CONVOLUTIONAL NETS HAVE VISUAL ILLUSIONS
(look similar to ours but they are not the same)

- EIGEN-ANALYSIS
- ARTIFICIAL PSYCHOPHYSICS

JESÚS MALO

ALEX GOMEZ
ADRIAN MARTÍN
JAVIER VAZQUEZ
MARCELO BERTALHIO

Universitat Pompeu Fabra
Barcelona
SEEING IS A SUBJECTIVE EXPERIENCE!
SEEING IS A SUBJECTIVE EXPERIENCE!
DO WE SEE THE SAME?
DO WE SEE THE SAME?

- Deep learning $\rightarrow$ Neuroscience
- Neuroscience $\rightarrow$ Deep learning

Goal functions
Psychophysics
VISUAL ILLUSIONS?
VISUAL ILLUSIONS?

* Do they have low-level illusions?

* If yes, similar to ours?

* Quantify similarity

* General reason for the similarity?

  - Initialization
  - Nonlinearity
  - Training data
  - Task & specific architect

* Reason for differences
CONVOLUTIONAL NETS HAVE VISUAL ILLUSIONS
CONVOLUTIONAL NETS HAVE VISUAL ILLUSIONS
(look similar to ours but they are not the same)
CONVOLUTIONAL NETS HAVE VISUAL ILLUSIONS
(LOOK SIMILAR TO OURS BUT THEY ARE NOT THE SAME)

- EIGEN-ANALYSIS
- ARTIFICIAL PSYCHOPHYSICS
CONVOLUTIONAL NETS HAVE VISUAL ILLUSIONS

(LOOK SIMILAR TO OURS BUT THEY ARE NOT THE SAME)

- EIGEN-ANALYSIS

- ARTIFICIAL PSYCHOPHYSICS

JESÚS MALO

ALEX GÓMEZ
ADRIÁN MARTÍN
JAVIER VÁZQUEZ
MARCELO BERTALHIO

Universitat Pompeu Fabra
Barcelona
1. Illusions in humans & networks
   - 1.1 - Color & Brightness illusions
   - 1.2 - Learning low-level vision tasks
   - 1.3 - Results I: Artificial Physiology
   - 1.4 - Results II: Artificial Psychophysics

2. Results II: Artificial illusions from "artificial" CSFs
   - 2.1 - Linearization analysis
   - 2.2 - Jacobian, eigenvalues & eigenvectors
   - 2.3 - Artificial CSFs

3. Discussion & conclusions
<table>
<thead>
<tr>
<th>a) Dungeon illusion</th>
<th>b) Hong-Shevell rings</th>
<th>c) White illusion</th>
<th>d) Luminance grad.</th>
<th>e) Chevreul</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Dungeon Illusion" /></td>
<td><img src="image2" alt="Hong-Shevell Rings" /></td>
<td><img src="image3" alt="White Illusion" /></td>
<td><img src="image4" alt="Luminance Grad" /></td>
<td><img src="image5" alt="Chevreul" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>a) Dungeon illusion</th>
<th>b) Hong-Shevell rings</th>
<th>c) White illusion</th>
<th>d) Luminance grad.</th>
<th>e) Chevreul</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Dungeon Illusion" /></td>
<td><img src="image2" alt="Hong-Shevell Rings" /></td>
<td><img src="image3" alt="White Illusion" /></td>
<td><img src="image4" alt="Luminance Grad" /></td>
<td><img src="image5" alt="Chevreul" /></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>a) Dungeon illusion</td>
<td>b) Hong-Shevell rings</td>
<td>c) White illusion</td>
<td>d) Luminance grad.</td>
<td>e) Chevreul</td>
</tr>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
</tbody>
</table>

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Dungeon illusion</td>
<td>b) Hong-Shevell rings</td>
<td>c) White illusion</td>
<td>d) Luminance grad.</td>
<td>e) Chevreul</td>
</tr>
<tr>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Illusions in Convolutional Neural Networks

STIMULI FOR EXPERIMENTS

a) Dungeon illusion
b) Hong-Shevell rings
c) White illusion
d) Luminance grad.
e) Chevreul

Ware & Cowan Vis. Res. 82
ASYMMETRIC COLOR MATCHING

Modify till match

Inductor

Match

Test

Inductor
Illusions in Convolutional Neural Networks

CNNs LEARNING LOW-LEVEL VISION TASKS

Input $i$

Network (visual task)

Response $r = m(i, \theta)$

Supervised Learning

Visual Task & Loss Function

$L(\theta) = |i_o - m(i, \theta)|_2$

Learning

- Denoising
- Deblurring
- Restoration
Illusions in Convolutional Neural Networks

CNNs LEARNING LOW-LEVEL VISION TASKS

Supervised Learning

Visual Task & Loss Function

\[ L(\theta) = |i_o - m(i, \theta)|_2 \]

Learning
- Denoising
- Deblurring
- Restoration

Architectures
- Shallow
- Deep
- Very Deep

2-layers 5x5 kernels
4-layers 10x10 kernels
20-layers [zhang et al. CVPR 17][Tao CVPR 18]
Illusions in Convolutional Neural Networks

CNNs LEARNING LOW-LEVEL VISION TASKS

Input $i$

Network (visual task)

Response $r = m(i, \theta)$

Supervised Learning

Visual Task & Loss Function

$L(\theta) = |i_o - m(i, \theta)|_2$

Original Stimulus $i_o$

Learning

- Demoisling
- Deblurring
- Restoration

Architectures

- Shallow
- Deep
- Very Deep

Data Sets

- Russowsky IJCV 15
- Vazquez et al. Percept. 09
- Malo et al. Neur. Comp. 12

Architectures

- 2-layers 5x5 Kernels
- 4-layers 10x10 Kernels
- 20-layers [Zhang et al. CVPR 17][Tao CVPR 18]
Artificial Physiology: Measure responses
Artificial PHYSIOLOGY: Measure responses

Artificial PSYCHOPHYSICS: Modify input to achieve match in response
Illusions in Convolutional Neural Networks

Results Artificial Physiology

Shallow

Deep

Very Deep

DN-NET

DB-NET

RestoreNET

Zhang et al.
**Illusions in Convolutional Neural Networks**

### SHALLOW

<table>
<thead>
<tr>
<th>Model</th>
<th>Dungeon</th>
<th>Hong-Shevell</th>
<th>White</th>
<th>Gradient</th>
<th>Chevreul</th>
</tr>
</thead>
<tbody>
<tr>
<td>DN-NET</td>
<td>In</td>
<td>Out-L, Out-R</td>
<td></td>
<td>In</td>
<td>Out-L, Out-R</td>
</tr>
<tr>
<td>R</td>
<td>0.58</td>
<td>0.77, 0.27</td>
<td>0.58</td>
<td>0.39, 0.72</td>
<td>1</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>0.38, 0.76</td>
<td>1</td>
<td>0.74, 0.53</td>
<td>0.5</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0.47, 0.51</td>
<td>0</td>
<td>0.44, 0.37</td>
<td>0</td>
</tr>
</tbody>
</table>

### DB-NET

<table>
<thead>
<tr>
<th>Model</th>
<th>Dungeon</th>
<th>Hong-Shevell</th>
<th>White</th>
<th>Gradient</th>
<th>Chevreul</th>
</tr>
</thead>
<tbody>
<tr>
<td>In</td>
<td>Out-L, Out-R</td>
<td>In</td>
<td>Out-L, Out-R</td>
<td>In</td>
<td>Out-L, Out-R</td>
</tr>
<tr>
<td>R</td>
<td>0.58</td>
<td>0.68, 0.22</td>
<td>0.58</td>
<td>0.21, 0.44</td>
<td>1</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>0.36, 0.75</td>
<td>1</td>
<td>0.62, 0.39</td>
<td>0.5</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0.34, 0.44</td>
<td>0</td>
<td>0.25, 0.21</td>
<td>0</td>
</tr>
</tbody>
</table>

### RestoreNET

<table>
<thead>
<tr>
<th>Model</th>
<th>Dungeon</th>
<th>Hong-Shevell</th>
<th>White</th>
<th>Gradient</th>
<th>Chevreul</th>
</tr>
</thead>
<tbody>
<tr>
<td>In</td>
<td>Out-L, Out-R</td>
<td>In</td>
<td>Out-L, Out-R</td>
<td>In</td>
<td>Out-L, Out-R</td>
</tr>
<tr>
<td>R</td>
<td>0.58</td>
<td>0.77, 0.24</td>
<td>0.58</td>
<td>0.3, 0.67</td>
<td>1</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>0.36, 0.76</td>
<td>1</td>
<td>0.64, 0.41</td>
<td>0.5</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0.43, 0.53</td>
<td>0</td>
<td>0.38, 0.34</td>
<td>0</td>
</tr>
</tbody>
</table>

### DEEP

**Results**

<table>
<thead>
<tr>
<th>Model</th>
<th>Dungeon</th>
<th>Hong-Shevell</th>
<th>White</th>
<th>Gradient</th>
<th>Chevreul</th>
</tr>
</thead>
<tbody>
<tr>
<td>In</td>
<td>Out-L, Out-R</td>
<td>In</td>
<td>Out-L, Out-R</td>
<td>In</td>
<td>Out-L, Out-R</td>
</tr>
<tr>
<td>R</td>
<td>0.58</td>
<td>0.77, 0.27</td>
<td>0.58</td>
<td>0.42, 0.63</td>
<td>1</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>0.38, 0.72</td>
<td>1</td>
<td>0.57, 0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0.5, 0.57</td>
<td>0</td>
<td>0.56, 0.49</td>
<td>0</td>
</tr>
</tbody>
</table>

### Zhang et al.

<table>
<thead>
<tr>
<th>Model</th>
<th>Dungeon</th>
<th>Hong-Shevell</th>
<th>White</th>
<th>Gradient</th>
<th>Chevreul</th>
</tr>
</thead>
<tbody>
<tr>
<td>In</td>
<td>Out-L, Out-R</td>
<td>In</td>
<td>Out-L, Out-R</td>
<td>In</td>
<td>Out-L, Out-R</td>
</tr>
<tr>
<td>R</td>
<td>0.58</td>
<td>0.74, 0.28</td>
<td>0.58</td>
<td>0.31, 0.58</td>
<td>1</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>0.39, 0.7</td>
<td>1</td>
<td>0.5, 0.42</td>
<td>0.5</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0.49, 0.56</td>
<td>0</td>
<td>0.49, 0.53</td>
<td>0</td>
</tr>
</tbody>
</table>
VARIANCE NOT AN ISSUE

a) Dongpan Mask
b) Hopf-Twist Rings
c) White Illusion
d) Luminance grad.
e) Chevreul

[Graphs and charts showing data for RestoreNET and Deep RestoreNET with various comparisons between Darker and Lighter conditions.]

V  V  V  V  V
ReLU vs Sigmoid NOT AN ISSUE
Relu vs Sigmoid: Not an Issue
Illusions in Convolutional Neural Networks

RESULTS ARTIFICIAL PSYCHOPHYSICS

SHALLOW

DN-NET

DB-NET

RestoreNET

DEEP

Deep DN-NET

Deep DB-NET

Deep RestoreNET

HUMANS

Human observers

Zhang

VERY DEEP
VARIANCE NOT AN ISSUE
* CNNs do have visual illusions

* Visual illusions of CNNs are similar but not like ours
  - Artificial physiol. ⇒ shifts in resp. u ok 75%
  - Artificial psychophys. ⇒ Assimilation vs contrast

* Complexity of architecture is an issue
* CNNs do have visual illusions

* Visual illusions of CNNs are similar but not like ours
  - Artificial physiol. $\Rightarrow$ shifts in resp. $\sim$ 75%
  - Artificial psychophys. $\Rightarrow$ Assimilation vs contrast

* Complexity of architecture is an issue
RESULTS II: Artificial illusions from "artificial" CSFs

2.1 - Linearization analysis

2.2 - Jacobian, eigenvalues & eigenvectors

2.3 - Artificial CSFs
RESULTS II: Artificial illusions from "artificial" CSIs

\[ r = m(i, \theta) \]

- Input: \( i \)
- Response: \( r \)

\( m(i, \theta) \)
RESULTS II: Artificial illusions from "artificial" CSIs

\[ r = m(i, \theta) \]

\[ r = m(0 + i, \theta) \approx m(0, \theta) + \nabla_i m(0) \cdot i \]

\[ r = \nabla_i m(0) \cdot i \]

Jacobian at 0
RESULTS II: Artificial illusions from "artificial" CSFs

1 - Linearization analysis

\[ r = m(i, \theta) \]

input \[ \rightarrow \] response

\[ r = m(0 + i, \theta) \approx m(0, \theta) + \nabla_i m(0) \cdot i \]

\[ r = \nabla_i m(0) \cdot i \]

- Analytic: Martinez, Halo et al., PLOS 18
- Autograd
- Linear regression \( \nabla_i m(0) = R \cdot J^t \)

\[ R = \begin{bmatrix} r^{(1)} r^{(2)} \cdots r^{(m)} \end{bmatrix} \]

\[ I = \begin{bmatrix} i^{(1)} i^{(2)} \cdots i^{(n)} \end{bmatrix} \]
RESULTS II: Artificial illusions from "artificial" CSs

1. Linearization analysis

Example 1

Original STIMULUS
Input STIMULUS
Degrad. = 33 %
Actual RESPONSE
Degrad. = 20 %
Approximated RESPONSE
Degrad. = 18 % Fract. Resp. = 91 %

Example 2

STIMULUS
RESPONSE
Linearized RESPONSE
Fract. Resp. = 83 %
RESULTS II: Artificial illusions from "artificial" CSFs

Jacobian

Impulse Response of $G_{out}$ channel on $R_{in}$
Impulse Response of $G_{out}$ channel on $G_{in}$
Impulse Response of $G_{out}$ channel on $B_{in}$

$D_{in} =$

ON-OFF-like cells!
RESULTS II: Artificial illusions from "artificial" CSFs

Jacobian

Impulse Response of $G_{out}$ channel on $R_{in}$

Impulse Response of $G_{out}$ channel on $G_{in}$

Impulse Response of $G_{out}$ channel on $B_{in}$

ON-OFF-like cells!

$D_{in} = \{R_{in}, G_{in}, B_{in}\}$

Eigen vectors
($\sim$ Fourier & opponents)

Eigen values
($\sim$ low-pass)

Color eigen-directions
$RG - YB$!
Figure 14: Accumulated spectra of eigenfunctions decomposed in their intrinsic color space and weighted by eigenvalues. Limited frequency resolution is due to the fact that this result comes from small $16 \times 16$ image blocks. This may give rise to artifacts in the spectra.
RESULTS II: Artificial illusions from "artificial" CSFs

ARTIFICIAL SPATIAL FILTERS

Denoise

Deblur

Restore

ARTIFICIAL SPECTRAL SENSITIV.

HUMAN SPECTRAL SENSITIV.

HUMAN CONTRAST SENSITIV. (CSFs)

Achromatic CSF

Chromatic CSF (RG)

Chromatic CSF (YB)
2) RESULTS II: Artificial illusions from "artificial" CSFs

ARTIFICIAL SPATIAL FILTERS

ARTIFICIAL SPECTRAL SENSITIV.

HUMAN SPECTRAL SENSITIV.

HUMAN CONTRAST SENSITIV. (CSFs)

ACROMATIC CSF

CHROMATIC CSF (RG)

CHROMATIC CSF (YB)


**3 Discussion & Conclusions**

- CNNs for low-level vision develop a number of human-like features
  - On-off cells
  - Opponent color channels RG, YB
  - Human-like spatial filters (CSFs) in achromatic RG, YB

---

A. Gómez, J. Malo, et al. (2019) "Visual Illusions also deceive Convolutional Neural Networks: Analysis and Implications".
**Discussion & Conclusions**

- CNNs for low-level vision develop a number of human-like features:
  - On-off cells
  - Opponent color channels RG, YB
  - Human-like spatial filters (CSFs) in Achrom. RG, YB

* Why? = error minimization goal (Wiener restoration) ⇒ fit to spectrum of natural images

---

**Discussion & Conclusions**

* CNNs for low-level vision develop a number of human-like features
  - On-off cells
  - Opponent color channels RG, YB
  - Human-like spatial filters (CSFs) in achrom. RG, YB

* Why? = error-minimization goal (Wiener restoration) $\Rightarrow$ fit to spectrum of natural images

* $(\text{Generality/illusions})$ vs $(\text{Specificity/no-illusions})$
  - simple architect.
  - complex architect.

---

* \((\text{Generality/illusions}) \text{ vs } (\text{Specificity/no-illusions})\) (simple architect.) vs (complex architect.)

Table 3: Nonlinearity, Performance & Illusion Strength (Denoising)

<table>
<thead>
<tr>
<th></th>
<th>Shallow</th>
<th>Deep</th>
<th>Zhang et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fract. Lin. Resp.</td>
<td>90%</td>
<td>93%</td>
<td>84%</td>
</tr>
<tr>
<td>Error NonLin.</td>
<td>11.2%</td>
<td>10.8%</td>
<td>7.5%</td>
</tr>
<tr>
<td>Error Linear</td>
<td>12.5%</td>
<td>12.1%</td>
<td>15.9%</td>
</tr>
<tr>
<td>Illusion Strength</td>
<td>++</td>
<td>+++</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4: Nonlinearity, Performance & Illusion Strength (Deblurring)

<table>
<thead>
<tr>
<th></th>
<th>Shallow</th>
<th>Deep</th>
<th>Tao et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fract. Lin. Resp.</td>
<td>88%</td>
<td>93%</td>
<td>94%</td>
</tr>
<tr>
<td>Error NonLin.</td>
<td>17.1%</td>
<td>16.3%</td>
<td>15.0%</td>
</tr>
<tr>
<td>Error Linear</td>
<td>19.0%</td>
<td>17.1%</td>
<td>15.6%</td>
</tr>
<tr>
<td>Illusion Strength</td>
<td>++</td>
<td>+++</td>
<td>++</td>
</tr>
</tbody>
</table>

RestoreNet:
+ Cleaner functions tuned to regular patterns
+ Right and strong chromatic directions

Zhang:
- more noisy functions
- no achromatic channel
- weaker contrast in color directions
  (achrom. channel mixed with chrom. channels)

Tao:
- wrong functions
  (more blobs than oriented patterns)
- wrong chromatic direct.
  (and really weak: virtually no color sensit.)
A deep learning framework for neuroscience

Blake A. Richards¹,²,³,⁴,⁴²*, Timothy P. Lillicrap⁵,⁶,⁴², Philippe Beaudoin⁷, Yoshua Bengio¹,⁴,⁸,*

Systems neuroscience seeks explanations for how the brain implements a wide variety of perceptual, cognitive and motor tasks. Conversely, artificial intelligence attempts to design computational systems based on the tasks they will have to solve. In artificial neural networks, the three components specified by design are the objective functions, the learning rules and the architectures. With the growing success of deep learning, which utilizes brain-inspired architectures, these designed components have increasingly become central to how we model, engineer and optimize complex artificial learning systems. Here we argue that a greater focus on these components would also benefit systems neuroscience. We give examples of how this optimization-based framework can drive theoretical and experimental progress in neuroscience. We contend that this principled perspective on systems neuroscience will help to generate more rapid progress.
Focus on the objective functions! (the "why" question)

**A deep learning framework for neuroscience**

Blake A. Richards, Timothy P. Lillicrap, Philippe Beaudoin, Yoshua Bengio

Systems neuroscience seeks explanations for how the brain implements a wide variety of perceptual, cognitive and motor tasks. Conversely, artificial intelligence attempts to design computational systems based on the tasks they will have to solve. In artificial neural networks, the three components specified by design are the objective functions, the learning rules and the architecture of the networks. With the growing success of deep learning, which utilizes brain-inspired architectures, these three designed components have increasingly become central to how we model, engineer and optimize complex artificial learning systems. Here we argue that a greater focus on these components would also benefit systems neuroscience. We give examples of how this optimization-based framework can drive theoretical and experimental progress in neuroscience. We contend that this principled perspective on systems neuroscience will help to generate more rapid progress.
Focus on the objective functions! (the "why" question)

A deep learning framework for neuroscience
Blake A. Richards1,2,3,4,*, Timothy P. Lillicrap5,6,7, Philippe Beaudoin1, Yoshua Bengio4,4,1,3

Systems neuroscience seeks explanations for how the brain implements a wide variety of perceptual, cognitive, and motor tasks. Conversely, artificial intelligence attempts to design computational systems based on the tasks they will have to solve. In artificial neural networks, the three components specified by design are the objective functions, the learning rules, and the architectures. With the growing success of deep learning, which utilizes biology-inspired architectures, these three designed components have increasingly become central to how we model, engineer, and optimize complex artificial learning systems. Here we argue that a greater focus on these components would also benefit systems neuroscience. We give examples of how this optimization-based framework can drive theoretical and experimental progress in neuroscience. We contend that this principled perspective on systems neurosciences will help to generate more rapid progress.

* CNN brightness & color illusions seem to come from error min. (whenever the net is general enough)
  
  * Illusions in humans may come from error min too!)
  
  Attrk & Li 1992
  
  * Caution with blind OVER-fitting Martinez & Molo 2019
  
  Laperre & Molo 15
QUESTIONS?

COMMENTS?