

Pooling Layers

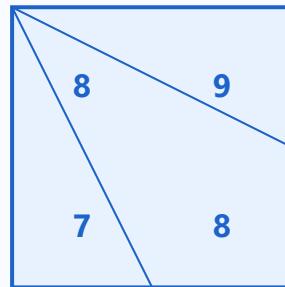
Max vs Average Pooling

Input (4×4)

2	8	4	1
5	3	9	2
7	1	6	4
3	2	5	8

2×2 Pool

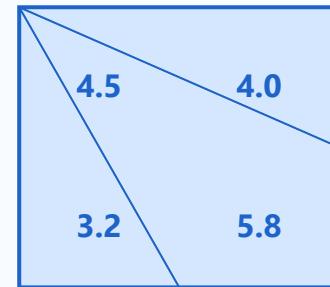
Max Pooling (2×2)



Same Input

4×4 matrix

Average Pooling (2×2)



Max Pooling

Selects maximum value from each window. Preserves strongest activations and provides translation invariance

Average Pooling

Computes average of values in each window. Smoother downsampling, often used before classification layers

Global Pooling

Reduces entire feature map to single value per channel.
Eliminates need for fixed input sizes

Adaptive Pooling

Outputs fixed size regardless of input dimensions.
Automatically adjusts pooling window and stride

Modern Alternatives

Strided convolutions can replace pooling while learning optimal downsampling patterns. Some architectures eliminate pooling entirely