

O'REILLY®

Искусственный интеллект и компьютерное зрение

Реальные проекты
на Python, Keras
и TensorFlow



Анирад Коул
Сиддха Ганджу, Мехер Казам

Краткое содержание

Вступление	20
Глава 1. Обзор ландшафта искусственного интеллекта	32
Глава 2. Что на картинке: классификация изображений с помощью Keras	69
Глава 3. Кошки против собак: перенос обучения с помощью Keras в 30 строках кода	83
Глава 4. Создание механизма обратного поиска изображений. Эмбеддинги	112
Глава 5. От новичка до мастера прогнозирования: увеличение точности сверточной нейронной сети	151
Глава 6. Увеличение скорости и эффективности TensorFlow: удобный чек-лист	183
Глава 7. Практические инструменты, советы и приемы	225
Глава 8. Облачные API для компьютерного зрения: установка и запуск за 15 минут	237
Глава 9. Масштабируемый инференс в облаке с помощью TensorFlow Serving и KubeFlow	274
Глава 10. ИИ в браузере с TensorFlow.js и ml5.js	304
Глава 11. Классификация объектов в реальном времени в iOS с Core ML	338
Глава 12. Not Hotdog на iOS с Core ML и Create ML	379

Глава 13. Шазам для еды: разработка приложений для Android с помощью TensorFlow Lite и ML Kit.....	400
Глава 14. Создание приложения Purrfect Cat Locator с помощью TensorFlow Object Detection API.....	445
Глава 15. Как стать творцом: ИИ в краевых устройствах	487
Глава 16. Моделирование беспилотного автомобиля методом сквозного глубокого обучения с использованием Keras.....	519
Глава 17. Создание беспилотного автомобиля менее чем за час: обучение с подкреплением с помощью AWS DeepRacer	553
Приложение. Краткое введение в сверточные нейронные сети	596
Об авторах	605
Иллюстрация на обложке.....	608

Оглавление

Вступление.....	20
Разработчикам.....	21
Специалистам по данным.....	22
Студентам.....	22
Преподавателям	23
Энтузиастам робототехники	23
Обзор глав	23
Условные обозначения.....	27
Использование исходного кода примеров.....	27
Благодарности.....	28
Коллективные благодарности.....	28
Личные благодарности	30
Глава 1. Обзор ландшафта искусственного интеллекта.....	32
Извинения.....	33
Настоящее вступление	34
Что такое ИИ?	34
Мотивирующие примеры	35
Краткая история ИИ	37
Захватывающее начало.....	38
Холодные и мрачные дни	39
Проблеск надежды	40
Как глубокое обучение вошло в моду.....	43
Рецепт идеального решения задачи глубокого обучения	46
Датасеты.....	47
Архитектура модели	50
Фреймворки.....	52
Ответственный ИИ	59
Предвзятость	60
Прозрачность и объяснимость.....	63
Воспроизводимость	64
Устойчивость	64
Конфиденциальность	65

Итоги	66
Часто задаваемые вопросы	66
Глава 2. Что на картинке: классификация изображений с помощью Keras	69
Введение в Keras	70
Классификация изображений	71
Исследование модели.....	75
Датасет ImageNet.....	76
Зоопарк моделей.....	78
Карты активации классов	79
Итоги	82
Глава 3. Кошки против собак: перенос обучения с помощью Keras	
в 30 строках кода	83
Адаптация предварительно обученных моделей к новым задачам	84
Неглубокое погружение в сверточные нейронные сети	85
Перенос обучения.....	87
Тонкая настройка.....	88
Сколько слоев выбрать для тонкой настройки.....	89
Создание специального классификатора методом переноса обучения с использованием Keras	90
Организация данных	91
Создание пайплайна обработки данных.....	94
Количество классов.....	94
Размер пакета.....	95
Аугментация данных	96
Определение модели	99
Обучение модели	100
Настройка параметров обучения	100
Запуск обучения.....	101
Тестируем модель	102
Анализ результатов	103
Для дальнейшего чтения	110
Итоги	111
Глава 4. Создание механизма обратного поиска изображений.	
Эмбеддинги	112
Сходство изображений.....	113
Извлечение признаков	116
Поиск сходств.....	119

Визуализация кластеров изображений с помощью t-SNE.....	123
Увеличение скорости поиска сходств	126
Длина векторов признаков.....	126
Уменьшение длины векторов признаков с помощью метода главных компонент.....	128
Масштабирование поиска сходства с помощью метода приближительных ближайших соседей	132
Бенчмарк метода приближительных ближайших соседей	133
Какую библиотеку выбрать?.....	134
Создание синтетического датасета	135
Полный перебор	135
Appnoy.....	136
NGT	137
Faiss	137
Повышение точности с помощью тонкой настройки.....	138
Тонкая настройка без полносвязных слоев	141
Сиамские сети для распознавания лица однократным (one-shot) обучением.....	142
Примеры из практики	144
Flickr	144
Pinterest	145
Двойники знаменитостей	146
Spotify.....	146
Описание изображений	148
Итоги	150
Глава 5. От новичка до мастера прогнозирования: увеличение точности сверточной нейронной сети	151
Что понадобится для работы	152
TensorFlow Datasets.....	153
TensorBoard.....	154
Инструмент What-If	157
tf-explain.....	161
Стандартные приемы для экспериментов с машинным обучением	163
Проверка данных	164
Разбиение данных на обучающую, проверочную и контрольную выборки	164
Ранняя остановка	165
Воспроизводимость экспериментов.....	166

Пример сквозного пайплайна глубокого обучения.....	166
Простой пайплайн переноса обучения	166
Простой пайплайн создания сети.....	169
Влияние гиперпараметров на точность	169
Сравнение переноса обучения и обучения с нуля.....	170
Влияние количества слоев для тонкой настройки при переносе обучения.....	171
Влияние объема данных на перенос обучения.....	172
Влияние скорости обучения	173
Влияние оптимизатора	174
Влияние размера пакета	175
Влияние изменения размеров	176
Влияние изменения соотношения сторон на перенос обучения.....	177
Инструменты автоматизации настройки моделей для достижения максимальной точности.....	178
Keras Tuner.....	178
AutoAugment	180
AutoKeras	181
Итоги	182
Глава 6. Увеличение скорости и эффективности TensorFlow: удобный чек-лист	183
Голодание GPU	183
nvidia-smi.....	184
Профилировщик TensorFlow + TensorBoard	186
Как использовать этот чек-лист.....	187
Чек-лист настроек производительности.....	188
Подготовка данных	188
Чтение данных	188
Аугментация данных	188
Обучение	189
Инференс.....	189
Подготовка данных.....	190
Сохраните данные в формате TFRecord	190
Уменьшите размеры исходных данных.....	191
Используйте TensorFlow Datasets	191
Чтение данных	192
Используйте tf.data	192
Организуйте предварительное извлечение данных	193

Организуйте параллельную обработку на CPU	194
Организуйте параллельный ввод/вывод и обработку.....	194
Разрешите недетерминированный порядок следования данных	194
Кэшируйте данные	195
Включите экспериментальные оптимизации.....	196
Автоматическая настройка значений параметров.....	197
Аугментация данных	198
Используйте GPU для аугментации.....	198
Обучение	200
Используйте автоматическую смешанную точность	200
Используйте пакеты большого размера.....	201
Используйте значения, кратные восьми	203
Определите оптимальную скорость обучения.....	203
Используйте <code>tf.function</code>	205
Переобучите и научите обобщать	207
Установите оптимизированный программный стек для поддержки оборудования	209
Оптимизируйте количество потоков, выполняющих на CPU параллельно.....	210
Используйте более производительное оборудование	212
Используйте распределенное обучение.....	212
Изучите отраслевые бенчмарки.....	214
Инференс	216
Используйте эффективную модель	216
Используйте квантование модели	219
Прореживайте модели	222
Используйте совмещенные операции.....	223
Включите сохранение состояния GPU	223
Итоги	224
Глава 7. Практические инструменты, советы и приемы	225
Установка.....	225
Обучение	227
Модель	229
Данные	230
Защищенность	233
Обучение и исследования.....	234
Последний вопрос	236

Глава 8. Облачные API для компьютерного зрения: установка и запуск за 15 минут.....	237
Ландшафт API распознавания образов	239
Clarifai	239
Microsoft Cognitive Services	240
Google Cloud Vision	240
Amazon Rekognition.....	242
IBM Watson Visual Recognition	242
Algorithmia.....	244
Сравнение API для распознавания образов.....	245
Предлагаемые услуги	245
Стоимость.....	246
Точность.....	247
Предвзятость	249
Подготовка и использование облачных API	252
Обучение собственного классификатора.....	255
Основные причины плохой работы классификатора	260
Сравнение качества работы собственных классификаторов в разных API ...	261
Настройка производительности облачных API.....	264
Влияние изменения разрешения на API разметки изображений.....	265
Влияние сжатия на API разметки изображений.....	265
Влияние сжатия на API оптического распознавания символов	266
Влияние изменения разрешения на API оптического распознавания символов	266
Примеры.....	267
New York Times	268
Uber	268
Giphy	268
OmniEarth.....	271
Photobucket.....	271
Staples	271
InDro Robotics	272
Итоги	273
Глава 9. Масштабируемый инференс в облаке с помощью TensorFlow Serving и KubeFlow	274
Ландшафт услуг прогнозирования с помощью ИИ	275
Flask: создание собственного сервера	277
Создание REST API с помощью Flask.....	277

Развертывание модели Keras в Flask.....	278
Плюсы использования Flask	279
Минусы использования Flask.....	280
Желаемые качества системы производственного уровня	280
Высокая доступность	280
Масштабируемость.....	281
Низкая задержка.....	281
Географическая доступность	282
Обработка сбоев.....	283
Мониторинг.....	283
Управление версиями модели	283
A/B-тестирование	284
Поддержка нескольких библиотек машинного обучения	284
Google Cloud ML Engine: управляемый стек облачных услуг ИИ	284
Плюсы использования Cloud ML Engine	285
Минусы использования Cloud ML Engine.....	285
Создание API классификации	285
TensorFlow Serving	292
Установка.....	293
KubeFlow.....	294
Пайплайны.....	297
Инструменты управления.....	297
Установка.....	298
Соотношение цены и производительности.....	300
Анализ затрат на управляемый стек Inference-as-a-Service	300
Анализ затрат на создание собственного стека	302
Итоги	303
Глава 10. ИИ в браузере с TensorFlow.js и ml5.js.....	304
Библиотеки машинного обучения на JavaScript: краткая история.....	305
ConvNetJS	306
Keras.js	307
ONNX.js	308
TensorFlow.js	308
Архитектура TensorFlow.js	310
Использование предварительно обученных моделей с TensorFlow.js.....	312
Преобразование модели для использования в браузере	314
Обучение в браузере	314
Извлечение признаков	315

Сбор данных.....	317
Обучение	317
Нагрузка на GPU.....	319
ml5.js	320
PoseNet	322
pix2pix.....	325
Сравнительный анализ и практические рекомендации.....	330
Размер модели.....	331
Время инференса	331
Примеры.....	333
Дирижер	333
TensorSpace.....	334
Metacar.....	335
Классификация фотографий в Airbnb	336
GAN Lab	336
Итоги	337
Глава 11. Классификация объектов в реальном времени в iOS с Core ML.....	338
Жизненный цикл разработки искусственного интеллекта для мобильных устройств	340
Краткая история Core ML	342
Альтернативы фреймворку Core ML.....	343
TensorFlow Lite.....	344
ML Kit	344
Fritz	345
Архитектура машинного обучения Apple	345
Предметно-ориентированные фреймворки	346
ML Framework	346
Оптимизированные примитивы ML	346
Приложение для распознавания объектов в реальном времени.....	347
Конвертация моделей в формат Core ML	354
Конвертация из формата Keras.....	354
Конвертация из формата TensorFlow.....	354
Развертывание динамической модели	356
Обучение на устройстве	357
Федеративное обучение	358
Анализ качества моделей.....	358
Бенчмарк моделей на iPhone.....	359

Оценка энергопотребления	363
Оценка нагрузки	365
Уменьшение размера приложения	368
Не внедряйте модели в приложение	368
Используйте квантование	369
Используйте Create ML	370
Примеры приложений	371
Magic Sudoku	371
Seeing AI	372
HomeCourt	373
InstaSaber + YoPuppet	374
Итоги	377
Глава 12. Not Hotdog на iOS с Core ML и Create ML	379
Сбор данных	381
Подход 1: поиск готового или создание своего датасета	381
Подход 2: загрузка изображений с помощью расширения Fatkun для браузера Chrome	382
Подход 3: загрузка с помощью Bing Image Search API	385
Обучение модели	386
Подход 1: с помощью инструментов с веб-интерфейсом	386
Подход 2: с помощью Create ML	390
Подход 3: тонкая настройка с использованием Keras	396
Конвертация модели с использованием Core ML Tools	397
Создание приложения для iOS	397
Что можно сделать дальше	398
Итоги	399
Глава 13. Шазам для еды: разработка приложений для Android с помощью TensorFlow Lite и ML Kit	400
Цикл разработки приложения для классификации блюд	401
Обзор TensorFlow Lite	403
Архитектура TensorFlow Lite	406
Конвертация модели в формат TensorFlow Lite	407
Создание приложения для распознавания объектов	408
ML Kit + Firebase	415
Классификация объектов в ML Kit	417
Использование своих моделей в ML Kit	417
Модели, размещенные в облаке	419

А/В-тестирование моделей, размещенных в облаке	423
Использование эксперимента в коде	426
TensorFlow Lite в iOS	427
Оптимизация производительности.....	428
Квантование с помощью TensorFlow Lite Converter	428
Набор инструментов TensorFlow для оптимизации моделей	428
Fritz	429
Целостный взгляд на цикл разработки мобильных приложений ИИ.....	432
Как собирать данные?	433
Как размечать данные?	434
Как обучить модель?	434
Как конвертировать модель в формат для мобильных устройств?.....	434
Как оптимизировать производительность модели?.....	435
Как повысить привлекательность для пользователей?	435
Как развернуть модель?	436
Как оценить успешность модели?	436
Как совершенствовать модель?.....	436
Как обновить модель на телефонах пользователей?	437
Самосовершенствующаяся модель	437
Примеры из практики	439
Lose It!.....	439
Режим портретной съемки на телефонах Pixel 3	441
Распознавание голоса от Alibaba	442
Определение контуров лица с помощью ML Kit.....	442
Сегментация видео в реальном времени в YouTube Stories	443
Итоги	444
Глава 14. Создание приложения Purrfect Cat Locator с помощью TensorFlow Object Detection API.....	445
Виды задач компьютерного зрения	446
Классификация	447
Локализация	447
Обнаружение.....	447
Сегментация.....	448
Способы обнаружения объектов.....	450
Использование готовых облачных API обнаружения объектов.....	451
Использование предварительно обученных моделей.....	453
Получение модели.....	453

Тест-драйв модели	454
Развертывание на устройстве	455
Создание своей модели обнаружения объектов без программирования	456
Развитие технологии обнаружения объектов	460
Вопросы производительности.....	462
Ключевые термины в обнаружении объектов	463
Intersection over Union.....	464
Mean Average Precision.....	465
Non-Maximum Suppression.....	465
Создание своих моделей с помощью TensorFlow Object Detection API	466
Сбор данных.....	467
Разметка данных	469
Предварительная обработка данных.....	473
Исследование модели	474
Обучение	476
Конвертация модели.....	478
Сегментация изображений.....	479
Примеры из практики	481
Умный холодильник	481
Подсчет толпы.....	482
Распознавание лиц в приложении Seeing AI	483
Беспилотные автомобили.....	484
Итоги	486
Глава 15. Как стать творцом: ИИ в краевых устройствах.....	487
Обзор краевых устройств ИИ	488
Raspberry Pi.....	489
Intel Movidius Neural Compute Stick.....	491
Google Coral USB Accelerator.....	492
NVIDIA Jetson Nano.....	494
FPGA + PYNQ	496
Arduino.....	499
Сравнение краевых устройств для ИИ.....	501
Raspberry Pi.....	504
Ускорение с помощью Google Coral USB Accelerator.....	506
NVIDIA Jetson Nano	508
Сравнение производительности краевых устройств	511
Примеры из практики	512

JetBot	512
Билеты на проезд в метро за приседания	514
Сортировщик огурцов	515
Что изучать дальше	517
Итоги	517
Глава 16. Моделирование беспилотного автомобиля методом сквозного глубокого обучения с использованием Keras.....	519
Краткая история автоматизации вождения.....	521
Глубокое обучение, автономное вождение и проблема данных.....	522
«Hello, World!» в автономном вождении: управление в моделируемой среде	525
Инструменты и требования	525
Исследование и подготовка данных	528
Определение области интереса	530
Аугментация данных	532
Дисбаланс датасета и стратегии вождения	533
Обучение модели автопилота	538
Генератор данных.....	539
Определение модели	542
Развертывание модели автопилота	547
Что изучать дальше	550
Расширение датасета.....	551
Обучение на последовательных данных.....	551
Обучение с подкреплением	552
Итоги	552
Глава 17. Создание беспилотного автомобиля менее чем за час: обучение с подкреплением с помощью AWS DeepRacer	553
Краткое введение в обучение с подкреплением	554
Почему для изучения обучения с подкреплением выбран беспилотный автомобиль?.....	555
Практика глубокого обучения с подкреплением с DeepRacer.....	557
Создание первой модели обучения с подкреплением	560
Шаг 1: создание модели.....	561
Шаг 2: настройка процесса обучения	562
Шаг 3: обучение модели	570
Шаг 4: оценка качества модели.....	571
Обучение с подкреплением на практике.....	572

Как происходит обучение с подкреплением?	572
Теория обучения с подкреплением	577
Алгоритм обучения с подкреплением в AWS DeepRacer	579
Кратко о порядке глубокого обучения с подкреплением на примере DeepRacer	581
Шаг 5: улучшение модели обучения с подкреплением	582
Гонки на автомобиле AWS DeepRacer	587
Создание трека	587
Шаблон трека для AWS DeepRacer с одним поворотом	588
Запуск модели на автомобиле AWS DeepRacer	589
Автономное вождение AWS DeepRacer	589
Что изучать дальше	592
Лига DeepRacer	592
AWS DeepRacer с расширенными возможностями	592
Олимпиада автопилотов с искусственным интеллектом	592
DIY Robocars	593
Roborace	594
Итоги	595
Приложение. Краткое введение в сверточные нейронные сети	596
Машинное обучение	596
Персептрон	596
Функции активации	597
Нейронные сети	598
Обратное распространение ошибки	600
Недостатки нейронных сетей	600
Желаемые свойства классификатора изображений	601
Свертка	601
Объединение	602
Структура сверточной сети	602
Что изучать дальше	604
Об авторах	605
Основные авторы	605
Приглашенные авторы	606
Иллюстрация на обложке	608