

Посвящаю эту книгу моей семье.
Светлой памяти моих родителей, Артура и Мэри Дэвис.
Моей жене Джоан за любовь, терпение, поддержку и вдохновение.
Моим детям Элизабет, Саре и Марион и внукам Джасперу, Джерому, Еве,
Таре и Пиа за то, что принесли бесконечную радость в мою жизнь!
— *Рой Дэвис*

Эта книга посвящается студентам, коллегам, друзьям и членам семьи,
которые мотивировали, направляли и поддерживали меня разными
способами, – всех невозможно перечислить поименно.
Моей жене Келли и детям Ханне и Мэтту – особая благодарность
и признательность за вашу любовь и вдохновение.
— *Мэтью Терк*

Содержание

От издательства	17
Список соавторов	18
О редакторах	20
Предисловие	21
Глава 1. Кардинальные переменны 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. Многоакурсное 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 многоакурсного 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	137
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 потерь	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-мат множественное 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.....

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.....

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.....

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	390
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

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

От издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте www.dmkpress.com, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу dmkpress@gmail.com; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу http://dmkpress.com/authors/publish_book/ или напишите в издательство по адресу dmkpress@gmail.com.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг, мы будем очень благодарны, если вы сообщите о ней главному редактору по адресу dmkpress@gmail.com. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательства «ДМК Пресс» и Elsevier очень серьезно относятся к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу электронной почты dmkpress@gmail.com.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

Список соавторов

Сатьянараяна Аакур, факультет информатики, Государственный университет Оклахомы, Стиллуотер, Оклахома, США

Йогеш Балахи, факультет информатики и UMACS, Мэрилендский университет, Колледж-Парк, Мэриленд, США

Хан Цай, Массачусетский технологический институт, Кембридж, Массачусетс, США

Чжаовой Цай, Amazon Web Services, Пасадена, Калифорния, США

Андреа Кавалларо, Центр интеллектуального восприятия, Лондонский университет Королевы Марии, Лондон, Соединенное Королевство

Рама Челлаппа, факультеты электроники, вычислительной техники и биомедицинской инженерии, Университет Джона Хопкинса, Балтимор, Мэриленд, США

Дондон Чен, Microsoft Cloud & AI, Редмонд, Вашингтон, США

Э. Р. Дэвис Ройал Холлоуэй, Лондонский университет, Эгам, графство Суррей, Соединенное Королевство

Михаэль Фельсберг, Лаборатория компьютерного зрения, факультет электроники, Линчепингский университет, Линчёпинг, Швеция; Инженерная школа Университета Квазулу-Натал, Дурбан, Южная Африка

Корнелия Фермюллер, Университет Мэриленда, Институт перспективных компьютерных исследований, Центр компьютерных наук и инженерии Ирибе, Колледж-Парк, Мэриленд, США

Эфстратиос Гаввес, Институт информатики при Амстердамском университете, Амстердам, Нидерланды

Дипак Гупта, Институт информатики при Амстердамском университете, Амстердам, Нидерланды

Сонг Хан, Массачусетский технологический институт, Кембридж, Массачусетс, США

Ганг Хуа, Wormpex AI Research, Белвью, Вашингтон, США

Али Краяни, DITEN, Генуэзский университет, Генуя, Италия

Цзи Линь, Массачусетский технологический институт, Кембридж, Массачусетс, США

Лучио Марсенаро, DITEN, Генуэзский университет, Генуя, Италия

Майкл Мейнорд, Университет Мэриленда, факультет компьютерных наук, Центр компьютерных наук и инженерии Ирибе, Колледж-Парк, Мэриленд, США

Умберто Микьели, кафедра информационных технологий, Университет Падуи, Падуя, Италия

Рами Мунир, кафедра вычислительной техники и технологии, Университет Южной Флориды, Тампа, Флорида, США

Хиен Нгуен, факультет электроники и вычислительной техники, Хьюстонский университет, Хьюстон, Техас, США

- Чанги О**, Центр интеллектуального восприятия, Лондонский университет Королевы Марии, Лондон, Соединенное Королевство
- Суджой Пол**, Google Research, Бангалор, Индия
- Карло Регаццони**, DITEN, Генуэзский университет, Генуя, Италия
- Амит Рой-Чоудхури**, факультет электроники и вычислительной техники, Калифорнийский университет, Риверсайд, Калифорния, США
- Судип Саркар**, кафедра компьютерных наук и технологии, Университет Южной Флориды, Тампа, Флорида, США
- Джулия Славик**, DITEN, Генуэзский университет, Генуя, Италия
- Раду Тимофте**, Лаборатория компьютерного зрения, ETH Zürich, Цюрих, Швейцария
- Марко Тольдо**, кафедра информационных технологий, Университет Падуи, Падуя, Италия
- Хасан Угайл**, Центр цифровой обработки визуальной информации, Университет Брэдфорда, Брэдфорд, Великобритания
- Нуно Васконселос**, Калифорнийский университет в Сан-Диего, факультет электроники и вычислительной техники, Сан-Диего, Калифорния, США
- Алессио Зомперо**, Центр интеллектуального восприятия, Лондонский университет Королевы Марии, Лондон, Соединенное Королевство
- Пьетро Зануттиг**, кафедра информационных технологий, Университет Падуи, Падуя, Италия
- Кай Чжан**, Лаборатория компьютерного зрения, ETH Zürich, Цюрих, Швейцария

О редакторах

Рой Дэвис – почетный профессор факультета машинного зрения в Роял Холлоуэй, Лондонский университет. Он работал над многими аспектами зрения, от обнаружения признаков и подавления шума до робастного сопоставления образов и реализации практических задач зрения в реальном времени. Область его интересов включает автоматизированный осмотр объектов, наблюдение, управление транспортными средствами и раскрытие преступлений. Он опубликовал более 200 статей и три книги: *Machine Vision: Theory, Algorithms, Practicalities* (1990 г.), *Electronics, Noise and Signal Recovery* (1993 г.) и *Image Processing for the Food Industry* (2000 г.); первая из них не теряет популярности на протяжении 25 лет, а в 2017 г. вышло ее значительно расширенное пятое издание под названием *Computer Vision: Principles, Algorithms, Applications, Learning*. Рой является членом IoP и IET, а также старейшим членом IEEE. Он входит в редакционные коллегии журналов *Pattern Recognition Letters*, *Real-Time Image Processing*, *Imaging Science and IET Image Processing*. Он получил степень доктора наук в Лондонском университете; в 2005 г. был удостоен титула почетного члена BMVA, а в 2008 г. стал лауреатом премии Международной ассоциации распознавания образов.

Мэтью Тёрк – президент Технологического института Toyota в Чикаго (TTIC) и почетный профессор Калифорнийского университета в Санта-Барбаре. Его исследовательские интересы охватывают компьютерное зрение и взаимодействие человека с компьютером, включая такие темы, как автономные транспортные средства, распознавание лиц и жестов, мультимодальное взаимодействие, компьютерная фотография, дополненная и виртуальная реальность и этика ИИ. Он был главным организатором или ведущим нескольких крупных конференций, включая конференцию IEEE по компьютерному зрению и распознаванию образов, мультимедийную конференцию ACM, конференцию IEEE по автоматическому распознаванию лиц и жестов, международную конференцию ACM по мультимодальному взаимодействию и Зимнюю конференцию IEEE по приложениям компьютерного зрения. Он получил несколько наград за лучшую исследовательскую работу, а также различные премии и награды ACM, IEEE, IAPR и почетную премию Фулбрайта-Nokia за 2011–2012 гг. в области информационных и коммуникационных технологий.

Предисловие

Миновало почти десятилетие с тех пор, как произошел прорыв в разработке и применении *глубоких нейронных сетей* (deep neural network, DNN), и их последующий прогресс можно почти без преувеличения назвать выдающимся. Правда, этому прогрессу значительно способствовало появление специального оборудования в виде мощных графических процессоров; кроме того, возникло понимание, что *сверточные нейронные сети* (convolutional neural network, CNN) составляют важнейшую архитектурную основу, в которую можно встроить такие функции, как ReLU, упаковку, полностью связанные слои, распаковку и обратную свертку. По сути, все эти подходы помогли вдохнуть реальную жизнь в глубокие нейросети и резко расширить возможности их использования, поэтому первоначальный почти экспоненциальный рост их использования сохранился на весь последующий период. Мало того, что мощь нейросетевых технологий была впечатляющей, их применение значительно расширилось: от первоначального акцента на быстрое определение местоположения объекта и сегментацию изображения – и даже семантическую сегментацию – до применений, относящихся к видео, а не просто к анализу статичного изображения.

Было бы неправильно утверждать, что все развитие компьютерного зрения с 2012 г. было связано исключительно с появлением DNN. Свою роль сыграли и другие важные методы, такие как обучение с подкреплением, обучение с переносом, самообучение, лингвистическое описание изображений, распространение меток и такие приложения, как обнаружение новизны и аномалий, раскрашивание и отслеживание изображений. Тем не менее многие из упомянутых методов и области их применения получили новые стимулы и были пересмотрены и улучшены благодаря чрезвычайно быстрому внедрению DNN.

В этой книге мы попытались оценить, какие изменения произошли в области компьютерного зрения за минувшее десятилетие, насыщенное драматическими переменами. Сейчас самое время задаться вопросом, где мы находимся сейчас и насколько прочна база глубокого нейронного и машинного обучения, на которую опирается современное компьютерное зрение. Было ли это продуманное последовательное движение или слепой отчаянный рывок вперед? Не упускаем ли мы важные возможности и можем ли мы заглядывать в будущее с уверенностью, что движемся в правильном направлении? Или это тот случай, когда каждый исследователь может придерживаться своей собственной точки зрения и обращать внимание только на то, что представляется необходимым для его прикладной области, и если это так, то не ускользает ли от нас что-то важное при столь ограниченном подходе?

На самом деле есть и другие фундаментальные вопросы, на которые нужно найти ответ. Например, это сложный вопрос о том, до какой степени возможности глубокой нейросети можно повышать за счет качества обучающих данных; этот вопрос, по-видимому, применим к любому альтернативному

подходу, основанному на машинном обучении, независимо от того, относится ли он к DNN. Вряд ли фундаментальные ограничения нейросети зависят от того, каким способом ее обучали – обучение с подкреплением, самообучение или что-то другое. И обратите внимание, что люди вряд ли являются примером того, что можно каким-либо образом избежать интенсивного обучения; их способность к обучению с переносом лишь подтверждает, насколько эффективным может быть процесс обучения.

В этой книге мы стремимся не только представить передовые методики и подходы в области компьютерного зрения, но и разъяснить основополагающие принципы; мы взяли на себя роль преподавателей и, прежде чем представить читателю самые последние достижения, хотим сформировать у него понимание общей картины. Поэтому *глава 1* посвящена основам компьютерного зрения. Она начинается с детального анализа ранних подходов к компьютерному зрению, включая обнаружение признаков, обнаружение объектов, трехмерное зрение и появление DNN; далее мы переходим к визуальному слежению за объектами, которое рассматривается как пример прикладной области, где решающую роль могут играть DNN. Эта глава самая длинная в книге, потому что мы должны пройти путь с нуля до современных достижений; кроме того, она готовит почву для понимания ключевых идей и методов, описанных выдающимися экспертами в остальных главах.

Как будет показано в *главе 1*, обнаружение объектов – одна из самых сложных задач компьютерного зрения. В частности, эффективная система должна преодолевать такие проблемы, как искажение масштаба, окклюзия, переменное освещение, сложный фон и все факторы изменчивости, связанные с миром природы. *Глава 2* описывает различные методы и подходы, на которых основаны последние достижения. К ним относятся *слияние видимых областей* (region-of-interest pooling), *многозадачные потери* (multitask losses), *сети для предложения регионов* (region proposal networks), *привязки* (anchors), *каскадное обнаружение и регрессия* (cascaded detection and regression), *многомасштабные представления признаков* (multiscale feature representations), *методы дополнения данных* (data augmentation techniques), *функции потерь* (loss functions) и многое другое.

В *главе 3* подчеркивается, что недавние успехи в области компьютерного зрения в значительной степени связаны с появлением огромных массивов тщательно размеченных данных, необходимых для обучения моделей. В ней рассматриваются методы, которые можно использовать для обучения моделей распознавания на основе таких данных, требующие ограниченной ручной обработки. Помимо уменьшения количества размеченных вручную данных, необходимых для обучения моделей распознавания, необходимо снизить уровень подкрепления с сильного на слабый, в то же время разрешая релевантные запросы от оракула. Дан обзор теоретических основ и экспериментальных результатов, которые помогают достичь этого.

В *главе 4* рассматриваются вычислительные проблемы глубоких нейронных сетей, которые затрудняют их развертывание на оборудовании с ограниченными ресурсами. В ней обсуждаются методы сжатия моделей и поиска нейронной архитектуры, ориентированной на оборудование, с целью повышения эффективности глубокого обучения, уменьшения размера и ускоре-

ния нейронных сетей. В главе показано, как использовать *отсечение коэффициентов* (parameter pruning) для удаления избыточных весов, *факторизацию низкого ранга* (low-rank factorization) для уменьшения сложности, *квантование весов* (weight quantization) для уменьшения точности весов и размера модели, а также *дистилляцию знаний* (knowledge distillation) для переноса знаний из «черного ящика» больших моделей в меньшие.

В главе 5 обсуждается, как *глубокие генеративные модели* (deep generative models) пытаются восстановить низкоразмерную структуру целевых визуальных моделей. В ней показано, как использовать глубокие генеративные модели для достижения более управляемого синтеза визуальных паттернов посредством условной генерации изображения. Ключом к достижению этой цели является «распутывание» визуального представления, когда предпринимаются попытки разделить различные управляющие факторы в скрытом пространстве встраивания. Представлены три тематических исследования по *переносу стиля* (style transfer), *визуально-языковой генерации* (vision-language generation) и *синтезу лица* (face synthesis), чтобы проиллюстрировать, как этого добиться в условиях обучения без подкрепления или при слабом подкреплении.

Глава 6 посвящена актуальной проблеме реального мира – *распознаванию лиц* (face recognition). В ней обсуждаются современные методы, основанные на глубоком обучении, которые можно применять даже к неполным изображениям лица. В главе показано: (а) как создаются необходимые архитектуры глубокого обучения; (б) как такие модели можно обучать и тестировать; (с) как можно использовать *точную настройку* (fine tuning) предварительно обученных сетей для определения эффективных сигналов распознавания с полными и частичными данными о лице; (d) какие успехи достигнуты за счет последних разработок в области глубокого обучения; (е) каковы текущие ограничения методов глубокого обучения, используемых для распознавания лиц. В главе также упомянуты некоторые из нерешенных проблем в этой области.

В *главе 7* обсуждается важнейший вопрос о том, как перенести обучение из одной области данных в другую. Сюда относятся методы, основанные на дифференциальной геометрии, *разреженном представлении* (sparse representation) и глубоких нейронных сетях. Они делятся на два широких класса – дискриминационные и генеративные подходы. Первые включают обучение модели классификатора с использованием дополнительных потерь, чтобы сделать исходное и целевое распределения признаков похожими. Вторые используют генеративную модель для выполнения адаптации предметной области (домена): обычно междоменная генеративная состязательная сеть обучается для сопоставления образцов из исходного домена с целевым, а модель классификатора обучается на преобразованных целевых изображениях. Такие подходы проверяются на задачах междоменного распознавания и семантической сегментации.

В *главе 8* мы возвращаемся к задаче адаптации предметной области в контексте семантической сегментации, когда глубокие сети испытывают потребность в огромном количестве размеченных данных для обучения. Глава начинается с обсуждения различных уровней, на которых может осуществ-

ляться адаптация, и стратегий их достижения. Затем рассматривается задача непрерывного обучения семантической сегментации. Хотя эта задача является относительно новой областью исследований, интерес к ней быстро растет, и уже представлено множество различных сценариев. Они подробно описаны вместе с подходами, необходимыми для их решения.

Вслед за главой 1 в *главе 9* вновь подчеркивается важность визуального отслеживания как одной из основных классических проблем компьютерного зрения. Цель этой главы – дать обзор развития области, начиная с алгоритма Лукаса–Канаде и согласованных фильтров и заканчивая подходами, основанными на глубоком обучении, а также переходом к сегментации видео. Обзор ограничен целостными моделями для общего отслеживания в плоскости изображения, и особое внимание уделяется дискриминационным моделям, трекеру MOSSE (minimum output sum of squared errors, минимальная выходная сумма квадратов ошибок) и DCF (discriminative correlation filters, дискриминационные корреляционные фильтры).

Глава 10 развивает концепцию визуального отслеживания объектов еще на один шаг и концентрируется на долгосрочном отслеживании. Чтобы успешно справиться с этой задачей, отслеживание объектов должно решать серьезные проблемы, связанные с *распадом модели* (model decay), то есть с ухудшением качества модели из-за нарастающей погрешности, а также с исчезновением и появлением цели. Успех глубокого обучения оказал большое влияние на подходы к отслеживанию визуальных объектов, поскольку автономное обучение *сиамских трекеров* (Siamese tracker) помогает устранить распад модели. Однако, чтобы избежать возможности потери отслеживания в тех случаях, когда внешний вид цели значительно меняется, сиамские трекеры могут воспользоваться встроенными инвариантностями и эквивариантностями, допускающими вариации внешнего вида, не усугубляя распад модели.

В последние годы крепнет уверенность в том, что в динамичной среде видео и движущихся объектов – особенно когда идет речь о *распознавании действий и поведения* (action/behavior recognition) – жизненно важную роль играет понимание когнитивных функций человека. Обоснованность этого предположения полностью подтверждают следующие две главы. В *главе 11* описывается ориентированная на действия структура, которая охватывает несколько временных масштабов и уровней абстракции. Нижний уровень детализирует *характеристики объекта*, который совершает различные действия; средний уровень моделирует *отдельные действия*, а самый высокий уровень моделирует *деятельность*. Упор на использование характеристик понимания, геометрии, онтологий и ограничений, основанных на физике, позволяет избежать чрезмерного обучения характеристикам внешнего вида. Чтобы объединить восприятие на основе сигналов с *символьными знаниями* (symbolic knowledge), векторизованные знания согласовываются с визуальными признаками. Глава также включает обсуждение понятий *действия* (action) и *деятельности* (activity).

В *главе 12* рассматривается проблема *временной сегментации событий* (temporal event segmentation). Достижения когнитивной науки демонстрируют подходы к разработке высокоэффективных алгоритмов компьютерного зрения для пространственно-временной сегментации событий в видео

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru