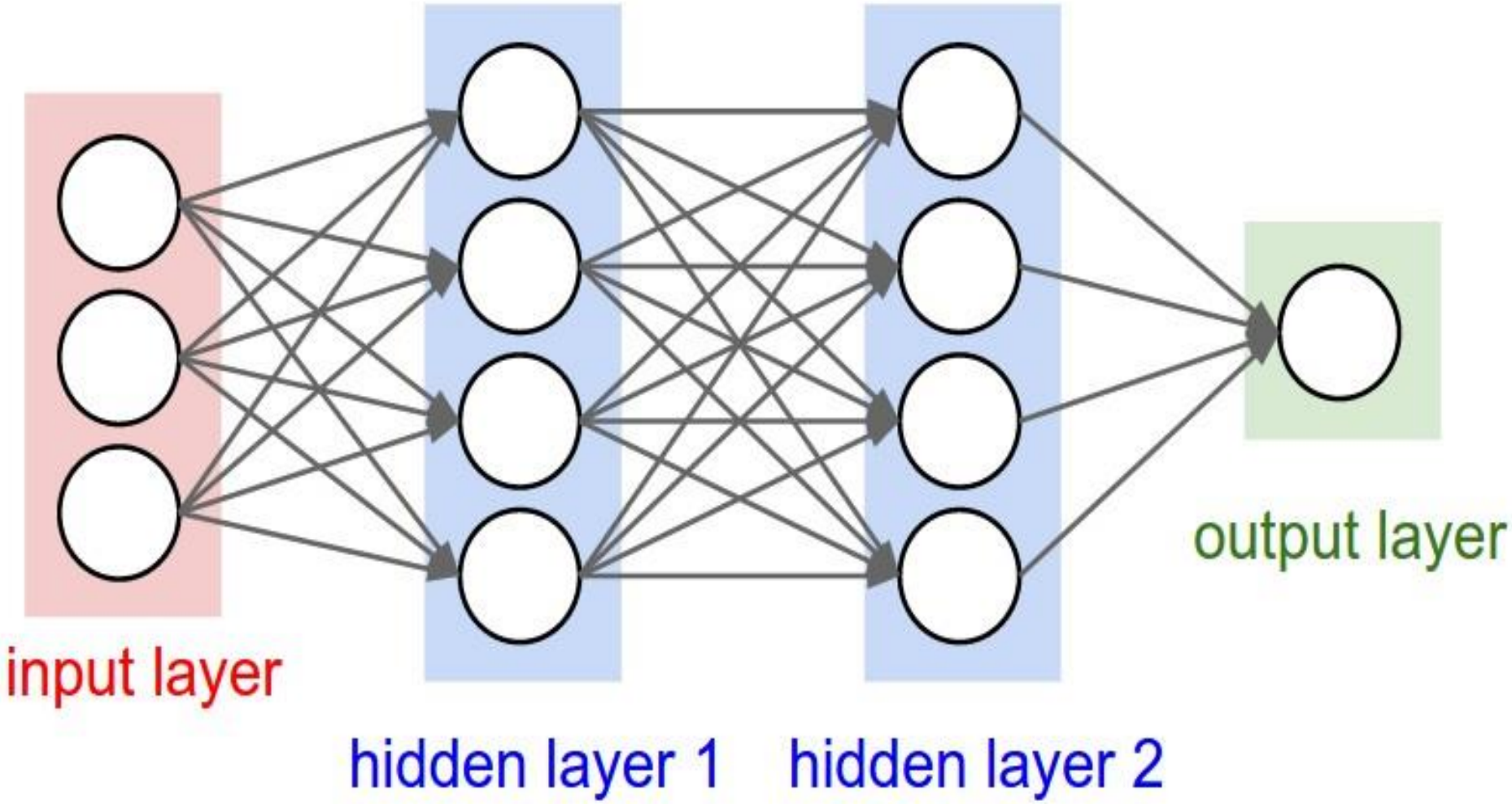
Deep Learning

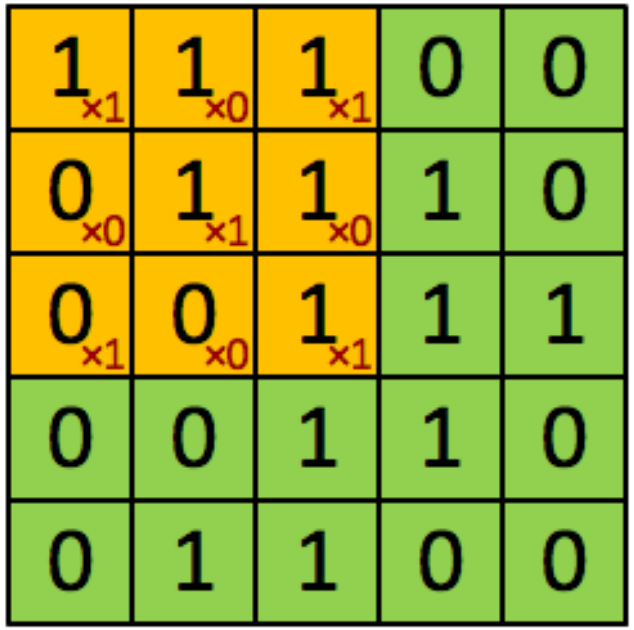
Multi-layer perceptron



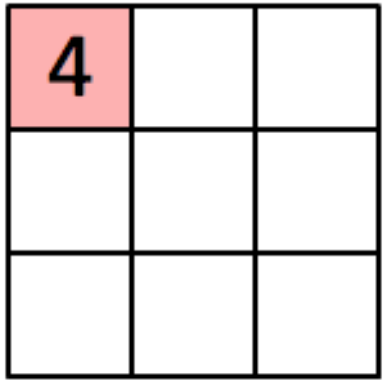
David Rumelhart, Geoffrey Hinton and Ronald Williams, 1986

Convolution

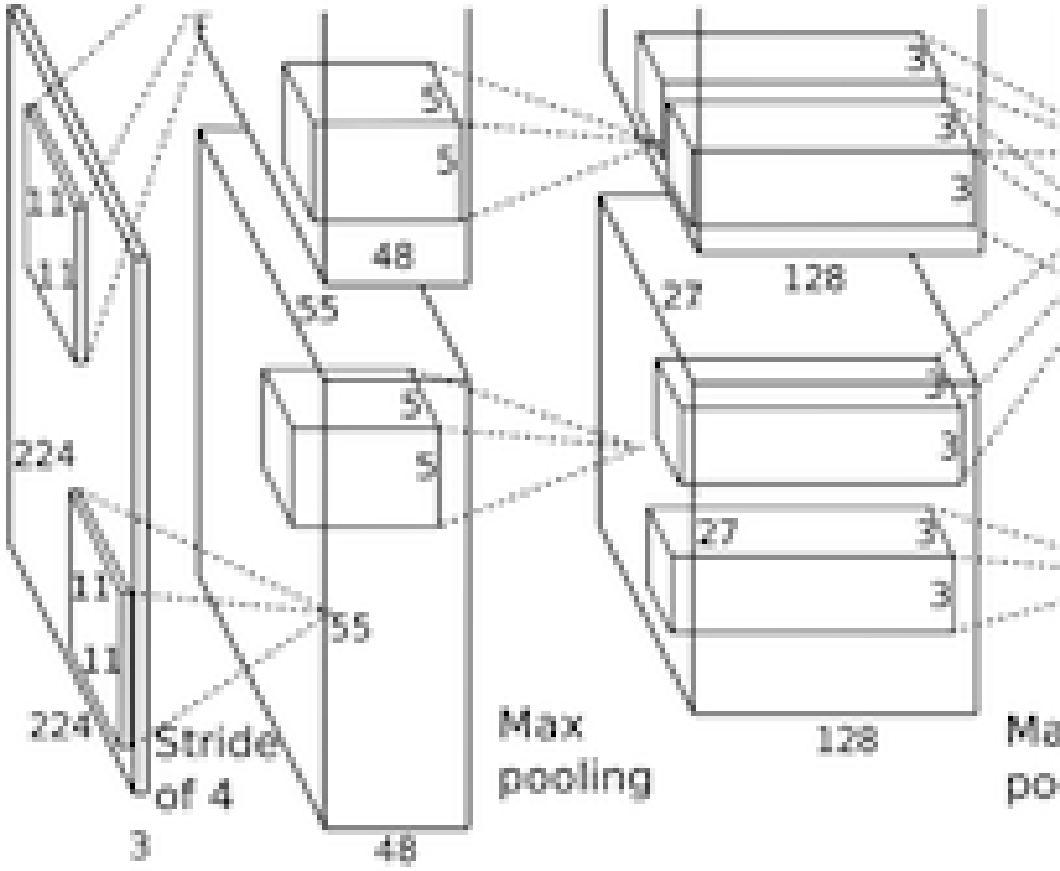
Convolution



Image



Convolved Feature



Convolution

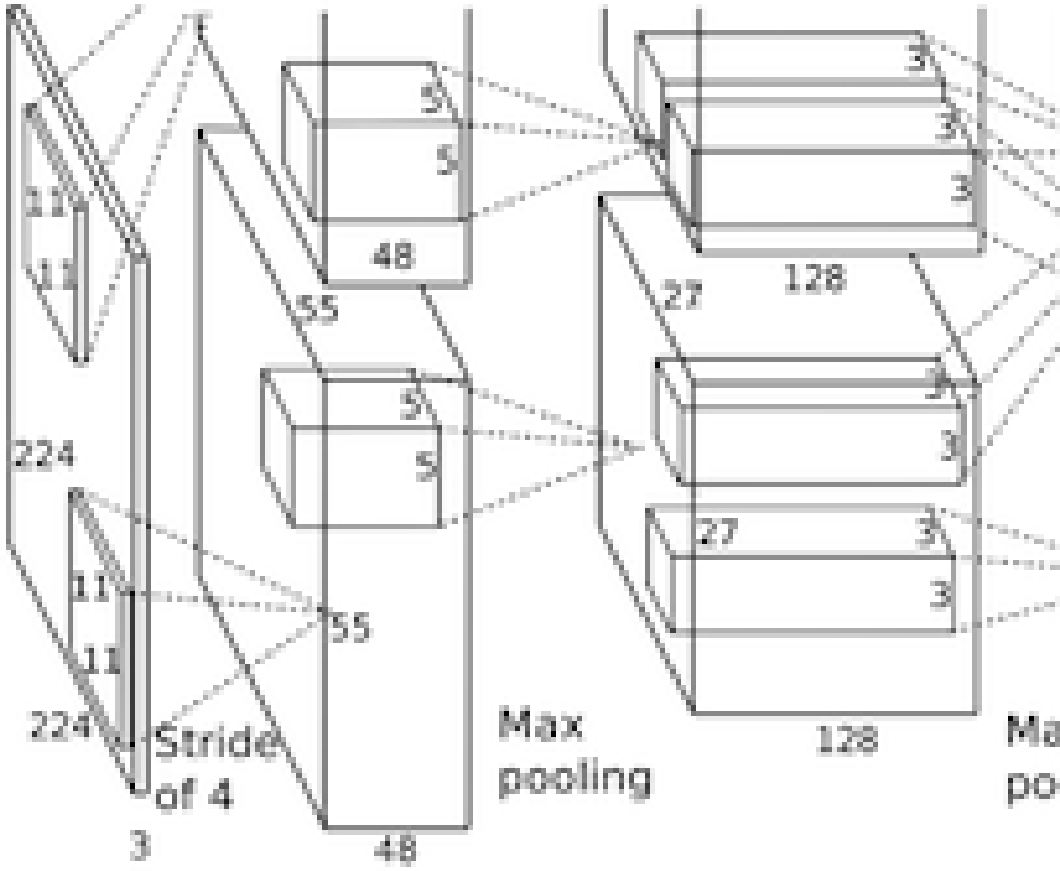
$1_{\times 1}$	$1_{\times 0}$	$1_{\times 1}$	0	0
$0_{\times 0}$	$1_{\times 1}$	$1_{\times 0}$	1	0
$0_{\times 1}$	$0_{\times 0}$	$1_{\times 1}$	1	1
0	0	1	1	0
0	1	1	0	0

Image

4		

Convolved Feature

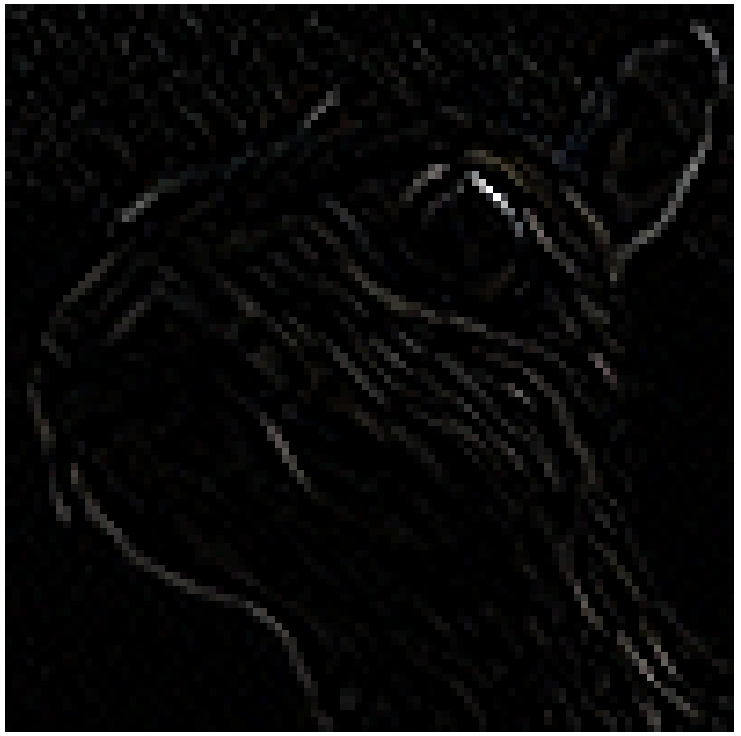
Kernel 3x3, Stride 1



Convolution examples

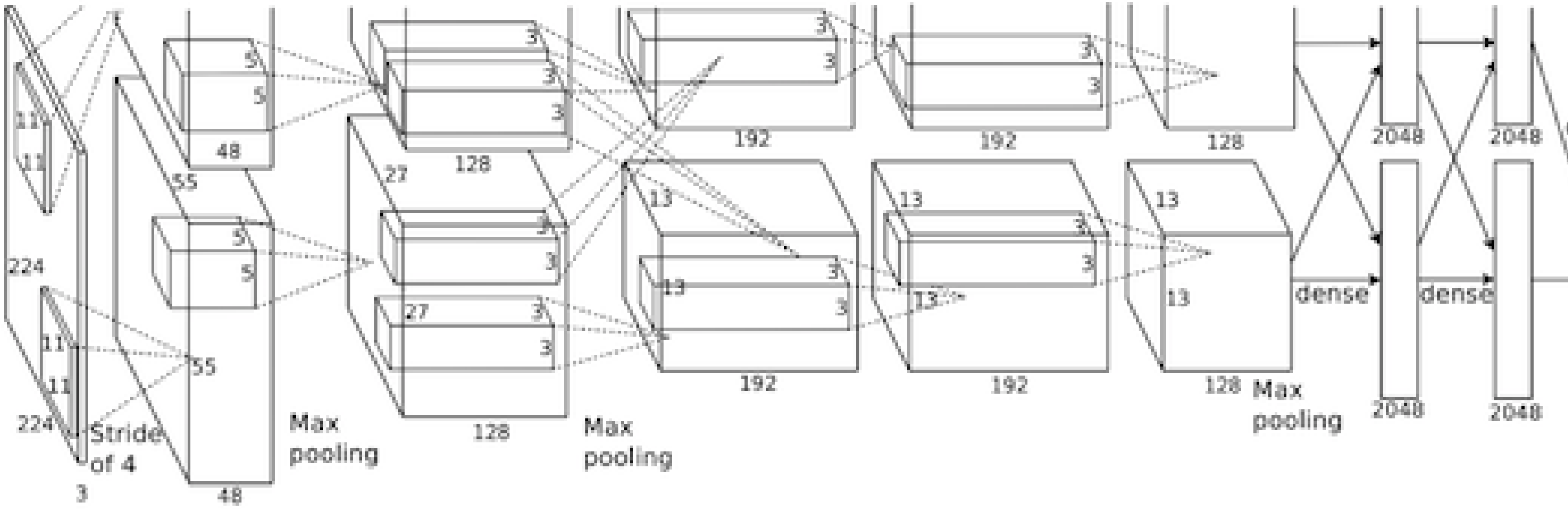
$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

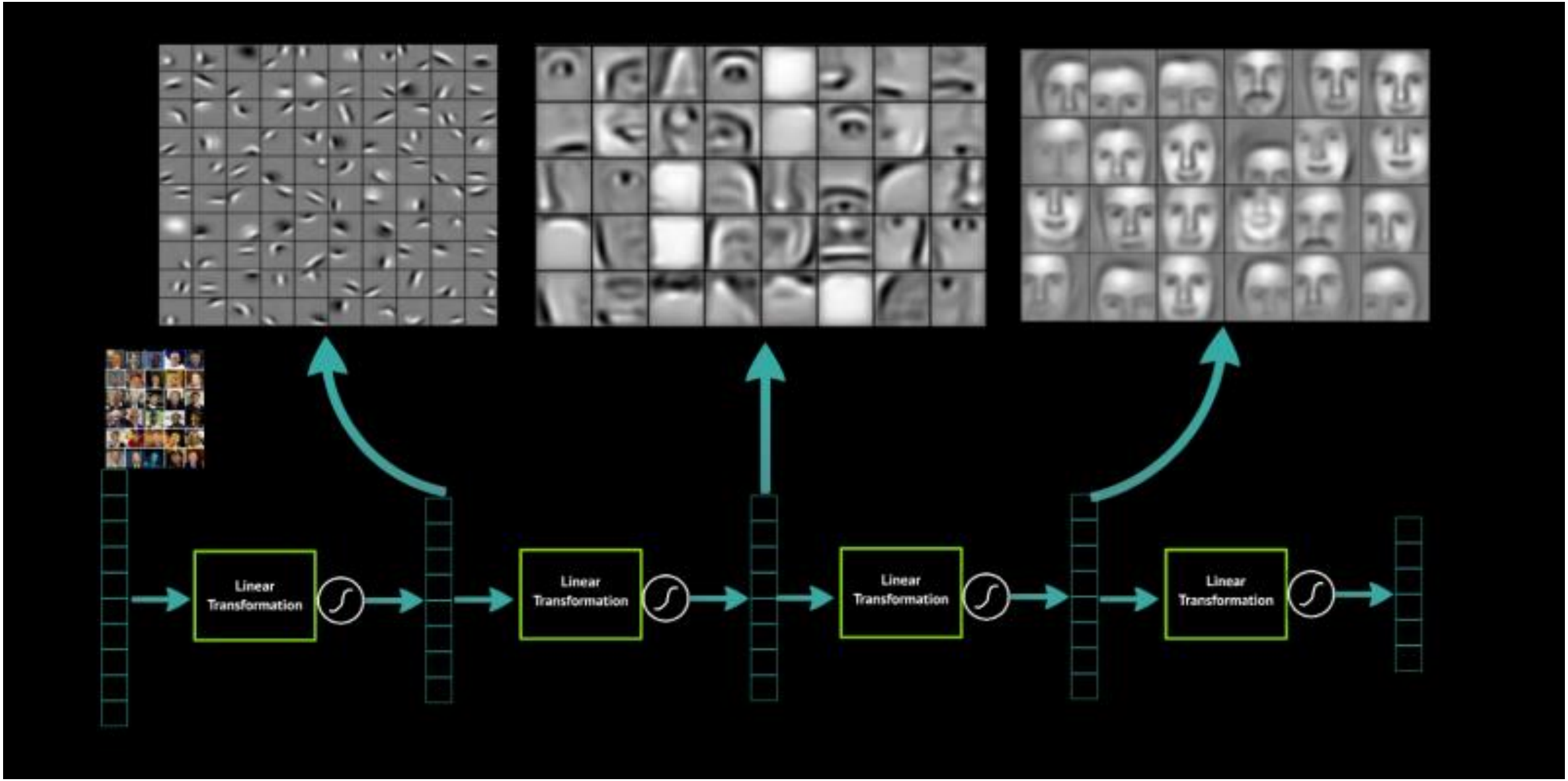


$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

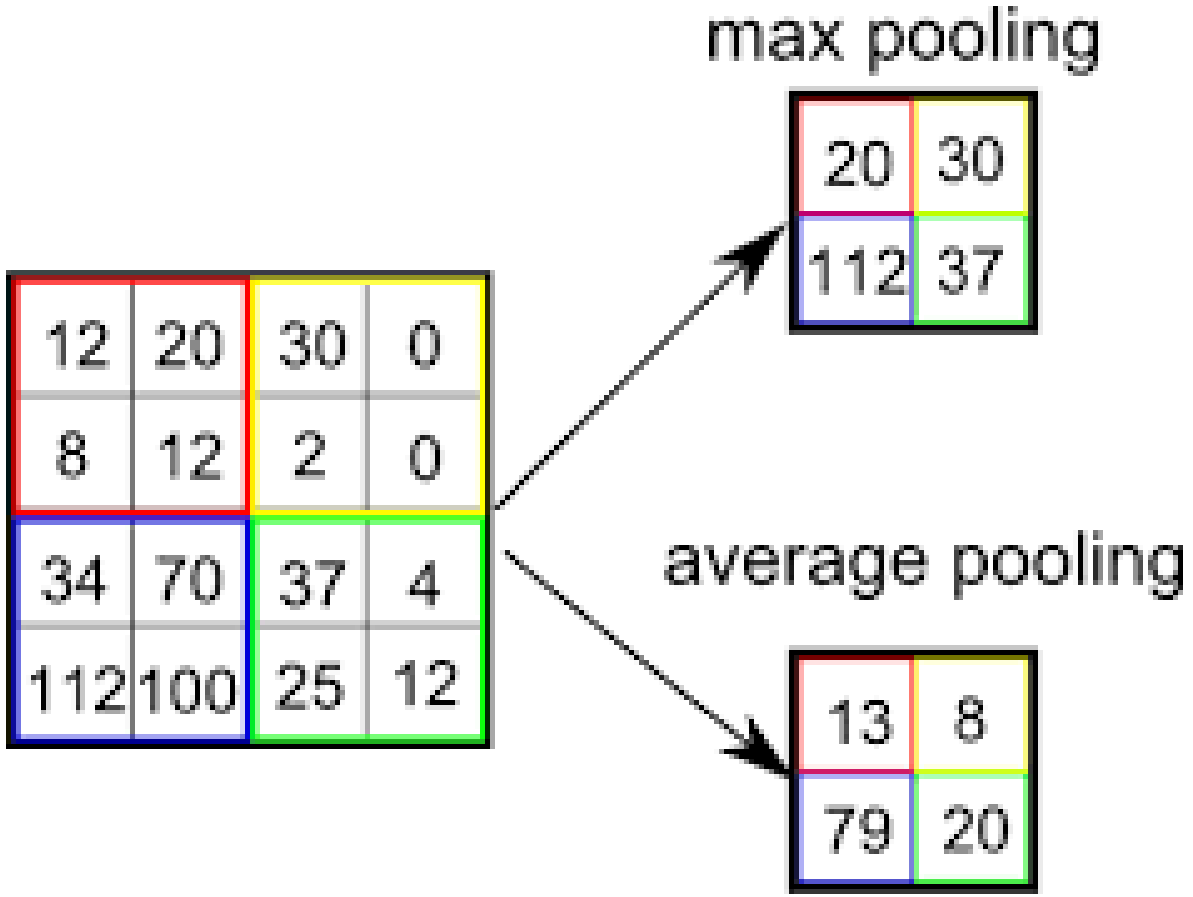
Next convolutions



Convolution layers



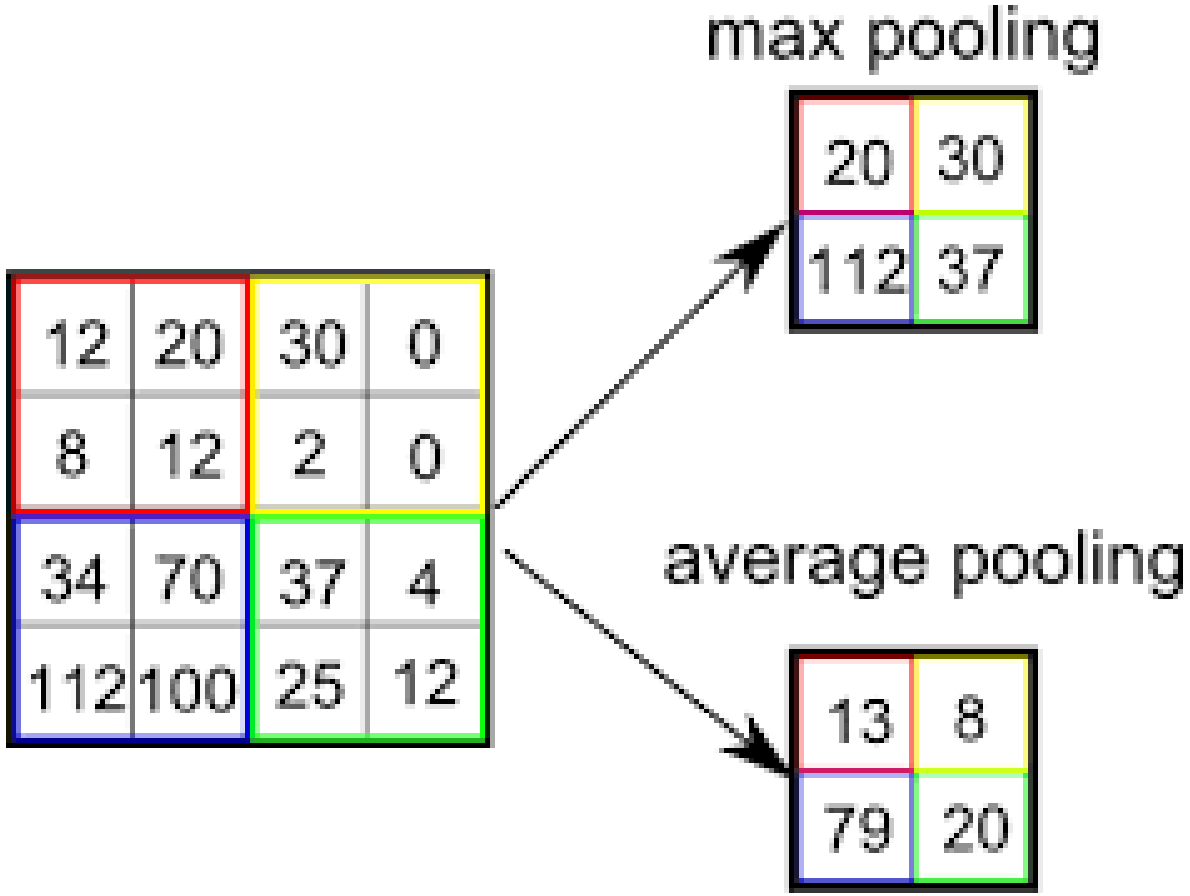
Pooling


max pooling
20 30
112 37
12 20 30 0
8 12 2 0
34 70 37 4
112 100 25 12
average pooling
13 8
79 20

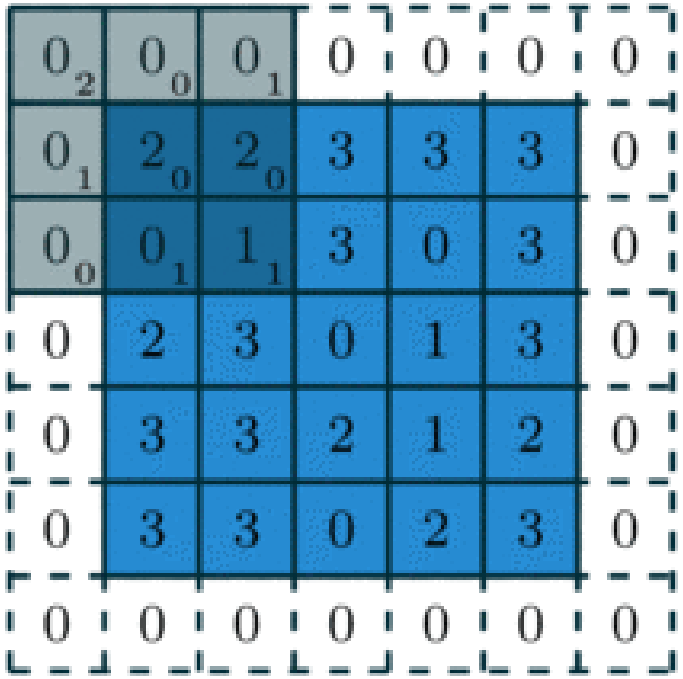
Pooling

Kernel 2x2

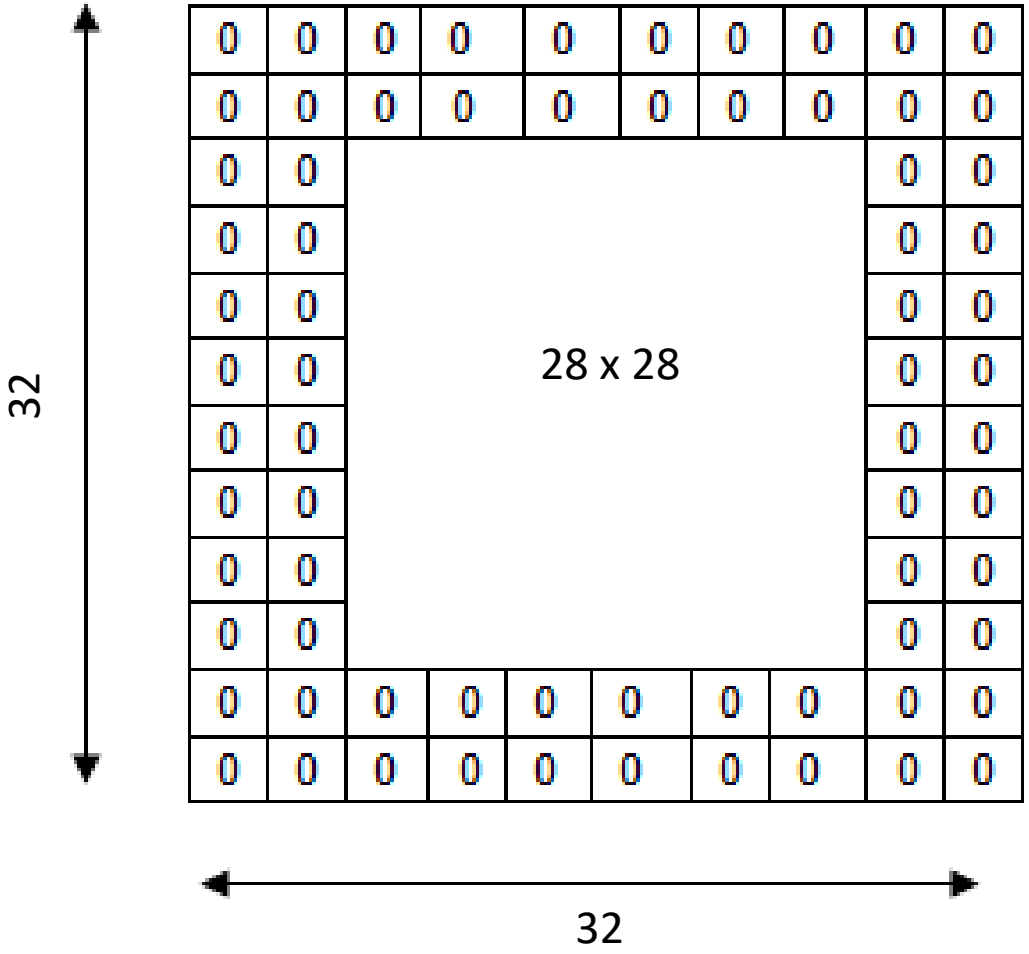
Stride 2



Padding

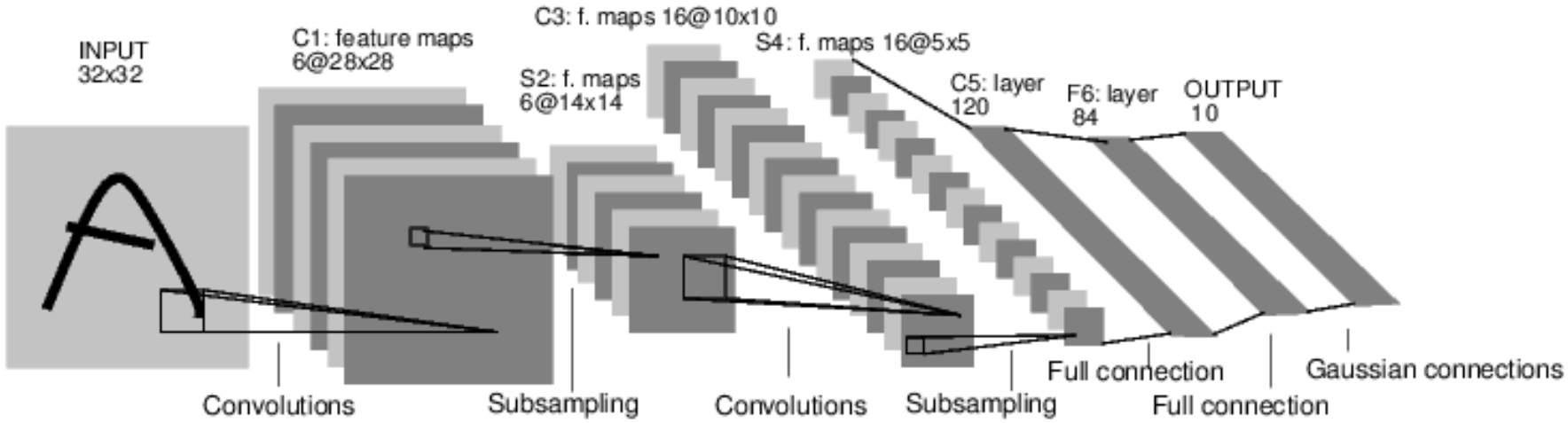


1	6	5
7	10	9
7	10	8



Deep learning

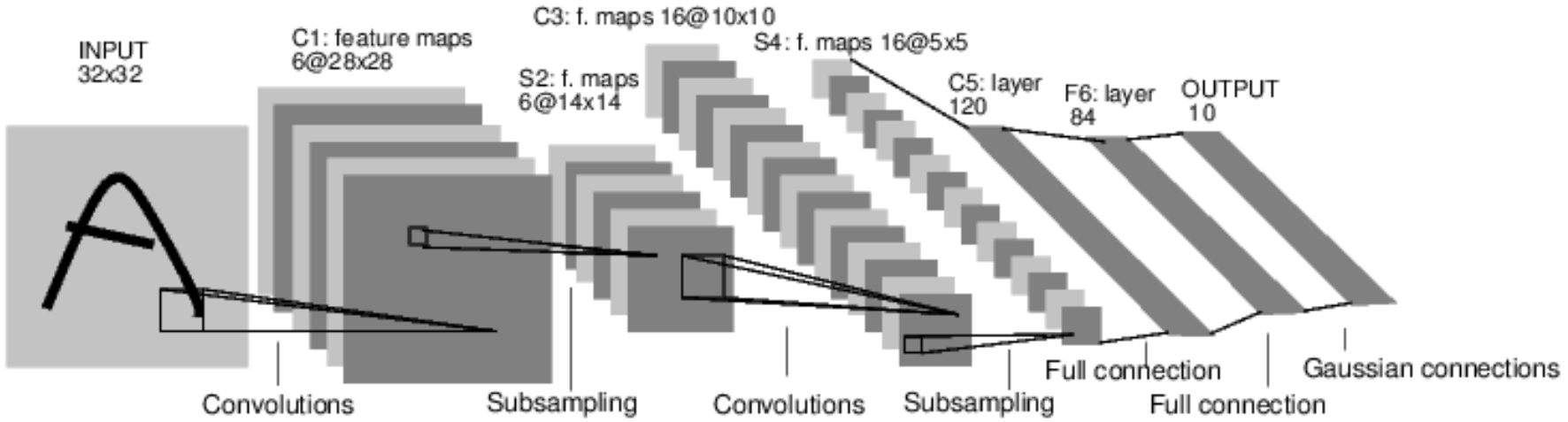
LeNet-5 (1998)



(5x5)x6 + (5x5x6)x16 + (16x5x5)x120 + 120x84 + 84x10 = 61470

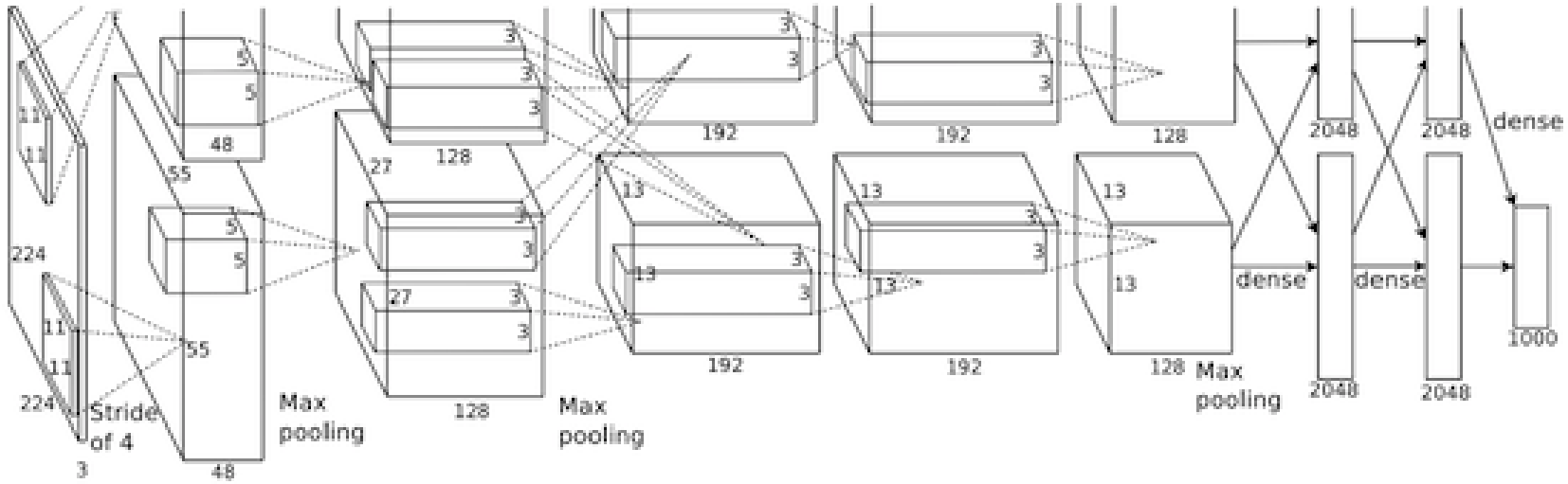
Deep learning

LeNet-5 (1998)



(5x5)x6 + (5x5x6)x16 + (16x5x5)x120 + 120x84 + 84x10 = 61470

AlexNet (2012)

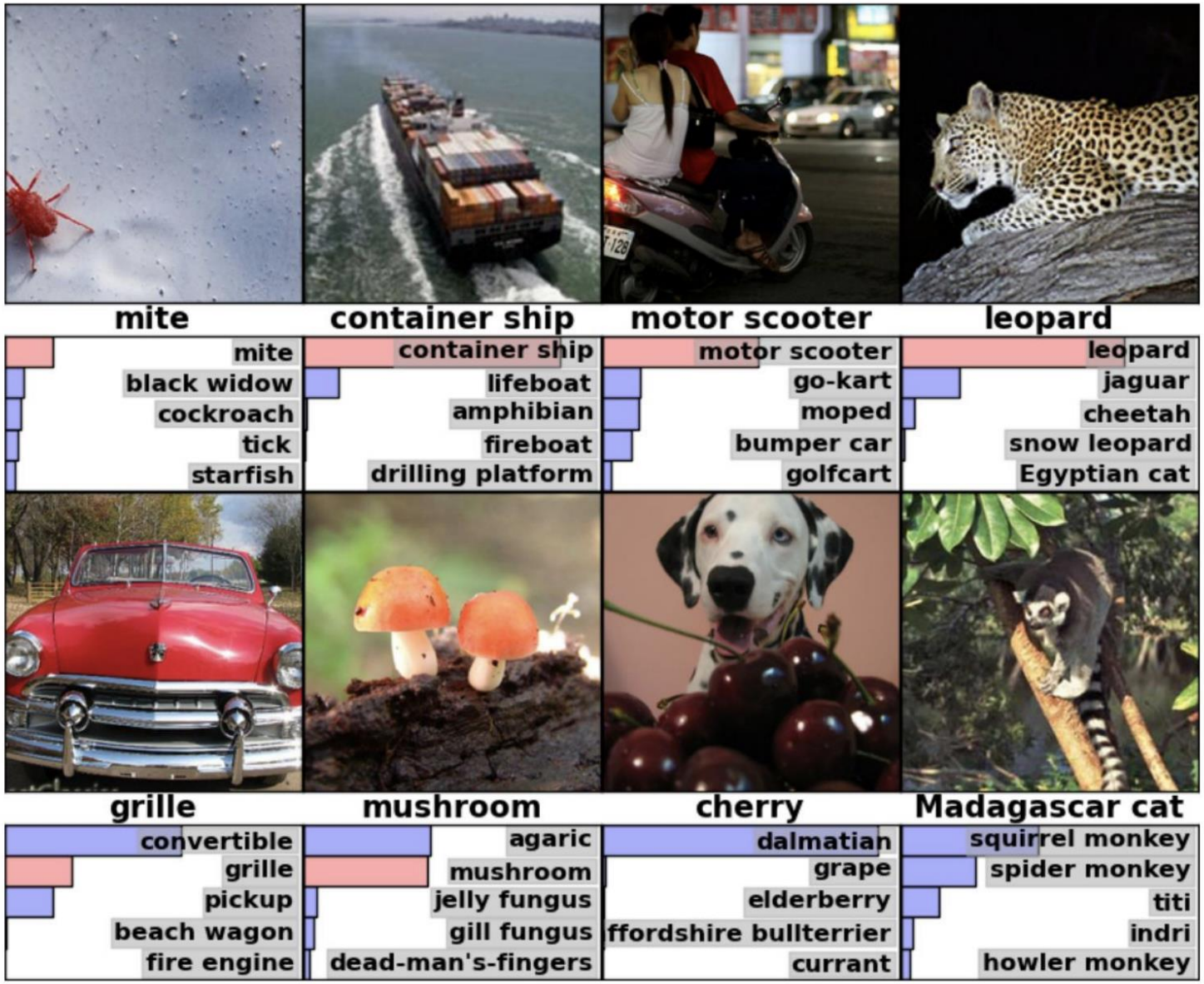


(11x11x3)x48 + (5x5x48)x128 + (3x3x128)x192x2 + (3x3x192)x192 + (3x3x192)x128 + (13x13x128)x2048x2 + 2048x2048x2 + 2048x1000 = 100 207 632

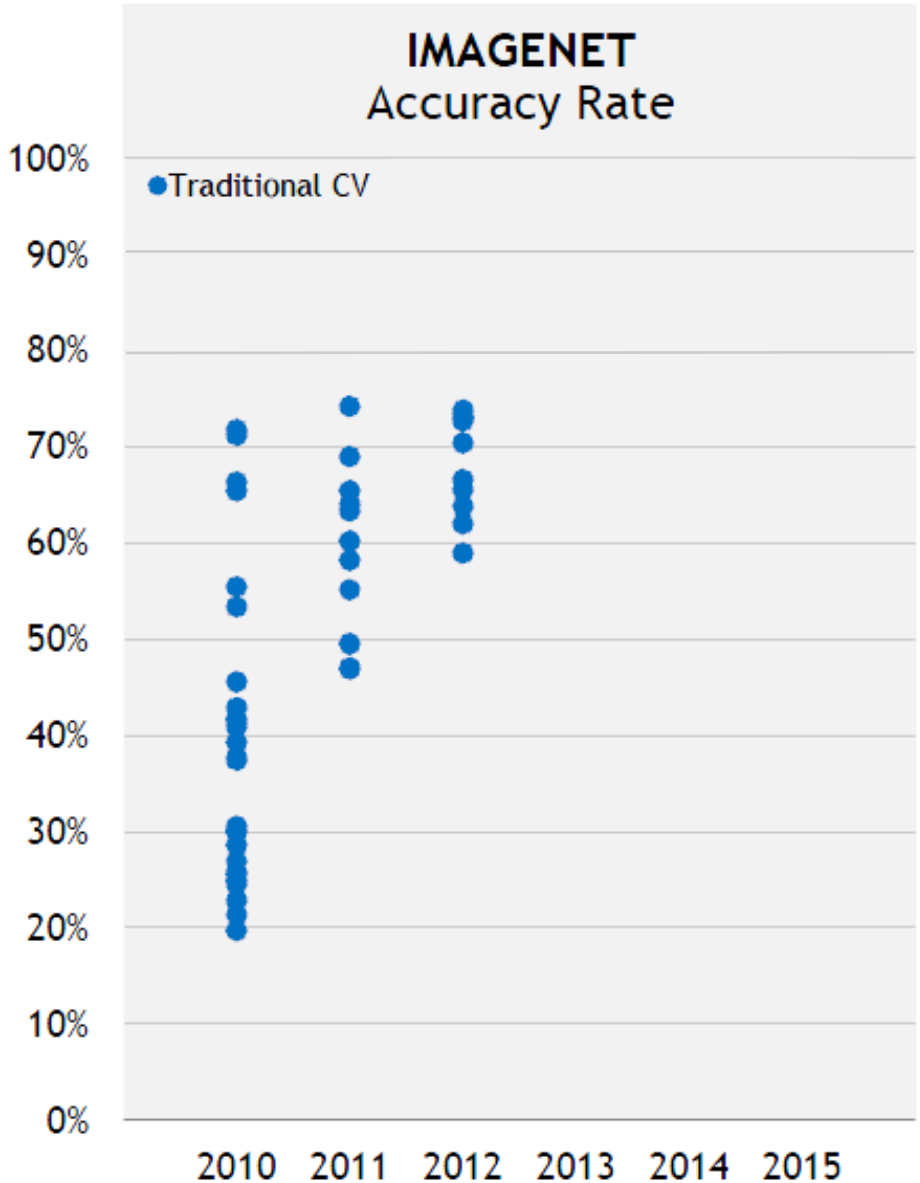
ImageNet Challenge

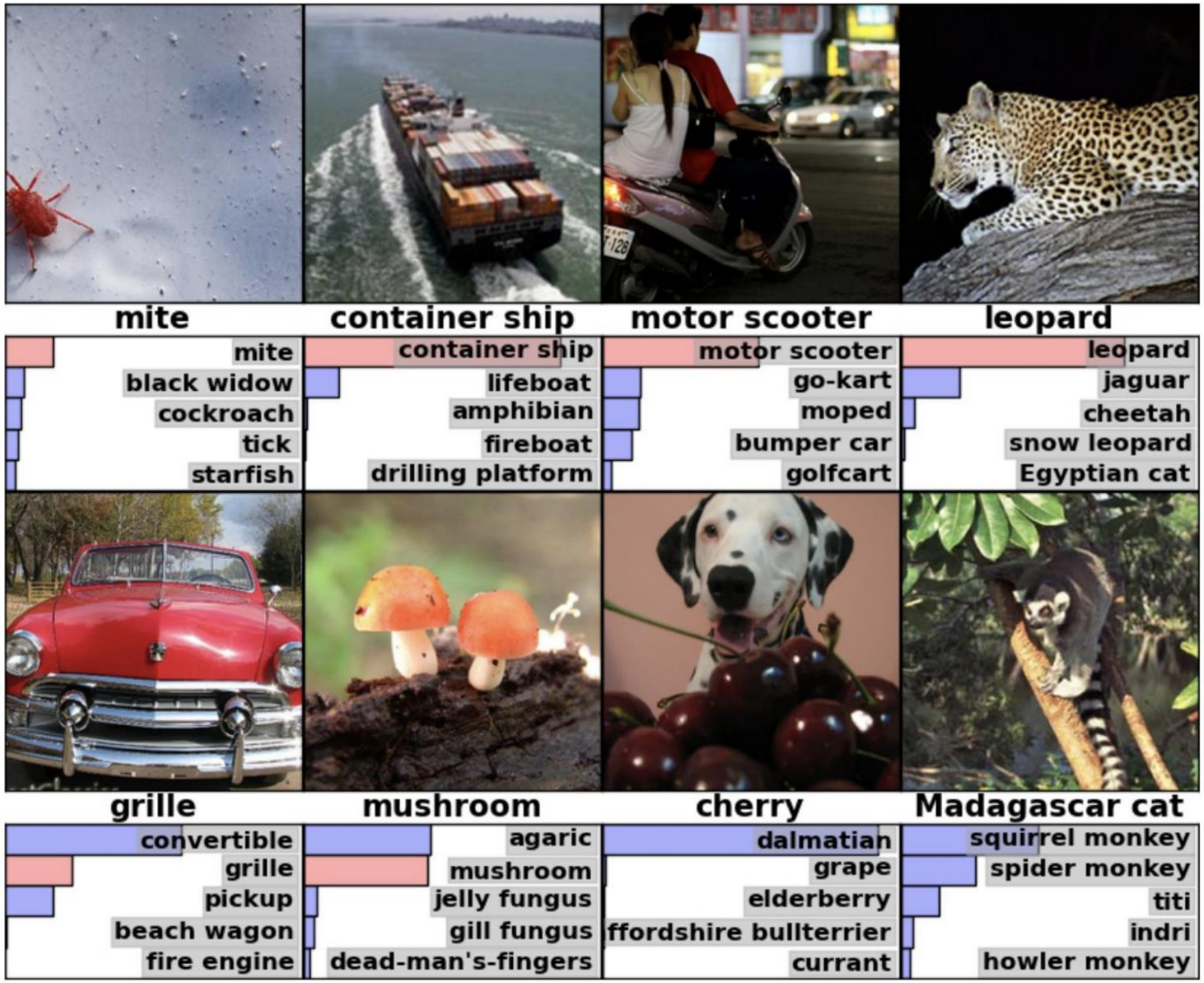
IMAGENET

- 1,000 object classes (categories).
- Images:
  - 1.2 M train
  - 100k test.

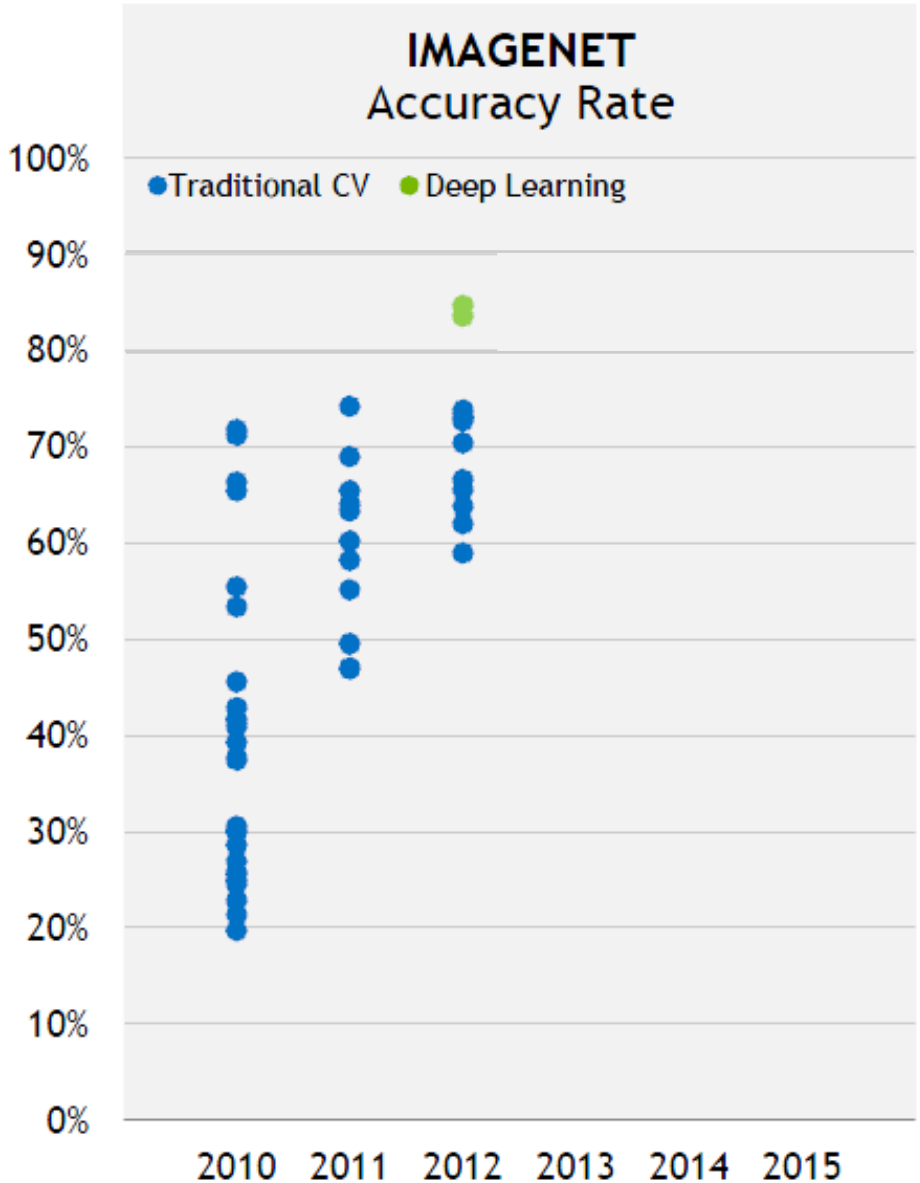


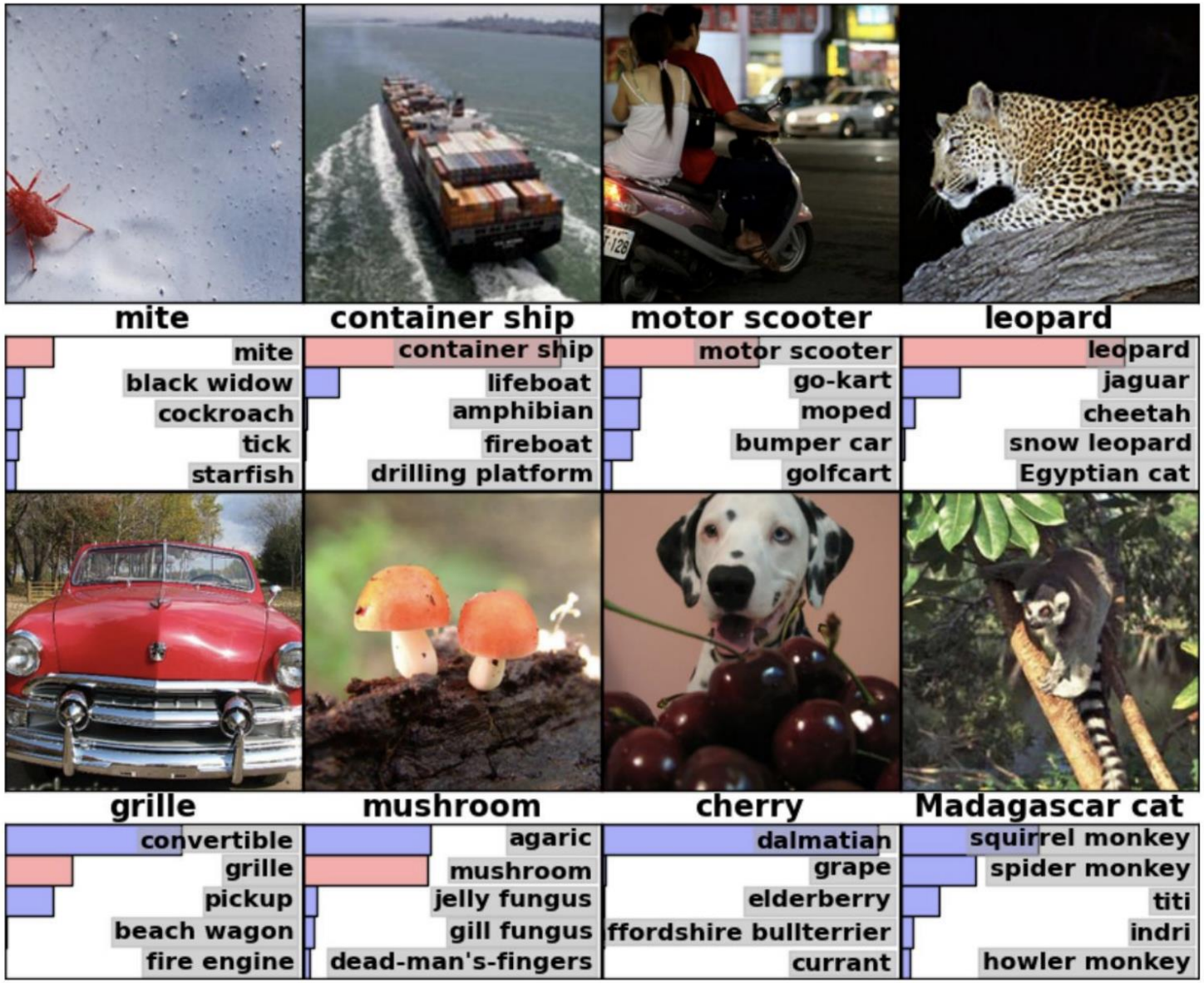
ImageNet Challenge



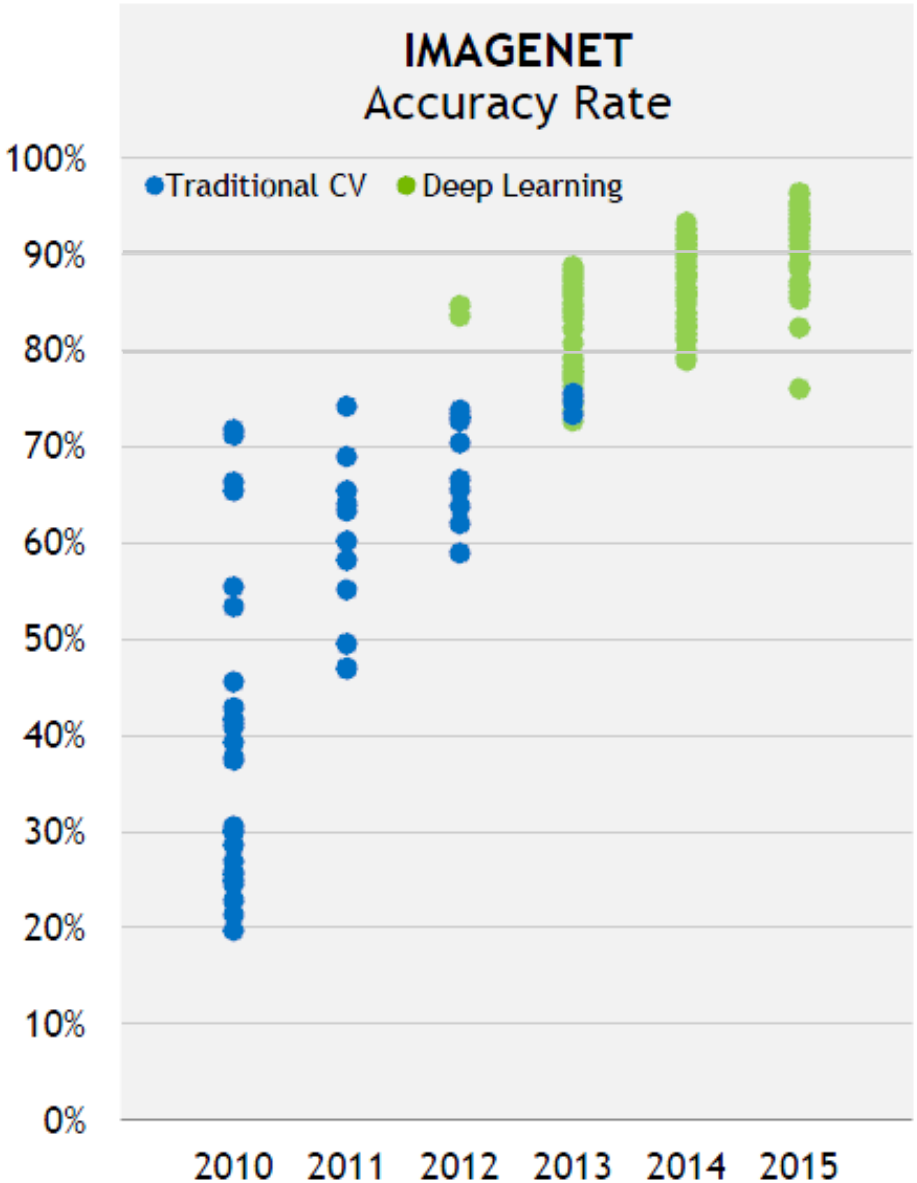


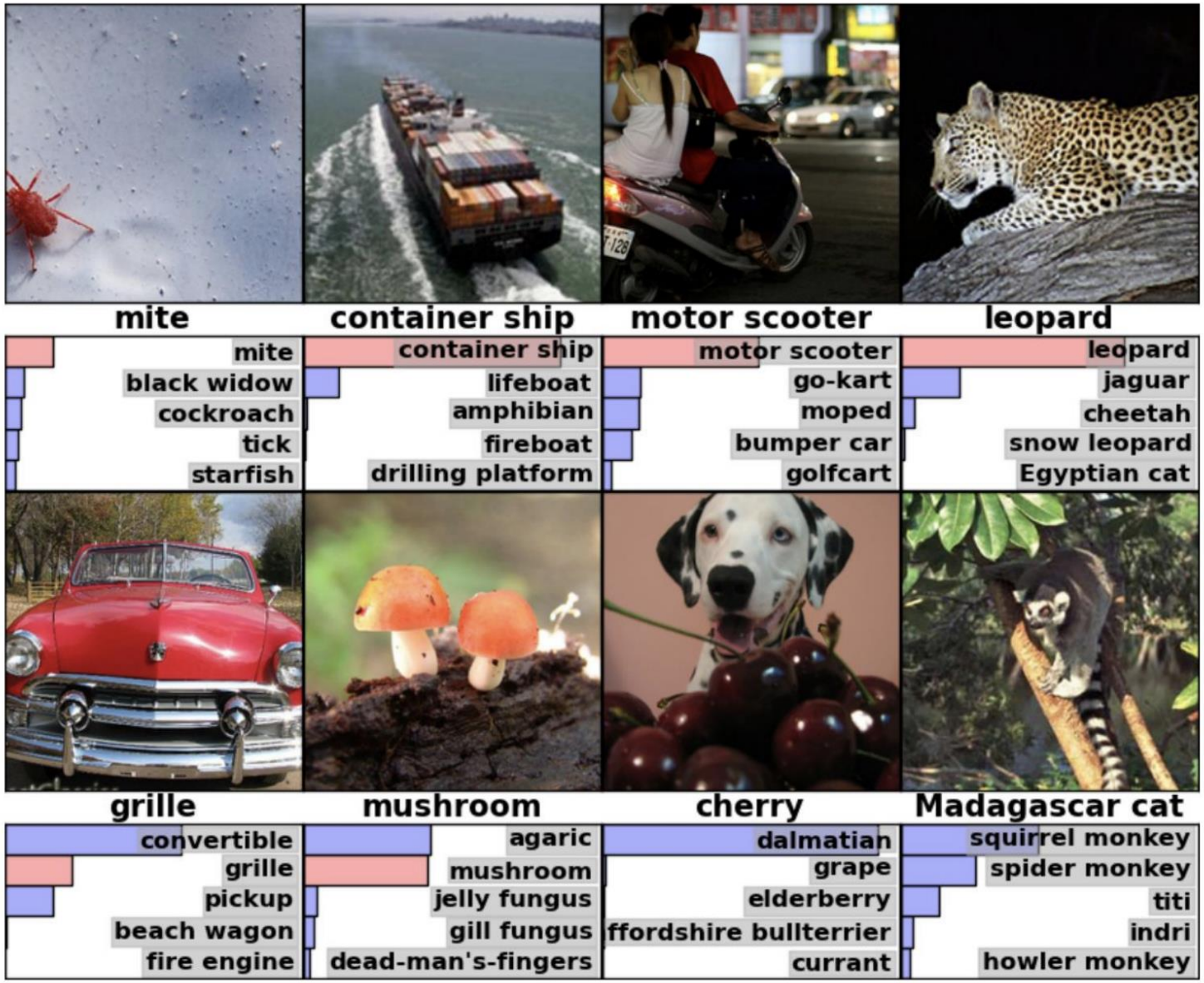
ImageNet Challenge





ImageNet Challenge





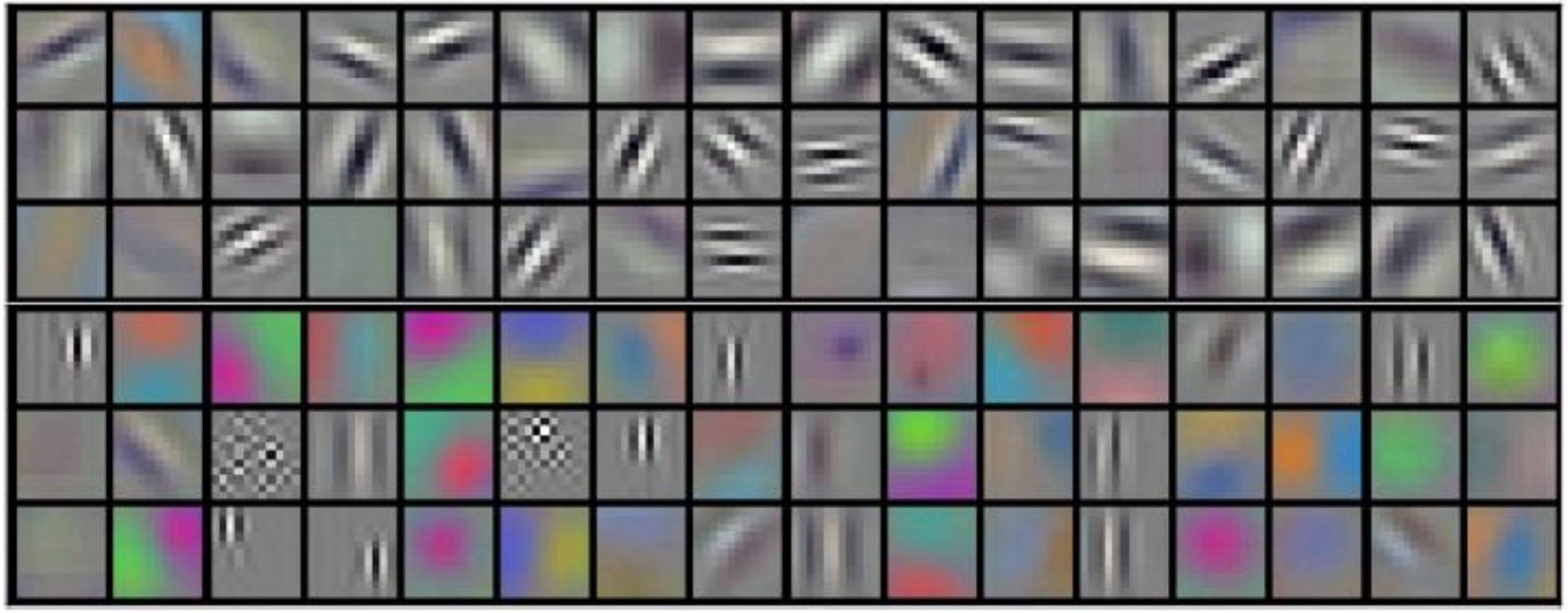
AlexNet

- Scale all images to 256x256, then take random 224x224 patches and mirror them.
- Subtract average pixel value from each pixel.
- ReLU ($f(x) = \max(0, x)$)
- Dropout 0.5
- Batch size 128, SGD with momentum (0.9), L2 weight decay ($\lambda = 0.0005$)

AlexNet, first convolution

GPU1

GPU2



“The kernels on GPU 1 are largely color-agnostic, while the kernels on on GPU 2 are largely color-specific. This kind of specialization occurs during every run and is independent of any particular random weight initialization (modulo a renumbering of the GPUs)”

Visualization of convolutions

Deep Visualization Toolbox

yosinski.com/deepvis

#deepvis


Jason Yosinski


Jeff Clune


Anh Nguyen


Thomas Fuchs


Hod Lipson

Cornell University

University of Wyoming

Jet Propulsion Laboratory
California Institute of Technology

Very Deep Convolutional Networks (VGGNet), 2014

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

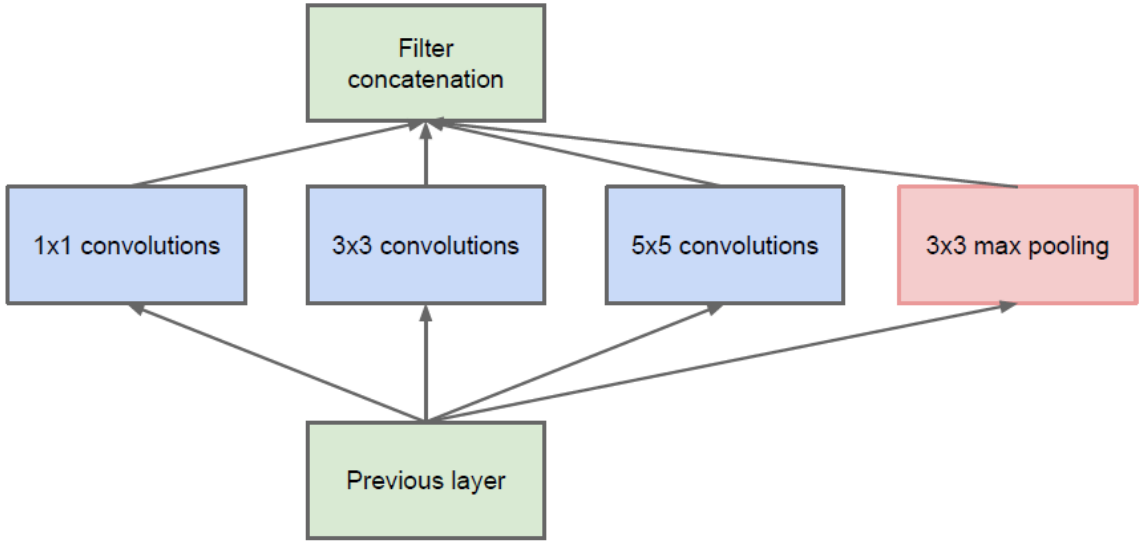
Very Deep Convolutional Networks (VGGNet), 2014

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

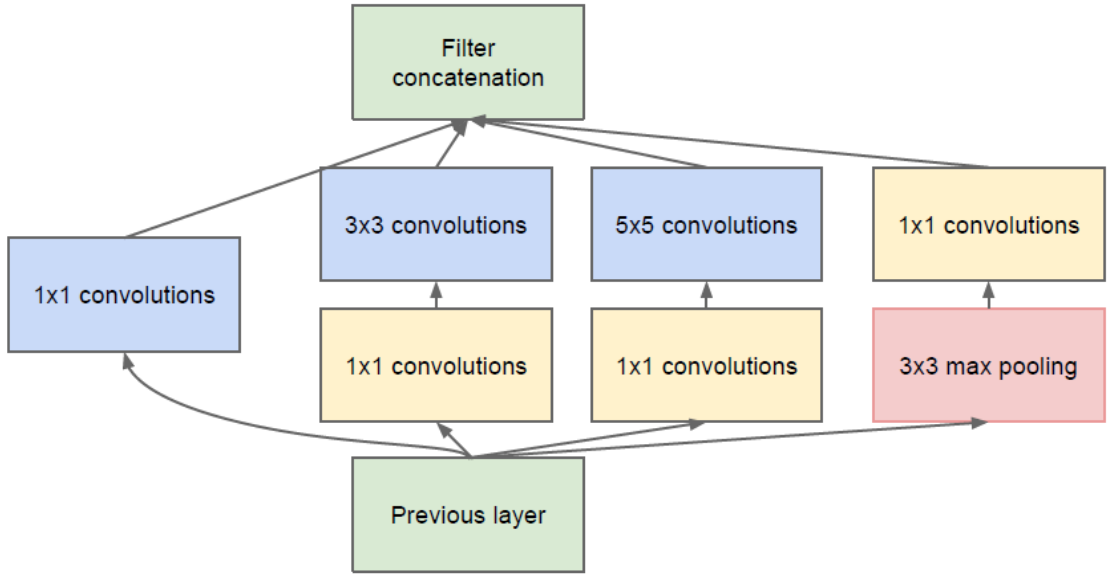
ImageNet Challeng result:

VGGNet (D+E) – 7.3% (2^{nd})

Inception (Google 2014)



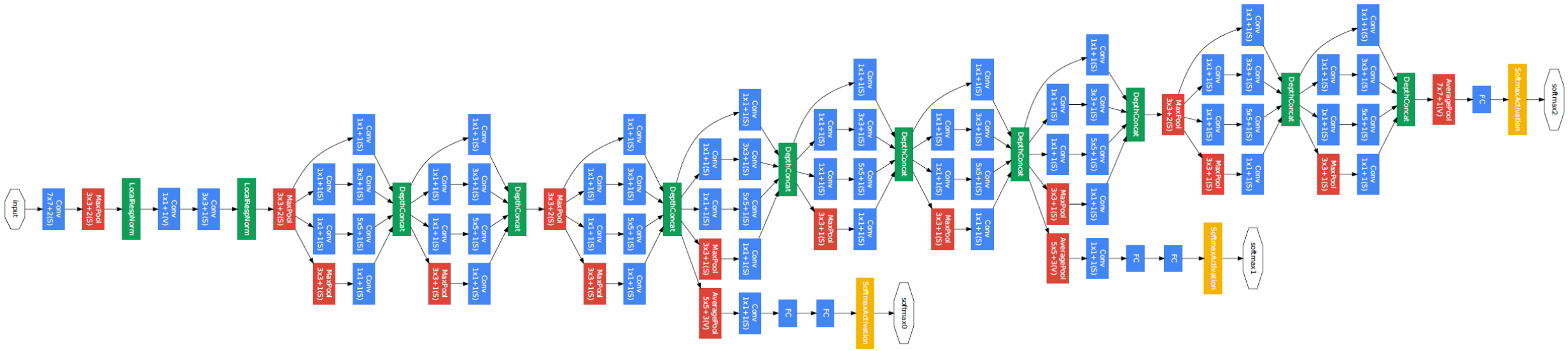
(a) Inception module, naïve version



(b) Inception module with dimension reductions

Using 1x1 convolutions to decrease a number of maps in layers.

GoogLeNet



Convolutions

MaxPooling

Concatenation

FC

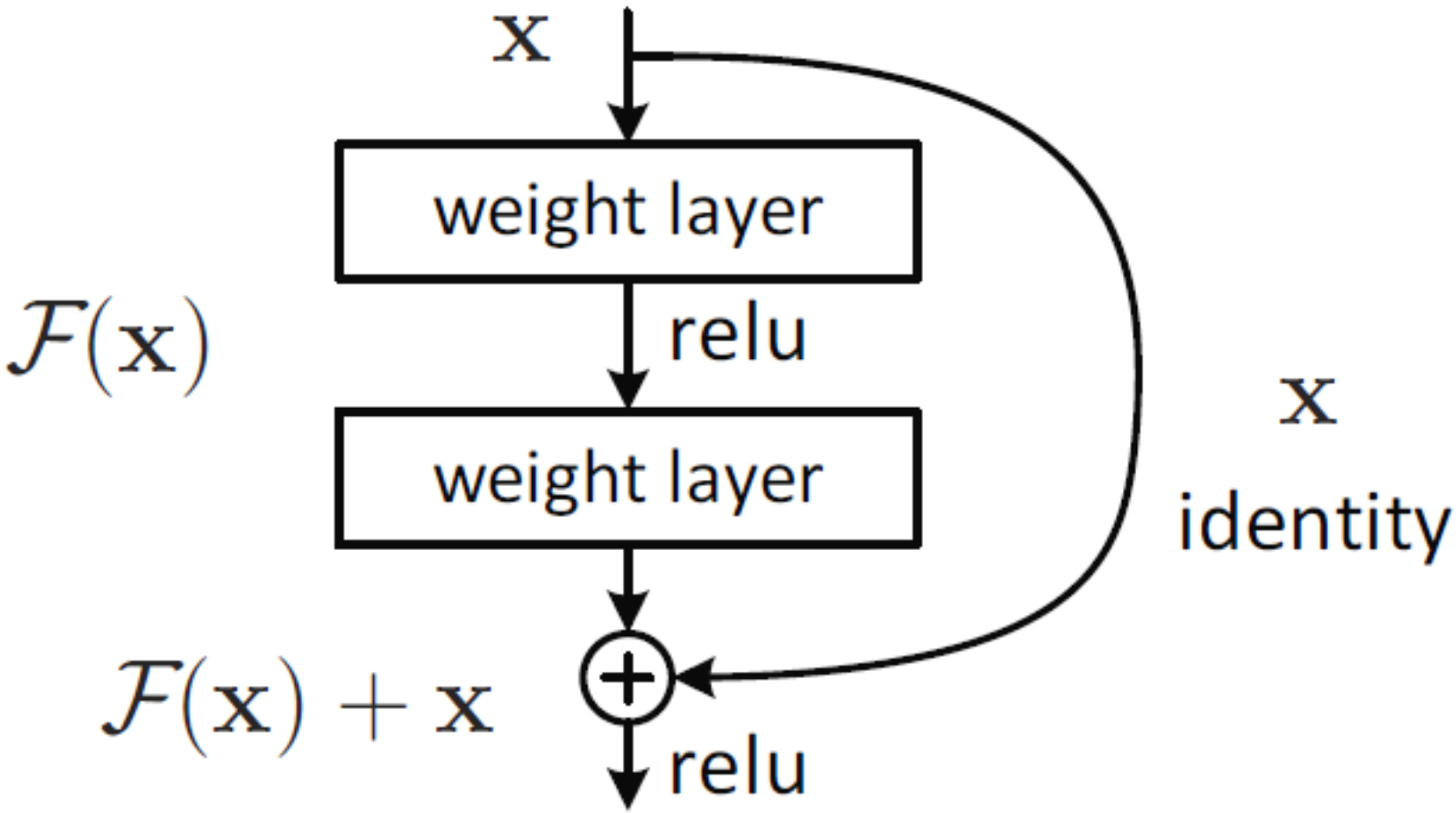
AveragePooling

Softmax

GoogLeNet (2014)

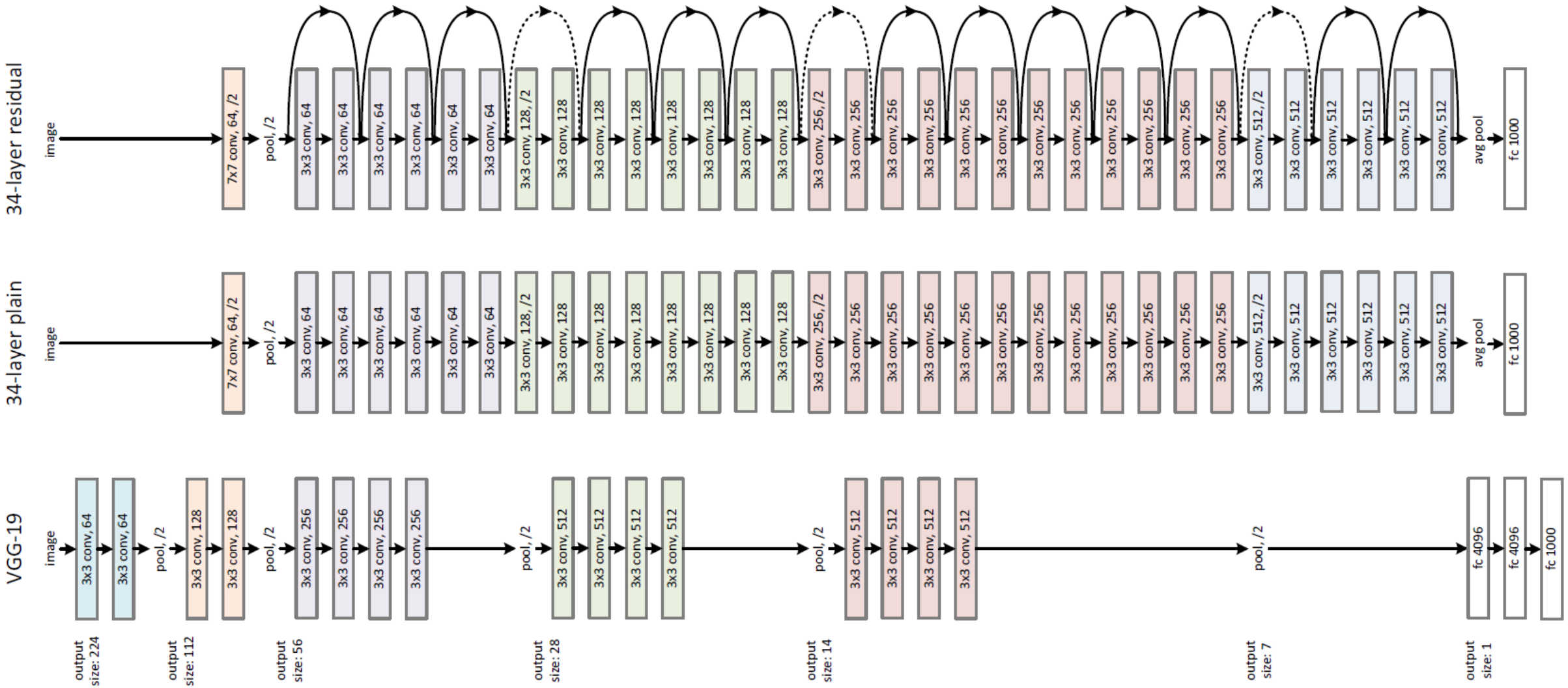
Team	Year	Place	Error (top-5)
SuperVision	2012	1st	16.4%
Clarifai	2013	1st	11.7%
MSRA	2014	3rd	7.35%
VGG	2014	2nd	7.32%
GoogLeNet	2014	1st	6.67%

Residual Learning



Now called “skip-connections”.

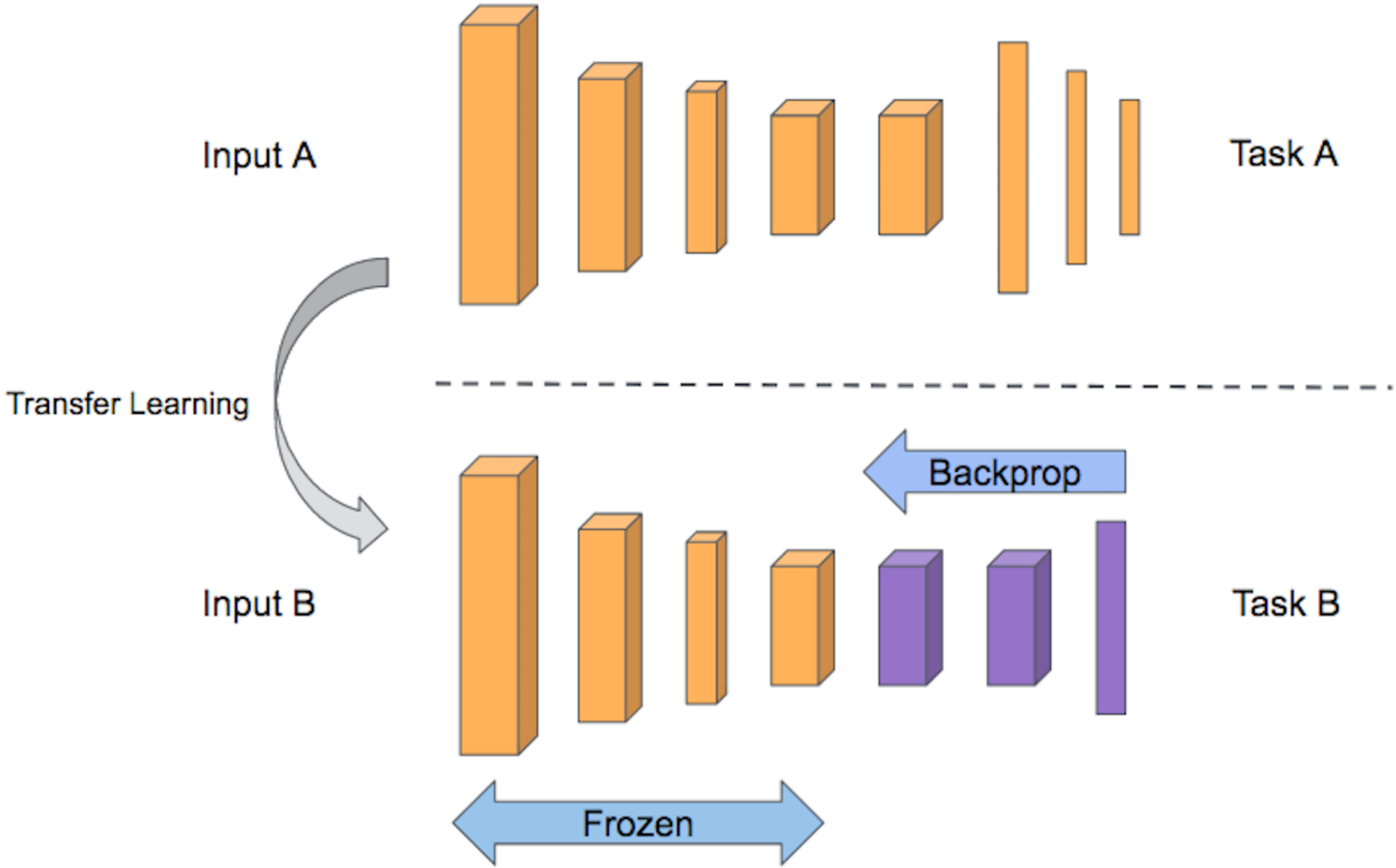
ResNet (2015)


VGG-19
image
3x3 conv, 64
3x3 conv, 64
pool, /2
3x3 conv, 128
3x3 conv, 128
pool, /2
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
pool, /2
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
pool, /2
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
pool, /2
fc 4096
fc 4096
fc 1000
output size: 224
output size: 112
output size: 56
output size: 28
output size: 14
output size: 7
output size: 1
34-layer plain
image
7x7 conv, 64, /2
pool, /2
3x3 conv, 64
3x3 conv, 128, /2
3x3 conv, 128
3x3 conv, 256, /2
3x3 conv, 256
3x3 conv, 512, /2
3x3 conv, 512
avg pool
fc 1000
34-layer residual
image
7x7 conv, 64, /2
pool, /2
3x3 conv, 64
3x3 conv, 128, /2
3x3 conv, 128
3x3 conv, 256, /2
3x3 conv, 256
3x3 conv, 512, /2
3x3 conv, 512
avg pool
fc 1000

ResNet

method	top-5 err. (**test**)
VGG [41] (ILSVRC'14)	7.32
GoogLeNet [44] (ILSVRC'14)	6.66
VGG [41] (v5)	6.8
PReLU-net [13]	4.94
BN-inception [16]	4.82
ResNet (ILSVRC'15)	**3.57**

Transfer learning and finetuning


Input A
Task A
Transfer Learning
Backprop
Input B
Task B
Frozen

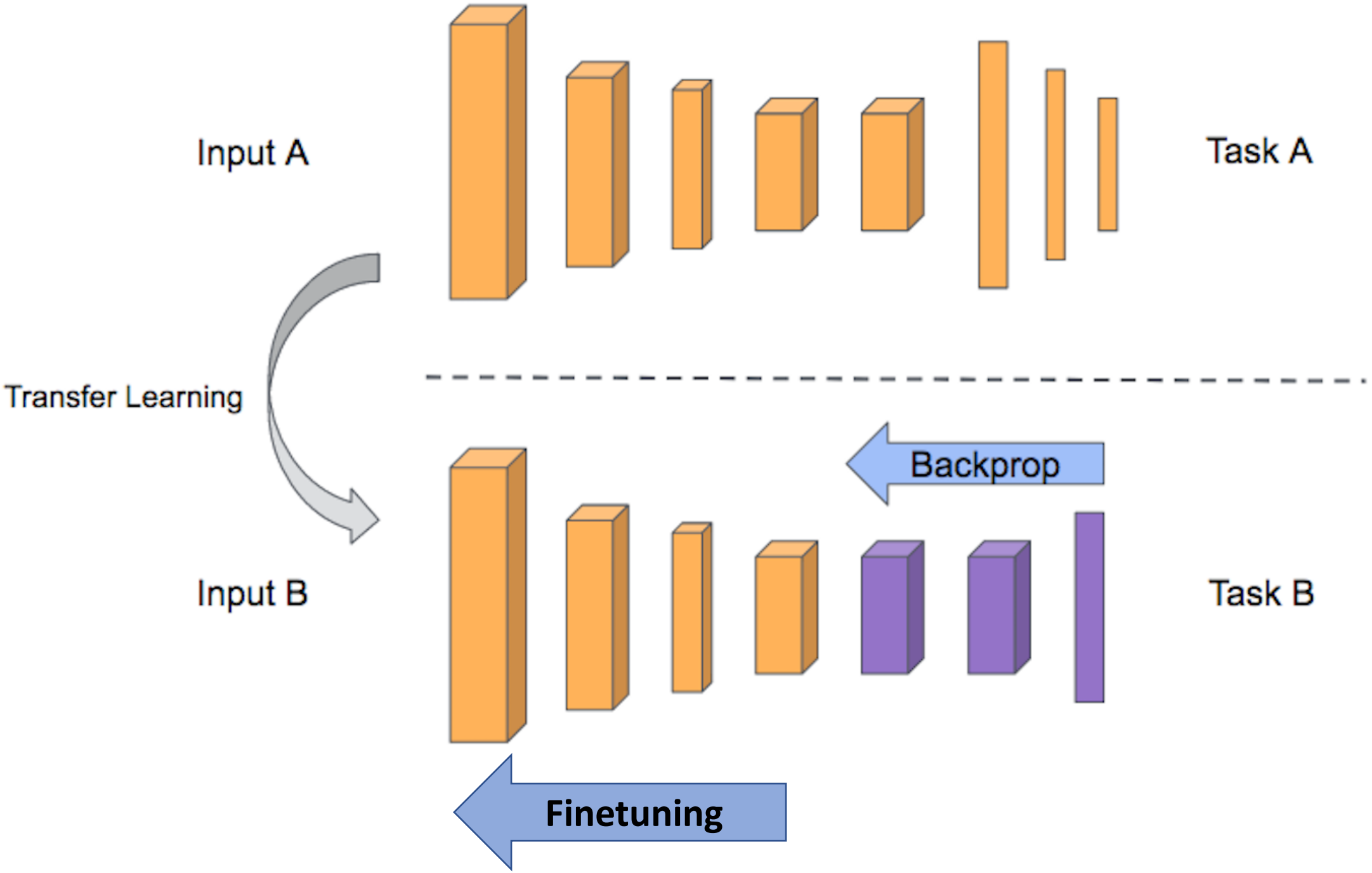
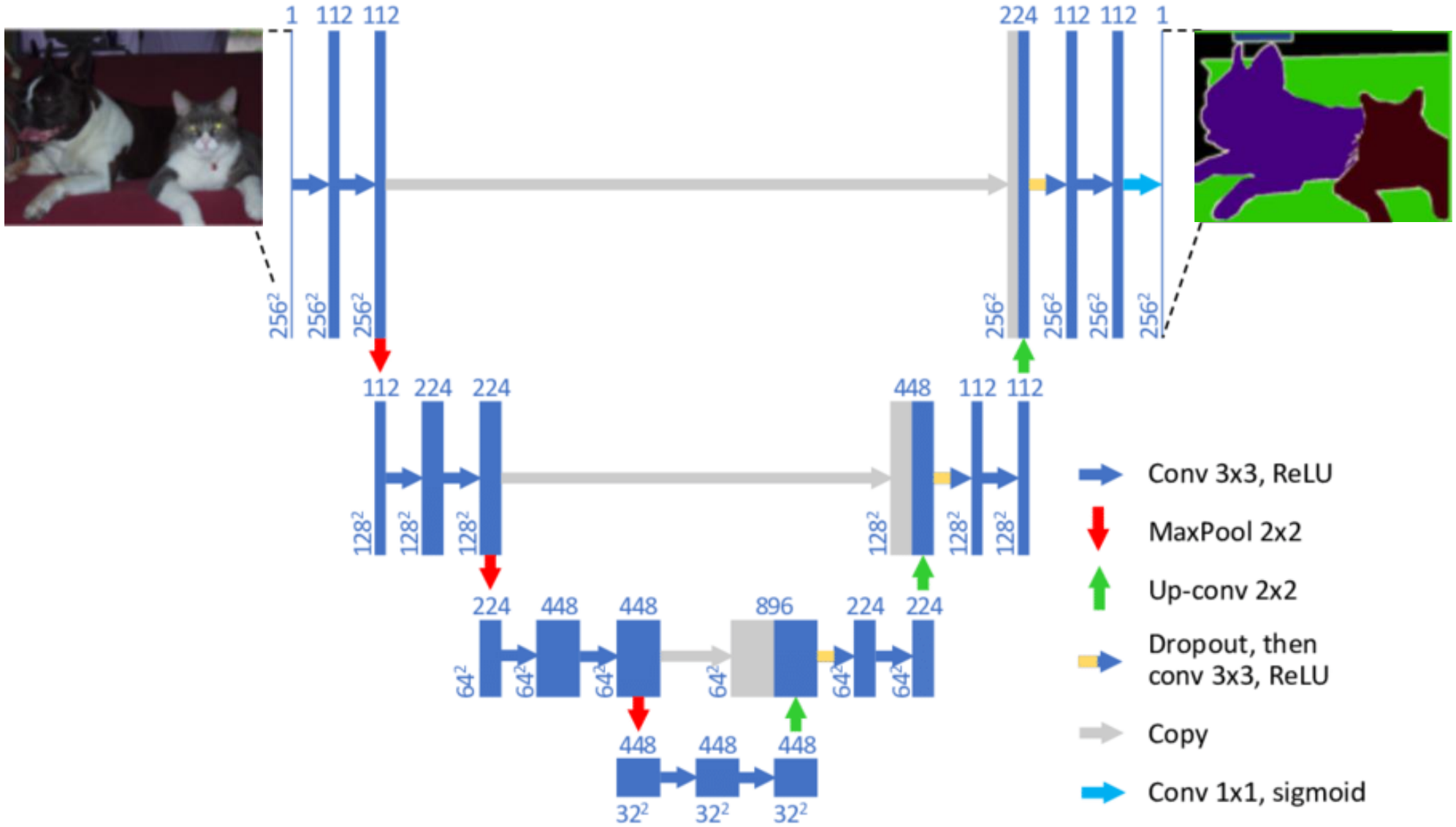
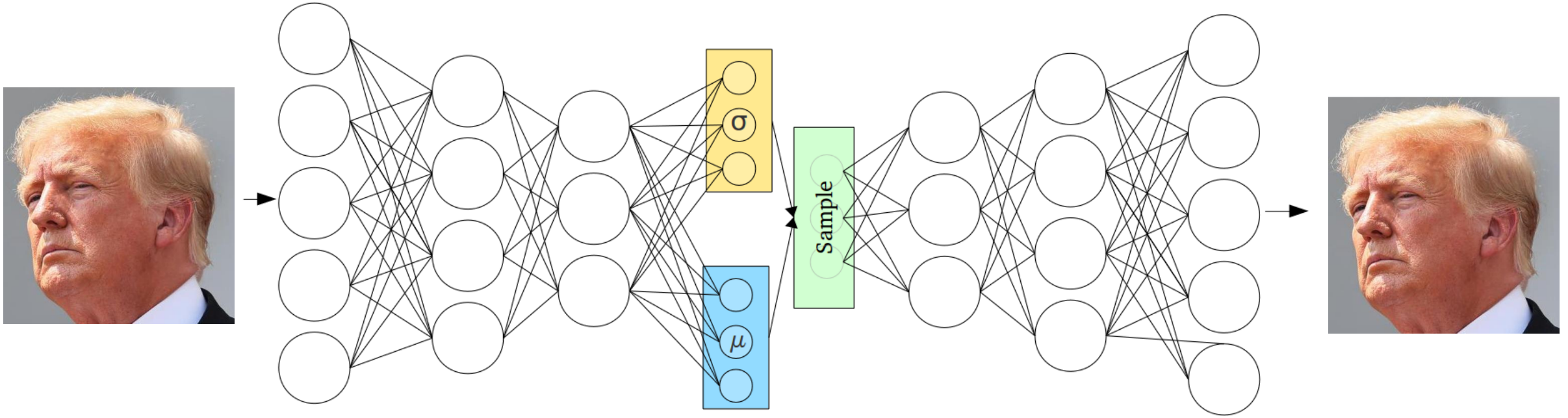

Input A
Task A
Transfer Learning
Backprop
Input B
Task B
Finetuning

Image segmentation and U-Net

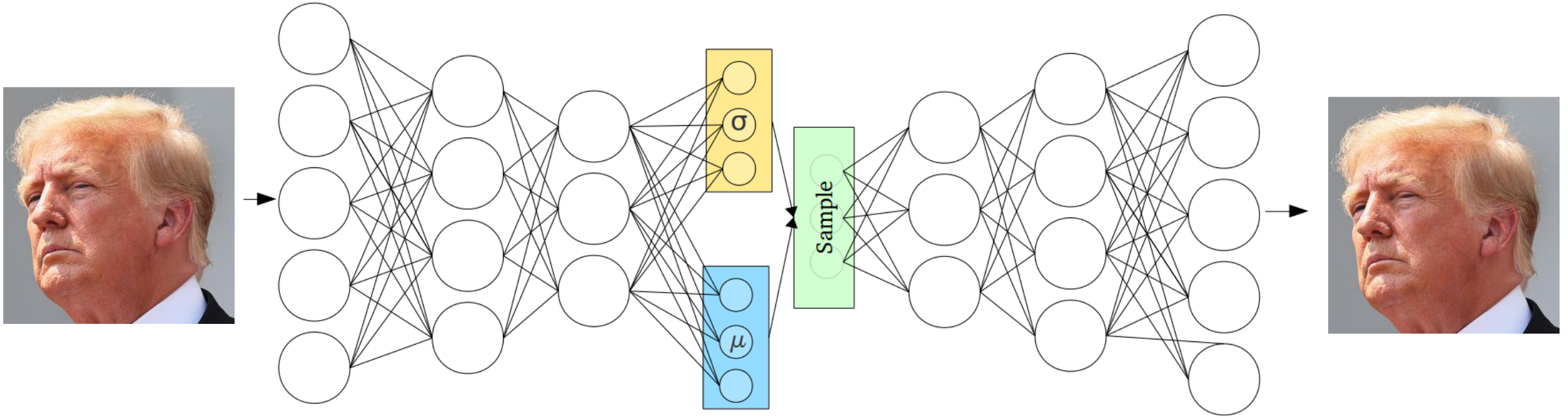


Variational Autoencoder (VAE)

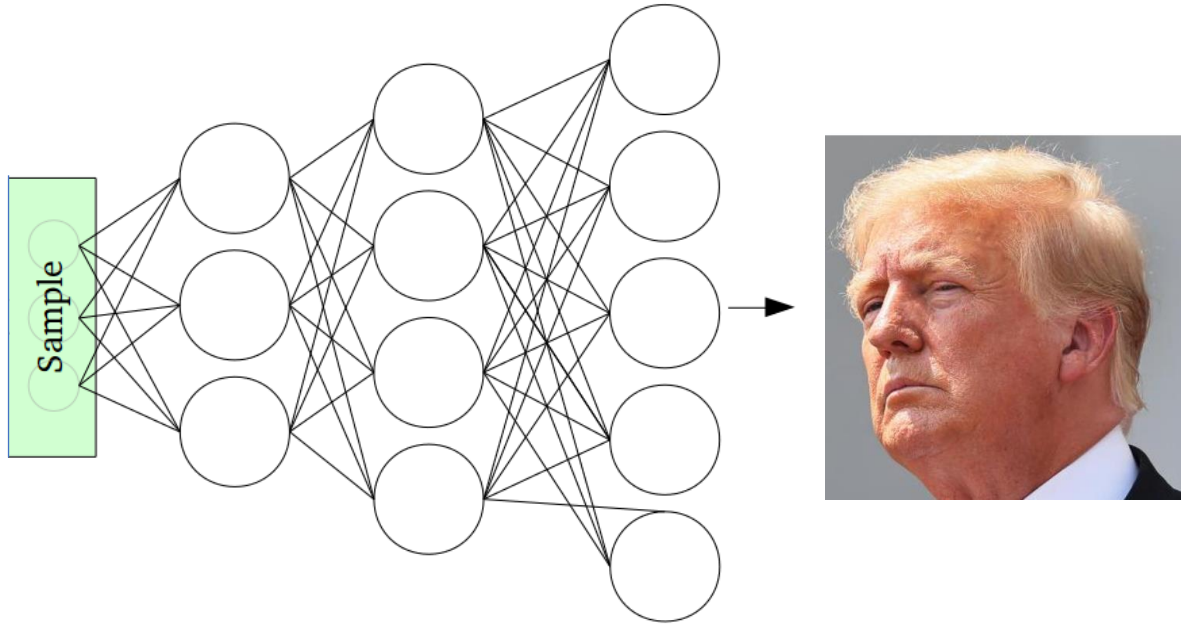


Variational Autoencoder (VAE)

Variational Autoencoder (VAE)



Generative Adversarial Network (GAN)



Generative Adversarial Network (GAN)