A **Convolutional Neural Network (CNN)** is a type of [Deep Learning neural network](https://www.geeksforgeeks.org/neural-networks-a-beginners-guide/) architecture commonly used in Computer Vision.

Convolutional Neural Network (CNN) is the extended version of artificial neural networks (ANN) which is predominantly used to extract the feature from the grid-like matrix dataset. For example visual datasets like images or videos where data patterns play an extensive role.

Different types of Neural Networks are used for different purposes, for example for predicting the sequence of words we use [**Recurrent Neural Networks**](https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/) more precisely an [LSTM](https://www.geeksforgeeks.org/understanding-of-lstm-networks/), similarly for image classification we use Convolution Neural networks.

**Neural Networks: Layers and Functionality**

In a regular Neural Network there are three types of layers:

1. **Input Layers:** It’s the layer in which we give input to our model. The number of neurons in this layer is equal to the total number of features in our data (number of pixels in the case of an image).
2. **Hidden Layer:** The input from the Input layer is then fed into the hidden layer. There can be many hidden layers depending on our model and data size. Each hidden layer can have different numbers of neurons which are generally greater than the number of features. The output from each layer is computed by matrix multiplication of the output of the previous layer with learnable weights of that layer and then by the addition of learnable biases followed by activation function which makes the network nonlinear.
3. **Output Layer:** The output from the hidden layer is then fed into a logistic function like sigmoid or softmax which converts the output of each class into the probability score of each class.
4. **CNN Architecture**
5. Convolutional Neural Network consists of multiple layers like the input layer, Convolutional layer, Pooling layer, and fully connected layers.
6. 
7. *Simple CNN architecture*
8. The Convolutional layer applies filters to the input image to extract features, the Pooling layer downsamples the image to reduce computation, and the fully connected layer makes the final prediction. The network learns the optimal filters through backpropagation and gradient descent.

**Advantages and Disadvantages of Convolutional Neural Networks (CNNs)**

**Advantages of CNNs:**

1. Good at detecting patterns and features in images, videos, and audio signals.
2. Robust to translation, rotation, and scaling invariance.
3. End-to-end training, no need for manual feature extraction.
4. Can handle large amounts of data and achieve high accuracy.

**Disadvantages of CNNs:**

1. Computationally expensive to train and require a lot of memory.
2. Can be prone to overfitting if not enough data or proper regularization is used.
3. Requires large amounts of labeled data.
4. Interpretability is limited, it’s hard to understand what the network has learned.

The MNIST DATSETS contains a large collection of handwritten digits that is commonly used for training various image processing systems

**Structure of MNIST dataset**

The MNIST dataset is a collection of 70,000 handwritten digits (0-9), with each image being 28x28 pixels. Here is the dataset information in the specified format:

* **Number of Instances:** 70,000 images
* **Number of Attributes:** 784 (28x28 pixels)
* **Target:** Column represents the digit (0-9) corresponding to the handwritten image
* **Pixel 1-784:** Each pixel value (0-255) represents the grayscale intensity of the corresponding pixel in the image.
* The dataset is divided into two main subsets:
  1. **Training Set:** Consists of 60,000 images along with their labels, commonly used for training machine learning models.
  2. **Test Set:** Contains 10,000 images with their corresponding labels, used for evaluating the performance of trained models

CIFAR-10 Dataset as it suggests has 10 different categories of images in it. There is a total of 60000 images of 10 different classes naming **Airplane**, **Automobile**,**Bird**, **Cat**, **Deer**,**Dog**,**Frog**,**Horse**,**Ship**,**Truck**. All the images are of size 32×32. There are in total 50000 train images and 10000 test images.

**EXP -10**

Regularization is a technique used to prevent overfitting by adding a penalty term to the loss function, discouraging the model from assigning too much importance to individual features or coefficients.

[Dropout](https://www.geeksforgeeks.org/dropout-in-neural-networks/) is a regularization technique which involves randomly ignoring or "dropping out" some layer outputs during training, used in deep [neural networks](https://www.geeksforgeeks.org/neural-networks-a-beginners-guide/) to prevent [overfitting](https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/).

Dropout is implemented per-layer in various types of layers like dense fully connected, convolutional, and recurrent layers, excluding the output layer. The dropout probability specifies the chance of dropping outputs, with different probabilities for input and hidden layers that prevents any one neuron from becoming too specialized or overly dependent on the presence of specific features in the training data.

CIFAR-10 Dataset as it suggests has 10 different categories of images in it. There is a total of 60000 images of 10 different classes naming **Airplane**, **Automobile**,**Bird**, **Cat**, **Deer**,**Dog**,**Frog**,**Horse**,**Ship**,**Truck**. All the images are of size 32×32. There are in total 50000 train images and 10000 test images.

**EXP-11**

The VGG-16 model is a convolutional neural network (CNN) architecture that was proposed by the Visual Geometry Group (VGG) at the University of Oxford. It is characterized by its depth, consisting of 16 layers, including 13 convolutional layers and 3 fully connected layers. VGG-16 is renowned for its simplicity and effectiveness, as well as its ability to achieve strong performance on various computer vision tasks, including image classification and object recognition

"ResNet" stands for "Residual Network" - a type of neural network architecture that utilizes "residual connections" to effectively train very deep networks by allowing information to skip layers, thus overcoming the vanishing gradient problem and enabling better performance with increased network depth

AlexNet is a convolutional neural network (CNN) architecture that uses eight layers to classify images. It was designed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton in 2012.

Layers

* **Convolutional layers**: AlexNet has five convolutional layers.
* **Fully connected hidden layers**: AlexNet has two fully connected hidden layers.
* **Fully connected output layer**: AlexNet has one fully connected output layer.

Activation function

* AlexNet uses the Rectified Linear Unit (ReLU) activation function instead of the sigmoid function. ReLU is less complicated than sigmoid or tanh, which helps speed up learning.