

УДК 004.8
ББК 32.81
К63

К63 Компьютерное зрение. Современные методы и перспективы развития / ред. Р. Дэвис, М. Терк; пер. с англ. В. С. Яценкова. – М.: ДМК Пресс, 2022. – 690 с.: ил.

ISBN 978-5-93700-148-1

Эта книга рассказывает о передовых методах компьютерного зрения. Показано, как искусственный интеллект обнаруживает признаки и объекты, на каких данных он обучается, на чем основано распознавание лиц и действий, отслеживание аномалий. Особое внимание уделяется методам глубокого обучения. Все ключевые принципы проиллюстрированы примерами из реальной практики.

Книга адресована исследователям и практикам в области передовых методов компьютерного зрения, а также тем, кто изучает эту технологию самостоятельно или в рамках вузовского курса.

УДК 004.8
ББК 32.81

This Russian edition of *Advanced Methods and Deep Learning in Computer Vision* (9780128221099) by E.R. Davies and Matthew Turk is published by arrangement with Elsevier Inc.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

ISBN 978-0-12-822109-9 (англ.)
ISBN 978-5-93700-148-1 (рус.)

© Elsevier Inc., 2022
© Перевод, оформление, издание,
ДМК Пресс, 2022

Содержание

От издательства	17
Список соавторов	18
О редакторах	20
Предисловие	21
 Глава 1. Кардинальные перемены в области компьютерного зрения	27
1.1. Введение. Компьютерное зрение и его история	27
1.2. Часть А. Обзор операторов низкоуровневой обработки изображений	31
1.2.1. Основы обнаружения краев	31
1.2.2. Оператор Кэнни	33
1.2.3. Обнаружение сегмента линии	34
1.2.4. Оптимизация чувствительности обнаружения	35
1.2.5. Работа с изменениями интенсивности фона	37
1.2.6. Теория, сочетающая согласованный фильтр и конструкции с нулевым средним	37
1.2.7. Структура маски (дополнительные соображения).....	38
1.2.8. Обнаружение угла	40
1.2.9. Оператор «особой точки» Харриса	41
1.3. Часть В. Локализация и распознавание двухмерных объектов	43
1.3.1. Подход к анализу формы на основе центроидного профиля	43
1.3.2. Схемы обнаружения объектов на основе преобразования Хафа	46
1.3.3. Применение преобразования Хафа для обнаружения линий	50
1.3.4. Использование RANSAC для обнаружения линий	51
1.3.5. Теоретико-графовый подход к определению положения объекта	54
1.3.6. Использование обобщенного преобразования Хафа для экономии вычислений	57
1.3.7. Подходы на основе частей	59
1.4. Часть С. Расположение трехмерных объектов и важность неизменности	60
1.4.1. Введение в трехмерное зрение.....	60
1.4.2. Неоднозначность положения при перспективной проекции.....	64
1.4.3. Инварианты как помощь в трехмерном распознавании	68
1.4.4. Кросс-коэффициенты: концепция «отношения коэффициентов»	69
1.4.5. Инварианты для неколлинеарных точек	71
1.4.6. Обнаружение точки схода	73
1.4.7. Подробнее о точках схода	75
1.4.8. Промежуточный итог: значение инвариантов	76
1.4.9. Преобразование изображения для калибровки камеры	77
1.4.10. Калибровка камеры.....	80

1.4.11. Внутренние и внешние параметры	82
1.4.12. Многоракурсное зрение	83
1.4.13. Обобщенная геометрия стереозрения	84
1.4.14. Существенная матрица	85
1.4.15. Фундаментальная матрица	87
1.4.16. Свойства существенной и фундаментальной матриц	88
1.4.17. Расчет фундаментальной матрицы	88
1.4.18. Усовершенствованные методы триангуляции	89
1.4.19. Достижения и ограничения многоракурсного зрения	90
1.5. Часть D. Отслеживание движущихся объектов	90
1.5.1. Основные принципы отслеживания	90
1.5.2. Альтернативы вычитанию фона	94
1.6. Часть E. Анализ текстур	98
1.6.1. Введение	98
1.6.2. Основные подходы к анализу текстур	99
1.6.3. Метод Лоуза на основе энергии текстуры	101
1.6.4. Метод собственного фильтра Аде	103
1.6.5. Сравнение методов Лоуза и Аде	105
1.6.6. Последние разработки	106
1.7. Часть F. От искусственных нейронных сетей к методам глубокого обучения	106
1.7.1. Введение: как ИНС превратились в СНС	106
1.7.2. Параметры, определяющие архитектуру CNN	109
1.7.3. Архитектура сети AlexNet	110
1.7.4. Архитектура сети VGGNet Симоняна и Зиссермана	113
1.7.5. Архитектура DeconvNet	116
1.7.6. Архитектура SegNet	118
1.7.7. Применение глубокого обучения для отслеживания объектов	120
1.7.8. Применение глубокого обучения в классификации текстур	124
1.7.9. Анализ текстур в мире глубокого обучения	128
1.8. Часть G. Заключение	129
Благодарности	130
Литературные источники	130
Об авторе главы	135

Глава 2. Современные методы робастного обнаружения объектов

2.1. Введение	137
2.2. Предварительные положения	139
2.3. R-CNN	141
2.3.1. Внутреннее устройство	141
2.3.2. Обучение	142
2.4. Сеть SPP-Net	142
2.5. Сеть Fast R-CNN	143
2.5.1. Архитектура	144
2.5.2. Пулинг ROI	144

2.5.3. Многозадачная функция потери	145
Классификация	145
Регрессия ограничивающей рамки	145
2.5.4. Стратегия тонкой настройки	146
2.6. Faster R-CNN.....	146
2.6.1. Архитектура.....	147
2.6.2. Сети прогнозирования регионов	147
2.7. Каскадная R-CNN.....	149
2.7.1. Каскадная архитектура R-CNN.....	150
2.7.2. Каскадная регрессия ограничивающей рамки.....	151
2.7.3. Каскадное обнаружение.....	152
2.8. Представление разномасштабных признаков.....	152
2.8.1. Архитектура MC-CNN.....	154
2.8.1.1. Архитектура.....	154
2.8.2. Сеть FPN	155
2.8.2.1. Архитектура.....	156
2.9. Архитектура YOLO	158
2.10. Сеть SSD	159
2.10.1. Архитектура.....	159
2.10.2. Обучение	160
2.11. RetinaNet.....	161
2.11.1. Фокальная потеря.....	161
2.12. Производительность детекторов объектов	162
2.13. Заключение	163
Литературные источники	164
Об авторах главы.....	165

Глава 3. Обучение с ограниченным подкреплением – статические и динамические задачи

3.1. Введение.....	168
3.2. Контекстно-зависимое активное обучение	168
3.2.1. Активное обучение.....	169
3.2.2. Важность контекста активного обучения	172
3.2.3. Фреймворк контекстно-зависимого активного обучения	174
3.2.4. Практическое применение.....	177
3.3. Локализация событий при слабой разметке.....	180
3.3.1. Архитектура сети	183
3.3.2. k-матричное обучение.....	183
3.3.3. Сходство совместных действий.....	184
3.3.4. Практическая реализация	186
3.4. Семантическая сегментация с использованием слабой разметки	189
3.4.1. Слабые метки для классификации категорий.....	191
3.4.2. Слабые метки для выравнивания признаков.....	192
3.4.3. Оптимизация сети	194
3.4.4. Получение слабой разметки.....	195
3.4.5. Применения.....	196
3.4.6. Визуализация выходного пространства.....	198

3.5. Обучение с подкреплением со слабой разметкой для динамических задач.....	199
3.5.1. Обучение прогнозированию подцелей.....	202
3.5.2. Предварительное обучение с учителем	204
3.5.3. Практическое применение.....	204
3.6. Выводы.....	207
Благодарности	209
Литературные источники	209
Об авторах главы.....	215

Глава 4. Эффективные методы глубокого обучения..... 216

4.1. Сжатие модели.....	216
4.1.1. Прореживание параметров	217
4.1.2. Низкоранговая факторизация	220
4.1.3. Квантование	221
4.1.4. Дистилляция знаний.....	225
4.1.5. Автоматическое сжатие модели.....	226
4.2. Эффективные архитектуры нейронных сетей	230
4.2.1. Стандартный сверточный слой	231
4.2.2. Эффективные сверточные слои.....	231
4.2.3. Разработанные вручную эффективные модели CNN.....	232
4.2.4. Поиск нейронной архитектуры	236
4.2.5. Поиск нейронной архитектуры, ориентированной на оборудование	239
4.3. Заключение	246
Литературные источники	246

Глава 5. Условная генерация изображений и управляемая генерация визуальных паттернов..... 254

5.1. Введение	254
5.2. Изучение визуальных паттернов: краткий исторический обзор.....	258
5.3. Классические генеративные модели.....	260
5.4. Глубокие генеративные модели.....	261
5.5. Глубокая условная генерация изображений	266
5.6. Разделенные представления в управляемом синтезе паттернов	267
5.6.1. Разделение визуального содержания и стиля	267
5.6.2. Разделение структуры и стиля.....	274
5.6.3. Разделение личности и атрибутов	277
5.7. Заключение.....	284
Литературные источники	284

Глава 6. Глубокое распознавание лиц с использованием полных и частичных изображений..... 289

6.1. Введение	289
6.1.1. Модели глубокого обучения.....	291

6.2. Компоненты системы глубокого распознавания лиц	297
6.2.1. Пример обученной модели CNN для распознавания лиц	298
6.3. Распознавание лиц с использованием полных изображений лица	301
6.3.1. Проверка подобия с использованием модели FaceNet	303
6.4. Глубокое распознавание неполных изображений лица	304
6.5. Обучение специальной модели для полных и частичных изображений лица	307
6.5.1. Предлагаемая архитектура модели	309
6.5.2. Фаза обучения модели	309
6.6. Заключение	310
Литературные источники	312
Об авторе главы	313

Глава 7. Адаптация домена с использованием неглубоких и глубоких нейросетей, обучаемых без учителя

7.1. Введение	314
7.2. Адаптация домена с использованием многообразия	316
7.2.1. Адаптация домена без учителя с использованием произведения многообразий	317
7.3. Адаптация домена без учителя с использованием словарей	319
7.3.1. Общий словарь доменной адаптации	321
7.3.2. Совместная иерархическая адаптация домена и изучение признаков	325
7.3.3. Инкрементное изучение словаря для адаптации предметной области без учителя	330
7.4. Адаптация домена с использованием глубоких сетей, обучаемых без учителя	334
7.4.1. Дискриминационные подходы к адаптации предметной области	335
7.4.2. Генеративные подходы к адаптации домена	338
7.5. Заключение	346
Литературные источники	346
Об авторах главы	352

Глава 8. Адаптация домена и непрерывное обучение семантической сегментации

8.1. Введение	353
8.1.1. Формальная постановка задачи	355
8.2. Адаптация домена без учителя	356
8.2.1. Формулировка задачи адаптации домена	358
8.2.2. Основные подходы к адаптации	359
8.2.2.1. Адаптация на входном уровне	360
8.2.2.2. Адаптация на уровне признаков	361
8.2.2.3. Адаптация на уровне выхода	362
8.2.3. Методы адаптации домена без учителя	362
8.2.3.1. Состязательная адаптация домена	362
8.2.3.2. Генеративная адаптация	366

8.2.3.3. Несоответствие классификатора	368
8.2.3.4. Самостоятельное обучение	369
8.2.3.5. Многозадачность	372
8.3. Непрерывное обучение	373
8.3.1. Формулировка задачи непрерывного обучения	374
8.3.2. Особенности непрерывного обучения в семантической сегментации	376
8.3.3. Методы поэтапного обучения	378
8.3.3.1. Дистилляция знаний	378
8.3.3.2. Замораживание параметров	380
8.3.3.3. Геометрическая регуляризация на уровне признаков	380
8.3.3.4. Новые направления	381
8.4. Заключение	382
Благодарности	382
Литературные источники	382
Об авторах главы	389

Глава 9. Визуальное отслеживание движущихся объектов

9.1. Введение	390
9.1.1. Определение задачи отслеживания	390
9.1.2. Затруднения при отслеживании	391
9.1.3. Обоснование методики	392
9.1.4. Историческая справка	393
9.2. Методы на основе шаблонов	394
9.2.1. Основы	394
9.2.2. Показатели качества модели	396
9.2.3. Нормализованная кросс-корреляция	398
9.2.4. Чисто фазовый согласованный фильтр	399
9.3. Методы последовательного обучения	400
9.3.1. Фильтр MOSSE	401
9.3.2. Дискриминативные корреляционные фильтры	403
9.3.3. Подходящие признаки для DCF	405
9.3.4. Отслеживание в масштабном пространстве	406
9.3.5. Пространственное и временное взвешивание	408
9.4. Методы, основанные на глубоком обучении	410
9.4.1. Глубокие признаки в DCF	411
9.4.2. Адаптивные глубокие признаки	413
9.4.3. DCF сквозного обучения	414
9.5. Переход от отслеживания к сегментации	416
9.5.1. Сегментация видеообъектов	416
9.5.2. Генеративный метод VOS	417
9.5.3. Дискриминативный метод VOS	419
9.6. Выводы	420
Благодарности	421
Литературные источники	422
Об авторе главы	429

Глава 10. Длительное отслеживание объекта на основе глубокого обучения	430
10.1. Введение	431
10.1.1. Трудности отслеживания видеообъектов	432
10.1.1.1. Видовые проблемы отслеживания	432
10.1.1.2. Проблемы машинного обучения при отслеживании	433
10.1.1.3. Технические проблемы при отслеживании	435
10.2. Краткосрочное визуальное отслеживание объекта	435
10.2.1. Неглубокие трекеры	436
10.2.2. Глубокие трекеры	438
10.2.2.1. Отслеживание на основе корреляционного фильтра	438
10.2.2.2. Отслеживание на основе некорреляционных фильтров	440
10.3. Долгосрочное визуальное отслеживание объекта	441
10.3.1. Устаревание модели при длительном отслеживании	442
10.3.2. Исчезновение и повторное появление цели	446
10.3.3. Долгосрочные трекеры	446
10.3.3.1. Предварительное обучение и сиамские трекеры	446
10.3.4. Инвариантность и эквивариантность представления	452
10.3.4.1. Инвариантность при отслеживании	452
10.3.4.2. Эквивариантность при отслеживании	454
10.3.4.3. Эквивариантность переноса	456
10.3.4.4. Эквивариантность вращения	458
10.3.4.5. Эквивариантность масштаба	461
10.3.4.6. Эффективность сиамских трекеров	464
10.3.4.7. Гибридное обучение с сиамскими трекерами	464
10.3.4.8. Последовательное обучение помимо сиамских трекеров	467
10.3.5. Наборы данных и тесты	468
10.4. Заключение	468
Литературные источники	469
Об авторах главы	473

Глава 11. Обучение пониманию сцены на основании действий	474
11.1. Введение	474
11.2. Аффордансы объектов	476
11.2.1. Зачем аффордансы нужны компьютерному зрению?	477
11.2.2. Первые исследования на тему аффордансов	479
11.2.3. Обнаружение, классификация и сегментация аффордансов	480
11.2.3.1. Обнаружение аффордансов по геометрическим признакам	480
11.2.3.2. Семантическая сегментация и классификация по изображениям	482
11.2.4. Аффорданс в контексте распознавания действий и обучения роботов	484
11.2.4.1. Распознавание действий	484
11.2.4.2. Изучение аффордансов в зрении роботов	485

11.2.5. Промежуточный итог – изучение аффордансов	486
11.3. Функциональный анализ манипуляций	487
11.3.1. Активное взаимодействие между познанием и восприятием	487
11.3.2. Грамматика действий	488
11.3.2.1. Различные реализации грамматики	490
11.3.2.2. Являются ли грамматики выразительными и лаконичными описаниями?	491
11.3.3. Модули для понимания действий	491
11.3.3.1. Захватывание: важный признак для понимания действий	491
11.3.3.2. Геометрические факторы для робастизации	494
11.3.4. Проблематика понимания деятельности	495
11.4. Понимание функциональной сцены посредством глубокого обучения с помощью языка и зрения	496
11.4.1. Атрибуты в обучении без ознакомления	498
11.4.2. Общие пространства для встраивания	499
11.4.3. Построение семантических векторных пространств	502
11.4.3.1. word2vec	502
11.4.4. Общие пространства представления и графовые модели	503
11.5. Перспективные направления исследований	505
11.6. Выводы	507
Благодарности	508
Литературные источники	508
Об авторах главы	513

Глава 12. Сегментация событий во времени

с использованием когнитивного самообучения	515
12.1. Введение	516
12.2. Теория сегментации событий в когнитивной науке	518
12.3. Вариант 1: однопроходная сегментация во времени с использованием предсказания	521
12.3.1. Извлечение и кодирование признаков	523
12.3.2. Рекуррентное прогнозирование для прогнозирования признаков	524
12.3.3. Реконструкция признаков	525
12.3.4. Функция потерь при самообучении	525
12.3.5. Механизм стробирования на основе ошибок	526
12.3.6. Адаптивное обучение для повышения робастности	527
12.3.7. Промежуточный итог	529
12.3.7.1. Наборы данных	529
12.3.7.2. Метрики оценки	529
12.3.7.3. Вариативные исследования	530
12.3.7.4. Количественная оценка	531
12.3.7.5. Качественная оценка	533
12.4. Вариант 2: сегментация с использованием моделей событий на основе внимания	534
12.4.1. Извлечение признаков	536

12.4.2. Модуль внимания	537
12.4.3. Функция потерь, взвешенная по движению.....	537
12.4.4. Результаты	538
12.4.4.1. Набор данных.....	539
12.4.4.2. Критерии оценки.....	539
12.4.4.3. Вариативные исследования	540
12.4.4.4. Количественная оценка.....	542
12.4.4.5. Качественная оценка	542
12.5. Вариант 3: пространственно-временная локализация с использованием карты предсказательных потерь	544
12.5.1. Извлечение признаков.....	544
12.5.2. Иерархический стек предсказания	546
12.5.3. Потеря предсказания	547
12.5.4. Извлечение каналов действий.....	548
12.5.5. Результаты	548
12.5.5.1. Данные	548
12.5.5.2. Показатели и базовые уровни	549
12.5.5.3. Количественная оценка.....	550
12.5.5.4. Качественная оценка	554
12.6. Другие подходы к сегментации событий в компьютерном зрении.....	556
12.6.1. Методы на основе обучения с учителем	556
12.6.2. Методы на основе частичного обучения с учителем	557
12.6.3. Методы на основе обучения без учителя	557
12.6.4. Методы на основе самообучения	558
12.7. Выводы	559
Благодарности	560
Литературные источники	560
Об авторах главы.....	567

Глава 13. Вероятностные методы обнаружения аномалий в данных временных рядов с использованием обученных моделей для мультимедийных самосознательных систем 568

13.1. Введение	569
13.2. Базовые понятия и текущее положение дел	571
13.2.1. Генеративные модели	571
13.2.2. Модели динамической байесовской сети (DBN).....	571
13.2.3. Вариационный автокодировщик	573
13.2.4. Типы аномалий и методы обнаружения аномалий	574
13.2.5. Обнаружение аномалий в данных низкой размерности.....	577
13.2.6. Обнаружение аномалий в многомерных данных.....	578
13.3. Архитектура вычисления аномалии в самосознательных системах	579
13.3.1. Общее описание архитектуры	579
13.3.2. Модель обобщенной динамической байесовской сети (GDBN).....	581
13.3.3. Алгоритм логического вывода в реальном времени.....	584
13.3.4. Измерения мультимодальных аномалий	586
13.3.4.1. Дискретный уровень.....	588

13.3.4.2. Непрерывный уровень	588
13.3.4.3. Уровень наблюдения	589
13.3.5. Использование обобщенных ошибок для непрерывного обучения.....	589
13.4. Пример: обнаружение аномалий в мультисенсорных данных от автомобиля с самосознанием.....	590
13.4.1. Описание условий эксперимента.....	590
13.4.2. Обучение модели DBN	591
13.4.3. Многоуровневое обнаружение аномалий.....	592
13.4.3.1. Задача объезда пешеходов.....	593
13.4.3.2. Задача разворота	594
13.4.3.3. Аномалии на уровне изображения	596
13.4.3.4. Оценка обнаружения аномалий.....	596
13.4.4. Аномалии проприоцептивных сенсорных данных	598
13.4.5. Дополнительные результаты	599
13.5. Выводы.....	600
Литературные источники	600
Об авторах главы	603

Глава 14. Методы PnP и глубокой развертки

для восстановления изображения	605
14.1. Введение	605
14.2. Алгоритм полуквадратичного разделения (HQS)	609
14.3. Глубокое восстановление изображения по методу PnP	610
14.3.1. Предварительное изучение глубокого шумоподавителя CNN	612
14.3.1.1. Шумоподавляющая сетевая архитектура	613
14.3.2. Методика обучения	614
14.3.3. Результаты удаления шума	615
14.3.3.1. Удаление шума с изображений в градациях серого.....	615
14.3.3.2. Удаление шума с цветного изображения.....	616
14.3.4. Алгоритм HQS для PnP IR	617
14.3.4.1. Алгоритм полуквадратичного разделения (HQS).....	617
14.3.4.2. Общая методика настройки параметров.....	617
14.3.4.3. Периодический геометрический самосогласованный ансамбль	618
14.4. Восстановление изображения методом глубокой развертки.....	619
14.4.1. Сеть глубокой развертки.....	620
14.4.1.1. Модуль данных \mathcal{D}	620
14.4.1.2. Модуль приора \mathcal{P}	620
14.4.1.3. Модуль гиперпараметров \mathcal{H}	621
14.4.2. Сквозное обучение	622
14.5. Эксперименты	622
14.5.1. Устранение размытия изображения	623
14.5.1.1. Количественные и качественные результаты.....	624
14.5.1.3. Промежуточные результаты.....	625
14.5.2. Сверхразрешение одиночного изображения (SISR).....	627

14.5.2.1. Количественное и качественное сравнение.....	628
14.6. Заключение	632
Благодарности	633
Литературные источники	633
Об авторах главы.....	638

Глава 15. Атаки на визуальные системы и защита

от злоумышленников	640
15.1. Введение.....	640
15.2. Определение проблемы	641
15.3. Свойства состязательной атаки	643
15.4. Типы возмущений.....	644
15.5. Сценарии атаки	645
15.5.1. Целевые модели	645
15.5.1.1. Модели для задач, связанных с изображениями.....	648
15.5.1.2. Модели для видеозадач	649
15.5.2. Наборы данных и метки	651
15.5.2.1. Наборы данных изображений	651
15.5.2.2. Наборы видеоданных	652
15.6. Обработка изображений	654
15.7. Классификация изображений.....	655
15.7.1. Белый ящик, ограниченные атаки	655
15.7.2. Белый ящик, атаки на основе контента.....	659
15.7.3. Атаки методом черного ящика	659
15.8. Семантическая сегментация и обнаружение объектов	661
15.9. Отслеживание объекта	662
15.10. Классификация видео	664
15.11. Защита от состязательных атак противника	666
15.11.1. Обнаружение атаки	666
15.11.2. Маскировка градиента.....	668
15.11.3. Устойчивость модели	670
15.12. Выводы.....	672
Благодарность.....	673
Литературные источники	673
Об авторах главы.....	682

Предметный указатель.....	683
----------------------------------	------------