CONVOLUTIONAL NETS HAVE VISUAL ILLUSIONS

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

- EIGEN-ANALYSIS
- ARTIFICIAL PSYCHOPHYSICS

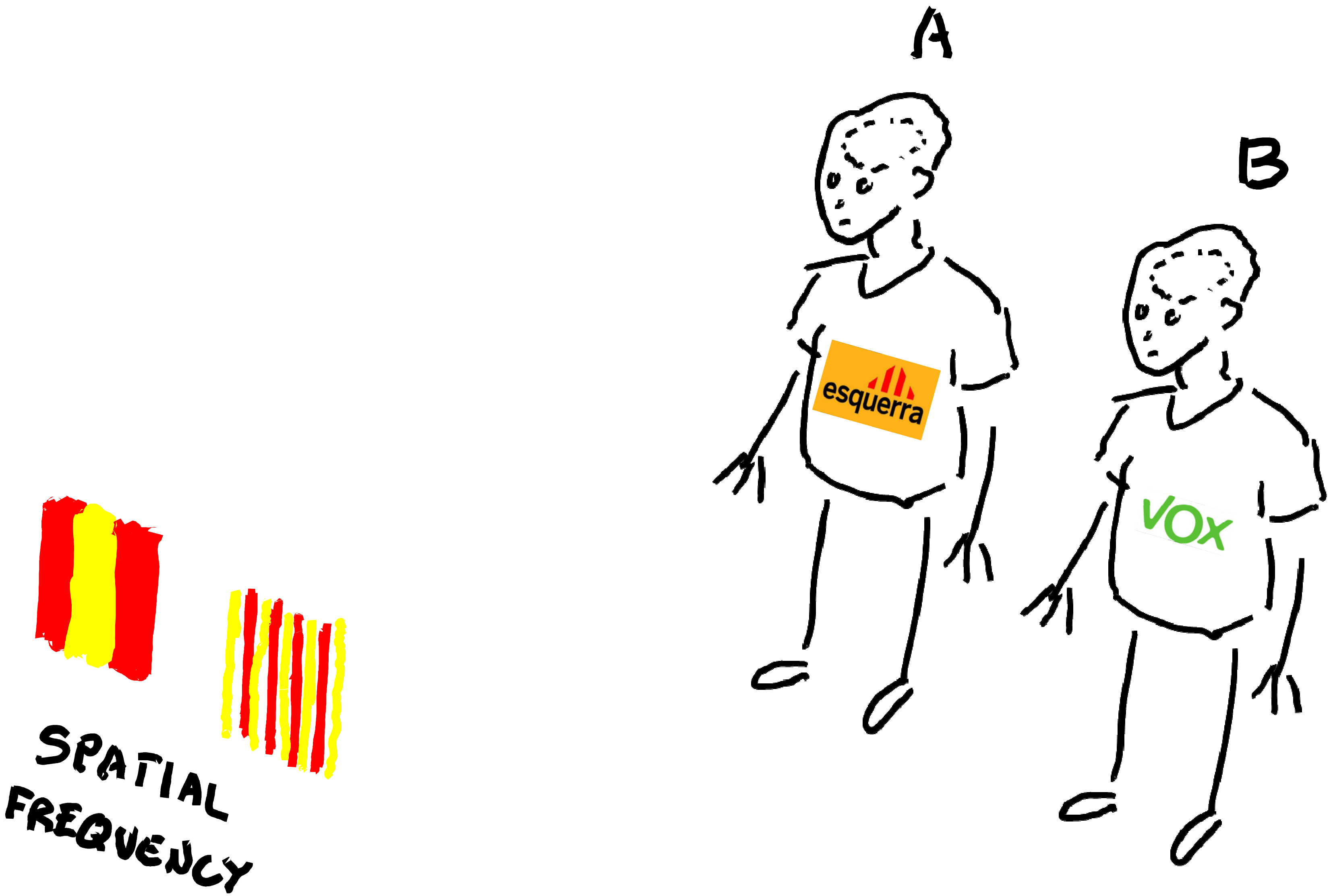
JESÚS MALO

UNIVERSITAT DE VALÈNCIA

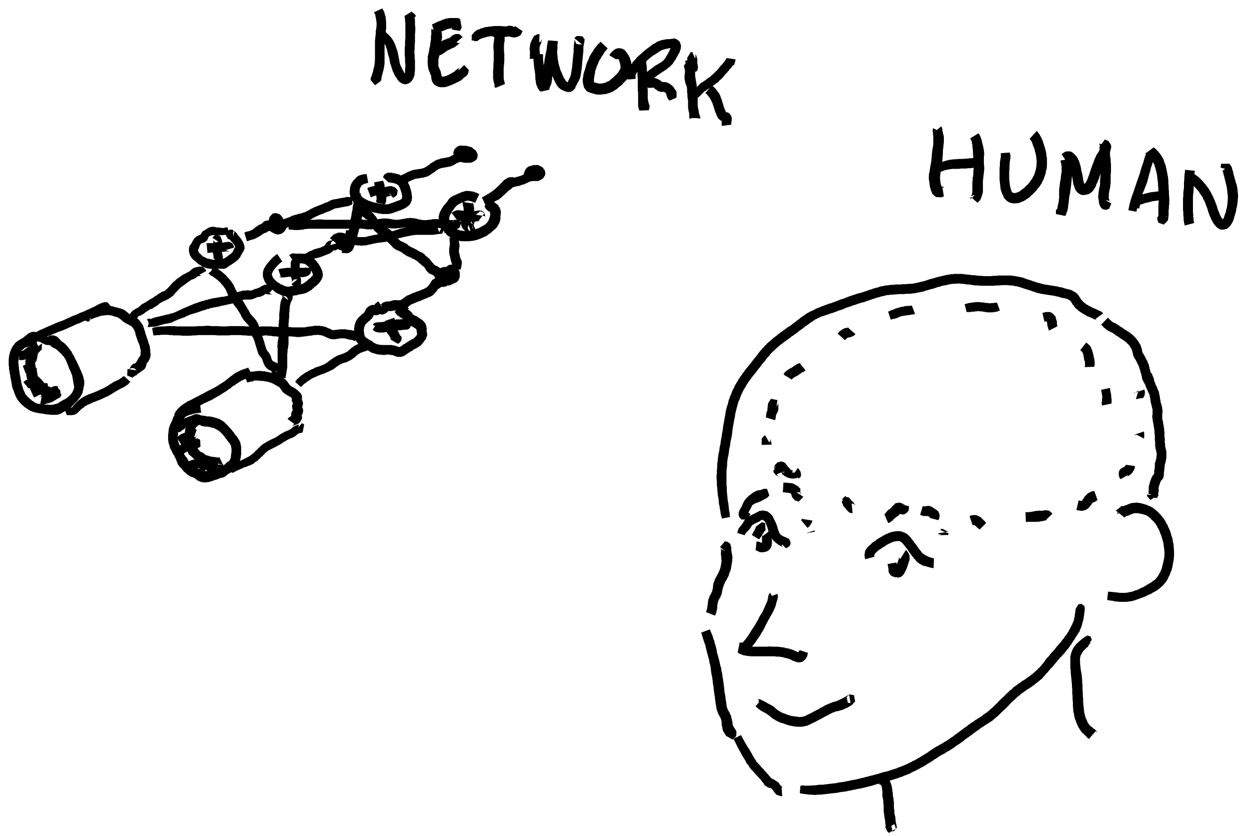
ALEX GOMEZ
ADRIAN MARTIN
JAVIER VAZQUEZ
MARCELO BERTALMIO

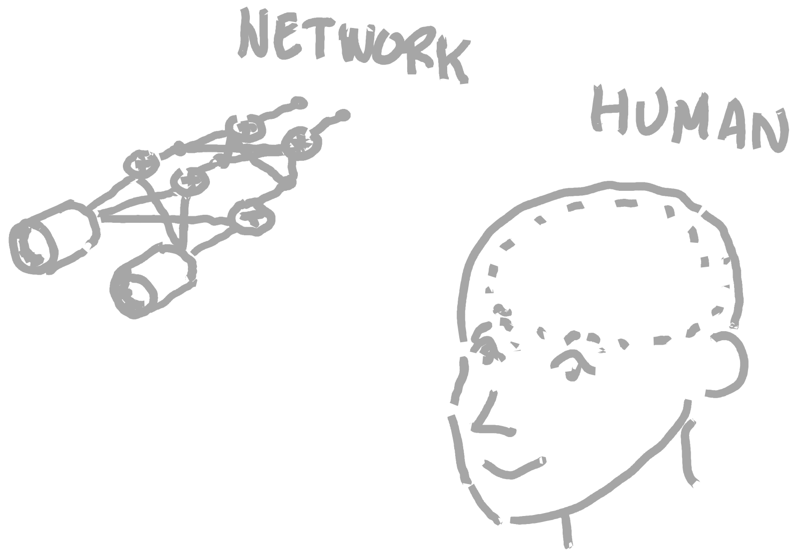
Universitat
Pompeu Fabra
Barcelona

SEEING IS A SUBJECTIVE EXPERIENCE !

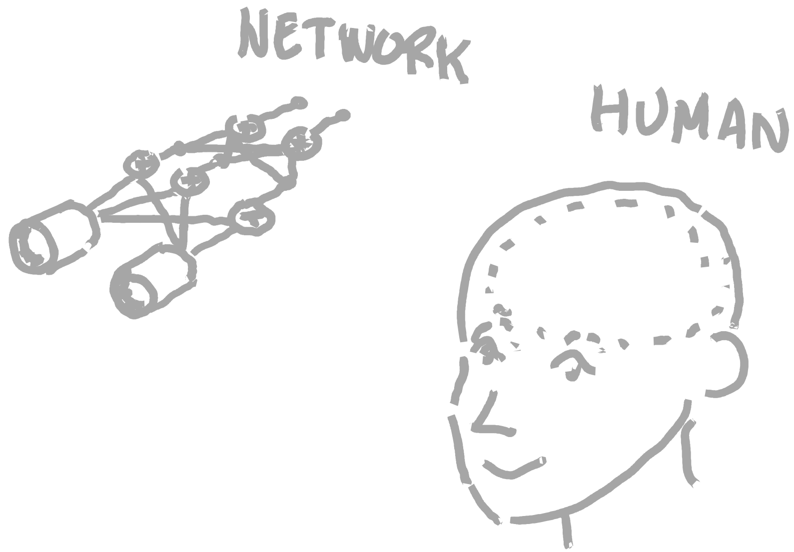


SEEING IS A SUBJECTIVE EXPERIENCE !





DO WE SEE THE SAME ?



DO WE SEE THE SAME ?

- Deep learning → Neuroscience Goal functions
- Neuroscience → Deep learning Psychophysics

VISUAL ILLUSIONS ?

VISUAL ILLUSIONS ?

* Do they have low-level illusions?
* If yes, similar to ours?
* Quantify similarity
* General reason for the similarity?
* Reason for differences
  - Initialization
  - Nonlinearity
  - Training data
  - Task & specific architect.

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

VNIVERSITAT DE VALÈNCIA

ALEX GOMEZ
ADRIAN MARTIN
JAVIER VAZQUEZ
MARCELO BERTALMIO

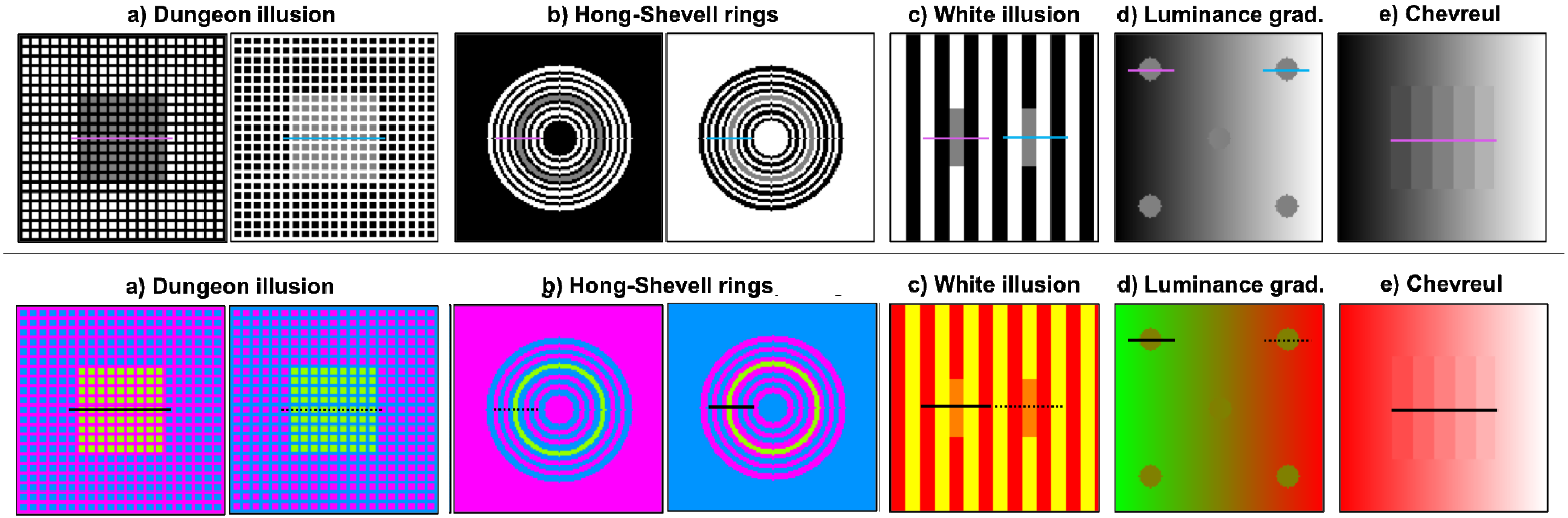
Universitat Pompeu Fabra Barcelona

1. Illusions in humans & networks
   - 1.1 - Color & Brightness illusions
   - 1.2 - Learning low-level vision tasks
   - 1.3 - Results I
     - Artificial Physiology
     - Artificial Psychophysics

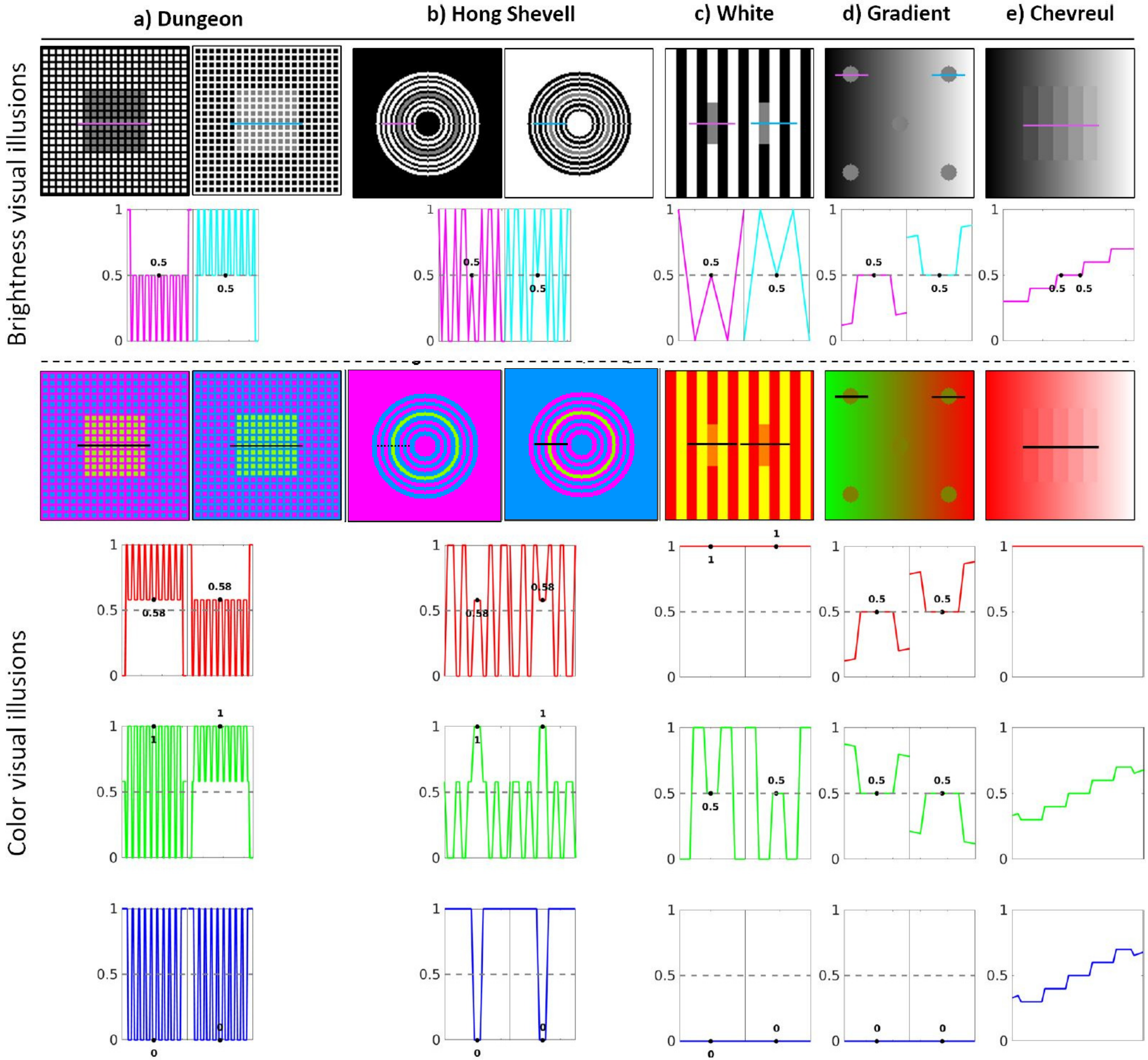
2. RESULTS II: Artificial illusions from "artificial" CSFs
   - 2.1 - Linearization analysis
   - 2.2 - Jacobian, eigenvalues & eigenvectors
   - 2.3 - Artificial CSFs

3. Discussion & conclusions

1 Illusions in Convolutional Neural Networks STIMULI FOR EXPERIMENTS

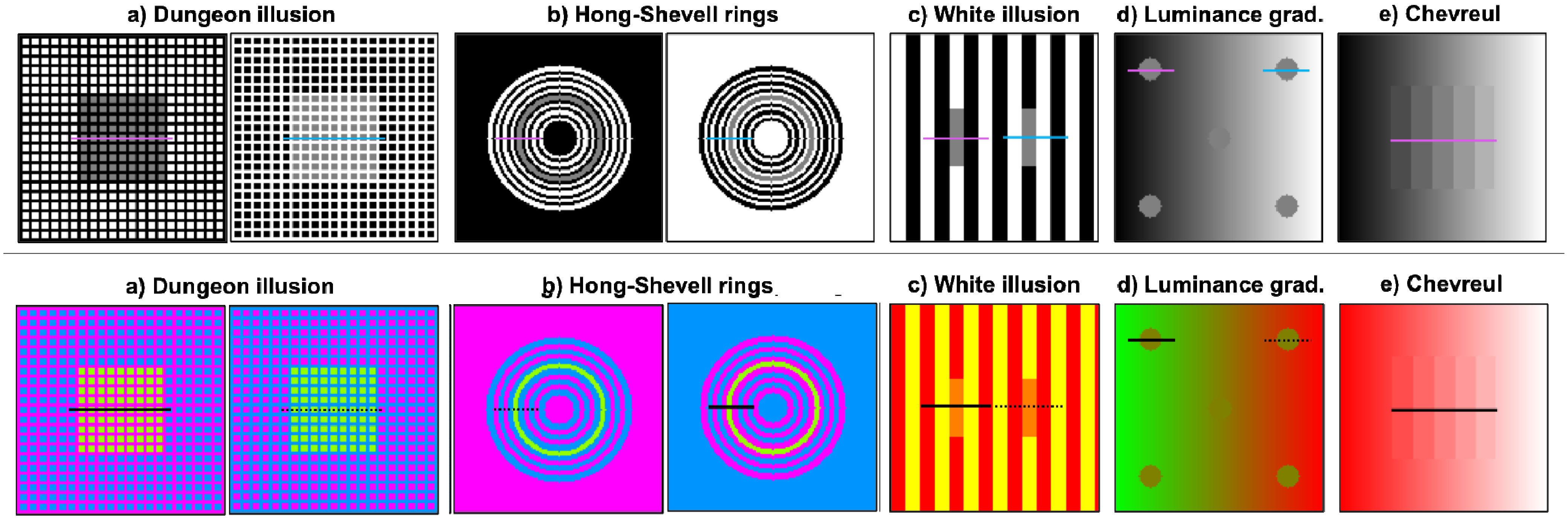


1

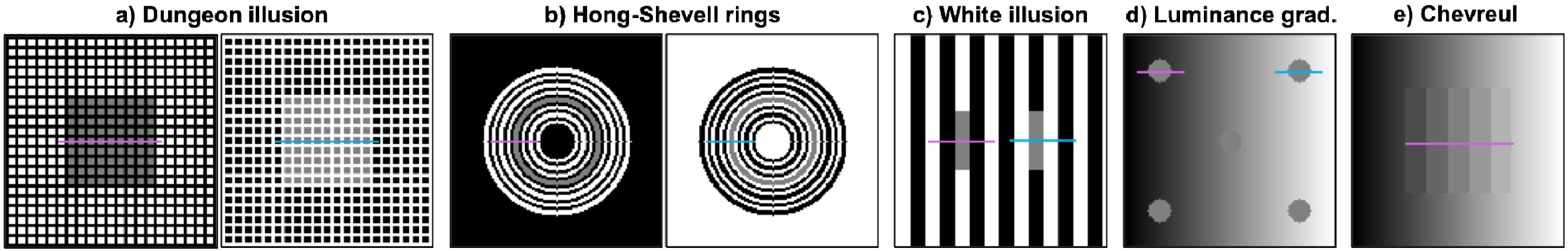


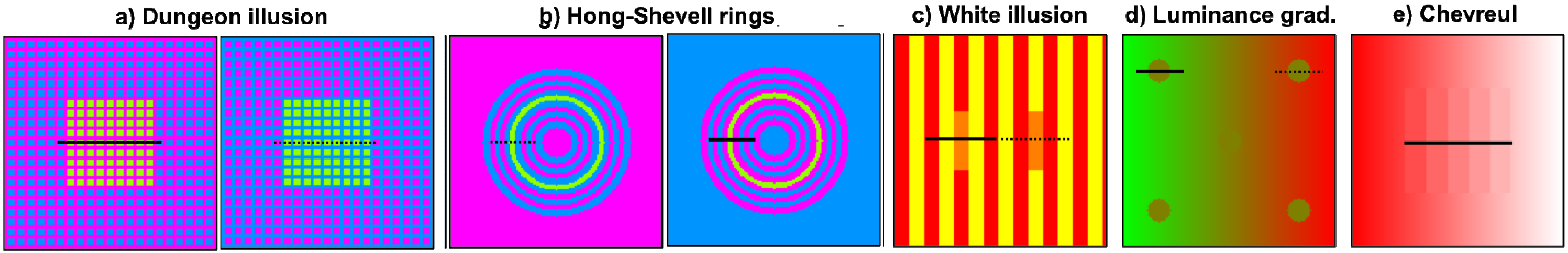
a) Dungeon
b) Hong Shevell
c) White
d) Gradient
e) Chevreul
Brightness visual illusions
Color visual illusions

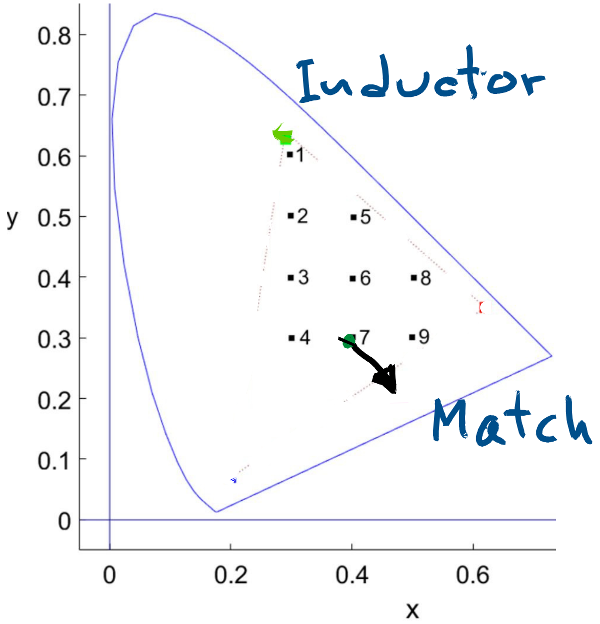
1 Illusions in Convolutional Neural Networks   STIMULI FOR EXPERIMENTS



1 Illusions in Convolutional Neural Networks — STIMULI FOR EXPERIMENTS

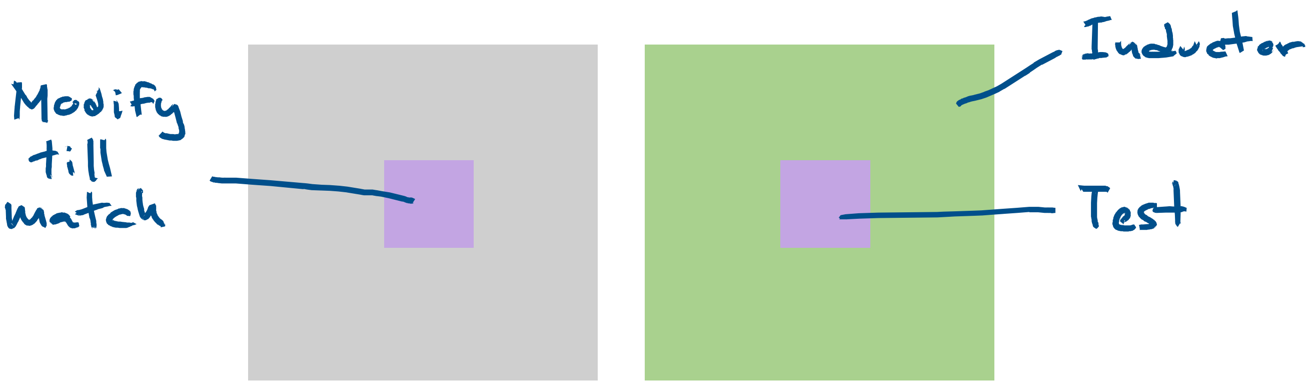






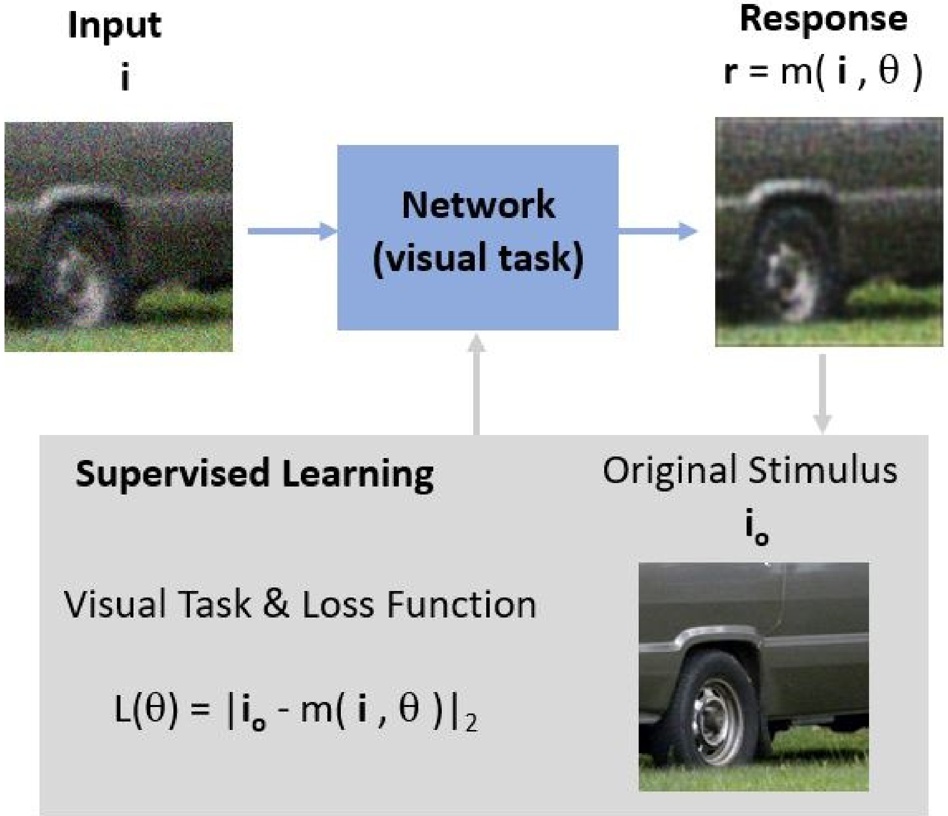
Ware & Cowan Vis. Res. 82

ASYMMETRIC COLOR MATCHING



① Illusions in Convolutional Neural Networks

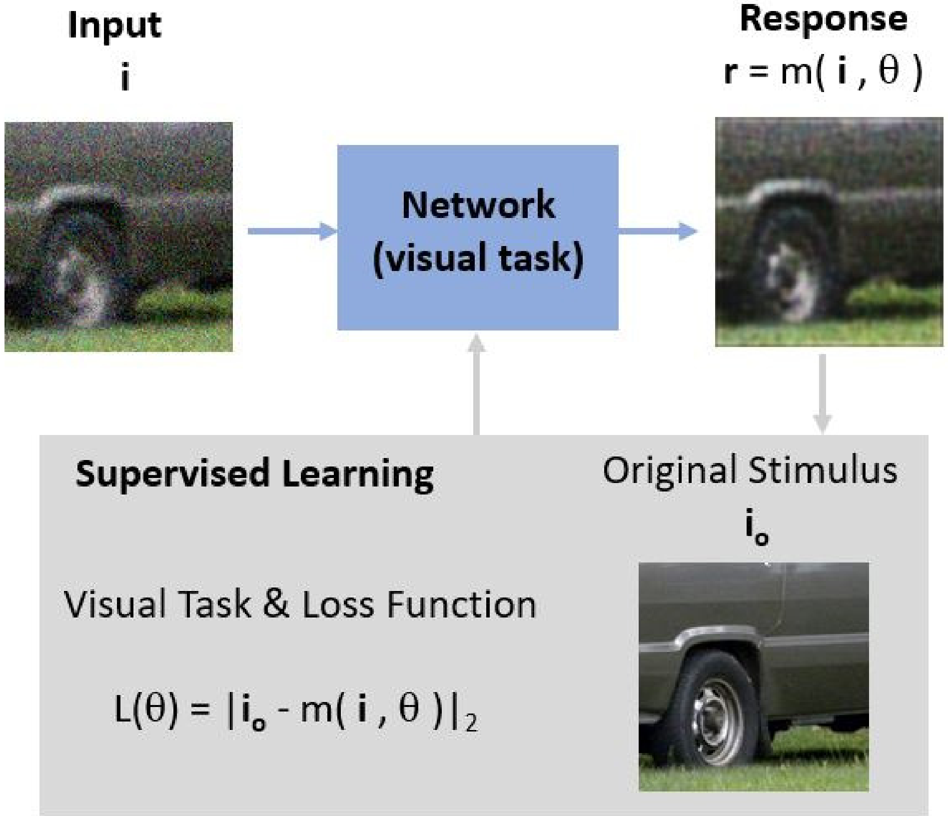
CNNs LEARNING LOW-LEVEL VISION TASKS



Learning
- Denoising
- Deblurring
- Restoration

1 Illusions in Convolutional Neural Networks

CNNs LEARNING LOW-LEVEL VISION TASKS

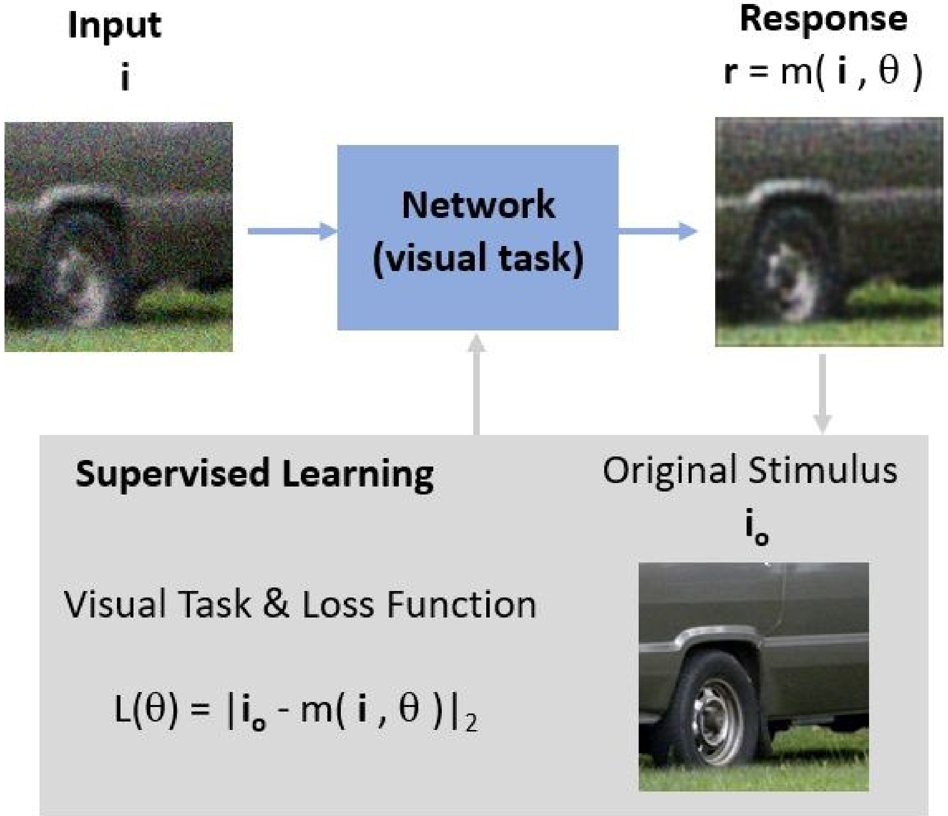


Learning
- Denoising
- Deblurring
- Restoration

Architectures
- Shallow — 2-layers 5x5 kernels
- Deep — 4-layers 10x10 kernels
- Very Deep — 20-layers [Zhang et al. CVPR 17][Tao CVPR 18]

1 Illusions in Convolutional Neural Networks

CNNs LEARNING LOW-LEVEL VISION TASKS



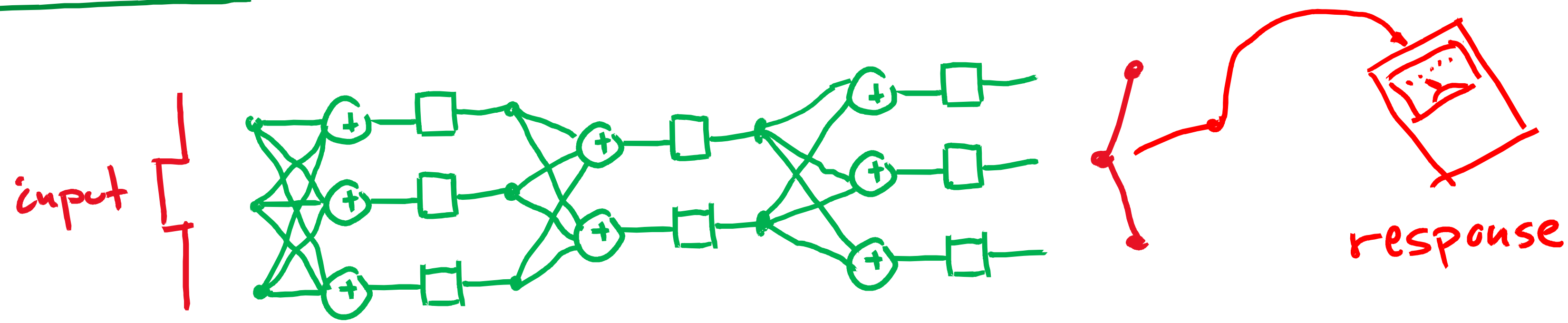
Learning
- Denoising
- Deblurring
- Restoration

Data sets
- Russakovsky IJCV 15
- Vazquez et al. Percept. 09
- Malo et al. Neur. Comp. 12

Architectures
- Shallow — 2-layers 5x5 kernels
- Deep — 4-layers 10x10 kernels
- Very Deep — 20-layers [Zhang et al. CVPR 17] [Tao CVPR 18]

1 Illusions in Convolutional Neural Networks ARTIFICIAL PHYSIOLOGY VS PSYCHOPHYSICS

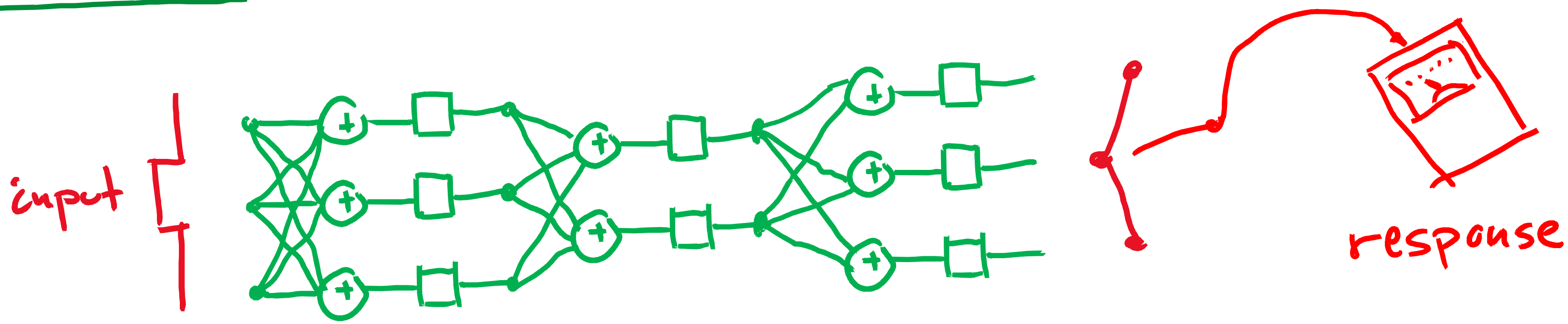
Artificial PHYSIOLOGY: Measure responses



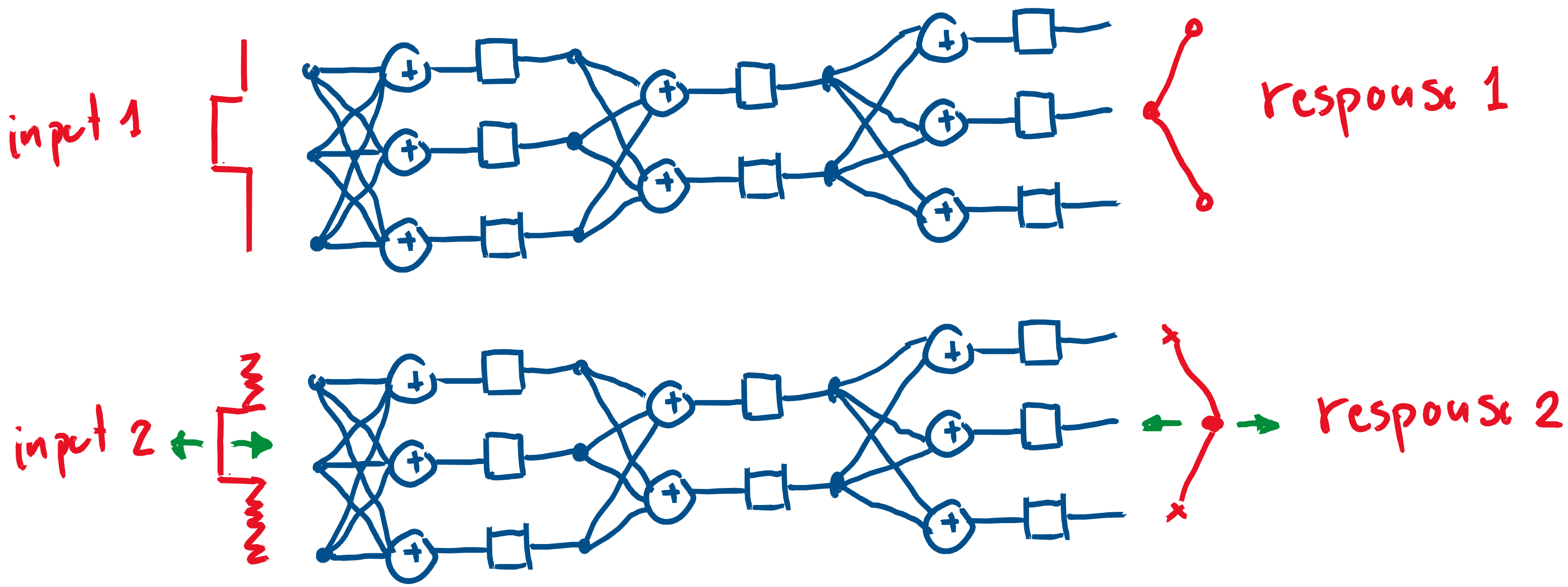
1 Illusions in Convolutional Neural Networks

ARTIFICIAL PHYSIOLOGY VS PSYCHOPHYSICS

Artificial PHYSIOLOGY: Measure responses

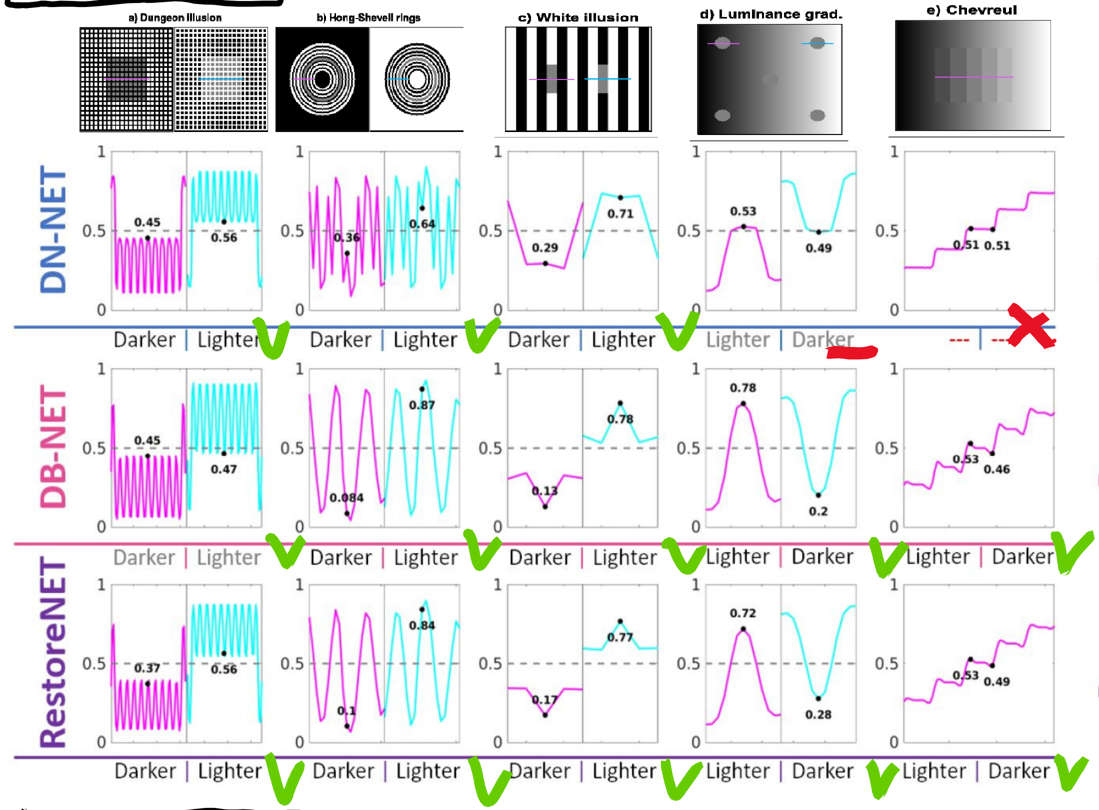


Artificial PSYCHOPHYSICS: Modify input to achieve match in response

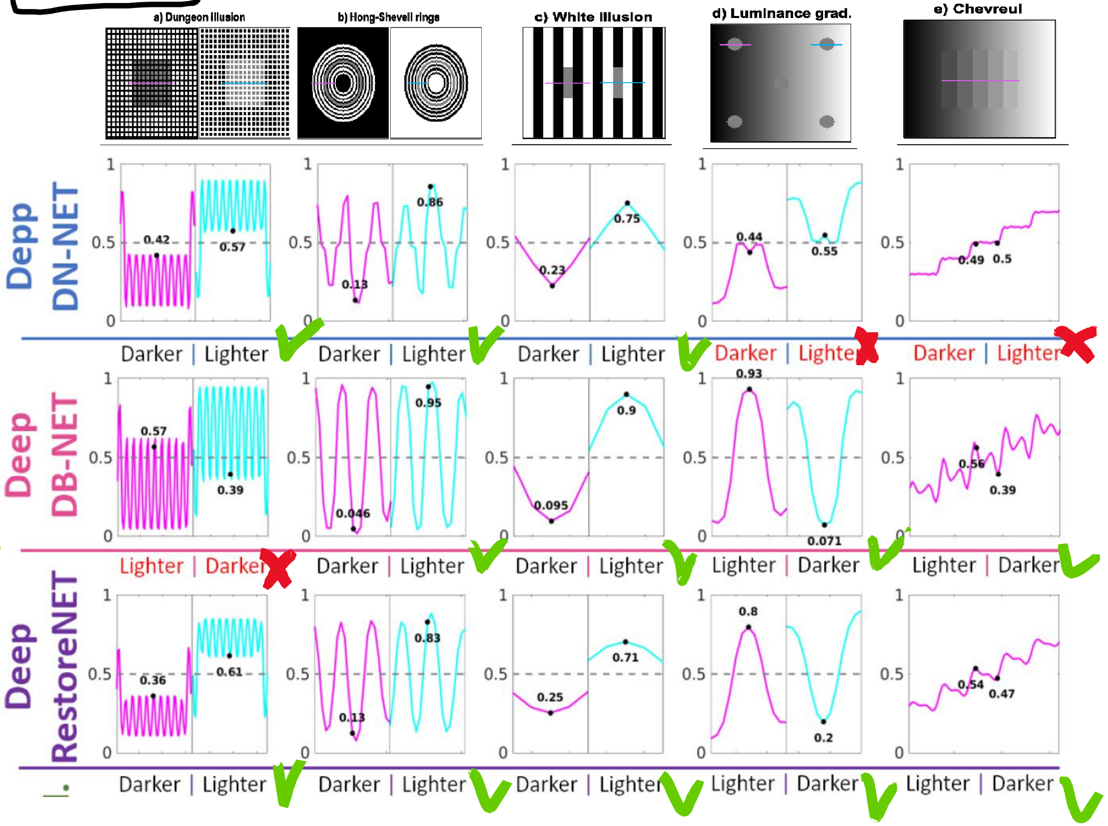


1 Illusions in Convolutional Neural Networks — Results ARTIFICIAL PHYSIOLOGY

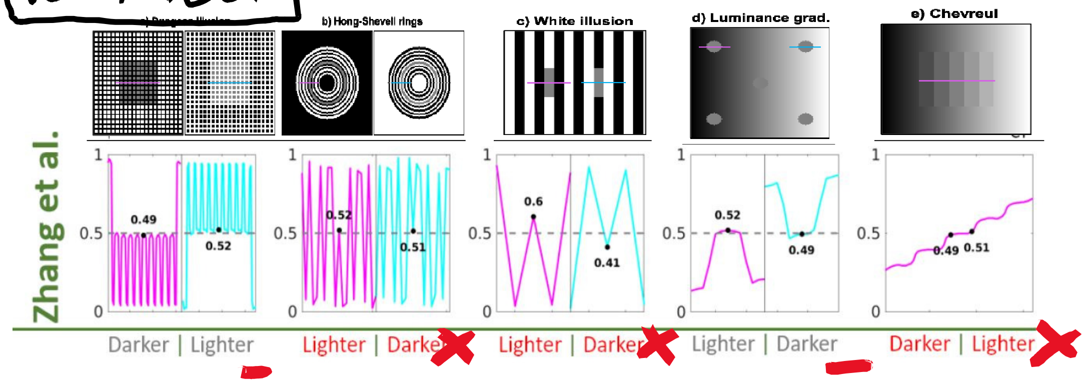
SHALLOW

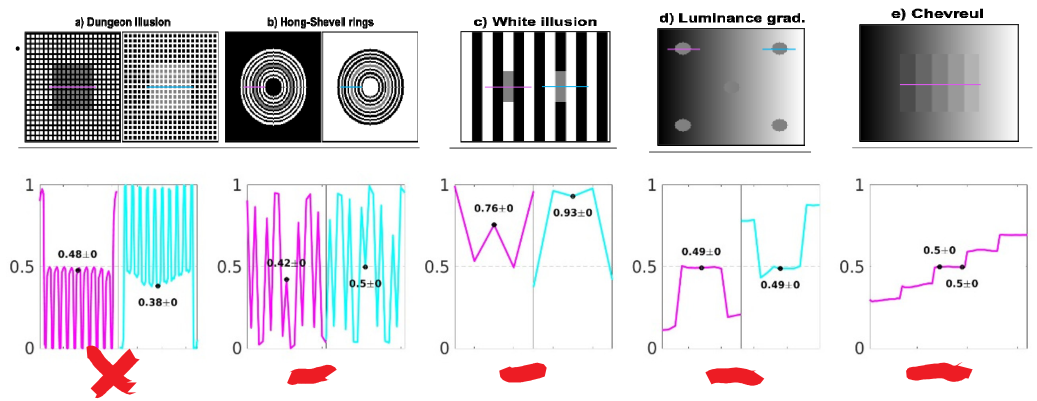


DEEP



VERY DEEP

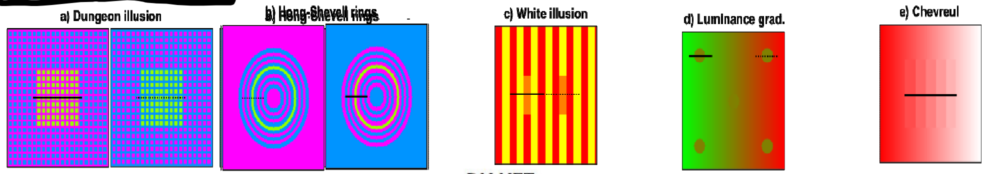




1 Illusions in Convolutional Neural Networks

Results ARTIFICIAL PHYSIOLOGY

SHALLOW



DN-NET

	Dungeon			Hong-Shevell			White			Gradient			Chevreul		
	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R
R	0.58	0.77	0.27	0.58	0.39	0.72	1	0.85	0.92	0.5	0.43	0.61	1	0.92	0.91
G	1	0.38	0.76	1	0.74	0.53	0.5	0.64	0.36	0.5	0.63	0.36	0.5	0.5	0.51
B	0	0.47	0.51	0	0.44	0.37	0	0.074	0.13	0	0.079	0.079	0.5	0.49	0.51

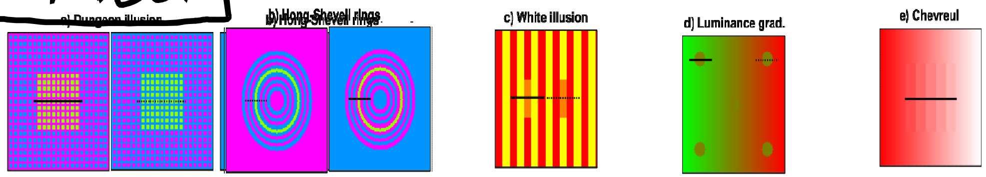
DB-NET

	Dungeon			Hong-Shevell			White			Gradient			Chevreul		
	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R
R	0.58	0.68	0.22	0.58	0.21	0.44	1	0.81	0.94	0.5	0.37	0.64	1	0.92	0.9
G	1	0.36	0.75	1	0.62	0.39	0.5	0.55	0.39	0.5	0.6	0.38	0.5	0.47	0.45
B	0	0.34	0.44	0	0.25	0.21	0	0.058	0.12	0	0.078	0.075	0.5	0.48	0.46

RestoreNET

	Dungeon			Hong-Shevell			White			Gradient			Chevreul		
	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R
R	0.58	0.77	0.24	0.58	0.3	0.67	1	0.85	0.93	0.5	0.36	0.64	1	0.92	0.9
G	1	0.36	0.76	1	0.64	0.41	0.5	0.55	0.39	0.5	0.6	0.37	0.5	0.49	0.48
B	0	0.43	0.53	0	0.38	0.34	0	0.057	0.11	0	0.073	0.075	0.5	0.48	0.47

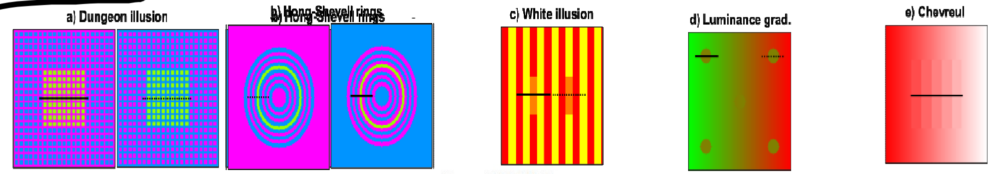
VERY DEEP



Zhang et al.

	Dungeon			Hong-Shevell			White			Gradient			Chevreul		
	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R
R	0.58	0.59	0.56	0.58	0.58	0.59	1	1	1	0.5	0.52	0.49	1	1	1
G	1	0.98	0.98	1	0.99	0.98	0.5	0.5	0.49	0.5	0.49	0.51	0.5	0.48	0.51
B	0	0.027	0.027	0	0.02	0.02	0	0.012	0.012	0	0.012	0.012	0.5	0.48	0.51

DEEP



Deep DN-NET

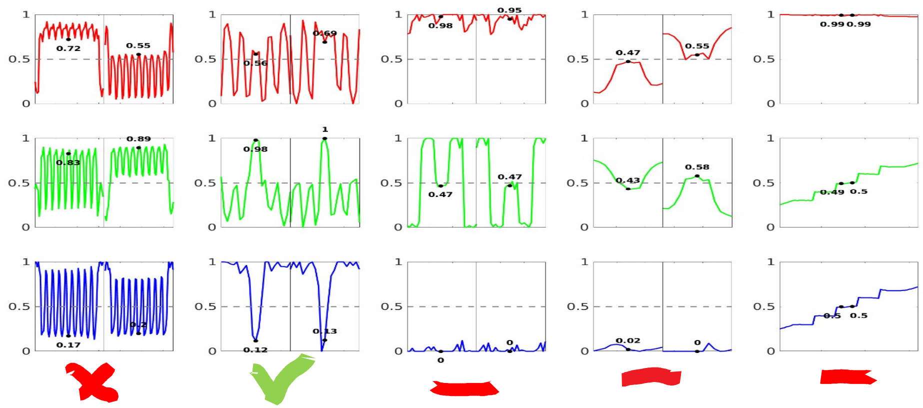
	Dungeon			Hong-Shevell			White			Gradient			Chevreul		
	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R
R	0.58	0.77	0.27	0.58	0.42	0.63	1	0.94	0.97	0.5	0.47	0.49	1	0.96	0.95
G	1	0.38	0.72	1	0.57	0.4	0.5	0.61	0.38	0.5	0.54	0.48	0.5	0.49	0.49
B	0	0.5	0.57	0	0.56	0.49	0	0.028	0.056	0	0.027	0.03	0.5	0.52	0.52

Deep DB-NET

	Dungeon			Hong-Shevell			White			Gradient			Chevreul		
	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R
R	0.58	0.67	0.29	0.58	0.36	0.45	1	0.89	0.97	0.5	0.5	0.45	1	0.95	0.92
G	1	0.27	0.72	1	0.51	0.25	0.5	0.48	0.47	0.5	0.56	0.41	0.5	0.52	0.45
B	0	0.41	0.57	0	0.52	0.36	0	0.024	0.094	0	0.068	0.052	0.5	0.53	0.46

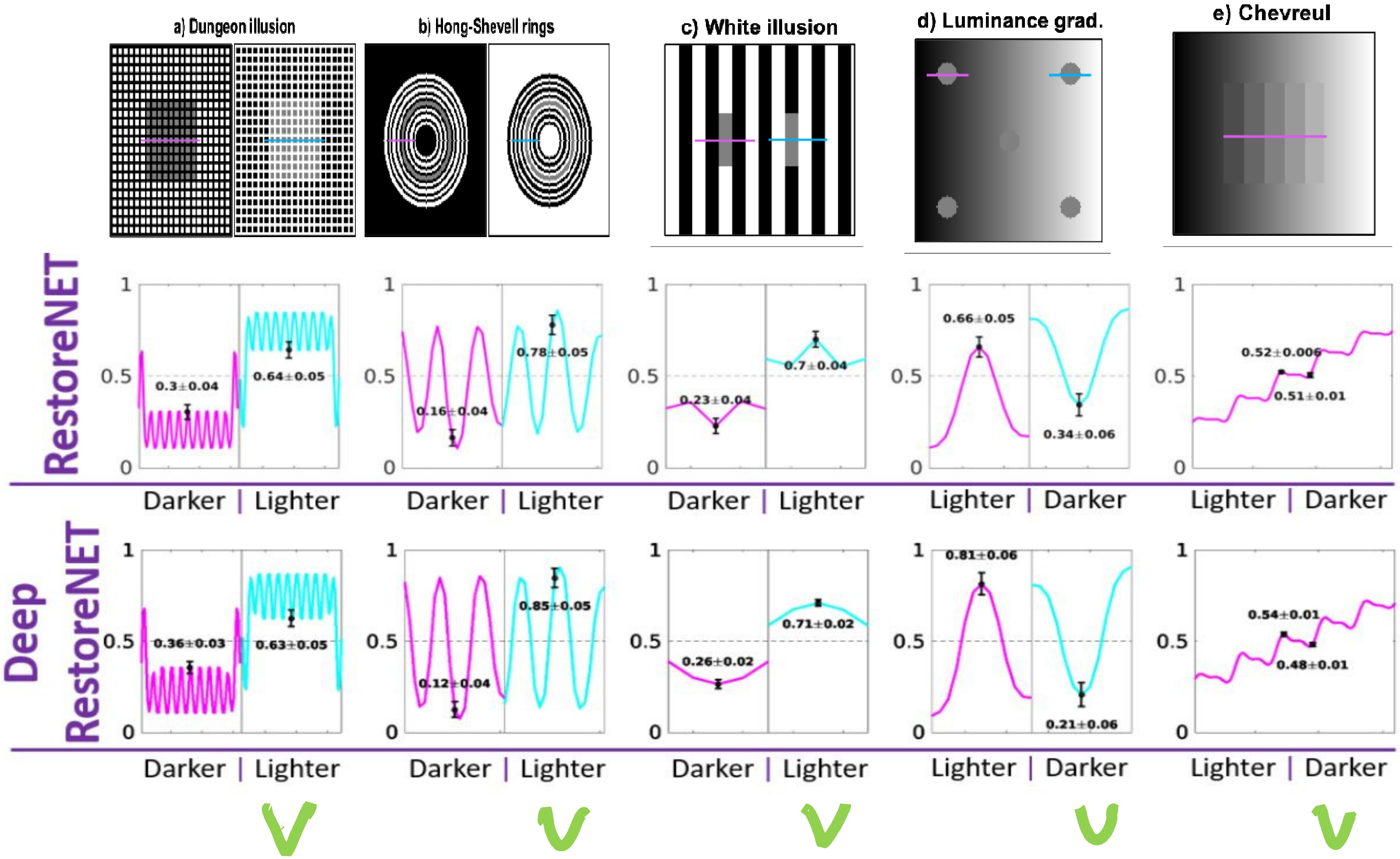
Deep RestoreNET

	Dungeon			Hong-Shevell			White			Gradient			Chevreul		
	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R	In	Out-L	Out-R
R	0.58	0.74	0.28	0.58	0.31	0.58	1	0.93	0.97	0.5	0.48	0.55	1	0.93	0.91
G	1	0.39	0.7	1	0.5	0.42	0.5	0.51	0.39	0.5	0.48	0.46	0.5	0.53	0.5
B	0	0.49	0.56	0	0.49	0.53	0	0.036	0.075	0	0.074	0.073	0.5	0.5	0.48



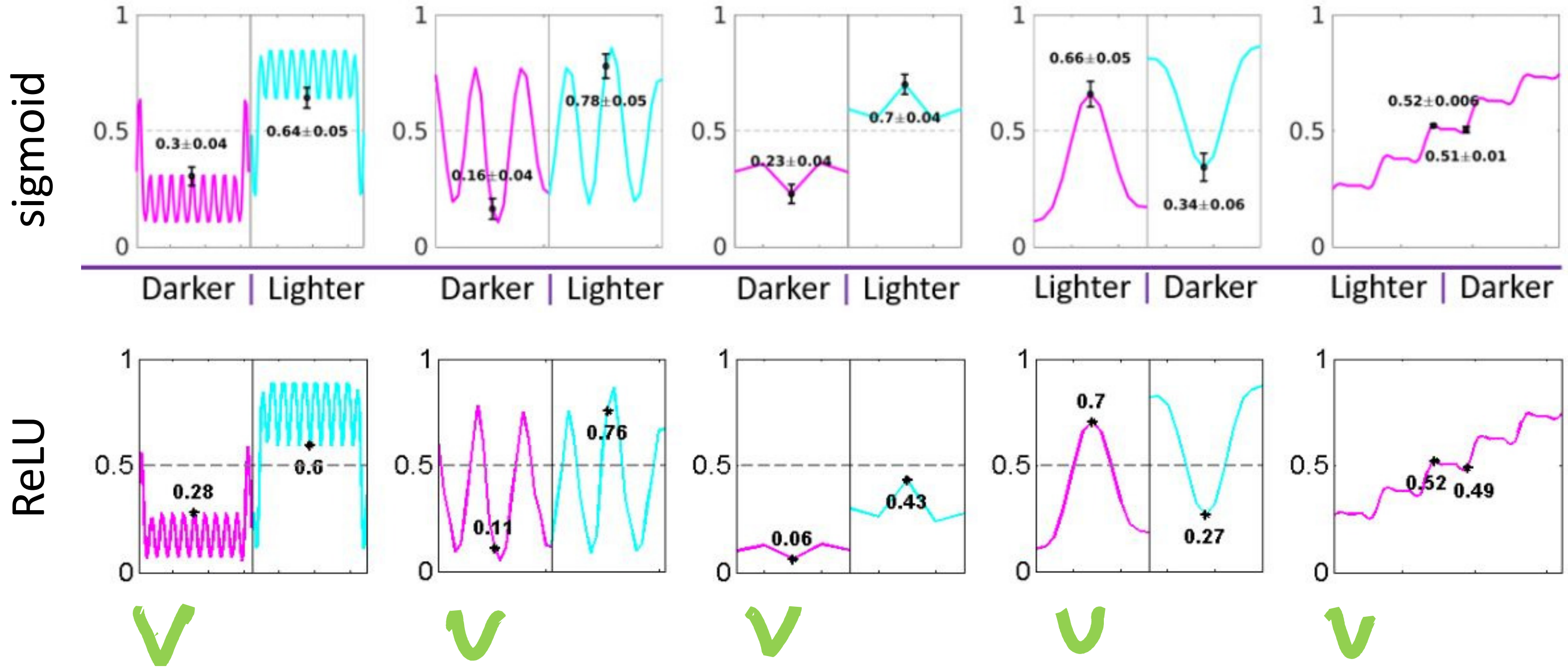
1

VARIANCE NOT AN ISSUE



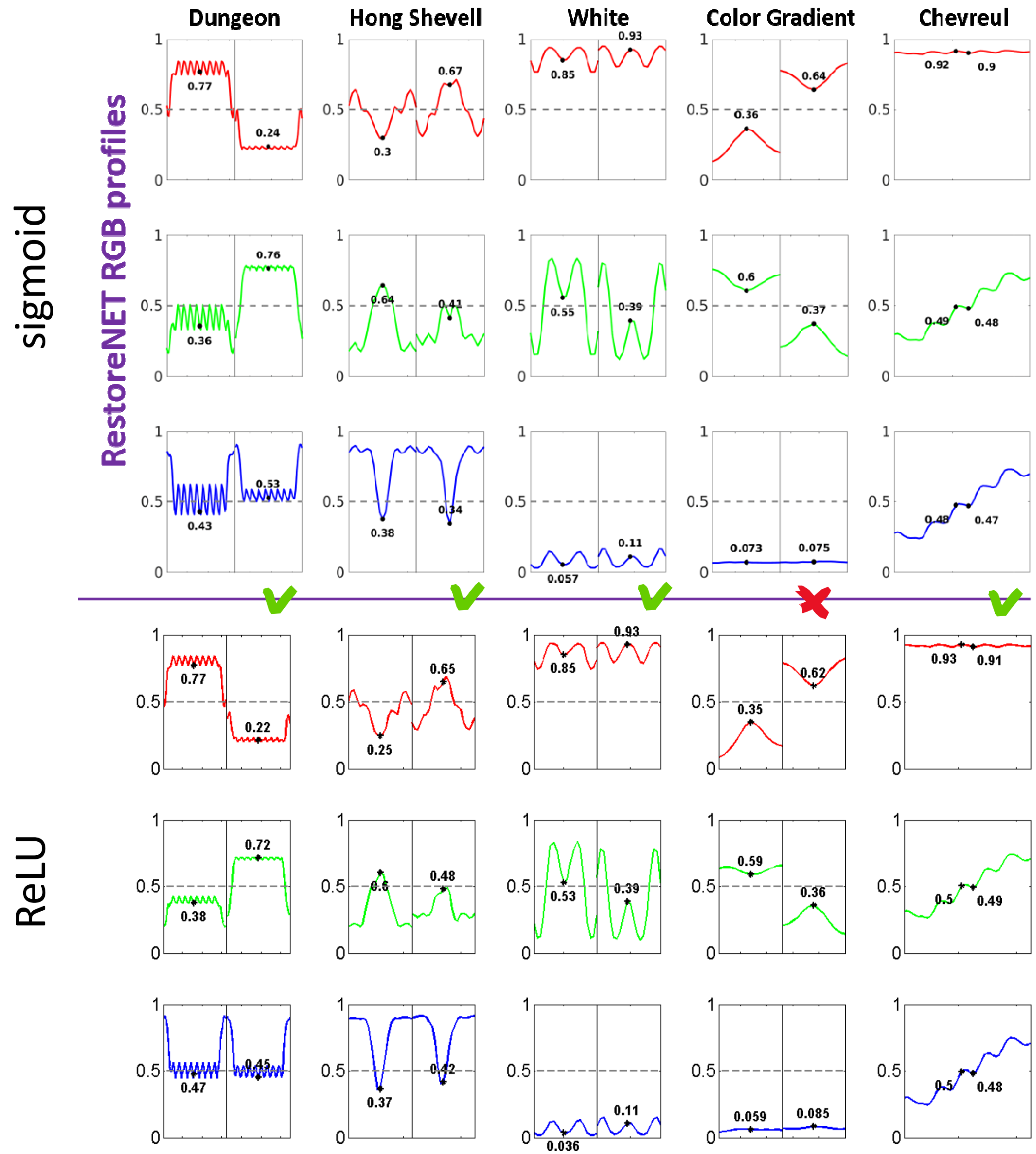
①

Relu vs Sigmoid NOT AN ISSUE



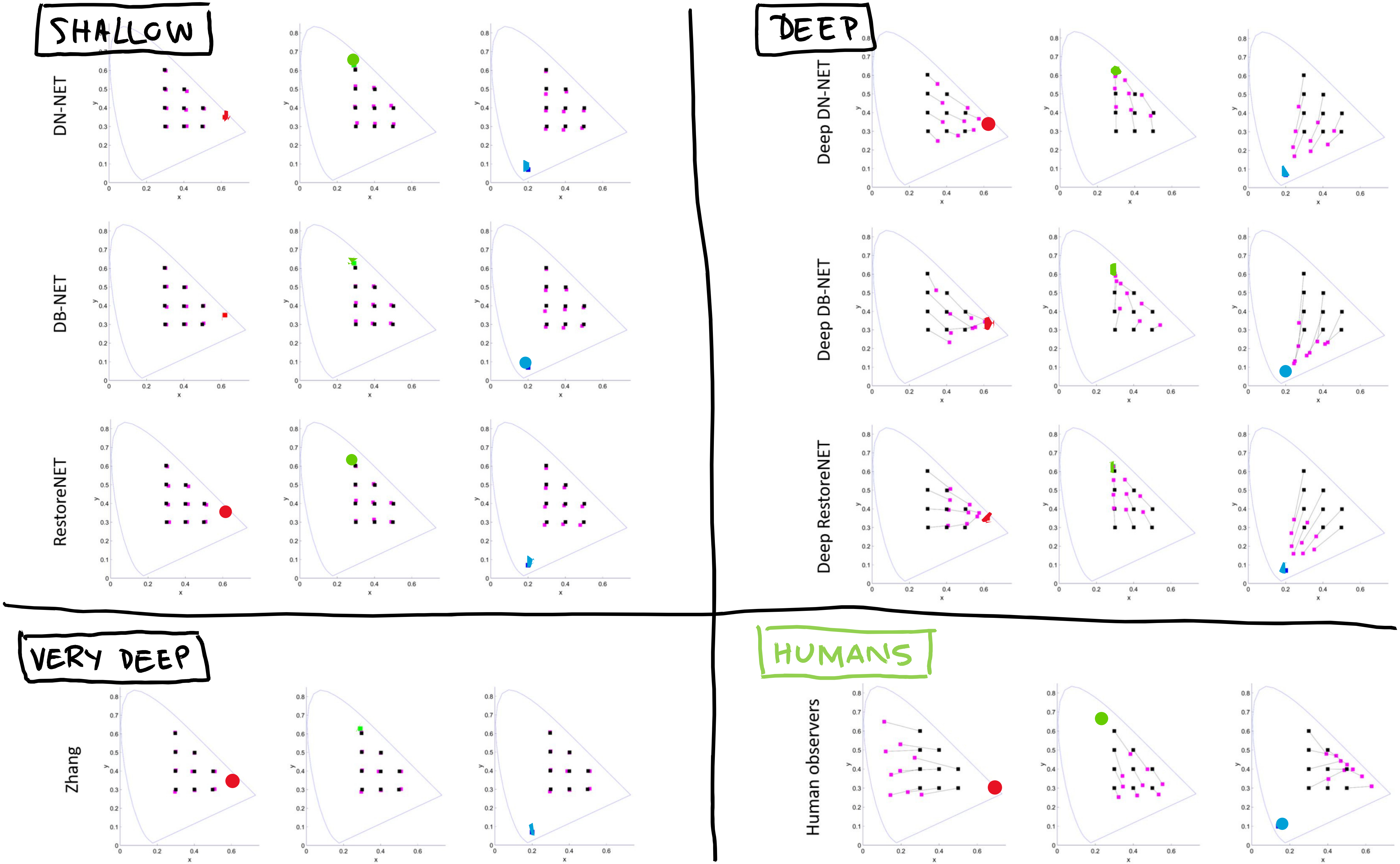
1

Relu vs Sigmoid NOT AN ISSUE



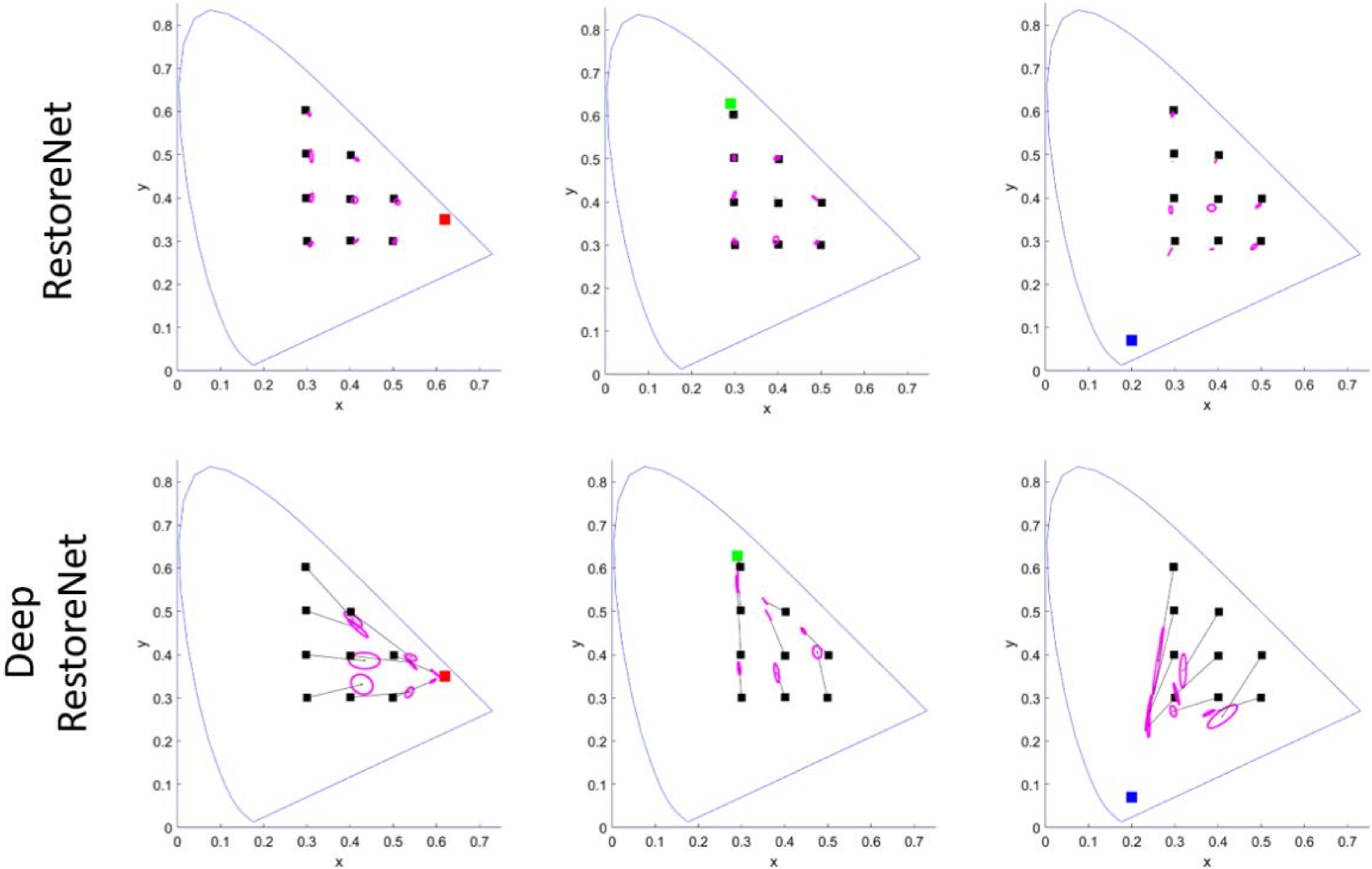
1 Illusions in Convolutional Neural Networks

Results ARTIFICIAL PSYCHOPHYSICS



1

VARIANCE NOT AN ISSUE



1 Illusions in Convolutional Neural Networks

SUMMARY { ARTIFICIAL PHYSIOLOGY
ARTIFICIAL PSYCHOPHYSICS

* CNNs do have visual illusions

* Visual illusions of CNNs are similar but not like ours

  - Artificial physiol. ⇒ shifts in resp. ~ ok 75%
  - Artificial psychophys. ⇒ Assimilation vs contrast

* Complexity of architecture is an issue

1 Illusions in Convolutional Neural Networks

SUMMARY { ARTIFICIAL PHYSIOLOGY
ARTIFICIAL PSYCHOPHYSICS

* CNNs do have visual illusions

* Visual illusions of CNNs are similar but not like ours

  - Artificial physiol. ⇒ shifts in resp. ~ ok 75%
  - Artificial psychophys. ⇒ Assimilation vs contrast

* Complexity of architecture is an issue

WHY?

② RESULTS II: Artificial illusions from "artificial" CSFs

2.1 - Linearization analysis

2.2 - Jacobian, eigenvalues & eigenvectors

2.3 - Artificial CSFs

② RESULTS II: Artificial illusions from "artificial" CSFs .1 - Linearization analysis

$$i \xrightarrow{m(i,\theta)} r$$

input response

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

2 RESULTS II: Artificial illusions from "artificial" CSFs .1 - Linearization analysis

$$\underset{\text{input}}{i} \xrightarrow{m(i,\theta)} \underset{\text{response}}{r}$$

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

$$r = m(0 + i, \theta) \simeq \cancelto{0}{m(0, \theta)} + \nabla_i m(0) \cdot i$$

$$r = \underbrace{\nabla_i m(0)}_{\text{Jacobian at } 0} \cdot i$$

2 RESULTS II: Artificial illusions from "artificial" CSFs .1 - Linearization analysis

$$\underset{\text{input}}{i} \xrightarrow{m(i,\theta)} \underset{\text{response}}{r}$$

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

$$r = m(0+i,\theta) \simeq \cancelto{0}{m(0,\theta)} + \nabla_i m(0) \cdot i$$

$$r = \underbrace{\nabla_i m(0)}_{\text{Jacobian at } 0} \cdot i$$

- Analytic Martinez, Malo et al. PLOS 18
- Autograd
- Linear regression $\nabla_i m(0) = R \cdot I^{+}$

$R = [\, r^{(1)} \; r^{(2)} \; \ldots \; r^{(n)} \,]$ $I = [\, i^{(1)} \; i^{(2)} \; \ldots \; i^{(n)} \,]$

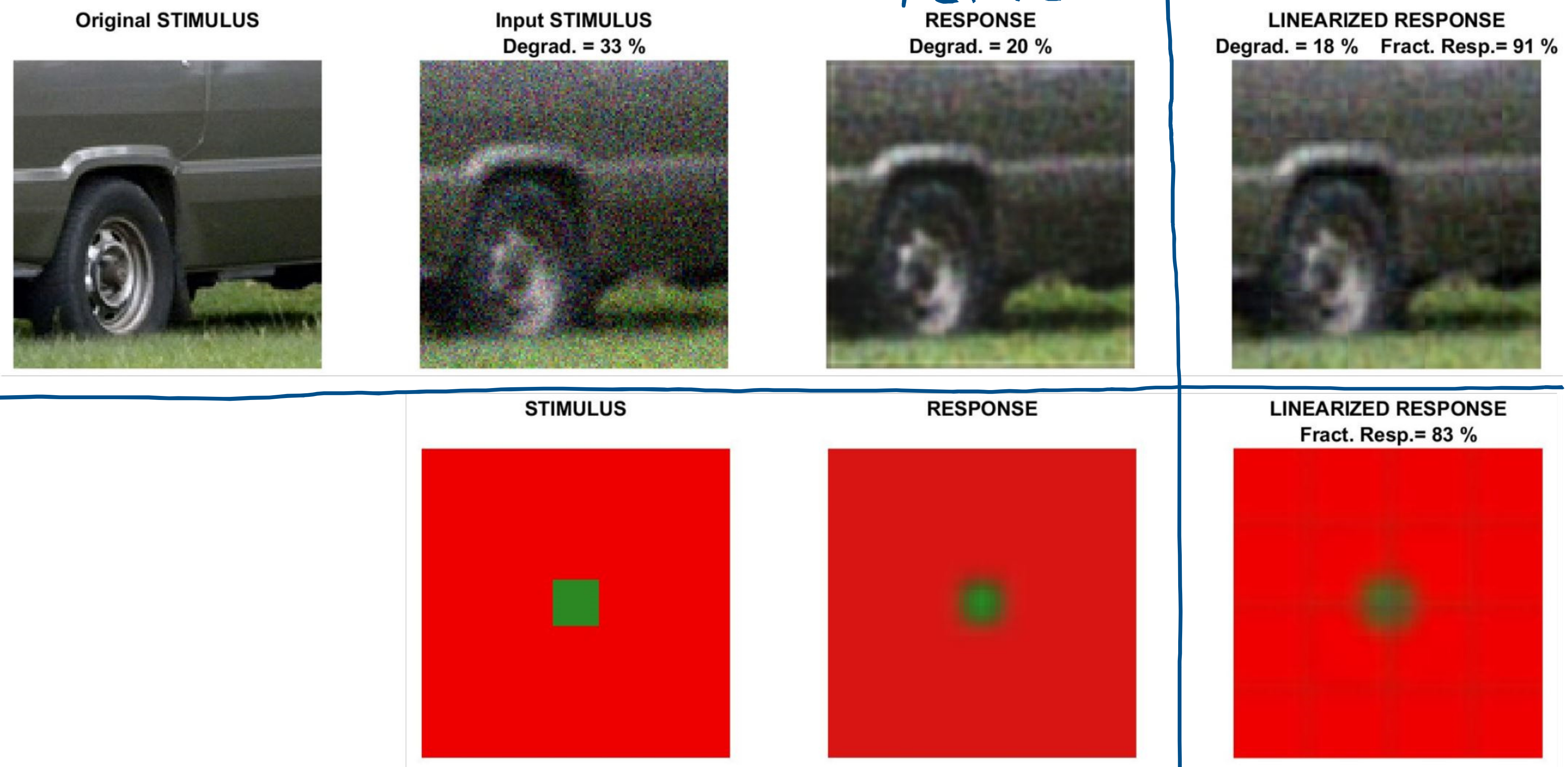
② RESULTS II: Artificial illusions from "artificial" CSFs .1 – Linearization analysis

ACTUAL

APPROXIM.

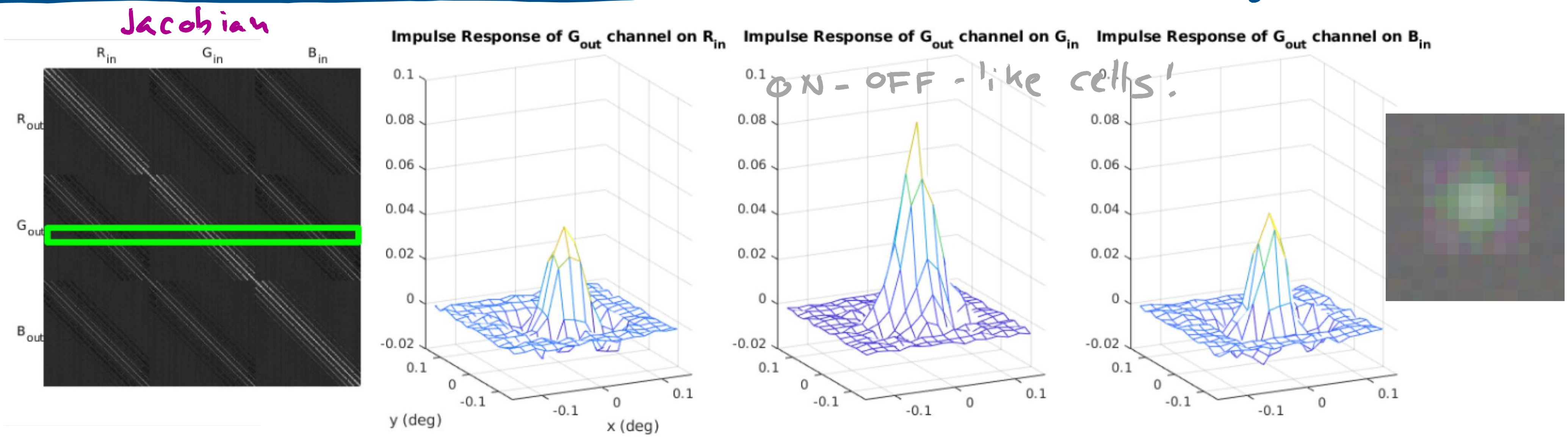
Example 1

Example 2



② RESULTS II: Artificial illusions from "artificial" CSFs .2 – Jacobian, eigenvalues & eigenvectors

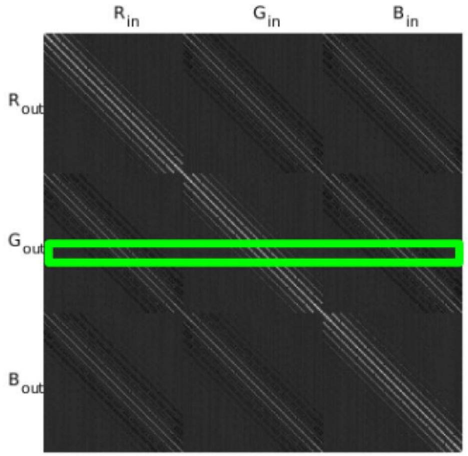
$D_i m =$

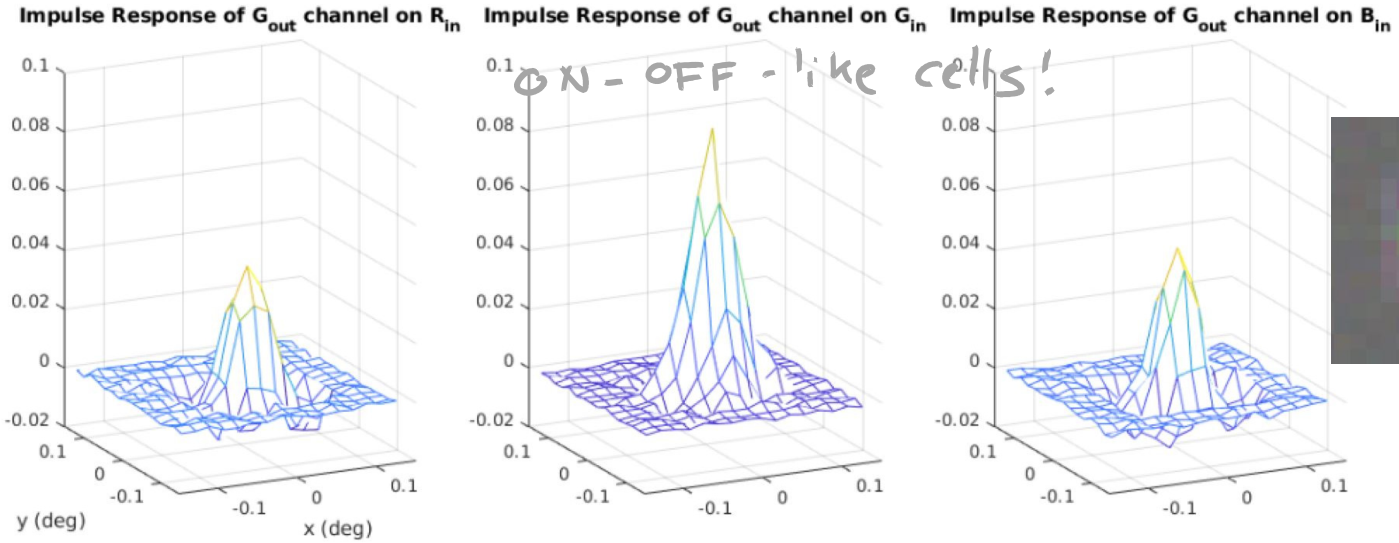


2 RESULTS II: Artificial illusions from "artificial" CSFs .2 - Jacobian, eigenvalues & eigenvectors

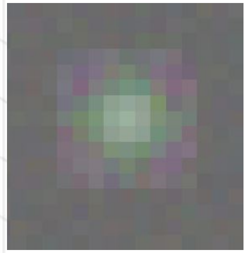
Jacobian

$D_i m =$

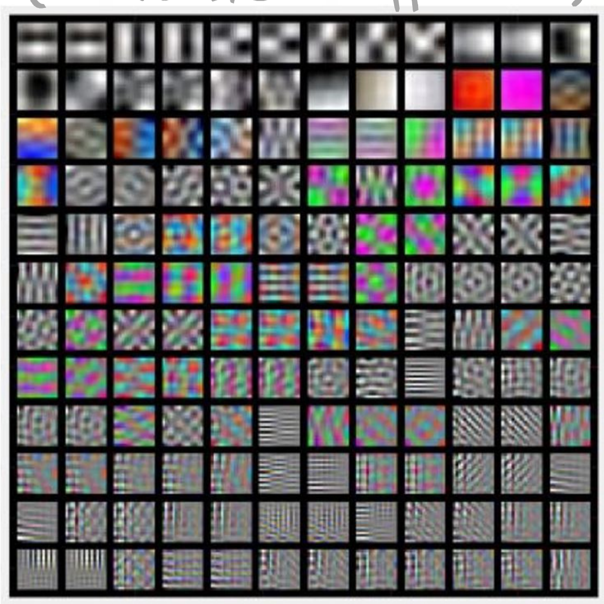




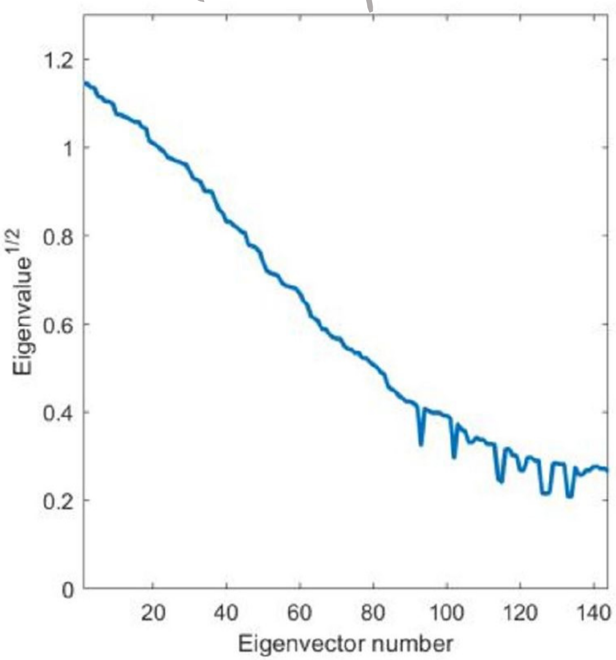
ON-OFF-like cells!



Eigenvectors (~ Fourier & opponents)

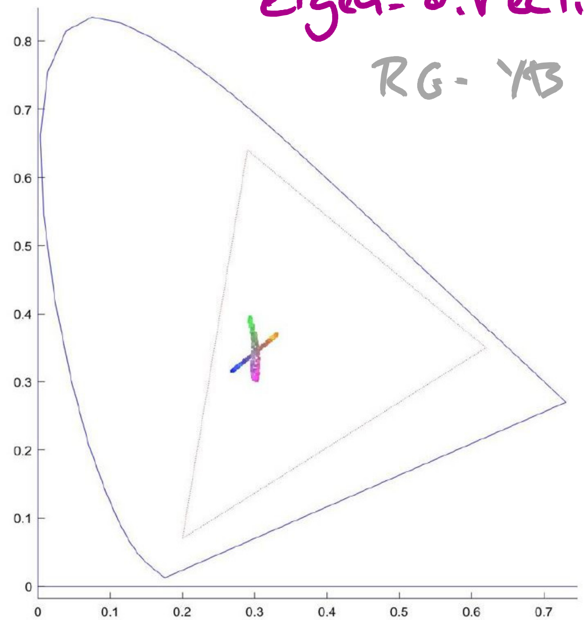


Eigenvalues (~ low-pass)



Color eigen-directions

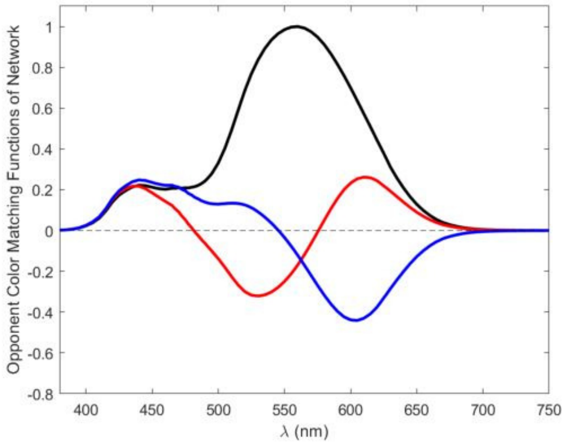
RG - YB!



2 RESULTS II: Artificial illusions rom "artificial" CSFs .3 - Artificial CSFs

ARTIFICIAL SPATIAL FILTERS

ARTIFICIAL SPECTRAL SENSITIV.



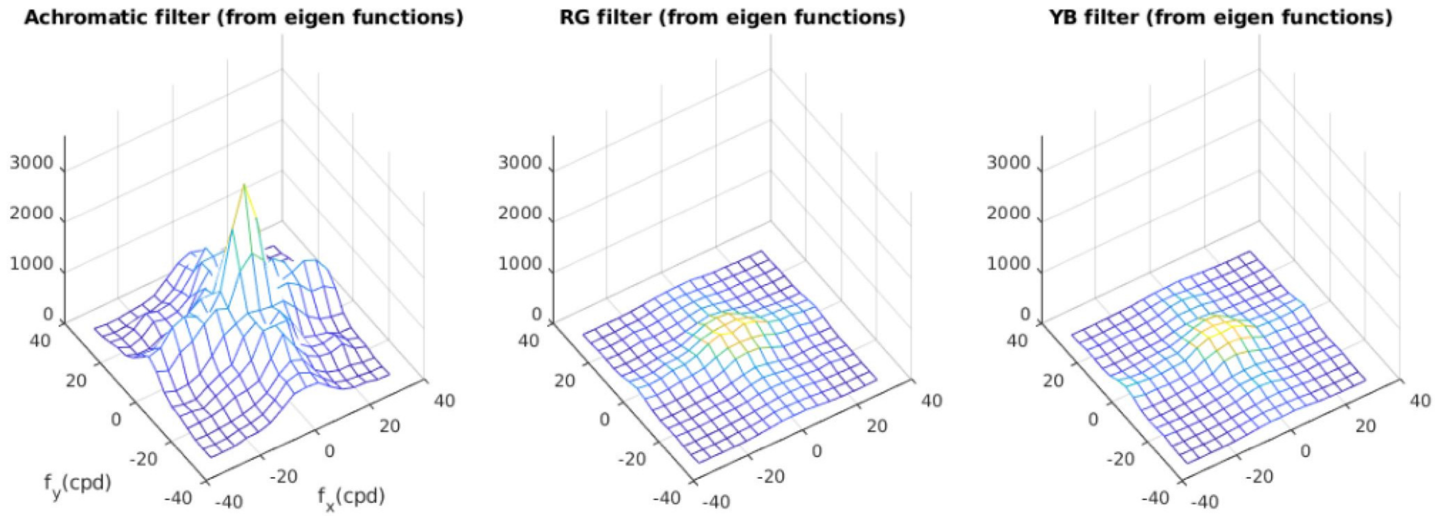
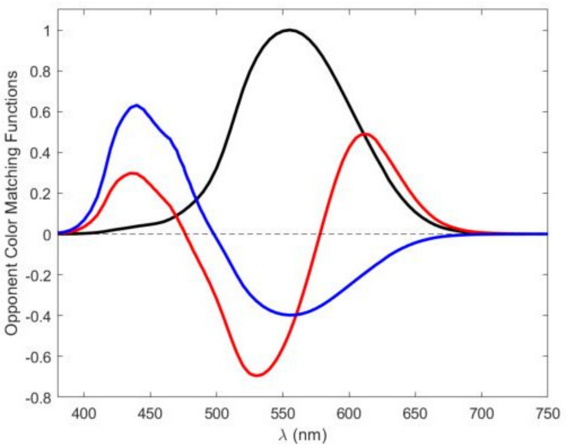
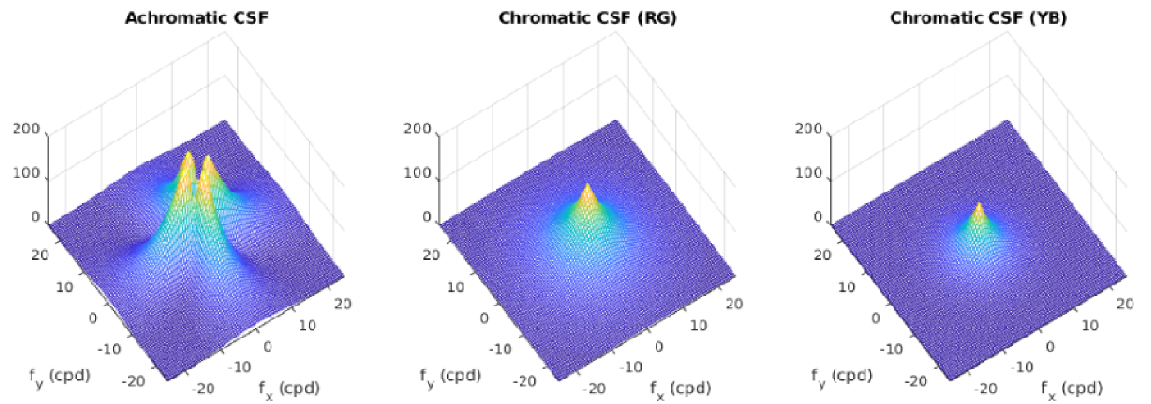


Figure 14: Accumulated spectra of eigenfunctions decomposed in their intrinsic color space and weighted by eigenvalues. Limited frequency resolution is due to the fact that this result comes from small 16 × 16 image blocks. This may give rise to artifacts in the spectra.

HUMAN SPECTRAL SENSITIV.



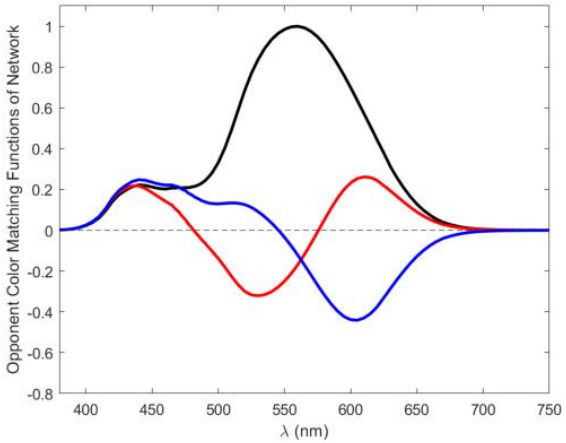
HUMAN CONTRAST SENSITIV. (CSFs)



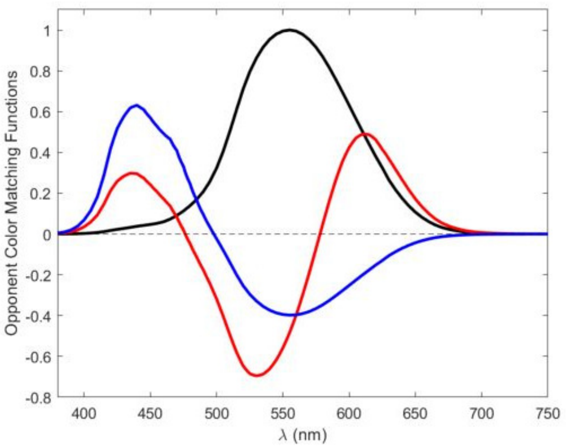
2 RESULTS II: Artificial illusions from "artificial" CSFs .3 – Artificial CSFs

ARTIFICIAL SPATIAL FILTERS

ARTIFICIAL SPECTRAL SENSITIV.



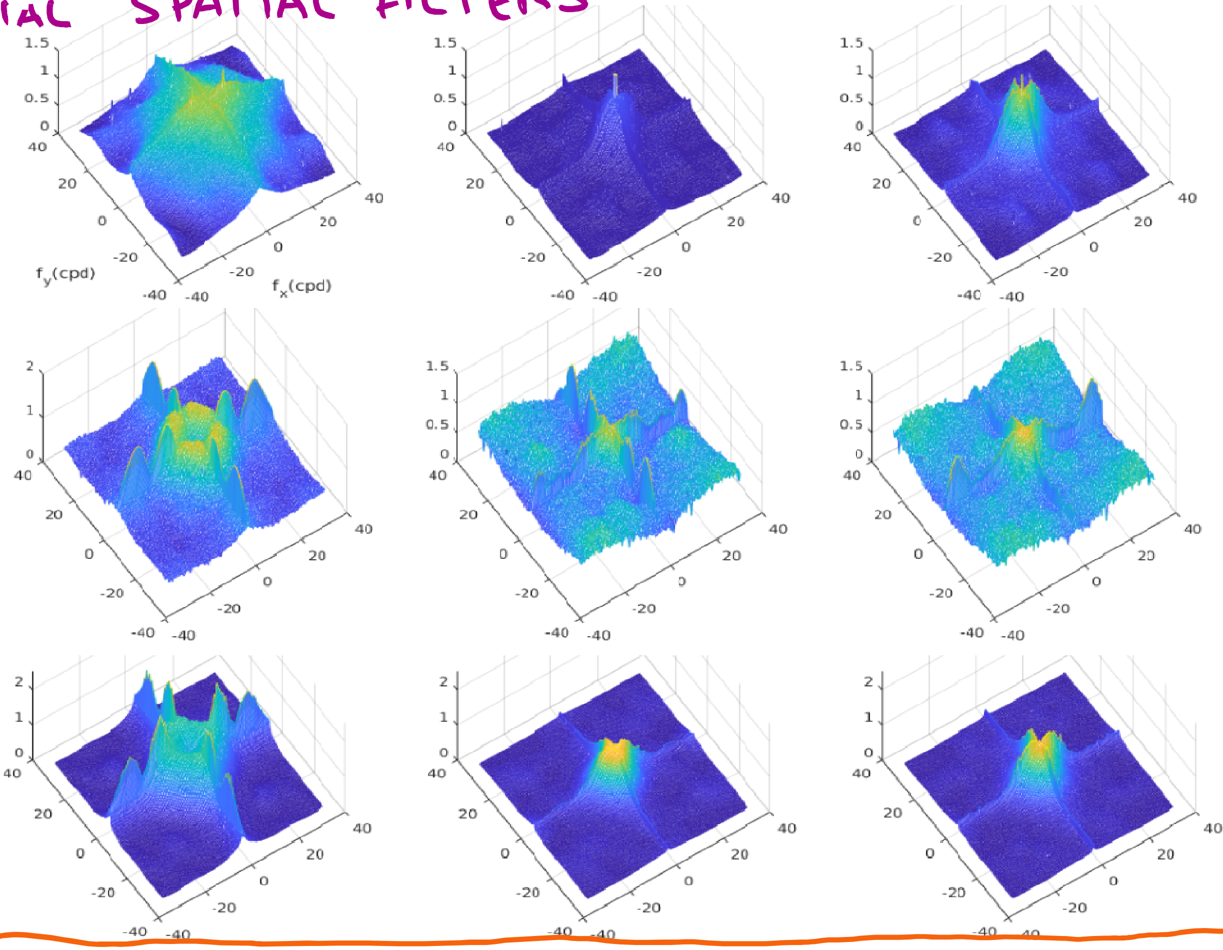
HUMAN SPECTRAL SENSITIV.



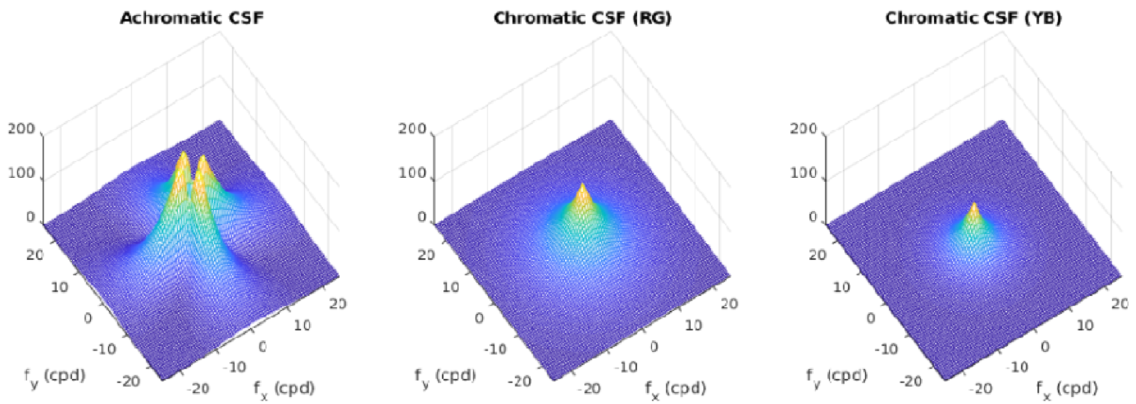
Denoise

Deblur

Restore



HUMAN CONTRAST SENSITIV. (CSFs)

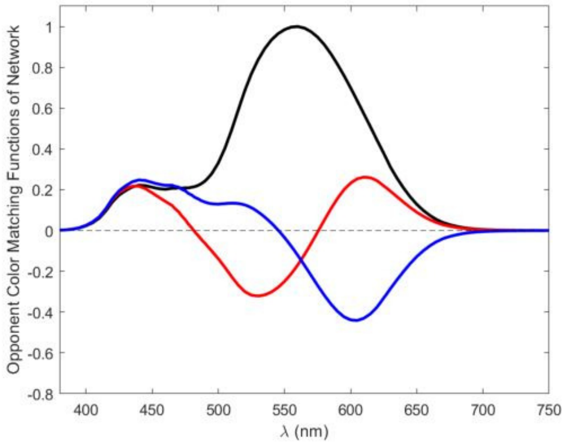


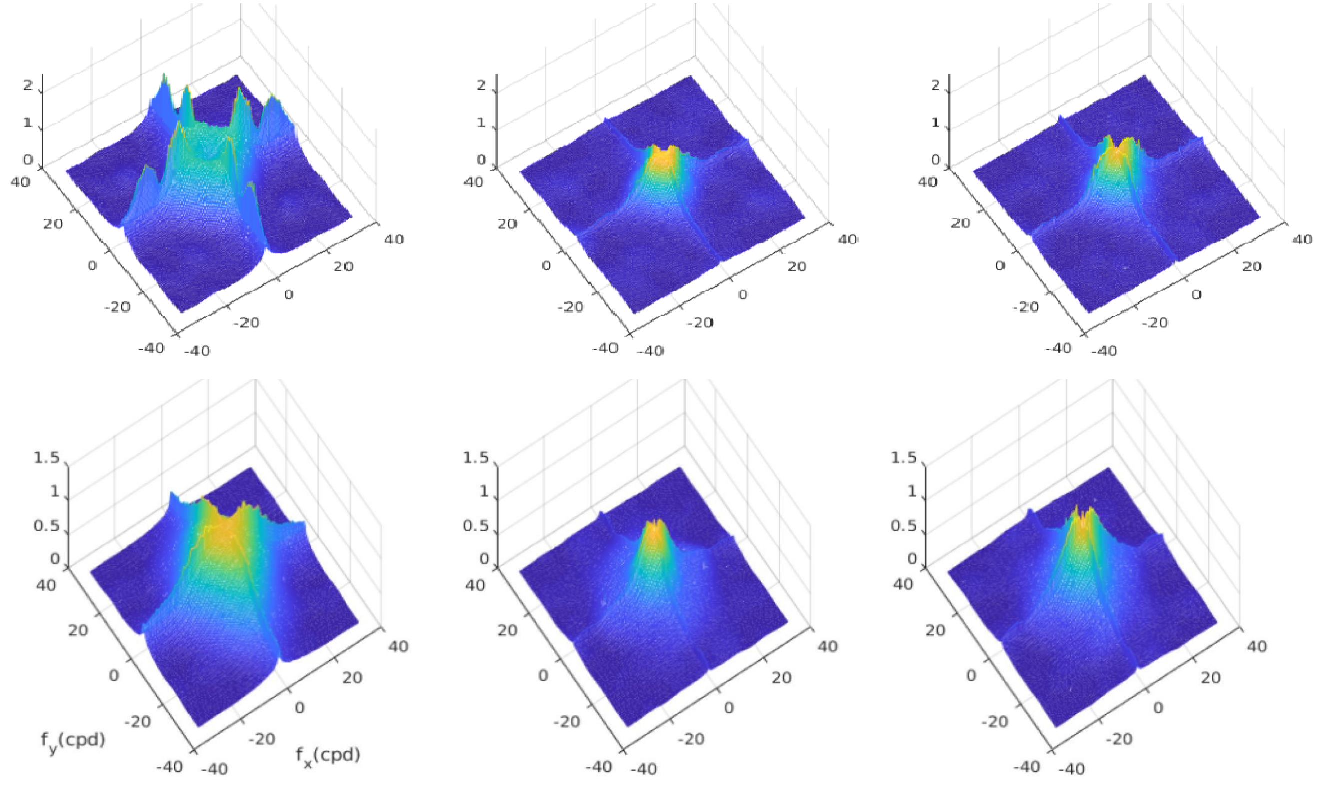
② RESULTS II: Artificial illusions from "artificial" CSFs .3 – Artificial CSFs

ARTIFICIAL SPATIAL FILTERS

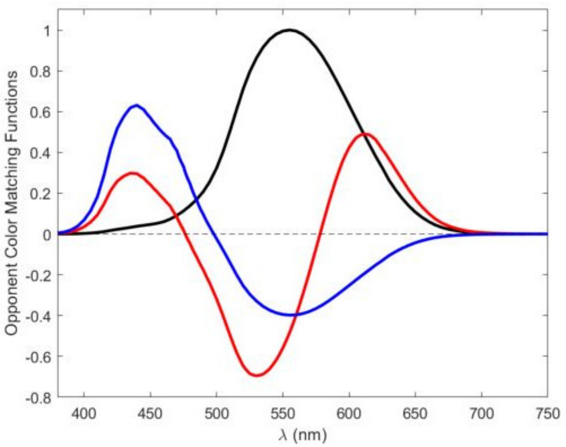
DATA NOT AN ISSUE

ARTIFICIAL SPECTRAL SENSITIV.

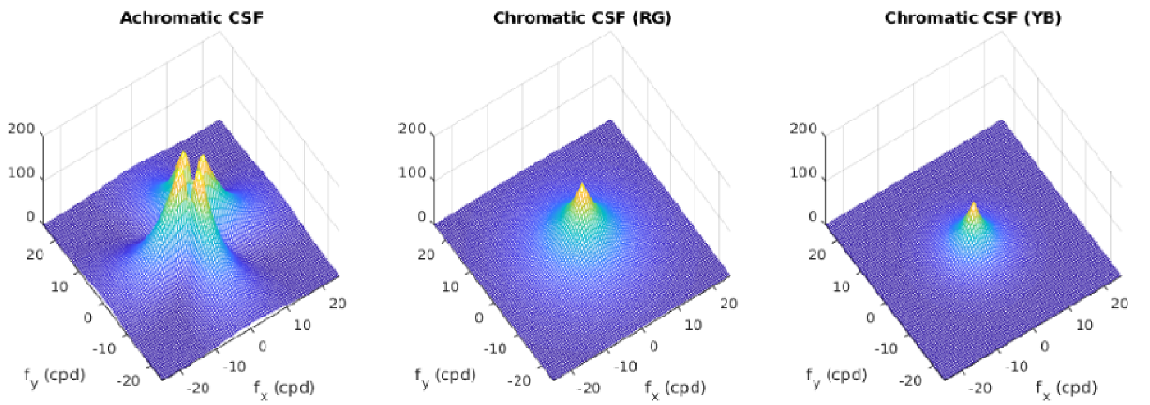




HUMAN SPECTRAL SENSITIV.



HUMAN CONTRAST SENSITIV. (CSFs)



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 Achrom. RG, YB

A. Gómez, J. Malo, et al. (2019) "Visual Illusions also deceive Convolutional Neural Networks: Analysis and Implications".
Submitted to **Vision Research** https://arxiv.org/abs/1912.01643

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 Achrom. RG, YB

* Why? ≡ error-minimization goal (Wiener restorat.) ⇒ fit to spectrum of natural images

  Atick Li Neur. Comp. 92
  Malo et al. IEEE TIP 06

A. Gómez, J. Malo, et al. (2019) "Visual Illusions also deceive Convolutional Neural Networks: Analysis and Implications". Submitted to **Vision Research** https://arxiv.org/abs/1912.01643

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 Achrom. RG, YB

* Why? ≡ error-minimization goal (Wiener restorat.) ⇒ fit to spectrum of natural images — Atick Li Neur. Comp. 92; Malo et al. IEEE TIP 06

* (Generality/illusions, simple architect.) vs (Specificity/no-illusions, complex architect.)

A. Gómez, J. Malo, et al. (2019) "Visual Illusions also deceive Convolutional Neural Networks: Analysis and Implications". Submitted to **Vision Research** https://arxiv.org/abs/1912.01643

3 DISCUSSION & CONCLUSIONS

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

	Shallow	Deep	Zhang et al.
Fract. Lin. Resp.	90%	93%	84%
Error NonLin.	11.2%	10.8%	7.5%
Error Linear	12.5%	12.1%	15.9%
Illusion Strength	++	+++	-

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

	Shallow	Deep	Tao et al.
Fract. Lin. Resp.	88%	93%	94%
Error NonLin.	17.1%	16.3%	15.0%
Error Linear	19.0%	17.1%	15.6%
Illusion Strength	++	+++	+

* (Generality/illusions simple architect.) vs (Specificity/no-illusions complex architect.)

A. Gómez, J. Malo, et al. (2019) "Visual Illusions also deceive Convolutional Neural Networks: Analysis and Implications". Submitted to **Vision Research** https://arxiv.org/abs/1912.01643

3

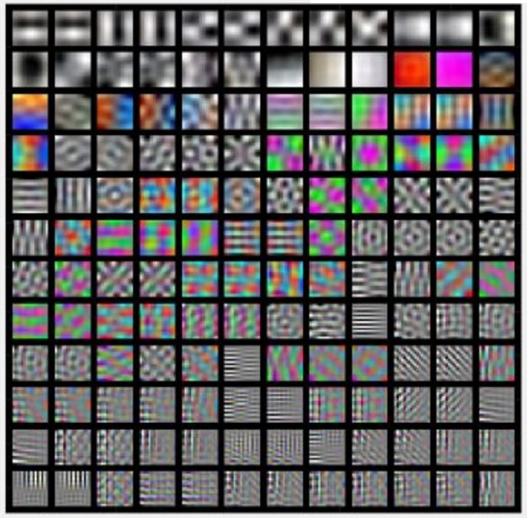
Table 3: Nonlinearity, Performance & Illusion Strength (Denoising)

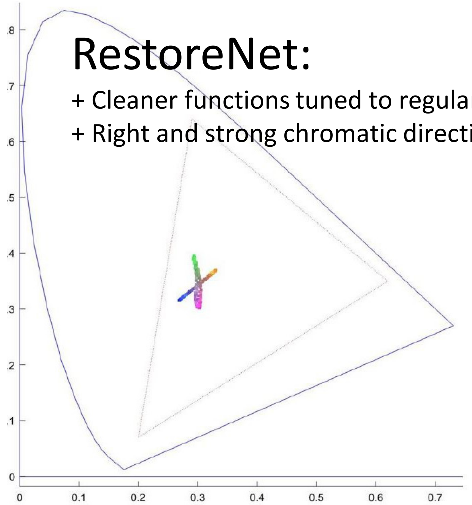
	Shallow	Deep	Zhang et al.
Fract. Lin. Resp.	90%	93%	84%
Error NonLin.	11.2%	10.8%	7.5%
Error Linear	12.5%	12.1%	15.9%
Illusion Strength	++	+++	-

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

	Shallow	Deep	Tao et al.
Fract. Lin. Resp.	88%	93%	94%
Error NonLin.	17.1%	16.3%	15.0%
Error Linear	19.0%	17.1%	15.6%
Illusion Strength	++	+++	++

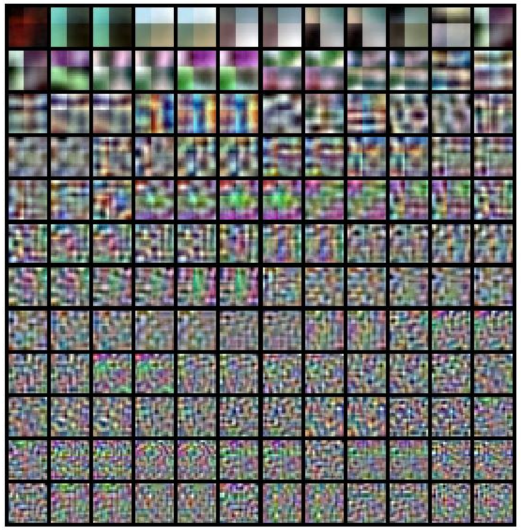
(Generality/illusions simple architect.) vs (Specificity/no-illusions complex architect.)

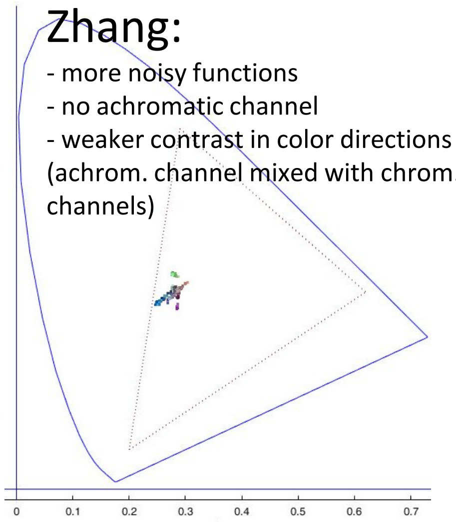




RestoreNet:

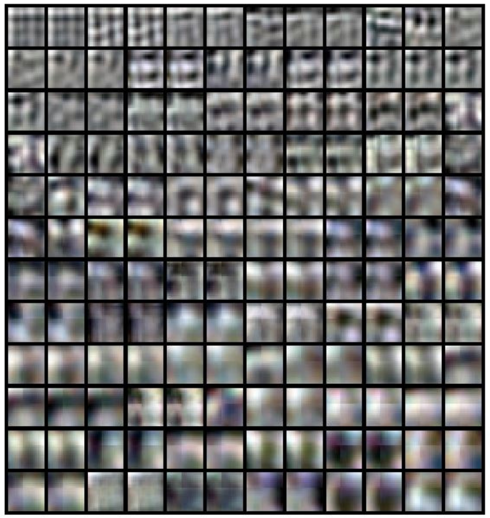
+ Cleaner functions tuned to regular patterns
+ Right and strong chromatic directions

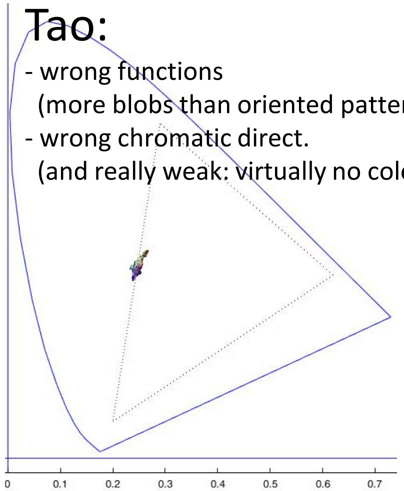




Zhang:

- 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.)

③ DISCUSSION & CONCLUSIONS

* Focus on the objective functions! (the "why" question)

nature neuroscience

FOCUS | PERSPECTIVE

https://doi.org/10.1038/s41593-019-0520-2

A deep learning framework for neuroscience

Blake A. Richards[1,2,3,4,42]*, Timothy P. Lillicrap[5,6,42], Philippe Beaudoin[7], Yoshua Bengio[1,4,8],

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 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.

NATURE NEUROSCIENCE | VOL 22 | NOVEMBER 2019 | 1761–1770 | www.nature.com/natureneuroscience

③ DISCUSSION & CONCLUSIONS

* Focus on the objective functions! (the "why" question)

nature neuroscience

FOCUS | PERSPECTIVE

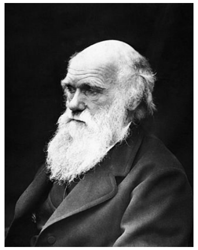
https://doi.org/10.1038/s41593-019-0520-2

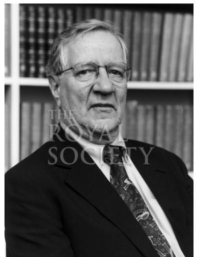
A deep learning framework for neuroscience

Blake A. Richards[1,2,3,4,42]*, Timothy P. Lillicrap[5,6,42], Philippe Beaudoin[7], Yoshua Bengio[1,4,8],

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 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.

NATURE NEUROSCIENCE | VOL 22 | NOVEMBER 2019 | 1761–1770 | www.nature.com/natureneuroscience





③ DISCUSSION & CONCLUSIONS

* Focus on the objective functions! (the "why" question)

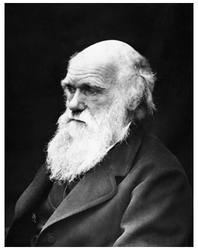
nature neuroscience FOCUS | PERSPECTIVE
https://doi.org/10.1038/s41593-019-0520-2

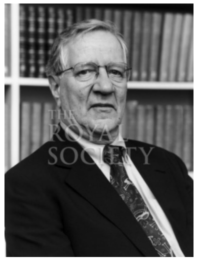
A deep learning framework for neuroscience

Blake A. Richards[1,2,3,4,42*], Timothy P. Lillicrap[5,6,42], Philippe Beaudoin[7], Yoshua Bengio[1,4,8],

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 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.

NATURE NEUROSCIENCE | VOL 22 | NOVEMBER 2019 | 1761–1770 | www.nature.com/natureneuroscience





* 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! } Attick & Li 1992, Laparra & Malo 15
* Caution with blind OVER-fitting. Martinez & Malo 2019

QUESTIONS?

COMMENTS?