



Machine Learning

Chapter 10: Deep Learning

Fall 2021

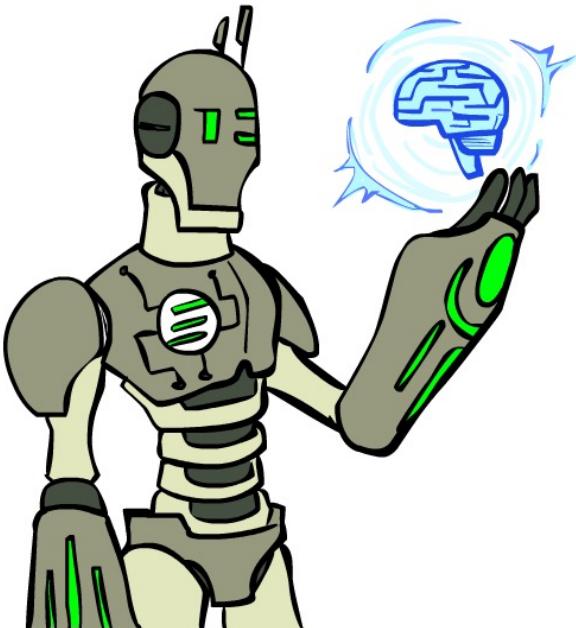
Instructor: Xiaodong Gu



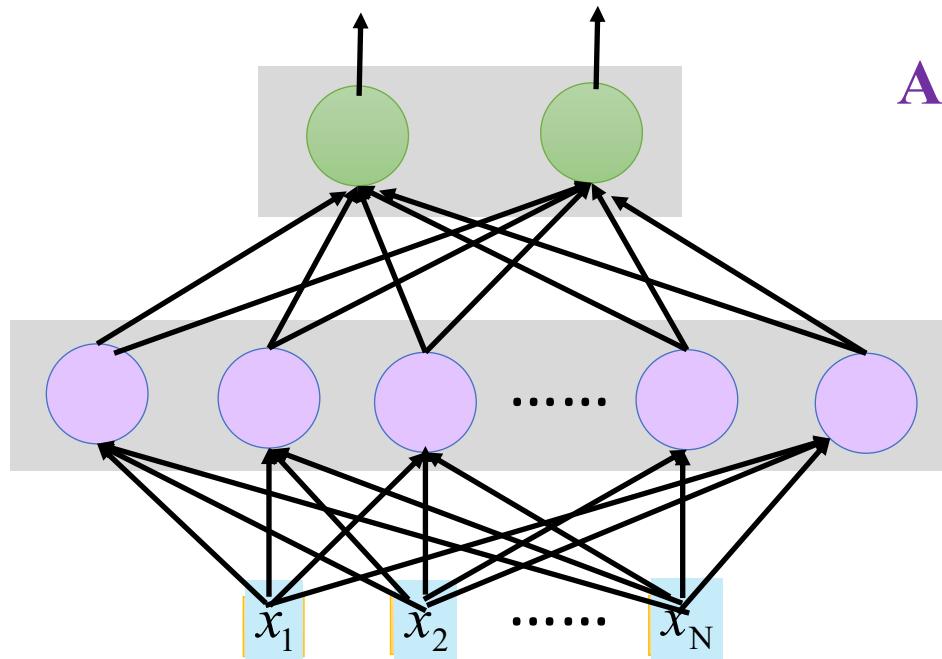
Today



- Deep Learning



Recall: Neural Networks



Any continuous function f

$$f : R^N \rightarrow R^M$$

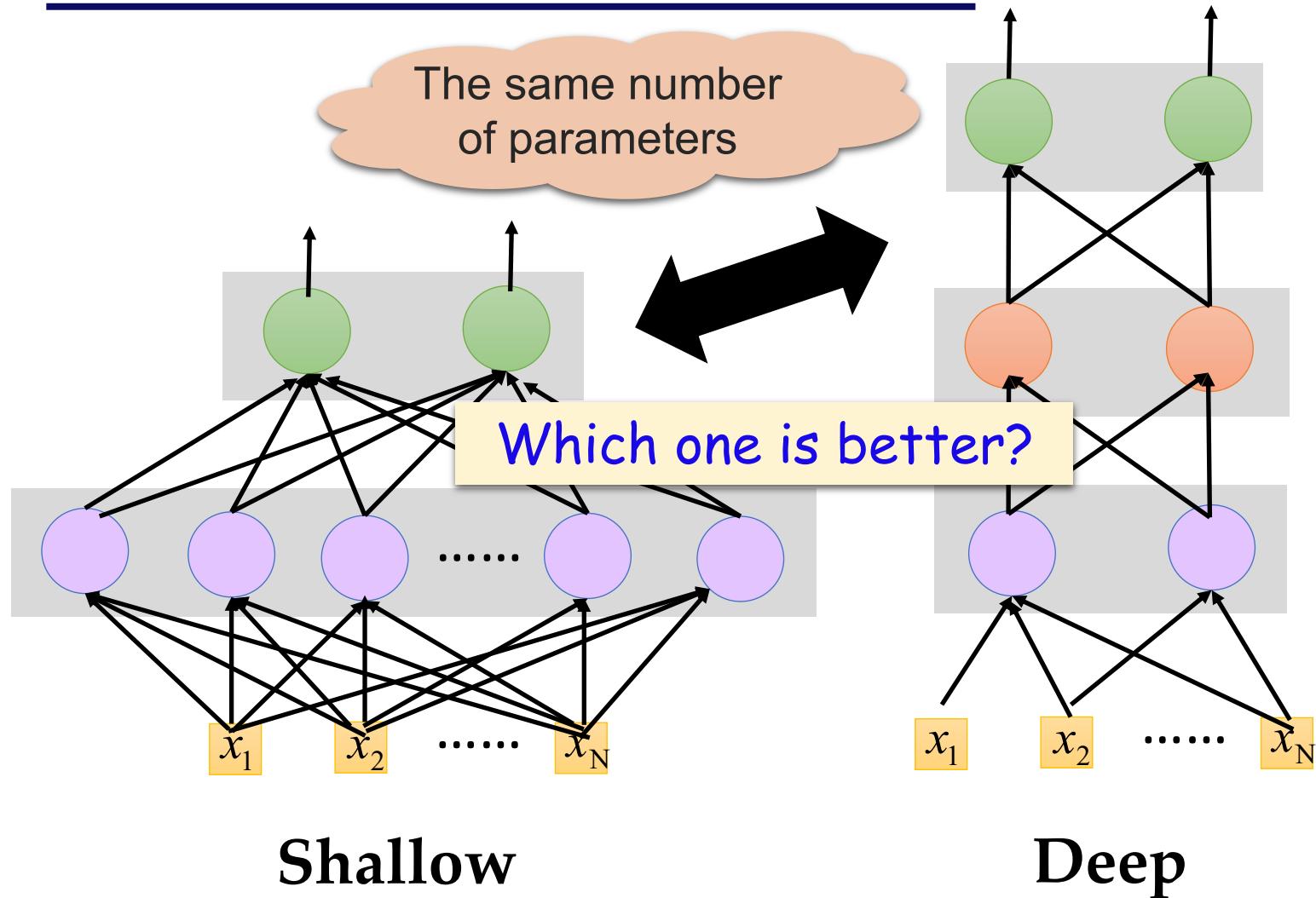
can be realized by an NN
with one hidden layer.

(given **enough** hidden neurons)

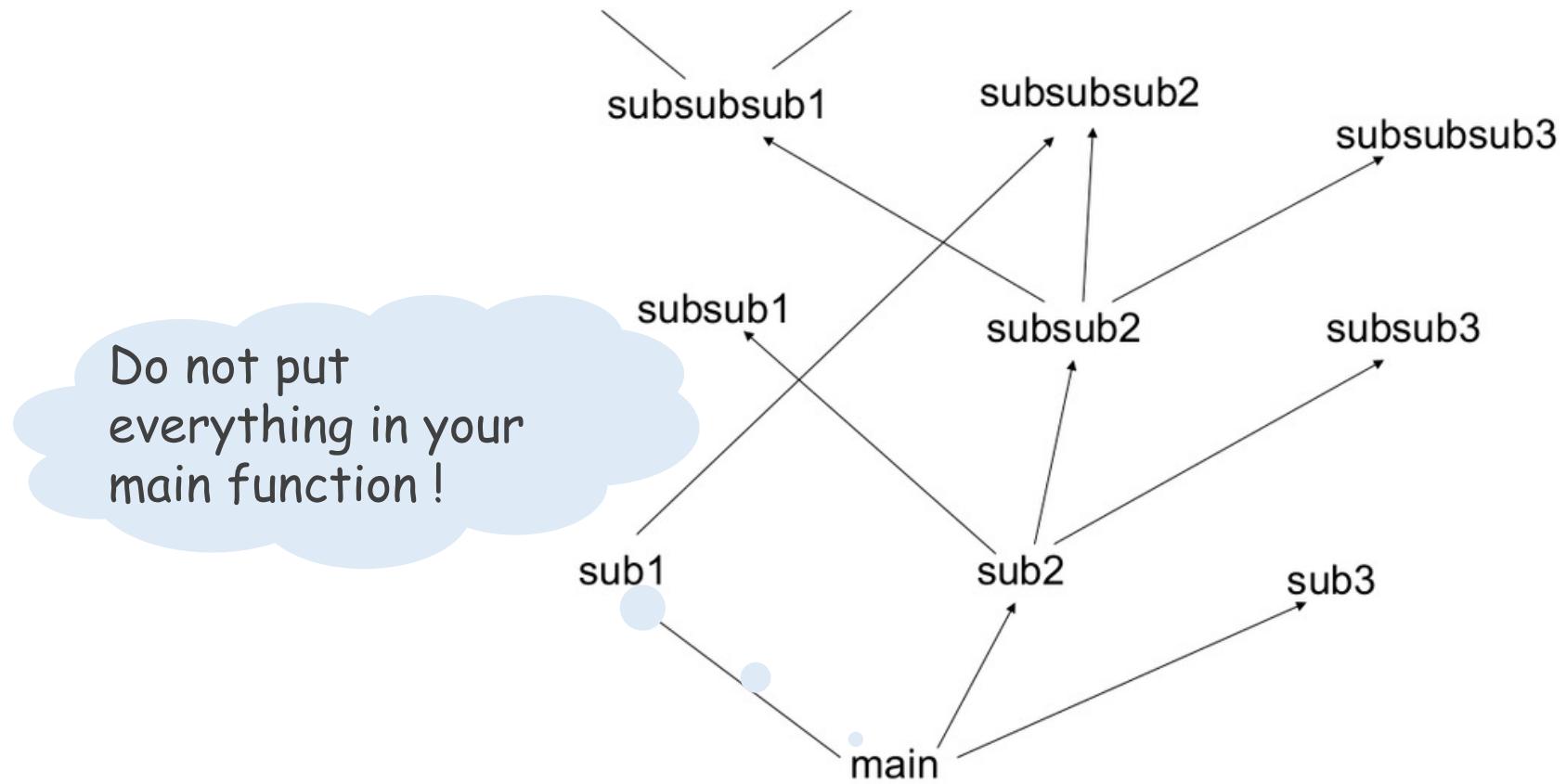
NNs with one hidden layer
can represent **every** function.

So?

The Deeper, The Better?

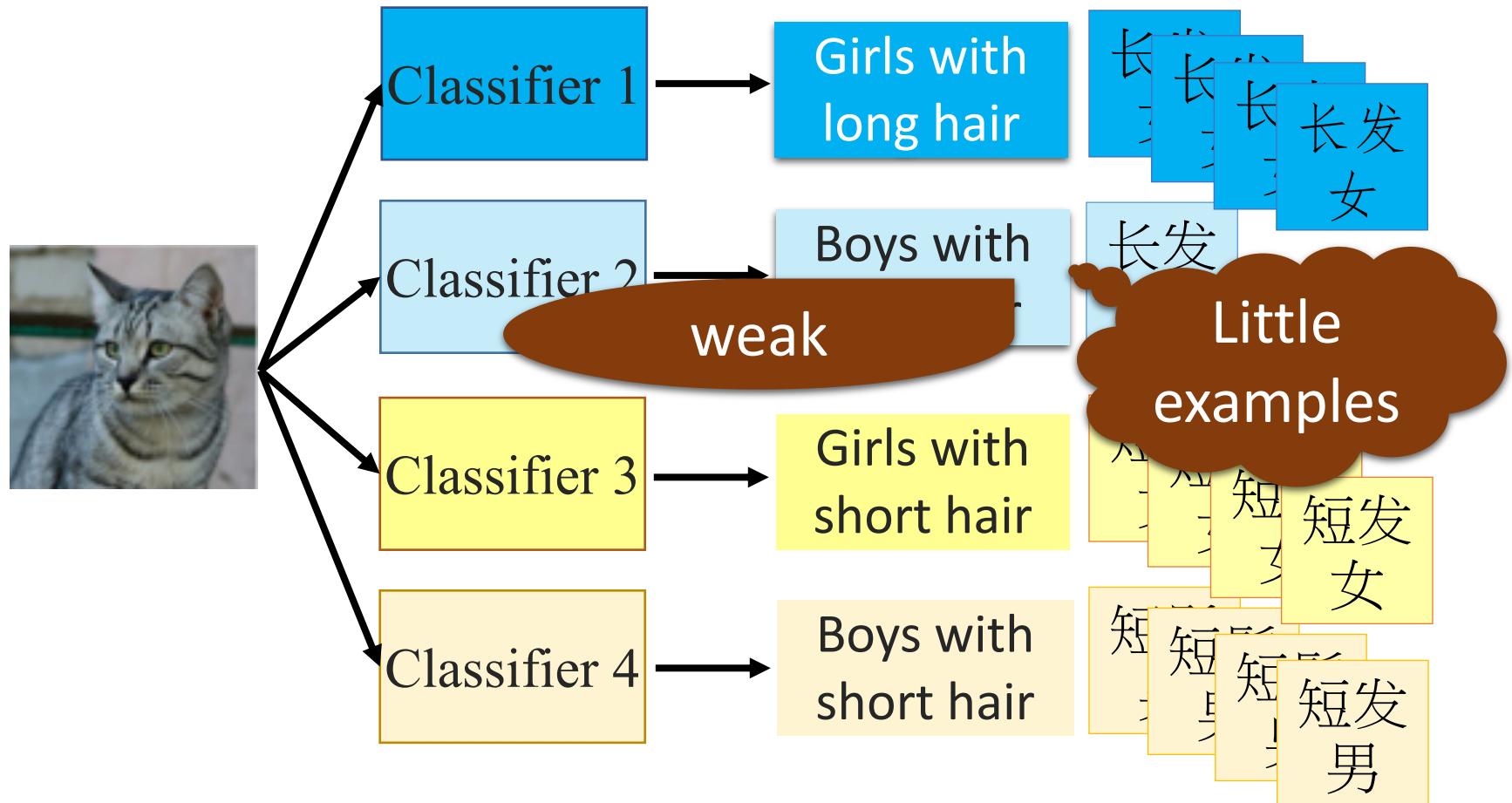


Idea: Modularization



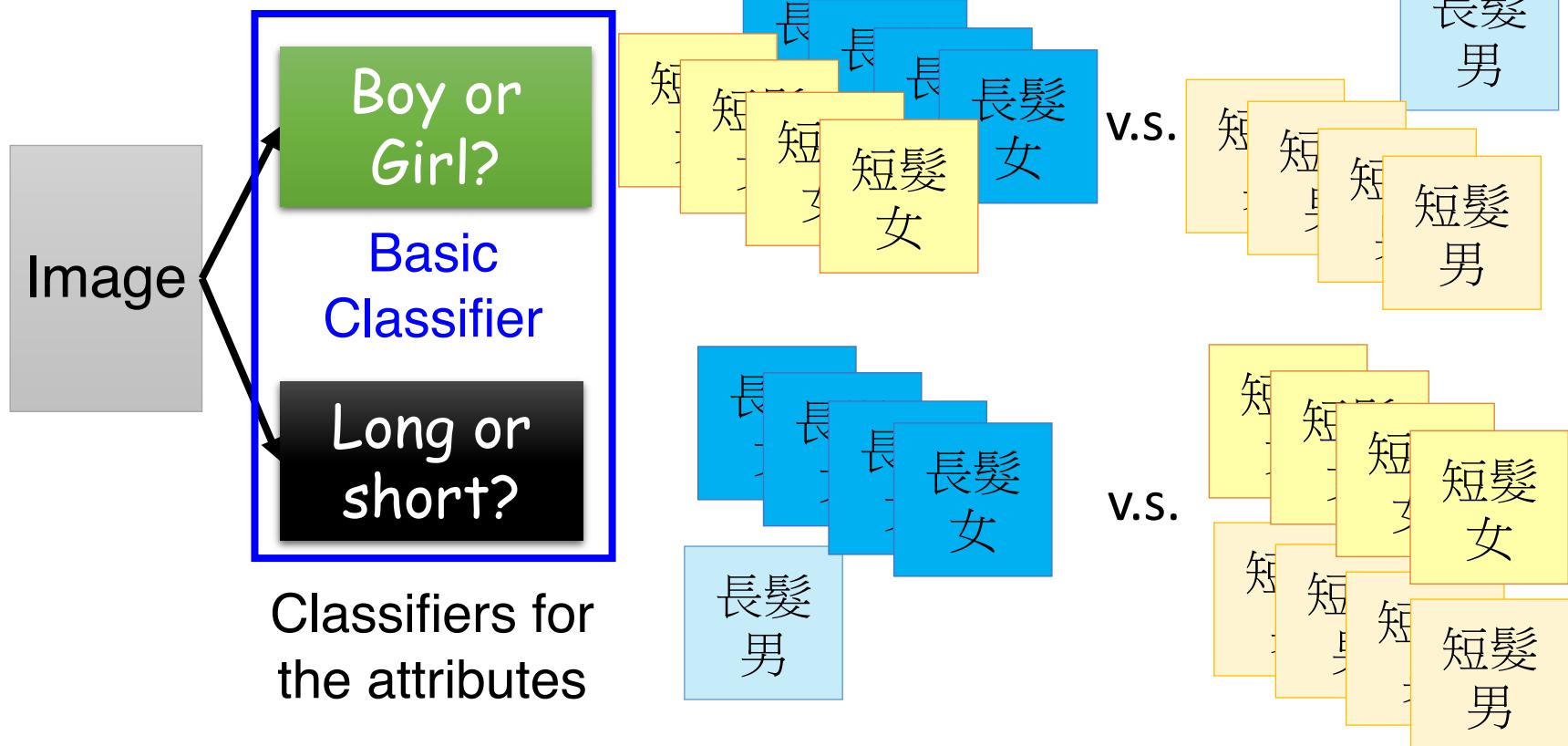
<http://rinuboney.github.io/2015/10/18/theoretical-motivations-deep-learning.html>

Idea: Modularization



Idea: Modularization

Each basic classifier can have sufficient training examples.

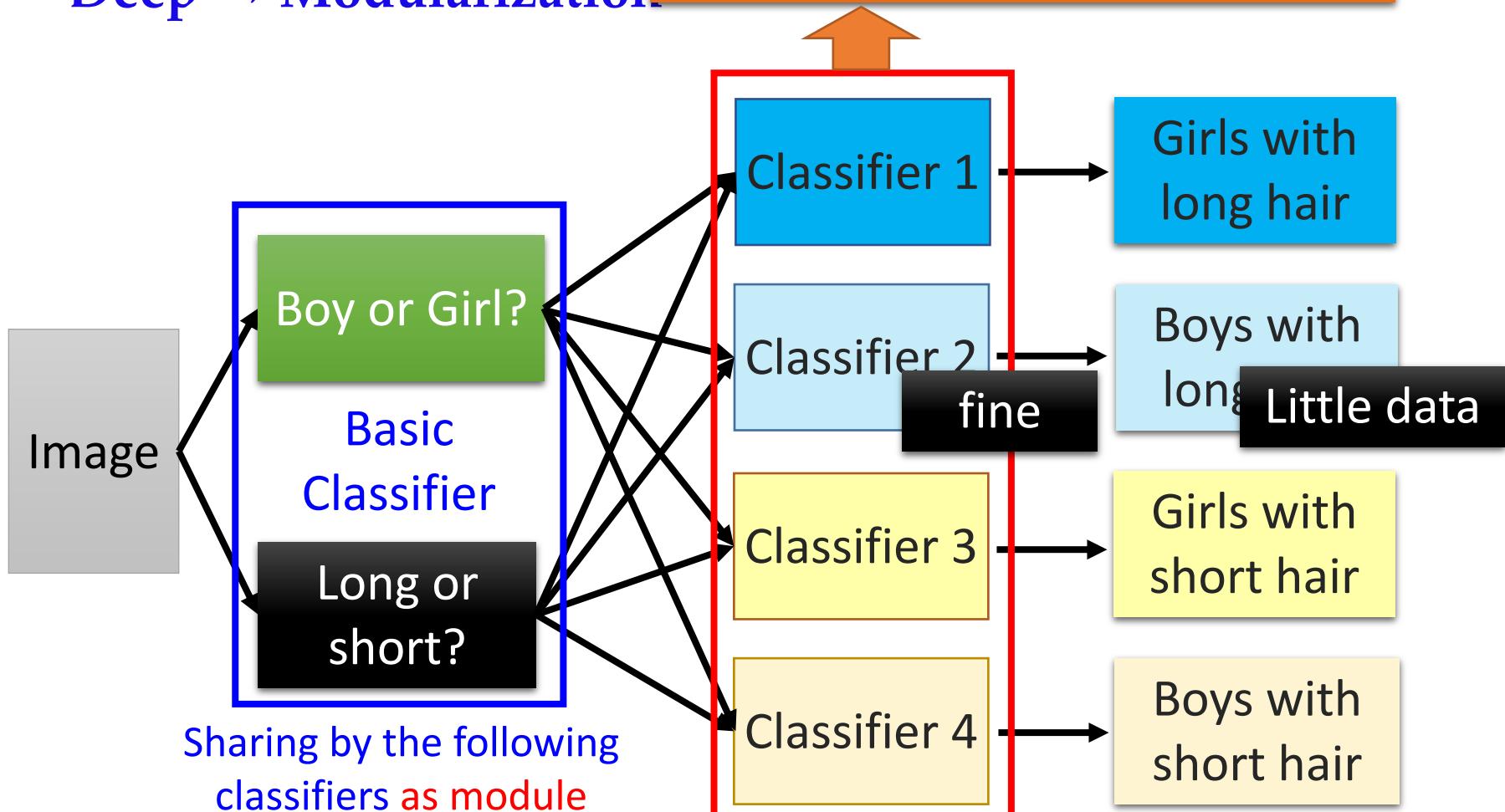


Idea: Modularization



- ## • Deep → Modularization

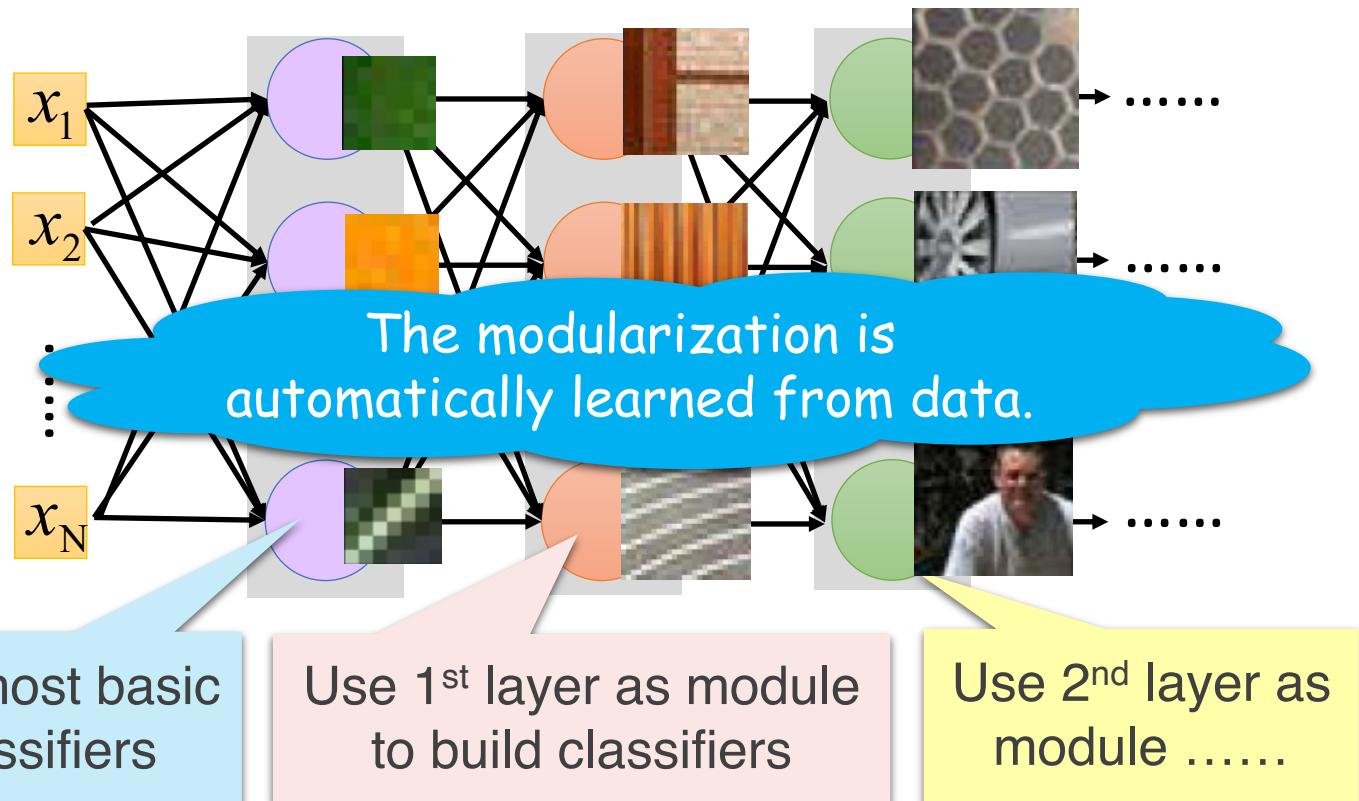
can be trained by little data



Modularization - Image



- Deep → Modularization → Less training data?

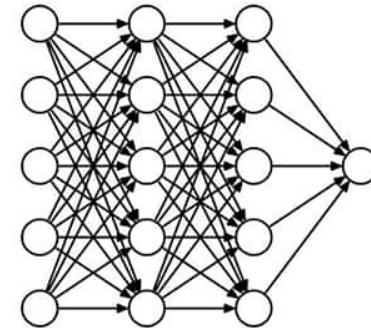


Zeiler, M. D., & Fergus, R. Visualizing and understanding convolutional networks. In *ECCV 2014*

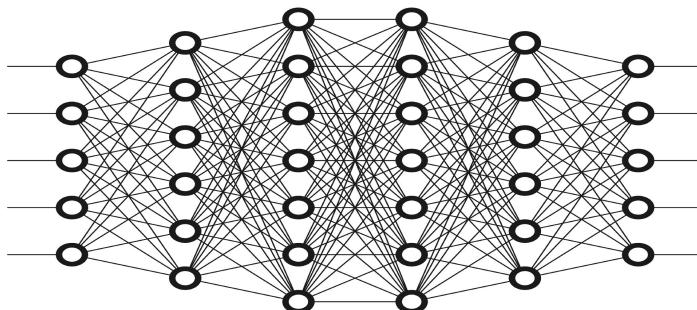
The Deeper, The Better



So, shallow network can represent any functions.



However, using deep structures is more effective.



Reference for the reason: <http://neuralnetworksanddeeplearning.com/chap4.html>



Deep Learning

What is Deep Learning?



- Deep learning (deep machine learning, or deep structured learning, or hierarchical learning, or sometimes DL) is a branch of machine learning based on a set of algorithms that attempt to model high-level abstractions in data by using model architectures, with complex structures or otherwise, composed of multiple non-linear transformations.

What is Deep Learning ?



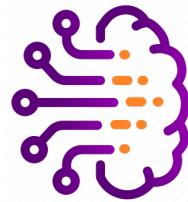
Artificial Intelligence

Computer systems that perform tasks that would usually require human intelligence



Machine Learning

Statistical techniques that learn from a series of inputs and outputs



Deep Learning

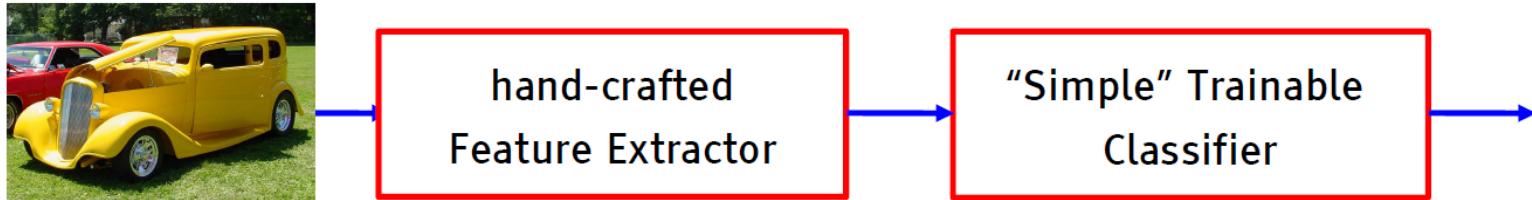


Machine learning algorithms that enable self learning of data representations

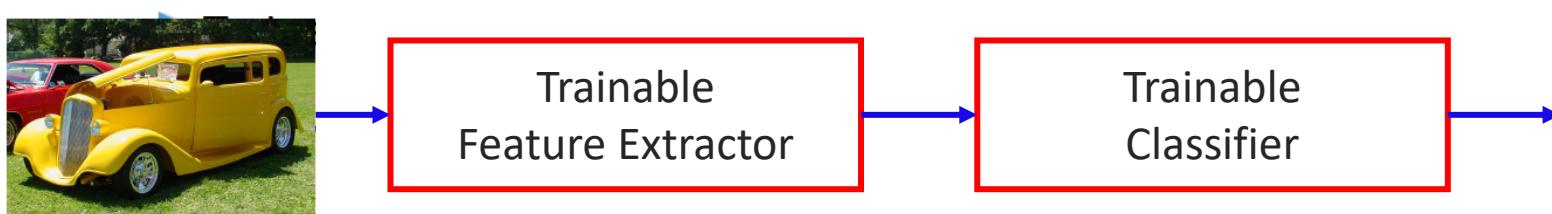
Deep Learning = Learning Representations/Features of Data



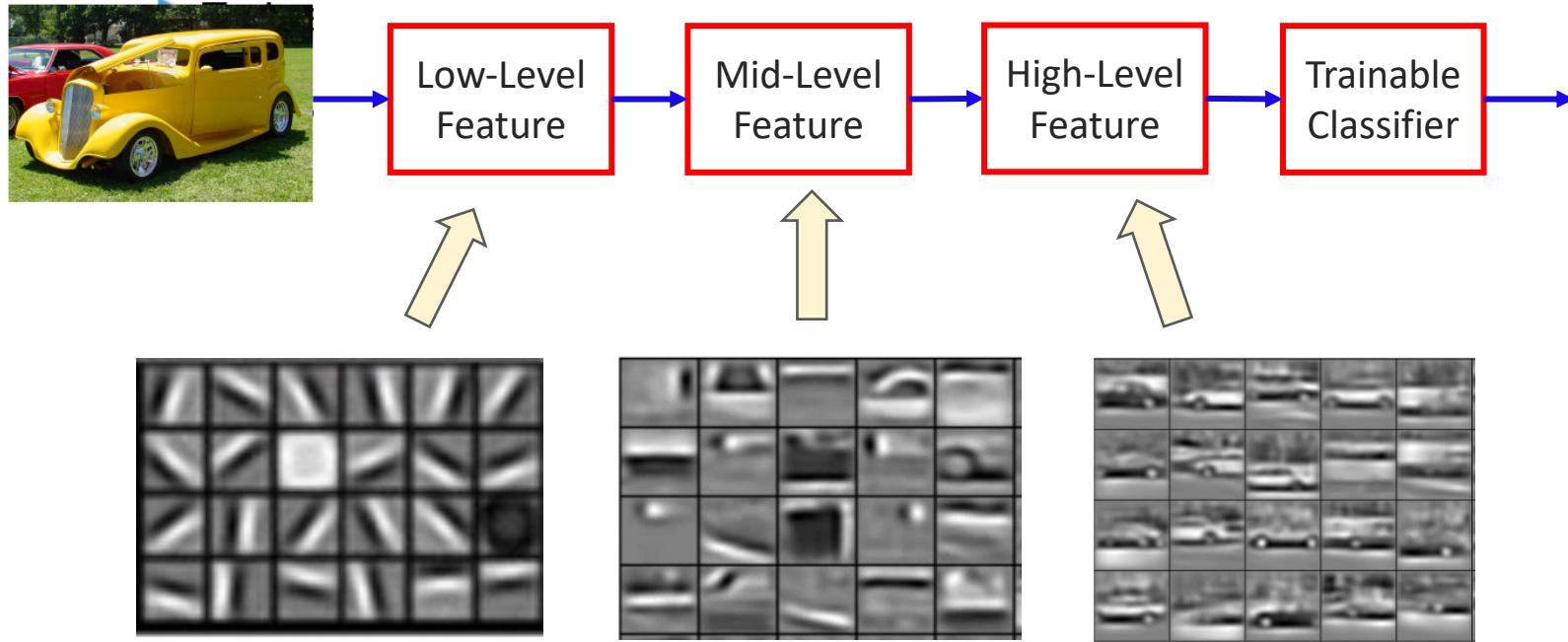
- The traditional model of pattern recognition (since the late 50's)
 - fixed/engineered features + **trainable classifier**



- Deep Learning/ Feature learning / End-to-end Learning
 - **trainable** features + **trainable** classifier



Deep Learning = Learning Hierarchical Representations



Hierarchical Representations with Increasing Level of Abstraction



- **Image recognition**

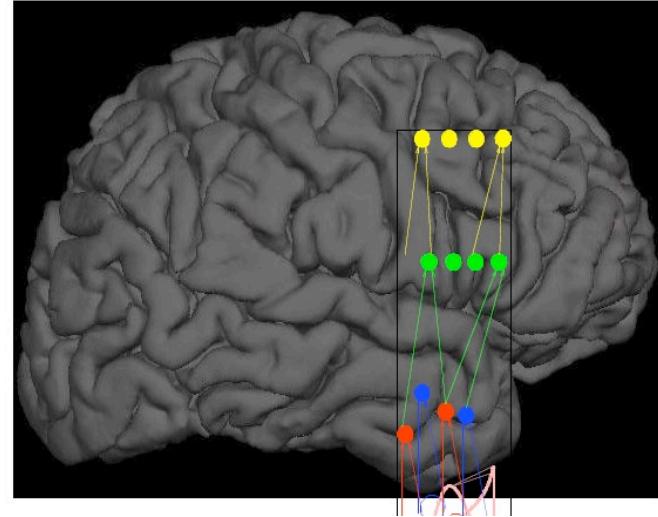
pixel → edge → texton → motif → part
→ object

- **Speech**

sample → spectral band → sound
→ phone → word...

- **Text**

character → word → phrase → clause
→ sentence → paragraph → document



Human brains also has a deep architecture.

Humans first learn simpler concepts and then compose them into complicated ones.

From concepts to consciousness?



- The worlds of different species.

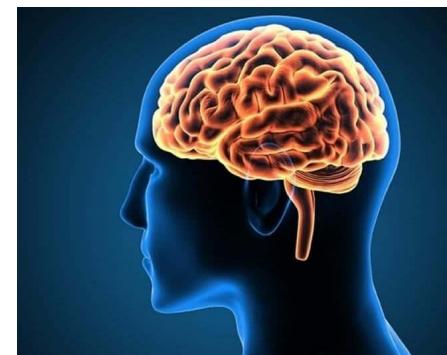
leaves,
wetness,...



self-awareness,
color, taste,
movement



happiness, sorrow,
anxiety, depression



(complexity of neurons)

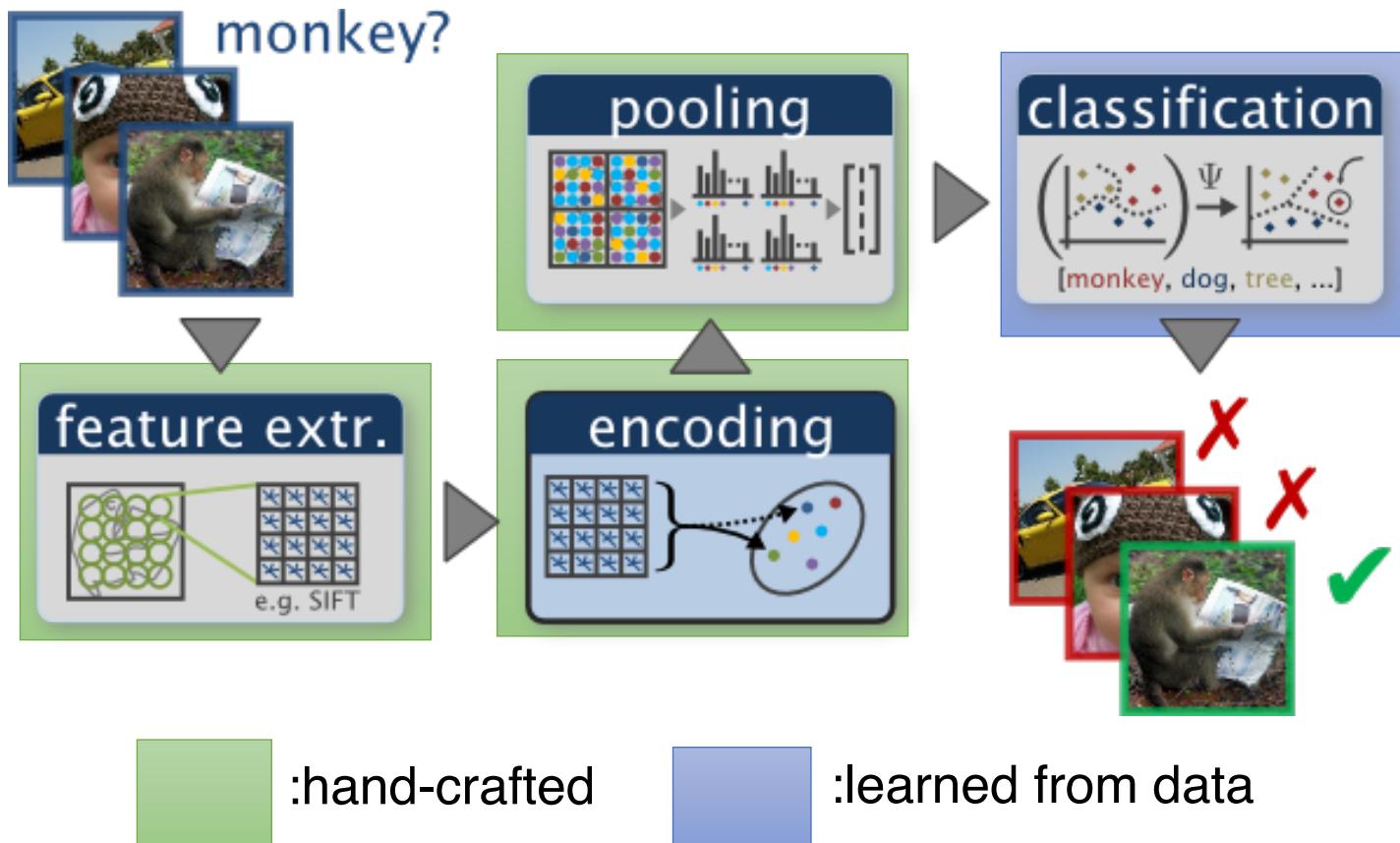
Life is about hierarchical representations

End-to-end Learning



- Shallow Approach

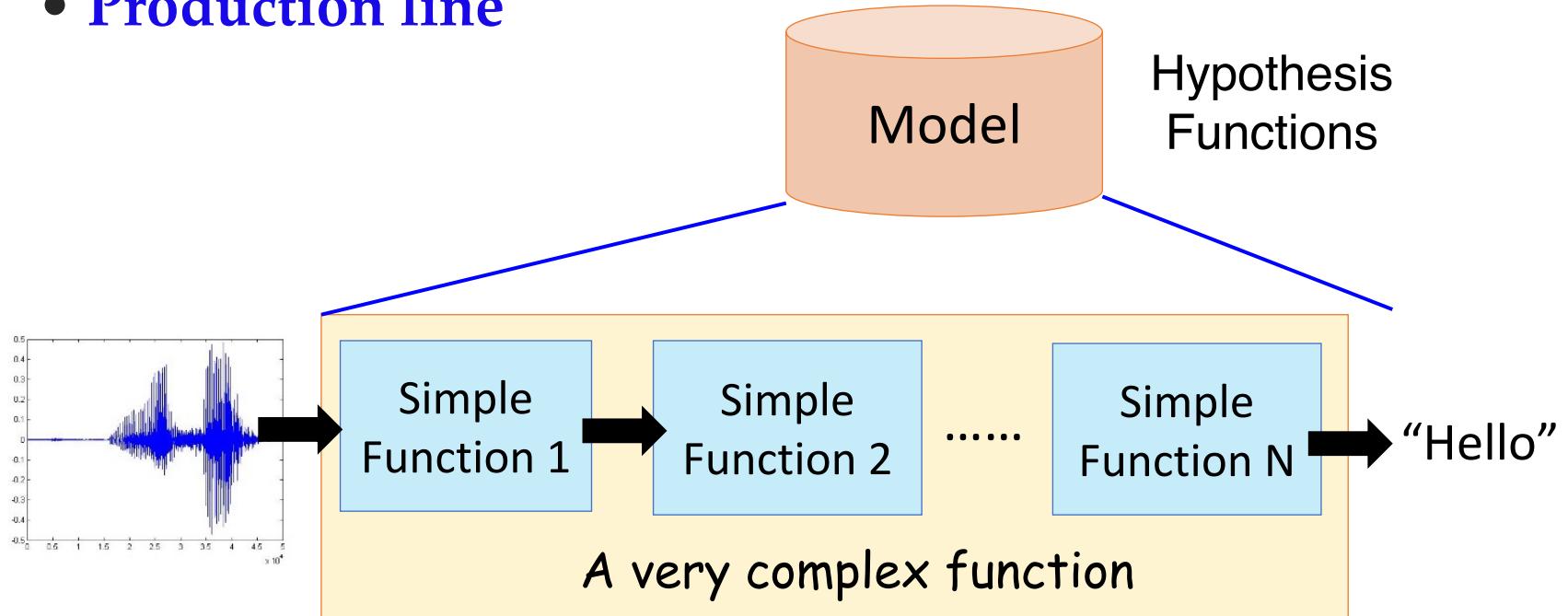
http://www.robots.ox.ac.uk/~vgg/research/encoding_eval/



End-to-end Learning

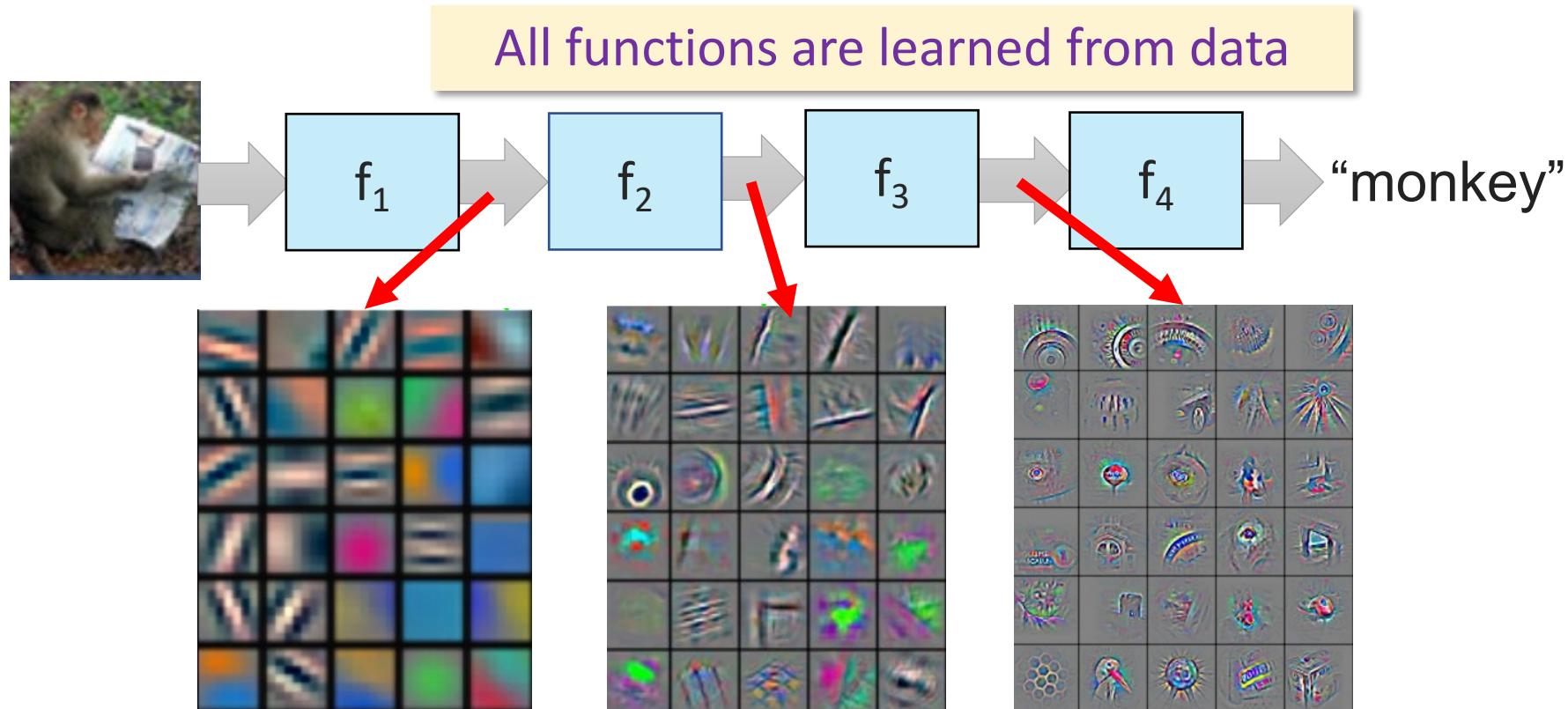


- Production line



End-to-end training: What each function should do is learned automatically.

End-to-end Learning

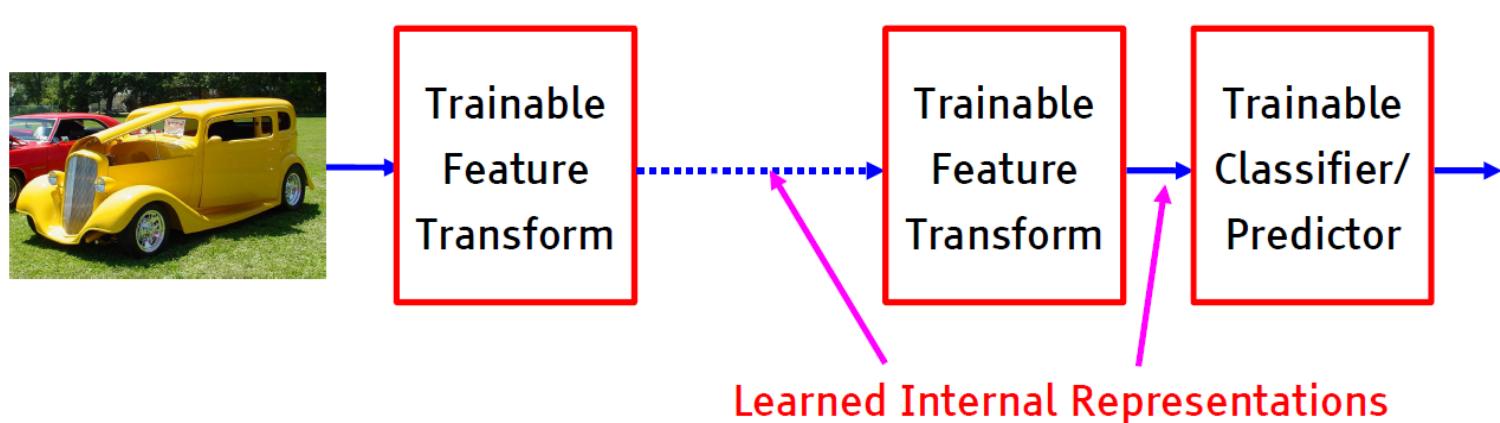


Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014* (pp. 818–833)

End-to-end Learning



- A hierarchy of trainable feature transforms
 - each module transforms its input representations into a higher-level one.
 - high-level features are more global and more invariant
 - low-level features are shared among categories



All in all: Why Deep Learning?



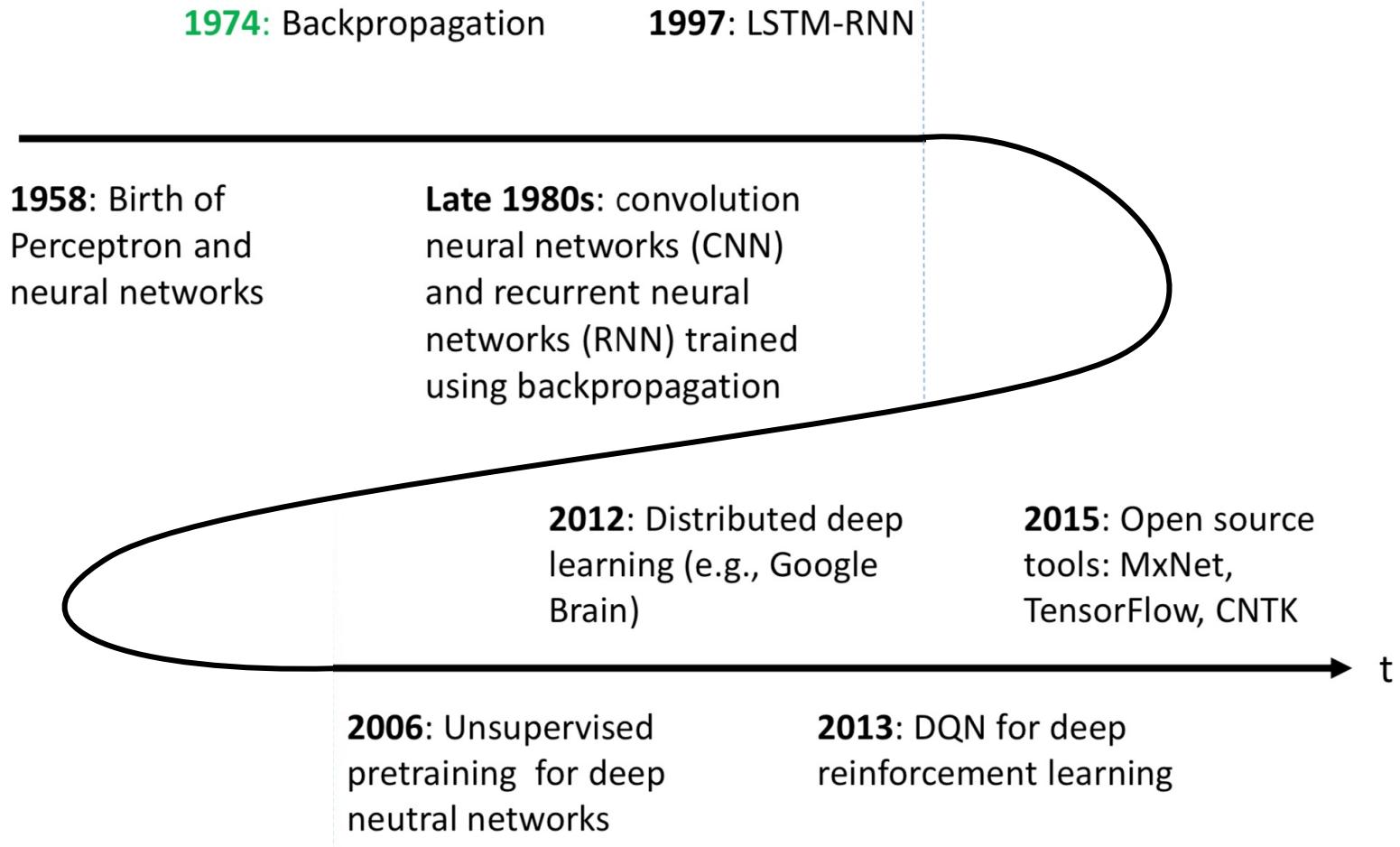
- huge amount of training data – **Big Data**
- sufficient computational power – **Big Machine (GPU and Cloud)**
- highly complicated models – **Big Model**
 - Deep structure reduces the number of parameters while achieving high model complexity
 - Layered structure is very natural

Image: Pixel → edge → texton → motif → part → object

Speech: Sample → spectral band → sound → phone → word...

Text: Character → word → phrase → clause → sentence → paragraph
→ document

History of Deep Learning



Success of Deep Learning



Microsoft Research shows a promising new breakthrough in speech translation technology

Facebook's DeepFace facial recognition technology has human-like accuracy

Join thousands of others, and sign up for TechCrunch's newsletter.

you@exa

by Lee Munson on Facebook June 12, 2014

Posted Nov 12, 2014

This is a guest post by Lee Munson, a software engineer at Facebook.

Facial recognition has been around for many years. The vast majority of people can identify eyes, a mouth and pretty much the same basic recognition features.

Total accuracy, however, has come by - even us positively identifying a **97.53% of the time**.

Certain groups have exceeded that level of

The visualization consists of six small grid-based diagrams labeled a through f. Diagrams a, b, and c show 'Value network' evaluations from a value net, with black dots representing one class and blue dots representing another. Diagrams d, e, and f show 'Policy network' evaluations from a policy net, also with black and blue dots. Colored circles (red, green, blue) highlight specific points of interest or errors in the evaluations.

AlphaGo

GLOBAL EDITION | PAPER | MAGAZINE | BUSINESS REPORTS | LISTS | EVENTS | MORE | CONNECT | LOG IN / JOIN

MIT Technology Review

EmTech MIT
Nov 2-4, 2016
MIT Media Lab Cambridge

VIEW

Emerging Technology From the arXiv

June 12, 2015

Deep Learning Machine Beats Humans in IQ Test

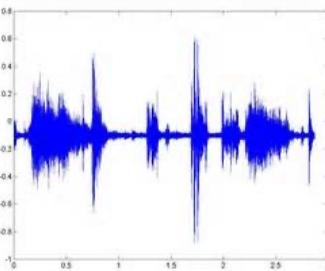
Computers have never been good at answering the type of questions that humans can. Now a deep learning system has done it.

A diagram illustrating a neural network architecture for an IQ test. It shows two input layers, w_{1+1} and w_{2+2} , each with three nodes. These are processed by softmax functions to produce probability distributions. The resulting vectors are then subtracted ($-$) to produce an embedding of w_3 . This embedding is then compared with a target vector v_T .

DONATE STUFF. CREATE JOBS.

FIND YOUR LOCAL GOODWILL

Applications in Various Domains



Speech

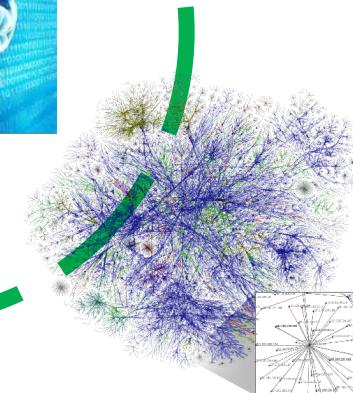


Image

But in a certain number of old time there was of St. Martin
broke out of Kent, which the Romans drove in Britain.
In hoc Ecclesia conuenire primi gallorum, certe millet factum
permit. In this church they began first to assemble them self
They were say Mass to breake, and to baptize. It is plaine that these were
Iulians, and reported to us, who believed, and were of them baptizel. Wondred
folk so
that other
men
was this the
the lands they took with them by commandement of S. Gregorius
the service
of the
gods he
Churches

Here is a great bulle, and no Cowne. If empfe we
then had the here prouice sufficient. Firste, I will traue
particularly by them felues, and in the ende, will shewe
lance, as it may be gathered by Tertullian, Optigen, &
suche other olde writers.

Natural language



Network



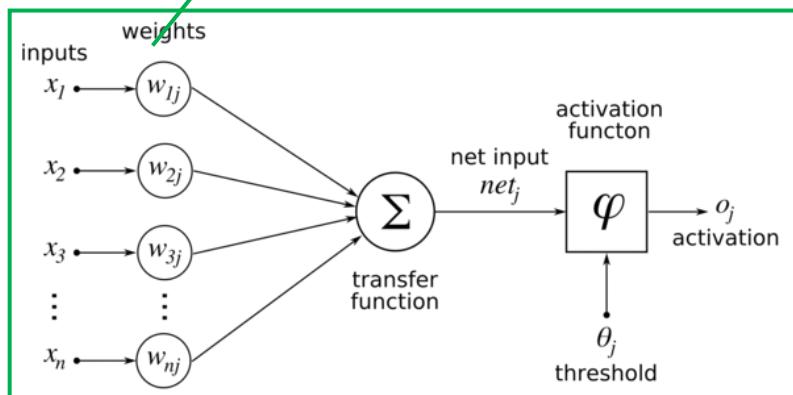
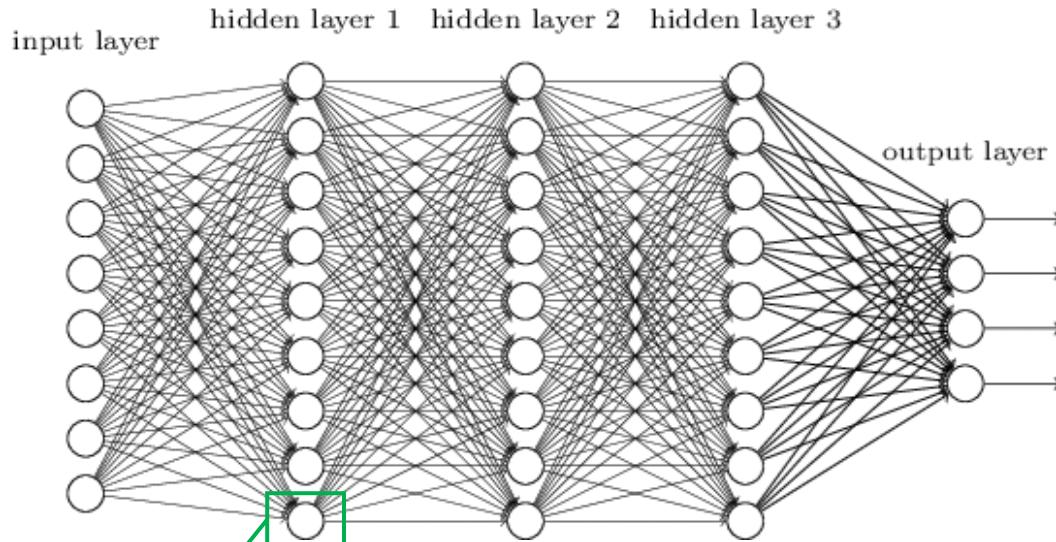
Google
Knowledge Graph

Deep Neural Networks



- Feedforward neural networks
- Convolutional neural networks
- Recurrent neural networks

Deep (Feedforward) Neural Networks

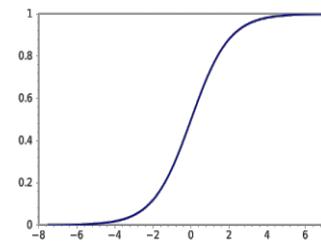


Loss functions:

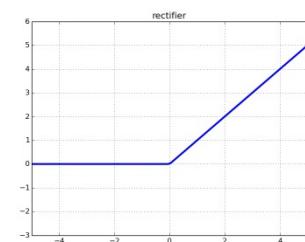
- Square error
- Cross entropy
- Hinge loss
- Ranking loss
- ...

Training:
Backpropagation

Activation functions



Sigmoid



ReLU

Other Structures?



- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks (RNN)
- Transformer
- ...