

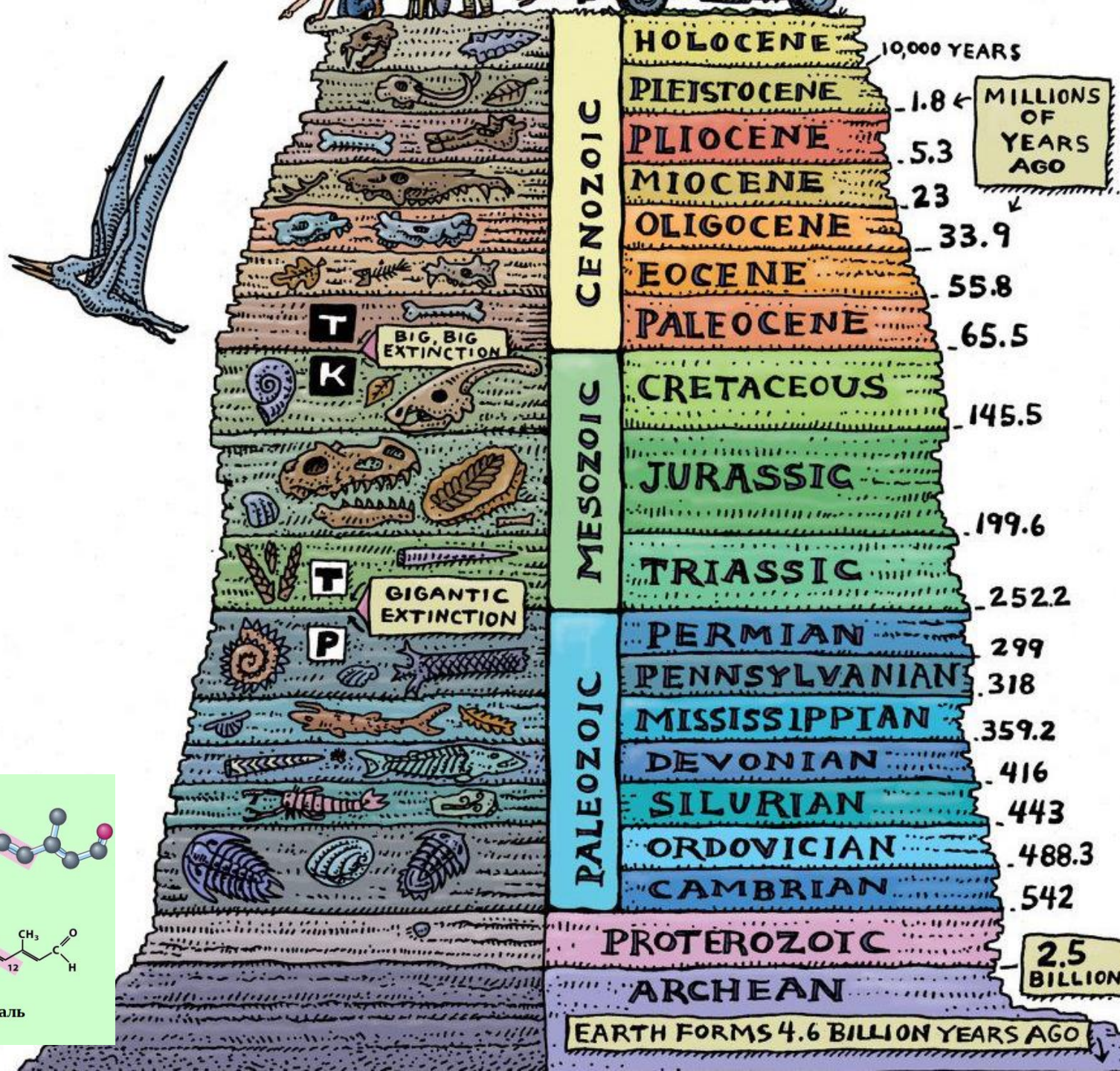
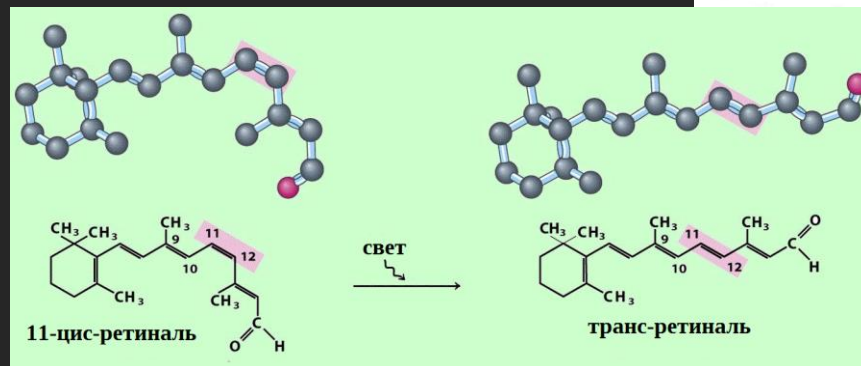


Эволюция и устройство зрения в природе и в машинах

Александр Бовырин

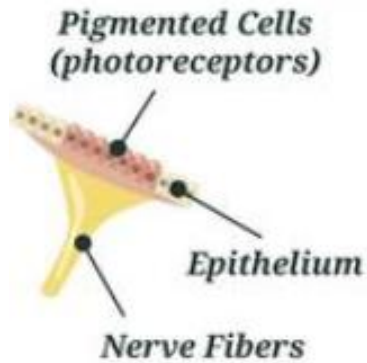


Когда
появилось
зрение?

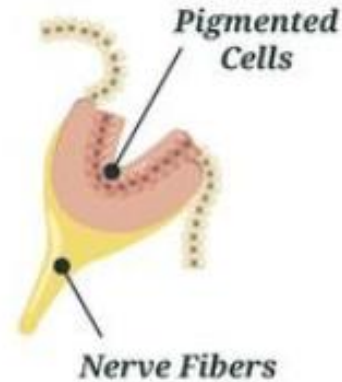


Эволюция глаза

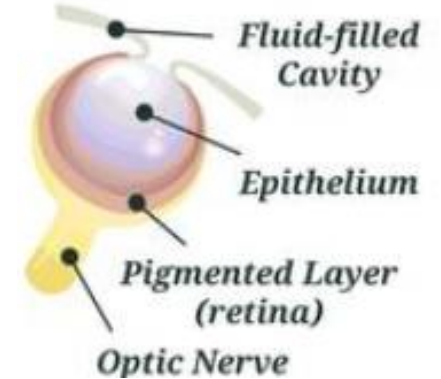
1. Светло или темно?



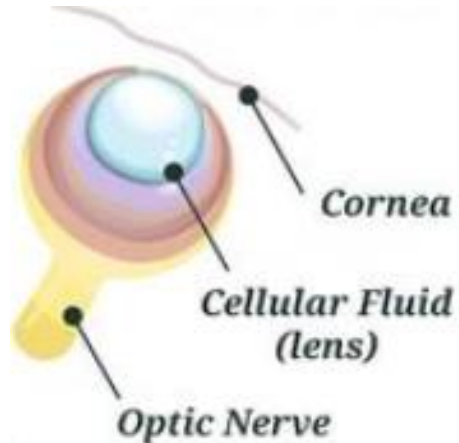
2. Где светло?



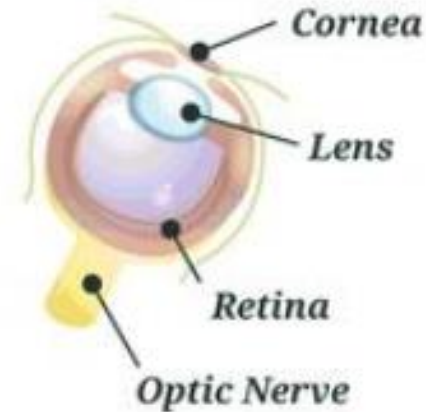
3. Камерный глаз.
Появляются расплывчатые образы.



4. Камерный глаз.
Появляются расплывчатые образы.



5. С линзой изображение становится резким.



Такое разное зрение...



Почему хищники в основном обладают бинокулярным зрением?

Почему нехищникам выгоднее панорамное зрение?



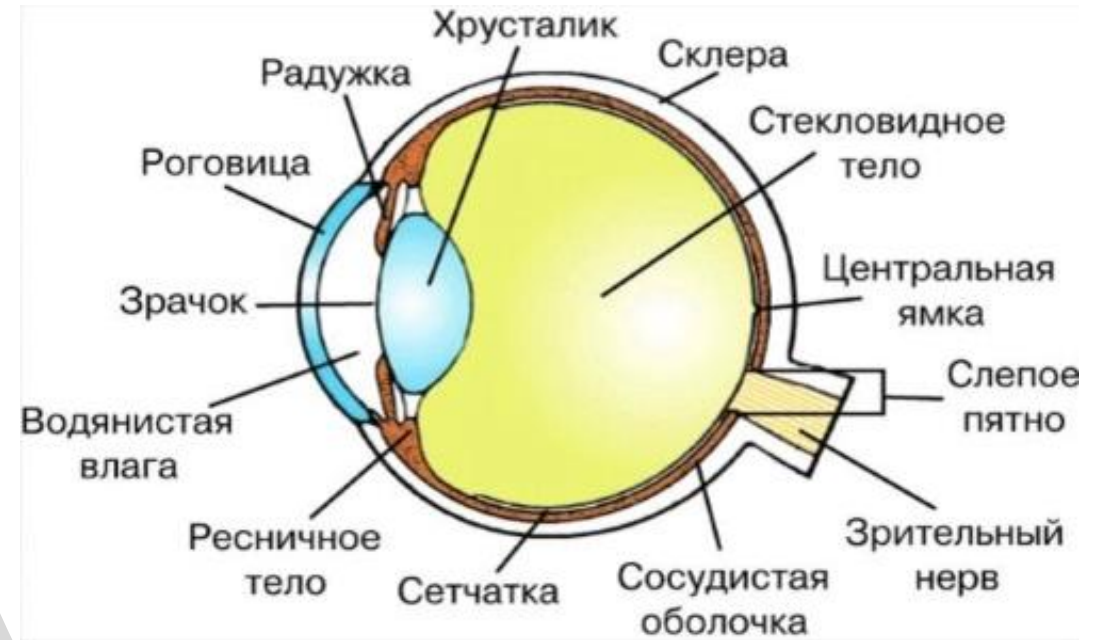
Почему глаза некоторых видов светятся в темноте?

Животные имеют разные наборы фоторецепторных белков, и их соответствующие спектры поглощения отличаются от человеческих белков. Поэтому восприятие света и изображений у человека и животных отличается. Некоторые [насекомые](#), например, могут видеть в ультрафиолетовом диапазоне, а животные с одним типом опсинов видят окружающий мир черно-белым.



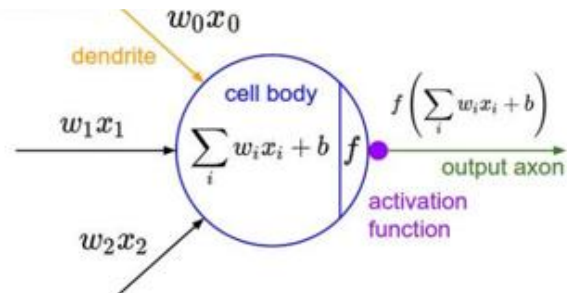
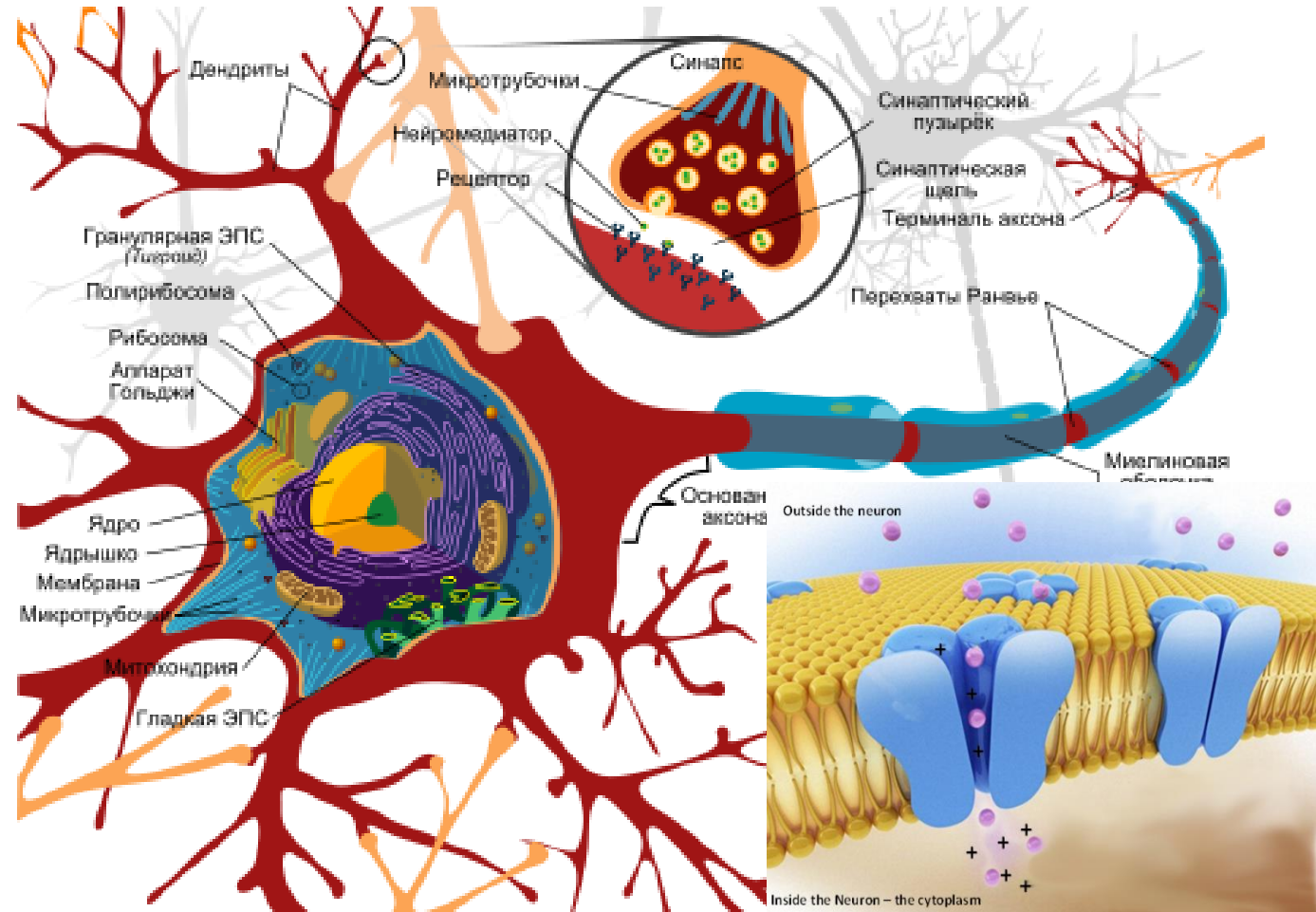
Глаз человека

- 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.



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 (2-30ms).
- Для передачи импульса требуется ~1 ms (зависит от типа нейрона и длины аксона)



<1 ms	+	+	+	+	+	+	+	+	+
	-	-	-	-	-	-	-	-	-
Axon	-	-	-	-	-	-	-	-	-
	+	+	+	+	+	+	+	+	+

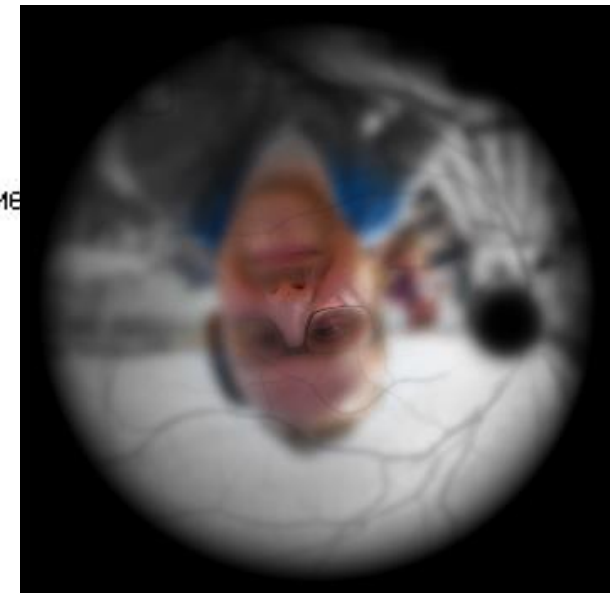
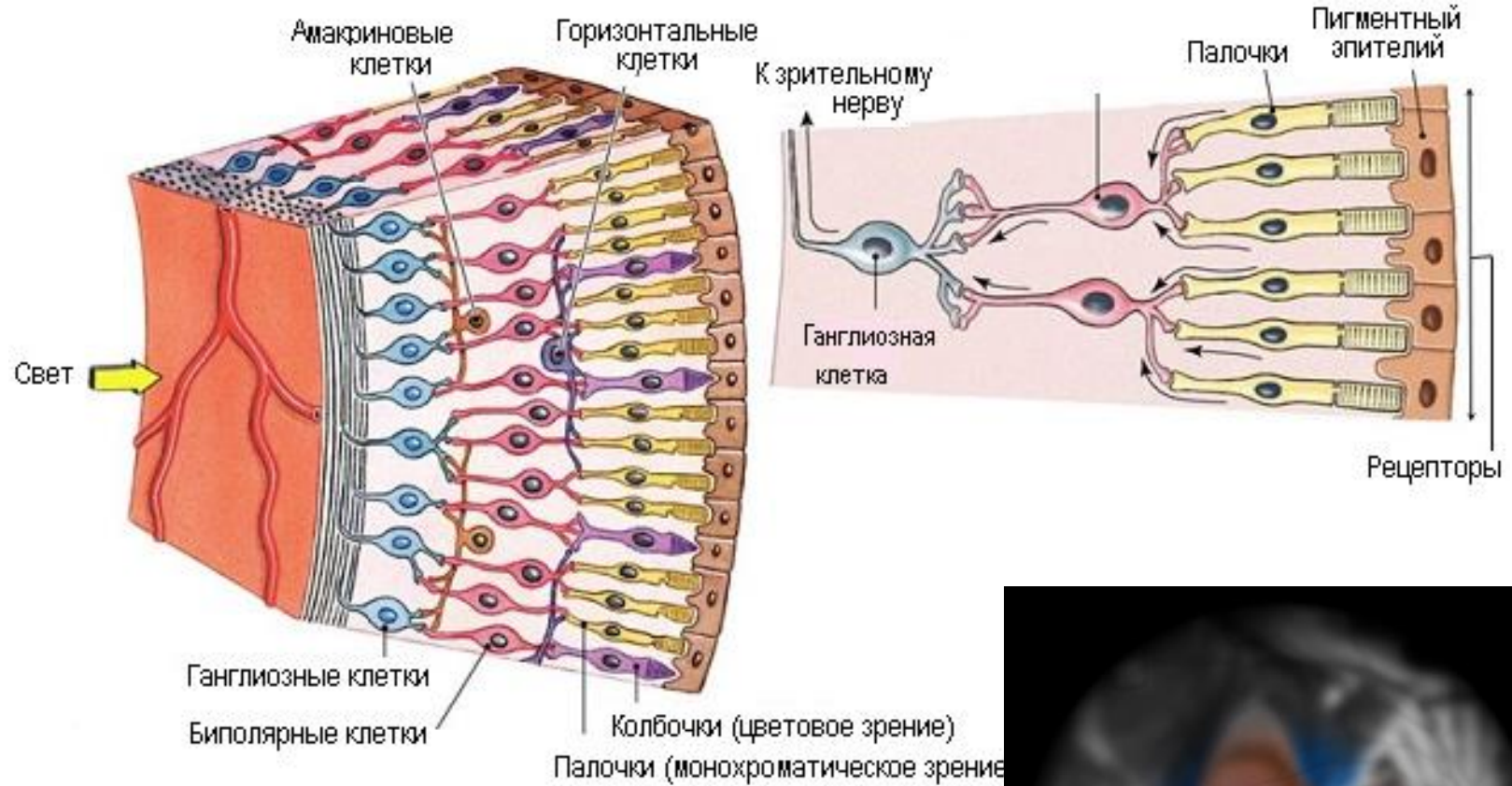
Мембрана нейрона содержит каналы и рецепторы, которые их открывают или закрывают. Есть белковые насосы, которые откачивают лишние ионы из клетки, восстанавливая потенциал.

Память

- Кратковременная память обусловлена выбросом нейромедиатора, который усиливает синаптическую передачу.
- При формировании долговременной памяти нейроны отращивают новые окончания, приобретают новые связи, усиливают старые.
- Было обнаружено, что у скрипачей и виолончелистов область коры, отвечающая за пальцы левой руки, которыми они зажимают струны, в два раза больше, чем в мозге нем музыканта.
- Если многократно вызывать у нервной системы привыкание, то нейроны, наоборот, втягивают имеющиеся окончания, а их связи становятся неактивными.

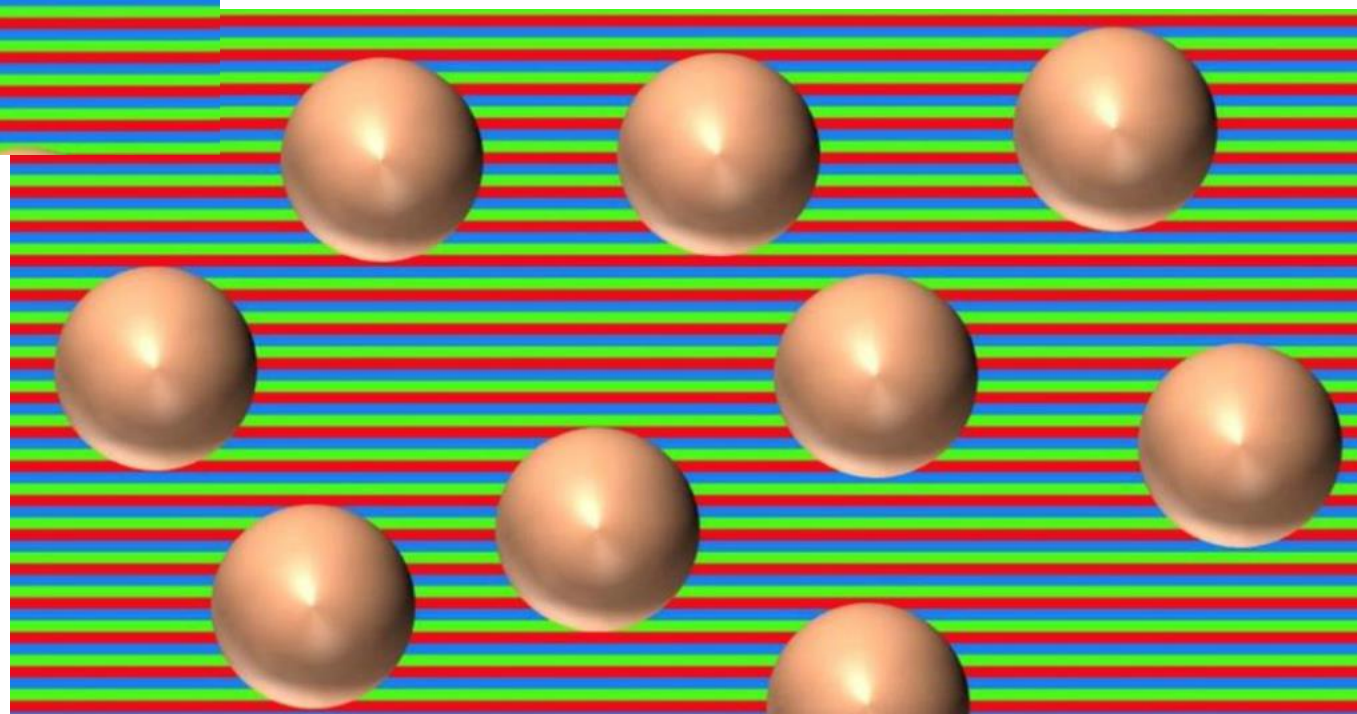
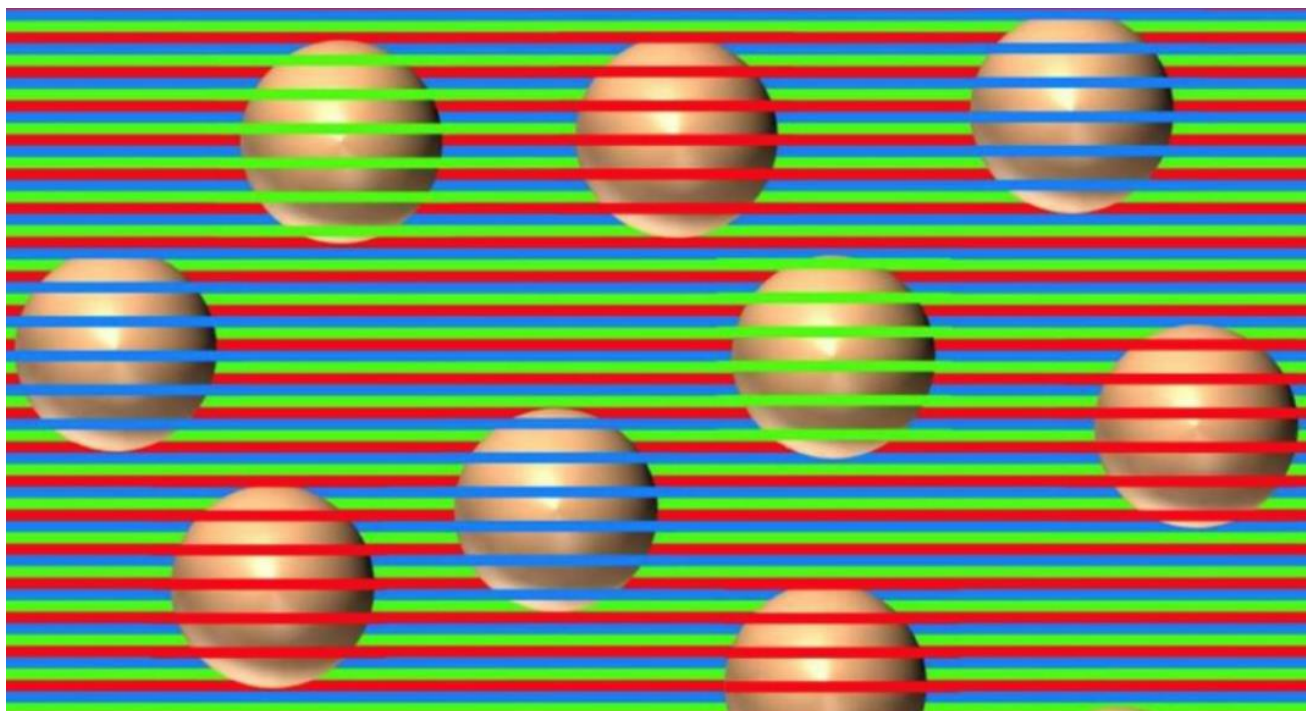
Сетчатка – simple IPU & edge detector

- 400-750 нм
- ~10 млн колбочек
- ~100 млн палочек
- Достаточно 1 фотона для реакции
- Рецептивное поле в центре меньше.
- Соседние нейроны взаимно гасят друг друга
- Мы видим лишь края объектов. И то не долго.
(потому что есть постоянные микро движения глаз - саккады).



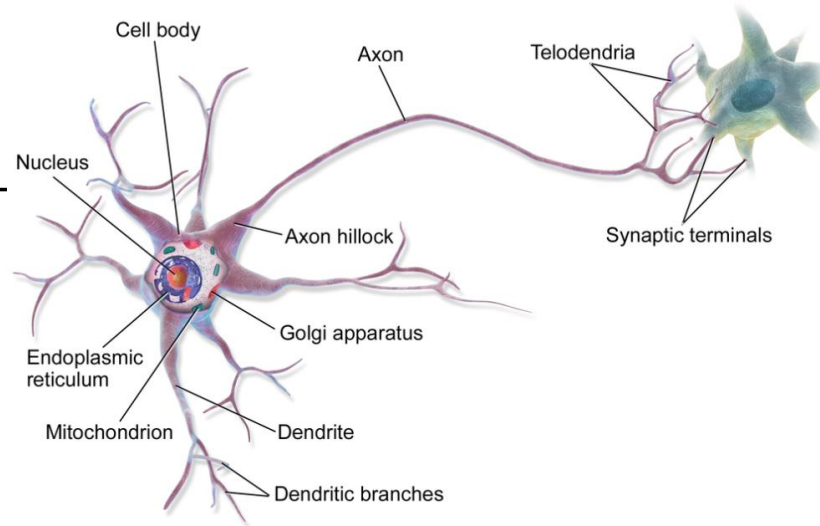
Слепое пятно





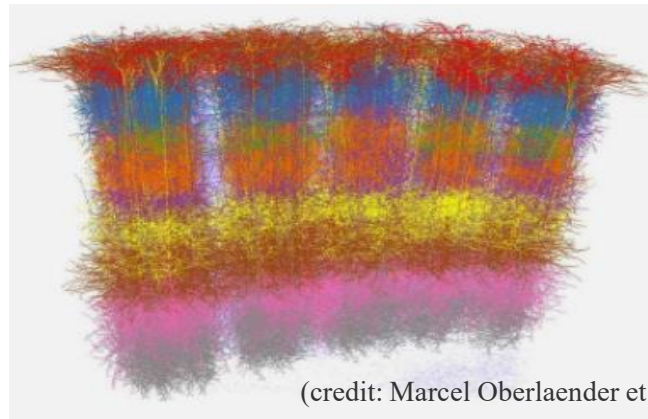
Каждую секунду мозг решает много задач : *восприятие*, *контроль органов*, *движение*, *обучение*, *разговор*, ...

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



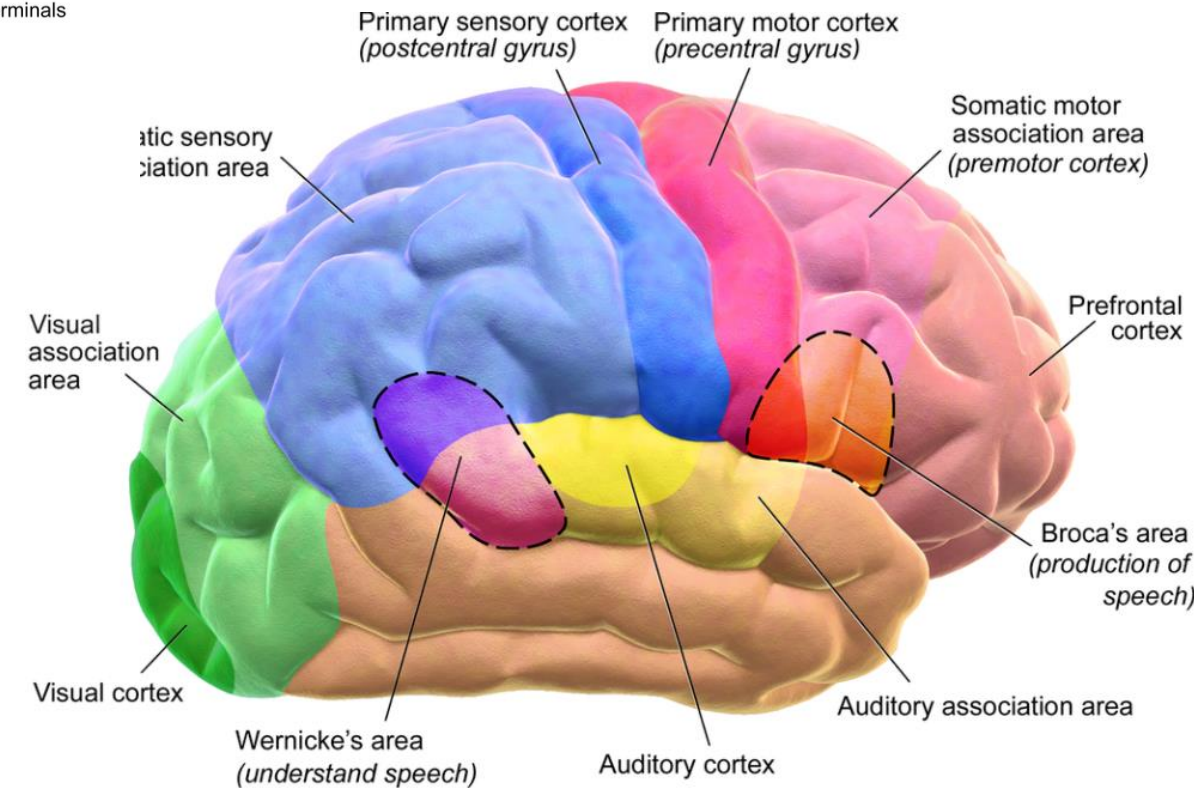
(credit: [BruceBlais](#))

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



(credit: Marcel Oberlaender et al.)

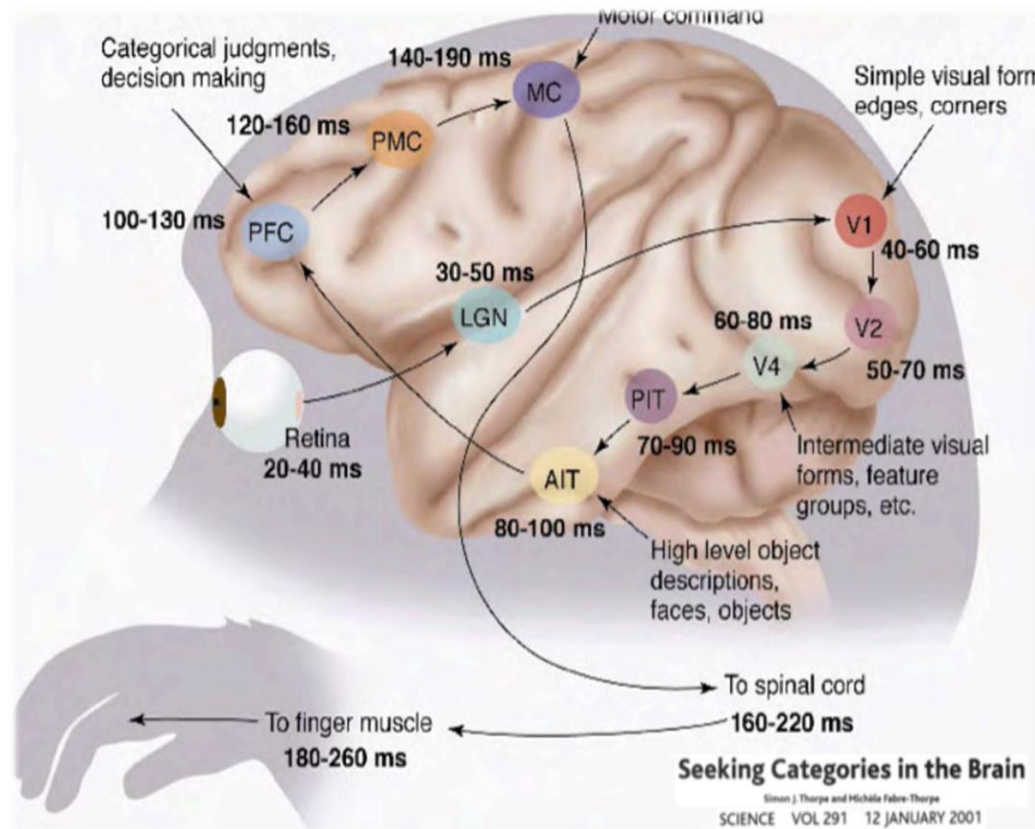
3. Мозг разделен на функциональные поля (visual, hearing, emotions, motion...).



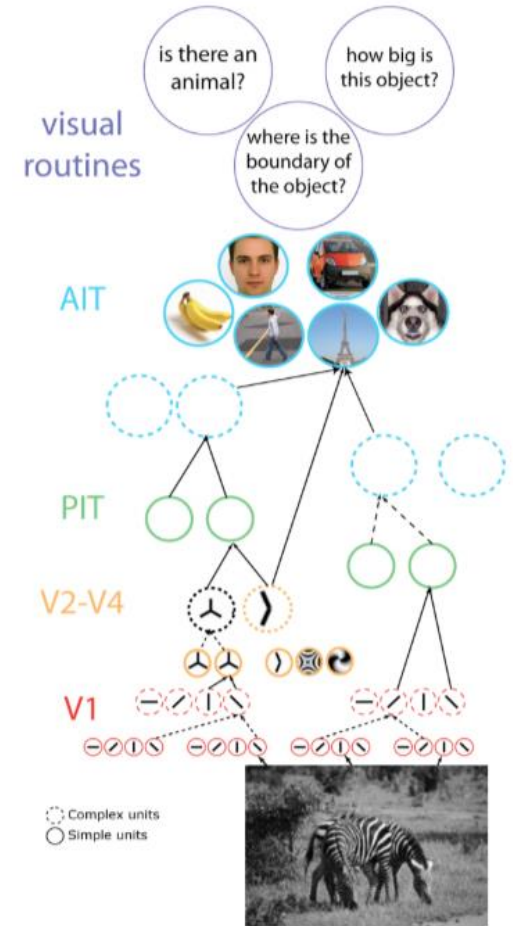
Credit: Blausen.com staff (2014). "[Medical gallery of Blausen Medical 2014](#)".

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.

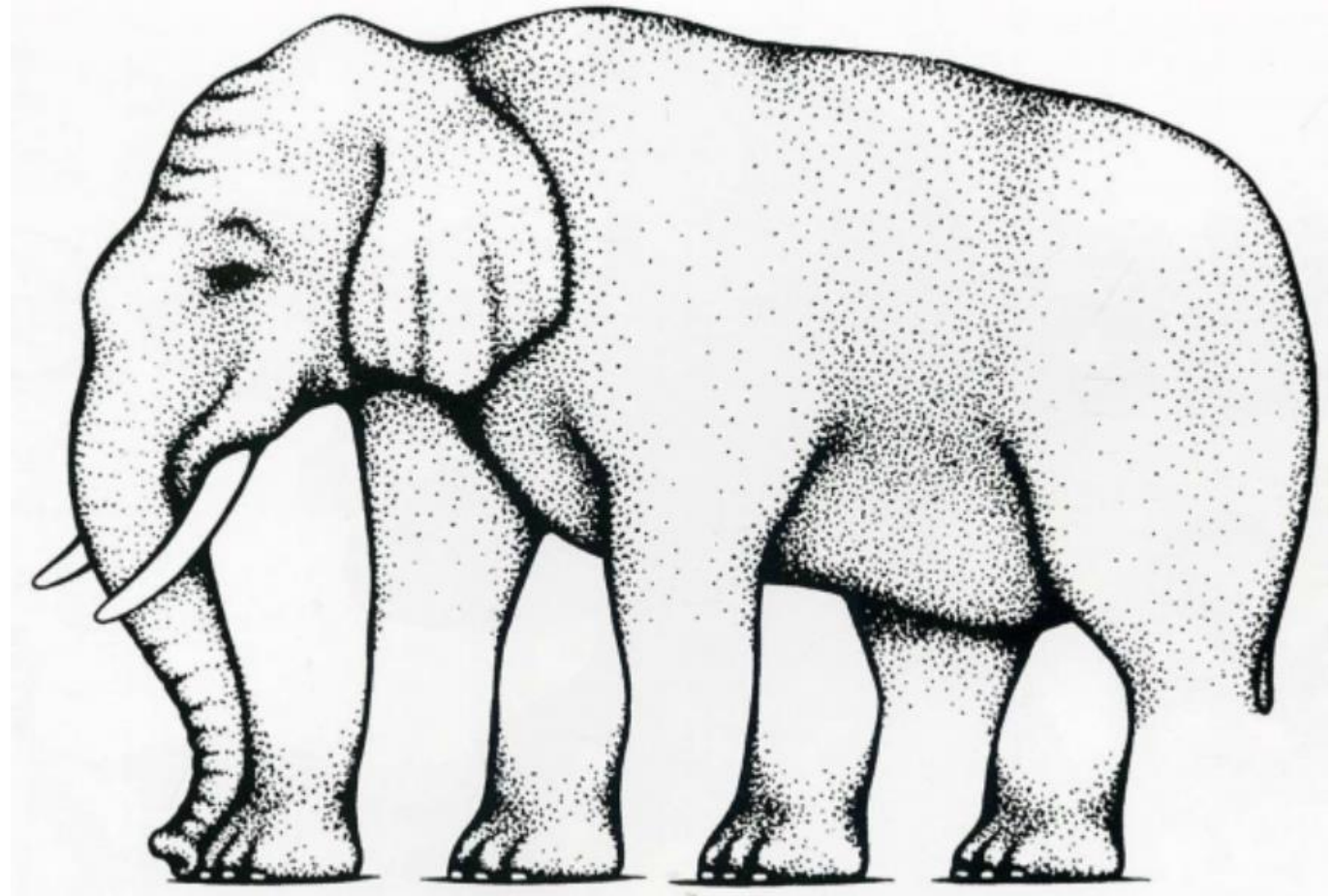


(credit: Simon Thorpe)



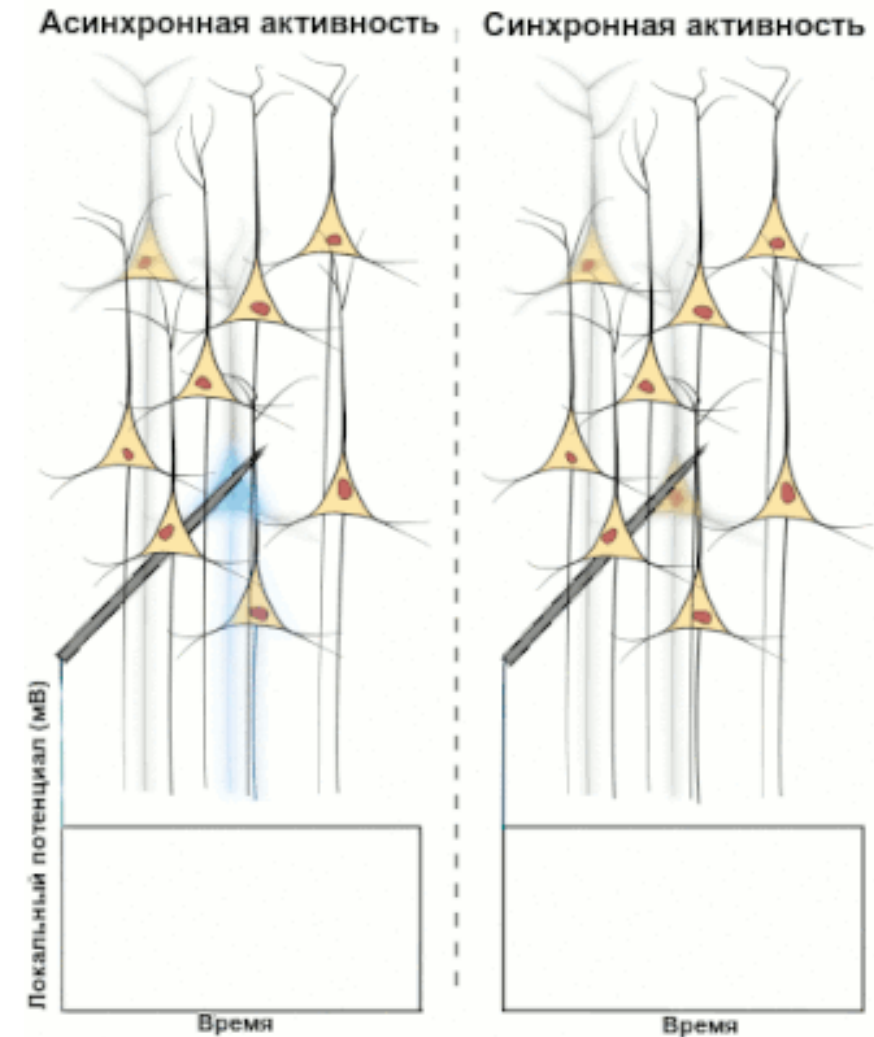
Целостность зрительного восприятия

- Мозг стремится к целостности.
- Нейроны “высших” уровней предсказывают что будет, активируются и посылают сигнал нейронам низших уровней. И те уже знают на что обратить 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 😊.

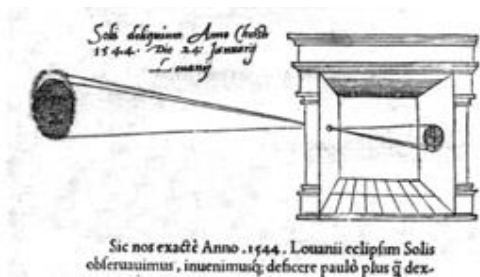
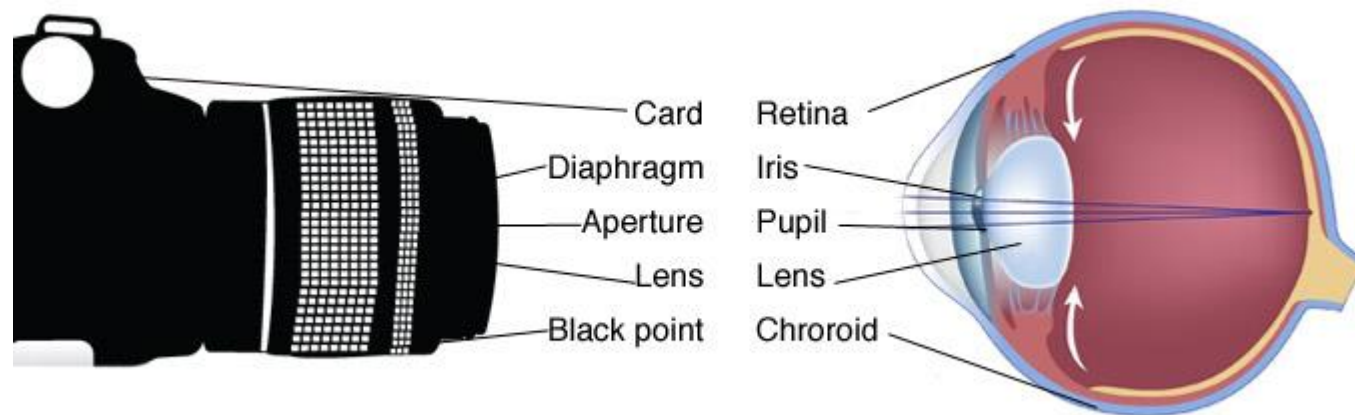


Импульсная передача и синхронная активность

- Если две удаленные популяции нейронов имеют синхронизированные коллективные колебания, то момент повышенной активности одной популяции совпадет с высокой возбудимостью другой, обеспечивая высокую вероятность передачи информации.
- Нейроны [зрительной коры](#), рецептивное поле которых содержит объект внимания, сильнее синхронизированы с локальным потенциалом (на γ -частоте), чем нейроны, которые реагируют на другие объекты вне внимания. Эта синхронизация повышает частоту срабатывания нейрона и его чувствительность (похоже на трэкинг).



Эволюция камер



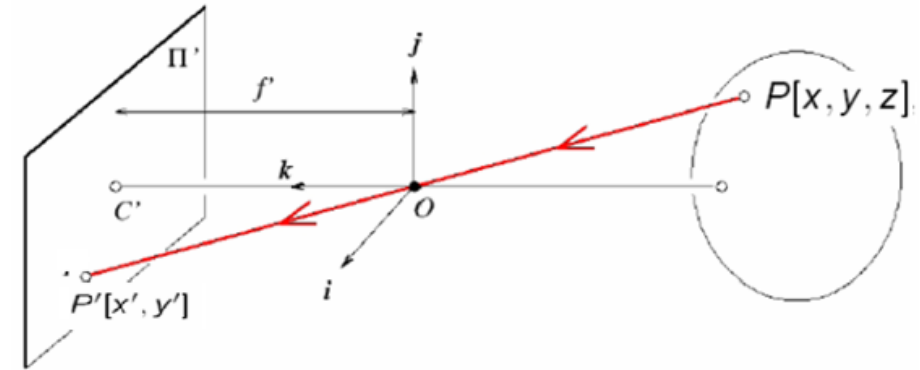
Не только RGB

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



Камера-обскура

- Точка проецируется в точку.
- Линия проецируется в линию.
- Углы не сохраняются.
- Параллельные линии могут пересекаться.



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

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

Линия проходящая через точку (X_0, Y_0, Z_0) в направлении (U, V, W) проецируется в точку $P_k = (f (X_0 + kU) / (Z_0 + kW), f (Y_0 + kV) / (Z_0 + kW))$.

При $k \rightarrow \infty$ получаем “точку схода” $P_{\infty} = (f U / W, f V / W)$.

Эволюция алгоритмов компьютерного зрения

We perceive this:



picture from:


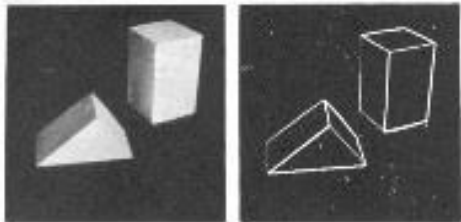
Gary Bradski, Adrian Kaehler, "Learning OpenCV:
Computer Vision with the OpenCV Library"

But the camera sees this:

194	210	201	212	199	213	215	196	178	198	182	209
188	189	190	221	209	205	191	187	147	119	129	183
114	126	140	188	176	165	152	148	170	106	78	88
87	103	113	154	143	142	149	153	173	181	87	87
102	112	106	171	122	139	152	147	128	84	58	66
94	93	79	104	103	124	129	113	107	87	69	67
68	71	69	98	89	92	98	95	89	88	76	67
41	58	68	89	63	43	88	82	88	76	75	83
28	43	69	75	56	41	51	73	55	70	63	44
58	58	37	69	70	70	73	74	53	68	58	37
72	59	53	66	84	92	84	74	57	72	63	42
47	61	55	65	75	78	78	73	59	75	69	58

Как все начиналось: 196х

Machine Perception Of Three-Dimensional Solids
Massachusetts Institued of Technology
Lincoln Laboratory
22 May 1963
Lawrence G. Roberts



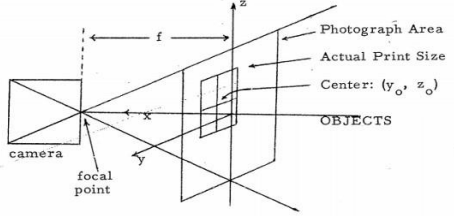


Figure 1: Camera Transformation

Photograph Area
Actual Print Size
Center: (y₀, z₀)
OBJECTS
camera
focal point

$ax + by = c$

$a = \sum x$	$\sum y^2$	$-\sum y$	$\sum xy$
$b = \sum y$	$\sum x^2$	$-\sum x$	$\sum xy$
$c = \sum x^2$	$\sum y^2$	$-\sum xy$	$\sum xy$

1966:
1st chatbot

```
=====
EEEEEEEE L      IIIIII ZZZZZZZ AAA
E      L      I      Z      A  A
EEEEEE  L      I      Z      A  A
E      L      I      Z      AAAAAAA
EEEEEEEE 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...
=====
```

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

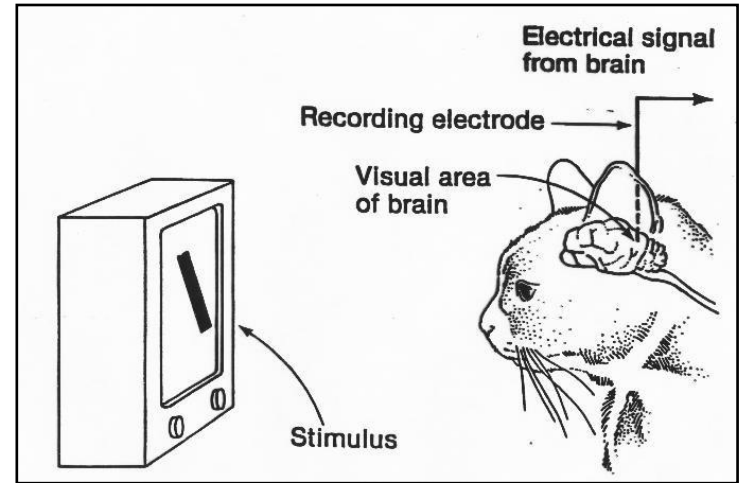


1966:IBM's speech recognition

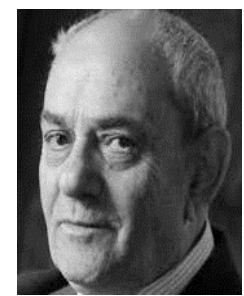
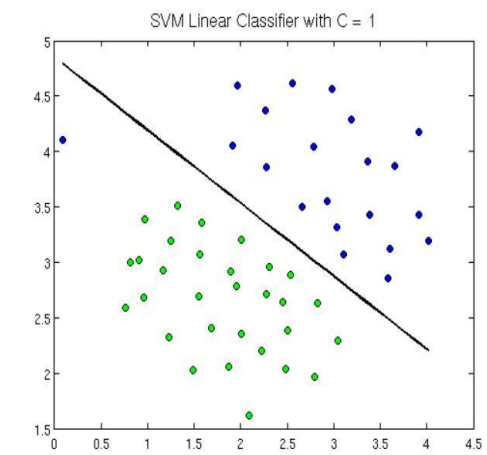


1965: Herbert Simon predicts that **"Machines will be capable, within 20 years, of doing any work a man can do"**
AI became an academic discipline.

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



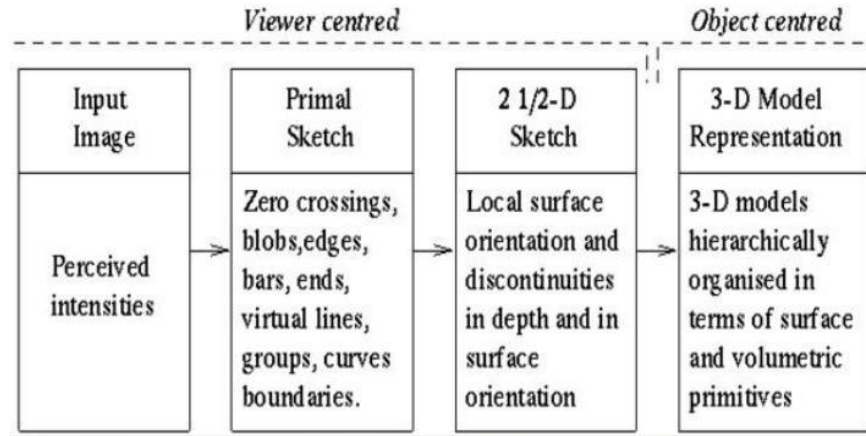
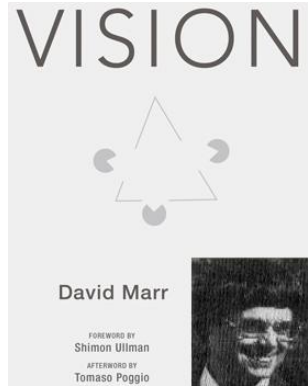
1963: Vladimir Vapnik: Support Vector Networks



197x:

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

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



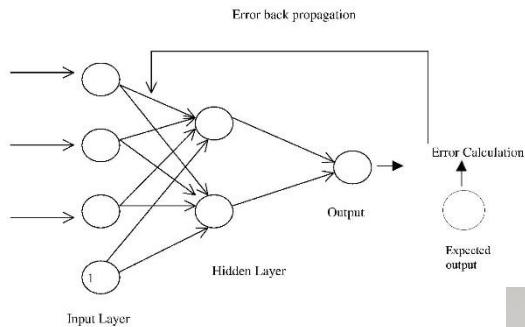
http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/GOMES1/marr.html

1974: Paul Werbos' backpropagation algorithm for neural networks

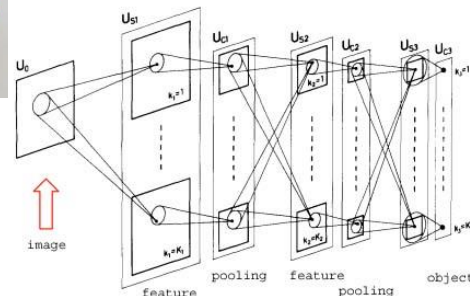


BEYOND REGRESSION:
NEW TOOLS FOR PREDICTION AND ANALYSIS
IN THE BEHAVIORAL SCIENCES

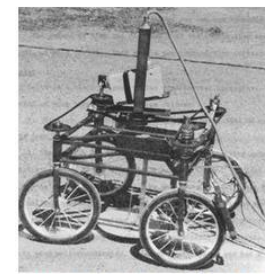
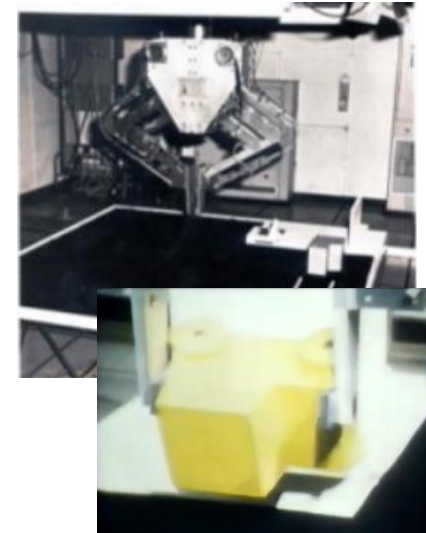
Harvard University
Cambridge, Massachusetts
August, 1974



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")



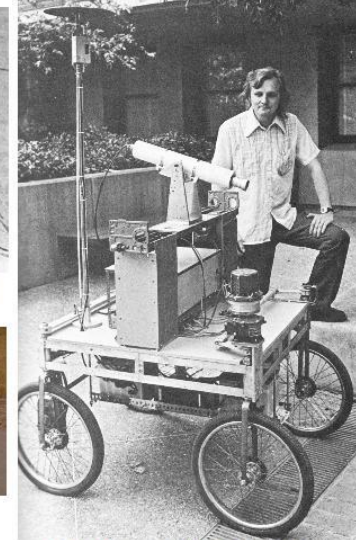
1973: Freddy – робот с машинным зрением



Jim Adams' "cart" in 1961



Rodney Schmidt's "cart" in 1971



Hans Moravec's "cart" in 1977

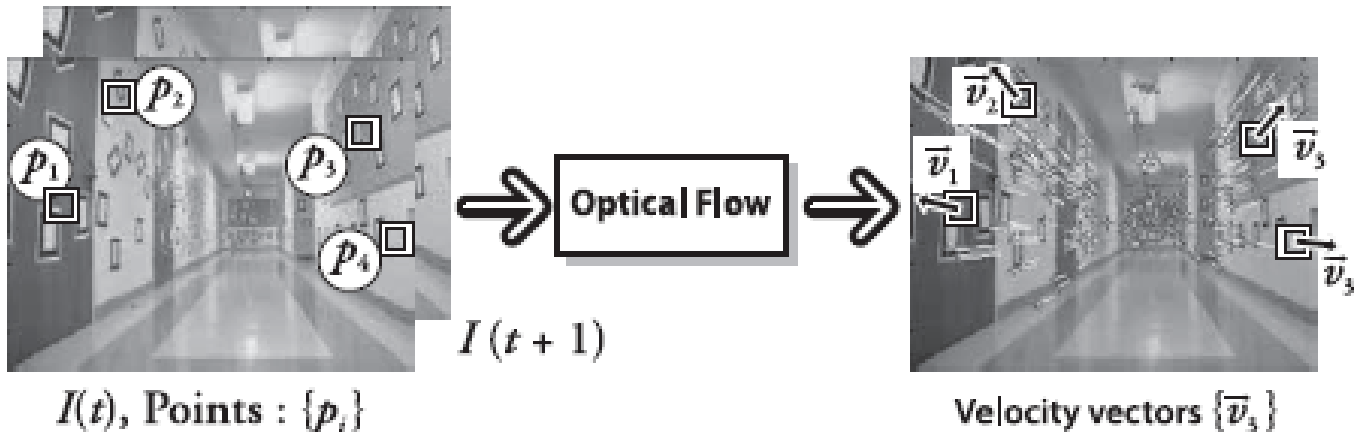
1974, [Kurzweil Computer Products](#) offered their first optical character recognition (OCR)



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

198x:

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



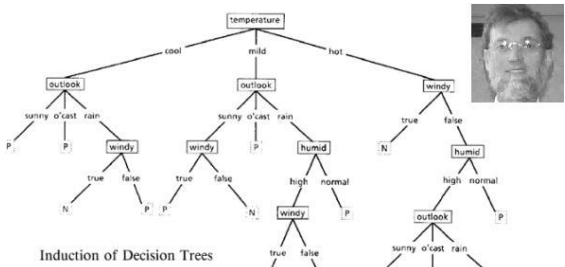
1986: classical [Canny edge detector](#)



1987: **EigenFace** algorithm



1985: Ross Quinlan's ID3 for decision trees analysis



1988: Dean Pomerleau's self-driving vehicle ALVINN



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

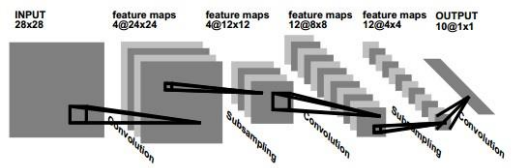
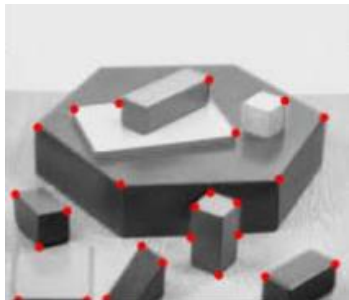


Figure 1: Architecture of LeNet 1

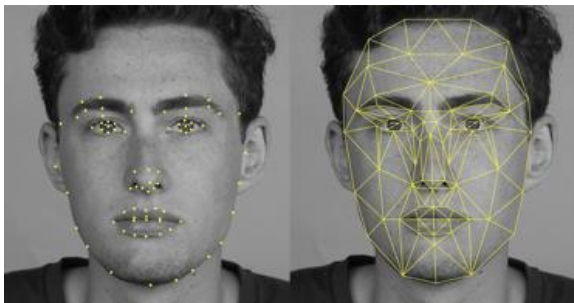


199x:

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



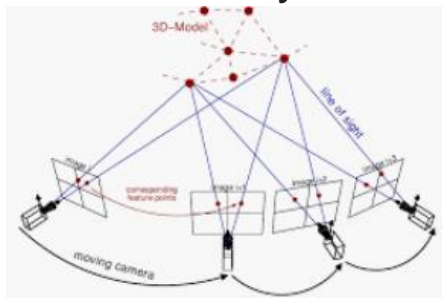
1995: Cootes et al.
Active Shape Modelling



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

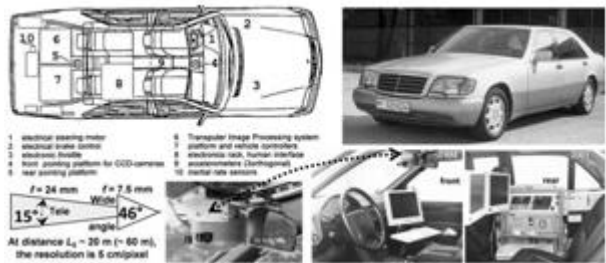


1999: Triggs et al.,
"Bundle Adjustment —
A Modern Synthesis"



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

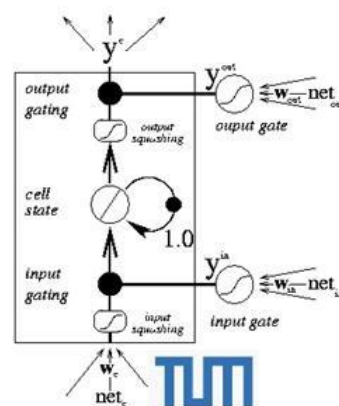
1994: Ernst Dickmanns' self-driving car drives more than 1,000
kms near the airport Charles-de-Gaulle in Paris



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



Long Short-Term Memory



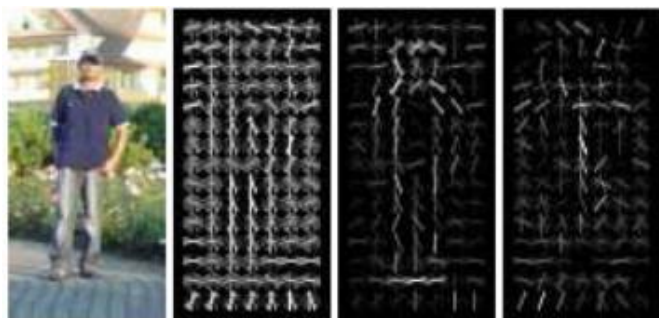
1997: Jeürgen Schmidhuber's
and Sepp Hochreiter's Long
Short Term Memory (LSTM)
model

200x:

2001: Viola/Jones face detector



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

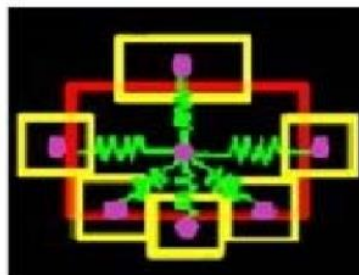
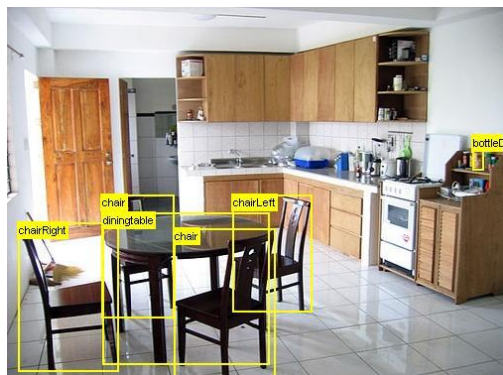


2005: driverless car Stanley wins
DARPA's Grand Challenge

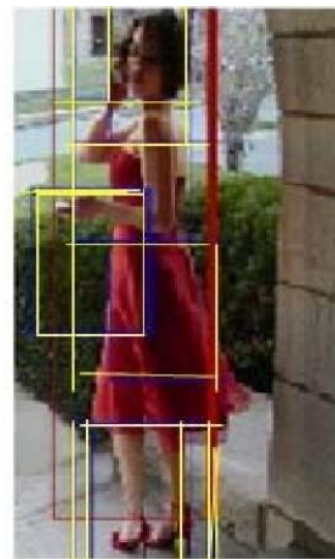


2006: Scott Hassan founds robot startup Willow Garage

2007, The PASCAL Visual Object Classes Challenge 2007



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

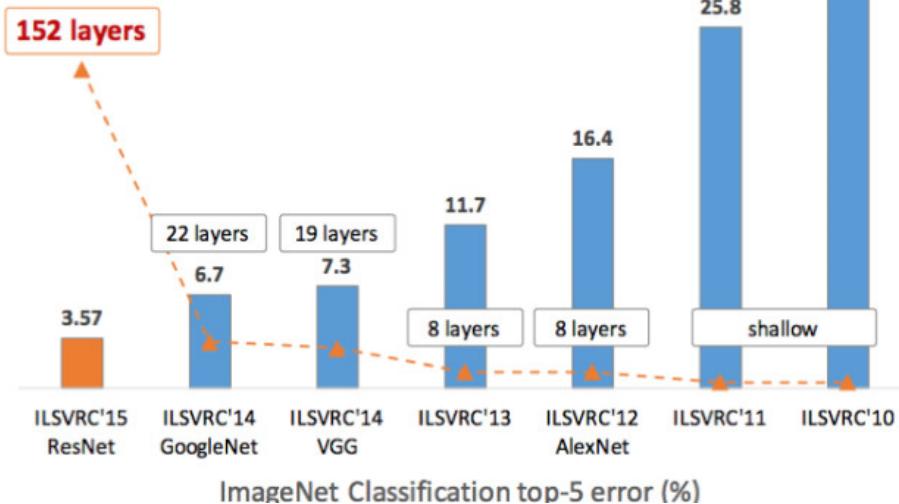


201x – Deep Learning Era

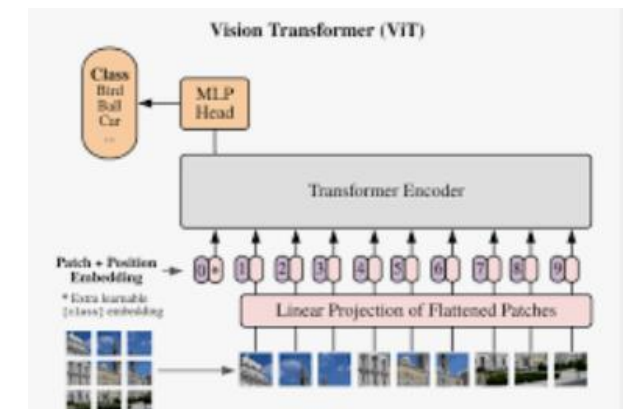
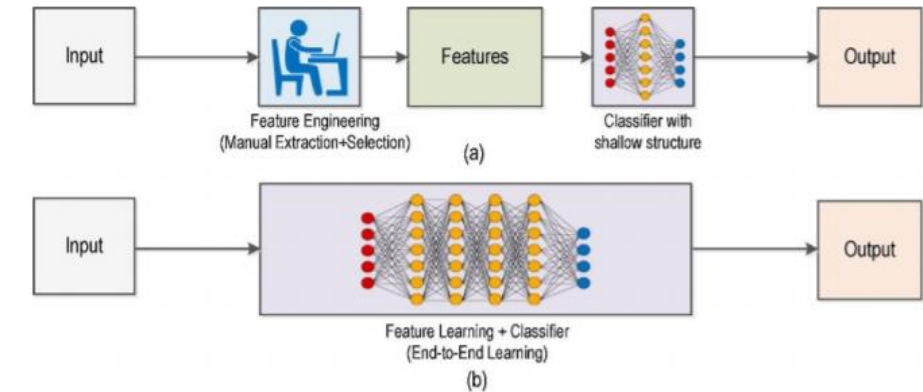
- 2010: The **ImageNet** Large Scale Visual Recognition Competition
- 2012 [AlexNet won ImageNet](#)



Revolution of Depth



- 2017, Ashish Vaswani et al, **Attention is all you need** -> **Visual Transformers**



Видеть невидимое

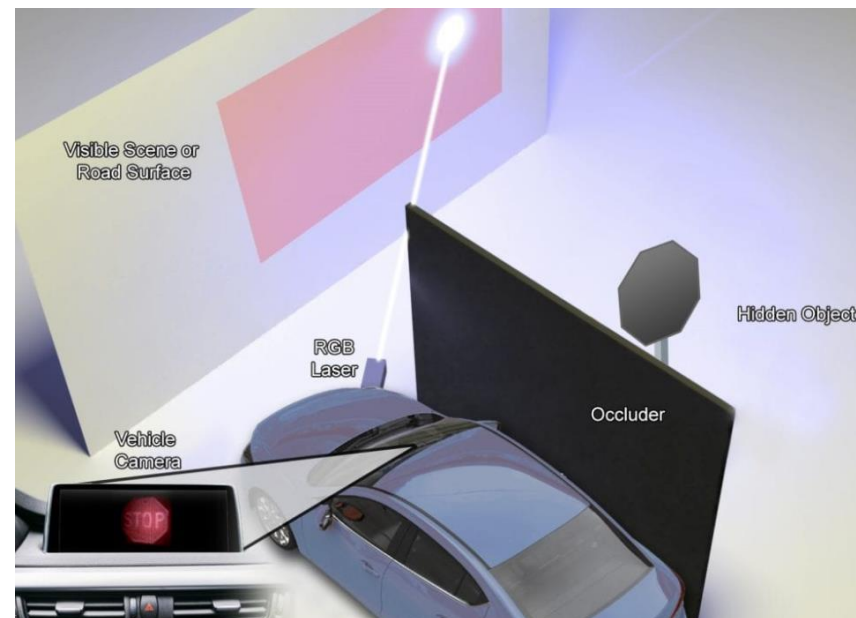
CVPR 2019 Best Paper Award:



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

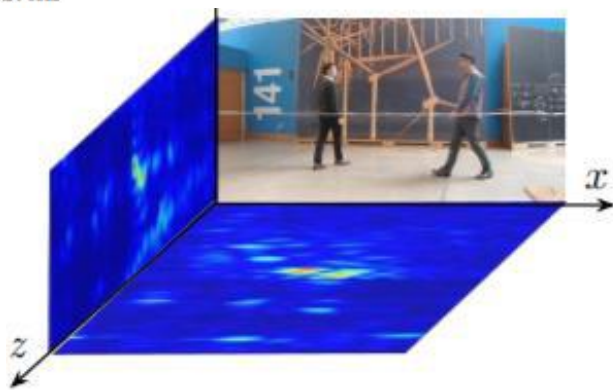
Shumian Xin¹, Sotiris Niousias^{2,3}, Kiriakos N. Kutulakos², Aswin C. Sankaranarayanan¹, Srinivasa G. Narasimhan¹, and Ioannis Gkioulekas¹

¹Carnegie Mellon University ²University of Toronto ³University College London



Through-Wall Human Pose Estimation Using Radio Signals

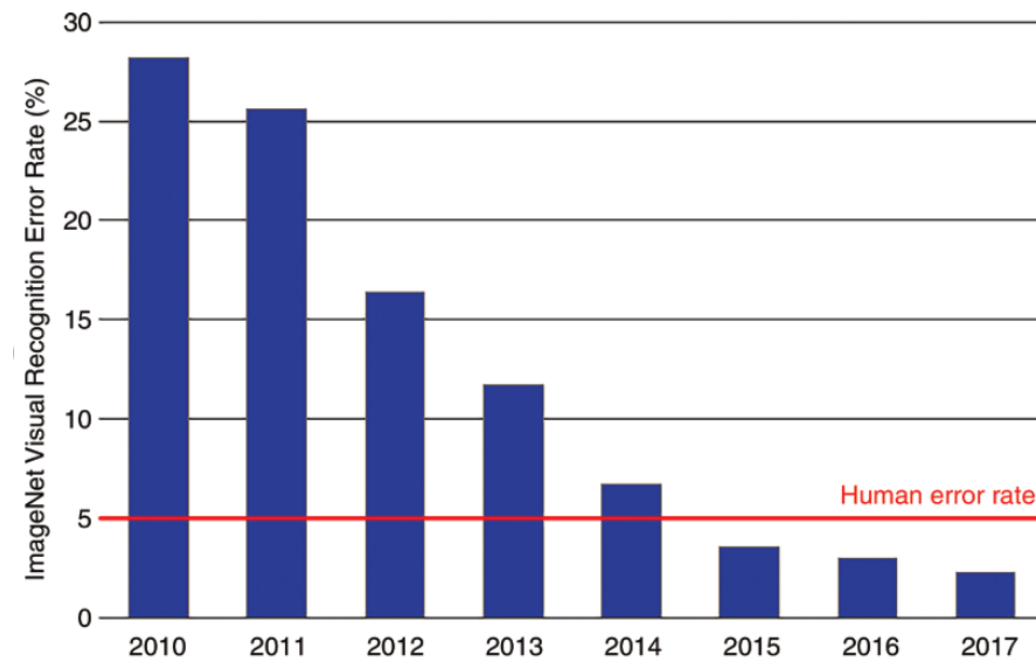
Mingmin Zhao Tianhong Li Mohammad Abu Alsheikh Yonglong Tian Hang Zhao
Antonio Torralba Dina Katabi
MIT CSAIL



Сегодня

Алгоритмы компьютерного зрения:

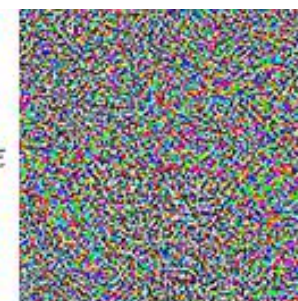
- Лучше человека классифицируют изображения и видео.
- Лучше человека распознают человека.
- Рисуют и подделывают (DeepFake).
- Видят невидимое.
- НО!



"panda"

57.7% confidence

+ ϵ



=



"gibbon"

99.3% confidence



Умный Ганс

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

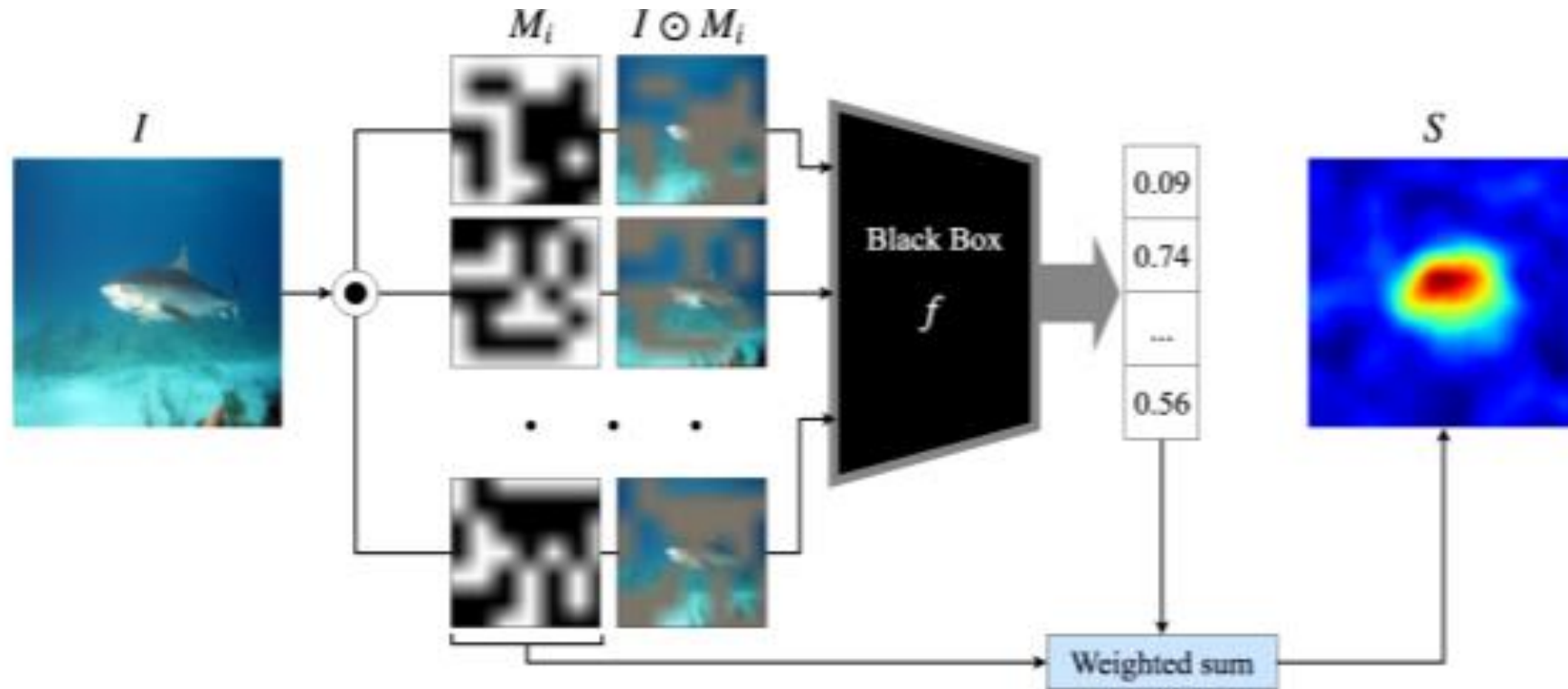
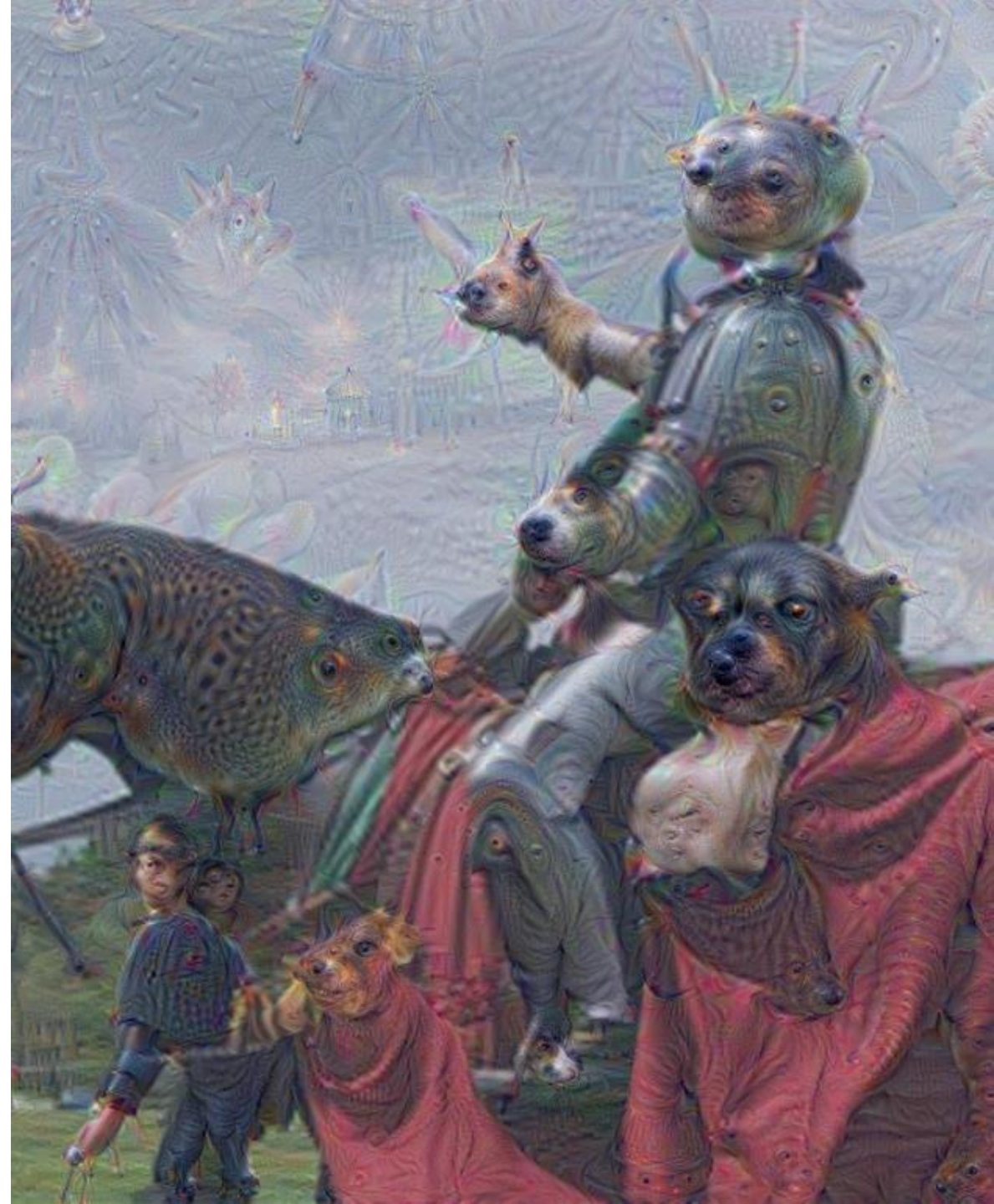


Figure 3: Overview of RISE: Input image I is element-wise multiplied with random masks M_i and the masked images are fed to the base model. The saliency map is a linear combination of the masks where the weights come from the score of the target class corresponding to the respective masked inputs.

Deep Dream

- Можно визуально оценить на что реагирует нейрон (или группа нейронов).
- Для этого не меняя веса сети, с помощью градиентной оптимизации меняем пиксели картинки, чтобы они максимизировали выход конкретного нейрона.



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”.