ATML Tutorial 07

Advanced Topics in Machine Learning
31.03.2020
Givi Meishvili

Today

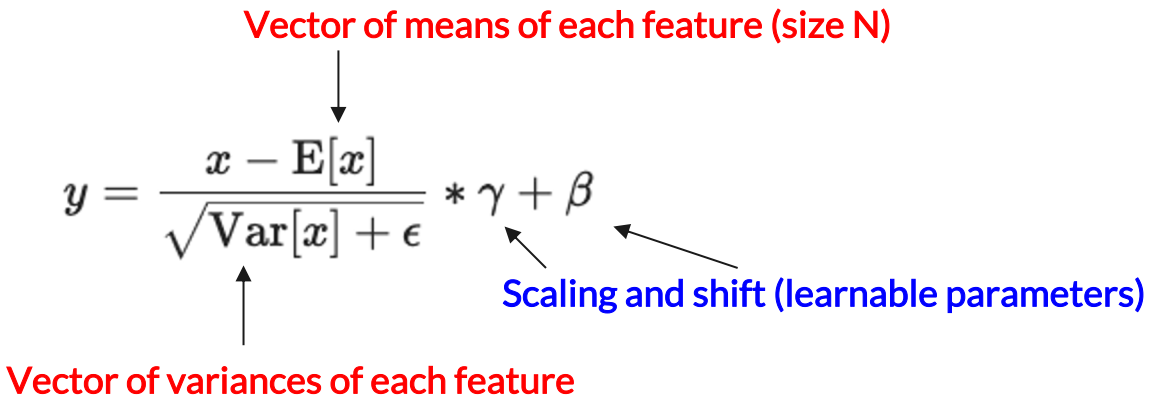
- Batch normalization
- Convolution layers

Batch Normalization

Normalize input to intermediate layers

x - data of size MxN

- M - mini batch size - number of training examples
- N - number of features



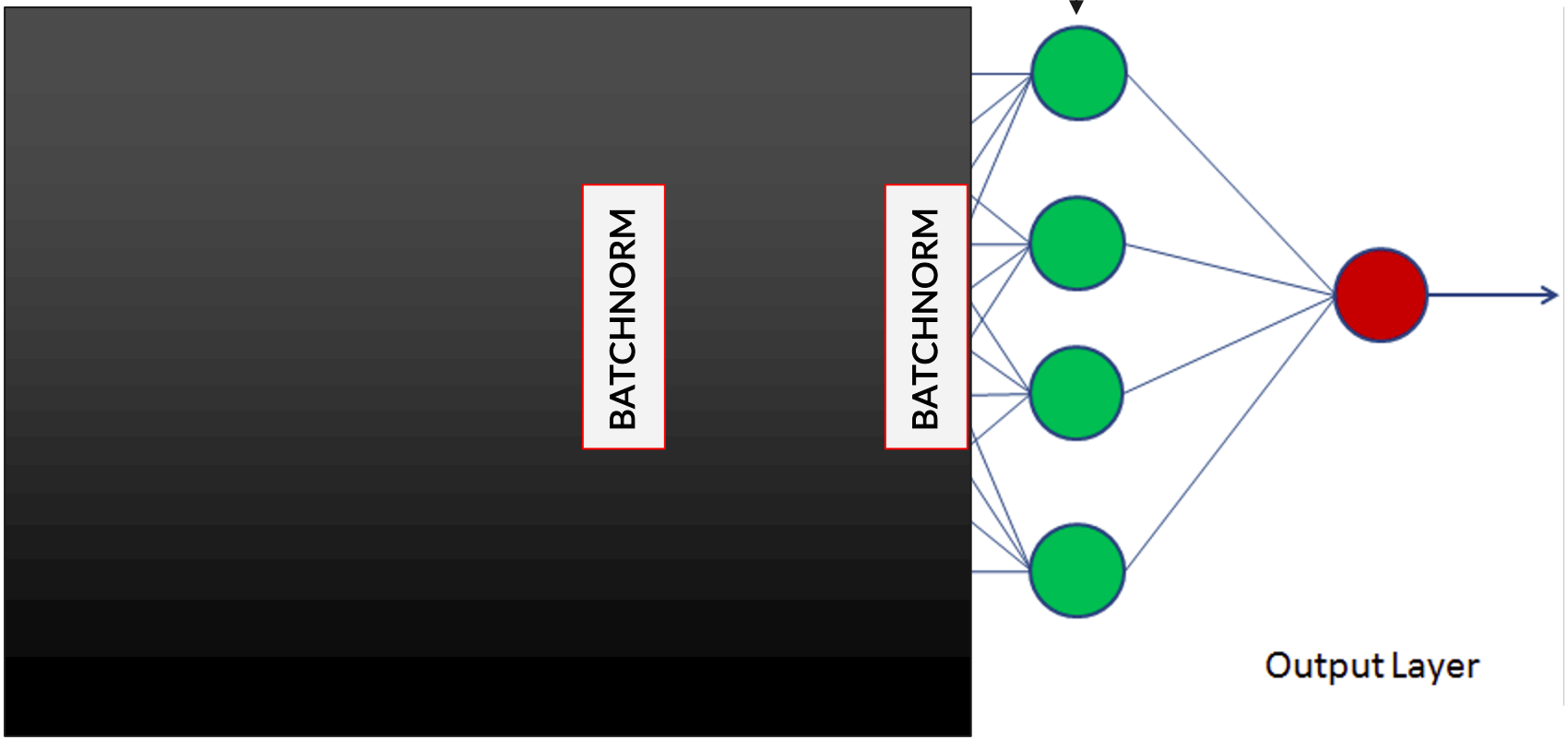
$$y = \frac{x - \mathrm{E}[x]}{\sqrt{\mathrm{Var}[x] + \epsilon}} * \gamma + \beta$$

Batch Normalization

As we update previous layers, input to this layer can be shifted that's not the expected input

Now the input to each layer is normalized shift is reduced layers can learn more "independently"

This layer learns to work for a certain input

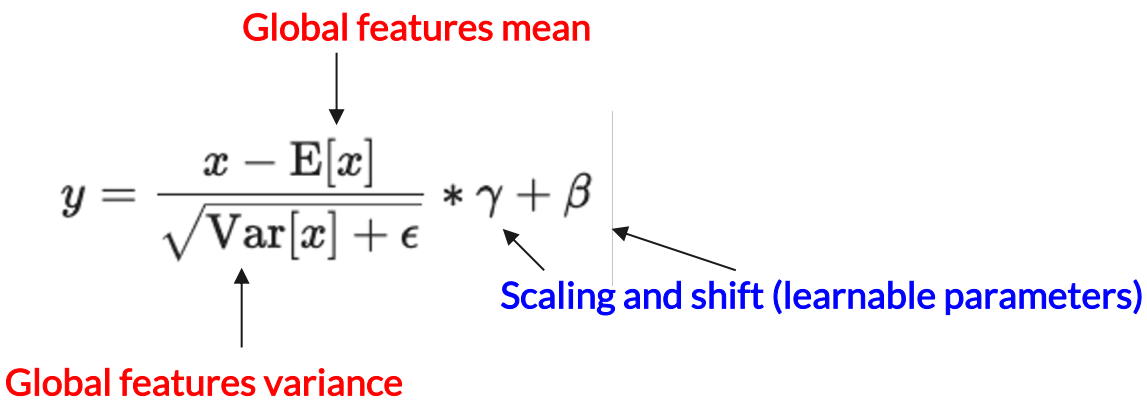


Batch Normalization - inference

In training phase: we compute mean and variance within a mini-batch

In evaluation phase: we want to get the same result regardless of batch size; we don't want any randomness that comes from sampling a mini-batch

Solution: during training, global running mean and variance are computed



Convolutions

Before: MLP - look at the entire image at once
- spatial information is lost
- very sensitive to shifts etc.

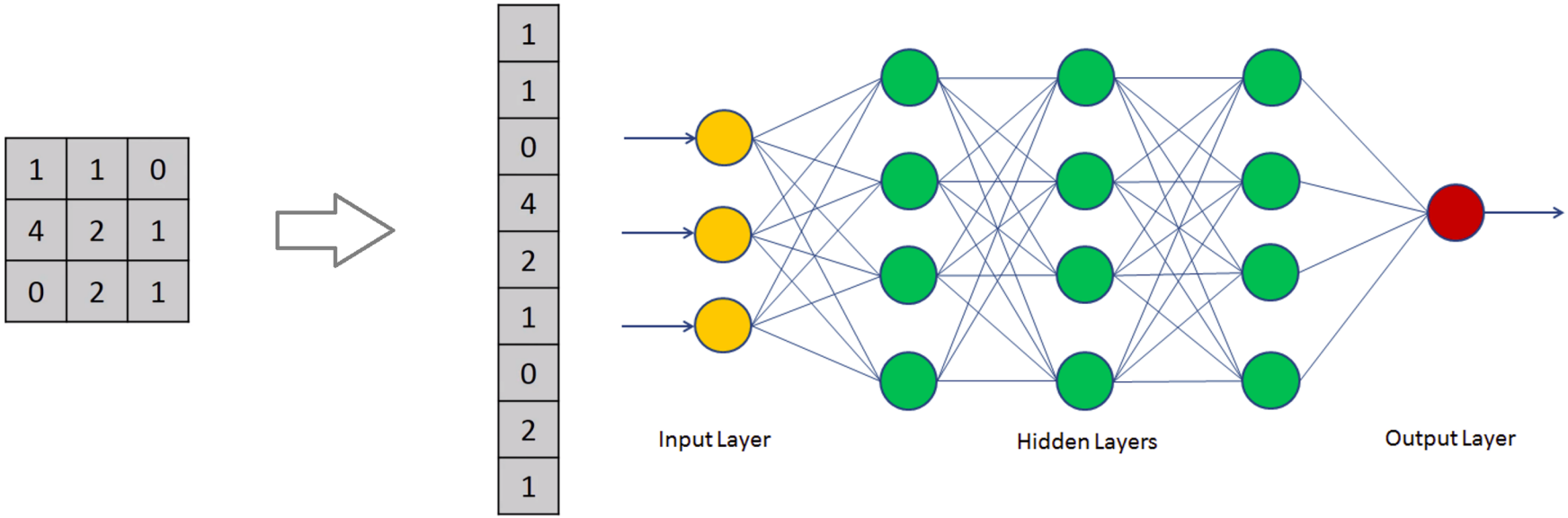


Image **Flattened image** **MLP**

Convolutions

Convolutions: Learn to detect local features
Multiply a small kernel (filter) at every image location

Example: edge detector

Input image

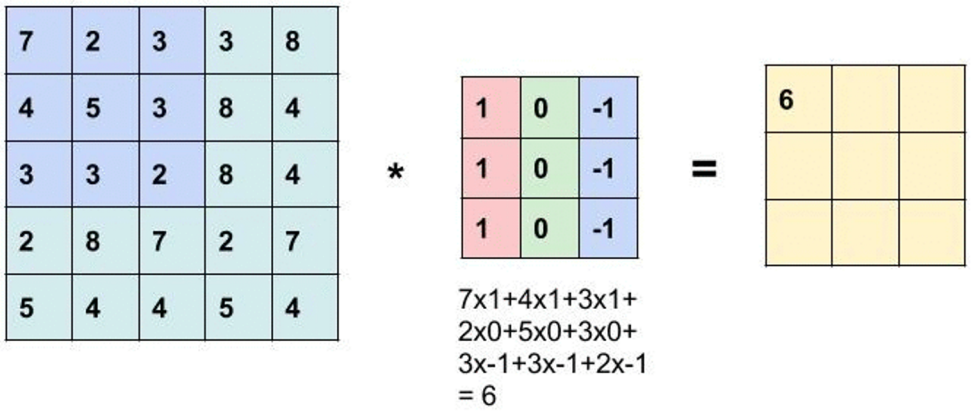
Convolution Kernel

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Feature map

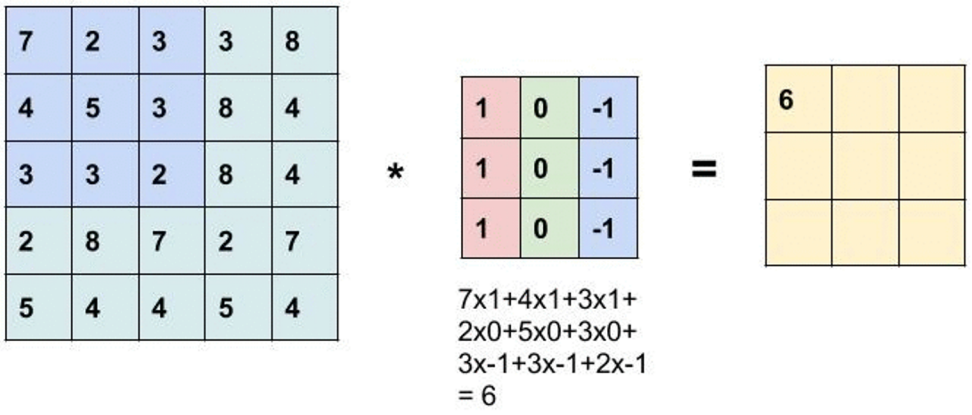
Convolutions

Convolutions: Learn to detect local features



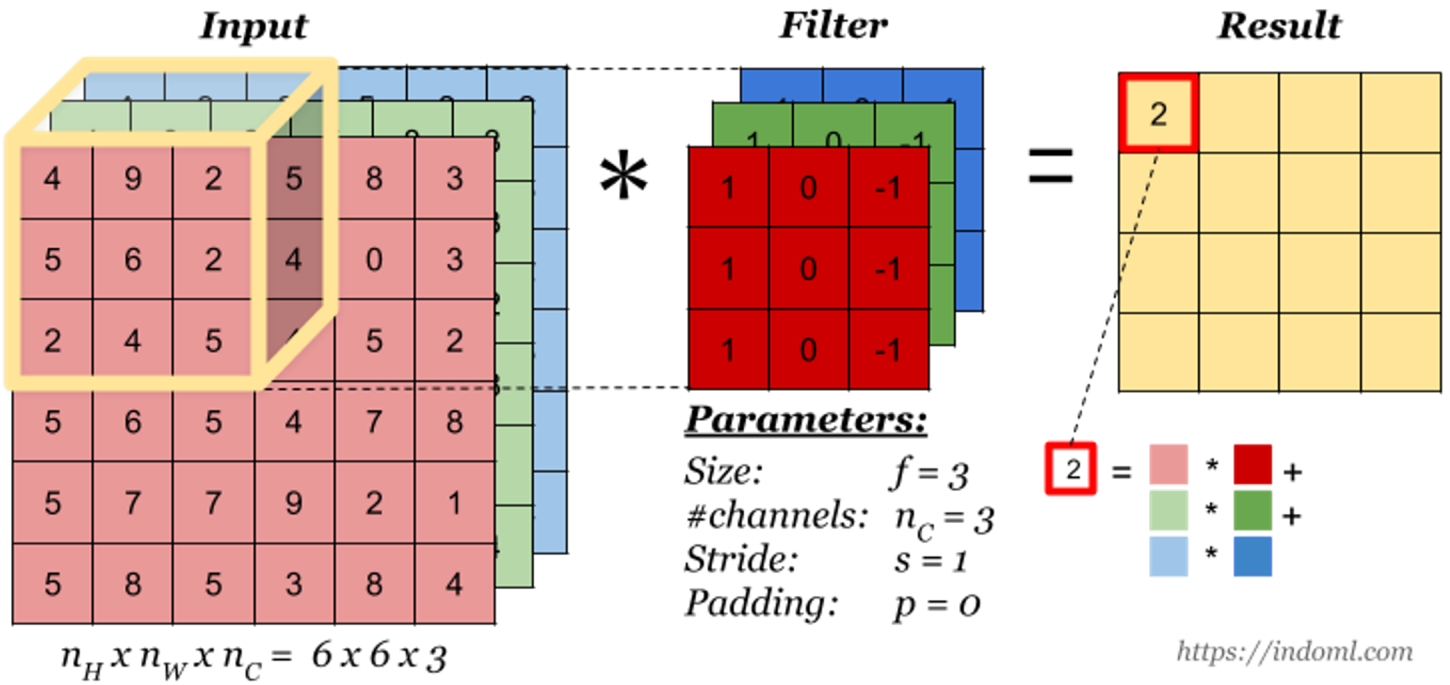
Convolutions

Convolutions: Learn to detect local features



Convolutions

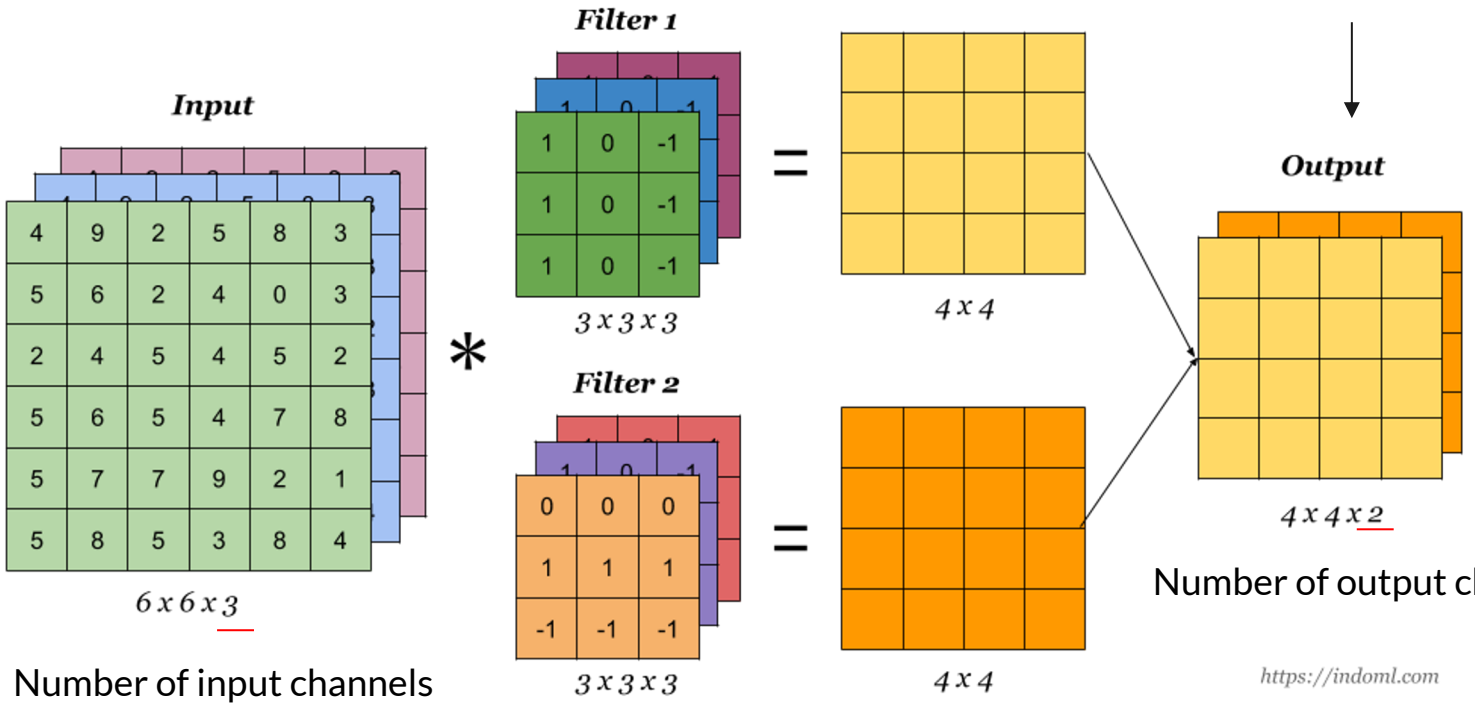
We can apply convolutions on volumes (e.g. RGB image)
Filter depth must match input depth



Convolutions

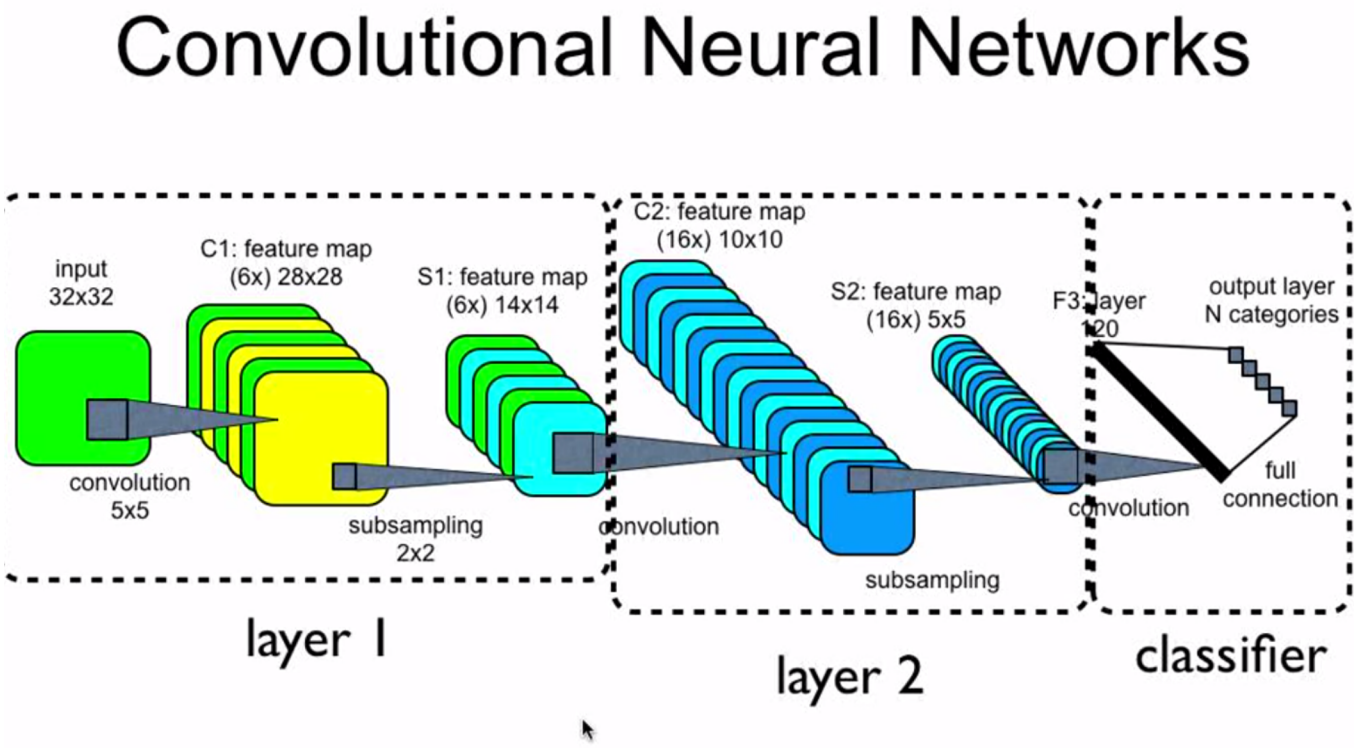
We can use multiple filters

Feature map -
"image" with information
detected features



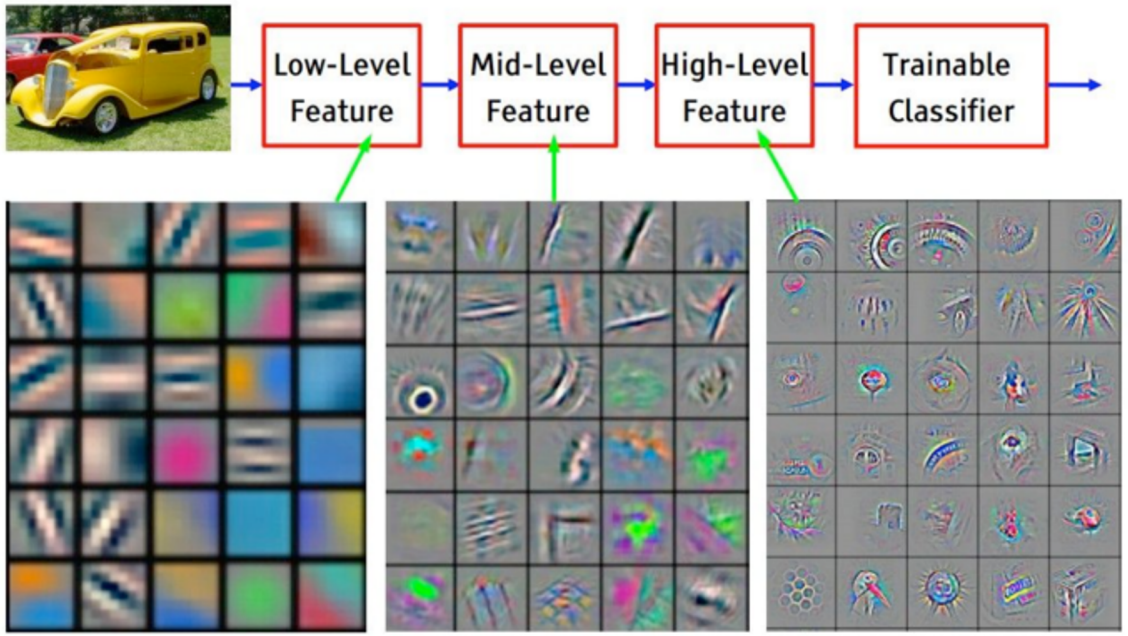
Convolutions

We can stack convolution layeres



Convolutions

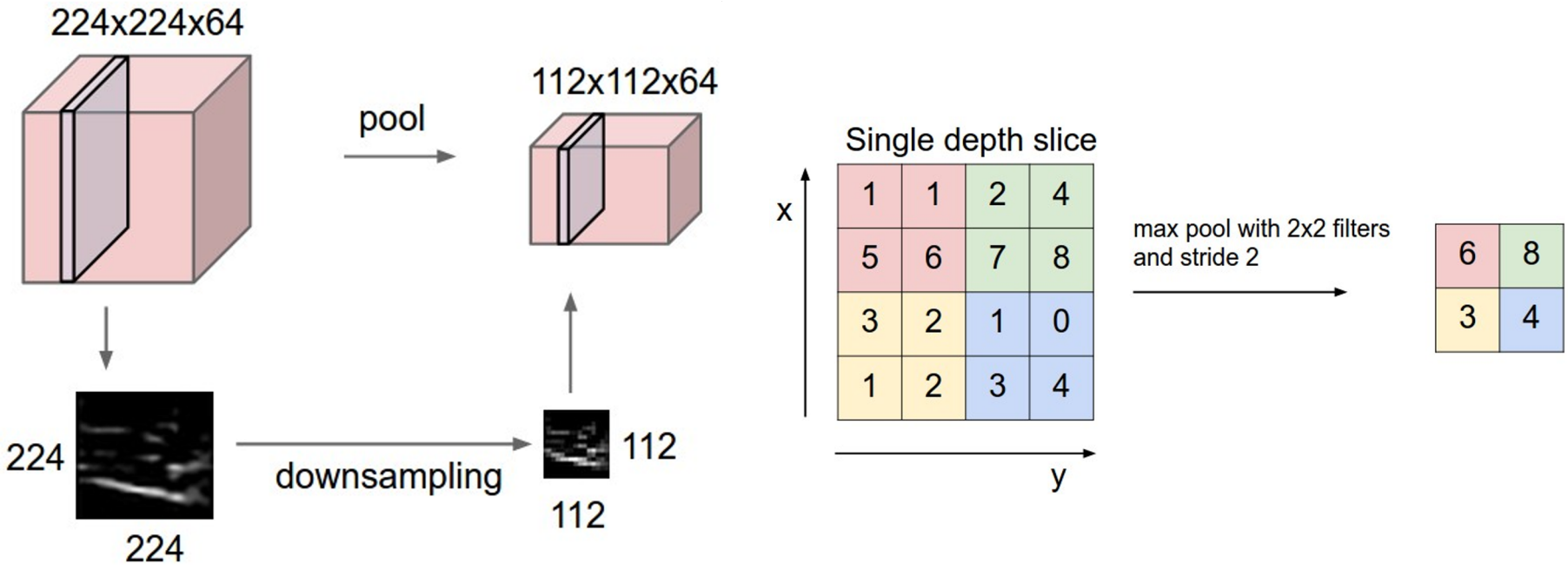
Learning hierarchy of features



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Pooling layers

Commonly used after some convolutional layers to reduce feature map size



http://cs231n.github.io/convolutional-networks/#pool