

Affective Computing

Olga Perepelkina
HSE, 2020

Тема 2. Теории эмоций
Тема 3. Эмоции и лицевые экспрессии

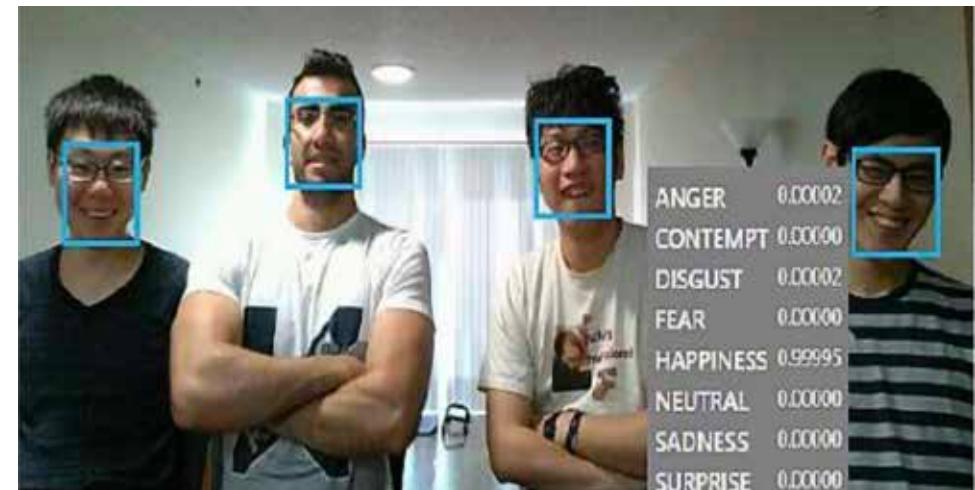
План лекции

- Идеи проектов и ресурсы
- Теории эмоций
- Искусственные нейросети
- Компьютерное зрение
- Практика: Microsoft API

Примеры проектов и идеи

Group Emotion-Recognition to Investigate Group vs. Individual Emotion

- Проверяли гипотезу о том, что эмоции человека усиливаются, если он их испытывает в группе по сравнению с тем, если он один.
- Показали похожие по интенсивности видео как для отдельных испытуемых, так и для групп, и проверили выражение их эмоций.



Reverse engineering affect recognition

- Создали большой синтетический датасет.
У виртуальных аватаров случайнym
образом генерировалась мимика, а с
помощью программ распознавания
выражения лиц делалась разметка.



Anger



Sad

Еще идеи

- Исследование (обзор) того, как Affective computing (например, распознавание эмоций) используется в индустрии.
- Этические вопросы вокруг Affective Computing:
 - Источники искажения (Potential for bias)
 - Источники манипуляции (Potential for manipulation)

Стажировка в NDL 2019

- Как алгоритмы распознавания эмоций справляются с видео, на которых люди чихают, кашляют и тд
- Ресерч: культурные различия в автоматическом распознавании эмоций (сравнение работы алгоритмов на датасетах представителей разных культур)
- Сбор данных для алгоритма распознавания фрода

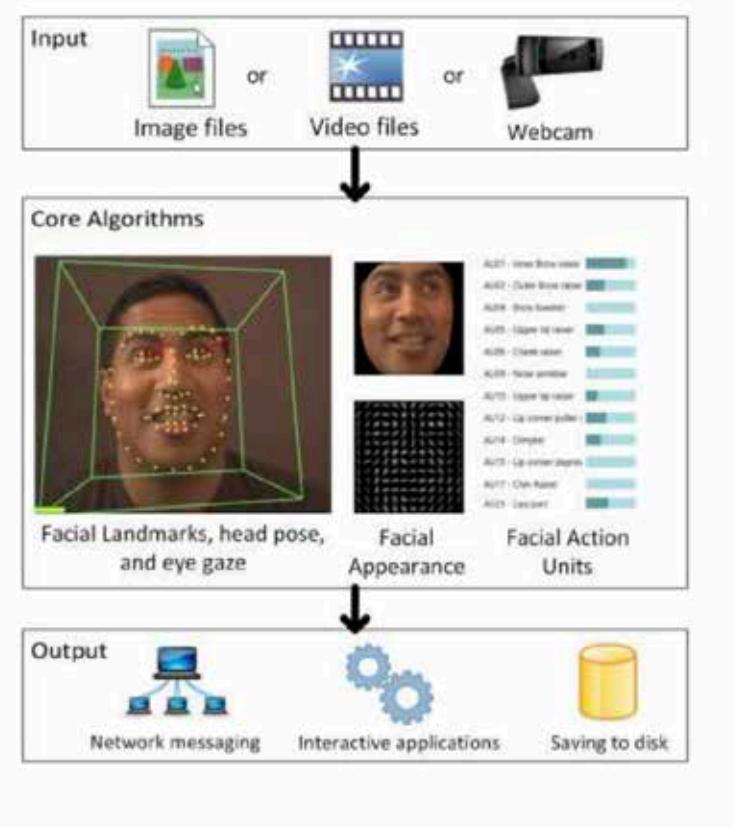
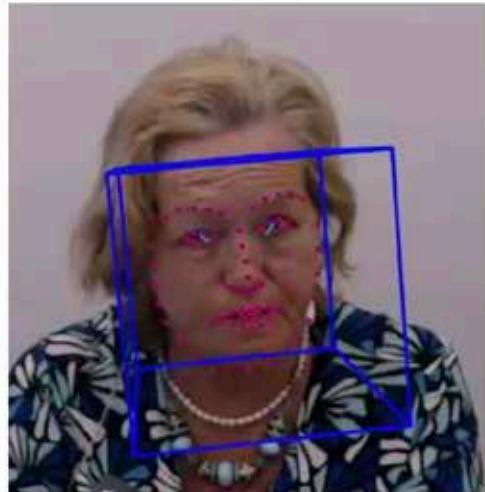
Ресурсы

Facial Expression Analysis

OpenFace

High-quality freely-available
facial action unit recognizer

- <https://github.com/TadasBaitrusaitis/OpenFace>



Voice Analysis

COVAREP

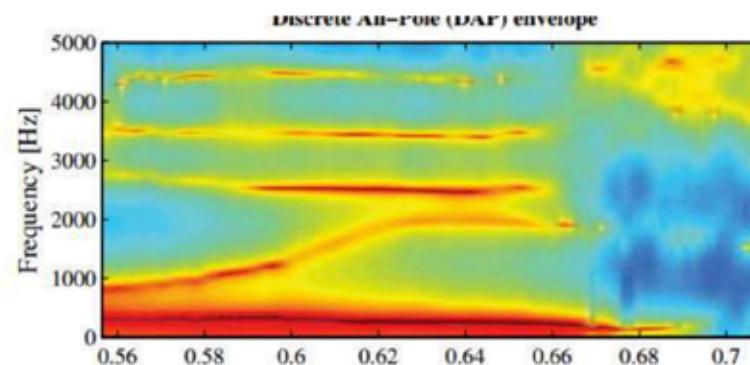
Freely-available voice analysis toolkit

- <http://covarep.github.io/covarep/>

OpenEAR

Another freely-available voice analysis tool

- <https://sourceforge.net/projects/openart/>



Text Analysis

LIWC

Calculates the degree to which people use different categories of words associated with emotion and other mental states

- <http://liwc.wpengine.com/>

Text Analysis, Crawling and Interpretation Tool

Freely-available tool for classifying social media text

- <http://tacit.usc.edu/index.html>

Sentiment analysis tools

Large collection of tools maintained by Rada Mihalcea

- <https://web.eecs.umich.edu/~mihalcea/downloads.html>

BERT, word2vec,

Databases: RAMAS



Multimodal emotion & behavior
data: video, audio, motion
capture, psychophysiology
(PPG, GSRI)



150 videos in the Russian
language



6 emotions: joy, anger,
sadness, disgust, fear, surprise



13 acted scenarios



5 male and 5 female actors,
aged 18-28



Each video file marked by at
least 5 annotators

Databases

<http://bit.ly/2RM4R90>

(можно попросить у меня)

Теории эмоций (доклад)

<http://bit.ly/2NVOdCv>

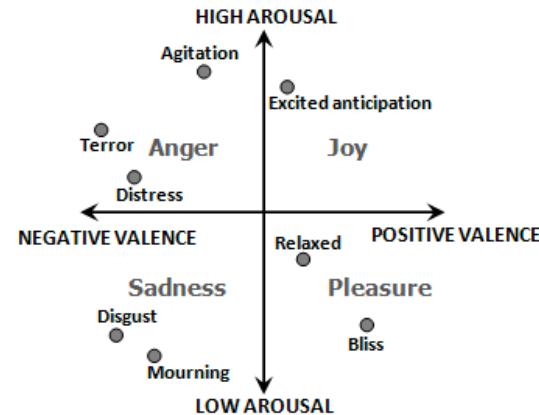
Теории эмоций (продолжение)

Теории эмоций

- Различные теории подчеркивают разные аспекты:
 - Теории оценки (appraisal) подчеркивают, что когнитивные процессы предшествуют эмоциям
 - Теории дискретных эмоций подчеркивают физиологические и экспрессивные следствия эмоций
- Исследователи Affective computing склонны опираться на различные теории в зависимости от аспектов, на которых они сосредоточены
- Например, методы распознавания эмоций часто опираются на дискретную теорию эмоций и избегают моделей оценки.

Дискретные/непрерывные

- Базовые/дискретные эмоции
- Эмоции как дискретные категории, биологически фиксированные, универсальные для всех людей и многих животных
- Ekman
- Непрерывные
 - Эмоции – это комбинации нескольких психологических измерений
- Russell, Feldman Barrett



Дискретные



Discrete



Disgust



Happiness



Fear



Sadness



Surprise

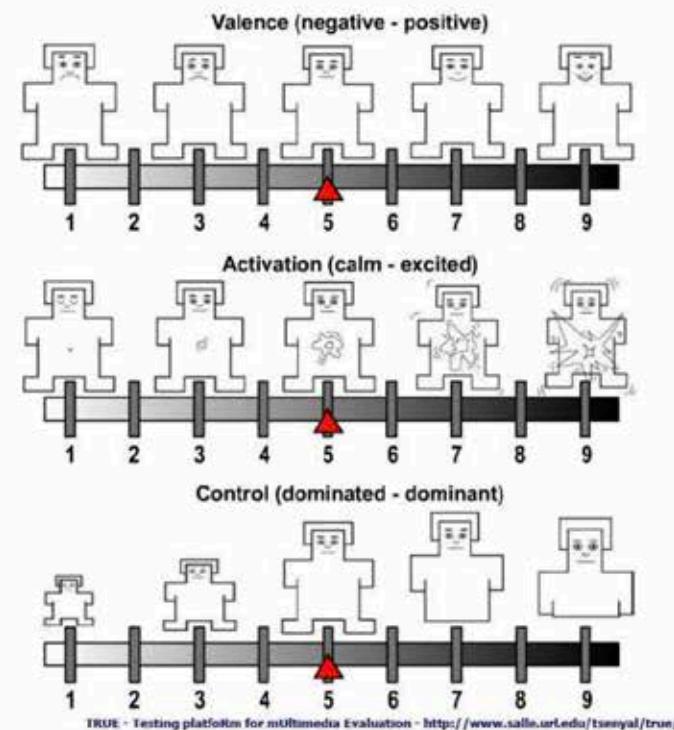


Anger

Непрерывные



Continuous

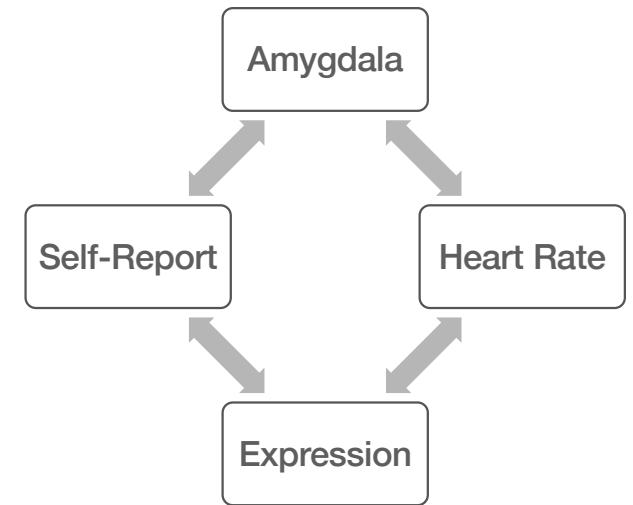


Эмоции как «атом»/«молекула»

- «Атом»
- Эмоциональные компоненты тесно связаны и могут рассматриваться как цепь, связывающая стимулы и реакцию
- Ekman
- «Молекула»
- Эмоции определяются свободной конфигурацией различных компонентов
- Russell, Feldman Barrett

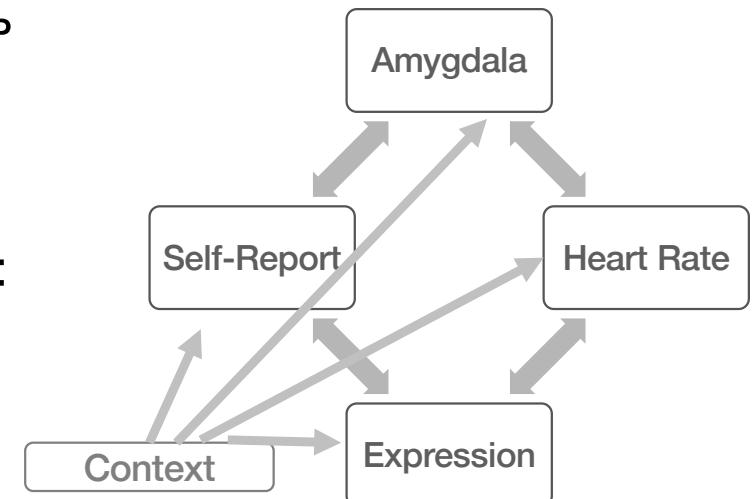
ЭМОЦИИ КАК «АТОМ»

- Если эмоция является «атомной цепью», все ее компоненты должны быть жестко соответствовать друг другу
- То есть экспрессия лица, физиологическая реакция и ощущаемые эмоции должны быть четко соответствовать друг с другом
- Эмоция может относиться ко всей «атомной цепи», но может быть измерена любым из компонентов
- Измеренные экспрессии должны предсказывать физиологию и ощущаемые эмоции.
- Мультимодальное распознавание должно работать наилучшим образом
- => **Если один из компонентов показывает эмоцию X, можем смело утверждать, что она есть**



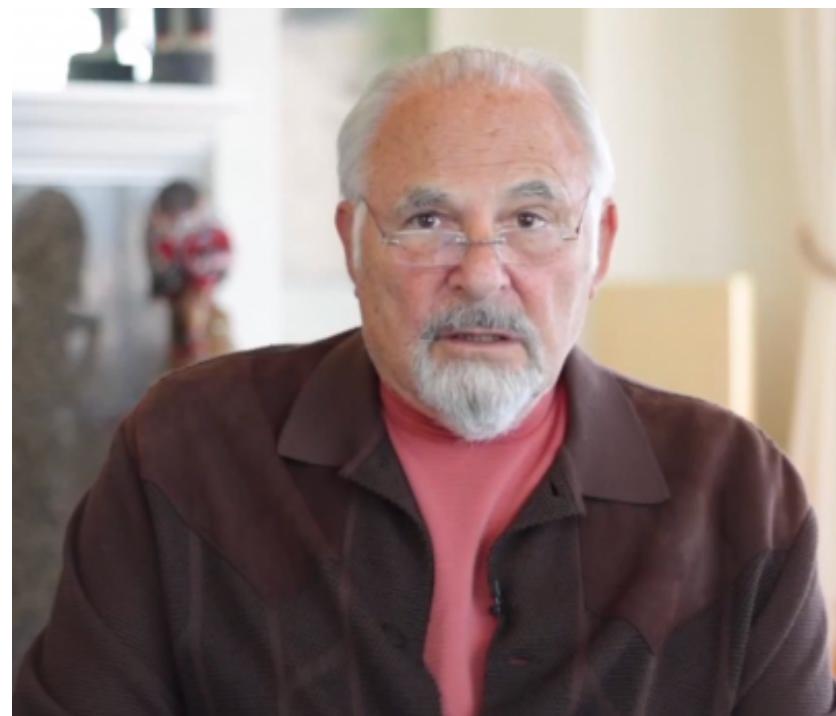
Эмоции как «молекула»

- Если эмоции это «молекулы», то ее компоненты жестко не соответствуют друг другу
- Компоненты влияют друг на друга, но могут быть не синхронизированы
- Лицевые экспрессии не должны точно отражать физиологию и ощущаемые эмоции
- Конструктивистские теории (Фельдман Барретт): эмоции - это ярлык, который мы присваиваем нашему ощущаемому физиологическому состоянию
- =>**Нельзя предполагать, что активация в одном компоненте отражает и лежит в основе эмоции X**



Теория базовых эмоций

- Лицевые экспрессии соотносятся с определенными эмоциями.
- Эти конфигурации являются общими для всех людей, независимо от их пола, расы, образования, культуры и т.д.
- Эта теория оказалась очень влиятельной на протяжении последних **50 лет**.





?



?



?

МГУ, наши дни...



- Эмоции – это «сложные системные психологические образования, включенные в различные виды психической деятельности и базирующиеся на разных потребностях».
- «В современной психологии принято выделять основные, или базальные, эмоции и высшие эмоции».
- «К числу базальных эмоций относятся: радость, горе, гнев, интерес, отвращение, презрение, удивление, стыд, вина» (согласно Изарду). «Другие авторы называют другое число базальных эмоций».
- «Общепризнанного "списка" базальных эмоций пока не существует».

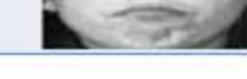
Издание 2013 года

State	Example Photo	Action Units	Physical Description	State	Example Photo	Action Units	Physical Description
Amusement		6 + 7 + 12 + 25 + 26 + 53	Head back, Duchenne smile, lips separated, jaw dropped	Fear		1 + 2 + 4 + 5 + 7 + 20 + 25	Eyebrows raised and pulled together, upper eyelid raised, lower eyelid tense, lips parted and stretched
Anger		4 + 5 + 17 + 23 + 24	Brows furrowed, eyes wide, lips tightened and pressed together	Happiness		6 + 7 + 12 + 25 + 26	Duchenne smile
Boredom		43 + 55	Eyelids drooping, head tilted (not scorable with FACS; slouched posture, head resting on hand)	Interest		1 + 2 + 12	Eyebrows raised, slight smile
Confusion		4 + 7 + 56	Brows furrowed, eyelids narrowed, head tilted	Pain		4 + 6 + 7 + 9 + 17 + 18 + 23 + 24	Eyes tightly closed, nose wrinkled, brows furrowed, lips tight, pressed together, and slightly puckerred
Contentment		12 + 43	Smile, eyelids drooping	Pride		53 + 64	Head up, eyes down
Coyness		6 + 7 + 12 + 25 + 26 + 52 + 54 + 61	Duchenne smile, lips separated, head turned and down, eyes turned opposite to head turn	Sadness		1 + 4 + 6 + 15 + 17	Brows knitted, eyes slightly tightened, lip corners depressed, lower lip raised
Desire		19 + 25 + 26 + 43	Tongue shown, lips parted, jaw dropped, eyelids drooping	Shame		54 + 64	Head down, eyes down
Disgust		7 + 9 + 19 + 25 + 26	Eyes narrowed, nose wrinkled, lips parted, jaw dropped, tongue shown	Surprise		1 + 2 + 5 + 25 + 26	Eyebrows raised, upper eyelid raised, lips parted, jaw dropped
Embarrassment		7 + 12 + 15 + 52 + 54 + 64	Eyelids narrowed, controlled smile, head turned and down (not scorable with FACS; hand touches face)	Sympathy		1 + 17 + 24 + 57	Inner eyebrow raised, lower lip raised, lips pressed together, head slightly forward

Keltner, D., Sauter, D., Tracy, J., and Cowen, A. (2019)

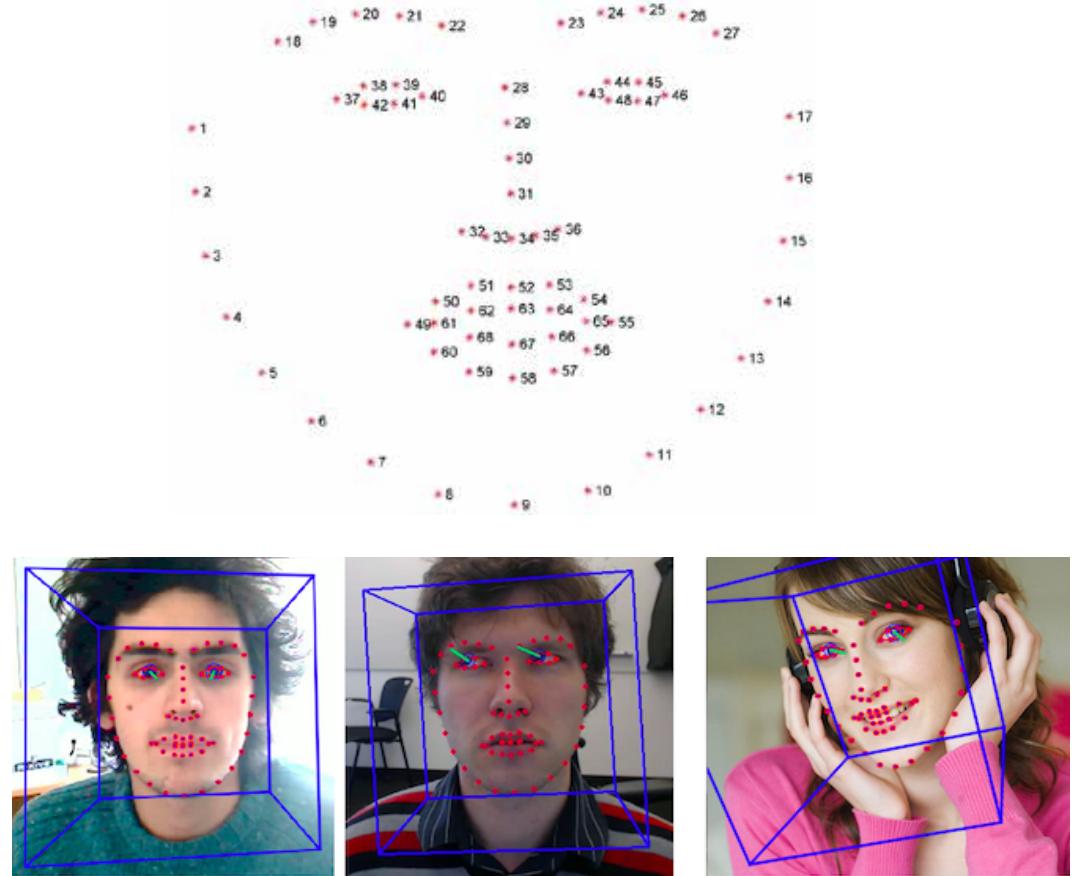
Action units

- Facial Action Coding System (FACS, Ekman et al., 2002).
- FACS описывает определенные лицевые движения (action units) одной или нескольких мышц.
- Система ничего не утверждает про эмоции, это описание поведения.

AU	Description	Facial Muscles (Type of Activation)		AU	Description	Facial Muscles (Type of Activation)	
1	Inner Brow Raiser	<i>Frontalis (pars medialis)</i>		18	Lip Puckerer	<i>Incisivillabii superioris and incisivillabii inferioris</i>	
2	Outer Brow Raiser	<i>Frontalis (pars lateralis)</i>		20	Lip Stretcher	<i>Risorius with platysma</i>	
4	Brow Lowerer	<i>Corrugator supercilii, depressor supercilii</i>		22	Lip Funneler	<i>Orbicularis oris</i>	
5	Upper-Lid Raiser	<i>Levator palpebrae superioris</i>		23	Lip Tightener	<i>Orbicularis oris</i>	
6	Cheek Raiser	<i>Orbicularis oculi (pars orbitalis)</i>		24	Lip Pressor	<i>Orbicularis oris</i>	
7	Lid Tightener	<i>Orbicularis oculi (pars palpebralis)</i>		25	Lips Part	<i>Depressor labii inferioris or relaxatio of mentalis, or orbicularis oris</i>	
9	Nose Wrinkle	<i>Levatorlabii superioris alaqueanasi</i>		26	Jaw Drop	<i>Masseter, relaxed temporalis and internal pterygoid</i>	
10	Upper-Lip Raiser	<i>Levatorlabii superioris</i>		27	Mouth Stretch	<i>Pterygoids, digastric</i>	
11	Nasolabial Deepener	<i>Zygomaticus minor</i>		28	Lip Suck	<i>Orbicularis oris</i>	
12	Lip-Corner Puller	<i>Zygomaticus major</i>		41	Lid Droop		
13	Cheeks Puffer	<i>Levatoranguli oris</i>		42	Slit		
14	Dimpler	<i>Buccinator</i>		43	Eyes Closed		
15	Lip-Corner depressor	<i>Depressor anguli oris</i>		44	Squint		
16	Lower-Lip depressor	<i>Depressor labii inferioris</i>		45	Blink		
17	Chin Raiser	<i>Mentalis</i>		46	Wink		

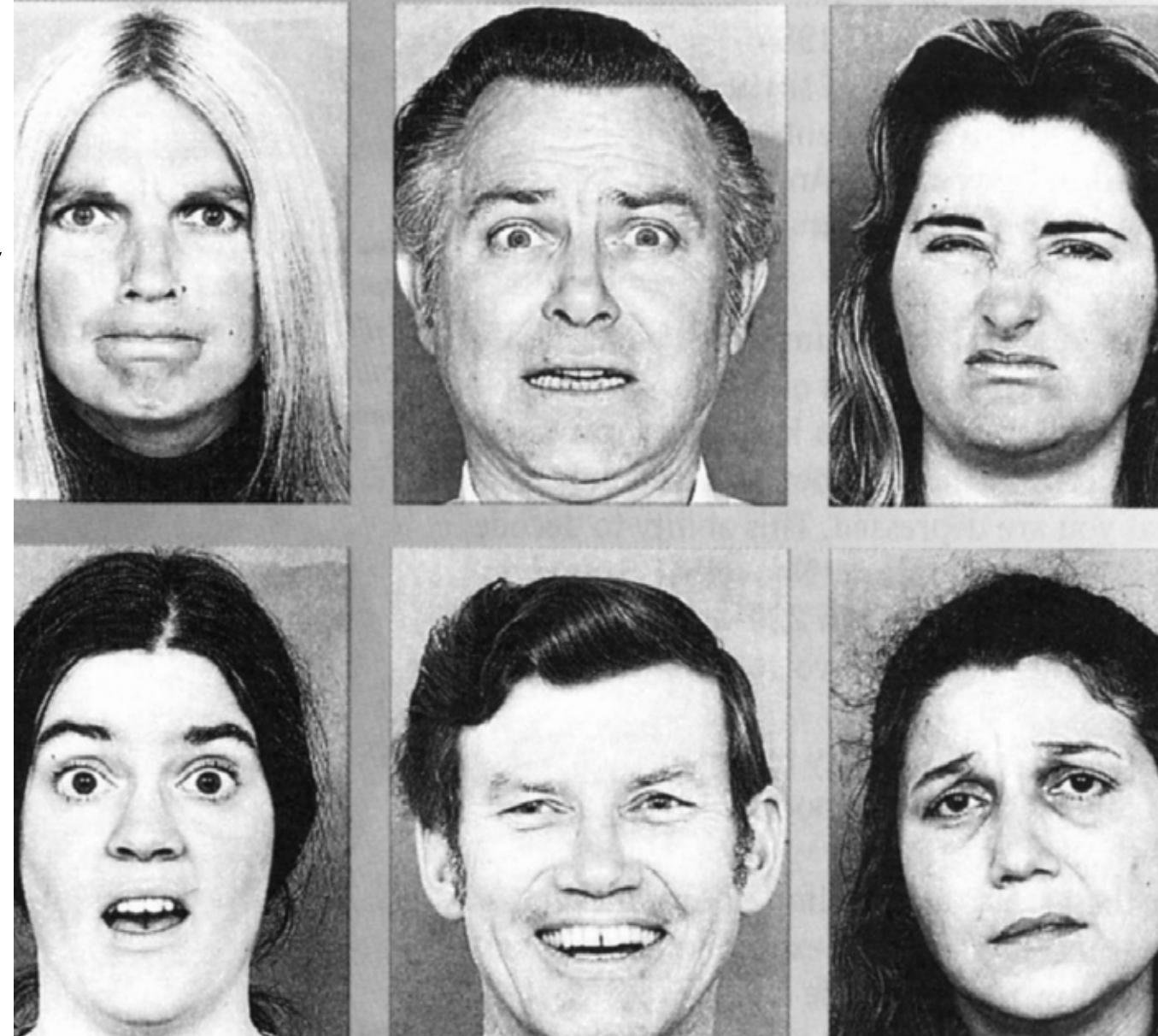
Action units – automatic detection

- Автоматические системы распознавания FACS достаточно точны (>90% acc).
- Особенno точны, если тестируются на тех же данных (~99%), менее точны – если на других (~90%).
- OpenCV dlib (facial landmarks).
- OpenFace



«Читаем» ЭМОЦИИ ПО ЛИЦУ

- "Эмоции
универсальны"
- "Эмоции
выражаются на лице
однозначно"



Microsoft Affectiva FaceReader

...

Lore ipsum dolor sit amet consectetur

Microsoft Project Oxford

Emotion API

Subscription Key Management

Select a scenario:

- Detect emotion using a stream
- Detect emotion using a URL**
- Detect emotion in a Video

Detect emotion by a specified image URL

Please type or paste an URL to an image below, and click [Detect]

Detect

The top 3 emotions will be shown in the list box. For more details on other emotions, please see the status.

Detection Done

Happiness:1.000000 Surprise:0.000000 Anger:0.000000
Happiness:1.000000 Surprise:0.000000 Anger:0.000000
Surprise:0.786557 Neutral:0.186621 Sadness:0.015479

```
[10:13:46.362032]: Emotion[2]
[10:13:46.366081]: .FaceRectangle = left: 219, top: 185, width: 66, height: 66
[10:13:46.366081]: Anger : 1.863633E-05
[10:13:46.366081]: Contempt : 4.439726E-05
[10:13:46.366081]: Disgust : 0.0002682077
[10:13:46.381711]: Fear : 0.01100761
[10:13:46.381711]: Happiness: 4.157415E-06
[10:13:46.381711]: Neutral : 0.1866214
[10:13:46.397337]: Sadness : 0.01547852
```

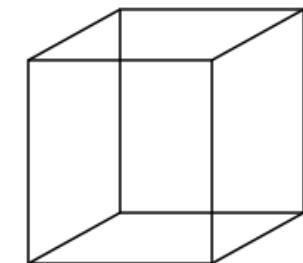
Microsoft will receive the images you upload and may use them to improve Face API and related services. By submitting an image, you confirm you have consent from everyone in it.

Mixed emotions

- Можем ли мы ощущать амбивалентные эмоции одновременно?
- **Bipolar account:** положительные и отрицательные эмоции находятся на противоположных полюсах одного континуума и не могут восприниматься одновременно.
- **Bivariate account:** положительные и отрицательные эмоции исходят из независимых биологических процессов, и могут возникать одновременно. Однако, их возникновение нестабильное и кратковременное.
 - Russell J. A. Mixed emotions viewed from the psychological constructionist perspective //Emotion Review. – 2017. – Т. 9. – №. 2. – С. 111-117.

Mixed emotions

- **Vacillation hypothesis** (гипотеза колебания): может восприниматься только одно состояние в один момент времени.
- Как с восприятием куба Неккера, ощущение одновременности двух противоположных состояний – может быть просто неверным истолкованием быстрой смены этих состояний.
- Экспериментально изучали эту гипотезу: нажимали на правую кнопку – когда счастье, на левую – когда грусть, на обе – одновременно при просмотре *bittersweet* фильмов. Чаще всего нажимают одновременно (то есть похоже на опровержение гипотезы).
- Однако переключения могут быть очень быстрыми
 - Larsen J. T. Holes in the case for mixed emotions //Emotion Review. – 2017. – Т. 9. – №. 2. – С. 118-123.



Mixed vs Blended emotions

- Mixed:
 - одновременное восприятие противоположных по валентности, эмоциональная амбивалентность.
 - Возникают на короткое время
 - “Bittersweet” ситуации.
- Blended: смешанные базовые эмоции (напр., joy + fear, amusement + disgust и т.д.)

Human facial expressions

- Лицевые выражения – инструмент коммуникации
- Сигнализируют об эмоциональном состоянии
- Сообщают о намерениях (intentions) и позициях (attitudes)
 - Ekman P., Friesen W. V. Unmasking the face: A guide to recognizing emotions from facial cues. – 1975.

Human facial expressions

- Люди модифицируют свои спонтанные лицевые выражения в зависимости от межличностных взаимодействий
 - (e.g. Buck et al., 1992; Ekman and Friesen, 1969; France and Hecht, 2005; Manstead et al., 2005; Wagner and Lee, 2005; Wagner and Smith, 1991)
- Для этого они используют различные стратегии: *fake* (ненастоящие), *inhibited* (скрытые, подавленные) и *masked* (замаскированные) expressions.
- Способность контролировать эмоциональные экспрессии (т.е. подавлять, изменять и симулировать выражения эмоций) – это один из навыков, относящихся к эмоциональному интеллекту.
 - Niewiadomski R. Ł., Pelachaud C. Affect expression in ECAs: Application to politeness displays //International journal of human-computer studies. – 2010. – Т. 68. – №. 11. – С. 851-871.

Human facial expressions

- Simple displays: спонтанные выражения эмоций, которые можно описать одним словом-лейблом (напр., гнев или презрение).
- Complex displays: комбинации нескольких простых выражений (например, суперпозиция двух выражений) или сознательно модифицированное выражение (подавление или маскирование).
 - Masked: anger masked by happiness
 - Superposed: anger and fear одновременно
 - Fake: simulated surprise
 - Inhibited: suppressed disappointment

Complex facial expressions

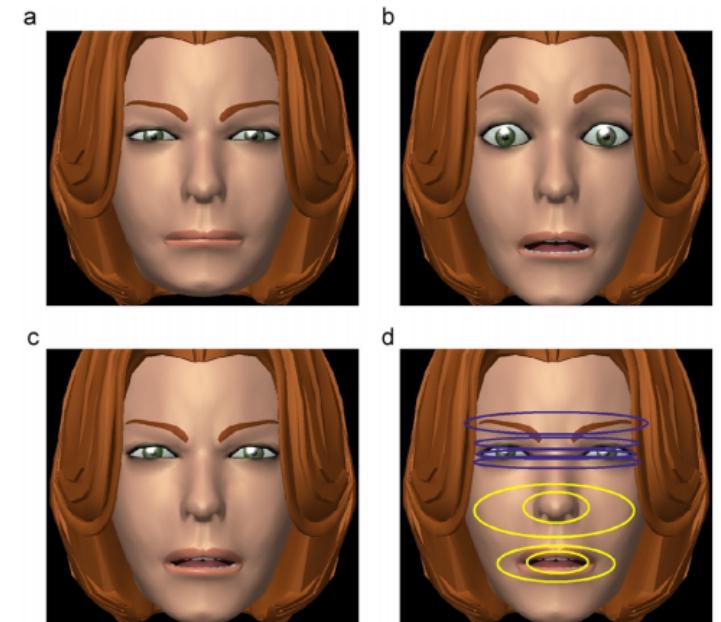
- Люди способны различать выражения реальной (felt) эмоции и ненастоящей (fake) эмоции.
 - Ekman, P., Friesen, W., 1969. The repertoire of nonverbal behavior: categories, origins usage and coding. *Semiotica* 1, 49–98.
- Люди не способны сознательно контролировать все свои лицевые мышцы.
- Выражения ощущаемых эмоций связаны со специфическими лицевыми признаками (напр., поднятие бровей при грусти). Эти признаки отсутствуют в ненастоящих эмоциях, т.к. их сложно сделать сознательно.
- **Inhibition hypothesis:** те же элементы лицевой экспрессии, которые сложно показать в случае отсутствия эмоции, также сложно подавить в случае ощущаемой эмоции.

Complex facial expressions

- Комплексные выражения лиц получаются из **композиции выражений различных областей лица**.
- Экман определил 18 различных экспрессий суперпозиций попарных 6 базовых эмоций.
- **Не каждая комбинация** эмоций верхней и нижней частей лица **правдоподобна** (например, грусть + счастье: грусть проявляется в верхней части лица, счастье – в нижней, наоборот – неправдоподобно).

Superposition

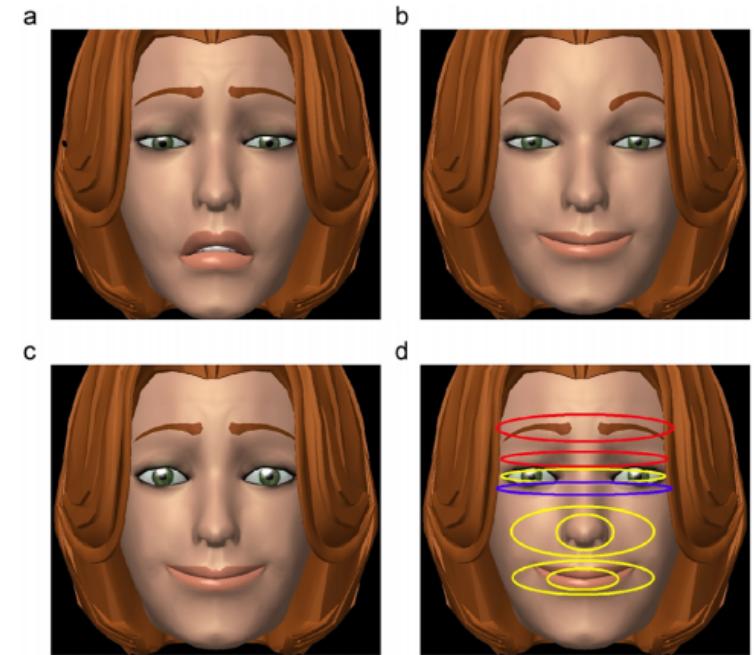
- Суперпозиция возникает, когда две эмоции ощущаются в одно и то же время.
- Результирующее лицевое выражение – выражение обеих эмоций.
- Суперпозиция обычно проявляется как комбинация нижней части лица одной эмоции и верхней части – другой эмоции.



Anger + Fear

Masking

- Маскировка возникает, когда ощущаемая эмоция не должна быть продемонстрирована по каким-либо причинам;
- Вместо нее демонстрируется другое эмоциональное выражение.
- Ощущаемая эмоция «протекает» через маскированную согласно inhibition hypothesis.
- Верхние части лица сложнее контролировать.

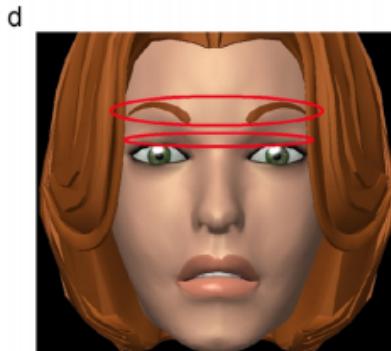


Disappointment masked by happiness

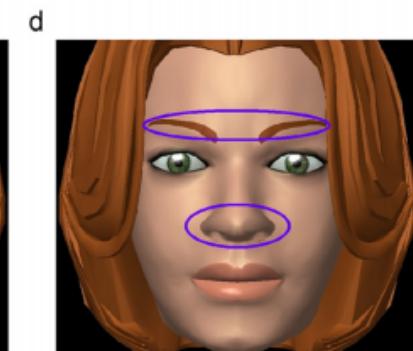
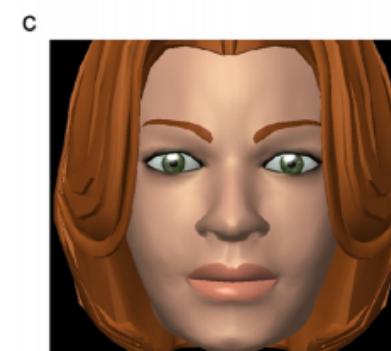
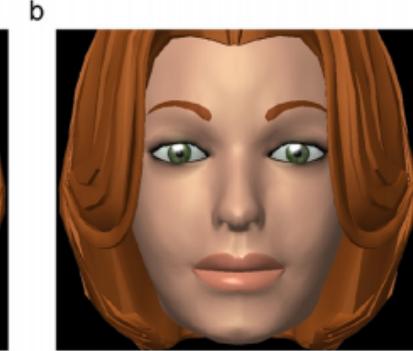
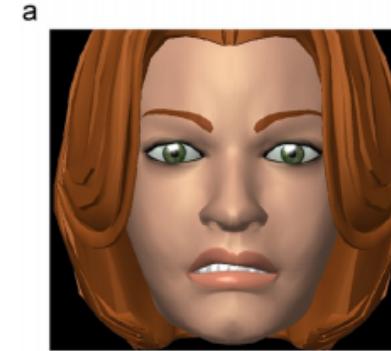
Fake and inhibited expressions

- Подавление возникает, когда человек пытается избежать демонстрации своих эмоций.
- В этом случае он пытается не показывать никаких эмоций.
- Подавленная эмоция просачивается сквозь нейтральное выражение.
- Ненатуральная эмоция возникает, когда человек пытается выразить эмоцию, которую он не испытывает в данный момент.

Fake and inhibited expressions

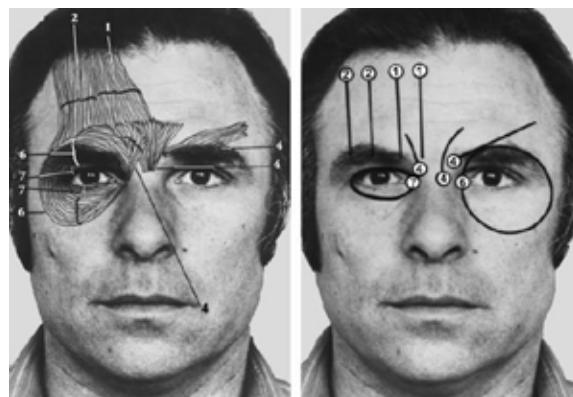


Fake distress
(страдание)



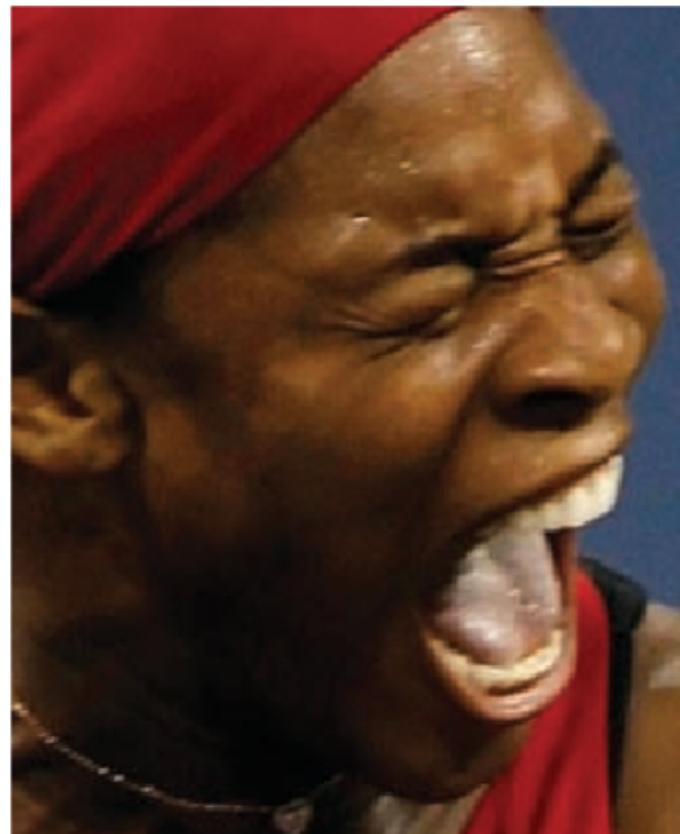
**Inhibited
contempt**

“Emotion AI”





**This woman
is in pain**



Or not?

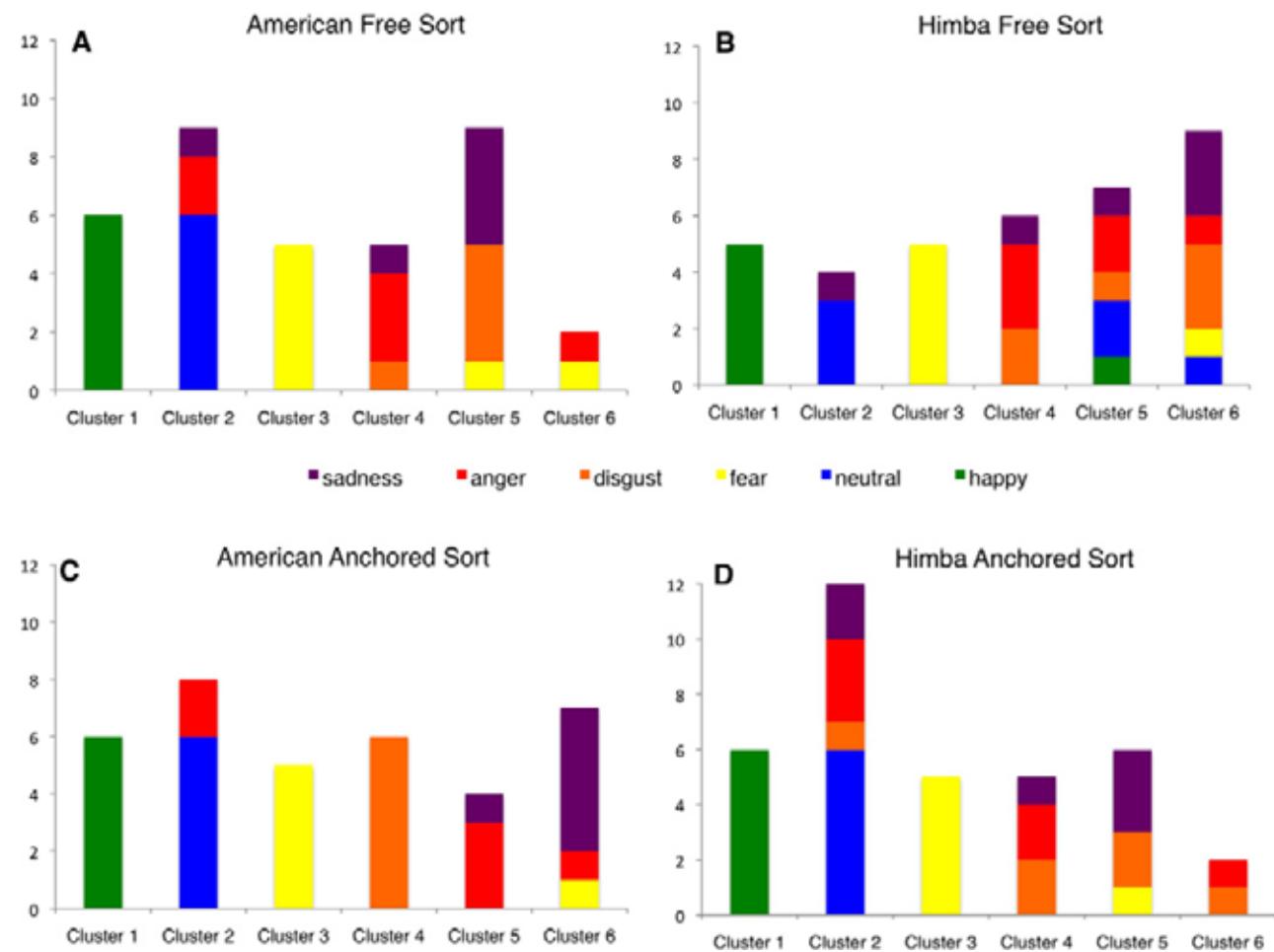
**Serena Williams has
won the US Open
championship**



Universality of emotions?

- Методические ограничения предыдущих экспериментов
- Если попросить использовать любые слова, универсальность исчезает!

[Gendron M. et al., 2014;
Barrett L. et al., 2011]



Surprise?







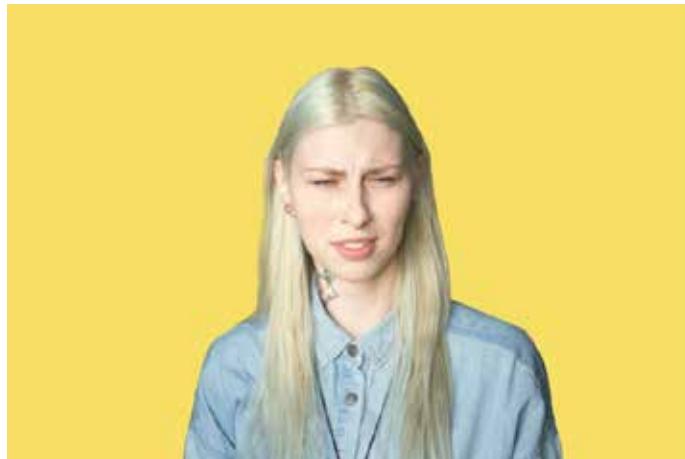
**Disgust?
Contempt?**



**Happiness?
Coquetry?**



**Anger?
Despair?**





«страх»



«отвращение»



«злость»

"many-to-many mappings"

- Люди не двигают лицами рандомно (конечно же), но лицевые конфигурации — это не «отпечатки» эмоций, которые не зависят от **контекста, личности, культуры**.
- Пример: нахмуренность может быть экспрессией злости, но это не значит, что нахмуренность — это выражение для злости. Злость может быть выражена и другими способами, а нахмуренность может значить что-то еще (**many-to-many mappings**).

Критика от Лизы Барретт

- Эмоциональные экспрессии более вариативны, чем предполагалось, системы не были разработаны, чтобы учитывать эту **вариативность и контекст**.

Контекст

- Культурный контекст
- Специфическая ситуация
- Индивидуальный опыт человека (person's learning history)
- Сиюминутное физиологическое состояние
- Временной контекст — то, что произошло секунду назад

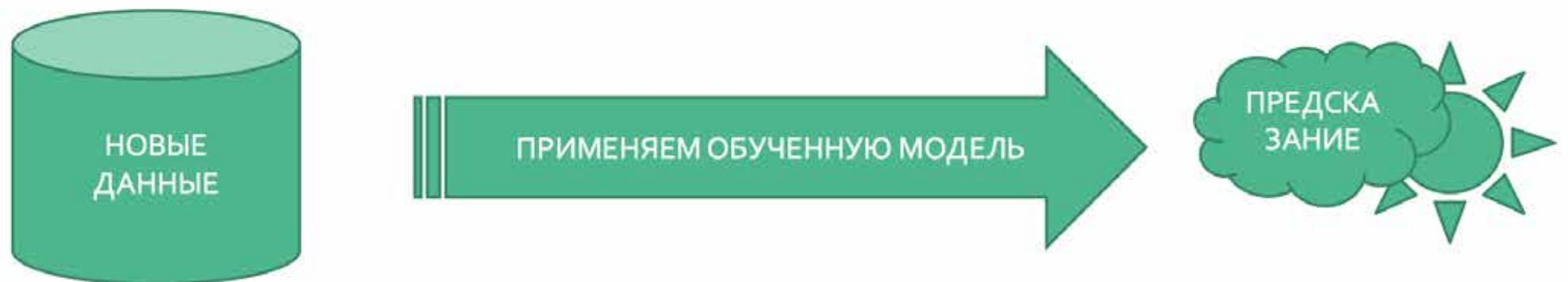
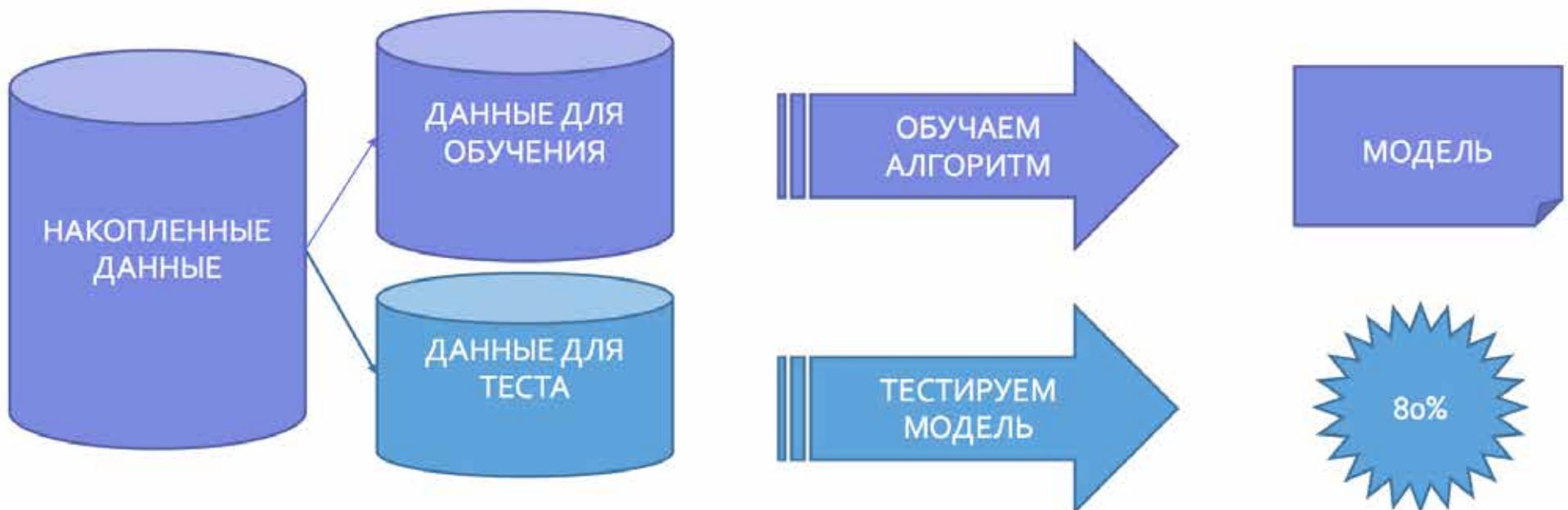
Критика от Лизы Барретт

- Технологичные компании тратят миллионы долларов, чтобы построить системы и устройства, которые будут «читать эмоции по лицам».
- Однако Лиза предлагает называть эти технологии не детекторами эмоциональных экспрессий, а **детекторами движений лица**.

Предложения от Лизы Барретт

- В идеальном случае должно быть записано много видео человека в естественных условиях. И алгоритмы должны автоматически вытащить проявления, такие как движения лица, походку, тон голоса. + записанные движения глаз, ЭЭГ и др.
- Тогда научимся классифицировать эмоции с высокой надежностью (*reliability*) и чувствительностью (*specificity*).
- Добавить еще туда физического контекста (где человек находится, есть ли другой человек рядом, как он реагирует и тд). Это она называет **Multimodal observations**.
- Люди должны разметить видео (*self-reports*). Предлагает категории: **valence**, **arousal**, **appraisals** (описания, как воспринимается ситуация) и **emotion-related goals**.
- Должна быть вариативность в ситуациях и социальных проявлениях.

Искусственные нейросети



Искусственные нейронные сети

- Искусственные нейросети – это **одновременно** наиболее **старая** (1950-е) и наиболее **новая** (1990-е) область искусственного интеллекта.

Искусственные нейронные сети

- Зачем нужны нейросети? Почему не использовать простые алгоритмы?
- Простые алгоритмы хорошо работают только в случае, если мы очень хорошо выбрали (извлекли) признаки из данных.
- Нейросети – мощный инструмент, умеют находить сложные (нелинейные) закономерности в данных.

Признаки – что это такое?

- Это всевозможные характеристики объектов выборки, которые мы можем **извлечь** из данных
- Есть отдельное направление – извлечение хороших признаков из данных (feature engineering)

Признаки

- Например, у нас есть выборка движений руки.
- Хотим классифицировать жесты глухонемых.
- Датасет:

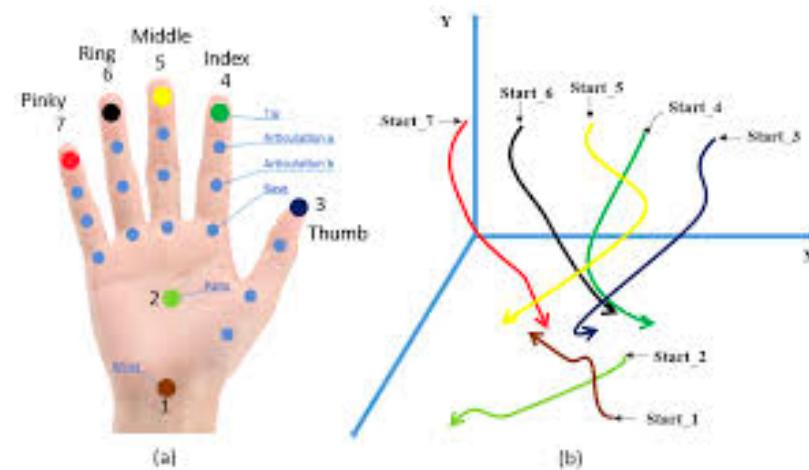
Движения каждого пальца в каждый момент времени (признаки);

Жест (лейбл).



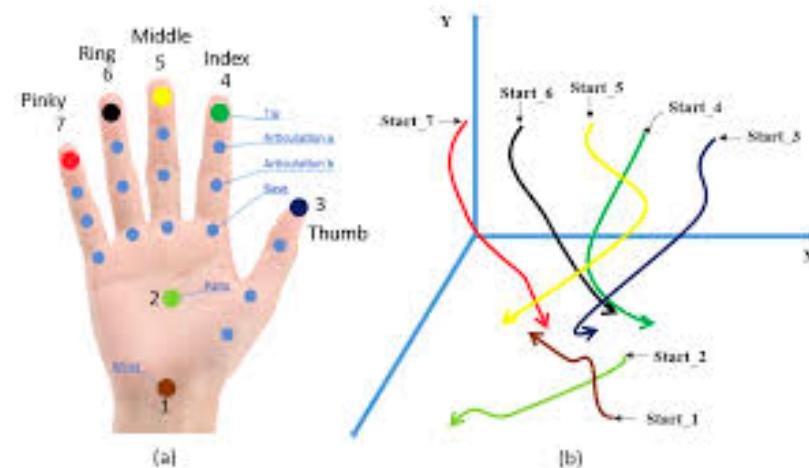
Признаки

- Какие можем извлечь признаки?

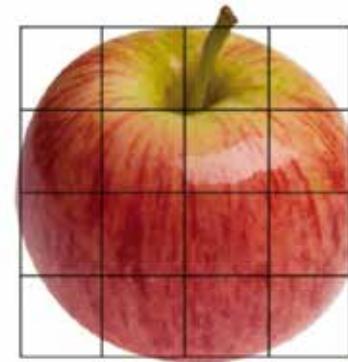
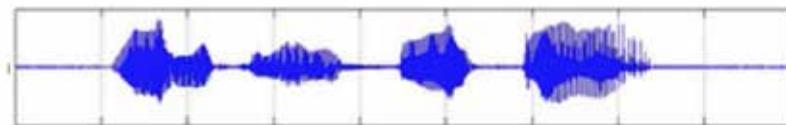


Признаки

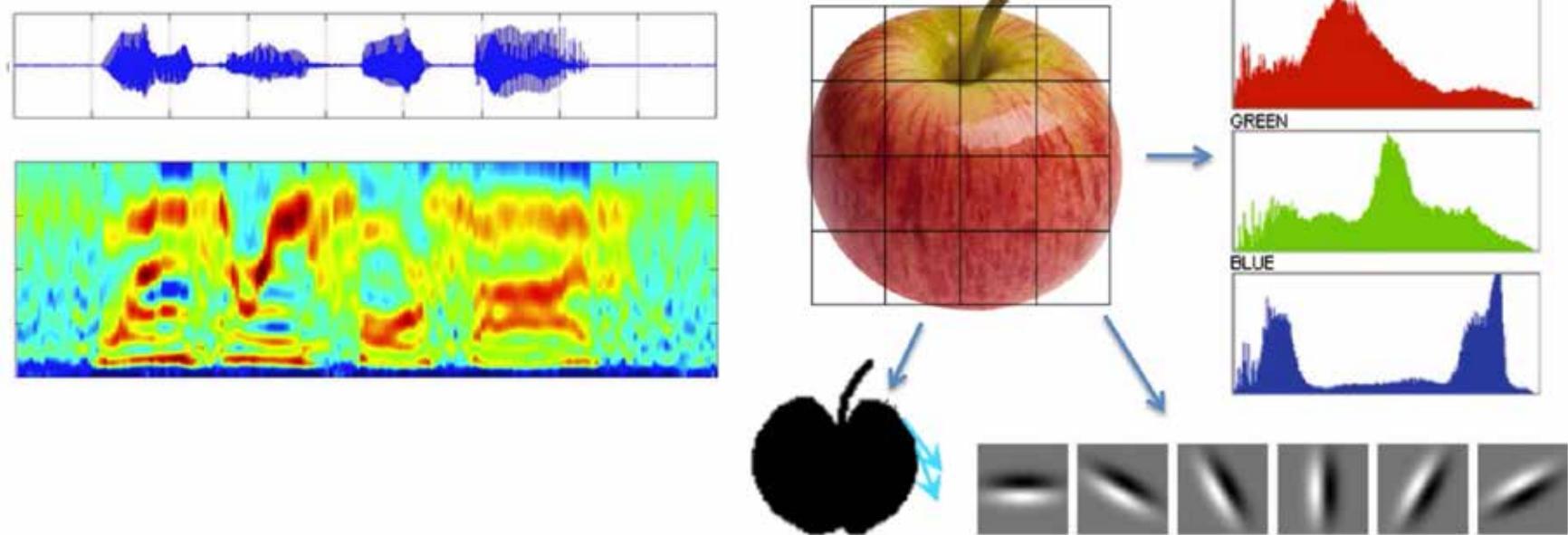
- Какие можем извлечь признаки?
- Скорость движения пальцев
- Расстояние, пройденное пальцами
- Ускорение
- Гладкость движений
- Траектории движений
- Повороты и вращения
- И т.д.



Признаки

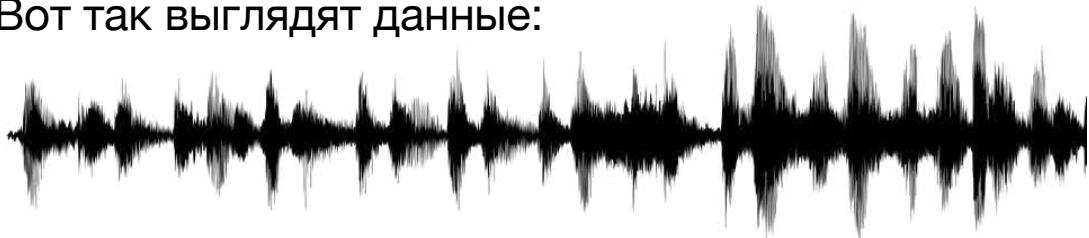


Признаки

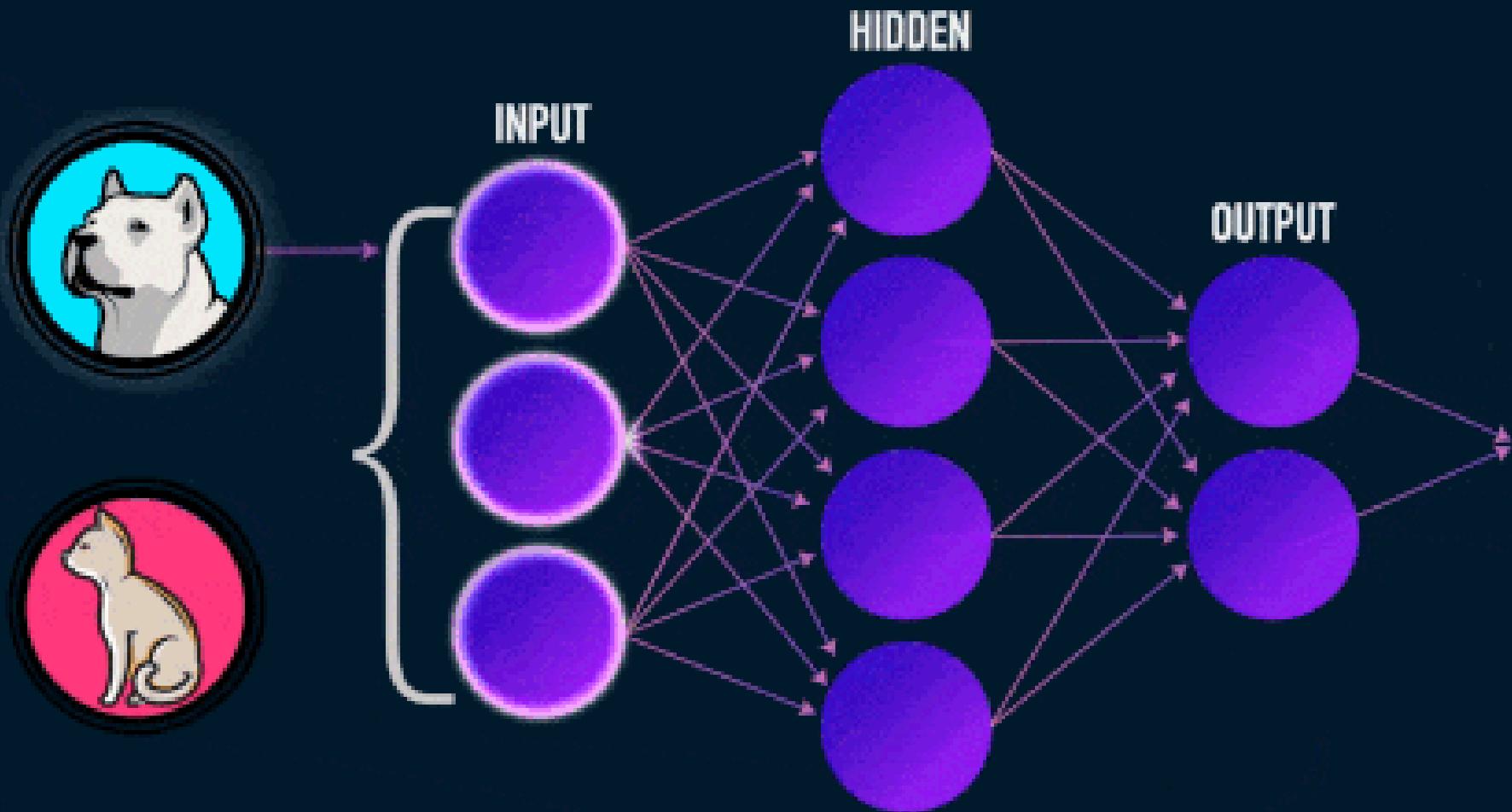


Нейросети

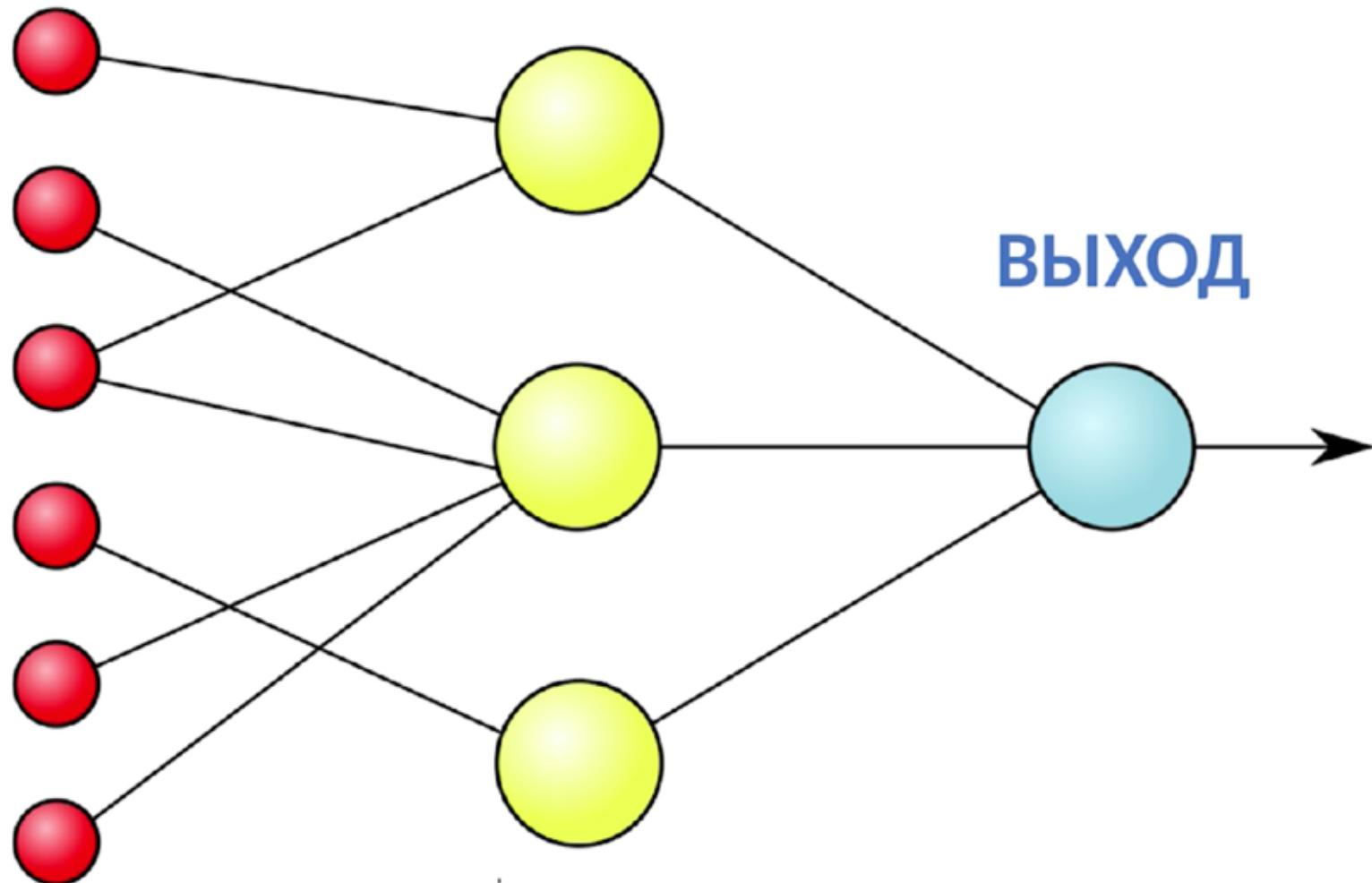
- Пример: у нас есть звук (хотим по звуку классифицировать жанры музыки – рок, поп, классика).
- Вот так выглядят данные:



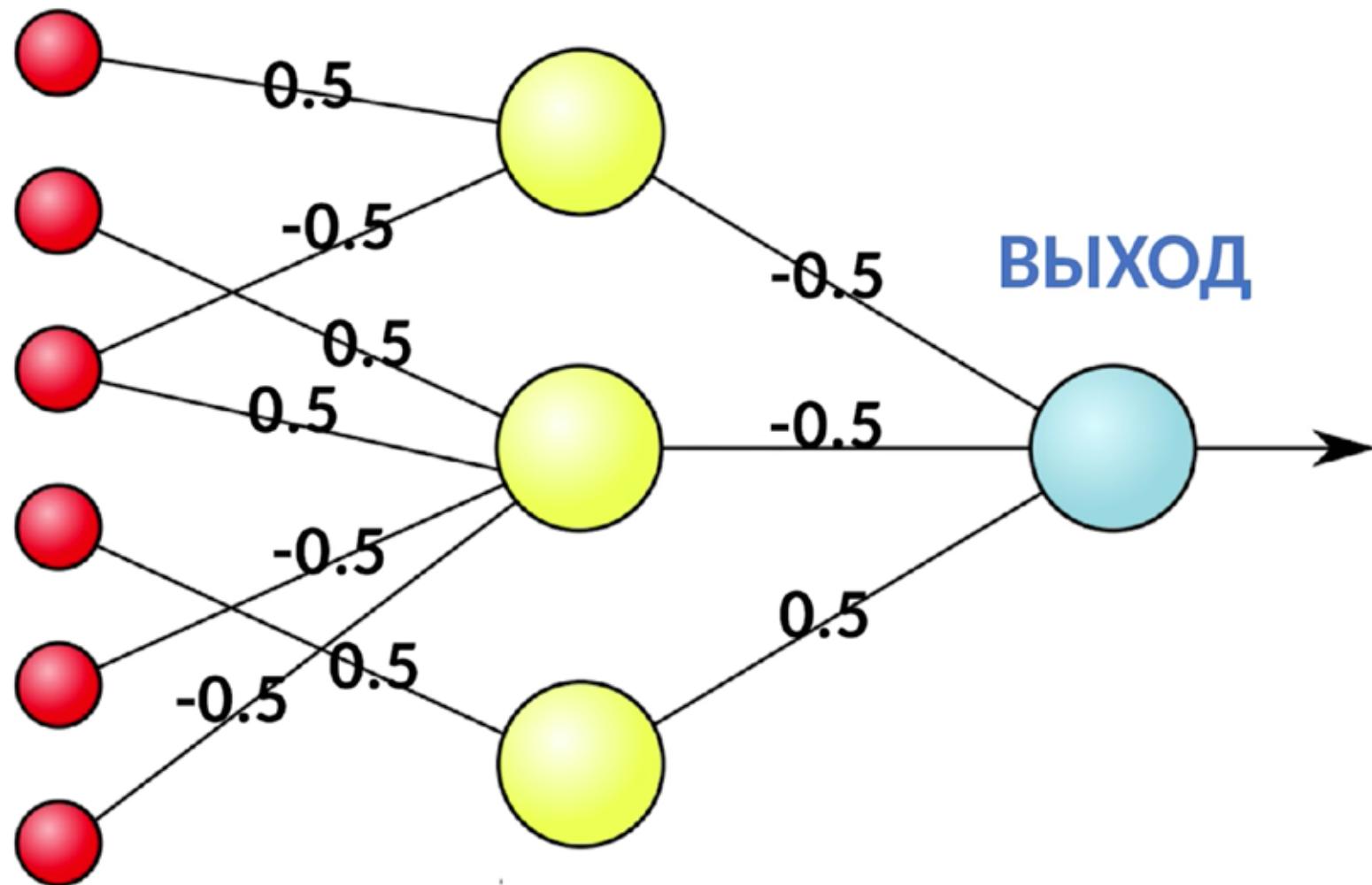
- Если их просто засунуть в классификатор, он ничего не выучит. Слишком сложно.
- Другой пример: картинка (хотим классифицировать фрукты – яблоки, бананы). Если просто засунем картинку (пиксели) в классификатор, он ничему не выучится. Нужно извлекать признаки.
- Можем ли создать алгоритм, который сам научится извлекать признаки? Да, так умеют нейросети.



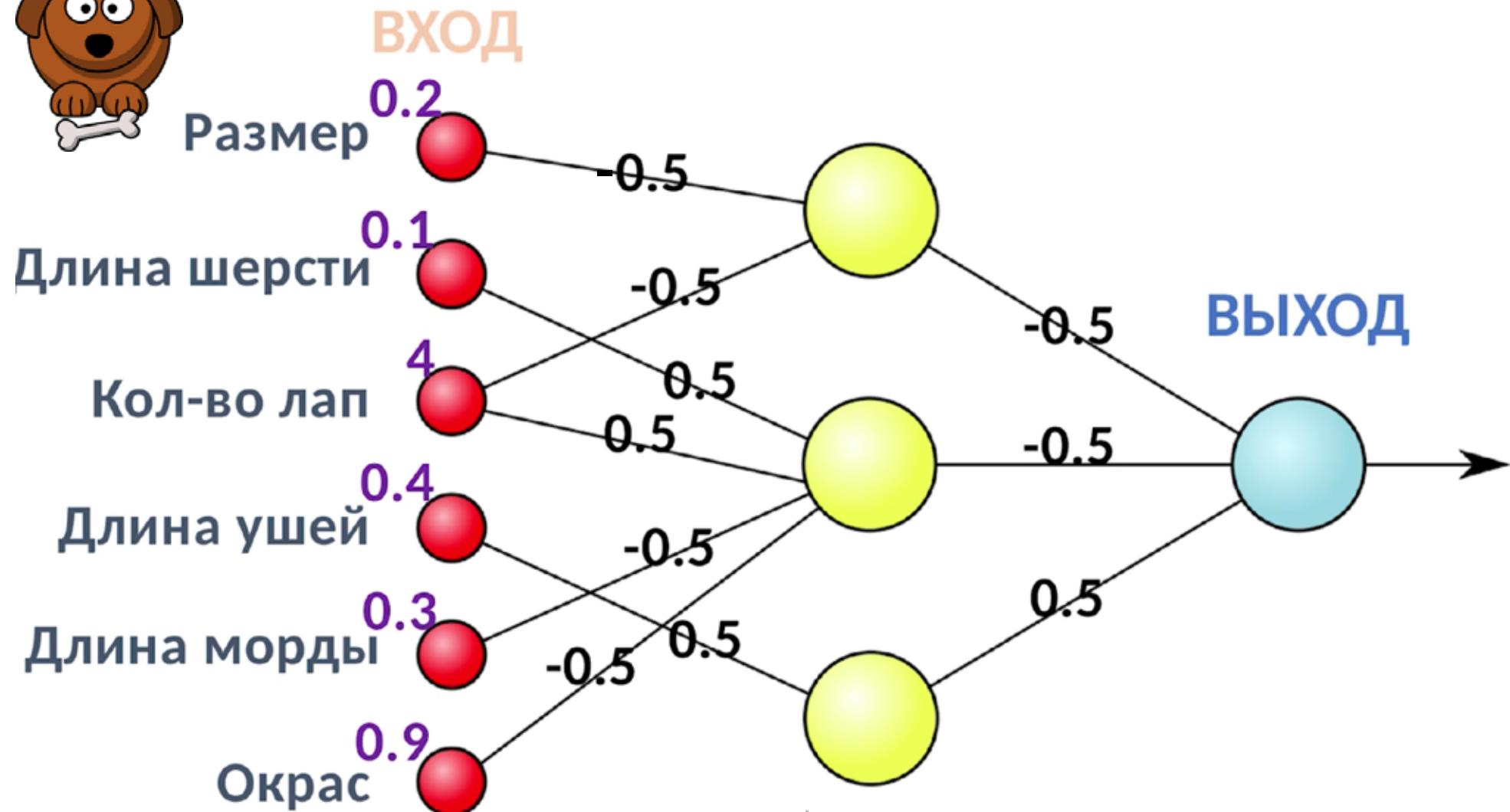
ВХОД

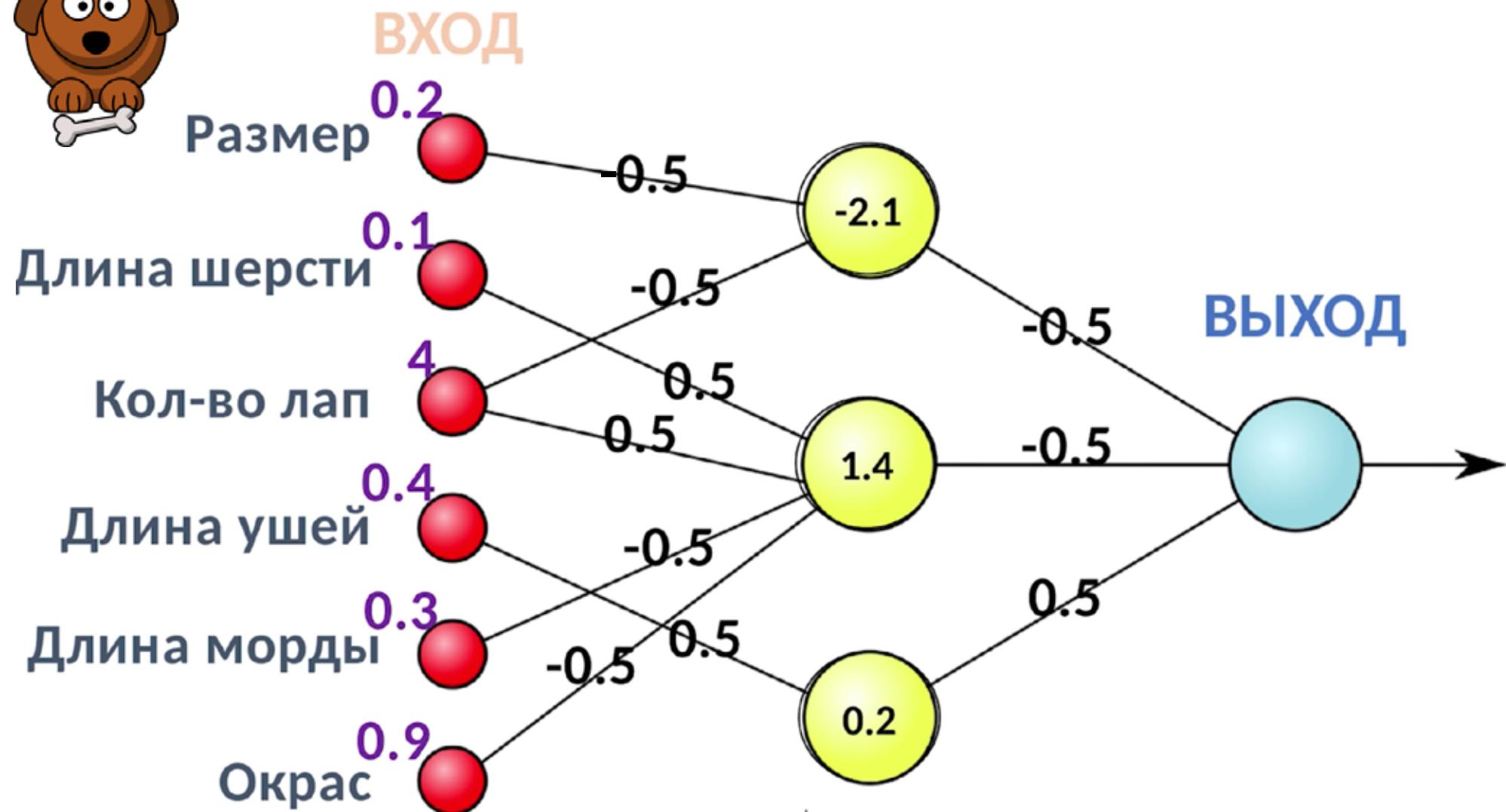


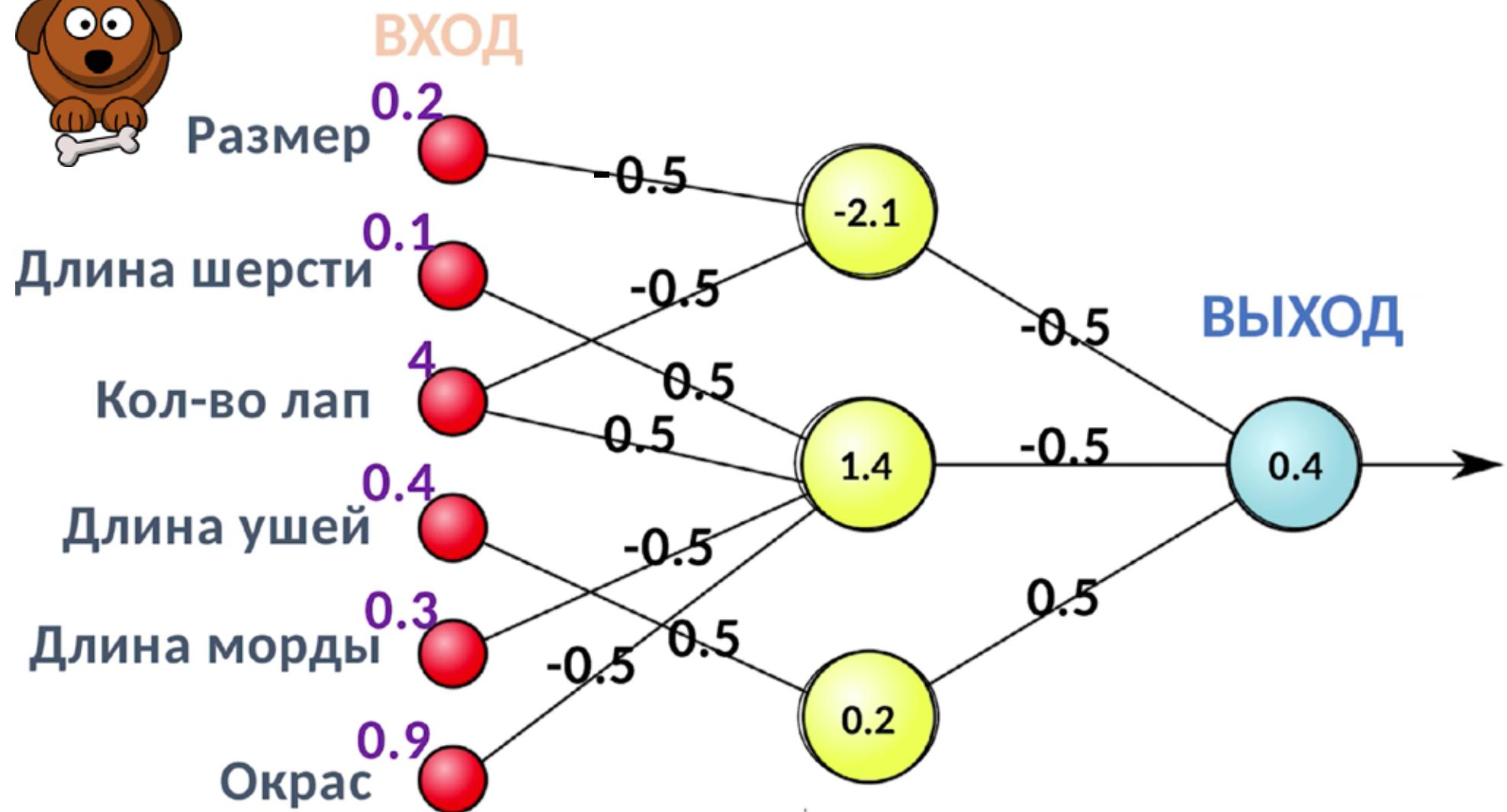
ВХОД

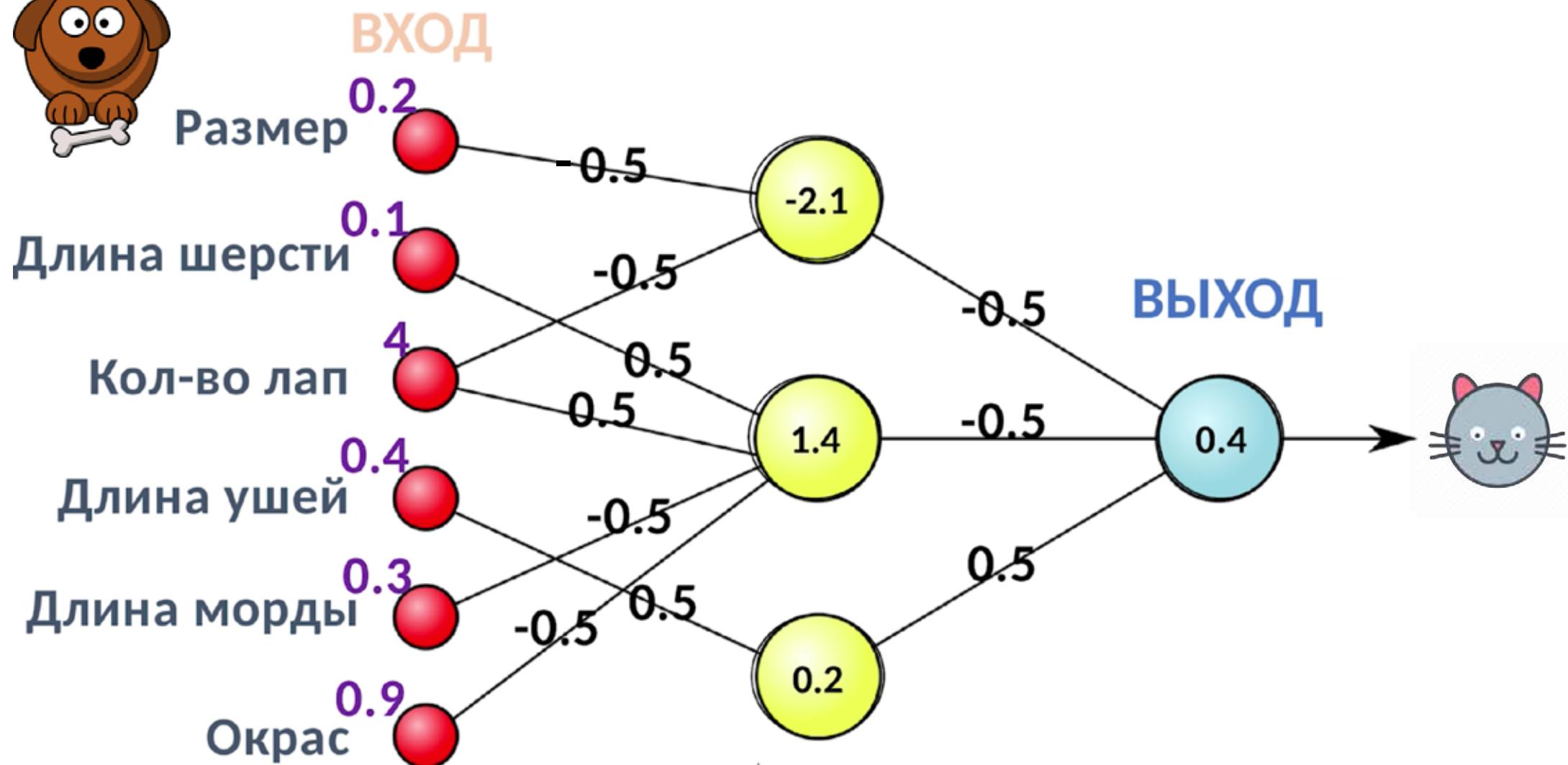


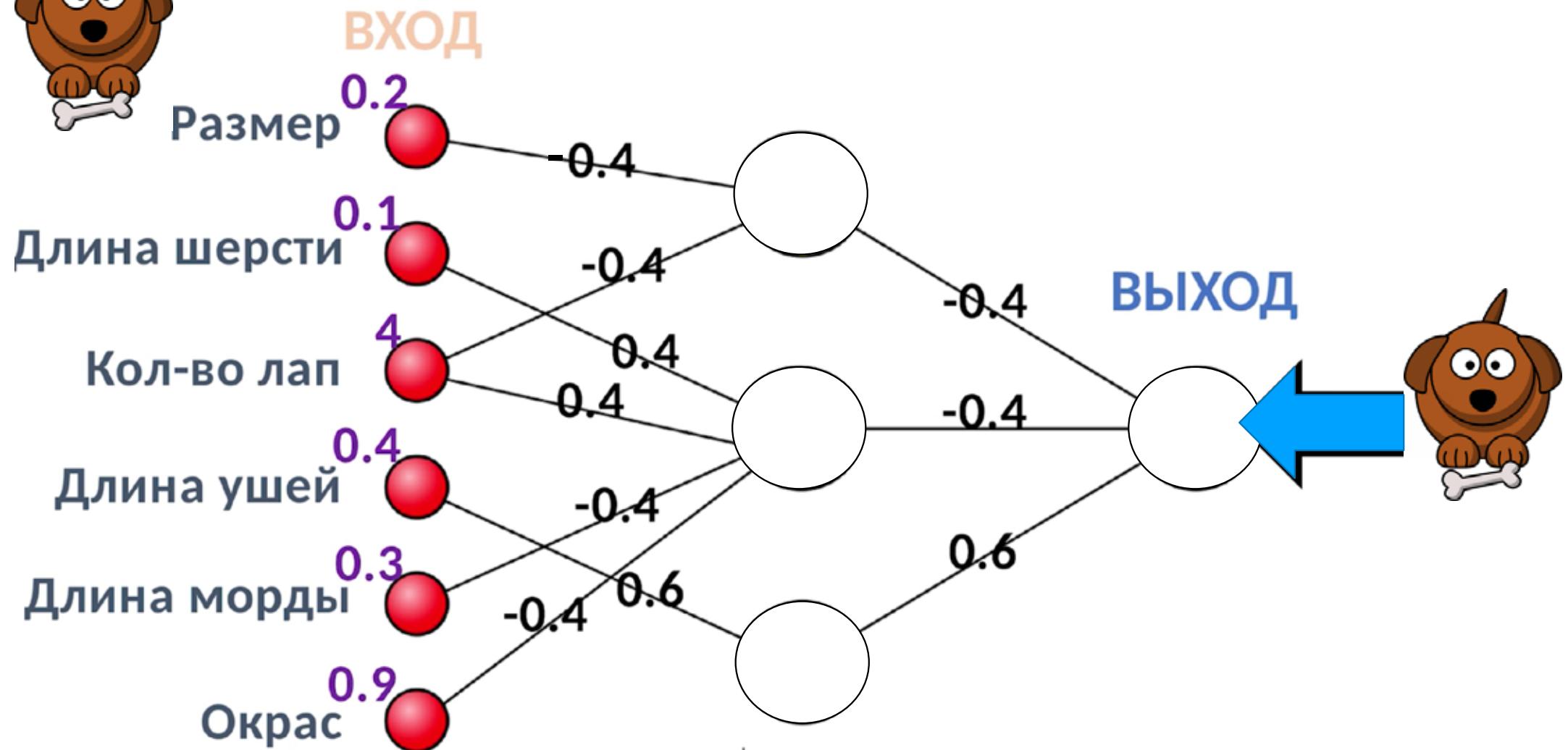
ВЫХОД











- В нейросети есть входные и выходные нейроны.
- Есть связи между ними, у каждой связи есть вес. Вес – это просто число, которое присваиваем связи между нейронами.
- Входные нейроны принимают информацию о признаках в нашем датасете.
- На выходе получается число (0 – кот, 1 – собака).
- Пока что нейросеть не обучена: подаем первую картинку на вход нейросети
- Вычисляем характеристики нашей картинки.
- Эти характеристики попадают на нейроны, потом на связи и на выходной нейрон – как сигнал распространяется в реальных нейронах.
- Входные признаки умножаются на веса каждой связи, суммируются и попадают в следующие нейроны.
- Точно также на следующем слое. Получили итоговое значение на выходе. К чему оно ближе – к 0 или 1?
- Нейросеть ошиблась, это для нее обратная связь, что нужно изменить веса связей и получить другие значения.
- Много раз покажем ей котиков и собачек, и тогда нейросеть обучится – то есть скорректирует веса в связях, и будет предсказывать правильно

Какие бывают нейронные сети?

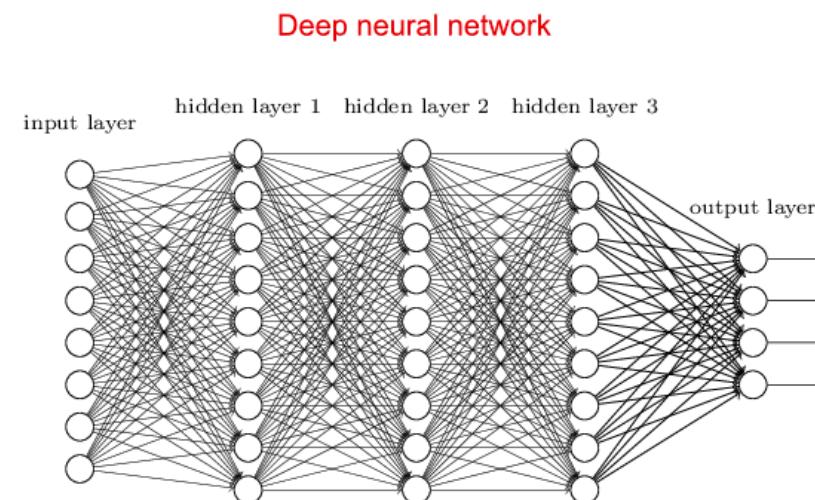
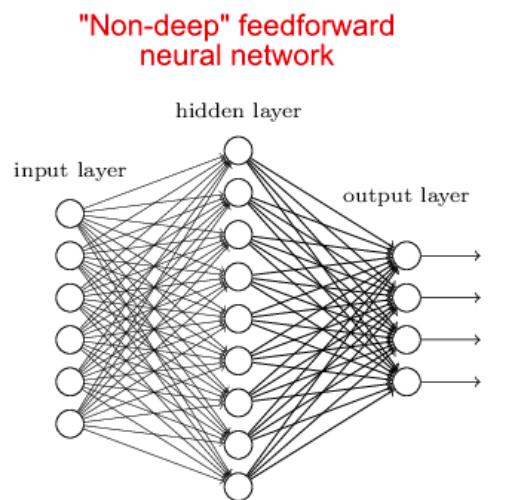
- Сверточные нейронные сети
- Рекуррентные нейронные сети
- Пораждающие состязательные сети
- Нейронные сети в обучении с подкреплением
- Нейронные сети с вниманием
- Нейронные сети с памятью

Нейросети и глубокое обучение: прорывы

- В распознавании речи (2010-е гг) – Microsoft, Google, IBM
- В распознавании изображений (сверточные сети) – распознавание символов, видео с камер слежения (2010-е гг), лиц (2014 г – Facebook).
- В обучении с подкреплением (Alpha Go) (2016 г), в роботике, в беспилотных автомобилях – Tesla, Google, Яндекс

Глубокое обучение

- Глубокое обучение – это обучение при помощи **глубоких нейронных сетей**. Глубокие нейросети имеют большое количество скрытых слоев.



Глубокое обучение

- **Универсальная теорема аппроксимации:** искусственная нейронная сеть с одним скрытым слоем может аппроксимировать любую непрерывную функцию многих переменных с любой точностью. Зачем же тогда нужны глубокие нейронные сети?
- Оказалось, что глубокие нейронные сети могут делать это **быстрее и эффективнее**.
- Глубокие нейронные сети применяются во многих областях: компьютерном зрении, обработке естественного языка, разработке роботов и беспилотных автомобилей.

Компьютерное зрение

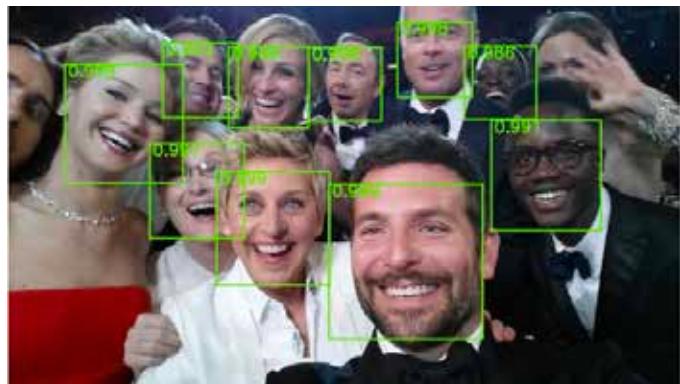
Компьютерное зрение (computer vision)

- Компьютерное зрение — это такая дисциплина (раздел искусственного интеллекта), которая занимается извлечением информации из изображений.
- Изображения могут быть разного типа. Это могут быть фотографии, видео, наборы фотографий или медицинский снимок из магнитно-резонансного томографа.

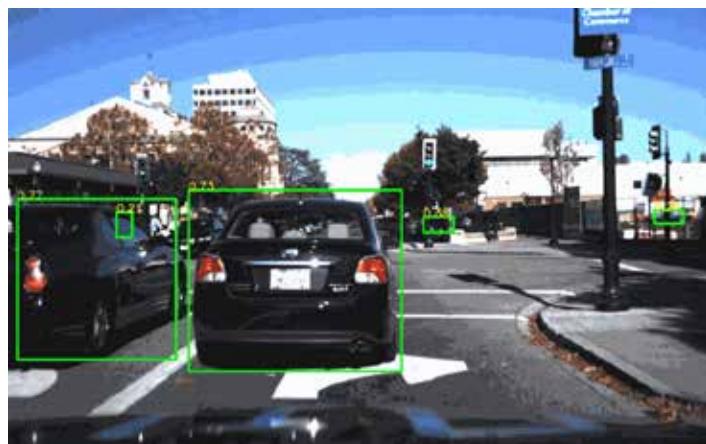
Computer vision



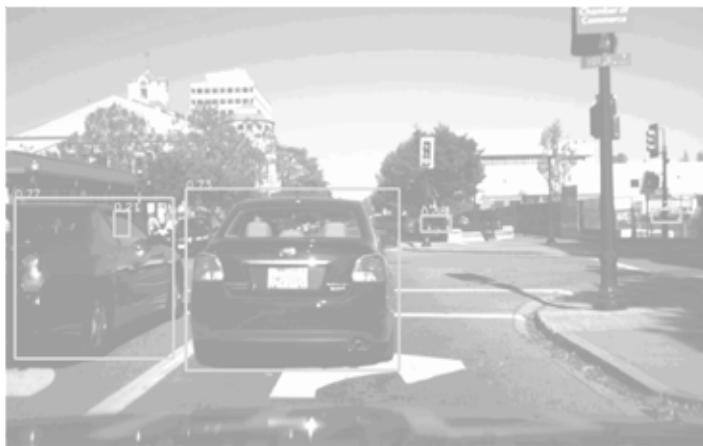
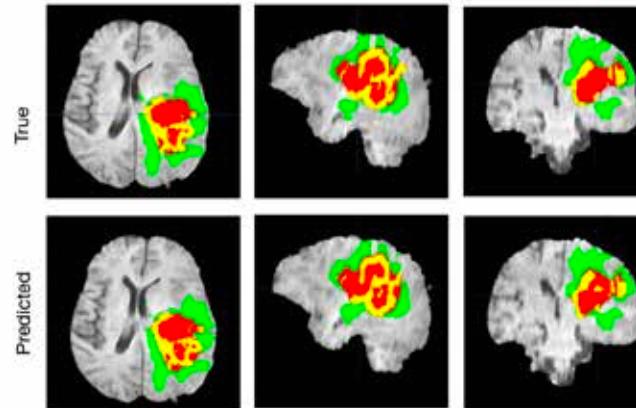
Computer vision



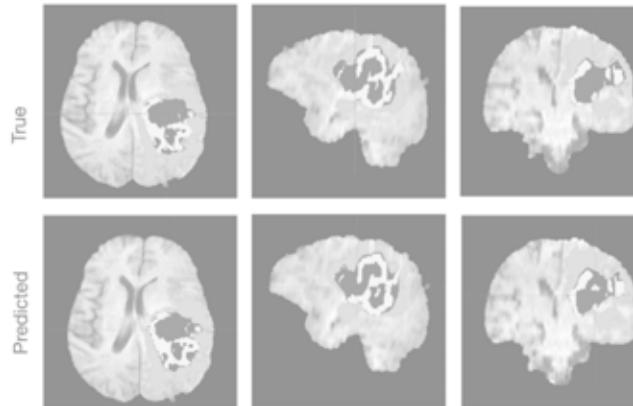
Computer vision



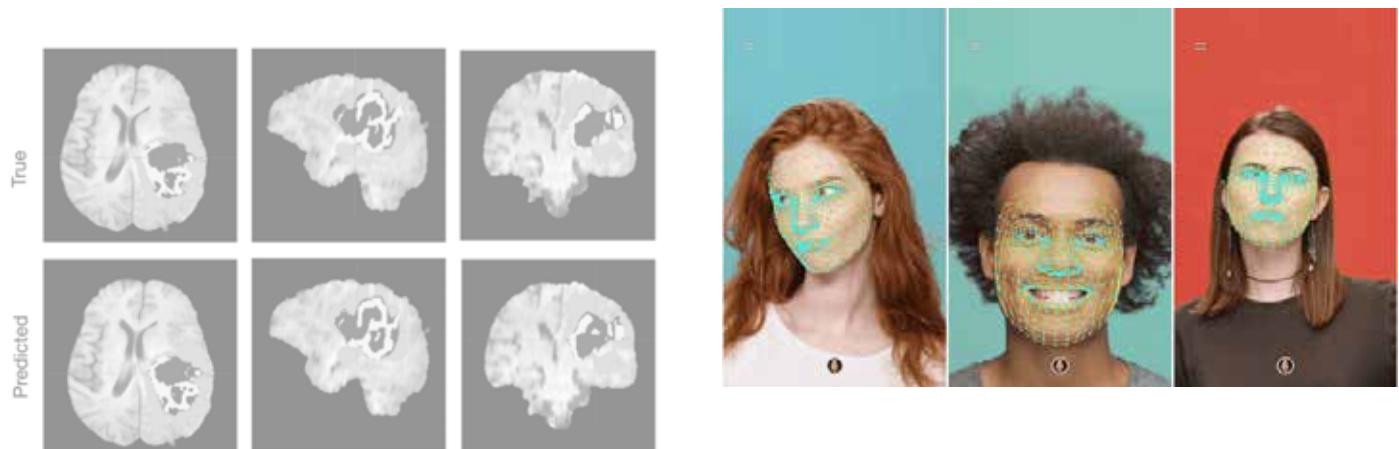
Computer vision



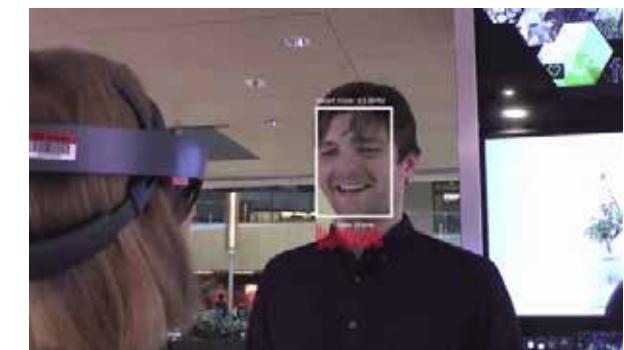
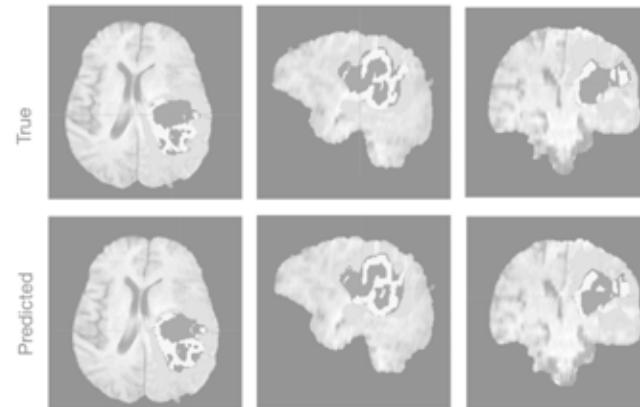
Computer vision



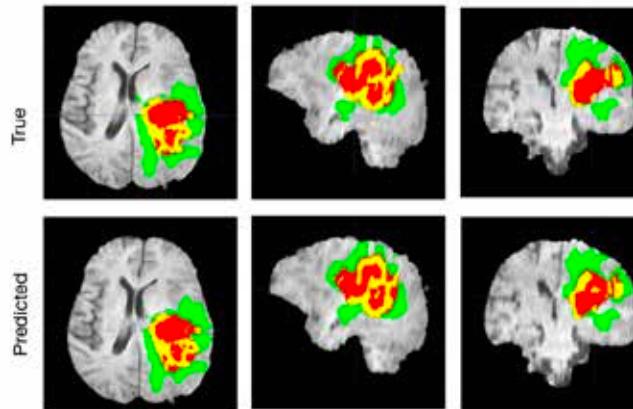
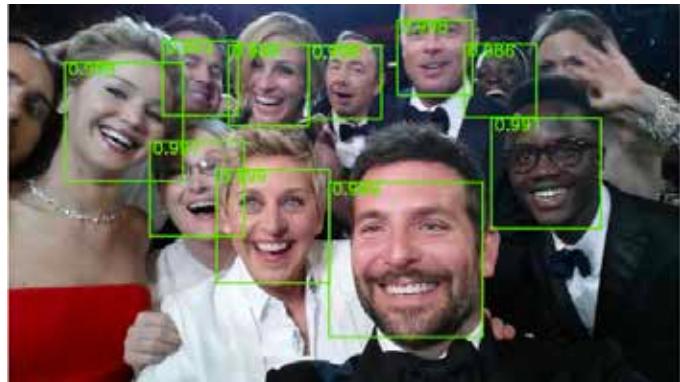
Computer vision



Computer vision



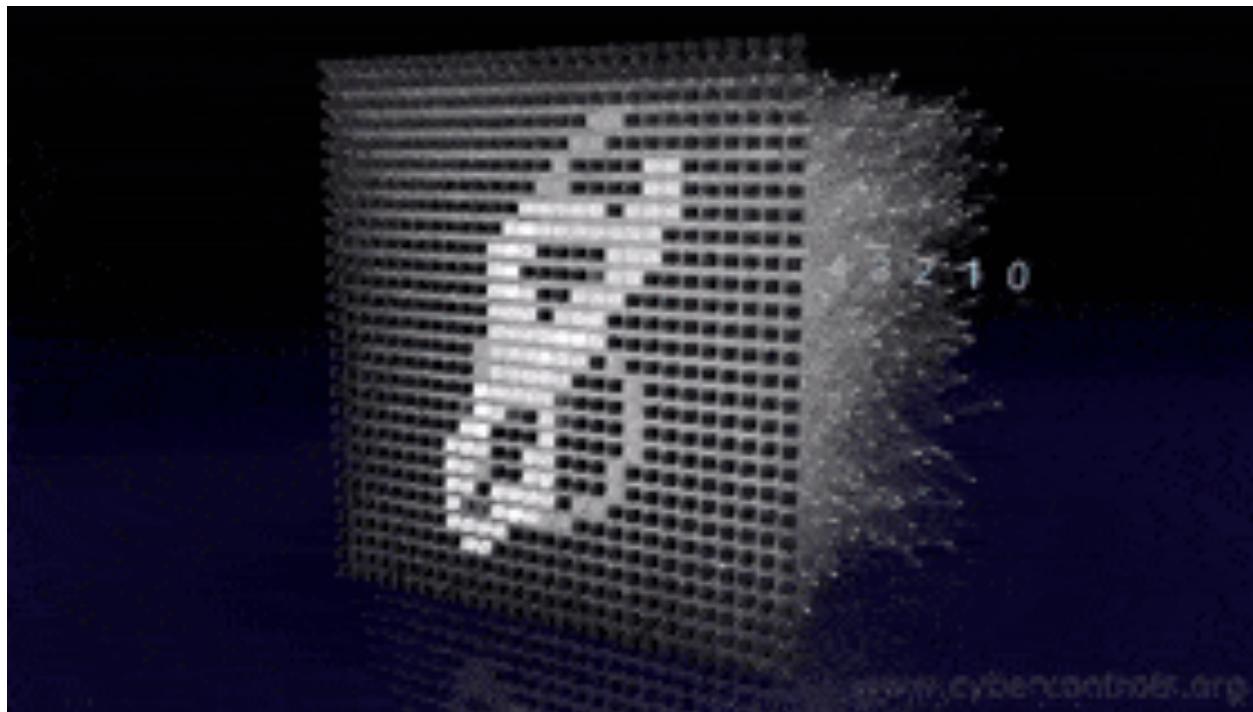
Computer vision



Задачи компьютерного зрения

- Распознавание
- Идентификация
- Восстановление 3D формы по 2D изображениям
- Отслеживание движения
- Сегментация изображений
- *И много других!*

Сверточные нейронные сети





What We See

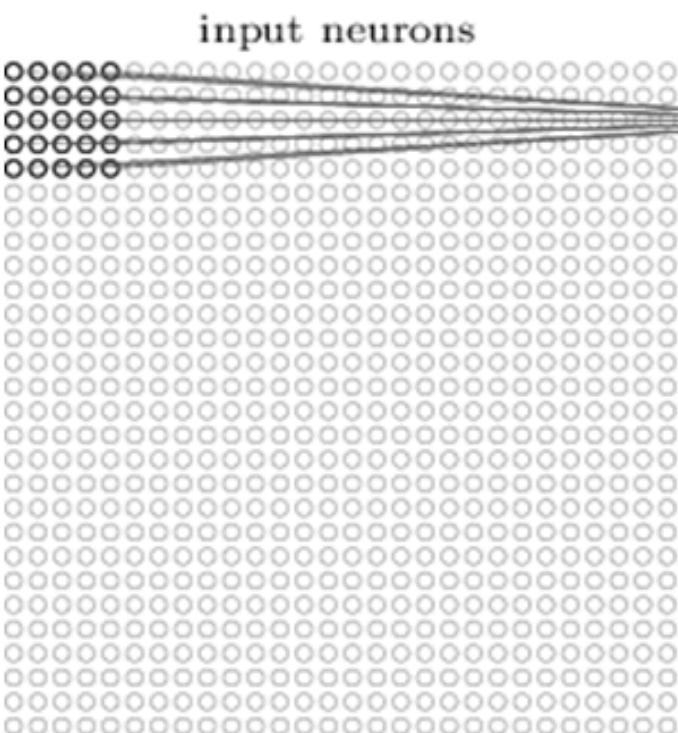
08 02 22 97 38 15 00 40 00 75 04 05 07 78 52 12 50 77 91 08
49 49 99 40 17 81 18 57 60 87 17 40 98 43 69 48 04 56 62 00
81 49 31 73 55 79 14 29 93 71 40 67 53 88 30 03 49 13 36 65
52 70 95 23 04 60 11 42 69 24 68 56 01 32 56 71 37 02 36 91
22 31 16 71 51 67 63 89 41 92 36 54 22 40 40 28 66 33 13 80
24 47 32 60 99 03 45 02 44 75 33 53 78 36 84 20 35 17 12 50
32 98 81 28 64 23 67 10 26 38 40 67 59 54 70 66 18 38 64 70
67 26 20 68 02 62 12 20 95 63 94 39 63 08 40 91 66 49 94 21
24 55 58 05 66 73 99 26 97 17 78 78 96 83 14 88 34 89 63 72
21 36 23 09 75 00 76 44 20 45 35 14 00 61 33 97 34 31 33 95
78 17 53 28 22 75 31 67 15 94 03 80 04 62 16 14 09 53 56 92
16 39 05 42 96 35 31 47 55 58 88 24 00 17 54 24 36 29 85 57
86 56 00 48 35 71 89 07 05 44 44 37 44 60 21 58 51 54 17 58
19 80 81 68 05 94 47 69 28 73 92 13 86 52 17 77 04 89 55 40
04 52 08 83 97 35 99 16 07 97 57 32 16 26 26 79 33 27 98 66
88 36 68 87 57 62 20 72 03 46 33 67 46 55 12 32 63 93 53 69
04 42 16 73 38 25 39 11 24 94 72 18 08 46 29 32 40 62 76 36
20 69 36 41 72 30 23 88 34 62 99 69 82 67 59 85 74 04 36 16
20 73 35 29 78 31 90 01 74 31 49 71 48 86 81 16 23 57 05 54
01 70 54 71 63 51 54 69 16 92 33 48 61 43 52 01 89 19 67 48

What Computers See



What We See

14	34	44	54	64	74	84	94	104	114	124	134	144	154	164	174	184	194	204
15	35	45	55	65	75	85	95	105	115	125	135	145	155	165	175	185	195	205
16	36	46	56	66	76	86	96	106	116	126	136	146	156	166	176	186	196	206
17	37	47	57	67	77	87	97	107	117	127	137	147	157	167	177	187	197	207
18	38	48	58	68	78	88	98	108	118	128	138	148	158	168	178	188	198	208
19	39	49	59	69	79	89	99	109	119	129	139	149	159	169	179	189	199	209
20	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200	200
21	41	51	61	71	81	91	101	111	121	131	141	151	161	171	181	191	201	201
22	42	52	62	72	82	92	102	112	122	132	142	152	162	172	182	192	202	202
23	43	53	63	73	83	93	103	113	123	133	143	153	163	173	183	193	203	203
24	44	54	64	74	84	94	104	114	124	134	144	154	164	174	184	194	204	204
25	45	55	65	75	85	95	105	115	125	135	145	155	165	175	185	195	205	205
26	46	56	66	76	86	96	106	116	126	136	146	156	166	176	186	196	206	206
27	47	57	67	77	87	97	107	117	127	137	147	157	167	177	187	197	207	207
28	48	58	68	78	88	98	108	118	128	138	148	158	168	178	188	198	208	208
29	49	59	69	79	89	99	109	119	129	139	149	159	169	179	189	199	209	209
30	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200	200	200
31	51	61	71	81	91	101	111	121	131	141	151	161	171	181	191	201	201	201
32	52	62	72	82	92	102	112	122	132	142	152	162	172	182	192	202	202	202
33	53	63	73	83	93	103	113	123	133	143	153	163	173	183	193	203	203	203
34	54	64	74	84	94	104	114	124	134	144	154	164	174	184	194	204	204	204
35	55	65	75	85	95	105	115	125	135	145	155	165	175	185	195	205	205	205
36	56	66	76	86	96	106	116	126	136	146	156	166	176	186	196	206	206	206
37	57	67	77	87	97	107	117	127	137	147	157	167	177	187	197	207	207	207
38	58	68	78	88	98	108	118	128	138	148	158	168	178	188	198	208	208	208
39	59	69	79	89	99	109	119	129	139	149	159	169	179	189	199	209	209	209
40	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200	200	200	200

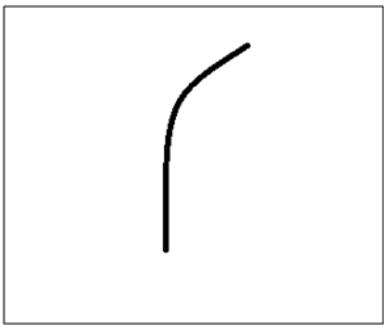


first hidden layer

Visualization of 5×5 filter convolving around an input volume and producing an activation map

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

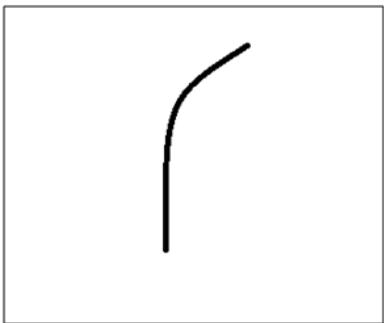
Pixel representation of filter



Visualization of a curve detector filter

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

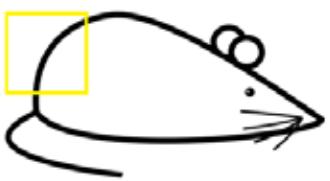
Pixel representation of filter



Visualization of a curve detector filter



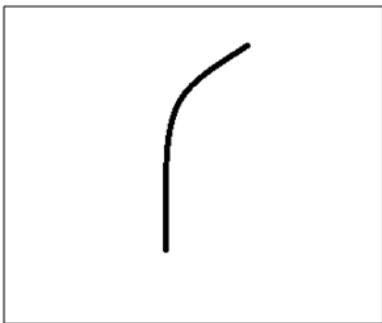
Original image



Visualization of the filter on the image

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

Pixel representation of filter



Visualization of a curve detector filter

0	0	0	0	0	0	30
0	0	0	0	50	50	50
0	0	0	20	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0

Visualization of the receptive field

0	0	0	0	0	0	30
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

Pixel representation of the receptive field

*

0	0	0	0	0	0	30
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

Pixel representation of filter

$$\text{Multiplication and Summation} = (50 \cdot 30) + (50 \cdot 30) + (50 \cdot 30) + (20 \cdot 30) + (50 \cdot 30) = 6600 \text{ (A large number!)}$$



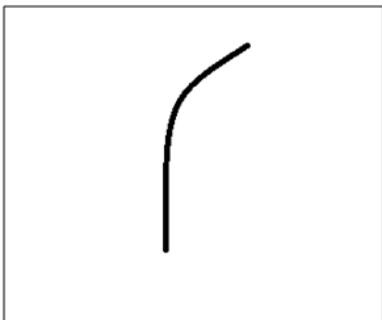
Original image



Visualization of the filter on the image

0	0	0	0	0	0	30	0
0	0	0	0	30	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	0	0	0	0	0

Pixel representation of filter



Visualization of a curve detector filter



Original image



Visualization of the filter on the image



Visualization of the receptive field

0	0	0	0	0	0	30	0
0	0	0	0	50	50	50	0
0	0	0	20	50	0	0	0
0	0	0	50	50	0	0	0
0	0	0	50	50	0	0	0
0	0	0	50	50	0	0	0
0	0	0	50	50	0	0	0

0	0	0	0	0	0	30	0
0	0	0	0	30	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	0	0	0	0	0

Pixel representation of the receptive field

*

0	0	0	0	0	0	30	0
0	0	0	0	30	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	0	0	0	0	0

Pixel representation of filter

$$\text{Multiplication and Summation} = (50 \cdot 30) + (50 \cdot 30) + (50 \cdot 30) + (20 \cdot 30) + (50 \cdot 30) = 6600 \text{ (A large number!)}$$



Visualization of the filter on the image

0	0	0	0	0	0	0	0
0	40	0	0	0	0	0	0
40	0	40	0	0	0	0	0
40	20	0	0	0	0	0	0
0	50	0	0	0	0	0	0
0	0	50	0	0	0	0	0
25	25	0	50	0	0	0	0

Pixel representation of receptive field

0	0	0	0	0	0	30	0
0	0	0	0	30	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	0	0	0	0	0

Pixel representation of filter

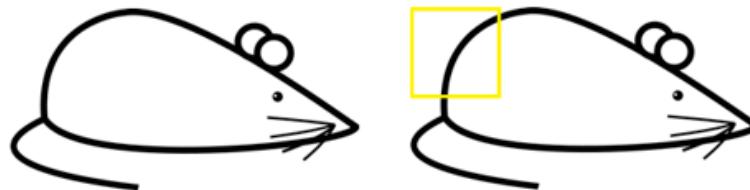
$$\text{Multiplication and Summation} = 0$$

Пример ядра с обученным признаком

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0



Входное изображение



В этом месте, где желтое окно, будет большой отклик (сигнал), что говорит о наличие этого признака на изображении.

Пример ядра с обученным признаком

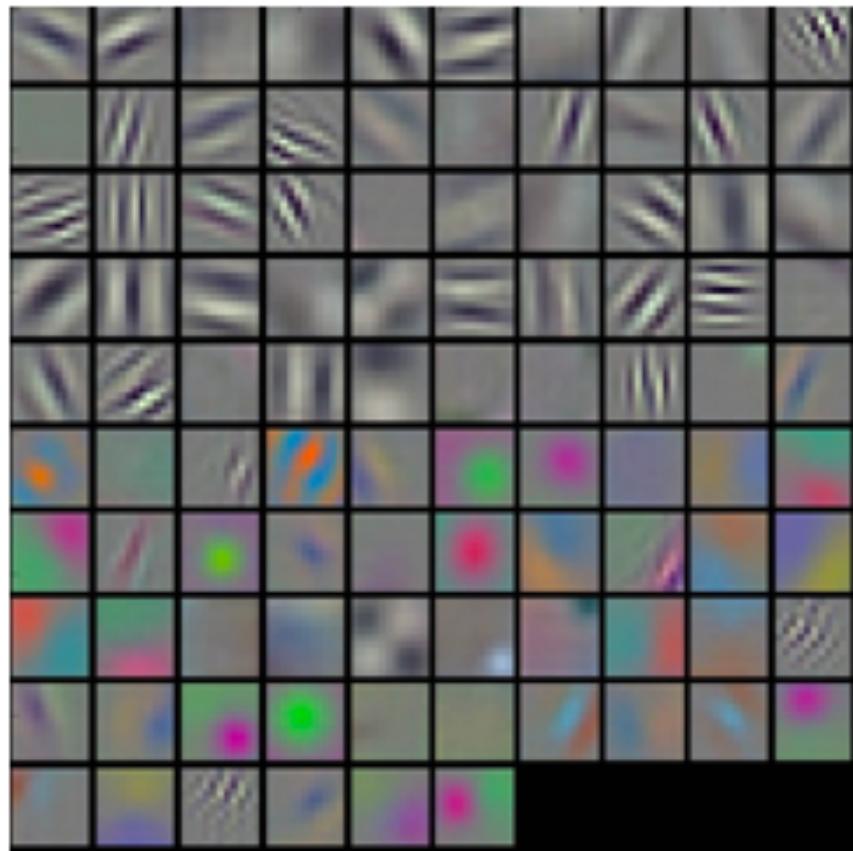
0	0	0	0	0	0	30	0
0	0	0	0	30	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	0	0	0	0	0



Входное изображение



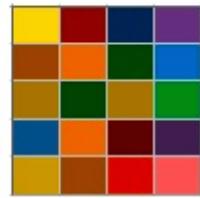
В этом месте, где желтое окно, будет большой отклик (сигнал), что говорит о наличие этого признака на изображении.

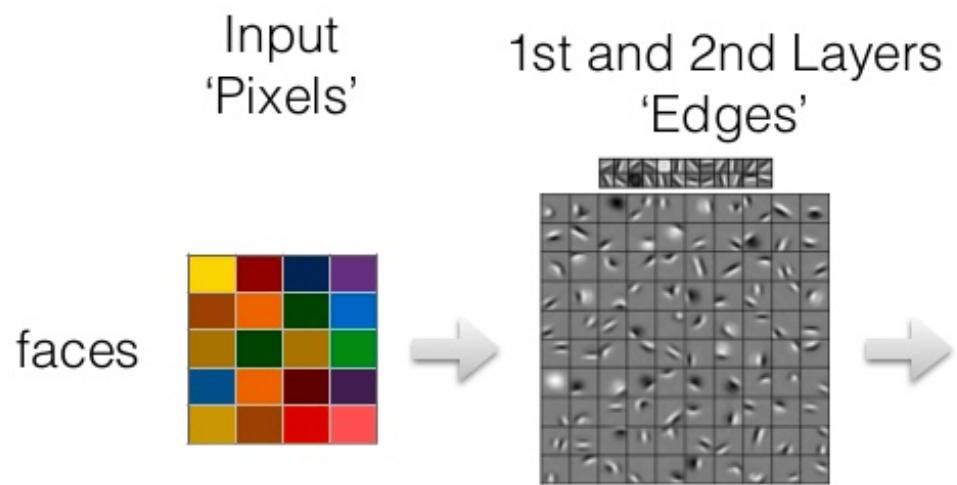


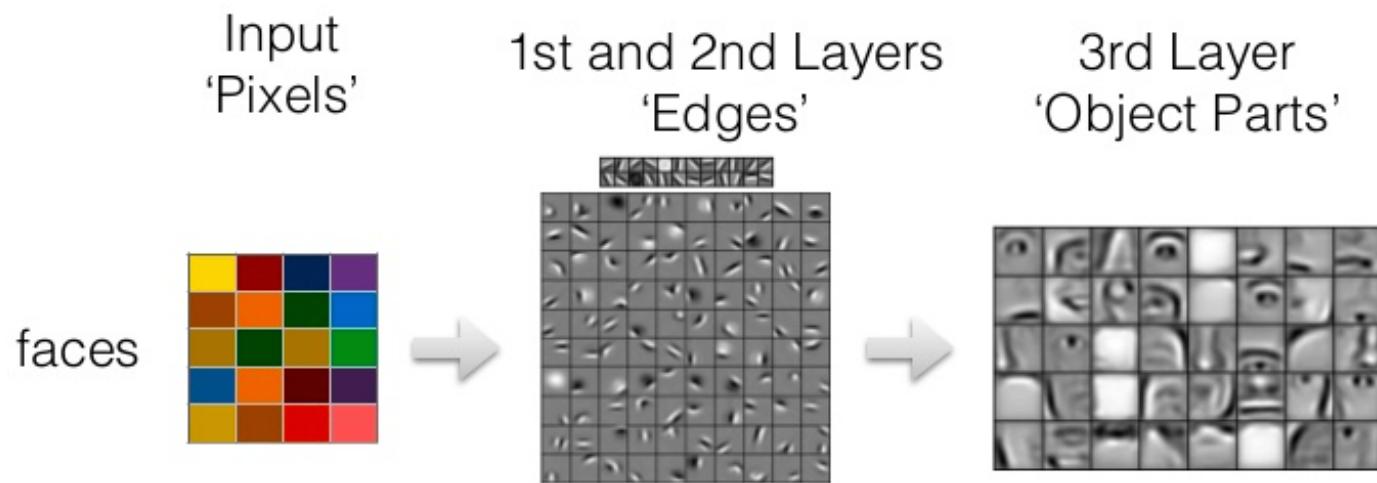
Visualizations of filters

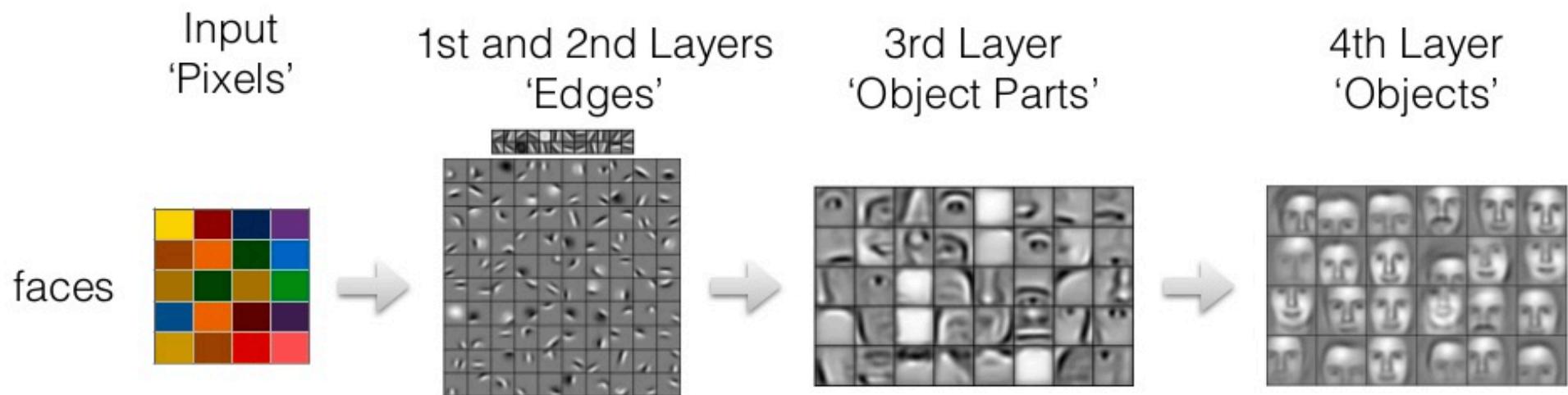
Input
'Pixels'

faces

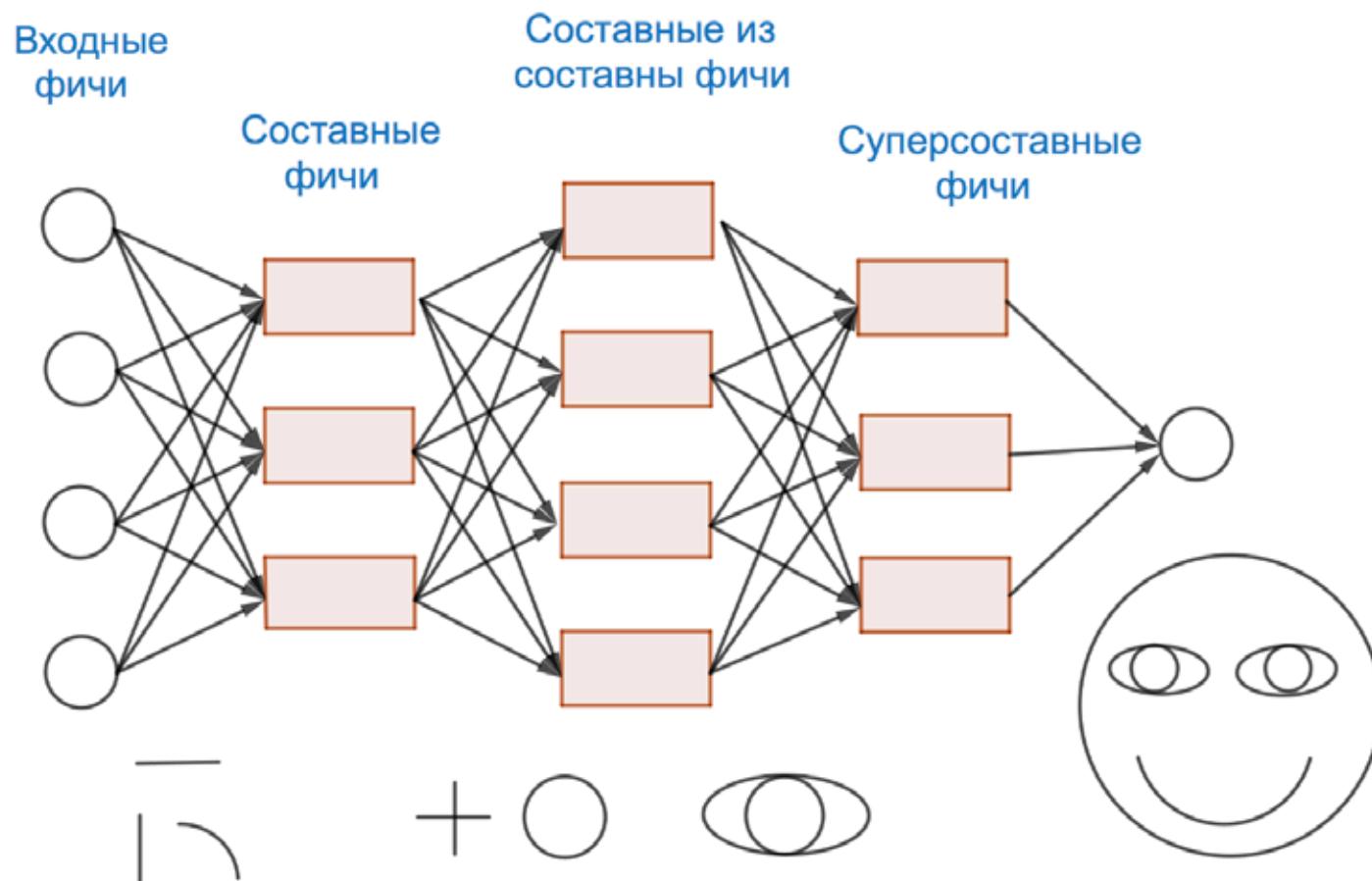








Что выучивают нейросети



Автоматическое распознавание эмоций по лицу

Faces communicate far more than emotion



- **Identity**
- **Age**
- **Race**
- **Gender**

Emotion recognition: challenges

- Accuracy of face detection, feature tracking, and expression recognition influenced by:
 - Background
 - Presence of other people



Emotion recognition: challenges

- Ракурс

Impacts human interpretation as well

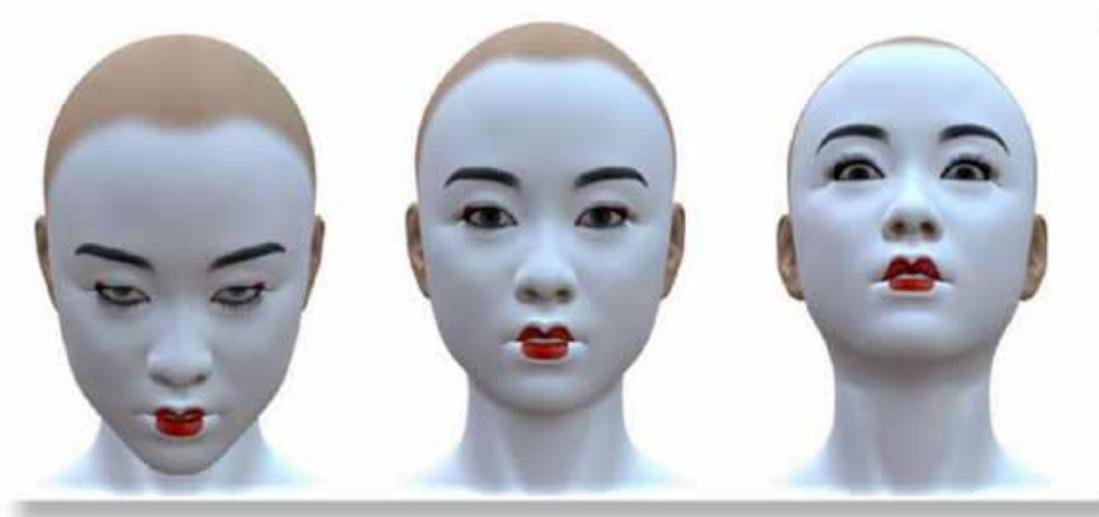
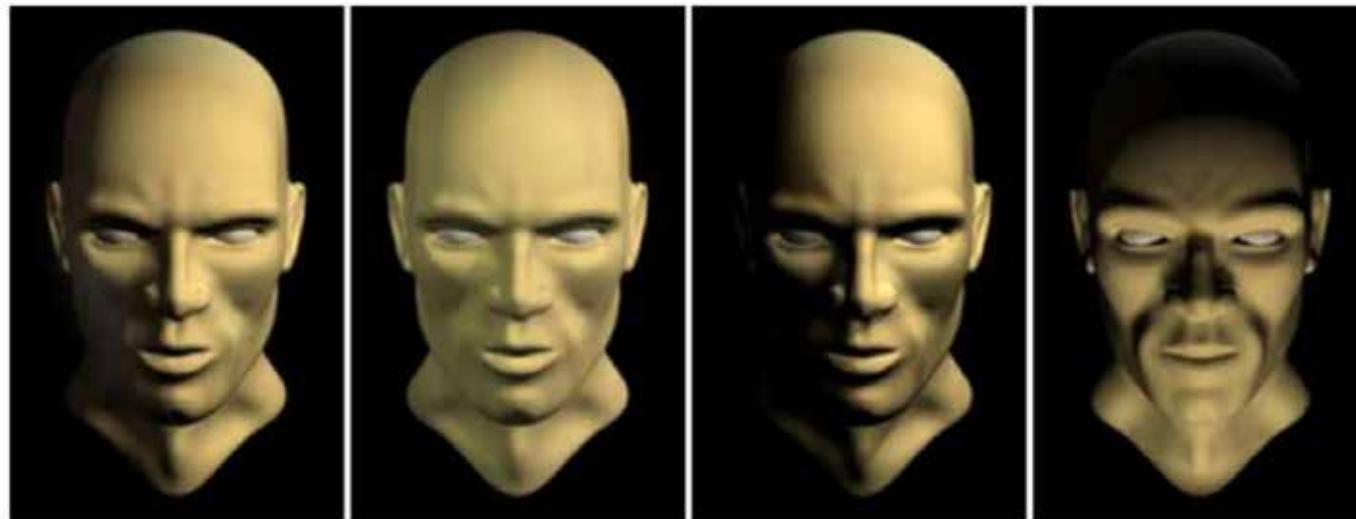


Image: Arvid Kappas

Emotion recognition: challenges

- Освещение

Cinematography emphasizes how light can create emotion



Emotion recognition: challenges

- Другие модальности



a



b

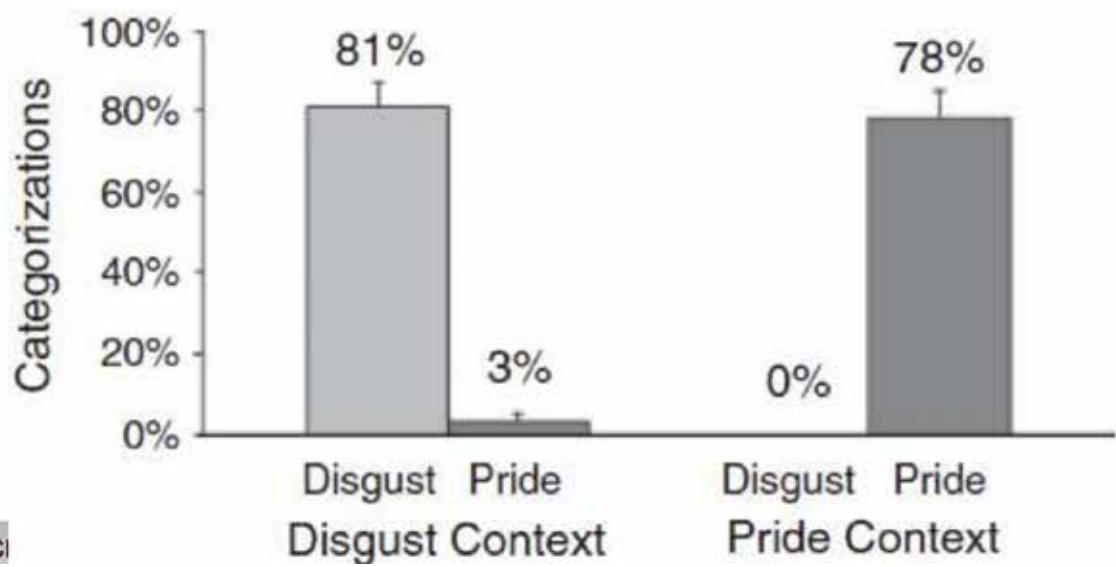


c



d





Практика

<http://bit.ly/36tsKHv>

HW_1

- Грибкова – метрики больше 100%
- Клепицкий – комменты?
- Степанищева – комменты?
- Ермаков – комменты?
- Самоукина – комменты?
- Антонов – комменты?
- Мурзякова – комменты?
- Иванина – только одна модель, комменты?
- Минаева – не приложила файлик
- Кто не прислал – присылайте!
- Дедлайн (**финальный!**): 31 января

HW_2

- Доделать практическое задание
- Прислать отчет на почту o.perepelkina@neurodatalab.com
- Дедлайн: 31 января