Deep learning

1.1. From neural networks to deep learning

François Fleuret

https://fleuret.org/dlc/

UNIVERSITÉ DE GENÈVE

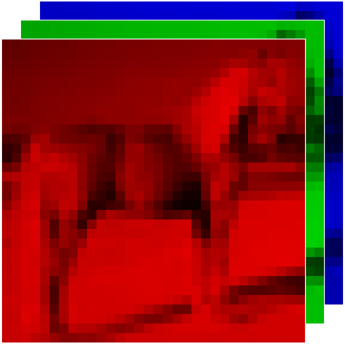
Many applications require the automatic extraction of "refined" information from raw signal (e.g. image recognition, automatic speech processing, natural language processing, robotic control, geometry reconstruction).

(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.datasets import CIFAR10
>>> cifar = CIFAR10('./data/cifar10/', train=True, download=True)
Files already downloaded and verified
>>> x = torch.from_numpy(cifar.data)[43].permute(2, 0, 1)
>>> 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.

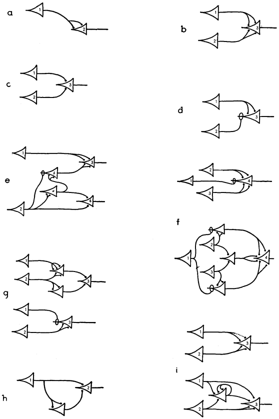
There are strong connections between standard statistical modeling and machine learning.

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

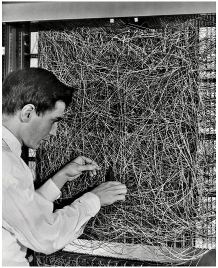
"Artificial neural networks" pre-dated these approaches, and do not follow this dichotomy. They consist of "deep" stacks of parametrized processing.

From artificial neural networks to "Deep Learning"

Networks of "Threshold Logic Unit"



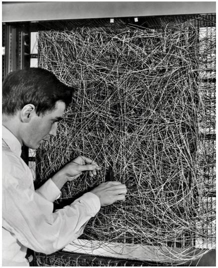
(McCulloch and Pitts, 1943)



Frank Rosenblatt working on the Mark I perceptron (1956)

1949 – Donald Hebb proposes the Hebbian Learning principle (Hebb, 1949).

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

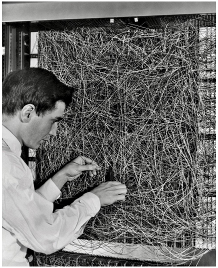


Frank Rosenblatt working on the Mark I perceptron (1956)

1949 – Donald Hebb proposes the Hebbian Learning principle (Hebb, 1949).

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

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



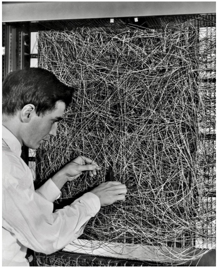
Frank Rosenblatt working on the Mark I perceptron (1956)

1949 – Donald Hebb proposes the Hebbian Learning principle (Hebb, 1949).

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 (Hubel and Wiesel, 1962).



Frank Rosenblatt working on the Mark I perceptron (1956)

1949 – Donald Hebb proposes the Hebbian Learning principle (Hebb, 1949).

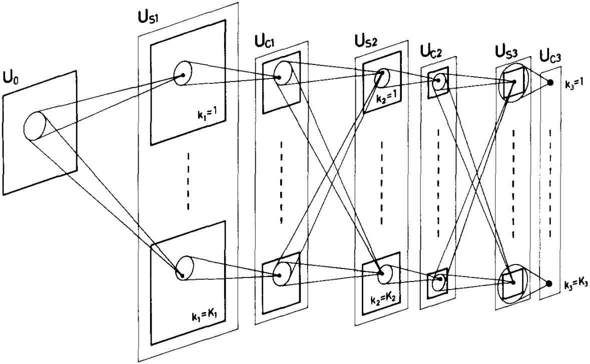
1951 – Marvin Minsky creates the first ANN (Hebbian learning, 40 neurons).

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 (Hubel and Wiesel, 1962).

1982 – Paul Werbos proposes back-propagation for ANNs (Werbos, 1981).

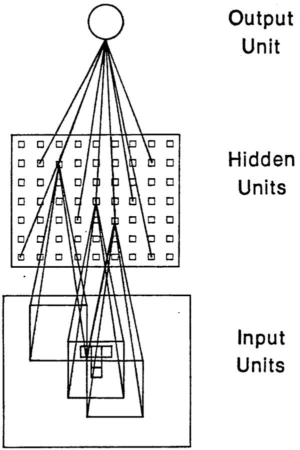
Neocognitron



(Fukushima, 1980)

This model follows Hubel and Wiesel's results.

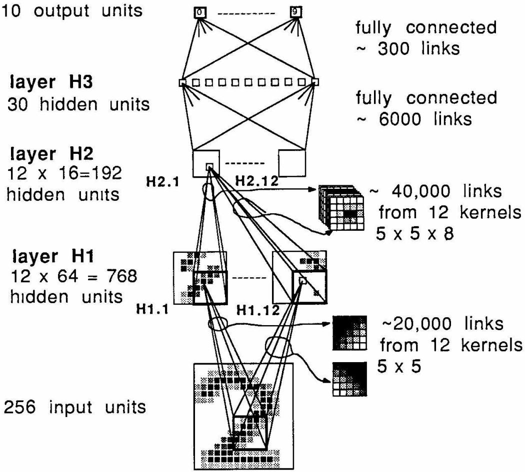
Network for the T-C problem



Trained with back-prop.

(Rumelhart et al., 1988)

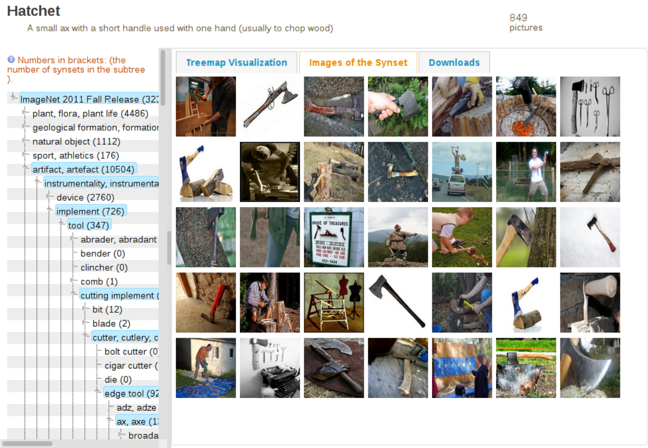
LeNet family



(LeCun et al., 1989)

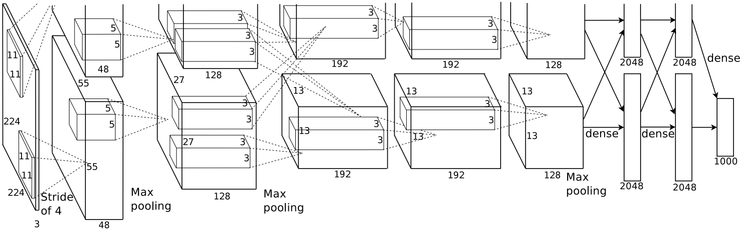
ImageNet Large Scale Visual Recognition Challenge.

Started 2010, 1 million images, 1000 categories



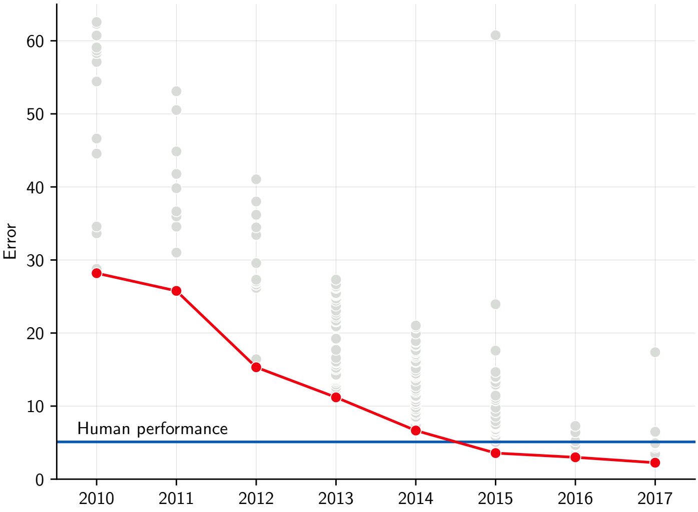
(http://image-net.org/challenges/LSVRC/2014/browse-synsets)

AlexNet

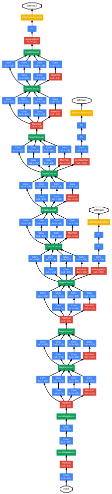


(Krizhevsky et al., 2012)

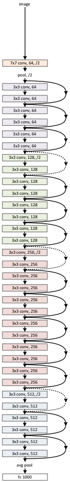
Top-5 error rate on ImageNet



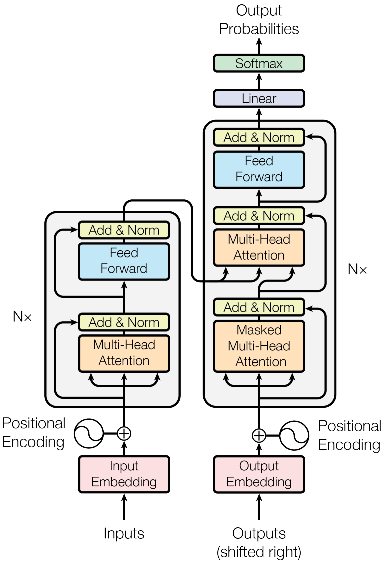
(Gershgorn, 2017)



GoogleNet (Szegedy et al., 2015)



ResNet (He et al., 2015)



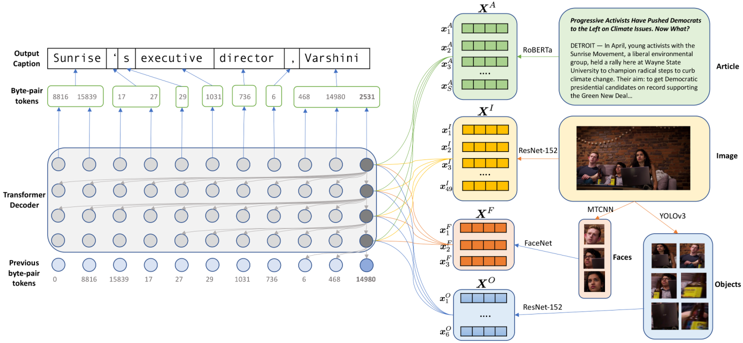
(Vaswani et al., 2017)

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.



(Tran et al., 2020)

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.

D. Gershgorn. **The data that transformed AI research—and possibly the world**, July 2017.

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

D. O. Hebb. The organization of behavior: A neuropsychological theory. Wiley, 1949. ISBN 0-8058-4300-0.

D. Hubel and T. Wiesel. **Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex**. Journal of Physiology, 160:106–154, 1962.

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

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

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.

A. Tran, A. Mathews, and L. Xie. **Transform and tell: Entity-aware news image captioning**. In Conference on Computer Vision and Pattern Recognition (CVPR), pages 13035–13045, 2020.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin. **Attention is all you need**. CoRR, abs/1706.03762, 2017.

P. J. Werbos. **Applications of advances in nonlinear sensitivity analysis**. In Proceedings of the 10th IFIP Conference, pages 762–770, 1981.