Ch

1 Introduction

2 Human Eye

3 Color 1 & 2

4 Texture Perception

5 Motion Perception

6 Depth Perception
- i) Brain
- ii) Neuroimaging

7 Grouping

8 Object detection & Scene Perception

9 Lesions & simulation

10 Face perception

11 Attention

12 Development
- i) Alignment
- ii) computational models
- iii) Dario Zanca

13 Computational models: Supervision, Generalisation & inductive bias

Paper 1

Paper 2

Exam
60 min

Prev Exam
60 min
22 questions
10 pages

Introduction:

ML/computer vision

Build vision systems
which are data efficient
& compute rich percepts

Understand human perception
with computational models.

computational
cognitive
Neuroscience

CH-1 Human Eye

- similarities b/w eye & camera
- Low visual acuity
- effects of low visual acuity.

Important parts of the Eye

- sclera: white outermost layer
- choroid: intermediate "
- Retina: inner most " [rods & cones]
- optic nerve: blind spot
- lens controlled by ciliary muscle
- Lens
- Fovea
- pupil: dialate/constrict to adjust light entering.

- Fovea: darker due to high density of cones in this region
- optic disc: no vision in this area. blind spot placed b/w the eye near the nose.

Camera	Eye
- spectrum of light visible to both is different	
- no aging/fatigue	- affected by age/fatigue
- mono-vision	- stereo vision
- refresh rate can vary	- refresh rate varies b/w 30 to 60 FPS.
- could have different designs.	- fixed design (~120° wide)
- field of view (can vary)/resolution is different	
- uses lens	- uses lens
- adjustable focus	- adjustable focus
- RGB sensors	- 3 types of cones

Rods: for night vision [No color, only intensity]
- respond under low luminance
- low visual acuity

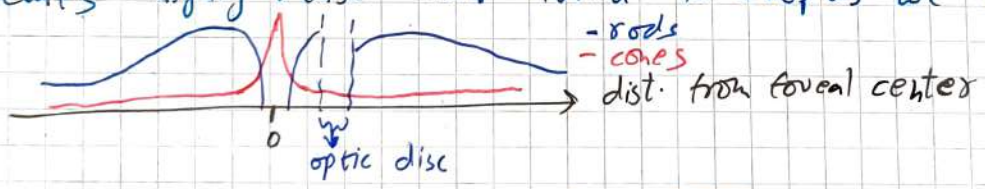
Cones: specialized for daylight vision [color]
- respond under high luminance
- high visual acuity

Ganglion neuron layer
↓
Bipolar neuron layer
↓
Rod & cone layer
↓
pigmented epithelium layer of retina
↓
choroid
↓
sclera

What is advantage of having
1) rods & cones at bottom
2) rods & cones at top
Which design is better?

	At bottom	At top
Adv	Protection, light filtering and optimization, longetivity of photo receptors efficient sig. processing	direct light capture w/o passing through other layers
disadv	scattering due to passing through other layers	more glare, increased oxidative damage

- Inhomogeneous photoreceptor distribution.
  Rods density high around fovea & drop as we move away
  Cones highly dense near fovea & drop as we move away.



- acuity = resolution / blurr of camera
- field of view: central → paracentral → near peripheral → mid peripheral → far peripheral (camera sensor)
- both eyes have same focal point.
- visual acuity: measure of the ability of the eye to distinguish shapes & details of objects at a given dist.
- hyper acuity: condition/state where vision is more than active allows eye to detect very small differences in spatial alignment / position, even beyond resolution of photoreceptors in retina.

allows to distinguish sub-pixel resolution

Why does hyperacuity happen?

- infants are born with low visual acuity & improves as they age.
- congenital cataract: lens opacity present at birth, surgery in 1st 10 weeks. No visual i/p until then

Computational Approach to study acuity

Hypothesis: low acuity leads to
- holistic representations (bigger receptive field)
- improved performance
- better generalisation

Modelling: CNN → curriculum learning.

Ch2 - Color

Pass white light through prism to get spectrum of light.

Difficulty in color science arives from perception, can be Physics but what matters most is how we pecieve color.

Color on diff. screens & print is similar but underlying mechanism is complex & diff.

Camera, screen capture/produce the spectrum.

Light is electromagnetic radiation occupying the visible band of the spectrum. Exists as oscillations of diff. freq/wavelength.

Radio - Microwave - IR - Visible - UV - X-Rays - Gamma rays

longer wavelength ← → shorter wavelength

lower freq. ← → higher freq

↳ range of visible radiation = light

Measuring light

- Spectral Power Distribution: how the energy is distributed over wavelength. (W/nm) → watts per nanometer

  Amount of light present at each wavelength.

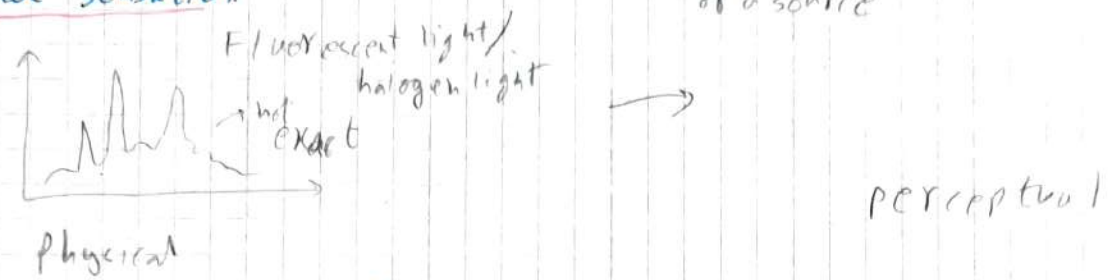
  Not really important as eye adjusts to diff. illumination conditions to pecieve color. (Sunny day vs Cloudy)

Color is a sensation that arise from light energy of diff. wavelengths. It is a phenomenon of human perception & not a universal property.

shades of gray have same spectral power throughout all wavelengths

if power is concentrated in some wavelength, then it is some color.

Problem of color science: Build a model for human color perception that maps physical light description to perceptual color sensation.
↳ given spectrum of a source



Eye = measurement device [low level behaviour modelled as a light-measuring device]. It measures light by **photoreceptors** in the retina.

The way we percieve color is not only through the eyes. Eye is only the begining. once it senses light & transforms it into neural impulses further **processing happens in the brain** & thats when **notion of color starts**.

HVS : Human Visual System [starts with eye & ends in brain]

Retina ≈ camera film

Rods & cones are types of **light receptive cells** [not evenly distr. over retina]
↳ sensitive to diff. wavelength of light

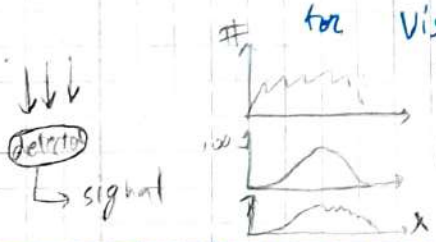
Rods: brightness sensing & are more sensitive.

Rods have faster response time than cones.

No rods & cones at **blind spot** where nerve tissue connects to the eye (optic disc).

<u>Simple light detector</u> / 1 camera pixel counts photons & **depending on wavelength it might miss a photon & not register it.**

- Produces **scalar value** when it lands on it
- each photon has **prob. of being detected** depending on wavelength.
- no way to tell diff. b/w signals caused by light of diff. wavelengths
- model works for semiconductors (in digital camera) or for visual photopigments (such as human eye).



$n(\lambda)$ → incident photons/unit wavelength

$p(\lambda)$ → detection efficiency

$X = \int n(\lambda)\, p(\lambda)\, d\lambda$ → detected photons/unit wavelength

- Same math for power distr.

$s(\lambda) \rightarrow$ spectral power distribution (SPD) of light entering the detector

$r(\lambda) \rightarrow$ spectral response/sensitivity of the detector

$$X = \int \underbrace{s(\lambda)}_{\text{i/p spectrum}} \underbrace{r(\lambda)}_{\text{sensitivity}} d\lambda$$

$X \rightarrow$ measured signal by the detector

$$X = \int s(\lambda) r(\lambda) d\lambda \quad \text{or} \quad x = s \cdot r$$

- discretise spectrum
- use vectors of s & r
- instead of integral use inner product. [elementwise multiplication & addition]

$$\int s(\lambda) r(\lambda) d\lambda \approx \sum_i \underbrace{\tilde{s}[i] \tilde{r}[i] \Delta\lambda}_{\text{dot product}}$$

Cones	Produce resp
S	r_S
M	r_M
L	r_L

$\rightarrow$ neural response when integrating SML curves with λ.

$\rightarrow$ results in tristimulus visual system

$\rightarrow$ color sensitivities is nonlinear & not evenly distributed.

$\rightarrow$ more sensitive to red-ish / brown-is tones [evolutionary development]

$\rightarrow$ color blindness: can't tell diff. redish & brown tones
red & green

$\rightarrow$ codes for pigmentation are on X chromosomes, hence more chance of males (XY) having colorblindness as they have only 1 X chromosome.

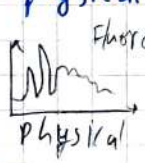
$\rightarrow$ Tetrachromats: women have 2 X chromosomes. & some have 4 sensitivity curves.

$\rightarrow$ metamerism

$$S = \int r_S(\lambda) s(\lambda) d\lambda = r_S \cdot s$$
$$M = \int r_M(\lambda) s(\lambda) d\lambda = r_M \cdot s$$
$$L = \int r_L(\lambda) s(\lambda) d\lambda = r_L \cdot s$$

Colorimetry: answer to problem of color science

- Maps physical light description to perceptual color sensation.



$\longrightarrow$ $S = r_S s$, $M = r_M s$, $L = r_L s$ perceptual

1) Colorimetry takes a spectrum (function)
2) Eye produces 3 numbers
Throwing away lot of information

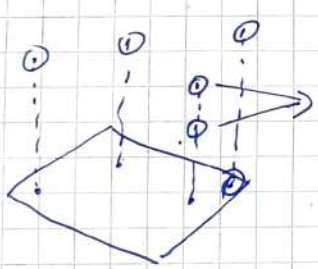
Possible to have 2 spectra that have same SML responses/ tristimulus values. The 2 spectra are called metamers.

Geometric interpretation

- dot product is a projection.
- projecting high dim. vector (spectrum) onto 3 vectors.
  Differences $\perp^r$ to all 3 vectors (i.e. in 4th dim) are not detectable

2 spectra that produce same resp are metamers.

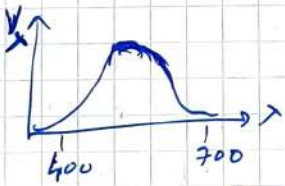
imaging 3D - i/p spectra with 2D representation



metamers → produce same resp in 2D space.

- Luminance: (brightness) overall magnitude of visual resp. to a spectrum

Luminous efficiency function V_λ describes eye's ability to detect light at each wavelength.



$Y = r_Y \cdot S$; $r_Y = V_\lambda$ V_λ → is linear comb. of S, M, L

- chromaticity: Whats left after luminance is factored out
  objective specification of quality of color regardless of luminance
  scaling spectrum up/down leaves chromaticity alone
- dominant wavelength: many colors created by white + spectral color.
  (hue)
- Purity: ratio of pure color to white in matching mixture
  [Colorfulness / saturation]

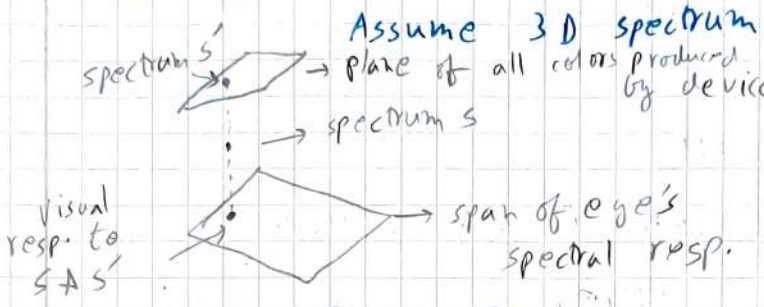
Color reproduction: make a spectrum S match to RGB monitor
Find spectrum that projects to same point in visual color space is a good reproduction.

- Find a spectrum that monitor can produce that is metamer of S.

Mixing colors

- Additive color : RGB $\quad R+G=Y$, $R+G+B=W$ $\quad$ Add energy in
- subtractive colors : take energy out

Color reproduction:



compute combination of r, g, b that project to same visual resp. as s.

$$\begin{bmatrix} S \\ M \\ L \end{bmatrix} = \begin{bmatrix} - r_S - \\ - r_M - \\ - r_L - \end{bmatrix} \begin{bmatrix} | \\ s \\ | \end{bmatrix} \quad \text{or} \quad V = M_{SML} \cdot s$$

Color reproduction as linear algebra

$$s_a(\lambda) = R\, s_r(\lambda) + G\, s_g(\lambda) + B\, s_b(\lambda)$$

$$\begin{bmatrix} | \\ s_a \\ | \end{bmatrix} = \begin{bmatrix} | & | & | \\ s_R & s_G & s_B \\ | & | & | \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad \text{or} \quad s_a = M_{RGB}\, C$$

$\rightarrow$ color

s_a is metamer of s.

given a color on screen how does a human see it?

- feed C to display
- display produces s_a
- eye looks at s_a & produces V

$$V = M_{SML}\, M_{RGB}\, C$$

$$\begin{bmatrix} S \\ M \\ L \end{bmatrix} = \begin{bmatrix} r_S s_R & r_S s_G & r_S s_B \\ r_M s_R & r_M s_G & r_M s_B \\ r_L \cdot s_R & r_L s_G & r_L s_B \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

goal: $M_{SML}\, \tilde{s} = M_{SML}\, \tilde{s}_a$

$$\Rightarrow M_{SML}\, \tilde{s} = M_{SML}\, M_{RGB}\, C$$

$$C = \underbrace{(M_{SML}\, M_{RGB})^{-1}\, M_{SML}}_{\text{color matching matrix for RGB}}\, \tilde{s}$$

Subtractive color: CMY K [water colors]

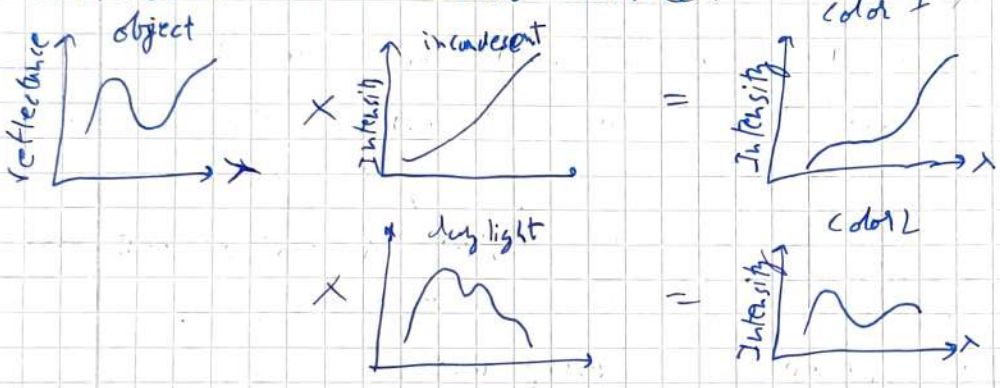
Primaries Color-2

Remove light from spectrum.

- printing
- absorb all other colors other than what is reflected.
  Ex: red chalk reflects red & absorbs G & B.
- spectrum that eye sees depends on what light is shined on the object as well.

Incandescent light → tungsten filament

Reflection from colored surface:



Color constancy: Brain compensates for illumination conditions.
A color remains the same

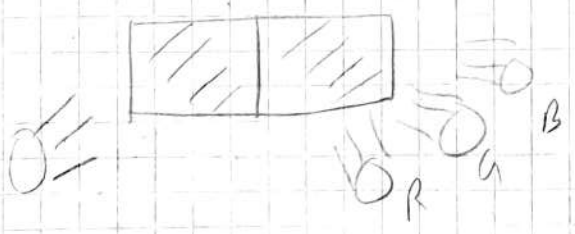
- produce color / desired spectrum by substracting from white light. (via absorption by pigments)
- Approximately 1-R, 1-G, 1-B

Color spaces: - Need 3 numbers to specify color
What are these 3 numbers?

Gamut: range of colors produced by mixing 3 diff. colors.

Tristimulus color matching Experiment:

Observer views a bipartite field & adjusts intensities of 3 primary lights to match appearance of test light



- if a dial reaches 0 & still more of that light needs to be removed then that light is added to the test light.
  This is the -ve amount we refer to.

Universal color space: X Y Z [standardise by CIE]

- based on 3 imaginary primaries X, Y & Z

$$S = XX + YY + ZZ$$

- imaginary → realisable by spectra that are -ve at some wavelengths
- Properties · Any stimulus can be matched with +ve X, Y, Z
- separates out luminance : X, Z have 0 luminance, so Y tells lumin. by itself.

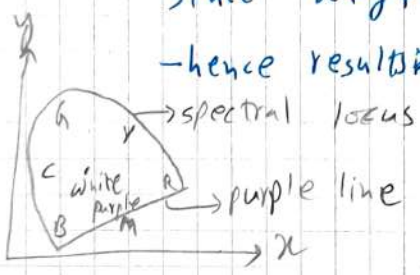
Advantages ① ② uses only +ve values

luminance : Y

chromaticity $x = \frac{X}{X+Y+Z}$; $y = \frac{Y}{X+Y+Z}$ $z = \frac{Z}{X+Y+Z}$

since $x+y+z = 1$ we only need to record 2 of 3.

- hence results in (x, y, Y) coordinates.
  - Y → non-normalised



→ monitors/printers cant produce all visible colors

Standardized RGB (sRGB) [mostly used]

- makes monitor RGB standard
- other color devices simulate that monitor by calibration
- sRGB space usable as interchange space
- gamut is still limited
- defines nonlinear distortion of how RGB coeff. are mapped to 8 bit values
- gamma has to be undone 2.2 ⇒ $\gamma = 2.2$

Assume D65 illumination, look at what is spectrum is reflected & then write down RGB values.

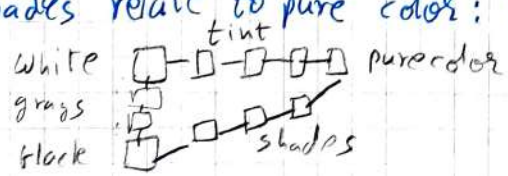
- Apple's true tone display uses camera to detect illumination & adjusts display/colors.

Perceptually organised color spaces

Artists refer to color as tints, shades & tones of pure pigments

- tint: mixture with white
- shade: mixture with black
- tones: mixture with GRAY
- gray: no color at all

- tints & shades relate to pure color: same color but lighter/darker/paler

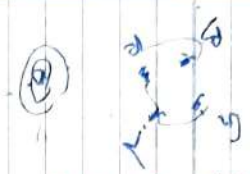


→ refers to the dominant color.

Hue: kind of color regardless of attributes

colorimetric correlate: dominant wavelength

artist " : chosen pigment color



opponent color theory
- mixture of opposite side colors isn't found

Saturation: colorfulness

colorimetric correlate: purity

artist " : fraction of paint from colored tube

Lightness: overall amount of light

colorimetric correlate: luminance

artist " : tint are lighter & shades are darker

- in x, y, Y space Y = lightness
- hue & saturation are then **polar coordinates** for chromaticity.
  [starting at white & go in dir. of hue]
  ↳ how far we go depends on purity/saturation.

<u>Perceptual dim. of color</u>

- brain seems to encode color early using 3 axes

  white-black, red-green, yellow-blue

  This ax is is lightness

  these 2 axes are hue & saturation

opponent color theory

B R
G — Y

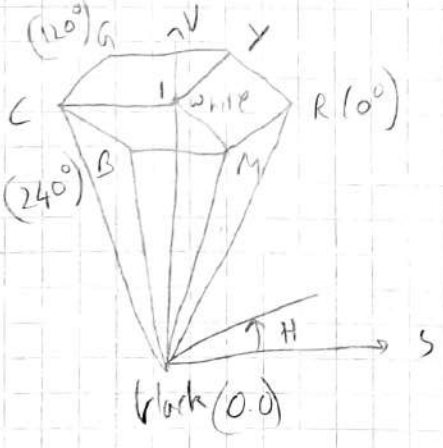
we see yellow-green, reddish-yellow but never yellowish blue or reddish green

- evidence 1: we can have light green, dark green, yelow green or blue green but never reddish green.
- evidence 2: after images [sig. from retina is transfered to visual cortex through nerves, by encoding different signals, like a basis tx before it is transmitted]

RGB as a 3D space

Cube

RGB to HSV



hue: uses angle 0 to 360

saturation: 0 to 1

value : 0 to 1

- LAB space: more uniform

Ch 3 - Texture Perception

Summary

- textures help perceive objects, materials & shapes
  (Branding examples) ↳ properties of material
- Texture is crutial for distinguishing objects from background
- in some cases texture is more significant than color for object differentiation. [object segmentation]
- subtle changes in texture help define obj. boundaries ↗ or texture boundaries
- our visual system processes texture differences for obj segmentation.
- Attention affects texture recognition & how easily we detect boundaries

Visual system & Processing of texture

- hemisphere covering parts of visual field, not 1 eye.
  Left field → right brain Right field → left brain

- Retina processing: compress information before sending it to the brain. Processing also enhances contrast for better perception
- Visual Cortex (V1): area of cerebral cortex that processes visual information. Hubel & Wiesel's experiments on cats showed how neurons respond to different orientation.
- Eccentricity in V1 mapping: visual system magnifies central vision more than peripheral vision. Foveated region has more resolution.

After effects in Visual Perception

- tilt after effect: viewing tilted lines can make straight lines appear tilted afterwards.
- Spatial freq. channels: Brain processes textures using different spatial freq. filters

Fourier Tx in Visual Perception

↗ 2 imgs embedded in 1 depending on scaling we see either
→ neurons looking for freq. & not edges

- Visual signals decomposed into sine waves.
- Brain uses spatial freq. analysis to decode textures.

Feature detection & Neural Processing

2 types of cells Hubel & Wiesel found out

- simple cells detect lines & edges
- complex " " / integrate info. to process textures

→ sensitive to some freq. than others
→ can be changed after staring for some time at a freq.

Modern research suggests interconnected processing rather than just feed-fwd networks

Hubel & Wiesel experiment [electrode in V1 of a cat]

- projected patterns of light & dark onto screen while [stimulus] recording neural responses in the visual cortex
- tried light spots initially but got weak responses
- discovered moving edges & bars of light caused stronger neuron activation.
- Feature detection: Neurons in V1 respond selectively to specific orientations of edges & lines in visual field
- Lead to discovery of orientation-selective neurons.
- simple cells respond to lines at particular orientation in a specific location.
- complex cells respond to oriented lines anywhere in the larger receptive fields.
- Hierarchical processing: retina & lateral geniulate nucleus (LGN) detect simple patterns, which are further processed in V1 to detect more complex features.

Feature visualisation

[how neural n/w see & process features]

- neural n/w break down images into simple patterns (edges, textures & shapes) similar to how V1 processes visual i/p.
- Feature visualization: is a technique to what specific neurons in ANN respond to by optimising images that strongly activate those neurons. Helps visualise what each layer of DNN detects.
- insights: early layers = edges & textures
  mid layers = shapes & patterns
  final " = specific obj. & concepts
- similar to Hubel & Wiesel's exp. neurons respond to oriented edges, artificial neurons in CNNs detect similar patterns at diff. processing levels.
- improves AI interpretability making deep learning less of a black box.

Hallucinations & Psychedelic visual experience

- Psychedelics affect visual cortex, disrupting normal perception
- hallucinations are not random: involve geometric patterns, textures & enhanced color perception
- increased activity in V1 & V4 (color processing areas) enhance texture & pattern recognition.
- hallucinations resemble naturally occuring visual processing, suggesting that perception is construct of the brain rather than a direct reflection of reality.

Neural adaption: is a gradual decrease over time in responsiveness of the sensory system to a constant stimulus. [tilt aftereffect]

Attention / focus: is concentration of awareness on some phenomenon to the exclusion of other stimuli.

- Different features & way we process make some texture boundaries to ~~process~~ detect difficult.
  - visual field ends up on diff. parts of the eye, through LGN & then to primal visual cortex.

Retinotopy: is mapping of visual i/p from retina to neurons, particularly those within the visual system.

- MAE: **motion after effect**
  is a visual illution experienced after viewing a moving visual stimulus for a time with stationary eyes & then fixating on a stationary stimulus.
- Spatial freq. channels: we percieve lines with certain width & mag. more easily than others.
  can become bad at recognising freq. after long exposure due to neural fatigue.
  - **spike triggered average** (STA): brute forcing patterns to neurons. show image to cat & record if neurons fire/not. Weight image by how strong firing was & then take avg. of all images where firing was more.
    Weight it by no. of neurons which fired.
  - results on Gabor filters [averaged images]

- Retinotopy: mapping of visual i/p from retina to neurons in visual system
- cortical magnification
- V1 preserves spatial properties; used for texture perception.

Eccentricity: means distance from fovea

Ch4 - Motion Perception

Summary

- Motion perception (Why is motion important) is crucial for visual understanding
- helps in object grouping, 3D perception & self movement awareness

Types of Motion perception

- Apparent motion:

- **Real motion**: physical movement of objects or self
- **Induced motion**: illusion of movement due to background movement. Ex neighbor train moving
- **Autokinetic motion**: small involuntary eye movement creating motion illusion. Ex: when he looked star & he feelit is moving.
- **Motion aftereffects**: Perception of motion after prolonged exposure to moving stimuli. - **Apparant motion**: Rings seen to move but not moving (Ex: movies).
- Motion helps in detecting hidden objects and to group.
  Ex: bird hidden in random lines becomes visible when it moves.
- brain interpolates missing frames creating smooth perception.

Motion parallax & depth perception

- **Motion parallax**: Nearby objects move faster than distant objects, helping depth perception.
  Ex: when moving in a train close trees pass quickly, while distant mountains move slowly.
- Motion allows better understanding of spatial relationships.

Self-movement Perception

- pilots use motion cues to navigate (Ex: optical flow where all points converge toward a goal) → Perception of own movement
  Experiment: moving a room while a person stands still creates illusion of movement, affecting balance.

Visual perception & motion interpretation

- different interpretation of same stimuli (Ex: blinking dots perceived as circular or linear movement) → solve global movement over local movement → complex movement is combination of local movement
- brain uses context & prior knowledge to interpret motion.

Motion detection mechanisms

- **Reichardt detector**: neural model detecting motion dir. & speed.
  Limitations: needs a 2nd mechanism to filter out stationary signals.

Motion detection in the brain

- V1 neurons: sensitive to motion direction and speed
- **Aperture problem**: local motion detection may misinterpret overall movement
- **Optical flow**: Tracks movement by analyzing pixel displacement over time.

→ sparse optical flow: tracks key-points
→ Dense " " : Analyses all pixels for motion detection

Brain Processing of Motion: From V1 [visual cortex there are 2 streams dorsal & ventral]

- Dorsal stream ("Where"): processes motion perception [finds representation of the scene]
- Ventral stream ("what"): identifies objects
- MT & MST areas: critical for motion processing
  → dorsal stream

Eye movements

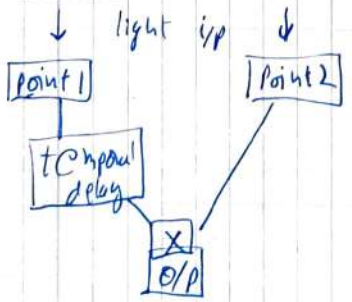
- Saccades: fast eye jumps from 1 ing location to other; occur ~3-4 times per second
- micro-saccades: tiny involuntary eye movements. gives more resolution & helps 3D perception
  when we try to fixate at a point
- smooth pursuit: eyes follow moving objects but cannot smoothly track without stimulus
- saccadic masking: Brain ignores motion during saccades
  to avoid motion blurr
  Ex: cannot see eye movements in a mirror

Motion perception & brain damage

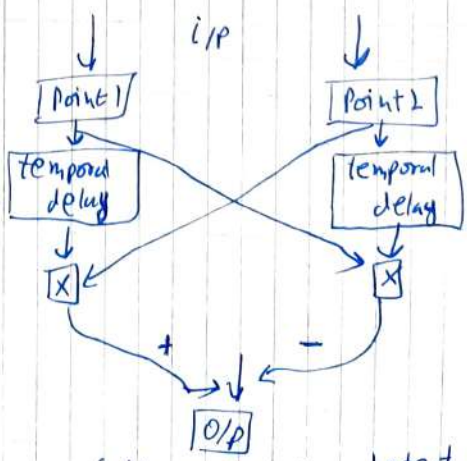
- Efference copy: brain predicts eye movement & ignores visual input during movement. As the brain itself controls movement of the eye
  Experiment: poking eye creates an illusion of motion, providing efference copy mechanism.
- Brain damage in MT/MST regions can cause motion perception disorders Ex: feeling const. motion.
- Optical flow, motion parallax & neural mechanisms contribute to accurate motion perception.

MT → Middle Temporal Area [also called V5] → motion dir & speed
MST → Medial Superior Temporal Area → optical flow processing
Depth perception: ability to perceive the world as a 3D space.

Reichhardt detector



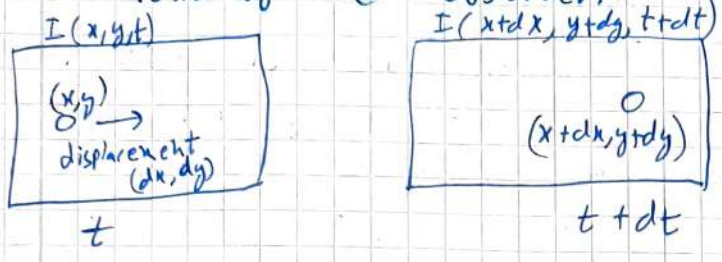
Point 1 → Point 2
& specific to temporal delay
else wont be able to detect motion.



- full sol. to detect no motion
- checks if both points are blinking & there is no motion

Aperture problem arises due to limited receptive field.

optical flow : pattern of motion in a visual scene due to relative motion of the observer. & camera.



take a step in time & see by how much a pixel/point moved.

- flow Net: estimates the optical flow
  takes old image, new image & outputs image with the flow. or how pixels move.
  Can be modelled with a U-net.
  encoder with skip connections.

Motion constancy: ~~points closer appear to move faster~~ ?
points appear to be moving at same speed but in reality they move at diff. speeds

How is motion removed when eye is moved? [saccadic masking]
- brain moves eye so it ignores visual i/p when it knows that it is moving the eye
- CV way: optical flow & correct

efference copy: internal copy of outflowing movement producing sig. generated by organism's motor system.
↳ cone b/w MT / MST.

damage in this region causes const. motion & motion sickness in such patients

→ optical flow looks only for apparent motion.

Ch5- Depth Perception

- Depth perception helps us understand spatial relationship between objects.
- Motion parallax also plays role in depth cues.
- Ex of depth perception: objects closer to observer appear to move faster than distant ones. Ex: when in a train nearby cows appear to pass faster while distant mountains stay in view longer.
  Similarly when walking in a room foreground objects move faster, than background objects

Challanges in Depth perception

- 2D images on retina lack inherent depth information.
- Ambiguity arises when interpreting depth from single perspective.
- Ex: cube in 2D image can have multiple valid depth interpretations
- Facial images also present depth ambiguities, seen in variations in nose length in selfies.

Ames room illusion

- Ames room manipulates depth perception using distorted architecture.
- Is a specially designed room, make people appear to shrink or grow as they move within it.

Art & depth perception

- Artists use perspective and shading to create 3D illusions on flat surfaces.
- These illusions rely on fixed viewpoint; from other angles, images look distorted.

Depth perception & visual illusions: → our brain has priors

- our brain prefers simpler depth solutions, even if incorrect
- monocular and binocular depth cues contribute differently to depth perception.

Inference

- our brains use prior experiences to infer depth.
- objects are often interpreted on learned expectations

Ex: people we meet online may seem diff. in person due to assumptions made based on limited visual cues.

Binocular cues

- Binocular disparity: each eye sees slightly different image, helping judge depth.
- Horopter: region where objects appear at same depth for both eyes.
- Random dot stereograms: demonstrates how disparity alone can create a perception of depth. works with vergence

Divergence

- vergence (eye positioning): Angle between eyes adjusts to focus on near or distant objects.
- Vergence works only for short distances; at greater distances it becomes less effective.

Motion based & static depth cues

- Motion parallax: obj. move at diff. speed relative to observer
- Perspective & Ponzo illusion: ||le lines converging create an illusion of depth.
- size constancy: objects of same size appear different due to distance cues.

Shading & shadows for depth cues

- shading: light dir. influences depth perception.
- shadows: position & shape help determine objects height & placement.
  Ex: inverted shading can make a bump look like a dent.

Occlusion, blurr & atmospheric depth cues.

- occlusion: objects blocking others indicate depth relationships.
- blurr (tilt-shift effect): Focus changes create artificial depth perception.
- Atmospheric cues: distant objects appear hazier due to light scattering.
- occlusion is strongest depth cue in most situations
- stereo vision is useful within ~2 meters for precise tasks
  Ex: pouring liquid All world we can interact with needs real depth perception.
- film & photography techniques use forced perspective to create size illusions.
- Jig saw:

Which depth cue work at diff. depths?

[Horopter is a 3D surface in space where objects appear at same depth for both eyes, serves as a reference plane for judging depth based on binocular disparity, where obj in front/behind it appear to have diff. depths]

Ames room is a distorted room that creates an optical illusion.

- Inverse graphics engine also chooses simpler sol. to model 3D from 2D representation.

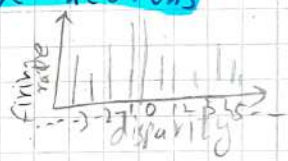
- Depth cues
  - Monocular
    - static
      - perspective
        - Linear
        - size of known object
        - Texture
        - height in the field of view
      - occlusion
      - light & shadows
      - Atmospheric
      - Accomodation
    - Dynamic (motion based cues)
  - Binocular
    - vergence
    - Retinal disparity

- Binocular disparity: cue to determine depth/distance of an object.

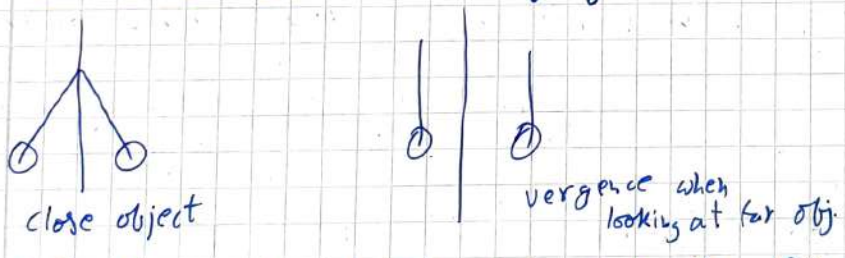
Points on Horopter have same distance on retina on both eyes.

How is this implemented? By disparity selective neurons

disparity → by how many degrees images on left/right eye differ.



Vergence: simultaneous movement of eyes in binocular vision.



works for nearby objects but not for far away objects, est. of vergence becomes difficult, beyond 5m.

Do binocular cues work on their own? Yes. Experiment: Random dot stereogram.

- dots shifted in 1 eye to the other.
- when this is seen it is percieved as 3D.

- Real depth perception: vergence & disparity (Binocular)

Ch-6 - Grouping

- Grouping is process of organising visual elements into meaningful structures.
- Relation to mid-level vision: takes local measurements from low level vision ↓ line receptive fields and combines them into more complex shapes for obj recognition.

Gestalt theory

- Gestalt : German term meaning pattern, form or shape.
- Concept: The whole is more than sum of its parts. Principle of totality.
- Gestalt psychologists focused on grouping rules in perception.
- The Gestalt laws of grouping
  1) Proximity: objects close to each other percieved as a group
  2) Similarity: objects sharing color, shape or texture tend to be grouped
  3) Closure: Mind fills in missing information to perceive complete objects.
  4) Symmetry: symmetrical elements are seen as belonging ~~to each other.~~ together
  5) Common fate: objects moving in same direction are grouped. what moves together belong together
  6) Continuity: eye follows smooth, continuous path rather than abrupt changes.

  Ex: European wedding dance trick where identical dress colors cause incorrect grouping of body parts. [Demonstrates power of similarity in grouping]
- Gestalt Principles explain how color, shape, closure, symmetry & motion influence perception.

  Ex: Bird swarms perceived as 1 entity due to common fate.
- Logos use Gestalt principles to create memorable designs.

  Ex: IBM logo uses continuity with evenly spaced lines
- Continuity v/s closure → gaps in objects are mentally filled in (if missing part is mentally completed)

  ↳ eyes follow a smooth path (if broken shape still appear continuous)

Neural basis for continuity & closure

- V1 neurons respond to line orientation & interact with neighboring neurons
- neurons suppress responses when conflicting sig. are present.

Visual processing layers

- V1 (primary visual cortex): detects simple edges & lines
- V2 (secondary visual cortex): processes complete shapes & illusions

- local association fields help perceive object continuity.

Visual processing in V2

- illusionary contours: perception of shapes without actual edges.
- Neurons in V2 fire in response to these imaginary contours.

Ex: Pac man illusion forming a Δ^{le}

Takeaways

- Grouping in vision helps organise perception
- Gestalt laws explain how elements are visually grouped.
- Neural basis for grouping occurs in V1 & V2.
- Visual processing extends from simple edge detection to complex object recognition.

Grouping is done by mid-level vision,

Mid-level vision involves taking local measurements made by receptive fields early in the visual system & combining them in various ways.

- Things similar tend to group together, appearing to be part of same thing.
- illusory contours or subjective contours are visual illusions that evoke perception of an edge w/o luminance/color change across the edge. [closure]
- in closure we dont follow any line/contour we follow.
- Association fields: neighbouring neurons communicate with each other. Integrates info across neighbouring filters tuned to similar orientation.

[change α, β & find diff. lines]

Low $\beta \rightarrow$ easy to find

high $\beta \rightarrow$ more curved line & diff. find

- Tanaka - IT neurons: tried to establish an alphabet

show diff. parts of diff objects & find neurons which are activated more/less

- we have neurons in V2 selective for illusionary contours
- 3D percepts also possible from illusionary contours.

Ch7 - Object detection & scene perception

- Focus is on object & scene perception transitioning from basic visual elements (lines & textures) to complete image perception.
- Brain processes images almost instantly, yet distinguishing objects can be complex.

Human visual Perception & Brain Processing

- EEG experiment: brain signals measured when participants see images.
- Finding: difference in brain signals appear ~150 ms after viewing an animal image.
- Processing stages: Retina → LGN → V1 → V2 → V4 → IT (Inferior Temporal cortex)
- Key insight: decision-making takes longer due to motor response delays, not visual processing speed.

Technological comparison & processing speed

- AI v/s humans: CNNs and AI models can process images faster
- Hardware factors: AI speed depends on GPU/FPGA implementations
- Limitations: AI may be faster but does not solve perception same way like humans do.

Challenges in visual processing

- Visual i/p complexity: brain processes millions of pixels to identify objects.
- Ambiguity: plane might look like a bird causing confusion
- Face recognition: Frontal v/s side-view faces appear different but belong to the same individual.
- Key challenge: AI struggles to move beyond pixel-based similarity to content based recognition.

Annotator performance on Image net

- Image-net, human annotators: 95% accuracy; 5% error
- Challenge: Humans struggle with fine-grained categorization (Ex: similar dog breeds)
- Key insight: AI surpasses human performance only in benchmarks, not in general object recognition.

CNN v/s Human Brain

- CNNs & Transformers are SOA
- Adv: AI models easier to study than the brain (no ethical concerns) (lesion studies easier on ANN)

Intro to visual recognition

- Humans recognise objects regardless of viewpoint/rotation. (invariant representation)

How does brain achieve this invariance?

- 2 theories of object recognition

i) View-based Models: brain stores multiple 2D images of an object

ii) 3D representation theory: brain builds a 3D structural model of objects.
» objects built from geons [3D building blocks for diff. objects]

- Debate: some evidence supports both models, but true mechanism is unknown.

Challenges in obj. recognition theories

- Mental rotation experiment: people struggle to recognize rotated objects.
- implies brain may not store full 3D models, challenging the 3D theory.
- view-based models require massive memory for storing all possible views. Experiment: rotated string of cubes identified by view based models but not by RBC [Recognition By components]

Experiment: recognise chairs not just visually but functionally (i.e. can we sit on them)

issue: AI & humans perception struggle with contextual recognition.

Complexity of visual perception

- context matters: shadows, occlusions & spatial relationships affect perception.
- Neural networks struggle with real world complexity beyond training data.

- Processing of image is different for an animal v/s non-animal
- V1: edges & contours
- V4: grouping
- IT: faces

Motion control:

PFC → PMC → MC

categories decision — motion control

- vision processing speed is shorter than motor responses.
- to tell individuals apart / to tell animal or not-animal we have to go away from image space to some representation where pose doesn't matter, but interested in label of image & not pixels.
- We are looking for this representation / space where we can distinguish objects.

Structural 3D representation

1) look at an image of an object
2) Use that image to describe what 3D shapes make up the object
3) While at it, describe how those parts go together
4) Now look at another image & if description matches then they are same thing.

Recognition by components: uses 3D - alphabet to form diff. objects. [Geons are individual alphabets]

view based models determining invariance prove above model wrong.

- humans perform bad based on how much the obj has to be rotated mentally.

view based models

1) learn about obj. by looking at them alot
2) store all images related to obj in album dedicated for it.
3) when new img is encountered check photo albums to see if there is anything that is similar to what you see now.
4) If decent match found, check which album it was in.

3D based schemes	View based schemes
strengths: 1) storage efficiency 2) view invariance	1. no reconstruction required 2. simple matching
weakness: 1) reconstruction req. 2) view invariance	1. huge memory req.

↳ more plausible as it explains many situations like with occlusions, shadows.

- to percieve the world it takes longer & takes more than 1 forward pass.

Ch 8 Lesions & stimulation

- Brain studies typically follow a sequence
  i) understanding diff. brain regions
  ii) using imaging techniques to locate functions
  iii) studying lesions (brain damage) to determine causality
- Lesion & brain damage studies help understand brain function by analysing impairment (functional deficit) due to: Accidents, strokes (localized brain damage), tumors and cancer, surgeries (Ex: epilepsy treatment)
  The studies reveal whether brain activity is necessary for a function or just correlated.

- surgeries & experimental approaches [only evidence for causal mapping of brain regions]
  - some surgeries involve implanting electrodes in the brain for medical or research purposes.
  - these procedures provide critical insights into brain function but raise ethical concerns.
  - studies in animals help but lack ability to verify subjective perception.

Historical cases

- Phineas Gage
  iron rod through left frontal lobe
  personality & behaviour dramatically changed
  demonstrates role of frontal lobe in personality
- other ex: stroke patients losing specific abilities help map functions to brain regions.

Vision & brain lesions

- Lesions in different parts of brain or visual pathway impact vision in distinct ways.
- Visual pathway damage
  Damage at different processing stages affect specific areas of vision.
- Kluver-Bucy syndrome
  Temporal lobe damage leads to
  - hypersexuality
  - visual agnosia (inability to recognise objects)
  - loss of emotional responses
- Patient DF
  Temporal lobe damage caused object agnosia. [This is where IT is]
  couldn't recognise/copy objects but could draw from memory.

Dorsal v/s ventral pathways

- Ventral (what!) : Inferior Temporal Cortex (IT) damage affects object recog.
- Dorsal (Where!) : Parietal cortex damage impairs spatial awareness & movement tracking.

Lesion studies in AI

- some researchers compare brain lesion methods to AI models.

Question is, if neuroscientists can't understant a μ-processor with similar methods, can they truly understand the brain?

Transcranial magnetic stimulation (TMS)

- Non invasive method to stimulate brain regions. and to find which part does what. Uses a magnetic coil.

Effects: - can temporarily disable or activate specific areas.
- used for studying **motor & perceptual functions**

Limitations: - can stimulate only surface level (cortical) areas
- less precise than direct brain implants

Ex: Stimulating motor cortex can make body parts move involuntarily.

Direct brain stimulation & visual prosthetics

Goal: use electrical stimulation to restore vision in blind individuals.

Methods: - Stimulate V1 (primary visual cortex) to create artificial vision
- visual prosthetics: implantable devices aiming to restore sight

Challenges: - limited success in real world applications
- only few devices have received regulatory approval

Neuralink & brain implants
+
focuses on brain implants

Challenges: - brain rejects foreign implants more than other organs
- ethical concerns & skepticism from neuro scientists
- safety concerns in early animal trials

Stimulation in AI

- researchers use **feature visualization** to understand neural networks.
- similar to **how brain stimulation reveals specific visual responses**

Neuroimaging is set of techniques to measure & visualize aspects of the nervous system.

Lesion: Abnormality in tissue of an organ.

Stroke: death of a region of brain cells due to poor blood flow.

Electroencephalography: Electrophysical monitoring method to record electrical activity of the brain.

Two-streams hypothesis: Model of neural processing of vision and hearing.

Phosphene is phenomenon of seeing light without light entering the eye. Electrostatic generator & group of people holding hands.

TMS: - changing magnetic field induces current at specific area of the brain, through electromagnetic induction
- used to measure activity & function of specific brain circuits in humans.

- V1 is retinotopic: stimulate V1 directly if intermediate visual pathway is damaged

- Jigsaw: what was the idea in the 3 articles?
current state?

Ch 9 - Face perception

- face perception is a key area in visual system research
- focuses on how faces are processed in the brain mainly in the IT and visual area V4.
- Faces hold social & cognitive significance, making their perception a specialized field of study.

Task:

- Face recognition involves identifying individuals similar to an object recognition task.
- Highlight is speed & accuracy in recognising well known faces.

Challenges in face recognition

- recognition gets harder with older/ less familiar images.
- context & familiarity influence recognition difficulty.

Robustness

- despite occlusions / aging, face recognition remains highly robust
- which implies specialised processing mechanisms.

Super recognizers and social context

- ↳ individuals with exceptional face recognition abilities
- social context affects recognition and memory of faces.
- face memory varies across individuals & may be influenced by innate ability / experience.

Experiments & Recognition challanges

- studies on challenges such as inverted faces reveal unique cognitive processing for face perception.
- faster, more accurate recognition of upright faces suggest face-specific mechanisms

Importance of faces in social interaction

- faces provide identity verification, emotional cues & social attributes
- convey info. about age, health & attractiveness
- Theory of mind (ToM): facial expressions help infer other's thoughts and emotions.
  - Prejudice & bias: 1st impression based on facial features can lead to unconscious biases.

Other race effect

cause: ↓ exposure to diversity during development

- People struggle to recognise faces of races they are less exposed to.
- This effect is malleable with experience; exposure improves recognition

Is face perception special?

↳ finding: Face perception is experience dependant & not purely innate

- are faces processed differently from objects?
- factors like viewpoint / illumination changes, expressions & identity recognition make faces unique.
- Fusiform Face Area (FFA) is a specialized brain region for face processing.

- Experiments compare face perception to object recognition. [BOLD responses]
  They try to determine role of FFA in distinguishing faces.
- studies with Mooney faces (low-contrast face illusions) show FFA activation
- AI models and multimodal experiments support specialized processing.
- brain stimulation in epilepsy patients alters face perception.
- Lesions in FFA lead to prosopagnosia (face blindness), where individuals cannot recognize faces.
  ↳ These experiments confirm FFA's role in face perception.
- some researchers argue faces are not uniquely special
- PCA analysis of neural networks layers suggest faces fall into general object categories, not in any special regions in the brain.
- Alternative perspectives propose face perception is part of broader shape or object recognition.

- Thatcher effect: distorted faces appear normal upside-down
- demonstrates how brain reconstructs distorted or blurred faces.
- face perception remains functional even with heavy distortion.
- indicates holistic nature of face perception.
- holistic face recognition relies on integrated features rather than isolated parts.
- certain facial features hold more weight in recognition.

Challenges in face recognition:
- Unfamiliar face recognition is significantly harder than recognizing familiar faces.
- changes in hairstyle, lighting / angle can impact accuracy.

Face space & statistical models [one of the hypothesis of how faces are represented in the brain]
- concept of face space, a statistical model explains how the brain encodes faces
- Face space model: is a mathematical model where faces are represented in a multi-dimensional space.
- concept: each face is plotted based on its features, making recognition statistical rather than categorical
- Face aftereffects: if we stare at a distorted face, a normal face appears distorted in the opposite direction. misidentification of avg face as anti face after staring at an extreme face
  ↳ implies brain encodes faces relative to an average face, not as absolute templates.

Why are faces important?

- identity
- expression (ToM) → helps estimate what others are thinking
- Attributes (Age, health, attraction)
- prejudice & bias
  - ability developed at a young age
  - other race effect is a type of bias.

How to find face selective neurons?

Monitor neurons in a brain region by showing different stimuli. can use FMRI scans.

Area found is the FFA.
↳ part of human visual system specializing in face recognition.

How to verify if it is only for faces?

Check for stimuli like
- human body parts
- only front view / other views of faces

Region of interest approach to find if region has higher attention to faces than other stimuli.

1. Localize face area individually in each subject
2. Measure response in this area in new scan.

Debate (articles) yes/no FFA is special?

Ch 10- Attention

Magic tricks
- Magic exploits attention misdirection
- humans struggle to focus on the right elements at the right time.

Ex: color changing cups experiment - many fail to notice key changes.

Gaze and visual attention

- foveated region in the eye has high visual acuity. [core attention mechanism]
- eye movement helps direct attention
- gaze tracking is key method in studying vision

Application: driver monitoring in cars

role of gaze in attention

- gaze control influences perception
- infrared-based gaze tracking improves accuracy.
- Humans instinctively detect where others are looking.
- white sclera in humans aids in social communication
- hyperacuity: humans can detect subtle gaze shifts.

Innovation in text reading - spritz

- spritz technology: reduces eye movement in reading.
- aims to enhance reading speed on small screens
- company failed due to open-source leaks of it's algorithm.

Eye tracking & Applications

- eye tracking experiments reveal task based gaze patterns
- Applications - web design (placing elements in attention zones)
  - Saliency prediction (predicting focus points in images)

Cognitive load

Why need attention?

- humans can't process all visual info simultaneously
- selective attention helps filter out irrelevant data
- visual search tasks demonstrate attention limitations [find vertical red line]
- attentia is required due to limitations in our visual processing capacities.

Visual processing

- fovea play important role in sharp vision
- preprocessing in the retina discards some visual data.
- Multitasking inefficiency: brain processes one action at a time.

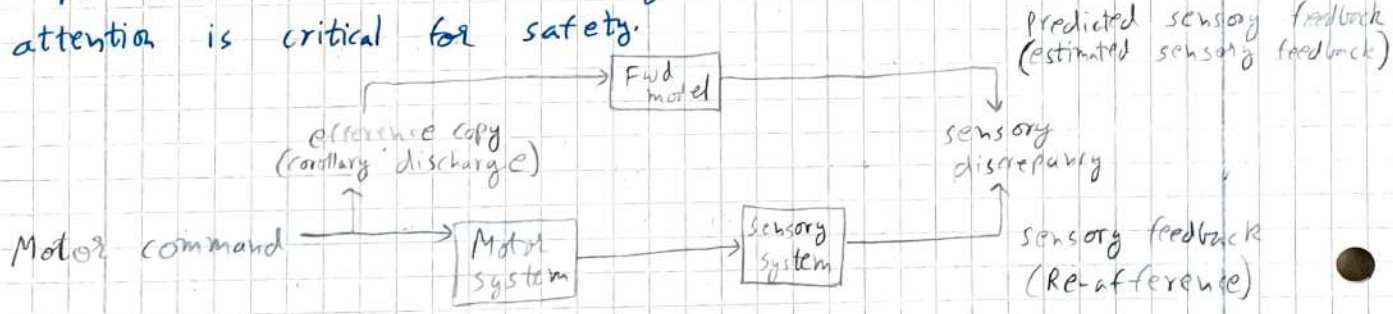
Change blindness

- inability to notice changes in scenes
- experiments show motion is key for detecting change [Motion cues are crucial]
- inverse rendering: idea is that the brain reconstructs the scene rather than storing it fully.

- Sparse visual representation: we only perceive what we focus on
- Just-in-time vision: world acts as external memory [we perceive it when needed, no need to store info in brain]
  Ex: refrigerator light looks like it is always on
- inverse rendering (full scene understanding) v/s sparse perception ?
- overt v/s covert attention
  → can be reflective or controlled
  Overt: moving eyes to focus. Others can see this type of attention.
  Covert: mentally shifting attention without eye movement
  → has to be actively done
- Experiments on FFA
  FFA is activated not just by seeing faces but also by thinking about them.
  Task specific attention changes FFA activity.
  Brain activity differs based on task demands, even with same visual i/p.
- Attention n/w in brain: Fronto-parietal attention network directs focus across tasks. Related to multi-demand system for cognitive flexibility

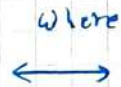
<u>Attention deficits & real-world implications</u>

- experiments on distracted driving show dangers of multitasking.
- attention is critical for safety.



→ move visual acuity around

- eye tracking can be done through web-cam but not accurate, insted use IR light as then eye is easy to detect then.
- eye gaze can be estimated by humans very accurately & we are very sensitive to small deviations/changes. Comes from hyperacuity.
- contrast b/w pupil, iris & retina, which makes it easy for gaze estimation.
- illusions where different lines of same length ~~of line~~ looks shorter/longer.



eye tracking experiments confirm eye movements were faster/slower hence speed of eye might influence the perceived length.

Saliency map is an image that highlights region on which people's eyes focus first or the most relevant regions for ML models

- **Inattentional blindness**: unable to see something in plain view
  cont. changes like in motion can be seen but with flashes what we see earlier is forgotten & changes in new img can't be easily seen.
- **change blindness** is a perceptual phenomenon that occurs when a change in a visual stimulus is introduced but is not noticed by the observer.
  flashes b/w images do not create motion signal & hence area cause of change blindness.
- peripheral vision is **almost always ignored. unless** there is some **motion** & we then change our attention i.e. move our eye to focus.

Change blindness

We see changes when

- something changes & creates a motion signal
- motion is exogenous cue; draws attention

(flicker/occlusion)

- if something interferes with drawing of attention then attention doesn't go to change.
  hence we don't see change as our attention is distracted.

Fronto-Parietal n/w [Domain general area]

Active when - shifting attention from one region to other
one feature to other
- engaging in a difficult task

Balint syndrome: when this n/w is not functioning, steering/controlling attention is difficult.

Ch 11 - Development

- Focus is on Face perception in the brain
- Alan Turing's imitation game: instead of creating adult intelligence, build child-like learning system.
- IS knowledge innate or acquired through experience?

Visual experience before birth

- Traditional belief: vision starts after birth when eye open.
- counter: Pre-birth visual experience occurs in the womb.
- Possible early visual stimuli:
  - Eye muscle movements for training
  - light penetration through womb (ex: bright light creating visual sig.)
  - Retinal waves: spontaneous neural activity in the retina.

Retinal waves & Pre-birth sensory learning

- Retinal waves act as pre-training for vision
- found in animals & likely humans.
- helps in developing visual processing before exposure to real world images
- Research: CNNs pretrained on retinal waves can develop features similar to real images.

Face Processing & brain development

- Is face recognition innate / learned?

2 theories or hypothesis: a) Minimal inductive bias (could be as simple as architecture): Learning occurs through exposure to moving faces. (Ullman's "Movers" theory) → hands touch objects & move it.

b) Innate templates: Face processing mechanisms are mostly built-in and refined over time.

- Evidence for innate recognition: [initial state]
  - infants as young as 1hr old prefer faces over non-face objects
  - looking time studies show babies look longer at faces than scrambled versions. [new borns look at something that is more surprising or new for longer]

Experiments on new borns:

- babies recognize faces even with minimal experience
- Tracking gaze & head movement shows preference for face like stimuli.
- Experiments on face occlusions, inversion & missing features reveal an understanding of facial structures.

Face perception in Monkeys

- New born monkeys initate human facial movements within days.
- suggest a strong innate component in face perception.
- comparison to human infants where facial imitation starts later (~1 month)

Development of Face Recognition skills

- recognition ability improves from infancy to adulthood
- Perceptual narrowing
  - infants recognise both human & monkey faces equally well.
  - by 9 months, ability to distinguish monkey faces declines, reinforcing specialisation in human faces.
- Development happens into adulthood, particularly in
  - expression recognition (micro-expressions, emotions)
  - Distinguishing diverse faces (ex: cross-race recognition)
  - which means it is not innate & improves with experience

Studying brain development in babies

- challenges
  - babies do not follow instructions, making brain imaging difficult.
  - studies use fMRI, EEG & looking time experiments.
- Findings
  - Early brain structure is mostly developed at birth, but myelination (process of insulating neurons for faster processing) continues for years.
  - Neural plasticity is highest in early years; brain damage recovery is better in infants than adults.

Role of innate v/s experience in brain development

- studies on congenital blindness
  - blind individuals develop category-selective areas for objects through touch and language.
  - suggests brain can adapt to different sensory inputs.
- Rewiring experiments (ex: ferrets) → animal born prematurely with eye closed ↳ no visual exp.
  - → connect eyes to auditory cortex
  - visual cortex rewired to auditory cortex can still process vision.
  - supports idea that connectivity influences structure
  - shows → architecture has huge affect on function of brain region

Brain structure

- experiments with face-deprived monkeys:
  - if monkeys lack visual exposure to faces, they do not develop face-specific brain regions.
  - suggests visual experience is crucial for forming face-recognition areas

Vision restoration & learning

congenital cataract → lens opacity at birth.

- congenital cataracts if not treated early (within 10 weeks), normal face recognition does not fully develop.
- supports critical period hypothesis for vision development

Computational Approaches

- Use pre-trained networks for visual learning.
  - inspired by retinal waves & early visual stimuli
  - Neural n/ws trained on fractals develop similar visual features as those trained on real images.

Fractals: complex geometric patterns that repeat at different scales

Takeaways

- Face perception starts early, possibly before birth
- Innate structures exist, but experience plays major role in refinement
- brain adapts dynamically, with perceptual narrowing - shaping abilities.
- critical periods determine success of vision restoration
- CNNs can simulate early visual learning.

Inductive bias of a learning algorithm is set of assumptions that the learner uses to predict outputs of given inputs that it has not encountered. Ex: CNN architecture does convolutions as that is the only operation it is built to do.; Training data

Empiricists: all knowledge comes from experience.

Brain development

At birth - most neurons in human brains are generated prenatally
- most long range structural connections are in place.

During 1st 1-2 years of life
- brain doubles in volume in 1st year
- cortical thickness & surface area increases sharply
- complexity of neurons in cortex & no. of synapses increase greatly.

Myelin - Fatty substance surrounding nerve cell axons to insulate them and increase transmission speed.

Min. inductive bias: what minimum construction is required for kids to pay attention to faces or learn. Ex: only architectures & no priors needed.

only prior needed

How to investigate if a skill is innate or learnt?

1) What is initial state? [skills at a new born]

2) How does it change/learn? when do we acquire the skill

3) What are causal relations? [injuries, brain damage]

Longitudinal study: study with repeated observations over time.

Tractography: 3D visualization of nerve tracts via diffusion MRI.
What fires together are connected together.

Multimodal i/p helps face perception develop.

Computational approach

hypothesis: - low initial acuity leads to holistic representation. (bigger receptive field)
- improved performance
- better generalization

Modelling: CNN + curriculum learning

Ch 1 - Computational Models, Supervision, generalization & Inductive bias

- Study Computational models + visual perception
- focus: understanding computational models & their applications in AI & human vision.

- aim of researchers is to mimic human vision through computational models.
- early focus was on deep reinforcement learning (Ex: AlphaGo) → AI that plays Go
- shift towards foundation models and LLMs, but these didn't solve all challenges of intelligence & we have moves past even this peak.
- Activity: Play Atari frostbite game w/o instructions [to show how bad AI is & how good a human is]
  observation of human intuition and learning v/s AI performance
- AI systems struggle to generalize, while humans adapt quickly
- Humans learn game rules & objectives faster than AI
- AI takes longer to reach human-level performance.
- Highlighting the challenge of RL in understanding context.

- Tools for understanding the brain / reverse engineering the brain
  - Mathematics
  - computational Models (CNNs, transformers)
  - Neural networks & symbolic AI
  - cognitive psychology
  - experience building models from both cognitive science & AI perspectives
  - perspectives on how different approaches integrate (ex: statistics, symbolic & statistical AI/ML, neural N/ws, cognitive psychology)
- Neuroscience approaches: from studying single neurons to higher level cognitive functions.

Levels of Analysis in brain functionality

David Marr's idea was that the bottom up approach is not the way to study the brain.

- Marr's 3 levels of Analysis: → figure out what problem & how the brain solves it on computational level & then use it on any hardware.

a) Computational theory: What problem is being solved?
b) Representation & Algorithm: How is the problem solved?
c) Hardware implementation: Physical system that executes the solution.

Ex: Physics engines simulate real world mechanics.

Mimicking Nature in engineering: computational problem: how to fly?

- Misconception: early attempts at flight mimicked bird wings
- True breakthrough: understanding aerodynamics & fixed wing flight
- Lesson: Understanding the problem at computational level is key.

Wright brothers didn't confuse flying with flapping. Algo: fixed wings + speed + control
Impl: engine, propeller, canvas & strings

Algorithmic V/s implementation levels in Aviation

- airplane design remained algorithmically unchanged since Wright brothers.
- Advancements happened at implementation level (jet engines, materials)
- similar issues in AI - algorithmic limitations hinder progress.

AI & levels of Analysis

- AI follows a similar pattern: many improvements are in implementation (hardware, data size) rather than algorithms.
- current AI lacks efficient learning algorithms like humans.

Limitations of Neural networks

- strengths: Learn from large datasets, computational efficiency.
- weakness: Poor generalization, slow learning, high energy consumption
- LLMs struggle with causal reasoning.

Challenges in AI generalisation & learning.

- AI fails to generalize from limited data
- struggles with forming goals & sub-goals
- suggests teacher-learner models for better AI development

- Important to fit models to data without overfitting / underfitting.

Model fitting

- overfitting: model memorises data, fails to generalise
- underfitting: model is too simple to capture patterns.
- choosing the right loss function is crucial.

overfitting is not necessarily bad. Neural n/w overfit mostly but with large data it works well.

Hypothesis space & errors:

- Expanding hypothesis space can reduce errors but introduce new ones.
- Approximation V/s optimisation errors in model training

Inductive bias in animals Example: rats drinking poisoned/non poisoned water with diff. taste or bg. noise.

- rats associate sickness with taste, not sound, highlighting built in biases that shape learning.
- Implications for AI: what should be learned V/s what should be innate? Ex: i) choice of model architecture ii) choice of loss/cost func. iii) Any other design choice

Generalization V/s overfitting

↳ within distr. V/s out of distr. generalization
Interpolation V/s Extrapolation

- AI must generalize beyond training data
- Occam's razor: Prefer simpler models for better generalisation. simplest explaination is the best one. Motivation for regularisation.
  ↳ methods for detecting deepfakes include inverse rendering
- inverse rendering reconstructs scenes to understand them.
  ↳ use in vision science (Ex: CAPTCHA recognition)

Challenges in vision

- Many factors affect visual perception (lighting, texture, camera angle)
- AI struggles with interpolating & extrapolating vision tasks.
- AI struggles with extreme lighting condition.
- Light & shadow perception impact face recognition.
- Physics-based models may improve AI generalization
- AI struggles with low-resolution & black and white images. Ex- Mooney faces test human & AI face recognition.

- 3D face reconstruction using probabilistic models V/s human perception
- face reconstruction can also use deep learning.

Challenges
- 3D face models require minimal data for learning.
- Quality control required in model reconstruction.

- Implementation level question:
  neuron = hardware in brain
  algorithm on top is neuron as to work on this hardware.

Hubel & wiesel did exactly this. investigated single neurons & what they did together.
given receptive field of a neuron (i.e. HW) what algo does it implement & which computational problem does this algo solve? (i.e. bottom-up approach)

- Gabor features were found by spike triggered averages.

Marr level analysis for obj recognition:

a) ~~What~~ recognition of obj or estimation of 3D scene with this obj

b) What is representation of obj in our brain? View point based / 3D methods

How does brain compute mappings & what are the algos?

c) How can we implement it on hardware?

Ex:

	Calc.	Physics engine
computational	Def: natural nos laws of arithmetic	Def: mass, force laws of mechanics $F=ma$...
↓ Algo	Base 10 + carrying Bas 2 + carrying Hexa decimal Roman	collision detection + resolution
↓ H/w	Digital/analog circuits Abacus	GPU TPU CPU

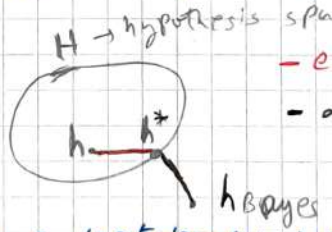
What levels of Marr levels our project uses?

Paper 1
Paper 2

- Neural networks are **good at**
  - learning spatitemporal patterns
  - responding quickly once learning is complete

Not good at
- cause-effect relations
- learning from others
- forming goals & sub-goals

Model selection:



- estimation error
- approximation error

Assumed that there is a best-in-class hypothesis, i.e. h^* s.t. $R(h^*) =$

Sources of error

- **Approximation error** exists because **actual model isn't in the hypothesis space**
  Ex: Using linear & sinusoidal functions but actual func. is something else, then sinusoid might fit data well but actual func. isn't in hypothesis space.
- sample complexity of ML algorithm represents no. of training samples required for successfully learning the target function.
- even within hypothesis space we might not find our perfect model, due to **optimisation problem**. [**estimation error**]

There exists a trade-off b/w the 2 errors

- **Hypothesis space is limited by model assumptions, inductive bias.**

Generalisation: **Problem** arises for **extrapolation** with a overfitted model. as it **doesn't generalise** well for out of distr. data. or it is better

to use a simpler model that doesn't fit data well but generalises well.

- 3D Morphable Model: is a generative technique that uses methods of statistical shape analysis to model 3D objects.

It is a simple model that can be used to represent faces.

Assumptions: face contains 2 components: shape S & texture T

hence use a linear combination of S & T, i.e. it is a linear model.

Use PCA so we don't over-represent certain faces.

Different approaches to detect deepfakes

1) Train a network to say if it is a deepfake or not
2) Use inverse graphics, to see if light is consistent & see if the scene is plausible.

Guest Lecture - Brain

Complexity of the brain

- brain contains 100 billion neurons (10^{11} neurons)
- each neuron has 10,000 synapses, totaling 10^{15} synapses
- Possible connections 10^{22}, but only 1 in 10 million is realised
- possible mind states exceeds no. of protons in the universe.

- A **connectome** is a **comprehensive map** of neural connections in the brain. It represents the wiring diagram of how neurons & brain regions are interconnected.

Nervous System overview

The Nervous system

- Central Nervous System (CNS)
  - consists of brain & spinal cord
  - body's master control unit
  - Spinal cord
    - connects CNS to PNS
  - Brain stem
    - connects brain to spinal cord
  - Brain
- Peripheral Nervous System (PNS)
  - connects CNS to body
  - body's link to the world
  - Autonomic Nervous System
    - regulates
    - involuntary processes
      - Ex: heart rate, digestion, pupil contraction
    - Sympathetic
      - Fight/flight response
    - Parasympathetic
      - calms the body
      - conserves energy
  - Somatic Nervous System
    - sensory info from sensory organs to CNS
    - relays motor commands from CNS to muscles
    - controls voluntary movements

Major brain structures

- Cerebral cortex:
  - Highest level of cognitive functions [Language, Math, reading]
  - voluntary movements
  - conscious perception
  - semantic memory: is a type of long-term memory that stores general knowledge about the world, concepts & facts not linked to personal experience.
- Limbic system (emotion & memory)
  - **Amygdala**: controls Fear, anxiety, aggression. responsible for emotional responses.
  - **Hippocampus**: Spatial Navigation, episodic memory, memory formation, organization of thoughts; Brain's ram

- **Thalamus**: **Gatekeeper** for sensory i/p to the cortex
  acts as **relay station** of sensory i/p to cerebral cortex
  Thalamo-cortical loops are neural circuits that connect thalamus & cerebral cortex.
  relaystation → receive, process & transmit info b/w diff. regions. (thalamus & cerebral cortex)

- **Basal ganglia**: (Motor control, reinforcement learning)
  "**servo-mechanism**" for cerebral cortex
  - **shortcut from sensory association areas to motor control**
  - trained, complex i/p/o/p mappings.

  - When learning to drive a car; switch from cerebral cortex to basal ganglia
  - basal ganglia bypass entire layers of neurons

- **Cerebellum**: Temporal coordination of movement
  error correction of movement
  Motor skill learning
  - Extremely regular structure: climbing fibres, parallel fibres

Organisation of cerebral cortex

Gyri & Sulci: Folds that increase surface area
↓ high ↓ valley (low)

information flow
Layer 4 – i/p layer
↓
layer 2/3 – hidden layer
↓
layer 5/6 – o/p layer

- layer 4: receives info from thalamus
- layer 2/3 & 5/6: Process & send o/p to diff. areas

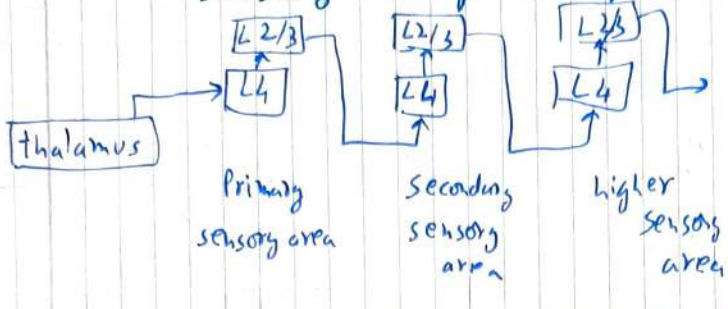
Nissl stain → for cell bodies of neurons
Golgi stain → dendrites & axons of a random subset of neurons.

- **cortical mapping**
  - **Somatotopy**: Body regions [mapping of diff. body regions to diff. brain areas]
  - **Retinotopy**: vision [maps **field of vision** to diff. parts of brain] primary visual cortex
  - **Tonotopy**: hearing [maps diff. freq. to diff. parts of brain] auditory cortex

Cortical circuits & connectivity

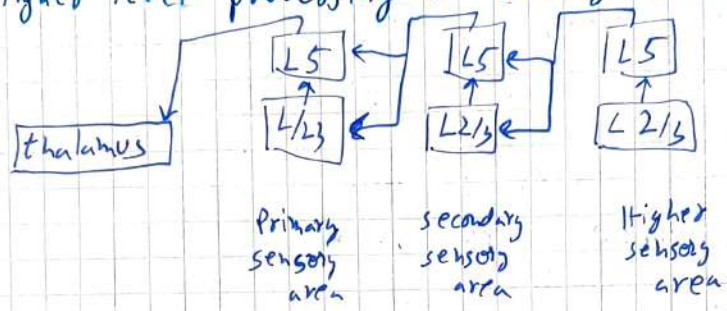
- **Feed fwd** connections: (**bottom up info flow**)

sensory → high level processing



- **Feed-back** connections (**top down info flow**)

Higher level processing → sensory modulation.



- Lateral connections: Inter hemispheric communication

Brain Lateralization & function

Lateralization is tending for neural functions or cognitive processes to be specialized to one side of the brain.

Homologue cortex areas at both sides have different functions

- Left hemisphere represents RHs of body & vice-versa

Left hemisphere	Right hemisphere
- Language	- spatial skills
- problem solving	- holistic processing / Ill^e processing
- seq. processing	- face recognition
- right body motor control	- left body motor control
- reading	- drawing, face recognition
- right visual field	- left visual field

- Non-human ex: birds use left hemisphere for food discrimination

- **Asymmetry enhances efficiency & specialization**
  - maybe a more efficient & flexible design principle
  - reduces redundancy across hemispheres
  - allows for no-cost extension: extends cortical capacity.

Guest Lecture- Neuroimaging

Neuroimaging involves techniques to visualise structure & function of the brain.

- used in medical research & clinical applications to study the brain-diseases & cognitive processes.

Types of neuroimaging
- structural
- Functional
- Multimodal

Neuro-imaging techniques
- Structural
  - CT [Computerized Tomography]
  - MRI [Magnetic Resonance Imaging]
- Functional
  - EEG Electro-encephalo-graphy
  - fNIRS Functional near IR spectrography
  - fMRI functional MRI
  - MEG Magneto encephalo-graphy
- Molecular
  - PET Positron emission tomography
  - SPECT Single photon emission CT

1) Structural Neuroimaging (x, y, z)
- Useful for medical applications where spatial resolution is important.

CT scans
- use x-ray radiations
- Provides faster scans & better visualisation of tissue structures, organs & bones
- useful for good spatial resolution in medical applications

Structural MRI scans
- Offers better spatial res. & soft tissue contrast
- avoids radiation exposure
- Challenges:
  - segmentation: separating white/gray matter & cerebrospinal fluid.
  - Tumor boundary detection (edema, core, enhancing core)

2) Functional Neuroimaging (x, y, z, t)

i) fMRI (Functional Magnetic Resonance Imaging)
- measures brain activity using Blood Oxygen Level Dependent (BOLD) signals
- tracks oxygenated v/s deoxygenated blood to identify active brain regions
- challenges: big data management due to high voxel counts and multi-dim. arrays.

ii) fNIRS
- measures brain activity by detecting changes in blood oxygenation using near-infrared light

iii) EEG (Electroencephalography)
- captures electrical activity in the brain
- high temporal resolution but limited spatial resolution

iv) MEG (Magnetoencephalography)

Measures magnetic activity in the brain using magnetically shielded rooms or OPM-MEG devices.

3) Multimodal Neuroimaging

combines different techniques for enhanced data analysis

EEG-fMRI or EEG-fNIRS

Mobile Neuroimaging: Portable neuroimaging devices enable studies outside of lab env., expanding possibilities of research.

Brain-Computer Interfaces (BCIs)

- BCIs allow direct communication between brain & external devices.
- integration with machine learning enhances predictive capabilities & real time analysis. Preprocessed neuroimaging data is converted into numpy arrays for ML applications.

Data formats for neuroimaging: standardized format BIDS

Challenges

1) Managing large datasets
2) Resolving spatio-temporal resolution issues across modalities
3) Combining multimodal data effectively.

Guest Lecture - Bridging Gap B/w human & machine visual attention

Human Eye Anatomy & function

- Fovea in the retina has highest concentration of cones, enabling sharp vision.
- peripheral vision is less precise, necessiating const. eye movement to focus the fovea on points of interest.

Types of eye movements

Saccades, smooth pursuit

- Vergence: reflexive alignment of eyes for targets at varying dist.
- Vestibulo ocular movement: stabilize vision by compensating for head movements.
  - → involuntary

Pupil size fluctuations

- influenced by brightness, near fixation & psychological responses
- Linked to neuromodulatory systems affecting arousal, attention & perception.

Eye Tracking technology

- uses near IR light & high res cameras
- relies on PC-PR principle for calibration. (Pupil Center - Corneal Reflection)

~~Types~~ - Eye-tracking is process of using sensors to locate & record features of the eye to estimate where, what & how an individual is looking.

- Pupil is black disk easy to detect
- Cornea layer reflects IR light

Types of eye trackers

- screen based: high sampling freq; remote setup
- wearable: VR/glasses allowing natural movement
- webcam-based: simple, no special hardware required.

Computational models of human visual attention

simulating human attention enables advancements in image/video compression, marketing, image captioning & quality metrics.

Goal: dev. models mimicking human cognitive capabilities

Saliency Maps: highlight regions where people's eyes focus first or express pixel relevance.

Scan path modeling with Gravitational ~~fields~~ models:

- inspired by gravitational laws; visual characteristics act as virtual masses.
- **Mechanism of inhibition of return** favours complete scene exploration.

Neural networks exhibiting human-like visual attention

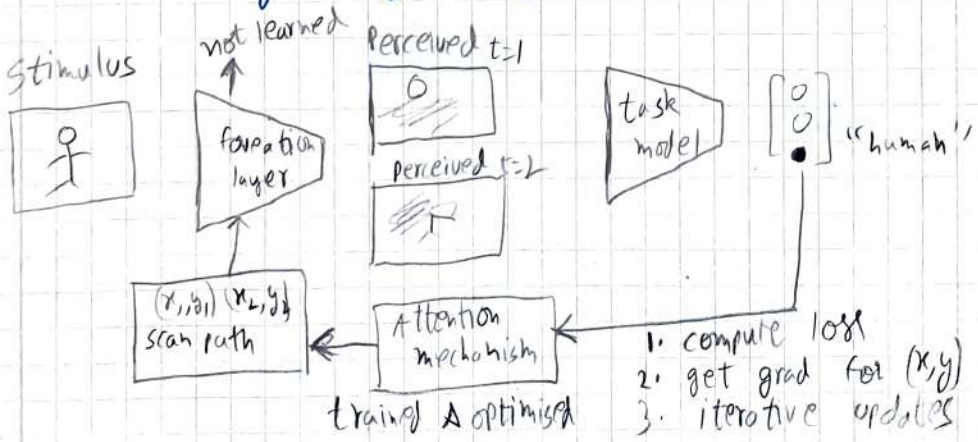
Diff. b/w humans & machines

- machines access full-res. stimuli at all times
- humans use eye movements to process limited portions of stimuli selectively.

- Neural Visual Attention (NeVA):
  - incorporates biologically inspired constrains like **foveated vision into NNs**
  - **Avoids hand-crafted features**; instead use differentiable foveation layers & attention mechanisms.
  - **Training** involves iterative updates to optimise attention mechanisms.

Advantages: - improves efficiency, interpretability & robustness of machine vision systems.

- Moves beyond descriptive or task specific models to generalized approaches

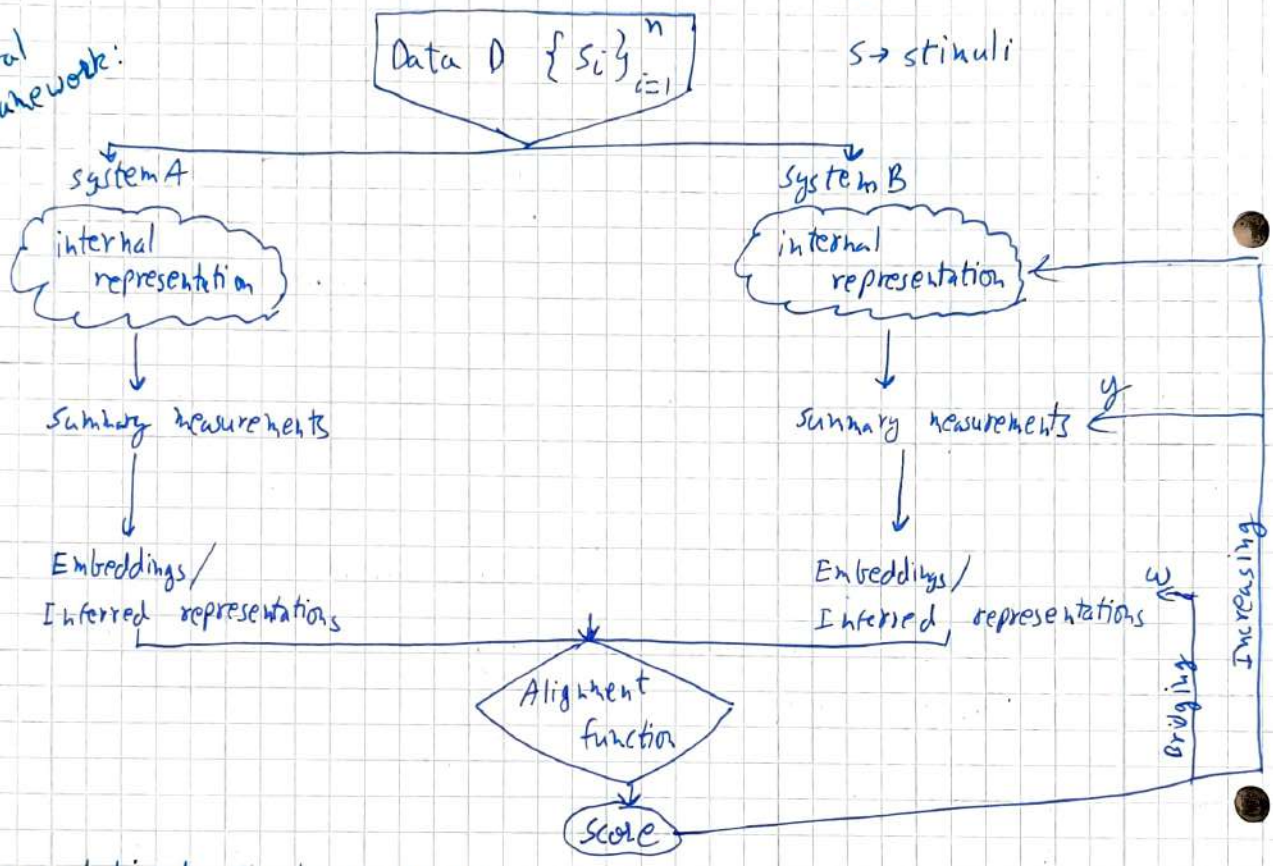


Guest Lecture Alignment

Objective: understand how neural networks (NNs) process stimuli & whether they align with human brain representations.

Answers key questions like
- Can NNs help us understand the brain?
- Are NN representations similar to those in human brains?
- How do NNs compare to other brain models?
- To what extent can NNs model brain processing?

General framework:



Representational Similarity Analysis

- RSA compares patterns of activity (in human brains to NNs) to asses representational similarity.
- RSA matrix is symmetric (upper Δ^{le} = lower Δ^{le})
- can be visualised as an undirected graph with edge strengths representing similarity.

Steps to create a data matrix

1) Extract $\hat{\beta}$ estimates from generalised Linear Model (GLM)
2) Flatten β maps into voxel patterns
3) Assemble the data matrix using consistent regressors for all stimuli.
   output: stimuli x stimuli matrix showing correlations across voxels.

Advantages of RSA
- simple to implement & interpret
- No strong assumptions about data.

Disadvantages
- can't control for covariates (there might be factors correlated to eachother)
- assumes equal weights for all predictors
- can't model non-linear effects.

Hence use GLM [Generalised Linear Model]

GLM

Purpose: Models BOLD response in fMRI experiments as a function of neural stimuli.

fMRI signal decomposed into func. of parameter estimates + residual = error

Components:

Parameter β estimates, capture contribution of each predictor to the observed signal.

Integration with RSA:
1) Parcel brain into given regions using an atlas
2) Construct Representational Dissimilarity Matrices (RDMs) for a specific region.

NN v/s Human brain representation

- Input images into NN layers & compare RDMs at each layer with brain RDMs.
- evaluate how similar/diff. the representations are b/w NNs & human brain.

Encoding Models
- Predict neural responses based on external stimuli or task variables.
- can synthesize the neural resp. to any image of choice; but can test only against a particular set of images for which ground truth is present.

Centered Kernel Alignment (CKA)

- Measures similarity b/w 2 matrices, focusing on structural alignment while ignoring irrelevant factors like global scaling/shifts.
- useful for comparing Non-linear representations in brains & NNs.
- applied in improving brain-inspired AI models.
- enhances interpretability.

Linear Shape Metric (LSM)

$$LSM(A, B) = \frac{\|A^T B\|}{\|A\| \, \|B\|}$$

- Denomenator ensures metric is independent of magnitudes of AA and BB.

- scale invariant due to normalization
- effective for low-rank matrix comparisons & latent dimensions.
- suitable for applications requiring computational efficiency.

Paper-1

- study investigates effects of high visual acuity on face recognition development, particularly in children treated for congenital cataracts.
- Proposes starting life with reduced acuity helps in forming spatial integration strategies necessary for face processing.

Key findings & methods used

1) children treated for congenital cataracts show face recognition Deficits

Method: case study of child (RK) who had cataracts removed at age 4.5. He underwent visual tests, including acuity, contrast sensitivity & face discrimination.

Result: RK had normal vision but showed significant impairment in face identification. Aligns with prev. ~~case~~ clinical studies where children exhibited similar deficits when treated late for cataracts.

2) High Initial Visual Acuity (HIA) hypothesis

Method : Literature review of past research on congenital cateracts

Result : Normal infants have low visual acuity, encouraging them to process broader spatial features.

In contrast, children treated ~~with~~ for cataracts often start with sharp vision, potentially skipping the spatial integration phase.

3) NN simulations support HIA hypothesis

Method: CNNs were trained on face images with diff. resolutions.

2 training approaches used

i) high res. images from the start

ii) Blurred images 1st, gradually increasing res.

Result: Networs trained with blurred images first, developed

- larger receptive fields
- better performance in face recognition tasks
- improved generalization across diff. image resolutions.

In contrast n/ws trained with high res. images 1st showed poor generalization & struggled with face recognition.

4) Gradual visual acuity development is evolutionarily Beneficial

Method: analysis of receptive field sizes in CNNs trained under diff. conditions.

Result: n/ws trained with progression from blurry to sharp images had better face processing, mirroring human dev.

suggests evolutionary advantage to natural progression of acuity in infants.

5) Implications for clinical & AI applications

Method: comparitive analysis of face recognition perf. across diff. CNN training regimes.

Result:

- Clinically: temporary optical blurr post-surgery might help children develop better spatial integration skills.
- AI : training face recognition models with progressive res. learning could enhance their robustness & generalization.

- even brief periods of visual deprivation can lead to face identification deficits.
- non-human primates with controlled deprivation face same deficits.
- children with high initial acuity are biased towards local processing.
- HIA hypothesis: Abnormally high retinal acuity experienced by children ~~with~~ treated for cataracts, compared to normal newborns, may result in compromised spatial integration.
- training done by manipulating blurr of training/test data.
- 4 instances of CNN trained
  i) blurr to high res
  ii) high res. to blurred
  iii) blurr - blurr
  iv) high res - high res.

Resulted in 3 primary results

1) Blurred images in training data consistently increased the size of RFs irrespective of when in training regimen it was introduced

~~High res - blurr~~ blurr-highres : didn't shrink size of RFs.

2) even when initial phase of training on blurred images is small, it is sufficient to induce creation of large RFs that are stable over subsequent training epochs with high res. images.

3) order in which images of diff. res. are presented affects subsequent classification perf.

Blurr - high res : produces best & generalized classification perf.
it has highest AUC & lowest absolute slope

High res - blurr : yield poor perf. when tested on high-res images although in aggregate it has been trained on same set of images with only the order varying.

Sim 1

i/p : grayscale face images

400 images
duplicated & blurred [9 pixel | blurr kernel]

- Each sample image queried against rest of the database, matched 2604 locations with 20 box sizes per location.
- L1 (norm) dist. calc. b/w given face patch & all other corresp. face patches.
- Z-score used to compare within ID dist to across ID dist.
- Averaged Z-score found
- At each location of patch, change patch size to calc. Z-score as function of query box size.
- compare these functions across high-res v/s blurred images to see how diff. patch sizes contribute to acheiving same Z-score, depending on image resolution.

Sim 2

i/p : face cropped color-images (100x100px) converted to grayscale

- gaussian filt with diff. σ used for blurring.
- Alexnet architecture with dropout (50%) fined LR
- To analyse RFs of CNN, shape of Gabor wavelets in kernels of 1st conv. layer were investigated.
- geometric properties quantified using dist b/w 2 elipses (1 +ve & other -ve)
- use thresholding, erosion & dialation to extract individual ellipses.