SUSTech
Southern University of Science and Technology

神经计算与控制实验室
Neural Computing & Control Lab

Brain Intelligence and Artificial Intelligence

人脑智能与机器智能

Lecture 2 – Visual System

Quanying Liu （刘泉影）

SUSTech, BME department

Email: liuqy@sustech.edu.cn

BME5012, Fall term, 2024

Lecture 1 Recap

1. BI inspires AI.
   I. Biological neuron → Artificial neuron
   II. Animal vision → CNN
2. AI helps understand BI.
   I. Neural / Behavioral data analysis
   II. Explain the necessity of observed brain structures or functions
   III. Resemblance between BI and AI
3. Levels/scales: molecular, synapse, neurons, networks, maps, system, CNS
4. Learning in the brain vs Learning in AI

Reverse-engineer the brain

to understand the design principles of brain

The Marr's three levels of explanation

top-down

<u>Level 1: Computation theory</u>

- What is the problem to be solved?
- What are the inputs and outputs to the computation?
- What is the goal, and what is the logic by which it is carried out?

<u>Level 2: Algorithmic</u>

- How is the information represented and processed to achieve the computational goal?

<u>Level 3: Implementation</u>

- How is the computation realized in physical or biological hardware?

The Marr's three levels of explanation

top-down

Level 1: Computation theory

- What is the problem to be solved?
- What are the inputs and outputs to the computation?
- What is the goal, and what is the logic by which it is carried out?

Level 2: Algorithmic

- How is the information represented and processed to achieve the computational goal?

Level 3: Implementation

- How is the computation realized in physical or biological hardware?

Bottom-up

Lecture 2 – Visual System

0. Marr's three levels of explanation

1. Evolution of the eye

2. Function of the visual system

3. Structure of the eye

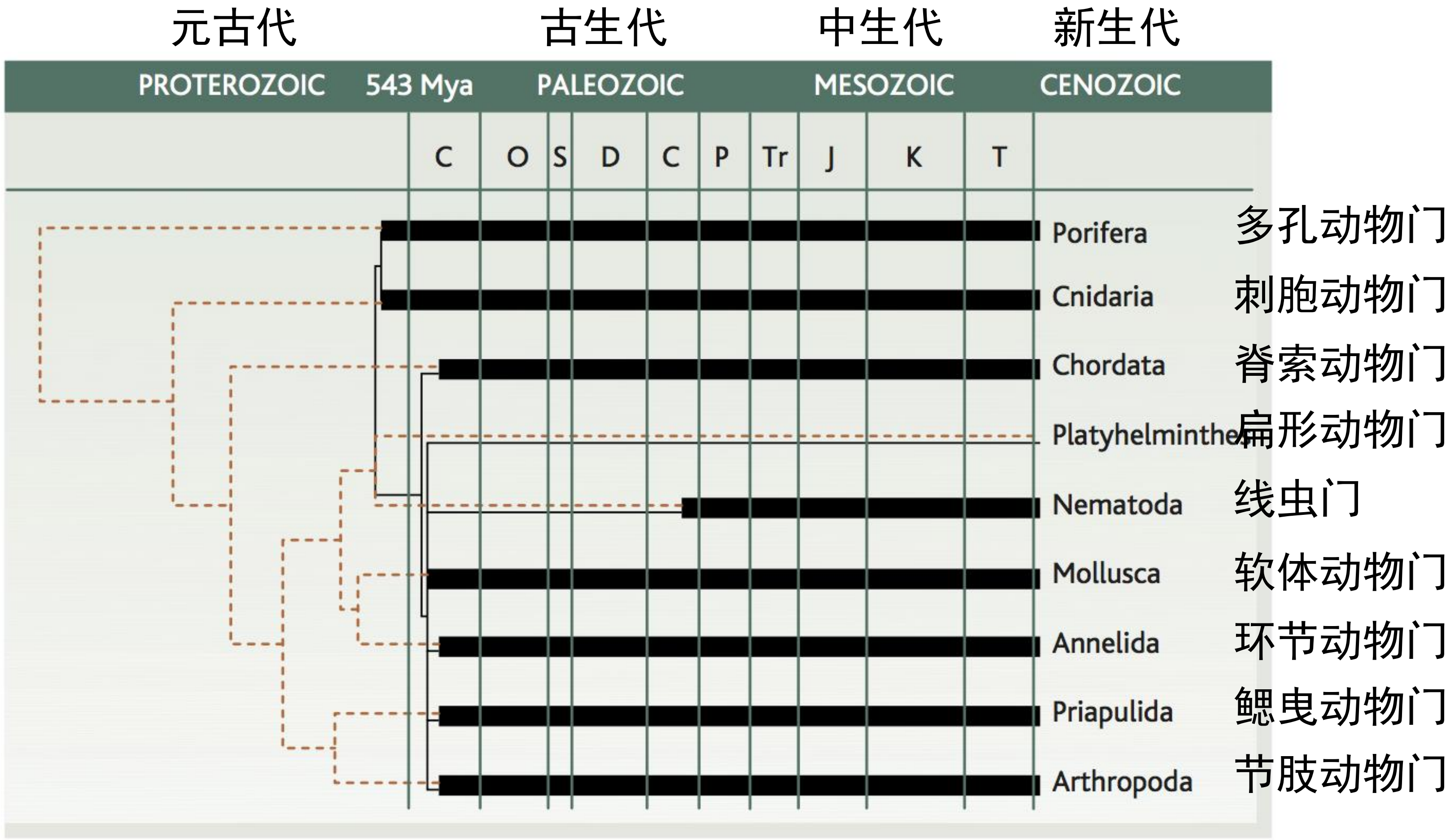
4. Photoreceptors

5. Information integration by ganglion cell

6. **Visual pathways**: photoreceptors, interneurons, ganglion cells, LGN, V1, ventral/dorsal streams…

7. Some discussions about **BI & AI in visual system**

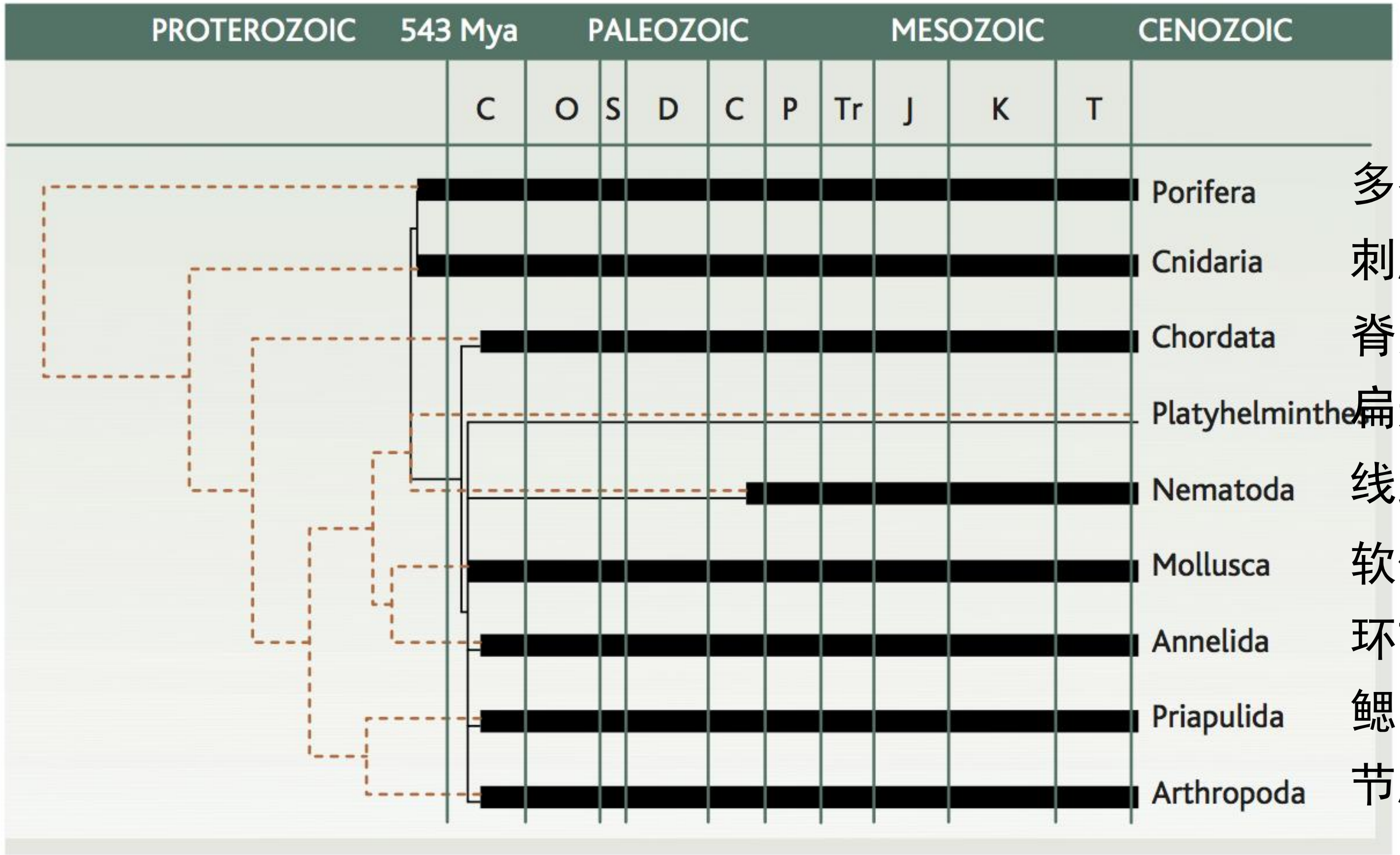
Animal Evolution



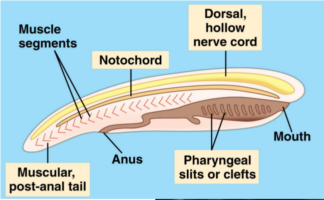
Jermiin, Poladian, Charleston (2005). Is the ”Big Bang” in Animal Evolution Real? *Science*

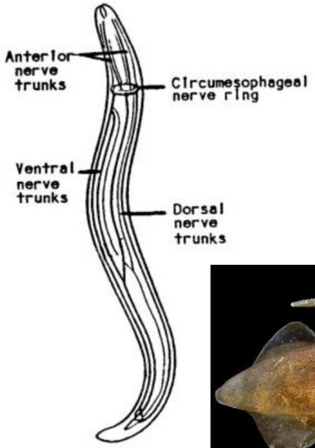
Animal Evolution

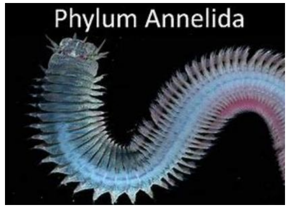
元古代 古生代 中生代 新生代



多孔动物门
刺胞动物门
脊索动物门
扁形动物门
线虫门
软体动物门
环节动物门
鳃曳动物门
节肢动物门

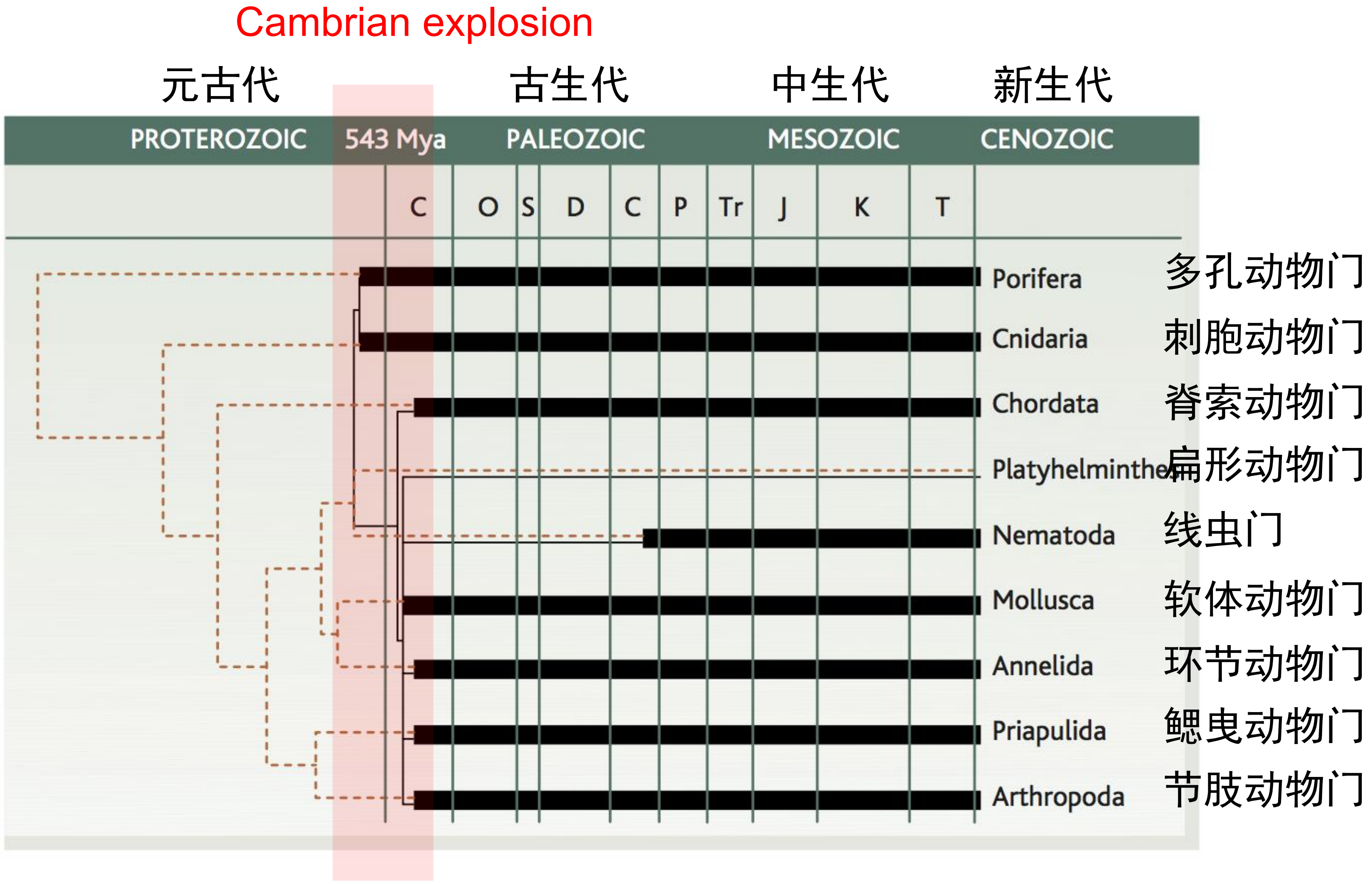


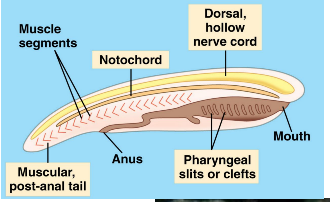


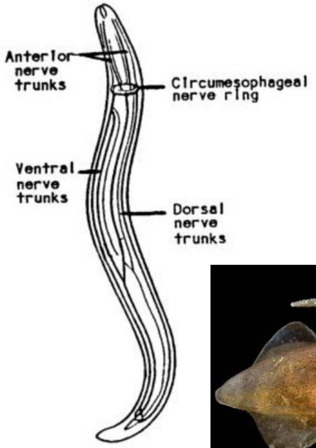


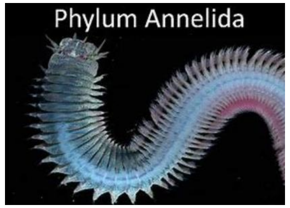
Jermiin, Poladian, Charleston (2005). Is the ”Big Bang” in Animal Evolution Real? *Science*

Cambrian explosion (寒武纪生命大爆发)



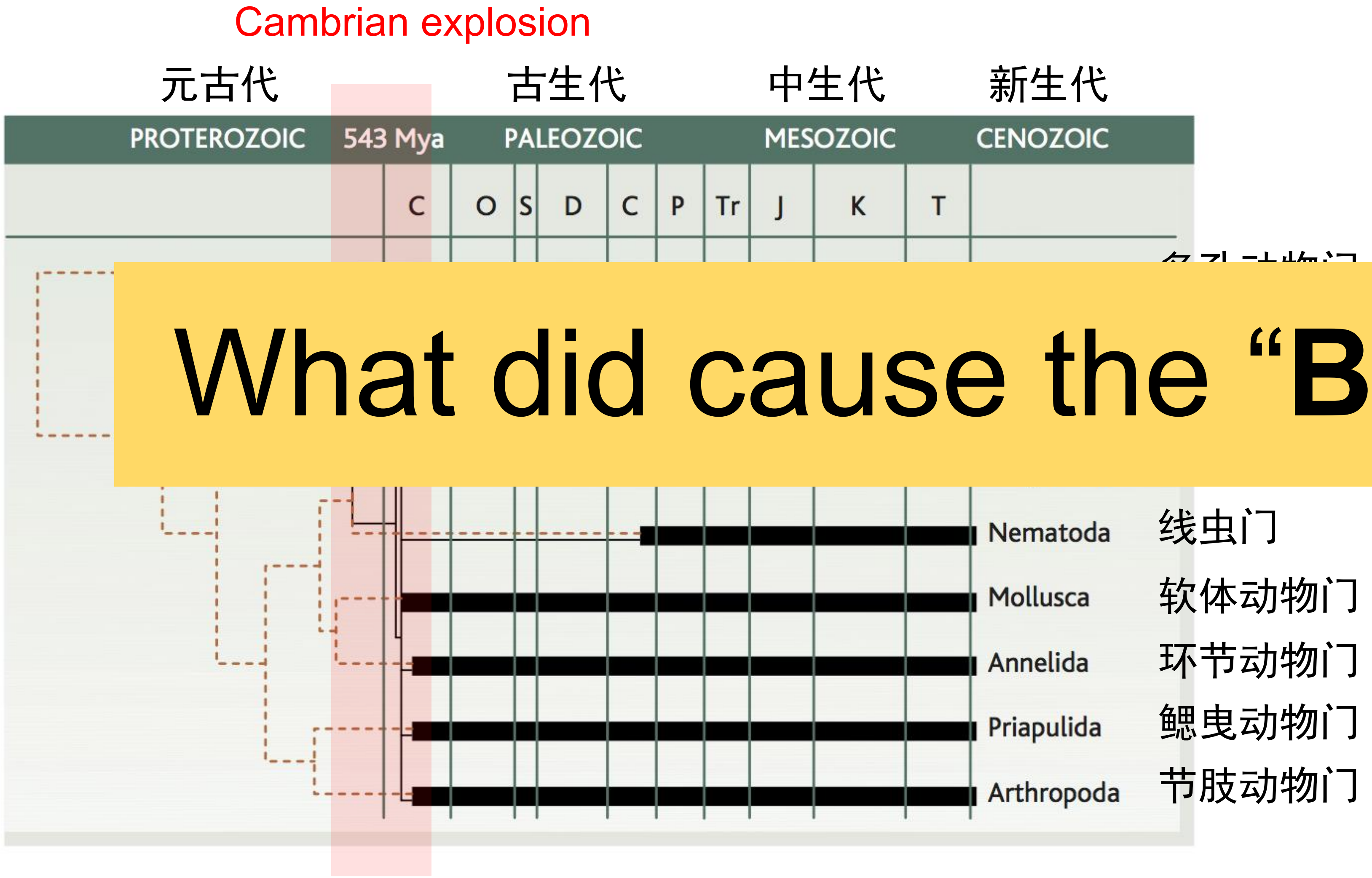




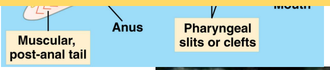


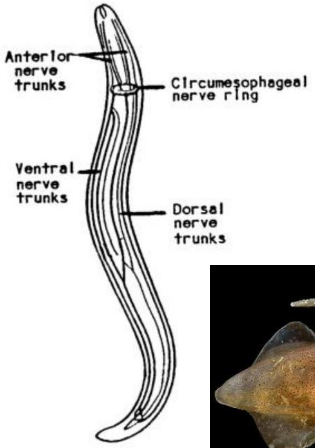
Jermiin, Poladian, Charleston (2005). Is the "Big Bang" in Animal Evolution Real? *Science*

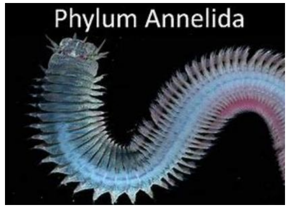
Cambrian explosion (寒武纪生命大爆发)



What did cause the “**Big Bang**”?

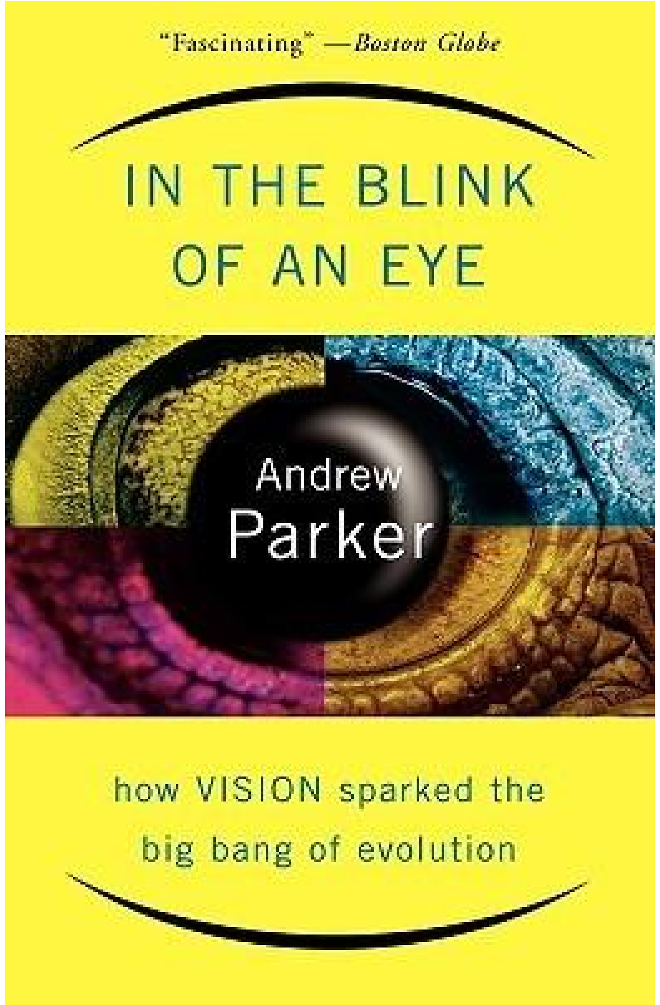






Jermiin, Poladian, Charleston (2005). Is the ”Big Bang” in Animal Evolution Real? *Science*

Evolution of the Eye



Andrea Parker
In the Blink of an Eye

Eyes first appeared approximately 543 million years ago during the Cambrian period—the geological period that marks the rapid increase in biodiversity.

Oxford zoologist Andrew Parker proposed "Light Switch Theory'", suggesting that it was the development of **vision** in primitive animals that caused the explosion.

Precambrian creatures were **unable to see**, making it impossible to find friend or foe.

With the evolution of the **eye**, the size, shape, color, and behavior of animals was suddenly revealed for the first time. Once the lights were "turned on," all animals had to either adapt or die, and in a geological instant, the world became a very different place.

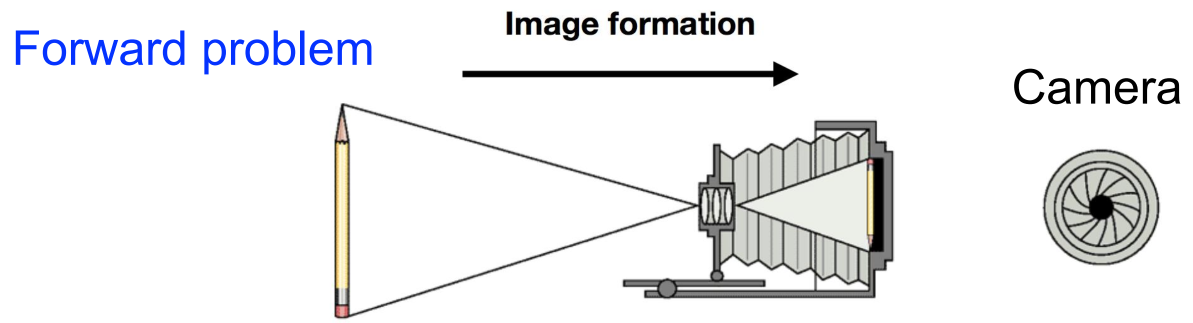
Function of the Visual System

The visual system is a part of the central nervous system (CNS) that gives organisms the ability to **detect**, **process**, **interpret** information from **visible light**, with the goal of building a **representation** of the surrounding environment.

Function of the Visual System

The visual system is a part of the central nervous system (CNS) that gives organisms the ability to **detect**, **process**, **interpret** information from **visible light**, with the goal of building a **representation** of the surrounding environment.

The visual system is **NOT** a camera.



Function of the Visual System

The visual system is a part of the central nervous system (CNS) that gives organisms the ability to **detect**, **process**, **interpret** information from **visible light**, with the goal of building a **representation** of the surrounding environment.

The visual system is **NOT** a camera.

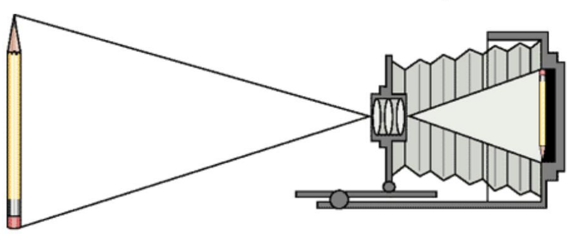
The visual system solves the inverse problem.

Based on the neural activity, to **infer** the image we see:

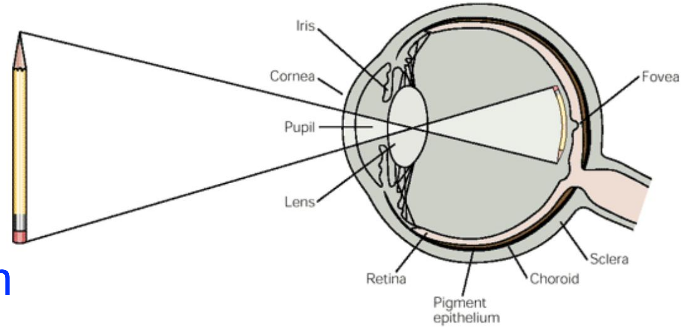
- Detect
- Process
- Interpret

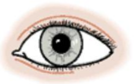
Forward problem

Image formation



Camera



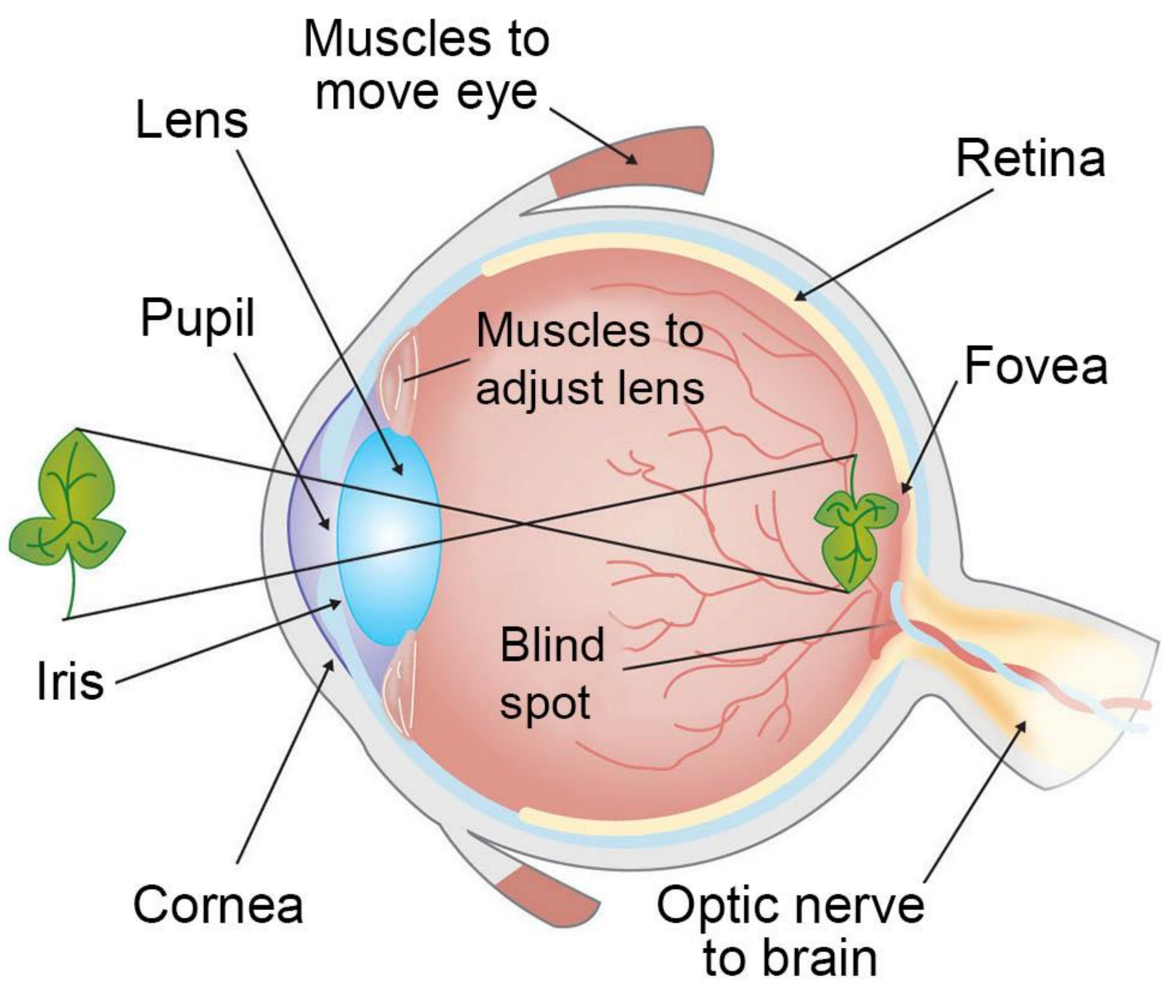


Inverse problem

Vision (brain)

Eye + LGN + V1 + …

Structure of the Eye



Lens + Cornea: help focus light onto the eye, like camera lenses.

<u>Photoreceptors</u> (感光细胞) in **Retina**: convert light energy into neuronal activity, like camera sensors.

Pupil: to adjust based on illumination of light, like aperture of the camera.

Optic nerve: transmit the image signal to the brain, like camera electronics.

Eye can **clean itself**, by producing tears, blinking.

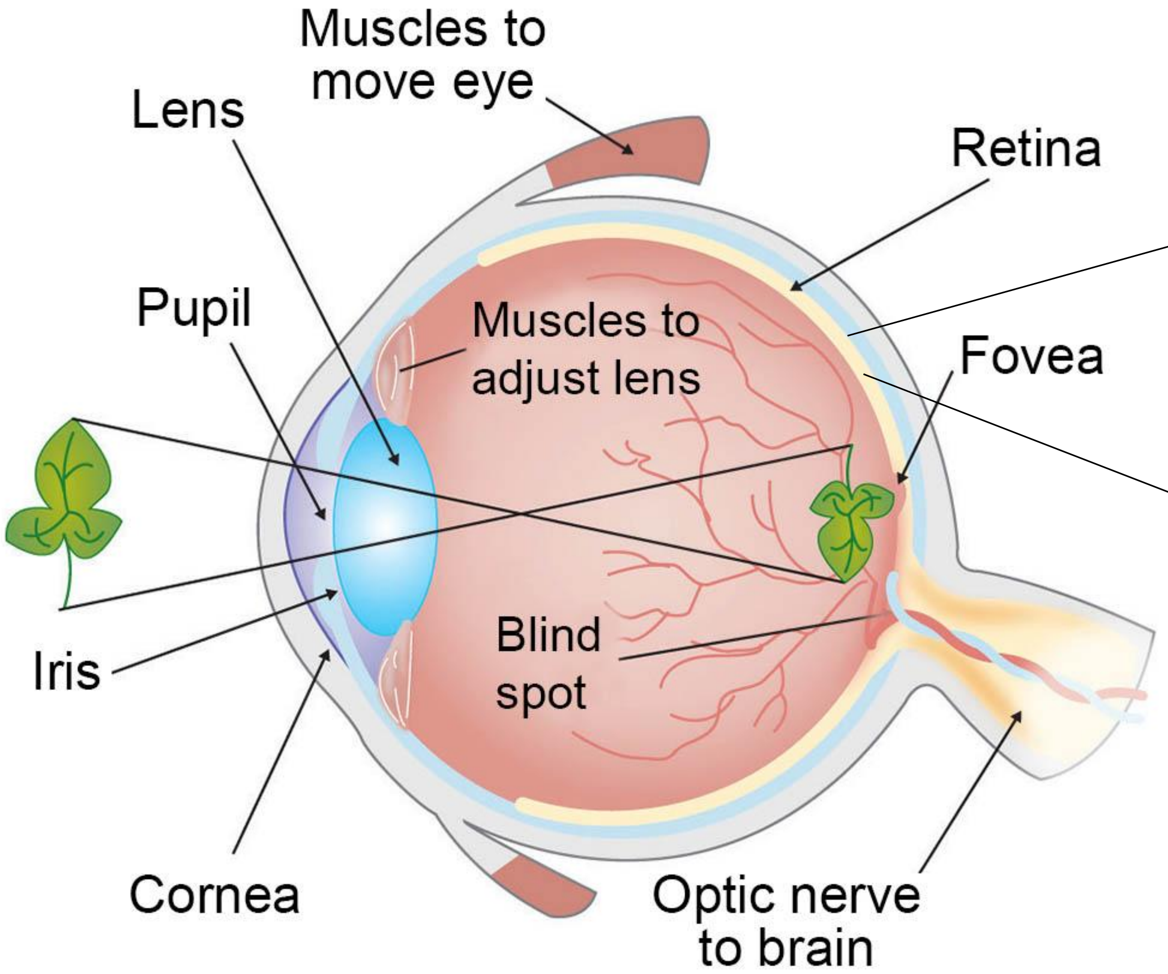
Eye can even **repair itself** in the event of modest damage to its optic.

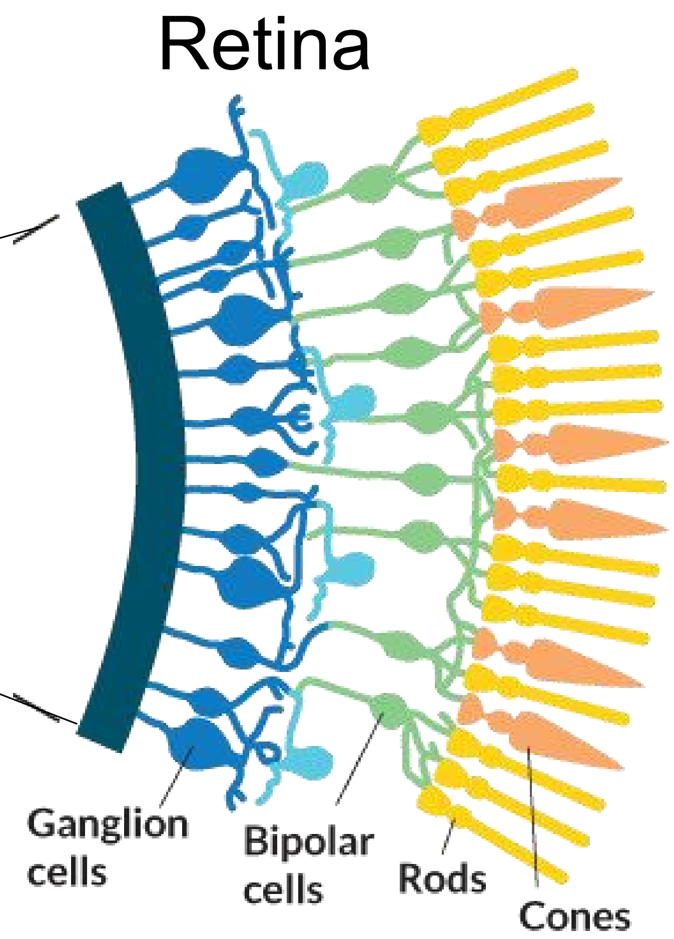
The structures of the eye are **different** across species.

The insects has many **lenses** arrayed across the eye.

Not all animal have **retinas** like ours.

Structure of the Eye



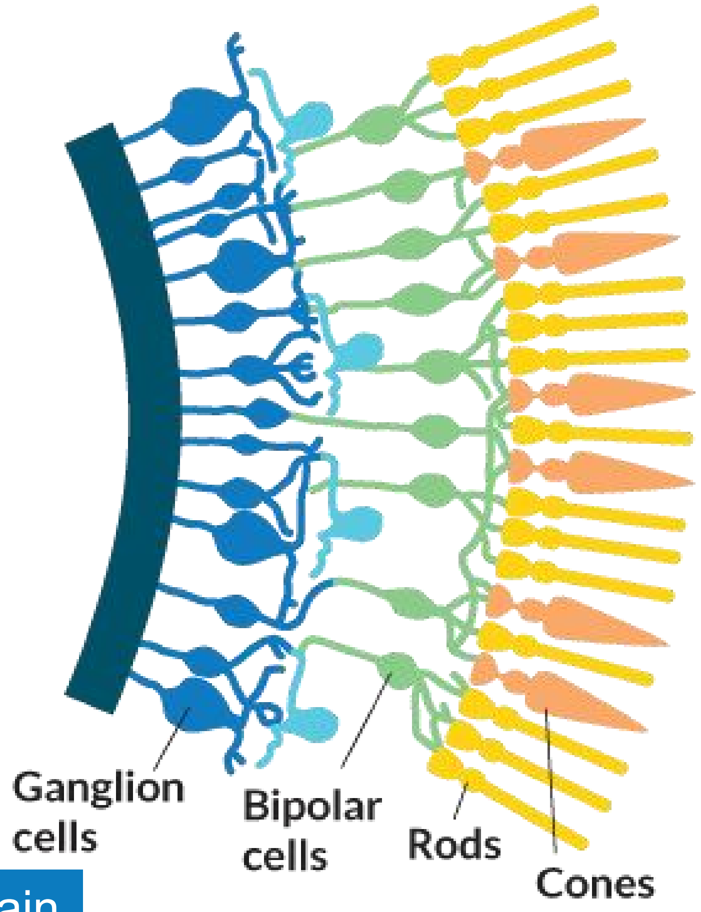


Diversity in the cell types

Photoreceptors: **2** types (rods and cones)

Interneurons (eg., bipolar cells): **> 50** types

Retinal ganglion cells: > 30 types



Output into brain

Internal transmission

Photoreception

Arranged in an inside-out fashion

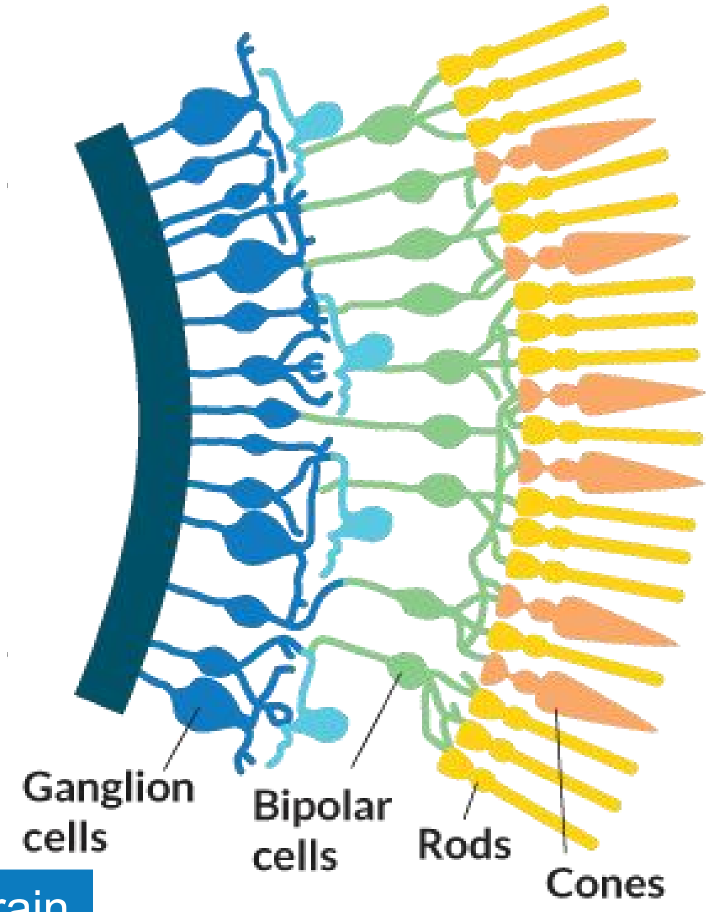
Diversity in the cell types

Photoreceptors: **2** types (rods and cones)

Interneurons (eg., bipolar cells): **> 50** types

Retinal ganglion cells: > 30 types

Q: Do you think this diversity is well designed by nature, or happens by chance?



Output into brain

Internal transmission

Photoreception

Arranged in an inside-out fashion

Diversity in the cell types

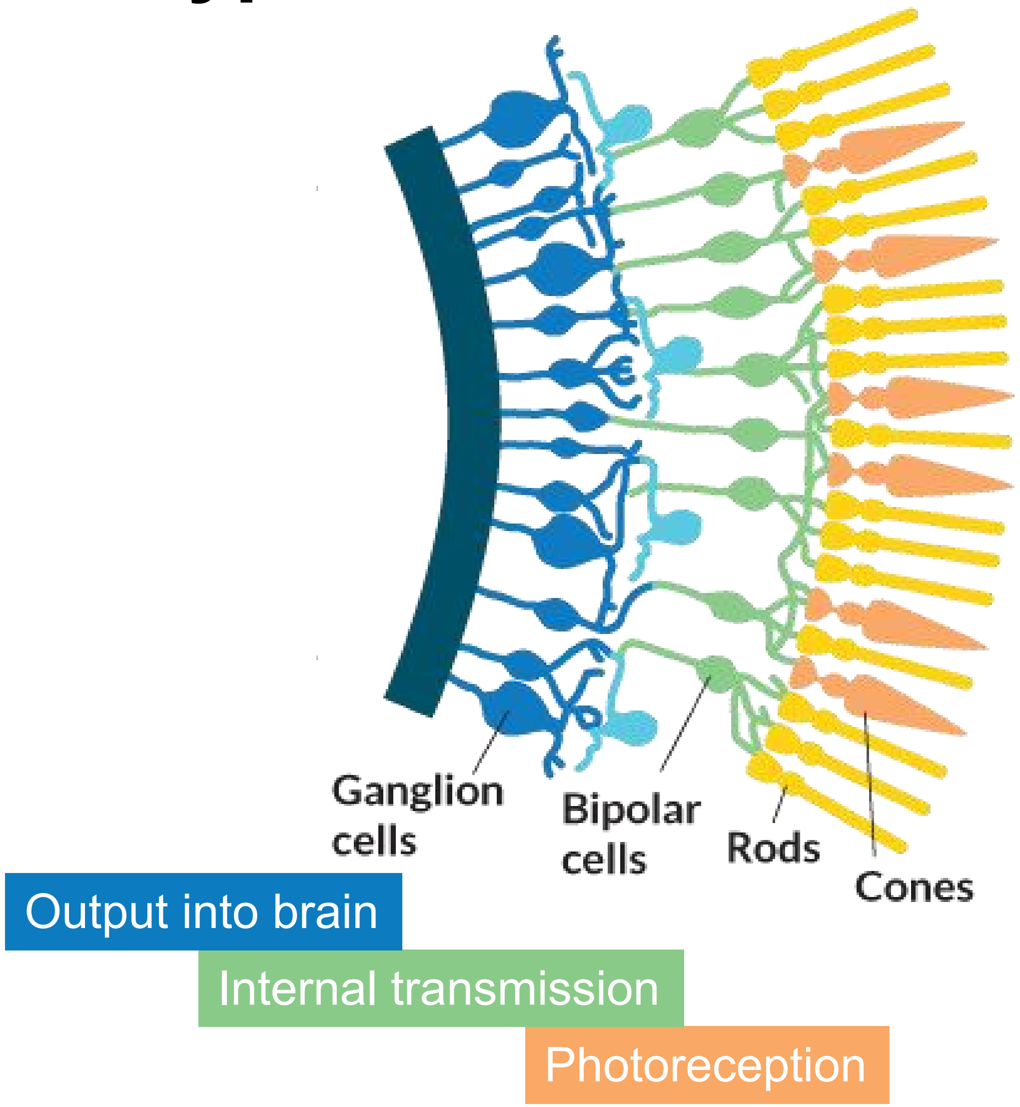
Photoreceptors: **2** types (rods and cones)

Interneurons (eg., bipolar cells): **> 50** types

Retinal ganglion cells: > 30 types

Q: Do you think this diversity is well designed by nature, or happens by chance?

We do not know.
But I personally believe it is well designed.



Arranged in an inside-out fashion

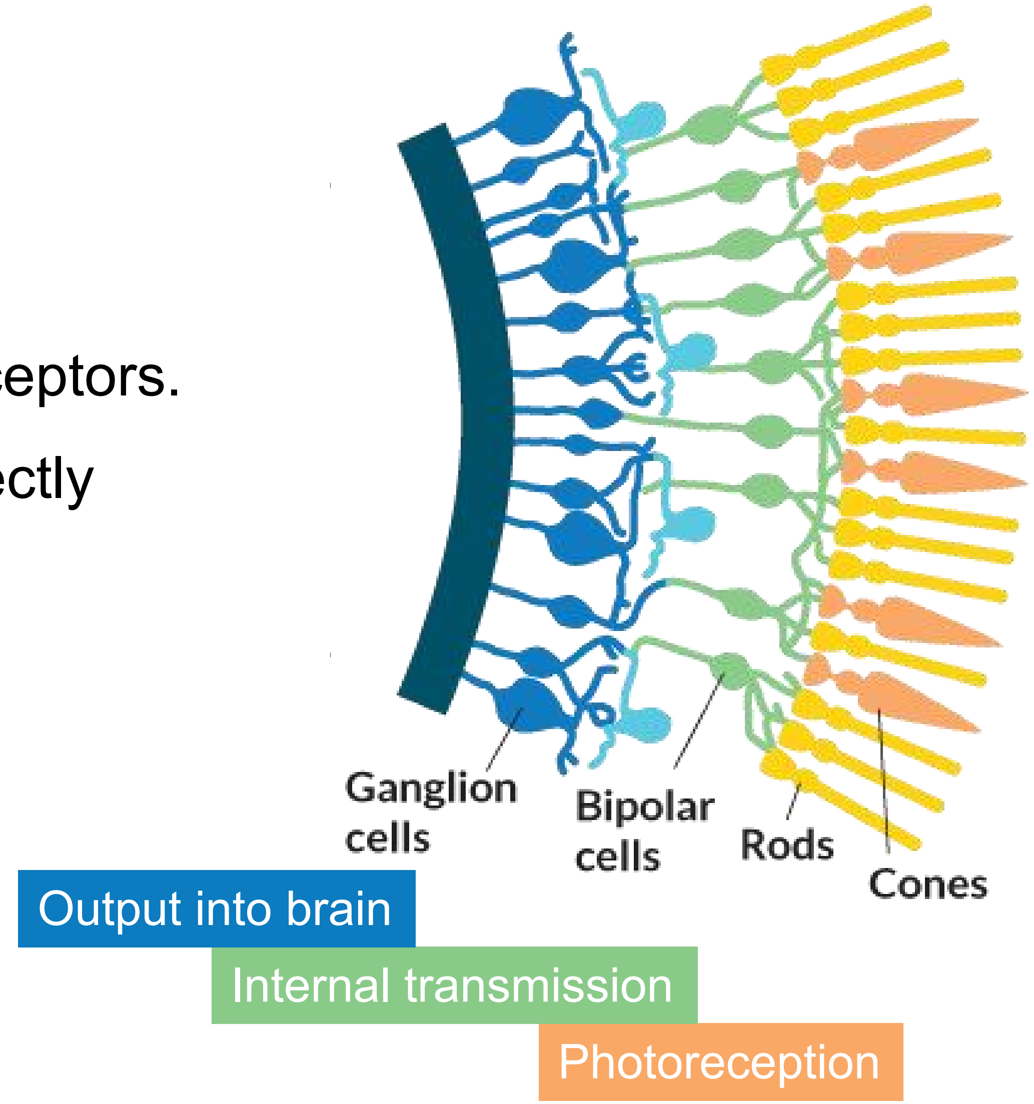
Photoreceptors

Photoreceptors: **2** types (rods and cones)

In 2002, scientists found a **new** type of photoreceptors.

The newly discovered cells turn light energy directly into brain signals.

The signals govern the body's 24-hour clock.



Arranged in an inside-out fashion

Berson et al (2002). Phototransduction by Retinal Ganglion Cells That Set the Circadian Clock, *Science*

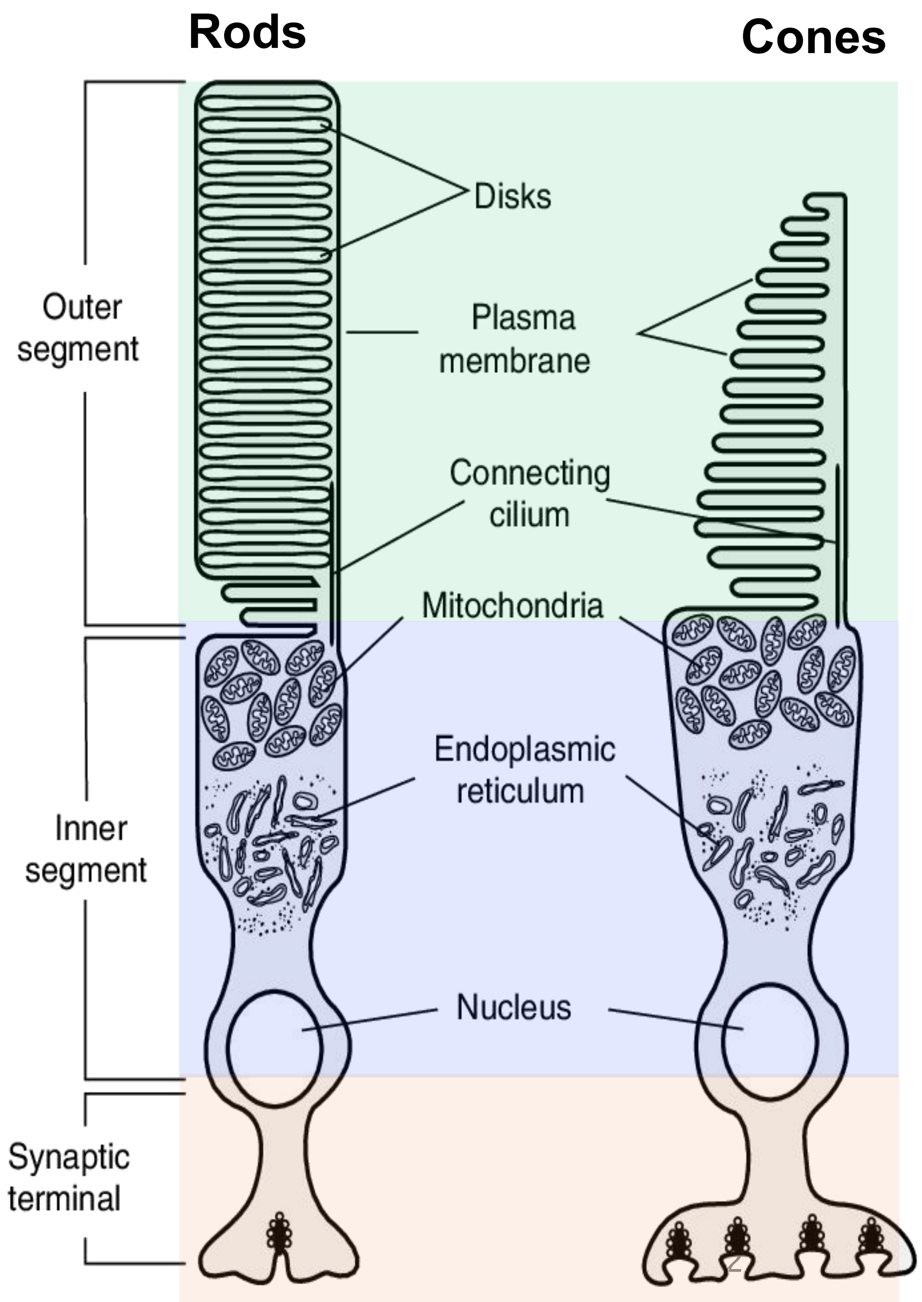
Rods and Cones

Rods (120 millions): more sensitive to light than the cones, not sensitive to color. (**night**)

Cones (6~7 millions): color sensitivity, concentrated in the central yellow spot known as the macula. (**day**)

Signal conversion

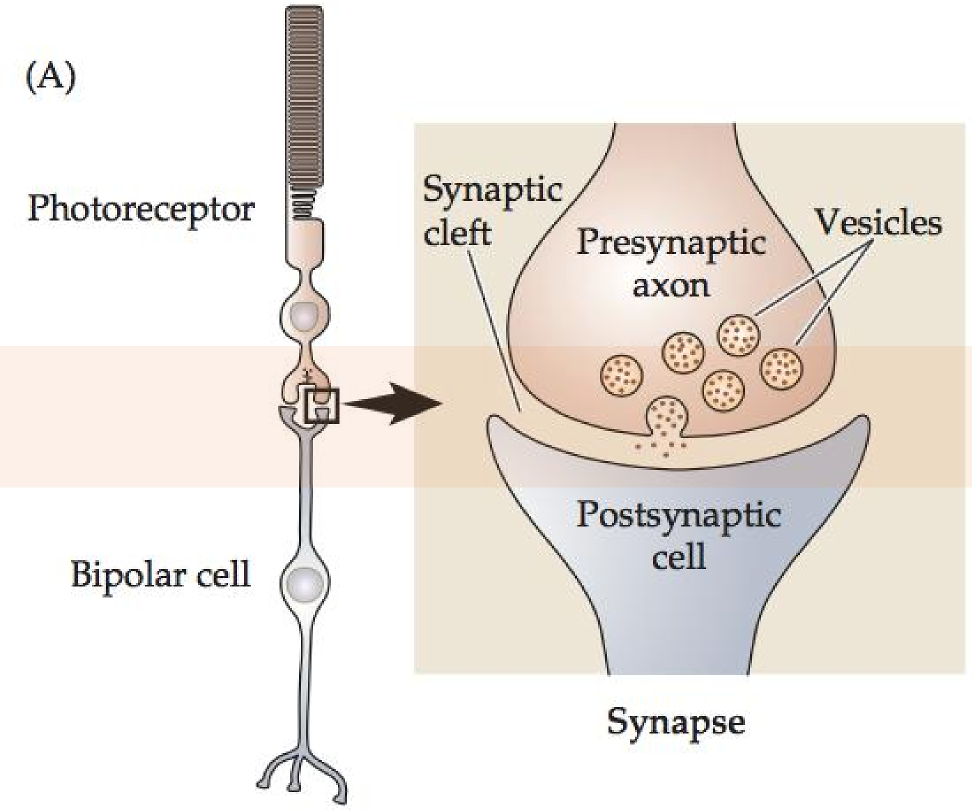
- Outer segment: Electromagnetic → Chemical
- Inner segment: Chemical → Electrical
- Synaptic terminal: Electrical → Chemical



Rods and Cones

Rods (120 millions): <u>more sensitive to light</u> than the cones, not sensitive to color. (**night**)

Cones (6~7 millions): <u>color sensitivity</u>, concentrated in the central yellow spot known as the macula. (**day**)



Signal conversion

- Outer segment: Electromagnetic → Chemical
- Inner segment: Chemical → Electrical
- Synaptic terminal: Electrical → Chemical

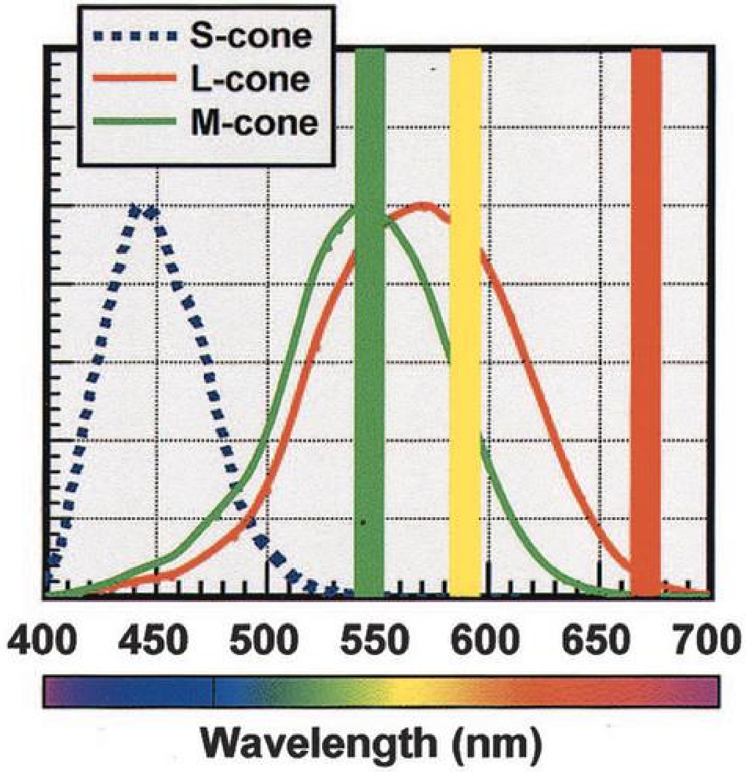
Human: one type of Rods, three types of Cones

Rods: have a peak sensitivity at 498 nm wavelength.

Short-wave cones: 420-440 nm, detecting blue light

Middle-wave cones: 530-550 nm, detecting green light

Long-wave cones: 565-580 nm, detecting red light



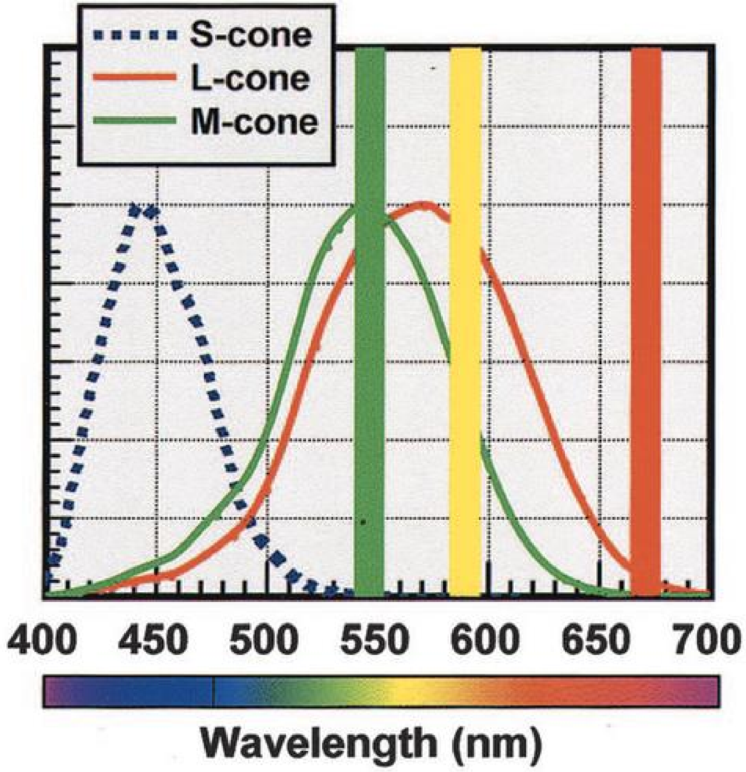
Human: one type of Rods, three types of Cones

Rods: have a peak sensitivity at 498 nm wavelength.

Short-wave cones: 420-440 nm, detecting blue light

Middle-wave cones: 530-550 nm, detecting green light

Long-wave cones: 565-580 nm, detecting red light



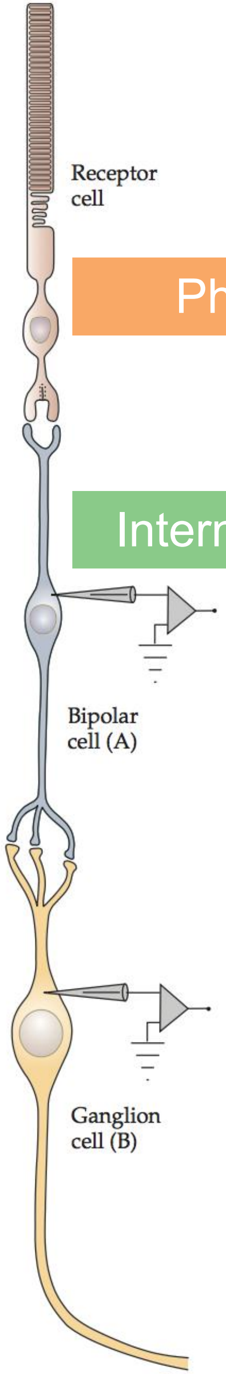
The color represented as a combination of these 3 types of cones, just as the way of **RGB color system**.

Color blindness: Lack one type of cones

Red-Green blindness lacks m-cone.

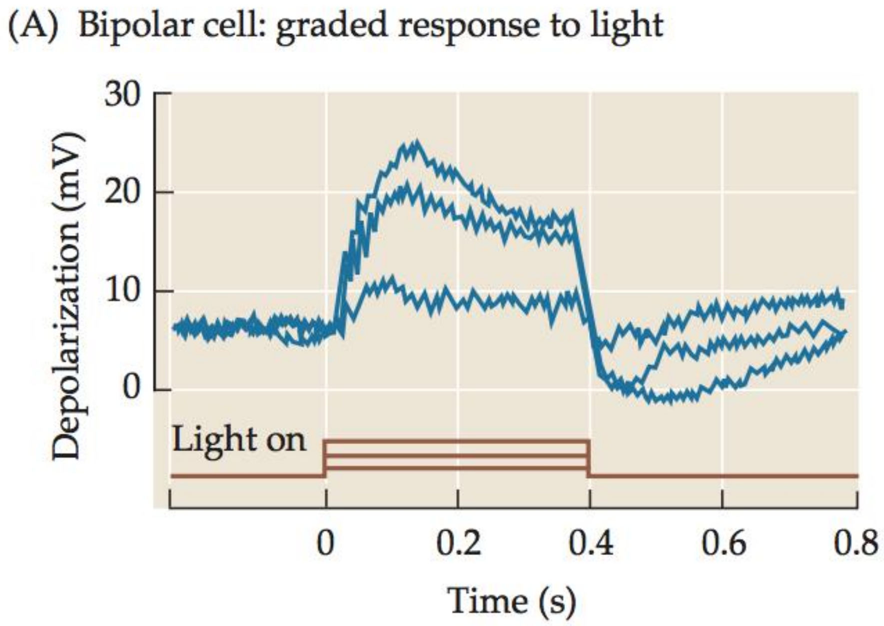
Fish and birds have 4 types of cones used for vision.

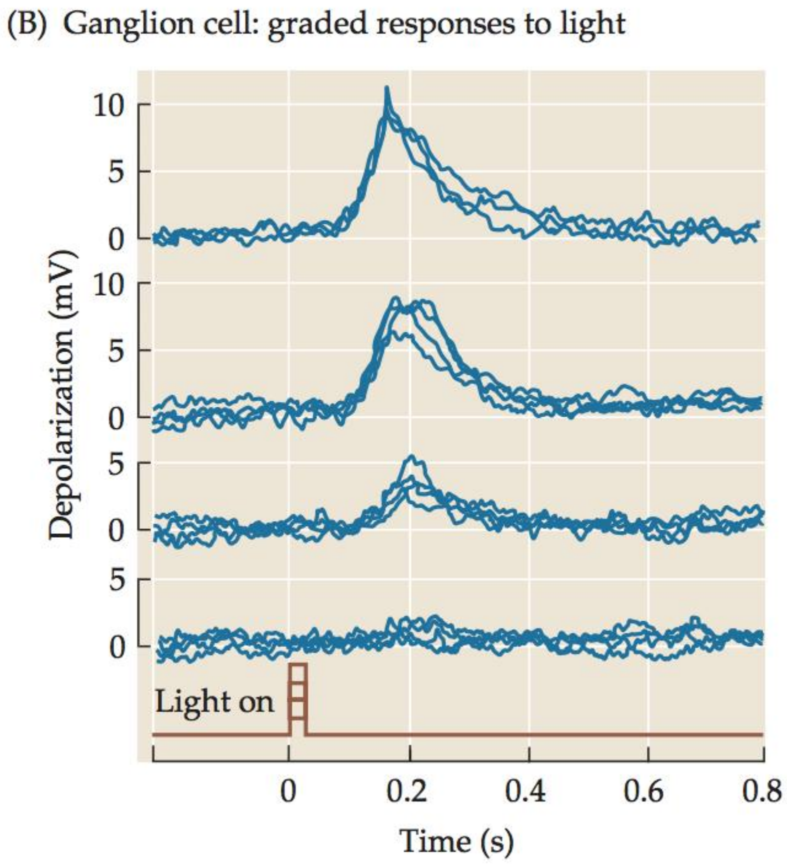
Electrical signals are generated ...



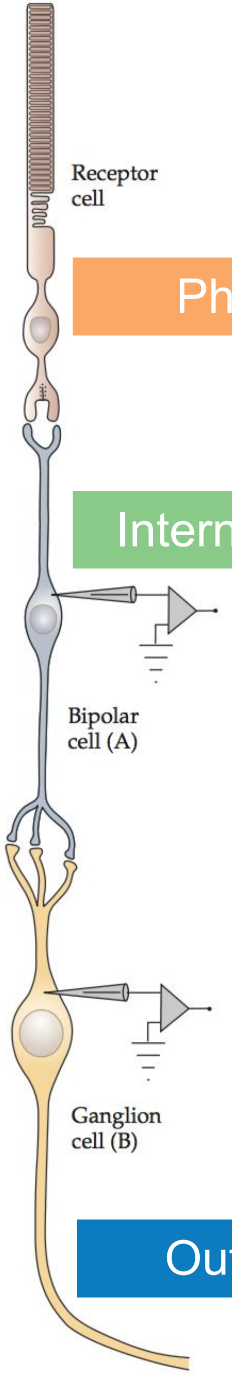
Photoreception

Internal transmission





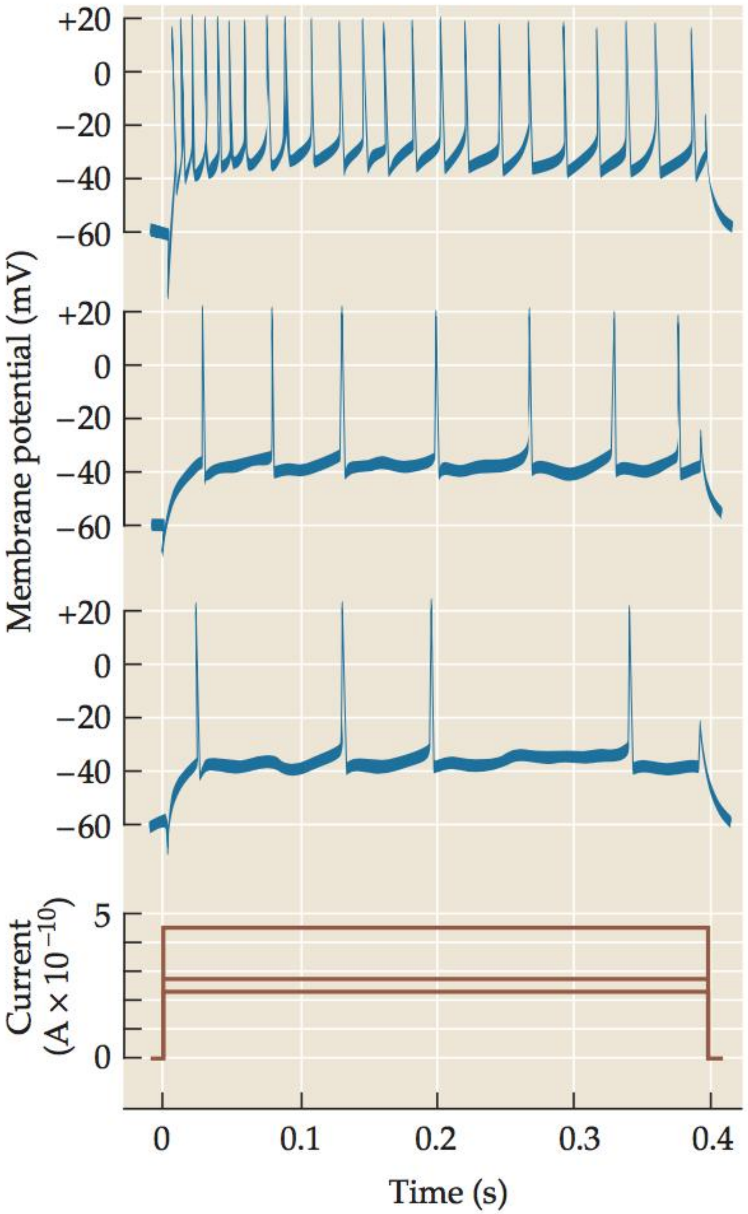
Action Potentials are generated by Ganglion cells.



Photoreception

Internal transmission

Output into brain

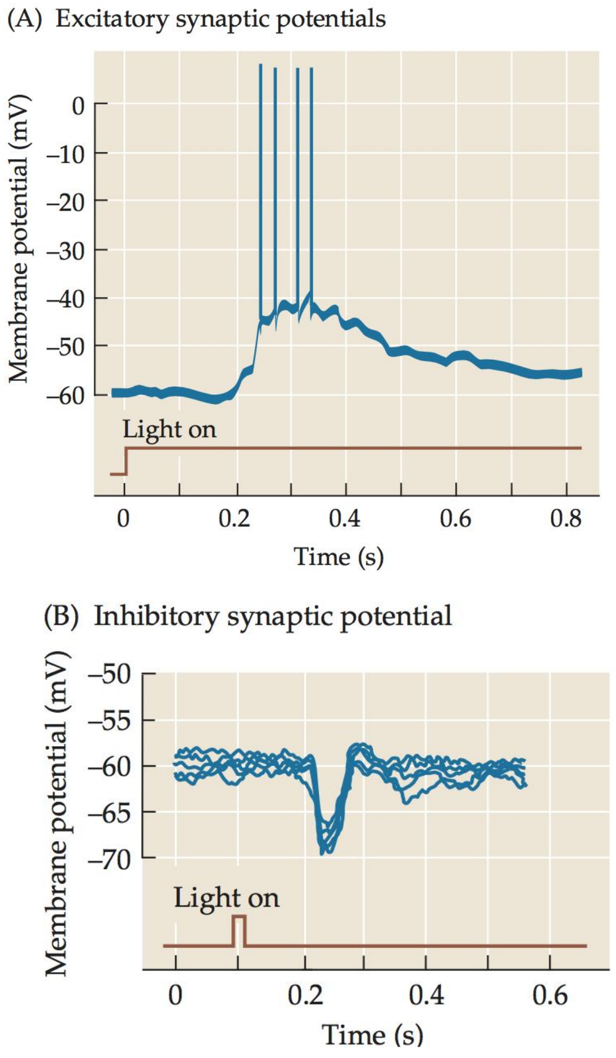


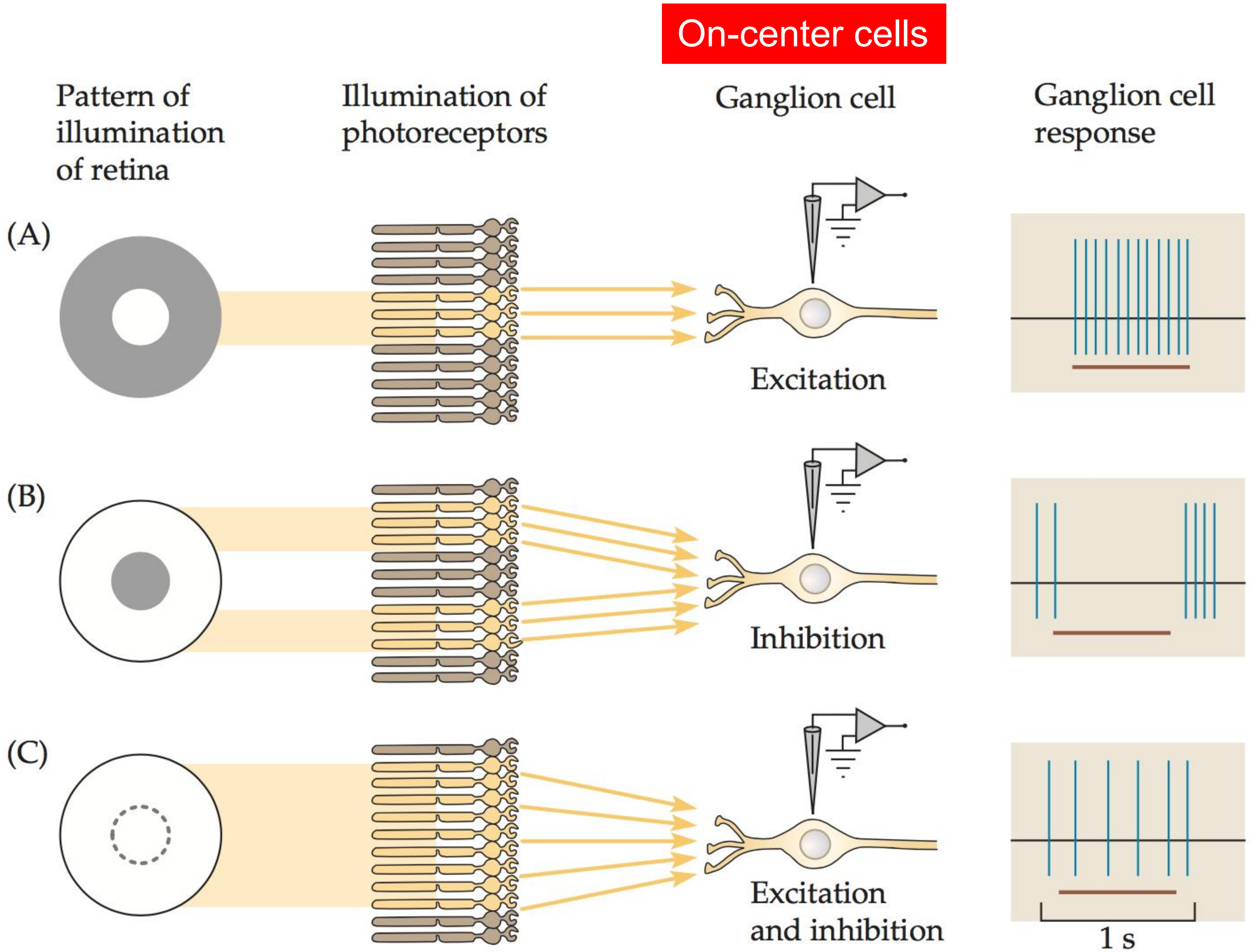
Why action potentials?

- Higher signal to noise ratio
- Travel long distance

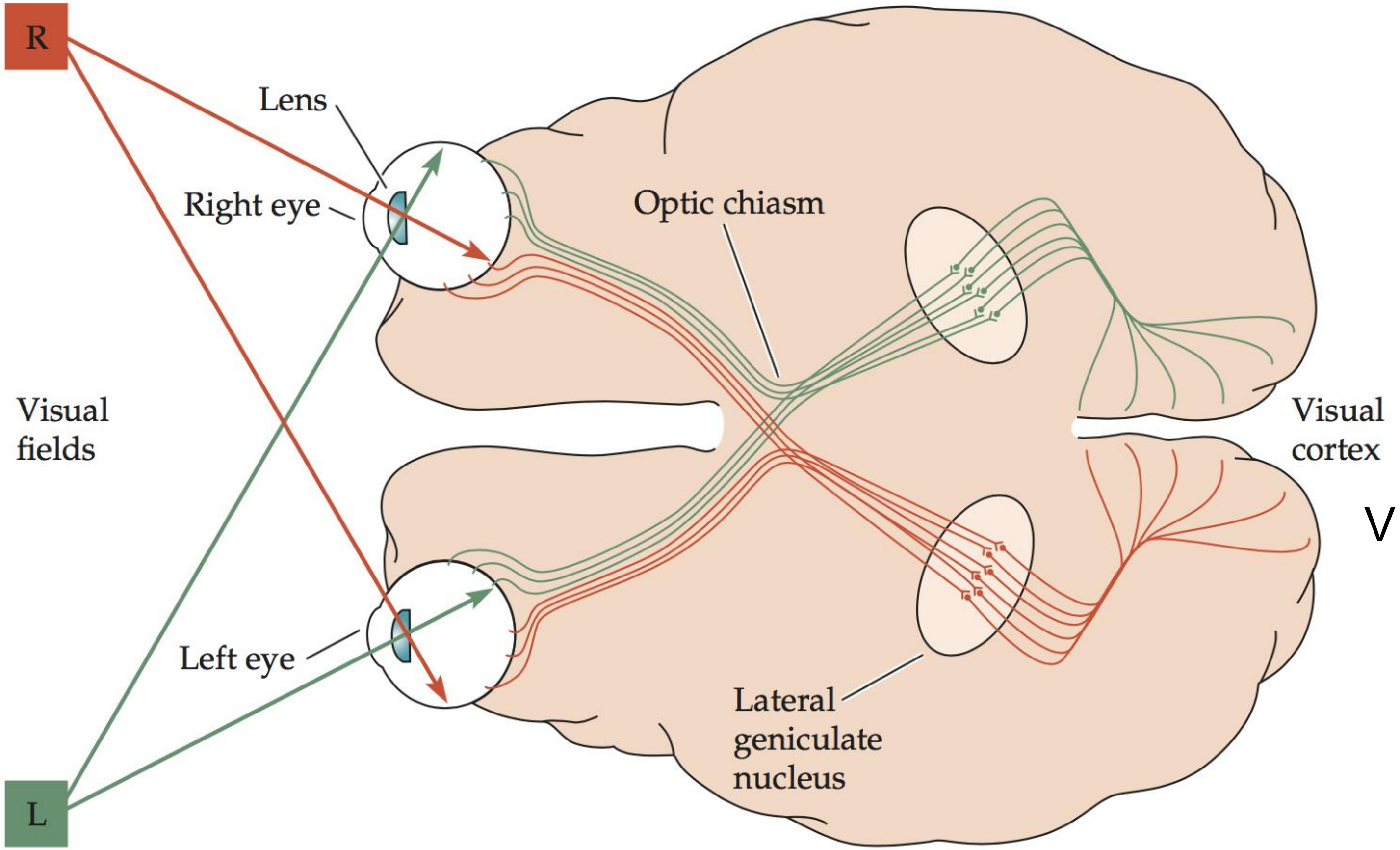
Action potentials are necessary to carry information along the length of a cell.

Information Integration by Ganglion Cell





Visual pathways: from the eye to the cerebral cortex

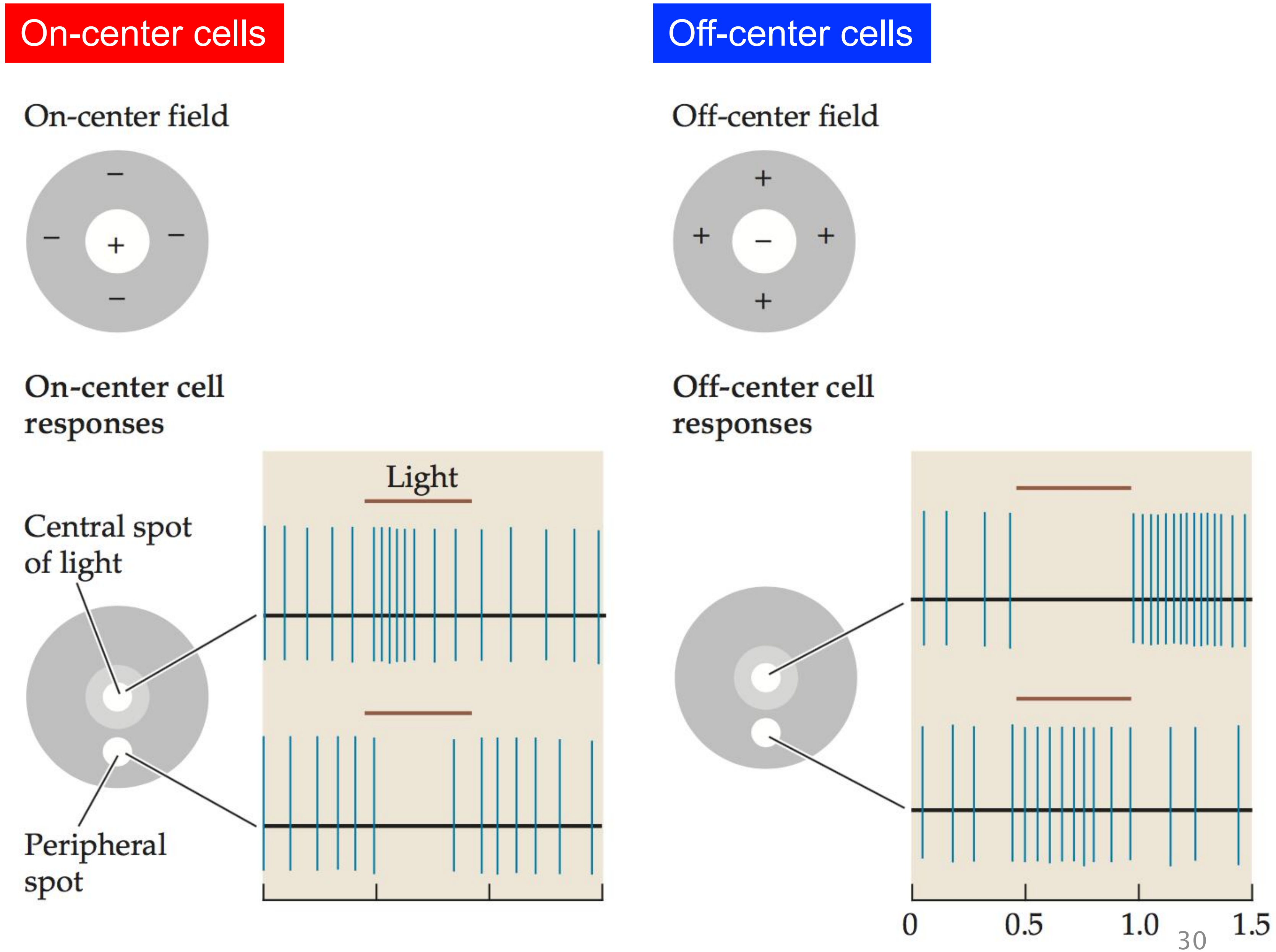


V1

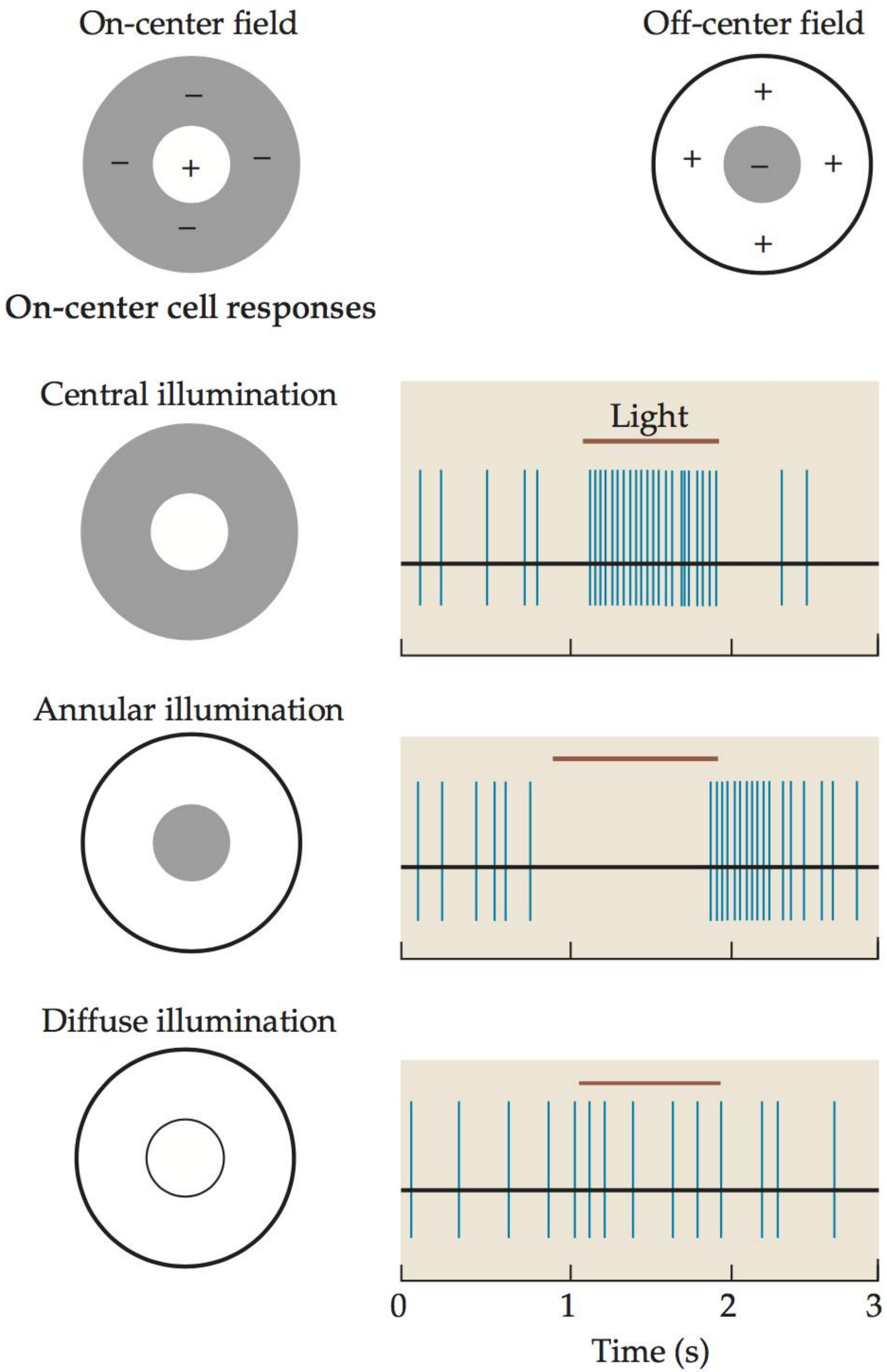
LGN in thalamus

- Rods, cones
- Interneurons
- Ganglion cells
- Optic fiber
- LGN
- V1
- Ventral / dorsal pathways

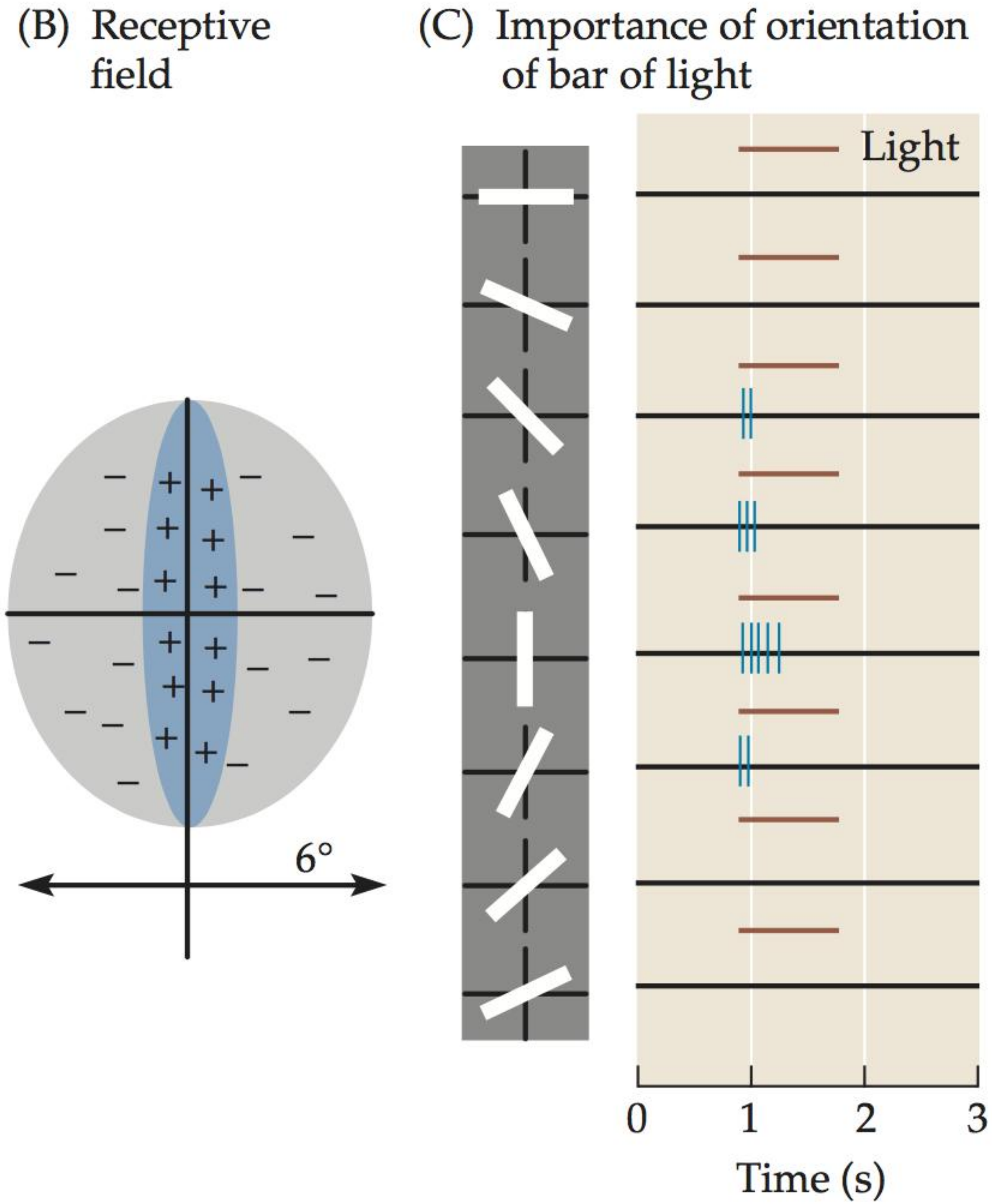
Ganglion Cells



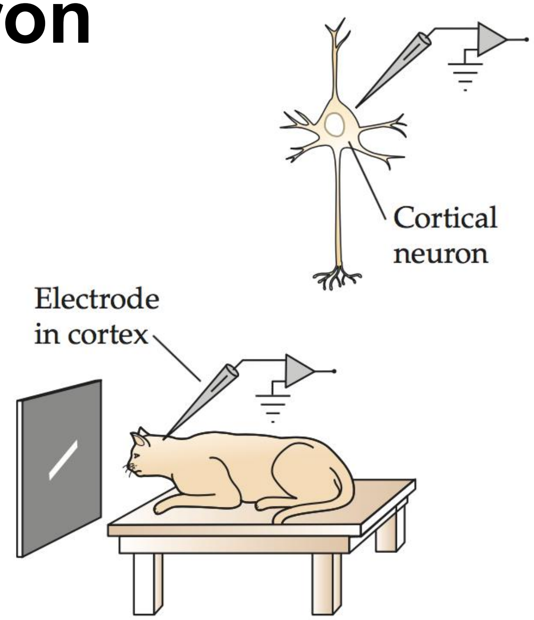
LGN Cells

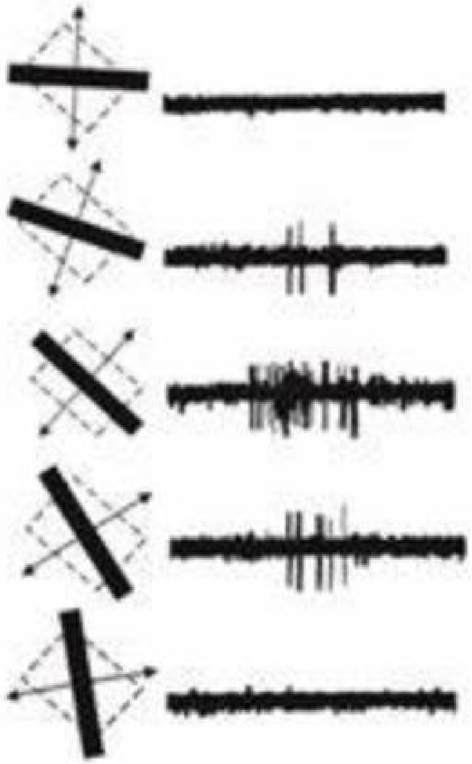

On-center field
Off-center field
On-center cell responses
Central illumination
Light
Annular illumination
Diffuse illumination
0
1
2
3
Time (s)

A Simple Cell in Cat's Striate Cortex (V1)

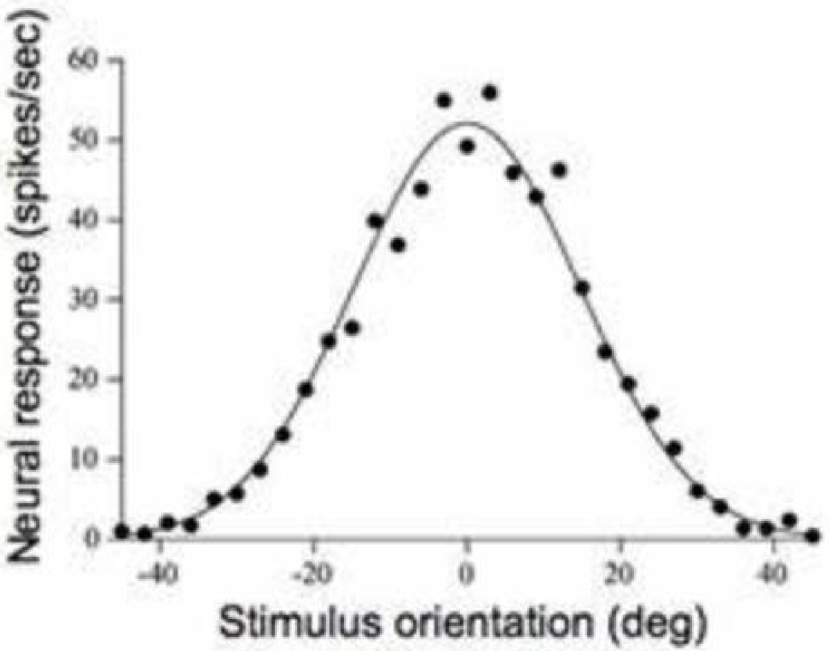


V1 neuron





Tuning curve



David H. Hubel &
Torsten N. Wiesel

Hubel, D. H. 1982. *Nature* 299: 515–524.

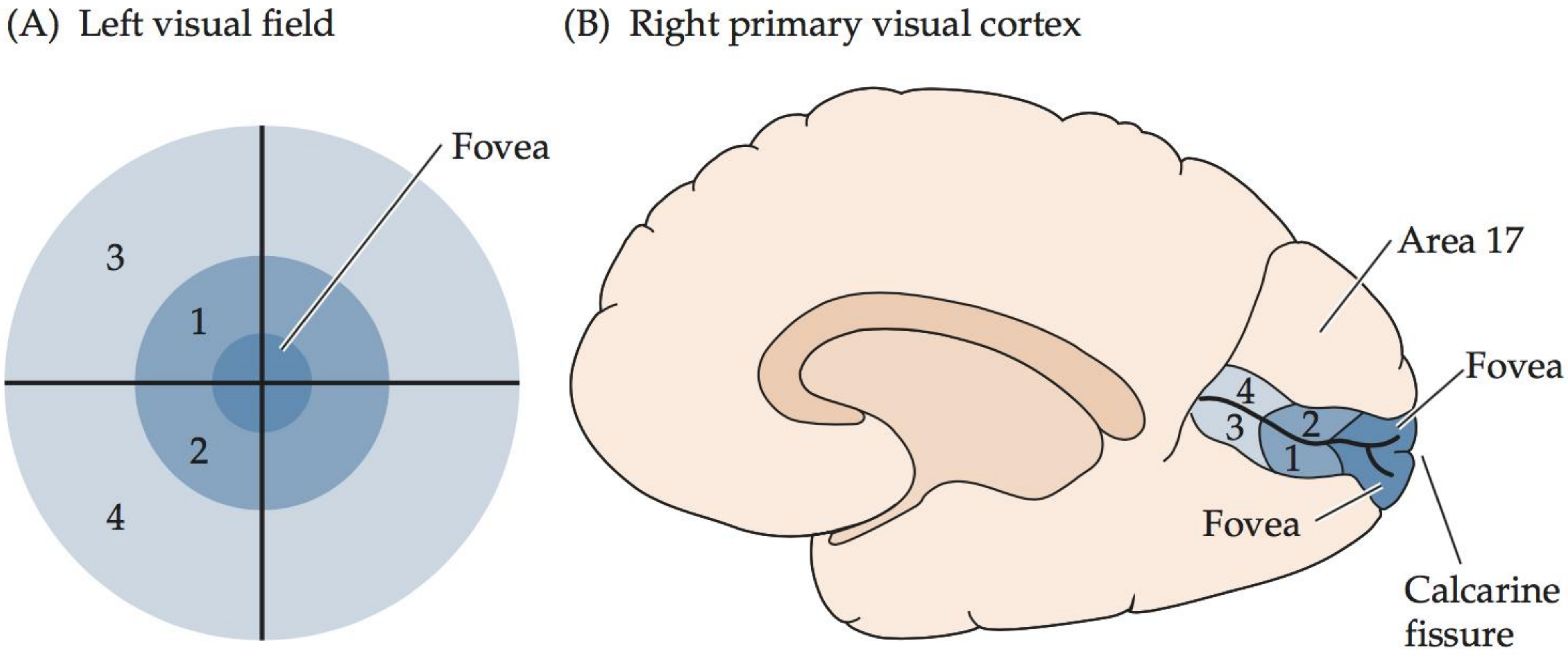
Hubel, D. H., and Wiesel, T. N. 1959. *J. Physiol.* 148: 574–591.

Hubel, D. H., and Wiesel, T. N. 1962. *J. Physiol.* 160: 106–154.

Hubel, D. H., and Wiesel, T. N. 1968. *J. Physiol.* 195: 215–243.

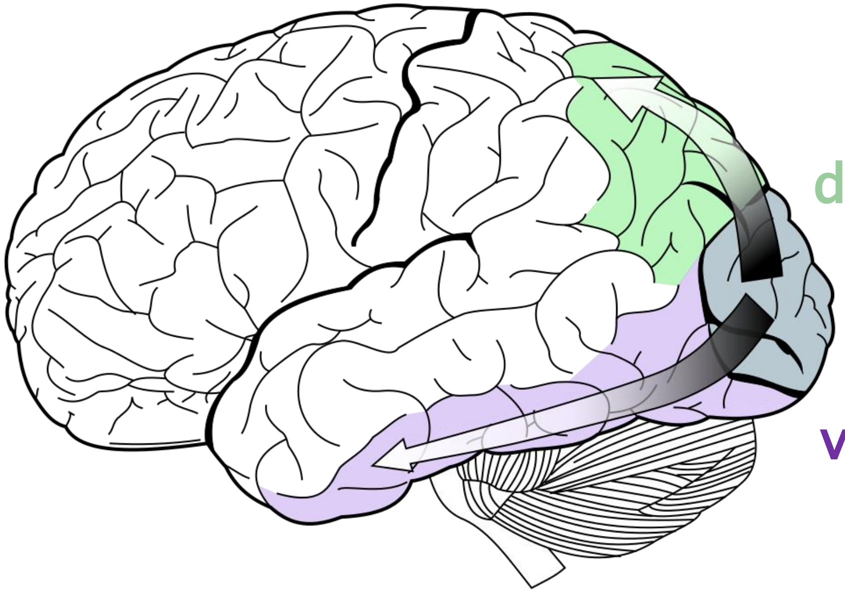
Nobel Prize for Physiology or Medicine in **1981**

Visual Field Map of the Cortex



Perceptual integration

-- the dorsal and ventral streams



dorsal 'where' pathway

- the motion & spatial location
- V1, V2, V3, MT (V5), MST & inferior parietal cortex

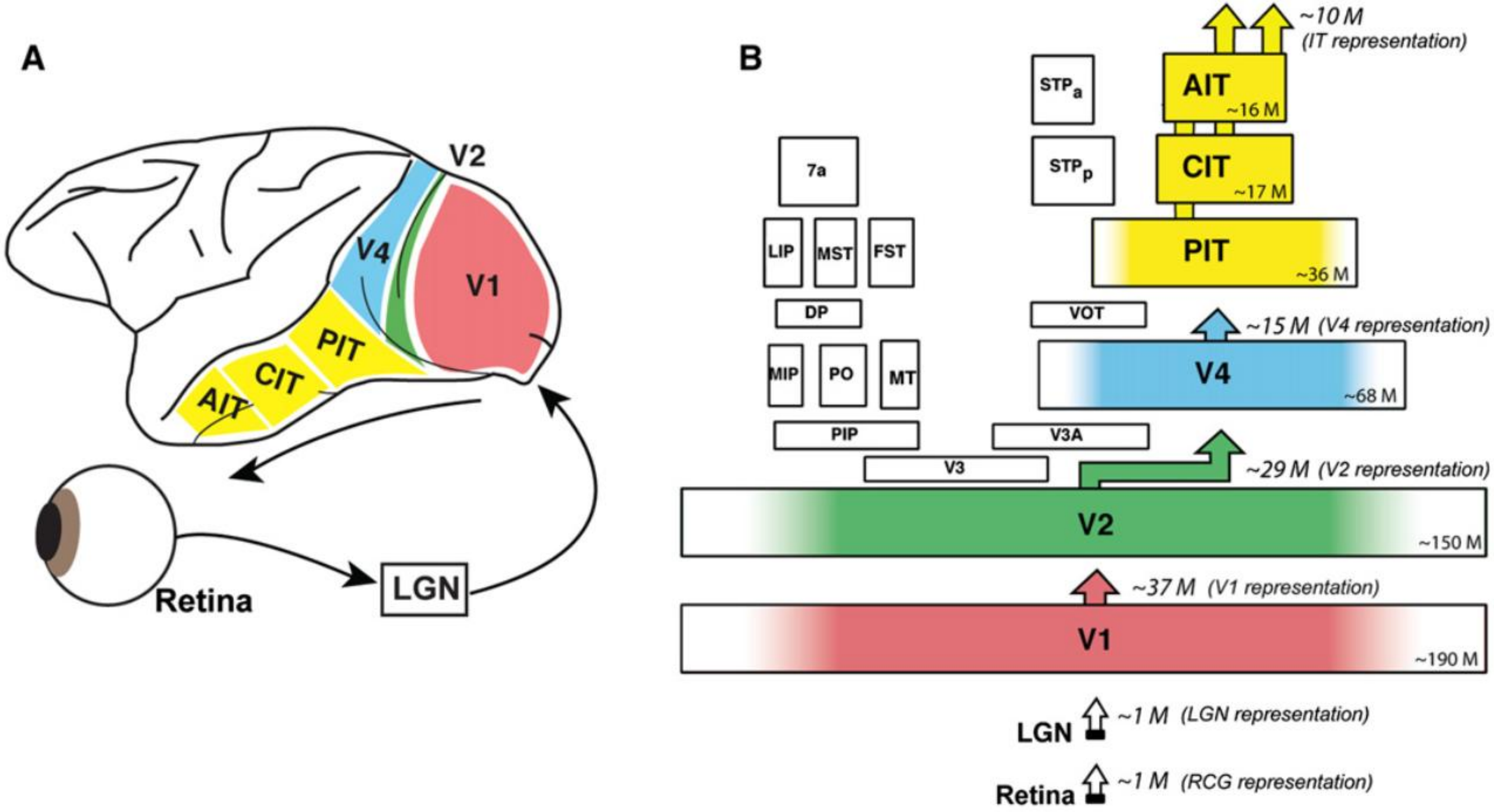
&

ventral 'what' pathway

- the detailed features, form & object identity
- V1, V2, V4 & inferior temporal areas

Goodale MA, Milner AD (1992). Separate visual pathways for perception and action. *Trends Neurosci*

The Ventral Visual Pathway: for object recognition

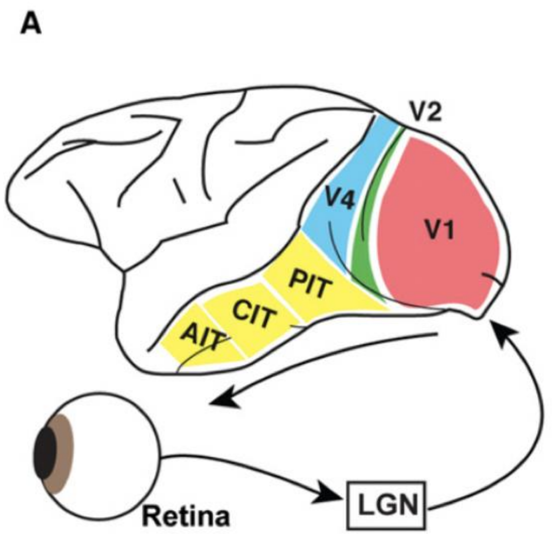


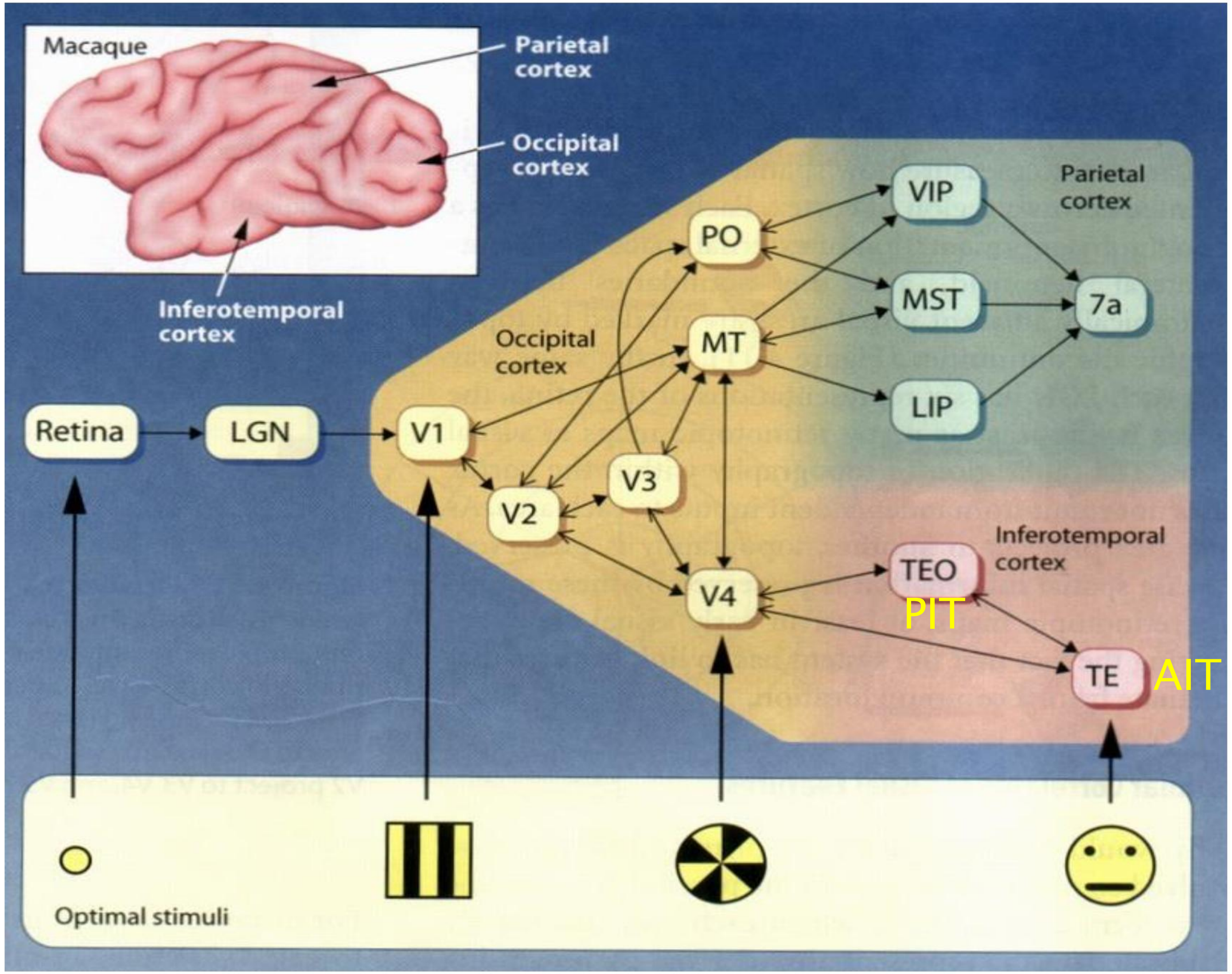
Anterior Inferior Temporal

Lateral Geniculate Nucleus

(A) Ventral stream cortical area locations in the macaque monkey brain, and flow of visual information from the retina.
(B) Each area is plotted so that its size is proportional to its cortical surface area (Felleman and Van Essen, 1991). Approximate total number of neurons (both hemispheres) is shown in the corner of each area (M = million). The approximate dimensionality of each representation (number of projection neurons) is shown above each area, based on neuronal densities (Collins et al., 2010), layer 2/3 neuronal fraction (O'Kusky and Colonnier, 1982), and portion (color) dedicated to processing the central 10 deg of the visual field (Brewer et al., 2002). Approximate median response latency is listed on the right (Nowak and Bullier, 1997; Schmolesky et al., 1998).

Along the Ventral Visual Pathway, feature extraction from simple to complex.





深度学习简介

深度学习的一些基本概念

- 训练**有监督学习**模型，需要大量的 (输入,输出) 数据pair
- 构造一个损失函数
- 通过Gradient Descent更新参数，使得损失函数最小化
- DL的学习过程，就是使损失函数最小的过程。最佳的参数，得到最小的损失

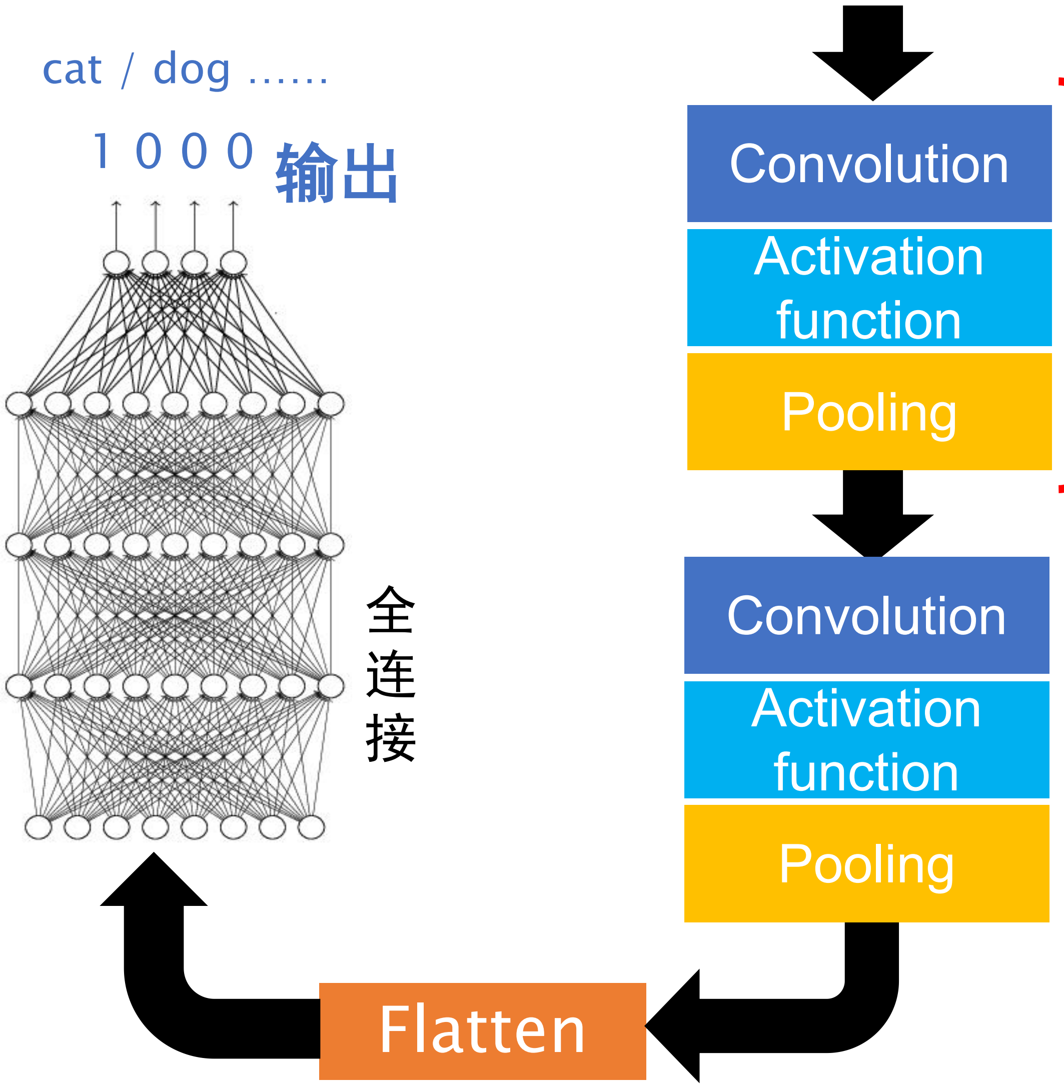
深度学习的主要研究内容

- 数据（input, output）
- 模型（网络架构）
- 损失函数（如何定义损失）
- 学习算法（如何优化损失）

输入

cat / dog ......

1 0 0 0 输出



These blocks can repeat many times

Some Activation Functions

1. Binary Step Function
2. Sigmoid / Logistic
3. TanH / Hyperbolic Tangent
4. Softmax
5. ReLU (Rectified Linear Unit)
6. Leaky ReLU
7. Parametric ReLU
8. …

The encoding / decoding model

External world (image, sound … stimuli)

Neural data (10^{12} neurons, 10^2 brain regions)

Cognitive process

Senses (感觉)
Motion (运动)
Emotion (情绪)
Attention (注意力)
Cognition (认知)

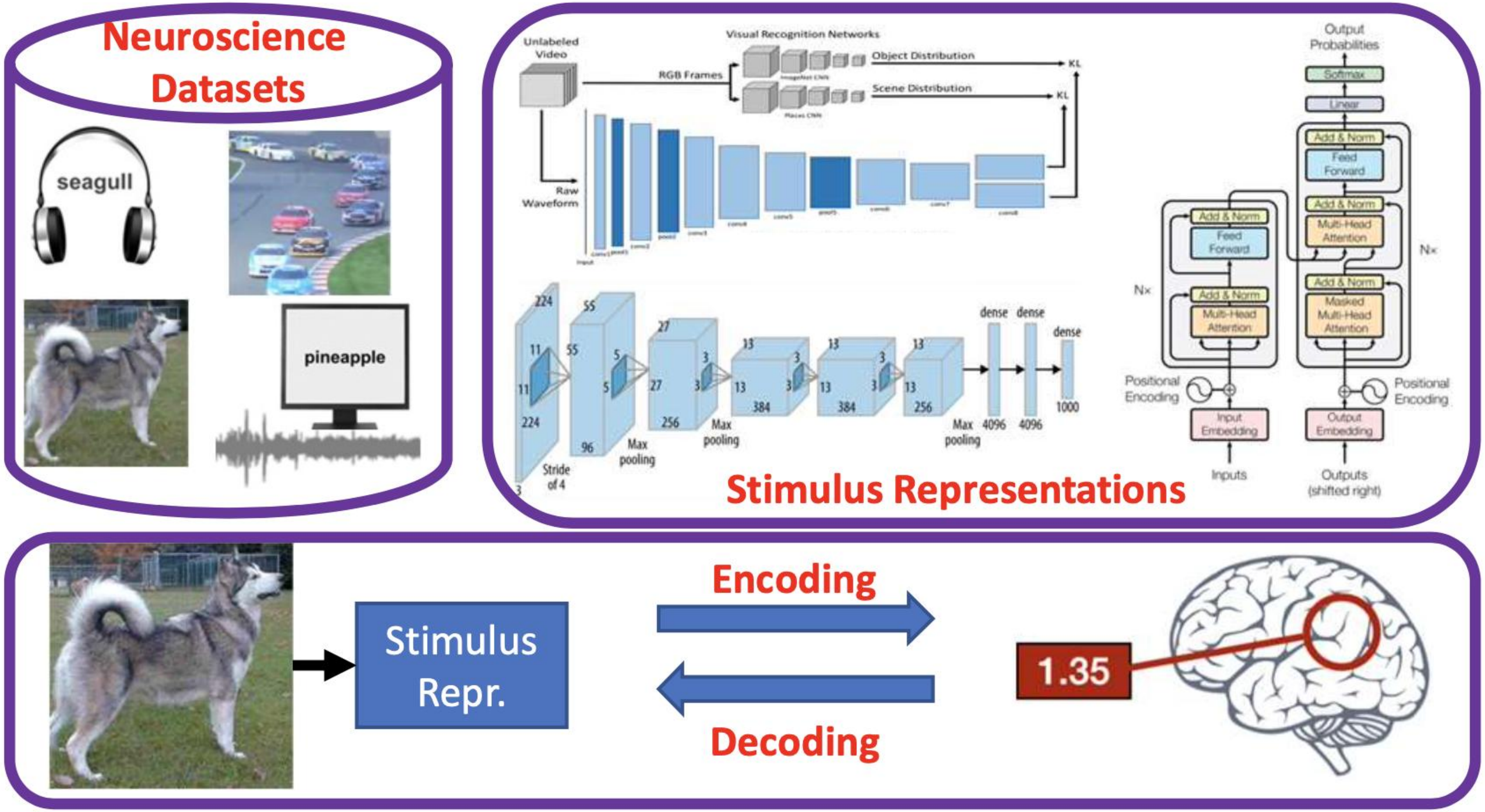
Neural encoding model: $dx/dt = f(x,u)$

- Given the input u, and current neural data $x(t)$, predict the future dynamics $x(t+1)$

Neural decoding model: $s = g(x)$

- Given neural data x, decode the cognitive state s
- **Brain-computer interfaces (BCI)** largely rely on the neural decoding model.
- There are various approaches for neural decoding, depending on the design of task, the brain areas involved, etc.

Brain encoding / Brain decoding



数据集

stimulus-neural data

刺激表征模型

stimulus-->vector

编码模型

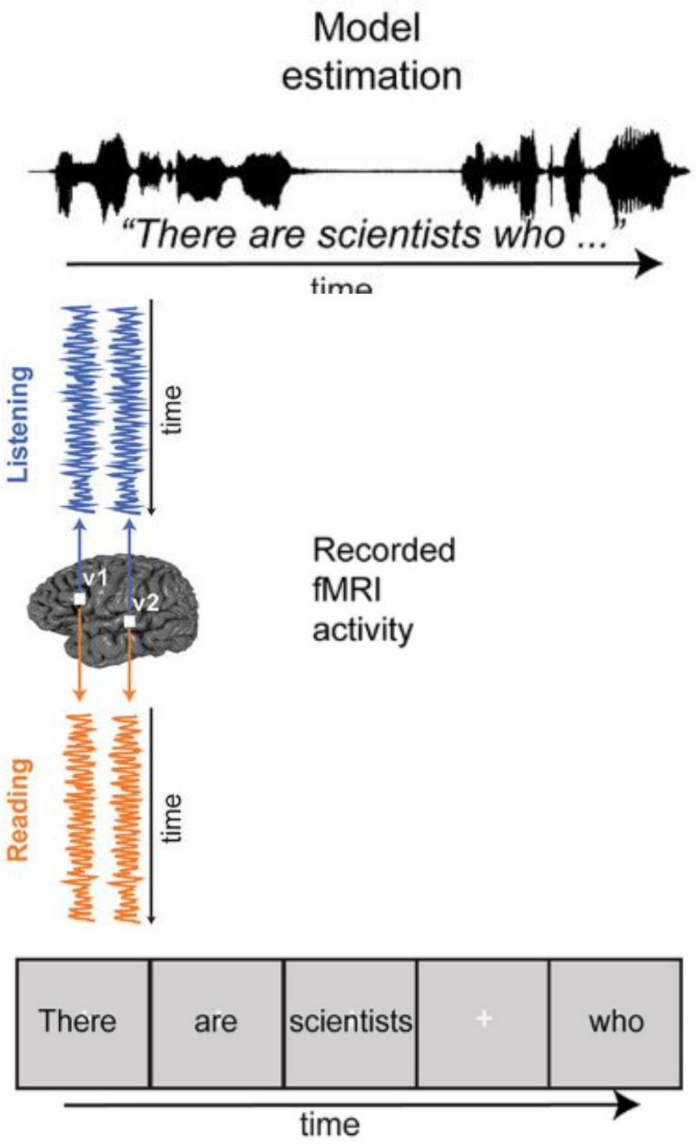
刺激表征-->神经信号

解码模型

神经信号-->刺激表征

Oota et al (2023) Deep Neural Networks and Brain Alignment: Brain Encoding and Decoding

Naturalistic Brain Dataset



BoLD5000

ImageNet

SSfMRI-1250 Images

VIM-1 1870-Images

自然刺激材料

文本

图片

音频

视频

刺激类型

特定语义

特定情绪

特定视觉特征

Naturalistic Brain Dataset

	Dataset	Authors	Type	Lang.	Stimulus	|S|	Task
Text	Harry Potter	[Wehbe *et al.*, 2014]	fMRI/ MEG	English	Reading Chapter 9 of Harry Potter and the Sorcerer's Stone	9	Story understanding
		[Handjaras *et al.*, 2016]	fMRI	Italian	Verbal, pictorial or auditory presentation of 40 concrete nouns, four times	20	Property Generation
		[Anderson *et al.*, 2017a]	fMRI	Italian	Reading 70 concrete and abstract nouns from law/music, five times	7	Imagine a situation with noun
	ZuCo	[Hollenstein *et al.*, 2018]	EEG	English	Reading 1107 sentences with 21,629 words from movie reviews	12	Rate movie quality
	240 Sentences with Content Words	[Anderson *et al.*, 2019]	fMRI	English	Reading 240 active voice sentences describing everyday situations	14	Passive reading
	BCCWJ-EEG	[Oseki and Asahara, 2020]	EEG	Japanese	Reading 20 newspaper articles for ∼30-40 minutes	40	Passive reading
	Subset Moth Radio Hour	[Deniz *et al.*, 2019]	fMRI	English	Reading 11 stories	9	Passive reading and Listening
Visual		[Thirion *et al.*, 2006]	fMRI	-	Viewing rotating wedges (8 times), expanding/contracting rings (8 times), rotating 36 Gabor filters (4 times), grid (36 times)	9	Passive viewing
	Vim-1	[Kay *et al.*, 2008]	fMRI	-	Viewing sequences of 1870 natural photos	2	Passive viewing
	Generic Object Decoder	[Horikawa and Kamitani, 2017]	fMRI	-	Viewing 1,200 images from 150 object categories; 50 images from 50 object categories; imagery 10 times	5	Repetition detection
	BOLD5000	[Chang *et al.*, 2019]	fMRI	-	Viewing 5254 images depicting real-world scenes	4	Passive viewing
	Algonauts	[Cichy *et al.*, 2019]	fMRI/ MEG	-	Viewing 92 silhouette object images and 118 images of objects on natural background	15	Passive viewing
	NSD	[Allen *et al.*, 2022]	fMRI	-	Viewing 73000 natural scenes	8	Passive viewing
	THINGS	[Hebart *et al.*, 2022]	fMRI/ MEG	-	Viewing 31188 natural images	8	Passive viewing
Audio		[Handjaras *et al.*, 2016]	fMRI	Italian	Verbal, pictorial or auditory presentation of 40 concrete nouns, 4 times	20	Property Generation
	The Moth Radio Hour	[Huth *et al.*, 2016]	fMRI	English	Listening eleven 10-minute stories	7	Passive Listening
		[Brennan and Hale, 2019]	EEG	English	Listening Chapter one of Alice's Adventures in Wonderland (2,129 words in 84 sentences) as read by Kristen McQuillan	33	Question answering
		[Anderson *et al.*, 2020]	fMRI	English	Listening one of 20 scenario names, 5 times	26	Imagine personal experiences
	Narratives	[Nastase *et al.*, 2021]	fMRI	English	Listening 27 diverse naturalistic spoken stories. 891 functional scans	345	Passive Listening
	Natural Stories	[Zhang *et al.*, 2020]	fMRI	English	Listening Moth-Radio-Hour naturalistic spoken stories.	19	Passive Listening
	The Little Prince	[Li *et al.*, 2021]	fMRI	English	Listening audiobook for about 100 minutes.	112	Passive Listening
	MEG-MASC	[Gwilliams *et al.*, 2022]	MEG	English	Listening two hours of naturalistic stories. 208 MEG sensors	27	Passive Listening
Video	BBC's Doctor Who	[Seeliger *et al.*, 2019]	fMRI	English	Viewing spatiotemporal visual and auditory videos (30 episodes). 120.8 whole-brain volumes (∼23 h) of single-presentation data, and 1.2 volumes (11 min) of repeated narrative short episodes. 22 repetitions	1	Passive viewing
	Japanese Ads	[Nishida *et al.*, 2020]	fMRI	Japanese	Viewing 368 web and 2452 TV Japanese ad movies (15-30s). 7200 train and 1200 test fMRIs for web; fMRIs from 420 ads.	52	Passive viewing
	Pippi Langkous	[Berezutskaya *et al.*, 2020]	ECoG	Swedish/ Dutch	Viewing 30 s excerpts of a feature film (in total, 6.5 min long), edited together for a coherent story	37	Passive viewing
	Algonauts	[Cichy *et al.*, 2021]	fMRI	English	Viewing 1000 short video clips (3 sec each)	10	Passive viewing
	Natual Short Clips	[Huth *et al.*, 2022]	fMRI	English	Watching natural short movie clips	5	Passive viewing
	Natual Short Clips	[Lahner *et al.*, 2023]	fMRI	English	Watching 1102 natural short video clips	10	Passive viewing
Other Multimodal	60 Concrete Nouns	[Mitchell *et al.*, 2008]	fMRI	English	Viewing 60 different word-picture pairs from 12 categories, 6 times each	9	Passive viewing
		[Sudre *et al.*, 2012]	MEG	English	Reading 60 concrete nouns along with line drawings. 20 questions per noun lead to 1200 examples.	9	Question answering
		[Zinszer *et al.*, 2018]	fNIRS	English	8 concrete nouns (audiovisual word and picture stimuli): bunny, bear, kitty, dog, mouth, foot, hand, and nose; 12 times repeated.	24	Passive viewing and listening
	Pereira	[Pereira *et al.*, 2018]	fMRI	English	Viewing 180 Words with Picture, Sentences, word clouds; reading 96 text passages; 72 passages. 3 times repeated.	16	Passive viewing and reading
		[Cao *et al.*, 2021]	fNIRS	Chinese	Viewing and listening 50 concrete nouns from 10 semantic categories.	7	Passive viewing and listening
	Neuromod	[Boyle *et al.*, 2020]	fMRI	English	Watching TV series (Friends, Movie10)	6	Passive viewing and listening

Naturalistic Brain Dataset

	Dataset	Authors	Type	Lang.	Stimulus	|S|	Task
Text	Harry Potter	[Wehbe *et al.*, 2014]	fMRI/ MEG	English	Reading Chapter 9 of Harry Potter and the Sorcerer's Stone	9	Story understanding
		[Handjaras *et al.*, 2016]	fMRI	Italian	Verbal, pictorial or auditory presentation of 40 concrete nouns, four times	20	Property Generation
		[Anderson *et al.*, 2017a]	fMRI	Italian	Reading 70 concrete and abstract nouns from law/music, five times	7	Imagine a situation with noun
	ZuCo	[Hollenstein *et al.*, 2018]	EEG	English	Reading 1107 sentences with 21,629 words from movie reviews	12	Rate movie quality
	240 Sentences with Content Words	[Anderson *et al.*, 2019]	fMRI	English	Reading 240 active voice sentences describing everyday situations	14	Passive reading
	BCCWJ-EEG	[Oseki and Asahara, 2020]	EEG	Japanese	Reading 20 newspaper articles for ~30-40 minutes	40	Passive reading
Video	BBC's Doctor Who	[Seeliger *et al.*, 2019]	fMRI	English	Viewing spatiotemporal visual and auditory videos (30 episodes). 120.8 whole-brain volumes (~23 h) of single-presentation data, and 1.2 volumes (11 min) of repeated narrative short episodes. 22 repetitions	1	Passive viewing
	Japanese Ads	[Nishida *et al.*, 2020]	fMRI	Japanese	Viewing 368 web and 2452 TV Japanese ad movies (15-30s). 7200 train and 1200 test fMRIs for web; fMRIs from 420 ads.	52	Passive viewing
	Pippi Langkous	[Berezutskaya *et al.*, 2020]	ECoG	Swedish/ Dutch	Viewing 30 s excerpts of a feature film (in total, 6.5 min long), edited together for a coherent story	37	Passive viewing
	Algonauts	[Cichy *et al.*, 2021]	fMRI	English	Viewing 1000 short video clips (3 sec each)	10	Passive viewing
	Natual Short Clips	[Huth *et al.*, 2022]	fMRI	English	Watching natural short movie clips	5	Passive viewing
	Natual Short Clips	[Lahner *et al.*, 2023]	fMRI	English	Watching 1102 natural short video clips	10	Passive viewing
Other Multimodal	60 Concrete Nouns	[Mitchell *et al.*, 2008]	fMRI	English	Viewing 60 different word-picture pairs from 12 categories, 6 times each	9	Passive viewing
		[Sudre *et al.*, 2012]	MEG	English	Reading 60 concrete nouns along with line drawings. 20 questions per noun lead to 1200 examples.	9	Question answering
		[Zinszer *et al.*, 2018]	fNIRS	English	8 concrete nouns (audiovisual word and picture stimuli): bunny, bear, kitty, dog, mouth, foot, hand, and nose; 12 times repeated.	24	Passive viewing and listening
	Pereira	[Pereira *et al.*, 2018]	fMRI	English	Viewing 180 Words with Picture, Sentences, word clouds; reading 96 text passages; 72 passages. 3 times repeated.	16	Passive viewing and reading
		[Cao *et al.*, 2021]	fNIRS	Chinese	Viewing and listening 50 concrete nouns from 10 semantic categories.	7	Passive viewing and listening
	Neuromod	[Boyle *et al.*, 2020]	fMRI	English	Watching TV series (Friends, Movie10)	6	Passive viewing and listening

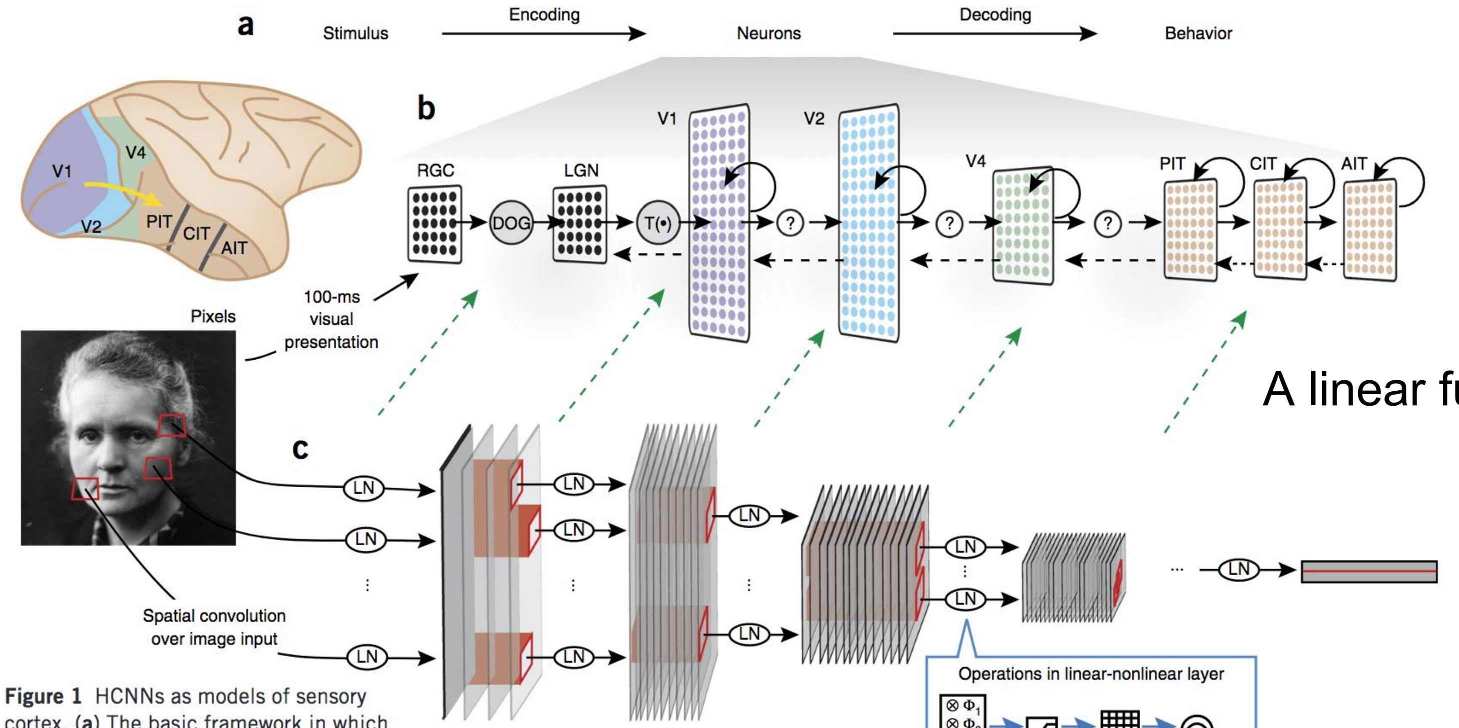
What can we do with these datasets?

Any ideas?

Q1:

Can we predict the neural signals based on the input image?

Image → neural activity

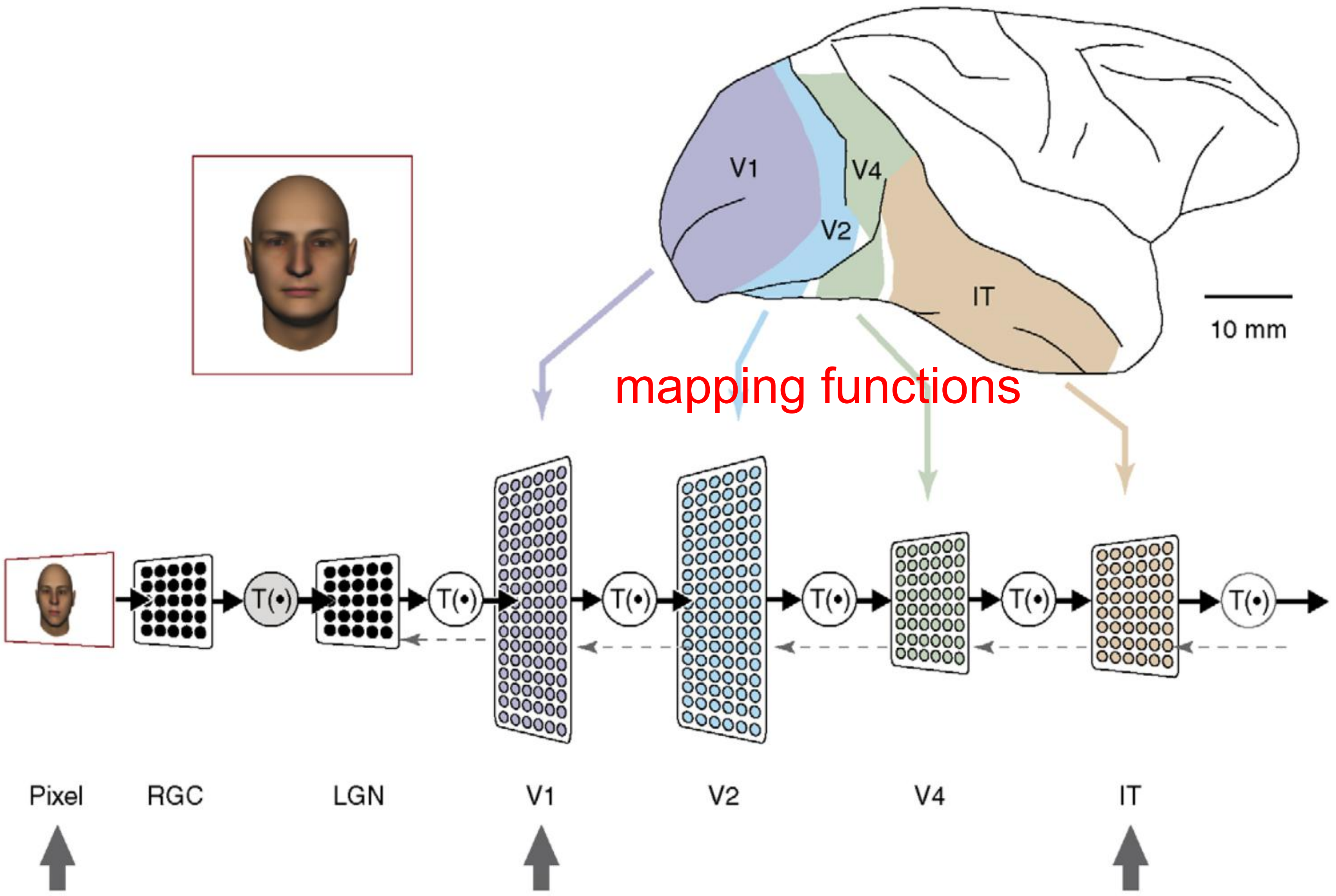


A linear function

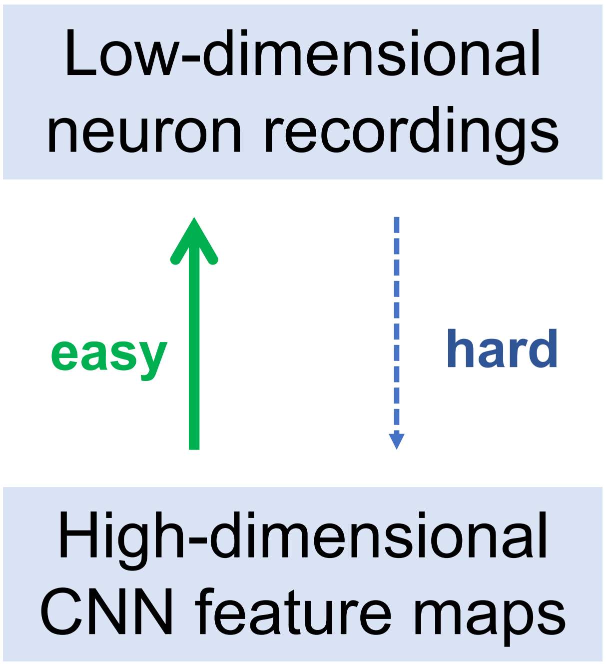
Figure 1 HCNNs as models of sensory cortex. (**a**) The basic framework in which sensory cortex is studied is one of encoding—the process by which stimuli are transformed into patterns of neural activity—and decoding, the process by which neural activity generates behavior. HCNNs have been used to make models of the encoding step; that is, they describe the mapping of stimuli to neural responses as measured in brain. (**b**) The ventral visual pathway is the most comprehensively studied sensory cascade. It consists of a series of connected cortical brain areas (macaque brain shown). PIT, posterior inferior temporal cortex; CIT, central; AIT, anterior; RGC, retinal ganglion cell; LGN, lateral geniculate nucleus. DoG, difference of Gaussians model; T(•), transformation. (**c**) HCNNs are multilayer neural networks, each of whose layers are made up of a linear-nonlinear (LN) combination of simple operations such as filtering, thresholding, pooling and normalization. The filter bank in each layer consists of a set of weights analogous to synaptic strengths. Each filter in the filter bank corresponds to a distinct template, analogous to Gabor wavelets with different frequencies and orientations; the image shows a model with four filters in layer 1, eight in layer 2, and so on. The operations within a layer are applied locally to spatial patches within the input, corresponding to simple, limited-size receptive fields (red boxes). The composition of multiple layers leads to a complex nonlinear transform of the original input stimulus. At each layer, retinopy decreases and effective receptive field size increases. HCNNs are good candidates for models of the ventral visual pathway. By definition, they are image computable, meaning that they generate responses for arbitrary input images; they are also mappable, meaning that they can be naturally identified in a component-wise fashion with observable structures in the ventral pathway; and, when their parameters are chosen correctly, they are predictive, meaning that layers within the network describe the neural response patterns to large classes of stimuli outside the domain on which the models were built.

Yamins (2016) Nature Neurosci

Biological neural representation resembles to the artificial neural representation.



Technical limits



DiCarlo and Cox, TiCS, 2007

Brain score: how well existing models explain the neural data

http://www.brain-score.org/#leaderboard

Sort by average score

Rank	Model	average	V1	V2	V4	IT	behavior	engineering	Deng2009-top1
1	efficientnet-b0 *Tan et al., 2019*	.442	.215	.317	.556	.547	.573		
2	efficientnet-b6 *Tan et al., 2019*	.435	.263	.295	.563	.541	.513		
3	efficientnet-b2 *Tan et al., 2019*	.434	.213	.317	.569	.547	.526		
4	efficientnet-b4 *Tan et al., 2019*	.434	.228	.286	.575	.543	.535		
5	CORnet-S *Kubilius et al., 2018*	.417	.294	.242	.581	.423	.545	.747	.747
6	vgg-19 *Simonyan et al., 2014*	.408	.347	.341	.610	.248	.494	.711	.711
7	resnet-50-robust *Santurkar et al., 2019*	.408	.378	.365	.537	.243	.515		
8	resnet-101_v1 *He et al., 2015*	.407	.266	.341	.590	.274	.561	.764	.764
9	vgg-16 *Simonyan et al., 2014*	.406	.355	.336	.620	.259	.461	.715	.715
10	resnet-152_v1 *He et al., 2015*	.405	.282	.338	.598	.277	.533	.768	.768

Sort by V4 score

Model	average	V1	V2	V4	IT	behavior	engineering	Deng2009-top1
vgg-16 *Simonyan et al., 2014*	.406	.355	.336	.620	.259	.461	.715	.715
vgg-19 *Simonyan et al., 2014*	.408	.347	.341	.610	.248	.494	.711	.711
xception *Chollet et al., 2016*	.384	.245	.306	.610	.249	.508	.790	.790
densenet-169 *Huang et al., 2016*	.404	.281	.322	.601	.274	.543	.759	.759
resnet-50-pytorch *He et al., 2015*	.399	.289	.317	.600	.259	.528	.752	.752
resnet-101_v2 *He et al., 2015*	.404	.274	.332	.599	.263	.555	.774	.774
resnet50-SIN_IN *Geirhos et al., 2019*	.404	.282	.324	.599	.276	.541	.746	.746
densenet-201 *Huang et al., 2016*	.402	.277	.325	.599	.273	.537	.772	.772
resnet-152_v1 *He et al., 2015*	.405	.282	.338	.598	.277	.533	.768	.768
resnet50-SIN_IN_IN *Geirhos et al., 2019*	.397	.275	.321	.596	.273	.523	.767	.767

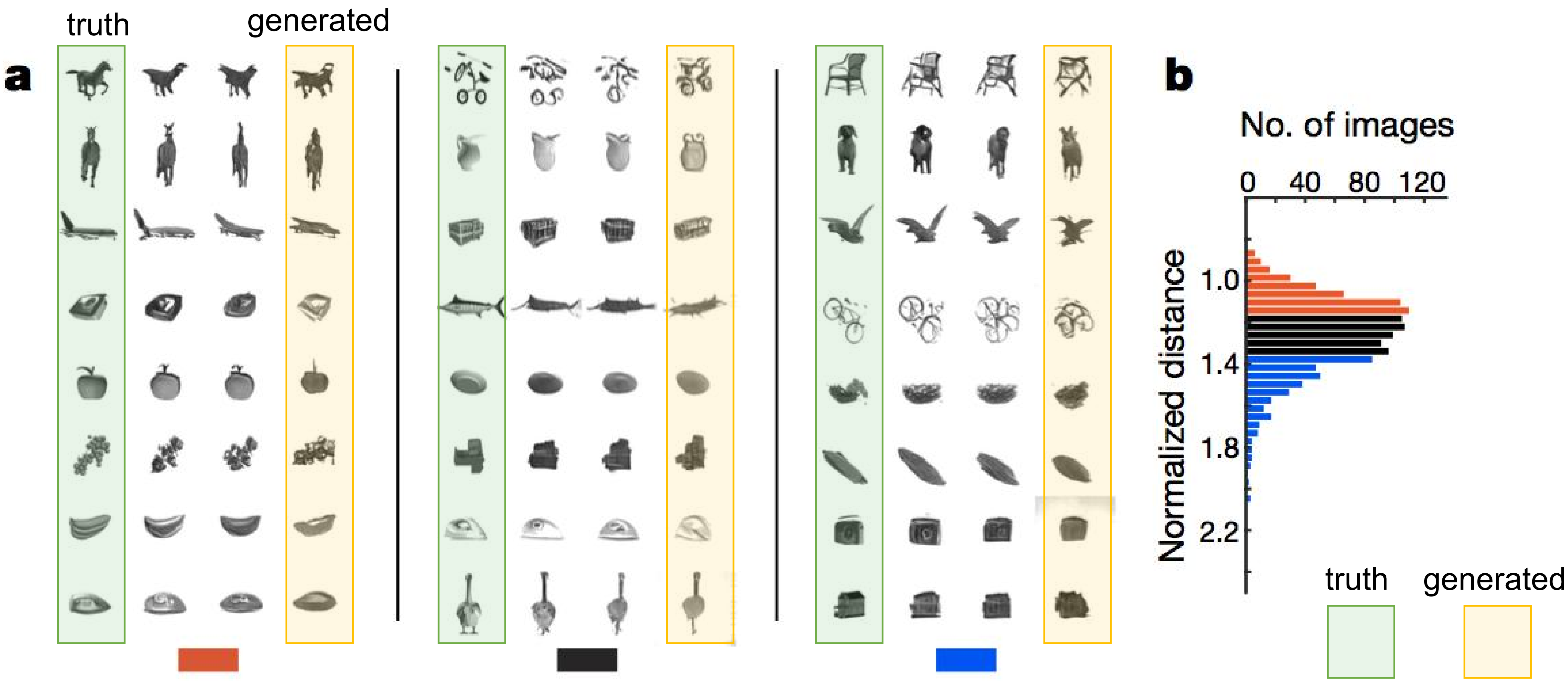
Q2:

Can we reconstruct the image based on neural signals?

neural activity → image

We will show you some detailed studies/algorithms in the following weeks.

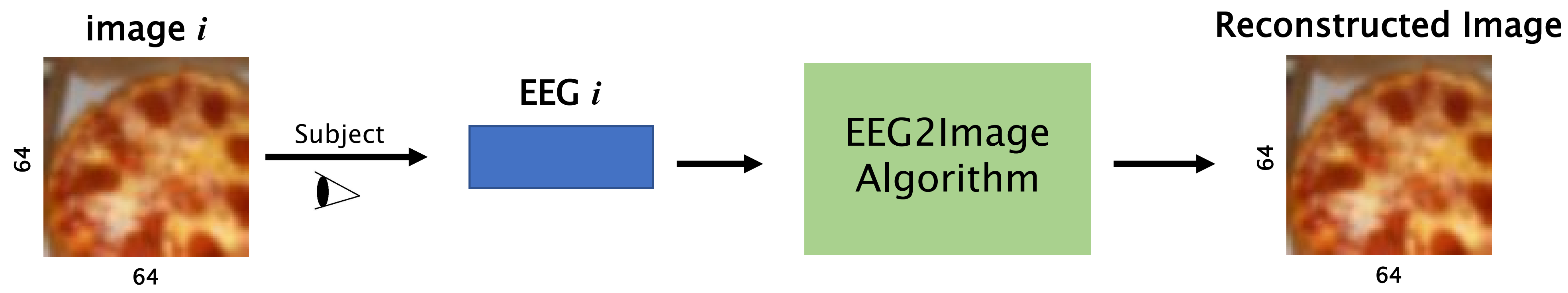
<u>Reconstruct images</u> based on IT neuronal activity via GAN



Pinglei Bao, Liang She, Mason McGill & Doris Y. Tsao (2020). A map of object space in primate inferotemporal cortex. *Nature*

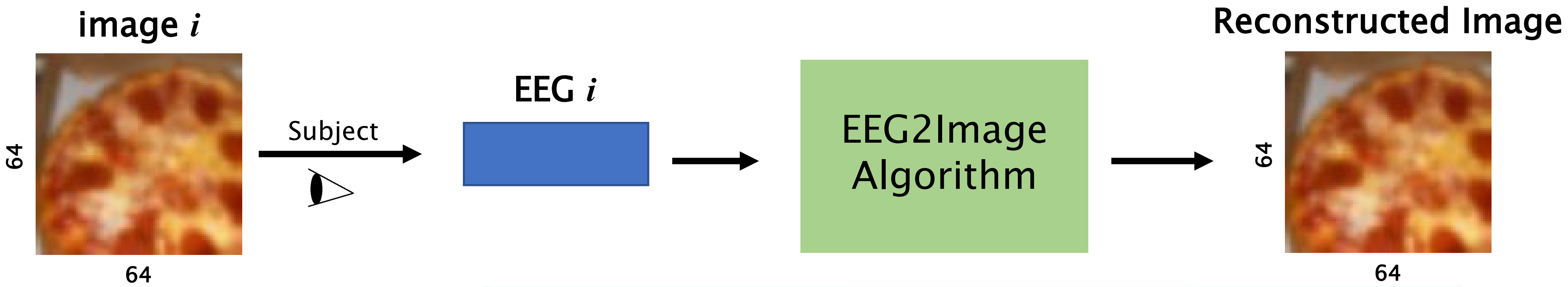
Reconstruct images based EEG activity via GAN

Framework

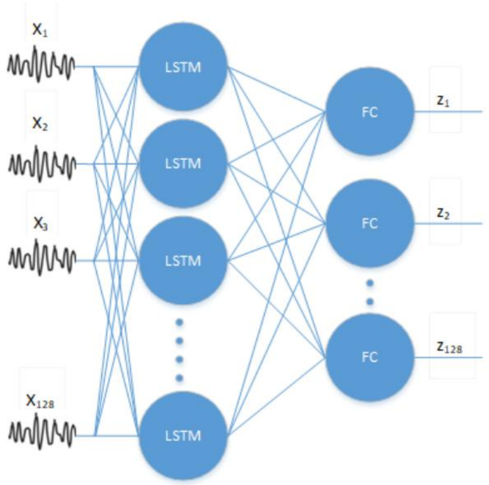


Reconstruct images based EEG activity via GAN

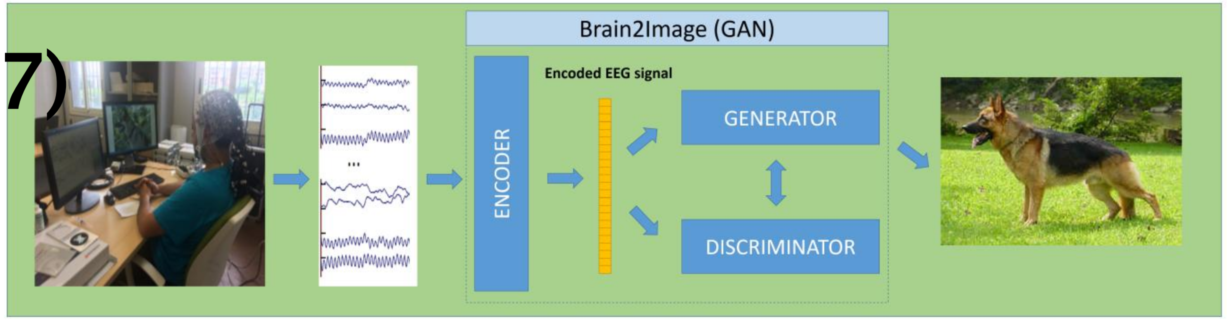
Framework



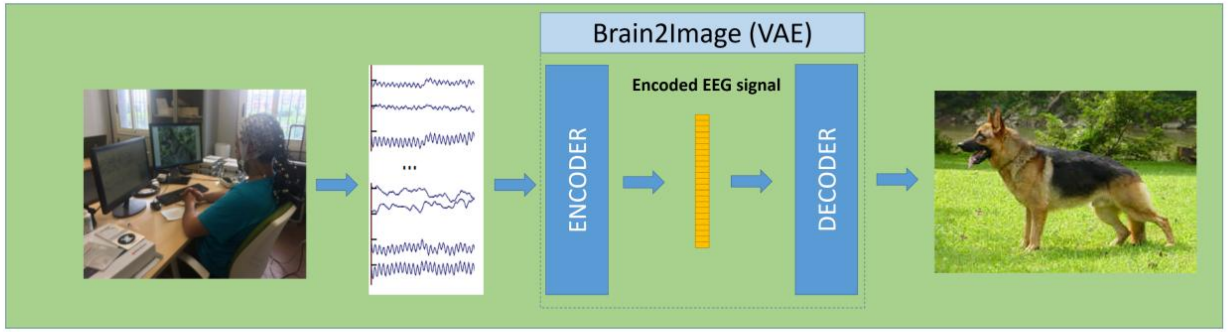
The first work (ACMMM 2017)



GAN version

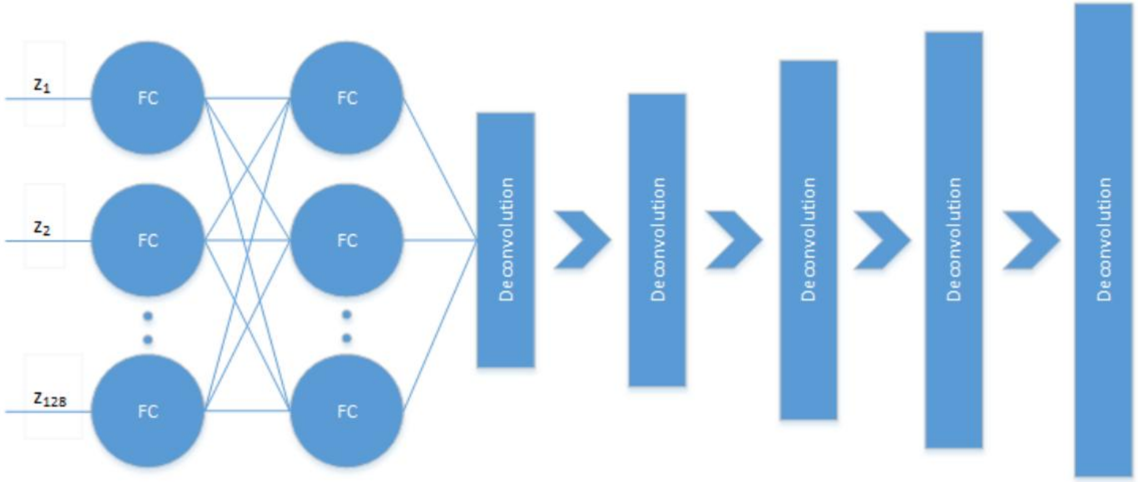


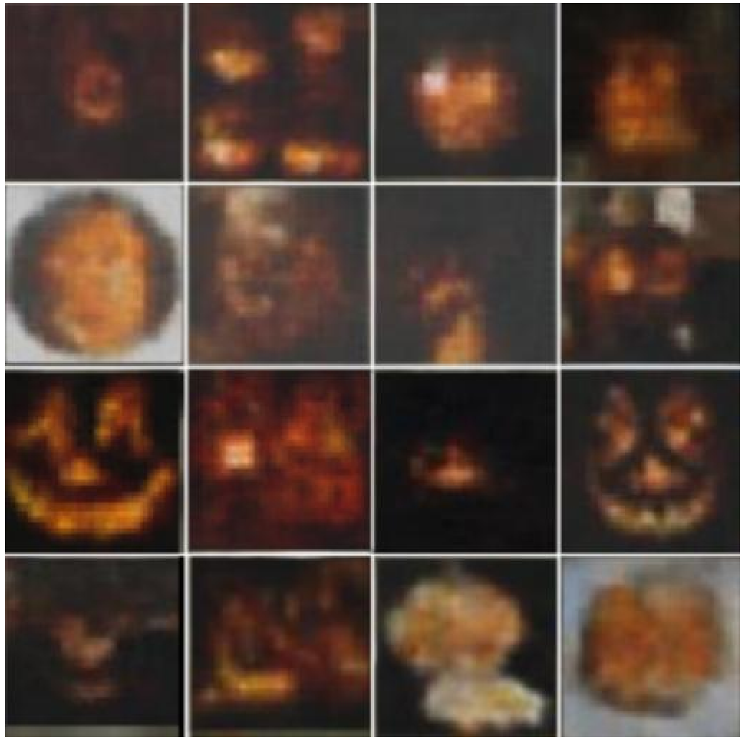
VAE version



Kavasidis, Isaak, et al. "Brain2image: Converting brain signals into images." *Proceedings of the 25th ACM international conference on Multimedia*. 2017.

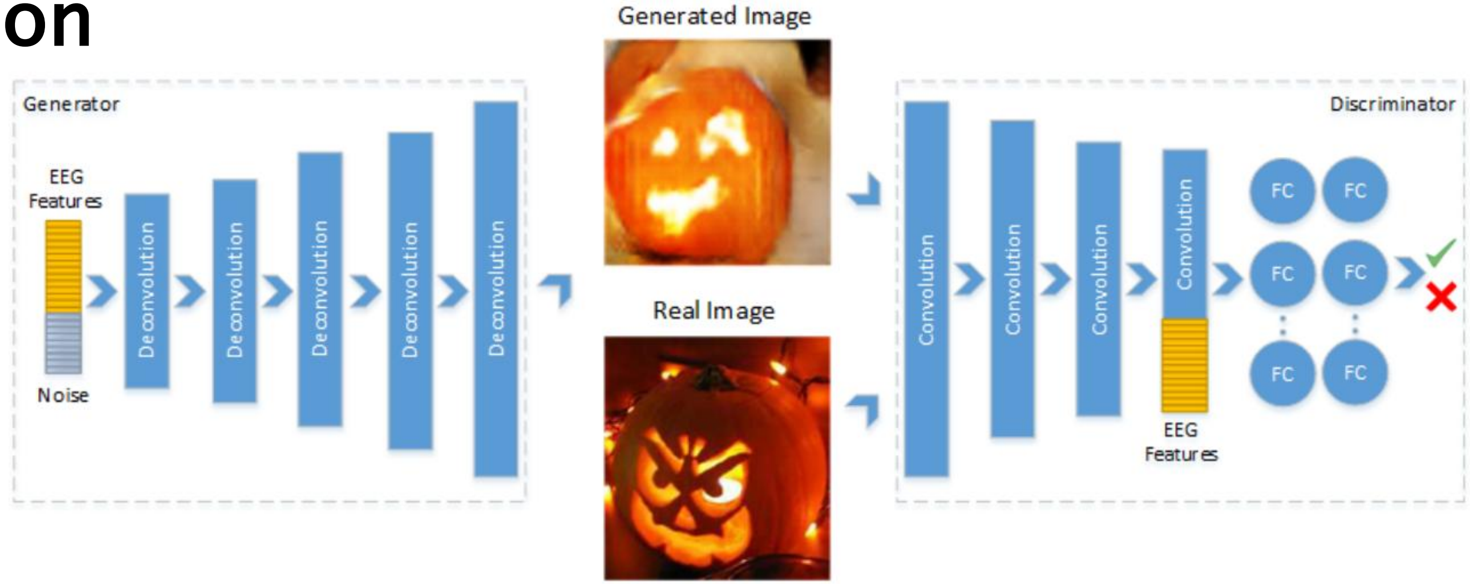
Brain2image: VAE version





Kavasidis, Isaak, et al. "Brain2image: Converting brain signals into images." *Proceedings of the 25th ACM international conference on Multimedia*. 2017.

Brain2image: GAN version



A work in IJCAI 2019

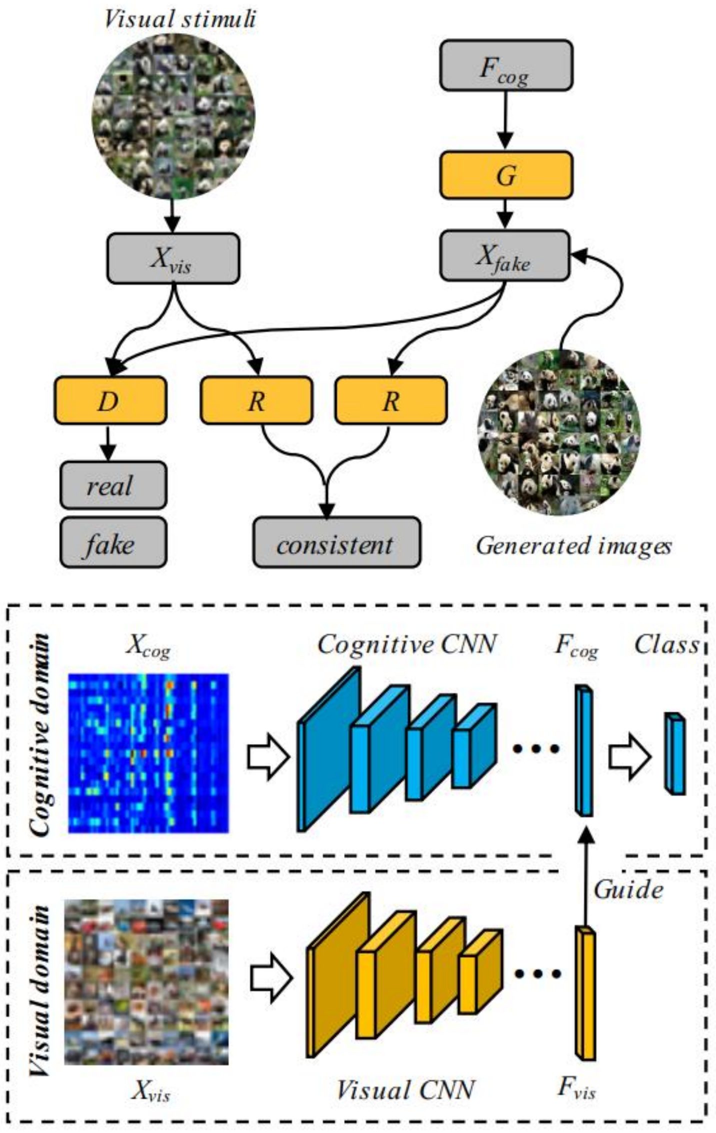


Figure 3: Visual-guided EEG classification.

Idea:
Guide EEG decoding via pre-trained CNN representation

#1 Classification: $L = -\sum_{j=1}^{N} y_j \log S_j + (F_{cog} - F_{vis})^2$

#2 Loss_d:

perceptual loss: $\| f(X_{vis}) - f(X_{fake}) \|_2^2$

f in the first FCN layer

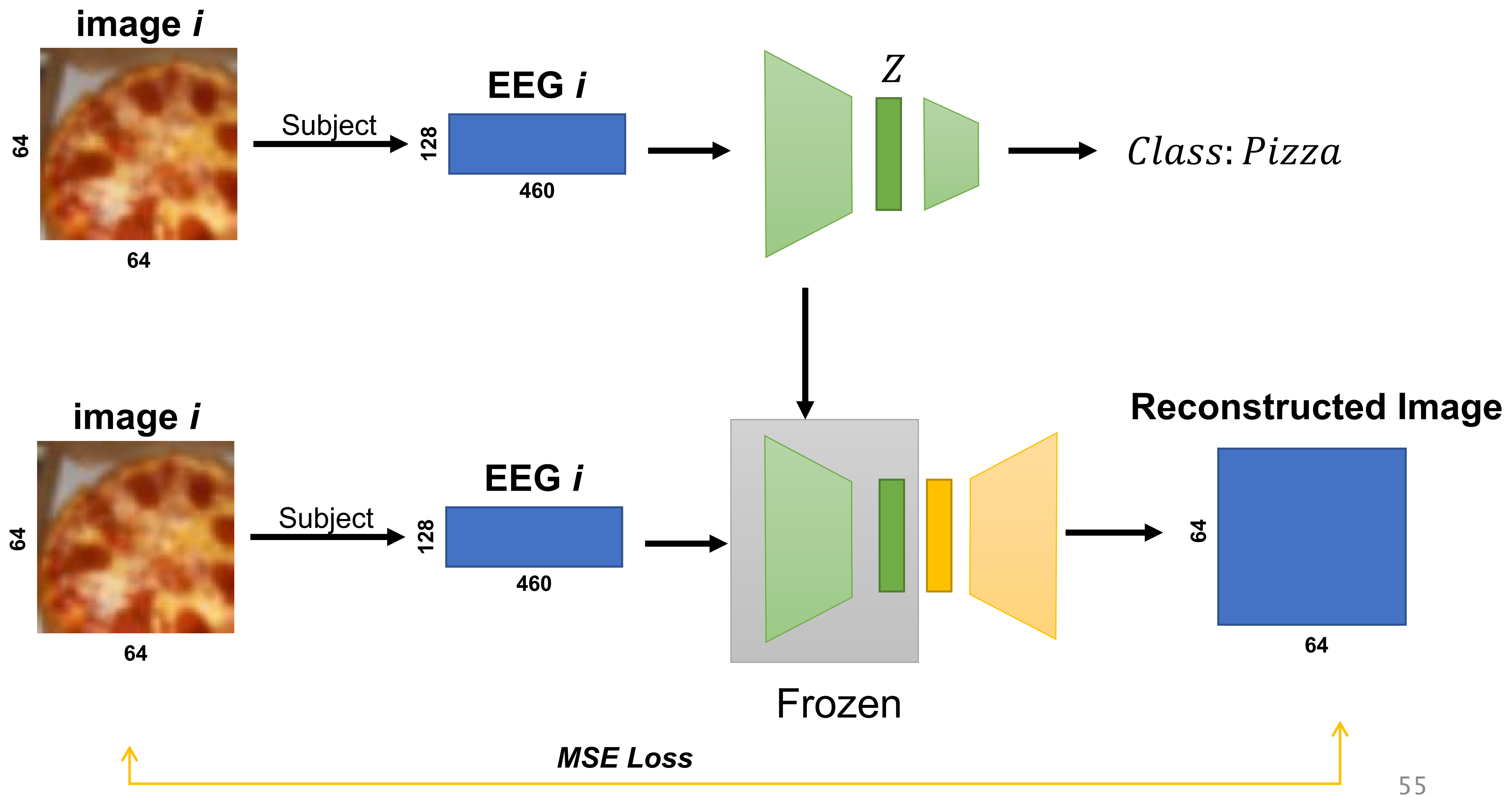
semantic loss: softmax cross entropy

Core idea:
Make EEG features more similar to pre-trained CNN features

Jiao, Zhicheng, et al. "Decoding EEG by Visual-guided Deep Neural Networks." IJCAI. 2019.

NCC lab tried to replicate the results.   But ....

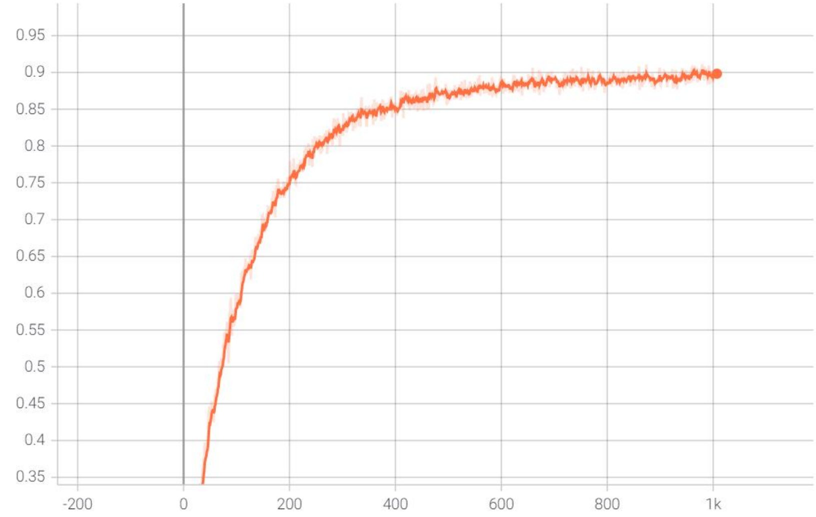
Our attempt 1



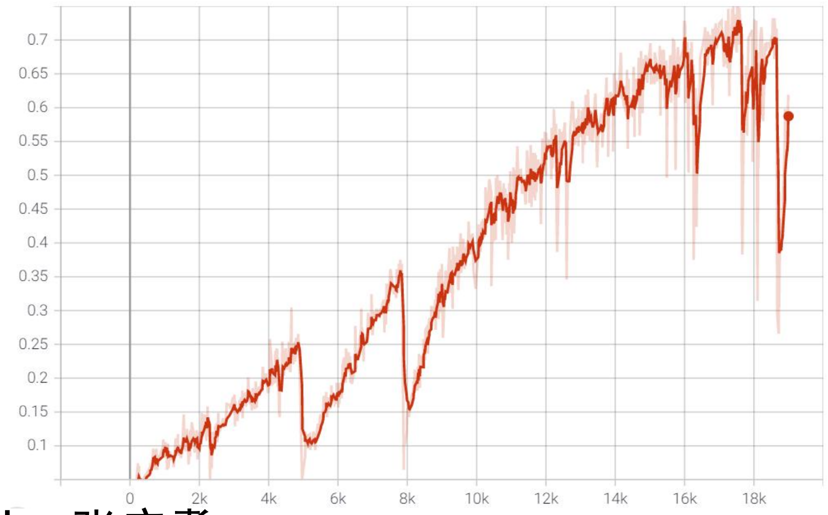
by 张庆翥

Attempt 1: Results

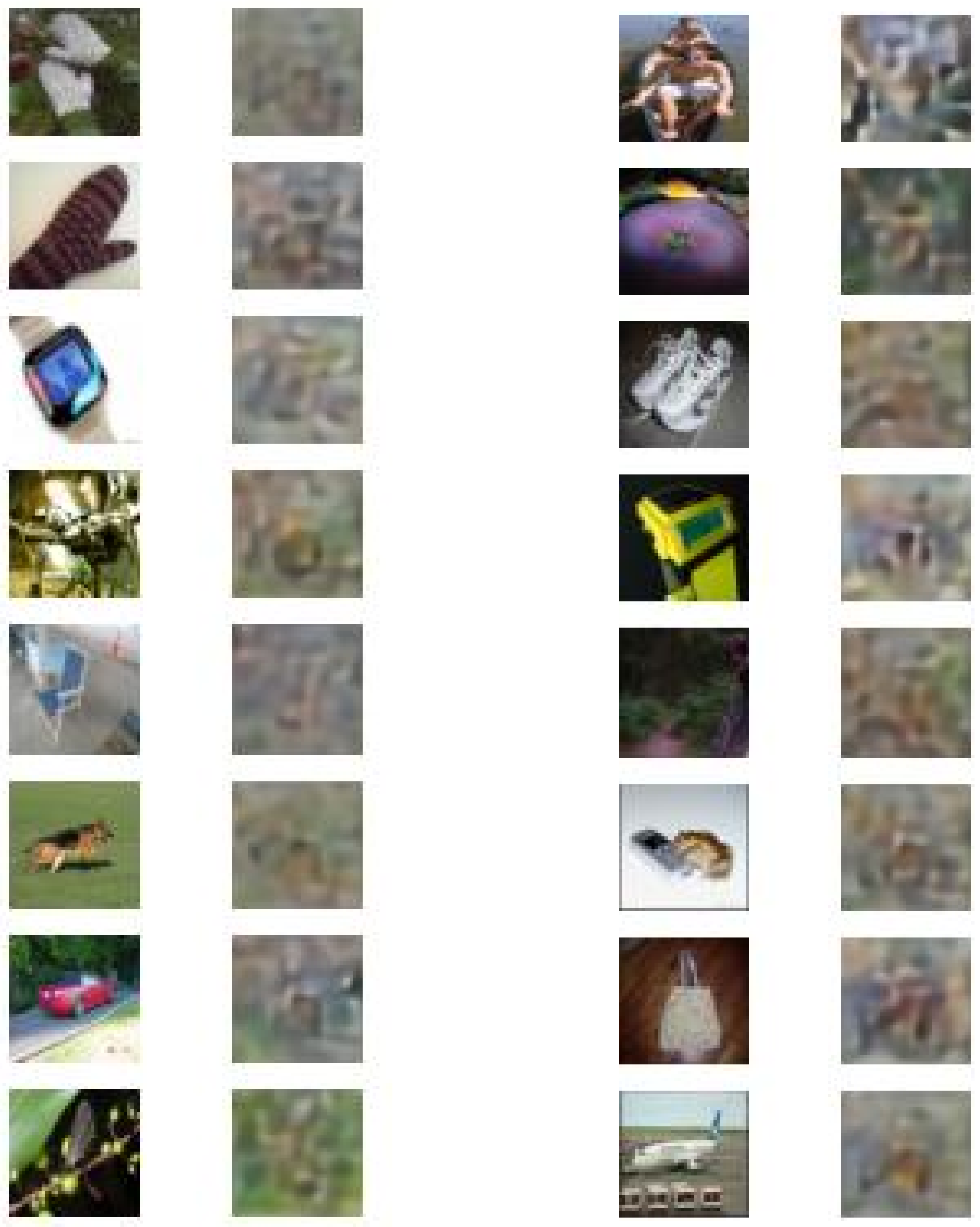
EEG classification accuracy using CNN encoder



EEG classification accuracy using LSTM encoder



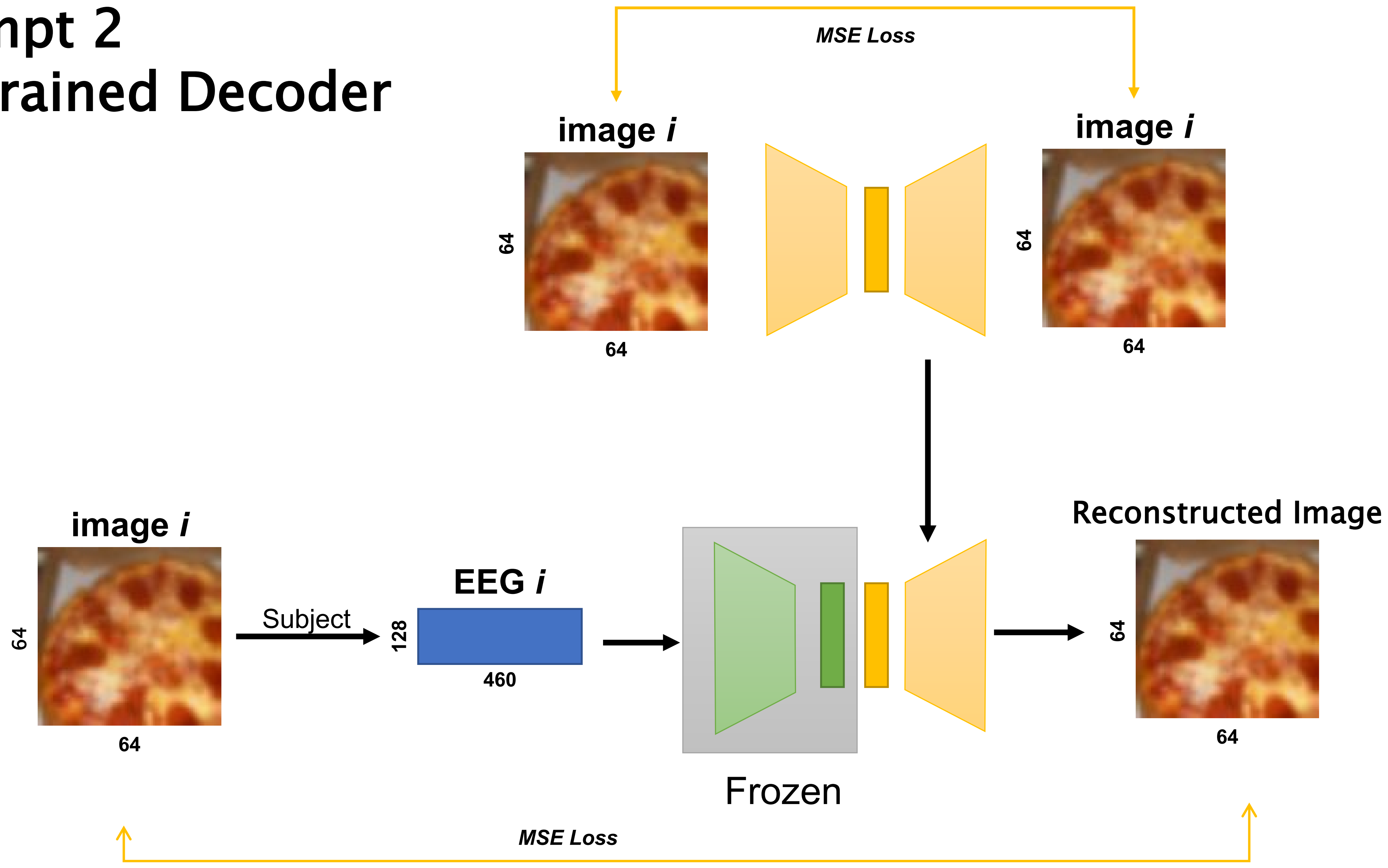
by 张庆翥



Train

Test

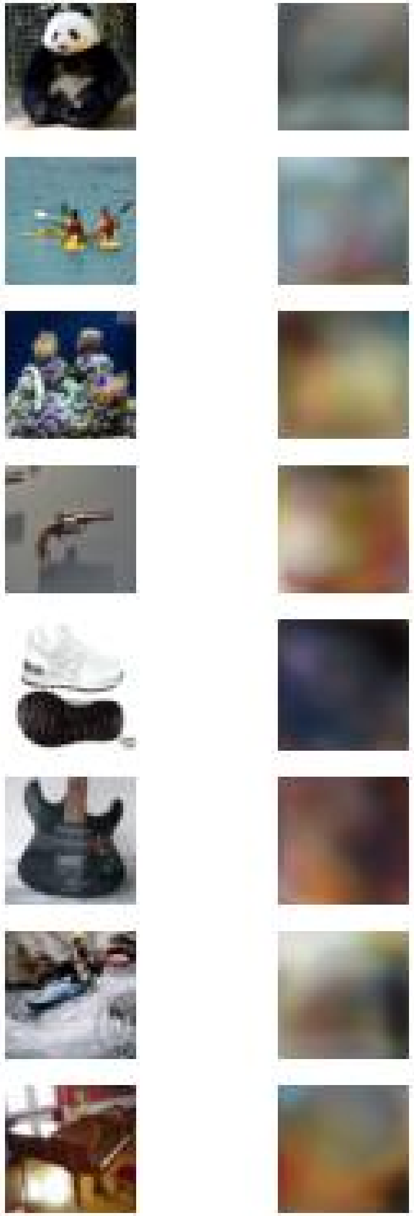
Attempt 2
Pre-trained Decoder



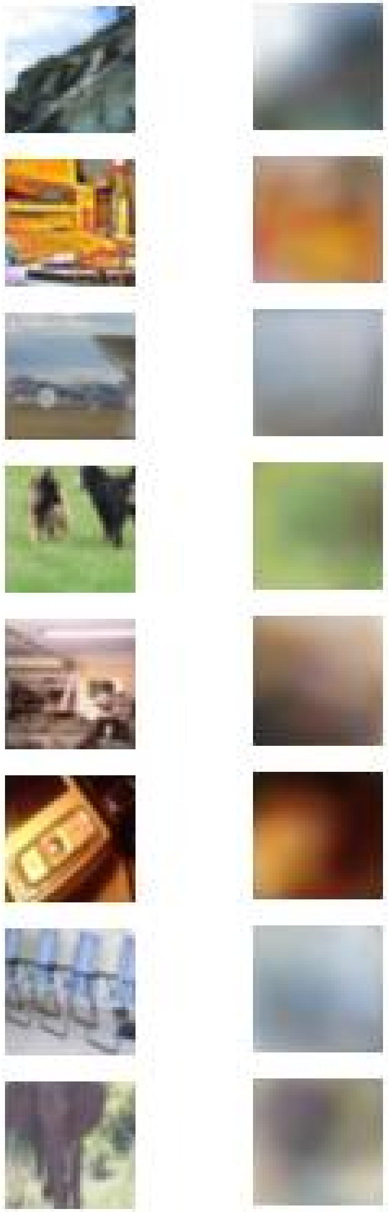
by 张庆翥 2022 summer

Attempt 2: Results

Autoencoder: Image to Image



Test



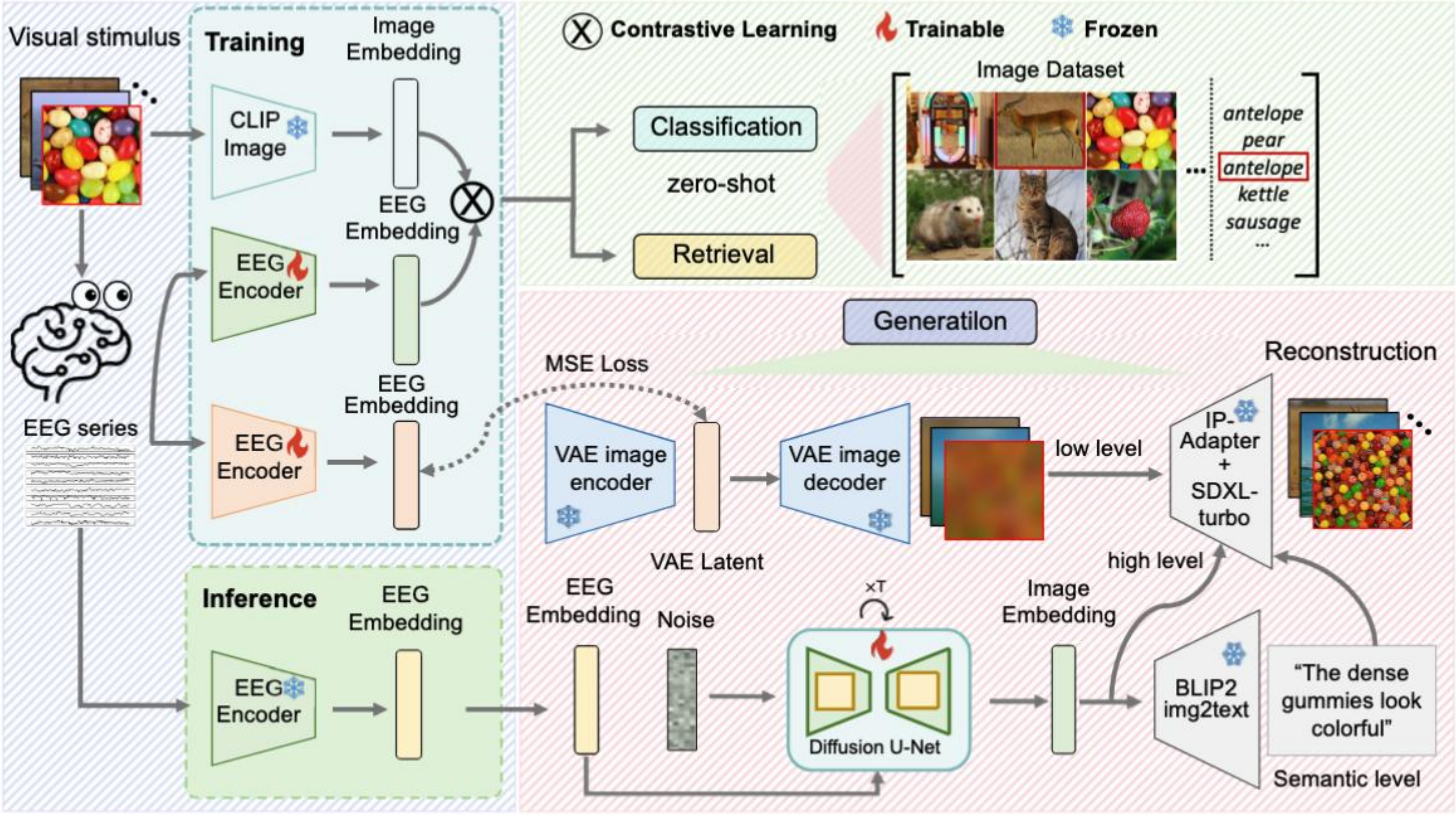
Overfitting
On training set

by 张庆翥

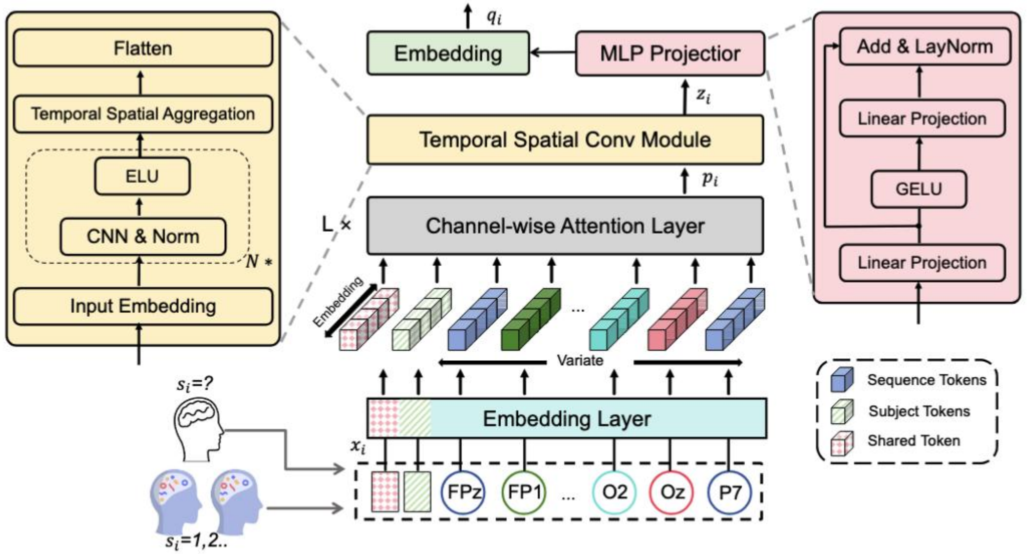
Attempt 3
EEG encoder (ATM)-CLIP Alignment


李东洋 魏晨

脑电的编码器与CLIP对齐，利用脑电的表征进行分类和重构



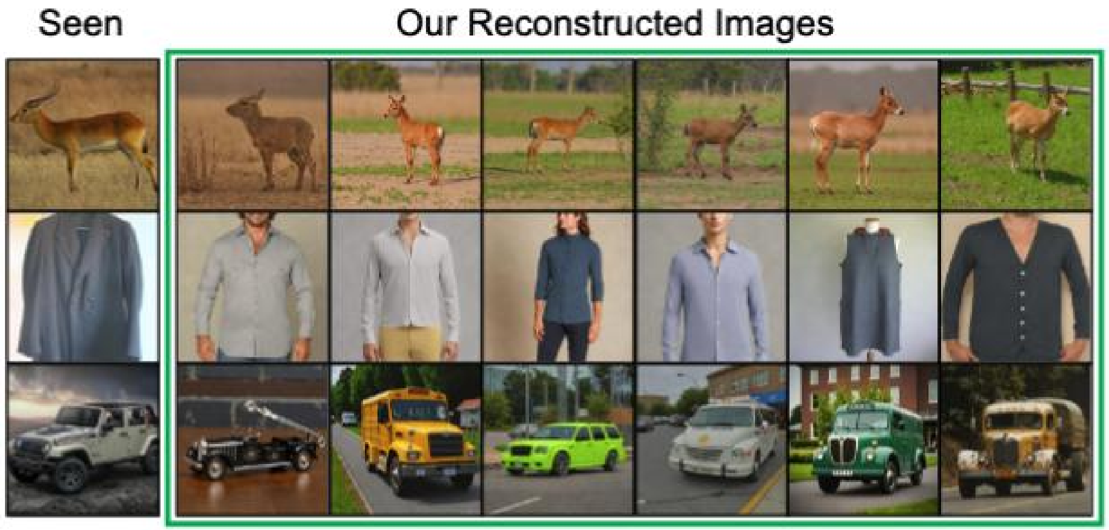
EEG encoder
Adaptive Thinking Mapper (ATM)



by 李东洋 2024 summer

Attempt 3: Results


李东洋 魏
晨



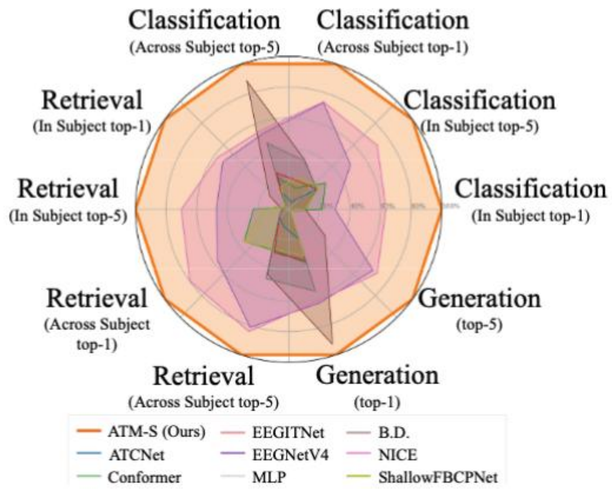
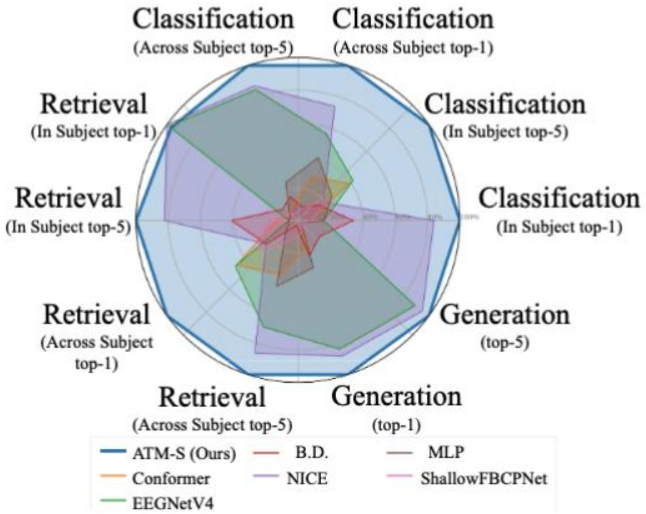


Table 1: Quantitative assessments of the reconstruction quality for EEG, MEG, and fMRI. For detailed explanations of the metrics.

	Low-level		High-level			
Dataset ↑	SSIM ↑	AlexNet(2) ↑	AlexNet(5) ↑	Inception ↑	CLIP ↑	SwAV ↓
NSD-fMRI [4]	0.366	0.962	0.977	0.910	0.917	0.410
NSD-fMRI [33]	0.356	0.942	0.962	0.872	0.915	0.423
NSD-fMRI [41]	0.308	0.917	0.974	0.936	0.942	0.369
THINGS-MEG [4]	0.327	0.695	0.753	0.593	0.700	0.630
THINGS-MEG (averaged) [4]	0.336	0.736	0.826	0.671	0.767	0.584
THINGS-MEG (Ours)	0.340	0.613	0.672	0.619	0.603	0.651
THINGS-EEG (Ours)	**0.345**	**0.776**	**0.866**	**0.734**	**0.786**	**0.582**

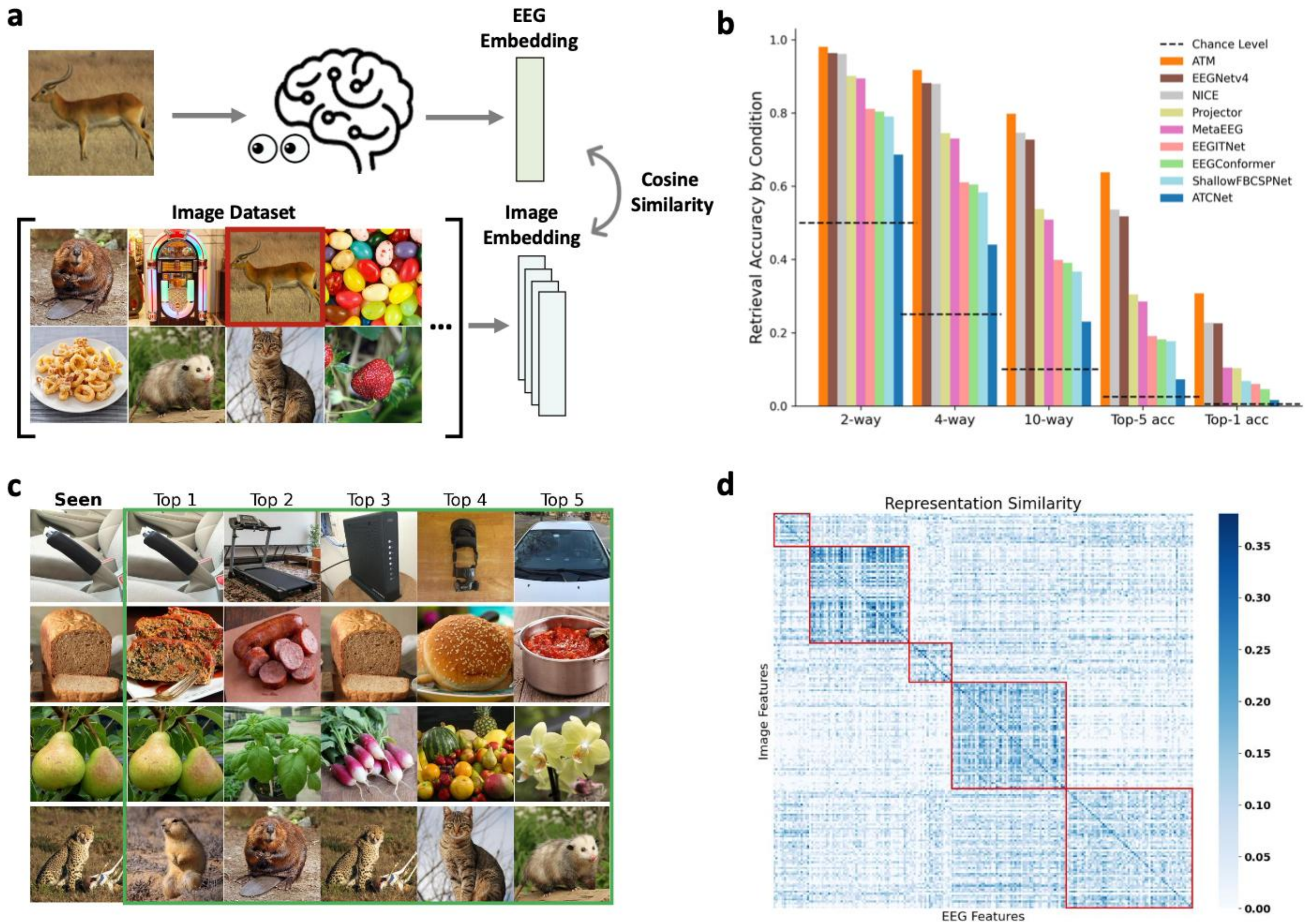


by 李东洋 2024 summer

基于脑电的视觉重构

李东洋　魏晨

基于脑电的视觉解码



Li, Wei, et al., (2024)　https://arxiv.org/abs/2403.07721

代码tutorial讲解

B站账号：NCC_lab

Q3:

Can BI (e.g. mechanism of visual system) help improve AI’s performance?

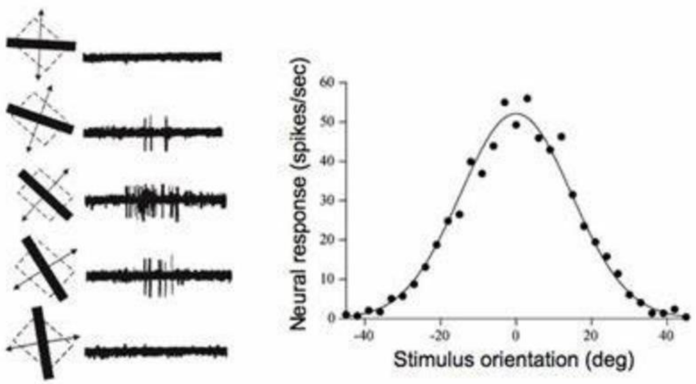
BI inspires AI

VOneNet 的设计思路为：通过使用一个精心设计的V1模块来替换CNN的第一层，使得CNN的活动模式更像真实的初级视觉皮层，以提升模型对图像扰动的鲁棒性。其核心的 V1模块 便是*VOneBlock*。

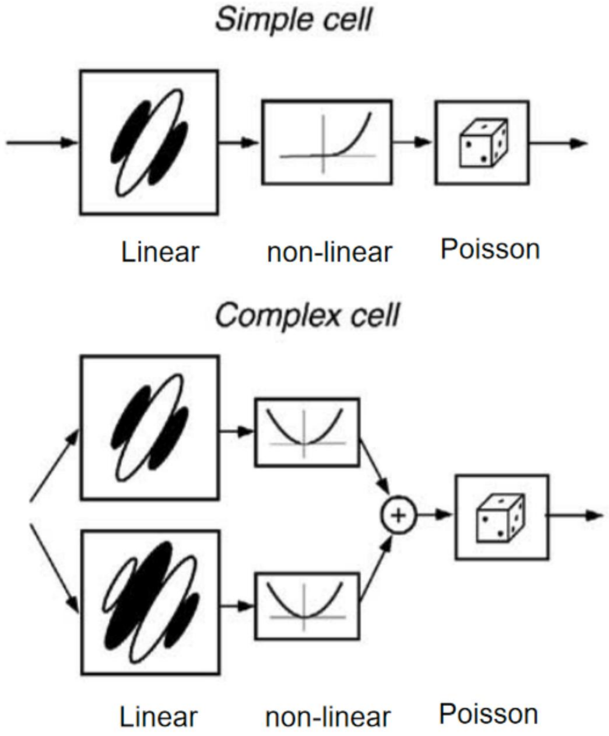
VOneNet 的设计思路为：通过使用一个精心设计的V1模块来替换CNN的第一层，使得CNN的活动模式更像真实的初级视觉皮层，以提升模型对图像扰动的鲁棒性。

其核心的 V1模块 便是*VOneBlock*。

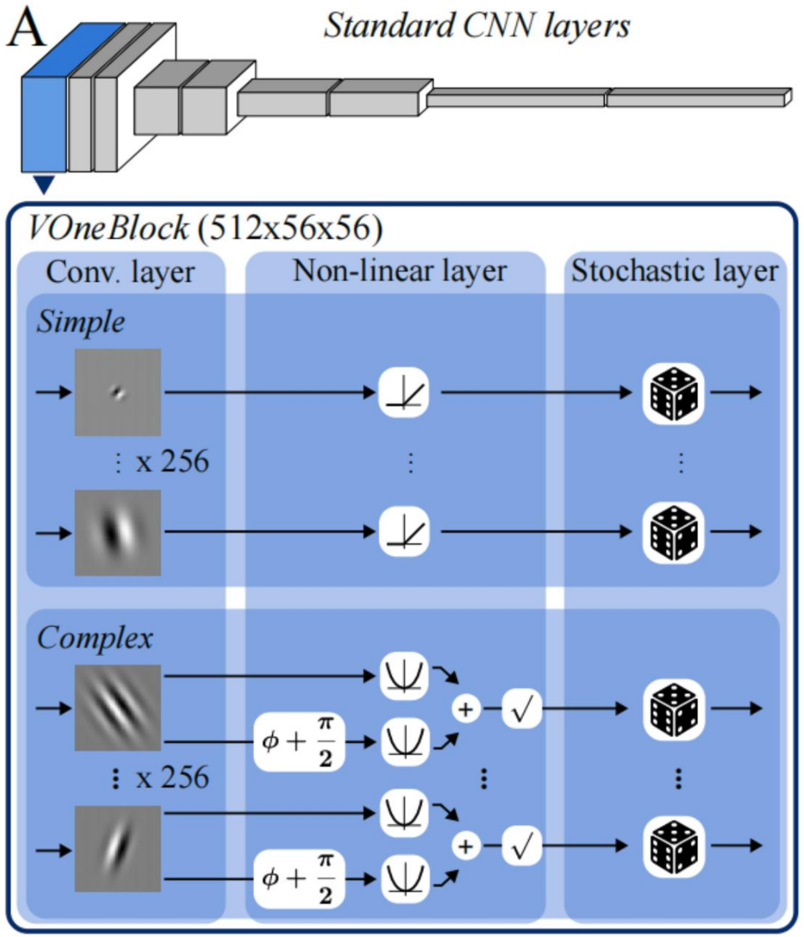
在V1区中，存在着两种重要的细胞：**复杂细胞**和**简单细胞**，它们由Wisel和Hubel在上世纪60年代共同发现。
这两种细胞的共同点在于：它们都对具有特定方向、频率和对比度的正弦光栅信号有偏好性



Rust等人在2005年提出了建模简单细胞和复杂细胞的**Linear-Nonlinear-Poisson模型 (LNP模型).**



在CNN中设计一个**V1模块**，使得CNN第一层具有大脑V1神经元的性质



What are the benefits from VOneNet? → Robustness to perturbations

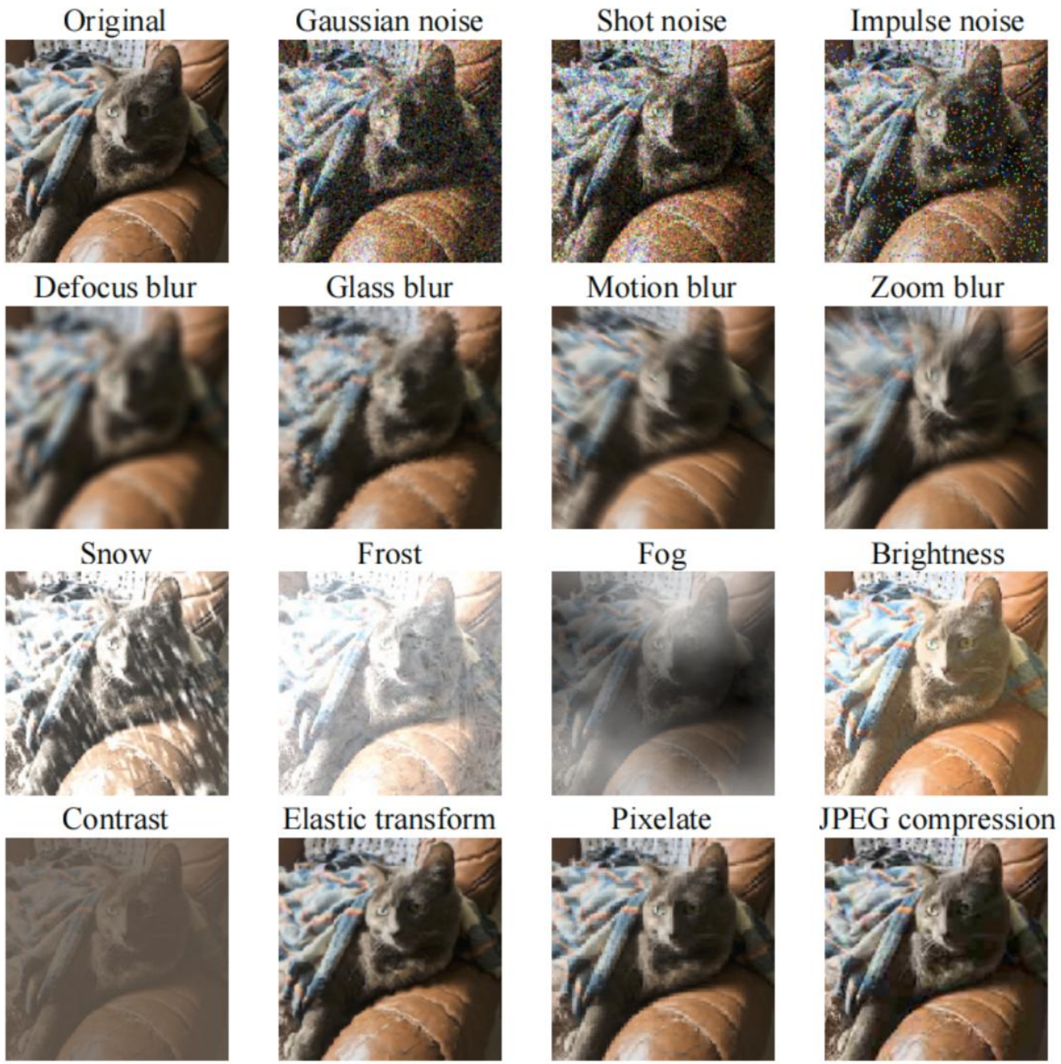
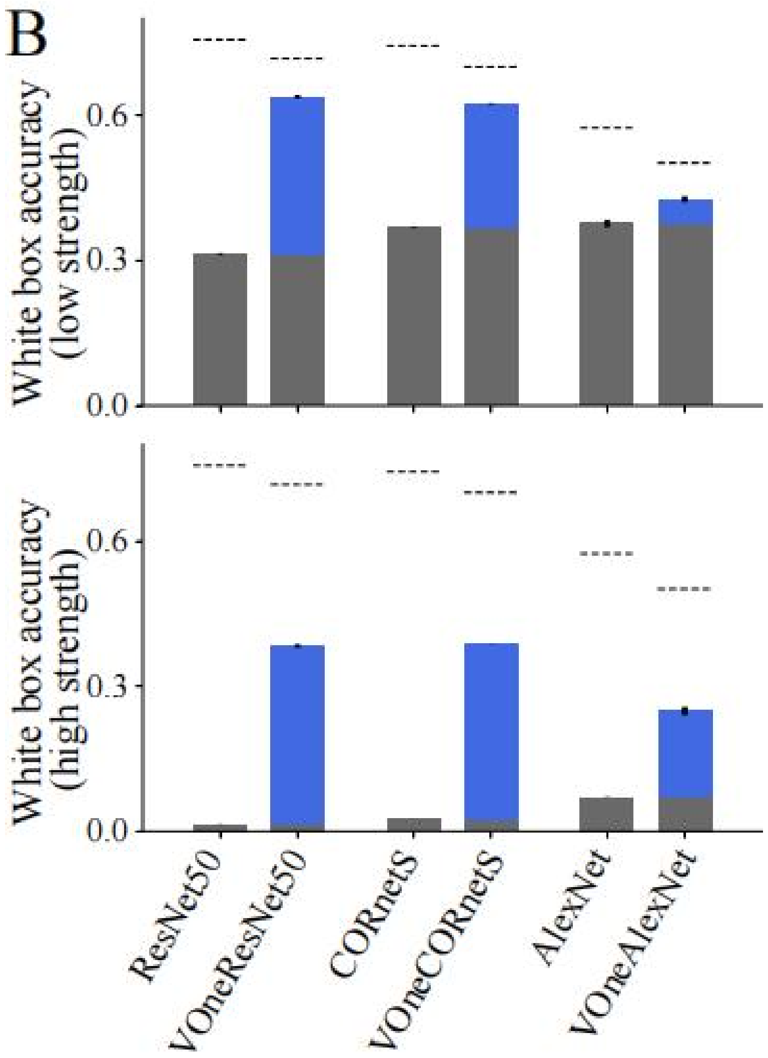


图 3. 十五种不同的图像损坏方式。



More details

Scan to Follow

https://mp.weixin.qq.com/s/izbA_ZaOOTGdYZ3UCauSDA

文章分享 ｜ 基于灵长类初级视觉皮层启发的CNN前端组件可以改善视觉模型对图像干扰的鲁棒性

Original NCC lab 神经计算与控制实验室 2022-04-28 15:03 Posted on 广东

收录于合集
#文献阅读
24个 >

/ 简介 /

本推文介绍的是2020年发表在NeurIPS的文章“**Simulating a Primary Visual Cortex at the Front of CNNs Improves Robustness to Image Perturbations**”[1]。在文章中，作者提出了一种可衡量CNN人工神经元激活与V1脑区真实的神经元活动之间可解释性的指标“**V1 explained variance**”，并发现了V1 explained variance与视觉模型的对抗鲁棒性成正比关系。基于此发现，作者设计了一种受大脑神经元**LNP模型**[2]启发的CNN前端组件**VOneBlock**。

实验结果表明，在添加了VOneBlock后，CNN模型对两种常用的图像干扰手段（即white-box attack和image corruption）的鲁棒性都有明显提升。

该VoneNet模型代码见Github：https://github.com/dicarlolab/vonenet

The Marr's three levels of explanation

top-down

Level 1: Computation theory

- What is the problem to be solved?
- What are the inputs and outputs to the computation?
- What is the goal, and what is the logic by which it is carried out?

Level 2: Algorithmic

- How is the information represented and processed to achieve the computational goal?

Level 3: Implementation

- How is the computation realized in physical or biological hardware?

Bottom-up

Reverse-engineer our visual system

to understand the design principles of vision

Future directions:

- Design eye-like camera
- Design AI for computer vision
- Explain visual phenomenon
- Explain/Treat visual disease
- Human vision enhancement
- Decode/encode neural signals

Summary of Lecture 2 – Visual System

0. Marr's three levels of explanation
1. Evolution of the eye
2. Function of the visual system
3. Structure of the eye
4. Photoreceptors
5. Information integration by ganglion cell
6. **Visual pathways**: photoreceptors, interneurons, ganglion cells, LGN, V1, ventral/dorsal streams…
7. Some discussions about **BI & AI in visual system**

Recommended materials

Some materials are from the textbook

- From Neuron to Brain (Ch1–Ch3)

Recommended book

- Principles of Neural Design, by Peter Sterling and Simon Laughlin

- Dapello, J., Marques, T., Schrimpf, M., Geiger, F., Cox, D., & DiCarlo, J. J. (2020). Simulating a primary visual cortex at the front of CNNs improves robustness to image perturbations. *Advances in Neural Information Processing Systems*, *33*, 13073–13087.

https://proceedings.neurips.cc/paper/2020/file/98b17f068d5d9b7668e19fb8ae470841-Paper.pdf