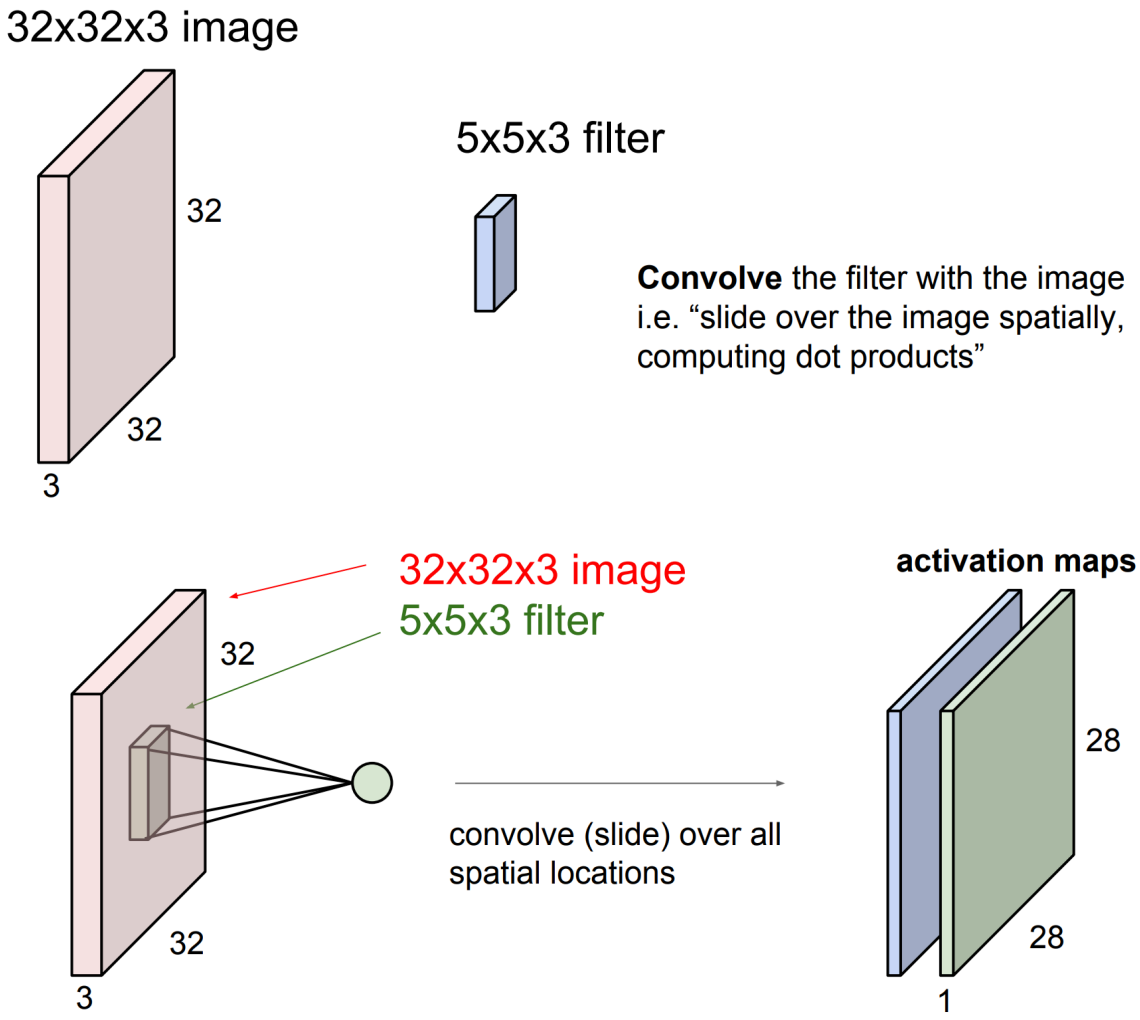
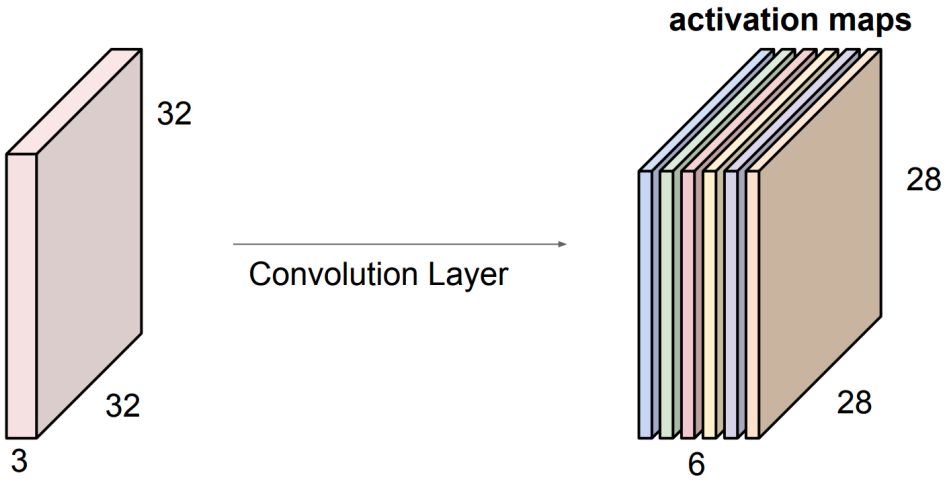
CNN architecture

Convolution layer



For example, if we had 6 5x5 filters, we'll get 6 separate activation maps:



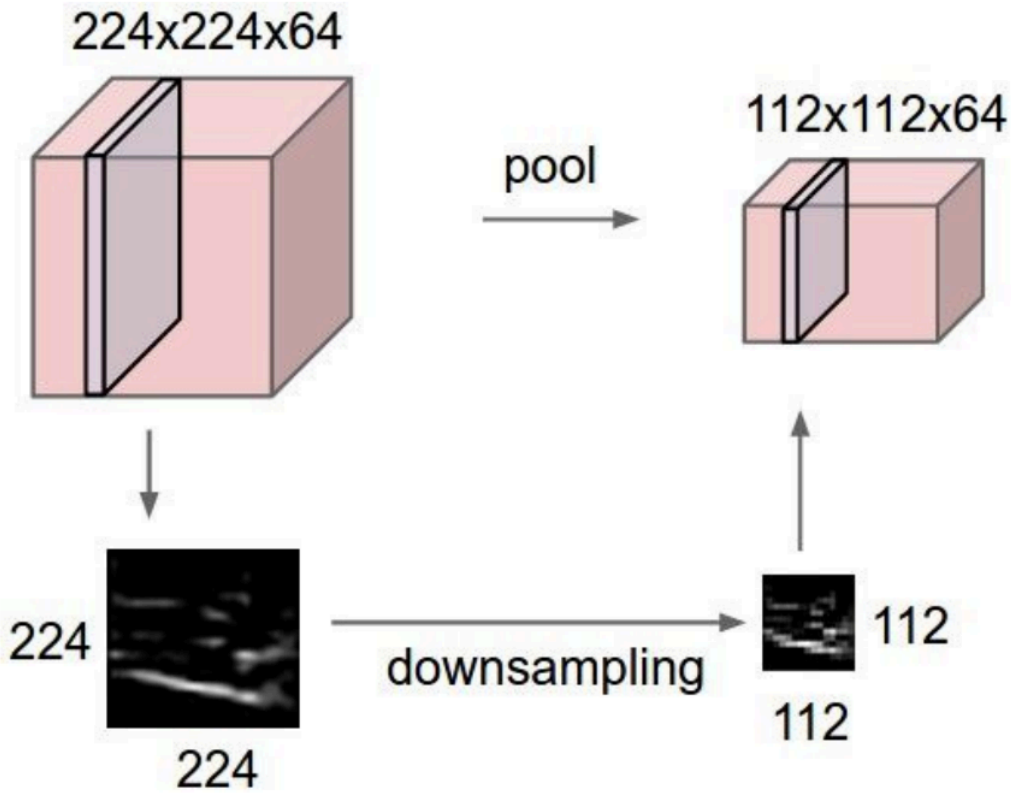
We stack these up to get a "new image" of size 28x28x6!

CONV layers with stride 1, filters of size FxF, and zero-padding with (F-1)/2.

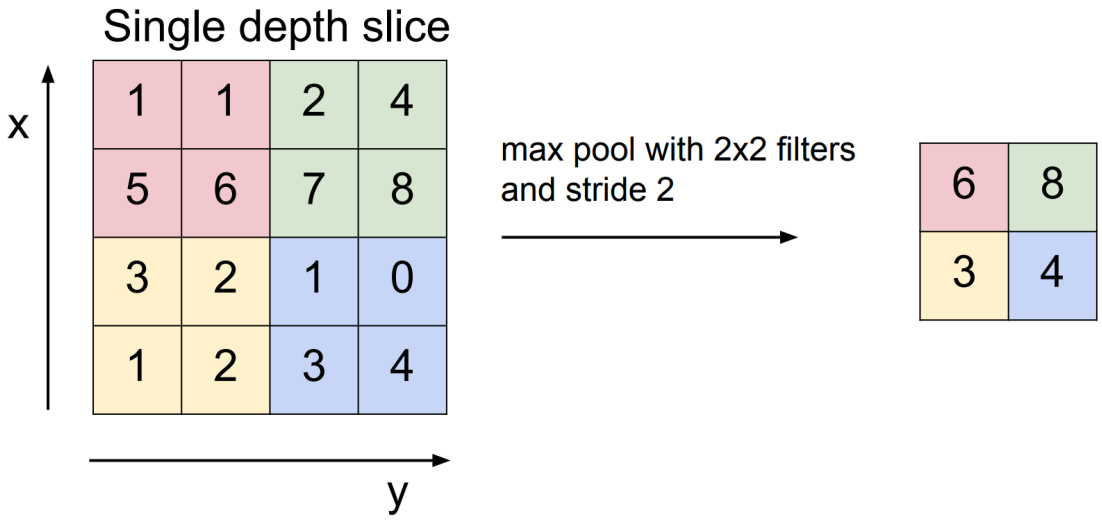
- **F = 3 => zero pad with 1**
- **F = 5 => zero pad with 2**
- **F = 7 => zero pad with 3**

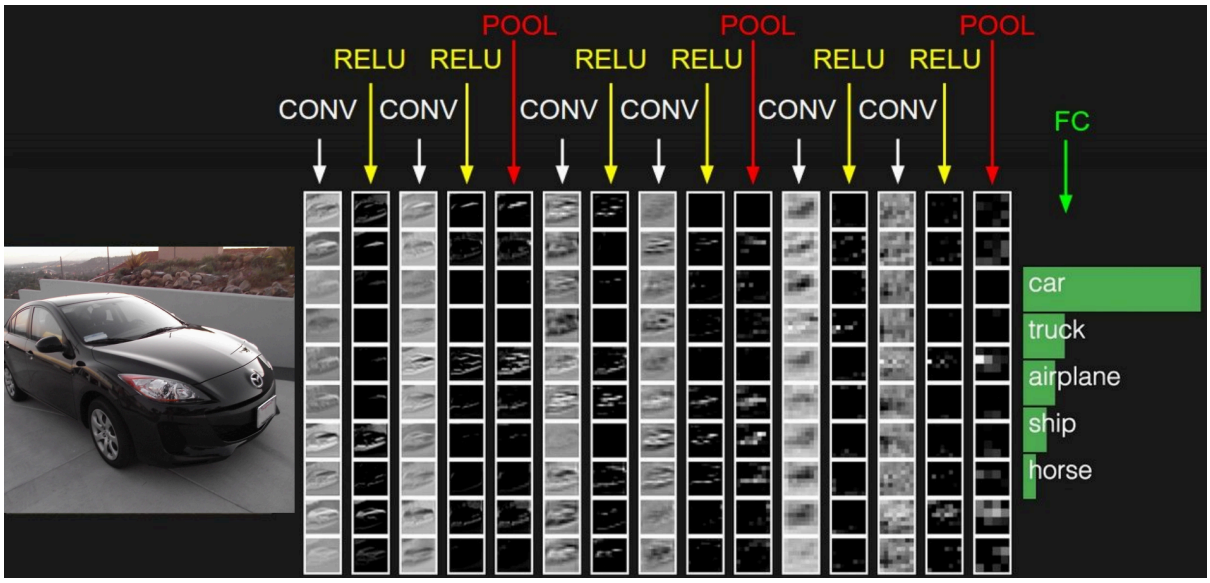
Pooling layer

- **makes the representations smaller and more manageable**
- **operates over each activation map independently**


224x224x64
pool
112x112x64
224
224
downsampling
112
112

MAX POOLING


Single depth slice
x
y
1 1 2 4
5 6 7 8
3 2 1 0
1 2 3 4
max pool with 2x2 filters
and stride 2
6 8
3 4


POOL
POOL
POOL
RELU
RELU
RELU
RELU
RELU
RELU
CONV
CONV
CONV
CONV
CONV
CONV
FC
car
truck
airplane
ship
horse