

Всем моим учителям – настоящим, прошлым и будущим

Содержание

От издательства	11
Предисловие	12
Благодарности	17
Сокращения	19
Глава 1. Знакомство с миром зрения	21
1.1. Эволюция зрительной системы	22
1.2. Будущее зрения	24
1.3. Почему зрение – не такая простая вещь?.....	26
1.4. Четыре важнейшие особенности зрительного восприятия	27
1.5. Путешествия и приключения фотона	30
1.6. Манипуляции с визуальной системой.....	32
1.7. Функции цепей зрительной коры.....	34
1.8. К нейронным коррелятам зрительного сознания	37
1.9. К разработке теории зрительного восприятия.....	39
1.10. Обзор главы.....	42
Литература	42
Глава 2. Путешествие фотона. Статистика естественного изображения и сетчатка	43
2.1. Естественные изображения имеют свои особенности.....	44
2.2. Эффективное кодирование за счет выделения дополнительных ресурсов там, где они необходимы	46
2.3. Визуальный мир – медленный.....	47
2.4. Наши глаза находятся в непрерывном движении	48
2.5. Сетчатка извлекает информацию из света	51
2.6. Требуется время, чтобы информация достигла зрительного нерва	58
2.7. Зрительные нейроны реагируют на определенную область поля зрения	58
2.8. Оператор разности гауссиан извлекает важную информацию и игнорирует однородные области	60
2.9. Зрительные нейроны реагируют на изменения.....	62
2.10. Движемся дальше, к другим отделам мозга.....	63
2.11. Цифровая камера vs глаз.....	64
2.12. Обзор главы.....	66
Литература	66
Глава 3. Феноменология зрения	68
3.1. Образ внешнего мира, рожденный нашим восприятием, – это совсем не то, что видят наши глаза.....	68

3.2. Зрительный образ зависит от адекватной группировки частей изображения с помощью определенных правил	70
3.3. Целое может быть больше, чем сумма его частей.....	72
3.4. Зрительная система компенсирует искажения в изображениях объектов	72
3.5. Финальная сборка: построение целого из видимых частей	76
3.6. Визуальное распознавание совершается очень быстро.....	78
3.7. Пространственный контекст также имеет значение	83
3.8. Ценность визуального опыта	85
3.9. Люди примерно одинаковые, куда бы вы ни пошли, с некоторыми исключениями.....	87
3.10. Зрение многих животных ничуть не хуже нашего	88
3.11. Обзор главы.....	91
Литература	92

Глава 4. Генерация зрительных образов и изменение зрительного восприятия посредством повреждений и электрической стимуляции мозга.....

4.1. Корреляции и причинность в неврологии	94
4.2. Арсенал инструментов для изучения функциональной роли различных областей мозга у животных.....	95
4.3. Некоторые инструменты для изучения функциональной роли областей человеческого мозга	99
4.4. Частичные повреждения первичной зрительной коры головного мозга приводят к локализованным скотомам	101
4.5. Пути «что» и «где»	105
4.6. Повреждения дорсального потока на пути «где».....	105
4.7. Нижняя височная кора критически важна для распознавания визуальных объектов у обезьян	107
4.8. Повреждения, ведущие к нарушению распознавания зрительных образов у человека	108
4.9. Инвазивная электрическая стимуляция мозга человека.....	112
4.10. Электрическая стимуляция зрительной коры приматов.....	117
4.11. Обзор главы.....	121
Литература	121

Глава 5. Приключения в terra incognita. Исследование нейронных цепей вентрального зрительного потока.....

5.1. О неокортексе	124
5.2. Связь с первичной зрительной корой и обратная связь.....	125
5.3. Золотой стандарт временного разрешения для исследования нейронов	128
5.4. Нейроны в первичной зрительной коре головного мозга избирательно реагируют на полосы, показанные в разной ориентации.....	129
5.5. Сложные нейроны демонстрируют толерантность к изменениям положения стимулов	131

5.6. Соседствующие нейроны проявляют схожие свойства.....	134
5.7. Количественное феноменологическое описание откликов нейронов в первичной зрительной коре.....	135
5.8. Простая модель ориентационной избирательности в первичной зрительной коре.....	136
5.9. Другие сюрпризы V1.....	138
5.10. Разделяй и властвуй.....	140
5.11. Невозможно исчерпывающе изучить реакцию на все возможные визуальные стимулы.....	141
5.12. Мы живем визуальным прошлым: латентность реакции вдоль вентрального потока увеличивается.....	143
5.13. Увеличение размера воспринимающего поля вдоль вентрального зрительного потока.....	145
5.14. Что предпочитают нейроны за пределами V1?.....	146
5.15. Мозг конструирует интерпретацию мира: случай иллюзорных контуров.....	147
5.16. Красочный V4.....	149
5.17. Модуляция внимания.....	149
5.18. Обзор главы.....	150
Литература.....	151

Глава 6. От высших уровней обработки зрительного сигнала к распознаванию образов.....

6.1. Зона с хорошей коммуникацией.....	153
6.2. ИТС-нейроны демонстрируют избирательность к формам.....	153
6.3. Избирательность вентральной зрительной коры человека.....	156
6.4. Чего хотят нейроны ИТС <i>на самом деле?</i>	158
6.5. Нейроны ИТС демонстрируют толерантность к преобразованиям объектов.....	160
6.6. Нейроны могут завершать формы частично видимых объектов.....	161
6.7. Информационные технологии выходят на передний план.....	162
6.8. Нейроны ИТС больше озабочены формой, чем смыслом.....	166
6.9. Адаптация нейронных реакций.....	168
6.10. Представление визуальной информации при отсутствии визуального стимула.....	170
6.11. Цели, поставленные в задаче, модулируют нейронные реакции.....	171
6.12. Роль опыта в формировании настройки предпочтений нейронов.....	173
6.13. Мост между зрением и распознаванием образов.....	174
6.14. Обзор главы.....	175
Литература.....	176

Глава 7. Нейробиологически подобные вычислительные модели.....

7.1. Зачем нужны вычислительные модели?.....	177
7.2. Модели одиночных нейронов.....	180
7.3. Модели сетей.....	185

7.4. Сетевые модели с частотой срабатывания нейронов	189
7.5. Операция свертки.....	189
7.6. Сети Хопфилда.....	192
7.7. Нейронные сети могут решать задачи распознавания.....	195
7.8. Экстремальный биологический реализм: проект «Blue Brain»	197
7.9. Обзор главы	198
Литература	199

Глава 8. Учим компьютеры видеть

8.1. Краткое описание и определения	200
8.2. Общие темы в моделировании вентрального потока обработки зрительного сигнала.....	204
8.3. Арсенал моделей	205
8.4. Общая схема решения задачи распознавания образов	208
8.5. Восходящие иерархические модели вентрального зрительного потока	209
8.6. Изучение весовых параметров	212
8.7. Базы данных названий	218
8.8. Перекрестная проверка необходима	221
8.9. Предупреждение: множество параметров!.....	222
8.10. Известный пример: распознавание цифр в сети прямого распространения, обученной методом градиентного спуска.....	223
8.11. Глубокая сверточная нейронная сеть в действии	224
8.12. Ошибаться свойственно и людям, и алгоритмам	229
8.13. Прогнозирование движений глаз	232
8.14. Прогнозирование частоты возбуждения нейронов.....	237
8.15. Все модели неточны, но некоторые полезны	239
8.16. Горизонтальные и нисходящие сигналы при распознавании образов.....	240
8.17. Предиктивное кодирование	241
8.18. Обзор главы.....	246
Литература	247

Глава 9. К миру с разумными машинами, которые смогут видеть и интерпретировать мир вокруг нас.....

9.1. Визуальный тест Тьюринга.....	249
9.2. Компьютерное зрение повсюду.....	252
9.3. Добавление временной информации с помощью видео.....	257
9.4. Основные этапы классификации объектов.....	258
9.5. Приложения алгоритмов компьютерного зрения для классификации реальных объектов физического мира	261
9.6. Компьютерное зрение может помочь людям со зрительной дисфункцией.....	267
9.7. Глубокие сверточные нейронные сети работают и за пределами зрения	269
9.8. Генераторы изображений и GAN	269

9.9. DeepDream и XDream: настройки вычислительных блоков и биологических нейронов.....	271
9.10. Размышления о перекрестной проверке и экстраполяции.....	273
9.11. Состязательные изображения.....	276
9.12. Обманчиво простые задачи, бросающие вызов алгоритмам компьютерного зрения.....	278
9.13. Вызовы, которые стоят впереди.....	280
9.14. Обзор главы.....	285
Литература.....	286
Глава 10. Зрительное сознание.....	287
10.1. Неполный список возможных ответов.....	289
10.2. Поиск NCC: нейронные корреляты сознания.....	294
10.3. Осознанный образ должен быть «явным».....	295
10.4. Экспериментальные подходы к изучению зрительного восприятия.....	296
10.5. Нейронные корреляты зрительного восприятия во время бинокулярного соперничества.....	301
10.6. Требования к NCC.....	304
10.7. Интегрированная теория информации.....	305
10.8. Обзор главы.....	308
Литература.....	309
Предметный указатель.....	310

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 Cambridge University Press 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.

Предисловие

Мне выпала честь вести занятия по биологическому и компьютерному зрению для студентов и аспирантов вот уже почти десять лет. Курс состоит из 10 лекций и активного обсуждения классической и современной литературы в этой области. За это время мой курс прошли множество выдающихся студентов; они сочетали академические, социальные занятия и занятия спортом; выигрывали стипендии, расставались со своими парнями и подругами или женились на своих возлюбленных; ставили эксперименты в лабораториях; проходили стажировку в отраслях компьютерного зрения; получали свои первые академические отзывы; публиковали свои первые статьи и в конце концов закончили обучение. После окончания учебы многие студенты продолжают нести факел исследовательской деятельности в области зрения; другие стали философами; некоторые создали свои стартапы; некоторые стали звездами в этой области; некоторые решили искать другие варианты карьеры.

На протяжении многих лет общение со студентами воодушевляло меня, заставляло думать, пересматривать некоторые из моих убеждений и побуждало меня объяснять некоторые вещи по-новому. Эта книга является следствием всех этих взаимодействий, всех этих итераций и всех этих обсуждений. В рассказе, который я пишу, слышны голоса, надежды, недоумения и вопросы всех этих студентов.

Когда я писал эти страницы, я имел в виду широкую аудиторию читателей. К нашему классу присоединялись студенты, представляющие широкий спектр специальностей (на нашем жаргоне это называется специализацией). Они представляли два основных направления – нейробиологию и информатику. Однако у нас были также философы, физики, математики, молекулярные биологи, психологи, экономисты, историки, инженеры-электрики и те, кто занимался статистикой. У нас также был студент из отдела общественного здравоохранения и талантливый студент, изучающий кино. Они пришли, возможно, потому, что заблудились в первый день занятий, либо сделали это целенаправленно. Я надеюсь, что ученые из области информатики, математики и физики будут рады узнать о достижениях в нашем постижении зрительного восприятия и поведения. Я также надеюсь, что нейробиологи, биологи и психологи будут рады узнать, как можно научить компьютеры видеть окружающий их мир. Цель этой книги не в том, чтобы дать исчерпывающее обсуждение всех возможных свойств зрения – количество научных публикаций по зрению огромно. За последние два десятилетия было опубликовано больше исследований зрения, чем их общее число за предыдущие два тысячелетия. Цель этой книги – разжечь аппетит студентов к исследованиям в области зрения.

Боюсь, что не смогу отдать должное всем увлекательным работам в области зрения. Каждая из затронутых здесь тем заслуживает отдельной книги. Действительно, есть целые книги, посвященные темам из гл. 2, например

«Сетчатка» Джона Доулинга. Некоторые книги обсуждают темы зрительно-моторной координации, изложенной в гл. 3, например книга Дейла Первса «Почему мы видим то, что делаем». Всего лишь одно из нарушений зрения, представленных в гл. 4, было расширено до целой книги («Визуальная агнозия» Марты Фарах). Томазо Поджио ведет отличный курс, охватывающий темы гл. 8. Я мог бы продолжить перечисление многих замечательных специализированных книг, которые расширяют материал других глав.

Вместо отдельного обсуждения различных подходов к изучению зрения моя цель в этой книге состоит в том, чтобы навести мосты между ними, соединить последние достигнутые результаты в исследованиях нейрофизиологии зрения с последними разработками в области компьютерных визуальных технологий.

В последнее десятилетие произошла революция в области изучения зрения. Теперь у нас есть инструменты, позволяющие исследовать мозг с беспрецедентным разрешением. Мы можем приступить к построению подробных коннектомов¹, с высоким разрешением описывающих, кто с кем общается в нейронном сообществе. Мы можем одновременно слушать активность сотен или даже тысяч отдельных нейронов. Мы можем включать и выключать определенные нейронные схемы обратимым образом с помощью впечатляющих новых методов, представленных Эдом Бойденом и Карлом Дайсеротом. В то же время впервые за короткую, но захватывающую историю развития компьютерного зрения у нас есть алгоритмы, которые достаточно хорошо работают при решении различных задач распознавания образов. Вычислительные модели зрения быстро становятся стандартным инструментом для экспериментаторов. Более того, нейробиологические открытия воспаляют воображение компьютерных экспертов и подталкивают их к созданию все более сложных моделей, дающих разумное первичное приближение к зрительно-моторной координации и нейрофизиологии зрения.

Вследствие огромного количества литературы и моей скромной дидактической цели по установлению связей между различными областями науки о зрении многие важные и захватывающие темы в этой книге пришлось опустить. Заранее прошу прощения за некоторые серьезные тематические пробелы. Одна огромная тема, которая здесь по большей части игнорируется, – это зрение различных представителей животного мира. Я сосредоточился на работе со зрением грызунов, кошек и особенно обезьян и людей в ущерб интереснейшей теме зрения многих других видов, таких как мухи, саламандры и саранча. Разнообразие зрительных систем вызывает растерянность и завораживает, и я лишь поверхностно коснулся этого великолепия в гл. 3. Я убежден, что мы должны исследовать зрение самых разных видов, чтобы понять, как работают их нейронные цепи обработки изображений, и у нас есть чему поучиться у иных видов.

Другая тема, занимающая значительную часть литературы, которая осталась без внимания, связана с неинвазивными исследованиями человеческого мозга: поле работ в этой области настолько обширно, что об этих исследо-

¹ Коннектом – полное описание структуры связей в нервной системе того или иного организма. – *Прим. ред.*

ваниях наверняка написано множество книг. Однако методы исследований развиваются так быстро, что я подозреваю, что мои записи быстро устареют, и новые, более точные исследования заменят старые в самом ближайшем будущем. Другой аспект человеческого зрения, упущенный здесь, – это клиническая практика в офтальмологии. Когда я говорю людям, что изучаю зрение, они сразу же думают о глазах, корректирующих хрусталиках, катаракте и глаукоме. Еще одна область человеческого зрения связана с эстетикой и искусством. Моя замечательная коллега, Маргарет Ливингстон, написала восхитительную книгу, в которой все, что мы знаем о зрительной коре головного мозга (обсуждается в гл. 5 и 6), связывается с восприятием и интерпретацией произведений искусства («Искусство и восприятие. Биология зрения»). Я хотел бы призвать читателей отважиться на изучение всех этих других свойств зрения. Описанное здесь обеспечивает основу для более глубокого изучения и более специализированной работы в этой области.

Я сознательно попытался установить как можно больше связей между различными главами. Мне особенно нравится соединять биологические и вычислительные схемы. Биологическое зрение – продукт миллионов лет эволюции. Нет причин изобретать велосипед при разработке компьютерных вычислительных моделей. Мы можем узнать, как биология решает задачи зрения, и использовать эти решения в качестве вдохновения для создания лучших алгоритмов. Обратное также верно. Гениальные разработки в области компьютерного зрения могут помочь нам понять, что искать в нейронных цепях и как моделировать сложные взаимодействия между нейронами.

Последовательность глав следует траектории, по которой я провел студентов в своем классе. Это повествование начинается с обсуждения основных ограничений зрения и определения типов задач, которые зрительная система должна решать (гл. 1–3), потом переходит к тому, как нейробиология решает эти задачи (гл. 2–6), и, наконец, собирает все это вместе с помощью вычислительных моделей (гл. 7–9). Хотя я считаю такой порядок дидактически эффективным, читатели и учителя могут предпочесть создавать свои собственные учебные программы и следовать альтернативными маршрутами.

Я завершаю книгу главой, в которой делается попытка соединить биологию, информатику и философию, обсуждая особенно таинственное свойство зрения: опыт сознательного восприятия. Страсть к этой проблеме я унаследовал от научного руководителя моей докторской диссертации Кристофа Коха. Читателям, интересующимся сознанием, предлагается изучить многочисленные трактаты профессора Коха по этой теме, в частности «В поисках сознания» и «Исповедь романтического редукциониста».

Еще одним следствием обилия литературы в этой области является то, что страницы книги можно легко заполнить длинным списком ссылок. Я долго решал, включать ли цитаты для каждого утверждения или нет. Как ученый я привык оправдывать каждое утверждение, приводя данные или цитируя соответствующие источники. Как учитель я был обеспокоен тем, что ссылки нарушат поток текста и отпугнут студентов. Хотя я не изучал этот феномен тщательно, у меня сложилось впечатление, что количество цитируемых ра-

бот, к которым читатели обращаются, обратно пропорционально количеству ссылок на них. Поэтому я сократил список ссылок строго до пяти на главу. Эти пять ссылок никоим образом не означают исчерпывающее описание научных работ в данной области. Я попытался совместить классические труды с недавними, интересными и актуальными работами. В веб-материалах, сопровождающих эту книгу, я предоставляю более широкий список ссылок для каждой главы (<https://klab.tch.harvard.edu/publications/Books/BiologicalAndComputerVision/TableOfContents.html>). Однако этот более широкий список никоим образом не претендует на полноту. Эти ссылки – всего лишь приглашение углубиться в темы, представленные в соответствующей главе.

Веб-материал также включает ссылки на соответствующие видеоматериалы по каждой главе. Во многих случаях видео может передавать материал способами, которые невозможны на печатной странице. Более того, молодое поколение особенно любит учиться по видео. Если читатели найдут какой-либо учебный материал, имеющий отношение к содержанию этой книги, или у них появятся другие комментарии о книге, я был бы признателен за обратную связь, чтобы, возможно, добавить этот материал к списку. Я не могу гарантировать, что добавлю их все, отчасти из-за предполагаемой обратной зависимости между количеством ссылок и тем, переходят ли люди по ним или нет. Тем не менее я могу гарантировать, что отнесусь к вашим предложениям очень серьезно. Таким образом, эти страницы открывают двери для обсуждения проблем зрения, приглашая читателей начать диалог с автором.

В этом диалоге я не уклоняюсь от обсуждения спорных или спекулятивных идей в темах, которые далеки от согласования. Наука – это не просто собрание теорий и фактов. Чтобы понять природу света в физике, потребовалось много шагов между теориями и экспериментами, когда эксперименты правили теориями и новые теории ставили новые экспериментальные задачи. Точно так же исследования в области зрения, несомненно, полны темами, которые остаются предметом интенсивных дискуссий. Я считаю полезным познакомить студентов и читателей с тем, как проводились эксперименты, каковы были ключевые идеи и гипотезы, как ученые могут ошибаться и исправлять свои представления, основываясь на новых эмпирических результатах. Я надеюсь, что включение такой интерпретации результатов может помочь передать динамичный и живой характер дискуссий, в отличие от простого набора фактов.

Открывая диалог с читателями, освещая новые открытия, что нам известно и что неизвестно, я был бы особенно счастлив, если бы эта книга вдохновила новые поколения ученых принять вызов и помочь нам понять, как работает зрение. Я был бы счастлив, если бы эти страницы вдохновили смелых молодых ученых доказать, что некоторые из представленных здесь предположений ошибочны. Я надеюсь, что других эти страницы побудят решить многие еще открытые вопросы в этой области.

Возможность объединения исследований нейронных цепей со стороны нейробиологии и разработок вычислительных моделей, на них основанных, все еще находится в зачаточном состоянии. Несмотря на недавние заявления в известных газетах, проблема зрения до сих пор не решена. В последнее вре-

мя в компьютерном зрении наблюдается впечатляющий прогресс, и теперь у нас есть алгоритмы, которые могут решать определенные задачи зрительного распознавания на том же уровне, что и люди, а в некоторых случаях даже превосходят человеческие возможности. Однако мы все еще довольно далеки от окончательного решения проблемы зрения. В законах физики нет никаких препятствий, нет фундаментальных ограничений, согласно которым мы не можем создавать алгоритмы и машины, которые будут видеть и интерпретировать визуальный мир все лучше и лучше. Я убежден, что мы *решим* эту проблему. Лучшее – впереди.

Благодарности

В некоторых фильмах ученый изображен как одинокое, часто злое седовласое существо, заключившее себя в тесную келью, в ожидании, когда яблоко озарения упадет ему на голову и до него внезапно дойдет, как все это работает. Моя жизнь как ученого имеет мало общего с таким голливудским стилем проведения научных исследований. Наука особенно привлекательна для меня из-за ее объективных и тщательных усилий по открытию Истины с большой буквы. Наука не возникает в вакууме, а, напротив, является частью социальной конструкции, и я надеюсь, что яблоко озарения еще упадет на мою голову. Я невероятно признателен множеству людей, которые так или иначе помогли собрать эти страницы воедино. Эта книга была бы невозможна без идей и помощи очень большого числа людей.

Во-первых, я хотел бы поблагодарить всех учеников курса биологического и компьютерного зрения. Их вечное любопытство, их кажущиеся наивными, но пронизательные способы заставить меня переосмыслить то, что я пытался объяснить, и их постоянное сомнение в наших предположениях и утверждениях держали меня в напряжении и продвигали эту работу вперед. Я не могу отдать должное всем из них, но я хотел бы упомянуть нескольких выделяющихся студентов, в том числе Кристен Фанг, Григория Гитчаунц, Кэтрин Харрисон, Хавьера Масиса, Оливию МакГиннис, Дэниела Ротшильда, Шона Салливана, Кеннета Шинозуку, Дункана Стотерса, Аннабель Тао, Салони Вишвакарм и Уилл Сяо.

Помощники учителя в классе – Джозеф Олсон, Уилл Сяо и Ючэнь Сяо – также сыграли важную роль в этой работе. Мы разбиваем каждый семинар на лекцию и обсуждение основной литературы по теме. Ассистенты учителя всегда задают провокационные вопросы, критикуют литературу и вовлекают студентов в увлекательные дискуссии о том, почему и как проводилось конкретное исследование. Некоторые коллеги также читали гостевые лекции на семинарах, и их способ представления материала тоже оказал огромное влияние. Я хотел бы поблагодарить Фредерико Азеведо, Андрея Барбу, Ксавье Буа, Дэвида Кокса, Камиллу Гомес-Лаберже, Тилля Хартманна, Лейлу Исик, Кохитию Кару, Джие Ким, Билла Лоттера, Диего Мендозза-Халлидея, Кена Накаямю, Карлоса Понсе, Сарит Спиро, Ханьлинь Тан, Каспера Винкеуна и Уилла Сяо.

Еще одним постоянным источником вдохновения были студенты и постдоки моей лаборатории. Особо хочу упомянуть Игала Агама, Арджуну Бансала, Ксавье Бойа, Калин Буйя, Кендру Бербанк, Стивена Каспера, Прабаха Гангпадхья, Камиллу Гомес-Лаберже, Стефана Гжековски, Шаши Гупту, Уолтера Хардести, Элеонору Яселли, Лейлу Исик, Джие Ким, Филя Кунке, Гарретта Лама, Даниэла Лопес Мартинеса, Билла Лоттера, Радхику Мадхавана, Томаса Микони, Пранава Мишру, Шарлотту Моэрманн, Кэндис Росс, Мартинаа Шримпф, Нимрода Шахама, Джедедию Сингер, Ханлин Тан, Бена Цуу, Каспера Винкена, Джерри Вана, Фараназа Вика, Эрика и Кевина Ву, Уилла

Сяо, Эге Юмусака и Мэнми Чжан. Многие из них оставили здесь полезные комментарии и критику. Они также внесли выдающийся научный вклад, который отражен в обсуждаемой здесь работе.

В наши дни, когда требования становятся все более строгими, я вижу, как мои коллеги постоянно торопятся уложиться в сроки и жонглируют своим вкладом в научное исследование. Поэтому для меня большая честь вспомнить о том, что я в долгу перед многими из моих друзей и коллег, которые были достаточно любезны, чтобы прочитать эти страницы, исправить мои ошибки, предложить дополнительные материалы и дать критические комментарии. Следующие были особенно любезны: Араш Афраз, Катарина Бендц, Ксавье Бойс, Альфонсо Карамазза, Стивен Каспер, Венцян Чен, Майкл До, Джон Даулинг, Кохитидж Кар, Талия Конкл, Билл Лоттер, Ричард Масланд, Дэвид Мазумдер, Лиад Мудрик, Карлос Понсе, Нао Цучия, Уилл Сяо, Яода Сюй и Мэнми Чжан. Больше всего я признателен Эллен Хилдрет. Она внесла большой вклад в исследования компьютерного зрения и возглавляет образовательные инициативы в нашем Центре мозга, разума и машин. Она прочитала всю книгу и помогла исправить ее своими подробными замечаниями, пронизательными вопросами и продуктивными комментариями.

Наконец, я бесконечно благодарен Центру мозга, разума и машин (СВММ), который спонсируется Национальным научным фондом. СВММ стал вдохновляющим и влиятельным домом для мозгового штурма новых идей и синергетического расширения границ нейробиологии и исследований искусственного интеллекта, объединив большое количество талантливых ученых из самых разных областей.

Спасибо вам всем!

Сокращения

В тексте используются сокращения. Вот некоторые из них.

AI – искусственный интеллект (гл. 1 и 9).

Alexnet – известная глубокая сверточная нейронная сеть для распознавания образов (Крижевский и др., 2012); другие архитектуры глубоких сверточных нейронных сетей, такие как **VGG**, **ResNet** и **Inception**, также упоминаются в гл. 8 и 9.

CLEVR – композиционный язык и база данных элементарных визуальных ссылок (Джонсон и др., 2016, гл. 9).

DCNN – глубокая сверточная нейронная сеть (гл. 8).

DeepDream – метод визуализации изображений, который приводит к высокой активации блоков в нейронной сети (гл. 9).

GAN – генеративная состязательная сеть (Гудфеллоу и др., 2014, гл. 9).

HMAX – вычислительная модель зрительного распознавания (Ризенхубер и Поггио, 1999, гл. 8).

ImageNet – большая база изображений, используемая для обучения и тестирования алгоритмов компьютерного зрения (Русаковский и др., 2014, гл. 8 и 9).

IOR – механизм запрета возврата, предотвращающий постоянное попадание движений глаз в наиболее заметную часть изображения (проиллюстрировано на рис. 8.9).

ITC – нижняя височная кора (гл. 4 и 6).

IVSN – сеть инвариантного визуального поиска (гл. 8, рис. 8.9).

L2 norm – евклидово расстояние между двумя векторами, обычно используемое для определения ошибки в машинном обучении (в отличие от **L1 norm**, которая представляет собой сумму абсолютных значений каждого компонента вектора расстояния) (гл. 8).

LGN – латеральное колочатое ядро, часть таламуса, которая получает информацию от сетчатки и проецируется на первичную зрительную кору головного мозга (гл. 2).

LSTM – рекуррентная нейронная сеть Long Short-Term Memory (Хорхрайтер и Шмидхубер, 1997, гл. 8).

MNIST – база данных, состоящая из изображений рукописных цифр (гл. 8).

MSCOCO – база данных сегментированных объектов (Лин и др., 2015, гл. 9).

Неокогнитрон – вычислительная модель зрительного распознавания (Фукусима, 1980, гл. 8).

PredNet – глубокая сверточная нейронная сеть с предиктивным кодированием (Лоттер и др., 2017, гл. 8).

QR-код – код быстрого доступа – матричный штрих-код, легко читаемый смартфонами (гл. 9).

ReLU – выпрямительный линейный блок (гл. 7).

ResNet – архитектура глубокой сверточной нейронной сети (Хи и др., 2015, гл. 8).

RGC – ганглиозные клетки сетчатки, выходные нейроны сетчатки (гл. 2).

ROC – кривая рабочих характеристик приемника (Грин и Светс, 1966, гл. 9).

Softmax – функция, которая преобразует вектор входных чисел в распределение вероятностей, которое в сумме дает 1, где каждое выходное значение пропорционально экспоненте входных чисел (гл. 8).

SVM – машина опорных векторов (гл. 6).

tSNE – t-распределенное стохастическое встраивание сети (Ван дер Маартен и Хинтон, 2008, гл. 8).

UCF101 – большая база данных видео, используемая для обучения и тестирования алгоритмов распознавания действий (Соомро и др., 2012, гл. 9).

V1 – первичная зрительная кора; определены другие визуальные области, такие как **V2**, **V3**, **V4** и **V5**, также известные как область **MT** (мозолистое тело. – *Прим. ред.*) (гл. 4 и 5).

WTA – механизм «победитель получает все» (например, тот, что показан на рис. 8.9).

XDream – расширяет DeepDream с эволюцией в реальном времени для максимизации активации, алгоритм для беспристрастной проверки настройки нейронов (Понсе и др., 2019, гл. 8).

Глава 1

Знакомство с миром зрения

Понимание того, как работает мозг, является величайшей научной задачей нашего времени, возможно, самой большой проблемой всех времен. Мы отправили космические корабли заглянуть за пределы нашей Солнечной системы, изучаем далекие галактики, чтобы строить теории о происхождении Вселенной. Мы построили мощные ускорители, чтобы исследовать тайны субатомных частиц. Мы раскрыли секреты наследственности, скрытые в миллиардах пар нуклеотидов ДНК. Но нам еще предстоит выяснить, как работают эти полтора килограмма мозговой ткани внутри наших черепов, давая нам возможность заниматься всем этим – физикой, биологией, музыкой, литературой и политикой. Переговоры и взаимодействия примерно 100 млрд нейронов в нашем мозге отвечают за нашу способность интерпретировать сенсорную информацию, ориентироваться, общаться, чувствовать и любить, принимать решения и строить планы на будущее, учиться. Понимание того, как нейронные цепи обеспечивают когнитивные функции, изменит нашу жизнь: это поможет нам облегчить нарушения психического здоровья, от которых страдают сотни миллионов людей, это приведет к созданию машин с настоящим искусственным интеллектом, которые будут так же умны или даже умнее, чем мы, и это откроет нам двери к окончательному пониманию того, кто мы есть.

Для понимания работы мозга мы сосредоточимся на одном из самых тонких элементов нейронных механизмов, когда-либо существовавших: зрительной системе. За небольшую долю секунды мы можем, лишь мельком увидев изображение, получить значительный объем информации. Например, мы можем взглянуть на рис. 1.1 и ответить на бесконечную серию вопросов о нем: *кто там, что там, где это место, какая стоит погода, сколько там людей и что они делают?* Мы даже можем сделать обоснованные предположения о том, что там происходит, включая описание *отношений между людьми на картинке, того, что произошло раньше или что, возможно, произойдет потом.* В основе ответов на эти вопросы лежит наша способность к распознаванию зрительных образов и интеллектуальным выводам на основе распознанных зрительных образов.



Рис. 1.1 ❖ Мы можем визуально интерпретировать изображение с первого взгляда. Кто изображен? Что здесь находится? Где это находится? Что они делают? Что будет дальше? Это всего лишь несколько из бесконечного числа вопросов, на которые мы можем ответить после нескольких сотен миллисекунд просмотра нового изображения

Наша замечательная способность интерпретировать сложные пространственно-временные последовательности входящей информации, которые мы можем условно приписать отчасти «здоровому смыслу», не требует, чтобы мы садились и решали сложные дифференциальные уравнения. Даже четырехлетний ребенок может достаточно точно ответить на часть вопросов, изложенных выше, дети более старшего возраста могут ответить на все или большинство из них, и многие виды животных, не только человек, также могут быть обучены правильно распознавать и демонстрировать понимание многих характеристик представленной им визуальной сцены. Кроме того, для получения такой важной информации из изображения нам требуется всего несколько сотен миллисекунд. Несмотря на то что у нас есть компьютеры, которые превосходно справляются с такими задачами, как решение сложных дифференциальных уравнений, компьютеры по-прежнему не могут сравниться с человеческими способностями ответить на простые вопросы о содержании изображения.

1.1. ЭВОЛЮЦИЯ ЗРИТЕЛЬНОЙ СИСТЕМЫ

Зрение необходимо для решения большинства повседневных задач, включая ориентацию, чтение и общение. Чтение этого текста включает в себя рас-

познавание образов. Дорога до дома предполагает обнаружение пешеходов, автомобилей и маршрутов передвижения. Зрение имеет решающее значение для узнавания наших друзей и расшифровки их эмоций. Поэтому нетрудно представить, что расширение зрительной коры сыграло значительную роль в эволюции млекопитающих в целом и приматов в частности. Вероятно, что эволюция алгоритмов распознавания образов на основе данных визуального канала информации привела к улучшению возможности приспособления и выживания за счет улучшенной ориентации на местности, различения друга и врага, дифференциации еды и яда, а также умения адекватно истолковывать сигналы социальных взаимодействий. В отличие от тактильных и вкусовых сигналов и, в некоторой степени, даже слуховых сигналов, визуальные сигналы приносят знания из обширных и далеких в пространстве областей. Хотя обонятельные сигналы также могут распространяться на большие расстояния, скорость распространения и содержание химических сигналов куда меньше, чем скорость света и та информация, которую он может нести.

Способность биологических организмов улавливать свет – древняя. Например, многие бактерии используют свет для фотосинтеза, предшественника аналогичного процесса, улавливающего энергию в зеленых растениях. Что особенно поражает в зрении, так это возможность использования света для сбора информации о мире. Эволюционное преимущество, обеспечиваемое обработкой визуальной информации, настолько очевидно, что это привело зоолога Эндрю Паркера к предложению так называемой «Теории включения света», объясняющей быстрое увеличение числа и разнообразия видов на Земле.

Около 500 млн лет назад, в ранний кембрийский период, произошел необычайный всплеск количества различных видов живых существ¹. Примерно в то же время окаменелости свидетельствуют о появлении первого вида, имеющего глаза, – трилобитов (рис. 1.2). Трилобиты – это вымершие членистоногие (дальние родственники насекомых и пауков), которые завоевали мир и были широко распространены на протяжении примерно 300 млн лет. Теория включения света утверждает, что появление глаз и взрыв разнообразия видов животных – не простое совпадение. Некоторые исследователи утверждают, что глаза появились прямо перед кембрийским взрывом. Глаза позволили одному удачливому раннему трилобиту или его предку собирать информацию издалека, обнаруживая присутствие добычи или хищника, таким образом наделяя его особым преимуществом перед другими существами без глаз, которым для выживания приходилось полагаться на более медленные и грубые каналы информации. С появлением этих новых органов, глаз, началась эволюционная гонка вооружений между добычей и хищниками, чтобы собирать информацию об окружающем мире и прятаться от опасности либо выслеживать добычу с помощью этих новых, пристальных и мощных датчиков. Внезапно формы тела, текстуры и цвета стали яркими, мощными и опасными. Кажется вероятным, что формы и цвета тела начали меняться, чтобы избежать обнаружения с помощью начальных версий камуфляжа, что, в свою очередь, привело к тому, что зрение стало более острым и лучшим

¹ Так называемый «кембрийский взрыв». – Прим. ред.

в своей способности обнаружения, стало более чувствительным к движению и к тонким изменениям благодаря способности лучше различать формы. Да будет свет! И пусть свет будет использоваться для передачи информации.



Рис. 1.2 ❖ Отпечаток трилобита в окаменелости, возрастом около 500 млн лет. У трилобитов, подобных изображенному на этой картинке, были сложные глаза, скорее всего, не слишком отличающиеся от тех, что встречаются у современных видов беспозвоночных, таких как мухи. Трилобиты распространились по всему миру и были одним из самых заметных видов около 300 млн лет. Автор фото – Dwergenpaartje, CC BY-SA 3.0

1.2. Будущее зрения

Перенесемся на несколько сотен миллионов лет вперед. Трудно недооценить фундаментальную роль зрения в эволюции человека. Задолго до появления языка, в том виде, в котором он известен сегодня, зрение играло решающую роль в общении, интерпретации эмоций и намерений, в облегчении социальных взаимодействий. Способность с помощью зрения определять закономерности в положении луны, солнца и звезд привела к предсказанию сезонных изменений, которые в конечном итоге привели к развитию сельского хозяйства, трансформировав кочевые общества в оседлые, породив предшественники городов будущего. Искусство, символы и, в конечном итоге, развитие письменной речи также в значительной степени основывались на возможностях зрительного распознавания образов.

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru