Machine Learning

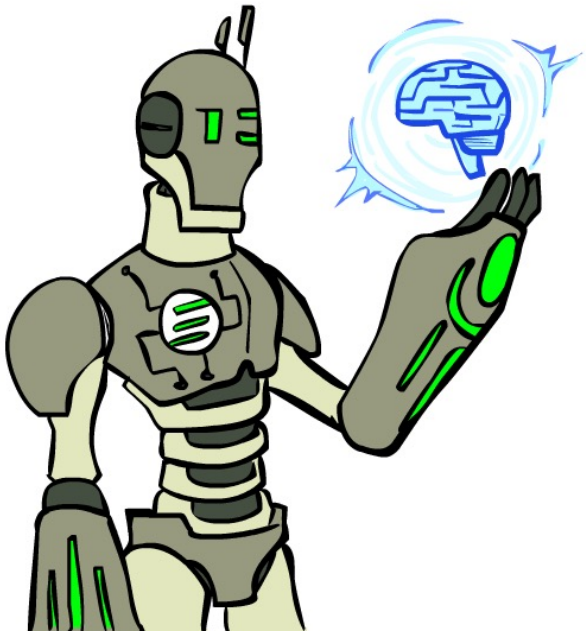
Lecture 9: Deep Learning

Fall 2023

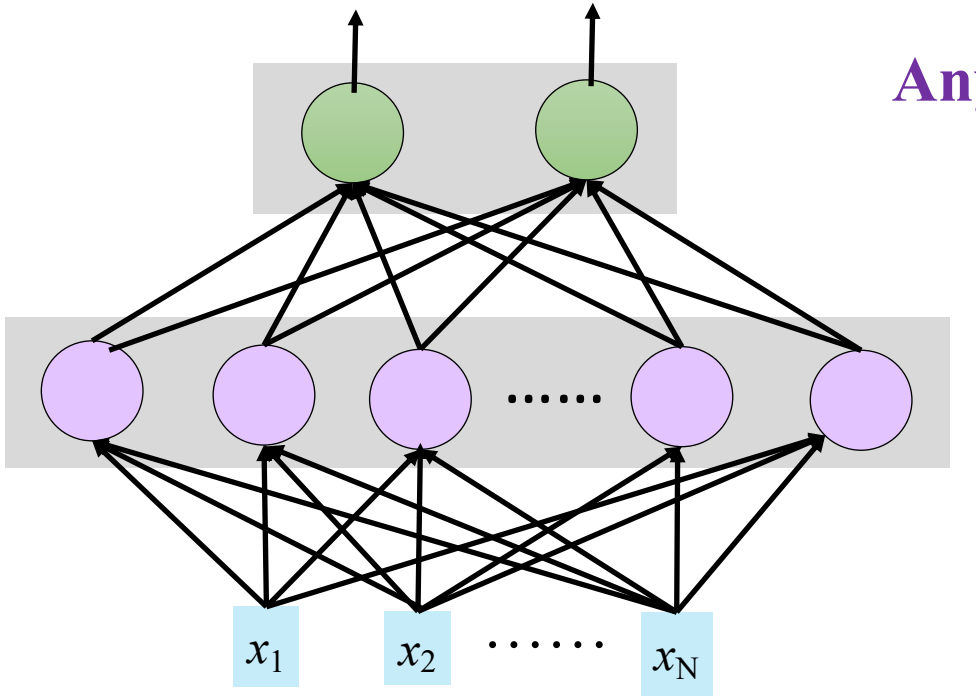
Instructor: Xiaodong Gu

Today

- Introduction to Deep Learning



Recall: Neural Networks

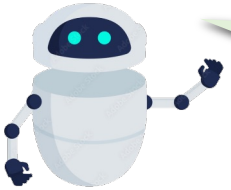


Any continuous function f

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

can be realized by an NN with one hidden layer.

(given **enough** hidden neurons)

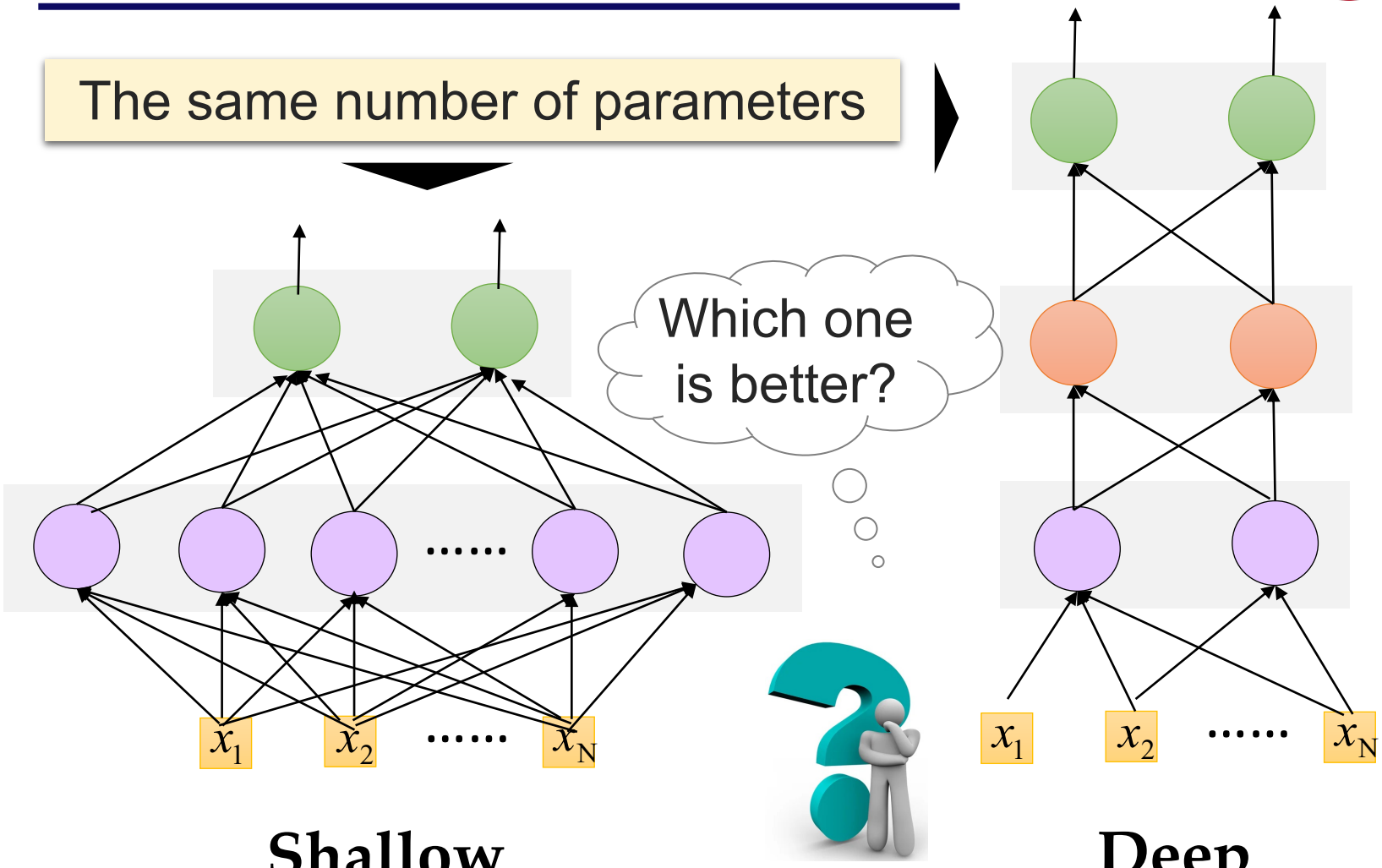


NNs can represent **every** function.

How?

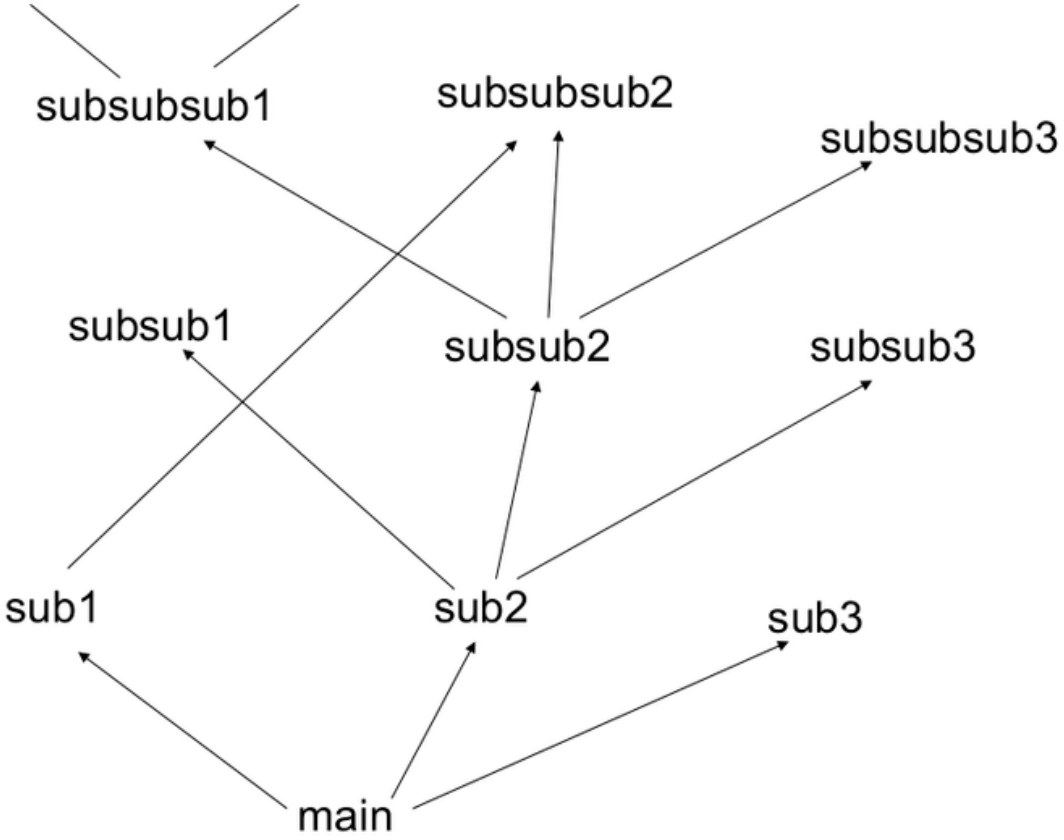
The Deeper, The Better?

The same number of parameters



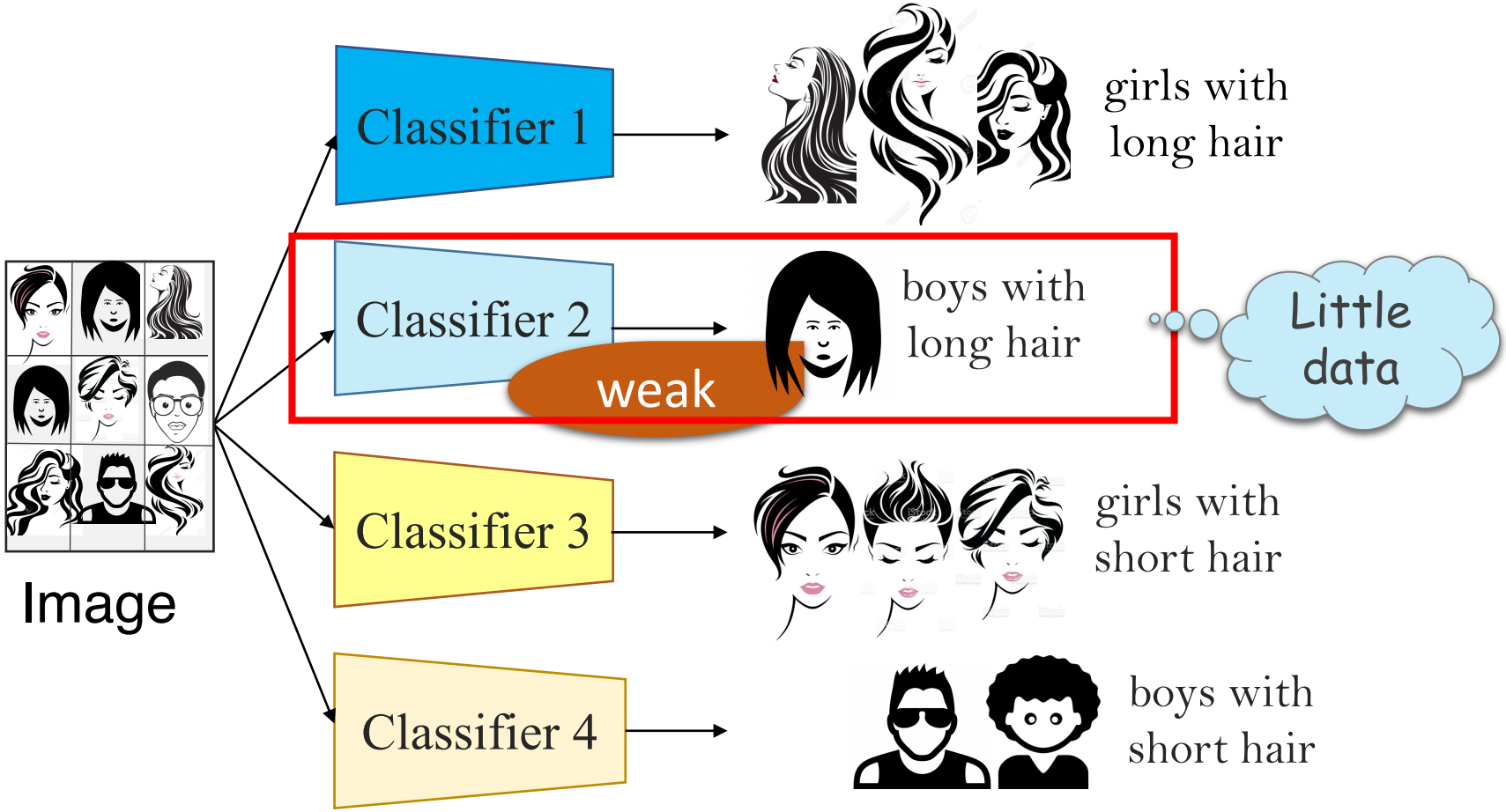
Idea: Modularization

Do not put everything in your main function!

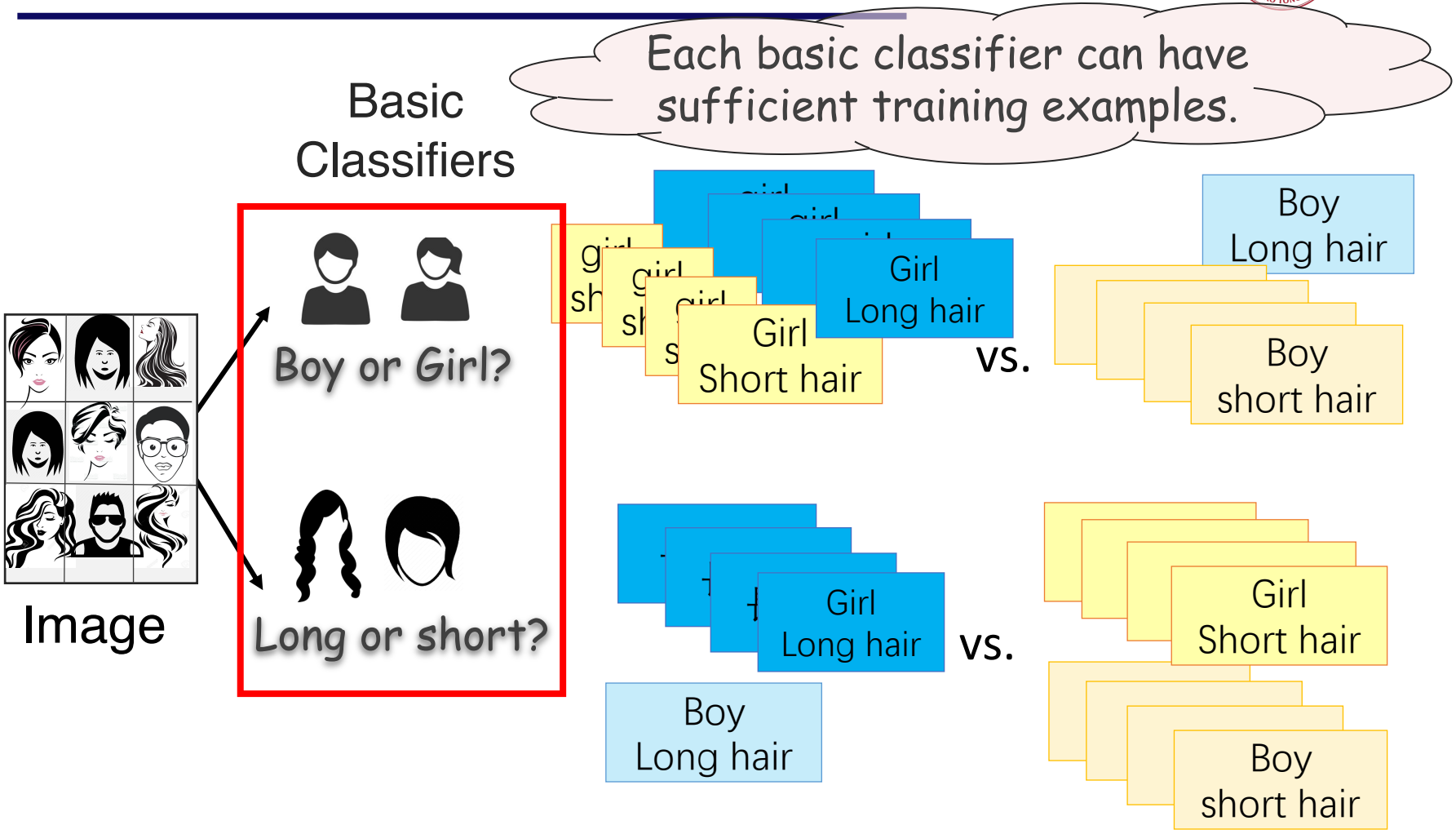


http://rinuboney.github.io/2015/10/18/theoretical-motivations-deep-learning.html

Idea: Modularization

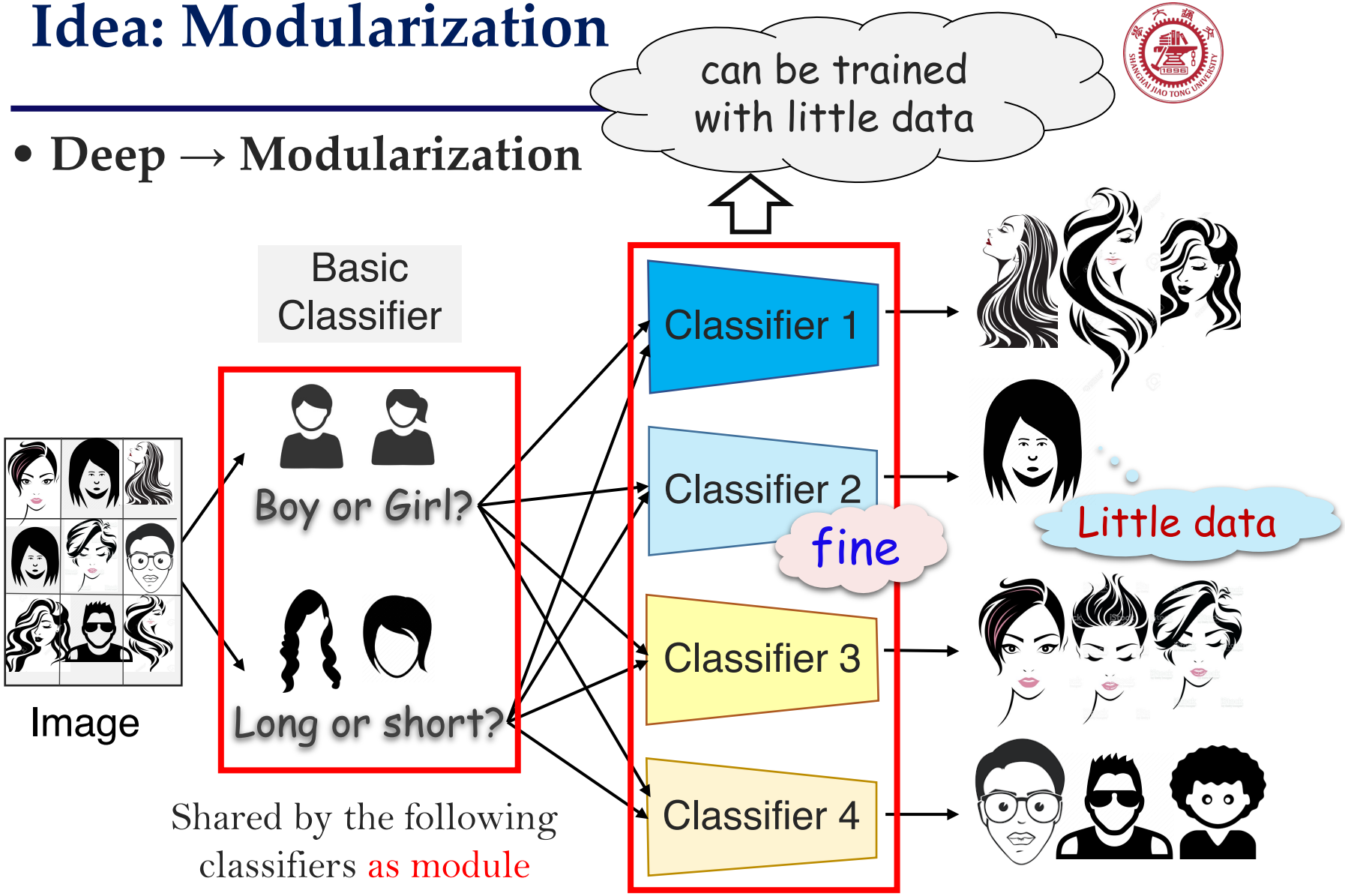


Idea: Modularization



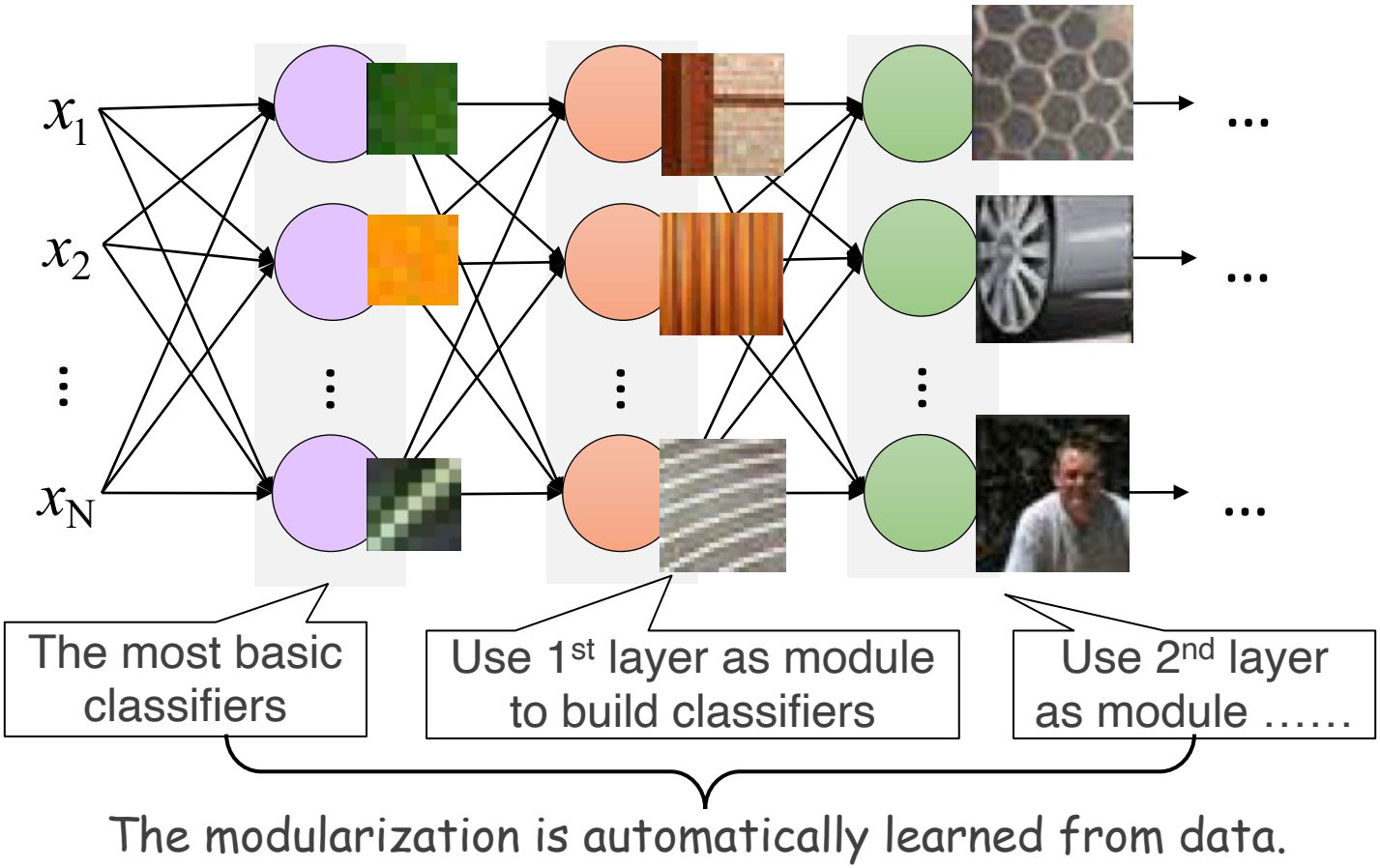
Idea: Modularization

- Deep → Modularization

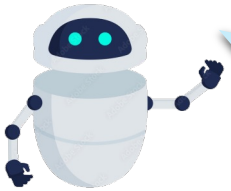


Modularization - Image

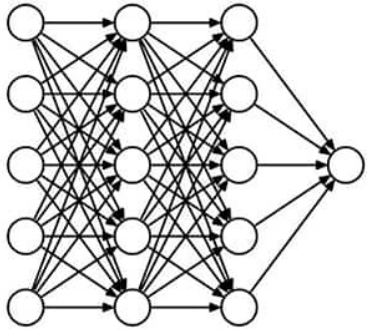
- Deep → Modularization → Less training data?



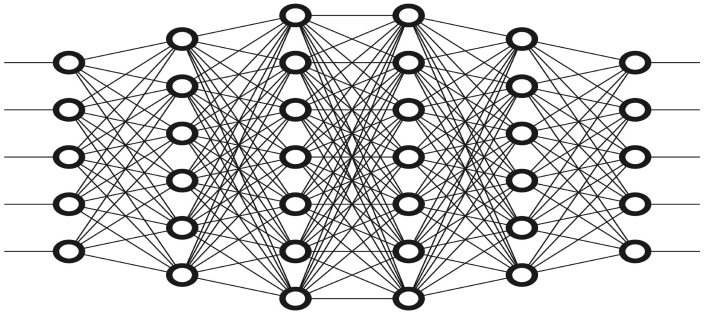
The Deeper, The Better

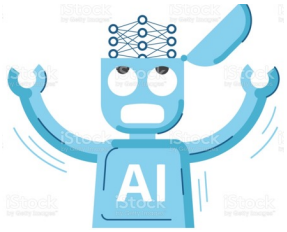


So, shallow networks can represent any functions.



However, using deep structures is more effective.





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

Deep Learning

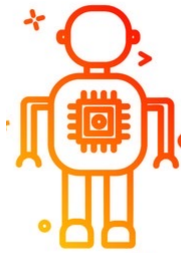
What is Deep Learning?

- **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

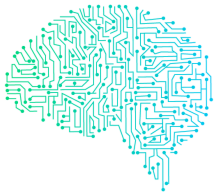
Computer systems that perform tasks that would usually require human intelligence



Machine Learning

Statistical techniques that learn from a series of inputs and outputs

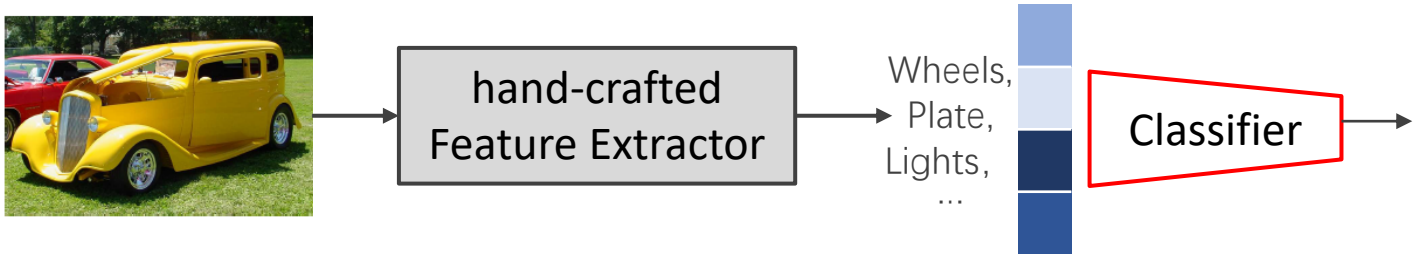
Deep Learning



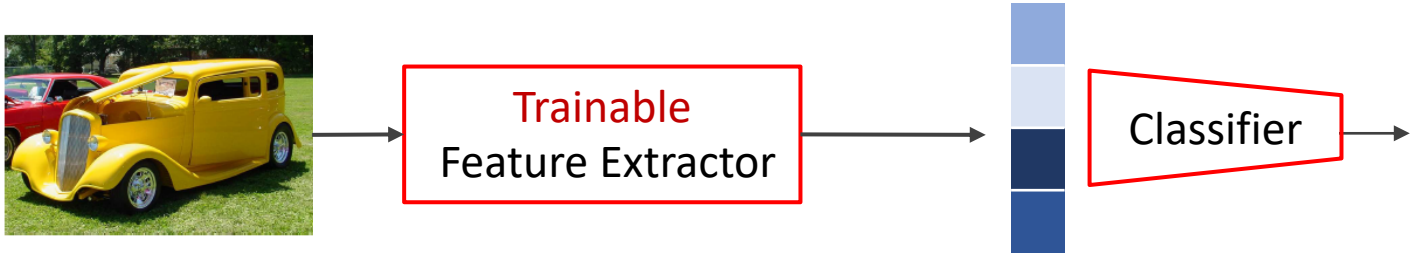
Machine learning algorithms that enable self learning of **data representations**

DL = Learning Feature Representations

- Traditional Machine Learning for Pattern Recognition
  - ▶ **fixed/engineered** features + **trainable** classifier

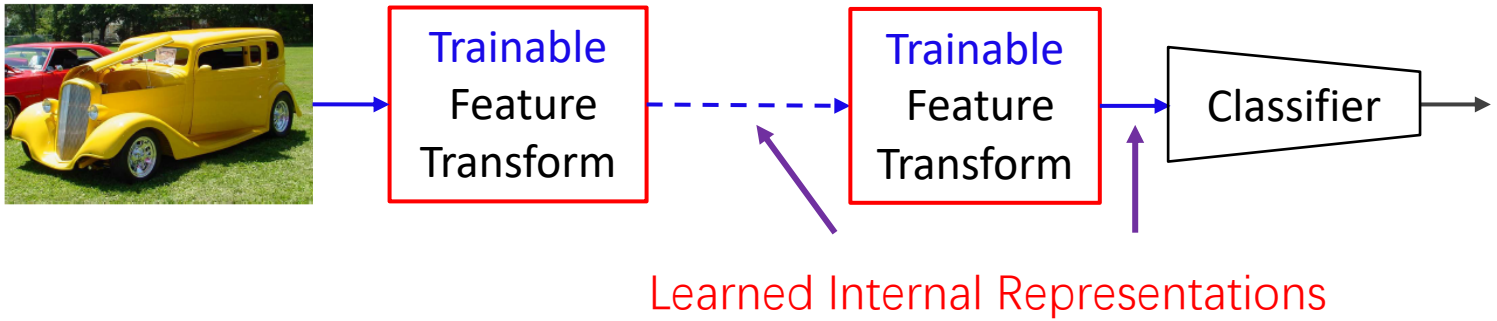


- Deep Learning
  - ▶ **trainable** features + **trainable** classifier

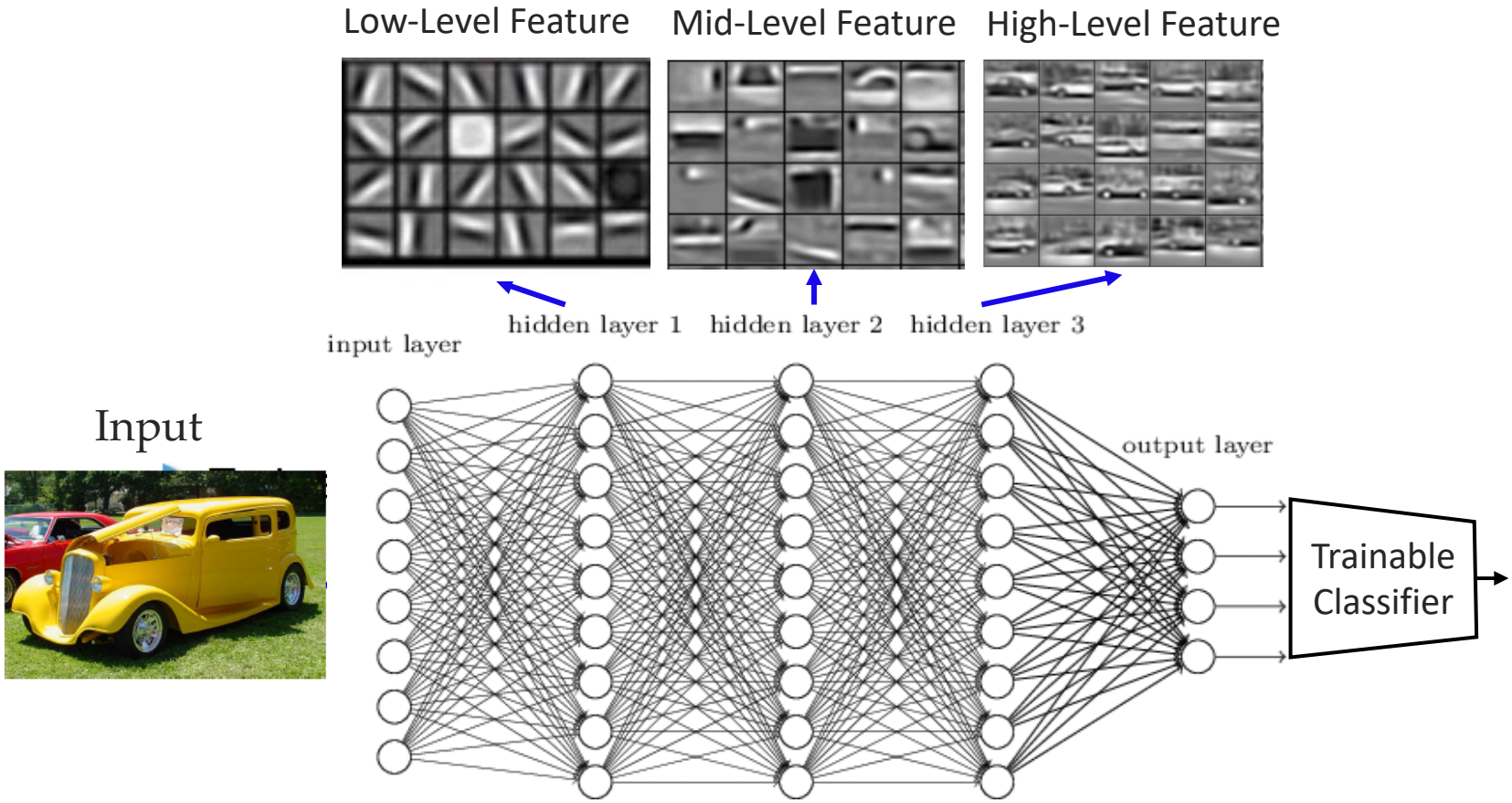


DL = Learning Hierarchical Representations

- **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



DL = Learning Hierarchical Representations



Hierarchical Representations with Increasing Level of Abstraction

- **Image recognition**

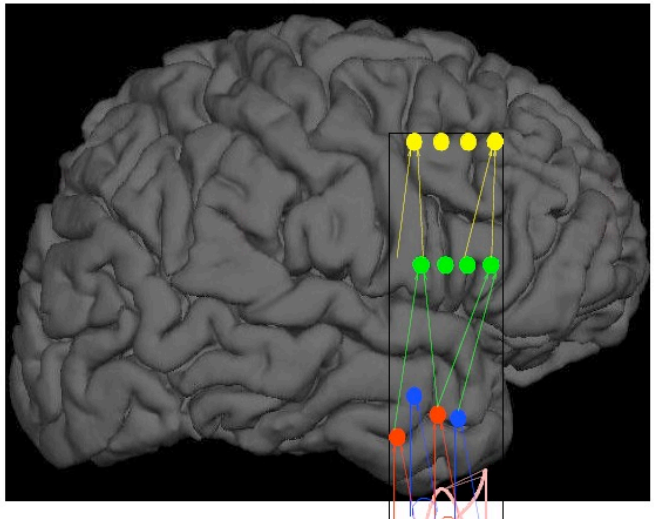
  pixel → edge → texton → motif → part → object

- **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.

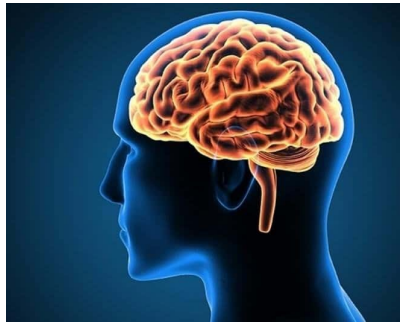
Hierarchical Representations in the Nature

- **The worlds of different species.**

sunshine,
wetness,…

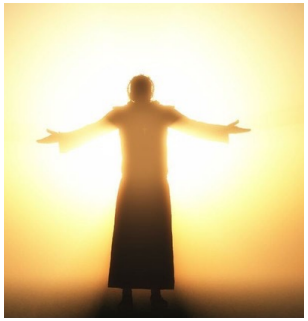
self-awareness、
color, taste,
movement

happiness,
sorrow, anxiety,
depression



?

…

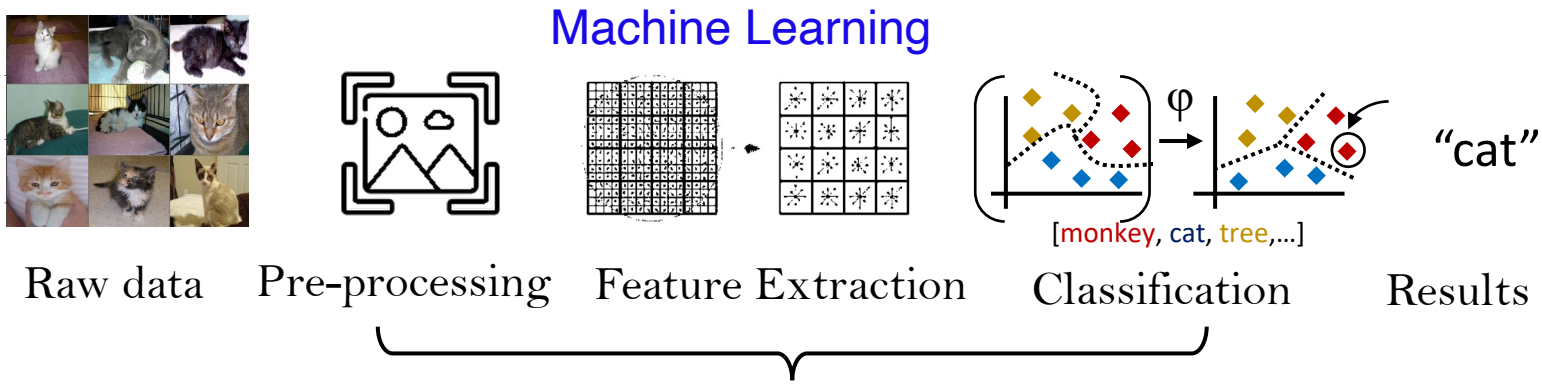


(complexity of neurons)

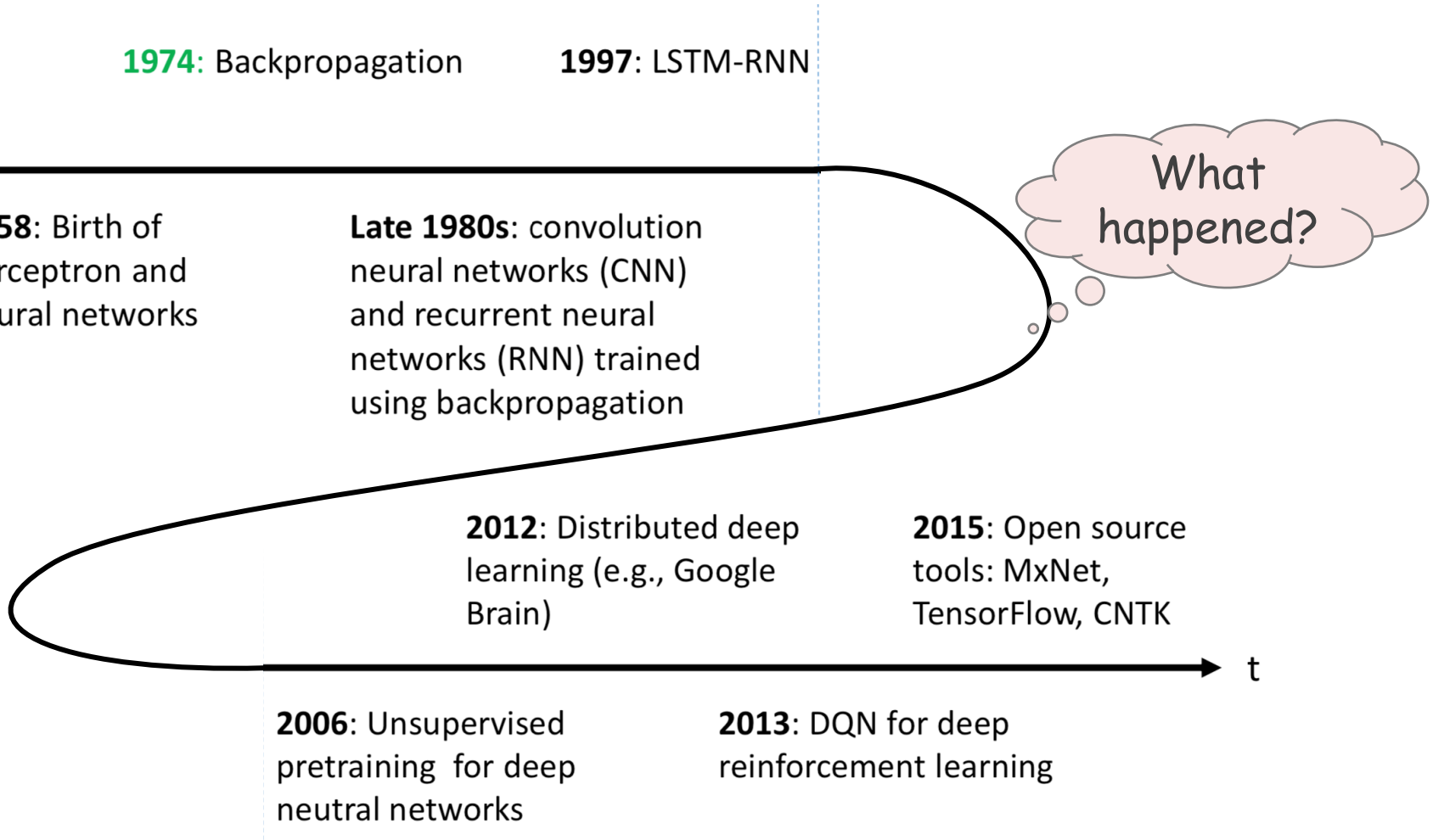
Maybe life is about hierarchical representations

DL = End-to-end Learning

- the model learns all the steps between the initial input phase and the final output result.



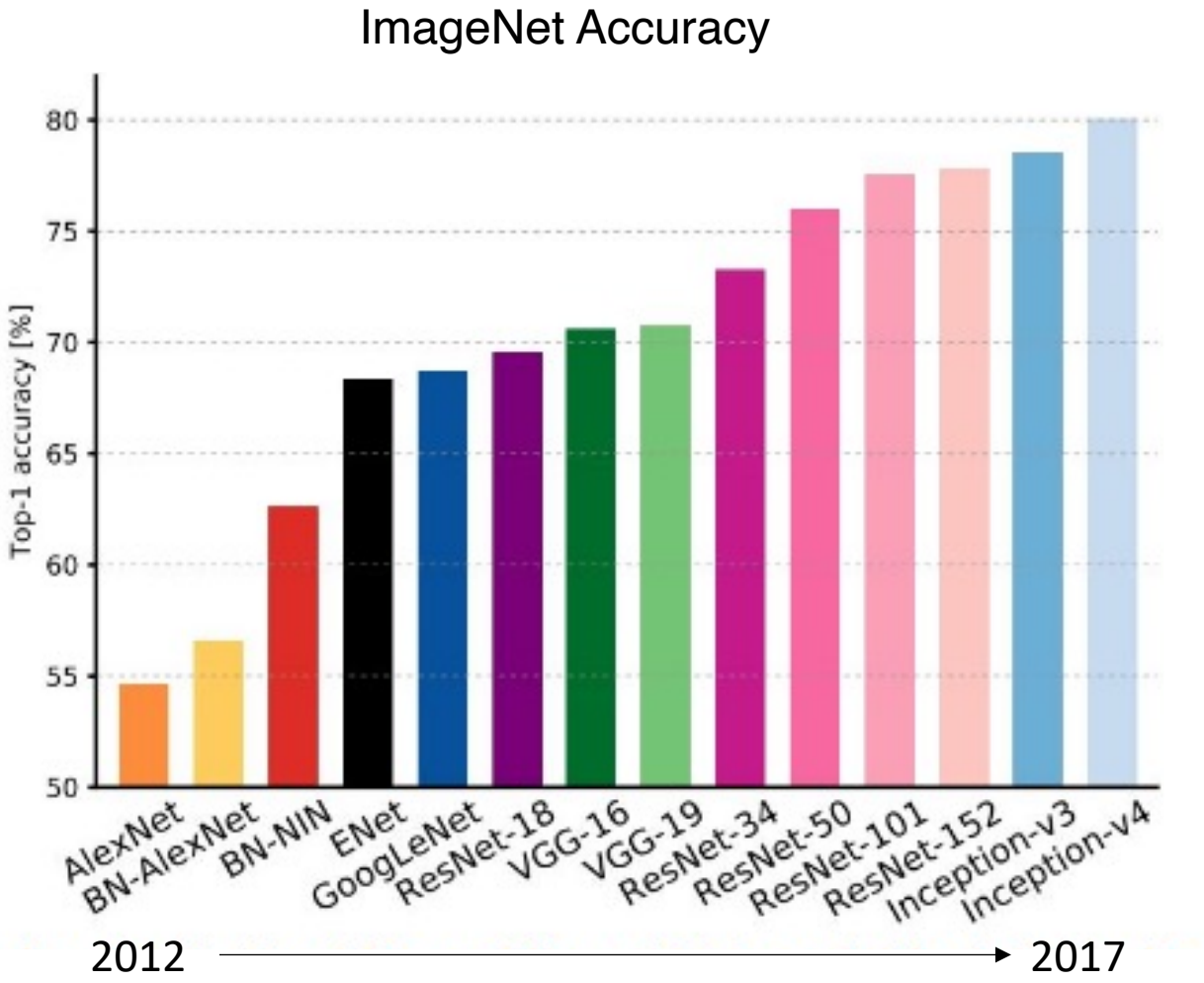
History of Deep Learning



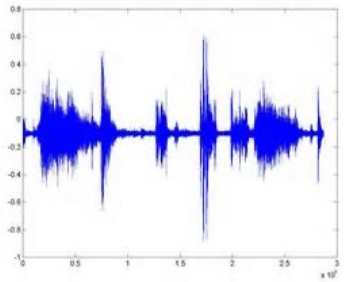
Why is deep learning blossoming eventually?

- 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

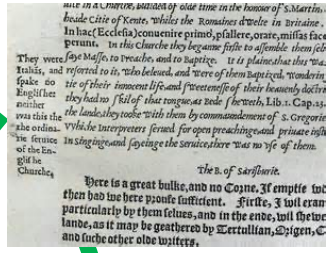
The Success of Deep Learning



Applications in Various Domains



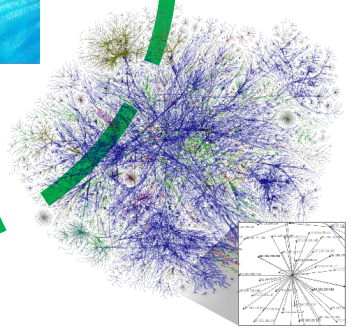
Speech



Natural language

Deep Learning

Image

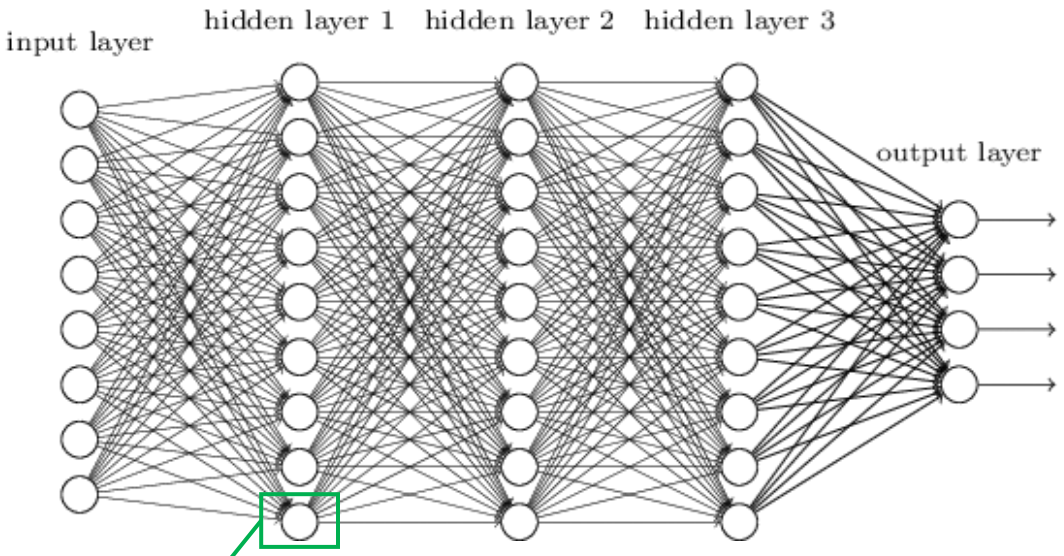


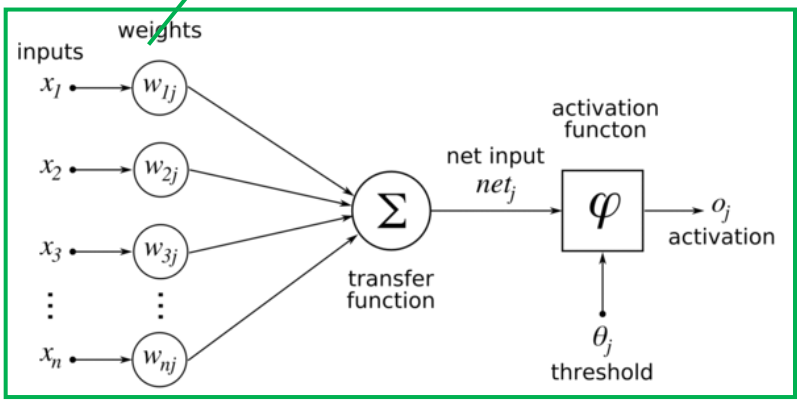
Network

Deep Neural Networks

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

Deep (Feedforward) Neural Networks





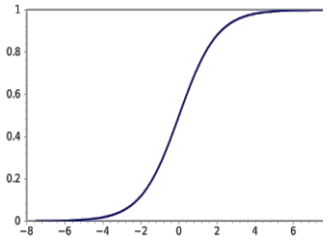
Loss functions:

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

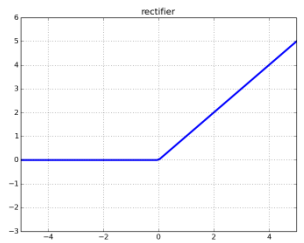
Training:

Backpropagation

Activation functions

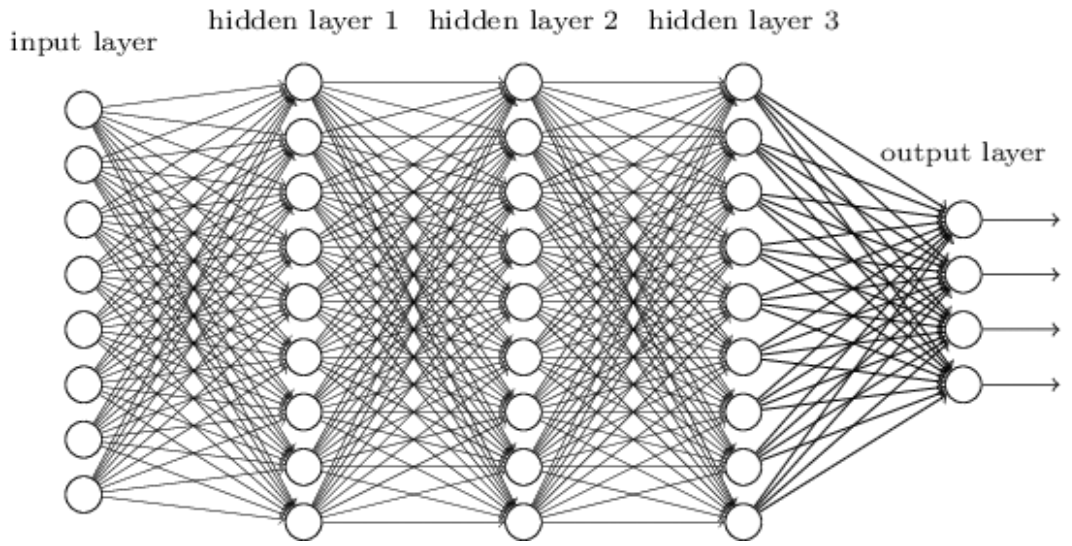


Sigmoid



ReLU

Limitations of Feedforward DNNs



But feedforward DNN is not widely used in practice.

- Hard to train (too many parameters, gradient vanishing, etc.)
- Input invariant
- …

Usually used as intermediate layers (e.g., nonlinear transformation)

Instead, We Use

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