Learning Image Features

Yann LeCun
The Courant Institute of Mathematical Sciences
And Center for Neural Science
New York University
How do we learn vision and perception?
- From the image of an airplane, how do we extract a representation that is invariant to pose, illumination, background, clutter, object instance....
- How can a human (or a machine) learn those representations by just looking at the world?

How can we learn visual categories from just a few examples?
- I don't need to see many airplanes before I can recognize every airplane (even really weird ones)
Vision occupies a big chunk of our brains

1/3 of the macaque brain

[from Van Essen]
Vision is very fast and the visual cortex is hierarchical

The ventral (recognition) pathway in the visual cortex
The Primate's Visual System is Deep (LGN->V1->V2->V4->IT)

- The recognition of everyday objects is a very fast process.
  - The recognition of common objects is essentially “feed forward.”
  - But not all of vision is feed forward.

- Much of the visual system (all of it?) is the result of learning
  - How much prior structure is there?

- If the visual system is deep (around 10 layers) and learned

- What is the learning algorithm of the visual cortex?
  - What learning algorithm can train neural nets as “deep” as the visual system (10 layers?).
  - Unsupervised vs Supervised learning
  - What is the loss function?
  - What is the organizing principle?
  - Broader question (Hinton): what is the learning algorithm of the neo-cortex?
Can we devise learning algorithms to train a “deep” artificial visual system, and other artificial perception systems.

How can we learn the structure of the world?
- How can we build/learn internal representations of the world that allow us to discover its hidden structure?
- How can we learn internal representations that capture the relevant information and eliminates irrelevant variabilities?

How can a human or a machine learn internal representations by just looking at the world?

Can we find learning methods that solve really complex problems end-to-end, such as vision, natural language, speech....?
The raw input is pre-processed through a hand-crafted feature extractor.

- The features are not learned.

- The trainable classifier is often generic (task independent), and “simple” (linear classifier, kernel machine, nearest neighbor,.....)

- The most common Machine Learning architecture: the Kernel Machine.
“Modern” Object Recognition Architecture in Computer Vision

Example:

- Edges + Rectification + Histograms + SVM [Dalal & Triggs 2005]
- SIFT + classification

Fixed Features + “shallow” classifier
Biologically-inspired models of low-level feature extraction
- Inspired by [Hubel and Wiesel 1962]
- Many feature extraction methods are based on this
- SIFT, GIST, HoG, Convolutional networks.....
“State of the Art” architecture for object recognition

Example:
- SIFT features with Spatial Pyramid Match Kernel SVM [Lazebnik et al. 2006]
- Fixed Features + unsupervised features + “shallow” classifier
Good Representations are Hierarchical

In Language: hierarchy in syntax and semantics
- Words -> Parts of Speech -> Sentences -> Text
- Objects, Actions, Attributes... -> Phrases -> Statements -> Stories

In Vision: part-whole hierarchy
- Pixels -> Edges -> Textons -> Parts -> Objects -> Scenes
Deep Learning: learning a hierarchy of internal representations

From low-level features to mid-level invariant representations, to object identities

Representations are increasingly invariant as we go up the layers

using multiple stages gets around the specificity/invariance dilemma
Plan of the Tutorial

Simple methods for supervised learning
- Energy-based learning
- Perceptron, logistic regression, SVM

Deep Supervised Learning
- Backpropagation

Architectures for Image Recognition
- Local feature extractors, SIFT, HoG
- Vector quantization and feature pooling

Trainable Architectures for Image Recognition: Feature Learning
- Supervised Convolutional Networks

Unsupervised Deep Learning, Energy-Based Models
- Predictive Sparse Decomposition

Applications
- Face/pedestrian detection, object recognition, image segmentation, obstacle detection for robots.