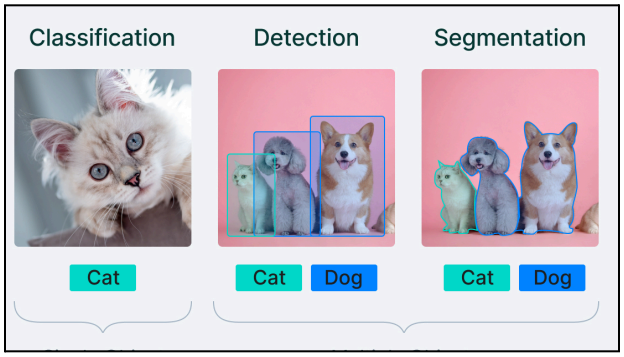
What is Computer Vision?

Computer vision CNN, or convolutional neural network, is a type of deep learning algorithm specifically designed for image and video recognition tasks. It uses a combination of convolutional layers, pooling layers, and fully connected layers to analyze and classify visual data.



Why use CNN over MLP?

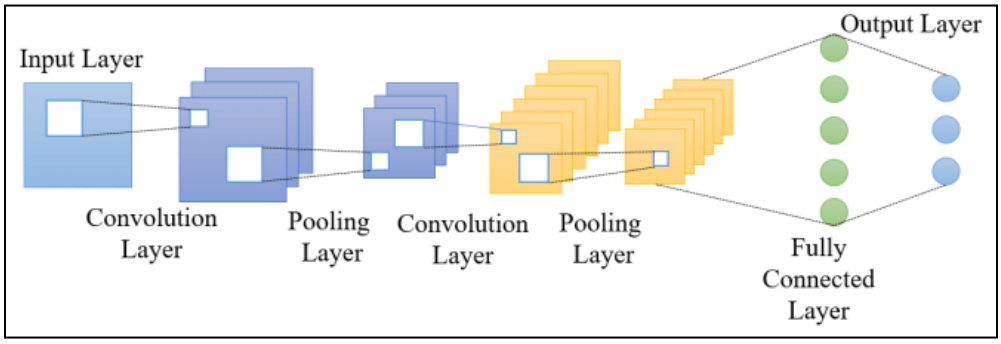
- **Feature extraction:** CNNs are particularly well-suited for computer vision tasks because they can automatically learn and extract meaningful features from images and videos, while traditional machine learning algorithms like MPLs require manual feature engineering.

- **Handling large datasets:** CNNs can handle large datasets and can be trained on a large set of image data, while MPLs are not that efficient in handling large datasets.

- **Handling image variations:** CNNs can handle variations such as different lighting conditions, angles, and rotations, while MPLs can struggle with these variations.

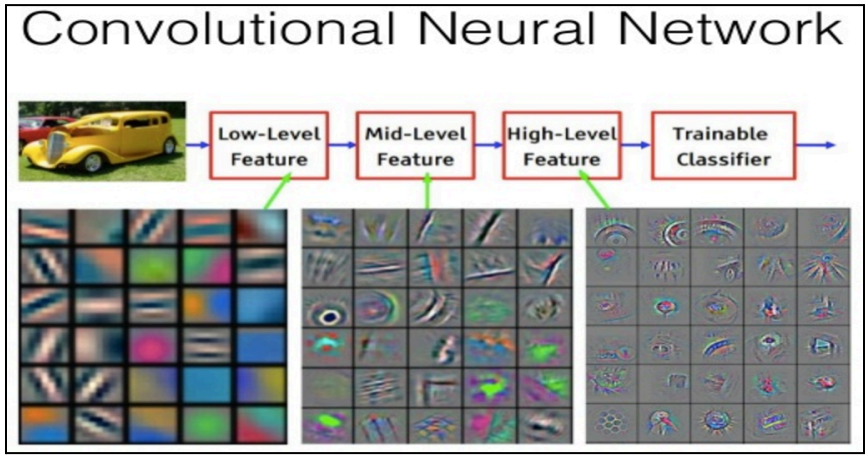
What is a Convolutional Neural Network?

A Convolutional Neural Network (CNN) is a deep learning model used for image and signal processing, where convolutional and pooling operations are applied to input data to extract features and make predictions.



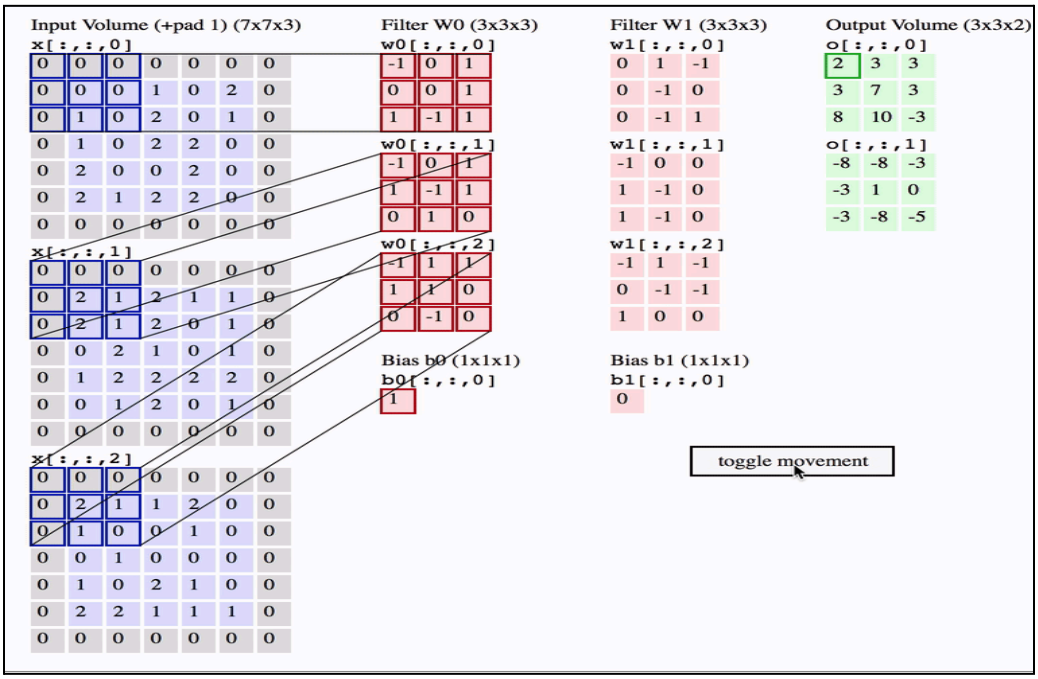
CNN Feature levels

In a Convolutional Neural Network (CNN), features can be classified into different levels based on their complexity and abstraction, such as low-level features (e.g. edges, textures), mid-level features (e.g. parts of objects), and high-level features. (e.g. whole objects, scenes).



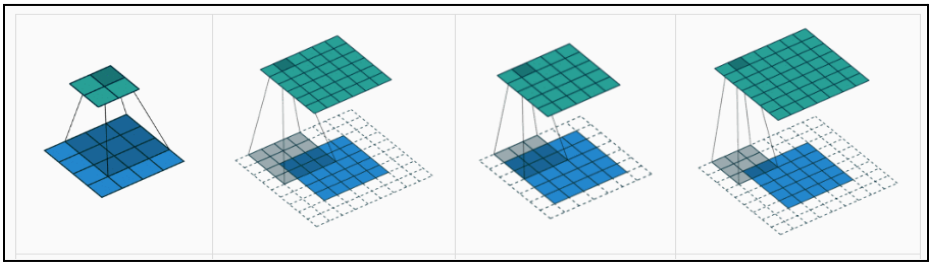
What is a Convolutional operation?

A kernel in a CNN is a small matrix used for performing a convolution operation which involves element-wise multiplication followed by addition and activation (such as ReLU) in order to extract both low-level features in the starting layers and high-level features in the top layers for feature extraction and transformation in convolutional layers.



What is Padding?

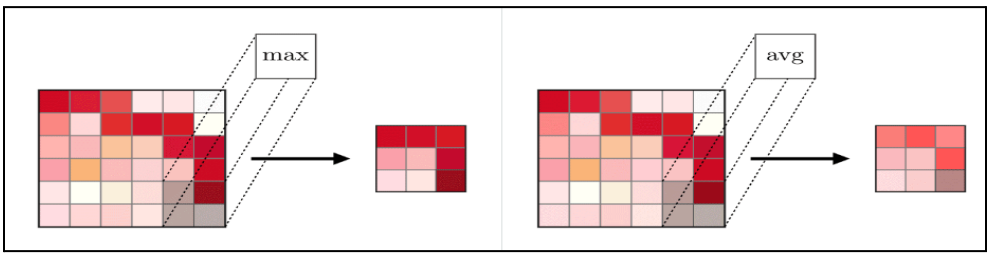
Valid and same padding in CNNs are methods of handling the spatial dimensions of input and output matrices during the convolution operation, with "valid" reducing the output size and "same" preserving the input size in the output.



What is Pooling?

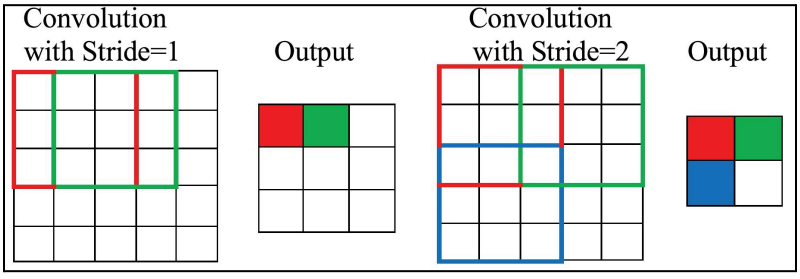
Pooling layers provide an approach to down-sampling feature maps by summarizing the presence of features in patches of the feature map. This has the effect of making the resulting down-sampled feature maps more robust to changes in the position of the feature in the image, referred to by the technical phrase “local translation invariance.”

Max- Pooling - Each pooling operation selects the maximum value of the current view.
Average-pooling - Each pooling operation averages the values of the current view



What is stride?

In Convolutional Neural Networks (CNNs), strides are the number of pixels the convolution filter moves when scanning over the input, determining the size of the output feature map.



Shape of the output layer

The shape of the output in a Convolutional Neural Network (CNN) is determined by the size of the input, the size of the filters used in the convolutional layers, the stride of the convolutional operations, and the padding applied to the input.

The formula for the shape of the output layer in a CNN is determined by the input shape, the kernel size, the stride and the padding

$$output = \frac{input - kernel\_size + 2 * padding}{stride} + 1$$

Number of Parameters

For Convolution layer with k filters of size f x f and the input channel of image c:
Total Parameter = k x (f x f x c_in) + k

For a Fully Connected layer with n neurons and number of input I:
Total Parameter = n x (I + 1)

By calculating the number of parameters in each layer and summing them up, you can obtain the total number of parameters in the entire CNN.