PMI Lab
Pattern Recognition and Machine Intelligence Lab

한국외국어대학교
HANKUK UNIVERSITY OF FOREIGN STUDIES

DEEP LEARNING BASICS

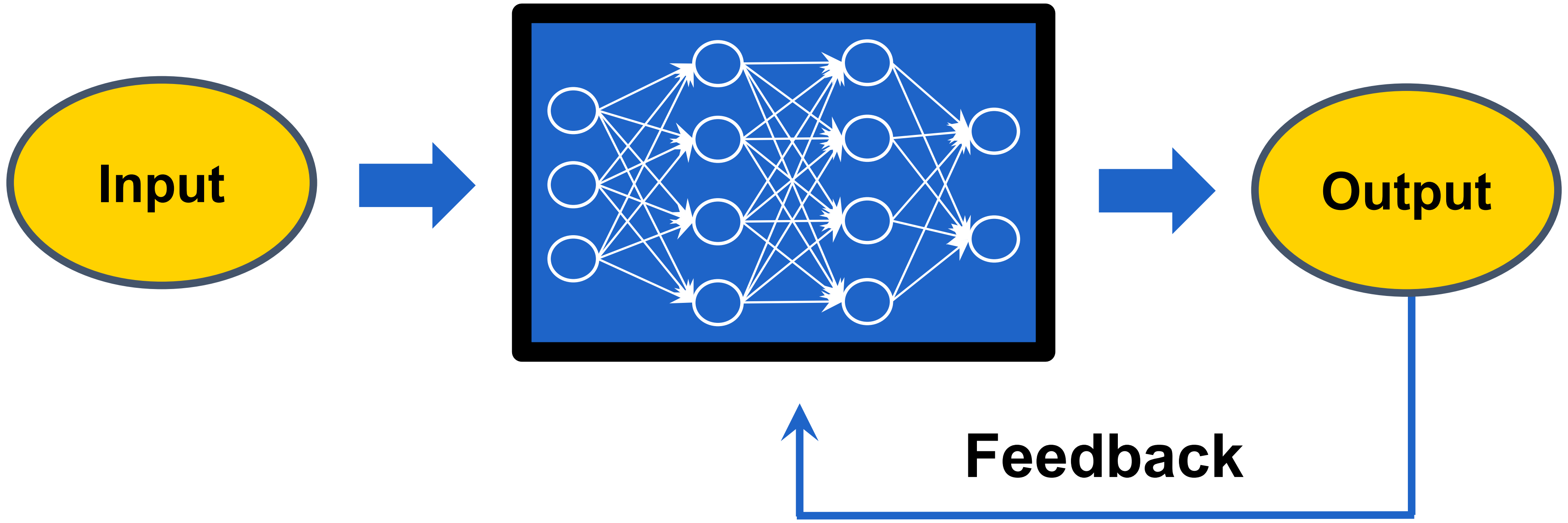
2020

Prof. Jae Young Choi

DEEP LEARNING BASICS

DEEP MACHINE LEARNING

- **Algorithms able to automatically construct expert knowledge**
  - typically by applying artificial neural networks (NNs) to huge amounts of data



Self-learning, black-box systems

- **Alternative to manual construction of expert knowledge**
  - time consuming and thus expensive

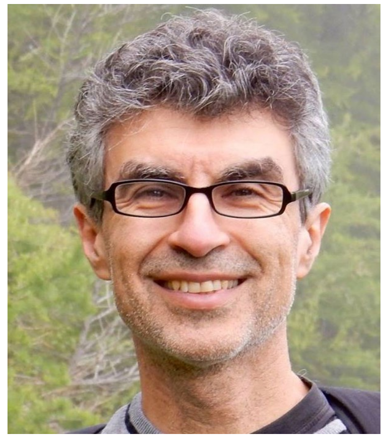
Slide credit: W. D. Neve

POPULARITY OF DEEP LEARNING

Geoff Hinton

Andrew Ng

Yann LeCun



Yoshua Bengio

Intel is paying more than **$400 million** to buy deep-learning startup Nervana Systems (Aug 2016)

Apple Acquires Machine Learning Startup Turi For **$200 Million** (Aug 2016)

Google Acquires Artificial Intelligence Startup DeepMind For More Than **$500M** (Dec 2014)

Deep Learning Enterprise Software Spending to Surpass **$40 Billion** Worldwide by 2024 (May 2016)

Twitter pays up to **$150M** for Magic Pony Technology, which uses neural networks to improve images (Jun 2016)

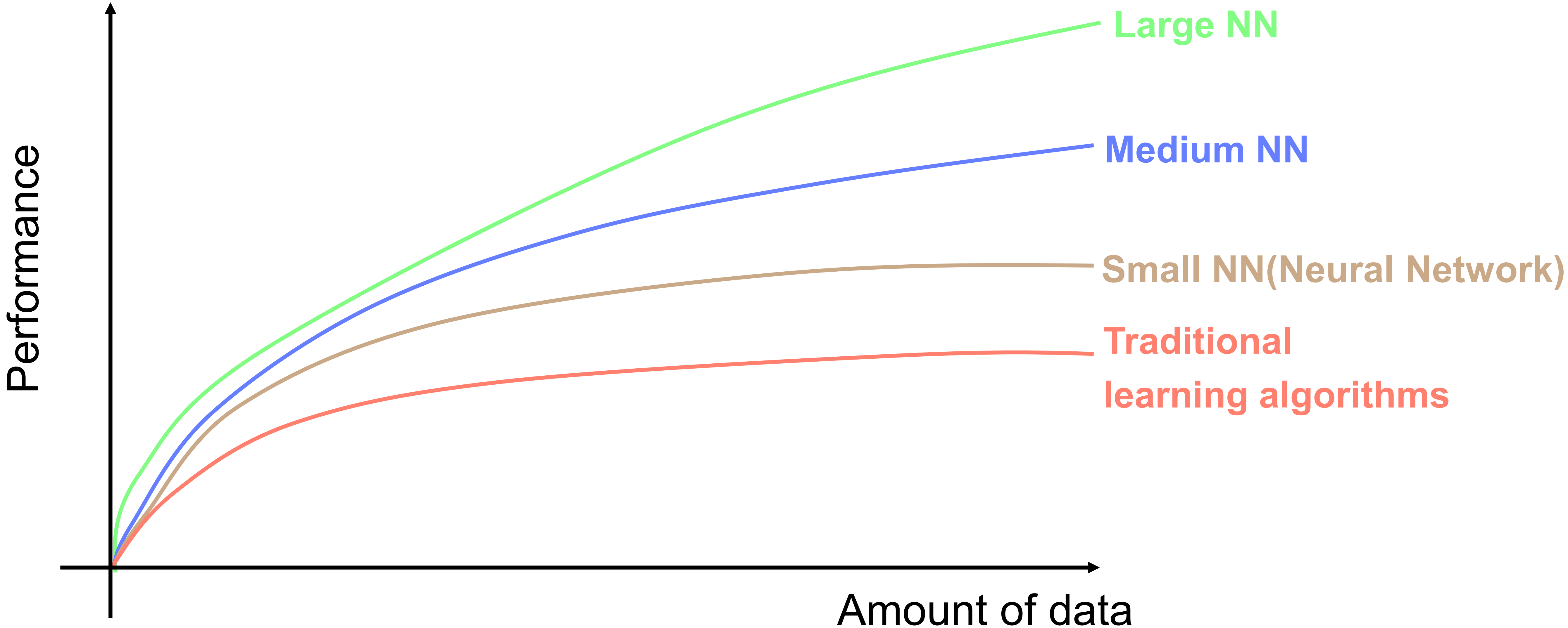
Slide credit: W. D. Neve

WHY DEEP LEARNING?

- **Availability of cheap and massive computational power**
  - GPU computing
  - cloud computing
- **Availability of large data sets**
  - social media applications
  - sensor output (Internet of Things / Internet of Services)
- **New algorithmic techniques**
  - dropout
  - rectified linear units (ReLU)
  - Layer-wise training

Slide credit: W. D. Neve

SCALE DRIVING DEEP LEARNING PROGRESS



THE RISE OF END-TO-END LEARNING

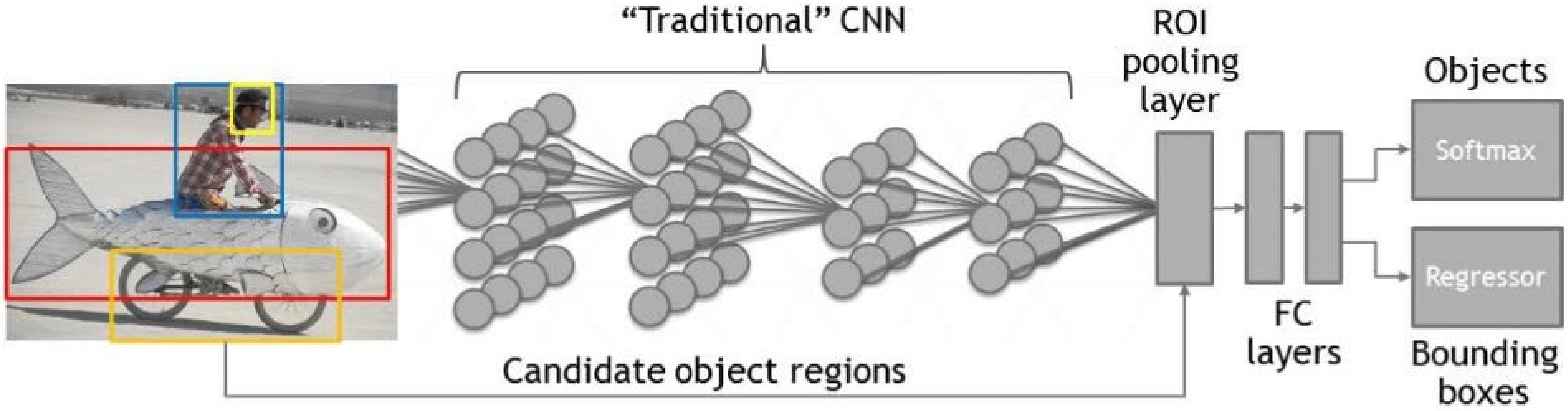
Learning with integer or real-valued outputs:

Problem	X	Y
Spam classification	Email	Spam/Not spam(0/1)
Image recognition	Image	Integer label
Housing price prediction	Features of house	Price in dollars
Product recommendation	Product & user features	Chance of purchase

Learning with complex (e.g., string valued) outputs:

Problem	X	Y	Example
Image captioning	Image	Text	Mao et al., 2014
Machine translation	English text	French text	Suskever et al., 2014
Question answering	(Text,Question) pair	Answer text	Bordes et al., 2015
Speech recognition	Audio	Transcription	Hannun et al., 2015
TTS(Texture-To-Speech)	Text features	Audio	van der Oord et al., 2016

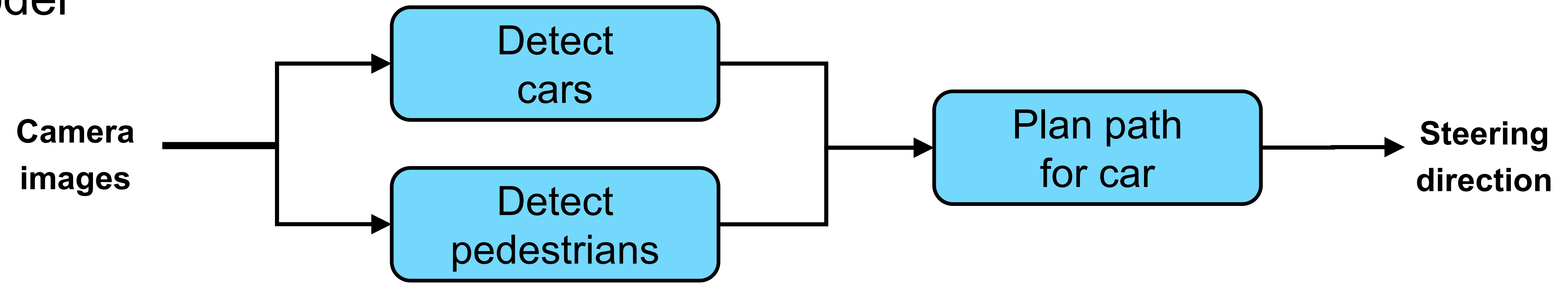
THE RISE OF END-TO-END LEARNING EXAMPLE APPLICATIONS : LOCALIZATION



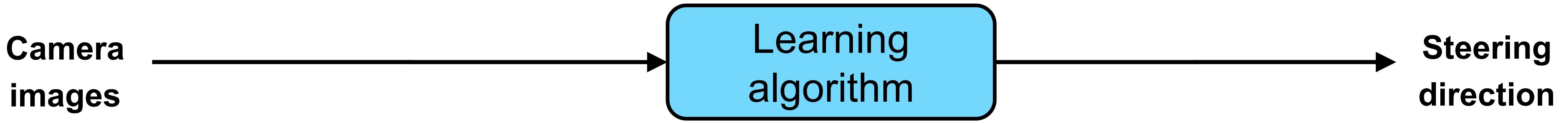
(Fast) Region based Convolutional Networks (R-CNN)
Ross Girshick, Microsoft Research
https://github.com/rbgirshick/fast-rcnn

END-TO-END LEARNING: AUTONOMOUS DRIVING

Traditional model



End-to-end learning

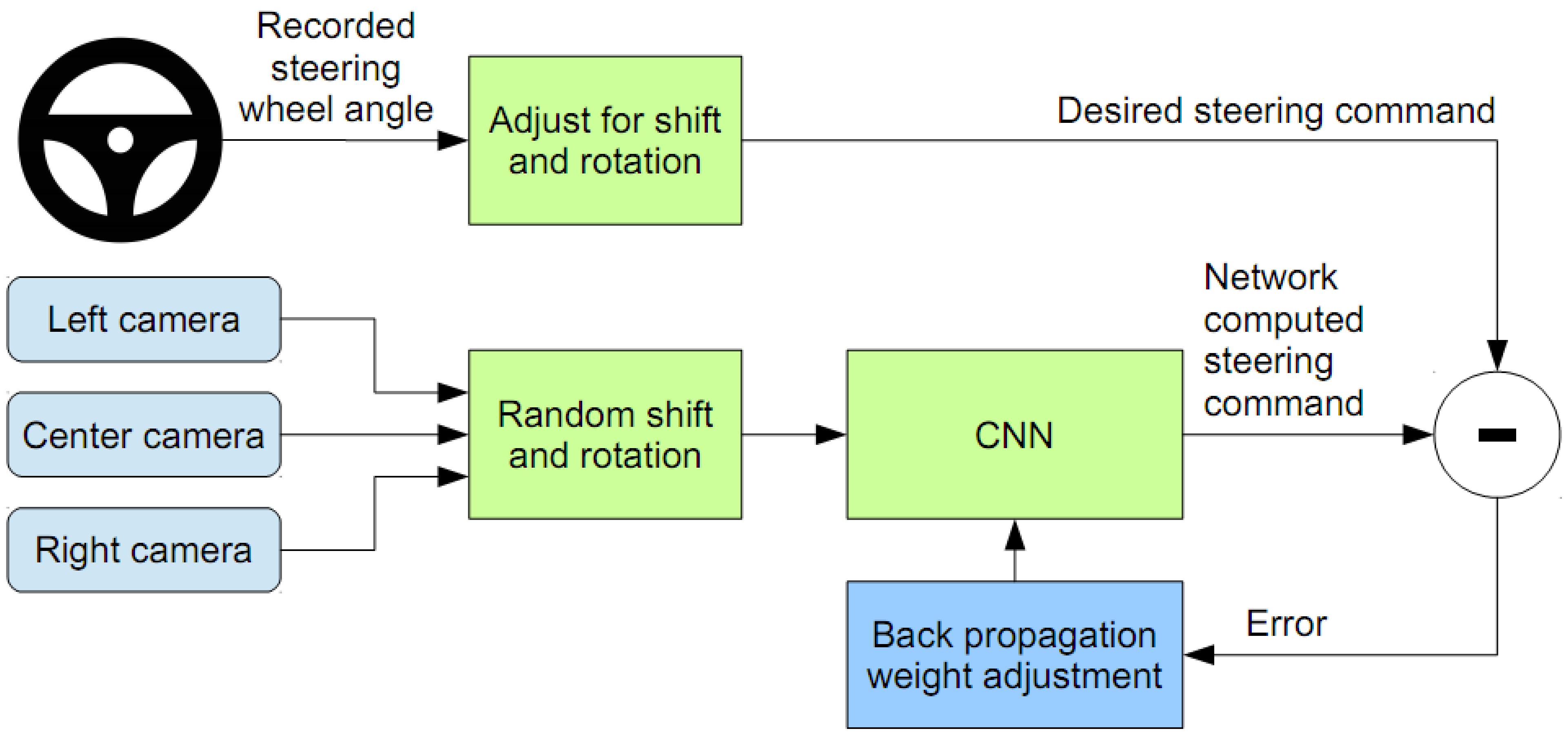


Given the safety-critical requirement of autonomous driving and thus the need for extremely high levels of accuracy, a pure end-to-end approach is still challenging to get to work

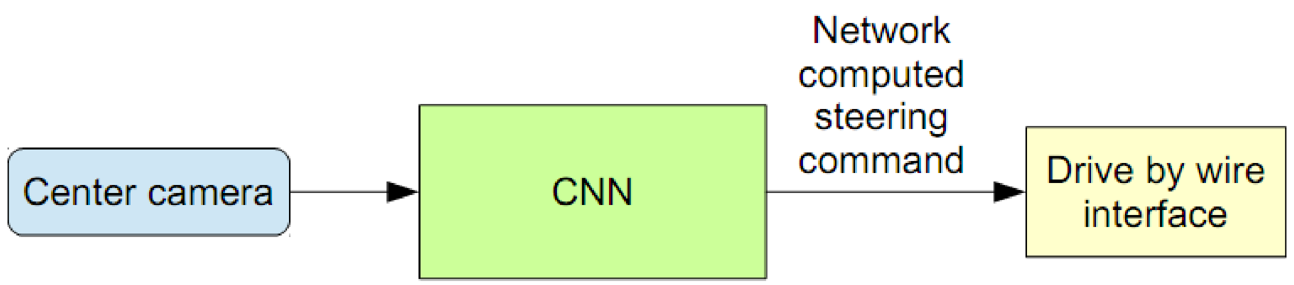
End-to-end works only when you have enough (x,y) data to learn function of needed level of complexity

END-TO-END LEARNING: AUTONOMOUS DRIVING

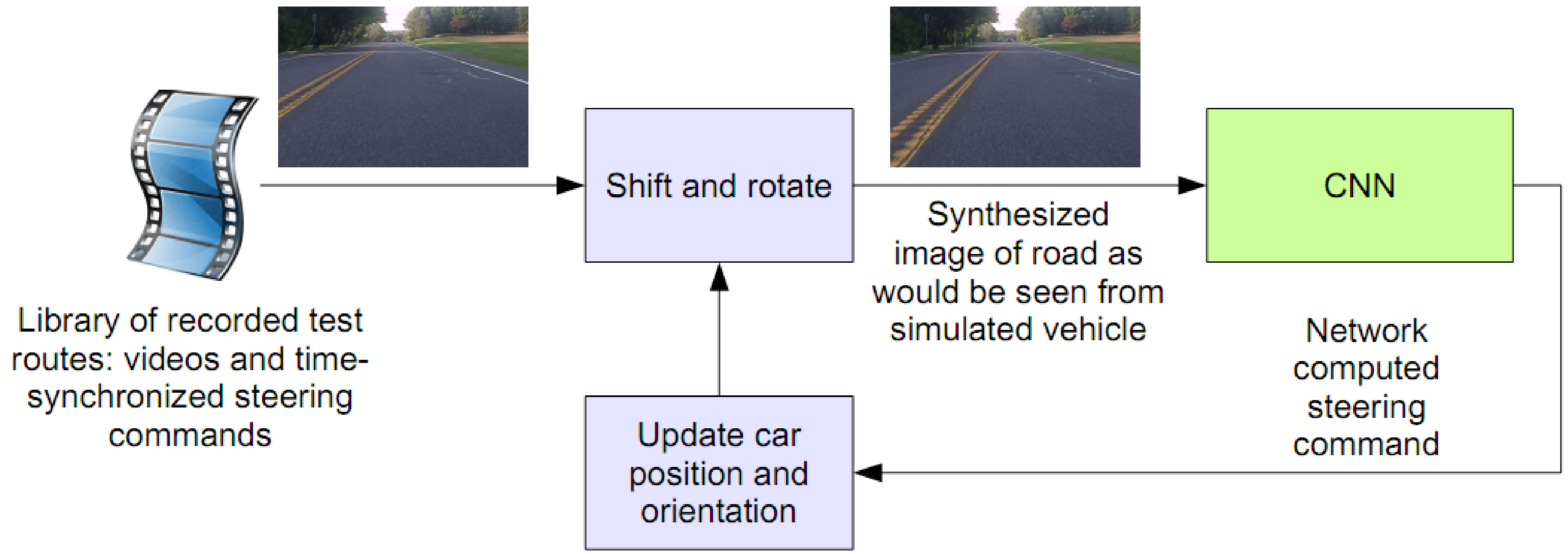
End to End Learnin for Self-Driving Cars.PDF



Training convolutional neural network



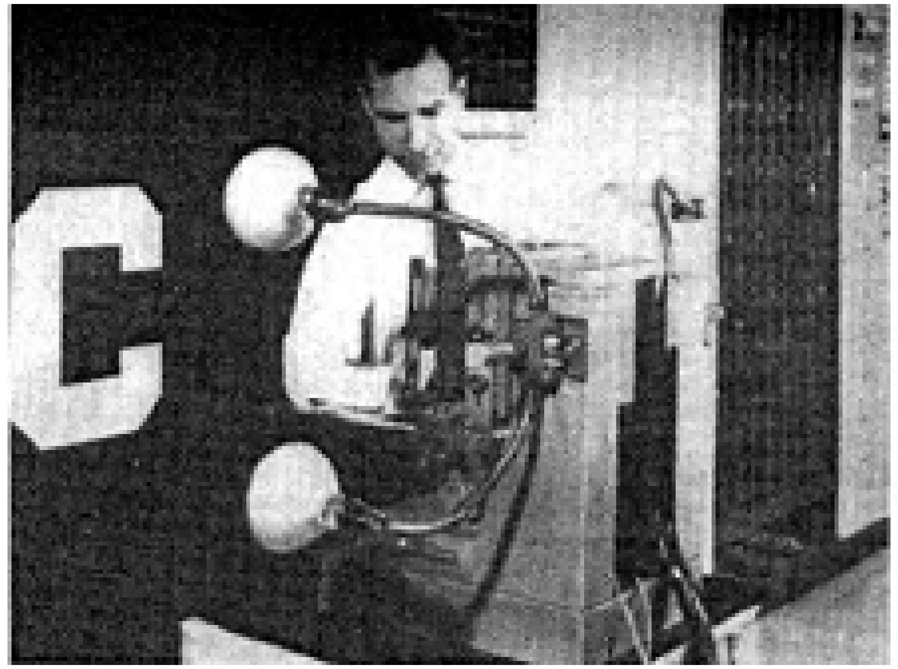
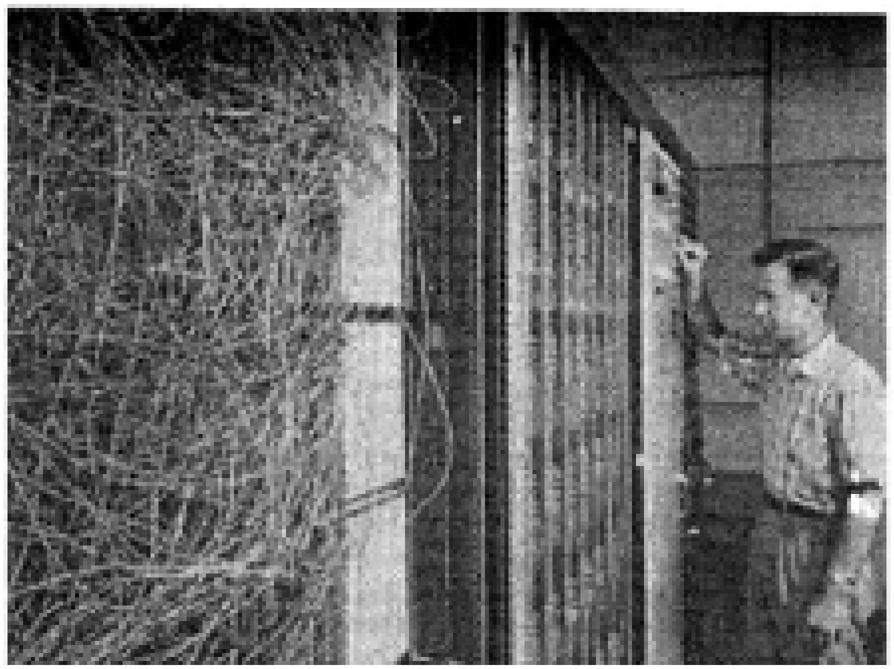
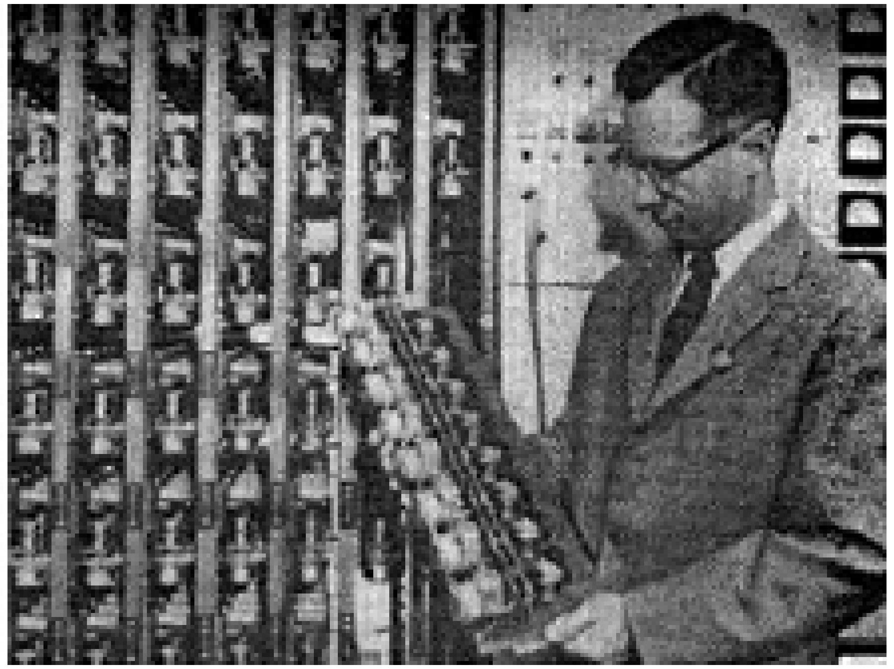
Trained network is used to generate steering commands



Driver simulator

Neuroscience inspired early works on Machine Learning & AI

- The perceptron (Frank Rosenblatt at Cornell University, 1957

- The traditional model of pattern recognition (since the late 50’s)
  - Fixed/engineered features (or fixed kernel) + trainable classifier

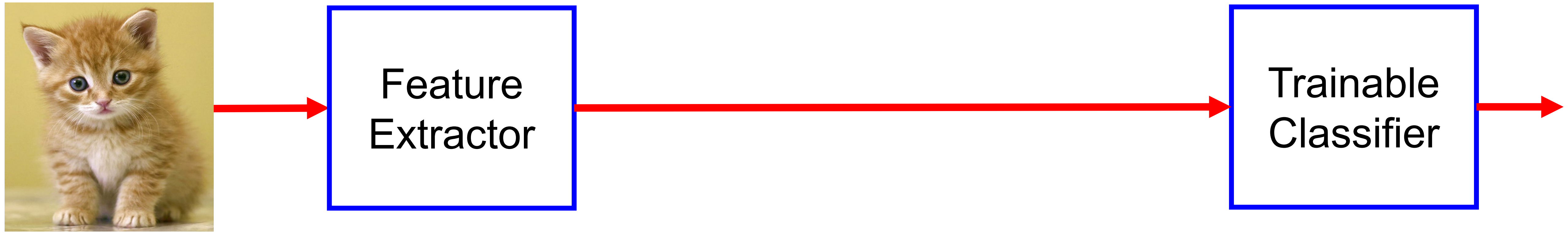
hand-crafted
Feature Extractor

“Simple” Trainable
Classifier

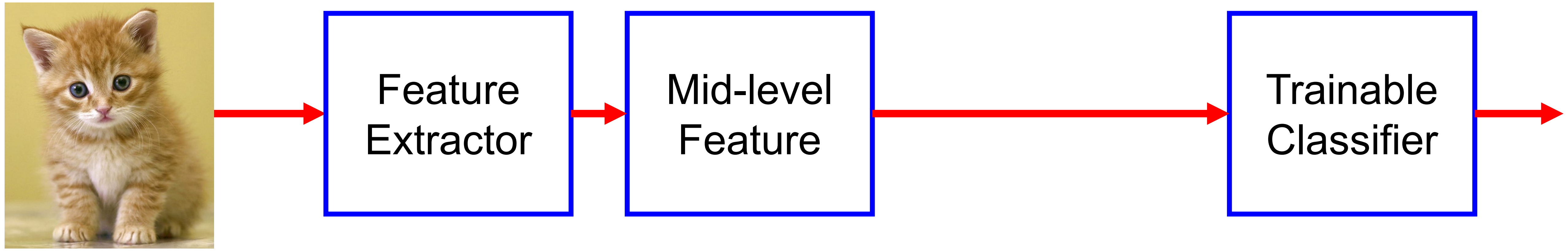
Slide credit: Y. LeCun

Deep Learning = The entire Machine is Trainable

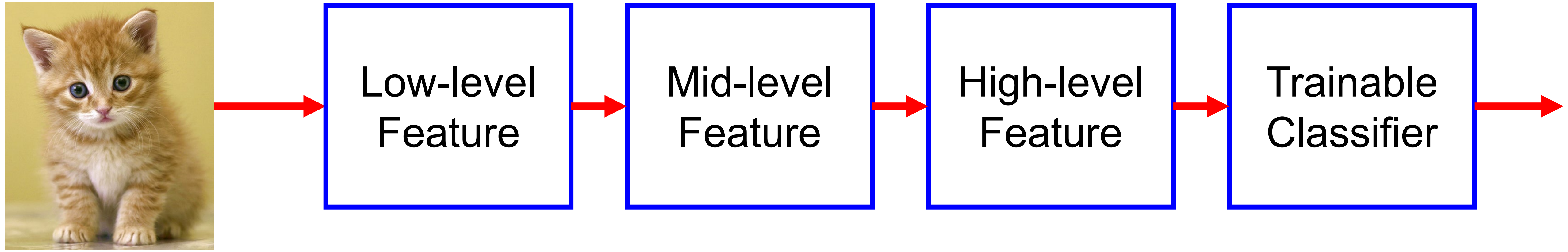
- Traditional Pattern Recognition: Fixed/Handcrafted Feature Extractor



- Mainstream Modern Pattern Recognition: Unsupervised mid-level feature



- Deep Learning: Representations are hierarchical and trained

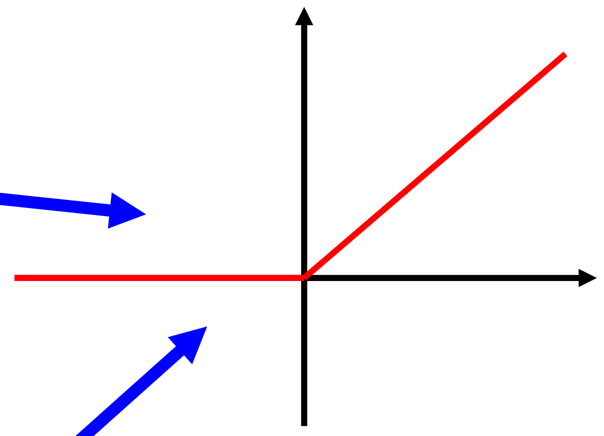


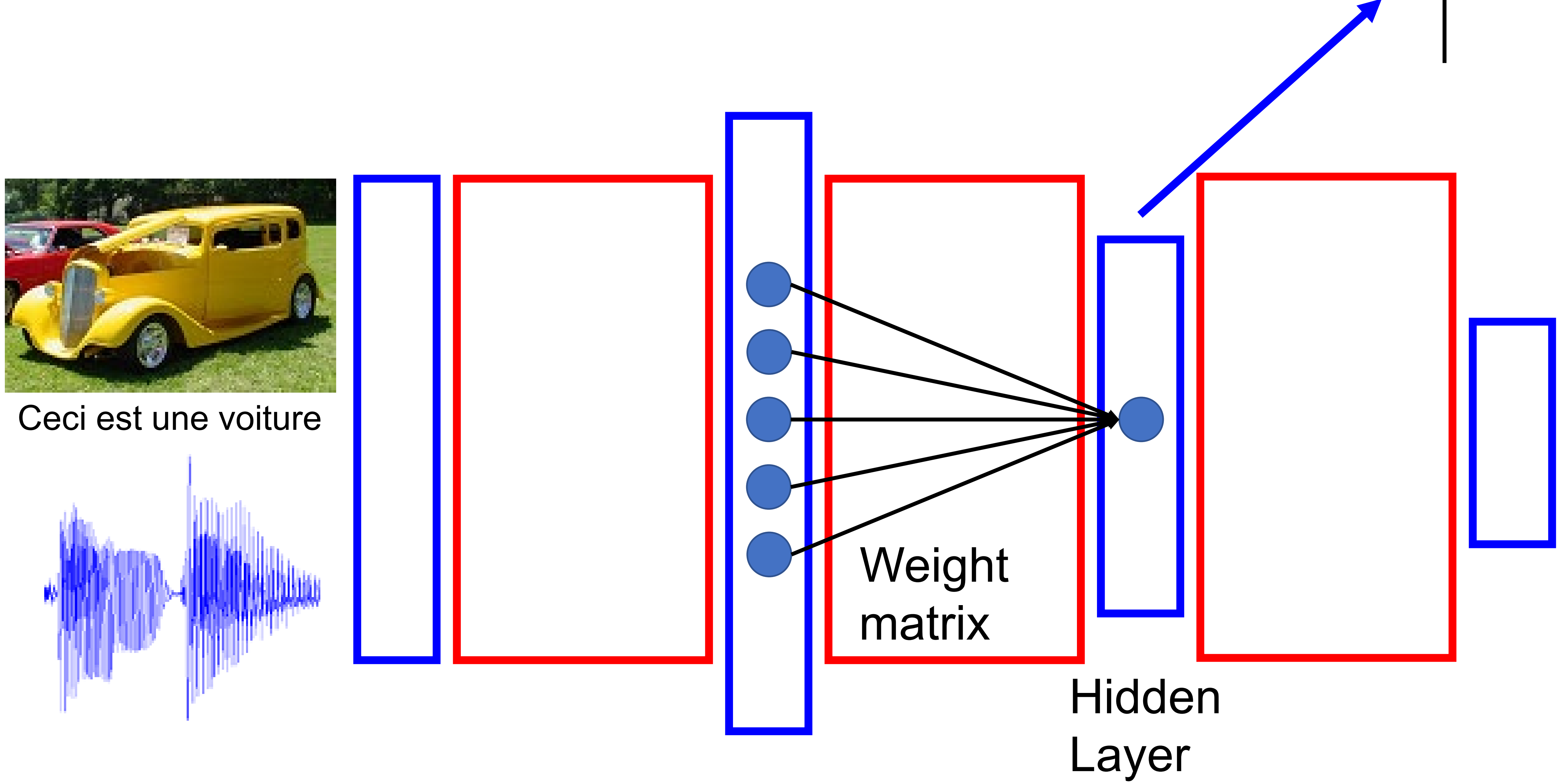
Slide credit: Y. LeCun

Multi-Layer Neural Nets

- Multiple Layers of simple units
- Each units computes a weighted sum of its inputs
- Weighted sum is passed through a non-linear function
- The learning algorithm changes the weights

$ReLU(x) = max(x, 0)$

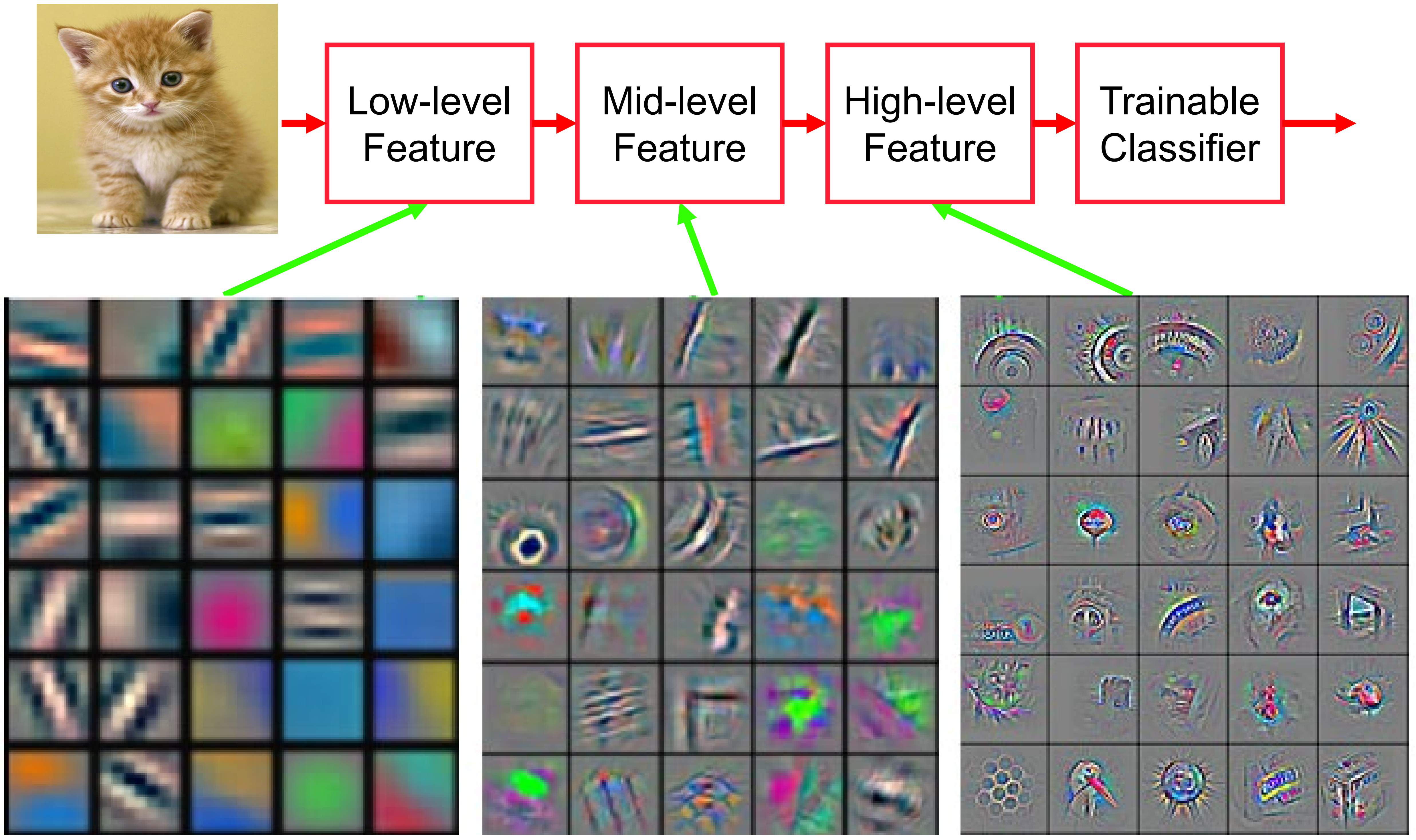




Slide credit: Y. LeCun

Deep Learning = Learning Hierarchical Representations

- It's deep if it has more than one stage of non-linear feature transformation

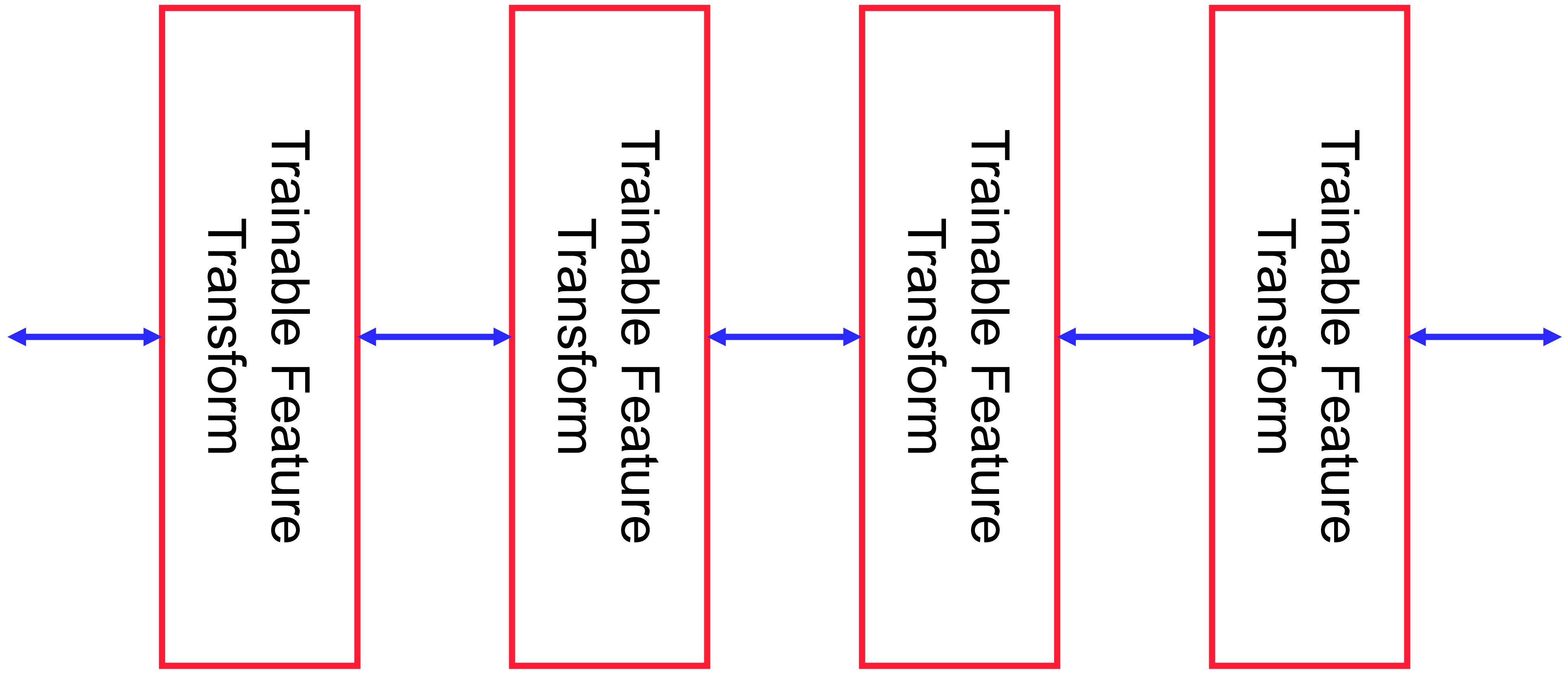


Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Slide credit: Y. LeCun

Trainable Feature Hierarchy

- Hierarchy of representations with increasing level of abstraction
- Each stage is a kind of trainable feature transform
- Image recognition
  - Pixel → edge → texton → motif → part → object
- Text
  - Character → word → word group → clause → sentence → story
- Speech
  - Sample → spectral band → sound → phoneme → word

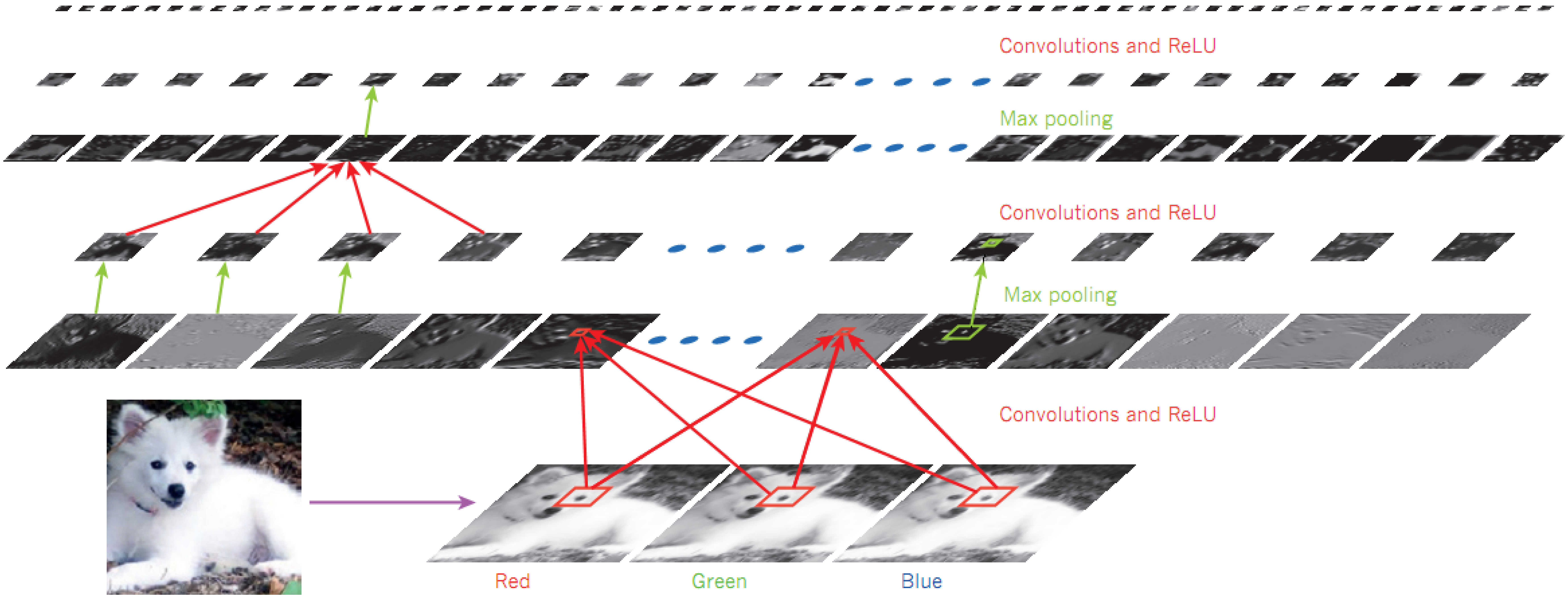


Slide credit: Y. LeCun

Milestone Paper for Deep Learning

Article *in* Nature

2015-lecun.pdf

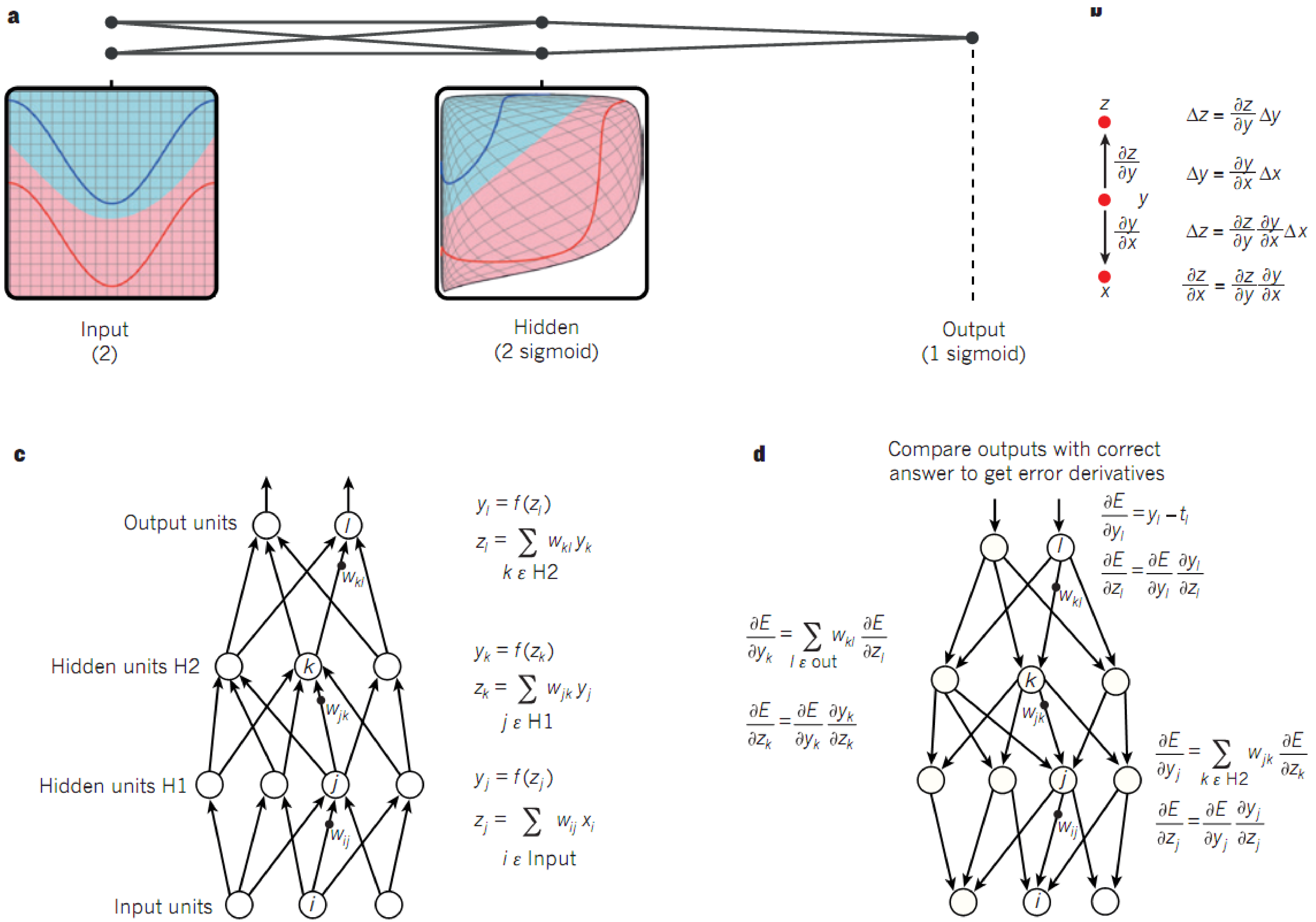


Inside a convolutional network

Slide credit: Y. LeCun

Milestone Paper for Deep Learning

Article *in* Nature



Multilayer neural networks and backprogation

Slide credit: Y. LeCun