

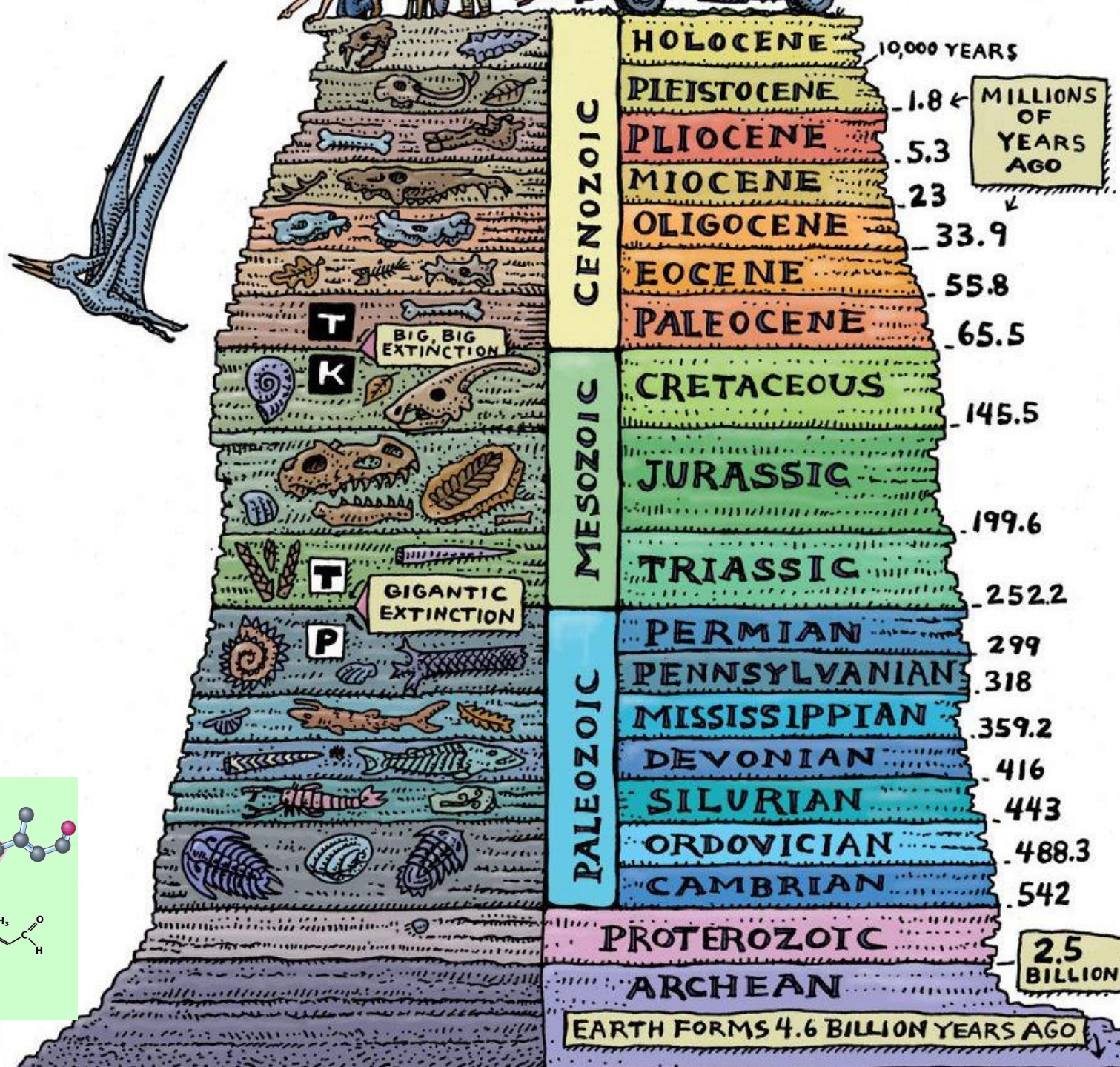
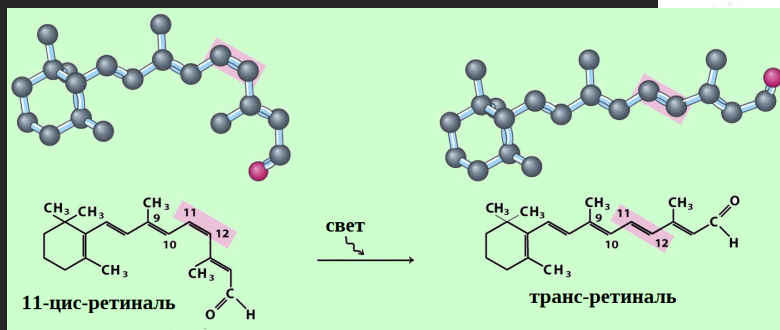


Эволюция зрения

Александр Бовырин

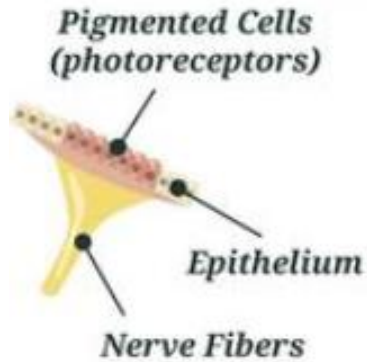


Когда и
зачем
появилось
зрение?

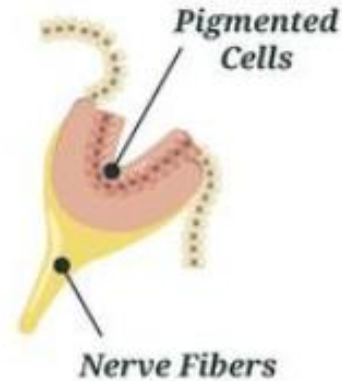


Эволюция глаза

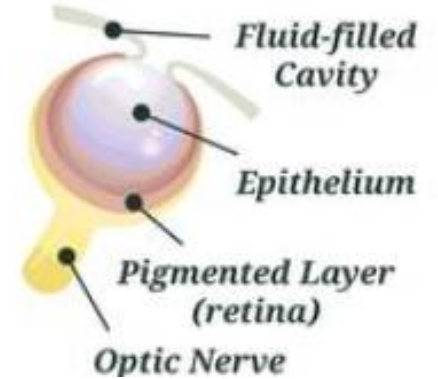
1. Светло или темно?



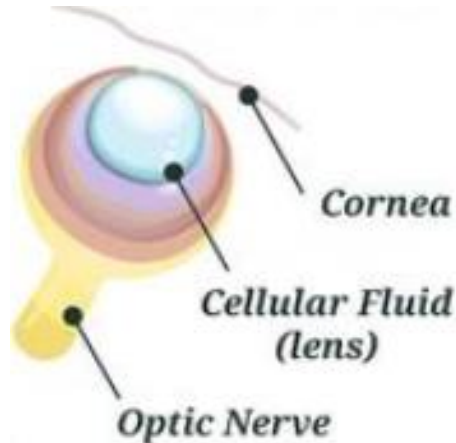
2. Где светло?



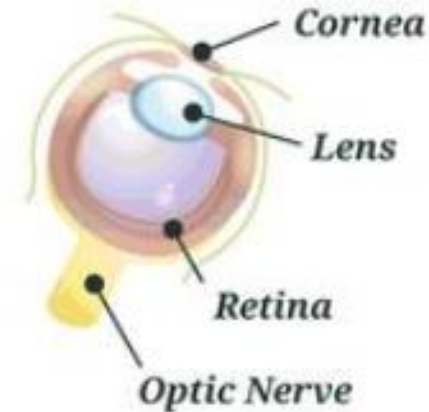
3. Камерный глаз.
Появляются расплывчатые образы.



4. Камерный глаз.
Появляются расплывчатые образы.



5. С линзой изображение становится резким.



Биноккулярное зрение и зрение в темноте



Почему хищники в основном обладают биноккулярным зрением?

Почему не-хищникам выгоднее панорамное зрение?

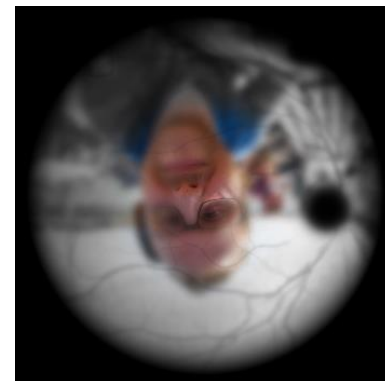
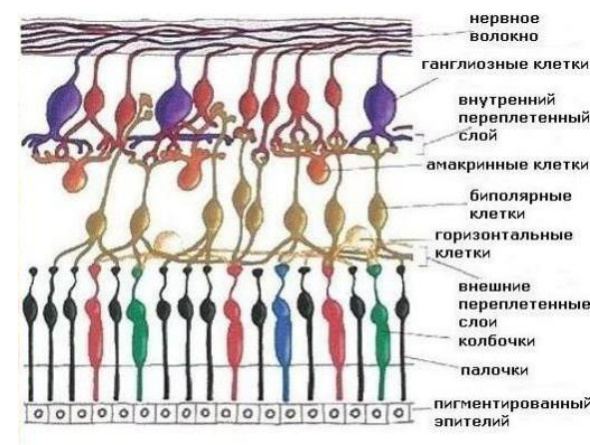


Почему глаза некоторых видов светятся в темноте?

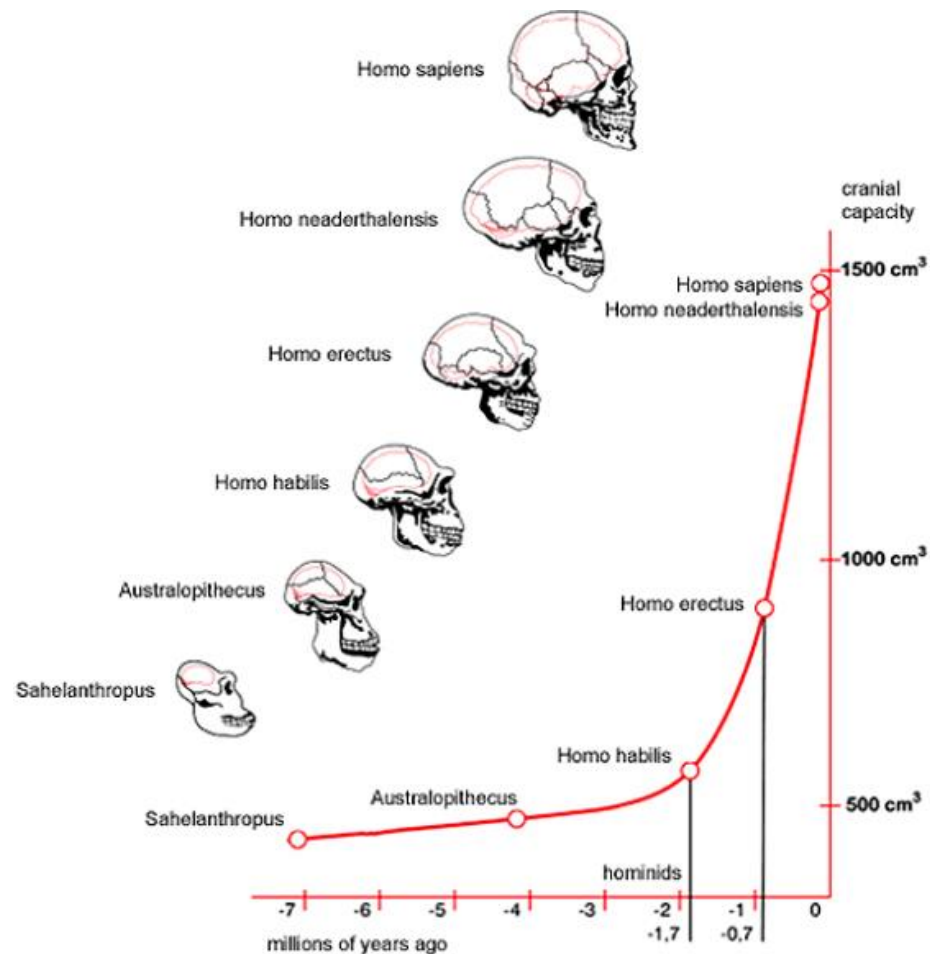


Глаз человека

- Зрачок 1..8 мм
- 100 млн палочек и 5 млн колбочек
- Видимый свет 330 нм -730 нм
- Распределение палочек не однородно. (средняя плотность $1,8 \cdot 10^5$ на 1 мм).
- Есть слепое пятно - где зрительный нерв
- Палочки могут реагировать на один фотон, но многие могут быть связаны только одним нейроном.
- У человека и приматов обнаружены колбочки с тремя разными кривыми спектральной чувствительности, максимумы которых у человека находятся в синей, зелёной и красной областях спектра.
- FPS (frames per second): лягушка 15—20 гц, человек — до 50—60 гц, муха — до 250—300 гц.

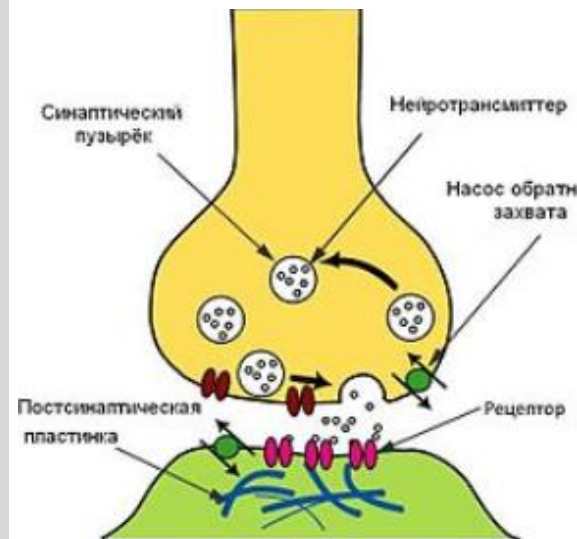
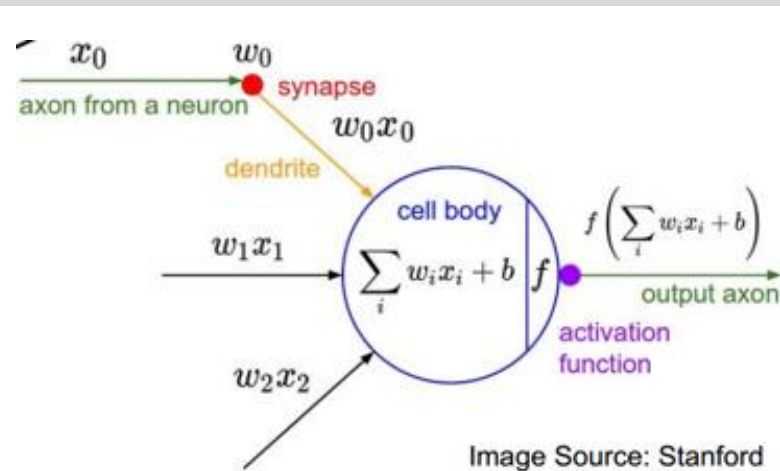
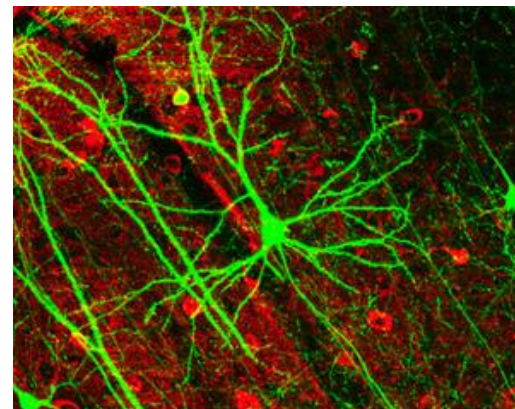
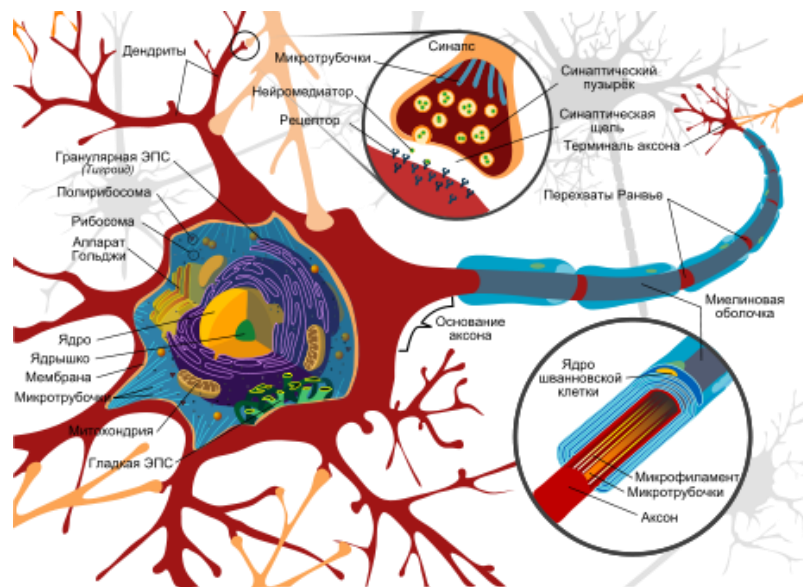


Эволюция мозга



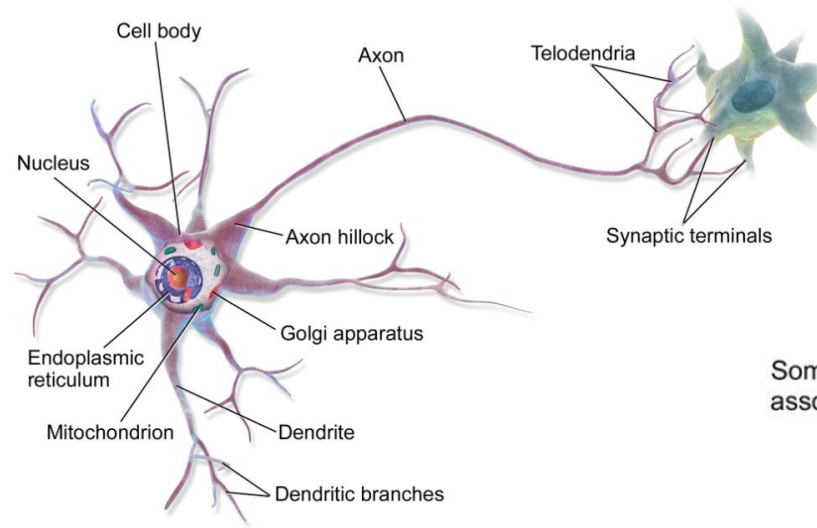
Нейрон

- Электрически возбудимая клетка, которая обрабатывает, хранит и передает информацию с помощью электрических и химических сигналов.
- Нейрон передает сигнал импульсами.
- После возбуждения нейрон на некоторое время оказывается в состоянии абсолютной рефрактерности
- Для передачи импульса требуется ~ 1 ms (зависит от типа нейрона и длины аксона)



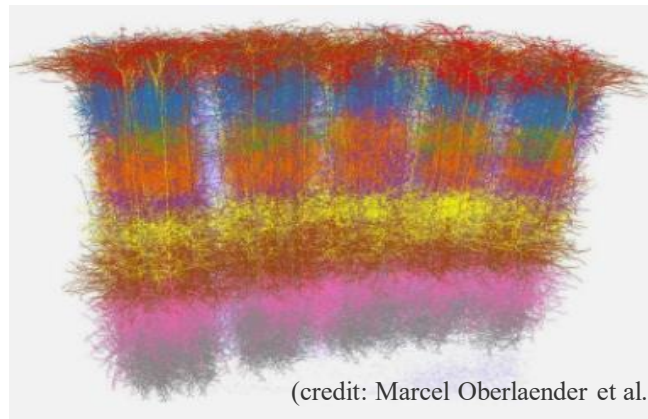
Каждую секунду мозг решает много задач : *восприятие*, *контроль органов*, *движение*, *обучение*, *разговор*, ...

1. В мозге ~**80 миллиардов** нейронов. Каждый нейрон – небольшой компьютер, распознающий множество паттернов.



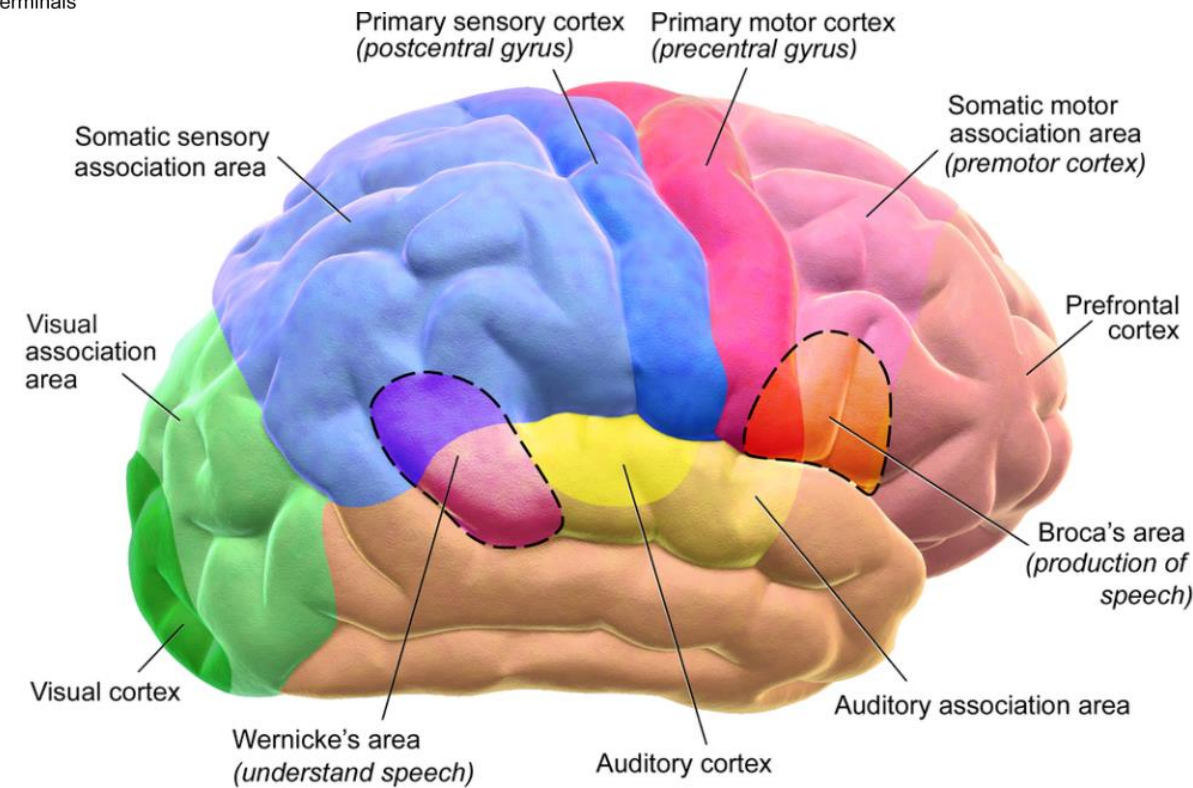
(credit: [BruceBlaus](#))

2. Нейроны объединяются в кортикальные колонки – каждая решает свою задачу распознавания и передачи информации.



(credit: Marcel Oberlaender et al.)

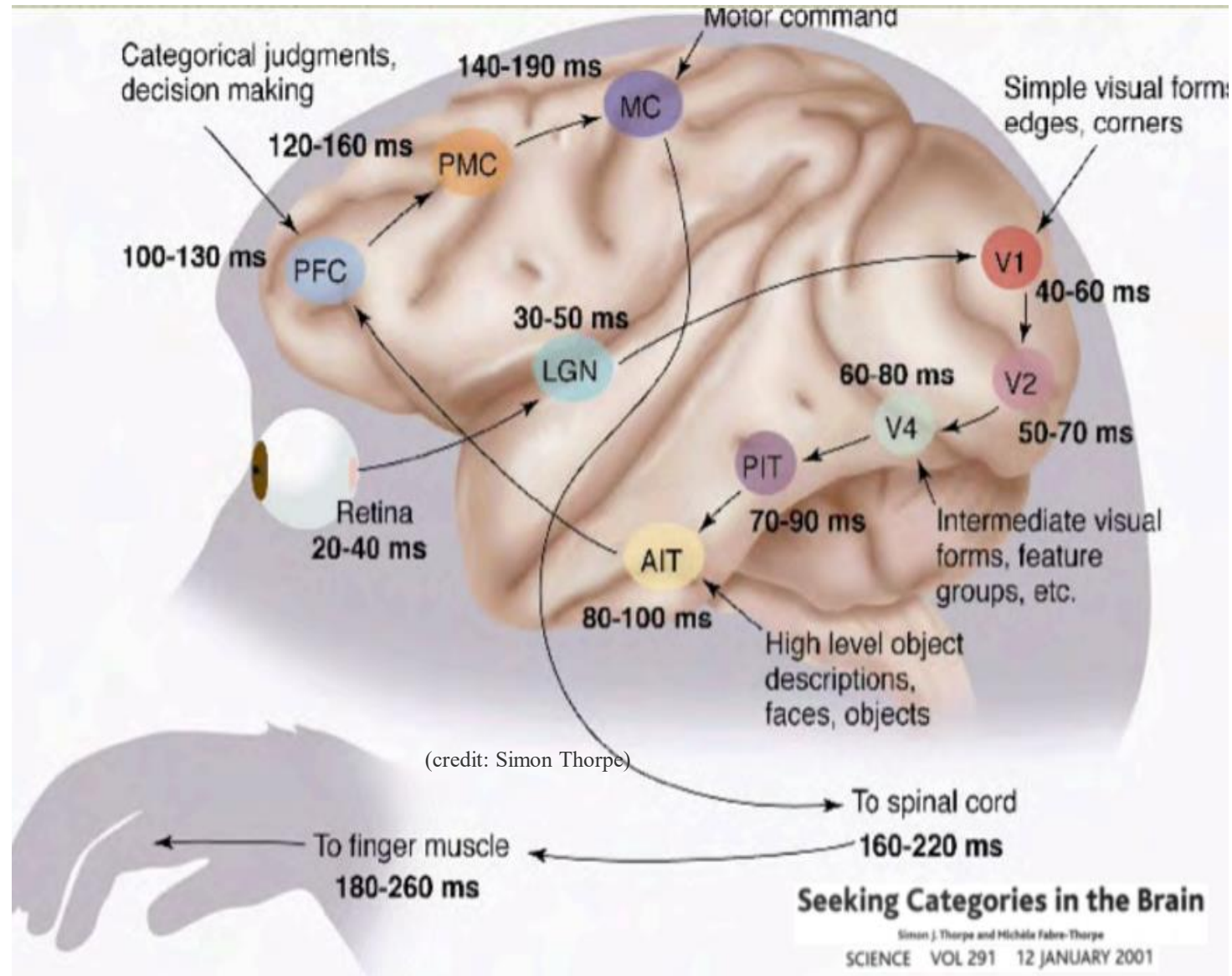
3. Мозг разделен на функциональные поля (visual, hearing, emotions, motion...).



Credit: Blausen.com staff (2014). "[Medical gallery of Blausen Medical 2014](#)".

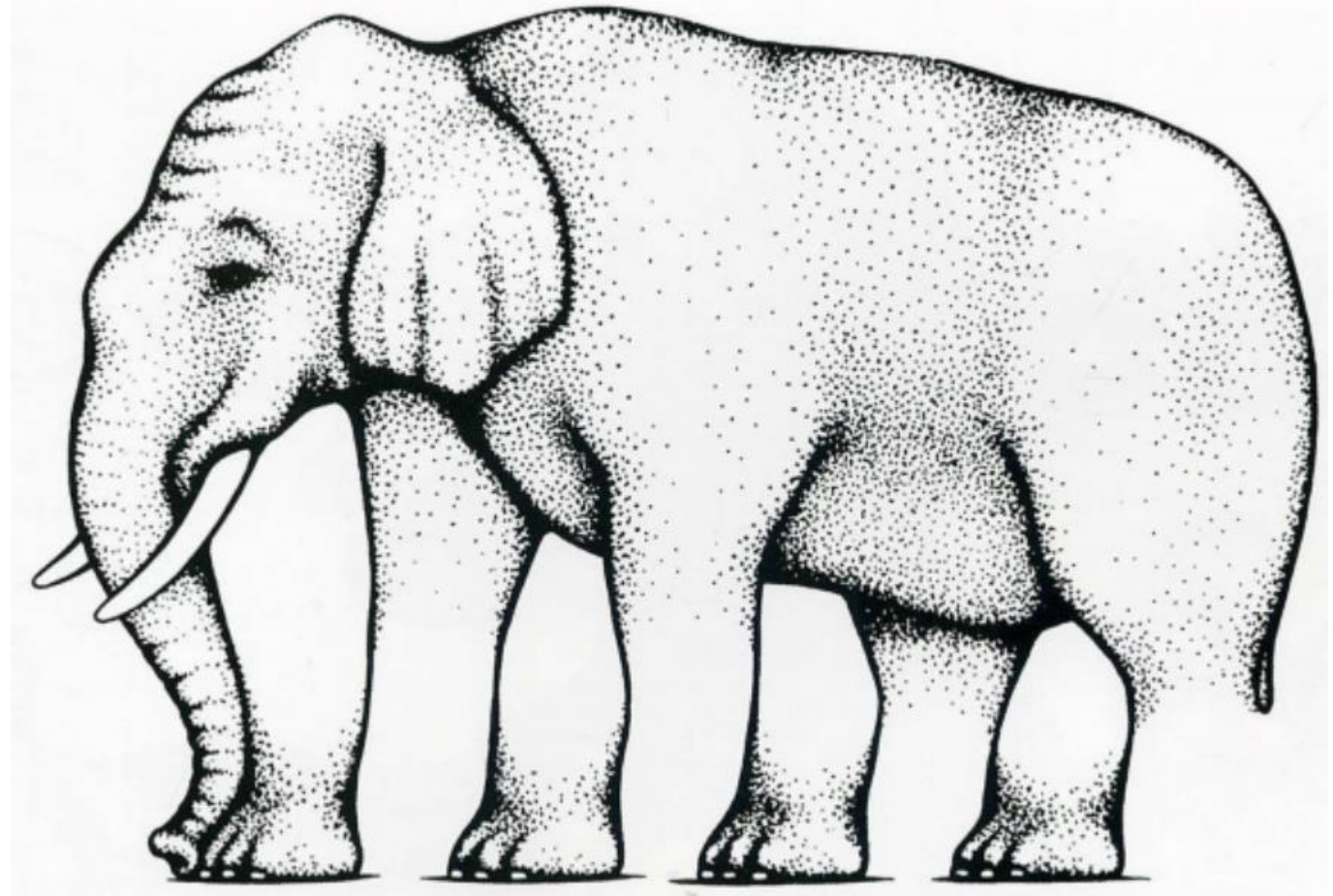
Зрительная система

- Иерархическая организация полей нейронов.
- Каждое поле можно рассматривать как слой нейронов, который извлекает все более сложную информацию из потока данных.
- V1 – первичная визуальная кора: по сигналам с сетчатки распознает цветовые пятна, особые области и изменения интенсивности в разных направлениях.
- V2, V3, V4 : бинакулярное зрение, простые геометрические формы, движение.
- Inferior Temporal Cortex (PIT, AIT areas) - is a final stage in the ventral cortical visual system. Нейроны “узнают” лица, эмоции, руки, жесты и другие сложные объекты.

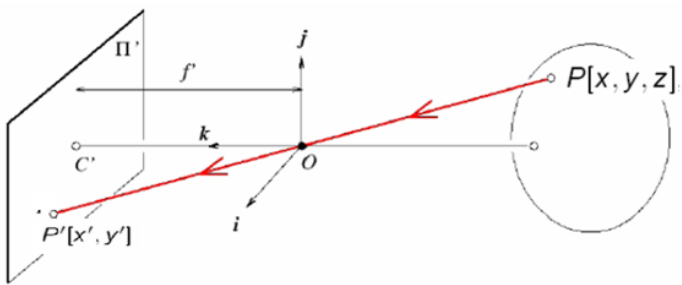


Целостность зрительного восприятия

- Мозг стремится к целостности.
- Нейроны “высших” уровней постоянно предсказывают что будет, активируются и посылают сигнал нейронам низших уровней. И те уже как бы “знают” на что обратить большее внимание.
- Через цепочки обратной связи мозг постоянно сравнивает предсказания с реальностью и корректирует модель.
- Если предсказание сильно, а сигналы с сетчатки не отчетливы, то мозг может и проигнорировать реальность! В какой-то степени мы живем в собственной галлюцинации 🖥️



Эволюция камер



$$x' = f' \frac{x}{z}$$

$$y' = f' \frac{y}{z}$$

Не только RGB

- Инфракрасное зрение
- Термальное зрение
- Рентгеновское зрение
- Терагерцовое зрение
- 3D зрение, стерео
- ...

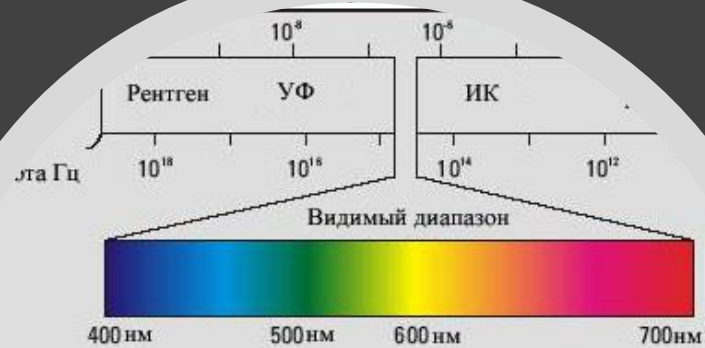
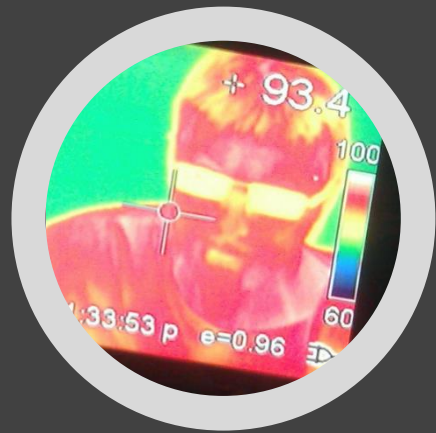


рис. 1



Эволюция алгоритмов компьютерного зрения

Как все начиналось: 196х

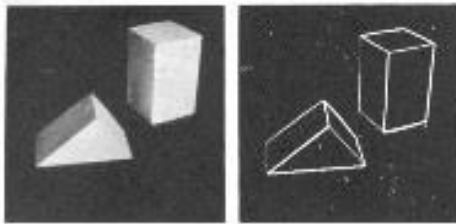
Machine Perception Of Three-Dimensional Solids

Massachusetts Institued of Technology

Lincoln Laboratory

22 May 1963

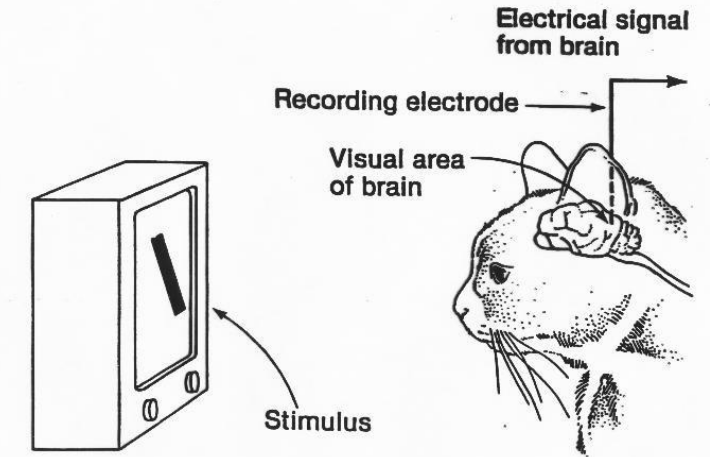
Lawrence G. Roberts



1959: An apparatus that allowed transforming images into grids of numbers (176x176)



1959: Receptive fields of single neurons in the cat's striate cortex



1966: IBM's speech recognition



1963: Vladimir Vapnik: Support Vector Networks

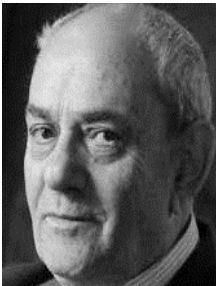
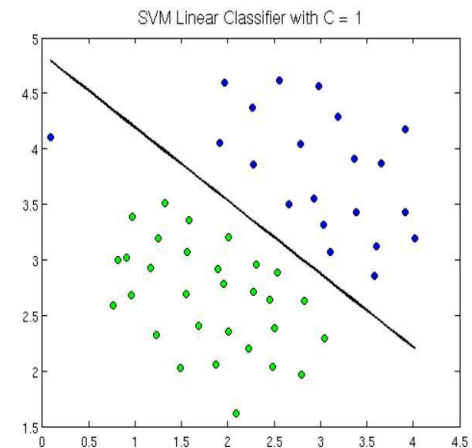


Figure 1: Camera Transformation

$$ax + by = c$$

$$\begin{aligned} a &= \sum x^2 - \frac{(\sum x)^2}{n} \\ b &= \sum y^2 - \frac{(\sum y)^2}{n} \\ c &= \sum xy - \frac{(\sum x)(\sum y)}{n} \end{aligned}$$

1966:
1st chatbot

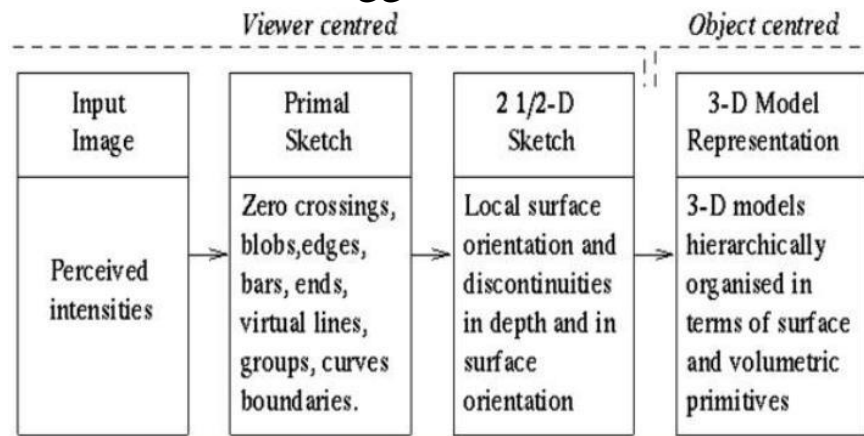
```
=====
EEEEEEE L      IIIIII ZZZZZZ   AAA
EEEEEE  L      I      Z      A  A
EEEEEE  L      I      Z      A  A
EEEEEE  L      I      Z      A  A
EEEEEEE LLLLLLL IIIIII ZZZZZZ   A  A
=====
ELIZA > (PLEASE TYPE IN ALL CAPS) WHAT'S YOUR NAME DEAR ?
AMIT
ELIZA > HI, I'M ELIZA. WHAT DO YOU WANT TO TALK ABOUT ?
AMIT > CAN I TALK ABOUT MY PROBLEM ?
ELIZA > SURE...!
AMIT > DUE TO ON GOING EXAMS IN COLLEGE, I HAVE A LOT OF STRESS.
ELIZA > PLEASE RELAX A LITTLE. SLEEP WELL.
AMIT > THANKS FOR YOUR ADVICE.
ELIZA > NO MENTION.
AMIT > BYE
ELIZA > BYE AND KEEP IN TOUCH...
=====
```

1965: Herbert Simon predicts that "Machines will be capable, within 20 years, of doing any work a man can do"
AI became an academic discipline.

197x:

1971: The "Stanford cart" – автономная машина, объезжающая препятствия

1979: David Marr's and Tomaso Poggio's "2 1/2 sketch"



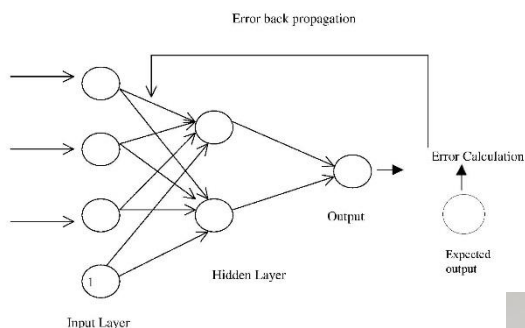
http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/GOMES1/marr.html

1974: Paul Werbos' backpropagation algorithm for neural networks

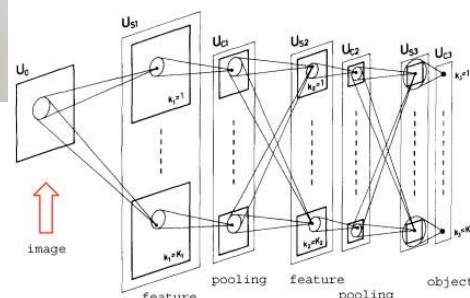


BEYOND REGRESSION:
NEW TOOLS FOR PREDICTION AND ANALYSIS
IN THE BEHAVIORAL SCIENCES

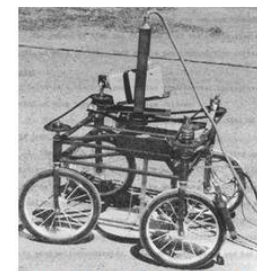
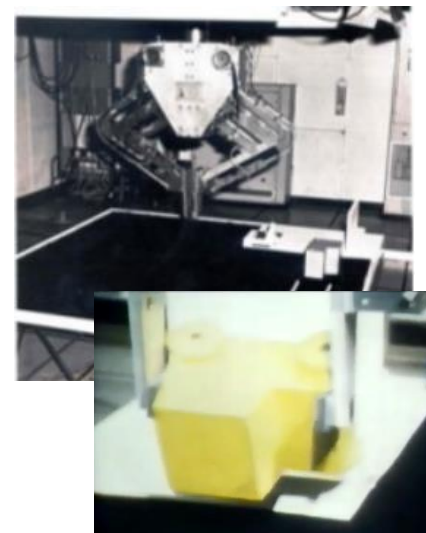
Harvard University
Cambridge, Massachusetts
August, 1974



1979: Kunihiro Fukushima's convolutional neural network ("Neocognitron - A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position")



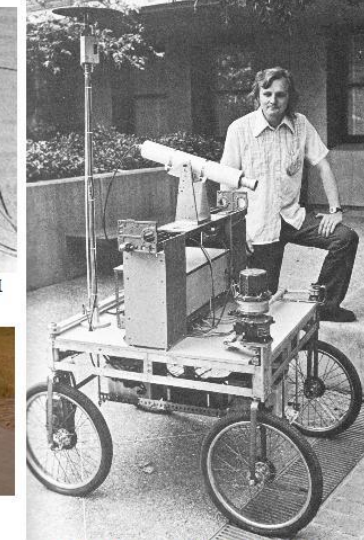
1973: Freddy – робот с машинным зрением



Jim Adams' "cart" in 1961



Rodney Schmidt's "cart" in 1971



Hans Moravec's "cart" in 1977

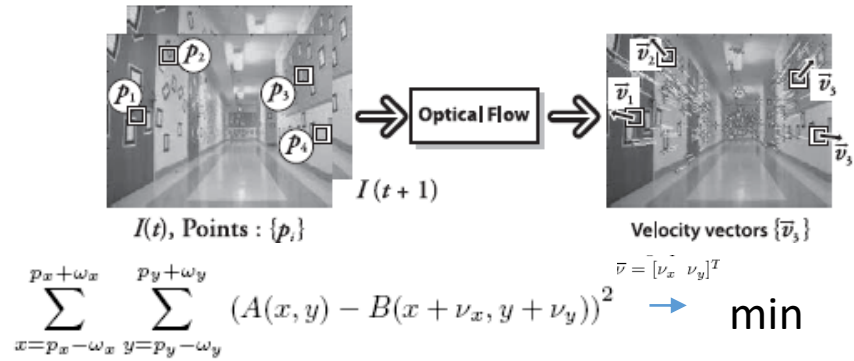
1974, [Kurzweil Computer Products](#) offered their first optical character recognition (OCR)



The growing criticism of the AI and computer vision technology resulted in "AI winter".

198x:

1981: famous
[Lukas-Kanade](#)
[optical flow](#)
[algorithm](#)



1986: classical [Canny edge detector](#)

1987: EigenFace algorithm



1989: Yann LeCun's convolutional
neural network for handwritten-digit
recognition (LeNet-1)

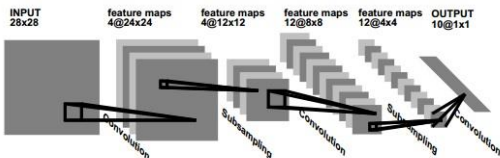
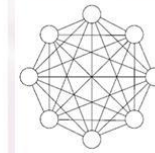


Figure 1: Architecture of LeNet 1

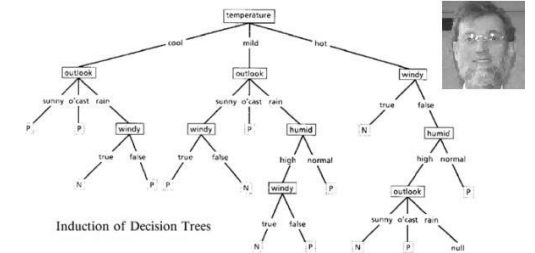


1982: John Hopfield describes a
new generation of neural
networks, based on recurrence

Hopfield networks
Multidirectional data flow
Total integration between input and
output data
All neurons are linked between
themselves
Trained with or without supervision



1985: Ross Quinlan's ID3 for
decision trees analysis

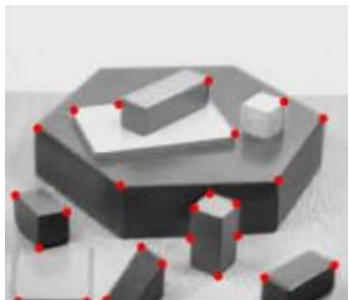


1988: Dean Pomerleau's self-driving vehicle ALVINN



199x:

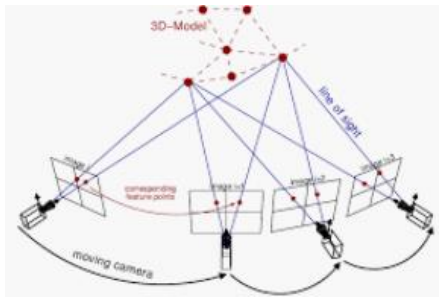
1994: Shi and C. Tomasi.
Good Features to Track.



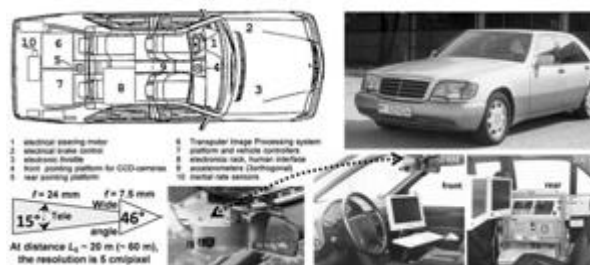
1999: David Lowe, Object recognition
from local scale-invariant features



1999: Triggs et al.,
"Bundle Adjustment —
A Modern Synthesis"



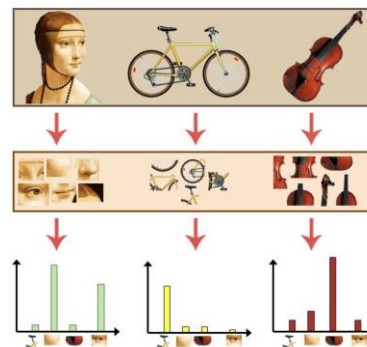
1994: Ernst Dickmanns' self-driving car drives more than 1,000
kms near the airport Charles-de-Gaulle in Paris



1997: IBM's "Deep Blue" chess
machine beats the world's chess
champion, Garry Kasparov

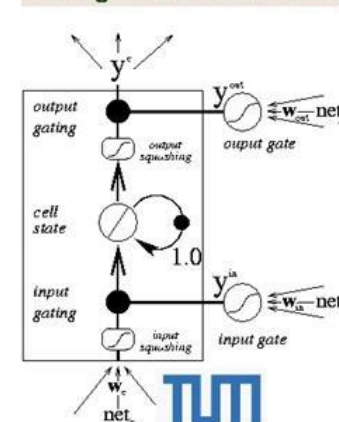


Object → Bag of 'words'



OpenCV (Open Source Computer Vision) -
a popular computer vision library started by
Intel in 1999.

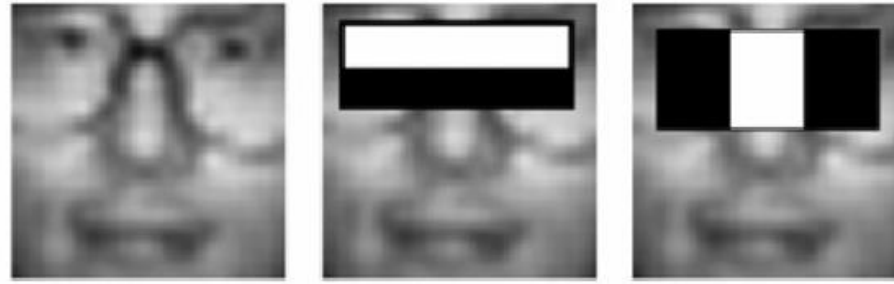
Long Short-Term Memory



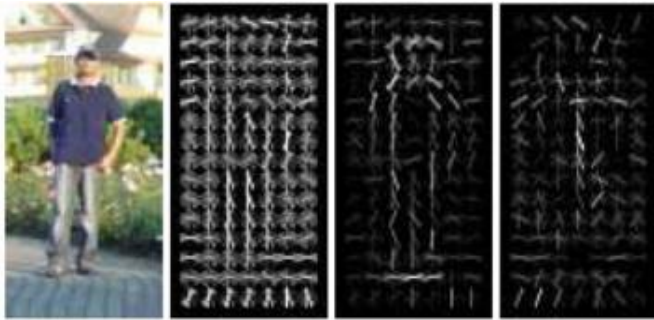
1997: Jurgen Schmidhuber's
and Sepp Hochreiter's Long
Short Term Memory (LSTM)
model

200x:

2001: Viola/Jones face detector



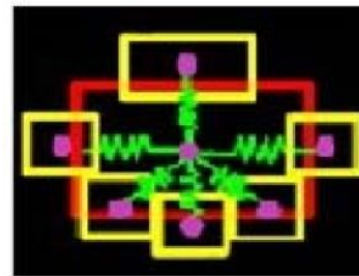
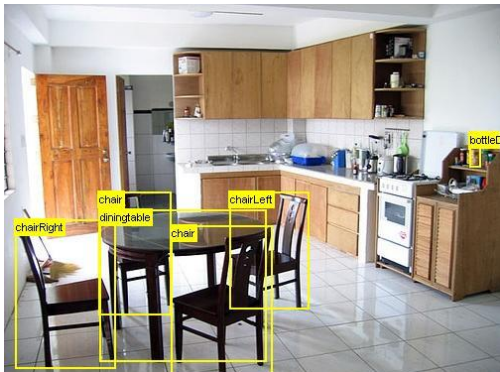
2005, Dalal, Triggs, [Histograms of Oriented Gradients for Human Detection](#)



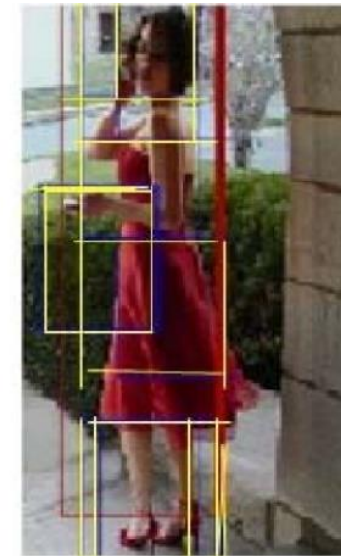
2005: driverless car Stanley wins DARPA's Grand Challenge



2007, The PASCAL Visual Object Classes Challenge 2007



2009: Felzenszwalb et al, [Deformable Part Model](#)

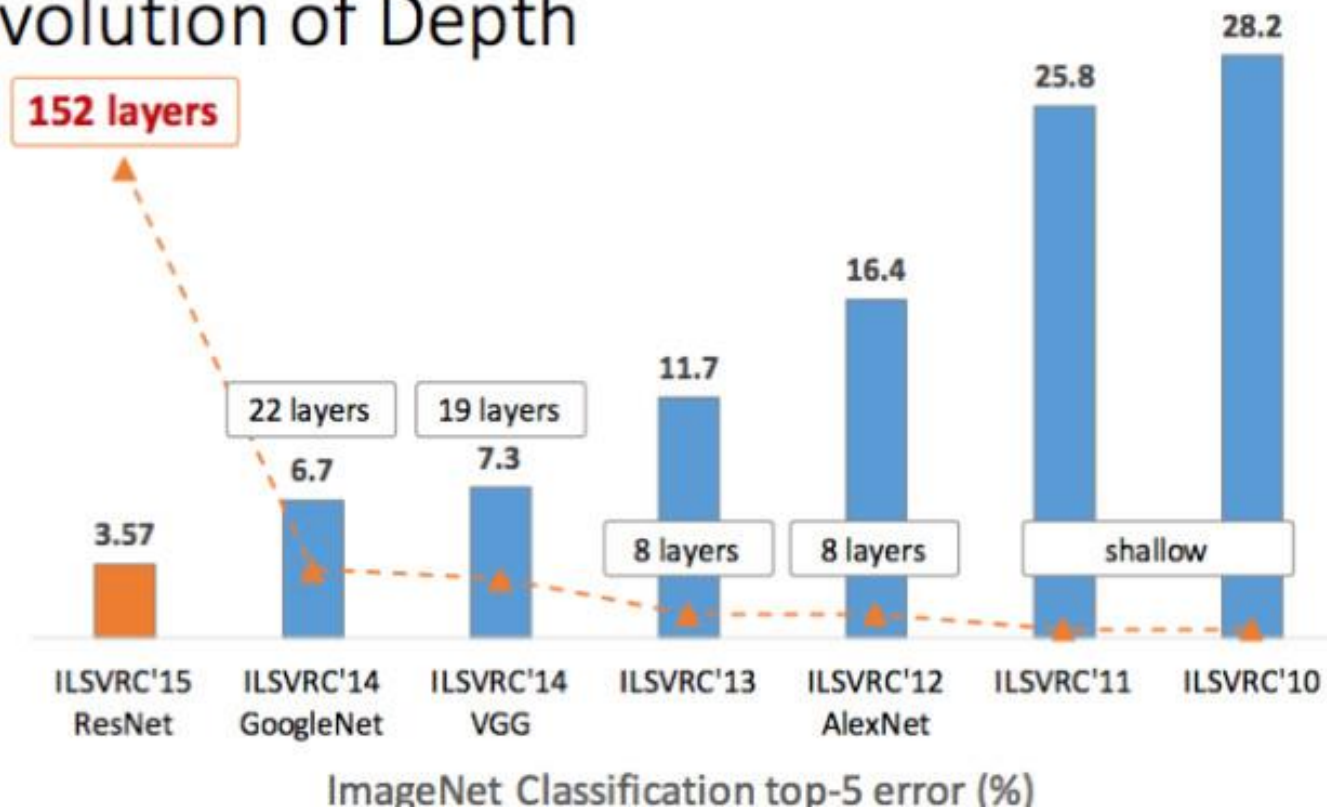


201x

- 2010: The ImageNet Large Scale Visual Recognition Competition
- 2012 when [AlexNet won ImageNet](#)
- 2014, Goodfellow et al, **Generative Adversarial Networks**



Revolution of Depth

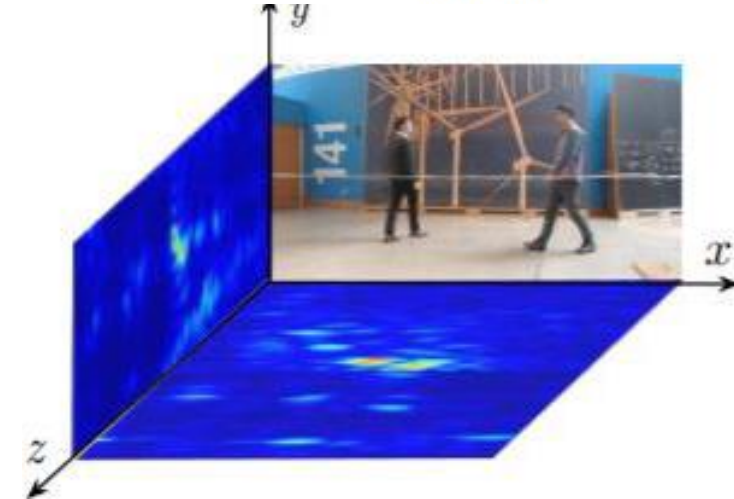


Видеть невидимое



Through-Wall Human Pose Estimation Using Radio Signals

Mingmin Zhao Tianhong Li Mohammad Abu Alsheikh Yonglong Tian Hang Zhao
Antonio Torralba Dina Katabi
MIT CSAIL

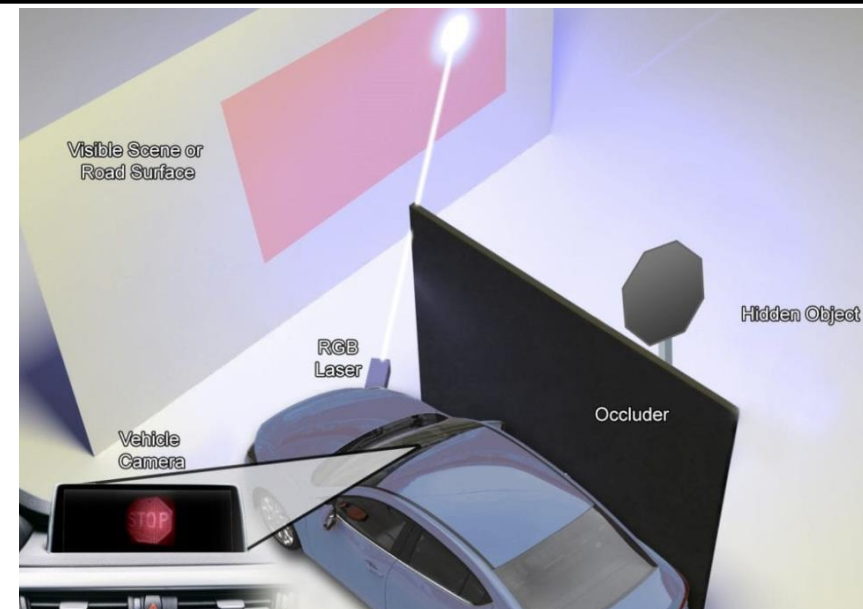


CVPR 2019 Best Paper Award:



A Theory of Fermat Paths for Non-Line-of-Sight Shape Reconstruction

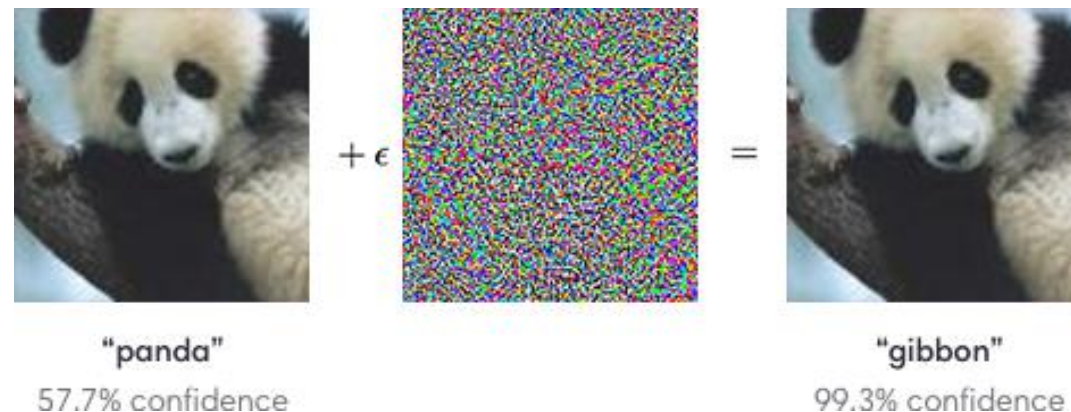
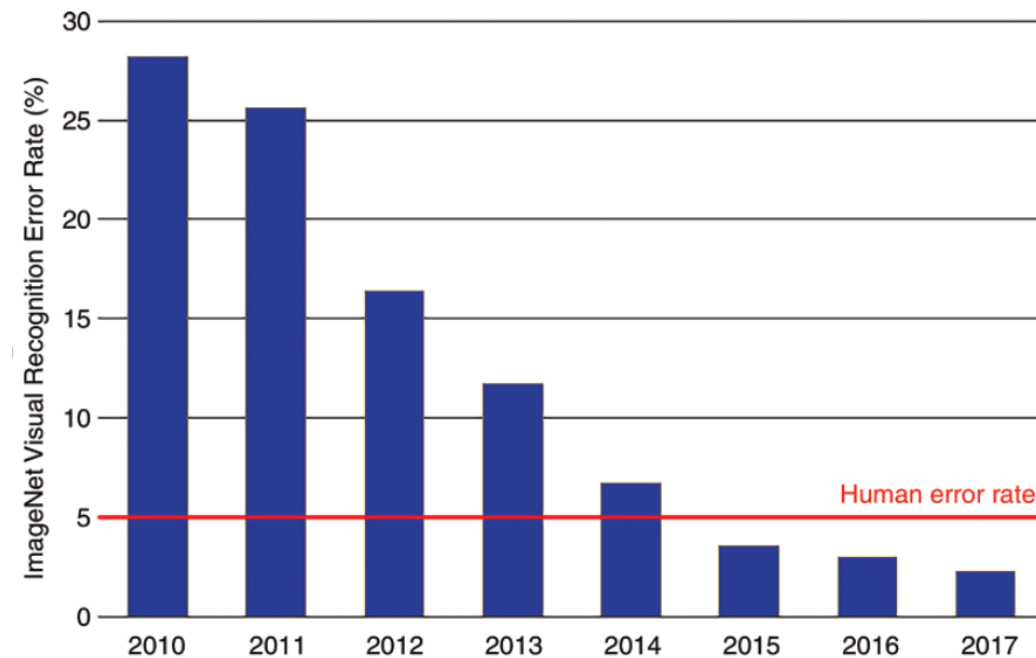
Shumian Xin¹, Sotiris Nousias^{2,3}, Kiriakos N. Kutulakos², Aswin C. Sankaranarayanan¹, Srinivasa G. Narasimhan¹, and Ioannis Gkioulekas¹
¹Carnegie Mellon University ²University of Toronto ³University College London



Сегодня

Алгоритмы компьютерного зрения:

- Лучше человека классифицируют изображения и видео.
- Лучше человека распознают человека.
- Рисуют и подделывают (DeepFake).
- Видят невидимое.
- НО!





Умный Ганс

RISE: Randomized Input Sampling for Explanation of Black-box Models

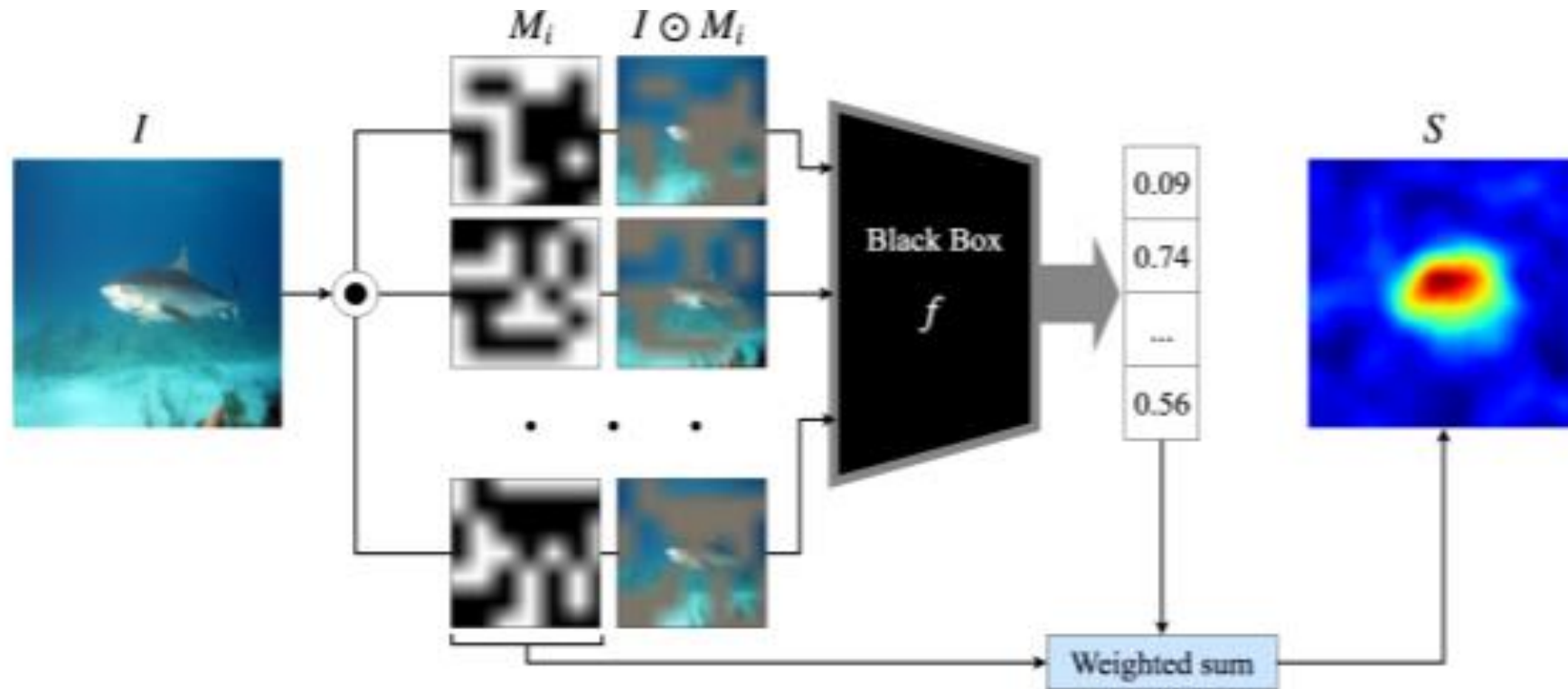
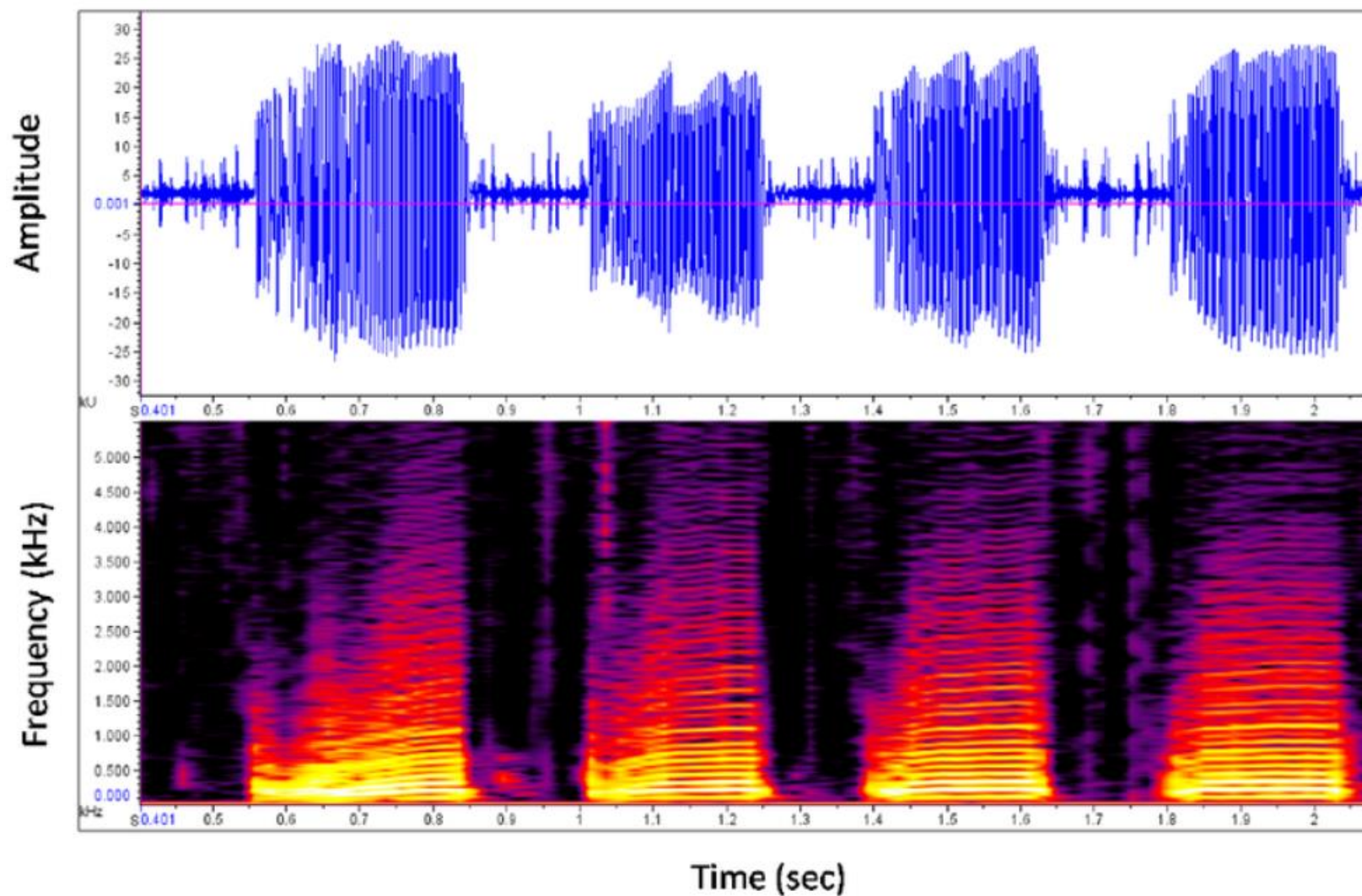


Figure 3: Overview of RISE: Input image I is element-wise multiplied with random masks M_i and the masked images are fed to the base model. The saliency map is a linear combination of the masks where the weights come from the score of the target class corresponding to the respective masked inputs.

Кросс-модальность алгоритмов



- Аудио сигнал с помощью преобразования Фурье можно перевести в картинку и применить CNN.

Заключение

- Появление “полноценного” зрения в природе возможно спровоцировало эволюционный взрыв. То же будет и в технике?
- Технологическая эволюция зрения ускоряется с каждым годом. Эволюция алгоритмов: от оптимизма (196х) к еще большему оптимизму (201х).
- Компьютерное зрение уже решает некоторые задачи лучше человека. Но эти решения могут быть не устойчивы к шуму и к исключительным случаям.
- Глубокая нейронная сеть - “черный ящик”.

Умное стекло

- Computation is performed by a host material with numerous inclusions that appear as bubbles and impurities embedded in the glass. These inclusions scatter light in both forward and backward directions to differentiate images.
- The scattering mixes the input light in a manner that can be compared to linear matrix multiplication in a digital artificial neural network (ANN).
- The locations and shapes of the inclusions can be compared to weight parameters in digital ANNs.
- At the output, the optical energy is concentrated in well-defined locations, which, for example, can be interpreted as the identity of the object in the image.

