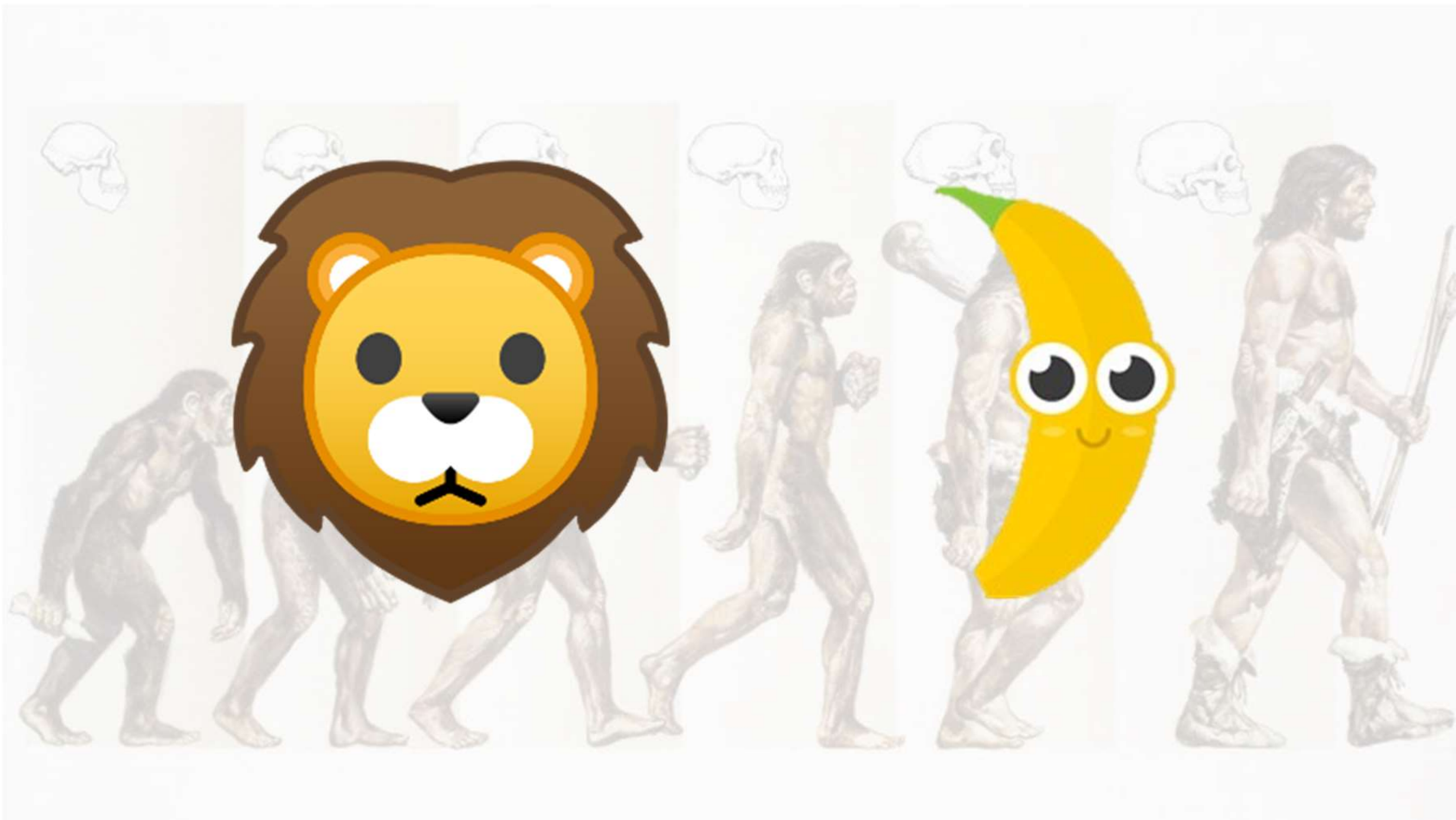
Perception and Interfaces (COMP0160) 2022/23

Visual Object Recognition

Tobias Ritschel



Overview

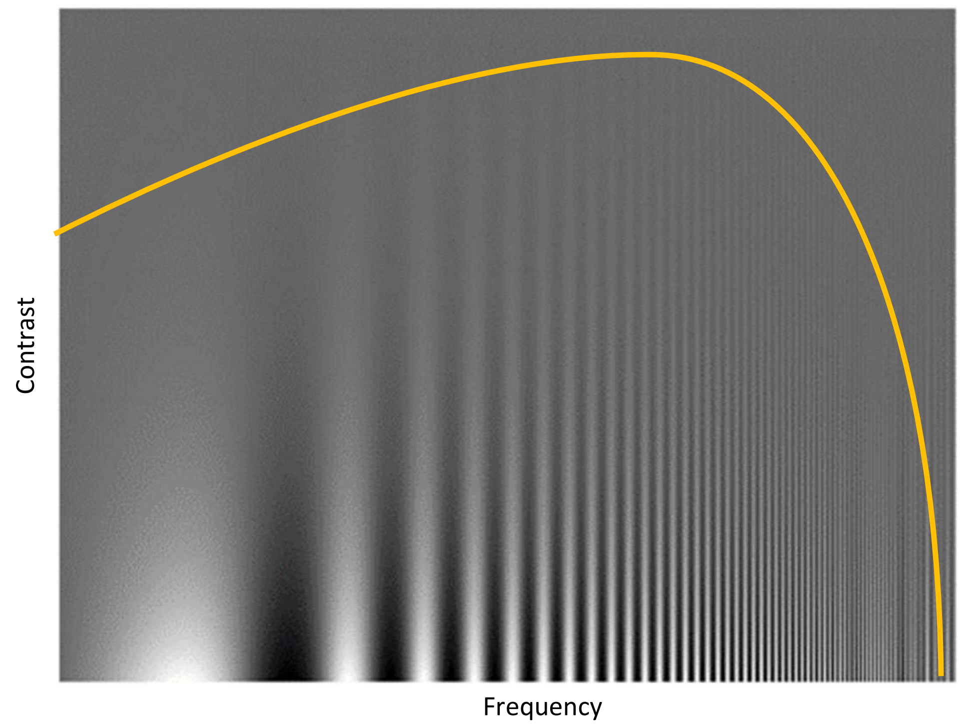
- We will take four different views on this problem:
  1. Low-level
  2. Phenomenological
  3. Neuro-physiological
  4. Computational

Low-level

- Hard facts about the limits of …
- Spatial resolution
- Temporal resolution
- Chromatic perception
- Luminance adaptation
- Peripheral vision

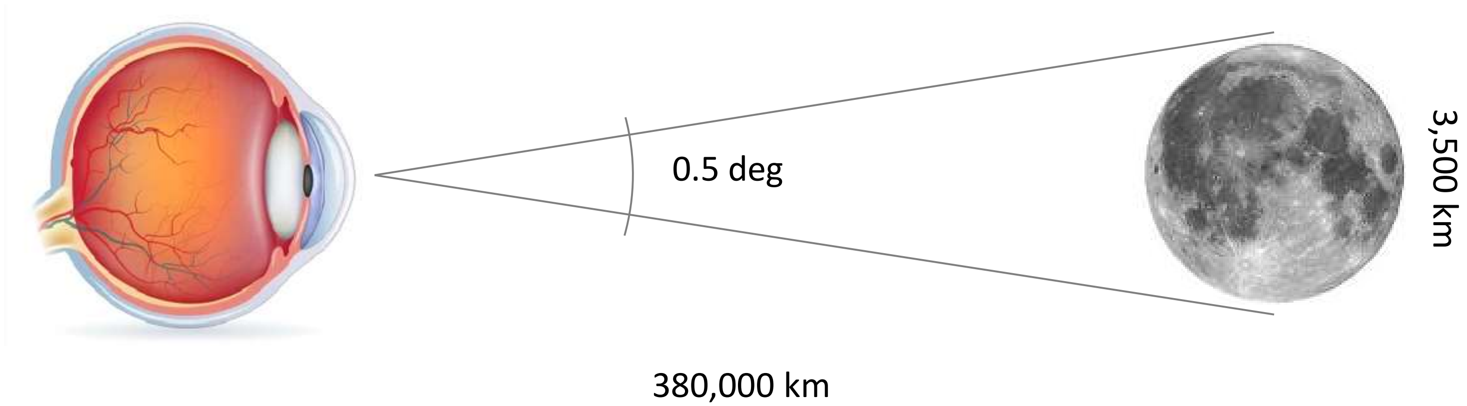
Spatial resolution

- Limited by
  - Optics
  - Photoreceptor density 感光器
- Campbell-Robson chart



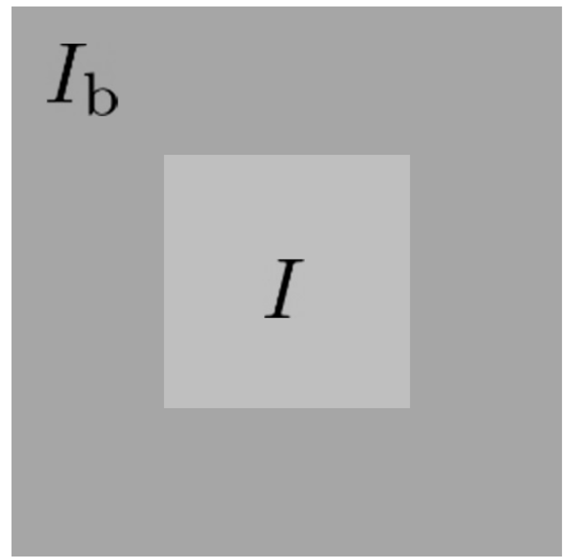
Visual angle

- Don’t ask about size, distance or pixels or stuff, ask about **visual angle** in degree
- Changes-of-something are in **cycles per degree**



Luminance contrast

- How much something is different from something else
- Weber contrast $\frac{I - I_\mathrm{b}}{I_\mathrm{b}}$
- Example: (110-100)/100 = 0.1
- Mitchelson contrast $\frac{2(I_1 - I_2)}{I_1 + I_2}$



Weber’s law

- The fraction by which a stimulus needs increment to be perceived is constant

Low stimulus

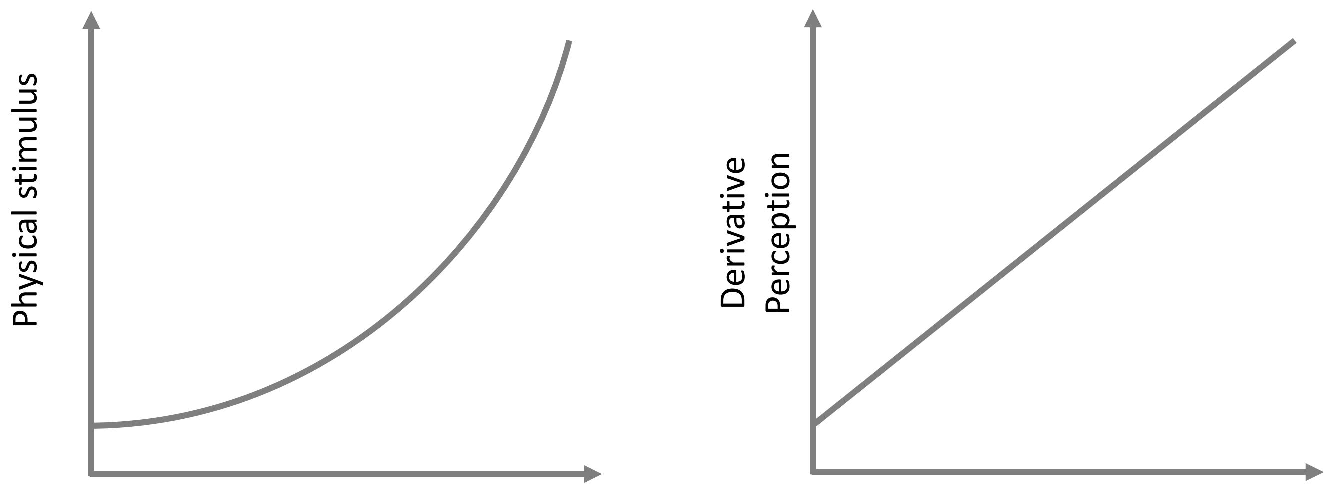
x1.01, not perceived x1.1, perceived

High stimulus

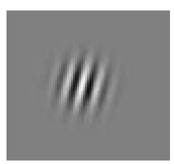
x1.01, not perceived x1.1, perceived

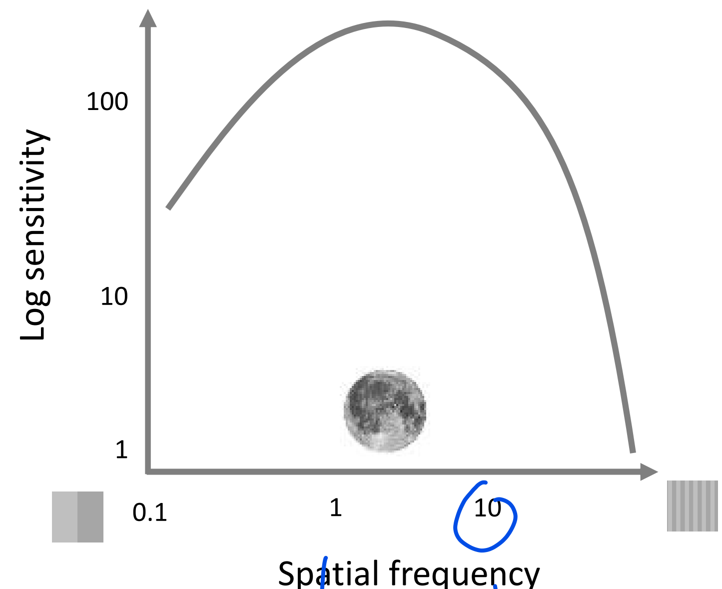
Relation to log

- To get a constant increase in response
- We need an increasing change of stimulus
- Like interest rates in finance or *R* number in CoViD



Contrast Sensitivity Function (CSF)

- For Gabor patches
- **Spatial frequency**:
  How often it changes per visual angle
  Sensitivity is to 1/contrast-threshold
- **Contrast-threshold**:
  How much to add so that 70% can see
- Stimulus: 



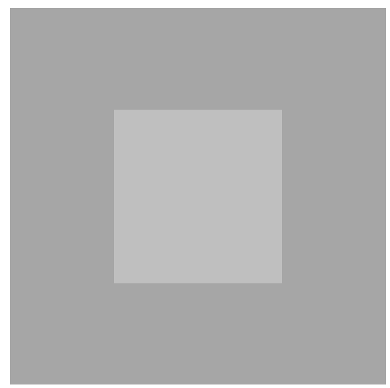
视1°变10

视1°变1

Suprathreshold vs detection

- Adding something to nothing (**Detection**)
  - There is never “nothing” in reality, so “very small”
- Adding something-to-something (**Supra-threshold**)

Detection



Supra-threshold

Retinal illuminance

- Invariant unit is **trolands** that is retinal illuminance

+

~

+

Weak light

Large pupil

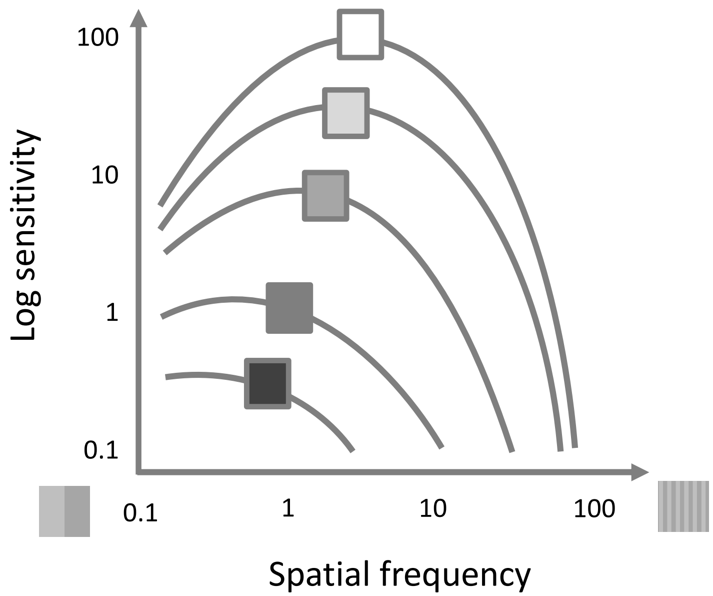
has equivalent troland

Strong light

Small pupil

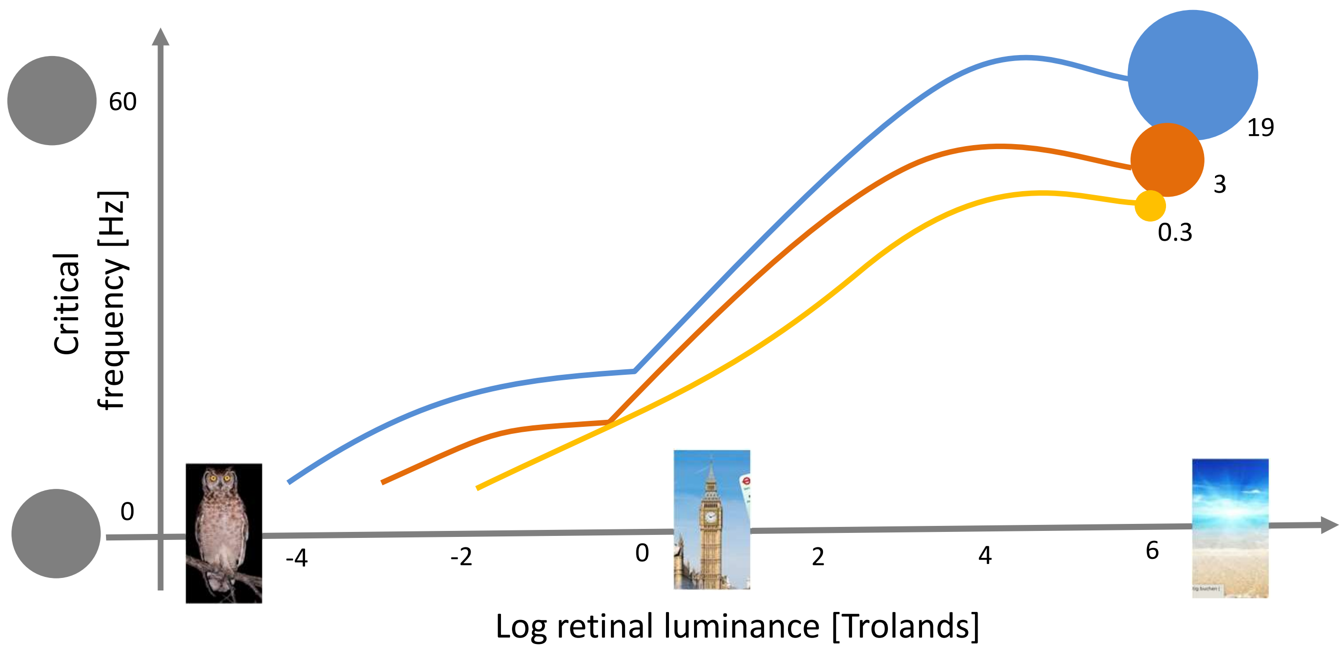
Suprathreshold CSF

- There is one CSF for every base retinal luminance
- Roughly:
  - When to dark, low freq is better
  - The brighter, the more there is a preferred freq around 3



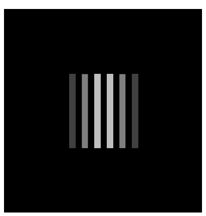
Flicker fusion

- Brightness changes quicker than threshold not discerned
- Depends heavily on retinal illuminance and size

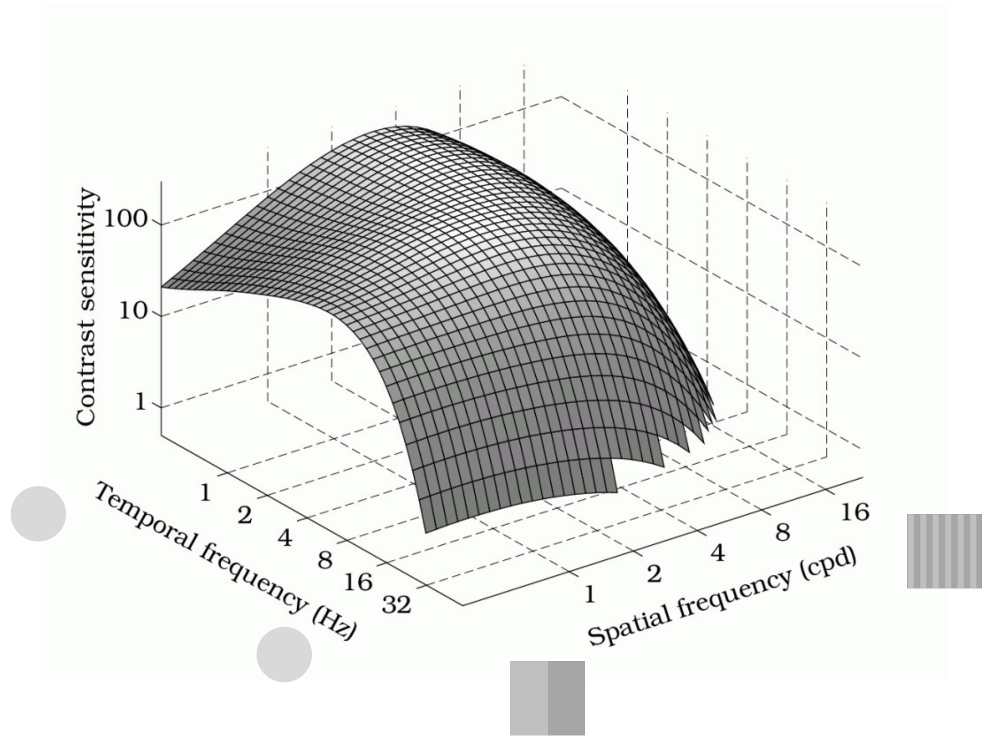


Spatio-temporal

- Sinusoisal
- Changing at temporal freq
- Changing at spatial freq
- Luminance detection threshold
- At 900 Troland

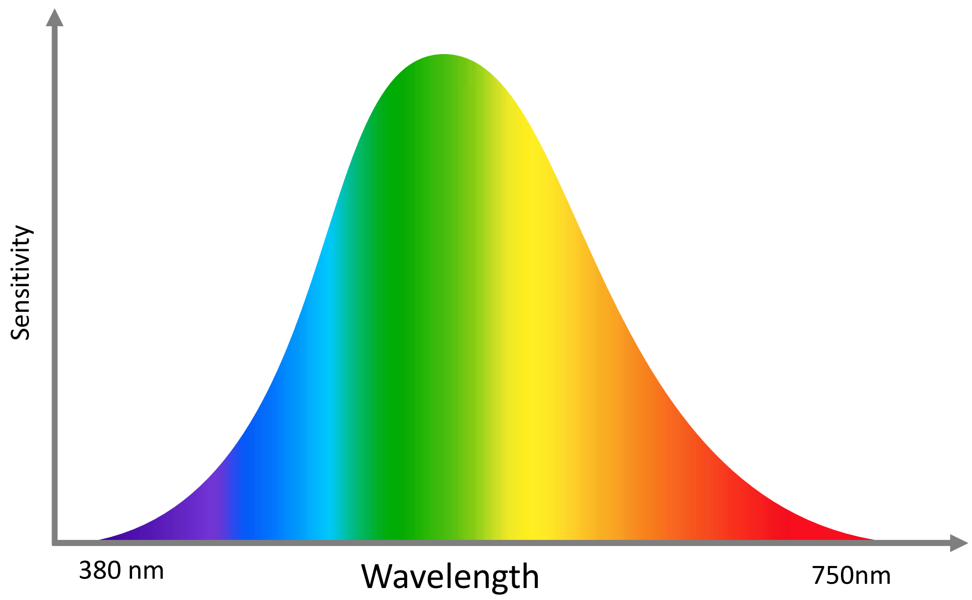


Stimulus



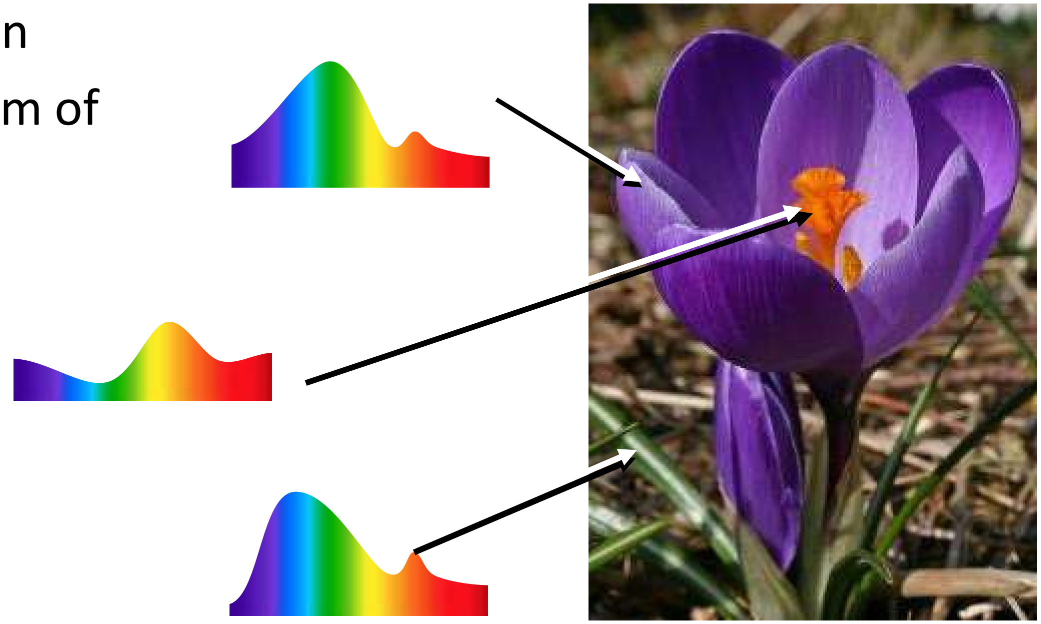
Luminance sensitivity

- Luminance sensitivity depends on wavelength



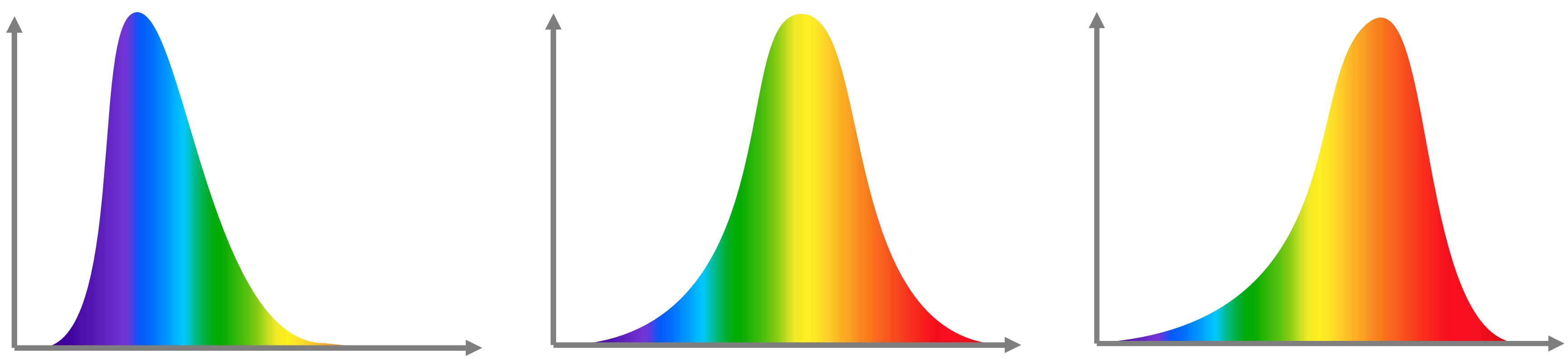
Chromatic perception

- Humans perceive color
- “Color“ is our name for a sensation
- The physical quality is the spectrum of electromagnetic radiation



Trichromatism 三色性

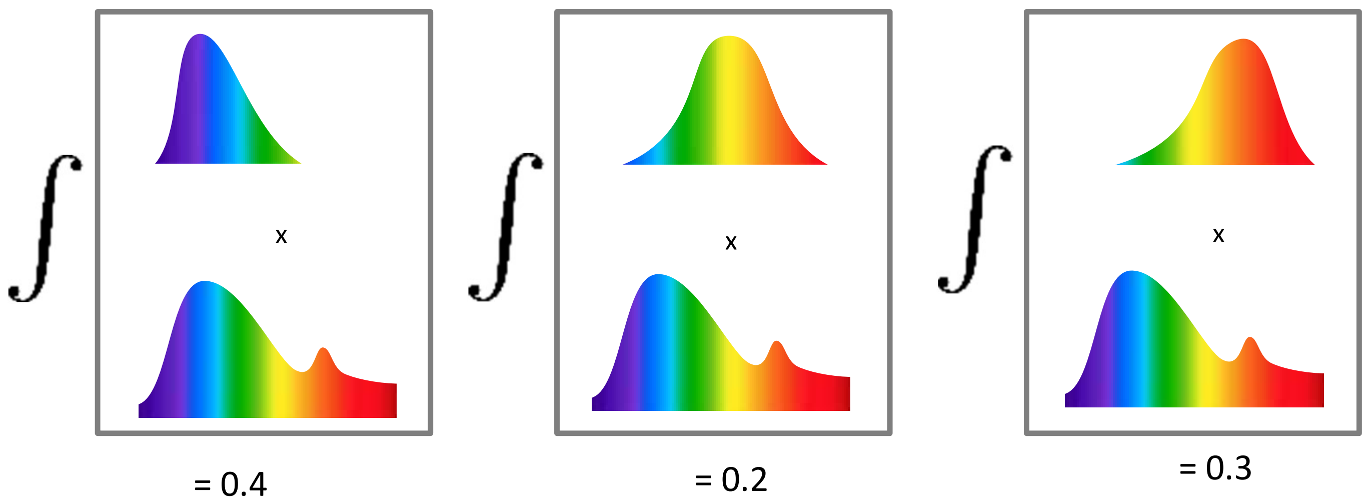
- Color sensitivity is three bases in function space
- No, this isn't RGB



Short Medium Long

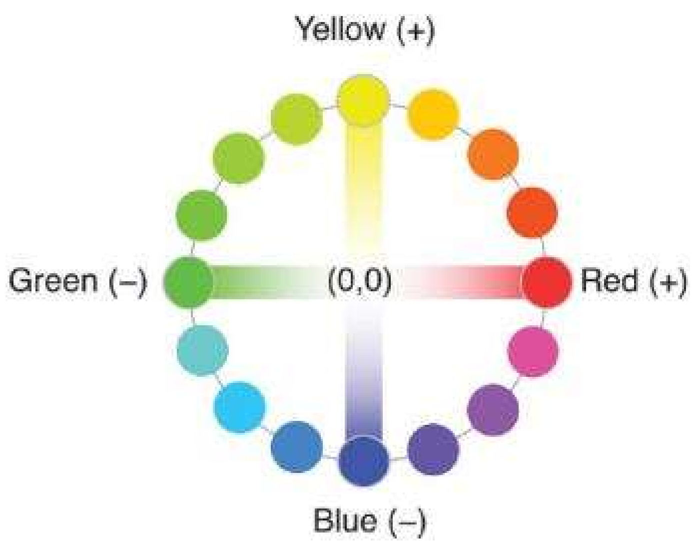
Trichromatism

- Sensation is dot product of a spectrum and the basis



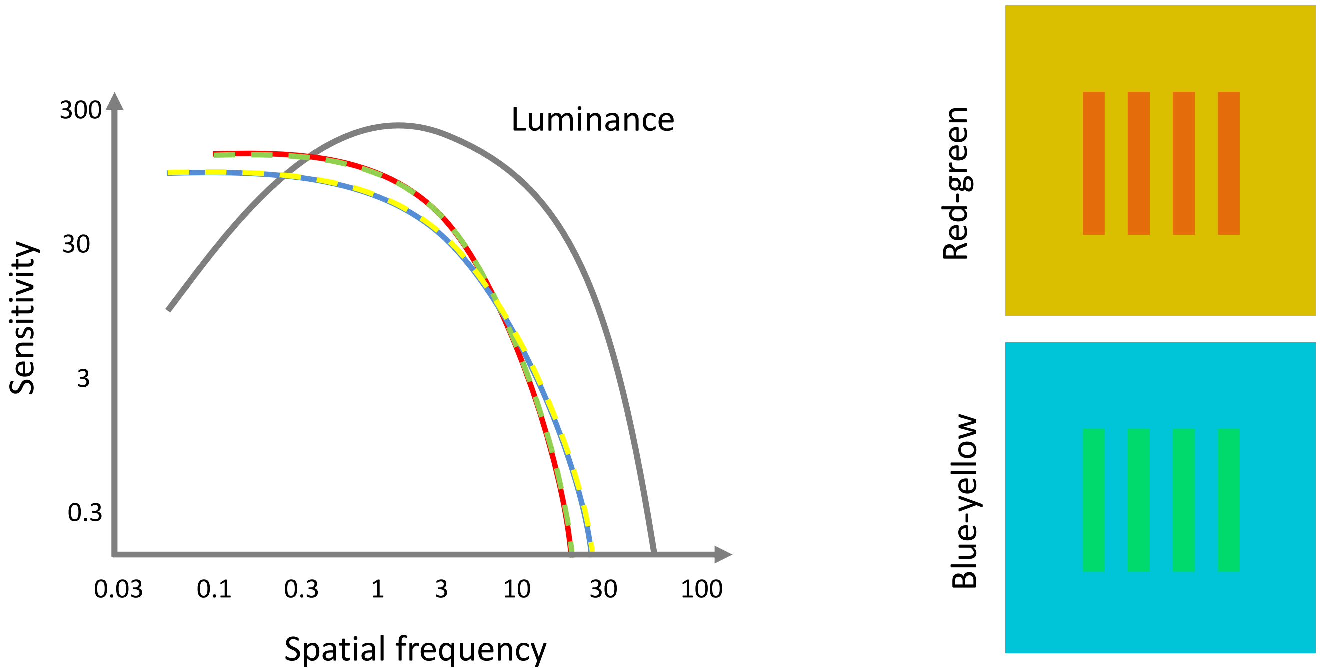
Color opponency and luminance

- These three responses are mapped into
  - A one-dimensional sensation of luminance
  - A two-dimensional sensation of chrominance
- We see physiology of this later



Chromatic sensitivity function

- How much do you need to add to a spatial colored pattern to see the change

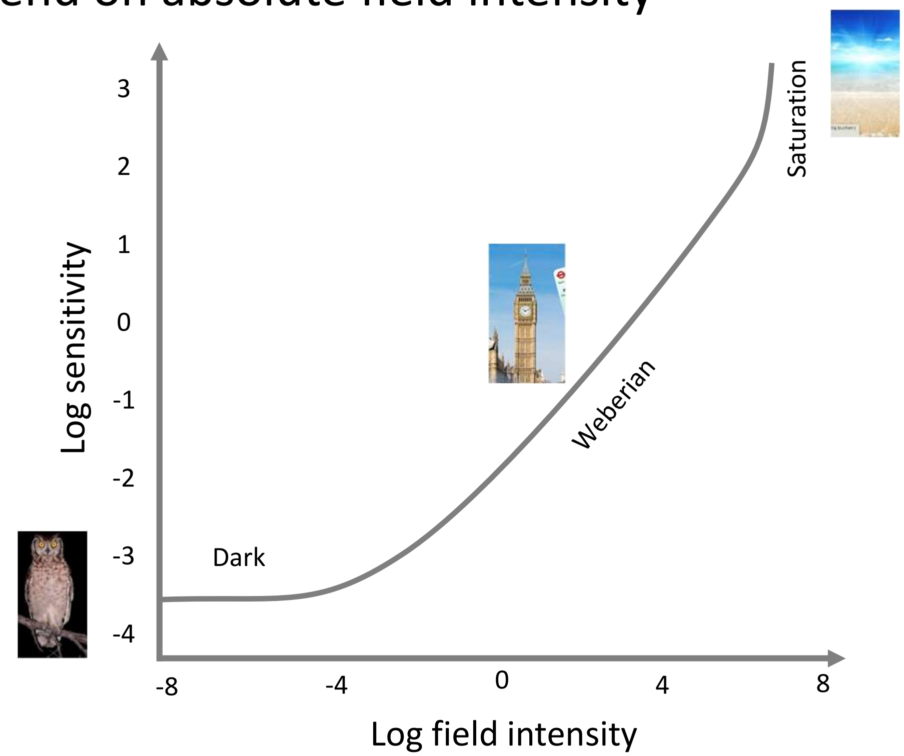


Luminance Adaptation

- HVS adapts to the average physical luminance
- Allows us to see in day and night, ten orders of magnitude
- Several details change across that range, we later see how
- At one point in time, only two orders of magnitude

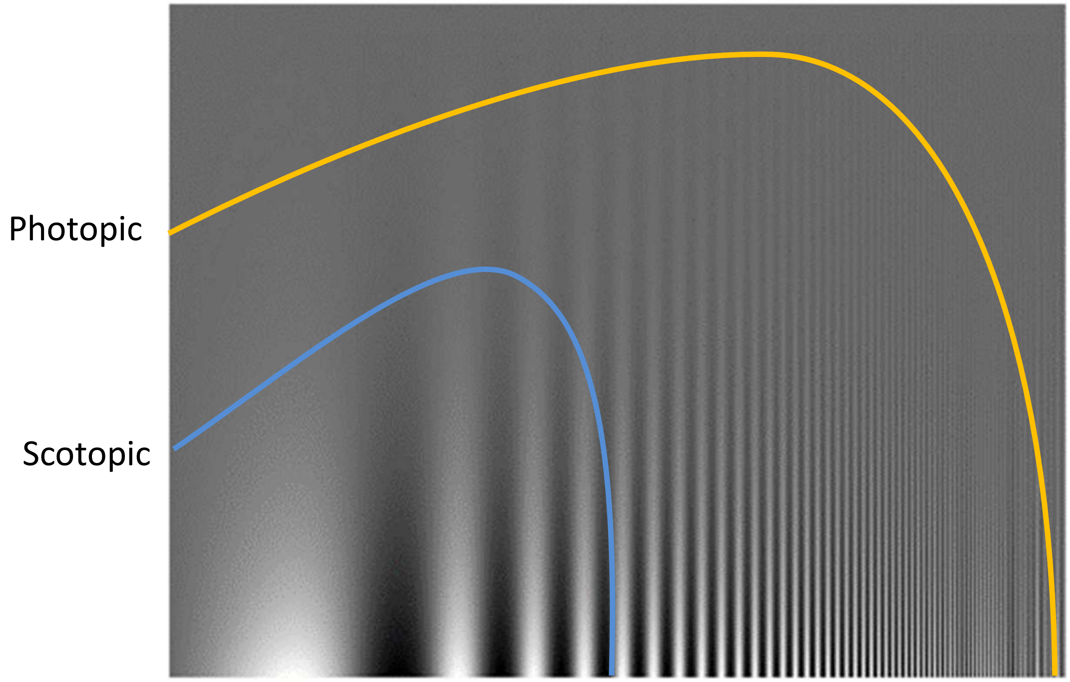
Luminance Adaptation

- Threshold depend on absolute field intensity



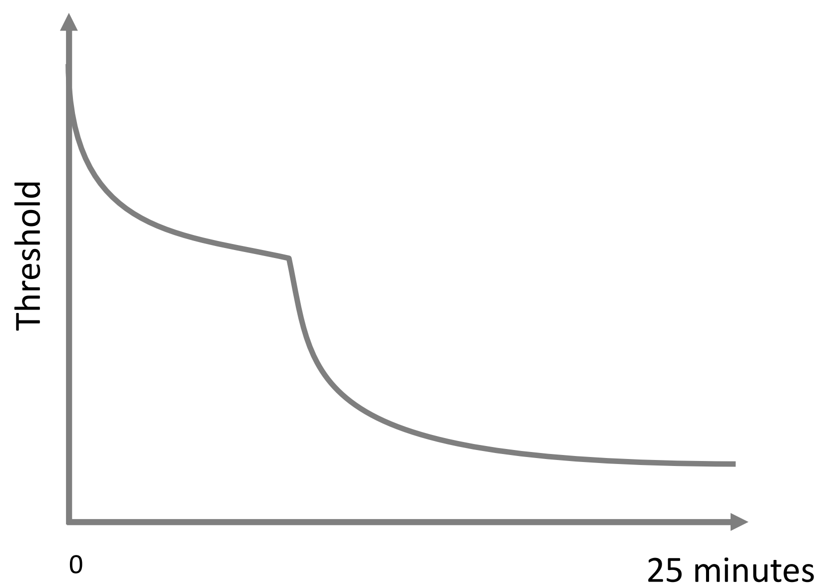
Spatial contrast at night

- We see fewer spatial details at night
- Subtle value changes missing
- High freq missing



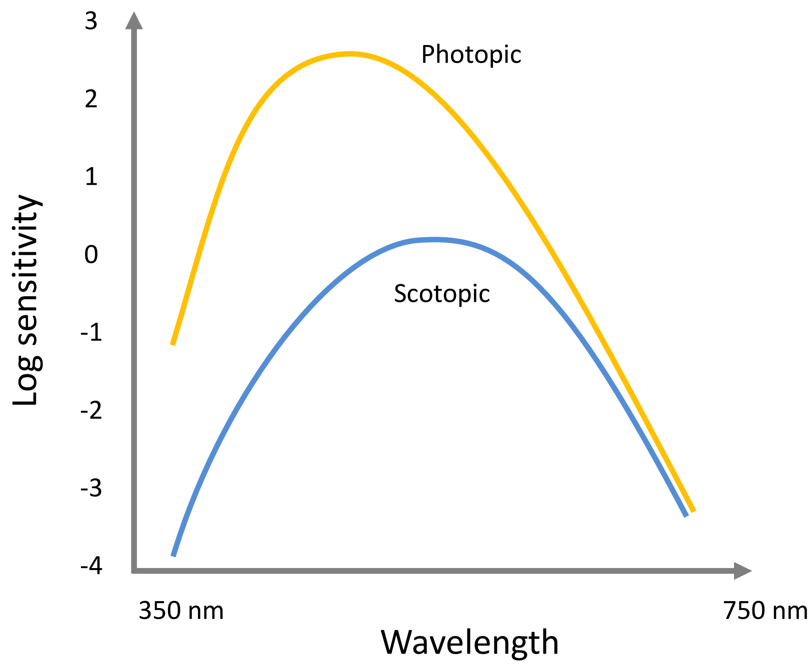
Time course of adaptation

- Adaptation takes considerable time to be effective



Color perception in the dark

- The luminance efficiency function shifts its peak
- Purkinje shift

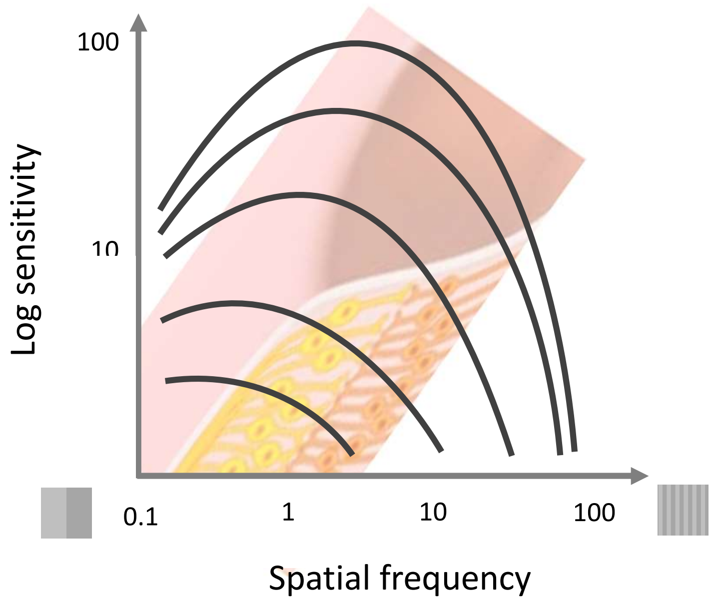


Peripheral vision

- We perceive more details in the center of our visual field (more later as to why)
- I do not draw this as blur as it is not blur

Foveation (Virso and Rovamo 1979)

- We perceive more details in the center of our visual field (more later as to why)
- Fovea perceives
  - More higher frequencies
  - More details in value



Literature

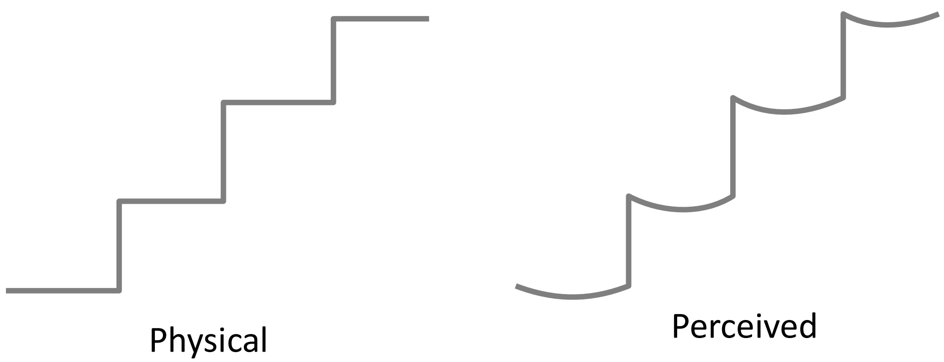
- Wandell, Brian A. "**Foundations of vision.**" *Sinauer Associates*, 1995.
- Campbell, Fergus W., and John G. Robson. "**Application of Fourier analysis to the visibility of gratings.**" *The Journal of physiology* 197.3 (1968): 551.
- Van Nes, Floris L., and Maarten A. Bouman. "**Spatial modulation transfer in the human eye.**" *JOSA* 57.3 (1967): 401-406.
- Virsu, V., and J. Rovamo. "**Visual resolution, contrast sensitivity, and the cortical magnification factor.**" *Experimental brain research* 37.3 (1979): 475-494.

Phenomenological 现象学

- What kinds of phenomena in visual perception exist?
- Less a solution to anything, more a shopping list of what to account for
- Illusions
- Depth cues
  - Monocular
  - Binocular
  - Fusion
- Gist

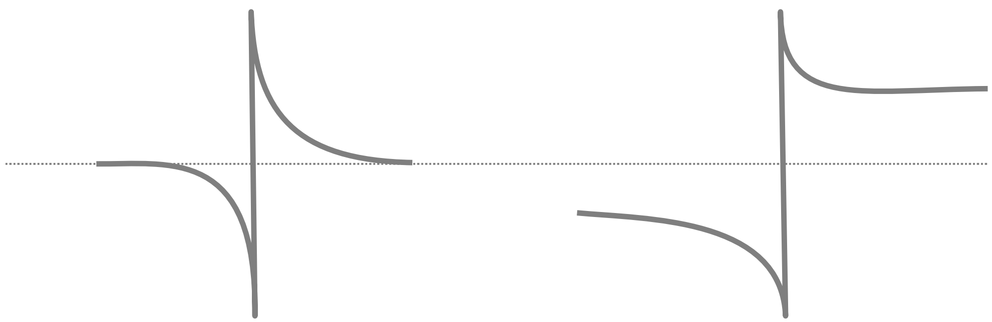
Mach bands

- Piecewise constant
- Perceived as curved



Cornsweet illusion

- Two constant colors with wedge at the junction
- Perceived as different brightness



Physical

Perceived

Depth cues

- We want to understand **how far away** the lion is
- Monocular
  - Pictorial
    - Relative size
    - Occlusion
    - Aerial perspective
  - Non-pictorial
    - Lens accommodation
- Binocular
  - Vergence
  - Binocular disparity

Familiar size

- We know a banana is smaller than a lion
- Hard to imagine this is a huge banana in the sky and a tiny lion, no?

Familiar size

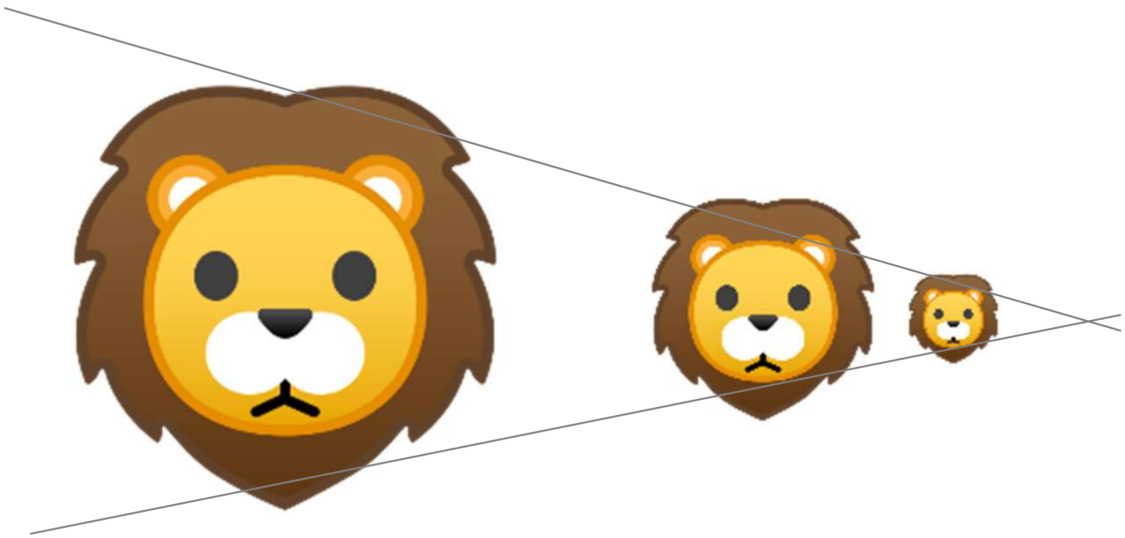
- Works better with natural images


Relative size

- Larger objects of same familiar size perceived closer

Linear perspective

- Linear perspective enforces.
- Objects on perspective lines are same-size



Size over horizon

- Adding a horizon, the pictorial relation is enforced further

Occlusion

- Closer objects occlude more far-away objects
- Works extremely well across all distances
- Only ordinal

Density

- Density is associated with distance
- A bit weaker

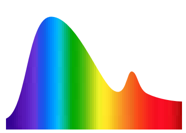
Moon illusion

- Why does the moon look so large on the horizon?
- It gets silly … Godzilla banana

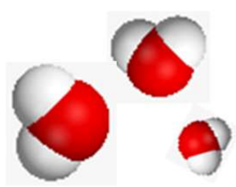
空气透视

Aerial perspective

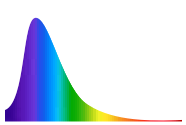
- Light is scattered in the atmosphere
- Scatter different at different wavelength
- Longer light path, more scattering
- More blue, more path, more distance



Spectrum



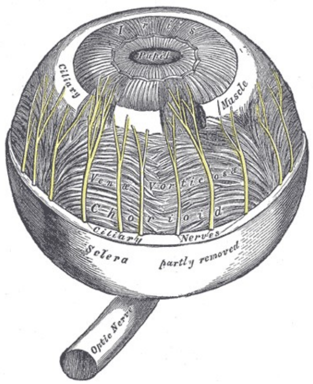
Water



Scattering

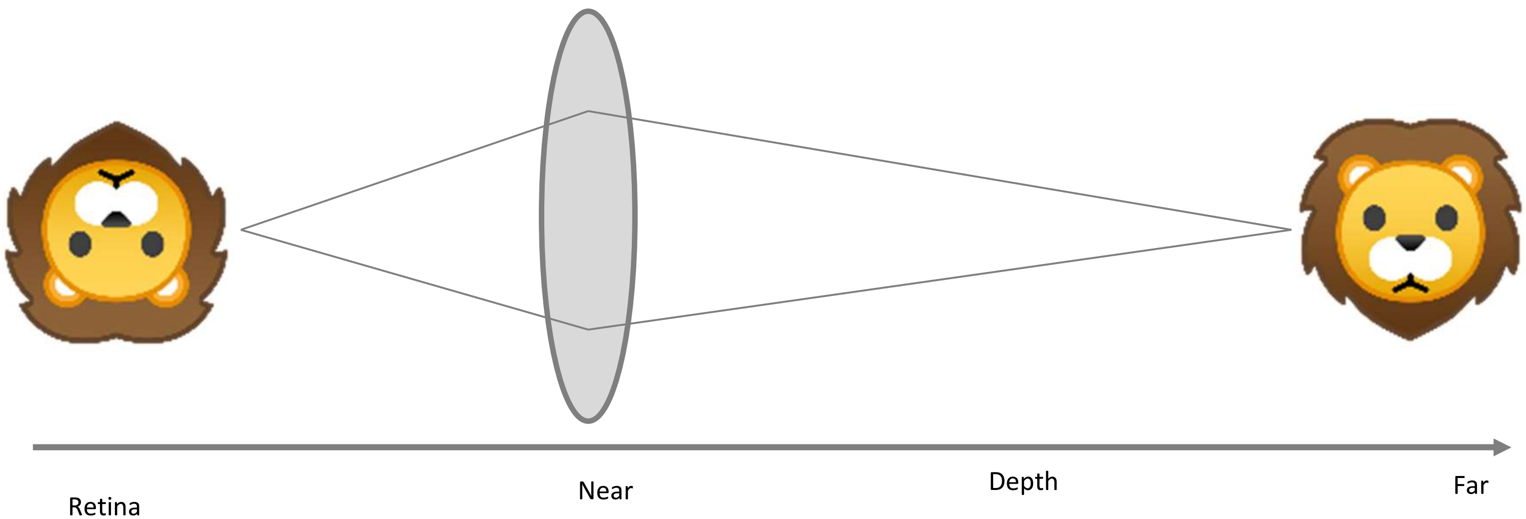
Lens Accommodation

- Monocular but not pictorial
- Objects at a certain distance will be sharp
- Objects at other distances will be blurry



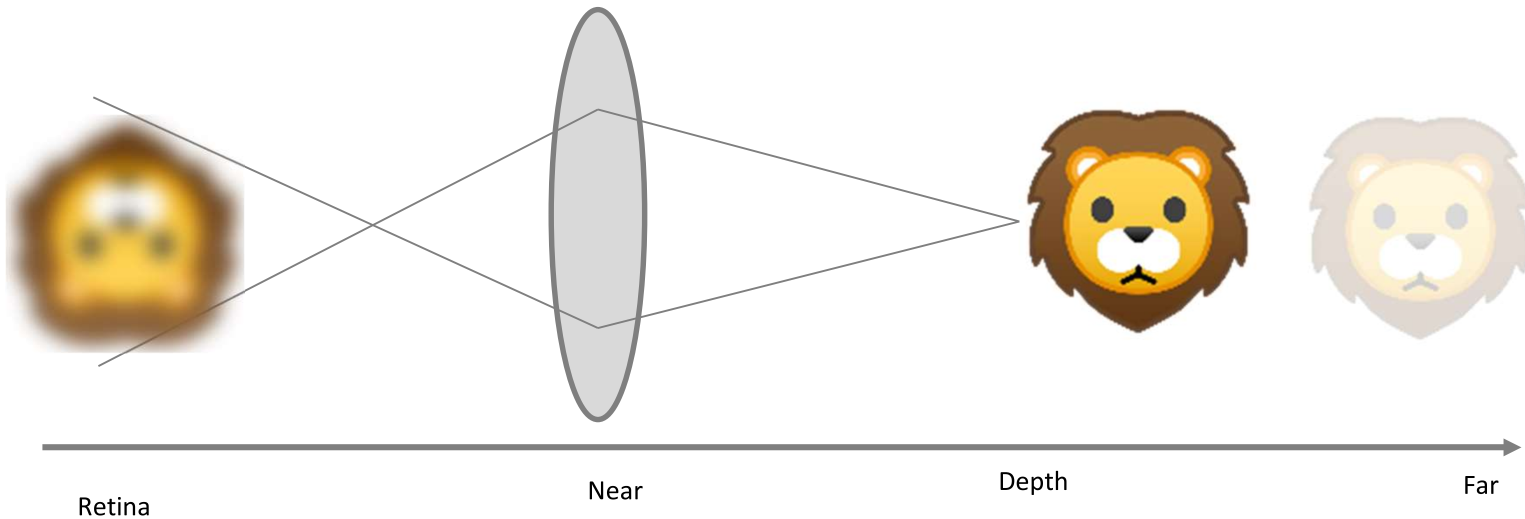
Accommodation: How it works

- Rays from a point at some depth map to a distance-dependent area



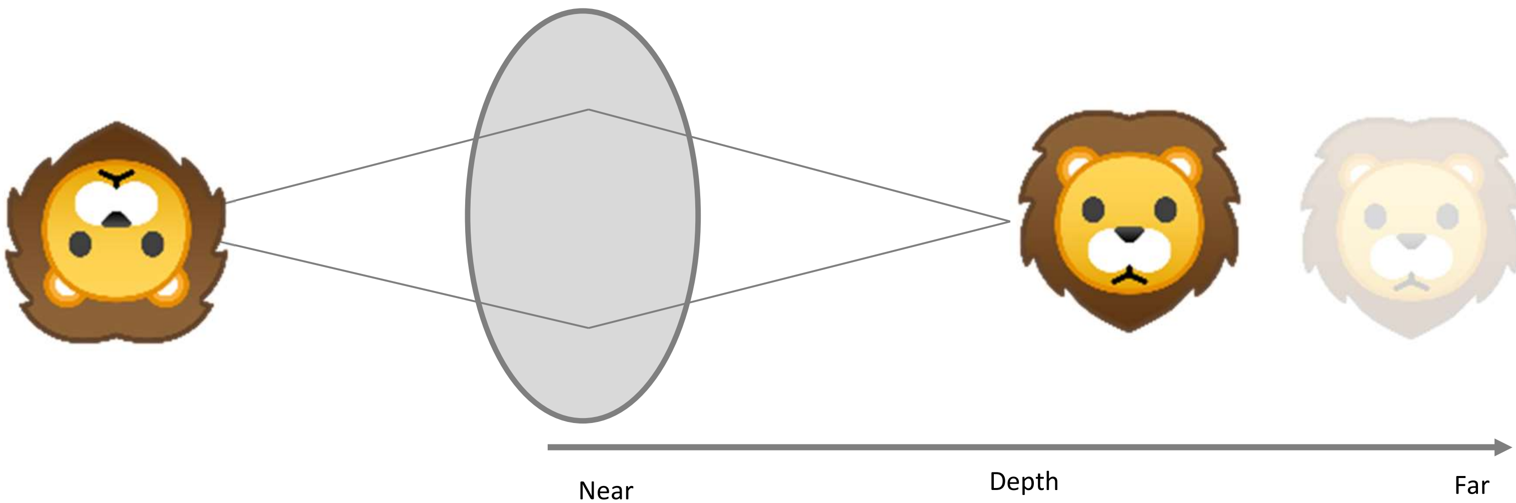
Accommodation: How it works

- Rays from a point at some depth map to a distance-dependent area
- Area is blur
- From lens state and blur, we can compute depth



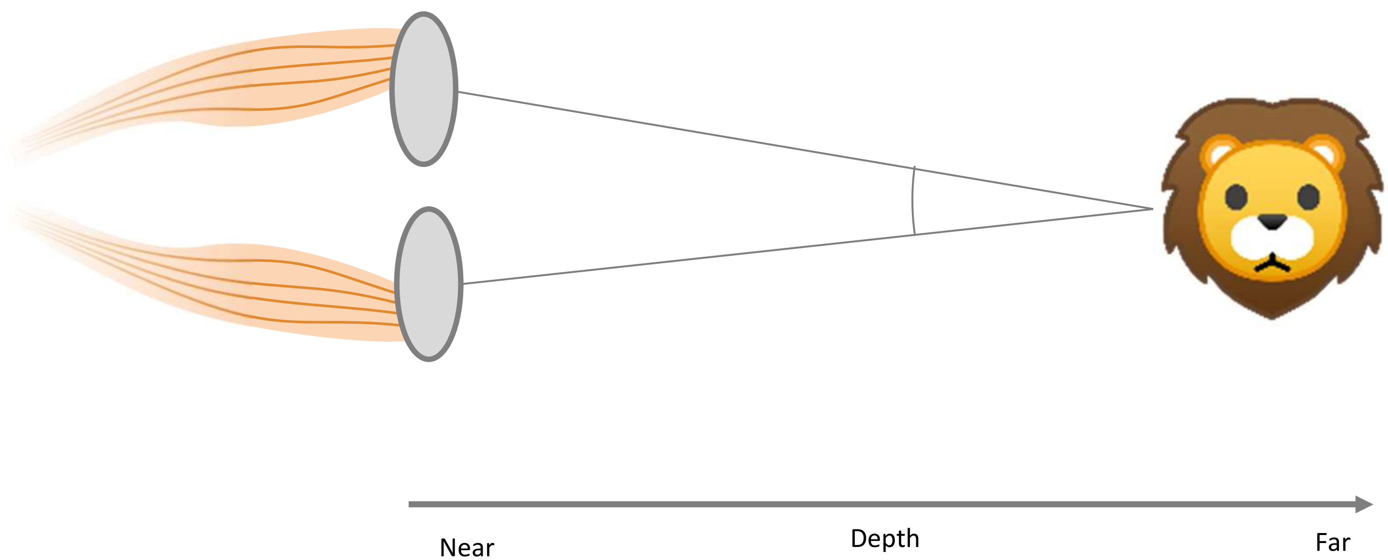
Accommodation: How it works

- Lens is flexible
- At different state, different things are in focus



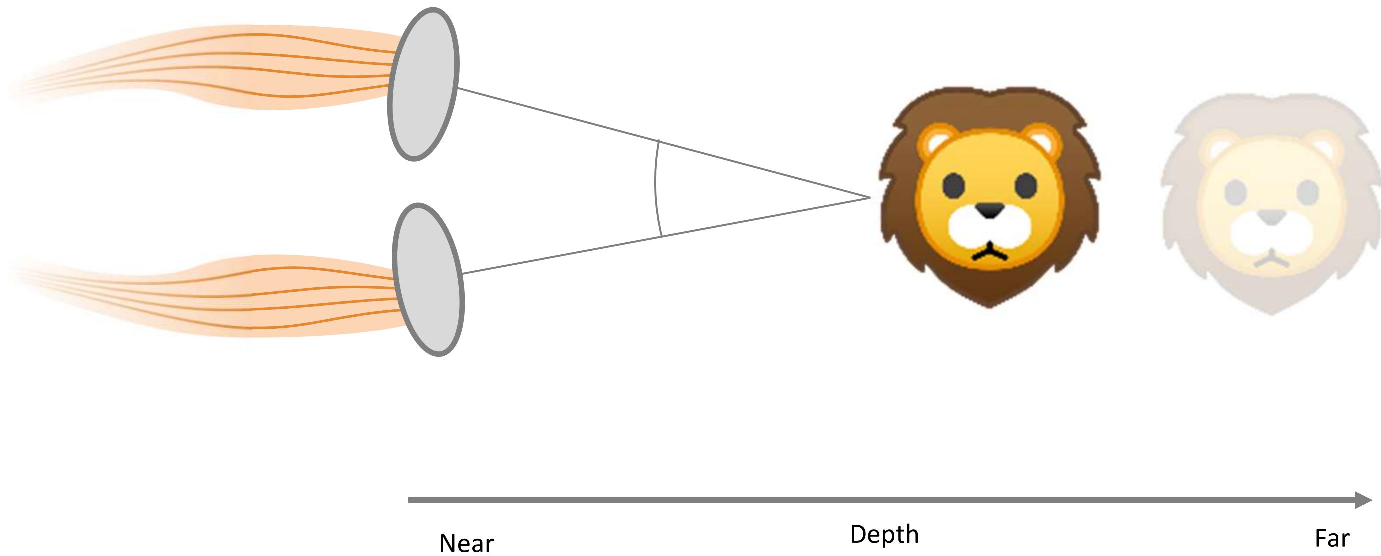
Binocular Convergence

- Eyes form different angles fixating 注视
- Muscles measure that angle



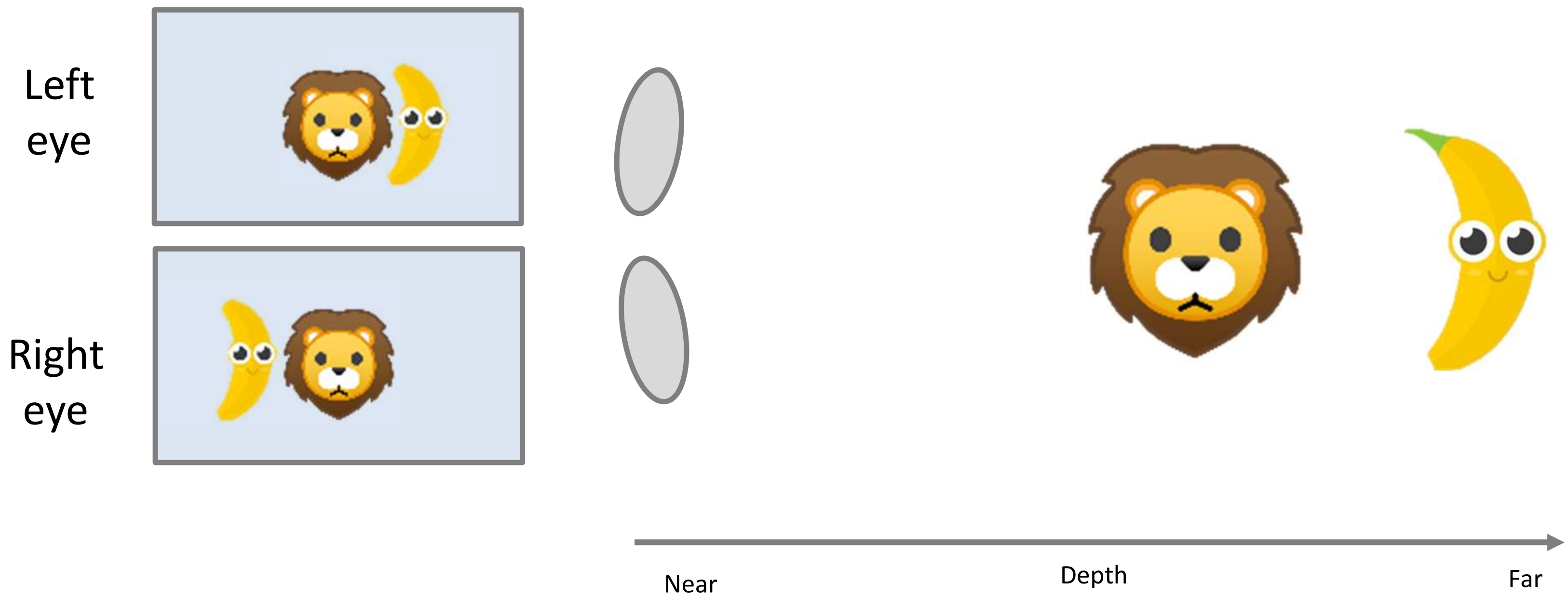
Binocular Convergence

- Eyes form different angles fixating
- Muscles measure that angle



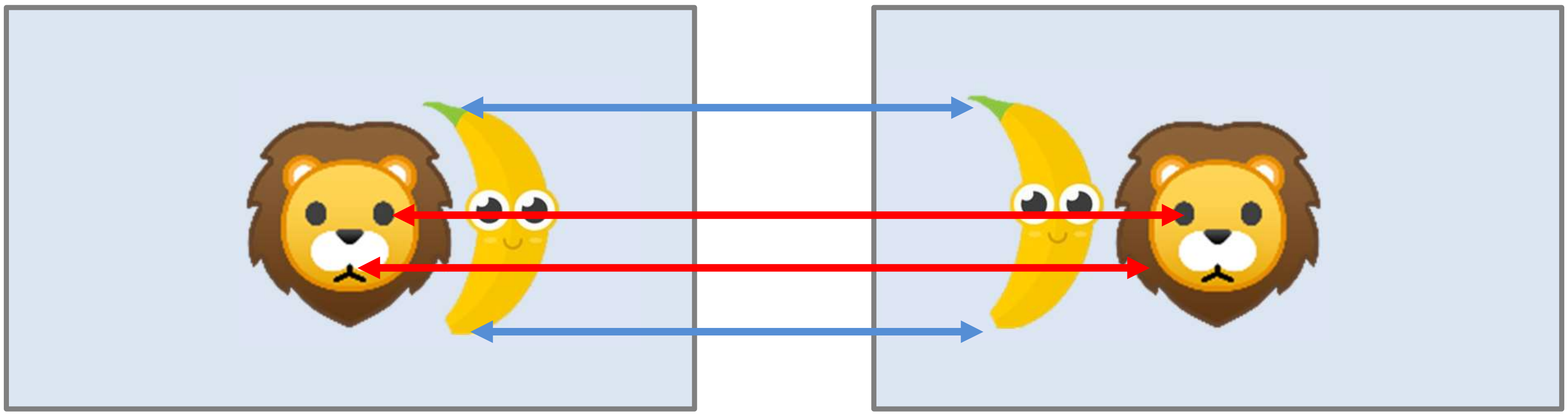
Binocular disparity

- Signed difference between retinal locations of same world points depends on depth



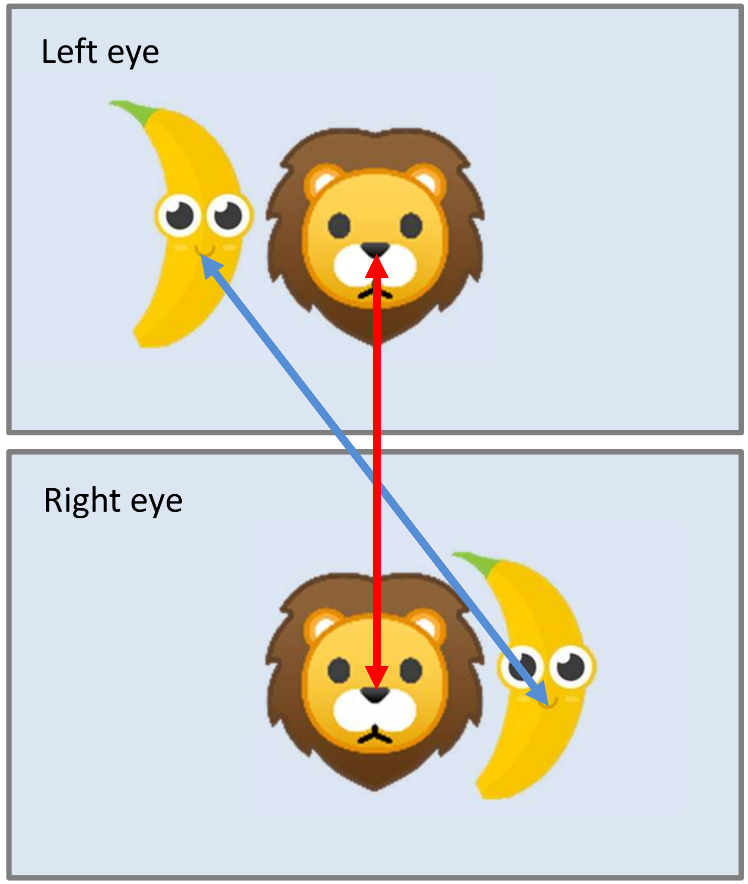
Binocular disparity

- Requires to **match** points
- Same distance, same depth
- Depends on vergence



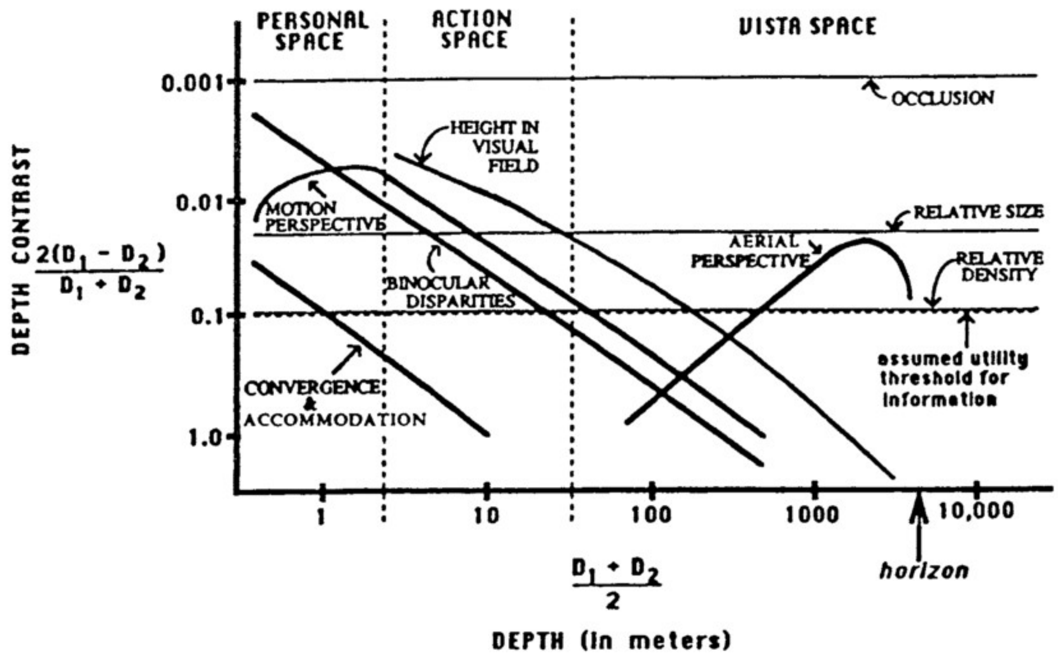
Binocular disparity

- Another way to see it
- Horopter is surface of points you verge on
- Objects at horopter
  - Do not change position between eyes
  - Lion here
- Objects away from it
  - Do change
  - Bana here



(Cutting and Vishton, 1995)

- Different depth cues are differently effective at different absolute scales



Fusion (Landy 1995)

human eye system

- The cues are **fused**
- HVS has a notion of **confidence**
- Cues are fused Bayesian

Depth of cue i

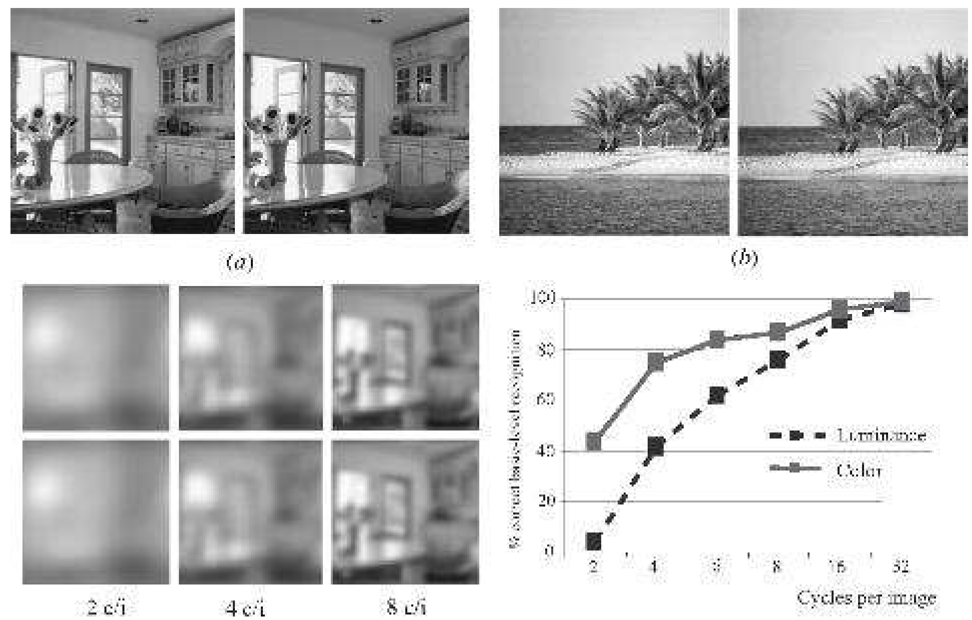
$$\frac{1}{\sum \sigma_i} \sum \frac{z_i}{\sigma_i}$$

Sum over cues

Confidence of cue i

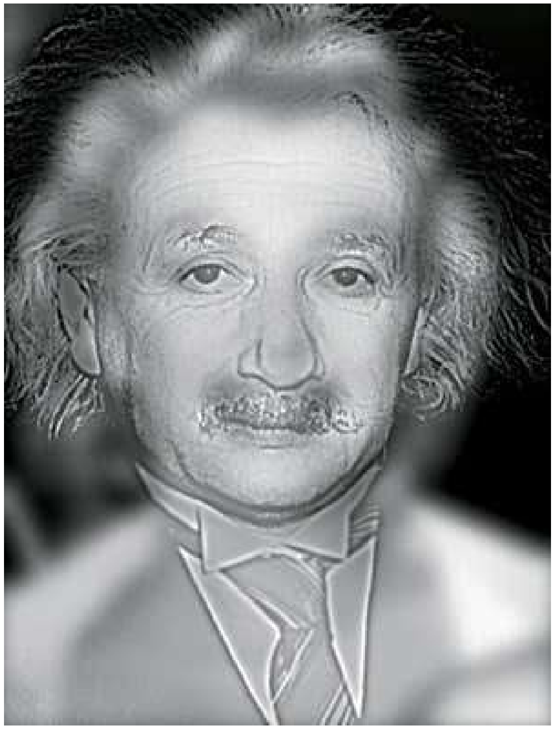
Gist (Olivia 1995)

- Show an image of a category extremely quickly
- Humans can say what it is, like landscape vs city
- Mostly texture perception



Hybrid images (Olivia 1995)

- Image that look like scene A from far
- And like image B from a near
- Produced by choosing optimal frequencies for that distance



Literature

- Landy, Michael S., et al. "**Measurement and modeling of depth cue combination: in defense of weak fusion**." *Vision research* 35.3 (1995): 389-412.
- Cutting, James E., and Peter M. Vishton. "**Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth.**" *Perception of space and motion*. Academic Press, 1995. 69-117.
- Oliva, Aude. "**Gist of the scene.**" *Neurobiology of attention. Academic press*, 2005. 251-256.

Neuro-physiological

- What is going on neurologically when we recognize?
- Pupil
- Retina
- Receptive fields
- LGN/Optical Chiasm
- Visual Cortex
- Bigger picture: Invariance

Pupil

- Controls how much light falls onto the retina
- Part of the adaptation

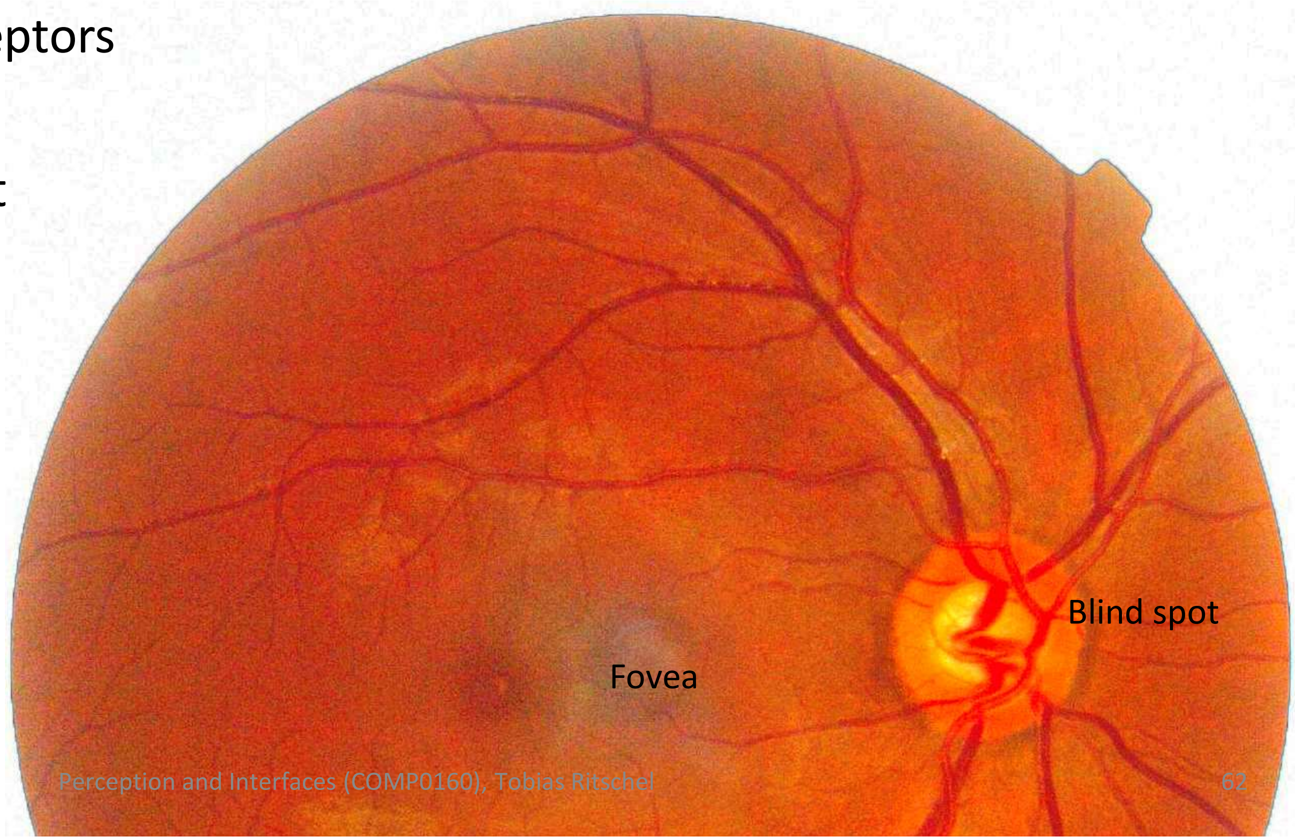
Large pupil
High field intensity

Medium pupil
Normal field intensity

Large pupil
Low field intensity

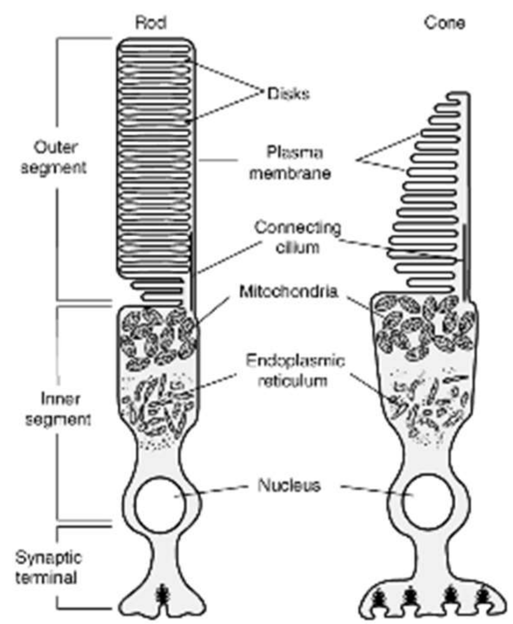
Retina

- Converts light into nerve impulses
- Covered by photoreceptors
- Denser in the fovea
- None in the blind spot



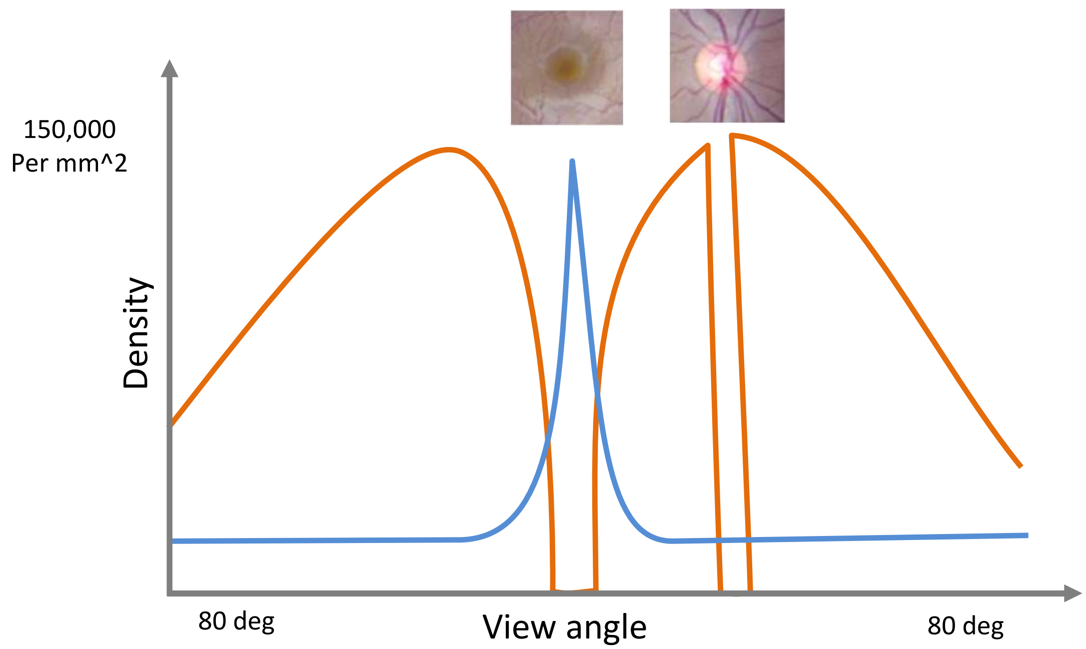
Photoreceptors

- Two kinds of photoreceptors
- **Rods**
  - Luminance
  - Day and night
- **Cones**
  - Color (so three kinds-of)
  - Day-only
- Ganglion cells
  - Not for image formation, circadian
- Adaptation in part done by flipping between these



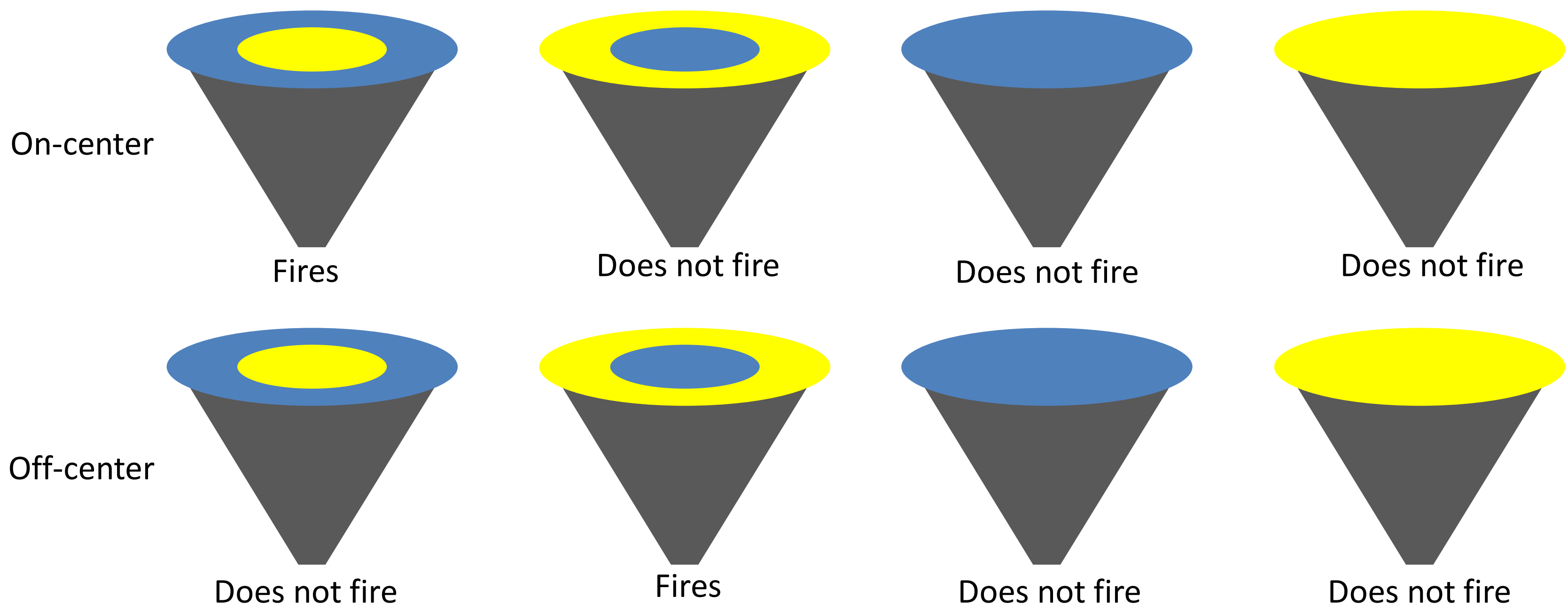
Foveation

- Physiological reason for foveal vision is receptor density



Receptive field

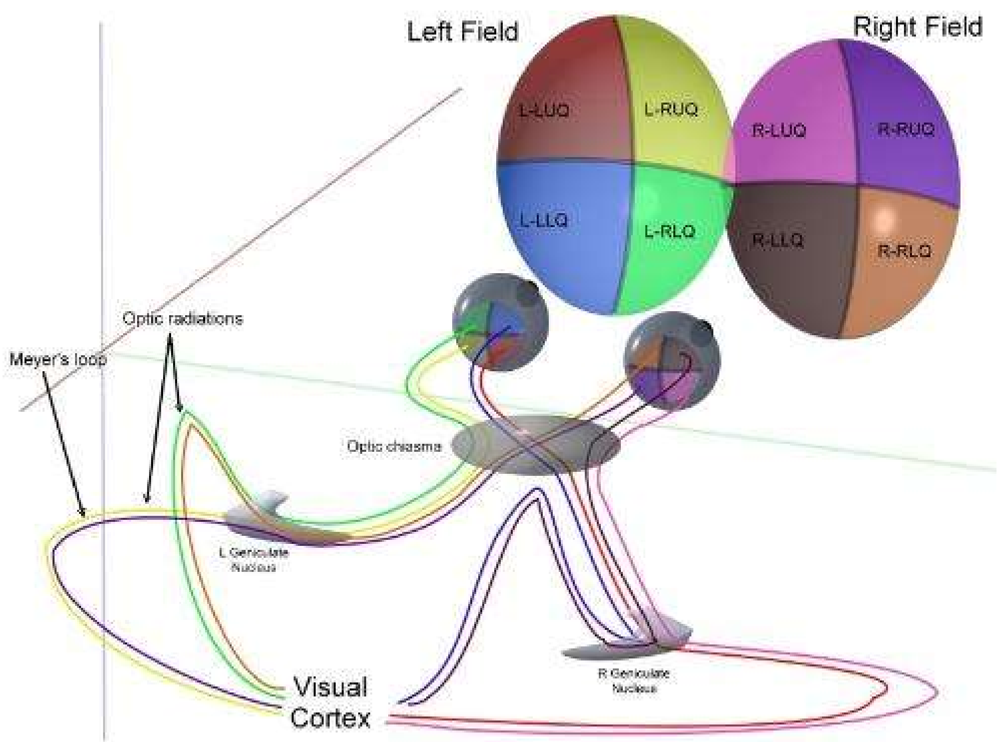
- Photoreceptors are combined on-site in the retina



No Sparsification: Eat much

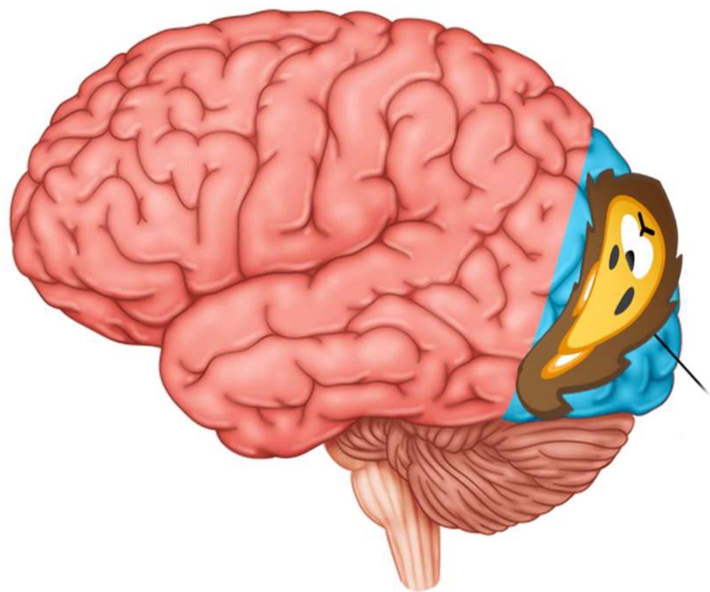
Sparsification: Eat little

LGN/Optical Chiasm



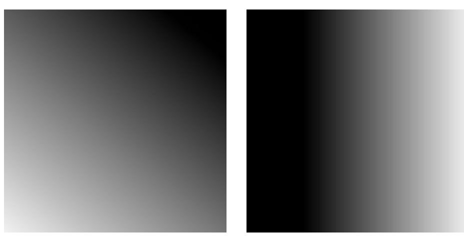
Visual cortex

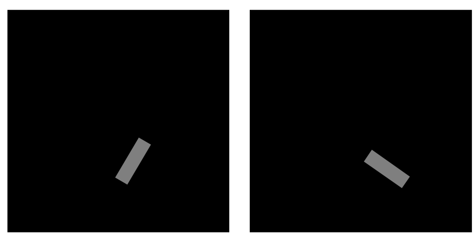
- Images literally get projected onto a part of the brain

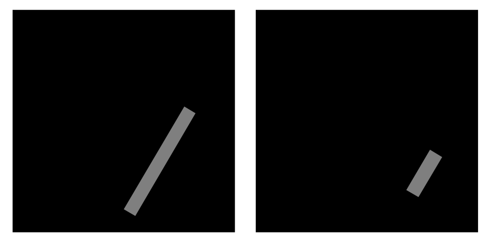


Cortical receptive fields

- Cortex selects frequencies and orientations of patterns
- Three levels
  - Simple
  - Complex
  - Hypercomplex

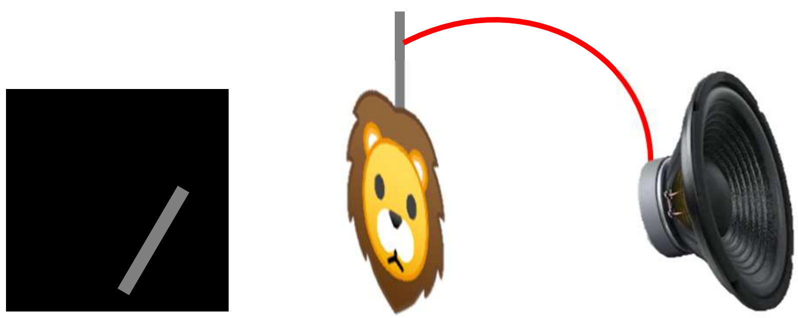






Experiment (Hubel & Wiesel, 1959)

- Important experiment
  - Anesthesized cat
  - Looks at a screen
  - Electrode capture cortical activity
  - Connected to loudspeaker



Cortical map (Virsu and Romavo 1979)

- (Foveal) areas with higher receptor density are represented larger in the cortex

Retinal
Light field

Cortical
representation

Literature

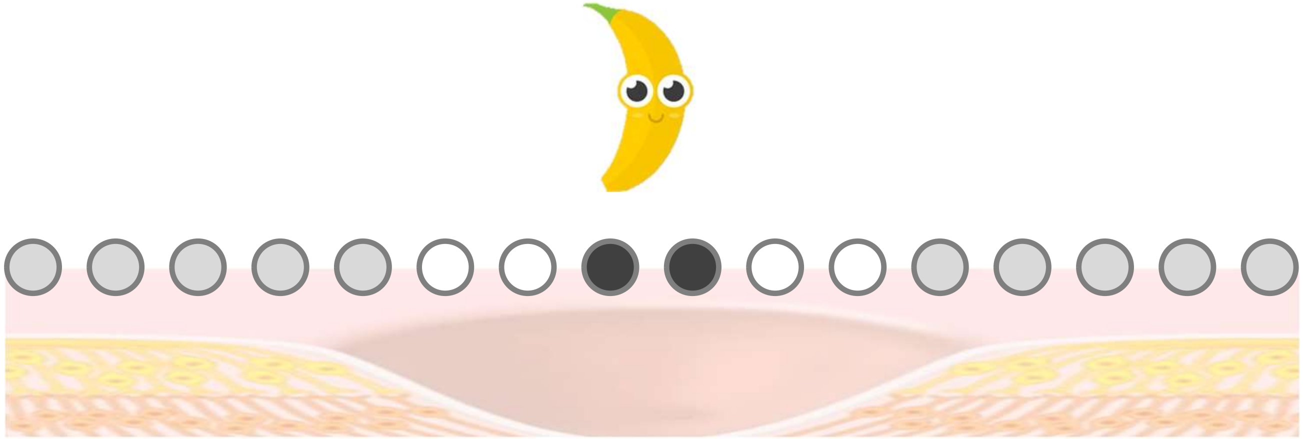
- Hubel, David H., and Torsten N. Wiesel. "**Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.**" *The Journal of physiology* 160.1 (1962): 106.

Computational

- How can I model these steps using a computer
- Neural network
- Cognitron
- Neocognitron (Pooling)
- Convolutional neural network

Simplification

- Consider a simplification:
  - 16 photoreceptors in 1D
  - Monocular
  - Monochromatic

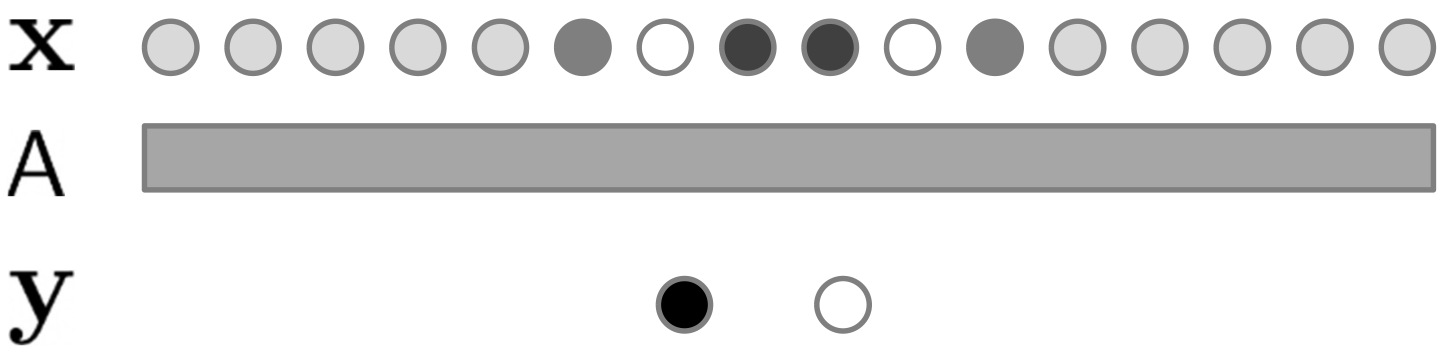


Simplification

- Simplification:
  - 16 photoreceptors in 1D
  - Monocular
  - Monochromatic

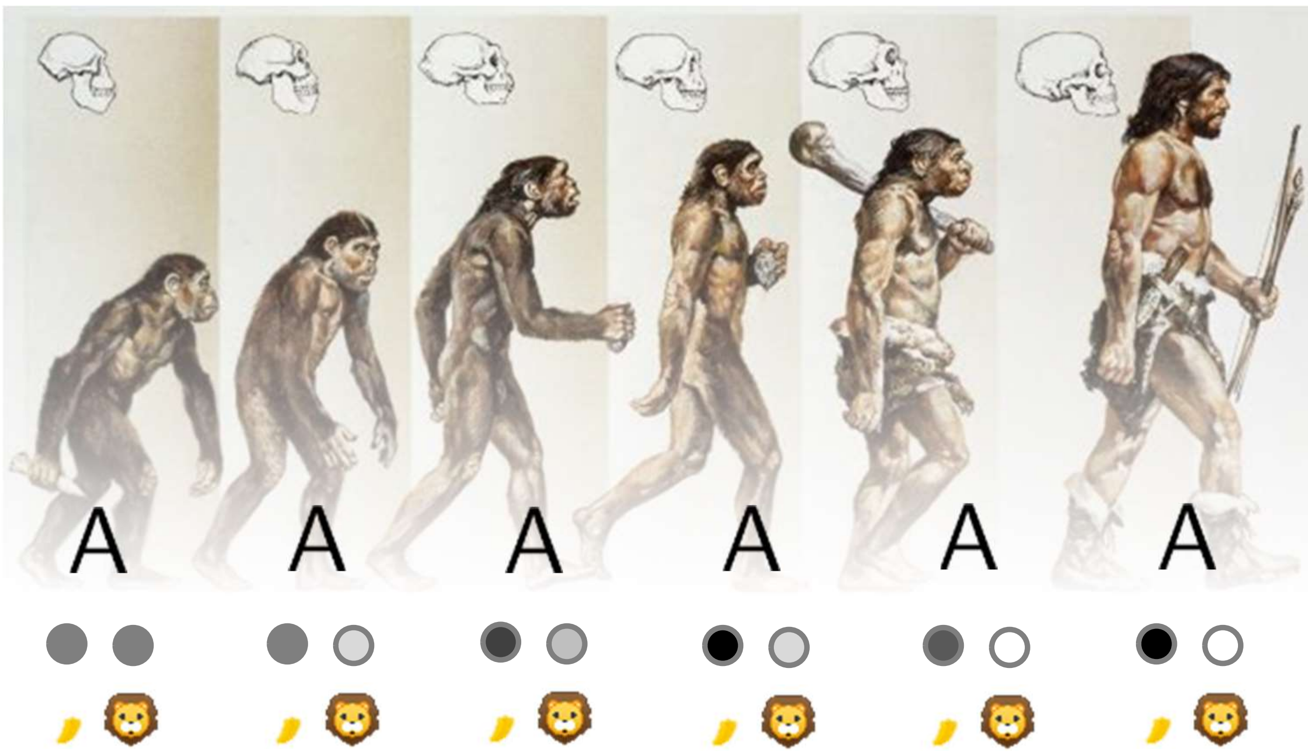
Neural network

- Input is vector
- Output is vector
- Vector-to-vector is matrix



- Probability of being a ….

Evoluion / learning



Evolution / learning

- Evolution or learning optimizes the relation of input and output over all items
- The target is unknown, but if a solution does not meet it, it will reproduce less

$$\operatorname{argmin}_{\mathsf{A}} \sum_i ||\mathbf{y}_i - \mathsf{A}\mathbf{x}_i||$$

Backprop (Rummelhart et al. 1986)

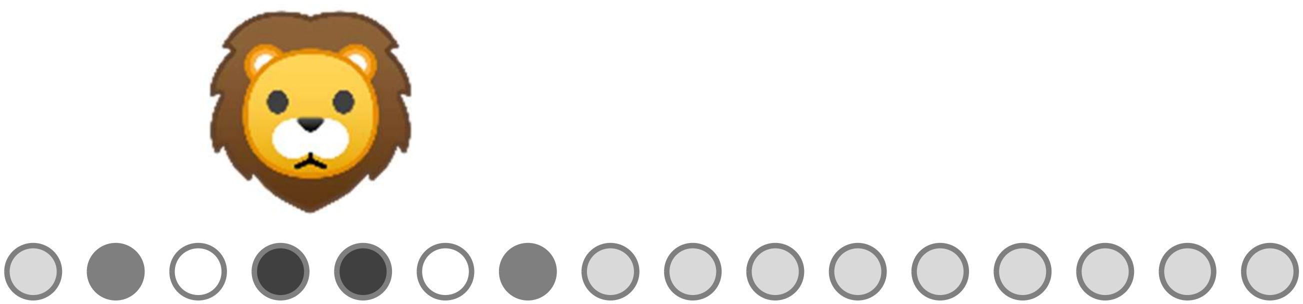
- If you have any function with parameters
- And you know for every input what the output should be

```
For all inputs
 Pass input through the function
 Compute difference between desired and current result (loss)
 Change the parameters so that the loss is reduced
```

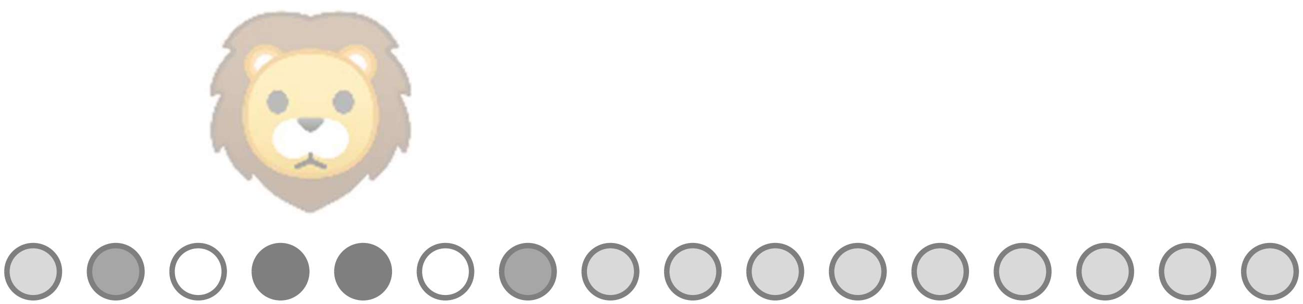
- This is not biologically plausible
- This is evolutionary plausible

Grandmother / lion cell

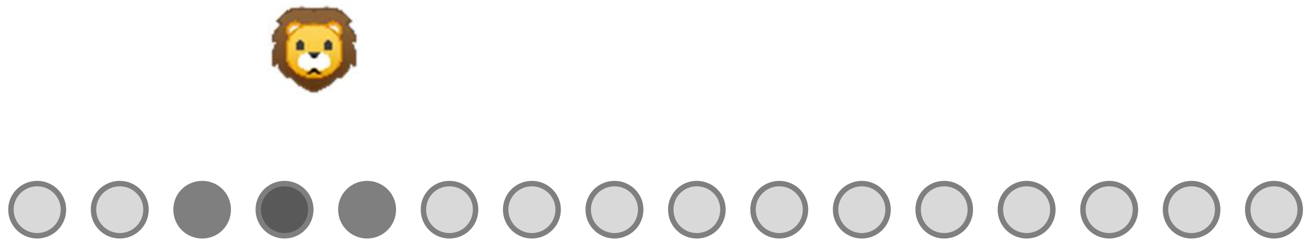
Grandmother / lion cell: Translation



Grandmother / lion cell: Brightness

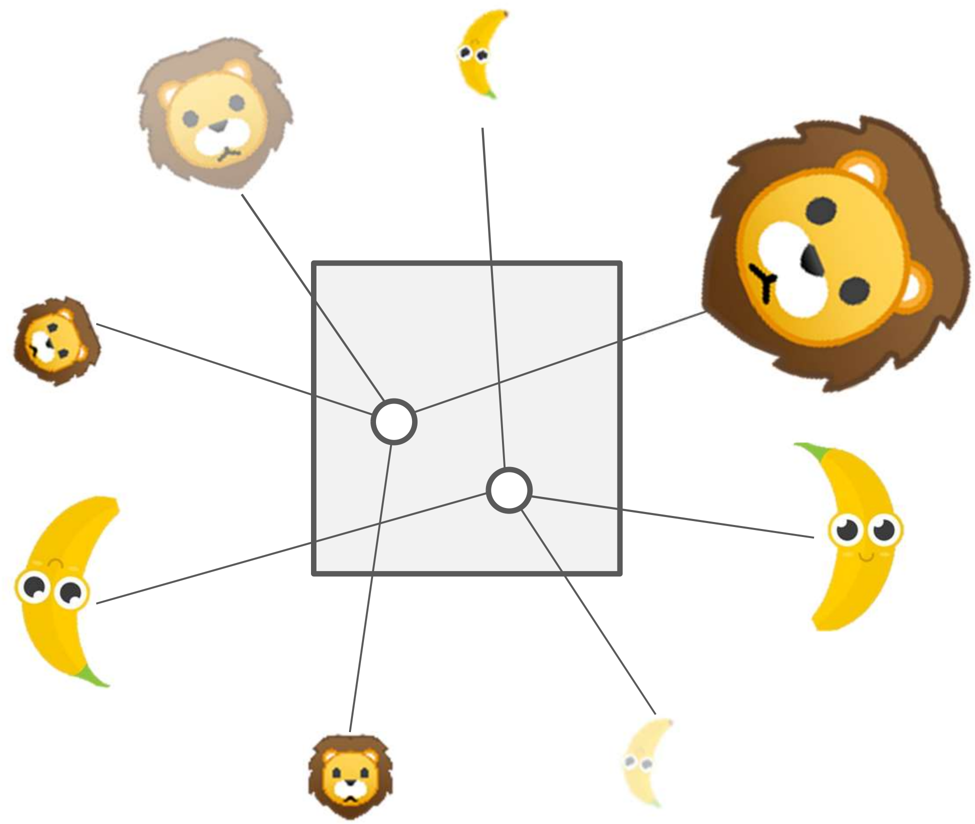


Grandmother / lion cell: Scale



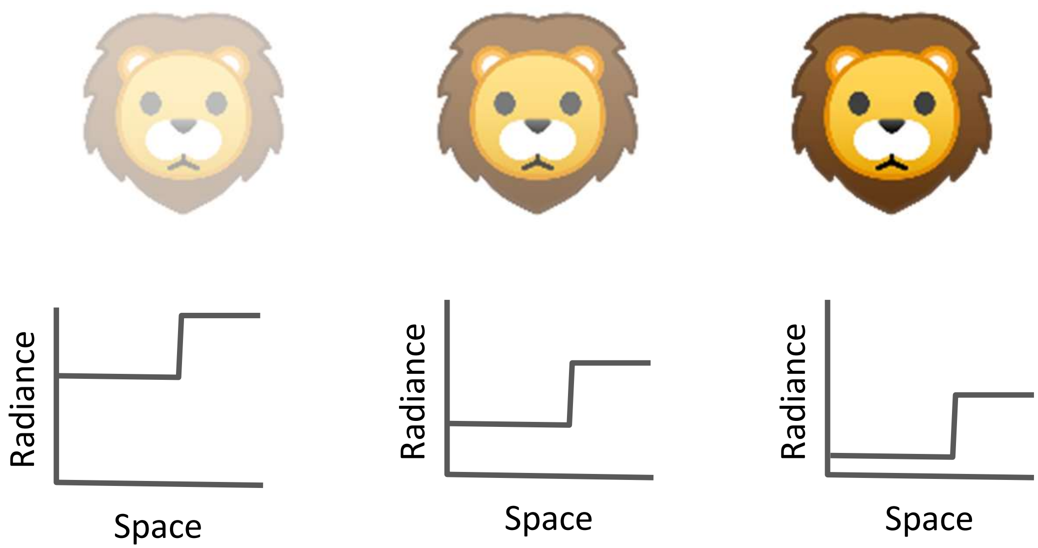
Invariance

- Problem: Response not invariant under transformations
  - Translation
  - Scale
  - Rotation
  - Perspective
  - Brightness
  - Etc
- Solution
  - HVS is a fat complex mapping to produce this invariance



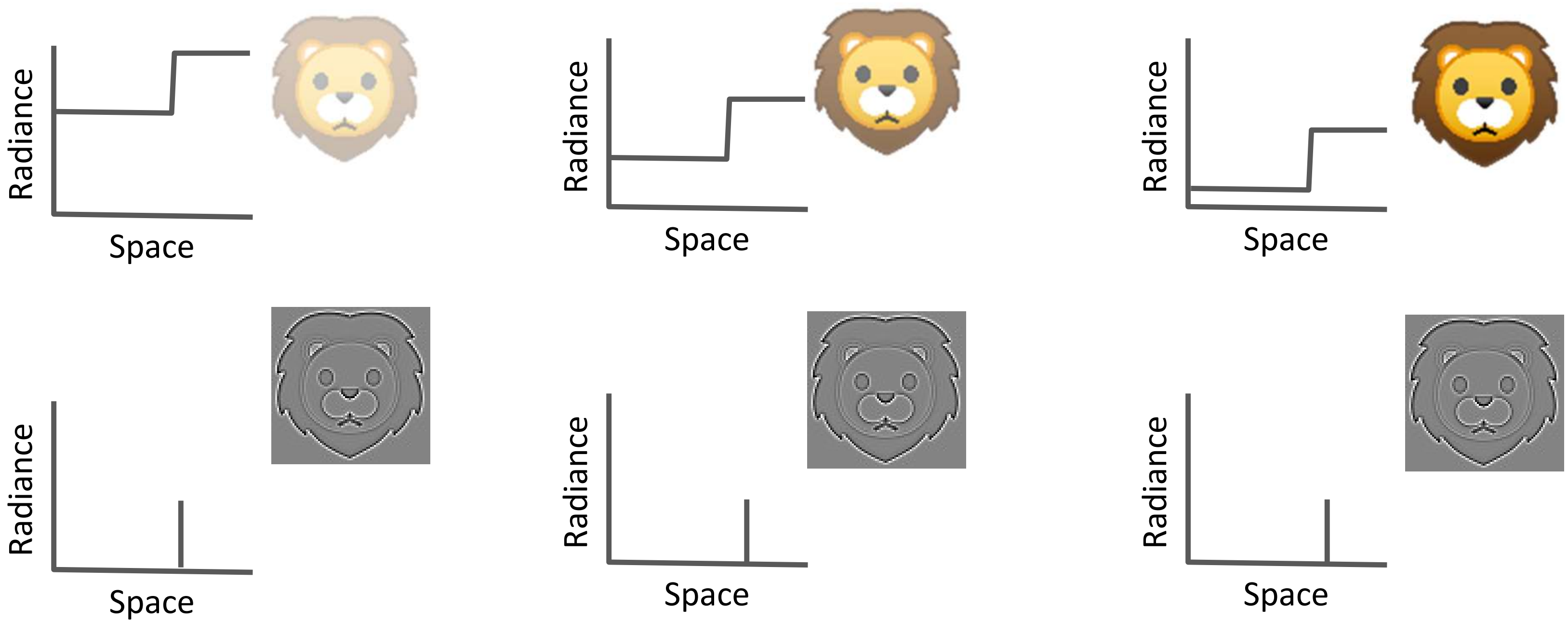
Simple example: brightness

- All these three predators should have the same response



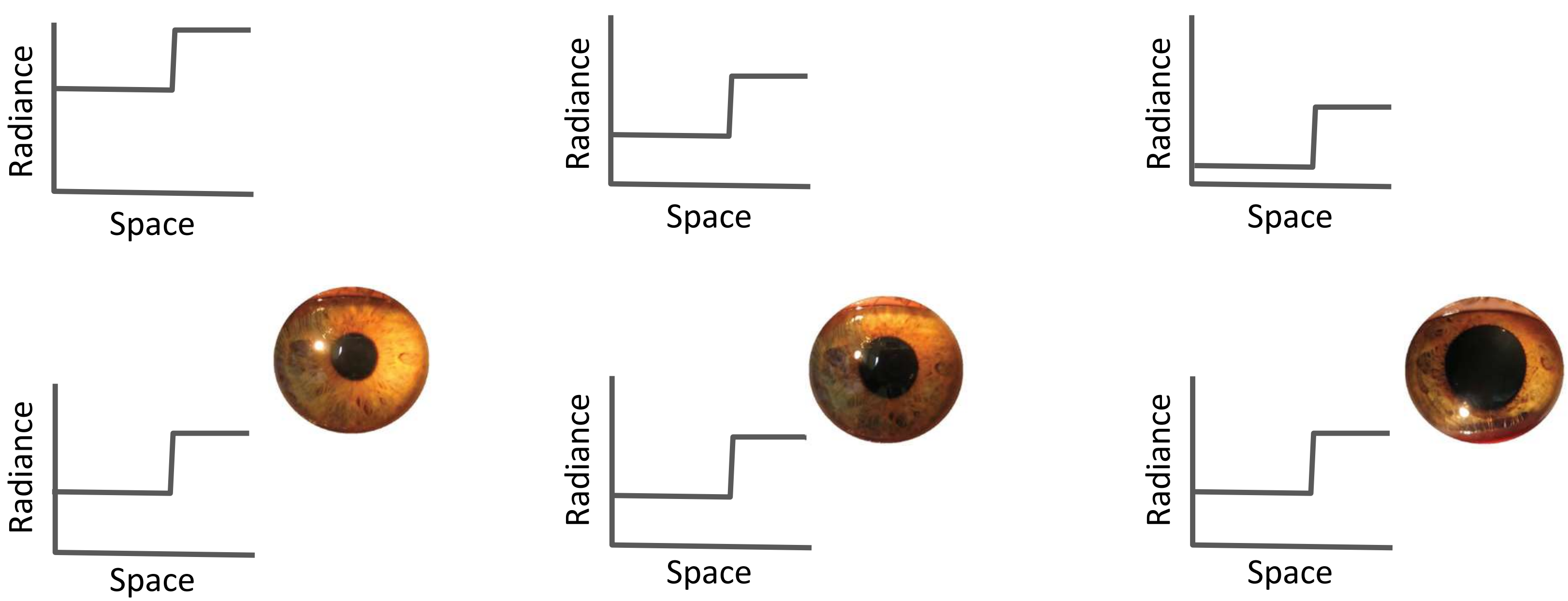
Solution 1: Edge filters
(Marr and Hildreth, 1980)

- Edge filtering alone already does this



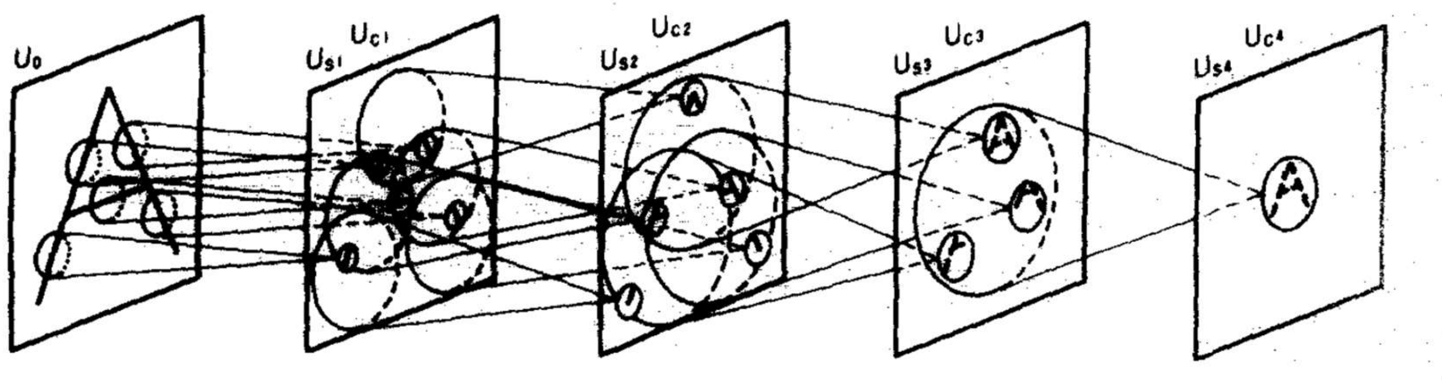
Solution 2: Adaptation

- Simply changing the pupil achieves this



Neocognitron (Fukushima, 1982)

- Idea: Pooling
- In a spatial area, **count** how often a feature is present
- A-example:
  - Check if things are present in a combination
  - Don’t care so much, where exactly



Convolutional neural networks (CNNs)

- Training it with convolutions (LeCun et al. 1989)
- Stacking many such things (Krizhevsky et al. 2012)
- Use multiple resolutions (Ronneberger et al. 2015)

Literature

- Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. **"Learning representations by back-propagating errors**." *Nature* 323.6088 (1986): 533-536.
- Fukushima, Kunihiko, and Sei Miyake. "**Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition.**" *Competition and cooperation in neural nets*. Springer, Berlin, Heidelberg, 1982. 267-285.
- Marr, David, and Ellen Hildreth. "**Theory of edge detection.**" Proceedings of the Royal Society of London. Series B. Biological Sciences 207.1167 (1980): 187-217.
- LeCun, Yann, et al. "**Handwritten digit recognition with a back-propagation network.**" *Advances in neural information processing systems* (1989).
- Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "**U-net: Convolutional networks for biomedical image segmentation.**" *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.