Scene Statistics
Part 2

Odelia Schwartz

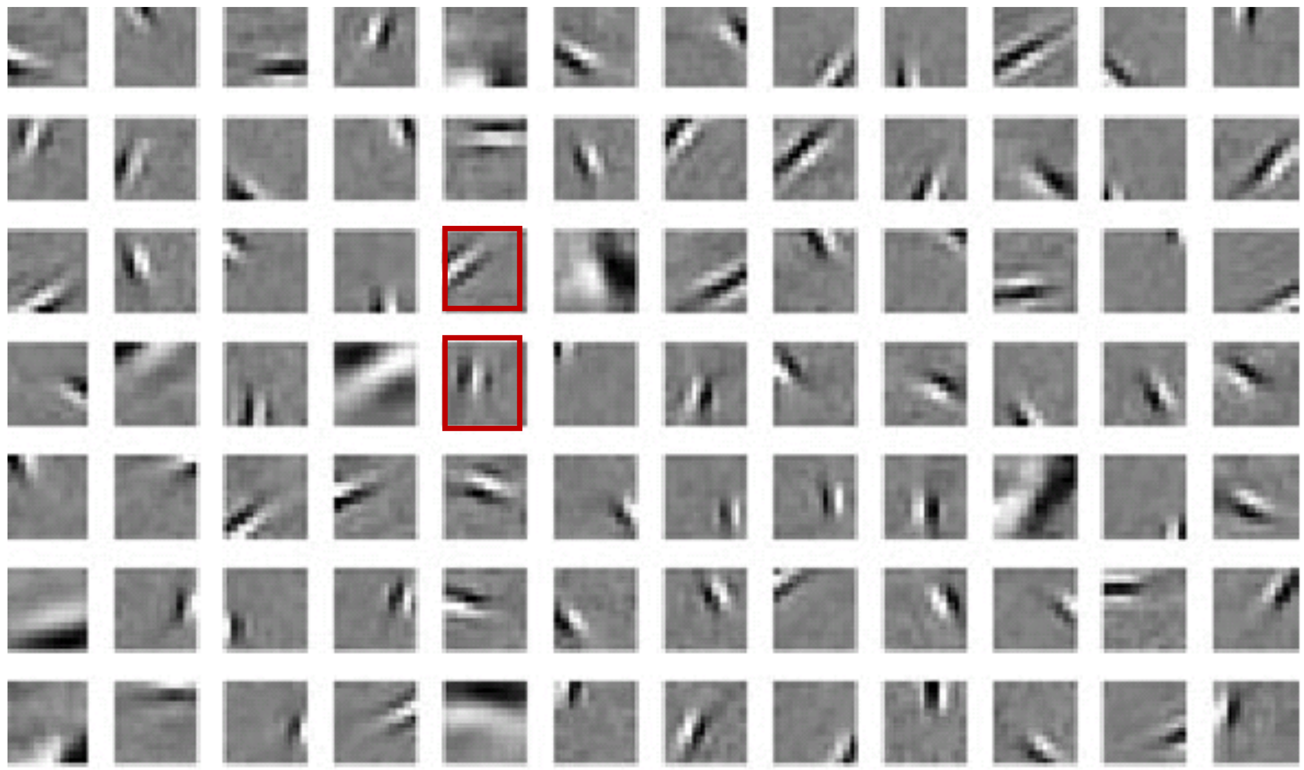
2019

Summary

- We’ve considered bottom-up scene statistics, efficient coding, and relation of linear transforms to visual filters

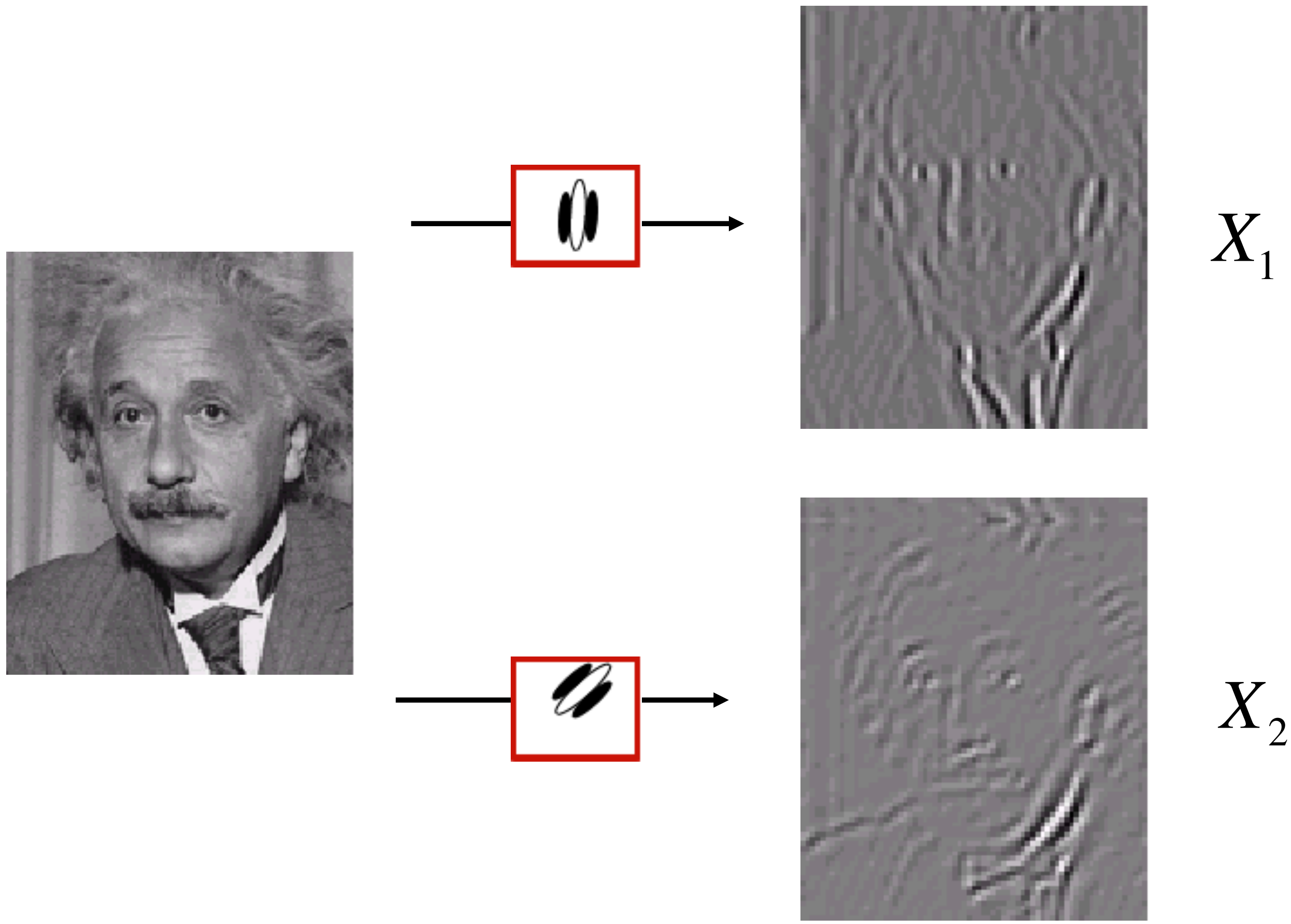
- This class: going beyond learning V1 like linear filters

Beyond linear

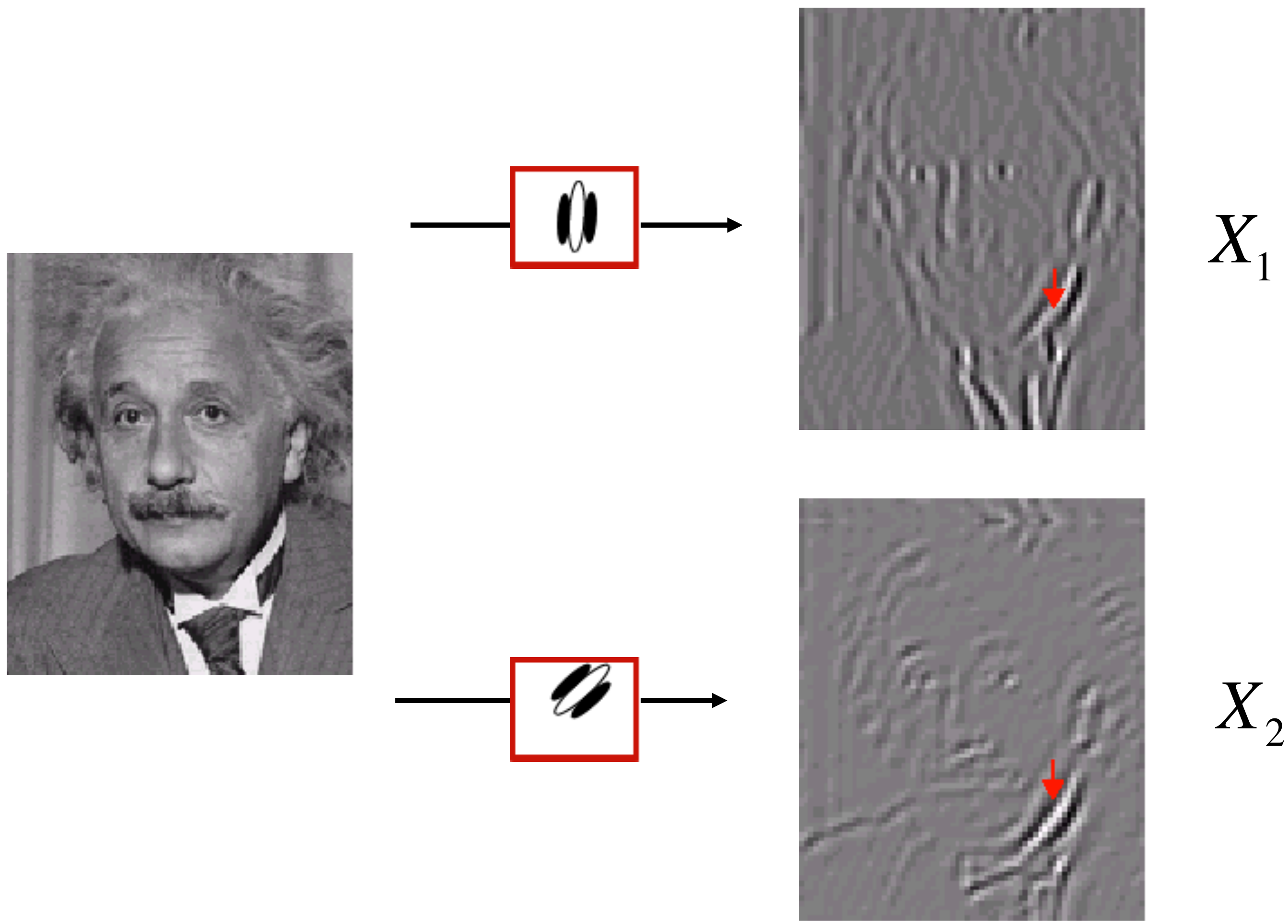


- Filter responses as independent as possible assuming a linear transform
- But are they independent?

Bottom-up Joint Statistics

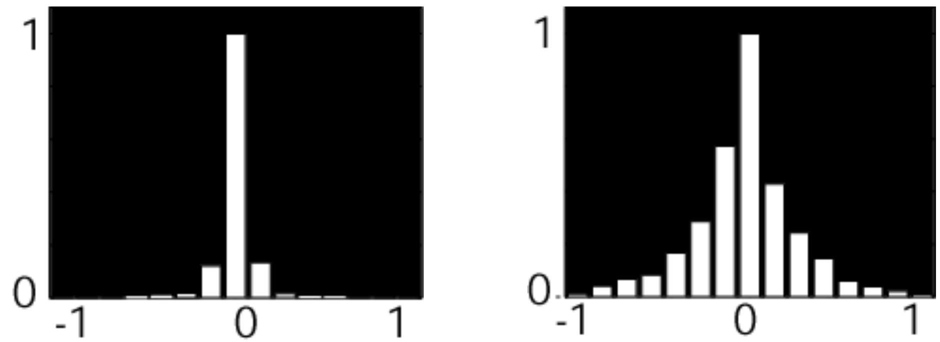


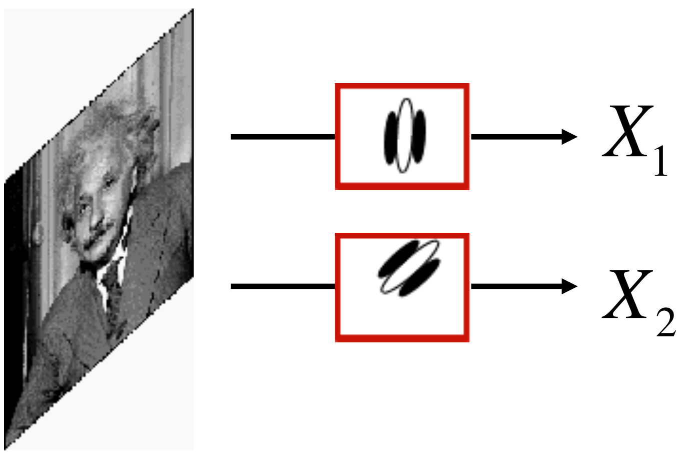
Bottom-up Joint Statistics



Bottom-up Joint Statistics

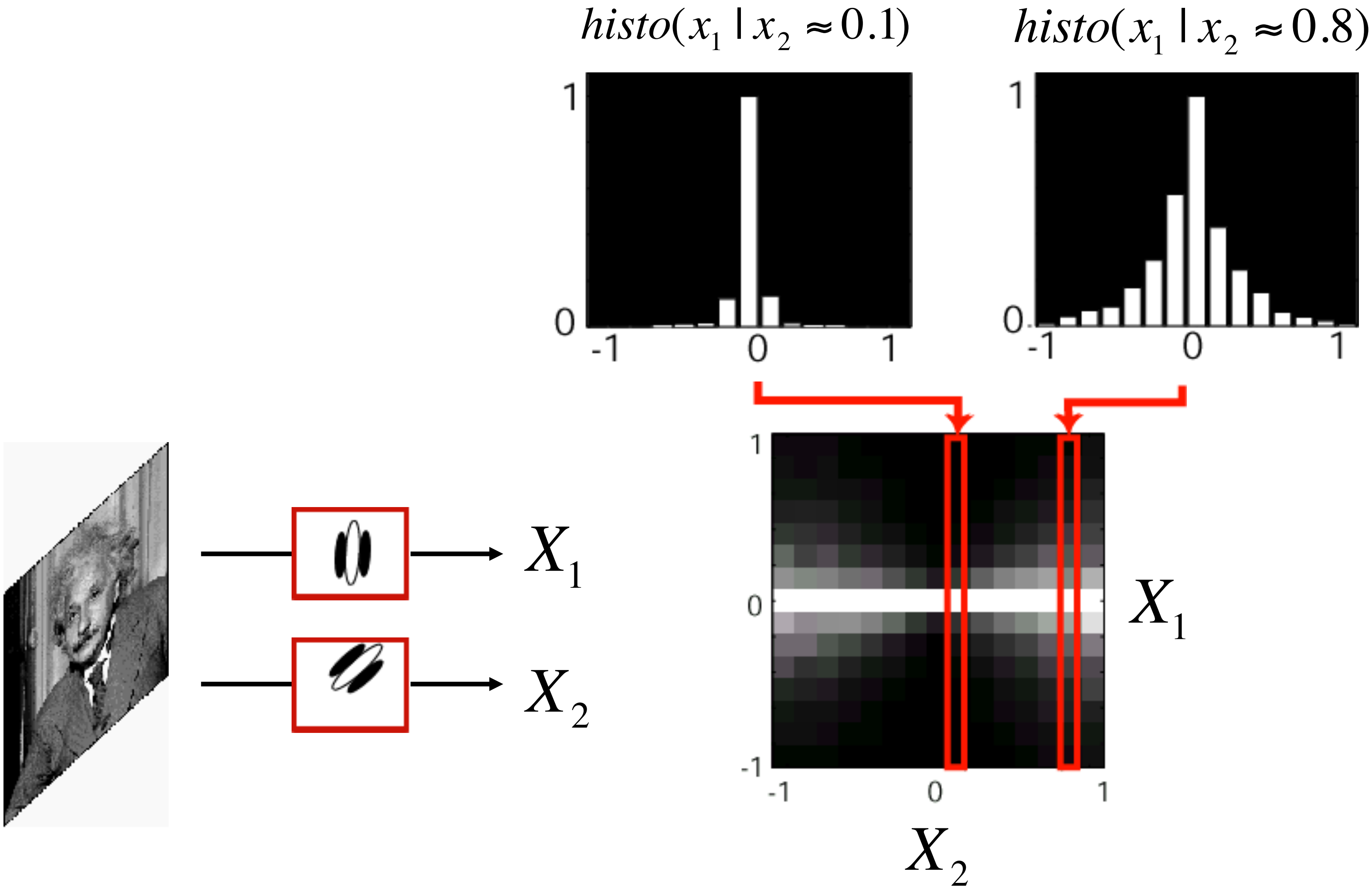
$histo(X_1 \mid X_2 \approx 0.1)$ $\quad$ $histo(X_1 \mid X_2 \approx 0.8)$





Are X_1 and X_2 statistically independent?

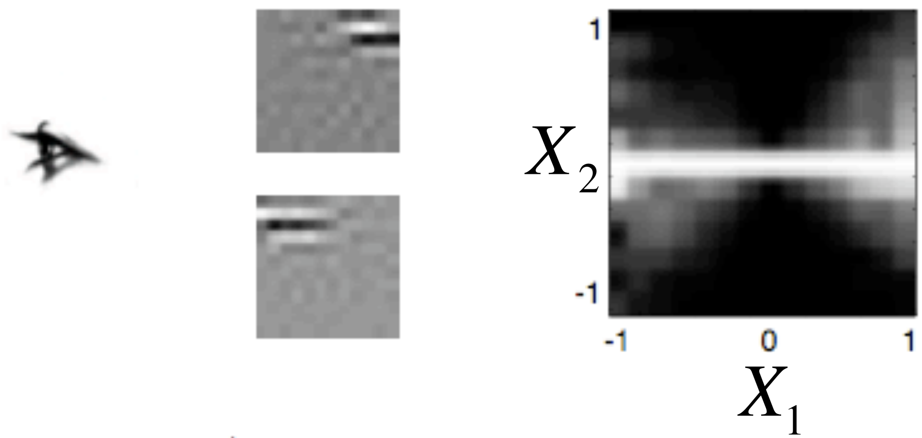
Bottom-up Joint Statistics

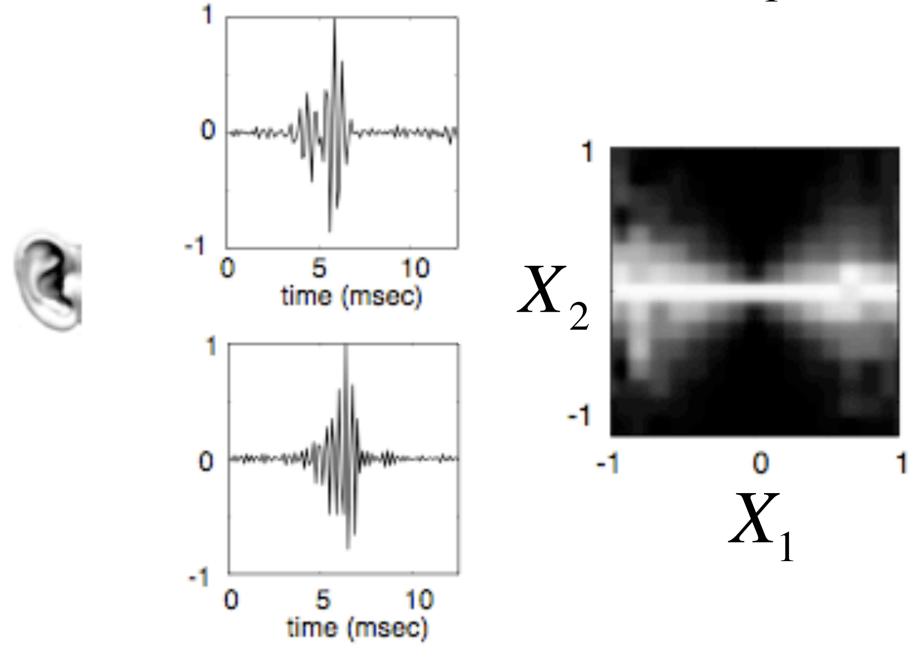


X_1 and X_2 are not statistically independent

Schwartz and Simoncelli, 2001

Bottom-up Joint Statistics



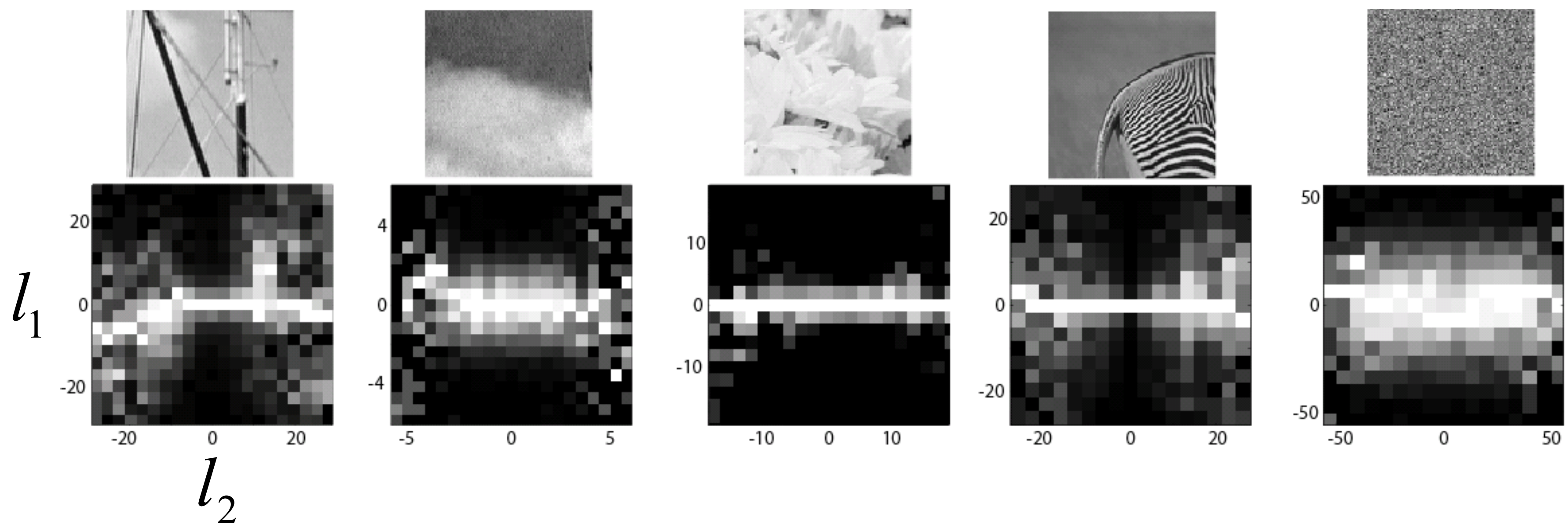


Bottom-up Statistics

Filter pair and different image patches…

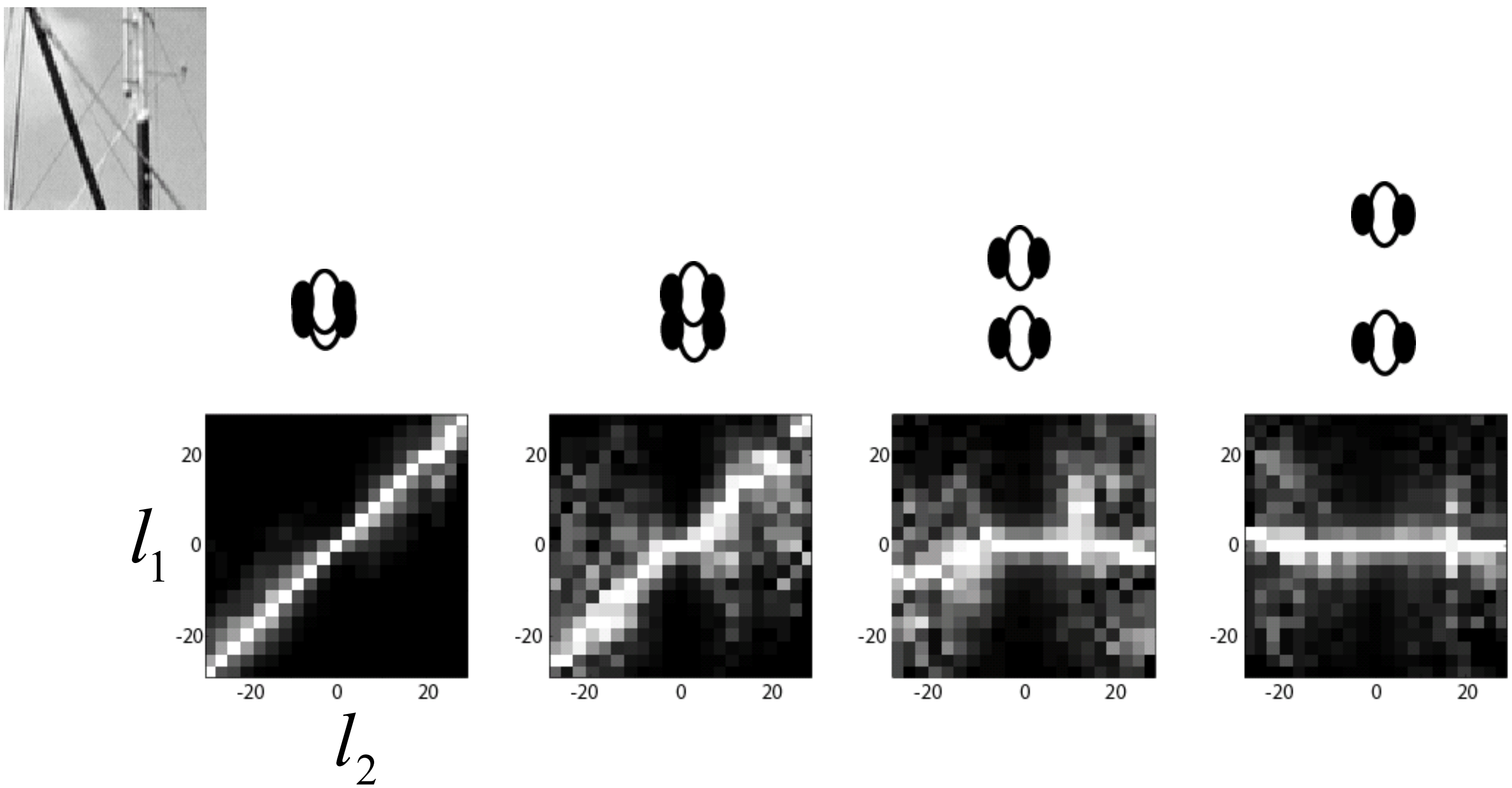
$\longrightarrow X_1$

$\longrightarrow X_2$



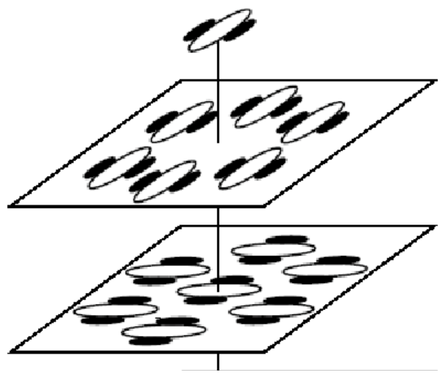
Bottom-up Statistics

Image patch and different filter pairs…



Modeling filter coordination

Modeling filter coordination in images

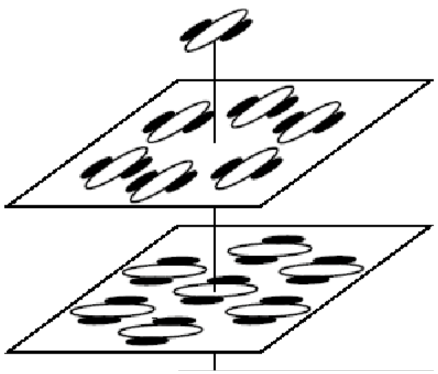


- Learning how more complex representations build up from the structure of dependencies in images

- Reducing dependencies further via nonlinear: divisive normalization – linking to spatial context effects (later)

Modeling filter coordination

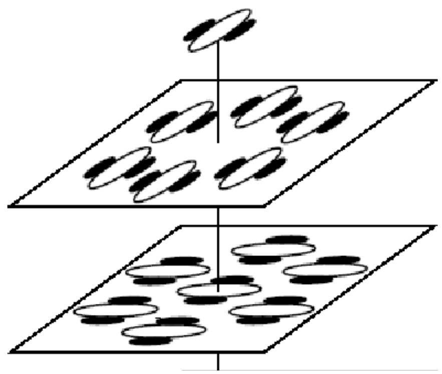
Modeling filter coordination in images



What kind of complex representations?

Modeling filter coordination

Modeling filter coordination in images



What kind of complex representations?

1. In V1, eg complex cells
2. Higher visual areas

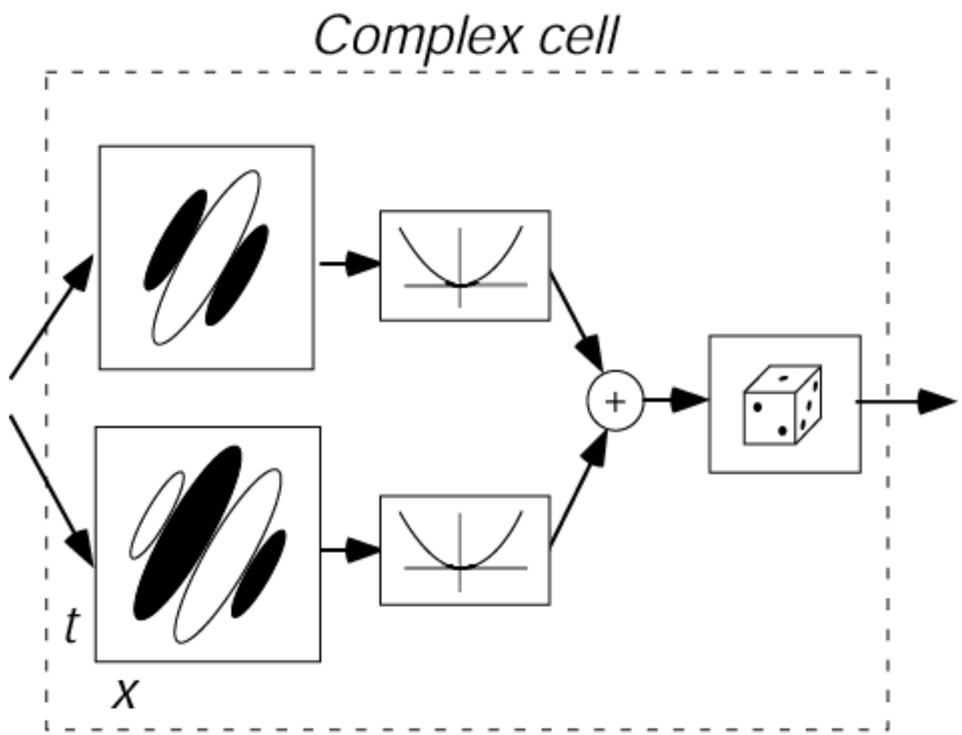
Modeling filter coordination

Modeling filter coordination in images

First what we know; then learning from dependencies in images

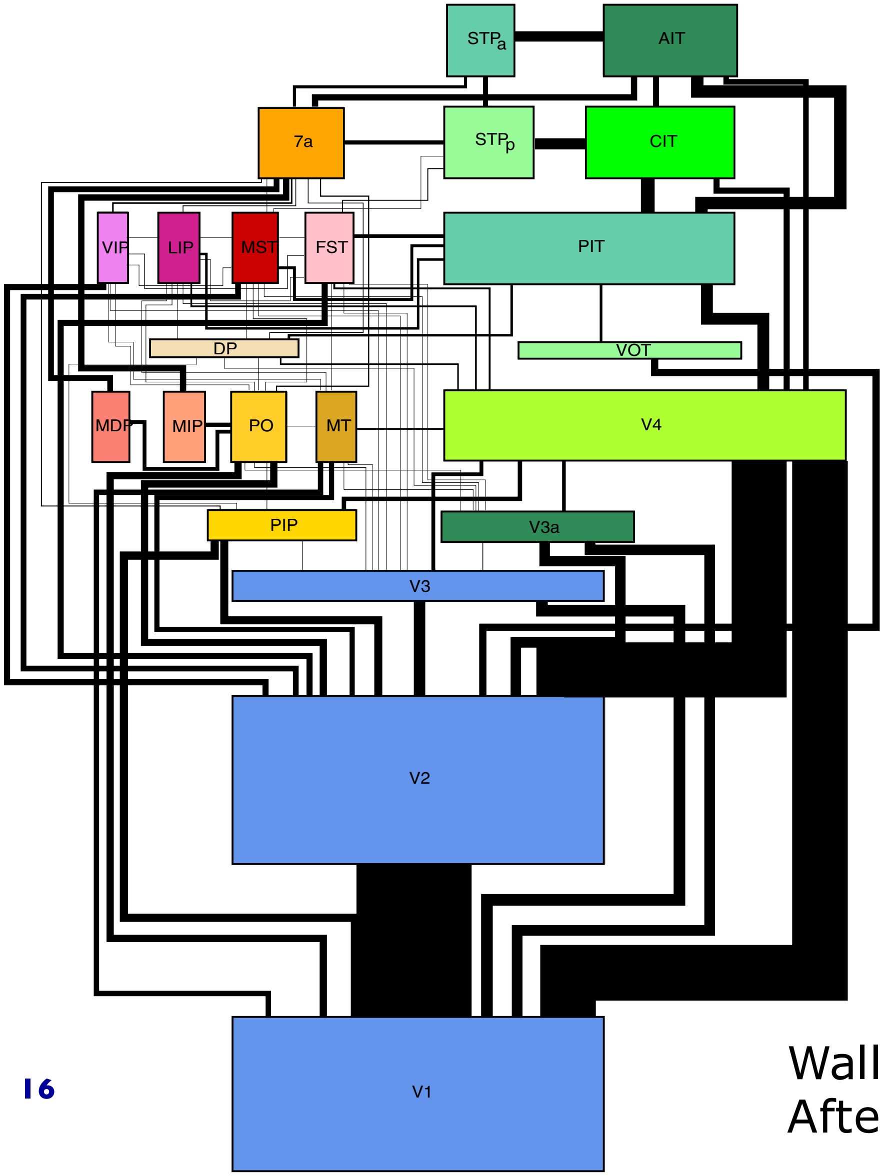
More complex representations

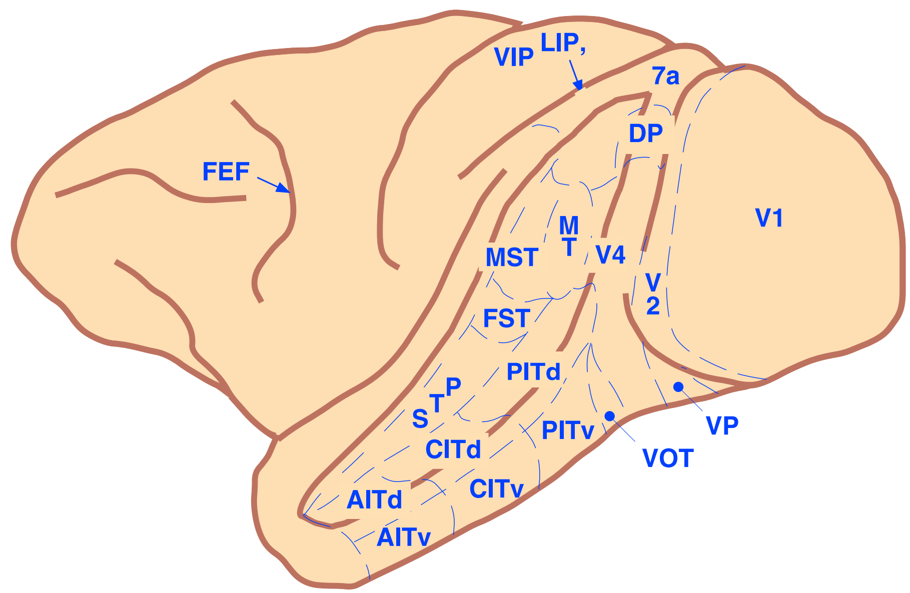
In primary visual cortex (capturing an invariance)



Adelson & Bergen (1985)

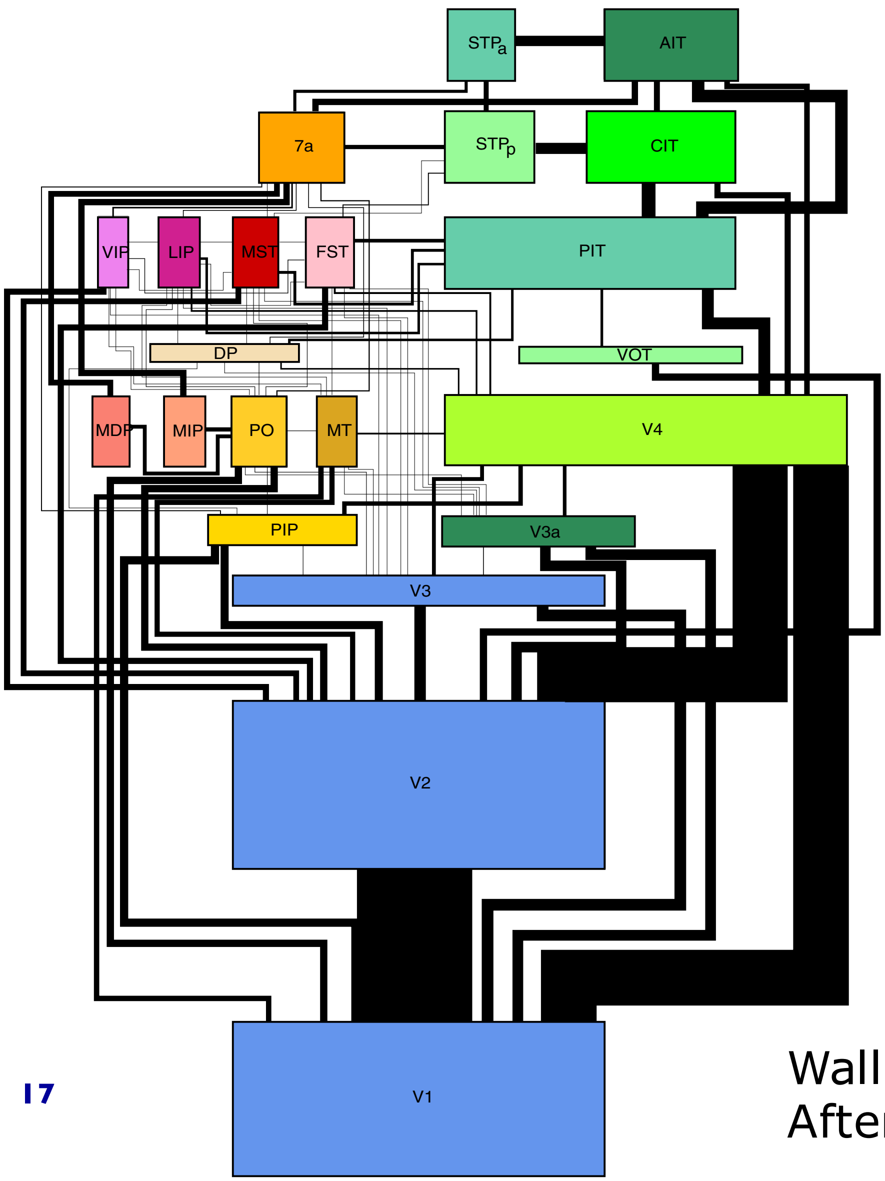
Beyond Primary Visual Cortex





Wallisch and Movshon 2008;
After Felleman and Van Essen, 1991

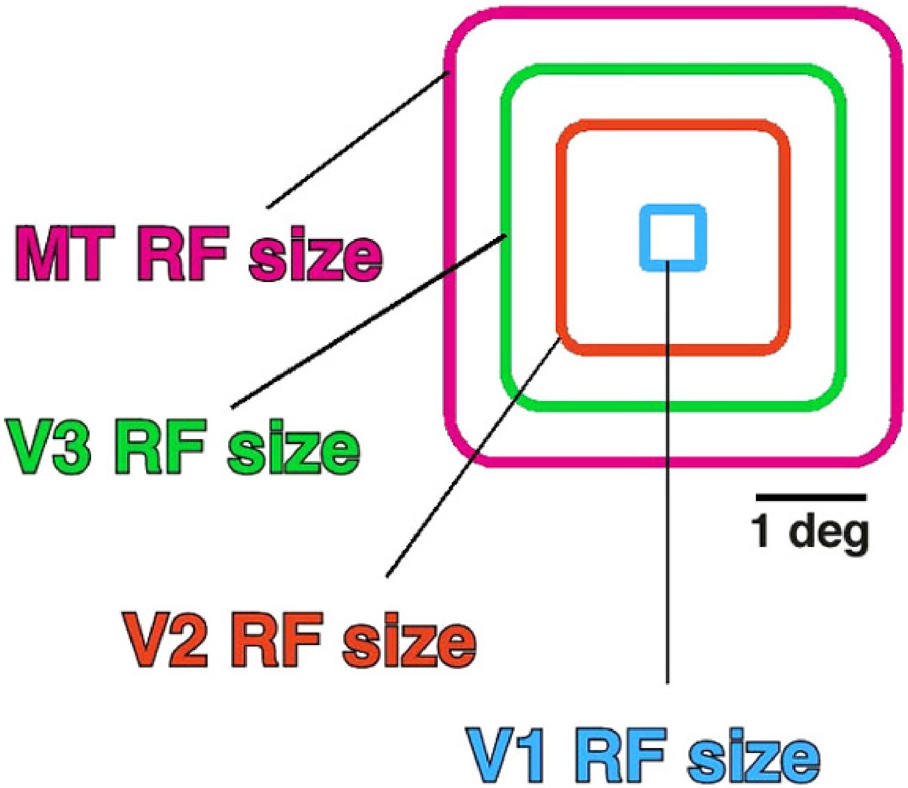
Beyond Primary Visual Cortex



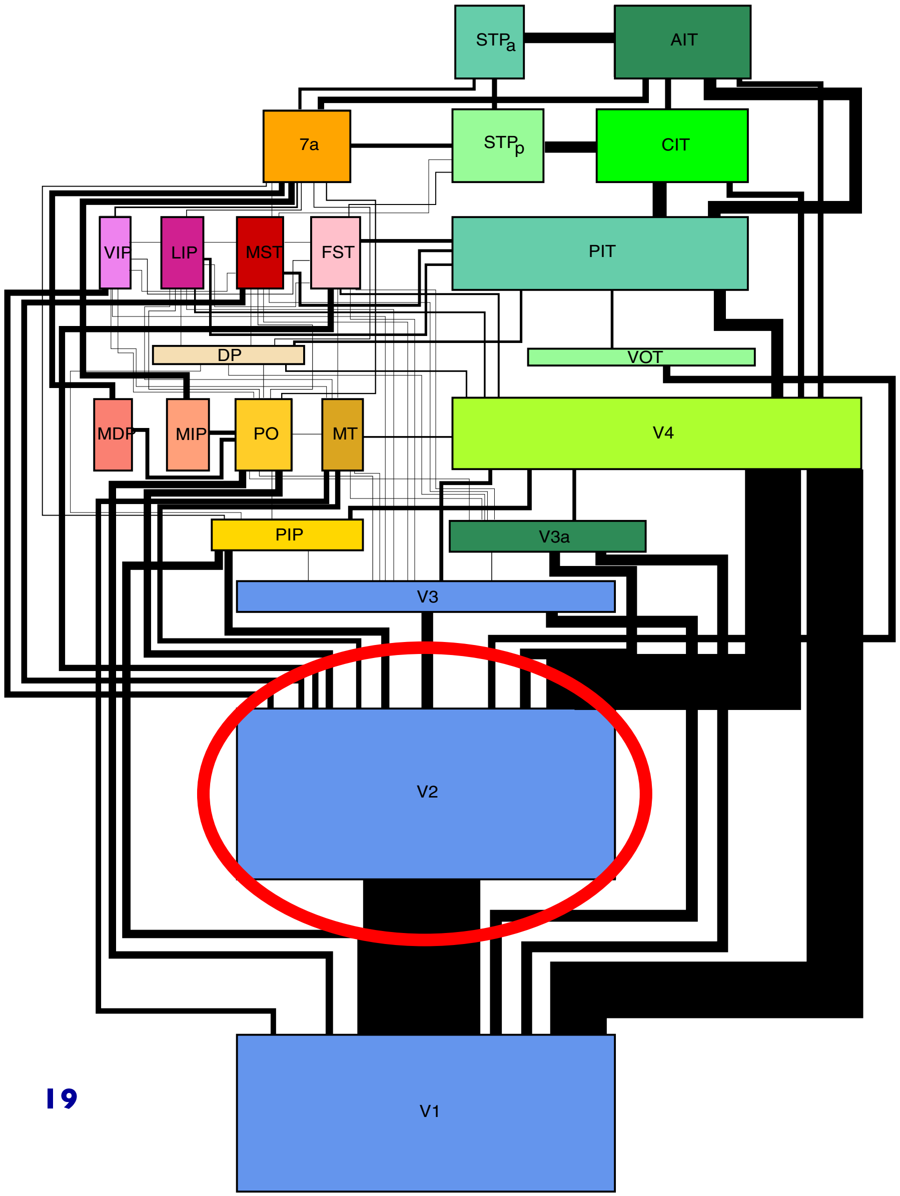
"each area is drawn with a size proportional to its cortical surface area, and the lines connecting the areas each have a thickness proportional to the estimated number of fibers in the connection. The estimate is derived by assuming that each area has a number of output fibers proportional to its surface area and that these fibers are divided among the target areas in proportion to their surface areas."

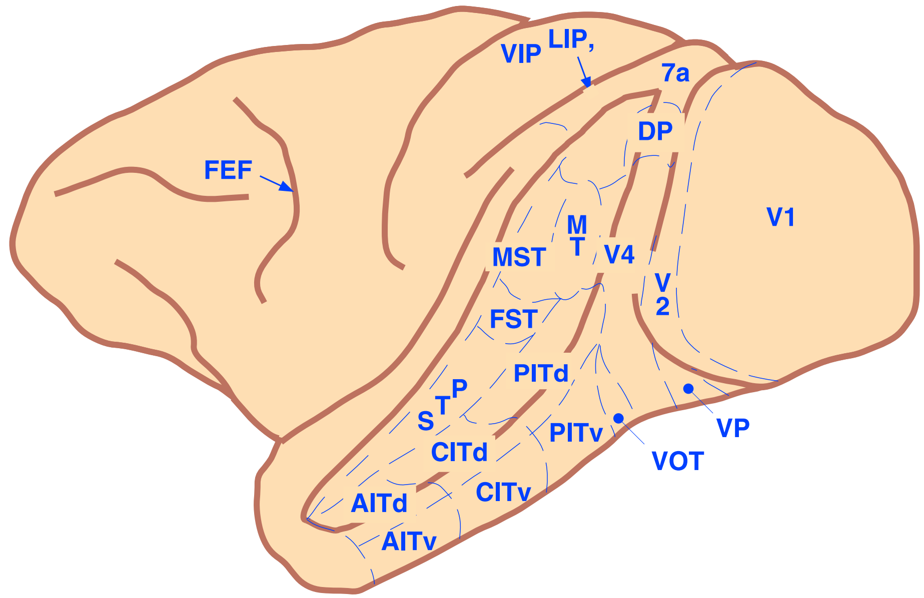
Wallisch and Movshon 2008;
After Felleman and Van Essen, 1991

RF size increases at higher levels



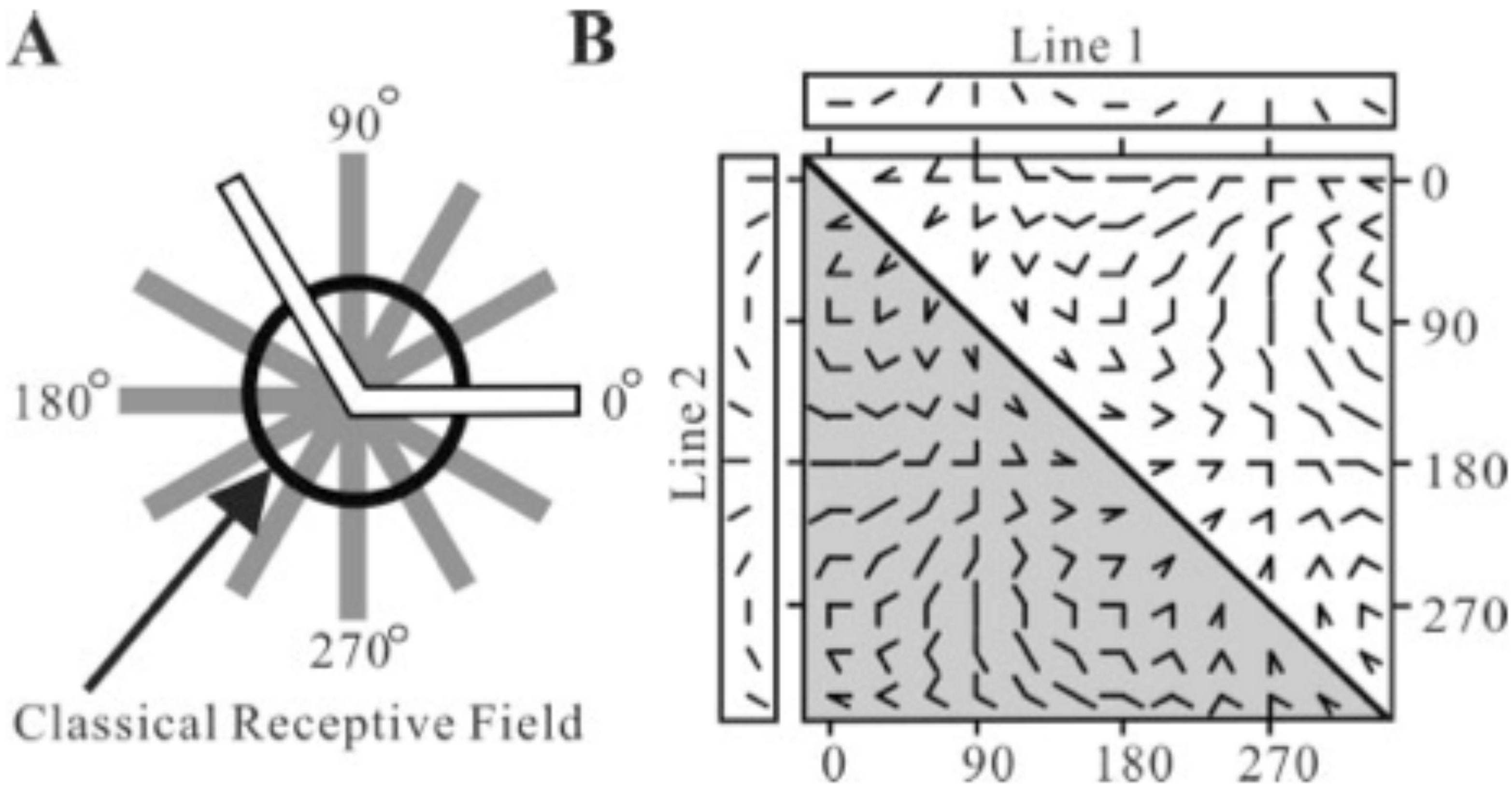
Beyond Primary Visual Cortex





More complex representations

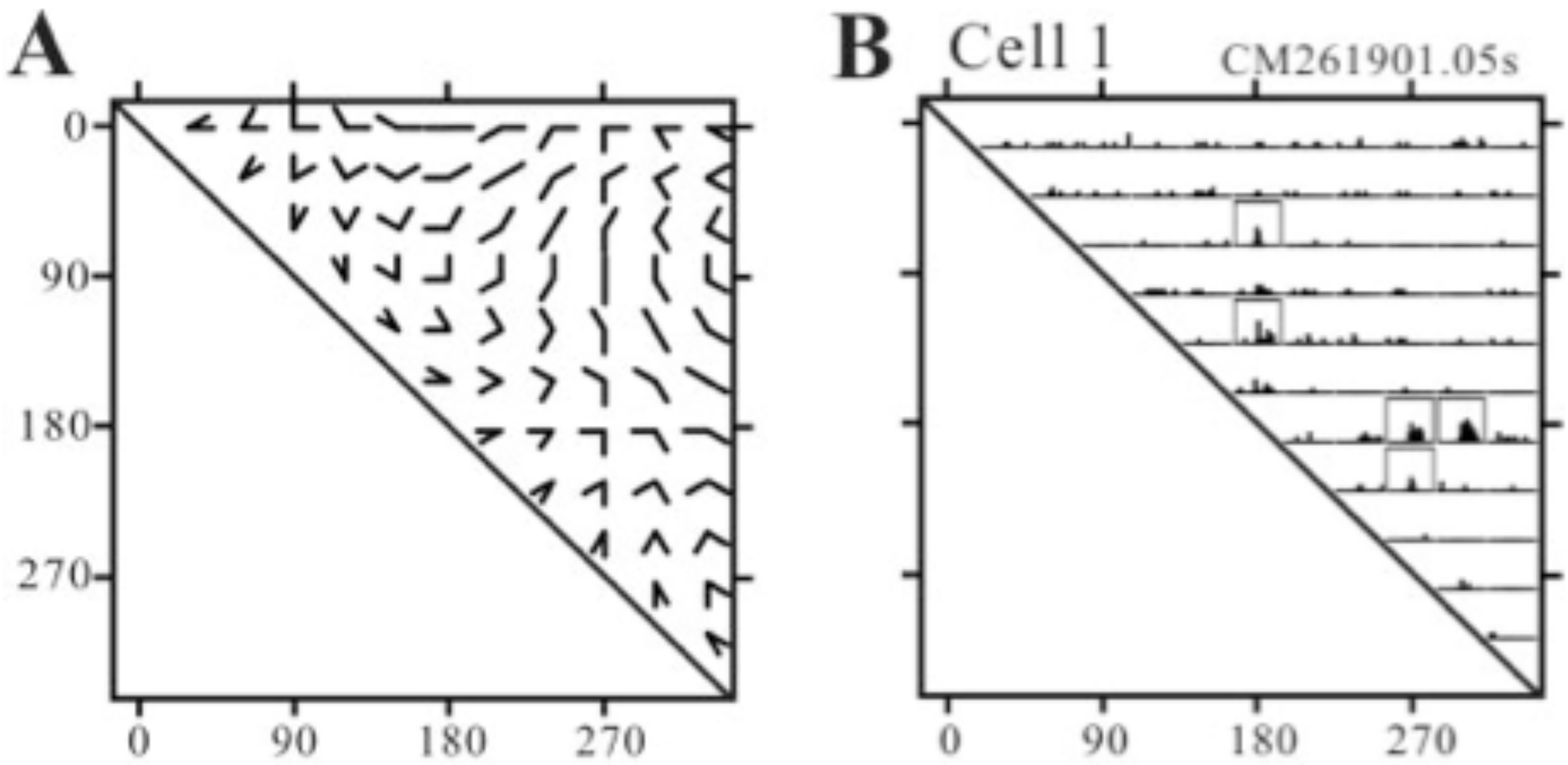
Example of V2 neurophysiology



Ito and Komatsu, 2005

More complex representations

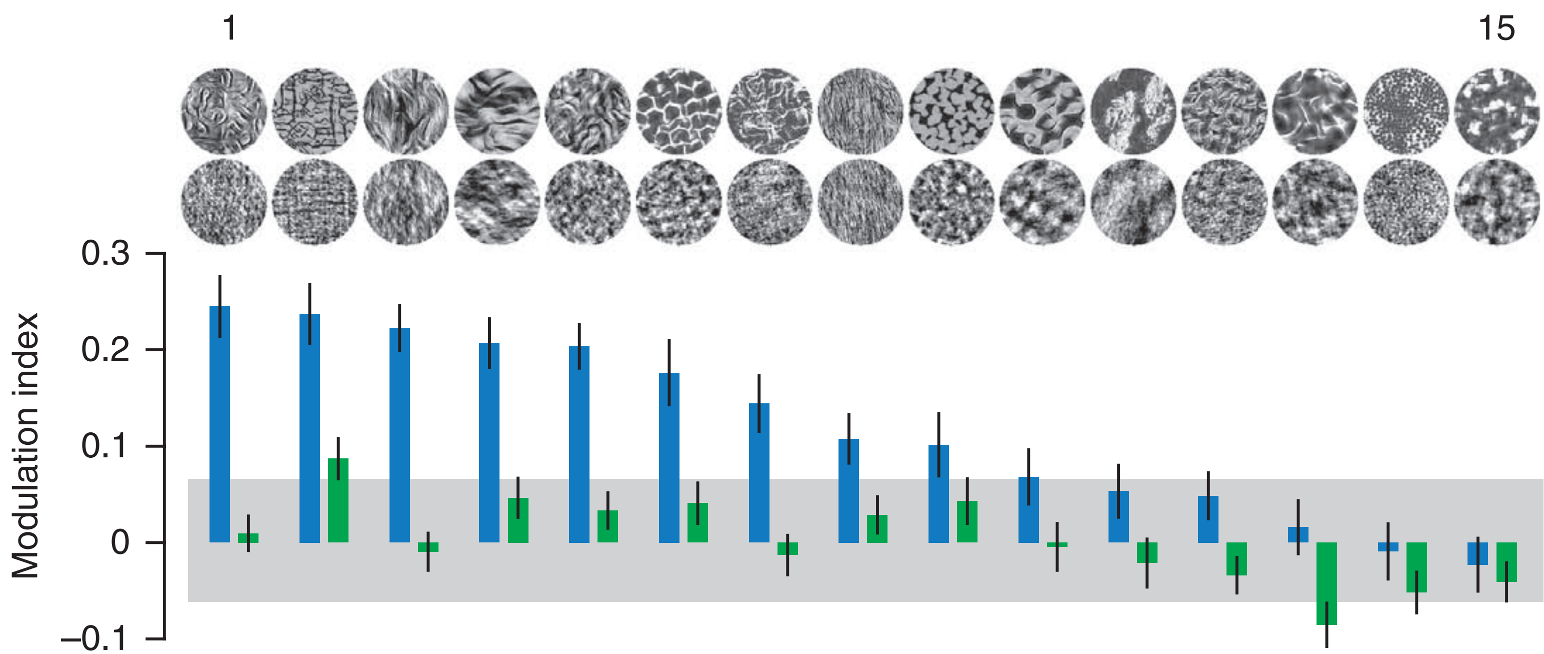
Example of V2 neurophysiology



Ito and Komatsu, 2005

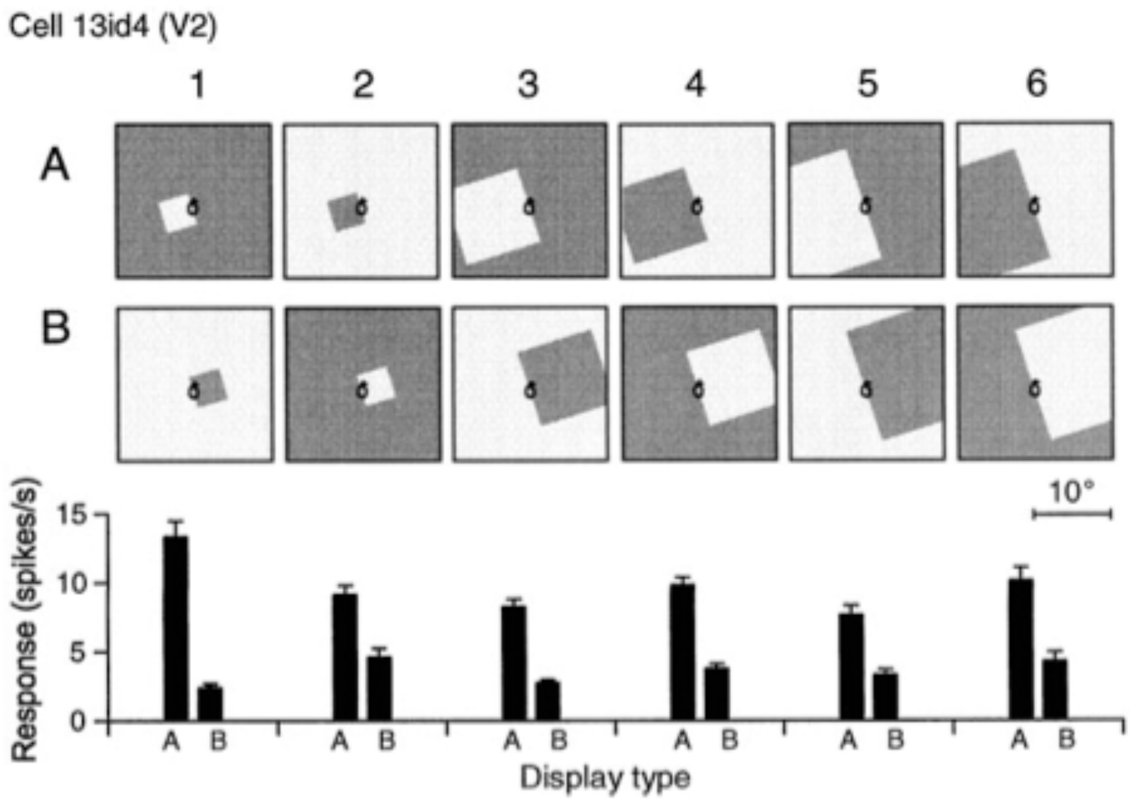
More complex representations

Example of V2 neurophysiology



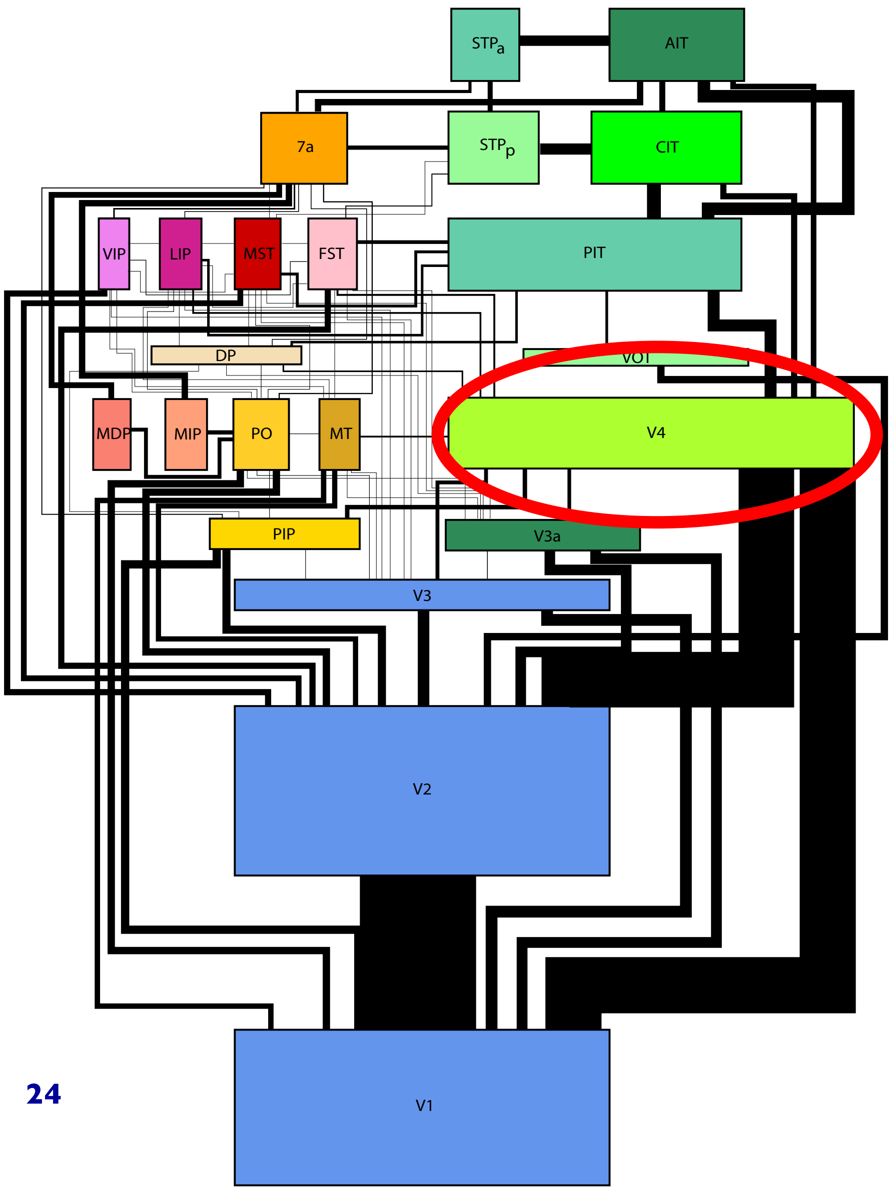
Freeman, Ziemba, Heeger, Simoncelli, Movshon 2013

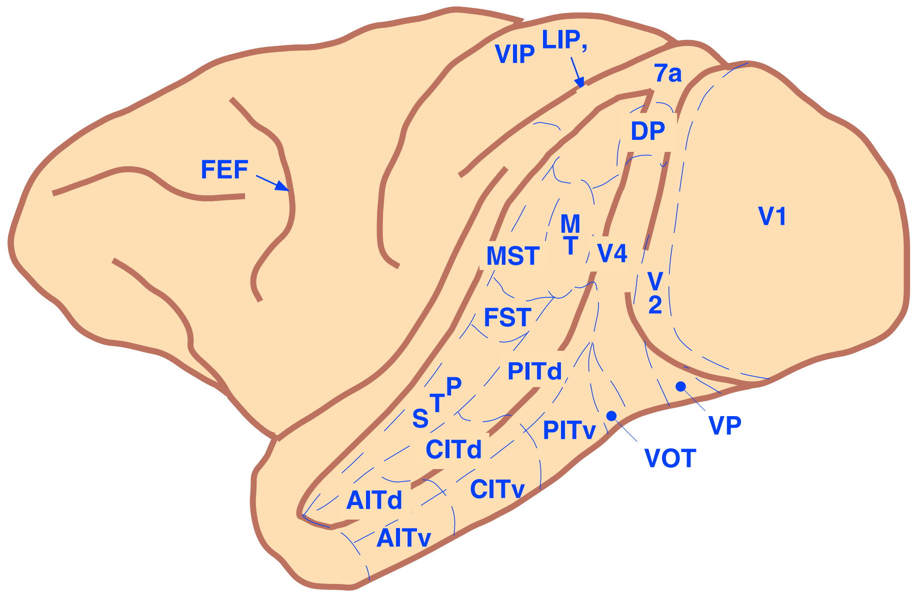
More complex: Figure ground



Zhou et al. von der Heydt, 2000; Zhaoping 2005

Beyond Primary Visual Cortex

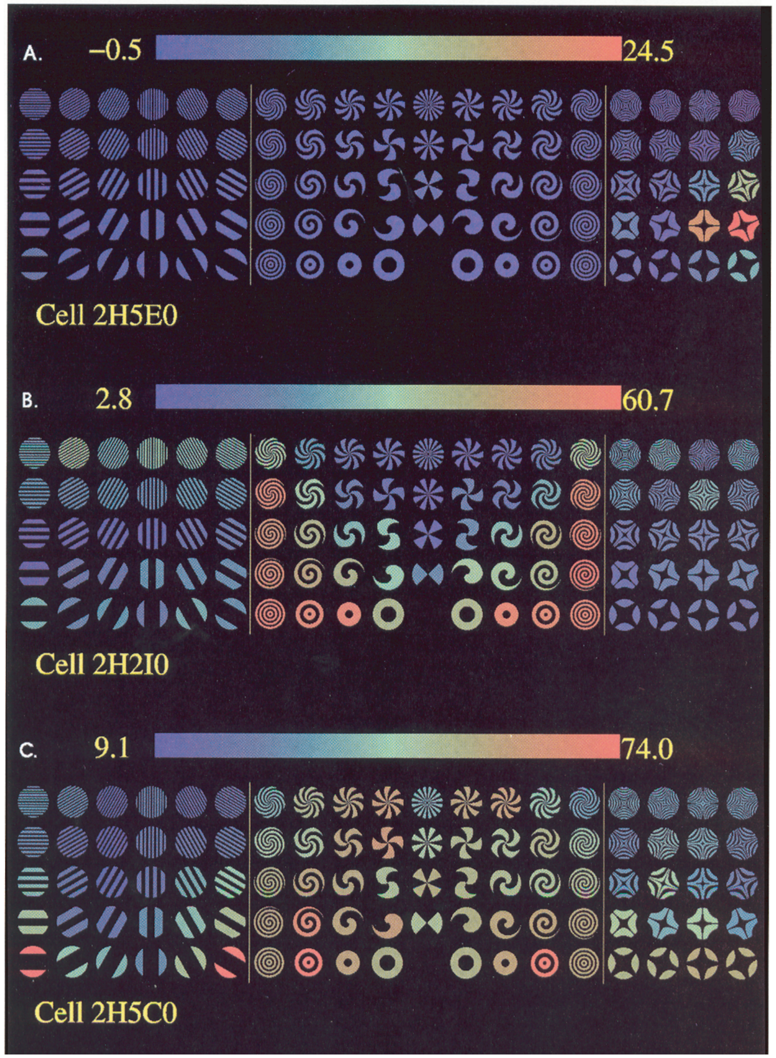




From Adam Kohn

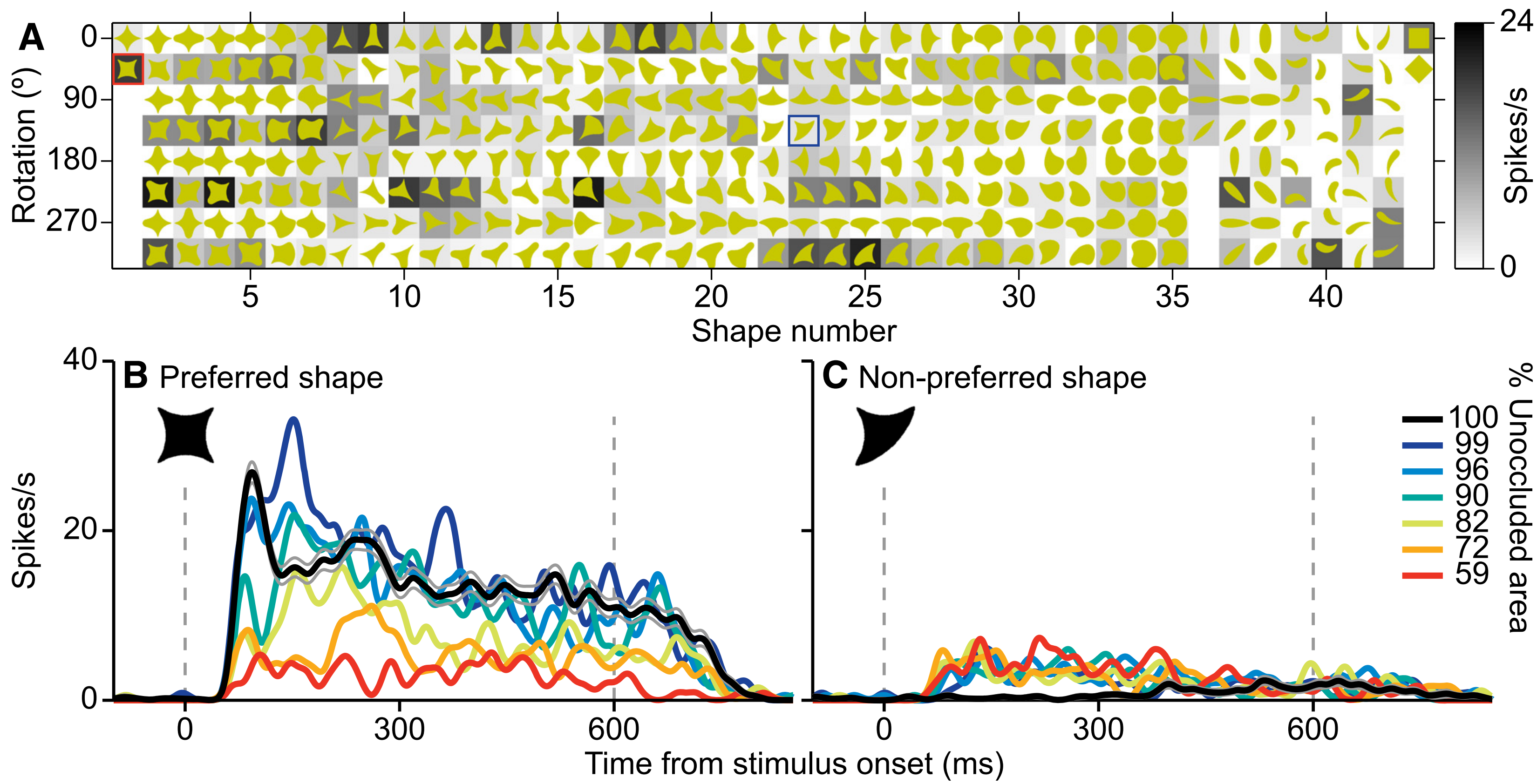
More complex representations

Example of V4 neurophysiology



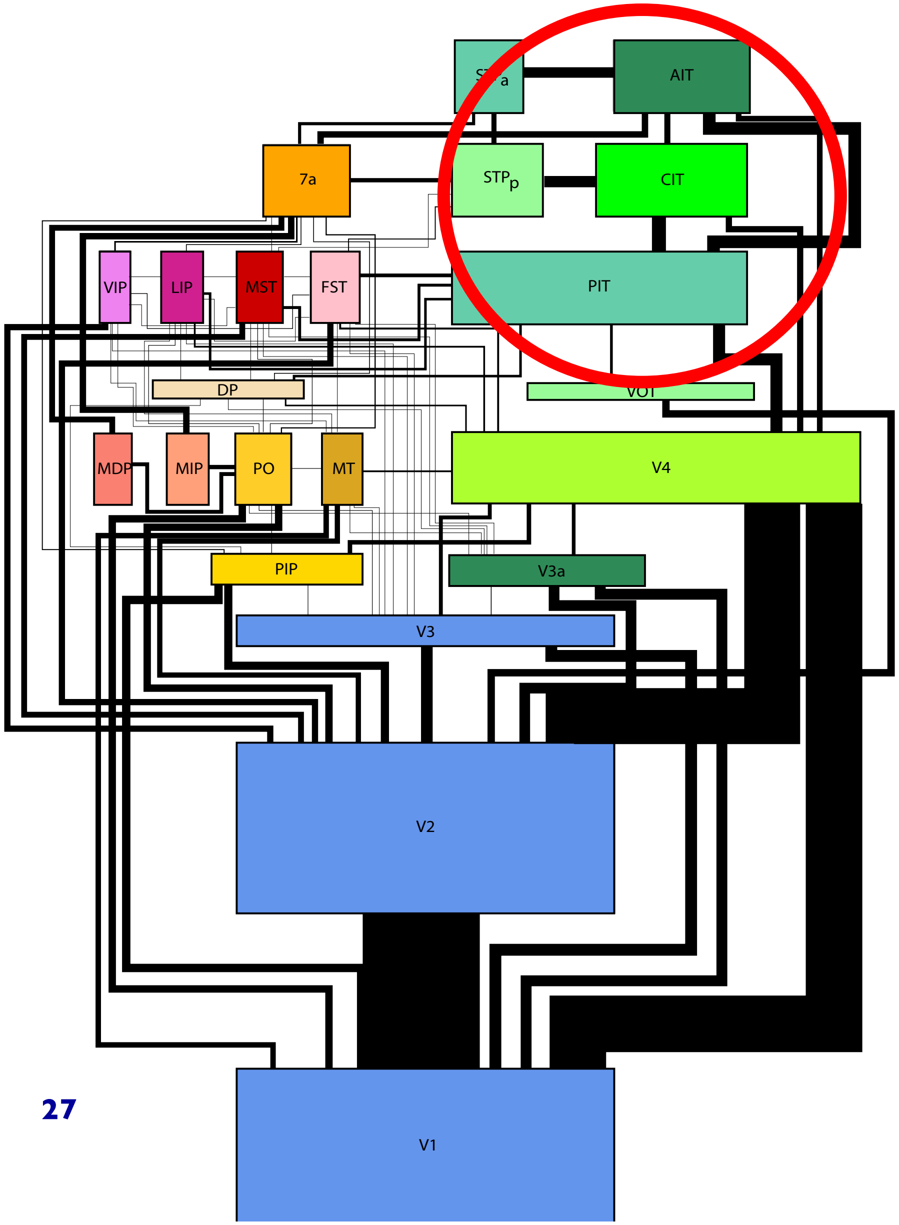
More complex representations

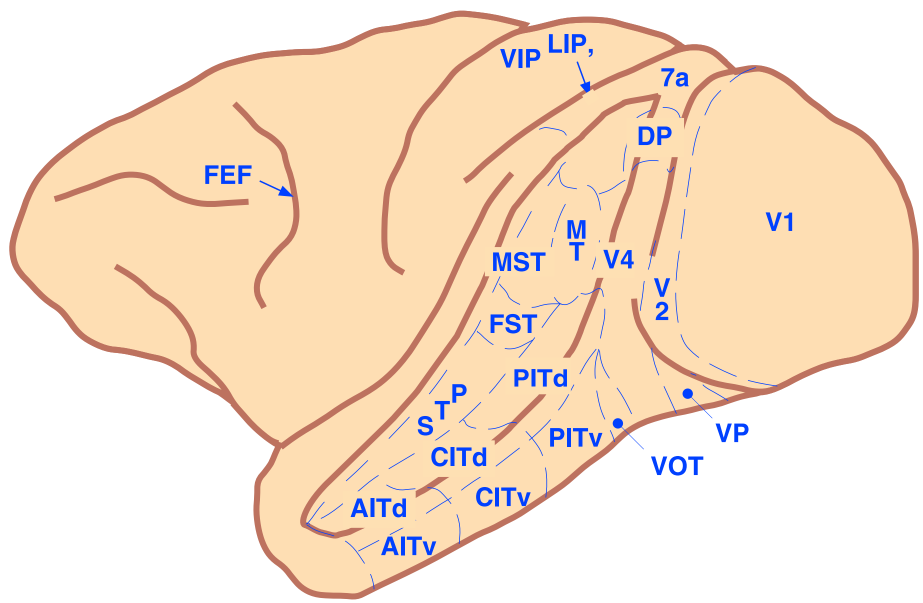
Example of V4 neurophysiology



Pasupathy lab (Kosai et al. 2014)

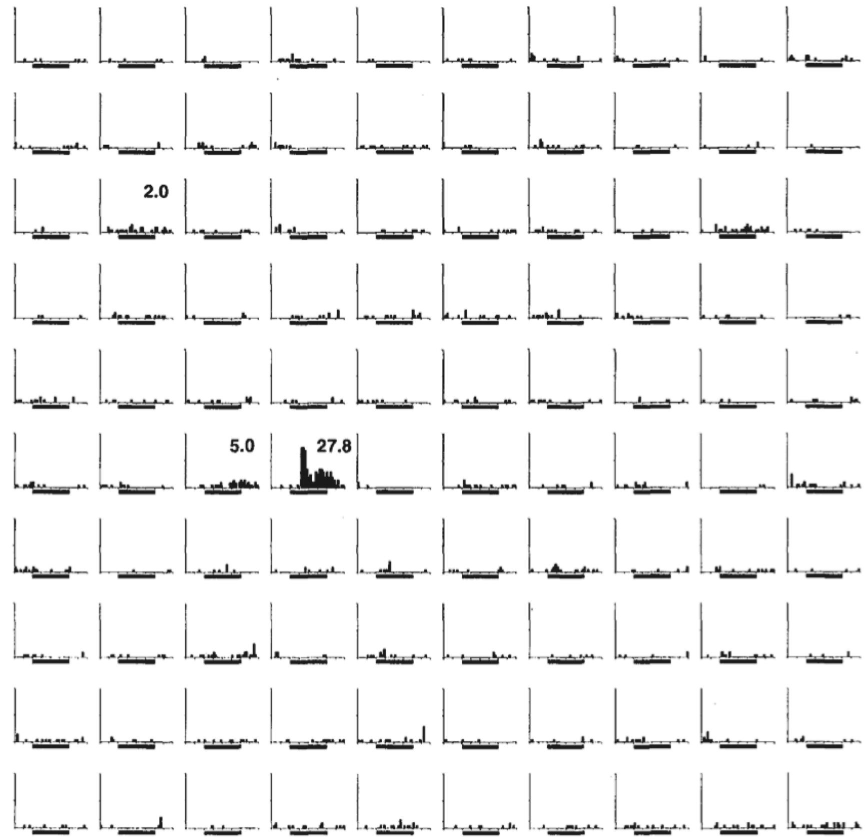
Beyond Primary Visual Cortex



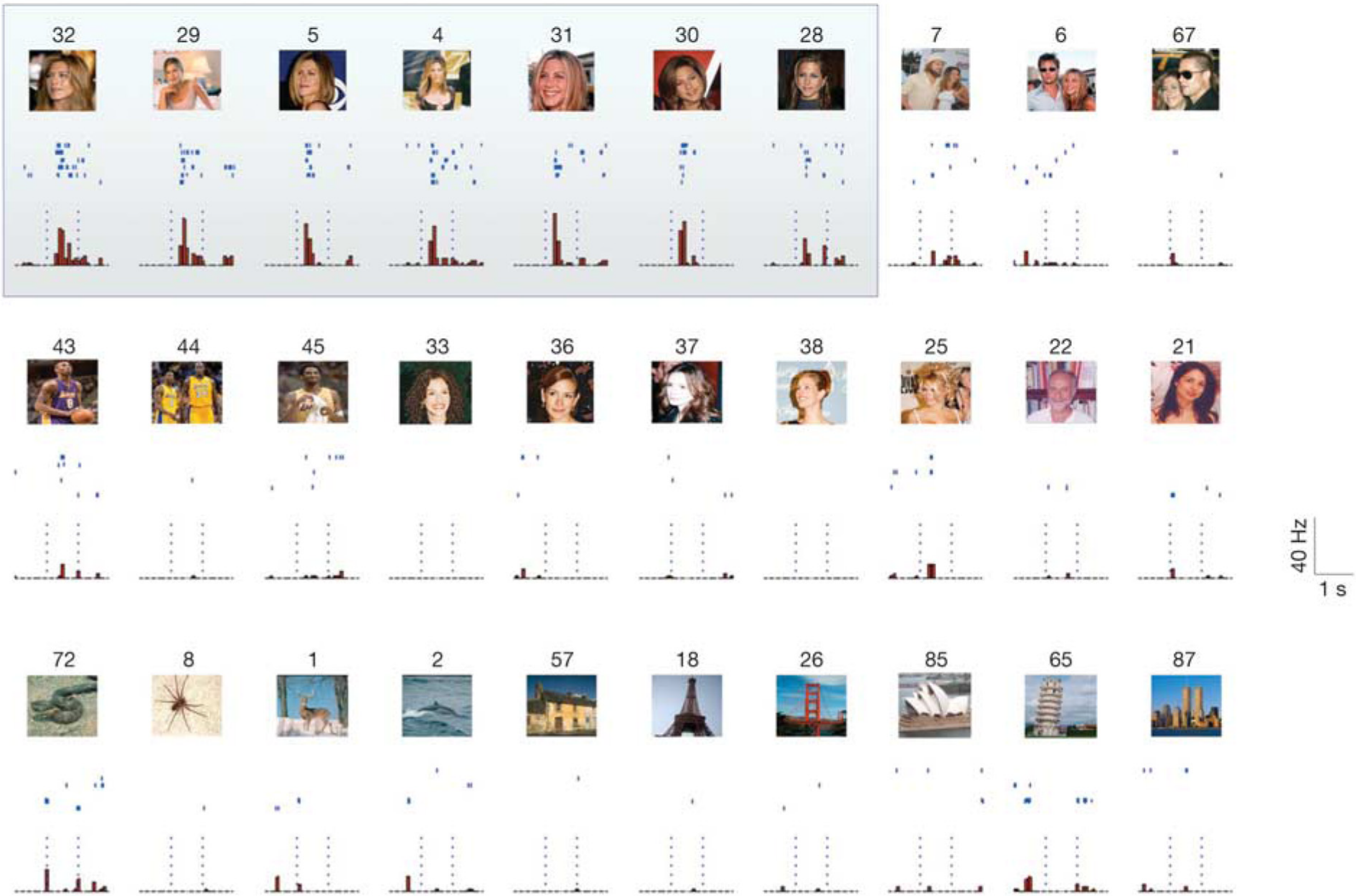


From Adam Kohn

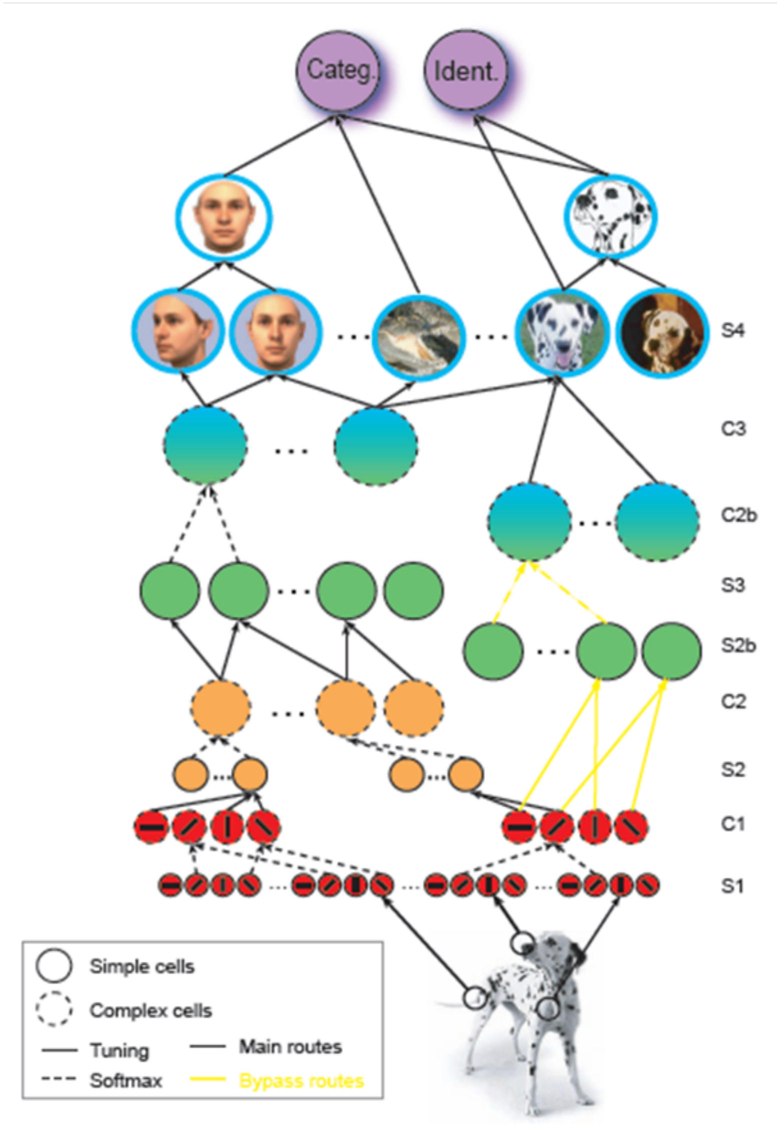
More complex representations



More complex representations



Selectivity and tolerance increase at higher levels



Reisenhuber and Poggio

More complex representations

What about learning from natural images beyond V1 like filters ?

Types of learning?

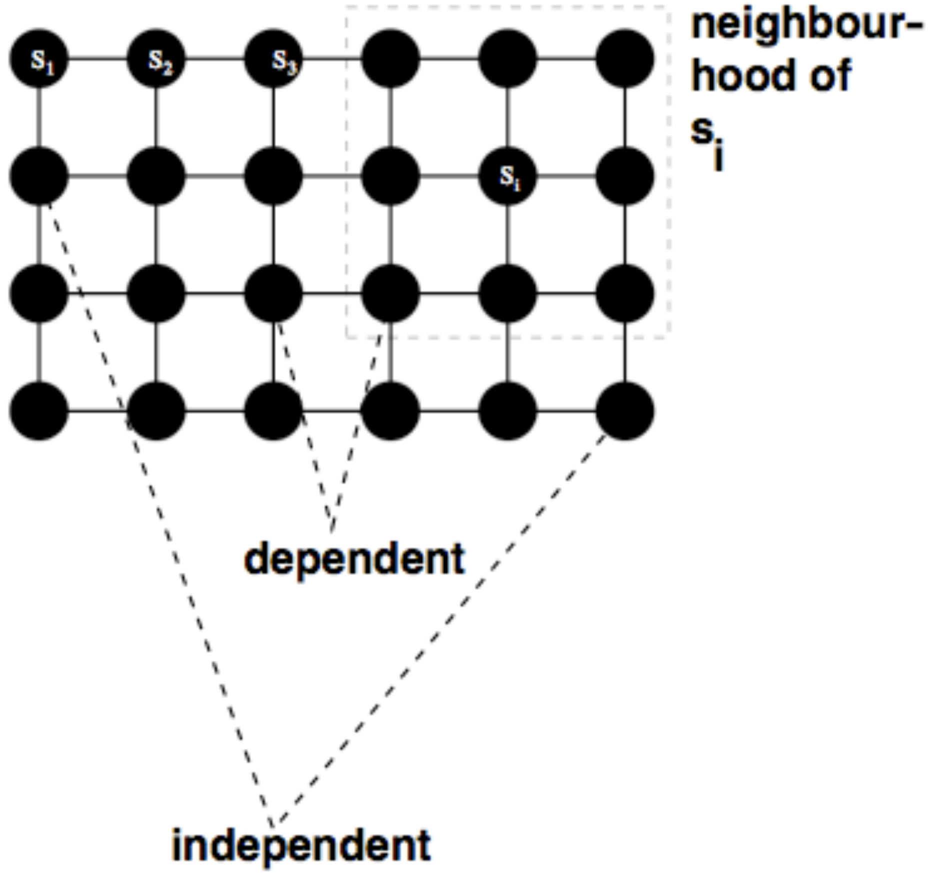
Types of learning

- Unsupervised
- Supervised, discriminative
- (Reinforcement learning)

Deep learning and unsupervised

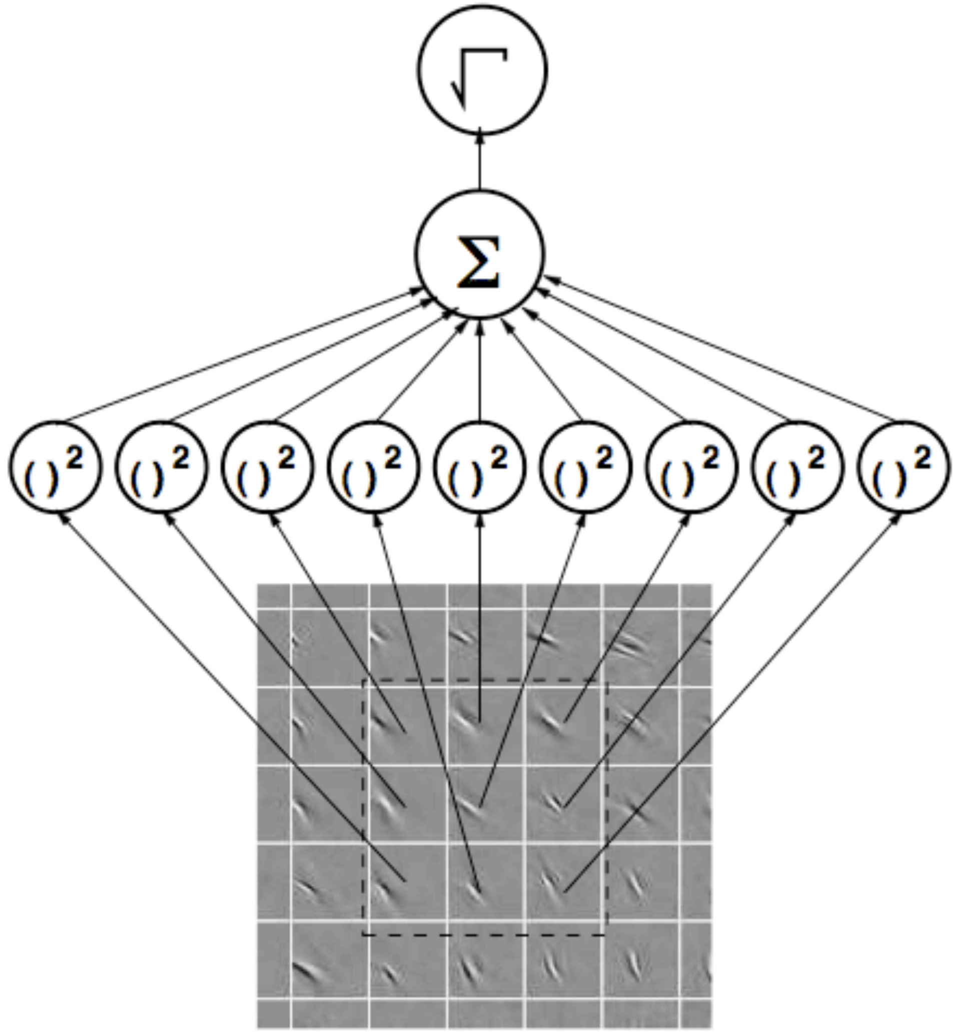
- Some work on learning hierarchy across several layers with unsupervised approaches
- Large scale supervised, discriminative learning has had success in scene recognition in recent years (eg, with Krizhevsky et al. 2012) from the machine learning perspective, and some studies have started linking to cortical processing

Extensions to ICA



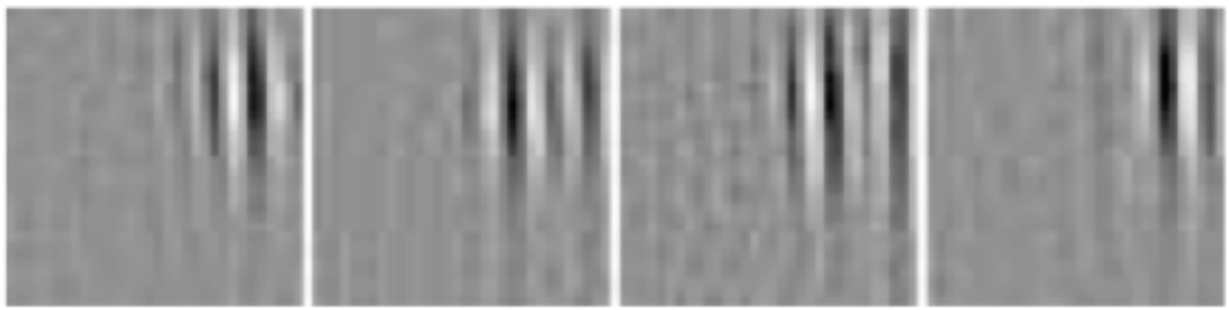
- from Hyvarinen and Hoyer; relax independence assumption; nearby units no longer independent; but different neighborhoods independent of one another...

Extensions to ICA



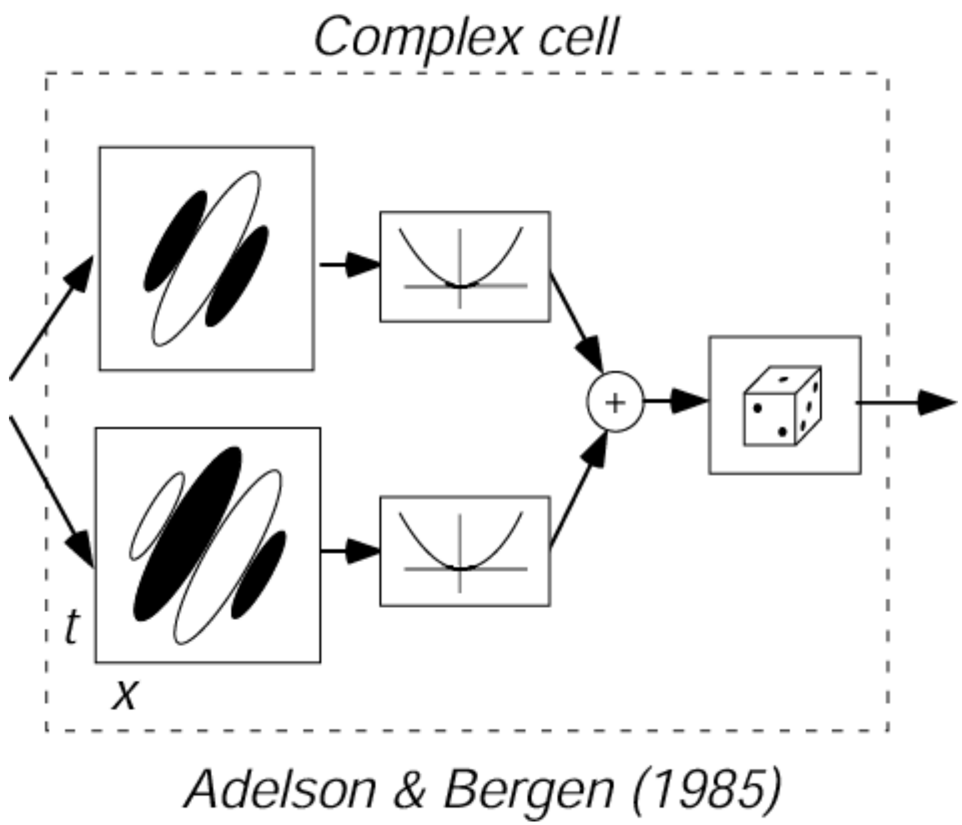
- Hyvarinen and Hoyer

Extensions to ICA



- Hyvarinen book: shown smaller group of dependent filters

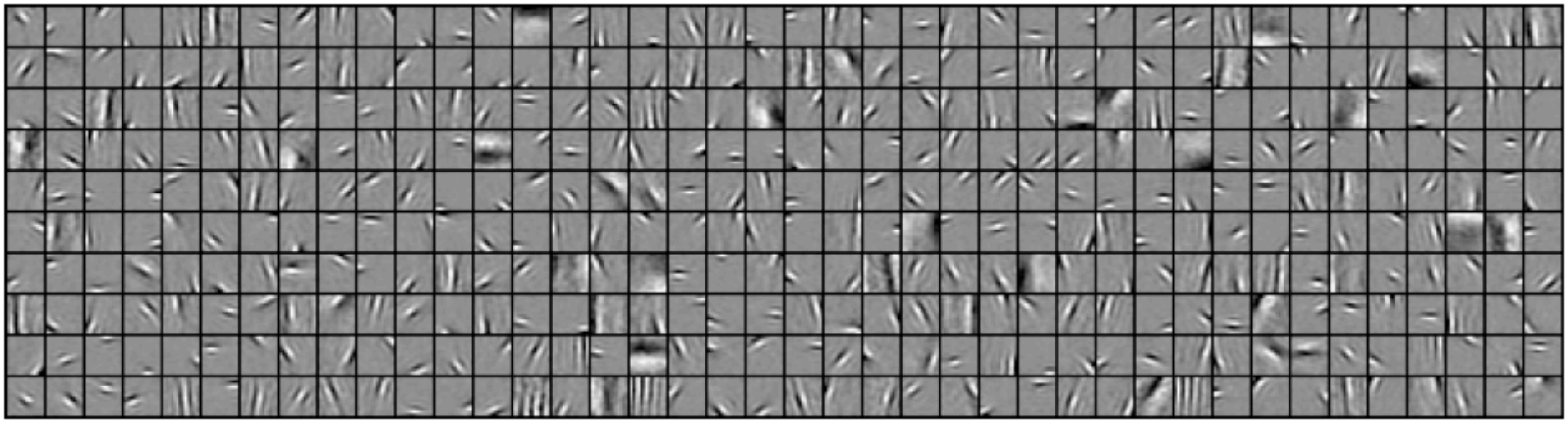
Complex cell



38

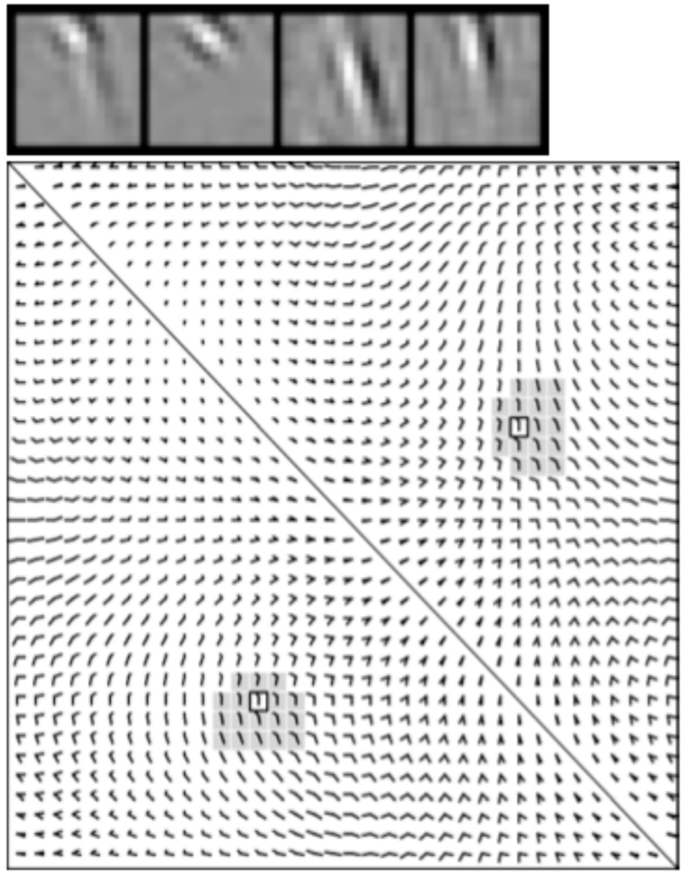
Relates to complex cells and invariances…

Unsupervised learning



Lee, Ekanadham, NG, 2007:
- 2-layer sparse coding (first layer)

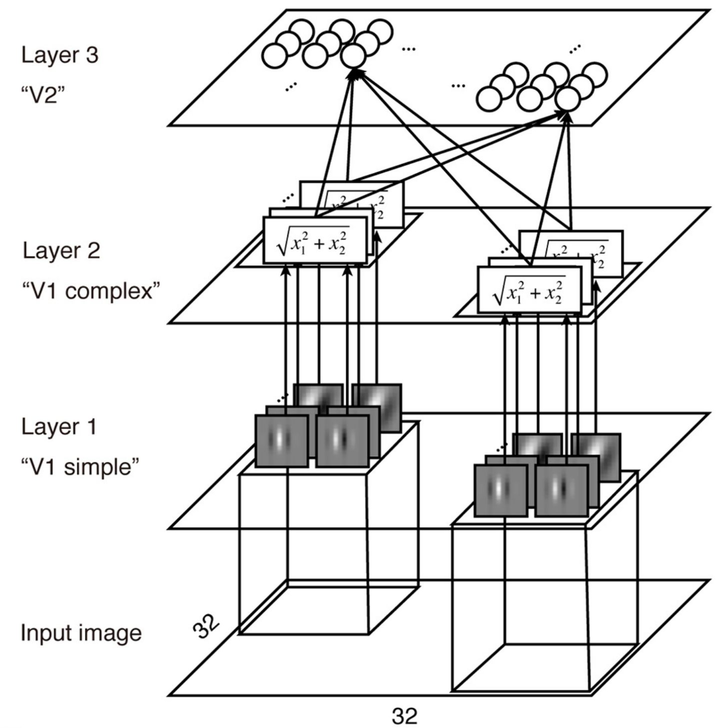
Unsupervised learning

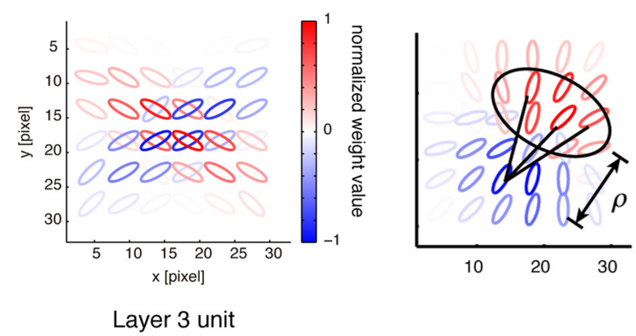


Lee, Ekanadham, NG, 2007:

- 2-layer sparse coding (second layer)

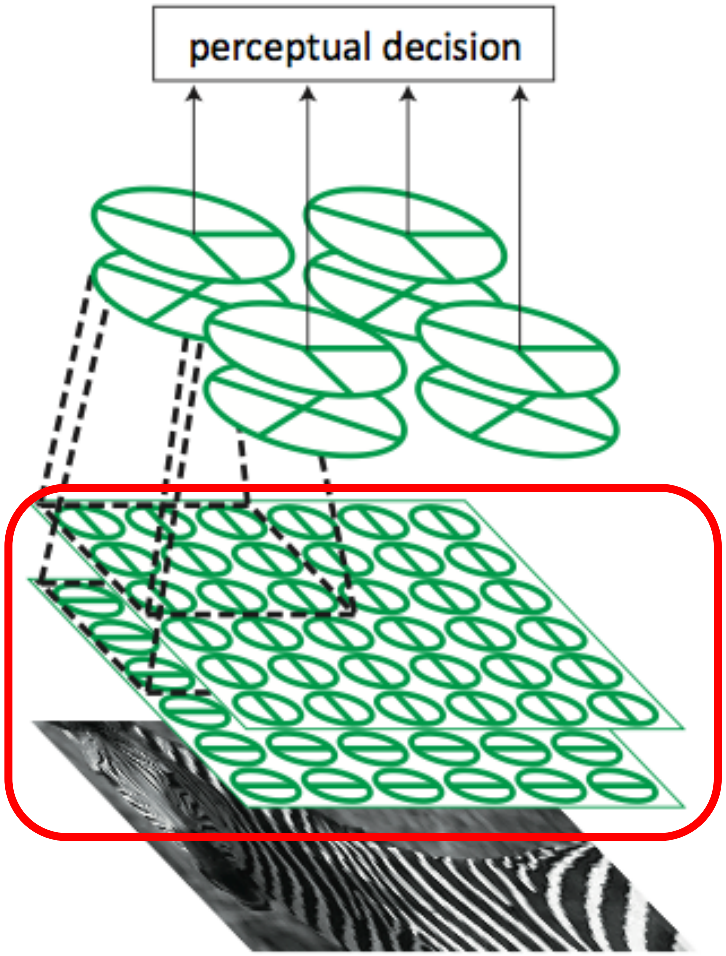
Unsupervised learning





- Hosoya, Hyvarinen, 2015
- Significant dimensionality reduction via PCA before expansive ICA on "complex cells"

Optimal normalization in first layer can help unsupervised learning of next layer



V2 model units

Linear transform
(e.g., PCA)

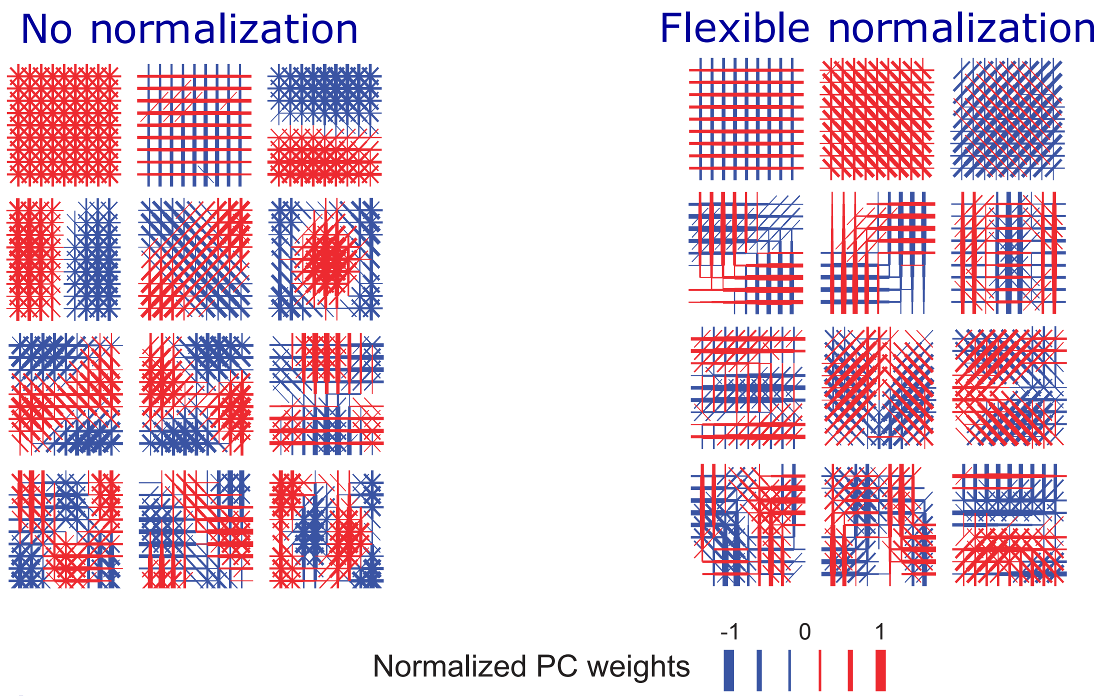
V1 model units

Nonlinear transform
(e.g., flexible divisive normalization)

Cagli, Schwartz, 2013

Optimal normalization in first layer can help unsupervised learning of next layer

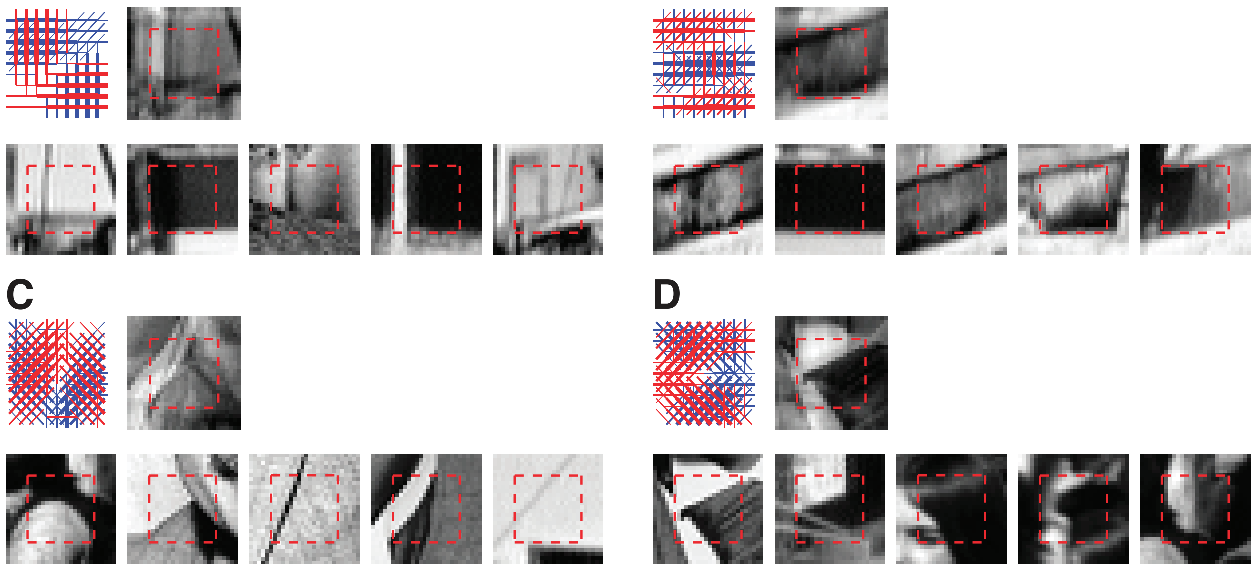
- Flexible normalization in V1 model units results in more sophisticated V2 units than with standard or no normalization



Cagli, Schwartz, 2013

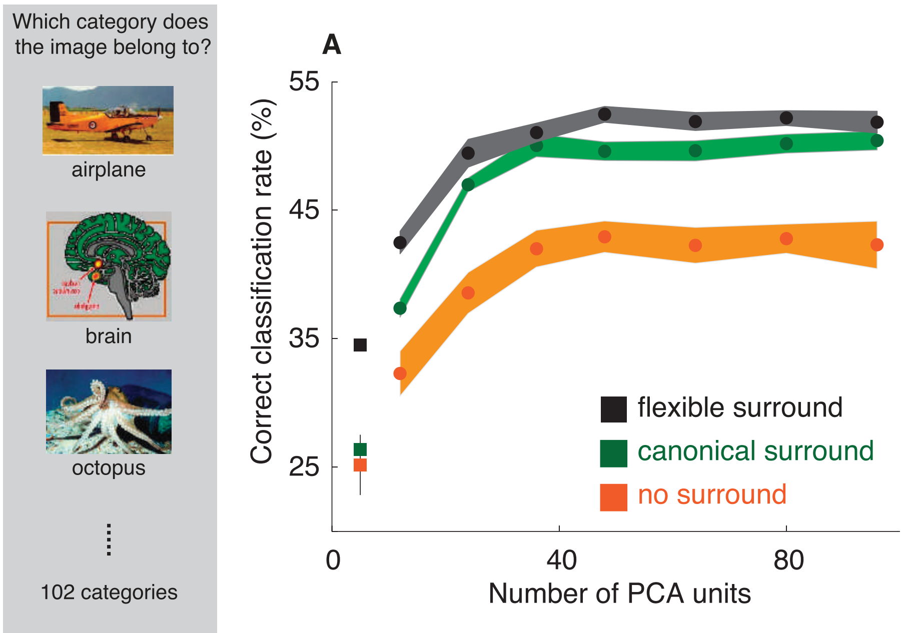
Optimal normalization in first layer can help unsupervised learning of next layer

- Flexible normalization in V1 model units results in more sophisticated V2 units than with standard or no normalization



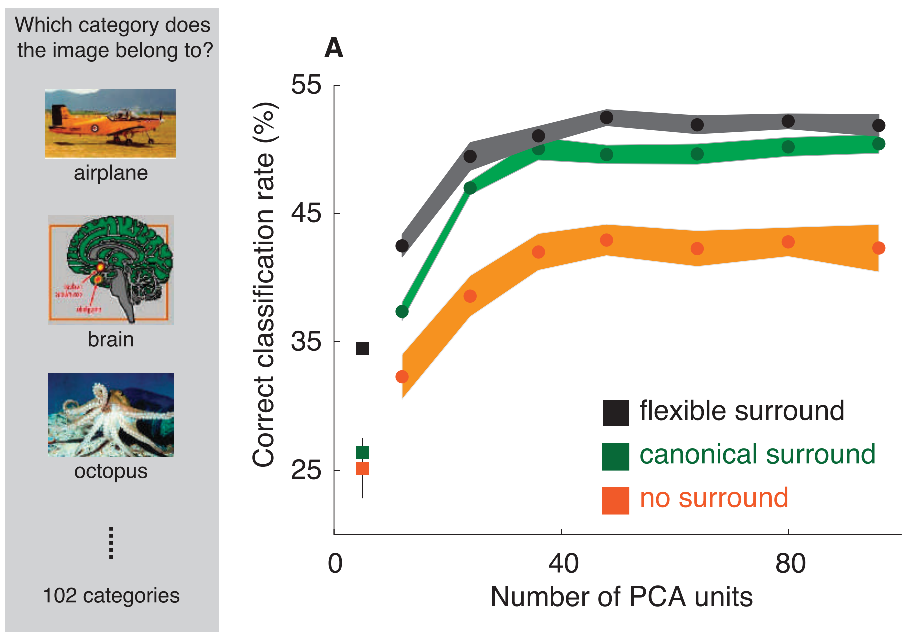
 Cagli, Schwartz, 2013 (also Bowren, Sanchez Giraldo, Schwartz, VSS 2019; see also V2 model of Hosoya, Hyvärinen, 2015)

Flexible normalization and perceptual tasks: recognition



Cagli, Schwartz, 2013

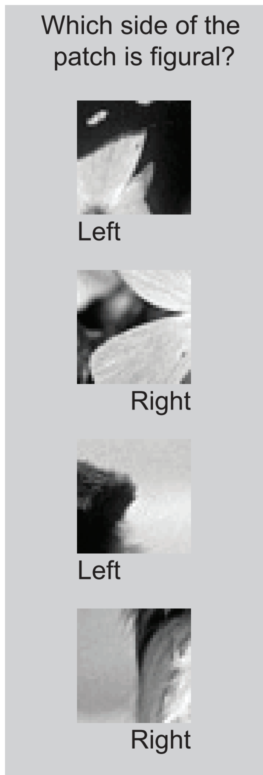
Flexible normalization and perceptual tasks: recognition

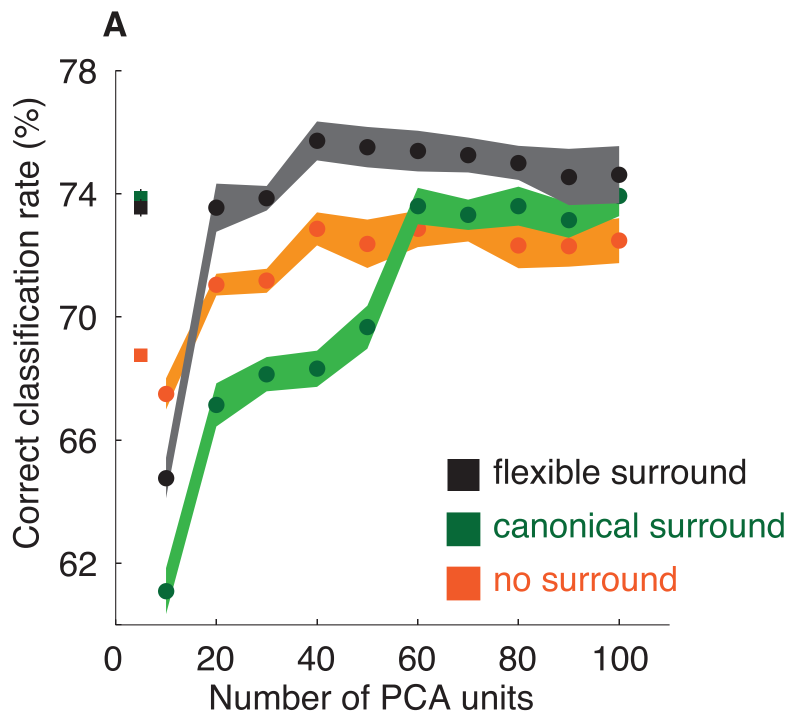


mean probability that center and surround were dependent = 0.78

Cagli, Schwartz, 2013

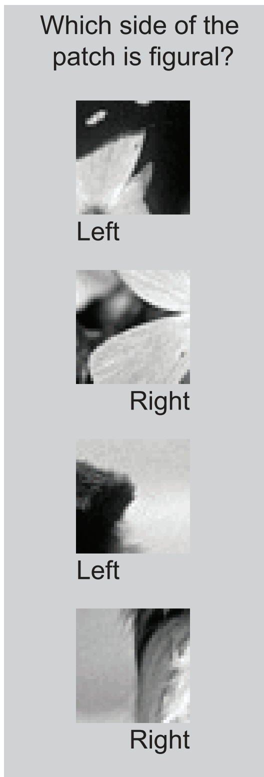
Flexible normalization and perceptual tasks: figure-ground classification

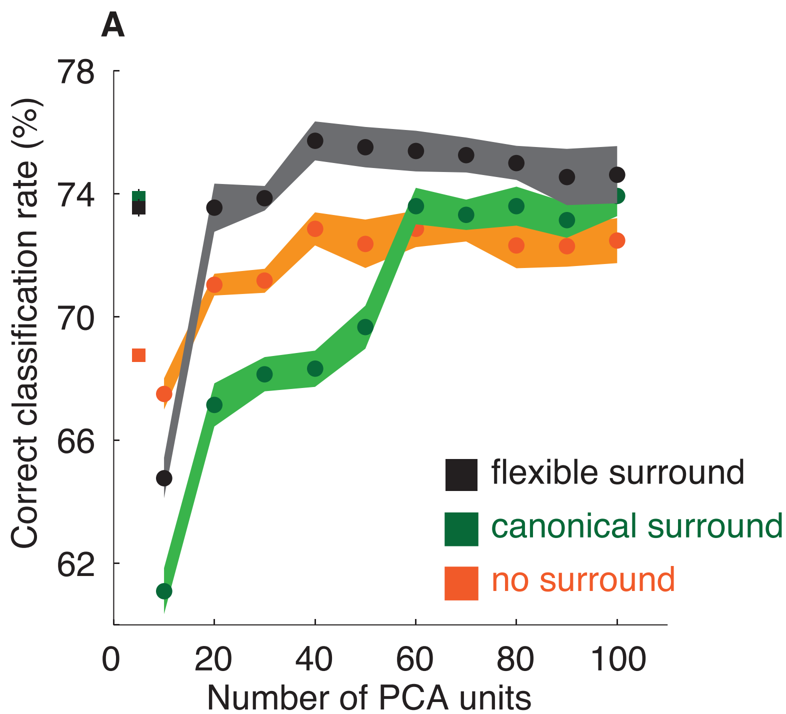




Cagli, Schwartz, 2013

Flexible normalization and perceptual tasks: figure-ground classification





mean probability that center and surround were dependent = 0.62

Cagli, Schwartz, 2013

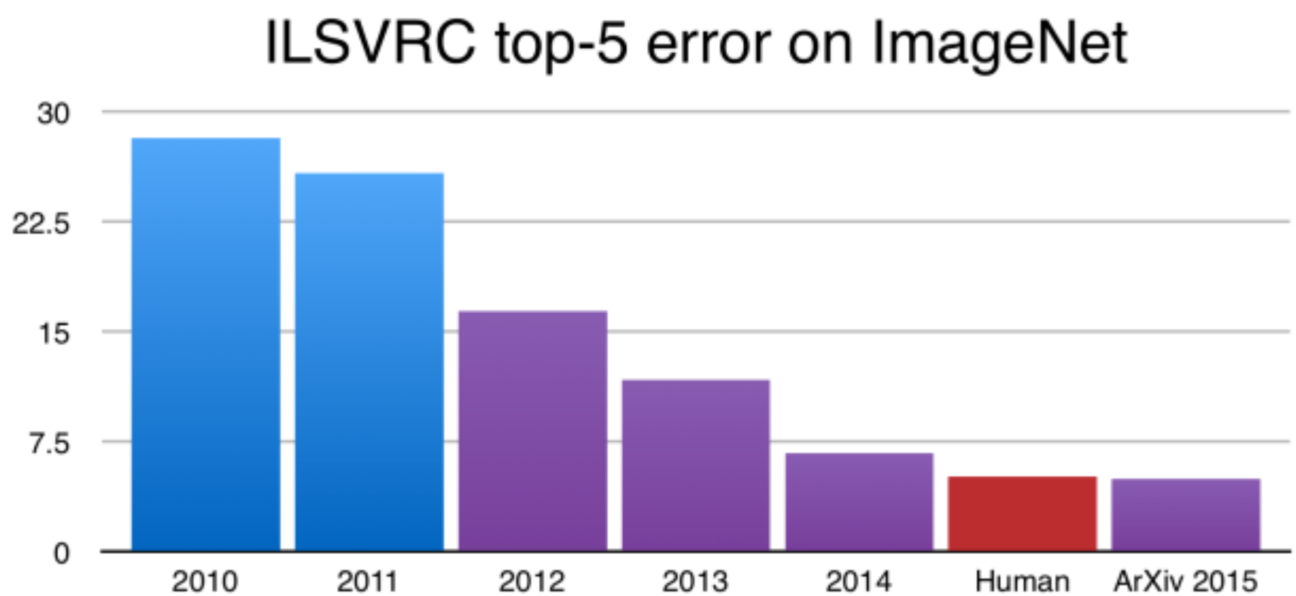
Hierarchical ICA

- Everything we have seen thus far: Unsupervised Learning
- There is no supervision about what object is in the image (eg, car versus tree)

Deep learning and unsupervised

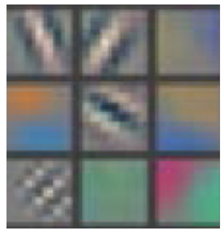
Large scale supervised, discriminative learning has had success in recent years (eg, with Krizhevsky et al. 2012)

"Neural networks are an old idea, so what is new now?"

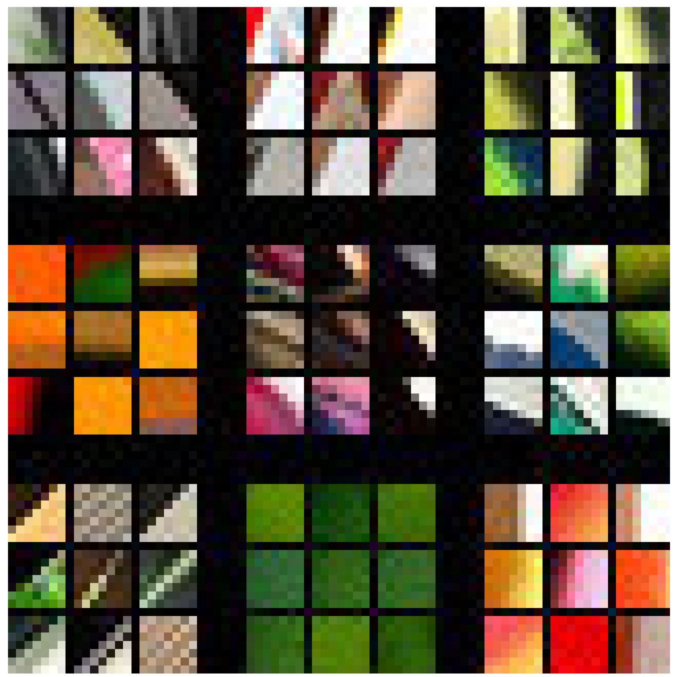


Taken from https://devblogs.nvidia.com/parallelforall/mocha-jl-deep-learning-julia/

Deep networks: supervised more layers

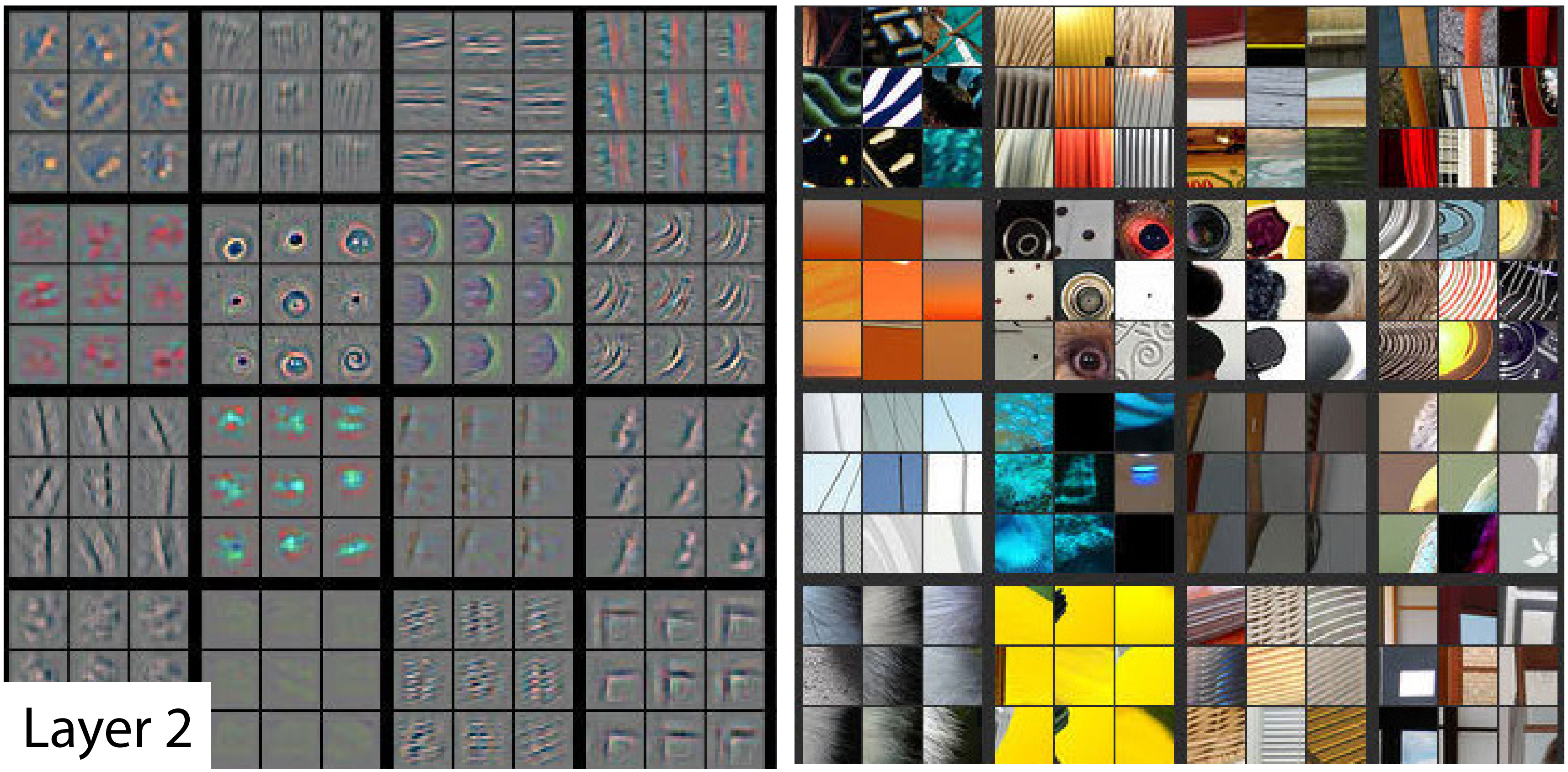


Layer 1



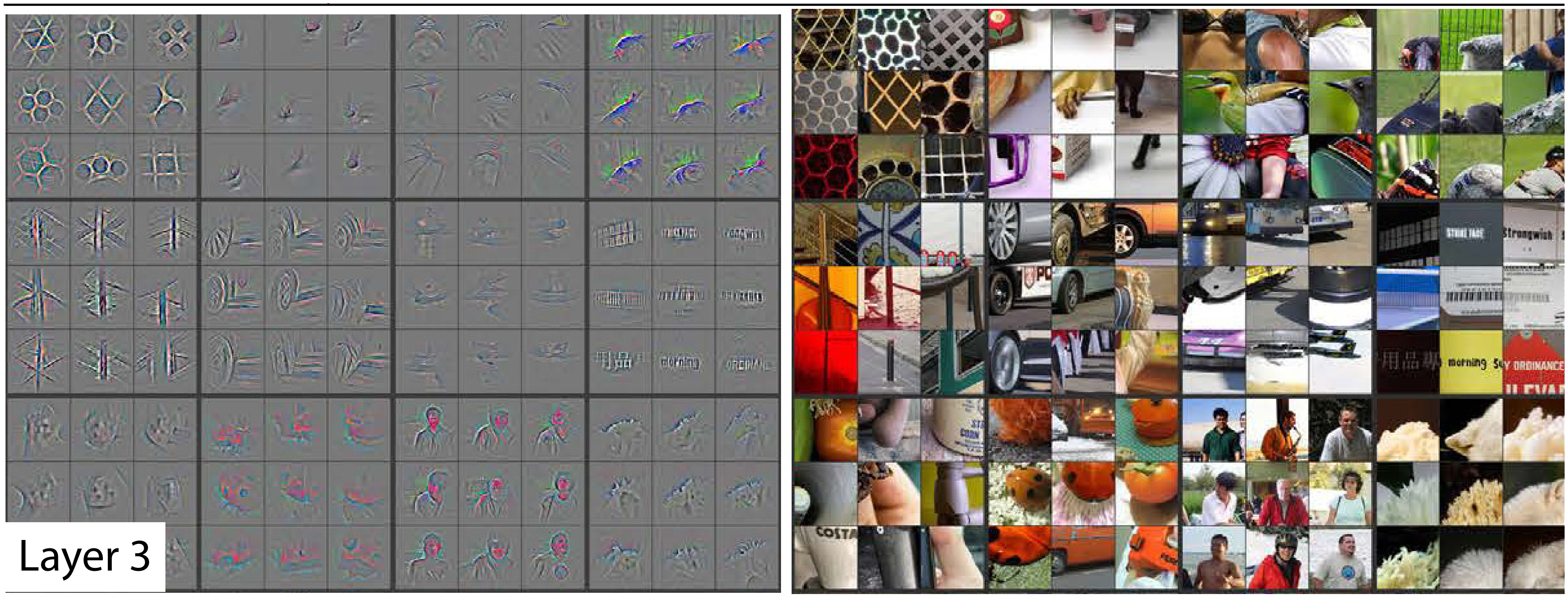
Zeiler, Fergus 2014

Deep networks: supervised more layers



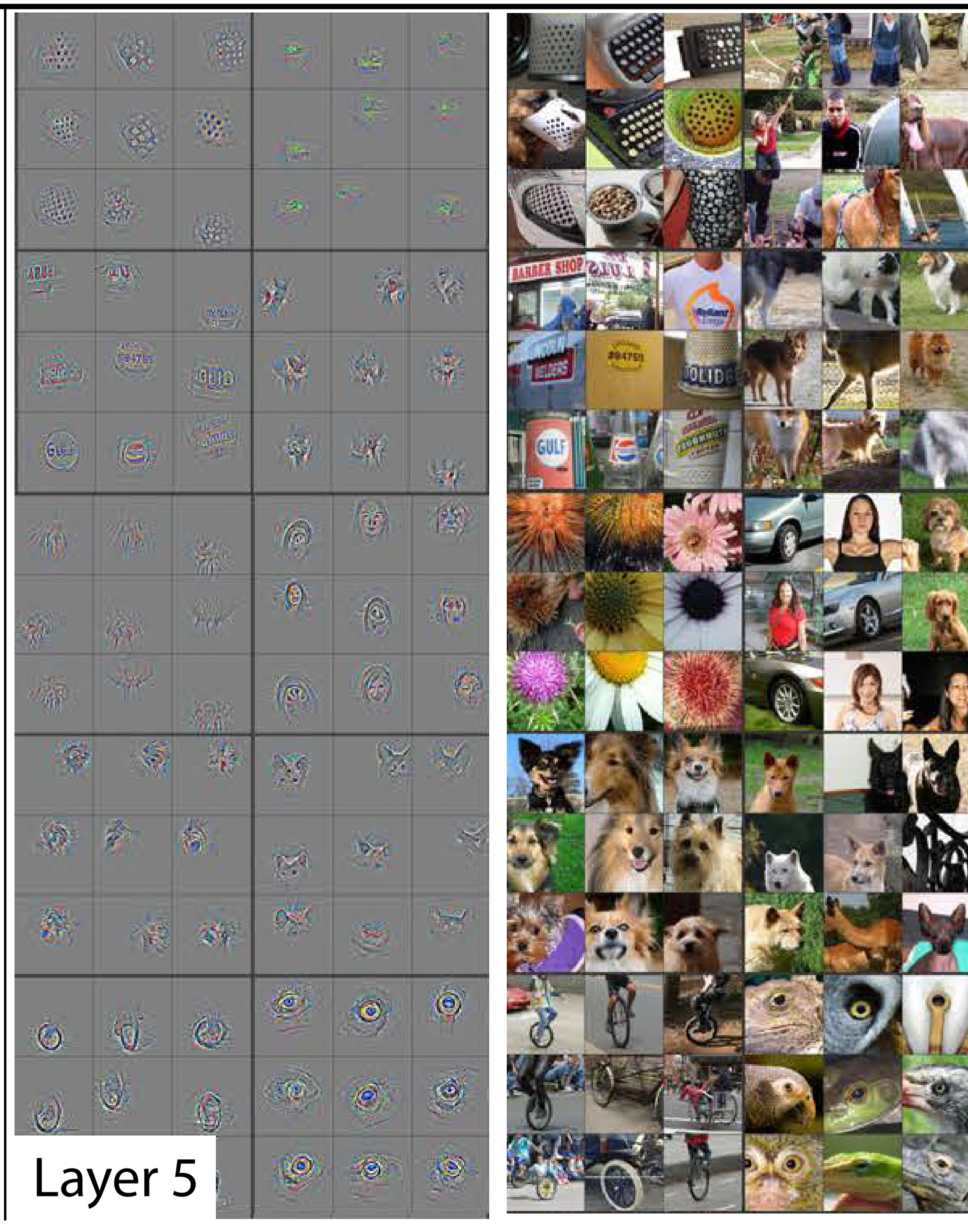
53 Zeiler, Fergus 2014

Deep networks: supervised more layers



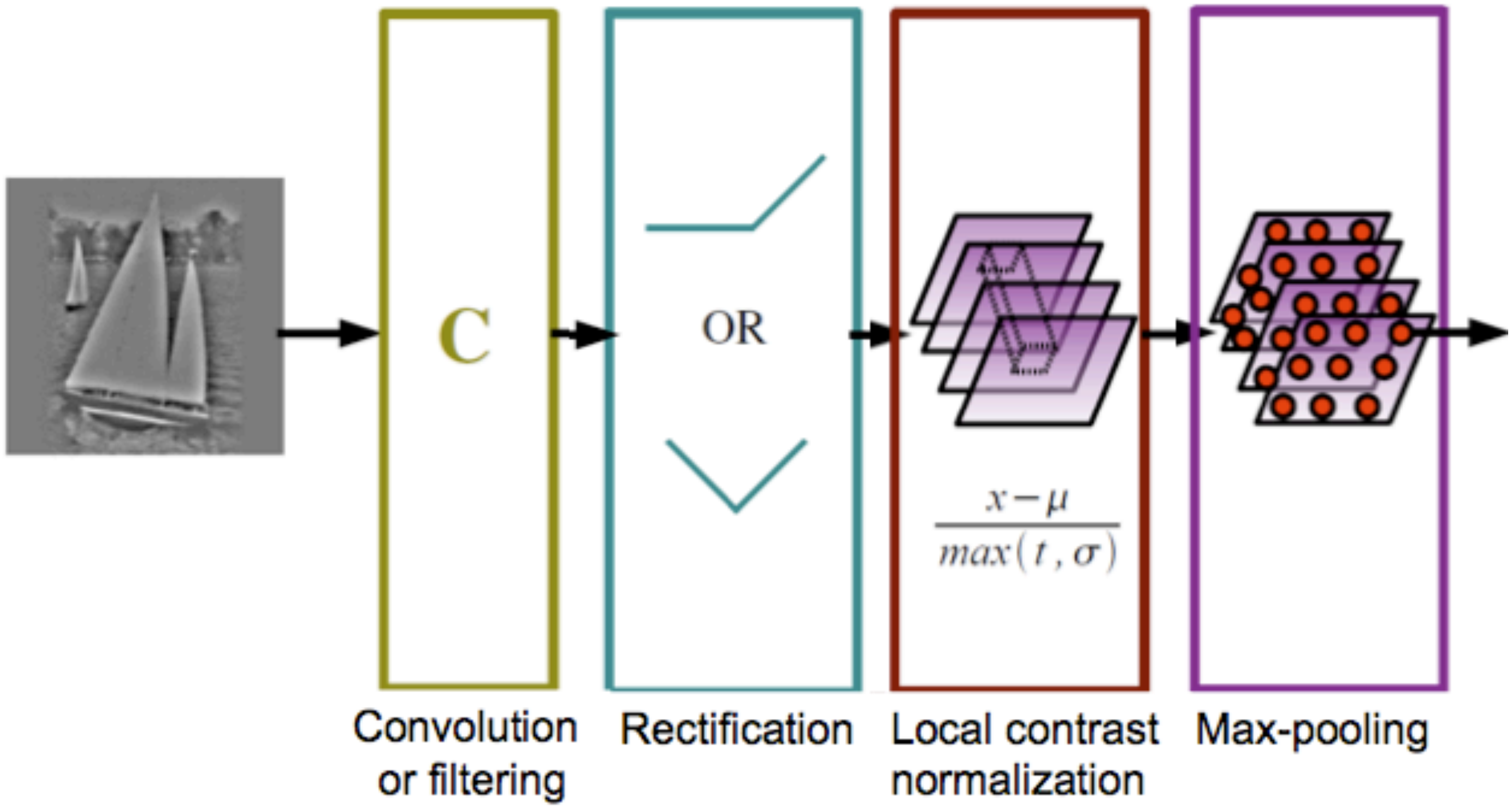
Zeiler, Fergus 2014

Deep networks: supervised more layers



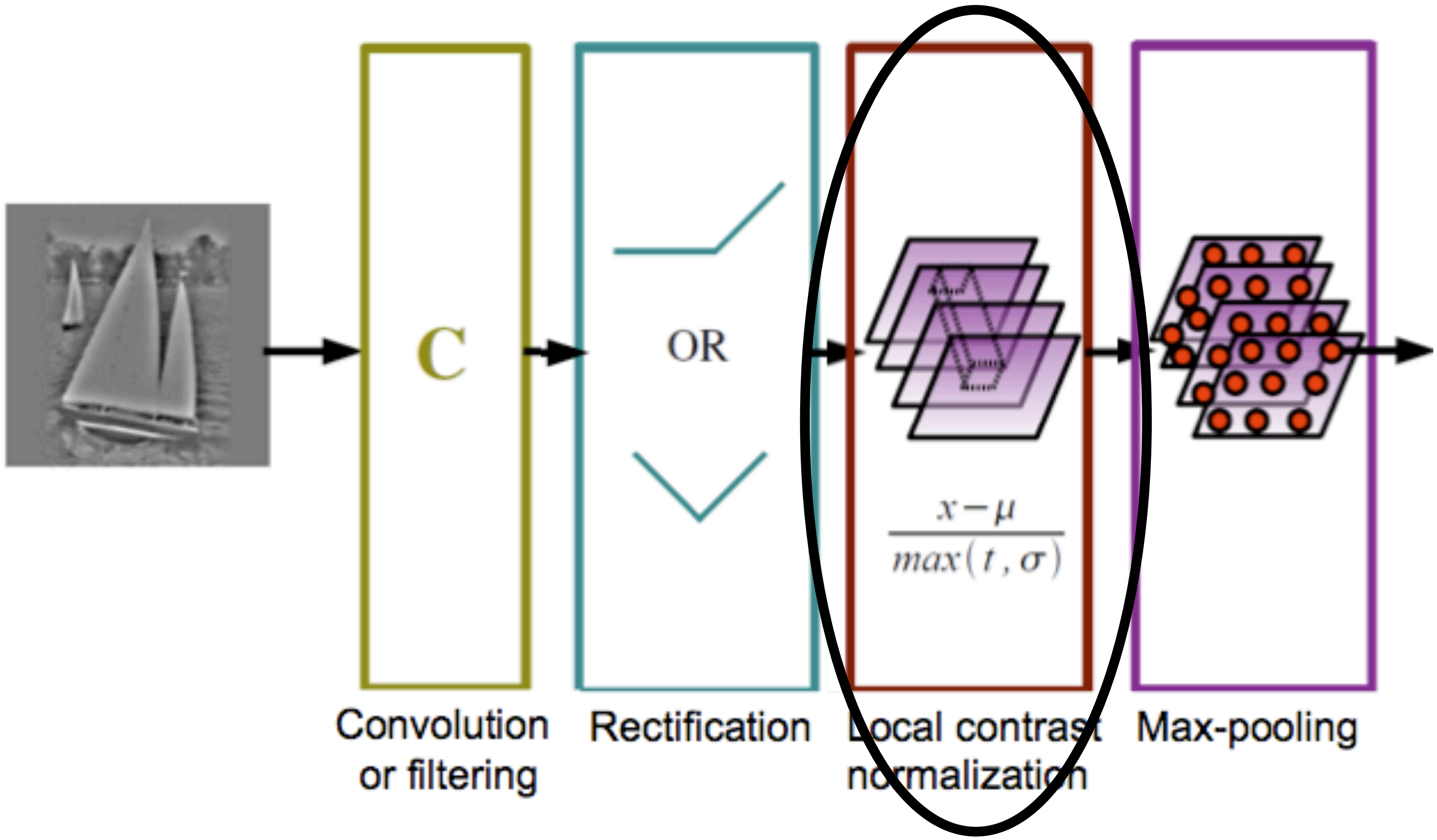
Zeiler, Fergus 2014

Deep networks: nonlinearities



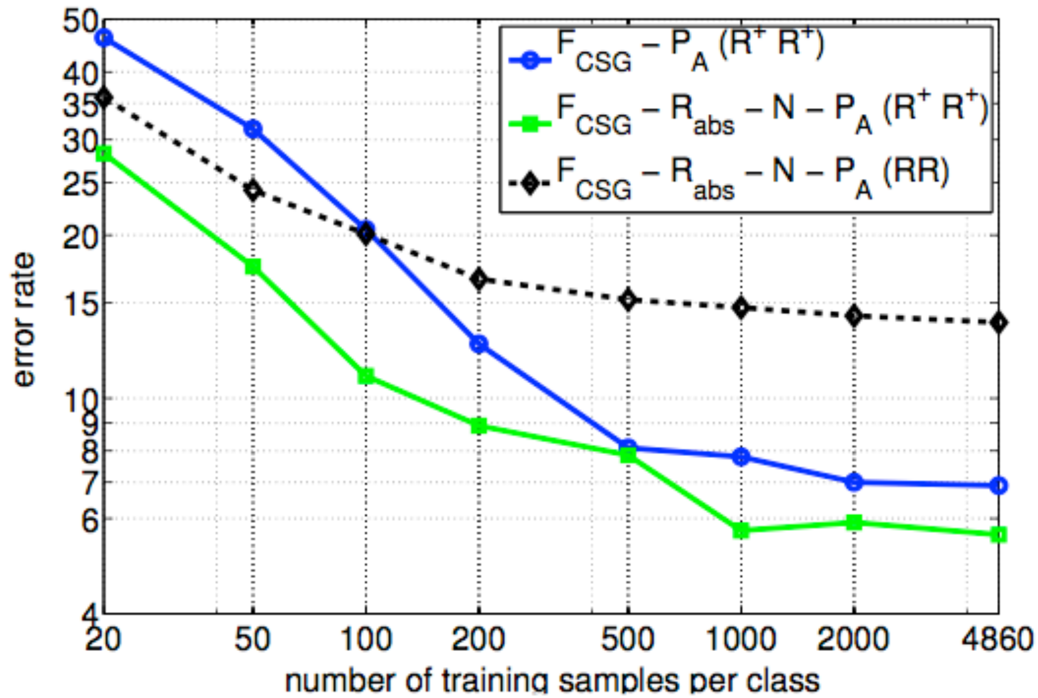
The importance of nonlinearities (From Lee NIPS 2010 workshop; Jarrett, LeCun et al. 2009)

Deep networks: nonlinearities



The importance of nonlinearities (From Lee NIPS 2010 workshop; Jarrett, LeCun et al. 2009)

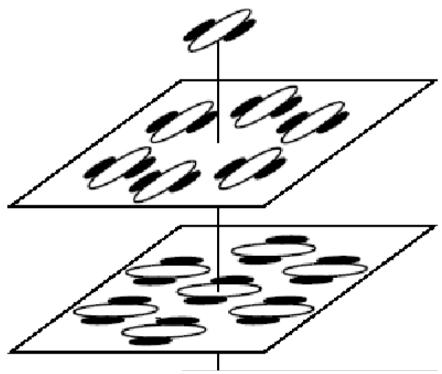
Deep networks: nonlinearities



The importance of nonlinearities (Jarrett, LeCun et al. 2009)

Scene statistics

Modeling filter coordination in images



- Learning how more complex representations build up from the structure of images

- Reducing dependencies further via divisive normalization – linking to spatial context effects (later)