Convolutional Neural Networks

An Overview

Guilherme Folego

2016-10-27

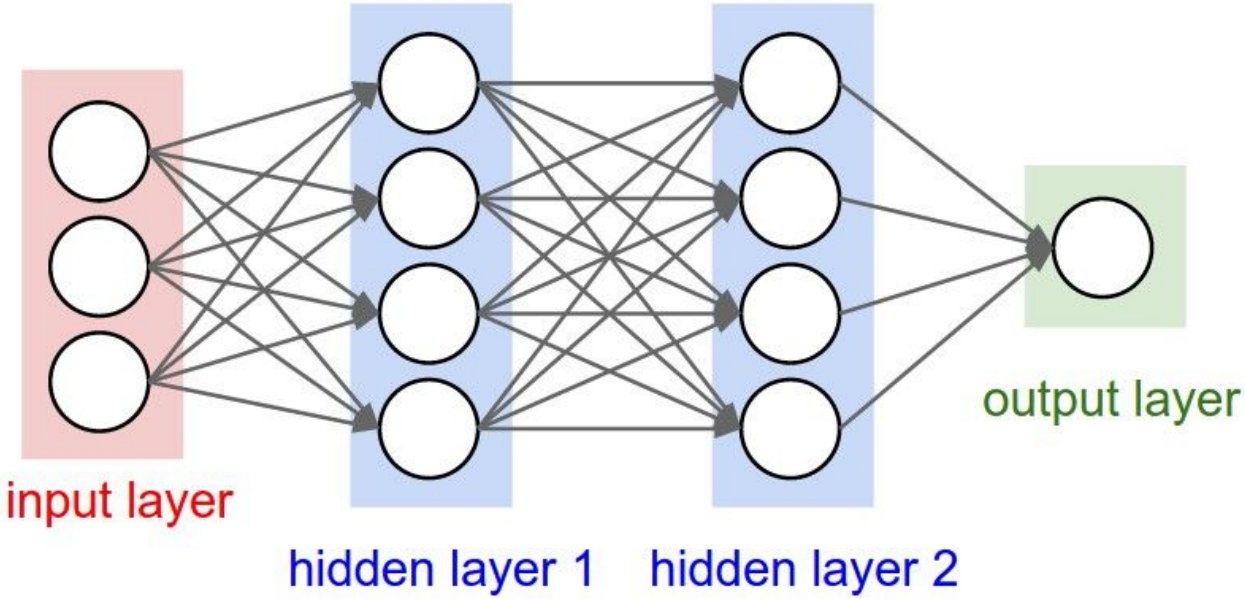
Objectives

What is a Convolutional Neural Network ?

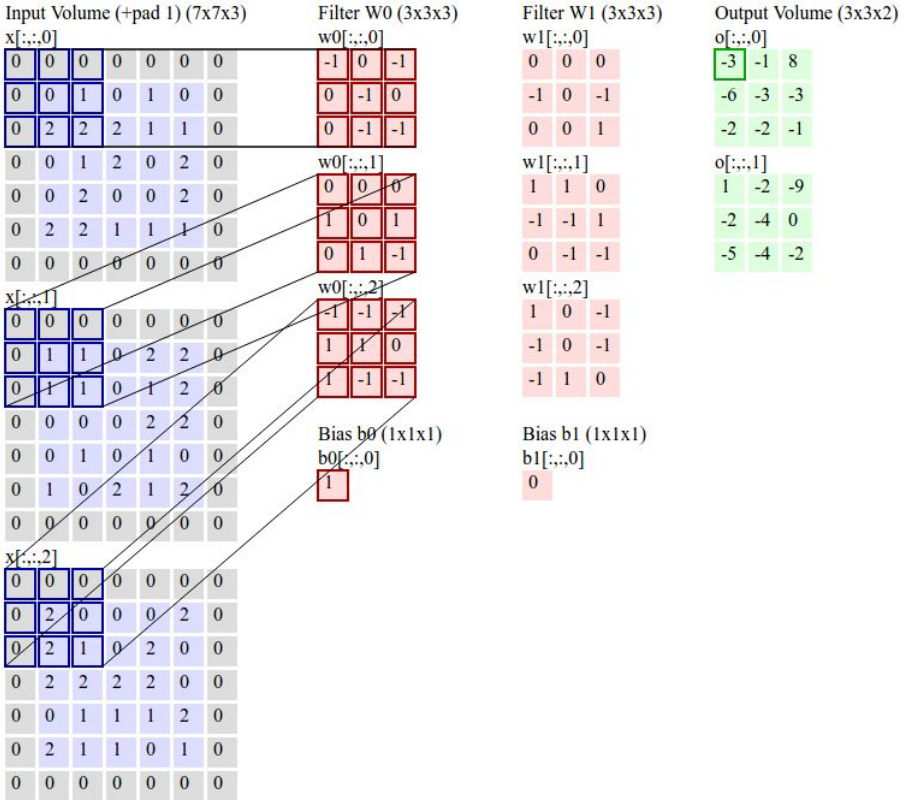
What is it good for ?

Why now ?

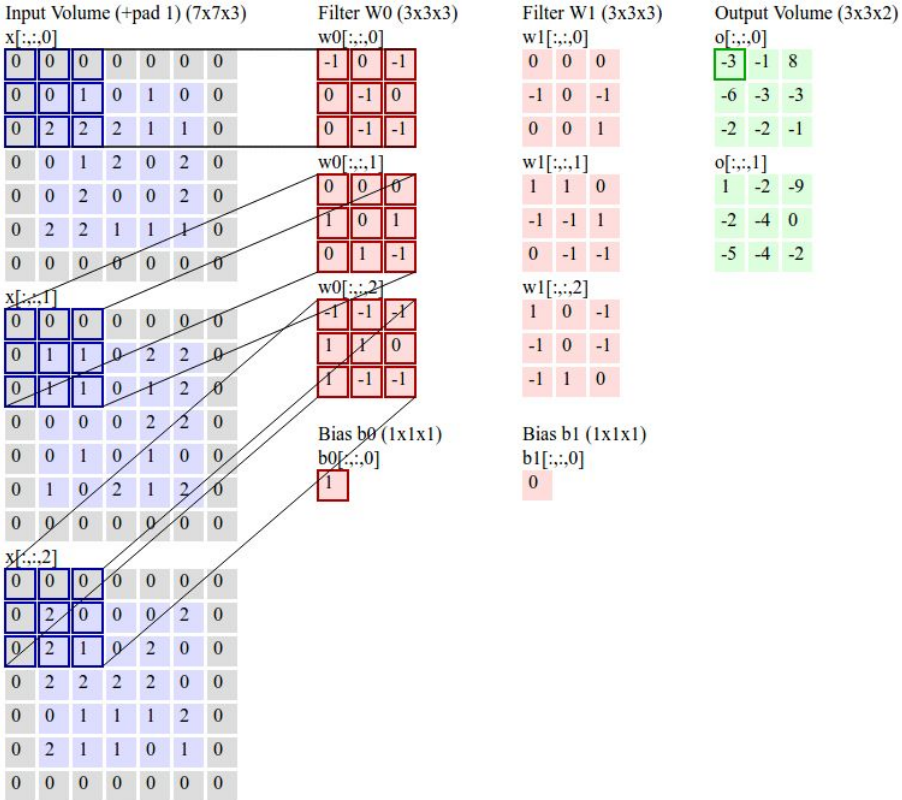
Neural Network



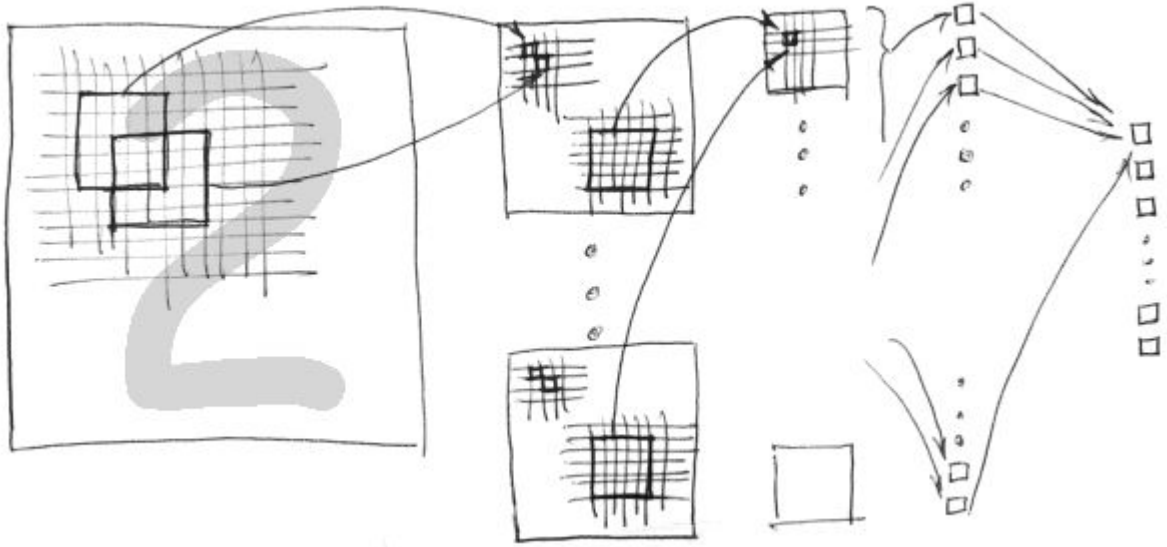
Convolutional Neural Network



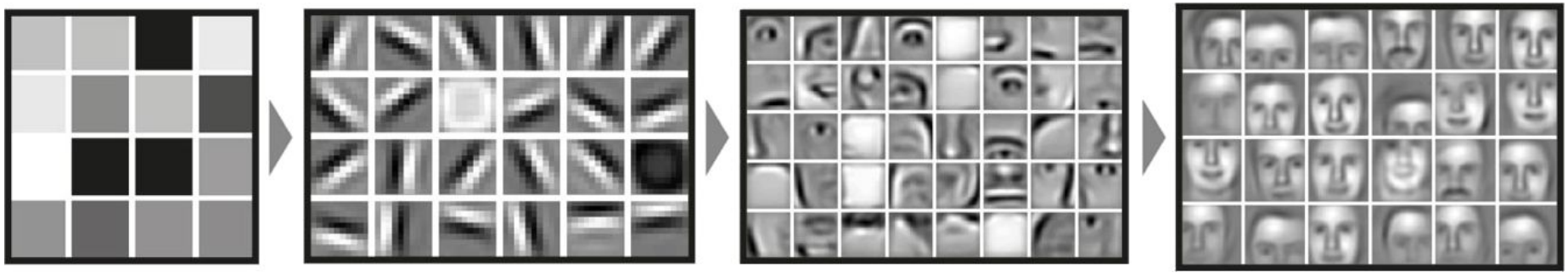
Convolutional Neural Network



Convolutional Neural Network



Convolutional Neural Network



LeNet

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W. and Jackel, L.D., ***1989. Backpropagation applied to handwritten zip code recognition.*** *Neural computation, 1(4), pp.541-551.*

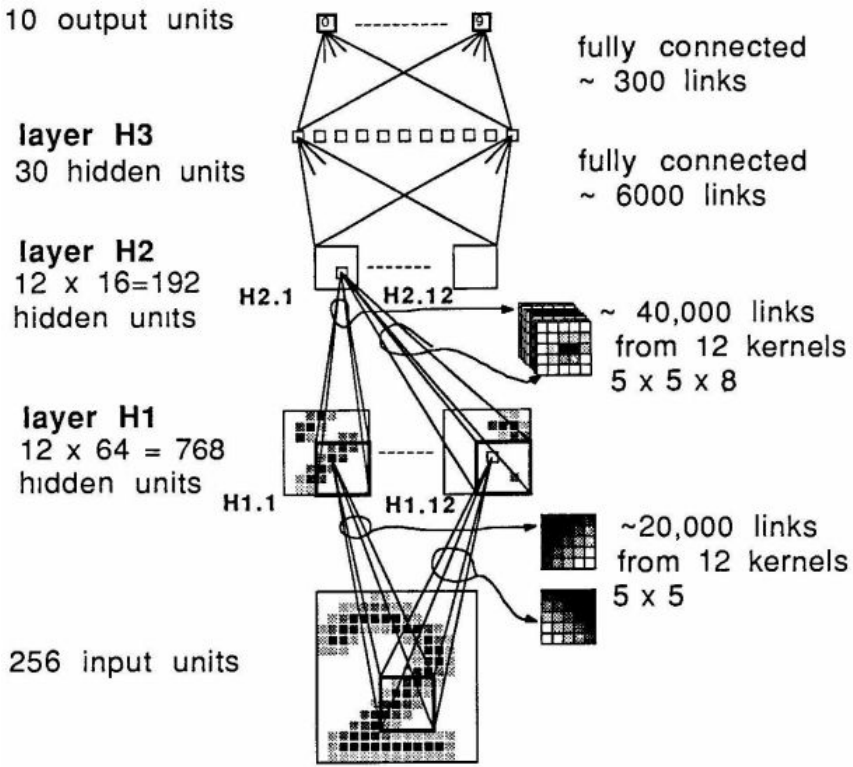
Google Scholar: Cited by 1846

LeNet

Highlights

- *"In summary, the network has 1,256 units, 64,660 connections, and 9,760 independent parameters."*
- *"… our training times were 'only' 3* ***days****"*
- *"We used an off-the-shelf board that contains 256 kbytes of local memory and … 25 MFLOPS"*
- *"This work points out the necessity of having flexible 'network design' software tools that ease the design of complex, specialized network architectures"*

LeNet


10 output units
0
9
fully connected
~ 300 links
layer H3
30 hidden units
fully connected
~ 6000 links
layer H2
12 x 16=192
hidden units
H2.1
H2.12
~ 40,000 links
from 12 kernels
5 x 5 x 8
layer H1
12 x 64 = 768
hidden units
H1.1
H1.12
~20,000 links
from 12 kernels
5 x 5
256 input units

LeNet-5

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., ***1998. Gradient-based learning applied to document recognition.*** *Proceedings of the IEEE, 86(11), pp.2278-2324.*

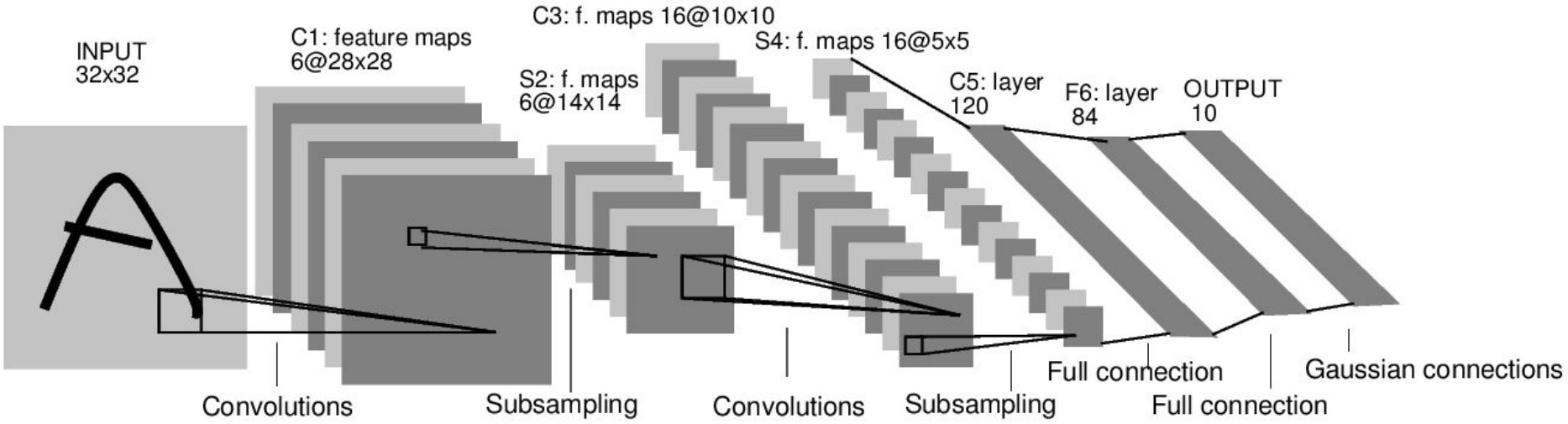
Google Scholar: Cited by 5964

LeNet-5

Highlights

- *Deployed commercially, reading "several million checks per day" (about 15% of all checks in the USA at the time)*
- *Introduced LeNet-5, arguably the most used CNN for teaching the subject or demonstrating a framework*
- *"Database: the Modified NIST set" (now known as MNIST, with about 60,000 images)*

LeNet-5


INPUT
32x32
C1: feature maps
6@28x28
S2: f. maps
6@14x14
C3: f. maps 16@10x10
S4: f. maps 16@5x5
C5: layer
120
F6: layer
84
OUTPUT
10
Convolutions
Subsampling
Convolutions
Subsampling
Full connection
Full connection
Gaussian connections

“AI winter” for neural nets in the 90’s

The Deep Learning Conspiracy

Around 2006, some papers on CNN started emerging …

- *CIFAR & "The Deep Learning Conspiracy"*
- *LeCun, Y., Bengio, Y., and Hinton, G. E.*

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., ***2009, June. Imagenet: A large-scale hierarchical image database.*** *In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on (pp. 248-255). IEEE.*

Google Scholar: Cited by 2964

ImageNet

And the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) started in 2010

- *"ImageNet is a dataset of over 15 million labeled high-resolution images belonging to roughly 22,000 categories."*
- *"ILSVRC uses a subset of ImageNet with roughly 1,000 images in each of 1,000 categories. In all, there are roughly 1.2 million training images, 50,000 validation images, and 150,000 testing images."*

Krizhevsky (SuperVision)

Krizhevsky, A., Sutskever, I. and Hinton, G.E., ***2012. Imagenet classification with deep convolutional neural networks.*** *In Advances in neural information processing systems (pp. 1097-1105).*

Google Scholar: Cited by 7153

Krizhevsky (SuperVision)

Highlights

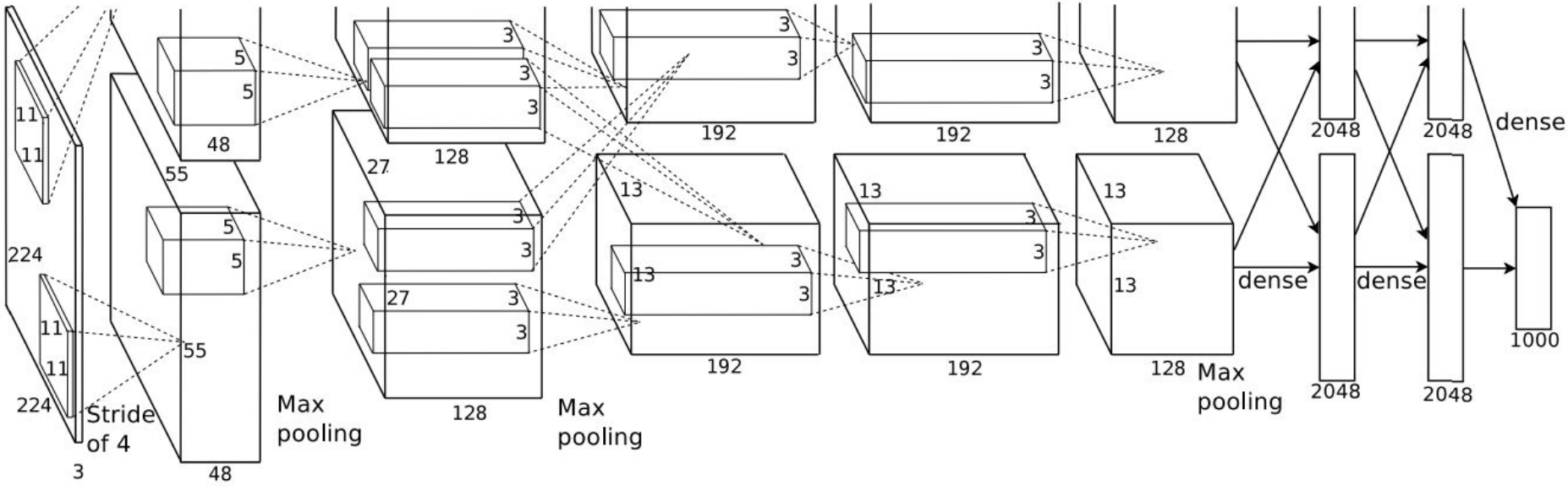
- *This paper completely changed the scenario*
  - *The first deep convolutional neural network entry in ILSVRC*
  - *Nearly half the error rate of the second-best entry*
    - *15.3% vs. 26.2%*
- *Network named SuperVision*
- *Code released: cuda-convnet*

Krizhevsky (SuperVision)

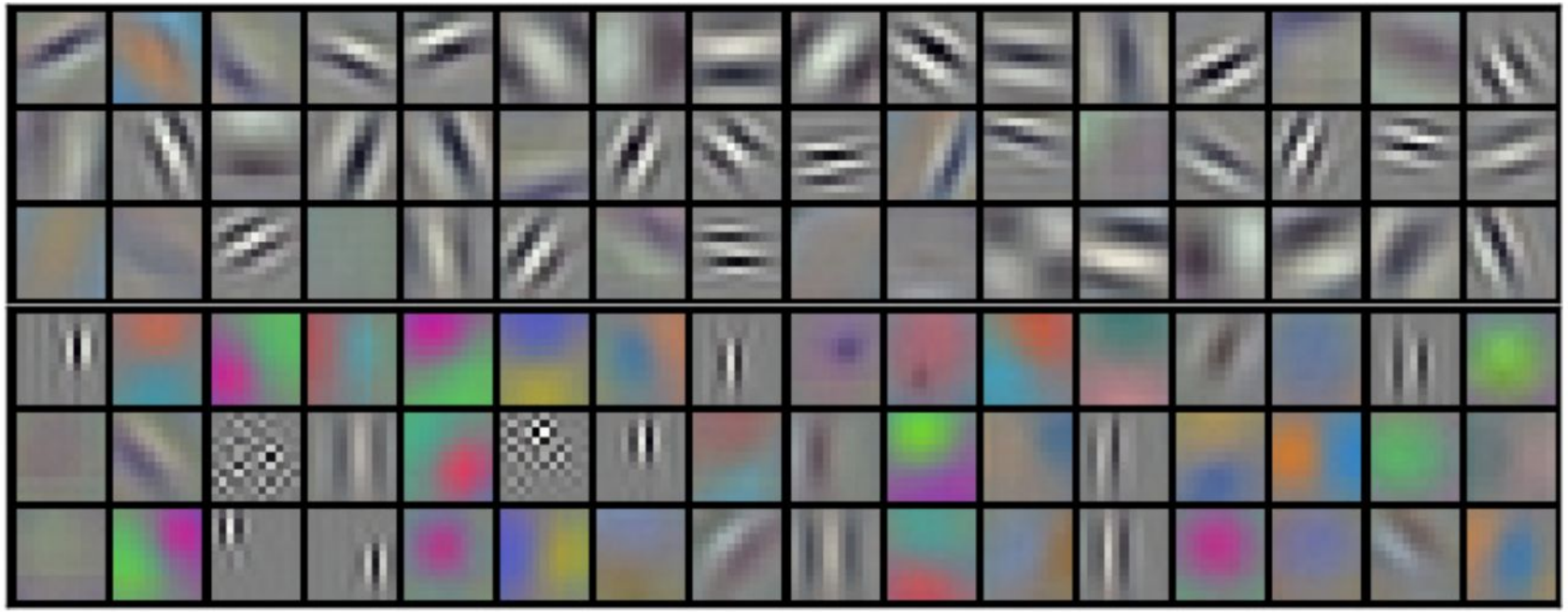
Highlights

- *Network's size is limited by the amount of memory available*
- *Between five and six days to train on two GTX 580 3GB GPUs [1,581,100 MFLOPS]*
- *"All of our experiments suggest that our results can be improved simply by waiting for faster GPUs and bigger datasets to become available."*

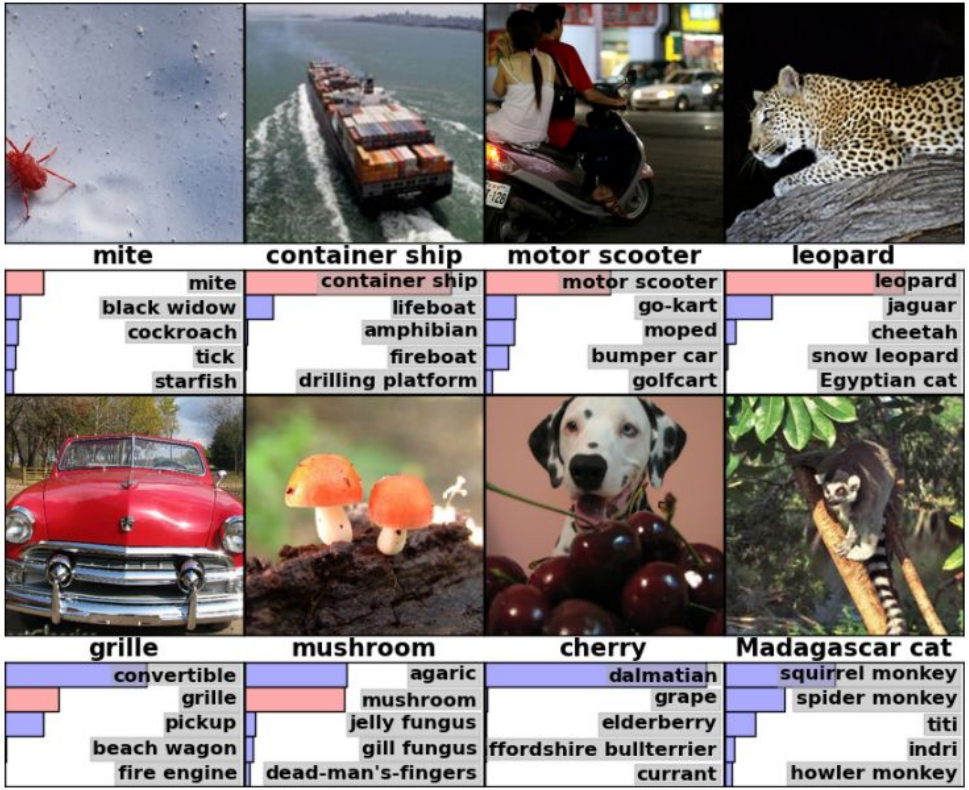
Krizhevsky (SuperVision)



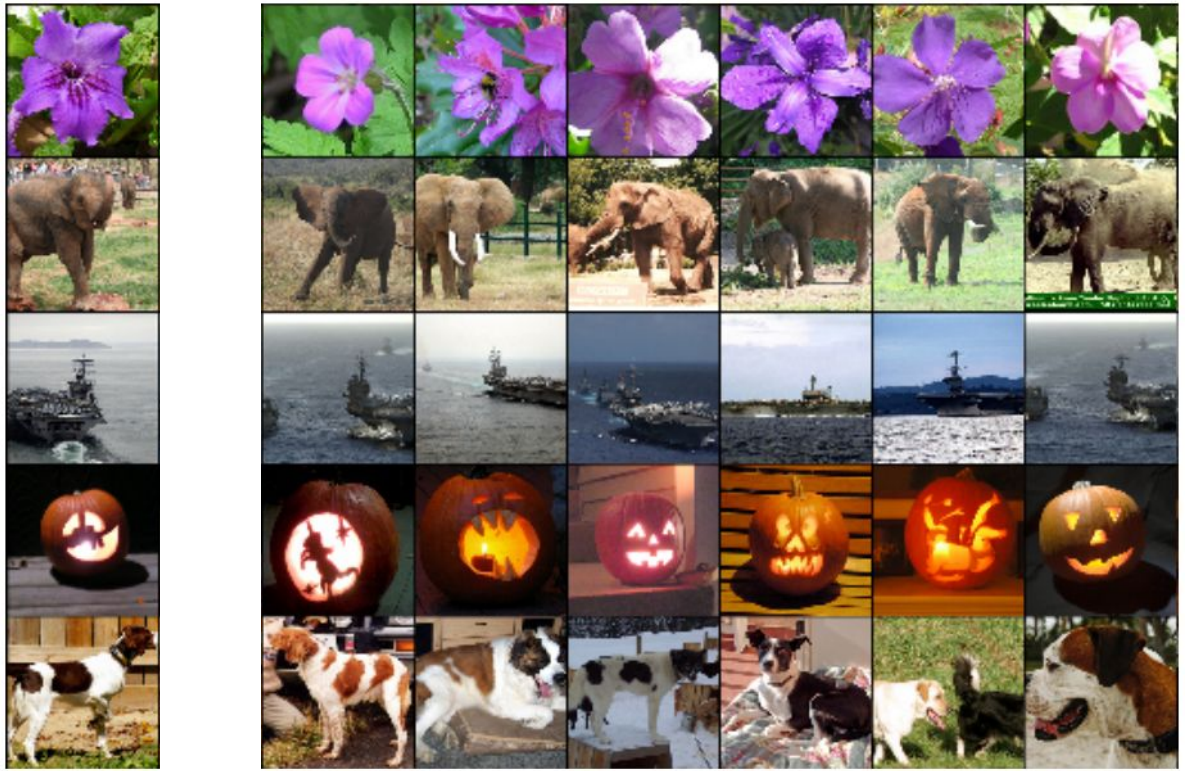
Krizhevsky (SuperVision)



Krizhevsky (SuperVision)



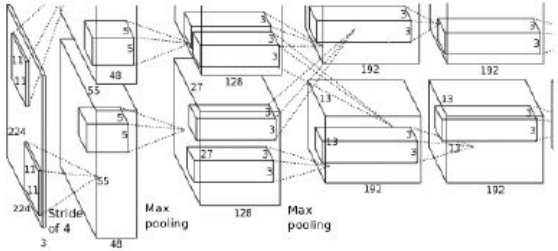
Krizhevsky (SuperVision)



The Deep Learning "Computer Vision Recipe"

Big Data: ImageNet

+



Deep Convolutional Neural Network

+

Backprop on GPU

=

Learned Weights

OverFeat

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R. and LeCun, Y., ***2013. Overfeat: Integrated recognition, localization and detection using convolutional networks.*** *arXiv preprint arXiv:1312.6229.*

Google Scholar: Cited by 943

OverFeat

Highlights

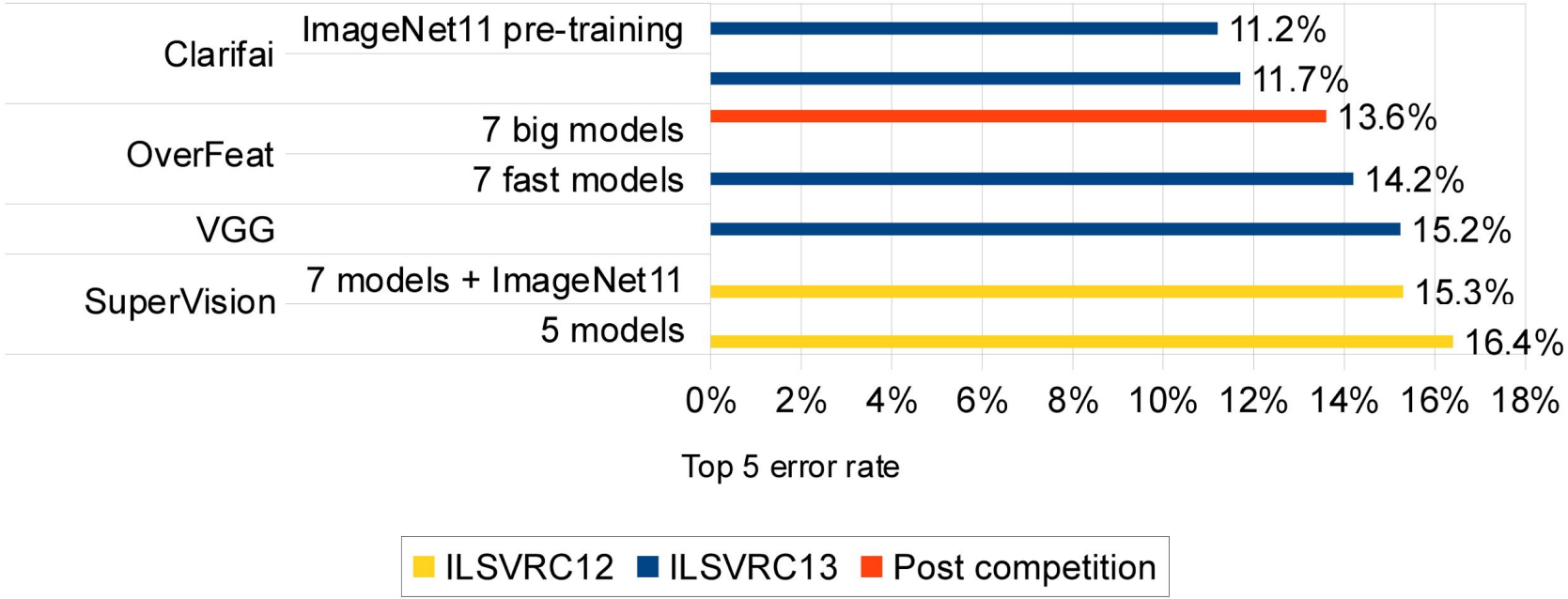
- *Improved on previous results*
  - *Winner of the localization task*
  - *Very competitive results on the detection and classification tasks*
- *Network named OverFeat*
- *Code released*
- ***Network weights released!***

OverFeat

model	# parameters (in millions)	# connections (in millions)
Krizhevsky	60	-
fast	145	2810
accurate	144	5369

Table 4: **Number of parameters and connections** for different models.

OverFeat


Clarifai
ImageNet11 pre-training
11.2%
11.7%
OverFeat
7 big models
13.6%
7 fast models
14.2%
VGG
15.2%
SuperVision
7 models + ImageNet11
15.3%
5 models
16.4%
0%
2%
4%
6%
8%
10%
12%
14%
16%
18%
Top 5 error rate
ILSVRC12
ILSVRC13
Post competition

Transfer Learning

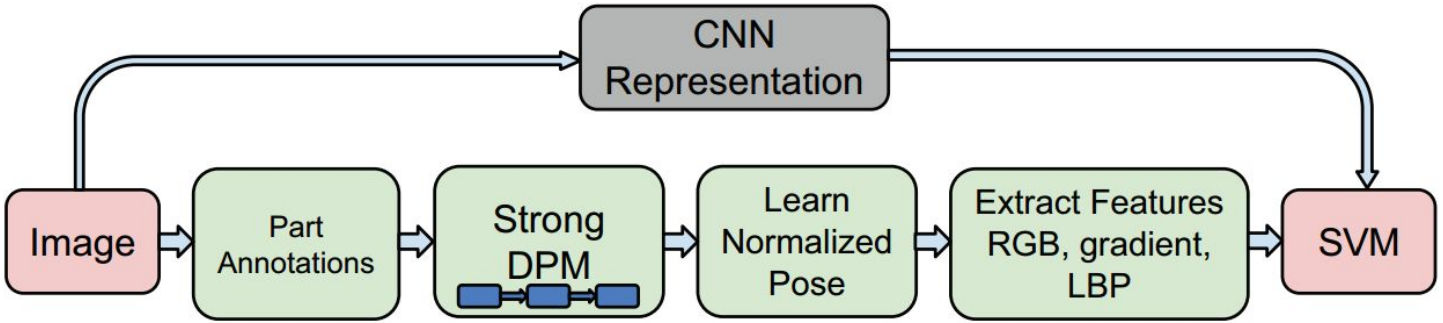
Sharif Razavian, A., Azizpour, H., Sullivan, J. and Carlsson, S., ***2014. CNN features off-the-shelf: an astounding baseline for recognition.*** *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 806-813).*

Google Scholar: Cited by 728

Transfer Learning

Highlights

- *"The results are achieved using a linear SVM classifier (or L2 distance in case of retrieval) applied to a feature representation of size 4096 extracted from a layer in the net."*

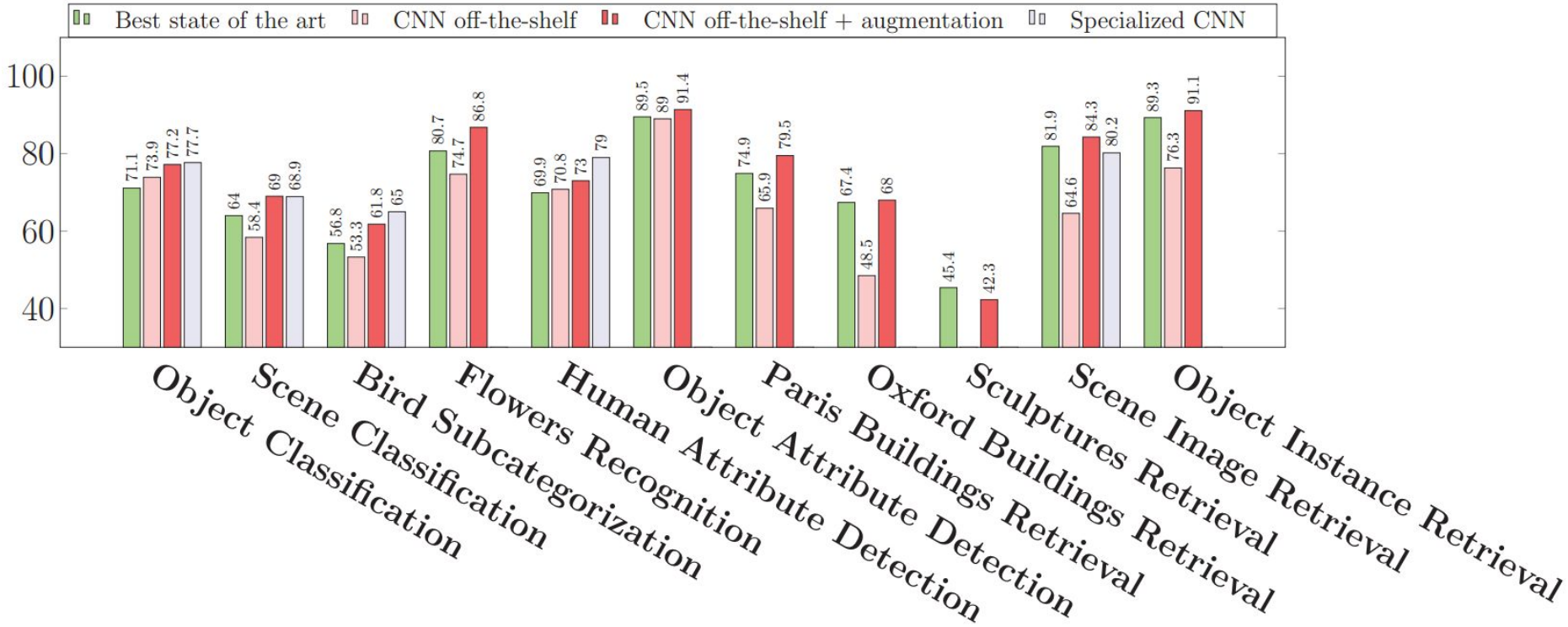


Transfer Learning

Highlights

- ***"The results strongly suggest that features obtained from deep learning with convolutional nets should be the primary candidate in most visual recognition tasks."***

Transfer Learning



Transfer Learning

Penatti, O. A., Nogueira, K. and dos Santos, J. A., ***2015. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?.*** *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 44-51).*

Google Scholar: Cited by 32

Micael Cabrera Carvalho's dissertation
http://www.bibliotecadigital.unicamp.br/document/?code=000956410

VGG

Simonyan, K. and Zisserman, A., ***2014. Very deep convolutional networks for large-scale image recognition.*** *arXiv preprint arXiv:1409.1556.*

Google Scholar: Cited by 2162

VGG

Highlights

- *Improved on previous results*
  - *First place in the localization task*
  - *Second place in the classification task*
- *Network named VGG*
- *Network architecture is very uniform*
- *Code based on Caffe framework*
- ***Network weights released!***

VGG

Highlights

- *"On a system equipped with four NVIDIA Titan Black GPUs [5,120,600 MFLOPS], training a single net took 2–3* ***weeks*** *depending on the architecture."*

VGG

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					

VGG

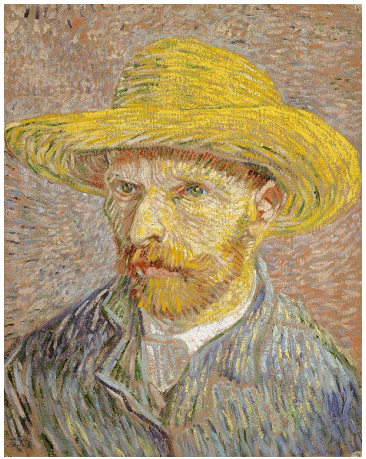
Table 2: **Number of parameters** (in millions).

Network	A,A-LRN	B	C	D	E
Number of parameters	133	133	134	138	144

Van Gogh

UNICAMP

Folego, G., Gomes, O. and Rocha, A., ***2016.***
From Impressionism to Expressionism:
Automatically identifying van Gogh's paintings.
In Image Processing (ICIP), 2016
IEEE International Conference on (pp. 141-145).



Artistic Style

Gatys, L.A., Ecker, A.S. and Bethge, M., ***2015. A neural algorithm of artistic style.*** *arXiv preprint arXiv:1508.06576.*

Google Scholar: Cited by 91

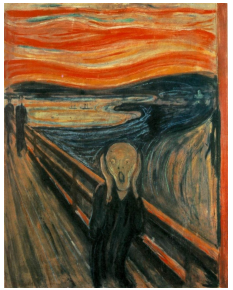
Artistic Style

Highlights

- *Based on VGG network architecture and weights*
- *“The key finding of this paper is that the representations of content and style in the Convolutional Neural Network are separable. That is, we can manipulate both representations independently to produce new, perceptually meaningful images.”*

Artistic Style

Artistic Style

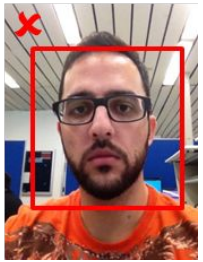


Artistic Style

Driver's Licence vs. Selfie

Folego, G., Angeloni, M. A., Stuchi, J. A., Rocha, A., Godoy, A., ***2016. Cross-Domain Face Verification: Matching ID Document and Self-Portrait Photographs.***

Accepted at the XII Workshop on Computer Vision (WVC 2016)



GoogLeNet

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., ***2015. Going deeper with convolutions.*** *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-9).*

Google Scholar: Cited by 1672

GoogLeNet

Highlights

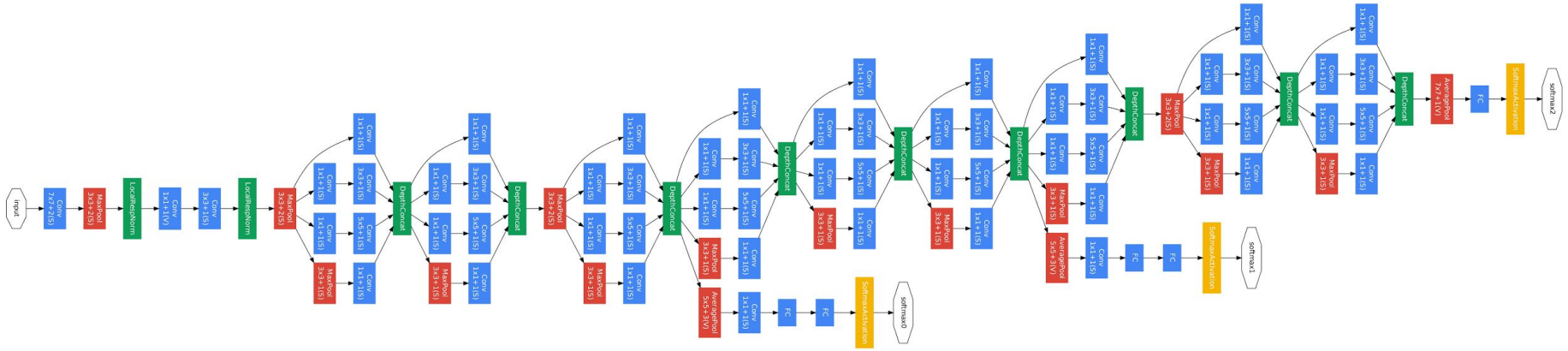
- *"... called **GoogLeNet**, a 22 layers deep network, ..."*
- *"For most of the experiments, the models were designed to keep a computational budget of 1.5 billion multiply-adds at inference time, so that they do not end up to be a purely academic curiosity, but could be put to real world use, even on large datasets, at a reasonable cost."*
- *"GoogLeNet networks were trained using the DistBelief distributed machine learning system ..." (lots of CPUs)*

GoogLeNet

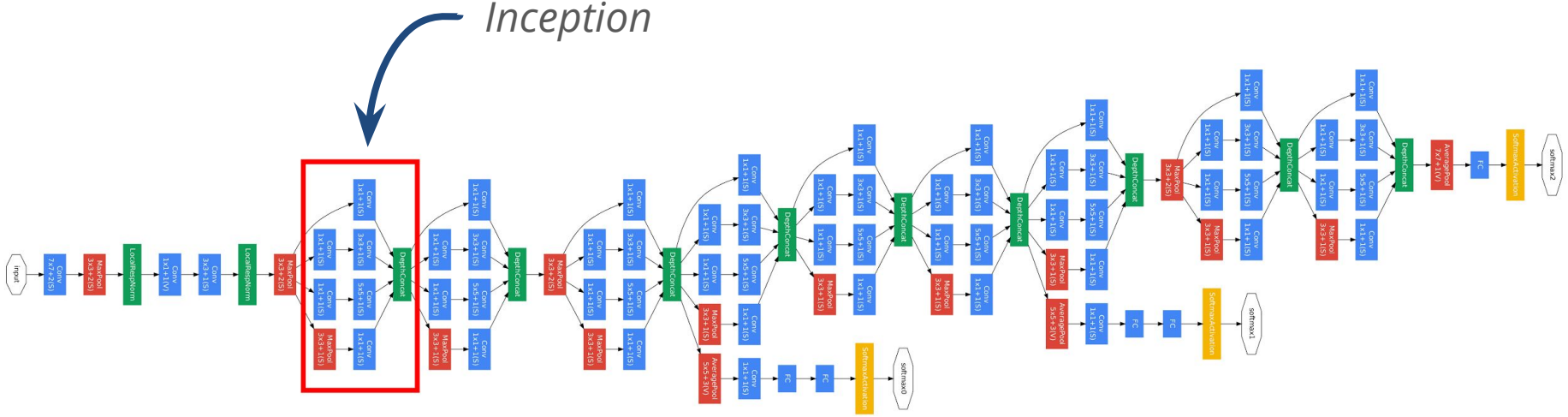
Highlights

- *The first reference is a meme*

GoogLeNet


input
Conv 7x7+2(S)
MaxPool 3x3+2(S)
LocalRespNorm
Conv 1x1+1(V)
Conv 3x3+1(S)
LocalRespNorm
MaxPool 3x3+2(S)
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
MaxPool 3x3+1(S)
Conv 1x1+1(S)
DepthConcat
Conv 3x3+2(S)
Conv 5x5+2(S)
MaxPool 3x3+2(S)
AveragePool 5x5+3(V)
FC
SoftmaxActivation
softmax0
softmax1
softmax2

GoogLeNet


Inception
input
Conv 7x7+2(S)
MaxPool 3x3+2(S)
LocalRespNorm
Conv 1x1+1(V)
Conv 3x3+1(S)
LocalRespNorm
MaxPool 3x3+2(S)
Conv 1x1+1(S)
Conv 3x3+1(S)
Conv 5x5+1(S)
MaxPool 3x3+1(S)
DepthConcat
Conv 1x1+1(S)
Conv 3x3+2(S)
Conv 5x5+1(S)
MaxPool 3x3+1(S)
Conv 1x1+1(S)
FC
SoftmaxActivation
softmax0
softmax1
softmax2
AveragePool 5x5+3(V)
AveragePool 7x7+1(V)

GoogLeNet

Figure 1: Two distinct classes from the 1000 classes of the ILSVRC 2014 classification challenge. Domain knowledge is required to distinguish between these classes.

GoogLeNet

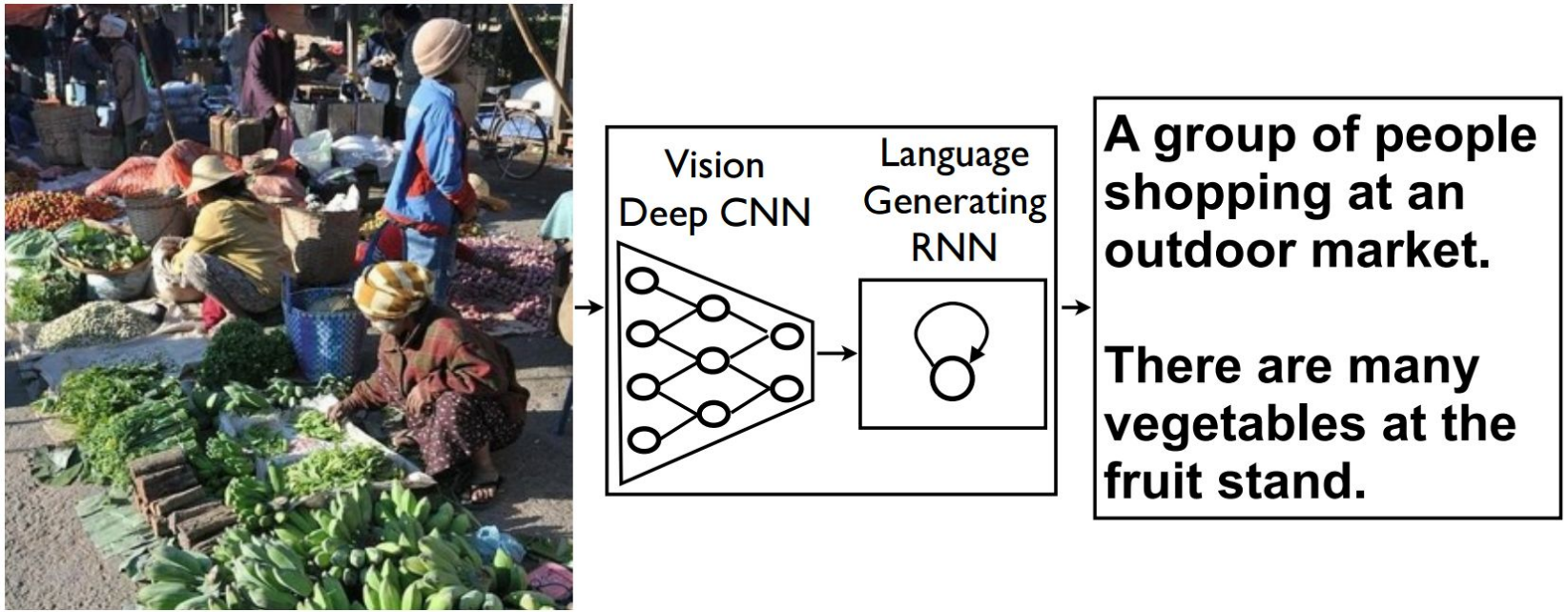
Team	Year	Place	Error (top-5)	Uses external data
SuperVision	2012	1st	16.4%	no
SuperVision	2012	1st	15.3%	Imagenet 22k
Clarifai	2013	1st	11.7%	no
Clarifai	2013	1st	11.2%	Imagenet 22k
MSRA	2014	3rd	7.35%	no
VGG	2014	2nd	7.32%	no
GoogLeNet	2014	1st	6.67%	no

Show and Tell

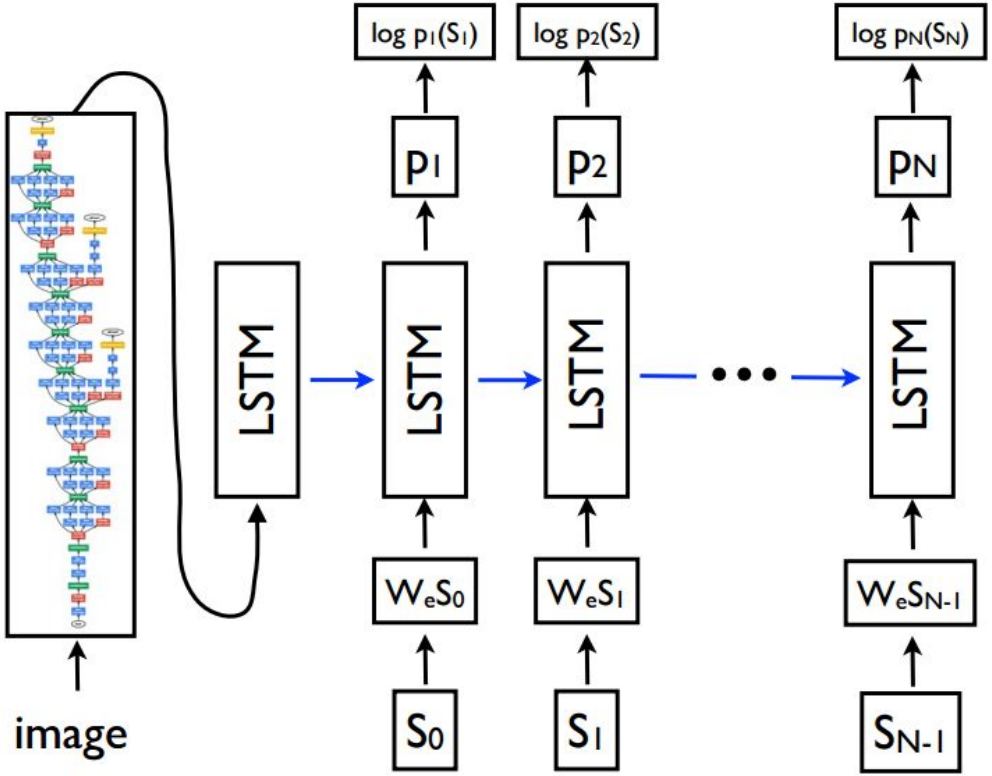
Vinyals, O., Toshev, A., Bengio, S. and Erhan, D., ***2015. Show and tell: A neural image caption generator.*** *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3156-3164).*

Google Scholar: Cited by 518

Show and Tell



Show and Tell



Show and Tell

A person riding a motorcycle on a dirt road.

Two dogs play in the grass.

A skateboarder does a trick on a ramp.

A dog is jumping to catch a frisbee.

A group of young people playing a game of frisbee.

Two hockey players are fighting over the puck.

A little girl in a pink hat is blowing bubbles.

A refrigerator filled with lots of food and drinks.

A herd of elephants walking across a dry grass field.

A close up of a cat laying on a couch.

A red motorcycle parked on the side of the road.

A yellow school bus parked in a parking lot.

Describes without errors	Describes with minor errors	Somewhat related to the image	Unrelated to the image

ResNet

He, K., Zhang, X., Ren, S. and Sun, J., ***2015. Deep residual learning for image recognition.*** *arXiv preprint arXiv:1512.03385.*

Google Scholar: Cited by 562

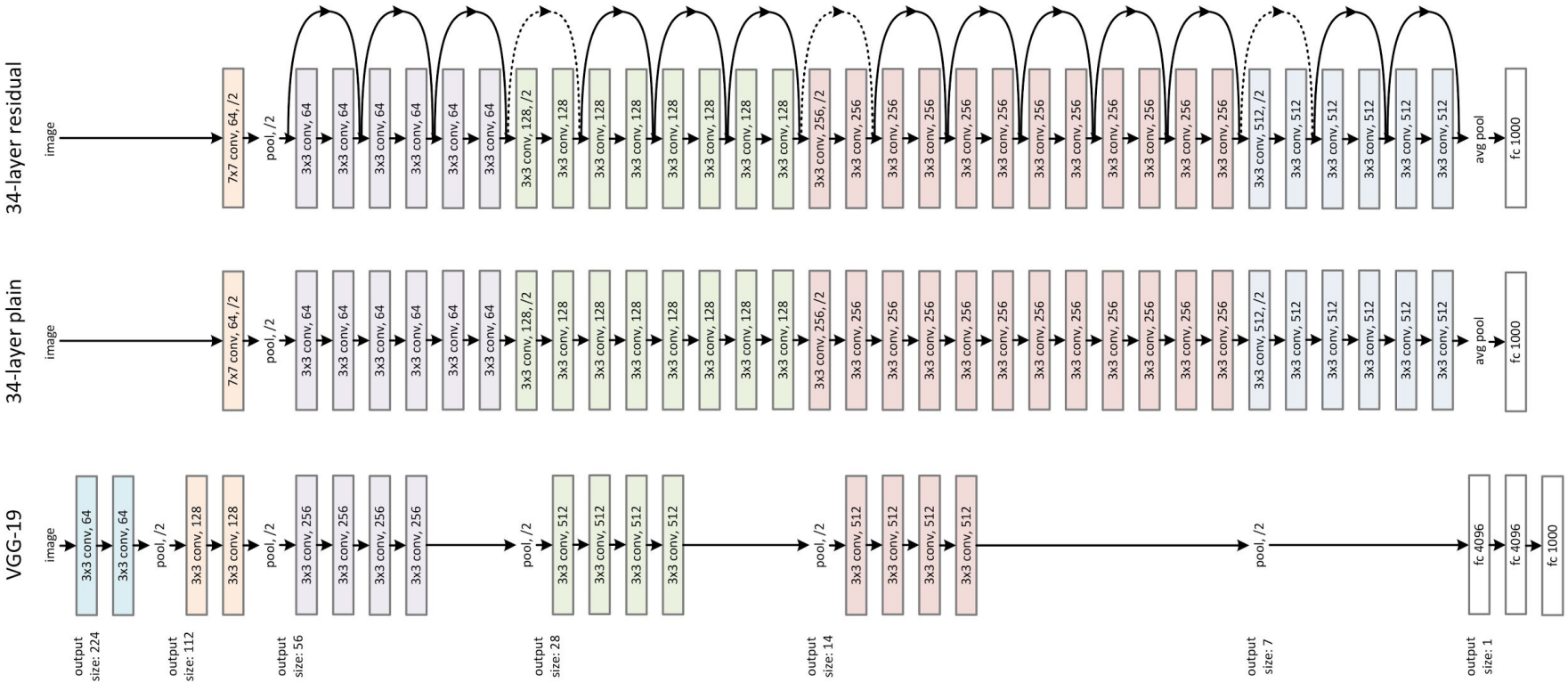
ResNet

Highlights

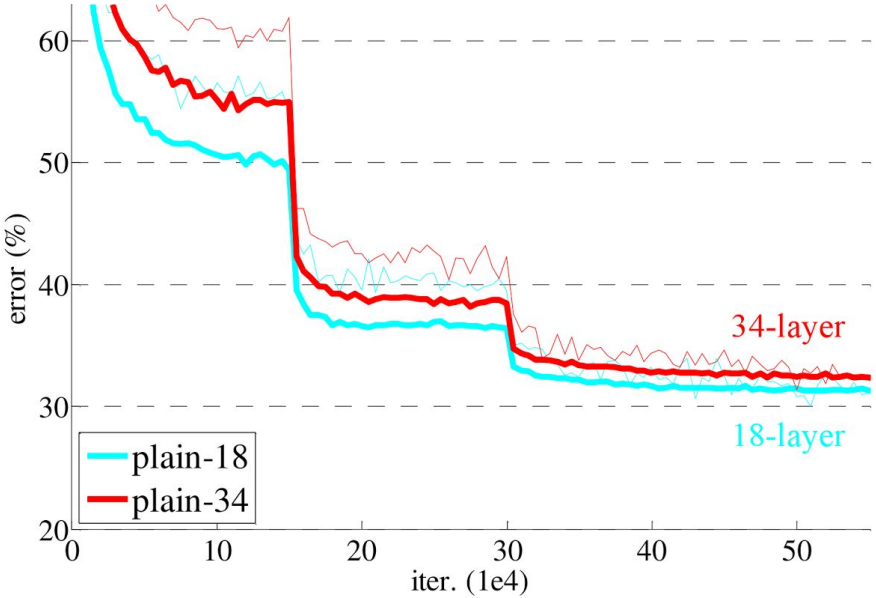
- *First place in ILSVRC 2015 classification, localization, and detection tasks*
- *"*Is learning better networks as easy as stacking more layers?*"*
- *"There exists a solution* by construction *to the deeper model: the added layers are* identity *mapping, and the other layers are copied from the learning shallower model."*
    - Degradation *problem*

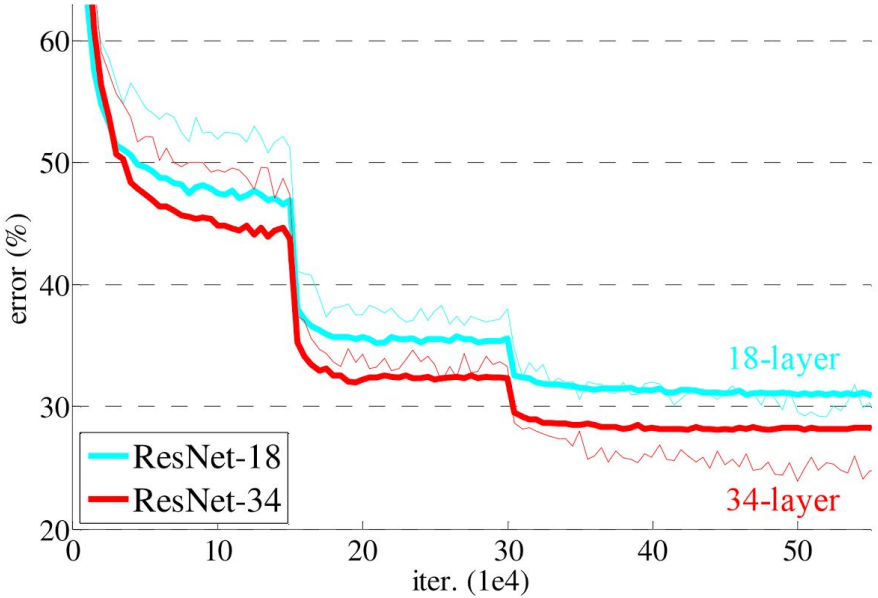
ResNet

19.6 B FLOPS vs *3.6 B*



ResNet


60
50
40
30
20
error (%)
0
10
20
30
40
50
iter. (1e4)
plain-18
plain-34
34-layer
18-layer


60
50
40
30
20
error (%)
0
10
20
30
40
50
iter. (1e4)
ResNet-18
ResNet-34
18-layer
34-layer

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**

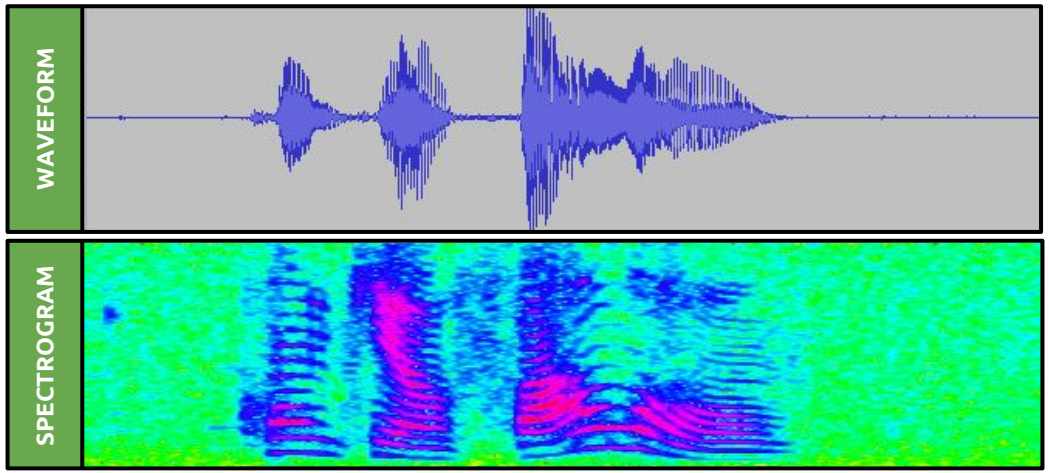
Table 5. Error rates (%) of **ensembles**. The top-5 error is on the test set of ImageNet and reported by the test server.

Object Recognition with Text-to-Speech

In Portuguese

Voice Activity Detection

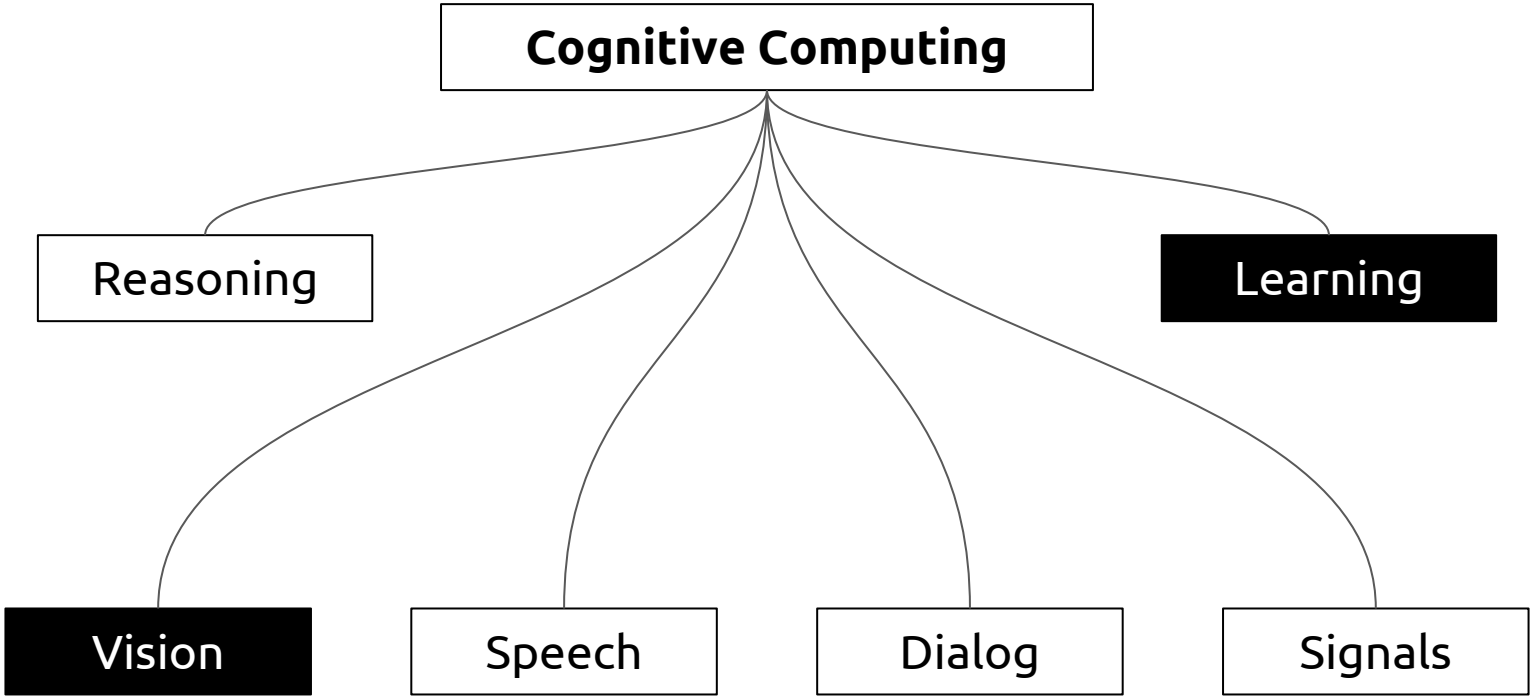
Silva, D. A., Stuchi, J. A., Violato, R. P. V., Cuozzo, L. G. D., ***2016. Exploring Convolutional Neural Networks for Voice Activity Detection.***
Accepted at Cognitive Technologies, CPqD Research Series – Springer



Alzheimer's Disease

Computer-aided diagnosis for Alzheimer's disease using 3D convolutional neural networks

Cognitive Computing



Conclusions

Deep Learning is rapidly evolving the machine learning field

Convolutional Neural Networks are key to this advance in the computer vision field

Lots of good data are necessary

Recent technologies are accessible

References

CS231n Convolutional Neural Networks for Visual Recognition

- *https://cs231n.github.io/*

***Deep Learning**, Yoshua Bengio, Ian Goodfellow, Aaron Courville, MIT Press, In preparation.*

- *http://www.deeplearningbook.org/*

A 'Brief' History of Neural Nets and Deep Learning

- *http://www.andreykurenkov.com/writing/a-brief-history-of-neural-nets-and-deep-learning/*

www.cpqd.com.br

TURNING
INTO REALITY

Guilherme Folego
gfolego@cpqd.com.br