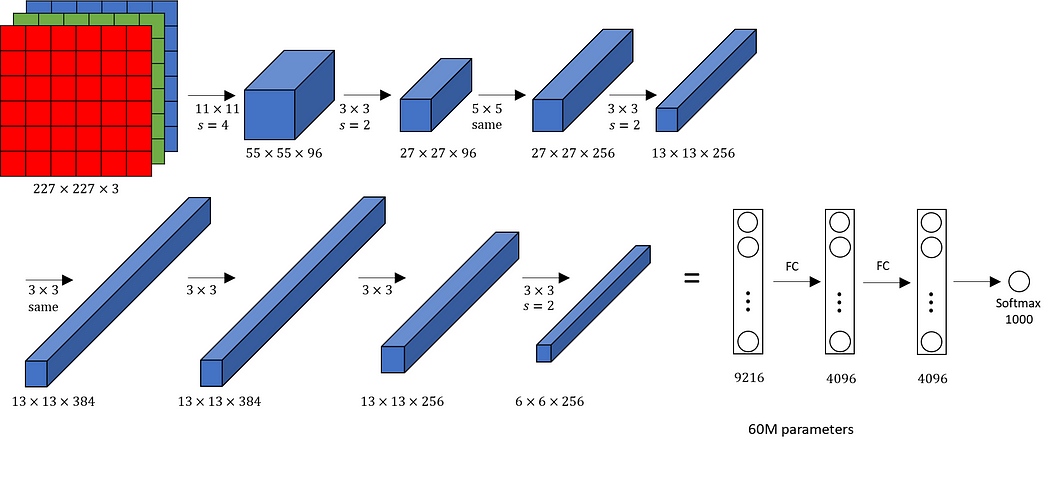
Deep Learning Assignment-1

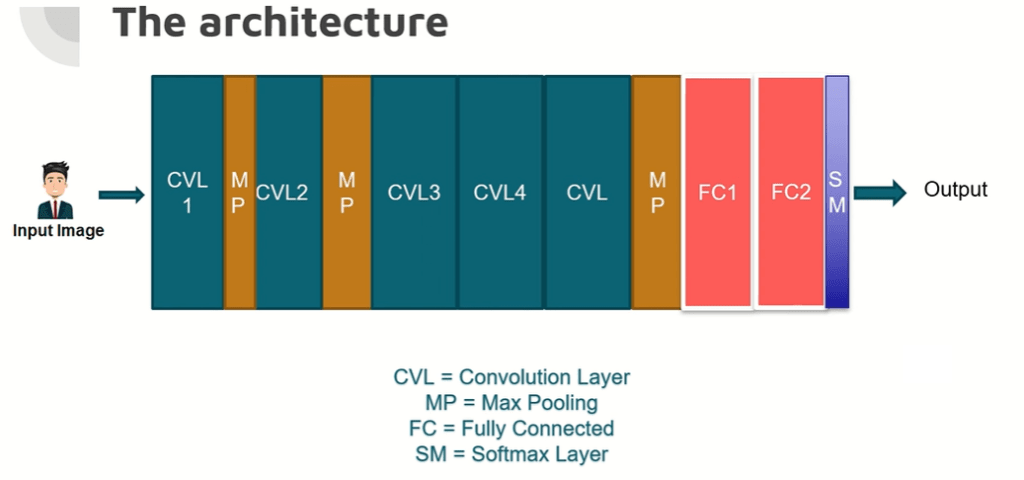
**AlexNet:**

* Alexnet is the name given to a [**Convolutional Neural Network**](https://www.analyticsvidhya.com/blog/2018/12/guide-convolutional-neural-network-cnn/)**Architecture**that won the LSVRC competition in **2012**.
* LSVRC**(**Large Scale Visual Recognition Challenge) is a competition where research teams evaluate their algorithms on a huge dataset of labeled images (**ImageNet**) and compete to achieve higher accuracy on several visual recognition tasks.
* This made a huge impact on how teams approach the completion afterward.

**Alexnet Architecture:**



* The AlexNet architecture consists of eight layers in total.
* The first five layers are convolutional layer.
* The sizes of the convolutional filters are 11×11, 5×5, 3×3, 3×3, and 3×3 for the respective convolutional layers.
* Some of the convolutional layers are followed by max-pooling layers, which help reduce spatial dimensions while retaining important features.
* The activation function used in the network is the Rectified Linear Unit (ReLU), known for its superior performance compared to sigmoid and tanh functions.
* After the convolutional layers, there are three fully connected layers.
* The network’s parameters can be tuned based on the training performance.
* The AlexNet can be used with transfer learning, utilizing pre-trained weights on the ImageNet dataset to achieve exceptional performance.



**Key components of AlexNet Architecture**

**1.Convolutional Neural Network (CNN):**

* AlexNet is a deep Convolutional Neural Network (CNN) architecture designed for image classification tasks. CNNs are specifically suited for visual recognition tasks, leveraging convolutional layers to learn features from images hierarchically.

**2.Architecture:**

* AlexNet consists of eight layers, with the first five being convolutional layers and the last three being fully connected layers. The convolutional layers are designed to extract relevant patterns and features from input images, while the fully connected layers perform the classification based on those features.

**3.ReLU Activation:**

* [Rectified Linear Unit (ReLU)](https://www.nomidl.com/deep-learning/what-is-relu-and-sigmoid-activation-function/) activation functions are used after each convolutional and fully connected layer. ReLU introduces non-linearity, enabling the network to model more complex relationships in the data.

**4.Max Pooling:**

* Max pooling layers are applied after certain convolutional layers to reduce spatial dimensions while retaining essential features. This downsampling process helps reduce computation and controls overfitting.

**5.Local Response Normalization:**

* Local Response Normalization (LRN) is implemented to enhance generalization by normalizing the output of a neuron relative to its neighbours. This creates a form of lateral inhibition, making the network more robust to variations in input data.

**6.Dropout:**

* AlexNet uses dropout regularization during training, where random neurons are dropped out during forward and backward passes. This technique prevents overfitting and improves the model’s generalization performance.

**7. Batch Normalization:**

* Batch Normalization is applied to normalize the outputs of each layer within a mini-batch during training. It stabilizes and accelerates the training process, allowing for higher learning rates and deeper architectures.

**8.Softmax Activation:**

* The final layer of AlexNet uses the softmax activation function to convert the model’s raw output into class probabilities. This allows the network to provide a probability distribution over the possible classes for each input image.

**9.Training and Optimization:**

* AlexNet is trained using stochastic gradient descent with momentum. The learning rate is adjusted during training, and data augmentation techniques are applied to increase the diversity of the training dataset.

**10.ImageNet Competition:**

* AlexNet achieved significant success when it participated in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. Its superior performance helped popularize deep learning and CNNs for image classification tasks. A group of rectangular objects with numbers and symbols

  Description automatically generated

### **Convolution and Maxpooling Layers**

* Then we apply the first convolution layer with 96 filters of size 11X11 with stride 4. The activation function used in this layer is relu. The output feature map is 55X55X96.
* In case, you are unaware of how to calculate the output size of a convolution layer.

output= ((Input-filter size)/ stride)+1

* Also, the number of filters becomes the channel in the output feature map.
* Next, we have the first Maxpooling layer, of size 3X3 and stride 2. Then we get the resulting feature map with the size 27X27X96.
* After this, we apply the second convolution operation. This time the filter size is reduced to 5X5 and we have 256 such filters. The stride is 1 and padding 2. The activation function used is again relu. Now the output size we get is 27X27X256.
* Again we applied a max-pooling layer of size 3X3 with stride 2. The resulting feature map is of shape 13X13X256.
* Now we apply the third convolution operation with 384 filters of size 3X3 stride 1 and also padding 1. Again the activation function used is relu. The output feature map is of shape 13X13X384.
* Then we have the fourth convolution operation with 384 filters of size 3X3. The stride along with the padding is 1. On top of that activation function used is relu. Now the output size remains unchanged i.e 13X13X384.
* After this, we have the final convolution layer of size  3X3 with 256 such filters. The stride and padding are set to one also the activation function is relu. The resulting feature map is of shape 13X13X256.
* So if you look at the architecture till now, the number of filters is increasing as we are going deeper. Hence it is extracting more features as we move deeper into the architecture. Also, the filter size is reducing, which means the initial filter was larger and as we go ahead the filter size is decreasing, resulting in a decrease in the feature map shape.
* Next, we apply the third max-pooling layer of size 3X3 and stride 2. Resulting in the feature map of the shape 6X6X256.