The “What” Pathway And the Fast Fourier Transform

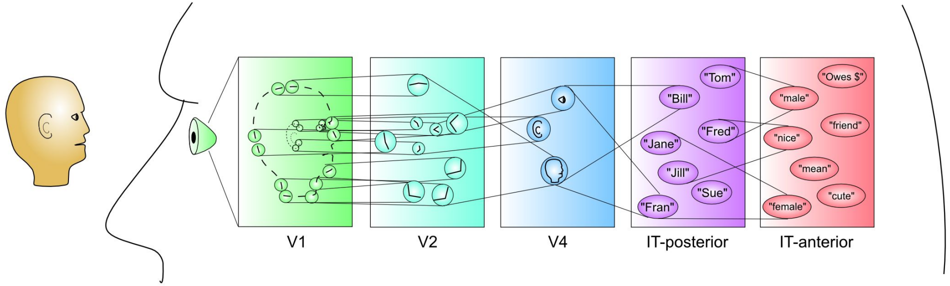
Andrew Valesky

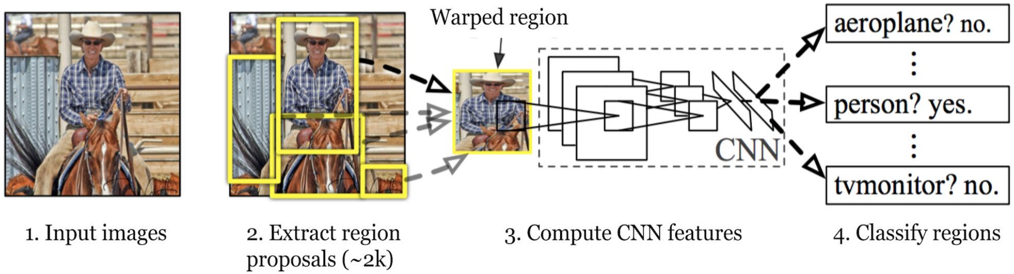
Overview

- The “What” Pathway
- Capsules
- Information Entropy
- The Fast Fourier Transform (FFT)
- The FFT and the “What” Pathway

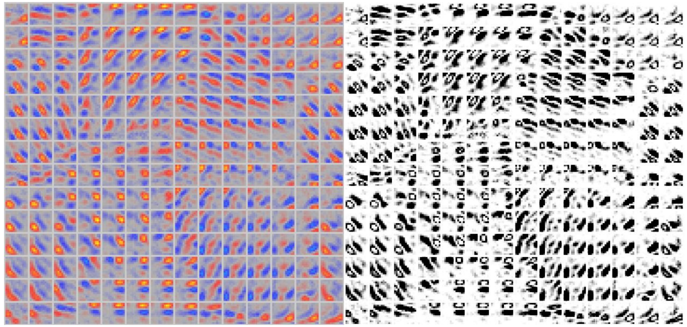
The What Pathway

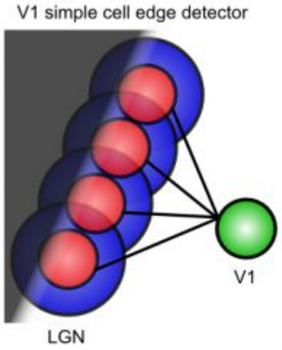
- **V1** --encodes the image in terms of oriented edge detectors that respond to edges along different angles of orientation.
- **V2** -- encodes combinations of edge detectors to develop a vocabulary of intersections and junctions, along with many other basic visual features
- **V4** -- detects more complex shape features, over an even larger range of locations (and sizes, angles, etc).
- **IT-posterior** (PIT) -- detects entire object shapes, over a wide range of locations, sizes, and angles.
- **IT-anterior** (AIT) -- this is where visual information becomes extremely abstract and **semantic** in nature -- it can encode all manner of important information about different people, places and things.





The What Pathway





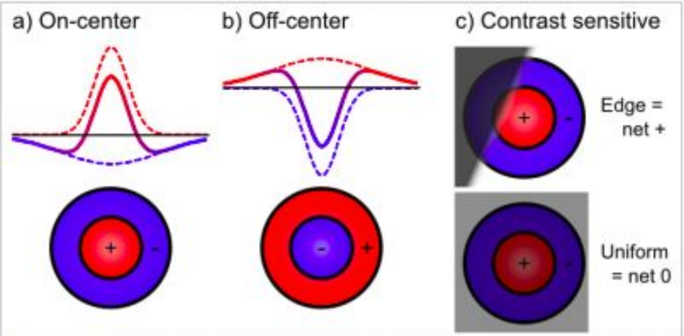
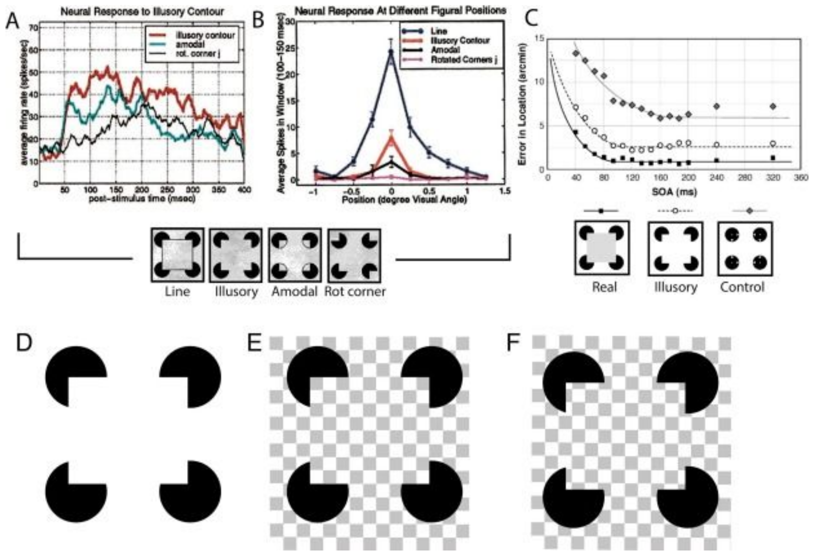


Figure 6.2: How the retina compresses information by only responding to areas of contrasting illumination, not solid uniform illumination. The response properties of retinal cells can be summarized by these Difference-of-Gaussian (DoG) filters, with a narrow central region and a wider surround (also called center-surround receptive fields). The excitatory and inhibitory components exactly cancel when both are uniformly illuminated, but when light falls more on the center vs. the surround (or vice-versa), they respond, as illustrated with an edge where illumination transitions between darker and lighter.

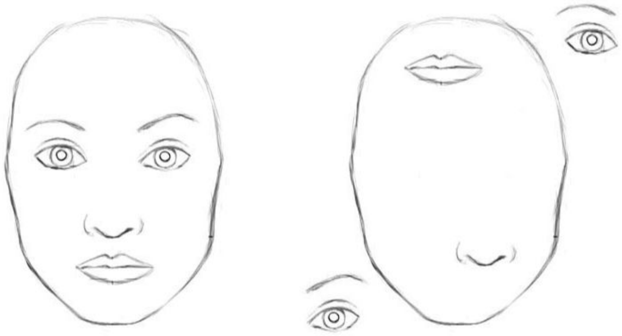
Capsules

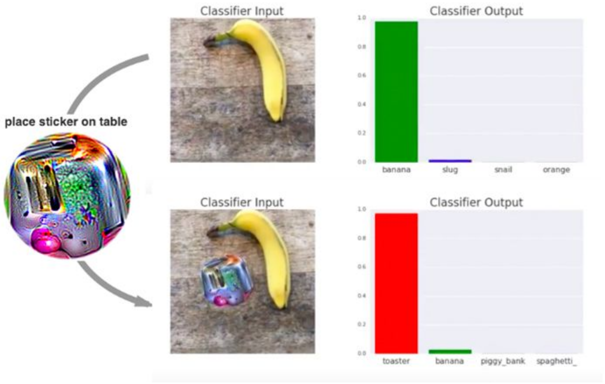
- Modules
- Contour interpolation: A case study in Modularity of Mind
  - Keane, Brian P
    - Jeffrey Fodor

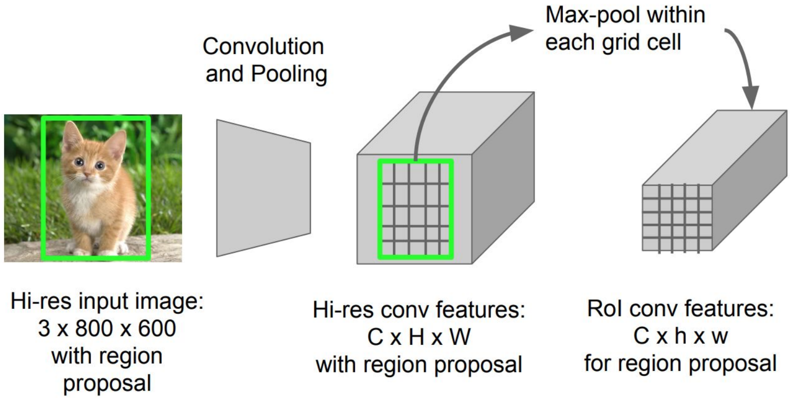


Capsules

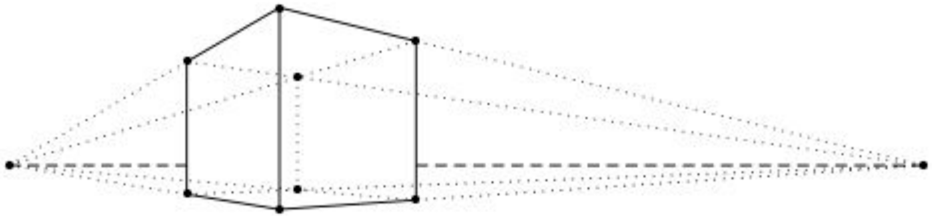
- Geoffrey Hinton
- Max Pooling is terrible
- Encode More Information

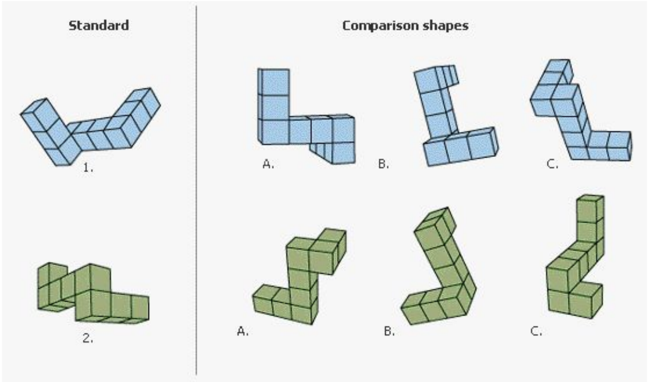






Capsules





Capsule vs. Traditional Neuron			
Input from low-level capsule/neuron		vector($\mathbf{u}_i$)	scalar(x_i)
Operation	Affine Transform	$\widehat{\mathbf{u}}_{j\|i} = \mathbf{W}_{ij}\mathbf{u}_i$	–
	Weighting	$\mathbf{s}_j = \sum_i c_{ij}\widehat{\mathbf{u}}_{j\|i}$	$a_j = \sum_i w_i x_i + b$
	Sum		
	Nonlinear Activation	$\mathbf{v}_j = \frac{\lVert\mathbf{s}_j\rVert^2}{1+\lVert\mathbf{s}_j\rVert^2}\frac{\mathbf{s}_j}{\lVert\mathbf{s}_j\rVert}$	$h_j = f(a_j)$
Output		vector($\mathbf{v}_j$)	scalar(h_j)

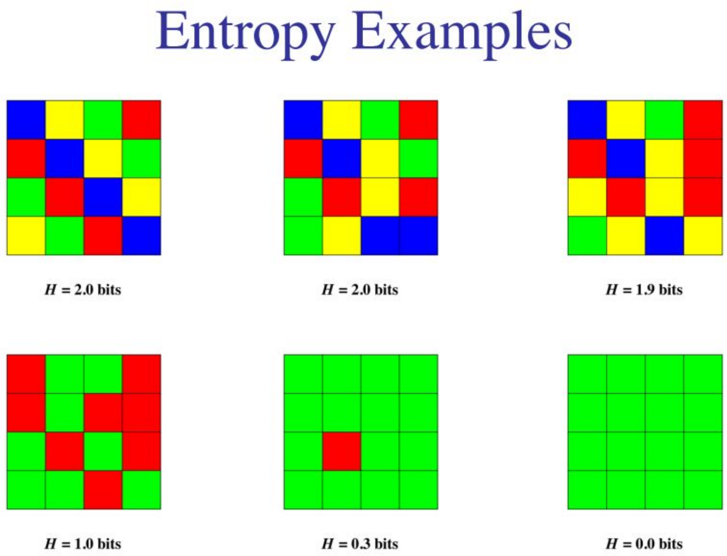
$$\mathbf{v}_j = \frac{\lVert\mathbf{s}_j\rVert^2}{1+\lVert\mathbf{s}_j\rVert^2}\frac{\mathbf{s}_j}{\lVert\mathbf{s}_j\rVert}$$

additional “squashing” — unit scaling

Information Entropy

- “Humans and Deep Networks Largely Agree on Which Kinds of Variation Make Object Recognition Harder”
  - Kheradpisheh, Saeed R. et al
  - August 2016
- Compared Humans and DCNN (Deep Convolutional Neural Net)
- The three-dimensional object models are constructed by O'Reilly et al. (2013)
- Humans and Computers Agree

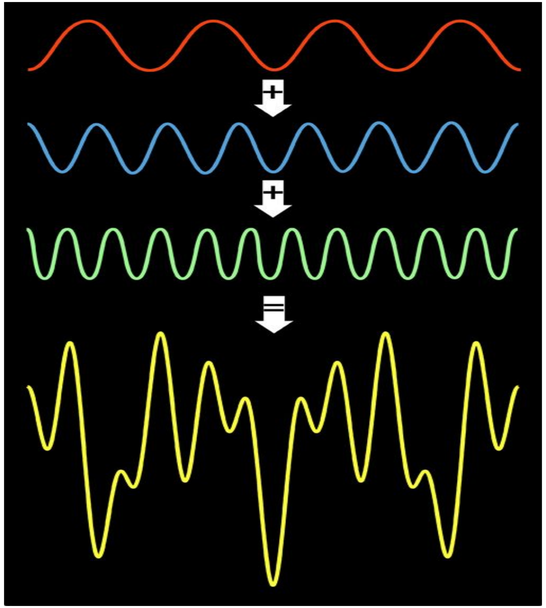
Information Entropy

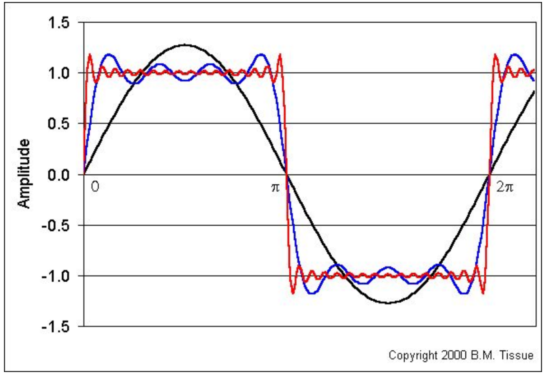


Information Entropy

FFT

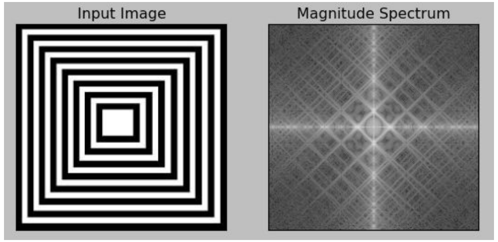
- “Differences in velocity-information processing between two areas in the auditory cortex of mustached bats”
  - Teng & Suga
  - July, 2017
- Bats use frequency decomposition for all information
- Not all bats are blind





FFT

- Fast Fourier Transform
  - Spatial to Frequency
  - Measure of Intensity
  - Black & White
    - 0 - 255
  - Red, Green, Blue (RGB)
    - R,G,B = 0-255
    - $I = \frac{1}{3}(R + G + B)$



$$F(x) = \sum_{n=0}^{N-1} f(n) e^{-j2\pi(x\frac{n}{N})}$$

$$f(n) = \frac{1}{N}\sum_{n=0}^{N-1} F(x) e^{j2\pi(x\frac{n}{N})}$$

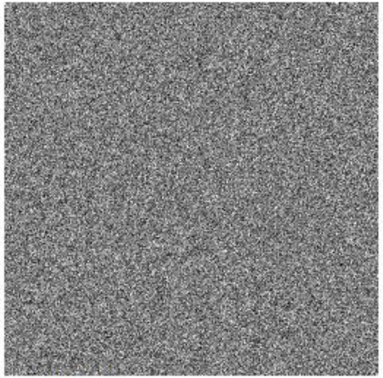
$$f(x) = \int_{-\infty}^{\infty} F(u) e^{j2\pi ux} du$$

$$f(x) = \int_{-\infty}^{\infty} F(u) e^{j2\pi ux} du$$

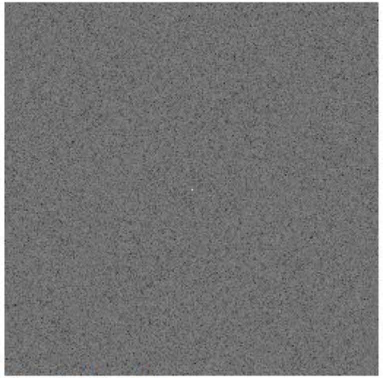
$$F(x,y) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f(m,n) e^{-j2\pi(x\frac{m}{M}+y\frac{n}{N})}$$

$$f(m,n) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} F(x,y) e^{j2\pi(x\frac{m}{M}+y\frac{n}{N})}$$

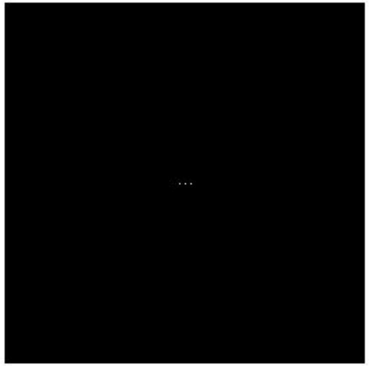
FFT

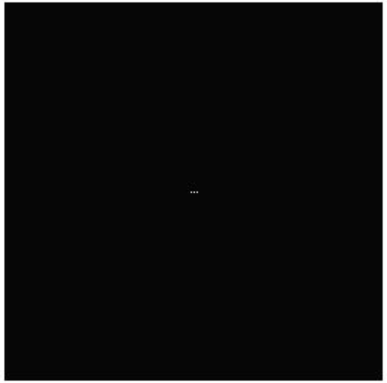


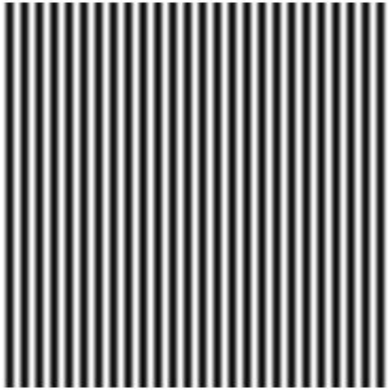
Gaussian Noise

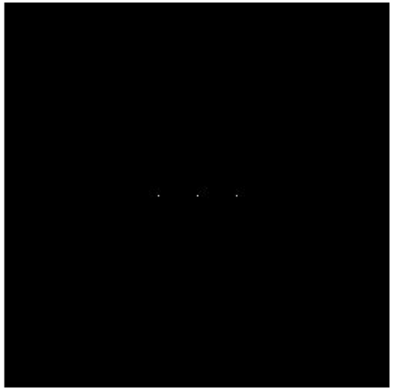


Gaussian Noise FFT

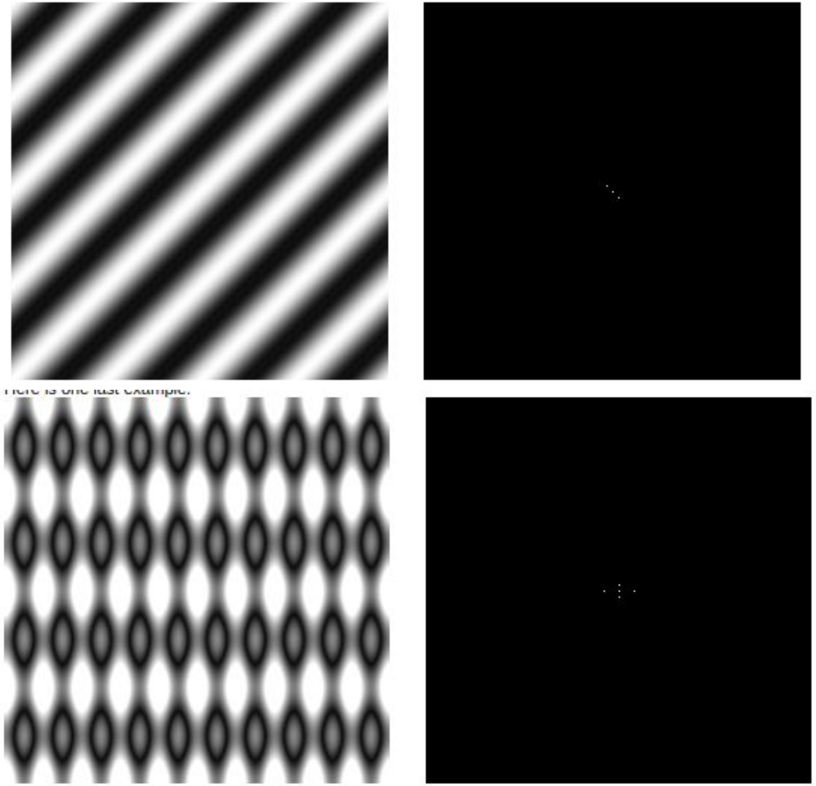


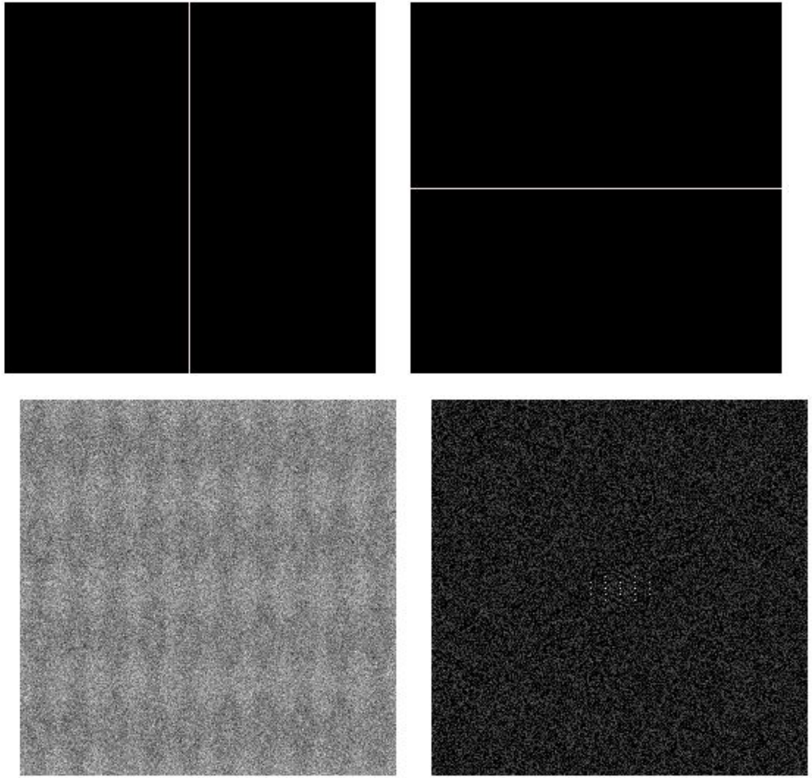




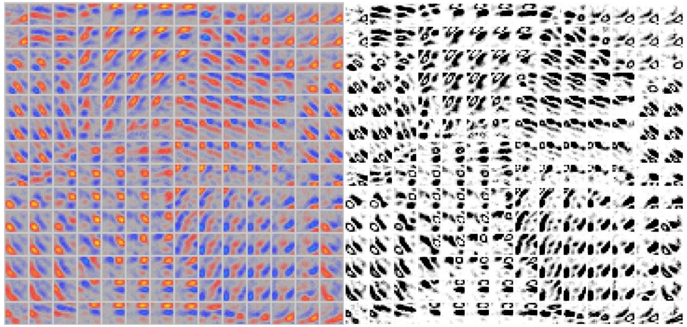


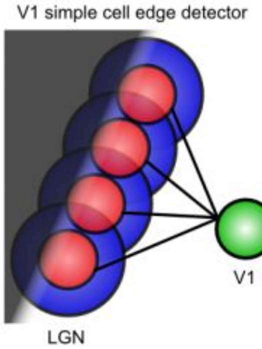
FFT





The What Pathway





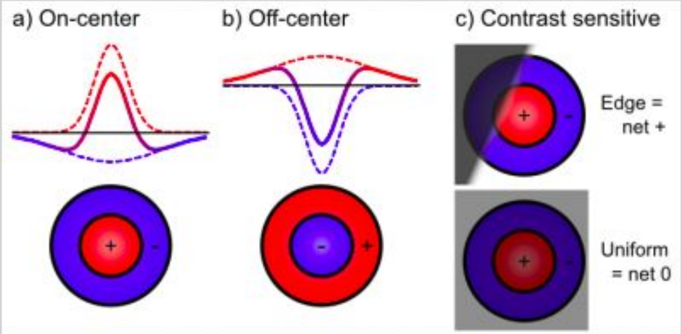


Figure 6.2: How the retina compresses information by only responding to areas of contrasting illumination, not solid uniform illumination. The response properties of retinal cells can be summarized by these Difference-of-Gaussian (DoG) filters, with a narrow central region and a wider surround (also called center-surround receptive fields). The excitatory and inhibitory components exactly cancel when both are uniformly illuminated, but when light falls more on the center vs. the surround (or vice-versa), they respond, as illustrated with an edge where illumination transitions between darker and lighter.

FFT and the “What” Pathway

- Gaussian Filter (GLPF)

$$D(u,v) = [(u - M/2)^2 + (v - N/2)^2]^{1/2}$$

- Gaussian Low Pass Filter (GLPF)

$$H(u,v) = e^{-D^2(u,v)/2\sigma^2}$$

- Gaussian High Pass Filter (GLPF)

$$H(u,v) = 1 - e^{-D^2(u,v)/2D_0^2}$$

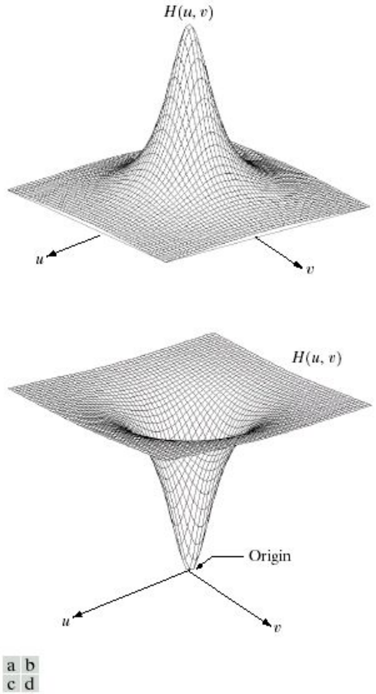
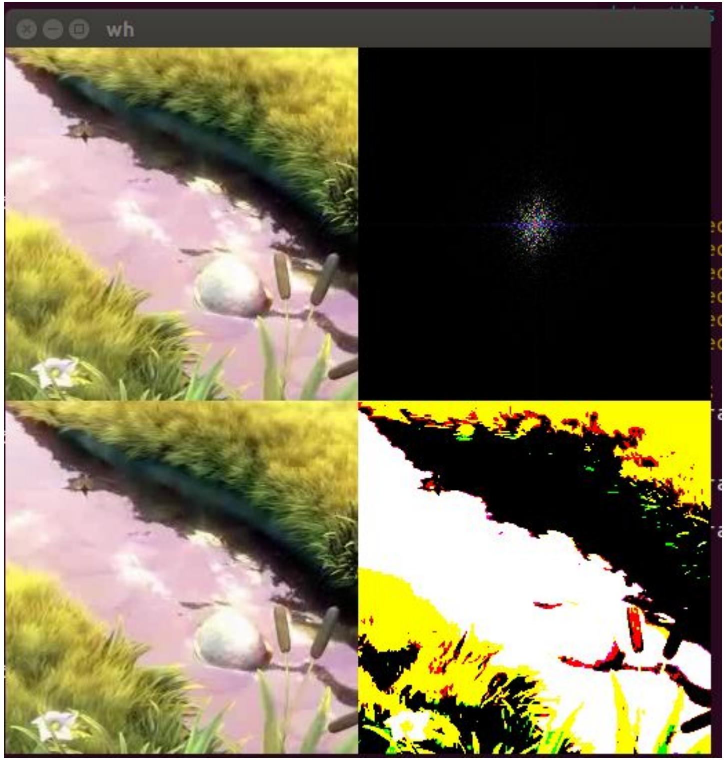
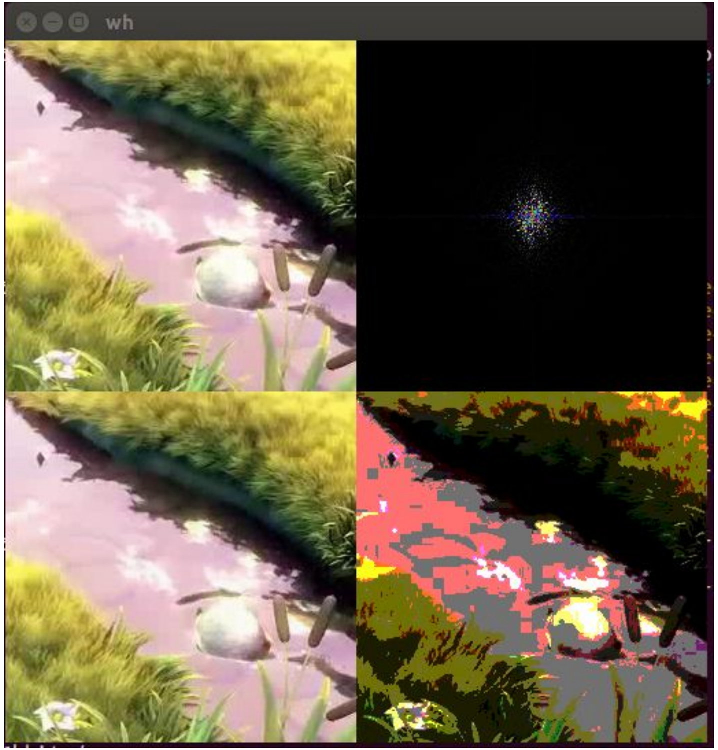


FIGURE 4.7 (a) A two-dimensional lowpass filter functior
(c) A two-dimensional highpass filter function. (d) Result

FFT and the "What" Pathway



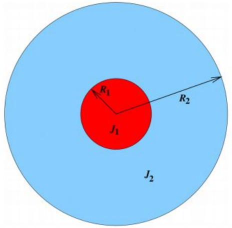


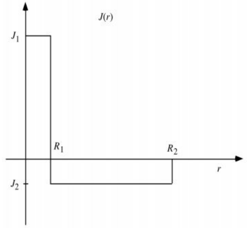
FFT and the “What” Pathway

Differentiation & Pattern Formation

- A central problem in development: How do cells differentiate to fulfill different purposes?
- How do complex systems generate spatial & temporal structure?
- CAs are natural models of intercellular communication

Interaction Parameters





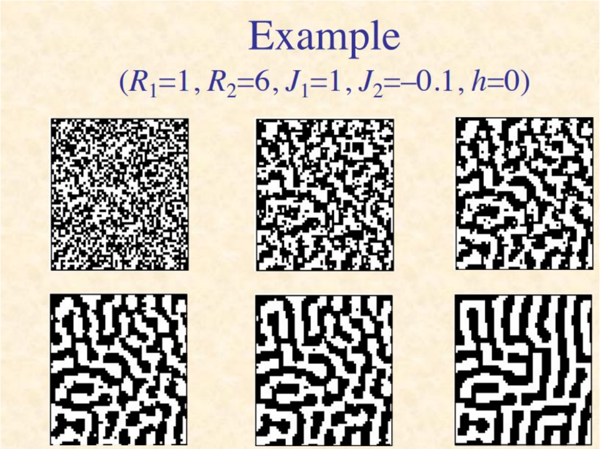
- R_1 and R_2 are the interaction ranges
- J_1 and J_2 are the interaction strengths

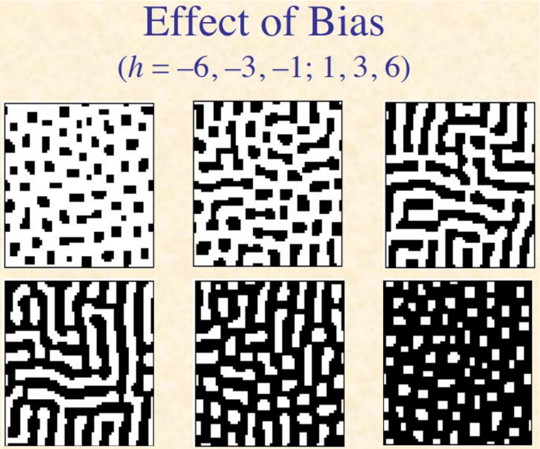
CA Activation/Inhibition Model

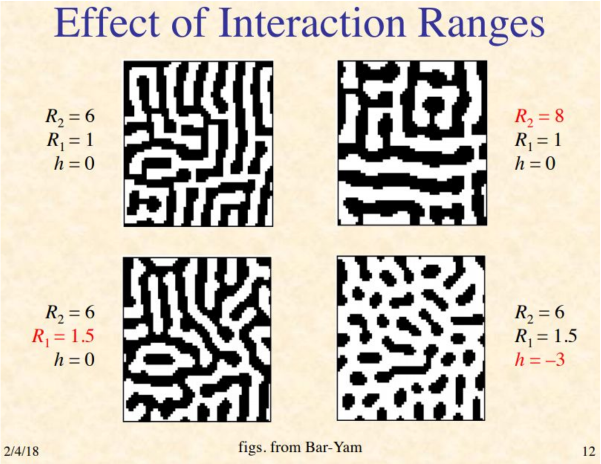
- Let states $s_i \in \{-1, +1\}$
- and h be a bias parameter
- and r_{ij} be the distance between cells i and j
- Then the state update rule is:

$$s_i(t+1) = \text{sign}\left[h + J_1 \sum_{r_{ij} < R_1} s_j(t) + J_2 \sum_{R_1 \le r_{ij} < R_2} s_j(t)\right]$$

FFT and the "What" Pathway







FFT and the "What" Pathway

$$D(u,v) = [(u - M/2)^2 + (v - N/2)^2]^{1/2}$$

$$F(x,y) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f(m,n) e^{-j2\pi(x\frac{m}{M}+y\frac{n}{N})}$$

$$f(m,n) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} F(x,y) e^{j2\pi(x\frac{m}{M}+y\frac{n}{N})}$$

Ideal Low Pass Filter

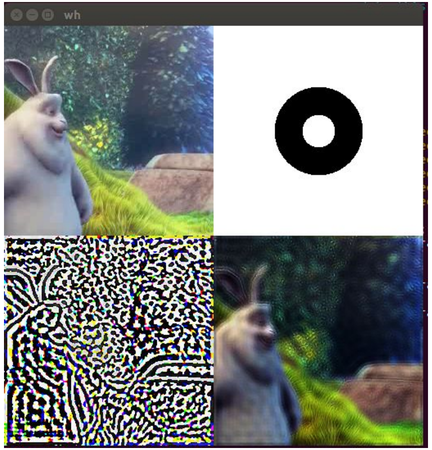
$$H(u,v) = \begin{cases} 1 & \text{if} \quad D(u,v) \le D_0 \\ 0 & \text{if} \quad D(u,v) > D_0 \end{cases}$$

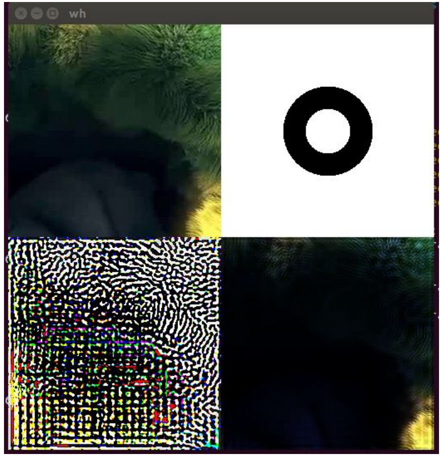
Ideal High Pass Filter

$$H(u,v) = \begin{cases} 0 & \text{if} \quad D(u,v) \le D_0 \\ 1 & \text{if} \quad D(u,v) > D_0 \end{cases}$$

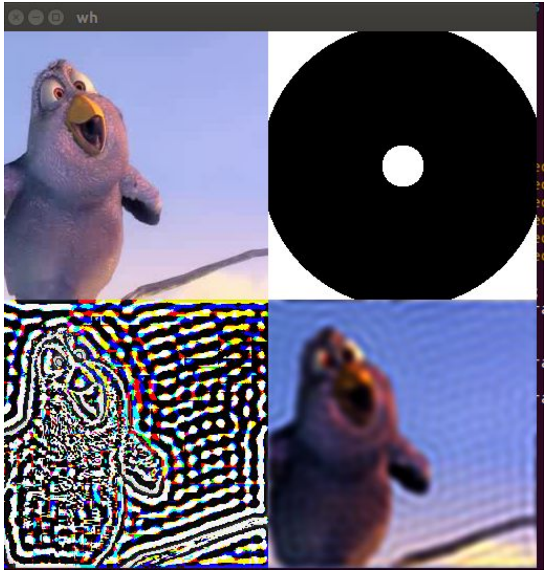
Image = Original - Transform

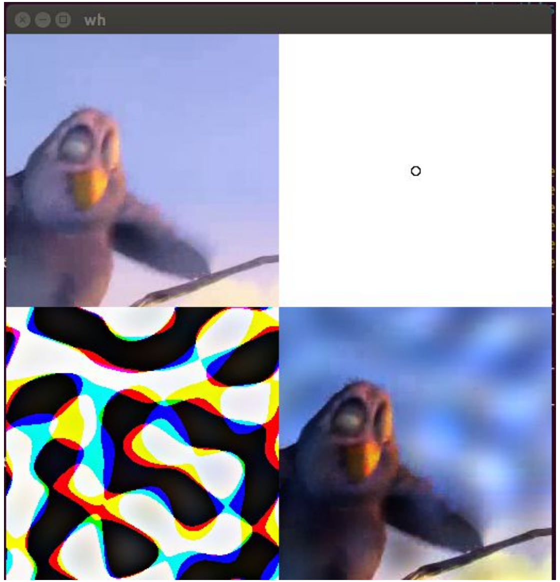
FFT and the "What" Pathway

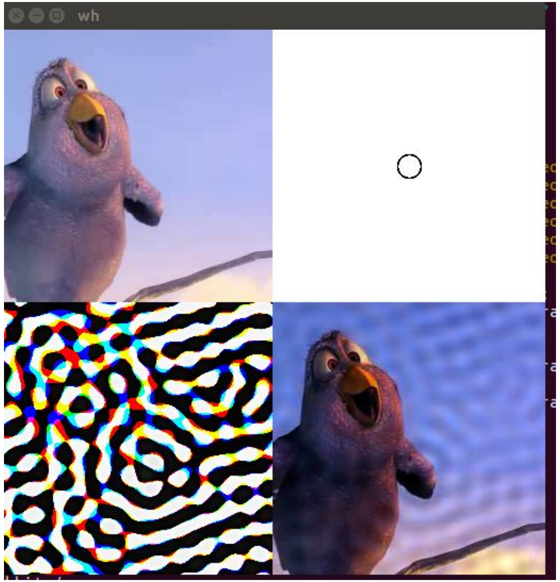




FFT and the “What” Pathway







Conclusions

- Capsules? Yes, I agree.
- FFT?
  - Animal Camouflage mimics the difference in spatial frequency
  - The Effect of Bias creates a different appearance.
  - Does Camouflage correlate to the hunter?
    - Will two different species converge to the same “look” based on being hunted by the same animal?

Work Cited

https://medium.com/ai%C2%B3-theory-practice-business/understanding-hintons-capsule-networks-part-i-intuition-b4b559d1159b

https://medium.com/ai%C2%B3-theory-practice-business/understanding-hintons-capsule-networks-part-ii-how-capsules-work-153b6ade9f66

https://www.frontiersin.org/articles/10.3389/fncom.2016.00092/full

https://web.eecs.utk.edu/~mclennan/Classes/420-527/handouts/Part%202B%20Patn%20Form.pdf

https://www.sciencedirect.com/science/article/pii/S0010027718300167

https://www-ncbi-nlm-nih-gov.proxy.lib.utk.edu:2050/pubmed/28456093

Work Cited Image

https://psychlopedia.wikispaces.com/mental+rotation

https://tex.stackexchange.com/questions/78145/how-can-i-draw-a-transparent-cube-in-3d-perspective

https://gizmodo.com/this-simple-sticker-can-trick-neural-networks-into-thin-1821735479

https://www.facebook.com/ThatManSloth/

http://www.qsimaging.com/ccd_noise_interpret_ffts.html

https://en.wikipedia.org/wiki/HSL_and_HSV