

## Pooling Layer

The pooling operation involves sliding a two-dimensional filter over each channel of feature map and summarizing the features lying within the region covered by the filter.

For a feature map having dimensions  $n_h \times n_w \times n_c$ , the dimensions of output obtained after a pooling layer is

$$(n_h - f + 1) / s \times (n_w - f + 1) / s \times n_c$$

where,

- $n_h$  - height of feature map
- $n_w$  – width of feature map
- $n_c$  – number of channels in the feature map
- $f$  - size of filter
- $s$ -stride length

A common CNN model architecture is to have a number of convolution and pooling layers stacked one after the other.

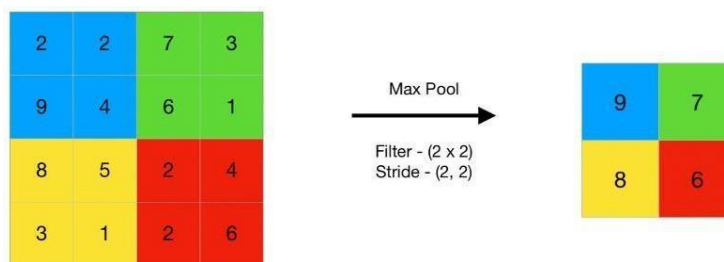
Pooling layers are used to reduce the dimensions of the feature maps. Thus, it reduces the number of parameters to learn and the amount of computation performed in the network.

The pooling layer summarizes the features present in a region of the feature map generated by a convolution layer. So, further operations are performed on summarized features instead of precisely positioned features generated by the convolution layer. This makes the model more robust to variations in the position of the features in the input image.

### Types of Pooling Layers:

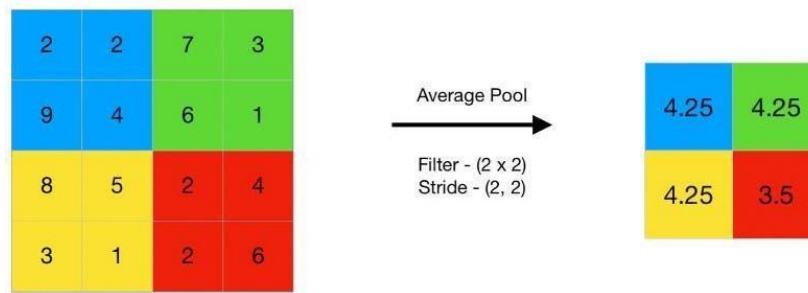
- **Max Pooling**

Max pooling is a pooling operation that selects the maximum element from the region of the feature map covered by the filter. Thus, the output after max-pooling layer would be a feature map containing the most prominent features of the previous feature map.



- **Average Pooling**

Average pooling computes the average of the elements present in the region of feature map covered by the filter. Thus, while max pooling gives the most prominent feature in a particular patch of the feature map, average pooling gives the average of features present in a patch.



- **Global Pooling**

Global pooling reduces each channel in the feature map to a single value. Thus, an  $n_h \times n_w \times n_c$  feature map is reduced to  $1 \times 1 \times n_c$  feature map. This is equivalent to using a filter of dimensions  $n_h \times n_w$  i.e. the dimensions of the feature map. Further, it can be either global max pooling or global average pooling.

- **Global Average Pooling**

Considering a tensor of shape  $h \times w \times n$ , the output of the Global Average Pooling layer is a single value across  $h \times w$  that summarizes the presence of the feature. Instead of downsizing the patches of the input feature map, the Global Average Pooling layer downsizes the whole  $h \times w$  into 1 value by taking the average.

- **Global Max Pooling**

With the tensor of shape  $h \times w \times n$ , the output of the Global Max Pooling layer is a single value across  $h \times w$  that summarizes the presence of a feature. Instead of downsizing the patches of the input feature map, the Global Max Pooling layer downsizes the whole  $h \times w$  into 1 value by taking the maximum.

In convolutional neural networks (CNNs), the pooling layer is a common type of layer that is typically added after convolutional layers. The pooling layer is used to reduce the spatial dimensions (i.e., the width and height) of the feature maps, while preserving the depth (i.e., the number of channels).

- The pooling layer works by dividing the input feature map into a set of non-overlapping regions, called pooling regions. Each pooling region is then transformed into a single output value, which represents the presence of a particular feature in that region. The most common types of pooling operations are max pooling and average pooling.
- In max pooling, the output value for each pooling region is simply the maximum value of the input values within that region. This has the effect of preserving the most salient features in each pooling region, while discarding less relevant information. Max pooling is often used in CNNs for object recognition tasks, as it helps to identify the most distinctive features of an object, such as its edges and corners.
- In average pooling, the output value for each pooling region is the average of the input values within that region. This has the effect of preserving more information than max pooling, but may also dilute the most salient features. Average pooling is often used in CNNs for tasks such as image segmentation and object detection, where a more fine-grained representation of the input is required.

Pooling layers are typically used in conjunction with convolutional layers in a CNN, with each pooling layer reducing the spatial dimensions of the feature maps, while the convolutional layers extract increasingly complex features from the input. The resulting feature maps are then passed to a fully connected layer, which performs the final classification or regression task.

### Advantages of Pooling Layer

**Dimensionality reduction:** The main advantage of pooling layers is that they help in reducing the spatial dimensions of the feature maps. This reduces the computational cost and also helps in avoiding over-fitting by reducing the number of parameters in the model.

**Translation invariance:** Pooling layers are also useful in achieving translation invariance in the feature maps. This means that the position of an object in the image does not affect the classification result, as the same features are detected regardless of the position of the object.

Feature selection: Pooling layers can also help in selecting the most important features from the input, as max pooling selects the most salient features and average pooling preserves more information.

### **Disadvantages of Pooling Layer**

- Information loss: One of the main disadvantages of pooling layers is that they discard some information from the input feature maps, which can be important for the final classification or regression task.
- Over-smoothing: Pooling layers can also cause over-smoothing of the feature maps, which can result in the loss of some fine-grained details that are important for the final classification or regression task.
- Hyperparameter tuning: Pooling layers also introduce hyperparameters such as the size of the pooling regions and the stride, which need to be tuned in order to achieve optimal performance. This can be time-consuming and require some expertise in model building.

