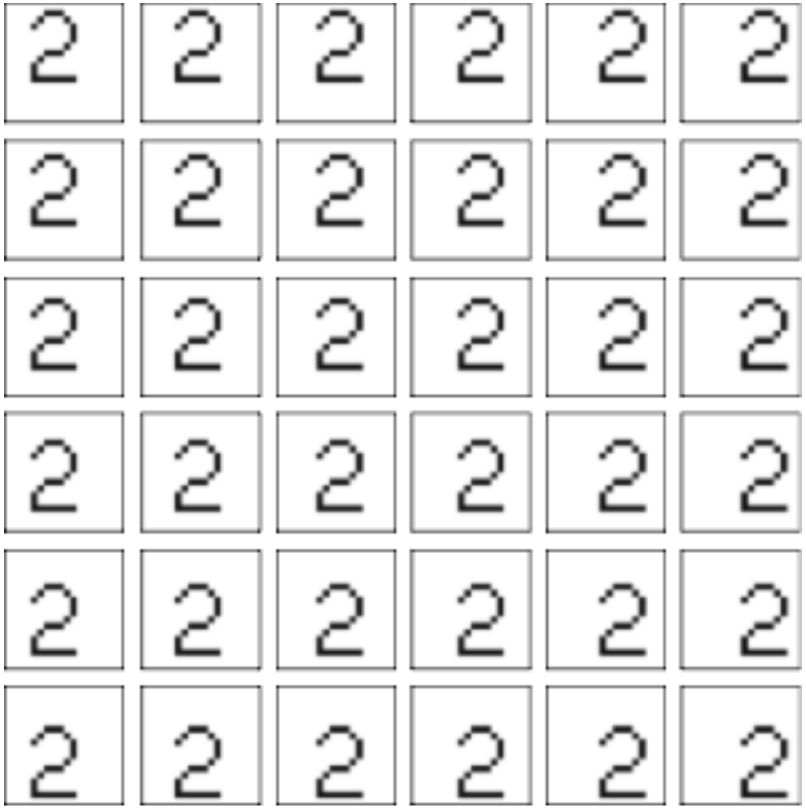
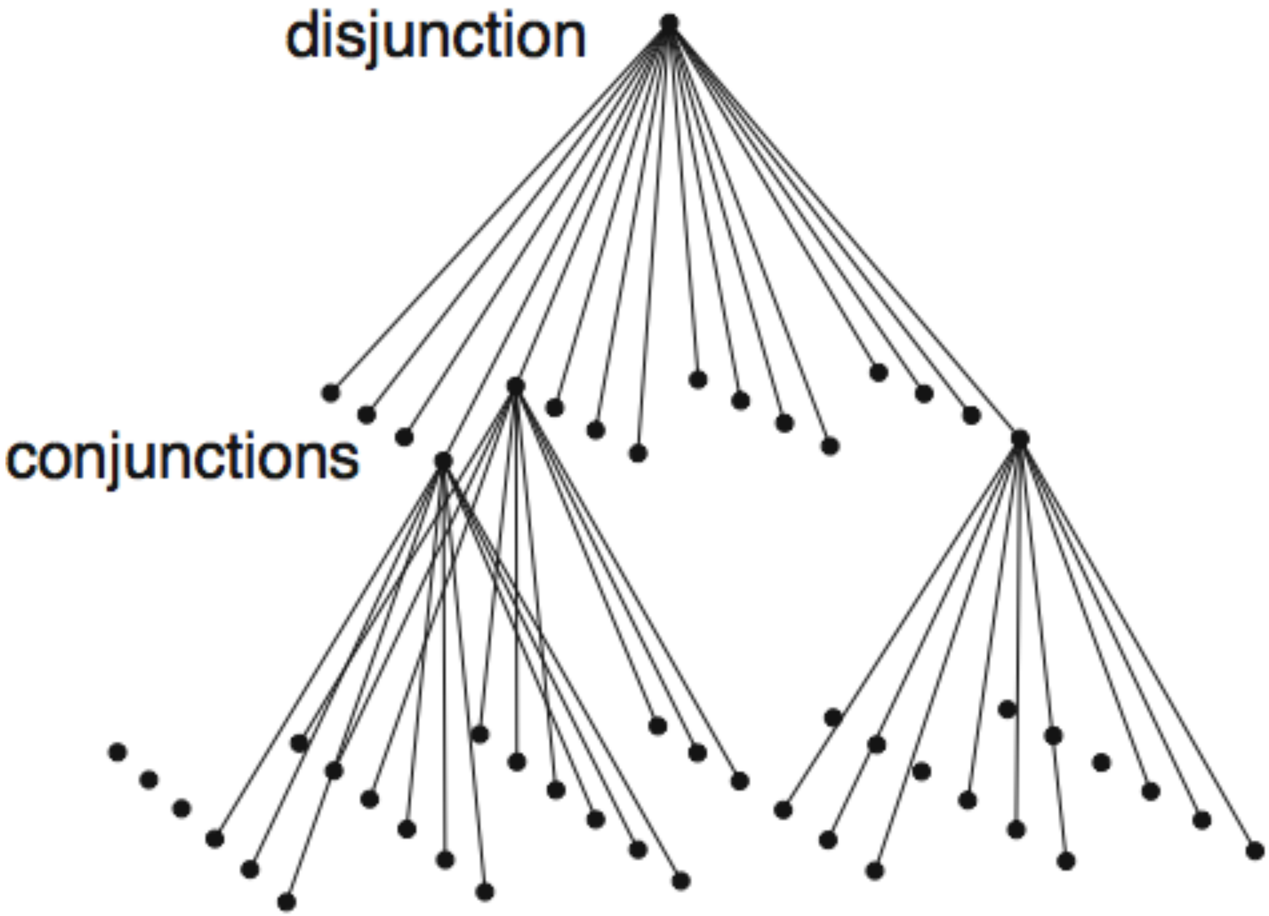
Pooling

Sebastian Seung

Translation-invariant recognition



DNF network



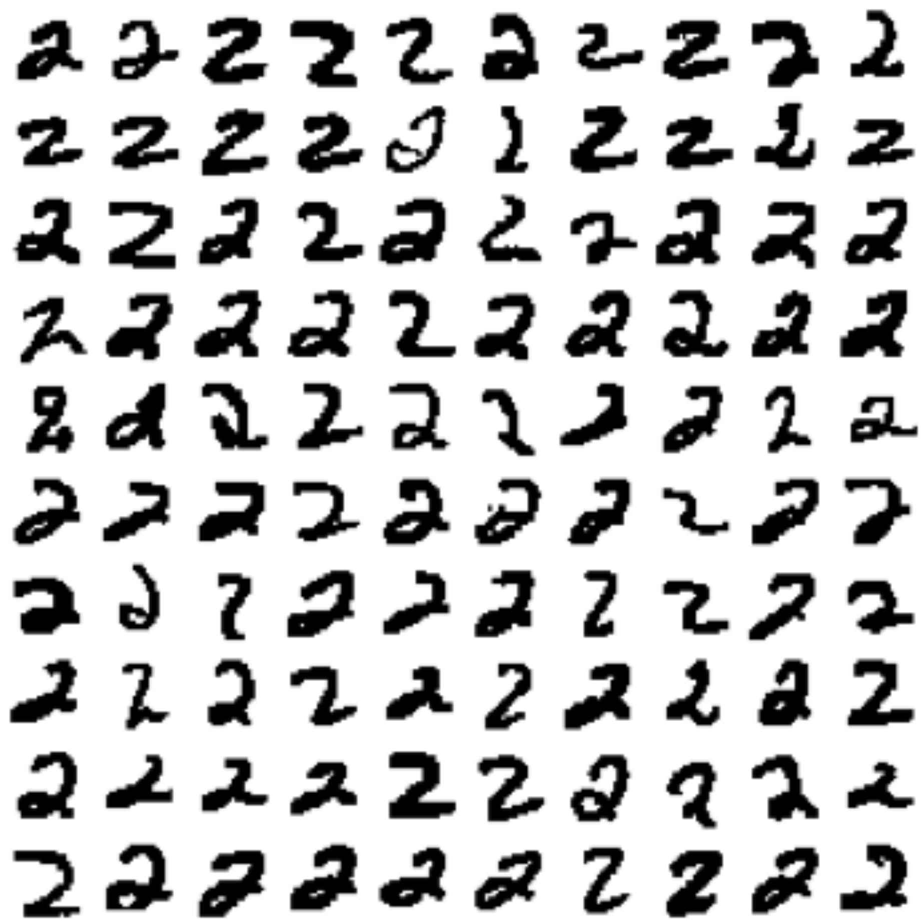
The images of a single object are highly variable.

- Illumination
- Scale
- Translation
- Rotation
- Deformation

Objects in the same class are variable.

- Many different objects are called chairs.

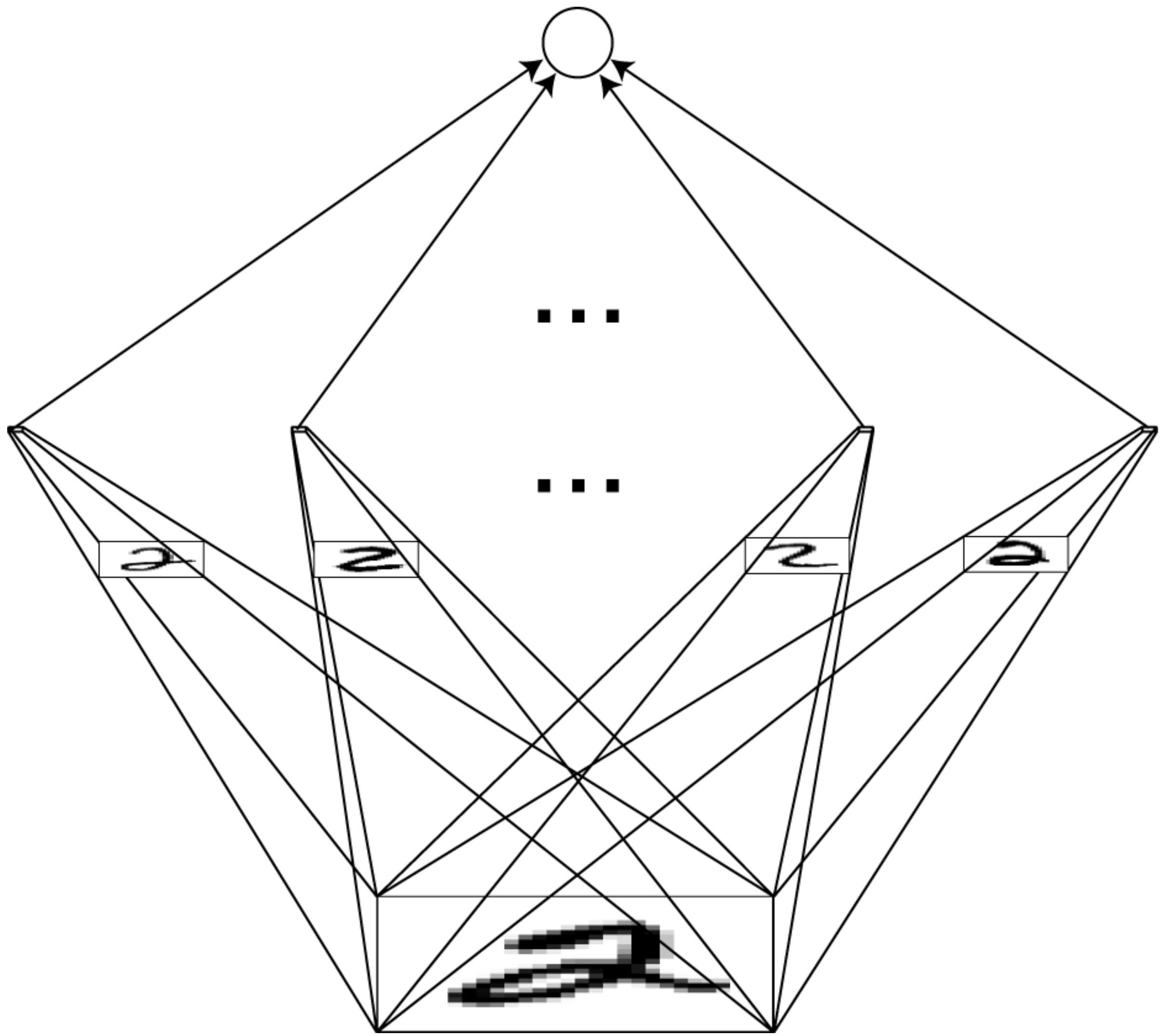
Handwritten digits are variable



Exhaustive enumeration

- Generate all possible images.
- Ask a human expert to classify them.
- Create an MLP using the DNF construction.

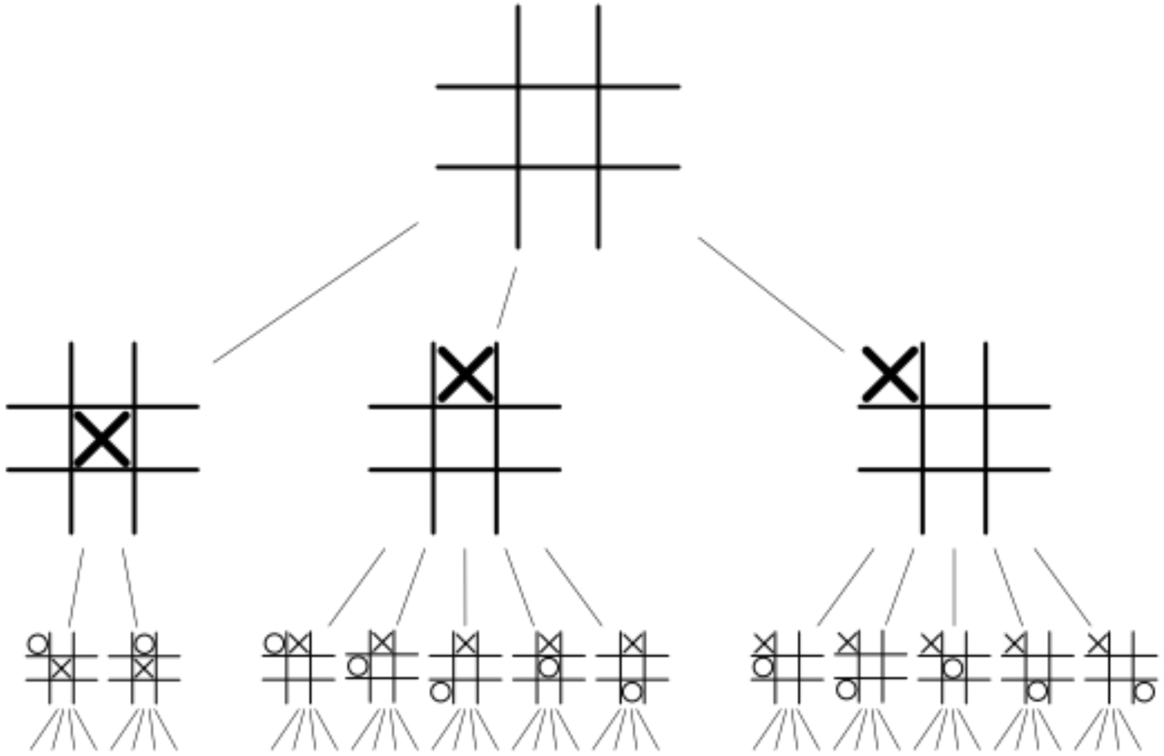
DNF construction again



Exhaustive enumeration for games

- Game tree
- Possible outcomes: win, lose, draw
- Minimax algorithm
  - work backwards in time

Game tree



see also http://xkcd.com/832/

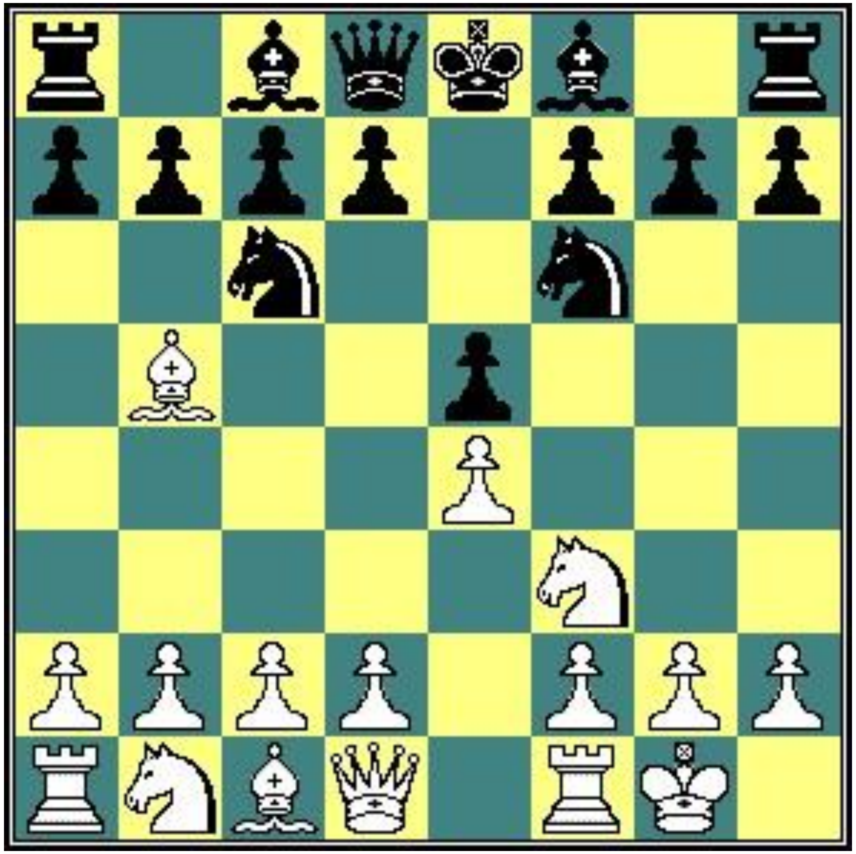
Rough measure of complexity

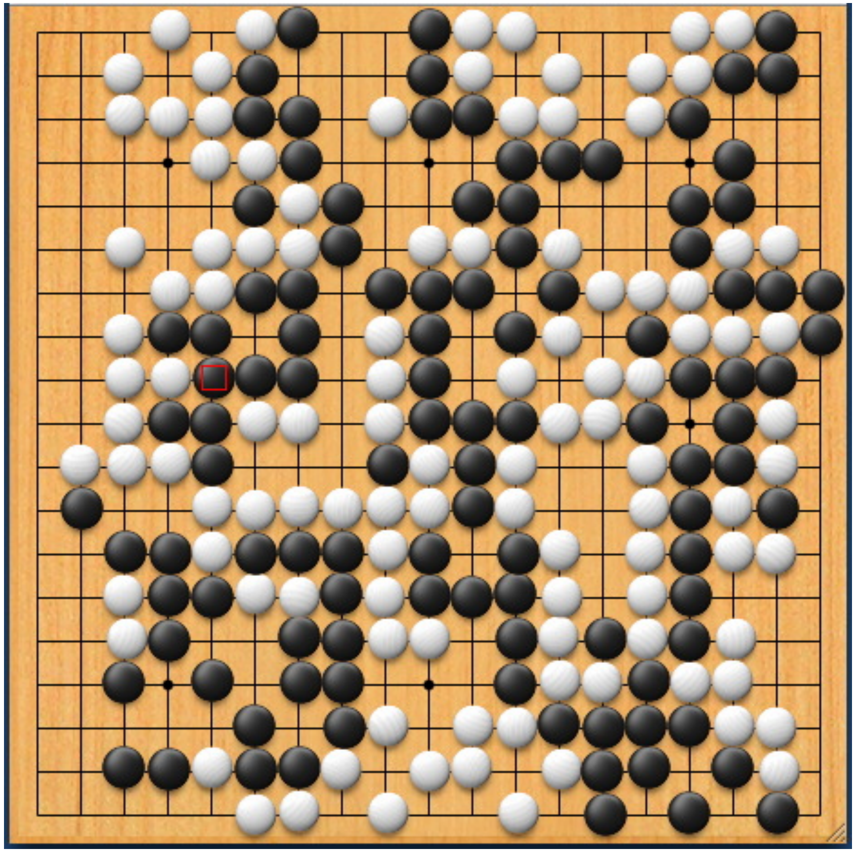
- Chess
  - game tree has 35^{100} nodes
  - 10^{40} configurations
- Megapixel image
  - $2^{1000000}$ binary images
  - $256^{1000000}$ gray scale images

Heuristic search

- Reduce complexity by limiting the time horizon of the search
- Minimax search using evaluation function
- Pruning and other tricks

Chess vs. go





Complexity

	vision	chess
humans	easy	hard
computers	hard	easy

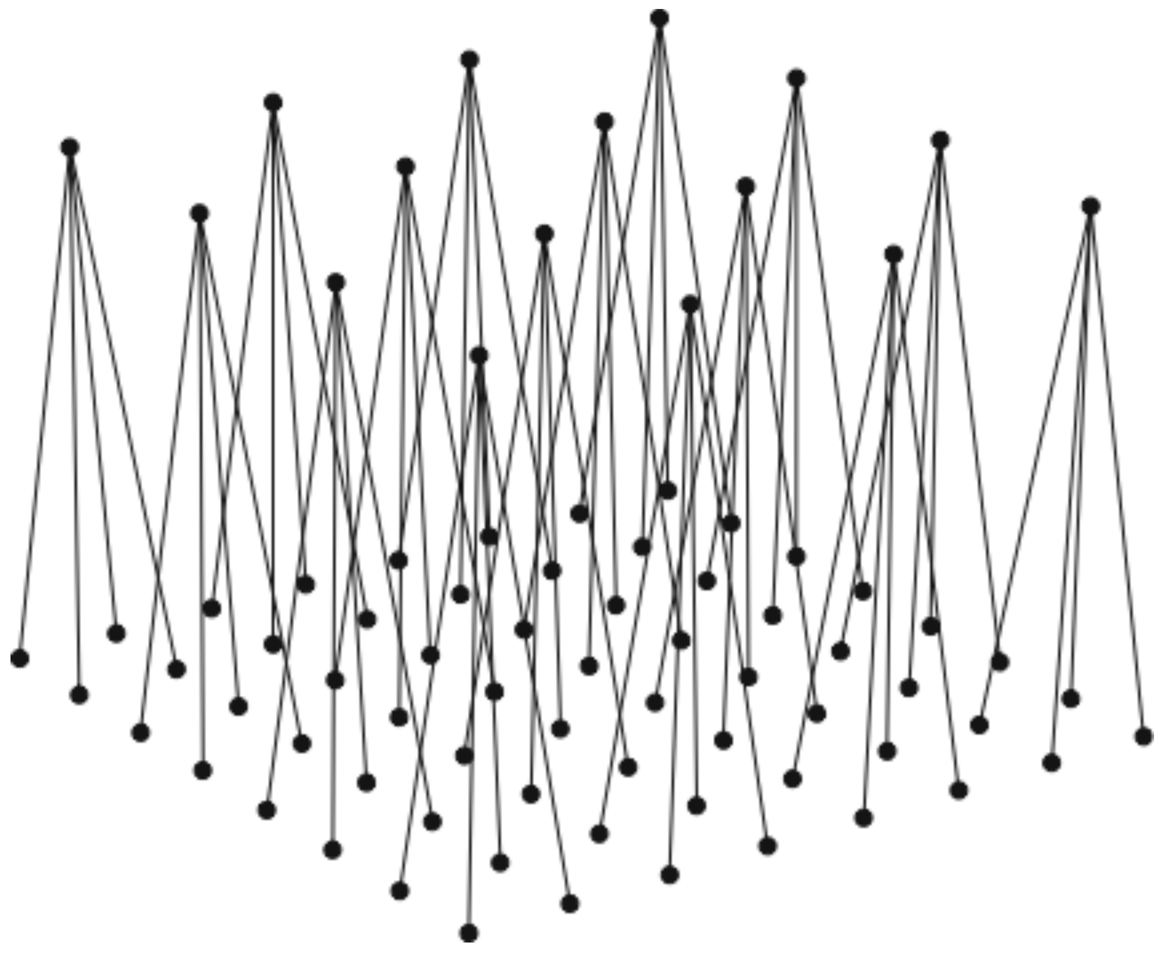
Selective invariance

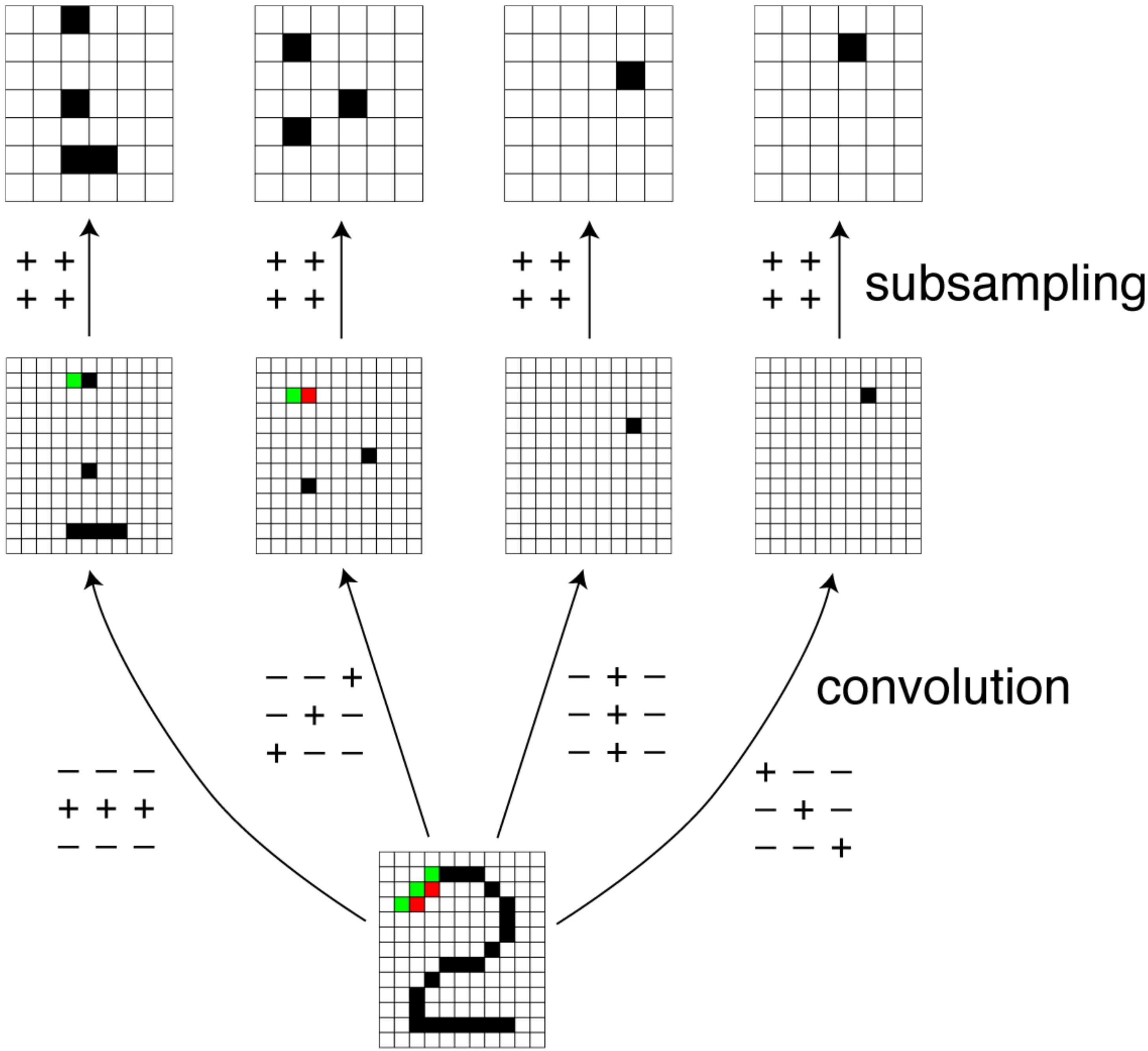
- Selectivity
  - sensitive to inter-class variations
  - i.e. chairs vs. non-chairs.
- Invariance
  - insensitive to intra-class variations
  - i.e. chair 1 vs. chair 2

Hierarchical perceptron model

- Alternation of layers
  - convolution for selectivity
  - subsampling for invariance
- Gradient of selectivity and invariance
  - Neocognitron
  - LeNet
- Hierarchy of feature maps

Subsampling network

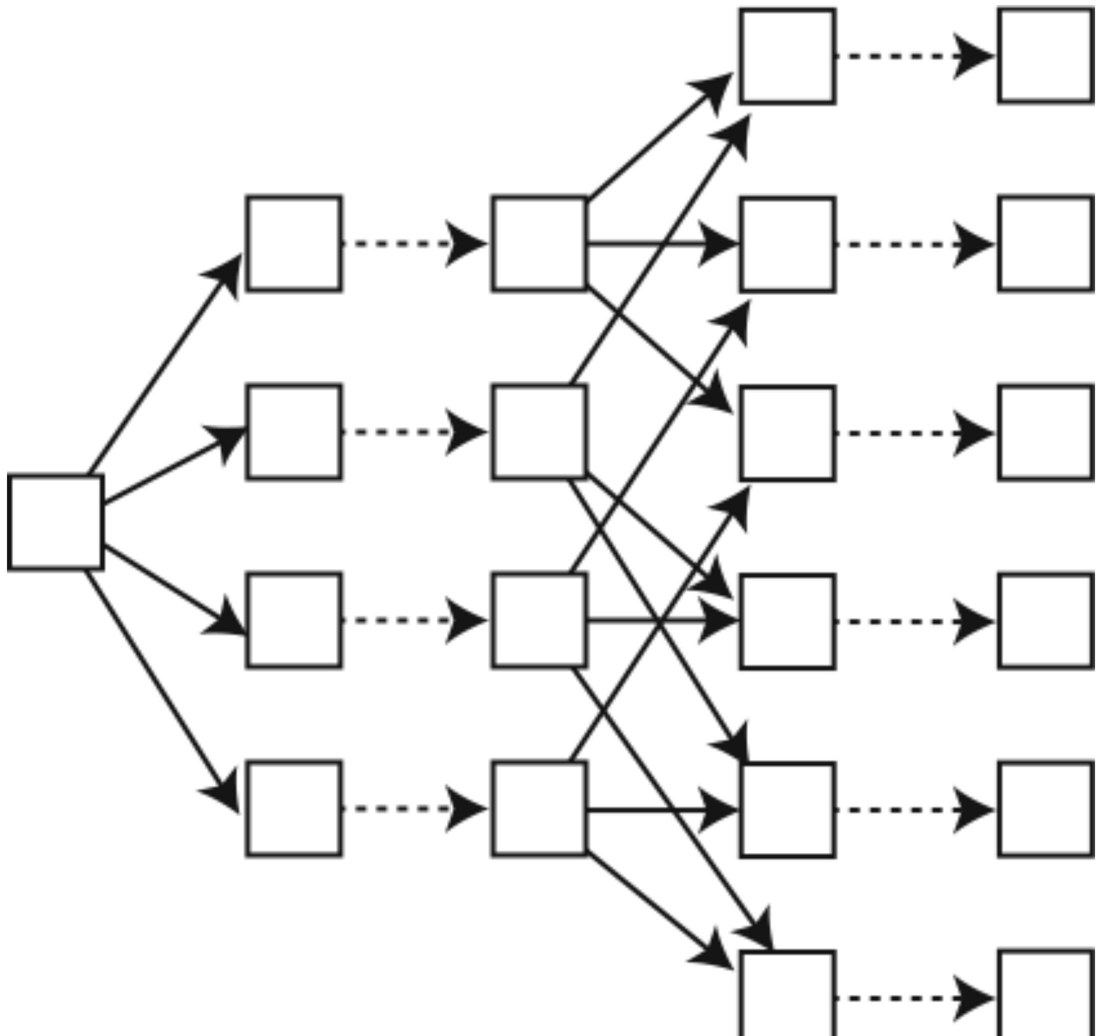



subsampling
convolution

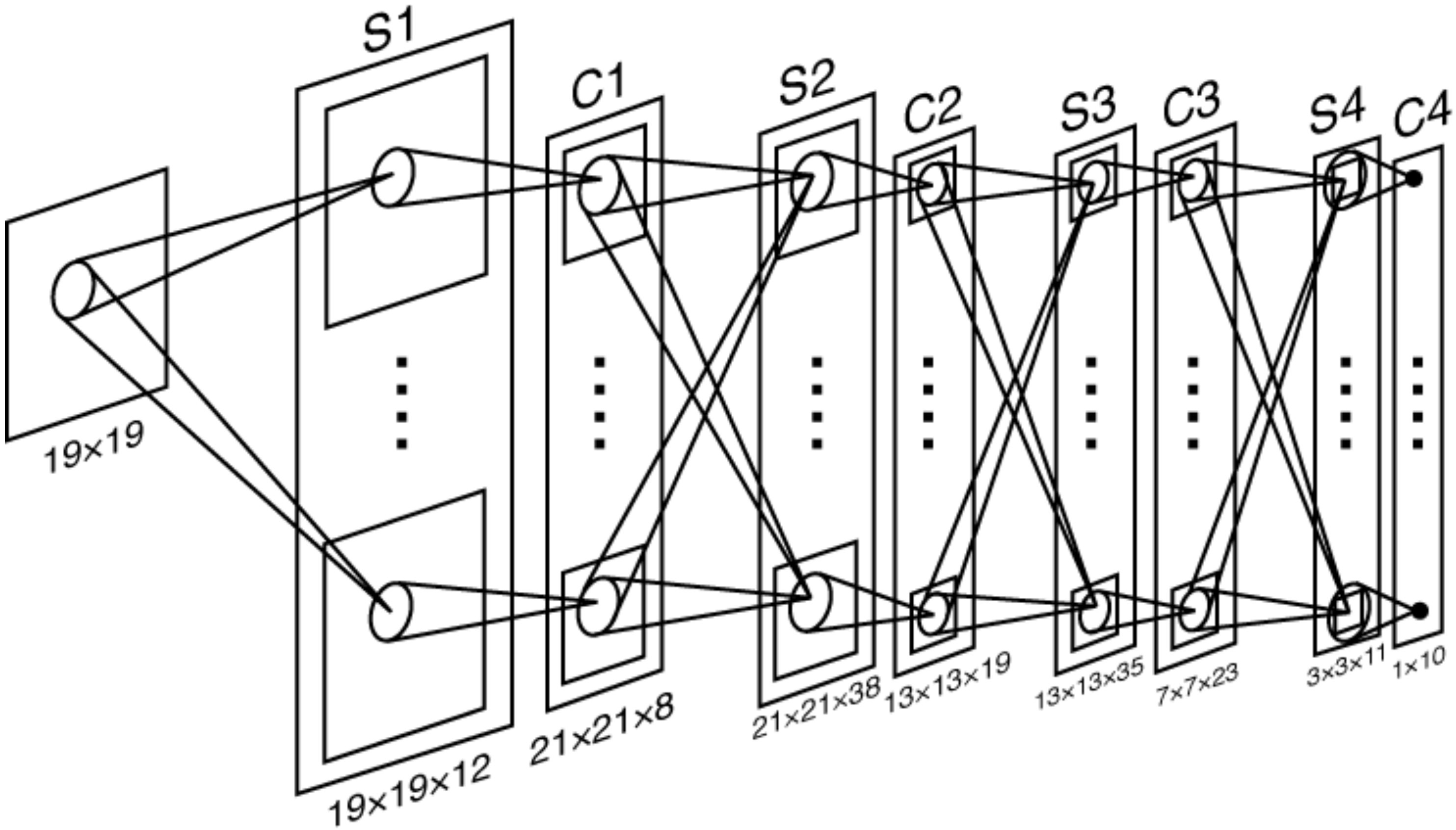
Other invariance

- Subsampling
- Downsampling by averaging
- Max pooling

Alternating convolution and subsampling

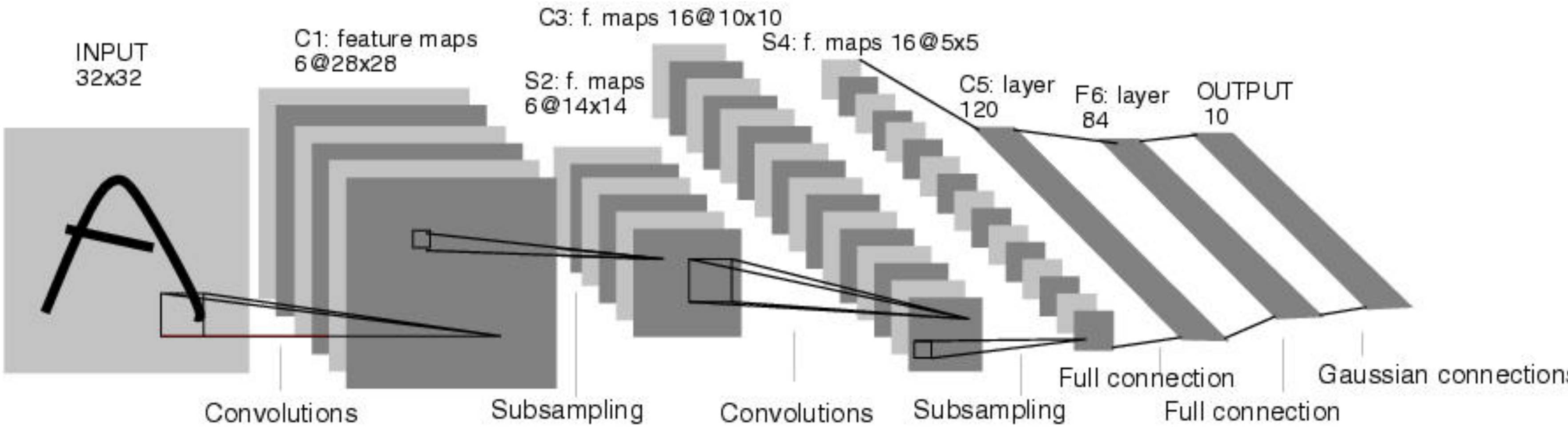


Neocognitron



- Fukushima (1980)

LeNet



LeCun et al. (1989)

Gradient of feature selectivity

- Preferred stimuli of neurons become more complex.
- Receptive fields become larger.
- More invariance to translation, etc.

When is a hierarchical perceptron more efficient?

When the intermediate results of recognition can be shared.

LeNet is fast

- Fast relative to other architectures for recognition
- Speed is the result of an efficient representation
  – relatively small number of synapses

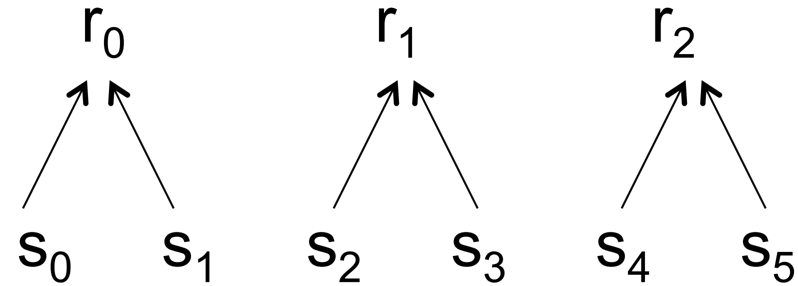
Average pooling (1D)

- Assume zero-based indexing

$$r_i = f\left(w \sum_{j=0}^{n-1} s_{ni+j} + b\right)$$

- $s_0, \ldots, s_{N-1}$ $\quad r_0, \ldots, r_{N/n-1}$

- E.g. for N=6, n=2



Average pooling

- Equivalent to:
  - convolution with filter containing only ones
  - downsampling

Max pooling (1D)

- Forward pass (zero-based indexing)

$$r_i = \max_{j=0,\ldots,n-1} \{s_{ni+j}\}$$

- Backward pass routes delta at output to the winning input

$$\delta^s_{ni+k} = \begin{cases} \delta^r_i, & k = \operatorname{argmax}_{j=0,\ldots,n-1}\{s_{ni+j}\} \\ 0, & \text{otherwise} \end{cases}$$