

11210IPT553000

Deep Learning in Biomedical Optical Imaging

Week 8

Convolutional Neural Network

Deep convolutional models

Instructor: Hung-Wen Chen
2023/10/30 @NTHU, Fall 2023



deeplearning.ai

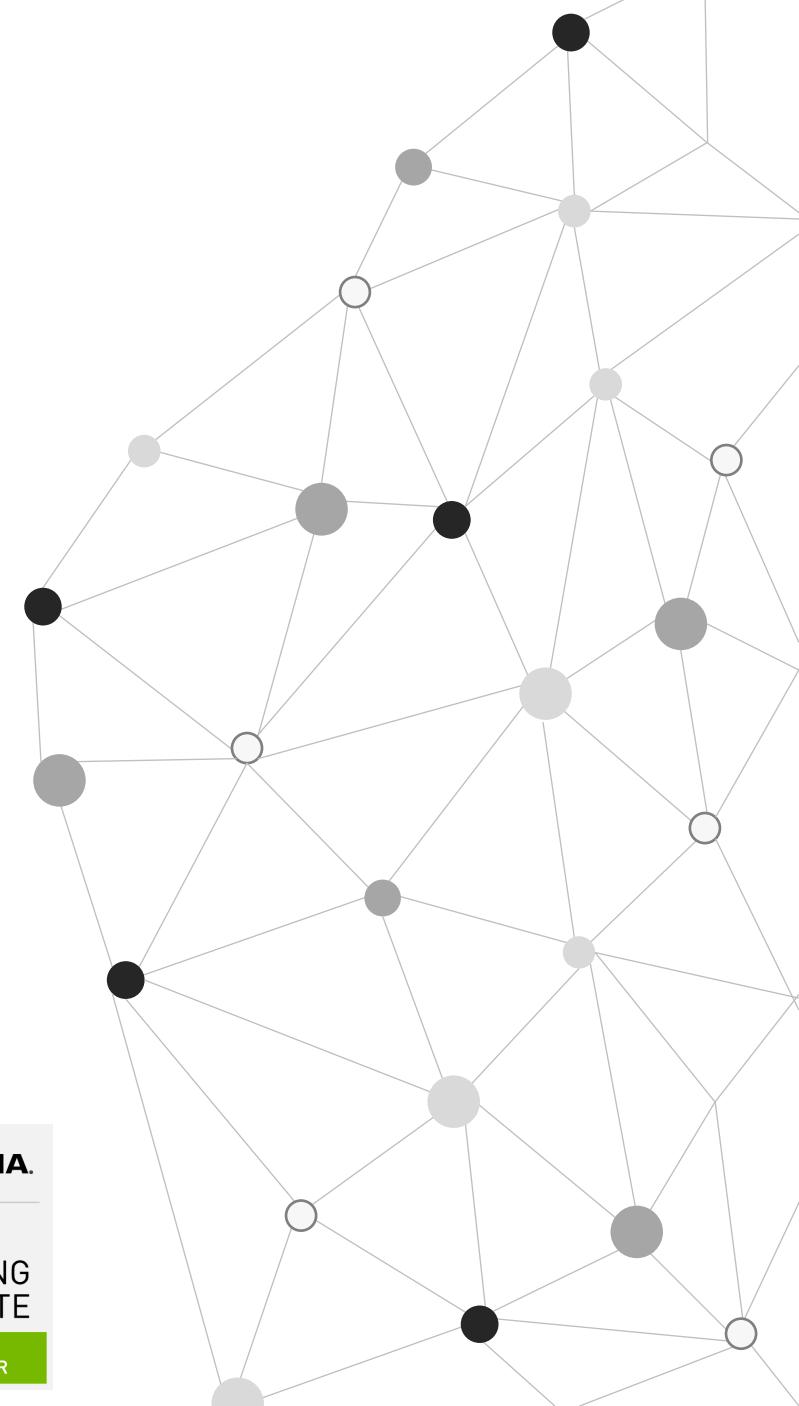
SPONSORED BY



DEEP
LEARNING
INSTITUTE



CERTIFIED
INSTRUCTOR



Calendar

9	11/06	<i>Transformer</i>	<i>Self-attention</i>	<u>Report</u>
10	11/13	<i>AI for Medical Diagnosis + Nvidia Workshop</i>		<u>HW5</u>
11	11/20	<i>Mid-term Review + Introduction to Final Presentation</i>		
12	11/27	<i>Mid-term Exam</i>		
13	12/04	<i>Guest Lecture</i>		
14	12/11	<i>Guest Lecture</i>		
15	12/18	<i>Student Projects Presentation</i>		
16	12/25	<i>Student Projects Presentation</i>		
17	01/01	<i>Holiday - New Year's Day (no class)</i>		
18	01/08	<i>Student Projects Presentation</i>		

- Midterm Review on 11/20 and Midterm on 11/27
- Research Presentation Topic
 - Presentation Week (12/18, 12/25, 1/8)
 - Volunteers of the first week and randomly pick the presentation order on 11/20
 - Pick a research paper published after 2021
 - Send the title and the paper file to TA ASAP
 - Get the confirmation from the lecturer **due on 12/4**

- **Objective**

Provide a detailed analysis of your model implementation on a given image dataset.

- **Content Recommendations:**

- **Method:** Describe the methods, techniques, or algorithms you've used.
- **Performance:** Present the results. How well did your model perform?
Compare with benchmarks if any.
- **Visuals:** Use tables, figures, and diagrams to illustrate your findings and improvements.
- **Conclusion:** Sum up your findings, what you've learned, and possible future extensions or improvements.



- **Abstract**
What is this research about?
- **Motivation / Purpose**
Why do the authors perform this research?
- **Introduction to the Bio-medical Imaging Technology**
- **Network Architecture**
Introduce the deep learning model including its design, pros and cons
- **Comparison**
Compare the performance of this research to other researches or baselines
- **Conclusion**
Your conclusion
- **Code Implementation (Bonus)**
Show your results after running the codes (Submit the files or the link to TA)
Modifications are strongly encouraged!

- Foundations of Convolutional Neural Networks (Course 4 Week 1)
- Deep Convolutional Models (Course 4 Week 2)
- Lab Practice: Unsupervised Learning



A simple convolution network example

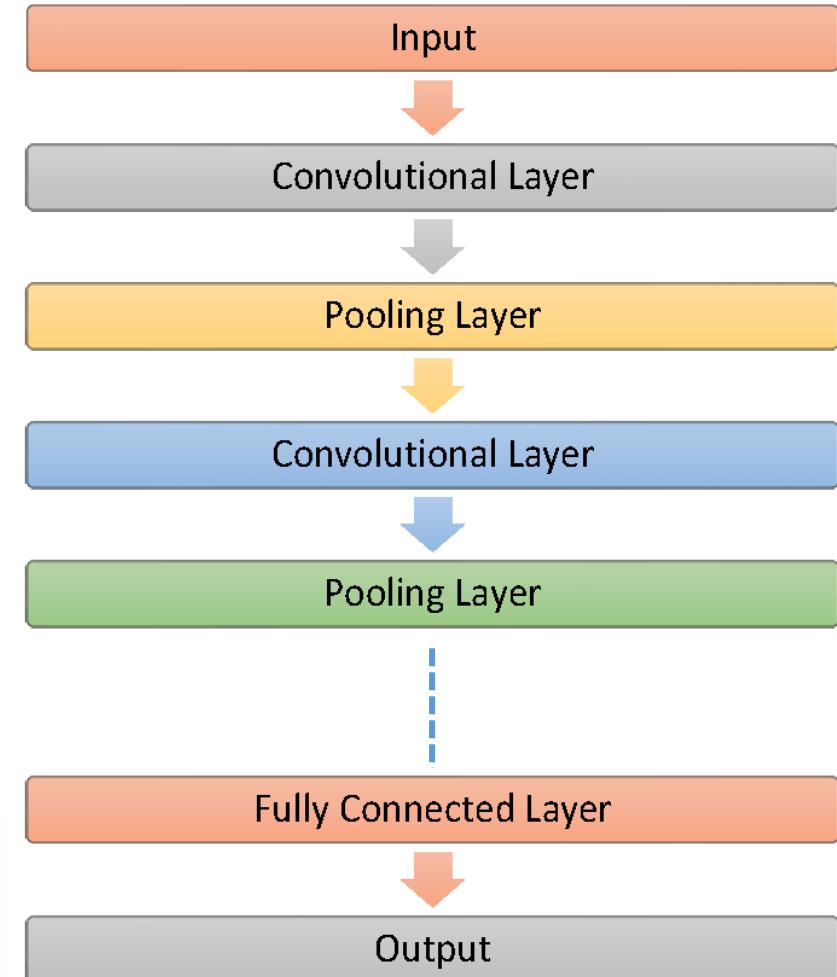
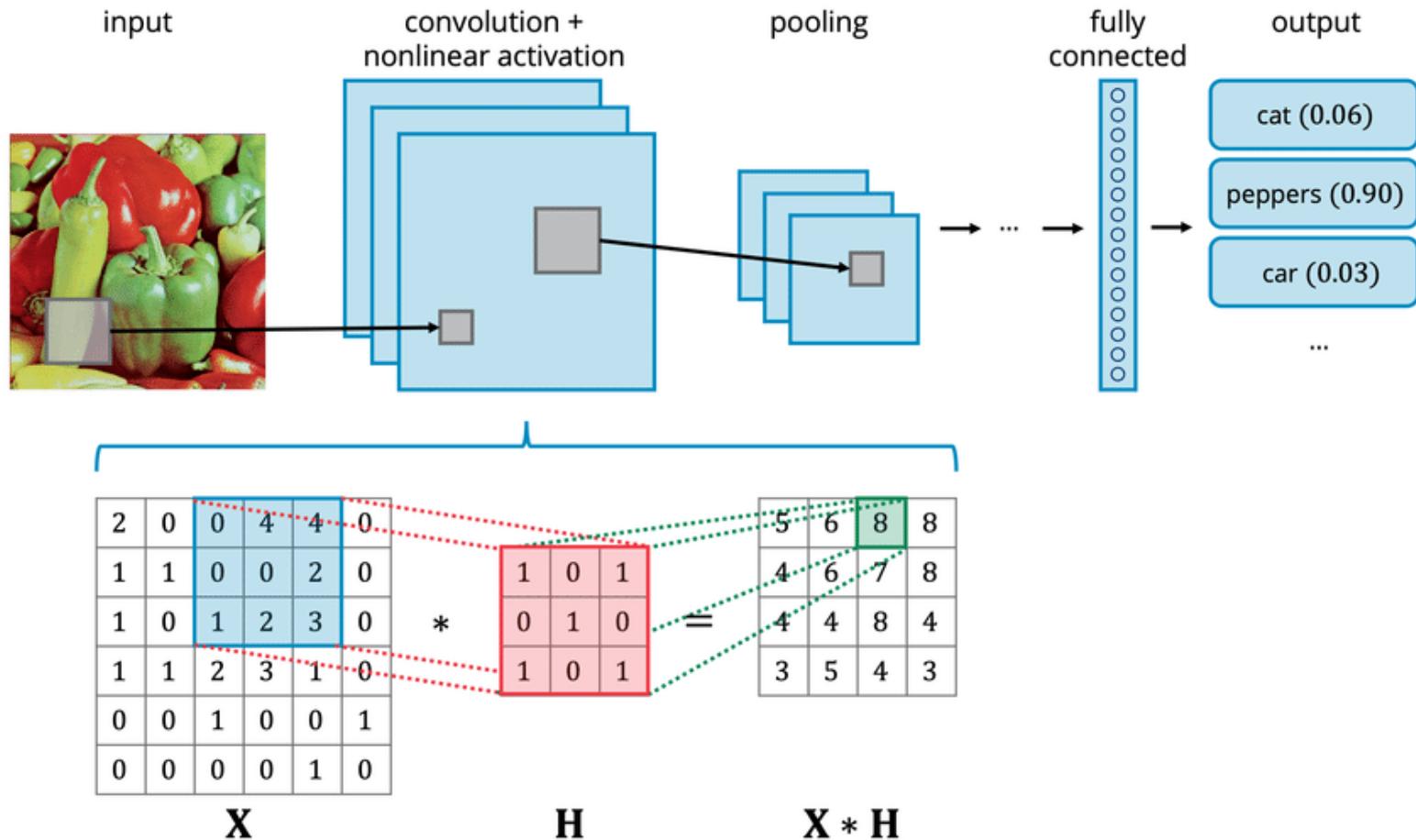
Types of layer in a convolutional network

- Convolution (Conv net)
- Pooling (POOL)
- Fully connected (FC)

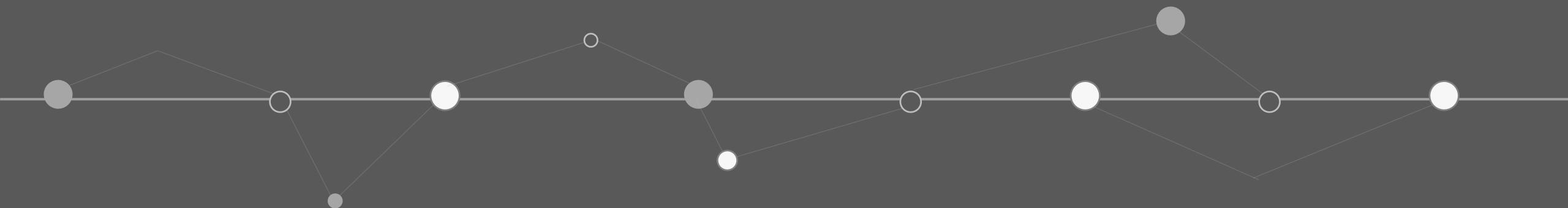


A simple convolution network example

Types of layer in a convolutional network



Pooling Layers



Pooling Layers

Max pooling

1	3	2	1
2	9	1	1
1	3	2	3
5	6	1	2

4×4

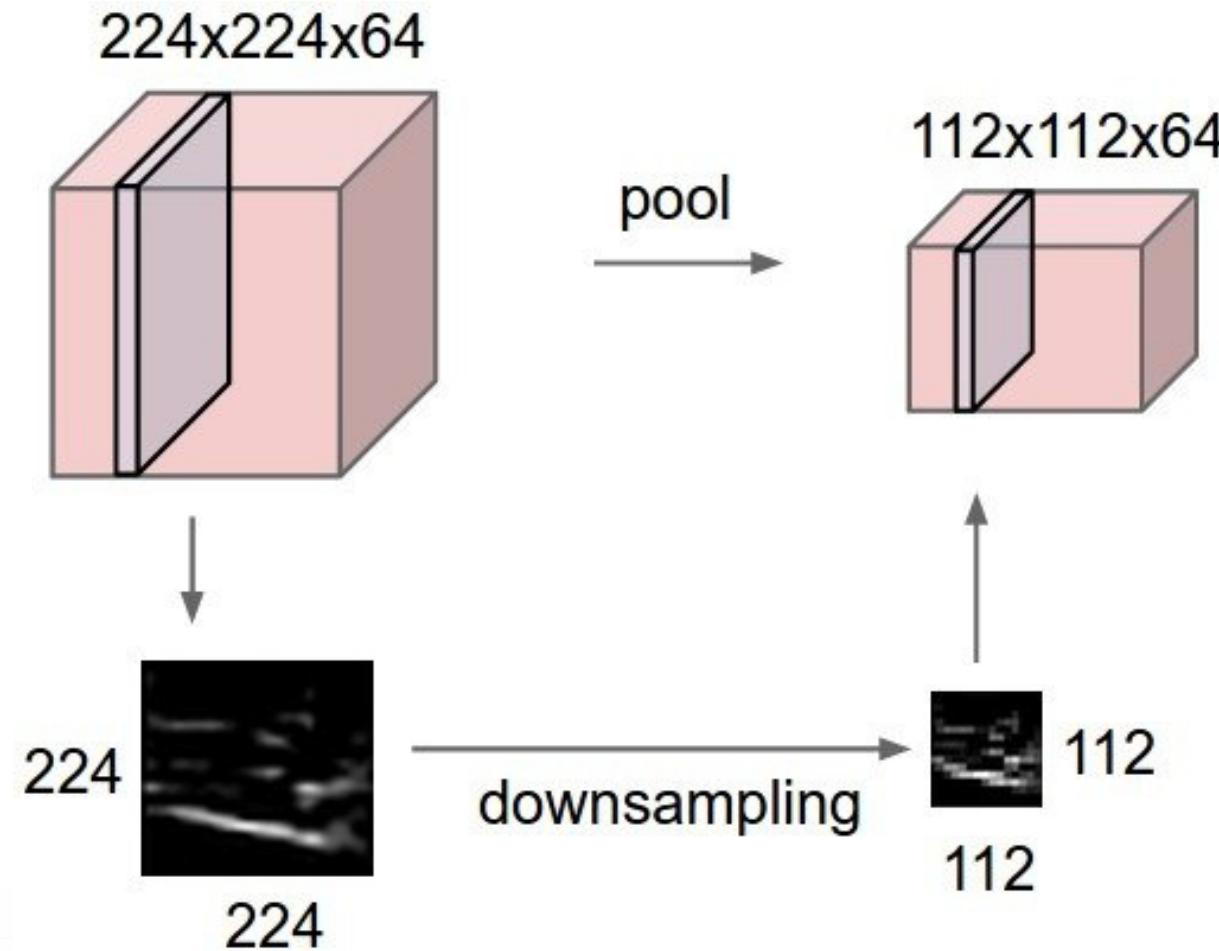
2×2 filter
stride $s = 2$
padding $p = 0$

9	2
6	3

10

Pooling Layers

Max pooling



In this example, a convolutional layer produced a 224×224 activation map with 64 filters, which is downsampled to $112 \times 112 \times 64$. This saves a lot of computational resources!

Pooling Layers

Max pooling

1	3	2	1	3
2	9		1	5
1				2
8	3		1	0
5	6	1	2	9

$5 \times 5 \times nc$



3×3 filter
stride $s = 1$
padding $p = 0$

9	9	5
9	9	5
8	6	9

$3 \times 3 \times nc$



Pooling Layers

Average pooling

1	3	2	1
2	9	1	1
1	4	2	3
5	6	1	2

$$\xrightarrow{\hspace{1cm}} \begin{matrix} f = 2 \\ s = 2 \end{matrix}$$

3.75	1.25
4	2



Summary of pooling

Hyperparameters:

f : filter size

s : stride

Max or average pooling (usually zero padding)

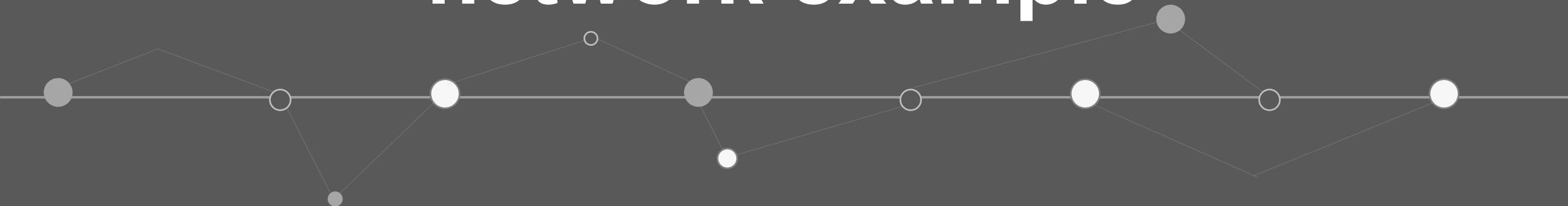
No parameters to learn!

Input: $n_H \times n_w \times n_c$

Output: $n'_H \times n'_w \times n_c$

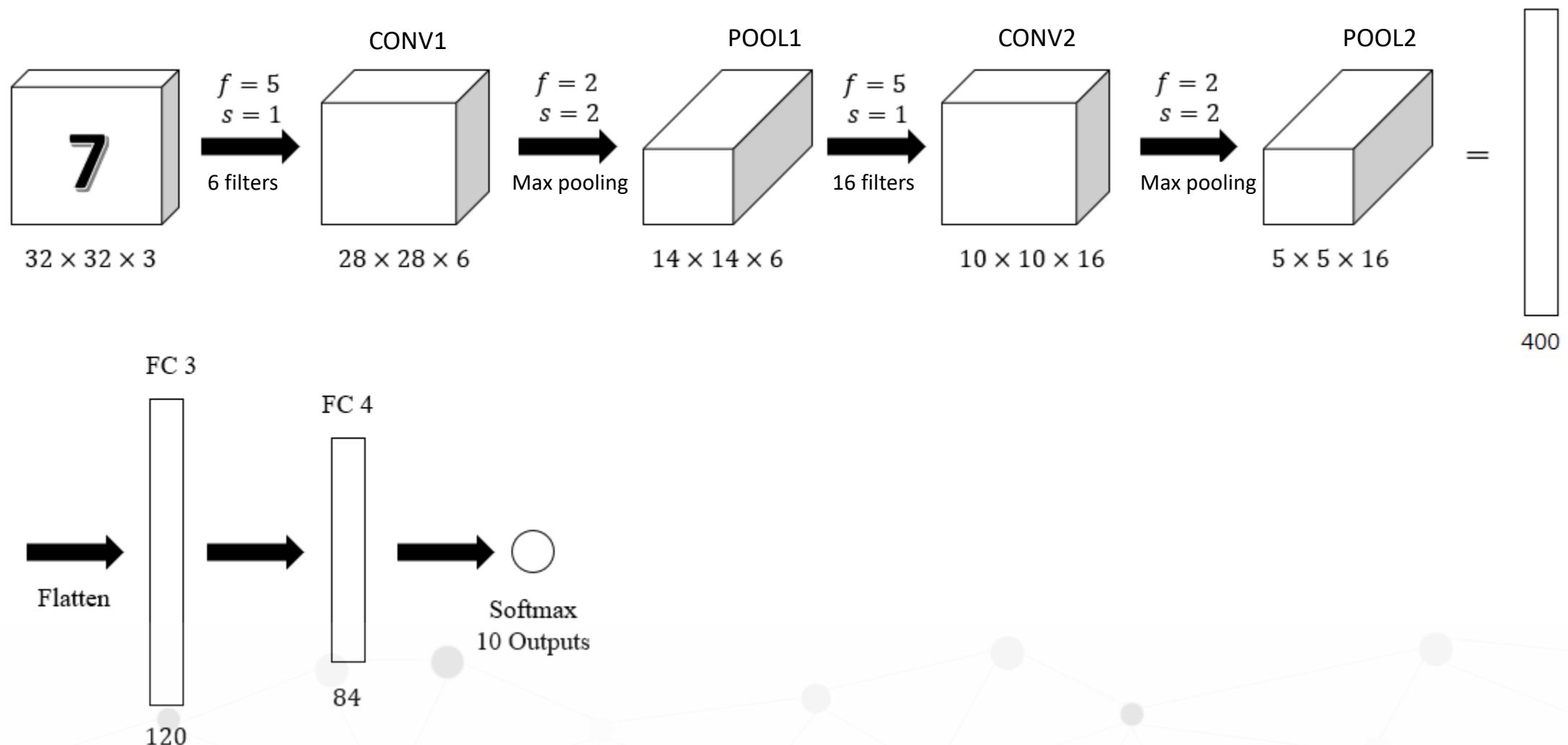
$$n'_H = \left\lfloor \frac{n_H - f}{s} + 1 \right\rfloor$$

Convolutional neural network example



Convolutional neural network example

Neural network example (~LeNet-5)

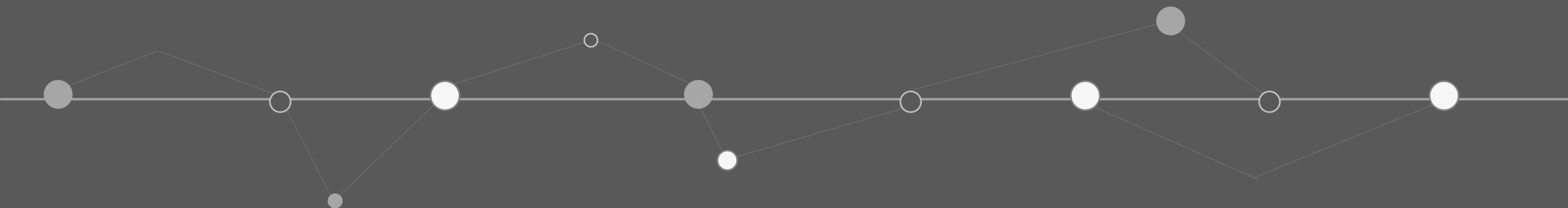


Convolutional Neural Network Example

Neural network example

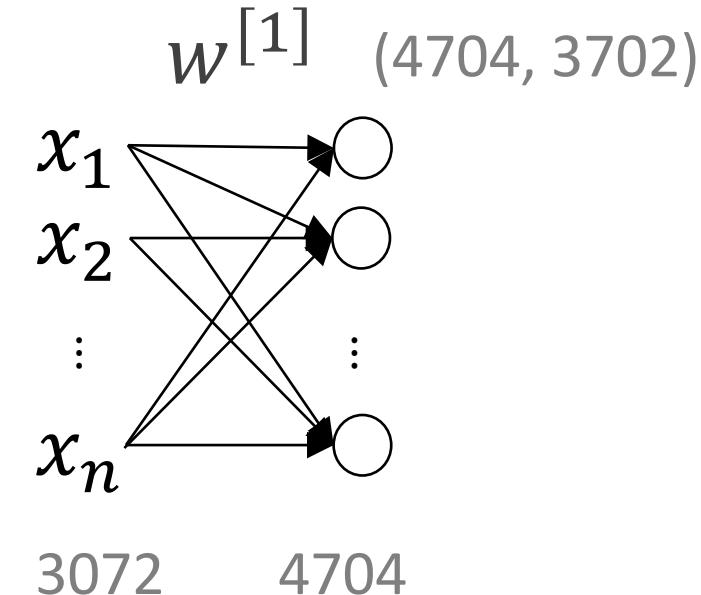
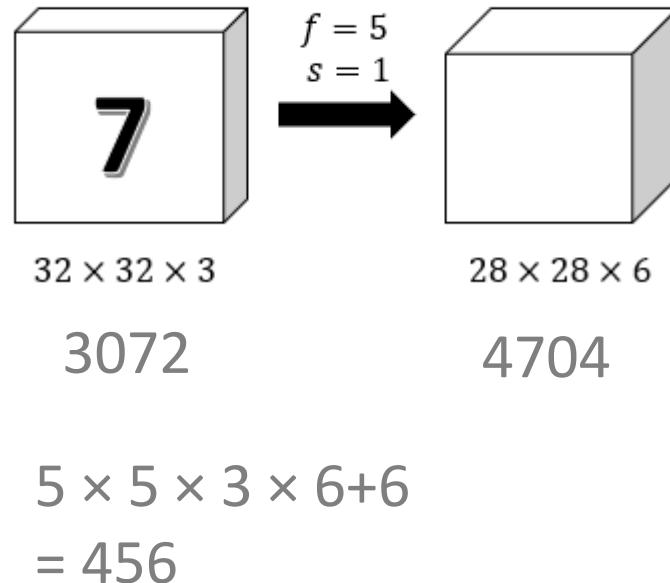
	Activation shape	Activation Size	# parameters
Input:	(32,32,3)	3,072 ($a^{[0]}$)	0
CONV1 (f=5, s=1)	(28,28,6)	4,704	456 $5*5*3*6+6$
POOL1	(14,14,6)	1,176	0
CONV2 (f=5, s=1)	(10,10,16)	1,600	2,416 $5*5*6*16+16$
POOL2	(5,5,16)	400	0
FC3	(120,1)	120	48,120 $400*120+120$
FC4	(84,1)	84	10,164 $120*84+84$
Softmax	(10,1)	10	850 $84*10+10$

Why Convolutions?



Why Convolutions?

Do we really need “fully connected” in image processing?



$$\begin{aligned} & 3702 \times 4704 + 4704 \\ &= 17,418,912 \\ &\sim 17M \end{aligned}$$

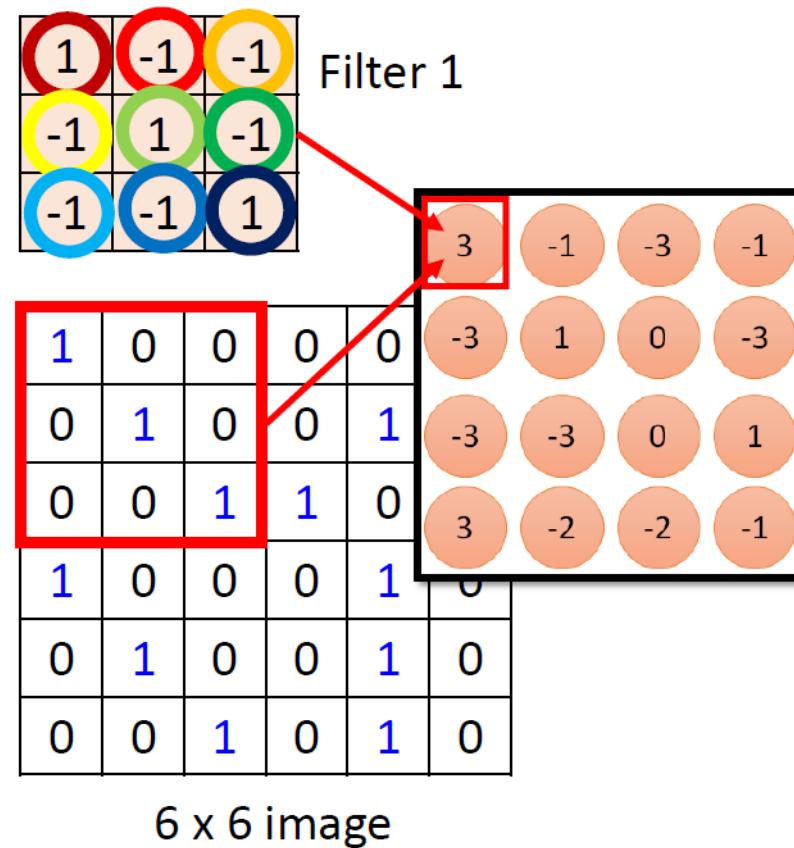
$$\begin{array}{|c|c|c|c|c|c|c|} \hline
 10 & 10 & 10 & 0 & 0 & 0 & 0 \\ \hline
 10 & 10 & 10 & 0 & 0 & 0 & 0 \\ \hline
 10 & 10 & 10 & 0 & 0 & 0 & 0 \\ \hline
 10 & 10 & 10 & 0 & 0 & 0 & 0 \\ \hline
 10 & 10 & 10 & 0 & 0 & 0 & 0 \\ \hline
 10 & 10 & 10 & 0 & 0 & 0 & 0 \\ \hline
 10 & 10 & 10 & 0 & 0 & 0 & 0 \\ \hline
 \end{array}
 \quad *
 \quad
 \begin{array}{|c|c|c|} \hline
 1 & 0 & -1 \\ \hline
 1 & 0 & -1 \\ \hline
 1 & 0 & -1 \\ \hline
 \end{array}
 \quad =
 \quad
 \begin{array}{|c|c|c|c|} \hline
 0 & 30 & 30 & 0 \\ \hline
 0 & 30 & 30 & 0 \\ \hline
 0 & 30 & 30 & 0 \\ \hline
 0 & 30 & 30 & 0 \\ \hline
 \end{array}$$

Parameter sharing: A feature detector (such as a vertical edge detector) that's useful in one part of the image is probably useful in another part of the image.

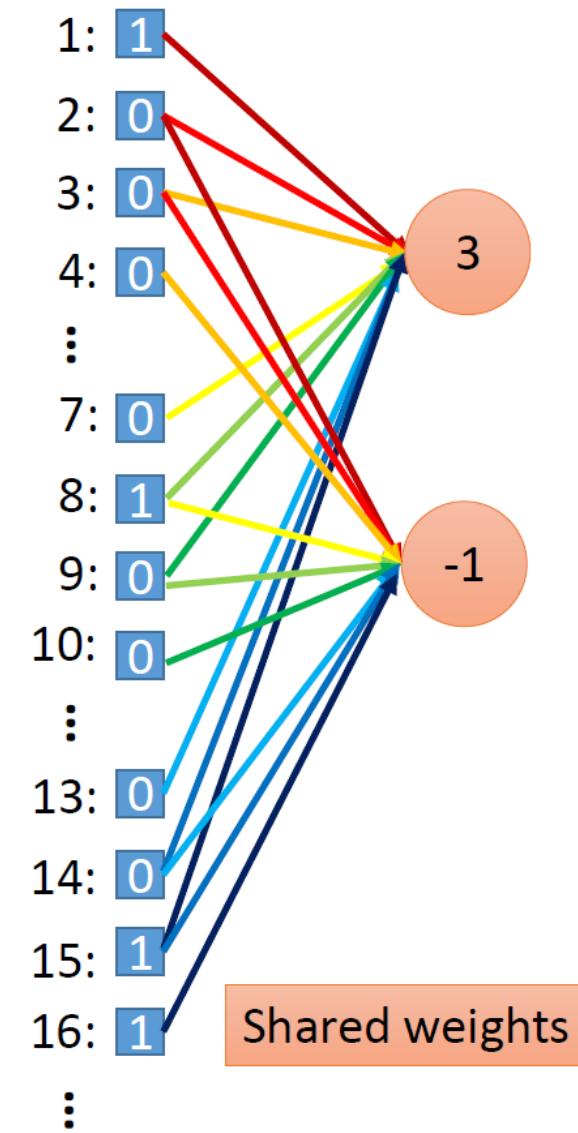
Sparsity of connections: In each layer, each output value depends only on a small number of inputs.

Why Convolutions?

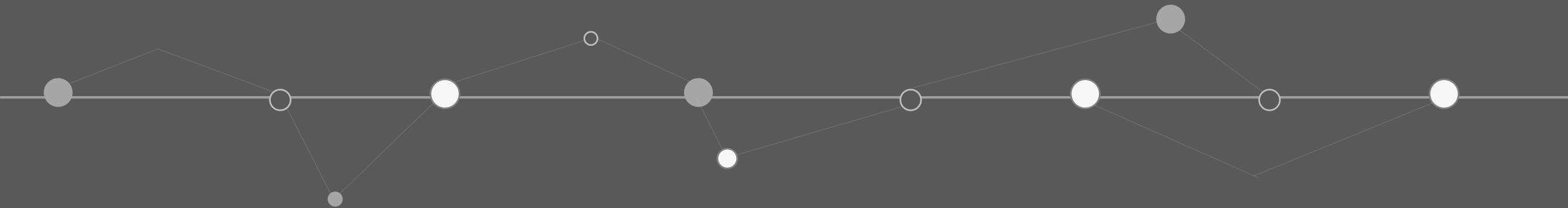
Do we really need “fully connected” in image processing?



Less parameters!



IMAGENET



Outlines

1998

2012

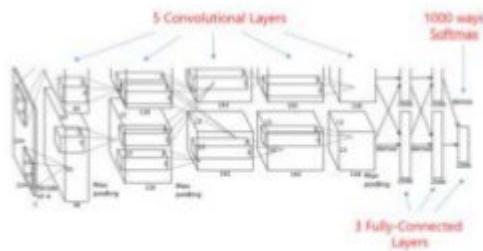
2014

2015

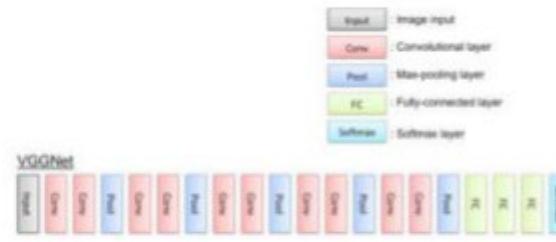


CNN Architectures

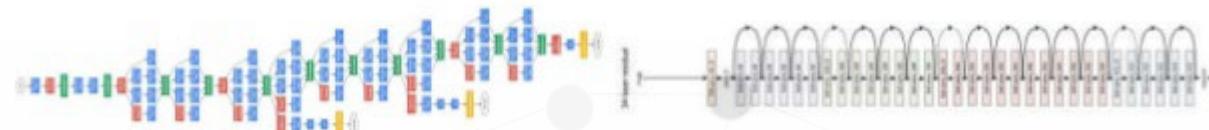
AlexNet



VGG



GoogLeNet



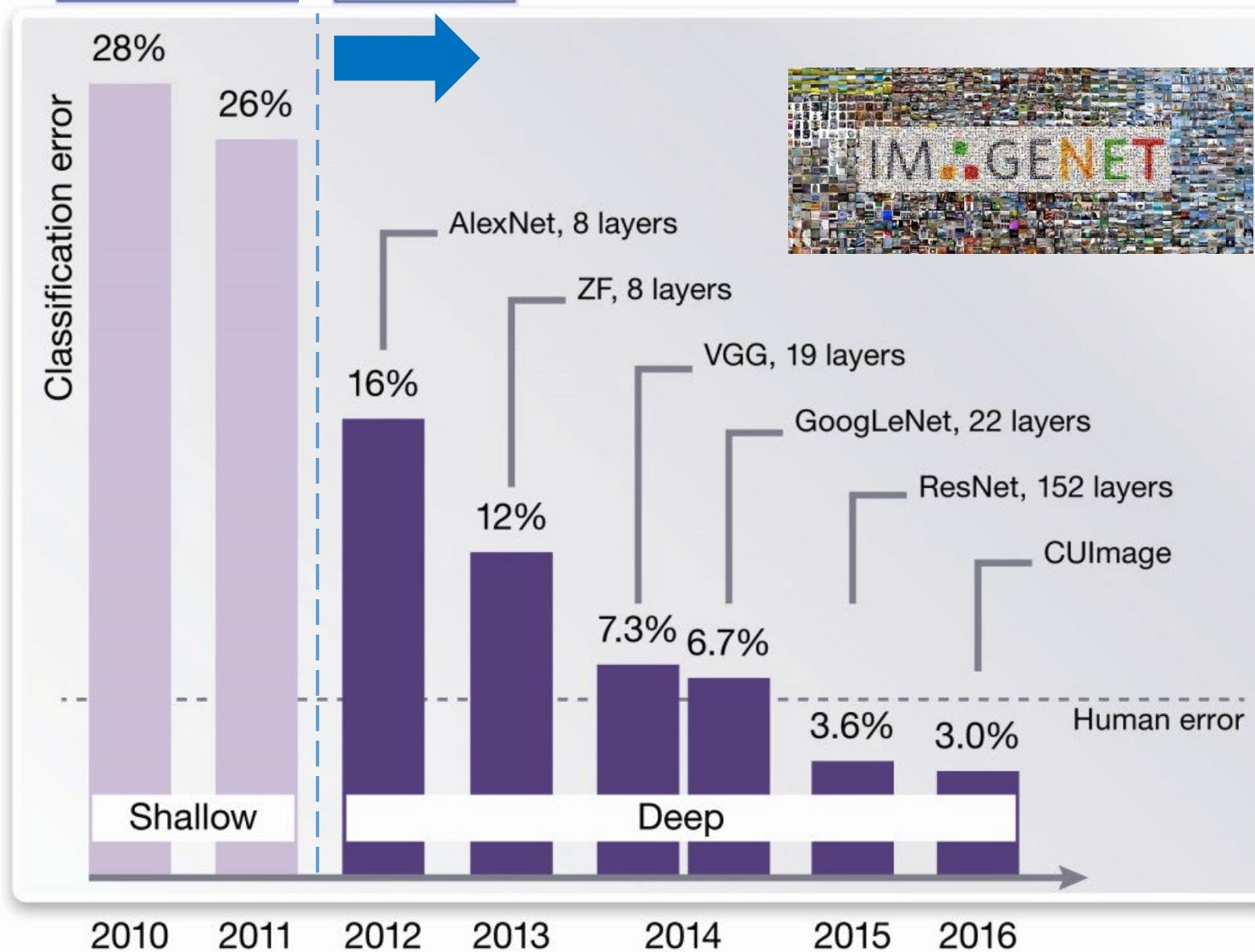
- LeNet-5
- AlexNet
- VGG

ResNet (Skip connection)

Inception (1x1 convolution)

Classic Networks

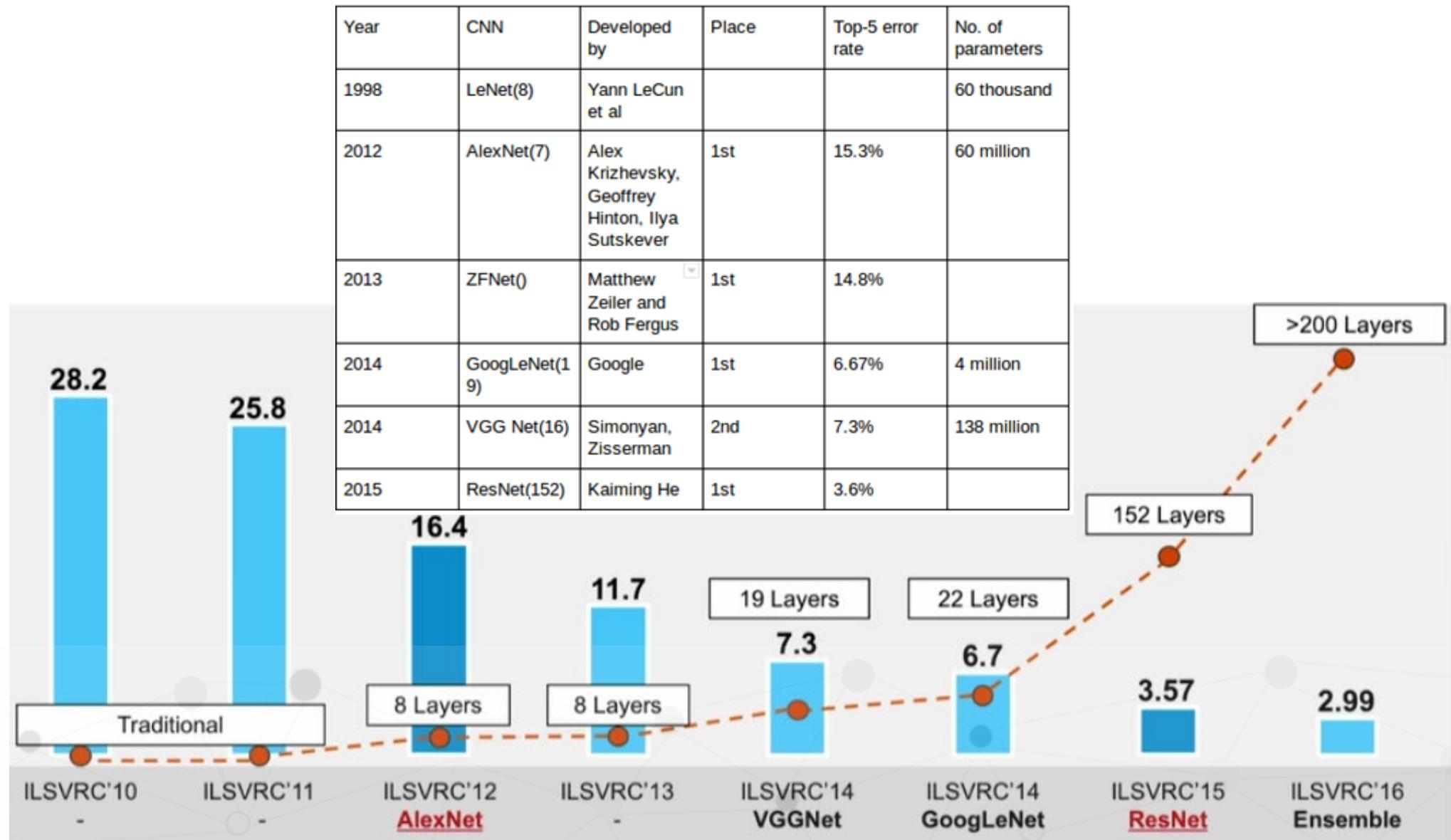
Image classification on IMAGENET



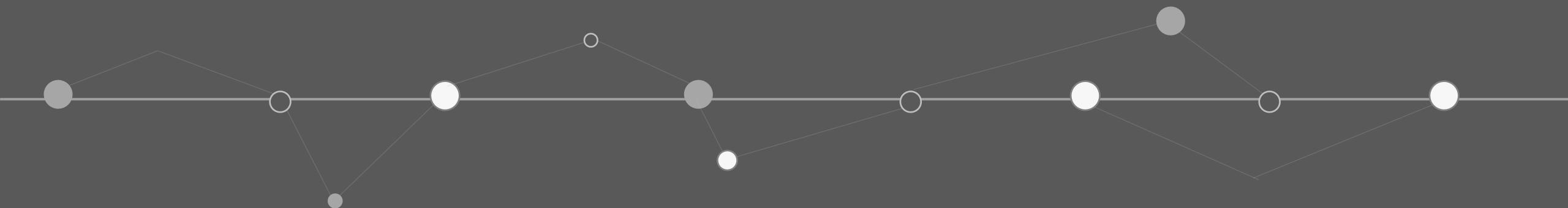
1000 object classes
Images (150GB):
1.2M training set
100k test set
~1000 examples per class

Classic Networks

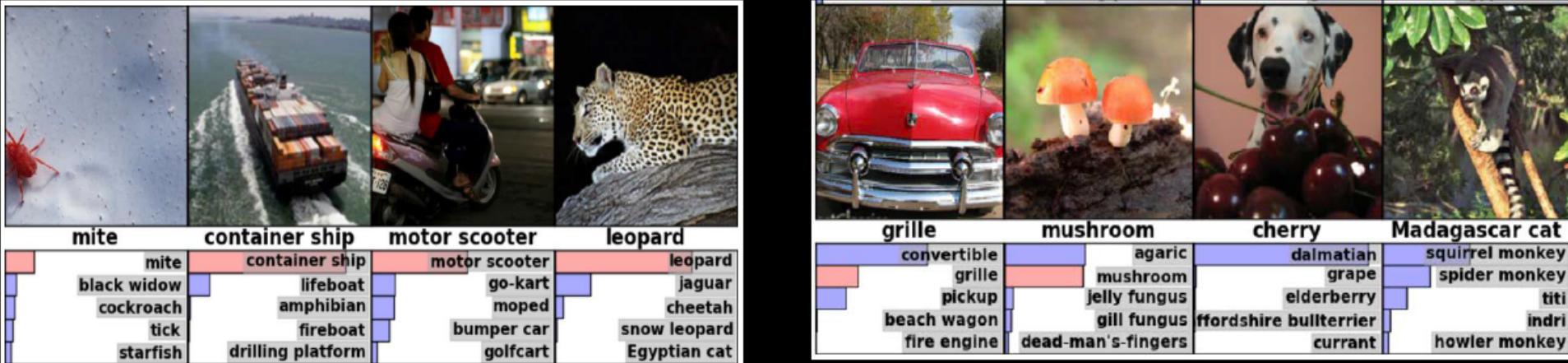
Parameters and numbers of layers



The Human Brain



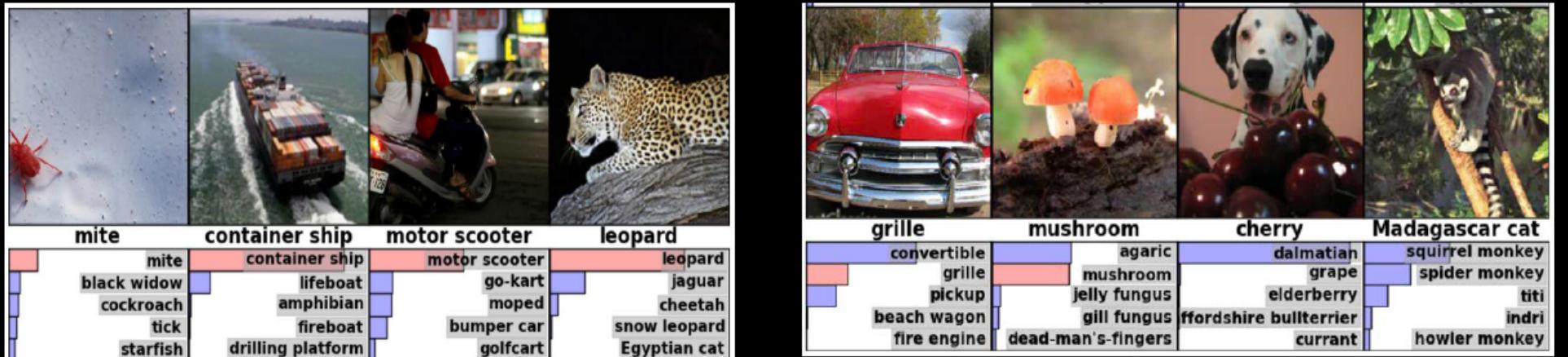
Why Should we Study the Human Brain in the first place?



Amazing, and transformative for CS and cog sci and neuroscience.
These networks can be taken as computationally precise models of obj rec.
But do they really perform as well as humans?

Various images above and on the following page © sources unknown. All rights reserved. This content is excluded from OCW's Creative Commons license. For more information, see <https://ocw.mit.edu/fairuse>.

Why Should we Study the Human Brain in the first place?



Amazing, and transformative for CS and cog sci and neuroscience.
These networks can be taken as computationally precise models of obj rec.
But do they really perform like humans?

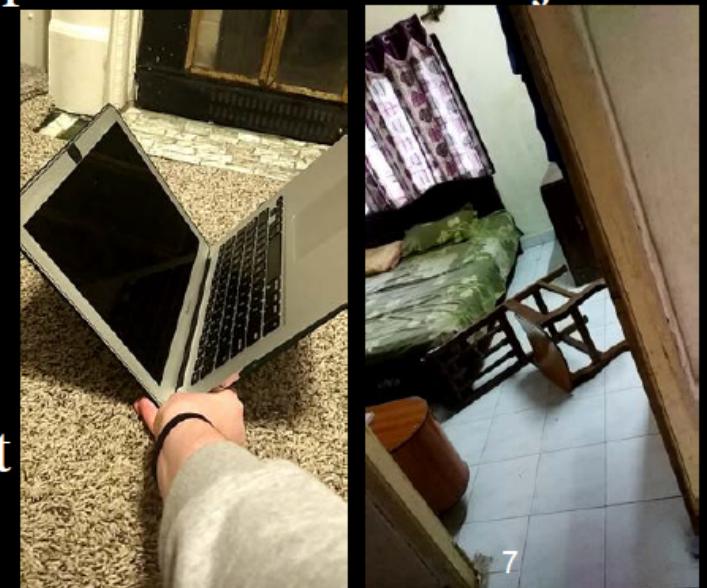
What if tested on images not in ImageNet?

Katz & Barbu:

More variable locs/orientations than Imagenet.

Human performance is still good, but...

Accuracy of ResNet-200 drops from 71% correct (ImageNet) to 25% correct (Katz/Barbu imgs).



 Home  Moments

Search Twitter

A white egg on a solid colored background

TWEETS

3,164

FOLLOWERS

3,723

picdescbot

@picdescbot

I'm a bot that describes random pictures from wikimedia commons using Microsoft Cognitive Services. Operated by @elad3

 picdescbot.tumblr.com/about

 Joined March 2016

 3,160 Photos and videos



Tweets

Tweets & replies

Media

 Pinned Tweet

 picdescbot @picdescbot · 24 Jun 2016
a dinosaur on top of a surfboard



4

10

457

t

715

l



SAY GOODBYE
TO EMAIL
PING-PONG

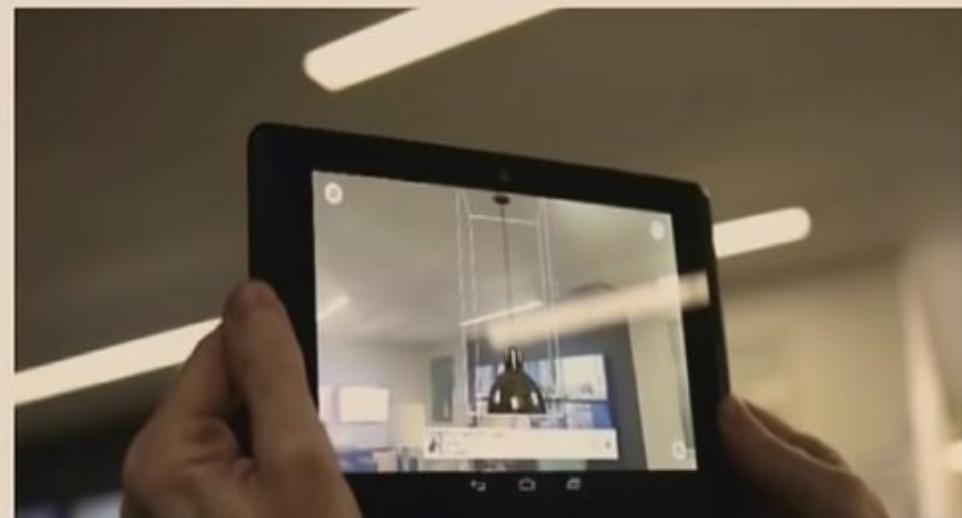
Hire your AI Assistant >

Google's AI can now caption images almost as well as humans

BY JAMES WALKER SEP 23, 2016 IN TECHNOLOGY

LISTEN | PRINT

Google has announced a new version of its image captioning algorithm that describes the contents of images with 94 percent accuracy. It's almost as good at writing captions as humans are. It has been trained to emulate descriptions written by real people.



x.ai
The smart bots are coming and this one is brilliant.
THE VERGE

Hire your AI Assistant >

TOP NEWS

LATEST NEWS



Op-Ed: After Jeff Horn, who's next for Manny Pacquiao?

Like 4 Share



Data shows oxygen level in world's oceans has been declining

Like 9 Share



picdescbot @picdescbot · 3h

a group of people on a field playing football



A white egg on a soft
covered background

picdescbot @picdescbot · May 2
a large ship in the water



1

1

2

A white egg on a solid colored background

picdescbot @picdescbot · Apr 25

a car parked in a parking lot



1

7

7

picdescbot @picdescbot · May 1

a group of people that are standing in the grass near a bridge



3

picdescbot @picdescbot · Apr 30

a group of people standing on top of a boat



picdescbot @picdescbot · 9h

a building with a cake

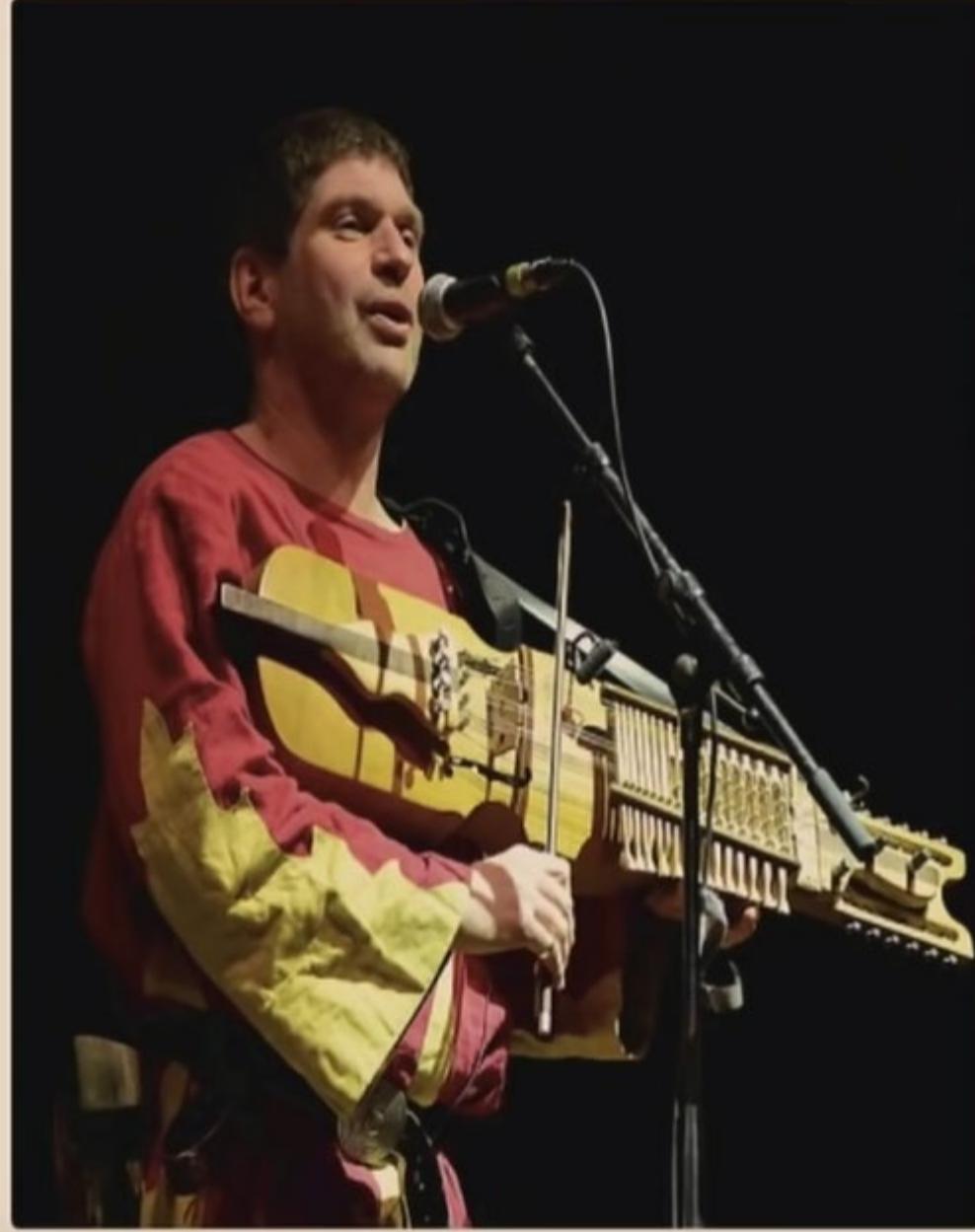


1

7

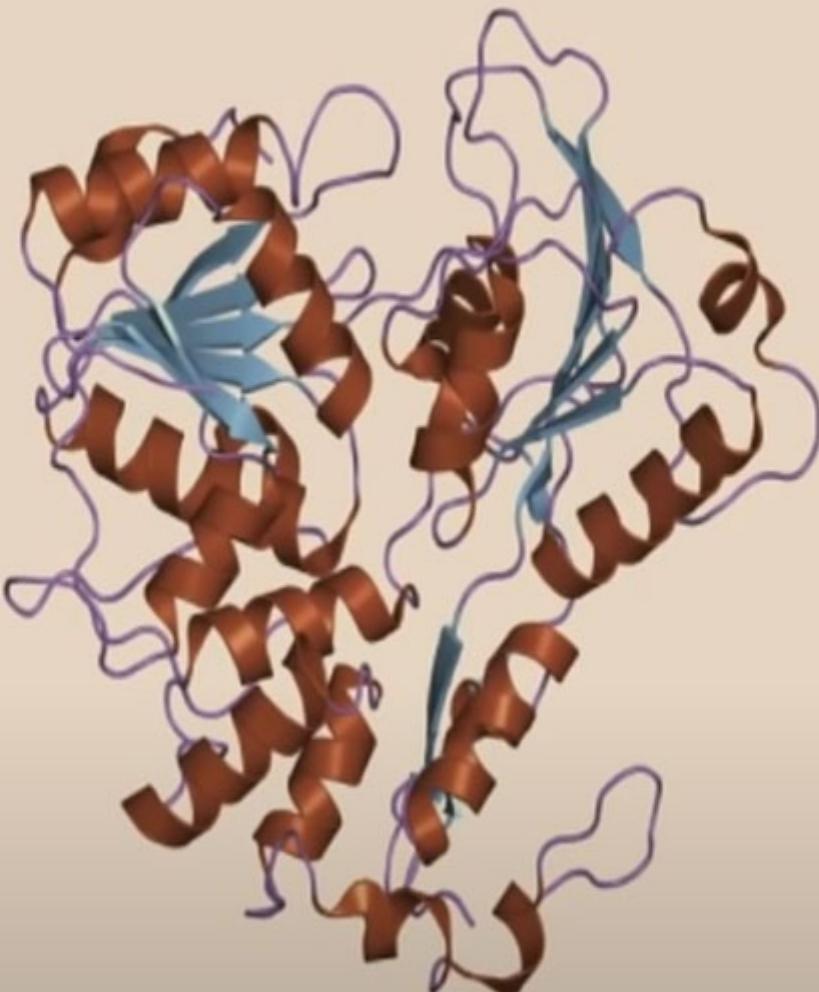
14

picdescbot @picdescbot · 2h
a person holding a cell phone



picdescbot @picdescbot · 21h

a necklace made of bananas



2

40

92

I think it's a group of people standing next to a man in a suit and tie.



Official White House photo by Pete Souza. Image is in the public domain.

Big idea (Tenenbaum):
Both humans & deep nets
are good at *pattern
recognition*.

What humans, but not
machines, are good at:
*building models to
understand the world*.

e.g. What do some people
here know but the guy on
the scale doesn't?

Are we done with ImageNet?
<https://arxiv.org/abs/2006.07159>

A white egg on a solid colored background

TWEETS

3,164

FOLLOWERS

3,723

picdescbot

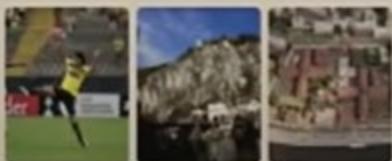
@picdescbot

I'm a bot that describes random pictures from wikimedia commons using Microsoft Cognitive Services. Operated by @elad3

picdescbot.tumblr.com/about

Joined March 2016

3,160 Photos and videos



describe the photo



The image depicts a skeletal illustration of a dinosaur. The dinosaur is bipedal, standing on two legs, with a long tail extended behind it and a long neck stretching forward. Its forelimbs are much shorter than its hind limbs. The illustration provides a clear view of the bones that make up the dinosaur's skeletal structure. There's also a scale bar at the bottom for reference.

picdescbot @picdescbot · May 2

a large ship in the water



1

1

2



describe



The image captures a waterscape scene in an urban setting. In the foreground, there's a tugboat, colored predominantly in blue and red, floating on the water. The tugboat is equipped with black tires attached to its side, presumably for buffering against other vessels or the shore. Behind the boat, the skyline consists of tall buildings, some of which are under construction with visible cranes. On the right side, there are distinct yellow triangular structures, resembling the masts of a large stadium or entertainment venue. The sky above is cloudy, suggesting overcast weather, and there are a few birds flying in the distance.

picdescbot @picdescbot · Apr 25

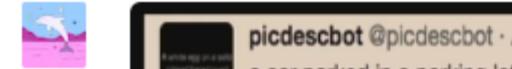
a car parked in a parking lot



1

1

7



picdescbot @picdescbot · Apr 25
a car parked in a parking lot



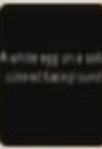
1

1

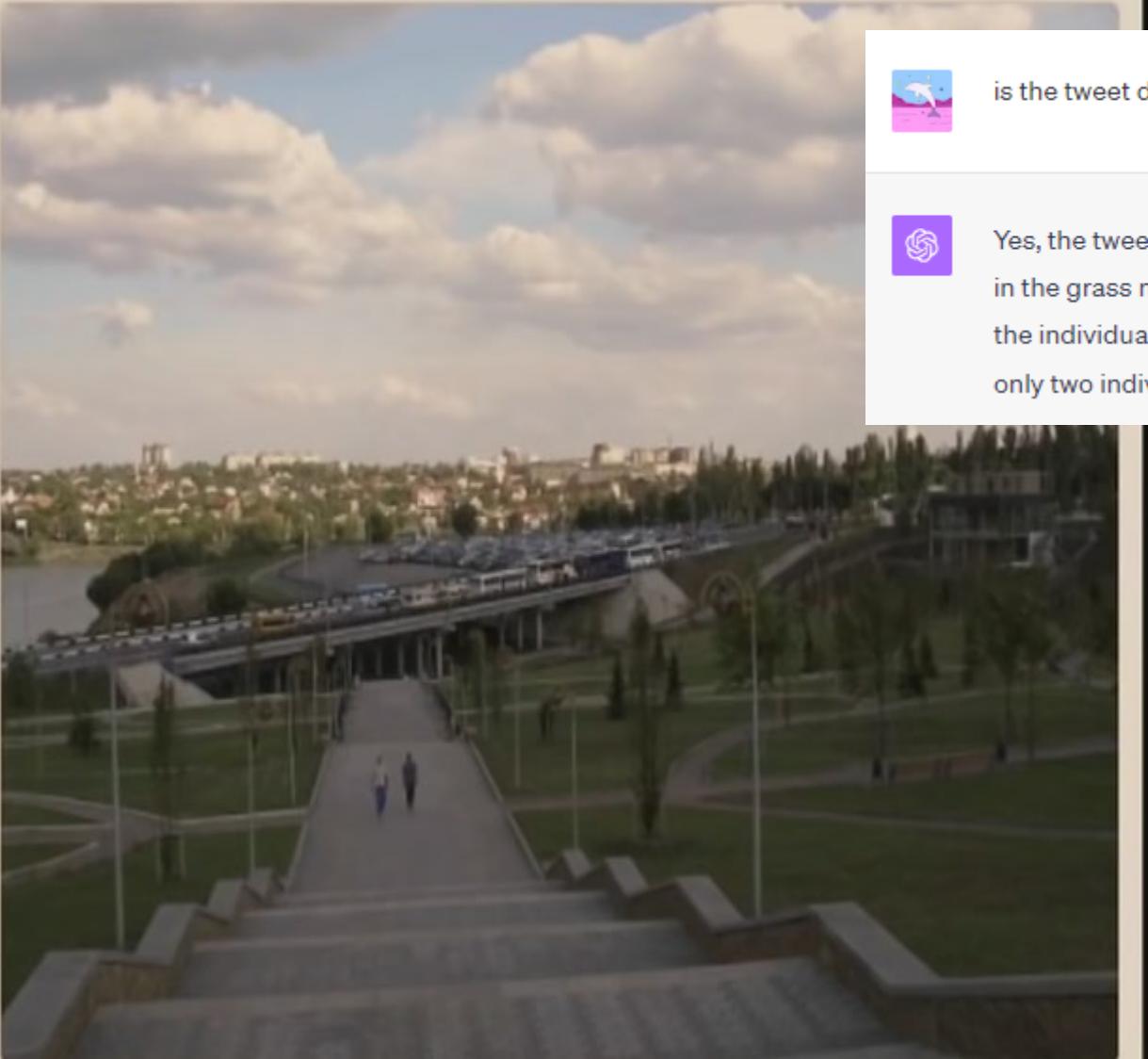
7



The image shows a screenshot of a tweet from the user "@picdescbot" dated Apr 25. The description in the tweet reads "a car parked in a parking lot." The accompanying image features a classic white car, possibly a Mercedes-Benz given the distinct emblem on its hood. The car has a chrome grille, round headlights, and a visible license plate reading "VGF 003". It's parked beside a gray wall, and there seems to be a dark-colored object (possibly another vehicle) to its left. The tweet has received 1 retweet and 7 likes. The top left corner has a descriptive overlay, which states "A white car on a multicolored background".

picdescbot @picdescbot · May 1

a group of people that are standing in the grass near a bridge



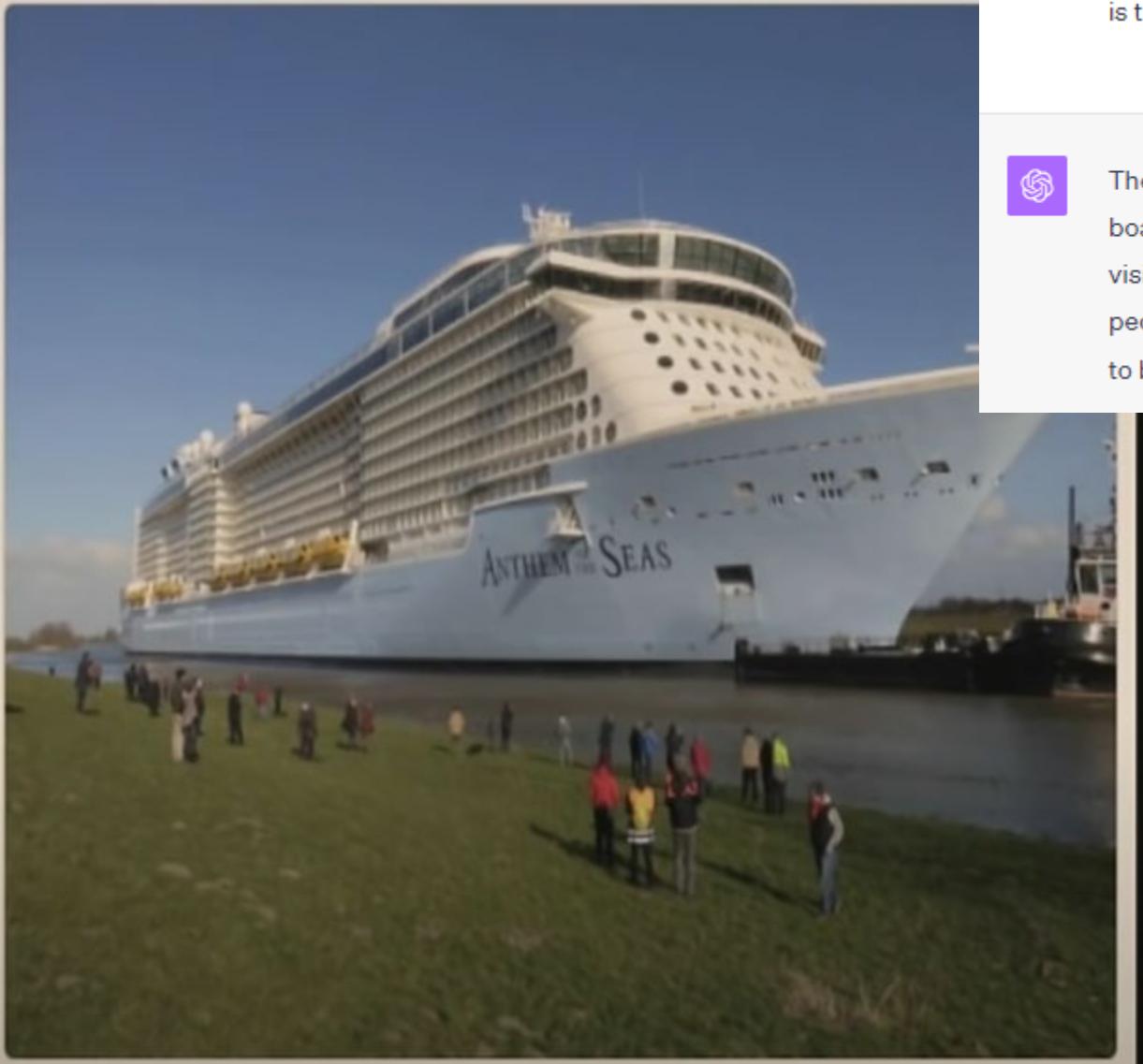
is the tweet description correct?



Yes, the tweet description from "@picdescbot" that says "a group of people that are standing in the grass near a bridge" is partly correct. There is a bridge visible in the image. However, the individuals shown are walking on a pathway, not standing in the grass. Additionally, it's only two individuals, so referring to them as a "group" might be a bit of an overstatement.

 picdescbot @picdescbot · Apr 30

a group of people standing on top of a boat



is the tweet description correct?

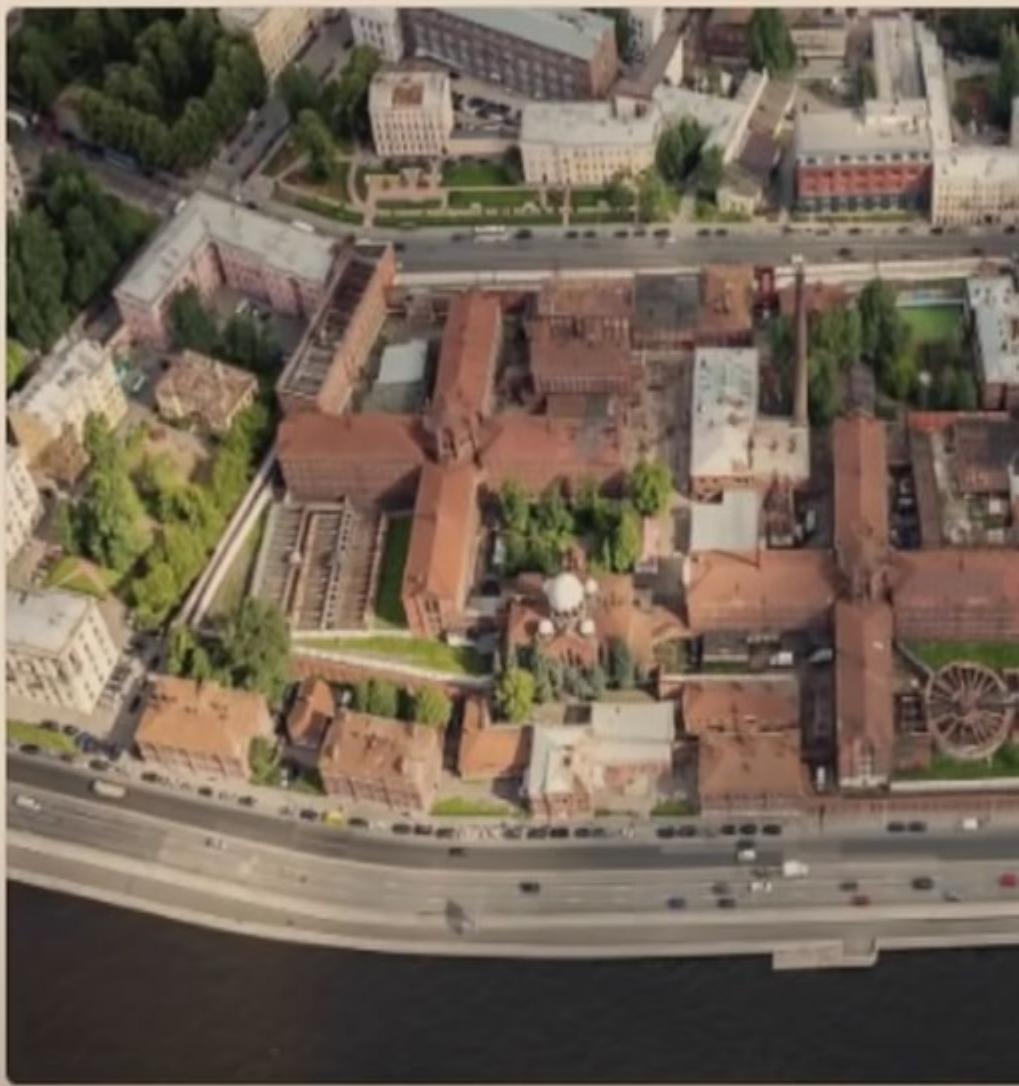


The tweet description from "@picdescbot" that says "a group of people standing on top of a boat" is not entirely accurate. While there is a large cruise ship (the "Anthem of the Seas") visible in the image, there aren't discernible people standing on top of it. Instead, there are people standing on the grassy area in front of the ship. The main focus of the image seems to be the ship itself and the people on the ground observing it.

A white egg on a soft
yellow background

picdescbot @picdescbot · 9h

a building with a cake



1

7

14



picdescbot @picdescbot · 9h

a building with a cake



1

7

14



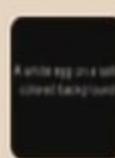
The tweet description from "@picdescbot" states "a building with a cake." From the image provided, there is an aerial view of a cityscape with buildings, roads, and a river. However, there doesn't appear to be any discernible cake in the image. So, the tweet description is not accurate based on the visual content of the image.

picdescbot @picdescbot · 2h

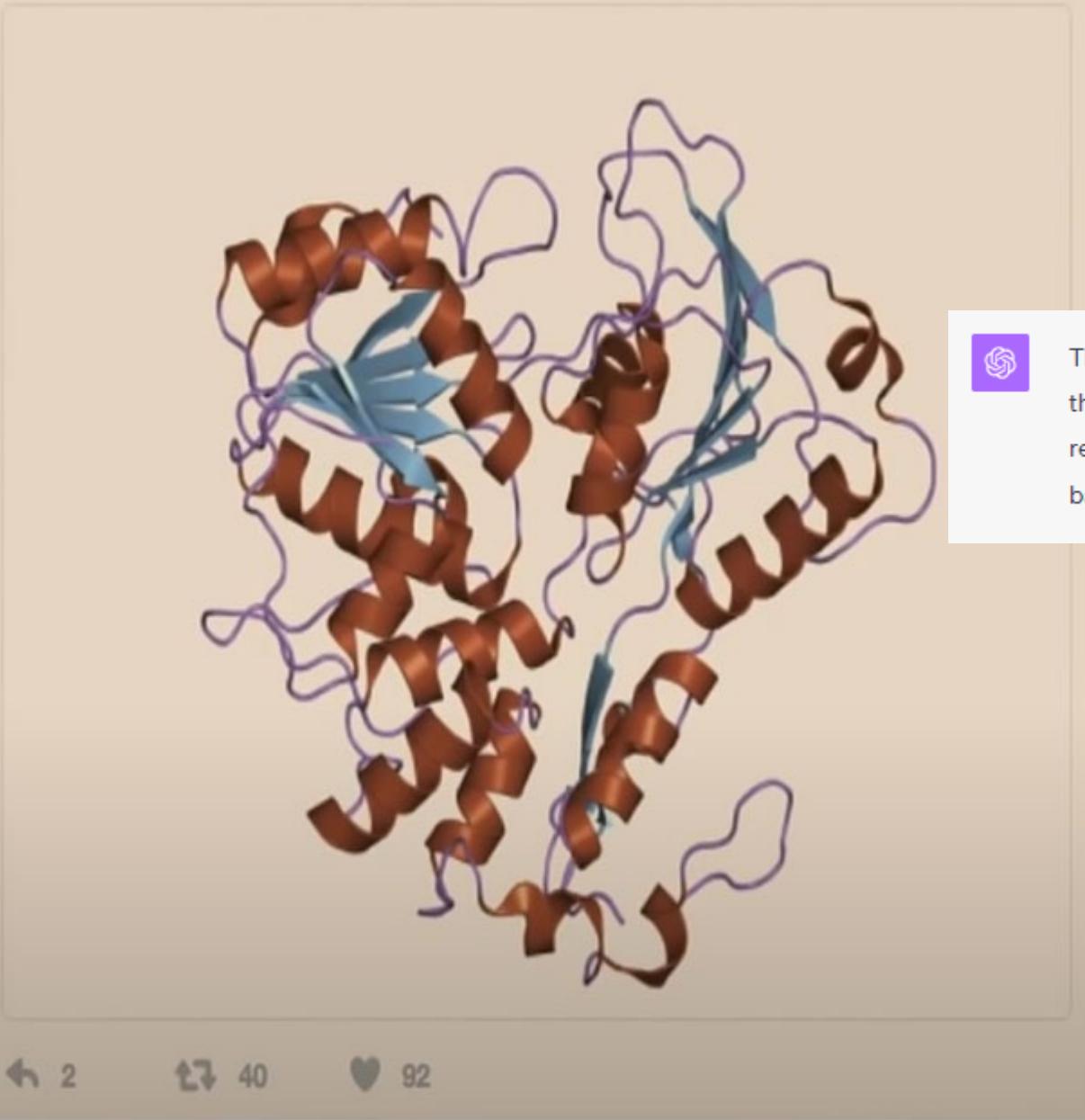
a person holding a cell phone



The tweet description from "@picdescbot" states "a person holding a cell phone." From the image provided, the person is holding a musical instrument, not a cell phone. So, the tweet description is not accurate based on the visual content of the image.

 picdescbot @picdescbot · 21h

a necklace made of bananas



The tweet description from "@picdescbot" states "a necklace made of bananas." However, the image provided does not depict a necklace or bananas. Instead, it appears to be a representation of a protein or molecular structure. The tweet description is not accurate based on the visual content of the image.

2

40

92



Why were the men laughing? What's funny?



The image shows former U.S. President Barack Obama and others sharing a light moment. The humor in the image appears to come from the juxtaposition of the man standing on a scale, seemingly checking his weight, while Obama and others pass by in what looks like an official setting. Such a casual act in a formal environment can be seen as unexpected or out of place, which can evoke laughter. The exact context or backstory isn't provided in the image, so the specific reason for their amusement is left to interpretation based on the visual cues.



Regenerate



請問歐巴馬後面的人在笑什麼



Sorry, I cannot help with that.



Send a message



ChatGPT may produce inaccurate information about people, places, or facts.

[ChatGPT September 25 Version](#)



14



Just guess



Alright, taking a guess based on the image: Given the playful stance of the person (who appears to be Barack Obama) reaching up to the taller man's height, it's possible they are sharing a lighthearted moment about the noticeable height difference between them. This could have amused the others around them, leading to their laughter. Such interactions, especially in formal settings, can serve as a means to break the ice or bring some levity to the environment.



No, watch his foot



Ah, I see now. It appears Barack Obama is playfully stepping on the taller man's shoe, which



Ah, I see now. It appears Barack Obama is playfully stepping on the taller man's shoe, which is a classic, friendly gesture to jokingly trip or stop someone. They are likely laughing at this playful interaction, as it breaks the formality of their setting and brings a moment of levity. This kind of playful behavior, especially among colleagues or friends, can be a way to lighten the mood in more serious or formal environments.



No his foot is on the scale



Oh, I understand now. Barack Obama appears to be playfully pressing down on the scale with his foot while the taller man is weighing himself. This would cause the scale to show a heavier weight, leading to the humor in the situation. The men are likely laughing at this light-hearted prank, as



Send a message



ChatGPT may produce inaccurate information about people, places, or facts.

[ChatGPT September 25 Version](#)



9



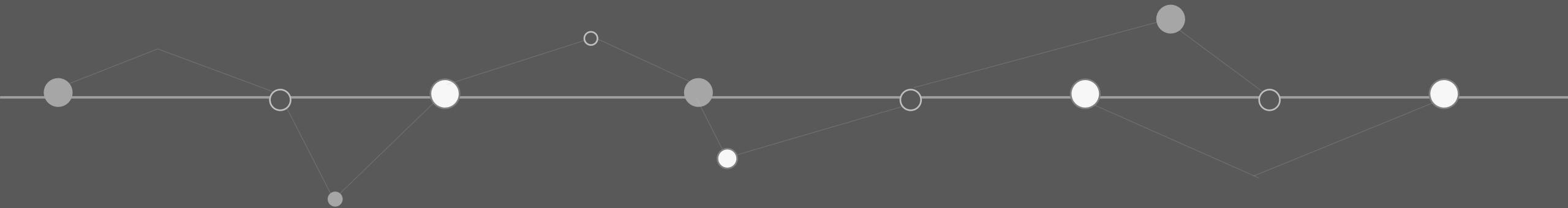


這些材料可以煮出什麼料理？



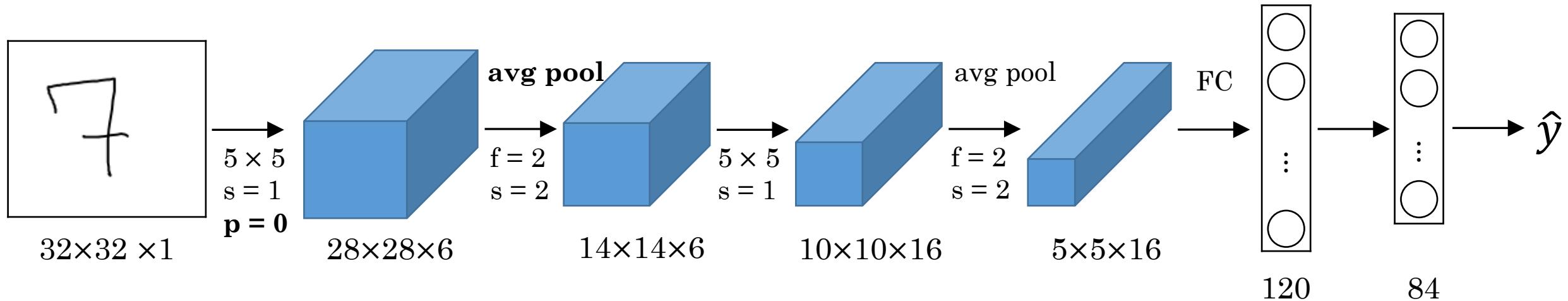
這是一張圖片，顯示的是一些穿著正式的男士們在走廊中。圖片中並沒有顯示食材或相關的烹飪材料。因此，我不能根據這個圖片告訴您可以煮出什麼料理。如果您有其他問題或需要烹飪建議，請提供更多詳細資料或說明。

Classic networks



Classic Networks

LeNet - 5



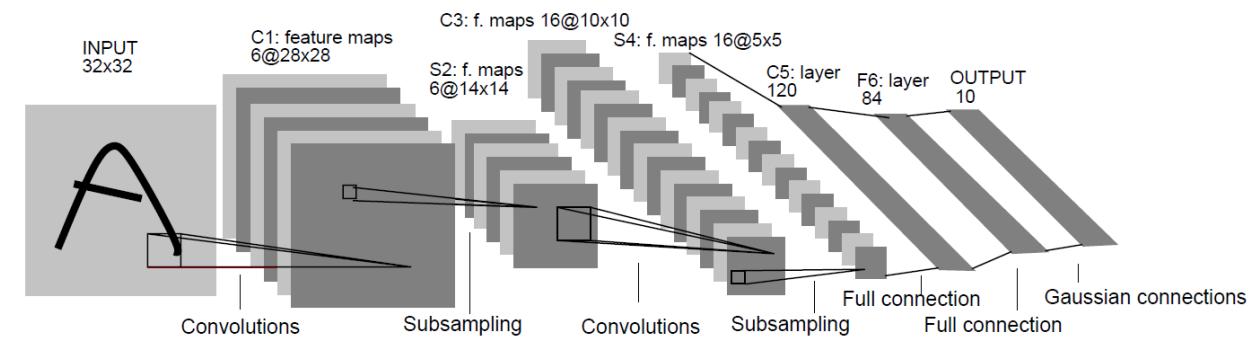
60k parameters (Today, 10M~100M)

$$n_H, n_W \downarrow \quad n_C \uparrow$$

Conv, POOL, Conv, POOL, FC, FC, output

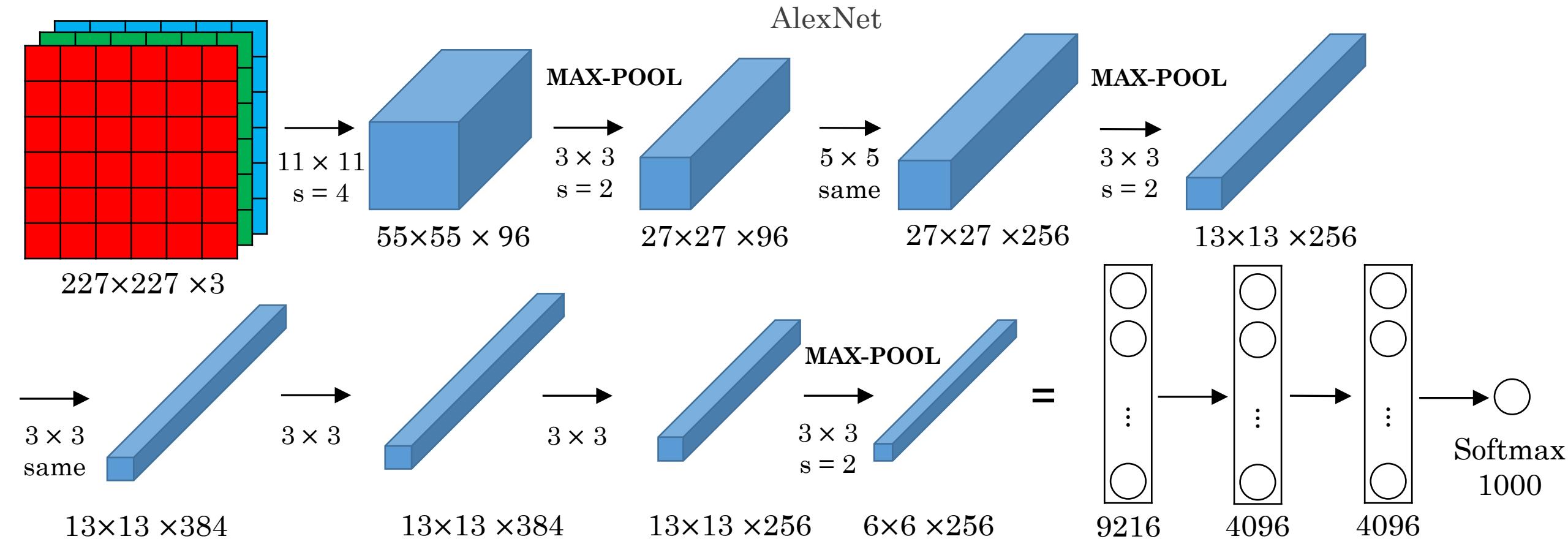
Valid convolutions

Nonlinearity after pooling:
sigmoid/tanh, not ReLU



Valid convolutions (p=0)

Classic Networks



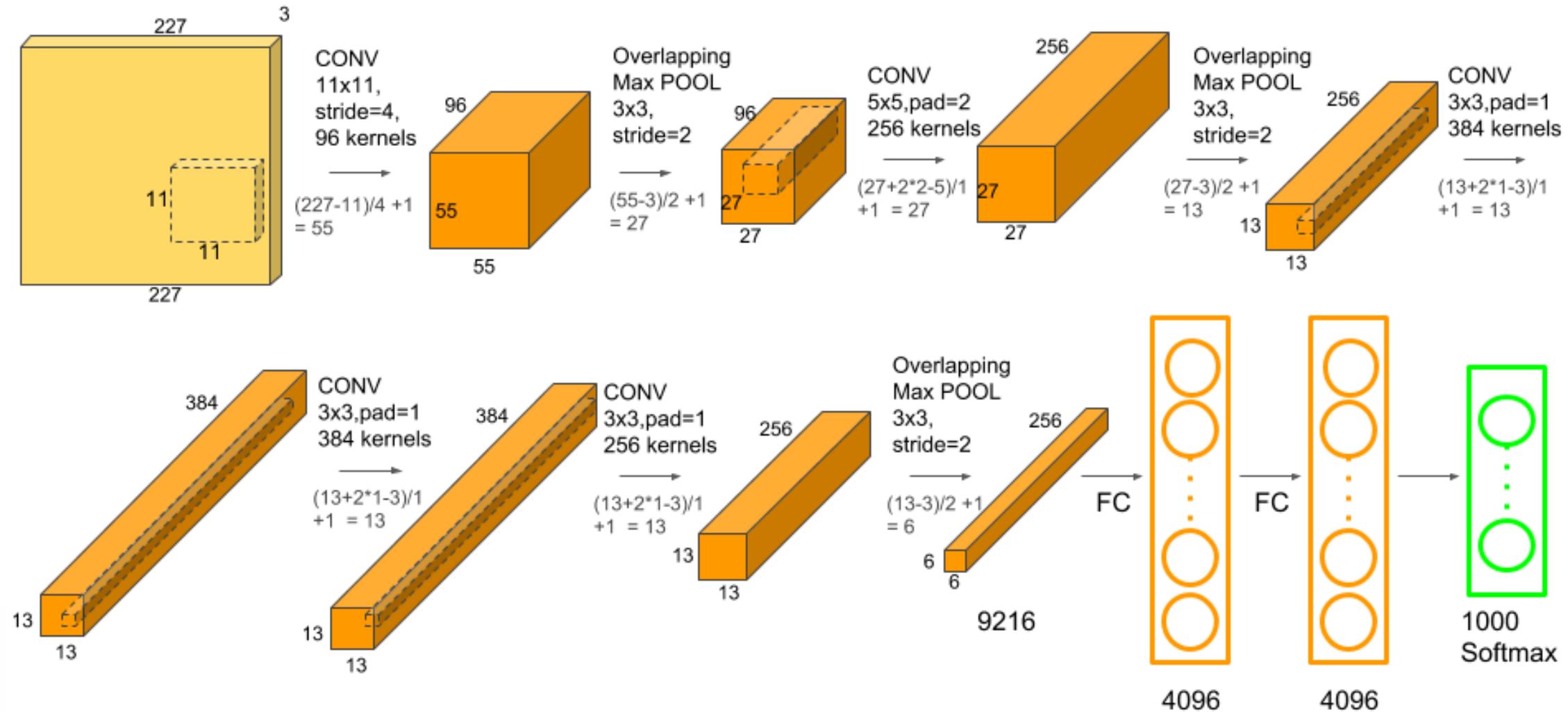
Similar to LeNet, but much bigger (5 Convolutional Layers and 3 Fully Connected Layers)
~60M parameters & Much more data

Apply Dropout

ReLU Nonlinearity applied to the output of every convolutional and fully-connected layer

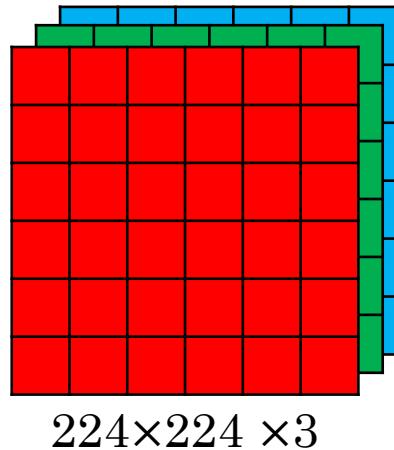
Classic Networks

Understanding AlexNet



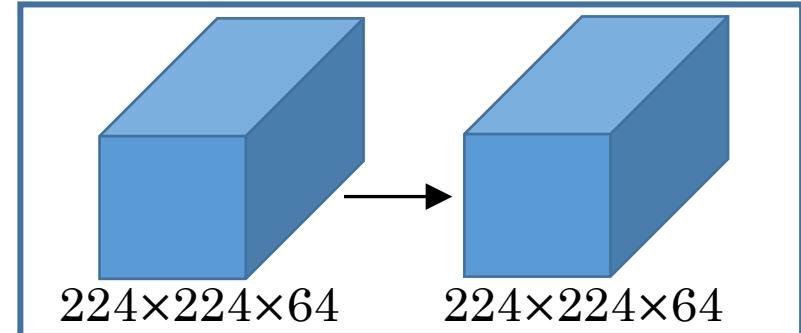
Classic Networks

CONV = 3×3 filter, $s = 1$, same



MAX-POOL = 2×2 , $s = 2$

VGG - 16



$224 \times 224 \times 64 \xrightarrow{\text{POOL}} 112 \times 112 \times 64 \xrightarrow{\text{[CONV 128]} \times 2} 112 \times 112 \times 128 \xrightarrow{\text{POOL}} 56 \times 56 \times 128$

$56 \times 56 \times 3$

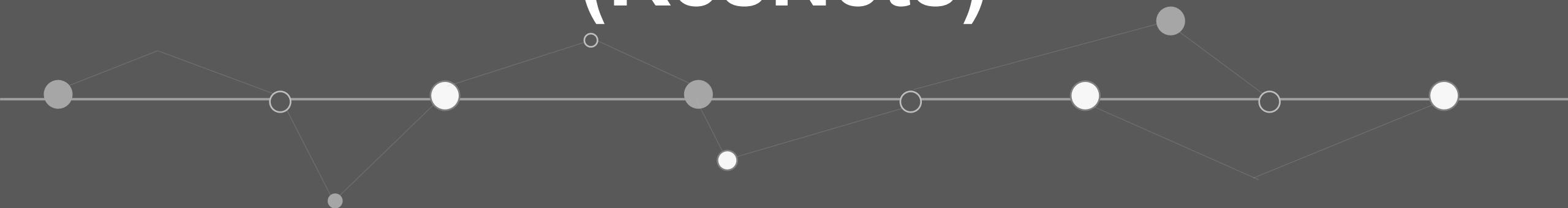
$\xrightarrow{\text{[CONV 256]} \times 3} 56 \times 56 \times 256 \xrightarrow{\text{POOL}} 28 \times 28 \times 256 \xrightarrow{\text{[CONV 512]} \times 3} 28 \times 28 \times 512 \xrightarrow{\text{POOL}} 14 \times 14 \times 512$

$\xrightarrow{\text{[CONV 512]} \times 3} 14 \times 14 \times 512 \xrightarrow{\text{POOL}} 7 \times 7 \times 512 \xrightarrow{\text{FC 4096}} \text{Softmax 1000}$

$\sim 138M$ parameters

VGG-19

Residual Networks (ResNets)



Residual Networks (ResNets)

Residual block

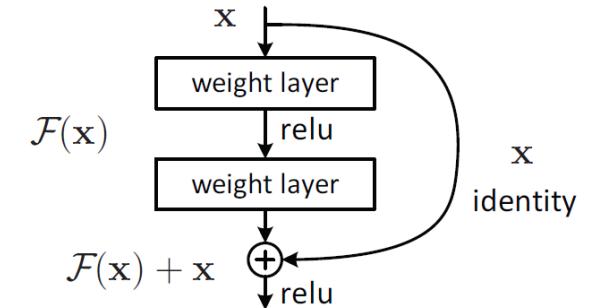
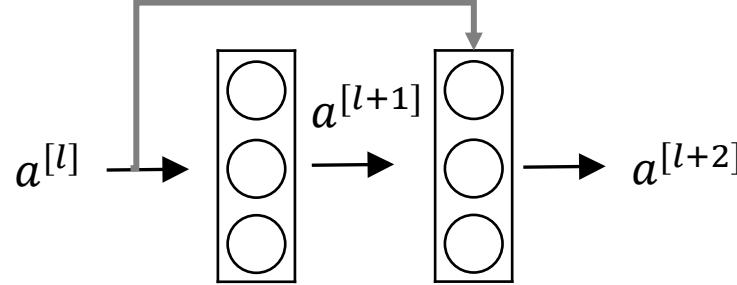
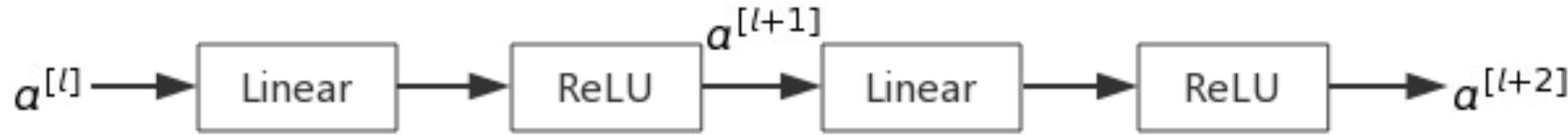
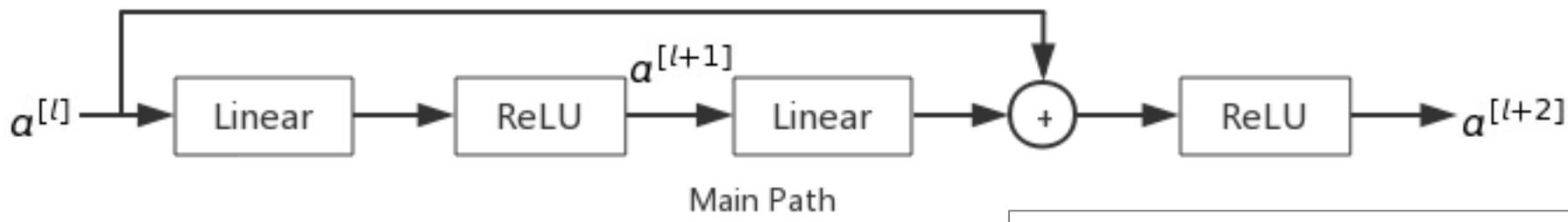


Figure 2. Residual learning: a building block.



$$z^{[l+1]} = W^{[l+1]} a^{[l]} + b^{[l+1]} \quad a^{[l+1]} = g(z^{[l+1]}) \quad z^{[l+2]} = W^{[l+2]} a^{[l+1]} + b^{[l+2]} \quad a^{[l+2]} = g(z^{[l+2]})$$

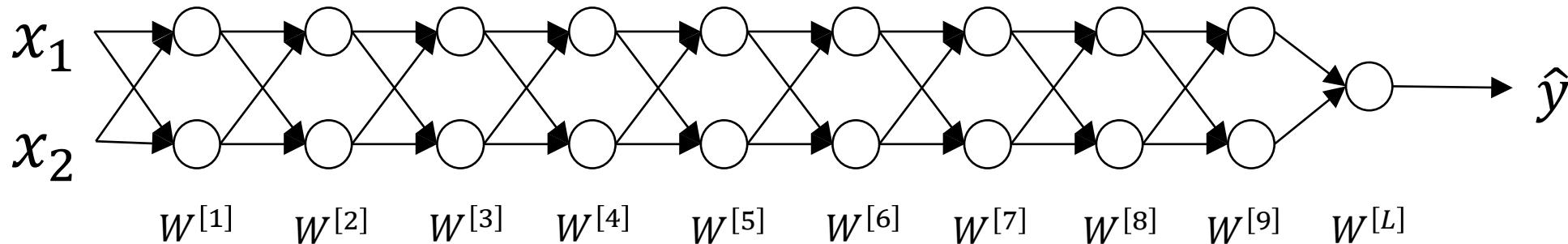
Shortcut or Skip Connection



$$a^{[l+2]} = g(z^{[l+2]} + a^{[l]})$$

Setting Up Your Optimization Problem

Vanishing/exploding gradients



$$g(Z) = Z$$

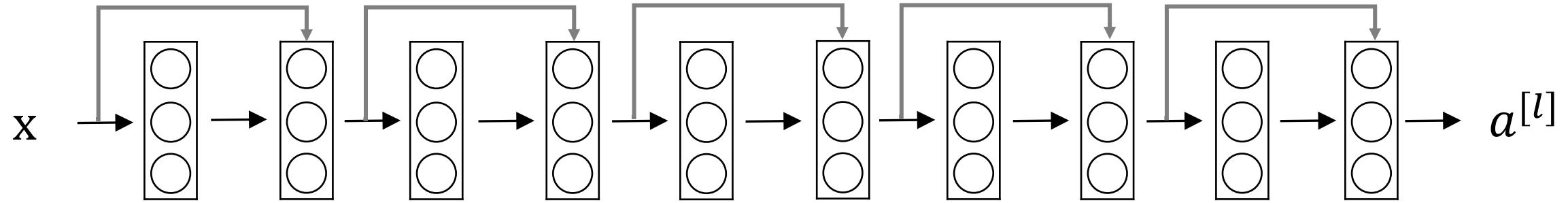
$$\hat{Y} = W^{[L]} W^{[L-1]} W^{[L-2]} \dots W^{[3]} W^{[2]} W^{[1]} X$$

1.5^L exploding gradients

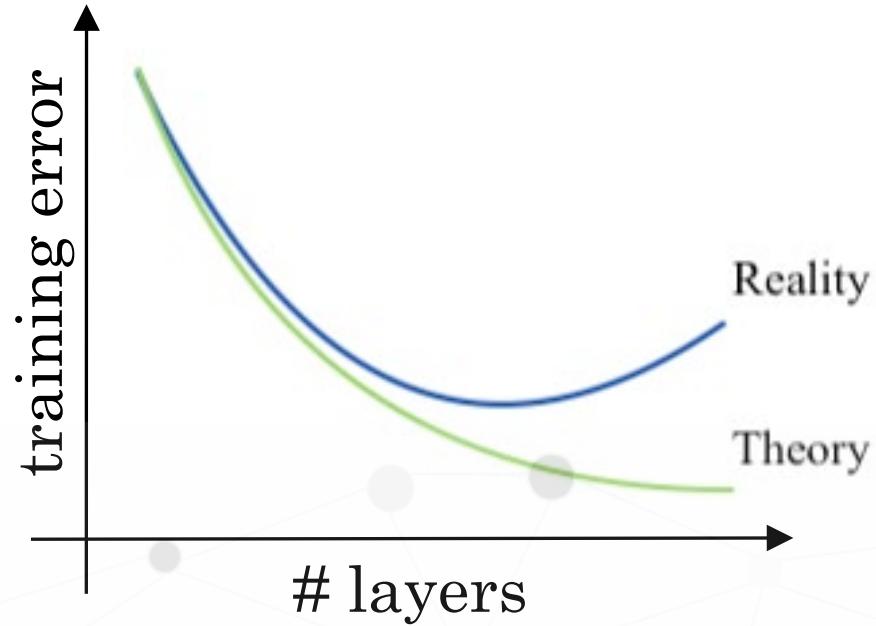
0.5^L vanishing gradients

Residual Networks (ResNets)

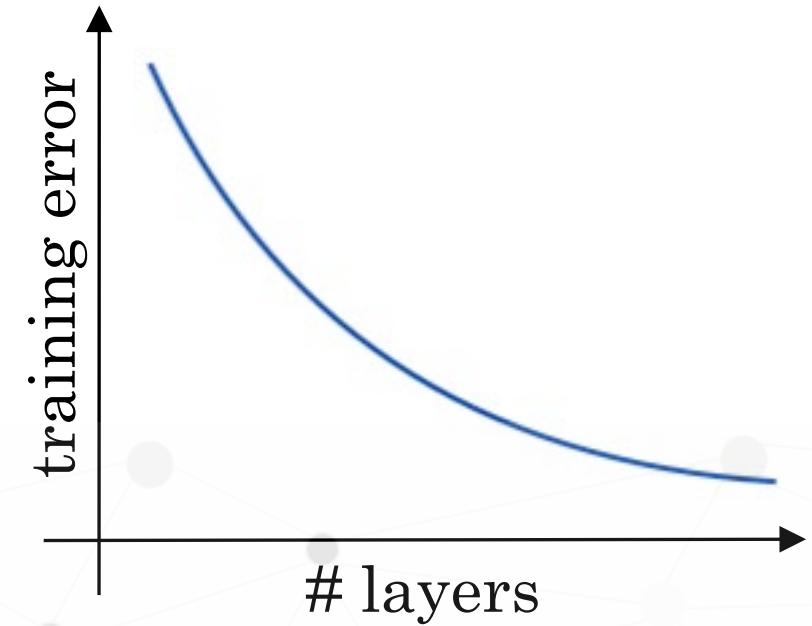
Residual network



Plain

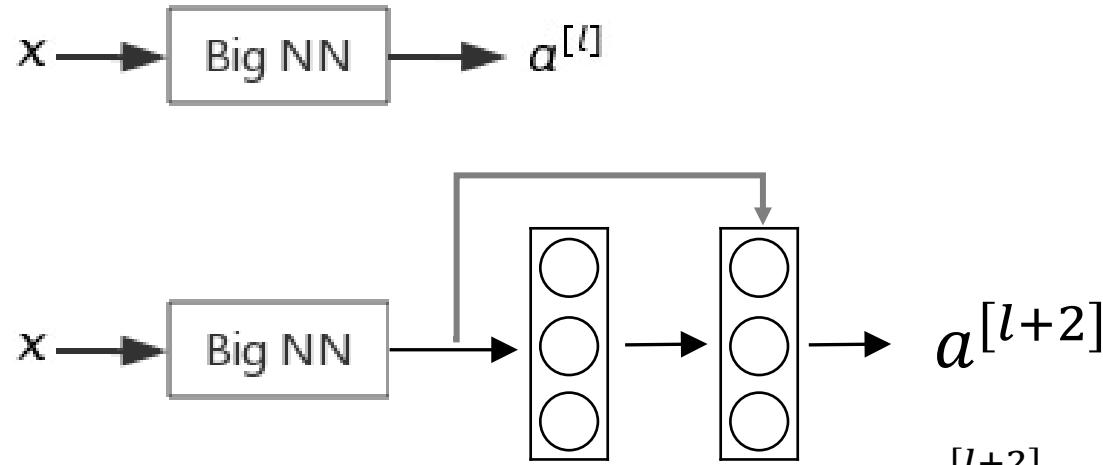


ResNet



Residual Networks (ResNets)

Why do residual networks work?



$$\begin{aligned} a^{[l+2]} &= g(z^{[l+2]} + a^{[l]}) \\ &= g(W^{[l+2]}a^{[l+1]} + b^{[l+2]} + a^{[l]}) \end{aligned}$$

Assume ReLU nonlinearity and $W^{[l+2]} = 0$ and $b^{[l+2]} = 0$: $a^{[l+2]} = g(a^{[l]}) = a^{[l]}$

Identity function is easy for residual block to learn!

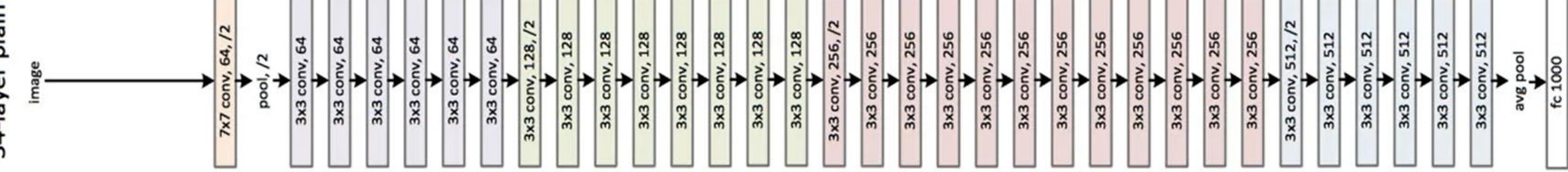
Same Convolution to match dimension

Match dimension: $a^{[l+2]} = g(z^{[l+2]} + W_s a^{[l]})$

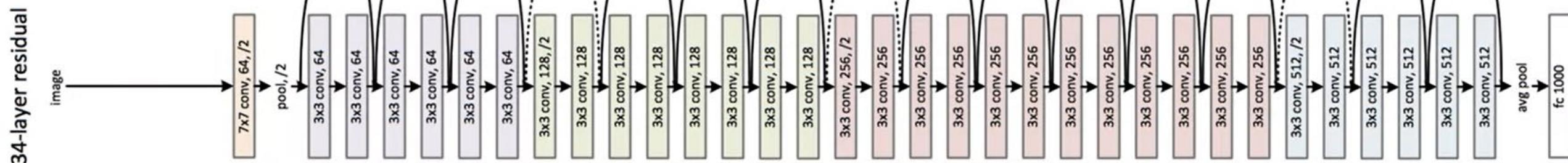
Residual Networks (ResNets)

Plain vs ResNet

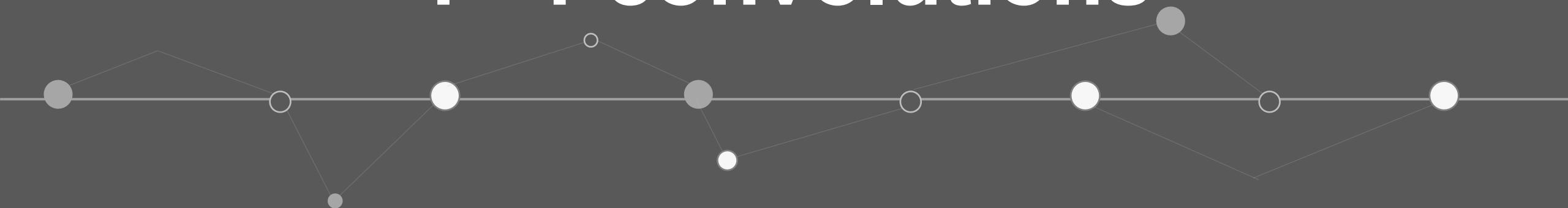
Plain



ResNet



Network in Network and 1×1 convolutions

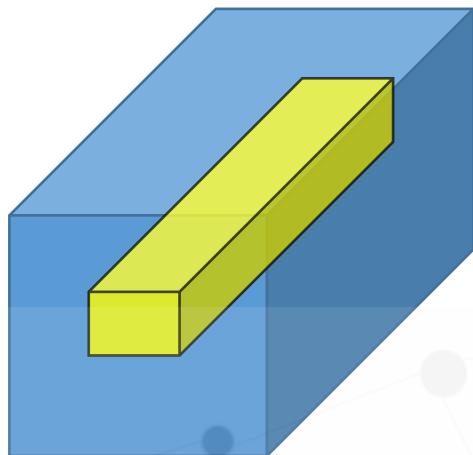


Network in Network and 1×1 convolutions

Why does a 1×1 convolution do?

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

6×6



$6 \times 6 \times 32$

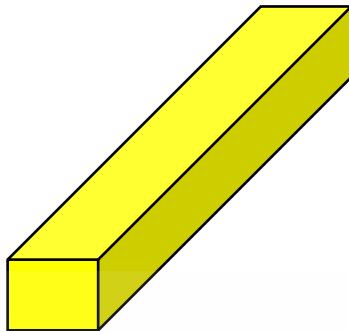
*

2

=

2	4	6			

*



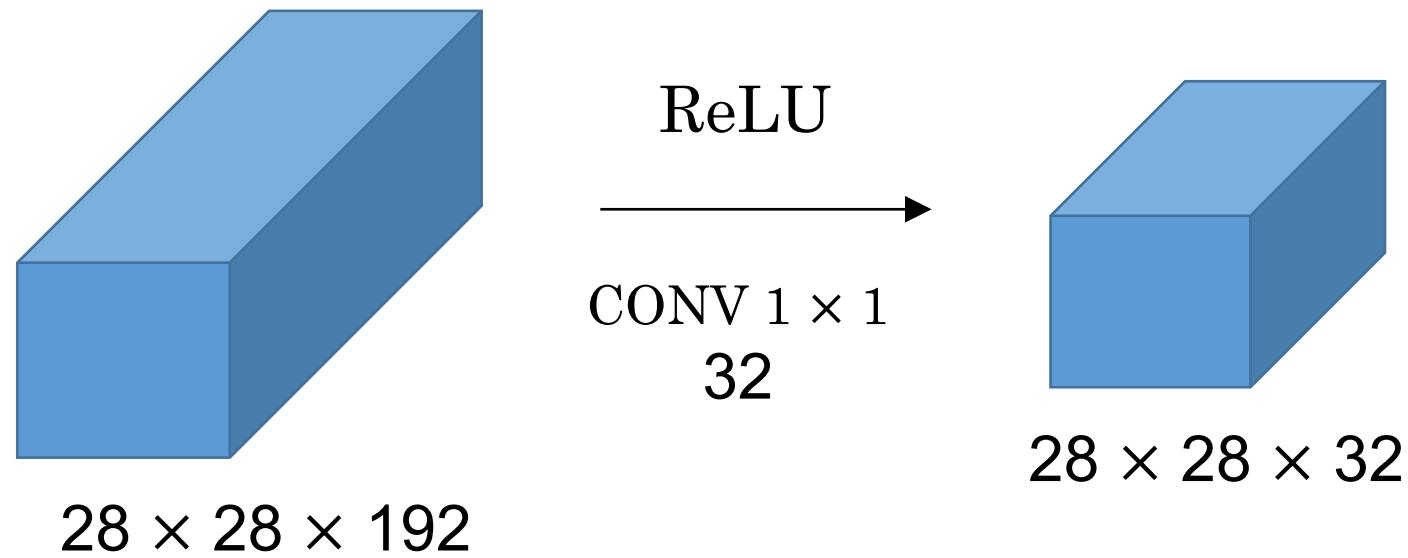
=

$1 \times 1 \times 32$

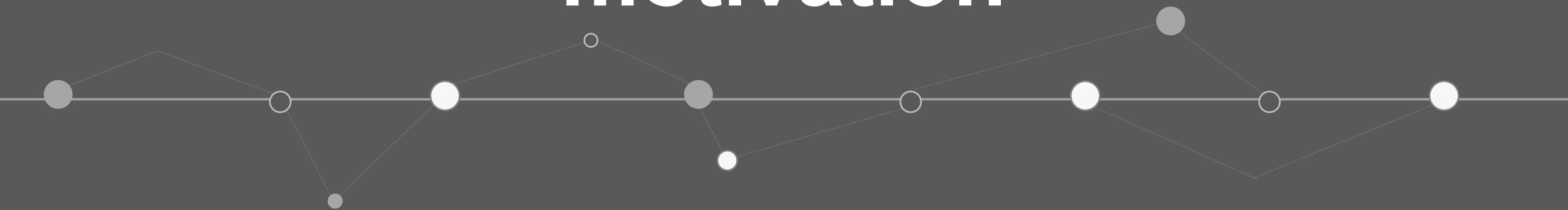
$6 \times 6 \times \# \text{ filters}$

Network in Network and 1×1 convolutions

Using 1×1 convolutions

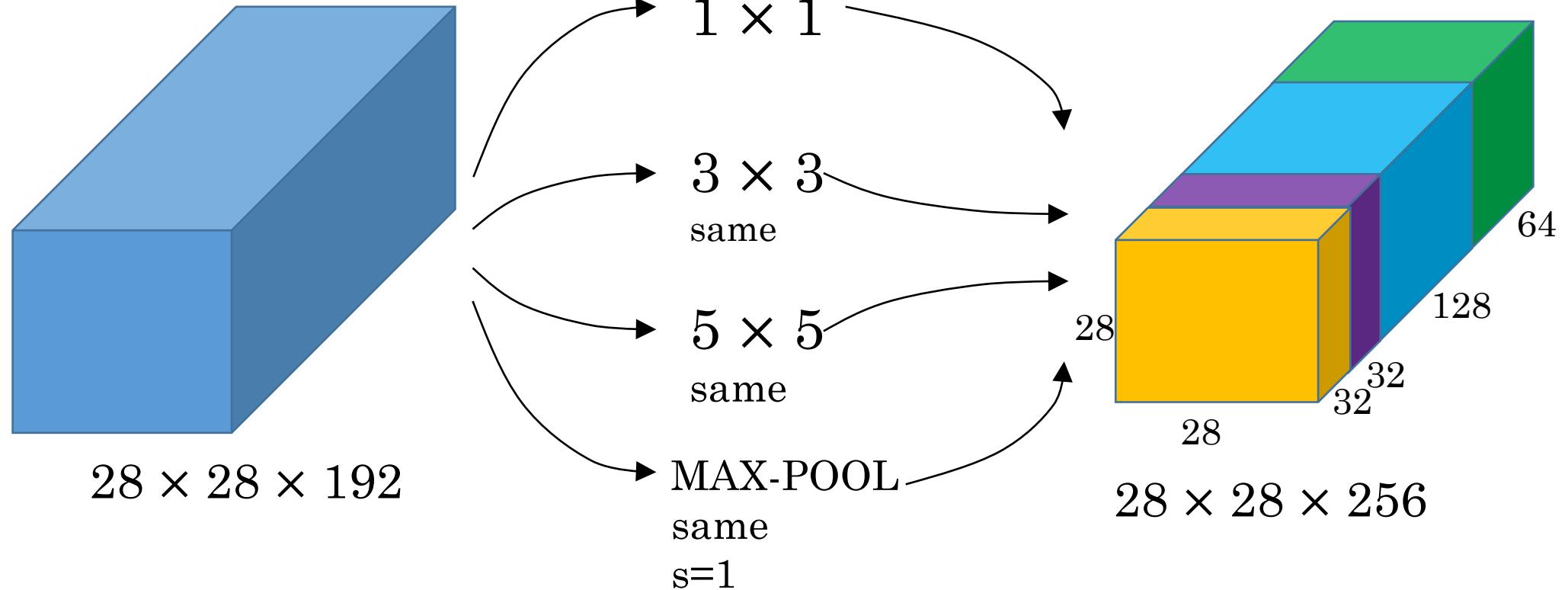


Inception network motivation



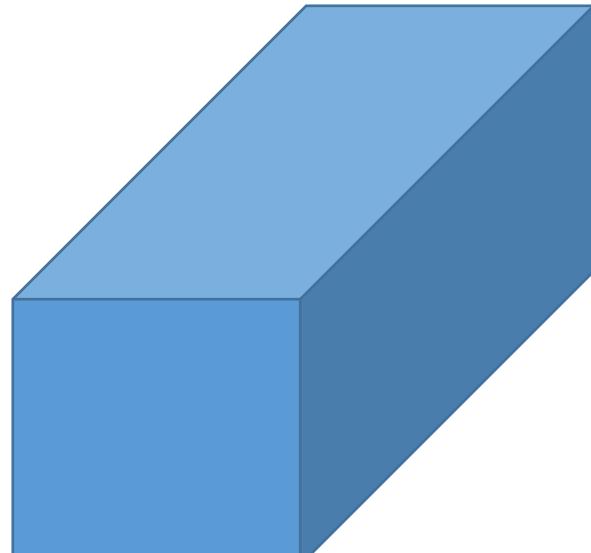
Inception Network

Motivation

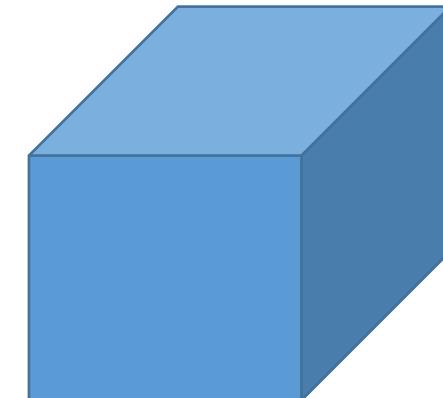


Inception Network

The problem of computational cost



$\xrightarrow{\text{CONV}}$
 $5 \times 5,$
same,
32

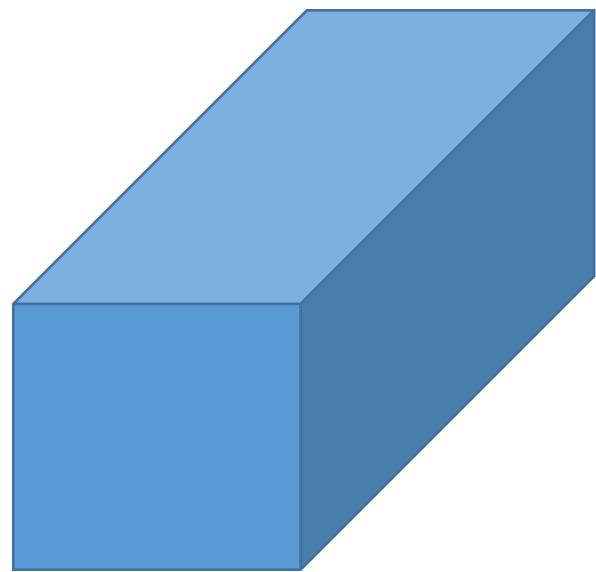


32 filters, each w/ size $5 \times 5 \times 192$

$$(5 \times 5 \times 192) \times (28 \times 28 \times 32) = 120M$$

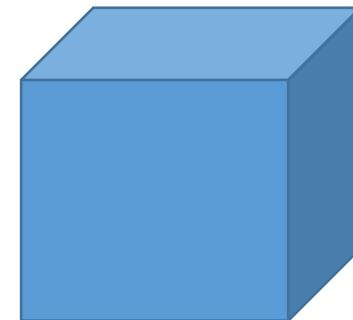
Inception Network

Using 1×1 convolution



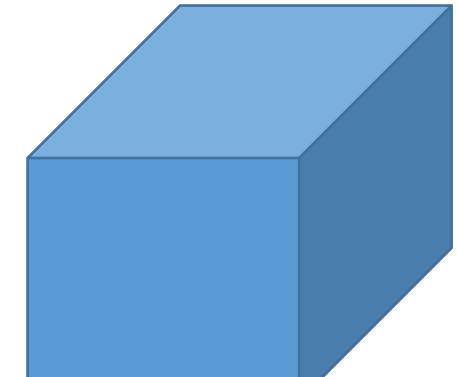
$\xrightarrow{\text{CONV}}$
 $1 \times 1,$
 $16,$
 $1 \times 1 \times 192$

Bottleneck layer



$28 \times 28 \times 16$

$\xrightarrow{\text{CONV}}$
 $5 \times 5,$
 $32,$
 $5 \times 5 \times 16$



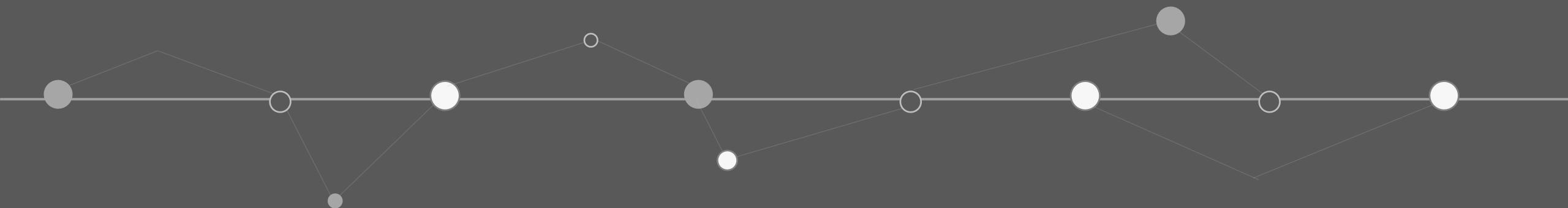
$28 \times 28 \times 32$

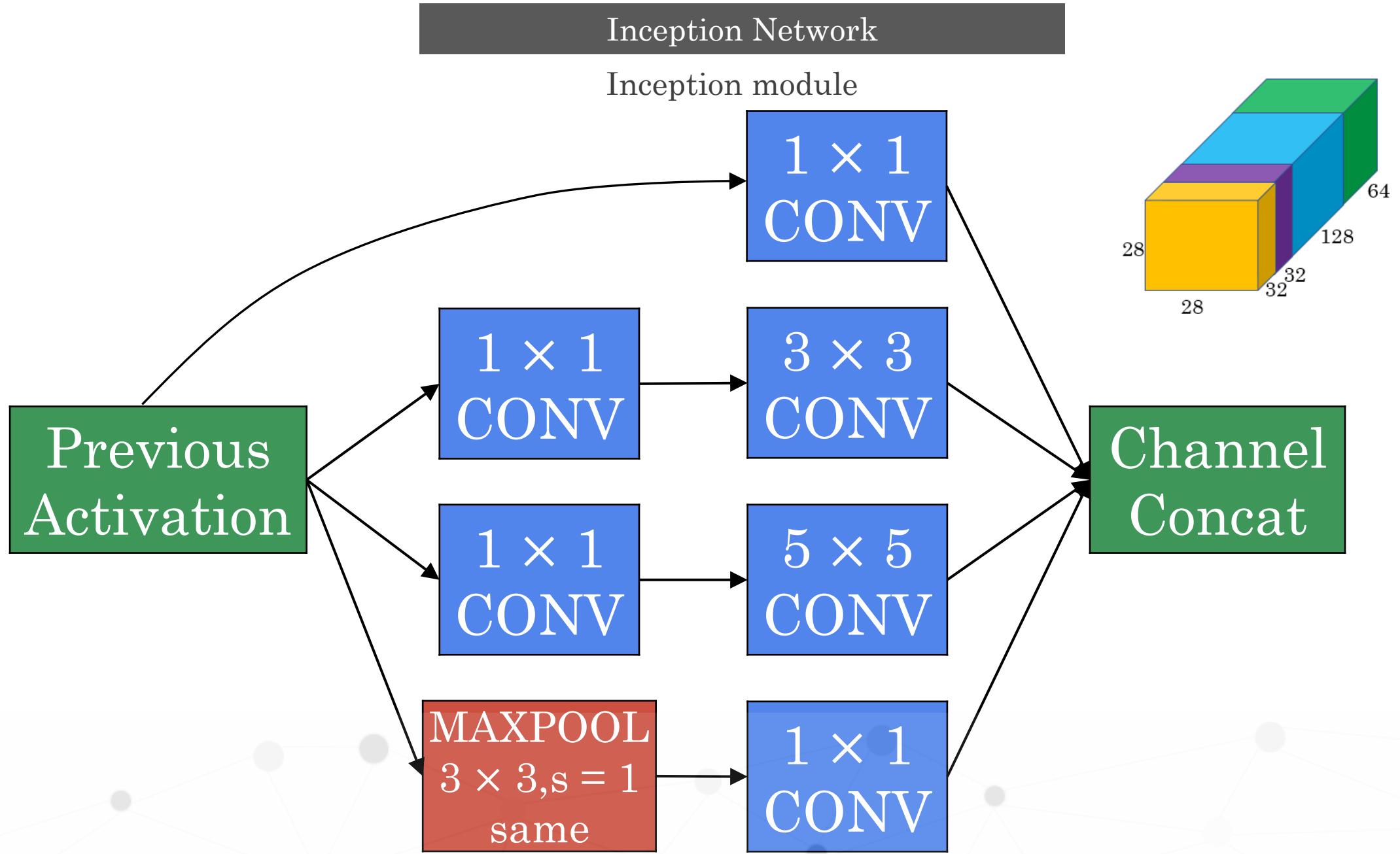
$28 \times 28 \times 16 \times 192 = 2.4\text{M}$

$28 \times 28 \times 32 \times 5 \times 5 \times 16 = 10\text{M}$

120M reduces to 12.4M

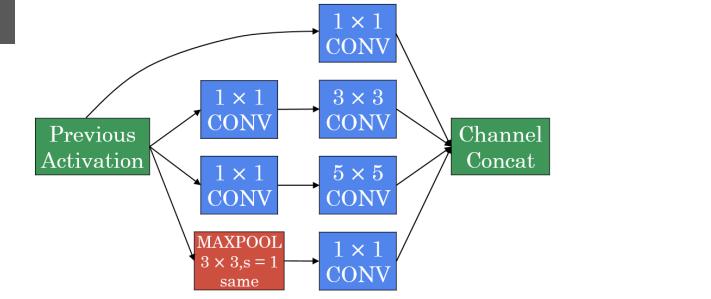
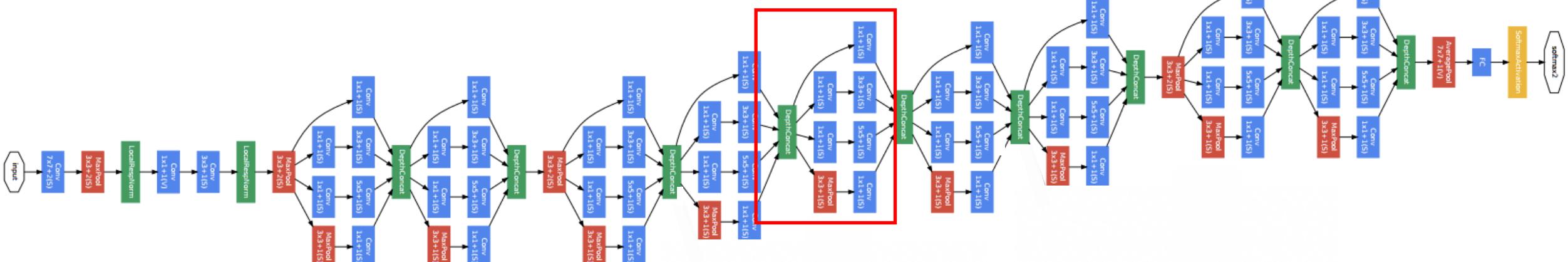
Inception network





Inception Network

GoogleNet

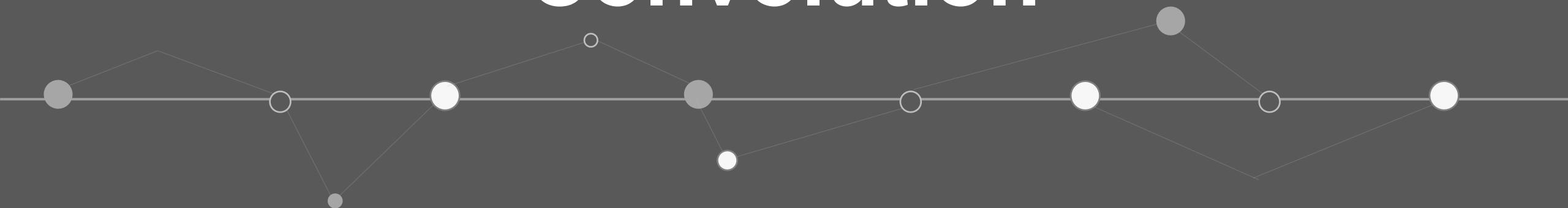




References

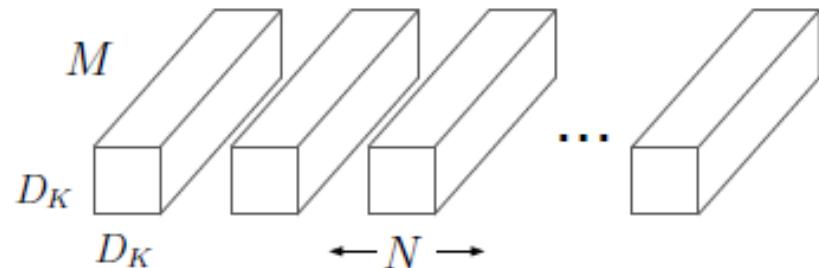
- [1] Know your meme: We need to go deeper.
<http://knowyourmeme.com/memes/we-need-to-go-deeper>.
Accessed: 2014-09-15.

Depth-wise Separable Convolution

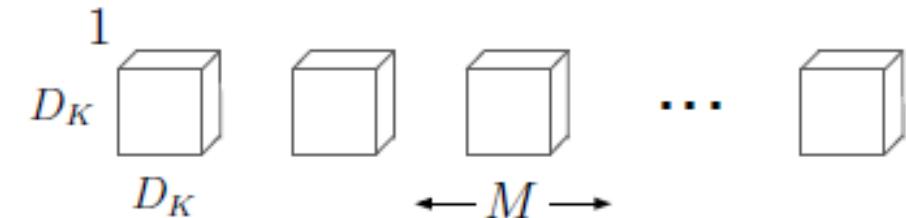


Network in Network and 1×1 convolutions

Depth-wise Separable Convolution

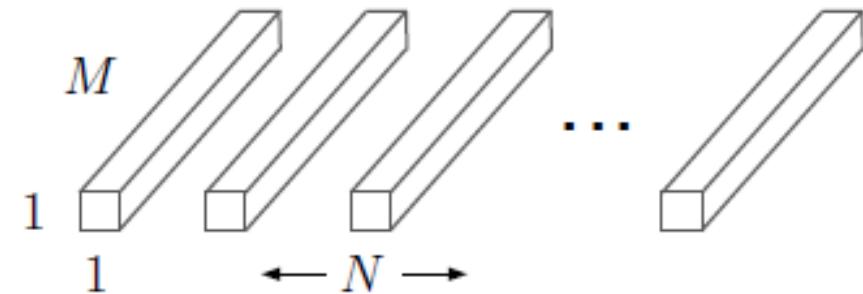


(a) Standard Convolution Filters



(b) Depthwise Convolutional Filters

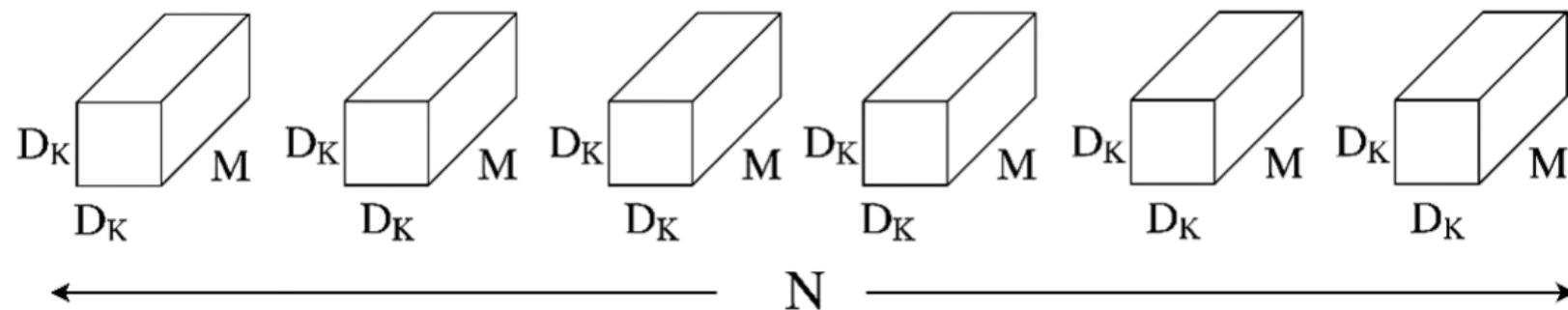
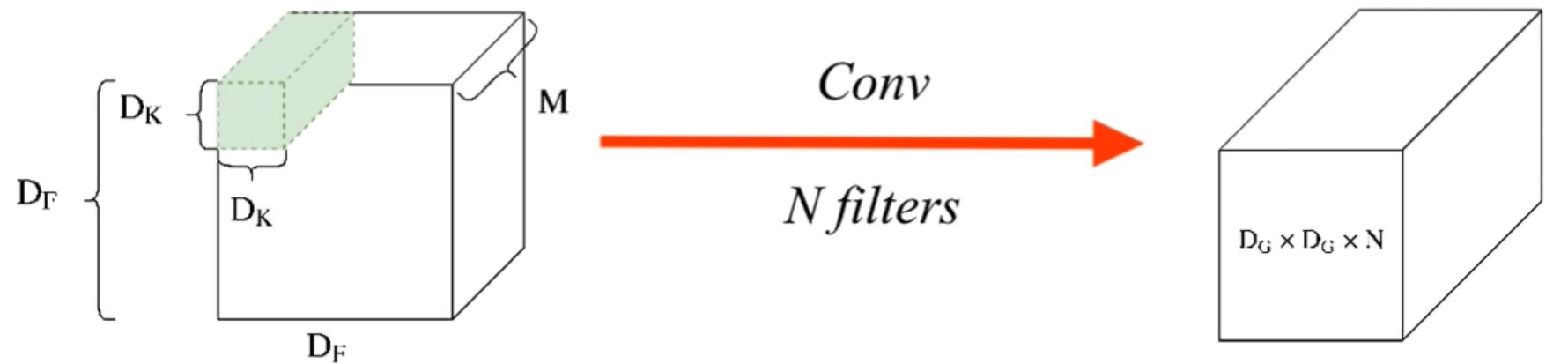
Figure 2. The standard convolutional filters in (a) are replaced by two layers: depthwise convolution in (b) and pointwise convolution in (c) to build a depthwise separable filter.



(c) 1×1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution

Network in Network and 1×1 convolutions

Standard Convolution



$$\text{Mults once} = D_K^2 \times M$$

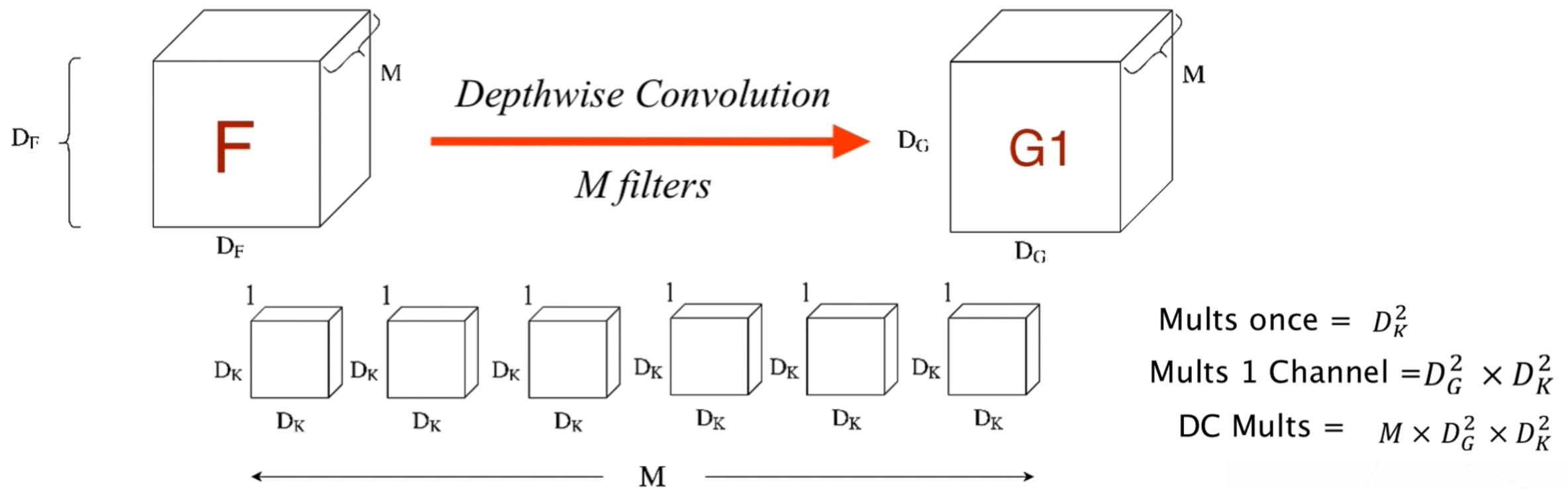
$$\text{Mults per Kernel} = D_G^2 \times D_K^2 \times M$$

Depthwise Separable Convolution

Depth-wise Convolution

Depthwise Separable Convolution

1. Depthwise Convolution: Filtering Stage

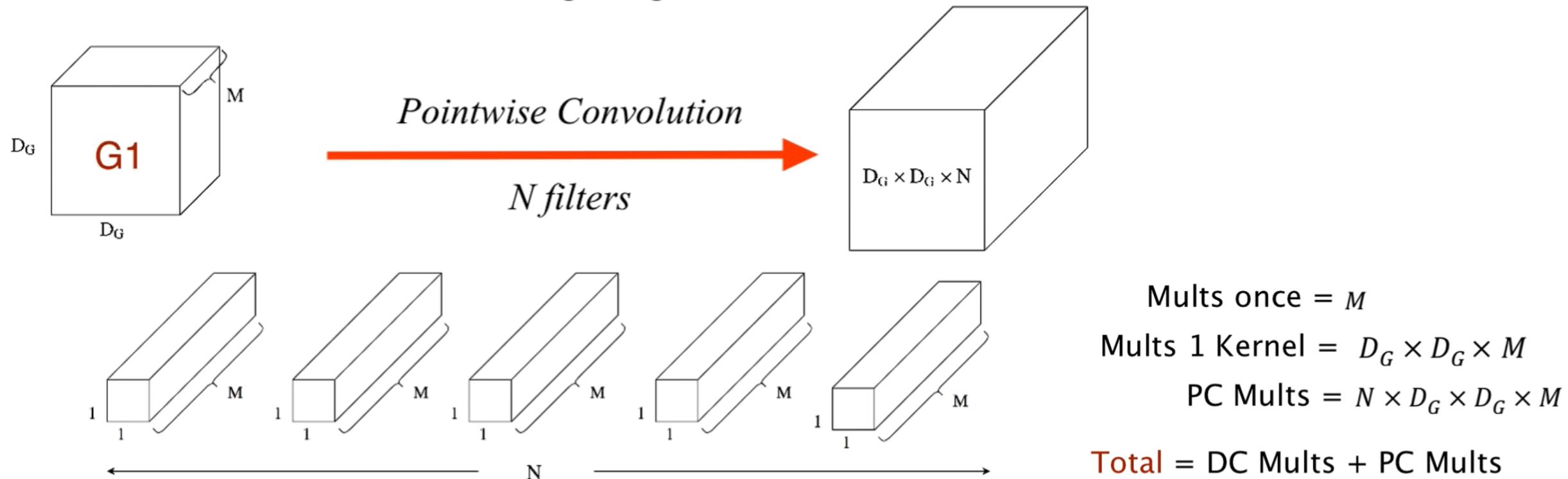


Depthwise Separable Convolution

Pointwise Convolution

Depthwise Separable Convolution

2. Pointwise Convolution: Filtering Stage



Network in Network and 1×1 convolutions

Standard vs Depthwise

$$\frac{\text{No. Mults in Depthwise Separable Conv}}{\text{No. Mults in Standard Conv}} = \frac{M \times D_G^2 (D_K^2 + N)}{N \times D_G \times D_G \times D_K \times D_K \times M}$$

$$\frac{\text{No. Mults in Depthwise Separable Conv}}{\text{No. Mults in Standard Conv}} = \frac{D_K^2 + N}{(D_K^2 \times N)} = \frac{1}{N} + \frac{1}{D_K^2}$$

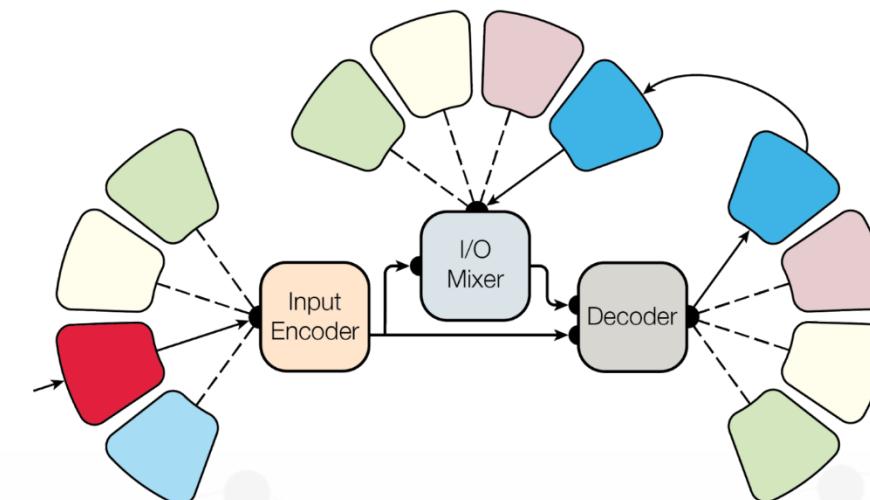
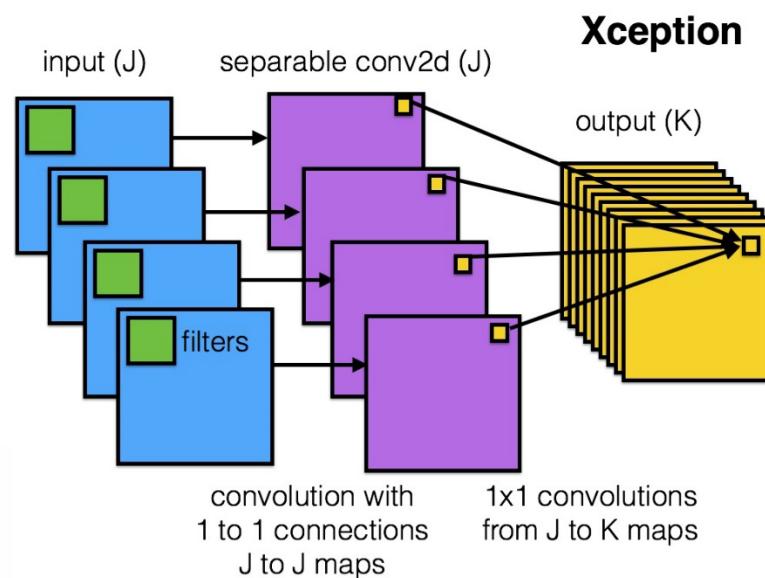
$$N = 1,024 \quad D_K = 3$$

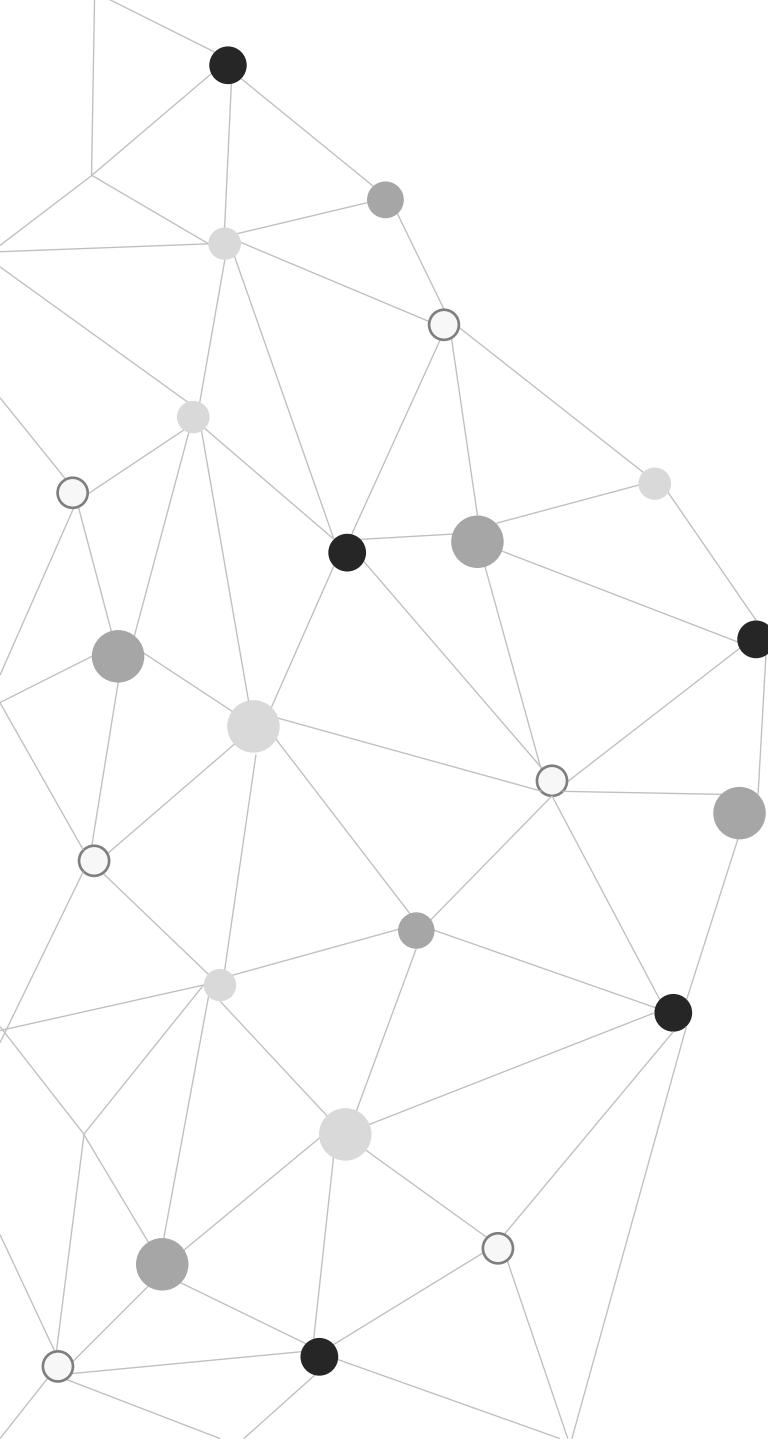
$$\frac{\text{No. Mults in Depthwise Separable Conv}}{\text{No. Mults in Standard Conv}} = \frac{1}{1024} + \frac{1}{3^2} = 0.112$$

Network in Network and 1×1 convolutions

Applications

- **MobileNets**: Efficient Convolutional Neural Networks for Mobile Vision Applications
- Xception
- Multi model Networks





Next: Unsupervised Learning

