Computer Vision (7CCSMCVI / 6CCS3COV)

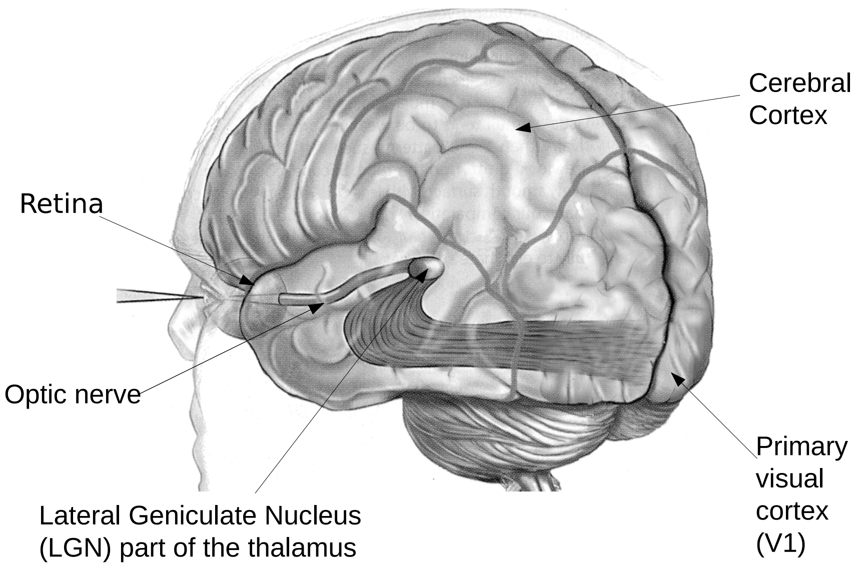
Recap

- **Image formation**
- **Low-level vision**
  - Artificial
    - Convolution:
      smoothing, differencing, edge detection
    - Multi-scale image representations:
      Gaussian and Laplacian image pyramids
  - Biological ←Today
- **Mid-level vision**
- **High-level vision**

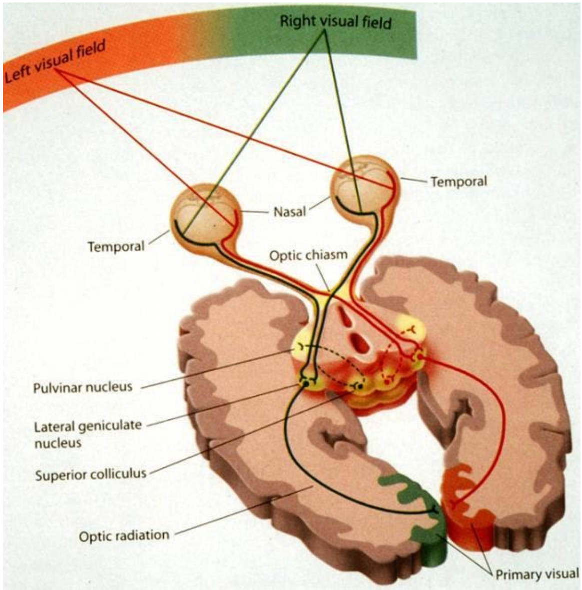
Today

- Biological Visual System basics
- Primary visual cortex (V1)
  - physiology
    – input selectivities / classical receptive fields
  - anatomy
    – layers, hypercoloumns, and maps
  - models of orientation selectivity
    – gabor filters
  - extra-classical receptive field properties
    – contextual influences / change detection

Human visual system (beyond the retina)



Pathways from Retina to Cortex



- The right visual field (RVF) projects to left sides of each retina
- ganglion cells from the left side of the left eye projects to the left LGN
- ganglion cells from the left side of the right eye cross over at the optic chiasm and go to the left LGN
- Hence, the RVF projects to the left LGN
- The left LGN projects to the left V1 (striate cortex)
- NOTE: It is not the case that the right eye goes to the left V1, it is the RVF (as seen by both eyes).

What does LGN do?

Lateral geniculate nucleus transmits information from retina to cortex.

Traditionally viewed as merely a relay station.

Current evidence suggests it does more than just relay information.

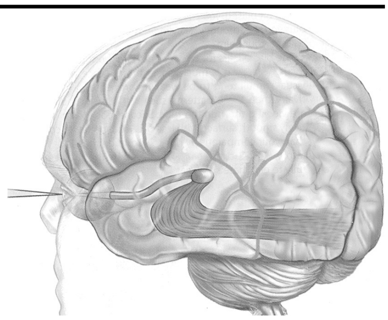
However, what computation occurs in the LGN is not currently known.

LGN cells have centre-surround RFs, like retinal ganglion cells.

What does the Cerebral Cortex do?

All higher cognitive functions

- Perception
- Knowledge
- Learning
- Language
- Memory
- Reasoning
- Decision Making

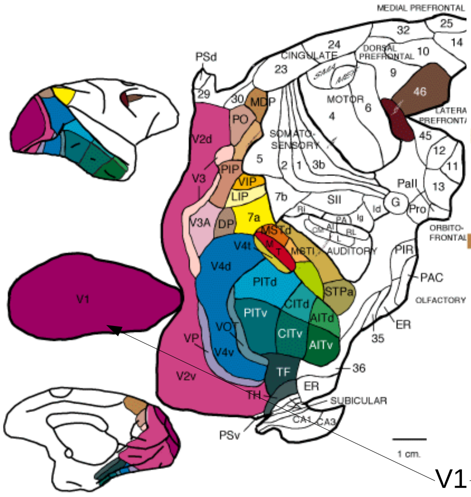


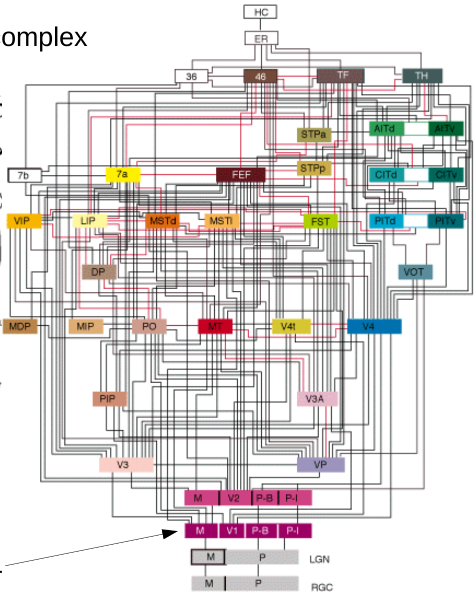
Basic facts:

- A folded sheet approx. 1.7mm thick, with an area of approx. $0.25m^2$
- Contains 10^{10} neurons (= approx. 10^5 neurons/mm^3) and $4x10^{13}$ synapses
- Is about 2% of human body mass but accounts for 20% of human energy consumption
- Approx. 50% of cerebral cortex is devoted to vision
- Is divided into areas which specialise in different functions. Approx. 30 areas are devoted to different aspects of visual information processing.

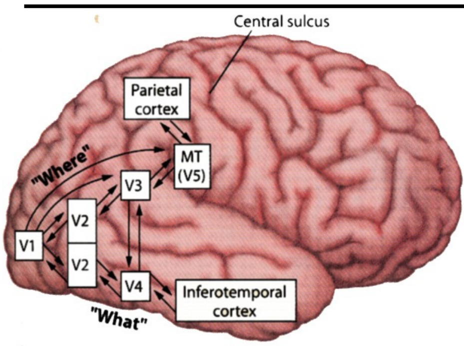
The Cortical Visual System: areas

Lots of distinct cortical areas. Very complex interconnectivity.





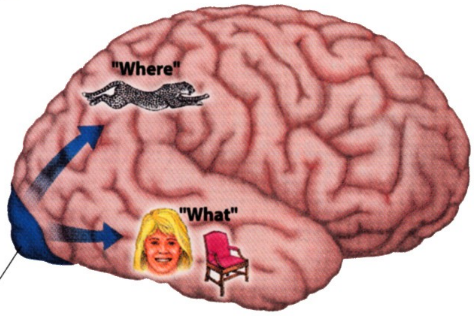
The Cortical Visual System: pathways



"What" and "Where" pathways

Hierarchically organised:

- simple, local, RFs at V1
- complex, large, RFs in higher areas



Where (or How):

- V1 to parietal cortex
- spatial / motion information

What

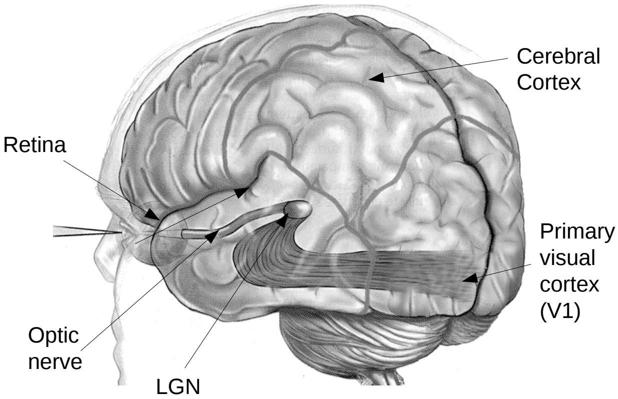
- V1 to inferotemporal cortex
- identity / category information

Cortical area V1

LGN neurons mainly project to the primary visual cortex

Primary visual cortex is also known as:

- V1
- striate cortex
- area 17



V1 therefore performs the initial (low-level) processing on the incoming information.

Processing is carried out by cells with a larger range of receptive field properties than are found in the retina and LGN.

V1 RFs: range of selectivities

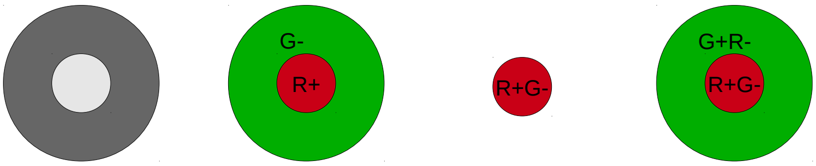
V1 cells have receptive fields selective for:

- colour
- orientation
- direction of motion
- spatial frequency
- eye of origin
- binocular disparity
- position

Some of these properties are similar to those of cells in the LGN and retinal ganglion cells (e.g., colour, eye of origin, position).

Other properties are seen for the first time in V1 (e.g., orientation selectivity and binocular disparity, direction of motion).

V1 RFs: centre-surround / colour



Broad band centre-surround | Colour opponent centre-surround | Colour opponent centre only | Double-opponent

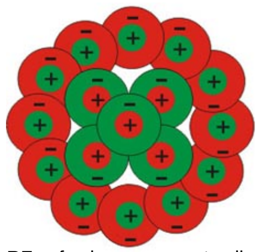
Double-opponent have larger receptive fields than simple opponent cells

Centre-surround / colour opponent RFs are similar to those in the LGN/retina and perform the same function: detecting <u>changes in intensity</u> or <u>regions of uniform colour</u>.

Double-opponent (DO) cells are not seen prior to V1.

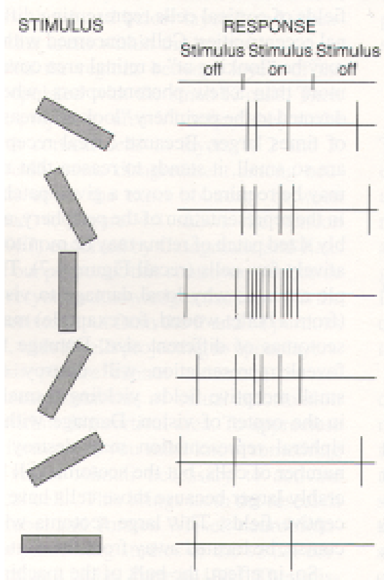
The input to a DO cell likely comes from a number of single opponent cells.

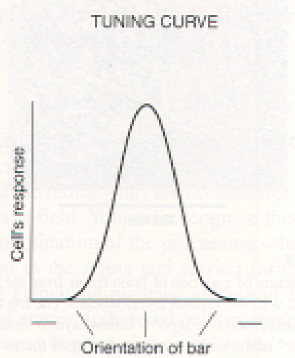
They detect locations where colour changes (e.g. where centre is red and background green).



RFs of colour opponent cells providing input to DO cell

V1 RFs: orientation selectivity

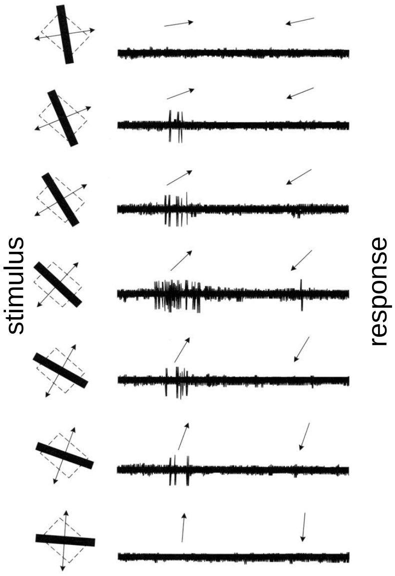




In contrast to retinal and LGN cells, many V1 cells display a marked preferences to a particular orientation.

Orientation selectivity is sharply tuned with responses falling to zero when a line is tilted ±15-30° from the “preferred orientation”.

V1 RFs: orientation (simple cells)



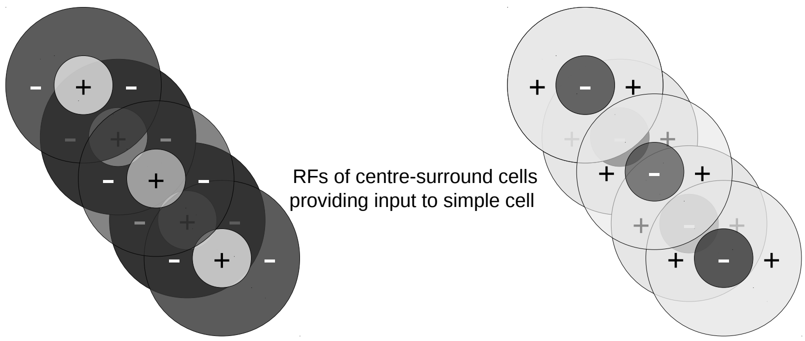
Optimum response to an appropriately oriented stimulus placed at a certain position within the receptive field.

This particular example is also direction selective as it produces a strong response only when the stimulus is at a particular angle and moving in a particular direction through a particular location.

V1 RFs: orientation (simple cells)

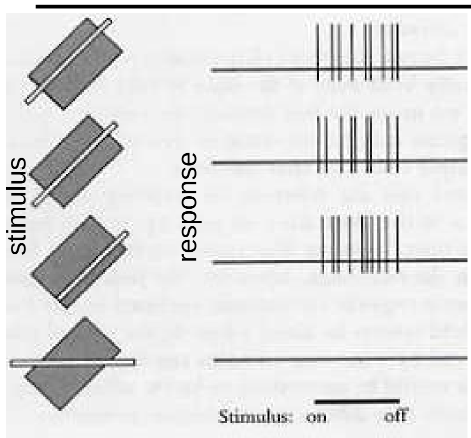
Simple cells act as edge and bar detectors.

Simple cell RFs are formed by pooling the output of several centre-surround cells, whose receptive fields lie along a common line.



Simple cells can be selective for light bars on a dark background, or dark bars on a light background, or coloured bars depending on the properties of the centre-surround cells they receive input from.

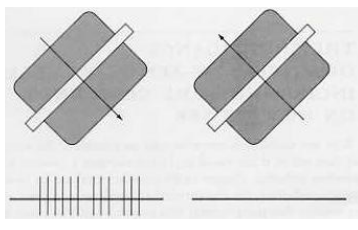
V1 RFs: orientation (complex cells)



Optimum response to an appropriately oriented stimulus placed anywhere within the receptive field.

Complex cells thus have a greater positional invariance compared to simple cells.

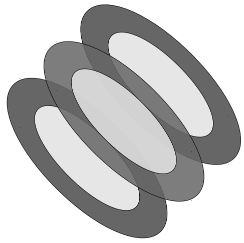
Usually they respond most strongly to moving bars/edges and can be direction selective:



V1 RFs: orientation (complex cells)

Complex cells act as edge and bar detectors with some tolerance to location.

Complex cell RFs are formed by pooling the output of several simple cells, whose receptive fields are parallel to each other.

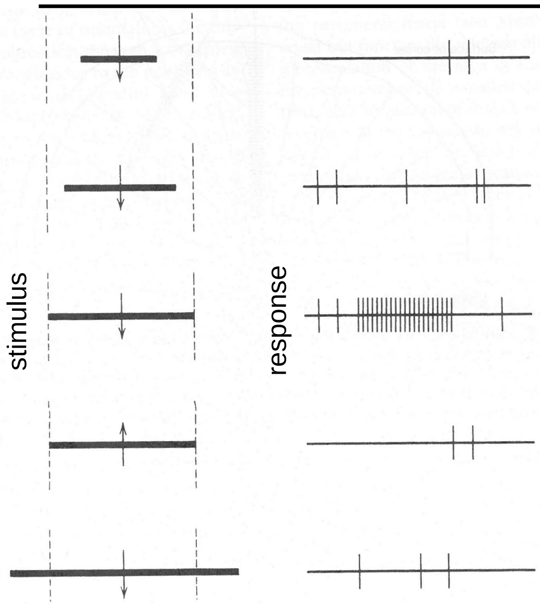


RFs of simple cells providing input to complex cell

Another way to say this is that the RFs of the simple cells lie in the same location but prefer stimuli at different spatial phases.

No matter what the spatial phase of the stimulus at least one simple cell will respond and drive the response of the complex cell.

V1 RFs: orientation (hyper-complex cells)



Optimum response depends not only upon stimulus orientation but also on length.

Maximum response occurs when the length of the bar matches the width of the receptive field (i.e. these cells are selective for length as well as orientation)

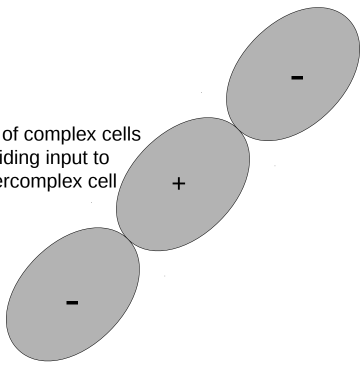
This is called “End-stopping”.

V1 RFs: orientation (hyper-complex cells)

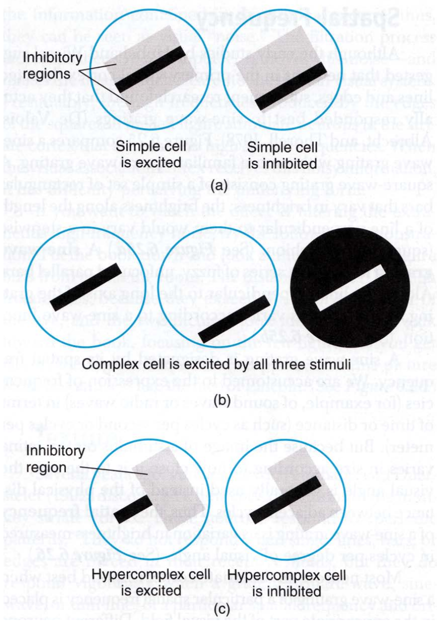
Hypercomplex cell RFs can be formed by pooling the output of two or more complex cells, whose receptive fields are collinear.

Complex cells at the ends need to inhibit the hypercomplex cell to produce end-stopping.

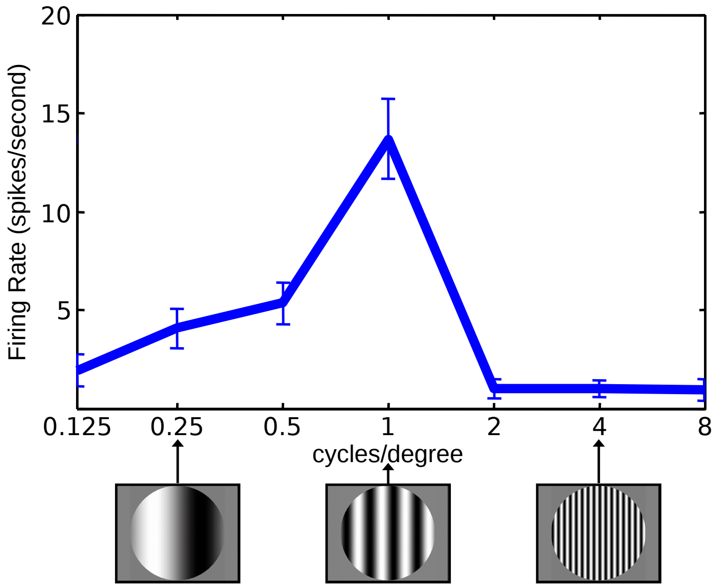
RFs of complex cells providing input to hypercomplex cell



V1 RFs: orientation selectivity (summary)



V1 RFs: spatial frequency tuning



V1 RFs: eye of origin

Monocular cells

Retinal and LGN cells receive input from one eye only.

Cells in V1 that receive the input from LGN are also monocular.

Binocular cells

Other cells in V1 are binocular: they receive input (via monocular cells) from the two eyes.

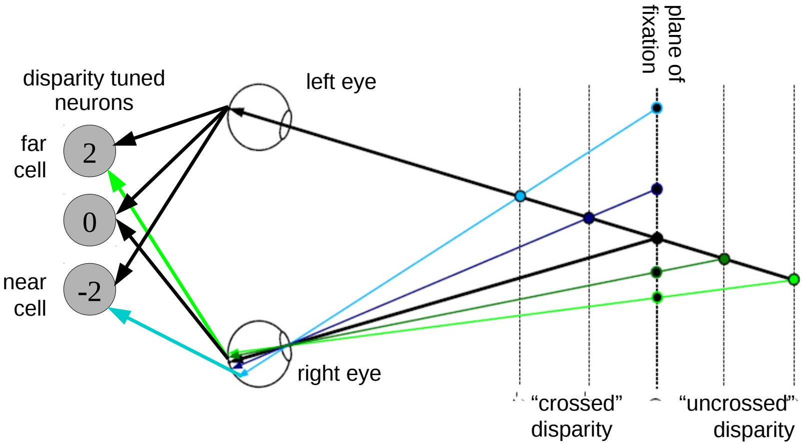
Binocular cells have two receptive fields (Left eye & Right eye) and these are matched in type (e.g. same direction of motion preference, same orientation preference, both simple or both complex).

Hence, binocular cells respond maximally when corresponding regions in each eye are stimulated by stimuli of similar appearance.

V1 RFs: disparity

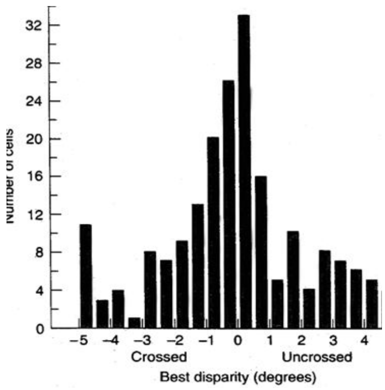
Many binocular cells are tuned to disparity.

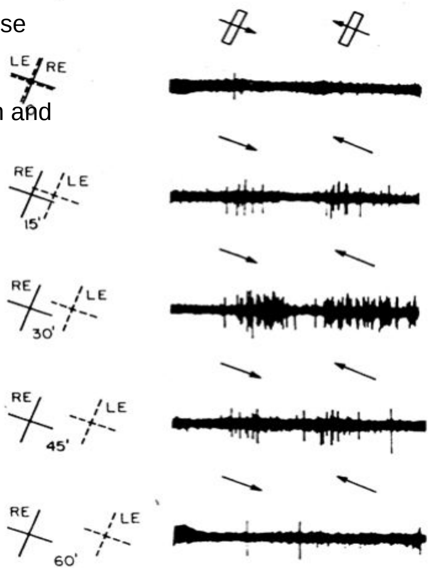
They encode the depth of the stimulus at a particular location.



V1 RFs: disparity

e.g. for this neuron, maximum response occurs when the line is presented simultaneously to both eyes, with the correct orientation, direction of motion and the correct binocular disparity.





V1 RFs: conjunctions

V1 cells have receptive fields selective for:

- colour
- orientation
- direction of motion
- spatial frequency
- eye of origin
- binocular disparity
- position

Some cells are responsive to more than one of these properties

e.g. orientation and direction of motion,

orientation and colour

All are selective for at least one of these properties **plus** position (as the stimulus needs to be at a particular location on the retina).

V1 organisation: Hypercolumns

All neurons with RFs at the same location on the retina are located in the same region of V1.

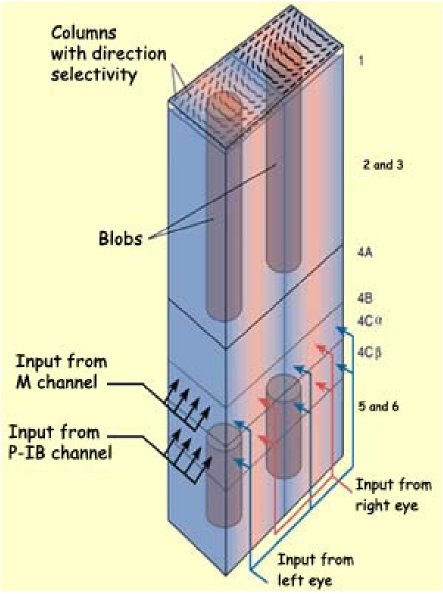
Approx. 1mm^2 of V1 is required to represent the whole range of RF types for one particular position.

Such a 1mm^2 area of V1 is called a “hypercolumn”

Hence, each hypercolumn contains the requisite neural machinery to simultaneously analyse multiple attributes of an image (colour, orientation, direction of motion, spatial frequency, eye of origin, binocular disparity) falling on a localised region of the retina.

Internally each hypercolumn contains substructures both across the cortical surface, and perpendicular to the surface (layers).

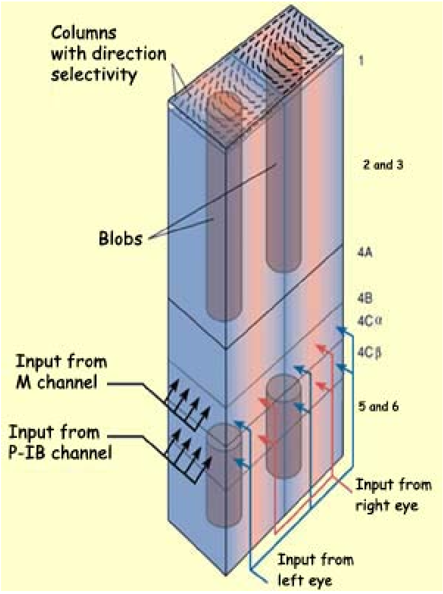
V1 organisation: cortical layers



As with the rest of the cerebral cortex, V1 consists of six layers and several sub layers which are arranged in bands parallel to the surface of the cortex.

Neurons with different properties are located in different layers.

V1 organisation: cortical layers

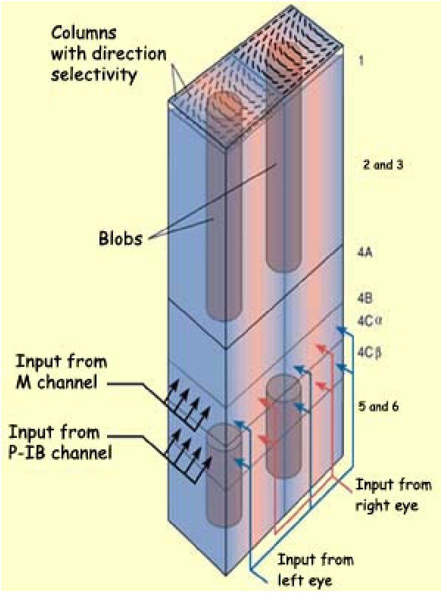


Input from the LGN arrives in L4.

- L4 cells have RFs similar to LGN and retinal ganglion cells (i.e. centre-surround, colour opponent), but there are also some simple cells.
- Each L4 cell receives input from one eye only, and hence code for eye-of-origin (i.e. they are monocular).

Layer 4 sends connections to upper layers (L2/3) and lower layers (L5/6).

V1 organisation: cortical layers

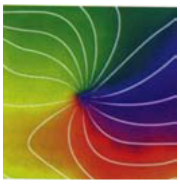


Superficial (L2/3) and deep (L5/6) layers are split into regions called "(CO) blobs" and "inter-blobs"

- Blobs contain cells sensitive to colour, including double opponent cells. Each hypercolumn contains two blobs; one right eye dominant, the other left.
- Inter-blobs contain cells sensitive to spatial frequency, orientation and direction of motion (simple, complex and hypercomplex cells) and cells that receive input from both eyes (binocular cells, sensitive to disparity).

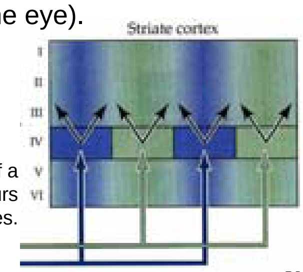
V1 organisation: Hypercolumns

The orientation preference of neurons within a hypercolumn changes in a continuous fashion forming a "pin-wheel" around each blob.



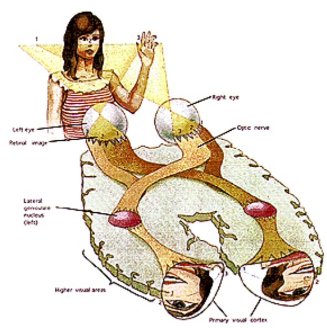
A view looking down on the surface of a hypercolumn. Colours represent different orientation preferences.

The ocular preference of neurons within a hypercolumn changes from monocular in L4 to binocular in other layers (although cells in other layers may still have some preference for one eye).



A view looking at a cross-section of a hypercolumn from the side. Colours represent different eye preferences.

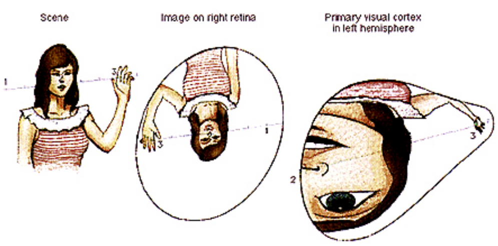
V1 organisation: retinotopic maps



Adjacent hypercolumns analyse information from adjacent areas of retina.

Hence, the spatial position of the ganglion cells within the retina is preserved by the spatial organisation of the neurons within V1.

This spatial layout is called **retinotopic** organization because the topological organization of the receptive fields in V1 parallels the organization of the retina.



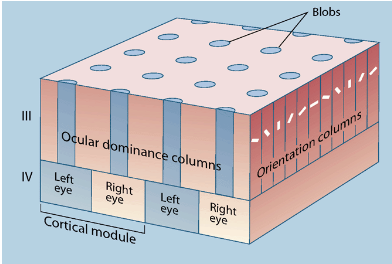
The map is distorted primarily due to cortical magnification of central vs. peripheral areas.

There is more cortical area (more hypercolomns) devoted to the fovea than peripheral vision.

This mirrors the fact that the vast majority of retinal ganglion cells are devoted to the fovea.

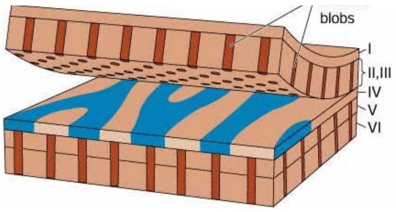
V1 organisation: retinotopic maps

The RF preferences of neurons vary smoothly within each hypercolumn and across adjacent hypercolumns.

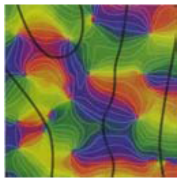


An idealised schematic showing the orientation changes, ocular preferences and blobs in 4 adjacent hypercolumns.

In reality neurons are not arranged so orderly and hypercolumns blur into each other.

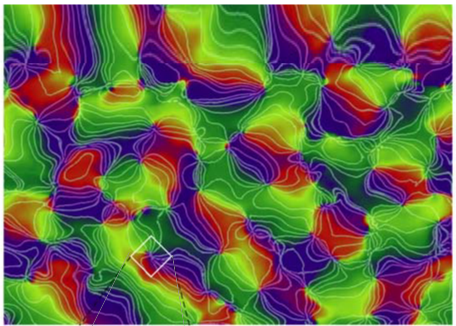


More realistic views of ocular preference changes (left) with orientation preference changes superimposed (right) for adjacent hypercolumns

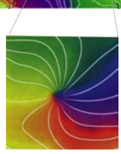


V1 organisation: retinotopic maps

However, across the whole surface of V1 there are retinotopically organised maps for:

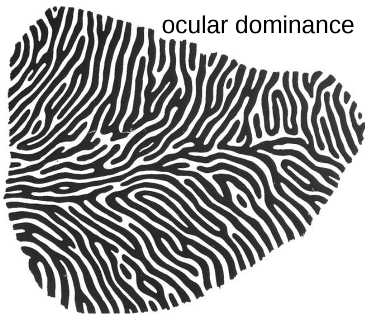


orientation preference



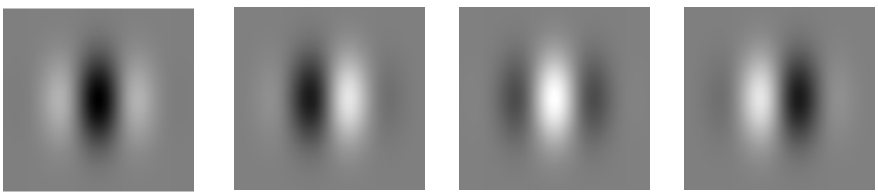
one hypercolumn

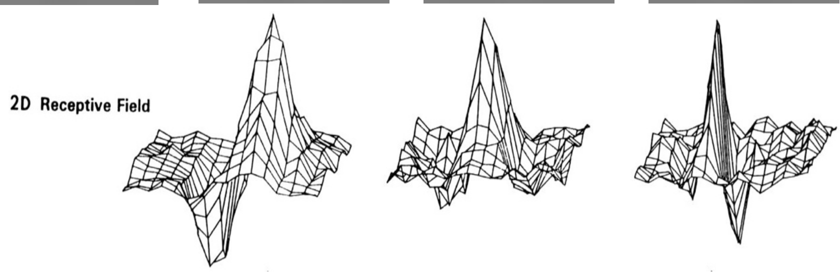
ocular dominance



Gabor: model of orientation selective RFs

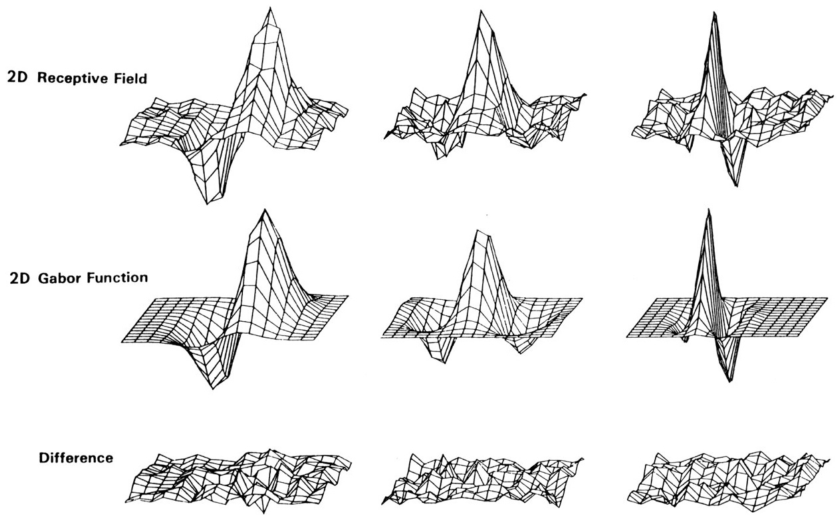
Simple cells RFs actually look more like this:





These are very similar to a mathematical function called a Gabor.

Gabor: model of orientation selective RFs



Gabor: model of orientation selective RFs

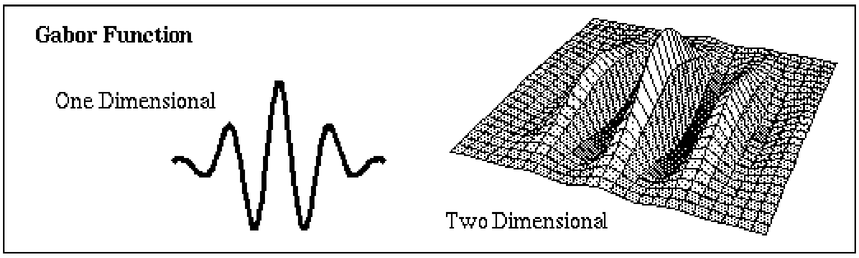
A 2-D Gabor function is a Gaussian multiplied by a sinusoid.

$$G(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos(2\pi x' f + \psi)$$

where:

$$x' = x\cos\theta + y\sin\theta$$

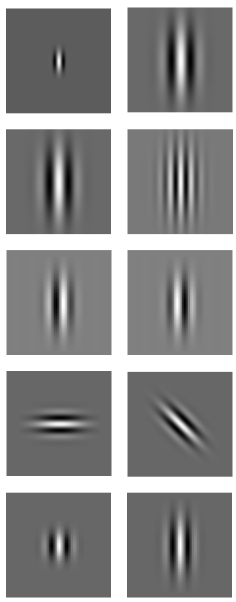
$$y' = -x\sin\theta + y\cos\theta$$



Gabor: model of orientation selective RFs

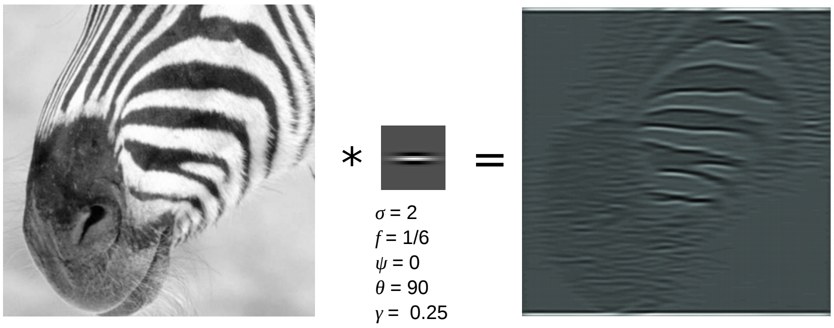
The Gabor function has 5 parameters:

- σ = the variance of the Gaussian function
- f = the frequency of the sinusoidal function
- ψ = the phase of the sinusoidal function
- θ = the orientation
- γ = the spatial aspect ratio



Gabor: model example

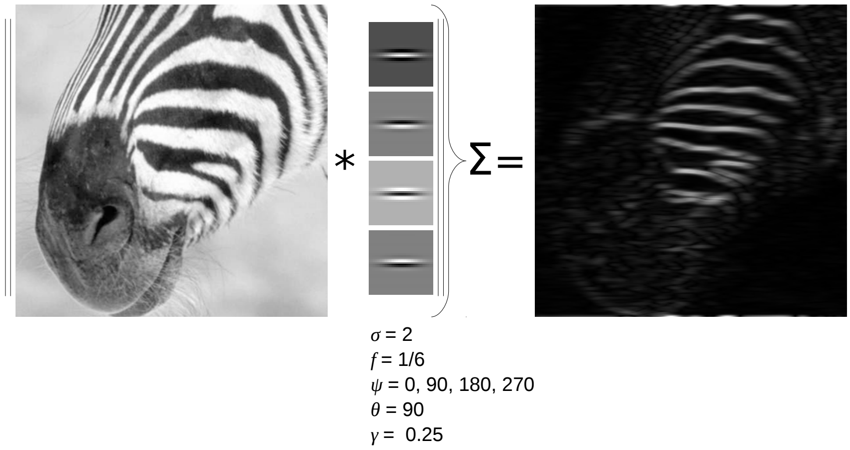
To simulate the activity of **simple cells** in response to an image, we can convolve the image using a Gabor function as the mask.



This provides a crude model of the responses of all simple cells selective for the same orientation, spatial frequency and phase across all cortical hypercolumns.

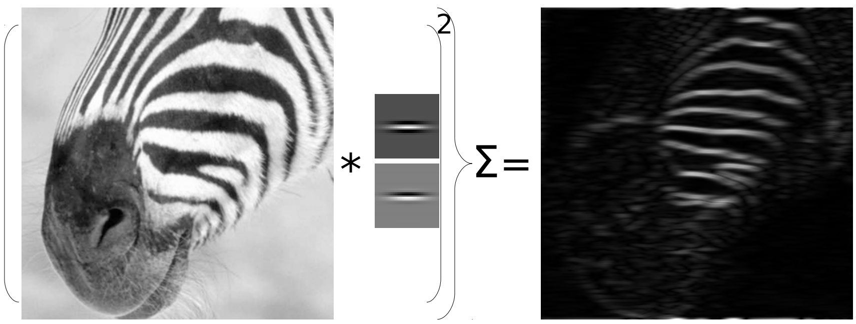
Gabor: model example

To simulate the activity of **complex cells** in response to an image, we can pool the outputs generated by convolving the image with Gabor functions of varying phases.



Gabor: model example

To simulate the activity of **complex cells** in response to an image, we can pool the outputs generated by convolving the image with Gabor functions of varying phases.

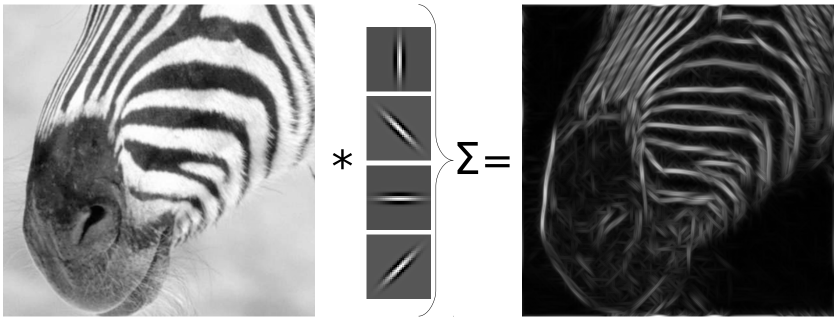


A mathematically simpler model (called the **energy model**) takes the square root of the sum of squared output of a quadrature pair of Gabor filters.

Result is invariant to phase.

Gabor: model example

To detect edges at multiple orientations, we can sum the outputs from simulated complex cells at multiple orientations.



$\sigma = 2$
$f = 1/6$
$\psi = 0, 90, 180, 270$
$\theta = 0, 45, 90, 135$
$\gamma = 0.25$

Multiscale Gabors: wavelet transforms

To detect edges at multiple spatial scales, we can use Gabor functions with different frequencies and variance.

Convolving a signal (e.g. an image) with a family of similar masks sensitive to different frequencies is known as a "Continuous Wavelet Transform".

The "wavelet" is the mathematical function used to generate the masks (e.g. a Gabor).

Hence, we can think of the simple cells in V1 as performing a continuous wavelet transform.

Gabors: as image components

We can think of an image being made up of a superposition of various image components or elementary features.

e.g. A letter **I** as a combination of three lines (Gabor image components):

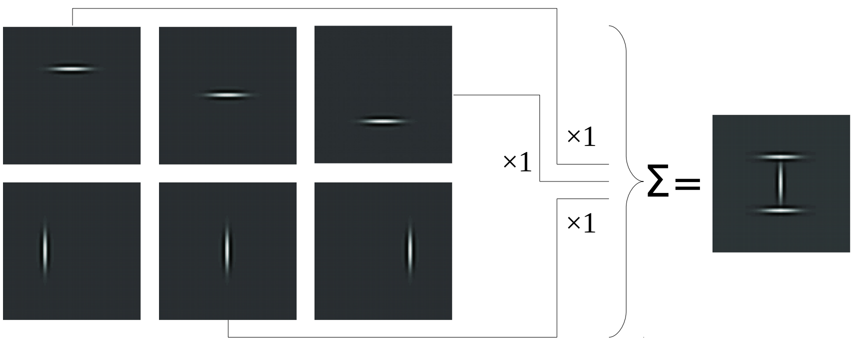


image components (A) activations (y) image (x)

Gabors: as image components

Combining different components in different proportions can generate numerous different images.

e.g. A letter H as a combination of three lines (Gabor image components):

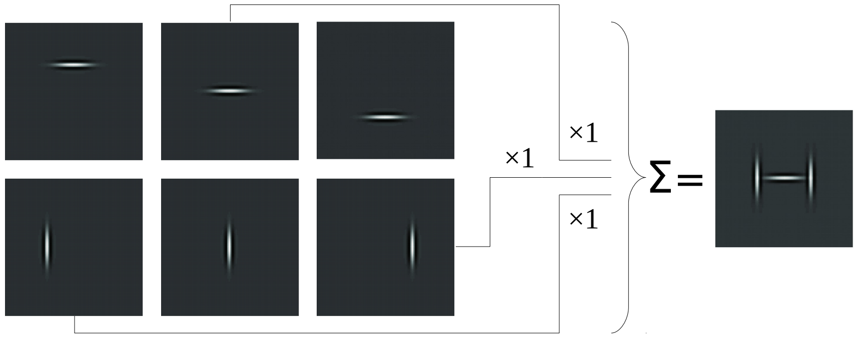


image components (A) activations (y) image (x)

Image components

This is a very general concept, e.g.:

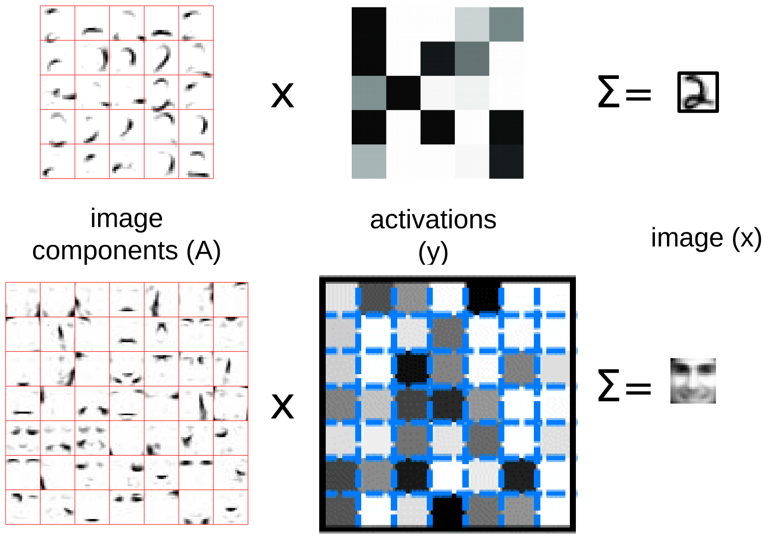


Image components

Mathematically, we can express this idea as:

$$A\,y \approx x$$

where:

A = is an m by n matrix of weight values the columns of which represent components (or basis vectors) into which the images can be decomposed

y = is a n by 1 vector describing the activation of each component in the image

x = is a m by 1 vector representing the image

(Note an image is represented by a vector rather than an array)

Image components

If we have lots of images, then this equation becomes:

$$AY \approx X$$

where:

A = is an m by n matrix of weight values the columns of which represent components (or basis vectors) into which the images can be decomposed

Y = is a n by p matrix each column of which contains the activation of each component in the corresponding image

X = is a m by p matrix each column of which contains the pixel values for one image

(Note an image is represented by a vector rather than an array)

Image components

Generally we know X, and want to determine how this set of images can be represented, i.e. we want to find the factors A and Y of X.

$$AY \approx X$$

Finding the factors A and Y is an ill-posed problem (there are lots of possible solutions).

By placing additional constraints on the factors A and Y different solutions, with different properties, can be found.

There are several standard "Matrix factorisation" methods that can find A and Y under different constraints. e.g.:

- Principal Component Analysis (PCA)
- Independent Component Analysis (ICA)
- Non-negative Matrix Factorization (NMF)

Gabors as components of natural images

What are the components of natural images?
If we create X by randomly selecting small image patches

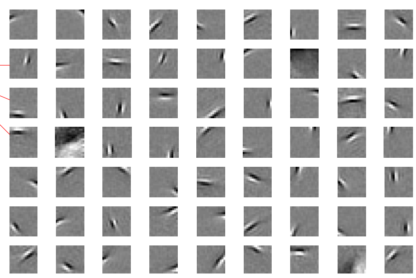
X

Then find the factors using two constraints:

1. That information be preserved (i.e. that $||X\text{-}AY||$ is minimised)

2. The representation be sparse (i.e. the number of non-zero values in each column of Y is minimised)

A



Then the columns of A are very similar to Gabor functions.

Gabors therefore appear to capture the intrinsic structure of natural images

Gabors as efficient code

The sparsity constraint used to find the components of natural images, means that only a few components are present in each image.

If components are represented by neurons, then this means that only a few neurons need to be active in order to represent an image.

Hence, Gabor RFs result in an efficient code which minimises the number of neuronal spikes needed to transmit a given signal.

Image components: Computer Vision Applications

Image Compression

The ability to reconstruct an image using image components allows an image to be efficiently encoded.

e.g. jpeg

Reconstructs each image patch as a linear combination of a set of image components

image components are cosine functions not Gabor functions

$$A\,y \approx x$$

y requires less storage space than x

activations are quantized rather than sparse

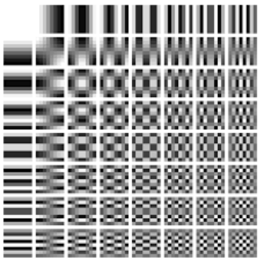


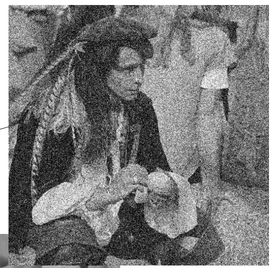
Image components: Computer Vision Applications

Image Denoising

Reconstruct a noisy image using image components produces a reconstructed image that has less noise.

image components are Gabor functions (or curvelets, wedgelets, or other functions)

$$A\,y \approx x$$



activations are found that produce the most accurate reconstruction of the noisy image (i.e. that minimise $||x-Ay||$)

$$A\,y =$$

Image components: Computer Vision Applications

Image Inpainting

Reconstruct missing parts of an image using a sparse-subset of image components that represent non-corrupted image patches

image components learnt from non-corrupted parts of image.

$A\,y \approx x$

activations are found that produce the most accurate reconstruction (that minimise $||x\text{-}Ay||$)

$A\,y =$

V1 RFs: non-classical

Classical Receptive Field (cRF) = the region of visual space that can elicit a response from a neuron.

However, this response can be modulated by stimuli appearing outside the neurons classical receptive field.

The region of visual space that can influence the response is called the non-classical receptive field (ncRF).

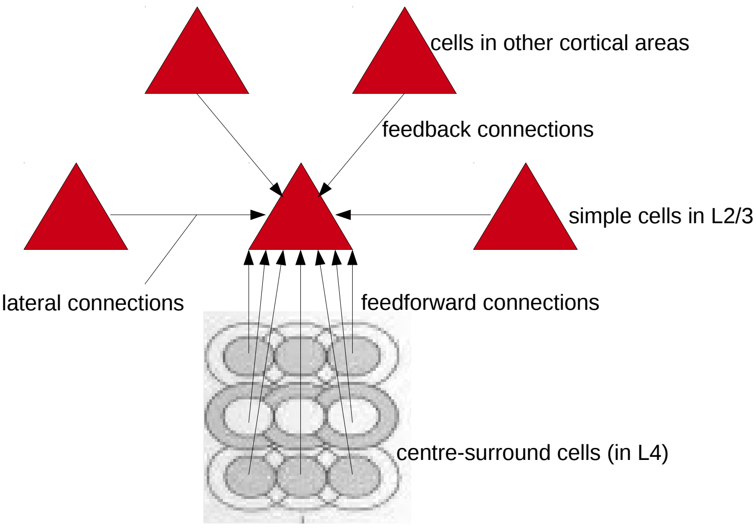
The result is that neuronal responses are influenced by context, i.e., neuronal activity at one location depends on activity at possibly distant locations.

Contextual influences on neuronal activation (in general) give rise to contextual influences on perceptual inference:

V1 RFs: non-classical

cRF = Feedforward connections

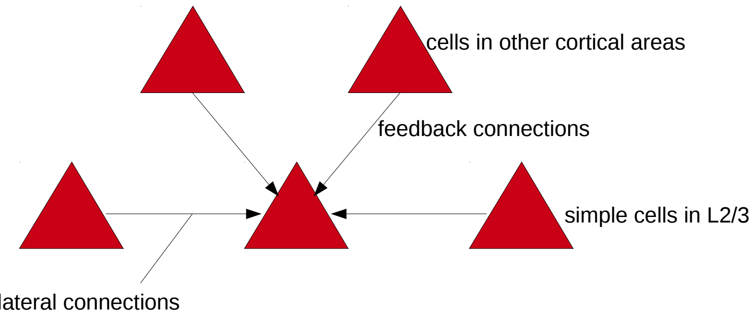
ncRF = lateral and feedback connections



V1 RFs: non-classical

cRF = Feedforward connections

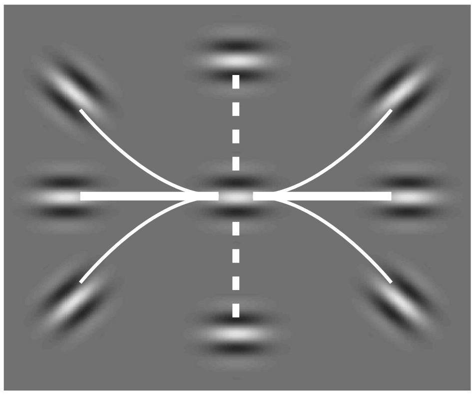
ncRF = lateral and feedback connections



Lateral and feedback connections can generate both excitatory and inhibitory effects.

V1: lateral connections

excitation from collinear/ co-circular neighbours (solid lines)

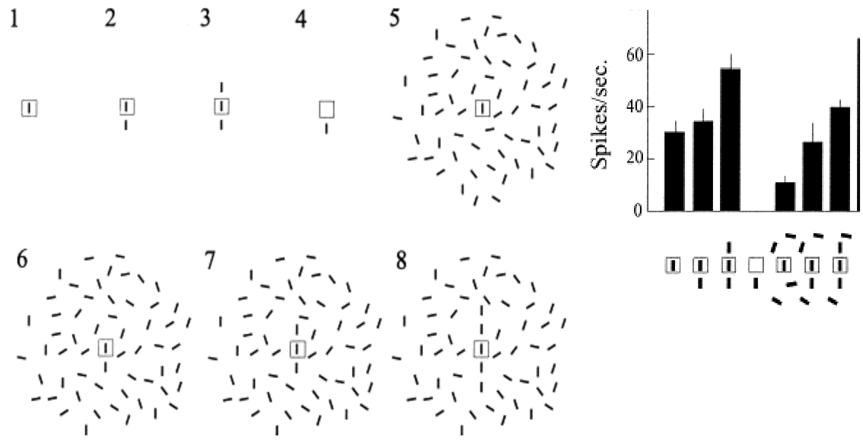


inhibition from parallel iso-orientated neighbours (dashed lines)

Thick lines indicate strongest connections (greatest influence)

This pattern of lateral connectivity is sometimes called the “association field”

V1: contextual influences

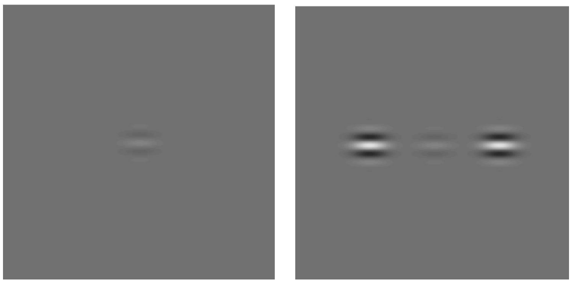


- context has no effect in isolation
- context can enhance response
- context can reduce response

V1: contextual influences

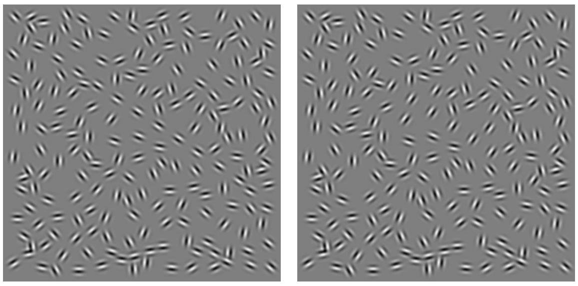
Collinear Facilitation

Which of the central Gabor patches is easier to see?



Contour Integration

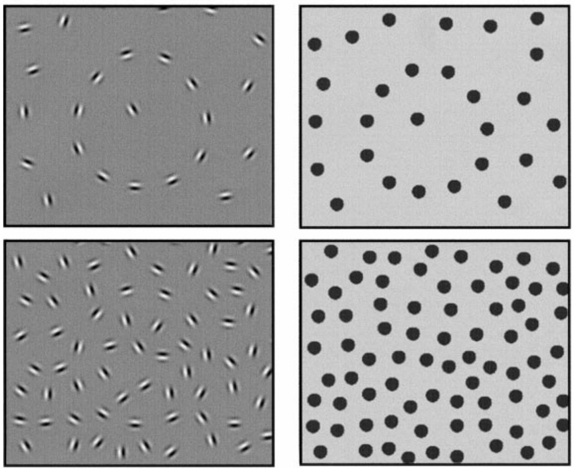
Which of these images contains a straight line?



V1: contextual influences

Contour Integration

Which of these images contains a circle?

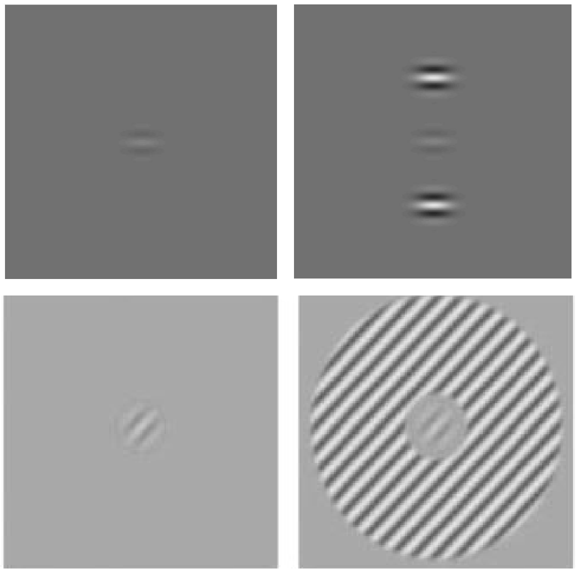


Lateral excitatory connections enhance responses to aligned elements, making these easier to perceive.

V1: contextual influences

Surround Suppression

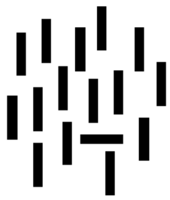
Which of the central Gabor patches is easier to see?

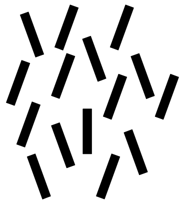


V1: contextual influences

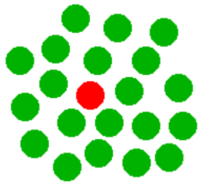
<u>Pop-out</u>

Which is the odd element in each figure?

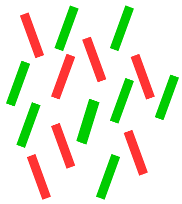




dissimilar distractors



feature search



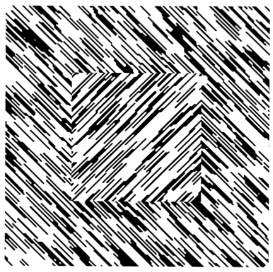
conjunction search

Lateral inhibitory connections suppress responses to similar elements, making dissimilar elements easier to perceive.

V1: contextual influences

<u>Texture Segmentation</u>

Where is there a boundary in each figure?

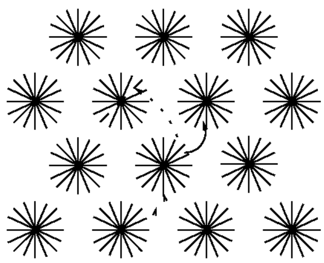


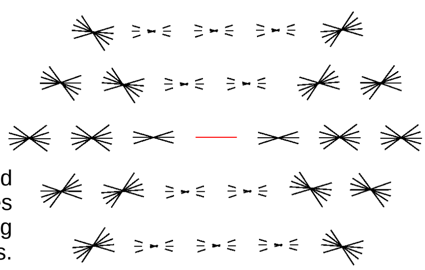
Similar to pop-out, but with two regions. Lateral inhibitory connections suppress responses to similar elements. At the boundary between dissimilar elements neurons receive less inhibition making it easier to perceive.

Modelling contextual influences in V1

A neural network model:

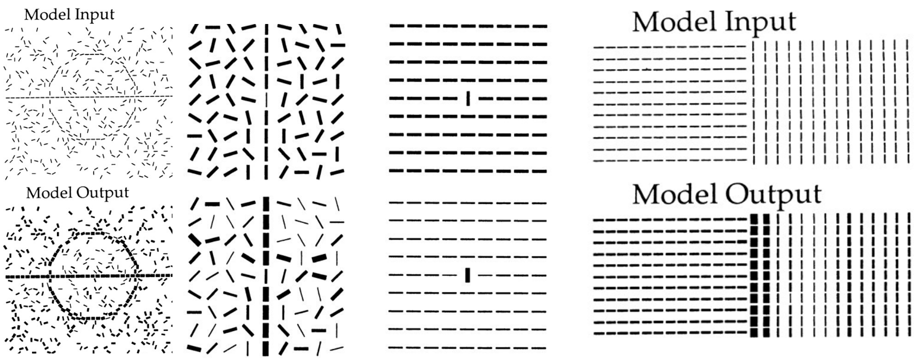
- neurons have feedforward weights selective for oriented bars
- multiple neurons at each location represent different orientations
- lateral excitatory and inhibitory connections between neurons with the same connectivity pattern as found in V1.





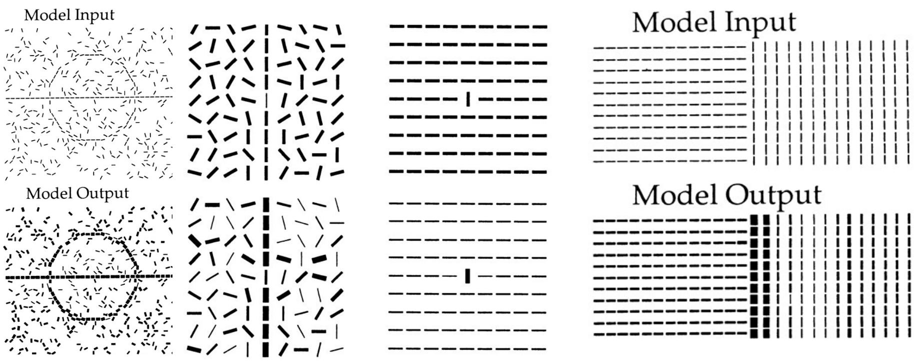
Sources of lateral connections received by central neuron. Solid (dashed) lines indicate RFs of neurons sending excitatory (inhibitory) connections.

Modelling contextual influences



contour integration pop-out texture segmentation

Modelling contextual influences



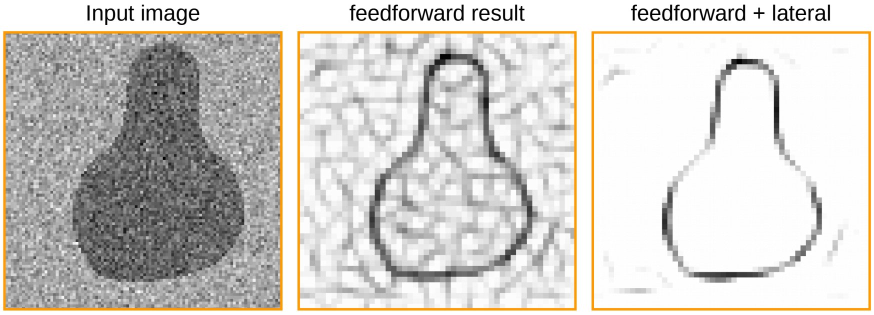
In general, contextual influences in V1 operate to highlight (increase the "saliency" of) important or conspicuous locations in the image, e.g.

- smooth contours against background noise,
- odd items within a homogeneous background,
- boundaries between textured regions.

Modelling contextual influences

Noise suppression

Input image feedforward result feedforward + lateral



Summary

Cortical Visual system is very complex

Consists of hierarchies of processing stages (cortical areas)

1st stage is V1

Summary

cRFs (pattern of feedforward connections)

large variety (colour, orientation, motion, frequency, disparity,...)

mapped across cortical surface

- a patch of V1 represents a particular spatial location
- neighbouring patches represent neighbouring locations

A hypercoloumn is a patch which contains all the RFs for the same location

Gabor functions

- provide good model of orientation selectivity (convolution of an image with a Gabor performs edge detection)
- provide efficient codes for natural images

Summary

ncRFs (pattern of lateral, and feedback, connections)

enables stimuli outside a neuron's cRF to modulate its response

nodes with similar cRFs at neighbouring locations connected via:

- inhibitory connections (generate pop-out, texture segmentation)
- excitatory connections (generate contour enhancement)

increases salience of locations where image changes