



Компьютерное зрение. Введение

Лекция 1

Лектор: Демяненко Я.М.,
ФИИТ, 2026



Что это такое и почему это сложно?



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

Компьютер видит

Магия???



Мы видим

Компьютерное зрение — это область ИИ, которая обучает компьютеры интерпретировать и понимать визуальный мир

Тест Тьюринга

«To see means to know what is where by looking»
David Marr, Vision, 1982

«Тест Тьюринга» — компьютер должен ответить на любой вопрос об изображении, на который может ответить человек

❌ Лёгкий вопрос: "Есть ли там люди?"

✅ Сложный вопрос: «Почему мужчина слева остановился?»

Зрение — это ИИ-полная задача



Где используется CV?

Рынок CV к 2030: > \$200 млрд

 Медицина (диагностика)

 Автопилот (пешеходы)

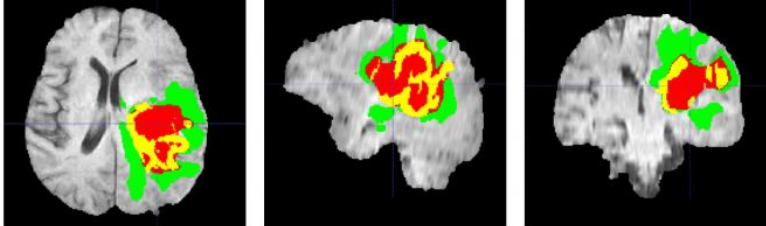
 AR/Фильтры (навигация)

 Безопасность (биометрия)

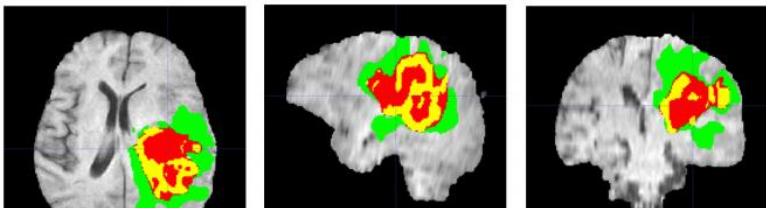
 Роботы (навигация)

 Искусство (генерация)

True



Predicted



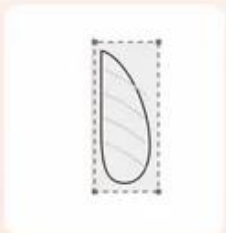
CV в медицине



-  ИИ находит 95% опухолей (врач — 92%)
-  Скорость: 2 секунды против 15 минут
-  ИИ НЕ ЗАМЕНЯЕТ врача, а помогает ему
- "Мы лечим пациентов, а не картинки"

Разметка — это спор врачей

Границу находят каждый по-своему, и это надо разбирать руками



Прямоугольник на всё лёгкое
сеть учит фон вместе с находкой



Контур трих врачей
верить можно пересечением

Автономный транспорт

Камера

Нейросеть

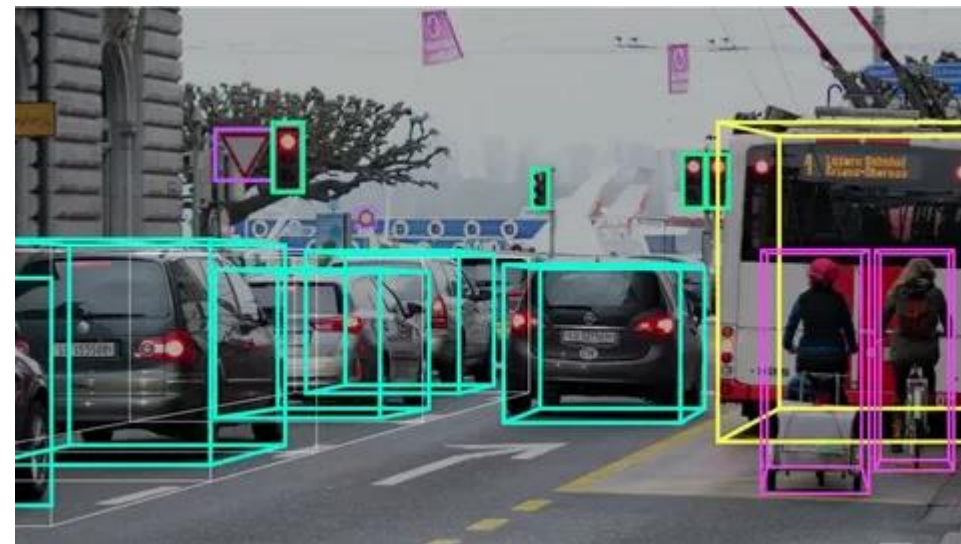
3D-сцена

Управление

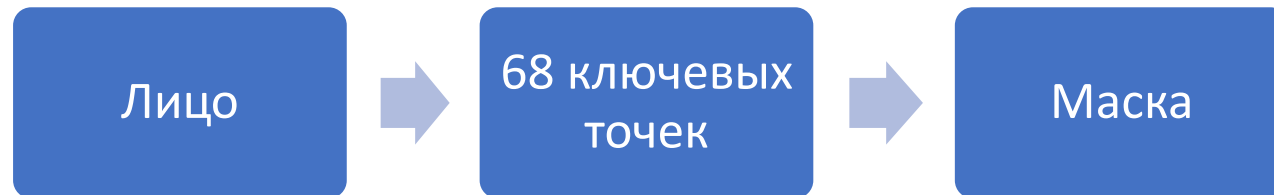
Что должен уметь автопилот:

- Распознавать пешеходов (днём/ночью/в тумане)
- Читать знаки и разметку
- Оценивать расстояние до объектов

Ошибка = жизнь человека → надёжность > 99.9999%



Дополненная реальность

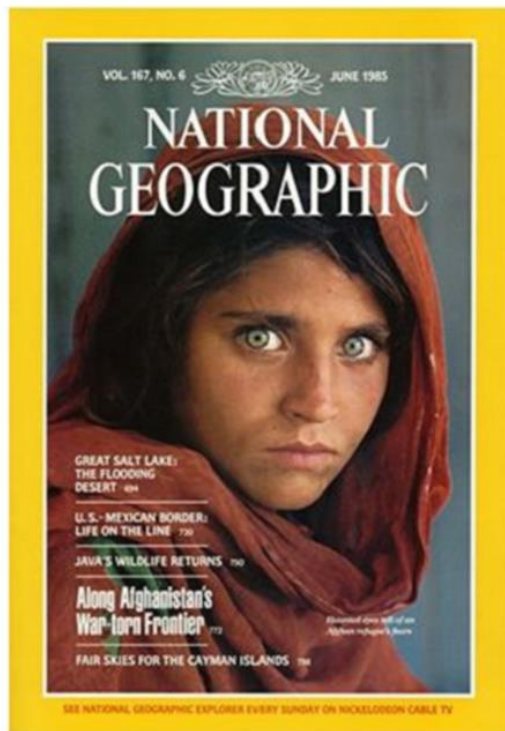


Как это работает:

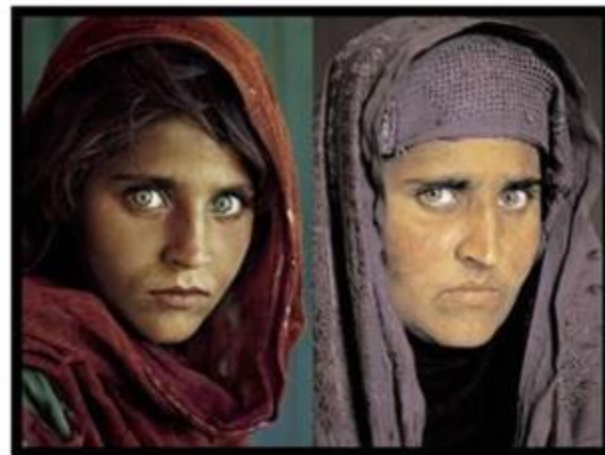
1. Найти лицо (Viola-Jones / CNN)
2. Найти ключевые точки (глаза, нос, рот)
3. Привязать 3D-маску к этим точкам



Распознавание

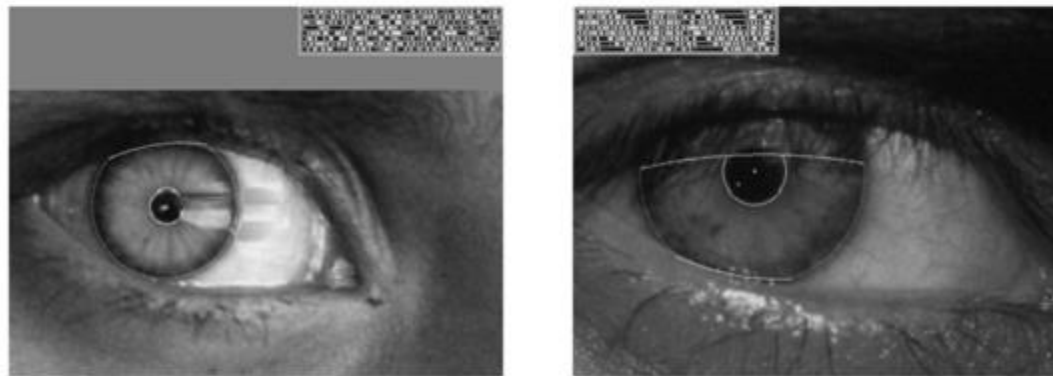


Незнакомка. Кто она?



Биометрия

"Как девушку из Афганистана идентифицировали по радужке глаза"



Source: S. Seitz

Безопасность



Технологии:

- Сравнение лиц (1:N поиск)
- Анализ поведения (трекинг)
- Обнаружение аномалий (прыжок на рельсы)

А это совсем просто сделать

Обнаружение изменяющихся областей видео, анализ их формы и динамики изменения (обычно для систем безопасности)

Робототехника

Задачи:

- SLAM (одновременная локализация и картографирование)
- Распознавание объектов (что брать?)
- Оценка позы (как взять?)



Генерация изображений

ШКАЛА
ВРЕМЕНИ:

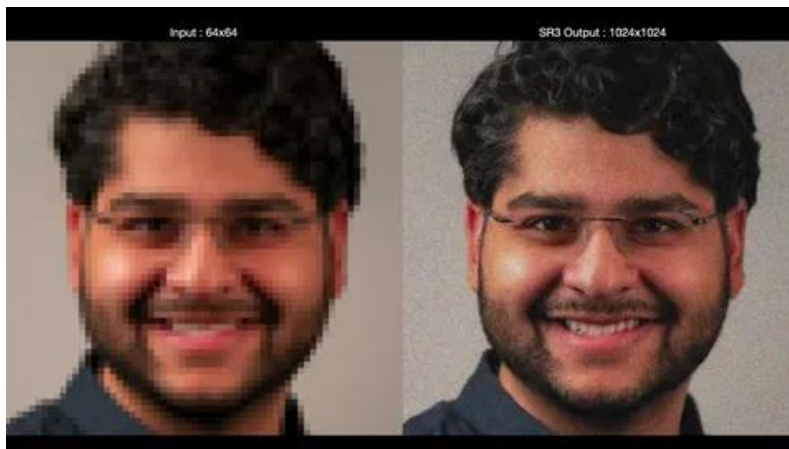
- 2014 (размытое лицо)
- 2021 (реалистичное)
- 2024 (фотореализм)

Модели:

GAN (2014) —
первые попытки

Diffusion (2021) —
Stable Diffusion,
Midjourney

Deepfakes —
этическая
проблема



Тренд: "Нарисовать" можно что угодно текстом



Распознавание текста — OCR

Этапы:

- Детекция текстовых областей
- Сегментация символов
- Классификация символов (CNN)

Где используется:

- сканеры документов,
- переводчик камеры



Анализ спутниковых снимков

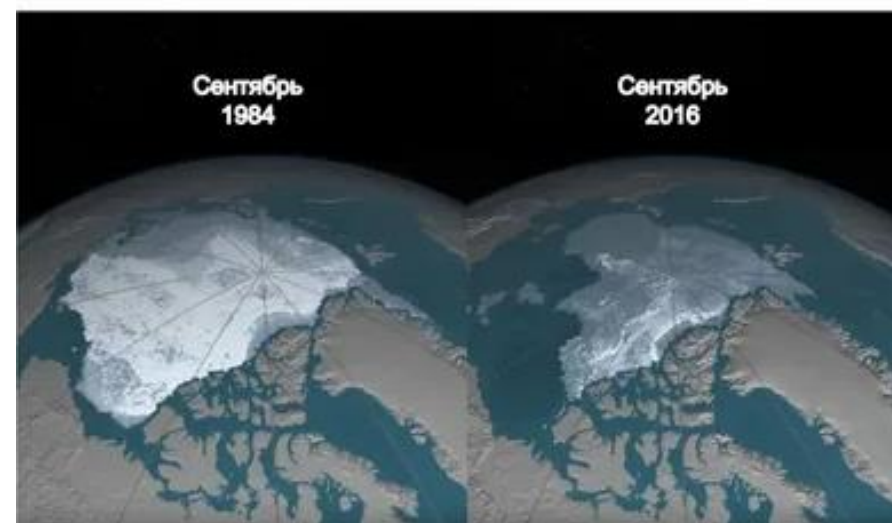
- ✓ Таяние ледников
- ✓ Вырубка лесов
- ✓ Урожайность полей



Задача: найти ИЗМЕНЕНИЯ на больших территориях



Таяние ледников, как следствие усиления парникового эффекта



Поиск конкретных объектов

запрос
(Тауэрский мост)



поиск



найжены
другие фото

Чем отличается от классификации/детекции:

- Классификация: "это мост"
- Детекция: "мост находится здесь"
- Поиск экземпляра: "это ТОТ ЖЕ мост, что на запросе?"



ГДЕ ПРИМЕНЯЕТСЯ:



Google Lens / Поиск по фото в AliExpress



Туристические приложения
(определение достопримечательностей)



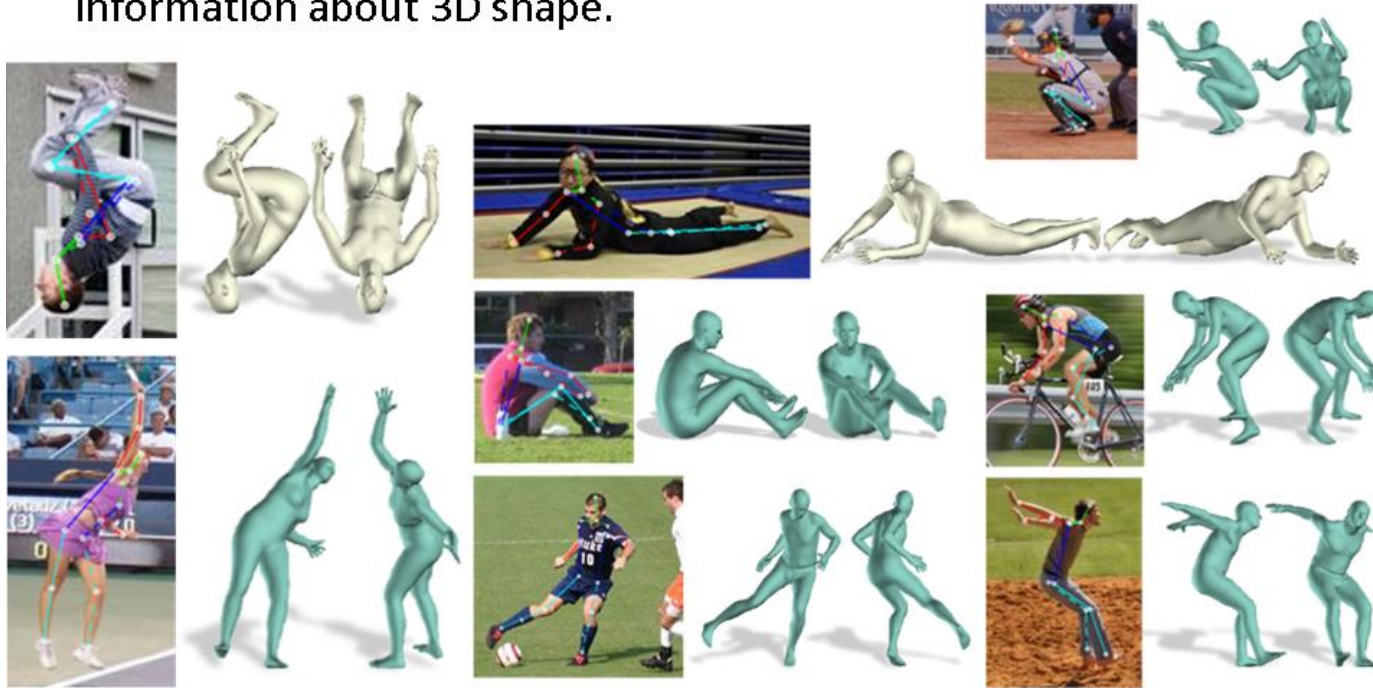
Правоохранительные органы (поиск человека в базе камер)



Социальные сети (защита авторских прав)

Human detection и Pose estimation

In addition to pose, 2D joints convey information about 3D shape.



Bogo, Kanazawa, Lassner, Gehler, Romero, Black, Keep it SMPL: Automatic estimation of 3D human shape and pose from a single image, ECCV'16

Слайд с презентации М. Black

Где нужно:

- Спорт (анализ техники)
- VR/AR (управление без контроллера)
- Медицина (реабилитация)

Ключевые точки: 17-25 суставов

3D модели и захват движения

Технологии:

- Структурированный свет (Kinect)
- Стереозрение (две камеры)
- LiDAR (лазерный сканер)

несколько
камер



3D-модель
человека



Применение: кино (Аватар), игры, биомеханика

Описание по изображению

curvy
feminine
attractive
hourglass



big
heavysset
stocky
short torso



lean
petite
skinny
small



short legs
short
short torso
small



Почему мы это изучаем?



КРУТО:

- Много практических задач
- Рост рынка → работа
- Сплав математики + программирование
- Много нерешённых проблем



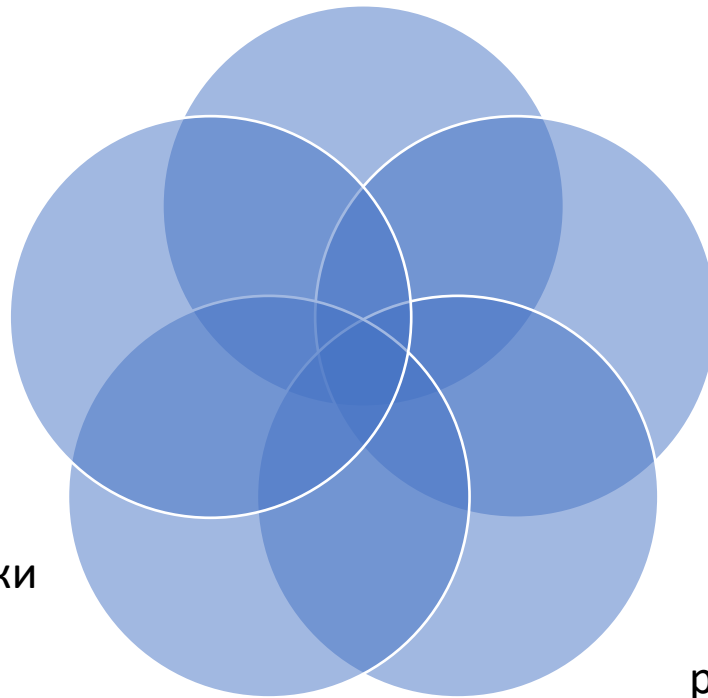
СЛОЖНО:

- 25% мозга — зрение
- ИИ-полная задача
- Семантический разрыв
- Нечёткая задача

Что нужно знать?

основы обработки
изображений

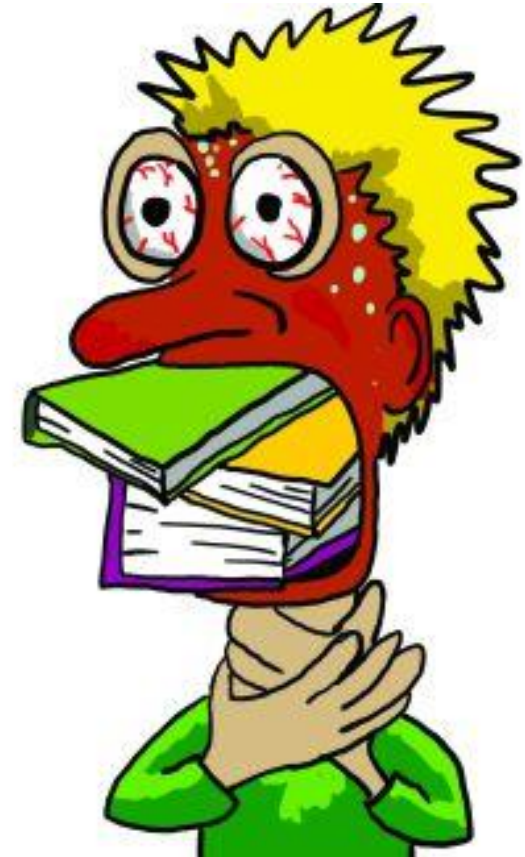
для верности
можно добавить к
списку оптику,
физику и теорию
обработки сигналов



немного математики
– **линейную**
алгебру,
геометрию,
дифуры и теорвер

методы машинного
обучения

принципы
распараллеливания
вычислений



Как называют то же самое?

Термин	Простыми словами	Пример
Фотограмметрия	Измеряем расстояния по фото	Карты
Компьютерное зрение	Понимаем сцену	Автопилот
Машинное зрение	Промышленные задачи	Контроль качества
Обработка изображений	Фотошопим	Убираем шум
Распознавание образов	Ищем закономерности	Распознаём цифры, буквы, отпечатки

Мы изучаем Computer Vision (самый широкий)

Объект CV — типы данных

 ЧТО МЫ АНАЛИЗИРУЕМ?

1.  2D изображения (фото, кадры)
2.  Видео (кадры + время)
3.  3D-данные (Kinect, LiDAR, стерео)



Мы видим

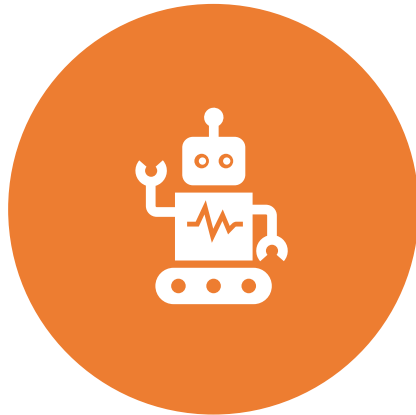
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

Компьютер видит

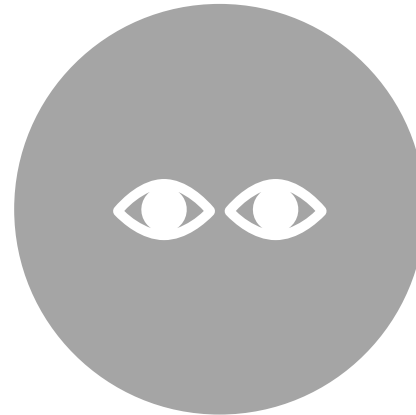
Source: S. Narasimhan

Итог: ВСЁ сводится к ЦИФРОВОМУ ИЗОБРАЖЕНИЮ (матрица пикселей)

Краткая история и эволюция подхода



НАЧАЛО ЦИФРОВОЙ ОБРАБОТКИ
ИЗОБРАЖЕНИЙ (ДО 1960)



КЛАССИЧЕСКОЕ КОМПЬЮТЕРНОЕ
ЗРЕНИЕ (1960-2000):



РЕВОЛЮЦИЯ ГЛУБОКОГО
ОБУЧЕНИЯ (2012 - НАСТОЯЩЕЕ
ВРЕМЯ):

История — до 1960



1920 год

- Первая передача изображения по кабелю между Лондоном и Нью-Йорком
- Время: неделя → 3 часа
- Качество: 5 градаций яркости



В 1929г. – уже 15 градаций

Это был первый шаг к цифровому изображению

Классическое компьютерное зрение (1960-2010):

Главная идея:
Инженеры
ПРИДУМЫВАЛИ
признаки вручную

[ПРИМЕРЫ:
выделение краёв,
углов, SIFT-точки]

Проблема:
"Семантический
разрыв"

Как описать КОТА
через линии и
углы? →
НЕВОЗМОЖНО

Прорыв: Viola-
Jones (2001) —
первый ML-
детектор лиц

Дамоклов меч, 1968

Вес: огромный (подвешивали к потолку)

Графика: только каркасные модели

Сейчас: Meta Quest — 500 грамм



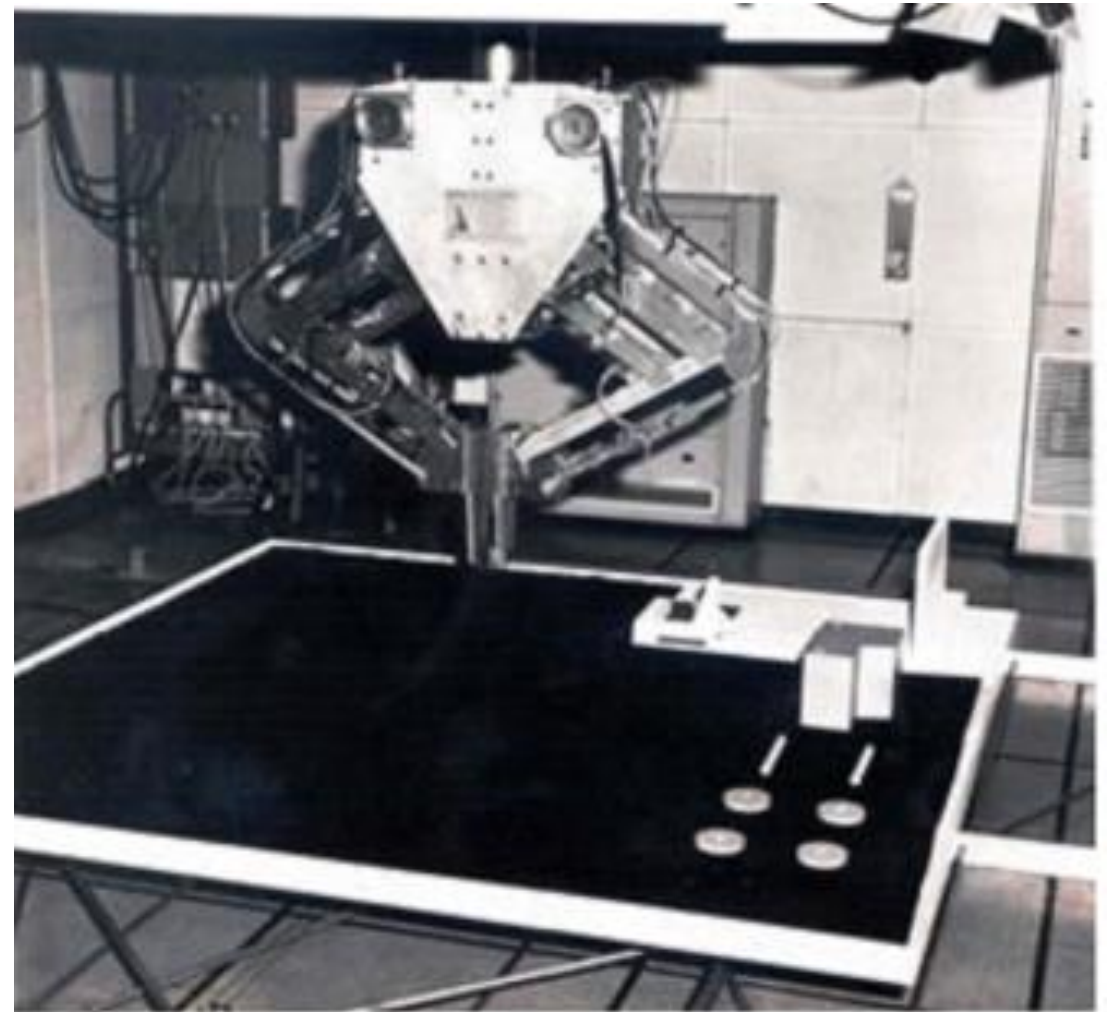
первый шлем виртуальной реальности

Freddy II, 1973

Мог собирать простые объекты

Ориентировался по камере

Медленно, но это был прорыв





Детектор лиц Viola-Jones (2001)



Viola-Jones:

- Работал в реальном времени (15 fps)
- Использовал простые признаки Хаара
- Первый успех ML в CV

Аватар (2009) — Захват движения

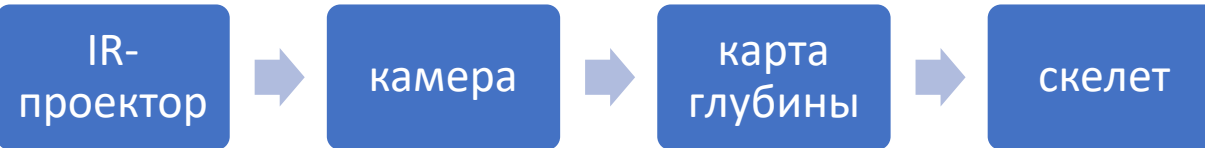
Как это сделали:

- Камеры снимали актёров в костюмах с маркерами
- CV восстанавливал 3D-позы
- Анимация накладывалась на 3D-модели

Прорыв в зрелищности!



Kinect (2010) — CV для ВСЕХ



Почему это важно:

- Первое массовое устройство с 3D-зрением
- Стоило \$150 (вместо \$10000)
- Показало: CV может быть в каждом доме



Первая потребительская система взаимодействия с компьютером с помощью жестов

Революция глубокого обучения (2012 – настоящее время)

Что изменилось:

- Раньше: инженеры ПРИДУМЫВАЛИ признаки (SIFT, HOG, SURF)
- Теперь: нейросети УЧАТСЯ признакам из данных

Ключевые вехи:

-  2012 — AlexNet (скачок точности на ImageNet: 74% → 84%)
-  2015 — ResNet (обучение глубиной >100 слоёв)
-  2017 — Attention / Transformer (из текста в изображения)
-  2021 — DALL-E / Stable Diffusion (генерация по тексту)
-  2024 — Sora (генерация видео)

Результат:

CV из лабораторий перешёл в ПРОДУКТЫ (Tesla, Face ID, Google Lens, Midjourney)

Главные вопросы + типы ответов

Вопрос 1: ЧТО ЭТО?

Ответ: СЕМАНТИЧЕСКАЯ информация

Задача: КЛАССИФИКАЦИЯ

"это кот", "это машина", "это пейзаж"

Вопрос 2: ГДЕ ЭТО?

Ответ: ^{Вопрос}МЕТРИЧЕСКАЯ информация (координаты, рамка)

Задача: ДЕТЕКЦИЯ

"кот в левом верхнем углу"

Вопрос 3: КАКОЙ ФОРМЫ/РАЗМЕРА?

Ответ: МЕТРИЧЕСКАЯ информация (3D-координаты)

Задача: 3D-РЕКОНСТРУКЦИЯ / МЕТРИЧЕСКОЕ ЗРЕНИЕ

"расстояние до кота — 2 метра"

ИТОГ: ВСЕ задачи CV сводятся к этим 3 вопросам!

Четыре главные задачи компьютерного зрения

КЛАССИФИКАЦИЯ	ДЕТЕКЦИЯ	СЕМАНТИЧЕСКАЯ СЕГМЕНТАЦИЯ	СЕГМЕНТАЦИЯ ЭКЗЕМПЛЯРОВ
[кот]	[рамки на улице]	[пиксели покрашены]	[разные люди разным цветом]
Что это?	"то и где?	Какой пиксель к какому классу относится?	Какой экземпляр?
Ответ: 1 метка на всё фото	ImageNet вокруг объектов	Ответ: маска (каждый пиксель)	Ответ: маска (каждый объект отдельно)
ImageNet	YOLO, R-CNN	Mask R-CNN	Mask R-CNN

Другие важные задачи

ТРЕКИНГ	3D-РЕКОНСТРУКЦИЯ	ОЦЕНКА ПОЗЫ	ГЕНЕРАЦИЯ
слежение за мячом/игроком	3D-модель по фото	скелет человека	DALL-E: "кот в космосе"
Где объект сейчас и куда он движется?	Как выглядит объект в 3D?	Как расположены части тела?	Можно ли создать новое изображение?
SORT, DeepSORT	COLMAP, NeRF	OpenPose, MMPose	Stable Diffusion DALL-E, Midjourney
Спорт, игры, видеонаблюдение	Кино, AR/VR, карты, робот	VR/AR, медицина, спорт, биомеханика	Искусство, дизайн, реклама

запрос
(Тауэрский
мост)



поиск



найжены
другие фото

Поиск экземпляра объекта

Чем отличается от классификации/детекции:

- Классификация: "это мост"
- Детекция: "мост находится здесь"
- Поиск экземпляра: "это ТОТ ЖЕ мост, что на запросе?"

КЛАССИЧЕСКИЙ ПОДХОД (до 2015):

Запрос → SIFT-точки → Сопоставление →
Ранжирование (ручные признаки,
геометрическая проверка)

СОВРЕМЕННЫЙ ПОДХОД (сейчас):

Запрос → CNN → Вектор (Embedding) →
Косинусное расстояние
(нейросеть сама учит признаки, сравнение за
миллисекунды)



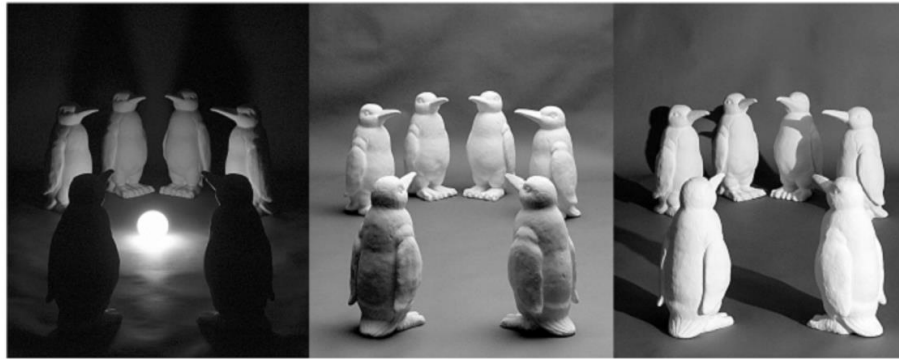
Сложности: ракурс, освещение, масштаб,
перекрытие, время суток

Почему зрение — это сложно?

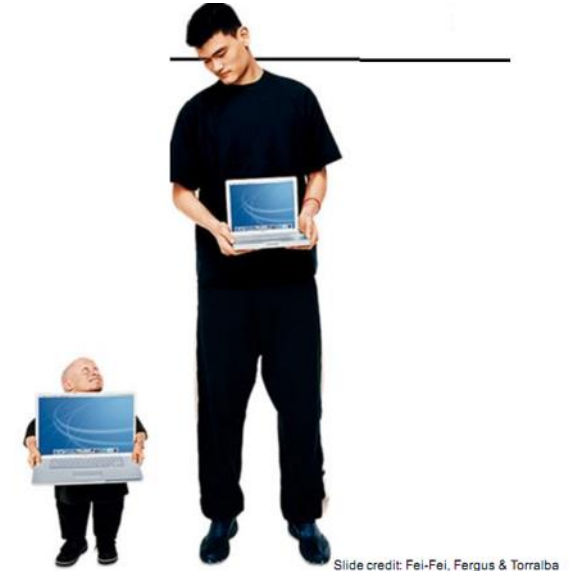
Точка наблюдения



Освещение



Масштаб



Один и тот же объект выглядит ПО-РАЗНОМУ!

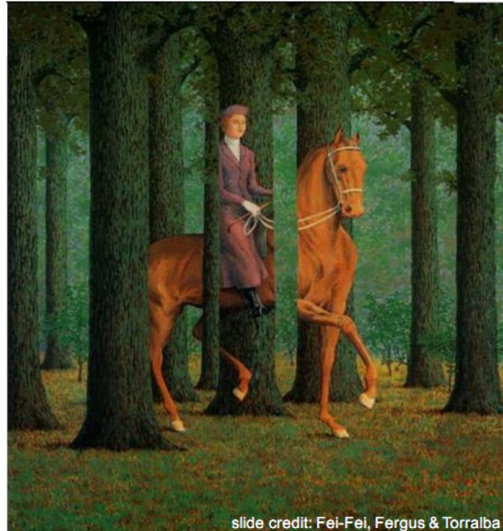
- ▶ Точка съёмки: вид сверху, снизу, сбоку — объект неузнаваем
- ▶ Освещение: день, ночь, тень, контражур — меняются цвета и тени
- ▶ Масштаб: крупный план vs общий план — теряются детали

Наш мозг справляется с этим легко. Компьютеру — сложно

Сложности — объект

Перекрытие

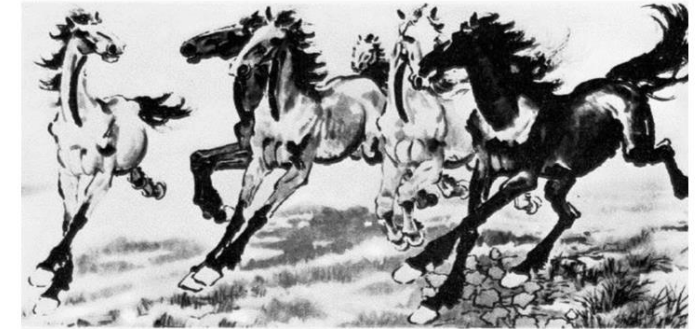
Magritte, 1957



Движение



Деформация



Xu, Beihong 1943

Slide credit: Fei-Fei, Fergus & Torralba

- ▶ Деформация: объект меняет форму (животные бегут/лежат)
- ▶ Перекрытие: объект частично скрыт другим объектом
- ▶ Движение: размытие, изменение позы за долю секунды

Внутриклассовая изменчивость



Класс — это не одна форма, а МНОЖЕСТВО!

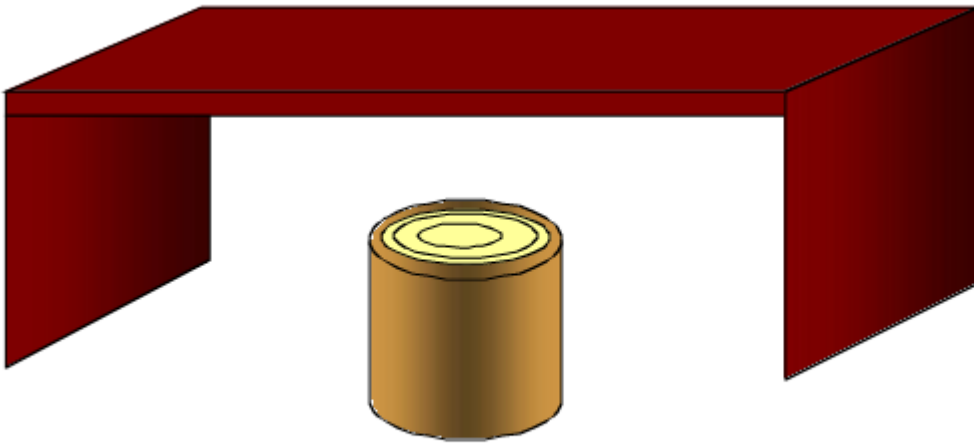
Внутриклассовая изменчивость: все стулья разные, все собаки разные, все люди разные

Алгоритм должен обобщать: видеть РАЗНОЕ как ОДНО

Slide credit: Fei-Fei, Fergus & Torralba

Сложности — сцена

Контекст



~~Полно~~

Стул

Маскировка



- Маскировка: объект сливается с фоном (защитная окраска)
- Контекст: объект может оказаться в неожиданном месте (кот на улице — это норма, кот в космосе — странно)

Контекст может КАК ПОМОЧЬ, так и ЗАПУТАТЬ!

Сложности или возможности?

- Изображение запутывает, но дает много подсказок
- Наша задача – интерпретировать подсказки



Image source: J. Koenderink

Подсказки

- Цвет
- Контур
- Текстура
- Тени и освещение
- Контекст (окружение объекта)



Априорные знания об объекте

- Размеры
- Контекст
- Типичный диапазон распределения конкретной характеристики
- Эталонные примеры

Резюме

- Зрение изначально нечеткая задача
- Разные 3D сцены могут давать одинаковое 2D изображение
- Необходимы априорные знания о структуре и свойствах мира
- Нужно сопоставлять наблюдения и априорные знания

Эволюция сложности задач в CV

1 LOW-LEVEL (Работа с "Сырыми" пикселями)

Задача: Сделать картинку КАЧЕСТВЕННЕЕ.

Вход: Картинка с шумом. Выход: Чистая картинка.

Инструменты: Фильтры (Гаусс, Собель, CLAHE).

Пример: Фотошоп.

Сегодня: Это готовые функции в OpenCV, мы их учим, чтобы понимать, как работает цифровой сигнал.

2 MID-LEVEL (Работа с "Частями" объектов)

Задача: Найти КЛЮЧЕВЫЕ ТОЧКИ и ГРАНИЦЫ.

Вход: Картинка. Выход: Координаты точек (края, углы).

Инструменты: Классика (SIFT, HOG) или Внутренние слои CNN.

Пример: Разметка лица точками.

Сегодня: Мы учим это, чтобы понять, как ИИ видит мир, а не просто верим в чёрный ящик.

3 HIGH-LEVEL (Работа со "Смыслом")

Задача: Ответить на вопрос пользователя.

Вход: Картинка. Выход: Метка, Рамка, Маска, Текст.

Инструменты: Глубокое обучение (CNN, Transformers).

Пример: Автопилот, Диагностика.

Сегодня: Это цель вашей будущей работы.

А зачем это нам сейчас? (Реалии рынка)

Low-level (Обработка сигналов)

-  Кто делает: Embedded-инженеры, ML-инженеры на старте.
-  Где: Камера в телефоне (HDR, Ночной режим).
-  В курсе: Основа. 1-я часть курса.

Mid-level (Извлечение признаков)

-  Кто делает: Уже почти никто (руками).
-  Где: Внутри нейросетей (как "внутреннее зрение" ИИ).
-  В курсе: 2-я часть. Чтобы понимать, как учатся сети.

High-level (Принятие решений)

-  Кто делает: Сеньоры, CV-инженеры, Исследователи.
-  Где: Автопилоты, Медицина, Генерация (DALL-E).
-  В курсе: 3-я часть. Ваша итоговая цель.

Заключение

1 ЗРЕНИЕ — НЕЧЁТКАЯ ЗАДАЧА

Разные 3D-сцены → одинаковые 2D-картинки

Пример: проекция 3D-мира на 2D-сенсор

2 НУЖНЫ АПРИОРНЫЕ ЗНАНИЯ О МИРЕ

Мы не начинаем с нуля — мы ЗНАЕМ законы физики, размеры объектов, контекст

3 ГЛАВНЫЙ НАВЫК:

СОПОСТАВЛЯТЬ НАБЛЮДЕНИЯ и АПРИОРНЫЕ ЗНАНИЯ

4 CV — ЭТО СПЛАВ:

- 📐 Математика (линал, геометрия, теорвер)
- 💻 Программирование (Python, C++, оптимизация)
- 🧠 Машинное обучение (нейросети, классика)
- ⚡ Высокопроизводительные вычисления (GPU, CUDA)



Литература Модуль 1 (Введение и классика):

1. Р. Гонсалес, Р. Вудс. Цифровая обработка изображений. R.Gonzalez, R. Woods Digital Image Processing
2. Л. Шапиро, Дж. Стокман. Компьютерное зрение Shapiro L., Stockman G. Computer Vision
3. Дэвид Форсайт, Жан Понс. Компьютерное зрение. Современный подход David A. Forsyth, Jean Ponce. Computer Vision: A Modern Approach
4. **R.Szeliski «Computer vision: Algorithm and applications» 2022 Электронная версия доступна бесплатно на [сайте автора http://szeliski.org/Book/](http://szeliski.org/Book/)**
5. Фундаментальный учебник по машинному обучению (Базис для CV): Kevin P. Murphy. "Probabilistic Machine Learning: An Introduction" (2022) и "Probabilistic Machine Learning: Advanced Topics" (2023).

Литература Модуль 2 (Глубокое обучение):

1. Simon J.D. Prince. "Understanding Deep Learning" (2023)
2. Aurelien Geron. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" (3rd ed. 2022). Главы, посвященные CNN и обработке изображений.