ЛЕКЦИЯ 10
МЕТОДЫ И ТЕХНОЛОГИИ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Цель лекции

Изучить основные понятия компьютерного зрения, способы представления изображения в цифровой форме, этапы анализа изображений, методы выделения границ и областей, задачи классификации, детекции, сегментации и распознавания объектов; рассмотреть роль нейронных сетей, особенности 2D- и 3D-зрения и практические примеры применения технологий компьютерного зрения.

Понятие компьютерного зрения

Компьютерное зрение (Computer Vision) — направление искусственного интеллекта, связанное с получением, обработкой и интерпретацией визуальной информации. В учебнике В.Н. Бондарева и Ф.Г. Аде компьютерное зрение рассматривается как область, направленная на методы понимания изображений и формирование на их основе символического описания реальности.

Изображения, поступающие в систему компьютерного зрения, могут быть различными: двумерными чёрно-белыми или цветными изображениями, стереоизображениями, диапазонными изображениями от датчиков глубины, последовательностями кадров и трёхмерными данными. Поэтому компьютерное зрение охватывает не одну отдельную задачу, а широкий класс методов анализа визуальной информации.

Общая схема системы компьютерного зрения

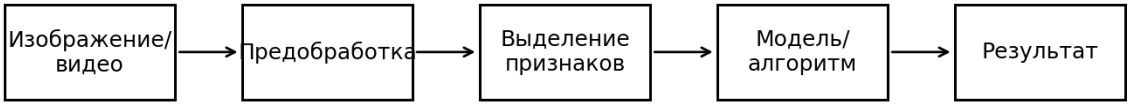


Рисунок 1 – Общая схема обработки визуальной информации

Изображение как цифровые данные

Для компьютера изображение первоначально представляет собой массив числовых значений. В чёрно-белом изображении каждому пикселю соответствует значение интенсивности. В цветном изображении пиксель обычно представлен несколькими компонентами, например каналами RGB.

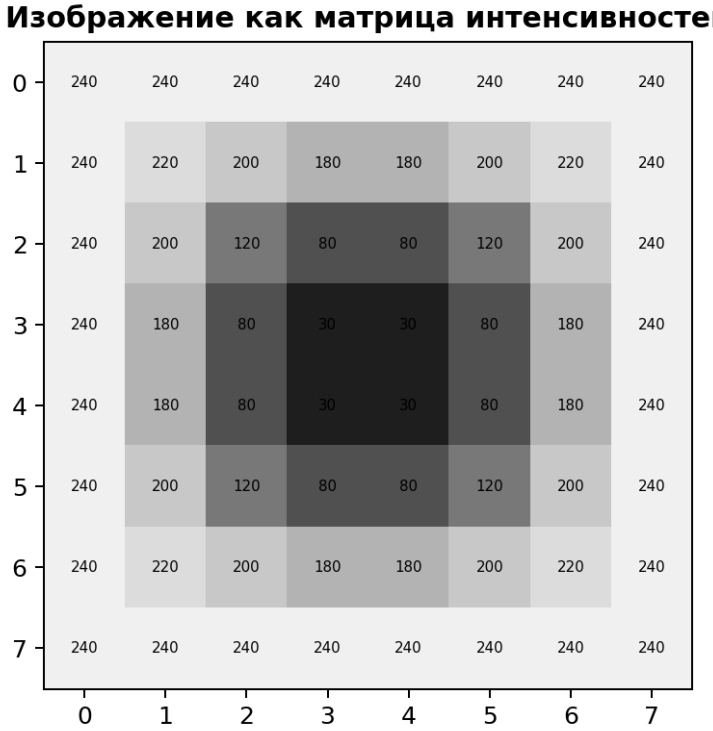


Рисунок 2 – Упрощённое представление изображения как матрицы значений

Если использовать 8 бит для одного канала интенсивности, значение пикселя может находиться в диапазоне от 0 до 255. При этом 0 соответствует минимальной интенсивности, а 255 — максимальной. Цветное изображение можно рассматривать как несколько матриц одинакового размера, каждая из которых описывает отдельный цветовой канал.

Тип данных	**Представление**	**Пример**
Чёрно-белое изображение	Матрица H×W	Рентгеновский снимок или техническое изображение
Цветное изображение	Тензор H×W×3	Фотография
Видео	Последовательность кадров	Видеонаблюдение
Изображение глубины	Значение расстояния для каждой точки	3D-датчик или depth-камера
Облако точек	Набор 3D-координат	Лидар, датчик глубины

Почему понимание изображения является сложной задачей

Учебник подчёркивает, что понимание изображения для человека кажется естественным, но для компьютера является сложной задачей. Одна из причин состоит в том, что двумерное изображение является проекцией трёхмерного мира. При проецировании часть информации, прежде всего глубина и пространственные отношения, теряется.

Другие сложности связаны с шумом датчиков, изменением освещения, тенями, отражающими свойствами поверхности, ракурсом, масштабом, частичным перекрытием объектов и необходимостью обрабатывать большие объёмы данных в реальном времени.

Фактор	**Как влияет на анализ**
Освещение	Один объект может иметь разные яркость и цвет
Ракурс	Форма объекта на изображении изменяется

Масштаб	Объект занимает разное число пикселей
Шум	Искажает значения пикселей и границы
Перекрытие	Часть объекта скрыта другим объектом
Движение	Необходимо учитывать последовательность кадров
Проекция 3D→2D	Теряется часть пространственной информации

Предобработка изображений

Перед распознаванием изображение часто подвергается предварительной обработке. Цель предобработки — уменьшить влияние шума, привести данные к ожидаемому формату и подготовить их к дальнейшему анализу.

В классических системах могут применяться сглаживание, фильтрация, изменение контраста, нормализация яркости и геометрические преобразования. В учебнике в качестве примера уменьшения шума рассматривается сглаживание изображения свёрткой с двумерным гауссовским окном.

Операция	**Назначение**
Изменение размера	Приведение всех изображений к единому размеру
Нормализация	Приведение значений к удобному диапазону
Сглаживание	Снижение влияния высокочастотного шума
Повышение контраста	Улучшение различимости деталей
Геометрические преобразования	Поворот, масштабирование, выравнивание

Выделение границ изображения

Граница — область резкого изменения яркости или другого визуального признака. Выделение границ является одной из базовых операций классического компьютерного зрения. Она позволяет перейти от массива пикселей к более структурированному описанию формы объектов.

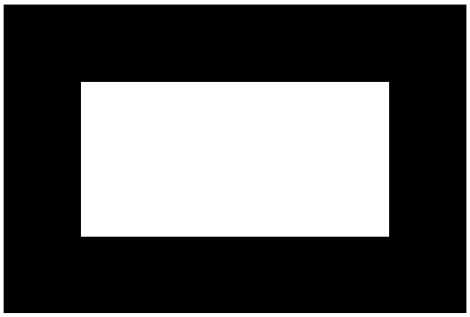


Рисунок 3 – Упрощённый пример выделенной границы объекта

Классический подход заключается в анализе изменений интенсивности по соседним пикселям. После обнаружения отдельных граничных элементов возникает задача объединения их в связанные линии и контуры.

От граничных элементов к контурам

Если обнаружены отдельные точки границы, необходимо установить, какие из них принадлежат одной линии или одному объекту. Для этого применяются методы группировки

и поиска. Один из классических методов — преобразование Хафа, позволяющее искать геометрические структуры, например прямые линии, через переход в пространство параметров.

Преобразование Хафа

Преобразование Хафа используется для обнаружения объектов или контуров, которые можно описать параметрической моделью. Наиболее известный пример — поиск прямых линий на изображении.

Каждая граничная точка голосует за множество возможных параметров линии. В параметрическом пространстве области с большим числом голосов соответствуют вероятным линиям исходного изображения.

Шаг	Содержание
1	Получить карту границ
2	Для каждой граничной точки сформировать допустимые параметры модели
3	Накопить голоса в пространстве параметров
4	Найти максимумы голосования
5	Восстановить найденные линии или другие геометрические объекты

Выделение областей изображения

Альтернативой анализу контуров является выделение областей. В этом случае алгоритм объединяет соседние пиксели, обладающие сходными характеристиками, например яркостью, цветом или текстурой.

Один из классических принципов — наращивание областей. Начальная точка или небольшой набор точек постепенно расширяется за счёт соседних пикселей, удовлетворяющих заданному критерию сходства. Такие методы применяются и к трёхмерным данным, например при сегментации облаков точек.

Основные задачи современного компьютерного зрения

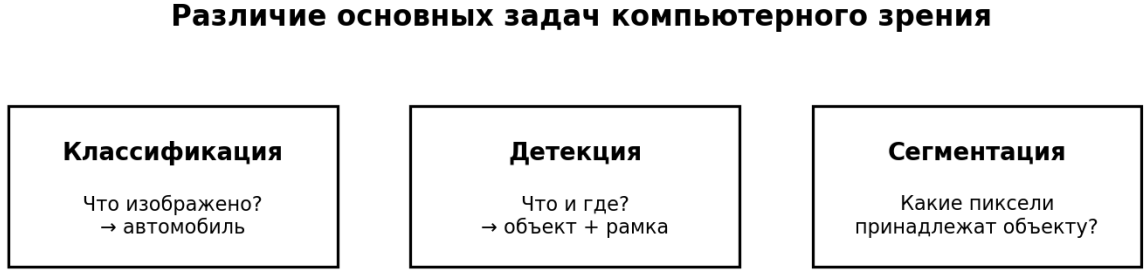


Рисунок 4 – Классификация, детекция и сегментация

Задача	Что требуется определить	Пример
Классификация изображения	Какой объект или класс представлен в кадре	malware screenshot / normal interface
Детекция объектов	Какие объекты присутствуют и где они расположены	человек, автомобиль, устройство

Сегментация	Какие пиксели принадлежат определённому объекту или области	выделение дефекта
Распознавание лица	Соответствует ли лицо известному человеку или шаблону	контроль доступа
Оценка позы	Положение ключевых точек объекта	положение человека
Отслеживание	Положение объекта во времени	движение объекта в видеопотоке

Классификация изображений

Классификация изображения назначает всему изображению одну или несколько категорий. ***Например***, система может определить, изображён ли на фотографии автомобиль, человек или другое состояние объекта.

В классических подходах сначала выделялись признаки изображения, после чего применялся классификатор. Современные глубокие нейронные сети способны обучать визуальные признаки непосредственно из данных.

Типовой процесс включает формирование размеченного набора изображений, разделение на train/validation/test, нормализацию, обучение модели и оценку качества на независимых данных.

Детекция объектов

Детекция объектов является более сложной задачей, чем классификация. Необходимо определить не только класс объекта, но и его положение на изображении. Обычно результат представляется ограничивающей рамкой (bounding box) и меткой класса.

Например, система видеонаблюдения может обнаружить на одном кадре несколько автомобилей и людей. Для каждого объекта модель должна определить координаты рамки и категорию.

Сегментация

Сегментация выполняет классификацию на уровне пикселей. При семантической сегментации каждому пикселю присваивается класс. При instance segmentation дополнительно различаются отдельные экземпляры объектов одного класса.

Тип	**Результат**	**Пример**
Semantic segmentation	Класс каждого пикселя	дорога / здание / человек
Instance segmentation	Отдельная маска каждого объекта	каждый человек выделен отдельно
Binary segmentation	Объект / фон	дефект / нормальная поверхность

Распознавание объектов

В классической компьютерной системе распознавание объектов может основываться на глобальных или локальных признаках. Учебник Бондарева и Аде рассматривает 2D-методы с глобальными признаками, локальными признаками и графами отношений, а также пример 3D-распознавания.

Глобальные признаки описывают объект целиком. Локальные признаки описывают отдельные характерные части — углы, отверстия, участки контура. Локальный подход полезен, когда объект частично перекрыт: для распознавания может быть достаточно видимой части, содержащей характерные признаки.

2D- и 3D-компьютерное зрение

Характеристика	2D-зрение	3D-зрение
Основные данные	Пиксели изображения	Глубина, стереоданные, облака точек
Координаты	x, y	x, y, z
Типичные задачи	Классификация, детекция, контуры	Оценка формы, положения и глубины
Источники	Камера	Стереокамера, depth-камера, лидар
Применение	Фото, видео, видеонаблюдение	Робототехника, навигация, манипуляция

Стереозрение и глубина

Для восстановления глубины можно использовать изображения, полученные двумя камерами с известным взаимным расположением. Одна и та же точка трёхмерной сцены отображается в двух изображениях с некоторым смещением — диспаритетом. На основе этого смещения и параметров камер можно оценить глубину.

Для удалённых точек диспаритет становится малым. Чтобы получить трёхмерную структуру сцены, требуется найти соответствующие точки на изображениях обеих камер.

Облака точек

Датчики глубины могут формировать набор трёхмерных точек. Такой набор называют облаком точек. Каждая точка имеет координаты x, y, z, но первоначально связи между точками могут отсутствовать.

Далее система должна объединить точки в поверхности или области. В приведённом в учебнике примере облака точек сегментируются на плоские участки, после чего строятся гипотезы объектов и оценивается возможность их захвата манипулятором робота.

Свёрточные нейронные сети CNN

Свёрточные нейронные сети являются основным классом глубоких моделей для анализа изображений. Свёрточный слой применяет небольшие обучаемые фильтры к локальным участкам изображения и формирует карты признаков.

Идея локального анализа хорошо соответствует структуре изображений: важны не только значения отдельных пикселей, но и локальные комбинации — границы, текстуры, формы. В глубоких CNN ранние слои формируют простые признаки, а последующие — более сложные визуальные представления.

Компонент CNN	Назначение
Convolution	Выделение локальных признаков
Activation	Добавление нелинейности
Pooling / downsampling	Снижение пространственного размера
Dense / classifier head	Формирование итогового класса или прогноза

Пример классификации изображений в TensorFlow/Keras

```
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
```

```
model = keras.Sequential([
    layers.Input(shape=(128, 128, 3)),
    layers.Conv2D(32, 3, activation='relu'),
    layers.MaxPooling2D(),
    layers.Conv2D(64, 3, activation='relu'),
    layers.MaxPooling2D(),
    layers.Flatten(),
    layers.Dense(64, activation='relu'),
    layers.Dense(3, activation='softmax')
])

model.compile(
    optimizer='adam',
    loss='sparse_categorical_crossentropy',
    metrics=['accuracy']
)

model.fit(train_ds, validation_data=val_ds, epochs=15)
```

В примере сеть получает цветное изображение 128×128 пикселей и формирует вероятности трёх классов. Это учебная схема; реальная архитектура и параметры должны определяться экспериментально.

Подготовка изображений для обучения

Операция	**Зачем нужна**
Resize	Единый размер входа модели
Normalization	Стабильный диапазон значений
Augmentation	Искусственное разнообразие обучающих примеров
Train/validation/test split	Корректная оценка обобщения
Проверка классов	Контроль дисбаланса и ошибок разметки

Аугментация может включать допустимые повороты, отражения, сдвиги и изменения масштаба. Она должна сохранять смысл класса. Например, не каждое геометрическое преобразование допустимо в техническом или медицинском изображении.

Transfer Learning

Для многих задач не требуется обучать глубокую сеть с нуля. Можно использовать модель, предварительно обученную на большом наборе изображений, и адаптировать её к новой задаче. Этот подход называется transfer learning.

Преимущества — уменьшение требований к объёму размеченных данных и сокращение вычислительных затрат. Однако необходимо проверить, насколько исходная область данных близка к новой задаче.

Оценка классификации изображений

Для классификации изображений применяются уже знакомые метрики accuracy, precision, recall, F1-score и confusion matrix. При дисбалансе классов общая accuracy может быть недостаточной.

В многоклассовой задаче важно анализировать ошибки отдельно по каждому классу. Например, если система хорошо распознаёт нормальные изображения, но пропускает редкий опасный дефект, высокая общая точность не означает пригодность модели.

Оценка детекции объектов

Для детекции необходимо оценивать не только правильность класса, но и точность локализации объекта. Часто используется мера IoU — Intersection over Union, то есть отношение площади пересечения, предсказанной и эталонной рамок к площади их объединения.

На основе порога IoU определяют корректность обнаружения, после чего рассчитывают precision и recall детектора. Для сравнения детекторов часто используется средняя точность по классам, однако конкретная схема оценки зависит от принятого протокола.

Распознавание лиц

Распознавание лиц включает несколько различных подзадач: обнаружение лица на изображении, выравнивание, формирование числового представления лица и сравнение с известными образцами.

Подзадача	**Результат**
Face detection	Координаты лица
Face alignment	Нормализованное положение лица
Feature extraction	Вектор признаков лица
Verification	Ответ: один и тот же человек или нет
Identification	Выбор личности из набора известных

Распознавание лиц связано с биометрическими и персональными данными. Поэтому при разработке таких систем необходимо учитывать требования конфиденциальности, права доступа, защиту шаблонов и риск ошибочной идентификации.

Видеопоток и отслеживание объектов

Видео добавляет временное измерение. Система должна не только обнаруживать объект в отдельном кадре, но и связывать его положения между кадрами. Эта задача называется tracking.

Пример: в системе видеонаблюдения один и тот же автомобиль последовательно появляется на множестве кадров. Детектор находит объект, а алгоритм отслеживания поддерживает его идентификатор во времени.

Компьютерное зрение в информационной безопасности

Применение	**Визуальные данные**	**Задача**
Контроль физического доступа	Камеры	Обнаружение и распознавание лиц при законном и разрешённом использовании
Мониторинг защищаемой зоны	Видео	Детекция объектов и событий
Анализ экранов и интерфейсов	Скриншоты	Классификация состояния или визуальных признаков
Защита документов	Изображения документов	Распознавание структуры, признаков подделки или повреждения

Анализ QR/штрих-кодов	Изображения	Обнаружение и декодирование визуального маркера

При использовании видеонаблюдения и биометрии необходимо учитывать требования законодательства и политики организации. Техническая возможность распознавания объекта не означает автоматической допустимости его применения.

Компьютерное зрение в промышленности и робототехнике

В основном компьютерное зрение связывает с системами технического зрения роботов. В робототехнике визуальная система не только распознаёт объект, но и оценивает его положение, ориентацию и возможность взаимодействия.

Например, робот-манипулятор получает данные от нескольких датчиков глубины, объединяет трёхмерные точки, сегментирует поверхности, формирует гипотезы объектов и выбирает подходящую область для захвата.

Классические методы и глубокое обучение

Критерий	**Классические методы**	**Глубокое обучение**
Признаки	Задаются алгоритмом или проектируются вручную	Обучаются из данных
Примеры	Границы, Хаф, контуры, локальные признаки	CNN и современные нейросетевые модели
Данные	Могут работать при меньших наборах	Часто выигрывают при большом объёме данных
Интерпретация	Легче связать результат с явным признаком	Представления сложнее интерпретировать
Гибридный подход	Классические этапы могут сочетаться с ML	Нейросеть может быть частью более широкой CV-системы

Типичные ошибки при разработке системы компьютерного зрения

– Оценивать модель только на обучающей выборке.
– Смешивать почти одинаковые изображения одного объекта между train и test.
– Игнорировать изменение освещения, ракурса и качества камеры.
– Применять аугментацию, которая меняет смысл класса.
– Оценивать многоклассовую модель только общей accuracy.
– Использовать изображения с персональными или биометрическими данными без необходимых прав и мер защиты.
– Считать детекцию и классификацию одной и той же задачей.
– Предполагать, что модель, обученная на одном типе камер или объектов, автоматически сохранит качество в другой среде.

Пример полного конвейера: классификация изображений

Пусть требуется классифицировать изображения на три категории. Сначала формируется размеченный набор данных и проверяется корректность меток. Затем изображения приводятся к одинаковому размеру, значения пикселей нормализуются, а данные разделяются на train, validation и test.

В качестве базовой модели можно использовать небольшую CNN. После обучения анализируются loss и accuracy на train и validation, затем выполняется независимая проверка на test. Если модель переобучается, применяются аугментация, регуляризация, Dropout, ранняя

остановка или transfer learning. В заключение строится confusion matrix и анализируются типичные ошибочные изображения.

Пример полного конвейера: детекция объектов

В задаче детекции кроме класса требуется разметка местоположения объектов. Для каждого обучающего изображения задаются рамки и категории. После обучения детектор выдаёт набор рамок, вероятностей и классов.

При проверке необходимо анализировать ложные обнаружения, пропущенные объекты и точность рамок. В практической системе также оцениваются скорость обработки кадров и устойчивость к изменению условий съёмки.

Основные выводы

– Компьютерное зрение преобразует цифровые изображения и видео в содержательную информацию о сцене.

– Изображение является числовым массивом, а цветное изображение — многоканальным тензором.

– Классические этапы компьютерного зрения включают предобработку, выделение границ, контуров и областей.

– Преобразование Хафа относится к классическим методам поиска параметрических геометрических структур.

– Основные современные задачи — классификация, детекция, сегментация, распознавание и отслеживание объектов.

– 3D-зрение использует стереоданные, изображения глубины и облака точек.

– CNN позволяют автоматически обучать иерархические визуальные признаки.

– При оценке CV-моделей необходимо учитывать качество разметки, независимость test и особенности конкретной задачи.

– Применение биометрических и визуальных данных требует мер безопасности, конфиденциальности и правовой обоснованности.

Контрольные вопросы

1. Что понимается под компьютерным зрением?
2. Как изображение представляется в памяти компьютера?
3. Чем чёрно-белое изображение отличается от цветного с точки зрения структуры данных?
4. Почему восстановление трёхмерной сцены по одному изображению является сложной задачей?
5. Какие факторы затрудняют анализ изображений? Для чего выполняется предобработка изображения?
6. Что такое граница изображения и какова цель выделения контуров?
7. В чём основная идея преобразования Хафа?
8. Чем выделение областей отличается от анализа границ?
9. Чем классификация изображения отличается от детекции объектов?
10. Что такое семантическая сегментация?
11. Что такое instance segmentation?
12. Что понимается под распознаванием объекта?
13. Какие применения компьютерного зрения возможны в информационной безопасности?
14. Чем классические методы CV отличаются от глубокого обучения?

15. Какие ошибки при разделении train/test могут привести к завышенной оценке качества? Почему условия съёмки важны для обобщения модели?

16. Какие риски необходимо учитывать при работе с биометрическими изображениями?

17. Как выглядит типовой конвейер классификации изображений?

Список источников

1. Бондарев В.Н., Аде Ф.Г. Искусственный интеллект: учебное пособие. Севастополь, 2002. Глава 10 «Компьютерное зрение»: основные понятия, системы компьютерного зрения, выделение границ и областей, преобразование Хафа, 2D- и 3D-методы распознавания.

2. Беликова Е.К., Гудилина И.С., Круглова М.А., Саратовская Л.Б., Ямпольская А.В. English Reader in Artificial Intelligence = Хрестоматия по искусственному интеллекту на английском языке. Учебное пособие. Москва, 2024. – 171 с.

3. Бессмертный И.А., Авдюшина А.Е., Кугаевских А.В., Королева Ю.А., Рущенко Н.Г. Системы искусственного интеллекта. Методическое пособие. Санкт-Петербург: Университет ИТМО, 2024. – 281 с.

4. Андреянов Н.В., Евдокимова Т.С., Павлов А.Д., и др. Методы искусственного интеллекта. Учебное пособие. Казань: КНИТУ-КАИ, 2024. – 392 с.