Deep Learning

Introduction

Applications

MIT Technology Review

10 BREAKTHROUGH TECHNOLOGIES 2013

Introduction | The 10 Technologies | Past Years

Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.

http://www.technologyreview.com/featuredstory/513696/deep-learning/

AI Tasks

- High dimensional input space

- Speech - Music
- Images – Movies - Vision
- Natural Languges – Text
- ...

Speech Recognition

- Speech-to-text
- Error rate on the Switchboard benchmark droped half due to using deep learning

  see review article by Hinton et. al. 2012
- Current commerical systems are based on Deep Leaning
  - Apple Siri, Google's voice search, Bing, wit.ai etc.
- Use of Deep Learning in combination with phonems, HMMs, etc.

Deep Speech

- new Publication of Dec.2014 from Baidu (also see Forbes):

Awni Hannun et. al. Deep Speech: Scaling up end-to-end speech recognition,

  - based on recurrent neural networks
  - better performance and more robust to noise
  - No phonems, no Hidden Markov Models, no handcrafed features etc.
  - Directly works on spectrogram

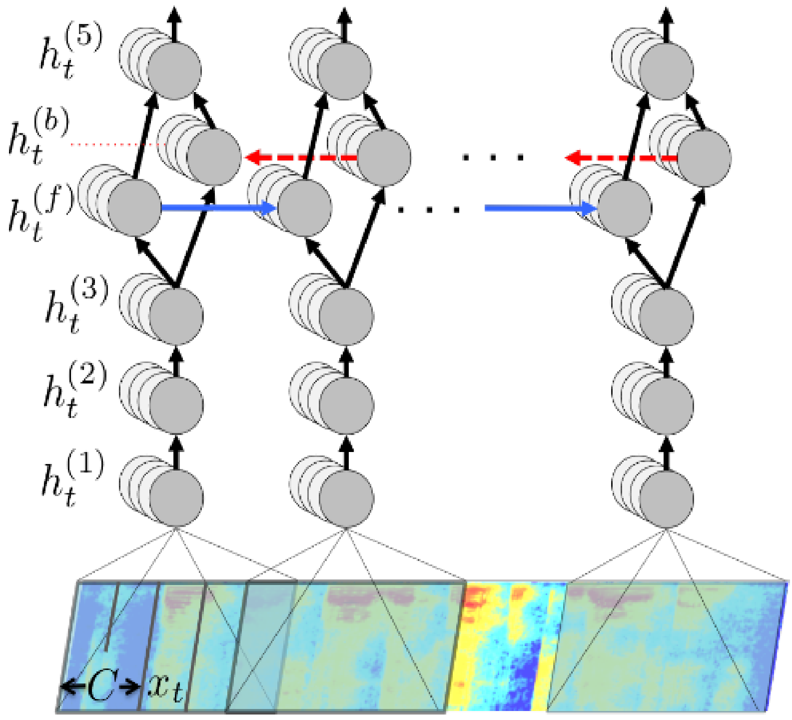


Figure 1: Structure of our RNN model and notation.

figure from Awni Hannun et. al. Deep Speech: Scaling up end-to-end speech recognition, http://arxiv.org/abs/1412.5567

Object Recognition

- ImageNet object LSVRC 2010 recognition benchmark
  - 1000 different classes
  - Competition
    - error rate with convolutional neural networks 15.3% Krizhevsky et.al. 2012
    - second best solution (no deep learning) error rate: 26.1%

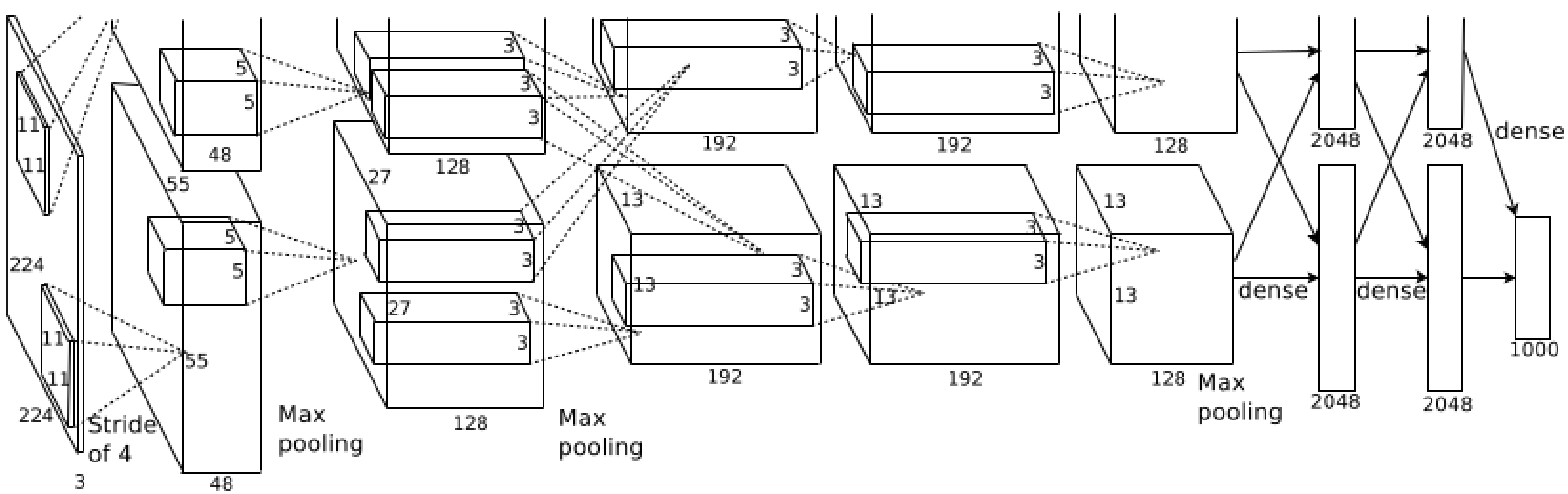
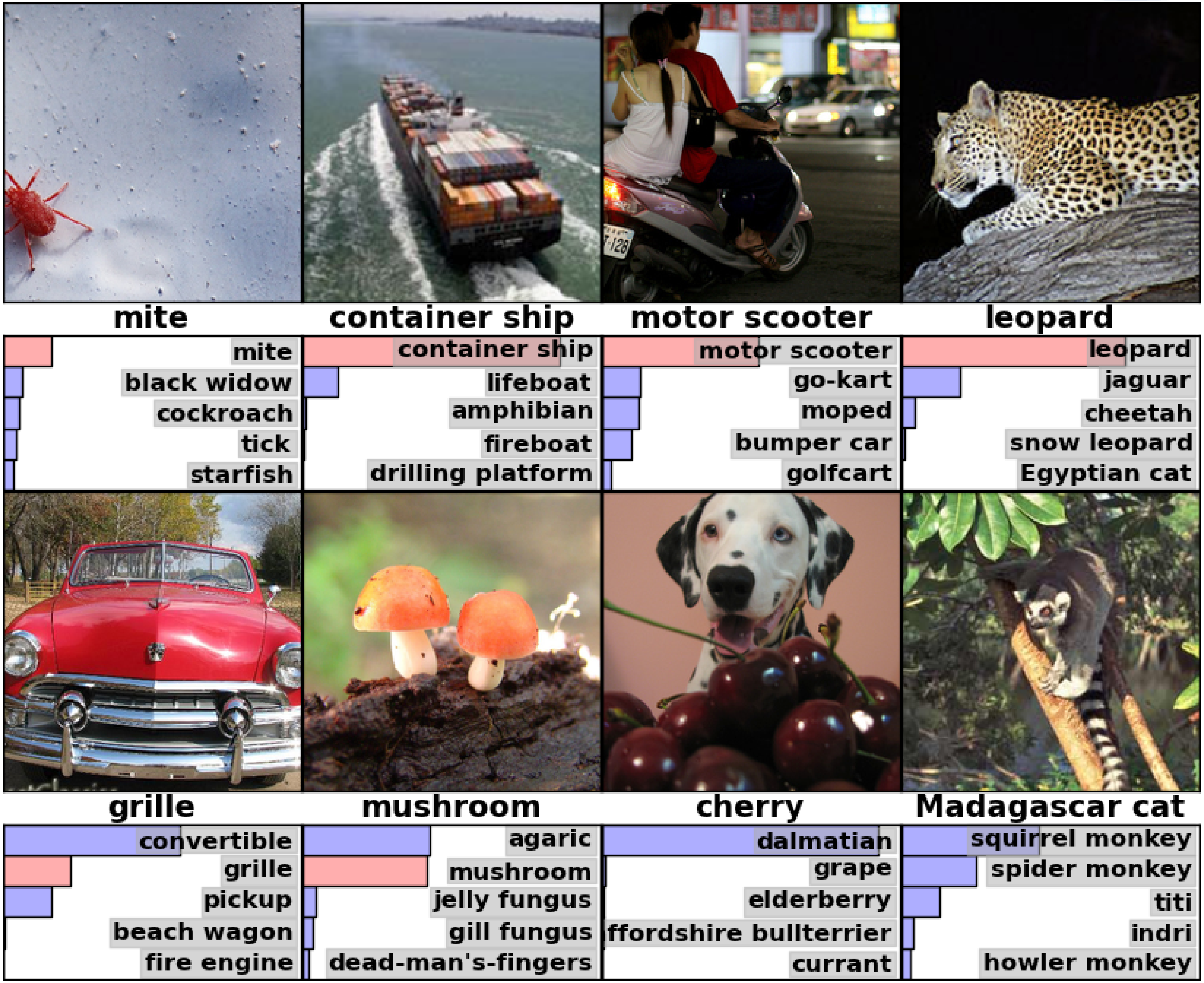


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities



from Krizhevsky et.al. 2012

- Convolutional Neural Networks are now the state-of-the-art in image classification
- Szegedy et.al. 2014 GoogLeNet error rate: for ILSVRC 2014: 6.5%
  - deep network with 22 layers

(a) Siberian husky (b) Eskimo dog

Figure 1: Two distinct classes from the 1000 classes of the ILSVRC 2014 classification challenge.

- now 4.94%: K. He et.al.: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification

Natural Language Processing

- Neural Probabilistic Language Models
  - Bengio et. al. 2003
  - on char level - rnn: Sutskever et. al. 2012
  - Mikolov et. al. 2012 Recurrent Neural Networks
- Word Embeddings
  - Mikolov et. al.
  - Pennington et. al.

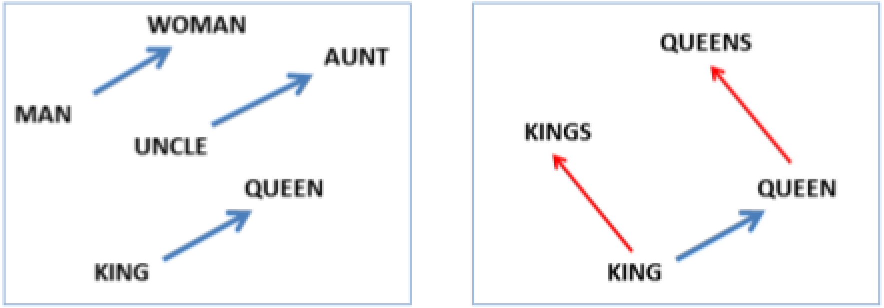


Figure 2: Left panel shows vector offsets for three word

- Recursive neural networks, e.g. for parsing, sentiment analysis (Socher) etc.

(Neural) Machine Translation

- Example: Skype Online Translation
  http://www.skype.com/en/translator-preview/
  http://www.technologyreview.com/news/534101/something-lost-in-skype-translation/
  https://gigaom.com/2014/05/28/skype-will-soon-get-real-time-speech-translation-based-on-deep-learning/
  http://research.microsoft.com/en-us/news/features/translator-052714.aspx
- Lisa Lab (Bengio): Bahdanau et. al., K. Cho et. al.
- Sutskever I. et. al.2014 (google)
  - 5 stacked LSTM hidden layers

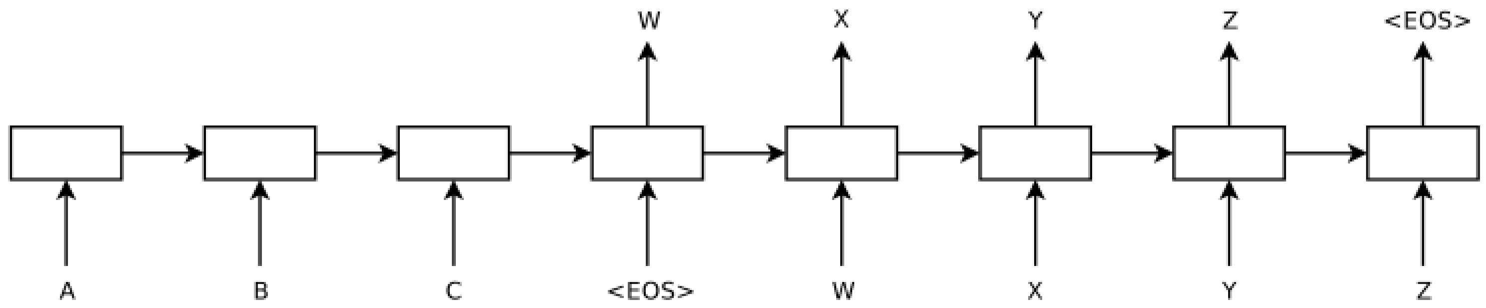


Figure 1: Our model reads an input sentence "ABC" and produces "WXYZ" as the output sentence. The

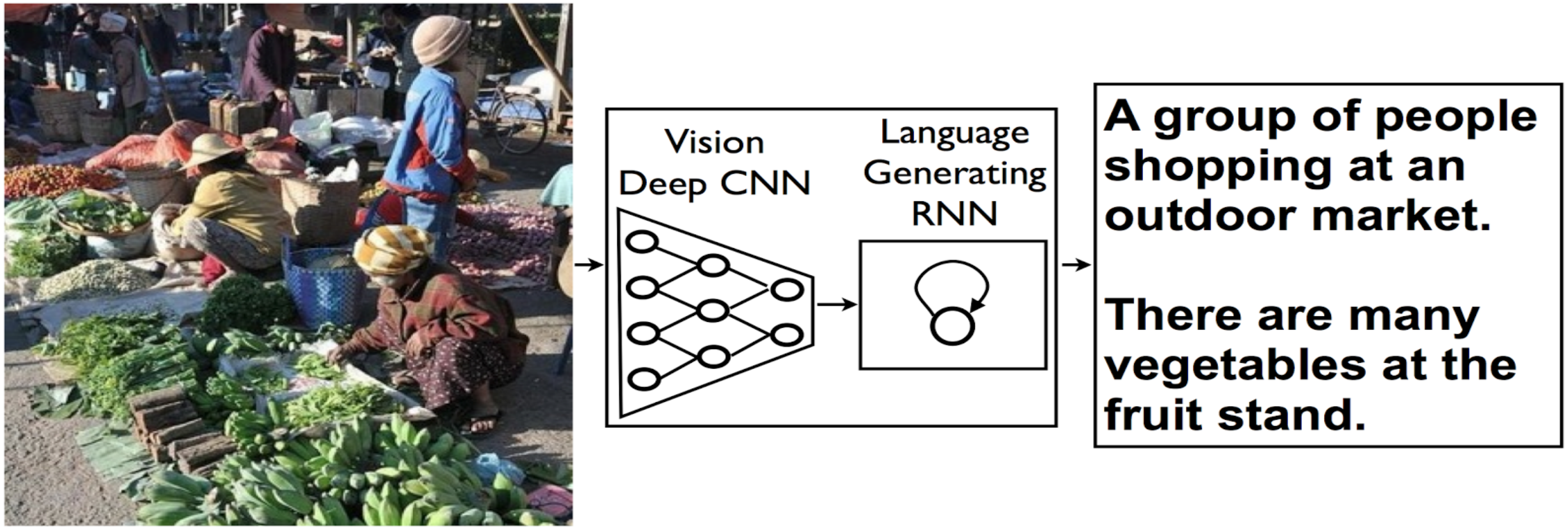
Automatic image captions

- Karpathy et. al.
- Fang et. al

  also see (with a list of more papers in this field:
  http://blogs.technet.com/b/machinelearning/archive/2014/11/18/rapid-progress-in-automatic-image-captionin

- Kiros et.al.
- Vinyals et. al.

  Output of a high level representation of a Convolutional Neural Network as input to an LSTM recurrent neural Network



Further applications

- Self Driving Cars
  - traffic sign detection – Ciresan et. al. 2012
  - pedestrian detection - Sermanet et. al. 2013
- Image segmentation
  - Couprie et. al., Farabet et.al.
- Search / Information Retrieval
  - Image Retrival
  - Music retrieval
- Robotics Yang et. al.
- Face Detection: Y. Taigman, DeepFace

Further applications cont.

- Deep reinforcement Learning: Example Playing Video Games
  - Playing Atari with Deep Reinforcement Learning
  - Human-level control through deep reinforcement learning

- Drug Discovery: https://gigaom.com/2015/03/02/google-stanford

- Molecular Properties: http://quantum-machine.org/documents/qm-icm

Further applications cont.

- Fraud detection (by PayPal)
- Question Answering
  - –

General Articles about Deep Learning in the press

- Technology review
  - http://www.technologyreview.com/featuredstory/51
- BBC Report
  - http://www.bbc.co.uk/programmes/p02kmqt1

History of NN and Deep Learning

- 1943, McCulloch Neuron
- Late 1950: Perceptron learning rule

  only linear separable problems solvable => rule base expert systems
- First revival of NN mid 1980: Backpropagation Algorithm: works well for one *hidden layer*
- 1998 Convolutional Neural Networks of Yann LeCun : "First" Deep Architecture, with efficient learning
- 2006 Stacking of Restricted Boltzmann Machines to learn feature hierachies by Hinton
- Since 2007 second revival of Neural Networks: Deep Learning – with many applications

Success in applications?
Why now?

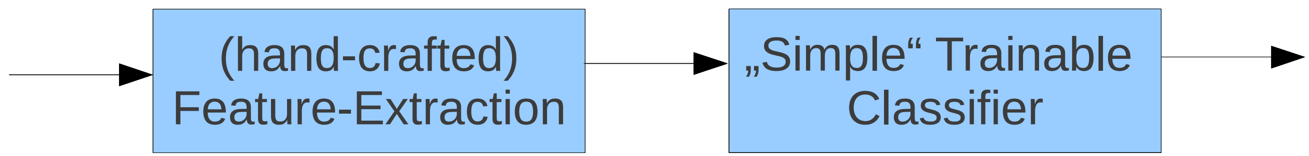
- Computational Power: Training on GPUs

  (graphic cards)
- Big Data
  - Labeled and unlabeled (semi-supervised)
- Better Algorithms
  - Learning of neural networks with many hidden layers

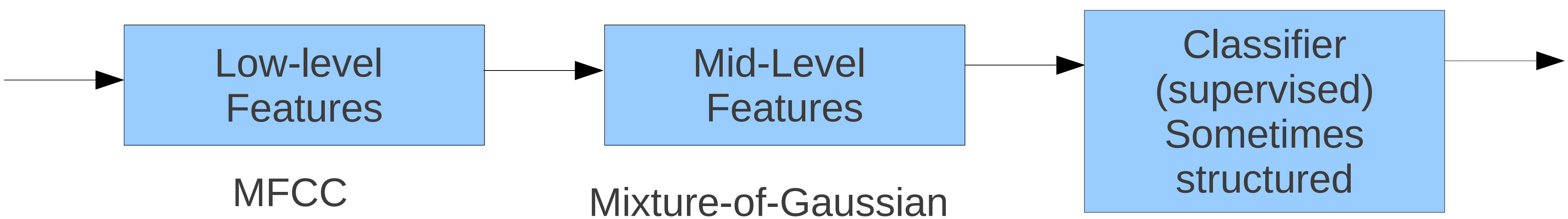
Mainstream Recognition Systems

nach [LeCun]

- traditional machine learning



- traditional approach in image processing und speech recognition
  - *mid-level features, often trained unsupervised*

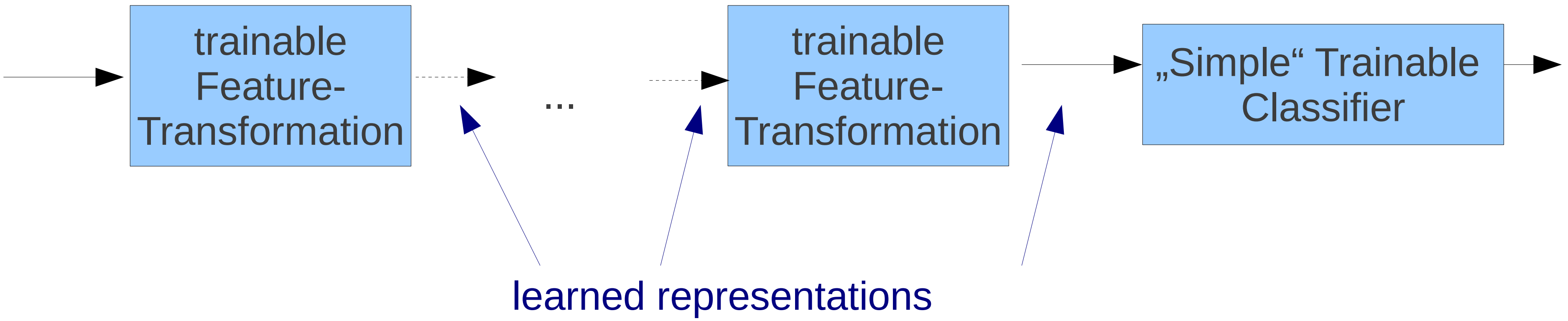


MFCC
SIFT
HoG
LBP

Mixture-of-Gaussian
K-means
Sparse Coding

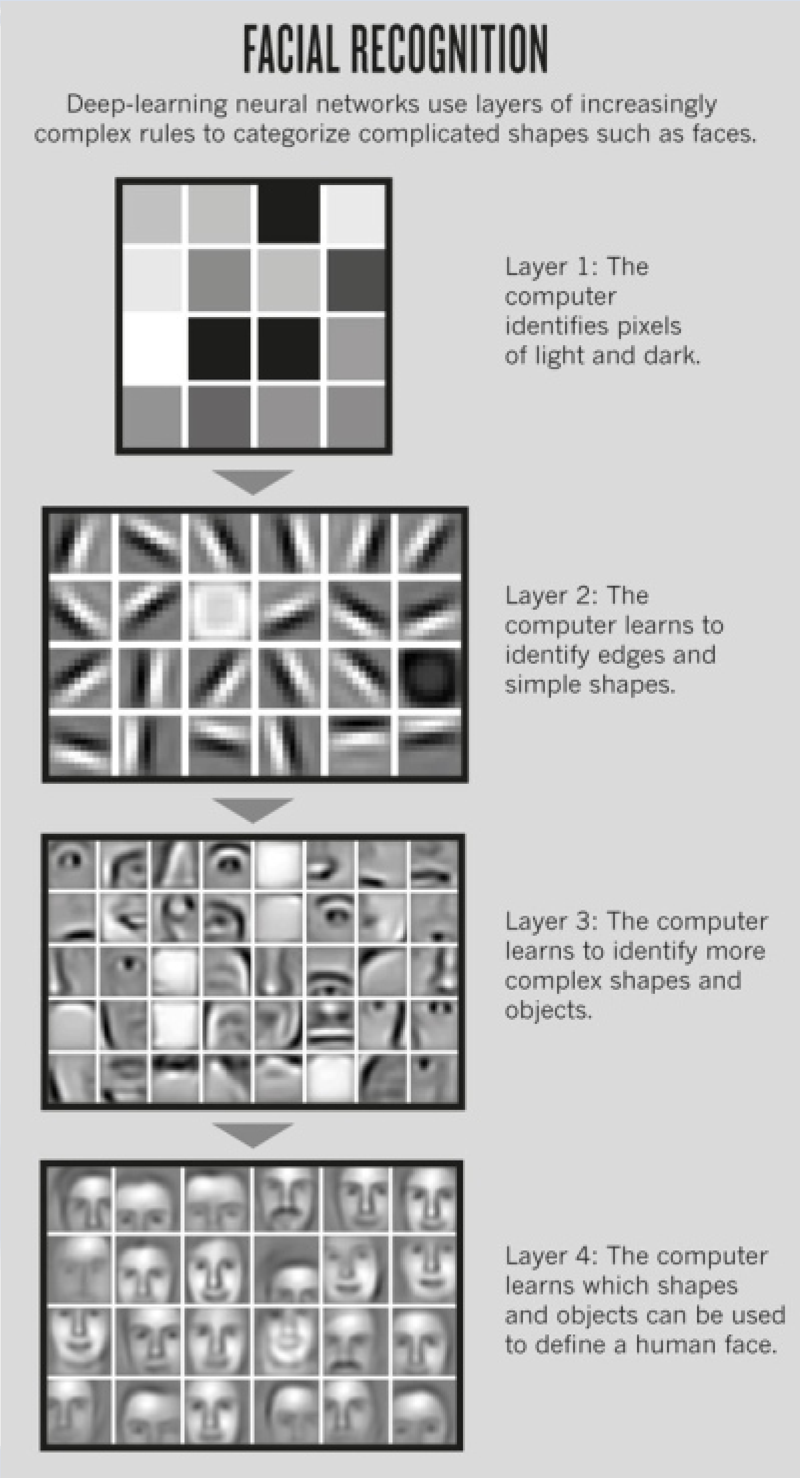
Trainable Feature Hierachies

- higher representations
  - independent „explaining factors“ are separeted



What is deep learning (DL)?

- Learning of multiple levels of hierarchical representations
  - DL is a subfield of representation learning
  - Feature Learning
- Higher layer have increasing complexity/abstraction
  - Deep architecture: multiple levels of feature learning
- If there are "good" representations using a supervised predictor on top on them is easy.

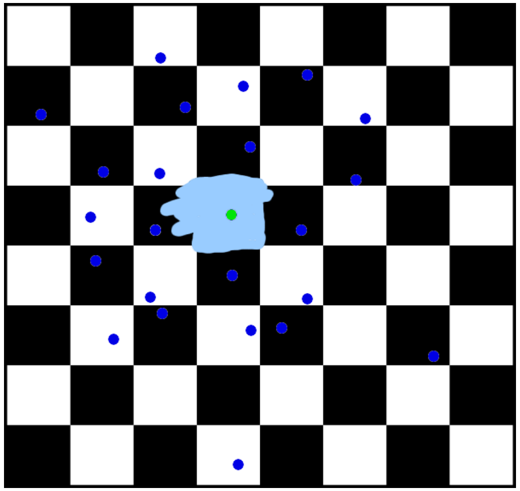


Example of a feature hierachy

figure from A. Ng.

- AI tasks
- high dimensional input space
  - if the target function varies a lot (many ups and downs) smoothness prior is not enough.

    A classifier with **non-distributed representation**, e.g. svn with gausian kernel, needs exponentially (with num dims) many examples because of the curse of dimensionality



further assumptions are necessary

Generalizing locally

- a state for each distinguishable region
- Number of distinguishable regions linear to the number of states
- exponentially ineffient

(with number of dimension of the input space)

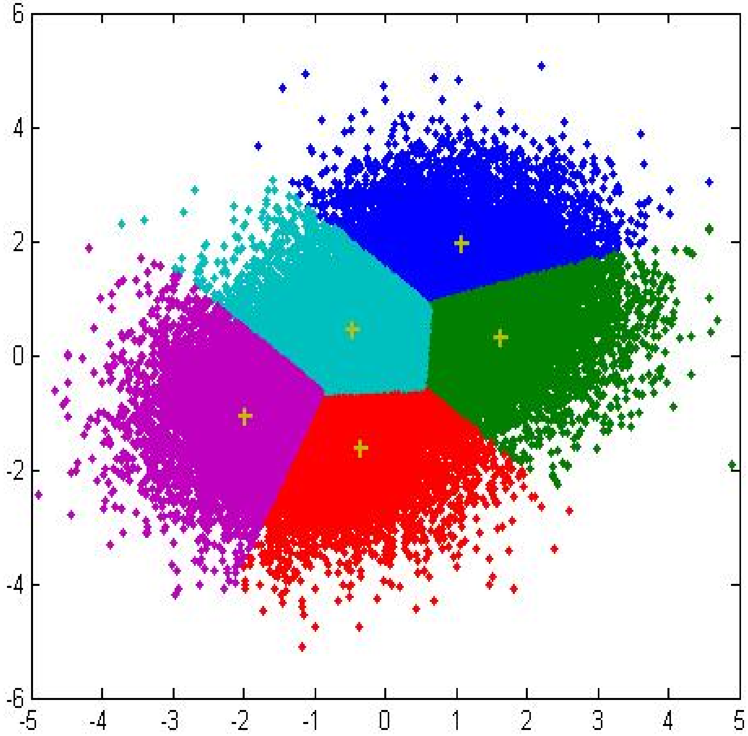
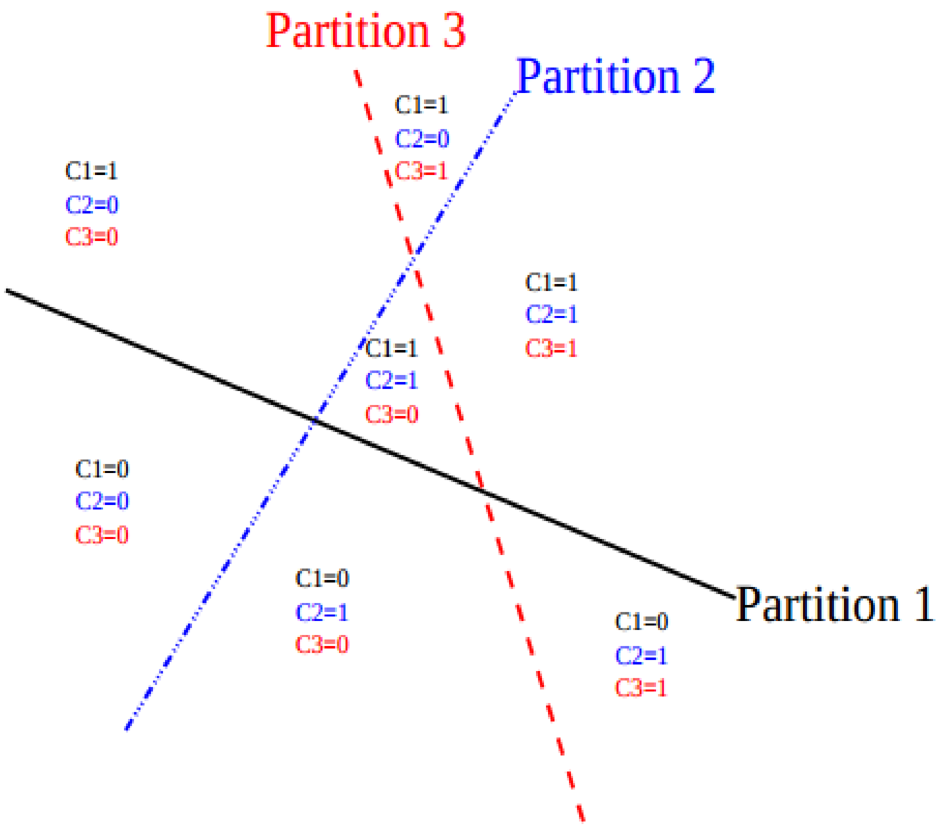


figure from [Be09]

Distributed Representation



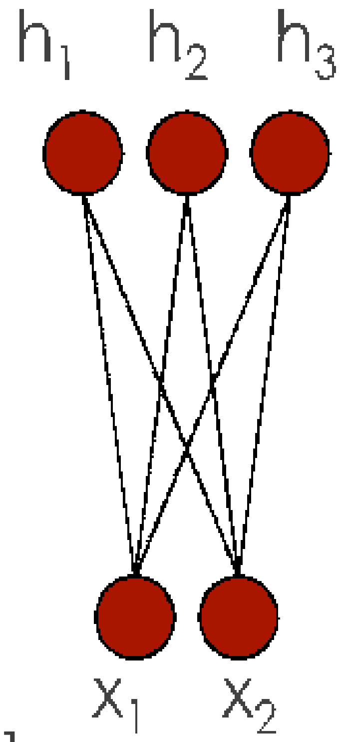
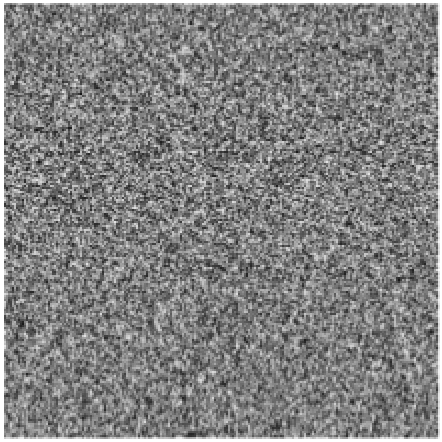


figure from [Be09]

allows non local generalization

Manifold Hypothesis

- Data space extrem high dimensional
- Natural data "lives" in low-dimensional (non-linear) manifolds
- Because variables in natural data are mutually dependent
- Example:
  - photos/pictures vs. random pixels



from Hyvärinen et. al. Natural Image Statistics, Springer Verlag 2009

Manifold

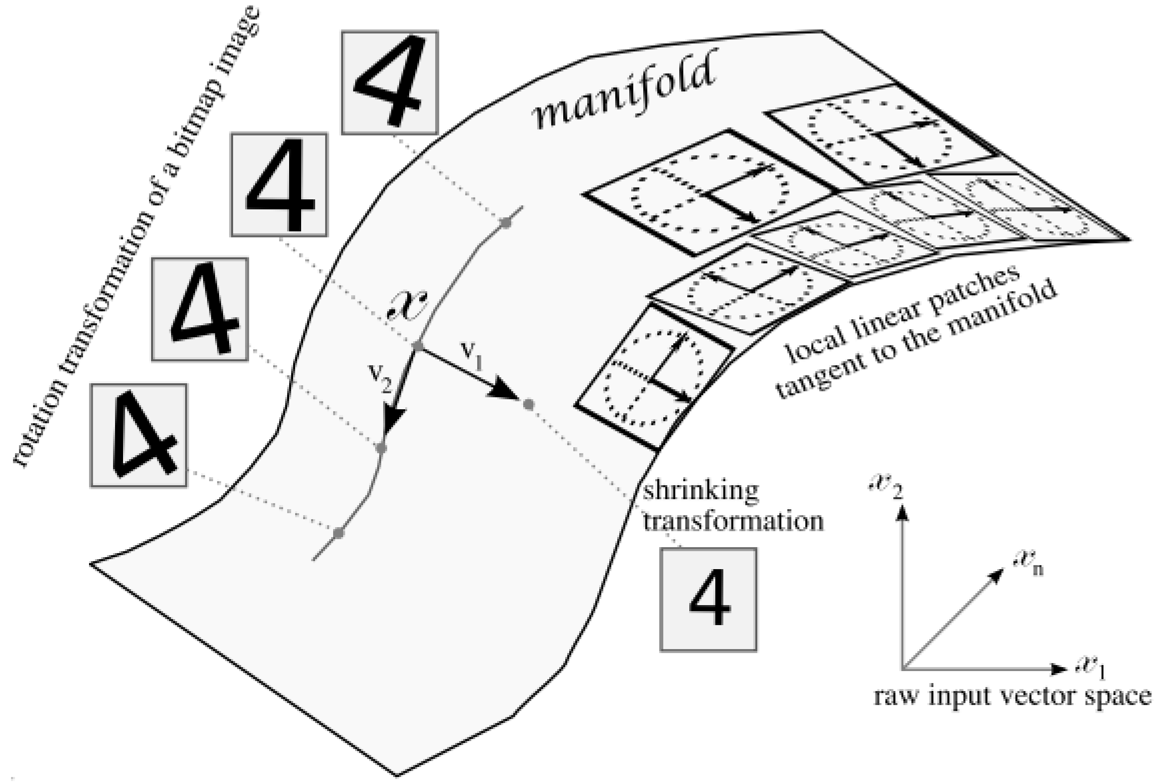
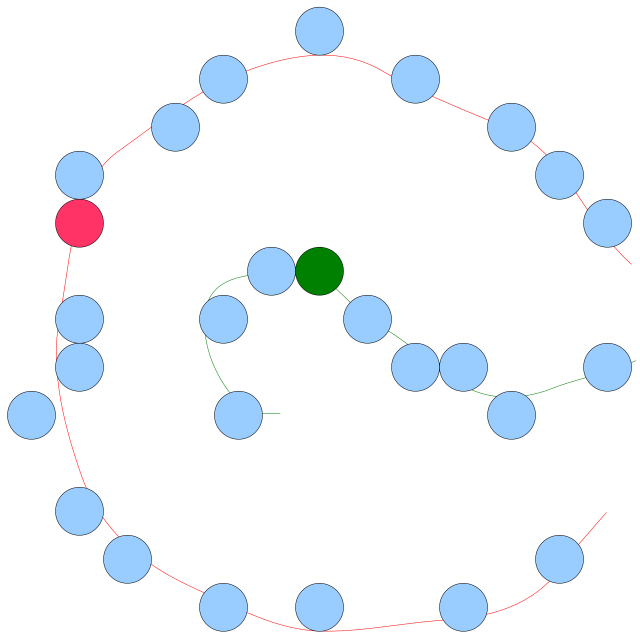


figure from [Be09]

Semi Supervised Learning

- Assumption: Knowing p(x) helps for p(y|x)
- Clustering Hypothesis (different classes are separated by low probability density)



"transfer learning"

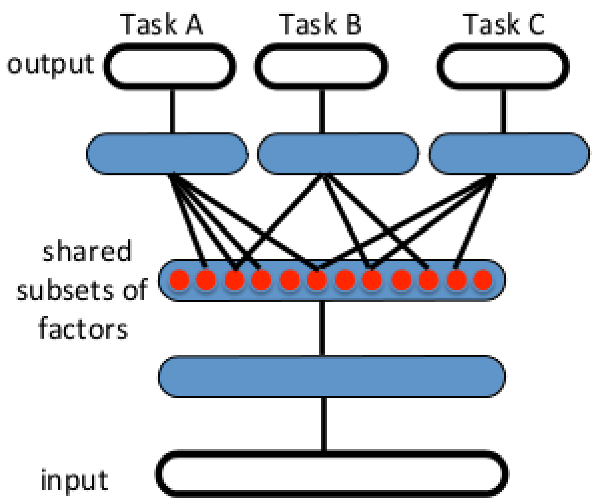


Fig. 1. Illustration of representation-learning discovering explanatory factors (middle hidden layer, in red), some explaining the input (semi-supervised setting), and some explaining target for each task. Because these subsets overlap, sharing of statistical strength helps generalization.

from Representation Learning:
A Review and New Perspectives
Yoshua Bengio, Aaron Courville, Pascal Vincent,
http://arxiv.org/abs/1206.5538

Depth helps

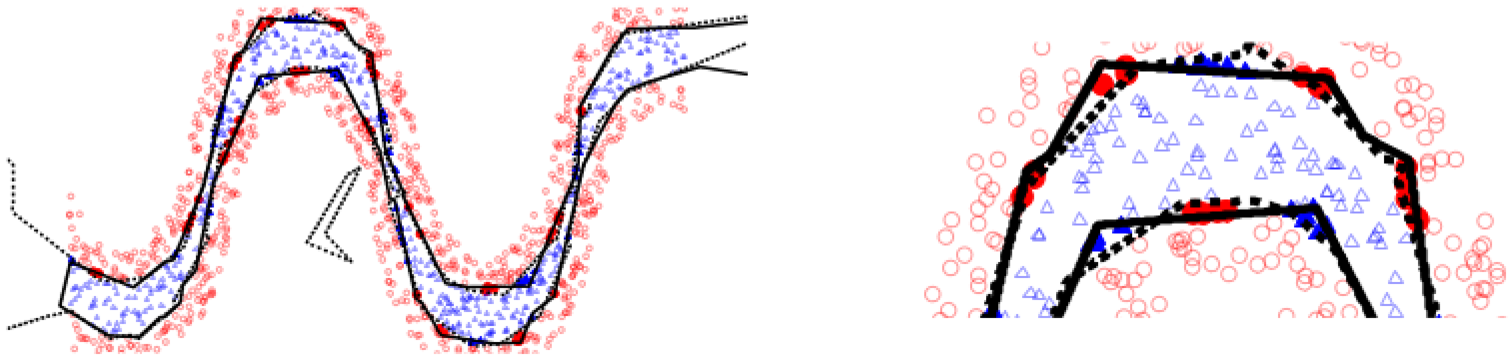


Figure 1: Binary classification using a shallow model with 20 hidden units (solid line) and a deep model with two layers of 10 units each (dashed line). The right panel shows a close-up of the left panel. Filled markers indicate errors made by the shallow model.

from
On the number of linear regions of deep neural networks.
G.Montufar, R. Pascanu, K. Cho, and Y. Bengio. NIPS 2014
http://arxiv.org/abs/1402.1869

Literature

- [Be09] Yoshua Bengio, Learning Deep Architectures for AI, Foundations and Trends in Machine Learning, 2(1), pp.1-127, 2009.
- [Be12] Yoshua Bengio, Aaron Courville, Pascal Vincent Representation Learning: A Review and New Perspectives, Arxiv, 2012
- Yoshua Bengio, Ian Goodfellow, Aaron Courville, Deep Learning, MIT Press, In preparation

Links to Reading Lists etc.

- http://deeplearning.net/reading-list/
- http://ufldl.stanford.edu/wiki/index.php/UFLDL_Recommended_R
- http://web.eecs.umich.edu/~honglak/teaching/eecs598/schedule.h