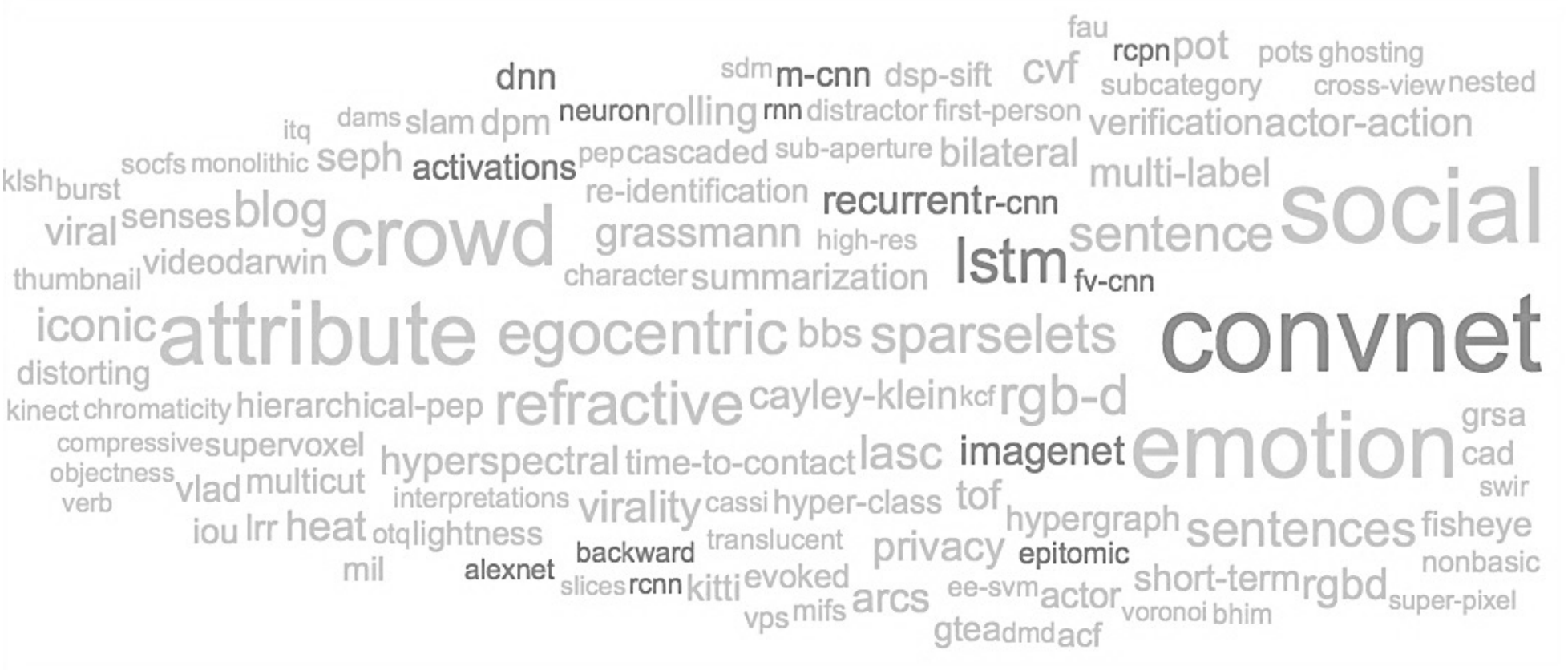
Natalia Neverova
INSA-Lyon, LIRIS CNRS

Introduction to deep learning

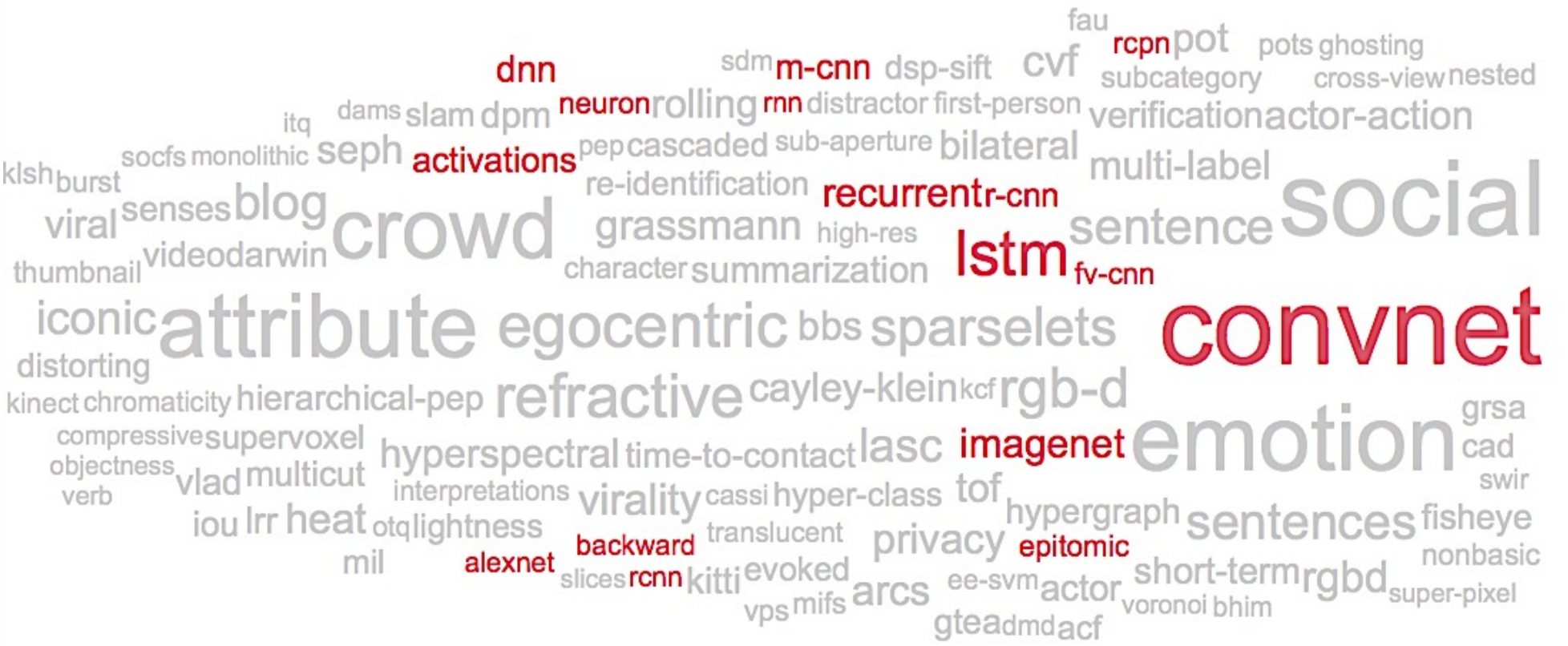
{{ softshake }} 2015
http://soft-shake.ch
@GENÈVE


UNIVERSITY of GUELPH
CHANGING LIVES IMPROVING LIFE
INSA
INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON
LIRIS
Google

Computer vision 2015:
what changed since ten years ago?



Computer vision 2015:
what changed since ten years ago?



What is the idea?

An example: classical supervised learning setting

x ⟶ y

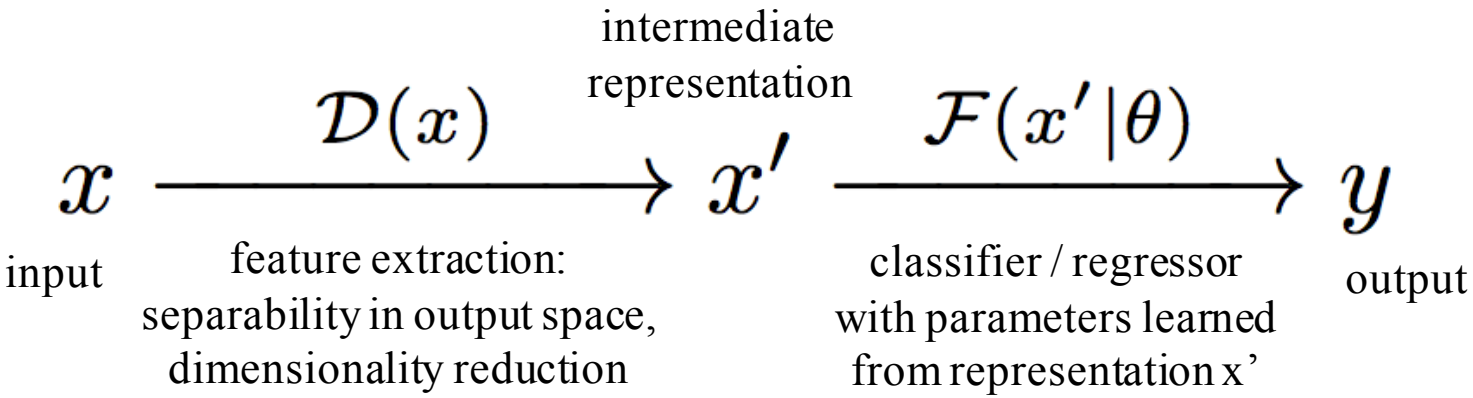
input output

object recognition ⟶ {dog, cat, pokemon}

super resolution ⟶

What is the idea?

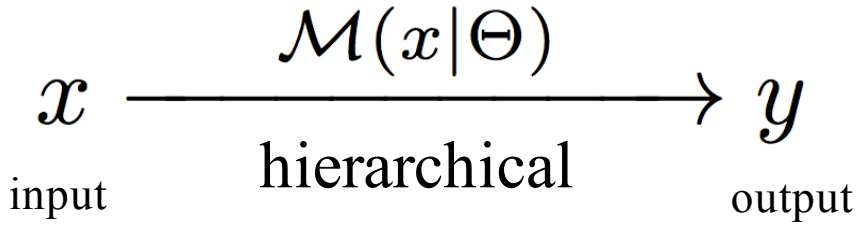
An example: classical supervised learning setting



where $\mathcal{D}(x)$...

What is the idea?

deep *learning* of data representations



joint learning
of representations with
increased levels of
abstraction
+ classification or
regression

What is the idea?

biologically inspired model

huge amount of training samples

general and suitable for any input

supervised, unsupervised and reinforcement learning

What has been achieved?

loosely biologically inspired models

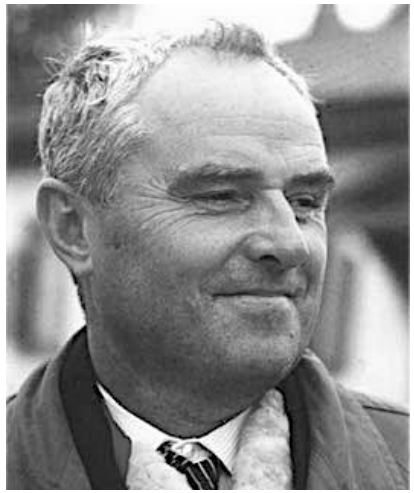
huge amount of training samples when available

general and suitable for many kinds of inputs after adaptation

supervised learning in a product

unsupervised and reinforcement learning – work in progress

A bit of history



early 1960s

Alexey Ivakhnenko

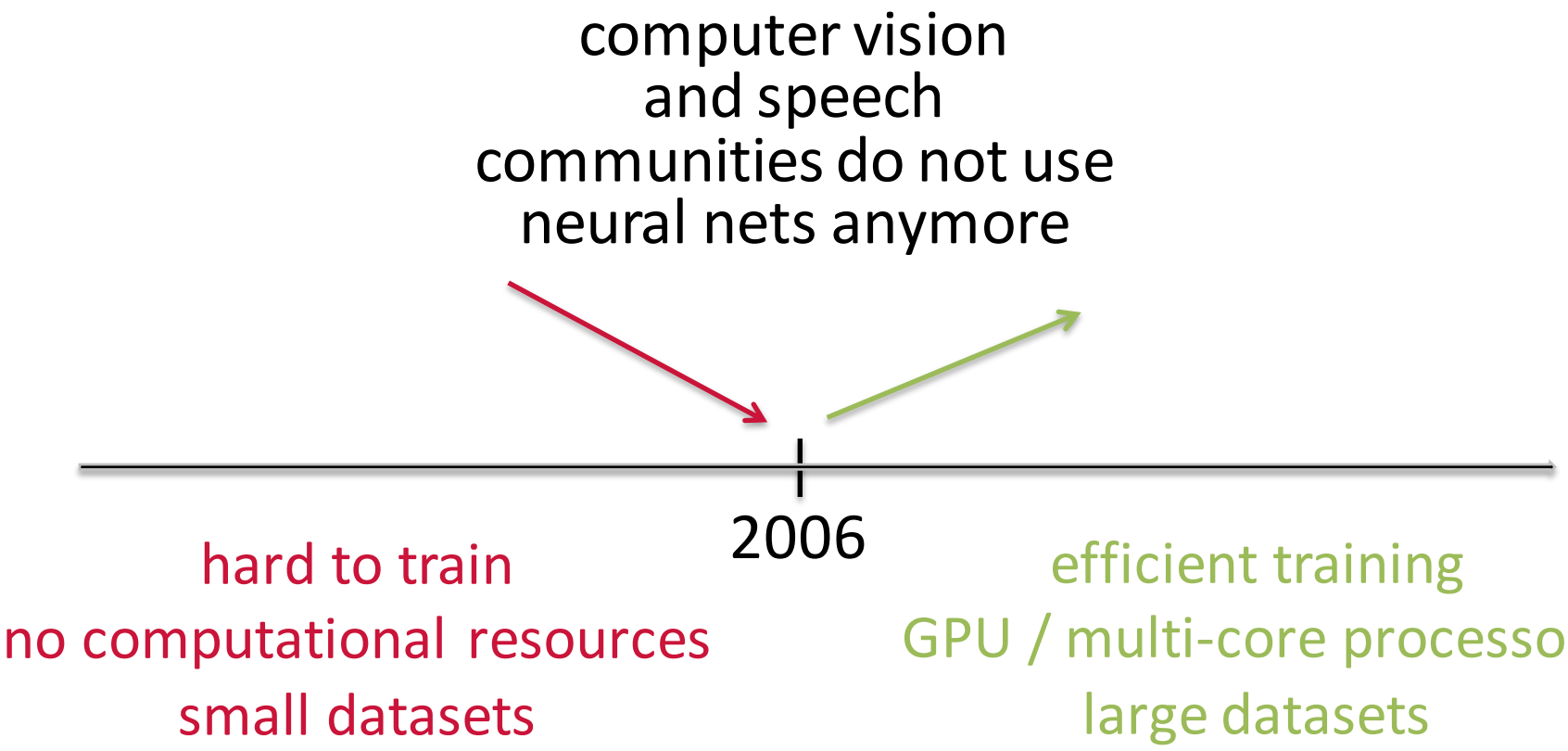
first works on deep neural networks

1986

Geoffrey Hinton

backpropagation algorithm in its current form

A bit of history



A bit of history

2011

Microsoft

breakthrough
in speech recognition

2012

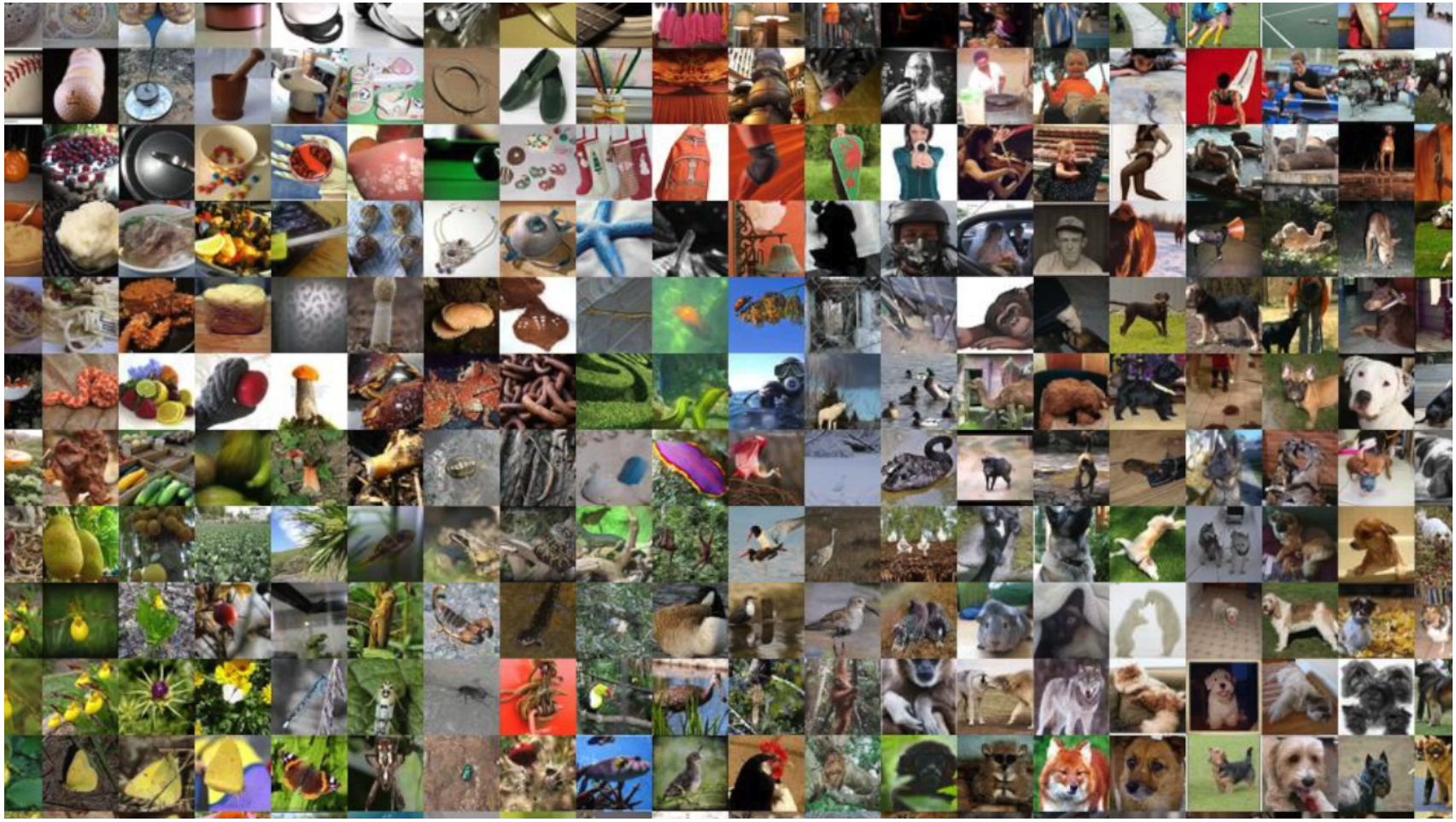
Geoffrey Hinton

breakthrough
in computer vision

IMAGENET

IMAGENET

WordNet hierarchy: 21841 classes
14,197,122 images, 1,034,908 annotated objects
crowd source annotations



IMAGENET

Object recognition

Method	Team	Year	Error
Hand crafted	University of Tokyo	2012	0.2617
AlexNet	University of Toronto	2012	0.1531
Multiple neural nets	Clarify	2013	0.1120
GoogLeNet	**Google**	**2014**	**0.0666**

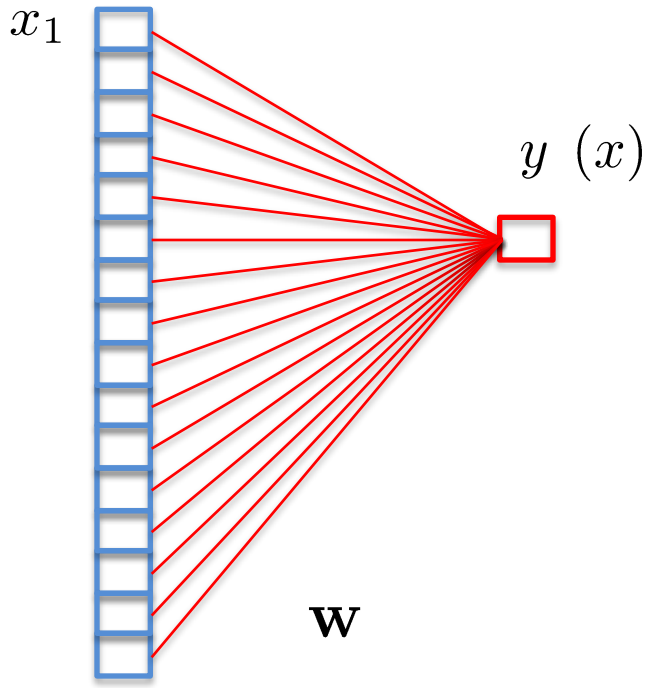
Human performance?

deep learning 6.7%, human 5.1% error

	Deep learning correct	Deep learning wrong
Human correct	1352/1500 	72/1500 • Objects very small or thin • Abstract representations • Image filters
Human wrong	46/1500 • Fine-grained recognition • Class unawareness • Insufficient training data	30/1500 • Multiple objects • Incorrect annotations

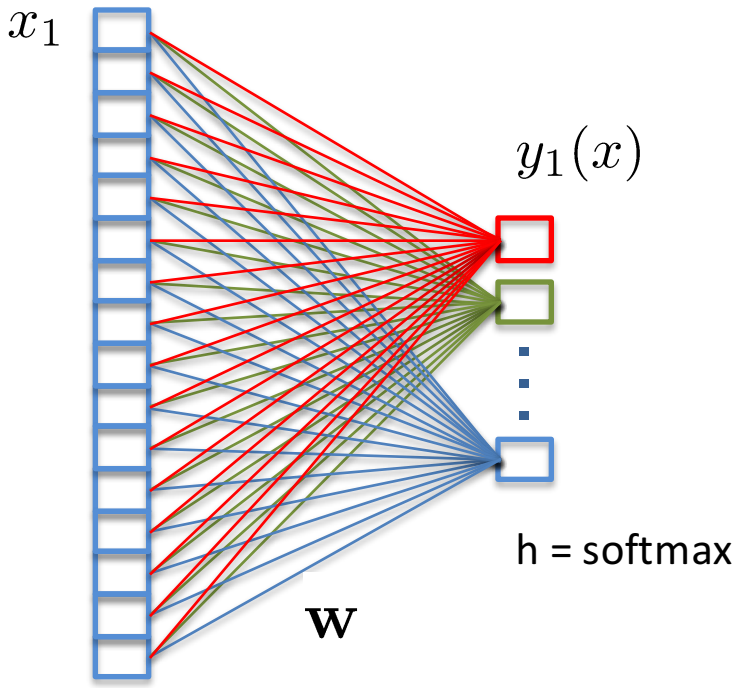
Technology overview

Perceptron

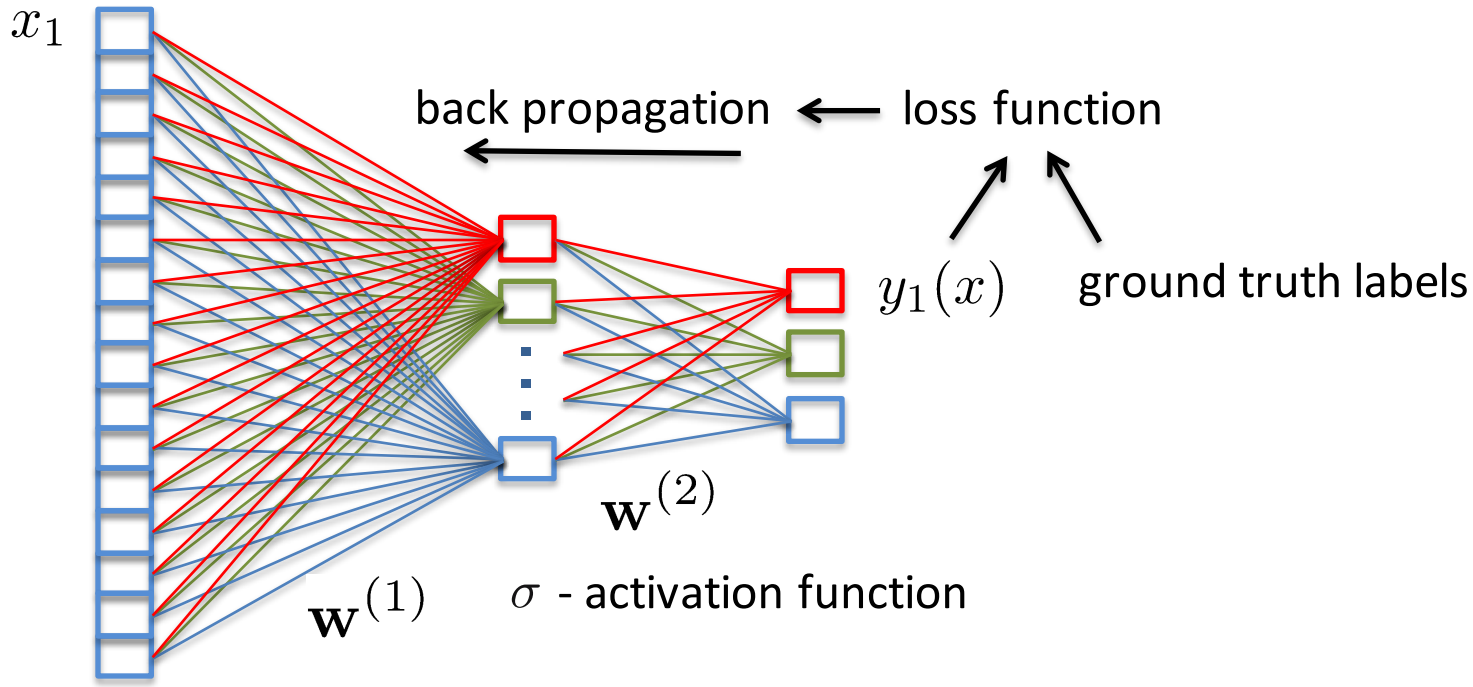


input layer, i output layer

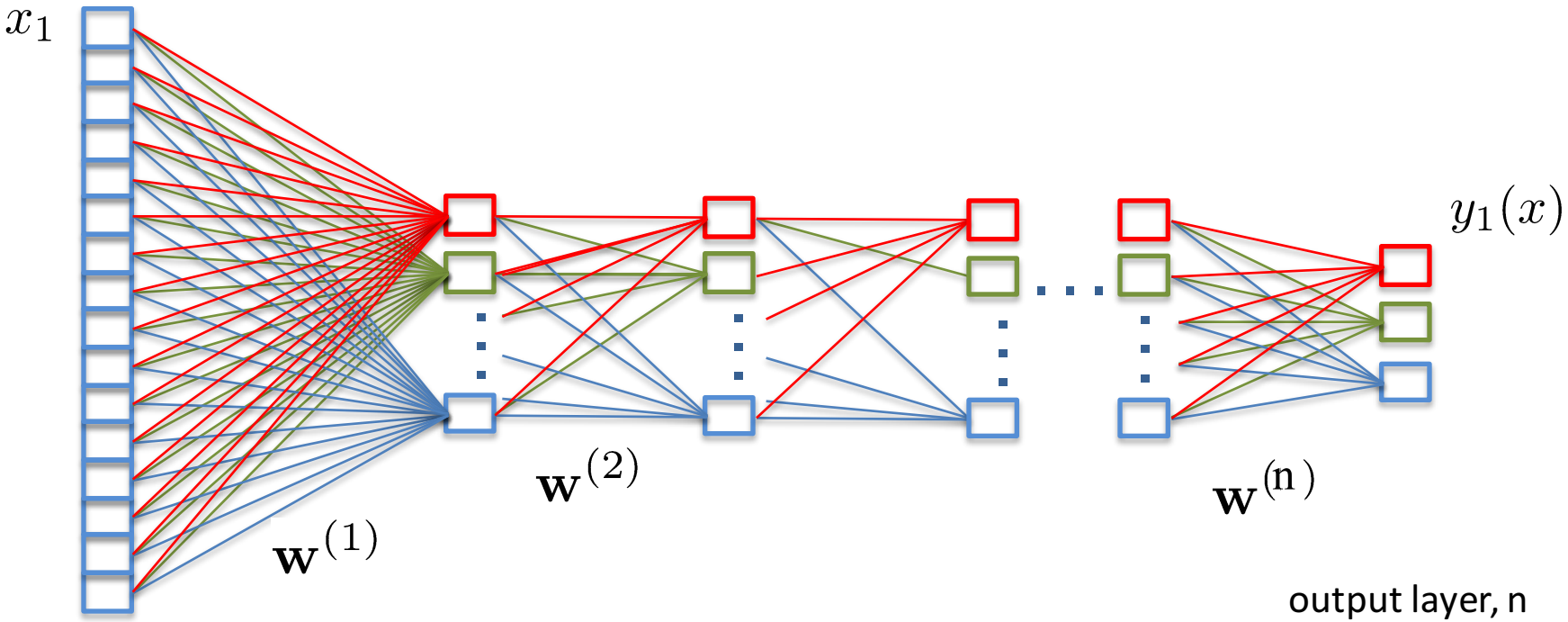
Logistic Regression



Multi-Layer Perceptron (MLP)



Deep Neural Network (DNN)



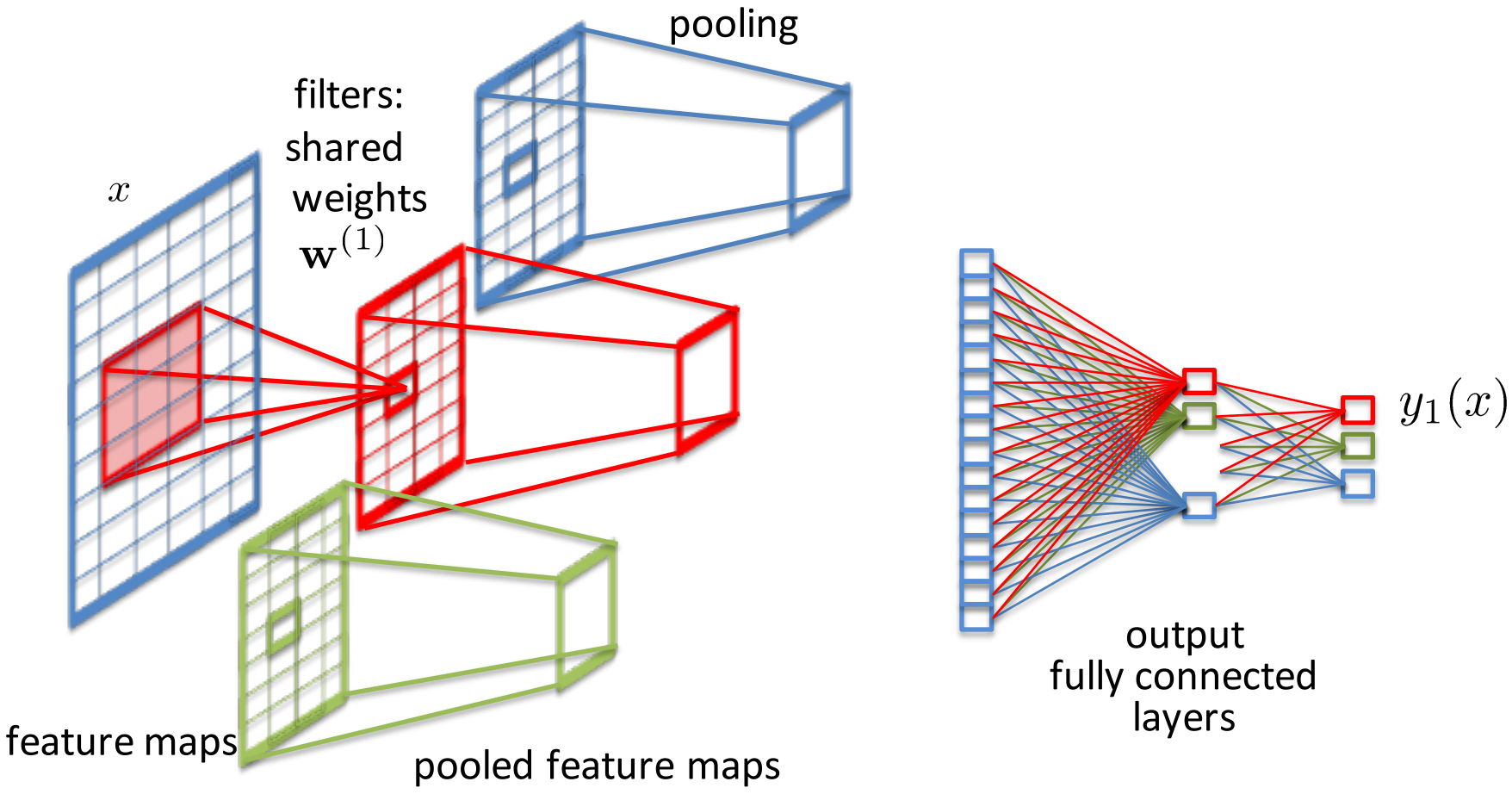
input layer, i hidden layer, j hidden layer, k

What is wrong with fully connected deep neural networks?

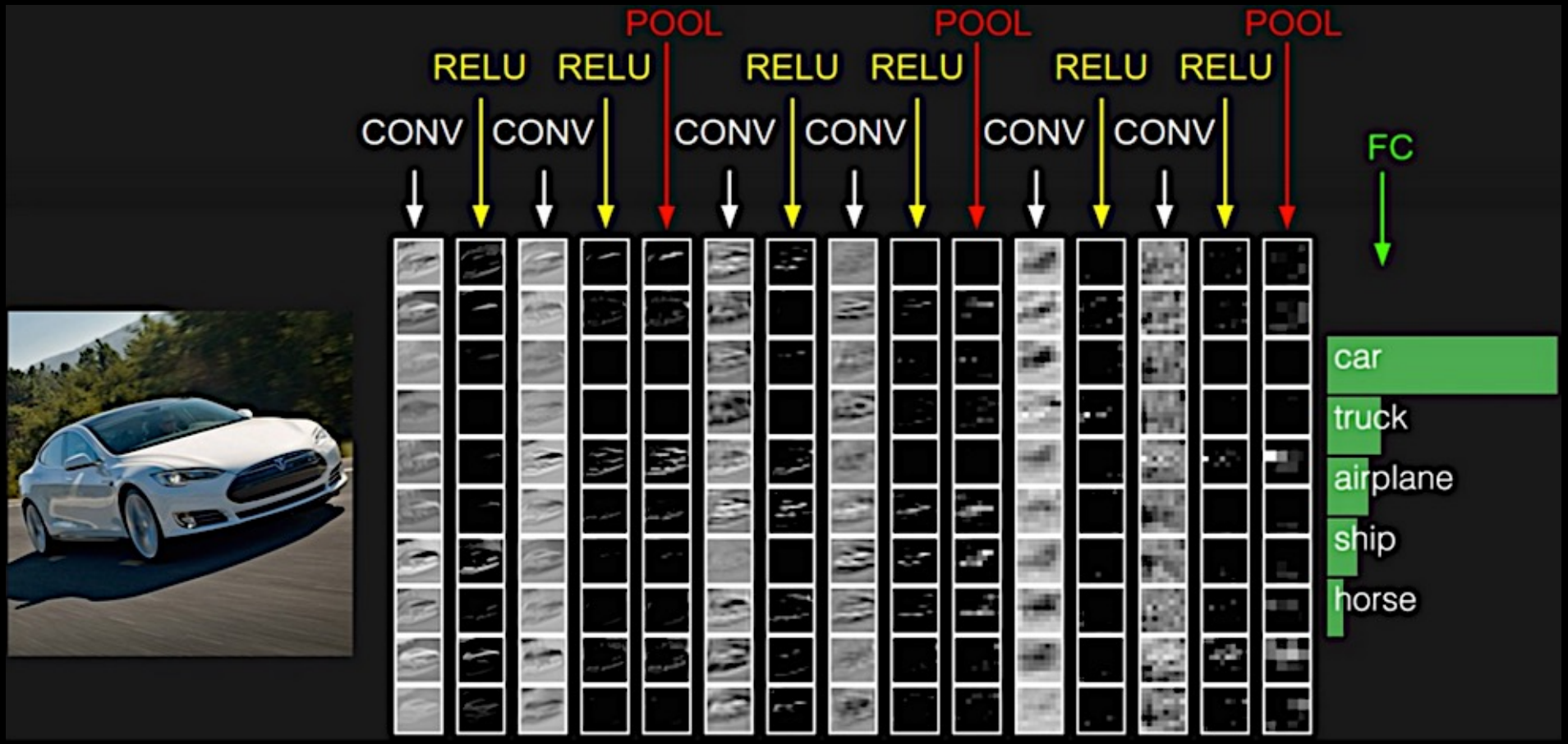
Too many parameters:
MLP[1000-1000-1000] - 2 000 000 parameters
require too much training data

Wasteful and do not generalize:
no spatial invariance for images

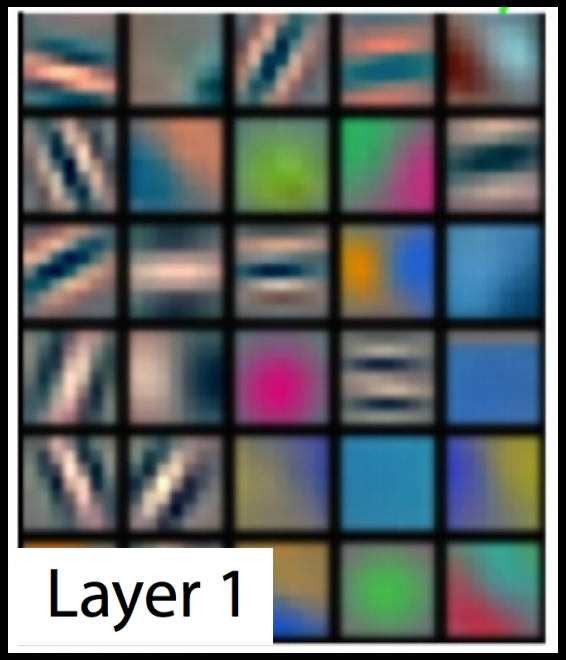
Convolutional Neural Network (CNN, Convnet)



What do we learn?

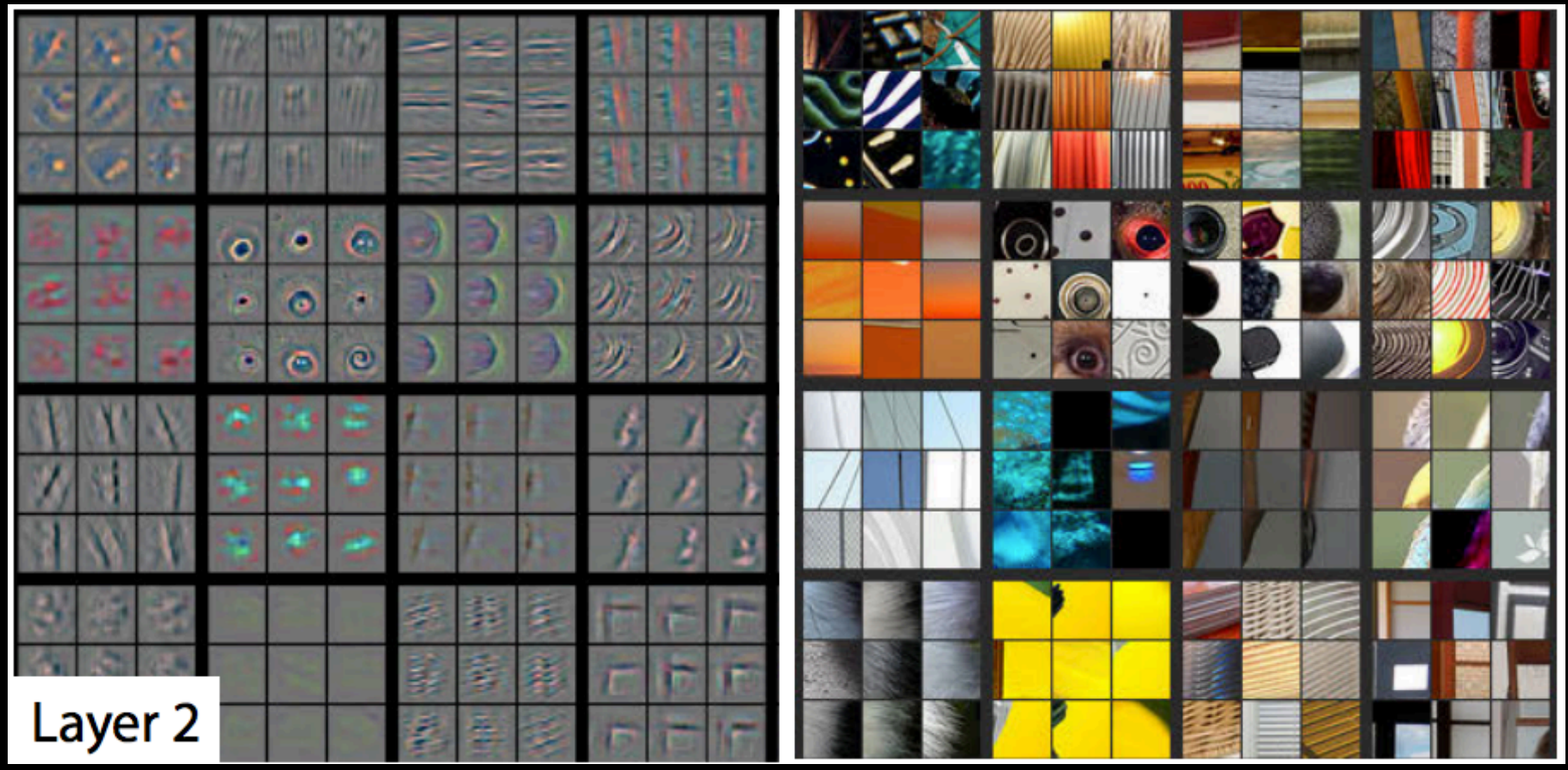


What do we learn?



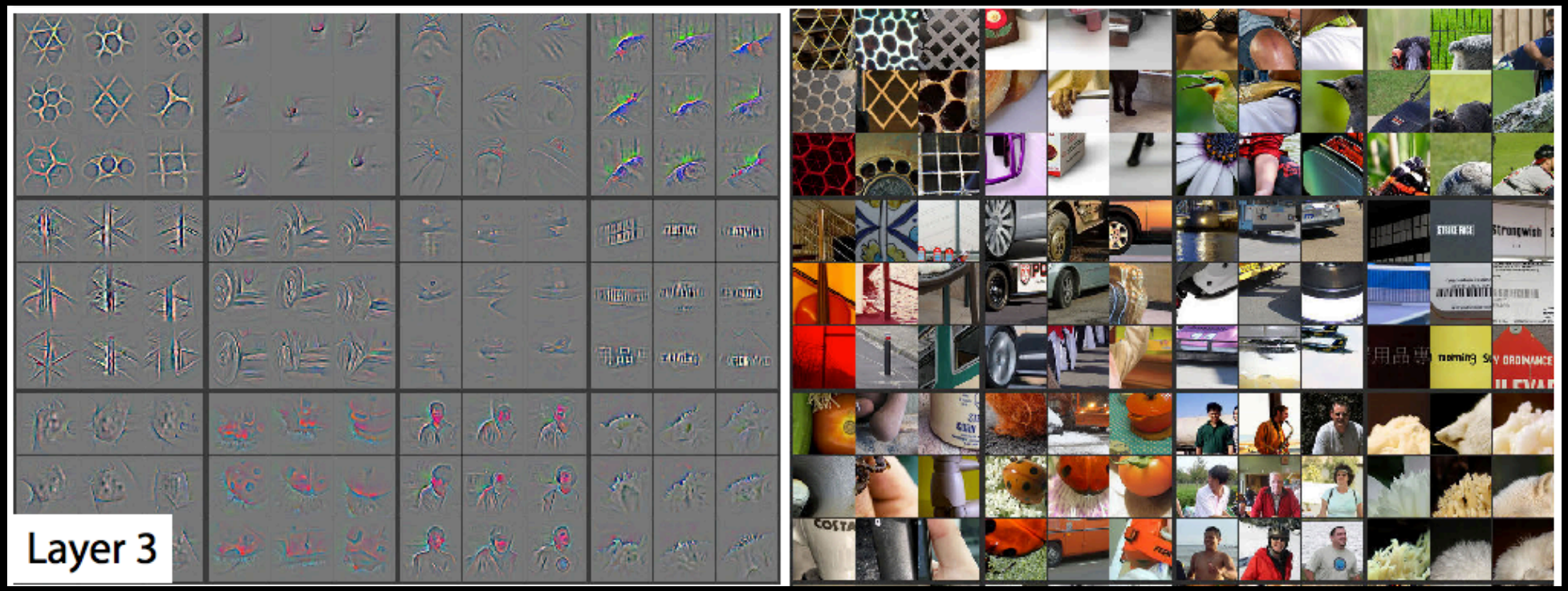
[Zeiler, Fergus, 2013]

What do we learn?



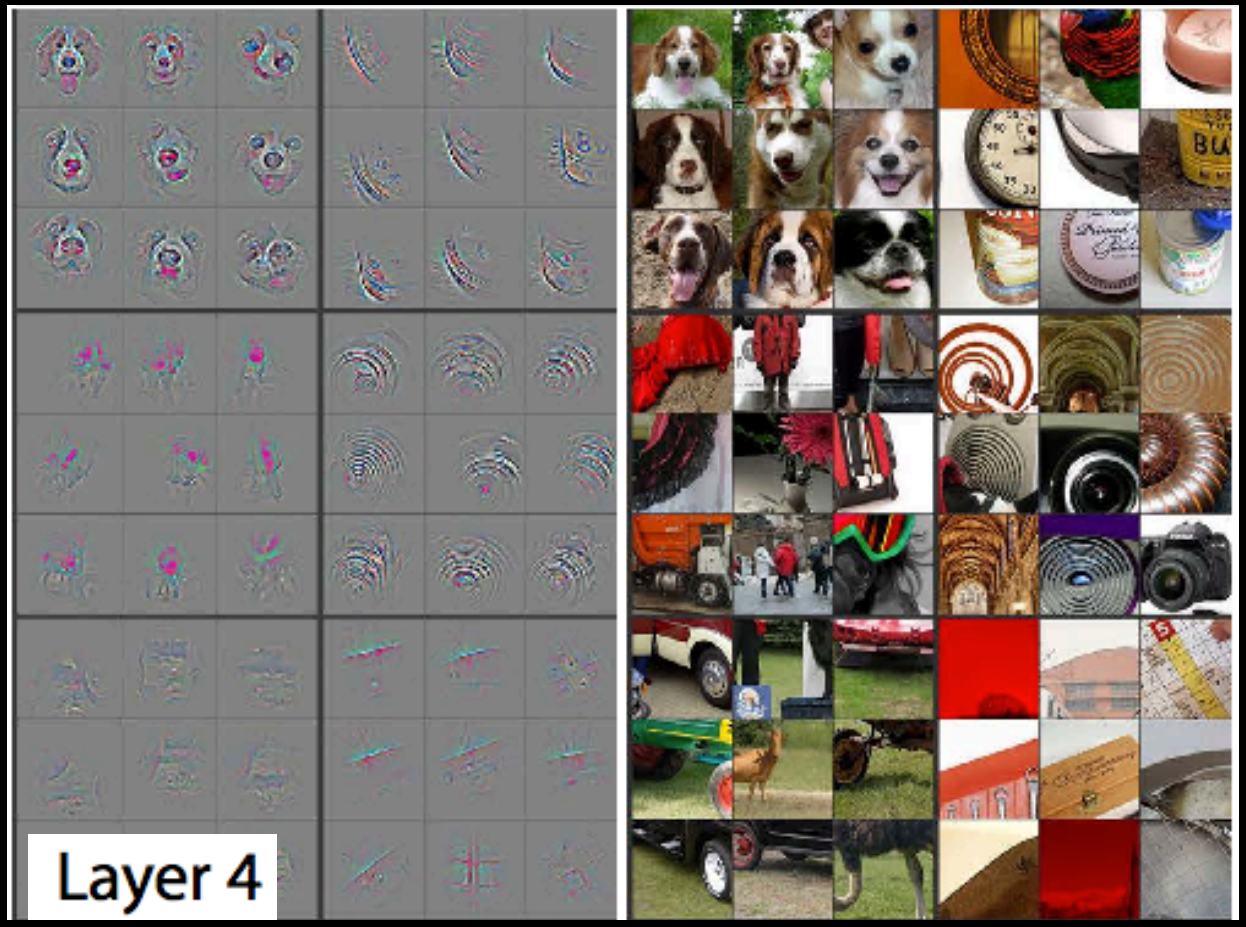
[Zeiler, Fergus, 2013]

What do we learn?



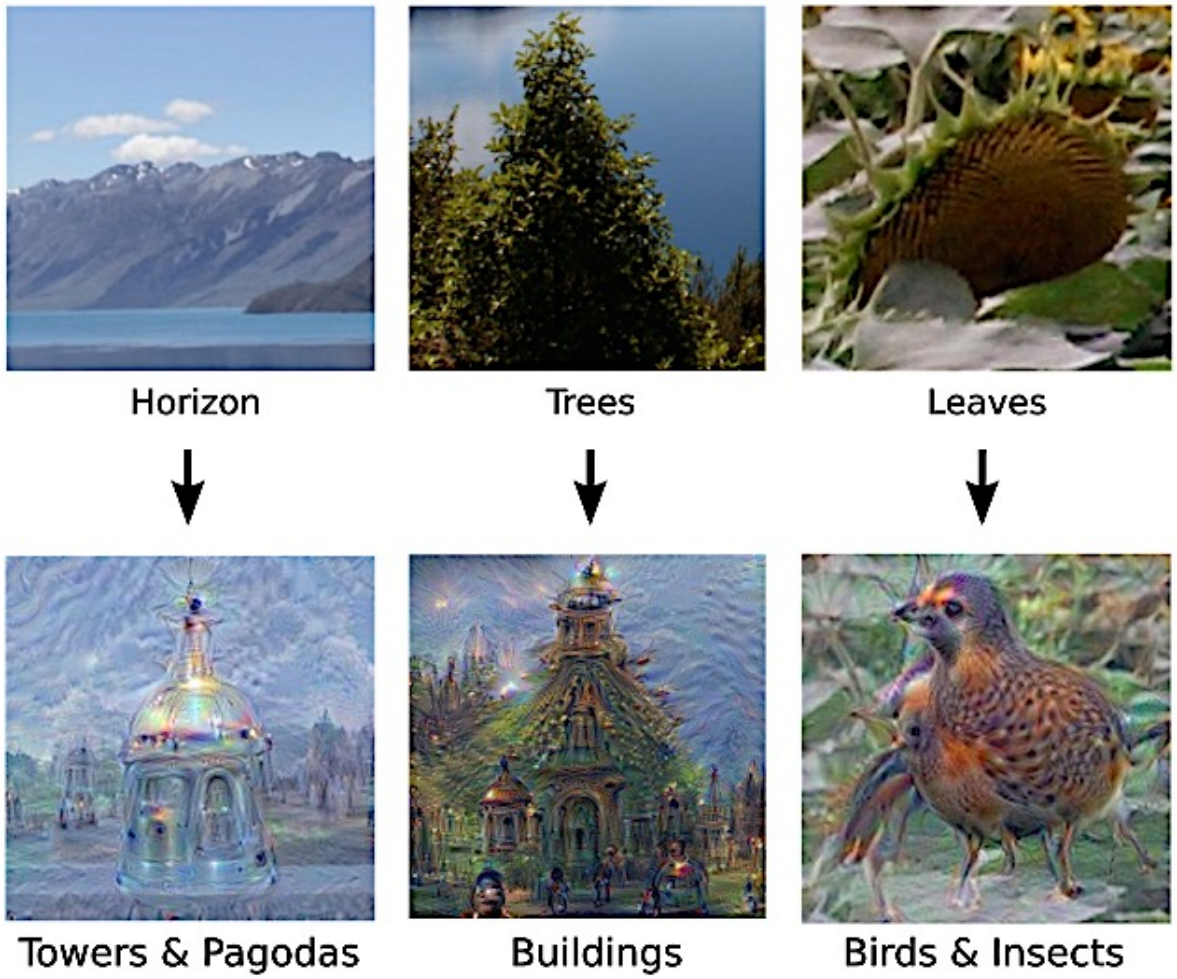
[Zeiler, Fergus, 2013]

What do we learn?

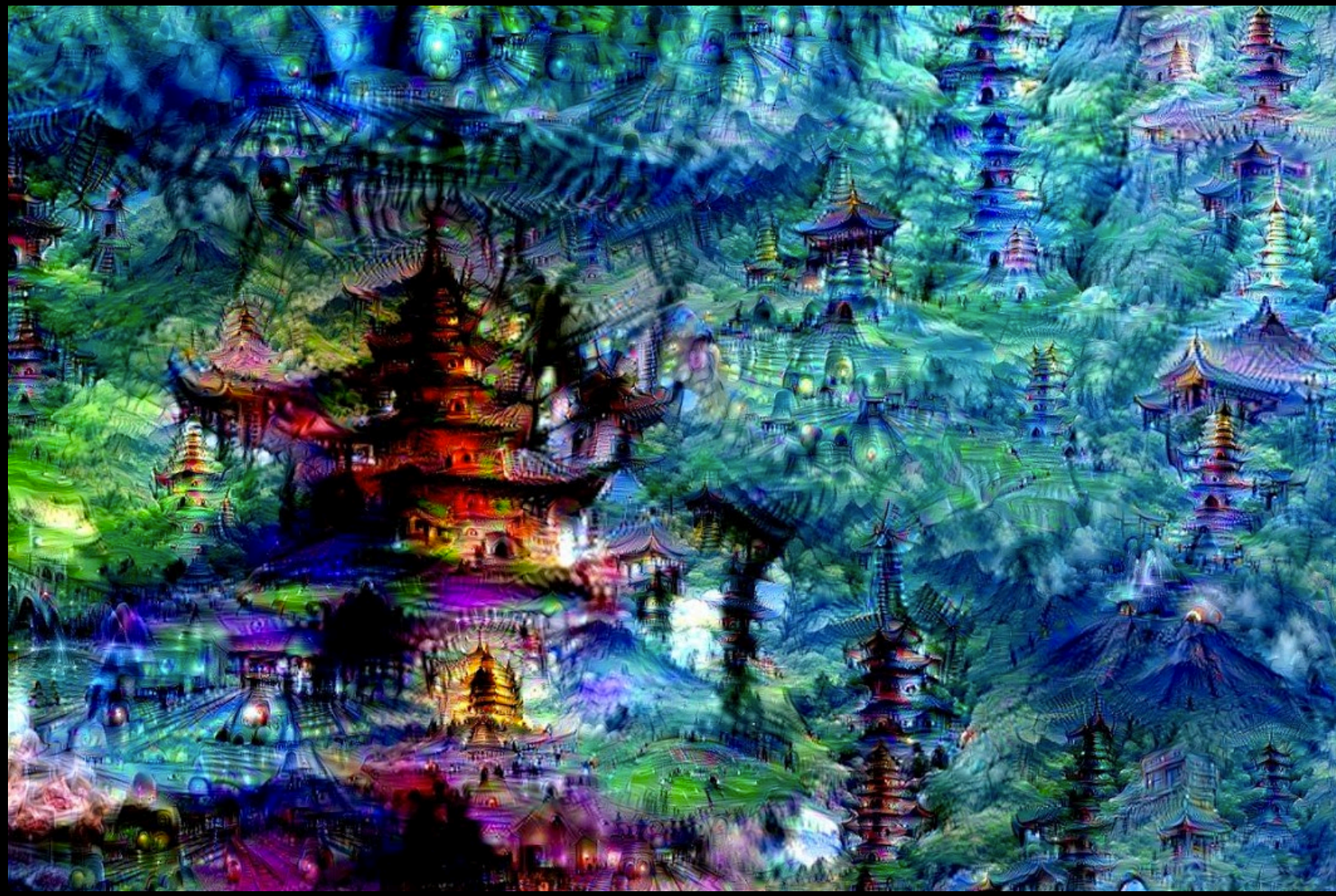


[Zeiler, Fergus, 2013]

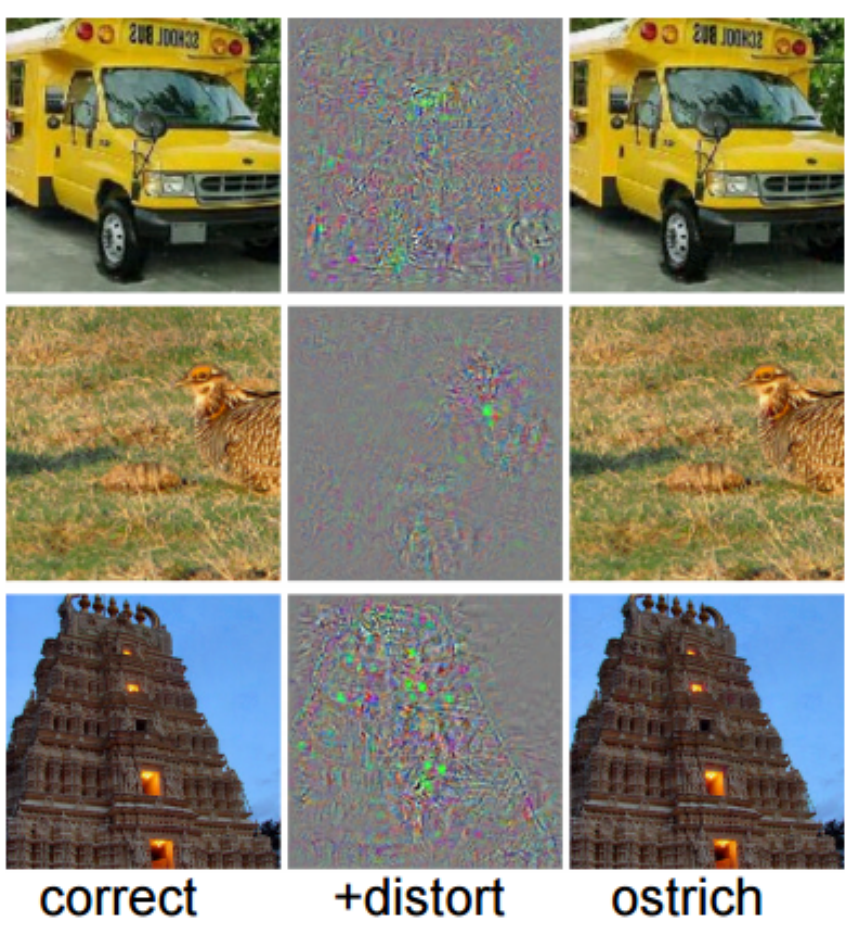
“Deep dreaming”

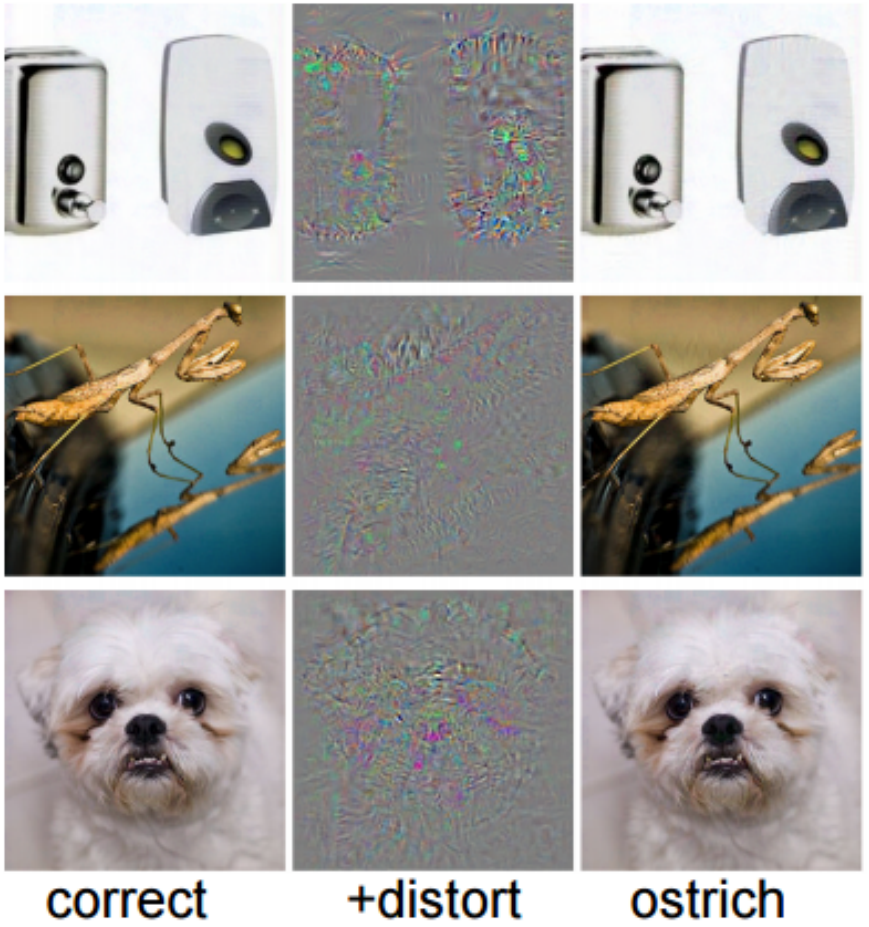


[http://googleresearch.blogspot.fr/]



Intriguing properties of ConvNets





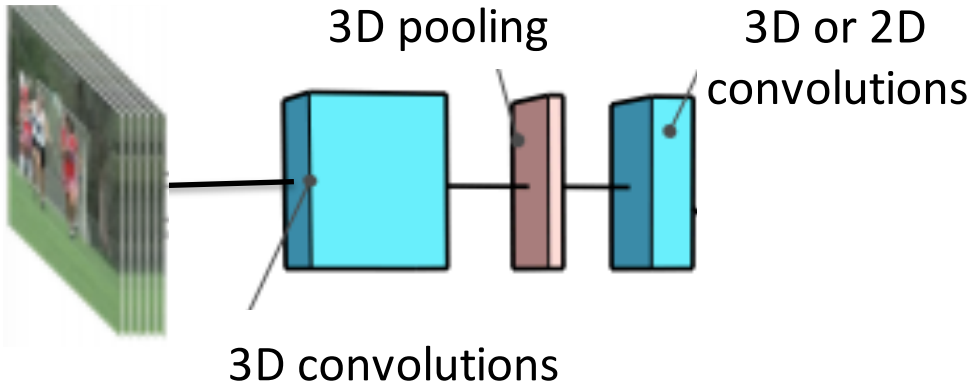
[Szegedy et al, 2013]

Practical issues with babysitting neural networks

- Data augmentation
- Data normalization
- Architecture optimization
- Loss function
- Weight initialization
- Learning rate and its decay schedule
- Overfitting: regularization, early stopping
- Momentum

What about sequential data?

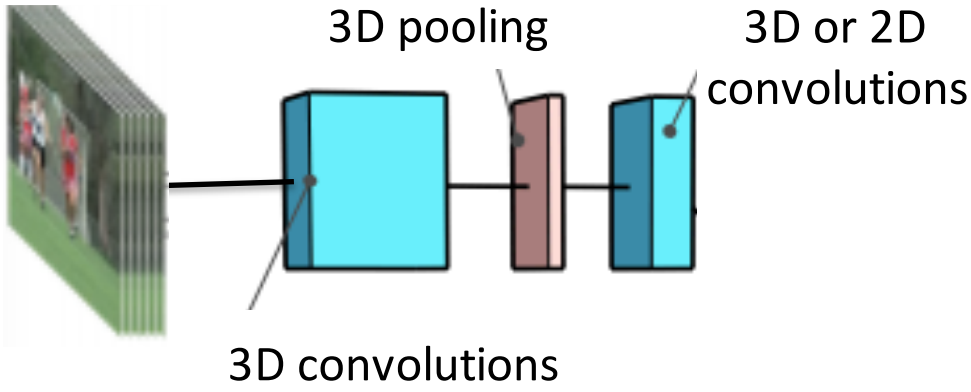
Video: spatio-temporal blocks (fixed length, simple patterns)



What about sequential data?

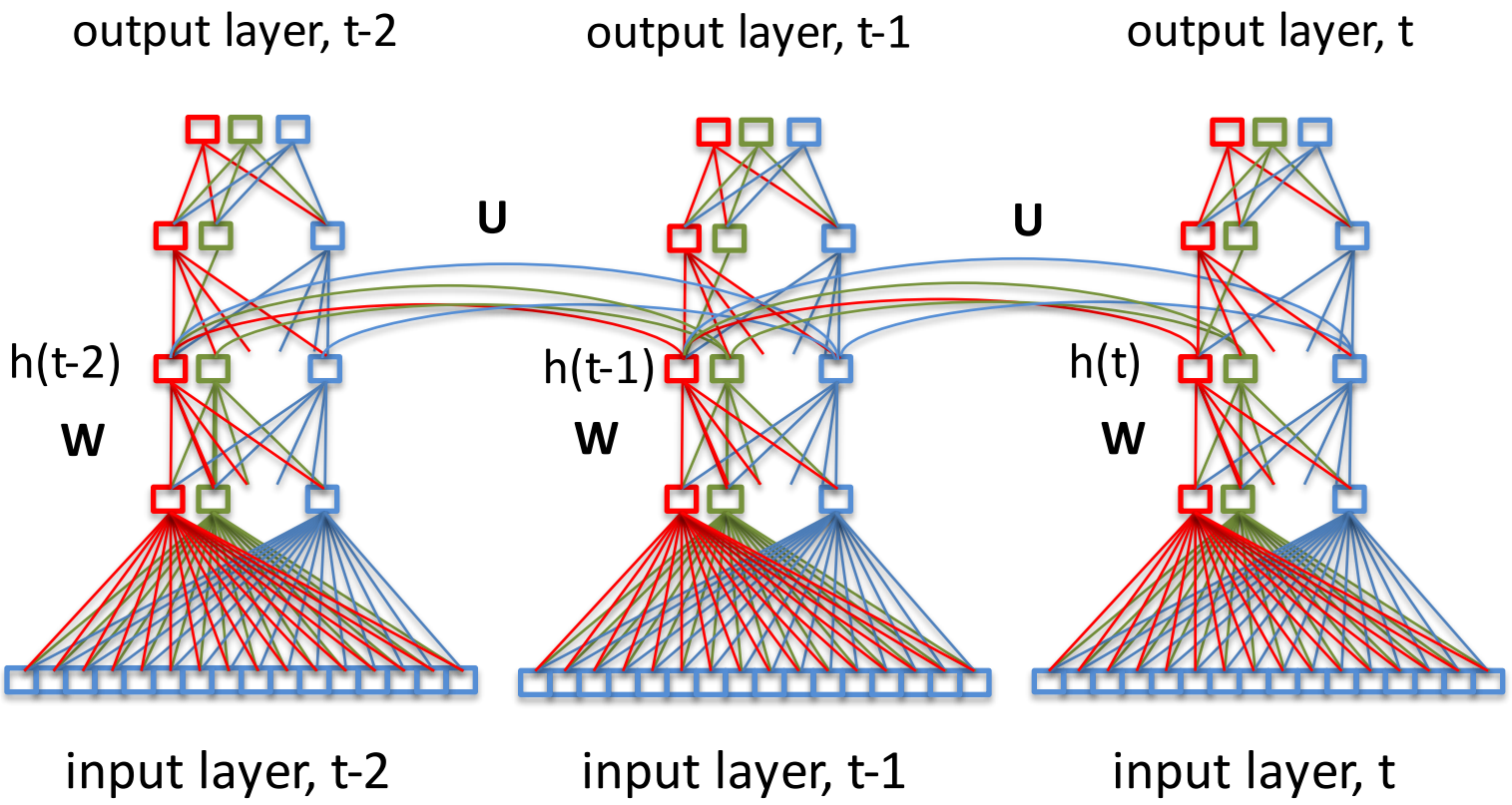
What about sequential data?

Video: spatio-temporal blocks (fixed length, simple patterns)



Alternatively: temporal modeling
of short- and long-term dependencies

Recurrent Neural Networks (RNN)



Leo Tolstoy's "War and Peace"

Iteration 100:

```
tyntd-iafhatawiaoihrdemot  lytdws  e ,tfti, astai f ogoh eoase rrranbyne 'nhthnee e
plia tklrgd t o idoe ns,smtt   h ne etie h,hregtrs nigtike,aoaenns lng
```

Iteration 500:

```
we counter. He stutn co des. His stanted out one ofler that concossions and was
to gearang reay Jotrets and with fre colt otf paitt thin wall. Which das stimn
```

Iteration 2000:

```
"Why do what that day," replied Natasha, and wishing to himself the fact the
princess, Princess Mary was easier, fed in had oftened him.
Pierre aking his soul came to the packs and drove up his father-in-law women.
```

[http://karpathy.github.io/2015/05/21/rnn-effectiveness/]

To start with…

Université de Montréal

Theano: flexible academic Python-based framework

http://deeplearning.net/

```
# convolve input feature maps with filters
conv_out = conv.conv2d(
    input=input,
    filters=self.W,
    filter_shape=filter_shape,
    image_shape=image_shape
)

# downsample each feature map individually, using maxpooling
pooled_out = downsample.max_pool_2d(
    input=conv_out,
    ds=poolsize,
    ignore_border=True
)
```

Facebook AI Research

Torch7: MATLAB-like environment (C++ / Lua)

http://torch.ch/

Supervised and unsupervised neural networks, optimization, graphical models, image processing

```
model = nn.Sequential()

model:add( nn.SpatialConvolution(3,16,5,5) )
model:add( nn.Tanh() )
model:add( nn.SpatialMaxPooling(2,2,2,2) )
model:add( nn.SpatialContrastiveNormalization(16, image.gaussian(3)) )
```

Berkeley
UNIVERSITY OF CALIFORNIA

Google

Caffe: C/C++ deep learning framework

http://caffe.berkeleyvision.org/

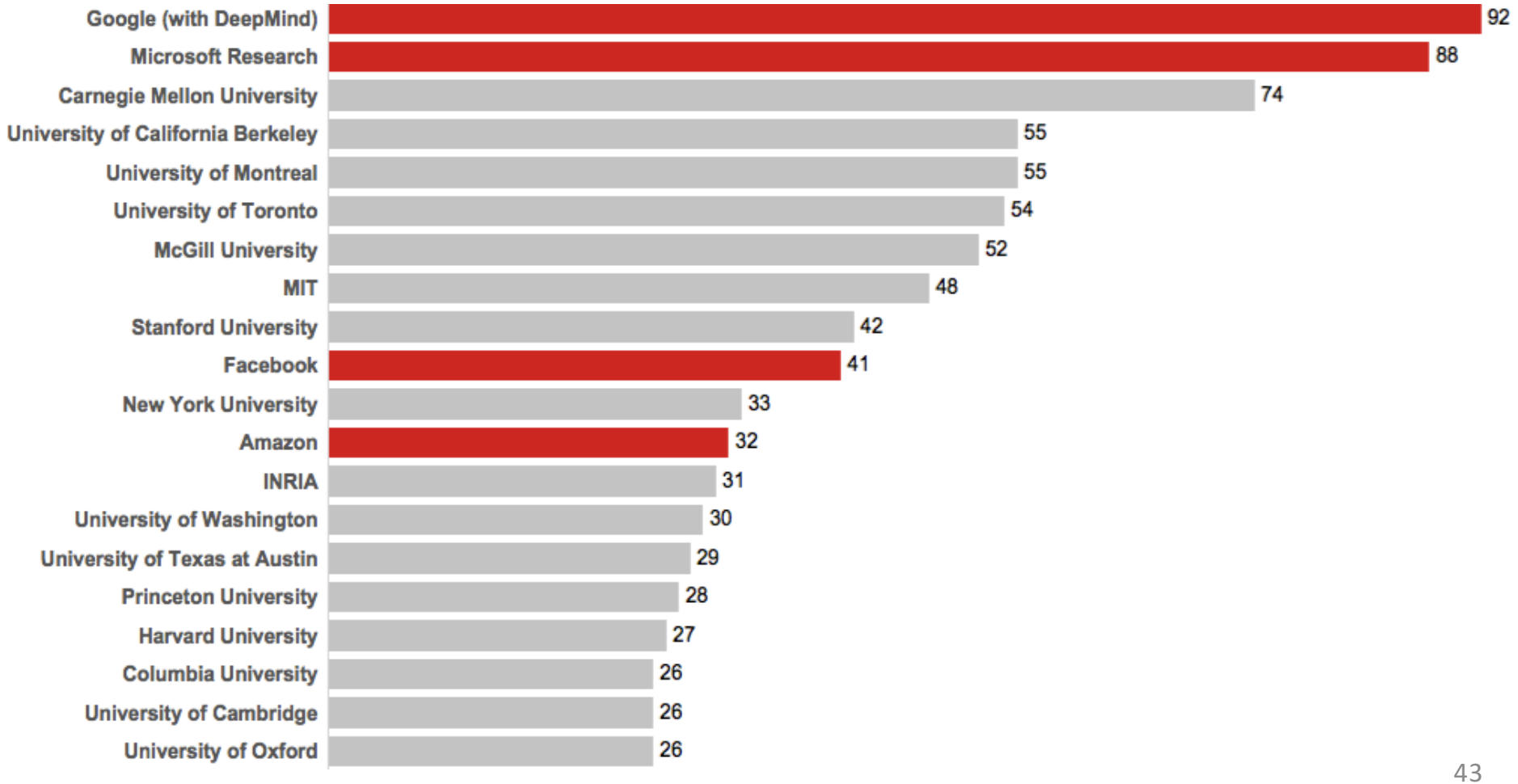
standard networks configurable without hard-coding,

fast and portable to mobile platforms,

large collection of pretrained models available

(widely used in industrial applications)

Deep learning: academia vs industry

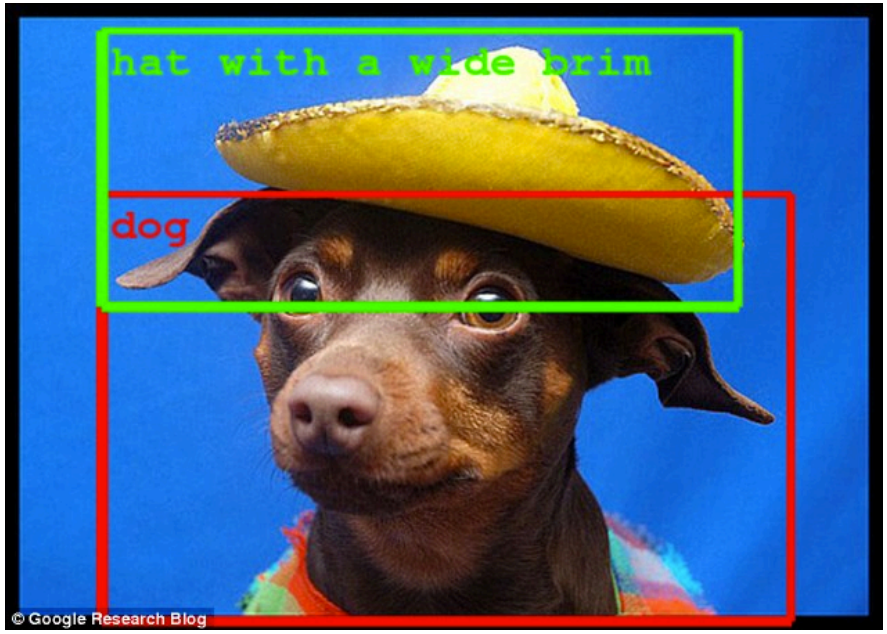


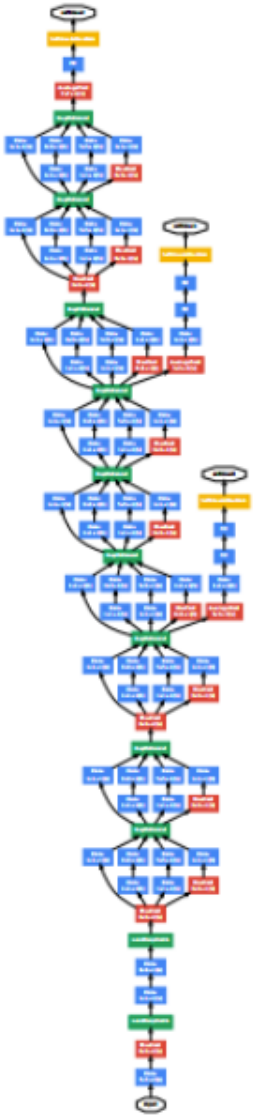
Applications

Google

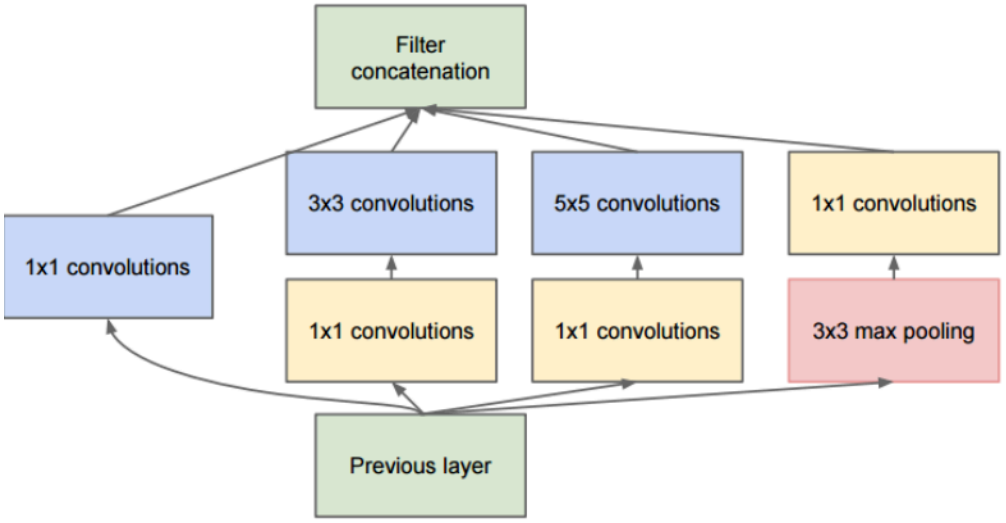
Object recognition and localization

Deep learning based visual search engine announced in 2013
Internal dataset with 100 000 000 labelled images and 18 000 classes





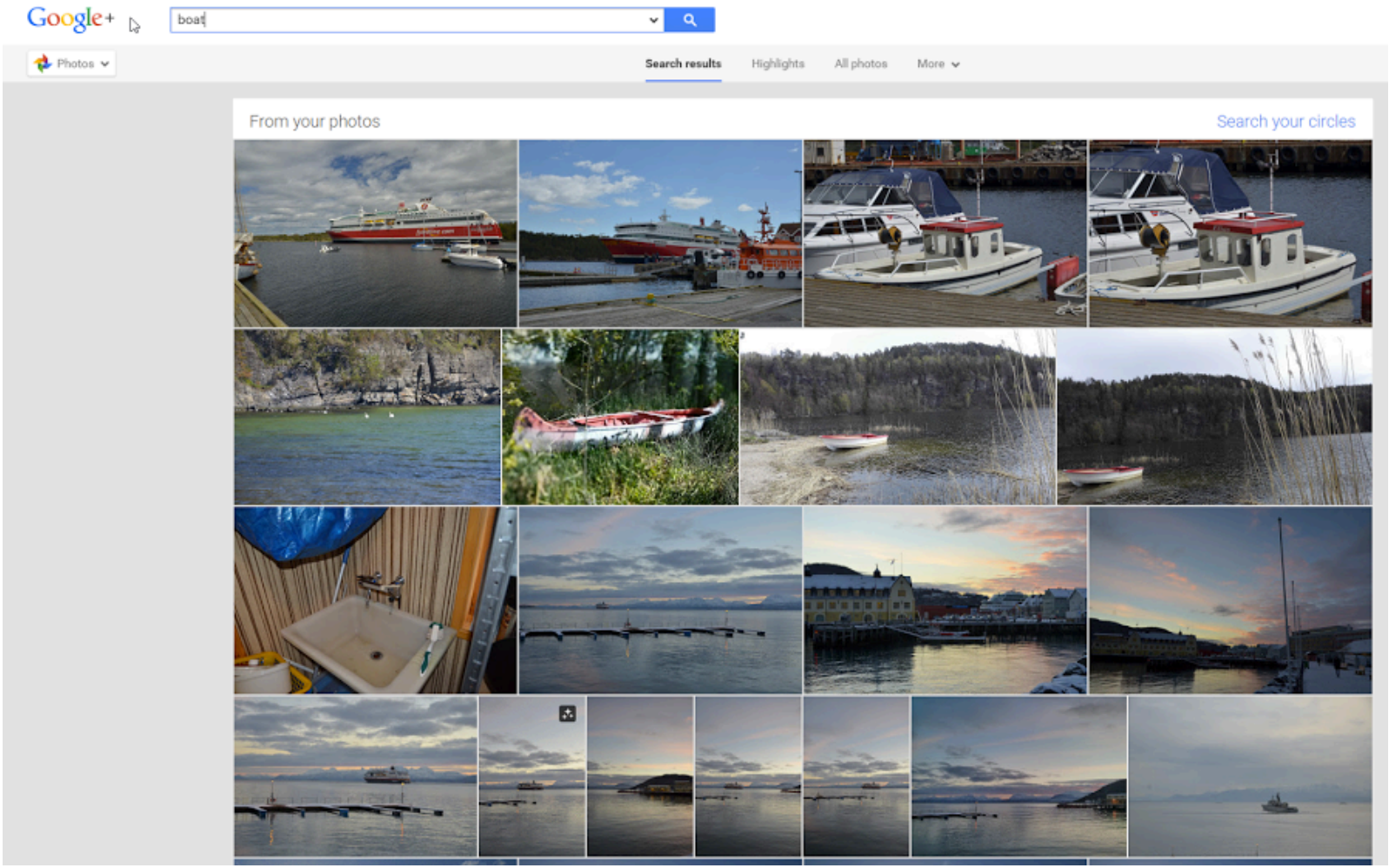
Inception model (GoogLeNet)



22 layers with parameters
27 layers including pooling
about 100 building blocks

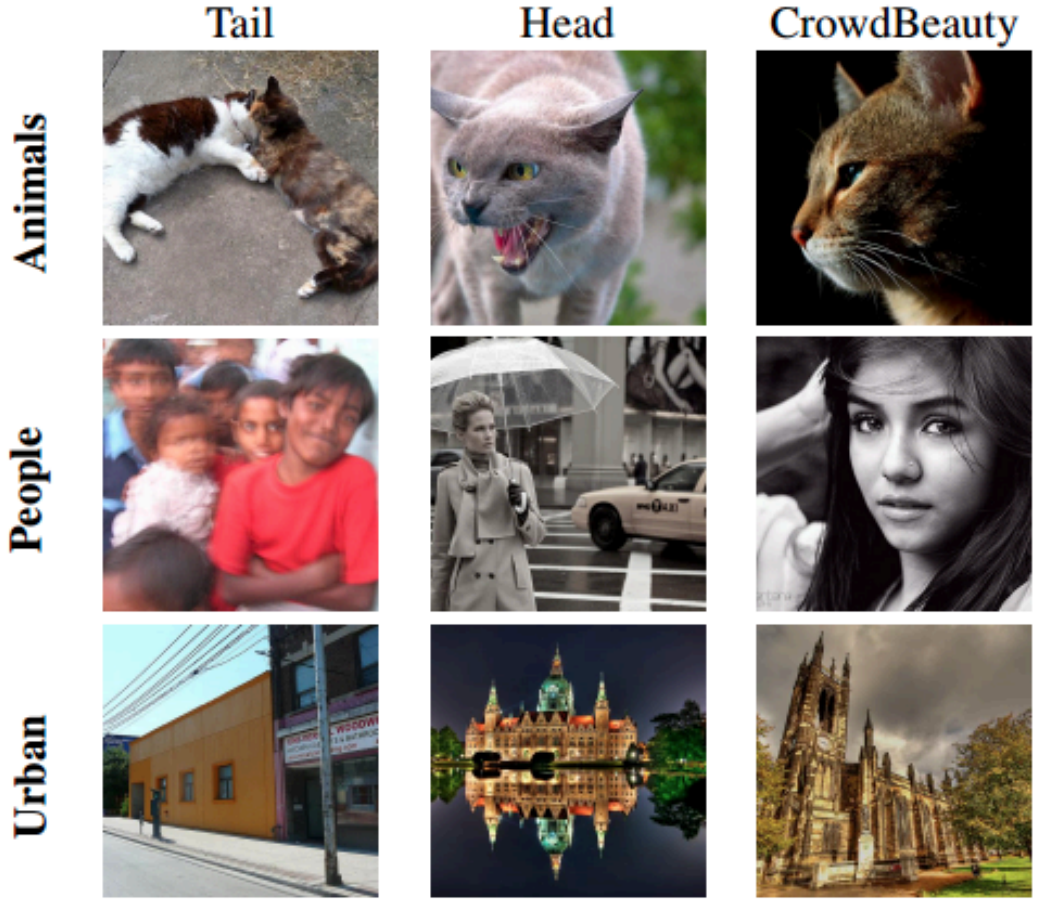
[Szegedy et el, CVPR 2015]

photos.google.com



flickr™ YAHOO!®

Computational aesthetics and video creativity



Facebook AI Research

DeepFace:
Closing the gap to human-level performance in face verification

Nicole Kidman Nicole Kidman

Jacquline Obradors Julie Taymor

97.35% accuracy on the Labeled Faces in the Wild dataset

(27% improvement over the state of the art)

[Taigman et al, CVPR 2014]

Microsoft Research

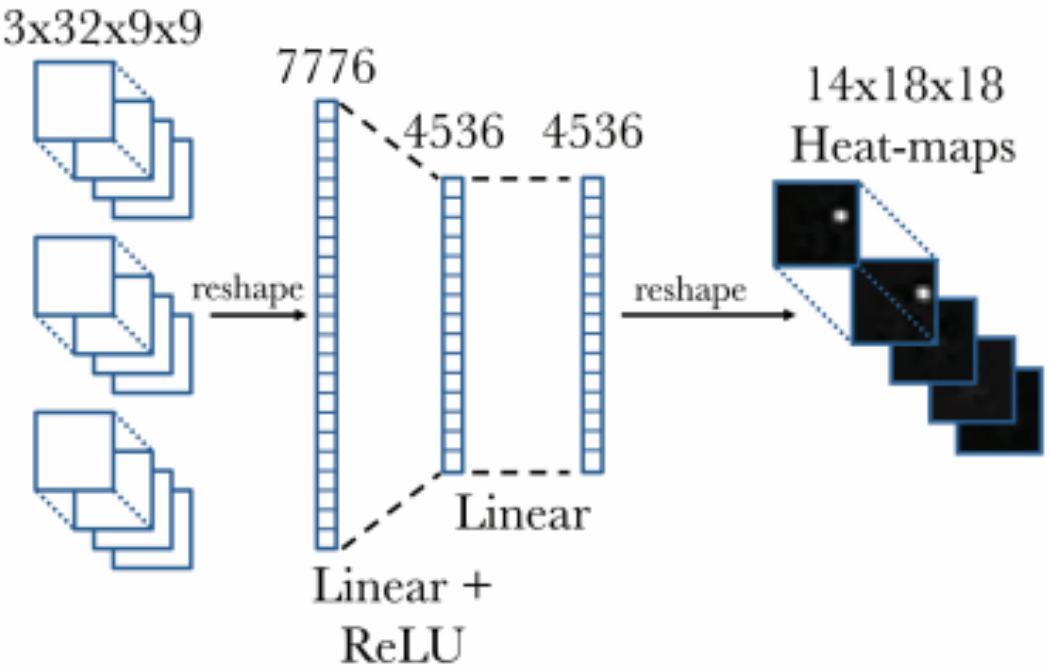
Age estimation

how-old.net

Automatic navigation: scene labeling

[Byeon et al, CVPR 2015]

Human-computer interaction: hand segmentation and pose estimation

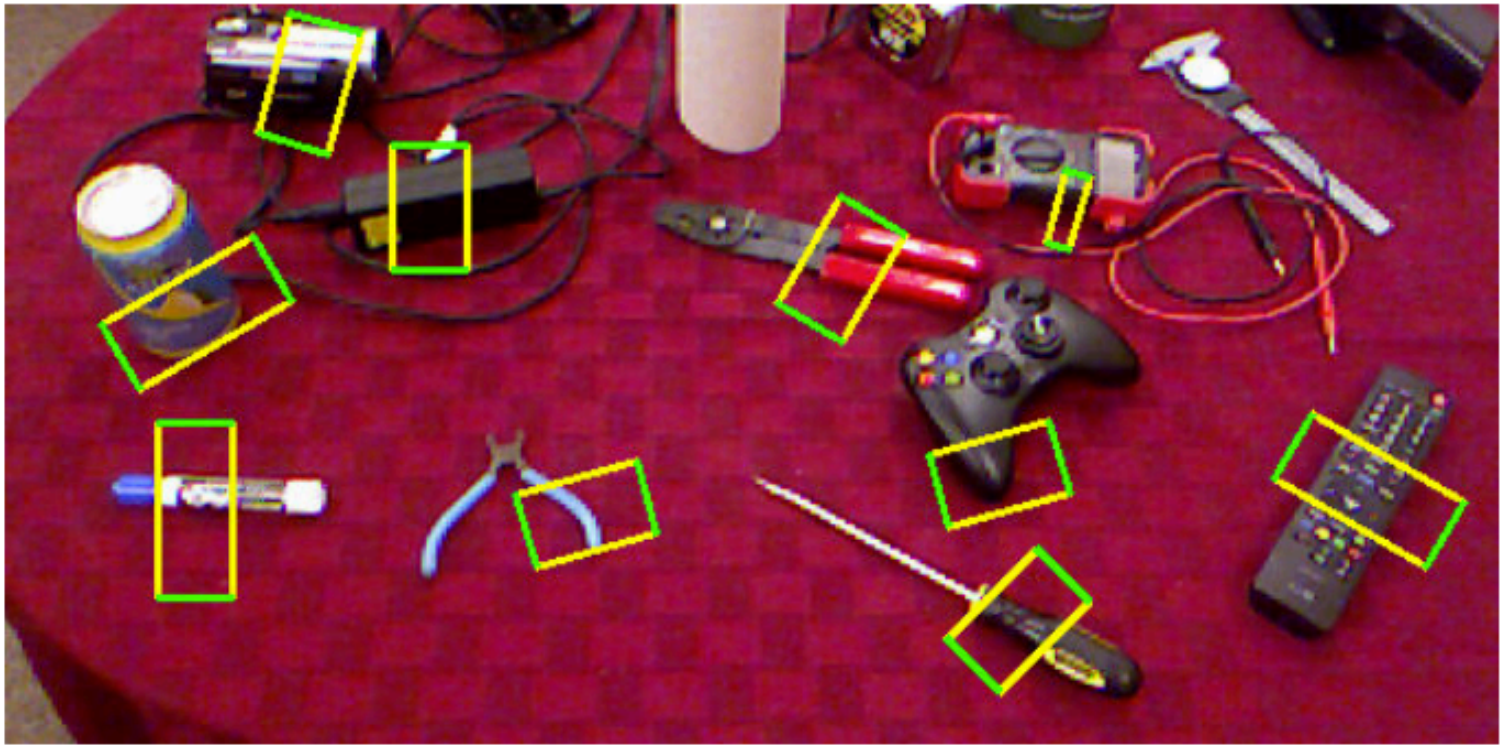


[Tompson et al, SIGGRAPH 2014]

Human pose estimation

[Tompson et al, CVPR 2015]

Deep learning for robotics: navigation, human-robot interaction, detecting grasps

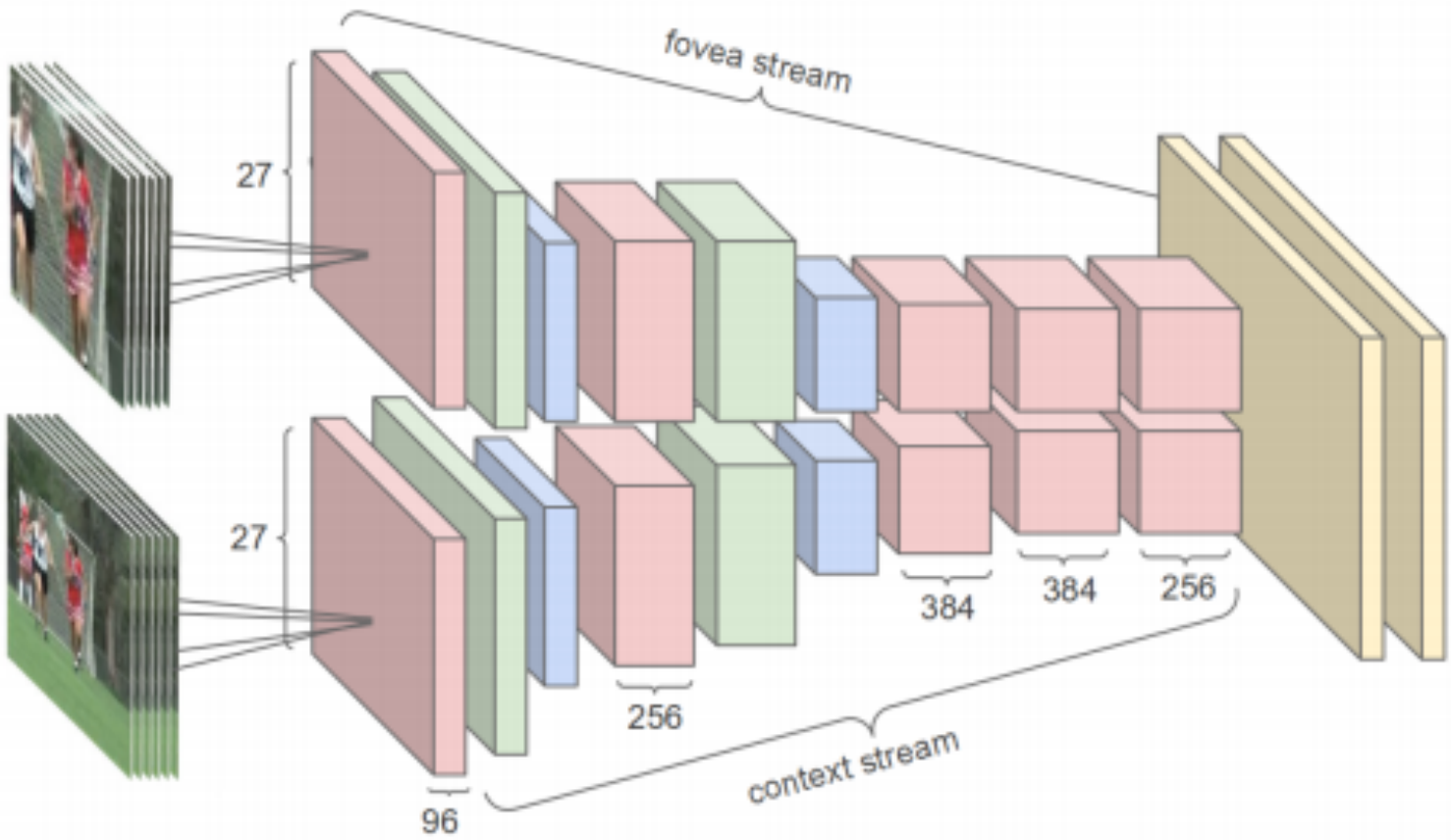


[Lenz et al, IJRR 2014]

Multi-modal deep learning: gesture recognition

[Neverova et al, PAMI 2015]

Video classification



[Karpathy et al, CVPR 2014]

Facebook AI Research

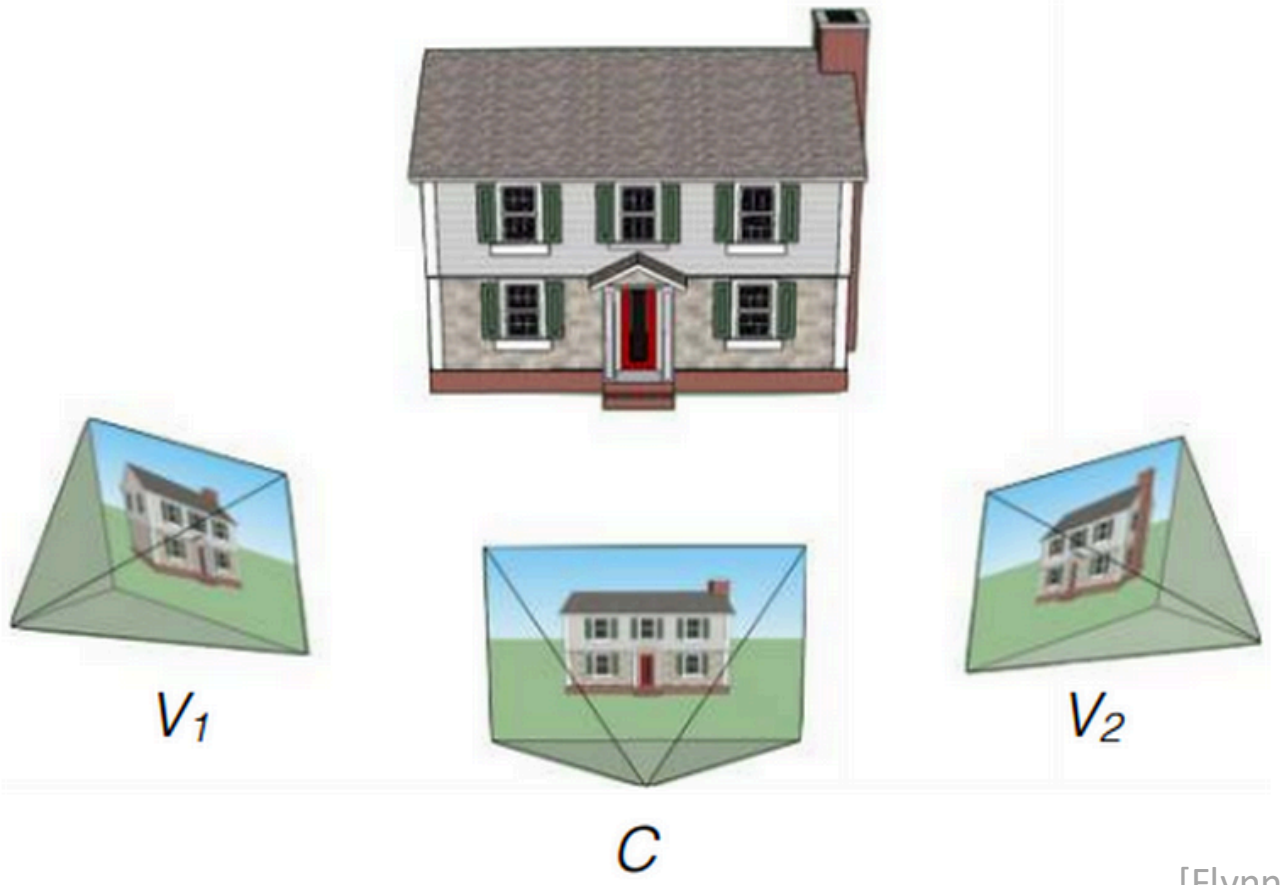
Action recognition from video

Natural image generation

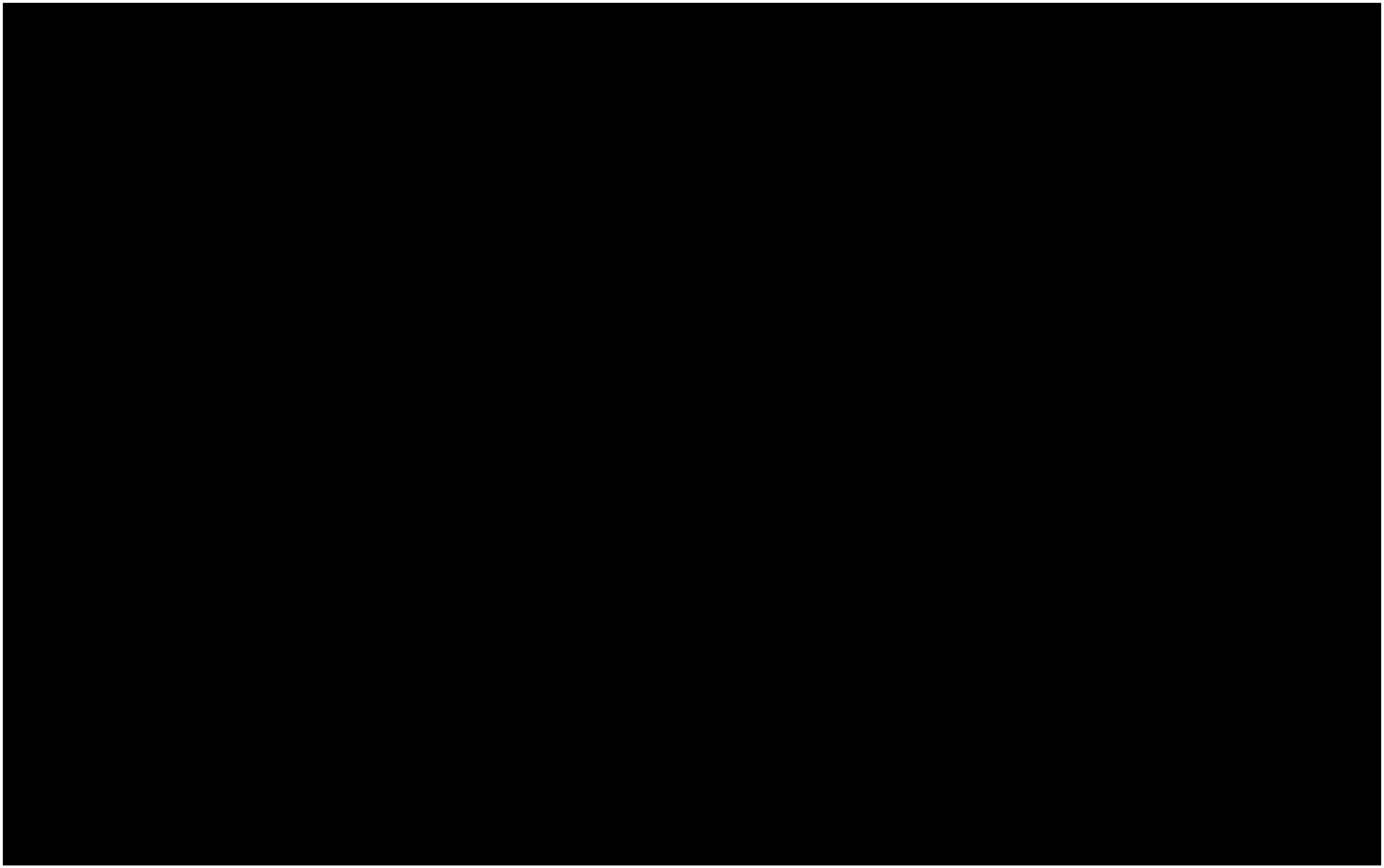
[Denton et al, 2015]

Google

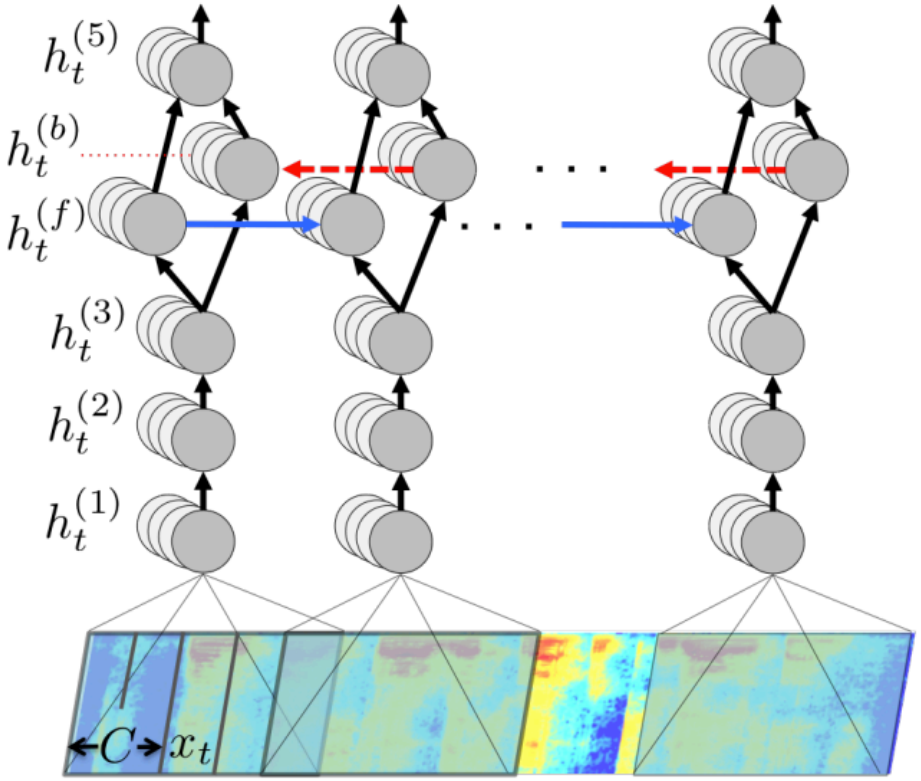
DeepStereo: View synthesis



[Flynn et al, 2015]



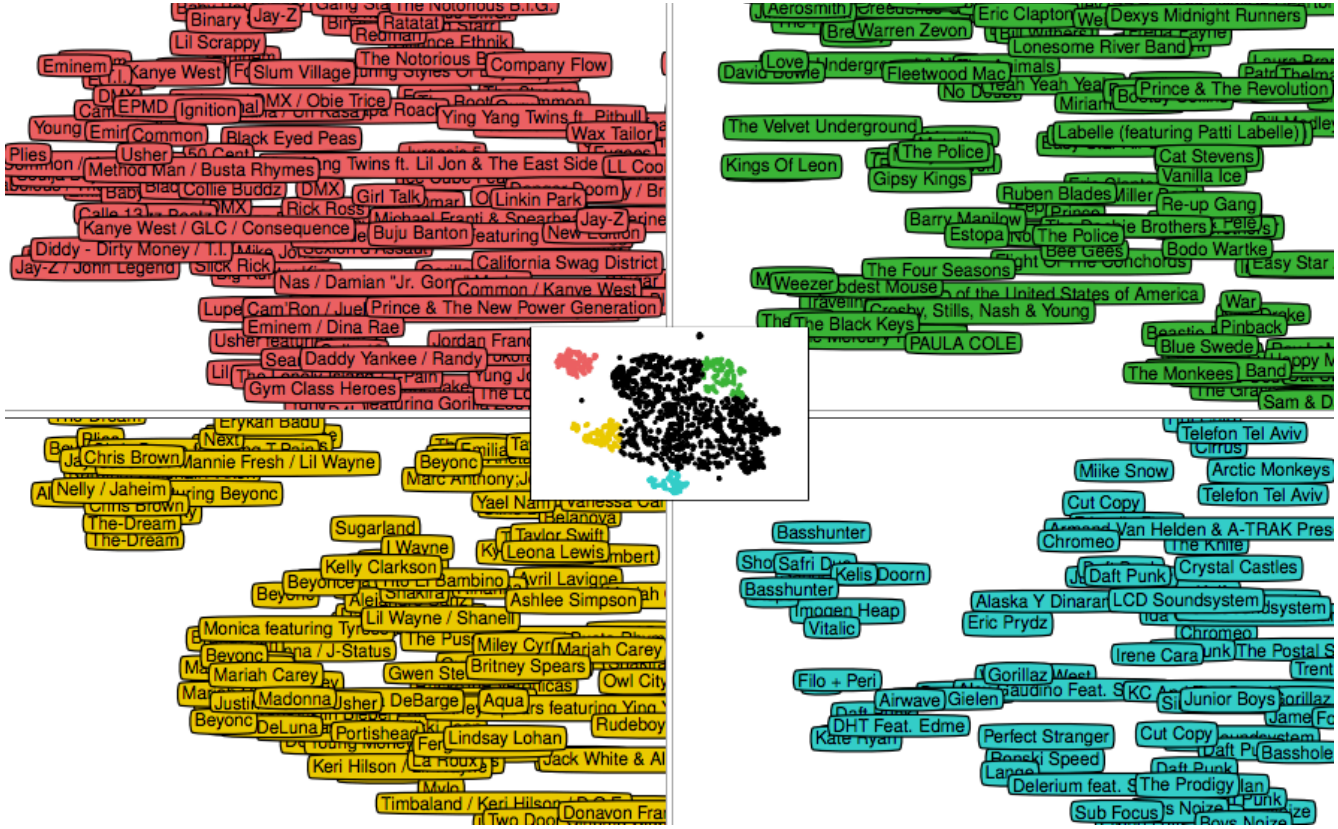
Deep Speech



System	Combined (176)
Apple Dictation	26.73
Bing Speech	22.05
Google API	16.72
wit.ai	19.41
Deep Speech	**11.85**

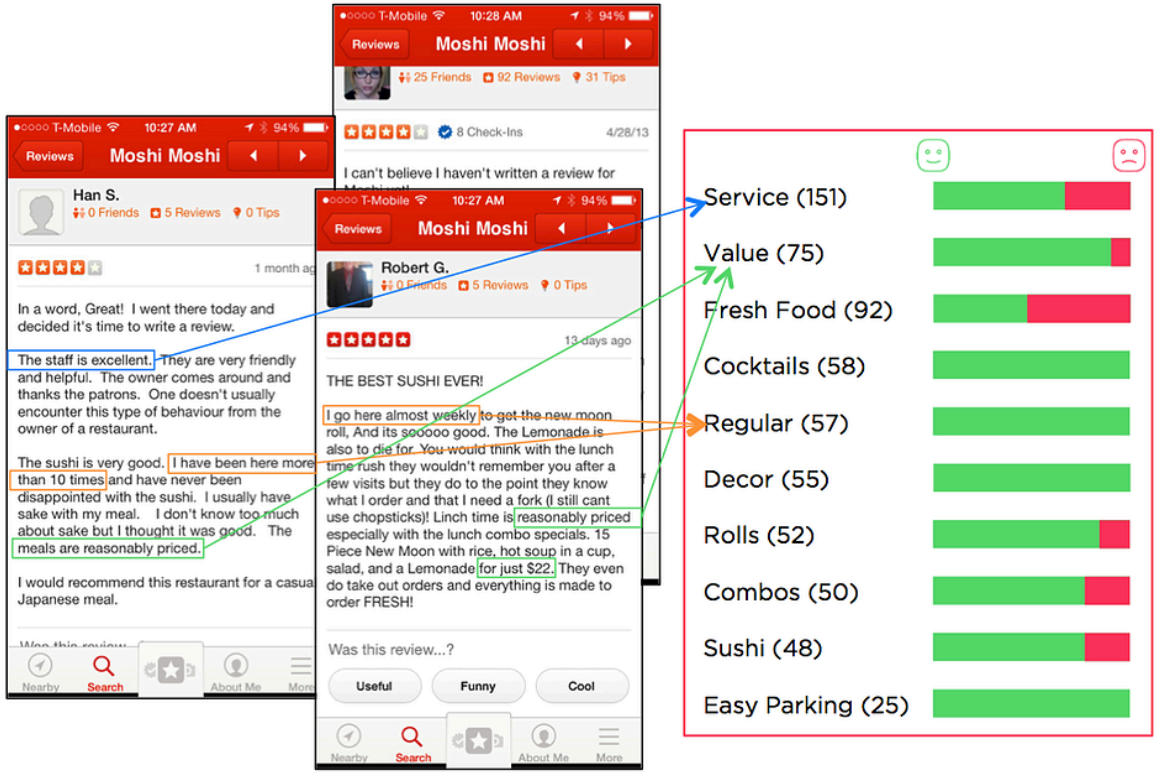
[Hannun et al, 2014]

Musical intelligence



[Oord, Dieleman, Schrauwen, NIPS 2013]

Text understanding



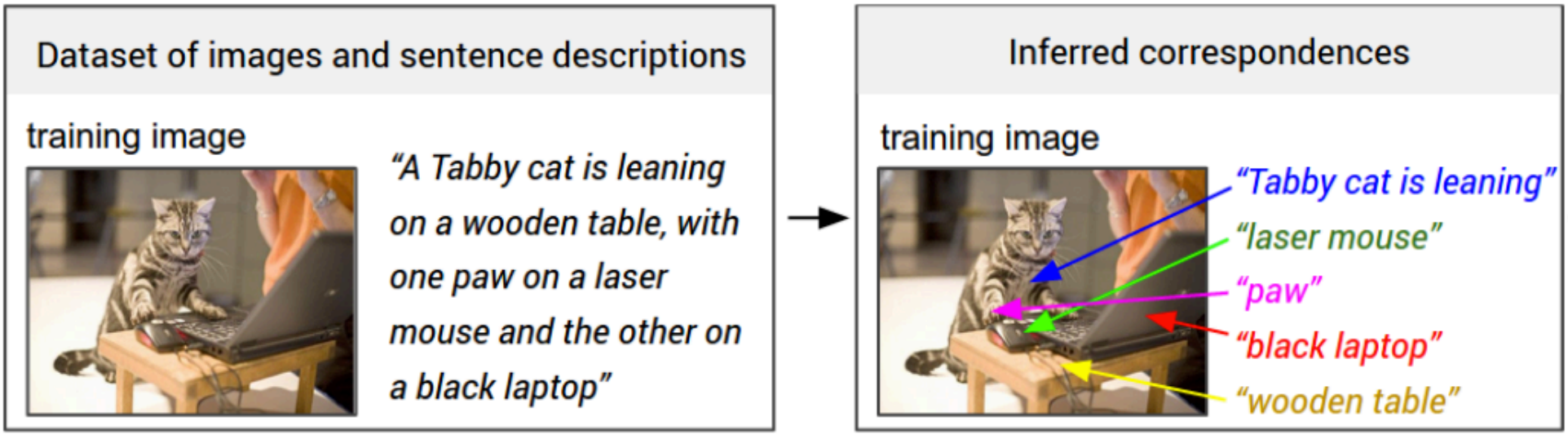
Facebook AI Research

Memory networks: question answering

Joe went to the kitchen. Fred went to the kitchen. Joe picked up the milk.
Joe travelled to the office. Joe left the milk. Joe went to the bathroom.
Where is the milk now? A: office
Where is Joe? A: bathroom
Where was Joe before the office? A: kitchen

[Weston et al, ICLR 2015]

Multimodal deep learning: image captioning



[Karpathy et al, CVPR 2015]

Image captioning

"man in black shirt is playing guitar."

"girl in pink dress is jumping in air."

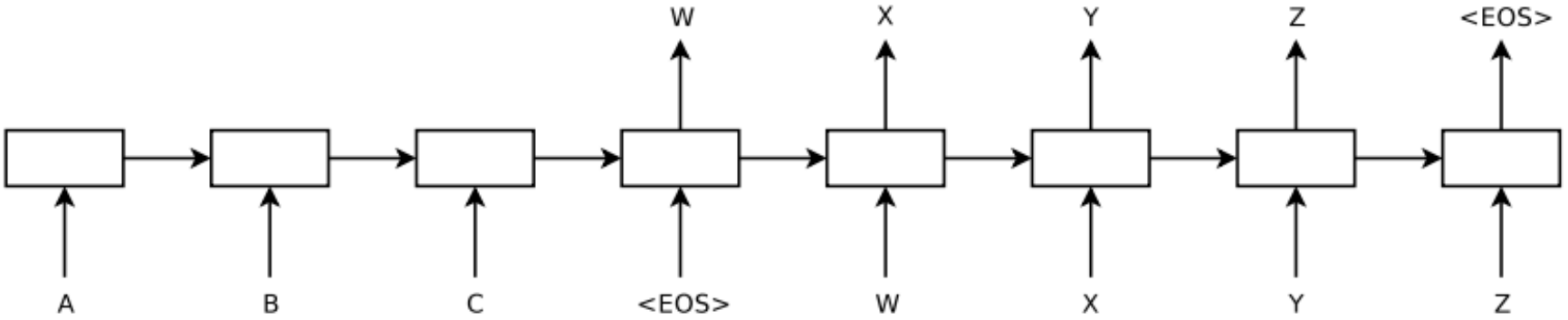
"black cat is sitting on top of suitcase."

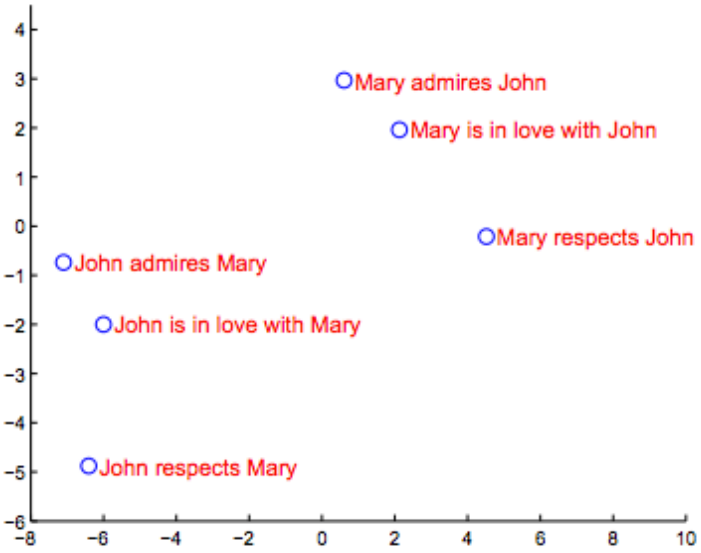
"a cat is sitting on a couch with a remote control."

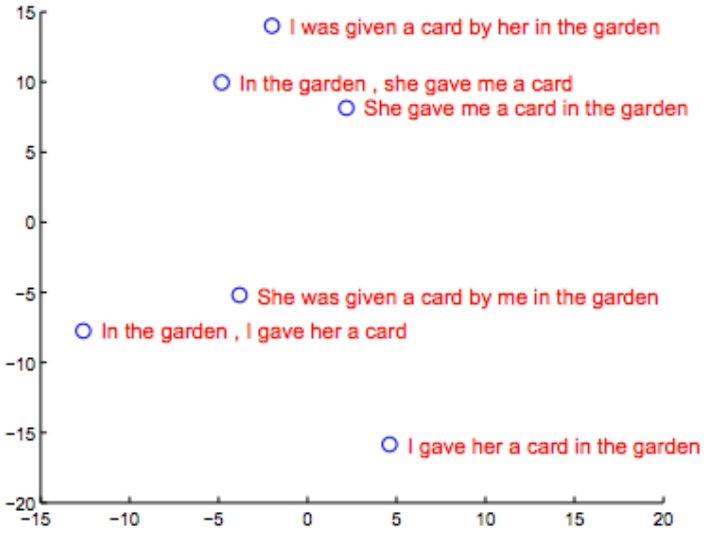
"a woman holding a teddy bear in front of a mirror."

[Karpathy et al, 2015]

Machine translation



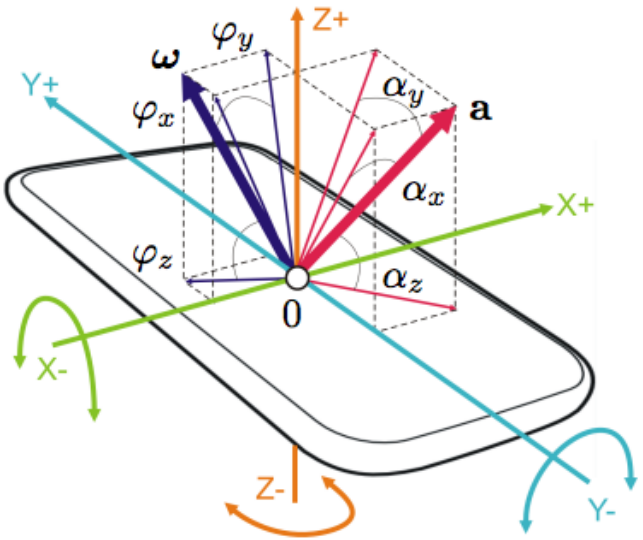


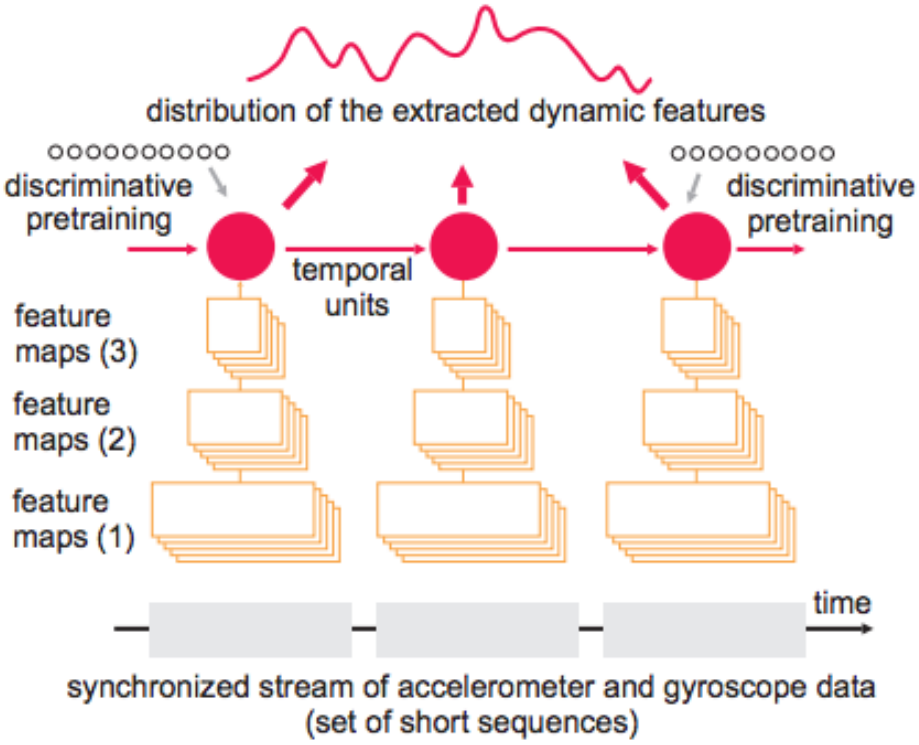


[Sutskever, Vinyals, Le, NIPS 2014]

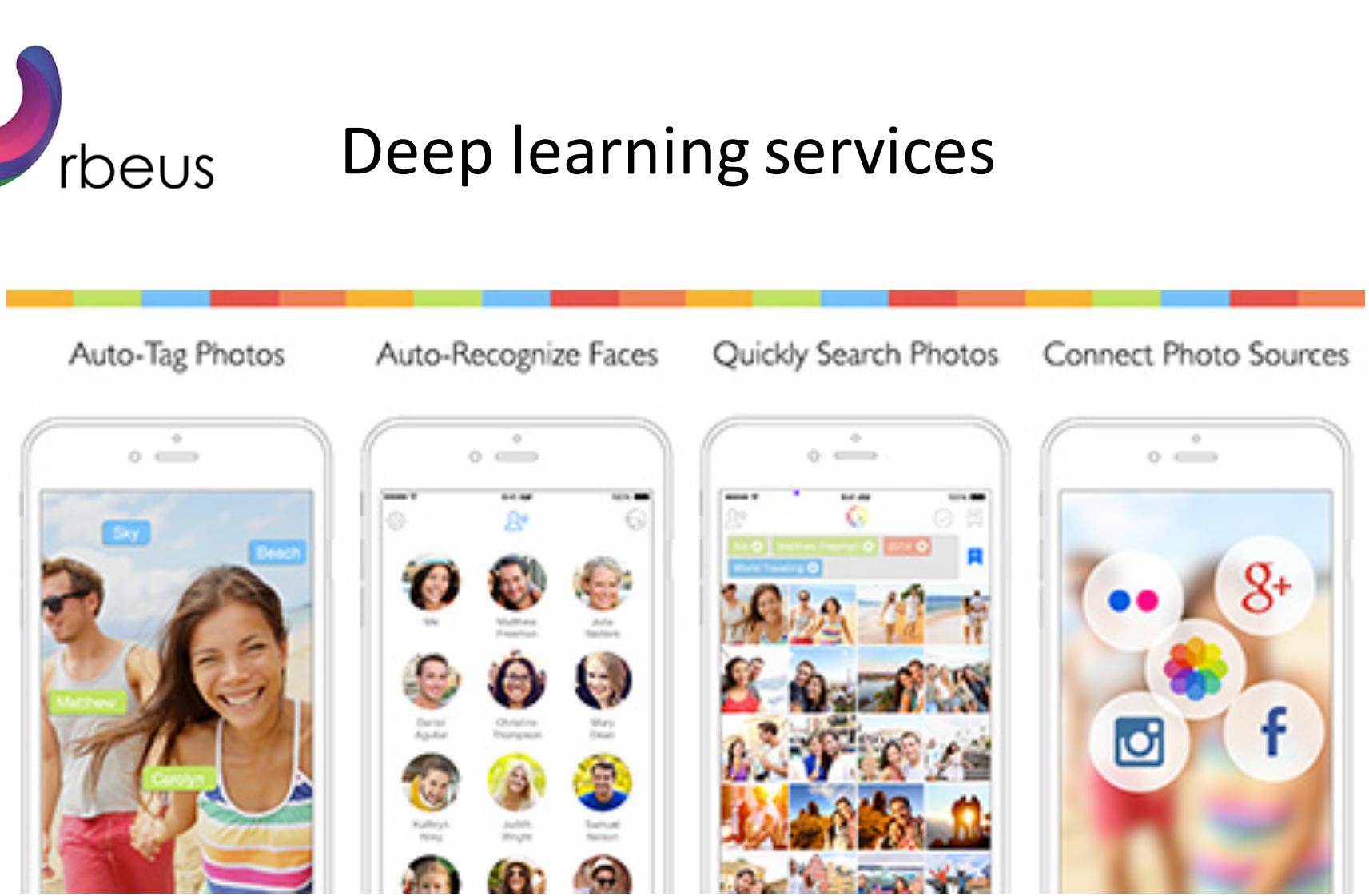
Google

Mobile authentication



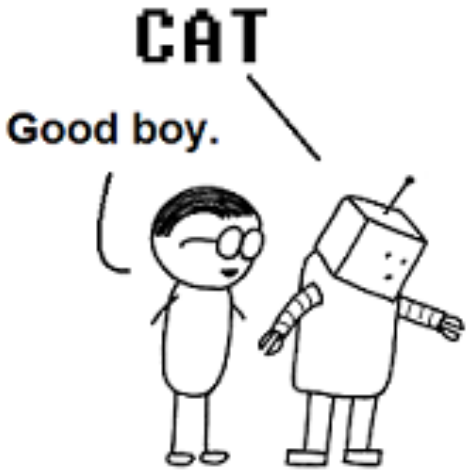


[Neverova et al, forthcoming publication]


rbeus
Deep learning services
Auto-Tag Photos
Auto-Recognize Faces
Quickly Search Photos
Connect Photo Sources
Sky
Beach
Matthew
Carolyn

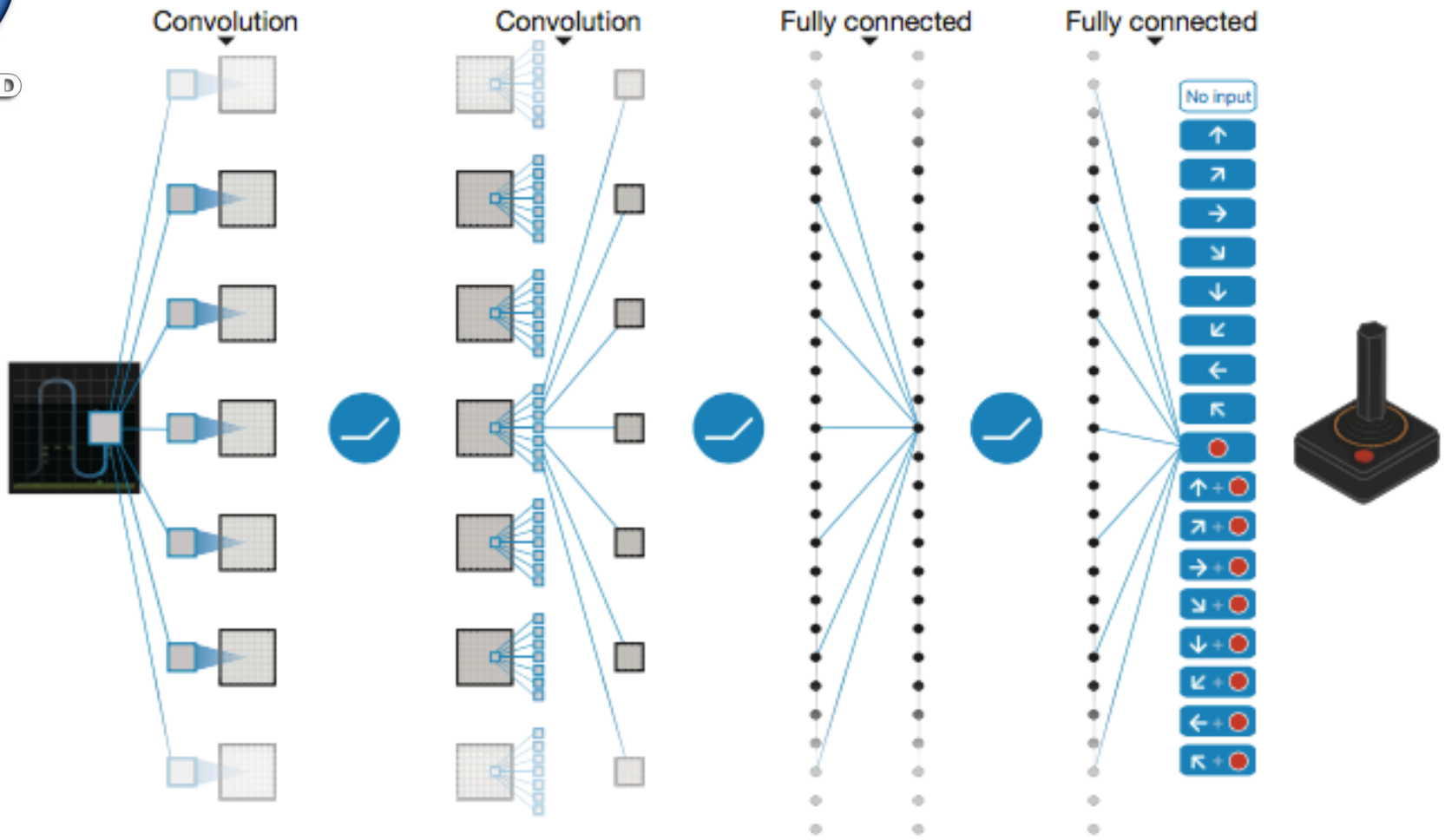
clarifai

Unsupervised deep learning?



[Le et al, ICML 2012]

Reinforcement deep learning?



[Mnih et al, Nature 2015]

Google DeepMind's
Deep Q-learning

Reinforcement deep learning?

Deep Sensorimotor Learning

rll.berkeley.edu/deeplearningrobotics

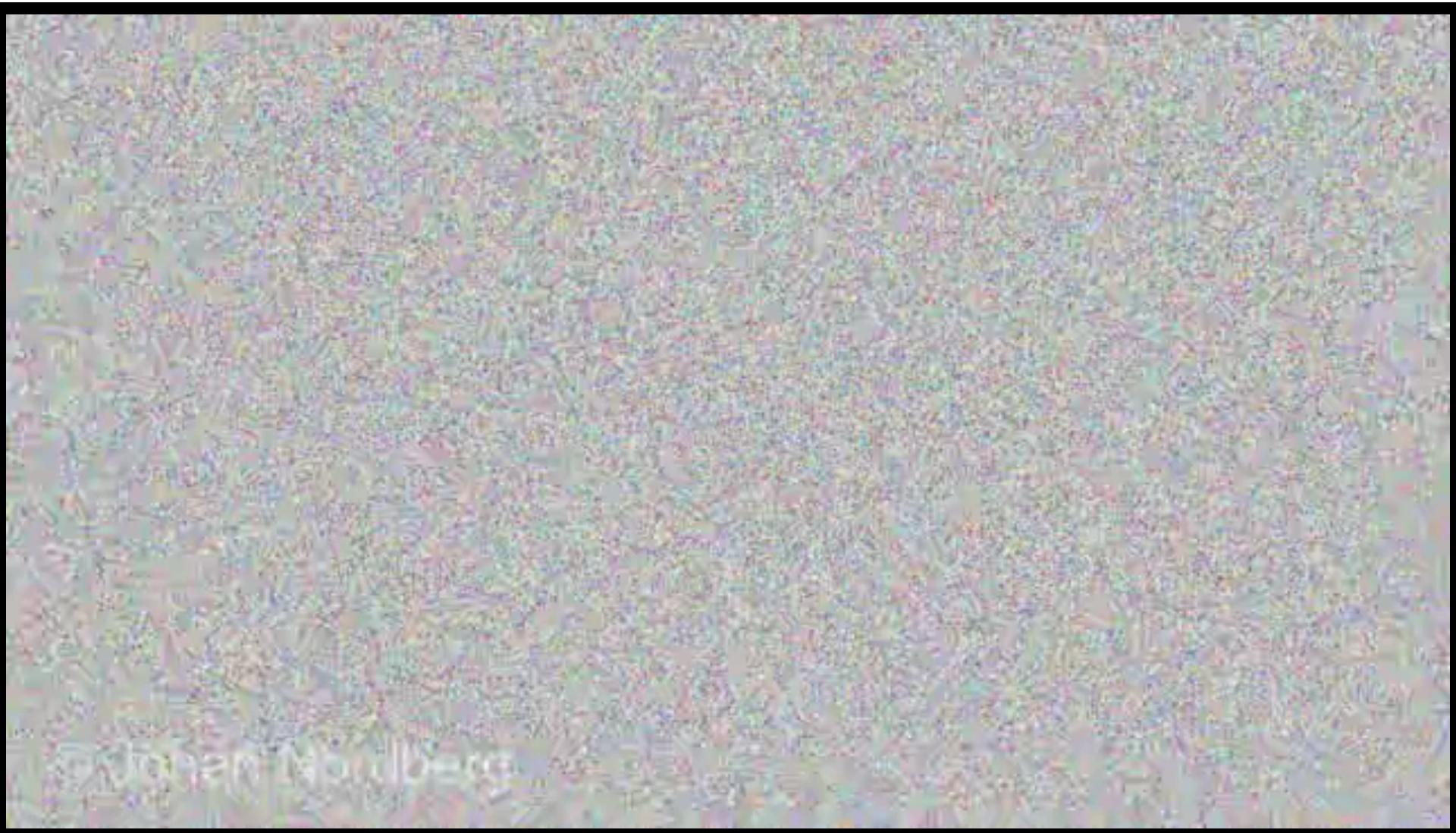
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

All roads lead to deep learning?

- Deep learning methods are general and work well with completely different kinds of data (text, sparse data).
- Depends on large data sets, which are not available in many domains.
- Hungry for computational resources.
- Requires certain expertise to train.

Future?

- Deep learning: big data + computational resources + efficient training + model complexity
- Deep learning chips, cloud services, specific and general software libraries
- Deep supervised learning in action, unsupervised and reinforcement learning coming soon
- Applications: “understanding” text and visual context, question answering, more efficient temporal models, generating content, new data types



Natalia Neverova, natalia.neverova@liris.cnrs.fr