EE-559 – Deep learning

1.1. From neural networks to deep learning

François Fleuret

https://fleuret.org/ee559/

Fri Jan 4 17:38:42 UTC 2019

RESEARCH INSTITUTE

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Why learning

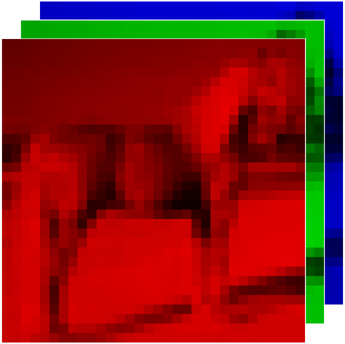
Many applications require the automatic extraction of “refined” information from raw signal (*e.g.* image recognition, automatic speech processing, natural language processing, robotic control, geometry reconstruction).

(ImageNet)

Our brain is so good at interpreting visual information that the "semantic gap" is hard to assess intuitively.

Our brain is so good at interpreting visual information that the "semantic gap" is hard to assess intuitively.

This:  is a horse



```
>>> from torchvision import datasets
>>> cifar = datasets.CIFAR10('./data/cifar10/', train=True, download=True)
Files already downloaded and verified
>>> x = torch.from_numpy(cifar.train_data)[43].transpose(2, 0).transpose(1, 2)
>>> x.size()
torch.Size([3, 32, 32])
>>> x[:,:4,:8]
tensor([[[ 99,  98, 100, 103, 105, 107, 108, 110],
         [ 100, 100, 102, 105, 107, 109, 110, 112],
         [ 104, 104, 106, 109, 111, 112, 114, 116],
         [ 109, 109, 111, 113, 116, 117, 118, 120]],

        [[ 166, 165, 167, 169, 171, 172, 173, 175],
         [ 166, 164, 167, 169, 169, 171, 172, 174],
         [ 169, 167, 170, 171, 171, 173, 174, 176],
         [ 170, 169, 172, 173, 175, 176, 177, 178]],

        [[ 198, 196, 199, 200, 200, 202, 203, 204],
         [ 195, 194, 197, 197, 197, 199, 200, 201],
         [ 197, 195, 198, 198, 198, 199, 201, 202],
         [ 197, 196, 199, 198, 198, 199, 200, 201]]], dtype=torch.uint8)
```

Extracting semantic automatically requires models of extreme complexity, which cannot be designed by hand.

Techniques used in practice consist of

1. defining a parametric model, and
2. optimizing its parameters by "making it work" on training data.

Extracting semantic automatically requires models of extreme complexity, which cannot be designed by hand.

Techniques used in practice consist of

1. defining a parametric model, and
2. optimizing its parameters by “making it work” on training data.

This is similar to biological systems for which the model (*e.g.* brain structure) is DNA-encoded, and parameters (*e.g.* synaptic weights) are tuned through experiences.

Extracting semantic automatically requires models of extreme complexity, which cannot be designed by hand.

Techniques used in practice consist of

1. defining a parametric model, and
2. optimizing its parameters by "making it work" on training data.

This is similar to biological systems for which the model (*e.g.* brain structure) is DNA-encoded, and parameters (*e.g.* synaptic weights) are tuned through experiences.

Deep learning encompasses software technologies to scale-up to billions of model parameters and as many training examples.

There are strong connections between standard statistical modeling and machine learning.

There are strong connections between standard statistical modeling and machine learning.

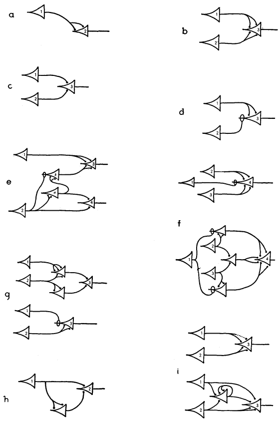
Classical ML methods combine a "learnable" model from statistics (*e.g.* "linear regression") with prior knowledge in pre-processing.

There are strong connections between standard statistical modeling and machine learning.

Classical ML methods combine a “learnable” model from statistics (*e.g.* “linear regression”) with prior knowledge in pre-processing.

“Artificial neural networks” pre-dated these approaches, and do not follow that dichotomy. They consist of “deep” stacks of parametrized processing.

From artificial neural networks to “Deep Learning”



Networks of "Threshold Logic Unit"

(McCulloch and Pitts, 1943)

1949 – Donald Hebb proposes the Hebbian Learning principle.

1951 – Marvin Minsky creates the first ANN (Hebbian learning, 40 neurons).

1949 – Donald Hebb proposes the Hebbian Learning principle.

1951 – Marvin Minsky creates the first ANN (Hebbian learning, 40 neurons).

1958 – Frank Rosenblatt creates a perceptron to classify 20×20 images.

1949 – Donald Hebb proposes the Hebbian Learning principle.

1951 – Marvin Minsky creates the first ANN (Hebbian learning, 40 neurons).

1958 – Frank Rosenblatt creates a perceptron to classify 20×20 images.

1959 – David H. Hubel and Torsten Wiesel demonstrate orientation selectivity and columnar organization in the cat's visual cortex.

1949 – Donald Hebb proposes the Hebbian Learning principle.

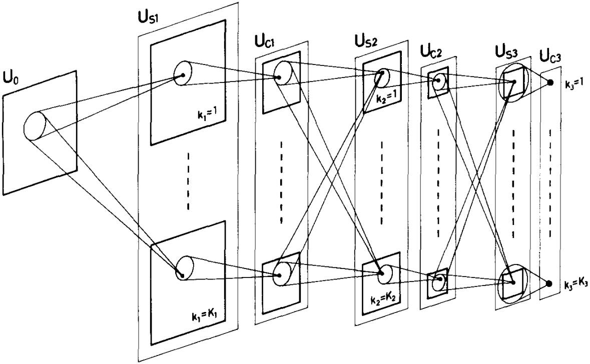
1951 – Marvin Minsky creates the first ANN (Hebbian learning, 40 neurons).

1958 – Frank Rosenblatt creates a perceptron to classify 20×20 images.

1959 – David H. Hubel and Torsten Wiesel demonstrate orientation selectivity and columnar organization in the cat's visual cortex.

1982 – Paul Werbos proposes back-propagation for ANNs.

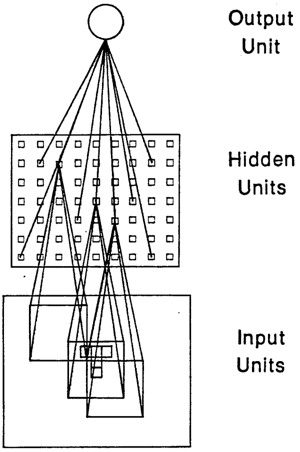
Neocognitron



Follows Hubel and Wiesel's results.

(Fukushima, 1980)

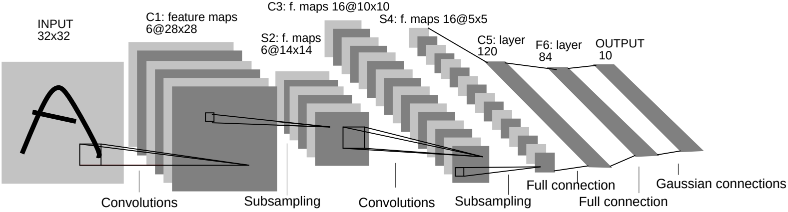
Network for the T-C problem



Trained with back-prop.

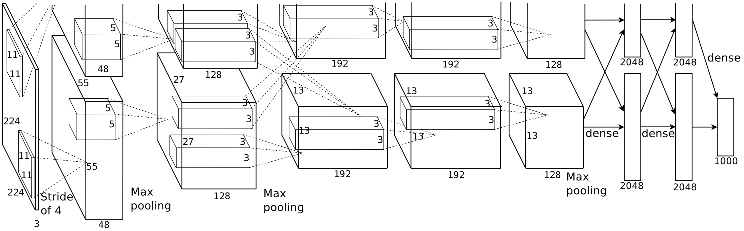
(Rumelhart et al., 1988)

LeNet-5



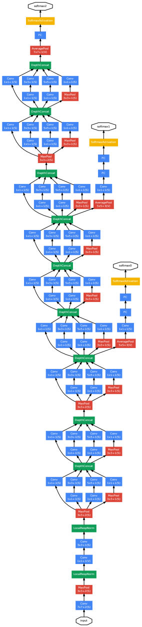
(leCun et al., 1998)

AlexNet



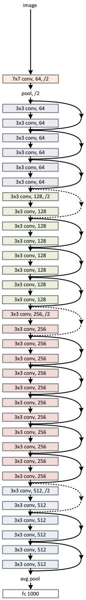
(Krizhevsky et al., 2012)

GoogLeNet



(Szegedy et al., 2015)

Resnet



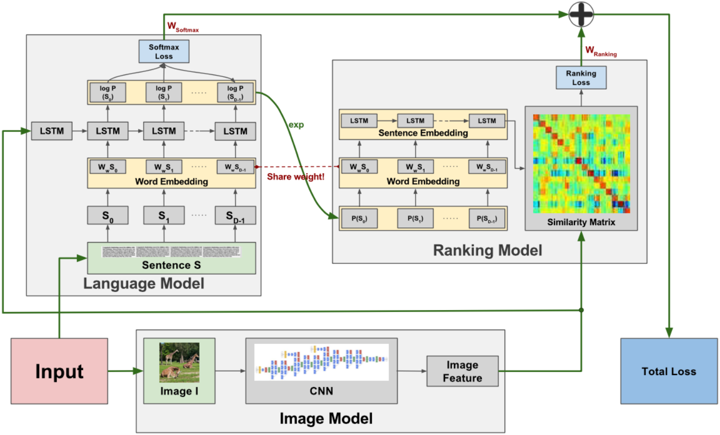
(He et al., 2015)

Deep learning is built on a natural generalization of a neural network: **a graph of tensor operators**, taking advantage of

- the chain rule (aka "back-propagation"),
- stochastic gradient decent,
- convolutions,
- parallel operations on GPUs.

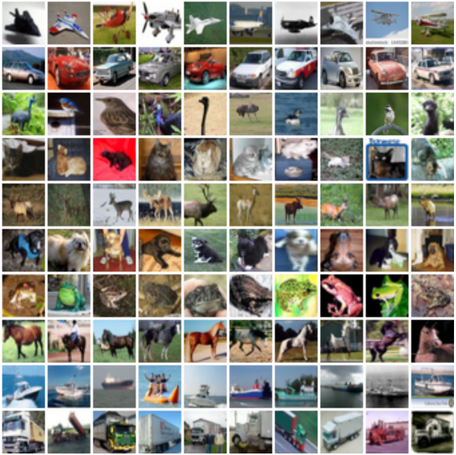
This does not differ much from networks from the 90s

This generalization allows to design complex networks of operators dealing with images, sound, text, sequences, etc. and to train them end-to-end.



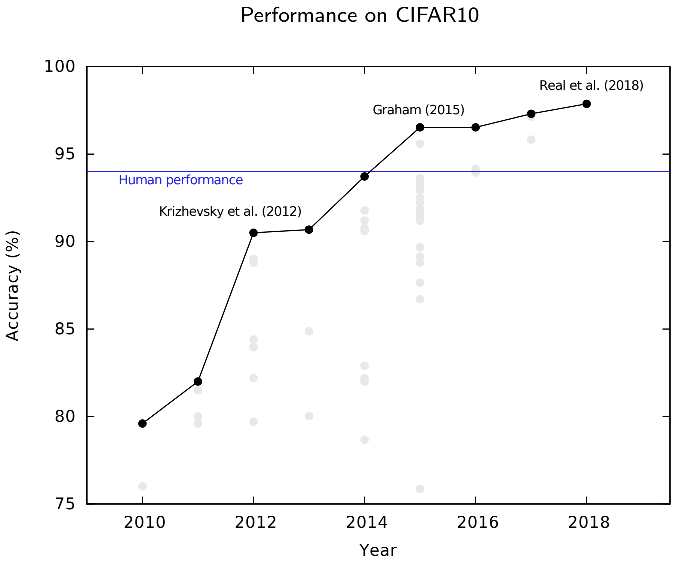
(Yeung et al., 2015)

CIFAR10



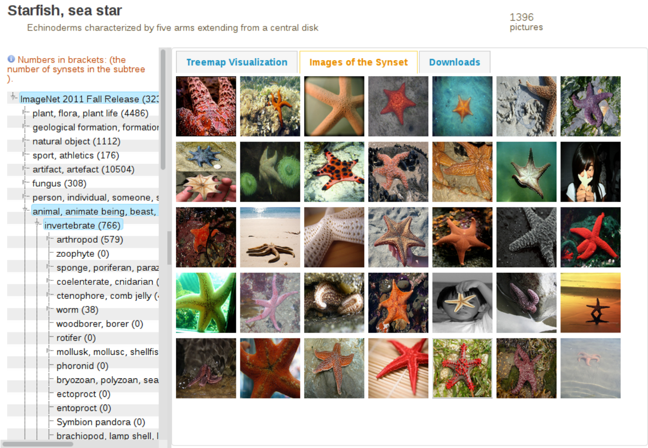
32×32 color images, 50k train samples, 10k test samples.

(Krizhevsky, 2009, chap. 3)


Performance on CIFAR10
100
95
90
85
80
75
2010
2012
2014
2016
2018
Year
Accuracy (%)
Human performance
Krizhevsky et al. (2012)
Graham (2015)
Real et al. (2018)

ImageNet Large Scale Visual Recognition Challenge.

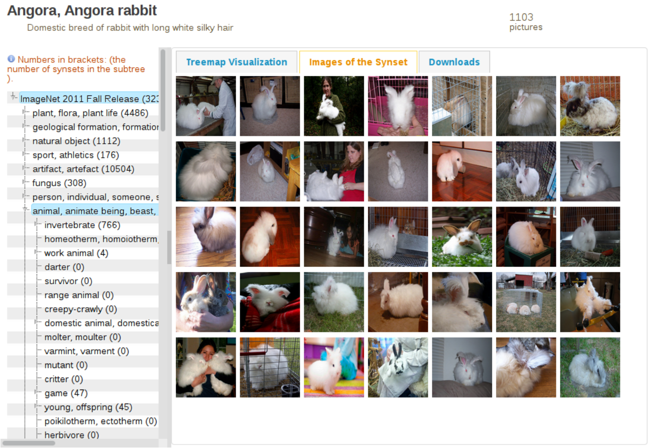
1000 categories, $>$ 1M images



(http://image-net.org/challenges/LSVRC/2014/browse-synsets)

ImageNet Large Scale Visual Recognition Challenge.

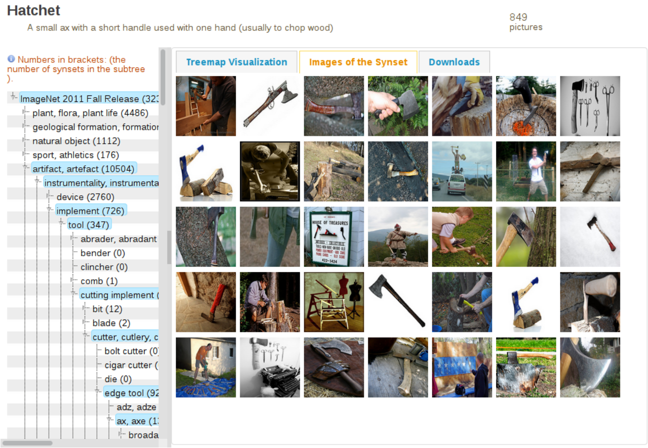
1000 categories, > 1M images



(http://image-net.org/challenges/LSVRC/2014/browse-synsets)

ImageNet Large Scale Visual Recognition Challenge.

1000 categories, $>$ 1M images



(http://image-net.org/challenges/LSVRC/2014/browse-synsets)

method	top-1 err.	top-5 err.
VGG [41] (ILSVRC'14)	-	8.43†
GoogLeNet [44] (ILSVRC'14)	-	7.89
VGG [41] (v5)	24.4	7.1
PReLU-net [13]	21.59	5.71
BN-inception [16]	21.99	5.81
ResNet-34 B	21.84	5.71
ResNet-34 C	21.53	5.60
ResNet-50	20.74	5.25
ResNet-101	19.87	4.60
ResNet-152	**19.38**	**4.49**

Table 4. Error rates (%) of **single-model** results on the ImageNet validation set (except † reported on the test set).

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.

(He et al., 2015)

The end

References

K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4): 193–202, April 1980.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

A. Krizhevsky. Learning multiple layers of features from tiny images. Master's thesis, Department of Computer Science, University of Toronto, 2009.

A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *Neural Information Processing Systems (NIPS)*, 2012.

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

W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

D. E. Rumelhart, G. E. Hinton, and R. J. Williams. *Neurocomputing: Foundations of Research*, chapter Learning Representations by Back-propagating Errors, pages 696–699. MIT Press, 1988.

C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

S. Yeung, O. Russakovsky, G. Mori, and L. Fei-Fei. End-to-end learning of action detection from frame glimpses in videos. *CoRR*, abs/1511.06984, 2015.