Deep networks: The good, the bad, and the ugly

The future: Neuroscience-inspired deep nets

Thomas Serre
Brown University
Cognitive Linguistic & Psychological Sciences
Carney Institute for Brain Science

ANR-3IA Artificial and Natural Intelligence Toulouse Institute
Rapid categorization
Ventral stream of the visual cortex

- Very fast latency <50-70 ms
- Category information <90-100 ms
- Primarily feedforward (with limited feedback)
- 5-8 processing stages

Image source: Jim Dicarlo

Cauchoix* Crouzet* Fize & Serre NeuroImage 2016; see also Hung Kreiman et al 2005
Computational models

- Neocognitron, VisNet, SpikeNet, ConvNet, HMAX, etc
- Feedforward cascade + 3 operations (linear filtering, non-linear rectification and pooling)

See Serre Ann Rev of Vision Science 2019 for a review
Computational models

- Neocognitron, VisNet, SpikeNet, ConvNet, HMAX, etc
- Feedforward cascade + 3 operations (linear filtering, non-linear rectification and pooling)

See Serre Ann Rev of Vision Science 2019 for a review
Computational models

- Neocognitron, VisNet, SpikeNet, ConvNet, HMAX, etc
- Feedforward cascade + 3 operations (linear filtering, non-linear rectification and pooling)
Computational models

- Neocognitron, VisNet, SpikeNet, ConvNet, HMAX, etc
- Feedforward cascade + 3 operations (linear filtering, non-linear rectification and pooling)

See Serre Ann Rev of Vision Science 2019 for a review
Computational models

- Consistent with anatomy and physiology
- Consistent with H and non-H primate decisions during rapid categorization

See Serre Ann Rev of Vision Science 2019 for a review
The good
The deep net revolution

Object recognition (ImageNet)

face recognition (MegaFace)

See Serre Ann Rev of Vision Science 2019 for a review
The deep net revolution

Source: Yamins & DiCarlo 2016
The bad
The deep net revolution
A petaflop/s-day (pfs-day) consists of performing $10^{15}$ neural net operations per second for one day, or a total of about $10^{20}$ operations.
The deep net revolution

Source: http://sqlml.azurewebsites.net/category/artificial-intelligence/deep-learning
Deep net limitations

Unsustainable growth
Deep net limitations

<table>
<thead>
<tr>
<th></th>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glasses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bike</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tie</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bug</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reliability

- Experiment 1 vs. Experiment 2: $\rho = 0.55^{***}$
- Experiment 1: $\rho = 0.79^{***}$
- Experiment 2: $\rho = 0.60^{***}$
Help the AI recognize this image before time runs out!
The top-5 scoring players by Sun Apr 23 2017 12:00:00 EST win a gift card! See the Scoreboard tab for details.

Click and then brush with your mouse to reveal image parts best describing:
king snake, or kingsnake

Your score: 0.00  |  High score: 113950.85

Our AI will evolve after this many images are played:
Deep net limitations

Reliability
- LRP $\rho = 0.33^{**}$
- CAM $\rho = 0.1$, ns

Sensitivity $\rho = 0.03$, ns
The uggly
DNN limitations

“Object transplanting”
Limited generalization

Graph showing the accuracy of different detectors over the years, with a significant drop in performance for ObjectNet from 2016 to 2017, indicating a 40-45% drop in accuracy.
Generalizing to unseen degradations
Limited generalization

(a) Super-human performance
(b) Super-human performance
(c) Chance level performance
Limited generalization
Limited generalization

Sort-of-CLEVR dataset (Santoro et al 2017)
Limited generalization

- train: test colour (cyan) present
- train: test shape (square) present
- test: novel colour × shape combination (cyan square) present

"Relational network"
One-shot learning
In newborn ducklings

- Computational evidence that same-different recognition task trivially solved w/ perceptual grouping mechanisms
The future
Attention and cortical feedback

Alamia Luo Kim Ricci & Serre VanRullen in prep
Driving DCNs towards human-like recognition

Objective 1: Object classification: This is a truck

Objective 2: Learn attention mask (feature-based and spatial)
Help the AI recognize this image before time runs out!

The top-5 scoring players by Sun Apr 23 2017 12:00:00 EST win a gift card! See the Scoreboard tab for details.

Click and then brush with your mouse to reveal image parts best describing:
king snake, or kingsnake

Your score: 0.00  |  High score: 113950.85

Our AI will evolve after this many images are played:

0 1000 2000 3000 4000 5000
# Driving DCNs towards human-like recognition

<table>
<thead>
<tr>
<th></th>
<th>top-1 err</th>
<th>top-5 err</th>
<th>maps</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE-ResNet-50</td>
<td>66.17</td>
<td>42.48</td>
<td>64.36**</td>
</tr>
<tr>
<td>ResNet-50</td>
<td>63.68</td>
<td>40.65</td>
<td>43.61</td>
</tr>
<tr>
<td>GALA-ResNet-50 no ClickMe</td>
<td>53.90</td>
<td>31.04</td>
<td>64.21**</td>
</tr>
<tr>
<td>GALA-ResNet-50 w/ ClickMe</td>
<td><strong>49.29</strong></td>
<td><strong>27.73</strong></td>
<td><strong>88.56</strong></td>
</tr>
<tr>
<td>Class</td>
<td>GALA w/ ClickMe maps</td>
<td>GALA no ClickMe maps</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------------------</td>
<td>----------------------</td>
<td></td>
</tr>
<tr>
<td>Semi</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>Bus</td>
<td><img src="image3" alt="Image" /></td>
<td><img src="image4" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>Plane</td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>Boat</td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>Lion</td>
<td><img src="image9" alt="Image" /></td>
<td><img src="image10" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>Horse</td>
<td><img src="image11" alt="Image" /></td>
<td><img src="image12" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>Eagle</td>
<td><img src="image13" alt="Image" /></td>
<td><img src="image14" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>Collie</td>
<td><img src="image15" alt="Image" /></td>
<td><img src="image16" alt="Image" /></td>
<td></td>
</tr>
</tbody>
</table>

**ClickMe maps** and **Gradient Δ** shown for different layers: Layer 27, Layer 33, Layer 39.
Feedback mechanisms

- Deep feedforward nets (CNNs) have become orders of magnitude deeper than our visual system.
- Feedback plays key role in visual perception beyond feedforward sweep.
Feedback mechanisms

- Visual processing builds “depth” dynamically through recurrent connections
Feedback mechanisms

- Visual processing builds “depth” dynamically through recurrent connections.
Feedback mechanisms

classical receptive field (CRF)

extra-classical receptive field (eCRF)

Top-down

Horizontal

Feedforward

To higher area
Recurrent neural circuit model

\[ C^{(1)}_{xyk} = (W^I \ast H^{(2)})_{xyk} \]
\[ C^{(2)}_{xyk} = (W^E \ast H^{(1)})_{xyk} \]

\[ \eta \dot{H}^{(1)}_{xyk} + \epsilon^2 H^{(1)}_{xyk} = \left[ \xi X_{xyk} - (\alpha H^{(1)}_{xyk} + \mu) C^{(1)}_{xyk} \right]_+ \]
\[ \tau \dot{H}^{(2)}_{xyk} + \sigma^2 H^{(2)}_{xyk} = \left[ \gamma C^{(2)}_{xyk} \right]_+ \].

Mely Linsley & Serre Psych Review 2018
CRF and eCRF mechanisms

- cross-orientation normalization
- surround suppression (multi-modality)
- feature-selective surround suppression
- size tuning

Busse, Wade & Carandini (2009)
DeAngelis, Freeman & Ohzawa (1994)
Cavanaugh, Bair & Movshon (2002)
Bradley & Andersen (1998)
Trott & Born (2015)

Xing, Shapley, Hawken, Ringach (2005)
Recurrent neural circuit model
Recurrent neural circuit model
Recurrent network model

2 opponent surround regions
near = excitation
far = inhibition

tuned connections from surround

asymmetry of excitation and inhibition
Recurrent neural circuit model
Recurrent neural circuit model

- e.g., repulsion in tilt illusion

Neurophysiology: Gilbert & Wiesel 1990; Thomas et al 2002; Kusunoki et al 2006; Li et al 1999
Modeling: Grossberg et al 2011
A new taxonomy of contextual phenomena

<table>
<thead>
<tr>
<th>separate</th>
<th>competition</th>
<th>cooperation</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Diagram" /></td>
<td><img src="image2.png" alt="Diagram" /></td>
<td><img src="image3.png" alt="Diagram" /></td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>A+C</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td></td>
</tr>
</tbody>
</table>

**References**

- **orientation tilt effect** (Goddard et al. 2008)
- **depth induction** (Westheimer 1986; Westheimer & Levi 1987)
- **orientation tilt effect** (Gibson & Radner 1937; O’Toole & Wenderoth 1977)
- **color induction** (Klauke & Wachtler 2015)
- **motion direction tilt effect** (Marshak & Sekuler 1979)
- **motion speed induction** (Loomis & Nakayama 1973; Noman et al. 1996; Baker et al. 2010)
- **enhanced color induction** (Monnier & Shevell 2003; Shevell & Monnier 2005)
From a neural circuit...
... to an ML module

\[
C_{xyk}^{(1)} = (W^I \ast H^{(2)})_{xyk}
\]

\[
C_{xyk}^{(2)} = (W^E \ast H^{(1)})_{xyk}
\]

\[
\eta \dot{H}_{xyk}^{(1)} + \epsilon^2 H_{xyk}^{(1)} = \left[ \xi X_{xyk} - (\alpha H_{xyk}^{(1)} + \mu) C_{xyk}^{(1)} \right]_+
\]

\[
\tau \dot{H}_{xyk}^{(2)} + \sigma^2 H_{xyk}^{(2)} = \left[ \gamma C_{xyk}^{(2)} \right]_+
\]
Horizontal GRU (hGRU)
Disentangling neural mechanisms for perceptual grouping

Pathfinder
Positive
Negative

Shorter curves

Longer curves

cABC
Positive
Negative

Correlated transformation

Independent transformation

Easy
Intermediate
Hard

Kim* Linsley* Thakkar & Serre ICLR 2020
Berkeley image segmentation dataset
Recurrent neural circuits for contours detection
Recurrent neural circuits for contours detection

Human Response (O’Toole and Wenderoth, 1977)

Perceived shift in orientation

Repulsion

Attraction

Center-Surround orientation difference

Models’ perceived shift in orientation

γ-Net
(Ours)

BDCN
(He et al., 2019)

Center-Surround orientation difference
Recurrent neural circuits for contours detection

\[ \Delta = 0.043 \pm 0.003 \]
\[ T = 13.570, p < 0.001 \]
Model has learned orientation-tuned CRFs matched in size to V1

...and purely modulatory extra-classical receptive fields...

...which combine to contribute surround suppression
V1-like stimulus competition
(Busse, Wade, & Carandini 2009)

A

0%

6%

12%

50%

0%

12%

V1 population responses

Circuit population responses
Tuned surround suppression
(Trott & Born, 2015)
Feature-selective surround suppression (Trott & Born, 2015)
Optogenics on the circuit model

eCRF connectivity is tuned

Chettih & Harvey 2019
Optogenics on the circuit model

Chettih & Harvey 2019
Results on large-scale computer vision challenges

(a) image  (b) semantic segmentation
(c) instance segmentation  (d) panoptic segmentation

Panoptic Quality (PQ)

<table>
<thead>
<tr>
<th>Model</th>
<th>Memory (GB)</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBP FPN</td>
<td>9</td>
<td>783,552</td>
</tr>
<tr>
<td>BPTT FPN</td>
<td>11</td>
<td>783,552</td>
</tr>
<tr>
<td>RBP FPN</td>
<td>12</td>
<td>783,552</td>
</tr>
<tr>
<td>Feedforward FPN-ResNet 50</td>
<td>13</td>
<td>+783,552 parameters</td>
</tr>
</tbody>
</table>
Feedforward

C-RBP
“Coloring” emerges from stable recurrent circuits
“Coloring” emerges from stable recurrent circuits
“Coloring” emerges from stable recurrent circuits
Summary

- New breed of highly-recurrent neural nets inspired by cortical circuits found in visual cortex
- Tilt illusion (and others!) as byproduct of neural circuits to help biological visual systems achieve robust and efficient perception
  - 2 opponent surround regions (exc. vs. inh)
  - “Tuned” (like-to-like) connections from eCRF
  - Asymmetry of exc. and inh.
- SOTA accuracy for contours detection, neural tissue segmentation and panoptic segmentation
Acknowledgments

Drew Linsley
Matt Ricci
Lakshmi Govindarajan
Alekh Ashok
Rex Liu
Junkyung Kim (now DeepMind)
David Mely (now Google Brain)