Artificial and robotic vision

Spring 2013

Lecture 3: convolutional neural networks and models of vision systems

e-Lab

Eugenio Culurciello

Convolutional Neural Networks

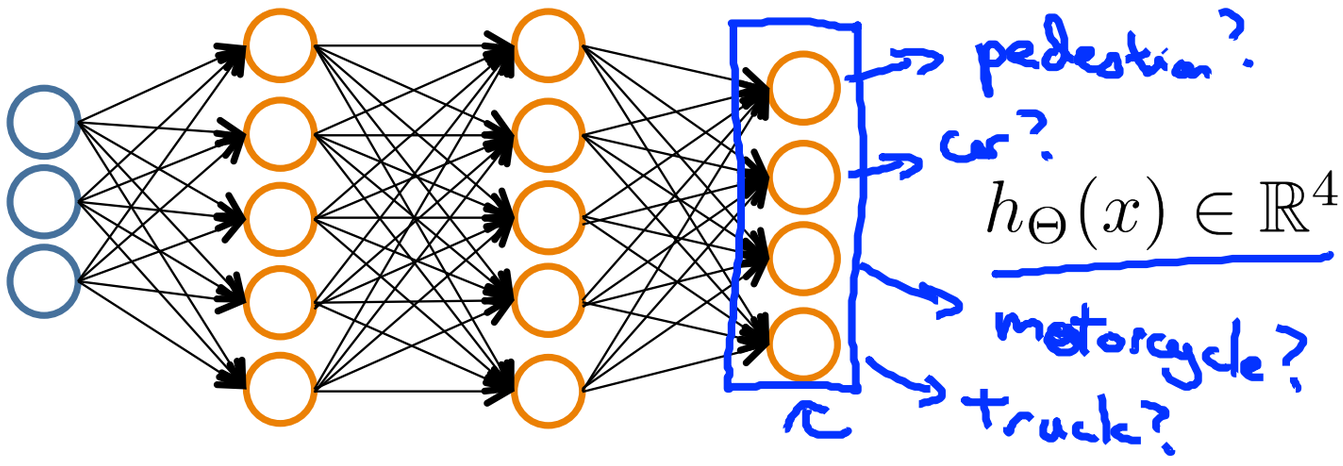
Multiple output units: One-vs-all.

Pedestrian

Car

Motorcycle

Truck



ConvNets are the main workhorse of this class, the main components of Deep Networks and Deep Learning... and of Big Data :-)

Convolutional Neural Networks

Gradient-Based Learning Applied to Document Recognition

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner

Abstract—

Multilayer Neural Networks trained with the backpropagation algorithm constitute the best example of a successful Gradient-Based Learning technique. Given an appropriate network architecture, Gradient-Based Learning algorithms can be used to synthesize a complex decision surface that can classify high-dimensional patterns such as handwritten characters, with minimal preprocessing. This paper reviews various methods applied to handwritten character recognition and compares them on a standard handwritten digit recognition task. Convolutional Neural Networks, that are specifically designed to deal with the variability of 2D shapes, are shown to outperform all other techniques.

Real-life document recognition systems are composed of multiple modules including field extraction, segmentation, recognition, and language modeling. A new learning paradigm, called Graph Transformer Networks (GTN), allows such multi-module systems to be trained globally using Gradient-Based methods so as to minimize an overall performance measure.

Two systems for on-line handwriting recognition are described. Experiments demonstrate the advantage of global training, and the flexibility of Graph Transformer Networks.

A Graph Transformer Network for reading bank check is also described. It uses Convolutional Neural Network character recognizers combined with global training techniques to provides record accuracy on business and personal checks. It is deployed commercially and reads several million checks per day.

Keywords— **Neural Networks, OCR, Document Recognition, Machine Learning, Gradient-Based Learning, Convolutional Neural Networks, Graph Transformer Networks, Finite State Transducers.**

Nomenclature

- GT Graph transformer.
- GTN Graph transformer network.
- HMM Hidden Markov model.
- HOS Heuristic oversegmentation.
- K-NN K-nearest neighbor.
- NN Neural network.
- OCR Optical character recognition.
- PCA Principal component analysis.
- RBF Radial basis function.
- RS-SVM Reduced-set support vector method.
- SDNN Space displacement neural network.
- SVM Support vector method.
- TDNN Time delay neural network.
- V-SVM Virtual support vector method.

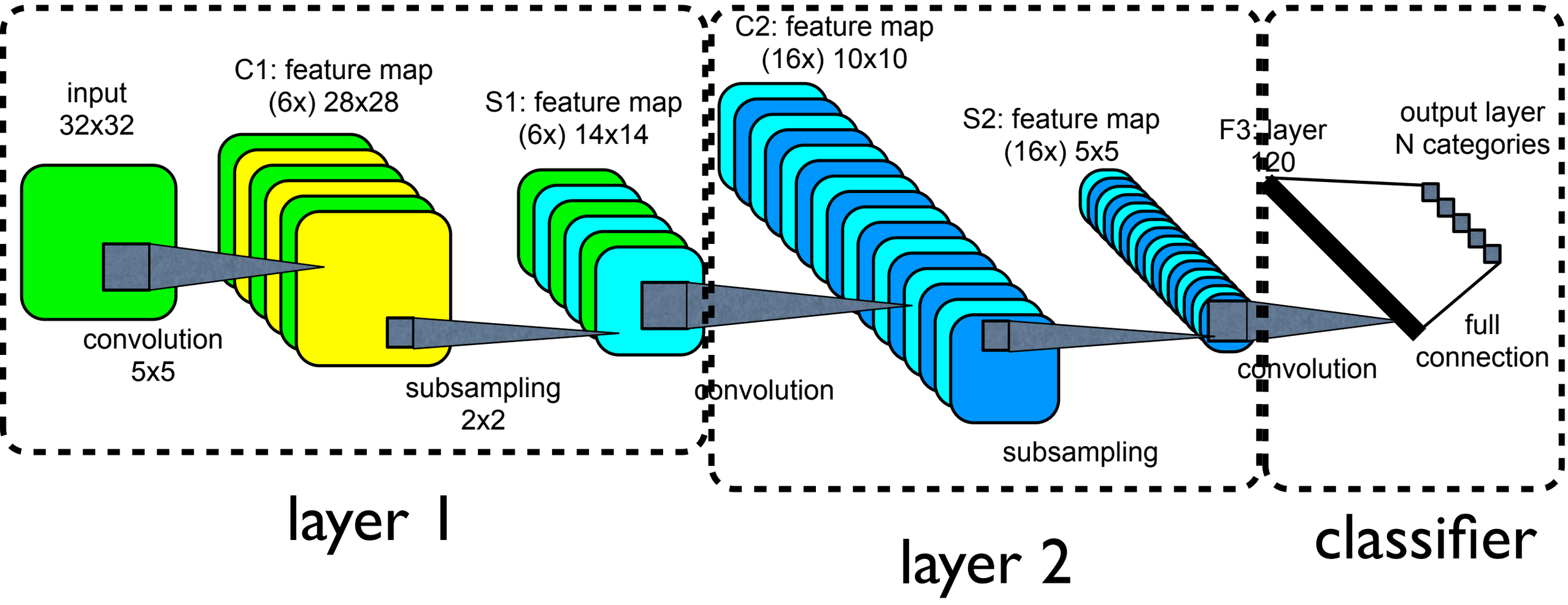
I. Introduction

Over the last several years, machine learning techniques, particularly when applied to neural networks, have played an increasingly important role in the design of pattern recognition systems. In fact, it could be argued that the availability of learning techniques has been a crucial factor in the recent success of pattern recognition applications such as continuous speech recognition and handwriting recognition.

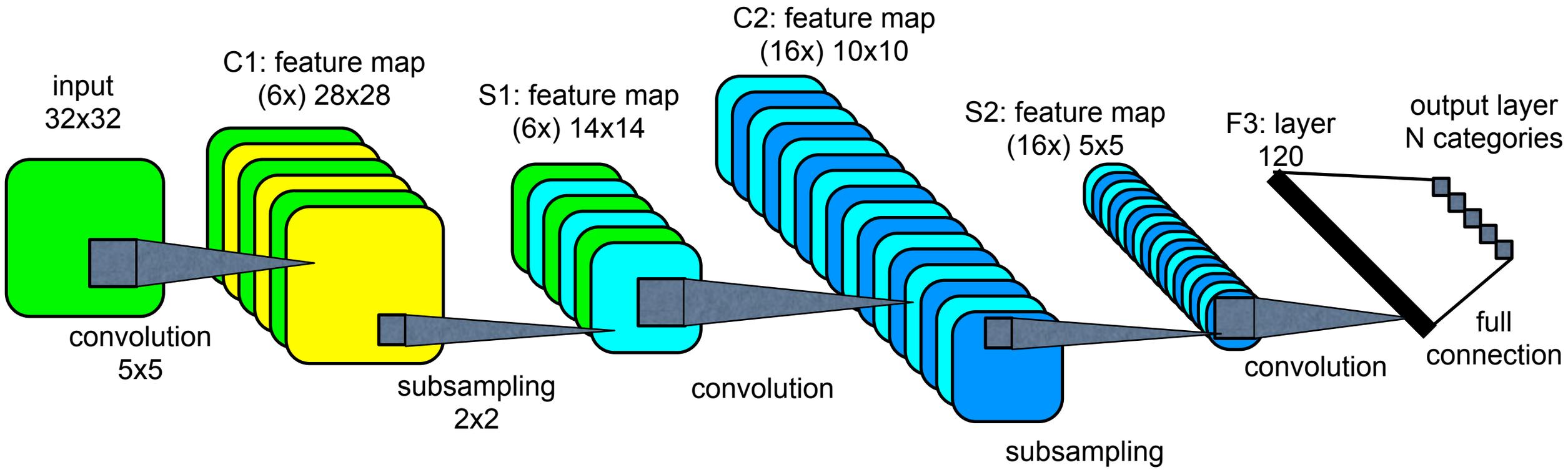
The main message of this paper is that better pattern recognition systems can be built by relying more on automatic learning, and less on hand-designed heuristics. This is made possible by recent progress in machine learning and computer technology. Using character recognition as a case study, we show that hand-crafted feature extraction can be advantageously replaced by carefully designed learning machines that operate directly on pixel images. Using document understanding as a case study, we show that the traditional way of building recognition systems by manually integrating individually designed modules can be replaced by a unified and well-principled design paradigm, called *Graph Transformer Networks*, that allows training all the modules to optimize a global performance criterion.

Since the early days of pattern recognition it has been known that the variability and richness of natural data, be it speech, glyphs, or other types of patterns, make it almost impossible to build an accurate recognition system entirely by hand. Consequently, most pattern recognition systems are built using a combination of automatic learning techniques and hand-crafted algorithms. The usual method of recognizing individual patterns consists in dividing the system into two main modules shown in figure 1. The first module, called the feature extractor, transforms the input patterns so that they can be represented by low-dimensional vectors or short strings of symbols that (a) can be easily matched or compared, and (b) are relatively invariant with respect to transformations and distortions of the input patterns that do not change their nature. The feature extractor contains most of the prior knowledge and is rather specific to the task. It is also the focus of most of the design effort, because it is often entirely hand-crafted. The classifier, on the other hand, is often general-purpose and trainable. One of the main problems with this ap-

Convolutional Neural Networks



Convolutional Neural Networks



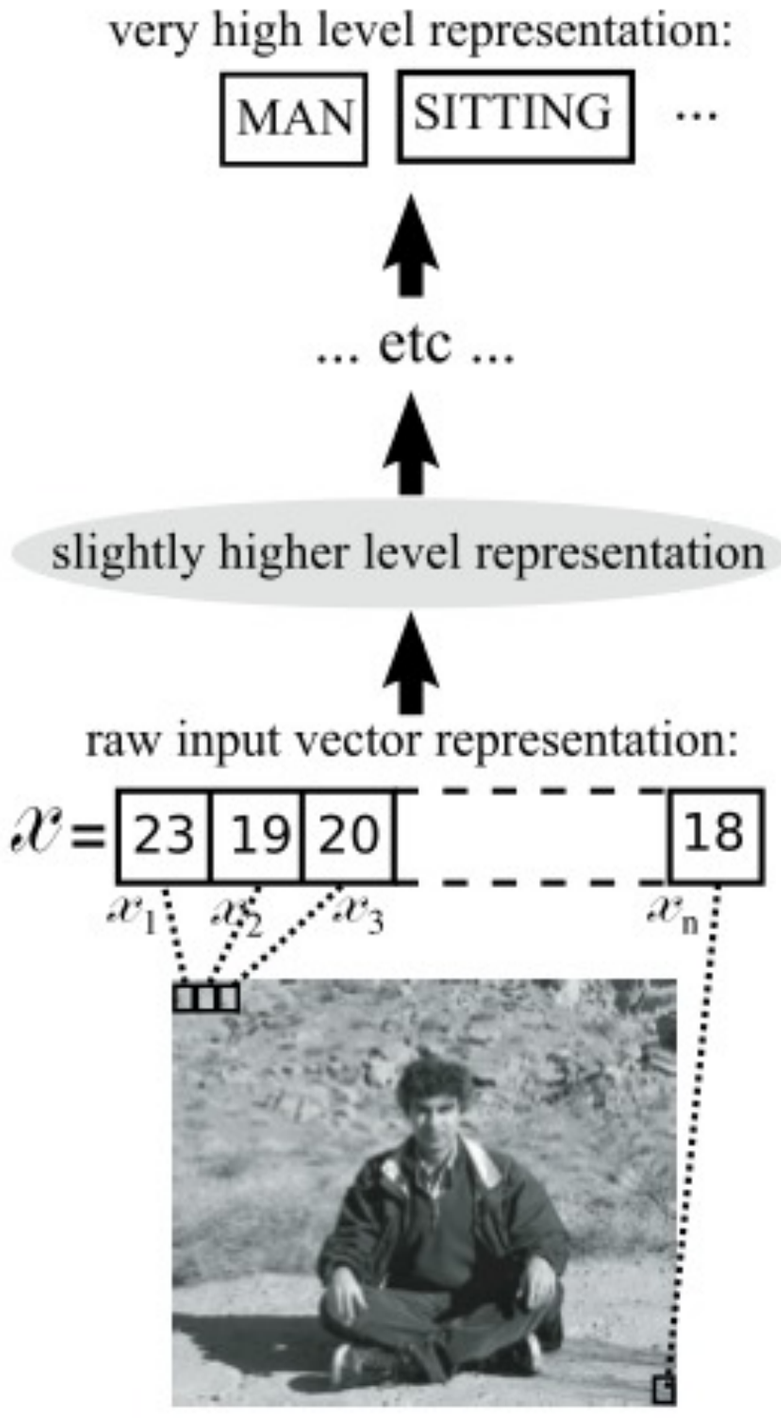
CNN or ConvNets are / have the property of:

- hierarchical
- invariance

e-Lab

why multi layer neural networks?

hierarchical

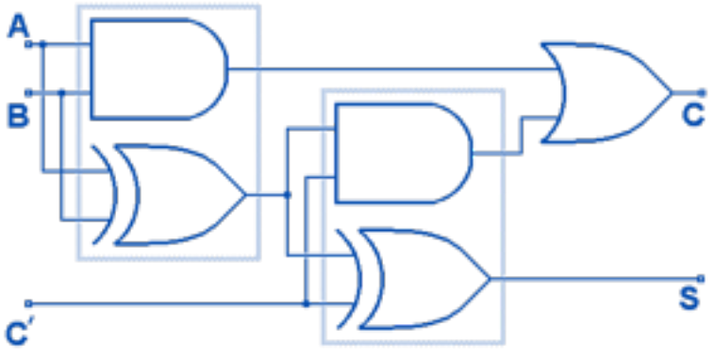


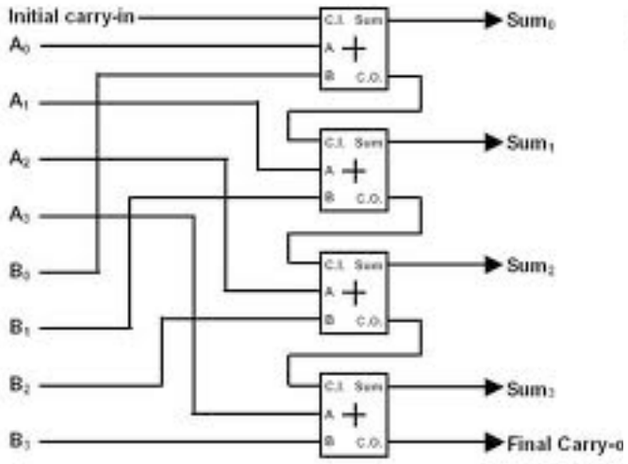
raw input, edges, local shapes, object parts, etc.

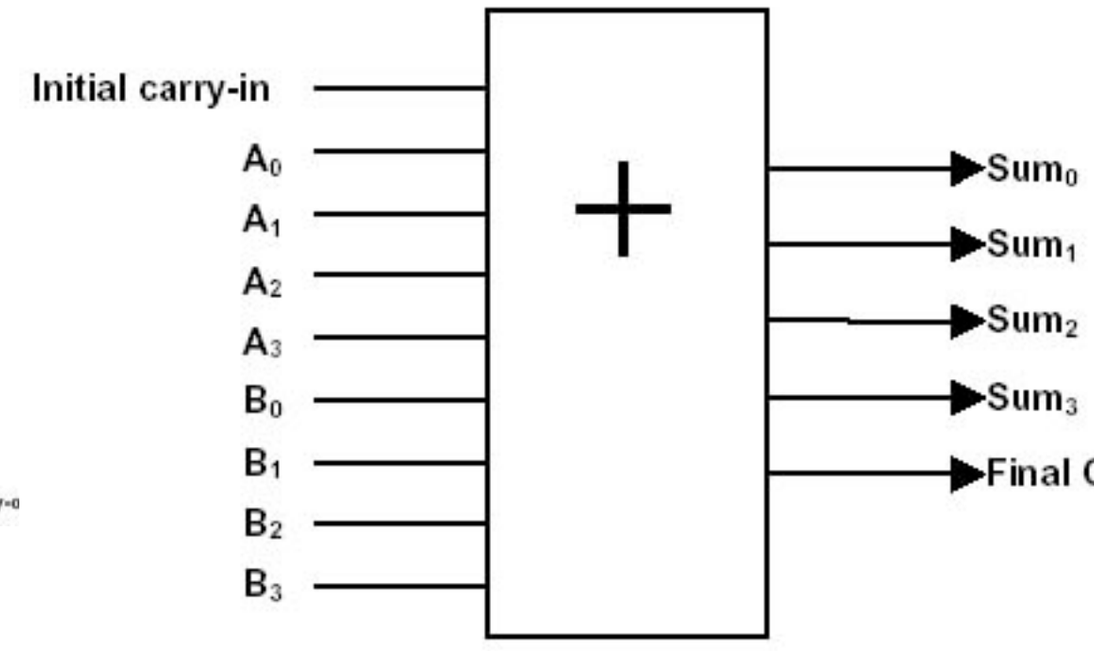
e-Lab

why multi layer neural networks?

hierarchical







functions of functions of functions...

output = f(g(h(...)))

more compact representation

more efficient

invariance and redundancy

- images: pixels in space
- videos: pixels space/time
  - data redundancy

- local information:
  - receptive fields

- shared processing
  - shared weights

e-Lab

invariance

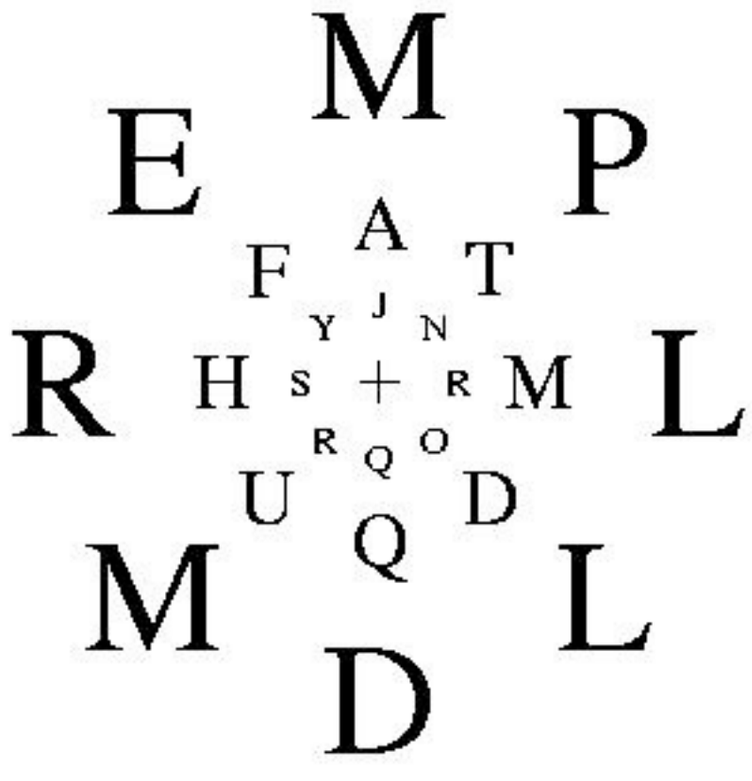
recognition invariant to:

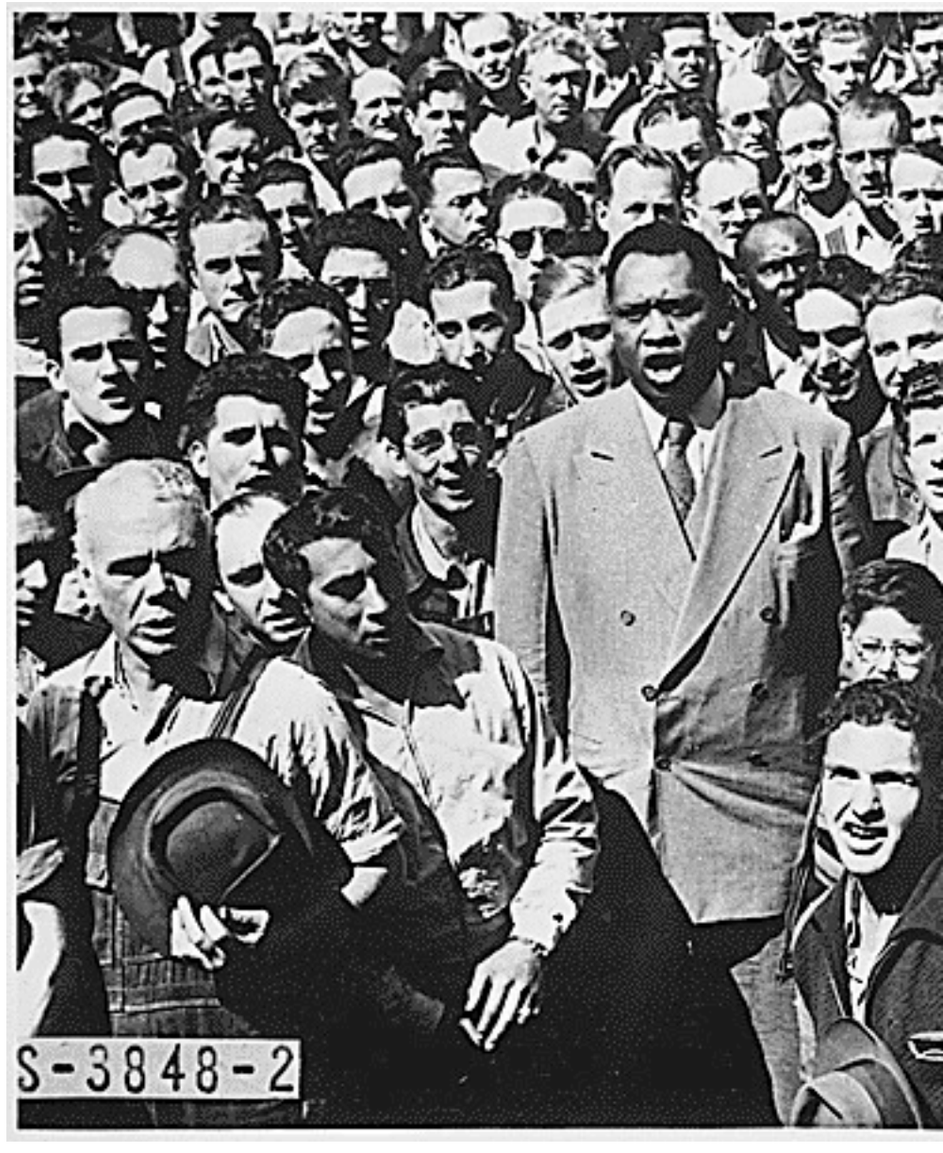
- position

invariance

recognition invariant to:

- size





e-Lab

invariance

recognition invariant to:

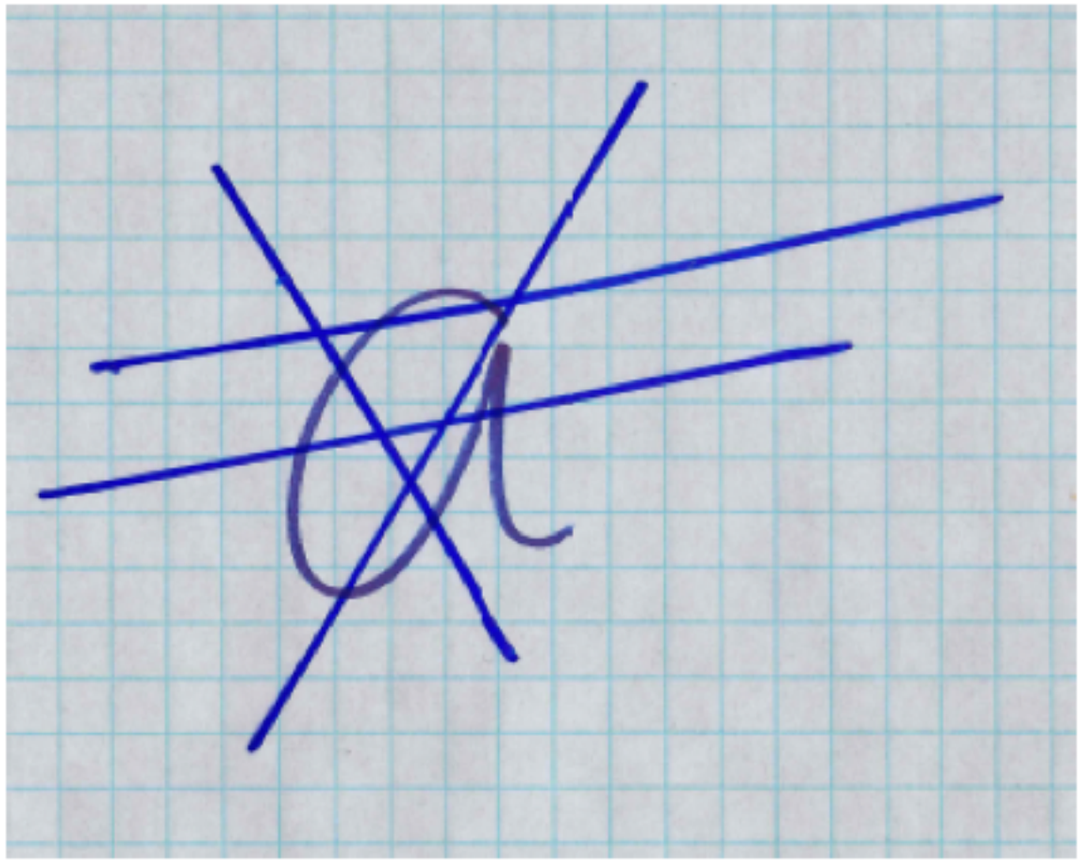
- rotation ~
- pose / angle

e-Lab

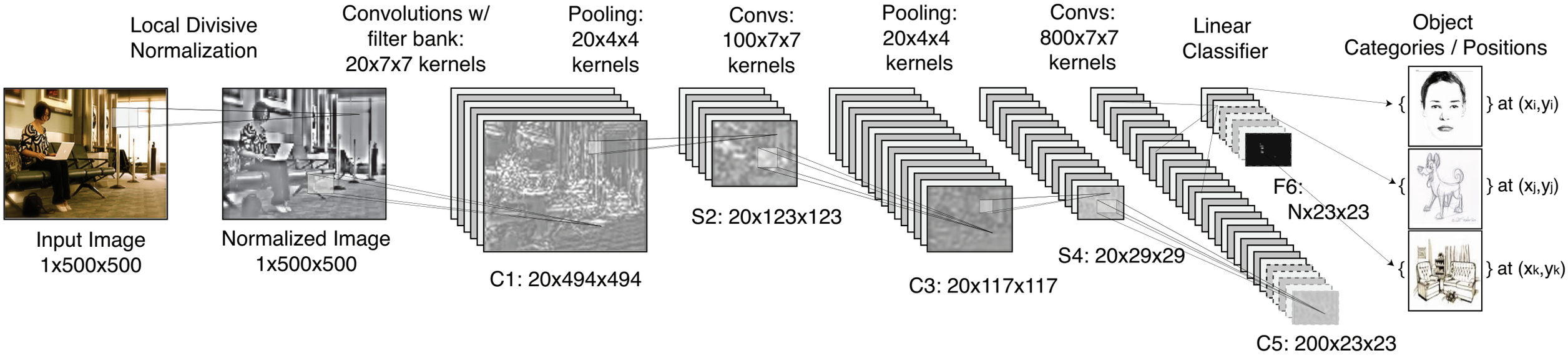
invariance

recognition invariant to:

- noise, distortion



multi-layer neural networks



details of convnet in next lecture by Ayse

cost functions

$$J(\Theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} \log h_\theta(x^{(i)})_k + (1 - y_k^{(i)}) \log(1 - h_\theta(x^{(i)})_k)\right]$$
$$+\frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(\Theta_j^{(l)})^2$$

$$\min_{\Theta} J(\Theta)$$

all neural networks (and ConvNets!) TRAINING
minimizes a cost function
eg.: minimize difference of net out to examples

e-Lab

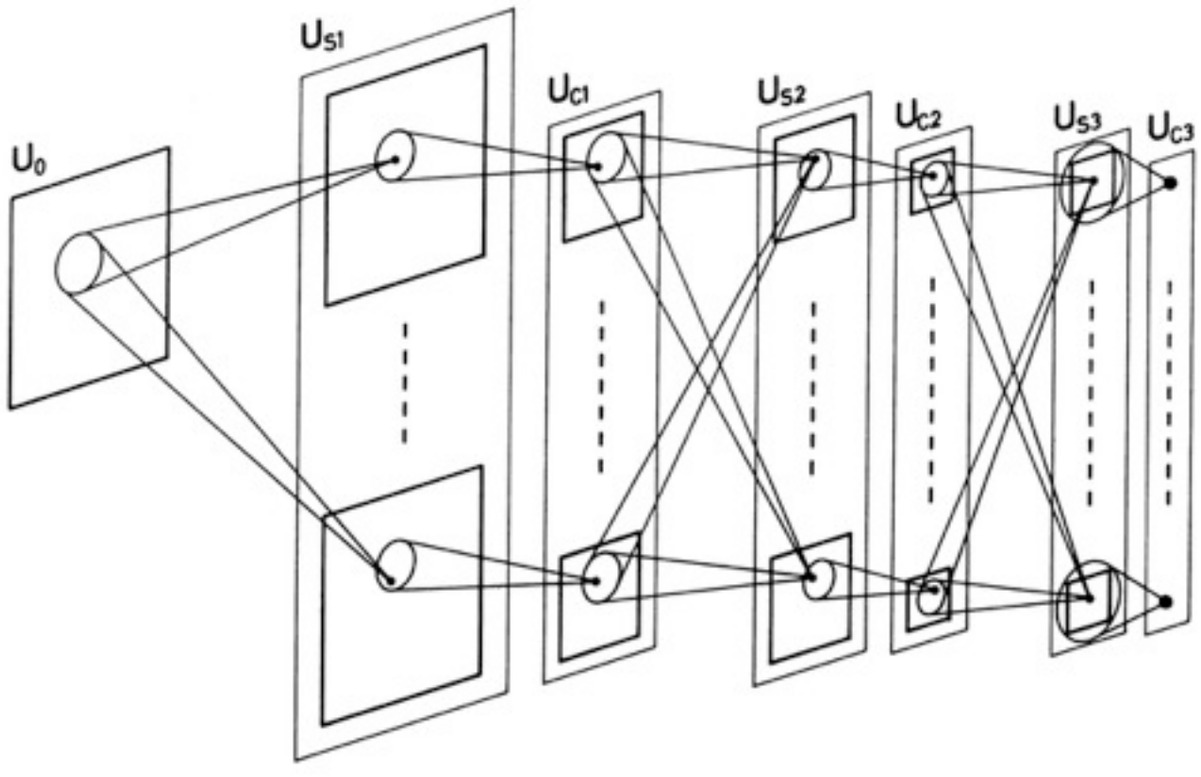
other models

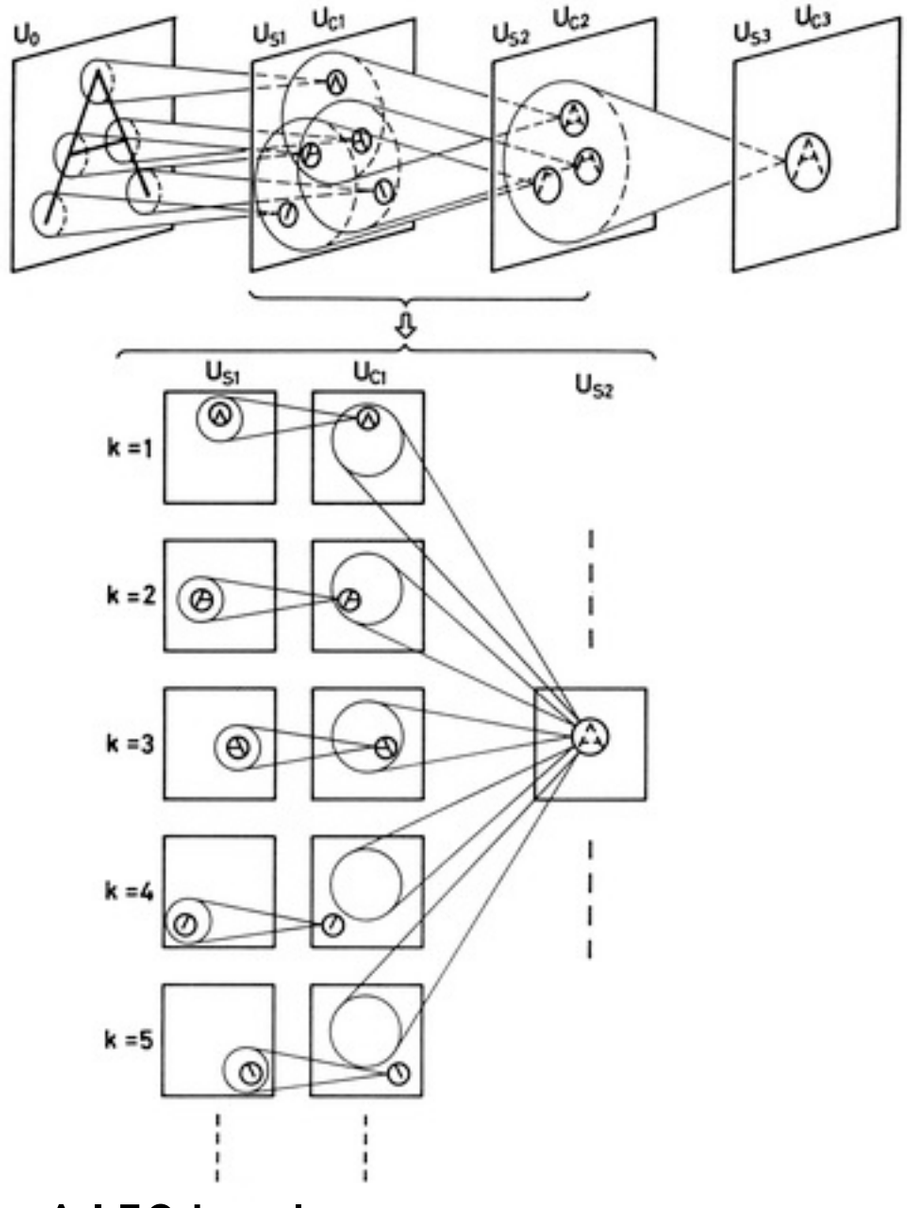
Neocognitron: A Model for Visual Pattern Recognition

Kunihiko Fukushima

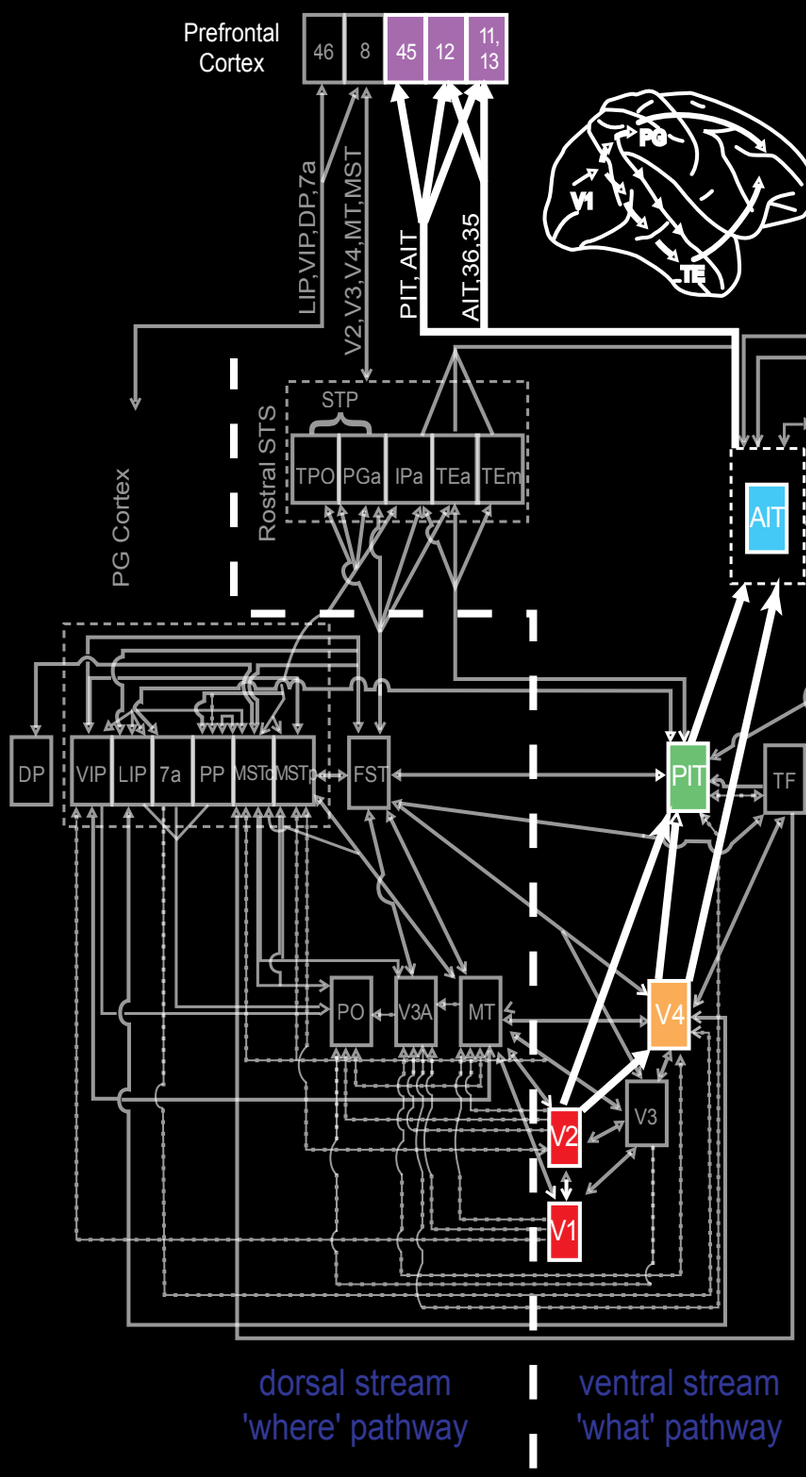
Introduction

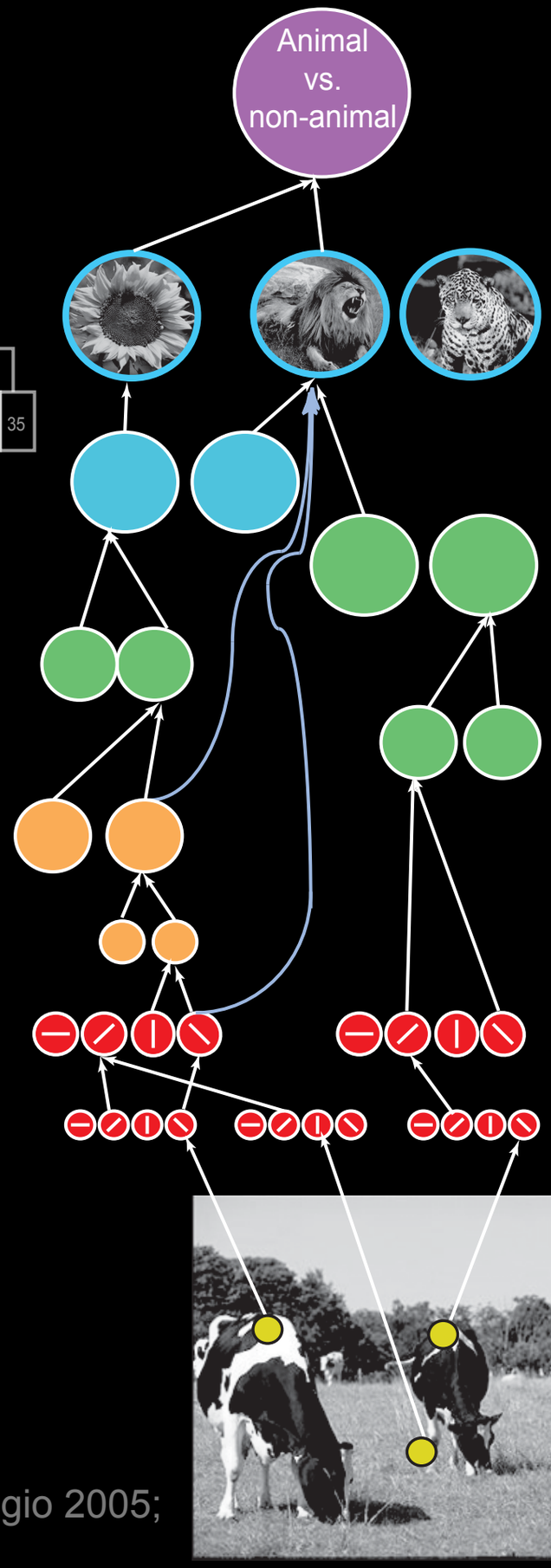
The *neocognitron* (Fukushima, 1980, 1988b, 1991) is a neural network model for deformation-resistant visual pattern recognition.





https://mitpdev.mit.edu/library/erefs/arbib/Arbib_A159.html





(Riesenhuber & Poggio 1999 2000; Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005; Serre Oliva & Poggio 2007)

✦V1:

- Simple and complex cells tuning properties (Schiller et al 1976; Hubel & Wiesel 1965; Devalois et al 1982)
- MAX operation in subset of complex cells (Lampl et al 2004)

✦V4:

- Tuning for two-bar stimuli (Reynolds Chelazzi & Desimone 1999)
- MAX operation (Gawne et al 2002)
- Two-spot interaction (Freiwald et al 2005)
- Tuning for boundary conformation (Pasupathy & Connor 2001)
- Tuning for Cartesian and non-Cartesian gratings (Gallant et al 1996)

✦IT:

- Tuning and invariance properties (Logothetis et al 1995)
- Differential role of IT and PFC in categorization (Freedman et al 2001 2002 2003)
- Read out data (Hung Kreiman Poggio & DiCarlo 2005)
- Average effect in IT (Zoccolan Cox & DiCarlo 2005; Zoccolan Kouh Poggio & DiCarlo in press)

✦Human behavior:

- Rapid animal categorization (Serre Oliva Poggio 2007)