

УДК 004.8
ББК 32.81
К62

Крейман Г.

К62 Биологическое и компьютерное зрение / пер. с англ. И. Л. Люско; под ред. Т. Б. Киселевой, Т. И. Люско. – М.: ДМК Пресс, 2022. – 314 с.: ил.

ISBN 978-5-93700-100-9

Книга посвящена разработкам искусственного интеллекта в части компьютерного зрения и объединяет подходы, позволяющие рассмотреть эту тему на пересечении биологии и математического моделирования.

В первой части книги рассматриваются биологические основы зрения: читатель знакомится с различными аспектами зрительного восприятия и узнает, какие структуры нейронов его обеспечивают, какие процессы происходят в мозге при распознавании образов. Во второй части представлены различные структуры математических моделей – нейронных сетей, и принципы их работы. Затрагиваются, в частности, глубокие сверточные нейронные сети, машинное обучение и генеративные состязательные сети. Последняя глава посвящена исследованию самого феномена сознания.

Издание рассчитано на широкую аудиторию, интересующуюся искусственным интеллектом и компьютерным зрением, в частности тем, как машины учат видеть и интерпретировать визуальный мир.

УДК 004.8
ББК 32.81

This translation of *Biological and Computer Vision* is published by arrangement with Cambridge University Press. Copyright © 2021 by DMK Press Publishing under license by No Starch Press Inc. All rights reserved.

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

ISBN 978-1-108-48343-8 (англ.)
ISBN 978-5-93700-100-9 (рус.)

© Gabriel Kreiman, 2021
© Перевод, оформление, издание,
ДМК Пресс, 2022

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

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.

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

Дополнительный контент на <http://bit.ly/2TqTDt5>.

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.

Эволюция зрительной системы плохо изучена и остается интересной темой для дальнейших исследований. Будущее ее будет не менее увлекательным. Хотя размышлять о биологических изменениях зрения у животных в эволюционных временных масштабах довольно сложно, легче представить, что можно было бы достичь в ближайшем будущем в более короткое время с помощью машин с подходящими камерами и вычислительными алгоритмами для обработки изображений. Мы вернемся к будущему зрения в гл. 9; в качестве тизера давайте кратко рассмотрим машины, которые могут достигать и, возможно, превосходить возможности человека в решении визуальных задач. Такие машины могут сочетать в себе высокоскоростные видеодатчики с высоким разрешением, которые передают информацию компьютерам, в реальном времени реализующим сложные симуляции реальности¹, подобные зрительной функции мозга.

Машины могут скоро преуспеть в решении задач распознавания лиц до уровня, когда банкомат будет приветствовать вас по вашему имени без необходимости ввода пароля, когда вам может уже не понадобится ключ для входа в ваш дом или чтобы сесть в автомобиль, где ваше лицо может стать вашей кредитной картой и вашим паспортом. Беспилотные автомобили, приводимые в движение алгоритмами машинного зрения, покинули страницы научной фантастики и вышли на наши улицы. Компьютеры также могут овладеть способностью анализировать изображения для поиска объектов в интернете по фото или видео (в отличие от ключевых слов и текстовых дескрипторов). Врачи могут все больше и больше полагаться на визуальные системы анализа рентгеновских снимков, МРТ и других изображений до такой степени, что диагностика на основе анализа изображений становится одной из областей информатики. Следующие поколения будут удивляться тому, что когда-то мы позволяли людям принимать ошибочные диагностические решения. Классификацией далеких галактик или открытием различных видов растений и животных могли бы руководить алгоритмы распознавания, а не астрономы или биологи.

Пройдя дальше в область научной фантастики, можно придумать интерфейсы мозг–машина, которые можно было бы имплантировать в человеческий мозг для того, чтобы дать возможность в той или иной степени видеть людям с нарушением зрения или слепым. Пока мы занимаемся этим, почему бы также не использовать такие интерфейсы для качественного улучшения зрения у людей с нормальным зрением, чтобы дать людям возможность видеть на 360°, наблюдать инфракрасный или ультрафиолетовый диапазон спектра, видеть сквозь непрозрачные объекты, такие как стены, или непосредственно наблюдать удаленные события?

Когда возникли дебаты о возможности того, смогут ли компьютеры когда-нибудь играть в шахматы против людей, большинство людей отнеслись к такой возможности скептически. Но сегодня даже простые компьютеры могут побеждать искушенных любителей шахмат, а продвинутые компьютеры могут побеждать даже мировых чемпионов. В последнее время компьютеры также преуспели в древней и сложной игре го. Несмотря на очевидный факт,

¹ Воссоздание реальности на основе имеющейся информации. – *Прим. ред.*

что большинство людей могут распознавать предметы намного лучше, чем они могут играть в шахматы или го, визуальное распознавание с вычислительной точки зрения является более сложной задачей, чем эти игры. Однако, возможно, мы не слишком далеки от построения точных вычислительных приближений к визуальным системам, в которых мы сможем доверять глазам компьютеров так же или даже больше, чем своим собственным. Вместо «верь глазам своим» девизом будущего может стать «верь глазам компьютера своего».

1.3. ПОЧЕМУ ЗРЕНИЕ – НЕ ТАКАЯ ПРОСТАЯ ВЕЩЬ?

Представление о том, что зрение является более сложным с точки зрения вычислений, чем игра в го, может показаться нелогичным. В конце концов, двухлетний ребенок может открыть глаза и быстро распознать и интерпретировать свое окружение, чтобы ориентироваться в комнате и схватить своего любимого плюшевого мишку, который может быть наполовину прикрыт другими игрушками. Он не умеет играть в го. Он определенно не прошел через миллионы часов обучения с подкреплением, которые должны были пройти машины нейронных сетей, чтобы научиться играть в го. У него было около десяти тысяч часов визуального опыта. Эти десять тысяч часов в основном проходят без присмотра; большую часть времени поблизости были взрослые, но, по большому счету, эти взрослые не предоставляли непрерывной информации о названиях объектов или непрерывных сигналов вознаграждения и наказания (конечно, были и награды, но они, скорее всего, составляли небольшую часть его визуального обучения).

Почему компьютерам так сложно выполнять задачи распознавания образов, которые нам кажутся такими простыми? Зрительная система приматов превосходно распознает образы, даже когда эти картинки радикально меняются от одного экземпляра к другому. Рассмотрим схему простых линий на рис. 1.3. Эти рукописные символы легко распознать, даже если на уровне пикселей все они значительно различаются. Каждый из них состоит всего лишь из считанных штрихов. С реальными сценами и объектами задача намного сложнее. Представьте себе множество возможных вариантов изображений, сделанных на площади Сан-Марко в Венеции (рис. 1.1), и то, как визуальная система может их интерпретировать, причем без особого труда. Любой объект может давать *бесконечное* количество проекций на сетчатку глаза. Эти вариации включают изменения масштаба, положения, точки обзора и освещения, а также другие вариации. Казалось бы, без особых усилий наши визуальные системы могут связывать все эти изображения с конкретным объектом.

Идентификация конкретных объектов – лишь одна из важных функций, которые должна решать визуальная система. Она может оценивать расстояния до объектов, прогнозировать, куда они движутся, определять идентичность объектов, которые частично или сильно закрыты от взгляда или камуфлированы, определять, какие объекты ближе, а какие дальше, и делать

обоснованные предположения относительно состава и веса объектов. Визуальная система может даже делать прогноз о намерениях объектов в случае живых существ. Во всех этих ситуациях зрение ставит нам некорректно поставленную задачу в том смысле, что здесь возможно наличие более одного решения, согласующегося с заданной схемой изображения.

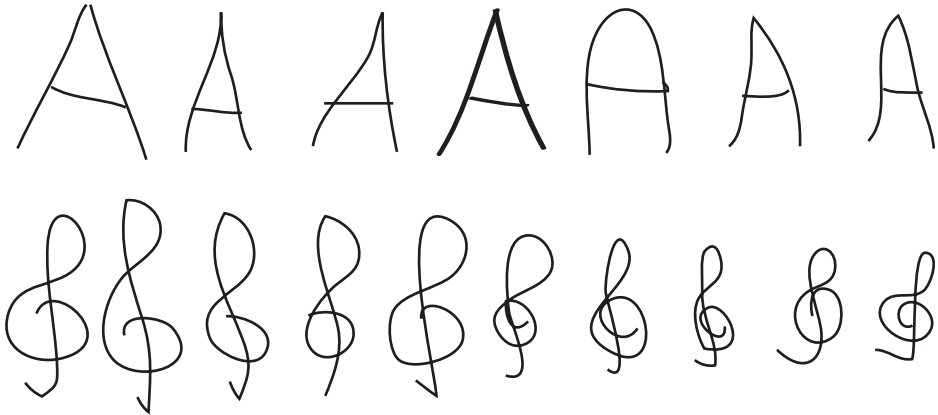


Рис. 1.3 ❖ Любой объект может быть представлен в самых разных образах. Несмотря на то что мы можем легко распознать эти образы, существует большая разница между различными визуализациями каждого образа на уровне пикселей

1.4. ЧЕТЫРЕ ВАЖНЕЙШИЕ ОСОБЕННОСТИ ЗРИТЕЛЬНОГО ВОСПРИЯТИЯ

Чтобы объяснить, как зрительная система справляется с идентификацией образов, нам необходимо учесть четыре важнейшие особенности зрительного распознавания: *избирательность*, *толерантность*, *скорость* и *вместимость*.

Избирательность восприятия включает в себя способность различать образы, очень похожие на уровне пикселей. Примеры исключительной избирательности визуальной системы – распознавание лиц и чтение. В обоих случаях визуальная система может различать очень близкие входные данные, если мы сравниваем их непосредственно на уровне пикселей. Тривиальный и бесполезный способ реализации избирательности в вычислительном алгоритме – запомнить все пиксели изображения (рис. 1.4А). При обнаружении точно тех же пикселей компьютер сможет «распознать» изображение. Компьютер будет чрезвычайно избирательным, потому что он не будет реагировать на любое другое похожее изображение. Проблема этой реализации в том, что ей не хватает *толерантности*.

Под *толерантностью* восприятия понимается способность распознавать объект, несмотря на множественные преобразования его изображения. На-

пример, мы можем распознавать объекты, даже если они представлены в другом положении, масштабе, с другого ракурса, с другим контрастом, освещением или в другом цвете. Мы даже можем распознавать объекты, в которых изображение претерпевает нежесткие преобразования, например изменения, которые претерпевает лицо при улыбке. Тривиальный, но бесполезный способ реализации толерантности – построить модель, которая будет давать плоскую реакцию – всегда говорить «да», независимую от входных данных. Она будет демонстрировать толерантность к преобразованиям изображения, но не будет демонстрировать избирательности к разным формам (рис. 1.4В). Сочетание *избирательности* и *толерантности* (рис. 1.4С), возможно, является ключевой проблемой при разработке алгоритмов компьютерного зрения для задач распознавания. Рассмотрим реальный пример: беспилотный автомобиль должен выборочно отличать пешеходов от многих других типов объектов, независимо от того, насколько высоки эти пешеходы, во что они одеты, что они делают или что держат в руках¹.

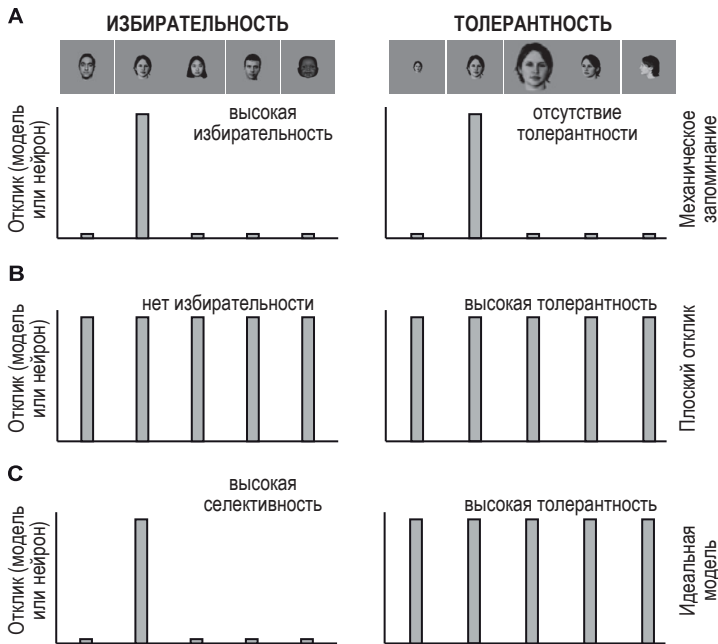


Рис. 1.4 ❖ Тривиальный (и бесполезный) подход к моделированию распознавания образов. Две простые модели, которые легко реализовать, но пользы от них немного. Модель механического запоминания (А) имеет исключительную избирательность, но невысокую толерантность. Напротив, модель плоской реакции (В) имеет хорошую толерантность, но плохую избирательность. (С) Идеальная модель должна сочетать избирательность и толерантность

¹ Известен случай, когда алгоритм распознавания пешеходов на изображении с камеры автомобиля опознал как человека собаку (видимо, он решил, что это человек, передвигающийся на четвереньках). – Прим. ред.

Учитывая комбинаторный взрыв количества изображений, воспринимаемых как один и тот же «объект», можно представить, что для зрительного распознавания требуется много лет обучения в школе. Конечно, это далеко не так. Задолго до того, как первоклассник начинает изучать основы сложения и вычитания (довольно тривиальные задачи для компьютеров), он уже достаточно хорошо владеет визуальным распознаванием образов – задачей, которую он может выполнить с первого взгляда. Объекты можно легко распознать в потоке других объектов, показываемых всего лишь в течение 100 мс каждый. Наблюдатели могут сделать движение глаз, чтобы указать на присутствие объекта в двухальтернативном задании с принудительным выбором, примерно через 200 мс после демонстрации визуального стимула. К тому же данные ЭЭГ с кожи головы, как и инвазивные регистрации потенциалов человеческого мозга, показывают, что распознавание может происходить уже через ~150 мс после появления стимула. *Скорость* зрительного распознавания ограничивает количество вычислительных шагов, которые любая теория распознавания может использовать для учета эффективности распознавания. Безусловно, зрение не останавливает свою работу через 150 мс. Многие особенности зрительного восприятия проявляются в течение сотен миллисекунд, а эффективность распознавания при выполнении сложных задач улучшается с увеличением времени демонстрации сигнала. Однако базовое понимание изображения или основных объектов в изображении может быть достигнуто всего за ~150 мс. Мы обозначаем этот режим как *быстрое визуальное распознавание*.

Одним из способов достижения прогресса в объединении избирательности, толерантности и скорости было сосредоточение внимания на алгоритмах, зависящих от объекта или категории объектов. Примером такого подхода может быть разработка алгоритмов для обнаружения автомобилей в естественных сценах с использованием особенностей автомобилей и сцен, в которых они обычно появляются. Другой пример – распознавание лиц. Некоторые из этих эвристических алгоритмов, зависящих от категорий и содержания, полезны, и мозг может научиться ими пользоваться. Например, если большая часть изображений голубого цвета, предполагается, что на фоне изображения может быть небо, тогда априорные вероятности увидеть машину будут низкими (машины обычно не летают), а априорные вероятности увидеть птицу будут высокими (птиц часто можно увидеть на фоне голубого неба). Мы обсудим закономерности визуального мира и статистику естественных изображений в гл. 2. Несмотря на эти корреляции, в более общем сценарии механизм зрительного распознавания должен сочетать избирательность, толерантность и скорость для огромного диапазона объектов и изображений. Например, в китайском языке более трех тысяч иероглифов. Оценки *вместимости* системы зрительного распознавания человека существенно различаются в разных исследованиях. В некоторых исследованиях приводятся цифры, которые значительно превышают десять тысяч наименований.

Таким образом, теория зрительного распознавания должна учитывать высокую избирательность, толерантность, скорость и вместимость зрительной системы. Несмотря на очевидную непосредственность зрения, объединение этих четырех ключевых характеристик является совсем не простой задачей.

1.5. ПУТЕШЕСТВИЯ И ПРИКЛЮЧЕНИЯ ФОТОНА

Решение некорректно поставленной задачи выбора среди бесконечного числа возможных интерпретаций сцены, тем более допускающее трансформацию объекта способом распознавания в течение всего лишь 150 мс, кажется обескураживающим. Как мозг совершает этот подвиг? Начнем с общего обзора обработки зрительного сигнала в мозге.

Свет попадает на сетчатку после того, как отражается объектами окружающей среды. Паттерны света, падающего на наши глаза, далеко не случайны, и статистика естественного изображения этих узоров играет важную роль в развитии и эволюции зрительной системы (гл. 2). В сетчатке глаза (retina) свет преобразуется в электрический сигнал специализированными клетками фоторецепторов. Информация обрабатывается сетчаткой посредством каскада вычислений, прежде чем она переходит в структуру, называемую таламусом, а оттуда – в кору головного мозга. Кора головного мозга управляет последовательностью шагов визуальных вычислений, преобразовывая изначальный сигнал, который несут фотоны, в зрительные образы. Некоторые модели зрительного распознавания рассматривают сетчатку как аналог попиксельного представления в матрице цифровой камеры. Цифровая камера – это слишком упрощенное описание вычислительной мощности сетчатки глаза, но оно проникло в общий жаргон, введенный производителями, которые хвастаются мониторами «Retina display».

В наши дни коммерчески доступные мониторы нередко отображают несколько миллионов пикселей. Имеющиеся в продаже цифровые камеры также могут похвастаться даже большим – десятками миллионов пикселей. Количество пикселей в таких устройствах приблизительно или даже превышает количество первичных сенсоров в некоторых биологических сетчатках; например, сетчатка человека содержит ~6,4 млн «колбочек» и ~110 млн «палочек» (мы обсудим эти рецепторы в гл. 2). Несмотря на эти технологические достижения, электронные камеры все еще отстают от биологических глаз по таким важным свойствам, как адаптация к яркости, обнаружение движения, фокусировка, энергоэффективность и скорость.

Результирующий сигнал от сетчатки передается во множество областей мозга, включая верхний бугорок, супрахиазматическое ядро¹ и таламус. Поверхностные слои верхнего бугорка можно рассматривать как древний зрительный мозг. Действительно, для многих видов, у которых нет коры, верхний холмик (называемый у этих видов зрительным каналом) является местом, где происходит основная обработка зрительного сигнала. Супрахиазматическое ядро играет центральную роль в регулировании циркадного ритма. У людей есть внутренние дневные часы, которые работают чуть быстрее, чем обычные 24-часовые часы, отсчитывающие сутки², и воспринимаемые сетчаткой световые потоки в конце концов через супрахиазматическое ядро помогают модулировать и регулировать этот цикл.

¹ Ядро передней части гипоталамуса. – *Прим. ред.*

² Эндогенный циркадный ритм человека составляет в среднем 24 часа 11 минут. – *Прим. ред.*

Основной зрительный канал передает информацию от сетчатки к части таламуса, называемой латеральным колленчатым ядром (LGN). LGN проецируется на первичную зрительную кору (V1), расположенную в задней части нашего мозга. Без первичной зрительной коры люди в основном слепы, что подчеркивает критическую важность пути передачи информации в кору для большинства функций зрения. Исследователи называют этапы обработки в сетчатке LGN и V1 «ранним зрением» (гл. 5). Обработка в первичной зрительной коре – это только первая стадия обработки зрительной информации в коре. Исследователи обнаружили десятки областей, ответственных за различные особенности зрения (их фактическое количество все еще остается предметом споров и зависит от того, что подразумевается под «областью»). Важным способом изображения этих множественных областей и их взаимосвязей является диаграмма, предложенная Феллеманом и Ван Эссенем и представленная на рис. 1.5. Неподготовленному глазу эта диаграмма своей сложностью может показаться ошеломляющей, мало чем отличающейся от принципиальных схем, обычно используемых в электронике. Мы углубимся в эту диаграмму более подробно в главах 5 и 6 и обсудим области и связи, которые играют решающую роль в зрительном восприятии.

Несмотря на кажущуюся сложность нейронной схемы на рис. 1.5, эта схема – чрезмерное упрощение реальной схемы подключения нейронов. В первых, каждый прямоугольник на этой диаграмме содержит миллионы нейронов. Есть много разных типов нейронов. Расположение нейронов внутри каждого прямоугольника можно описать с помощью шести основных слоев коры (некоторые из которых, в свою очередь, состоят из нескольких слоев) и топографического расположения нейронов внутри и между слоями. Во вторых, мы еще далеки от описания *всех* связей в зрительной системе. Одним из захватывающих достижений последнего десятилетия является разработка методов тщательного исследования подключения нейронных схем с высоким пространственным и временным разрешением.

Для маленького животного, такого как червь длиной 1 миллиметр с причудливым названием *Caenorhabditis elegans*, мы уже несколько десятилетий знаем подробную схему связи каждого из его 302 нейронов благодаря работе Сидни Бреннера (1927–2019). Однако кора головного мозга – это совершенно другое дело, с плотностью нейронов в десятки тысяч нейронов на квадратный миллиметр. Героические усилия ученых в растущей области «коннектомики» в настоящее время дают первое представление о том, как одни нейроны соединены с другими нейронами коры головного мозга. Основные сюрпризы в нейроанатомии, скорее всего, будут связаны с использованием новых инструментов, использующих преимущества высокой специфичности молекулярной биологии.

Наконец, даже если бы мы знали все связи каждого отдельного нейрона в зрительной коре, это знание не сразу раскрыло бы нам его вычислительные функции (но знание о связях все равно было бы чрезвычайно полезно). В отличие от электрических цепей, в которых мы понимаем каждый элемент, а общую функцию можно оценить путем внимательного изучения электрической схемы, многие нейробиологические факторы делают соотношение структуры и функции нетривиальным.

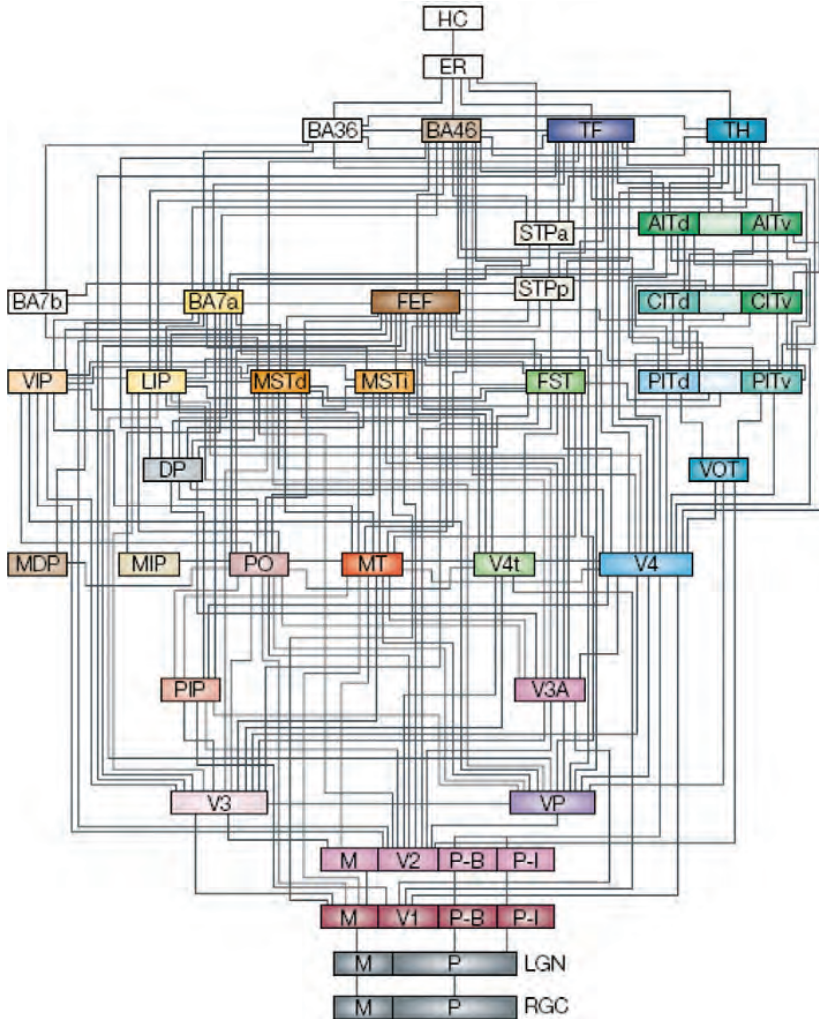


Рис. 1.5 ❖ Приключения фотона.

Принципиальная схема коммуникаций зрительной системы приматов.

Адаптировано из книги Феллемана и Ван Эссена, 1991

1.6. МАНИПУЛЯЦИИ С ВИЗУАЛЬНОЙ СИСТЕМОЙ

Один из способов узнать, как работает какое-нибудь устройство, – это разобрать его, удалить некоторые части и заново оценить функциональность. Например, если мы снимем акустические динамики с автомобиля, он по-прежнему будет нормально работать как средство передвижения, но мы не сможем слушать музыку. Если вынуть аккумулятор, машина не заведется. Удаление частей также является важным способом изучения зрительной системы. Исследователи рассматривают поведенческие дефекты, которые

проявляются при повреждении тех или иных частей мозга в ходе исследований на животных.

Помимо экспериментов с животными, повреждения мозга людей, возникших в результате несчастных случаев, также могут дать важные подсказки относительно функций различных частей зрительного пути (а также про другие аспекты восприятия). Действительно, фундаментальная роль первичной зрительной коры головного мозга в зрении была обнаружена при изучении повреждений мозга. Поднимаясь по зрительной системе за пределы первичной зрительной коры, повреждения могут дать специфические поведенческие дефекты. Например, наблюдатели, которые страдают от редкого, но хорошо известного состояния, называемого *прозопагнозией*, обычно демонстрируют серьезное нарушение способности распознавать лица (гл. 4).

Одна из проблем при интерпретации последствий повреждений головного мозга человека состоит в том, что они часто охватывают большие области мозга и не ограничиваются нейроанатомически или нейрофизиологически определенными локусами. Более контролируемые исследования были выполнены с животными, включая грызунов, кошек и обезьян, для изучения поведенческих дефектов, возникающих после повреждения определенных частей зрительной коры. Являются ли эффекты, вызванные повреждениями, специфичными для одной сенсорной модальности, или они мультимодальны? Насколько избирательны нарушения зрения? Можно ли отделить эффект обучения от следствий различного представления стимула? Что такое нейроанатомический код? Мы вернемся к этим вопросам в гл. 4.

Другой важный путь изучения функций мозга – это изучение эффектов внешней активации определенных мозговых цепей. Один из известных способов сделать это – с помощью электростимуляции ввести в кору головного мозга ток. Грубые методы электростимуляции частей коры часто нарушают обработку зрительного сигнала и повторяют эффекты ограниченного физического повреждения. Одно из преимуществ электростимуляции состоит в том, что его эффект можно быстро убрать, просто прекратив ее, убрав потенциал с электрода. Таким образом, можно изучать одно и то же животное, выполняющее ту же самую задачу, в условиях электрической стимуляции определенной нейронной цепи и без стимуляции. Интересно, что в некоторых случаях более совершенные формы электростимуляции могут привести не к нарушению обработки, а вместо этого к усиленной реакции на определенные типы информации. Например, есть часть мозга, называемая средней височной корой (МТ), которая получает сигналы от первичной зрительной коры и расположена рядом с центром диаграммы на рис. 1.5. Нейроны в этой области играют важную роль в способности различать направление движущихся объектов. Введение локализованного электрического тока в область МТ у макак может исказить восприятие животным того, как движутся предметы в их визуальном мире. Другими словами, можно напрямую создавать зрительные мысли о движении, раздражая субпопуляции нейронов в области МТ (гл. 4). В сочетании с тщательными поведенческими измерениями электрическая стимуляция может дать представление о том, как внешнее воздействие на данный кластер нейронов может влиять на восприятие.

Также существует долгая история исследований электростимуляции у людей с эпилепсией. Нейрохирургам приходится принимать решения о возможности резекции эпилептогенной ткани для лечения припадков. Перед процедурой резекции нейрохирурги используют электрическую стимуляцию, чтобы изучить функцию ткани, которая должна подвергнуться резекции. Знаменитый американско-канадский нейрохирург Уайлдер Пенфилд (1891–1976) был среди пионеров в использовании этой техники для картирования функций мозга. Одно из его знаменитых открытий – это карта «гомункула» сенсомоторного мира: определение топографического расположения областей, где электрическая стимуляция приводит к определенным движениям подопытных или сообщению от них о тактильных ощущениях в ногах, ступнях, пальцах ног, на туловище, языке и лице. Точно так же подопытные сообщают, что при электрической стимуляции первичной зрительной коры видели локализованные вспышки света.

Насколько специфичны эффекты электростимуляции? При каких условиях возбуждение нейронов причинно связано с восприятием? Сколько нейронов и какие типы нейронов активируются при электростимуляции? Как эффекты стимуляции зависят от времени, продолжительности и интенсивности электростимуляции? Мы вернемся к этим вопросам в гл. 4.

1.7. ФУНКЦИИ ЦЕПЕЙ ЗРИТЕЛЬНОЙ КОРЫ

Золотым стандартом для изучения функций цепей мозга является имплантация микроэлектрода (или нескольких микроэлектродов) в интересующую экспериментатора область (рис. 1.6). Микроэлектрод – это тонкий кусок металла, обычно диаметром около 50 мкм, который может регистрировать изменения напряжения, часто во внеклеточной среде. Этот метод был введен Эдгаром Адрианом (1889–1977) в начале 1920-х годов для исследования активности отдельных нервных волокон. Регистрация записей с них требовала разумного использования доступной в то время электроники, чтобы иметь возможность усиливать небольшие перепады напряжения, которые характеризуют электрическую связь внутри нейронов. Эти внеклеточные регистрации электрического потенциала (в отличие от гораздо более сложных внутриклеточных) позволяют исследователям контролировать активность одного или нескольких нейронов, находящихся в непосредственной близости от электрода (в окрестности ~200 мкм) и с временным разрешением в миллисекунды.

Многие неинвазивные методы нацелены на изучение того, что происходит в мозге весьма косвенным образом, путем измерения сигналов, которые имеют слабую корреляцию с совокупной активностью миллионов различных клеток. Результаты этих методов, скорее всего, включают в себя косвенную оценку активности не только нейронов, но и множества других клеток, присутствующих в головном мозге. Что еще хуже, некоторые неинвазивные методы усредняют активность в течение многих секунд, что в несколько тысяч

раз дольше, чем реальные изменения активности, происходящие в мозге. В качестве аналогии представьте социолога, интересующегося тем, что люди в Париже думают об изменении климата; он может опросить многих людей, что трудоемко, но довольно точно (эквивалентно инвазивным регистрациям отдельных нейронов), или же он может усреднить общее количество звука, производимого во всем городе за целую неделю, что намного проще, но не очень информативно (эквивалент неинвазивным измерениям).

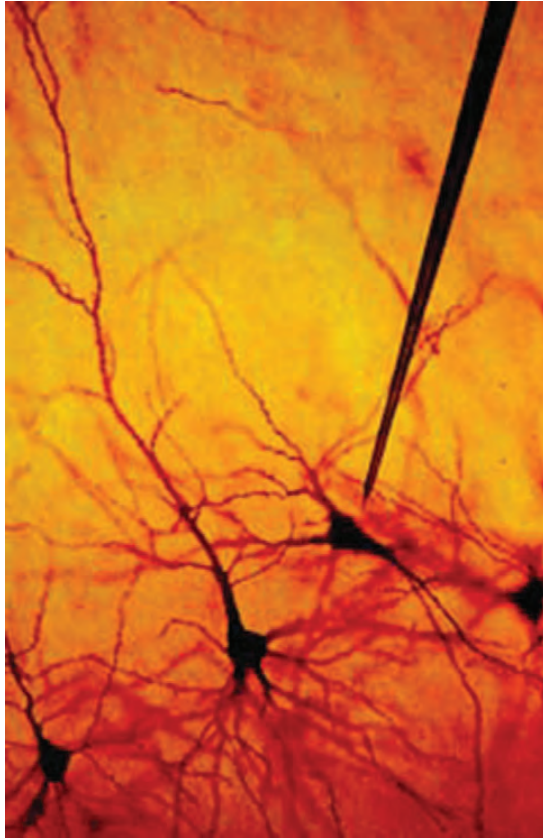


Рис. 1.6 ❖ Прослушивание активности отдельных нейронов с помощью микроэлектрода. На фото – электрическая регистрация с электрода из микропровода. Воспроизведено по Хьюбел, 1995

Регистрация активности нейронов показала, какие типы зрительных стимулов вызывают наибольшее возбуждение в различных областях мозга. Одним из самых ранних открытий были открытия восприимчивого поля нейронов сетчатки, LGN, и первичной зрительной коры. Рецептивное поле определяется как область внутри зрительного поля, где нейронный ответ может быть вызван визуальной стимуляцией (рис. 2.9, гл. 2). Зрительные нейроны разборчивы: они не реагируют на изменения освещения в любой

части зрительного поля. Каждый нейрон отвечает за представление ограниченной области визуального пространства. Вместе все нейроны в данной области мозга образуют карту всего зрительного поля, то есть карту доступной части зрительного поля (например, люди не имеют визуального доступа к тому, что происходит позади них). Размер этих рецептивных полей единичных нейронов обычно увеличивается, начиная от сетчатки до таких областей, как нижняя височная кора, расположенных в верхней части диаграммы на рис. 1.5.

Для активации зрительных нейронов необходимы пространственные и временные изменения освещения в пределах рецептивного поля нейрона. Однако не все световые паттерны равнозначны. Нейроны сильно возбуждаются в ответ на определенные зрительные стимулы и не обращают внимания на другие. В классическом нейрофизиологическом эксперименте Дэвид Хьюбел (1926–2013) и Торстен Визель вставили микроэлектрод таким образом, чтобы регистрировать отклики отдельных нейронов в определенной области первичной зрительной коры головного мозга кошки. После подачи различных зрительных стимулов они обнаружили, что нейрон возбуждается особенно сильно, когда в пределах его рецептивного поля была представлена полоса определенной ориентации. Отклик был слабее, когда полоса имела другую ориентацию. Это предпочтение определенной ориентации является отличительной чертой значительной части нейронов первичной зрительной коры (гл. 5).

Открытие Хьюбела и Визеля вдохновило поколения ученых, занимающихся зрением, использовать электроды, помещенные в зрительную кору для изучения предпочтительных стимулов в различных областях мозга. При регистрациях из других частей зрительной коры исследователи описали нейроны, которые демонстрируют усиленные ответы на стимулы, движущиеся в определенных направлениях, нейроны, которые предпочитают сложные формы, такие как фрактальные узоры или лица, и нейроны, которые особенно чувствительны к цветовому контрасту.

Как возникает избирательность к сложным формам, и какие вычислительные преобразования превращают простую структуру рецептивного поля на уровне сетчатки в более сложные формы? Насколько устойчивы зрительные реакции коры при трансформациях стимулов, подобных показанным на рис. 1.3? Насколько быстро нейроны зрительной коры отвечают на новые стимулы? Что такое нейронный код, или какие особенности нейронных ответов лучше отражают входные стимулы? Какие биологические схемы и механизмы сочетают селективность и инвариантность? Главы 5 и 6 посвящены исследованиям нейрофизиологических реакций зрительной коры.

Зрение – это гораздо больше, чем фильтрация и обработка изображений для их распознавания. Особенно важна обработка зрительного сигнала, потому что она определяет восприятие; она связывает внешний мир с воспоминаниями, текущими целями и внутренними моделями мира. Полная интерпретация изображения, такого как рис. 1.1, и способность ответить на бесконечное количество вопросов об изображении предполагают мост между зрением и восприятием, что мы обсудим в гл. 6.

1.8. К НЕЙРОННЫМ КОРРЕЛЯТАМ ЗРИТЕЛЬНОГО СОЗНАНИЯ¹

Сложный каскад взаимосвязанных процессов в визуальной системе приводит к богатому и субъективному восприятию объектов и сцен вокруг нас. Мы не очень хорошо знаем, как напрямую оценивать субъективное восприятие извне. Откуда мы знаем, что то, что один человек называет красным, совпадает с восприятием красного цвета другим? Некоторое время назад в СМИ шли бурные дискуссии о цвете платья; фотография стала настолько популярной, что теперь известна как «Платье» (рис. 1.7). Некоторые клянутся, что платье сине-черное. Для меня это так же загадочно, как если бы мне сказали, что у этих людей тридцать пальцев на правой руке. Зачем кому-то из честных людей пытаться убедить меня, что это явно бело-золотое платье на самом деле сине-черное? И все же некоторые люди видят платье как бело-золотое, а другие – как отчетливо сине-черное.



Рис. 1.7 ❖ Вирусная фотография платья

Восприятие – в глазах наблюдателя. Если быть более точным, восприятие находится в мозге наблюдателя. Если бы мы работали только на уровне восприятия, не общаясь, мы бы никогда не догадались, что люди могут видеть одно и то же платье такими радикально разными способами. Чтобы косвенно получить доступ к субъективному восприятию, нам нужно изучить поведение субъектов исследования. Платье подчеркивает, что мы не должны позволять нашей субъективной оценке руководить научной повесткой дня. Наша интуиция подвержена ошибкам, о чем мы будем говорить снова и снова.

¹ В этой книге автор использует понятие «зрительное сознание», мало распространенное в российской научной литературе. Для адекватного понимания этого термина сошлемся на понимание сознания, которое приводится в книге Стивена Пинкера: «Модель сознания – это синтез когнитивистики, аффективной и когнитивной нейронауки, социальной и эволюционной психологии и других наук о человеческой природе, которые я анализировал в книгах «Как работает мозг» (How the Mind Works), «Чистый лист» (The Blank Slate) и «Субстанция мышления» (The Stuff of Thought). Согласно такому толкованию, сознание – это комплексная система когнитивных и эмоциональных способностей, интегрированных в мозг, которая обязана своей базовой конструкцией процессам эволюции» (Стивен Пинкер. «Все лучшее в нас»). – Прим. ред.

Целая область с замечательным названием *психофизика* посвящена тщательной количественной оценке поведения как способу оценки зрительного восприятия (гл. 3). Далее мы исследуем, где, когда и как быстро наблюдатели воспринимают различные образы, составляющие субъективную интерпретацию окружающего их мира. Мы также обсудим, почему мозг легко обмануть визуальными иллюзиями. Поведенческие (бихейвиоральные) измерения станут основным ограничением в построении теории обработки изображений.

Зрительное восприятие, конечно же, находится не в пальцах ног и даже не в сердце, как считали некоторые наши предки. Большинство ученых согласятся с тем, что субъективные чувства и образы возникают в результате активности нейронных цепей в головном мозге. Гораздо меньше согласия имеется относительно механизмов, ответственных за субъективные ощущения. «Где», «когда» и особенно «как» так называемых *нейронных коррелятов сознания* составляют область активных исследований и горячих споров. Исторически сложилось так, что многие неврологи избегали исследований в области сознания как слишком запутанной темы или слишком далекой от того, что, как мы считали, стоит серьезных затрат времени и усилий. Однако в последние годы ситуация начала меняться. Хотя все еще очень далекие от разрешения проблемы, систематические и строгие подходы, основанные на нейробиологии, могут однажды раскрыть ответ на одну из величайших загадок нашего времени, а именно на *физическую основу сознательного восприятия*.

По нескольким практическим причинам основы субъективного восприятия в основном (но не исключительно) изучались в области зрения. Были приложены героические усилия по изучению нейронных коррелятов зрительного восприятия с использованием животных. Превалирующая экспериментальная схема предполагает отделение зрительного сигнала от воспринимаемого образа. Например, в мультистабильных восприятиях (как на рис. 1.8) одна и та же вводная картинка может привести к двум различным интерпретациям. В этих условиях исследователи спрашивают, какие нейронные события коррелируют с различными субъективными образами.

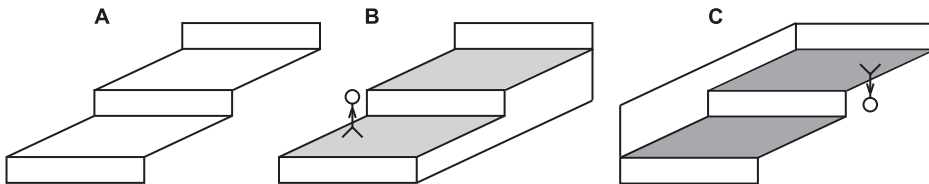


Рис. 1.8 ❖ Бистабильное восприятие. (А) Изображение можно интерпретировать двумя разными способами. (В) В одной версии человек поднимается по лестнице. (С) Другая версия предполагает перевернутый мир

Стало ясно, что возбуждение нейронов во многих частях мозга *не коррелирует* с воспринимаемым образом. В этом, пожалуй, тривиальном примере активность сетчатки важна для зрения, но зрительный образ не возникает до тех пор, пока через несколько синапсов активность не достигает более

высоких слоев зрительной коры головного мозга (гл. 10). Нейрофизиологические, нейроанатомические и теоретические соображения предполагают, что субъективное восприятие коррелирует с активностью в областях, находящихся дальше первичной зрительной коры. Точно так же исследователи предложили верхнюю границу с точки зрения того, где в визуальной иерархии могут быть цепи, участвующие в субъективном восприятии. Хотя повреждения, ограниченные гиппокампом и лобной корой (которые, как считается, лежат в основе памяти и ассоциаций), вызывают серьезные когнитивные нарушения, эти повреждения не затрагивают зрительное восприятие в значительной степени. Таким образом, нейрофизиология и исследования повреждений ограничивают нейронные цепи, участвующие в субъективном зрительном восприятии, несколькими областями, которые на разных стадиях участвуют в обработке зрительной информации вдоль вентральной коры. Несколько нейрофизиологических исследований показывают, что при восхождении вверх через вентральную зрительную кору степень корреляции между активностью нейронов и воспринимаемыми зрительными образами увеличивается.

Как можно изучать зрительное сознание с помощью научных методов? Какие области мозга, цепи и механизмы отвечают за зрительное сознание? Каковы функции зрительного сознания? Какие животные проявляют сознание? Могут ли машины иметь сознание? В гл. 10 дается первое представление о том, что известно (и что неизвестно) в этих увлекательных темах.

1.9. К РАЗРАБОТКЕ ТЕОРИИ ЗРИТЕЛЬНОГО ВОСПРИЯТИЯ

Ричард Фейнман (1918–1988), лауреат Нобелевской премии по физике из Калифорнийского технологического института, как известно, заявил, что понимание устройства означает, что мы должны уметь его построить. Мы стремимся разработать теорию зрения, которая может объяснить, как люди и другие животные воспринимают и интерпретируют окружающий мир. В одной из основополагающих работ по зрению Дэвид Марр (1945–1980) определил три уровня понимания, которые мы можем представить: (1) какова функция зрительной системы¹, (2) как ведет себя зрительная система при разных данных на входе и в разных обстоятельствах и (3) как биологический объект реализует эти функции и поведение.

Успешная теория зрения должна поддаваться вычислительной реализации, и в этом случае мы можем напрямую сравнивать выходные данные вычислительной модели с показателями поведенческих характеристик и регистрациями активности нейронов. Полная теория будет включать информацию из исследований повреждений, нейрофизиологических регистраций электрической активности нейронов, психофизических исследований

¹ Также используется термин «зрительный анализатор». – *Прим. ред.*

и исследований электростимуляции. В гл. 7 и 8 представлены современные подходы к построению вычислительных моделей и теорий зрительного распознавания.

При отсутствии полного понимания схемы подключения и лишь скудных знаниях о нейрофизиологических реакциях важно задуматься, стоит ли вообще думать о построении теорий. Но это не только полезно, на самом деле *важно* развивать теории и воплощать их в жизнь с помощью вычислительных моделей, чтобы продвигать эту область вперед. Вычислительные модели могут объединять существующие данные из разных лабораторий, методов и условий экспериментов и помогать согласовывать на первый взгляд несопоставимые наблюдения. Математические модели могут формализовать знания и предположения и предоставить количественный, систематический и строгий путь к изучению вычислений в зрительной коре головного мозга. Хорошая модель должна основываться на эмпирических данных и, в свою очередь, давать нетривиальные и экспериментально проверяемые прогнозы. Эти прогнозы можно оценить эмпирически, чтобы подтвердить, опровергнуть или расширить модели. Даже опровержение моделей несет положительный характер. Демонстрация того, что модель неверна, является прогрессом и помогает нам создавать более совершенные модели.

Как мы строим и тестируем вычислительные модели? Как нам справиться с нехваткой знаний и огромным количеством параметров, которые часто требуются? Какие приближения и абстракции можно делать? Если будет слишком много упрощений, мы можем упустить важные особенности задачи. Но при слишком малом упрощении мы можем провести десятилетия, увязнув в несущественных деталях.

В качестве простой аналогии рассмотрим, как в доньютоновскую эпоху физики пытались охарактеризовать движение объекта при приложении к нему силы. В принципе, один из этих ученых мог бы подумать о многих переменных, которые могут повлиять на движение объекта, включая форму объекта, его температуру, время дня, материал объекта, поверхность, на которой он стоит, и точное положение точки приложения силы. По-видимому, мы должны быть благодарны за отсутствие компьютеров во времена Ньютона: не было возможности запускать сложные симуляции машинного обучения, которые включали бы все эти несущественные переменные, чтобы понять красоту линейной зависимости между силой и ускорением. С другой стороны, чрезмерное упрощение (игнорирование массы объекта в данном случае) сделало бы модель бесполезной. Центральной целью вычислительной нейробиологии является достижение правильного уровня абстрагирования для каждой задачи, решение Златовласки, которое не является ни излишне подробным, ни слишком упрощенным. Альберт Эйнштейн (1879–1955) ссылаясь на модели, которые максимально просты, но не проще, чем необходимо¹.

Особенно интересным практическим следствием построения теорий зрения является возможность научить компьютеры видеть (гл. 8 и 9). Мы по-

¹ Важнейшим методом решения многих физических задач является построение правильной «физической модели» – упрощенной версии физической системы или процесса, – сохраняющей их главные свойства. – *Прим. ред.*

стоянно используем зрение для решения множества повседневных задач. Если мы сможем научить компьютеры некоторым уловкам, используемым зрением, тогда машины смогут помочь нам решить эти задачи, а возможно даже смогут решить многие из этих задач быстрее и лучше, чем мы. Последнее десятилетие стало свидетелем впечатляющего роста доступности алгоритмов компьютерного зрения для решения многих задач распознавания образов. От телефона, который может распознавать лица, до компьютеров, которые могут помочь врачам в диагностике рентгеновских снимков, до автомобилей, которые могут обнаруживать пешеходов, до классификации изображений растений или галактик – список увлекательных приложений продолжает расти.

В гл. 9 будет представлен обзор современных подходов к компьютерному зрению для решения различных задач зрительного восприятия. Люди по-прежнему превосходят компьютеры во многих визуальных задачах, но разрыв между людьми и машинами быстро сокращается. Мы доверяем машинам вычислять квадратный корень из семи с любым количеством десятичных знаков, но у нас еще нет такого же уровня строгости и эффективности в автоматическом распознавании образов. Однако для многих реальных приложений такая точность может не требоваться. В конце концов, люди тоже совершают ошибки в распознавании. Мы можем довольствоваться алгоритмом, который делает меньше ошибок, чем люди при выполнении той же задачи. Например, при автоматическом распознавании лиц на фотографиях правильное определение 99 % лиц может быть очень хорошей установкой качества решения. Слепые люди были бы рады использовать устройства для определения того, куда они направляются, даже если их мобильное устройство зрения может распознавать только часть зданий в данном месте.

Алан Тьюринг (1912–1954), известный британский математик, который помог расшифровать коды, используемые нацистами для секретных сообщений, и который считается одним из отцов-основателей информатики, предложил простой тест, чтобы оценить, насколько умна машина. В контексте зрения представьте, что у нас есть две комнаты с закрытыми дверями. В одной из комнат находится человек; в другой комнате находится машина, которую мы хотим протестировать. Мы можем передать любую картину в ту или иную комнату и можем задать любые вопросы по картине. Машина и человек возвращают ответы на машинописном листе бумаги, так что мы не можем различить их по голосу или почерку, и других возможностей сделать это нет. Основываясь только на вопросах и ответах, нам нужно решить, в какой комнате находится машина, а в какой – человек. Если для любой картинке и любого вопроса о картинке мы не можем определить, какие ответы приходят от машины, а какие от человека, мы говорим, что машина прошла визуальный тест Тьюринга.

Дразняще, захватывающе и, возможно, немного страшно думать, что в течение нашей жизни мы сможем построить компьютеры, которые пройдут по крайней мере некоторые ограниченные формы визуального теста Тьюринга. Теория Эндрю Паркера предполагает, что животная жизнь в том виде, в каком мы ее знаем, началась с «включения света», вызванного появлением первых глаз на Земле. Возможно, мы близки к очередному ароморфозу – включению

зрения у машин. Вполне вероятно, что жизнь радикально изменится, когда машины смогут видеть мир так, как мы. Возможно, мы на горизонте второго кембрийского взрыва – эпохи, в которой могут появиться новые гибридные виды существ с машинно-расширенным зрением, в которой мы сможем доверять машинному зрению больше, чем своим собственным глазам, и где машины будут прокладывать путь к открытиям так же, как наше зрение руководило нами на этом пути на протяжении последних тысячелетий.

1.10. ОБЗОР ГЛАВЫ

- Теория включения света утверждает, что именно появление глаз вызвало «кембрийский взрыв» – резкое увеличение числа и разнообразия видов животных.
- Теория зрительного распознавания должна учитывать четыре основных свойства зрения: избирательность, толерантность, скорость и большую емкость.
- Повреждения головного мозга и электрическая стимуляция дают возможность вмешательства в зрение и, таким образом, начинают раскрывать функциональную архитектуру, отвечающую за обработку изображений.
- Изучение активности отдельных нейронов в зрительной системе открывает дверь для понимания нейронных вычислений, ответственных за преобразование пикселей в зрительный образ.
- Зрение находится в мозге наблюдателя. Восприятие – субъективно.
- Поиск механизмов сознания требует выявления нейронных коррелятов субъективного восприятия.
- Вдохновленные и ограниченные нейрофизиологической функцией, нейроанатомическими схемами и исследованиями повреждений, мы можем обучать компьютеры видеть и интерпретировать мир таким же образом, как это делают люди.

ЛИТЕРАТУРА

См. <http://bit.ly/2TqTDt5> для получения дополнительных ссылок.

- Hubel, D. (1979). The brain. *Scientific American* 241: 45–53.
- Marr, D. (1982). *Vision*. San Francisco: Freeman Publishers.
- Parker, A. (2004). *In the blink of an eye: how vision sparked the big bang of evolution*. New York: Basic Books.
- Poggio, T.; and Anselmi, F. (2016). *Visual cortex and deep networks*. Cambridge, MA: MIT Press.
- Ullman, S. (1996). *High-level vision*. Cambridge, MA: MIT Press.