**CHANNEL ATTENTION**

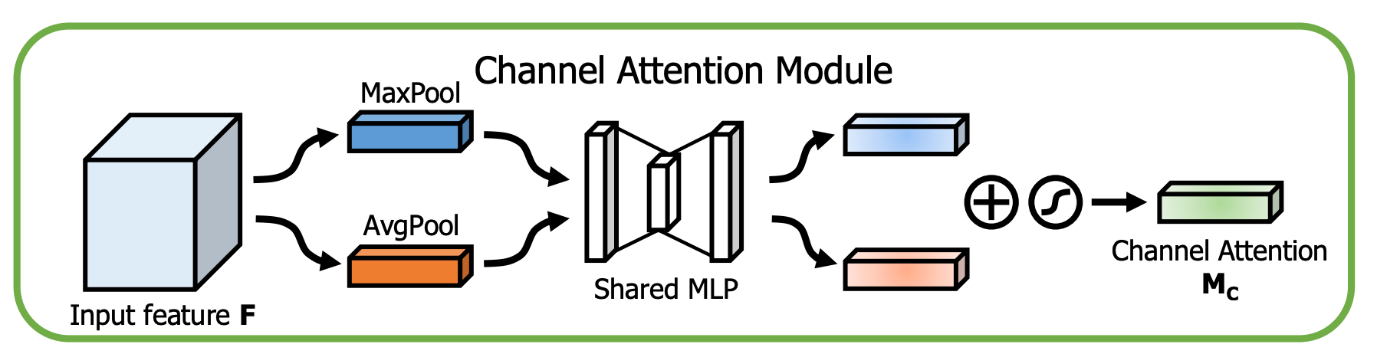
Convolutional Neural Networks (CNNs) use channel attention techniques to highlight the key characteristics in each channel in an effort to maximize network performance. Reweighting the channels according to how relevant they are to the job at hand allows for this.Convolutional Neural Networks (CNNs) employ channel attention to dynamically emphasize the significance of various feature channels, hence improving their performance. Through better feature representation and enhanced generalization to new, unseen data, this method enables the network to concentrate on the most important aspects in the incoming data. Channel attention makes the network dynamically adaptable, more resilient to changes in the data, and more efficient in situations where each channel's priority is determined by the input.

**CHANNEL ATTENTION USED IN CNN**

Neural networks use channel attention techniques to enhance the model's ability to focus on the most important components within each channel of the input data. By giving different weights to different channels, these techniques highlight more informative qualities and suppress less relevant ones. Because of its selective emphasis, the model performs better overall for tasks like object identification, segmentation, and picture recognition. Moreover, channel attention can maximize computational resources, reducing memory use and perhaps speeding up inference. These strategies provide the network greater flexibility in responding to shifting circumstances and allow it to dynamically shift its emphasis in response to incoming input.

**WORKING PROCEDURE**

The working procedure of the channel attention is as follows



**EXPLANATION**

Input Feature: The process starts with input feature maps from the convolutional neural network's (CNN) previous layer. Numerous channels (e.g., several filters applied to the input image) are included in these feature maps.

Pool Management: Max Pooling- Using the spatial dimensions of height and width as a reference, this operation determines the maximum value from each channel. The final product is a vector where each element represents the highest value found in that channel. The average value of every channel across all spatial dimensions is determined by the average pooling operation. The outcome is a vector, where each element denotes the average value for that particular channel.

Fout = Fin⋅σ(MLP(AvgPool=Fin)+MaxPool"(Fin)))

The max-pooled and average-pooled vectors are fed into the same multilayer perceptron, or MLP, in a shared model. A small feed-forward neural network with two fully connected layers is called a shared MLP. It trains a set of weights for each channel using the pooled vectors. First Fully Connected Layer: This layer records inter-channel interactions and reduces dimensionality. ReLU Activation: To introduce non-linear behavior, this non-linear activation function is used after the first fully connected layer. The dimensionality is adjusted in the second fully connected layer to match the number of channels in the input feature map. To guarantee that the attention scores fall between 0 and 1, the second fully connected layer's output is subjected to the sigmoid activation function.

Channel Attention: The two vectors of attention scores (max-pooling and average-pooling) that comprise the shared MLP outputs are added together. The initial input feature map is reweighted using the sum of the attention ratings. The attention score of every channel in the feature map is doubled, emphasizing the most important channels and underplaying the less important ones.

**BENEFITS OF USING CHANNEL ATTENTION IN CNNS**

It highlights the most important characteristics, improving accuracy in tasks like object identification and picture categorization. Based on the input data, they modify the focus, and it returns. enhances task-specific performance and increases resilience to variances. With very few adjustments, it is readily integrated into CNN structures that are already in place.