DLSG #4 - Convolutional Neural Networks

DLSG

25 November, 2016

Agenda

Housekeeping

- Next session?
- Feedback so far?

Big Schedule Review

- Introduction to Keras
- (Artificial) Neural Networks and their training
- Convolutional Neural Networks

- Recurrent Neural Networks
- Neural Networks in Computer Vision
- Neural Networks in Natural Language Processing

News

- iSee: Using deep learning to remove eyeglasses from faces `https://blog.insightdatascience.com/isee-removing-eyeglasses-from-faces-using-deep-learning.86qdd0mqu`

News

- iSee: Using deep learning to remove eyeglasses from faces `https://blog.insightdatascience.com/isee-removing-eyeglasses-from-faces-using-deep-learning.86qdd0mqu`
- Lip Reading Sentences in the Wild `https://www.youtube.com/watch?v=5aogzAUPilE`

News

- iSee: Using deep learning to remove eyeglasses from faces `https://blog.insightdatascience.com/isee-removing-eyeglasses-from-faces-using-deep-learning.86qdd0mqu`
- Lip Reading Sentences in the Wild `https://www.youtube.com/watch?v=5aogzAUPilE`
- Interactive convnet features visualization for Keras `https://github.com/jakebian/quiver` `https://www.youtube.com/watch?edit=vd&v=tgRW3BRi_FA`

News

- iSee: Using deep learning to remove eyeglasses from faces `https://blog.insightdatascience.com/isee-removing-eyeglasses-from-faces-using-deep-learning.86qdd0mqu`
- Lip Reading Sentences in the Wild `https://www.youtube.com/watch?v=5aogzAUPilE`
- Interactive convnet features visualization for Keras `https://github.com/jakebian/quiver` `https://www.youtube.com/watch?edit=vd&v=tgRW3BRi_FA`
- Image-to-Image Translation with Conditional Adversarial Nets `https://phillipi.github.io/pix2pix/`
- ...

ConvNets

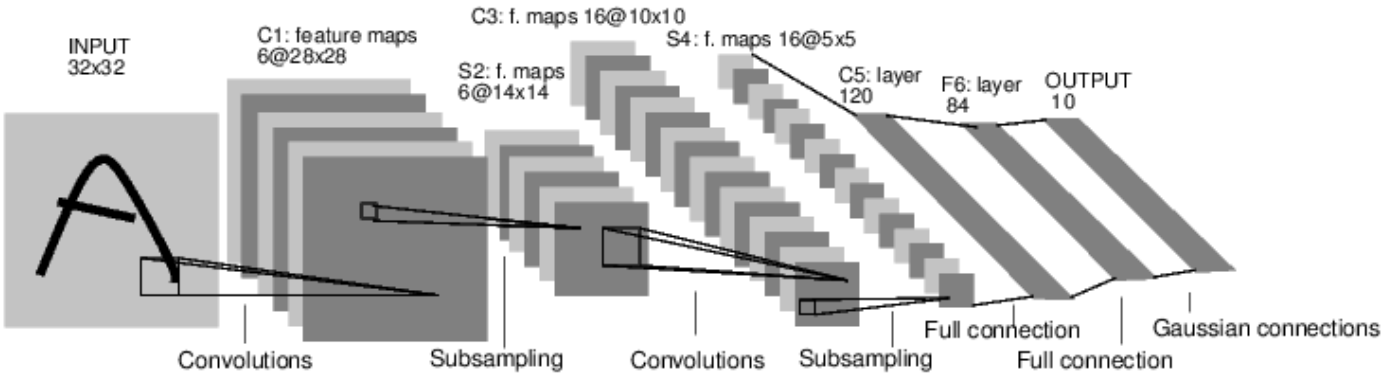
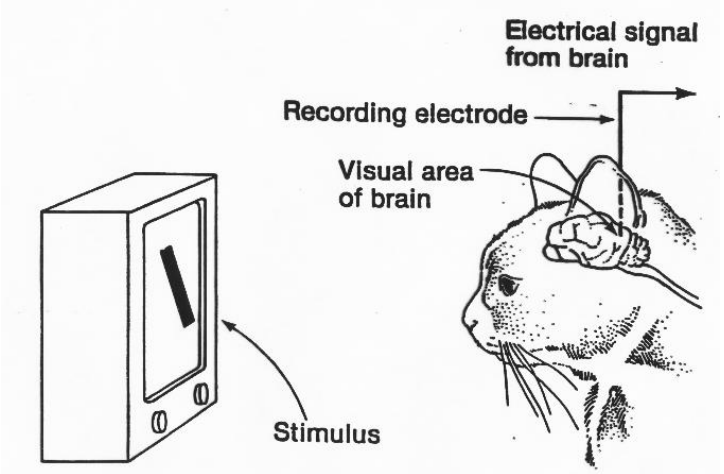
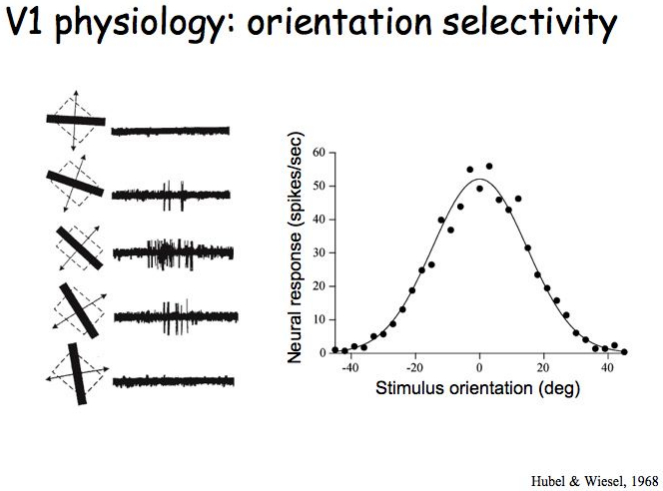


Figure: LeNet [LeCun et al., 1998]

Hubel & Wiesel

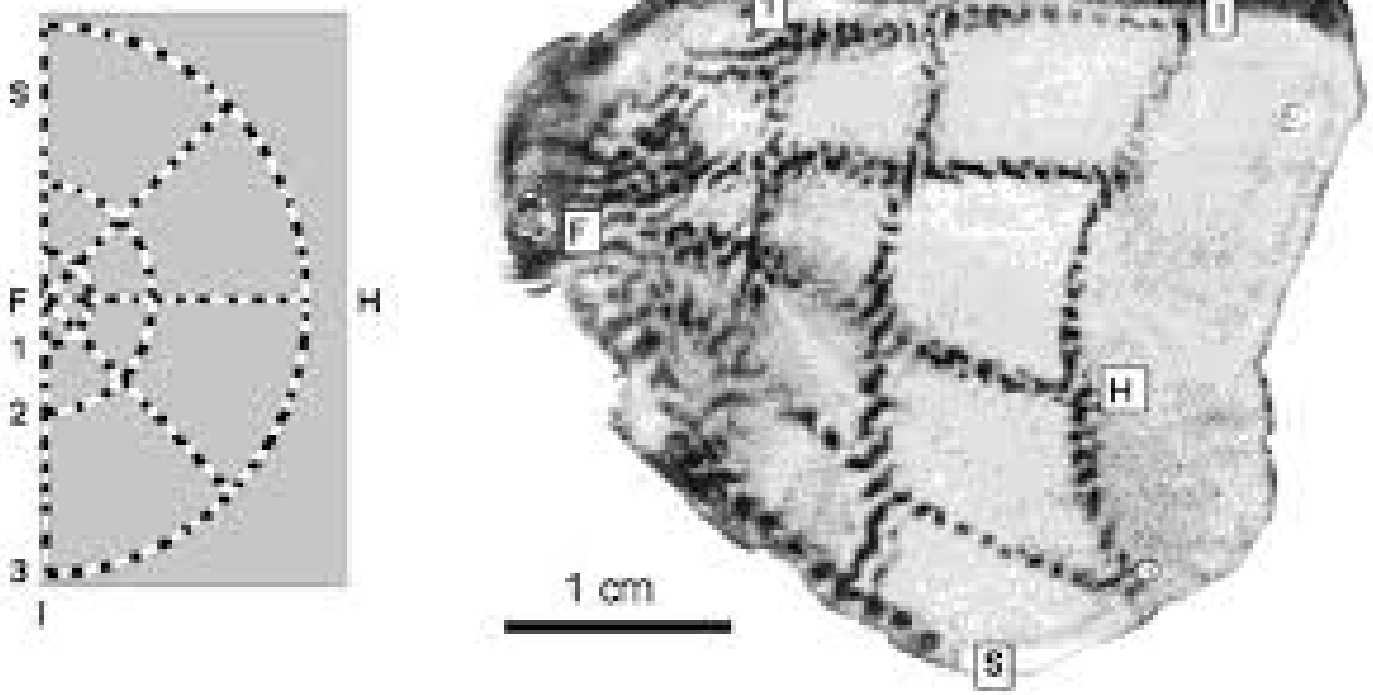
1959 - Receptive fields of single neurones in the cat's striate cortex
1962 - Receptive fields, binocular interaction and functional architecture in the cat's visual cortex





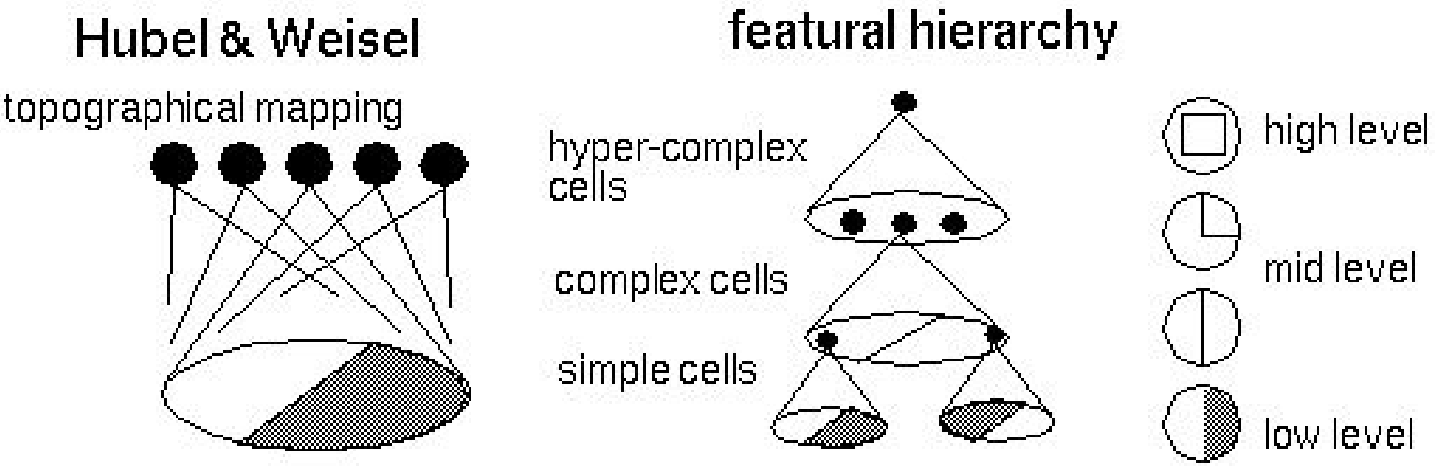
A bit of history

Topographical mapping in the cortex: nearby cells in cortex represented nearby regions in the visual field



A bit of history

Hierarchical organization



A bit of history

Neurocognitron [Fukushima 1980]: “sandwich” architecture (SCSCSC...) simple cells: modifiable parameters complex cells: perform pooling

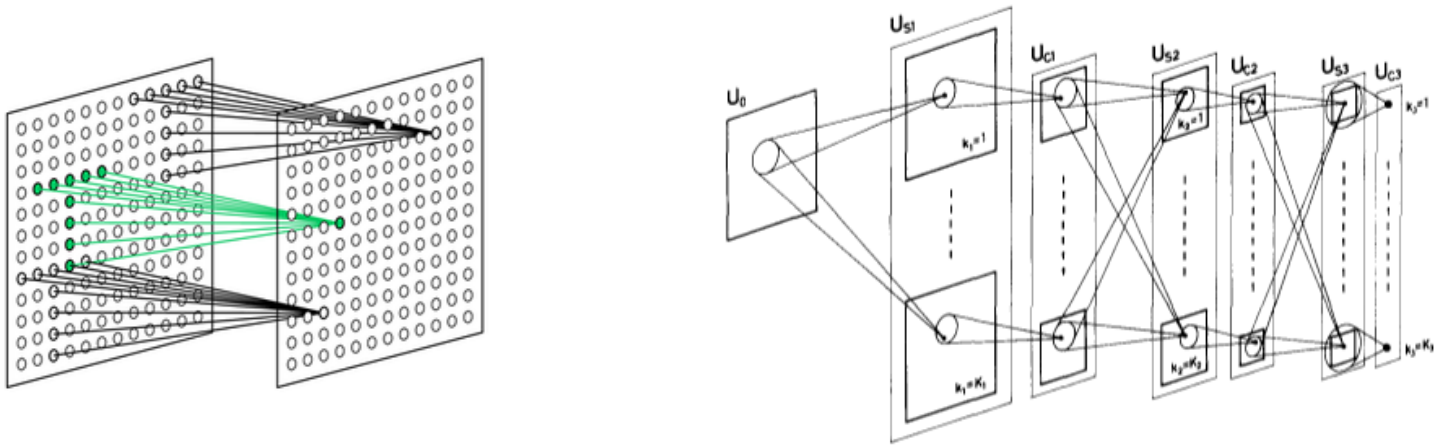


Figure: Neurocognitron [Fukushima 1980]

LeNet

Gradient-based learning applied to document recognition

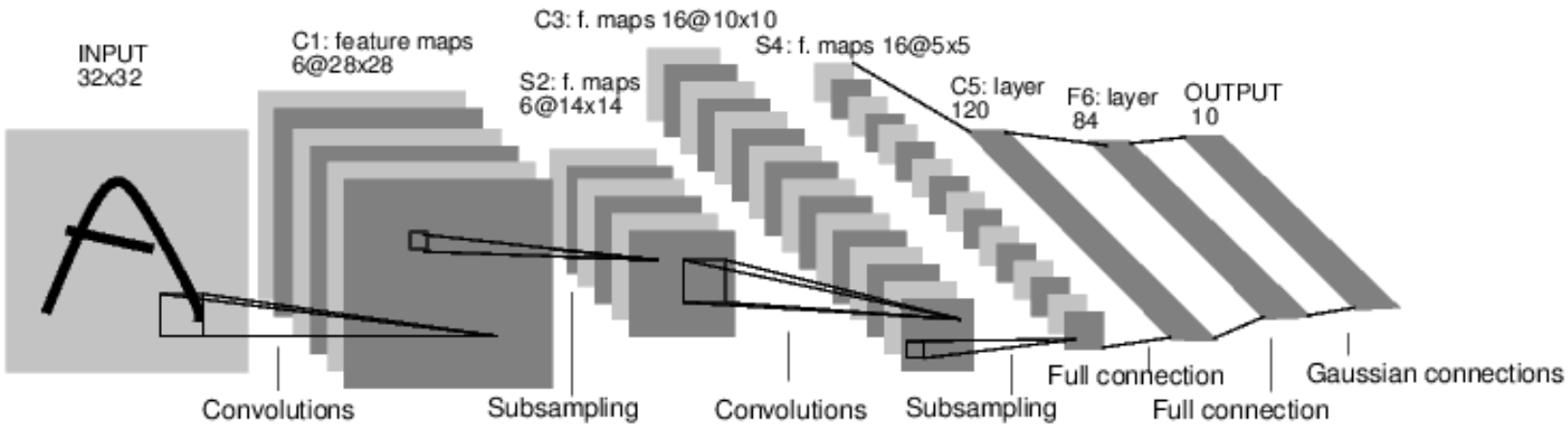


Figure: LeNet [LeCun, Bottou, Bengio, Haffner 1998]

AlexNet

ImageNet Classification with Deep Convolutional Neural Networks

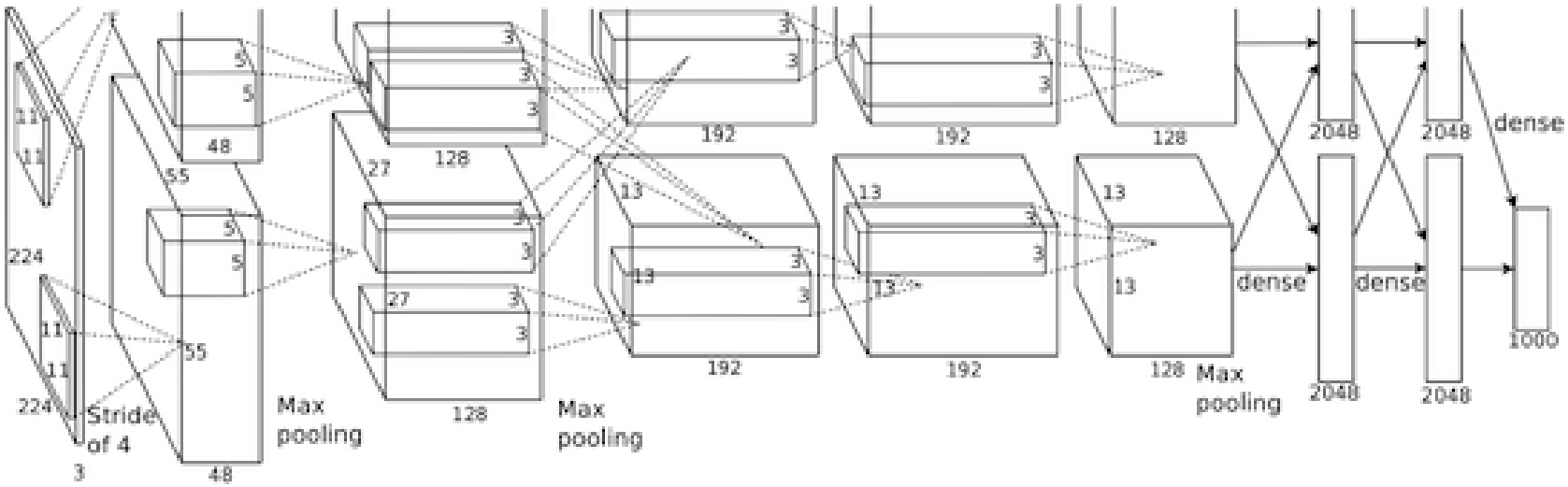


Figure: AlexNet [Krizhevsky, Sutskever, Hinton, 2012]

ResNet

Depth Revolution

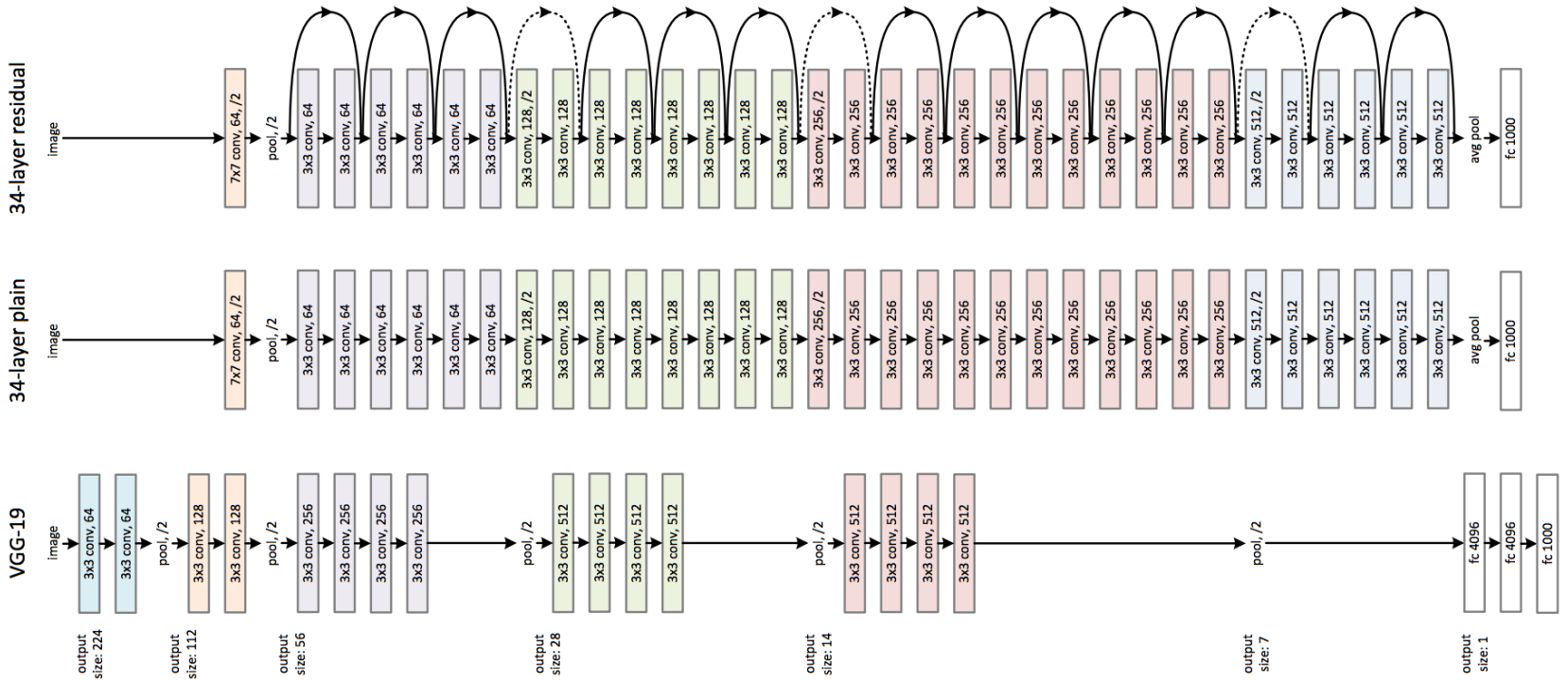


Figure: ResNet [Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, 2015]

ConvNets today

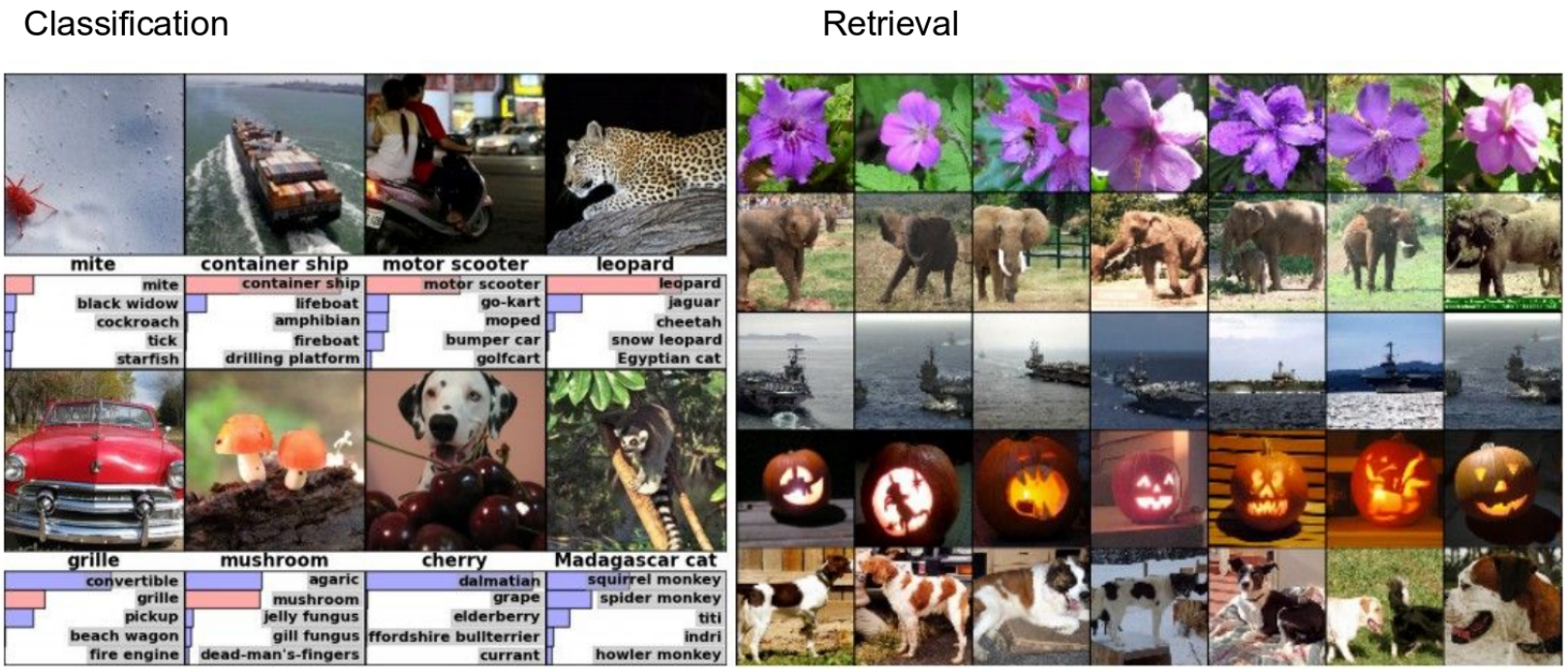


Figure: [Krizhevsky 2012]

ConvNets today

Figure: [Faster R-CNN: Ren, He, Girshick, Sun 2015] Detection Segmentation & [Farabet et al., 2012]

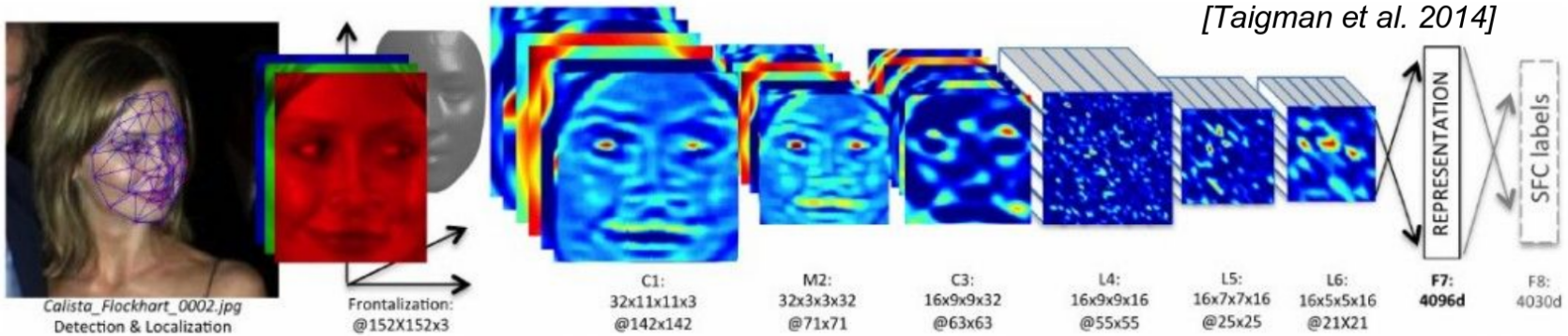
ConvNets today

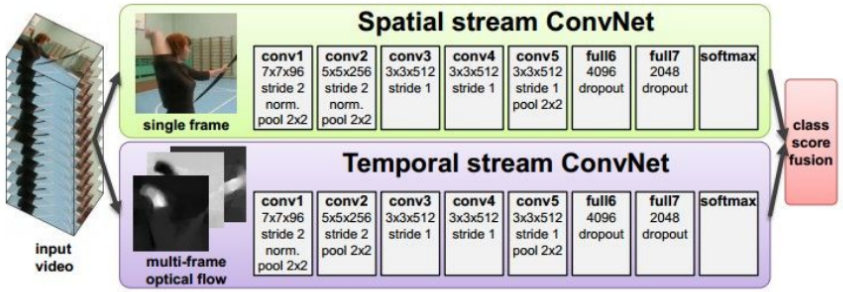
NVIDIA Tegra X1

Figure: Self driving cars

ConvNets today

[Taigman et al. 2014]





[Simonyan et al. 2014]

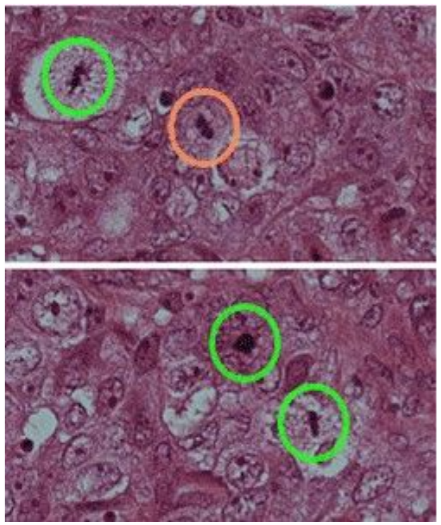
[Goodfellow 2014]

ConvNets today

[Toshev, Szegedy 2014]

[Mnih 2013]

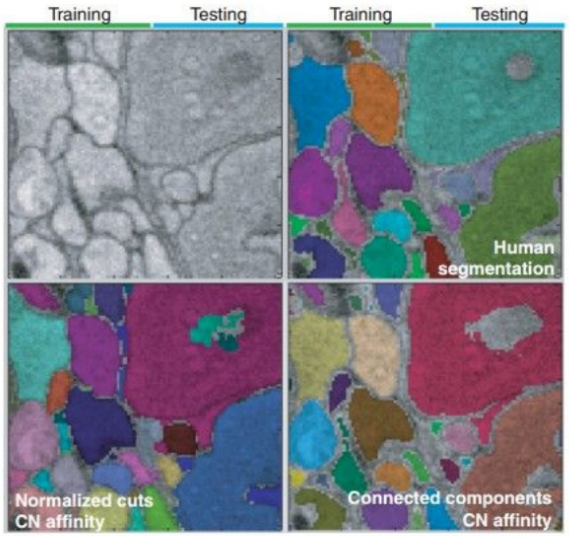
ConvNets today



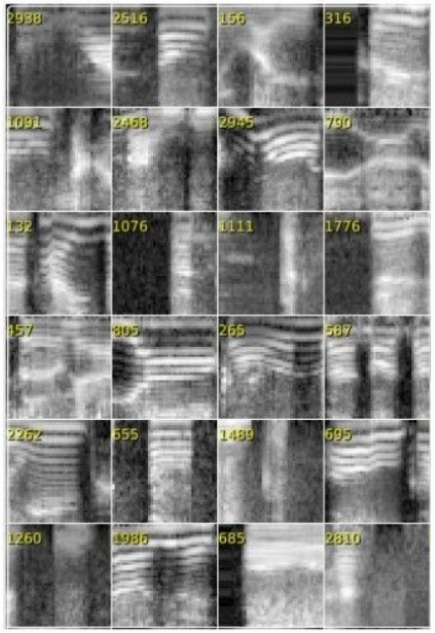
[Ciresan et al. 2013]

[Sermanet et al. 2011]
[Ciresan et al.]

ConvNets today



[Turaga et al., 2010]



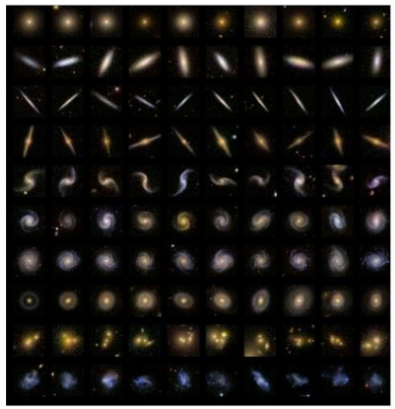
I caught this movie on the Sci-Fi channel recently. It actually turned out to be pretty decent as far as B-list horror/suspense films go. **Two guys (one naive and one loud mouthed a **) take a road trip to stop a wedding but have the worst possible luck when a maniac in a freaky, make-shift tank/truck hybrid decides to play cat-and-mouse with them.** Things are further complicated when they pick up a ridiculously whorish hitchhiker. What makes this film unique is that the combination of comedy and terror actually work in this movie, unlike so many others. The two guys are likable enough and there are some good chase/suspense scenes. Nice pacing and comic timing make this movie more than passable for the horror/slasher buff. **Definitely worth checking out.**

I just saw this on a local independent station in the New York City area. **The cast showed promise but when I saw the director, George Cosmotos, I became suspicious. And sure enough, it was every bit as bad, every bit as pointless and stupid as every George Cosmotos movie I ever saw.** He's like a stupid man's Michael Bey – with all the awfulness that accolade promises. There's no point to the conspiracy, no burning issues that urge the conspirators on. We are left to ourselves to connect the dots from one bit of graffiti on various walls in the film to the next. Thus, the current budget crisis, the war in Iraq, Islamic extremism, the fate of social security, 47 million Americans without health care, stagnating wages, and the death of the middle class are all subsumed by the sheer terror of graffiti. A truly, stunningly idiotic film.

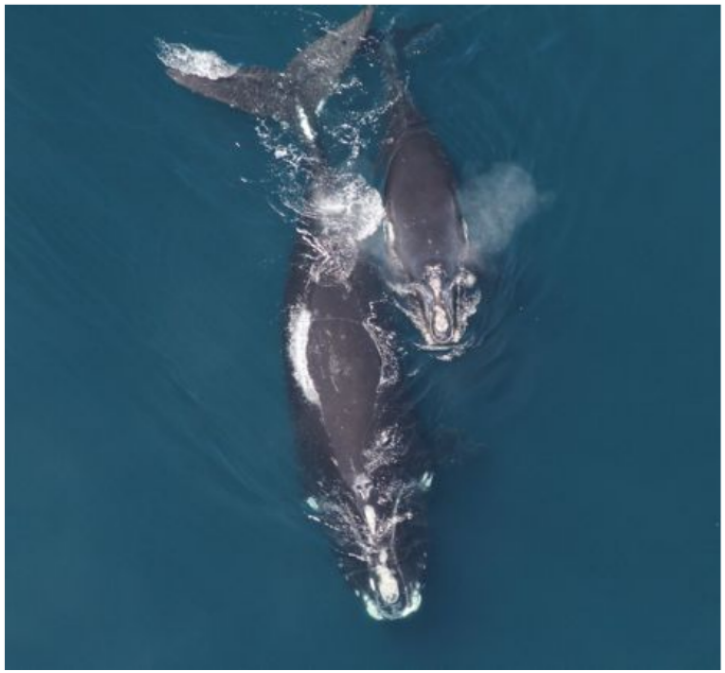
Graphics is far from the best part of the game. **This is the number one best TH game in the series.** Next to Underground. **It deserves strong love. It is an insane game.** There are massive levels, massive unlockable characters... it's just a massive game. **Waste your money on this game. This is the kind of money that is wasted properly.** And even though graphics suck, thats doesn't make a game good. Actually, the graphics were good at the time. Today the graphics are crap. WHO CARES? As they say in Canada, This is the fun game, aye. (You get to go to Canada in THPS3) Well, I don't know if they say that, but they might. who knows. Well, Canadian people do. Wait a minute, I'm getting off topic. This game rocks. Buy it, play it, enjoy it, love it. It's PURE BRILLIANCE.

The first was good and original. I was a not bad horror/comedy movie. So I heard a second one was made and I had to watch it . What really makes this movie work is Judd Nelson's character and the sometimes clever script. **A pretty good script for a person who wrote the Final Destination films and the direction was okay.** Sometimes there's scenes where it looks like it was filmed using a home video camera with a grainy - look. Great made - for - TV movie. **It was worth the rental and probably worth buying just to get that nice eerie feeling and watch Judd Nelson's Stanley doing what he does best.** I suggest newcomers to watch the first one before watching the sequel, just so you'll have an idea what Stanley is like and get a little history background.

[Denil et al. 2014]



ConvNets today



Whale recognition, Kaggle Challenge

Mnih and Hinton, 2010

ConvNets today

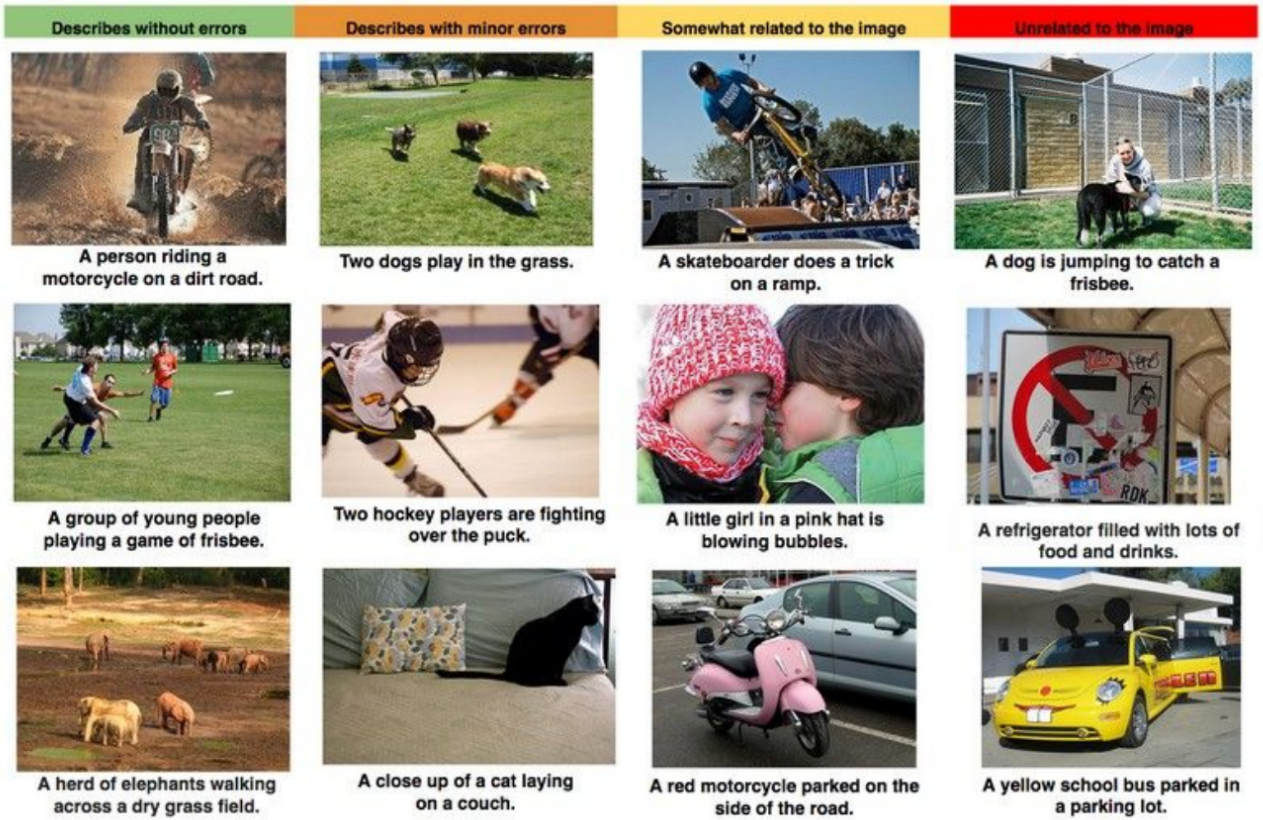
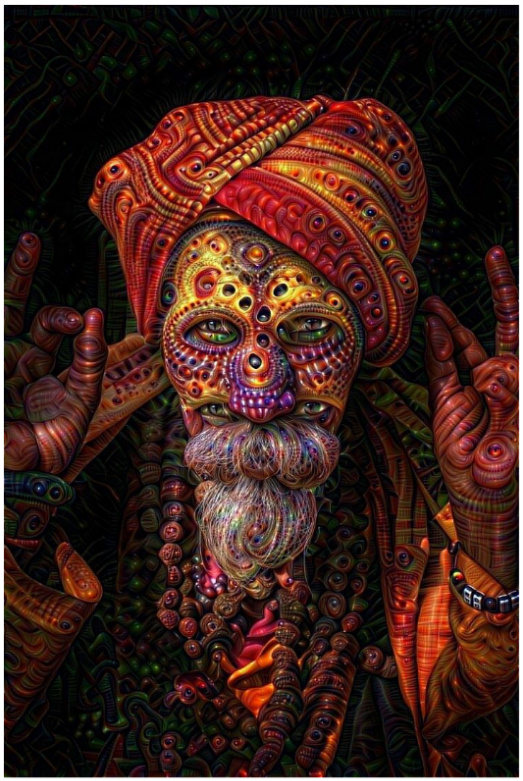
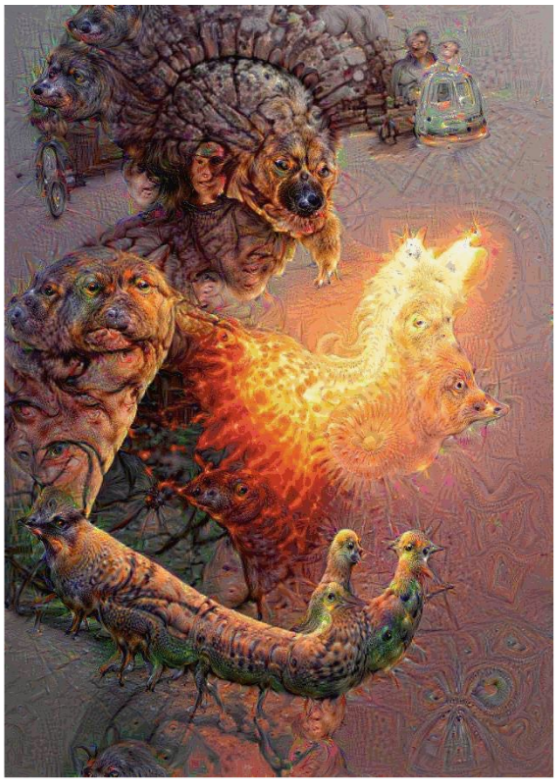


Image Captioning

[Vinyals et al., 2015]

ConvNets today

reddit.com/r/deepdream

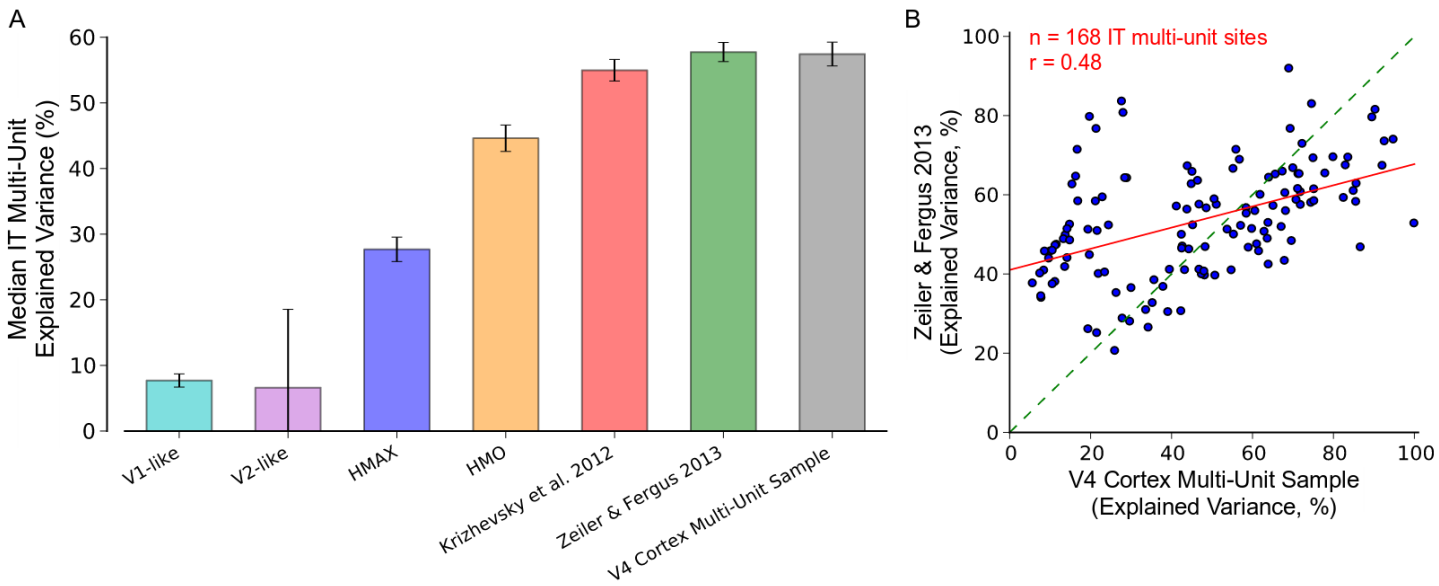


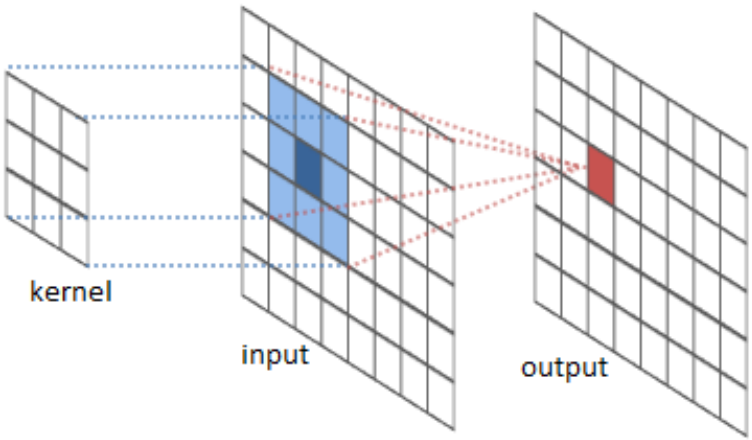
Figure: Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition [Cadieu et al., 2014]

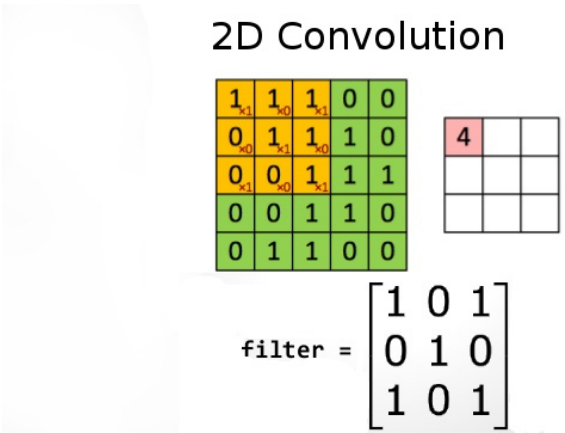
Convolution

So what is "convolution"

Convolution

So what is "convolution"





Convolution

$$\sum_{i=1}^{n} x_i w_i$$

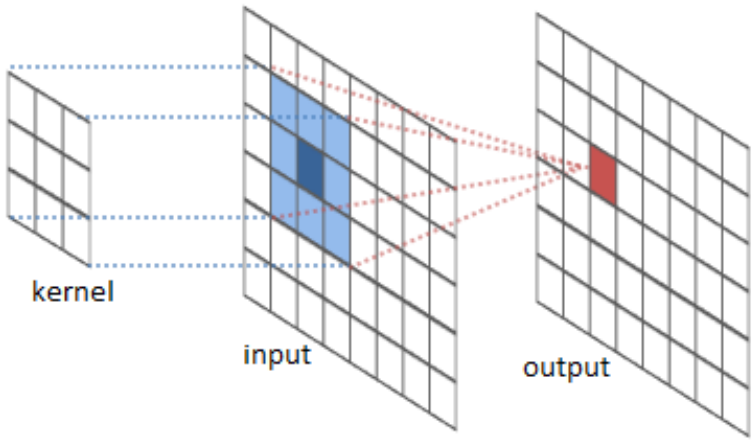
Convolution

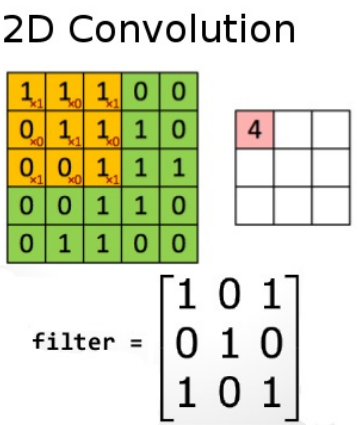
$$\sum_{i=1}^{n} x_i w_i = X \cdot W$$

Convolution

$$(f * g)[n] = \sum_{m=-\infty}^{\infty} f(m)g(n-m)$$

Convolution



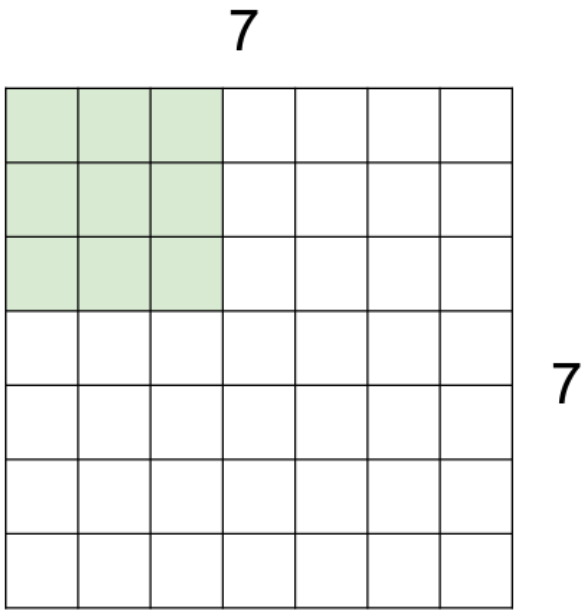


Convolution

We don't have to go with stride 1

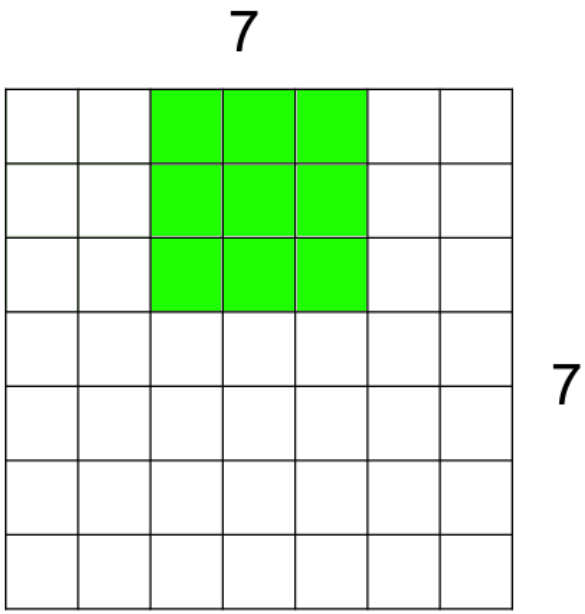
Convolution

We don't have to go with stride 1



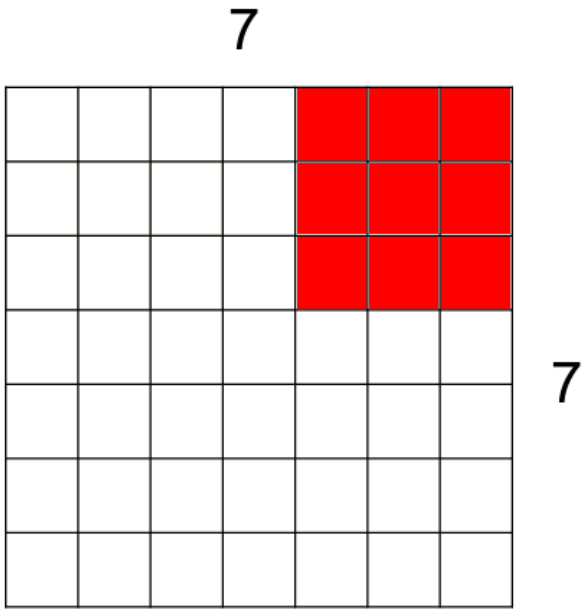
7x7 input (spatially)
assume 3x3 filter
applied **with stride 2**

Convolution



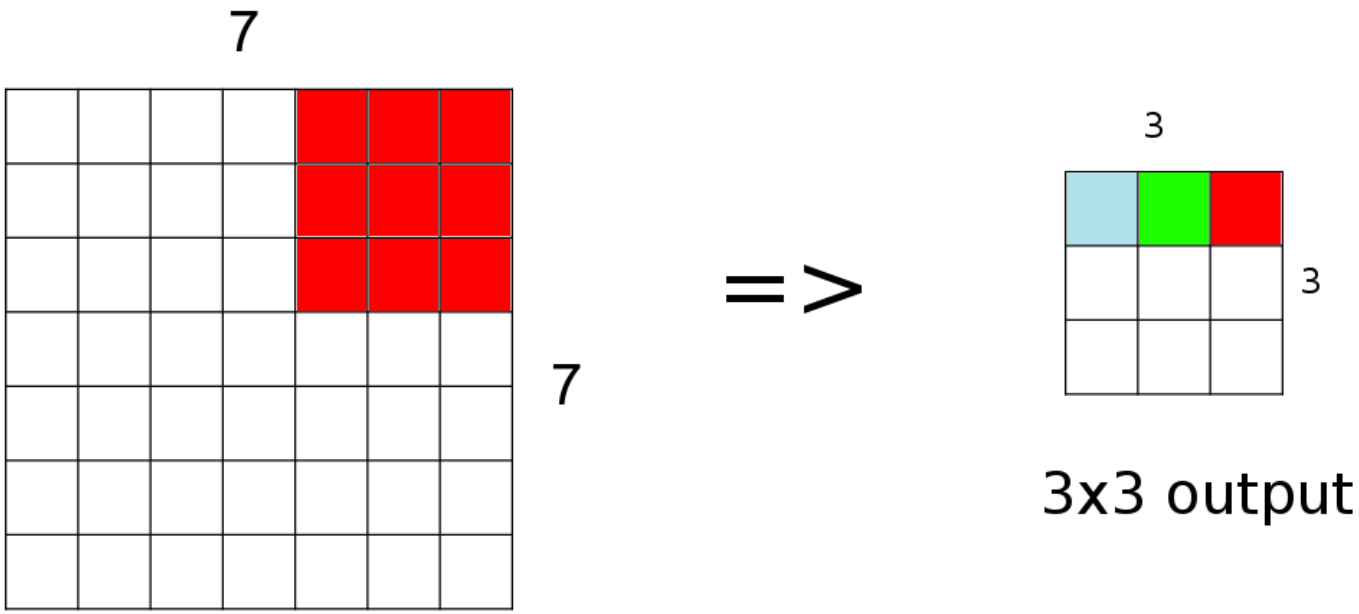
7x7 input (spatially)
assume 3x3 filter
applied **with stride 2**

Convolution

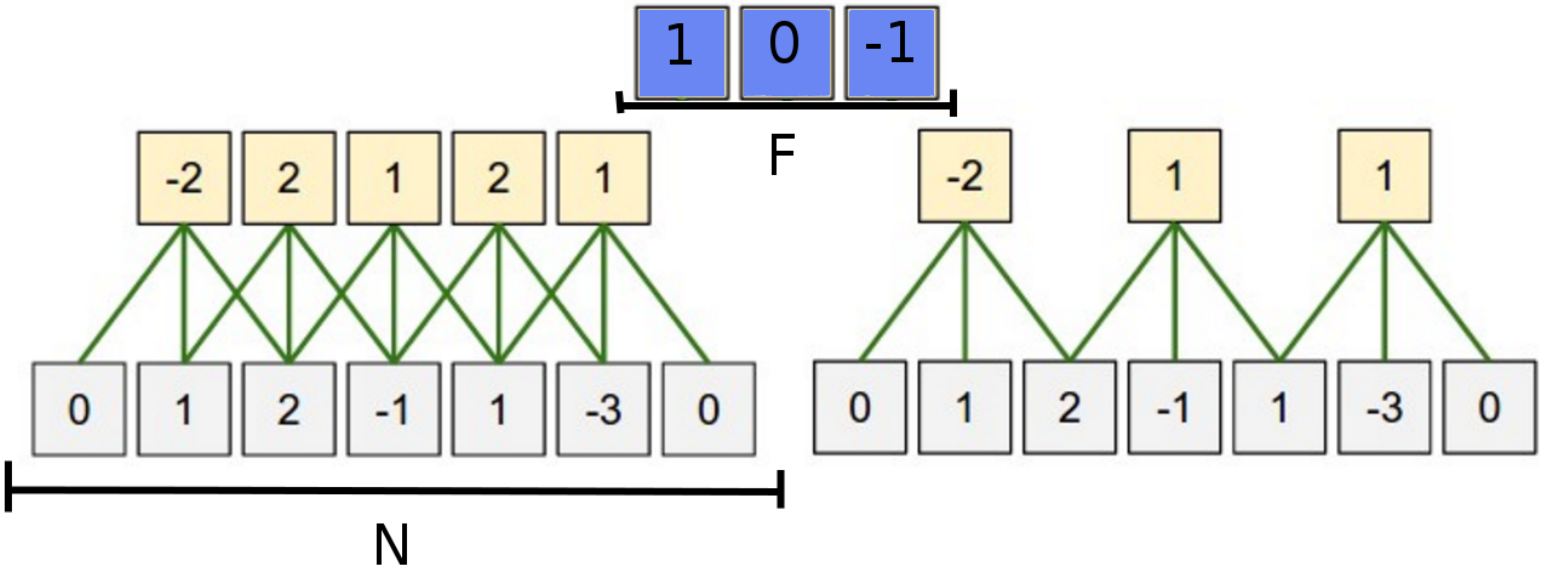


7x7 input (spatially)
assume 3x3 filter
applied **with stride 2**

Convolution



Convolution



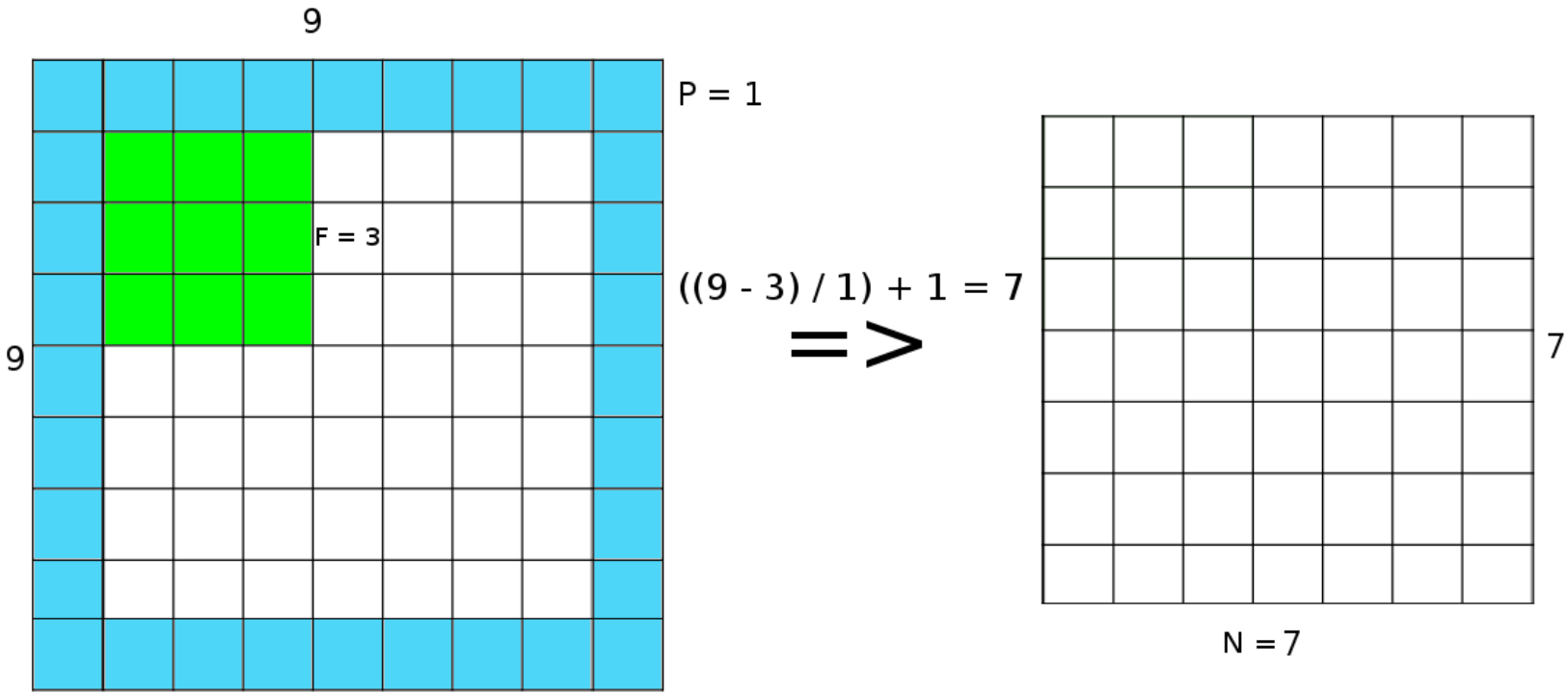
Output size: $\frac{N-F}{stride} + 1$

Convolution

What if I want to keep spatial dimension?

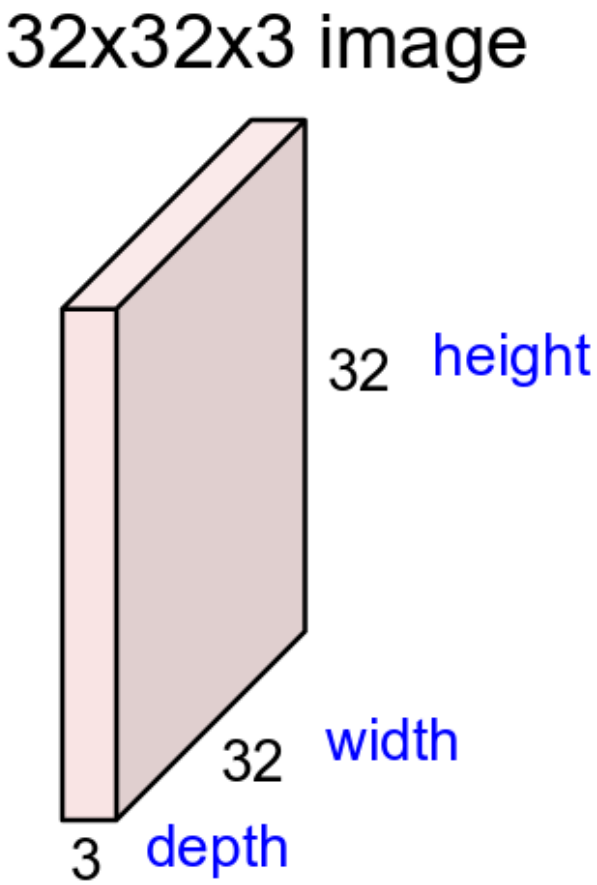
Convolution

What if I want to keep spatial dimension? Pad the input!

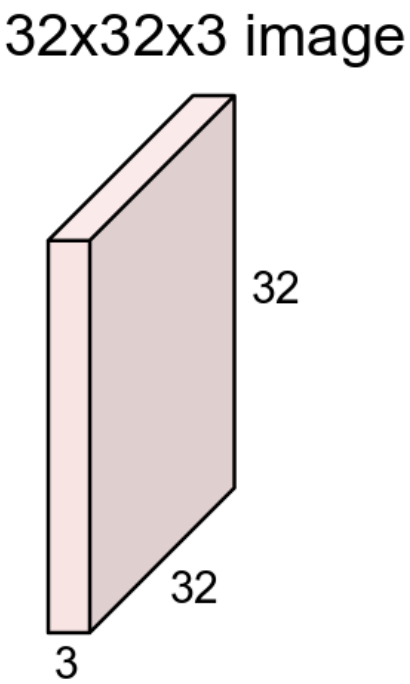


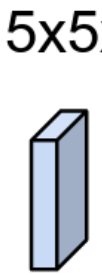
Output size: $\frac{N-F+2P}{stride} + 1$

Convolution layer



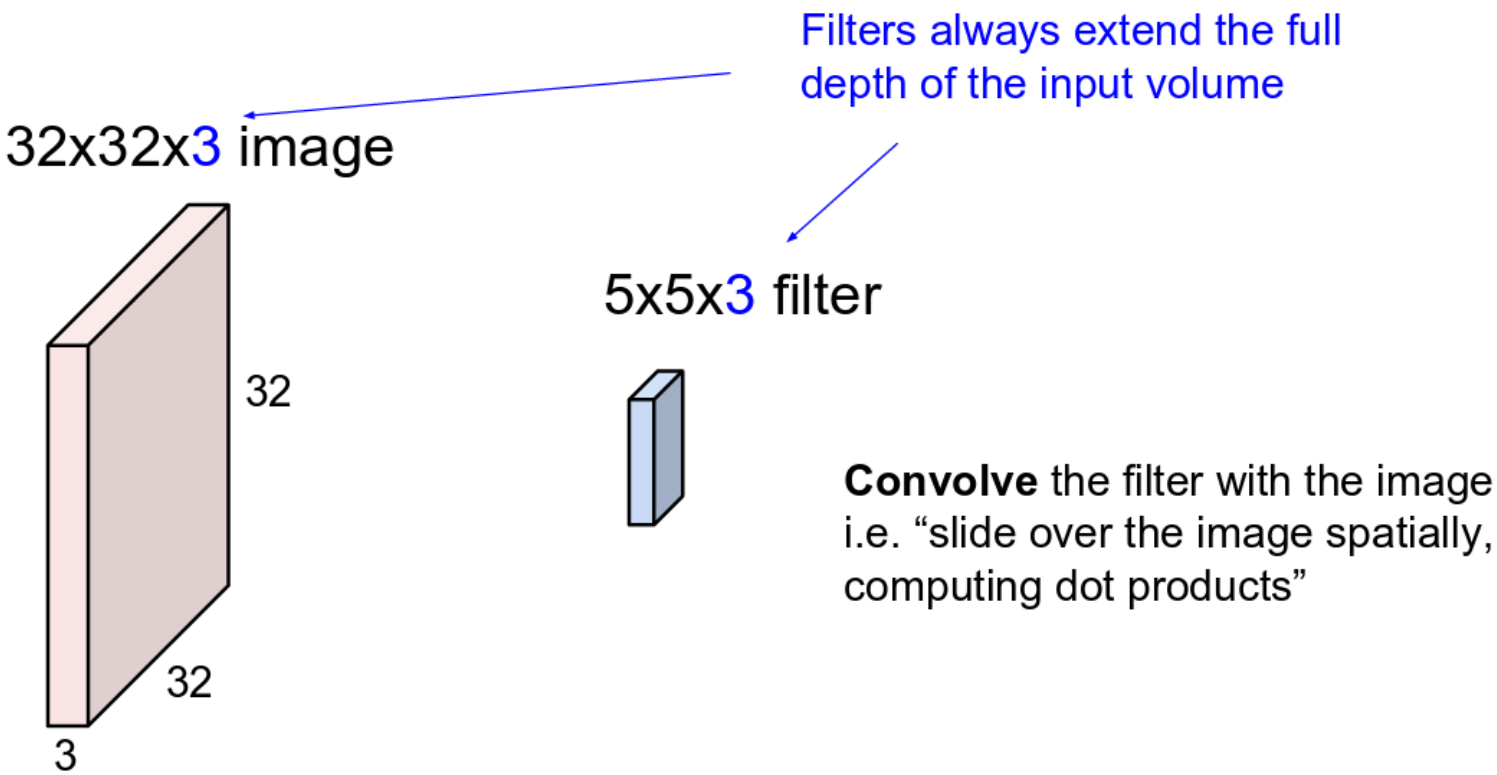
Convolution layer



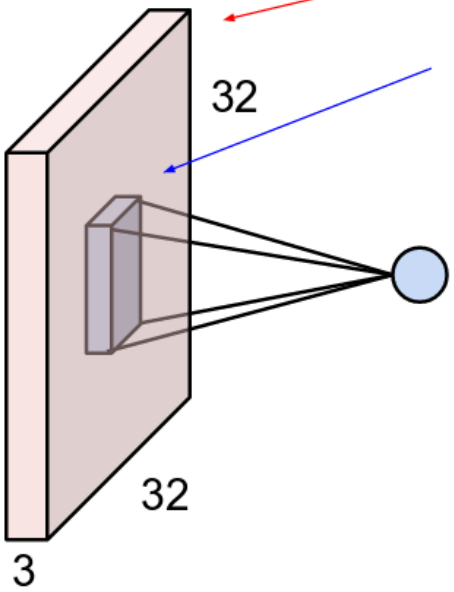


Convolve the filter with the image
i.e. “slide over the image spatially,
computing dot products”

Convolution layer



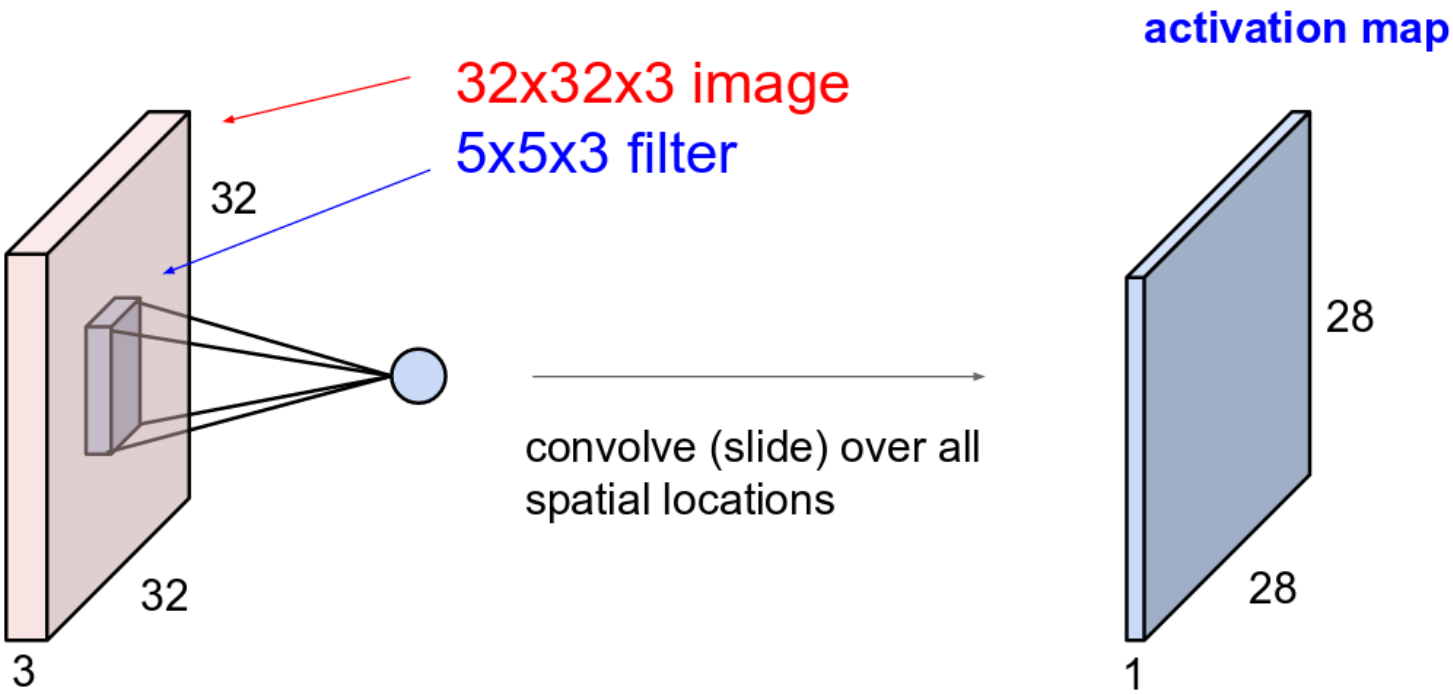
Convolution layer



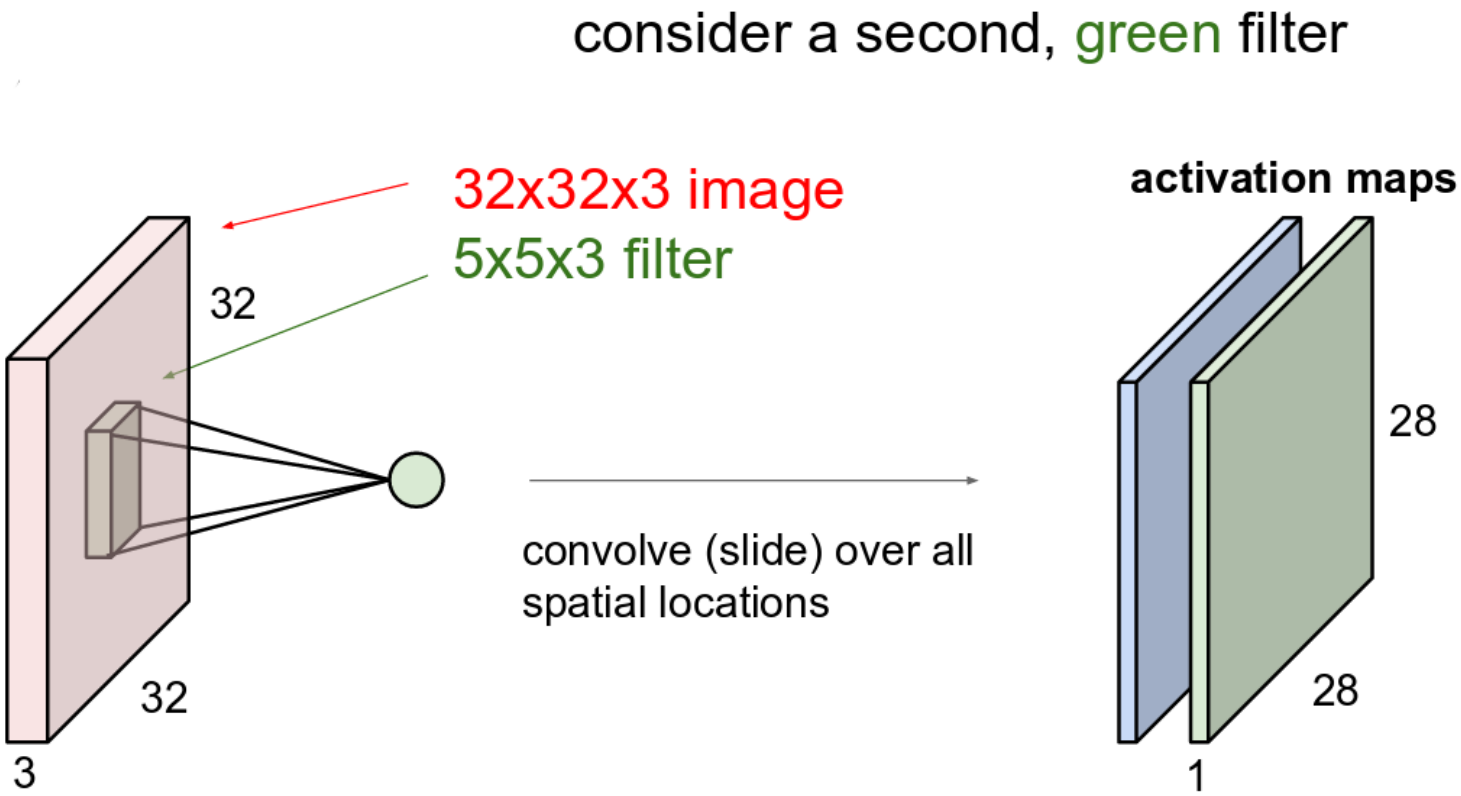
1 number:
the result of taking a dot product between the filter and a small 5x5x3 chunk of the image
(i.e. 5*5*3 = 75-dimensional dot product + bias)

$$w^T x + b$$

Convolution layer

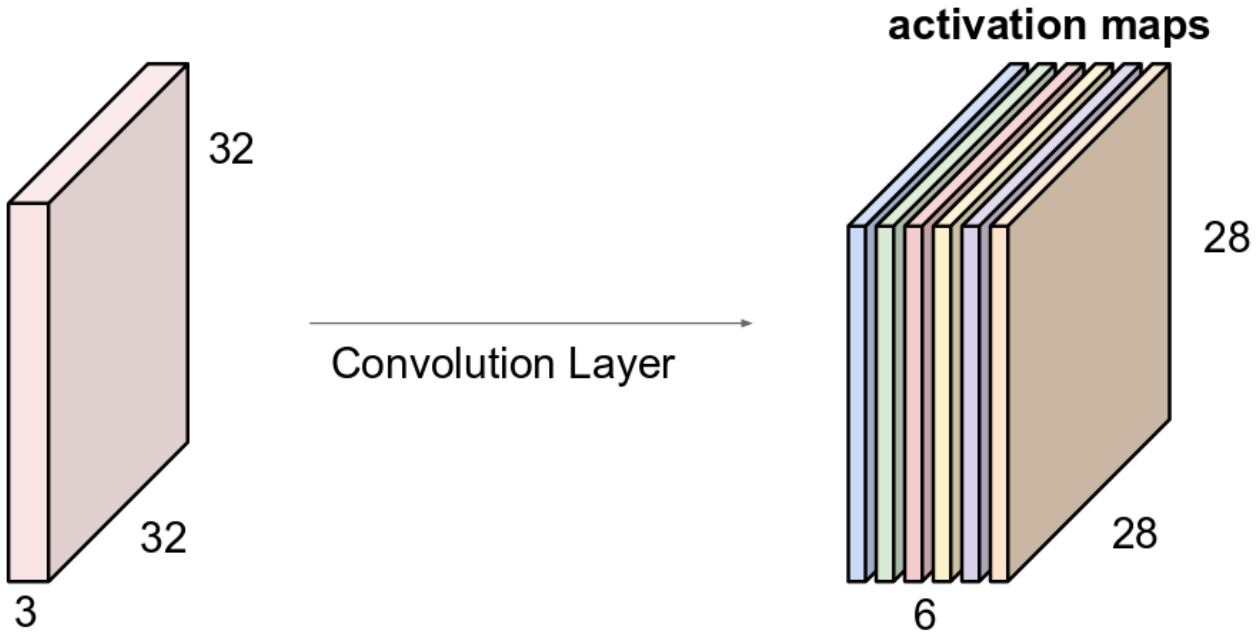


Convolution layer



Convolution layer

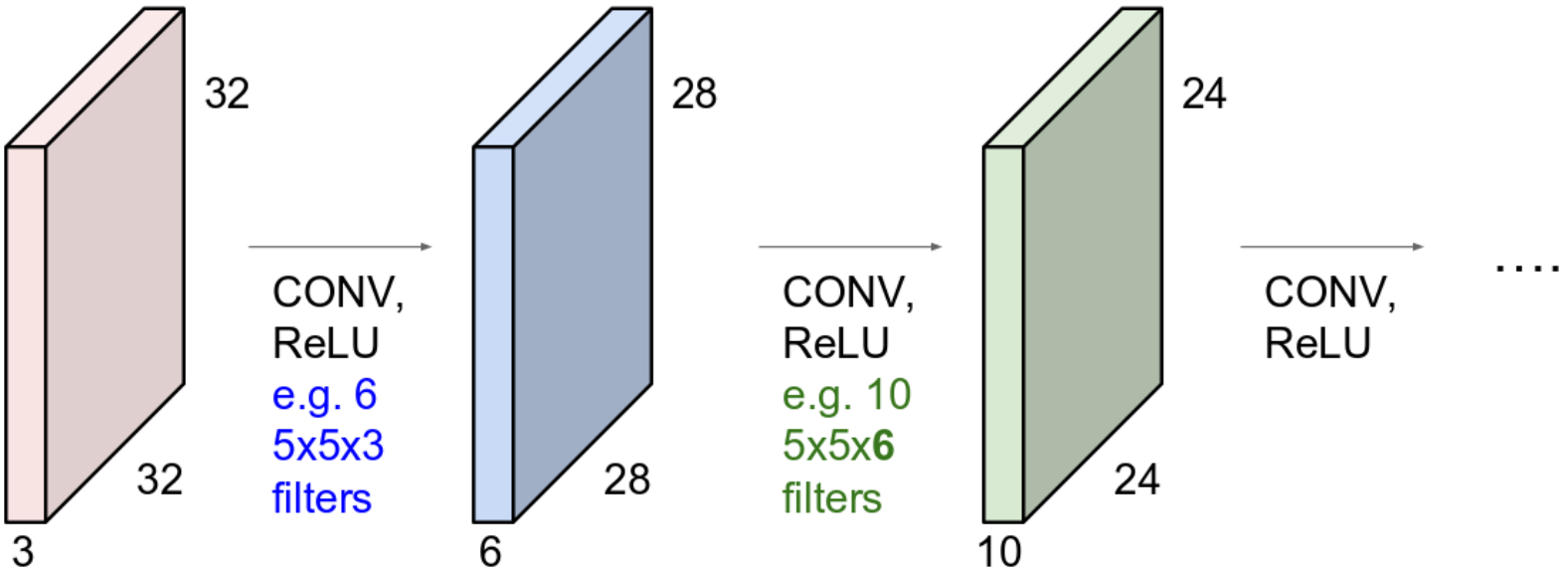
For example, if we had 6 5x5 filters, we'll get 6 separate activation maps:



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

Convolution layer

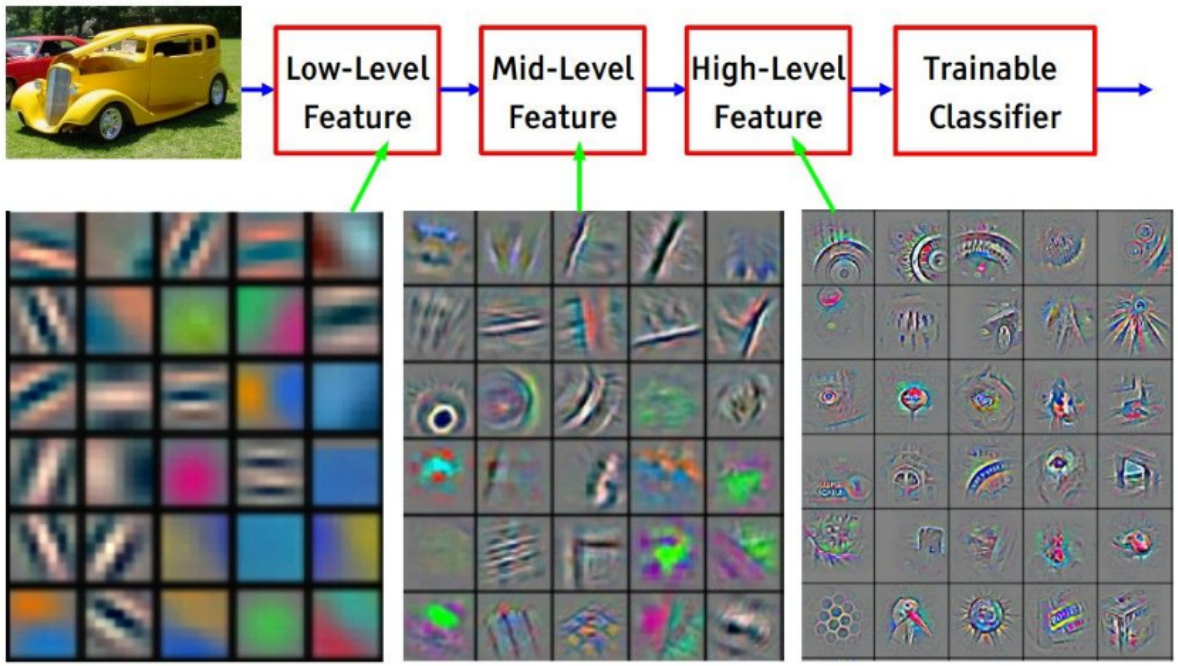
Preview: ConvNet is a sequence of Convolutional Layers, interspersed with activation functions



Convolution layer

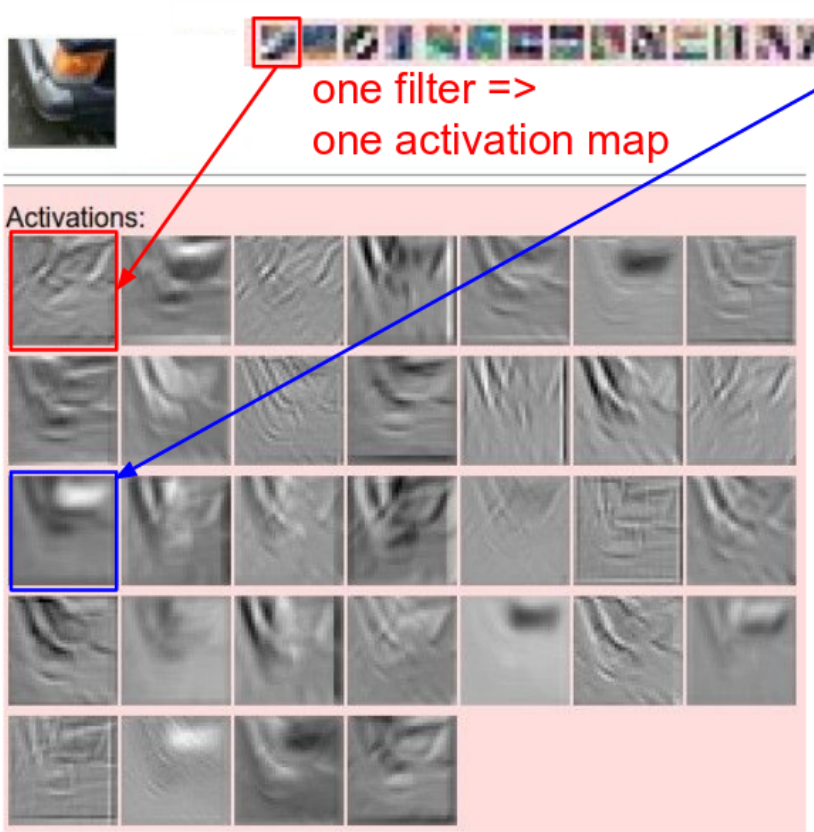
Preview

[From recent Yann LeCun slides]



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

Convolution layer



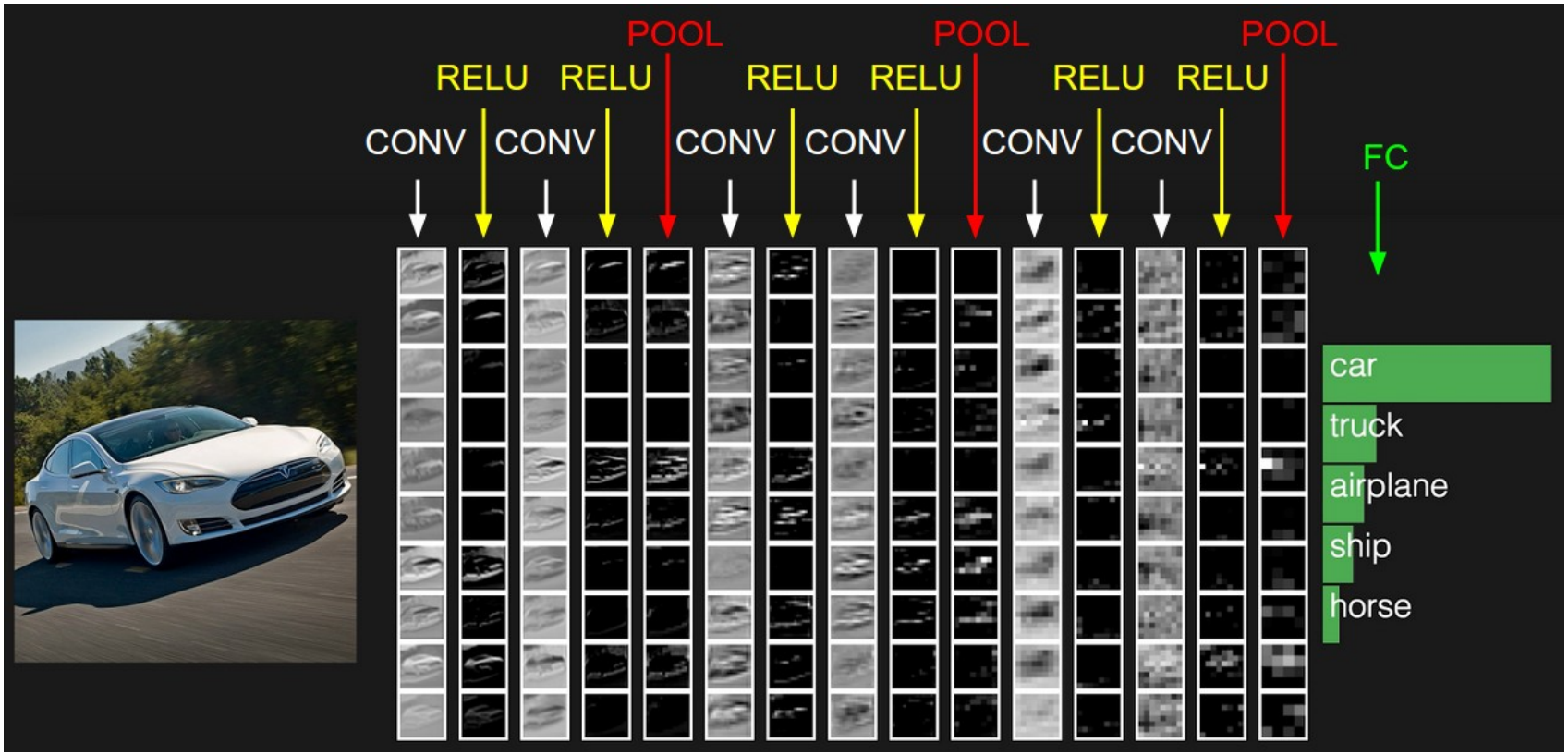
example 5x5 filters
(32 total)

We call the layer convolutional because it is related to convolution of two signals:

$$f[x,y] * g[x,y] = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f[n_1,n_2] \cdot g[x-n_1, y-n_2]$$

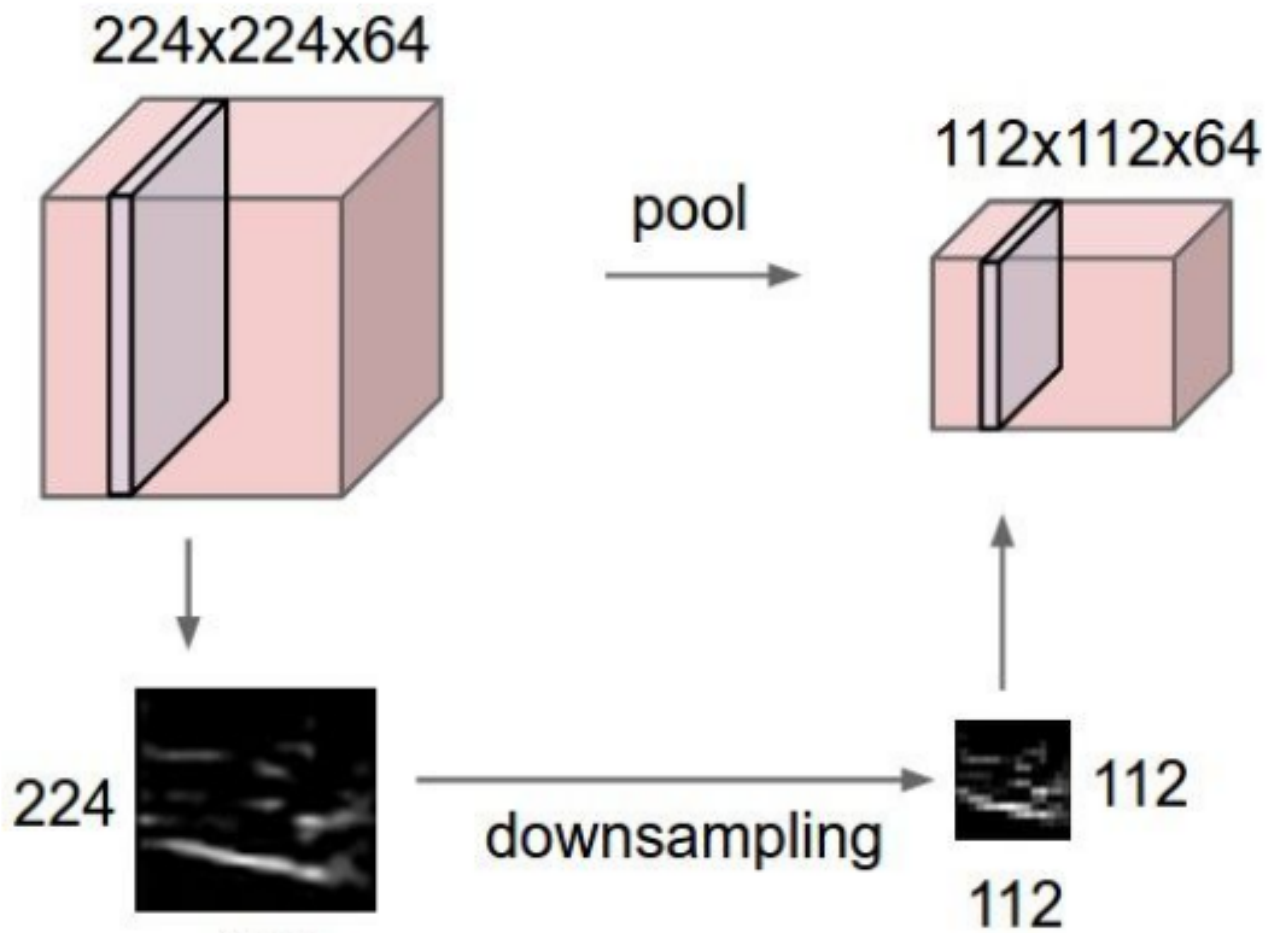
elementwise multiplication and sum of a filter and the signal (image)

Convolutional Neural Network

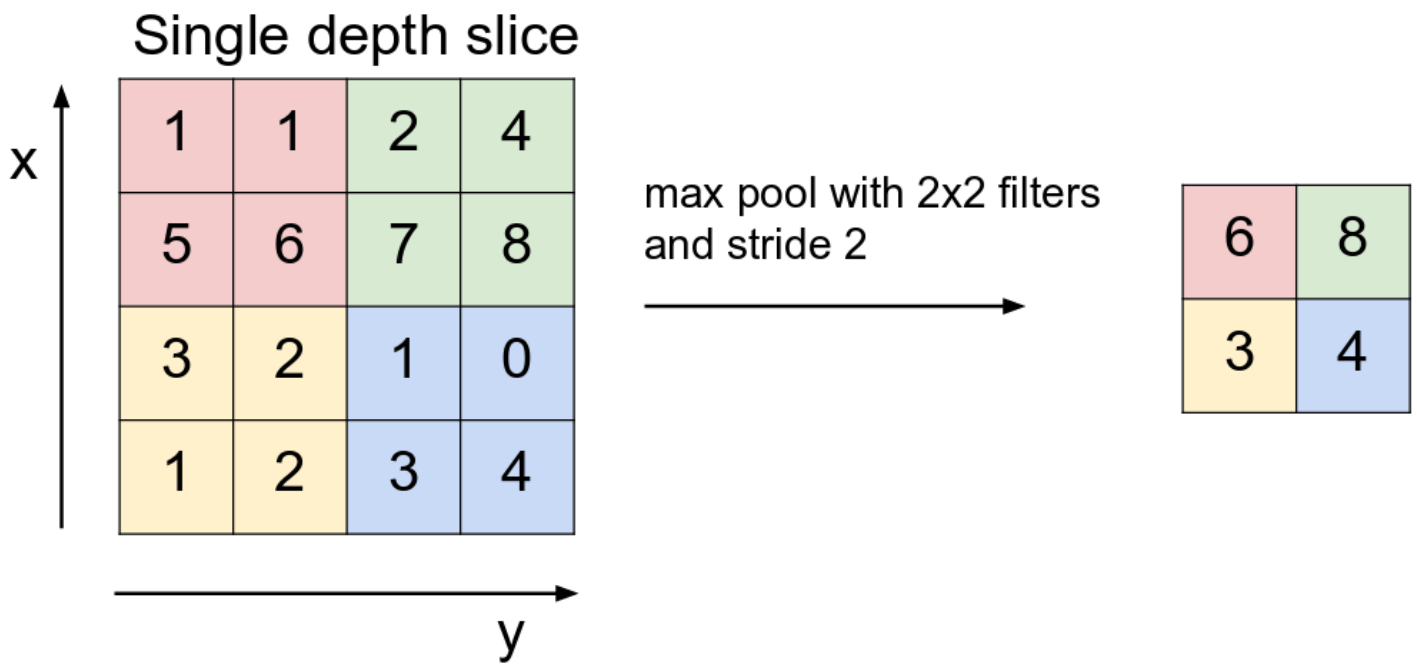


Pooling layer

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

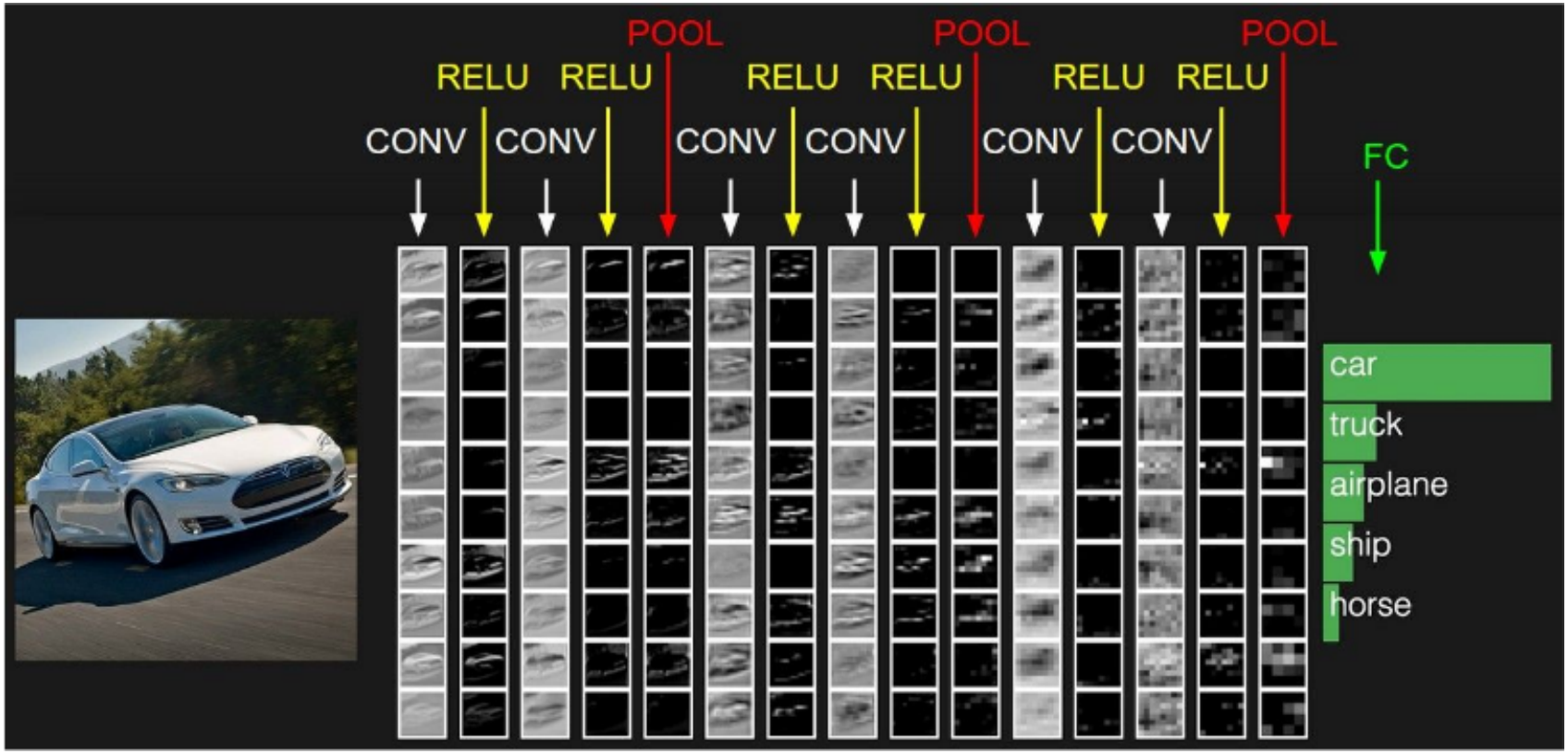


Max pool


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

Fully Connected layers

Contains neuurons that connect to the entire input volume, as in ordinary NN



Final words

Final words

- `http://dlsg.naiveneuron.com`
- ...

References

- `cs231n.stanford.edu/slides/winter1516_lecture6.pdf`
- `cs231n.stanford.edu/slides/winter1516_lecture7.pdf`
- `cs231n.github.io/`
- IRC server *freenode* - channel *#naiveneuron*