6.819 / 6.869: Advances in Computer Vision

MIT COMPUTER VISION

Basics of Image Processing III:

Image Operations for ConvNet & Image/Dataset Statistics

Instructor: Aude Oliva

Lecture TR 9:30AM – 11:00AM
(Room 34-101)

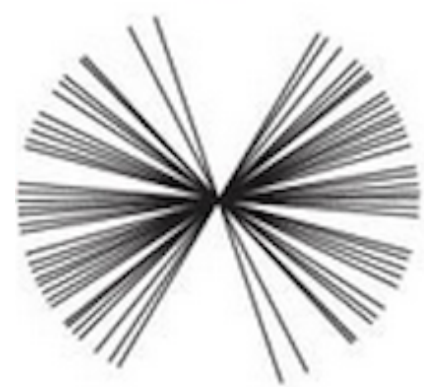
Website:
http://6.869.csail.mit.edu/fa15/

Selective rearing experiment

Blackmore & Cooper (1970) Development of the brain depends on the visual environment

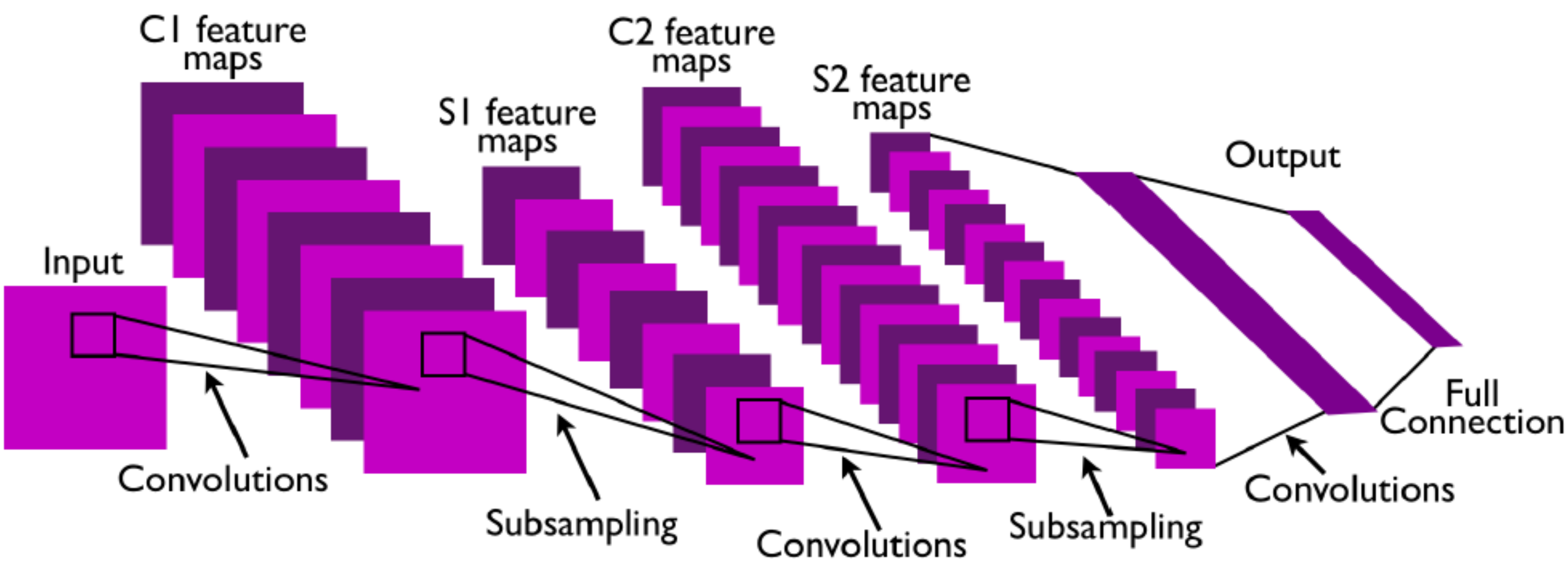
Vertically reared cat

Horizontally reared cat



Distribution of optimal orientations for 72 cells in the early visual area of the reared cat

Why did we learn these image operations?

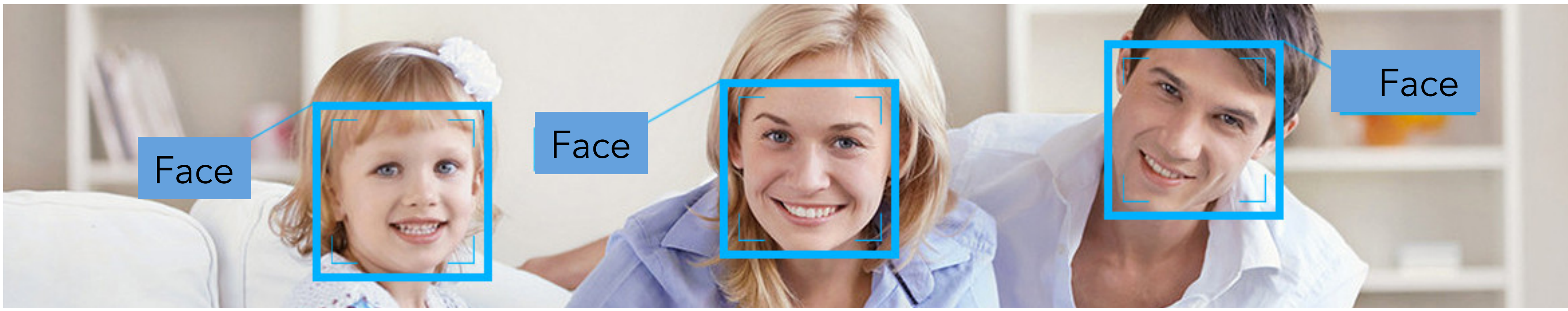


A ConvNet architecture with two feature maps

From Aysegul Dundar's lecture

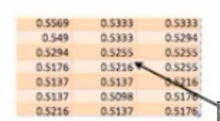
How to detect a face?

Introduction to Convolutional Neural Network Image Operations



You need two groups of Images
Faces and Background (no faces)

How can it detect a face?

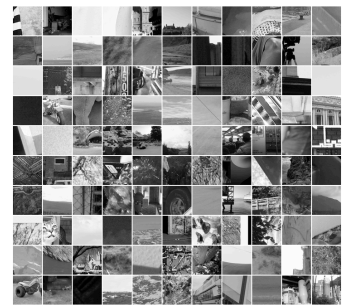


0.5569	0.5333	0.5333
0.549	0.5333	0.5294
0.5294	0.5255	0.5255
0.5176	0.5216	0.5255
0.5137	0.5137	0.5216
0.5137	0.5098	0.5176
0.5216	0.5137	0.5176

32 x 32 input images

Faces

Background



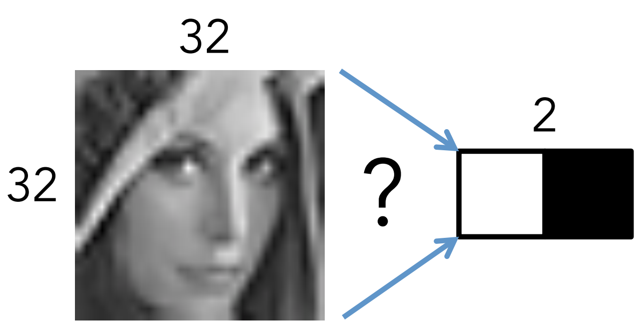
Output from the convnet

+1 -1

First category is white: face

-1 +1

Second category is white: background



Training: give the input and labeled output so it can explore the **features** that make a face look like a face

Let's go through the network

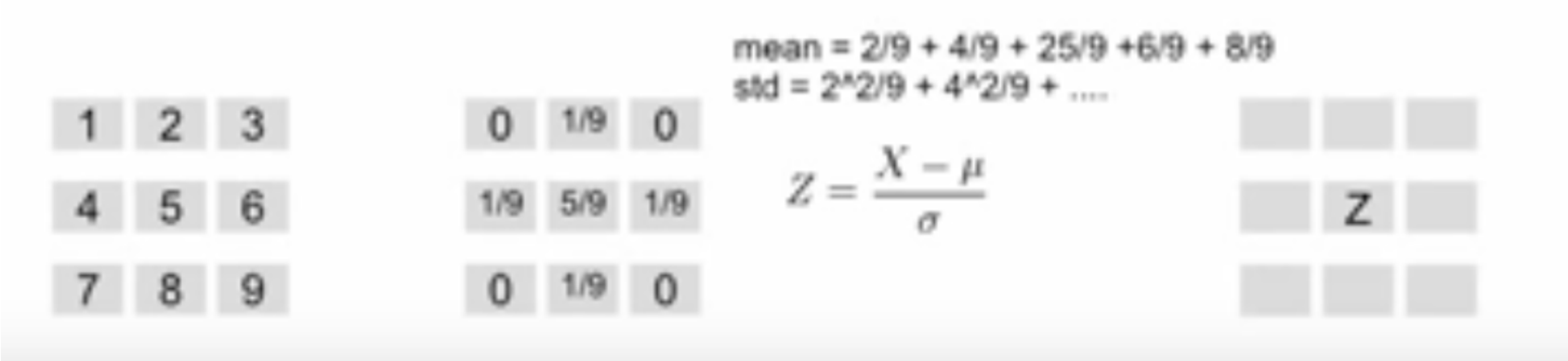
32

32

1st layer

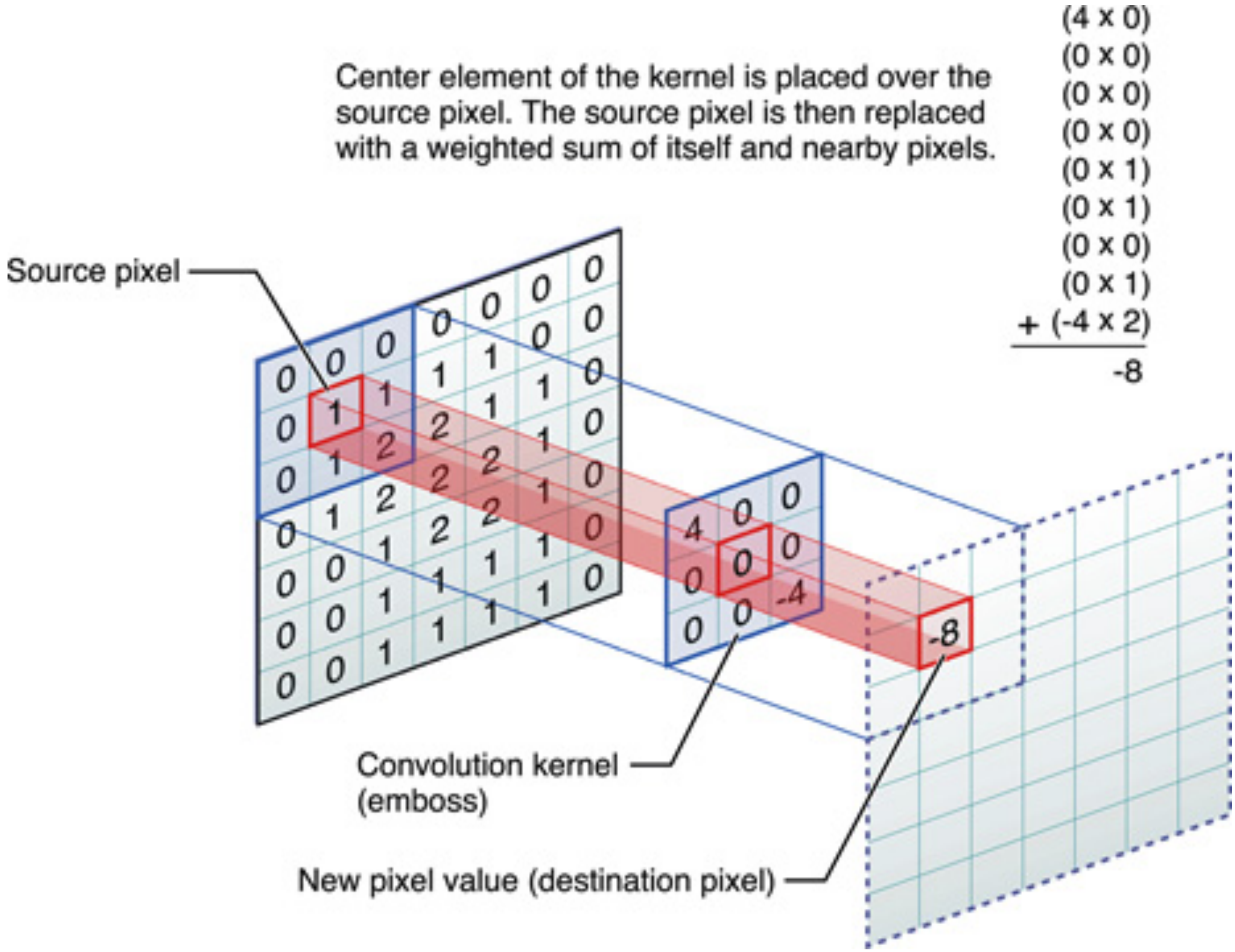
Spatial contrast normalization

Y. Lecun, I. Kanter, and S. Solla, "Eigenvalues of covariance matrices: Applications to neural network learning," Phys. Rev. lett., vol 67, no 18, pp. 669-687, Aug. 1993

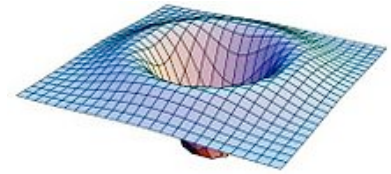


http://setosa.io/ev/image-kernels/

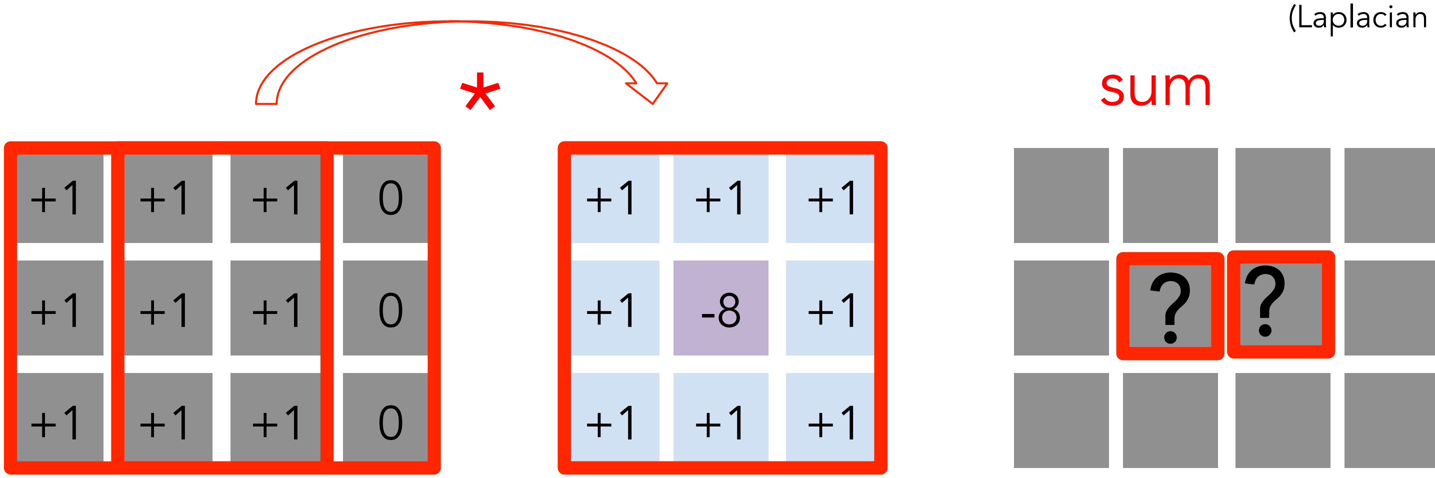
Convolution



Convolution Example



Mexican hat edge detector filter (kernel)
(Laplacian of Gaussian filter)

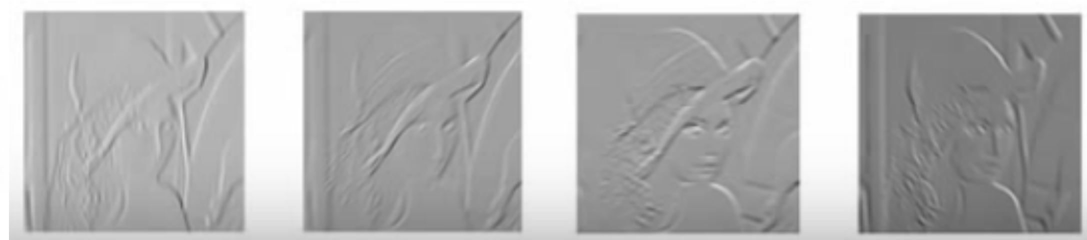


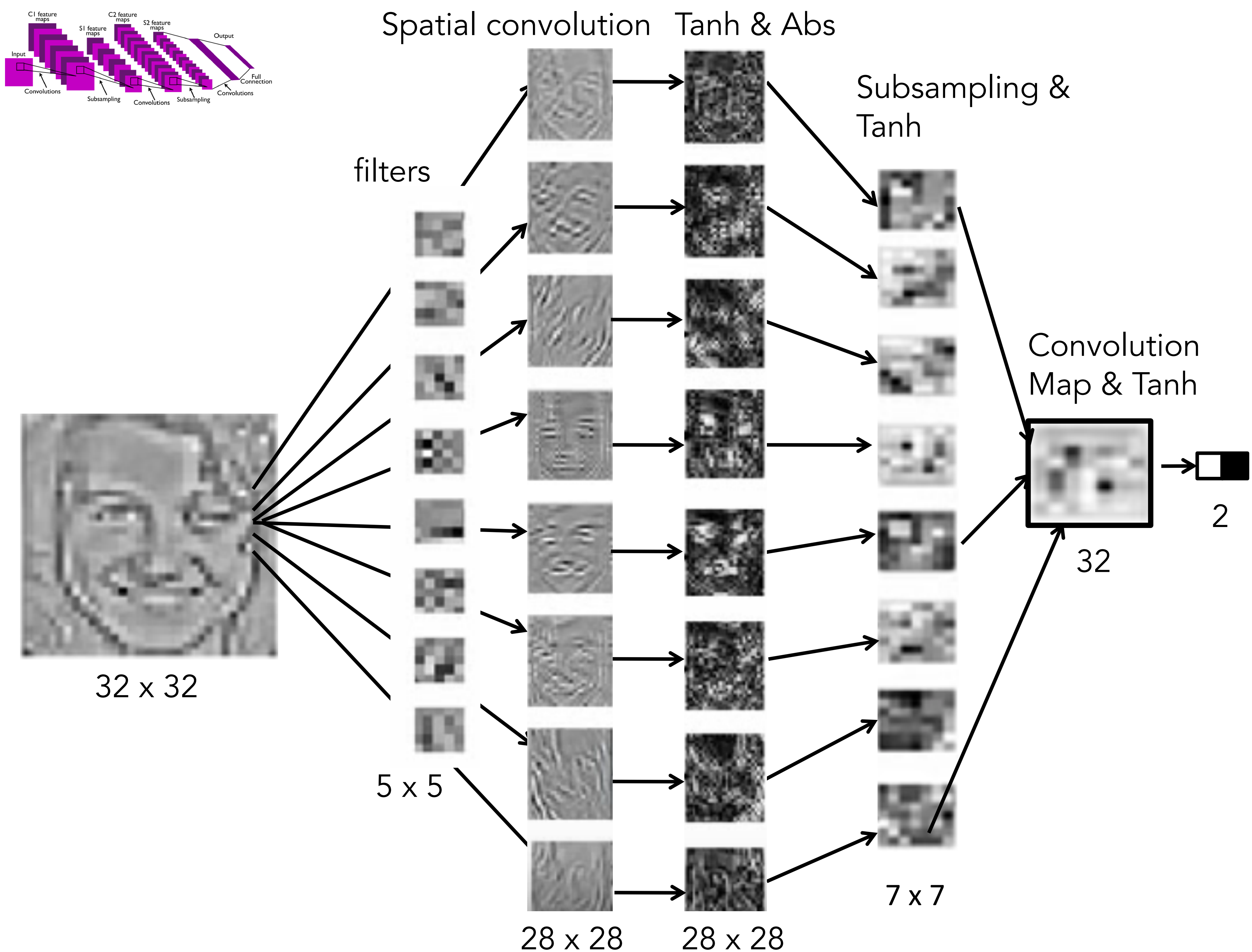
Input image

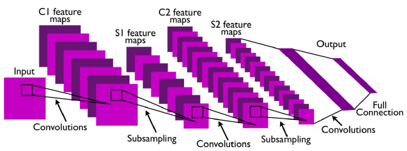
Kernel

Output

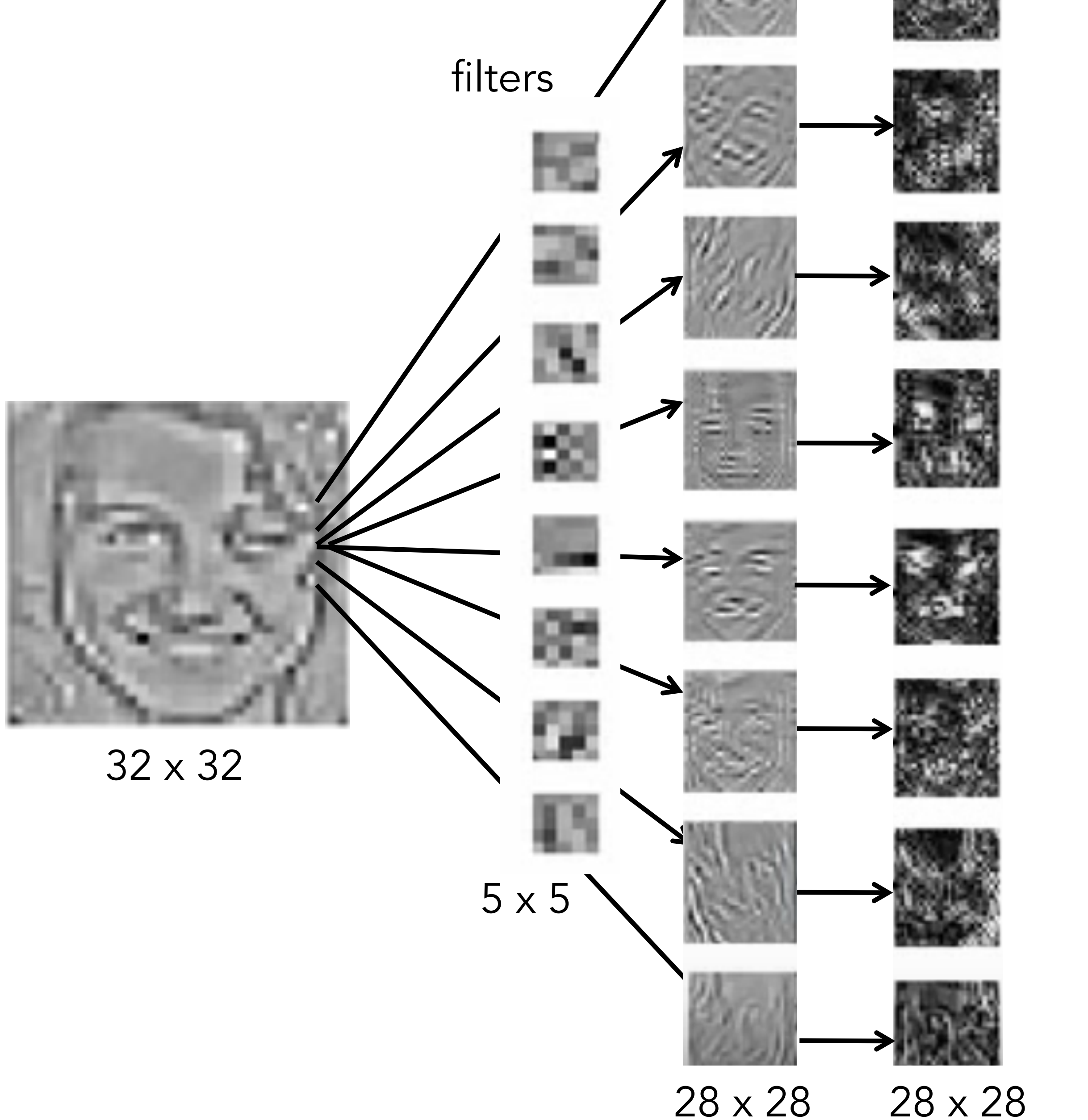
Gabor filters



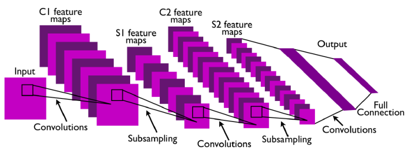

C1 feature maps
S1 feature maps
C2 feature maps
S2 feature maps
Input
Output
Full Connection
Convolutions
Subsampling
Convolutions
Subsampling
Convolutions
Spatial convolution
Tanh & Abs
Subsampling & Tanh
filters
Convolution Map & Tanh
32 x 32
5 x 5
28 x 28
28 x 28
7 x 7
32
2

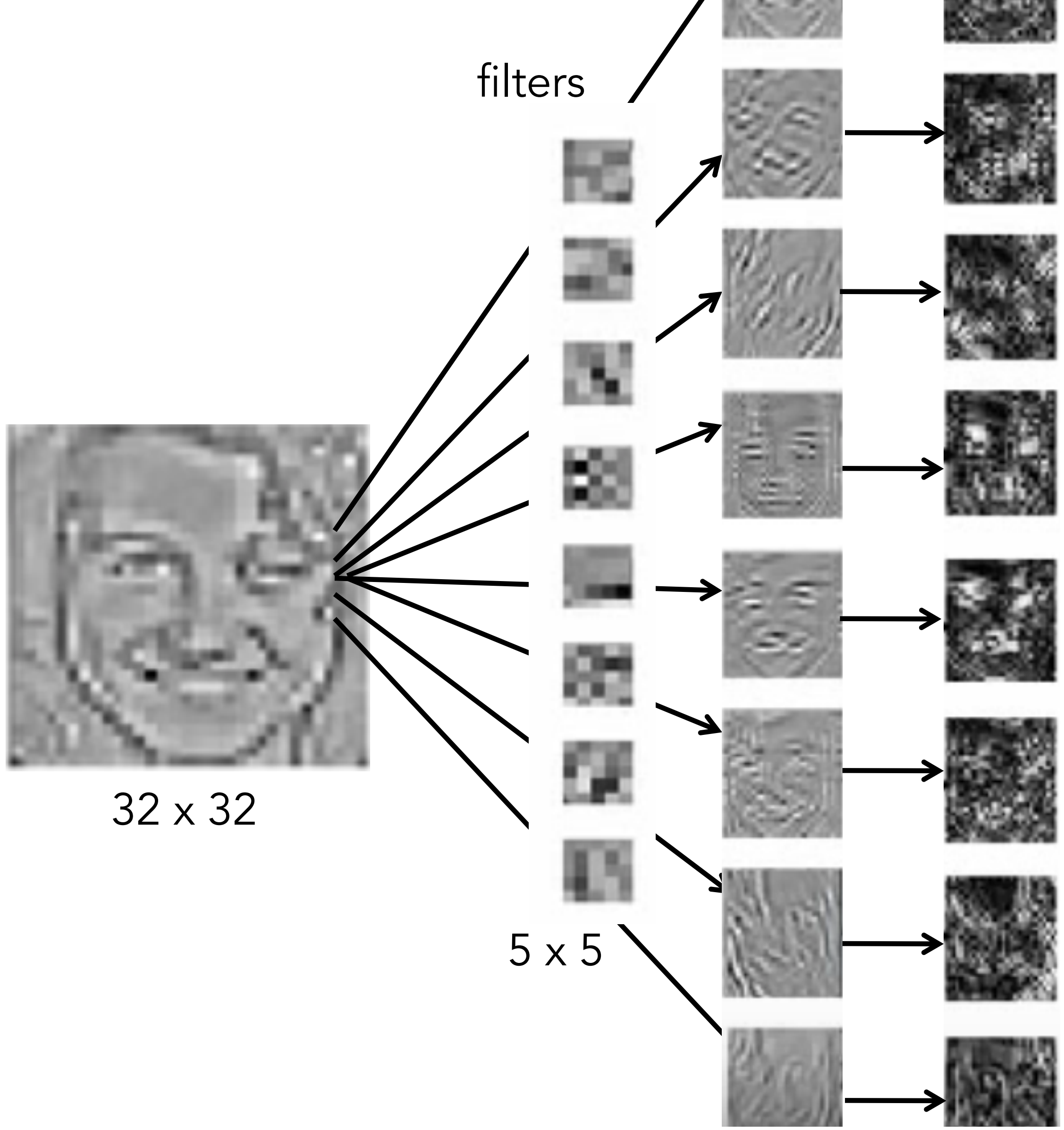


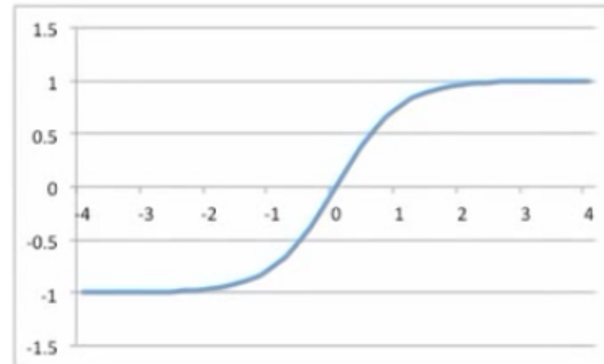
C1 feature maps
S1 feature maps
C2 feature maps
S2 feature maps
Output
Input
Convolutions
Subsampling
Convolutions
Subsampling
Convolutions
Full Connection



Spatial convolution
Tanh & Abs
filters
32 x 32
5 x 5
28 x 28
28 x 28





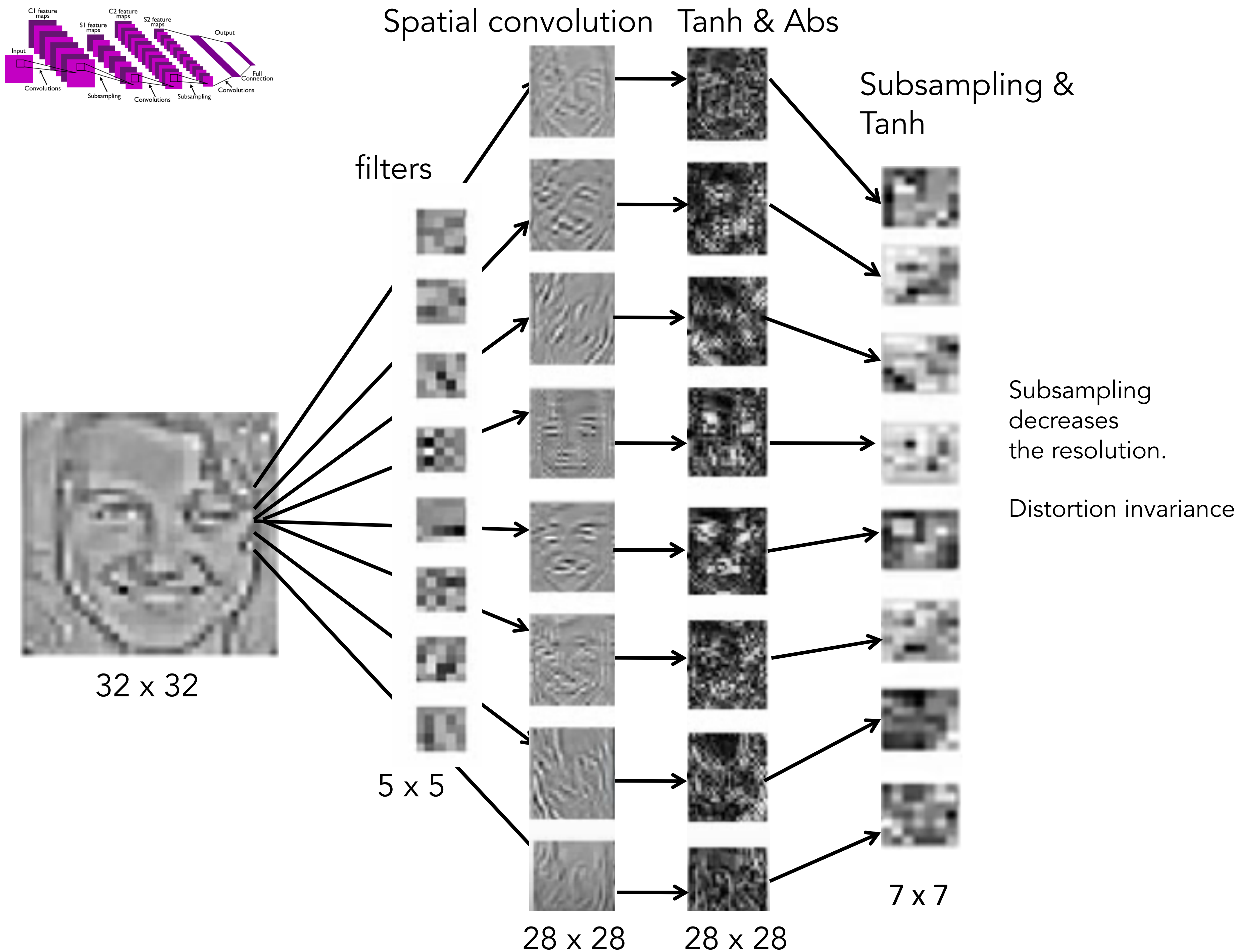


Tanh & Abs = very important
Factor to improve accuracy

a) The polarity of features is often
Irrelevant to recognize parts, objects

b) The rectification eliminates
cancellations between neigh-boring
filter outputs when combined with
average pooling

(Biologically plausible)



C1 feature maps
S1 feature maps
C2 feature maps
S2 feature maps
Output
Input
Full Connection
Convolutions
Subsampling
Convolutions
Subsampling
Convolutions
Spatial convolution
Tanh & Abs
Subsampling &
Tanh
filters
32 x 32
5 x 5
28 x 28
28 x 28
7 x 7
Subsampling
decreases
the resolution.
Distortion invariance

What is Subsampling?

508 x 508 pixels

down-sampled by 2 x 2

254 x 254 pixels

Spatial subsampling:
1) Compute the average
2) Multiplies it by a trainable coefficient
1) Adds a training bias

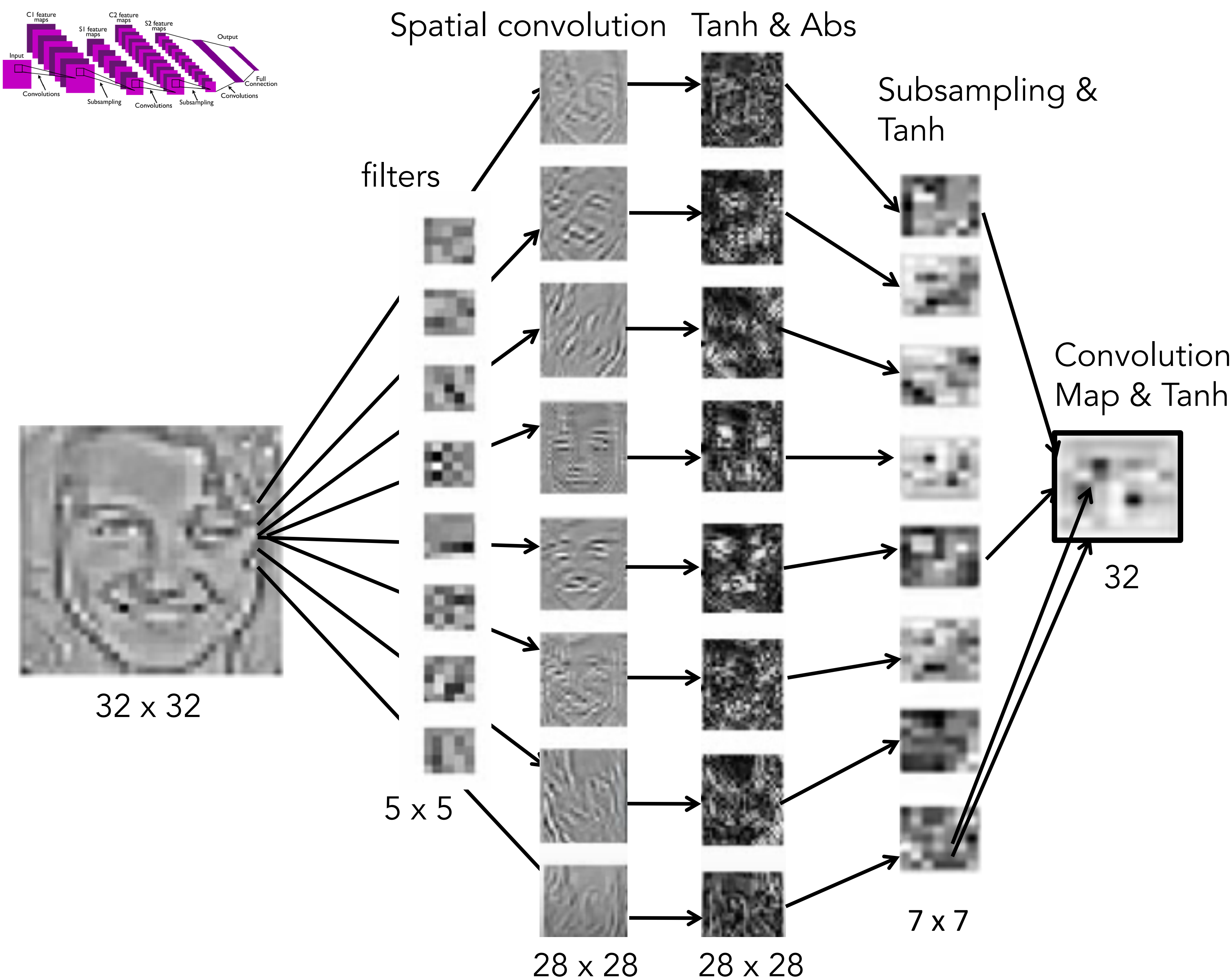
+1	+1
+1	+1

down-sampled by 2 x 2

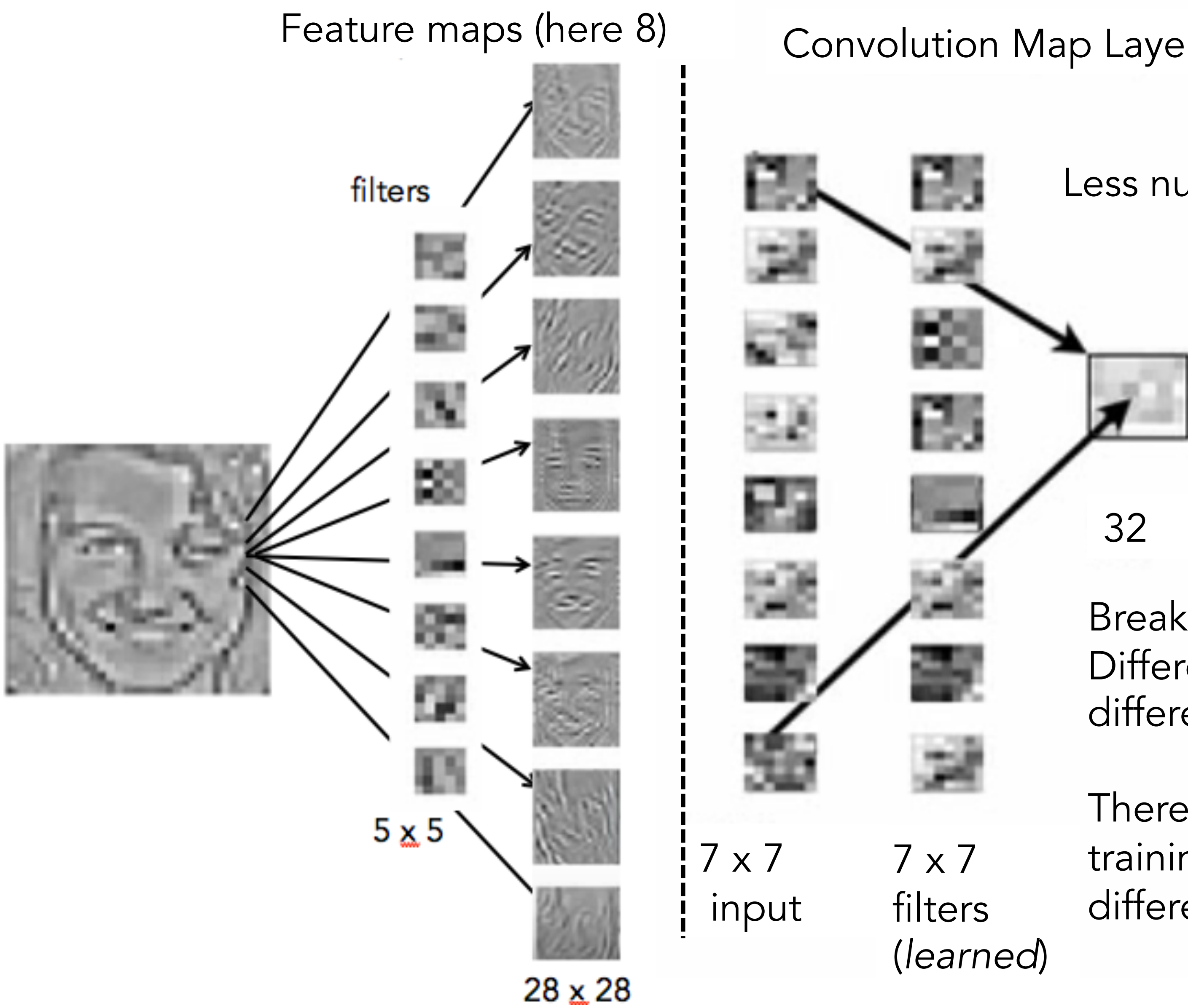
+1 * 0.2 + 0.5 = 0.7

weight

Additive bias


C1 feature maps
S1 feature maps
C2 feature maps
S2 feature maps
Output
Input
Full Connection
Convolutions
Subsampling
Convolutions
Subsampling
Convolutions
Spatial convolution
Tanh & Abs
Subsampling & Tanh
filters
Convolution Map & Tanh
32 x 32
5 x 5
28 x 28
28 x 28
7 x 7
32

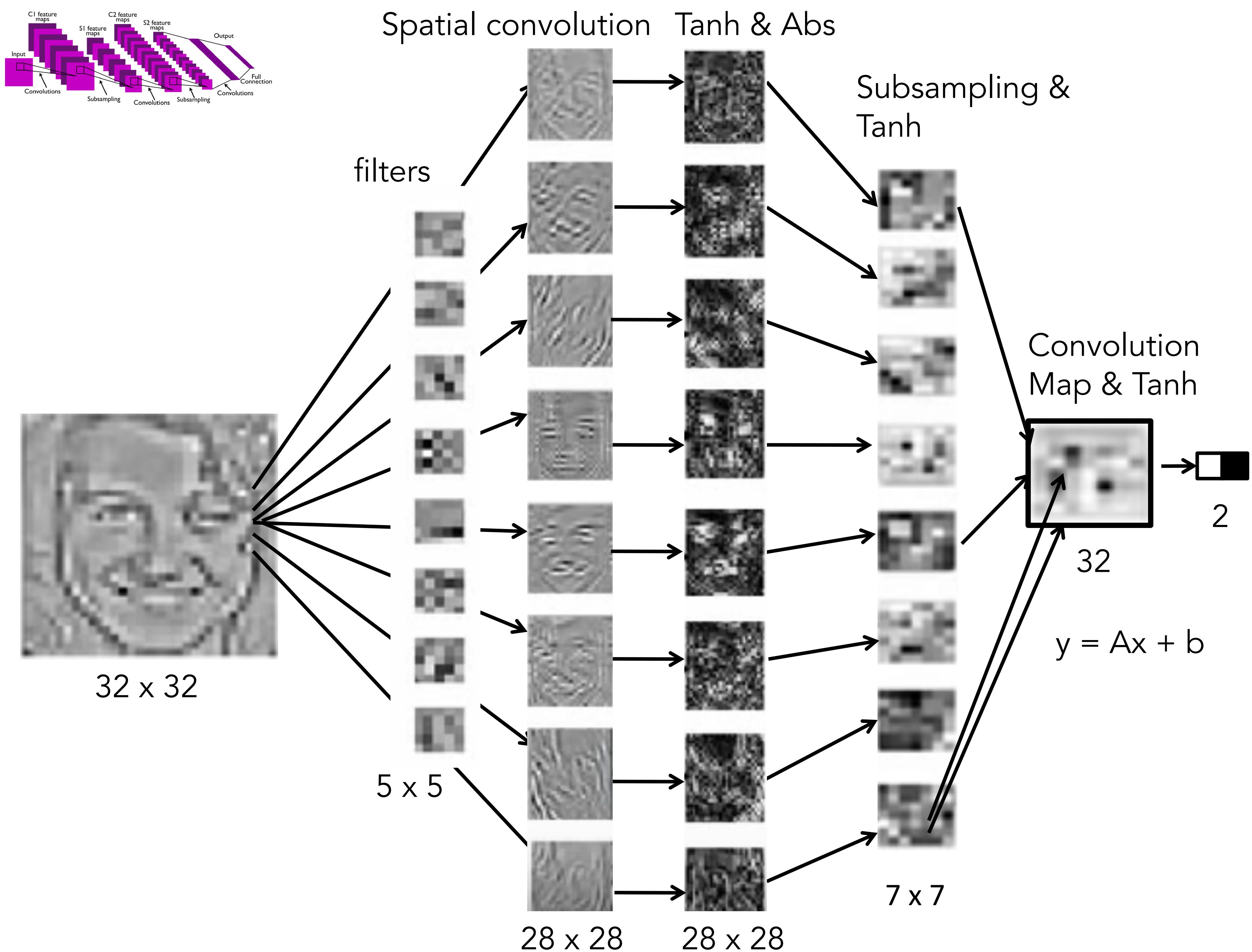
What is Convolution Map?



Less number of connections

Breaks the symmetry:
Different maps receive
different feature maps

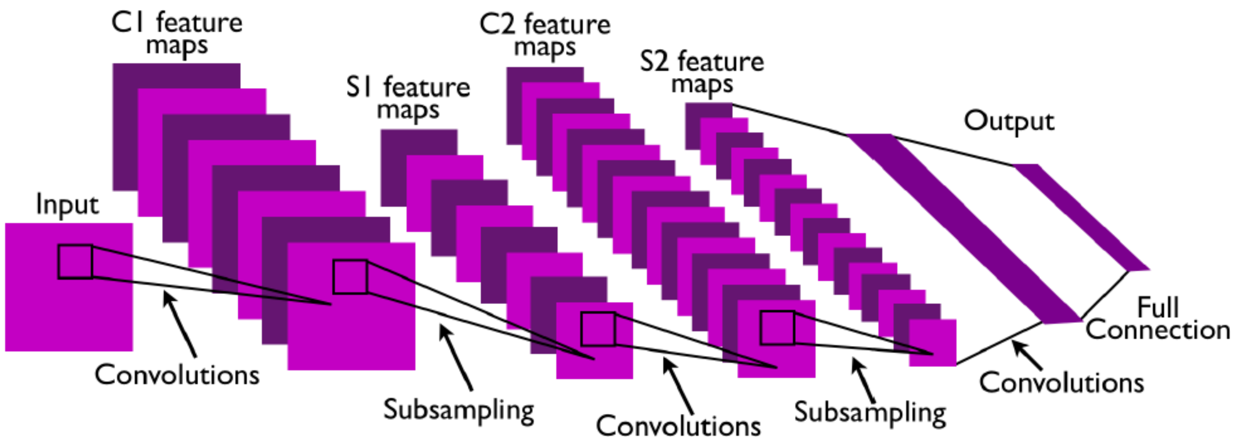
Therefore during the
training, they can explore
different features

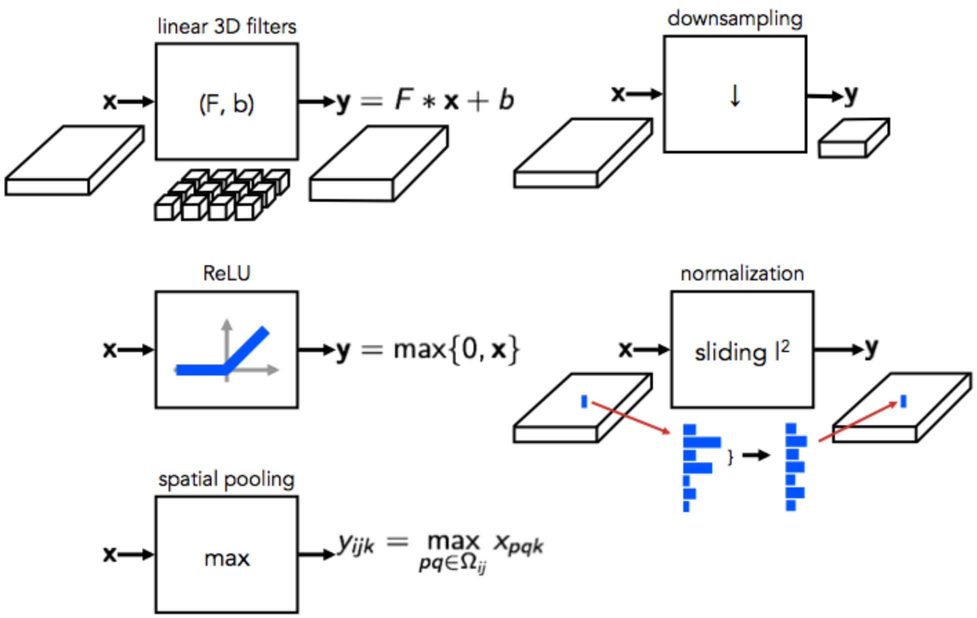

C1 feature maps
S1 feature maps
C2 feature maps
S2 feature maps
Output
Input
Full Connection
Convolutions
Subsampling
Convolutions
Subsampling
Convolutions
Spatial convolution
Tanh & Abs
Subsampling & Tanh
filters
Convolution Map & Tanh
32 x 32
5 x 5
28 x 28
28 x 28
7 x 7
32
2
y = Ax + b

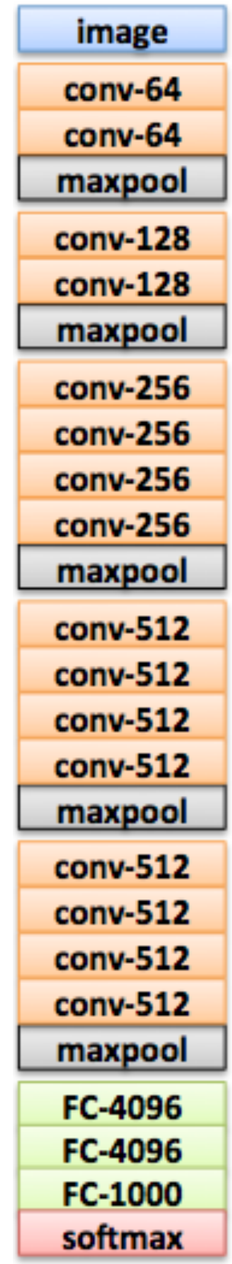
Does the face have to be 32 x 32 ?
Solution: Pyramid

32 x 32
box

CNN: Many components

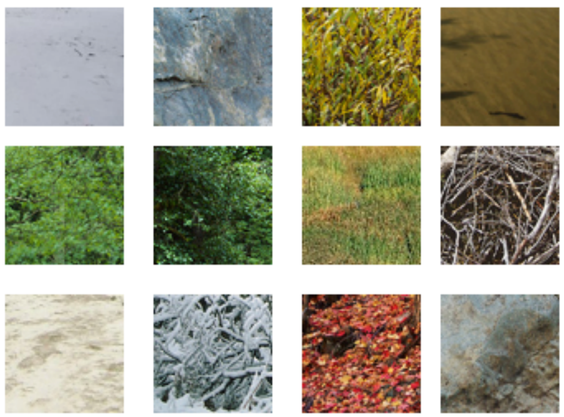


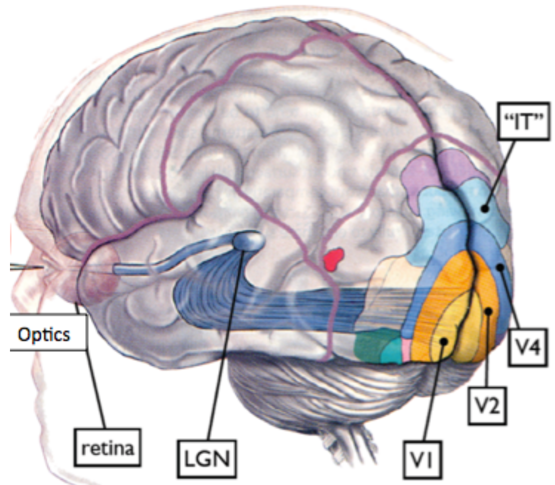


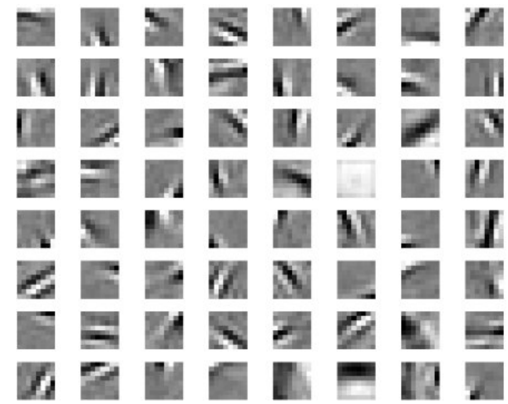


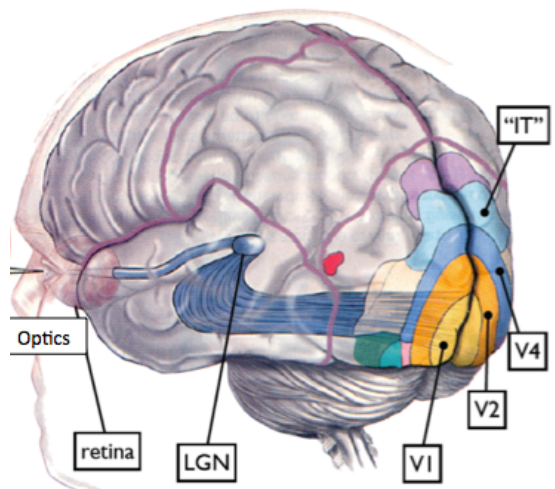
19-layer

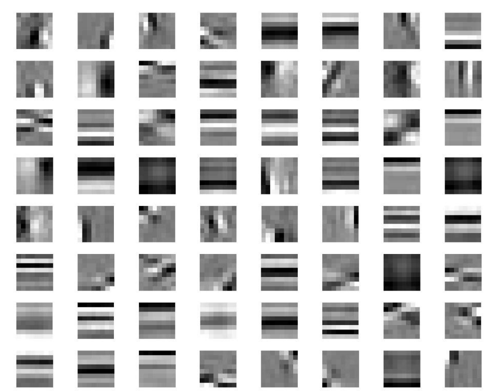
Real world image statistics shape the units' network











I- Low level Natural Image Statistics

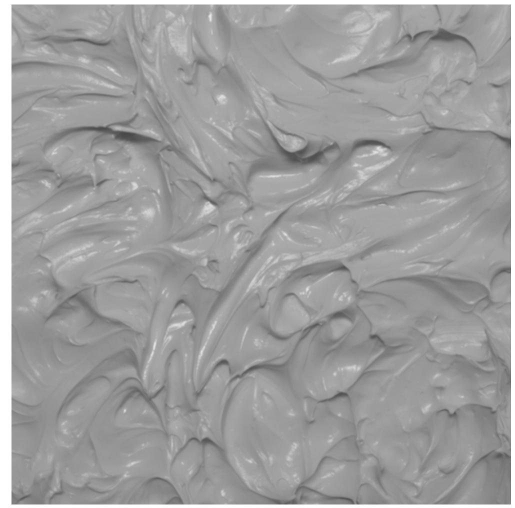
- Every picture is a natural image. But **some processes are going to be more likely than others** in building the structures that one observer is going to see.

- Computational investigations of the statistical structures of natural images suggest that the receptive fields of V1 cells may be optimized for extracting the structure information of natural images

I-1- Material Perception & Object Recognition

Figure 4-1: The bagel on the left and the doughnut on the right have similar shapes and are easy to distinguish. Is this material recognition or object recognition? *(Image source: Flickr)*

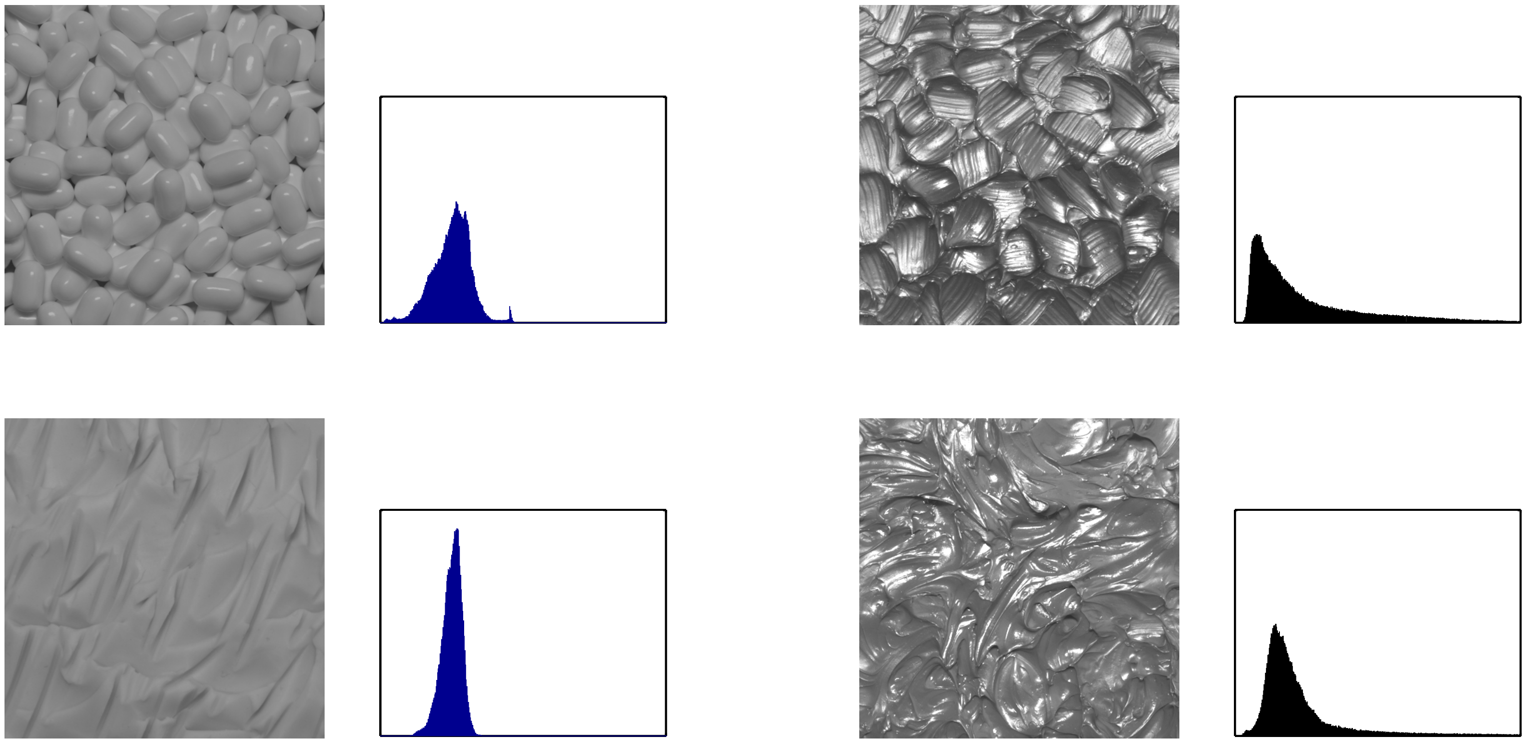
Figure 4-2: The Oreo cookie on the left is made of knit wool whereas the ones on the right are genuine. Both cookies have similar shape and reflectance properties, a fact that may confuse machines but not humans. *(Image source: Flickr)*



You can tell black stucco from white stucco?

Deeper shadows and higher contrast in black surfaces.

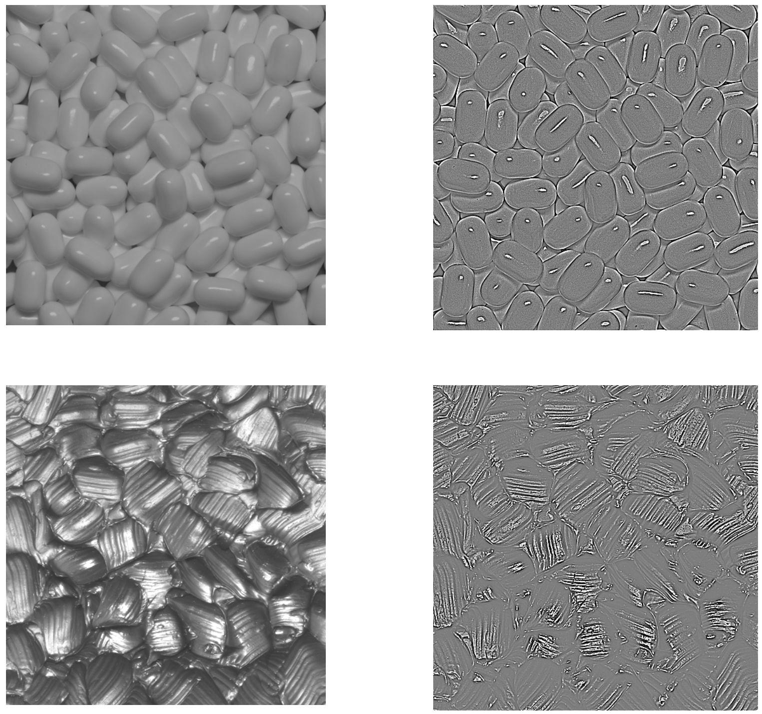
Luminance histograms of white and black surfaces look different

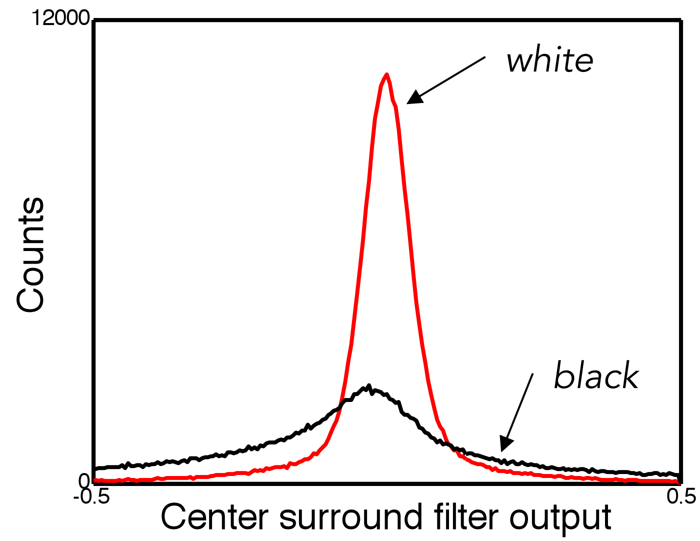


Statistics like moments or percentiles capture the differences in histograms e.g. standard deviation, skewness, 90th percentile.

(c.f. Nishida & Shinya JOSA 1998, Adelson et al VSS 2004, Motoyoshi et al VSS 2005)

Filter outputs look different too.

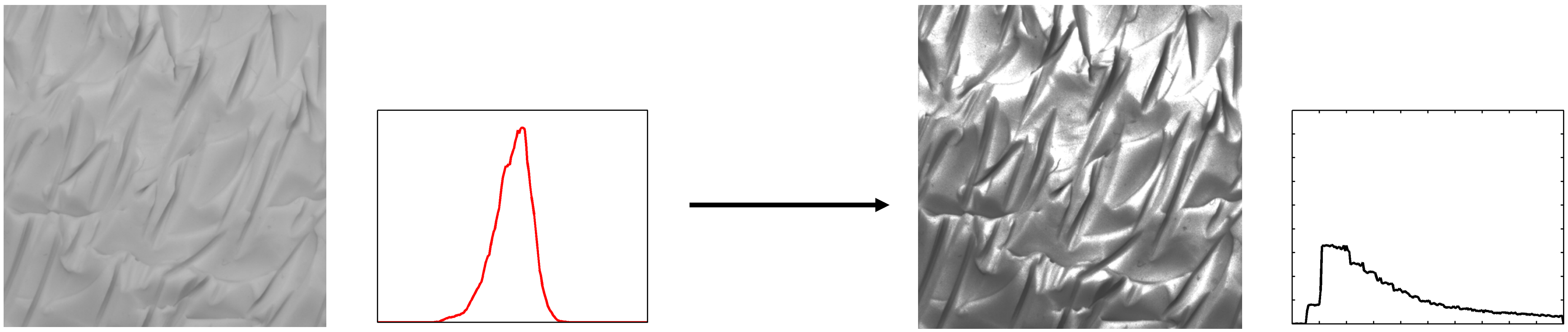




Filters pick up on the deep shadows, bright specularities and higher local contrast of black materials.

Statistics of filter output histogram can be used to discriminate white and black surfaces e.g. standard deviation, skewness, 10^{th} percentile etc.

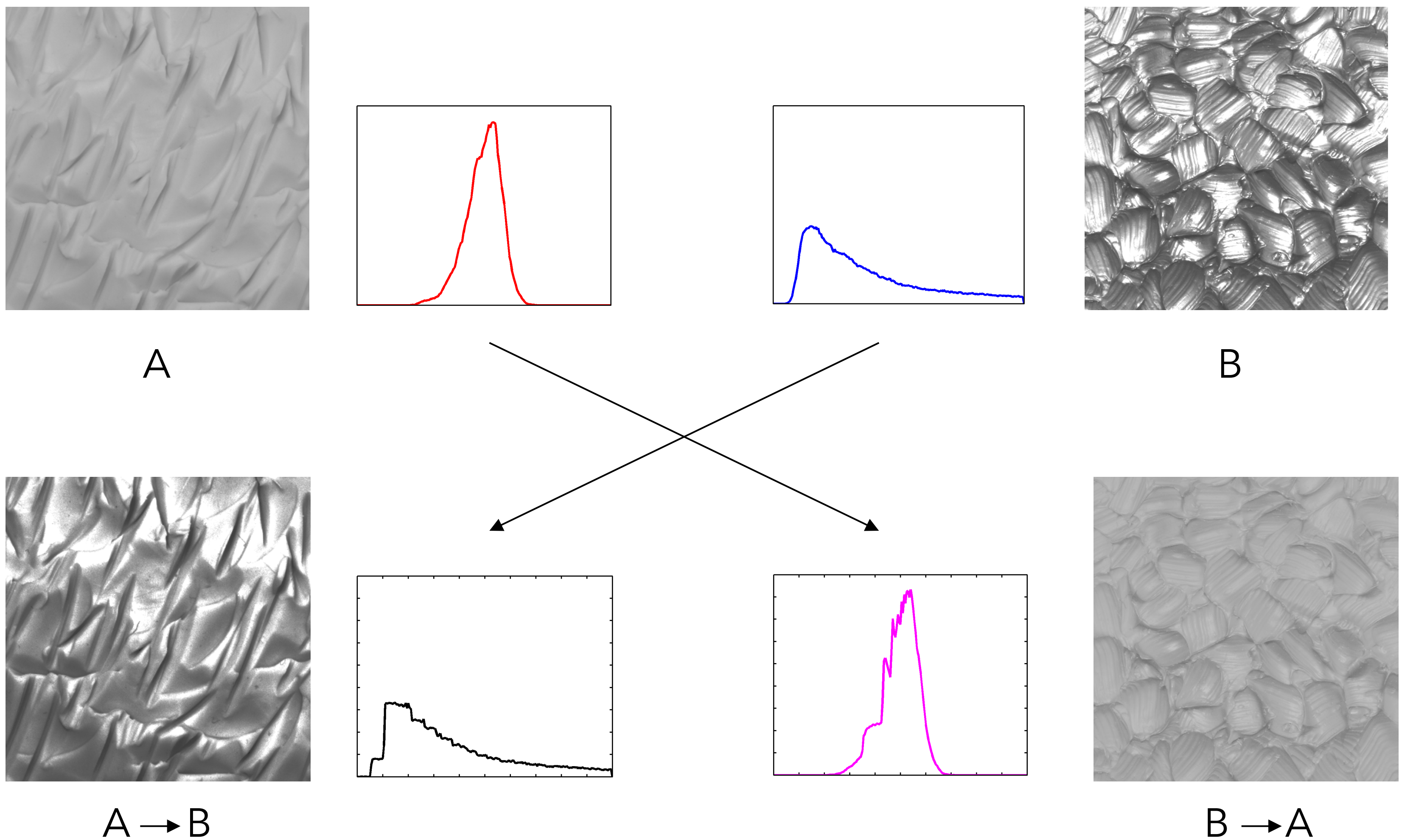
Manipulating histograms changes surface appearance



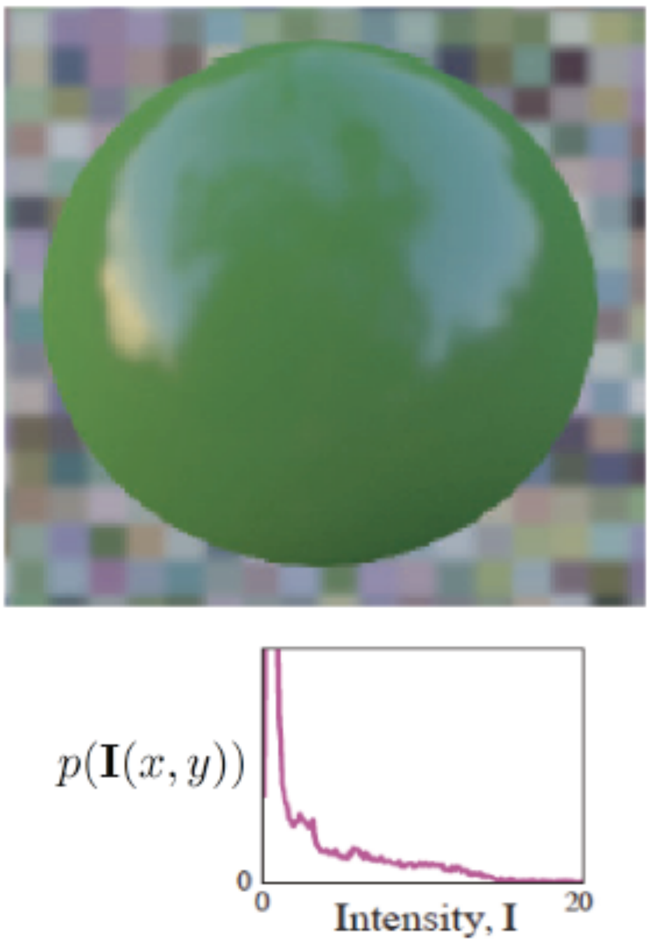
Changing the shape of the luminance histogram alters the lightness, and thus surface appearance

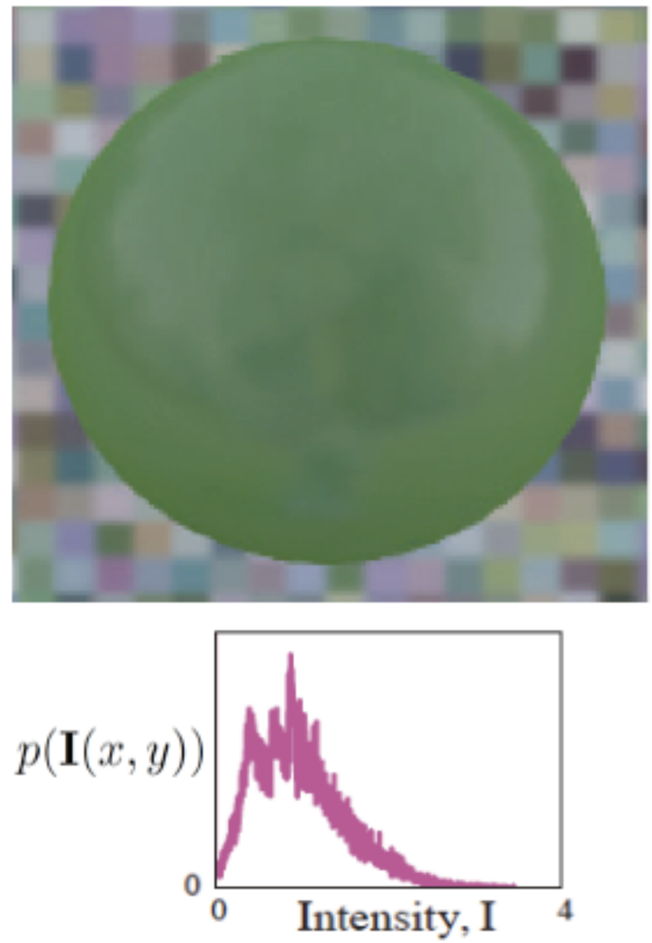
➢ Image based statistics like moments and percentiles are diagnostic of diffuse reflectance. Altering these statistics of an image changes the surface appearance.

Effect of manipulating image statistics on perception



The importance of distribution of intensities





I-2. Contour grouping from natural image statistics

- How well do the contour integration preferences of human vision actually mirror the characteristics of natural images ?

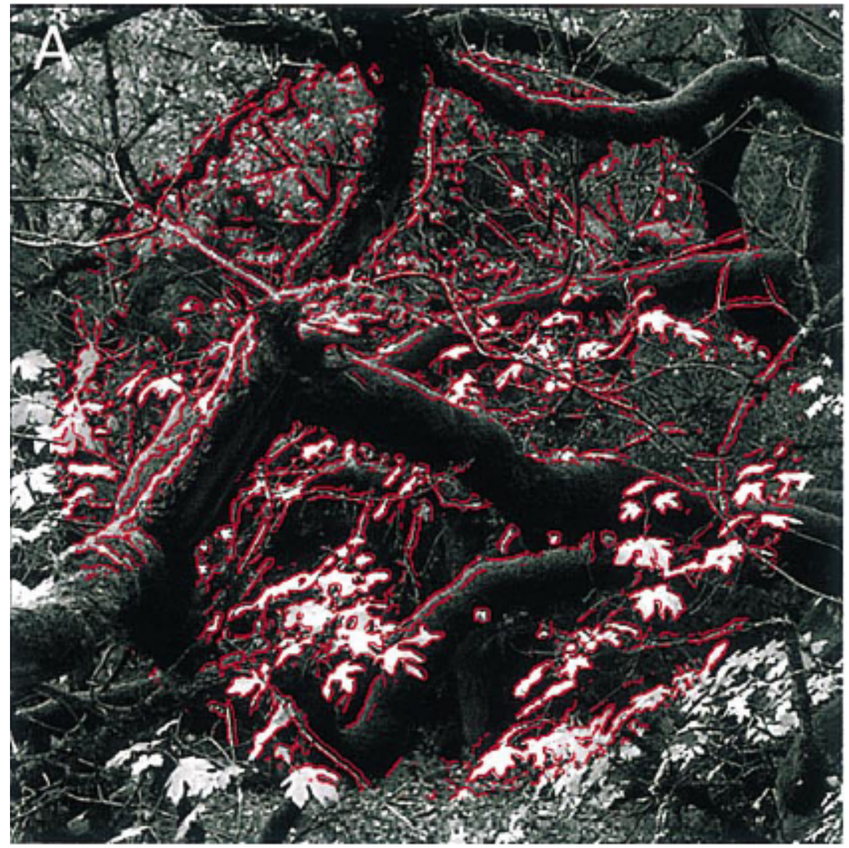
- Hypothesis: the development of contour integration mechanisms is driven by the occurrence statistics of images encountered in the natural world.

Geisler et al (2001). Edges co-occurrence in natural images predicts contour grouping performances. Vision Research, 41, 711-724.

- Geisler measured the contour formation properties of images. Each image was displayed on a computer screen and people moved a cursor to select all the oriented elements that belonged together in a single shared contour.

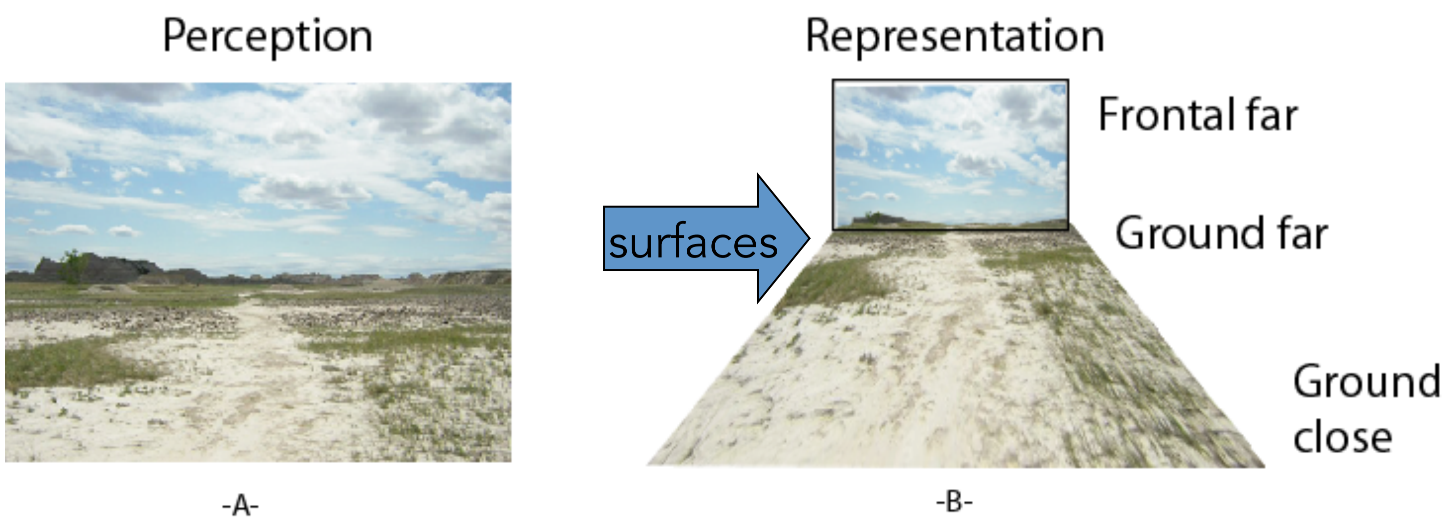
- They computed the orientation and position differences among all pairs of segments belonging to a same contour.

- Result: **Adjacent segments of any single natural contour tend to have very similar orientations**, but segments of the same contour that are further apart tend to have orientations disparate.



Geisler et al (2001). Edges co-occurrence in natural images predicts contour grouping performances. Vision Research, 41, 711-724.

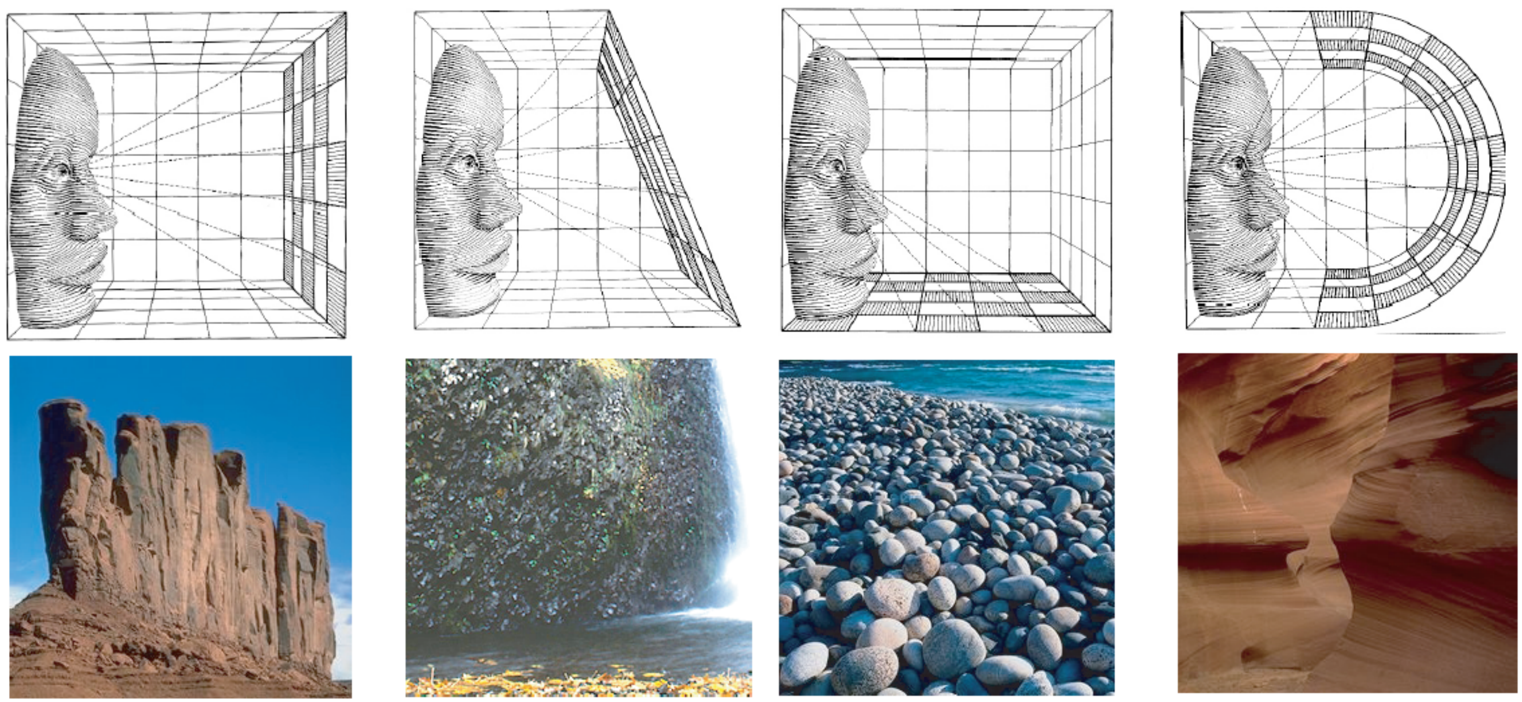
II- Mid-level Image Statistics



II-1 Texture Gradient

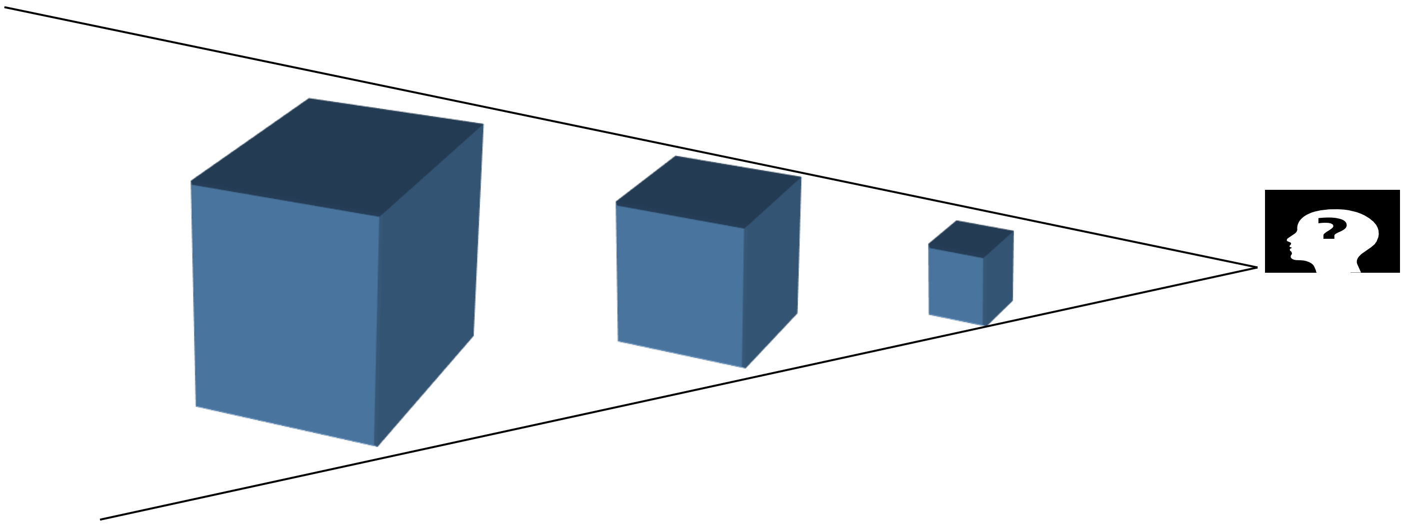
Texture gradient describes the correspondence between the pattern of a surface and the structure of the 3 D world.
There are several signature textural gradient: e.g. frontal surface project uniform gradients. Longitudinal surfaces such as floors and streets project gradient that diminish with greater distance from the observer.

Flat frontal vertical surface | Flat frontal slanting surface | Flat longitudinal ground surface | Rounded surface



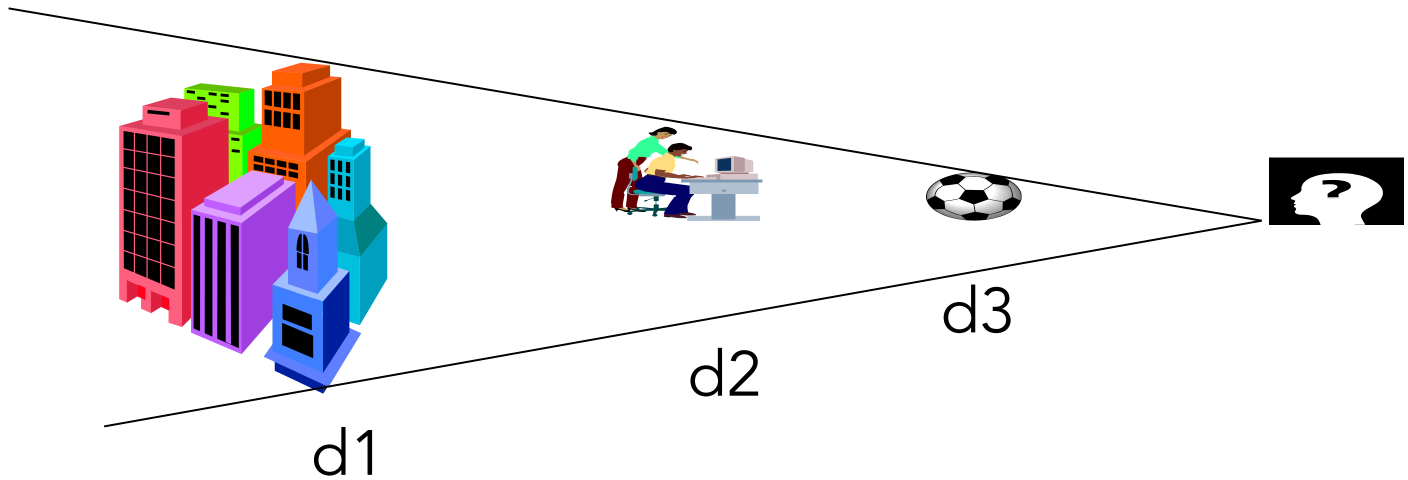
II-2 Depth Perception

Mean depth refers to a global measurement of the mean distance between the observer and the main objects and structures that compose the scene.



Stimulus ambiguity: the three cubes produce the same retinal image. Monocular information cannot give absolute depth measurements. Only relative depth information such as shape from shading and junctions (occlusions) can be obtained.

Depth Perception from Image Structure

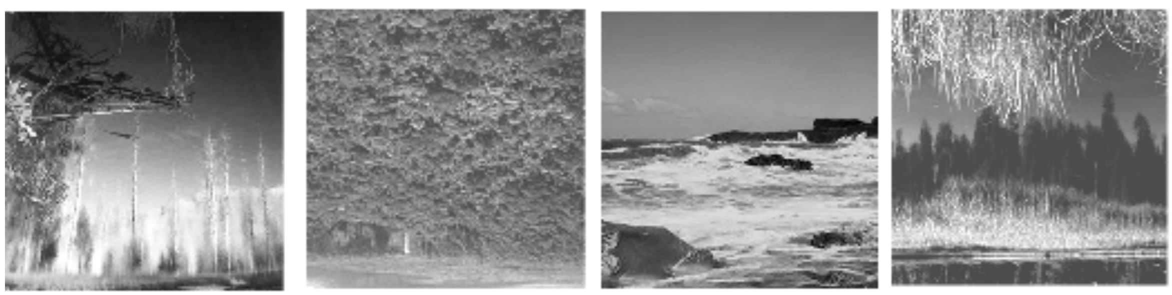


If d1>>d2>>d3 the structures of each view strongly differ. **Structure** provides monocular information about the scale (mean depth) of the space in front of the observer.

Torralba, A., & Oliva, A. (2002). Depth estimation from image structure. *IEEE Pattern Analysis and Machine Intelligence*, 24,1226-1238.

Close up view / Looking down

Large space / open space / looking at the horizon



The image inversion has two main effects:

1) Reverse lighting effects: mainly changes the interpretation of object/ ground affiliation

2) Inversion of spatial organization: it can produce in some cases large changes in the perceived *scale* of the image

Statistical Regularities of Depth

When increasing the size of the space, natural environment structures become larger and smoother.

Close-up → Very far

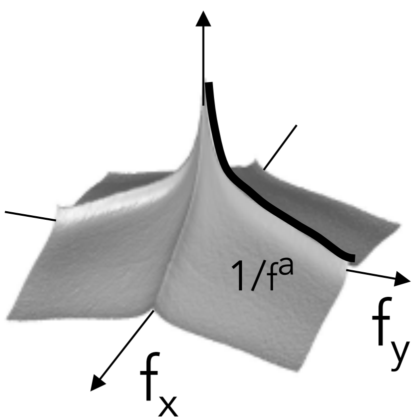
For man-made environments, the clutter of the scene increases with increasing distance: close-up views on objects have large and homogeneous regions. When increasing the size of the space, the scene "surface" breaks down in smaller pieces (objects, walls, windows, etc).

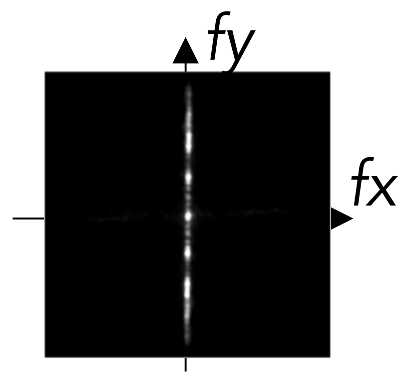
Torralba & Oliva. (2002). Depth estimation from image structure. IEEE Pattern Analysis and Machine Intelligence

Natural Image Statistics

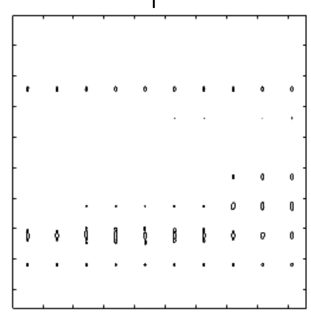
The group of natural images have particular second-order statistics (quantity of orientation, quantity of frequencies).

Fourier Power Spectrum





Power Spectrum

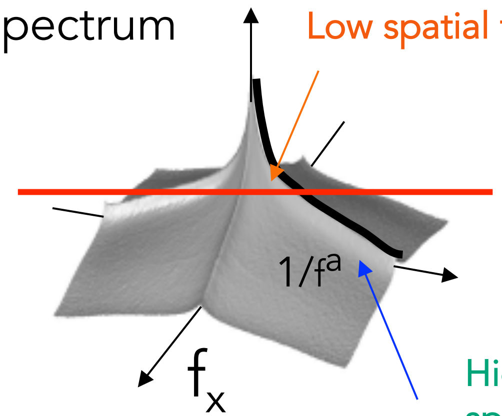


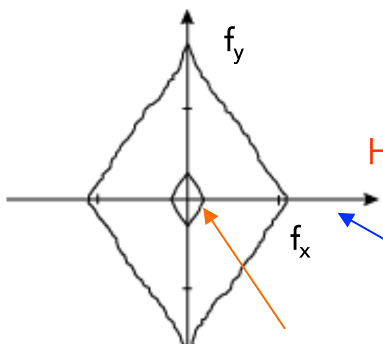
Spectrogram

Fourier Characteristics of Natural Images

GlobalLocalFourierSpectra/AveragePowerSpectrum.m

Fourier
Power spectrum





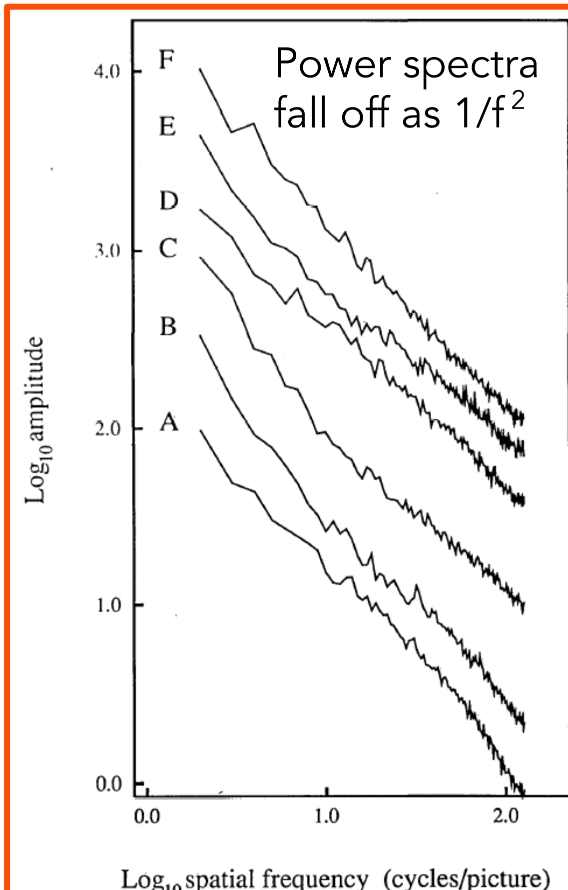
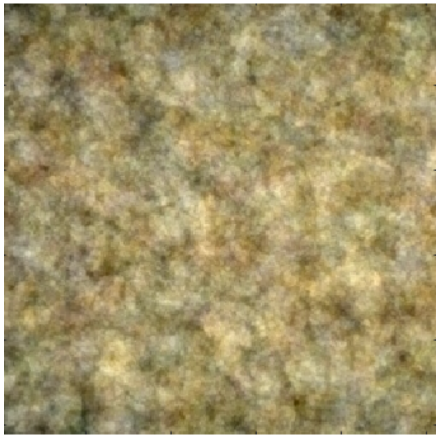


Fig. 8. Amplitude spectra for the six images A–F, averaged across all orientations. The spectra have been shifted up for clarity. On these log–log coordinates the spectra fall off by a factor of roughly $1/f$ (a slope of -1). Therefore the power spectra fall off as $1/f^2$.

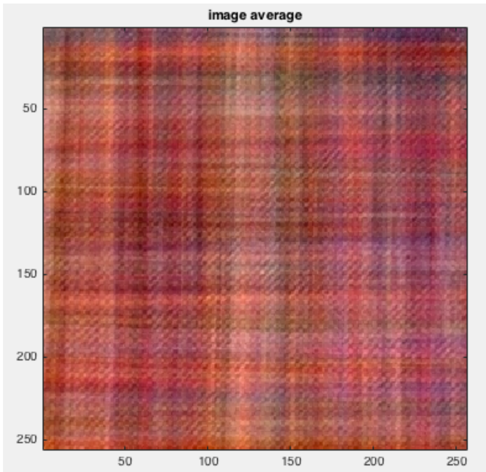
D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," J. Opt. Soc. Am. A **4**, 2379- (1987)

In a world of pebbles…



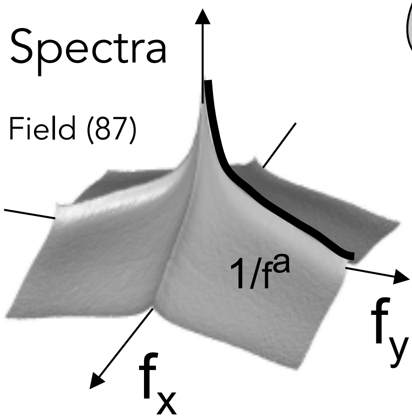
GlobalLocalFourierSpectra/AverageAndPowerSpectrum.m

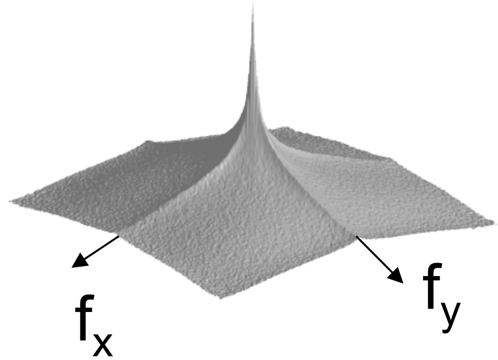
In a world of plaids…



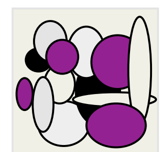
GlobalLocalFourierSpectra/AverageAndPowerSpectrum.m

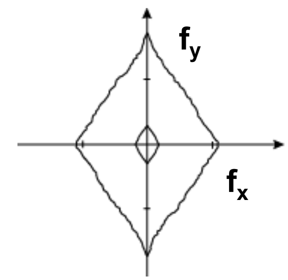
Statistics of Scene Categories



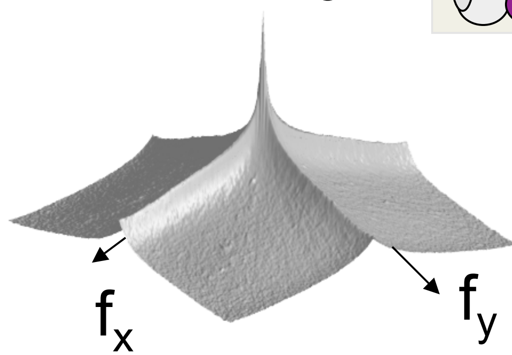


Natural scenes
(6000 images)

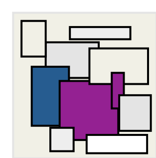


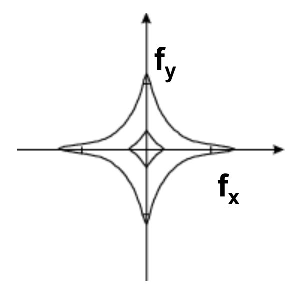


Natural scenes
spectral signature



Man-made scenes
(6000 images)





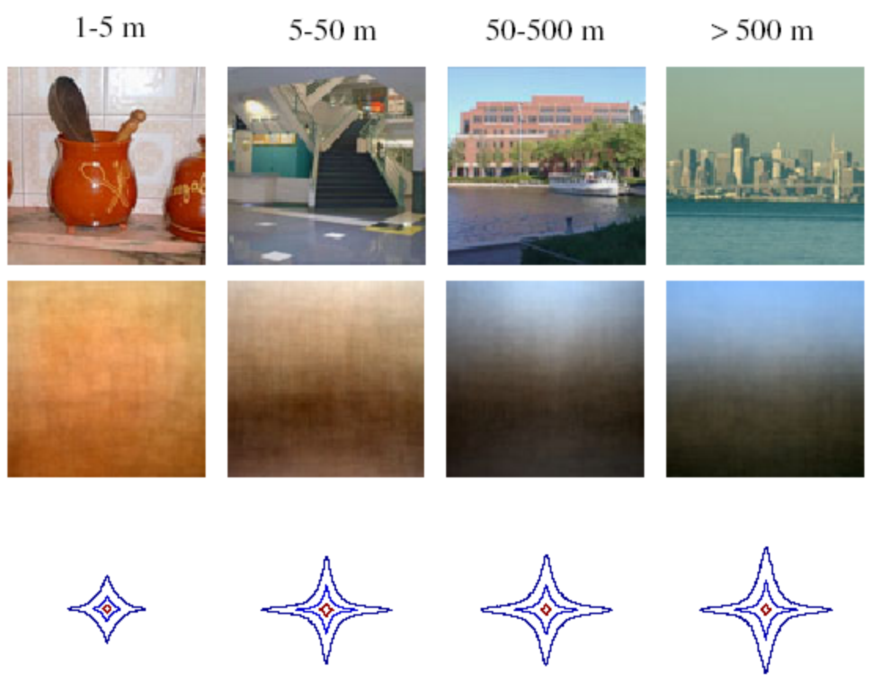
Man-made scenes
spectral signature

		H	O	V
Natural	α	1.98 (0.58)	2.02 (0.53)	2.22 (0.55)
	A	0.96 (0.40)	0.86 (0.38)	1 (0.35)
Man-made	α	1.83 (0.58)	2.37 (0.45)	2.07 (0.52)
	A	1 (0.32)	0.49 (0.24)	0.88 (0.29)

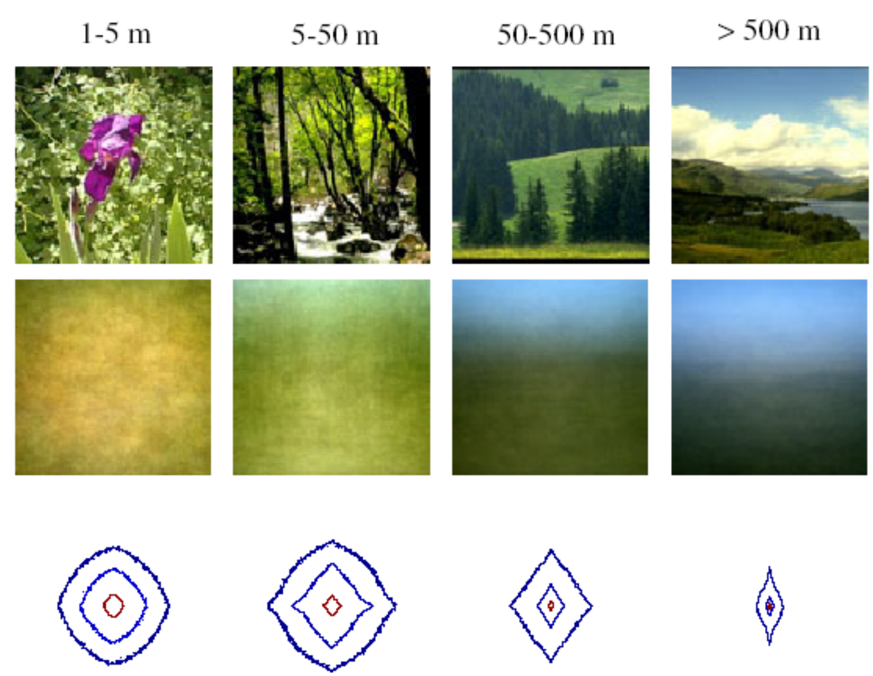
Torralba and Oliva, *Statistics of Natural Image Categories*. Network: Computation in Neural Systems 14 (2003) 391-412.

Statistics of Environments

Spectral signature of
man-made environments



Spectral signature of
natural environments

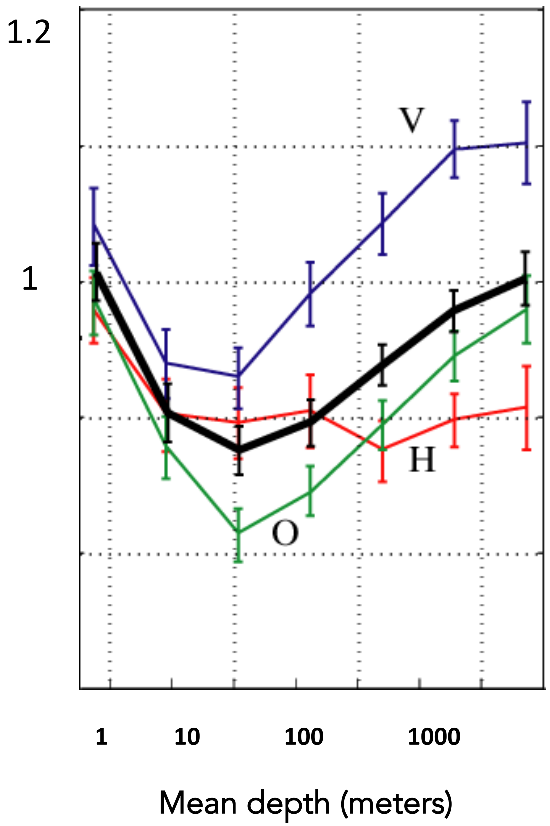


Spectral Regularities of Mean Depth

When increasing the size of the space, natural environment structures become larger and smoother.

For man-made environments, the clutter of the scene increases with increasing distance: close-up views on objects have large and homogeneous regions. When increasing the size of the space, the scene "surface" breaks down in smaller pieces (objects, walls, windows, etc).

Slope of the **magnitude spectrum** (Vertical, Horizontal, Oblique) with respect to the mean depth of the scene



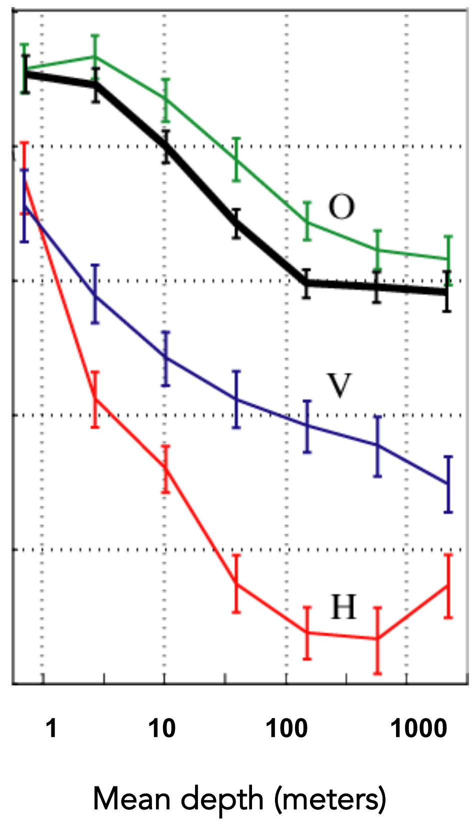
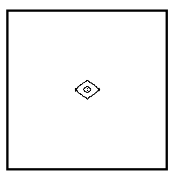


Image Statistics and Scene Scale

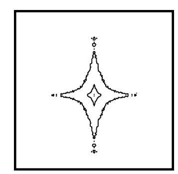
Close-up views

Large scenes

On average, low clutter



On average, highly cluttered



Viewpoint is unconstrained

Point view is
Strongly constrained

Image Scale vs. World Scene Scale

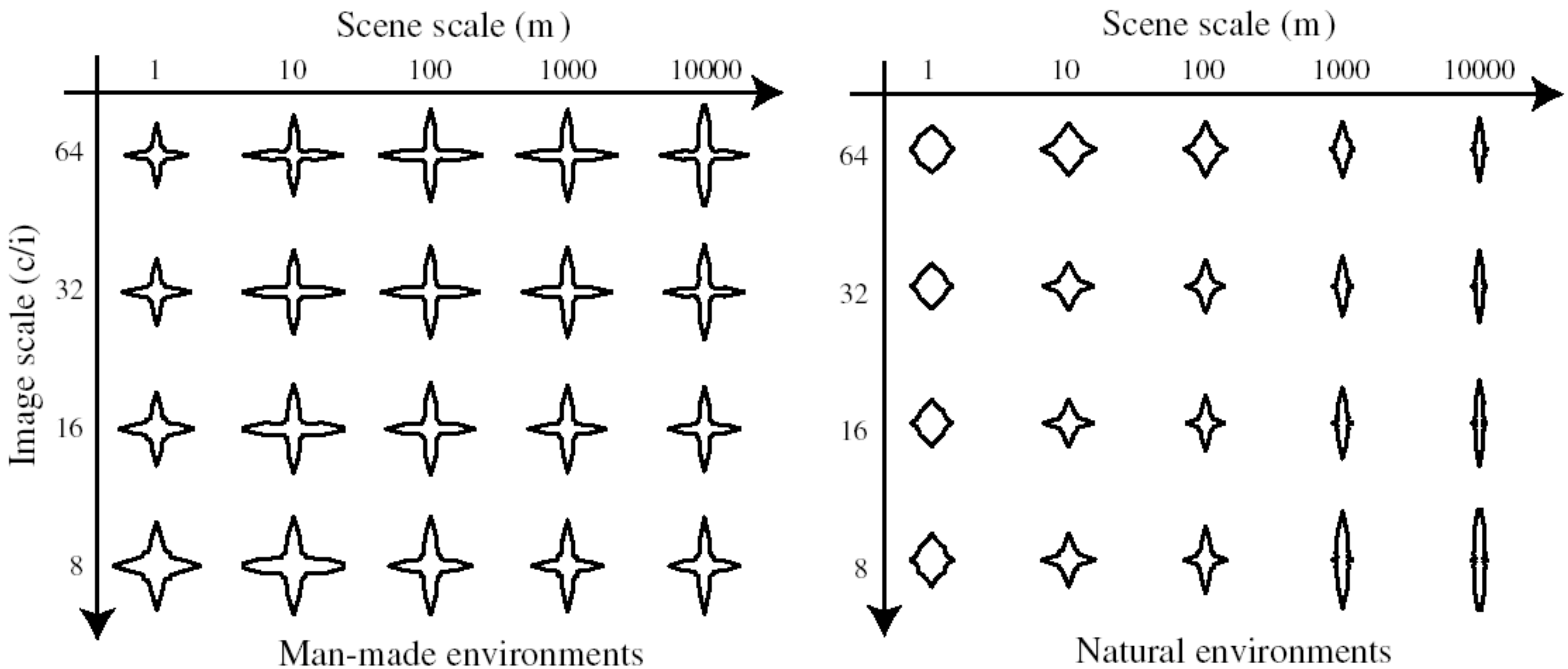


Figure 5. Polar plots of responses of multiscale oriented Gabor filters. The magnitude of each orientation corresponds to the total output energy averaged across the entire image. The energies are normalized across image scale by multiplying by a constant so that noise with $1/f$ amplitude spectrum has the same polar plots at all image scales.

Torralba and Oliva, *Statistics of Natural Image Categories*. Network: Computation in Neural Systems 14 (2003) 391-412.

Spatially Localized Statistics

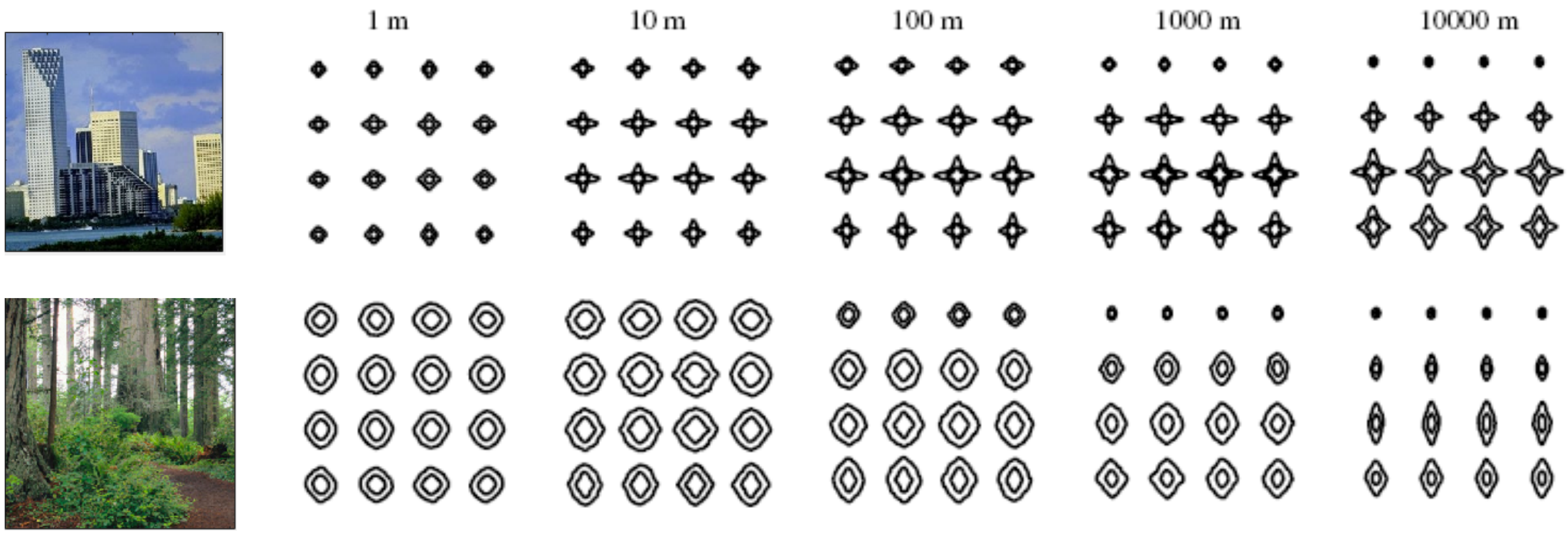
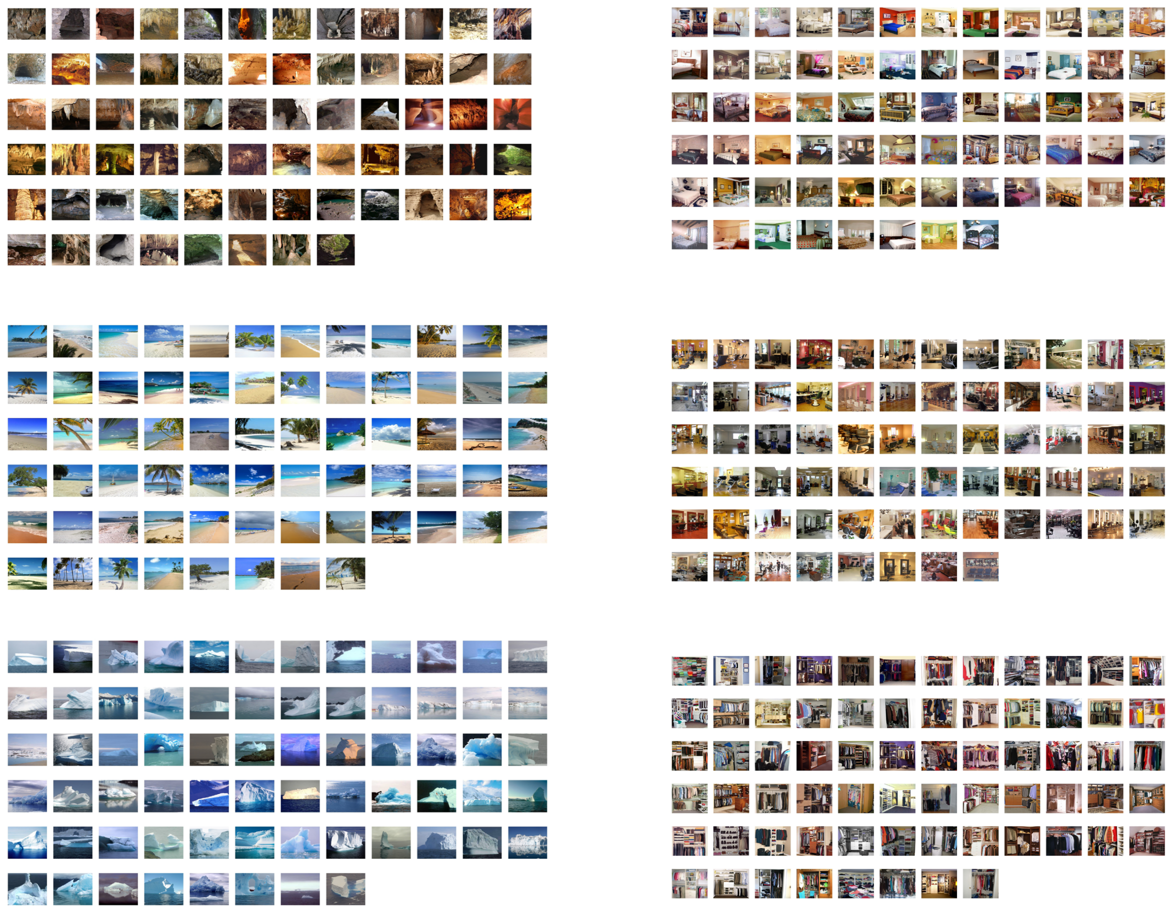


Image statistics become non-stationary as scene scale increases.

GlobalLocalFourierSpectra/LocalPowerSpectrum.m

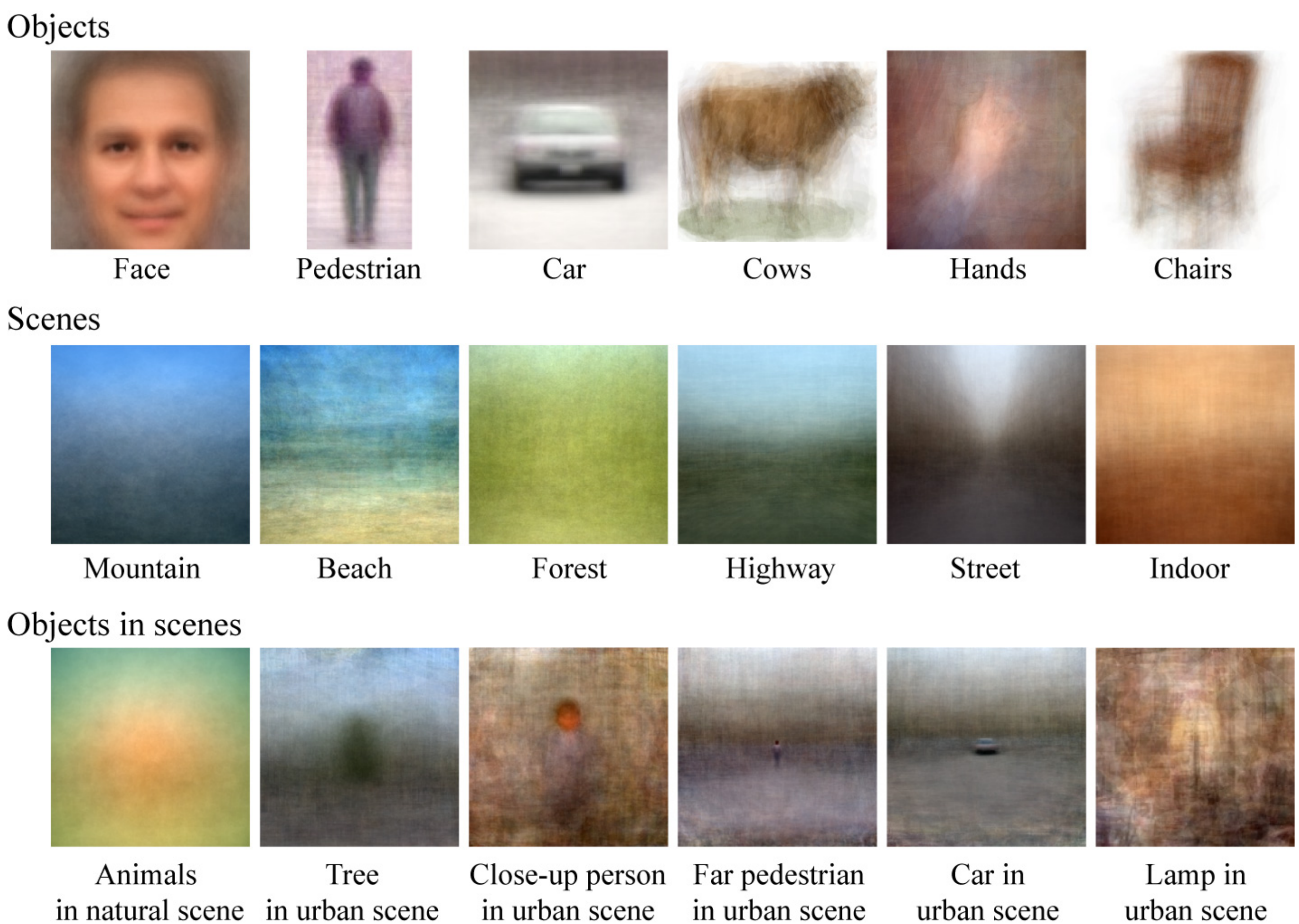
III - High level image statistics

There are lots of regularities.. Which ones are important ?



http://cvcl.mit.edu/MM/sceneCategories.html

Statistics of Categories of Natural Images



Averaged pictures of categories of objects, scenes and objects in scenes, computed with 100 exemplars or more per category. Exemplars were chosen to have the same basic level and viewpoint in regard to an observer. The group objects in scenes (third row) represent examples of the averaged peripheral information around an object centered in the image.

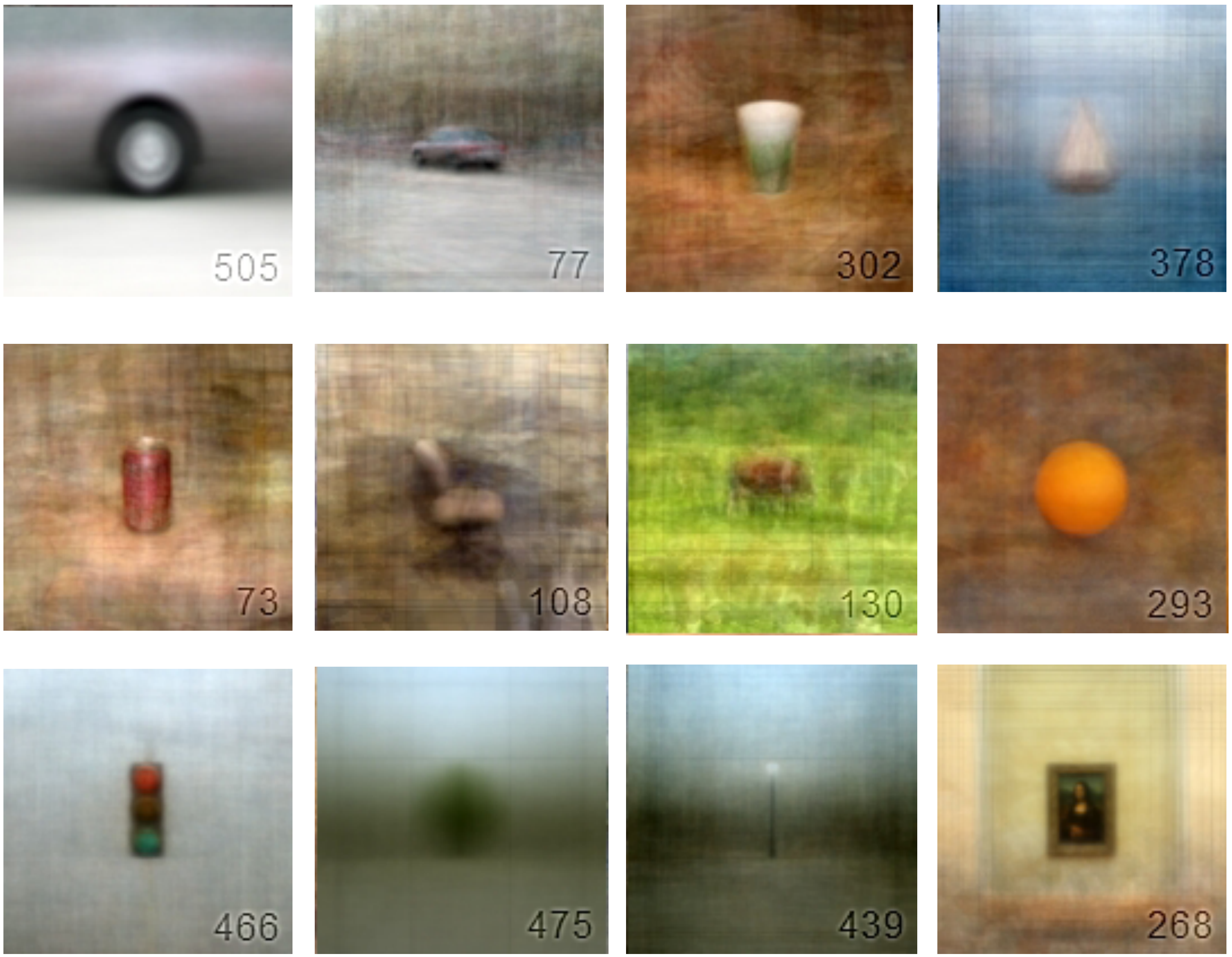
Statistical Regularities object-background

keyboard

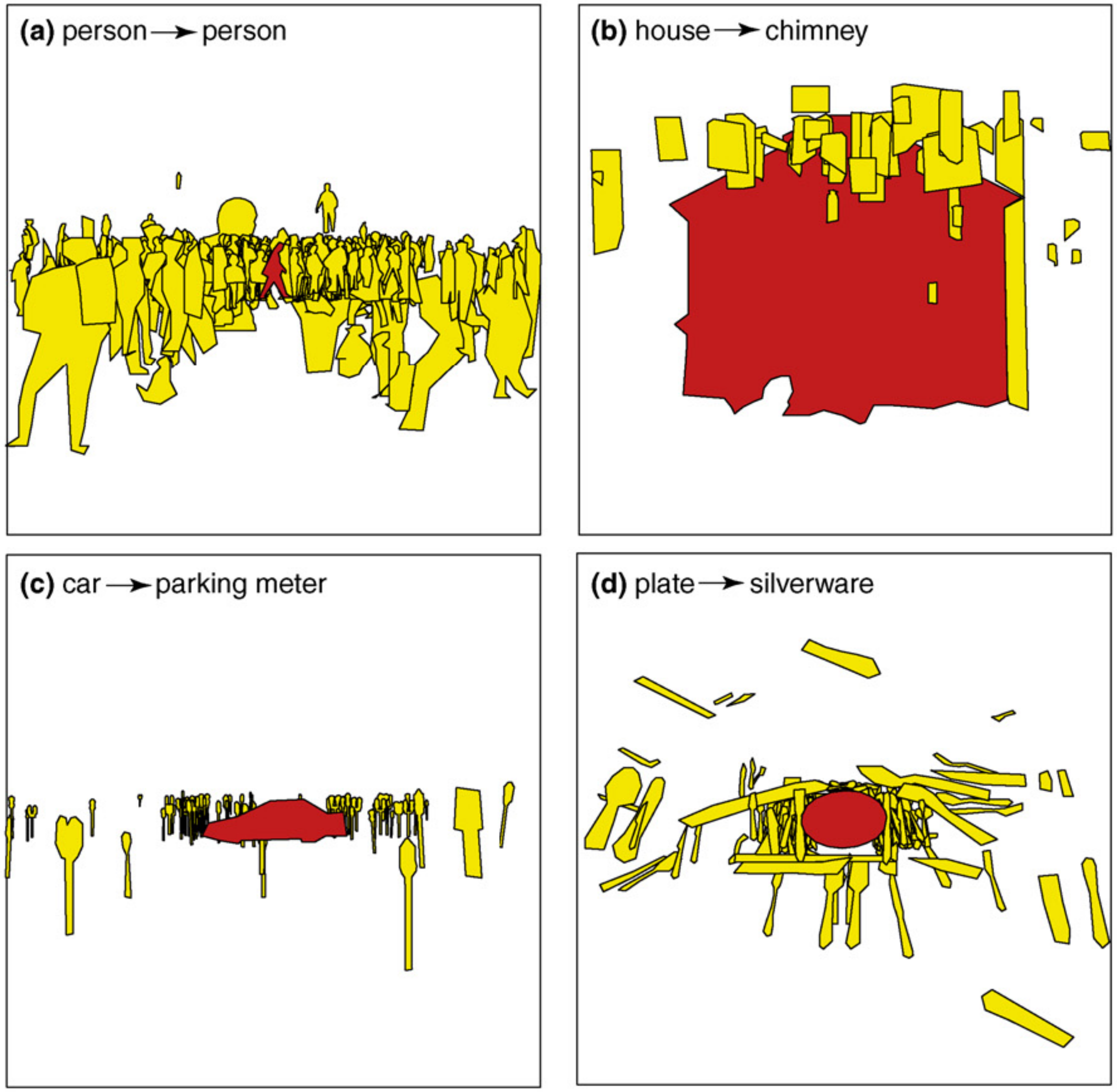
Fire hydrant

Oliva & Torralba (2007) TICS

Statistical Regularities object-background



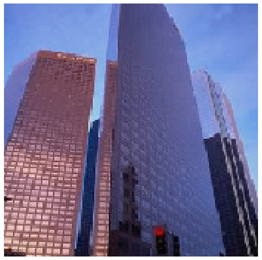
Statistical regularities object-object



Oliva & Torralba (2007) TICS

What is driving the scene regularities?

Physical processes that shape the environment?



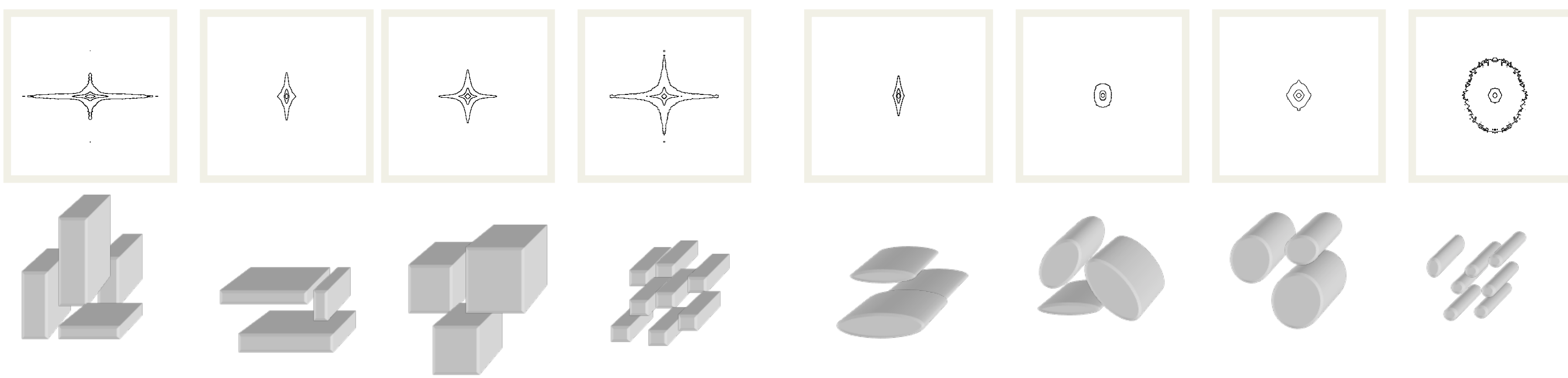
Restrictions on possible observer points of view?

Functional constraints of the scene?

Spectral Signature of semantic categories

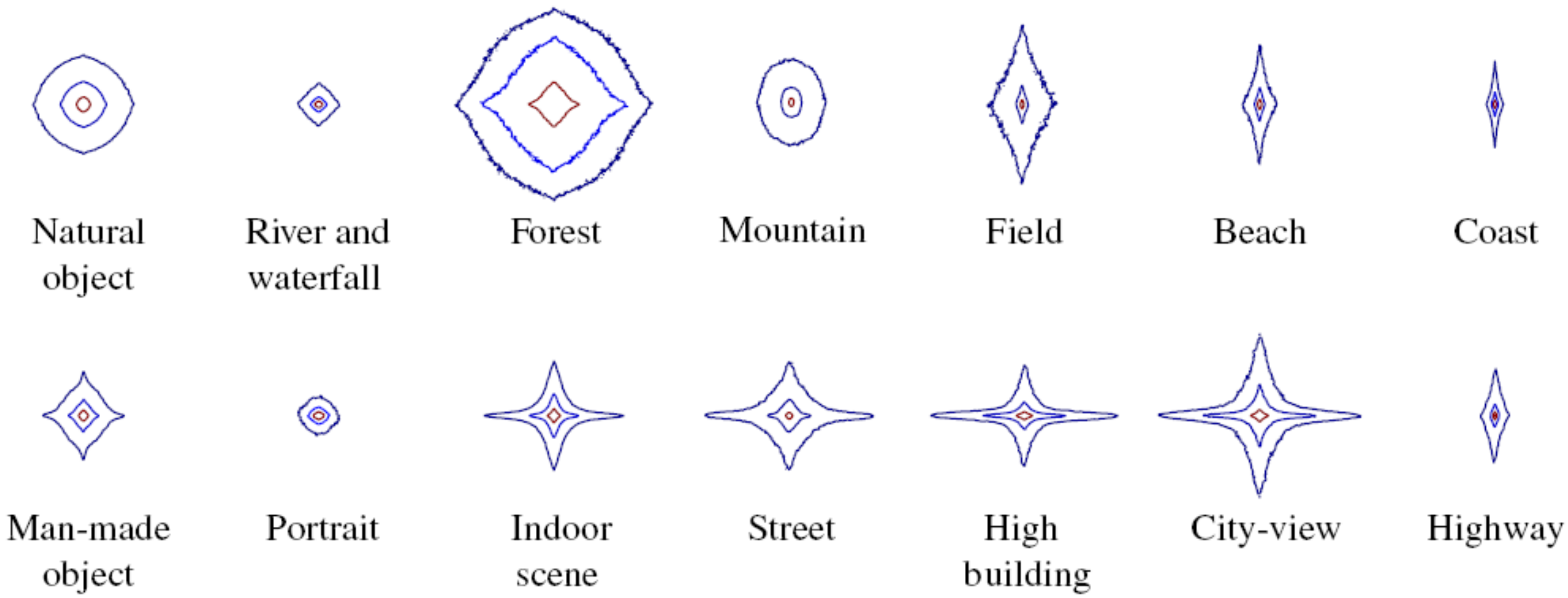
Man-made environments   Natural environments

Spectral signature of categories of man-made environments   Spectral signature of categories of natural environments



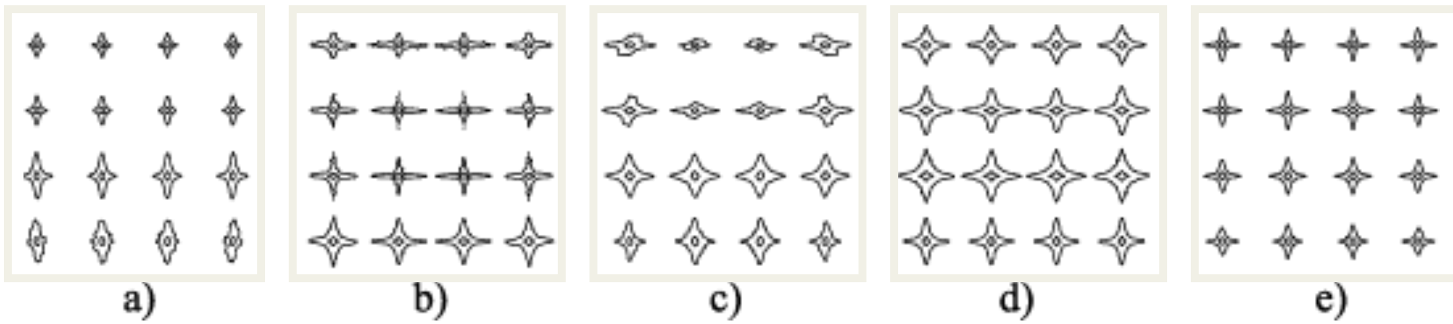
Torralba and Oliva, *Statistics of Natural Image Categories.* Network: Computation in Neural Systems 14 (2003) 391-412.

Basic-level scene spectral signatures

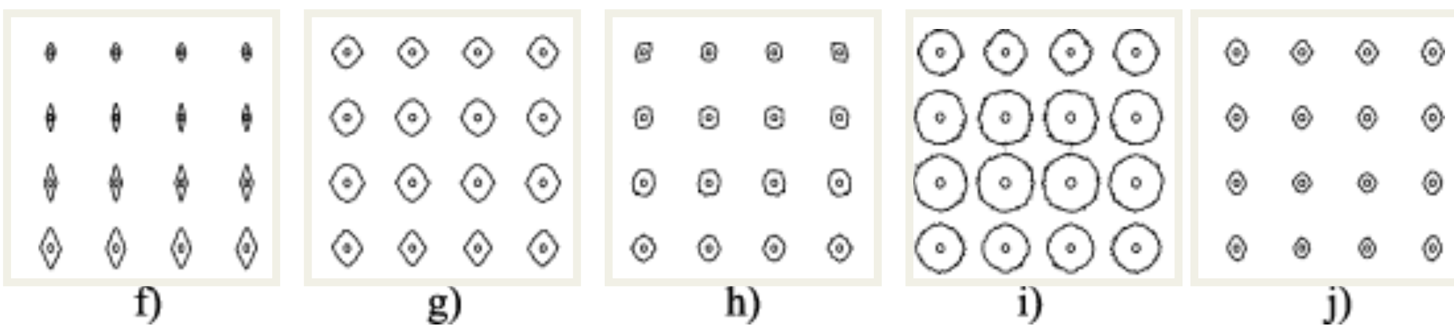


Spectral Layout Signature

Spectral layout signatures of several scene categories (averaged from hundreds of exemplars)



Open man-made scenes | Vertically structured scenes | Perspective views of streets | Far views of city center | Close up views of urban scenes



Open natural scenes | Closed natural scenes | Mountaneous landscapes | Enclosed forests | Close up views of rocks, water

Image statistics are non-stationary when considering specific scene categories.

GlobalLocalFourierSpectra/LocalPowerSpectrum.m