ICVSS 2013
Sicily ~ 14-20 July
International Computer Vision Summer School

# Deep Learning and Convolutional Networks in Vision

**Yann LeCun**
**New York University**

## Abstract

### I. Intro / Motivation

*Motivation*

- learning representations and features is the next big challenge in computer vision
- traditional model: features + shallow classifier
- why do representation need to be hierarchical?
- circuit complexity argument: most functions require mutliple steps
- why are kernel methods "shallow"?
- hierarchical representations are required for invariance
- deep architectures enable feature sharing

*What are good representations/features?*

- the manifold model of natural data
- disantangling the explanatory factors of variation
- embedding inputs into high dimensional space non-linearly

### II. deep and supervised models

*II.A module-based deep learning*

- backprop through modules
- complete example (SVHN with Torch)

*II.B backprop in practice*

- initialization
- stochastic gradient descent
- ill-conditioning issues
- vanishing/exploding gradient problem
- regularization
- parallelization

## III. Convolutional Nets: deep supervised learning in the real world

*III.A Motivation, basic architecture*

- convolution module, pooling module, contrast normalization
- complete example (SVHN with Torch)
- hardware implementations

*III.B Applying convnets*

- application: object recognition (Krizhevski, Zeiler...)
- application: robot driving
- application: scene parsing 2D, 3D, video (Farabet, Couprie)
- application: face/person/pedestrian detection (LeCun, Garcia, Sermanet, NEC)
- action recognition in videos (Taylor, Le)
- pose estimation (Taylor)
- application: volumetric image segmentation (Jain)

*III.C Deployed applications of convnets*

- check reading (AT&T), handwriting recognition (Microsoft), gender/age recognition (NEC), image tagging (Google, Baidu)

*III.D Connection with other methods*
- deconvolutional nets
- connection with "mainstream" approaches to object recognition
- scattering networks

## IV. unsupervised learning, Energy-Based Models

*IV.A Intro*

- basic model: layerwise unsupervised training
- common architecture: regularized auto-encoder
- when is unsupervised pre-training useful?

*IV.B Energy-based unsupervised learning*

- Training strategies and criteria for unsupervised training
- Sparse Coding
- sparse auto-encoders, predictive sparse decomposition

- application: pedestrian detection
- learning invariant features: IPSD, DrSAE

**V. Other Topics**

- Deep Learning and Structured Prediction
- open question and future directions

*Keywords: Deep Learning, Convolutional Networks*