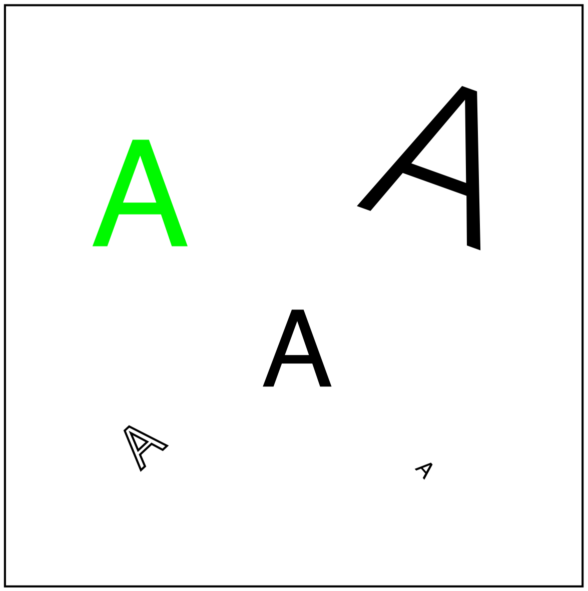
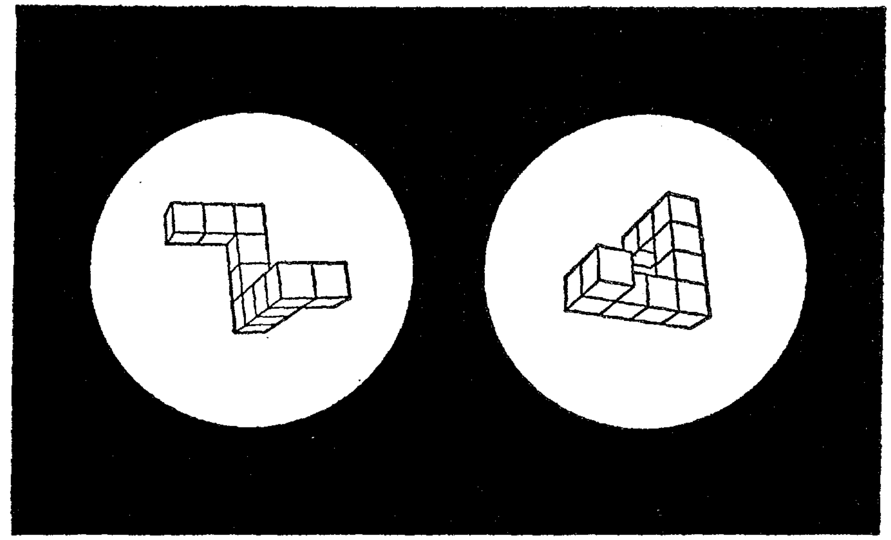
In search of invariance in brains and machines

Bruno Olshausen
Helen Wills Neuroscience Institute, School of Optometry
Redwood Center for Theoretical Neuroscience
UC Berkeley


Sophia Sanborn
Christian Shewmake
Redwood Center for Theoretical Neuroscience
April 2022

How do we see these things as the same?



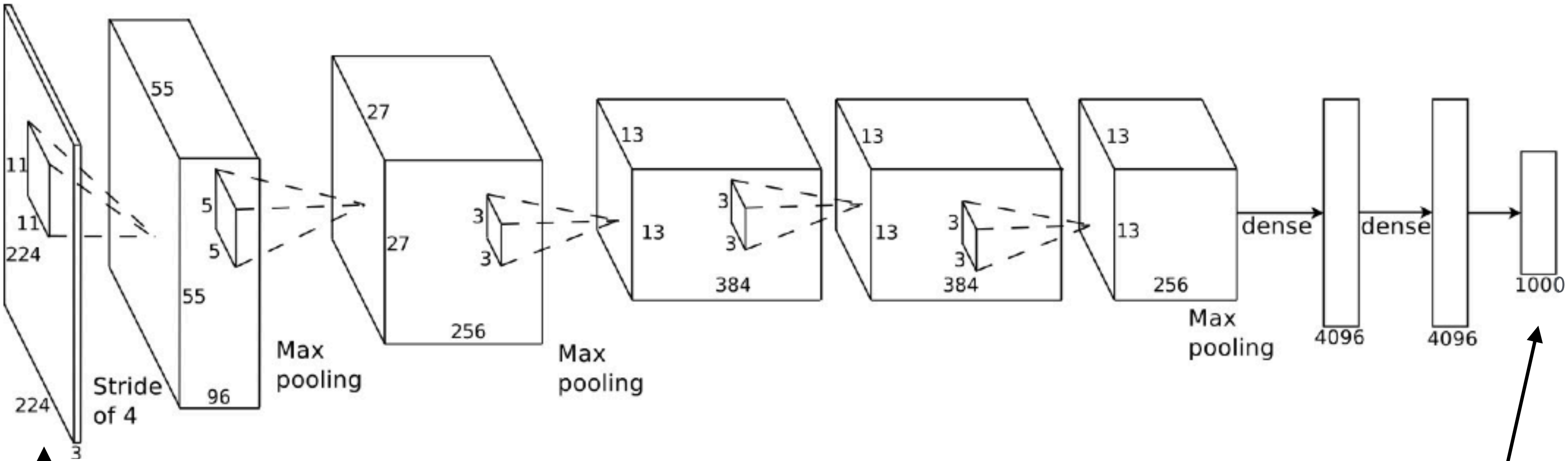


Shepard & Metzler (1971)

The image of a single object has 9 factors of variation:

- 3D position (3)
- 3D rotation (3)
- Photometric (3)

Assuming 100 distinct states for each yields $100^9 = 10^{18}$ variations.



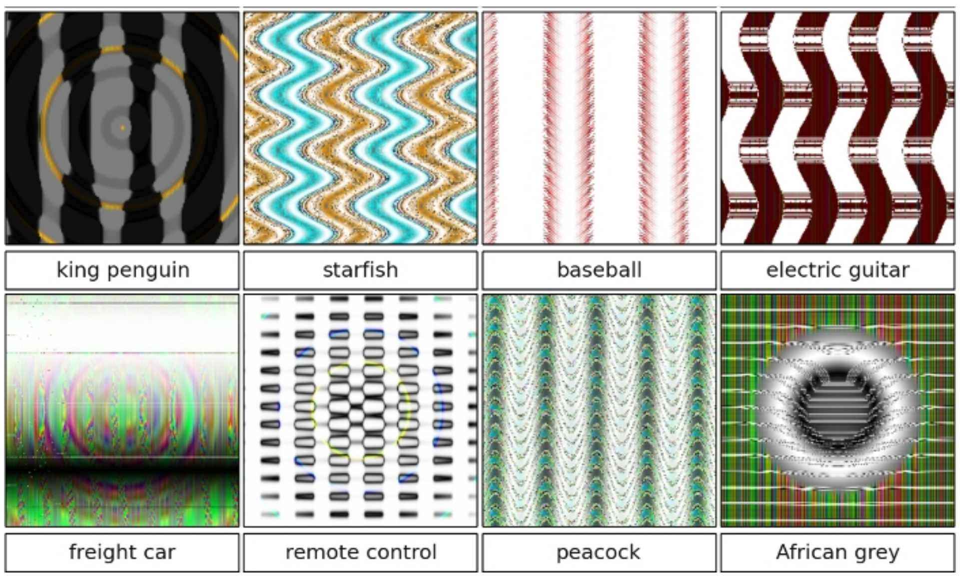
image

feature extraction and pooling

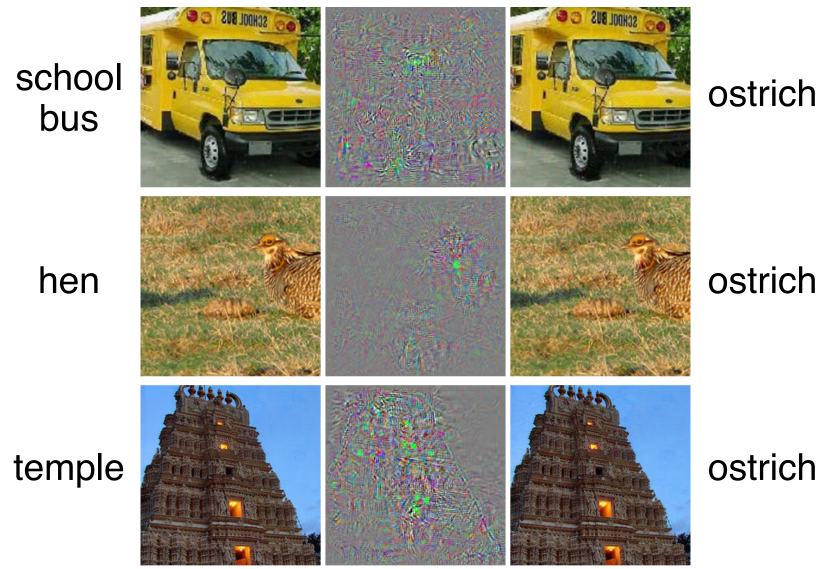
classification

‘cat’

The invariant representations produced by deep convnets have a *high false-positive rate*



easily fooled

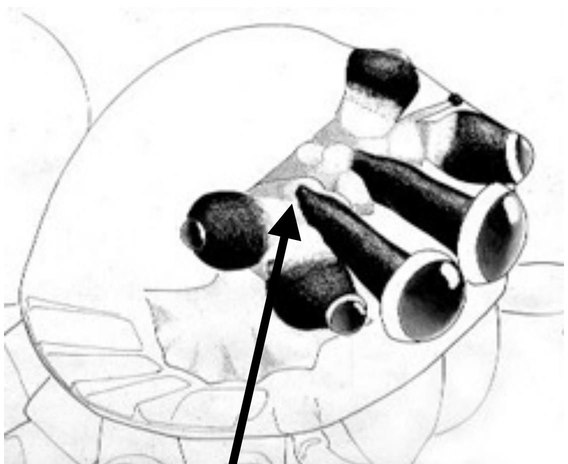


brittle

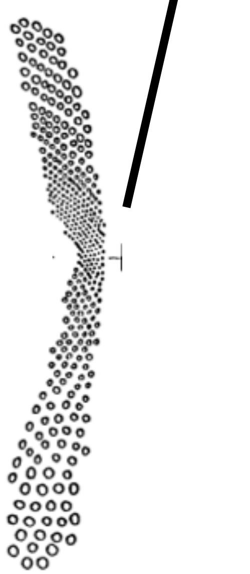
Jacobsen, J. H., Behrmann, J., Zemel, R., & Bethge, M. (2018).
Excessive invariance causes adversarial vulnerability. arXiv:1811.00401.

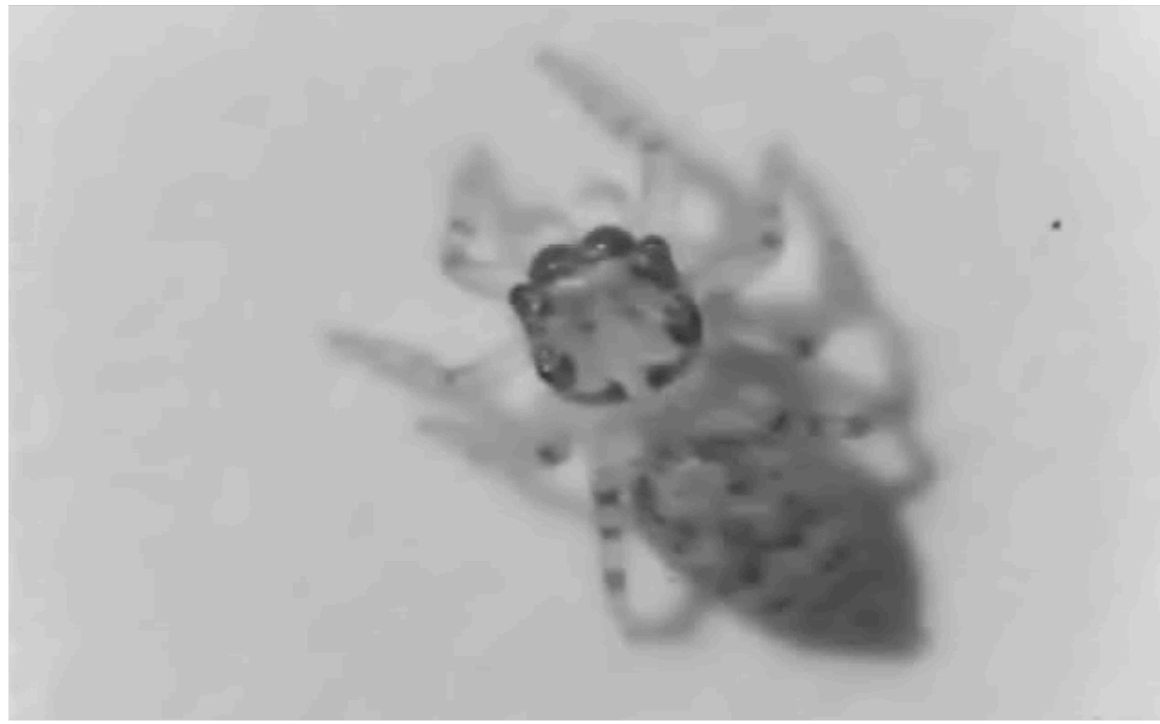
What is vision for? How did it evolve?

Vision in jumping spiders



(Wayne Maddison)

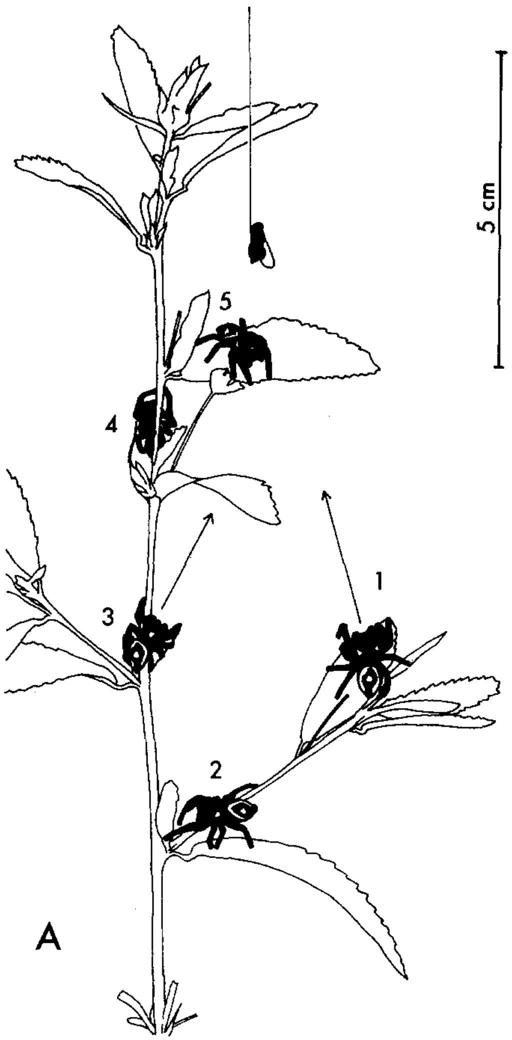


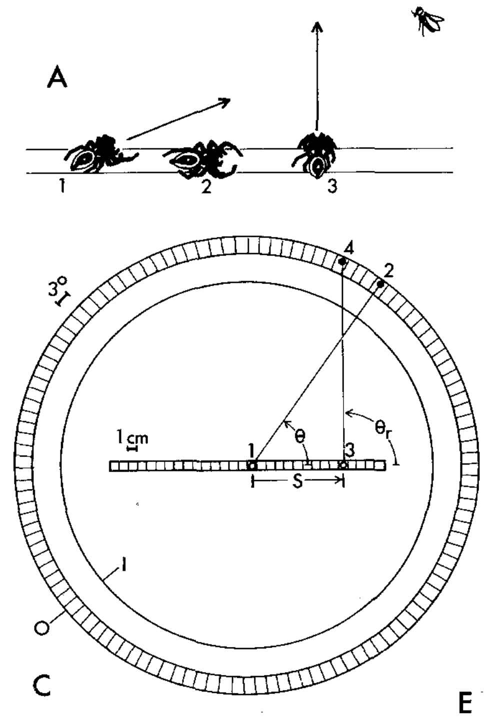


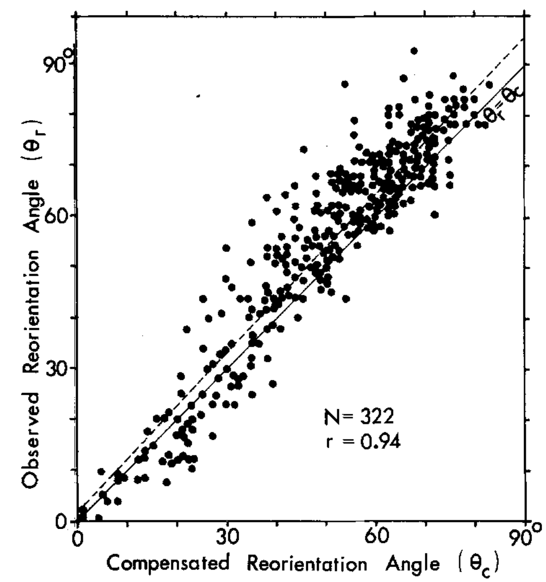
(Bair & Olshausen, 1991)

Orientation by Jumping Spiders During the Pursuit of Prey

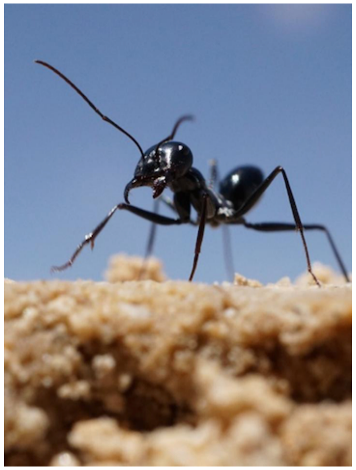
(D.E. Hill, 1979)

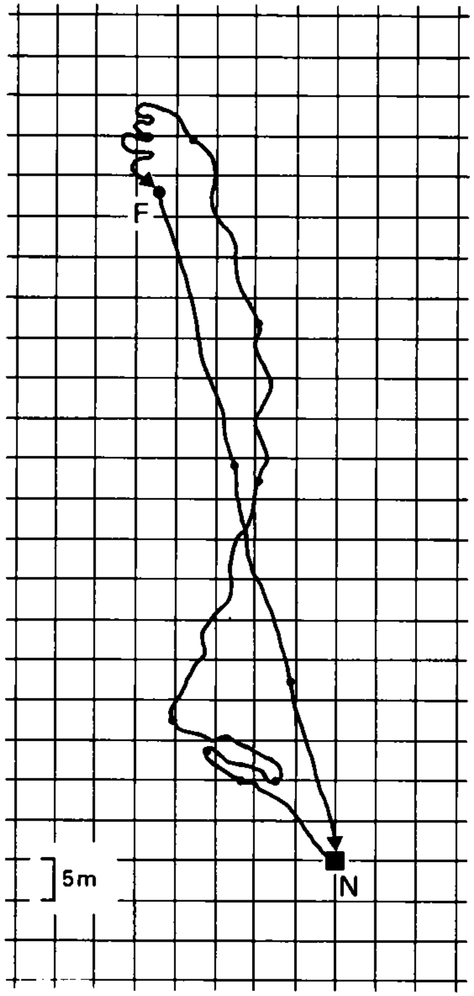






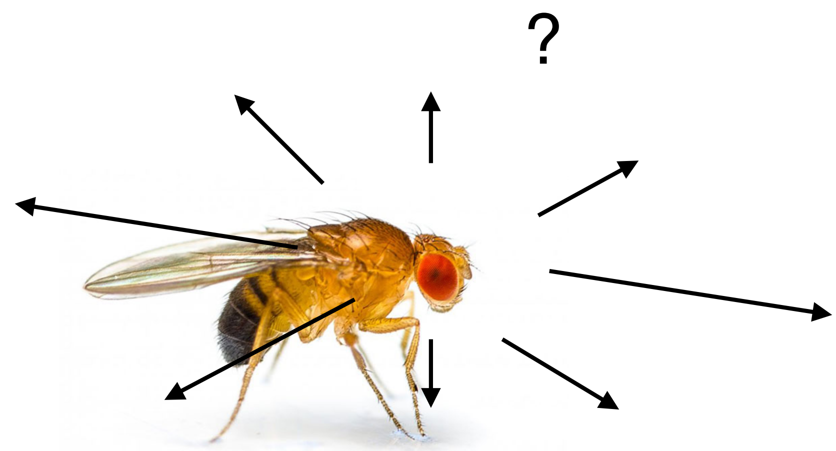
Path integration in desert ants

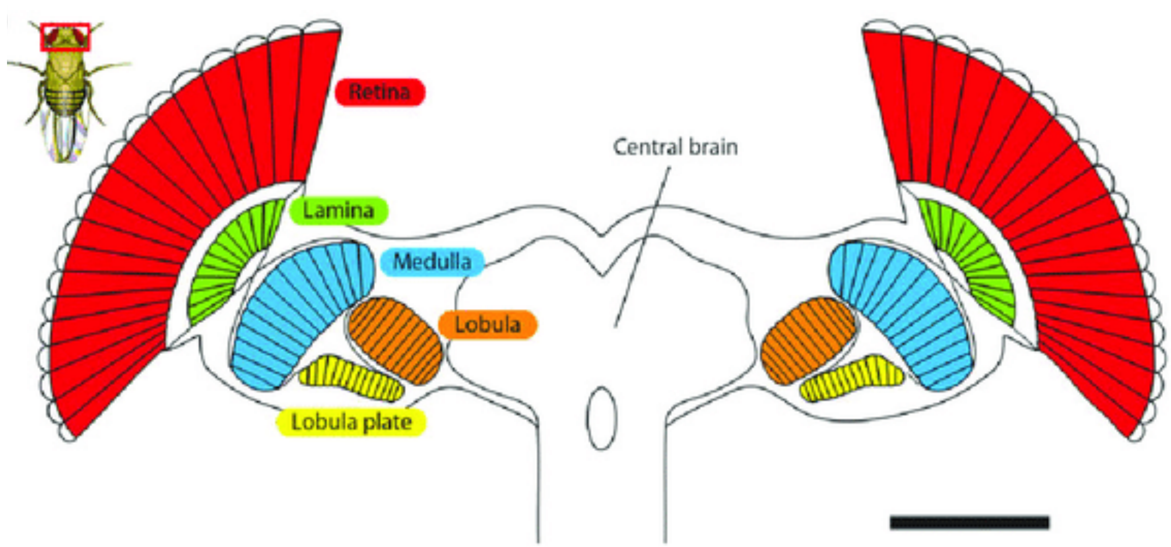




(R. Wehner, S. Wehner, 1986)

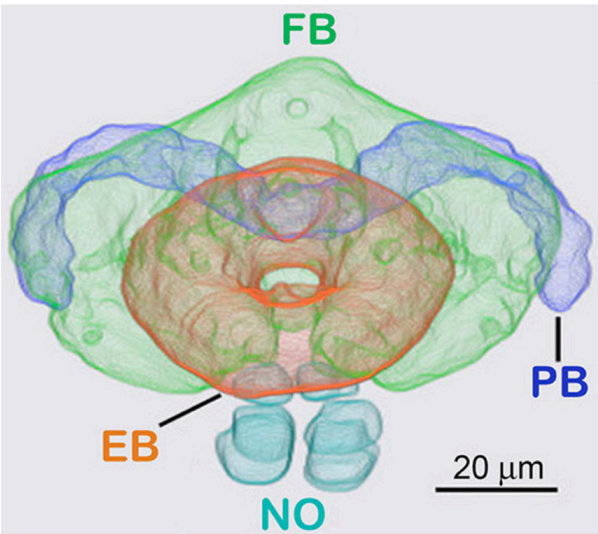
Navigation in fruit flies

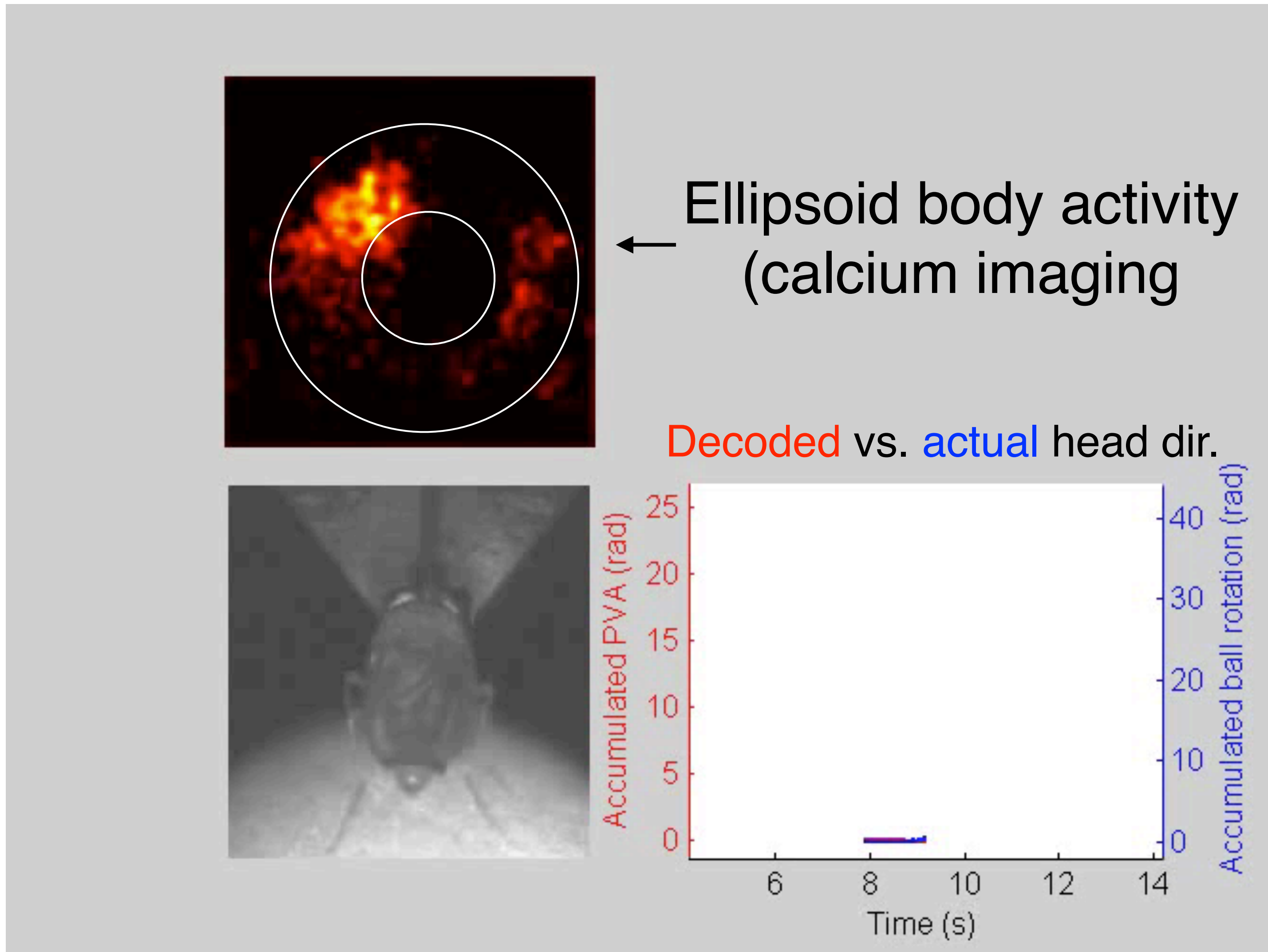




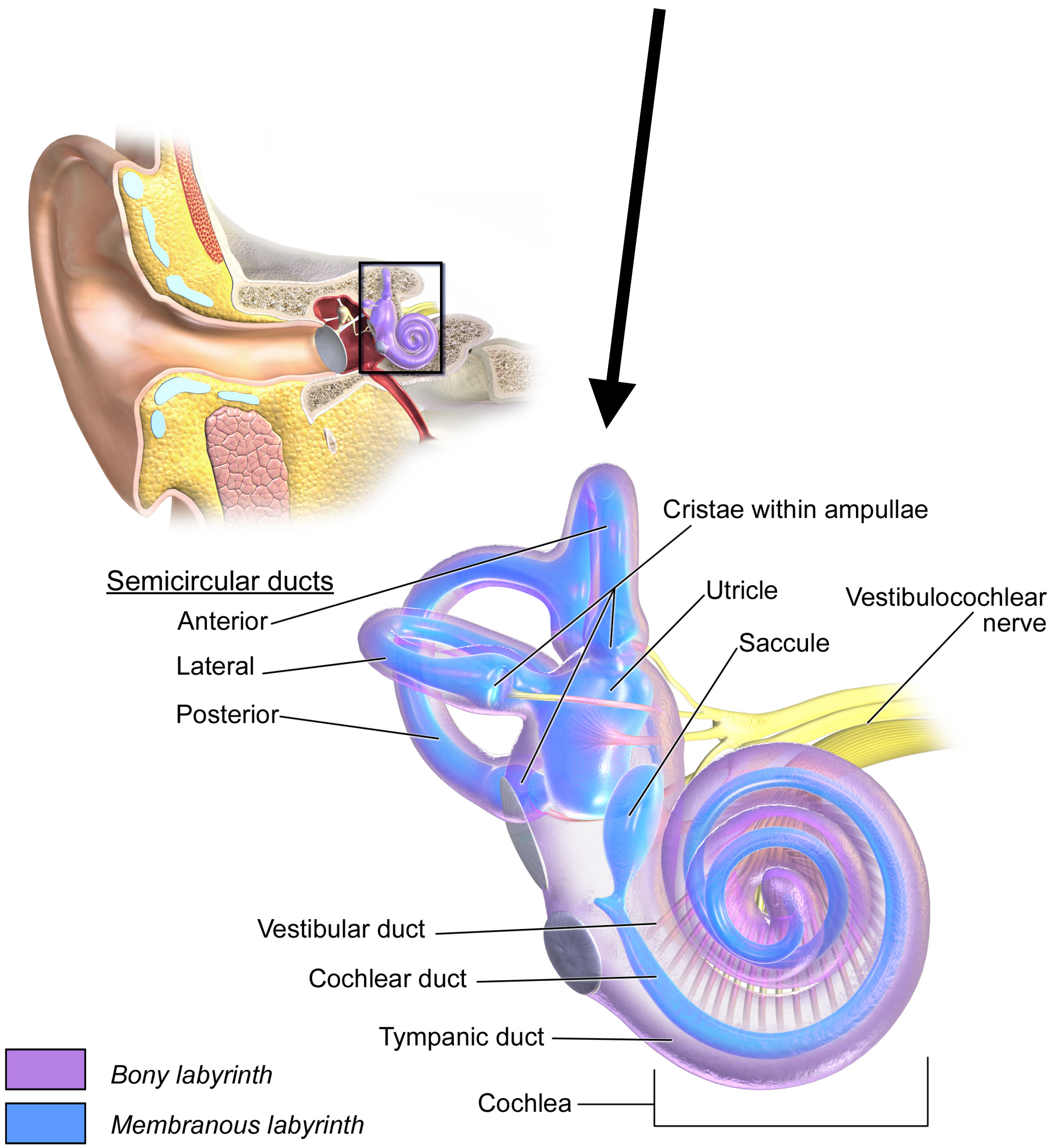
Head-direction cells in ellipsoid body of Drosophila

(Seelig & Jayaraman 2015)





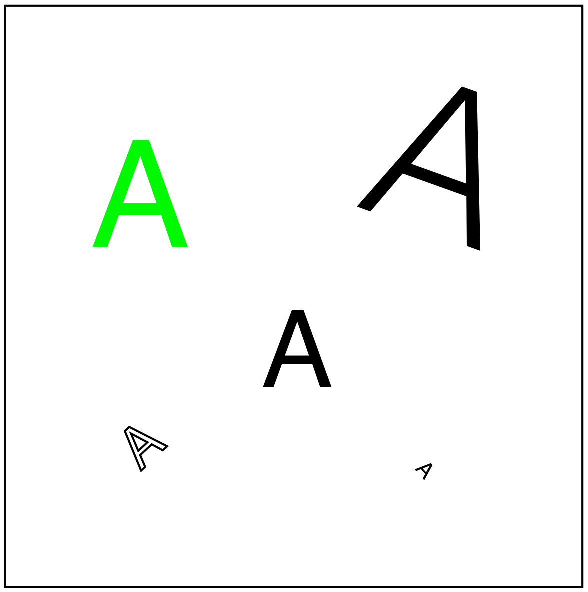
semicircular canals

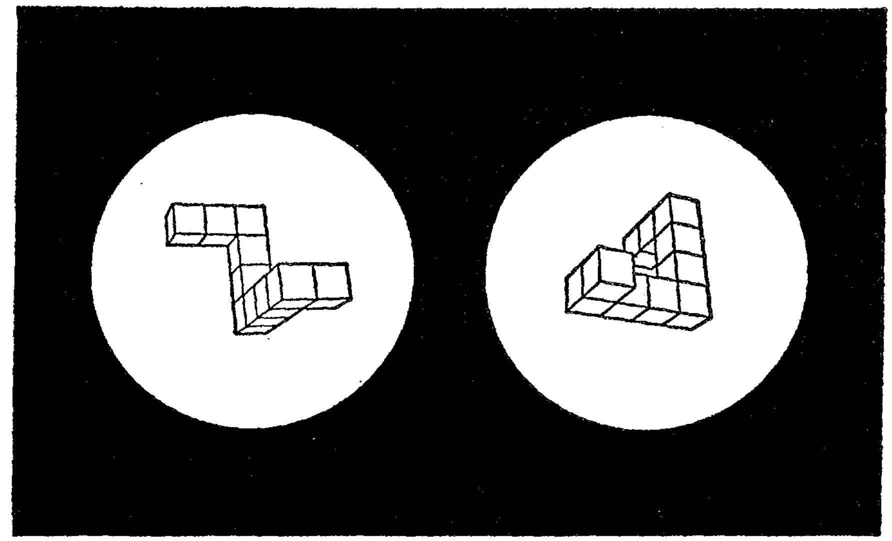


Perception of 3D shape from motion

Randomized dot motion

Our ability to see these as the same stems from our ability to infer the *transformation* between them.





Shepard & Metzler (1971)

How to compute transformations?

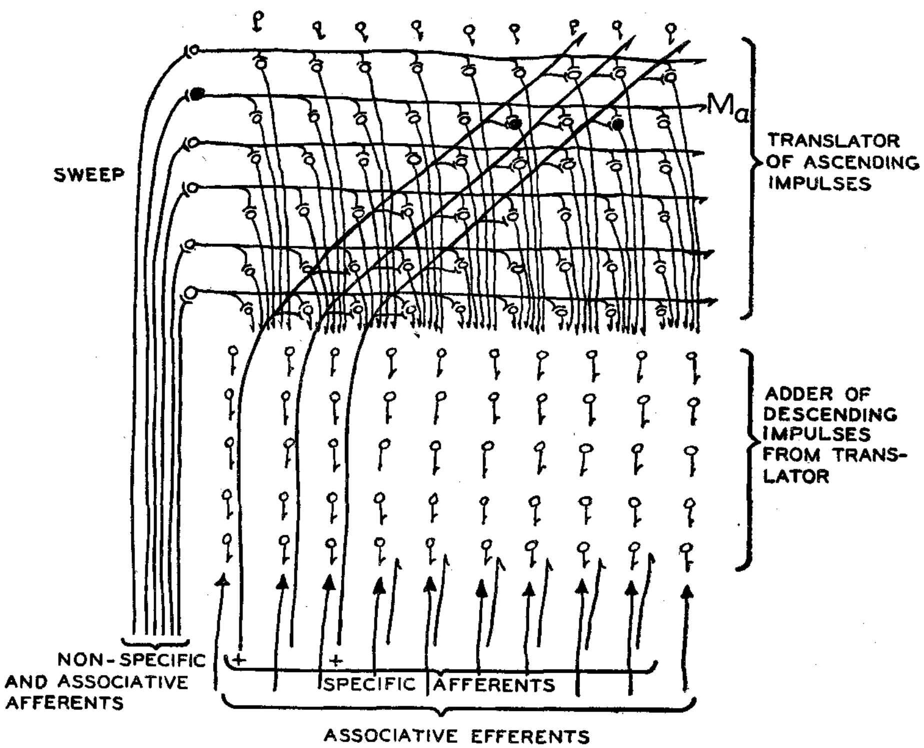
How does the brain do it?

Remapping via multiplicative gating

BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 9, 1947

HOW WE KNOW UNIVERSALS
THE PERCEPTION OF AUDITORY AND VISUAL FORMS

WALTER PITTS
JOHN SIMON GUGGENHEIM FELLOW FOR 1947
AND
WARREN S. MCCULLOCH
DEPARTMENT OF PSYCHIATRY, UNIVERSITY OF ILLINOIS COLLEGE OF MEDICINE AT THE ILLINOIS NEUROPSYCHIATRIC INSTITUTE, CHICAGO

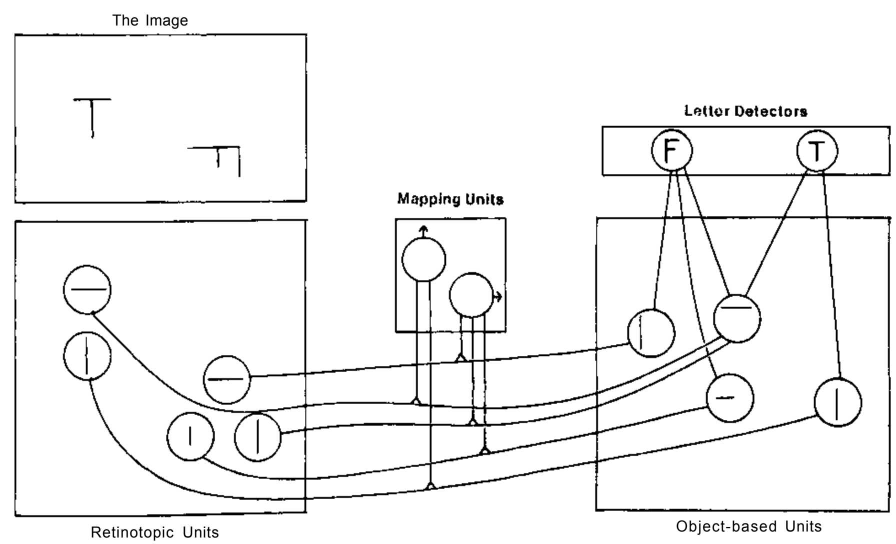


International Joint Conference on
Artificial Intelligence 1985

SHAPE RECOGNITION AND ILLUSORY CONJUNCTIONS

Geoffrey E. Hinton and Kevin J. Lang

Computer Science Department
Carnegie-Mellon University
Pittsburgh PA 15213



Bilinear models for factorizing 'form' and 'motion'

$$I(x) = \sum_{x'} T(x,x')\, I_0(x')$$

$$T(x,x') = \sum_k c_k\, \Psi_k(x,x')$$ transformation

$$I_0(x) = \sum_i a_i\, \phi_i(x)$$ shape

$$= \sum_{x'} \sum_k c_k \Psi_k(x,x') \sum_i a_i\, \phi_i(x')$$

$$= \sum_{i,k} a_i\, c_k\, B_{ik}(x)$$

$$B_{ik}(x) = \sum_{x'} \Psi_k(x,x')\, \phi_i(x')$$

shape (a_i) transformation (c_k)

Pitts & McCulloch (1947) - neural remapping circuits

Hinton (1981; 1985; 2011; 2017) - remapping frames of reference

Anderson & Van Essen (1987) - 'shifter circuits'

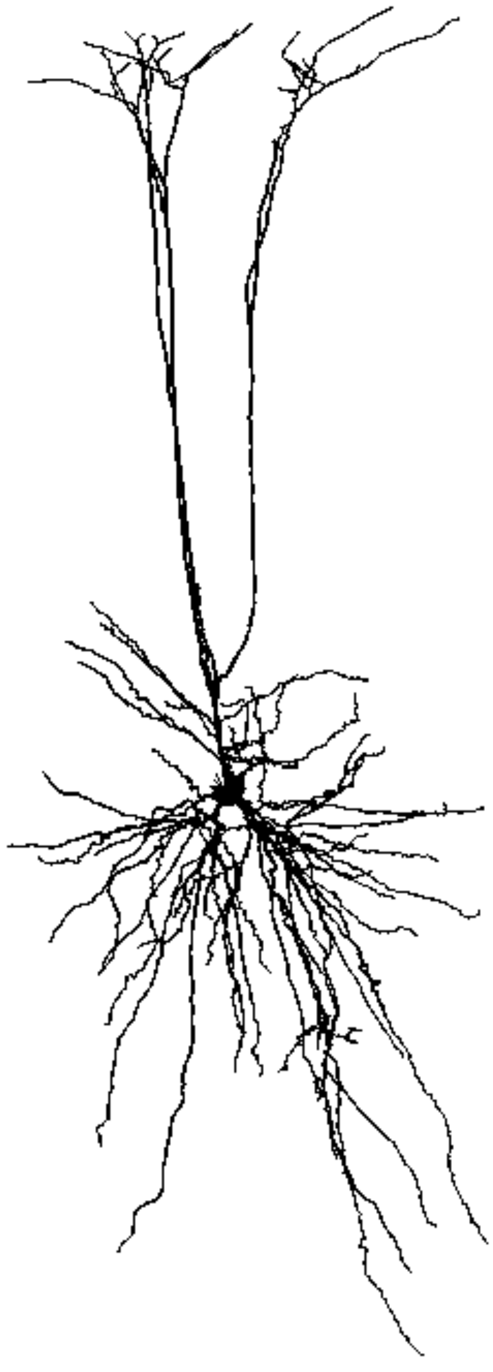
Olshausen, Anderson & Van Essen (1993) - dynamic routing

Tenenbaum & Freeman (2000) - separation of content and style

Arathorn (2002) - Map seeking circuits

Grimes & Rao (2005) - bilinear sparse coding

Memisevic & Hinton (2010) - higher-order Boltzmann machines



$$\sim \quad g\left(\sum_{i} w_i \, \Pi_{j \in G_i} \, x_j\right)$$

Lie groups for modeling *continuous* transformations

$$\mathbf{I}_s = \mathbf{T}(s)\,\mathbf{I}_0$$
$$= e^{\mathbf{A}s}\,\mathbf{I}_0$$

Zhang (1996) - head direction cells

Rao & Ruderman (1999) - learning translation and rotation

Miao & Rao (2007) - learning multiple transformations

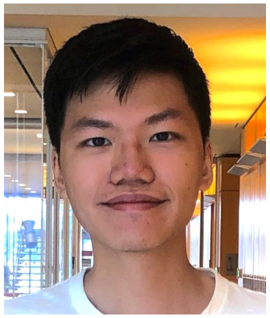
Sohl-Dickstein, Wang & Olshausen (2010) - learned from natural movies

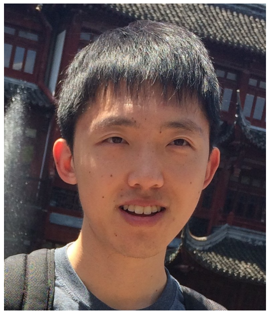
Culpepper & Olshausen (2010) - manifold transport operators

Cohen & Welling (2014) - posterior inference

Gklezakos & Rao (2017) - transformational sparse coding

Connor & Rozell (2023) - learning 3D transformations from 2D projections





Ho Yin Chau Yubei Chen Frank Qiu

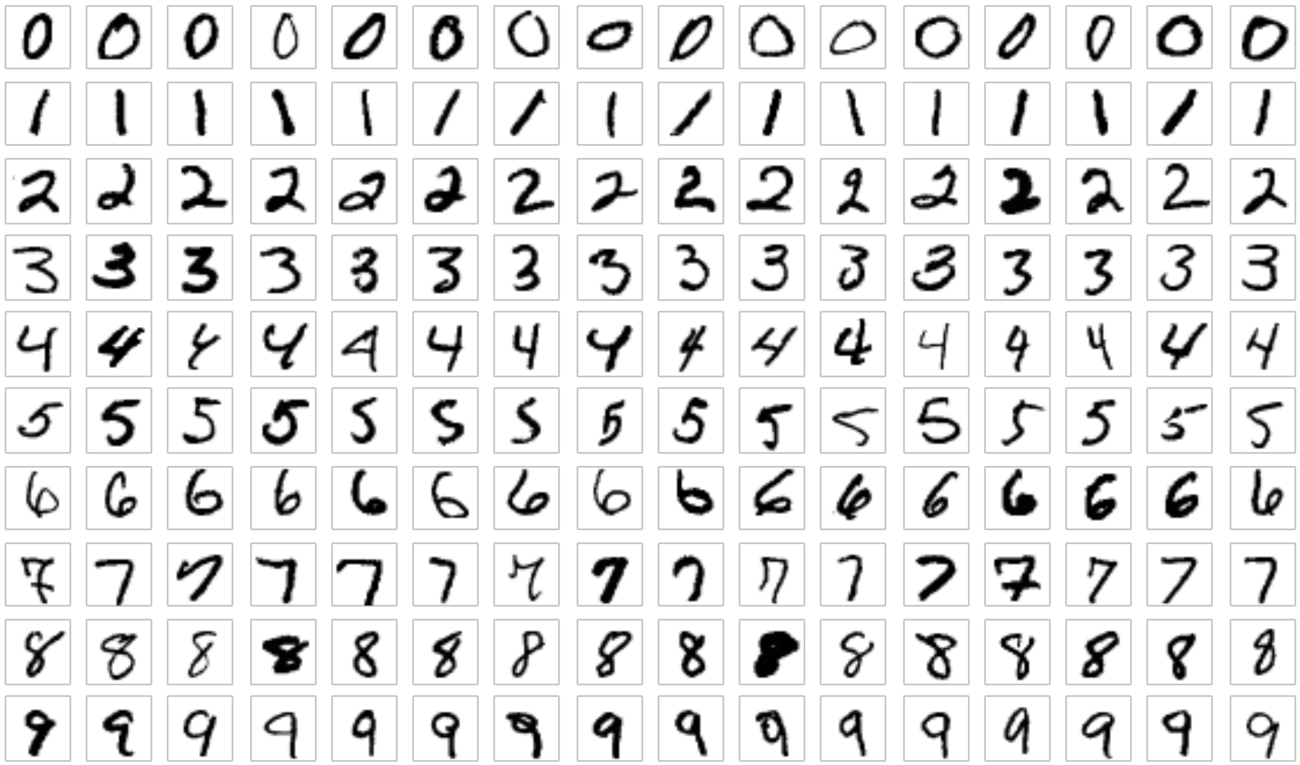
Disentangling Images with Lie Group Transformations and Sparse Coding.
NeurReps Workshop Proceedings, *NeurIPS 2022*.



Sophia Sanborn Christian Shewmake Chris Hillar

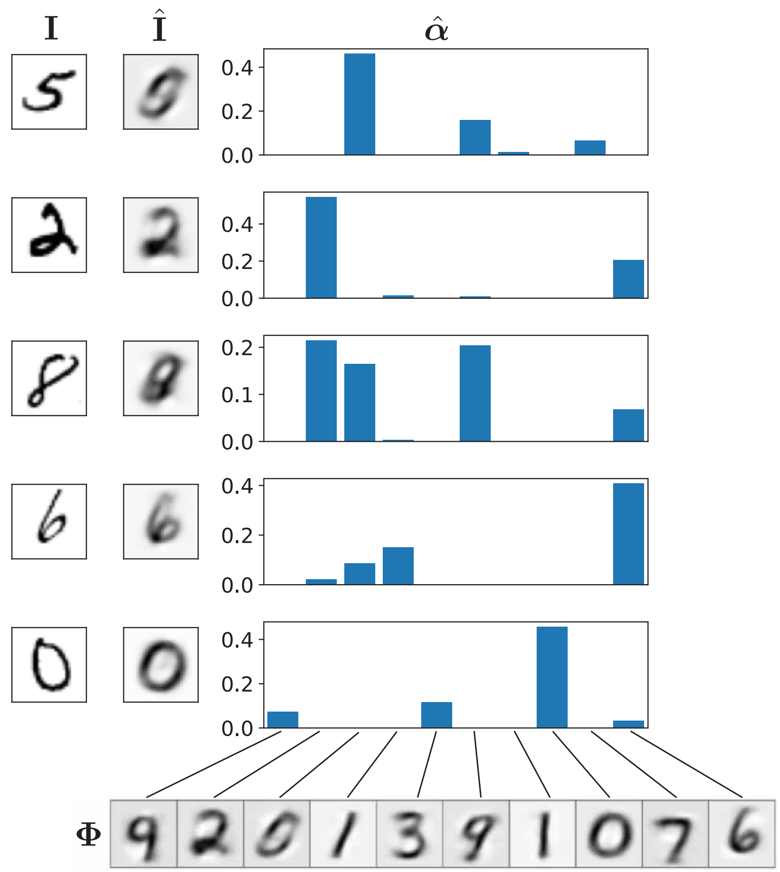
Bispectral Neural Networks. *ICLR 2023*

MNIST dataset



Sparse coding model trained on MNIST (dictionary size = 10)

$$\mathbf{I} = \mathbf{\Phi}\,\alpha + \epsilon$$



Learned $\mathbf{\Phi}$

Factorizing images with Lie group transformations and sparse coding

(Ho Yin Chau, Yubei Chen, Frank Qiu)

$$\mathbf{I} = \mathbf{T}(s)\,\mathbf{\Phi}\,\alpha + \epsilon$$
$$= e^{\mathbf{A}s}\,\mathbf{\Phi}\,\alpha + \epsilon$$

$$\begin{aligned} \mathbf{T}(s) &= e^{\mathbf{A}s} \\ &= \mathbf{W}\,e^{\mathbf{\Sigma}s}\,\mathbf{W}^T = \mathbf{W}\,\mathbf{R}(s)\,\mathbf{W}^T \end{aligned}$$

$$\mathbf{\Sigma} = \begin{bmatrix} 0 & -\omega_1 & & & \\ \omega_1 & 0 & & & \\ & & \ddots & & \\ & & & 0 & -\omega_{D/2} \\ & & & \omega_{D/2} & 0 \end{bmatrix} \qquad \mathbf{R}(s) = \begin{bmatrix} \cos(\omega_1 s) & -\sin(\omega_1 s) & & & \\ \sin(\omega_1 s) & \cos(\omega_1 s) & & & \\ & & \ddots & & \\ & & & \cos(\omega_{D/2} s) & -\sin(\omega_{D/2} s) \\ & & & \sin(\omega_{D/2} s) & \cos(\omega_{D/2} s) \end{bmatrix}$$

$$\boxed{\mathbf{I} = \mathbf{W}\,\mathbf{R}(s)\,\mathbf{W}^T\,\mathbf{\Phi}\,\alpha + \epsilon}$$

Learning

$$\nabla_{\boldsymbol{\theta}} \ln P_{\boldsymbol{\theta}}(\mathbf{I}) \approx \mathbb{E}_{\mathbf{s} \sim P_{\boldsymbol{\theta}}(\mathbf{s}|\mathbf{I}, \hat{\boldsymbol{\alpha}})} [\nabla_{\boldsymbol{\theta}} \ln P_{\boldsymbol{\theta}}(\mathbf{I}|\mathbf{s}, \hat{\boldsymbol{\alpha}})]$$

$$\hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} P_{\boldsymbol{\theta}}(\boldsymbol{\alpha}|\mathbf{I})$$

Inference

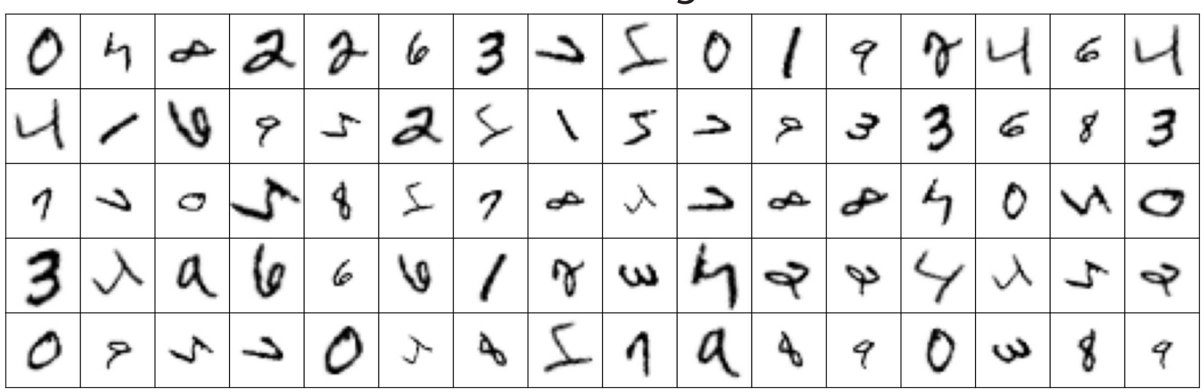
$$\hat{\alpha} = \arg\max_{\alpha} P_{\theta}(\alpha|\mathbf{I})$$

$$= \arg\max_{\alpha} [\langle \ln P_{\theta}(\mathbf{I} \,|\, \mathbf{s}, \alpha) \rangle_{q(\mathbf{s})} + \ln P_{\theta}(\alpha)]$$

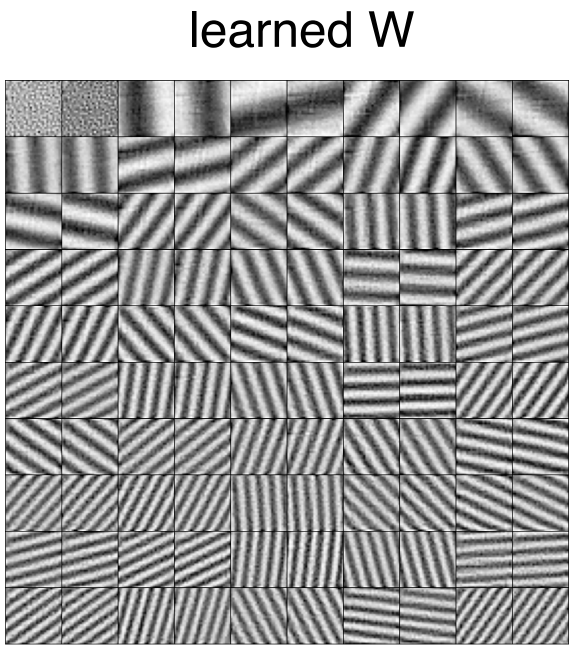
$$q(\mathbf{s}) \leftarrow P_{\theta}(\mathbf{s}|\mathbf{I}, \hat{\alpha})$$

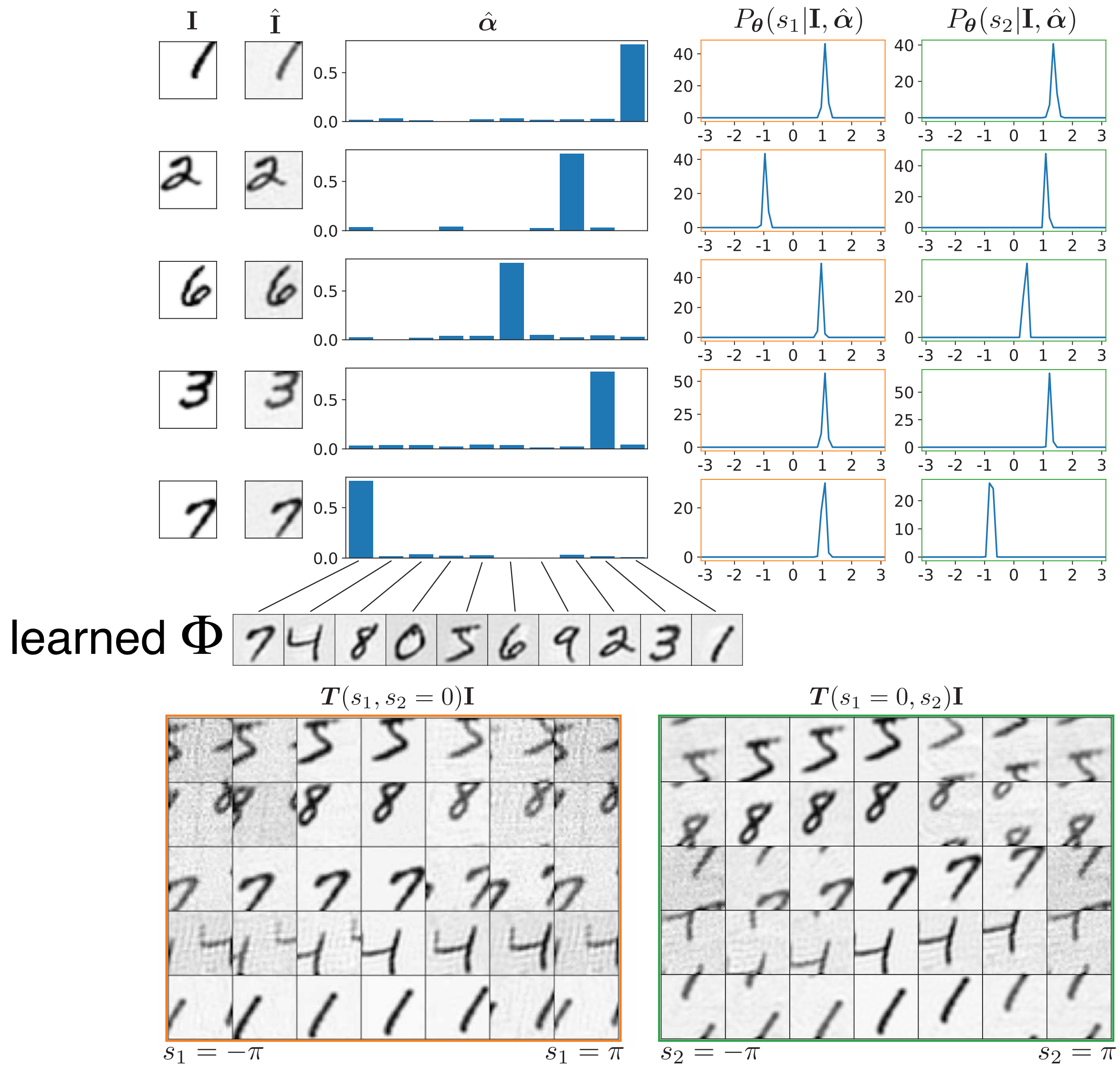
2D Translation Dataset

Rotation + Scaling Dataset



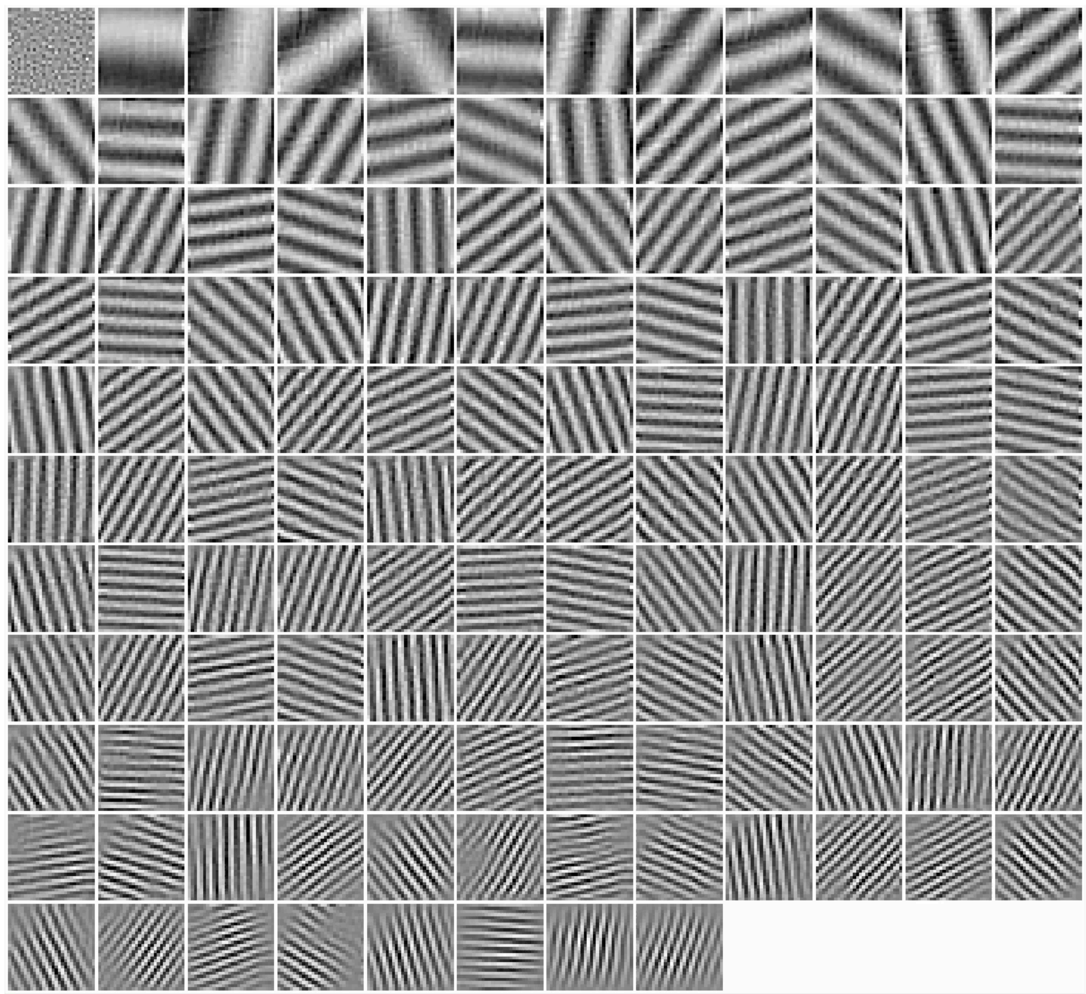
Results: 2D translation





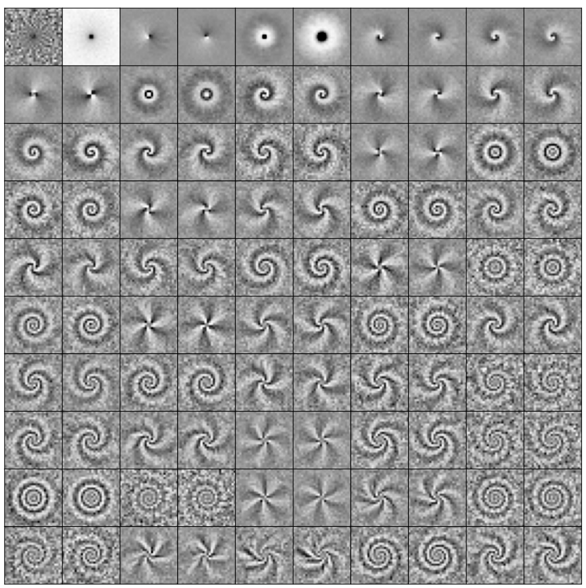
Results: 2D translation

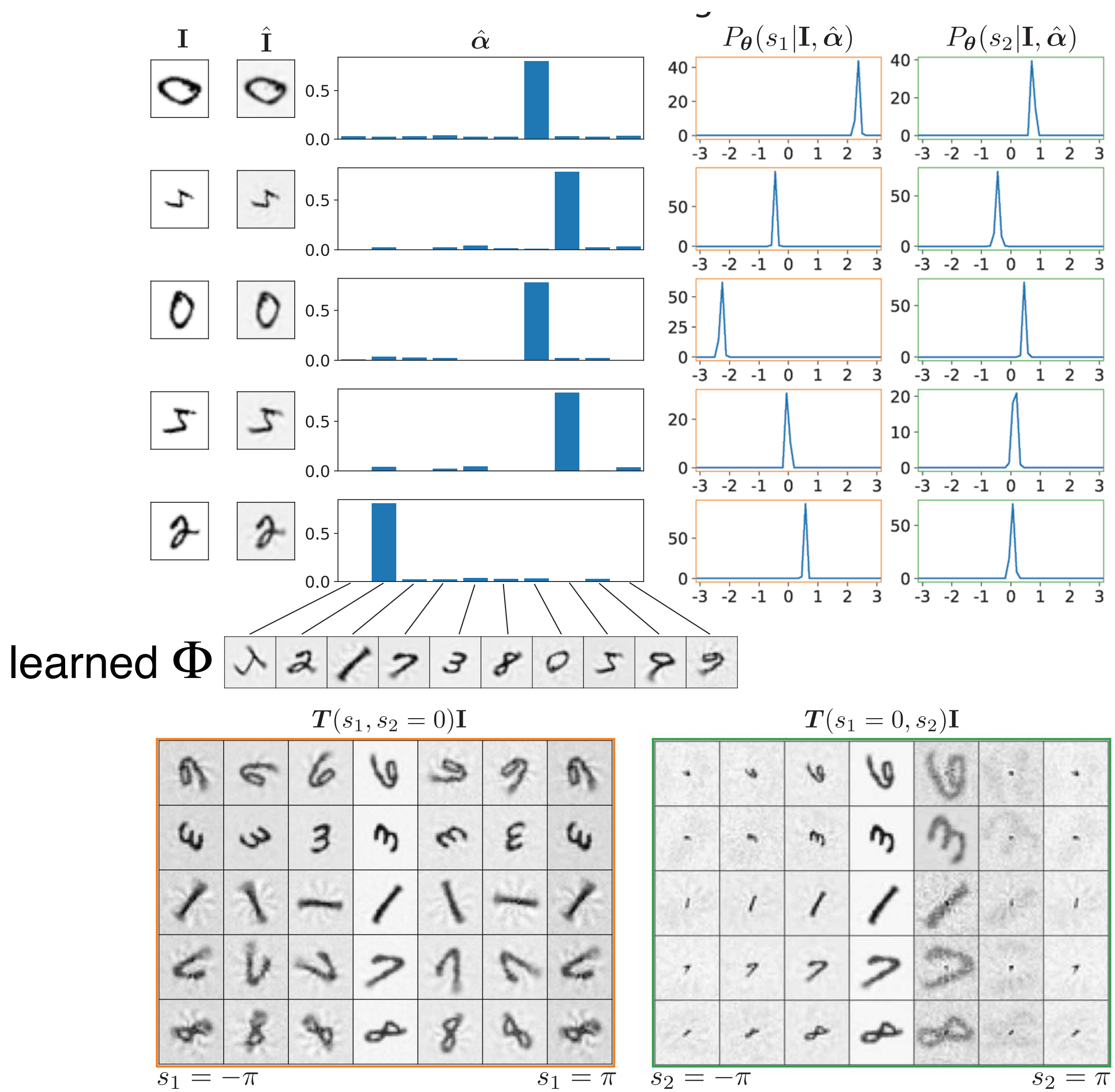
learned W



Results: rotation and scale

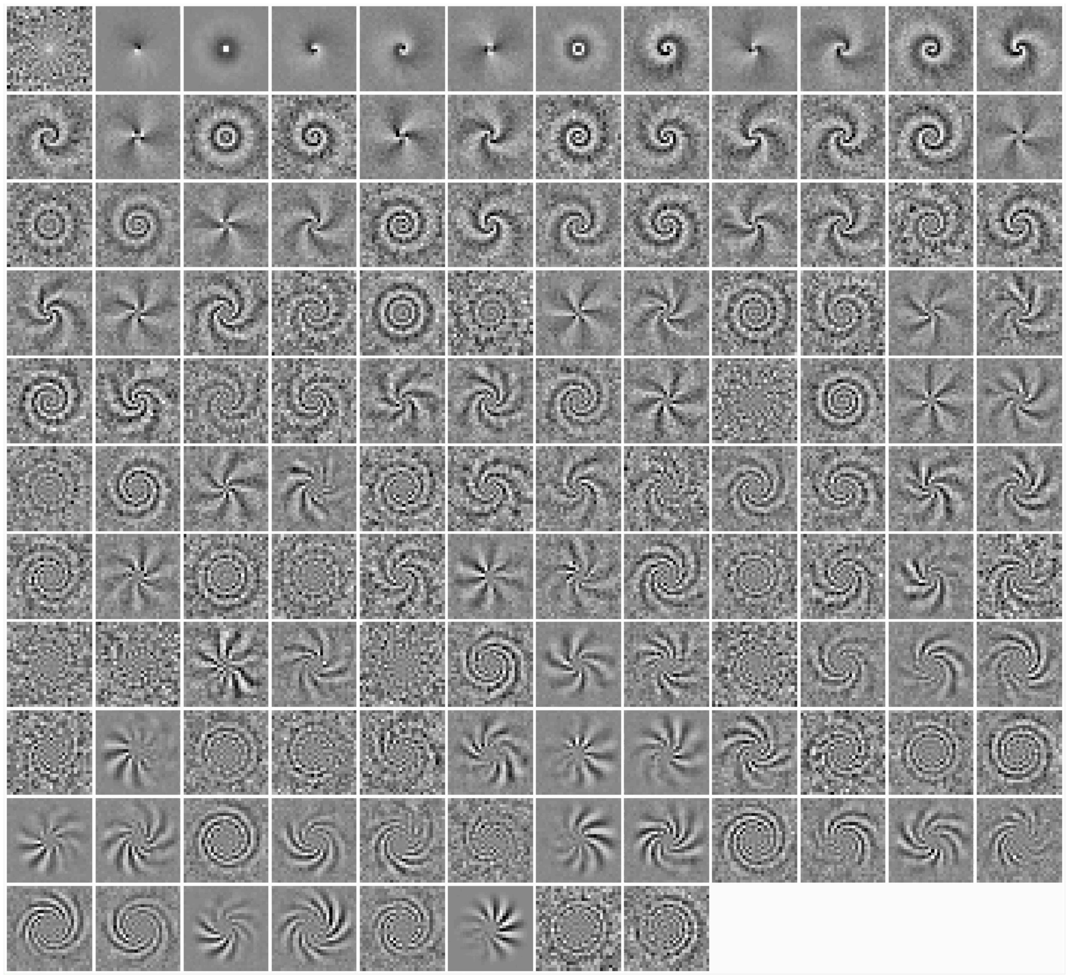
learned W





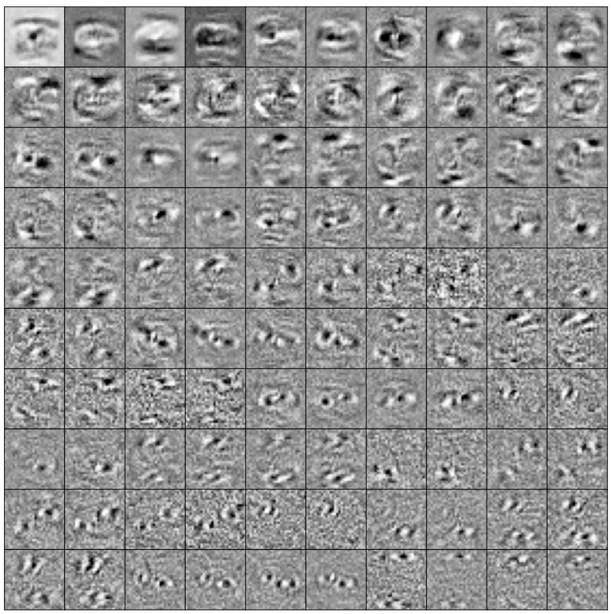
Results: rotation and scale

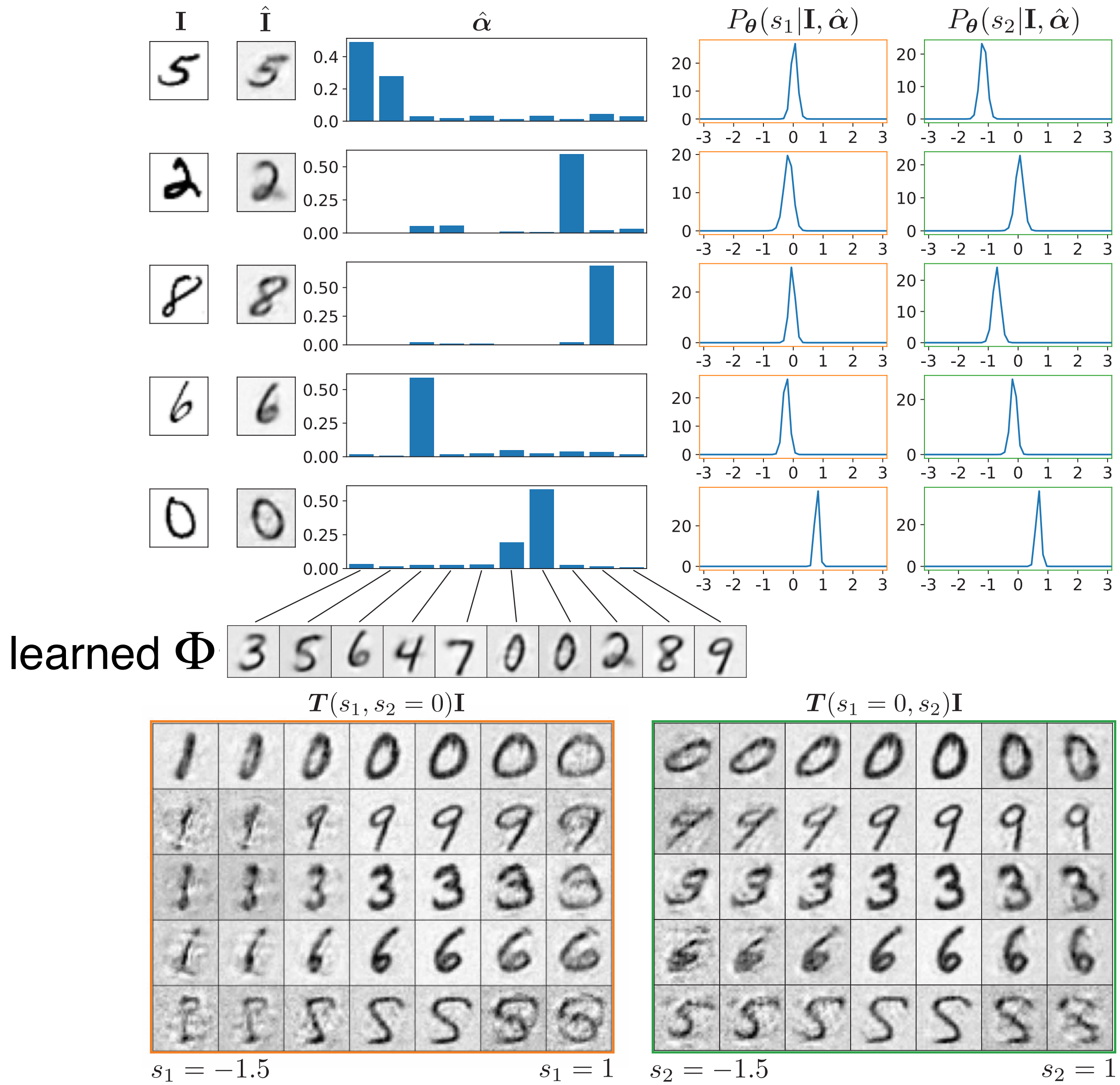
learned W



Results: full MNIST

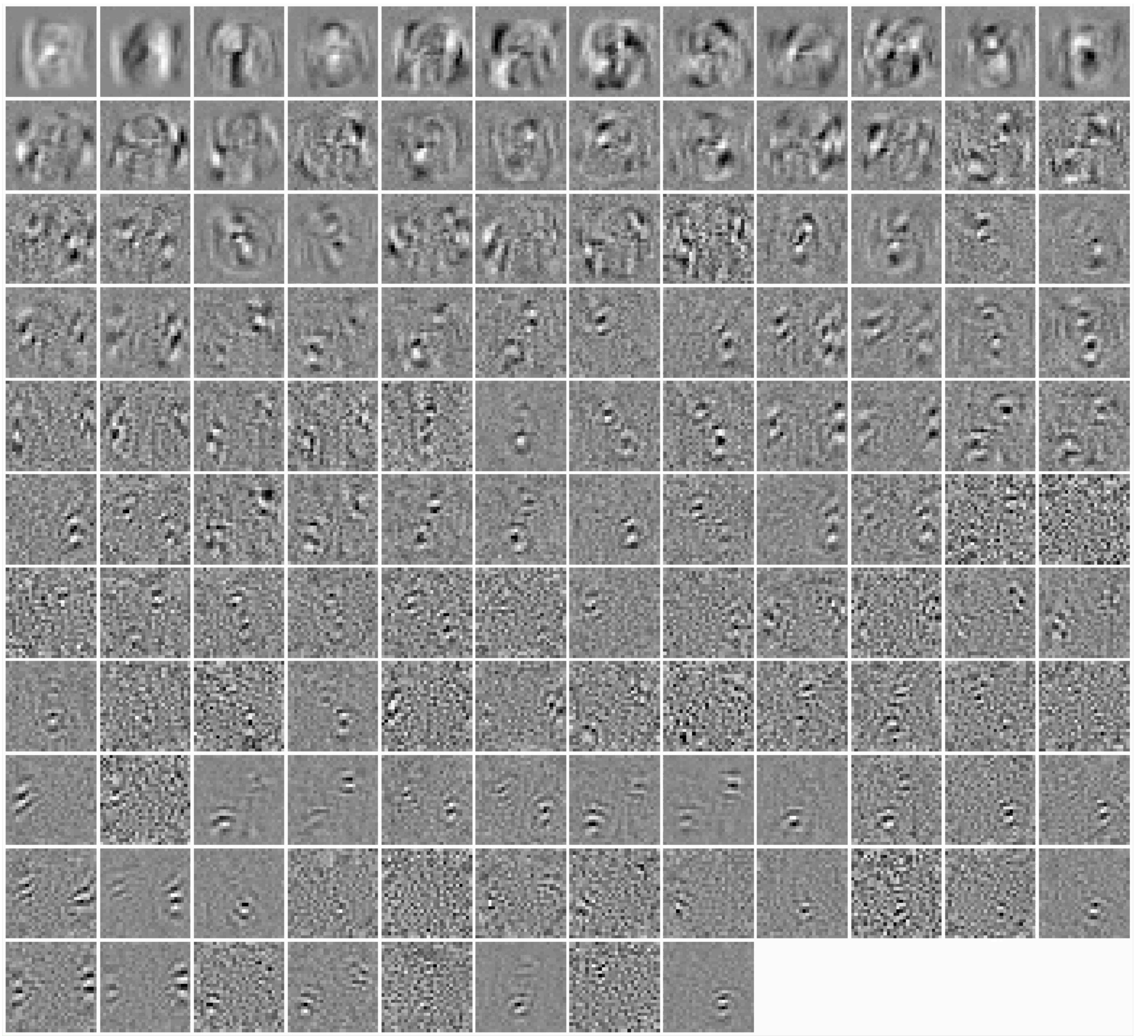
learned W





Results: full MNIST

learned W



The bispectrum

Fourier transform

$$\mathscr{F}\{f(x)\} \equiv \int f(x)\, e^{-j\omega x} dx$$

$$f(x) \longleftrightarrow \tilde{f}(\omega) = |\tilde{f}(\omega)|\, e^{j\,\phi(\omega)}$$

Power spectrum

$$C(\Delta x) = \langle f(x)\, f(x-\Delta x)\rangle_x \overset{\mathscr{F}}{\longleftrightarrow} |\tilde{f}(\omega)|^2 = \tilde{f}(\omega)\tilde{f}^*(\omega)$$

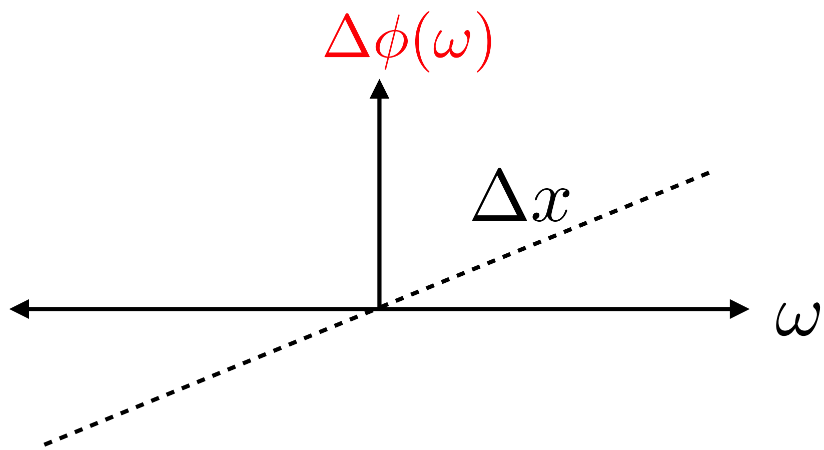
Bispectrum

$$C(\Delta x_1, \Delta x_2) = \langle f(x)\, f(x-\Delta x_1)\, f(x-\Delta x_2)\rangle_x \overset{\mathscr{F}}{\longleftrightarrow} B(\omega_1, \omega_2) = \tilde{f}(\omega_1)\tilde{f}(\omega_2)\tilde{f}^*(\omega_1+\omega_2)$$

Fourier shift theorem

$$f(x - \Delta x)$$

$$\mathscr{F}\{f(x - \Delta x)\} = e^{-j\overbrace{\omega \Delta x}^{\Delta\phi(\omega)}}\,\tilde{f}(\omega)$$



Power spectrum is invariant to shift (but excessively so)

Power spectrum

$$\begin{aligned}\tilde{f}(\omega)\tilde{f}^*(\omega) &= |\tilde{f}(\omega)|e^{j\phi(\omega)}\,|\tilde{f}(\omega)|e^{-j\phi(\omega)} \\ &= |\tilde{f}(\omega)|^2\end{aligned}$$

Power spectrum of shifted pattern

$$\begin{aligned}e^{-j\omega\Delta x}\tilde{f}(\omega)e^{j\omega\Delta x}\tilde{f}^*(\omega) &= |\tilde{f}(\omega)|e^{j(\phi(\omega)-\omega\Delta x)}\,|\tilde{f}(\omega)|e^{-j(\phi(\omega)-\omega\Delta x)} \\ &= |\tilde{f}(\omega)|^2\end{aligned}$$

The Importance of Phase in Signals

ALAN V. OPPENHEIM, FELLOW, IEEE, AND JAE S. LIM, MEMBER, IEEE

Invited Paper

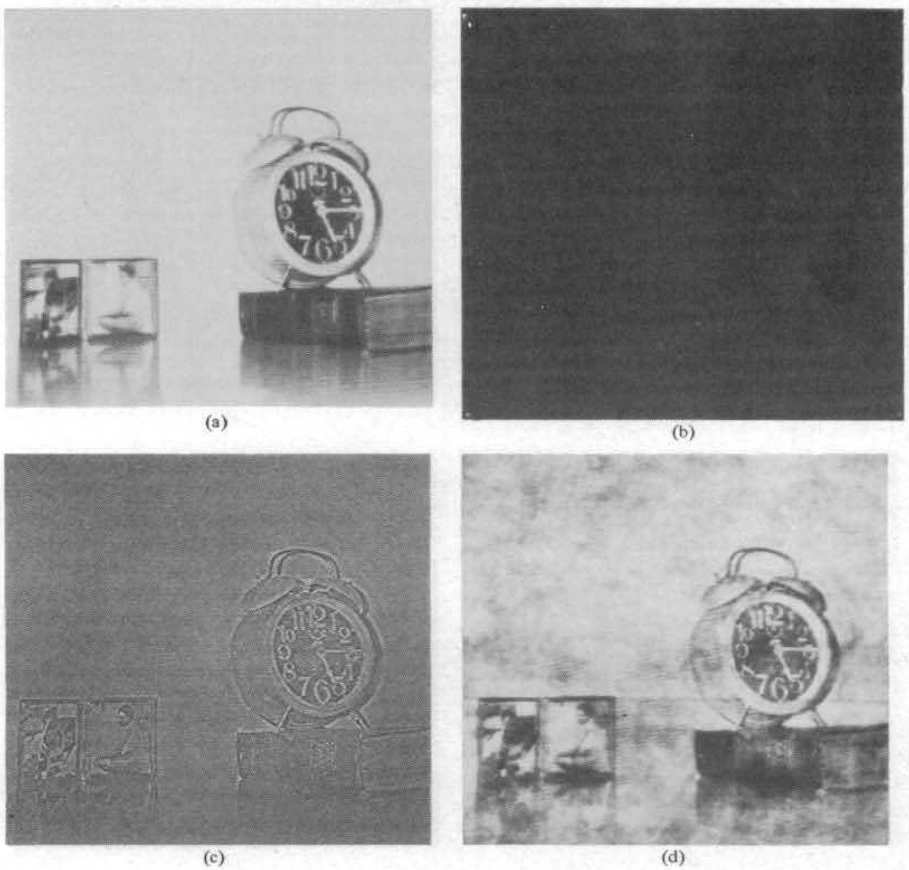


Fig. 2. (a) Original image. (b) Image synthesized from the Fourier transform magnitude of (a) and zero phase. (c) Image synthesized from the Fourier transform phase of (a) and unity magnitude. (d) Image synthesized from the Fourier transform phase of (a) and a magnitude averaged over an ensemble of images.

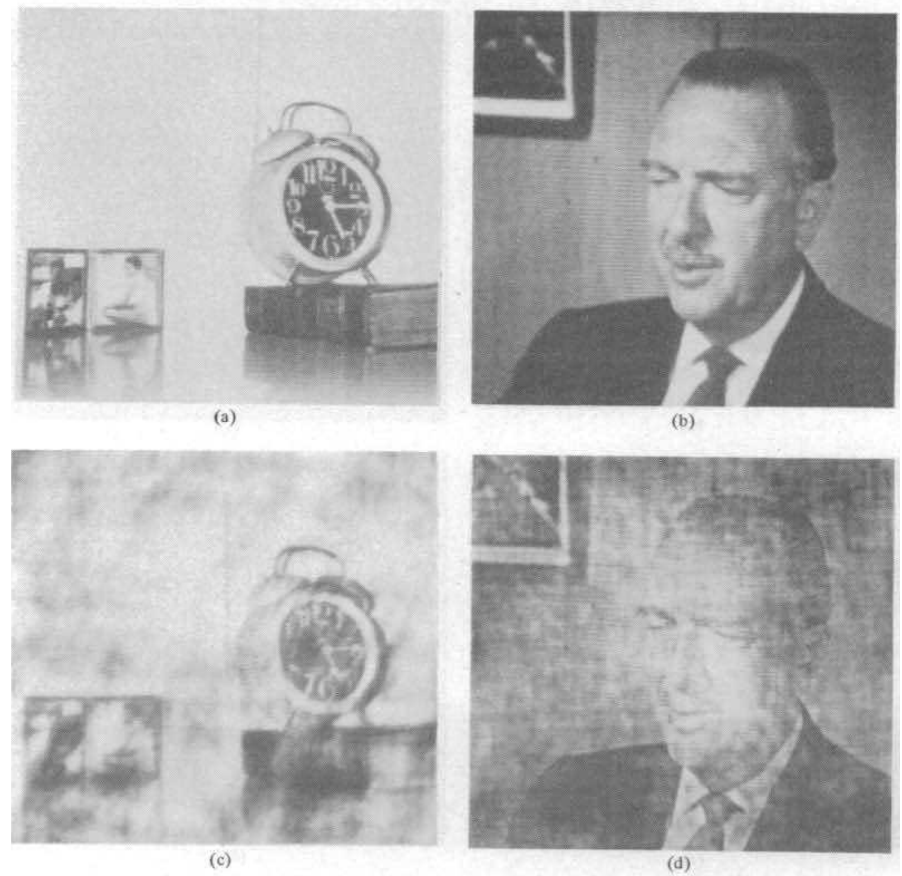


Fig. 3. (a) Original image A. (b) Original image B. (c) Image synthesized from the Fourier transform phase of image A and the magnitude of image B. (d) Image synthesized from the Fourier transform magnitude of image A and the phase of image B.

Bispectrum is invariant to shift (and unique)

Bispectrum

$$
\begin{aligned}
\tilde{f}(\omega_1)\tilde{f}(\omega_2)\tilde{f}^*(\omega_1+\omega_2) &= |\tilde{f}(\omega_1)|e^{j\phi(\omega_1)}\,|\tilde{f}(\omega_2)|e^{j\phi(\omega_2)}\,|\tilde{f}(\omega_1+\omega_2)|e^{j\phi(\omega_1+\omega_2)} \\
&= |\tilde{f}(\omega_1)||\tilde{f}(\omega_2)||\tilde{f}(\omega_1+\omega_2)|e^{j(\phi(\omega_1)+\phi(\omega_2)-\phi(\omega_1+\omega_2))} \\
&= |B(\omega_1,\omega_2)|e^{j(\phi(\omega_1)+\phi(\omega_2)-\phi(\omega_1+\omega_2))} \equiv B(\omega_1,\omega_2)
\end{aligned}
$$

relative phase

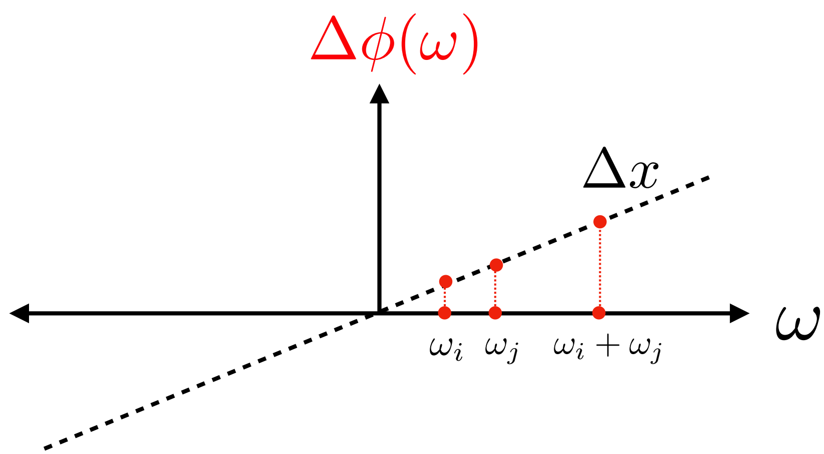
Bispectrum of shifted pattern

$$
\begin{aligned}
&e^{-j\omega_1\Delta x}\tilde{f}(\omega_1)\,e^{-j\omega_2\Delta x}\tilde{f}(\omega_2)\,e^{j(\omega_1+\omega_2)\Delta x}\tilde{f}^*(\omega_1+\omega_2) = \\
&\qquad |B(\omega_1,\omega_2)|\,e^{j(\phi(\omega_1)-\omega_1\Delta x)}\,e^{j(\phi(\omega_2)-\omega_2\Delta x)}\,e^{-j(\phi(\omega_1+\omega_2)-(\omega_1+\omega_2)\Delta x)} \\
&\quad = |B(\omega_1,\omega_2)|\,e^{j(\phi(\omega_1)+\phi(\omega_2)-\phi(\omega_1+\omega_2))}\,e^{j(-\omega_1\Delta x-\omega_2\Delta x)+(\omega_1+\omega_2)\Delta x)} \\
&\quad = B(\omega_1,\omega_2)
\end{aligned}
$$

Fourier shift theorem

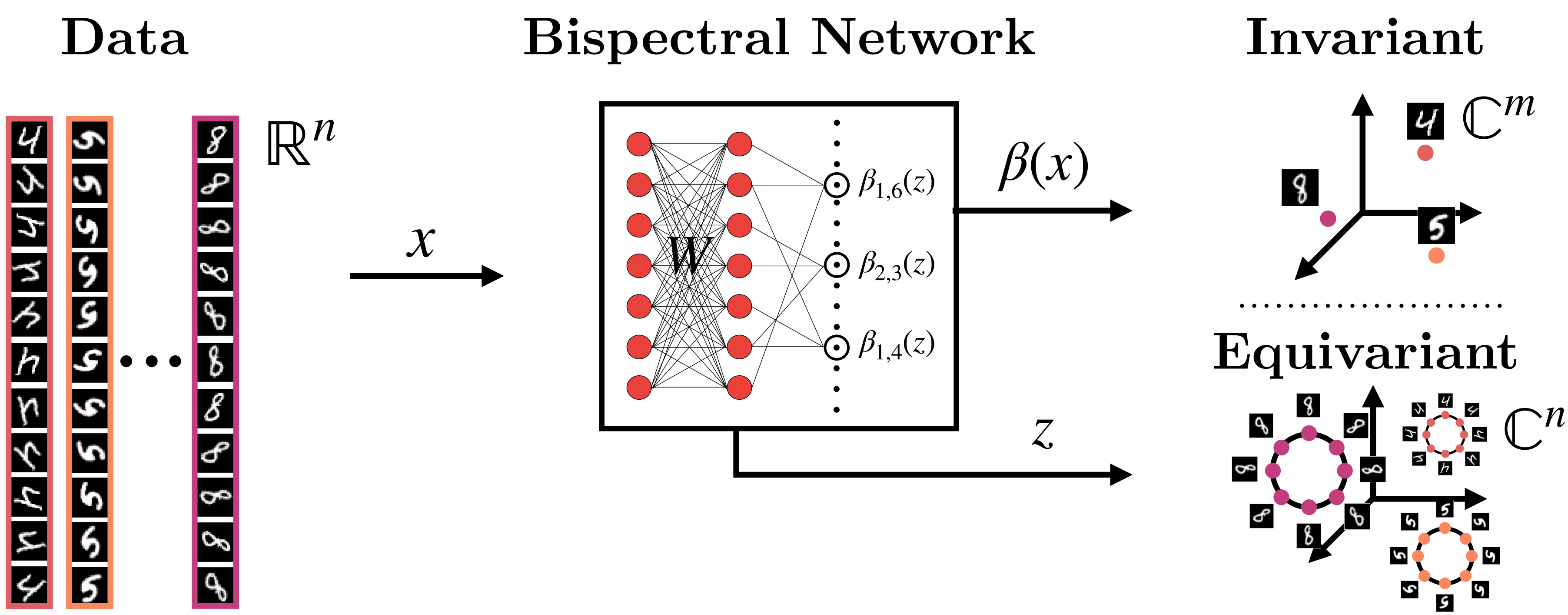
$$f(x - \Delta x)$$

$$\mathscr{F}\{f(x - \Delta x)\} = e^{-j\overbrace{\omega \Delta x}^{\Delta\phi(\omega)}} \tilde{f}(\omega)$$

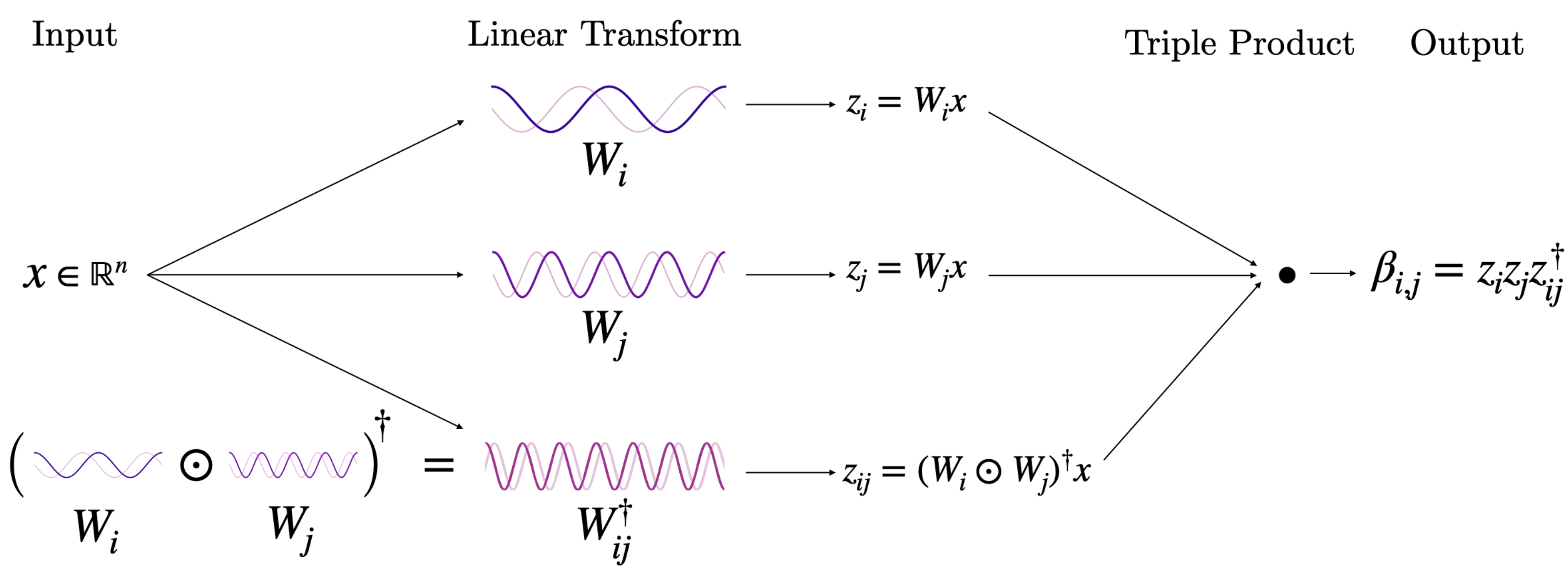


How to *learn* the *group* underlying the bispectrum?

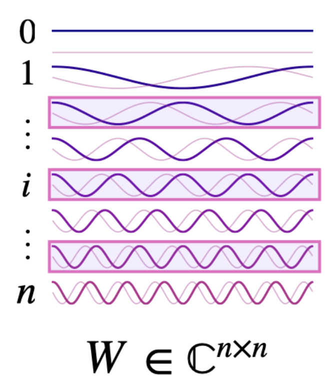
Learning the bispectrum from data



Bispectrum ansatz



Weight Matrix



Orbit separation loss

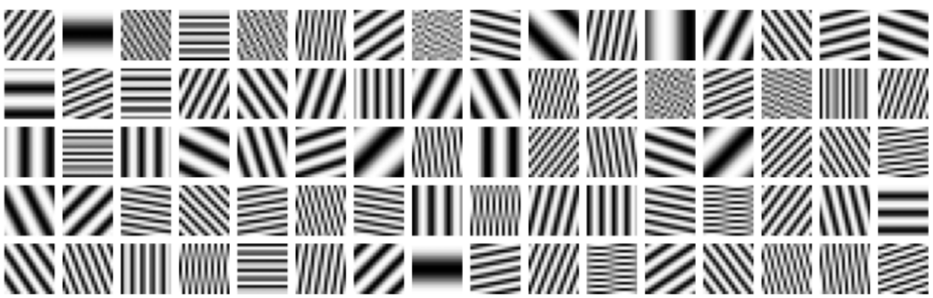
$$L(x_i) = \sum_{j|y_j=y_i} ||\bar{\beta}(x_i) - \bar{\beta}(x_j)||_2 + \gamma||x_i - W^\dagger W\, x_i||_2$$

↑ "be invariant" ↑ "keep W orthonormal"

Learned *W*
(trained on natural image patches)

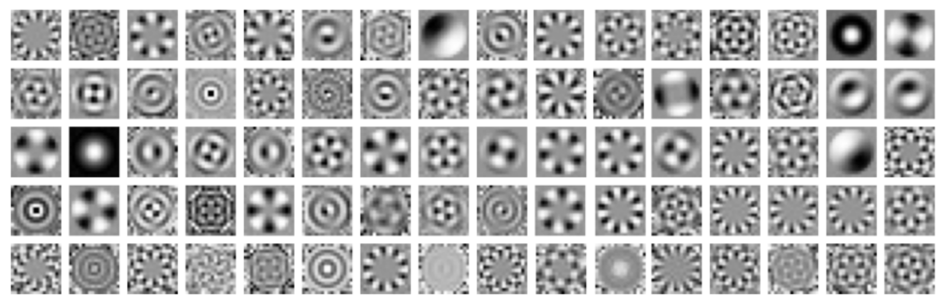
2D cyclic translation

$\mathbb{S}^1 \times \mathbb{S}^1$

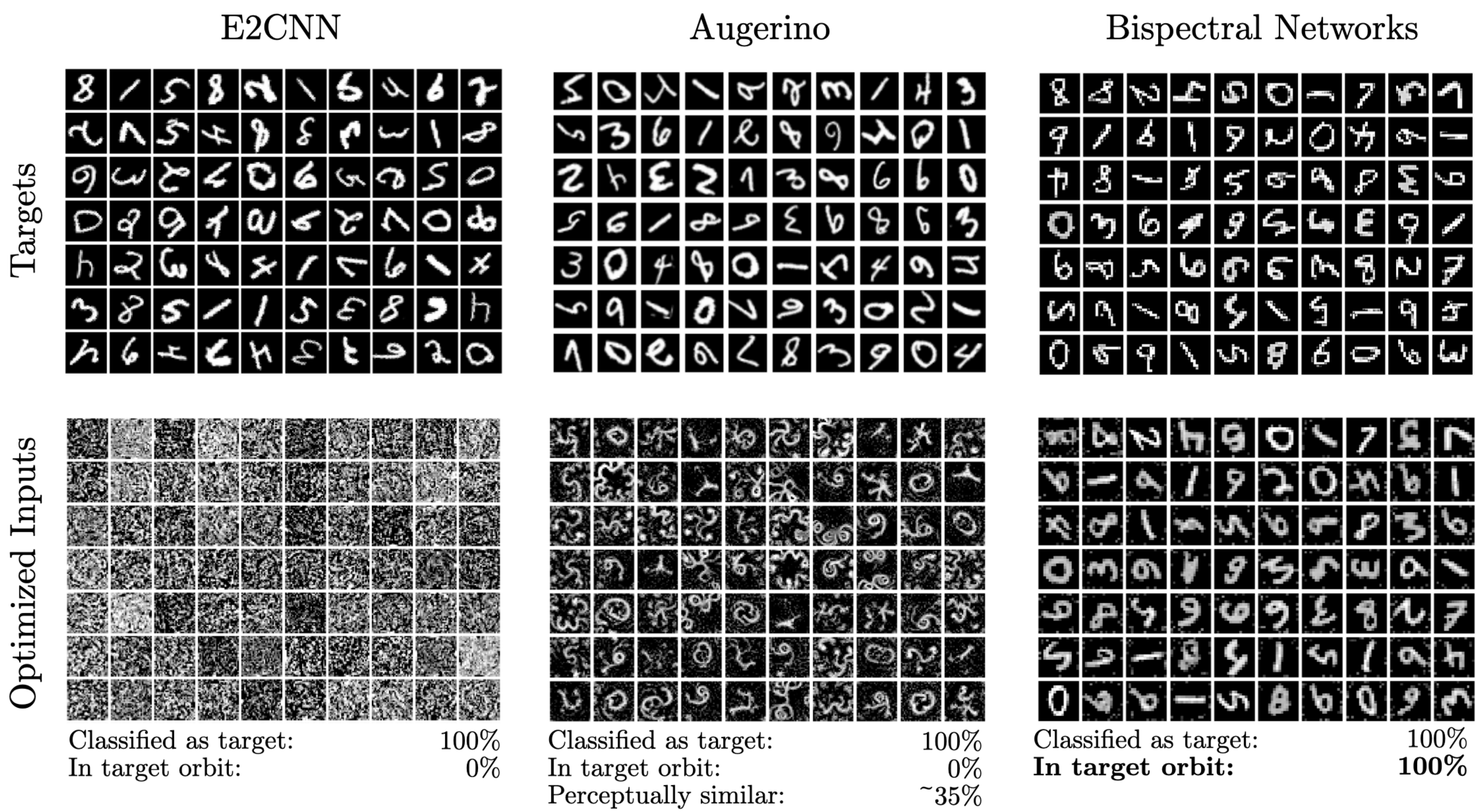


2D rotation

$SO(2)$



Robustness to adversarial perturbation



Main points

- Invariance - the ability to perceive shape independent of pose - evolved from the need to geometrically reason about the environment.
- Computing transformations is fundamental to enabling this.
- Lie groups provide a promising mathematical framework for modeling the neural computations underlying our ability to compute transformations.
- Representations may be learned from data, and could provide a new computational primitive for deep learning.