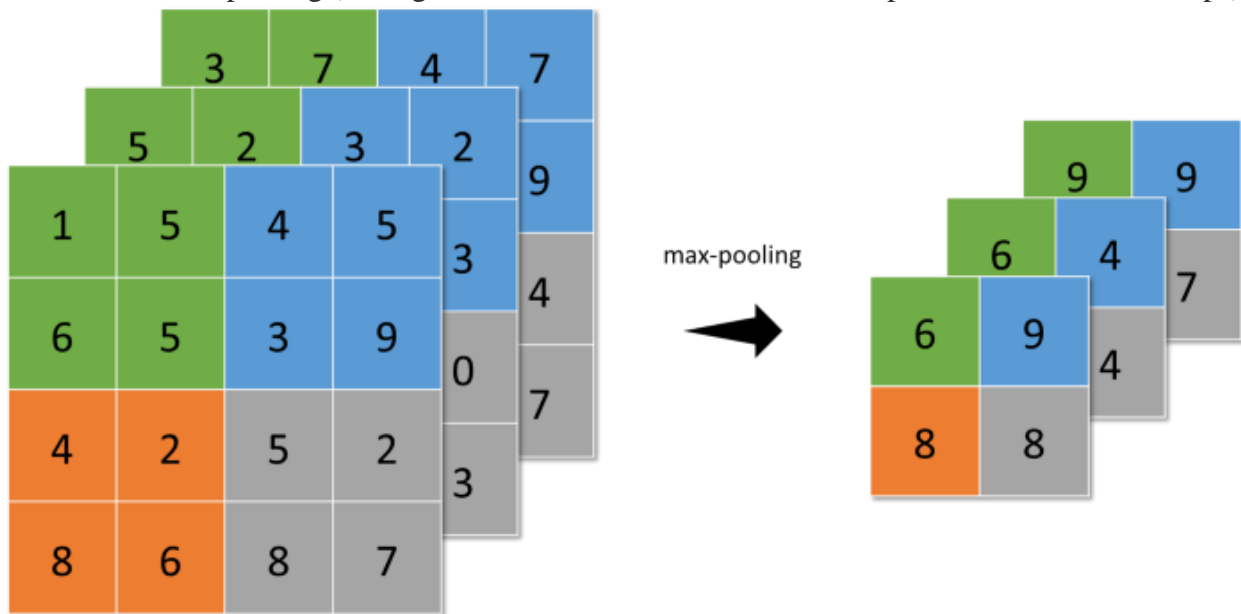POOLING LAYER

CNN allows the neural network to reuse parameters across different spatial locations of an image. Various choices of filters (also called kernels) could achieve different image operations: identity, edge detection, blur, sharpening, etc. The idea of CNN is to discover some interesting features of the image by introducing random matrices as a convolution operator.

After the image has been convolved with kernels, you're left with what's known as a *feature map*. You can convolve your feature maps again and again to extract features, however, this turns out to be incredibly computationally expensive. Instead, you can reduce the size of your feature maps using the *pooling layer*.

The pooling layer represents small regions of a feature map by a single pixel. Some different strategies can be considered:

- max-pooling (the highest value is selected from the N×N patches of the feature maps),

max-pooling with 2×2 patches applied on 3 filters with size 4×4 | Image by author

- average-pooling (sums up over N×N patches of the feature maps from the previous layer and selects the average value), and
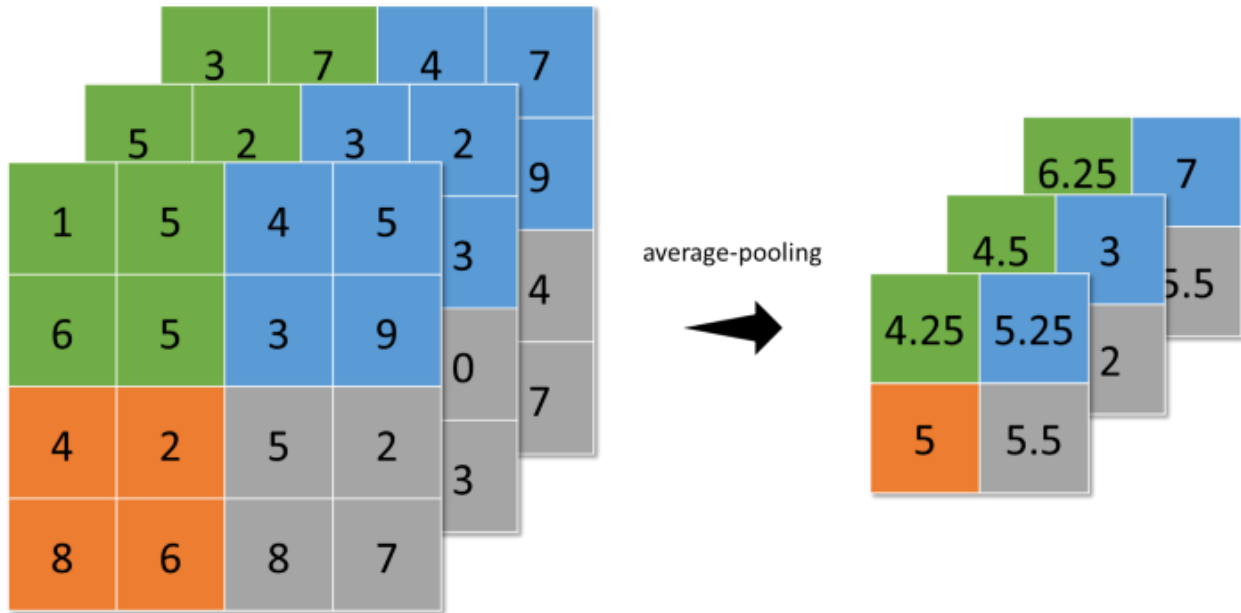
average-pooling with 2×2 patches applied on 3 filters with size 4×4 | Image by author

- global-average-pooling (similar to average-pooling but instead of using N×N patches of the feature maps, it uses all feature maps area at once).
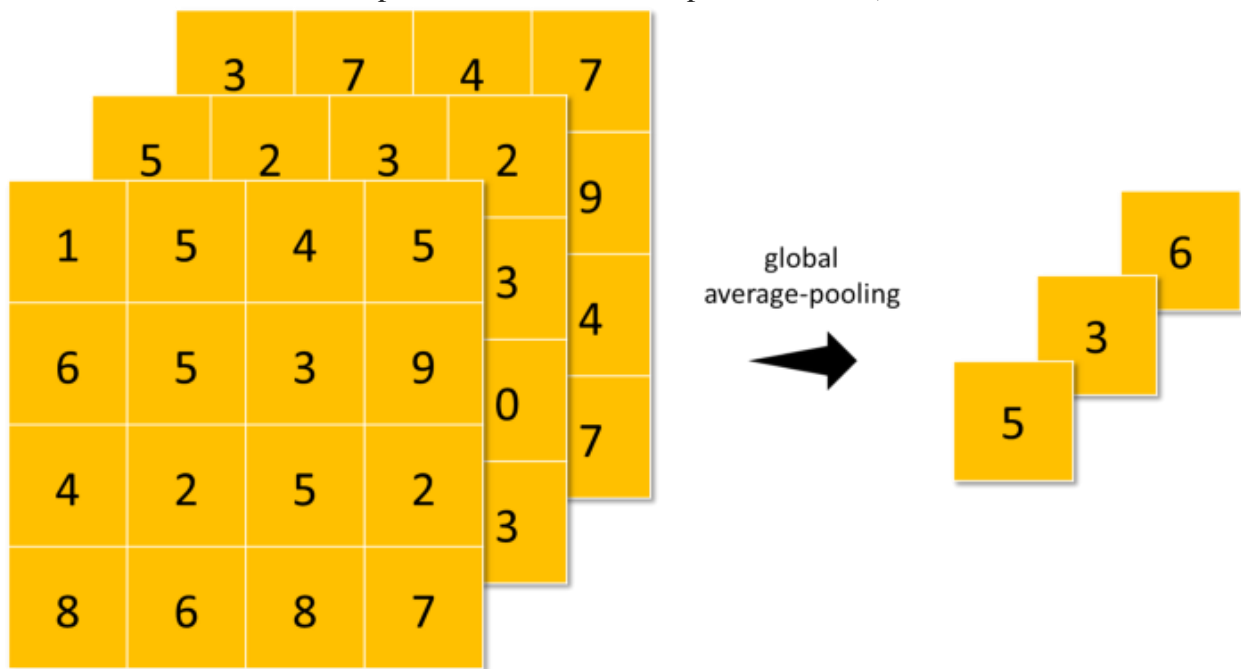
global-average-pooling applied on 3 filters with size 4×4 | Image by author

Pooling makes the neural network invariant to small transformations, distortions, and translations in the input image. A small distortion in the input will not change the outcome of pooling since you take the maximum/average value in a local neighborhood.