What’s Missing from Deep Learning?

Bruno Olshausen
Helen Wills Neuroscience Institute, School of Optometry
Redwood Center for Theoretical Neuroscience
UC Berkeley
"We’re at the beginning of a new day… This is the beginning of the AI revolution.”
— Jensen Huang, GTC Taiwan 2017
Among the most challenging scientific questions of our time are the corresponding analytic and synthetic problems: How does the brain function? Can we design a machine which will simulate a brain?

-- *Automata Studies*, 1956
The theory reported here clearly demonstrates the feasibility and fruitfulness of a quantitative statistical approach to the organization of cognitive systems. By the study of systems such as the perceptron, it is hoped that those fundamental laws of organization which are common to all information handling systems, machines and men included, may eventually be understood.” -- Frank Rosenblatt

A brief history of neural networks

1960's

\[ y = g(u) \]

\[ u = \sum_{i} w_i x_i \]
A brief history of neural networks

1980's

\[ u = \sum_{i} w_i x_i \]

\[ y = g(u) \]
A brief history of neural networks

2000’s

\[
\sum x_1 w_1 x_2 w_2 x_3 w_3 \ldots x_n w_n w_0 u \rightarrow g(y) = g(u) = \sum_i w_i x_i
\]
Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position

Kunihiko Fukushima

NHK Broadcasting Science Research Laboratories, Kinuta, Setagaya, Tokyo, Japan
Neocognitron: rationale

Fig. 5. An example of the interconnections between cells and the response of the cells after completion of self-organization.
Neocognitron: activation rule

\[ u_{sl}(k_l, n) = r_l \cdot \varphi \left[ 1 + \frac{\sum_{k_{l-1}=1}^{K_{l-1}} a_i(k_{l-1}, v, k_l) \cdot u_{C_{l-1}}(k_{l-1}, n + v)}{\sum_{v \in S_l} 1 + \frac{2r_l}{1 + r_l} \cdot b_l(k_l) \cdot v_{C_{l-1}}(n)} - 1 \right] \]

where

\[ \varphi[x] = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases} \]

Relu

\[ v_{C_{l-1}}(n) = \sqrt{\sum_{k_{l-1}=1}^{K_{l-1}} \sum_{v \in S_l} c_{l-1}(v) \cdot u_{C_{l-1}}^2(k_{l-1}, n + v)} \]
Neocognitron: learning rule

Let cell $u_{Sl}(k_l, \hat{n})$ be selected as a representative.

$$\Delta a_l(k_{l-1}, v, \hat{k}_l) = q_l \cdot c_{l-1}(v) \cdot u_{Cl-1}(k_{l-1}, \hat{n} + v),$$

Hebbian learning

From each S-column, every time when a stimulus pattern is presented, the S-cell which is yielding the largest output is chosen as a candidate for the representatives. Hence, there is a possibility that a number of candidates appear in a single S-plane. If two or more candidates appear in a single S-plane, only the one which is yielding the largest output among them is selected as the representative from that S-plane. In

Local WTA
Neocognitron: performance

Fig. 6. Some examples of distorted stimulus patterns which the neocognitron has correctly recognized, and the response of the final layer of the network

Fig. 7. A display of an example of the response of all the individual cells in the neocognitron
This isn’t a good model of perception
Relative spatial relationships are important.
Spatial phase, not amplitude, determines shape

Image

Amplitude spectrum
randomize local phase

randomize local amp.
Deep neural networks are easily fooled
(Nguyen, Yosinski & Clune 2014)
Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope*

AUDE OLIVA
Harvard Medical School and the Brigham and Women’s Hospital, 221 Longwood Ave., Boston, MA 02115
oliva@search.bwh.harvard.edu

ANTONIO TORRALBA
Department of Brain and Cognitive Sciences, MIT, 45 Carleton Street, Cambridge, MA 02139
torralba@ai.mit.edu

Figure 4
Figure 15. Organization of natural scenes according to the openness and ruggedness properties estimated by the WDSTs. The rank correlation was 0.82. When using the WDST the rank correlation was 0.87 that was close to the agreement among observers (0.90).

VI. Experimental Results

Each spatial envelope property corresponds to the axes of a multidimensional space into which scenes with similar spatial envelopes are projected close together.

Figures 15 and 16 show a random set of pictures of natural and man-made environments respectively projected in a two-dimensional space corresponding to the openness and ruggedness (or expansion for man-made environments) dimensions. Therefore, scenes close in the space should have the same (or very similar) membership category, whether the spatial envelope...
What’s missing?
\[ \neq g(\sum_i w_i x_i) \]
\[ g(\sum_{i} w_i \prod_{j \in G_i} x_j) \]
Single neuron recording ⇒ Single neuron thinking

What the Frog’s Eye Tells the Frog’s Brain*

J. Y. LETTVIN†, H. R. MATUSHANA‡, W. S. McCULLOCH¶, SENIOR MEMBER, IRE,
AND W. H. PITTS||

Summary—In this paper, we analyze the activity of single fibers in the optic nerve of a frog. Our method is to find what sort of stimulus causes the largest activity in one nerve fiber and then what is the exciting aspect of that stimulus such that variations in everything else cause little change in the response. It has been known for the past 20 years that each fiber is connected not to a few rods and cones in the retina but to very many over a fair area. Our results show that for it moves like one. He can be fooled easily not only by a bit of dangled meat but by any moving small object. His sex life is conducted by sound and touch. His choice of paths in escaping enemies does not seem to be governed by anything more devious than leaping to where it is darker. Since he is equally at home in water and on

factor. There are four types of fibers, each type concerned with a different sort of pattern. Each type is uniformly distributed over the visual image in terms of local pattern independent of average illumination. We describe the patterns and show the functional and anatomical separation of the channels. This work has been done on the frog, and our interpretation applies only to the frog.

anatomy of frog visual apparatus

The retina of a frog is shown in Fig. 1(a). Between the rods and cones of the retina and the ganglion cells, whose axons form the optic nerve, lies a layer of con-
Cortical circuits

- highly organized by layer
- layers are interconnected in a ‘canonical microcircuit’
- signals are strongly intermixed within layers 2/3

(Douglas and Martin, 2007)
Feedback is pervasive throughout the thalamo-cortical system

Diagram showing the flow of information from retina to LGN, then to V1, V2, V4, pulvinar, and IT.
Two specific proposals

1. Dynamic routing

2. Hierarchical Bayesian inference
Reference frame effects in perception

Diamond or square?
Which way are the triangles pointing?

From Attneave
Reference frames require *structured representations*

Hinton (1981)
Dynamic routing
(Olshausen, Anderson, Van Essen 1993)
Dynamic routing circuit
Dynamic routing: control

\[ I_{i_{out}} = \sum_j w_{i,j} I_{i_{in}}^j \]

\[ w_{i,j} = \sum_k c_k \Gamma_{i,j,k} \]
Dynamic routing: control

a.
\[ w_{ij} \]

b.

Input \[ j \] \[ i \] Output

\[ \text{window of attention} \]

c.
\[ \text{window of attention} \]

d.
\[ \text{window of attention (aliased)} \]
Pattern matching via dynamic routing

\[ \Gamma_{ijk} \]

\[ V_i \]

\[ I_i^{in} \]

Control

(other control units)
Pattern matching via dynamic routing
Dynamic routing in deep networks

Our model, layers 6,7: 8192 units
With layers 6,7: 4096 units
With removed layer 7
With removed Layers 3,4,6,7
Krizhevsky et al. 2012

Table 1. Various architectural changes to our ImageNet model.

<table>
<thead>
<tr>
<th>Error %</th>
<th>Top-1 Train</th>
<th>Top-1 Val</th>
</tr>
</thead>
<tbody>
<tr>
<td>44.7</td>
<td>81.5</td>
<td>71.6</td>
</tr>
<tr>
<td>47.3</td>
<td>82.7</td>
<td>72.6</td>
</tr>
<tr>
<td>50.3</td>
<td>84.3</td>
<td>74.3</td>
</tr>
<tr>
<td>51.9</td>
<td>85.3</td>
<td>75.0</td>
</tr>
<tr>
<td>45.7</td>
<td>81.7</td>
<td>71.7</td>
</tr>
<tr>
<td>47.3</td>
<td>82.7</td>
<td>72.7</td>
</tr>
<tr>
<td>50.3</td>
<td>84.3</td>
<td>74.3</td>
</tr>
<tr>
<td>51.9</td>
<td>85.3</td>
<td>75.0</td>
</tr>
<tr>
<td>45.7</td>
<td>81.7</td>
<td>71.7</td>
</tr>
<tr>
<td>47.3</td>
<td>82.7</td>
<td>72.7</td>
</tr>
<tr>
<td>50.3</td>
<td>84.3</td>
<td>74.3</td>
</tr>
<tr>
<td>51.9</td>
<td>85.3</td>
<td>75.0</td>
</tr>
<tr>
<td>45.7</td>
<td>81.7</td>
<td>71.7</td>
</tr>
<tr>
<td>47.3</td>
<td>82.7</td>
<td>72.7</td>
</tr>
<tr>
<td>50.3</td>
<td>84.3</td>
<td>74.3</td>
</tr>
<tr>
<td>51.9</td>
<td>85.3</td>
<td>75.0</td>
</tr>
</tbody>
</table>

(Zeiler & Fergus, 2013)
Visualization of filters learned at intermediate layers (Zeiler & Fergus 2013)

Layer 3

Figure 8. Visualization of features in a fully trained model. For layers 2-5 we show the top 9 activations in a random subset of feature maps across the validation data, projected down to pixel space using our deconvolutional network approach. Our reconstructions are not samples from the model: they are reconstructed patterns from the validation set that cause high activations in a given feature map. For each feature map we also show the corresponding image patches. Note: (i) the strong grouping within each feature map, (ii) greater invariance at higher layers and (iii) exaggeration of discriminative parts of the image, e.g. eyes and noses of dogs (layer 4, row 1, cols 1). Best viewed in electronic form.
Perception as inference
Is this the goal of perception?
What do these edges mean?

reflectance

shading

(p, q, r)
Vision as inference

World  Image  Model
Hierarchical Bayesian inference in visual cortex  
(Lee & Mumford, 2003)

Hierarchical Bayesian inference in visual cortex  
(Lee & Mumford, 2003)
output (y)

input data (x)

\[ y = f(x; w) \]
Main points

• Multilayer perceptrons were a good idea in 1960’s

• Neocognitron was a good idea in 1980’s

• The way forward:
  - identify the right problems to be solved
  - exploit the computational richness offered by real neurons and cortical circuits

• Two examples:
  - Dynamic routing
  - Hierarchical Bayesian inference