



Convolutional Neural Networks



Why ConvNets?

- 1D layers do not preserve spatial relationships
- Better for region intensive tasks such as localization and detection
- Better for feature extraction

General Overview of Convnet architecture

- Convolutional layers
- Pooling or subsampling layers
- Fully connected layers

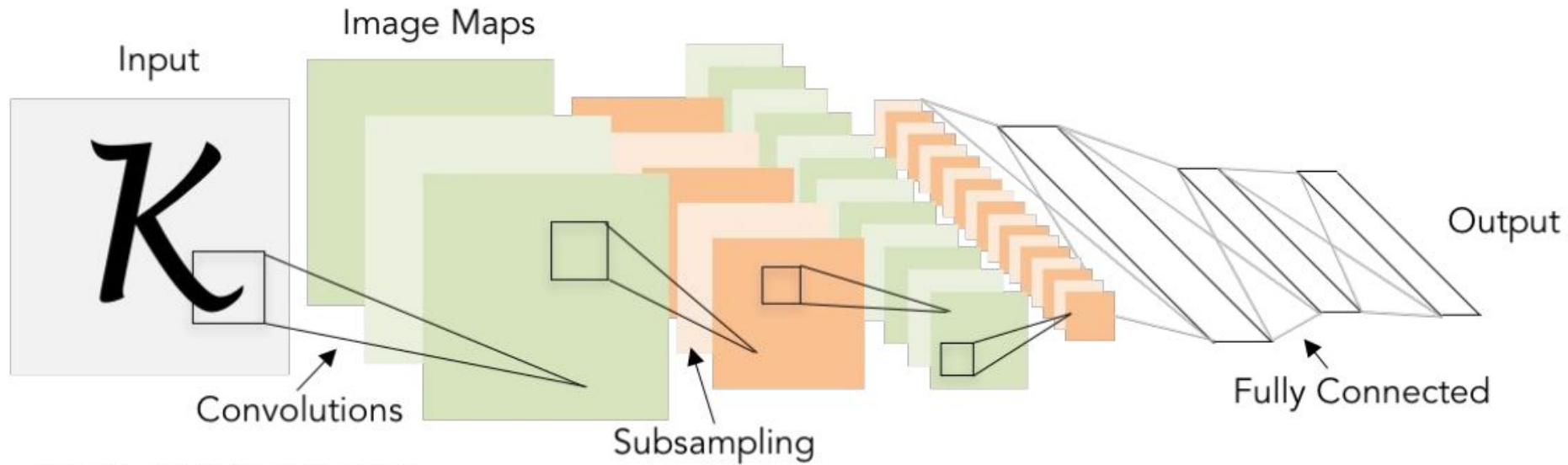
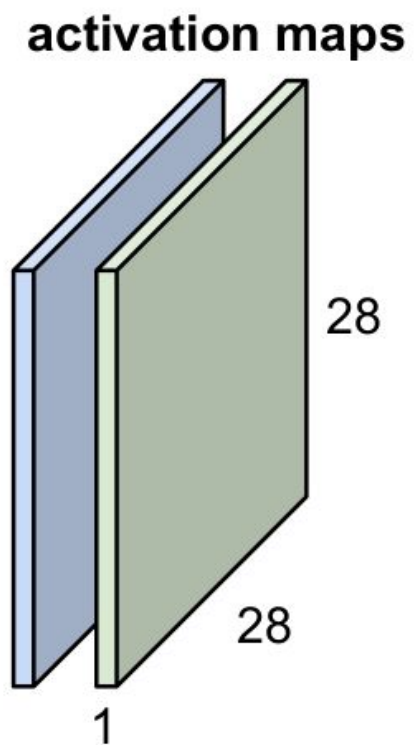
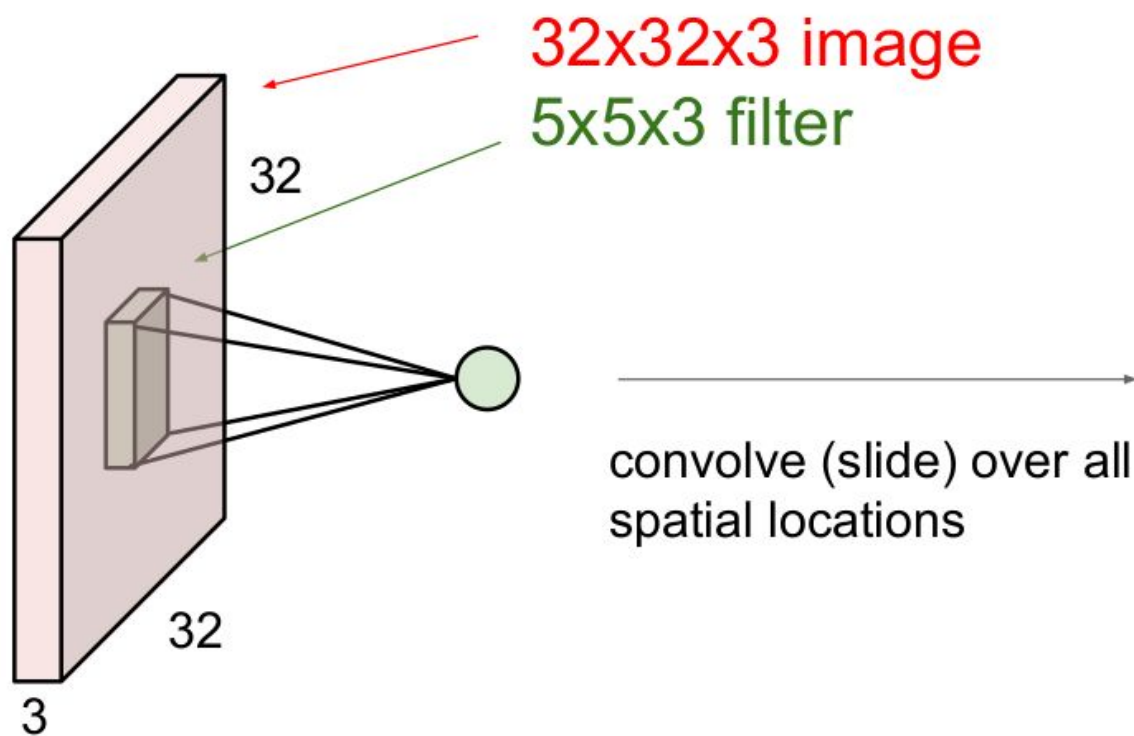


Illustration of LeCun et al. 1998 from CS231n 2017 Lecture 1

Convolutional layers

- **Filters** are sliding matrices of weights used to extract and detect features from the image.
- Each filter basically calculates the dot product between itself and the segment of the image of the same size.



Output size

Output size:
 $(N - F) / \text{stride} + 1$

Padding, New output size

Note that ,you can have different strides horizontally and vertically. You can use the following equations to calculate the exact size of the convolution output for an input with the size of (width = W , height = H) and a Filter with the size of (width = F_w , height = F_h):

$$\text{output width} = \frac{W - F_w + 2P}{S_w} + 1$$

$$\text{output height} = \frac{H - F_h + 2P}{S_h} + 1$$

Pooling

- Reducing the resolution by using only CONV layers is not recommended since
 - They might impact the amount of feature extraction possible.
 - Using CONV for dimension reduction also increases the number of parameters in the network
- To combat this, Pooling techniques serve to reduce the 2D dimension WITH 0 PARAMETERS
- Two popular kinds of pooling:
 - Max Pooling
 - Average Pooling

Single depth slice

x ↑

1	1	2	4
5	6	7	8
3	2	1	0
1	2	3	4

→ y

max pool with 2x2 filters
and stride 2



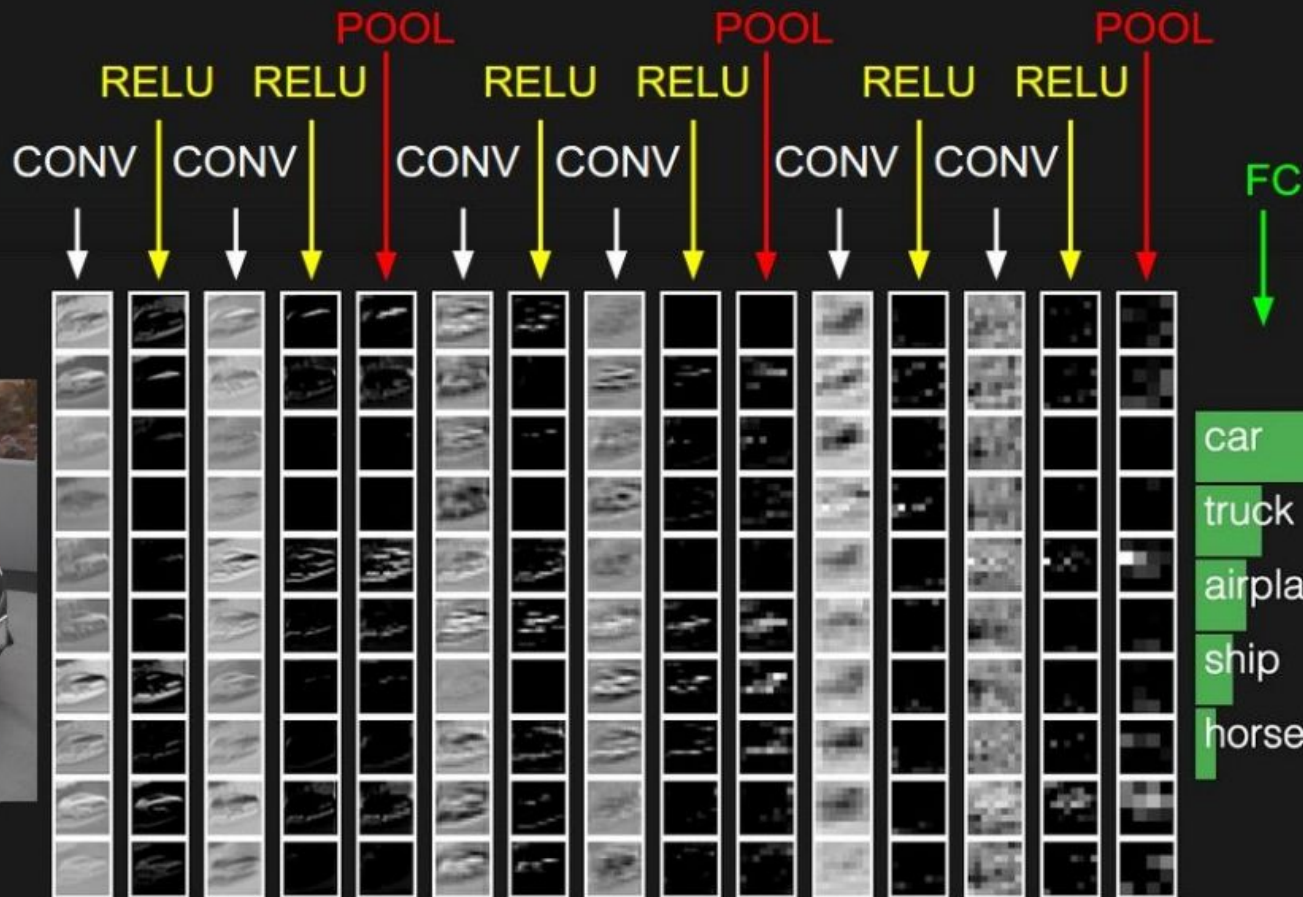
6	8
3	4

Fully connected layers

- WHAT YOU'VE ALREADY LEARNT
- Except you need to flatten before
- THAT'S ALL

The Whole Sequence

- A commonly used architecture style used is
- CONV RELU POOL -> CONV RELU POOL -> ...

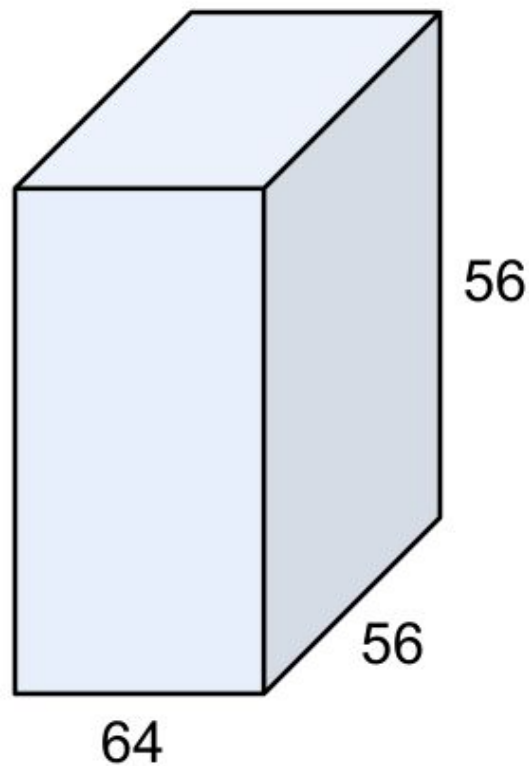


1 x 1 Convolution

- DOES THIS EVEN MAKE SENSE?

Yes

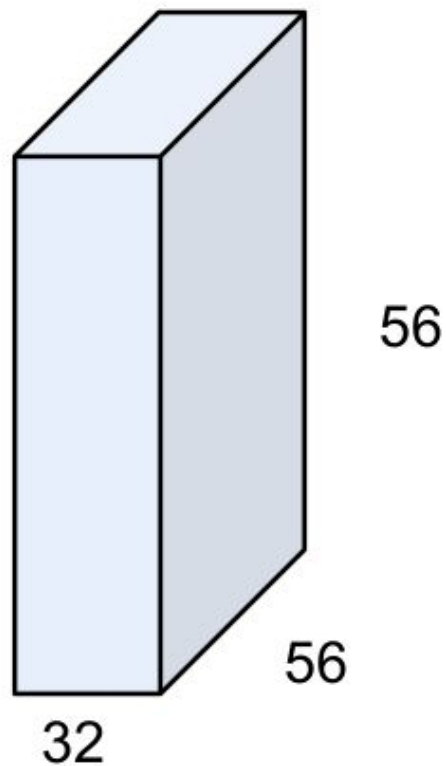
- It's used to reduce the 3D dimension (number of channels) of the feature map without affecting the width and length of the image.



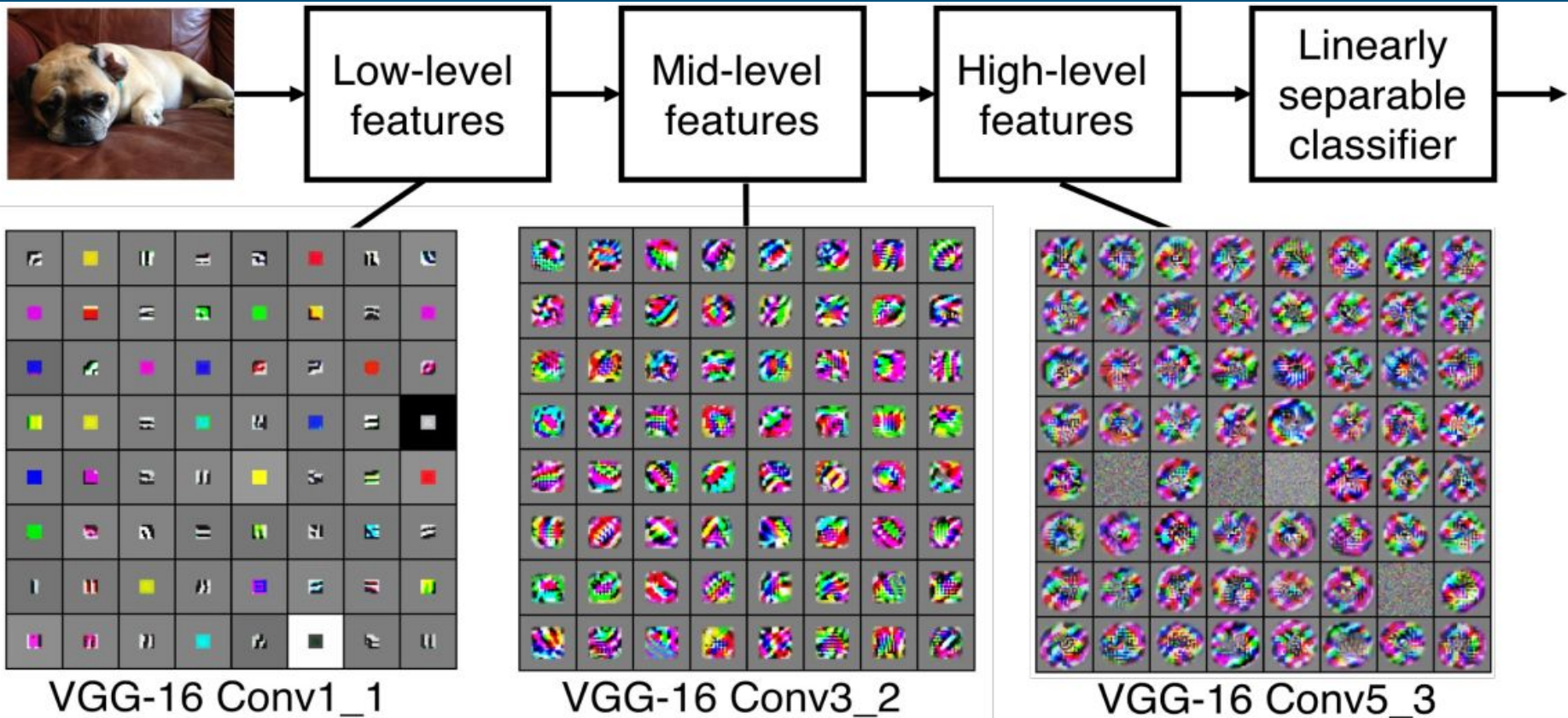
1x1 CONV
with 32 filters



(each filter has size
1x1x64, and performs a
64-dimensional dot
product)



How many CONV-RELU-POOL blocks should we use?



Up next
