Cognitive Neuroscience for AI Developers

Week 9 and 10 – Vision

Next week is **NO LIVE LECTURE!**

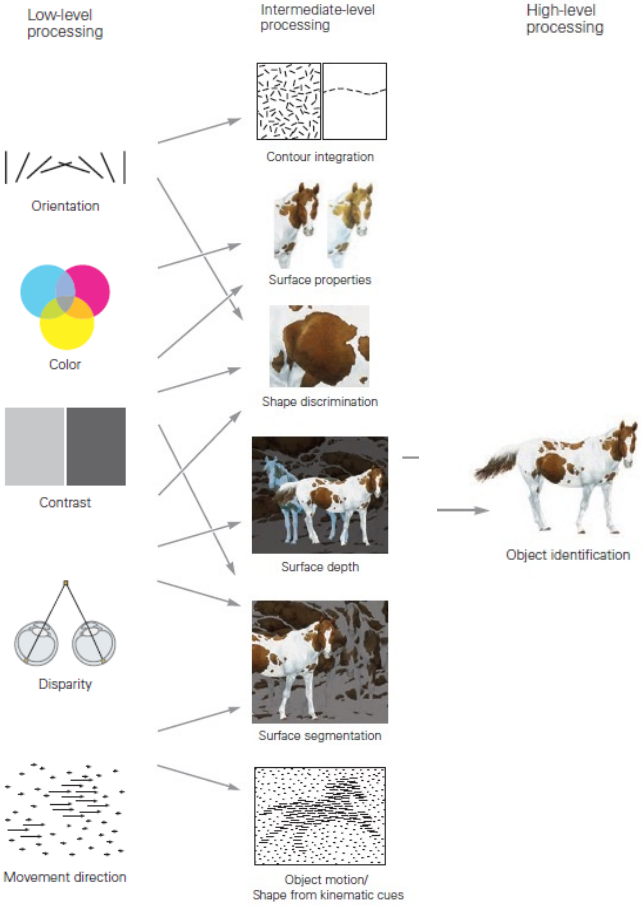
But there will be a video from last year and its content is exam relevant!

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF ENGINEERING

Exam: **August 1st, (Thursday)**

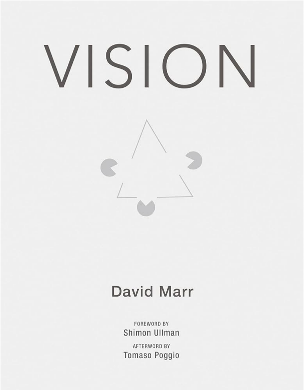
Time to be announced as early as possible, we are aware that people may write more than one exam on the same day. But most likely not at the same time.

A visual scene is analyzed in 3 levels



Marr‘s idea of vision

2D Primal Sketch	2.5D Sketch	3D model
Fundamental components	Textures, shades	continuous



Marr's tri-level hypothesis

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF ENGINEERING

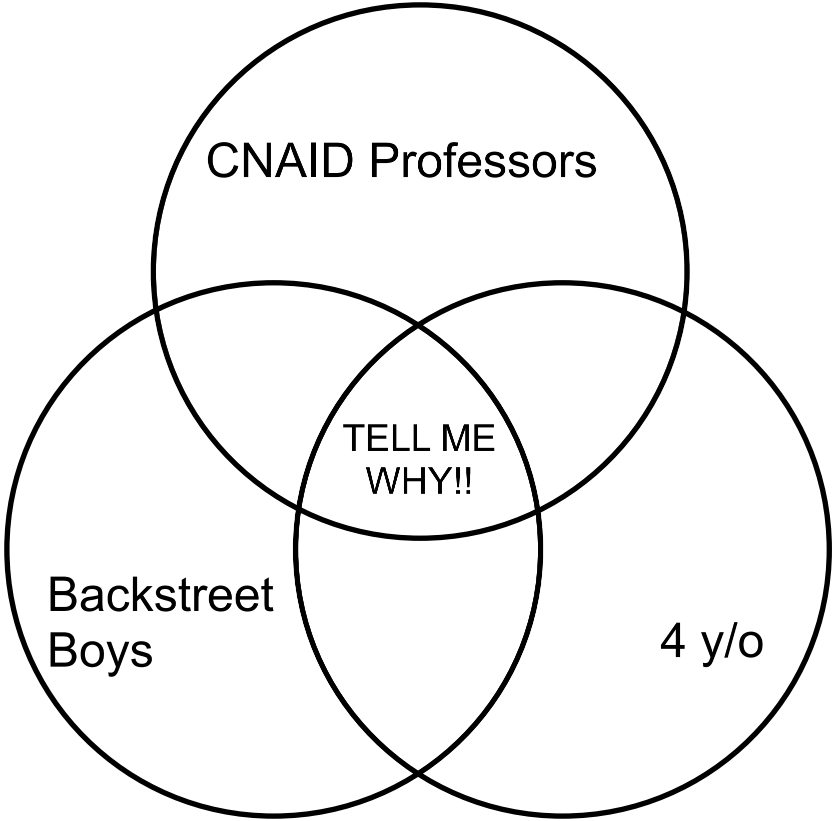
Computational level

What problem does it try to solve? **WHY?!**

Algorithmic/Rep. level

Implementation level

Why?!



Marr‘s tri-level hypothesis

Computational level

What problem does it try to solve? **WHY?!**

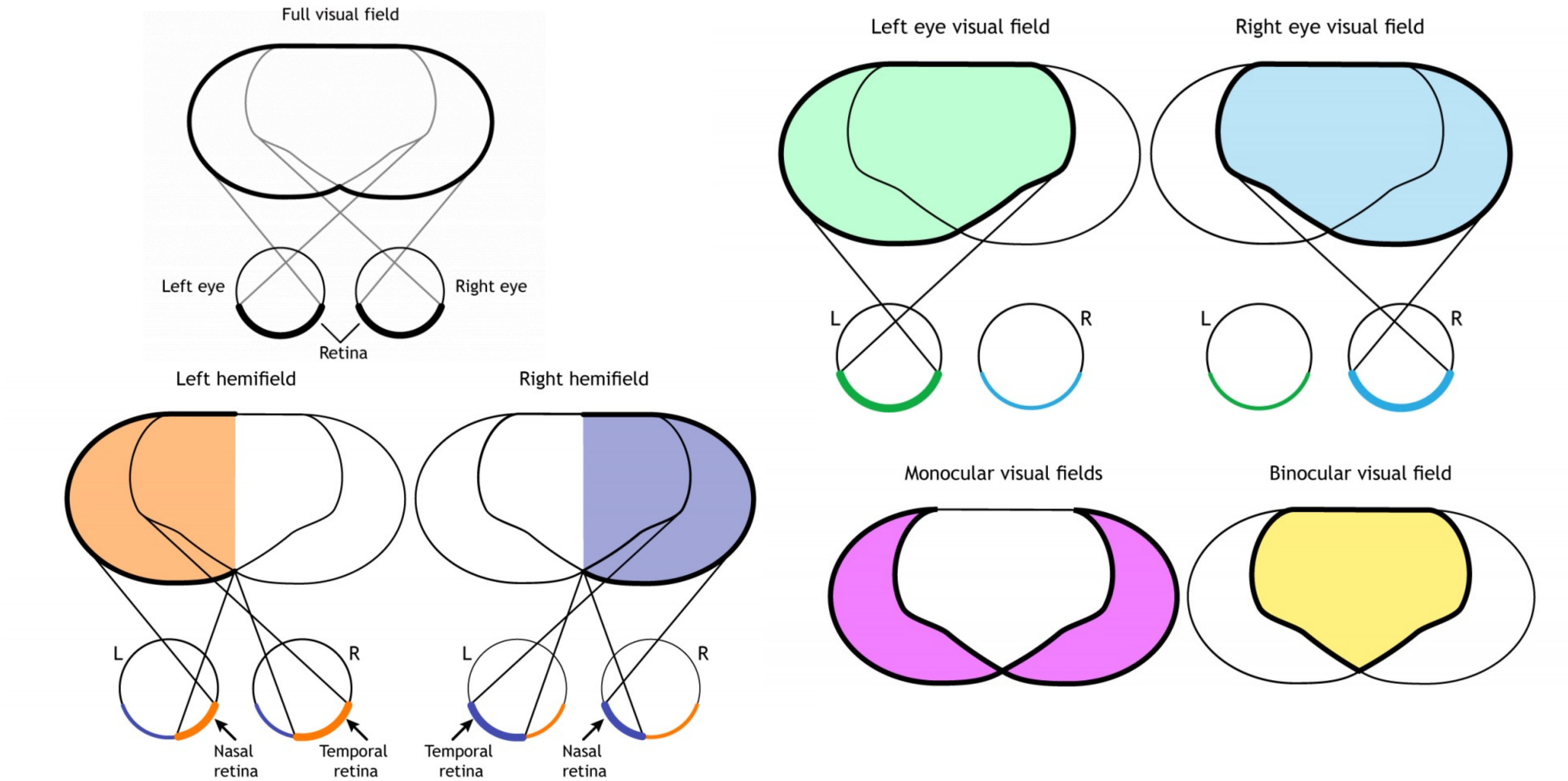
Algorithmic/Rep. level

HOW does it solve this? What are the processes?

Implementation level

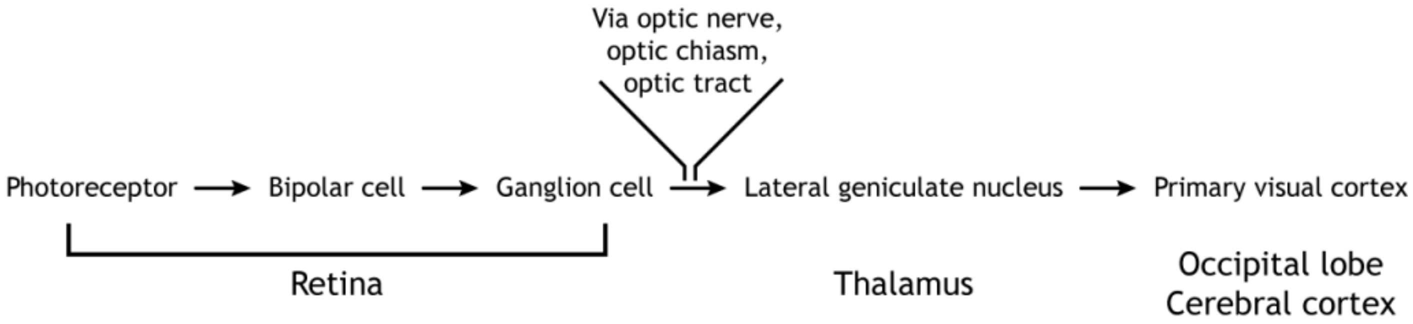
HOW is it implemented (… in biology?)?

What can we see?



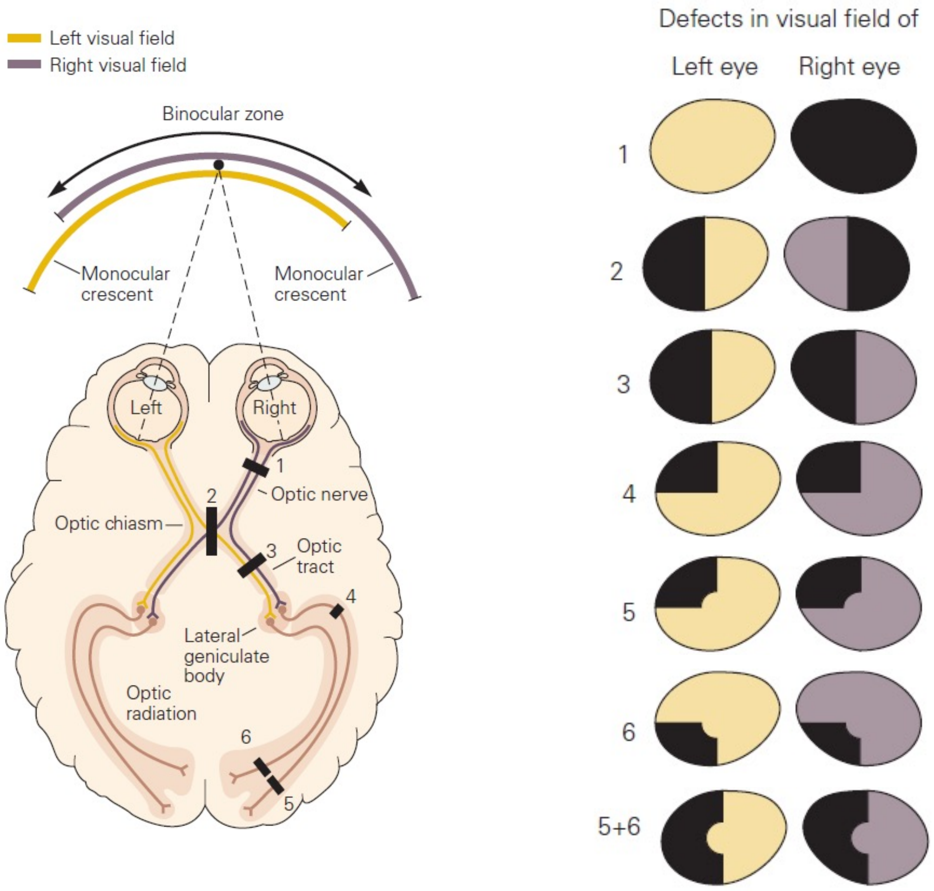
https://openbooks.lib.msu.edu/neuroscience/chapter/vision-central-processing/

Visual pathway

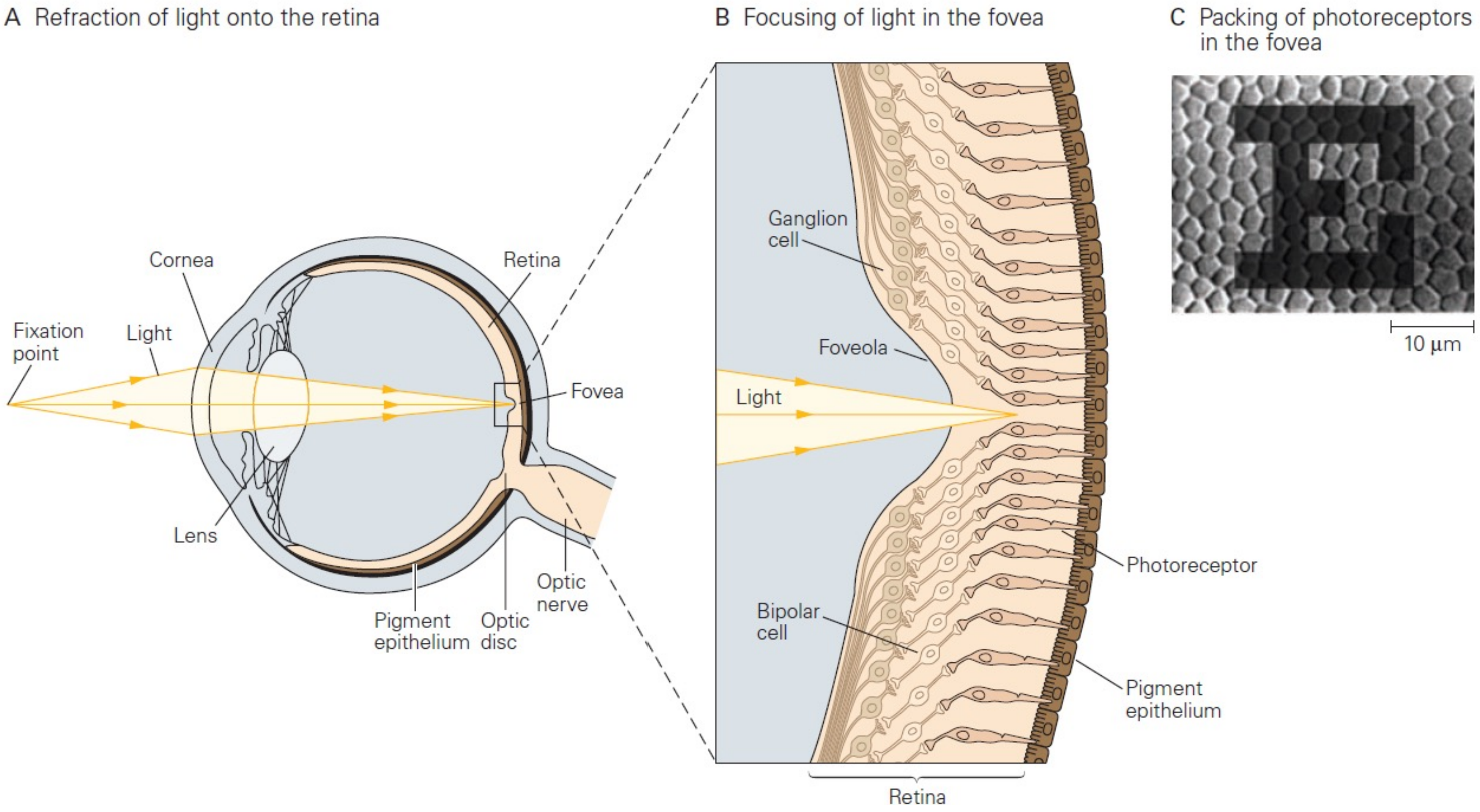


https://openbooks.lib.msu.edu/neuroscience/chapter/vision-central-processing/

Visual field along the visual pathway

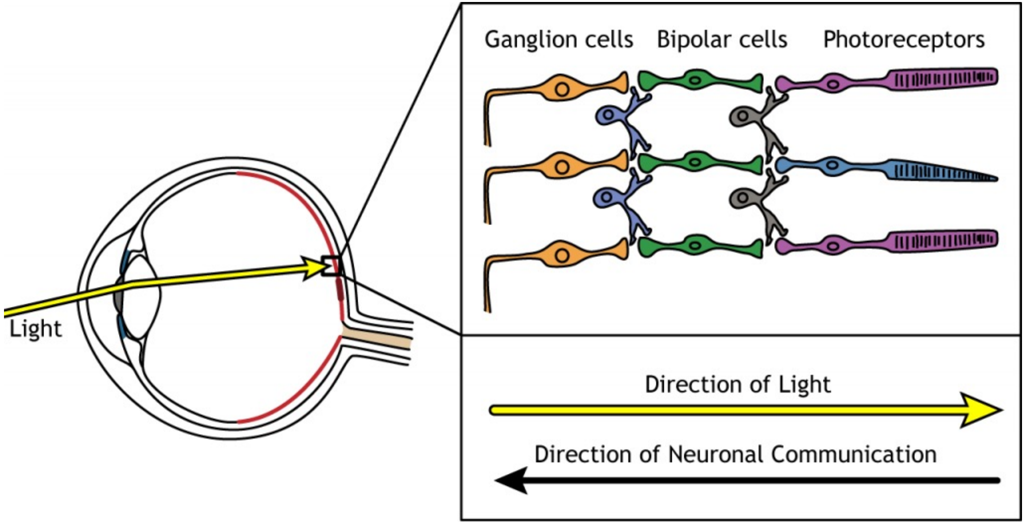


Low-level visual processing



Neural communication is against direction of light

vs. cephalopods

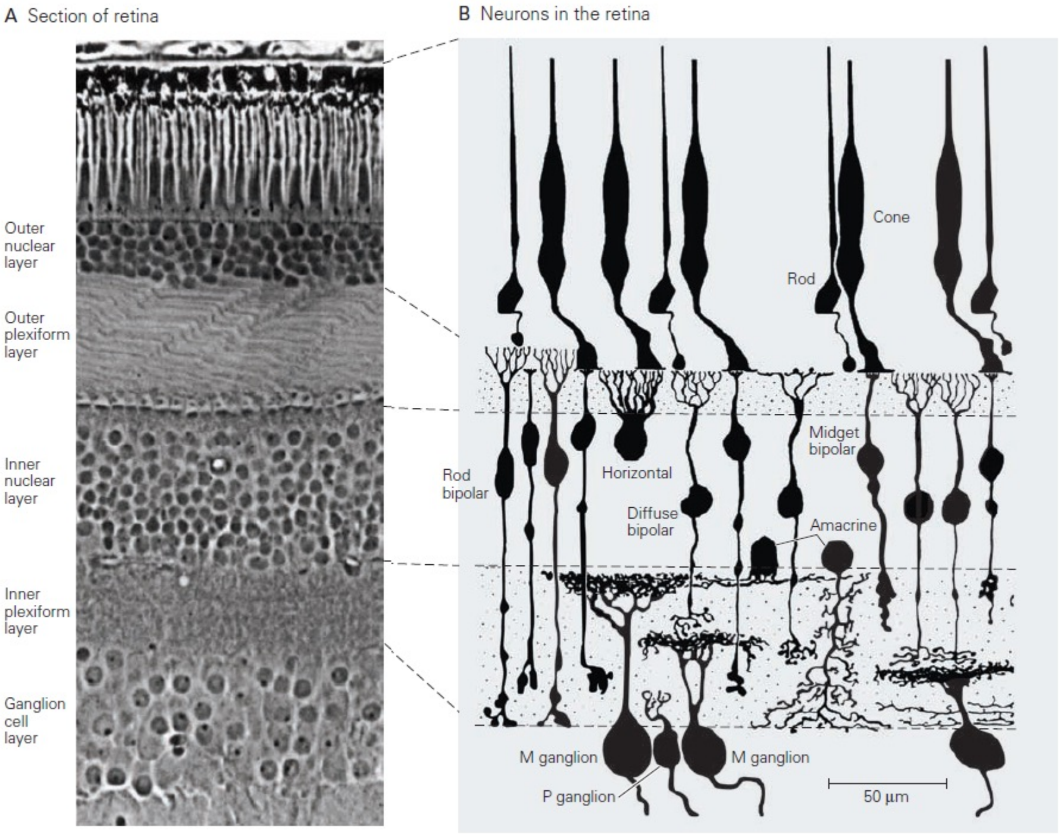


Why inverted in vertebrates?

- Save-spacing in inverted eyes against everted eyes (Kröger & Biehlmaier, 2009)
- Existing Retinal Pigment Epithelium (RPE) containing melanin absorbing light and avoiding back reflections
- Better maintenance of photoreceptors
- Light guides by ganglion cells

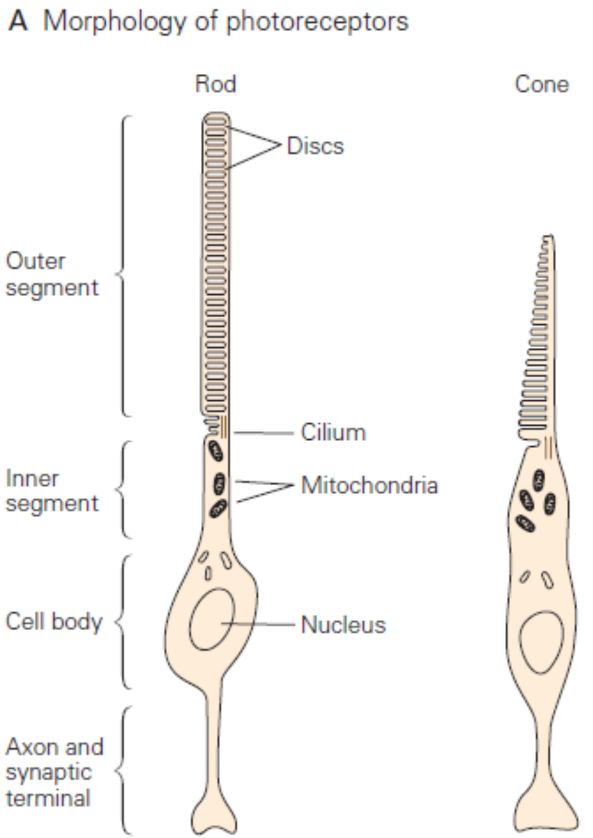
https://openbooks.lib.msu.edu/neuroscience/chapter/vision-the-retina/

Retina components

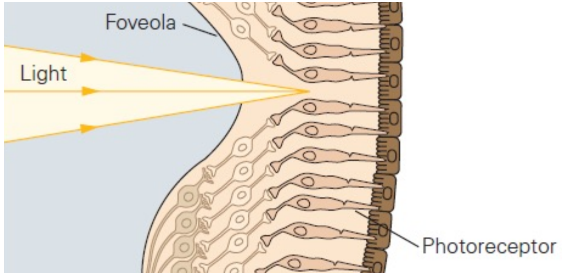


Rods and cones

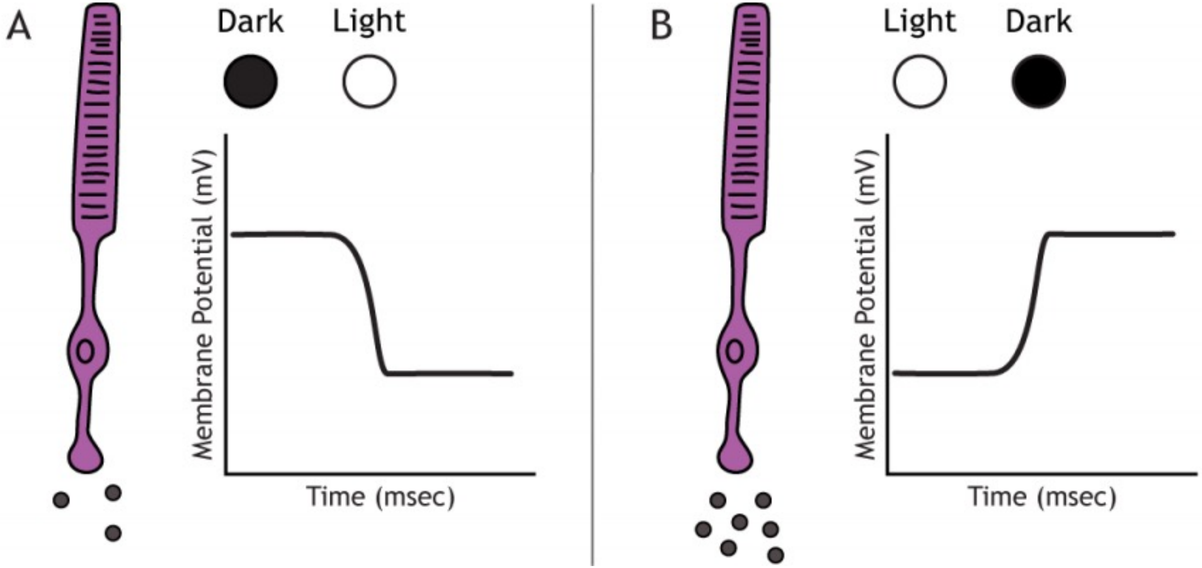
Black&white



Color

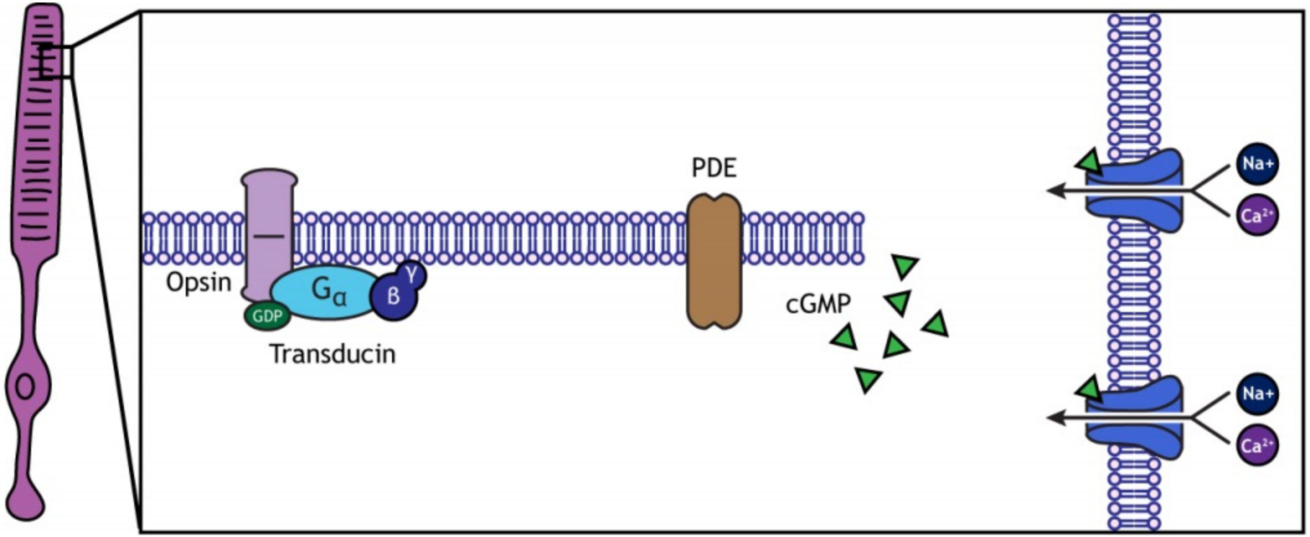


How does our „Analog-Digital-Converter“ work?



https://openbooks.lib.msu.edu/neuroscience/chapter/vision-the-retina/

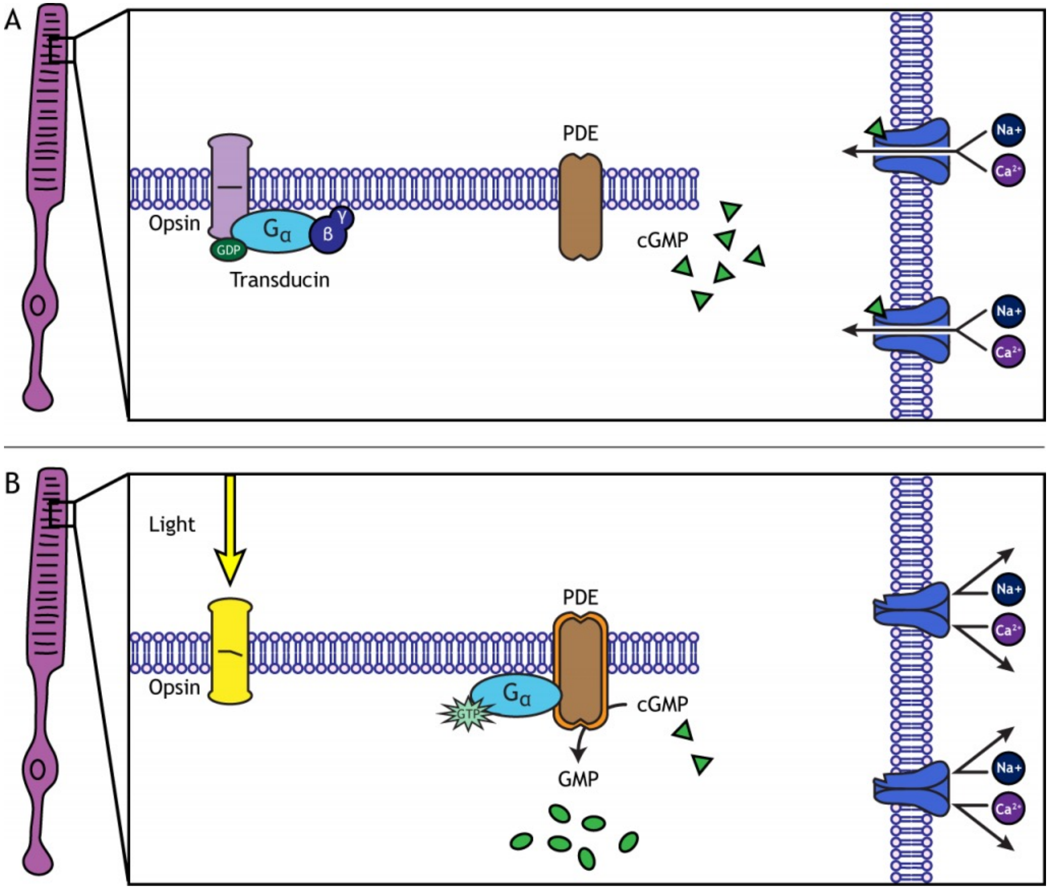
The light sensing cascade



Opsin: The light-sensing channel
PDE: Phosphodiesterase
cGMP: circular Guanosine monophosphate

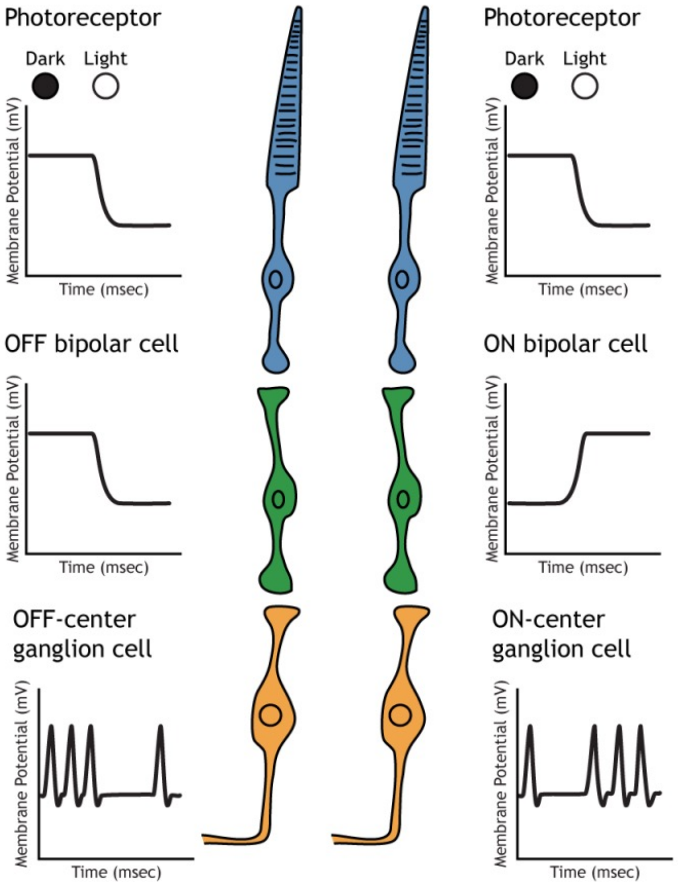
https://openbooks.lib.msu.edu/neuroscience/chapter/vision-the-retina/

Darkness vs Light



https://openbooks.lib.msu.edu/neuroscience/chapter/vision-the-retina/

Ganglion cell activity

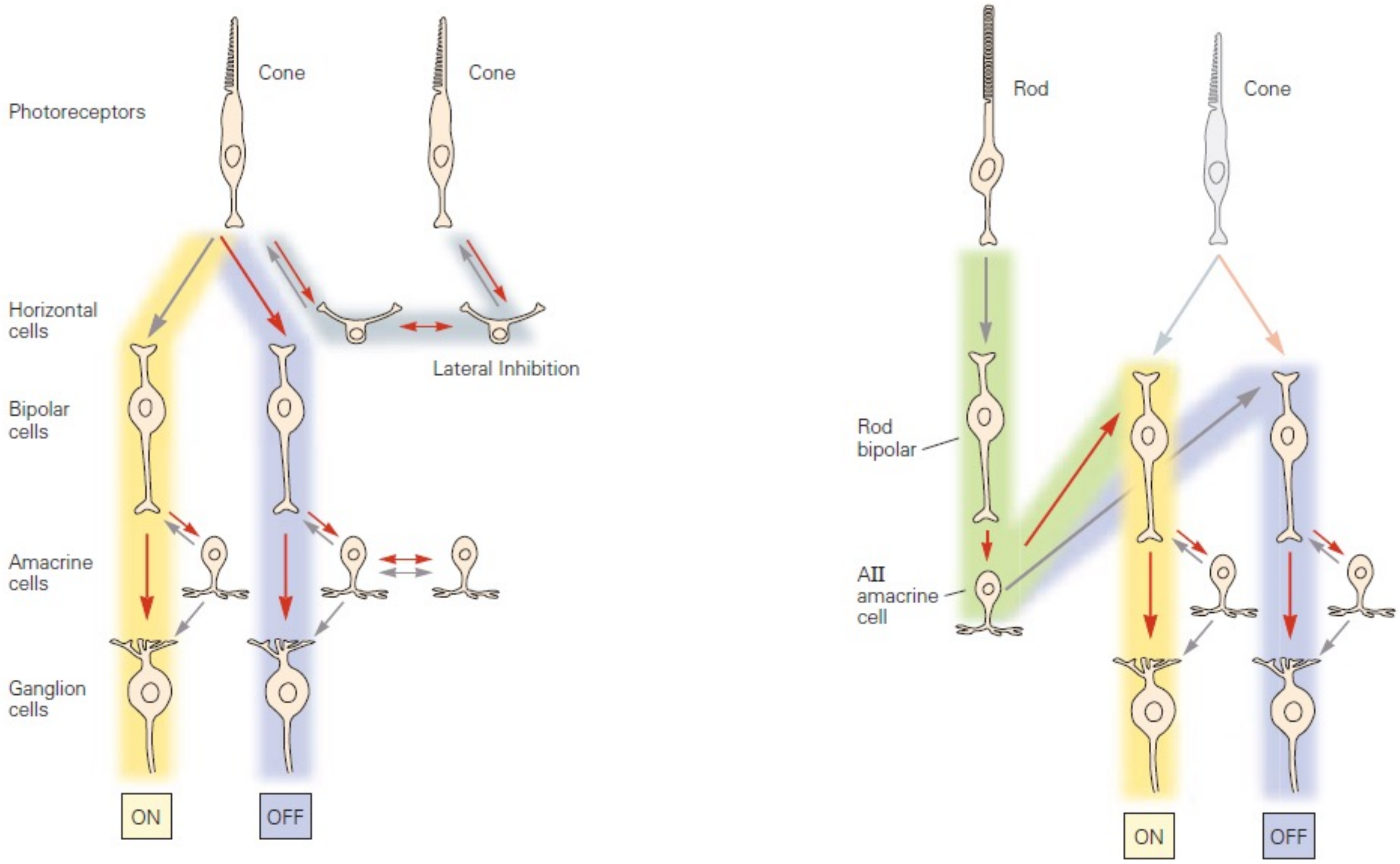


Bipolar cells can bypass signal or invert signal, can amplify signal

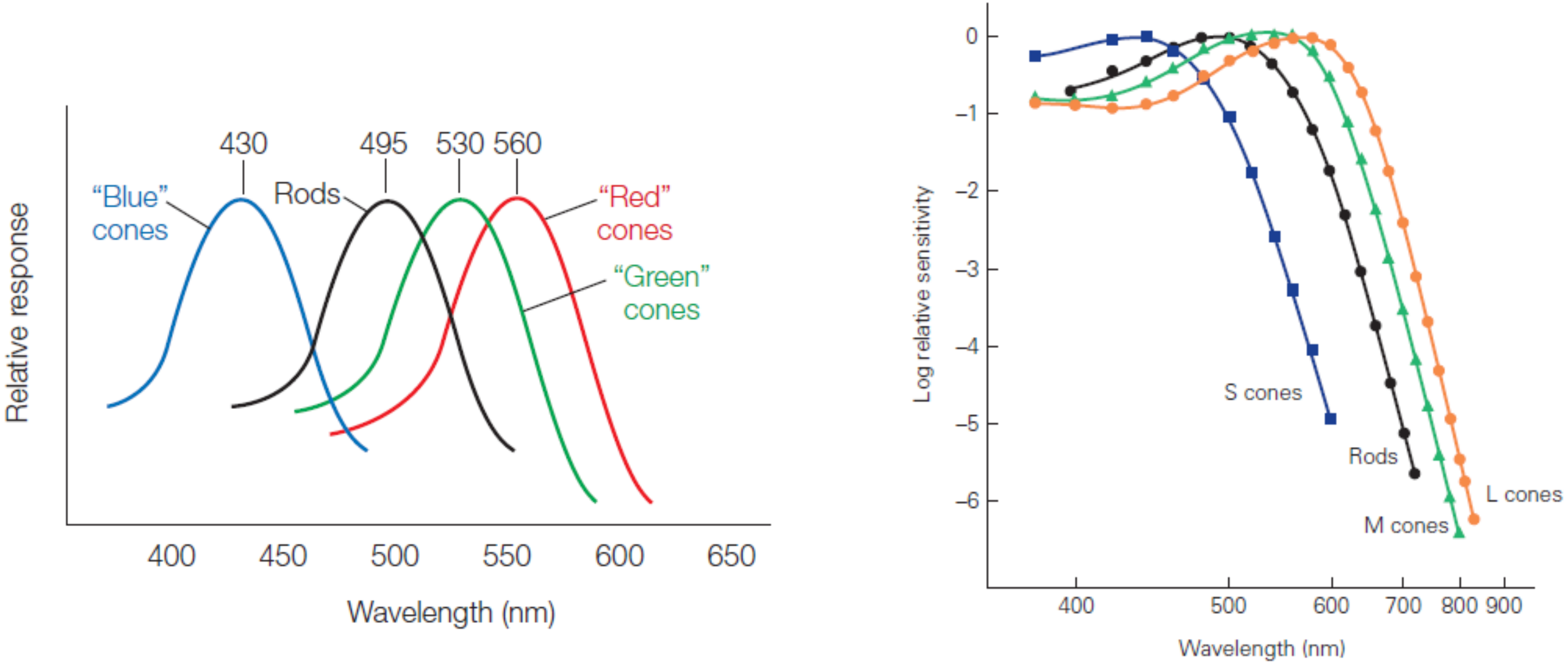
Ganglion cells encode signal using action potentials

https://openbooks.lib.msu.edu/neuroscience/chapter/vision-the-retina/

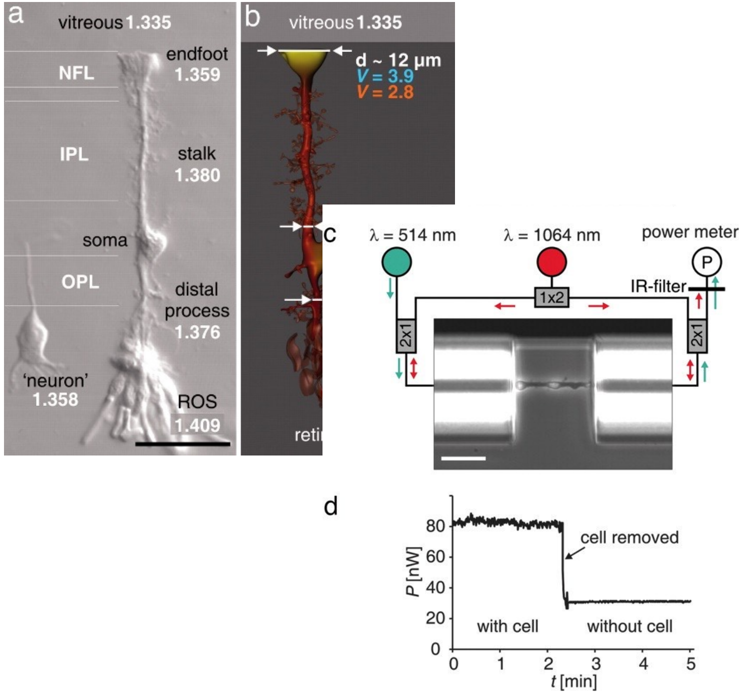
Retinal circuitry



Rod and cone spectral sensitivity

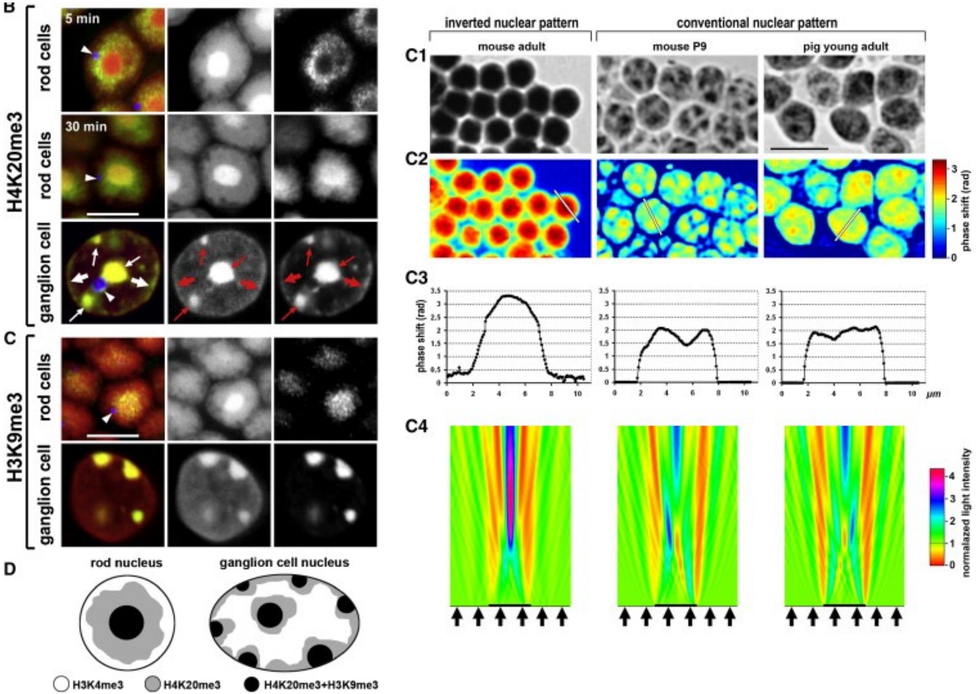


Retina physiology is fascinating

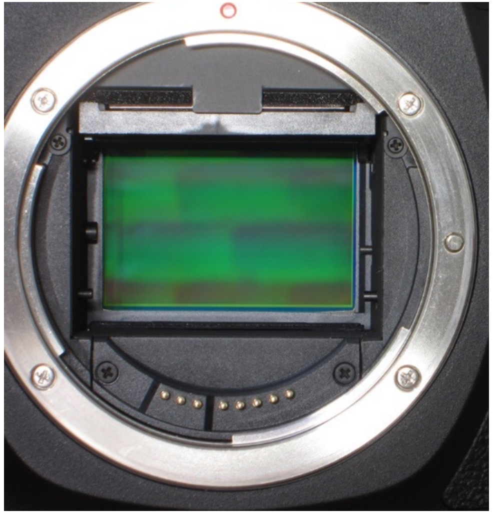


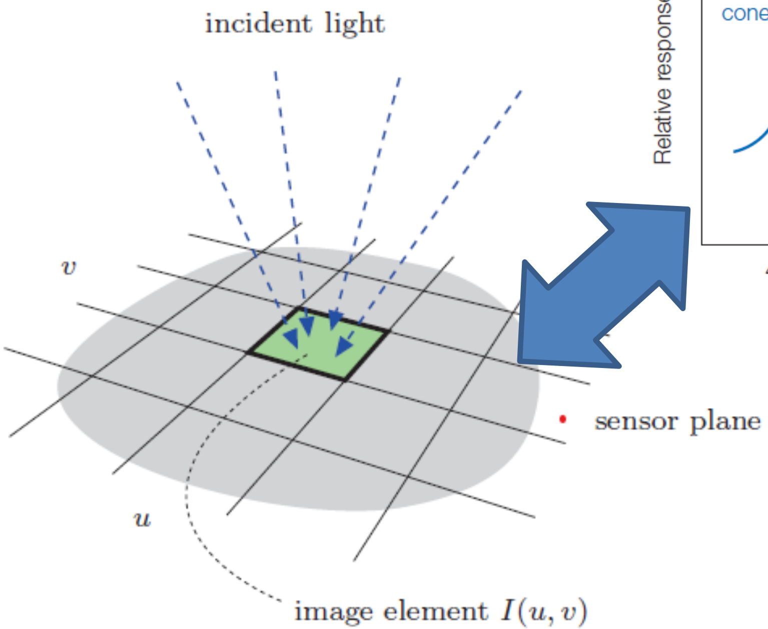
Müller cells are light guides

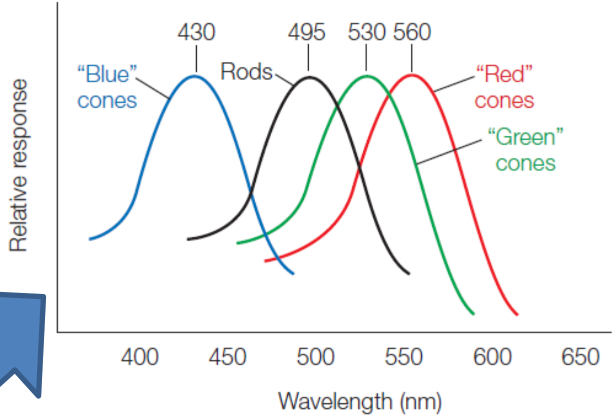
Inverted cell nucleus organization acts as a focusing lens



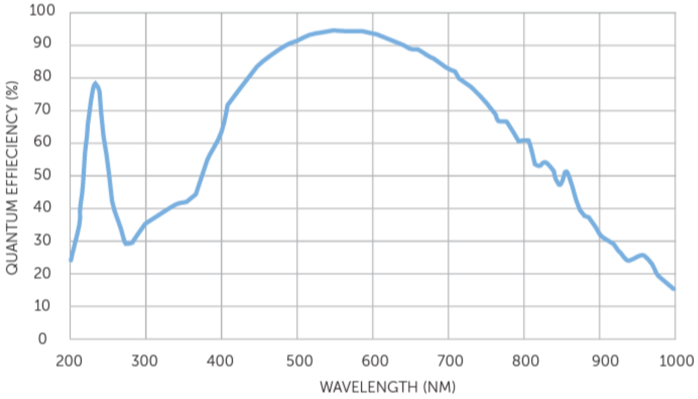
Generating color images for image processing tasks



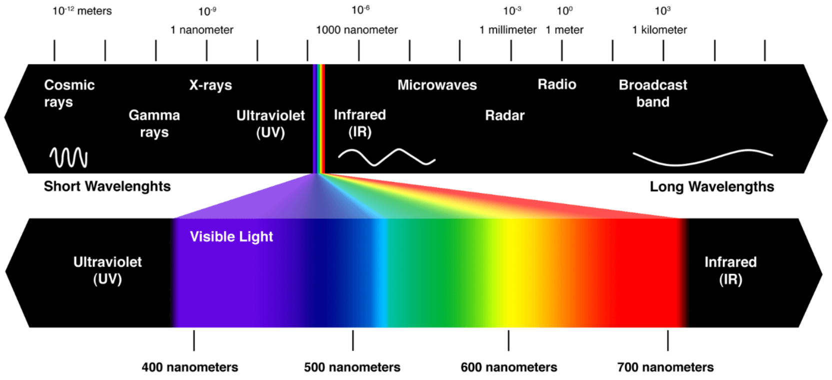




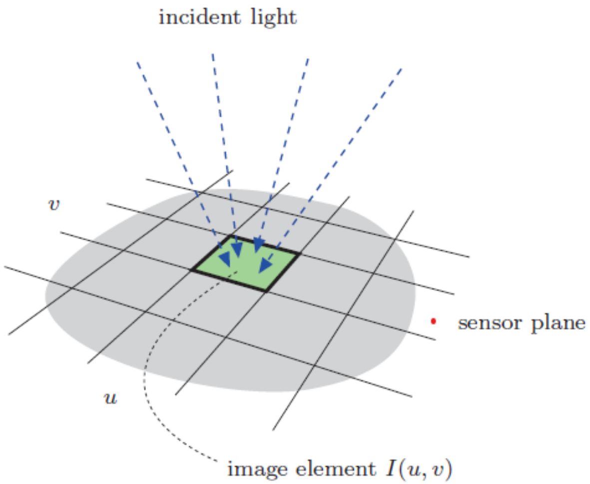
Spectral properties of a sCMOS camera



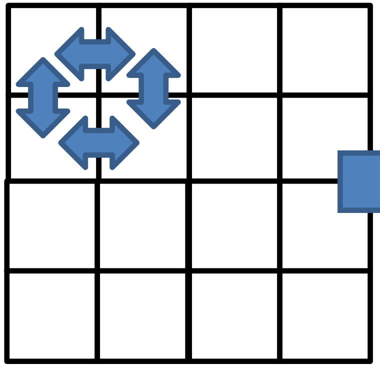
Frame Rate (PCIe interface)		
Array Size	16-bit	12-bit
1200 x 1200	40	80
1200 x 512	94	188
1200 x 256	188	374
1200 x 128	374	737



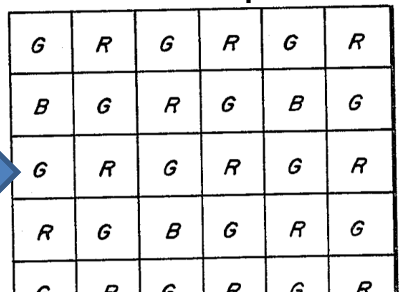
The Bayer pattern

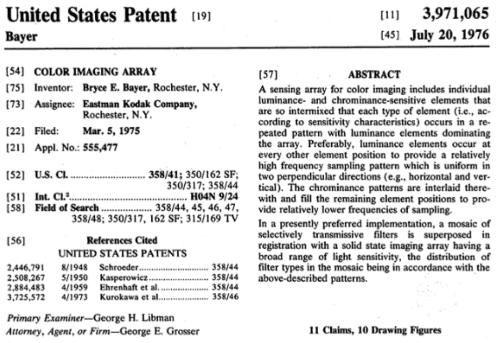


Adjacent pixels have high correlation!



Can we just put Small bandpass filter Over the pixels?!





United States Patent [19] [11] **3,971,065**

Bayer [45] **July 20, 1976**

[54] **COLOR IMAGING ARRAY**

[75] Inventor: **Bryce E. Bayer,** Rochester, N.Y.

[73] Assignee: **Eastman Kodak Company,** Rochester, N.Y.

[22] Filed: **Mar. 5, 1975**

[21] Appl. No.: **555,477**

[52] **U.S. Cl.** .... **358/41;** 350/162 SF; 350/317; 358/44

[51] **Int. Cl.²** **H04N 9/24**

[58] **Field of Search** .... 358/44, 45, 46, 47, 358/48; 350/317, 162 SF; 315/169 TV

[56] **References Cited**

UNITED STATES PATENTS

2,446,791	8/1948	Schroeder	358/44
2,508,267	5/1950	Kasperowicz	358/44
2,884,483	4/1959	Ehrenhaft et al.	358/44
3,725,572	4/1973	Kurokawa et al.	358/46

Primary Examiner—George H. Libman

Attorney, Agent, or Firm—George E. Grosser

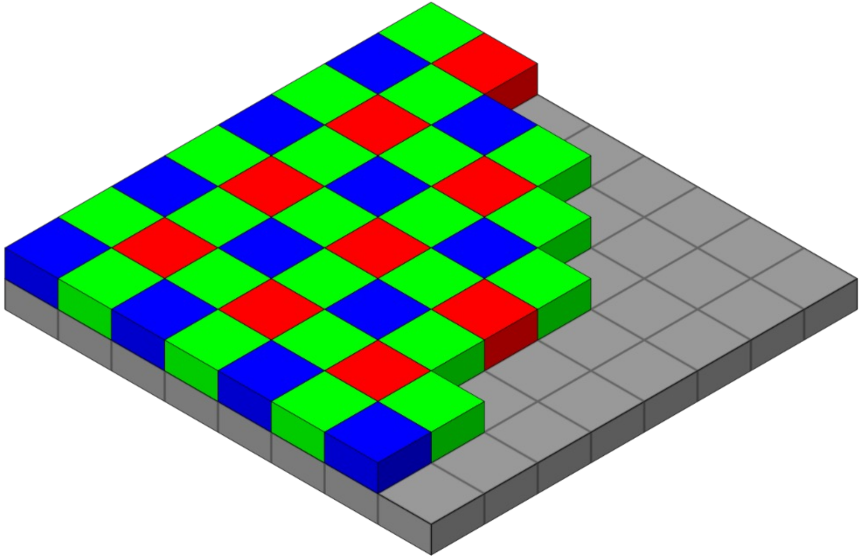
[57] **ABSTRACT**

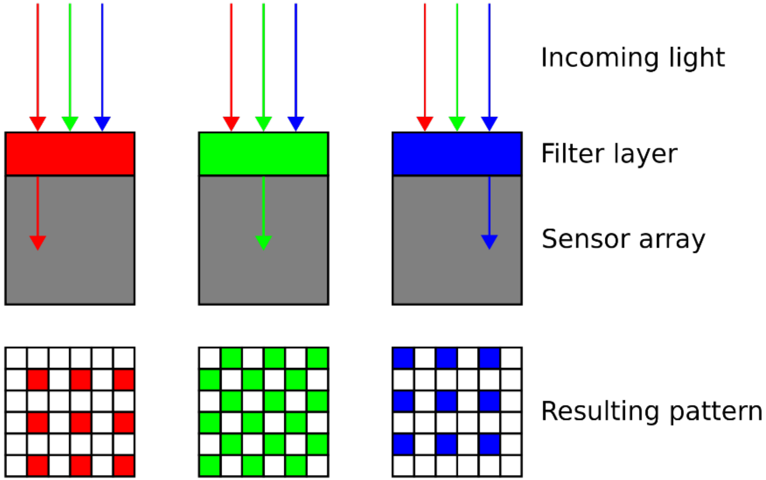
A sensing array for color imaging includes individual luminance- and chrominance-sensitive elements that are so intermixed that each type of element (i.e., according to sensitivity characteristics) occurs in a repeated pattern with luminance elements dominating the array. Preferably, luminance elements occur at every other element position to provide a relatively high frequency sampling pattern which is uniform in two perpendicular directions (e.g., horizontal and vertical). The chrominance patterns are interlaid therewith and fill the remaining element positions to provide relatively lower frequencies of sampling.

In a presently preferred implementation, a mosaic of selectively transmissive filters is superposed in registration with a solid state imaging array having a broad range of light sensitivity, the distribution of filter types in the mosaic being in accordance with the above-described patterns.

11 Claims, 10 Drawing Figures

Bayer pattern image generation

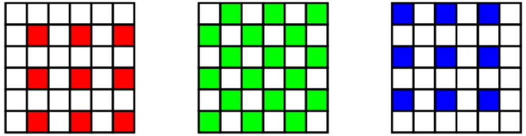




©Colin ML Burnett

Generate a color image

© Cmglee

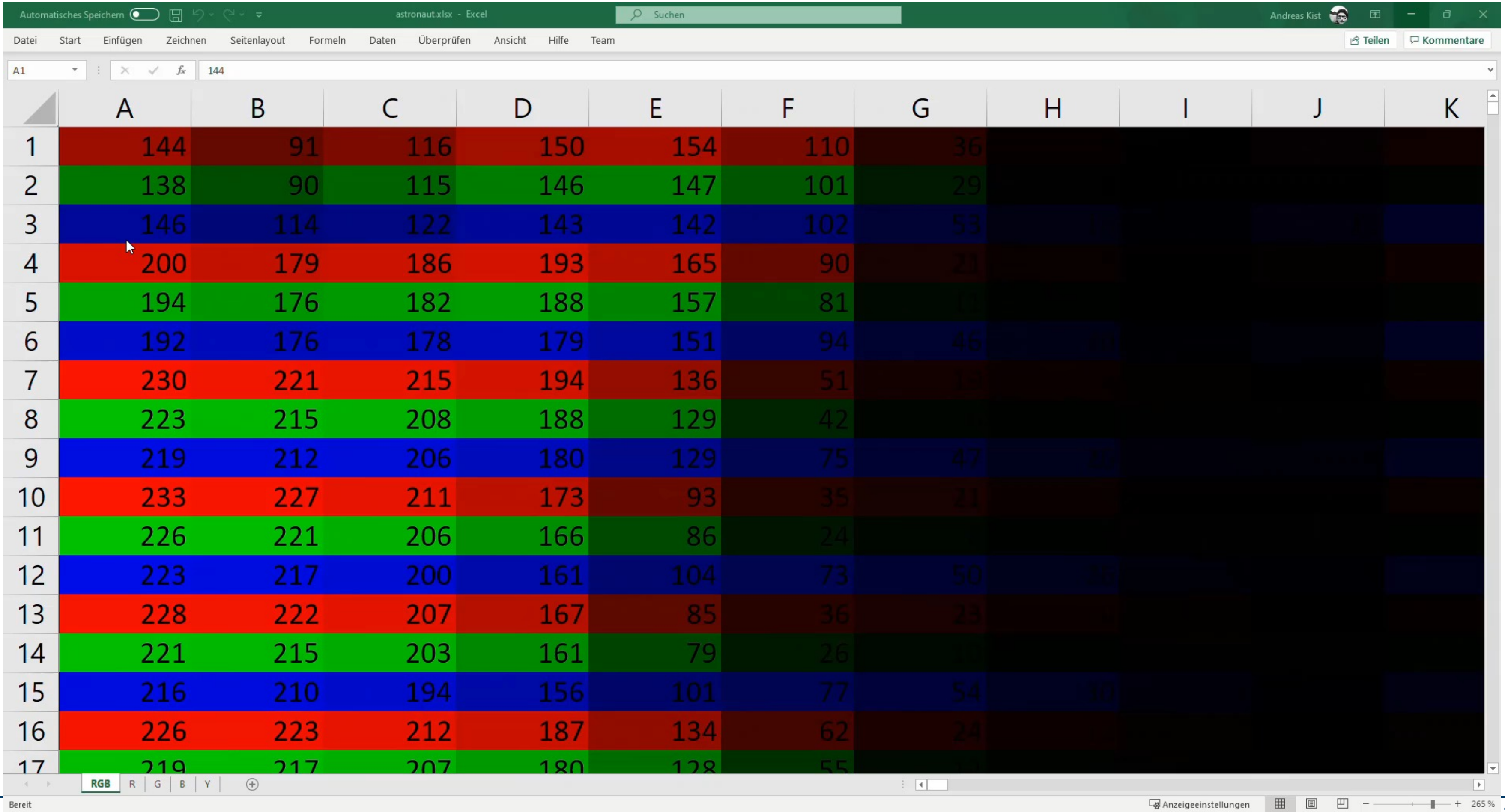


Demosaicing

Most frequent artifact: Moiré patterns

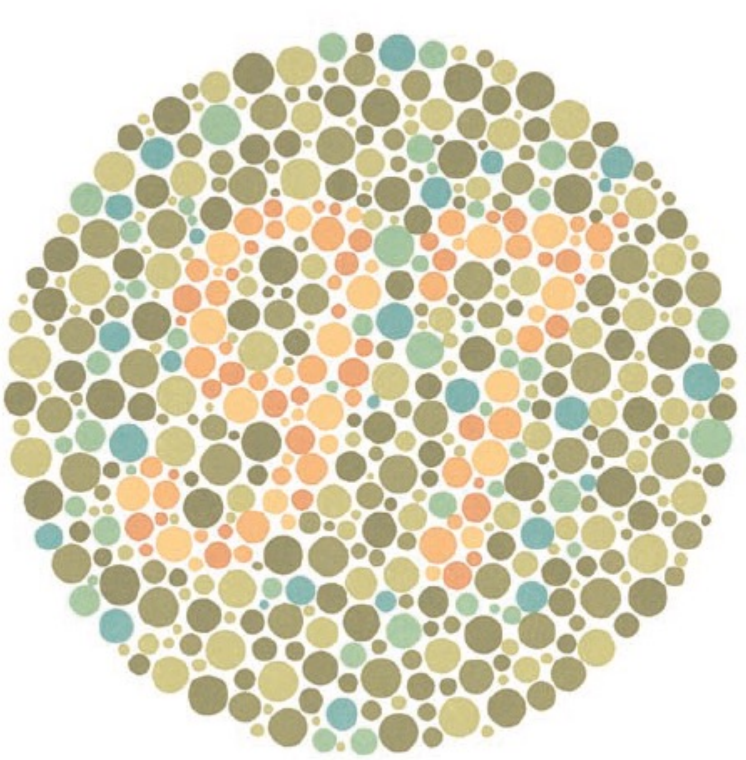
Images are Excel sheets

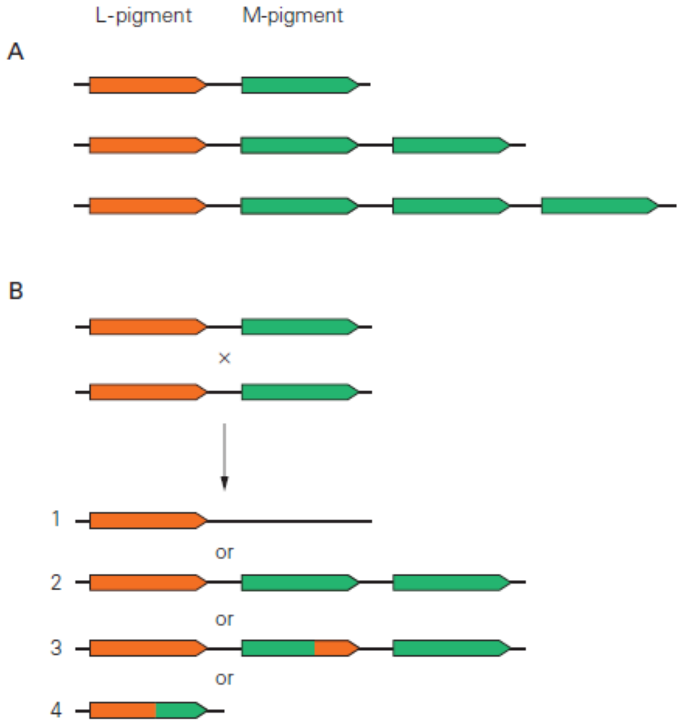
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF ENGINEERING



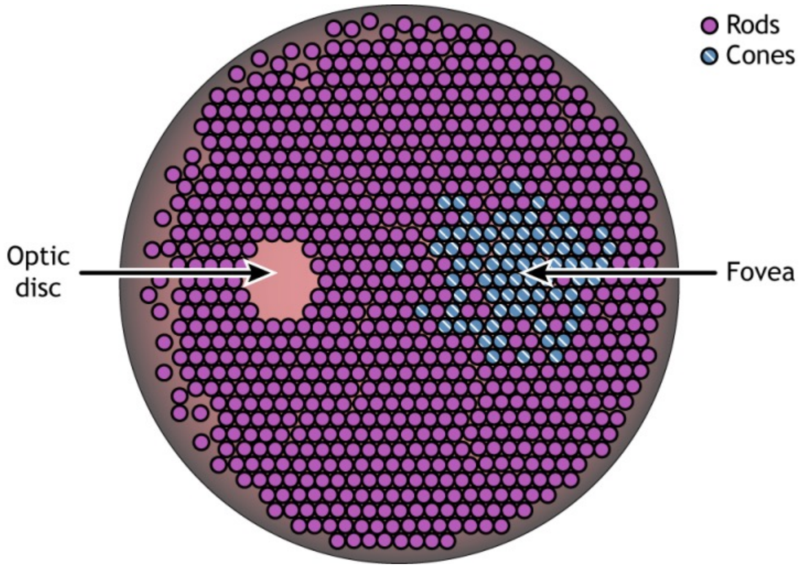
	A	B	C	D	E	F	G	H	I	J	K
1	144	91	116	150	154	110	36	[illegible]	[illegible]	[illegible]	[illegible]
2	138	90	115	146	147	101	29	[illegible]	[illegible]	[illegible]	[illegible]
3	146	114	122	143	142	102	53	[illegible]	[illegible]	[illegible]	[illegible]
4	200	179	186	193	165	90	21	[illegible]	[illegible]	[illegible]	[illegible]
5	194	176	182	188	157	81	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6	192	176	178	179	151	94	46	[illegible]	[illegible]	[illegible]	[illegible]
7	230	221	215	194	136	51	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
8	223	215	208	188	129	42	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9	219	212	206	180	129	75	47	[illegible]	[illegible]	[illegible]	[illegible]
10	233	227	211	173	93	35	21	[illegible]	[illegible]	[illegible]	[illegible]
11	226	221	206	166	86	24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
12	223	217	200	161	104	73	50	[illegible]	[illegible]	[illegible]	[illegible]
13	228	222	207	167	85	36	23	[illegible]	[illegible]	[illegible]	[illegible]
14	221	215	203	161	79	26	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
15	216	210	194	156	101	77	54	[illegible]	[illegible]	[illegible]	[illegible]
16	226	223	212	187	134	62	24	[illegible]	[illegible]	[illegible]	[illegible]
17	219	217	207	180	128	55	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

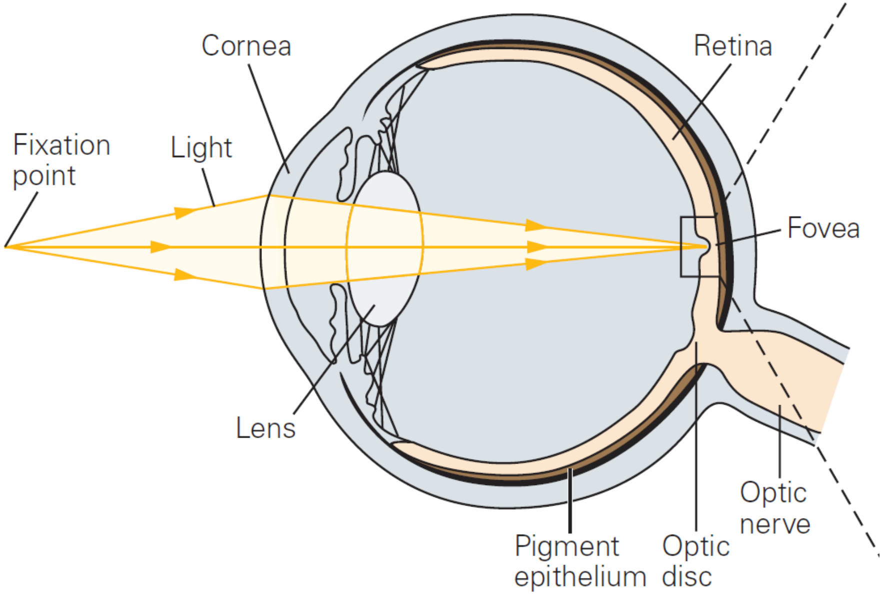
Color weak- and blindness



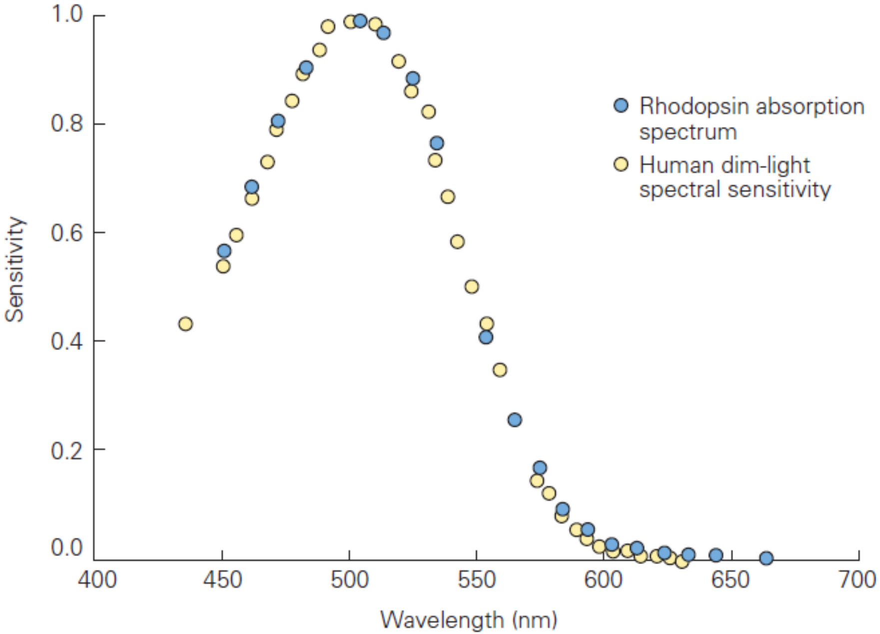


Our communication channel: the optic nerve

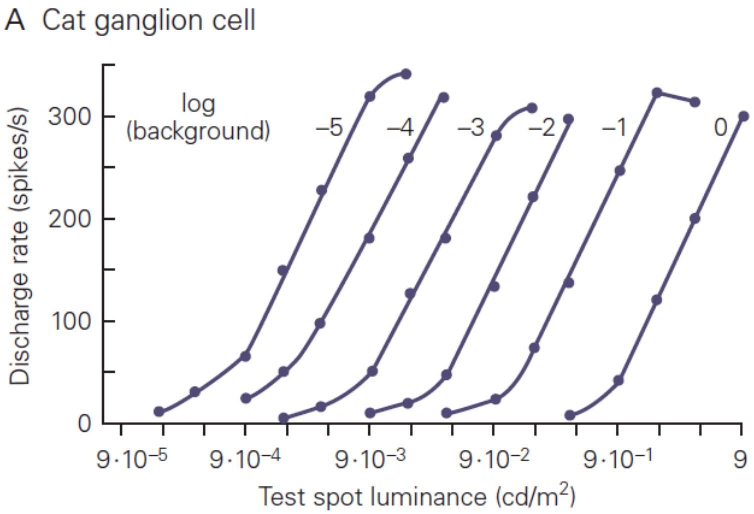


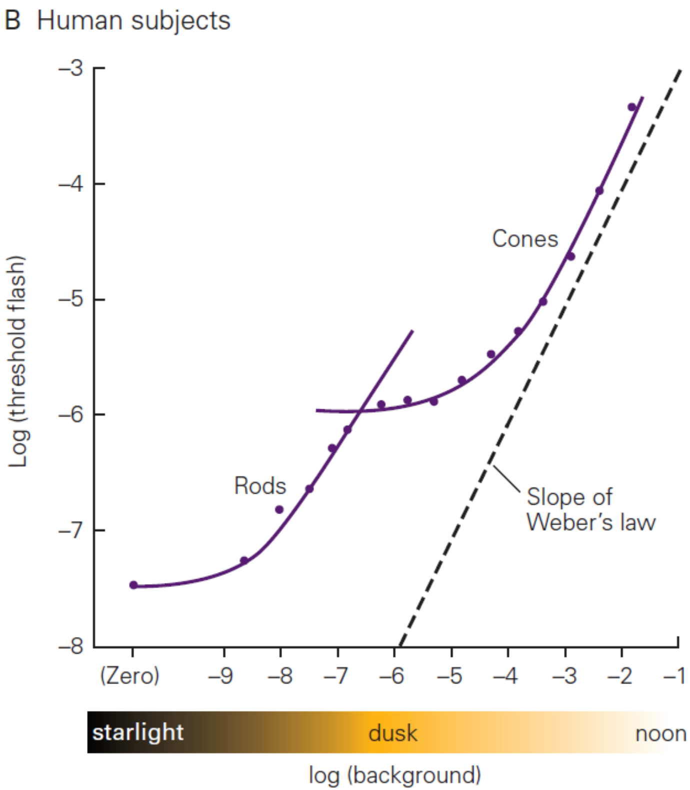


Our night vision

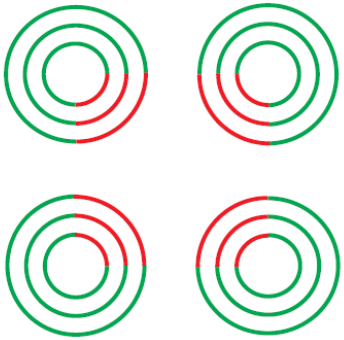


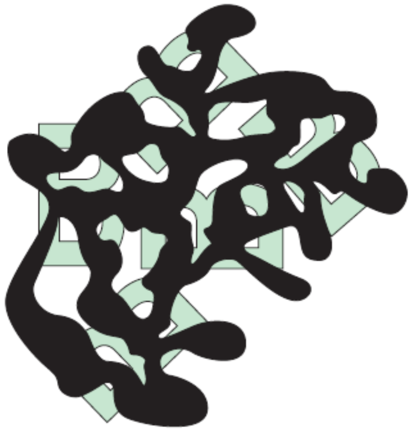
Light adaptation



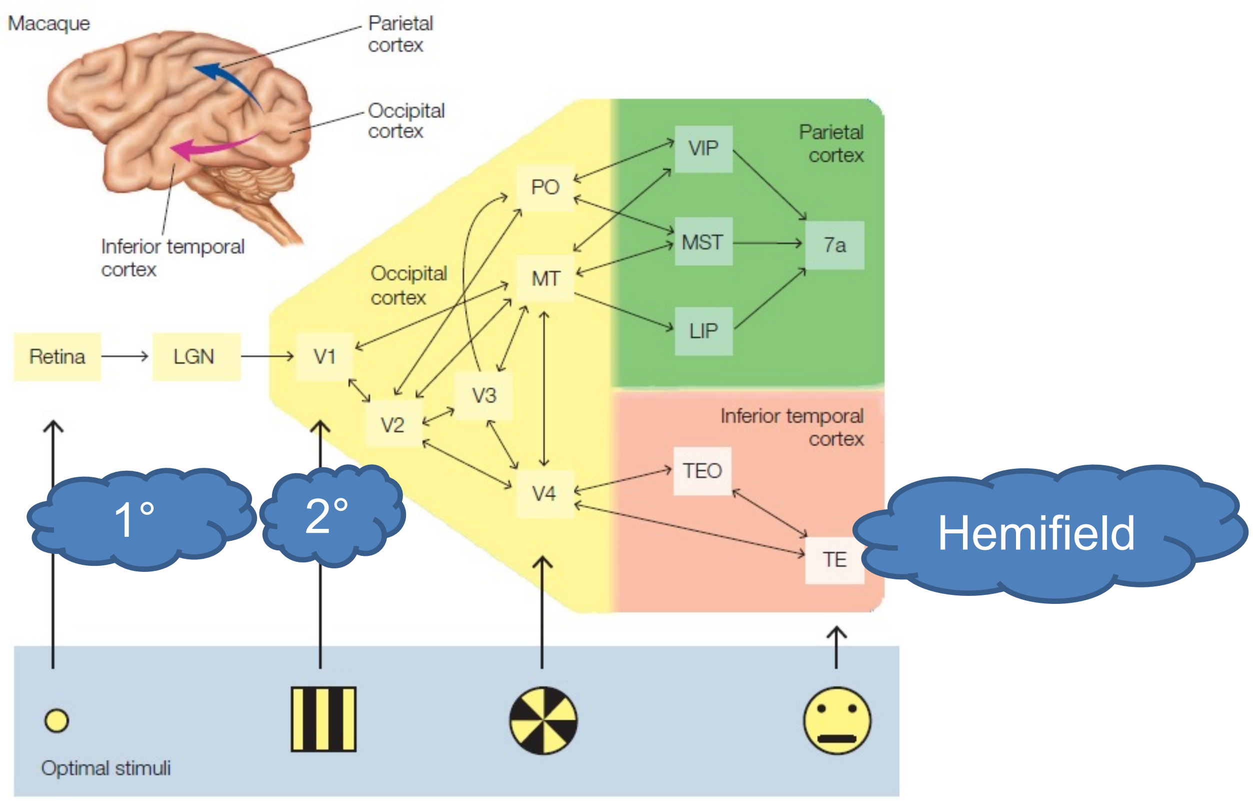


Visual perception

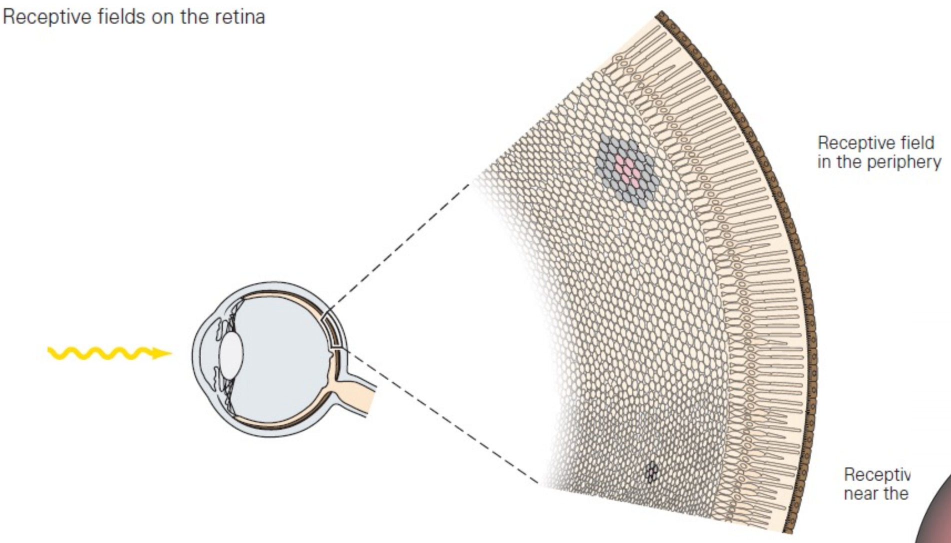


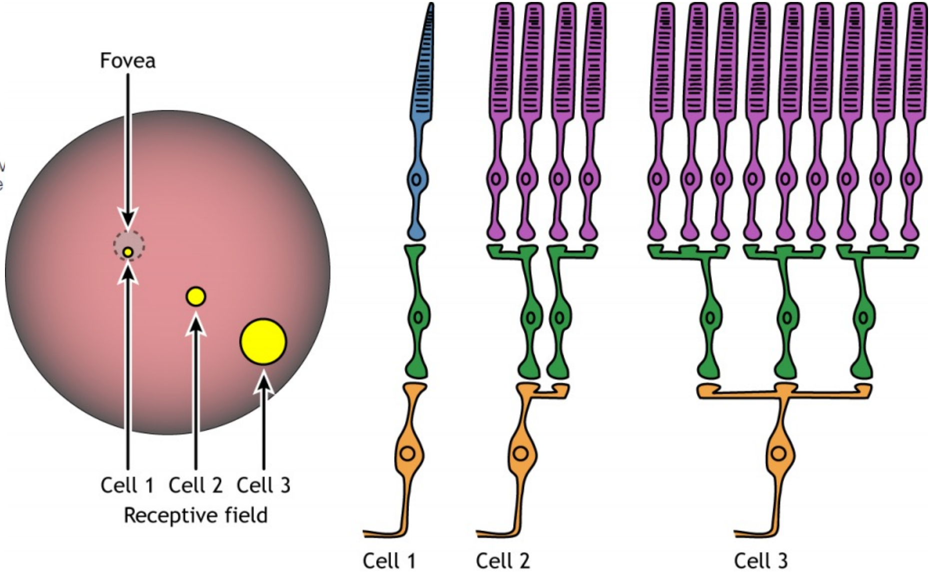


Layers of abstraction

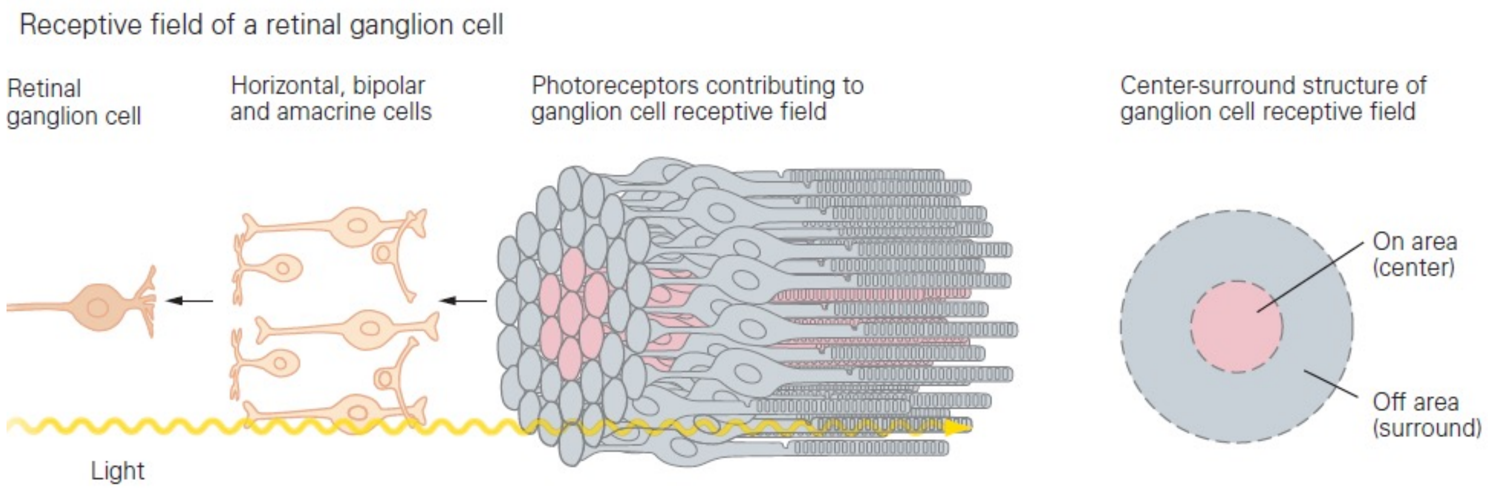


Generating a receptive field

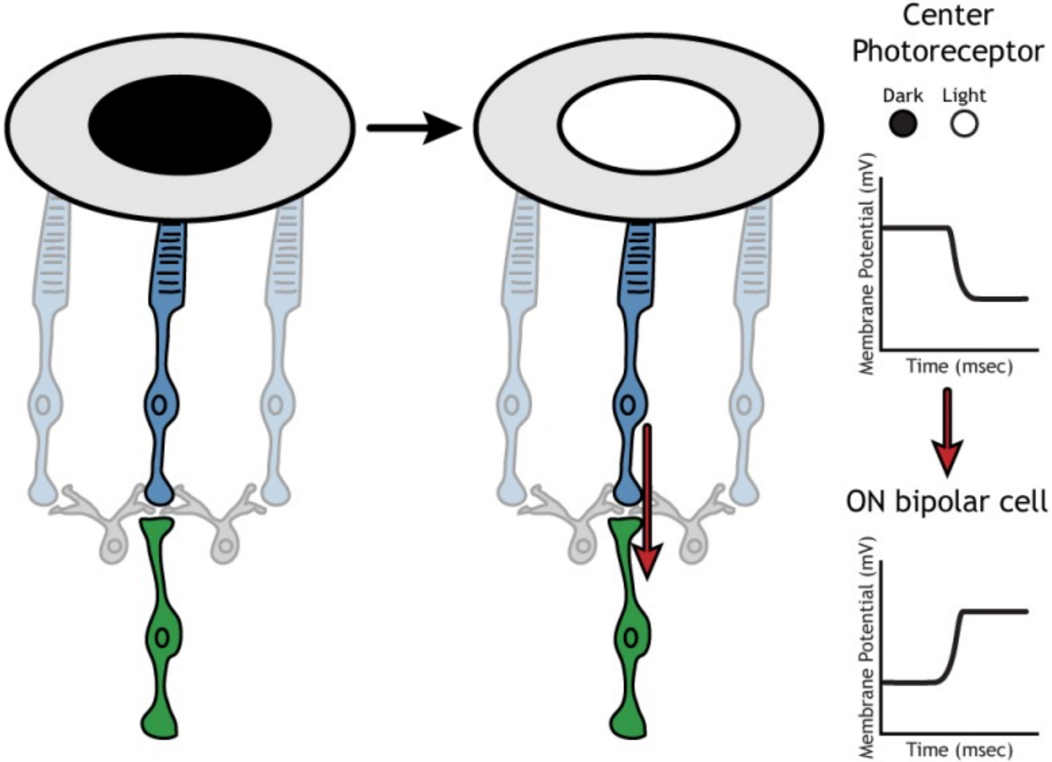




Generating a receptive field



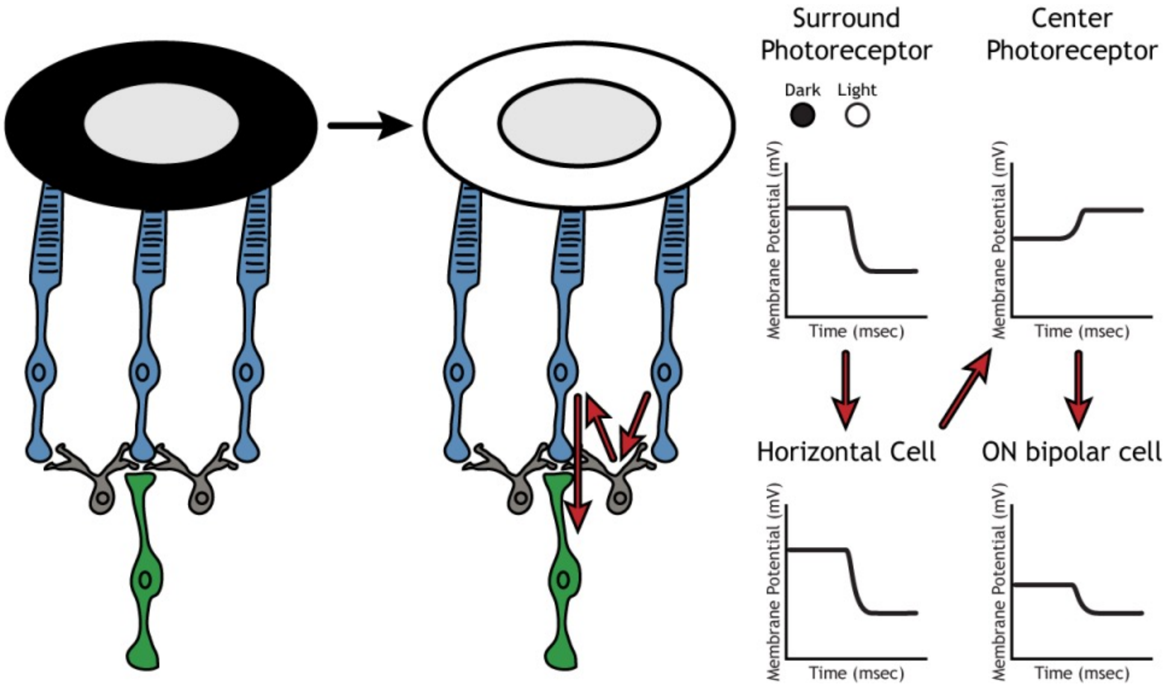
Examples for receptive field



The ON bipolar cell depolarizes upon hyperpolarization of the center photoreceptor

https://openbooks.lib.msu.edu/neuroscience/chapter/vision-the-retina/

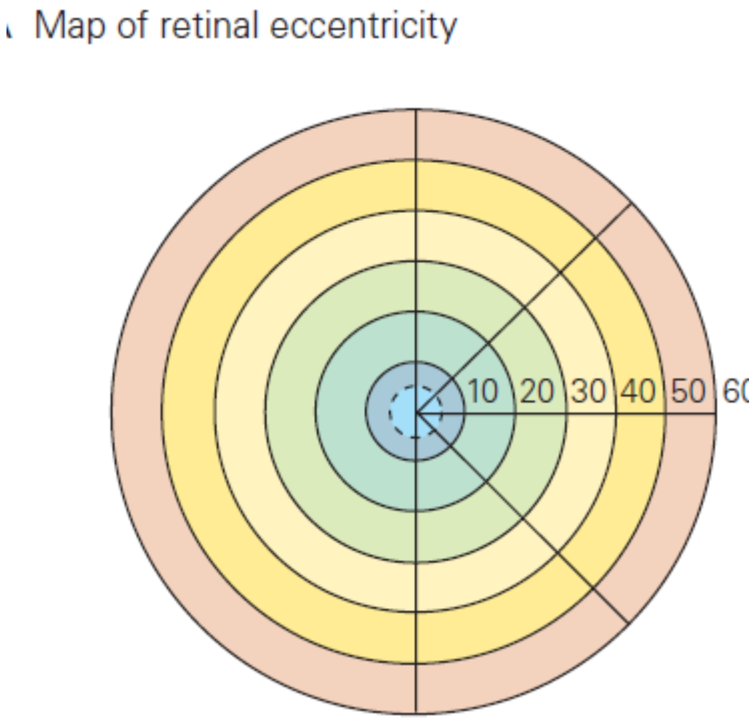
Examples for receptive fields

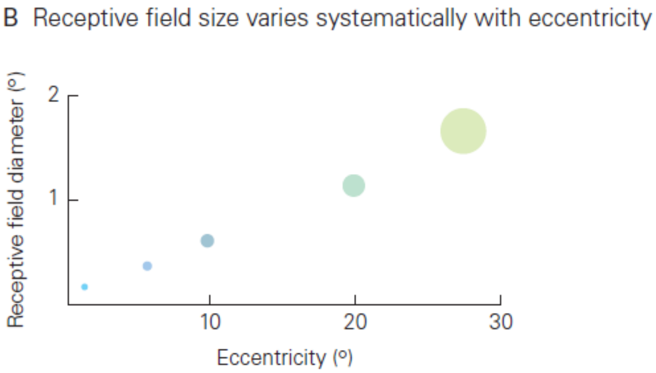


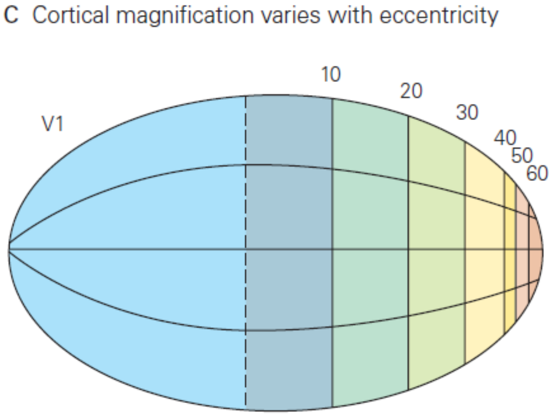
(i) the surrounding photoreceptor will hyperpolarize
(ii) The postsynaptic horizontal cell will hyperpolarize
(iii) The center photoreceptor depolarizes
(iv) The ON bipolar cell to hyperpolarize.

➔ Same as if the center photoreceptor experiences dark

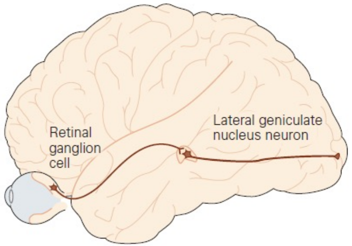
Receptive field changes with eccentricity

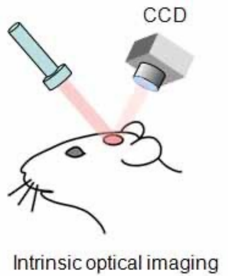






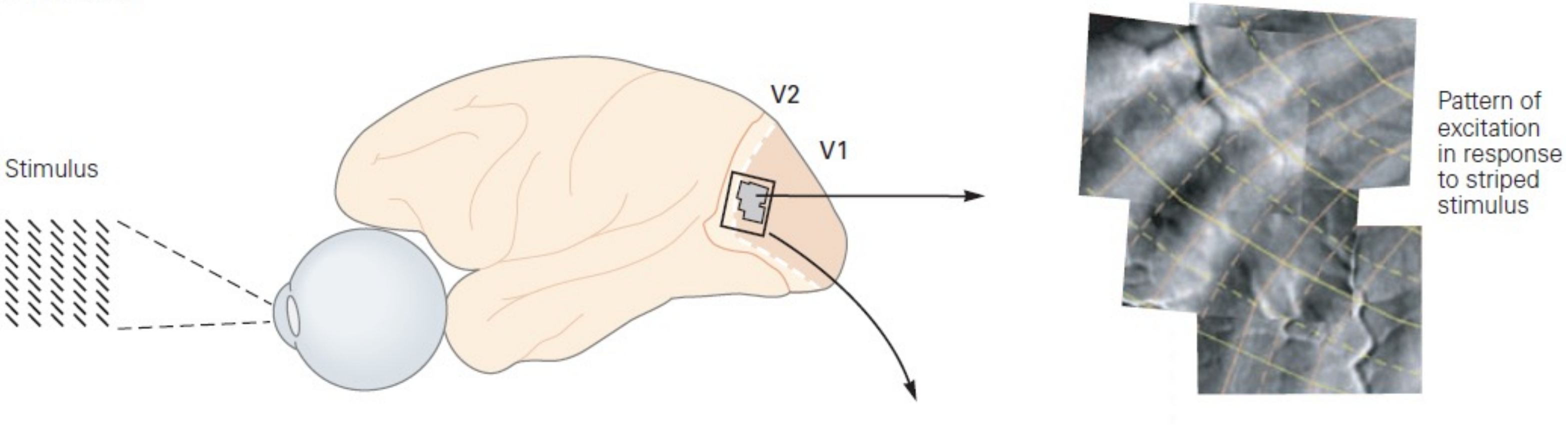
Retina produces visuotopic maps in V1



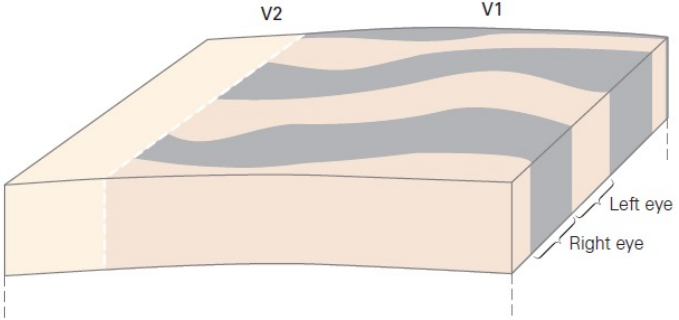


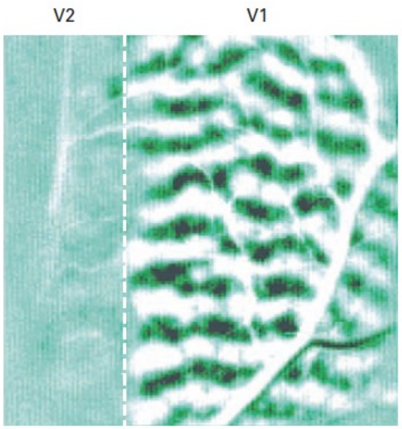
© Uniklinik Jena

Visuotopic map

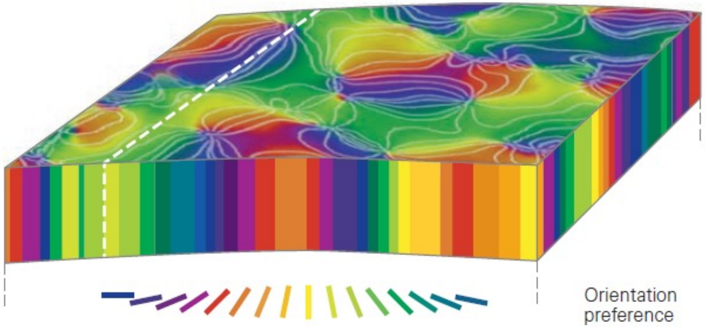


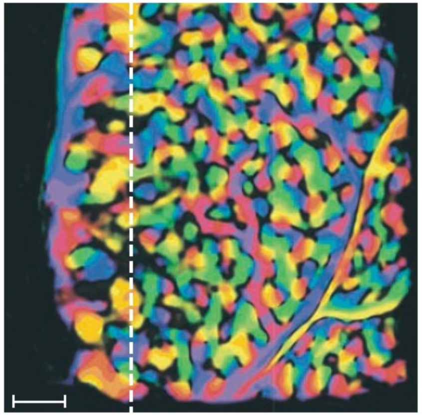
B Ocular dominance columns

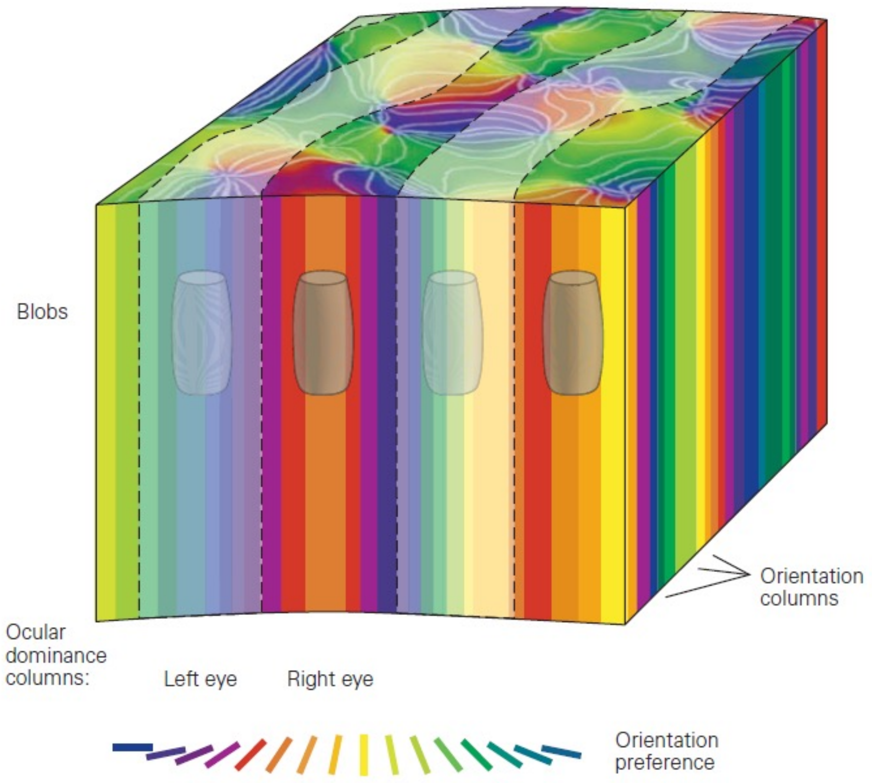




C Orientation columns

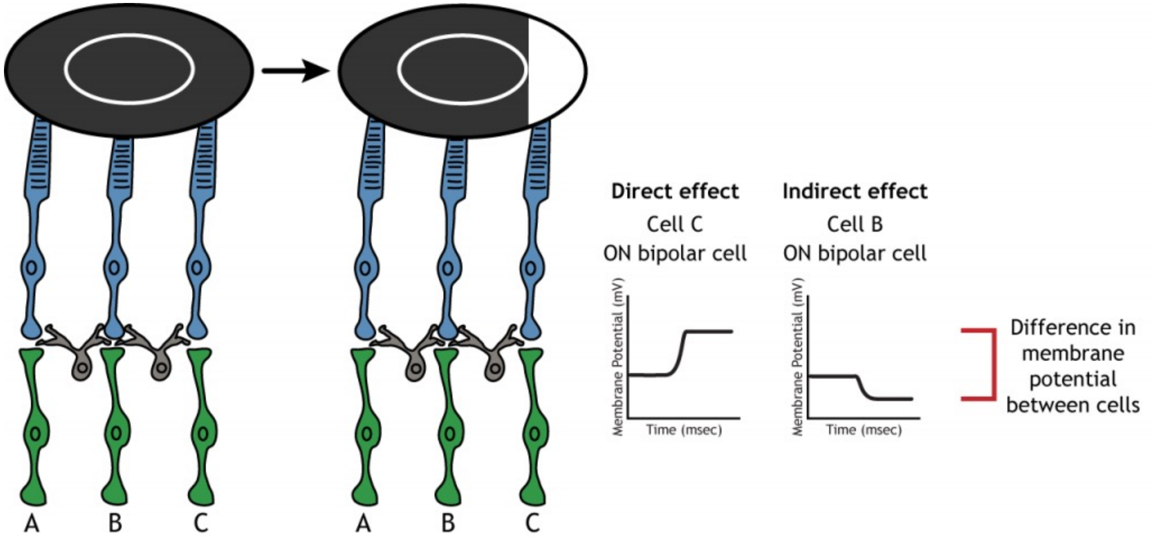






Blobs
Orientation columns
Ocular dominance columns:
Left eye
Right eye
Orientation preference

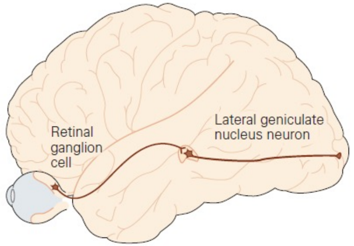
Edge detection using lateral inhibition



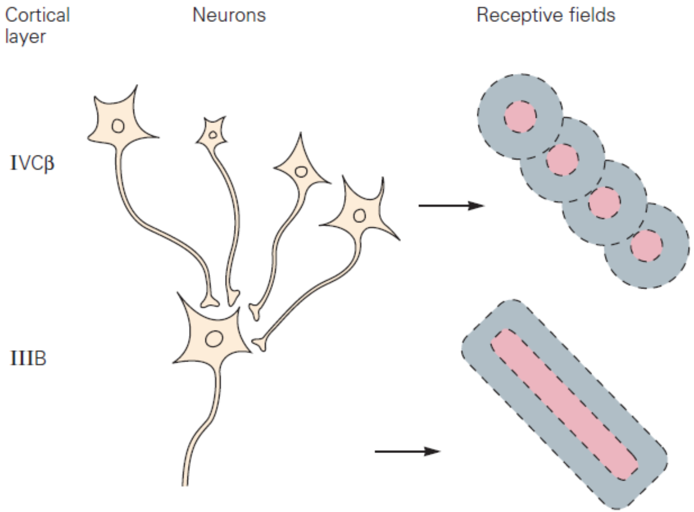
(i) The light will cause bipolar cell C to depolarize because of the direct synapse with the photoreceptor. (ii) The light will also cause bipolar cell B to hyperpolarize because of the indirect synapses through the horizontal cell. ➔ This hyperpolarization causes **a larger membrane potential difference** between cells B and C that would occur if the horizontal cells were absent. ➔ The larger membrane potential difference between the cells will enhance the perception between the dark and light side of the edge.

https://openbooks.lib.msu.edu/neuroscience/chapter/vision-the-retina/

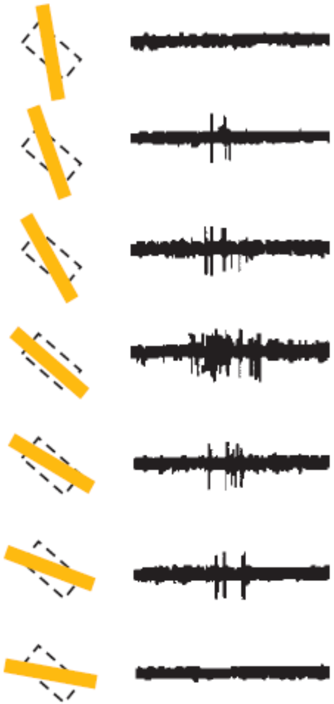
Direction selectivity



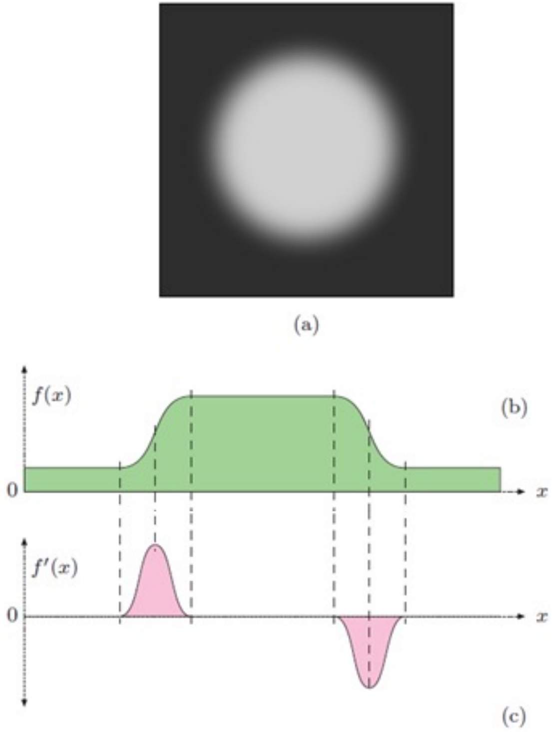
LGN

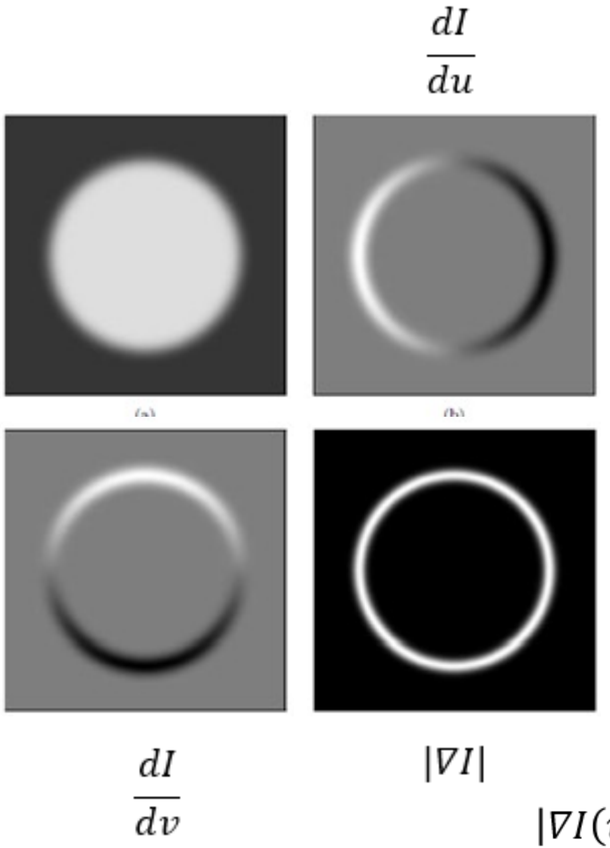


V1



Edge detection





$$|\nabla I(u,v)| = \sqrt{(\frac{\partial I}{\partial u})^2 + (\frac{\partial I}{\partial v})^2}$$

Linear filter

Just imagine a noisy pixel...

	12	11	11
v	13	255	10
	14	15	10
		u	

$$\xrightarrow{\frac{12+11+11+13+255+10+14+15+10}{9} = 39}$$

12	11	11
13	39	10
14	15	10

$$I'(u,v) = \frac{\sum_{i=-1}^{1}\sum_{j=-1}^{1} I(u+i,v+j)}{9}$$

1	1	1
1	1	1
1	1	1

Applying linear filter over whole image

$$\otimes \begin{array}{|c|c|c|}\hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array} =$$

CONVOLUTION

Kernel size

Increase of receptive field

Edge detection with linear filters

Prewitt-Filter

$$H_x^P = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \; and \quad H_y^P = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

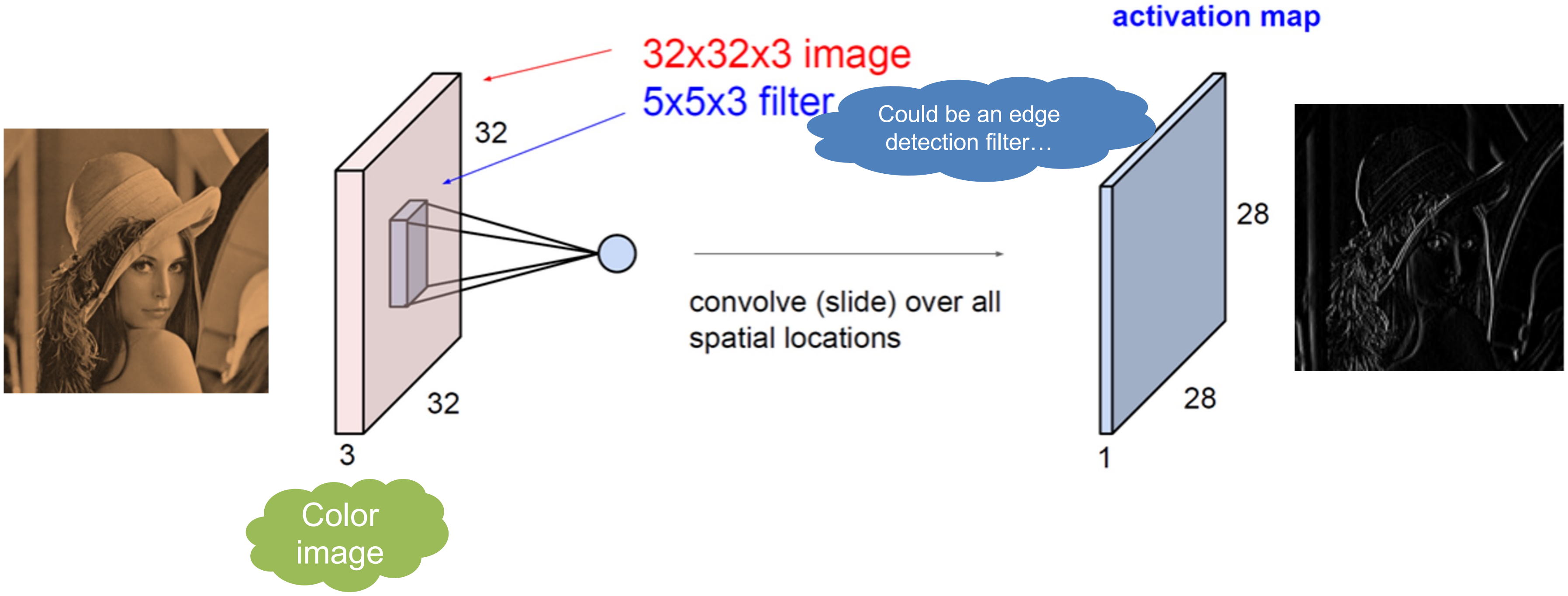
Sobel-Filter

$$H_x^S = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \; and \quad H_y^S = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$



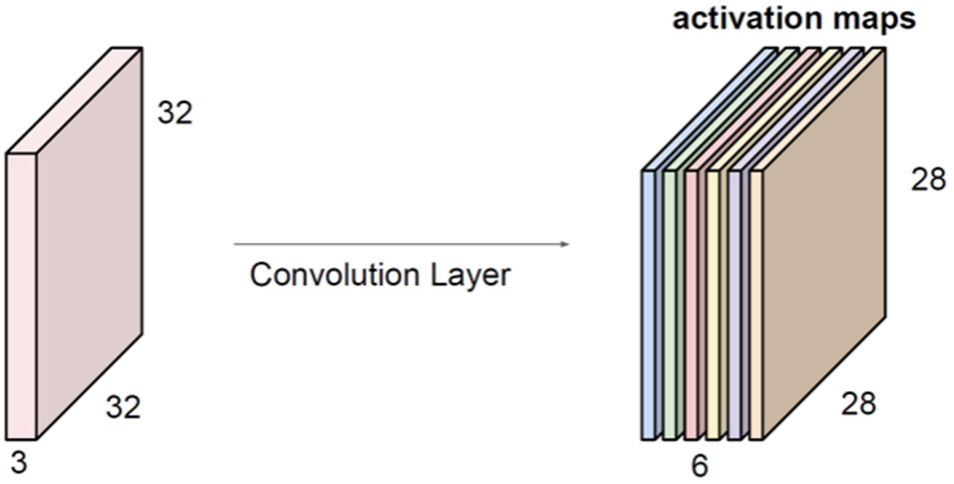


Convolutional neural networks



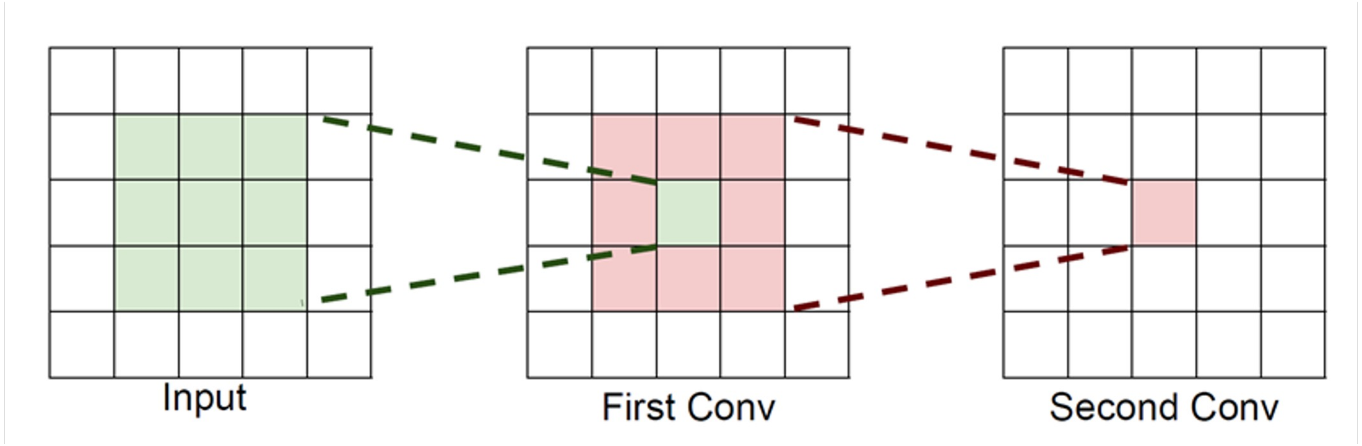
Many activation maps

For example, if we had 6 5x5 filters, we'll get 6 separate activation maps:

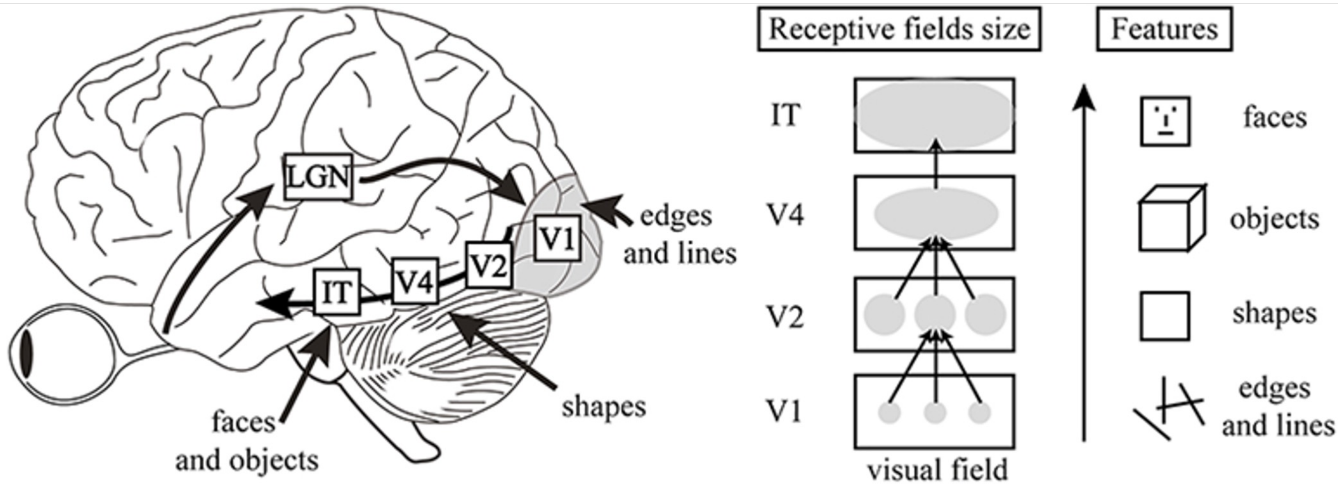


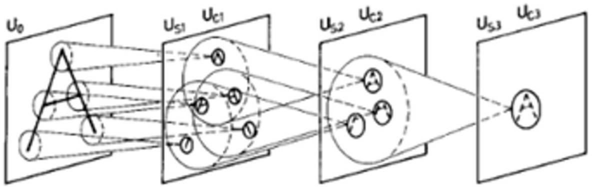
We stack these up to get a "new image" of size 28x28x6!

Multiple rounds of applying filters



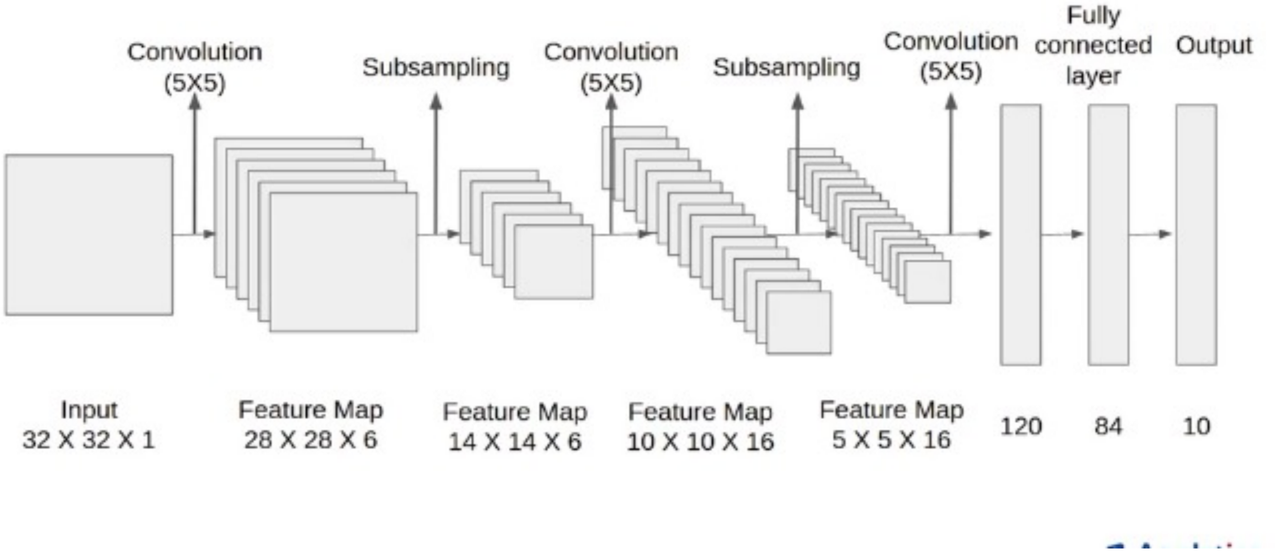
High-level features





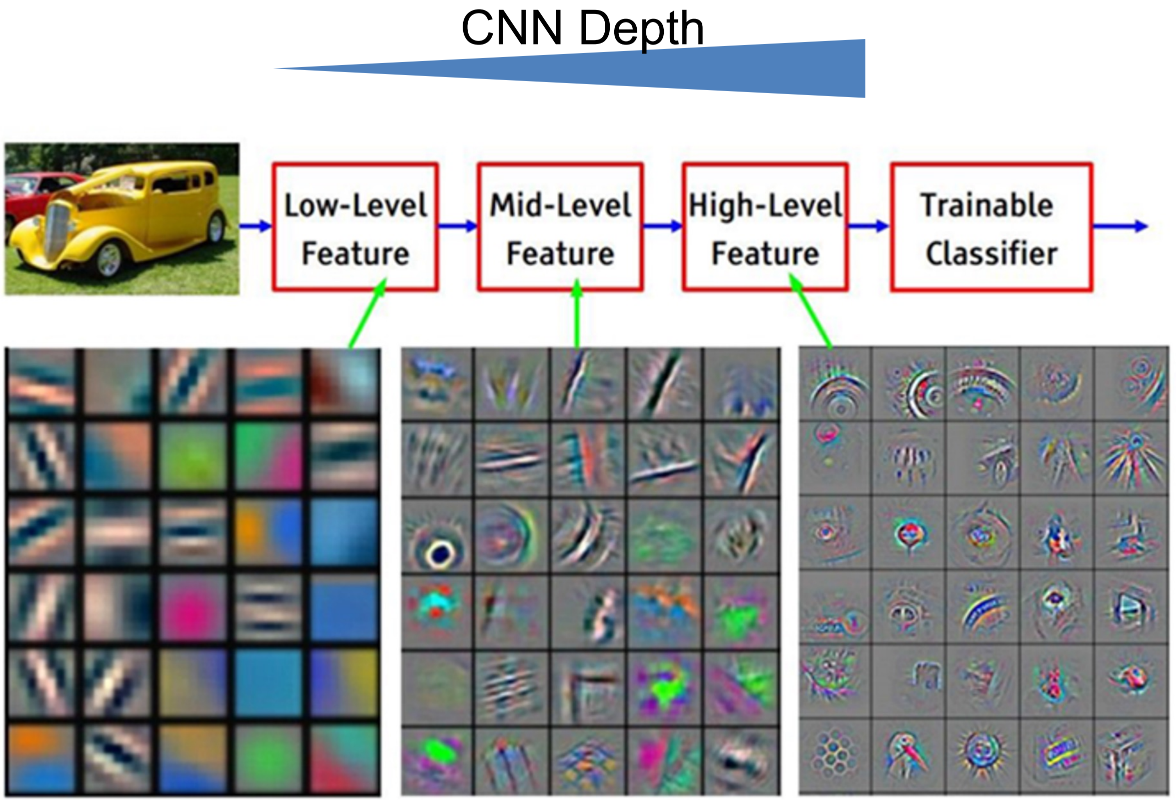
https://neurdiness.wordpress.com/2018/05/17/deep-convolutional-neural-networks-as-models-of-the-visual-system-qa/

LeNet-5



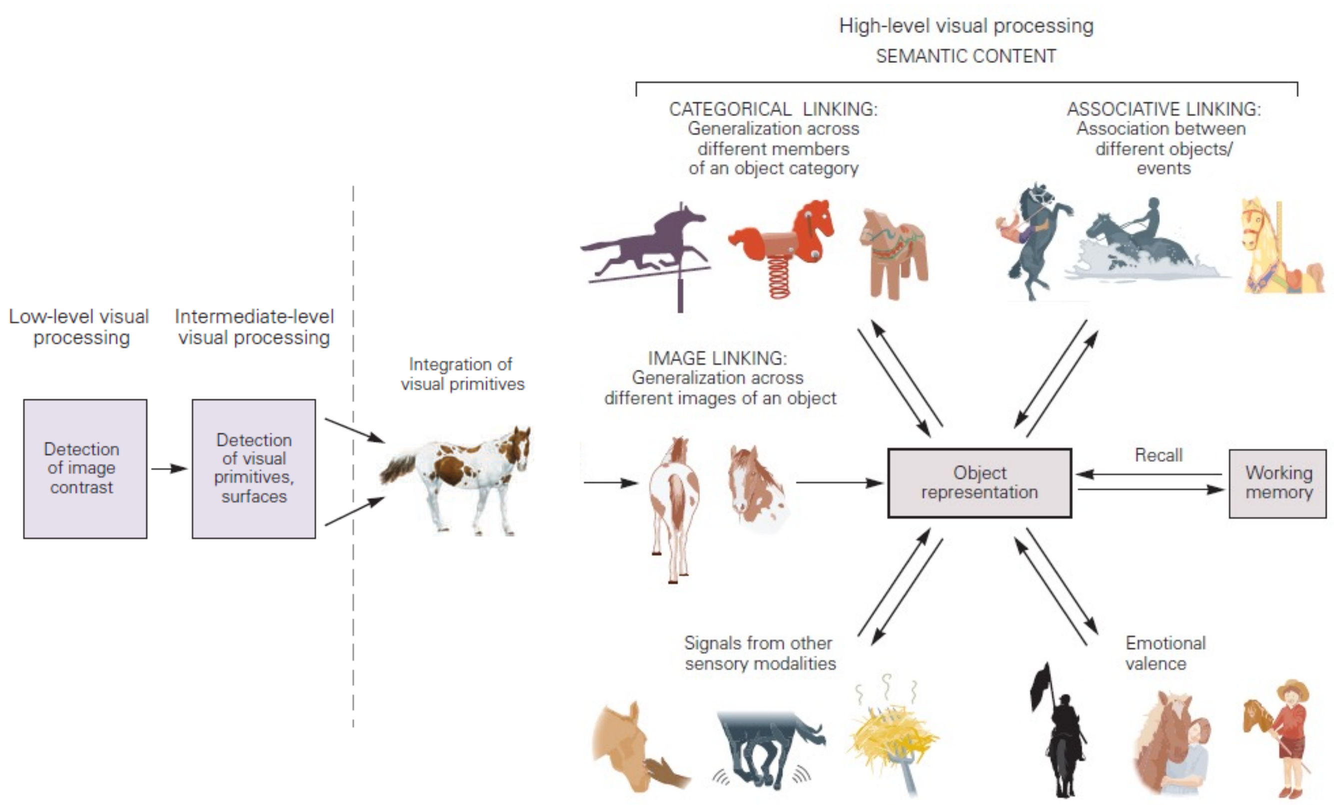
Convolution (5X5)
Subsampling
Convolution (5X5)
Subsampling
Convolution (5X5)
Fully connected layer
Output
Input 32 X 32 X 1
Feature Map 28 X 28 X 6
Feature Map 14 X 14 X 6
Feature Map 10 X 10 X 16
Feature Map 5 X 5 X 16
120
84
10

High level features in network depth

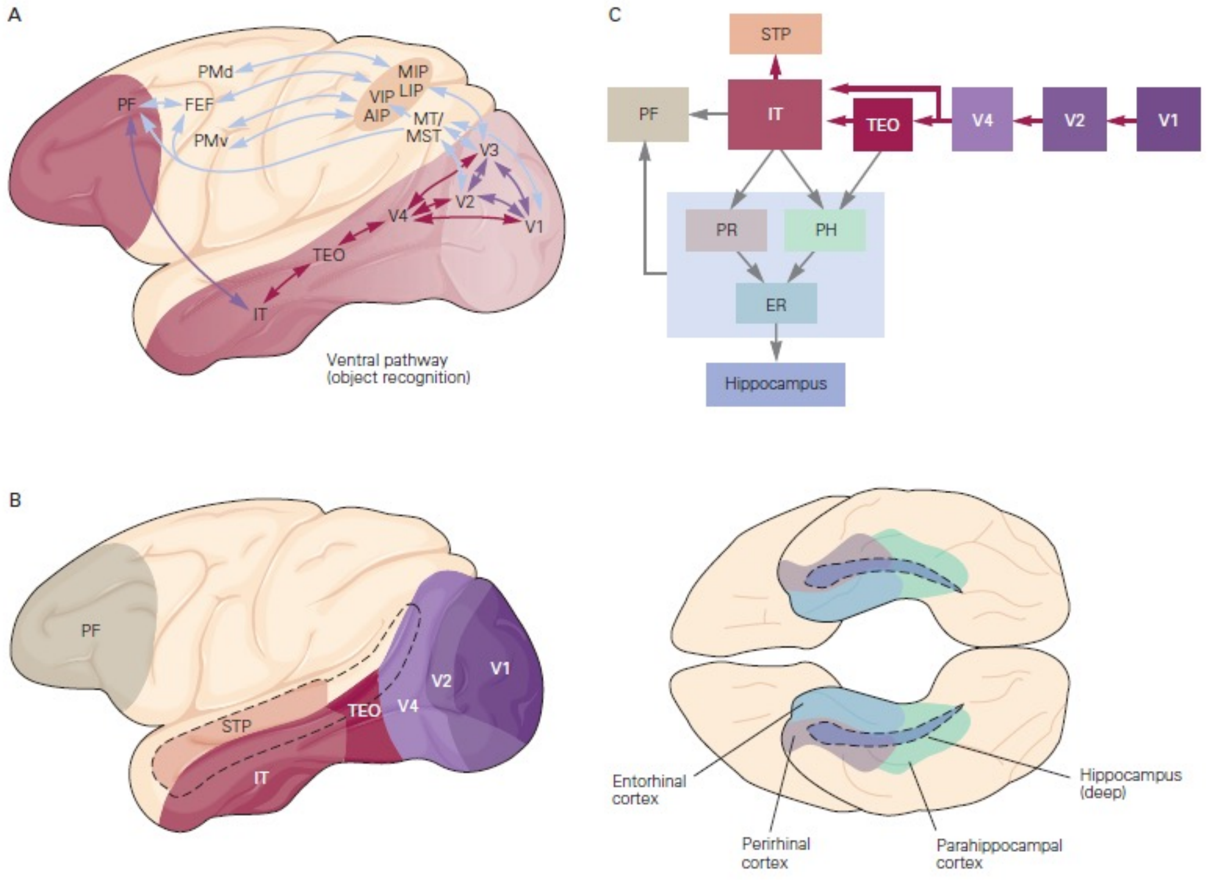


Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Object recognition



Object recognition pathway



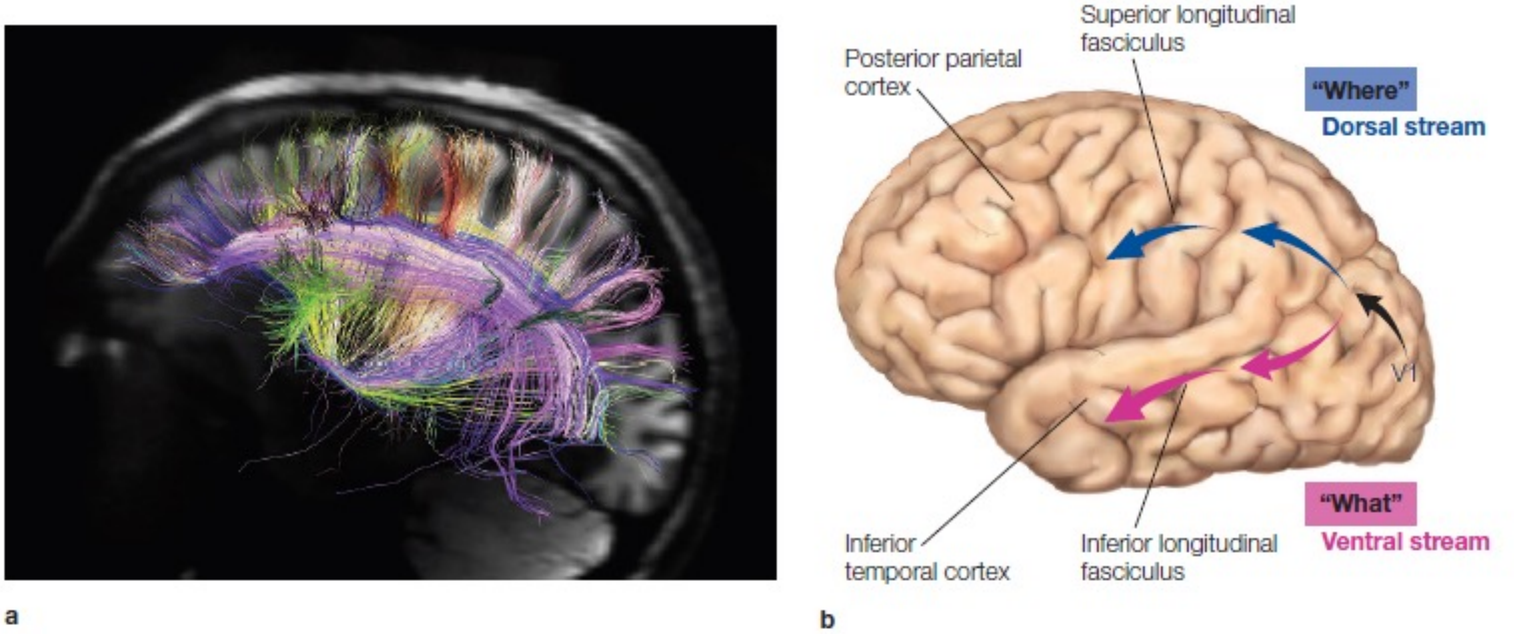
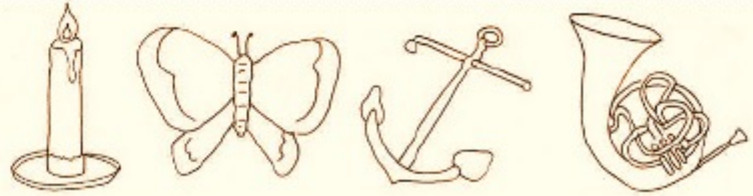
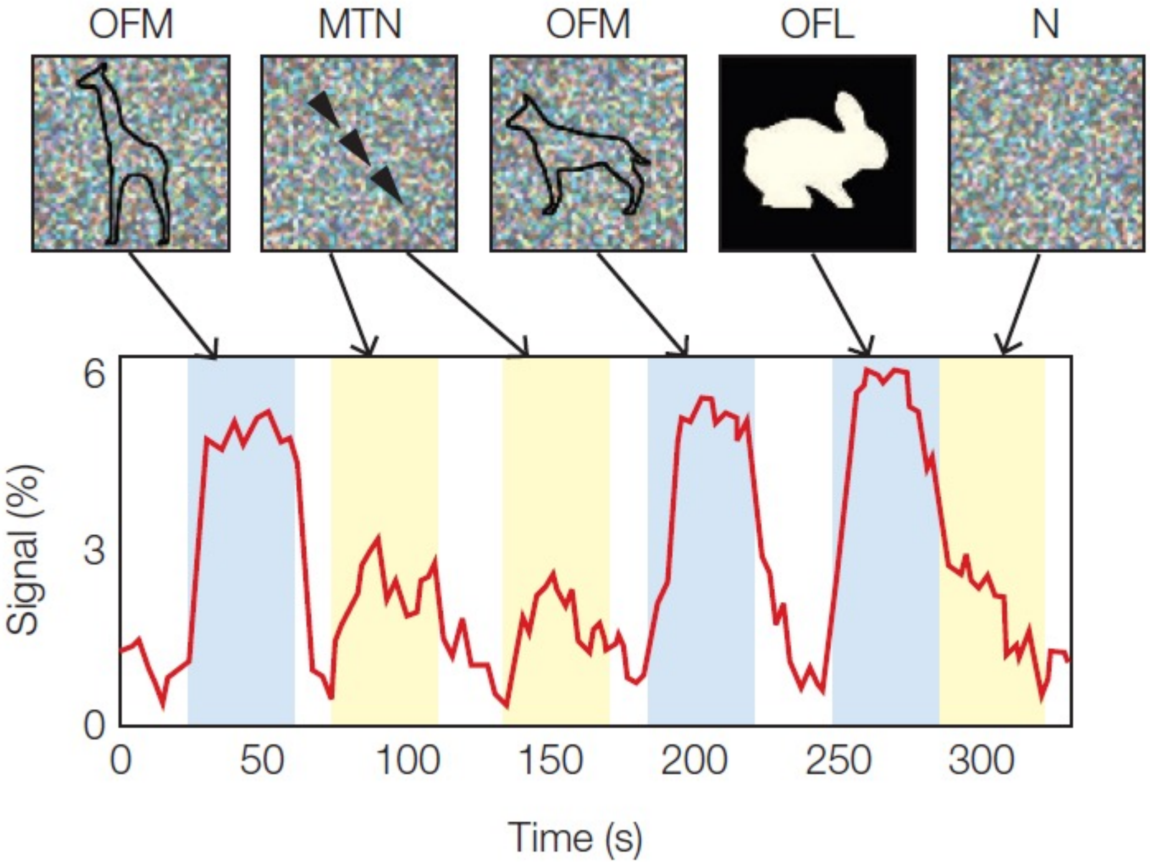


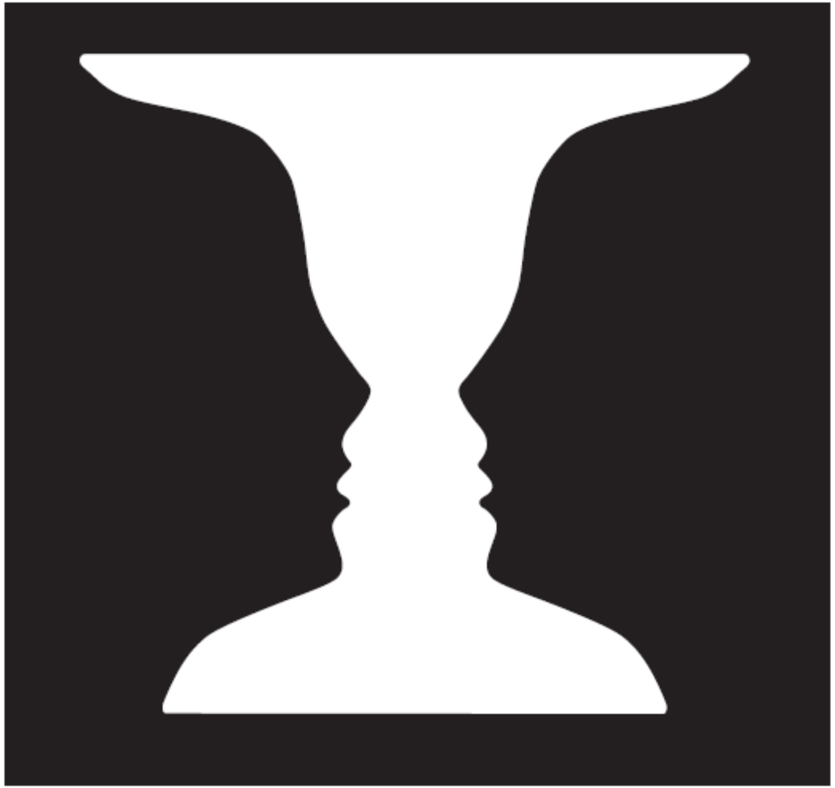
FIGURE 6.4 The major object recognition pathways.
(a) The longitudinal fasciculus, shown here in shades of purple. (b) The ventral "what" pathway terminates in the inferior temporal cortex, and the dorsal "where" pathway terminates in the posterior parietal cortex.

	Sample stimuli	Feature extraction	Shape description	Memory matching
Familiar		✓	✓	✓
Novel		✓	✓	
Scrambled		✓		

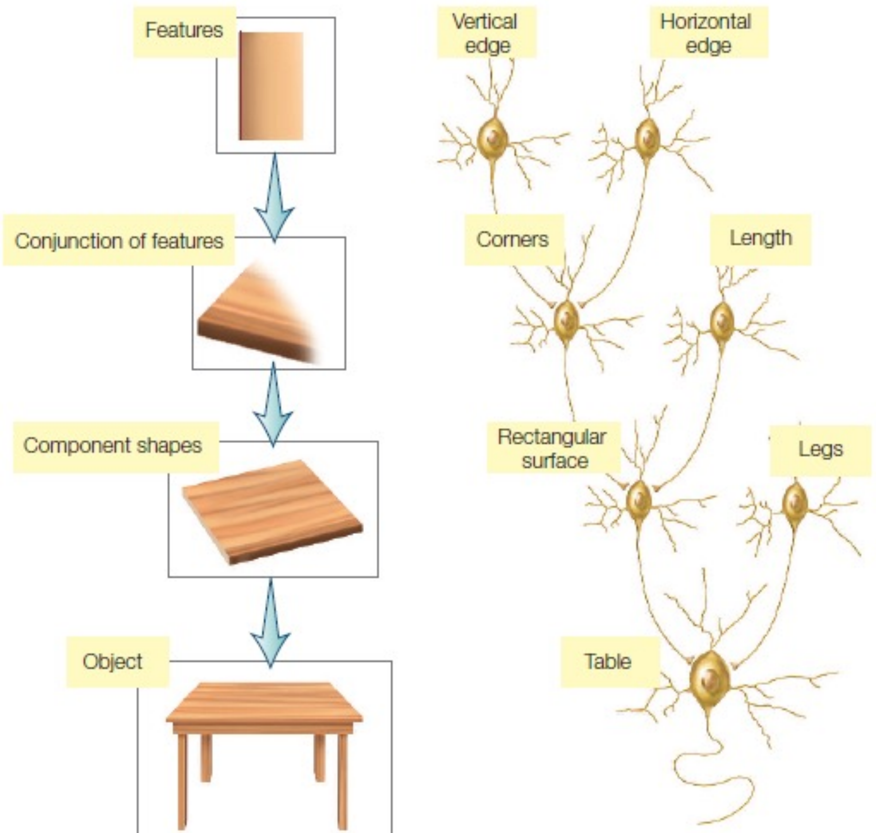
The lateral



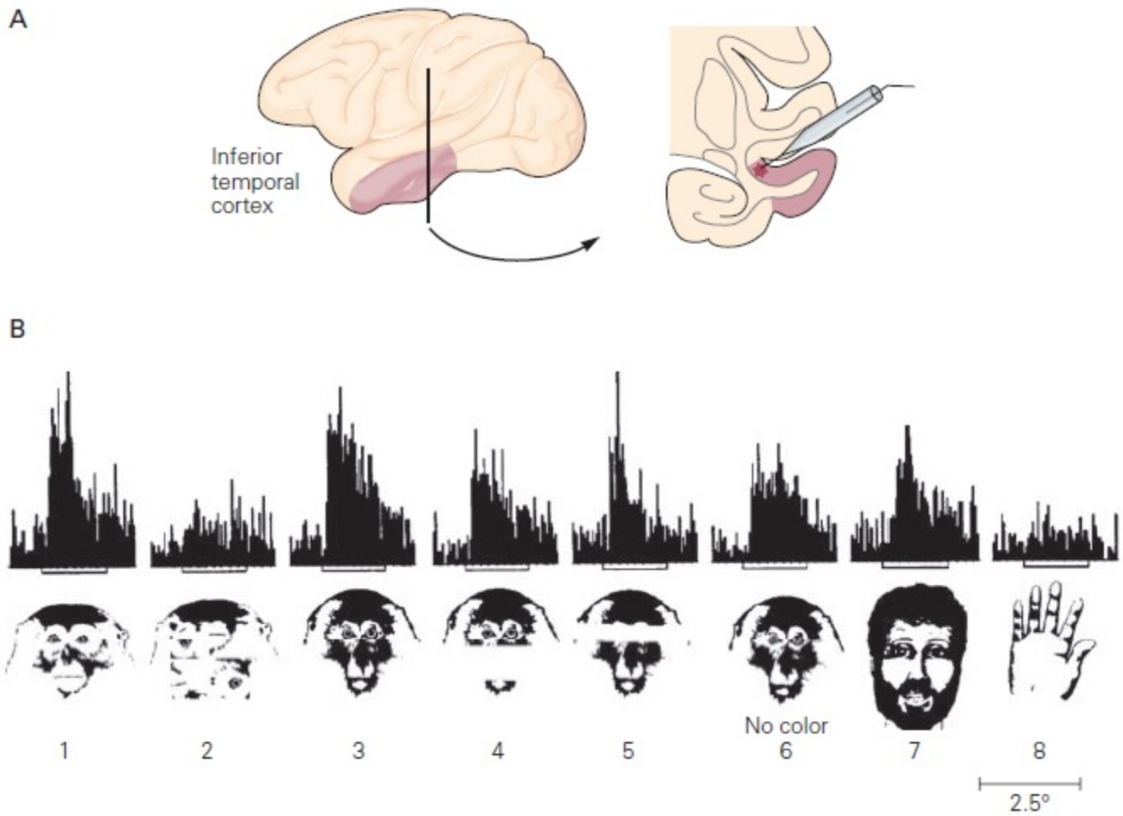
Multistable perceptions



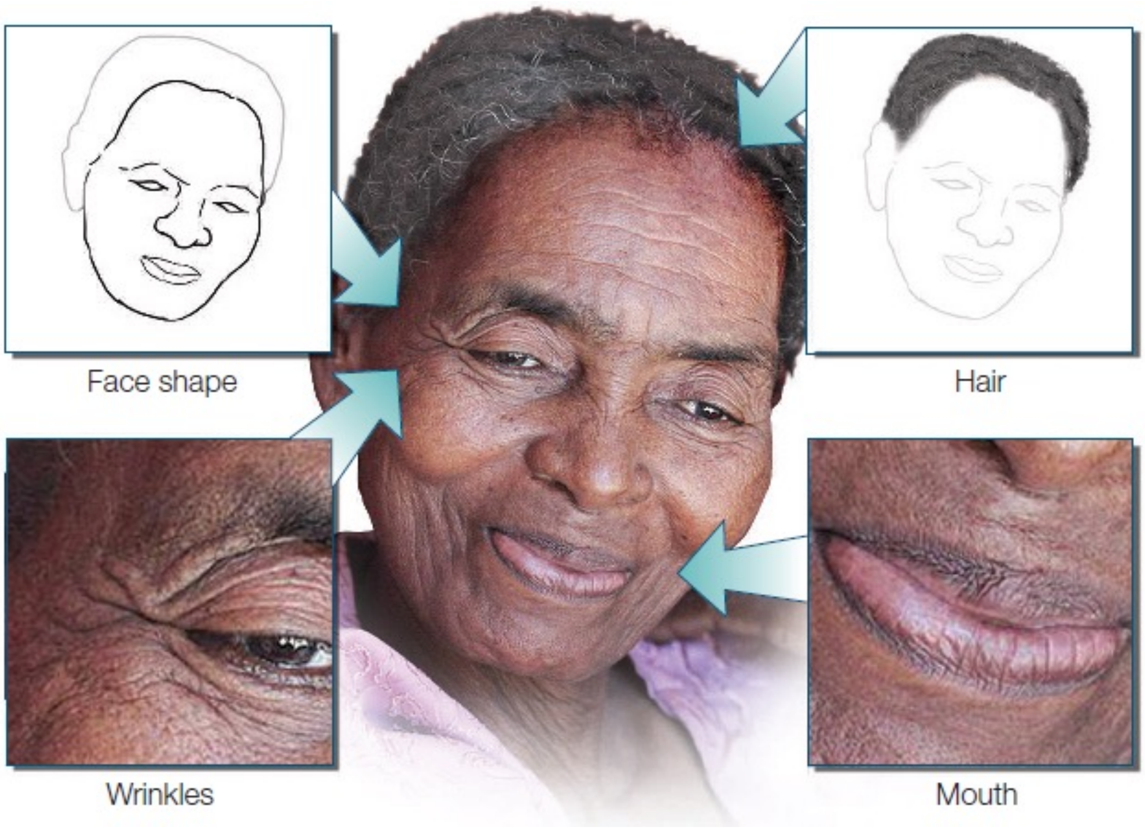
Grandmother Cells and Ensemble Coding


Features
Conjunction of features
Component shapes
Object
Vertical edge
Horizontal edge
Corners
Length
Rectangular surface
Legs
Table

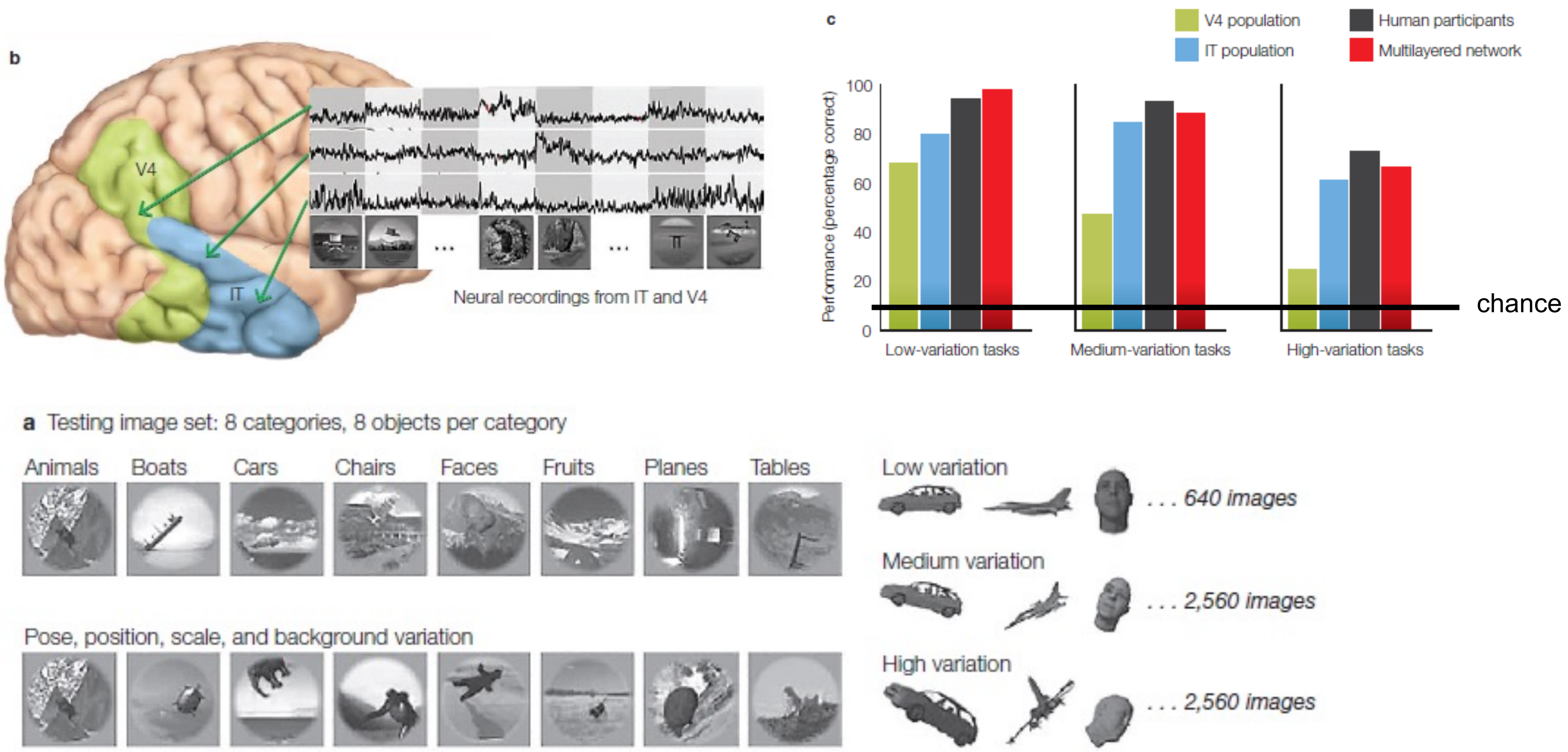
IT is our face detector



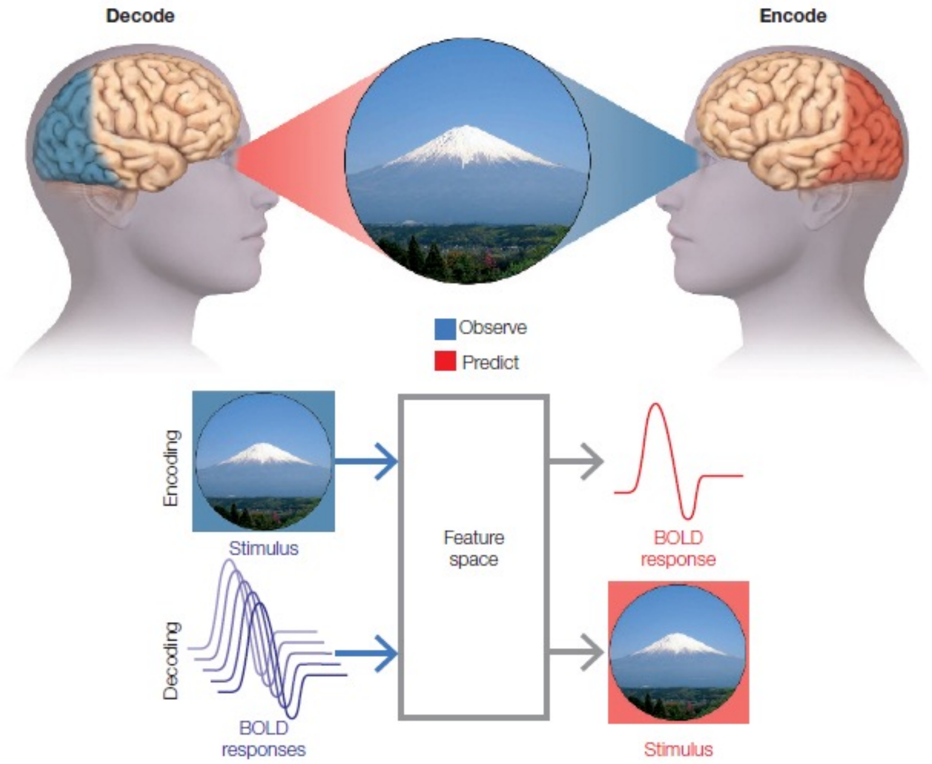
Ensemble Coding Hypothesis



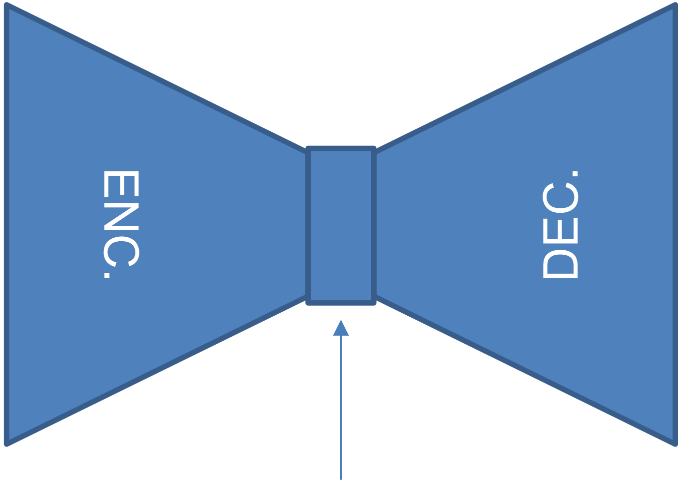
IT / V4 activity



Encoder-Decoder

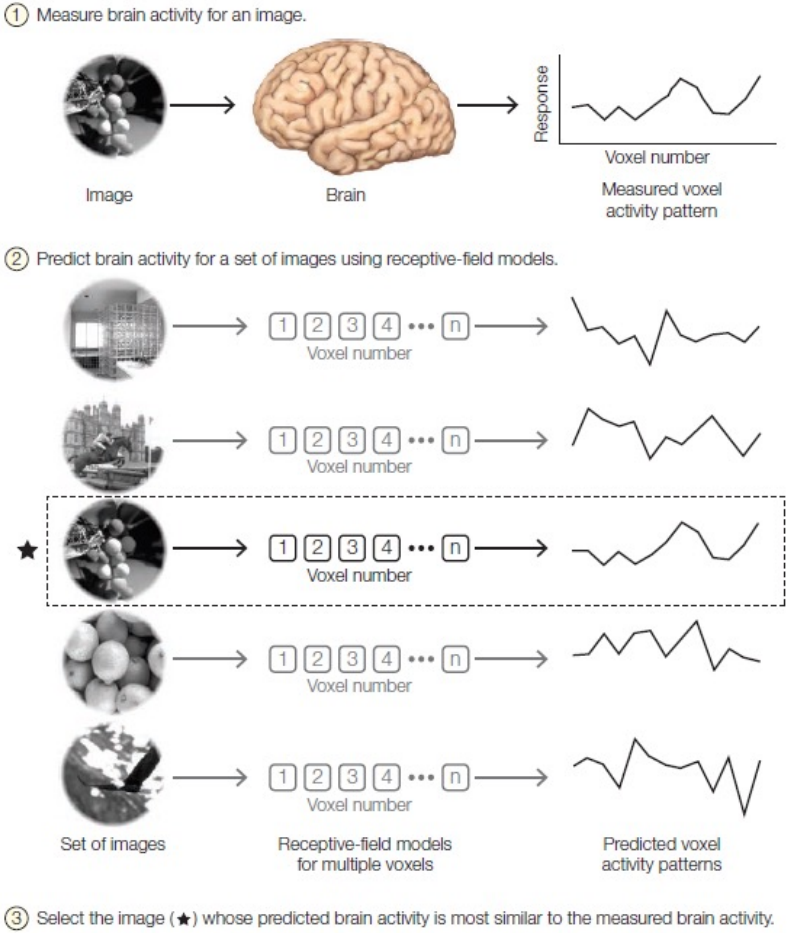


Encoder-Decoder-DNNs

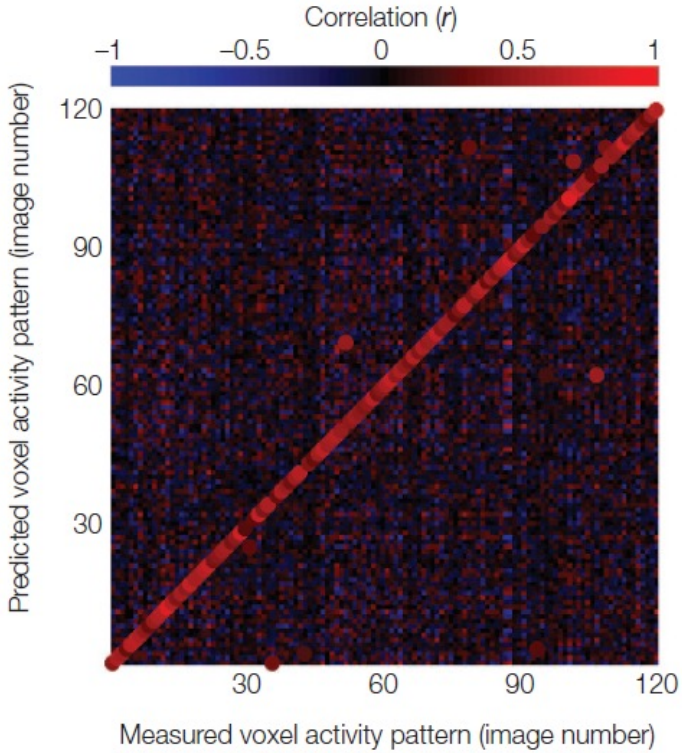


Latent/feature space

Decoding brain activity to natural images



Correlation coefficients



Face recognition by single neurons

Published: 23 June 2005

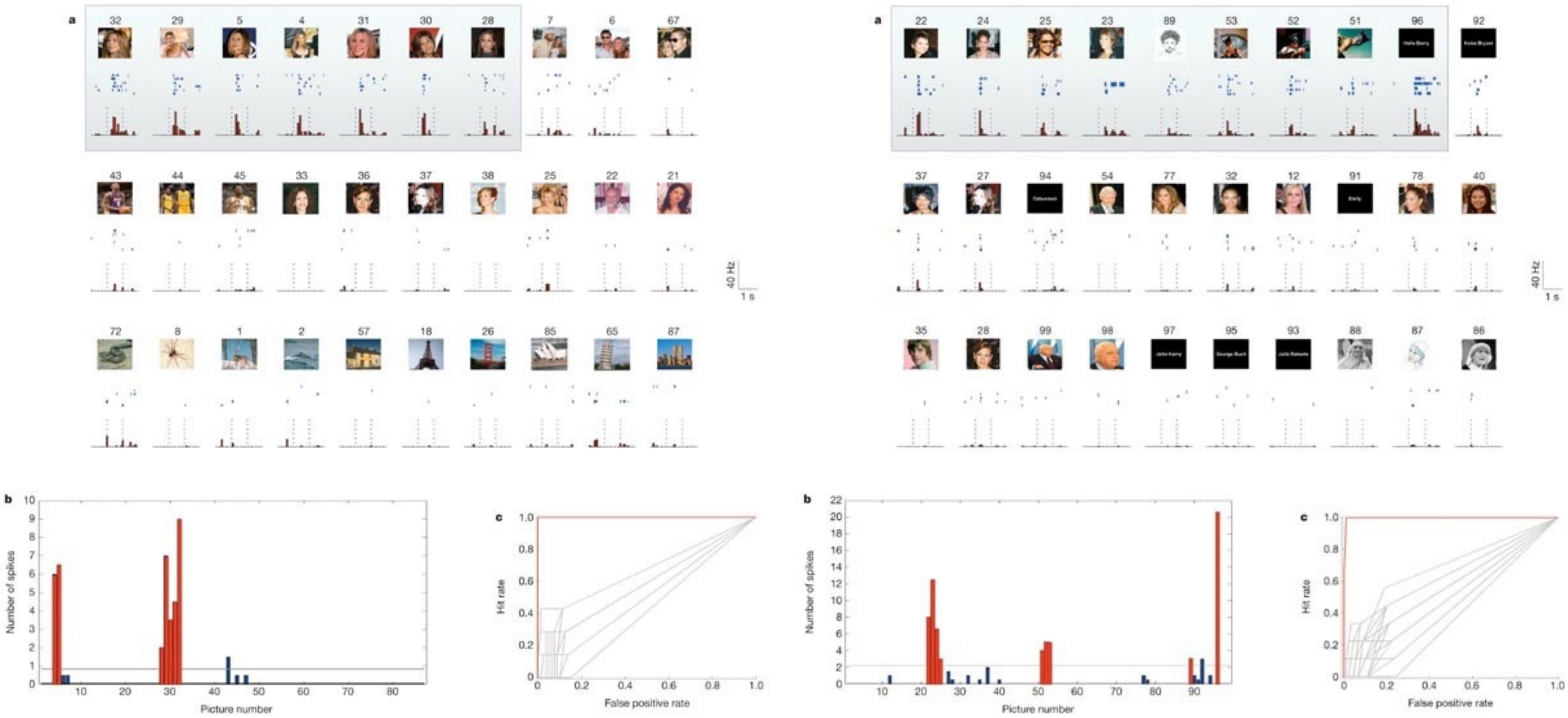
Invariant visual representation by single neurons in the human brain

R. Quian Quiroga, L. Reddy, G. Kreiman, C. Koch & I. Fried

Nature **435**, 1102–1107 (2005) | Cite this article

14k Accesses | **896** Citations | **320** Altmetric | Metrics

Jennifer Aniston and Halle Berry neurons



What is a face?

Dimensionality reduction

High-level,
Low-dim representation.

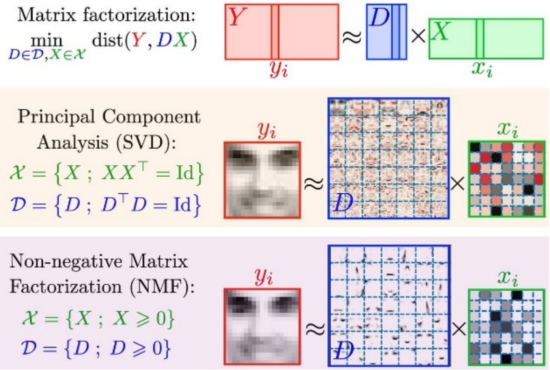
Published: 21 October 1999

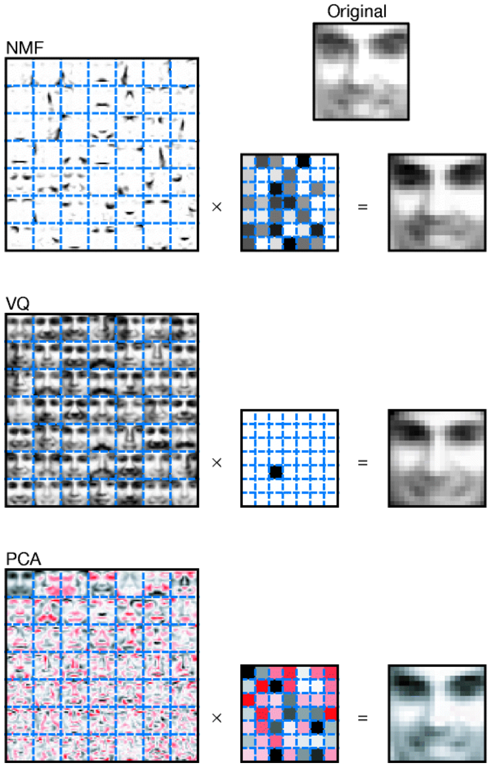
Learning the parts of objects by non-negative matrix factorization

Daniel D. Lee & H. Sebastian Seung

Nature **401**, 788–791 (1999) | Cite this article

56k Accesses | **7555** Citations | **48** Altmetric | Metrics



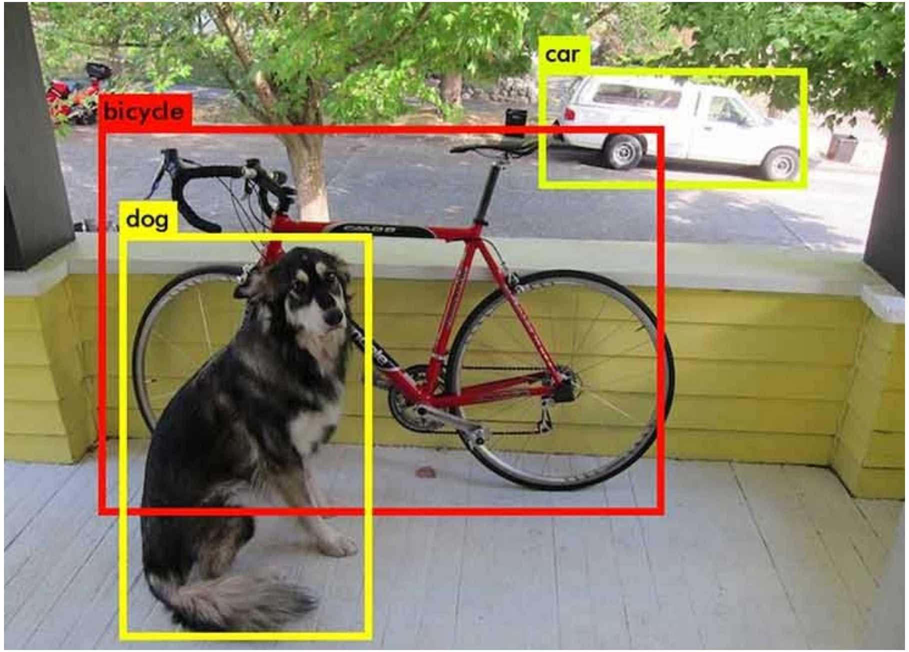


VQ → Vector Quantization

PCA+VQ Holistic, NMF: parts-based

Eigenfaces=„ghost“faces

Object detection

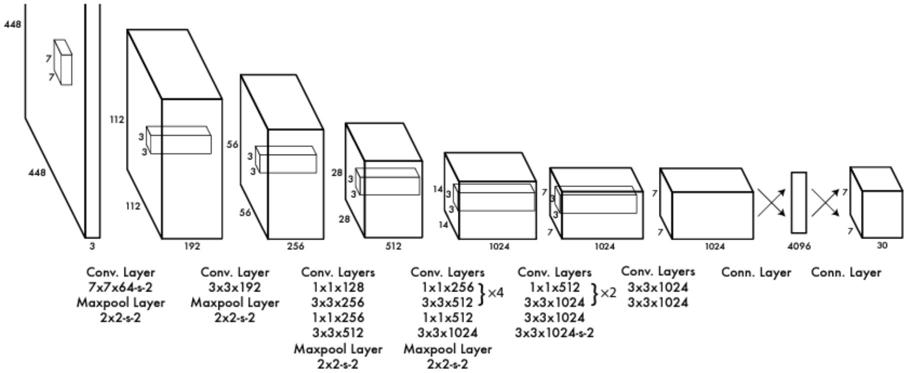


Find objects in an image and **identify their location** using a bounding box

One example: You Only Look Once (YOLO)

You Only Look Once:
Unified, Real-Time Object Detection

Joseph Redmon*, Santosh Divvala*†, Ross Girshick¶, Ali Farhadi*†
University of Washington*, Allen Institute for AI†, Facebook AI Research¶
http://pjreddie.com/yolo/



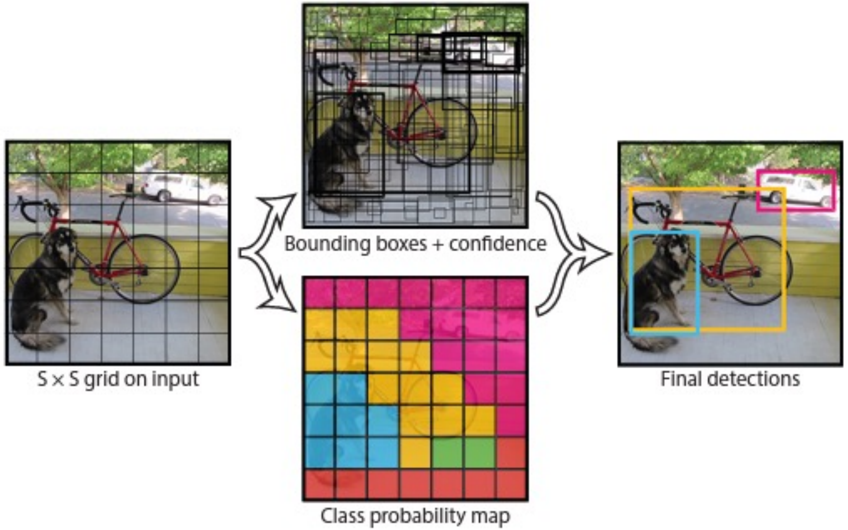


Figure 2: The Model. Our system models detection as a regression problem. It divides the image into an $S \times S$ grid and for each grid cell predicts B bounding boxes, confidence for those boxes, and C class probabilities. These predictions are encoded as an $S \times S \times (B * 5 + C)$ tensor.

[Introduction to YOLO](https://www.youtube.com/watch?v=MhftoBaoZpg)
https://www.youtube.com/watch?v=MhftoBaoZpg

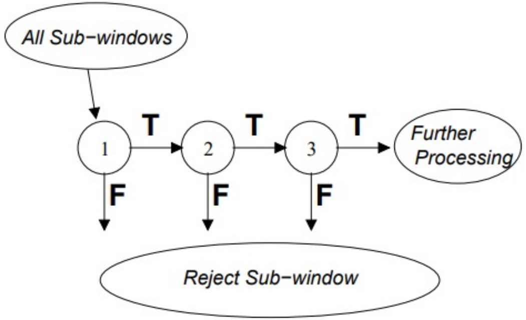
Face recognition using image processing

Accepted Conference on Computer Vision and Pattern Recognition 2001

Rapid Object Detection using a Boosted Cascade of Simple Features

Paul Viola
viola@merl.com
Mitsubishi Electric Research Labs
201 Broadway, 8th FL
Cambridge, MA 02139

Michael Jones
mjones@crl.dec.com
Compaq CRL
One Cambridge Center
Cambridge, MA 02142



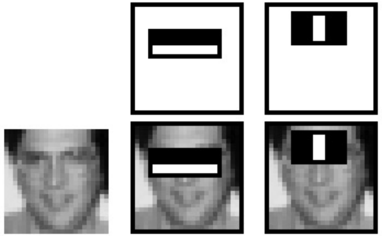


Figure 3: The first and second features selected by AdaBoost. The two features are shown in the top row and then overlayed on a typical training face in the bottom row. The first feature measures the difference in intensity between the region of the eyes and a region across the upper cheeks. The feature capitalizes on the observation that the eye region is often darker than the cheeks. The second feature compares the intensities in the eye regions to the intensity across the bridge of the nose.

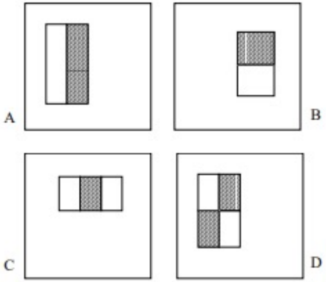
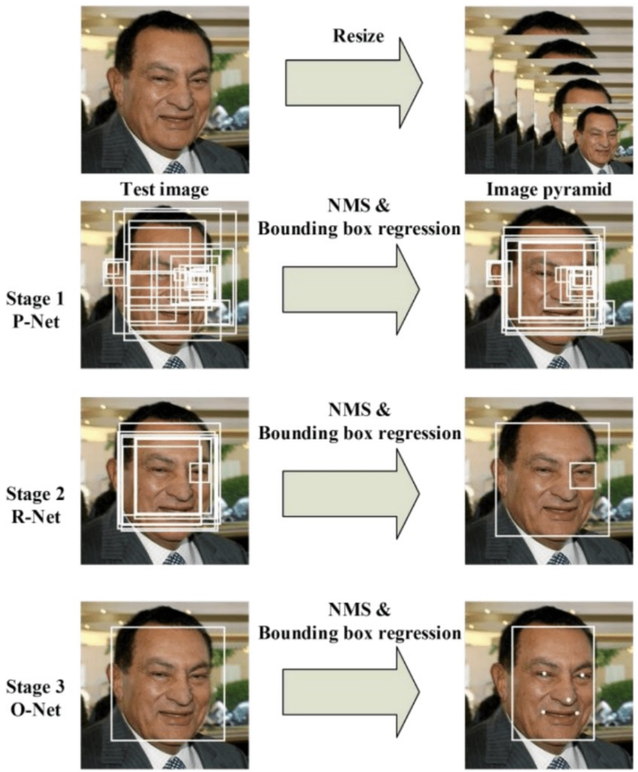


Figure 1: Example rectangle features shown relative to the enclosing detection window. The sum of the pixels which lie within the white rectangles are subtracted from the sum of pixels in the grey rectangles. Two-rectangle features are shown in (A) and (B). Figure (C) shows a three-rectangle feature, and (D) a four-rectangle feature.

Deep neural network face detection



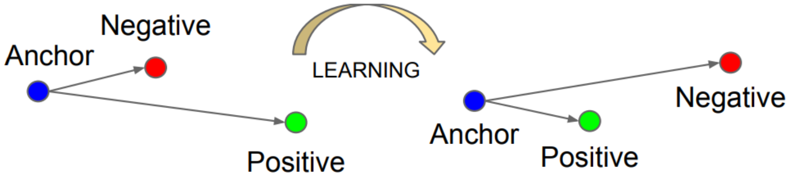
Joint face detection and alignment using multitask cascaded convolutional networks

K Zhang, Z Zhang, Z Li, Y Qiao - IEEE Signal Processing Letters, 2016 - ieeexplore.ieee.org

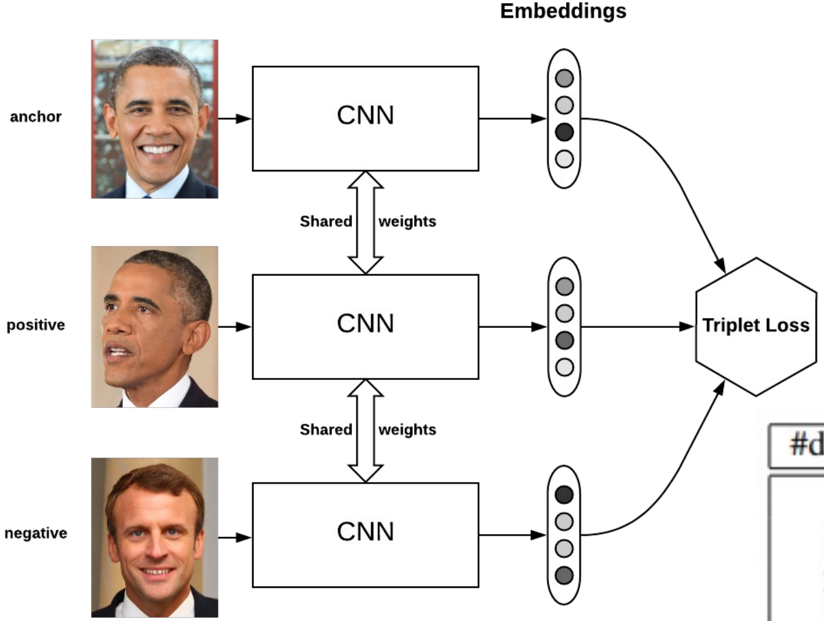
Face detection and alignment in unconstrained environment are challenging due to various poses, illuminations, and occlusions. Recent studies show that deep learning approaches can achieve impressive performance on these two tasks. In this letter, we propose a deep cascaded multitask framework that exploits the inherent correlation between detection and alignment to boost up their performance. In particular, our framework leverages a cascaded architecture with three stages of carefully designed deep convolutional networks to predict …

☆ 99 Zitiert von: 2717 Ähnliche Artikel Alle 4 Versionen

Face identification: FaceNet



[**Facenet**: A unified embedding for face recognition and clustering](#)
F Schroff, D Kalenichenko, J Philbin - Proceedings of the IEEE ..., 2015 - cv-foundation.org
Despite significant recent advances in the field of face recognition [DeepFace, DeepId2], implementing face verification and recognition efficiently at scale presents serious challenges to current approaches. In this paper we present a system, called **FaceNet**, that ...
☆ 99 Zitiert von: 8106 Ähnliche Artikel Alle 24 Versionen »



jpeg q	val-rate
10	67.3%
20	81.4%
30	83.9%
50	85.5%
70	86.1%
90	86.5%

#pixels	val-rate
1,600	37.8%
6,400	79.5%
14,400	84.5%
25,600	85.7%
65,536	86.4%

#dims	VAL
64	$86.8\% \pm 1.7$
128	$87.9\% \pm 1.9$
256	$87.7\% \pm 1.9$
512	$85.6\% \pm 2.0$

© Olivier Moindrot