

# Введение в компьютерное зрение

# Содержание

<b>1</b>	<b>Введение</b>	<b>2</b>
<b>2</b>	<b>Примеры практического применения</b>	<b>7</b>
2.1	Медицина . . . . .	7
2.2	Промышленность . . . . .	9
2.3	Системы видеонаблюдения . . . . .	10
2.4	Биометрия . . . . .	10
2.5	Индексирование и поиск изображений . . . . .	11
2.6	Ещё приложения . . . . .	12
<b>3</b>	<b>Задачи и наборы данных</b>	<b>13</b>
3.1	Задачи . . . . .	13
3.2	Наборы данных . . . . .	15

## 1 Введение

Наверное, сложно переоценить значимость зрительного восприятия для человека. Зрение необходимо для выполнения практически всех задач в нашей жизни, как повседневных, так и профессиональных. Поэтому, неудивительно, что задача компьютерного зрения столь популярна и востребована. Способность компьютера "видеть" анализировать визуальную информацию - один из самых распространённых примеров искусственного интеллекта.

Поскольку человек так хорошо и эффективно воспринимает и обрабатывает визуальную информацию, данные об окружающем нас мире мы часто регистрируем и сохраняем в виде цифровых изображений и видео. Сегодня камера есть в каждом мобильном телефоне, камеры встраиваются в бытовую и профессиональную технику. Мы можем снимать под водой, с воздуха, из космоса. Камеры вводят в человеческое тело, прикрепляют к диким животным. Объёмы мультимедиа информации растут с каждой секундой. Чтобы с этой информацией сделать что-то полезное, нужно научить компьютер обрабатывать и анализировать видео-данные, решать задачи, специфичные для той или иной области.

Давайте посмотрим на некоторые примеры областей применения компьютерного зрения.

- Диагностика опухоли головного мозга по МРТ-снимкам (изображениям, полученным при помощи магнитно-резонансной томографии).
- Автоматическое обнаружение подозрительных объектов с камер видеонаблюдения.
- Большой спектр задач для автономного управления автомобилем: распознавание разметки на дороге, дорожных знаков, других участников движения, и многие другие.
- Анализ спортивных матчей - например, распознавание точной позиции мяча.
- Биометрические системы, как, например, разблокировка телефона по лицу владельца.
- Перевод вывесок, указателей, меню в ресторане с иностранного языка.
- Визуальные системы контроля качества - одно из самых ранних и распространённых применений систем компьютерного зрения. Компьютер анализирует все ли составляющие детали на нужном месте.

- Ещё один пример - автоматизированные магазины Amazon Go, без касс и продавцов. При входе каждый покупатель регистрируется при помощи приложения на мобильном телефоне, и дальше камеры следят за ним по всему магазину, отмечая, что данный покупатель берет с полки и складывает себе в сумку.
- И последний пример на этом слайде - применение компьютерного зрения в сельском хозяйстве. Роботы и дроны на полях - реальность сегодняшнего дня. Роботы определяют степень зрелости урожая, предсказывают его объёмы, распознают наличие заболеваний, используются для автоматической прополки, сбора урожая.

Это всего лишь несколько примеров использования возможностей компьютерного зрения. Мы вернёмся к более подробному обсуждению этих и других приложений чуть позже. А сейчас давайте определим, что же такое компьютерное зрение.

Компьютерное зрение — это область науки, которая занимается алгоритмами и системами фиксации и анализа изображений и извлечения информации из видео-данных. Или, иными словами, машинами, которые могут видеть мир и воспринимать его так же, как человек. По аналогии со зрительной системой человека, система компьютерного зрения состоит из сенсора, фиксирующего изображение, и устройства извлечения информации из изображения. Камеры - это глаза системы компьютерного зрения, а компьютер и алгоритмы анализа и обработки изображений - мозг. На входе система компьютерного зрения, как и зрительная система человека, фиксирует изображение, а на выходе должна быть способна извлечь семантическую информацию, интерпретировать изображение. Например, понять, что это - изображение сада, что какая-то часть изображения соответствует мостику, а какая-то - деревьям, цветам. Людям это кажется простой задачей. Но научить этому компьютер не просто.

В данном курсе мы будем уделять основное внимание второй, и на мой взгляд основной, компоненте системы компьютерного зрения - алгоритмам и методам анализа изображений и видео.

В литературе можно встретить разные определения компьютерного зрения. Для примера приведем три.

- Computing properties of the 3D world from one or more digital images (by Trucco and Veri). В этом определении подразумевается, что вне зависимости от того, есть мы или нет, существует какой-то окружающий мир и его изображения, анализируя которые мы хотим что-то о нём понять. Это определение через имитацию зрительной системы человека. У нас есть сенсор — глаза, у нас есть преобразующее устройство — мозг, и мы воспринимаем мир при помощи анализа тех картинок, которые видим.

- Make useful decision about real physical objects and scenes based on the sensed images by Shapiro). Это определение ближе к робототехнике. Мы хотим принимать решения и делать выводы о реальных объектах вокруг нас на основе изображений, которые уловили сенсоры. К примеру, это определение идеально подходит под описание того, что делает робот-пылесос. Он принимает решение о том, куда ему дальше ехать и какой угол пылесосить на основании того, что он видит.
- The construction of explicit, meaningful decisions of physical objects from images (by Ballard and Brown). Наиболее общее определение из трех. Если опираться на него, мы хотим просто описывать явления и объекты вокруг нас на основе анализа изображений.

Подытоживая, можно сказать, что компьютерное зрение сводится к извлечению значимой информации из цифровых изображений.

Что из себя представляет цифровое изображение? Существует несколько типов цифровых изображений, самым распространенным из которых являются растровые изображения. Растровое черно-белое изображение - это двумерный массив чисел, или матрица пикселей. Каждый элемент массива кодирует интенсивность света в соответствующей пространственной позиции. Ноль соответствует чёрному цвету - у чёрного цвета нулевая интенсивность света. Как по массиву чисел компьютеру определить, что данное изображение - это портрет Эйнштейна? Это непросто.

Итак, задача компьютерного зрения - извлечь значимую информацию из цифрового изображения. Какую значимую информацию можно извлечь из изображения?

- Первое - семантическую информацию. Ответить на вопрос, что изображено на картинке, понять семантику сцены. Распознать и классифицировать объекты, определить их свойства и отношения. Например, распознать, что на фотографии слева изображён мужчина на мотоцикле с двумя собаками. Идентифицировать личность этого мужчины, породу собак, модель мотоцикла.

Помимо семантической информации, важное значение имеет также метрическая информация об объектах. Определение размеров объектов, расстояния между ними и их взаимного положения - важная подзадача компьютерного зрения. Метрическая информация необходима в робототехнике и системах навигации, в задачах 3D моделирования по снимкам.

Область компьютерного зрения тесно связана с многими другими дисциплинами. На сегодняшний день, большинство задач компьютерного зрения решается при помощи алгоритмов машинного обучения. Компьютерное

зрение также связано с когнитивной наукой, занимающейся моделированием когнитивных систем; с нейронауками, занимающимися изучением нейронных процессов. Компьютерное зрение широко используется в робототехнике, система компьютерного зрения - неотъемлемая компонента большинства роботов. Информационный поиск, анализ и синтез речи - тоже связанные дисциплины.

Обработка изображений, анализ изображений и компьютерная графика - очень тесно связанные области. Некоторые относят все три дисциплины к области компьютерного зрения.

В рамках курса мы будем говорить об алгоритмах обработки изображений. Именно они используются, когда мы повышаем контрастность, удаляем цвет или шум, применяем фильтры и т.д. Любые преобразования изображений, когда на входе и выходе алгоритма изображение — это обработка изображений.

Для анализа изображений характерно, имея на входе изображение, на выходе получить некую модель или набор признаков. То есть по исходному изображению вычислить некоторые числовые параметры, которые описывают это изображение. Например, гистограмму распределения уровней серого цвета, или координаты и классы объектов. В анализе изображений как результат мы получаем набор признаков.

Ещё одна смежная область — компьютерная графика, в которой генерируют изображение по модели или набору признаков.

Все это невозможно без использования знаний и алгоритмов из ещё целого ряда областей, таких как распознавание образов и статистике. Также можно сказать, что анализ изображений — это частный случай анализа данных, и под-область искусственного интеллекта. К смежной дисциплине можно отнести и нейропсихологию — для того чтобы понять, какие у нас есть возможности и как устроено восприятие картинок, полезно понимать, как эти задачи решает наш мозг.

Отсчёт истории компьютерного зрения, как области исследований, принято начинать с 60-х годов 20-го века. В качестве первой точки отсчёта часто приводят студенческий летний проект в группе искусственного интеллекта MIT 1966 года. Первые попытки обнаружения краёв и восстановления 3D-формы для простых объектов, "мира игрушечных кубиков" (Roberts 1965).

В 1970-х годах сформировался основной понятийный аппарат в области обработки изображений. Велись активные исследования алгоритмов обнаружения краёв, были предложены первые алгоритмы сопоставления стерео и анализа оптического потока.

В 80-х годах большое внимание уделялось математическому аппарату анализа изображений. Многие работы используют Марковские случайные поля для формулировки задач компьютерного зрения, что позволяет рассмат-

ривать эти задачи как задачи оптимизации. Сформировалась теория уровней представления изображений. Одной из основных вех этого этапа является книга Дэвида Марра «Зрение. Информационный подход к изучению представления и обработки зрительных образов».

В 90-х годах продолжают активно развиваться многие направления компьютерного зрения: определение формы объекта по движению, методы на основе оптического потока, определение трехмерной формы объекта по изображениям с разных ракурсов, сегментация изображений (алгоритмы *normalized cuts* и *mean shift*). Возможно самым значимым результатом этого десятилетия стало близкое взаимодействие областей анализа изображений и компьютерной графики - накладывание текстуры с реального изображения на 3D модель, построение трехмерных моделей объектов по снимкам с разных ракурсов, и автоматическая генерация новых ракурсов этих объектов. Первые примеры использования статистических методов - использование метода главных компонент для распознавания лиц.

В 2000-е годы связь между компьютерным зрением и компьютерной графикой продолжила развиваться: построение панорам по нескольким снимкам, создание HDR изображений примеры такого взаимодействия. Также на это десятилетие приходятся существенные сдвиги в алгоритмах распознавания, классификации изображений, информационного поиска. Было предложено большое количество методов извлечения признаков из изображений, которые затем подавались на вход алгоритмам машинного обучения для решения задач распознавания и классификации.

В последнем десятилетии несомненно самой значительной вехой в компьютерном зрении стало использование алгоритмов глубокого обучения (*deep learning*). Рост объёмов данных и развитие аппаратного обеспечения позволили использовать эти вычислительно сложные алгоритмы, которые за несколько лет привели к большому прогрессу во всех задачах компьютерного зрения. До методов глубокого обучения системы компьютерного зрения в лучшем случае работали в контролируемой среде, когда можно отрегулировать освещение, камеру, положение объекта, и т.п. В естественных условиях (неконтролируемой среде) был огромный разрыв в точности распознавания. Однако за последние несколько лет удалось добиться большого прогресса благодаря использованию методов глубокого обучения.

Но несмотря на прогресс последних лет, мы до сих пор водим автомобили самостоятельно, и даже к системам видеонаблюдения часто должен быть приставлен человек. Почему? Одна из ключевых проблем — семантический разрыв.

Человек, глядя на картинку, понимает её семантику. Компьютер же понимает цвет пикселей, умеет выделить текстуру и в конечном счёте отличить кирпичную стену от ковра и распознать на фотографии человека, но опреде-

лить намерения человека по выражению его лица, позе машина ещё может.

Помимо семантического разрыва, есть ещё ряд сложностей, с которыми сталкиваются разработчики систем компьютерного зрения.

- Один и тот же объект будет может выглядеть очень по-разному при различном освещении
- Изображение может быть очень разным в зависимости от позы модели
- Фоновый шум и маскировка
- Перекрытия
- Внутриклассовая изменчивость
- Ракурс, точка съёмки

## 2 Примеры практического применения

Мы обсудили, что такое компьютерное зрение и почему это сложно. Давайте теперь рассмотрим несколько областей применения анализа изображений и видео чуть более детально.

### 2.1 Медицина

Одной из основных (и самых первых) областей применения анализа изображений является медицина. Первые попытки автоматически анализировать медицинские снимки при помощи компьютера были предприняты на заре компьютерного зрения. И сегодня эта область остаётся одной из самых востребованных.

Медицинские изображения являются важным источником информации о внутренних органах человека. Они активно используются для диагностики, при планировании и проведении операций, для индивидуального моделирования последствий операций.

Существует множество способов получения медицинских снимков. Изображения получают при помощи электромагнитного излучения, рентгена, ультразвука, компьютерной томографии (КТ), магнитно-резонансной томографии (МРТ). Такое многообразие также говорит о том, насколько важной является визуальная информация в медицине. Большинство медицинской техники, которая используется для получения медицинских снимков, обладает встроенными возможностями автоматической обработки и анализа изображений.



Компьютерный анализ медицинских изображений применим буквально во всех областях медицины. Обнаружение переломов на рентгеновских снимках костей, анализ снимков грудной клетки для диагностики лёгочных заболеваний (например, пневмонии), анализ кровеносных сосудов при помощи ангиографии, обнаружение раковых клеток, диагностика мозга, моделирование зубов. Широкий круг задач в анализе медицинских изображений связан с дерматологией. Список можно продолжать очень долго.

На слайде примеры рентгеновских изображений.

- Рентген кисти руки.
- Рентген грудной клетки.
- Ангиограмма аорты. Для получения ангиограммы в артерию или вену вводят катетер, который продвигается в вдоль сосуда, пока не достигнет обследуемой зоны. Затем через катетер впрыскивается контрастное вещество, хорошо поглощающее рентгеновские лучи. Благодаря этому усиливается контраст рентгеновского изображения, что позволяет лучше определять аномалии кровоснабжения или места закупорки сосудов.
- Компьютерная томограмма головы.

Какого типа алгоритмы обычно используются для анализа медицинских снимков? По снимкам можно выявлять аномалии — отличается ли снимок данного пациента от снимка здорового человека?

При помощи алгоритмов классификации можно проводить диагностику заболеваний. Если у вас есть база снимков пациентов и известно, что первая аномалия встречается у здоровых людей, а вторая означает, что человек болен раком, то, основываясь на подобию изображений, можно помочь врачам с диагностикой заболеваний.

Анализ изображений используется для индивидуального моделирования организма и предсказания последствий лечения. Хотя мы все и похожи, каждый организм устроен индивидуально. Например, при помощи моделирования зубов пациент может увидеть, как именно его зубы будут меняться от недели к неделе, если поставить брекеты. Другой пример - моделирование последствий операции шунтирования кровеносного сосуда. Если человеку требуется соединить разорванный сосуд шунтом, то определить, где его ставить можно смоделировав по снимку кровеносную систему именно этого пациента и «вставив» шунт в этой модели. Так мы получим возможность увидеть, как изменится кровоток, и предсказать, как пациент будет себя чувствовать при разных вариантах проведения операции.

Анализ изображений незаменим для роботизированной хирургии, для неинвазивных и малоинвазивных операций, которые все больше и больше

заменяют открытые операции. Например, операции на мозге без вскрытия черепа - удаление опухолей при помощи сфокусированного ультразвука. Или различные операции с использованием эндоскопических инструментов и дистанционного управления этими инструментами. При таких операциях инструменты вводятся в тело через кожу или через анатомические отверстия. Операция проводится при непрямом наблюдении внутренних органов через инструменты, введенные в тело. И алгоритмы обработки и анализа изображений помогают хирургам видеть и оперировать.

## 2.2 Промышленность

Ещё одна очень востребованная область применения анализа изображений - системы контроля качества на производстве. Для данной прикладной задачи часто используют изображения, регистрируемые в видимом диапазоне (т.е. при помощи обычных оптических камер), и изображения, получаемые при помощи электронного микроскопа.

Компьютерное зрение используется для автоматического контроля выпускаемой продукции давно. Все примеры на слайде из книги 1992 года издания.

- Контроль наличия всех компонент на плате контроллера дисководов CD-ROM. Чёрный квадрат в правой верхней части демонстрирует отсутствие микросхемы.
- Контроль наличия всех таблеток в упаковке.
- Контроль за уровнем жидкости в бутылках.
- Контроль за качеством пластмассы - на данном изображении пластмассовая деталь с недопустимым количеством пузырьков воздуха. Обнаружение подобных аномалий составляет важную область промышленного контроля различных материалов, например тканей и дерева.
- Контроль за качеством кукурузных хлопьев - по цвету и наличию подгоревших хлопьев.

Первые алгоритмы работали на основе большого количества эвристик и правил. Алгоритмы создавались под определённое расположение камеры, под определённый продукт. Так что создание и внедрение таких автоматизированных систем контроля качества было оправдано только для больших производственных линий. Сегодня все больше и больше систем контроля качества используют машинное обучение и искусственные нейронные сети.

## 2.3 Системы видеонаблюдения

Ещё одна прикладная область, где необходимы алгоритмы анализа изображений и видео, это системы видеонаблюдения. Сегодня камеры наблюдают за нами повсюду: в аэропортах, на вокзалах, в метро, в магазинах, просто на улицах. При таком объёме видеоданных без автоматизированных систем анализа видеоряда не обойтись. Большинство камер устанавливается для обеспечения безопасности. Системы видеораспознавания позволяют оперативно распознать скопления людей, идентифицировать и зафиксировать факты их неадекватного движения (драки, падения, хаотичное движение, резкое ускорение, пересечение запрещённой зоны), обнаружить в контролируемой зоне оставленные или исчезнувшие предметы, выявить задымление или открытый огонь. Задачи по обнаружению подозрительных предметов или подозрительного поведения не так просты. Их сложность в том, что часто нельзя заранее дать описание того, что является подозрительным предметом. Невозможно предоставить все возможные примеры аномального поведения. Поэтому часто в таких системах моделируют нормальное поведение, и распознавание аномалий устроено на выявлении отклонений от нормы.

Помимо охранных целей, системы видеонаблюдения часто используются для изучения поведения людей. Например, при помощи установленных камер в магазине можно определить наиболее популярные отделы магазина и использовать эту информацию для более удобной планировки расположения товаров.

Системы видеонаблюдения позволяют автоматически собирать плату за проезд по платному участку дороги, штрафовать за превышение скорости, наблюдать за животным миром. Список задач бесконечен!

## 2.4 Биометрия

Ещё одним направлением, активно использующим компьютерное зрение, являются биометрические системы идентификации личности. Такие системы основаны на измерении уникальных биологических и физиологических характеристик, индивидуальных для каждого человека. Иногда эти характеристики ещё называют биологическим кодом человека.

Преимущество биометрических систем идентификации, по сравнению с традиционными (например, PIN-кодовыми системами или системами доступа по паролю), заключается в том, что идентифицируется не какой-то внешний предмет, принадлежащий человеку, а собственно сам человек. Анализируемые характеристики неразрывно связаны с человеком, их невозможно потерять, передать, забыть и крайне сложно подделать. К тому же эти характеристики практически не подвержены износу и не требуют замены или восстановления.

Среди основных биометрических технологий, работающих на основе компьютерного зрения - распознавание личности по отпечаткам пальцев, геометрии лица, геометрии кисти руки, радужной оболочке или сетчатке глаза, геометрии рисунка вен, по подписи и др.

Первые биометрические системы были достаточно дорогими и требовали значительных вычислительных ресурсов. Сегодня мы разблокируем компьютеры и телефоны по отпечатку пальца или геометрии лица.

Приведу ещё одну нашумевшую в СМИ историю, связанную с задачами биометрии. В Википедии даже есть статья, посвящённая этой истории. В 80-х годах фотограф National Geographic сфотографировал неизвестную афганскую девочку. Фотография стала очень известной после того, как появилась на обложке журнала National Geographic в июне 1985 года. Поскольку личность девочки была неизвестна, фотографию называли просто «Афганской девочкой» (Afghan Girl). Фотографию иногда сравнивают с портретом Моны Лизы кисти Леонардо да Винчи и называют «Афганской Моной Лизой». На протяжении 1990-2000 годов журналист предпринимал несколько попыток узнать имя девочки. Несколько женщин ошибочно объявляли себя «Афганской девочкой»; многие молодые люди называли ее своей женой. Через 17 лет после исходной фотографии фотограф нашел ее в Афганистане и её личность была подтверждена при помощи биометрии, показавшей полное соответствие радужной оболочки глаза изображению на фотографии.

## 2.5 Индексирование и поиск изображений

Индексирование и поиск изображений ещё одна важная задача, для решения которой требуются алгоритмы анализа изображений. Коллекции и архивы изображений бывают разные.

- Персональные. Например, в отпуске человек может сделать пару тысяч фотографий, с которыми потом нужно что-то делать.
- Профессиональные. Они насчитывают миллионы фотографий. Здесь тоже есть необходимость как-то их организовывать, искать, находить то, что требуется.
- Коллекции репродукций. Это тоже миллионы изображений. Сейчас у большого количества музеев есть виртуальные версии, для которых оцифровываются репродукции. Пока утопичная задача — поиск всех репродукций одного и того же автора. Человек по стилю может предположить, что видит, допустим, картины Сальвадора Дали. Было бы здорово, если бы этому научилась и машина.

Как упорядочить все эти коллекции? Какие ещё задачи можно решать? Можно построить систему навигацию по коллекциям, классифицируя их по

темам. Отдельно складывать мишек, отдельно слонов, отдельно апельсины — так, чтобы пользователю потом было удобно ориентироваться по этой коллекции.

Ещё одна задача — это поиск дубликатов. В двух тысячах фотографий из отпуска неповторяющихся значительно меньше. Мы любим экспериментировать, снимать с разной выдержкой, фокусным расстоянием и т.д., что в итоге нам даёт большое количество "нечётких" дубликатов, т.е. примерных копий. Кроме того, поиск по дубликатам может помочь обнаружить незаконное использование вашей фотографии, которую вы однажды могли выложить в интернете.

Ещё одна интересная задача — выбор лучшей фотографии. С помощью алгоритма можно понять, какая картинка больше всего понравится пользователю. Например, если это портрет, лицо должно быть освещено, глаза открыты, изображение должно быть чётким и т.д. В современных фотоаппаратах есть такая функция.

Ещё одна задача поиска — создание коллажей, т.е. подбор фотографий, которые будут хорошо смотреться рядом.

## 2.6 Ещё приложения

Есть ещё большое количество прикладных задач, где используется анализ изображений.

- Системы слежения и целенаведения в военной промышленности. Очевидными примерами являются обнаружение вражеских солдат и транспортных средств и управление ракетами. Наиболее совершенные системы управления ракетами посылают ракету в заданную область, вместо конкретной цели, а выбор целей производится, когда ракета достигает заданной области, основываясь на получаемых видеоданных.
- Автономные транспортные средства, как для военного применения, так и для автомобилестроения в целом. Уровень автономности изменяется от полностью автономных (беспилотных) до транспортных средств, где системы, основанные на компьютерном зрении, поддерживают водителя или пилота в различных ситуациях. Полностью автономные транспортные средства используют компьютерное зрение для навигации, то есть для получения информации о месте своего нахождения, для создания карты окружающей обстановки, для обнаружения препятствий. Примерами таких систем могут быть система предупредительной сигнализации о препятствиях на машинах, системы автономной посадки самолётов. Сегодня активно разрабатываются и тестируются системы автономного управления автомобилем.

- Компьютерное зрение также активно используется в киноиндустрии при создании спецэффектов, для совмещения игры живых актёров с компьютерной анимацией. Сегодня создатели фильмов могут записать эмоции настоящего актёра и наложить их на нарисованного персонажа или на лицо другого актёра. Компьютерное зрение сегодня - неотъемлемая часть абсолютного большинства компьютерных игр, приложений виртуальной и дополненной реальности.
- Зондирование земной поверхности с использованием видимого и инфракрасного диапазонов спектра - ещё одна важная область применения алгоритмов анализа изображений. Изображения населённых пунктов часто (и давно) используются для оценки роста численности населения, динамики загрязнения окружающей среды. По изображениям со спутников автоматически составляют карты местности. Наблюдение за погодой и составления прогнозов также является важным применением спутниковых изображений. Обнаружение и предсказание ураганов, обнаружение лесных пожаров, и даже изменение уровня поверхности земли как показатель истощения подземных вод - все это делается сегодня при помощи анализа спутниковых снимков земли.
- Оцифровка печатных документов, обнаружение и распознавание текста - ещё один пример применения алгоритмов анализа изображений. Сегодня существуют приложения, которые позволяют в режиме реального времени переводить указатели и таблички с иностранного языка.
- И, конечно, компьютерное зрение - неотъемлемая компонента системы навигации и управления роботов.

## 3 Задачи и наборы данных

### 3.1 Задачи

Мы рассмотрели большое число различных прикладных задач, в которых используется компьютерное зрение. Как же можно научить компьютер "понимать" изображения, чтобы можно было решать все эти задачи?

Давайте посмотрим на одно видео. Ваша задача - посчитать сколько раз передают мяч друг другу игроки в белом.

Заметили ли вы, что среди игроков прошёл человек, наряженный в костюм медведя? Давайте посмотрим видео ещё раз.

После того, как я вам сказала про медведя, Вы его, конечно заметили. Потому что Вы знали, что искать.

Это видео демонстрирует, что даже зрительная система человека обычно в каждый конкретный момент времени решает какую-то одну поставленную задачу, игнорируя сигналы, которые не относятся к решению этой задачи. Точно также при разработке приложений компьютерного зрения люди обычно решают какую-то одну конкретную поставленную задачу. Это существенно проще, чем пытаться решить абстрактную задачу "научить компьютер понимать изображение".

Какие же конкретные задачи чаще всего решаются в различных системах компьютерного зрения?

- Обнаружение лиц - найти на изображении все человеческие лица.
- Распознавание лиц - определить, какому человеку соответствует данное лицо.
- Идентификация лиц - определить, соответствует ли данное лицо какому-то конкретному человеку.
- Распознавание позы - определить какой позе из набора заранее заданных поз соответствует данная поза человека.
- Выделение областей, однородных по цвету или текстуре.
- Распознавание объектов - определить какому из заранее определённых классов относится данный объект на изображении.
- Семантическая сегментация - выделение областей изображения, относящихся к одному объекту.
- Классификация изображений - определение к какой категории относится изображение или изображённые на нем объекты, без локализации этих объектов на изображении. При помощи классификации изображений можно ответить на такие вопросы об изображении как изображён ли на картинке слон, или самолёт, или закат. Без указания где именно на картинке расположен слон или самолёт.
- Обнаружение и распознавание текста - определить, где на картинке изображён текст и распознать что там написано.

И даже такая общая задача, как семантическое описание изображения может быть решена как совокупность перечисленных ранее более узких задач. Например, давайте посмотрим на эту фотографию. Если я вас попрошу описать эту фотографию словами, скорее всего я услышу, что на фотографии изображена маленькая девочка, которая ест мороженное. Вы сможете описать во что она одета, где она находится.

можно составить похожее описание изображения автоматически при помощи компьютера, решив такие задачи, как распознавание объектов и лиц, определение пола и возраста человека, выделение однородных по цвету областей, распознавание действия, выделение текстуры.

## 3.2 Наборы данных

Очень важную роль в развитии области компьютерного зрения (как и во многих других областях анализа данных) играют общедоступные наборы данных, собранные для тестирования и сравнения различных алгоритмов, решающих какую-то конкретную задачу. Стандартные наборы данных необходимы для объективного сравнения различных подходов к решению одной и той же задачи. Данные необходимы для разработки алгоритмов, основанных на машинном обучении.

Давайте познакомимся с наиболее известными наборами данных, используемыми научным сообществом для разработки и тестирования все более и более совершенных алгоритмов анализа изображений.

**Lena** Наверное, самым известным тестовым изображением, использовавшимся для проверки и демонстрации различных алгоритмов обработки изображений, является «Лена». Тестовое изображение представляет собой оцифрованный портрет шведской модели, который в свою очередь является фрагментом разворота из журнала Playboy. Согласно Википедии, в 1973 году Александру Савчуку из Университета Южной Калифорнии для иллюстрации статьи на тему обработки изображений понадобился фотопортрет с хорошим динамическим диапазоном. Савчук отсканировал фрагмент постера из Playboy. Он использовал сканер с разрешением 100 линий на дюйм, и результирующее изображение получилось 512 на 512 точек. Вскоре эта картинка превратилась для индустрии в стандарт де-факто: на нём проверялись и отрабатывались всевозможные приёмы коррекции изображений, оттачивались новые алгоритмы обработки.

**MNIST** Одним из первых наборов изображений для тестирования алгоритмов машинного обучения является коллекция MNIST (сокращение от «Modified National Institute of Standards and Technology»). Она состоит из содержит образцы рукописного написания цифр. С использованием этого набора данных разрабатывались первые алгоритмы анализа изображений на основе нейронных сетей. Коллекция MNIST содержит 60000 изображений для обучения и 10000 изображений для тестирования размером 28x28 пикселей. Все изображения поделены на 10 классов, по числу цифр. Задача распознавания рукописных цифр на этом наборе данных - это классическая задача классификации. Необходимо по изображению определить к какому классу



оно принадлежит (т.е. какая из 10 цифр изображена). По сегодняшним меркам, и размер коллекции, и разрешение изображений очень небольшой. Но этот набор данных по-прежнему активно используется, в основном при обучении нейронным сетям. Запрограммировать нейронную сеть для обучения и тестирования на этом наборе данных, это как написать "Hello World" на новом языке программирования.

**CIFAR** CIFAR-10 и CIFAR-100 - тоже популярные наборы данных для тестирования нейронных сетей. По количеству и разрешению изображений они похожи на MNIST, но изображения более разнообразны. CIFAR-10 содержит картинки из 10 классов, таких как самолёт, машина, птица, и тому подобные. CIFAR-100 содержит картинки из 100 классов. Эти наборы данных также широко используются для тестирования алгоритмов классификации изображений.

**ImageNet** ImageNet - размеченная коллекция изображений, с созданием которой многие связывают начало эпохи глубокого обучения и успех глубоких нейронных сетей. Для обучения глубоких сетей нужно много данных. ImageNet - первый действительно большой набор аннотированных изображений. ImageNet содержит более 14 миллионов изображений (сравните с 60000 в MNIST и CIFAR!), разбитых на более чем 20000 классов.

С использованием этой коллекции данных с 2010 года проводится соревнование ILSVRC (англ. ImageNet Large Scale Visual Recognition Challenge — Кампания по широкомасштабному распознаванию образов в ImageNet), в рамках которого различные программные продукты ежегодно соревнуются в классификации и распознавании объектов и сцен в базе данных ImageNet. Для ILSVRC используется подмножество коллекции ImageNet, состоящее из около 1.2 миллиона изображений и 1000 классов.

**PASCAL VOC** До появления ImageNet и ILSVRC наиболее известным соревнованием по классификации изображений и обнаружению объектов было PASCAL VOC challenge (Pattern Analysis, Statistical Modelling and Computational Learning Visual Object Classes), проводившееся с 2005 по 2012 год. Коллекций изображений, собранные для этого соревнования, долгое время являлись де-факто стандартом для тестирования различных алгоритмов, особенно для алгоритмов обнаружения объектов и сегментации. PASCAL VOC в отличие от предыдущих обсуждавшихся коллекций содержит разметку на уровне объектов, когда на изображении размечены регионы с объектами.

<http://host.robots.ox.ac.uk/pascal/VOC/databases.html>

<http://host.robots.ox.ac.uk/pascal/VOC/images/tud3c.html>

**Caltech** Caltech101 и Caltech256 ещё две коллекции изображений для тестирования алгоритмов классификации изображений и распознавания объектов. Коллекции содержат изображения 101 и 256 категорий соответственно. Размер картинок примерно 200x300 пикселей - как в ImageNet и PASCAL VOC. Большинство категорий Caltech101 содержит около 50 изображений, но есть категории с намного большим числом картинок. Например в категориях "самолет" и "лицо" около 800 примеров.

К каждому изображению прилагается аннотация, содержащая координаты рамки в которой находится объект и класс объекта.

[http://www.vision.caltech.edu/Image\\_Datasets/Caltech101/](http://www.vision.caltech.edu/Image_Datasets/Caltech101/)

[http://www.vision.caltech.edu/Image\\_Datasets/Caltech256/](http://www.vision.caltech.edu/Image_Datasets/Caltech256/)

**MS COCO** Все рассмотренные нами ранее коллекции содержат изображения относительно небольшого размера. В то же время, большинство камер сегодня снимает с более высоким разрешением. И для ряда задач (например, распознавание объектов для автономных транспортных средств) важно уметь работать с изображениями с более высоким разрешением. MS COCO (Microsoft COCO: Common Objects in Context) содержит изображения с разрешением до 1000 пикселей вдоль большей стороны. Также большинство коллекций, о которых мы говорили до этого, не имеет разметки для семантической сегментации, когда для каждого пикселя изображения есть метка о его принадлежности к определённому объекту.

Коллекция MS COCO сегодня является стандартом для сравнения алгоритмов распознавания объектов и семантической сегментации. Она содержит 328000 изображений с объектами 91 категории. Всего в коллекции около 2500000 размеченных объектов.

<http://cocodataset.org/#home>

<https://arxiv.org/pdf/1405.0312.pdf>

**Cityscapes** Ещё один широко используемый набор изображений высокого разрешения - Cityscapes. Коллекция содержит изображения городских пейзажей, собранных в 50 городах Европы. Размер изображений 1024 x 2048 пикселей. В наборе данных 5000 изображений с детальной аннотацией объектов - на уровне отдельных пикселей, и около 20000 изображений с менее детальной аннотацией.

<https://www.cityscapes-dataset.com/>

<https://arxiv.org/pdf/1604.01685.pdf>

**Labeled Faces in the Wild** Помимо коллекций, содержащих картинки всевозможных категорий объектов, существуют также наборы данных для распознавания определённых категорий, имеющих особую значимость.

Так, существуют размеченные наборы для обнаружения пешеходов (<http://coding-guru.com/popular-pedestrian-detection-datasets/>), распознавания жестов, распознавания дорожных знаков.

Отдельного внимания всегда заслуживала задача обнаружения и распознавания лиц. Было создано большое количество размеченных коллекций для этих задач. Одной из наиболее широко используемых сегодня коллекций является Labeled Faces in the Wild:

<http://vis-www.cs.umass.edu/lfw/>.

Как следует из названия, коллекция содержит изображения лиц в "естественных условиях без контроля над освещением, позой человека и его головы. Это не портретная съёмка на монотонном фоне, а репортажные кадры. Коллекция содержит более 13000 изображений известных людей, собранных в интернете. Около 1680 личностей из коллекции имеют более одной фотографии.

**Медицинские изображения** Несмотря на то, что анализ медицинских изображений является одной из основных прикладных задач компьютерного зрения, общедоступных коллекций медицинских изображений не так много. В первую очередь это связано с персональным аспектом этих данных. Только в последние годы, когда стало ясно что доступность больших коллекций данных способна значительно ускорить развитие алгоритмов, начали появляться объёмные анонимизированные коллекции медицинских снимков. Первой такой коллекцией, получившей распространение среди групп машинного обучения, была ChestXray14, содержащая около 112,000 снимков грудной клетки с признаками 14 видов болезней. Недавно группа машинного обучения Стэнфорда выпустила ещё одну объёмную коллекцию CheXpert. Эта коллекция содержит 224316 снимков грудной клетки.

<http://academictorrents.com/details/557481faacd824c83fbf57dcf7b6da9383>

<https://stanfordmlgroup.github.io/competitions/chexpert/>

<https://arxiv.org/abs/1901.07031>