edges and textures, leading to a more simplified representation.

📌 Channel in Image

Channel represents a separate component of an image that stores intensity values for a specific colour or feature. The number of channels in an image depends on its colour model and depth.

✒️ Grayscale - Single-Channel, Pixel values range from 0 (black) to 255 (white) (for 8-bit images).

✒️ RGB - 3 channels, Red (R), Green (G), and Blue (B). Each pixel is a combination of the three colour intensities.

✒️ RGBA - 4 Channels, includes an Alpha (A) channel for transparency. Used in images requiring transparency effects.

✒️ CMYK - 4 Channels, Cyan (C), Magenta (M), Yellow (Y), and Black (K). Used in printing applications.