Introduction

M. Soleymani
Sharif University of Technology
Fall 2017

Course Info

- Course Number: 40-959 (Time: Sun-Tue 13:30-15:00 Location: CE 103)
- Instructor: Mahdieh Soleymani (soleymani@sharif.edu)
- TAs:
  - Mahsa Ghorbani (Head TA)
  - Seyed Ali Osia
  - Sarah Rastegar
  - Alireza Sahaf
  - Seyed Mohammad Chavoshian
  - Zeynab Golgooni
- Website: http://ce.sharif.edu/cources/96-97/1/ce979-1
- Office hours: Tuesdays 15:00-16:00

Materials

- Text book: Ian Goodfellow, Yoshua Bengio and Aaron Courville, *Deep Learning*, Book in preparation for MIT Press, 2016.

- Papers

- Notes, lectures, and demos

Marking Scheme

• Midterm Exam:	25%
• Final Exam:	30%
• Project:	5-10%
• Homeworks (written & programming) :	25-30%
• Mini-exams:	10%

Prerequisites

- Machine Learning
- Knowledge of calculus and linear algebra
- Probability and Statistics
- Programming (Python)

This Course

Goals:

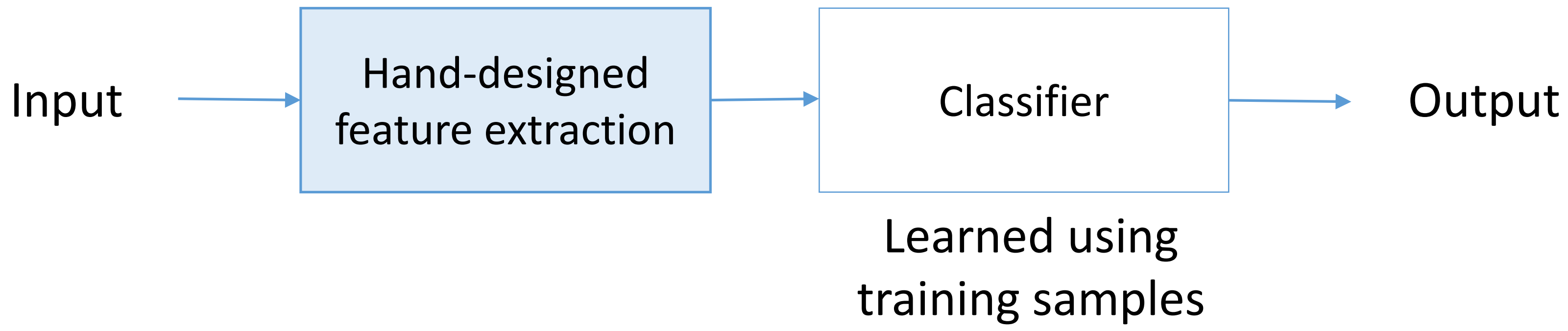
- Review principles and introduce fundamentals for understanding deep networks.
- Introduce several popular networks and training issues
- Develop skill at designing architectures for applications.

Deep Learning

- Learning a computational models consists of multiple processing layers
  - learn representations of data with multiple levels of abstraction.
- Dramatically improved the state-of-the-art in many speech, vision and NLP tasks (and also in many other domains like bioinformatics)

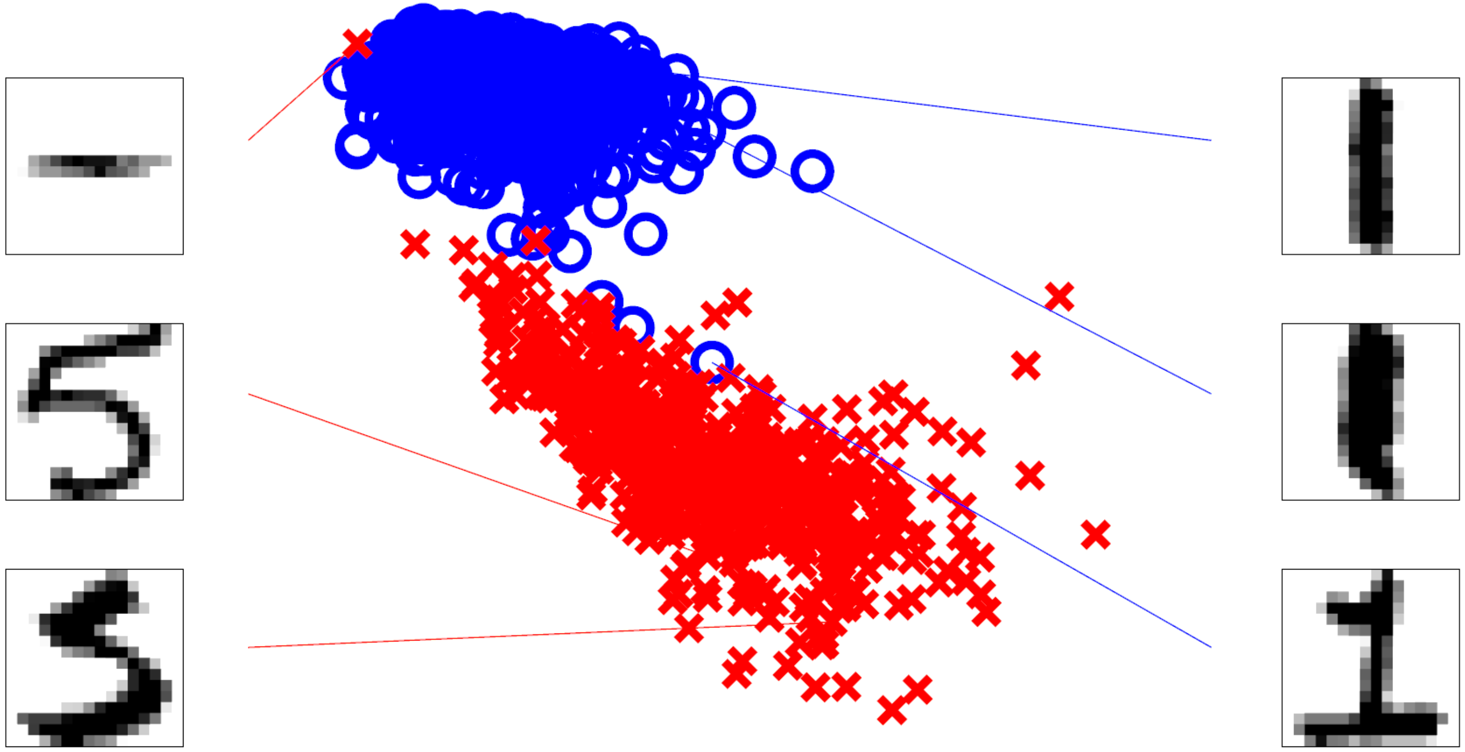
Machine Learning Methods

- Conventional machine learning methods:
  - try to learn the mapping from the input features to the output by samples
  - However, they need appropriately designed hand-designed features



Example

- x_1: intensity
- x_2: symmetry



[Abu Mostafa, 2012]

Representation of Data

- Performance of traditional learning methods depends heavily on the representation of the data.
  - **Most efforts were on designing proper features**

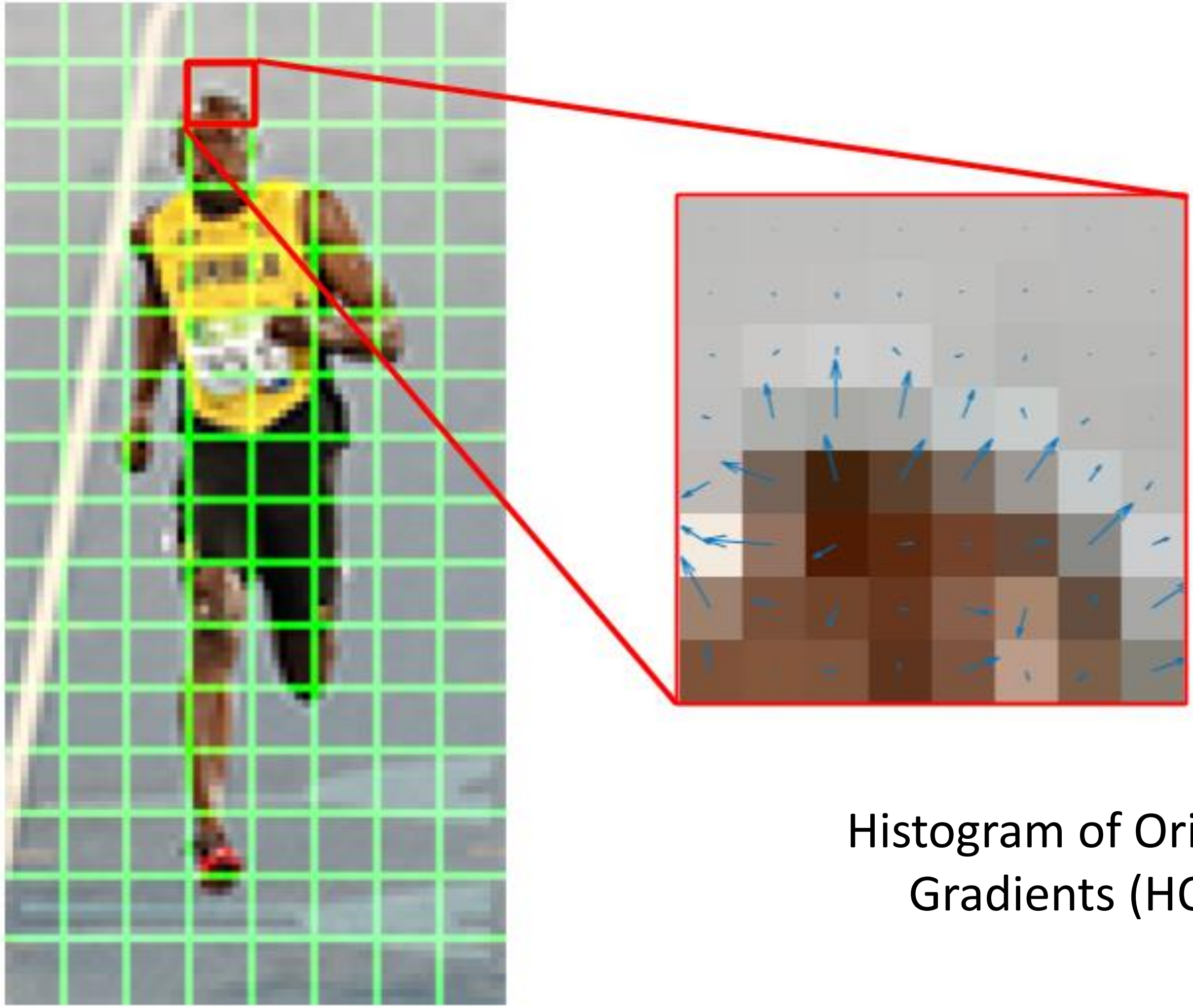
- However, designing hand-crafted features for inputs like image, videos, time series, and sequences is not trivial at all.
  - It is difficult to know which features should be extracted.
    - Sometimes, it needs long time for a community of experts to find (an incomplete and over-specified) set of these features.

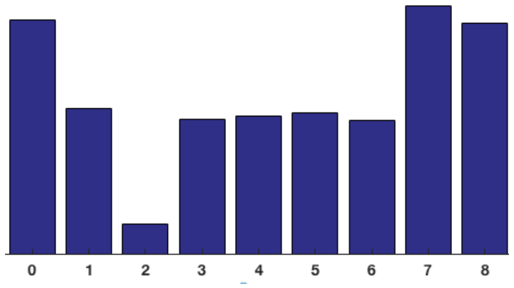
Hand-designed Features Example: Object Recognition

- Multitude of hand-designed features currently in use
  - e.g., SIFT, HOG, LBP, DPM

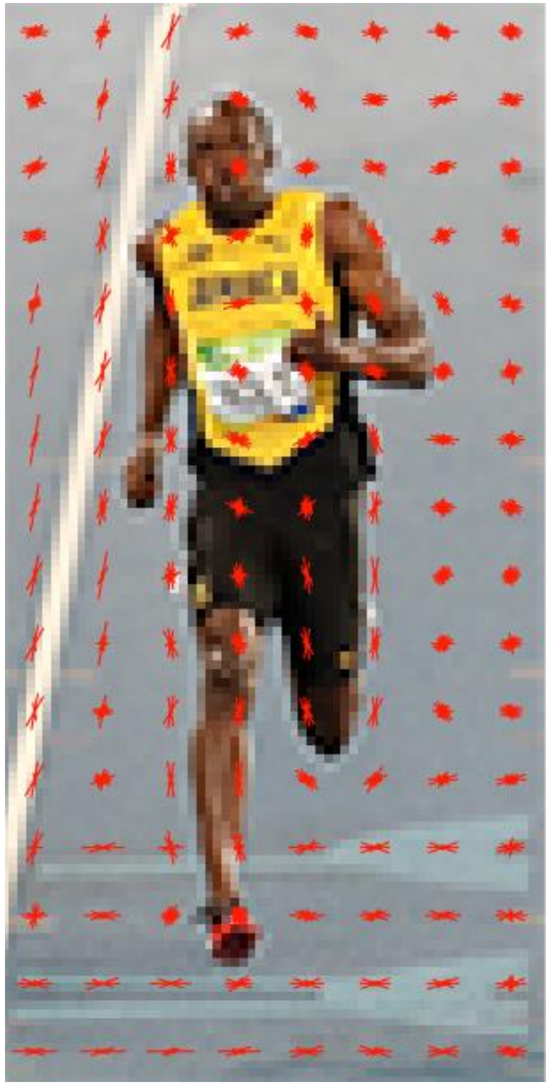
- These are found after many years of research in image and computer vision areas

Hand-designed Features Example: Object Recognition





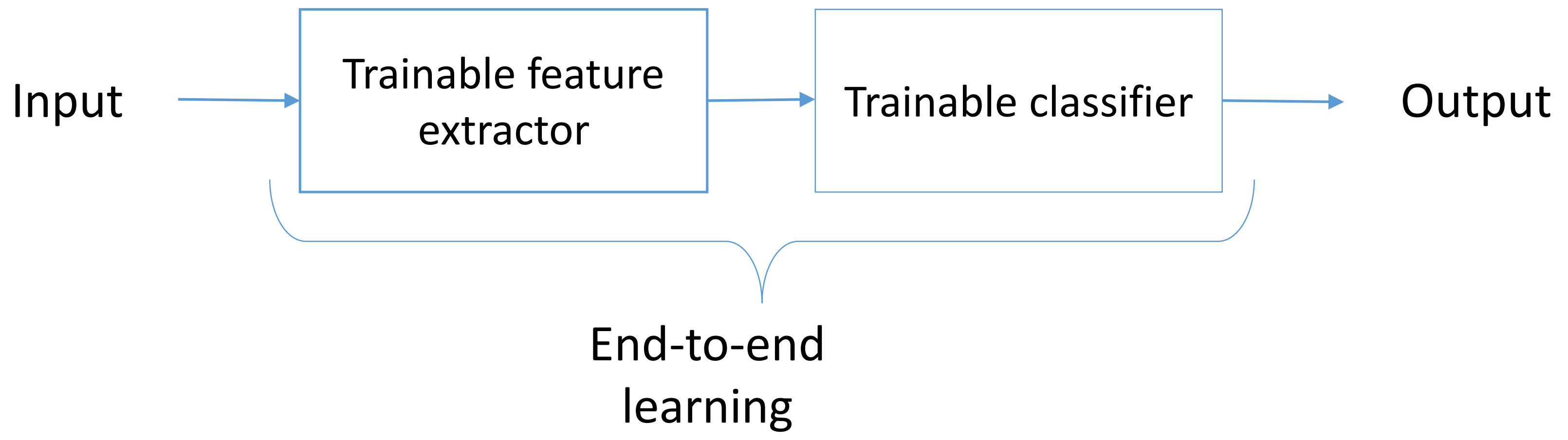
Histogram of Oriented
Gradients (HOG)



Source: http://www.learnopencv.com/histogram-of-oriented-gradients/

Representation Learning

- Using learning to discover both:
  - the representation of data from input features
  - and the mapping from representation to output



Previous Representation Learning Methods

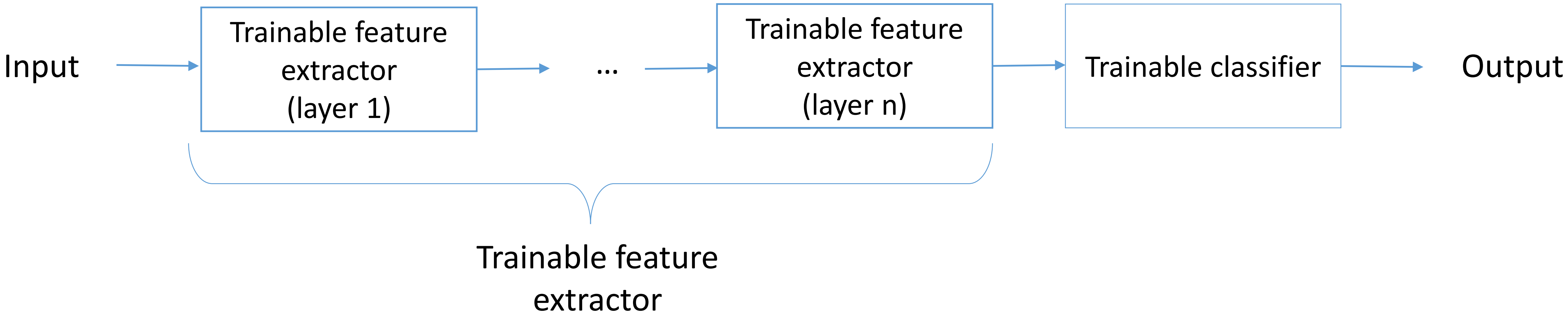
- Although metric learning and kernel learning methods attempted to solve this problem, they were shallow models for feature (or representation) learning

- Deep learning finds representations that are expressed in terms of other, simpler representations
  - Usually hierarchical representation is meaningful and useful

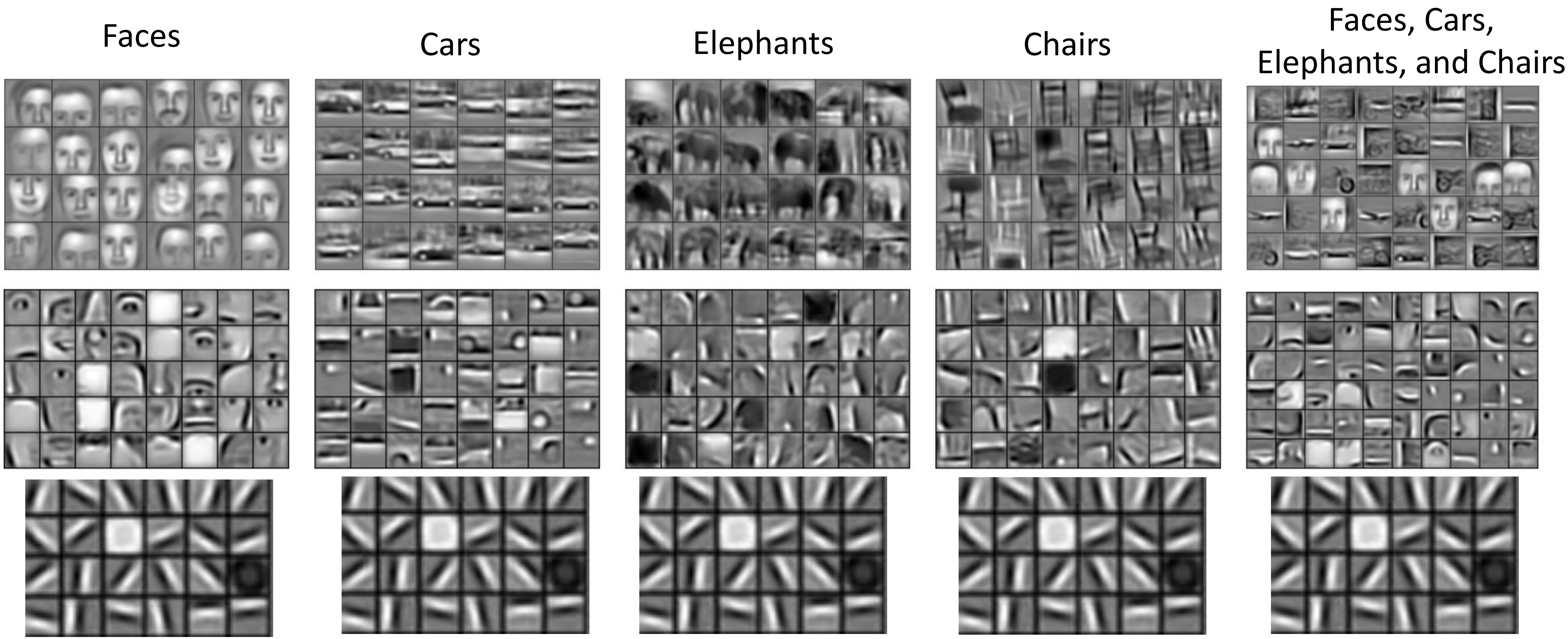
Deep Learning Approach

- Deep breaks the desired complicated mapping into a series of nested simple mappings
  - each mapping described by a layer of the model.
  - each layer extracts features from output of previous layer

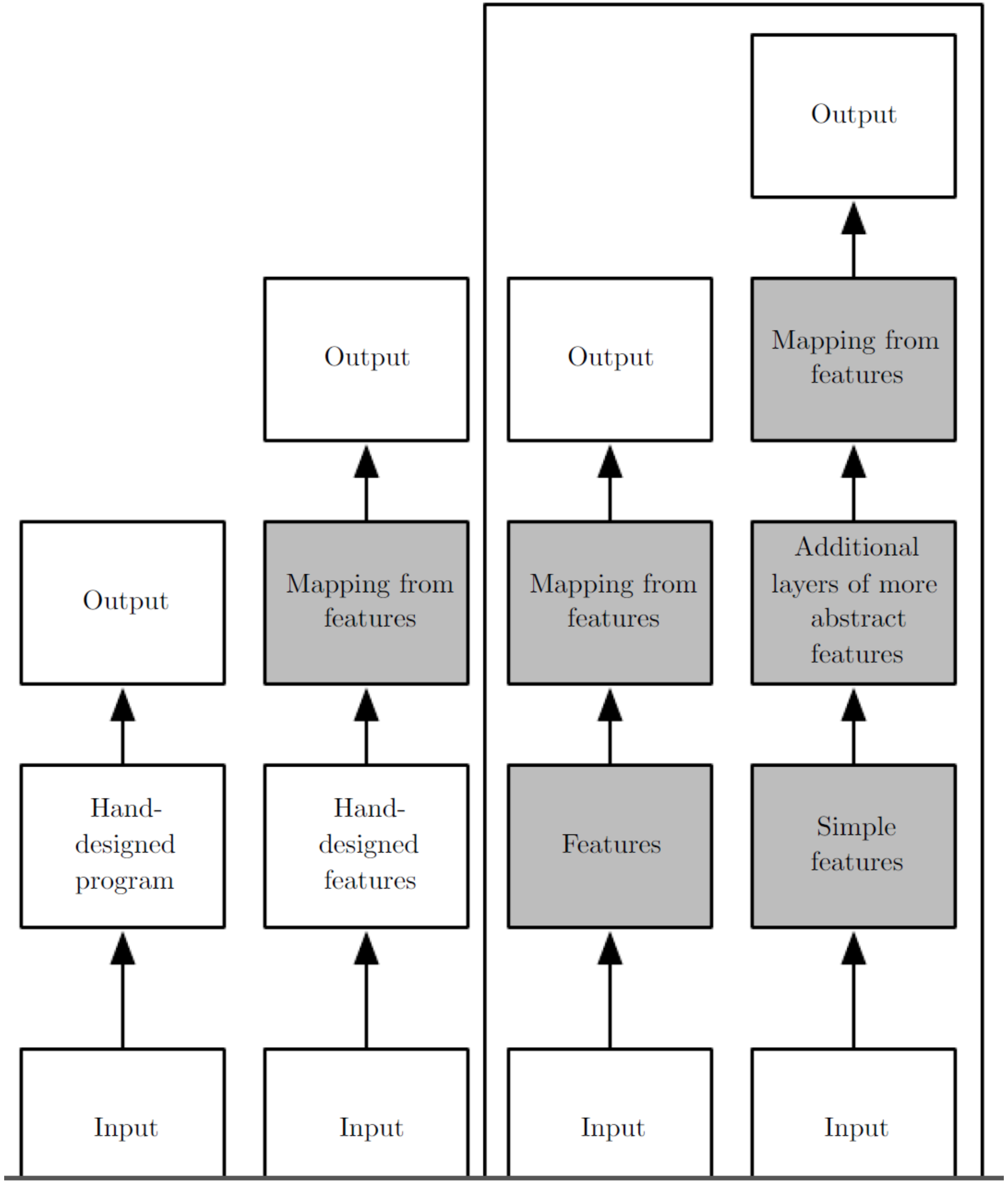
- shows impressive performance on many Artificial Intelligence tasks



Example of Nested Representation



[Lee et al., ICML 2009]



[Deep Learning book]

Deep Representations: The Power of Compositionality

- Compositionality is useful to describe the world around us efficiently
  - Learned function seen as a composition of simpler operations
  - Hierarchy of features, concepts, leading to more abstract factors enabling better generalization
    - each concept defined in relation to simpler concepts
    - more abstract representations computed in terms of less abstract ones.
  - Again, theory shows this can be exponentially advantageous

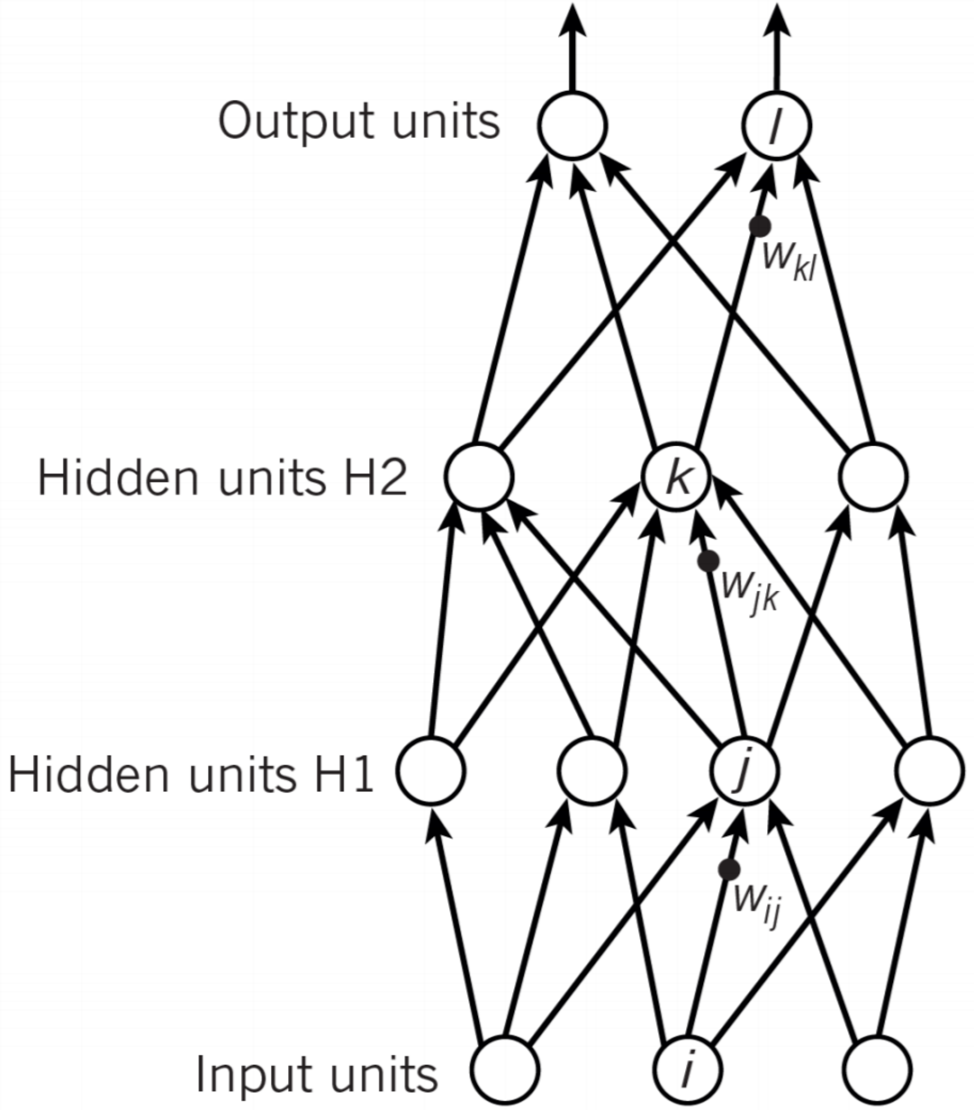
- Deep learning has great power and flexibility by learning to represent the world as a nested hierarchy of concepts

This slide has been adopted from: http://www.ds3-datascience-polytechnique.fr/wp-content/uploads/2017/08/2017_08_28_1000-1100_Yoshua_Bengio_DeepLearning_1.pdf

Feed-forward Networks or MLPs

- A multilayer perceptron is just a mapping input values to output values.
  - The function is formed by composing many simpler functions.
  - These middle layers are not given in the training data must be determined

Multi-layer Neural Network



$$y_l = f(z_l)$$

$$z_l = \sum_{k \,\varepsilon\, \mathrm{H2}} w_{kl}\, y_k$$

$$y_k = f(z_k)$$

$$z_k = \sum_{j \,\varepsilon\, \mathrm{H1}} w_{jk}\, y_j$$

$$y_j = f(z_j)$$

$$z_j = \sum_{i \,\varepsilon\, \mathrm{Input}} w_{ij}\, x_i$$

Example of f functions:

$$f(z) = \max(0, z)$$

[Deep learning, Yann LeCun, Yoshua Bengio, Geoffrey Hinton, Nature 521, 436–444, 2015]

Training Multi-layer Neural Networks

- Backpropagation algorithm indicate to change parameters
  - Find parameters that are used to compute the representation in each layer

- Using large data sets for training, deep learning can discover intricate structures

Deep Learning Brief History

- 1940s–1960s:
  - development of theories of biological learning
  - implementations of the first models
    - perceptron (Rosenblatt, 1958) for training of a single neuron.

- 1980s-1990s: back-propagation algorithm to train a neural network with more than one hidden layer
  - too computationally costly to allow much experimentation with the hardware available at the time.

- 2006 “Deep learning” name was selected
  - ability to train deeper neural networks than had been possible before
    - Although began by using unsupervised representation learning, later success obtained usually using large datasets of labeled samples

Why does deep learning become popular?

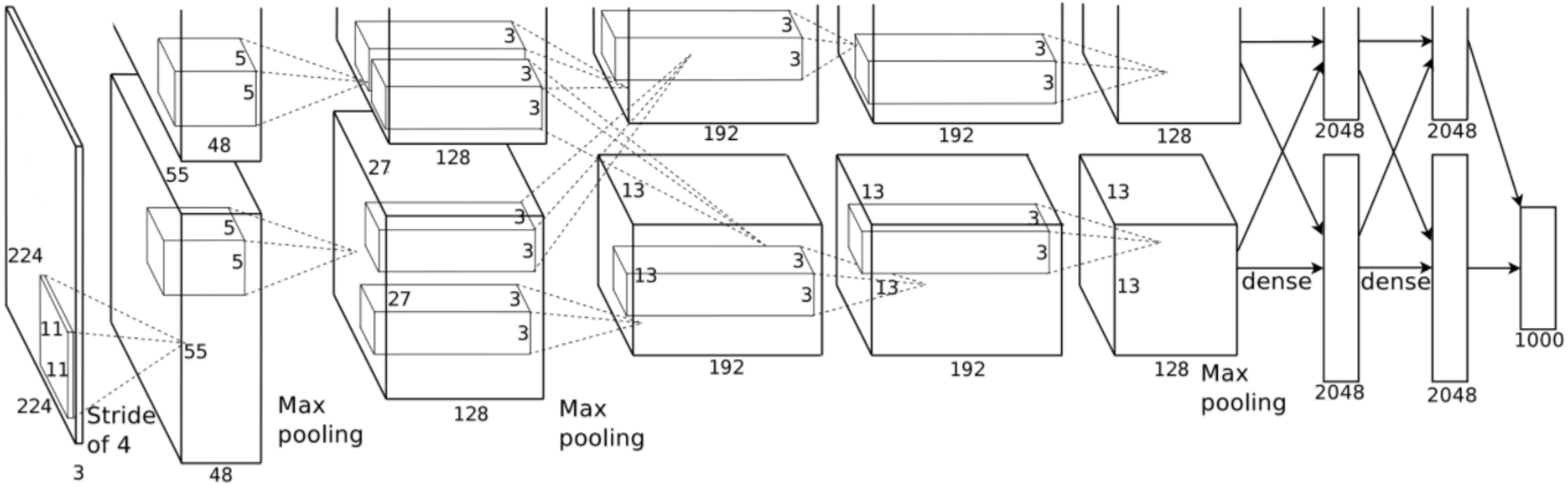
- Large datasets
- Availability of the computational resources to run much larger models
- New techniques to address the training issues

ImageNet

[Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009]

- 22K categories and 14M images
  - Collected from web & labeled by Amazon Mechanical Turk
- The Image Classification Challenge:
  - 1,000 object classes
  - 1,431,167 images
- Much larger than the previous datasets of image classification

Alexnet (2012)



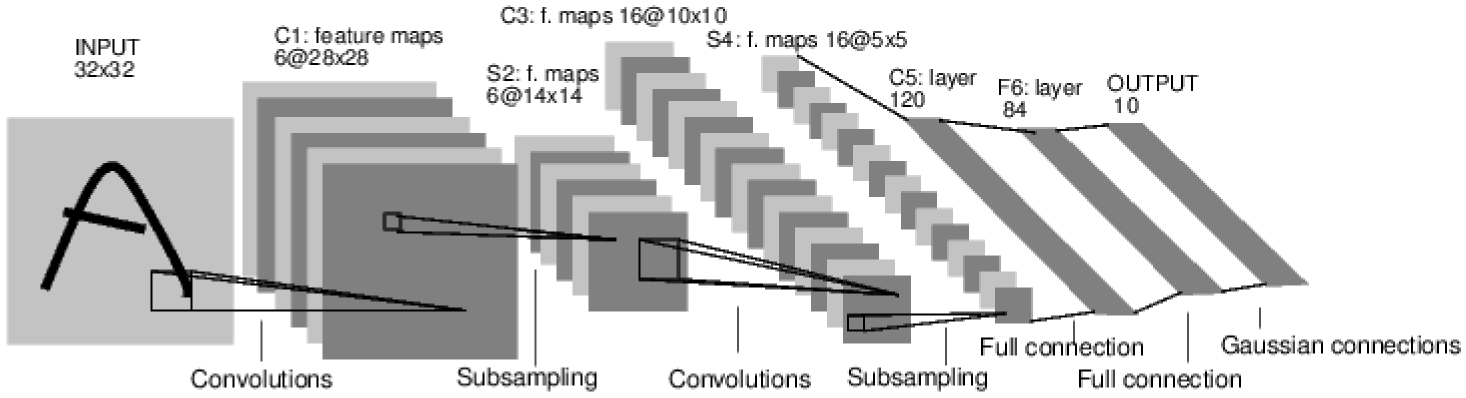
[Krizhevsky, Alex, Sutskever, and Hinton, Imagenet classification with deep convolutional neural networks, NIPS 2012]

- Reduces 25.8% top 5 error of the winner of 2011 challenge to 16.4%

CNN for Digit Recognition as origin of AlexNet

LeNet: Handwritten Digit Recognition (recognizes zip codes)

Training Sample : 9298 zip codes on mails



[LeNet, Yann Lecun, et. al, 1989]

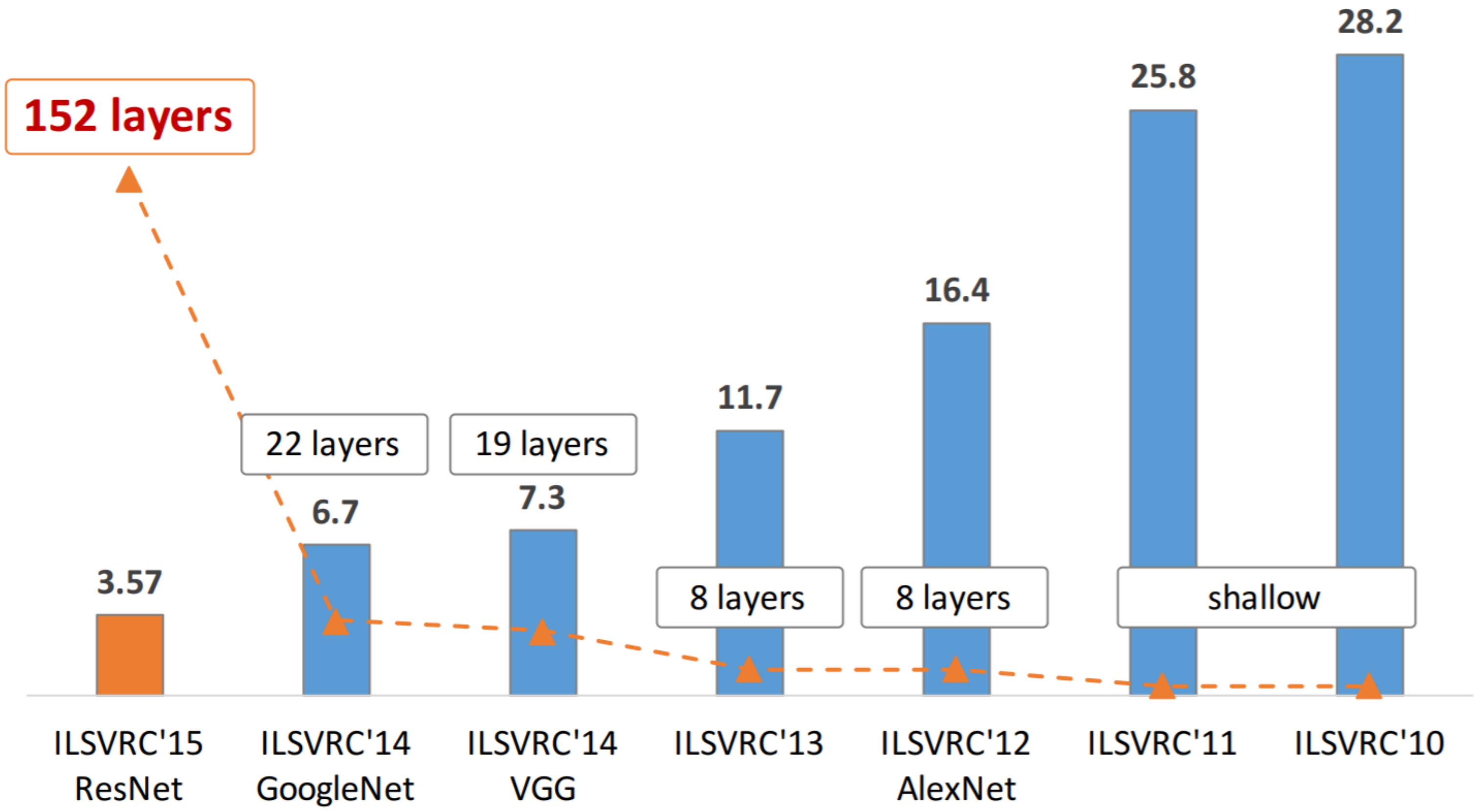
AlexNet Success

- Trained on a large labeled image dataset

- ReLU instead of sigmoids, enable training much deeper networks by backprop

Deeper Models Work Better

- 5.1% is the performance of human on this data set



ImageNet Classification top-5 error (%)

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". CVPR 2016.

Using Pre-trained Models

- We don’t have large-scale datasets on all image tasks and also we may not time to train such deep networks from scratch

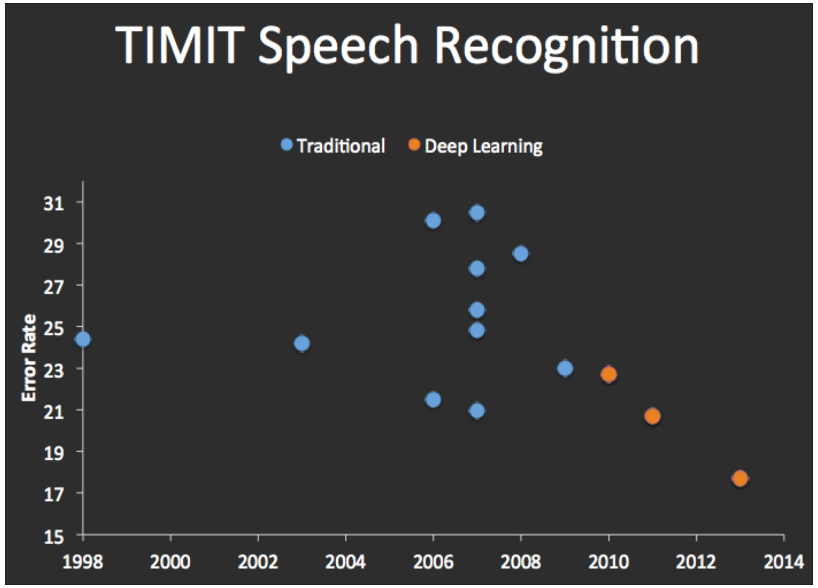
- On the other hand, learned weights for popular networks (on ImageNet) are available.

- Use pre-trained weights of these networks (other than final layers) as generic feature extractors for images

- Works better than handcrafted feature extraction on natural images

Speech Recognition

- The introduction of deep learning to speech recognition resulted in a sudden drop of error rates.

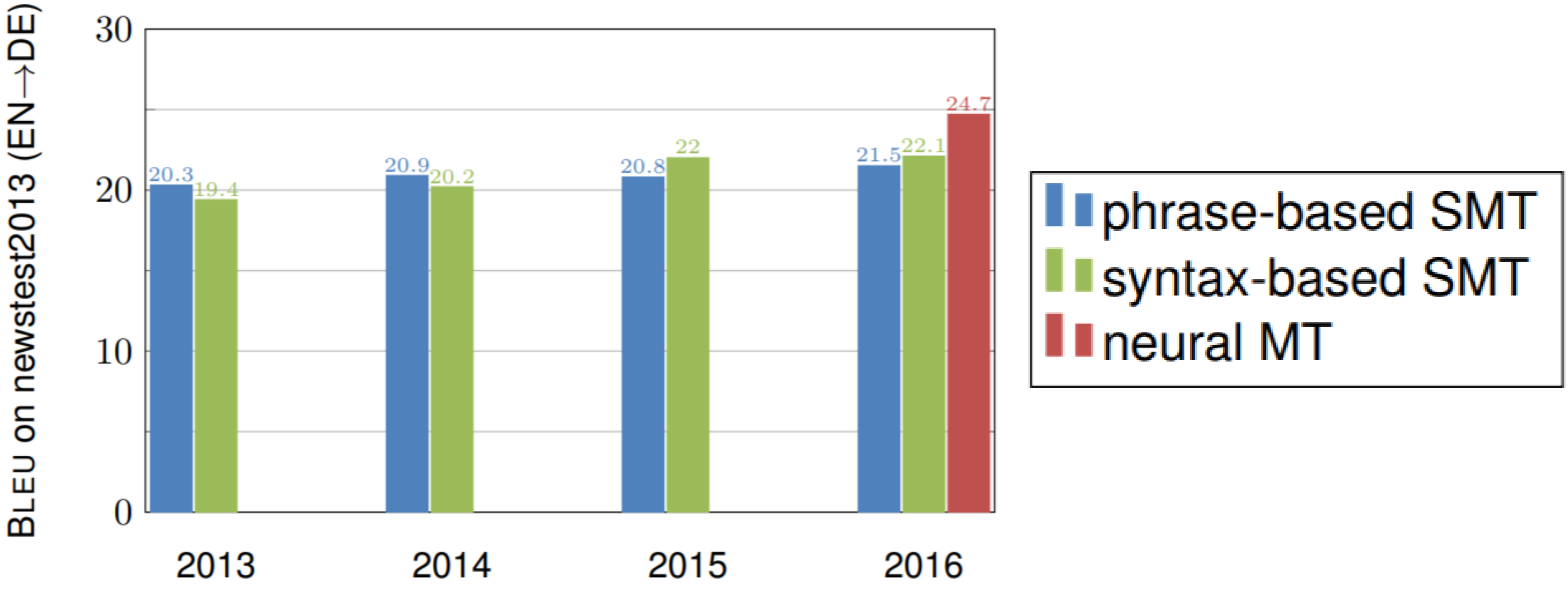


Source: clarifai

Text

- Language translation by a sequence-to-sequence learning network
  - RNN with gating units + attention

Edinburgh's WMT Results Over the Years



Source: http://www.meta-net.eu/events/meta-forum2016/slides/09_sennrich.pdf

Deep Reinforcement Learning

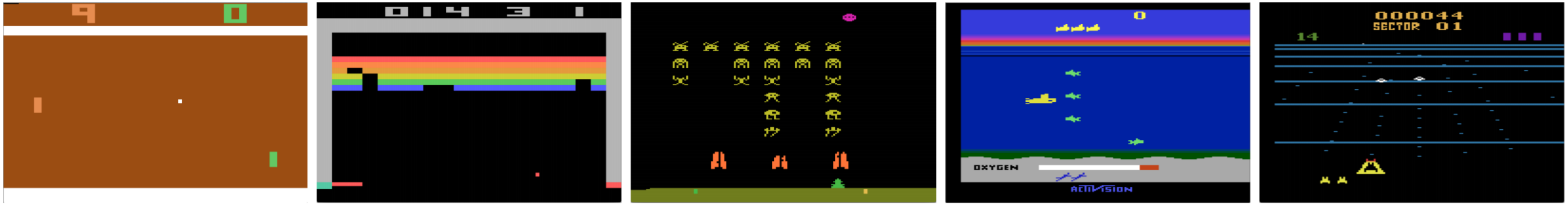
- Reinforcement learning: an autonomous agent must learn to perform a task by trial and error

- DeepMind showed that Deep RL agent is capable of learning to play Atari video games reaching human-level performance on many tasks

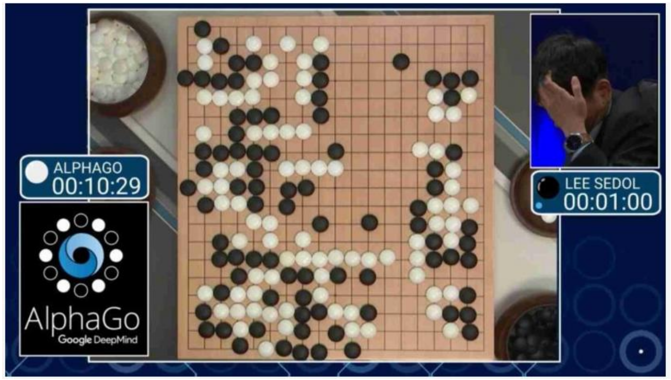
- Deep learning has also significantly improved the performance of reinforcement learning for robotics

Deep Reinforcement Learning

- DQN (2013): Atari 2600 games
  – neural network agent that is able to successfully learn to play as many of the games as possible without any hand-designed feature.



- Deep Mind's alphaGo defeats former world champion in 2016.



Source: https://gogameguru.com/alphago-shows-true-strength-3rd-victory-lee-sedol/

Generative Adversarial Networks

GANs to synthesize a diversity of images, sounds and text imitating unlabeled images, sounds or text

[Goodfellow, NIPS 2016 Tutorial, https://arxiv.org/pdf/1701.00160.pdf]

Memory Networks & Neural Turing Machines

- Memory-augmented networks gave rise to systems which intend to reason and answer questions

- Neural Turing Machine can learn simple programs from examples of desired behavior
  - They can learn to sort lists of numbers given examples of scrambled and sorted sequences.
  - This self-programming technology is in its infancy.

Questions

- Why deep learning approach?
  - Which makes it such popular (in comparison with traditional artificial neural networks)

- Future development
  - The road to general-purpose AI?

Still Far from Human-Level AI

- Industrial successes mostly based on supervised learning
  - Unsupervised and reinforcement learning are more important in human intelligence
    - Human outperforms machines at unsupervised learning
    - Discovering the underlying causal factors is much helpful
    - Human interact with the world not just observe it
- Learning superficial clues, not generalizing well outside of training contexts, easy to fool trained networks
- Still unable to discover higher-level abstractions at multiple time scales, very long-term dependencies
- Still relying heavily on smooth differentiable predictors (using backprop, the workhorse of deep learning)

This slide has been adapted from: http://www.ds3-datascience-polytechnique.fr/wp-content/uploads/2017/08/2017_08_28_1000-1100_Yoshua_Bengio_DeepLearning_1.pdf

Still Far from Human-Level AI

- We need sufficient computational power for models large enough to capture human-level knowledge

- Actually understanding language (also solves generating), requiring enough world knowledge / commonsense
  - Neural nets which really understand the notions of object, agent, action, etc.

- Large-scale knowledge representation allowing one-shot learning as well as discovering new abstractions and explanations by 'compiling' previous observations

- Many fundamental research questions are in front of us

This slide has been adapted from: http://www.ds3-datascience-polytechnique.fr/wp-content/uploads/2017/08/2017_08_28_1000-1100_Yoshua_Bengio_DeepLearning_1.pdf

Course Outline

- Introduction
- Machine Learning review and history of deep learning
- Multi-layer perceptrons and Backpropagation
- Convolutional neural networks (CNN)
- Recurrent neural networks (RNN)
- Deep reinforcement learning (Deep RL)
- Unsupervised deep methods
- Generative Adversarial networks (GAN)
- Variational Autoencoders (VAE)
- Advanced topics
- Applications

Applications We Enter

- Computer vision
- Text and NLP
- Control (Atari games)

Resource

- Deep learning book, Chapter 1.