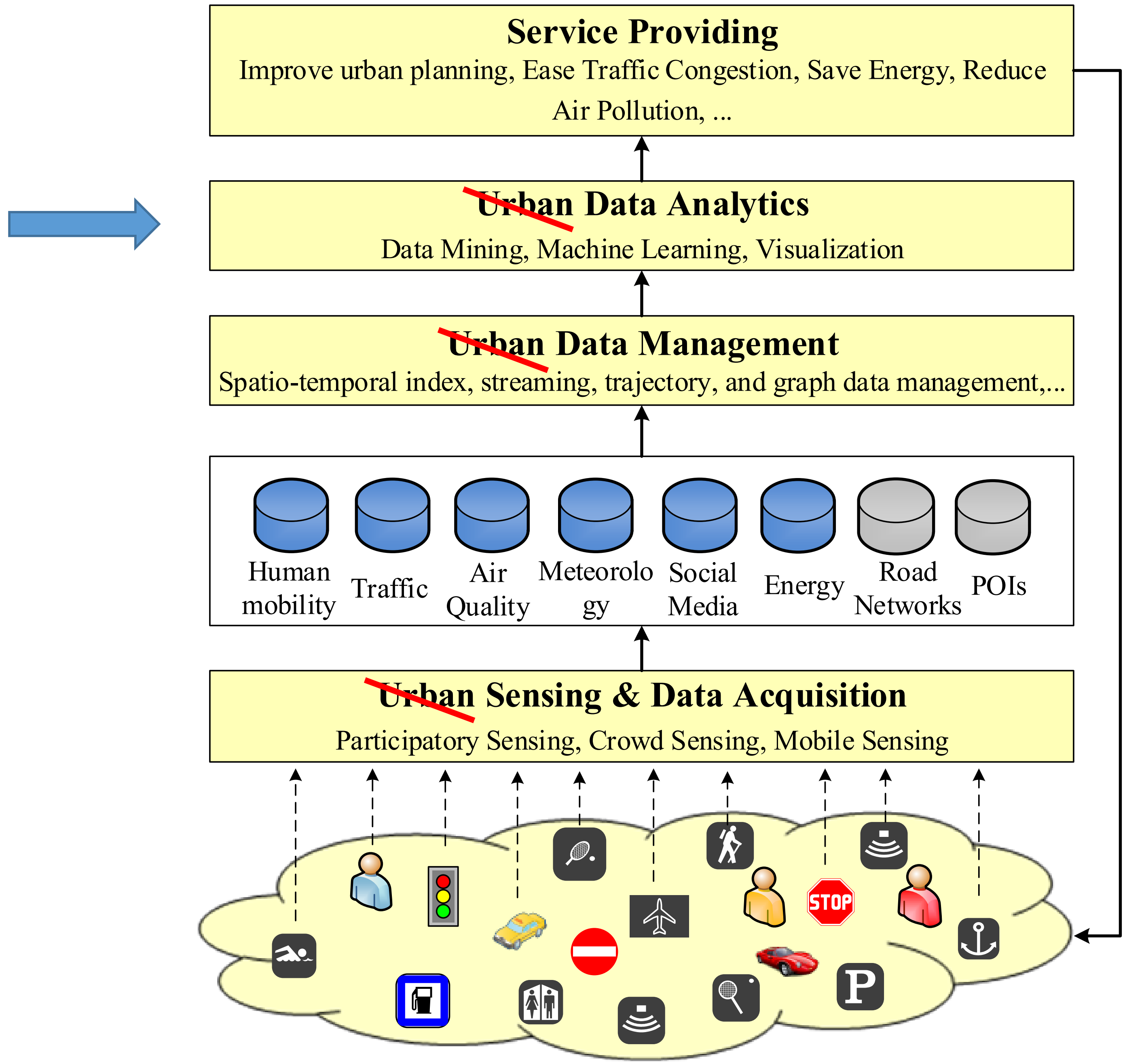
Welcome to

DS3010:
DS-III: Computational Data Intelligence
Why Deep Learning / Multi-Layer Perceptron?

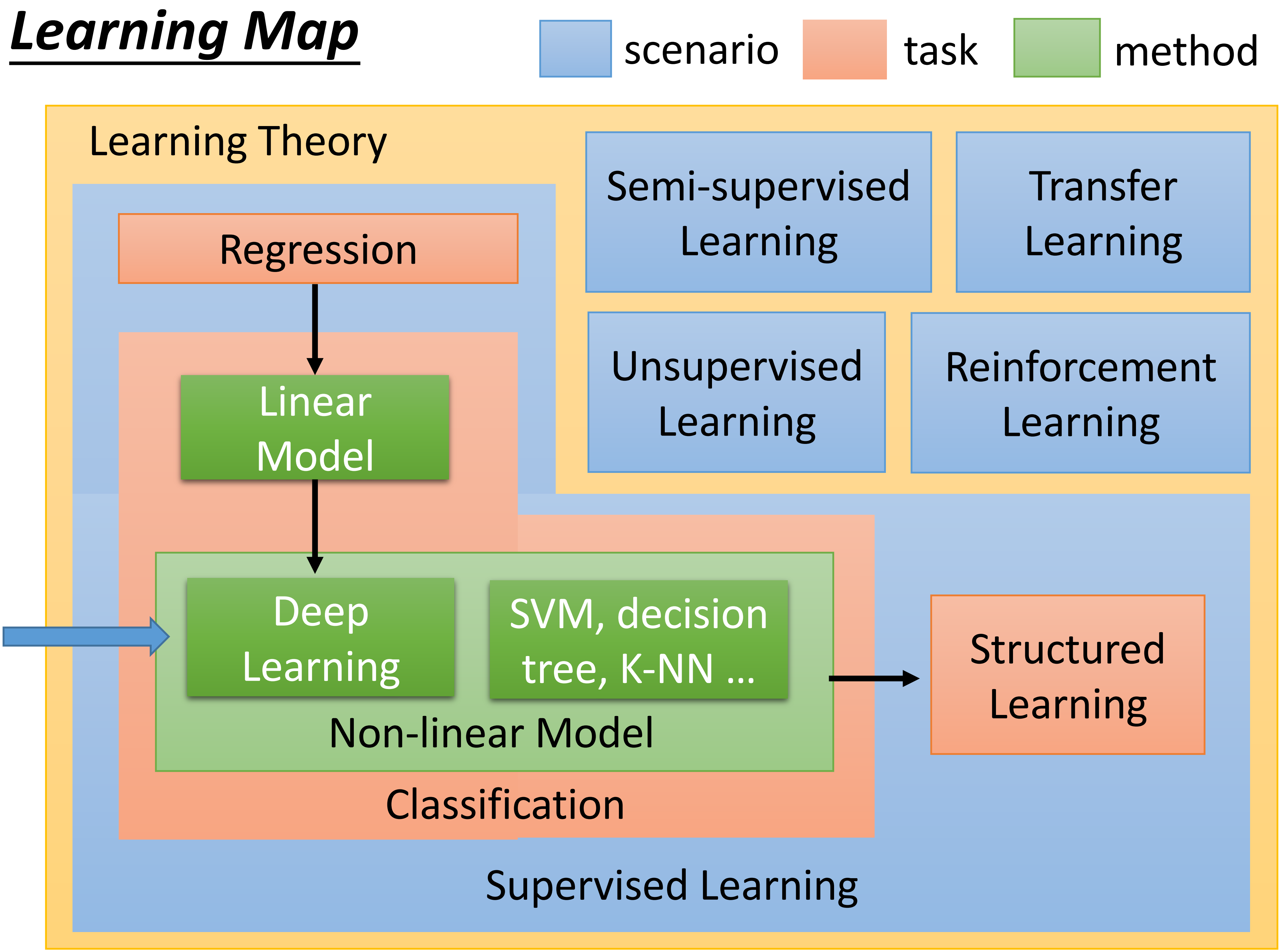
Prof. Yanhua Li

Time: 11:00am – 12:50pm M & R
Location: HL 114
D-term 2022

Data pipeline

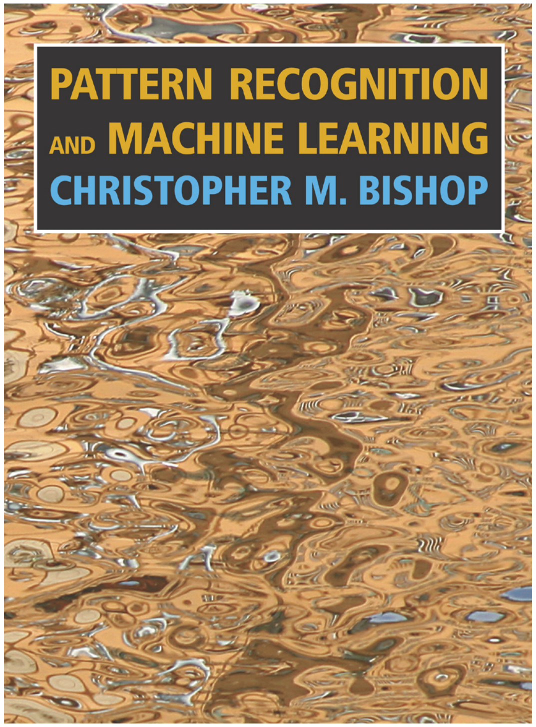


Urban Computing: concepts, methodologies, and applications.
Zheng, Y., et al. *ACM transactions on Intelligent Systems and Technology*.


Learning Map
scenario
task
method
Learning Theory
Semi-supervised Learning
Transfer Learning
Unsupervised Learning
Reinforcement Learning
Regression
Linear Model
Deep Learning
SVM, decision tree, K-NN …
Non-linear Model
Classification
Structured Learning
Supervised Learning

References

Classification
Multi-Layer
Perceptron / Deep
Learning



Bishop: Chapter 5.1

Why Deep Learning?

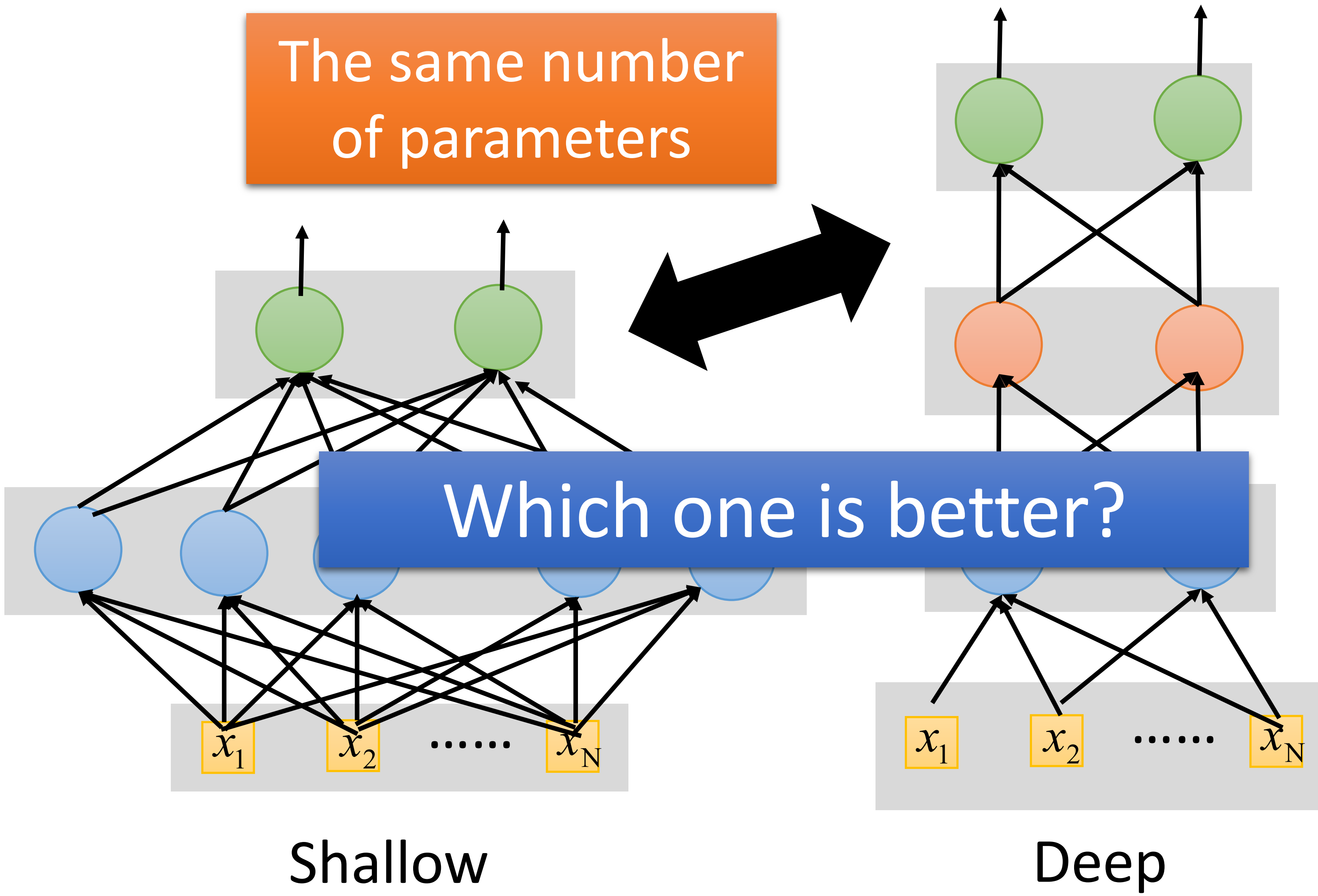
Deeper is Better?

Layer X Size	Word Error Rate (%)
1 X 2k	24.2
2 X 2k	20.4
3 X 2k	18.4
4 X 2k	17.8
5 X 2k	17.2
7 X 2k	17.1

Not surprised, more parameters, better performance

Seide, Frank, Gang Li, and Dong Yu. "Conversational Speech Transcription Using Context-Dependent Deep Neural Networks." *Interspeech*. 2011.

Fat + Short v.s. Thin + Tall



Fat + Short v.s. Thin + Tall

Layer X Size	Word Error Rate (%)	Layer X Size	Word Error Rate (%)
1 X 2k	24.2		
2 X 2k	20.4		
3 X 2k	18.4		
4 X 2k	17.8		
5 X 2k	17.2	1 X 3772	22.5
7 X 2k	17.1	1 X 4634	22.6
		1 X 16k	22.1

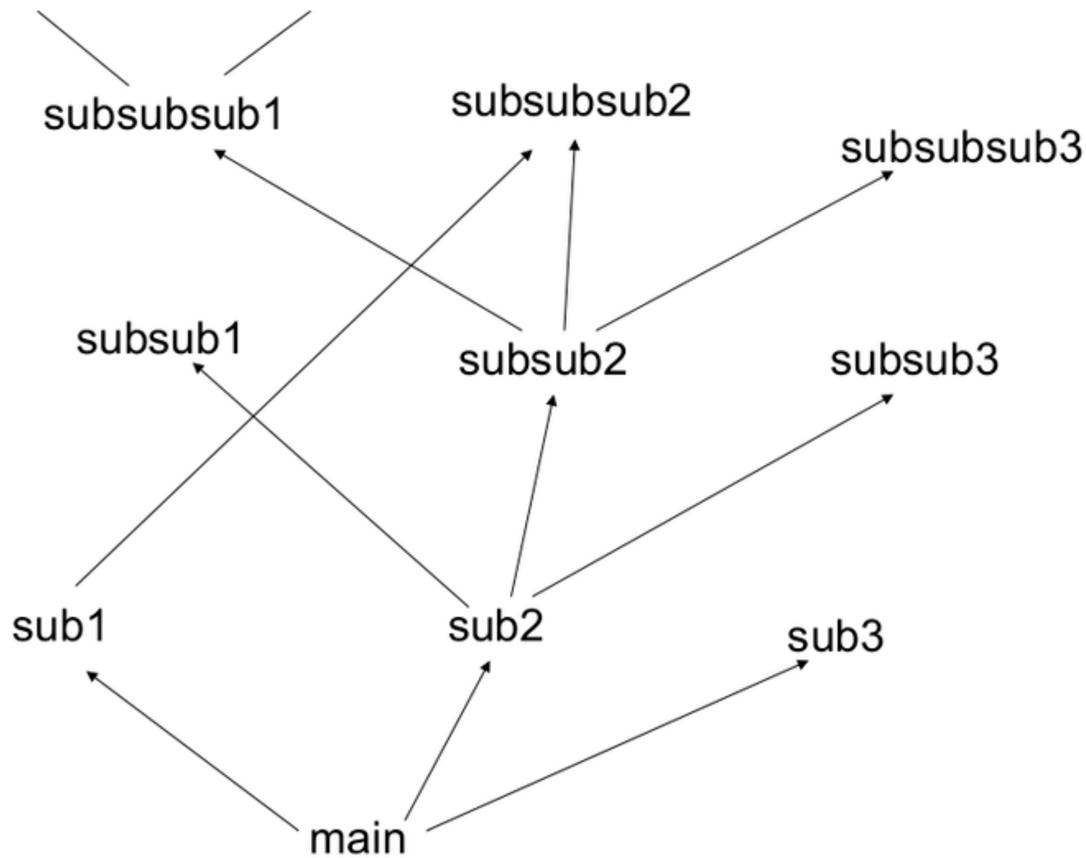
Why?

Seide, Frank, Gang Li, and Dong Yu. "Conversational Speech Transcription Using Context-Dependent Deep Neural Networks." *Interspeech*. 2011.

Modularization

- Deep → Modularization

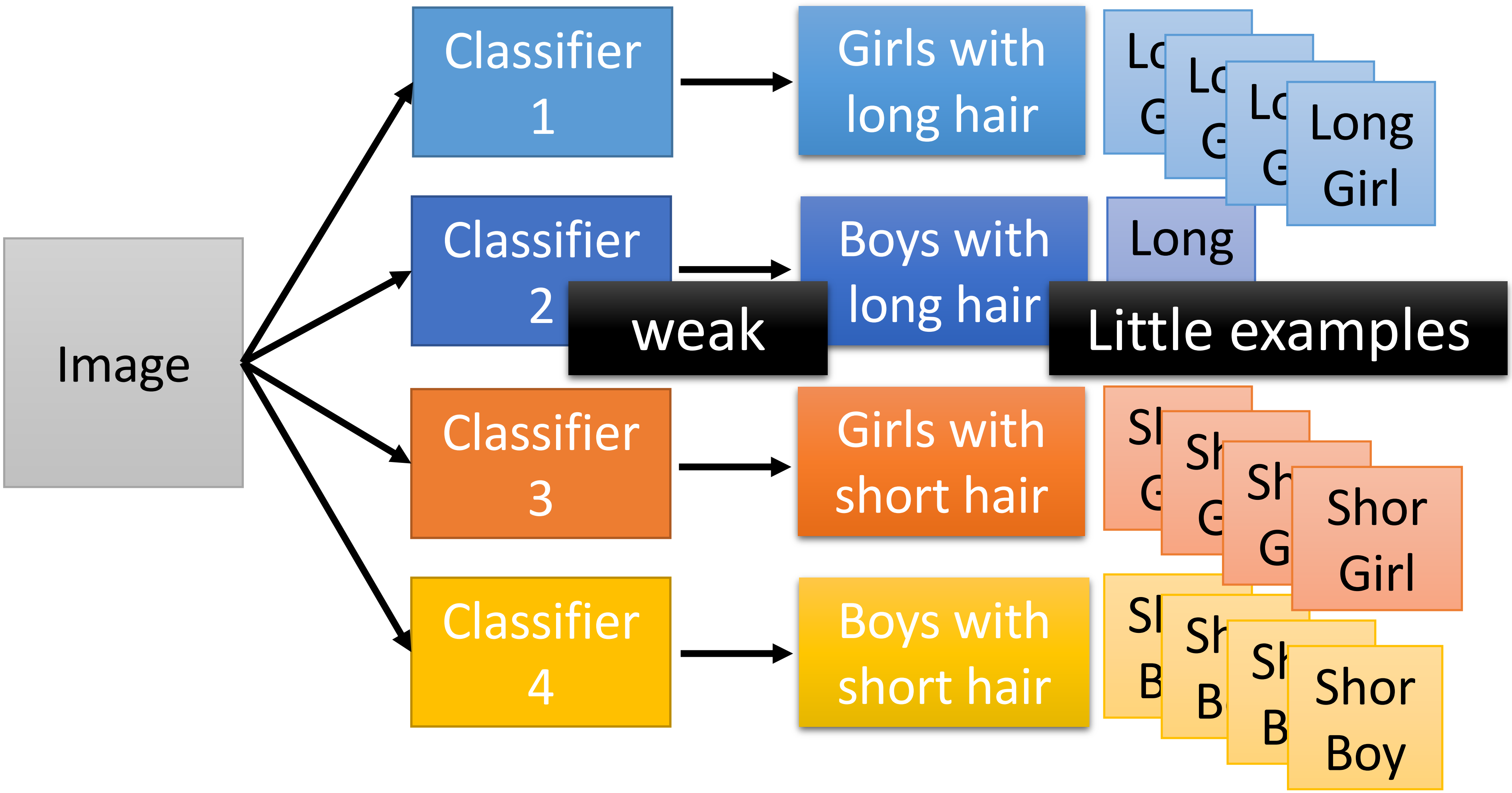
Don't put everything in your main function.



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

Modularization

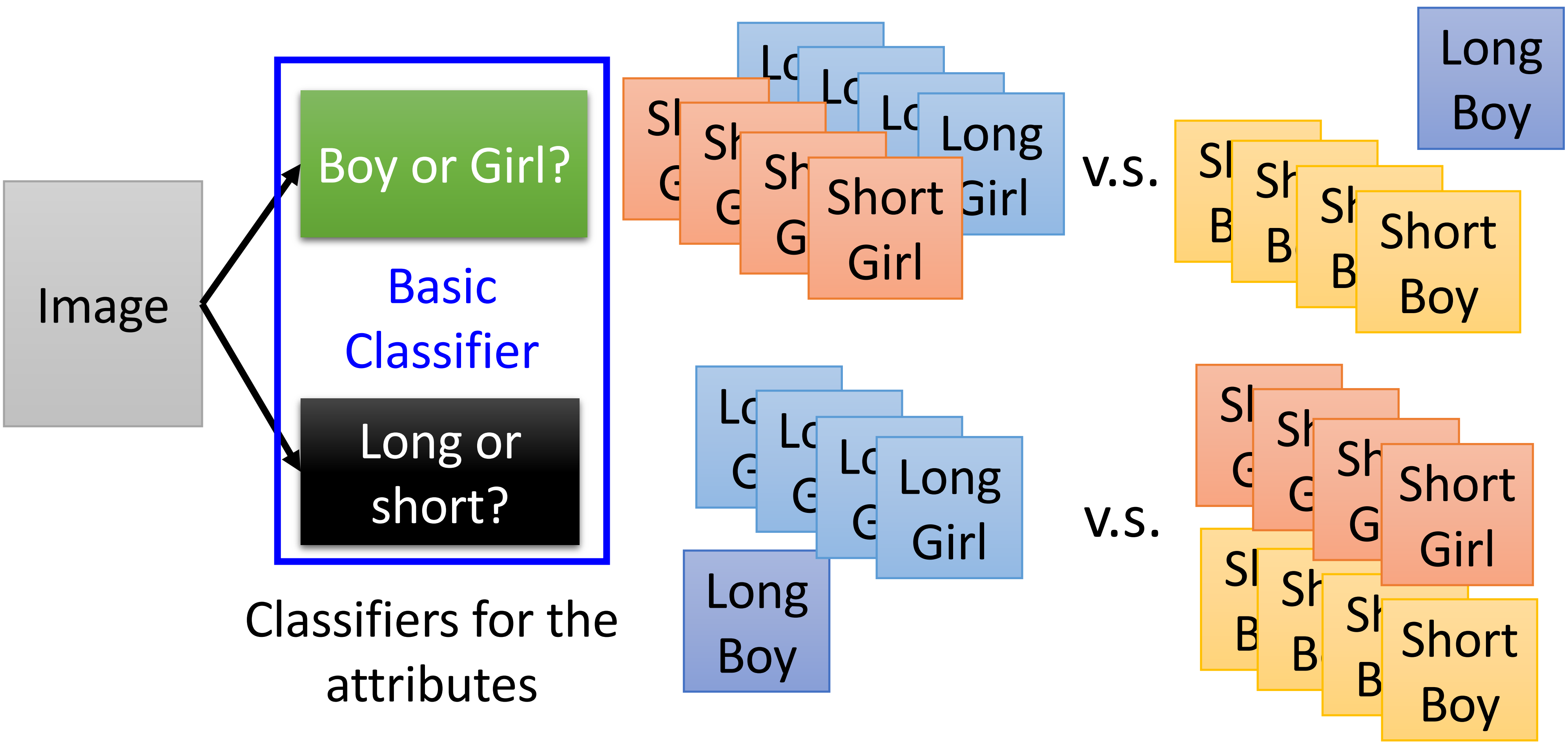
- Deep → Modularization



Modularization

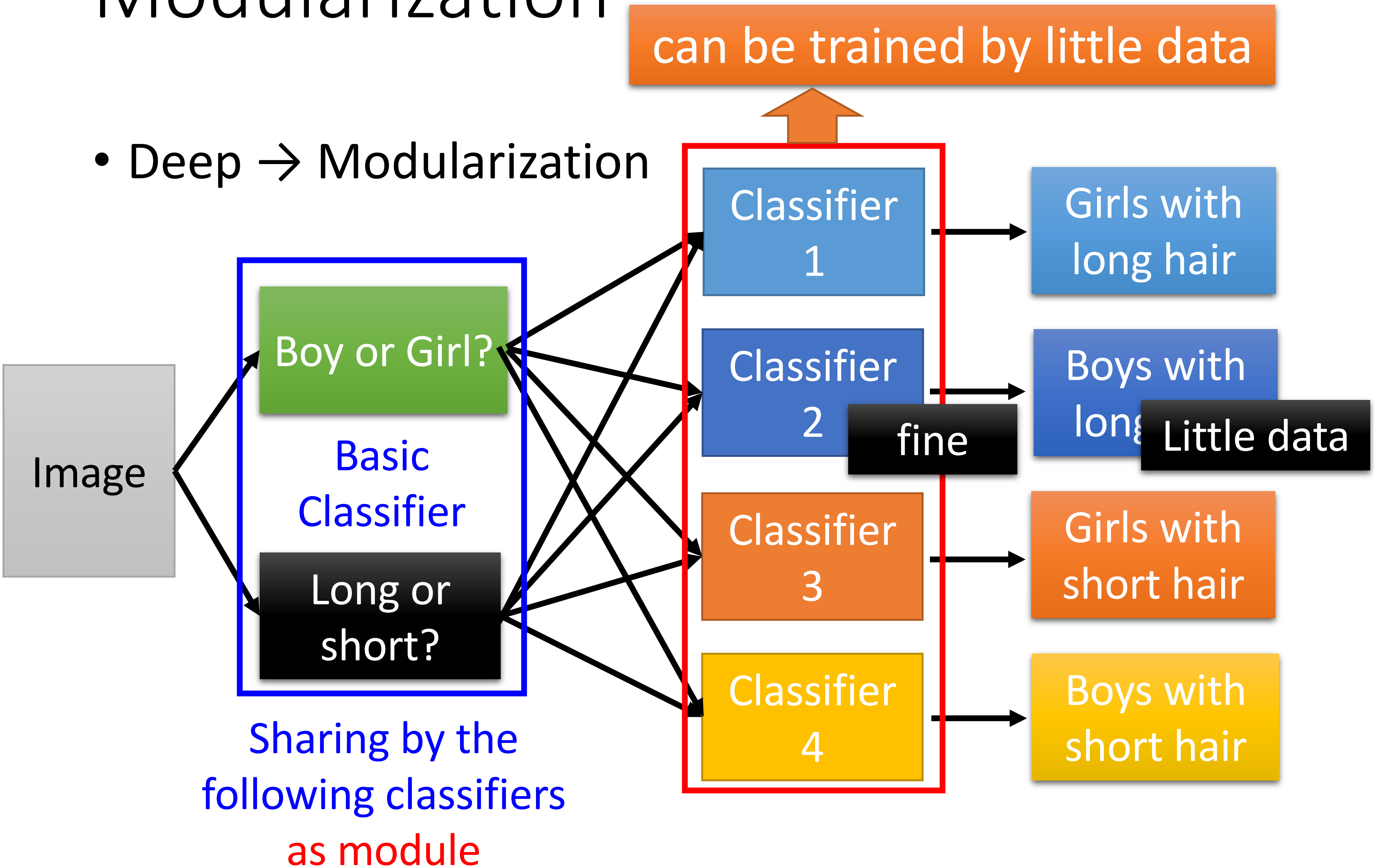
Each basic classifier can have sufficient training examples.

- Deep → Modularization



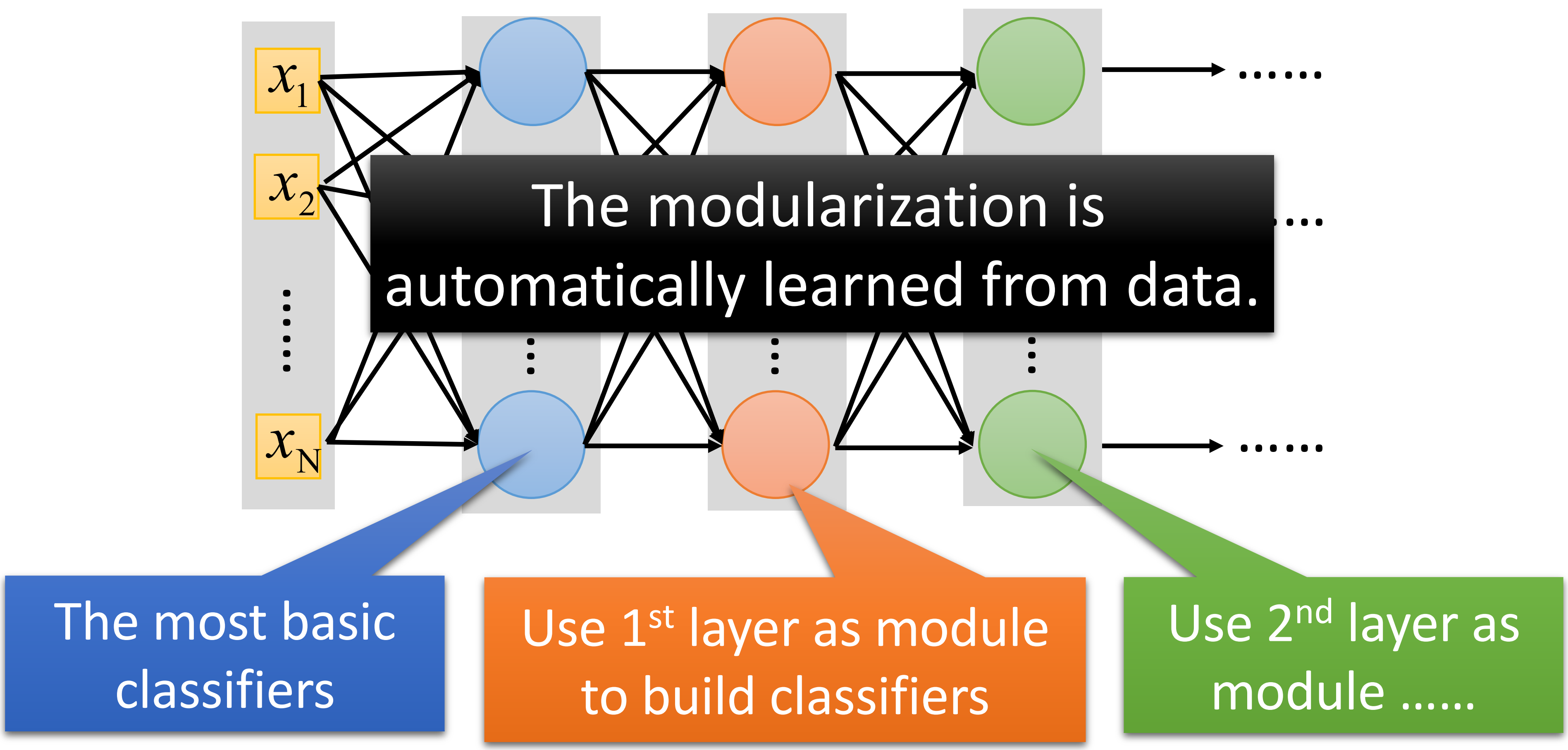
Modularization

- Deep → Modularization



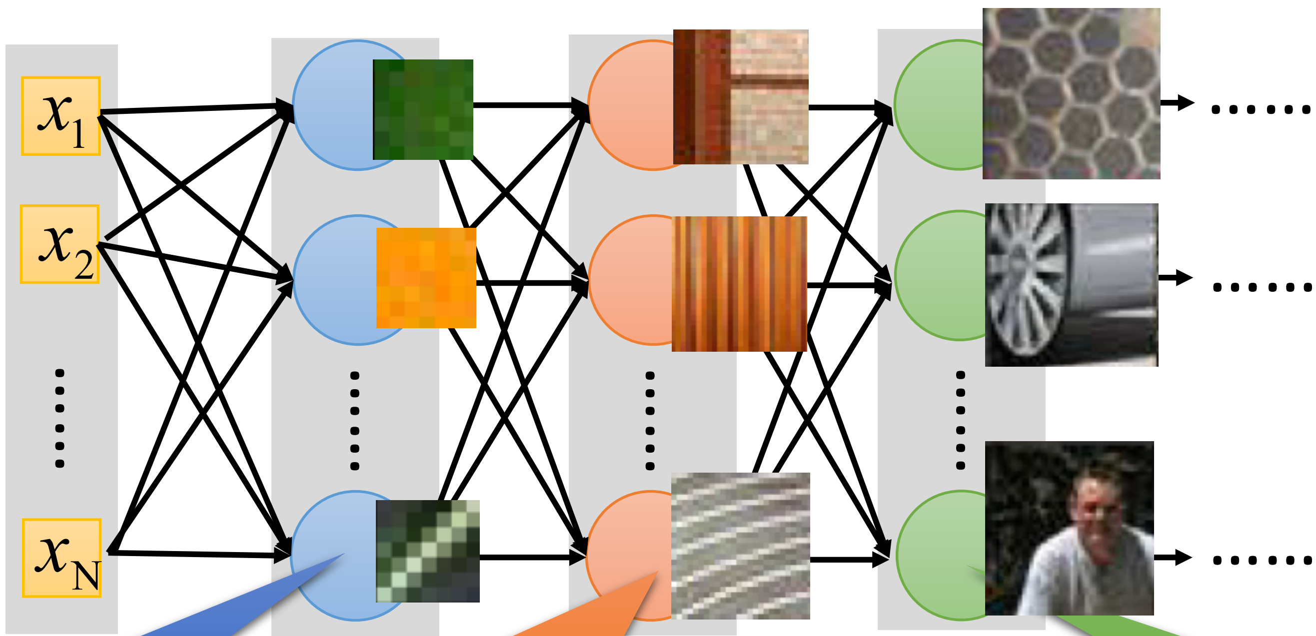
Modularization

- Deep → Modularization → Less training data?



Modularization - Image

- Deep → Modularization



The most basic classifiers

Use 1st layer as module to build classifiers

Use 2nd layer as module ……

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

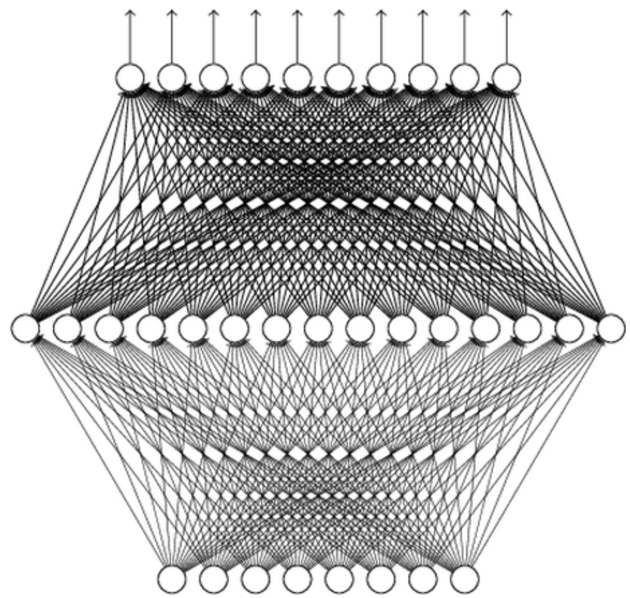
Universality Theorem

Any continuous function f

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

Can be realized by a network with one hidden layer

(given **enough** hidden neurons)

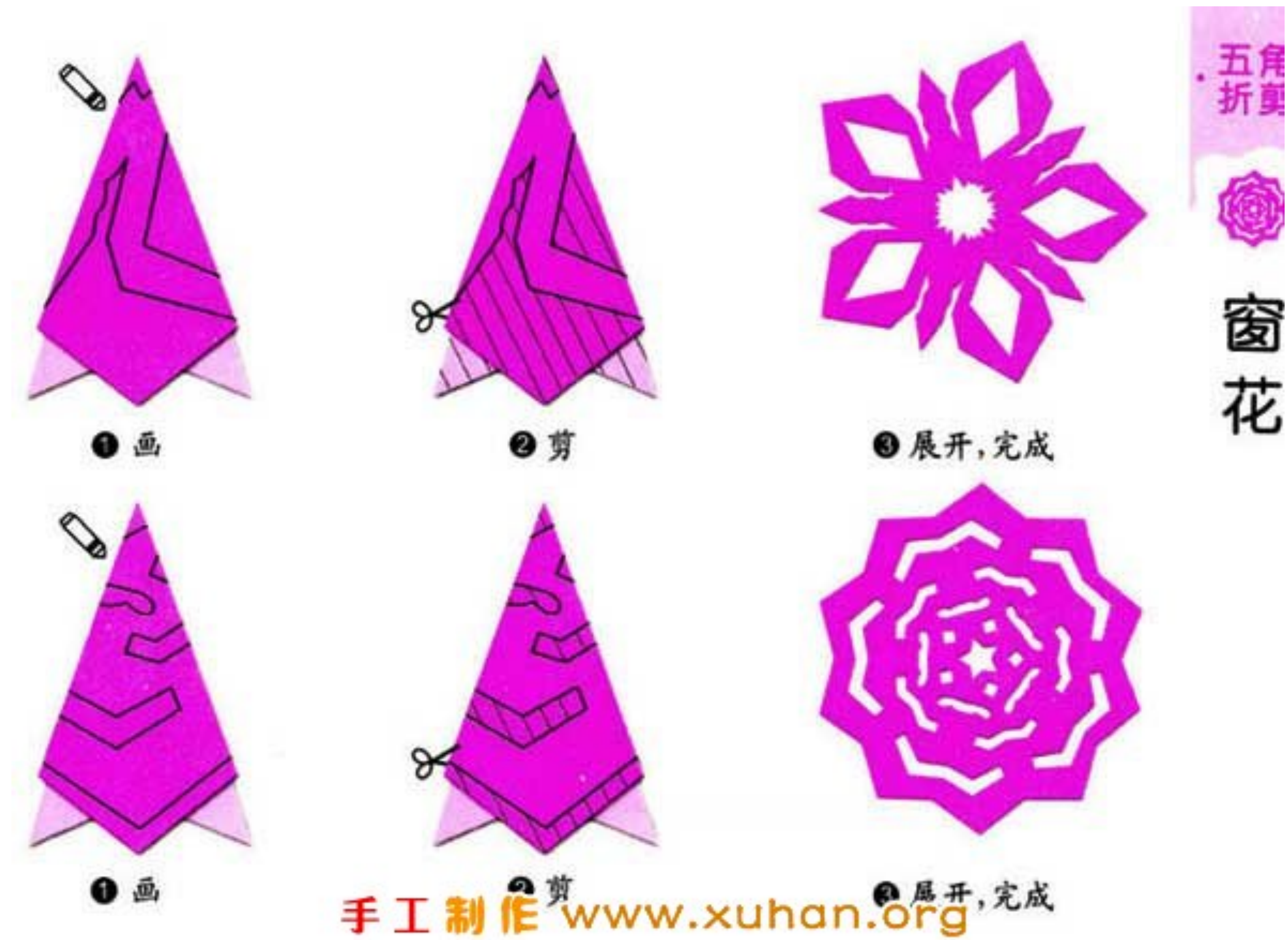


Reference for the reason: http://neuralnetworksanddeeplearning.com/chap4.html

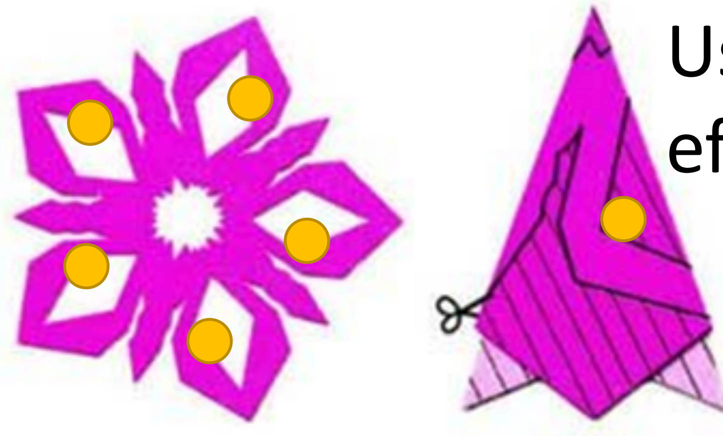
Yes, shallow network can represent any function.

However, using deep structure is more effective.

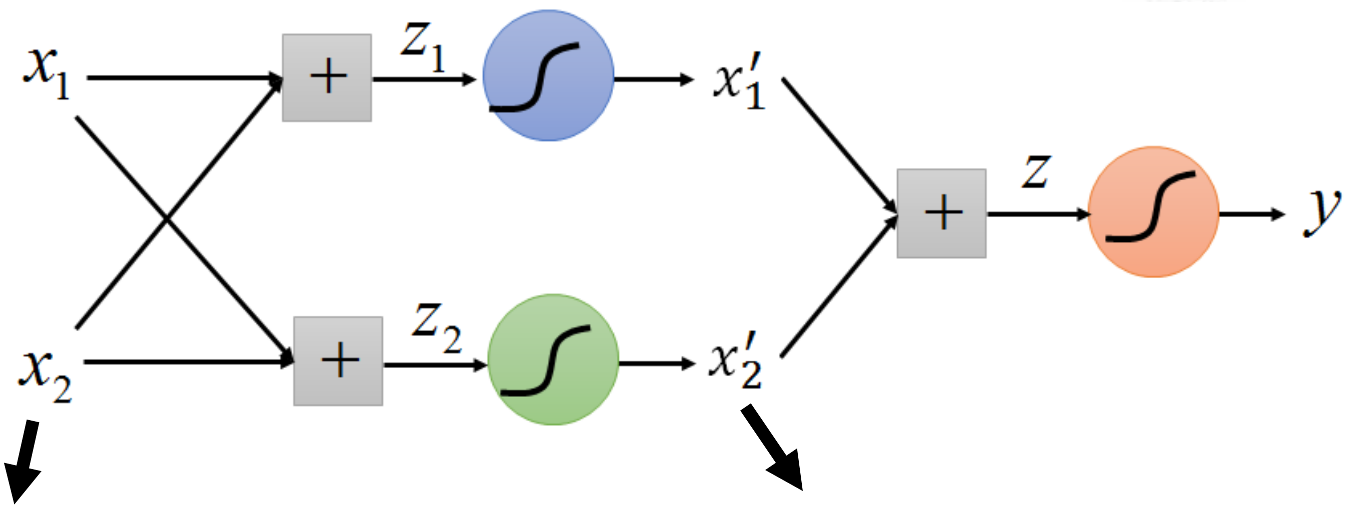
More Analogy

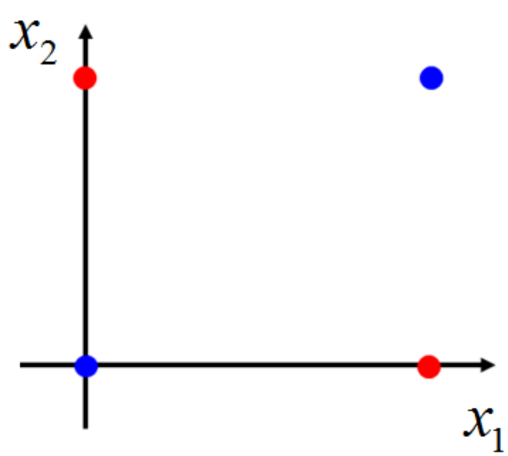


More Analogy

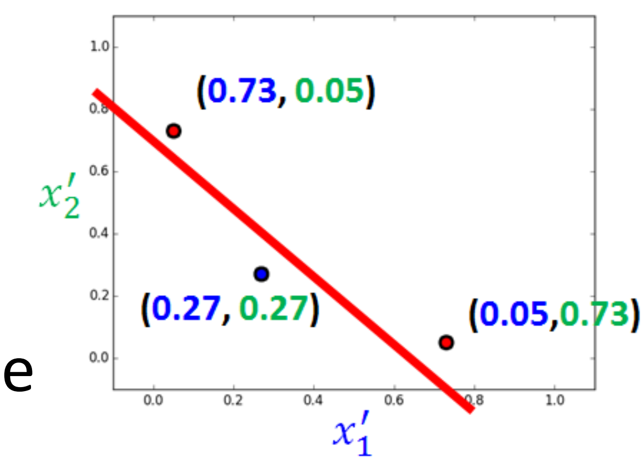


Use data effectively



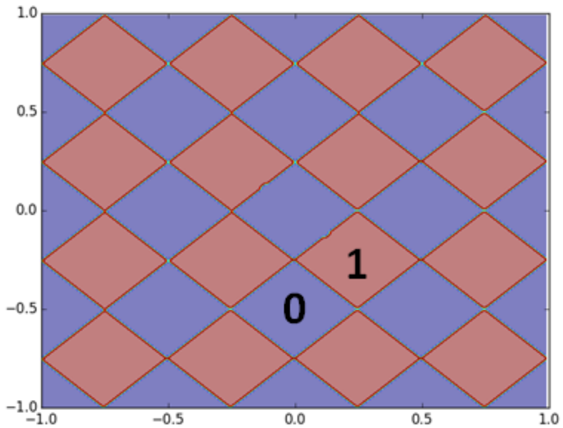


Folding the space



More Analogy - Experiment

$$f : R^2 \rightarrow \{0,1\}$$

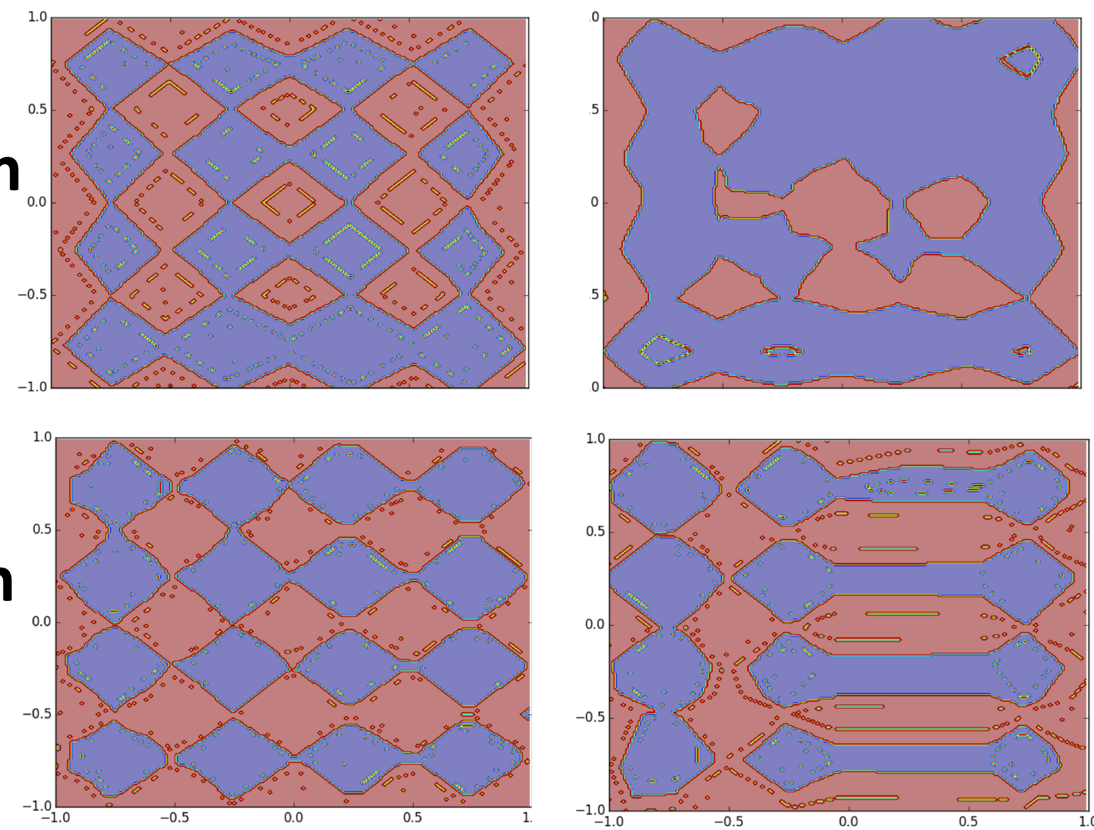


Different numbers of training examples

10,0000 | 2,0000

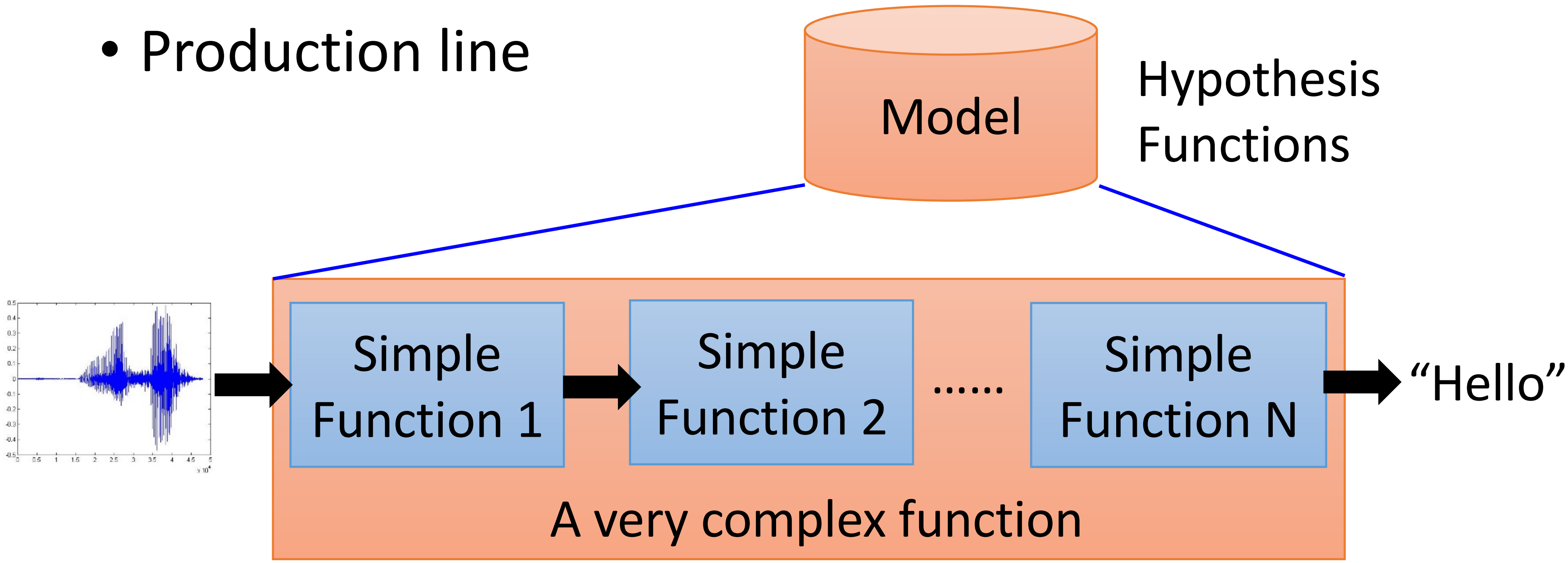
1 hidden layer

3 hidden layers



End-to-end Learning

- Production line

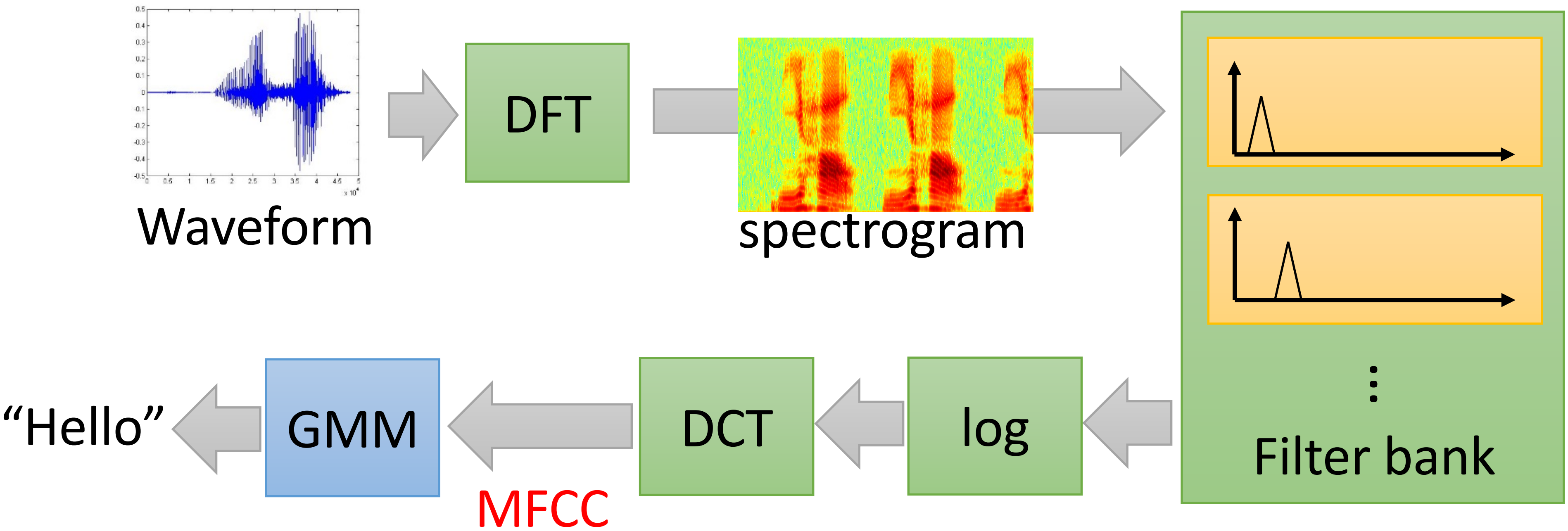


End-to-end training:

What each function should do is learned automatically

End-to-end Learning
- Speech Recognition

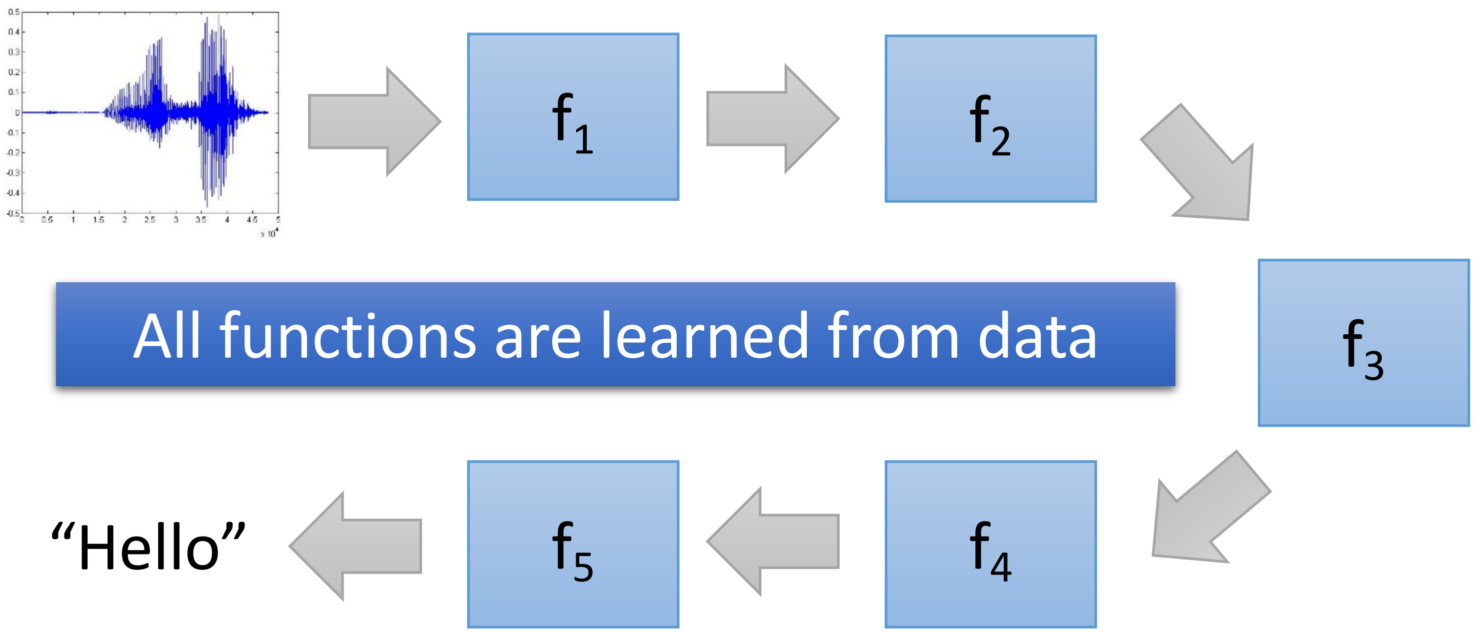
- Shallow Approach



Each box is a simple function in the production line:

End-to-end Learning
- Speech Recognition

- Deep Learning

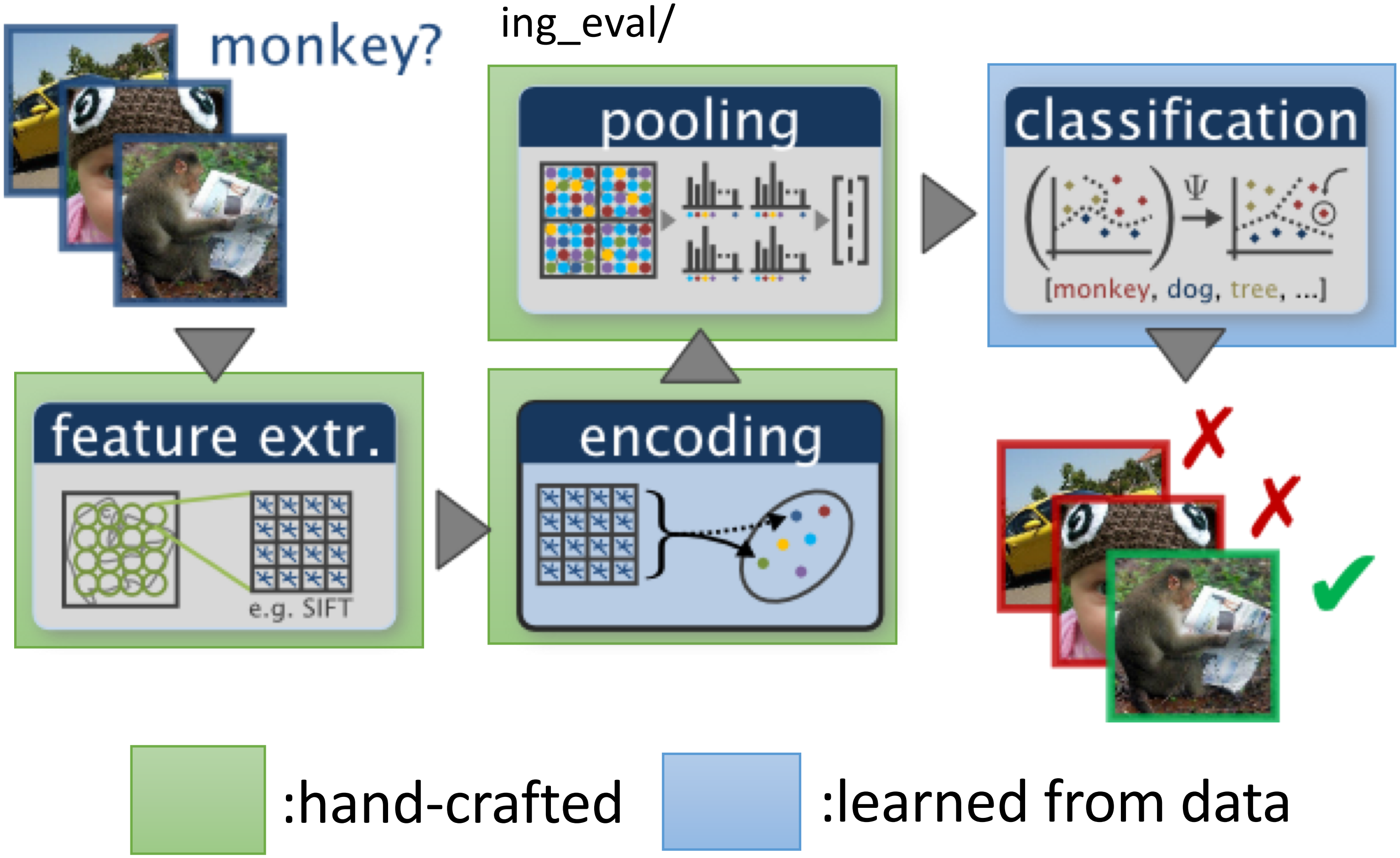


Less engineering labor, but machine learns more

End-to-end Learning
- Image Recognition

- Shallow Approach

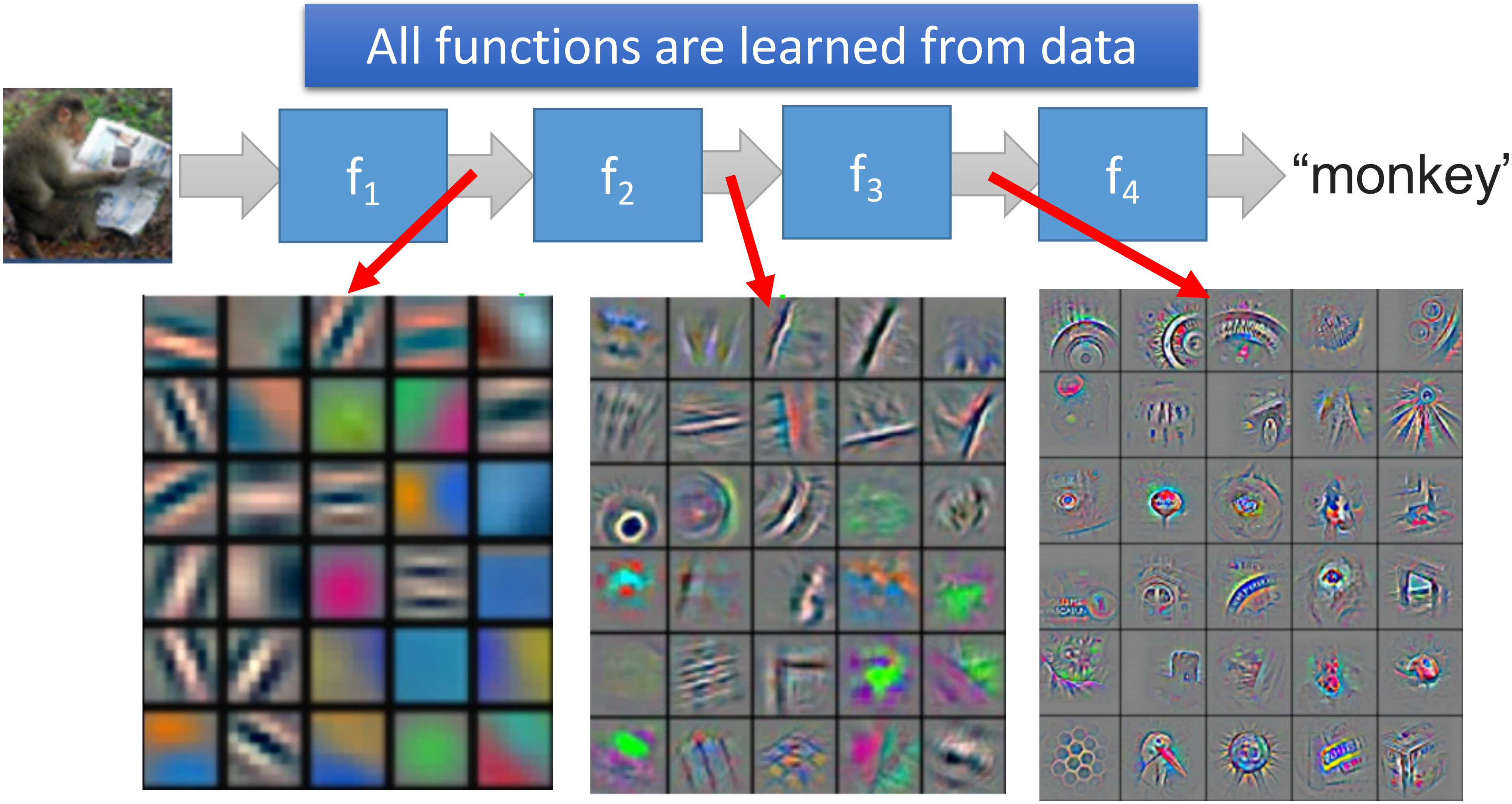
http://www.robots.ox.ac.uk/~vgg/research/encoding_eval/



End-to-end Learning - Image Recognition

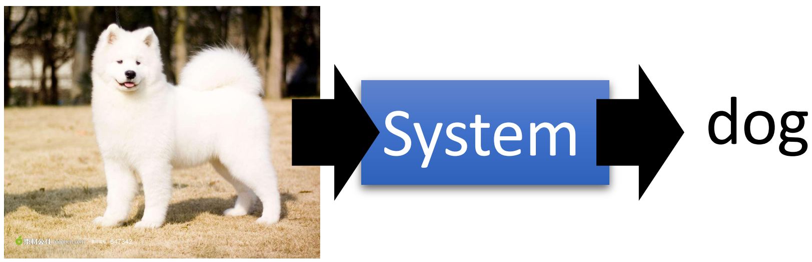
Reference: Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014* (pp. 818-833)

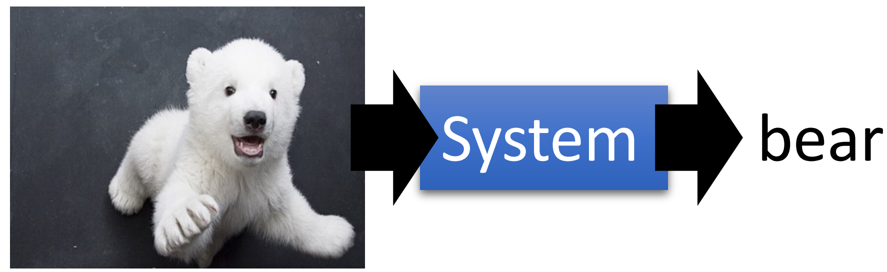
- Deep Learning



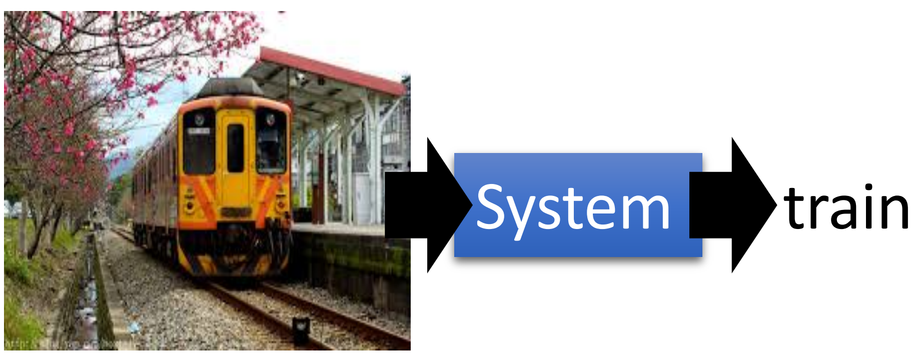
Complex Task …

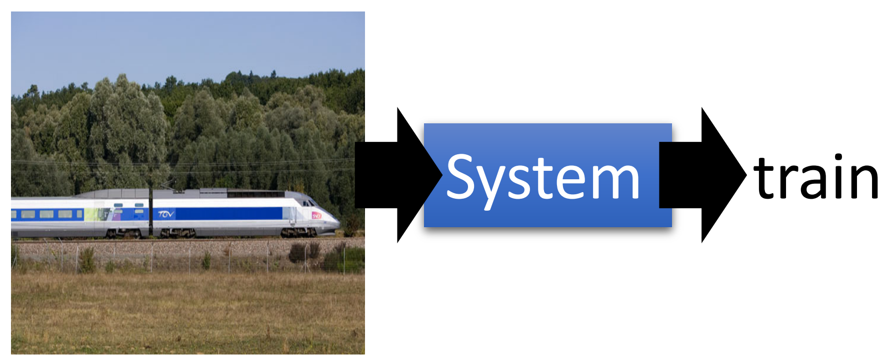
- Very similar input, different output





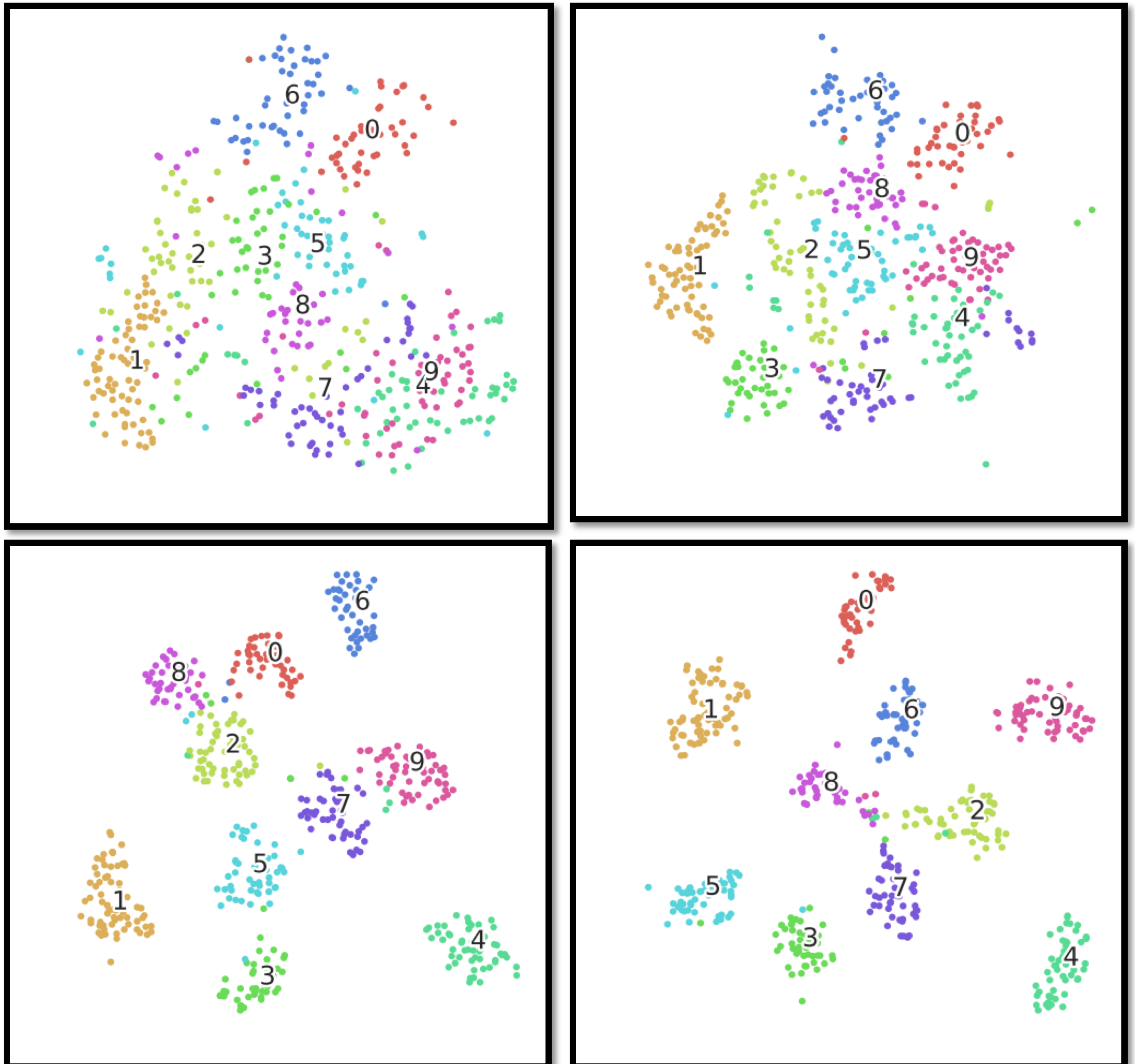
- Very different input, similar output





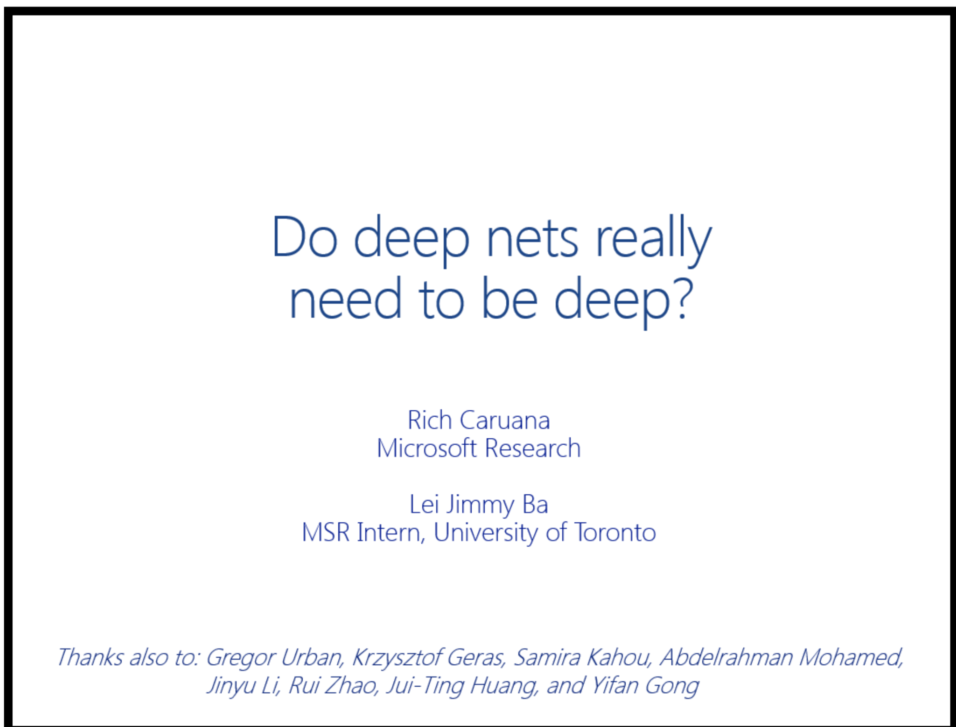
MNIST

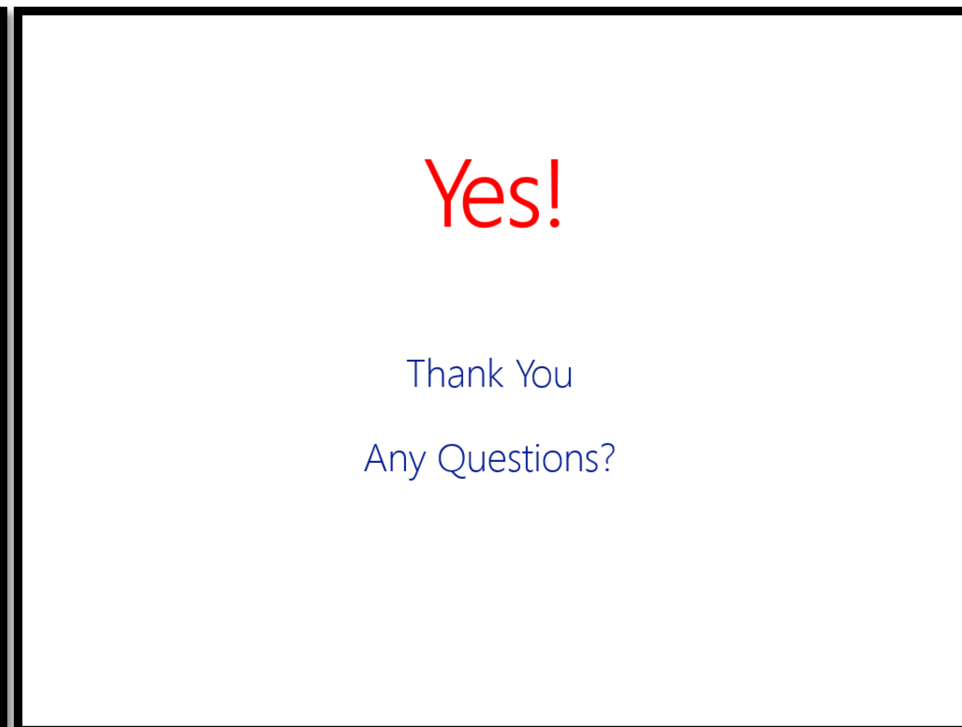
Complex Task ...



To learn more …

- Do Deep Nets Really Need To Be Deep? (by Rich Caruana)
- http://research.microsoft.com/apps/video/default.aspx?id=232373&r=1





keynote of Rich Caruana at ASRU 2015

To learn more …

- Deep Learning: Theoretical Motivations (*Yoshua Bengio*)
  - http://videolectures.net/deeplearning2015_bengio_theoretical_motivations/
- Connections between physics and deep learning
  - https://www.youtube.com/watch?v=5MdSE-N0bxs
- Why Deep Learning Works: Perspectives from Theoretical Chemistry
  - https://www.youtube.com/watch?v=kIbKHIPbxiU

Questions