



# Пчелы и AI



**Артём Курапов**

Engineer at Clarifai / Founder at Gratheon

# О себе

пишу веб приложения на php ts go ru  
многое повидал

Люблю простой деплой и надежный код

🐝 Раз пчеловодил что скорую вызывал

🤖 Не ML - глубоко не знаю

Занятно:

🔧 Хакатоню

🐈 Предпочитаю кошачих

🔧 Увлекаюсь EDC, часами, steam, умным домом

🙏 За opensource и прозрачные процессы

Критик (INTJ) с плохой памятью на имена



Senior Software Engineer

Clarifai

Jan 2023 - Present · 11 mos



OPUS





## Личный опыт

- 2019 - Сад = яблони = пчелы
- Курс пчеловодства Korpli Ametikool
- Купил 3 улья
- Купил 2 Buckfast семьи
- Один дикий рой вселился сам
- 2020 - 7 семей, фото-инспекции
- 2021 - 7 семей, переезд, плохая зима
- 2022 - 1 семья, 2 роя, шершни
- 2023 - 1 семья



# Содержание

- Пчелы
- Пчеловодство
- Проблемы
- Существующие решения
- **Grattheon.com**
- Machine Learning
- 🐝 Deep Learning для зрения
- 🐝 Научные работы по пчелам
- 🐝 Смысловые нейронки
- Модальности
- Будущее и Clarifai.com



Слайды презентации





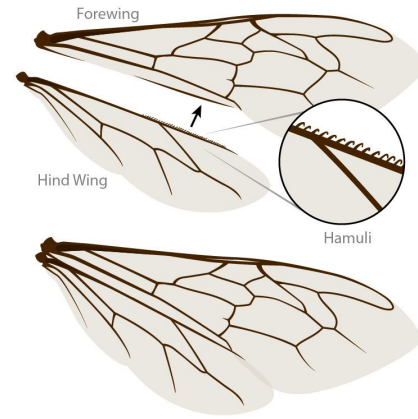


**Пчелы**

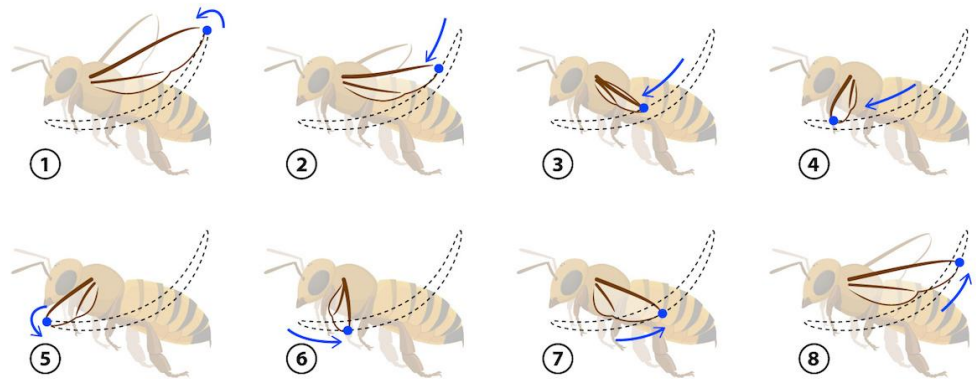


# Физиология / Крылья

- маленькие
- пара сцеплена крючками
- 230 взмахов в секунду
- улетают на 2-4 км от улья
- 20 - 65 км/ч



----- Wing path      ● Wing tip      → Direction of wing movement



# Физиология / Волоски

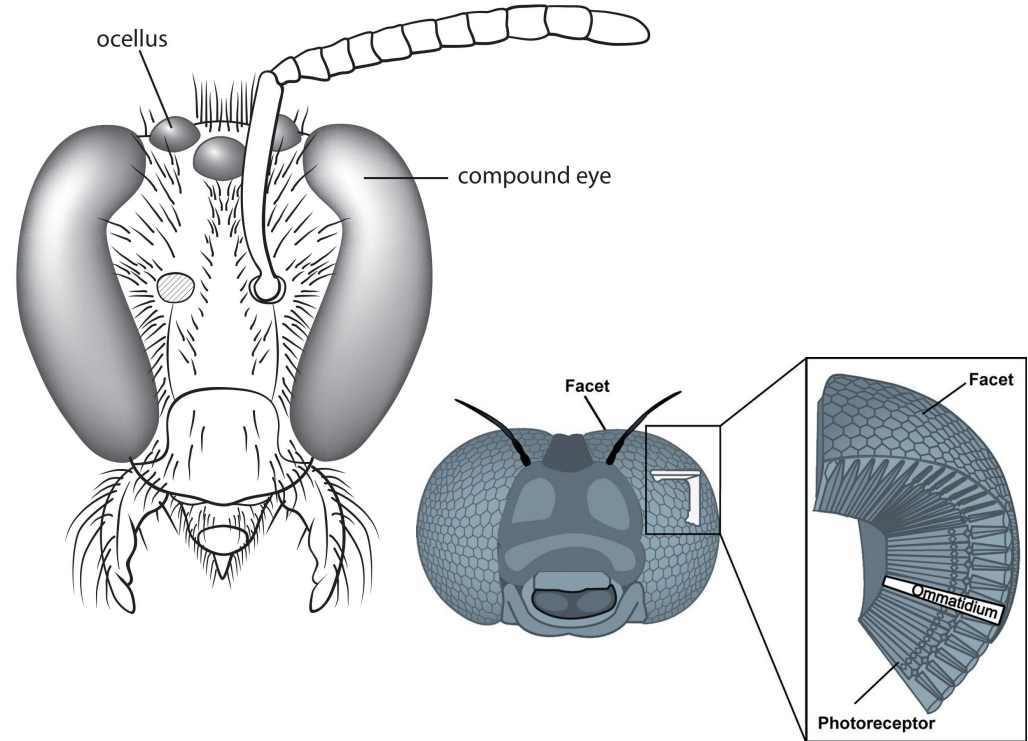
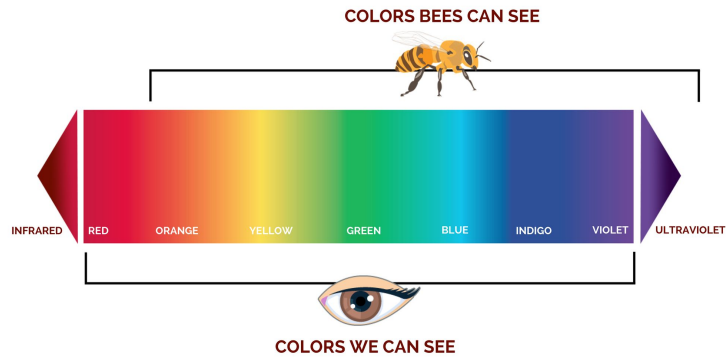
- осязание в темноте
- волосы + электростатика = магнит пыльцы
- в т.ч. на глазах





# Физиология / Глаза

- не видят красный
- трутни лучше видят



# Физиология / Глаза



Глаза рабочей пчелы

<https://www.flickr.com/photos/usgsbiml/8682047014/in/album-72157664305903459/>



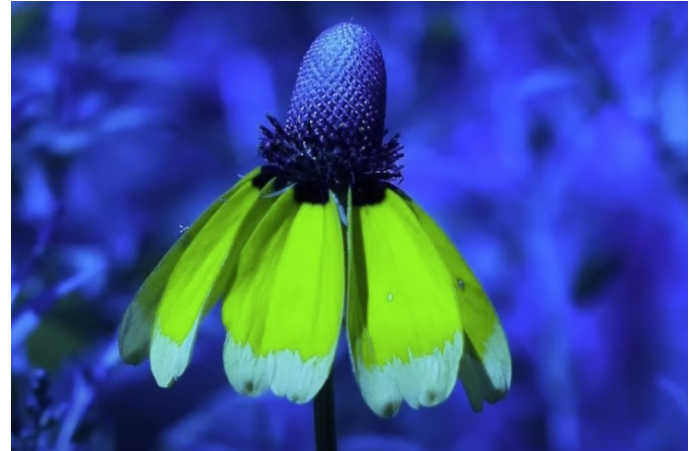
Глаза трутня

<https://www.flickr.com/photos/usgsbiml/14283379287>



# Физиология / Глаза

- растения стараются для пчел



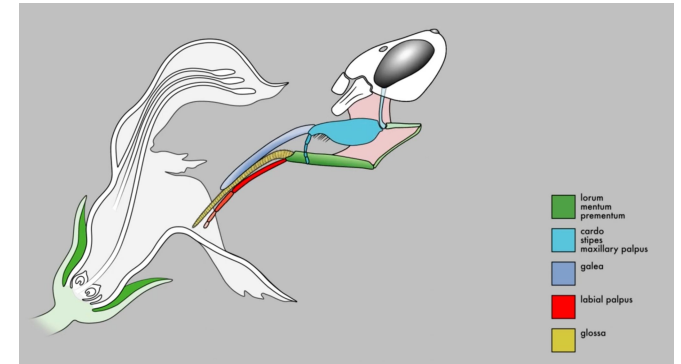
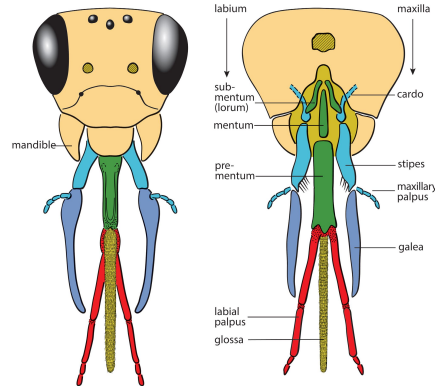


# Физиология / Рот

выдвижной механизм

трубочка для нектара

язык

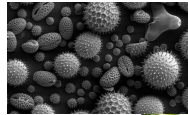
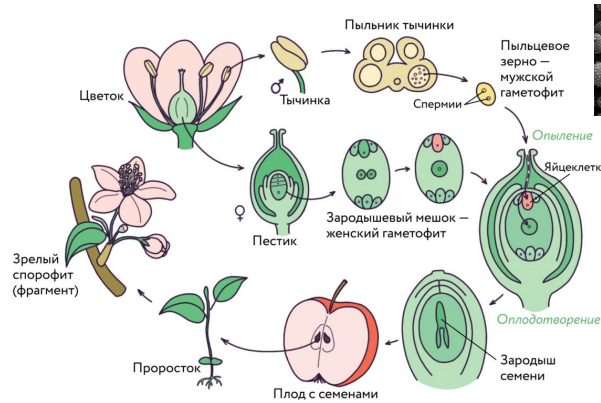


# Пчелы и цветы

- 🦖 Первые цветы - 170 млн лет назад
- Цветок - половой орган размножения
- 20 млн лет - общественные насекомые
- 22000 видов пчел
  - 70% живет под землей



Ископаемая пчела,  
нижний мел (145-100 млн)



Отель для одиночных пчел



Стекланный глаз

Фото: Клеберсон Бервиан



Пчела канудо, Бразилия

Фото: Ясмин Амиден

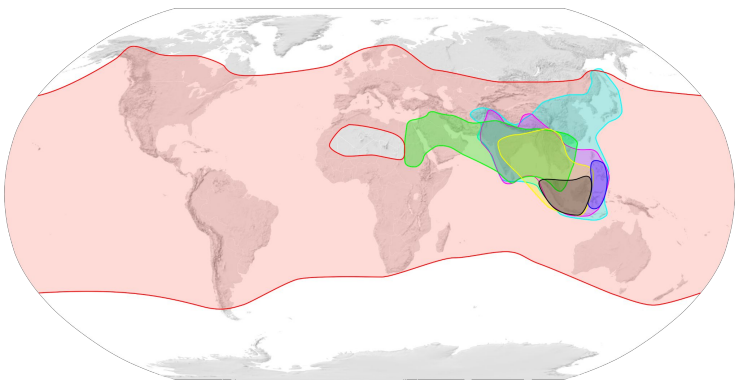




# Медоносные пчелы

~ 8 видов общественных медоносных пчел  
(10% от всех), например *Apis cerana*

~ Основной вид - *Apis Mellifera*  
~ 40 подвидов  
(*Apis Mellifera Buckfast*)



итальянская

карника

среднерусская = европейская темная

горная кавказская



Медоносная Пчела  
*Apis mellifera*



*Apis mellifera capensis*  
Подвид



*Apis mellifera scutellata*  
Подвид



*Apis mellifera iberiensis*  
Подвид



*Apis mellifera unicolor*  
Подвид



*Apis mellifera adansonii*  
Подвид



*Apis mellifera ligustica*  
Подвид



*Apis mellifera carnica*  
Подвид



*Apis mellifera mellifera*  
Подвид



*Apis mellifera caucasia*  
Подвид



# Половой диморфизм и специализация



рабочие пчелы

- чистка сот
- согревание яиц
- кормление расплода (пыльцой)
- строительство сот
- защита входа
- сбор нектара и пыльцы



матка

- откладывание яиц (до 2000 в сутки)
- одухотворение

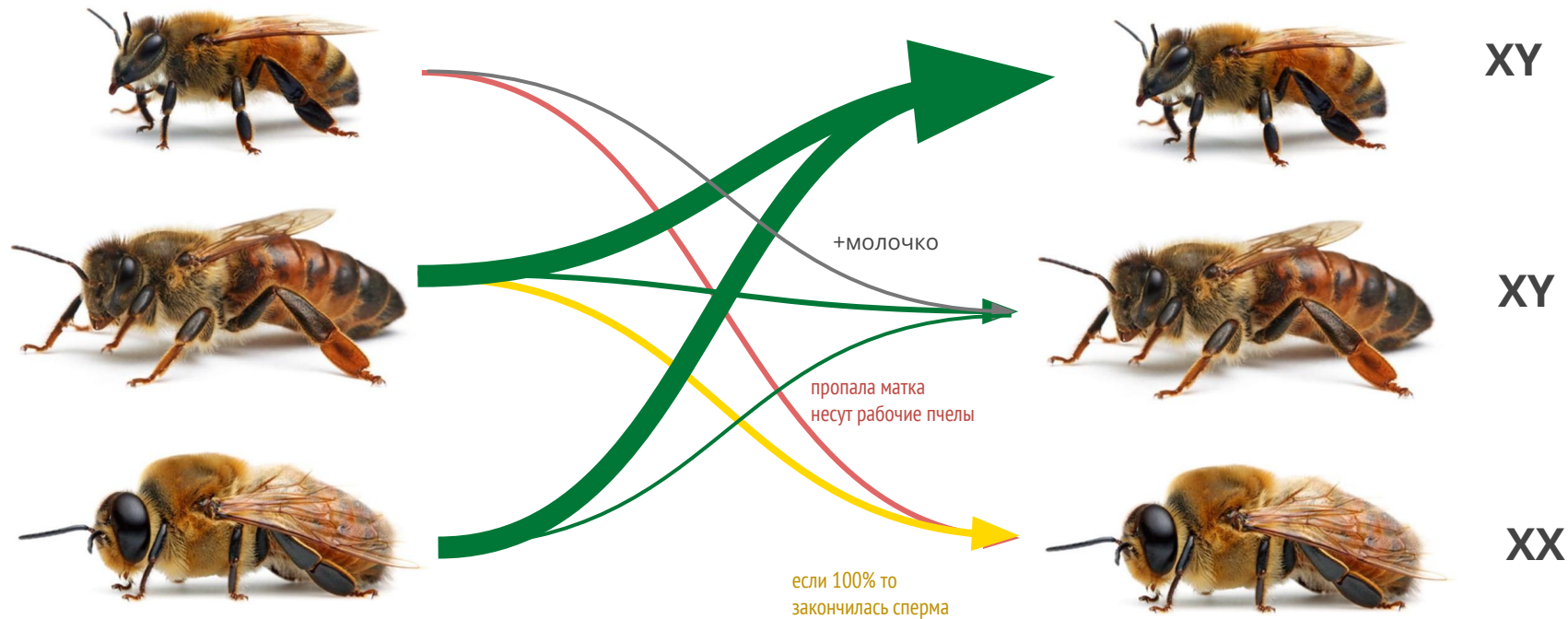


трутни

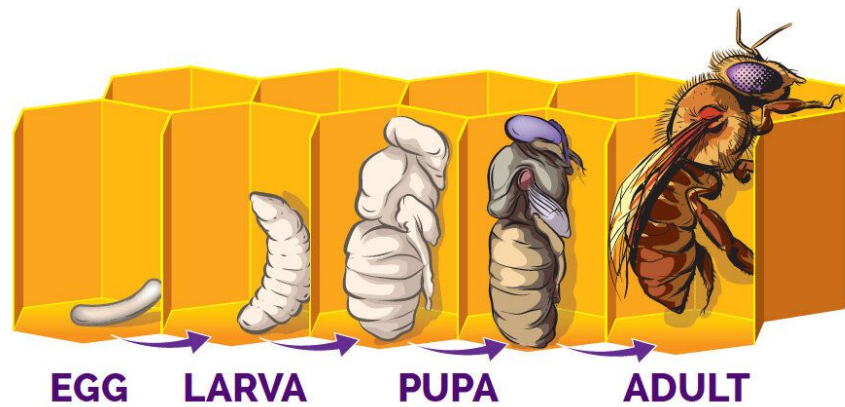
- оплодотворение чужих маток
- погибают в конце лета



# Гаплодиплоидия и непорочное зачатие



# Развитие



куколка, фото Marco Moretti





# Цикл жизни рабочих летом

- по 21 день

до “рождения” → взросления → в поле

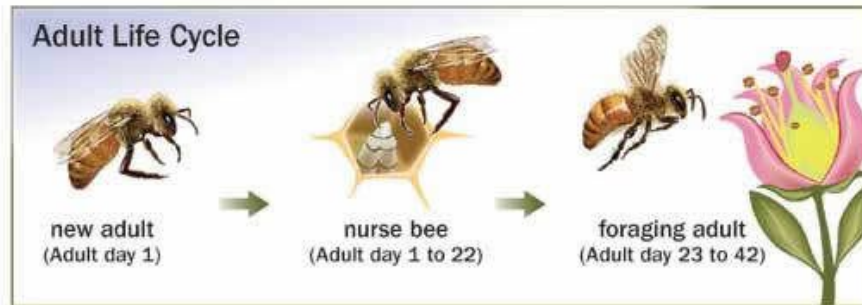


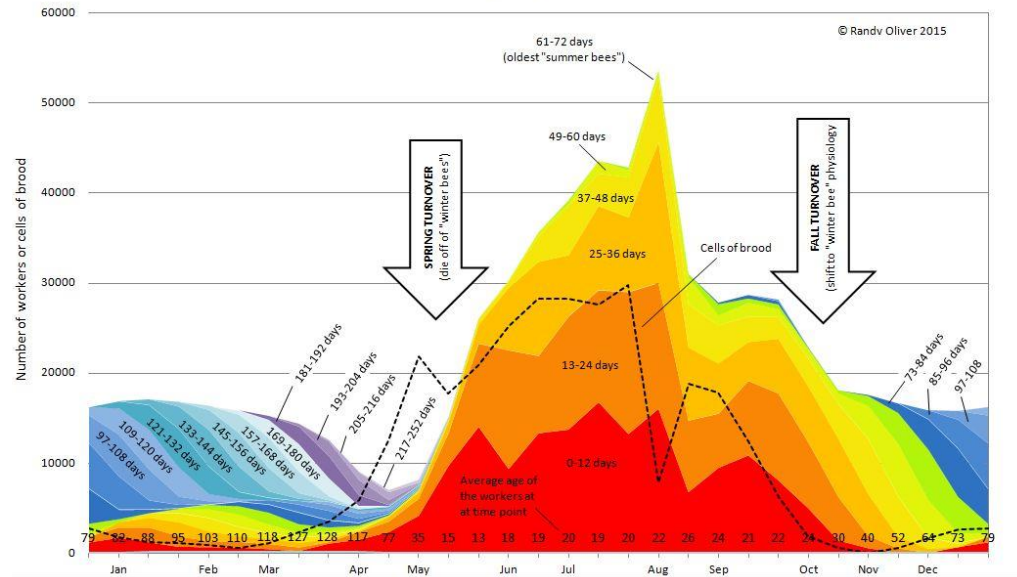
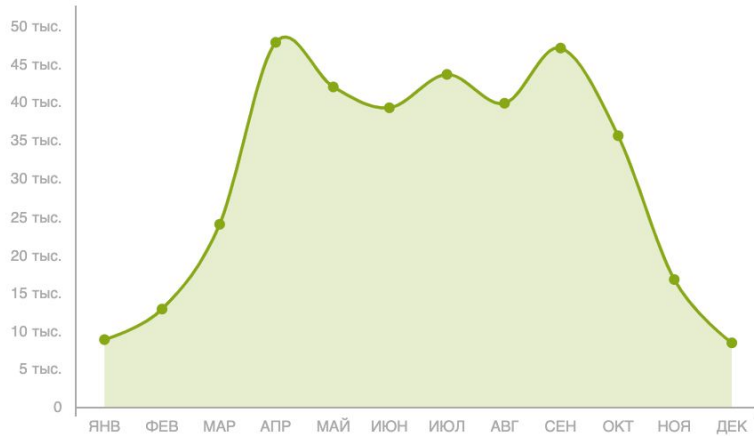
Illustration: Margareta Meyer





# Популяция в течение года

- 10-80 тыс рабочих пчел
  - время жизни варьируется
  - зависит от многих факторов
- сотни трутней
- 1 матка (3-5 лет)



Возраст и количество пчел в зависимости от времени года с зимовкой в сарае. Канада



# Роение

Естественное размножение супер-организма

Июнь-июль

15 мин






Рой забирает с собой мед

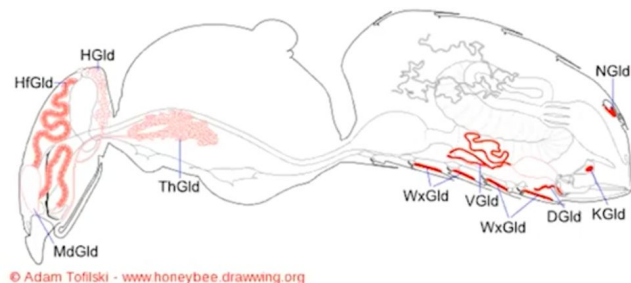
Очень заряженный старт



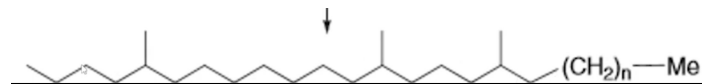
# Общение / запахи

> 50 обнаружено, 10 синтезировано  
спирт, духи, сигареты, пот раздражает

-  свой-чужой
-  матки QMP
  - подавляет репродуктивность работников
  - свита разносит запах по улью
-  полевые (межвидовые метки на цветах на 1ч, путь)
- расплод → “накорми меня”, “собери больше пыльцы”
- работники
  - специализация % полевых
  -  маяк дома
  -  опасности. isopentyl acetate + 24 компонентов



экзокринная система



# Общение танцем

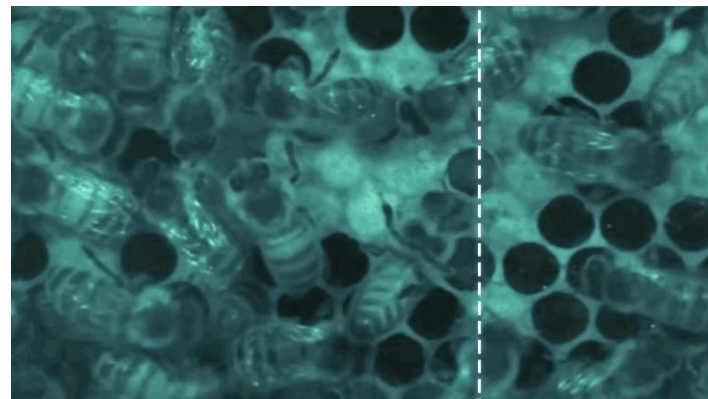
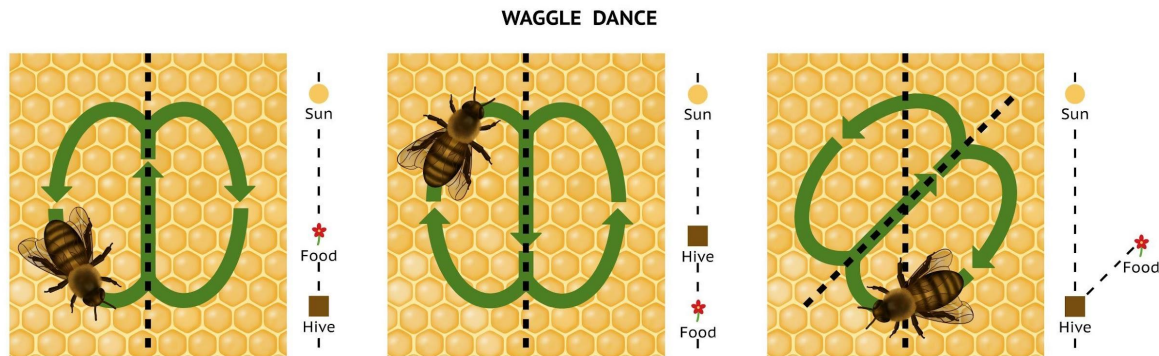
танец

Аристотель, Karl von Frisch

антенны

звук

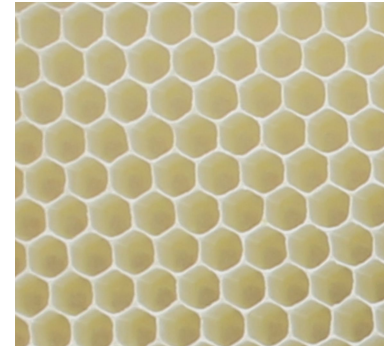
- раздражение пчеловодом
- начало и конец рабочего дня





# Воск

- Выделяется специальными железами
  - развиты к 12 - 18 возрастному дню
  - Строительный материал
  - минимальное количество для 3д сот
  - темнеет / загрязняется по мере использования
- Прекрасно горит



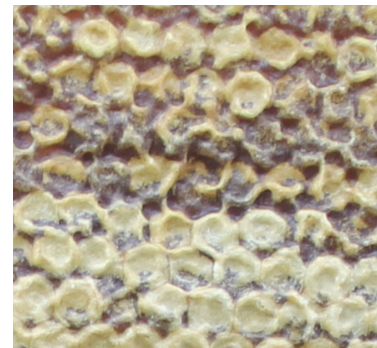
Пчела выделяет воск  
Фото A Spürgin





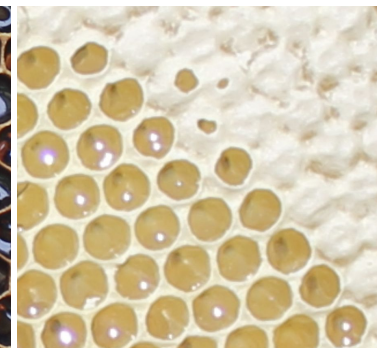
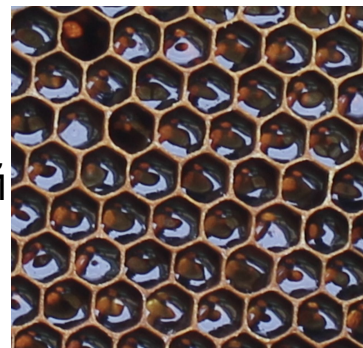
# Мед

- Способ долгосрочного хранения энергии
- Нектар собирается с растений с 50% влажностью
  - может забродить
- Переносится в специальном мешочке
- Вода испаряется < 20% влажность
- Закрывается восковой крышечкой
- до 100 кг / сезон



По источнику

- монофлерный, полифлерный, падевый



# Перга / Пчелиный хлеб

- Утрамбованная пыльца + выделения
- Полита нектаром/медом, молочной кислотой
- Ферментируется в безвоздушной среде
- в основном для расплода
- 40 кг/сезон



# Прополис

- клей / строительный материал
  - заделывать дыры от сквозняка
- смолы (50%), воск (30%)
- эфирные масла (10%), пыльца (5%)
- антибактериальные свойства



камедь

прополис



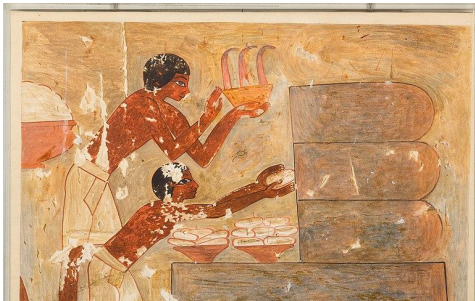




# Пчеловодство

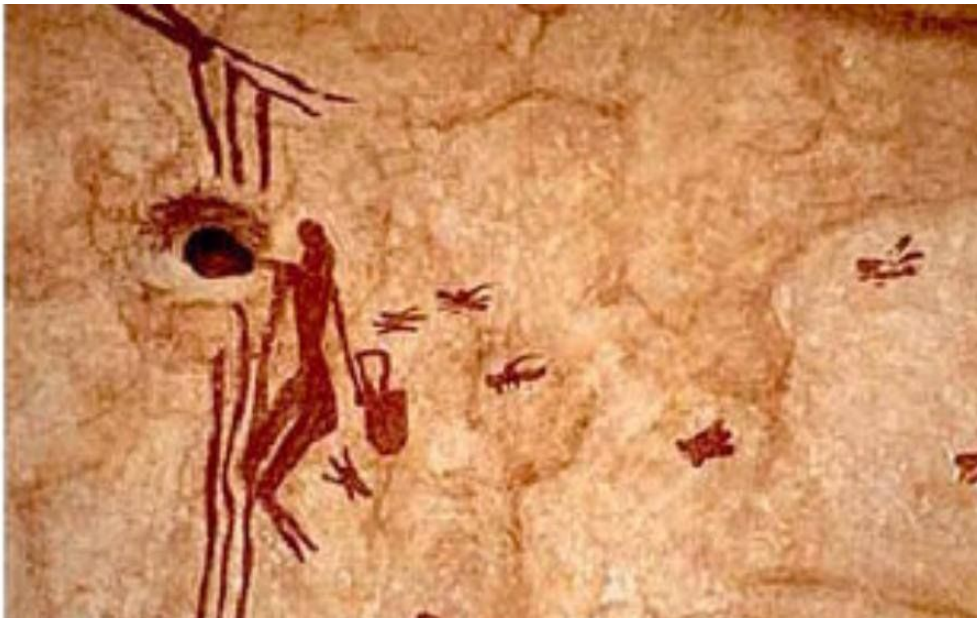


# История





# История пчеловодства / неолит



Араña Caves or the Spider Caves ~ Bicorp Валенсия, Испания  
**-8 тыс до нэ**



Сбор безумного меда. Племена Гурунг, Непал



# История пчеловодства / Древний Египет

”Я беру тебя в жены... и обещаю доставлять тебе ежегодно 12 банок меда”

*Apis mellifera lamarckii*



= bee



= honey



= beekeeper

мед = бит

пчелы = слезы солнца

улья из глины

кочевка пчел по Нилу

фараон - повелитель пчел



Традиционные глиняные



**Сбор меда.** В гробнице дворянина и визира Рекхмире. ~1450 до нэ (ТТ 100)



**Изображение пчелы** (*Apis mellifera lamarckii*) в Гробнице фараона Сети I (KV17), Долина Королей





# История пчеловодства / майя

*Melipona beecheii* - без жала, 1-2 л меда / год

апикультура => мелипоникультура  
рядом с домом ~ 8 колоний

**Ah-Mucen-kab, бог меда у Майя**  
Meliponiculture in Mexico:  
Problems and perspective for  
development  
José Javier G Quezada-Euán,  
William De Jesús May-Itzá, Jorge A  
González-Acereto



пчелиный домик (nahil-kab)  
[https://youtu.be/d\\_pioDxwY58](https://youtu.be/d_pioDxwY58)



дупла и каменные заглушки (hobones)



*Melipona beecheii* <https://www.inaturalist.org/observations/187547762>





# История пчеловодства / Индия

*"Цветок без пчелы подобен молодой деве без любовника"*

Несколько видов (*Apis cerana indica*, *Apis dorsata*, *Apis florea*, *Tetragonula iridipennis*)

Мадхава - апрель, Индра, Кришна, Вишну.

Отсюда мед и mead (медовуха)



*Apis dorsata*

<https://www.inaturalist.org/observations/192283450>



# Сапетки

Корзины

“Telling the bees”





# Бортъ / Русь, Башкирия

5-17 вв

Бортъ - деревья с дуплами  
Надо защищать от медведя  
Трудно залезть на дерево



Бортъ [https://kitapl.org.ru/bort\\_defence\\_rus](https://kitapl.org.ru/bort_defence_rus)





# Колоды

поближе к дому  
вырубка леса  
теплоизоляция  
почти естественная среда



# Японский безрамочный улей

*Apis cerana japonica*

- квадратное сечение
- довольно узкий 22см







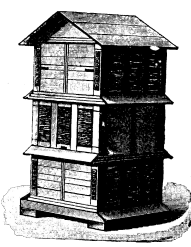
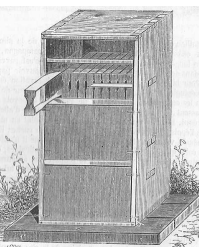


# Разговоры с пчелами

*Telling the bees* — традиция, в которой пчелам рассказывают о важных событиях, включая смерти, рождения, браки, отъезды и возвращения в семье пчеловода.



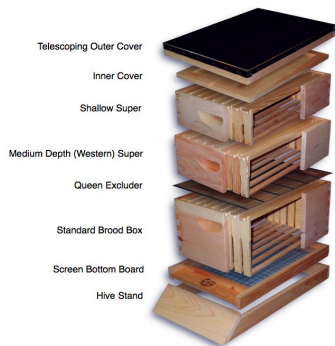
# История



1792- François Huber  
(Швейцария)  
стеклянный наблюдательный  
улей как книга

**1814 - Прокопович** (Российская  
империя/Украина)  
рамочность

1835 - Дзержон  
(Пруссия/Польша)  
8мм пространство  
передвижные рамки



Telescoping Outer Cover  
Inner Cover  
Shallow Super  
Medium Depth (Western) Super  
Queen Excluder  
Standard Brood Box  
Screen Bottom Board  
Hive Stand

**1851 - Лангстрот (US)**

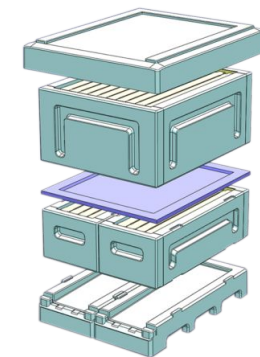
модульные прямоугольные корпуса  
вертикальный модульный улей, 8-10  
рамок  
передвижные рамки, подставка  
разделитель - матка внизу  
6-9 мм - пчелиное пространство над  
рамками  
корпуса надстраиваются (super) сверху

1869 - Amos Root (US) - предприниматель



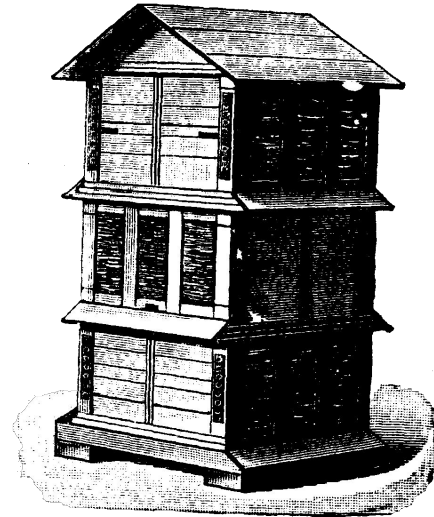
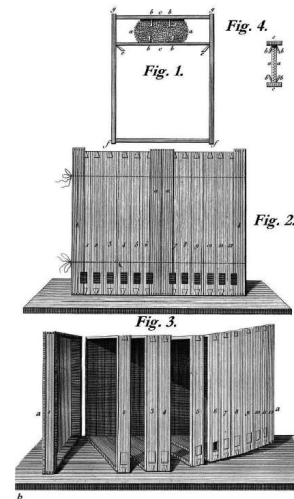
1853 - Берлепш  
**1874 - Дадан (US)**  
большое гнездо 10-12 рамок

**1950 - Варре (FRA)**  
30x30x21 см = квадратные корпуса  
9 рамок  
крыша с юбкой  
корпуса добавляются снизу =  
ротация сот



# История улья

- 1792- François Huber (Швейцария)
  - стеклянный наблюдательный улей как книга
- 1814 - Прокопович (Российская империя/Украина)
  - рамочность
- 1835 - Держон (Пруссия/Польша)
  - 8мм пространство
  - передвижные рамки





# История улья - Лангстрот/Рут

75% в мире  
рамка 48x23 см

- 1851 - Лангстрот (US)
  - модульные прямоугольные корпуса
  - вертикальный модульный улей, 8-10 рамок
  - передвижные рамки, подставка
  - разделитель - матка внизу
  - 6-9 мм - пчелиное пространство над рамками
  - корпуса надстраиваются (super) сверху
- 1869 - Amos Root (US) - предприниматель

Telescoping Outer Cover

Inner Cover

Shallow Super

Medium Depth (Western) Super

Queen Excluder

Standard Brood Box

Screen Bottom Board

Hive Stand



# История улья - Дадан и Варре

- 1853 - Берлепш
- 1874 - Дадан (US)
  - большое гнездо 10-12 рамок
- 1950 - Émile Warré (FRA)
  - 30x30x21 см = квадратные корпуса
  - 9 рамок
  - крыша с юбкой
  - корпуса добавляются снизу = ротация сот



warre hive  
фото timberbee.com



# История улья

- Горизонтальный улей / лежак
- Top bar
- Удобней человеку





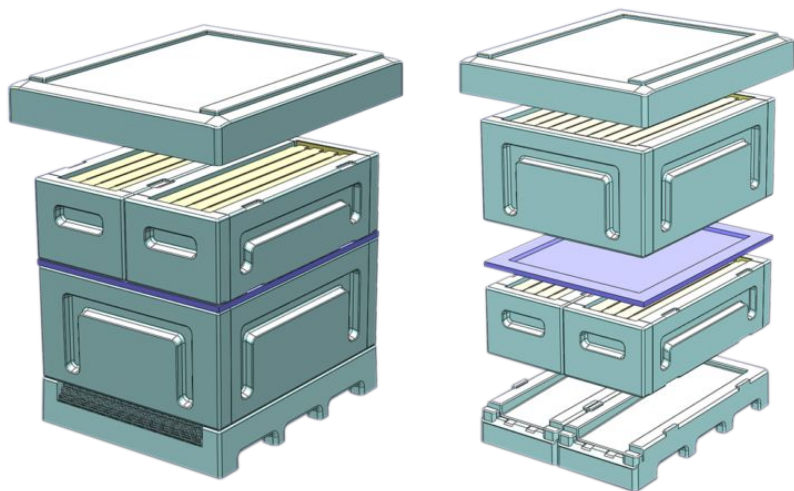
# Апидомик

апитерапия



# Современные улья

- полистирол ~120 EUR
- 10 рамок ~ 25 EUR
- легкий и теплый





# Современные улья

- Утепленная фанера
- ~160 EUR

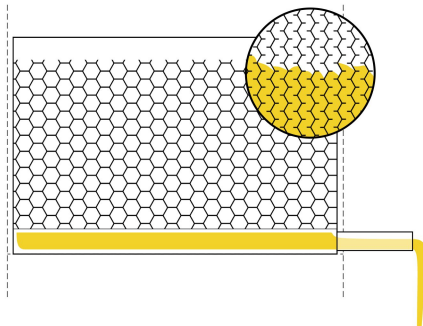


karlwood.ee



# Современные улья

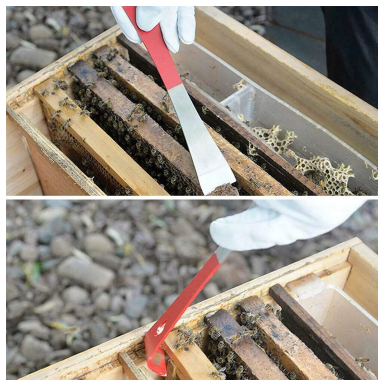
- Flow hive (2015) ~ 780 EUR
  - Австралия
  - пластик
  - кристаллизация
  - мед-по-заказу = нет инспекций
- Китайский аналог ~ 370 EUR





# Инструменты

- Медогонка ~ 600 EUR
- Дымарь ~ 35 EUR
- Костюм



# Промышленное пчеловодство

- Апилифт





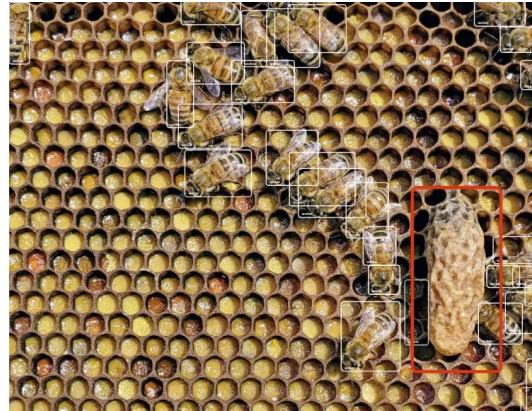
The background is a dense, abstract mosaic of irregular, overlapping shapes. The color palette is primarily composed of muted blue, burnt orange, and black, with some lighter, off-white or beige tones interspersed. The shapes are fragmented and angular, creating a complex, textured visual field. The overall effect is one of chaotic energy and intricate detail.

# Проблемы



# Неизвестность развития семьи

- матки нет?
  - добавить матку
- слабая семья?
  - объединять
- маточники?
  - разделять
- голодают?
  - добавить рамок
- болят?
  - лечить





# Заболевания и опасности пчел

- паразиты
  - **клещ варроа** (*Varroa destructor*)
  - трахейные клещи
- вирусы
  - хронический паралич (CBPV, IAPV, LSV)
  - colony collapse из-за IIV-6?
- бактерии
  - foulbrood (American / European)
  - nosema
- грибок
  - chalkbrood, stonebrood
- **шершни**, мыши, восковая моль, ульевого жука
- отравление пестицидами, замерзший расплод



Разрушение крыльев из-за вируса вызванного варроатозом. фото Philippe Psaila, science.org



# Напряжно

- тяжелые корпуса
- жарко
- в перчатках неудобно
- опасненько
  - дикие пчелы злятся сразу
- ограниченное время  
(днем и в солнечную погоду)
- регулярность инспекций





# Местоположение, время и масштабы

- пасека за городом
  - несколько пасек, далеко ездить
  - в лесу и с **медведями**
- в городе
  - но с шершнями
  - но с соседями и детьми
- не масштабируемо



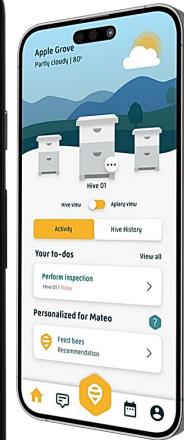
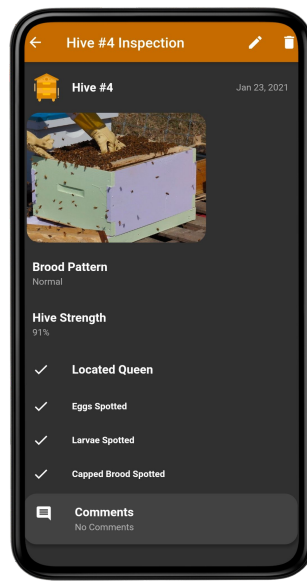


**Существующие решения**



# Приложения

- BeeScanning
- ApiZoom
- HiveTracks
- HiveBloom
- BeeQueenDetector
- apimanager
- apiary book



Дневники. Помогают само-организовываться



# IoT / Аналоговые метрики

- beep.nl - opensource
- beehero.io - 42M \$\$
- broodminder.com
- beelab.se
- intelligenthives.eu
- beehivemonitoring.com
- solutionbee.com
- beehivemonitoringusa.com
- osbeehives.com
- beesage.co

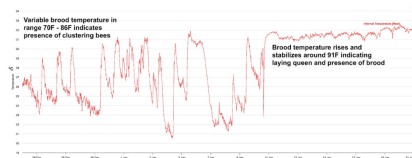


Fig. 2: Using Brood temperature to detect onset of laying queen in late winter/early spring





# IoT / Beehero.io

- Израиль
- Подняли 42 М
- Для фермеров
- точечное опыление (миндаля)
- аналоговые метрики
  - **звук**
  - температура
  - влажность

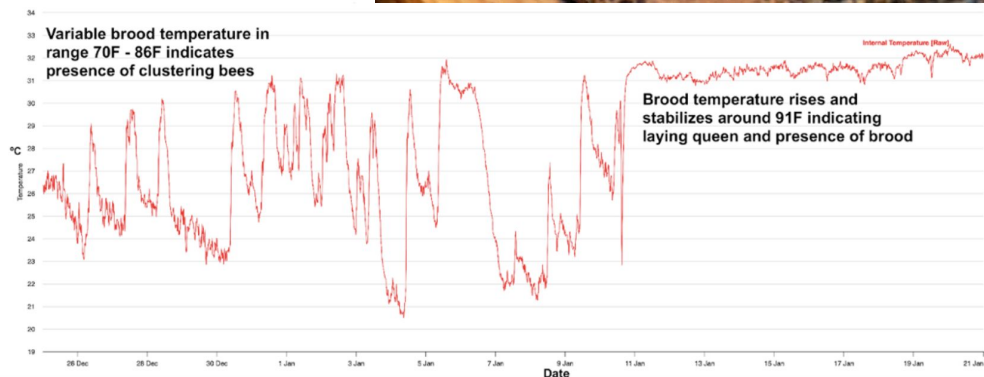


Fig. 2: Using Brood temperature to detect onset of laying queen in late winter/early spring



# BeeMate.buzz

- Австралия
- счетчик пчел 340 eur



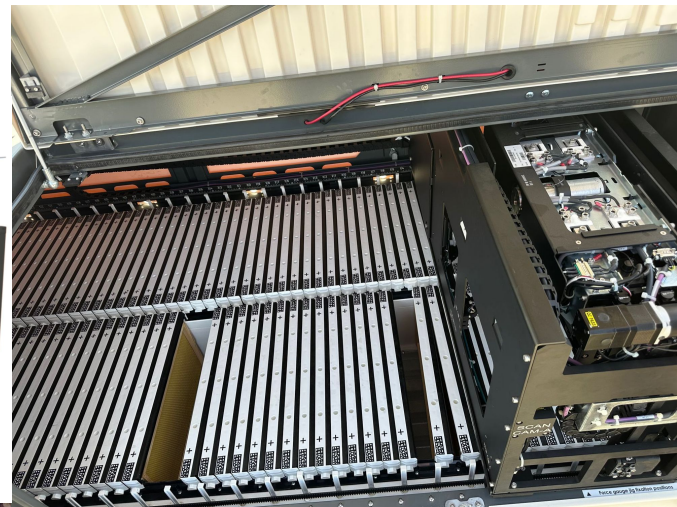
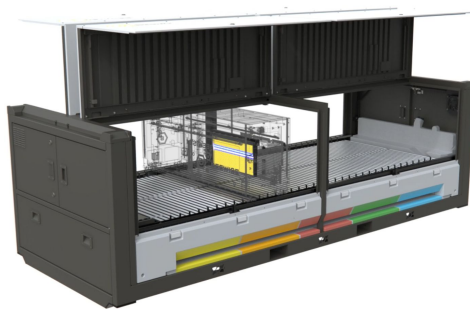
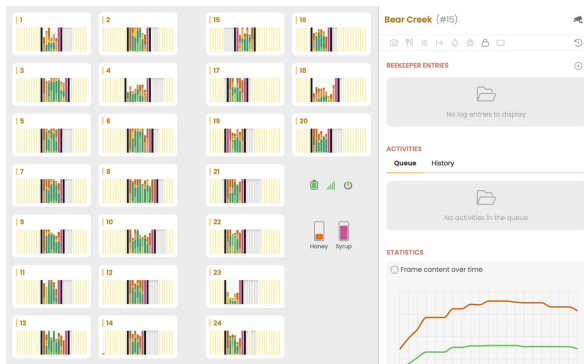
видео со входа beemate.buzz





# Beewise

- Израиль
- 12/24 колонии
- 400\$ / мес
- Подняли 120 М







[Gratheon.com](https://www.Gratheon.com)



# Grattheon.com

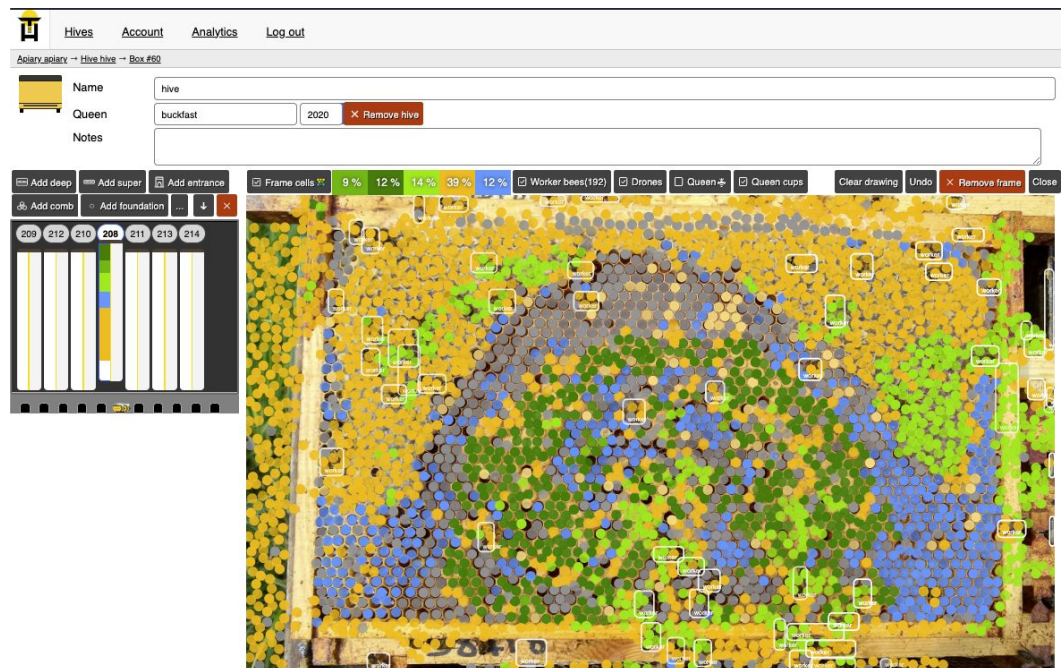
app

Автоматизация пчеловодства 🐝 с AI зрением, роботами 🤖 и спутниковыми данными 📡

- ✓ управление пасекой, ульями, фото рамок
- ✓ Нахождение объектов на фото
  - ✓ Матка и маточники
  - ✓ Пчелы
  - ✓ Соты
  - ✓ Клещи
- ✓ GraphQL API
- ✓ LLM помощник
- ✓ Мобильное приложение

OpenSource + API

- Тренировка на пользовательских данных
- Ручные инспекции
- Аналитика для поиска корреляций
- Оповещения







# Grattheon.com

app

bee  
counter

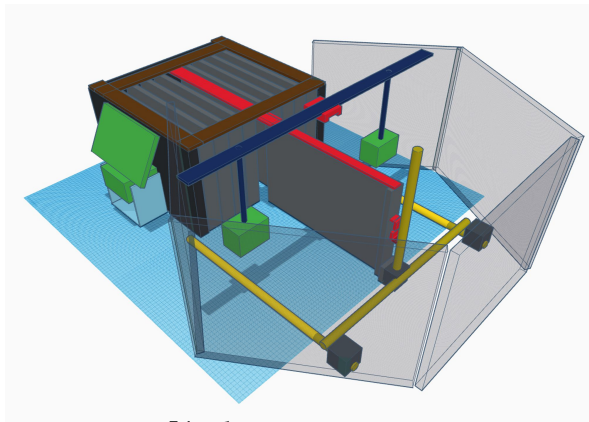
- Видео-мониторинг входа в улей
- Нахождение шершней
- Видео стрим/запись (HLS / MPEG-DASH)
- Обработка на устройстве / в облаке
- Интеграция с умным домом (home assistant / tuya)



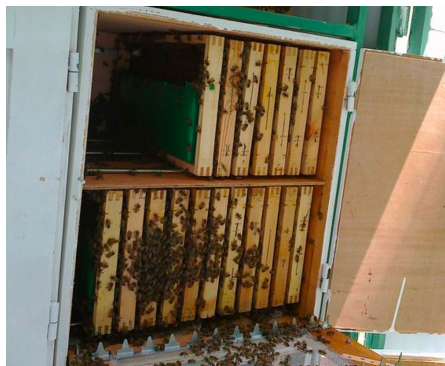
Испробован BeeAlarmed на своем улье



- Стационарный робот на одну семью
- Выдвижение рамок
- Отсылка фото для приложения



3d набросок компонентов



идея, выдвижные в сторону рамки



визуализация leonardo.ai





# Gratheon.com

app

bee  
counter

static robot

moving robot

- Двигающийся робот для инспекции пасеки
- На колесах или на рельсах
- Изъятие рамки с медом
- Перестановка рамки с расплодом между семьями
- Удаленное управление



визуализация leonardo.ai



swisslog рельсовый робот в логистическом центре



The image features a complex, repeating pattern of yellow industrial robotic arms, likely from a factory or manufacturing plant. The robots are arranged in a grid-like fashion, creating a dense and intricate visual texture. The background is a light blue, possibly representing a factory wall or sky. In the center of the image, there is a prominent black rectangular banner with the words "Machine learning" written in a bold, white, sans-serif font. The overall composition is highly detailed and visually busy, emphasizing the theme of automation and artificial intelligence.

# Machine learning



# Регрессия

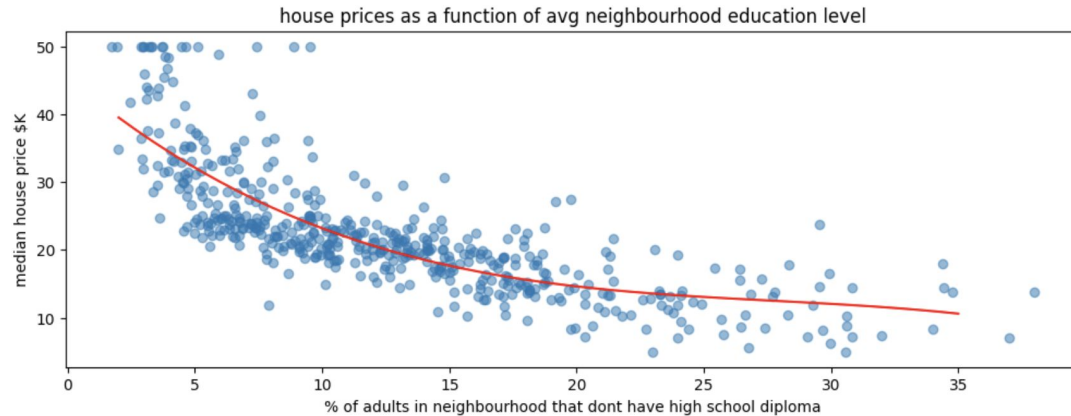
Обобщение, упрощение от набора данных до простой формулы и коэффициентов  
упорядоченность

```
import numpy as np
from sklearn.kernel_ridge import KernelRidge

model = KernelRidge(alpha=1, kernel='poly')
model.fit(boston.data[:,[12]], boston.target.flatten())
predictions = model.predict(np.linspace(2,35)[:,:np.newaxis])

plt.scatter(boston.data[:,12], boston.target, alpha=0.5)
plt.plot(np.linspace(2,35), predictions, c='red')
plt.ylabel('median house price $K')
plt.xlabel('% of adults in neighbourhood that dont have high school diploma')
plt.title('house prices as a function of avg neighbourhood education level')
```

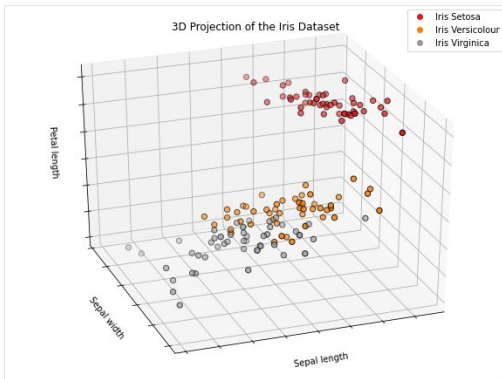
Text(0.5, 1.0, 'house prices as a function of avg neighbourhood education level')



# Классификация

Разделение массива по каким-то выделяемым признакам

классы неупорядоченные  
true/false тоже классы



```
from sklearn import mixture
import numpy as np

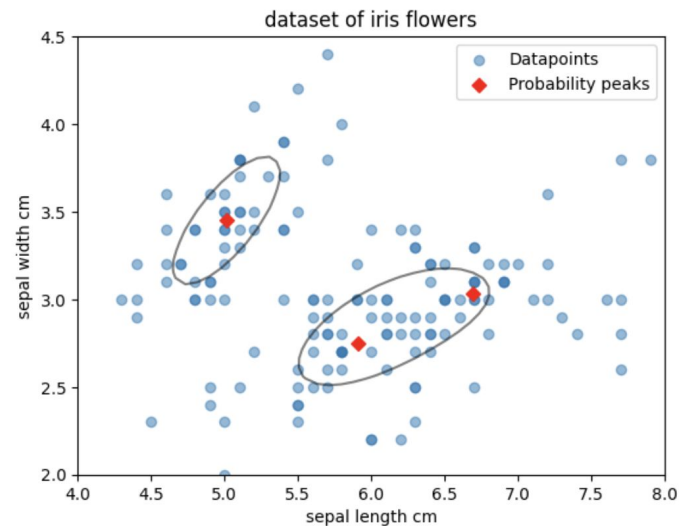
model = mixture.GaussianMixture(n_components=3, covariance_type='full')
model.fit(iris.data[:,0,1])

x,y = np.linspace(4.0, 8.0), np.linspace(2.0, 4.5)
X,Y = np.meshgrid(x,y)
Z = -model.score_samples(np.array([X.ravel(), Y.ravel()]).T).reshape(X.shape)

plt.contour(X,Y,Z, levels=np.logspace(0,10,1), cmap="gray", alpha=0.5)
plt.scatter(iris.data[:,0], iris.data[:,1], alpha=0.5)
plt.scatter(model.means_[0], model.means_[1], marker='D', c='r')

plt.ylabel("sepal width cm")
plt.xlabel("sepal length cm")
plt.title("dataset of iris flowers")
plt.legend(['Datapoints', 'Probability peaks'])
```

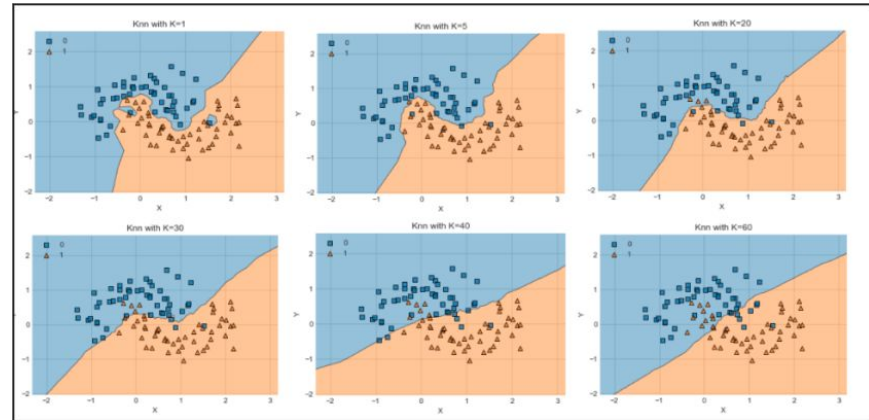
<matplotlib.legend.Legend at 0x13ceb8fa0>





# И много всего другого

Linear Regression  
Logistic regression  
Decision Trees  
Naive Bayes  
Support Vector Machines  
KNN  
Random Forest  
Ridge regression  
Lasso  
Elastic-Net  
Least Angle Regression  
Orthogonal Matching Pursuit  
Bayesian Regression  
Quantile Regression  
Perceptron  
Gradient boosting



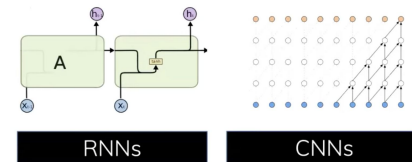
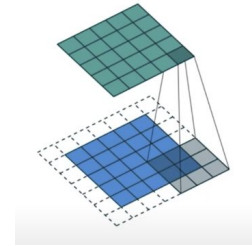
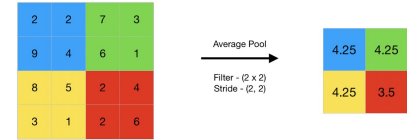
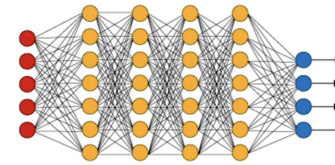
A topographic map of a mountain range, rendered in shades of brown, tan, and blue. The terrain is highly detailed, showing ridges, valleys, and a winding road that snakes through the lower elevations. A black banner with the text "Deep Learning" is centered over the map.

# Deep Learning



# Слои

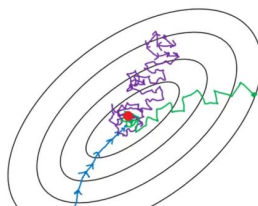
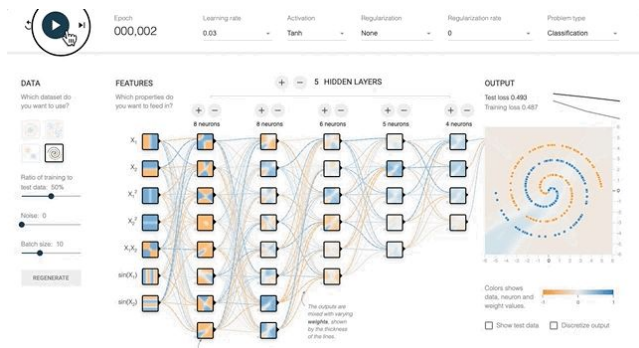
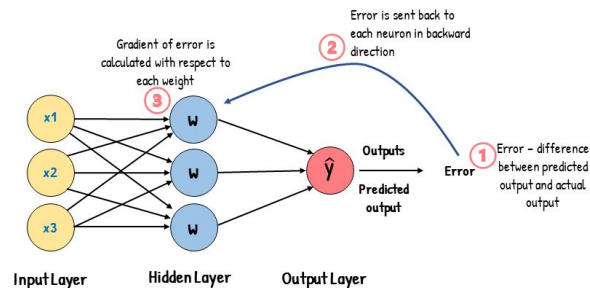
- полносвязные (**dense**, fully-connected)
- подвыборка (**pooling**) - уменьшение размерности
  - max, avg, global avg
- activation (**relu**, **softmax**, sigmoid..)
- сложение (add)
- сверточные (**convolution**, weight) - фильтрация через kernel для выделения фич
  - dilated convolution
  - сдвиг (**stride**) kernel по входу для скорости
  - padding
- нормализация (**batch norm**) - баланс сигнала в сети
- обратная связь (у **feed-forward** нет, recurrent есть)
- регуляризация - обобщение вместо зубрежки
  - dropout - случайное удаление узлов
  - augmentation, random crop, flip, noise, brightness



# Обратное распространение ошибки

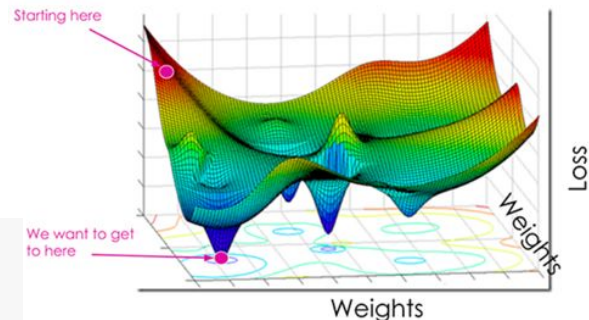
- Поиск наилучших весов через минимизации ошибки (loss) в N-мерном пространстве
  - Производная → Градиент
  - Векторы → Матрицы → Тензоры
- Гипер-параметры.
  - Как спускаться. Эпохи, batching
  - Когда остановиться. Overfitting, validation
- Узлы должны быть обучаемы (дифференцируемы)

## Backpropagation



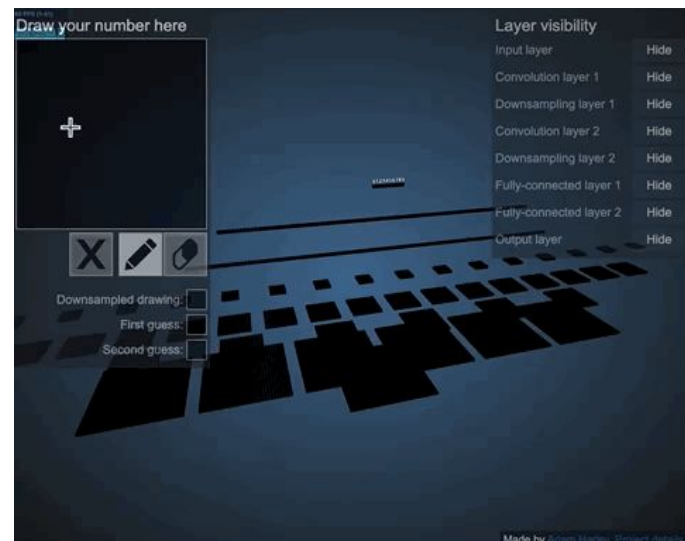
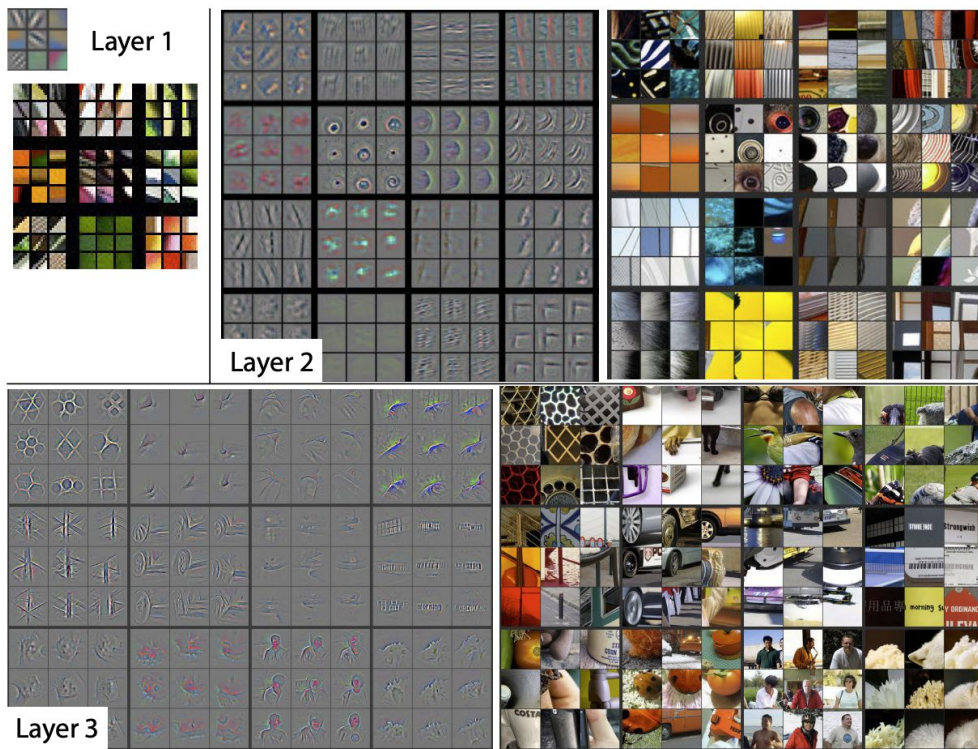
- Batch gradient descent
- Mini-batch gradient Descent
- Stochastic gradient descent

```
class MulConstant(Function):  
    @staticmethod  
    def forward(tensor, constant):  
        return tensor * constant  
  
    @staticmethod  
    def backward(ctx, grad_output):  
        # We return as many input gradients as there were arguments.  
        # Gradients of non-Tensor arguments to forward must be None.  
        return grad_output * ctx.constant, None
```





# Автообучение свойствам



[https://adamharley.com/nn\\_vis/cnn/3d.html](https://adamharley.com/nn_vis/cnn/3d.html)

Visualizing and Understanding Convolutional Networks

M. Zeiler, R.Fergus



# Наборы данных (datasets)

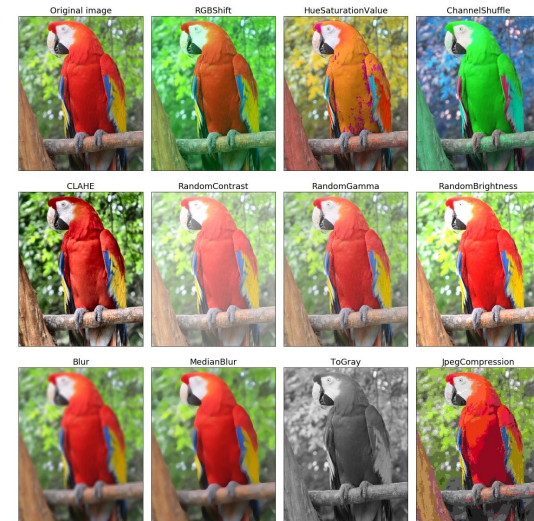
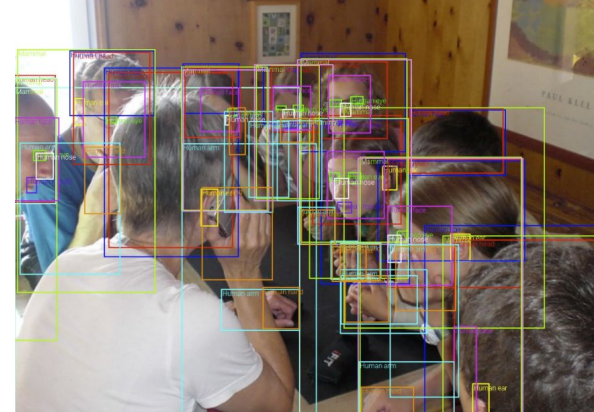
Отличаются:

- модальностью (картинка, текст..)
  - областью (общие vs биология vs электроника)
  - количеством и качеством картинок
  - классами
  - разрешением px
  - разметкой (class, bbox, segmentation, pose)
- PASCAL VOC12 11k
  - MNIST - 60k 28x28px
  - CIFAR-10 - 60k 32x32, 10 классов
  - COCO - 328k
  - iNaturalist - 675k, 5k классов
  - IMAGENET 14M
    - 1M с bbox
  - Cityscapes + 30 классов
  - OpenImages V6 - 9M imgs
  - JFT300M закрытый dataset 303M картинок, 18k классов

<https://paperswithcode.com/datasets>

<https://universe.roboflow.com/>

<https://huggingface.co/datasets>




albumentations lib →





# Прорывные визуальные нейронки

- 1998 LeNet-5. 7 слоев.
- 2012 AlexNet - 8 слоев. 81M параметров
  - 2013 ZFNet ← Clarifai 
  - 2014 VGGNet. 16 слоев. 138M.
  - 2014 GoogLeNet. 22 слоя → Inception v3
- 2015 ResNet. 152 слоя
- 2014 R-CNN → Fast R-CNN → Faster R-CNN
- 2015 YoLo
- 2015 U-Net
- 2017 DenseNet
- 2017 MobileNet
- 2019 EfficientNet



# Классификатор / LeNet-5 (1998)

MNIST dataset  
вход 28x28  
выход 10 классов  
Sigmoid  
5 слоев

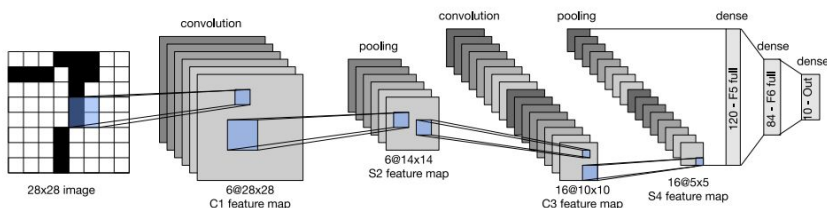


0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9



Gradient-Based Learning Applied to Document Recognition

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner



```
class LeNet(d2l.Classifier): #@save
    """The LeNet-5 model."""
    def __init__(self, lr=0.1, num_classes=10):
        super().__init__()
        self.save_hyperparameters()
        self.net = nn.Sequential(
            nn.LazyConv2d(6, kernel_size=5, padding=2), nn.Sigmoid(),
            nn.AvgPool2d(kernel_size=2, stride=2),
            nn.LazyConv2d(16, kernel_size=5), nn.Sigmoid(),
            nn.AvgPool2d(kernel_size=2, stride=2),
            nn.Flatten(),
            nn.LazyLinear(120), nn.Sigmoid(),
            nn.LazyLinear(84), nn.Sigmoid(),
            nn.LazyLinear(num_classes))
```





# Классификатор / AlexNet (2012)

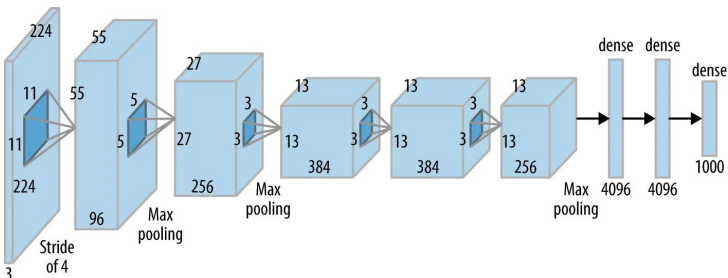
вход - 224x224px

выход - 1000 классов

ReLU

ImageNet dataset

8 слоев



ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky  
University of Toronto  
kriz2@cs.utoronto.ca

Ilya Sutskever  
University of Toronto  
ilya@cs.utoronto.ca

Geoffrey E. Hinton  
University of Toronto  
hinton@cs.utoronto.ca

```
class AlexNet(d2l.Classifier):
    def __init__(self, lr=0.1, num_classes=10):
        super().__init__()
        self.save_hyperparameters()
        self.net = nn.Sequential(
            nn.LazyConv2d(96, kernel_size=11, stride=4, padding=1),
            nn.ReLU(), nn.MaxPool2d(kernel_size=3, stride=2),
            nn.LazyConv2d(256, kernel_size=5, padding=2), nn.ReLU(),
            nn.MaxPool2d(kernel_size=3, stride=2),
            nn.LazyConv2d(384, kernel_size=3, padding=1), nn.ReLU(),
            nn.LazyConv2d(384, kernel_size=3, padding=1), nn.ReLU(),
            nn.LazyConv2d(256, kernel_size=3, padding=1), nn.ReLU(),
            nn.MaxPool2d(kernel_size=3, stride=2), nn.Flatten(),
            nn.LazyLinear(4096), nn.ReLU(), nn.Dropout(p=0.5),
            nn.LazyLinear(4096), nn.ReLU(), nn.Dropout(p=0.5),
            nn.LazyLinear(num_classes))
        self.net.apply(d2l.init_cnn)
```

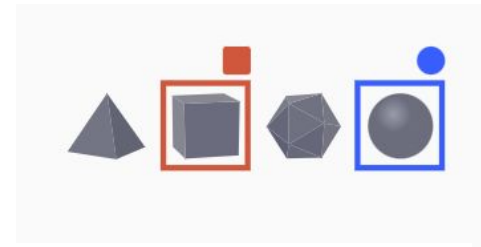
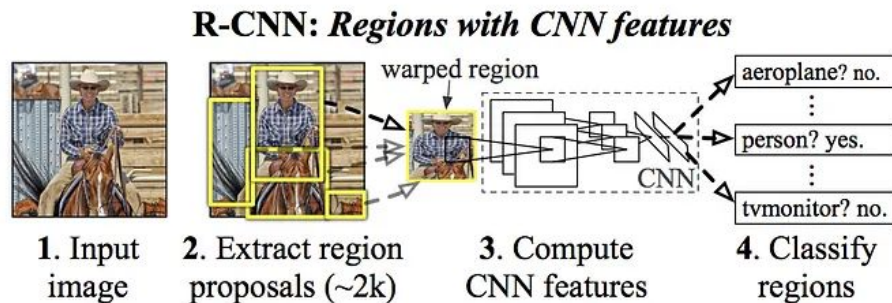


# Детектор / Region CNN (2014)

Найди интересные регионы (2000), для каждого региона, классифицируй, верни XY прямоугольники и вероятность

- очень медленно (47 сек на GPU)
- искажение регионов
- ТОЧНО но СЛОЖНО (много шагов)

Region-based-CNN → Fast R-CNN → Faster R-CNN.  
DensePose → Detectron2



Rich feature hierarchies for accurate object detection and semantic segmentation  
Tech report (v5)

Ross Girshick Jeff Donahue Trevor Darrell Jitendra Malik  
UC Berkeley  
{rgb, jdonahue, trevor, malik}@eecs.berkeley.edu





# Детектор / YoLo (2015)



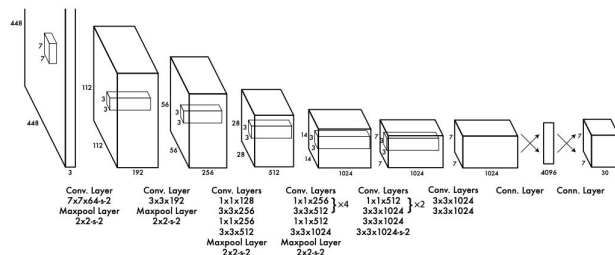
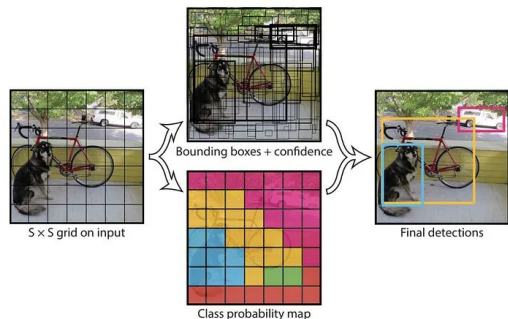
**You Only Look Once:**  
Unified, Real-Time Object Detection

Joseph Redmon\*, Santosh Divvala<sup>†</sup>, Ross Girshick<sup>‡</sup>, Ali Farhadi<sup>†</sup>  
University of Washington\*, Allen Institute for AI<sup>†</sup>, Facebook AI Research<sup>‡</sup>

- Разбивает картинку на части. В каждой части предсказывает координаты, класс и уверенность
- Быстрый. Плохо видит мелкие объекты
- Своя cost-функция и объединения bbox iou
- Больше разрешение 448x448, 24 conv слоя

⇒ DarkNet, Yolo9k, YoLo v5, pp-yolo, YoLo v8;  
SSD как альтернатива

Real-Time Detectors	Train	mAP	FPS
100Hz DPM [31]	2007	16.0	100
30Hz DPM [31]	2007	26.1	30
Fast YOLO	2007+2012	52.7	<b>155</b>
YOLO	2007+2012	<b>63.4</b>	45
<b>Less Than Real-Time</b>			
Fastest DPM [38]	2007	30.4	15
R-CNN Minus R [20]	2007	53.5	6
Fast R-CNN [14]	2007+2012	70.0	0.5
Faster R-CNN VGG-16[28]	2007+2012	73.2	7
Faster R-CNN ZF [28]	2007+2012	62.1	18
YOLO VGG-16	2007+2012	66.4	21



**Table 1: Real-Time Systems on PASCAL VOC 2007. Compar-**

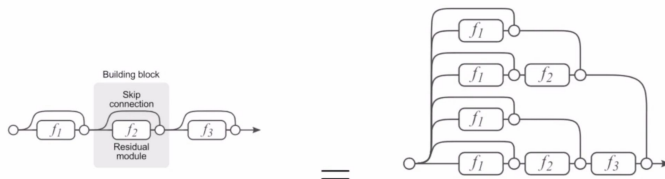
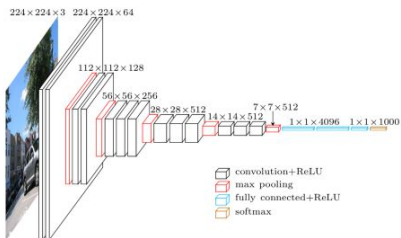
	YOLO							YOLOv2
batch norm?	✓	✓	✓	✓	✓	✓	✓	✓
hi-res classifier?		✓						✓
convolutional?			✓	✓	✓	✓	✓	✓
anchor boxes?			✓					✓
new network?				✓	✓	✓	✓	✓
dimension priors?					✓	✓	✓	✓
location prediction?					✓	✓	✓	✓
passthrough?						✓	✓	✓
multi-scale?							✓	✓
hi-res detector?								✓
VOC2007 mAP	63.4	65.8	69.5	69.2	69.6	74.4	75.4	76.8
								<b>78.6</b>



# Детектор / VGG → Residual Net (2015)



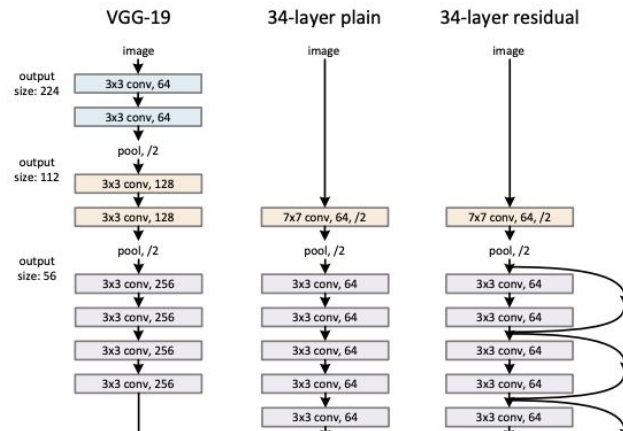
- Visual Geometry Group - 16 слоев. 138M параметров. 90% в dense слоях
- классический подход достиг своих пределов точности



## Deep Residual Learning for Image Recognition

Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun  
 Microsoft Research  
 {kahe, v-xiangz, v-shren, jiansun}@microsoft.com

- ResNet - 34 слоя. 21M параметров. 97% - convolution, 2% - dense
  - **промежуточные связи!**
  - пробовали 50 слоев, 150, 1200
  - Еще точнее, но нужен batch normalization
  - Надежные. Можно выбрасывать блоки
- Улучшения - ResNeXt, SENet, Boosted ResNet, RetinaNet, DenseNet, BiT
  - блоки перепрыгивают еще дальше

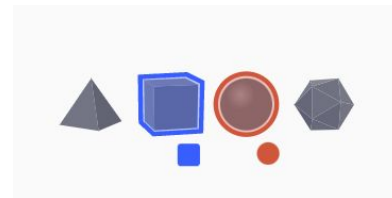




# Сегментация / U-Net (2015)

- encoded + decoder
- контур объекта, близкие объекты со слабыми границами
- Сегментация бывает разной
  - semantic, boundary, semantic instance
- upscale low-res фотографии
- diffusion models → Dall-E

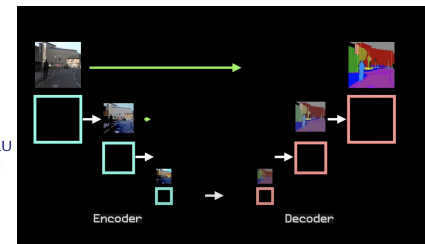
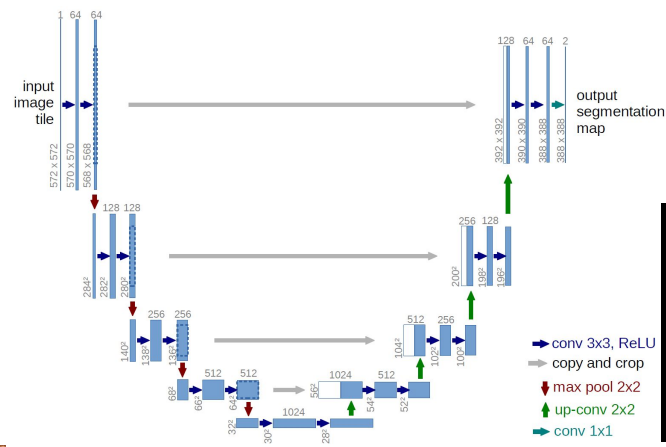
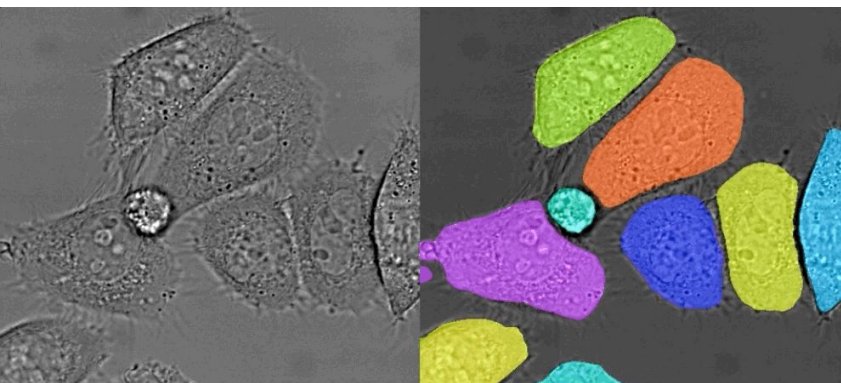
Ternaus, Mask R-CNN, DeepLab



U-Net: Convolutional Networks for Biomedical Image Segmentation

Olaf Ronneberger, Philipp Fischer, and Thomas Brox

Computer Science Department and BIOS Centre for Biological Signalling Studies,  
University of Freiburg, Germany





# Научные работы

# BeeAlarmed

- работа магистра, Германия
- только зеленый фон :(
- detector + classifier
  - судя по коду простой CNN
- 👍 opensource
- пчелы
  - с пыльцой
  - вентилирующие
  - с клещами
- ОСЫ
- LoRaWan



```
def build_pollen_branch(input_shape):
    """! Creates the branch that detects pollen packets
    """
    tmp_layer= layers.experimental.preprocessing.Rescaling(1./255)(input_shape)
    tmp_layer= Conv2D(32, (4, 4), padding="valid")(tmp_layer)
    tmp_layer= Activation("relu")(tmp_layer)
    tmp_layer= BatchNormalization(axis=CHAN_DIM)(tmp_layer)
    tmp_layer= MaxPooling2D(pool_size=(2, 2))(tmp_layer)

    tmp_layer= Conv2D(16, (3, 3), padding="valid")(tmp_layer)
    tmp_layer= Activation("relu")(tmp_layer)
    tmp_layer= MaxPooling2D(pool_size=(2, 2))(tmp_layer)

    tmp_layer= Flatten()(tmp_layer)
    tmp_layer= BatchNormalization(axis=CHAN_DIM)(tmp_layer)
    tmp_layer= Dense(1)(tmp_layer)
    tmp_layer= Activation("sigmoid", name="pollen_output")(tmp_layer)

    return tmp_layer
```





# LabelBee

US / Puerto Rico

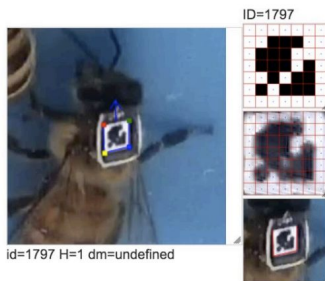
индивидуальный трекинг пчел на входе улья с QR кодами

Part Affinity Fields pose estimation

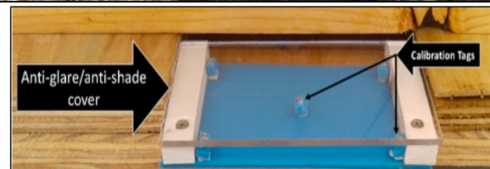
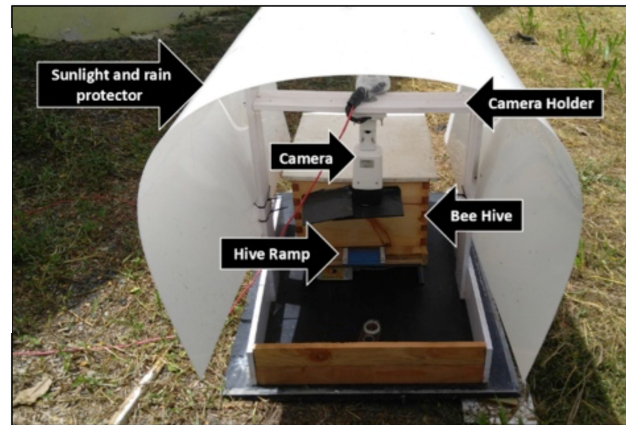
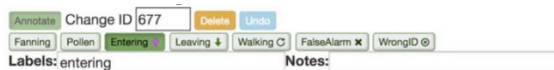
👍 opensource

2.6 TB / 8MP cam / 10 семей

см OpenPose для людей

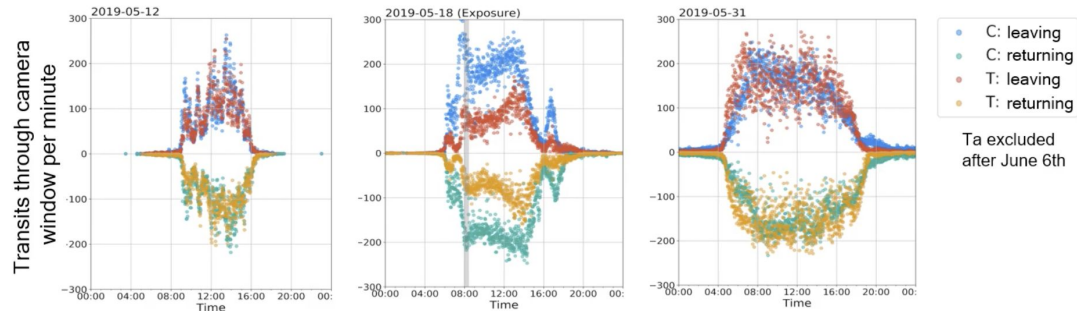


id=1797 H=1 dm=undefined

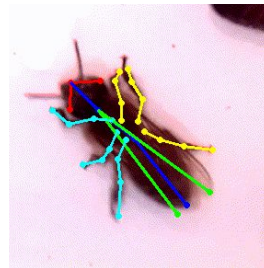
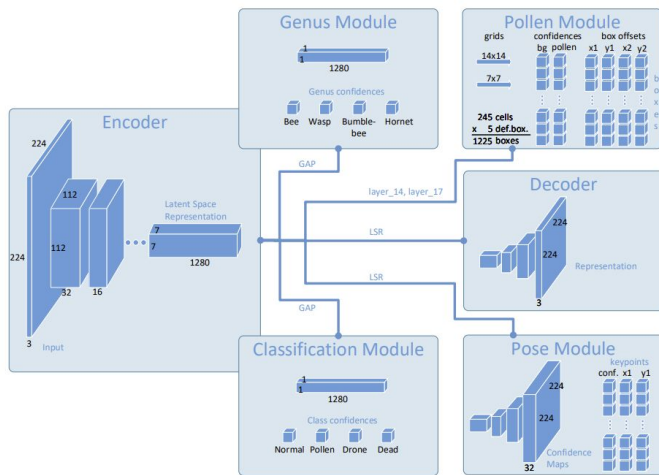


# Apic.ai / DeepBees

- подсчет пчел на входе
- MultiNet - несколько модулей
  - классификация видов (шершни)
  - классификация пчел
  - распознавание пыльцы -> MobileNet v2
  - поза
- фокус на анализе опыления
- 14 ТВ видео, 49 семей
- открыт только dataset



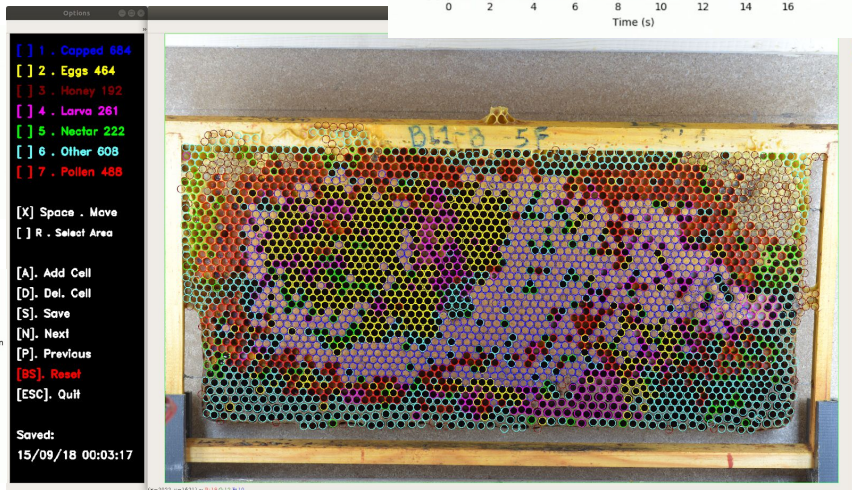
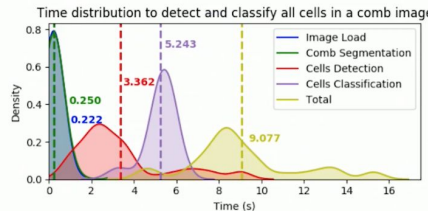
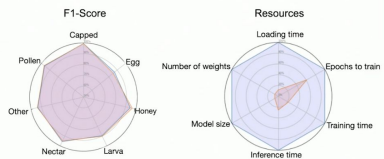
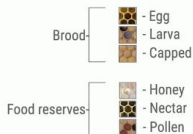
активность пчел до и после лечения <https://apic.ai/research.html>



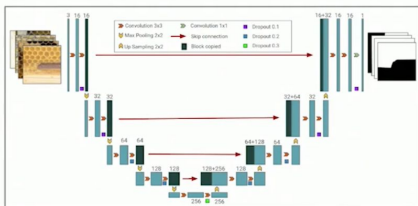


# DeepBee - Thiago, Metz et al.

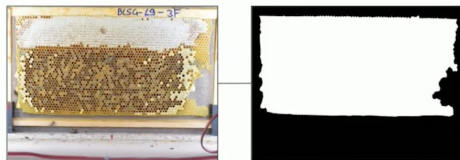
- contrast filters
- segmentation → U-net
- classification → MobileNet
- pro camera
- 2k frames
- ест 2 гб RAM
- ~ 5 сек
- 👍 opensource



CNN architecture - U-Net based

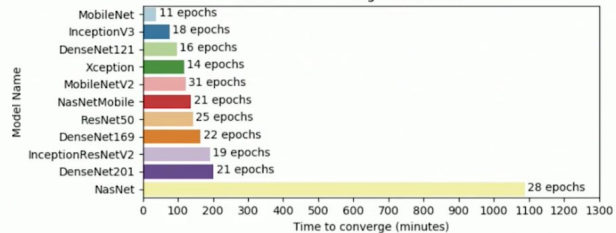


Dataset Creation



- ✗ DenseNet121
- ✗ DenseNet169
- ✓ DenseNet201
- ✓ InceptionResNetV2
- ✓ InceptionV3
- ✓ MobileNet
- ✓ MobileNetV2
- ✗ NasNet
- ✓ NasNetMobile
- ✗ ResNet50
- ✗ Xception

Time for the convergence of each model





# Bozek и др - Honeybee tracking I

трекинг пчел на рамке  
трудности  
быстрое движение  
препятствия видимости (occlusion)

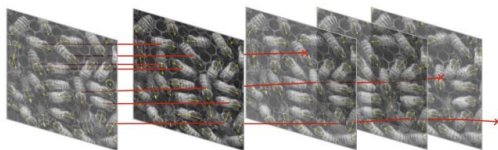
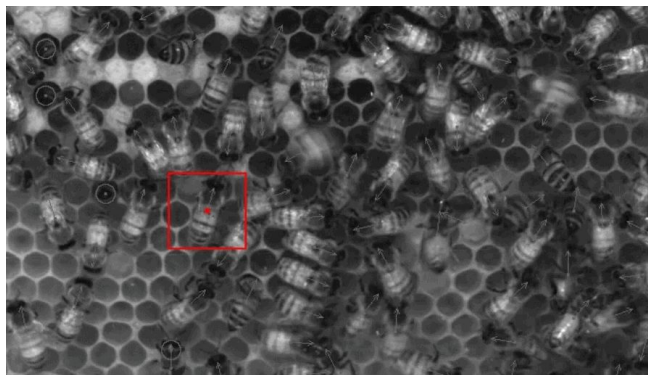
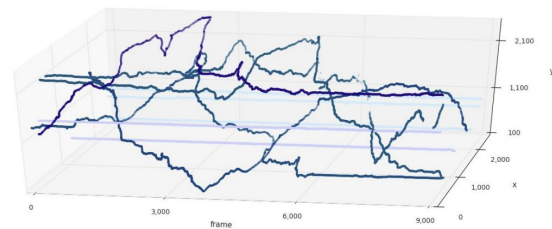
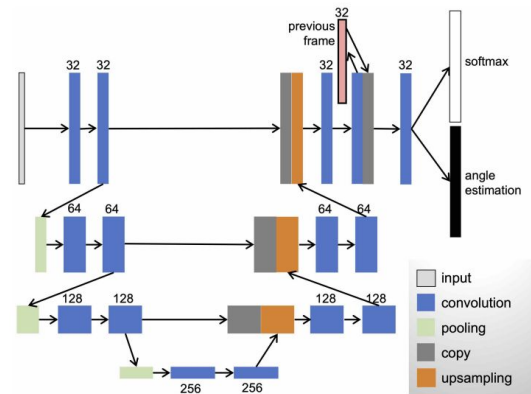


FIG. 2: Object detections are joined into short track fragments using a simple distance metric. We join the



траектория пчел на рамке и во времени

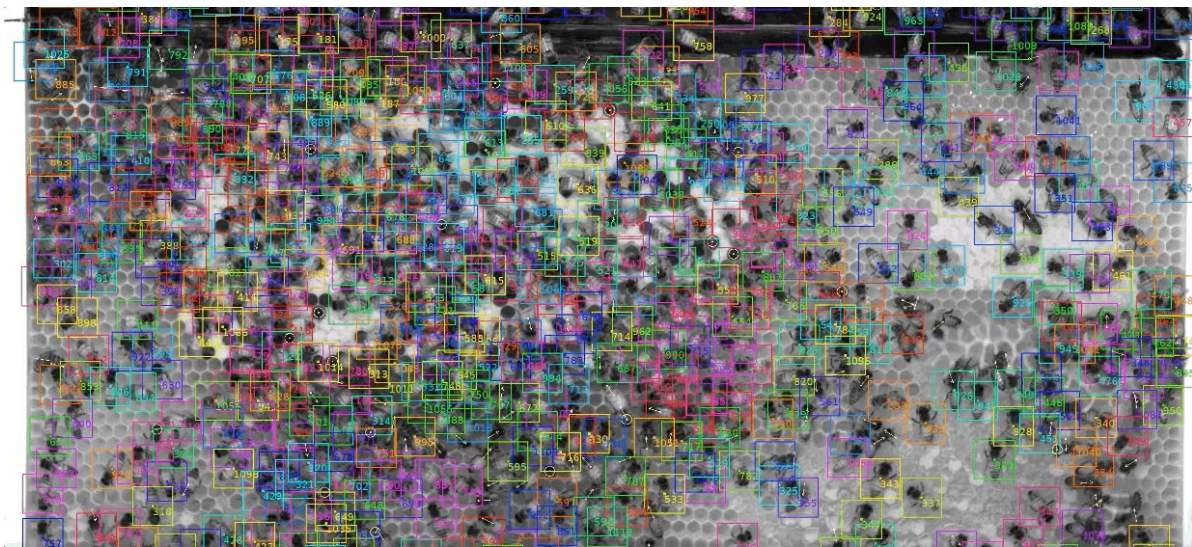
**Pixel personality for dense object tracking in a 2D honeybee hive**

Katarzyna Bozek, Laetitia Hebert, Yoann Portugal, Alexander S. Mikheyev & Greg J. Stephens



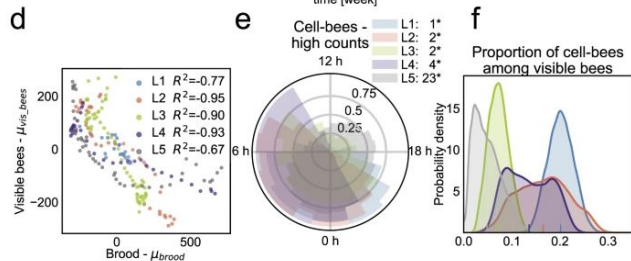
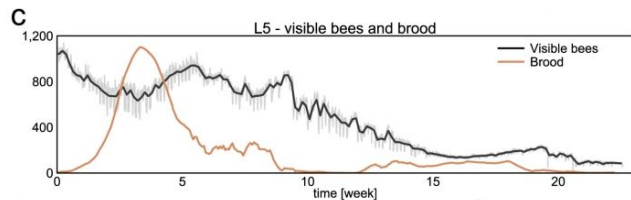
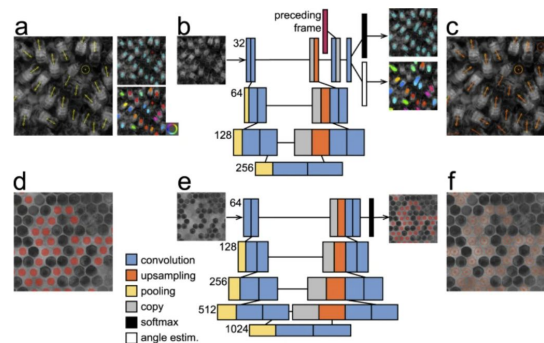
# Bozek и др - Honeybee tracking II

👍 opensource + dataset  
трекинг всей семьи



Towards dense object tracking in a 2D honeybee hive

Katarzyna Bozek, Laetitia Hebert, Yoann Portugal, Alexander S. Mikheyev & Greg J. Stephens



# BeeNet

Микс двух популярных архитектур

кода нет

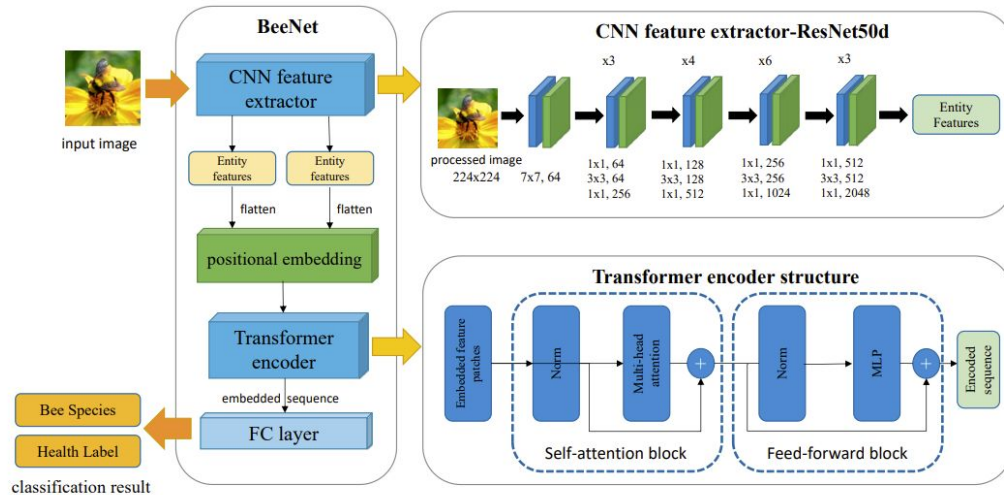


Table 1: Best CNN models and Vision Transformer against 3 bee data sets

Models/Data sets	Bee identification	Varroa detection	Pollen detection
ResNest	85.08	84.82	96.48
EfficientNet NoisyStudent	84.63	84.44	96.48
Vision Transformer 16x16	90.09	88.46	98.74
Vision Transformer 8x8	90.78	93.24	98.88
BeeNet	<b>92.45</b>	<b>94.50</b>	<b>99.18</b>



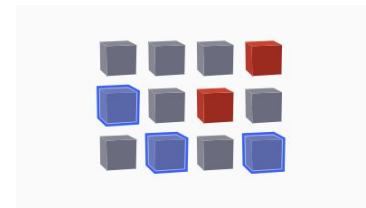


A detailed oil painting of a dense thicket of tree branches in autumn. The leaves are rendered in various shades of orange, yellow, and red, with some still showing green. The branches are dark brown and intricate, creating a complex web of lines. The background is a soft, textured blue, suggesting a clear sky. The overall style is impressionistic with visible brushstrokes.

**СМЫСЛ**

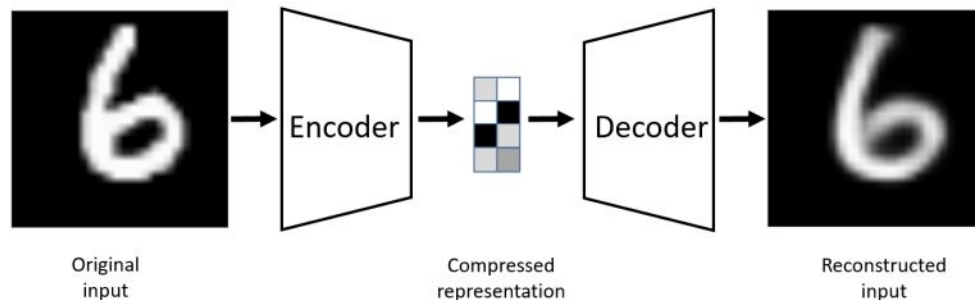
# Автокодировщик и Embeddings

- Autoencoder - обучается передачи ввода на вывод с уменьшением размерности в процессе
- В середине получаем **embedding** = latent vector)  
- float-векторов в **latent space** (512-мерный например)
- CNN, GPT, BERT - на последнем слое embedding → transfer learning
- В чистом виде генерируются предобученными моделями  
- word2vec, sentence2vec, doc2vec, glove, USE, CLIP, ImageBind



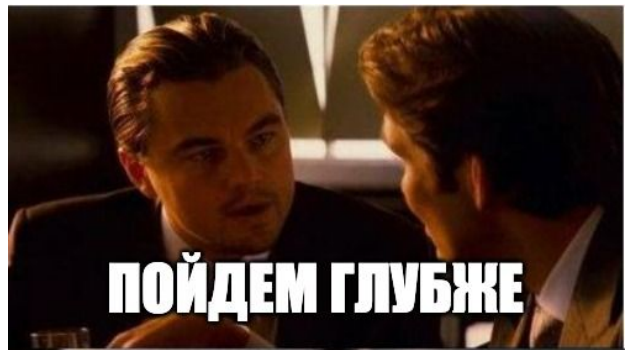
Use case:

- Компрессия при передаче по сети
- Очистка картинок от шума
- Увеличение frame rate
- Перевод между языками
- Кластеризация
- Семантический поиск
- Рекомендации
- Генерация картинок (из шума)



# Прорывные смысловые нейронки

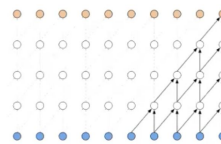
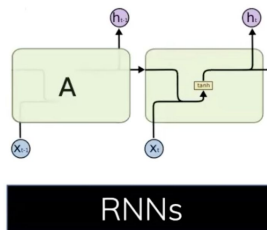
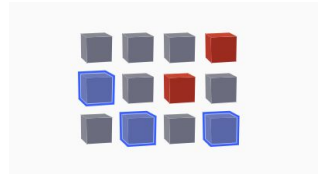
- 1986 Autoencoders
- 2003 Neural Probabilistic Language Model
- 2008 Word position
- 2013 Word2vec - word embeddings
- 2014 RNN → LSTM, GRU
- 2014 GAN. Две нейронки генерят картинки.  
Self-play
- 2015 Attention (Luong et al.)
- 2018 ELMo
- 2018 **Transformers** → BERT, GPT, PaLM
- 2020 Visual Transformers (ViT) → DINO
- 2021 S4
- 2021 RAG, CLIP





# Recurrent NN (2014)

- Use case
  - перевод текста (**google translate** - LSTM)
  - распознавание звука, временные ряды
- RNN = последовательные слои
  - есть скрытое состояние = память
- дорого тренировать
- ⇒ Gated Recurrent Unit
- Long Short-Term Memory + **attention**



Deep Visual-Semantic Alignments for Generating Image Descriptions

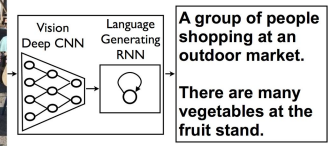
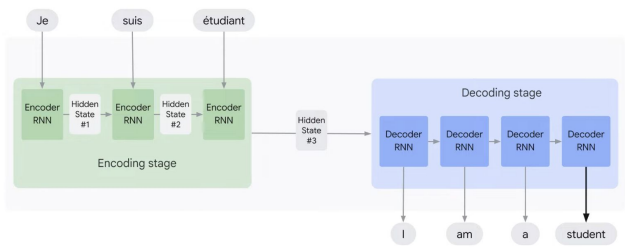
Andrej Karpathy    Li Fei-Fei  
 Department of Computer Science, Stanford University  
 {karpathy, feifeili}@cs.stanford.edu

Show and Tell: A Neural Image Caption Generator

Oriol Vinyals (Google)    Alexander Toshev (Google)    Samy Bengio (Google)    Dumitru Erhan (Google)  
 vinyals@google.com    toshev@google.com    bengio@google.com    dimitru@google.com

DEEP CAPTIONING WITH MULTIMODAL RECURRENT NEURAL NETWORKS (M-RNN)

Justin Mao (University of California, Los Angeles; Baidu Research)  
 mjhust@ucla.edu  
 Wei Xu & Yi Yang & Jiang Wang & Zhiheng Huang (Baidu Research)  
 {wei.xu, yangyi105, wangjiang03, huangzhiheng}@baidu.com  
 Alan Yuille (University of California, Los Angeles)  
 yuille@stat.ucla.edu

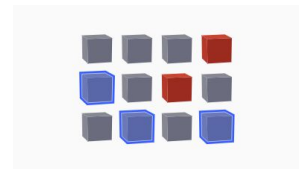
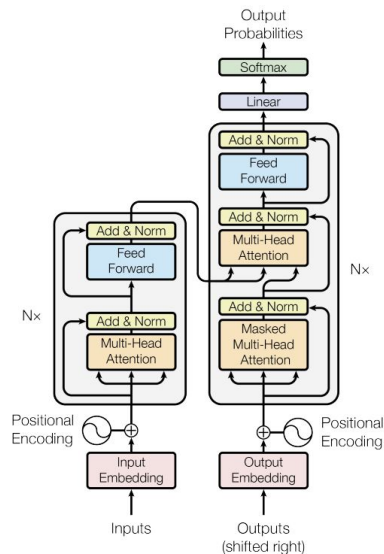


A group of people shopping at an outdoor market.  
 There are many vegetables at the fruit stand.



# Transformers (2017)

- Use case - перевода текста
- Механизм **внимания**
  - Абстрактизация нейронок = CPU Von Newman'a
  - Многоголовое внимание = многозначимость
  - Ест много памяти  $O(N^2)$  в зависимости от входа
- Оптимизируем/обучаем
  - Остаточные (residual) связи  $\Rightarrow$  backpropagation достает до первого слоя
  - Эффективно тренируется на GPU (параллелизация)
- <http://nlp.seas.harvard.edu/annotated-transformer/>



Attention Is All You Need

- Ashish Vaswani\*  
Google Brain  
vaswani@google.com
- Noam Shazeer\*  
Google Brain  
noam@google.com
- Niki Parmar\*  
Google Research  
niki@google.com
- Jakob Uszkoreit\*  
Google Research  
uszko@google.com
- Llion Jones\*  
Google Research  
llion@google.com
- Aidan N. Gomez\*<sup>†</sup>  
University of Toronto  
aidan@cs.toronto.edu
- Lukasz Kaiser\*  
Google Brain  
lukaszkaizer@google.com
- Illia Polosukhin\*<sup>†</sup>  
illia.polosukhin@gmail.com

### 2. Positional Encoding

Positional Input Embeddings

Positional Encoding

Time Step: 1, 2, 3, 4

$$PE(pos, 2i + 1) = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right)$$

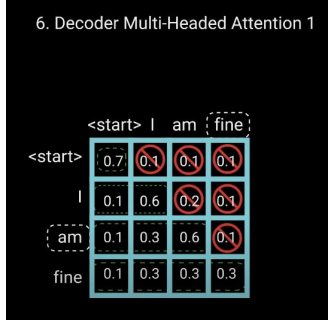
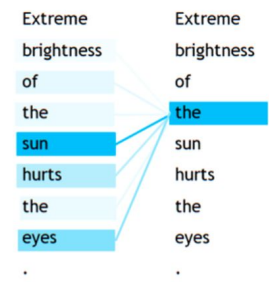
$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right)$$

### 3 - 4. Encoder Layer

### 3. Multi-headed Attention

#### 3.1. Self-Attention

	Hi	how	are	you
Hi	98	27	10	12
how	27	89	31	67
are	10	31	91	54
you	12	67	54	92



# Наследие трансформеров

- Bidirectional Encoder Representations from Transformers (BERT)
- Generative Pre-trained Transformer (GPT)
- скорость генерации vs контекст vs качество воспоминания

картинки из

Transformer models: an introduction and catalog. Amatriain et al

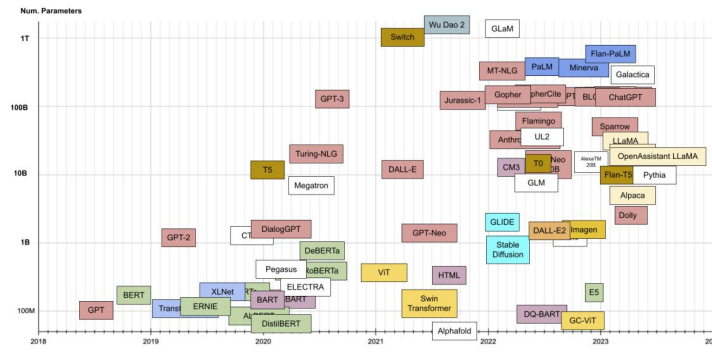
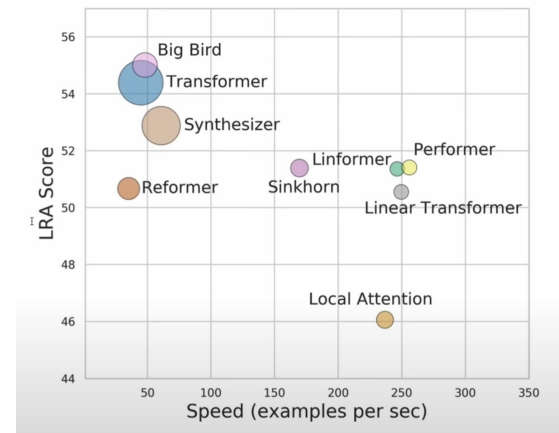
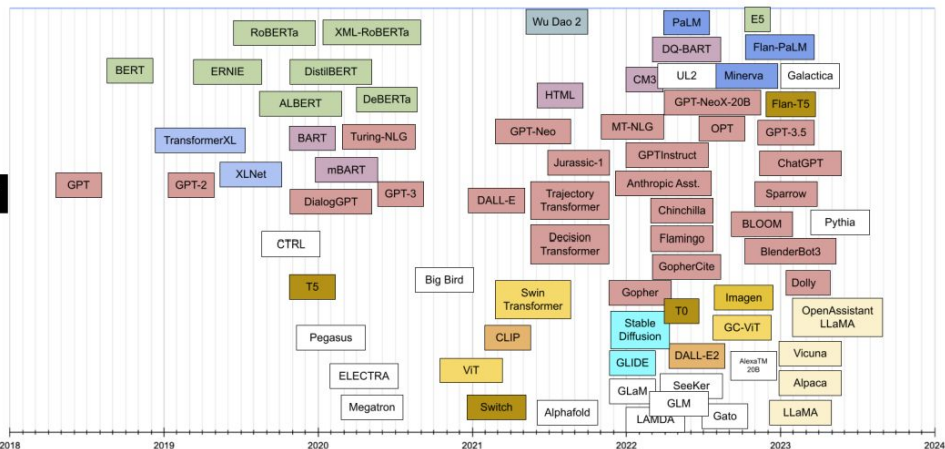


Figure 7: Transformer timeline. On the vertical axis, number of parameters. Colors describe Transformer family.

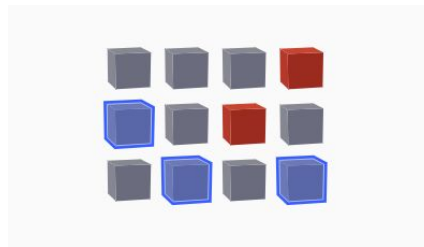




# Visual Transformer (2021)

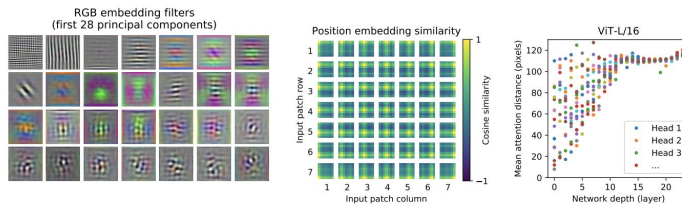
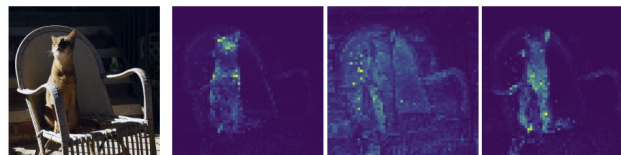
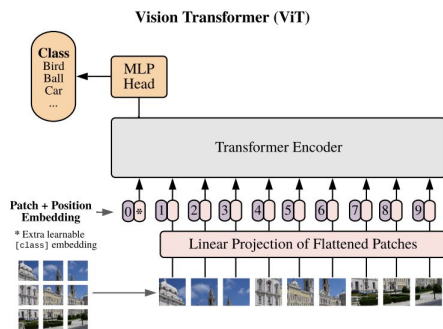
- Замена CNN на глобальные маски внимания
- 12-32 слоев, 12-16 головок внимания
- Превосходит ResNet152 в точности и скорости обучения при больших данных (ViT vs BigTransfer)
- делит исходную картинку на куски (16x16px patch = 196 кусков)
- добавляется позиционный embedding
- между patchами возникает матрица внимания

⇒ DeiT, CCT, CaiT, PVT, RadixNet, T2T-ViT, EfficientViT, MaxViT, DINOv2



AN IMAGE IS WORTH 16X16 WORDS:  
TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE

Alexey Dosovitskiy<sup>1,2</sup>, Lucas Beyer<sup>1</sup>, Alexander Kolesnikov<sup>1</sup>, Dirk Weissenborn<sup>1</sup>,  
Xiaohua Zhai<sup>1</sup>, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer,  
Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby<sup>1,2</sup>  
<sup>1</sup>equal technical contribution, <sup>2</sup>equal advising  
Google Research, Brain Team



Emerging Properties in Self-Supervised Vision Transformers

Mathilde Caron<sup>1,2</sup>, Hugo Touvron<sup>1,3</sup>, Ishan Misra<sup>1</sup>, Hervé Jegou<sup>1</sup>,  
Julien Mairal<sup>2</sup>, Piotr Bojanowski<sup>1</sup>, Armand Joulin<sup>1</sup>  
<sup>1</sup> Facebook AI Research, <sup>2</sup> Inria, <sup>3</sup> Sorbonne University



The background is a complex, organic pattern of marbled paper. It features a dense interplay of colors: vibrant blue, rich orange, and creamy white. The colors are swirled and blended together, creating a fluid, almost cellular appearance with many small, rounded shapes and intricate, vein-like structures. The overall effect is one of dynamic movement and organic complexity.

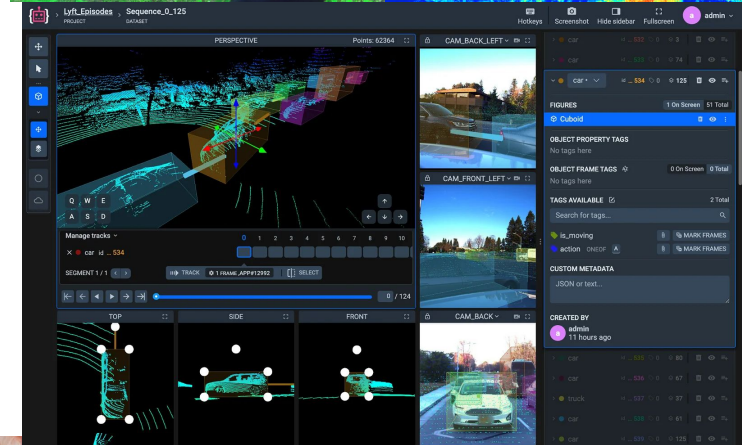
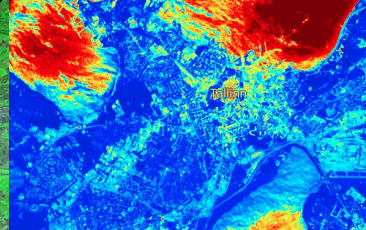
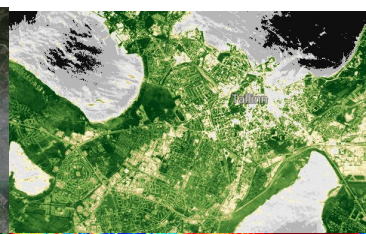
**Multi-modal**



# Модальности

## Разные типы данных

- временной ряд
    - температура
    - акции
    - звук
  - текст
    - код
    - графики
  - изображение
    - многослойное
    - звук
  - видео
  - облако точек
    - LiDaR
- координаты
  - погода
  - позы
  - 3д модель (для AR/VR и игр)
  - сердце, нервы, давление
  - прикосновения / haptic
  - эмоции человека
  - запахи
  - гены
  - молекулы
  - томография
  - ВСІ
  - управление роботом
  - управление протезом



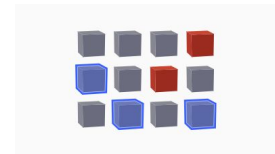
разметка облака точек LiDaR, интерфейс supervisely →





# фото → текст / CLIP (2021)

- Генератор мультимодальных embedding
- Связь картинок с текстом
  - но тренировка происходит не на буквальном тексте на возможных вариантах (T1..Tn) = на смысле. Как тест в школе
- более глубокое понимание картинки
- Contrastive Language-Image Pre-training



Learning Transferable Visual Models From Natural Language Supervision

Alec Radford<sup>1</sup> Jong Wook Kim<sup>2\*</sup> Chris Hallacy<sup>1</sup> Aditya Ramesh<sup>1</sup> Gabriel Goh<sup>1</sup> Sandhini Agarwal<sup>1</sup> Girish Sastry<sup>1</sup> Amanda Askell<sup>1</sup> Pamela Mishkin<sup>1</sup> Jack Clark<sup>1</sup> Gretchen Krueger<sup>1</sup> Ilya Sutskever<sup>1</sup>



Possible class names (comma-separated)

bee, mite, cell, larva

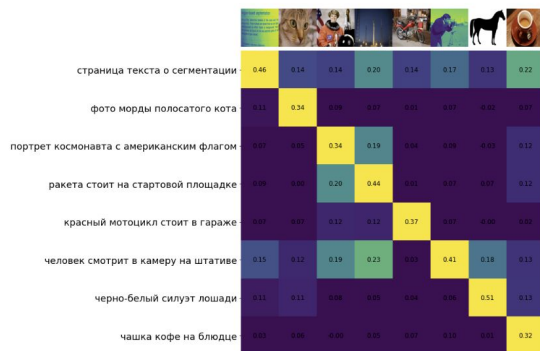
Compute

Computation time on gpu: 0.722 s

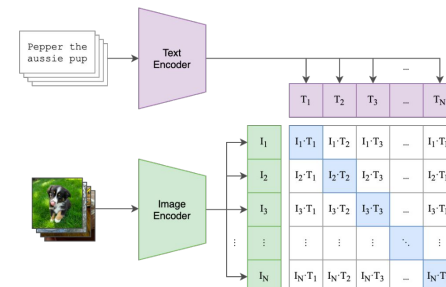
bee	0.955
mite	0.024
larva	0.021
cell	0.000

open-clip  
ru-clip

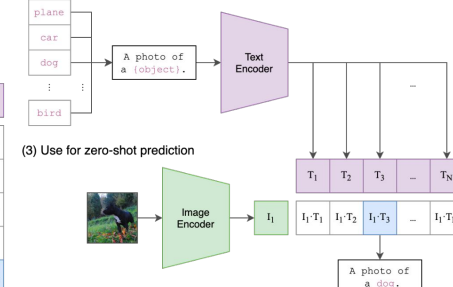
Cosine similarity between text and image features



(1) Contrastive pre-training



(2) Create dataset classifier from label text



(3) Use for zero-shot prediction



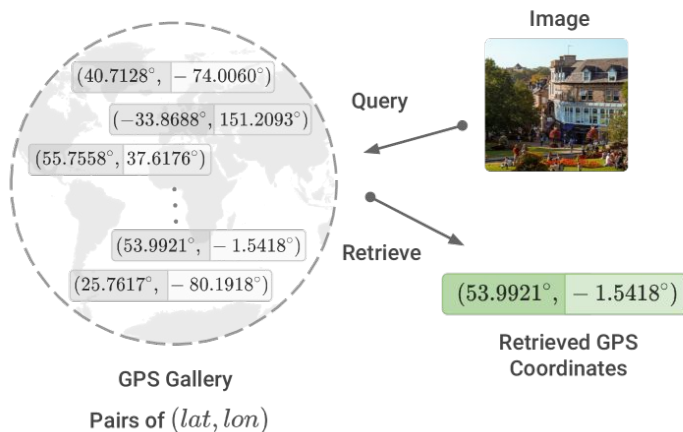
# φOTO → GPS (x,y)

## GeoCLIP: Clip-Inspired Alignment between Locations and Images for Effective Worldwide Geo-localization

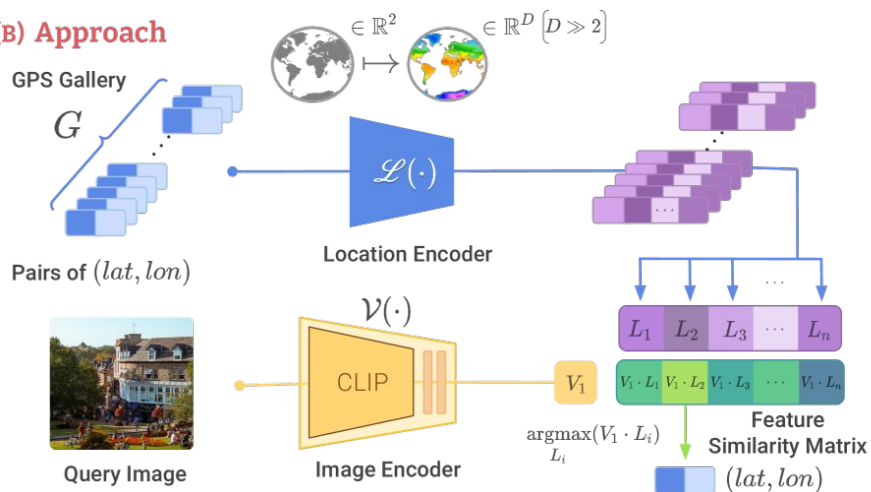
Vicente Vivanco Cepeda, Gaurav Kumar Nayak, Mubarak Shah

Center for Research in Computer Vision, University of Central Florida, USA  
{vicente.vivancocepeda, gauravkumar.nayak}@ucf.edu; shah@crvc.ucf.edu

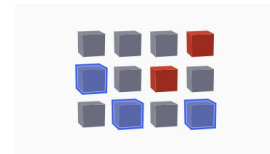
### (A) Problem Setup



### (B) Approach



# фото - текст / GLIP и OWL-ViT (2022)

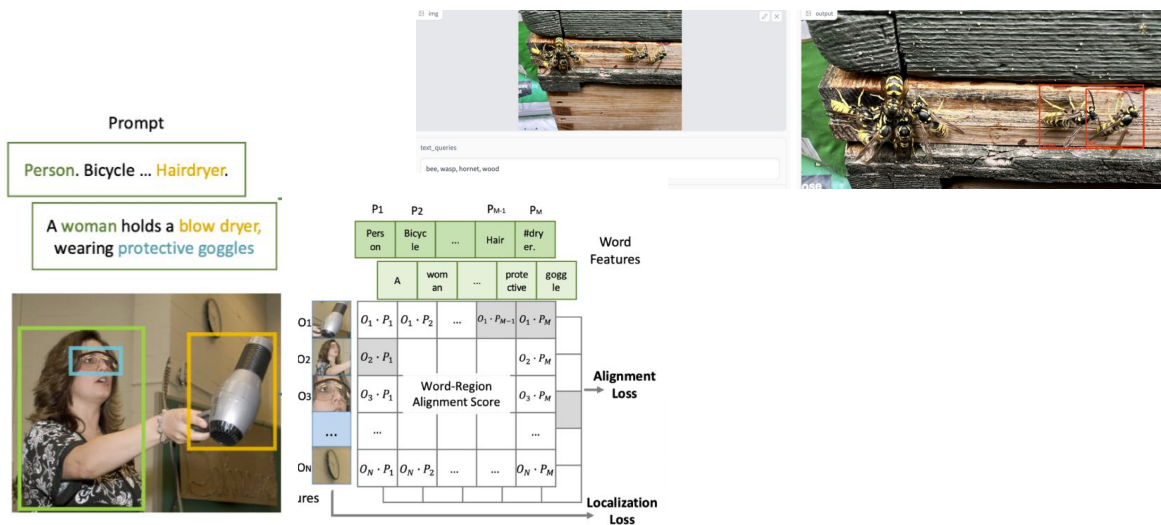


Смысловая связь текст  $\longleftrightarrow$  картинка

- модели сами возвращают классы в зависимости от текстового запроса
- точность плохая

## Simple Open-Vocabulary Object Detection with Vision Transformers

Matthias Minderer\*, Alexey Gritsenko\*, Austin Stone, Maxim Neumann, Dirk Weissenborn, Alexey Dosovitskiy, Aravindh Mahendran, Anurag Arnab, Mostafa Dehghani, Zhuoran Shen, Xiao Wang, Xiaohua Zhai, Thomas Kipf, and Neil Houlsby



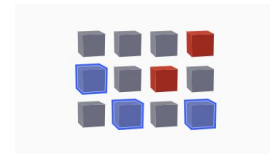
[a] I





# ImageBind (2023)

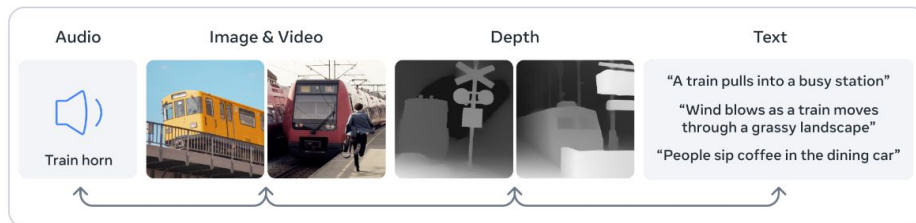
## мультимодальные embeddings



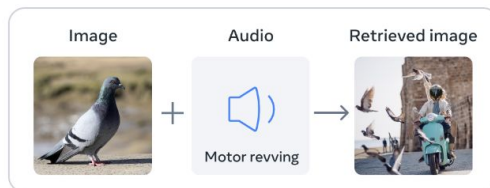
**IMAGEBIND: One Embedding Space To Bind Them All**

Rohit Girdhar\* Alaaeldin El-Nouby\* Zhuang Liu Mannat Singh  
Kalyan Vasudev Alwala Armand Joulin Ishan Misra\*  
FAIR, Meta AI

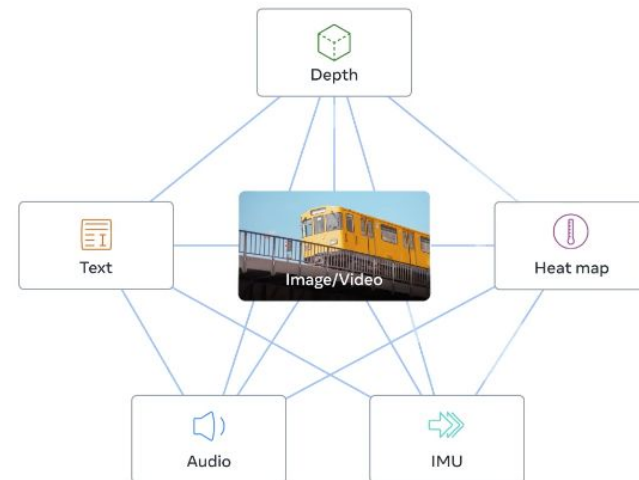
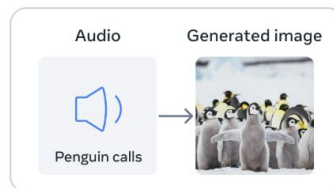
### Cross-modal retrieval



### Embedding-space arithmetic



### Audio to image generation



# N фото → фото / ClimaX (2023)

Предсказание погоды

Моделирование климата (при другом уровне CO2)

Pretraining на данных (foundational model)

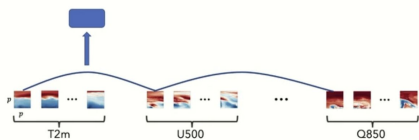
Местное предсказание погоды

CMIP6 datasets - очень разные данные  
температура, давление, влажность, CO2, SO2  
снятые в разное время, в разных точках

ViT как основа

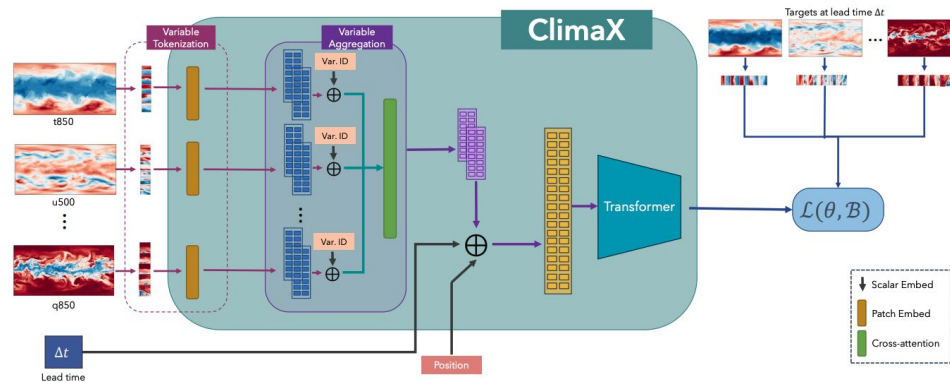
многоспектральный вход

альтернатива FourCastNet, Pangu-weather, GraphCast



## ClimaX: A foundation model for weather and climate

Tung Nguyen<sup>1</sup>, Johannes Brandstetter<sup>2</sup>, Ashish Kapoor<sup>3</sup>,  
Jayesh K. Gupta<sup>1,2</sup>, and Aditya Grover<sup>1</sup>  
<sup>1</sup>UCLA, <sup>2</sup>Microsoft, <sup>3</sup>Scaled Foundations



**Figure 2:** Pretraining phase of ClimaX. Variables are encoded using variable-separate tokenization, and subsequently aggregated using variable aggregation. Together with position embedding and lead time embedding those are fed to the ViT backbone.



# фото → сегментация / SAM (2023)

интерактивная сегментация  
кликами, текстом, маской или bbox

см также SegGPT - сегментация  
текстом



**Segment Anything**

Alexander Kirillov<sup>1,2,4</sup> Eric Mintun<sup>2</sup> Nikhila Rav<sup>1,2</sup> Hanzi Mao<sup>2</sup> Chloe Rolland<sup>3</sup> Laura Gustafson<sup>3</sup>  
Tete Xiao<sup>2</sup> Spencer Whitehead<sup>2</sup> Alexander C. Berg<sup>2</sup> Wan-Yen Lo<sup>2</sup> Piotr Dollár<sup>4</sup> Ross Girshick<sup>4</sup>





# фото → сегментация HIPIE (2023)

semantic segmentation, instance  
segmentation, panoptic  
segmentation, referring  
segmentation, and part/subpart  
segmentation

## Hierarchical Open-vocabulary Universal Image Segmentation

Xudong Wang<sup>1\*</sup> Shufan Li<sup>1\*</sup> Konstantinos Kallidromitis<sup>2\*</sup> Yusuke Kato<sup>2</sup>

Kazuki Kozuka<sup>2</sup> Trevor Darrell<sup>1</sup>

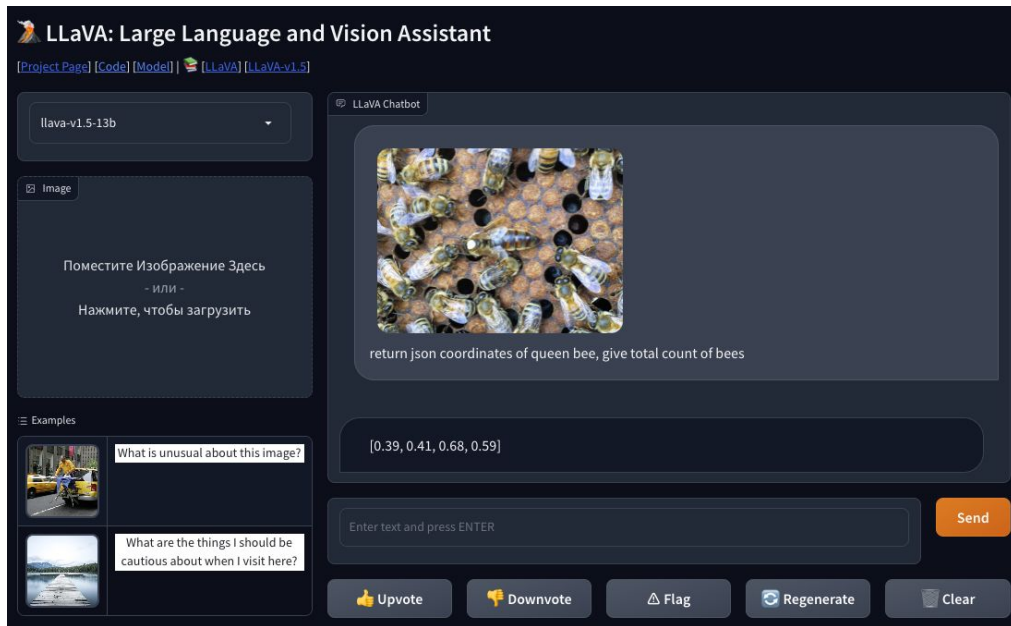
<sup>1</sup>Berkeley AI Research, UC Berkeley <sup>2</sup>Panasonic AI Research



# фото + текст → текст / LLaVA (2023)

Интерактивный prompting LLM  
о картинке - верни мне JSON  
матки

Не надо больше обучать CNN с  
нуля?

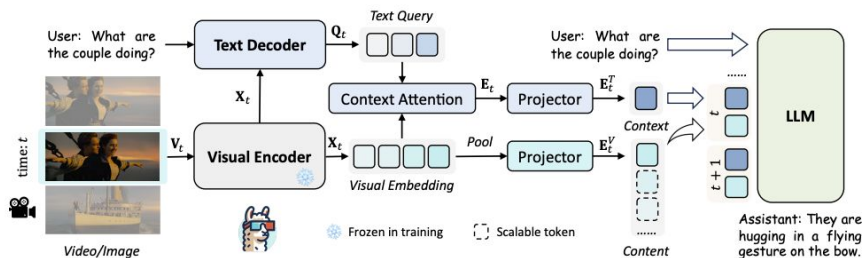


The screenshot displays the LLaVA chat interface. At the top, it says "LLaVA: Large Language and Vision Assistant" with links to "Project Page", "Code", "Model", and "[LLaVA] [LLaVA-v1.5]". The model selected is "llava-v1.5-13b". The chat area shows a user prompt: "Поместите Изображение Здесь - или - Нажмите, чтобы загрузить" (Place Image Here - or - Click to upload). An image of a beehive is uploaded. The user's prompt is "return json coordinates of queen bee, give total count of bees". The model's response is "[0.39, 0.41, 0.68, 0.59]". Below the chat area, there are "Examples" showing two prompts: "What is unusual about this image?" and "What are the things I should be cautious about when I visit here?". At the bottom, there are buttons for "Upvote", "Downvote", "Flag", "Regenerate", and "Clear".



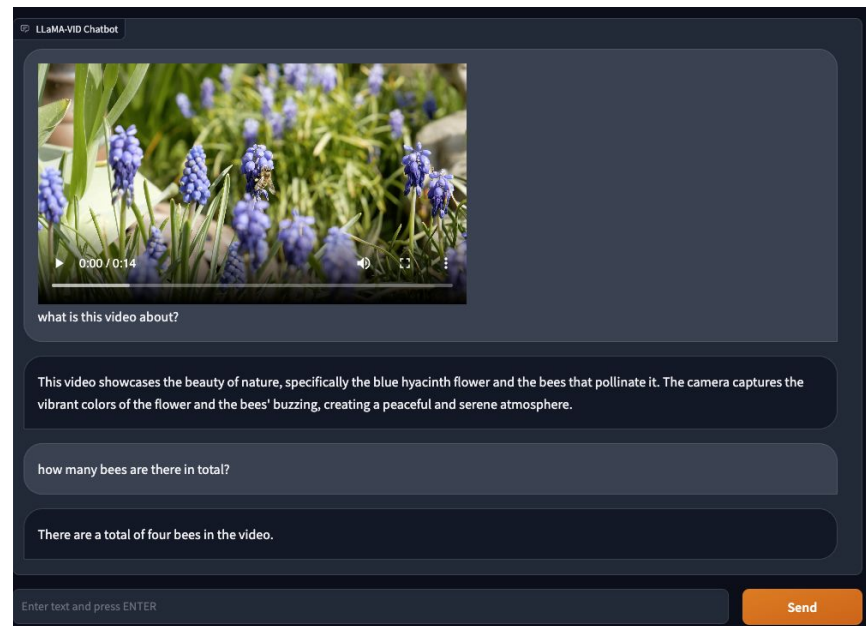
# ВИДЕО + ТЕКСТ LLaMa-VID (2023)

## Интерактивный prompting видео



LLaMA-VID: An Image is Worth 2 Tokens in Large Language Models

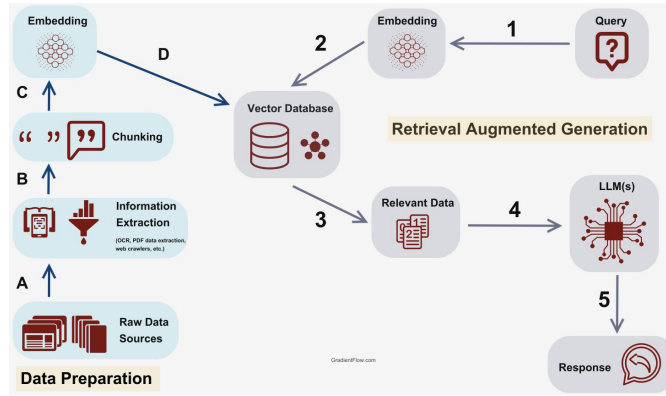
Yanwei Li<sup>1\*</sup> Chengyao Wang<sup>1\*</sup> Jiaya Jia<sup>1,2</sup>  
CUHK<sup>1</sup> SmartMore<sup>2</sup>





# ДОКУМЕНТ + ТЕКСТ → ТЕКСТ / RAG

- вместо дополнения контекста своим текстом..
- превращаем любые внешние документы (текст+таблицы+картинки) в embeddings
- при запросе к модели
  - ищем ближайшие векторы
  - дополняем контекст ответами
  - передаем в модель



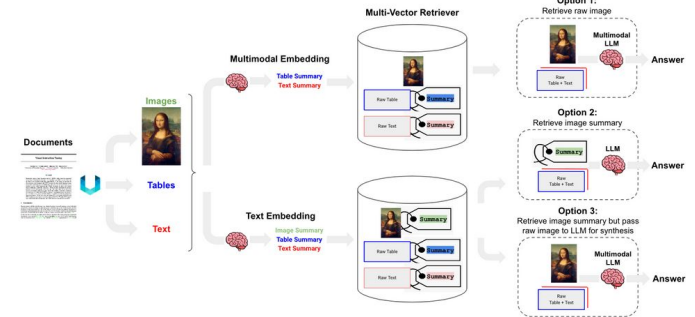
## Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks

Patrick Lewis<sup>†</sup>, Ethan Perez<sup>\*</sup>,

Aleksandra Piktus<sup>†</sup>, Fabio Petroni<sup>†</sup>, Vladimir Karpukhin<sup>†</sup>, Naman Goyal<sup>†</sup>, Heinrich Küttler<sup>†</sup>,

Mike Lewis<sup>†</sup>, Wen-tau Yih<sup>†</sup>, Tim Rocktäschel<sup>††</sup>, Sebastian Riedel<sup>††</sup>, Douwe Kiela<sup>†</sup>

<sup>†</sup>Facebook AI Research; <sup>‡</sup>University College London; <sup>\*</sup>New York University;





# фото + текст → код → текст / ViperGPT (2023)

ViperGPT: Visual Inference via Python Execution for Reasoning

Генерируем python код  
Исполняем с учетом что на входе

умеет видео  
chain of thought (разбиение на шаги)

Điđac Suris\*, Sachit Menon\*, Carl Vondrick  
Columbia University  
viper.cs.columbia.edu

Query: How many muffins can each kid have for it to be fair?



## Generated Code

```
def execute_command(image):  
    image_patch = ImagePatch(image)  
    muffin_patches = image_patch.find("muffin")  
    kid_patches = image_patch.find("kid")  
    return str(len(muffin_patches) // len(kid_patches))
```

## Execution

```
muffin_patches =  
image_patch.find("muffin")
```



```
kid_patches =  
image_patch.find("kid")
```



```
► len(muffin_patches)=8  
► len(kid_patches)=2
```

```
► 8//2 = 4
```

```
Result: 4
```

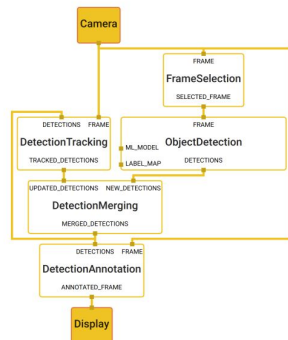
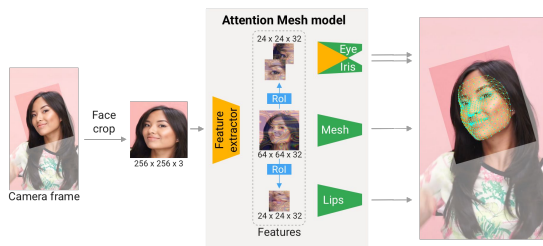




# MediaPipe (2019)

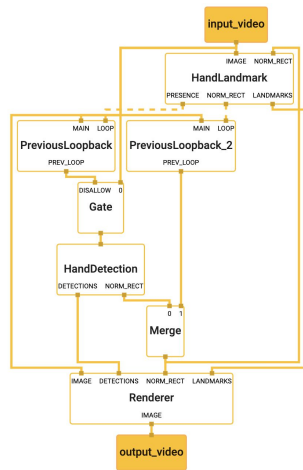
framework / модульный набор

- нахождение лиц
- нахождение жестов рук
- детект позы
- mesh лица



## MediaPipe: A Framework for Building Perception Pipelines

Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg and Matthias Grundmann  
Google Research





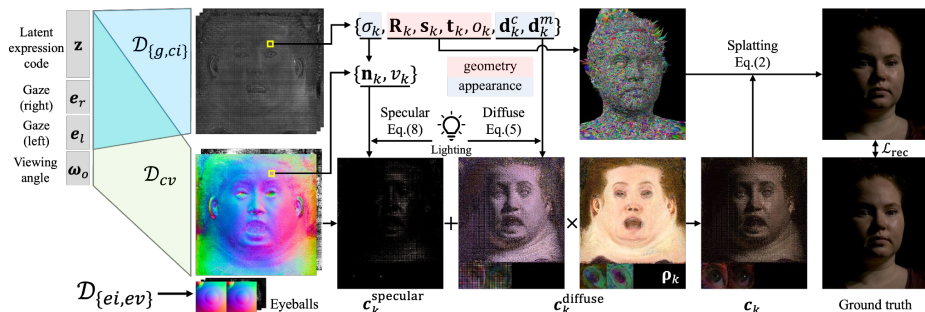
# глаза + фото + эмоция → 3d

Генерируем 3d аватары  
сжатый VR в realtime

## Relightable Gaussian Codec Avatars

Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, Giljoo Nam

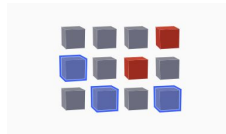
Codec Avatars Lab, Meta







# фото + текст → текст + фото / Gemini (2023)



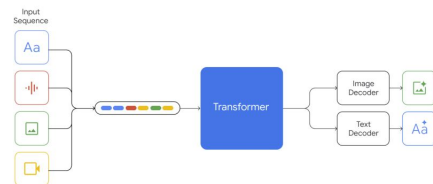
- expert in 57 subjects
- Multimodal-first
- 32k token context length - маловато
- tech report для маркетинга
- демо видео - монтаж

Question	LLaVA	BakLLaVA	Qwen-VL	CogVLM	GPT-4V	Gemini
How many coins do I have? (VQA)	Green	Green	Green	Green	Red	Green
Which movie is this scene from? (VQA)	Red	Red	Green	Yellow	Green	Green
Read text from the picture (Document OCR)	Red	Red	Green	Green	Green	Yellow
What is the price of Pastrami Pizza? (Document VQA)	Red	Red	Red	Green	Red	Green
How much tax did I pay? (Document VQA)	Red	Red	Red	Red	Red	Green
Find the dog. (Zero-Shot Object Detection)	Red	Red	Green	Green	Red	Yellow
Read the serial number. (OCR)	Red	Red	Green	Green	Green	Red

сравнение моделей - [blog.roboflow.com](http://blog.roboflow.com)

Google DeepMind

## Gemini: A Family of Highly Capable Multimodal Models



How about this one?



Gemini: I see a hand with two fingers extended, which is a common symbol for the number two.

But wait ... what if we asked Gemini to reason about all of these images together?



What do you think I'm doing? Hint: it's a game.  
Gemini: You're playing rock, paper, scissors.



The background is a dense, abstract pattern of organic, cell-like shapes. The colors are primarily shades of blue and yellow, with some lighter, almost white, areas. The shapes are irregular and interconnected, creating a complex, textured surface that resembles a microscopic view of a material or a biological structure. The overall effect is vibrant and dynamic.

**Будущее**



# Рост моделей - Mixture of Experts

Параллелизация / разделение части модели на несколько (8) экспертов

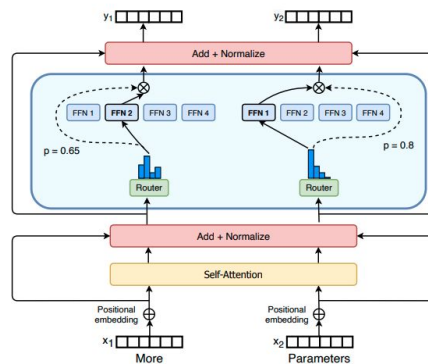
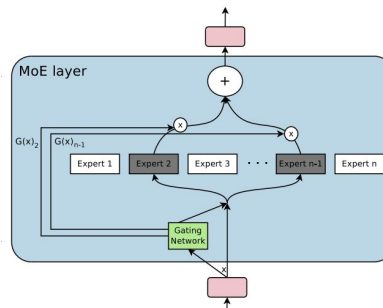
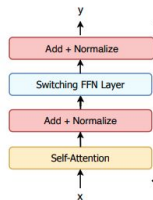
- быстрее inference
  - эксперты могут быть на разных GPU
  - эксперты специализируются на токенах, пунктуации, глаголах, союзах, визуальном описании
  - больше параметров в целом, еще больше VRAM
- Mistral 8x7B (80GB)

Рутер → эксперт  
Изначально для LSTM (2017)

В случае трансформеров - несколько FFN  
Тренировка одновременно обоих

Проблема - дисбаланс экспертов

Mistral / GPT-4  
<https://huggingface.co/blog/moe>

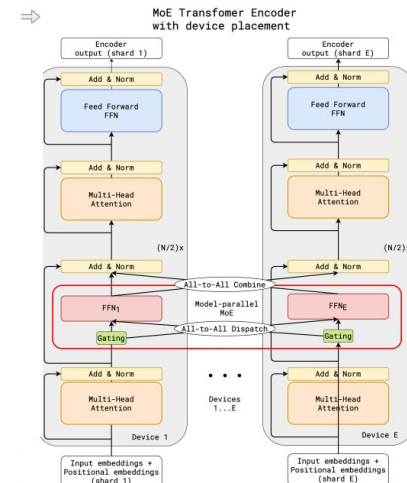


SWITCH TRANSFORMERS: SCALING TO TRILLION PARAMETER MODELS WITH SIMPLE AND EFFICIENT SPARSITY

William Fedus\*  
Google Brain  
lliamfedus@google.com

Barret Zoph\*  
Google Brain  
barret.zoph@google.com

Noam Shazeer  
Google Brain  
noam@google.com

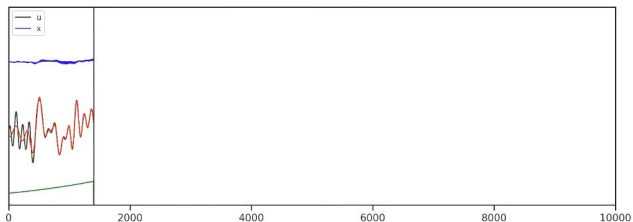


# Сжатие контекста - Structured State Space Models (2021)

- State Space Model (1960) для последовательности данных, по сути RNN. Vanishing gradient = плохо обучается
  - $u \rightarrow x \rightarrow y$
- Упрощает вычисление последовательностей, но с потерями
  - Звук, картинки
  - линейно использует память  $O(N)$
- в 5-60x быстрее трансформеров
- The Annotated S4

<https://srush.github.io/annotated-s4/>

Hippo  $\rightarrow$  S4  $\rightarrow$  S5  $\rightarrow$  Mamba (2023)  
Hyena, RetNet, RWKV



	Model	sMNIST	FMNIST	sCIFAR
<b>Transformers</b>	Transformer (Vaswani et al., 2017; Triih et al., 2018)	98.9	97.9	62.2
	CKConv (Romero et al., 2021)	99.32	98.54	63.74
	TritNet (Bai et al., 2019)	99.20	98.13	73.42
	TCN (Bai et al., 2018)	99.0	97.2	-
<b>CNNs</b>	LSTM (Hochreiter & Schmidhuber, 1997; Gu et al., 2020b)	98.9	95.11	63.01
	rLSTM (Triih et al., 2018)	98.4	95.2	72.2
	Dilated GRU (Chang et al., 2017)	99.0	94.6	-
	Dilated RNN (Chang et al., 2017)	98.0	96.1	-
	IndRNN (Li et al., 2018)	99.0	96.0	-
<b>RNNs</b>	expRNN (Lozcano-Casado & Martinez-Rubio, 2019)	98.7	96.6	-
	UR-LSTM	99.28	96.96	71.00
	UR-GRU (Gu et al., 2020b)	99.27	96.51	74.4
	LMU (Voelker et al., 2019)	-	97.15	-
	Hippo-RNN (Gu et al., 2020a)	98.9	98.3	61.1
<b>SSMs</b>	UNicoRNN (Rusch & Mishra, 2021)	-	98.4	-
	LMUFT (Chalkuri & Elasmith, 2021)	-	98.49	-
	LipshitzRNN (Erichson et al., 2021)	99.4	98.3	64.2
	<b>S4</b>	<b>99.63</b>	<b>98.70</b>	<b>91.13</b>

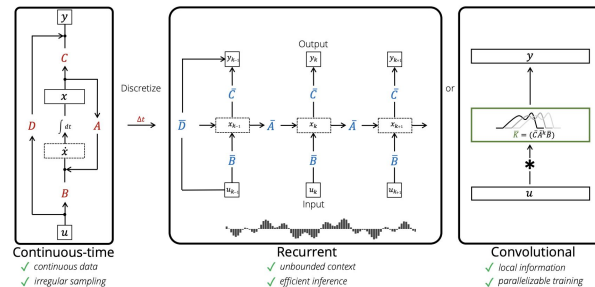
Benchmark spanning text, images, symbolic reasoning (length 1K-16K)

Model	LISTOps	TEXT	RETRIEVAL	IMAGE	PATHFINDER	PATH-X	AVG
Random	10.00	50.00	50.00	10.00	50.00	50.00	36.67
Transformer	36.37	64.27	57.46	42.44	71.40	✗	53.66
Local Attention	15.82	52.98	53.39	41.46	66.63	✗	46.71
Sparse Trans.	17.07	63.58	59.59	44.24	71.71	✗	51.03
Longformer	35.63	62.85	56.89	42.22	69.71	✗	52.88
Linformer	35.70	53.94	52.27	38.56	76.34	✗	51.14
Reformer	37.27	56.10	53.40	38.07	68.50	✗	50.56
Sinkhorn Trans.	33.67	61.20	53.83	41.23	67.45	✗	51.23
Synthesizer	38.99	61.68	54.67	41.61	69.45	✗	52.40
BigBird	36.05	64.02	59.29	40.83	74.87	✗	54.17
Linear Trans.	16.13	65.90	53.09	42.34	75.30	✗	50.46
Performer	18.01	65.40	53.82	42.77	77.05	✗	51.18
FNet	35.33	65.11	59.61	38.67	77.80	✗	54.42
Nystromformer	37.15	65.52	79.56	41.58	70.94	✗	57.46
Luna-256	37.25	64.57	79.29	47.38	77.72	✗	59.37
<b>S4</b>	<b>58.35</b>	<b>76.02</b>	<b>87.09</b>	<b>87.26</b>	<b>86.05</b>	<b>88.10</b>	<b>80.48</b>

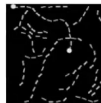
Efficiently Modeling Long Sequences with Structured State Spaces

Albert Gu, Karan Goel, and Christopher Ré

Department of Computer Science, Stanford University



Path-X

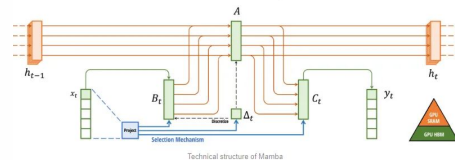


(a) A positive example.



(b) A negative example.

Selective State Space Model  
with Hardware-aware State Expansion



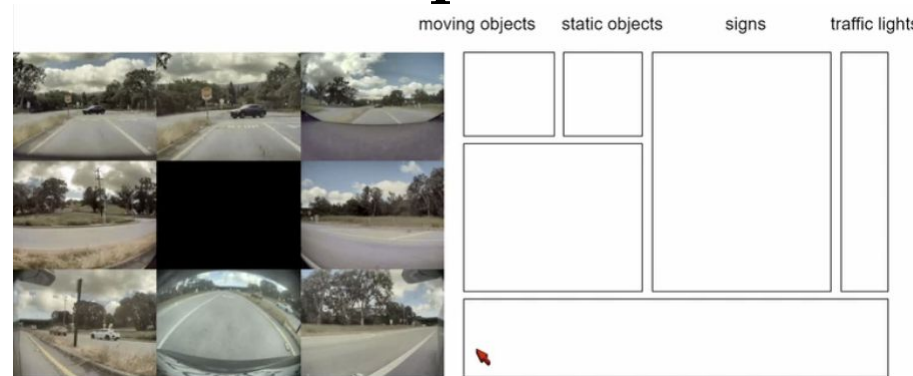
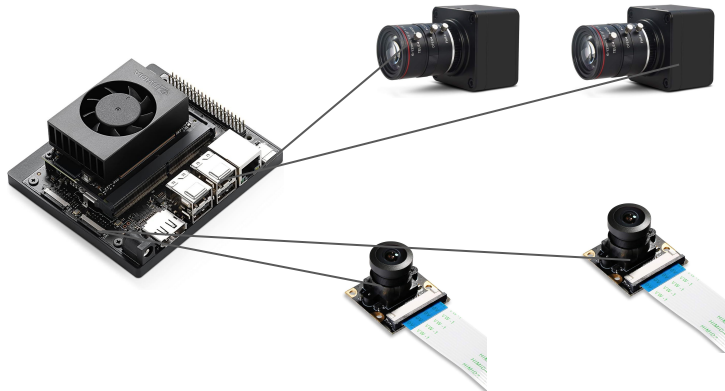
(c) Technical structure of Mamba



# Ограничения железа - монолит или микро- модели

100 разных задач - машины, знаки, путь,  
статичные объекты

камеры вместе тренируются  
задачи используют разные камеры  
ограниченность по GPU, требования по высокому  
FPS



камеры у Tesla → разные задачи и модели  
“Multi-Task Learning in the Wilderness”

монолит

- эффективность использования VRAM GPU
- подборка датасета под разные задачи оч тяжело из-за разного распределения классов
- остановка обучения у задач происходит в разное время

“микросервисы”

- простота обучения
- независимость команд
- неэффективность взаимосвязи GPU → RAM → Network
- AutoML?





# За кем следить



- Google DeepMind / Google Research → BARD, LaMDA, PaLM, Imagen, MusicLM, Gemini



- OpenAI → GPT, CLIP, Whisper, ImageGPT



- Anthropic → Claude



- Meta AI → LLaMa2, Galactica



- Microsoft Research → Orca, phi-2, LayoutLM, LLaVa, Kosmos, table-transformer, diberta, Speech-tts



- Cohere



- Mistral → Mixtral 8x7B



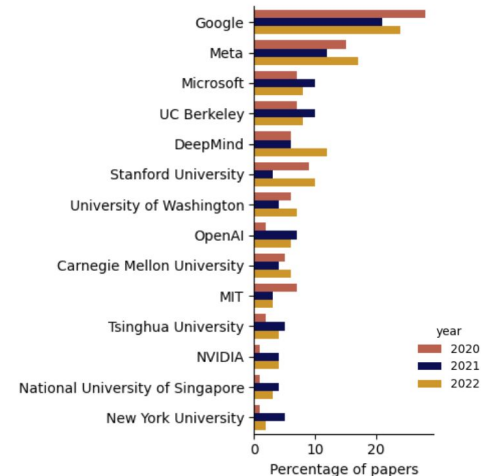
- NVIDIA → Nemotron, PeopleNet, SegFormer, Conformer, Megatron, BioBERT, NeVA, GenSLM, Vista3D

- Runway → Stable diffusion



- Beijing Academy of Artificial Intelligence → BAAI

- TII → Falcon



Lmsys.org Chatbot Arena Elo Ratings (based on blind voting by users) 17 December 2023

Company	Model	Elo Rating	License
OpenAI	GPT-4-Turbo	1233	Proprietary
OpenAI	GPT-4-0314	1191	Proprietary
OpenAI	GPT-4-0613	1157	Proprietary
Anthropic	Claude-1	1151	Proprietary
Anthropic	Claude-2.0	1130	Proprietary
Anthropic	Claude-2.1	1120	Proprietary
OpenAI	GPT-3.5-Turbo-0613	1116	Proprietary
Mistral	Mixtral-8x7b-Instruct	1116	Apache 2.0
Anthropic	Claude-Instant-1	1110	Proprietary
Open Source	Tulu-2-DPO-70B	1110	AI2 ImpACT
01 AI (China)	Yi-34B-Chat	1109	Yi License
Google	Gemini Pro	1106	Proprietary
OpenAI	GPT-3.5-Turbo-0314	1105	Proprietary
Open Source	WizardLM-70B-v1.0	1102	Llama 2
Open Source	Vicuna-33B	1096	Non-commercial
Open Source	Starling-LM-7B-alpha	1088	CC-BY-NC-4.0
Open Source	OpenChat-3.5	1077	Apache-2.0
OpenAI	GPT-3.5-Turbo-1106	1077	Proprietary
Perplexity	pplx-70b-online	1075	Proprietary
Meta	Llama-2-70b-chat	1074	Llama 2

Observations:

- OpenAI is still the king of LLMs, and their models are getting better
- Anthropic is not too far behind, but their models are getting worse
- Mistral is a very strong new entrant with the best Open Source model
- Google's Gemini Pro has not impressed, it is a ~GPT-3.5 model
- A lot of open source rely on Llama 2 or GPT-4 outputs for fine tuning
- Meta's Llama 2 itself is not a very highly performant model



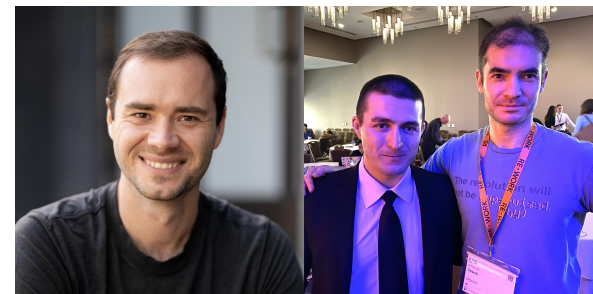
# У КОГО УЧИТЬСЯ



- MIT Deep Learning Basics → Lex Fridman
- MIT Introduction to Deep Learning | 6.S191
- Stanford CS221: Artificial Intelligence: Principles and Techniques
- Stanford CS229: Machine Learning Course
- Stanford CS231n: Deep Learning for Computer Vision
- Stanford CS25 - Transformers United
- Stanford CS224N: NLP with Deep Learning
- DeepMind RL Course by David Silver
- Harvard CS50's Introduction to Artificial Intelligence with Python



- Конференции - ICML, ICLR, CVPR, MLCON, ICCV
- Geoffrey Hinton, Jeremy Howard, Andrew Ng, Andrej Karpathy
- СПбГУ - Сергей Николенко - Глубокое обучение
- МГУ - Радолав Нейчев – Введение в глубокое обучение
- КФУ - Евгений Разинков - Глубокое обучение



# С чего начать

- практический проект
- модальность (текст, картинки, аудио, видео)
- данные → увеличить и сбалансировать
  - свои, ObjectNet, ImageNet, ILSVRC-2012, huggingface, roboflow, datasetninja)
- разметка
- модель → transfer-learning + fine-tuning. Pytorch.
- тренировка
  - оценить метрики (mAP), сравнить с другими моделями
- deploy



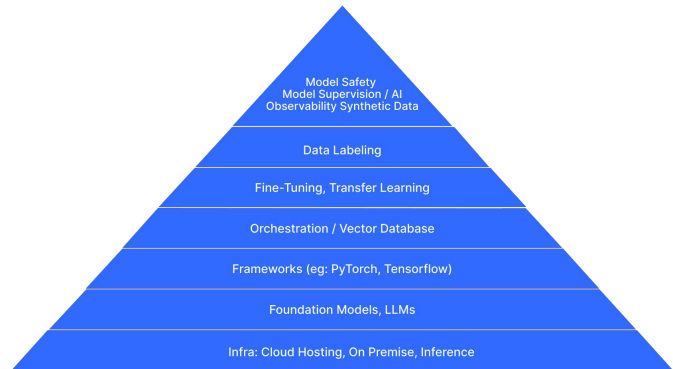
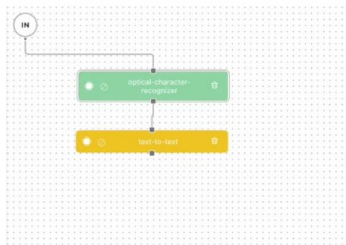


# С чего начать - Clarifai.com



Full-stack AI платформа для

- разметки данных (labeling)
- смыслового поиска (embeddings!)
- обучения моделей
- предикта данных по API
- связки моделей в мультимодальный workflow



Profile Community

tpessagno vehicle-detect

RESOURCES

- Inputs 207
- Datasets 5
- Concepts 16
- Models 9
- Workflows 4
- Collectors 2
- Labeler Tasks
- Settings

Predictions

vehicle-general Ver 2.1

BOUNDING BOXES

- vehicle
- 1 bbox : 0.957
- 2 bbox : 0.957
- 3 bbox : 0.957
- 4 bbox : 0.957
- 5 bbox : 0.953

MASKS

POLYGONS

12:54 - 15:14



# Инструментарий

jupyter  
google colab

kubeflow  
nvidia triton

Pytorch  
Tensorflow  
Keras  
CoreML  
SciKit-learn  
MXNet  
CNTK  
Caffe  
NLTK

ONNX = protobuf графа  
pickle - serialized python  
object



# AI инструментарий

- HuggingFace.co, Stability.ai, Inflection.ai, snorkel.ai, AI21, H2O.ai, scale.ai, gradio.app, octoml.ai, labelbox, supervisely, v7, mosaicml, databricks
- Nvidia Triton, Nvidia NeMo, lamini.ai, adept.ai, character.ai, coreweave, lambda labs





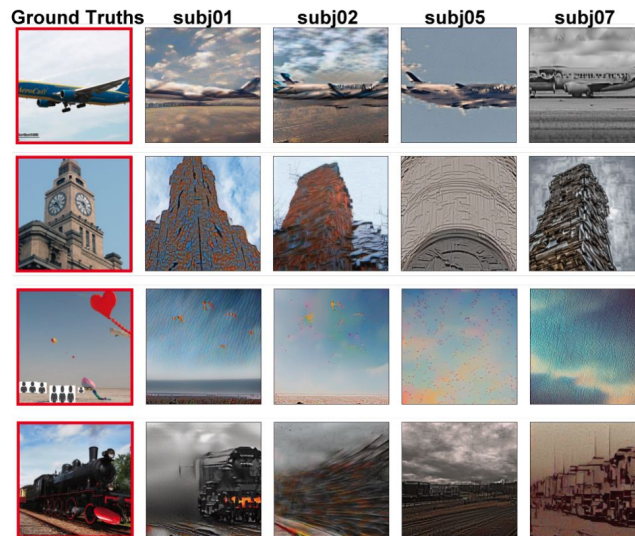
# MRI → фото / Чтение мыслей (2022)

- fMRI voxel на входе
- картинка на выходе
- CLIP → Stable Diffusion

<https://github.com/yu-takagi/Stable-DiffusionReconstruction>

High-resolution image reconstruction with latent diffusion models from human brain activity

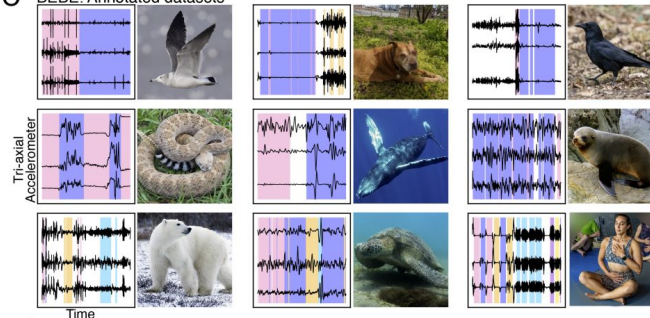
Yu Takagi<sup>1,2\*</sup> Shinji Nishimoto<sup>1,2</sup>



# Bio-logger Ethogram Benchmark

Мультимодальный датасет  
Разные виды животных

C BEBE: Annotated datasets



Ethogram subset (present in sensor chip)	Gull	Rattlesnake	Polar bear	Dog	Whale	Sea turtle	Crow	Seals	Human
Stationary	Stationary	Move	Head shake	Pant + lie on chest	Travel	Swim	At nest	Feed	Sit
Forage	Stationary	Stationary	Roll	Lie on chest	Explore	Feed	Stand	Travel	Stand
			Groom	Gallop		Glide			Lie
			Walk			Stay at surface			Walk
									Walk upstairs
									Walk downstairs

A BENCHMARK FOR COMPUTATIONAL ANALYSIS OF ANIMAL BEHAVIOR, USING ANIMAL-BORNE TAGS

Benjamin Hoffman* Earth Species Project benjamin@earthspecies.org	Maddie Cusiman* Earth Species Project maddie@earthspecies.org
Vittorio Baglione U. de León vbag@unileon.es	Daniela Canevari CNRS Borea dcan@unileon.es
Dominic L. DeSantis Georgia College & State U. dominic.desantis@gc.edu	Lorine Jeanet U. of Stellenbosch, African Institute for Mathematical Sciences lorine.jeanet@matmail.fr
Takuya Mackawa Osaka U. takuya.mackawa@acn.org	Vicente Mata-Silva U. Texas El Paso vmata@utep.edu
Eva Trapote Universidad de León etrap@unileon.es	Victor Moreno-González Universidad de León vmz@unileon.es
Antti Vehkaoja Tampere U. antti.vehkaoja@uni.fi	Outi Vainio U. of Helsinki outi.vainio@helsinki.fi
Ari Friedlander <sup>1</sup> U.C. Santa Cruz ari.friedlander@ucsc.edu	Ken Yoda Nagoya U. yoda.ken@nagoya-u.jp
	Katherine Zacarias Earth Species Project katie@earthspecies.edu
	Christian Rutz <sup>1</sup> Univ. of St. Andrews christian.rutz@st-andrews.ac.uk

Data Article

Description of movement sensor dataset for  
dog behavior classification



Antti Vehkaoja<sup>a,\*</sup>, Sanni Somppi<sup>b</sup>, Heini Törnqvist<sup>b,c</sup>,  
Anna Valldeoriola Cardó<sup>b</sup>, Pekka Mumpulainen<sup>a</sup>, Heli Väättäjä<sup>c,d</sup>,  
Päivi Majaranta<sup>c</sup>, Veikko Surakka<sup>c</sup>, Miimaaria V. Kujala<sup>b,c</sup>,  
Outi Vainio<sup>b,h</sup>

<sup>a</sup> Faculty of Medicine and Health Technology, Tampere University, P.O. Box 692, Tampere FI-33101, Finland  
<sup>b</sup> Department of Equine and Small Animal Medicine, University of Helsinki, P.O. Box 57, Helsinki FI-00014, Finland  
<sup>c</sup> Research Group for Emotions, Sociality, and Computing, Faculty of Information Technology and Communication Sciences, Tampere University, P.O. Box 100, Tampere FI-33014, Finland  
<sup>d</sup> Master School, Lapland University of Applied Sciences, Jokiväylä 11 B, Rovaniemi 96300, Finland  
<sup>e</sup> Department of Psychology, Faculty of Education and Psychology, University of Jyväskylä, P.O. Box 35, Jyväskylä FI-40014, Finland

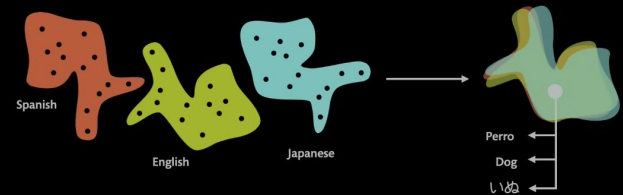
```

DogID, TestNum, t_sec, ABack_x, ABack_y, ABack_z, ANeck_x, ANeck_y, ANeck_z, GBack_X, GBack_y, GBack_z, GNeck_x, GNeck_y, GNeck_z, Task, Behavior_1, Behavior_2, Behavior_3, PolNT
16, 1, 0, 0.041504, 0.938965, -0.015137, -0.067871, -0.510254, -0.93457, -17.639161, -22.766115, 7.44629, -7.934571, 6.347657, 13.427735, <undefined>, <undefined>, <undefined>, <
16, 1, 0.01, 0.041992, 0.941895, -0.02002, -0.128906, -0.494141, -0.913086, -15.075685, -11.413575, 4.821778, -3.90625, 4.394532, 16.540528, <undefined>, Synchronization, <undef
16, 1, 0.02, 0.040527, 0.939453, -0.004395, -0.158691, -0.480469, -0.911133, -12.207032, -0.12207, 0.207617, -0.488281, -1.953125, 26.794435, <undefined>, Synchronization, <unde
16, 1, 0.03, 0.021484, 0.946289, 0.007813, -0.12207, 0.486816, -0.880371, -9.46045, 7.995606, 1.586914, 1.159668, -5.67627, 38.08594, <undefined>, Synchronization, <undefined>
16, 1, 0.04, -0.000977, 0.951172, 0.033691, -0.053711, -0.5, -0.807129, -8.361817, 14.587403, -1.037598, 4.577637, 4.089356, 41.503909, <undefined>, Synchronization, <undefined>
16, 1, 0.05, -0.006348, 0.955078, 0.038574, 0.019043, 0.463867, -0.749023, -9.826661, 19.653322, -5.493164, 12.939454, 20.81299, 32.53174, <undefined>, Synchronization, <undefi
16, 1, 0.06, 0.001953, 0.959473, 0.035156, 0.045898, -0.376953, -0.78125, -11.657716, 15.869142, -7.751465, 22.70508, 29.968264, 14.038807, <undefined>, Synchronization, <undefi
16, 1, 0.07, 0.013672, 0.958984, 0.030273, 0.028809, -0.307617, -0.941406, -13.671876, 5.859375, -8.361817, 23.925783, 19.470216, -5.920411, <undefined>, Synchronization, <undef
16, 1, 0.08, 0.02832, 0.941895, 0.04541, -0.014648, -0.349121, -1.100098, -16.113282, -1.220703, -8.483887, 12.084962, -8.300782, -20.202638, <undefined>, Synchronization, <unde
16, 1, 0.09, 0.028809, 0.934082, 0.06543, -0.061035, -0.480469, -1.129395, -18.859864, -0.183105, -7.8125, -3.845215, -37.475588, -24.475809, <undefined>, Synchronization, <unde
16, 1, 0.1, 0.03125, 0.932617, 0.061523, -0.086426, -0.598633, -1.083008, -22.155763, 7.56836, -8.789063, -10.681153, -42.968753, -20.324708, <undefined>, Synchronization, <unde
16, 1, 0.11, 0.022461, 0.950684, 0.045898, -0.086914, -0.634766, -0.996094, -24.230959, 12.145997, -10.009766, -3.90625, -30.700685, -13.122559, <undefined>, Synchronization, <unde
16, 1, 0.12, 0.012207, 0.980957, 0.030273, -0.071289, -0.599609, -0.931152, -22.094728, 7.934571, -11.169434, 6.225586, -5.187989, -7.507325, <undefined>, Synchronization, <unde
16, 1, 0.13, -0.011719, 1.030273, 0.016113, -0.101074, -0.54248, -0.925293, -17.211915, -0.427246, -13.427735, 10.620118, 11.962891, -4.943848, <undefined>, Synchronization, <un
16, 1, 0.14, -0.049805, 1.101563, -0.004883, -0.112793, -0.501465, -0.91748, -11.840821, -6.713868, -15.869142, 9.765626, 5.859375, -3.723145, <undefined>, Synchronization, <und
16, 1, 0.15, -0.062012, 1.206543, -0.023926, -0.087891, -0.48584, -0.919922, -7.629395, -5.737305, -20.446779, 7.324219, -2.319336, -6.835938, <undefined>, Synchronization, <und
16, 1, 0.16, -0.062012, 1.338379, -0.082031, -0.075195, -0.46875, -0.929199, -0.305176, -2.990723, -27.954103, 8.850098, -2.746582, -13.854981, <undefined>, Synchronization, <un
16, 1, 0.17, -0.003418, 1.433594, -0.068848, -0.078125, -0.444824, -0.922363, 10.314942, -4.760742, -34.442383, 12.756348, -1.831055, -20.202638, <undefined>, Synchronization, <un
    
```



# Latent space

- Международный язык
- Модель мира
- Мультимодальный
- Векторные БД как постоянное, независимое от модели хранилище
- Межвидовой язык
  - главное хороший encoder, decoder и маппинг



Using AI to Decode Animal Communication with Aza Raskin





# Наблюдайте за пчелами

ОНИ МОГУТ ВАМ ЧТО-ТО РАССКАЗАТЬ

