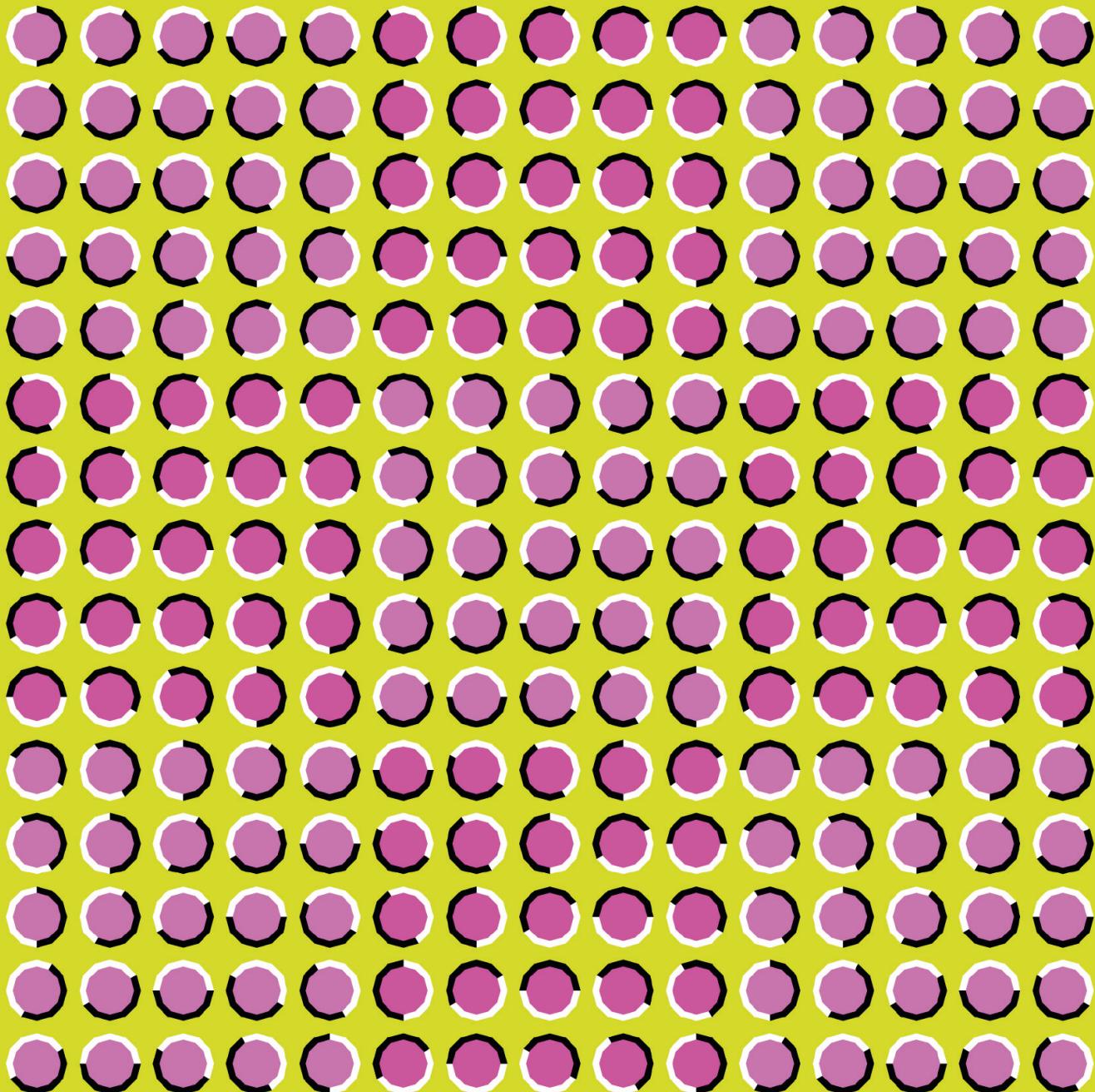


# БИОЛОГИЧЕСКОЕ И КОМПЬЮТЕРНОЕ ЗРЕНИЕ



ГЭБРИЕЛ КРЕЙМАН

Гэбриел Крейман

# **Биологическое и компьютерное зрение**

# Biological and Computer Vision

**GABRIEL KREIMAN**  
Harvard University



**CAMBRIDGE**  
UNIVERSITY PRESS

# Биологическое и компьютерное зрение

ГЭБРИЕЛ КРЕЙМАН

Гарвардский университет



Москва, 2022

**УДК 004.8****ББК 32.81****K62****Крейман Г.**

**K62** 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. 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

*Всем моим учителям – настоящим, прошлым и будущим*

# Содержание

<b>От издательства.....</b>	11
<b>Предисловие.....</b>	12
<b>Благодарности .....</b>	17
<b>Сокращения.....</b>	19
<b>Глава 1. Знакомство с миром зрения .....</b>	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
<b>Глава 2. Путешествие фотона. Статистика естественного изображения и сетчатка .....</b>	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
<b>Глава 3. Феноменология зрения.....</b>	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.....**

.....	93
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.....**

.....	123
5.1. О неокортике .....	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

## 8 ♦ Содержание

---

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 центрального потока увеличивается .....	143
5.13. Увеличение размера воспринимающего поля вдоль центрального 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
<b>Глава 6. От высших уровней обработки зрительного сигнала к распознаванию образов .....</b>	<b>152</b>
6.1. Зона с хорошей коммуникацией .....	153
6.2. ITC-нейроны демонстрируют избирательность к формам .....	153
6.3. Избирательность центральной зрительной коры человека .....	156
6.4. Чего хотят нейроны ITC <i>на самом деле?</i> ? .....	158
6.5. Нейроны ITC 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 ITC 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
<b>Глава 7. Нейробиологически подобные вычислительные модели .....</b>	<b>177</b>
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
 <b>Глава 8. Учим компьютеры видеть .....</b>	 200
8.1. Краткое описание и определения .....	200
8.2. Общие темы в моделировании центрального потока обработки зрительного сигнала.....	204
8.3. Арсенал моделей .....	205
8.4. Общая схема решения задачи распознавания образов .....	208
8.5. Восходящие иерархические модели центрального 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
 <b>Глава 9. К миру с разумными машинами, которые смогут видеть и интерпретировать мир вокруг нас.....</b>	 248
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

**10 ♦ Содержание**

---

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
<b>Глава 10. Зрительное сознание .....</b>	<b>287</b>
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
<b>Предметный указатель.....</b>	<b>310</b>

# 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](http://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](mailto: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/](http://dmkpress.com/authors/publish_book/) или напишите в издательство по адресу [dmkpress@gmail.com](mailto: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](mailto: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](mailto: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 Персса «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 коннектомов<sup>1</sup>, 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 Байденом 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 воспламеняют 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. За ранее 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 темы 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-

<sup>1</sup> Коннектом – полное описание структуры связей в нервной системе того или иного организма. – Прим. ред.

## 14 ♦ Предисловие

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, упомянутый 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 нарушают 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 поправляли теории 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-

**16 ♦ Предисловие**

---

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 проницательные 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 Гитчаунца, Кэтрин Харрисон, Хавьера Масиса, Оливию МакГиннис, Дэниела Ротшильда, Шона Салливана, Кеннета Шинозуку, Дункана Стоттерса, 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 Иsic, Кохитию Кару, Джие Ким, Билла Лоттера, Диего Мендозза-Халлидея, Кена Накаяму, 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, Кендума Бербанка, Стивена Каспера, Прабаха Гангпадхья, Камиллу Гомес-Лаберже, Стефана Гжековски, Шаши Гупту, Уолтера Хардести, Элеонору Яселли, Лейлу Иsic, Джие Ким, Фила Кунке, Гарретта Лама, Даниэла Лопес Мартинеса, Билла Лоттера, Радхику Мадхавана, Томаса Микони, Пранава Мишру, Шарлотту Моэрманн, Кэндис Росс, Мартина Шримпф, Нимрома Шахама, Джедедию Сингер, Ханлин Тан, Бена Цуу, Каспера Винкена, Джерри Вана, Фараназа Вика, Эрика и Кевина Ву, Уилла

## 18 ♦ Благодарности

---

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 часть вспомнить о том, что я в долгую 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, проницательными 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 коленчатое 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 (Фуксисма, 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).

**20 ♦ Сокращения**

---

**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 Маатен 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 **МТ** (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 «здравому 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, различия 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.

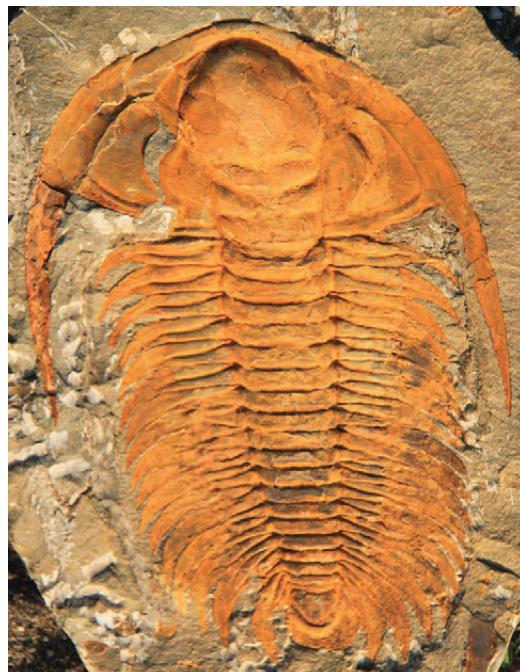
Способность биологических организмов улавливать свет – древняя. Например, многие бактерии используют свет для фотосинтеза, предшественника аналогичного процесса, улавливающего энергию в зеленых растениях. Что особенно поражает в зрении, так это возможность использования света для сбора информации о мире. 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 существ<sup>1</sup>. Примерно в то же время окаменелости свидетельствуют о появлении первого вида, имеющего глаза, – трилобитов (рис. 1.2). Трилобиты – это вымершие членистоны ( 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

<sup>1</sup> Так называемый «кембрийский взрыв». – Прим. ред.

## 24 ♦ Глава 1

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 реальности<sup>1</sup>, 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 понадобиться 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,

---

<sup>1</sup> 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».

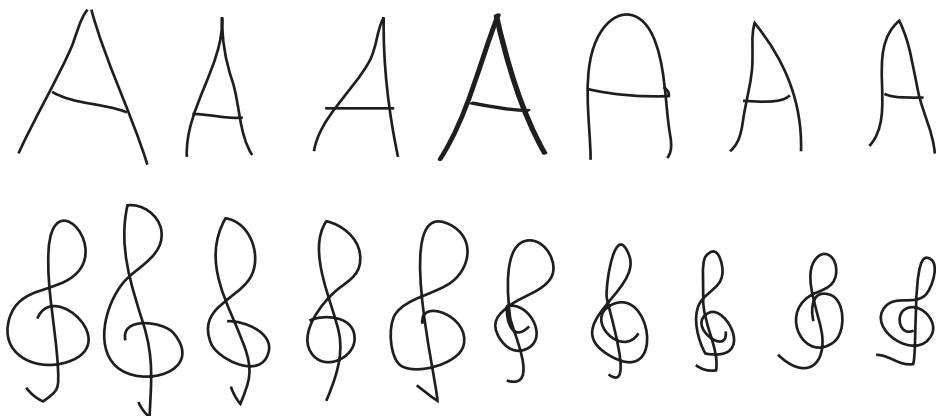
## 1.3. ПОЧЕМУ ЗРЕНИЕ – НЕ ТАКАЯ ПРОСТАЯ ВЕЩЬ?

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

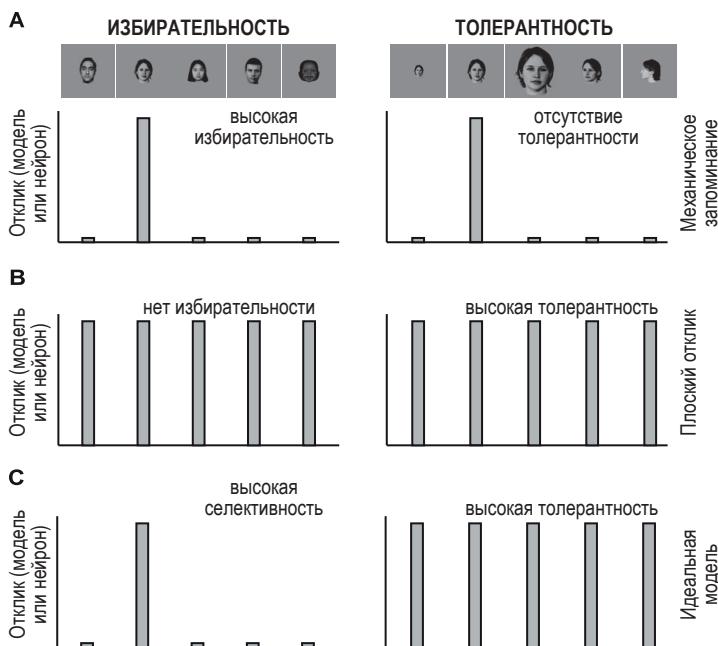
Чтобы объяснить, как зрительная система справляется с идентификацией образов, нам необходимо учесть четыре важнейшие особенности зрительного распознавания: *избирательность, толерантность, скорость и вместимость*.

*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-

## 28 ♦ Глава 1

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 руках<sup>1</sup>.



**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 (Б) имеет хорошую толерантность, но плохую избирательность. (С) Идеальная модель должна сочетать избирательность и толерантность

<sup>1</sup> 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 совершают 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 хващаются 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  $\sim 6,4$  млн «колбочек» и  $\sim 110$  млн «папочек» (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 ядро<sup>1</sup> 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, отчитывающие сутки<sup>2</sup>, и воспринимаемые сетчаткой световые потоки в конце концов через супрахиазматическое ядро помогают модулировать и регулировать этот цикл.

<sup>1</sup> Ядро передней части гипоталамуса. – Прим. ред.

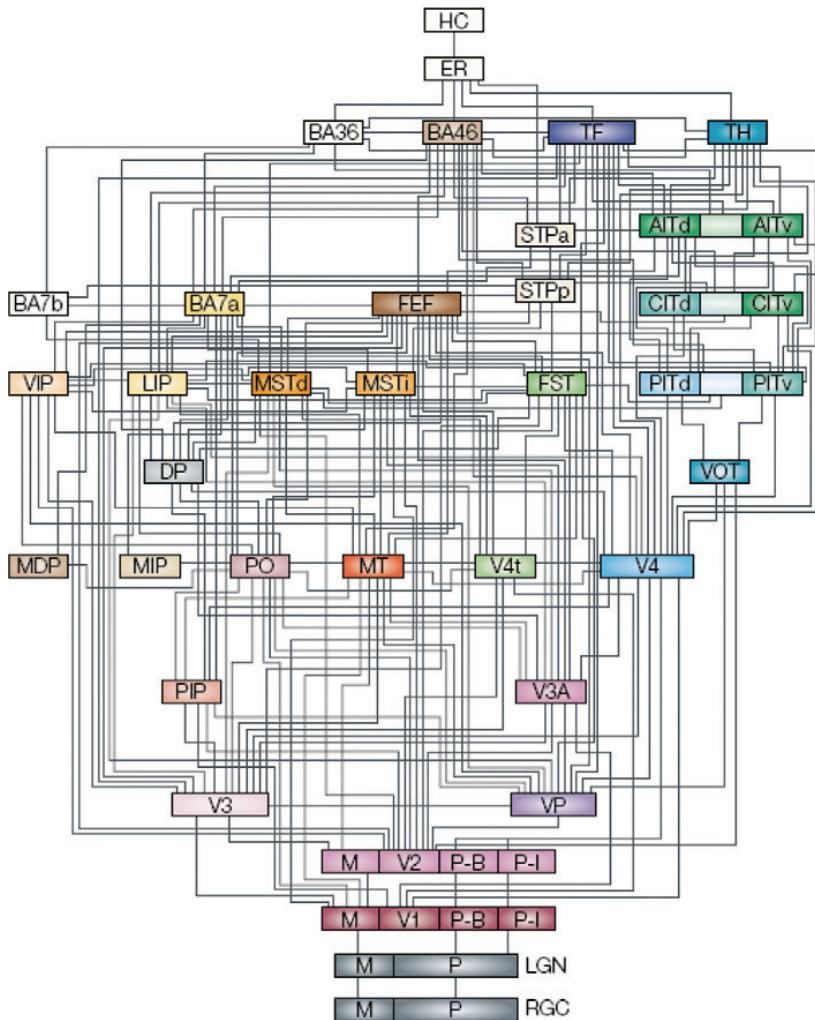
<sup>2</sup> Эндогенный циркадный ритм человека составляет в среднем 24 часа 11 минут. – Прим. ред.

Основной зрительный канал передает информацию от сетчатки к части таламуса, называемой латеральным коленчатым 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 Эссеном 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. Во-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 при-чудливым 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 американо-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 движением 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.

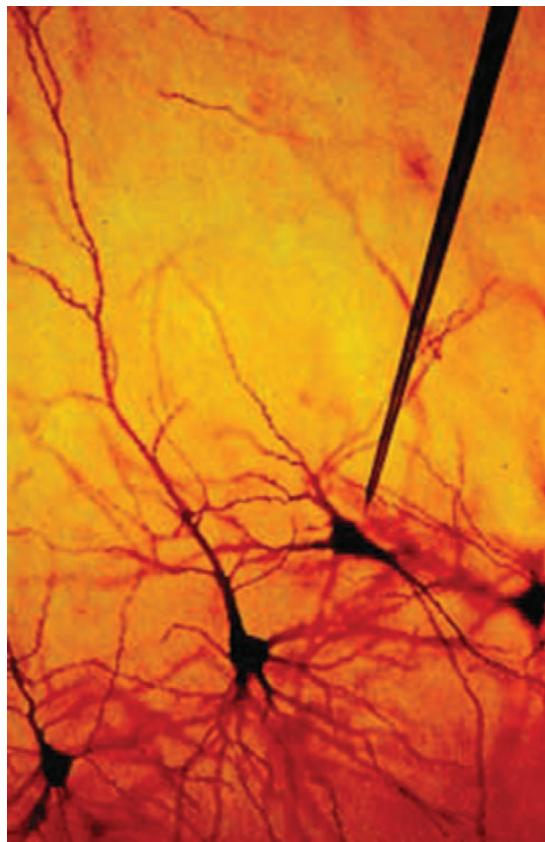
Насколько специфичны эффекты электростимуляции? При каких условиях возбуждение нейронов причинно связано с восприятием? Сколько нейронов и какие типы нейронов активируются при электростимуляции? Как эффекты стимуляции зависят от времени, продолжительности и интенсивности электростимуляции? Мы вернемся 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 стимулацией (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 СОЗНАНИЯ<sup>1</sup>

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.

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.



Рис. 1.7 ♦ Вирусная фотография платья

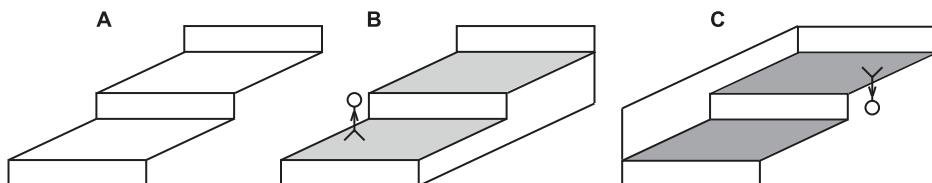
<sup>1</sup> В этой книге автор использует понятие «зрительное сознание», мало распространенное в российской научной литературе. Для адекватного понимания этого термина сошлемся на понимание сознания, которое приводится в книге Стивена Пинкера: «Модель сознания – это синтез когнитивистики, аффективной 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.

## 38 ♦ Глава 1

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 (бихевиоральные) 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 некоторым 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. (A) Изображение можно интерпретировать двумя разными способами. (B) В одной версии человек поднимается по лестнице. (C) 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 центральной коры. Несколько нейрофизиологических исследований показывают, что при восхождении вверх через центральную 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 системы<sup>1</sup>, (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

<sup>1</sup> 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 скучных 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 доныштоновскую 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) ссылался на модели, которые максимально просты, но не проще, чем необходимо<sup>1</sup>.

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

<sup>1</sup> 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 картинки 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.

Дополнительный контент вы можете скачать по ссылке <http://bit.ly/2TqTDt5>.

# Глава 2

---

## Путешествие фотона. Статистика естественного изображения и сетчатка

«И стал свет...» Зрение начинается, когда электроны, отраженные от объектов вокруг нас, попадают на сетчатку глазного дна. Хотя сейчас это может показаться нам довольно очевидным, человечеству потребовалось несколько столетий, если не больше, чтобы прийти к такому выводу. Разделение исследования оптики как раздела физики и зрительного восприятия как раздела нейробиологии произошло недавно. Представления о природе зрительного восприятия переплетались с представлениями об оптике на протяжении Античности и Средневековья. Гиганты мысли величины Платона (~428–~348 гг. до н. э.) и Евклида (~300 г. до н. э.) выдвигали теорию проекции, согласно которой тонкие лучи света, исходящие из глаз, либо достигали самих объектов, либо встречались на полпути с другими лучами света, исходящими от объектов, вызывая чувство зрения<sup>1</sup>. Различие между светом и зрением восходит к Аристотелю (384–322 гг. до н. э.), но не получило широкого признания до исследования свойств глаза Иоганном Кеплером (1571–1630).

Свет преобразуется в электрические сигналы клетками-фоторецепторами, что является одним из поразительных достижений эволюции; это позволяет организму быстро делать выводы об удаленных объектах и событиях в окружающей среде. Функция зрительной системы – быстро извлекать информацию о том, что происходит вокруг нас. Таким образом, структура окружающей среды играет решающую роль в определении схемы связей и физиологических реакций всей зрительной системы и знаменует собой начало нашего путешествия.

---

<sup>1</sup> Наиболее близкое к современному пониманию феномена зрения провозгласил Пифагор: свет – это поток частиц, которые излучают предметы; проникая в глаз человека, они приносят информацию о том, что нас окружает. – Прим. ред.

## 2.1. ЕСТЕСТВЕННЫЕ ИЗОБРАЖЕНИЯ ИМЕЮТ СВОИ ОСОБЕННОСТИ

Давайте рассмотрим цифровое изображение размером  $100 \times 100$  пикселей и ограничимся монохроматическим миром, где каждый пиксель может принимать 256 оттенков серого ( $0 =$  черный,  $255 =$  белый). Такие маленькие бесцветные участки изображения далеки от сложности реальной входящей визуальной информации. Тем не менее даже при этих ограничениях существует огромное количество возможных вариантов этого изображения. Есть 256 изображений с одним пикселием,  $256^2$  изображений с двумя пикселями и т. д. Всего существует  $256^{10\,000}$  возможных изображений с разрешением  $100 \times 100$  пикселей. Это число больше единицы, за которой следует 24 000 нулей: вариантов этой картинки больше, чем текущая оценка общего числа звезд во Вселенной.

Теперь возьмите цифровую камеру, довольно старую, с сенсором, состоящим всего из  $100 \times 100$  пикселей, переключите настройку на черно-белое изображение с восемью битами ( $2^8 = 256$ ) и начните снимать случайные изображения (рис. 2.1). Если вы снимаете одно изображение в секунду и потратите целую неделю на съемку кадров, не будете спать и не будете прерываться на еду, вы получите меньше миллиона картинок, микроскопическое количество из числа всех возможных. Однако даже с этим крошечным набором вы начнете замечать довольно любопытные закономерности. Распределение собранных вами кусочков *естественного* изображения имеет особые свойства, которые охватывают интересное подмножество всех возможных изображений такого размера и диапазона яркости.



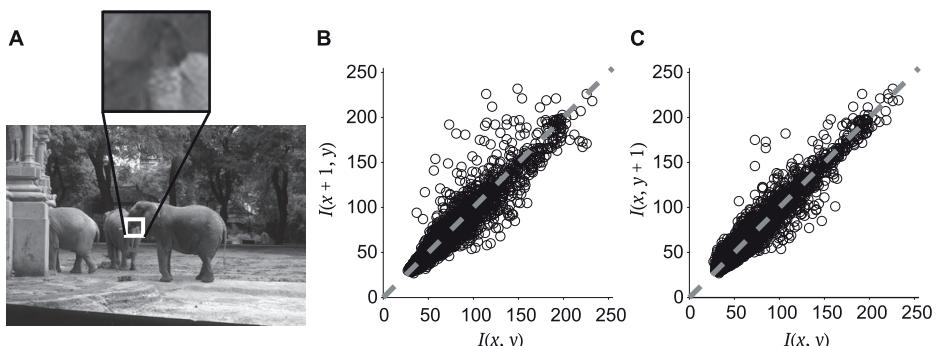
**Рис. 2.1** ♦ Естественные изображения имеют свои особенности. Шестьдесят четыре примера полутоновых элементов изображения размером  $100 \times 100$  пикселей, извлеченных из фотографий. Встречающиеся в природе элементы изображений такого размера составляют микроскопическое подмножество всех возможных элементов изображения размером  $100 \times 100$

В принципе, в естественном мире может появиться любая из  $256^{10\,000}$  картинок  $100 \times 100$  в оттенках серого. Однако существуют сильные корреляции и ограничения в том, как выглядят естественные изображения. Особенно бросается в глаза то, что существует тенденция к сильной корреляции между яркостью любых двух соседних пикселей (рис. 2.2). Другими словами, яркость серого в естественных изображениях обычно изменяется плавно и содержит области примерно одинаковой яркости. Эти поверхности разделены границами, которые представляют собой разрывы, где такие корреляции между соседними пикселями нарушаются; эти границы – скорее исключение, чем правило. Они играют значительную роль в зрении (гл. 5), но занимают небольшую часть площади изображения.

Одним из способов количественной оценки этих пространственных закономерностей является вычисление функции автокорреляции. Для упрощения рассмотрим изображение только в одном измерении. Если  $f(x)$  обозначает яркость в точке  $x$ , то функция автокорреляции  $A$  измеряет среднюю корреляцию яркостей пикселей как функцию расстояния  $\Delta$  между двумя точками:

$$A(\Delta) = \int f(x)f(x - \Delta)dx, \quad (2.1)$$

где интеграл берется по всему изображению (в данном случае отрезку). Это определение может быть легко распространено на изображения с большей размерностью и цветные<sup>1</sup>. Функция автокорреляции естественного изображения обычно показывает пик при небольшом расстоянии между пикселями, за которым следует постепенное снижение.



**Рис. 2.2 ♦** Мир довольно гладкий. Для небольшого участка размером  $100 \times 100$  пикселей из изображения в части А (белый прямоугольник, увеличенный на вставке) диаграммы рассеяния показывают яркость серого в позиции  $(x, y)$  в зависимости от яркости оттенков серого в положении  $(x + 1, y)$  (Б, смежный по горизонтали пиксель) или в положении  $(x, y + 1)$  (С, смежный по вертикали пиксель). Между яркостями соседних пикселей в естественных изображениях обнаруживается сильная корреляция

<sup>1</sup> Последнее фактически добавляет еще два измерения. – Прим. ред.

Связанный способ оценки пространственных корреляций в изображении – вычисление его спектра яркостей. Интуитивно можно преобразовать корреляцию из *пиксельного представления* в *частотное представление*. Обратите внимание, что здесь, когда мы говорим «частота», мы имеем в виду *пространственные* частоты – то есть насколько быстро величины меняются в пространстве. Если уровень велик на высоких частотах, это подразумевает существенные изменения на малых расстояниях между пикселями, как это можно наблюдать на границе. И наоборот, больший уровень на низких частотах подразумевает более гладкие изменения и плавность изменений в пиксельной области. Если  $P$  обозначает уровень, а  $f$  – пространственную частоту, естественные изображения обычно показывают, что уровень  $P$  уменьшается по  $f$  примерно как

$$P \sim 1/f^2. \quad (2.2)$$

В естественных изображениях на низких частотах наблюдается значительный больший уровень, чем на высоких частотах. Представленная здесь функция является степенной. Законы, определяемые степенной функцией, распространены во многих природных явлениях: размеры кратеров на Луне, частота употребления слов, площади перебоев в подаче электроэнергии, количество уголовных обвинений на одного осужденного и человеческие суждения об интенсивности стимулов – все они подчиняются степенному закону распределения. Важное свойство степенных законов – инвариантность масштаба. В частности, если  $P(f) = a \cdot 1/f^2$ , где  $a$  – константа, и мы умножим  $f$  на скаляр  $c$ ,  $f' = cf$ , то  $P(f') = a \cdot 1/(cf)^2 = a/c^2 \cdot 1/f^2 = a' \cdot 1/f^2$ , с новой константой  $a' = a/c^2$ . Если мы изменим масштаб изображения, его спектр уровней по-прежнему будет иметь ту же форму, что и в предыдущем уравнении.

## 2.2. ЭФФЕКТИВНОЕ КОДИРОВАНИЕ ЗА СЧЕТ ВЫДЕЛЕНИЯ ДОПОЛНИТЕЛЬНЫХ РЕСУРСОВ ТАМ, ГДЕ ОНИ НЕОБХОДИМЫ

Одна из причин, по которой мы заинтересованы в определении характеристик естественных изображений, – это предположение о том, что мозг особенно хорошо приспособлен для восприятия реального мира. Эта идея, известная как *принцип эффективного кодирования*, утверждает, что зрительная система специализируется на особенностях изображений, присущих окружающему нас миру. Если в любом типичном изображении присутствует только часть из  $256^{10000}$  возможных элементов изображения, разумно использовать большую часть нейронов для обработки именно этой части пространства возможных изображений. Эволюция накладывает ограничения на размеры мозга, и возникает соблазн предположить, что мозг не заполнен нейронами, кодирующими характеристики изображений, которые никогда не встречаются в естественном мире. Кроме того, мозг стоит дорого и с энергетической

точки зрения, поэтому имеет смысл тратить больше ресурсов там, где они более необходимы.

Зная структуру и свойства естественных изображений, можно выдвигать гипотезы, которые можно проверить, о предпочтениях нейронов, обрабатывающих визуальную информацию. Мы вернемся к этой теме, когда углубимся в нейронные схемы, участвующие в обработке визуальной информации (далее в этой главе, а также в разделе 6.12). Такая специализация для обработки естественных изображений может возникнуть как следствие эволюции (филогенеза) и как следствие обучения через визуальное взаимодействие с миром (онтогенеза). Вопрос о филогенезе и онтогенезе постоянно возникает при исследованиях практических функций мозга. Как и в других случаях дилеммы «филогенез – онтогенез», вполне вероятно, что оба варианта верны.

Некоторые особенности зрительной системы жестко запрограммированы, но зрительный опыт играет центральную роль в формировании настроек нейронов. Например, работа светочувствительных молекул в фоторецепторах жестко запрограммирована; мы не можем видеть цвета за пределами видимого спектра, независимо от того, насколько сильно на нас может действовать излучение с такими частотами. С другой стороны, изменение статистики различных видов воспринимаемых изображений может привести к изменению реакции нейронов на зрительные стимулы. Мы вернемся к вопросу о том, какие особенности нейронной схемы жестко запрограммированы, а какие пластичны, когда будем обсуждать зрительную кору (раздел 6.12). Можно сделать предположение, что пластичность увеличивается по мере продвижения по зрительной системе от основных сенсорных элементов к кортикалным реакциям. Согласно этой гипотезе, начальная обработка визуальной информации, обсуждаемая в этой главе, в основном является аппаратной<sup>1</sup>.

## 2.3. ВИЗУАЛЬНЫЙ МИР – МЕДЛЕННЫЙ

Визуальные свойства близлежащих локаций в мире природы обычно похожи друг на друга. В дополнение к этим пространственным корреляциям в мире природы существуют также сильные временные корреляции. Расширяя коллекцию фотографий естественного мира в разделе 2.1, представьте, что вы возвращаетесь в те же места и теперь собираете короткие видеоролики продолжительностью две секунды, удерживая камеру неподвижно. Поскольку камера не может двигаться, единственные изменения кадров в видео будут продиктованы движением объектов. Если вы используете камеру, которая снимает 30 кадров в секунду, в большинстве случаев соседние кадры в этих видеороликах будут выглядеть очень похожими. За некоторыми исключениями, объекты в мире движутся довольно медленно. Представьте гепарда или машину, движущихся с довольно большой ско-

<sup>1</sup> То есть жестко автоматической. – Прим. ред.

ростью – 80 км/ч. Если предположить, что у нашей камеры дистанция порядка 30 м проецируется на 2000 пикселей матрицы, гепард или машина будут перемещаться примерно на 30 пикселей от одного кадра к другому. Большинство объектов движутся с меньшей скоростью. Следовательно, временной спектр уровней яркости естественного мира также показывает пик на низких временных частотах с заметными изменениями, происходящими в течение десятков и сотен миллисекунд. Визуальный мир медленный и в основном непрерывный<sup>1</sup>.

В некоторых вычислительных моделях при разработке алгоритмов использовались преимущества непрерывности визуального ввода в соответствии с естественным зрением, распознающим объекты и их трансформации, – тема, к которой мы еще вернемся при обсуждении вычислительных расчетов при обучении визуальной системы (гл. 8). Поскольку движение довольно медленное и непрерывное, мы можем предположить, что последовательность изображений, доходящих до глаз, обычно содержит один и тот же объект, демонстрирующий относительно медленное изменение формы. Многочисленные примеры такого могут быть использованы для настройки модели на толерантность к преобразованиям, упомянутой в гл. 1. Идея использования временной непрерывности в качестве ограничения при обучении часто упоминается как «принцип медленности» (slowness principle).

## 2.4. НАШИ ГЛАЗА НАХОДЯТСЯ В НЕПРЕРЫВНОМ ДВИЖЕНИИ

Предположение, что камера совершенно неподвижна в предыдущем разделе, не совсем верно, если рассмотреть, что происходит в реальности. Для начала мы можем двигать головой, тем самым изменения информацию, попадающую в глаза. Однако движения головы довольно редкие и относительно медленные. Оказывается, что даже при совершенно неподвижной голове люди и другие приматы все время двигают глазами. Это наблюдение может показаться несколько нелогичным. Если вы не обратили внимания на движения своих глаз или не потратили время на изучение движений глаз другого человека, может возникнуть впечатление, что визуальный мир вокруг нас вообще не меняется в отсутствие движений внешних объектов или движений нашей головы. Однако весьма легко можно принять концепции, полученные в результате самоанализа, не подвергая сомнению наши предположения и не проверяя их результатами экспериментов.

В настоящее время в лаборатории без особых усилий можно достаточно точно зарегистрировать и измерить движения глаз; это было не всегда

---

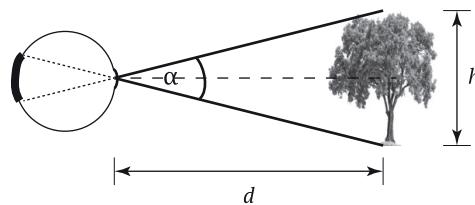
<sup>1</sup> Объекты в поле зрения достаточно редко возникают и исчезают мгновенно. – Прим. ред.

так, однако физики создали гениальные приспособления для фиксации этих быстрых движений глаз. На рис. 2.3 показан пример последовательности движений глаз во время представления статического изображения наблюдателю. Глаза некоторое время остаются направлены примерно на одно место, а затем быстро перескакивают на другое место, исследуя его, прежде чем снова отправиться в путешествие к новой цели. Эти быстрые скачки глаз называются *саккадами*, и обычно движение от исходного до конечного положения занимает несколько десятков миллисекунд. Относительно малоподвижные положения глаз между саккадами называются *фиксациями*.



**Рис. 2.3** ♦ Люди непрерывно двигают глазами. Траектория движения глаз испытуемого во время наблюдения изображения в течение 12 с. На этом рисунке показаны положения глаз, усредненные каждые 33 мс (красные кружки), а желтые линии саккад соединяют последовательные положения фиксации глаз. Угол обзора всего дисплея составляет примерно  $20 \times 30^\circ$

Во время рассматривания сцены наблюдатели обычно совершают скачки с изменением угла зрения глаз примерно на  $4^\circ$ . Градусы угла зрения являются наиболее подходящей и стандартной единицей измерения размеров и положения в поле зрения и отражают тот факт, что существует множество комбинаций размеров объектов и расстояний до глаза, которые образуют один и тот же угол (рис. 2.4). Угол обзора в один градус примерно соответствует длине большого пальца на расстоянии вытянутой руки. При естественном рассматривании сцены подопытные совершают саккады примерно каждые 250–300 мс.



**Рис. 2.4** ♦ Размеры объектов измеряются в градусах дуги поля зрения. Размер дерева характеризуется углом  $\alpha$  в глазном яблочке. Различные комбинации высот  $h$  и расстояний  $d$  дают одинаковый визуальный размер в градусах угла обзора

Интуитивное предположение о том, что наши глаза в основном неподвижны, просто ошибочно. Почему кажется, что мир не перескакивает с одной фиксации на другую несколько раз в секунду? Просмотр фильма, в котором камера перескакивает с одного ракурса на другой 3–4 раза в секунду, будет довольно утомительным. Мозг принимает те сигналы сетчатки, которые меняются пару раз в секунду, создавая иллюзию стабильности. Кроме того, саккады – одно из самых быстрых движений человеческого тела, достигающее максимальной скорости до  $900^\circ$  угла обзора в секунду. Учитывая типичную саккаду, перемещающую фиксацию на 10 см за 20 мс, получится примерно 20 км/ч; максимальная скорость может превышать 160 км/ч. В течение нескольких десятков миллисекунд, когда глаза перемещаются из одного места в другое, сенсорные входные сигналы меняются так быстро, что во время этой саккады практически невозможно что-либо воспринять. Каждый раз, совершая саккаду, мы практически не воспринимаем сенсорные сигналы в течение нескольких десятков миллисекунд. Однако мы обычно не осознаем эти промежутки. В нашем мозге есть механизм подавления восприятия во время саккад, поэтому мы воспринимаем мир как стабильный. Моргания происходят быстрее, чем саккады, они происходят примерно 15 раз в минуту и обычно делятся около 100–200 мс. Информация не поступает в наши глаза более 100 мс, примерно 15 раз в минуту, и все же мы обычно не замечаем морганий, если не обратим на них особое внимание. Подавление восприятия саккад, моргания и стабильность визуального мира в то время, когда глаза прыгают с одного места на другое, представляют собой убедительные примеры, показывающие, что наше субъективное восприятие мира является результатом обработки исходной информации. Зрительный образ представляет собой интерпретацию, построенную нашим мозгом на основе поступающей сенсорной информации в сочетании с ожиданиями и нашими общими знаниями о мире. То, что мы видим, – это не копия того, что диктуют глаза.

Схема фиксаций взгляда зависит от самой картинки, ближайшей истории и текущих целей. Характеристики изображения влияют на движения глаз: например, высококонтрастные области более заметны и имеют тенденцию привлекать внимание глаз. История предыдущих фиксаций также имеет значение: в среднем наблюдатели склонны избегать возвращения в то место, где

недавно фиксировался их взгляд, – явление, известное как *ингибирование возвращения*. Текущие цели также играют решающую роль: если вы ищете свою машину на стоянке, вы, скорее всего, сделаете больше фиксаций на машинах и близлежащих объектах того же цвета, что и ваша машина.

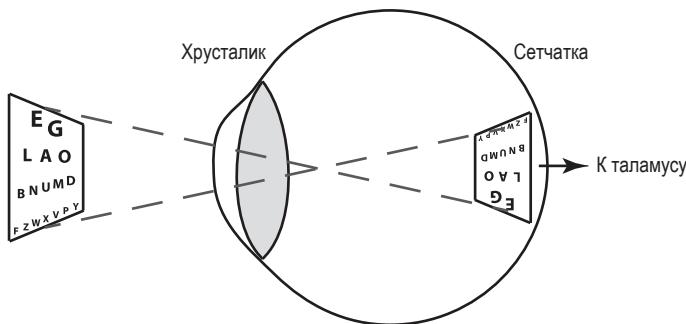
Если посмотреть на рис. 2.3, то в дополнение к баллистическим движениям глаз, охватывающим несколько градусов угла обзора и происходящим каждые 200–300 мс (саккады), есть также много других меньших и более быстрых движений глаз. Эти движения глаз называются *микросаккадами* и обычно охватывают доли зрительного градуса. Поскольку эти движения глаз происходят во время более или менее стабильной фиксации, их называют фиксационными движениями глаз. Большинство саккад являются непроизвольными (как отмечалось ранее, мы обычно даже не осознаем, что их делаем), но, конечно, мы можем произвольно управлять своими саккадами. В свою очередь, микросаккады тоже непроизвольны. Вместе с другими неподвижными движениями глаз эти небольшие изменения положения глаз могут играть решающую роль в предотвращении адаптации. Как мы увидим в разделе 6.9, при отсутствии какого-либо внешнего движения, движения головы или движения глаз нейроны быстро адаптируются к входным сигналам, снижая свою активность. На самом деле неожиданные эксперименты показали, что если изображение на сетчатке идеально стабилизировано – с помощью устройства, способного немного перемещать изображение для учета небольших движений глаз, – тогда изображение быстро перестает восприниматься. Другими словами, без постоянных движений глаз мы не смогли бы ничего видеть, кроме кратковременных изменений, связанных с движущимися объектами или движениями головы.

## 2.5. СЕТЧАТКА ИЗВЛЕКАЕТ ИНФОРМАЦИЮ ИЗ СВЕТА

Приключение обработки изображений в мозге начинается с преобразования фотонов в электрические сигналы в *сетчатке* глаза (retina). Благодаря своей доступности сетчатка является наиболее изученной частью зрительной системы. Преобразование света в электрические сигналы в сочетании с точными схемами обработки информации в сетчатке можно считать одним из величайших достижений эволюции. Способность преобразовывать световые узоры в информацию и строение глаза заставили Чарльза Дарвина (1809–1882) задуматься о том, могло ли произойти это чудо путем естественного отбора<sup>1</sup>. Изящные биохимические и электрофизиологические исследования смогли описать каскад передачи сигнала, ответственный за захват фотонов и использование энергии фотона для запуска ряда химических реакций, которые приводят к изменениям электрического потенциала в фоторецепторных клетках.

---

<sup>1</sup> Такие «чудесные» изобретения эволюции, происходящие в очень короткий в геологическом масштабе период времени, в биологии называются *ароморфозами*. Другой пример ароморфоза – возникновение фотосинтеза. – Прим. ред.



**Рис. 2.5 ♦** Хрусталик глаза переворачивает изображение.  
Как и во многих других типах линз, изображение переворачивается на 180°  
при фокусировке на сетчатку

Световая информация достигает глаза через линзу хрусталика. Когда свет достигает фокальной плоскости, совпадающей с положением сетчатки, изображение переворачивается (на 180°, рис. 2.5). Этот основной факт оптики иногда ставит в тупик тех, кто впервые задумывается о восприятии. Почему мы не видим все вверх ногами? Этот вопрос также мучил некоторые из самых ярких умов с тех пор, как были открыты основные принципы оптики. Не кто иной, как великий Леонардо да Винчи (1452–1519), ошибочно предположил, что мы не видим перевернутого изображения из-за второй линзы в глазу, снова инвертирующей изображение. Более того, Иоганн Кеплер (1571–1630), сыгравший центральную роль в развитии нашего знания о зрительном восприятии, четко описал переворот изображения в глазном яблоке и оставил проблему восприятия на совесть натурфилософов (в то время нечто среднее между физиками и философами). Другие философы предполагали, что новорожденные младенцы действительно видят объекты вверх ногами и что это восприятие в конечном итоге «корректируется» благодаря согласованию зрительных сигналов с осознанием. Эти философские идеи являются еще одним примером ошибочных интерпретаций, основанных на интроспективных моделях без привязки к реальным экспериментам: нет никаких доказательств того, что осознание необходимо для развития зрительной системы, способной определять верх и низ в окружающем мире.

Мы не видим объекты вверх ногами, потому что восприятие представляет собой реконструкцию внешнего мира нашим мозгом, основанную на схемах активности нейронов сетчатки. С того дня, как мы родились, наш мозг узнает, что определенный образец активации сетчатки – это то, как обстоят дела в мире. Мозг не знает, что находится вверху, а что внизу; это все электрические сигналы. Можно даже научить мозг приспособливаться к изображениям по другим правилам, надев очки, переворачивающие изображение. Приспособиться к таким очкам непросто, и это требует самоотверженности, но люди в очках с линзами, которые переворачивают мир вверх ногами, или в очках, которые зеркально отражают изображение слева направо, даже могут научиться ездить на велосипеде. После адаптации к этим новым правилам снятие очков вызывает неприятные ощущения; подопытным при-

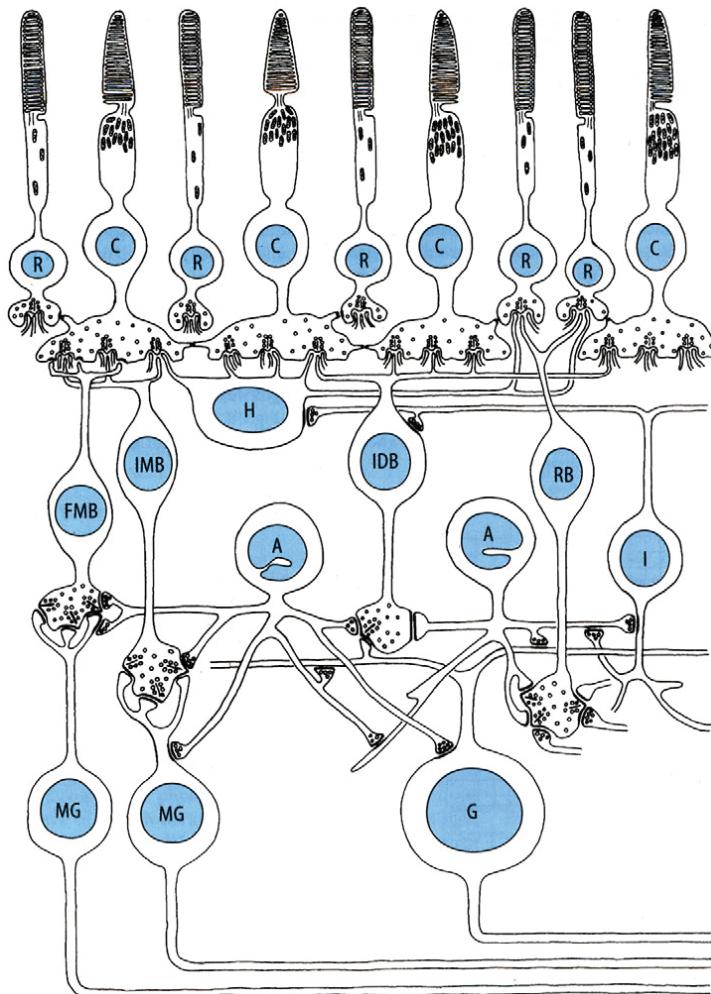
ходится заново учиться интерпретировать визуальный мир без переворота изображения. Сняв эти противные очки, заново научиться адаптироваться к естественному миру можно намного быстрее, чем длилась начальная тренировка по приспособлению мозга к перевернутому.

Сеть нейронов сетчатки – очень совершенная структура, которая завораживает нейробиологов уже более века. История исследований сетчатки тесно связана с историей нейробиологии и начинается с рисунков знаменитого Сантьяго Рамона-и-Кахаля (1852–1934). Сантьяго Рамон-и-Кахаль, которого считали отцом нейробиологии, умел рисовать и хотел стать художником. Однако у его родителей были другие планы; Рамон-и-Кахаль в конечном итоге последовал их совету и стал врачом. Получив медицинскую степень, он изучал методы окрашивания нервной ткани у великого Камилло Гольджи (1843–1926), с которым он участвовал в жестоком научном споре о фундаментальной структуре ткани мозга и с которым он разделил Нобелевскую премию 1906 г.

Вскоре сетчатка стала постоянной страстью Рамона-и-Кахаля. Сетчатка расположена на задней части глаза; у людей она имеет толщину примерно 250 мкм и охватывает площадь поверхности, соответствующей площади полусфера диаметром примерно 2,5 см. Сетчатка является частью центральной нервной системы: она происходит из тех же эмбриональных структур, которые дают начало остальной части мозга, и она обладает гематоэнцефалическим барьером, аналогичным такому же у остальной части мозга.

Схематическая диаграмма сетчатки на рис. 2.6 иллюстрирует типичную схему связи, состоящую из трех основных клеточных слоев (фоторецепторы, биполярные клетки и ганглиозные клетки), соединенных между собой двумя дополнительными промежуточными слоями (горизонтальные клетки и амакриновые клетки). У позвоночных животных свет должен пройти через все другие типы клеток, чтобы добраться до фоторецепторов, показанных вверху на рис. 2.6. Фоторецепторы бывают двух основных типов: палочки и колбочки. Всего у человека около  $10^8$  палочек; эти клетки очень чувствительны к свету и предназначены для улавливания фотонов в условиях низкой освещенности. Ночное видение зависит от палочек. Поскольку колбочки имеют широкую спектральную чувствительность, которая позволяет определять цвет, и поскольку колбочки гораздо менее чувствительны к слабому освещению, чем палочки, мы почти не видим цвета ночью. Палочки настолько чувствительны, что могут захватывать одиночный фотон видимой части спектра, энергия которого составляет всего лишь порядка  $10^{-19}$  Дж, и передавать его энергию дальше. Тщательные эксперименты достоверно показывают, что иногда люди могут регистрировать одиночные фотоны.

Помимо палочек, имеется около  $10^7$  колбочек, предназначенных для зрения в условиях яркого света. У большинства людей есть три типа колбочек: длинноволновые, чувствительные к длине волн света с максимумом  $\sim 560$  нм, средневолновые, чувствительные к диапазону с максимумом  $\sim 530$  нм, и коротковолновые, чувствительные к диапазону с максимумом  $\sim 420$  нм. Цветовое зрение зависит от активности колбочек. Некоторые люди страдают тем или иным видом цветовой слепоты, или дальтонизма, – в большинстве случаев из-за недостатка или даже полного отсутствия одного из



**Рис. 2.6 ♦ Схематическая диаграмма типов клеток и коммуникаций в сетчатке приматов.** На этой диаграмме свет приходит снизу и проходит через все слои, достигая фоторецепторов. R – фоторецепторы-палочки; C – фоторецепторы-колбочки; FMB – гладкие сверхмалые биполярные клетки; IMB – инвагинирующие сверхмалые биполярные клетки; H – горизонтальные клетки; IDB инвагинирует диффузные биполярные клетки; RB – стержневые биполярные клетки; I – внутренняя плексиформная клетка; A – амакриновые клетки; G – ганглиозные клетки; MG – сверхмалые ганглиозные клетки. Воспроизведено с разрешения Доулинга, 2012

этих типов колбочек; в редких случаях может отсутствовать более одного типа колбочек. Даже с двумя типами колбочек люди все равно могут видеть разные оттенки. Например, если людям не хватает коротковолновых колбочек, они все равно могут различать свет с длиной волны 400 нм и 500 нм из-за дифференциальных откликов, возникающих в колбочках, чувствительных к длинным и средним длинам волн. Таким образом, цветовая слепота – не-

правильное название, и его следует использовать только для людей, которые видят в оттенках серого, то есть людей, которые чувствительны лишь к интенсивности света без какого-либо цветового восприятия. Состояние, известное как *ахроматопсия*, вызванное повреждением мозга, но не глаза, может привести к полной цветовой слепоте, что блестяще описал известный британский невролог и писатель Оливер Сакс (1933–2015) в одной из своих книг (раздел 4.8).

Люди, которым не хватает одного типа колбочек, имеют определенные точки путаницы, то есть определенные комбинации длин волн, которые они не могут различить. Чтобы продемонстрировать эти цвета, которые они не могут различить, очень важно выровнять интенсивность света. В естественных условиях цвета часто коррелируют с различной интенсивностью, и, следовательно, люди с недостатком колбочек могут использовать эти признаки интенсивности, чтобы обойти свое пониженное разрешение в цветовом спектре. Тест *Исхара* – это распространенный способ оценки цветового дефицита, и в интернете есть множество таких тестов. Многие люди удивляются, когда проходят эти тесты и обнаруживают, что не могут различить определенные цветовые комбинации. Дефицит цветового зрения на самом деле довольно часто встречается у мужчин (примерно один из 12!) и гораздо меньше распространен у женщин (примерно у одной из 200). Политически некорректная шутка гласит, что женщины знают сотни цветов, а мужчины – только пять. Эта шутка не так уж и неверна для некоторых мужчин (хотя, строго говоря, даже с двумя типами колбочек можно различить много разных цветов).

Палочки и колбочки неравномерно распределены по сетчатке. В частности, есть область сетчатки, называемая *центральной ямкой* (*fovea*), область вокруг которой характеризуется особо высокой остротой зрения<sup>1</sup>. Область fovea имеет размер ~300 мкм, не содержит палочек и имеет высокую плотность колбочек с фантастическим общим числом – 17 500 колбочек. Эта высокая плотность приводит к высокому разрешению поля зрения, тем самым обеспечивая лучшее различение объектов в точке фиксации. Например, наша способность читать зависит именно от этой ямки: попробуйте фиксировать взгляд на букве «R» во второй строке на рис. 2.7. Затем попробуйте прочитать слово, которое находится на расстоянии пяти слов и двух строк ниже буквы «R», не двигая глазами. Плотность клеток и степень конвергенции от конусов к нижележащим нейронам уменьшается с эксцентризитетом, то есть с удалением от ямки. Кроме того, оптика хрусталика глаза имеет наилучшую передачу контраста именно в ямке. Из-за оптики глаза и неоднородности разрешающей способности сетчатки мы видим с высоким разрешением только в области центральной ямки (рис. 2.8В). Следовательно, саккады глаз позволяют точно направить центр фиксации для получения подробной информации. У людей с дегенерацией желтого пятна все больше и больше поражается фoveальная область, что приводит к ухудшению качества информации с высоким разрешением и в конечном итоге к появлению шумной<sup>2</sup> или размытой версии изображения (рис. 2.8С).

<sup>1</sup> Область вокруг центральной ямки называется желтым пятном. – Прим. ред.

<sup>2</sup> Имеется в виду «визуальный шум». – Прим. ред.

Reading depends strictly on foveal resolution. Try to fixate on the letter “R” shown here in large bold font. Make sure that you do not move your eyes away from the R. If you do, then your high resolution area rapidly shifts to whatever location you are fixating on. Once you are fixating, try to read a word that is four lines below the letter “R”. This task is basically impossible for us because the resolution drops sharply outside of the fovea. The notion that we can capture the entire visual scene at high resolution is merely an illusion created by our rapid eye movements and the fact that whenever we land on a particular location, it appears in high resolution!

**Рис. 2.7** ♦ Мы можем читать только тогда, когда изображение проецируется на центральную ямку (фовеальную область). Обратите внимание на большую букву R, выделенную жирным шрифтом, во второй строке и попробуйте прочитать слова в другой строке, не двигая глазами



**Рис. 2.8** ♦ В высоком разрешении мы видим только область вокруг фиксации внимания. (A) Оригинальная фотография. Если бы вы здесь находились, фиксируясь на месте, обозначенном знаком +, у вас возникла бы иллюзия, что все поле заполнено деталями. (B) Однако изображение, передаваемое сетчаткой в мозг, ближе к изображению A, с высоким разрешением в месте фиксации и все более размытым по направлению к периферии. Воспринимаемое нами изображение кажется ближе к A, чем к B, потому что мы постоянно двигаем глазами, выбирая новые места и рассматривая их с высоким разрешением. (C) Люди с дегенерацией желтого пятна видят шум или размытое изображение в центре в дополнение к обычной размытости периферии

Несмотря на то что в местах, удаленных от желтого пятна, разрешение уже, у нас есть иллюзия восприятия приблизительно одинакового разрешения во всем визуальном поле. Причиной этой иллюзии отчасти является движение глаз: каждый раз, когда мы передвигаем взгляд, мы фиксируемся на новом месте, которое появляется в высоком разрешении. Поэтому мы естественным образом предполагаем, что все поле зрения имеет одинаковое разрешение. Кроме того, очевидно, в мозге сохраняется информация о предыдущих фиксациях. Когда мы перемещаем взгляд в новое место, старое место фиксации теперь появляется на периферии с более низким разре-

шением. Однако версия с низким разрешением может быть ассоциирована с версией, хранящейся в рабочей памяти, о предыдущей фиксации с высоким разрешением.

В задней части каждого глаза есть область, в которой нет фоторецепторов. В этой области из глаза выходят аксоны выходных клеток сетчатки, ганглиозные клетки сетчатки (RGC). Люди не могут обнаружить свет, сфокусированный на этой области, соответственно, она называется *слепым пятном*. Самый простой способ обнаружить слепое пятно – закрыть один глаз, сосредоточиться на заданном удаленном месте и медленно перемещать указательный палец от центра к периферии до тех пор, пока часть пальца не исчезнет из поля зрения (но не полностью, что может подразумевать, что вы переместили палец полностью за пределы поля зрения). В интернете есть множество демонстраций, которые помогут обнаружить слепое пятно. Легенда гласит, что король Англии Карл II имел обыкновение развлекаться, перемещая взгляд так, что голова заключенного, приговоренного к обезглавливанию, оказывалась в слепом пятне; увидеть его без головы перед казнью он находил занимательным.

При нормальных обстоятельствах мы не осознаем слепое пятно, т. е. у нас есть субъективное ощущение, что мы можем видеть все поле зрения (даже с одним закрытым глазом). Учитывая, что мы обычно не воспринимаем слепое пятно, можно предположить, что на самом деле оно довольно мало. Однако в слепой зоне может поместиться проекция девяти полных лун! Как возможно полностью игнорировать такую большую пустую область зрительного поля? Обычно мы не осознаем это, потому что мозг заполняет и компенсирует отсутствие рецепторов в слепом пятне. Этот процесс заполнения еще раз подчеркивает то, что наши зрительные образы не являются буквальным отражением реальности, а скорее реконструкцией, придуманной нашим мозгом. Мы вернемся к понятию зрения как субъективной реконструкции, когда будем обсуждать зрительные иллюзии (раздел 3.1) и визуальный образ мира (гл. 10).

Информация от фоторецепторов передается на второй клеточный слой, состоящий из горизонтальных клеток, bipolarных клеток и амакриновых клеток, и, наконец, в ганглиозные клетки сетчатки (RGC). Сетчатка человека содержит примерно 6,4 млн колбочек, около 110 млн палочек и около миллиона ганглиозных клеток сетчатки. Таким образом, в среднем происходит конвергенция около 100 фоторецепторов к одной ганглиозной клетке, но это количество варьируется в зависимости от местоположения ее в сетчатке. Как отмечалось ранее, конвергенция минимальна в центральной ямке и более обширна на удаленной периферии. В ямке одна колбочка находится выше одной клетки RGC, а на периферии имеется около 15 колбочек на одну RGC и сотни палочек на одну RGC.

Рисунок 2.6 показывает упрощенную схему связи в сетчатке от фоторецепторов к горизонтальным и bipolarным клеткам, а затем к амакриновым и ганглиозным клеткам. Молекулярные и анатомические маркеры помогли определить различные типы горизонтальных и bipolarных клеток и еще больше типов амакриновых и ганглиозных клеток, каждый из которых участвует в специальных вычислениях для захвата различных характеристик вхо-

дящих изображений. Более того, серийная электронная микроскопия начинает прояснять коннектом сетчатки – то есть точную схему синаптических связей в сетчатке. Возможно, в недалеком будущем мы получим довольно полную анатомическую карту сетчатки.

## 2.6. ТРЕБУЕТСЯ ВРЕМЯ, ЧТОБЫ ИНФОРМАЦИЯ ДОСТИГЛА ЗРИТЕЛЬНОГО НЕРВА

На первый взгляд, зрение кажется мгновенным. Мы открываем глаза, и мир немедленно проявляется во всей своей красе. Однако мгновенного распространения сигнала не существует. Требуется время для каскада процессов, которые преобразуют поступающие фотоны в импульсную активность ганглиозных клеток сетчатки. Латентность реакции ганглиозных клеток сетчатки на вспышку стимула зависит от множества факторов, включая, среди прочего, предыдущую историю визуальной стимуляции, интенсивность вспышки стимула, его размер и цвет.

Аксоны ганглиозных клеток сетчатки, которые передают информацию остальной части мозга, вместе известны как зрительный нерв. В среднем от начала вспышки стимула требуется 30–50 мс, чтобы импульсы вышли из зрительного нерва и распространились вниз по остальному мозгу. Эта задержка дополнительно сочетается с вычислительным временем, необходимым для интерпретации информации в мозге, о чем будет подробнее рассказано в разделе 5.12. Из-за этих задержек то, что мы видим, отражает то, что происходило в мире в ближайшем прошлом. Задержки достаточно короткие, чтобы обмануть наше восприятие и позволить нам получить быструю оценку того, что происходит в мире.

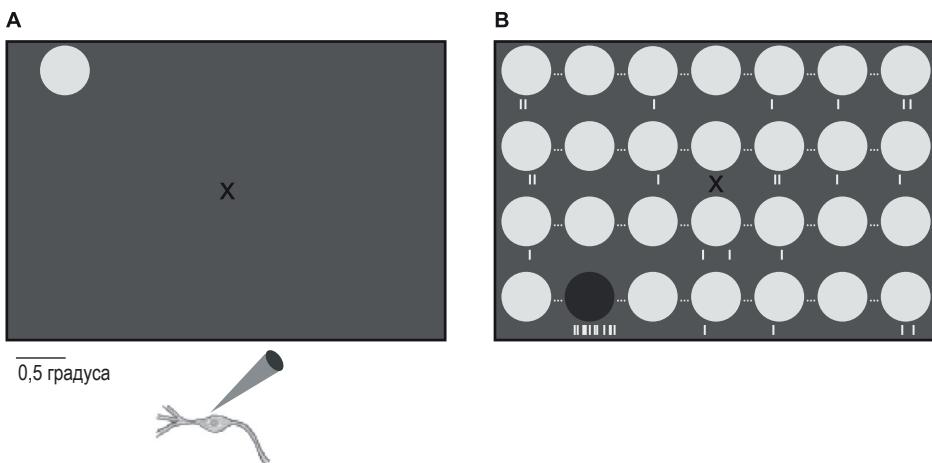
## 2.7. ЗРИТЕЛЬНЫЕ НЕЙРОНЫ РЕАГИРУЮТ НА ОПРЕДЕЛЕННУЮ ОБЛАСТЬ ПОЛЯ ЗРЕНИЯ

Как и большинство нейронов головного мозга, ганглиозные клетки сетчатки (RGC) передают информацию, испуская потенциалы действия<sup>1</sup>, также известные как спайки<sup>2</sup>. Клетки перед RGC в сетчатке составляют исключение из этого правила и обмениваются данными, используя градиентные сигналы напряжения, не испуская спайков. Чтобы понять, как RGC представляют визуальную информацию, нам нужно изучить, как различные входные данные отображаются в спайковых ответах. Функциональные свойства RGC широко исследовались с помощью электрофизиологических регистраций, которые

<sup>1</sup> Волны возбуждения, распространяющиеся по мембране клетки. – Прим. ред.

<sup>2</sup> Кратковременные колебания потенциала, регистрируемого электродом, в форме острого пика. – Прим. ред.

восходят к выдающимся работам Халдана Хартлайна (1903–1983), Горация Барлоу (1921–2020) и Штефана Куффлера (1913–1980). RGC (как и большинство нейронов зрительной коры) наиболее сильно реагируют на ограниченную область зрительного поля, называемую *рецептивным полем* (рис. 2.9). Рецептивное поле можно картировать, если давать стимул в разных местах и разных площадях, чтобы найти области, которые активируют нейроны. Нейроны также имеют тенденцию к спонтанному возбуждению, так что возникают небольшие нейронные ответы, даже когда сетчатка находится в полной темноте или когда стимул находится очень далеко от рецептивного поля. Другими словами, частота срабатывания нейронов не обязательно равна нулю при отсутствии зрительной стимуляции в пределах рецептивного поля. Следует подчеркнуть, что расположение рецептивного поля всегда определяется относительно точки фиксации, а не относительно фиксированного положения в пространстве. Если наблюдатели двигают глазами, положение рецептивного поля в окружающей среде изменяется, но положение относительно точки фиксации – нет.



**Рис. 2.9 ♦** Нейроны имеют локализованные рецептивные поля. (A) Световой стимул (белый кружок) вспыхивает в определенном месте во время регистрации активности нейрона фиксирующего животного («Х» обозначает точку фиксации). (B) Процедура повторяется в нескольких разных местах. Маленькие вертикальные отметки обозначают активность нейронов. Место максимальной активности (черный кружок) обозначает рецептивное поле нейрона. Размер стимула тоже изменяется, чтобы отобразить границы рецептивного поля. Нейрон также показывает низкую спонтанную скорость в других местах

Эти рецептивные поля покрывают все визуальное поле. Не двигая глазами, любое место в визуальном поле, где вы можете что-либо увидеть, подразумевает наличие RGC с рецептивным полем, охватывающим это место. Восприимчивые поля RGC организованы топографически – то есть соседние RGC в сетчатке представляют близлежащие места в визуальном поле.

Эта топография сохраняется в проекциях RGC на таламус, а оттуда на кору. Неравномерное распределение нейронов от ямки к периферии означает, что существует постоянная зависимость размера рецептивных полей от эксцентризитета. В центральной ямке существует взаимно однозначное соответствие между колбочками и RGC. Рецептивные в желтом пятне самые маленькие, а размеры рецептивных полей растут примерно линейно с эксцентризитетом. Большие восприимчивые поля на периферии – одна из основных причин, по которым у нас меньшее разрешение за пределами желтого пятна.

RGC, схематически показанный на рис. 2.9, увеличивает свою частоту срабатывания нейронов с увеличением яркости внутри воспринимающего поля. Этот тип клеток называется *клетки с on-центром*. Существуют также другие RGC, *клетки с off-центром*, которые увеличивают спайковую частоту при уменьшении яркости в центре их рецептивных полей.

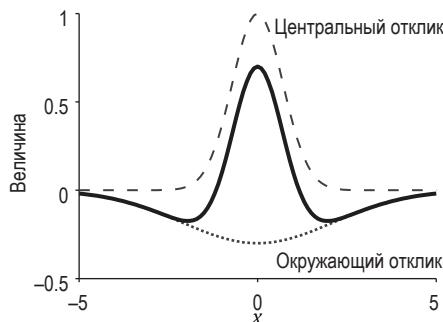
Активность RGC не отражает напрямую узор света, поступающий на сетчатку из-за искажений, вносимых хрусталиком глаза, из-за временных задержек и промежуточной обработки, вносимой предыдущими клеточными слоями, а также из-за зависимых от эксцентризитета вариаций конвергенции от фоторецепторов к RGC. Тем не менее можно сделать обоснованное предположение о поступающих визуальных стимулах, исследуя ответы RGC. У нас нет инструментов для регистрации активности каждого RGC. Современные технологии позволяют одновременно регистрировать деятельность лишь нескольких сотен RGC. Даже с такой небольшой популяцией можно восстановить довольно точную форму световых узоров, достигающих сетчатки.

## 2.8. ОПЕРАТОР РАЗНОСТИ ГАУССИАН ИЗВЛЕКАЕТ ВАЖНУЮ ИНФОРМАЦИЮ И ИГНОРИРУЕТ ОДНОРОДНЫЕ ОБЛАСТИ

Даже когда центр клетки с off-центром подвергается бомбардировке вспышкой высокой яркости, ее реакция будет модулироваться тем, что находится за пределами центра рецептивного поля. В частности, для большинства RGC идеально однородная белая стена с высокой яркостью не вызовет высокой активации. Рассмотрим следующий эксперимент: в центре рецептивного поля показан маленький однородный белый кружок, и нейрон возбуждается выше исходного уровня (рис. 2.9). Затем круг немного увеличивается, и нейрон показывает более высокую частоту срабатывания. Если мы продолжаем увеличивать размер круга, в какой-то момент частота достигает предельного значения. Дальнейшее увеличение круга ведет к снижению частоты срабатывания; это явление известно как *ингибиование окружения* (торможение окружения). Ингибиование окружения наблюдается не

только для RGC; оно распространено во всей зрительной системе. Нейроны с оп-центром особенно заинтересованы в пространственных изменениях, т. е. увеличении яркости внутри рецептивного поля в сочетании с уменьшением яркости вне рецептивного поля. Для нейронов с off-центром верно обратное.

Эта форма пространственного контекстно-зависимого типа реакции известна как центр-периферийные рецептивные поля и обычно моделируется как разность двух гауссовых кривых (рис. 2.10). Рассматривая клетку и предполагая, что центр рецептивного поля находится в местоположении  $x = 0, y = 0$ , активность нейронов в ответ на освещение в новой позиции  $x, y$  будет управляться возбуждающим компонентом, пропорциональным  $(1/2\pi\sigma_{cen}^2)e^{-(x^2+y^2)/2\sigma_{cen}^2}$ , где  $\sigma_{cen}$  отражает пространственную протяженность возбуждающего воздействия (пунктирная линия на рис. 2.10). Это возбуждение уравновешивается объемным торможением окружения, выраженным как  $(1/2\pi\sigma_{sur}^2)e^{-(x^2+y^2)/2\sigma_{sur}^2}$ , где  $\sigma_{sur}$  отражает пространственную протяженность тормозящего воздействия (пунктирная линия на рис. 2.10). Оператор разности гауссиан используется для описания структуры рецептивного поля RGC:



**Рис. 2.10** ♦ Рецептивное поле гауссианы в форме мексиканской шляпы. Рецептивное поле ганглиозных клеток сетчатки часто характеризуется как разница между центральной реакцией (пунктирная линия) и более широкой и слабой реакцией окружения (пунктирная линия), что приводит к форме «мексиканской шляпы» (сплошная линия)

$$D(x, y) = \pm \left( \frac{1}{2\pi\sigma_{cen}^2} e^{-\frac{x^2+y^2}{2\sigma_{cen}^2}} - \frac{B}{2\pi\sigma_{sur}^2} e^{-\frac{x^2+y^2}{2\sigma_{sur}^2}} \right), \quad (2.3)$$

где масштабирующий коэффициент  $B < 1$  контролирует относительную силу возбуждения и торможения, где  $\sigma_{sur} > \sigma_{cen}$  и где  $\pm$  соответствует клеткам с оп-центром и off-центром соответственно. Разница между двумя функциями дает кривую в форме «мексиканской шляпы» с пиком в центре и падением на краях. Биология полна неожиданностей и исключений. Отклики некоторых RGC не могут быть объяснены уравнением (2.3).

## 2.9. ЗРИТЕЛЬНЫЕ НЕЙРОНЫ РЕАГИРУЮТ НА ИЗМЕНЕНИЯ

Точно так же, как большой пространственно-однородный стимул не вызывает сильной активации, потому что нейроны настроены на обнаружение пространственных изменений, временные изменения тоже критичны. Постоянный стимул обычно не приводит к устойчивым нейронным ответам. Некоторые RGC реагируют на начало стимула, другие реагируют на спад, а третья реагируют на начало и спад. Во всех этих случаях реакция нейронов имеет тенденцию быстро адаптироваться, когда стимул остается постоянным и при отсутствии каких-либо других внешних изменений (при отсутствии движений глаз или головы). Некоторые нейроны поддерживают постоянный отклик выше базовой линии во время действия стимула после начального переходного процесса. У других частота возбуждения нейронов снижается до исходного уровня после начального переходного процесса. Следовательно, RGC чувствительны не только к пространственному, но и к временемному контексту.

Использование контекстной информации позволяет нейронам эффективно кодировать пространственные и временные изменения, не тратя обильных и энергетически дорогостоящих всплесков (спайков), чтобы поддерживать отклик на стимул, постоянный в пространстве или времени. Таким образом, закономерности в структуре визуального стимула, описанные в разделах 2.1 и 2.2, отражаются в свойствах возбуждения RGC.

Уравнение (2.3) может быть обобщено для количественного описания динамических реакций ганглиозных клеток сетчатки при предъявлении стимула, который начинается при  $t = 0$  и остается постоянным:

$$D(x, y, t) = \pm \left( \frac{D_{cen}(t)}{2\pi\sigma_{cen}^2} e^{-\frac{x^2+y^2}{2\sigma_{cen}^2}} - \frac{BD_{sur}(t)}{2\pi\sigma_{sur}^2} e^{-\frac{x^2+y^2}{2\sigma_{sur}^2}} \right), \quad (2.4)$$

где  $D_{cen}(t) = \alpha_{cen}^2 te^{-\alpha_{cen}t} - \beta_{cen}^2 te^{-\beta_{cen}t}$  описывает динамику центральной функции возбуждения, а  $D_{sur}(t) = \alpha_{sur}^2 te^{-\alpha_{sur}t} - \beta_{sur}^2 te^{-\beta_{sur}t}$  описывает динамику тормозящей функции окружения.

Уравнение (2.4) описывает внутреннюю динамику RGC при получении стимула, который остается постоянным. В дополнение к этим типам реакций некоторые RGC также сильно активируются стимулом, который движется в пределах рецептивного поля. Одним из таких типов клеток является *направленно-избирательный on-off RGC*, который демонстрирует усиленные ответы, когда стимул в пределах рецептивного поля движется в определенном направлении. Такие избирательные по направлению отклики также модулируются окружением: нейроны реагируют наиболее энергично, когда есть разница в направлении движения между рецептивным полем и его окружением. Все поле зрения, движущееся в одном направлении, является слабым стимулом для нейронов этого типа. Это контекстное вычитание помогает нейронам отличать движение внешних объектов от собственного движения.

Кроме того, положение границ поля зрения также приводит к прерываниям движения во время собственного движения относительно статической сцены. Чувствительные к движению RGC обычно имеют распространенные дендритные ветви и особенно многочисленны на периферии. Из-за этого легче обнаружить движущийся объект на периферии поля зрения, – наблюдение, которое вы можете легко проверить, если зафиксируете взгляд на любой букве этого текста, протянете руку на периферию поля зрения и сравните воспринимаемый вами образ руки, когда она неподвижна и когда она движется.

Скорости проводимости RGC-клеток были использованы для разделения *магнотеллюлярных* («крупных») клеток (RGC M-типа) и *парвотеллюлярных* клеток (RGC P-типа, также называемых сверхмалыми клетками). Клетки M-типа имеют распространенные дендритные ветви, высокую скорость проводимости, реагируют на слабоконтрастные стимулы, демонстрируют скоротечные ответы с быстрым переходным процессом (затуханием. – Прим. ред.) и имеют небольшую чувствительность к цвету. Напротив, клетки P-типа имеют небольшие дендритные ветви, обладают цветовой чувствительностью и, как правило, демонстрируют более устойчивые ответы и низкие скорости проводимости.

Продолжаются захватывающие исследования, направленные на выявление всех различных типов функциональных и структурных специализаций RGC; современные оценки предполагают, что существует десятки, если не больше, различных типов ганглиозных клеток, в зависимости от того, как точно определить «тип». За исключением центральной ямки, различные типы ганглиозных клеток распределены повсюду примерно одинаково, так что во всем поле зрения могут быть восприняты одни и те же характеристики внешнего стимула.

## 2.10. ДВИЖЕМСЯ ДАЛЬШЕ, К ДРУГИМ ОТДЕЛАМ МОЗГА

Основным местом назначения выходного сигнала ганглиозных клеток сетчатки является часть таламуса, называемая латеральным коленчатым ядром (LGN). Сетчатка также проецируется на супрахиазматическое ядро и верхний холмик среди многих других областей (анатомические исследования картировали более 40 областей мозга, которые получают сигналы от сетчатки). Супрахиазматическое ядро играет жизненно важную роль в регулировании циркадных ритмов, в то время как верхний бугорок представляет собой главный центр обработки изображений для многих видов животных, до эволюционного расширения коры головного мозга. Приматы могут распознавать объекты после повреждений верхнего холмика, но не после повреждений зрительных областей коры. Следовательно, главный путь зрительного сигнала – это путь от RGC до LGN и к коре головного мозга.

Как мы обсудим в разделах 5.17 и 6.11, по всей зрительной системе есть обширные *обратные проекции* (рис. 1.5). Если область А проецируется на область В, то в большинстве случаев область В также проецируется обратно в область А. Одним из немногих исключений из этого правила является со-

единение сетчатки глаза с LGN. Нет никаких соединений от LGN обратно к сетчатке. Следовательно, пути от фоторецепторов к клеткам RGC и LGN можно рассматривать как в основном прямые.

Таламус часто кратко называют воротами в кору, которые определяют, какой тип сенсорной информации ее достигнет. Рецептивные поля клеток LGN также демонстрируют центро-периферийную структуру, изображенную на рис. 2.10, и могут быть аппроксимированы уравнениями (2.3) и (2.4). Клетки таламуса нередко довольно несправедливо называют *ретрансляционными клетками*, утверждая идею о том, что таламус просто копирует выходной сигнал RGC и передает его в кору.

Одно очевидное различие между клетками RGC и LGN – это структура их связей. Хотя мы часто думаем о LGN преимущественно с точки зрения входных данных от RGC, существует большое количество обратных проекций из различных областей коры, преимущественно из первичной зрительной коры, на LGN. Не совсем понятно, как именно эти обратные связи модулируют реакцию на зрительные стимулы в LGN.

Как подавляющее большинство структур мозга, существуют две копии LGN, по одной в каждом полушарии. Правый LGN получает входные данные от обоих глаз, но только из левой полусфера поля зрения (в основном это часть поля зрения слева от точки фиксации), в то время как обратное верно для левого LGN. Правый глаз получает информацию из обоих полушарий и отправляет информацию правого полушария в LGN в левом полушарии, а информацию из левого полушария – в LGN в правом полушарии.

В LGN можно выделить шесть слоев. Слои 2, 3 и 5 получают *инсилатеральный* сигнал (т. е. информацию от глаза с той же половины тела). Слои 1, 4 и 6 получают *контралатеральный* сигнал (т. е. информацию от глаза на противоположной стороне тела). Таким образом, одна точка в пространстве представлена на шести различных картах на уровне LGN. Информация от правого и левого глаз не сливается в LGN. Слои 1 и 2 называются магноклеточными слоями и принимают входные данные от RGC M-типа. Слои 3–6 называются парвоцеллюлярными слоями и принимают входные данные от RGC P-типа. В LGN человека около 1,5 млн клеток. Таким образом, общая плотность нейронов LGN, распределенных по разным частям зрительного поля, сравнима с таковой в RGC, тогда как количество нейронов значительно увеличивается по мере того, как мы движемся к коре головного мозга.

Поскольку LGN и таламус в целом связаны с несколькими областями коры головного мозга, он занимает довольно уникальное положение для интеграции сенсорных входов с различными формами обрабатываемой информации по всей коре. Описание LGN как ретрансляционной структуры является грубым упрощением, и представление о LGN кардинально изменится по мере того, как мы больше узнаем о нейронных цепях и вычислениях в LGN.

## 2.11. ЦИФРОВАЯ КАМЕРА VS ГЛАЗ

В гл. 7–9 мы рассмотрим вычислительные модели обработки изображений. По большому счету, современные компьютерные модели начинаются с выво-

да обычной цифровой камеры, которая захватила изображение и представляет его как двумерную матрицу пикселей, каждый из которых закодирован в формате трехмерного цветового пространства (например, яркость красной, зеленои и синей составляющих цвета каждого пикселя). Однако даже лучшая современная цифровая камера пока несопоставима с сетчаткой по сложности серий вычислений.

Угол обзора цифровой камеры зависит от фокусного расстояния объектива. Для фокусного расстояния 17 мм (приблизительное расстояние от оптического центра хрусталика глаза до сетчатки) поле зрения составляет примерно 90°, тогда как поле зрения человека составляет почти 180°. По оценкам, разрешение человеческого глаза составляет порядка 500 мегапикселей, что намного больше, чем у некоторых из самых модных имеющихся в продаже цифровых фотоаппаратов.

Еще одно отличие состоит в том, что цифровые фотоаппараты примерно одинаковы по своей светочувствительности. Однако сетчатка гораздо лучше, чем самые лучшие современные камеры, в обработке изображения, полученного в условиях слабого освещения. Если вы когда-либо пробовали делать снимки ночью, вы, скорее всего, поняли, что делать цифровые снимки в условиях низкой освещенности непросто. Чтобы обойти эти проблемы, фотографы используют такие приспособления, как штативы, для стабилизации камеры и оставляют затвор камеры открытый на много секунд, минут или больше. Напротив, глаз может передавать точную информацию и помогать нам ориентироваться в лесу даже при свете звезд. Нам бы не хотелось ждать много секунд или минут, прежде чем мы сможем что-либо увидеть ночью. Один из приемов достижения этого заключается в том, что сетчатка может адаптироваться к условиям низкой освещенности и изменять свое усиление для достижения более высокой чувствительности. Глаз должен работать в условиях от сильного солнечного света вплоть до безлунных ночей, разница в яркости света которых составляет около девяти порядков. Эта адаптация требует времени, что можно ощутить, если выйти из темного места на солнечный свет или наоборот.

В дополнение к этой адаптации к средней освещенности яркость света может варьироваться в пределах трех единиц логарифмической шкалы уровней в пределах сцены. Сетчатка может приспособиться к этому благодаря механизмам адаптации, охватывающим различные пространственные и временные масштабы. Напротив, делать цифровые снимки в сцене с такой значительной разницей освещенности сложно: либо одна часть изображения окажется полностью темной, либо другая часть изображения окажется сильно переэкспонированной.

В цифровых камерах обычно отсутствуют многие сложные механизмы обнаружения движения и контекстной коррекции, описанные в этой главе для RGC. Для нас изображения редко бывают размытыми, тогда как в цифровых камерах необходимо реализовать множество дополнительных механизмов коррекции для получения четких изображений. Еще одно поразительное отличие заключается в том, как мы компенсируем спектральный состав источника света: мы никогда не видим, как это случается, на этих оранжевых фотографиях, которые зачастую дают нам цифровые камеры. Однако наи-

более разительным различием между биологическим зрением и цифровыми камерами является наличие чрезвычайно сложного вычислительного устройства для обработки выходных данных RGC, коры головного мозга, которые мы начнем изучать дальше.

## 2.12. ОБЗОР ГЛАВЫ

- Естественные изображения особенные: они пространственно-гладкие и медленно меняются во времени. Гипотеза эффективного кодирования утверждает, что нейронные ресурсы распределяются оптимально, чтобы соответствовать статистике входных данных окружающей среды.
- Положение и размеры в поле зрения измеряются в градусах угла обзора. Один градус примерно соответствует размеру вашего большого пальца на расстоянии вытянутой руки. Люди и другие приматы совершают частые движения глаз, называемые *саккадами*, охватывающие несколько градусов угла зрения и происходящие от трех до четырех раз в секунду.
- Фоторецепторы двух типов преобразуют свет в электрические сигналы для зрительного восприятия: *палочки* и *колбочки*. Палочки отвечают в первую очередь за ночное видение, а колбочки – за цветовое зрение при хорошем освещении.
- Ганглиозные клетки сетчатки передают сигнал к остальной части мозга.
- Ганглиозные клетки сетчатки отвечают за локализованную область зрительного поля, именуемую *рецептивным полем*.
- Центр фокуса проецируется на ямку (*fovea*), область, населенную колбочками, с более высокой плотностью клеток и меньшими размерами воспринимающего поля, что обеспечивает высокое разрешение проецируемого изображения.
- Центральные ганглиозные клетки сетчатки возбуждаются светом в пределах их рецептивного поля и подавляются светом в окружающей области. Их отклики можно описать функцией разности гауссиан.
- Информация от ганглиозных клеток сетчатки передается в латеральное коленчатое ядро таламуса.
- В грубом приближении глаз можно рассматривать как специализированную цифровую камеру, хотя глаза способны на многие сложные трюки, которые современные цифровые камеры выполнить не могут.
- Зрительный образ – это реконструкция, интерпретация, сделанная мозгом, вдохновленная сенсорным сигналом, но не буквальное отражение внешнего мира.

## ЛИТЕРАТУРА

См. <http://bit.ly/3aeW07Z> для получения дополнительных ссылок.

- Barlow, H. (1972). Single units and sensation: a neuron doctrine for perception. *Perception* 1, 371–394.

- Helmstaedter, M.; Briggman, K. L.; Turaga, S. C.; Jain, V.; Seung, H. S.; and Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* 500, 168–174.
- Kuffler, S. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology* 16, 37–68.
- Simoncelli, E.; and Olshausen, B. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience* 24, 193–216.
- Yarbus, A. (1967). *Eye movements and vision*. New York: Plenum Press.

Дополнительный контент вы можете скачать по ссылке <http://bit.ly/3aeW07Z>.

# Глава 3

---

## Феноменология зрения

Мы хотим понять нейронные механизмы, ответственные за зрительное восприятие, и мы хотим реализовать эти механизмы в вычислительных алгоритмах, которые подобны человеческим возможностям, а возможно, даже превосходят их. Чтобы построить такие машины с визуальным интеллектом, созданным по образу и подобию биологической системы зрения, нам в первую очередь необходимо определить возможности зрительного восприятия на поведенческом уровне. Какие типы образов можно распознать, когда и как? При каких условиях люди совершают ошибки при обработке изображений? Какой объем опыта взаимодействия с миром, и какие разновидности этого опыта требуются, чтобы научиться видеть? Чтобы ответить на эти вопросы, нам необходимо количественно оценить деятельность человека при выполнении хорошо контролируемых визуальных задач. Дисциплина с живописным и привлекательным названием *психофизика* направлена на строгое описание, количественную оценку и понимание поведения субъекта во время решения когнитивных задач.

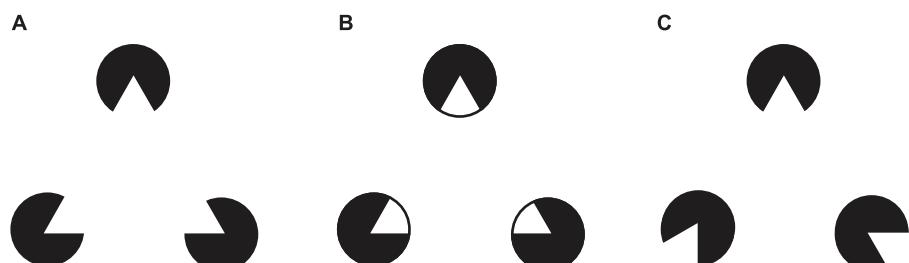
### 3.1. ОБРАЗ ВНЕШНЕГО МИРА, РОЖДЕННЫЙ НАШИМ ВОСПРИЯТИЕМ, – ЭТО СОВСЕМ НЕ ТО, ЧТО ВИДЯТ НАШИ ГЛАЗА

Как уже говорилось в разделе 2.5, ясно, что то, что мы в конечном итоге воспринимаем, является значительно преобразованной версией узора фотонов, падающих на нашу сетчатку. Наш мозг фильтрует и обрабатывает визуальные данные, чтобы понять физический мир вокруг нас, создавая интерпретацию, которая согласуется с нашим опытом. Представление о том, что наш мозг создает этот материал, может поначалу показаться нелогичным: воспринимаемое нами окружение – это достаточно разумное представление о внешнем мире, позволяющее нам ориентироваться, схватывать

объекты руками, предсказывать, что будет дальше, и определять, счастлив ваш друг или нет. Чрезвычайно заманчиво предположить, что наша зрительная система действительно передает идеальную буквальную визуализацию внешнего мира.

Зрительные иллюзии представляют собой убедительные примеры расхождения между тем, что существует в реальном мире, и тем, что мы в конечном итоге воспринимаем. В гл. 2 представлено несколько примеров расхождения между данными на входе и их восприятием: слепое пятно (раздел 2.5), полное исключение входных сигналов во время моргания и во время сверхбыстрых изменений входных данных, сопровождающих саккады (раздел 2.4). Во всех этих случаях наш мозг восполняет недостающую информацию.

Зрительные иллюзии – не исключение из правил; они иллюстрируют фундаментальный принцип, согласно которому наше восприятие – это реконструкция, конфабуляция, вдохновленная входящей визуальной информацией. В мире есть существенно важная информация, которую мы просто не видим. Например, мы не можем воспринимать нашими глазами информацию в ультрафиолетовой части светового спектра (в отличие от других животных, таких как мыши). Другой пример: у нашей остроты зрения есть предел: есть такие мелочи, как бактерии, которых мы не видим глазами. Есть вещи, которые существуют, но мы их не видим, а есть вещи, которых не существует, но мы их видим. Например, когда мы смотрим фильм, на экране отображается последовательность кадров очень быстро, со скоростью 24 или 25 кадров в секунду. Наш мозг не воспринимает каждый кадр из этой последовательности кадров. Вместо этого мозг интерпретирует то, что он видит, как движущиеся по экрану объекты. В качестве другого примера рассмотрим треугольник, показанный на рис. 3.1, известный как треугольник Канижа. Мы видим белый треугольник в центре изображения и можем проследить каждую из сторон этого треугольника. Однако этот треугольник построен из иллюзорных контуров: между краем одного Пакмана<sup>1</sup> и краем соседнего нет белого края треугольника. Треугольник существует только в нашем мозге.



**Рис. 3.1** ♦ Наш мозг создает вещи. (A) Мозг создает белый треугольник из неполной информации, показанной с помощью персонажей – Пакманов – на рисунке. Иллюзия разрушается, если закрыть круги (B) или повернуть Пакманов (C)

<sup>1</sup> Персонаж популярнейшей компьютерной игры прошлого «Pac-Man». – Прим. ред.

## 3.2. ЗРИТЕЛЬНЫЙ ОБРАЗ ЗАВИСИТ ОТ АДЕКВАТНОЙ ГРУППИРОВКИ ЧАСТЕЙ ИЗОБРАЖЕНИЯ С ПОМОЩЬЮ ОПРЕДЕЛЕННЫХ ПРАВИЛ

Наш мозг – это конфабуляторы, довольно полезные конфабуляторы<sup>1</sup>, которые следуют систематическим правилам для создания наших перцептивных миров. Одна из первых и основополагающих попыток установления основных принципов зрительного восприятия была предпринята немецкими философами и экспериментальными психологами в конце XIX века. Так называемые законы гештальта (на немецком языке «гештальт» означает форма) устанавливают элементарные ограничения на то, как образцы света интегрируются в перцептивные ощущения. Эти правила возникли из попыток понять основные принципы, которые приводят к интерпретации объектов как целых, а не как совокупность отдельных линий или элементов, которые их порождают. Эти законы группирования обычно резюмируются указанием на то, что образы – это нечто большее, чем простая сумма частей.

- *Разделение фигуры и фона.* Мы легко отделяем рисунок от фона на основе относительного контраста, размера, цвета и других свойств (рис. 3.2). Знаменитый художник М. К. Эшер (1898–1972) использовал этот аспект познания для создания неоднозначных образов, в которых фигура и фон сливаются в разных областях. Эволюция, вероятно, открыла важность отделения фигуры от фона при обнаружении добычи, что привело к появлению камуфляжа, когда фигура сливается с фоном, что затрудняет ее обнаружение.



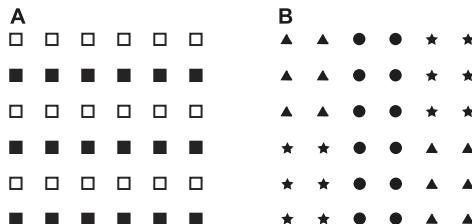
**Рис. 3.2 ♦ Разделение персонажа и фона.** Мы склонны отделять фигуру персонажа (здесь это бегущий человек белого цвета) от фона (здесь это однородный черный цвет)

- *Законченность.* Мы дополняем линии и экстраполируем их на полные известные формы или правильные фигуры. Мы склонны соединять разные части изображения, чтобы создать единую узнаваемую форму.

<sup>1</sup> Конфабуляция – ложное воспоминание. – Прим. ред.

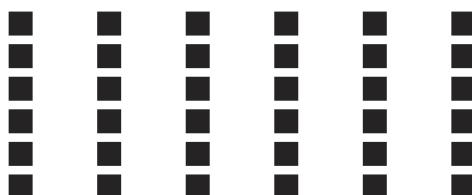
Например, наш мозг создает треугольник в середине изображения Каннижа из неполной информации (рис. 3.1).

- **Сходство.** Мы склонны группировать похожие объекты вместе. Сходство можно определить по форме, цвету, размеру, яркости и другим свойствам (рис. 3.3).



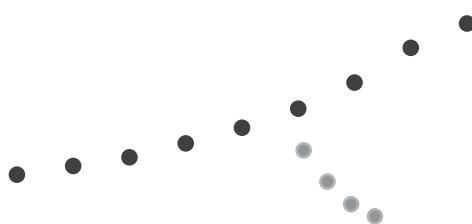
**Рис. 3.3 ♦** Группирование по сходству. Мы склонны группировать объекты с общими свойствами. **(A)** Мы воспринимаем горизонтальные линии, состоящие из черных квадратов, чередующихся с горизонтальными линиями, составленными из белых квадратов, группируя предметы по их цвету. **(B)** Мы воспринимаем пять различных групп, основанных на группировании по форме

- **Близость.** Мы склонны группировать объекты на основе их относительных расстояний (рис. 3.4). Близость – мощный сигнал, который часто может превзойти некоторые другие критерии группировки.



**Рис. 3.4 ♦** Группирование по близости.  
Мы воспринимаем эту фигуру как несколько вертикальных линий

- **Симметрия.** Мы склонны группировать симметричные изображения.
- **Последовательность.** Мы склонны продолжать регулярные последовательности (рис. 3.5).



**Рис. 3.5 ♦** Группирование по последовательности.  
Мы склонны считать, что темно-серые круги образуют непрерывную линию

- **Общая судьба.** Элементы с одинаковым направлением движения обычно группируются вместе. Движение – один из самых сильных и надежных сигналов для группирования и сегментации изображения, превосходящий другие критерии. Для этого животное, которое хочет замаскироваться под фон, должно оставаться неподвижным.

### **3.3. ЦЕЛОЕ МОЖЕТ БЫТЬ БОЛЬШЕ, ЧЕМ СУММА ЕГО ЧАСТЕЙ**

Правила группировки (гештальта) диктуют организацию элементов изображения в структуры более высокого порядка, новые интерпретируемые комбинации простых элементов. Демонстрация комбинации элементов, выходящих за рамки того, что можно различить по отдельным компонентам, называется *холистической обработкой* (основанной на целостности объекта или явления. – Прим. ред.). Особенno широко изучаемая форма холистической обработки – это интерпретация лиц.

Для подтверждения холизма обработки лица должны быть сделаны три основных наблюдения. Во-первых, эффект *инверсии* описывает, насколько сложно различить локальные изменения на лице, когда оно перевернуто. Иллюзия, известная как «эффект Тэтчер», иллюстрирует этот эффект: искаженные изображения премьер-министра Великобритании можно легко отличить от оригинала, когда они повернуты набок, правой стороной вверх, но не тогда, когда они перевернуты. Второе наблюдение, подтверждающее холизм обработки, – это *иллюзия составного лица*: объединение верхней части данного лица А и нижней части другого лица В создает новое лицо, которое кажется отличным от двух исходных лиц. Третий аргумент в пользу целостной обработки – *эффект частей и целого*: изменение локальной особенности лица искажает общее восприятие всего лица. Наблюдение за тем, что целое может быть больше, чем сумма его частей, не ограничивается лицами; опыт в других областях, включая идентификацию отпечатков пальцев или распознавание новых произвольных образов, также приводит к аналогичным холистическим эффектам.

### **3.4. ЗРИТЕЛЬНАЯ СИСТЕМА КОМПЕНСИРУЕТ ИСКАЖЕНИЯ В ИЗОБРАЖЕНИЯХ ОБЪЕКТОВ**

Понимание того, что интерпретация всего объекта – это не просто список его составляющих, затрудняет построение моделей распознавания, основанных на контрольном списке конкретных частей объекта. Еще одна серьезная проблема для этого типа модели распознавания на основе контрольных списков заключается в том, что часто некоторые части могут быть не видны или могут быть сильно искажены. Отличительной чертой зрительного рас-

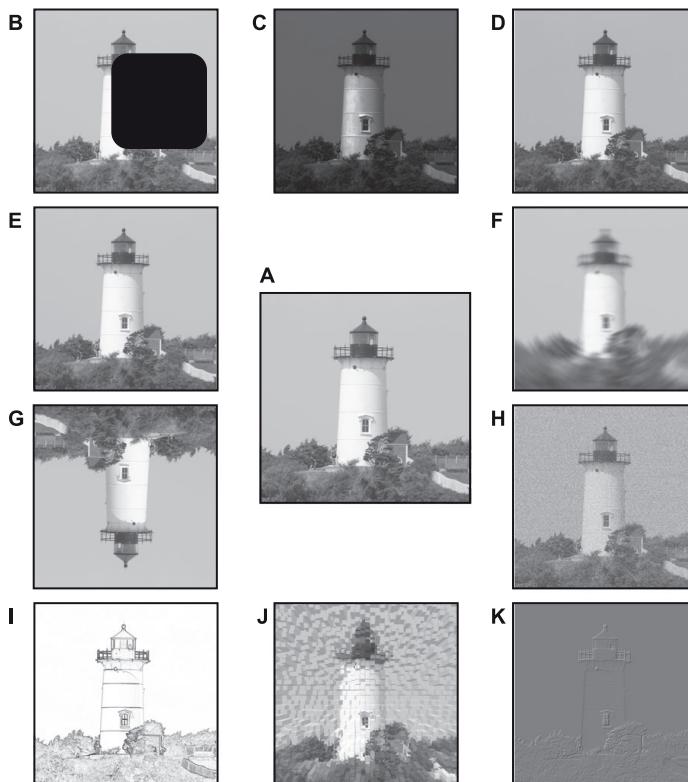
познавания является способность идентифицировать и классифицировать объекты, несмотря на большие трансформации изображения. Объект может отбрасывать бесконечное количество проекций на сетчатку из-за изменений положения, масштаба, поворота, освещения, цвета и других переменных. Эта толерантность к преобразованиям изображений имеет решающее значение для распознавания, она составляет одну из фундаментальных проблем в зрении (гл. 1) и, следовательно, является одной из ключевых целей вычислительных моделей (разделы 8.1 и 8.2). Возможности зрительного распознавания были бы совершенно бесполезны без возможности абстрагироваться от изменений изображения.

Чтобы дополнительно проиллюстрировать критическую роль толерантности к преобразованиям изображений в визуальном распознавании, рассмотрим простой алгоритм, который мы будем называть «машиной механического запоминания» (рис. 1.4). Этот алгоритм получает входные данные от цифровой камеры и отлично запоминает каждый пиксель. Он может помнить солнечные цветы Ван Гога; он может помнить селфи, сделанное две недели назад в понедельник в 14:30; он может точно помнить, как выглядела ваша машина три года назад в субботу в 17:01. Хотя такая экстраординарная память на основе пикселей может сначала показаться довольно примечательной, оказывается, что это представляет собой весьма сомнительный подход к распознаванию. Этот алгоритм не смог бы распознать вашу машину на парковке сегодня, потому что вы можете увидеть ее при другом освещении, под другим углом и с другим количеством пыли на кузове, чем на любой из фотографий в памяти. Проблема с машиной для механического запоминания прекрасно проиллюстрирована в рассказе аргентинского писателя Хорхе Луиса Борхеса под названием «Фунес памятливый». История повествует о злоключениях персонажа по имени Фунес, который приобрел абсолютную память из-за травмы мозга. Первоначальное восхищение Фунеса своей необыкновенной памятью вскоре угасает, когда он не может достичь визуальной неизменности, что проявляется, например, в том, что он не понимает, что собака в 15:00 – это та же собака, что и в 15:01, если смотреть под немногим другим углом. Борхес заключает: «Думать – значит забывать о различиях, обобщать, абстрагировать».

Наша визуальная система может абстрагироваться от различных вариантов изображения объекта для распознавания этого объекта (рис. 3.6), демонстрируя высокую степень устойчивости к изменениям некоторых свойств изображения, включая следующие:

- *толерантность к изменениям масштаба*, т. е. распознавание одного и того же объекта в разных размерах. В зрении размеры объектов обычно измеряются в градусах угла обзора (рис. 2.4). Теперь снова рассмотрим фигуру бегущего человека на рис. 3.2. Если вы держите страницу примерно на расстоянии вытянутой руки, человек будет иметь угловые размеры примерно  $2^\circ$ . Перемещение страницы все ближе и ближе приведет к многократному увеличению его размера, без малейшего ущерба для распознавания. Существуют ограничения на распознавание, обусловленные остротой зрения (если страница отодвинута слишком далеко), а также если изображение становится слишком большим (если страница касается вашего носа). Однако существует широкий диапазон

масштабов, в которых мы можем распознавать объект разного видимого размера как один и тот же;



**Рис. 3.6 ♦ Тolerантность в распознавании образов.**

Маяк легко узнатъ, несмотря на сильные изменения внешнего вида изображения

- *положение относительно фиксации* – то есть распознавание объекта, находящегося на разном расстоянии от точки фиксации внимания. Например, зафиксируйте взгляд на заданной точке, скажем на большом пальце правой руки. Следите за тем, чтобы не двигать глазами или большим пальцем. Затем переместите бегущего человека, изображенного на рис. 3.2, в другое положение. Вы по-прежнему можете распознать изображение в разных местах относительно точки фиксации. Как обсуждалось в разделе 2.5, острота зрения резко снижается по мере удаления от точки фиксации. Следовательно, если вы продолжаете отодвигать страницу от точки фиксации, в конечном итоге изображение бегущего человека станет неузнаваемым. Однако есть широкая область позиций, где распознавание по-прежнему работает;
- *плоское вращение*, то есть распознавание объекта, который поворачивается в одной плоскости<sup>1</sup> (рис. 3.6G). Вы можете узнатъ бегущего

<sup>1</sup> В плоскости, перпендикулярной направлению взгляда. – Прим. ред.

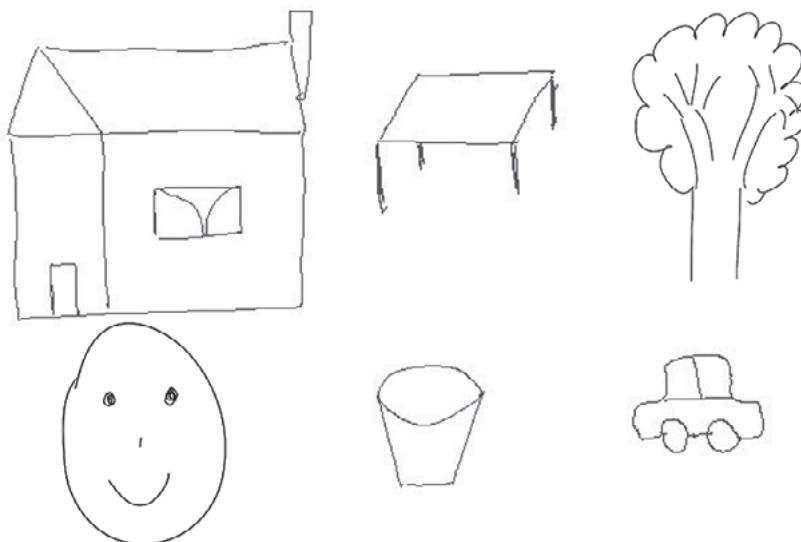
человека, даже если перевернете страницу вверх ногами или наклоните голову. Эффективность распознавания не полностью инвариантна к двумерному вращению, как упоминалось ранее в случае иллюзии Тэтчер;

- *трехмерное вращение*, то есть распознавание объекта с разных точек зрения. Распознавание показывает некоторую степень толерантности к трехмерному вращению объекта, но не совсем к нему инвариантно. Вращение в трехмерном мире представляет собой особенно сложную трансформацию, потому что типы характеристик, обнаруживаемых в объекте, могут довольно сильно зависеть от ракурса. В частности, некоторые объекты намного легче распознать с определенных канонических точек зрения, чем с других углов обзора;
- *цвет*. Во многих случаях объекты можно легко распознать на фотографии – будь то цветное, черно-белое фото или сепия (рис. 3.6Е). Цвет, безусловно, может добавить ценную информацию и может улучшить распознавание, но возможности распознавания довольно устойчивы к изменениям цвета;
- *освещение*. В большинстве случаев объекты можно легко идентифицировать независимо от того, освещены ли они слева, справа, сверху или снизу. Кроме того, восприятие в значительной степени устойчиво к изменениям интенсивности или спектрального состава источника света;
- *неожиданные трансформации*. В некоторой степени мы также можем идентифицировать объекты, подвергнувшиеся трансформациям, которых мы раньше не видели. Возможно, мы никогда не видели маяк, изображенный на рис. 3.6F или K. Способность экстраполировать на такие новые условия особенно примечательна и представляет собой серьезную проблему для компьютерных моделей зрительного распознавания.

Это лишь некоторые из бесчисленных преобразований, через которые объект может пройти с минимальным влиянием на узнавание; многие другие примеры показаны на рис. 3.6. Зрительная система также может вытерпеть многие типы некритических трансформаций, такие как распознавание лиц с изменениями в выражении лица, старением, макияжем или бритьем. Все примеры на рис. 3.6 зависят от идентификации маяка по его резким контрастным краям, но объекты можно легко идентифицировать даже без таких краев. Например, для определения формы объекта можно использовать признаки движения.

Интригующий пример толерантности – способность распознавать карикатуры и контурные рисунки (рис. 3.7). На уровне пикселей эти изображения мало похожи на реальные объекты, и тем не менее мы можем распознать их довольно эффективно, иногда даже лучше, чем реальные изображения. Возможно, что способность интерпретировать контурные рисунки, подобные изображениям на рис. 3.7, больше зависит от конкретного обучения распознаванию символов и определенных соглашений о том, как рисовать эти объекты, чем от визуального сходства форм с объектами, представленными на этих рисунках. В случае карикатур на лица художники улавливают основные узнаваемые черты человека, в отличие от символов и условностей в других

простых набросках, тем самым подчеркивая высокую степень инвариантности к преобразованиям изображения.



**Рис. 3.7** ♦ Распознавание карандашных скетчей. Мы можем идентифицировать объекты на этих контурных рисунках, несмотря на крайнюю простоту ли-ний и минимальную степень сходства с реальными объектами

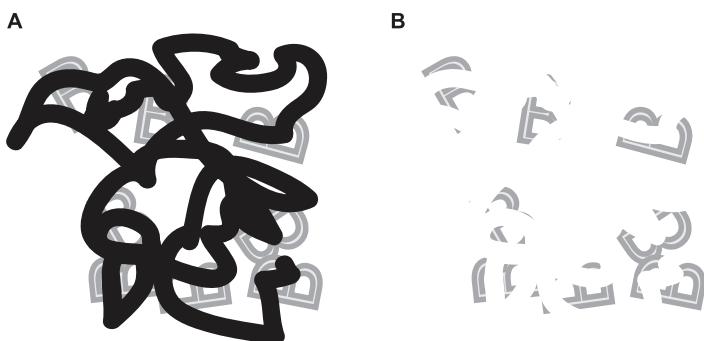
Во всех этих случаях распознавание устойчиво к изменениям изображения, но не полностью инвариантно к этим изменениям. Можно нарушить распознавание, изменив изображение. Таким образом, хотя многие исследователи ссылаются на инвариантное зрительное распознавание, лучшим термином, возможно, является зрительное распознавание, допускающее трансформации, – это подчеркивает, что мы не ожидаем полной инвариантности к любому изменению изображения.

## 3.5. ФИНАЛЬНАЯ СБОРКА: ПОСТРОЕНИЕ ЦЕЛОГО ИЗ ВИДИМЫХ ЧАСТЕЙ

Особенно сложной формой толерантности, которая довольно часто встречается в естественном зрении, является распознавание частично закрытых объектов. Глядя на окружающие нас объекты, мы часто имеем прямой доступ только к частичной информации из-за плохой освещенности или из-за того, что впереди находится другой объект. Расшифровка того, что такое объект, когда видны только его части, требует экстраполяции для получения законченных форм. Грубый пример окклюзии показан на рис. 3.6А. Маяк легко идентифицировать, даже если видно менее половины его пикселей.

Зрительная система обладает замечательной способностью делать выводы на основе неполной информации. Эта способность не является исключительной для зрения, но, скорее, она очевидна во многих других модальностях, включая понимание речи, искаженной шумом, или даже в более высоких областях познания, таких как восстановление истории по нескольким словам, напечатанным на странице, или расшифровка социальных взаимодействий по разрозненным частям информации.

Зрение – это некорректно поставленная задача, потому что решение у нее не единственное. В общем, может существовать бесконечное количество интерпретаций мира, согласующихся с данным изображением на сетчатке глаза. Бесконечность решений легко оценить в случае окклюзии. Существует бесконечное множество способов завершить контуры на основе частичной информации о форме. Например, на рис. 3.6А в маяке может быть большая дыра или за черным пятном может прятаться слон. Однако мы обычно интерпретируем изображение иначе. Несмотря на бесконечное количество возможных решений, зрительная система обычно опирается на единственную интерпретацию изображения, которая в большинстве случаев является правильной. Исследователи обращаются к амодальному завершению, когда есть явный окклюдер<sup>1</sup> (например, на рис. 3.8А), и к модальному завершению, когда создаются иллюзорные контуры для завершения объекта без окклюдера (например, рис. 3.1А). Частичное перекрывание наблюдаемого объекта приводит к выводу о некоей глубине пространства между силуэтами закрывающего объекта и объектом интереса. Такие выводы о глубине помогают создать объемное представление сцены. Окклюдер помогает интерпретировать закрытый объект, как это продемонстрировано в знаменитой иллюзии Брегмана с повернутыми буквами В (сравните рис. 3.8А с 3.8В).



**Рис. 3.8 ♦** Распознавание частично закрытых объектов. (A) Можно распознать повернутые буквы В, несмотря на неполноту информации о них. (B) Легче распознавать объекты, когда присутствует явный окклюдер (A), по сравнению с теми же частями объекта, когда окклюдер отсутствует (B)

Зрительная система может работать с крошечными объемами информации. Можно перекрыть до 80 % пикселей объекта с небольшим ухудшением

<sup>1</sup> Перекрывающий часть изображения объект. – Прим. ред.

качества распознавания. Распознавание зависит от того, какие особенности объекта закрыты. Некоторые части объекта более критичны для распознавания, чем другие. Один из подходов к исследованию того, какие части объекта являются более важными, состоит в том, чтобы представить объекты сквозь пузырьки, случайно расположенные на изображении, контролирующие, какие части объекта видны, а какие нет. Усредняя эффективность по нескольким экспериментам с распознаванием, можно оценить, какие особенности объекта приводят к улучшенному распознаванию, а какие особенности объекта предоставляют менее полезную информацию. Вместо того чтобы представлять изображение через окклюдер или обнаруживать особенности через пузырьки, существует альтернативный подход к изучению завершения формы, который заключается в уменьшении изображения путем обрезки или размытия до тех пор, пока оно не станет неузнаваемым. Используя этот подход, исследователи описали *минимальные изображения*, которые можно легко распознать, но которые становятся неузнаваемыми при дальнейшем уменьшении размера.

## 3.6. ВИЗУАЛЬНОЕ РАСПОЗНАВАНИЕ СОВЕРШАЕТСЯ ОЧЕНЬ БЫСТРО

Напомним, что зрительный образ, который у нас возникает, – это субъективная реконструкция, сделанная нашим мозгом, в соответствии с рядом феноменологических правил группировки элементов изображения. Наш мозг делает выводы, чтобы прийти к уникальному решению некорректно поставленной задачи, создавая представление, которое позволяет нам интерпретировать сцену и идентифицировать объекты и их взаимодействия. Учитывая сложность этого процесса, можно представить, что потребуется огромное количество вычислительного времени, чтобы увидеть что-либо. Напротив, распознавание *кажется* почти мгновенным.

Немецкий физик и врач Герман фон Гельмгольц (1821–1894) продемонстрировал, что передача сигналов в нервной ткани имеет конечную и измеримую скорость, что было довольно революционной мыслью в то время. Как мы обсуждали в разделе 2.6, мгновенного зрения не существует: даже преобразование входящих световых сигналов в выходные данные ганглиозных клеток сетчатки требует времени, порядка 40 мс. Последующая обработка изображения остальной частью мозга также требует дополнительного времени. Что примечательно, так это то, что вся обработка сенсорных входов, толерантность к трансформациям и выводы из неполной информации могут быть выполнены за малую долю секунды. Эта скорость весьма критична: зрение было бы гораздо менее полезным, если бы на получение ответа уходило несколько секунд (гл. 1).

Измерения времени реакции использовались для изучения механизмов восприятия с самого начала психофизики. Оно предоставило исследователям объективную оценку, в отличие от интроспективного анализа. Например,

эти измерения позволили психофизикам количественно оценить понятие компромисса между скоростью и точностью, очевидное во всех визуальных и других задачах и составляющее основу моделей принятия решений.

Одно из первоначальных исследований по измерению скорости зрения заключалось в демонстрации изображений в быстрой последовательности (известной в этой области как задача быстрой последовательной визуальной презентации). Наблюдатели могли интерпретировать каждое из отдельных изображений, даже когда объекты представлялись со скоростью восемь единиц в секунду. В настоящее время относительно легко отображать тестовые изображения на экране в течение коротких периодов времени, охватывающих десятки миллисекунд или даже более коротких временных масштабов. Раньше исследователям приходилось прибегать к остроумным решениям, чтобы картинки подавались только на определенное короткое время. В устройстве, изобретенном в 1859 году для быстрого воздействия световых сигналов, называемом *тахистоскопом*, используется проектор и затвор, аналогичные тем, что применяются в зеркальных фотокамерах с одним объективом. Впоследствии это устройство использовалось во время Второй мировой войны для обучения пилотов быстрому распознаванию силуэтов самолетов. Сложные объекты можно распознать при тахистоскопическом представлении быстрее, чем за 50 мс, даже при отсутствии каких-либо предварительных ожиданий или других знаний<sup>1</sup>.

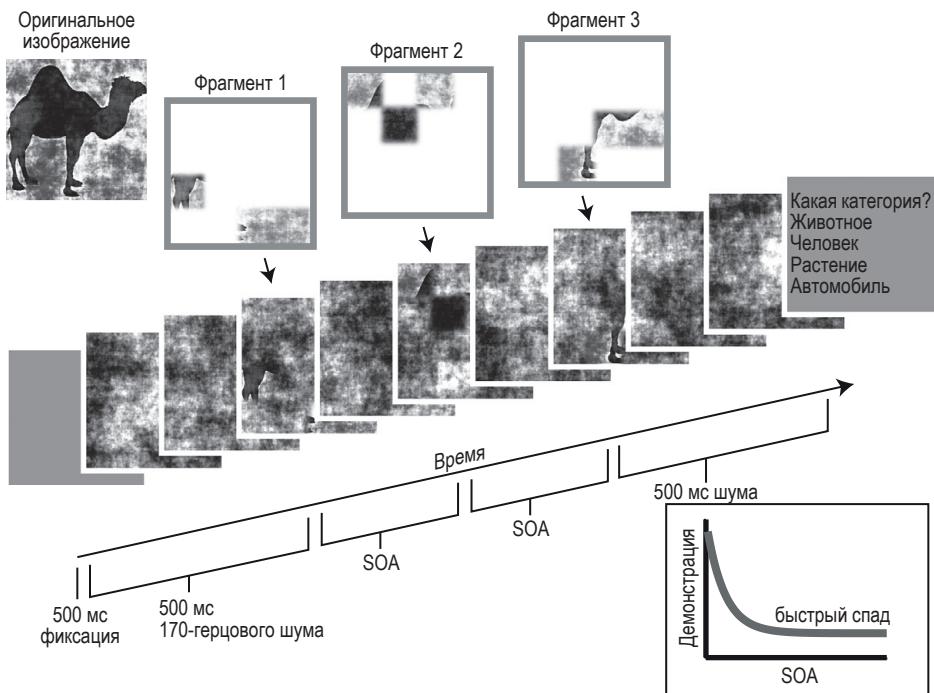
Время реакции, измеренное в ответ на визуальные стимулы, занимает намного больше, чем 50 мс. Чтобы вызвать любой тип ответа (нажатие кнопки, произнесение словесного ответа или движение глаз), требуется несколько шагов обработки, выходящих за рамки визуальной обработки, включая принятие решения и нейронные реакции для подготовки и выполнения поведенческой реакции. Пытаясь ограничить количество времени, необходимое для зрительного распознавания, Саймон Торп и его коллеги записали вызванные стимулом потенциалы ответа электроэнцефалографических сигналов (ЭЭГ) кожи головы, в то время как подопытные выполняли задачу категоризации типа «да/нет» на изображение животного. Субъектам показывали фотографию, на которой было либо животное, либо нет, и им предлагали нажимать клавишу всякий раз, когда они обнаруживают животное. Что именно измеряют эти сигналы ЭЭГ, остается неясным. Однако можно измерить минимальные напряжения, порядка нескольких микровольт, на коже головы и обнаружить изменения, вызванные предъявлением визуальных стимулов. Исследователи обнаружили, что ЭЭГ выявила сигнал примерно через 150 мс после начала действия стимула, который различался между испытаниями, когда было показано животное, и теми испытаниями,

<sup>1</sup> На самом деле лучше сравнивать тахистоскоп с обычным пленоочным кинопроектором, который также показывает нам серию разных кадров, прерываемых черным экраном шторкой обтюратора, со скоростью демонстрации, и соответственно продолжительностью показа каждой картинки, которую можно регулировать. (При обычной проекции с кинопленки время непрерывного показа каждого кадра составляет около 20 мс – 48 кадров в секунду, поскольку каждый кадр показывается дважды, для предотвращения восприятия «мигания» у зрителя. При обычной цифровой проекции оно составляет примерно 40 мс.) – Прим. ред.

когда животного не было. Неизвестно, составляет ли это измерение ЭЭГ визуальный сигнал, сигнал принятия решения, моторный сигнал или некоторую комбинацию всех этих типов процессов. Независимо от точной интерпретации этих измерений, результаты накладывают верхнюю границу для этой конкретной задачи распознавания; исследователи утверждали, что визуальное различие животных и не животных в естественных сценах должно произойти *быстрее* 150 мс. Подобные поведенческие и физиологические реакции наблюдались и у обезьян – макак. В соответствии с этой временной границей в другом исследовании подопытные должны были как можно скорее совершить саккаду в одно из двух альтернативных мест, чтобы отличить стимул «лицо» от стимула «не лицо». Саккады удобны для измерения времени поведенческой реакции, потому что они быстрее, чем нажатие кнопок или верbalный ответ. В среднем подопытным требовалось 140 мс от начала действия стимула, чтобы начать движение глаз в нужном направлении, при решении этой задачи. Эти наблюдения накладывают сильные ограничения на вычислительные механизмы, лежащие в основе визуальной обработки (см. раздел 8.2).

Такая скорость распознавания объектов также предполагает, что механизмы, интегрирующие информацию во времени, должны работать довольно быстро. В нормальных условиях просмотра все видимые части объекта достигают глаза более или менее одновременно (при отсутствии оклюзии и движения объекта). Нарушая такой синхронный доступ к частям объекта, можно исследовать скорость временной интеграции в зрении. В поведенческом эксперименте по количественной оценке скорости интеграции исследователи представляли различные части объекта асинхронно (рис. 3.9), например ломали Шалтая-Болтая и пытались снова собрать части вместе. В промежутках между демонстрацией частей объекта подопытным в течение заданного времени подавали шум, известный как асинхронность начала стимула (SOA). Исследователи предположили, что если бы между презентациями частей разных объектов был большой интервал (длинный SOA), наблюдатели не смогли бы интерпретировать, что это за объект. И наоборот, если бы части были представлены в непосредственной близости во времени, мозг мог бы интегрировать части вместе для единого восприятия объекта. Результаты показали, что подопытные могут интегрировать информацию с промежутком асинхронности около 30 мс.

Другой поразительный пример быстрой временной интеграции – это явление, известное как *анортоскопическое восприятие*, определяемое как интерпретация всего объекта в тех случаях, когда в данный момент времени видна только его часть. В классических экспериментах изображение показывается через щель. Изображение перемещается быстро, позволяя зрителю уловить только небольшую часть целого в любой момент времени. Мозг объединяет все кадры, создавая впечатление цельного движущего объекта. Восприятие движения из кадров в этом и связанных с ним экспериментах в конечном итоге вдохновило на создание фильмов, в которых последовательность слегка смешанных кадров, представленных с достаточной скоростью, интегрируется мозгом, чтобы вызвать непрерывное визуальное восприятие.



**Рис. 3.9 ♦** Завершение пространственно-временного распознавания объекта: наблюдатели могут интегрировать асинхронно представленную частичную информацию об объекте. Подопытным были асинхронно представлены разные части объекта (в этом примере – верблюда). В средней части схемы показана последовательность шагов в эксперименте. Подопытные фиксировали внимание в течение 500 мс, а затем наблюдали последовательность кадров, в которых фрагменты объекта были разделены асинхронностью начала стимула (SOA). Подопытные выполняли пятиальтернативную задачу категоризации с принудительным выбором. Подопытные смогли интегрировать информацию с асинхронностью около 30 мс

Сила временной интеграции также хорошо проиллюстрирована в экспериментах, где актер в черной одежде находится в полностью темной комнате. С помощью всего лишь нескольких световых точек можно сделать вывод о закономерностях движения актера. Подобные исследования показали, что можно динамически группировать и сегментировать информацию исключительно на основе временной интеграции.

Не все визуальные задачи выполняются так быстро. Как известно, найти иголку в стоге сена сложно. Поиск Уолдо<sup>1</sup> может быть несколько раздражающим и занимает несколько секунд или больше, в течение которых наблюдатели, как правило, будут двигать глазами несколько раз, последовательно исследуя различные части изображения. Некоторые зрительные задачи даже

<sup>1</sup> Персонаж фантастического рассказа Роберта Хайнлайна, страдающий миастенией глазных мышц. – Прим. ред.

без движения глаз требуют больше времени. Одним из примеров задачи, которая требует больше времени, даже при отсутствии саккад, является задача завершения формы, описанная в предыдущем разделе. Эксперименты, в которых подопытные имеют ограниченное время для обработки изображения, показывают, что завершение распознавания объектов с сильной окклюзией требует больше времени, чем распознавание полностью видимых аналогов объекта. Самыми простыми из этих экспериментов являются задачи с ограничением по времени, когда идентификация сильно закрытых объектов обычно отстает от распознавания полностью видимых объектов на 50–150 мс.

Другая ситуация, когда подопытные имеют ограниченное вычислительное время, – это эксперименты по *праймингу*. Прайминг относится к форме временной контекстной модуляции, при которой изображению А во времени предшествует другое изображение Р, называемое праймом. Если образ А зависит от Р, то говорят, что Р является прайм-образом А. Например, представление простого Р может повлиять на то, насколько хорошо или насколько быстро наблюдатели распознают стимул А. Прайминг не ограничивается визуальной областью. Представим пример совсем из другой области: вам дают прочитать список названий планет Солнечной системы и их спутников: Меркурий, Венера, Земля, Луна, Марс, Юпитер и т. д. Теперь попробуйте завершить следующее слово: M\*\*N. Вполне вероятно, что вы подумали о «moon», хотя слово «MEAN» было бы таким же хорошим ответом. Однако согласно Google Ngrams слово «mean» в английском языке встречается в три раза чаще по сравнению со словом «moon».

Следовательно, люди должны с большей вероятностью думать о «mean», а не о «moop»; однако предыдущая демонстрация, в которой перечислены названия планет, побудила читателя подумать о Луне.

Подобные эксперименты можно провести и в визуальной области, показывая сначала прайм-изображение вместо списка слов. Изменяя время демонстрации прайма, мы можем оценить, было ли распознано прайм-изображение, путем оценки его влияния на последующий образ основного изображения. Когда прайм Р представляет собой сильно перекрытый объект, величина эффекта прайминга зависит от временного интервала между Р и А. Если этот интервал меньше 50 мс, эффект прайминга исчезает, из чего мы предполагаем, что 50 мс было недостаточно для завершения распознавания и, следовательно, не имело никакого влияния на последующее распознавание основного изображения А.

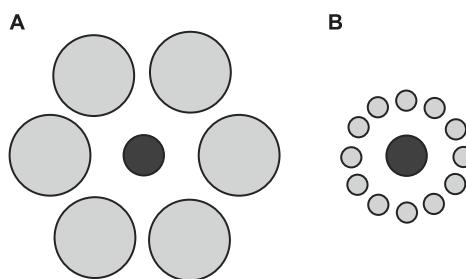
Наконец, еще один распространенный инструмент для уточнения времени обработки в арсенале психофизиков – *обратная маскировка*. В экспериментах с обратной маскировкой за стимулом А следует картинка шума В. Если интервал между А и В очень короткий, менее 20 мс, исходный стимул А практически невидим. При более длительных интервалах подопытные могут различить исходный стимул А, но распознавание не очень хорошее. Когда А представляет собой сильно окклюдирующий объект и шум В появляется примерно через 50 мс после А, становится сложно завершить распознавание образа А. Исследователи утверждают, что картинка шума прерывает вычисления, необходимые для завершения структуры. Если интервал между А

и шумом В превышает приблизительно 100 мс, эффект обратной маскировки исчезает. Эти различные типы экспериментов демонстрируют сходящиеся доказательства того, что объединение частей для завершения распознавания целого во время одной фиксации требует дополнительных вычислительных шагов, проявляющихся в более продолжительном времени реакции.

## 3.7. ПРОСТРАНСТВЕННЫЙ КОНТЕКСТ ТАКЖЕ ИМЕЕТ ЗНАЧЕНИЕ

Помимо временной интеграции, визуальное распознавание также использует возможность интеграции пространственной информации. Существенные особенности узнавания упускаются, если мы вырываем видение из контекста.

Некоторые зрительные иллюзии при зрительном распознавании демонстрируют сильные контекстные эффекты. В простой, но элегантной демонстрации воспринимаемый размер круга может сильно зависеть от размера соседних стимулов (рис. 3.10). Другой пример – иллюзия Мюллера-Лайера: воспринимаемая длина линии со стрелками на двух концах зависит от направления этих стрелок. Эти сильные контекстные зависимости показывают, что зрительная система пространственно интегрирует информацию, и восприятие локальных особенностей также может зависеть от окружения и даже от глобальных свойств изображения.



**Рис. 3.10 ♦ Контекст имеет значение.**  
Темный круг в центре справа кажется больше, чем слева,  
но на самом деле они одинакового размера

Такие контекстные эффекты не ограничиваются визуальными иллюзиями и психофизическими демонстрациями, подобными показанной на рис. 3.10. Повседневное зрение опирается на контекстную информацию. Взгляните на рис. 3.11 (не заглядывая на рис. 3.12): что это за объект в белой рамке? Обычно на этот вопрос сложно ответить с какой-либо степенью уверенности. Если вы не уверены, сделайте предположение. Запишите свои пять самых безумных догадок. Теперь посмотрите на рис. 3.12. Что за объект в белой рамке? Распознать тот же объект на рис. 3.12 гораздо проще! Несмотря на то что пиксели внутри белой рамки на обоих рисунках идентичны,

окружающая контекстная информация существенно изменяет вероятность правильного распознавания объекта. Можно представить, что наблюдатель может изучить несколько различных частей изображения, прежде чем зафиксировать внимание на объекте в белой рамке, чтобы определить, что это за объект. Однако в лабораторных экспериментах, где мы можем точно контролировать направление взгляда, подопытные демонстрируют заметное и быстрое улучшение распознавания, даже когда они сразу фиксируются на белом прямоугольнике, и все изображение исчезает до того, как подопытные успеют переместить область фиксации. Эти контекстные эффекты бывают быстрыми, зависят от объема контекста и могут быть, по крайней мере, частично вызваны представлением еще более простой и размытой версии фоновой информации. Эти эффекты также подчеркивают, что восприятие представляет собой интерпретацию сенсорных входов в свете временного и пространственного контекста.



**Рис. 3.11 ♦** Контекст имеет значение и в реальном мире. Что за объект в белой рамке? Предупреждение: не смотрите на следующий рисунок, прежде чем пытаться ответить на этот вопрос!

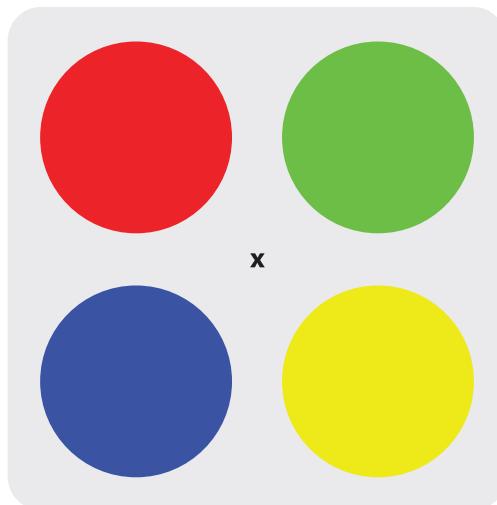


**Рис. 3.12 ♦** Контекст имеет значение и в реальном мире. Что за объект в белой рамке?

## 3.8. ЦЕННОСТЬ ВИЗУАЛЬНОГО ОПЫТА

На наше восприятие влияет предыдущий зрительный опыт во многих временных масштабах. Феномены, которые мы описали до сих пор, в том числе способность отличать животных от неодушевленных существ, обнаруживать лица и интегрировать пространственно-прерывистые фрагменты объектов, охватывают временные масштабы от десятков до нескольких сотен миллисекунд. Мы рассмотрели два примера временной интеграции, которые также охватывают от десятков до сотен миллисекунд: предварительная обработка и обратная маскировка.

Некоторые зрительные иллюзии и феномены демонстрируют сильные эффекты временного контекста в более длительных временных масштабах, охватывающих несколько секунд. Один из примеров – *зрительная адаптация*. Известным примером зрительной адаптации является эффект водопада: после того как вы смотрите на водопад в течение примерно 30 с, переводя взгляд на другие статические объекты, кажется, что эти объекты движутся вверх. Зрительная система привыкла к движению вниз, и вещи, которые не движутся, кажутся движущимися вверх, в направлении, противоположном тому, к чему мы привыкли, глядя на водопад. Адаптация не ограничивается движением. Подобные последствия можно наблюдать после адаптации к цветам, текстурам или объектам, таким как лица. Например, зафиксируйте внимание на «х» в центре рис. 3.13 примерно на 30 с; затем переместите взгляд на белую поверхность. Вы испытаете последействие: на белой поверхности появятся пятна цвета примерно в тех же местах сетчатки, что и круги на рис. 3.13, но в дополнительных цветах.



**Рис. 3.13 ♦** Цветовое последействие. Зафиксируйте центр «х», не двигая глазами, и медленно со-считайте до 30. Затем переместите глаза на белую поверхность. Что вы видите?

Роль опыта в восприятии выходит далеко за рамки секунд и минут. Опыт, полученный на протяжении всей жизни, тоже может существенно повлиять на то, как мы воспринимаем визуальный мир. Например, интерпретация изображения может сильно зависеть от того, видели вы это конкретное изображение раньше или нет. При первом воздействии на так называемую иллюзию далматинской собаки (рис. 10.2В) наблюдатели думают, что изображение состоит из черных и белых пятен. Однако, узнав собаку, наблюдатели могут сразу же интерпретировать сцену и снова узнать собаку в следующий раз. Несколько похожих изображений, созданных Крейгом Муни, обычно используются для оценки роли опыта в визуальном группировании.

Можно сказать, что неподготовленные наблюдатели не могут интерпретировать изображение далматинской собаки, но они могут научиться понимать изображение. В этом случае процесс обучения проходит довольно быстро: краткое объяснение или беглое отслеживание контуров собаки сразу же раскрывает содержание изображения. Интересно, что как только зритель узнает собаку, он может интерпретировать и иные части изображения.

Есть много других ситуаций, когда изображения могут казаться непонятными для наблюдателя без опыта. Возможно, вы видели клинические изображения, такие как рентгеновские снимки или снимки МРТ. Во многих случаях эти изображения для неподготовленного мозга не представляют из себя ничего, кроме странных текстур в оттенках серого (в качестве дополнительного комментария обратите внимание, что именно мозг обучен интерпретировать изображения, но не глаза; нет никакого смысла говорить о неподготовленном глазе!). Однако опытный врач может быстро интерпретировать изображение и поставить диагноз. Точно так же, если вы не читаете по-китайски, китайский текст будет выглядеть бессмысленным набором живописных иероглифов.

Другой аспект того, как наш опыт взаимодействия с миром влияет на наше восприятие, – это создание образа трехмерной структуры на базе двумерного изображения. Многие зрительные иллюзии основаны на интригующих трехмерных интерпретациях. Например, уличные художники создают поразительные иллюзии, которые передают потрясающую трехмерную сцену, если это двумерное изображение рассматривать под определенным углом. Даже когда мы знаем, что это иллюзии, они настолько сильны, что наш мозг, нагруженный многолетним опытом, не может не посыпать свои нисходящие когнитивные влияния, обеспечивая устойчивое объемное восприятие.

Другим примером того, как наши предвзятые, зависящие от опыта представления о трехмерном мире влияют на то, что мы видим, является иллюзия маски. Трехмерная маска лица поворачивается, и под определенными углами она кажется выпуклой, выступающей в сторону зрителя, тогда как под другими углами она будет казаться вогнутой и полой. Однако вогнутый вариант все равно воспринимается зрителем как выпуклое лицо. Существует сильная тенденция к интерпретации лица как выпуклого, скорее всего, потому, что мы редко, если вообще когда-либо, сталкиваемся с вогнутыми полыми версиями лица.

Лица всегда были особенно увлекательной областью изучения психологов. Распознавание и идентификация лиц подвержены тем же эффектам, зави-

сящим от опыта, что и другие визуальные стимулы. Например, психологи охарактеризовали эффект «другой расы», согласно которому людям труднее идентифицировать лица рас, с которыми у них нет опыта общения. Представьте себе человека, родившегося в Азии, который не имел контакта с западным миром ни лично, ни в кино, ни в каком-либо другом формате. Для этого человека все лица людей с Запада будут похожи. Верно и обратное: западным людям, которые не видели многих азиатских лиц, может быть трудно различить их. В качестве другого примера, не касающегося лиц, – пастух, который годами ухаживает за своими овцами, может довольно хорошо различать отдельных овец, в то время как для неподготовленного наблюдателя все они будут выглядеть одинаковыми.

### 3.9. Люди примерно одинаковые, куда бы вы ни пошли, с некоторыми исключениями

В предыдущих разделах и большей части литературы по психофизике в качестве прототипа для обсуждения свойств человеческого зрения мы представляли взрослого человека. В хорошем первом приближении основные наблюдения, описанные до сих пор, справедливы, независимо от пола человека, цвета кожи, религии, культурного происхождения и даже возраста, за исключением первых нескольких лет жизни. Люди видят мир примерно одинаково, куда бы мы ни пошли.

Из этого правила есть исключения. Одно исключение, обсуждавшееся в последнем разделе, связано с ролью опыта. Врачи могут увидеть структуру органа при изучении рентгеновского снимка, а пастух может узнать каждую свою овцу. Другие очевидные исключения – случаи, когда наше оборудование особенное или работает со сбоями. Например, как обсуждалось в разделе 2.5, у многих мужчин в сетчатке есть только два типа колбочек. Мы упомянули некоторые другие дисфункции зрения, включая амблиопию (снижение зрения на один глаз) и нистагм (повторяющиеся непроизвольные движения глаз). Многим людям требуются корректирующие очки для устранения проблем с аккомодацией хрусталика глаза. Альбинизм также приводит к проблемам со зрением в условиях яркого освещения. Как мы обсудим в гл. 4, есть еще корковые дисфункции, которые приводят к аномальному зрительному восприятию.

Возраст тоже имеет значение. С возрастом аккомодация хрусталика глаза может измениться; у некоторых людей развивается катаракта; другие страдают дегенерацией желтого пятна (рис. 2.8). Младенцы и очень маленькие дети также видят мир по-другому, не только из-за их опыта общения с миром, но и из-за того, что их зрительная система еще не полностью созрела. Люди не рождаются с полностью развитой зрительной системой. Острота зрения новорожденных составляет примерно от 7/30 до 7/60, что означает, что то, что они видят на расстоянии от 7 до 30 метров, примерно так же, как видят взрослые на расстоянии от 7 до 60 метров. В Соединенных Штатах по

закону человек с остротой зрения 7/30 или меньше считается человеком с ограниченным зрением.

После того как мы исключим все эти факторы, давайте рассмотрим двух людей примерно одного возраста, с примерно одинаковым визуальным опытом, без какого-либо дефекта зрения. Насколько различаются их представления о мире? В последнее время возрос интерес к пониманию индивидуальных различий в визуальном восприятии у людей с нормальным зрением. Хотя общие принципы, изложенные в этой главе, применимы, остается интересное количество вариаций в восприятии. Пример таких вариаций недавно был выдвинут на передний план во время довольно страстной дискуссии о цвете платья (гл. 1, рис. 1.7). Было обнаружено бимодальное распределение названий цветов, используемых людьми для описания платья.

Кроме того, были также исследования, подтверждающие изменчивость в других визуальных областях. Например, способность распознавать лица сильно различается: одни люди в этом особенно хороши, а другие – плохи. Переходя на более высокую психологическую территорию, красота находится в мозге наблюдателя: существуют значительные различия в визуальных эстетических предпочтениях.

## **3.10. ЗРЕНИЕ МНОГИХ ЖИВОТНЫХ НИЧУТЬ НЕ ХУЖЕ НАШЕГО**

В следующих главах мы углубимся в изучение мозга, чтобы узнать о вычислениях, производимых нейронами, отвечающими за зрительное восприятие. Легче исследовать структуру мозга животных, чем мозга людей. Поэтому большая часть обсуждения в следующих трех главах будет сосредоточена на мозге животных. Обратное верно в отношении поведения: у людей легче, чем у животных, изучать зрительно-моторную координацию. В этой главе основное внимание уделяется зрительно-моторной координации человека. Прежде чем мы приступим к изучению нейронных цепей мозга, важно спросить, обладают ли животные описанными выше удивительными свойствами зрения.

Почти все существующие виды животных, от мух до рыб, от птиц до грызунов и приматов, извлекли выгоду из преимуществ обработки изображений. Ночные животные, такие как мыши, койоты или летучие мыши, обладают хорошо развитым зрением. Многие подземные виды, такие как кроты, все-таки видят. Недавнее исследование так называемых слепых, или малых, кротов, которые считаются слепыми, потому что их глаза на протяжении всей жизни постоянно закрыты кожей, показало, что у них есть палочки, колбочки и ганглиозные клетки сетчатки, которые дают проекцию на остальную часть мозга. Исследования даже показали, что эти кроты реагируют на свет! Есть несколько видов животных, которые полностью слепы, в том числе некоторые виды пауков, рыб и червей. Однако полная слепота – исключение в животном мире.

Правило дивергенции в биологии: зрительная система обладает необычным набором вариаций. Мы не можем воздать здесь должное яркому арсеналу всех визуальных способностей, существующих в животном мире. Животные адаптировались к своей экологической нише и требованиям выживания, развивая специализированные способы использования визуальной обработки. Мы приведем лишь несколько примеров сходства и различий между зрением у животных и людей.

Некоторые свойства зрения животных отличаются от зрительных способностей человека. Люди ограничены видимой частью спектра (которая определяется так именно как видимая людьми!), тогда как другие виды могут воспринимать ультрафиолетовый свет (например, мыши, собаки, многие виды птиц<sup>1</sup>), а также инфракрасный свет (например, многие виды змей<sup>2</sup>). Хотя у большинства людей есть три типа колбочек (раздел 2.5), количество колбочек у животных широко варьируется в зависимости от вида. Некоторые виды имеют только один тип колбочек (например, различные летучие мыши и обыкновенный енот), другие животные имеют два типа колбочек (например, кошки и собаки), и есть даже виды с 16 (креветки-богомолы) или даже 20 видами колбочек (некоторые виды стрекоз). Каракатицы даже могут ощущать поляризацию света, чего люди не могут совсем.

Даже количество и положение глаз у животных сильно различается. У пауков от 8 до 12 глаз, у морских звезд – 5 глаз, а у мечехвостов – 10 глаз. Положение глаз определяет, какие области зрительного поля доступны животному. У улиток глаза в щупальцах; у морских звезд глаза расположены на каждом луче. Даже у видов с двумя глазами положение глаз играет решающую роль в зрении и демонстрирует разные варианты. Глаза, обращенные приблизительно вперед, означают, что центральные части поля зрения доступны обоим глазам, что дает возможность оценивать глубину по стереопсису (небольшая разница в изображении между двумя глазами<sup>3</sup>). С другой стороны, глаза, смотрящие вбок, обеспечивают более широкое поле зрения. В предельном случае бокового расположения глаз, как, например, у кроликов, в центре поля зрения есть слепое пятно. У людей есть приблизительно 120° бинокулярного поля и общее поле зрения приблизительно на 180° по горизонтали. Другие животные с двумя глазами, расположенными так, что они смотрят в стороны, могут иметь поле зрения более 300° (например, коровы, козы, лошади). Два глаза у людей по существу скреплены вместе, так что их положения сильно коррелированы (за исключением определенных случаев, таких как амблиопия). Напротив, другие виды, такие как хамелеон, могут двигать каждым глазом независимо, и поэтому они могут фокусироваться на двух совершенно разных местах поля зрения.

Разрешающая способность визуальной системы также демонстрирует огромное разнообразие у животных, таких как морские звезды, которые представляют собой всю визуальную сцену примерно с 200 пикселями, до таких видов, как хищные птицы, острота зрения которых превосходит чело-

<sup>1</sup> Бабочки и пчелы в числе прочих. – Прим. ред.

<sup>2</sup> А также комары, некоторые виды рыб и лягушек. – Прим. ред.

<sup>3</sup> То есть такие животные обладают бинокулярным зрением. – Прим. ред.

веческую. Ночные хищники имеют более чувствительное зрение, чем люди, в условиях низкой освещенности (например, совы, тигры, львы, ягуары, леопарды, гекконы).

Способность обнаруживать движение, возможно, является одним из немногих универсальных свойств зрительных систем, скорее всего, свидетельством важности реакции на движущихся хищников и добычу, а также на другие надвигающиеся опасности. Многие виды специализируются на быстром обнаружении изменений движения. Например, движения крыльев, вызванные визуальными стимулами, могут быть вызваны у стрекоз примерно через 30 мс после появления стимула, что быстрее, чем время, необходимое для выхода информации из сетчатки глаза человека.

Таким образом, зрительная система человека, сколь бы удивительной она ни была, определенно не уникальна. Существует несколько видов животных, которые демонстрируют «лучшее» зрение с точки зрения способности обнаруживать значимые для выживания вида особенности картины окружающего мира, причем то, что является значимым, конечно, зависит от вида. Наше зрение во многом определяет то, как мы этот мир воспринимаем. Без помощи других инструментов мы ограничены интерпретацией мира, основанной на наших чувствах, и мы часто настолько высокомерны или лишены воображения, что думаем – мир именно такой, каким мы его представляем. Краткий список свойств визуальной системы, изложенный ранее, подчеркивает, что наш взгляд на мир – это всего лишь ограниченное представление, что мы можем видеть то, что другие не могут, и наоборот. Мы не можем увидеть многие захватывающие зрелища в этом мире.

А как насчет свойств восприятия, описанных ранее в этой главе? Как выглядят правила гештальта для других видов? Могут ли животные также выполнять завершение рисунка? Расшифровать то, что видят животные, – непростая задача, требующая хорошо продуманных экспериментов и тщательного обучения. Обезьяны, особенно макаки-резусы, представляют собой один из основных видов, побуждающих интерес для изучения зрительной системы. Их глаза очень похожи на человеческие, и их можно научить выполнять сложные визуальные задания. У шимпанзе и шимпанзе-бонобо зрительная система еще больше похожа на человеческую, но они менее изучены, особенно с точки зрения свойств их мозга.

Обезьяны можно обучить выполнять несколько визуальных задач, в том числе распознавать наличие или отсутствие визуального стимула, сообщать направление движущегося стимула и определять, являются два стимула одинаковыми или нет. Обезьяны были обучены различать сложные объекты, включая лица, а также числовые символы. Они могут обводить линии и контуры. Они даже могут определить, что символ 7 соответствует семи элементам на экране и это количество больше числа 3. Обезьяны также могут научиться играть в простые видеоигры.

Насколько хорошо обезьяны и другие животные могут экстраполировать свои способности распознавания на новые стимулы, которые им незнакомы? Например, насколько их способности терпимы к типу преобразований изображений, описанному ранее в этой главе (рис. 3.6)? Мы можем определить несколько уровней все большей сложности и абстракции в способности вы-

полнять задачи визуального различения: (1) различение при наличии или отсутствии источника света; (2) механическая категоризация, например способность запоминать несколько образцов в классе объектов и отличать эти образцы от нескольких образцов другого класса; (3) открытая категоризация, расширяющая предыдущую способность на ситуации, когда существует обширное и, возможно, непрерывное количество образцов в категории; (4) задачи, в которых животные могут обнаруживать общие свойства разных образцов; (5) абстрактные соотношения, касающиеся связи между образцами, а также связи между категориями. Обезьяны-макаки, по-видимому, действительно способны к относительно сложному уровню абстракции, включая толерантную к трансформации визуальную категоризацию. После обучения с набором категорий визуальных объектов их поведение и характер ошибок напоминают человеческие при тестировании в тех же условиях. Однако некоторые задачи ставят под сомнение, насколько абстрактны внутренние представления обезьян о визуальном мире. Например, обезьян, преуспевающих в задаче визуального различия в верхнем левом зрительном поле, возможно, придется переобучать, чтобы выполнять ту же задачу в правом нижнем зрительном поле, в то время как люди будут быстро переносить свой навык от одной локации стимулов к другой. Этот недостаток экстраполяции может не отражать строго визуального различия между видами, но, возможно, он больше связан с инструкциями по выполнению задания и общением между исследователями и обезьянами.

За последнее десятилетие также возрос интерес к использованию грызунов, в частности мышей и крыс, для исследования зрения. При рассмотрении зрительной системы грызунов есть множество ценных возможностей, включая количество индивидов, которых можно исследовать, и доступность расширенного набора молекулярных инструментов. Тип задач визуального различения, которым обучены грызуны, ограничен по сравнению с поведенческим репертуаром обезьян-макак. Тем не менее крысы, похоже, действительно могут проводить базовые сравнения зрительных образов, даже с некоторой степенью экстраполяции на новые визуализации объектов с точки зрения размера, поворота и освещения.

## 3.11. ОБЗОР ГЛАВЫ

- Психофизика – это увлекательная область, которая занимается количественной оценкой поведения, включая время реакции, показатели эффективности и регистрацию движения глаз.
- Мозг создает вещи. Субъективное восприятие – это реконструкция, которая ограничена сенсорной информацией в свете предыдущего опыта. Зрительные иллюзии иллюстрируют диссоциацию между сенсорными входами и образами, рожденными восприятием.
- Гештальт-правила восприятия описывают, как мы группируем части изображения для создания объектов. К таким правилам относятся замкнутость, близость, сходство, разделение фигуры и фона, непрерывность и общая судьба.

- Зрительное распознавание образов показывает толерантность к большим преобразованиям изображения.
- Мы можем делать выводы из неполной информации – например, при распознавании частично закрытых объектов.
- Визуальное распознавание происходит быстро. Многие задачи распознавания решаются примерно за 150 мс.
- Наблюдатели могут интегрировать информацию, представленную асинхронно, но только в течение нескольких десятков микросекунд.
- Контекстная информация может помочь распознать объекты.
- Люди обычно сравнимы друг с другом в своих способностях распознавания и зрительного восприятия. Тем не менее существует индивидуальная вариативность, особенно когда речь идет о задачах, требующих большого опыта.
- Животные тоже имеют превосходное зрение, и очень важно изучать животных, чтобы выяснить его механизмы.

## ЛИТЕРАТУРА

См. <http://bit.ly/38buAhB> для получения дополнительных ссылок.

- Eagleman, D. M. (2001). Visual illusions and neurobiology. *Nature Reviews Neuroscience* 2: 920–926.
- Herrnstein, R. J. (1990). Levels of stimulus control: a functional approach. *Cognition* 37: 133–166.
- Nakayama, K.; He, Z.; and Shimojo, S. (1995). Visual surface representation: a critical link between lower-level and higher-level vision. In *Visual cognition*, ed. S. Kosslyn and D. Osherson. Cambridge: MIT Press.
- Thorpe, S.; Fize, D.; and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* 381: 520–522.
- Wolfe, J. M.; and Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience* 5: 495–501.

Дополнительный контент вы можете скачать по ссылке <http://bit.ly/38buAhB>.

# Глава 4

---

# Генерация зрительных образов и изменение зрительного восприятия посредством повреждений и электрической стимуляции мозга

Мы хотим понять, как нейронные цепи порождают феномен зрения. Для этого мы можем использовать микроэлектроды и нейрофизиологические регистрации, о которых рассказано в разделе 2.7. В случае сетчатки очевидно, где разместить микроэлектроды для проверки их функции. Однако в человеческом мозге около  $10^{11}$  нейронов, и у нас нет никаких инструментов, которые позволили бы нам записывать данные со всех них. Как нам выяснить, какие части мозга имеют отношение к зрению, чтобы мы могли изучать их на нейрофизиологическом уровне? Чтобы выяснить, как работает устройство, можно разобрать его, внимательно осмотреть его элементы, изучить функцию устройства после систематического удаления отдельных компонентов, собрать устройство заново и убедиться, что его первоначальная функция восстановлена. Инопланетянин, пришедший на Землю, заинтригованный тем, как работают автомобили, может обнаружить, что машина все еще может неплохо перемещаться после отключения радио, но двигатель не запустится без аккумулятора.

Попытки выяснить, как работает мозг, путем изучения поведенческих последствий ограниченных повреждений мозга были фундаментальным подходом с самого начала нейробиологических исследований. История науки о мозге восходит к знаменитому Хирургическому папирусу Эдвина Смита,

создание которого датируется XVII в. до н. э., который описывает симптомы двух воинов, получивших ранения в голову. Несмотря на это раннее описание, на протяжении веков люди считали средоточием сознания сердце. Философ Аристотель (384–322 до н. э.) считал, что цель мозга – охлаждение крови, и утверждал, что сердце содержит разумную душу. Именно изучение травм головы, переломов черепа и травм позвоночника вернуло место размещения психики в мозг.

Легче сказать, чем сделать, чтобы определить функцию различных структур мозга путем исследования последствий его повреждений. Во-первых, различные компоненты системы могут взаимодействовать друг с другом для выполнения любой конкретной задачи, так что удаление одного из них может привести к косвенным функциональным последствиям, помимо тех, которые непосредственно связаны с потерей поврежденной ткани. Во-вторых, может существовать значительная избыточность, так что другой компонент может взять на себя управление, тем самым затеняя фактическую функцию поврежденной области. В-третьих, удалить определенные части мозга не так-то просто. Несмотря на эти проблемы, было изучено многое о функции зрительных цепей путем повреждения мозга у людей и животных.

Помимо оценки роли повреждений определенных областей мозга в зритально инициируемом поведении для манипулирования изменениями функций мозга, существует метод введения электрического тока в мозг. Инвазивная стимуляция мозга может вызывать активность нейронов в областях исследования и помочь проверить гипотезы об участии нейронов этой области в обработке изображений.

## 4.1. КОРРЕЛЯЦИИ И ПРИЧИННОСТЬ В НЕВРОЛОГИИ

Как часто утверждается, корреляции не подразумевают обязательную причинно-следственную связь (*non causa pro causa* на латыни). Это простое логическое утверждение часто игнорируется, что приводит к путанице и неправильному толкованию причин и следствий в нейробиологии, а также во многих других дисциплинах. В новостях можно найти множество примеров такого неправильного толкования. Например, следующие утверждения, взятые из новостных статей, могут быть легко неверно истолкованы как подразумевающие причинную связь: «Курение связано с алкоголизмом»; «Девушки, которые смотрят мыльные оперы, чаще страдают расстройствами пищевого поведения»; «Люди, которые ходят в музеи, живут дольше». Верны ли эти утверждения или нет, здесь не имеет значения. Эти утверждения отражают корреляции, о которых сообщают журналисты, и читатели могут ошибочно сделать вывод о некоторой форме причинно-следственной связи. Медицинское сообщество не застраховано от этого заблуждения. Рассмотрим следующее утверждение: «У большинства детей аутизм диагностируется в возрасте от 18 месяцев до трех лет. Это тот же период, когда детям делают большое количество прививок». Люди видят взаимосвязь между получением прививок и диагнозом аутизм и предполагают, что прививки вызывают

аутизм. Корреляция между возрастом вакцинации и появлением синдромов аутизма не подразумевает какой-либо причинной связи между ними.

В следующих главах мы рассмотрим активность отдельных нейронов зрительной коры. Эти нейрофизиологические регистрации детектируют корреляцию между зрительными стимулами и нейронными ответами или корреляцию между нейронными ответами и визуально вызванным поведением. Выход за рамки этих корреляций для установления причинно-следственной связи – нетривиальный вопрос. Мы рассмотрим здесь два метода, которые помогут приблизиться к пониманию взаимосвязи между нейронной активностью в определенных цепях мозга и зрительным восприятием: повреждениями и электрической стимуляцией.

## 4.2. АРСЕНАЛ ИНСТРУМЕНТОВ ДЛЯ ИЗУЧЕНИЯ ФУНКЦИОНАЛЬНОЙ РОЛИ РАЗЛИЧНЫХ ОБЛАСТЕЙ МОЗГА У ЖИВОТНЫХ

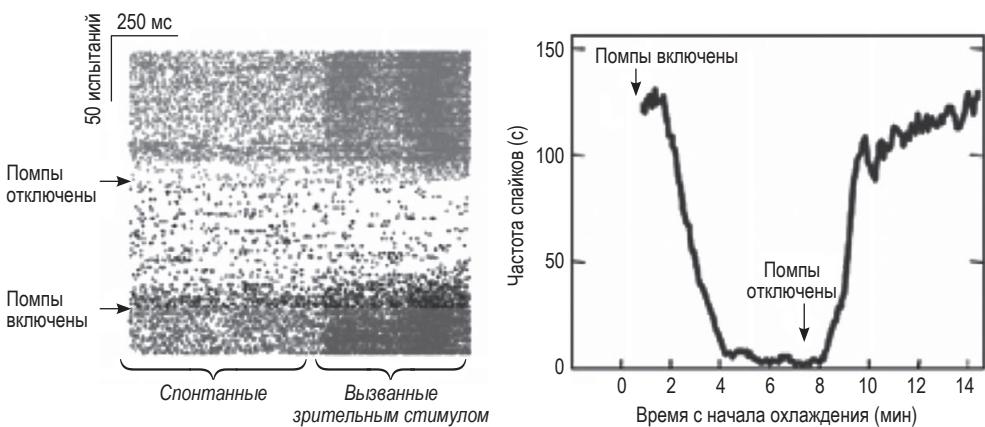
Исследователи используют несколько инструментов для изучения эффекта удаления областей мозга или подавления их активности, включая физические повреждения, эксперименты с охлаждением, фармакологическое вмешательство, избирательную абляцию, молекулярные инструменты, такие как отключения генов, и оптогенетику.

*Физические повреждения.* Одним из наиболее широко используемых инструментов для изучения функций мозга является исследование поведения пациентов с физическими повреждениями мозга. Также возможно вызвать повреждения мозга, вводя химические вещества, такие как нейротоксины. У животных исследователи могут удалить определенные области мозга, чтобы изучить нарушения поведения. Например, ганглиозные клетки сетчатки проецируются в первичную зрительную кору (через LGN) и в верхний холмик. Приматы с поражением верхнего холмика по-прежнему способны решать задачи зрительного распознавания, тогда как животные с поражением первичной зрительной коры – нет. В последующих исследованиях изучали функцию различных частей зрительной коры путем ее повреждения. Повреждения области, известной как средняя височная область (MT, также известная как область V5), приводят к серьезным нарушениям способности различать направление движения, тогда как повреждения нижней височной коры приводят к неспособности распознавать объекты.

Исследования повреждений на животных часто дают очень ценную информацию, но их не всегда легко интерпретировать. Во-первых, сложно сделать анатомически точные повреждения. Во-вторых, оценка поведения может быть нетривиальной: если животное не демонстрирует явных нарушений при решении набора заранее определенных задач, важные аномалии поведения все же могут быть упущены. Наконец, по определению, повреждения, определяемые анатомическими ориентирами, воздействуют на несколько

типов клеток и множество соединений, включая нервные волокна. В качестве приблизительной аналогии представьте себе удаление всего штата Массачусетс из Соединенных Штатов. Потеря экономической активности в Массачусетсе может иметь последствия для экономики в целом, которые трудно предсказать, потеря инфраструктуры автомагистралей и железнодорожных линий, проходящих через Массачусетс, может также повлиять на движение транспорта и другую экономическую активность в Новой Англии и других регионах, а удаление 114 колледжей может повлиять на возможность получения образования жителями других штатов. Возможны серьезные дефициты на рынке, но некоторые из них могут быть неочевидными, некоторые могут быть неуникальными для Массачусетса, иные могут возникнуть не сразу и могут потребовать времени для их оценки, а некоторые могут потребовать подробного знания того, что нужно искать.

**Охлаждение.** Активность нейронов резко снижается при понижении температуры всего головного мозга или его определенной области (рис. 4.1). В мозг можно вживить охлаждающие устройства, чтобы снизить локальную температуру на несколько градусов. Понижение температуры может затормозить активность в области, окружающей охлаждающий зонд. В отличие от физических повреждений, охлаждение – это времененная и обратимая процедура, поэтому одно и то же животное можно изучать до, во время и после инактивации нейронов.



**Рис. 4.1.** Охлаждение участка коры головного мозга может существенно снизить активность местных связей. (А) Растворные графики, показывающие каждый потенциал действия в виде точки, когда полоса многократно перемещалась по рецептивному полю нейрона в визуальной области V2 у бодрствующей обезьяны-макаки. Каждая строка представляет собой одну развертку продолжительностью 1,5 с, и последовательные развертки показаны снизу вверх. Стрелки указывают, когда насос включен для охлаждения области вблизи регистрирующего электрода и когда насос снова выключен. В течение нескольких минут после включения помпы активность снижается. (Б) Средняя визуально вызванная активность (за вычетом спонтанной активности) с течением времени. Зрительная активность исчезает в течение четырех минут после включения помпы. Измененное по Понс и др., 2008

**Фармакология.** Фармакологическое вмешательство также может быть использовано для обратимого торможения участков мозга. Самый известный вид химического вмешательства – это, пожалуй, общая анестезия, воздействие которой распространяется на большие области мозга, когда пациента «усыпляют». Также можно вводить ингибиторы нейронов, чтобы повлиять на активность в локальных цепях. Процедуры фармакологического подавления часто обратимы, и эффекты подавления исчезают, когда лекарства вымываются. Одной из наиболее распространенных форм фармакологического вмешательства является применение *мусцимола*. Мусцимол является мощным активатором особого типа рецептора главного тормозящего нейромедиатора мозга, гамма-аминомасляной кислоты (ГАМК). Это также один из основных психоактивных компонентов нескольких видов психотропных грибов. Мусцимол широко используется для локального подавления нервной активности. Другим распространенным примером является использование *лидокаина*, который действует, увеличивая инактивацию быстрых потенциал-управляемых натриевых каналов, что приводит к снижению или исключению возможности запуска потенциалов действия.

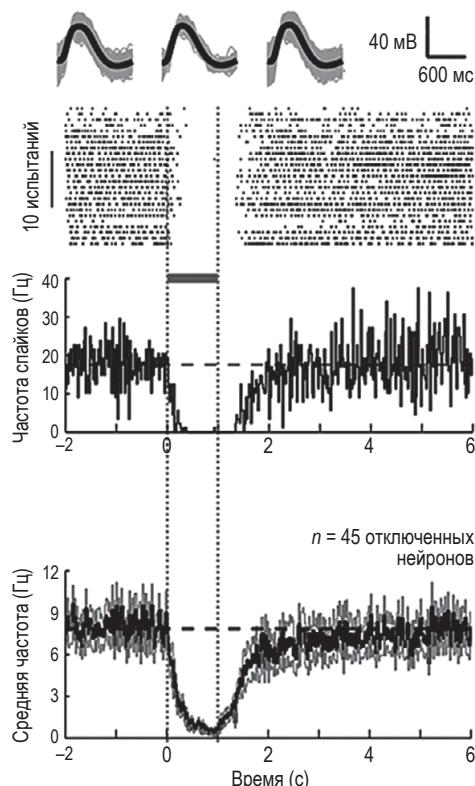
**Удаление клеток.** В нескольких героических исследованиях исследователи использовали инструменты с сильным увеличением для идентификации конкретных типов клеток, а затем удаления этих клеток по одной для изучения поведенческих последствий. Абляция клеток является необратимой процедурой, с помощью этого протокола трудно дезактивировать большие области, и это особенно сложный эксперимент из-за необходимости вручную оперировать клетками, которые нужно инактивировать. Эти усилия вознаграждаются уникальной возможностью для эксперимента, которую дает удаление отдельных клеток из схемы.

**Генные нокауты (отключение генов).** Чтобы описать отключение генов, давайте сначала кратко резюмируем основные принципы молекулярной биологии. Генетическая информация хранится в ДНК. Каждая клетка может экспрессировать разные гены двойной цепочки ДНК, а это означает, что эти гены преобразуются в молекулу-посредник, называемую РНК, а затем транслируются в белки. Мощные инструменты молекулярной биологии позволяют подавить экспрессию определенных генов путем их *отключения* и *включить* ее обратно. Эти методы позволяют исследователям изучать последствия «удаления» определенных генов, «добавления» их обратно либо изменения временной, или пространственной, структуры экспрессии определенных генов.

Более того, недавно разработанный метод, известный как *CRISPR*, позволяет исследователям редактировать отдельные нуклеотиды в определенных генах. Эти методы можно применять таким образом, чтобы животные рождались с измененными схемами экспрессии генов. В качестве альтернативы эти манипуляции также могут быть запрограммированы в индуцируемом формате, чтобы изменения вступали в силу только тогда, когда исследователь их включает. Такие молекулярные манипуляции традиционно были областью работы на мышах, и остается трудным, хотя и не невозможным, использовать эти методы в исследованиях приматов. Совсем недавно исследователи приматов обратили свое внимание на методы инъекции вирусов, которые могут

достичь высокой молекулярной специфичности без необходимости тратить время и деньги на разработку генных нокаутов.

**Оптогенетика.** Особенno интересным и многообещающим новым инструментом для того, чтобы заставить замолчать или активировать определенную популяцию нейронов, является *оптогенетика* (рис. 4.2). Оптогенетика, предложенная Эдом Бойденом и Карлом Дейссеротом, представляет собой преобразующий метод управления нейронной активностью интересующих типов клеток с беспрецедентным разрешением и контролем. Вкратце: нейроны генетически модифицируются путем инъекции вируса для экспрессии светочувствительного ионного канала. Этот ионный канал экспрессируется



**Рис. 4.2.** Выключение определенных групп нейронов с помощью оптогенетики. Активность молчащего нейрона во время и после 1 с воздействия светом в оптогенетическом эксперименте. Вверху: формы волны потенциала действия, полученные до освещения (слева), во время освещения (в центре) или после освещения (справа); показано среднее значение (черный), а также наложение необработанных сигналов (серый). В центре: активность отдельного нейрона, показанная в виде растрового графика спайков и в виде гистограммы мгновенной спайковой частоты, усредненной по испытаниям (внизу; размер клетки 20 мс). Внизу: гистограмма мгновенной частоты спайков, усредненной по всем отключенным одиночным нейронам, записанная после воздействия зеленого света в течение 1 с. Черная линия – среднее; серые линии – средняя стандартная ошибка (SE);  $n = 45$  отключенных нейронов. Измененное по Хан и др., 2011

только в определенных нейронах благодаря *промотору*, который управляет его экспрессией. Промотор – это область ДНК, обычно расположенная прямо перед самим геном, которая отвечает за контроль, когда и где данный ген будет активирован. Как только интересующие нейроны экспрессируют этот светочувствительный ионный канал, клетки готовы к манипулированию путем освещения ткани светом. Открытие некоторых ионных каналов может привести к возбуждению (деполяризации мембранны нейрона), тогда как открытие других ионных каналов может привести к торможению (гиперполяризации мембранны нейрона). Путем инъекции вируса, несущего ингибирующий канал, экспрессии этого канала только в субпопуляции нейронов и освещения ткани, можно отключить только определенные типы нейронов ограниченным по времени и обратимым образом.

При рассмотрении исследований, касающихся повреждений и принудительного торможения, необходимо сделать несколько замечаний. Во-первых, как отмечалось ранее, многие методы, такие как физические повреждения, представляют из себя удаление (или подавление) большого количества ткани. Поэтому в литературе часто встречается обсуждение вопроса о разделении локальных эффектов от эффектов «передающих волокон». Представьте, что аксоны, идущие из области А в область С, случайно проходят рядом с областью В. Поражение области В может также перерезать аксоны А → С. Исследователь, сделавший повреждение в области В, может сделать выводы о функциях области В. Однако может оказаться, что поведенческие последствия повреждения могут быть вообще связаны не с функцией области В, а являются функцией областей А или С, связь которых была нарушена при удалении В.

Еще одно замечание, которое следует сделать, касается немедленных и долгосрочных эффектов. Мозг обладает удивительной пластичностью. Со временем поведенческие эффекты повреждения в данной области мозга могут быть преодолены за счет компенсаторных корректировок в других областях мозга. Одним из таких потенциальных компенсаторных механизмов является наличие «копии» той же области мозга в другом полушарии. Многие (но не все) части мозга имеют аналоги в других полушариях. Эффекты односторонних повреждений (сделанных только в одном полушарии) могут быть замаскированы возросшей активностью в другом полушарии (если в плане эксперимента не приняты особые меры, чтобы учесть этот эффект).

## 4.3. НЕКОТОРЫЕ ИНСТРУМЕНТЫ ДЛЯ ИЗУЧЕНИЯ ФУНКЦИОНАЛЬНОЙ РОЛИ ОБЛАСТЕЙ ЧЕЛОВЕЧЕСКОГО МОЗГА

По очевидным этическим причинам большинство методов, обсуждаемых в разделе 4.2, нельзя применять в исследованиях на людях. Однако существует широкий спектр ситуаций, когда можно получить важную информацию о функциональной нейроанатомии человеческого мозга. Эти случаи воз-

никают из-за различных неврологических осложнений, несчастных случаев и ран. Мы упоминали Оливера Сакса в гл. 2; он был известным неврологом, который написал очень интересные отчеты о своих пациентах с самыми разными психическими расстройствами. В прологе к одной из своих книг, описывающих последствия повреждений человеческого мозга, он писал: «...чувствуя себя частично натуралистом, исследующим редкие формы жизни, частично антропологом, нейроантропологом в полевых условиях... Это рассказы не только о метаморфозах, вызванных неврологическими случаями, но и о путешествиях в альтернативные состояния бытия, другие формы психики, не менее человечные из-за того, что они сильно отличаются от нормы».

Пулевые ранения и раны, нанесенные другим оружием, предоставили важнейшую информацию о функциях зрительной коры головного мозга. Отравление угарным газом, а также некоторые вирусные инфекции, такие как энцефалит, часто вызывают серьезные нарушения зрения, особенно когда они возникают в височной доле. Травма головы, частичная асфиксия в первые недели жизни, опухоли и гидроцефалия (скопление спинномозговой жидкости) тоже могут привести к нарушениям зрения.

При исследованиях последствий повреждений всегда важно тщательно планировать эксперименты. В противном случае можно упустить даже заметные поведенческие дефекты, связанные с повреждениями. Рассмотрим, например, случай пациентов с синдромом расщепления мозга. Это пациенты с фармакологически устойчивой эпилепсией, которые страдают разрывом первичных нервных волокон, соединяющих два полушария, волокон мозолистого тела, как причиной, вызывающей это заболевание. Долгое время считалось, что у этих больных, у которых два полушария были практически разъединены (разъединены, но не полностью, потому что есть несколько других, меньших волоконных трактов, которые также соединяют полушария), не было никаких аномалий восприятия. Однако некоторые последствия этого синдрома стали очевидными только после того, как Роджер Сперри (1913–1994) разработал тщательные эксперименты, основанные на своем научном понимании нейроанатомии зрительной системы.

Сперри знал, что правое зрительное поле отображается на левое полушарие в зрительной коре, и наоборот. Как описано в разделе 2.10, важно различать правый и левый глаза и правое и левое *зрительные поля*: правое и левое зрительные поля (visual hemifield) определяются положением в визуальной сцене относительно точки фиксации. Таким образом, каждый раз, когда вы перемещаете глаза и фиксируете на новом месте, положение каждого поля меняется. Большую часть информации правое полушарие получает как от левого, так и правого глаза (а большая часть левого полушария также получает информацию от сетчаток как левого, так и правого глаза). Используя простой разделитель, Сперри разработал эксперимент, в котором визуальная информация об объекте достигала только правого полушария мозга (информация из левого зрительного поля). Таким образом, информация об объекте была недоступна левому полушарию. Поскольку для большинства правшей левое полушарие имеет решающее значение для речи, Сперри зафиксировал, что испытуемые не могут давать объектам имена. Сперри продемонстрировал, что подопытные не смогли эти объекты назвать. И наоборот, когда информа-

ция об объектах доходила только до левого полушария, у субъектов не было проблем с их наименованием. Сперри был удостоен Нобелевской премии за эту работу в 1981 году. Предыдущие исследования по этим предметам не смогли выявить каких-либо недостатков, потому что визуальная информация была представлена обоим полушариям, и поэтому исследователи не могли наблюдать проблем, связанных с отсутствием связи между ними.

Изучение «естественных повреждений» у пациентов сталкивается с другими проблемами в дополнение ко многим из тех, что обсуждались в предыдущем разделе о работе с животными. Многие исследования, проведенные с людьми, могут быть уникальными и трудновоспроизводимыми из-за крайней трудности найти человека с точно таким же повреждением. Есть множество отдельных тематических исследований. Эти исследования увлекательны и очень информативны. Тем не менее без воспроизведимости не всегда легко проследить или исследовать возникшие дефициты восприятия более подробно, как это можно сделать в исследованиях на животных. Кроме того, естественные повреждения не обязательно соответствуют анатомическим границам областей мозга, цитоархитектоническим или нейрофизиологическим критериям. Многие неврологические повреждения охватывают большие части коры и множество функционально разных областей. Случайный характер этих поражений затрудняет интерпретацию результатов из-за комбинации множественных прямых, косвенных и неспецифических эффектов повреждения. Еще одна трудность в исследованиях на людях заключается в том, что не всегда легко локализовать повреждение или аномалию головного мозга. Магнитно-резонансная томография (МРТ) и компьютерная томография (КТ) могут обнаруживать только определенные типы относительно крупномасштабных изменений мозга, но более тонкие эффекты обычно упускаются.

Несмотря на ограничения в исследовании повреждений мозга человека, преимущество неврологических исследований на людях перед исследованиями на животных заключается в доступности субъективных поведенческих отчетов. В некоторых случаях специфические визуальные дефекты после экспериментов с повреждением или подавлением активности на подопытных животных может быть трудно обнаружить из-за ограниченного характера способов оценки поведения. Поведенческая оценка у людей более проста. Например, люди могут просто прийти к врачу и напрямую рассказать ему о нарушениях своего зрения, во всех деталях.

## **4.4. ЧАСТИЧНЫЕ ПОВРЕЖДЕНИЯ ПЕРВИЧНОЙ ЗРИТЕЛЬНОЙ КОРЫ ГОЛОВНОГО МОЗГА ПРИВОДЯТ К ЛОКАЛИЗОВАННЫМ СКОТОМАМ**

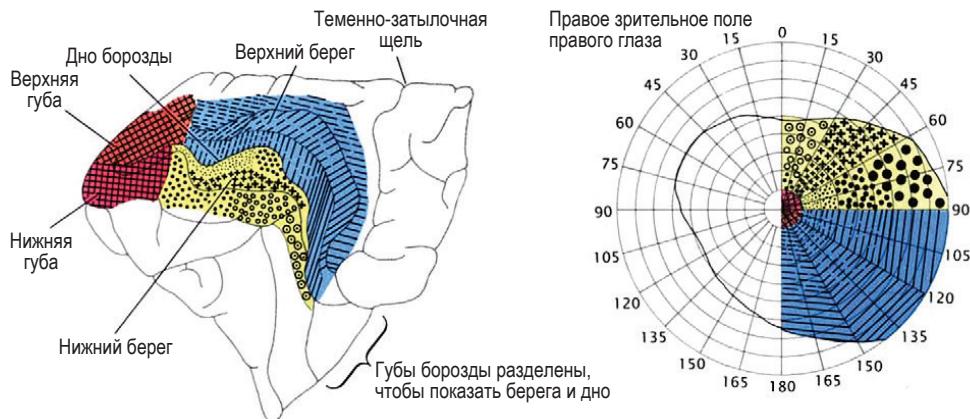
Научные исследования зрительной коры головного мозга, возможно, являются единственным положительным результатом бесчеловечных войн в начале XX века. Открытие первичной зрительной коры произошло вслед-

ствие тщательного изучения траекторий пуль, прошедших через человеческий мозг, и их поведенческих последствий во время Русско-японской войны и Первой мировой войны. В конце XIX века Герман Мунк (1839–1912) сообщил, что повреждение затылочной доли в одном полушарии у обезьян сделало животных слепыми в контралатеральном поле зрения (контралатеральный означает противоположную сторону). Во время войн начала XX века японский врач по имени Тацудзи Иноуэ (1881–1976) и два британских врача по имени Гордон Холмс (1876–1965) и Джордж Риддок (1888–1947) описали четкие и картированные нарушения поля зрения, контралатеральные повреждениям в затылочной коре. Новое оружие, появившееся во время этих войн, позволяло пулям проникать в череп на высокой скорости, не разрушая его полностью. Определенные траектории пули, проходящие через затылочную долю, заставили солдат мгновенно потерять сознание, но в конечном итоге они выздоровели.

Методические исследования показали, что пациенты, пострадавшие от ран в затылочной коре, были практически слепыми в ограниченной части поля зрения, что приводило к *визуальной скотоме*, слепому пятну в определенном месте относительно места фиксации взгляда. Поскольку положение описывается относительно точки фиксации, расположение скотомы в мире изменяется, когда субъект перемещает глаза. Локальное повреждение первичной зрительной коры привело к появлению слепых областей в поле зрения, и эффекты были очень похожи на те, которые наблюдались из-за локальных повреждений в некоторых областях сетчатки. Восприятие формы, цвета и, в меньшей степени, распознавание движений внутри скотомы обычно отсутствовали. Подобные эффекты часто встречаются при исследованиях сосудистых повреждений, опухолей и травм затылочной коры. Путем сопоставления зрительных дефектов с повреждениями головного мозга стало возможным составить карту зрительного поля в задней части затылочной доли, область, которая теперь известна как *первичная зрительная кора*, или V1 (рис. 4.3). Биологи любят придумывать названия генов, типов клеток и областей мозга; нет ничего необычного в том, что несколько имен относятся к одному и тому же. Зрительную область V1 иногда называют первичной зрительной корой, полосатой корой, калькариновой корой, областью 17 (у кошек), а также областью Бродмана 17, на основании цитоархитектонических карт немецкого невролога Корбиниана Бродмана (1868–1918), сделанных им в начале XX века, разделяющих кору на несколько областей. Роза, как ее ни назови, пахнет так же сладко.

Открытие первичной зрительной коры было вдохновляющим в нескольких смыслах. Во-первых, оно зафиксировало, как конкретная функция может быть локализована в определенном месте коры головного мозга. Вместе с исследованиями локализации языковых функций, восходящими к основополагающим работам Поля Брока (1824–1880), эти открытия подкрепили идею о том, что кора головного мозга – это не просто аморфный распределенный лист вычислительных элементов, но, скорее, это совокупность групп нейронов, расположенных в определенных местах, выполняющих определенную функцию. Во-вторых, выходя далеко за рамки локализации языковых функций, повреждения зрительной коры головного мозга показали довольно

подробную топографическую карту соответствия внешнего мира и мозга. В-третьих, эти исследования повреждений заложили основу для подробного анализа нервных цепей, который будет проведен только спустя несколько десятилетий и продолжается по сей день, что будет обсуждаться в гл. 5 и 6.



**Рис. 4.3.** Локальные слепые пятна (скотомы), вызванные повреждениями в первичной зрительной коре. (Слева) Уплощенная карта первичной зрительной коры в затылочной доле вокруг калькариновой щели. На карте разные места выделяются разными цветами для сравнения с отображением зрительного поля справа. (Справа) Карта зрительного поля, показывающая положение от центра фиксации (центр круга) при различных эксцентризитетах. Обратите внимание на непропорционально большую часть коры, на которую проецируется небольшая фoveальная область вокруг фиксации (красный). Воспроизведено с Гликштейн, 1988

Функциональный объем областей коры головного мозга, посвященной различным частям зрительного поля, не распределяется равномерно. Непропорционально большая область V1 предназначена для обработки центральной ямки (красная область на рис. 4.3); это увеличение фoveальной представления известно как эффект кортикального увеличения. Ямка составляет менее 0,1 % от общего размера поля зрения, но ее проекция занимает почти 10 % площади первичной зрительной коры. По мере того как мы перемещаемся от центра фиксации к периферии, площадь коры головного мозга на градус поля зрения становится все меньше. Изображение с высоким разрешением – это хорошо, но за это приходится платить. Для этого мозг должен быть на несколько порядков больше, а значит, тяжелее и дороже с энергетической точки зрения, чтобы представлять все поле зрения с тем же разрешением, что и ямка. Кортикальное усиление ямки в сочетании с быстрыми движениями глаз (раздел 2.4) и рабочей памятью (способность временно хранить информацию) обеспечивает элегантный компромисс для получения высокой остроты зрения при разумном размере мозга.

Зрительное поле отображается на кортекс топографическим способом: близлежащие места в коре представляют собой проекцию близлежащих мест

зрительного поля. Как отмечалось ранее, информация из правого полушария поля зрения отображается на левую зрительную кору, и наоборот. *Калькариновая борозда* разделяет первичную зрительную кору. Калькариновая борозда – это борозда, которая придает извилистую форму мозгу и позволяет сворачивать обширное пространство коры в плотный объем. Верхнее полушарие (желтым на рис. 4.3) отображается на нижнем берегу калькариновой борозды, и наоборот. Из-за такой упорядоченной топографической организации исследователь имеет возможность, при введении электрода для регистрации активности нейронов первичной зрительной коры (раздел 5.3), использовать анатомические ориентиры для локализации рецептивного поля нейрона (т. е. части зрительного поля зрения, области, которая активирует нейрон, рис. 2.9). Топографическое расположение, в данном случае обычно называемое ретинотопическим картированием, может привести к экономии пространства для определения всех соединений, идущих как в первичную зрительную кору, так и обратно. Кроме того, из-за такой топографии грубые измерения, которые усредняют активность нескольких соседних нейронов, могут все же выявить интересные свойства схемы, в отличие от ситуации, когда все нейроны расположены совершенно случайным образом. Такая топографическая организация также является свойством большинства, если не всех, других зрительных областей коры.

Несколько лет назад в сообществе специалистов по зрению было много волнений по поводу описания феномена, называемого *слепым зрением*. Как следует из названия, было замечено, что некоторые наблюдатели с глубокими повреждениями затылочной коры все же сохранили способность к определенной зрительно-моторной реакции в пределах скотомы. Было предложено несколько объяснений этих наблюдений, в том числе существование анатомических маршрутов, которые обходят V1 (например, те, которые соединяют LGN с другими зрительными кортикалыми областями), или наличие небольших неповрежденных островков в V1, которые нельзя увидеть в грубом масштабе магнитно-резонансных изображений, используемых для характеристики повреждений. Хотя нет никаких сомнений в базовой феноменологии остаточных зрительных способностей у пациентов с повреждением V1, диапазон зрительно-моторных реакций у них ограничен. Подопытные могли обнаруживать движение (это также наблюдалось в первоначальном исследовании солдат с ранениями затылочной коры, проведенном Риддоком в 1917 году), отличать день от ночи, приблизительно определять местонахождение источника света и описывать его цвет, а также другие грубые визуально обусловленные действия. Однако во всех случаях их способность к тонкому зрительному различию была утрачена.

Глубокий дефицит после повреждений V1 как у животных, так и у людей, в сочетании с проблемами при изучении зрительно-моторной координации у животных, побудил нескольких выдающихся исследователей в 1950-х годах утверждать, что V1 не только необходима, но и достаточна для зрительно-гого восприятия. В интересном историческом обзоре Чарльз Гросс приводит несколько ярких деклараций этой научной концепции, которая оказалась совершенно неверной: «У людей нет доказательств того, что какая-либо область коры, кроме области зрения 17 [это область V1], важна как первичная

способность видеть образы. Всякий раз, когда вопрос проверялся на животных, история была такой же»; «Функция зрения зависит от стриарной коры [другое название области V1] и ни от какой другой части коры головного мозга»; «Формирование образа и распознавание изображений – все происходит внутри области 17 ... связи области 17 минимальны».

## 4.5. Пути «ЧТО» и «ГДЕ»

Утверждение, что зрение ограничивается областью V1, оказалось совершенно ошибочным. Многочисленные исследования показали, что серьезные нарушения зрения могут быть приписаны повреждениям в областях коры за пределами V1. Одной из первых демонстраций того, что V1 не может определять все, было исследование так называемого синдрома Клювера–Бюси. После двустороннего удаления височной доли у макак в исходных отчетах описывалось множество поведенческих эффектов, включая потерю способности визуального различения, а также другие симптомы, такие как повышенная прирученность, гиперсексуальность и изменение пищевых привычек. В настоящее время считается, что широкий спектр симптомов является следствием массивных и плохо локализованных повреждений. Последующие и более уточненные исследования подтвердили, что повреждения нижней височной коры приводят к дефициту способности различать цвета и формы без всех других сопутствующих невизуальных проявлений.

Исследования вне V1 привели к пониманию фундаментального различия между повреждениями *вентральной* коры и повреждениями *дорсальной* коры. Вентральный путь ведет от V1 в область V4 и внутренние области височной коры (в основном вдоль самой правой части диаграммы на рис. 1.5). Дорсальный поток ведет от V1 к областям MT, MST и теменным структурам коры (в основном вдоль средней и левой части диаграммы на рис. 1.5). Повреждения вентральной зрительной коры приводят к неспособности распознавать форму, и, следовательно, этот путь часто называют путем «что». Повреждения дорсальной зрительной коры приводят к дефициту локализации объекта, и, следовательно, этот путь часто упоминается как путь «где». Как подчеркивается плотной связью на рис. 1.5, эти два пути на самом деле не являются независимыми, и между ними существует множество взаимосвязей.

## 4.6. ПОВРЕЖДЕНИЯ ДОРСАЛЬНОГО ПОТОКА НА ПУТИ «ГДЕ»

Типы дефицита, связанные с повреждениями вдоль дорсального зрительного потока, совершенно отличны от тех, которые связаны с повреждениями вдоль вентрального зрительного потока. Дорсальный путь в основном участвует в пространственной локализации объектов в окружающей среде и обнаружении движения объекта. Повреждения дорсального тракта могут привести

к акинетопсии, пространственному игнорированию, одностороннему пространственному игнорированию, оптической атаксии и симултанагнозии.

**Акинетопсия** относится к специфической неспособности различать зрительные движения. Это состояние можно сравнить с восприятием, вызываемым стробоскопическим светом на дискотеке. Субъект с акинетопсией может видеть отдельные кадры, но не видит траекторию движения. Это состояние также было воспроизведено у обезьян при поражении области МТ на дорсальном пути.

Зрительное одностороннее пространственное игнорирование (*hemineglect*) описывается как нарушение внимания и связано с повреждениями теменно-затылочного перехода. Оно определяется как неспособность реагировать на зрительные раздражители на стороне, противоположной стороне повреждения мозга. Например, человек может есть только с правой половины тарелки или может копировать только половину рисунка. Состояние довольно любопытное: люди с этим расстройством *не слепы* в одной половине поля зрения, о чем свидетельствует тот факт, что их можно заставить обнаружить и идентифицировать объекты, проецирующиеся на поврежденное полушарие. Более того, что особенно интересно, некоторые пациенты с пространственным игнорированием также терпят неудачу в задачах, связанных с извлечением информации из зрительных воспоминаний, связанных с местоположением. В известном эксперименте итальянца с левосторонним пространственным игнорированием попросили представить, что он стоит на площади Пьяцца дель Дуомо в Милане лицом к знаменитому собору, и рассказать, что он может вспомнить из этой сцены. Испытуемый не упомянул объекты и улицы с левой стороны этой точки обзора.

**Симултанагнозия** – это неспособность видеть более одного или двух объектов в сцене<sup>1</sup>. Иногда такие пациенты могут распознать только мелкие части объектов. Люди с симултанагнозией не могут интерпретировать визуальную сцену. Нарушение настолько изнурительно, что этих людей часто описывают как функционально слепых, несмотря на то что они демонстрируют идеальную остроту зрения по отношению к объекту, который они рассматривают в тот или иной момент времени.

Все эти состояния у людей встречаются крайне редко. Мы можем утверждать, что при всех этих условиях распознавание формы объекта остается неизменным. Из-за недостатка распознавания движений и игнорирования пространственных аспектов дорсальная зрительная кора описывается как путь *где*. Альтернативное, но не обязательно взаимоисключающее описание относится к дорсальной зрительной коре как к пути *действия*. Мелвин Гудейл и Дэвид Милнер описали пациентку с повреждением мозга, ограниченным в основном височной долей. У нее были серьезные нарушения в распознавании формы объекта, что мы опишем в следующем разделе. Однако, несмотря на неспособность распознавать объекты, она показала довольно замечательную способность взаимодействовать со многими объектами. Например, она

<sup>1</sup> Или нарушение способности распознавать как целое объекты, состоящие из нескольких частей, несмотря на возможность узнавать отдельные их элементы. – Прим. ред.

показала адекватную реакцию на объекты, которые не могла описать. Также она показала адекватные поведенческие реакции в зрительно-моторных задачах. Гудейл и Милнер предположили, что дорсальный путь особенно задействован в «зрении для действия», немедленном использовании визуальной информации для выполнения определенных действий, обусловленных визуальными стимулами. В отличие от этого режима действия, они предположили, что распознание объекта требует активности центрального потока и, в частности, височной доли.

## 4.7. Нижняя височная кора критически важна для распознавания визуальных объектов у обезьян

Путаница вокруг синдрома Клювера–Бюси иллюстрирует проблемы в интерпретации последствий серьезных повреждений коры головного мозга. Определение более точно локализованных повреждений, ограниченных нижней височной корой (ITC), показало, что двустороннее удаление ITC приводит к нарушению обучения зрителю различению, а также к дефицитам в сохранении информации о зрительных различиях, которые были изучены до повреждения. В типичных экспериментах обезьяны должны были научиться различать разные визуальные формы, чтобы получить награду. Животные с повреждениями ITC не справляются с этой задачей, но они все еще могут выполнять другие визуальные задачи, такие как изучение того, какое из двух визуальных местоположений связано с вознаграждением.

Серьезность дефицита обычно коррелирует с трудностью задания. Обезьяны все еще могут выполнять «легкие» задачи визуального различия после двусторонних повреждений ITC. Дефициты касаются объектов, визуальных образцов, размера, цвета и других свойств объекта. Нарушения в распознавании форм, определяемых движением или яркостью после повреждений ITC, также были описаны. Поведенческие дефекты ограничиваются визуальной областью и не влияют на различие, основанное на тактильных, обонятельных или слуховых сигналах. Ни «психическая слепота», ни другие социальные эффекты, описанные первоначально Клювером и Бюси, не были очевидны после двусторонних повреждений ITC, что дополнительно подчеркивает важность пространственно локализованных повреждений для адекватной интерпретации поведенческих дефектов. Эти недостатки зрительного распознавания формы сохраняются надолго.

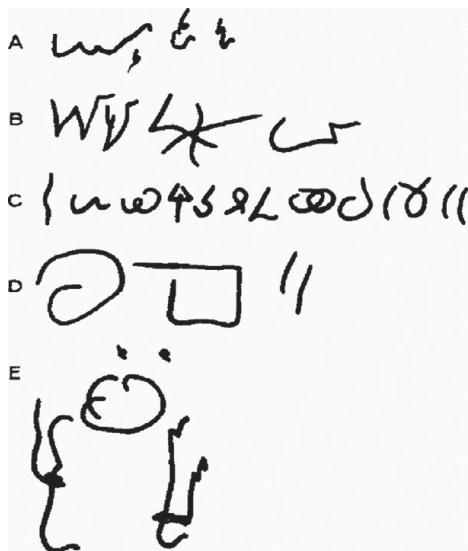
Изучая анатомические пути, описанные на рис. 1.5, мы видим, что существует множество способов передачи информации из одной точки в другую в зрительной коре головного мозга. Информация может гибко маршрутизироваться по схемам в зависимости от характера решаемой визуальной задачи. В отсутствие ITC некоторые легкие задачи могли быть решены путем направления информации из ранних зрительных областей коры в центры принятия решений и двигательные центры. Для других, более сложных за-

дач могут потребоваться расчеты, которые выполняются в более высоких областях, таких как ИТС.

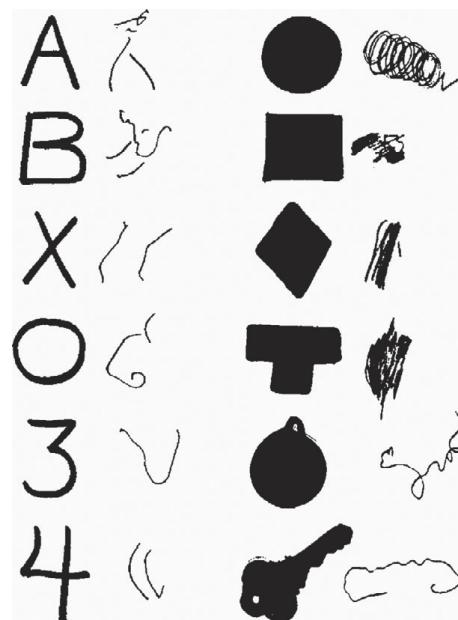
Точно так же, как синдром Клювера–Бюси может быть разделен на несколько синдромов с помощью более точно сделанных повреждений, вполне вероятно, что будущие, еще более точные повреждения внутри ИТС еще больше фракционируют дефекты распознавания объектов, преобладающие после двусторонней аблации ИТС. Действительно, намеки на этот тип специфичности очевидны в недавних элегантных работах, сочетающих фармакологию, оптогенетические манипуляции, нейронные регистраций и поведение у обезьян. Исследователи сосредоточили свое внимание на области ИТС с обилием нейронов, которые реагируют преимущественно на лица по сравнению с другими объектами (тема, к которой мы вернемся, когда будем изучать нейрофизиологические свойства нейронов ИТС в разделе 6.2). В той степени, в которой активность этих нейронов играет важную роль в задачах, которые зависят от распознавания формы лица, авторы предположили, что подавление некоторых локальных областей может нарушить поведенческие реакции в соответственных задачах распознавания. Чтобы оценить эту гипотезу, они обучили обезьян выполнять задание по гендерной дискриминации на основе изображений лиц. После того как животные были обучены, авторы инактивировали небольшие локальные районы ИТС. Эта инактивация проводилась с использованием либо оптогенетических манипуляций, либо фармакологического вмешательства. Подавление локальной нейронной активности привело к ухудшению способности обезьян к дискриминации по признаку пола. Эти поведенческие эффекты были обратимыми: функциональность возвращалась к нормальному уровню после отключения оптогенетического или фармакологического подавления нейронов. Эффекты также были специфичными: инактивация других областей мозга не приводила к таким поведенческим нарушениям. В целом исследования повреждений указывают на важную функцию ИТС, заключающуюся в способности зрительного различать различные формы. Такие исследования сыграли важную роль в определении направлений нейрофизиологических исследований свойств нейронов ИТС во время зрительного распознавания (раздел 6.2).

## **4.8. ПОВРЕЖДЕНИЯ, ВЕДУЩИЕ К НАРУШЕНИЮ РАСПОЗНАВАНИЯ ЗРИТЕЛЬНЫХ ОБРАЗОВ У ЧЕЛОВЕКА**

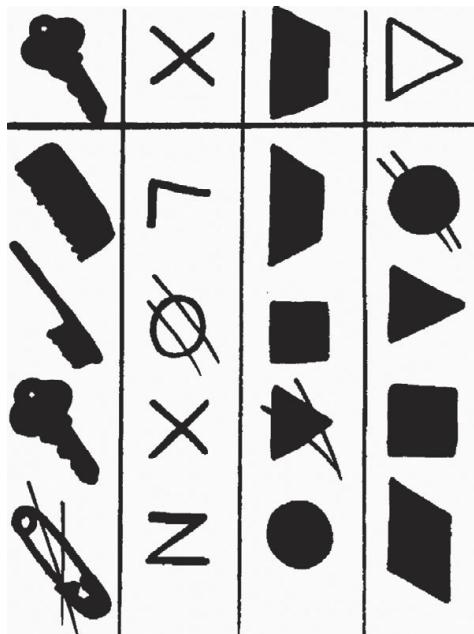
Из-за экспериментальных проблем гораздо меньше известно о последствиях повреждения вентральной зрительной коры головного мозга человека. Повреждения вокруг области V4 вентрального зрительного потока приводят к *ахроматопсии*, специфической неспособности распознавать цвета. Обратите внимание, что это состояние отчетливо отличается от типа дальтонизма сетчатки, описанного в разделе 2.5, который связан с отсутствием одного или нескольких типов колбочек. Повреждения в более высоких областях височной доли могут привести к разнообразным интригующим формам *агнозий* (*agnosia* в переводе с древнегреческого означает «незнание»). Рисунки 4.4–4.6



**Рис. 4.4** ♦ Пациент с агнозией визуальных форм, который изо всех сил пытается их рисовать. Пациенту было предложено нарисовать (A) свое имя; (B) буквы W, V, L, X и A; (C) числа от 1 до 11; (D) круг, квадрат и треугольник; (E) человека. Воспроизведено по книге Бенсона, 1969



**Рис. 4.5** ♦ Пациент также не способен копировать. Тому же пациенту было предложено скопировать буквы и цифры в столбце 1 и фигуры в столбце 3. Воспроизведено по книге Бенсона, 1969



**Рис. 4.6 ♦** Тот же пациент не может выполнить задачу визуального сопоставления формы. Пациента попросили указать, какой из объектов соответствует верхнему в каждом столбце. Воспроизведено по книге Бенсона, 1969

илюстрируют тип поведенческих дефектов, встречающихся при одном виде зрительной агнозии при выполнении трех различных задач. Эти рисунки иллюстрируют поведение 25-летнего пациента, который был обследован через семь месяцев после случайного отравления угарным газом. Пациент мог ориентироваться в больнице, в которую он был госпитализирован, он мог следить за движущимися объектами и мог называть цвета. На сетчатке нет аномалий. Его зрительные поля – то есть его способность обнаруживать вспышки света в разных местах относительно фиксации – были в основном нормальными.

Хотя элементарные зрительные функции были сохранны, пациент не мог назвать объекты, расположенные перед ним. Он все еще мог давать имена объектам, используя тактильные, обонятельные или слуховые сигналы, из чего можно предположить, что это нарушение не было связано с неспособностью воспроизводить речь или извлекать семантическую информацию об объектах. Он не мог рисовать определенные формы, в том числе написать свое имя или нарисовать человека (рис. 4.4). Неспособность рисовать определенные формы по запросу потенциально может быть приписана определенному нарушению в извлечении зрительных воспоминаний. Однако пациент также не мог копировать формы, используя шаблоны, представленные перед ним (рис. 4.5). Пациент мог ясно понимать язык, а также мог выполнять двигательные команды. Можно было бы утверждать, что задачи на рис. 4.4 и 4.5 основаны на умении рисовать, и люди могут иметь разные способности

рисовать. Тем не менее тот же пациент мог рисовать лучше и, конечно, умел писать свое имя до несчастного случая. Кроме того, пациент также безуспешно боролся с задачей визуального сопоставления форм, не требующей умения рисовать (рис. 4.6).

Как отмечалось ранее, на рис. 4.4–4.6 приведено описание одного случая. Не все пациенты с агнозиями имеют одинаковые нарушения. Например, некоторые пациенты могут отлично справиться с задачей копирования (рис. 4.5), но не с задачей рисования по памяти (рис. 4.4). Не ясно, следует ли описывать этот тип недостатка как нарушение зрения или, скорее, как языковой недостаток, проблему с визуальными образами или сложность в извлечении образов из памяти. Тип обобщенного дефицита, связанного с образами в сочетании с нормальным зрительным полем, адекватной речевой способностью и функцией памяти в других сенсорных модальностях, как показано пациентом на рис. 4.4–4.6, встречается редко, но, по-видимому, предполагает именно нарушение зрения.

В исследованиях такого типа часто участвуют отдельные или сразу несколько испытуемых, и очаги повреждения не всегда четко очерчены. Агнозии распознавания формы были подразделены на несколько групп в зависимости от того, считаются ли они в основном визуальными, связаны ли дефициты также с языком и влияют ли эти дефициты на манипулирование объектами или распознавание посредством других сенсорных модальностей. Один из вариантов расстройств называется *ассоциированной зрительной агнозией*. Это название присваивается случаям, когда копирование чертежа и сопоставление сложных образов остаются незатронутыми, но рисование по памяти и идентификация объектов значительно ухудшаются. Как упоминалось ранее, во многих из этих исследований не совсем ясно, что дефициты связаны именно с механизмом зрительного распознавания. Пациенты могут проявлять дефициты в назывании визуальных стимулов, их описании, использовании, извлечении из памяти – все задачи, которые могут зависеть от языковых способностей или иметь отношение к ним, – но не обязательно при сопоставлении двух похожих объектов на основе внешнего вида или их копировании. Другой вариант называется *аперцептивной зрительной агнозией*. Эти пациенты не могут давать имена, копировать или сопоставлять простые образы. Поскольку они не могут копировать или сопоставлять образы, эти случаи могут быть ближе к случаю с проблемами зрительного распознавания, описанными на рис. 4.4–4.6.

Важно отметить, что во многих случаях зрительной агнозии основная зрительная функция остается неизменной. Острота зрения, способность распознавать цвета, способность обнаруживать движение и другие зрительные функции, похоже, не страдают. Эта двойная диссоциация между «основными» зрительными функциями и зрительными способностями «более высокого уровня» согласуется с идеей иерархии вычислений, которая постепенно извлекает все более сложную информацию из изображения – от ранней обработки изображения в сетчатке, LGN и первичного зрительного восприятия коры головного мозга до полного структурирования информации об образе в нижней височной коре. Действительно, зрительные агнозии обычно связаны с двусторонним повреждением зрительных областей за пределами

области V1, обычно включая нижнюю височную кору. Как отмечалось в предыдущем разделе, многие из этих основных функций зрения могут быть реализованы через соединения на рис. 1.5, которые обходят ИТС.

Будут ли наблюдатели со зрительной агнозией испытывать трудности в задачах рисования и сопоставления образов для любого типа объекта? Насколько специфичны агнозии? Любопытно, что в нескольких исследованиях сообщалось о случаях агнозий, зависящих от категории объектов распознавания. Например, некоторые исследования сообщают о более значительном дефиците распознавания «живых существ». Другие исследования описывают неспособность распознавать животных, инструменты, слова или дорожные знаки. В одном исследовании сообщалось о пациенте с неспособностью назвать фрукты или овощи, представленные с помощью карандашных набросков или фотографий. В литературе по исследованиям дисфункций человека, касающихся симптомов нарушения зрения, приводятся заметные, а иногда весьма специфические дефициты в распознавании зрительных образов.

Одна конкретная форма агнозии получила особое внимание в литературе. *Прозопагнозия* (*proso* на древнегреческом означает «лицо») относится к неспособности визуально распознавать лица с сохранной способностью идентифицировать другие объекты и формы. Агнозия на лица также очень редка и обычно возникает после поражения мозга, вызванного инсультом в правой задней мозговой артерии. В некоторых исследованиях описана также врожденная форма прозопагнозии. Обычно поражаются веретенообразные и язычные извилины. Оливер Сакс описал крайнюю форму прозопагнозии в своей книге «Человек, который принял свою жену за шляпу». Люди с прозопагнозией часто могут узнавать людей по голосам, одежде, походке и другим характеристикам, но не по лицам или фотографиям их. Охват случаев, в которых эффекты должны быть описаны как специфические для лица, широко обсуждался. Некоторые авторы утверждают, что нарушение распознавания лиц может быть лучше понято как общая трудность в выявлении экземпляра из более общего класса со многими схожими особенностями и степенью умения оперировать этими особенностями.

## 4.9. ИНВАЗИВНАЯ ЭЛЕКТРИЧЕСКАЯ СТИМУЛЯЦИЯ МОЗГА ЧЕЛОВЕКА

Повреждения – не единственный способ изучить причинно-следственную (каузальную) функцию той или иной области мозга. Теперь рассмотрим другой способ вмешательства в работу мозга: введение электрического тока. Уайлдер Пенфилд был одной из ключевых фигур в инвазивных исследованиях человеческого мозга, работая с пациентами, страдающими эпилепсией. Будучи нейрохирургом, он понял, что имеет прямой доступ к внутренней работе человеческого мозга благодаря своему нейрохирургическому подходу к эпилепсии. Он много работал с пациентами, страдающими фармакологически устойчивой эпилепсией. У этих пациентов приступы не купируются

современными методами фармакологического вмешательства. В этих случаях одним из лучших подходов к устраниению припадков является удаление эпилептогенного очага, то есть части мозга, ответственной за возникновение припадков. Чтобы выполнить этот тип резекции, нейрохирург должен уметь четко локализовать эпилептогенный очаг, а также составить функциональную карту области, чтобы гарантировать отсутствие каких-либо других неблагоприятных когнитивных симптомов как следствие хирургического вмешательства. Руководствуясь этими клиническими потребностями, нейрохирурги инвазивно имплантируют пациентам несколько электродов; пациенты остаются в больнице около недели с установленными электродами, чтобы исследователи накопили достаточно данных. В течение этой недели можно исследовать функцию человеческого мозга с гораздо лучшим соотношением сигнал/шум, гораздо лучшим пространственным разрешением и гораздо лучшим временным разрешением, чем позволяет любой другой метод изучения активности человеческого мозга.

Поскольку у большинства этих пациентов эпилептогенный очаг удаляется, также возможно изучить поведенческие последствия удаления части мозга. Одно из основополагающих исследований Пенфилда описало роль медиальной височной доли в консолидации памяти у пациентов, перенесших двустороннее удаление гиппокампа и прилегающих областей. Удаление гиппокампа на обоих полушариях приводит к серьезным проблемам с консолидацией памяти: пациент может нормально видеть и визуально интерпретировать сцену, может вести нормальный разговор, рассуждать и выполнять различные когнитивные задачи. Фактически после разговора с таким пациентом в течение нескольких минут было бы трудно обнаружить что-либо необычное. Несмотря на то что гиппокамп находится на вершине анатомической схемы зрительной системы, представленной на рис. 1.5, все имеющиеся на сегодняшний день данные свидетельствуют о том, что гиппокамп не является визуальной областью. Отличительные характеристики пациентов с двусторонним иссечением гиппокампа проявляются только при рассмотрении системы памяти. Если бы вы поговорили с тем же пациентом на следующий день, он не вспомнил бы, что произошло во время вашей первой встречи. Из-за этих исследований в настоящее время выполняются только односторонние резекции.

Пенфилд также был одним из пионеров в проведении нейрофизиологических регистраций потенциалов с внутричерепных электродов в человеческом мозге. Кроме того, он также широко изучал поведенческие эффекты, вызванные электрической стимуляцией через субдуральные электроды<sup>1</sup>, когда испытуемые бодрствовали и с готовностью рассказывали о своих ощущениях. Электростимуляция – это стандартная процедура, которая обычно используется в больницах по всему миру. Поскольку в головном мозге нет болевых рецепторов, это безболезненная процедура, и пациенты могут бодрствовать во время операции на головном мозге, что часто оказывается весьма полезным с клинической точки зрения. В этих случаях очень важно работать с бодрствующими субъектами, чтобы иметь возможность кар-

<sup>1</sup> Находящиеся под твердой мозговой оболочкой. – Прим. ред.

тировать когнитивные функции перед резекцией. В частности, неврологов и нейрохирургов беспокоят языковые функции, которые часто находятся в непосредственной близости от эпилептогенных областей. Цель состоит в том, чтобы лечить эпилептические припадки, не влияя на другие когнитивные функции. Одно из самых известных открытий Пенфилда, основанное на этих исследованиях электростимуляции, – это *кортикальный гомункулус*, отображение человеческого тела в моторной и соматосенсорной областях вокруг прецентральной извилины.

После стимуляции данного места он просил испытуемых устно сообщать о своих ощущениях. Пенфилд определяет места стимуляции номерами. Например, в первый раз, когда он стимулировал электрод 5, пациент не ответил. После второго импульса стимуляции в том же месте пациент сказал: «Что-то слышу». В четвертый раз он сообщил, что слышал: «Слышу голоса людей». Пенфилд переключился на электрод 7. Первый импульс на электроде 7 вызвал следующую реакцию: «Как шаги, слышные по радио». После третьего импульса стимуляции на электроде 7 субъект объяснил: «Это было похоже на то, как будто ты в танцевальном зале, стоишь в дверном проеме – в спортзале – как в средней школе Ленвуд». Двадцать минут спустя Пенфилд вернулся к электроду 5, и испытуемый сообщил: «Голоса людей».

Некоторые из наблюдений записаны здесь дословно, чтобы проиллюстрировать захватывающие возможности с точки зрения вопросов, которые мы можем задать, получая прямые устные отчеты при стимуляции коры головного мозга человека. В то же время этот пример показывает, насколько сложно интерпретировать результаты этих увлекательных, но анекдотических отчетов. Что именно стимулируется в этих исследованиях? Сколько нейронов активировано? Какие типы нейронов активируются? Насколько воспроизводимы эффекты при многократном повторении? Как ответ на эти вопросы и обеспечение поведенческих отчетов зависят от параметров стимуляции, таких как длительность и интенсивность импульса? Как выводы зависят от показателей, используемых для оценки поведенческого результата? Что чувствует субъект во время электростимуляции? В какой степени на субъективный отчет влияет окружение (врачи, медсестры, больница)? Как мы можем сопоставить эти увлекательные отчеты, полученные с помощью электростимуляции, с нашим пониманием функций коры головного мозга? При интерпретации теряется много информации.

В некоторых случаях электроды помещают в области зрительной коры. В частности, при введении электродов в раннюю затылочную кору несколько исследователей продемонстрировали, что можно вызвать перцептивные световые вспышки, названные *фосфенами*. Пример такого эксперимента показан на рис. 4.7. После введения тока испытуемых просили сообщить форму и местоположение того, что они воспринимают. В большинстве случаев подопытный сообщал о том, что видел округлые вспышки света; в некоторых случаях, как в случае с электродом 9, субъект сообщал о том, что видел удлиненные линии. Мы кратко упомянули нейроны в первичной зрительной коре (V1), демонстрирующие настройку на полосы разной ориентации в гл. 1, и мы обсудим предпочтения нейронов более подробно в разделах 5.4 и 5.5. Основываясь на идее, что нейроны V1 возбуждаются ориентированными

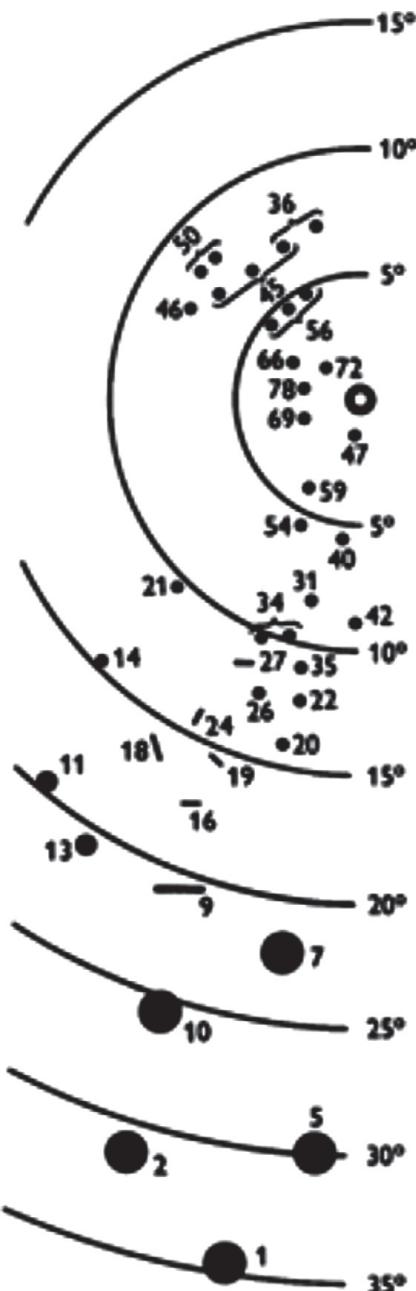
полосами, можно ожидать увидеть больше линий на рис. 4.7. Однако эти эксперименты с грубой стимуляцией, скорее всего, активируют огромное количество нейронов, охватывающих многие, если не все, ориентации. Поэтому неудивительно, что основной отчет не показывает восприятия таких особенностей.

Интересно, что восприятие, вызванное стимуляцией ранней зрительной коры головного мозга, согласуется с нашим пониманием топографической организации, полученным из описанных ранее исследований повреждений (рис. 4.3). Эта организация также согласуется с нейрофизиологическими регистрациями, которые мы обсудим в разделе 5.6. Во-первых, расположение ощущаемого фосфена в поле зрения зависит от точной области стимуляции. Эти фосфены локализованы, и это согласуется с идеей, что активируются несколько нейронов с перекрывающимися и ограниченными размерами рецептивного поля. Во-вторых, введение токов через близлежащие электроды (например, электроды 27, 31, 34 и 35 на рис. 4.7) запускает ощущения фосфена в близлежащих местах в поле зрения, как и следовало ожидать, исходя из топографической организации зрительной коры. В-третьих, приблизительный размер воспринимаемого фосфена увеличивается по мере удаления от фиксации, что согласуется с увеличением размеров воспринимающего поля в зависимости от эксцентриситета.

Следуя плодотворным исследованиям Пенфилда, несколько других исследователей использовали электрическую стимуляцию у пациентов с эпилепсией для картирования функций коры головного мозга человека. Например, исследователи описали множественные субъективные переживания, возникающие после стимуляции височной доли, включая зрительные иллюзии, как элементарные зрительные галлюцинации (фосфены), так и сложные зрительные галлюцинации. Помимо зрительных иллюзий, электрическая стимуляция височной доли вызывает множество других переживаний, включая страх, жажду, узнавание, чувство дежавю и воспоминания.

Элегантное исследование Мерфи и его коллег продолжило изучение взаимосвязи между электрической стимуляцией и нейрофизиологическими регистрациями. Они исследовали область, которая реагировала на цвета – точнее, на синий цвет, согласно нейронным регистрациям. Впоследствии они давали подопытным психофизическое задание, чтобы спросить, могут ли они определить эффекты электростимуляции. Наблюдатели сообщали о восприятии синего оттенка при электрической стимуляции, что согласуется с предсказаниями авторов, основанными на их нейрофизиологических данных.

Фосфены, зрительные галлюцинации и восприятие цвета являются примерами ощущений, вызываемых подачей тока в отсутствие сопутствующей визуальной стимуляции. Многие другие исследования были сосредоточены на оценке воздействия стимуляции на сопутствующий визуальный стимул. В таких случаях результаты показывают, что электрическая стимуляция обычно мешает реальному восприятию. Например, несколько исследований показали, что применение электростимуляции через электроды возле веретенообразной извилины искажает или ухудшает способность распознавать лица.



**Рис. 4.7 ♦** Генерация визуальных ощущений путем электростимуляции зрительной коры. Положение фосфенов (ощущений вспышки света) в поле зрения, вызванных электростимуляцией затылочной коры головного мозга человека. Центральный круг обозначает ямку, а числа обозначают электроды, через которые подавались импульсы электростимуляции. Кружки и полоски грубо обозначают размер и форму выделенных фосфенов. Воспроизведено по книге Бриндли и Левина, 1968

## 4.10. ЭЛЕКТРИЧЕСКАЯ СТИМУЛЯЦИЯ ЗРИТЕЛЬНОЙ КОРЫ ПРИМАТОВ

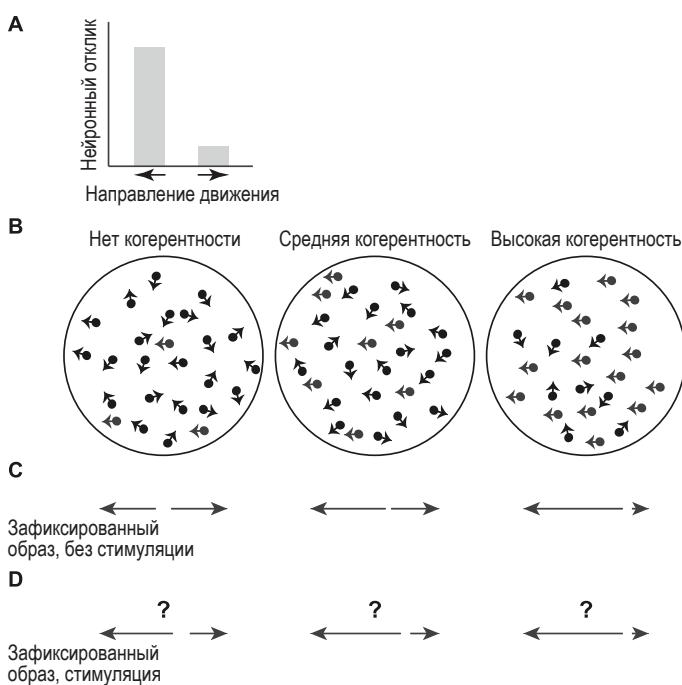
Возможность электрического стимулирования коры головного мозга человека весьма интересна, и многие наблюдения были увлекательными, но в основном анекдотичными из-за трудностей, связанных с небольшим количеством испытаний и большим числом электродов с грубым отображением нейронных ответов. Большинство исследователей использовали электрическую стимуляцию с помощью микроэлектродов, помещенных в зрительную кору головного мозга макак. Толщина микроэлектродов, применяемых в исследованиях на животных, меньше (около 50 микрон в диаметре), чем те, которые использовались на людях (около 2 мм в диаметре; микроэлектроды для стимуляции человеческого мозга применялись нечасто). Таким образом, количество нейронов, активируемых посредством электростимуляции в исследованиях на животных, хотя и очень велико, но меньше, чем количество нейронов, возбуждаемых в исследованиях на людях.

Одно из основополагающих исследований на обезьянах включало электрическую стимуляцию области МТ, которую мы представили в разделе 4.2 как область, критическую для распознавания движений при физических повреждениях мозга у обезьян, и в разделе 4.6 как вероятную область, ответственную за нарушение восприятия движения при акинетопии. Область МТ получает прямой входной сигнал из области V1 (а также входные сигналы из других областей, таких как V2), исходящий от магноклеточных слоев в LGN. Нейроны в области МТ избирательны в отношении направления движения в рецептивном поле; например, нейрон может сильно реагировать на полосу, перемещающуюся влево, но не на полосу, перемещающуюся вправо (рис. 4.8А).

Типичный стимул, используемый для возбуждения МТ-нейронов, представляет собой дисплей с изображением, представляющим множество точек, движущихся в случайных направлениях (рис. 4.8В). Определенный процент точек настроен на последовательное движение в одном направлении. В зависимости от процента когерентного движения стимул может вызвать сильное восприятие движения. Получается типичная симмоидная психометрическая кривая (как у людей, так и у обезьян), показывающая долю испытаний, в которых подопытный сообщает, что точки движутся в одном направлении, в зависимости от степени согласованности движения точек на показываемом изображении. Если 100 % точек движутся вместе в одном направлении, наблюдатели сообщают о движении в этом направлении во всех испытаниях. Если 0 % точек движутся когерентно (все точки движутся случайным образом), то подопытные сообщают о случайному движении в том или ином направлении (рис. 4.8С).

Команда Уильяма Ньюсома из Стэнфорда обучила обезьян сообщать о предполагаемом направлении движения, одновременно регистрируя активность нейронов в области МТ. Регистрируя активность нейронов в области МТ, исследователи начинали эксперимент, картируя предпочтительные

направления движений для возбуждения нейронов. В типичном эксперименте на экране появляется точка фиксации, и обезьяна должна зафиксироватьсья на ней; визуальный стимул показывается в течение 1 с, затем исчезает, и обезьяне необходимо указать направление, в котором двигались точки, в двух вариантах парадигмы принудительного выбора (например, совершая саккаду на одну из двух возможных целей). Направление движения в каждом эксперименте совпадает с предпочтительным направлением нейрона, так что точки перемещаются либо в направлении предпочтения нейронов, либо в противоположном.



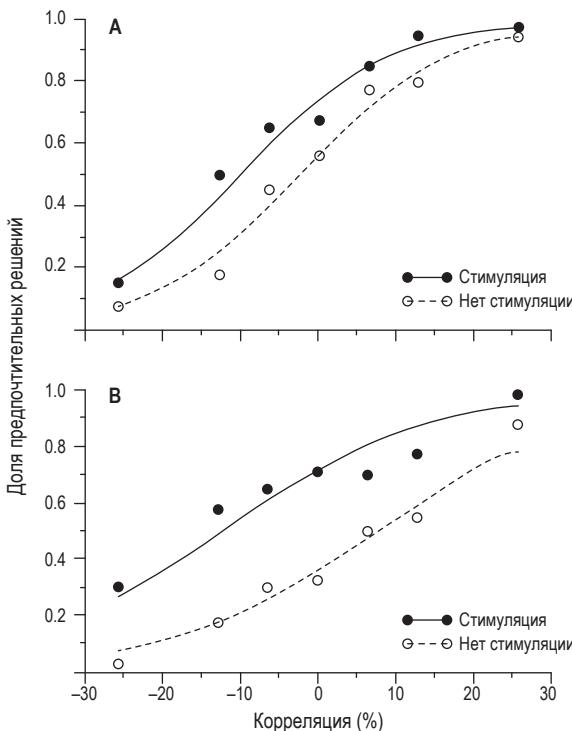
**Рис. 4.8 ♦** Схематическое представление эксперимента по электростимуляции в зоне МТ обезьяны-макаки. (A) Схематические ответы нейрона, который избирательно реагирует на движение влево. (B) Случайные точечные стимулы без когерентности (слева), с высокой когерентностью движения влево (справа) и средней когерентностью движения влево (в центре). (C) Зафиксированный образ в отсутствие электростимуляции. (D) Прогнозируемый образ при наличии стимуляции нейронов вокруг того, что показано в части А. Основано на работе

Зальцман и др., 1990

Основываясь на нейрофизиологических регистрациях, исследователи спросили, будет ли электрическая стимуляция по тому же микропроводу искажать визуально обоснованное поведение обезьяны в задаче различия движений и согласуется ли это искажение с предпочтениями нейронов. Чтобы ответить на этот вопрос, они применили очень короткие электрические импульсы (двухфазные прямоугольные импульсы с силой

тока 10 мкА, частотой 200 Гц и длительностью 0,2 мс). Электрическая стимуляция применялась в центре областей, где находился кластер нейронов с похожими предпочтениями движения в пределах ~150 мкм. Как и в других частях неокортекса, в области МТ существует топографическая организация предпочтений нейронов; то есть соседние нейроны в МТ обычно имеют сходные предпочтения к направлению движения. Эта топография, по-видимому, важна с точки зрения понимания эффектов электрической стимуляции, потому что активация многих локальных нейронов с аналогичными свойствами избирательности может привести к более сильным поведенческим эффектам, чем активация нейронов, которые пространственно организованы совершенно случайным образом в отношении их предпочтений. За правильные ответы обезьян вознаграждали. Результаты этих экспериментов показаны на рис. 4.9. В отсутствие микростимуляции (пустые кружки) обезьяны демонстрировали типичную приблизительно сигмоидную психометрическую кривую. Обезьяны сообщили о предпочтительном направлении движения в более чем 80 % испытаний, когда точки имели 30%-ную корреляцию в направлении предпочтения нейронов, и они сообщили о противоположном направлении движения в более чем 80 % испытаний, когда точки имели 30%-ную корреляцию в противоположном направлении. В условиях нулевой корреляции обезьяны сообщали о том и другом направлениях с вероятностью, близкой к 50 %. Примечательно, что при применении электрической стимуляции (заштрихованные кружки) произошел сдвиг психометрической кривой. Обезьяны сообщали о движении в предпочтительном направлении чаще (на 15 % чаще), чем при отсутствии электростимуляции. Это причинное усиление из-за электрической стимуляции является важным открытием, поскольку оно предоставило убедительные доказательства того, что нейрофизиологические регистрации выявили сигнал, который может быть преобразован в поведенческие решения при электрической стимуляции соответствующих нейронных цепей.

В аналогичном эксперименте Араш Афраз и его коллеги стимулировали нижнюю височную кору макак во время задачи зрительного распознавания. Исследования на повреждениях показывают, что область ИТС важна для зрительного распознавания образа (раздел 4.7). Эксперимент соответствовал структуре исследования Ньюсома на рис. 4.8 и 4.9. Поскольку нейроны в ИТС больше заинтересованы в сложных визуальных образах, а не в направлении движения, исследователи сравнивали реакции на лица с реакциями на другие образы. Выбор лиц в качестве одного из двух стимулов был важным методологическим моментом. Во-первых, может быть, легче научить обезьян распознавать двумерные изображения лиц по сравнению с другими образами. Во-вторых, может быть, существует больший кластер нейронов, реагирующих на образы, похожие на лица, по сравнению с другими образами. Исследователи представили лица и другие изображения без лица, залитые шумом. Уровень шума изменялся со 100 % (чисто шумовой стимул) до 20 %; шум в этом эксперименте играет такую же роль, как когерентность в экспериментах Ньюсома.



**Рис. 4.9 ♦** Результаты эксперимента по электростимуляции в зоне МТ макаки. Графики показывают поведенческую психометрическую функцию в распознавании предпочтительного направления движения нейронов в присутствии (заполненные кружки) или отсутствии (пустые кружки) электрической стимуляции (подробности см. в тексте). Воспроизведено по Зальцман и др., 1990

Как мы обсудим в разделе 6.2, нейроны ИТС демонстрируют визуально избирательные ответы; исследователи сосредоточились на участках, которые показали последовательное усиление реакции на лица в пределах области стимуляции приблизительно  $\pm 150$  мкм. Исследователи применили электрическую стимуляцию к этим областям и оценили пороговую величину шума, при которой обезьяны переставали различать лица. В среднем исследователям удавалось выявить примерно 10%-ное изменение в поведении обезьяны, увеличивающее количество раз, когда обезьяны сообщали о том, что видели лица (даже в тех случаях, когда информация о лицах была минимальной из-за шума). Кроме того, поведенческие эффекты, вызванные электрической стимуляцией, коррелировали со степенью избирательности нейронов: стимуляция более избирательных участков приводила к более сильным поведенческим искажениям. В целом в разных областях зрительной коры введение токов во многие нейроны, которые проявляют избирательность в отношении определенных свойств стимула, может смешать реакцию обезьяны в сторону сообщений о том, что они видят особенности, связанные с предпочтениями нейронов. Даже в тех случаях, когда стимул состоит из случайного шума, можно изменить поведение таким образом, что можно

достоверно предсказать соответствующий нейрофизиологический ответ. Эти эксперименты обеспечивают прочную причинно-следственную связь между специфической и избирательной нейронной активностью и зрительным восприятием. Кроме того, эти эксперименты представляют собой интригующую форму введения в мозг определенных визуальных мыслей.

## 4.11. ОБЗОР ГЛАВЫ

- Инактивация областей зрительной коры приводит к специфическим зрительным дефектам, начиная от локализованных скотом (первичная зрительная кора) и заканчивая нарушением распознавания сложных образов (нижняя височная кора).
- Без первичной зрительной коры наблюдатели практически слепы. Остаются только очень ограниченные базовые зрительные способности.
- Исследования повреждений определили два основных потока обработки зрительного сигнала: (1) *дорсальный путь* («где?»), который особенно важен для обнаружения движения, интерпретации местоположения стимула и пространственного воздействия на зрительные стимулы, и (2) *центральный путь* («что?»), который более важен с распознанием цветов и форм.
- Хотя повреждения мозга у людей трудно полностью интерпретировать из-за их редкости и случайного характера, они выявили множество интересных наблюдений, отображающих зрительные дисфункции в локализованных цепях мозга.
- Сообщалось о нескольких случаях *агнозии*, когда у субъектов возникали определенные проблемы с распознаванием образов при сохранении нормального зрения в остальном.
- Электрическая стимуляция в ранней зрительной коре человека приводит к восприятию *фосфенов*. Расположение и размер этих фосфенов соответствуют нашему пониманию топографической организации ранней зрительной коры.
- Стимуляция других частей зрительной коры головного мозга человека при одновременном предъявлении визуального стимула может привести к специфическим нарушениям восприятия.
- Эксперименты по микроэлектродной стимуляции на обезьянах показали, что можно изменить поведение животного таким образом, чтобы достоверно предсказывать нейрофизиологические ответы нейронов в стимулируемой области.

## ЛИТЕРАТУРА

См. <http://bit.ly/3abKVpP> для получения дополнительных ссылок.

- Boyden, E. S.; Zhang, F.; Bamberg, E.; Nagel, G.; and Deisseroth, K. (2005). Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neuroscience* 8: 1263–1268.

- Penfield, W. (1958). Some mechanisms of consciousness discovered during electrical stimulation of the brain. *Proceedings of the National Academy of Sciences of the United States of America* 44: 51–66.
- Sacks, O. (1995). *An anthropologist on Mars*. New York: Alfred A. Knopf.
- Sperry, R. (1982). Some effects of disconnecting the cerebral hemispheres. *Science* 217: 1223–1226.
- Ungerleider, L.; and Mishkin, M. (1982). Two cortical visual systems. In *Analysis of Visual Behavior*, ed. D. Ingle, M. Goodale, and R. Mansfield. Cambridge: MIT Press.

Дополнительный контент вы можете скачать по ссылке <http://bit.ly/3abKBpP>.

# Глава 5

---

## Приключения в *terra incognita*. Исследование нейронных цепей вентрального зрительного потока

Примерно в 1950-х годах многочисленные поведенческие эксперименты дали характеристику многим феноменологическим особенностям зрительного восприятия, требующим механистического объяснения (гл. 3). Исследования повреждений предоставили убедительные доказательства того, что повреждение ограниченных областей мозга приводит к определенным дефицитам визуальной обработки (гл. 4). Эти исследования указали на определенные области мозга, участвующие в процессе обработки изображений, особенно на первичную зрительную кору в задней части мозга. Кроме того, успешное использование микроэлектродных электрических регистраций привело к непосредственному пониманию функции нейронов в цепях сетчатки (гл. 2). Пришло время открыть черный ящик мозга и начать думать о том, как возникает феномен зрения в результате спайковой активности нейронов коры головного мозга<sup>1</sup>.

Ганглиозные клетки сетчатки проецируются в латеральное коленчатое ядро (LGN) в таламусе, а основная выходная проекция LGN передает визуальную информацию в первичную зрительную кору (зона V1; см. раздел 2.10), первую стадию кортикалной обработки визуальной информации. Из зоны V1 информация распространяется в большое количество зрительных областей коры, которые отвечают за преобразование пиксельного представле-

---

<sup>1</sup> Спайки – пики возбуждения, видимые как острые всплески на записи регистрируемого сигнала нейрона. – Прим. ред.

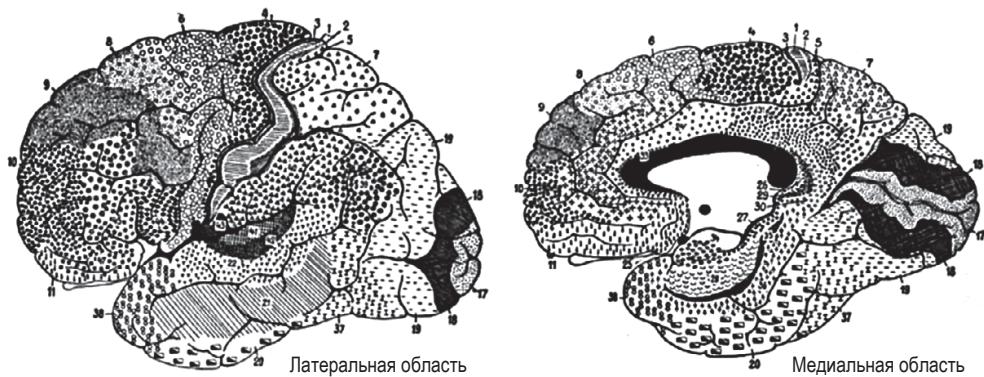
ния сенсорной информации фоторецепторов в богатые и сложные зрительные образы (гл. 1, рис. 1.5). Исследование и компьютерное моделирование зрительной коры головного мозга – это продолжающееся приключение, в котором отважные конкистадоры осмеливаются заглянуть внутрь самой сложной системы, когда-либо становившейся предметом научных исследований. Фундаментальные структурные и функциональные принципы этих вычислений начинают проявляться из часто кажущейся загадочной *terra incognita* зрительной коры. Эти основные принципы представлены в этой и следующей главах и составляют основу вычислительных моделей зрения, обсуждаемых в гл. 7–9.

## 5.1. О НЕОКОРТЕКСЕ

Неокортекс – это внешняя структура нервной ткани в головном мозге, которая, как считается, отвечает за распознавание. Приставка « neo » означает «новое», что следует понимать в эволюционных временных масштабах и контрастирует со старым палеокортексом, который включает обонятельную систему и гиппокамп. Неокортекс человека имеет толщину порядка 2–4 мм, составляет около 40 % массы мозга и содержит порядка  $10^{10}$  нейронов. Кора головного мозга показывает большое количество складок, так что она имеет площадь около 26 квадратных дециметров. Из-за его обширной поверхности и относительно небольшой глубины многие исследователи считают неокортекс квазидвумерной структурой. Самая заметная складка – это продольная трещина, разделяющая правое и левое полушария. Неокортекс человека имеет больше складок, чем у многих других млекопитающих; например, кора головного мозга мыши кажется относительно гладкой по сравнению с корой головного мозга человека. Механическое давление в сочетании с сильным ограничением по экономии коммуникаций и пространства, вероятно, было важным фактором в определении формы и складок коры на протяжении всей эволюции.

В довольно разумном приближении первого порядка *кора головного мозга*, или *корtekс*, – это именно кора головного мозга: окрашенные корковые ткани разных участков мозга очень похожи. Кроме того, окрашенные части коры головного мозга очень похожи даже у разных видов. Требуется знаток, чтобы отличить срез кортикальной ткани мыши от кортикальной ткани человека. Это сходство, возможно, для некоторых людей будет неожиданным. Эгоцентрические или антропоморфные соображения могут привести некоторых людей к мысли, что кора головного мозга человека существенно отличается от коры животных; в конце концов, мыши не играют в шахматы и не читают Шекспира. Грубое сходство в основной корковой структуре предполагает, что примерно одни и те же аппаратные средства можно комбинировать разными и интересными способами, чтобы объяснить когнитивные способности разных видов. В качестве грубой аналогии аналогичные транзисторы можно использовать для создания электронного калькулятора, смартфона и ноутбука.

При более глубоком исследовании кортикальных тканей специалисты могут различить кору разных видов животных. Кроме того, исследуя кору, можно также различить разные области мозга. Немецкий нейроанатом Корбинян Бродманн (1868–1918) разработал картирование мозга человека и обезьяны, а также многих других видов, только на основе морфологических цитоархитектонических признаков. Многие части неокортекса все еще обозначаются по номеру области Бродмана (рис. 5.1). Например, первичная зрительная кора V1 соответствует зоне 17 Бродмана. Нейрофизиологические исследования и исследования повреждений показали, что некоторые структурные подразделения, предложенные Бродманном, а также последующие нейроанatomические исследования коррелируют с функциональной специализацией. Попытки определить функциональность отдельных областей коры, особенно в сочетании с попытками привязать когнитивные функции к разным областям, имеют долгую и богатую историю, которая пишется до наших дней.



**Рис. 5.1** ♦ Кора головного мозга может быть разделена на несколько областей на основе цитоархитектонических признаков. В соответствии с ними Бродманн разделил неокортекс на несколько областей. Первичная зрительная кора соответствует зоне Бродмана 17, как показано на этой диаграмме. Согласно Википедии, исследования отдельных областей коры головного мозга в сочетании с попытками привязать когнитивные функции к различным областям имеют долгую и богатую историю, которая продолжается писаться до наших дней

## 5.2. СВЯЗЬ С ПЕРВИЧНОЙ ЗРИТЕЛЬНОЙ КОРОЙ И ОБРАТНАЯ СВЯЗЬ

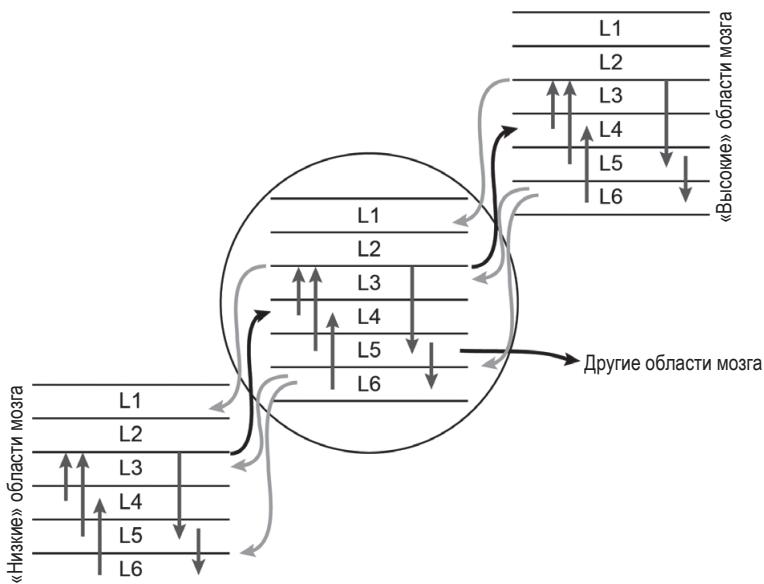
Первичная зрительная кора – это первый этап, на котором информация от двух глаз сходится к отдельным нейронам. Каждое полушарие в V1 представляет контрлатеральное поле зрения. Часть сетчатки, которая находится ближе к носу, называется *носовой*, а другая часть сетчатки – *височной*. Левая часть поля зрения (слева от центра фиксации) представлена носовой

частью сетчатки левого глаза и височной частью сетчатки правого глаза. Информация из носовой сетчатки левого глаза проходит через мозг и в конечном итоге представляется в правом полушарии в первичной зрительной коре. Информация от височной сетчатки на правом глазе перенаправляется в *перекрестье зрительных нервов*, а также в конечном итоге представляется в правом полушарии в первичной зрительной коре.

Как и большинство других аспектов нейроанатомии, первые рисунки первичной зрительной коры были сделаны Сантьяго Рамоном-и-Кахалем, о котором упомянуто в гл. 2. Базовая архитектура первичной зрительной коры оказалась примерно такой же, как и у других частей зрительного неокортекса. Неокортикальный пласт характеризуется шестью слоями, которые можно различить с помощью окрашивания по Нисслю, техники введения красителя в некоторые нейроны в данной области для различения ее структуры. Исключительное окрашивание здесь важно, потому что плотность нейронов в коре настолько велика, что при окрашивании всех нейронов с помощью стандартного окрашивания при микроскопии будет трудно что-либо различить. Шесть слоев характеризуются стереотипной структурой связности, которую часто называют *канонической кортикальной микросхемой*. За некоторыми исключениями – в конце концов, это биология, – эта каноническая схема связи характерна для разных визуальных областей, а также для различных сенсорных модальностей.

Связи между различными областями коры головного мозга часто описываются как «восходящие», «нисходящие» или «горизонтальные» – номенклатура, которая также используется для описания связности в архитектурах искусственных нейронных сетей (раздел 7.3, рис. 7.4). Отдельный нейрон будет проецировать только снизу вверх, или горизонтально, или давать нисходящие сигналы, но не все из них. Эти различные типы связей определяются на основе специфического слоя пре- и постсинаптических нейронов. Связи между зрительными корковыми областями и внутри них следуют стереотипной схеме, которая определяет то, какая область находится «выше по течению» или «ниже по течению», и, следовательно, какие связи являются восходящими или нисходящими (рис. 5.2). Соединения снизу вверх достигают уровня 4 – область LGN проецируется на пирамидные нейроны в слое 4 первичной зрительной коры. Слой 1 является наиболее поверхностным и содержит в основном дендриты и несколько тел нейронных клеток; тела нейронных клеток этих дендритных ветвей в основном расположены в слоях 2 и 3. Нисходящие соединения из других зрительных областей коры обычно заканчиваются в глубоких слоях 5 и 6, а также, в меньшей степени, в слоях 2 и 3. После того как вход LGN (или вход из «нижней области») поступает на уровень 4, информация перетекает из уровня 4 в слои 2 и 3, а затем на уровень 5 и уровень 6. Информация из уровня 6 обеспечивает обратные проекции в LGN (или в «нижнюю» визуальную область), а также передается обратно в слой 4.

Важным аспектом связности зрительной коры является то, что связи между областями почти всегда взаимообратны. Если область А обеспечивает ввод снизу вверх в область В, область В обеспечивает ввод сверху вниз в область А. Кроме того, эти взаимообратные связи количественно сопоставимы: количество проекций от А к В и от В к А примерно одинаково.



**Рис. 5.2** ♦ Канонические кортикальные цепи. Корковые связи через зрительную кору головного мозга следуют стереотипным схемам связности, проиллюстрированным здесь. L1–L6 относятся к шести кортикальным слоям. Связь «снизу вверх» между областями показана черным, связь «сверху вниз» между областями – светло-серым цветом, а соединения внутри области показаны средним серым цветом

Изучая схемы связи между слоями в нескольких областях мозга, исследователи составили приблизительную карту анатомических путей, посредством которых различные зрительные области взаимодействуют друг с другом (гл. 1, рис. 1.5). Основываясь на разделении связей на восходящие и нисходящие, можно организовать несколько различных визуальных областей мозга в приблизительную иерархическую структуру. Схема на рис. 1.5 представляет собой некоторое подобие иерархического описания анатомического потока информации в зрительной системе.

Чем больше мы изучаем взаимосвязи в зрительной коре, тем больше понимаем, что эта в общем стереотипная схема полна исключений. Существуют различия между видами, различия между зрительной корой и моторной корой и даже различия между разными зрительными кортикальными областями. Что еще больше усложняет ситуацию, эти слои, в свою очередь, могут быть подразделены на собственные слои, а шаблоны связности могут быть разными в зависимости от типов рассматриваемых нейронов. Например, мы начали этот раздел с утверждения, что первичная зрительная кора *примерно* похожа на другие зрительные области коры. Возможно, из-за своего уникального положения в получении большего количества прямых таламических входов, чем все другие визуальные области, V1 на самом деле толще, слой 4 имеет разное количество подслоев, а характер его входных и выходных сигналов также отличается от других зрительных областей.

В дополнение к вариациям в канонической схеме соединений между областями коры и между разными видами иерархическая природа зрительной коры не должна интерпретироваться слишком строго. Например, многочисленные «обходные» соединения отправляют информацию из области А в область С, не проходя через промежуточную область В (например, потоки информации идут от V1 к V2 и далее к V4, но есть также прямые соединения от V1 к V4). Несмотря на разницу, исключения и уточнения, основные принципы взаимодействия в зрительной коре сыграли важную роль в описании случаев очевидного безумия и вдохновили на разработку лучших вычислительных моделей, которые у нас есть сегодня (гл. 7 и 8).

Уместно сделать предостережение относительно номенклатуры, особенно для компьютерных ученых, привыкших думать о нейронных сетях. Биологи говорят о разных *областях* коры головного мозга, таких как V1, V2 и V4. Каждая из этих областей состоит из шести *слоев*, как описано ранее. В гл. 7 и 8 мы обсудим вычислительные модели визуальной обработки, которые часто относятся к вычислительным шагам, реализованным на разных *уровнях* (слоях). Эти вычислительные слои не следует путать с корковыми слоями, описанными здесь. Слой в нейронной сети не обязательно имеет прямой аналог в одном из шести слоев неокортикса в любой данной области мозга. Точное отображение между вычислительными слоями и областями мозга не всегда учитывается разработчиками моделей. Фактически во многих случаях люди думают о слое в нейронной сети как о потенциально эквивалентном целой области мозга в коре головного мозга. Мы обязательно вернемся к вопросу о сопоставлении вычислительных моделей с анатомией биологического объекта. На данный момент здесь мы имеем в виду слои в биологическом смысле, рассмотренном в предыдущем абзаце и на рис. 5.2. В дополнение к информации, передаваемой от одного слоя к другому в пределах визуальной области мозга, и информации, передаваемой между областями мозга, существуют обширные *горизонтальные связи*, посредством которых информация перемещается *внутри* слоя. Некоторые исследователи используют термин *рекуррентные соединения* (сети с обратной связью. – Прим. ред.) для обозначения как горизонтальных, так и нисходящих соединений, но концептуально более ясно использовать разные термины для этих двух различных типов трактов прохождения сигнала.

## 5.3. ЗОЛОТОЙ СТАНДАРТ ВРЕМЕННОГО РАЗРЕШЕНИЯ ДЛЯ ИССЛЕДОВАНИЯ НЕЙРОНОВ

У каждой задачи есть соответствующая шкала изучения: шкала Златовласки, так сказать – не слишком грубая, не слишком тонкая. Например, очень утомительно и сложно читать газету с помощью микроскопа (слишком высокое разрешение). Также крайне сложно читать газету на расстоянии 200 метров (слишком плохо видно). Для изучения мозга доступно множество методов, от выяснения трехмерной структуры определенных типов ионных каналов до

косвенного измерения сигналов, которые показывают некоторую степень корреляции с кровотоком, усредненным по грубому пространственному масштабу.

В случае неокортикальных цепей шкала Златовласки – это активность отдельных нейронов. Изучение трехмерной структуры каждого белка внутри нейрона эквивалентно попытке прочитать газету под микроскопом, но это может быть чрезвычайно полезно для других вопросов, таких как понимание кинетики и свойств ионных каналов в нейронной мембране. Изучение среднего количества крови, протекающей через половину кубического сантиметра коры головного мозга за несколько секунд, эквивалентно попытке различить тона типографской краски газеты на расстоянии 200 м, но это может быть полезно для других вопросов, таких как дифференциация общих и грубых свойств области коры головного мозга.

В дополнение к этой пространственной шкале для изучения нейронной активности существует также естественная временная шкала. Большинство нейронов общаются друг с другом, посылая электрические сигналы, называемые потенциалами действия, длительностью около 2 мс. Для большинства исследований достаточно изучить активность нейронов на уровне миллисекунд. За некоторыми исключениями (например, при очень малой разнице во времени между звуками, поступающими в левое и правое уши), использование временной шкалы с микросекундным разрешением не даст нам никакой дополнительной информации. В сутках 1440 минут, и, следовательно, аналогия изучения мозга в микросекундах вместо миллисекунд (множитель 1000) – это примерно как перечитывать одну и ту же газету каждую минуту. На другом конце спектра – методы, которые усредняют активность за несколько секунд, которые слишком грубы, чтобы прояснить вычисления в коре головного мозга. По аналогии с изучением мозга в масштабе нескольких секунд вместо миллисекунд (коэффициент 1000) можно было бы усреднить газету за три года.

Изучение активности неокортикальных цепей при разрешении активности нейронов в масштабе миллисекунд нетривиально и требует введения тонких микроэлектродных зондов прямо в интересующие области. Потенциалы действия нейронов приводят к изменениям электрического потенциала во внеклеточной среде. Можно усилить и измерить этот электрический потенциал во внеклеточном пространстве и измерить потенциалы действия, испускаемые отдельными нейронами. Методология была разработана Эдгаром Адрианом (1889–1977), и мы уже представили примеры измерения активности отдельных нейронов сетчатки в разделе 2.7.

## 5.4. НЕЙРОНЫ В ПЕРВИЧНОЙ ЗРИТЕЛЬНОЙ КОРЕ ГОЛОВНОГО МОЗГА ИЗБИРАТЕЛЬНО РЕАГИРУЮТ НА ПОЛОСЫ, ПОКАЗАННЫЕ В РАЗНОЙ ОРИЕНТАЦИИ

Первичная зрительная кора человека состоит примерно из 280 млн нейронов, расположенных в виде листа толщиной 2 мм, который занимает площадь в несколько десятков квадратных сантиметров. Существует больше

работ, посвященных нейрофизиологии первичной зрительной коры, чем о всех других областях зрительной коры, вместе взятых. Нейроны в первичной зрительной коре, в сетчатке и LGN (раздел 2.7), а также нейроны в других областях зрительной коры демонстрируют пространственно-ограниченные рецептивные поля; то есть они реагируют только на определенную часть зрительного поля (рис. 2.9). Ансамбль всех нейронов покрывает все визуальное поле. В среднем размер рецептивного поля нейронов в первичной зрительной коре больше, чем размер рецептивного поля в сетчатке и LGN, и обычно составляет от 0,5 до 1° угла зрения. Типичный нейрофизиологический эксперимент часто начинается с определения местоположения рецептивного поля исследуемого нейрона. После определения местоположения рецептивного поля используется батарея стимулов для исследования предпочтений нейронного отклика.

Первые результативные шаги, обеспечившие значительное продвижение в понимании нейрофизиологических реакций в первичной зрительной коре, были сделаны Торстеном Визелем (1924 г. р.) и Дэвидом Хьюбелом (1926–2013). История нейробиологии зрения вращается вокруг истории зрительного стимулирования. До эры Хьюбеля–Визеля исследователи изучали реакции первичной зрительной коры с использованием рассеянного света или точечных источников, которые успешно вызывали активность в сетчатке и LGN. Путем сочетания вдохновения и тщательного наблюдения, в поте лица, Хьюбел и Визель определили, что нейроны первичной зрительной коры наиболее сильно реагируют, когда в рецептивном поле нейрона представлена полоса определенной пространственной ориентации. Очень увлекательная история этого открытия была изложена в Нобелевской лекции Дэвида Хьюбела. Хьюбел и Визель не имели особо смелых гипотез о функции нейронов в зрительной коре до того, как приступили к этим исследованиям, скорее, у них были интуитивные догадки, что убедительные результаты будут получены, если смело помещать электроды в область V1. После особенно долгого дня регистрации активности нейрона в V1 они были разочарованы тем, насколько мало нейрон, казалось, заботился о наличии светлого или темного кольца внутри рецептивного поля. В те дни у них не было компьютеров для представления стимулов; вместо этого они использовали слайды, вставленные в проектор. Внезапно их наблюдательность привела исследователей к осознанию того, что нейрон дает всплеск активности каждый раз, когда они вставляют слайд в проектор. Край слайда, входящий и выходящий из проектора, вызывал активацию гораздо больше, чем содержание слайда. Взволнованные этим открытием, они продолжили эксперименты и обнаружили, что ориентация края, помещенного в рецептивное поле, имеет значение для нейрона: определенные углы ориентации приводят к гораздо большей активации, чем другие.

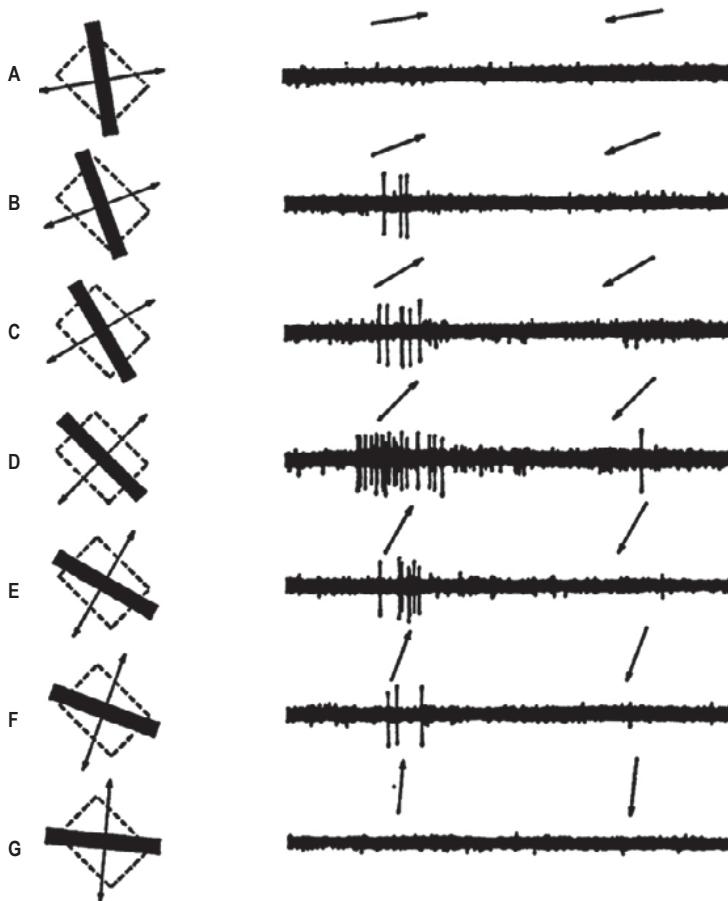
Типичный образец ответов нейронов, полученных в регистрациях V1, показан на рис. 5.3. В этом эксперименте ориентированная определенным образом полоска перемещалась в пределах рецептивного поля исследуемого нейрона. Направление движения было перпендикулярно ее ориентации. Различная ориентация вызвала резко различающееся количество потенциалов действия. Хотя количество потенциалов действия (или число спайков) – не

единственная переменная, которую можно использовать для определения ответа нейрона, оно обеспечивает простую и адекватную отправную точку для изучения предпочтений нейронов. Когда полоска находилась примерно под углом  $-45^\circ$  (рис. 5.3D), нейрон испускал больше импульсов, чем при любой другой ориентации. Более того, активность этого нейрона зависела также и от направления движения. Когда полоса двигалась к верхнему правому углу, нейрон был очень активен, тогда как в противоположном направлении движения активация была минимальной.

Далее Хьюбел и Визель охарактеризовали свойства нейронов V1 с точки зрения их топографии, ориентации, оптических предпочтений, цвета, направления движения и даже того, как эти свойства возникают и развиваются во время единичного эксперимента. Их открытие, получившее Нобелевскую премию, вдохновило поколения нейрофизиологов на изучение нейронных реакций во всей зрительной коре.

## 5.5. СЛОЖНЫЕ НЕЙРОНЫ ДЕМОНСТРИРУЮТ ТОЛЕРАНТНОСТЬ К ИЗМЕНЕНИЯМ ПОЛОЖЕНИЯ СТИМУЛОВ

В примере, показанном на рис. 5.3, нейрон V1 избирательно реагирует на движущуюся полосу. Нейроны в V1 также реагируют на вспышки статических стимулов. Насколько точным должно быть положение ориентированной полосы в рецептивном поле нейрона при подаче стимула, чтобы вызвать определенный отклик? Наблюдается различие между двумя типами нейронов в V1 в зависимости от того, насколько они придиличны к положению стимула в рецептивном поле: *простые* и *сложные* нейроны V1. Сложные нейроны менее чувствительны к точному положению полоски в рецептивном поле. При использовании решеток, содержащих несколько ориентированных полос с заданной пространственной периодичностью, сложные нейроны допускают большие изменения длины этого периода, чем простые клетки. Простые и сложные нейроны часто различаются отношением поддерживающего «постоянного тока» (DC) к их «переменному» ответу (AC), вызванному движущейся решеткой. Сложные нейроны показывают небольшое отношение AC/DC (обычно меньше 10), тогда как простые нейроны имеют большее отношение AC/DC (обычно больше 10). Другими словами, сложный нейрон демонстрирует более высокую степень *толерантности* к точному положению ориентированной полоски внутри рецептивного поля по сравнению с простым нейроном, величина реакции которого уменьшается, когда полоска смещается от предпочтительного положения (рис. 5.4). Этот переход от простого нейрона к сложному нейрону, демонстрирующему повышенную толерантность, вдохновил на разработку иерархических вычислительных моделей распознавания объектов, которые объединяют операции, подобно простым и сложным нейронам, как способ сохранения избирательности при достижении толерантности к изменению стимула (раздел 8.5).

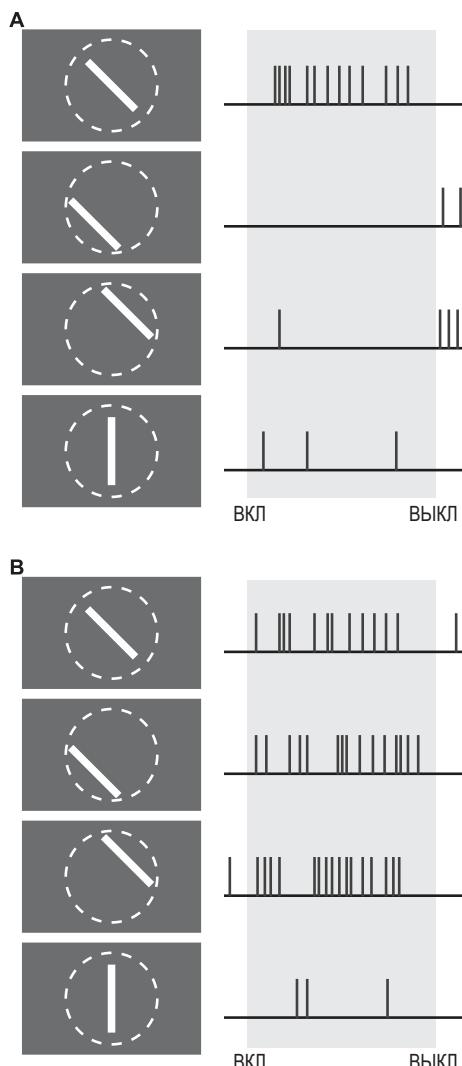


**Рис. 5.3** ♦ Пример ответа нейрона первичной зрительной коры головного мозга обезьяны. Физиологические реакции нейрона первичной зрительной коры на полосы разной пространственной ориентации. В этих примерах полоса перемещалась в направлении, перпендикулярном ее ориентации. Пунктирные линии слева указывают на воспринимающее поле, черный вытянутый прямоугольник – это ориентированная полоса, а стрелки указывают направление движения. Справа показаны записи ответа нейронов. Воспроизведено по работе Хьюбела и Визеля, 1968

Некоторые сложные нейроны также показывают «конечную остановку», что означает, что стимул является для них оптимальным, если он вместе с его концами находится полностью внутри рецептивного поля, в отличие от очень длинных полосок, конец которых расположен за его пределами<sup>1</sup>. Явление конечной остановки может быть понято как форма контекстной модуляции, когда образы в области, окружающей рецептивное поле (в данном случае продолжается линия на его крае или обрывается), влияют на реакцию на стимул внутри рецептивного поля. Такие влияния извне рецептивного поля не ограничива-

<sup>1</sup> Этот тип нейрона дает максимальный отклик в конце полосы и используется для уточнения положения ее края. – Прим. ред.

ются конечной остановкой. Нейроны V1 также демонстрируют ингибицию окружения, подобное ингибирующим эффектам света вокруг центра рецептивного поля для центральных ганглиозных клеток сетчатки, описанных в разделе 2.8 (рис. 2.10). В общем, нейроны V1 особенно чувствительны к пространственным изменениям, обнаруживая нарушение непрерывности визуального поля, а некоторые нейроны также обнаруживают, где находится ее край.



**Рис. 5.4** ♦ Сложные нейроны демонстрируют устойчивость к изменениям положения. (A) Схематическая диаграмма, показывающая ответы простого нейрона, который максимально реагирует на линию, ориентированную на  $-45^\circ$ , когда он расположен в центре рецептивного поля (вверху), но не при смещении позиции (строки 2, 3), или когда ориентация меняется (внизу). (B) Схематическая диаграмма, показывающая отклики сложного нейрона, который демонстрирует устойчивость к изменениям положения

## 5.6. Соседствующие нейроны проявляют схожие свойства

Нейроны первичной зрительной коры топографически организованы аналогично ситуации, описанной для сетчатки в разделе 2.7. Топография V1 унаследована от LGN: связи от LGN с первичной зрительной корой топографически организованы, что означает, что соседние нейроны в LGN отображаются на соседние нейроны в первичной зрительной коре. Нейроны V1 покрывают поле зрения с гораздо большей плотностью нейронов в фoveальной области. Эти нейрофизиологические наблюдения согласуются с типами скотом, наблюдавшим в случаях локализованных повреждений V1 (раздел 4.4), а также с местоположениями фосфенов, о которых сообщается в исследованиях при стимуляции в V1 (раздел 4.9).

Хьюбел и Визель открыли еще один аспект топографического размещения нейронов в V1, сравнив настройки предпочтений разных нейронов, записанные во время одного и того же сеанса использования электрода. Помимо общих свойств со своими двумерными соседями по кортикальному листу, нейроны также проявляют аналогичные виды реакций со своими соседями в третьем измерении, на другой глубине коры. Продвигая электрод в направлении, приблизительно касательном к поверхности коры, разные нейроны вдоль линии проникновения разделяют схожие предпочтения к ориентации стимула. Это наблюдение привело к понятию *столбчатой структуры*: нейроны в столбце имеют схожие предпочтения; нейроны в соседних столбцах демонстрируют непрерывное изменение своих предпочтений настройки ориентации.

Такая топография может иметь решающее значение для экономии коммуникаций, поскольку нейроны со схожими свойствами должны располагаться рядом друг с другом. В частности, интернейроны с короткими дендритами могут нуждаться в наличии своих мишенией поблизости. Однако если мы сохраним связь между нейронами нетронутой, мы, в принципе, могли бы изменить геометрию нейронов произвольным образом, сохранив при этом скорость вычислений. Таким образом, топография может быть в основном отделена от функции. Следовательно, плавное отображение свойств настройки в V1, вероятно, не является требованием к вычислениям V1. Недавние исследования показали, что такой уровень организации не может быть универсальным свойством. Первичная зрительная кора у мышей не имеет такого точного топографического отображения ориентационных предпочтений; геометрическое расположение настроек предпочтений описывается как «соль и перец».

Даже если эффективность вычислений не требует строго такой топографии, она может оказаться весьма полезной для исследователей. Например, методы регистрации с пониженным пространственным разрешением, которые усредняют активность многих нейронов, могут сильно зависеть от топографии (потому что усредненные ответы от нейронов, расположенных совершенно случайно, могут ничего не дать). По аналогичным причинам,

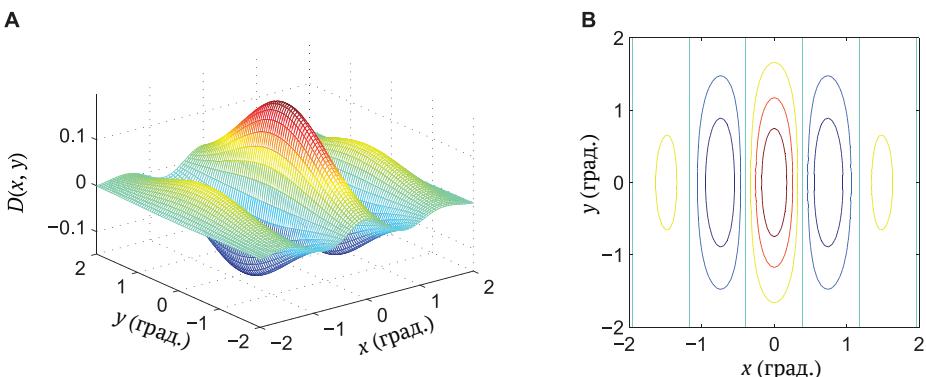
как обсуждалось в разделе 4.10, стимуляция многих нейронов с помощью введения тока также может зависеть от топографии.

## 5.7. КОЛИЧЕСТВЕННОЕ ФЕНОМЕНОЛОГИЧЕСКОЕ ОПИСАНИЕ ОТКЛИКОВ НЕЙРОНОВ В ПЕРВИЧНОЙ ЗРИТЕЛЬНОЙ КОРЕ

Пусть  $D(x, y)$  обозначает отклики нейрона, находящегося в позиции  $x, y$ . Структура рецептивного поля  $D(x, y)$  простых нейронов V1, настроенных на ориентацию, математически обычно описывается функцией Габора, то есть произведением функций экспоненты и косинуса:

$$D(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{\left[ -\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} \right]} \cos(kx - \phi), \quad (5.1)$$

где  $\sigma_x$  и  $\sigma_y$  контролируют пространственное распределение рецептивного поля,  $k$  контролирует пространственную частоту, а  $\phi$  – фазу. Функция Габора характеризуется удлиненной областью возбуждения, угол в которой соответствует предпочтению ориентации нейрона V1, а также окружающей ингибирующей области. Пример вида функции Габора показан на рис. 5.5.



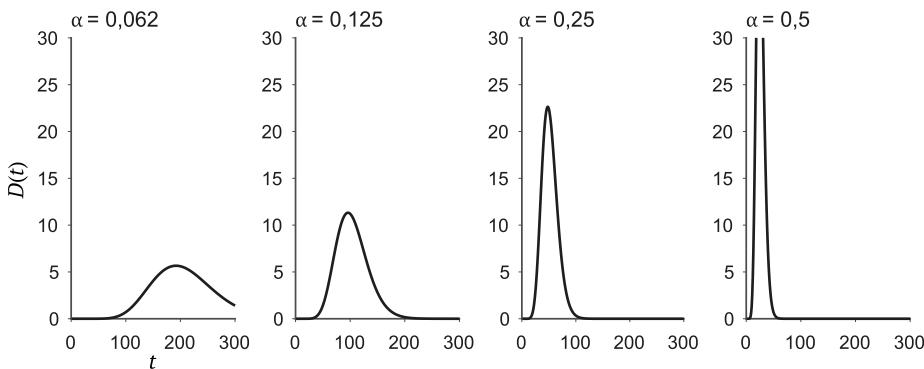
**Рис. 5.5** ♦ Пространственная структура рецептивных полей нейронов V1 обычно описывается функцией Габора – уравнение (5.1). (A) Иллюстрация функции Габора. (B) Контуровая диаграмма

Помимо пространственных характеристик рецептивного поля, важно охарактеризовать временную динамику ответов в V1. В большинстве случаев пространственные и временные показатели рецептивных полей в V1 можно рассматривать как приблизительно независимые; то есть их можно разделить без учета сложных взаимодействий между пространством и временем.

Временную функцию рецептивного поля можно описать следующим феноменологическим уравнением:

$$D(t) = \alpha e^{-\alpha t} [(\alpha t)^5 / 5! - (\alpha t)^7 / 7!] \quad (5.2)$$

для  $t \geq 0$  и 0 в противном случае. Это уравнение представляет собой причудливый способ согласовать быстрое и кратковременное увеличение спайковой частоты при запуске стимула в момент времени 0 (рис. 5.6). Параметр  $\alpha$  контролирует задержку и ширину временного рецептивного поля.



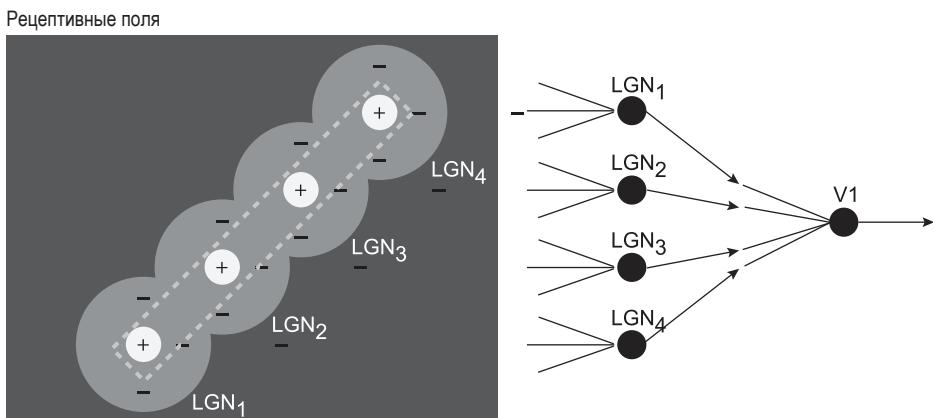
**Рис. 5.6** ♦ Временная структура рецептивных полей нейронов V1.  
Уравнение (5.2) показано для разных значений параметра  $\alpha$

## 5.8. ПРОСТАЯ МОДЕЛЬ ОРИЕНТАЦИОННОЙ ИЗБИРАТЕЛЬНОСТИ В ПЕРВИЧНОЙ ЗРИТЕЛЬНОЙ КОРЕ

Уравнение (5.1) дает феноменологическое описание структуры рецептивного поля. Благодаря замечательному интуитивному прорыву Хьюбел и Визель предложили простую и элегантную, биофизически правдоподобную модель того, как настройка ориентации может возникать из ответов нейронов с рецептивными полями LGN-типа (рис. 5.7). В их модели несколько нейронов LGN с центрально-периферийными рецептивными полями с круговой симметрией (рис. 2.10), расположенных вдоль линии, проецируются на простой нейрон V1. Таким образом, настройка ориентации осуществляется снизу вверх путем объединения входов правого набора нейронов LGN.

Последующая работа привела к появлению множества других моделей, и до сих пор продолжаются споры о том, в какой степени чисто восходящая модель Хьюбела–Визеля представляет собой единственный механизм, приводящий к избирательности ориентации в области V1. Тем не менее эта простая и элегантная интерпретация происхождения рецептивных полей V1 представляет собой замечательный пример того, как экспериментаторы могут придумать разумные и глубокие модели, которые объясняют полученные данные. Кроме того, основные идеи, примененные в этой модели,

были обобщены, чтобы объяснить формирование предпочтений нейронов для стимулов более сложной формы в других областях коры (раздел 8.5).



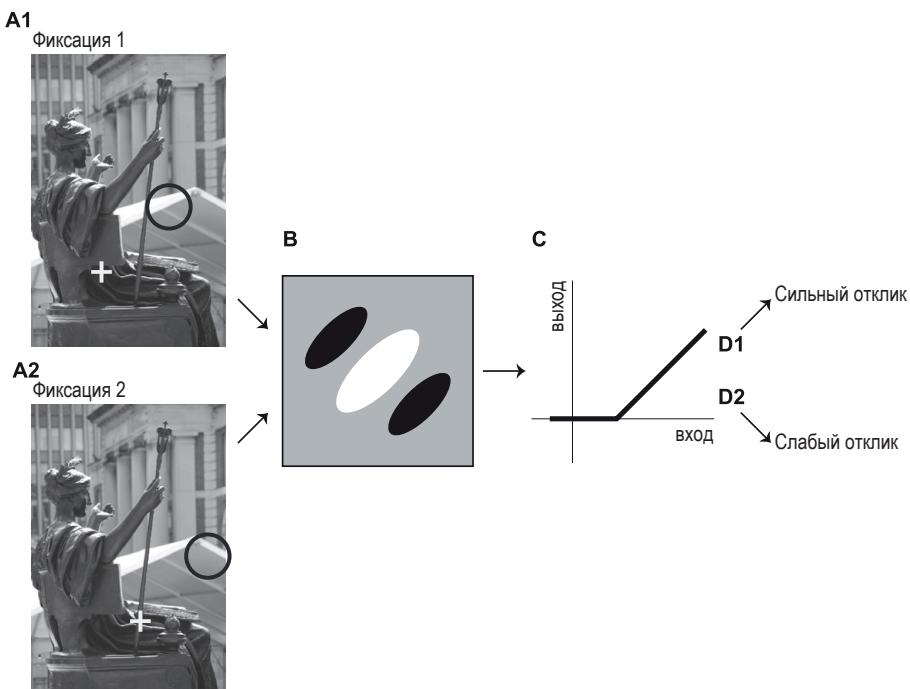
**Рис. 5.7** ♦ Установление ориентации путем объединения центро-периферийных нейронов. Схематическая диаграмма, показывающая, как несколько нейронов LGN с круговой структурой рецептивного поля могут быть объединены для возбуждения простого нейрона V1, который показывает настройку ориентации, при которой центры этих рецептивных полей адекватно выровнены. Измененное из Хьюбел и Визель, 1962

В дополнение к предпочтениям ориентации существует множество других свойств нейронов V1, которые также расположены топографически, включая их пространственные рецептивные поля, их окулярное доминирование (более сильные ответы на входные сигналы, исходящие от одного или другого глаза), их избирательность по направлению (более сильные ответы для конкретных направлений движения), несогласованность сетчатки (чувствительность к разнице положения объекта между правым и левым глазами, что используется для стереопсиса). Оказывается, все эти свойства также могут быть представлены как специфические схемы входов от LGN.

Расширяя свою модель ориентационных предпочтений простых нейронов как комбинацию выходных данных LGN (рис. 5.7), Хьюбел и Визель предположили, что ответы сложных нейронов могут возникать в результате нелинейной комбинации ответов нескольких простых нейронов с аналогичными предпочтениями ориентации, но с небольшим сдвигом рецептивных полей. Эти новаторские идеи об операции линейной фильтрации, вызывающей реакции простых нейронов в V1, с последующей нелинейной операцией объединения, приводящей к возбуждению сложных нейронов в V1, сыграли важную роль в новых разработках вычислительных моделей обработки изображения (раздел 8.5).

На рис. 5.8 схематически показано, как простой нейрон V1 будет реагировать на реальное изображение. Этот нейрон имеет рецептивное поле в верхней правой части поля зрения (черный кружок). На этом рисунке показаны две фиксации. При первой фиксации (A1) изображение внутри рецептив-

ногого поля похоже на предпочтительную ориентацию нейрона (B). После нелинейной функции активации (C) нейрон показывает сильный ответ (D1). Когда подопытный делает небольшое движение глаза вправо, приземляясь на фиксацию 2 (A2), изображение внутри рецептивного поля больше не напоминает предпочтительные характеристики нейрона, и реакция становится слабой (D2).



**Рис. 5.8** ♦ Покидаем простые схемы и выходим в реальный мир. Схематический пример того, как простой нейрон V1 ведет себя в реальном мире. Знак «+» указывает на место фиксации, а черный кружок указывает на место рецептивного поля. В **A1** изображение внутри рецептивного поля похоже на предпочтительную ориентацию нейрона (**B**), вызывающую сильный отклик (**D1**), тогда как в нижнем случае верно обратное

## 5.9. ДРУГИЕ СЮРПРИЗЫ V1

Несмотря на значительный объем работы по исследованию свойств нейронов первичной зрительной коры, многое еще предстоит объяснить. Некоторые искажения приводят к неполноте нашего представления о функции V1. Во-первых, многие из регистрирующих процедур на сегодняшний день, как правило, сосредоточены на нейронах, которые имеют более высокую частоту срабатывания, и их легче уловить с помощью внеклеточных регистраций. Интернейроны меньше, и их труднее регистрировать, чем более крупные

пирамидальные клетки. Кроме того, могут существовать «застенчивые» нейроны, которые ускользают от внимания исследователей. Во-вторых, типы стимулов, которые мы используем для исследования нейронных откликов, также имеют предвзятость (раздел 5.11, рис. 5.10). Возможно, в V1 есть нейроны, которые сильно реагируют на фиолетовые треугольники с подсолнухом наверху, но, что неудивительно, никто этого не проверял. Зачем кому-то проверять такой стимул? Ни одна из наших теорий не предполагает, что такой стимул будет особенно актуален для нейронов V1. Однако наши теории могут быть недостаточно объективными. Еще один важный момент, о котором следует помнить, – это то, что нейронные реакции в V1 часто исследуются у обезьян, которые не выполняют никаких визуальных задач, кроме фиксации. На ответы нейронов V1 могут воздействовать пространственный контекст, временной контекст, внутренние ожидания и требования задачи.

В последнее десятилетие наблюдается захватывающий рост исследований V1 мышей, многие из которых открыли нам глаза на мир, полный сюрпризов, даже в самом сердце наиболее изученной области коры. Многие эксперименты на мышах проводятся, когда животное бегает по мячу, что-то вроде упражнения на беговой дорожке, в то время как мышь просматривает демонстрируемый ей световой стимул или фильм. Одно из самых шокирующих открытий V1 за последнее десятилетие заключается в том, что скорость бега может сильно модулировать ответы нейронов V1. Один и тот же визуальный стимул может вызывать очень разные реакции в зависимости от того, неподвижно ли животное, бежит ли оно медленно или изо всех сил. Если это недостаточно поразительно, то вот вам еще: ответы этих нейронов V1 также можно модулировать при беге в темноте, при отсутствии какой-либо визуальной стимуляции мыши. Продолжая список интригующих наблюдений на мышах, найдены прямые связи первичной слуховой коры с V1, и, оказывается, возможно вызвать ответы в нейронах V1 с помощью акустических сигналов! Эти ответы слабее, чем реакции, запускаемые визуально, но это слуховой сигнал, управляющий наиболее визуальной частью коры. Отклонения усов также могут модулировать нейроны V1. И движения головы тоже.

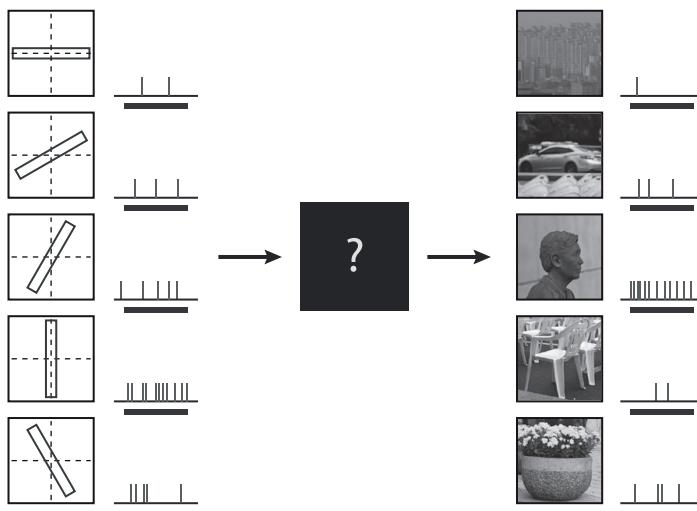
Остается неясным, распространяются ли какие-либо из этих наблюдений на обезьян, не говоря уже о людях. Нелегко проводить нейрофизиологические записи у бегающих обезьян, и очень сложно выполнять нейрофизиологические записи в V1 человека.

Насколько нам известно, нет сообщений о слуховых стимулах, модулирующих ответы V1 у обезьян (после контроля положения глаз, внимания и визуальных стимулов). Мозг грызунов намного меньше, чем мозг обезьяны макаки (*Mus musculus* и *Macaca mulatta* эволюционно разошлись примерно 75 млн лет назад), который, в свою очередь, меньше человеческого мозга (*Macaca mulatta* и *Homo sapiens* разошлись примерно 25 млн лет назад). Интроспективно кажется, что наш визуальный мир не меняется, когда мы ходим или бегаем. Однако могут существовать механизмы, которые компенсируют модулирующие реакции в V1 во время бега (помните, что мы не осознаем даже глобальные изменения картины мира, вызванные морганиями и движениями глаз, гл. 2). Слуховая кора, соматосенсорная кора и моторная кора у мышей ближе к V1, чем у обезьян и тем более у людей; у макак боль-

ше извилин, которые могут разделять эти области мозга, и еще больше их в мозге человека. Следует отметить, что все это предположения, и нам нужно будет оценить все эти возможности в нейрофизиологических экспериментах на обезьянах и людях. Мы должны держать свой разум открытым и ожидать много интересных сюрпризов в будущем.

## 5.10. РАЗДЕЛЯЙ И ВЛАСТВУЙ

Покидая первичную зрительную кору и поднимаясь по иерархии корковых вычислений, мы достигаем захватывающих и сбивающих с толку областей коры головного мозга, которые соединяют низкоуровневые зрительные функции со строительными блоками восприятия образов. В первичной зрительной коре есть нейроны, которые избирательно реагируют на линии разной ориентации (рис. 5.3). На другом конце визуальной иерархии находятся нейроны в нижней височной коре (ITC), которые выборочно реагируют на сложные образы и помогают нам идентифицировать стулья, лица и планеты (будет обсуждаться в разделе 6.2). Между V1 и представлением сложных образов объектов существует огромное пространство коры головного мозга, вовлеченнное в, казалось бы, волшебные преобразования, которые превращают ориентированные линии в сложные образы. Как перейти от ориентированных линий к распознаванию стульев, лиц и планет (рис. 5.9)?



**Рис. 5.9 ♦** Как кора головного мозга преобразует пиксели в визуальные образы? Посредством каскада вычислений в центральном визуальном потоке мозг может преобразовывать предпочтения простых свойств стимула, таких как настройка ориентации, в сложные признаки, такие как лица

Несмотря на героические усилия талантливых исследователей изучить реакции между первичной зрительной корой и высшими эшелонами ITC,

вентральная зрительная кора остается в основном *terra incognita*. Визуальная информация течет по вентральному зрительному потоку из V1 в области V2, V4, заднюю и переднюю части ИТС. Корковое пространство между V2 и ИТС представляет собой таинственный, соблазнительный и противоречивый ансамбль нейронов, функции которых остаются неясными и только начинают расшифровываться. Смелые исследователи, вооруженные вычислительными моделями, электродами и интуицией, начинают описывать предпочтения к повороту нейронов в областях V2, V3 и V4 относительно различных характеристик изображения, включая кривизну, несоответствие, цвет, текстуру и силуэты.

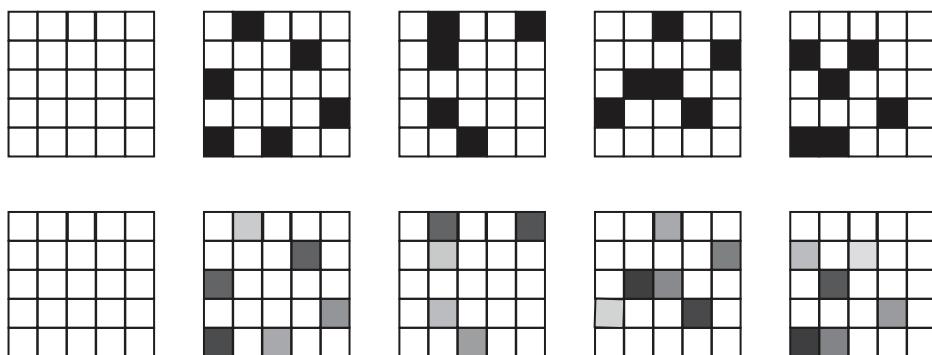
Чтобы решить сложную задачу интерпретации сцены, визуальная система, похоже, приняла стратегию «разделяй и властвуй». Вместо того чтобы пытаться придумать единственную функцию, которая будет преобразовывать линии в сложные зрительные образы за один шаг, вычисления, лежащие в основе визуального восприятия мира, реализуются каскадом в основном последовательных вычислений. Каждое из этих вычислений может быть обманчиво простым, и все же объединение таких шагов может привести к интересным и сложным результатам. В качестве грубой аналогии рассмотрим завод по производству автомобилей. Существует длинная последовательность специализированных областей, отделов и задач. Одна группа рабочих может принимать участие в приеме и заказе запчастей; другие могут специализироваться на сборке карбюратора; третий в покраске кузова. Автомобиль – результат всех этих последовательных и параллельных шагов. Чтобы понять весь механистический процесс, с помощью которого создается автомобиль, нам нужно глубже изучить каждый из этих специализированных этапов, не теряя связи с общей целью, достижению которой способствует каждый из них, то есть с конечным продуктом.

## 5.11. НЕВОЗМОЖНО ИСЧЕРПЫВАЮЩЕ ИЗУЧИТЬ РЕАКЦИЮ НА ВСЕ ВОЗМОЖНЫЕ ВИЗУАЛЬНЫЕ СТИМУЛЫ

Было бы неплохо иметь возможность описать предпочтения настройки нейронов вдоль вентрального зрительного потока аналогично настройке ориентации и функциям Гabora для нейронов V1. Было предпринято множество эмпирических попыток охарактеризовать нейронные предпочтения нейронов V2, V4 и ИТС, что привело к захватывающим открытиям. Как и в известной притче о слепых, пытающихся описать слона, касаясь отдельных частей, разные исследователи придумали несколько примеров того, как нейроны реагируют на углы, цвета, искривления и другие формы.

Одна из основных проблем при исследовании функции и предпочтений нейронов коры головного мозга состоит в том, что существует слишком много возможных изображений, а время регистрации ответа для данного нейрона ограничено. Учитывая современные методы, просто физически невоз-

можно исчерпывающе изучить большое количество возможных комбинаций различных стимулов, которые могут управлять нейроном. Рассмотрим простой сценарий, в котором мы представляем участки изображения размером  $5 \times 5$  пикселей, где каждый пиксель либо черный, либо белый (рис. 5.10, вверху). Таких стимулов  $2^{25}$  (более 33 млн). Если мы представляем каждый стимул в течение 100 мс и не допускаем никакого промежутка времени между стимулами, то для представления наблюдателю всех возможных комбинаций потребуется более пяти недель. Если мы позволим каждому пикселю иметь оттенки серого от 0 до 255 (рис. 5.10, внизу), то будет гораздо больше возможностей:  $256^{25}$  таких изображений (около  $10^{60}$  таких изображений!). Более того, проблема становится еще более неразрешимой, если мы допускаем использование трех цветов (красный, зеленый и синий) в изображениях размером более  $5 \times 5$  пикселей. Даже после ограничения нашего анализа плохо определенным подмножеством естественных изображений (раздел 2.1) мы все равно будем иметь астрономическое количество возможных изображений. Обычно мы можем проводить внеклеточные записи с одиночными (временно установленными) электродами в течение нескольких часов. Недавние экстраординарные усилия позволили отслеживать активность данного нейрона в течение года. Однако даже с такими постоянными электродами сложно удерживать животное в задаче визуального представления более нескольких часов в день. Таким образом, мы не можем записать ответы нейрона на все изображения.



**Рис. 5.10 ♦** Проклятие размерности в зрении. Используя современные методы, мы не можем исчерпывающе охватить все возможные стимулы. Здесь мы рассматриваем сетку  $5 \times 5$  возможных двоичных изображений (вверху) или возможных изображений в градациях серого (внизу). Даже для таких простых стимулов количество вариантов чудовищно велико

Из-за серьезных ограничений количества стимулов, которое в принципе возможно протестировать, исследователи часто прибегают к нескольким хитрым стратегиям, чтобы решить, какие стимулы использовать для исследования реакции корковых нейронов. Эти стратегии обычно включают комбинацию (i) вдохновений из предыдущих исследований (прошлое поведение нейронов в других исследованиях является хорошим предсказанием

того, как нейроны будут вести себя в новом эксперименте); (ii) интуитивных предположений относительно того, какие типы изображений могут иметь значение для нейронов, а какие – нет (например, многие исследователи утверждали, что объекты реального мира, такие как лица, должны иметь значение); (iii) статистики естественных стимулов (как обсуждалось в разделе 2.1, разумно предположить, что настройка нейронов формируется воздействием изображений в естественном мире); (iv) вычислительных моделей (более подробно обсуждаются в главах 7–9); (v) интуиции (нельзя недооценивать роль тщательного изучения и систематического наблюдения в сочетании с удачей). Комбинируя эти подходы, несколько исследователей пытались протестировать нейронный код на предмет различия визуальных образов вдоль вентральной зрительной коры.

## 5.12. Мы живем визуальным прошлым: латентность реакции вдоль вентрального потока увеличивается

Визуальная обработка выполняется очень быстро (раздел 3.6). В самом деле, как мы утверждали в гл. 1, скорость зрения, возможно, дала важные преимущества первым видам, приобретшим глаза, и, вполне может быть, была одной из ключевых причин, почему эволюция привела к развитию зрительных способностей. Однако даже если кажется, что мир материализуется перед нами, как только мы открываем глаза, как мы отмечали в разделе 2.6, обработка сигнала в сетчатке требует времени. Представление о том, что зрение работает мгновенно, – не более чем иллюзия. Для выхода сигналов из ганглиозных клеток сетчатки в таламус требуется от 30 до 50 мс, а для распространения сигналов через кору требуется еще больше времени.

Небольшая часть этого времени связана со скоростью распространения по дендритам и аксонам внутри нейрона. Однако задержки внутри нейрона относительно невелики. В частности, сигналы потенциала действия внутри аксонов, которые изолированы миелином, могут распространяться со скоростью около 100 м/с. Таким образом, сигналы от одиночного миелинизированного аксона могут, в принципе, пройти всю длину человеческого мозга, составляющую около 15 см, примерно за 1,5 мс. Дендриты обычно короче аксонов, и скорость распространения внутри дендритов также довольно высока. Основная причина, по которой зрение далеко не мгновенно, – это многочисленные вычисления и этапы интеграции в каждом нейроне в сочетании с синаптической передачей информации от одного нейрона к следующему, выполняемой по мультисинаптической связи коры.

На каждом этапе обработки в зрительной системе можно оценить время, которое требуется нейронам этой области коры головного мозга, чтобы понять, что испытуемому была представлена вспышка света. Латентность ответа на вспышку стимула в рецептивном поле нейрона увеличивается с ~45 мс

в LGN до ~100 мс в нижней височной коре (табл. 5.1). Также увеличивается средняя задержка в каждой области коры от сетчатки до LGN, до V1, до V2, до V4 и ITC. Эта прогрессия задержки сигнала еще больше укрепила представление о центральном потоке обработки как приблизительно иерархической и последовательной архитектуре. Каждый дополнительный этап обработки в центральном потоке добавляет в среднем ~15 мс вычислительного времени.

Следует подчеркнуть, что это только грубые значения, и существует значительная вариабельность от нейрона к нейрону в каждой области. Анализ нейронных записей обезьян под наркозом, проведенный Шмолоски и его коллегами, показал, что задержка в первичной зрительной коре головного мозга колеблется от 30 до 70 мс. Из-за этой неоднородности распределения латентных периодов ответа перекрываются, и самые быстрые нейроны в данной области (скажем, V2) могут срабатывать раньше, чем самые медленные нейроны в более ранней области (скажем, V1). Существует не только неоднородность задержки передачи сигнала от одного нейрона к другому в пределах данной визуальной области, но даже один и тот же нейрон может показывать разные задержки в зависимости от природы стимула. Например, задержки ответа имеют тенденцию быть обратно пропорциональны контрасту стимула. Понятие последовательной обработки является лишь грубым приближением. Однако задержки отклика представляют собой важное ограничение для количества возможных вычислительных шагов в зрительной системе.

**Таблица 5.1. Задержки отклика в различных областях мозга у обезьяны-макаки.  
Из Шмолоски и др., 1998**

Зона	Среднее (мс)	S. D. (мс)
Слой LGNd M	33	3.8
Слой LGNd P	50	8.7
V1	66	10.7
V2	82	21.1
V4	104	23.4
V3	72	8.6
MT	72	10.3
MST	74	16.1
FEF	75	13

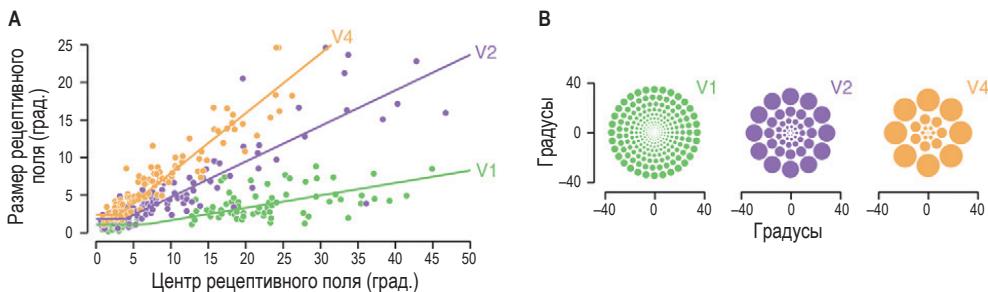
Из-за этих задержек мы постоянно живем в визуальном прошлом. Представление о том, что мы видим только прошлые события, особенно очевидно, когда мы рассматриваем далекие звезды. Световые сигналы, которые достигают Земли, давно покинули эти звезды. Хотя в случае с мозгом эта идея менее очевидна, она применима и к обработке изображений в нашей зрительной системе. Конечно, время, необходимое для того, чтобы свет отразился на данном объекте и достиг сетчатки, ничтожно мало, но, как обсуждалось ранее, распространение сигнала в мозге занимает порядка сотни микросекунд. Благодаря обучению мозг может учесть эти задержки, предсказывая, что произойдет дальше. Например, как игрок в пинг-понг может прореагировать на удар? Мяч может двигаться со скоростью около 50 км/ч (по-видимому,

мировой рекорд составляет около 112 км/ч), и, таким образом, мяч проходит через 3-метровую длину стола примерно за 200 мс. К тому моменту, когда противник должен отбить мяч обратно, в лучшем случае его зрительная кора обрабатывает сенсорные сигналы примерно с того времени, когда мяч прошел над сеткой, не говоря уже о том, что организация ответного движения также требует времени (сигналы должны успеть распространиться от сетчатки к центрам принятия решений в мозге, а затем оттуда к мышцам; все эти шаги требуют времени). Единственная возможность играть в пинг-понг и во многие другие виды спорта – использовать получаемую зрительную информацию в сочетании с предсказаниями, полученными исходя из собственного опыта. С помощью этих прогнозов игроки не только получают пользу из быстроты ответного удара, но также имеют возможность прибегнуть к другим стратегиям, например закручиванию мяча, чтобы сбить с толку противника.

## 5.13. УВЕЛИЧЕНИЕ РАЗМЕРА ВОСПРИНИМАЮЩЕГО ПОЛЯ ВДОЛЬ ВЕНТРАЛЬНОГО ЗРИТЕЛЬНОГО ПОТОКА

По мере того как мы поднимаемся по зрительной иерархии, одновременно с увеличением задержек становятся больше рецептивные поля (рис. 5.11). Они варьируются от долей градуса на начальных этапах (LGN, V1) до нескольких градусов или даже десятков градусов в высших эшелонах коры. Каждая область зрительной коры имеет полную карту зрительного поля; таким образом, центры рецептивных полей переходят от центральной ямки к периферии. Как обсуждалось для первичной зрительной коры, в каждой области размер рецептивного поля увеличивается по мере того, как мы удаляемся от центральной ямки. Фовеальная область всегда имеет лучшее разрешение среди всех областей сетчатки. Диапазон размеров рецептивного поля в пределах зрительной области также увеличивается со средним размером рецептивного поля. Распределение относительно узкое в первичной зрительной коре, но исследователи описали широкий диапазон размеров рецептивного поля в V4 и нижней височной коре. Коэффициент масштабирования между размером рецептивного поля и эксцентриситетом более выражен в V4, чем в V2, и более выражен в V2 по сравнению с V1.

Увеличение размера рецептивного поля от одной области к другой может быть естественным следствием операций, подобных объединению, в иерархической сети, что мы обсудим более подробно, когда представим вычислительные модели обработки сигнала в зрительной коре в разделе 8.2. Увеличение размера рецептивного поля обеспечивает несколько интересных свойств: (i) особый механизм отказа от точной позиционной информации в пользу (ii) извлечения визуальных признаков, которые показывают все большую степень инвариантности к точному положению или масштабу соответствующих визуальных признаков, и (iii) способность комбинировать формы из слегка смещенных мест для построения все более сложных декрипторов визуальных характеристик.



**Рис. 5.11** ♦ Размеры рецептивного поля увеличиваются с увеличением эксцентрикситета и вдоль вентрального потока. Кроме того, размер рецептивного поля увеличивается вдоль вентрального зрительного потока с фиксированным эксцентрикситетом. (А) Экспериментальные данные, основанные на нейрофизиологических записях макак. (Б) Схематическое изображение размеров рецептивного поля в областях V1, V2 и V4. Воспроизведено по работе Фримена и Симончелли, 2011

## 5.14. ЧТО ПРЕДПОЧИТАЮТ НЕЙРОНЫ ЗА ПРЕДЕЛАМИ V1?

Было проведено несколько систематических параметрических исследований предпочтений нейронов в областях V2 и V4. Эти исследования открыли двери для изучения сложных трансформаций вентрального зрительного потока. Несмотря на то что в нескольких интересных исследованиях сравнивали отклики в V1, V2 и V4, у нас еще нет четкой единой теории о том, что нейроны «предпочитают» в этих более высоких зрительных областях. Конечно, термин «предпочитать» – это антропоморфизм. Нейроны ничего не предпочитают. Они запускают спайки всякий раз, когда интегрированный входной сигнал превышает заданный порог. Исследователи часто говорят о предпочтениях нейронов в терминах того, какие типы изображений вызывают более высокую частоту их срабатывания.

Представление о том, что нейроны V1 предпочитают определенную настройку ориентации объекта, хорошо известно, даже если это объясняет только часть вариаций в ответах V1 на естественные стимулы. Значительно меньше согласия относительно типов элементов формы, которые закодированы в V2 и V4. Было проведено несколько исследований, в которых изучали ответы с помощью стимулов, более сложных, чем ориентированные полоски, и менее сложных, чем повседневные предметы. Эти стимулы включают в себя, среди прочего, синусоидальные решетки, гиперболические решетки, полярные решетки, углы, образованные пересекающимися линиями, а также кривые различных видов. Простые стимулы, такие как декартовы решетки, безусловно, могут вызывать ответы в V2 и V4. Как правило, нейроны в V2 и V4 могут более сильно управляться более сложными формами. Как обсуждалось ранее в контексте задержки, предпочтения тех или иных визуальных стимулов в V2 и V4 широко распространены.

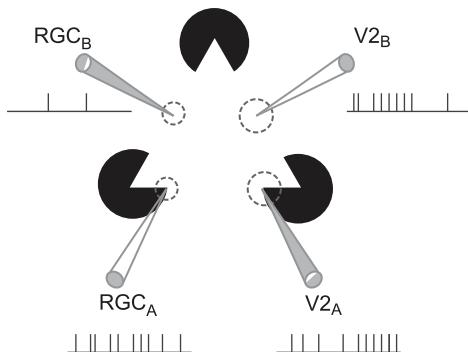
Возможно, одна из проблем заключается в том, что исследователи ищут объяснение предпочтений нейронного кодирования в терминах разговорных английских выражений, таких как ориентация, цвет или кривизна. Привлекательной идеей, которая набирает обороты, является представление о том, что нейроны в этих высших визуальных областях фильтруют входные данные с предыдущих этапов, чтобы производить сложные функции настройки, которые не поддаются описанию на разговорном языке. Нейрон может быть активирован зрительным образом, представляющим собой сложные формы и текстуры, которые нельзя просто определить как «угол» или «выпуклую кривую». В конечном счете языком природы является математика, а не английский или эсперанто. Свойства настройки нейронов не обязательно должны напрямую отображаться в краткой фразе на разговорном или литературном языке; мы вернемся к этой идеи в гл. 7 и 8, когда будем обсуждать вычислительные модели зрения.

## 5.15. МОЗГ КОНСТРУИРУЕТ ИНТЕРПРЕТАЦИЮ МИРА: СЛУЧАЙ ИЛЛЮЗОРНЫХ КОНТУРОВ

Всеобщее заблуждение – думать, что наши органы чувств дают достоверное представление о том, что действительно существует в мире. Это суждение можно легко опровергнуть путем изучения зрительных иллюзий. В разделе 3.1 мы утверждали, что наш мозг создает вещи, когда делает интерпретацию внешнего мира. Наш мозг «выдумывает вещи» – это означает, что должны быть нейроны, которые явно создают эти мысленные конструкции. Давайте вернемся к треугольнику Каниджа (рис. 5.12), где у нас возникает сильная иллюзия присутствия равностороннего треугольника посреди трех силуэтов персонажа компьютерной игры, Пакмана. Небольшие части сторон треугольника возле вершин составлены из настоящих черных контуров. Однако центр каждой стороны состоит из линии, которой на самом деле не существует. Эти линии представляют собой иллюзорные контуры, то есть края, созданные без какого-либо изменения яркости на изображении.

Относительно легко «обмануть глаз», за исключением того, что сам глаз обычно *не обманывается* в большинстве зрительных иллюзий. Зрительные иллюзии представляют собой ситуацию, когда наш мозг создает интерпретацию изображения, отличную от содержания пиксельного уровня картинки, представленной сетчаткой глаза. В большинстве таких иллюзий реакции ганглиозных клеток сетчатки (RGC) относительно хорошо соответствуют содержанию пиксельного уровня изображения. Рассмотрим регистрацию активности RGC, центр рецептивного поля которого соответствует положению А на рис. 5.12, справа вдоль одной стороны Пакмана. Внутри рецептивного поля присутствует изменение яркости, и мы ожидаем, что нейрон будет энергично возбуждаться в этом месте после сканирования фигуры Каниджа. Теперь рассмотрим RGC с центром рецептивного поля, расположенным в позиции В, прямо посередине стороны иллюзорного треугольника. Мы не ожидаем, что этот нейрон сработает выше базового уровня, потому что внутри

рецептивного поля стимул отсутствует. Другими словами, активность RGC не коррелирует с нашим восприятием образа. Если сетчатка не создает то, что мы видим, тогда что это создает? Кажется разумным предположить, что где-то должны быть нейроны, которые явно конструируют содержание нашего восприятия – в данном случае иллюзорный контур треугольника. Это явное представление является важнейшим постулатом, который мы обсудим еще раз более подробно, когда будем разбирать нейронные корреляты сознания в разделе 10.3.



**Рис. 5.12** ♦ Нейроны V2 могут нарисовать линии, которые существуют только в глазах наблюдателя. На рисунке показаны зрительная иллюзия треугольника Каниджи и схематическая визуализация нейрофизиологических записей четырех нейронов: двух ганглиозных клеток сетчатки (RGC) и двух нейронов V2. Когда рецептивные поля (серые пунктирные круги) охватывают участки, которые имеют реальный контур (A), нейроны RGC и V2 активно возбуждаются. Напротив, когда рецептивные поля охватывают иллюзорный контур (B), нейрон V2 активно стреляет спайками, но нейрон RGC запускает только несколько базовых спайков

Действительно, нейроны в области V2 реагируют на иллюзорные контуры (рис. 5.12). Нейрон V2, который предпочитает горизонтальные края, будет сильно возбужден, если его рецептивное поле находится в точке A, потому что там есть настоящая горизонтальная линия. Примечательно, что нейрон V2, который предпочитает левые края, также сработает, если его рецептивное поле находится в позиции B, где есть иллюзорное ребро треугольника. Нейроны V2 почти одинаково хорошо реагируют на иллюзорную линию или реальную линию. Реакция на иллюзорные контуры замечательна, потому что в рецептивном поле нейрона не происходит изменения яркости изображения. Следовательно, эти ответы указывают на форму контекстной модуляции, которая согласуется с субъективной интерпретацией границ. В V1 также есть нейроны, которые реагируют на иллюзорные контуры, но в V2 таких нейронов больше. Интересно, что отклики на иллюзорные контуры показывают небольшую задержку по сравнению с откликами на реальные линии. Эти задержки могут отражать необходимость дополнительных вычислительных шагов для вывода о наличии линии, когда ее на изображении нет.

## 5.16. КРАСОЧНЫЙ V4

Нейроны сетчатки (колбочки), LGN (парвоцеллюлярные нейроны) и первичной зрительной коры, особенно внутри так называемых пятен (blobs) в V1, чувствительны к цвету стимула в пределах их рецептивного поля. Нейроны в области V4 демонстрируют более сложную чувствительность к цветовым свойствам сигнала, чем в более ранних областях. Примечательным является то, что нейроны в V4 проявляют более постоянное восприятие цвета, в результате чего воспринимаемый цвет объекта относительно независим от больших изменений в общей освещенности, в отличие от реакций предыдущих отделов зрительной системы.

Существует множество зрительных иллюзий, основанных на явлении постоянства цвета. Банан обычно кажется желтым для наших глаз, независимо от того, видим ли мы его в полдень, ранним вечером или в свете лампы на кухне. Фактический спектр света, достигающего глаз, довольно сильно зависит от освещения окружающей среды, а колбочки на сетчатке сигнализируют о реальных длинах волн, отраженных от банана. Однако наше восприятие не учитывает освещение и интерпретирует банан как желтый. Считается, что интеграция цветовых сигналов, получаемых в пределах рецептивного поля, с сигналами окружения, для выполнения этого типа цветовой коррекции, имеет место в V1 и даже более определенно в нейронах V4. Ответы нейронов V4 лучше коррелируют с тем, как приматы воспринимают цвета. Более того, редкое явления коркового дальтонизма, известное как ахроматопсия, было связано с повреждением области V4 (раздел 4.8).

## 5.17. Модуляция внимания

Как отмечалось ранее в этой главе, нейроны вдоль вентральной зрительной коры получают множество нисходящих сигналов в дополнение к своим восходящим сигналам. Посредством этих нисходящих сигнальных механизмов активность нейронов вдоль вентральной зрительной коры может сильно модулироваться сигналами, выходящими за рамки определенного визуального содержания в пределах их рецептивных полей, включая пространственный контекст, временной контекст, ожидания и когнитивные влияния более высокого уровня, такие как цели выполняемой задачи.

Несмотря на большой интерес к таким нисходящим сигналам, было проведено гораздо больше исследований о роли восходящих сигналов в ответах нейронов. По крайней мере частично, этот дисбаланс связан с тем, что гораздо легче изменить то, что отображается на экране, чем изменить внутренние ожидания и цели животного.

Ярким примером изучения модулирующих сигналов сверху вниз при визуальной обработке является *пространственное внимание*. Один из способов привлечь внимание к одной части поля зрения – это двигать глазами. Однако эффекты пространственного внимания могут быть продемонстрированы и вне точки фиксации. Подопытный может смотреть в одно место и обращать

внимание на другое – явление, известное как скрытое внимание (в отличие от открытого внимания, которое является более распространенным сценарием, когда внимание уделяется области фиксации). Посредством серии изощренных методик исследователи смогли обучить животных задействовать скрытое пространственное внимание, что позволило им исследовать влияние пространственного внимания на нейроны с рецептивными полями за пределами центральной ямки зрительного нерва.

Животное приучают фиксироваться на центр экрана, и за движениями его глаз строго следят, чтобы гарантировать, что эффект внимания не включает в себя использование саккад. В некоторых испытаниях животное награждают за обнаружение визуального стимула в определенном месте справа, и это говорит животному о необходимости сосредоточить внимание на данной области зрительного поля, не нарушая фиксации. Чистоту эксперимента можно проверить путем случайного зондирования стимула, представленного в другом месте, и демонстрации того, что эффективность лучше (быстрее, точнее) в области интереса исследователя.

В этих экспериментальных условиях нейроны обычно демонстрируют усиление откликов, когда их рецептивное поле находится в фокусе внимания и частично в области предъявления визуального стимула. Другими словами, представьте нейрон в V2 с расположением рецептивного поля, которое находится прямо в центре экспериментальной области в некоторых испытаниях и вне этой области в других испытаниях. Нейрон будет реагировать на идентичный визуальный стимул с большим количеством спайков в тех пробах, когда внимание охватывает рецептивное поле. Эффект пространственного внимания не является однозначным, «или – или». Нейроны по-прежнему энергично реагируют на стимул, помещенный в их рецептивное поле, независимо от того, обращает ли животное внимание на это место. Концентрация внимания приводит к увеличению спайковой частоты примерно на 5–30 %. Величина этого эффекта внимания соответствует обратному иерархическому порядку и значительно сильнее в области V4 по сравнению с областью V1.

Нейронные ответы также можно модулировать специфическим образом, основанным на наличии каких-либо конкретных признаков. Вместо того чтобы обращать внимание на конкретное место, животное можно обучить обращать внимание на конкретный признак стимула, такой как красный цвет или вертикальные линии. Когда животное обращает внимание на предпочтительные особенности объекта, нейрон демонстрирует повышенную частоту спайков.

## 5.18. Обзор главы

- Визуальные вычисления происходят в шестислойной структуре неокортика.
- Кора головного мозга характеризуется стереотипными схемами связи от одной области к другой, образуя приблизительно канонические микроцепи.
- Золотым стандартом для изучения функции коры головного мозга является точное измерение активности отдельных нейронов.

- Нейроны первичной зрительной коры обнаруживают края объектов и демонстрируют ориентационно-специфическую реакцию, сильнее реагируя на полосу определенной ориентации в пределах рецептивного поля.
- Сложные нейроны первичной зрительной коры демонстрируют толерантность к точному положению предпочтительного стимула в рецептивном поле.
- Функция Габора феноменологически соответствует ответам нейронов V1.
- Механистическая модель утверждает, что рецептивные поля простых клеток V1 могут быть созданы путем адекватного сочетания выходов нескольких центро-периферийных нейронов латерального коленчатого ядра, взаимно расположенных определенным образом для фиксирования требуемой ориентации.
- Модель утверждает, что рецептивные поля сложных клеток V1 могут быть созданы путем адекватного комбинирования выходных сигналов простых клеток V1 с одинаковыми предпочтениями ориентации, но со слегка смещенными рецептивными полями.
- Зрительная кора использует стратегию «разделяй и властвуй», разделяя визуальную обработку на последовательность вычислений в десятках различных областей мозга, организованных в определенную иерархию.
- Поднимаясь по визуальной иерархии, нейроны демонстрируют увеличение размеров своих рецептивных полей, более сложные настройки предпочтений и более длительные задержки ответа.
- Нейроны в области V2 реагируют на иллюзорные контуры.
- Пространственный контекст, временной контекст и требования задачи, такие как, например, скрытое внимание, могут модулировать нейронные реакции вдоль вентральной зрительной коры.

## ЛИТЕРАТУРА

См. <http://bit.ly/2TpAg3w> для получения дополнительных ссылок.

- Carandini, M.; Demb, J. B.; Mante, V.; Tolhurst, D. J.; Dan, Y., et al. (2005). Do we know what the early visual system does? *Journal of Neuroscience* 25: 10577–10597.
- Hubel, D. H.; and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology* 195: 215–243.
- Kremkow, J.; Jin, J.; Wang, Y.; and Alonso, J. M. (2016). Principles underlying sensory map topography in primary visual cortex. *Nature* 533: 52–57.
- Markov, N. T.; Ercsey-Ravasz, M. M.; Ribeiro Gomes, A. R.; Lamy, C.; Magrou, L., et al. (2014). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cerebral Cortex* 24: 17–36.
- Schmolesky, M.; Wang, Y.; Hanes, D.; Thompson, K.; Leutgeb, S.; et al. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology* 79: 3272–3278.

Дополнительный контент вы можете скачать по ссылке <http://bit.ly/2TpAg3w>.

# Глава 6

---

## От высших уровней обработки зрительного сигнала к распознаванию образов

Нижняя височная кора головного мозга (ITC) – это высший эшелон визуального потока обработки информации о зрительных образах. Диаграмма Феллемана и Ван Эссена (гл. 1, рис. 1.5) помещает гиппокамп наверху<sup>1</sup>. В то время как зрительные реакции могут быть размещены в гиппокампе, люди с двусторонними повреждениями гиппокампа все еще могут видеть, и очень хорошо. Известный пример – пациент, известный как Х. М., у которого не было известного зрительного дефицита, но он дал начало целой области исследований памяти, основываясь на своей неспособности сохранять новые воспоминания. Гиппокамп не является особой визуальной областью мозга, он получает сигналы от всех сенсорных модальностей (гл. 4).

История того, как нижняя височная кора стала восприниматься и описываться как визуальная область, увлекательна и следует за усовершенствованием возможности производить более точные повреждения мозга и более точные поведенческие эксперименты. В отличие от повреждений гиппокампа, двусторонние повреждения ITC связаны с нарушением распознавания образов у макак (раздел 4.7); также это связано с некоторыми агнозиями у людей (раздел 4.8). Мы начинаем расшифровывать нейронный код, который интерпретирует визуальные сцены.

---

<sup>1</sup> Гиппокамп расположен в медиальных отделах височной доли и представляет собой две вложенные друг в друга согнутые полоски нервной ткани: зубчатую извилину и собственно гиппокамп (аммонов рог, или cornu Ammonis (CA)). В гистологическом плане кора гиппокампа относится к архикортексу, представленному тремя слоями нейронов. Кора височной доли в содействии с гиппокампом участвует в образовании долговременной памяти. – Прим. ред.

## 6.1. ЗОНА С ХОРОШЕЙ КОММУНИКАЦИЕЙ

Нижняя височная кора (ITC) охватывает цитоархитектонические области Бродмана 20 и 21 (рис. 5.1). ITC – это обширное пространство коры, которое обычно подразделяется на заднюю область (PIT), центральную область (CIT) и переднюю область (AIT). Биологи любят вводить людей в заблуждение, используя разные названия для одного и того же, явление, которое можно частично объяснить независимостью исследователей, работающих над смежными темами параллельно и придумывающих свою, новую номенклатуру для описания своих открытий. ITC также упоминается в литературе как области ТЕО и ТЕ. Степень функциональной специализации между различными частями ITC остается плохо изученной, и весьма вероятно, что нам придется подразделить ITC на множество различных промежуточных зон за пределами текущих грубых подобластей, основываясь на структуре связей, нейрофизиологических и вычислительных свойствах.

Как и в большинстве других частей коры, схемы связи ITC обширны и сложны. Когда мы опишем вычислительные модели зрения в гл. 7 и 8, будет очевидно, что большинство моделей представляют собой серьезное упрощение реальной структуры связей. ITC получает прямую топографически организованную информацию из областей V2, V3 и V4 вдоль центральной зрительной коры. ITC также получает часть сигналов из областей V3A, MT и MST, что подчеркивает взаимосвязь между дорсальным и центральным потоками (раздел 4.5). ITC проектируется обратно на V2, V3 и V4. Есть также межполушарные связи между ITC в правом и левом полушариях через главный набор нервных волокон, соединяющих два полушария, мозолистое тело.

ITC также имеет обширные проекции и принимает сигналы от невизуальных областей, включая (i) области, которые обеспечивают важные входные данные для системы памяти медиальной височной доли, такие как периринальная кора, парагиппокампальная извилина и энторинальная кора; (ii) области, участвующие в обработке эмоций, такие как миндалевидное тело; и (iii) области префронтальной коры, которые имеют отношение к принятию решений, планированию и рабочей памяти. Таким образом, с анатомической точки зрения ITC идеально подходит для интерпретации визуальных входов в контексте текущих целей и предыдущей истории, а также для передачи этой информации для принятия поведенческих решений и создания эпизодических воспоминаний.

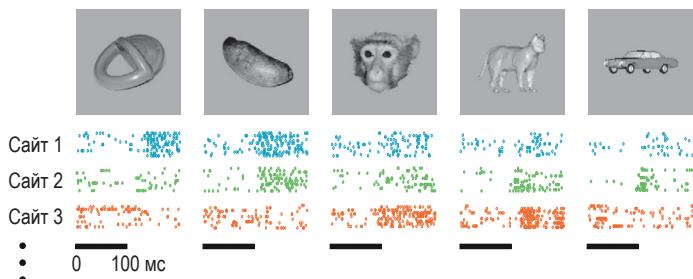
## 6.2. ITC-НЕЙРОНЫ ДЕМОНСТРИРУЮТ ИЗБИРАТЕЛЬНОСТЬ К ФОРМАМ

В течение последних пяти десятилетий героическая школа исследователей изучала отклики нейронов нижней височной коры (ITC) у обезьян из-за общего сходства их зрительной системы с системой человеческого зрения. Большинство, если не все, нейронов ITC демонстрируют визуально стиму-

лированную реакцию, энергично стреляя спайками в ответ на цвет, ориентацию, текстуру, направление движения и форму. Задние части ИТС демонстрируют грубую ретинотопную организацию и почти полное представление контралатерального поля зрения. Размеры рецептивного поля нейронов задней ИТС составляют примерно  $1,5\text{--}4^\circ$ ; в среднем рецептивные поля шире, чем в нейронах V4.

В областях, которые находятся ближе к передней части, вдоль ИТС ретинотопная организация выражена более слабо. Размеры рецептивного поля в более передних частях ИТС часто бывают большими. Оценки сильно различаются: от полей восприятия  $\sim 2^\circ$  до нейронов с полями восприятия, охватывающими несколько десятков градусов. Большинство рецептивных полей в передней ИТС включают фoveальную область.

Примеры ответов трех нейронов ИТС в ответ на пять изображений показаны на рис. 6.1. На этом рисунке каждое изображение повторялось 10 раз, и стохастичность нейронных ответов очевидна в неоднородных формах отклика от одного испытания к другому. Эта вариабельность от испытания к испытанию не является специфической для ИТС и преобладает во всей зрительной коре. Ведется активная дискуссия о происхождении этой изменчивости, которая, по-видимому, не присуща нейронам, но может представлять собой сетевой феномен, который отражает различные уровни внимания, ожидания, положения глаз и другие изменения в ходе эксперимента.



**Рис. 6.1** ♦ ИТС-нейроны избирательны. Пример откликов трех нейронов в нижней височной коре (помеченных «Сайт 1», «Сайт 2», «Сайт 3») на пять различных изображений в градациях серого. Каждая точка представляет собой спайк, каждая строка представляет отдельное повторение (10 повторений каждого демонстрируемого изображения), а горизонтальные черные линии обозначают продолжительность демонстрации изображения (время демонстрации 100 мс).

Данные Хунг и др., 2005

Несмотря на эту вариабельность от испытания к испытанию, есть несколько последовательных особенностей, которые очевидны в откликах нейронов на рис. 6.1. Все три нейрона показывают повышенную спайковую частоту, которая начинается примерно через 100 мс после начала стимула (примерно около конца горизонтальной линии, обозначающей продолжительность предъявления стимула). Эту задержку не следует интерпретировать как реакцию, вызванную продолжительностью стимула; если бы длительность стимула была больше, нейроны все равно начали бы активироваться примерно

через 100 мс после начала стимула. Эти 100 мс отражают задержку для всех вычислений, которые происходят в центральной зрительной коре, прежде чем сигнал достигает этих нейронов ИТС (раздел 5.12). Нейроны разборчивы в предпочтениях стимулов. Нейрон «Сайт 1» показал более сильную реакцию на первые два изображения (игрушка, еда) по сравнению с двумя последними изображениями (синтетическая визуализация кошки, машины). Напротив, нейрон «Сайт 3» показал повышенную реакцию на третье и четвертое изображения (морда обезьяны, кошка).

Исследователи эффективно протестировали реакцию нейронов ИТС на широкий спектр визуальных стимулов. Например, в некоторых исследованиях использовались параметрические дескрипторы в абстрактных образах. Логотип с коллегами обучил обезьян распознавать скрепки, образующие различные трехмерные формы, и впоследствии обнаружили нейроны, которые были избирательны для определенных конфигураций трехмерных объектов. Нейроны ИТС могут управляться изображениями машин, игрушек, лиц и фруктов.

Такой широкий диапазон предпочтений в ответах поначалу может показаться озадачивающим. Возможно, кто-то захочет предположить, что область, которая играет жизненно важную роль в распознавании объектов, содержит нейроны, которые специфически реагируют на объекты в реальном мире. Это могут быть нейроны для банана (то есть нейрон, который реагирует тогда и только тогда, когда исследователи показывают обезьяне изображение банана), нейроны арахиса, нейроны кресла, нейроны лица, нейроны скрепки, нейроны рук и нейроны спагетти с фрикадельками. В самом деле, если на мгновение проигнорировать часть «если и только если», можно обнаружить нейроны, избирательно активируемые этими типами изображений. Как показано в примерах на рис. 6.1, ответы не являются однозначными. Кажется, что нейроны ИТС активируются не только при представлении одного определенного типа объекта в реальном мире, с ответами базового уровня на все остальное. Вместо этого нейроны ИТС демонстрируют ступенчатую активацию с более сильными ответами на одни стимулы по сравнению с другими.

Не ясно, проявляют ли нейроны ИТС какое-либо особое отношение к естественным объектам, таким как бананы или лица. ИТС-нейроны могут иметь достаточно богатый словарь сложных признаков. Эти признаки могут использоваться для представления любого количества встречающихся в природе объектов аналогично формированию слов путем комбинирования различных букв или предложений путем комбинирования слов. Эти особенности можно найти во фрактальных узорах, скрепках, лицах и стульях. Мы вернемся к вопросу о количественном описании этих свойств и ответов нейронов ИТС, когда обсудим современные вычислительные модели визуальной обработки в гл. 7 и 8.

Как обсуждалось в случае нейронов в более ранних зрительных областях (разделы 2.7 и 5.6), на карте ответа ИТС есть четкая топография. Продвигая электрод по траектории, приблизительно касательной к коре, исследователи находят нейроны с аналогичной настройкой. Этот уровень организации может быть представлен «столбцами» нейронов со схожими предпочтениями.

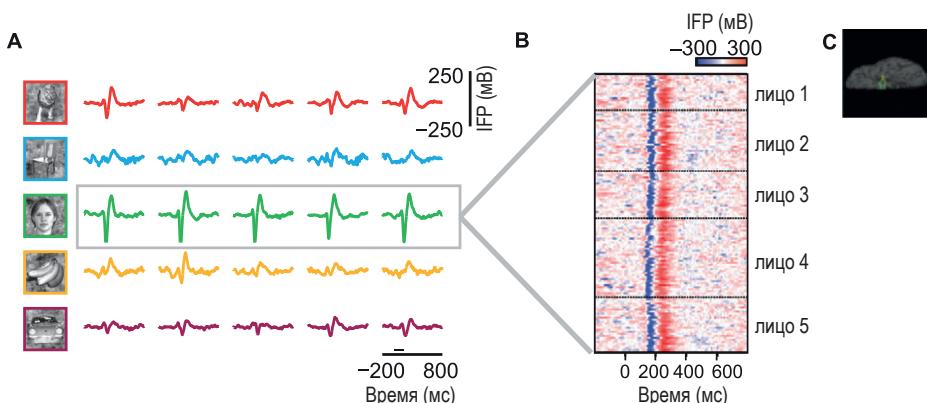
При движении по горизонтали соседние нейроны в ИТС также демонстрируют схожие, но не идентичные предпочтения.

## 6.3. ИЗБИРАТЕЛЬНОСТЬ ВЕНТРАЛЬНОЙ ЗРИТЕЛЬНОЙ КОРЫ ЧЕЛОВЕКА

Не так уж много известно о внутреннем механизме, обрабатывающем визуальную информацию в человеческом мозге. Основным источником информации о внутренней работе вентральной зрительной коры человека являются инвазивные нейрофизиологические регистраций у пациентов с эпилепсией, представленные работой Пенфилда, которую мы обсуждали в разделе 4.9. Некоторые пациенты с эпилепсией поддаются фармакологическому лечению. В случаях фокальной эпилепсии, не поддающейся лечению современными лекарствами, одним из главных методов терапии является хирургическое удаление эпилептогенного очага. В большинстве случаев эта хирургическая процедура требует прежде всего тщательного картирования активности мозга, чтобы определить источник сигнала, приводящего к судорогам, а также убедиться, что иссечение мозга не навредит когнитивной функции мозга. С этой целью нейрохирурги обычно вживляют электроды в мозг человека. Поскольку современные неинвазивные методы слишком грубы, чтобы картировать центр происхождения припадков, нейрохирурги обычно имплантируют несколько десятков электродов в различные области мозга в надежде точно определить и картировать момент возникновения припадка. После имплантации пациенты остаются в больнице примерно на одну неделю для наблюдения, предоставляя исследователям редкую и уникальную возможность тщательно изучить функцию человеческого мозга с высоким пространственно-временным разрешением и с высоким соотношением сигнал/шум по сравнению со всем, что можно сделать, используя неинвазивные технологии, такие как ЭКГ.

Расположение электродов строго определяется клиническими потребностями. Иногда эти электроды размещают вдоль вентральной зрительной коры. Пример визуально избирательных ответов в ИТС человека показан на рис. 6.2. Потенциалы внутричерепного электрического поля человека, то есть напряжение, регистрируемое на этих электродах, соответствуют многим отличительным чертам ИТС-ответов макак. Сигналы вдоль вентральной зрительной коры головного мозга человека также показывают ограниченные рецептивные поля, которые увеличиваются в размере от фoveальной области к периферии и от одной области к другой. Потенциалы поля регистрации вдоль вентральной зрительной коры человека также являются выборочными и градуированными (рис. 6.2А). Сигналы потенциала внутричерепного поля демонстрируют вариабельность от испытания к испытанию, однако визуально обусловленные ответы на стимулы можно легко оценить лишь в единичных испытаниях (рис. 6.2В). Было проведено гораздо больше нейрофизиологических исследований, изучающих реакцию обезьян, по сравнению

с исследованиями на людях. Многие тонкие детали реакций вентральной зрительной коры человека остаются неизученными. Например, насколько нам известно, никто не исследовал реакцию вентральной зрительной коры человека на фрактальные узоры или скрепки, как это показано в исследованиях на обезьянах.



**Рис. 6.2** ♦ ИТС человека также демонстрирует избирательность к форме. Пример электрода, описывающий физиологические реакции на демонстрацию 25 различных образцов объектов, принадлежащих к пяти различным категориям. (А) Ответы на каждый из 25 различных образцов (каждый цвет обозначает отдельную категорию изображений; каждая линия представляет ответ на отдельный образец). (Б) Растворный график, показывающий каждое испытание в ответах на пять образцов лиц. Каждая строка – это повторение; горизонтальные линии разделяют экземпляры; цвет показывает напряжение (см. шкалу справа). (С) Расположение электрода. Воспроизведено по Лю и др., 2009

Однако отсутствие доказательств не означает доказательство отсутствия. Насколько мы можем судить, ответы вдоль вентральной зрительной коры головного мозга человека демонстрируют избирательность к широкому спектру зрительных образов, так же, как и у их родственников – обезьян-макак.

Время ожидания ответа в человеческом мозге кажется немного больше, чем у макак, возможно, из-за большего размера мозга или потому, что требуется больше вычислительных шагов, прежде чем информация достигнет ИТС человека. В рамках скучных и предварительных нейрофизиологических данных, доступных сегодня, и по разумным соображениям многие свойства ИТС макак соответствуют свойствам ИТС человека.

Следует отметить, что не совсем понятно, как достоверно сравнивать области мозга и функциональные реакции между людьми и обезьянами (или любой другой парой видов, разделенных длительными периодами эволюционного расхождения). Прежде всего нам следует с осторожностью сравнивать спайки у обезьян с сигналами внутричерепного потенциала электрического поля у людей. Оказывается, что сигналы потенциала поля демонстрируют модели избирательности, аналогичные спайковым сигналам в ИТС обезьяны. Потенциальные ответы на более грубые поля несколько менее четки, чем

спайки, с точки зрения их способности разделить различные стимулы, возможно из-за усреднения по многим нейронам.

Более сложное рассмотрение связано с установлением строгой гомологии между видами. Кажется очевидным, что глаза обезьян гомологичны глазам человека. Кроме того, хотя нейроанатомические связи у людей остаются недостаточно ясными, весьма заманчиво предположить, что первичная зрительная кора обезьяны может быть гомологичной первичной зрительной коре головного мозга человека. По мере того как мы углубляемся вентральную зрительную кору за пределами V1, гомологии становятся более расплывчатыми. Независимо от того, сможем ли мы создать уникальную эволюционно точную взаимно однозначную карту между конкретными структурами мозга у разных видов, очевидно, ясно, что вентральная зрительная кора человека демонстрирует быстрые и избирательные реакции на сложные образы, которые качественно подобны тем, которые наблюдаются у обезьян.

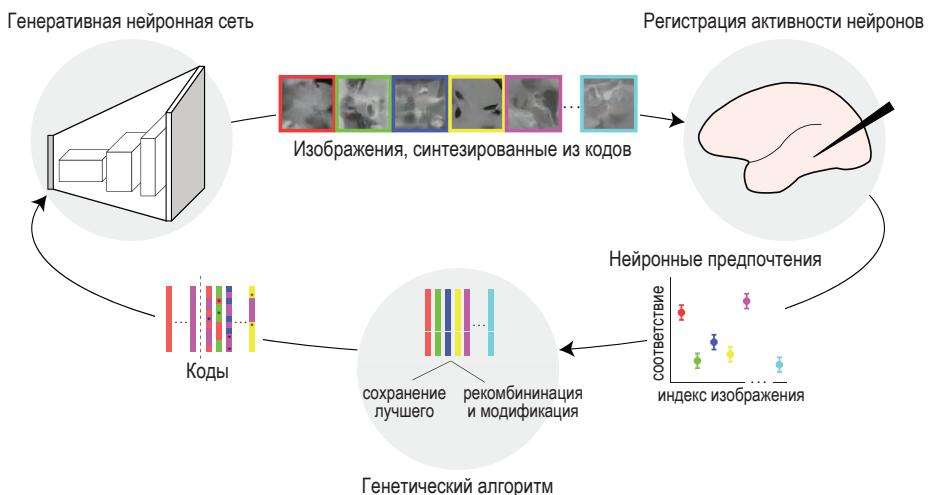
## 6.4. ЧЕГО ХОТЯТ НЕЙРОНЫ ИТС НА САМОМ ДЕЛЕ?

ITC-нейроны, кажется, реагируют на самые разные образы, которые исследователи использовали для проверки своих предпочтений в отношении стимулов. Время регистрации ограничено, и исследователям необходимо сделать выбор, какие стимулы использовать в эксперименте; мы рассмотрели эту проблему в разделе 5.11. Как правило, исследователи выбирают стимулы на основе сочетания примера из предыдущих исследований (если определенный тип стимула работал раньше для управления нейронами в данной области, он должен работать и сейчас) или интуиции, основанной на преобладании статистики естественных стимулов (вполне логично предположить, что нейроны могут давать сильные ответы на входящие стимулы, которые животное наблюдает ежедневно, или аргументировать предполагаемую эволюционную важность определенных классов стимулов). Кроме того, важные достижения в области свойств предпочтения нейронов были основаны на некотором смысле случайных открытиях.

Все эти эксперименты несут в себе потенциальные предубеждения, вносимые исследователями при выборе стимулов. Очевидно, мы можем обнаружить настройку только на те стимулы, которые исследуем. Даже название этого раздела имеет сильный антропоморфный оттенок. Нейроны на самом деле ничего не «хотят». Вопрос указывает на то, какие типы визуальных стимулов максимально активируют данный нейрон (в смысле запуска большего количества спайков). Как подчеркивалось в разделе 5.11, критическая трудность в выяснении предпочтений нейронов в ответах связана с *проклятием размерности*: слишком много возможных тестовых изображений и слишком мало времени на эксперимент.

Многообещающее направление исследований для выяснения предпочтений признаков в ITC включает изменение стимулов в реальном времени, продиктованное предпочтениями нейрона. Недавняя работа, основанная на этом подходе, предполагает, что нам, возможно, потребуется переосмыс-

лить нейронный код для признаков в ИТС (и, возможно, также в более ранних визуальных областях). Одно из первых применений этого подхода было разработано группой Чарльза Коннора, который позволил нейронам самим показывать то, что им нравится, а не демонстрировать реакцию на предвзятые стимулы. Недавняя работа Уилла Сяо заключалась в разработке вычислительного алгоритма, способного генерировать изображения на основе частоты срабатывания нейронов (рис. 6.3). Исследователи объединили генератор изображений и генетический алгоритм, основанный на спайковой частоте, в качестве функции соответствия, чтобы направлять эволюцию стимулов в реальном времени. В данном поколении ученые исследуют ответы на набор изображений. Изображения, которые вызывают высокую спайковую частоту, сохраняются, а остальные модифицируются и рекомбинируются генетическим алгоритмом в сочетании с алгоритмом генерации изображений.



**Рис. 6.3 ♦** Позволим нейронам раскрыть свои предпочтения. Подход к беспристрастному исследованию настройки нейронов. Генеративная нейронная сеть используется для создания изображений путем инвертирования модели зрительного распознавания (раздел 9.9, рис. 9.10). Синтетические изображения представлены при регистрации активности нейронов. Ответы нейронов используются в качестве индекса соответствия, чтобы направлять генетический алгоритм для выбора нового поколения улучшенных изображений. Воспроизведено по Понсе и др., 2019

В разделе 8.5 мы познакомимся с глубокими иерархическими моделями зрения, которые начинаются с пикселей и дают высокоуровневое функциональное представление изображения. Кроме того, в разделах 9.8 и 9.9 мы представим генеративные состязательные сети, которые создают изображения путем инвертирования глубокой иерархической модели. Генеративный алгоритм, развернутый Сяо и его коллегами, вдохновленными работой в области машинного обучения для создания генераторов изображений, по сути,

является перевернутой версией глубоких иерархических вычислительных моделей, начиная с высокоуровневых функций и заканчивая генерацией изображения.

Запустив этот генеративный вычислительный алгоритм при регистрации активности нейронов в ИТС, Сяо и его коллеги обнаружили изображения, вызывающие более высокую спайковую частоту, чем естественные изображения, которые использовались ранее для проверки ответов нейронов. Исследователи называют эти синтетические изображения «суперстимулами». Эти суперстимулы содержат натуралистические комбинации текстур и широких мазков, которые можно сравнить разве что с картинами абстракционистов. Фундаментальная новая концепция здесь заключается в том, что нейроны могут быть лучше всего активированы комбинациями сложных признаков, которые с трудом поддаются описанию словами. В отличие от антропоморфных описаний предпочтений характеристик в ИТС («этот нейрон любит лица», «этот нейрон любит стулья», «этот нейрон любит изогнутые формы»), новое направление исследований предполагает, что нейроны могут активироваться сложными визуальными образами, которые не поддаются определению на основе языка. Богатый базовый набор нейронов, настроенных на такие сложные признаки, способен позволить организму различать объекты реального мира, но базовый набор не обязательно должен быть основан на пиктограммах реальных объектов.

## 6.5. Нейроны ИТС демонстрируют толерантность к преобразованиям объектов

Как подчеркивается в разделах 1.4 и 3.4, важным свойством распознавания является способность узнавать объекты, несмотря на преобразования изображений на уровне пикселей (рис. 3.6). Поэтому интересно задать вопрос, сохраняется ли визуальная избирательность на нейронном уровне, как описано в предыдущих разделах, при преобразованиях изображений. Например, будет ли нейрон, показанный в верхнем ряду на рис. 6.1, продолжать выборочно реагировать на первые два объекта, если они будут показаны в другом масштабе, в другом положении относительно точки фиксации или в другом цвете?

ИТС-нейроны демонстрируют значительную степень толерантности к определенным преобразованиям объектов. Они обладают большими рецептивными полями и, следовательно, демонстрируют большую толерантность к изменениям положения объекта по сравнению с нейронами в более ранних частях центральной зрительной коры. ИТС-нейроны также демонстрируют аналогичные ответы, несмотря на существенные изменения размера стимулов на сетчатке. Толерантность не обязательно означает, что частота спайков в ответ на данный объект должна быть *идентичной* для различных вариантов трансформации объекта. Хотя на абсолютную частоту спайков трансформация, такая как изменение размера стимула, очевидно, влияет,

но предпочтения в порядке ранжирования между различными объектами – и, следовательно, относительные предпочтения стимулов – сохраняются. ITC-нейроны также демонстрируют определенную толерантность к вращению объекта. Кроме того, хотя большинство форм обычно определяют изменения яркости изображения, нейроны ITC еще реагируют на формы, определяемые другими сигналами. Например, форма может определяться шаблонами шума, которые движутся когерентным образом, или изменениями текстуры без скачка яркости между одной и другой.

Ярким примером толерантности к преобразованиям объекта были регистрации откликов отдельных нейронов медиальной височной доли (не ITC) у пациентов с эпилепсией. Регистрируя активность гиппокампа, энторинальной коры, миндалины и парагиппокампа извилины, исследователи обнаружили нейроны, которые дают ответы на несколько объектов в пределах одной семантически определенной категории. Они также обнаружили некоторые нейроны, которые демонстрируют значительную избирательность по отношению к отдельным людям или дорожным знакам. Например, один нейрон избирательно реагировал на изображения, на которых присутствовал бывший президент Билл Клинтон; другой нейрон предпочел снимки известной актрисы Дженифер Энистон. Примечательно, что изображения, которые вызывали реакцию этих нейронов, довольно сильно отличались друг от друга с точки зрения их пиксельного содержания – от черно-белого рисунка до цветных фотографий с разными позами и видами. Такое экстремальное сочетание избирательности и толерантности было описано не в областях ITC, а скорее в медиальной височной доле. Как отмечалось в начале этой главы, эти медиальные структуры височных долей получают зрительные сигналы, но не являются строго зрительными областями. Повреждение структур медиальной височной доли, по-видимому, не связано с каким-либо явным нарушением зрения или каким-либо другим дефицитом восприятия, а скорее с проблемами памяти. Следовательно, вполне вероятно, что эта комбинация избирательности и толерантности отражает *считывание активности из популяции нейронов ITC для преобразования сенсорных входов в эпизодические воспоминания*.

## 6.6. НЕЙРОНЫ МОГУТ ЗАВЕРШАТЬ ФОРМЫ ЧАСТИЧНО ВИДИМЫХ ОБЪЕКТОВ

Во время естественного наблюдения объекты нередко видны только частично из-за плохого освещения или из-за того, что перед ними есть другие объекты (раздел 3.5). В ранних зрительных областях с небольшими рецептивными полями окклюзия может покрывать всю часть зрительного поля, в котором заинтересован данный нейрон. Напротив, в более высоких визуальных областях с большими рецептивными полями окклюзия может только частично препятствовать сигналу, получаемому нейроном. Степень толерантности к преобразованиям объектов, описанная в предыдущем разделе, предпо-

лагает, что нейроны потенциально могут успешно обрабатывать входные данные, которые содержат только некоторые из предпочтительных признаков объекта.

Действительно, ИТС демонстрирует высокую степень устойчивости к окклюзии. Нейронные отклики в ИТС могут завершать образы и сохранять свою избирательность, даже когда более половины предпочтительных для нейрона особенностей объекта невидимы. Как на поведенческом уровне (раздел 3.5), так и на нейрофизиологическом уровне завершение формы требует дополнительного времени вычислений: задержки зрительных избирательных реакций, вызванных частично видимыми объектами, примерно на 50 мс больше, чем те, которые вызываются полностью видимыми. Эти наблюдения предполагают необходимость дополнительной обработки сигнала, чтобы сделать выводы на основе частичной информации. Мы вернемся к этому моменту в разделах 7.6 и 8.16, когда будем обсуждать вычислительные механизмы завершения формы.

В предыдущем разделе мы отметили, что толерантность к преобразованиям объекта не обязательно означает, что нейронные реакции на преобразованные версии объекта должны быть идентичными. Масштабирование, вращение, изменение цвета и другие преобразования могут изменить частоту срабатывания нейрона, а толерантность относится к поддерживаемой нейронной селективности. Таким же образом завершение форм не означает, что нейронные реакции на сильно закрытые объекты идентичны ответам на полностью видимые варианты; завершение формы на уровне нейронов указывает на то, что селективность сохраняется.

В то время как при определенных преобразованиях изображения, таких как изменение масштаба или положения, одни и те же признаки объекта остаются видимыми (хотя и в разных местах и в разных размерах), другие преобразования изображения, такие как трехмерное вращение или сильная окклюзия, изменяют то, какие элементы объекта видны, а какие нет. Поэтому неудивительно, что исчезновение некоторых особенностей объекта и появление новых деталей во время вращения может привести к разной частоте спайков. Что примечательно, так это то, что некоторые относительные предпочтения стимулов сохраняются в этих условиях, несмотря на существенные изменения на уровне пикселей.

## 6.7. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ВЫХОДЯТ НА ПЕРЕДНИЙ ПЛАН

Наблюдение за тем, что отдельные нейроны могут проявлять высокую степень избирательности и толерантности к преобразованиям изображения, не следует рассматривать как подразумевающее, что существует взаимно однозначное соответствие между активностью отдельного нейрона и распознаванием определенного объекта. Идея взаимно однозначной карты соответствия между нейронами и конкретными объектами ошибочно упоми-

нается как теория «бабушкиного нейрона» (*grandmother cell*)<sup>1</sup>. Система «один в один» была бы чрезвычайно громоздкой и хрупкой. Потеря одного нейрона могла бы привести к неспособности распознать этот конкретный объект. Кроме того, в большинстве случаев считающие нейроны зависят от сигналов сотен тысяч других нейронов и не могут надежно или исключительно управляться одним входом.

Как отмечалось в разделе 6.2, соседние нейроны в зрительной коре, как правило, демонстрируют аналогичные настройки избирательности. В настоящее время мы не можем контролировать активность каждого нейрона в локальной области, однако нахождение нейрона со специфической функцией настройки, скорее всего, означает существование в его окрестности большого количества нейронов с аналогичными свойствами. Идея «бабушкиного нейрона», придуманная Джерри Летвином в 1969 году, относилась ко всей популяции клеток с идентичными свойствами селективности и толерантности (в первоначальном описании он упоминал его скорее как «материнскую клетку», чем как «бабушкуну клетку»). Понимаемая в исходном определении, идея «бабушкиного нейрона», то есть популяции, скорее всего, соседних нейронов, которые проявляют избирательность и толерантность к похожим свойствам стимула, является адекватным описанием нейрональной настройки всей зрительной коры. Ганглиозные клетки сетчатки являются «бабушкиными» для изменений освещенности в редких определенных местах зрительного поля, первичные нейроны зрительной коры являются «бабушкиными» для ориентированных линий, а нейроны ИТС – «бабушкиными» для элементов сложной формы.

Хотя каждый нейрон предпочитает одни формы другим, объем информации, передаваемой отдельными нейронами об общей форме, ограничен. Кроме того, кажется, что в ответах нейронов в любом исследовании присутствует значительный «шум». Термин шум в некоторой степени неверен, поскольку он относится к изменчивости тайминга и количества спайков от испытания к испытанию, как показано на рис. 6.1. Является ли это настоящим шумом или частью сигнала и какова причина этих флюктуаций, остается предметом дискуссий. Для надпороговых стимулов восприятие образа достаточно устойчиво: вы можете тысячу раз взглянуть на форму буквы А, и она всегда будет выглядеть как А. Следовательно, где-то в мозге есть некий постсинаптический нейрон, который получает сигналы от капризных пре-синаптических нейронов, которые дают разные ответы на предположительно идентичные входные данные в некоем производимом эксперименте, и он сбрасывает со счетов эту вариабельность и расшифровывает то, что реально существует в окружающем мире.

Могут ли животные использовать нейронные ответы популяции несколько капризных ИТС-нейронов, чтобы различать объекты в единичных испытаниях? Критический упор делается на единичные испытания. В отличие от того, что делают многие исследователи при анализе нейронных регистраций, мозг не имеет возможности усреднять данные по испытаниям (нам не нужно

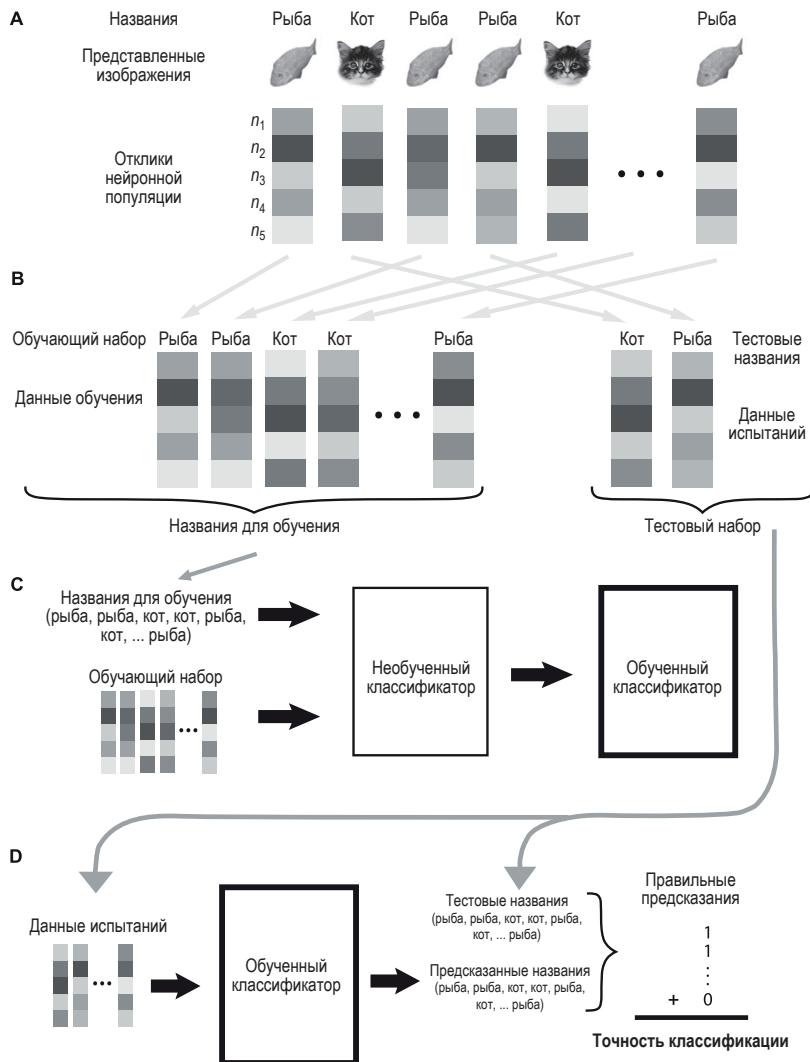
<sup>1</sup> Предполагается, что существует единичный нейрон, реагирующий только на конкретную сущность, например бабушку наблюдателя. Также его называют «нейрон Дженнifer Энистон». – Прим. ред.

смотреть на букву А десять раз, чтобы распознать ее). Мозг не обязан делать выводы из активности отдельного нейрона. Любой нейрон в коре головного мозга получает входные данные примерно от 10 000 других нейронов. Такая популяция могла бы показать интересные свойства, которые помогают решить или решают проблемы, связанные с интерпретацией выходного сигнала одного нейрона.

Чоу Хунг и его коллеги занялись этим вопросом, последовательно записав активность сотен нейронов ИТС и используя классификаторы машинного обучения для декодирования активности псевдопопуляции нейронов в отдельных испытаниях. Термин псевдопопуляция используется потому, что эти нейроны не регистрировались одновременно. Метод декодирования машинного обучения направлен на изучение карты между (i) схемами активности популяции нейронов в ответ на набор изображений и (ii) названиями объектов на этих изображениях (рис. 6.4). Рассмотрим эксперимент, в котором мы представляем изображения кошек или рыб. Пусть  $jx_i$  представляет активность нейрона  $i$  в ответ на изображение  $j$ . Например,  $x$  может представлять общее количество импульсов, испускаемых нейроном на заданном временном промежутке. Из-за задержки откликов ИТС (рис. 6.1) мы можем рассматривать окно от 100 до 300 мс после начала стимула. Популяционный ответ  $N$  нейронов на изображение  $j$  получится  $jx = [jx_1, \dots, jx_N]$ .

Если мы представим, что все эти входные сигналы могут проецироваться на данный постсинаптический нейрон, мы можем записать общий агрегированный сигнал в постсинаптический нейрон как взвешенную сумму всех этих сигналов:  $w_1x_1 + \dots + w_Nx_N$ . Эту сумму можно рассматривать как меру синаптической силы, воздействие, которое данный сигнал будет иметь на постсинаптический нейрон. Может ли такой нейрон, расположенный ниже по потоку сигнала, обнаружить присутствие кошки или рыбы? Мы можем построить детектор, который по активности популяции нейронов будет считывать, содержит ли изображение, показанное в данном испытании, кошку или рыбу. Мы установим порог для общих объединенных сигналов, для краткости  $g(w \cdot x)$ , где  $g$  обозначает нелинейную функцию, такую как сигмоида,  $w$  и  $x$  – векторы размерности  $N$ , определенные выше, а символ  $\cdot$  представляет скалярное произведение. Мы можем определить, что если  $g$  больше порогового значения, изображение содержит кошку, а если  $g$  меньше порога, то изображение содержит рыбу. Алгоритмы машинного обучения предлагают несколько хитрых способов выбора этих весов  $w$ , чтобы минимизировать количество ошибок классификации, которые делает алгоритм. Мы не будем вдаваться в подробности здесь, но, чтобы быть конкретными, можно представить, что мы используем классификатор машины опорных векторов (SVM) с линейным ядром, что является надежным способом выбора этих весов и методом, которому следует Чжоу Хунг и его коллеги. Этот метод можно распространить на многие категории, а не только на двоичную классификацию. Главный вывод состоит в том, что если можно изучить надежный и простой (например, линейный) классификатор, то псевдопопуляция нейронов содержит достаточную информацию о стимулах, которую можно легко извлечь с помощью биологически достоверных вычислений (скалярное произведение с последующей нелинейностью).

## От высших уровней обработки зрительного сигнала к распознаванию образов ✦ 165



**Рис. 6.4** ✦ Расшифровка популяционных откликов нейронов. Основные этапы обучения и тестирования классификатора. (A) Иллюстрация эксперимента, в котором изображения кошек и рыб показывались подопытному в случайному порядке, в то время как одновременные записи производились с пяти нейронов/каналов. Уровень оттенков серого обозначает активность каждого нейрона/канала. (B) Точки данных и соответствующие картинки выбираются случайнным образом для включения либо в обучающий набор, либо в тестовый набор. (C) Точки обучающих данных и обучающие картинки передаются неподготовленному классификатору, который «узнает», какая нейронная активность полезна для предсказания того, какое изображение было показано, – таким образом становясь «обученным» классификатором. (D) Тестовые данные передаются обученному классификатору, который прогнозирует, какие названия соответствуют каждой немаркированной точке тестовых данных. Эти предсказанные названия затем сравниваются с реальными тестовыми картинками (то есть фактическими названиями, которые были представлены при записи тестовых данных), и вычисляется процент правильных предсказаний, чтобы дать общую точность классификации. Измененное из работы Мейерса и Креймана, 2011

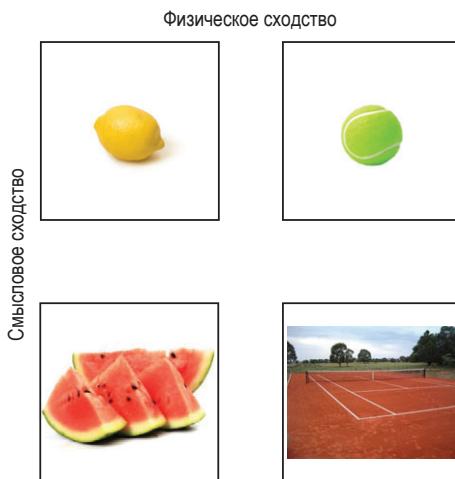
Используя этот подход, Хунг с коллегами обнаружили, что относительно небольшая группа нейронов ITC ( $N \sim 200$ ) может поддерживать категоризацию объектов довольно точно: до ~90 % точности в задаче, состоящей из восьми возможных категорий (где шанс – один из восьми). Кроме того, реакция псевдопопуляции может экстраполироваться на изменения в масштабе и положении объекта. Другими словами, можно подобрать значения  $w$ , используя отклики  $x_1$  на изображения в определенном масштабе, а затем использовать отклики  $x_2$  на изображения в другом масштабе для точного прогнозирования названий объектов. Таким образом, даже если каждый нейрон передает только зашумленную информацию о различиях формы, небольшая популяция нейронов может быть сильна в различении зрительных образов в отдельных испытаниях, даже если экстраполировать на трансформированные версии изображений, используемых для обучения.

## 6.8. НЕЙРОНЫ ITC БОЛЬШЕ ОЗАБОЧЕНЫ ФОРМОЙ, ЧЕМ СМЫСЛОМ

В предыдущем разделе мы рассмотрели, можно ли определить, какая категория объектов была представлена обезьяне, если считывать нейронную активность в ITC. Вместо того чтобы расшифровывать категорию объекта, можно поставить другую задачу – определить, какой именно образец был представлен обезьяне. Популяция нейронов ITC преуспевает и в этом вопросе. Количественное сравнение идентификации и категоризации образцов сложно, потому что эти две задачи не равны с точки зрения сложности. Во-первых, в эксперименте, описанном в предыдущем разделе, было восемь категорий и около 80 экземпляров. Поэтому даже чисто случайно определить категорию объекта легче, чем точно определить экземпляр. Выравнивание вероятностей может быть легко достигнуто путем случайной подвыборки и выбора только восьми экземпляров. Тем не менее это не очень хорошо решает более сложную проблему в данном типе сравнения: легче отличить изображение лица от изображения дома, чем различать два разных дома.

Имеют ли нейроны ITC какую-либо информацию о категории, или форма объекта является основной переменной, которая оценивается в ITC? Чтобы ответить на этот вопрос, нам нужно лучше определить, что мы подразумеваем под «категорией». Словесная категория обычно связана с семантическим названием. Один из способов отделить семантическую информацию от чистой информации о форме – рассмотреть объекты, которые физически похожи, но семантически различны, и наоборот (рис. 6.5). Например, лимон похож на теннисный мяч по цвету, размеру и приблизительной форме. Однако семантически лимон ближе к арбузу или дереву, а теннисный мяч семантически ближе к теннисному корту или теннисной ракетке. На сегодняшний день нет доказательств того, что нейроны ITC могут связывать теннисный мяч с теннисным кортом или лимон с арбузом. Вместо этого есть свидетельства того, что ответы нейронов ITC на физически похожие изображения

ближе друг к другу, чем ответы на семантически похожие, но физически разные объекты.



**Рис. 6.5** ♦ ITC-нейроны больше озабочены сходством формы, чем сходством смысла. Эти изображения имеют большее физическое сходство по горизонтали и большее семантическое сходство по вертикали. Ответы в ITC более точно отражают физические свойства стимула, включая цвет, размер и форму

Группа Эрла Миллера провела элегантную серию экспериментов, посвященных вопросу категоризации. Они создали синтетические образы кошек и собак и морфировали их между собой таким образом, чтобы они представляли собой непрерывный ряд образов, похожих на соседние, так что между соседними изображениями в этом ряду не было категориальной разницы вследствие их схожести. Они обнаружили, что ответы нейронов ITC лучше коррелируют со сходством формы, чем с категориальной принадлежностью. Они также записали отклики нейронов в префронтальной коре, которая является одной из мишней нейронов ITC. В отличие от нейронов ITC, ответы этих нейронов действительно отражали категориальные границы, зависящие от поставленной задачи.

Другой интригующий случай, когда ответы нейронов, казалось, были отделены от чистой информации о форме, – это случай записи нейронов медиальной височной доли человека, обсужденный ранее (раздел 6.5). Эти нейроны, кажется, несут семантическую информацию, которая выходит за рамки сходства физических форм, и эти нейроны получают прямую или косвенную информацию от передней ITC, но они не являются частью ITC.

Как неоднократно указывалось, отсутствие доказательств не должно трактоваться как доказательство отсутствия чего-то. Вполне возможно, что существует семантическая информация, которая в ITC может быть отделена

от чистой информации о форме, но пока нет четких доказательств этого. Семантическая информация – важнейший компонент того, как мы используем язык. Помимо медиальной височной доли и префронтальной коры, структуры, отвечающие за язык, возможно, содержат нейроны, которые реагируют на семантическую информацию. Более того, вполне вероятно, что такие семантические нейроны могут проецироваться обратно вентральную зрительную кору и модулировать или обострять визуально стимулированные ответы.

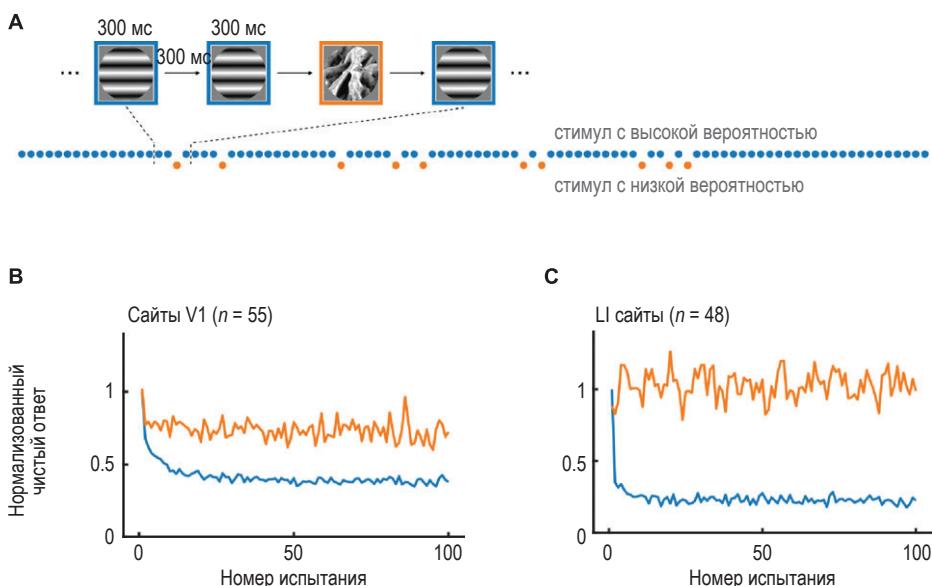
## 6.9. АДАПТАЦИЯ НЕЙРОННЫХ РЕАКЦИЙ

Нейроны зрительной коры особенно чувствительны к изменениям. Реакции нейронов динамически зависят от временного контекста. Временной контекст может кардинально изменить зрительное восприятие (раздел 3.8), как при иллюзорном восприятии восходящего движения после фиксации на водопаде, из-за адаптации. Как следствие адаптации ответы нейронов ITC, как и в более ранних частях зрительной коры, являются временными (разделы 2.9, 5.7 и 5.12). Если постоянный стимул отображается в течение многих секунд, нейронные реакции делятся всего несколько сотен миллисекунд.

Адаптация – это эволюционно сохраняющееся свойство обработки изображений, которое также преобладает в других сенсорных системах. Одна из функций адаптации – это, вероятно, экономия энергии за счет уменьшения количества спайков, вызываемых неизменным стимулом. По крайней мере частично, биофизические механизмы, лежащие в основе такого подавления, могут быть обусловлены внутренними изменениями в нейроне посредством временной модуляции его мембранный проводимости. Однако адаптация также очевидна в гораздо более длительных временных масштабах, чем при предъявлении единственного стимула. Например, воздействие адаптивного стимула приводит к снижению нейронной реакции на последующие предъявления тех же или похожих стимулов – явление, известное как подавление повторения. Повторения не обязательно должны быть близкими по времени. Подавление также очевидно, даже когда есть другие промежуточные стимулы, хотя его сила уменьшается с интервалом времени между повторениями.

Адаптация очевидна во многих временных масштабах. Как обсуждалось в разделах 2.9 и 5.7 (рис. 5.6), ответы нейронов обычно недолги и быстро затухают в течение одного испытания в масштабе сотен миллисекунд, даже если стимул остается на экране. Подавление повторения – это проявление адаптации в масштабе нескольких попыток, обычно происходящих в течение нескольких секунд. На рис. 6.6 показан пример эксперимента, в котором эффекты адаптации могут проявляться в течение нескольких минут. В так называемой «парадигме странностей» (oddball paradigm) данный стимул повторяется несколько раз (стимул с высокой вероятностью показан синим), тогда как другой стимул отображается редко (стимул с низкой вероятностью показан оранжевым цветом). На рис. 6.6В–С показаны средние популяционные ответы от нескольких нейронов в первичной зрительной коре го-

ловного мозга крысы (V1) и в более высокой зрительной области, называемой латерально-интермедиальной областью (LI). В то время как существует общее согласие относительно того, что составляет первичную зрительную кору у разных видов, менее очевидно, как установить гомологическую связь между высшими визуальными областями у разных видов, и, следовательно, номенклатура названий между видами расходится. Повторное предъявление стимула с высокой вероятностью приводит к резкому снижению нейронных ответов в течение эксперимента (синий). Напротив, стимул с низкой вероятностью вызывает большую реакцию, особенно в области LI. Этот эффект может помочь обнаружить новые стимулы или изменения в окружающей среде.



**Рис. 6.6** ♦ Нейронная адаптация увеличивает значимость новых стимулов. (A) Парадигма странностей, в которой один стимул представлен с высокой вероятностью (синий), а другой стимул – с низкой вероятностью (оранжевый). (B)–(C) Нормализованные средние популяционные ответы нейронов первичной зрительной коры головного мозга крысы (B) и латерально-интермедиальной области (C) в зависимости от числа испытаний для стимулов с низкой и высокой вероятностями. Измененное по Винкен и др., 2020

Адаптация происходит во всей зрительной системе. Последствия адаптации сильнее в более высоких областях, таких как ITC или область LI у крысы, по сравнению с более ранними нейронами, такими как V1, вероятно, из-за кумулятивных эффектов иерархического каскада нейронов, каждый из которых демонстрирует все более сильные эффекты адаптации, влияющие на следующий этап. Другими словами, адаптация приводит к снижению ответа в RGC и LGN, что, в свою очередь, подразумевает более слабый входной сигнал в V1, и это усугубляется внутренними эффектами адаптации в V1. Более слабые сигналы V1 приводят к уменьшенному входу в V2, что сочетается

с внутренними эффектами адаптации в V2, и так далее. Еще один эффект, который может способствовать усилению адаптации на более высоких стадиях, заключается в том, что более ранние области значительно чувствительнее к небольшим движениям глаз, что снижает сходство входных сигналов для длительных стимулов или повторений одного и того же стимула.

## 6.10. ПРЕДСТАВЛЕНИЕ ВИЗУАЛЬНОЙ ИНФОРМАЦИИ ПРИ ОТСУТСТВИИ ВИЗУАЛЬНОГО СТИМУЛА

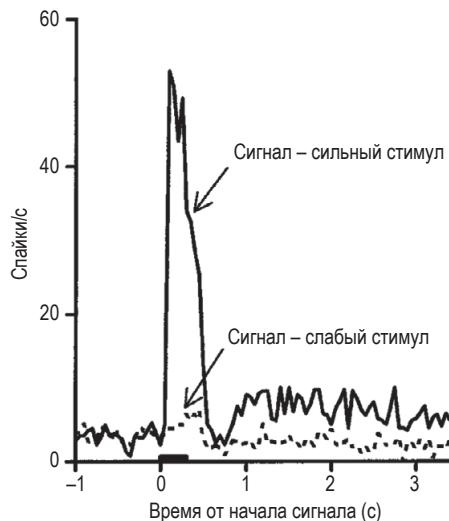
В процессе восприятия продолжительное воздействие стимула часто приводит к временному снижению чувствительности к его особенностям. Затяжные эффекты после удаления стимула называются последействиями, которые были описаны для широкого диапазона свойств стимулов от низкого до высокого уровня зрительной коры; эти эффекты считаются связанными с адаптацией.

Помимо последействия, воздействие стимула оставляет след в памяти, который позволяет подопытным вспомнить то, что они только что видели. Классическим экспериментом, используемым для изучения эффектов памяти в коротких временных масштабах, является задача отложенного сопоставления с выборкой. Объектам предоставляется изображение, которое исчезает через несколько секунд. После этой задержки показывается второе изображение, и подопытные должны указать, соответствует ли второй стимул первому (потому, что он идентичен, или потому, что это масштабированная или повернутая версия того же объекта, или потому, что они совпадают по цвету, или любое другое возможное совпадение свойств). Обычно период задержки представляет собой пустой экран. Чтобы наблюдатели могли выполнить эту задачу, нейроны где-то в мозге должны иметь возможность сохранять информацию о предыдущем стимуле во время демонстрации этого пустого экрана. Такая информация, хранящаяся в течение нескольких секунд, называется *рабочей памятью*.

Оказывается, что, хотя ответы нейронов в ИТС резко снижаются в отсутствие визуальной стимуляции, активность полностью не возвращается к исходному уровню (рис. 6.7). Вместо этого во время такой задержки нейроны ИТС сохраняют небольшую активность выше базовой. Кроме того, эта активность во время задержки соответствует стимулу: нейрон будет поддерживать более высокую активность, если его реакция на предыдущий стимул была выше.

Некоторые исследователи интерпретируют эти нейронные реакции в период задержки при отсутствии визуальной стимуляции как пример зрительного воображения. Они утверждают, что подопытные воображают образ стимула, чтобы сохранить его в памяти во время задержки. В той степени, в которой это соответствует действительности, кажется, что нейроны ИТС могут демонстрировать избирательную реакцию, которая соответствует внутреннему зрительному представлению животного, независимо от сенсорных входов. Непосредственно проверить эту идею непросто из-за сложности объ-

ективного выявления у животных мысленных зрительных образов. У людей несколько исследователей измерили нейронные корреляты таких образов, но эти реакции были исследованы в медиальной височной доле, а не в ИТС.



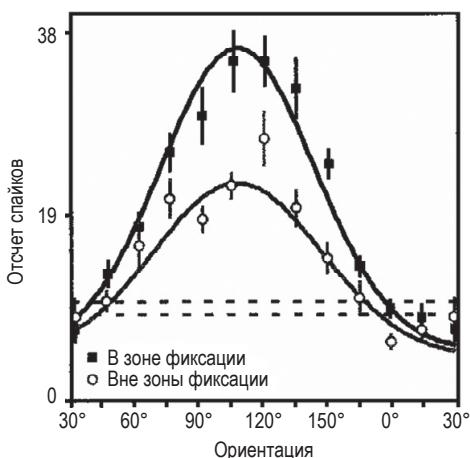
**Рис. 6.7** ♦ Избирательный ответ нейронов во время действия рабочей памяти. Ответы нейрона во время отложенной задачи сопоставления с образцом, когда сигнал был сильным стимулом (сплошной) или слабым стимулом (пунктир). Горизонтальная черная полоса обозначает продолжительность сигнала (300 мс). Воспроизведено по Челацици и др., 1998

Если пойти дальше, мы столкнемся с еще одной ситуацией, когда зрительные образы могут возникнуть в отсутствие сопутствующих внешних зрительных сигналов, – это сновидения. Люди часто сообщают о ярких зрительных образах, увиденных во сне. Остается определить, участвует ли в их создании зрительная кора. Мы вернемся к этому в разделе 10.4.

## 6.11. ЦЕЛИ, ПОСТАВЛЕННЫЕ В ЗАДАЧЕ, МОДУЛИРУЮТ НЕЙРОННЫЕ РЕАКЦИИ

Мы описали особенности откликов нейронов ИТС, как если бы они были одинаковыми и неизменными, но это далеко не так. Например, уменьшенный ответ на повторное предъявление одного и того же стимула (раздел 6.9, рис. 6.6) показывает, что нейронные реакции могут модулироваться временными особенностями задачи. Помимо ослабления ответа со временем, другие особенности текущих задач также могут модулировать ответы по всейентральной зрительной коре.

Одной из наиболее изученных форм модуляции нейронных ответов в зависимости от особенностей задачи является эффект внимания, представленный в разделе 5.17. Типичная схема изучения пространственного внимания состоит в том, чтобы научить обезьяну фиксироваться в середине экрана, уделяя скрытое внимание либо левой, либо правой половине экрана. В этих условиях обезьяны демонстрируют повышенную результативность и более быстрое время реакции во время задач визуального различения, когда стимул предъявляется в пределах локуса внимания, указанного в задаче. Когда внимание обезьяны отвлекается от места нахождения рецептивного поля, нейронный отклик меньше (рис. 6.8). Такая модуляция внимания очевидна во всем диапазоне предпочтений стимулов.



**Рис. 6.8** ♦ Пространственное внимание модулирует отклики в области V4. Модуляция кривой настройки нейрона V4 в ответ на решетки различной ориентации, когда животное обращает внимание на место рецептивного поля (закрашенные квадраты) или когда внимание отвлекается от места нахождения рецептивного поля (пустые кружки). Воспроизведено по МакАдамс и Маунселл, 1999

Другие особенности постановки задачи также могут модулировать нейронные реакции вдоль вентральной зрительной коры. Во время экспериментов по визуальному поиску субъект ищет определенный объект или определенную особенность (например, ищет Уолдо). Лаборатория Роберта Дезимоуна обучила обезьян искать красные ориентированные полосы. В этих условиях реакции нейронов на красные объекты усиливались во всем поле зрения. Другие типичные задачи включают в себя кратковременные вспышки изображений, во время которых наблюдатели должны осуществить принудительный выбор, присутствует ли на изображении конкретный целевой объект или нет. Здесь, опять же, испытания, содержащие целевой объект или категорию объекта, вызывают усиленные нейронные реакции. В разделе 3.7 мы описали две формы временной контекстной модуляции: первичную и обратную маскировки. Обе эти манипуляции также влияют на реакцию ИТС. В общем,

хотя содержимое того, что отображается на экране в определенный момент, является основными детерминантами ответов ИТС-нейронов, текущие цели, пространственный контекст, временной контекст и другие требования задач могут модулировать ответы по всей центральной зрительной коре.

## 6.12. Роль опыта в формировании настройки предпочтений нейронов

Нейронные реакции могут изменяться во время выполнения задачи в результате адаптации (раздел 6.9), влияния памяти (раздел 6.10) или постановки целей задачи (раздел 6.11). Нейронные реакции также могут изменяться в более длительных временных масштабах. Настройка предпочтений нейронов податлива и сильно зависит от «диеты» зрительных стимулов, которые демонстрируются подопытным животным. Многолетние дебаты по этой теме сосредоточены на относительной роли, которую природа и воспитание играют в формировании архитектуры зрительной коры и функций настройки нейронных реакций.

Генетика определяет базовую архитектуру зрительной системы во многом. Животные рождаются со зрительными структурами, такими как глаза, LGN и различные области коры. Хотя между видами существуют различия, шесть корковых слоев, а также их каноническая связь друг с другом, по-видимому, либо уже присутствуют при рождении, либо сформированы вскоре после него. Кроме того, существует небольшая, но четкая степень избирательности ориентации, которую можно измерить в первичной зрительной коре сразу после открытия глаз у хорьков, кошек и обезьян.

Зрелые свойства настройки нейронов являются результатом опыта. Несколько экспериментов показали, что входящие зрительные сигналы формируют настройку нейронов в зрелой первичной зрительной коре. Например, монокулярная депривация (исключение входных сигналов от одного глаза) приводит к расширению нейронных предпочтений для активного глаза в ущерб нейронам, отвечающим на входные данные от закрытого глаза. Воспитание в темноте ведет к нарушению настройки ориентации во всей первичной зрительной коре. Кроме того, эксперименты, в которых кошек выращивают в среде, где они подвергаются воздействию преимущественно вертикальных, а не горизонтальных линий, приводят к преобладанию нейронов V1, предпочитающих вертикально ориентированные полосы.

Учитывая, что даже ранние стадии корковой обработки зависят от визуального опыта, менее удивительно, что последующие стадии также можно модифицировать, изменения статистику визуальных стимулов. Как упоминалось ранее, нейроны ИТС макаки могут выборочно реагировать на образы, такие как скрепки, после того как обезьяне многократно показывают эти изображения. Понятно, что обезьяны не рождаются с нейронами, настроенными на форму скрепки. Кроме того, обезьян можно обучить распознавать символы, такие как цифры или буквы. После обучения нейроны ИТС также могут выборочно

реагировать на эти новые формы, и, опять же, такая настройка невозможна при рождении или без обучения. Предполагаемая этологическая значимость естественных стимулов, таких как лица, привела некоторых исследователей к предположению, что настройка на эти образы может быть врожденной. Однако тщательные эксперименты опровергли данную гипотезу. Если обезьяны выращиваются в среде без какого-либо контакта с лицами, исследователи не обнаруживают кластеров нейронов, настроенных на лица. Таким образом, современные данные свидетельствуют о том, что при развитии функций зрительного ответа генетика обеспечивает основную архитектуру и правила пластиичности, в то время как статистика окружающей среды направляет изменение функции настройки нейронов по всей зрительной коре.

Формирование настройки нейронов ИТС происходит не только во время развития, но и во взрослом возрасте. Эксперименты со скрепкой и числовыми символами проводились на взрослых особях обезьян, которым демонстрировали эти новые для них изображения в течение нескольких месяцев.

Настройку нейронов можно изменить гораздо быстрее. Например, скорее всего, если мы научимся распознавать символы на новом для нас языке или научимся узнавать нового человека, мы обнаружим изменения настройки нейронов в ИТС. Действительно, элегантные эксперименты на обезьянах показали, что можно изменить настройки ИТС-нейронов в течение сеанса регистрации, продолжающейся менее одного часа.

## 6.13. МОСТ МЕЖДУ ЗРЕНИЕМ И РАСПОЗНАВАНИЕМ ОБРАЗОВ

Обсуждаемые здесь исследования представляют собой неполный список примеров типов реакций, которые исследователи описывают в высших частях нижней височной коры. Несмотря на то что в данной области накоплено значительное количество таких примеров, существует острые необходимости объединить эти эмпирические наблюдения в целостную теорию зрительного распознавания, которой будут посвящены следующие главы.

Крайне важно разработать более количественные и систематические подходы для изучения предпочтений к отдельным признакам изображения в экстрастриарной зрительной коре (и других сенсорных модальностях). Методология, описанная в разделе 6.4, обеспечивает начальные шаги к беспристрастным способам исследования функций настройки нейронов в зрительной коре. В то же время мы должны стремиться описать предпочтения нейронов в количественном выражении, начиная с пикселей. На какие образы реагирует нейрон? Эта количественная теория должна позволить нам делать прогнозы и экстраполяции к новым зрительным образам. Недостаточно показать стимулы А и А”, а затем выполнить интерполяцию, чтобы предсказать ответы на А’. Если бы мы могли действительно охарактеризовать реакции нейрона, мы смогли бы предсказать ответы на любую другую

форму В. Точно так же, как неоднократно подчеркивалось, предпочтения характеристик неразрывно связаны с толерантностью к трансформациям объектов. Следовательно, мы должны уметь предсказывать реакцию нейронов на различные типы таких трансформаций. Для понимания того, как производятся вычисления и преобразования вентральной зрительной коре, требуется гораздо больше работы. Как перейти от ориентированных полосок к сложным образам, таким как лица? Большим шагом было бы взять один нейрон, скажем, в ITC и получить возможность исследовать свойства и ответы связанных с ним блоков V4, чтобы охарактеризовать преобразования сигнала от V4 к ITC.

Эта формулировка предполагает, что большая часть ответов ITC определяется их входными данными, полученными от V4. Однако мы должны помнить о сложных связях в коре и том факте, что нейроны ITC получают также множество других входных сигналов (повторяющиеся соединения, полученные по обходным путям данные из более ранних зрительных областей, обратные проекции из медиальной височной доли и префронтальной коры, а также соединения от дорсального зрительного пути). Очевидно, что существует множество неизведанных областей для отважных исследователей, которые осмелятся исследовать обширные территории экстрастриальной вентральной зрительной коры и вычисления в ней, связанные с обработкой зрительных образов. Еще одна зарождающаяся область перспективных исследований, которая все еще находится в зачаточном состоянии и потребует серьезной научной работы в ближайшем будущем, – это углубление нашего понимания того, как визуальная информация высокого уровня взаимодействует с остальной частью нашего восприятия окружающего мира.

## 6.14. ОБЗОР ГЛАВЫ

- Нижняя височная кора головного мозга (ITC) находится на вершине иерархии зрительной коры, получая сильные сигналы как от вентральных, так и от дорсальных областей коры и широко проецируясь на области, участвующие в формировании эпизодической памяти, принятии решений и когнитивном контроле.
- ITC-нейронные реакции обезьяны и человека избирательны к широкому диапазону образов, включая абстрактные узоры, бананы, стулья или лица.
- ITC-нейроны имеют обширный переполненный словарь признаков и больше озабочены формой, чем семантикой; они демонстрируют инвариантность к преобразованиям изображений.
- ITC-нейроны могут создавать законченные образы из частично видимых зрительных образов.
- Активность нейронных популяций в ITC в отдельных экспериментах можно использовать для декодирования информации об объектах с помощью линейных классификаторов.
- Нейронные отклики продолжают представлять некоторую зрительную информацию даже в отсутствие визуального стимула.

- Свойства настройки нейронов являются результатом опыта, следствием статистики визуального мира.

## ЛИТЕРАТУРА

См. <http://bit.ly/364H8WR> для дополнительной информации.

- Arcaro, M. J.; Schade, P. F.; Vincent, J. L.; Ponce, C. .R.; and Livingstone, M. S. (2017). Seeing faces is necessary for face-domain formation. *Nature Neuroscience* 20: 1404–1412.
- Freedman, D.; Riesenhuber, M.; Poggio, T.; and Miller, E. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291: 312–316.
- Hung, C. P.; Kreiman, G.; Poggio, T.; and DiCarlo, J. J. (2005). Fast read-out of object identity from macaque inferior temporal cortex. *Science* 310: 863–866.
- Liu, H.; Agam, Y.; Madsen, J. R.; and Kreiman, G. (2009). Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 62: 281–290.
- Logothetis, N. K., and Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience* 19: 577–621.

Дополнительный контент вы можете скачать по ссылке <http://bit.ly/364H8WR>.

# Глава 7

# Нейробиологически подобные вычислительные модели

Мы путешествовали по чудесной территории зрительной коры, исследуя свойства различных областей мозга и нейронных цепей, узнавая, как животные и их нейроны реагируют на зрительные стимулы и что происходит, когда различные части зрительной коры повреждаются или искусственно стимулируются. Пришло время вложить все эти биологические данные в теорию зрительного распознавания и реализовать эту теорию с помощью вычислительной модели, которая может видеть и интерпретировать мир. На пути к этой цели мы начнем с обсуждения того, как ученые описывают нейронные цепи с помощью вычислительных моделей и с определения основных свойств нейронных сетей.

## 7.1. ЗАЧЕМ НУЖНЫ ВЫЧИСЛИТЕЛЬНЫЕ МОДЕЛИ?

Я должен начать с признания, что я здесь весьма пристрастен. Построение количественных моделей совершенно *необходимо* для понимания. Я пошел бы еще дальше и заявил, что понимание *подразумевает* построение количественных, прогностических и фальсифицируемых моделей<sup>1</sup>. Для ученых сферы информатики, физиков или математиков это утверждение может показаться проповедью к новообращенным, потому что вычислительные модели обычно преподаются на курсах, а построение таких моделей – их повседневная задача. Однако слишком часто биологи или психологи смотрят на вычислительные модели с подозрением и задаются вопросом, зачем они нужны вообще. В учебных программах по биологии или психологии, как правило, отсутствуют примеры количественных моделей; вместо этого

---

<sup>1</sup> Фальсифицируемость – критерий научности, возможность экспериментального или иного опровержения теории. – Прим. ред.

концепции часто передаются через языковые структуры и графику, цель которых – описать идеи о том, как работает визуальная система.

Переход от словесных идей к формальным количественным описаниям – признак зрелости в этой области. Язык науки – математика, а не английский или эсперанто. Описания, не подкрепленные строгими математическими выкладками, часто бывают неточными, двусмысленными и подверженными ошибкам. Другая проблема вербальных моделей состоит в том, что они обычно не являются ложными, потому что определения слов недостаточно четко сформулированы, а значение слов может быть достаточно гибким, чтобы объяснить самые разные результаты. Еще более решительная версия этого утверждения была элегантно сформулирована Максом Тегмарком, известным астрофизиком Массачусетского технологического института, в его гипотезе о математической вселенной<sup>1</sup>.

В ходе формулирования гипотез, планирования экспериментов и интерпретации их результатов ученые неявно делают некие утверждения, рассматривают определенные интуитивные предположения для объяснения установленных фактов и затем пересекают через предположительно существующие логические связи. Количественные модели заставляют задуматься и формализовать эти гипотезы и предположения. Этот процесс явного изложения предположений может помочь нам лучше разрабатывать эксперименты, обнаруживать логические недостатки в наших построениях и лучше понимать результаты. Часто одни и те же или тесно связанные вопросы анализируются под разными углами, с использованием разных экспериментальных систем или одних и тех же систем в разных лабораториях. Ученые нередко применяют качественные описания наблюдений, и одни и те же слова можно интерпретировать по-разному, что приводит к бесполезным дискуссиям. Рассмотрим такие утверждения, как «мы записали высококачественную многоблочную активность», «нейрон был очень избирательным», «нейрон сильнее реагировал на лица, чем на другие стимулы» или «изображение было поразительно редким». Эти заявления полны двусмысленности.

Сравнивать результаты разных отчетов непросто. Количественные модели могут объединять наблюдения, полученные в результате разных экспериментов, измерений и методов, в разных лабораториях. Иногда на первый взгляд ничем не связанные наблюдения можно объединить вместе, используя общую теоретическую основу. Модель может указать на важные недостающие данные, важную информацию и подсказать возможные решающие эксперименты. Хорошая модель может привести к контринтуитивным экспериментальным предсказаниям. Часто экспериментаторы правильно либо ошибочно полагают, что могут сделать прогноз для следующей серии экспериментов, основываясь на своей интуиции; однако интуиция нередко дает

<sup>1</sup> Согласно этой гипотезе, наша внешняя физическая реальность является математической структурой. То есть физический мир является в определенном смысле математическим и достаточно сложен, чтобы содержать подструктуры, наделенные сознанием, которые могут субъективно воспринимать себя как существующие в физически реальном мире. – Прим. ред.

сбой (к сожалению). Яркие примеры того, как интуиция может потерпеть неудачу или наоборот, – это идея о том, что Солнце обращается вокруг Земли, корпускулярно-волновой дуализм и туннельный эффект в квантовой механике. В предыдущих главах мы обсуждали множество примеров ошибочной интуиции, в том числе идею о том, что зрение является мгновенным, что наше восприятие точно отражает то, что существует в реальном мире, и что весь мир вокруг нас имеет одинаково высокое разрешение. Сила абстракции имеет решающее значение для возможности экстраполировать и расширять границы знания за пределы ограничений, налагаемых нашими предубеждениями и интуицией.

Кроме того, количественная модель, реализованная посредством математического моделирования, может быть полезна с инженерной точки зрения (мы вернемся к этому в разделе 9.5). Например, рассмотрим задачу построения алгоритмов, которые будут принимать входные данные с цифровой камеры и распознавать объекты. Как мы скоро увидим, теоретическая модель, описывающая, как зрительная кора приматов распознает объекты, может привести к созданию вычислительных алгоритмов, с возможностью их широкого применения в реальном мире.

Иногда экспериментаторы боятся выдвигать количественные модели, считая, что построение таких моделей – прерогатива исключительно ученых, работающих в области информатики, или физиков. Я часто встречал блестящих ученых, которые, похоже на то, не хотели рисковать, вступив в чудесную страну вычислительных моделей и теоретической нейробиологии. Одной из причин этого может быть извечный страх перед математикой. В других случаях они могли полагать, что нужно быть «профессиональным теоретиком», чтобы строить количественные модели. Я категорически возражаю против этого мнения.

Некоторые из самых провокационных вычислительных моделей были созданы учеными, которые, скорее всего, не считают себя теоретиками и проводят большую часть своей жизни, совершенствуя тонкие эксперименты. Можно было бы предоставить длинный список изящных расчетов, выдвинутых экспериментаторами. Отличным примером модели, предложенной экспериментаторами, является предположение о том, как в первичной зрительной коре (V1) возникает настройка ориентации. Хьюбел и Визель, лауреаты Нобелевской премии, представленные в разделе 5.4, обнаружили, что нейроны V1 настроены на ориентацию полоски в пределах их рецептивных полей. Помимо описания эмпирических открытий, они предложили элегантную модель того, как может возникнуть такая настройка ориентации. Они рассмотрели модель с прямой связью, которая объединяет активность нескольких нейронов (блоков вычислительной модели) в LGN с круговыми рецептивными центрально-периферийными полями (рис. 5.7). Хьюбел и Визель предположили, что настройка ориентации в простых клетках в V1 возникает за счет комбинирования активности нескольких нейронов LGN с рецептивными полями, которые выровнены вдоль предпочтительной ориентации нейронов V1. С тех пор был проведен большой объем вычислительной работы для описания активности этих нейронов. Идеи Хьюбела и Визеля сыграли ключевую роль во вдохновении поколений экспериментаторов и теорети-

ков: современные вычислительные теории зорения уходят своими корнями в модели, предложенные Хьюбелом и Визелем.

## 7.2. МОДЕЛИ ОДИНОЧНЫХ НЕЙРОНОВ

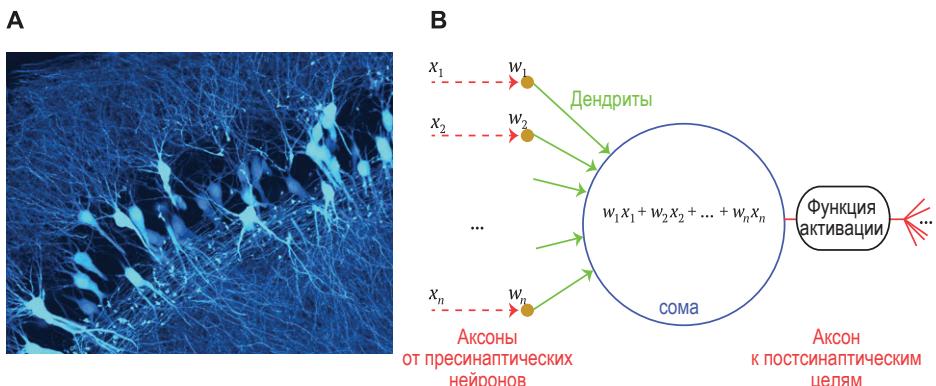
В основе вычислительных моделей функционирования мозга лежит фундаментальный «атом» вычислений – нейрон. Я оставляю слово *нейрон* для обозначения реальных биологических клеток и буду использовать слово блок (unit) для обозначения вычислительной абстракции того, что делает нейрон (хотя некоторые специалисты используют эти два термина как синонимы). Было предложено множество моделей для описания активности отдельного нейрона. Эти модели варьируются от использования операций фильтрации для описания частоты всплесков потенциала до моделирования, которое включает дендритные шипы и даже отдельные ионные каналы. Мы можем выделить несколько категорий моделей нейрона в порядке возрастания сложности: модели с фильтрами, модели «интегрировать и сработать», модели Ходжкина–Хаксли, многокомпонентные модели, модели, включающие дендритные выросты – шипы, и модели, включающие геометрию реальных схем связи нейронов.

По мере того как мы переходим от фильтрационных операций к реалистичной геометрии, биологическая точность модели значительно возрастает. Аналитические решения становятся все более сложными и часто отсутствуют, поскольку мы увеличиваем сложность модели (считается, что уравнение имеет *аналитическое решение*, если мы можем явно записать выражение в замкнутой форме, которое может быть вычислено за конечное число операций). Также наблюдается увеличение вычислительных затрат на моделирование по мере продвижения к более сложным моделям.

Более биологически точные модели не обязательно лучше, если дополнительный реализм достигается за счет слишком большой сложности, которая не имеет прямого отношения к решению поставленной задачи. Как однажды сказал знаменитый писатель Хорхе Луис Борхес: «Думать – значит забывать о различиях, обобщать, абстрагировать». Борхес проиллюстрировал этот момент восхитительным рассказом об абстракции и географических картах. Карта представляет собой простой повседневный пример того, как абстрактные модели могут быть чрезвычайно полезными. По определению, карта абстрагируется от многих деталей, чтобы раскрыть фундаментальные свойства, например как переместиться из точки А в точку В. Карта города в масштабе 1:1 (где каждый сантиметр в городе представлен сантиметром на карте) была бы намного более реалистичной и содержала бы все возможные детали. Такая карта в масштабе 1:1 занимала бы столько же места, сколько и сам город, и была бы не очень полезна для навигации. Биологические системы могут показаться устойчивыми к абстракции; эволюция заботится о приспособлении и не оптимизируется для интерпретации человеком. Случайные вариации, которые возникают в эволюционных масштабах времени, приводят к тому, что биологические системы все время нарушают «правила», и к развитию сложности, которая «просто работает».

Есть несколько вопросов, на которые нам нужно ответить, чтобы смоделировать активность нейрона. Ответы на эти вопросы зависят от того, какие конкретные особенности нейронных реакций мы хотим уловить. Рассмотрим простую аналогию из общей физики. Представьте, что мы хотим понять, как объект массы  $m$  – скажем, корова – будет ускоряться, если мы приложим силу  $F$ . Мы можем рассмотреть простую модель, которая предполагает, что объект является точечной массой, то есть что вся масса сосредоточена в точке приложения силы, – и примем модель с одним параметром  $F = m \cdot a$ . Мы прекрасно понимаем, что коровы – это не точечные массы; это предположение игнорирует геометрию коровы. Хотя это тривиальный момент, следует отметить, что эта модель с одним параметром не обеспечивает идеального описания движения коровы при наличии трения. Тем не менее эта простая модель может охватить основные составляющие задачи и даже помочь нам понять, что те же принципы, лежащие в основе движения коровы, также объясняют движение планет.

Аналогичным образом теоретики часто игнорируют геометрическую форму нейрона с его дендритами и аксонами (рис. 7.1A). Простая идеализация рассматривает блок как черный ящик, в котором принимаются и объединяются входные данные и определяется выход. Например, в модели Хьюбеля–Визеля, упомянутой ранее (раздел 7.1, рис. 5.7), можно смоделировать активность отдельных нейронов V1 как фильтр, действующий на визуальный входящий сигнал, и описать особенности ответов V1, не вдаваясь в детали дендритных вычислений, биофизики генерации потенциала действия или другие интересные свойства нейронов.



**Рис. 7.1** ♦ От реальных нейронов к вычислительным блокам. (A) Сеть нейронов гиппокампа, помеченных растворимым tdTomato<sup>1</sup> (Штрауб и Сабатини, 2016). (B) Типичный вычислительный блок (синий круг) получает входной сигнал от  $n$  пресинаптических блоков  $x_1, x_2, \dots, x_n$ . Каждый из этих сигналов умножается на синаптический вес  $w$ , который контролирует величину постсинаптического потенциала, запускаемого этим конкретным синапсом (оранжевые кружки). Дендриты (отм. зеленым) передают информацию соме (отм. синим), которая вычисляет взвешенную сумму входных данных. Нелинейная функция активации определяет выходной сигнал для заданного суммарного входного уровня. Этот вывод, в свою очередь, передается через аксоны другим блокам

<sup>1</sup> Флуоресцентный маркер. – Прим. ред.

В зависимости от вопроса в иных случаях может быть критичным рассмотреть несколько отделов нейрона, таких как сома, аксоны и дендриты. Различные вычисления могут выполняться в зависимости от расположения входов в дендрите, и, может быть, потребуется обратить внимание на точную трехмерную форму каждой отдельной ветви аксона и пространственное положение шипиков и синапсов на каждой ветви. Как сказал Эйнштейн: «Делайте вещи настолько простыми, насколько это возможно, но не проще этого».

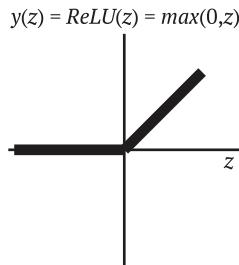
Полезная схематизация нейрона, которая широко используется в моделях нейронных сетей, проиллюстрирована на рис. 7.1В. Мы можем разделить нейрон на три основных отдела: дендриты, сома и аксон. Каждый дендрит получает входные данные от другого блока сети. Пресинаптическая активность обозначается как  $x_i$ , с  $i = 1, \dots, n$ , где  $n$  представляет собой общее количество входов. Активность каждого входного блока – это скалярная величина, которую можно грубо представить как спайковую частоту пресинаптического входа  $i$ . Воздействие данного пресинаптического входа  $i$  на интересующий нас блок зависит от весового коэффициента («веса»)  $w_i$ , который можно грубо представить как синаптическую силу между двумя блоками. В простейшем варианте каждый из этих входов считается независимым, и их вклады суммируются с соматическим потенциалом  $z$ :

$$z = w_1x_1 + w_2x_2 + \dots + w_nx_n. \quad (7.1)$$

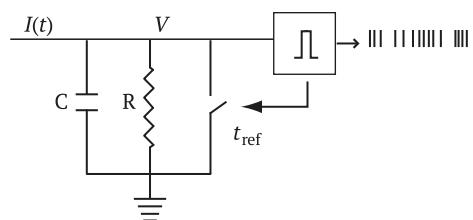
Далее сумма активности у передается через нелинейную функцию активации для формирования выходных данных. Эта нелинейность отражает представление о том, что спайковая частота не может быть меньше нуля. Она также может налагать ограничение на максимальную частоту срабатывания сигнала и может моделировать другие эффекты, такие как адаптация нейронов (уравнение 7.2 реализует только первое из этих ограничений). Особенно простой и часто используемой функцией активации является выпрямительный линейный блок (ReLU), схематически показанный на рис. 7.2:

$$y(z) = \max(0, z). \quad (7.2)$$

Затем результирующая активность у распространяется на все постсинаптические блоки. Нелинейность, подобная той, что в уравнении (7.2), играет решающую роль. Во-первых, есть целые семейства функций, которые нельзя применять без введения нелинейностей. Во-вторых, как мы скоро обсудим (раздел 7.4), мы хотим объединить множество блоков для построения нейронных сетей; выход  $y(z)$  будет давать входной сигнал для другого блока и так далее. Если в нашем распоряжении только линейные функции, то вместо того, чтобы иметь несколько уровней блоков, каждый из которых линейно суммирует предыдущие входные данные, мы могли бы также объединить все шаги в одну линейную операцию (математически, если  $y = Ax$  и  $z = By$ , тогда мы могли бы записать  $z = Cx$ ). Уравнение (7.2), несомненно, является чрезмерным упрощением, но часто оказывается полезным.



**Рис. 7.2** ♦ Выпрямительный линейный блок (ReLU). Простая нелинейность, очень популярная в моделях нейронных сетей. Активация блока представлена скалярной величиной, которую в общих чертах называют «частотой срабатывания» реального нейрона. Блок получает общий входной сигнал  $z$ , который можно представить как общий, суммированный потенциал в соме. Выходной сигнал блока выпрямляется таким образом, что отрицательные входные сигналы не приводят к активации, а положительный выходной сигнал линейно пропорционален  $z$



**Рис. 7.3** ♦ Блок «интегрировать и сработать» с утечкой. Модель такого блока представляет нейрон в виде цепи RC с конденсатором  $C$ , который объединяет входящие токи  $I(t)$  и резистор  $R$  утечки. Когда напряжение достигает определенного порога, происходит скачок потенциала (спайк), и напряжение сбрасывается. До следующего скачка длится период торможения  $t_{ref}$

Операции, показанные на рис. 7.1 и в уравнениях (7.1) и (7.2), не имеют внутренней динамики. Шагом выше по сложности является модель «интегрировать и сработать» с утечкой, которая восходит к 1907 году и, возможно, является одной из наиболее часто используемых схематизаций единичных блоков в вычислительной нейробиологии. Простейшим экземпляром модели «интегрировать и сработать» с утечкой является схема резистор–конденсатор (рис. 7.3). Ток  $I(t)$  интегрируется емкостью  $C$  и протекает через сопротивление  $R$ . Динамика внутриклеточного напряжения  $V(t)$  может быть описана следующим образом:

$$C \frac{dV(t)}{dt} = -\frac{V(t)}{R} + I(t). \quad (7.3)$$

Всякий раз, когда напряжение перескакивает пороговое значение, испускается спайк, напряжение сбрасывается, и устанавливается период абсолют-

ного торможения. Эта упрощенная версия реального нейрона отражает некоторые из наших самых главных интуитивных представлений о нейронной интеграции. Синаптические сигналы передаются от дендритов к соме, где информация интегрируется, и выходной потенциал действия генерируется, когда соматическое напряжение превышает пороговое значение. Эта модель не отражает несколько биофизических явлений, включая адаптацию частоты спайков, различные вычисления в разных отделах клетки, генерацию спайков вне сомы, субмиллисекундные события во время всплеска потенциала действия, геометрию нейронов и другие важные нюансы нейронов. Тем не менее модель «интегрировать и сработать» объясняет основные свойства того, как входы интегрируются, чтобы дать хороший сигнал на выходе.

Очень просто написать код для моделирования динамического поведения блоков «интегрировать и сработать». Ниже представлена простая (и не совсем корректная для дотошных программистов) реализация блока «интегрировать и сработать» на языке программирования под названием MATLAB.

```

1 V(1)=V_res;           % Initial resting voltage
2 for t=2:n             % For each time in the simulation from 2 to n
3     V(t)=V(t-1)+(dt/tau_m) * (E_L - V(t-1) + R_m * I_e(t));
                           % Change in voltage at time t
4     if (V(t)>V_th)    % If V(t) is above threshold V_th
5         spk(t)=1;       % Emit a spike
6         V(t)=V_res;   % And reset the voltage to a value V_res
7     end
8 end

```

Всего в нескольких программных строках можно смоделировать это простое дифференциальное уравнение первого порядка и описать всплески (spk) в ответ на произвольные входные токи (заданные  $I_e(t)$ ). В качестве примера мы можем установить  $E_L = -65$  мВ,  $V_{\text{res}} = E_L$ ,  $V_{\text{th}} = -50$  мВ,  $\tau_m = 10$  мс,  $R_m = 10$  МОм,  $n = 1000$  временных шагов и  $dt = 0,1$  мс. Мы можем играть с различными шаблонами входного сигнала (например, случайным входным сигналом типа  $I_e = 2 + 3 * \text{randn}(n, 1)$ ). Модель «интегрировать и сработать» может описывать некоторые из основных свойств мгновенного возбуждения корковых нейронов. Например, когда в пирамидальный нейрон первичной зрительной коры головного мозга кошки вводится ток, начальная частота срабатываний, вычисленная на основе первых двух спайков, может быть хорошо аппроксимирована моделью «интегрировать и сработать». Настоящие нейроны – более сложные устройства. Помимо других свойств, нейроны демонстрируют адаптацию, и возбуждение после первых двух спайков не очень хорошо описывается простой моделью «интегрировать и сработать» (но для описания адаптации могут быть внесены корректировки).

Блок «интегрировать и сработать» не фиксирует биофизические процессы в субмикросекундной динамике, описывающей форму потенциалов действия. Сделав смелый интуитивный прорыв, экспериментаторы Аллан Ходжкин (1914–1998) и Эндрю Хаксли (1917–2012) представили углубленное понимание генерации потенциалов действия. За эту работу, которая предшествовала биологической характеристике различных ионных каналов,

они получили Нобелевскую премию. *Модель Ходжкина–Хаксли* характеризует форму потенциала действия путем включения в схему главных натриевых и калиевых токов, которые ответственны за деполяризацию и реполяризацию мембранны нейрона:

$$I(t) = C \frac{dV}{dt} + \bar{g}_L(V - E_L) + \bar{g}_K n^4 (V - E_K) + \bar{g}_{Na} m^3 h (V - E_{Na}). \quad (7.4)$$

$E_L$ ,  $E_K$  и  $E_{Na}$  представляют собой обратные потенциалы утечки, калия и натрия соответственно;  $\bar{g}_L$  – проводимость утечки;  $\bar{g}_K n^4$  описывает зависящую от времени и напряжения калиевую проводимость;  $\bar{g}_{Na} m^3 h$  описывает зависящую от времени и напряжения натриевую проводимость.

Прямым продолжением было бы написать код для моделирования динамики в блоке модели Ходжкина–Хаксли. Она обеспечивает значительно более точное представление о динамике внутриклеточного напряжения по сравнению с более простыми моделями «интегрировать и сработать», а также широко используется при исследовании свойств нейронных сетей.

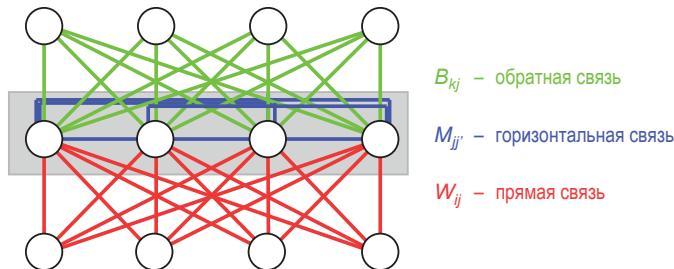
## 7.3. МОДЕЛИ СЕТЕЙ

Теперь, когда мы кратко описали семейство все более сложных моделей отдельного нейрона, мы собираемся существенно упростить каждый блок, вернувшись к представленному на рис. 7.1. Мы собираемся сместить акцент с отдельных блоков на свойства сетей взаимосвязанных блоков. Несмотря на то что каждый отдельный нейрон может выполнять интересные вычисления, визуальная избирательность, инвариантность и способность решать различные визуальные задачи возникают как следствие взаимодействий, происходящих на уровне сети. Мы будем рассматривать сети, состоящие из миллионов блоков (недавняя оценка подсчитала, что в области V1 макаки находится около 416 млн нейронов). Из-за вычислительной стоимости изучения сетей с большим количеством взаимосвязанных блоков, когда каждый из них сам выполняет свои собственные сложные операции вычисления, подавляющее большинство моделей нейронных сетей имеют дело с элементарными блоками.

Даже самые простые блоки, объединенные изящно, могут производить удивительно интересные вычисления. Коллективные вычисления относятся к возникающим функциональным свойствам группы взаимосвязанных нейронов. В конечном итоге, чтобы понять результат работы сложной системы, такой как мозг, нам нужно подумать о схемах блоков и их взаимодействиях. Интуиция часто дает сбой при рассмотрении активности сети в целом, и модели нейронных сетей могут помочь понять эти возникающие свойства сети. Чтобы изучить механику жидкости, можно абстрагироваться от детального описания столкновений и траекторий отдельных молекул и вместо этого охарактеризовать свойства жидкости, такие как температура и вязкость. Точно так же большинство моделей нейронных сетей идеализируют и упрощают составляющие ее блоки. Сети могут быть построены из простых электронных

блоков (операционные усилители заменяют нейроны; кабели, резисторы и конденсаторы заменяют аксоны, дендриты и синапсы). Динамику нейронных сетей также можно легко смоделировать на компьютере.

Типичная архитектура нейронной сети включает в себя размещение блоков слоями, которые последовательно обрабатывают информацию. Первый слой представляет собой входной слой, и мы часто думаем о конечном слое как о выходном (хотя можно также считывать информацию с любого из слоев). Трехуровневая сеть схематично проиллюстрирована на рис. 7.4. Сосредоточившись только на среднем слое (серый прямоугольник) и предположив, что нижняя часть диаграммы представляет вход, соединения, которые идут от нижнего уровня к среднему уровню, называются *восходящими* (показаны красным). Без каких-либо других подключений этот тип сети называется сетью с восходящей (*bottom-up*) или прямой связью (*feed-forward*). Простейшим примером сети прямой связи является *перцептрон* с одним входным слоем и выходом. Соединения между блоками на одном уровне называются *горизонтальными* (показаны синим цветом, иногда они также называются боковыми, или *латеральными*, соединениями). Связи от верхнего уровня к среднему известны как *нисходящие*, или *обратные*, связи (показаны зеленым). Некоторые исследователи для обозначения горизонтальных и нисходящих связей используют термин *рекуррентные связи*, но предпочтительно описывать эти горизонтальные и нисходящие связи по отдельности, поскольку они могут быть задействованы в различных вычислениях. Мощность соединений характеризуется силой или весом – здесь обозначается как  $W_{ij}$  для восходящих соединений,  $M_{jj'}$  для горизонтальных соединений и  $B_{kj}$  для нисходящих соединений. На схеме на рис. 7.4 блоки подключены по принципу «все ко всем»; то есть каждый блок нижнего уровня проецируется на каждый блок среднего уровня, и то же самое справедливо для всех других типов соединений. Связь не обязательно должна быть универсальной; некоторые из уровней силы соединения могут быть установлены на 0, чтобы указать на отсутствие соединения. Кроме того, в схеме на рис. 7.4 на каждом уровне четыре блока, и поэтому все индексы  $i$ ,  $j$  и  $k$  изменяются от 1 до 4, но это не обязательно; на каждом слое может быть разное количество блоков. На схеме основное внимание уделяется подключению к среднему уровню, но в общем случае также будут существовать дополнительные восходящие соединения от среднего уровня к верхнему уровню и нисходящие соединения от среднего уровня к нижнему уровню. Как правило, в нижнем слое не должно быть горизонтальных соединений; мы часто думаем о нижнем слое как о входном изображении. Точно так же, как правило, в верхнем слое не должно быть никаких горизонтальных соединений; мы часто думаем о верхнем слое как о выходе, что, возможно, указывает на присутствие различных классов объектов в изображении.

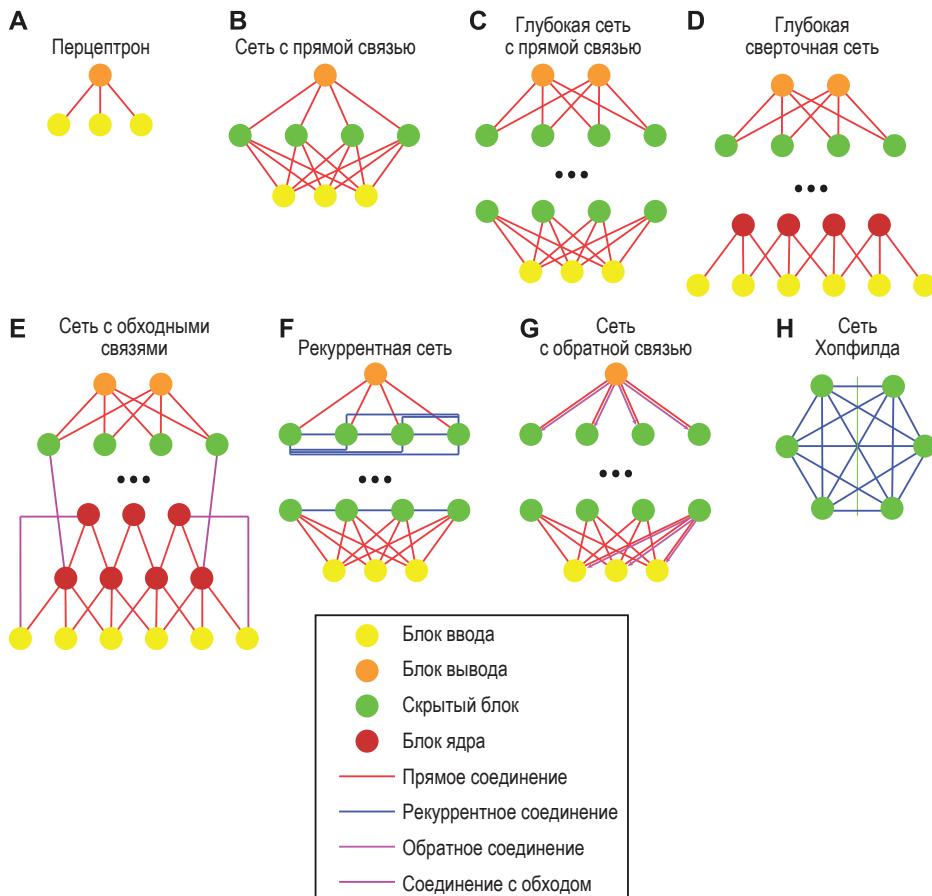


**Рис. 7.4** ◆ Прямые, горизонтальные и обратные связи в нейронных сетях. Модели нейронных сетей состоят из нескольких взаимосвязанных нейроноподобных блоков (здесь кружки), каждый из которых соответствует типам вычислений, показанным на рис. 7.1. Типичная архитектура нейронной сети состоит в том, чтобы упорядочить блоки по слоям. На этой диаграмме показаны три слоя. Предполагая, что вход находится внизу диаграммы, а выход находится вверху, мы можем различить прямые соединения (красный), горизонтальные соединения (синий) и обратные соединения (зеленый)

Блоки в нейронных сетях могут быть возбуждающими (положительные веса) или тормозящими (отрицательные веса). Один и тот же блок может возбуждать одни постсинаптические цели и тормозить (ингибиовать) другие. За исключением нескольких контрпримеров, это не тот случай в биологии, где отдельный нейрон обеспечивает либо возбуждающие, либо тормозящие сигналы, но не то и другое одновременно.

Рисунок 7.4 не является исчерпывающим описанием всех возможных способов соединения блоков в нейронной сети. В наиболее типичных сценариях блоки соединяются внутри слоя (горизонтальные соединения) или между соседними слоями (снизу вверх и сверху вниз). Однако также можно создать «обходные» соединения, которые пропускают определенный уровень – например, от нижнего уровня непосредственно к верхнему. На рис. 7.5 схематически показаны некоторые важные типы архитектуры нейронных сетей, которые были исследованы в научных работах. На этом рисунке показаны архитектуры, которые использовались для моделирования визуальных вычислений, но это представление не является исчерпывающим. На всех этих диаграммах в иллюстративных целях показаны только несколько блоков, но их может быть намного больше.

На рис. 7.5А–Е блоки организованы в каскад слоев, передающих информацию снизу вверх, аналогично примеру на рис. 7.4, с учетом только красных линий. Иерархическая организация в этих сетях примерно напоминает иерархическую организацию вычислений в зрительной коре (рис. 1.5), хотя даже один взгляд на рис. 1.5 показывает, что современные архитектуры охватывают лишь небольшую часть сложности реальной, биологической зрительной системы.



**Рис. 7.5** ♦ Семейство моделей нейронных сетей. Этот рисунок представляет некоторые важные типы моделей нейронных сетей. На каждой диаграмме нижний уровень обеспечивает ввод сигнала (желтый), а верхний уровень обеспечивает вывод (оранжевый). В иллюстративных целях показано только несколько блоков или слоев. Многоточия указывают на возможное присутствие некоторого количества промежуточных слоев. В большинстве случаев информация течет снизу вверх через прямые соединения (красный). В **F** есть дополнительные рекуррентные соединения (синий), а в **G** есть обратные связи, которые предоставляют информацию от верхнего уровня к нижнему уровню. Сеть в **H** – это архитектура другого типа, в которой все блоки находятся на одном уровне и взаимно связаны между собой по принципу «все ко всем»

## 7.4. СЕТЕВЫЕ МОДЕЛИ С ЧАСТОТОЙ СРАБАТЫВАНИЯ НЕЙРОНОВ

Сетевые модели с частотой срабатывания нейронов представляют собой простой, но поучительный класс нейронных схем. В простейшем случае рассмотрим схему прямой связи с  $N$  блоками, проецируемыми на данный выходной блок. Вектор  $x$  представляет входную активность. Мы можем думать о компонентах  $x$  как о частоте срабатывания (спайковой частоте) каждого входного блока. Скалярное значение  $y$  обозначает выходную частоту импульсов. Синаптическое ядро  $K_s$  описывает, как входная спайковая частота линейно преобразуется во входной ток для выходного блока. Теоретики часто представляют силу данного синапса  $i$  ( $i = 1, \dots, N$ ) скалярным значением  $w_i$ . Это значение может представлять собой комбинацию вероятности синаптического выброса пресинаптического нейрона и амплитуды постсинаптического потенциала (положительного или отрицательного), вызываемого входящими нейротрансмиттерами. Общий входной сигнал в выходной блок  $I_s$  определяется выражением

$$I_s = \sum_{i=1}^N w_i \int_{-\infty}^t d\tau K_s(t - \tau) x_i(\tau), \quad (7.5)$$

где  $w_i$  представляет собой вес или силу каждого синапса. Используя экспоненциальное ядро, динамику этой схемы можно описать следующим образом:

$$\tau_s \frac{dI_s}{dt} = -I_s + \sum_{i=1}^N w_i x_i. \quad (7.6)$$

Спайковая частота выходного блока обычно представляет собой нелинейную функцию общего входного тока:  $y = F(I_s)$ .  $F$  может быть сигмоидной функцией или функцией порогового выпрямления.

## 7.5. ОПЕРАЦИЯ СВЕРТКИ

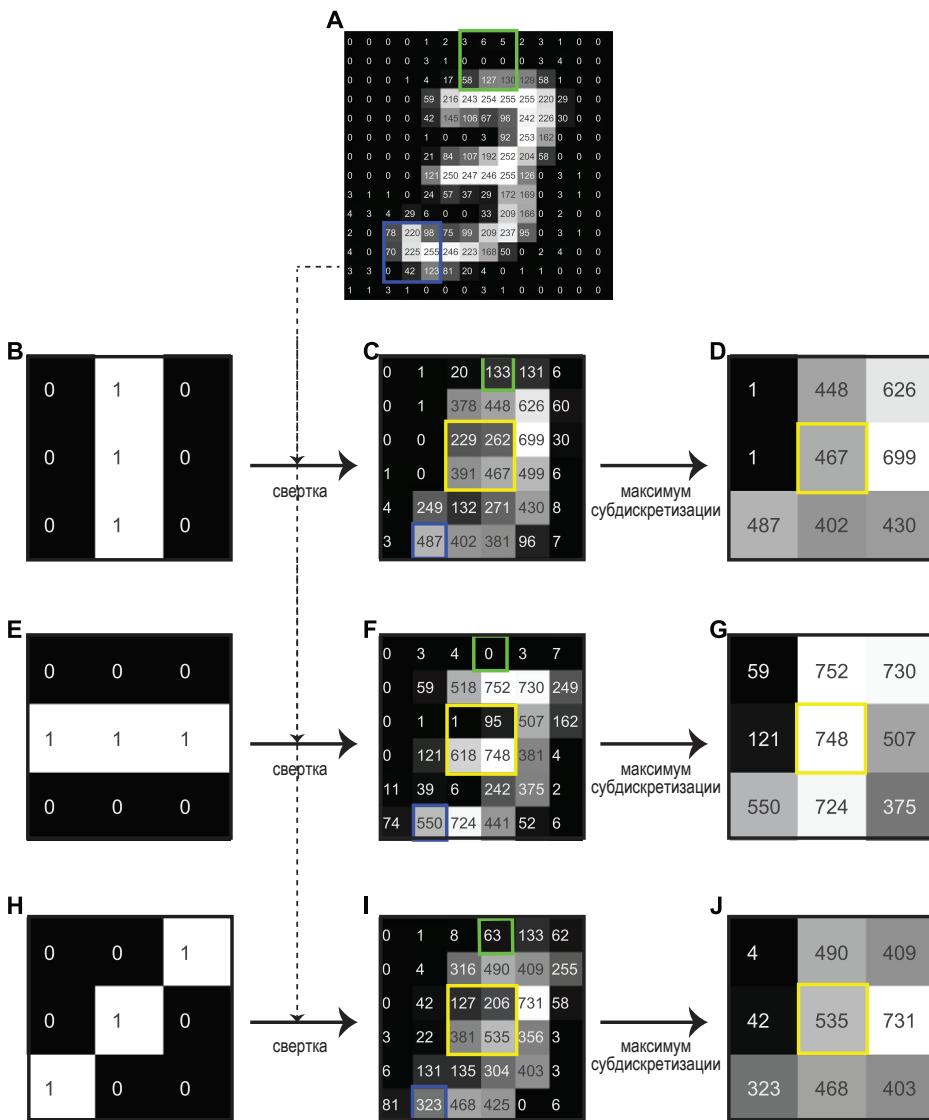
Одним из ключевых вычислительных компонентов обработки зрительного сигнала является то, что одна и та же операция обычно повторяется во всем визуальном поле. Например, мы находим нейроны в первичной зрительной коре, которые демонстрируют настройку ориентации с рецептивными полями, покрывающими все зрительное поле. Таким образом, вычислительная операция, которая фильтрует изображение для извлечения информации об ориентации, должна повторяться снова и снова по всему изображению. Этот тип операции легко реализуется посредством *операции свертки*.

Для двух функций  $f(t)$  и  $g(t)$  операция свертки (на английском – convolution; на латыни convolvere означает «сворачиваться вместе»), обозначенная символом  $*$  в следующем уравнении, определяется как интеграл от сигнала, отраженного, сдвинутого и умноженного на другой сигнал:

$$f(t) * g(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau. \quad (7.7)$$

При обработке изображения процесс свертки относится к применению данного фильтра по всему изображению (или всему предыдущему слою) и возврату результата в каждой позиции. Пример этого процесса показан на рис. 7.6. Для простоты здесь входное изображение в градациях серого представляет собой рукописную версию числа 3, уменьшенную до разрешения  $14 \times 14$  пикселей. Каждый пиксель имеет интенсивность (яркость) от 0 (черный) до 255 (белый). В общем, значение активации каждого блока не обязательно должно быть целым числом, и входное изображение может иметь три цвета, а не всего лишь один. Мы рассматриваем три возможных функциональных фильтра, показанных на рис. 7.6 **В** (вертикальный фильтр), **Е** (горизонтальный фильтр) и **Н** (диагональный фильтр). Для упрощения здесь веса фильтра  $3 \times 3$  пикселя состоят из нулей и единиц, но, опять же, в общем эти фильтры содержат действительные значения. Фильтр применяется к каждой позиции, и его значения умножаются на соответствующие значения изображения. Например, рассмотрим вертикальный фильтр и зеленый квадрат в верхней части изображения, содержащие значения 3, 6, 5 в первой строке, 0, 0, 0 во второй строке и 58, 127, 130 в третьей строке (рис. 7.6A). Мы получаем  $0 \times 3 + 1 \times 6 + 0 \times 5 = 6$  в первом ряду (числа, выделенные жирным шрифтом, взяты из фильтра),  $0 \times 0 + 1 \times 0 + 0 \times 0 = 0$  во втором ряду и  $0 \times 58 + 1 \times 127 + 0 \times 130 = 127$  в третьем ряду. Сложение этих трех чисел дает значение 133 в соответствующем зеленом квадрате на рис. 7.6C. Тот же процесс повторяется на всем изображении, в результате получается матрица, представленная на рис. 7.6C. Поскольку фильтр напоминает вертикальную линию, после соответствующей нормализации операция выделяет области входного изображения, содержащие пиксели, которые выглядят как короткие вертикальные линии. Точно так же фильтр на рис. 7.6E выделяет горизонтальные края, а фильтр на рис. 7.6H выделяет диагональные края.

Мы можем рассматривать эти фильтры как грубое приближение к простым нейронам в области V1, реагирующими на ориентированные линии (раздел 5.4). Следующим шагом в области V1 является суммирование сигналов от нескольких простых нейронов для создания сложного нейрона с аналогичной настройкой, но реагирующего более или менее независимо от положения предпочтительного признака в воспринимающем поле (раздел 5.5). Вдохновленные идеей простых и сложных нейронов, после свертки мы реализуем операцию субдискретизации, или объединения (pooling), которая объединяет несколько значений в окне. Эта операция субдискретизации увеличивает размер воспринимающего поля. Типичная операция субдискретизации состоит в том, чтобы взять максимум всех входных значений. Например, рассмотрим желтый квадрат в центре рис. 7.6C, состоящий из матрицы размерности  $2 \times 2$  со значениями 229, 262 в первой строке и 391, 467 во второй строке.



**Рис. 7.6** ♦ Основные операции в нейронных сетях. Изображение в градациях серого (изображение размером  $14 \times 14$  пикселей, представляющее рукописную цифру 3) подвергается операции свертки с помощью трех различных фильтров (**B, E, H**). В этом случае каждый из фильтров имеет размер  $3 \times 3$  пикселя, и здесь, для простоты, значения их интенсивности равны только 0 и 1. Операция свертки тут имеет «шаг», равный 2, что означает, что фильтр пропускает один пиксель при перемещении по изображению. Зеленый и синий квадраты на изображении (**A**) дают результаты, выделенные соответственно зеленым и синим квадратами после свертки в (**C, F, I**). Операция субдискретизации берет выходные данные свертки и извлекает максимум в блоках размером  $2 \times 2$ , также с шагом 2. Отмеченная желтым область после свертки соответствует желтой области в окончательном выводе в **D, G, J**.

Эти четыре числа объединяются с помощью операции `max`, чтобы получить 467 в соответствующем желтом квадрате на рис. 7.6D. Операция объединения дает неизменность позиции, обеспечивая высокую активность в любом из четырех местоположений.

Операции свертки и субдискретизации позволяют разработать систему этапов иерархического извлечения признаков объектов. В примере, показанном на рис. 7.6, все операции фиксированы. В общем, мы будем заинтересованы в разработке подходящих фильтров для решения конкретной задачи или, что еще лучше, в автоматическом обучении этих фильтров. После обучения решению визуальных задач последовательные слои свертки и субдискретизации в сети учатся извлекать из изображения все более сложные элементы, от краев до сложных форм и объектов. Мы вернемся к вопросу о том, как обучить нейронные сети определению весов, в разделе 8.6.

## 7.6. Сети Хопфилда

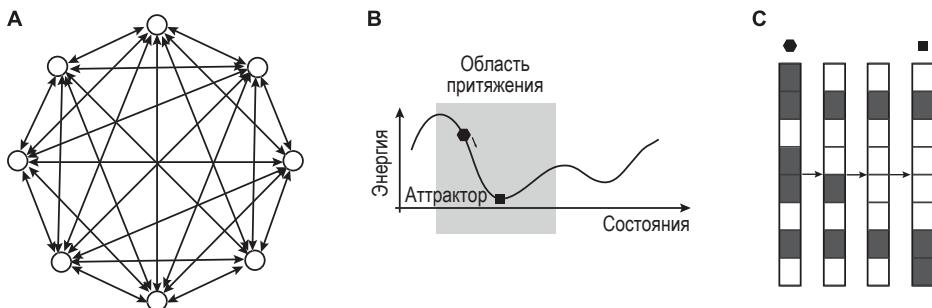
Динамика сетей прямого распространения довольно проста, информация переходит от одного уровня к другому. Более сложная динамика может быть получена в сетях с рекуррентным подключением. Простой, но богатый пример – это случай *рекуррентных сетей Хопфилда* (рис. 7.7). Что особенно привлекательно в этих сетях, так это то, что здесь есть новые свойства цепи, которые нелегко идентифицировать или описывать, рассматривая только отдельные блоки, не обращая внимания на их взаимодействия. Сеть Хопфилда может решать довольно сложные вычислительные задачи и обладает интересными свойствами, такими как устойчивость к возмущениям и способность завершать формы.

Самая базовая версия сети Хопфилда определяется одним слоем с бинарными блоками, которые связаны по принципу «все со всеми», с симметричными весами. На рис. 7.7A показан пример сети Хопфилда с восемью блоками. Пусть состояние блока  $i$  в момент времени  $t$  представлено как  $s_i(t)$ ; это состояние может принимать значения 0 или 1 для двоичной сети. Тогда состояние сети будет представлено вектором  $s(t) = [s_1(t), \dots, s_N(t)]$ , где  $N$  – общее количество блоков (рис. 7.7C). Самосоединений нет ( $w_{ii} = 0$ ), а блоки связаны, «все со всеми», – симметричным образом ( $w_{ij} = w_{ji}$ ). Следуя уравнениям (7.1) и (7.2), состояние каждого блока обновляется в соответствии с пороговой и взвешенной суммой входных данных от всех других блоков:

$$s_i(t+1) = \text{sign}\left(\sum_{j \neq i}^N w_{ij} s_j(t) - \theta\right), \quad (7.8)$$

где  $\theta$  – порог. В этом типе рекуррентной архитектуры интересно то, что можно определить функцию энергии (рис. 7.7B), задаваемую формулой

$$E(t) = -\frac{1}{2} \sum_{i,j} w_{ij} s_i(t) s_j(t) + \sum_i s_i(t) \theta_i. \quad (7.9)$$



**Рис. 7.7** ♦ Рекуррентные нейронные сети на основе атTRACTоров. (A) Схема сети Хопфилда из восьми блоков с подключением «все ко всем» и симметричной матрицей связности ( $w_{ij} = w_{ji}$ ). (B) Состояние сети характеризуется энергетической функцией с состояниями атTRACTора, определяемыми весовой матрицей. Начиная с энергетического состояния в «области притяжения» (серый прямоугольник), точка сети, представленная шестиугольником, идет вниз по энергетическому пространству к состоянию атTRACTора, представленному квадратом. (C) Пример эволюции состояния сети от начального состояния (шестиугольник) к атTRACTору (квадрату). Здесь каждый квадрат представляет активность блока (серый – включен, белый – выключен). Показаны три этапа обновления (стрелки). Сеть может выполнять завершение формы, потому что, когда она инициируется в состоянии, которое близко, но не идентично памяти (атTRACTору), ее динамические свойства перемещают ее состояние к атTRACTору

Можно показать, что эта функция энергии ограничена снизу и монотонно убывает в соответствии с динамикой, определяемой уравнением (7.8). Другими словами, сеть имеет состояния атTRACTора, к которым она будет сходиться при запуске в произвольных состояниях. Если сеть начинается с состояния, представленного шестиугольником на рис. 7.7В (состояние слева на рис. 7.7С), она будет динамически развиваться, всегда уменьшая свою энергию, пока не достигнет состояния атTRACTора, представленного квадратом на рис. 7.7В (состояние справа на рис. 7.7С).

Теперь предположим, что мы хотим сохранить серию графических форм  $\mu = 1, \dots, m$ , определенных состоянием каждого из блоков:  $\epsilon_1^\mu, \dots, \epsilon_N^\mu$ . Мы можем использовать правило обучения Хебба, чтобы вычислить веса блоков сети Хопфилда:

$$w_{ij} = \frac{1}{m} \sum_{\mu=1}^m \epsilon_i^\mu \epsilon_j^\mu. \quad (7.10)$$

Эти паттерны определяют состояния атTRACTора для сети. Если мы инициализируем сеть в каком-нибудь произвольном состоянии, до тех пор, пока это состояние находится в пределах области притяжения данного атTRACTора, состояние сети будет развиваться в сторону соответствующего атTRACTора (рис. 7.7В и С).

С точки зрения реализации, рекуррентную сеть с дискретными времененными шагами можно «развернуть», чтобы преобразовать ее в сеть с прямой

связью с общими весами. Например, три временных шага рекуррентной сети с восемью блоками могут быть реализованы как четырехуровневая сеть с прямой связью с восемью блоками на каждом уровне, со всеми соединениями, и где все веса от одного уровня к другому равны по слоям. Для рекуррентной сети Хопфилда отсутствие самоподключений подразумевает установку веса от блока  $i$  в данном слое к блоку  $i$  на следующем слое равным 0, а симметричная матрица связности подразумевает установку веса от блока  $i$  в данном слое к блоку  $j$  в следующем слое равным весу от блока  $j$  до блока  $i$  в следующем слое.

Несмотря на эту эквивалентность между рекуррентными и прямыми сетями, рекуррентное соединение дает ряд преимуществ. Во-первых, для рекуррентной сети требуется меньшее количество блоков (если  $T$  – количество повторяющихся шагов, количество блоков в сети с прямой связью равно  $T + 1$ , умноженному на количество блоков в рекуррентной сети). В биологии размер мозга имеет большое значение из-за ограничений по весу и особенно из-за энергетических ограничений. Мозг особенно затратен с точки зрения энергии. Соображения о размере и потреблении энергии также могут иметь значение для определенных вычислительных приложений, таких как реализация алгоритмов компьютерного зрения в смартфоне. Во-вторых, рекуррентная сеть тоже требует меньших весов (снова в  $T + 1$  раз). Количество весов также важно с точки зрения ограничений по размеру в биологии.

Кроме того, важнейшим преимуществом рекуррентных сетей является их вычислительная гибкость. В рекуррентной сети архитектура не требует заранее установленного количества шагов  $T$ . Некоторые задачи могут быть сложнее и требуют калькуляций с большим количеством шагов, тогда как другие задачи могут быть проще и требуют меньшего количества шагов. Напротив, сеть с прямой связью предлагает жесткую структуру, в которой вычисления всегда должны проходить через все уровни  $T + 1$ . Чтобы добавить сети гибкости и обойти эту проблему, некоторые сети прямого распространения включают в себя обходные соединения, при которых обработка информации может пропускать определенные уровни (рис. 7.5Е). Достижение полной гибкости сети Хопфилда через обходные соединения требует подключения каждого уровня ко всем другим уровням, что приведет к огромному увеличению количества весов. Большинство моделей глубоких нейронных сетей включают лишь небольшое подмножество всех возможных обходных соединений.

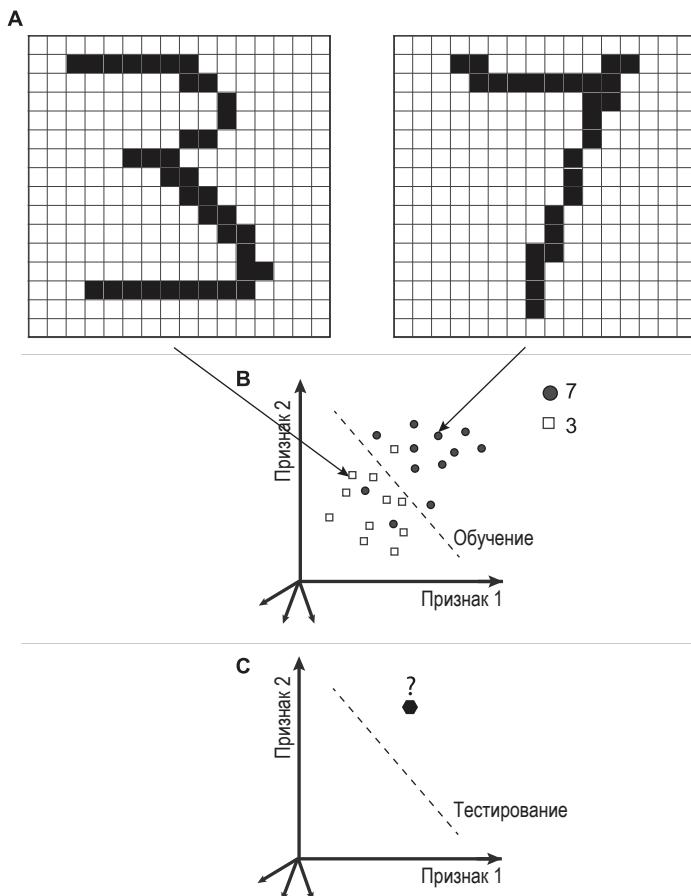
Одно из критических замечаний по поводу сетей Хопфилда состоит в том, что нет доказательств связи «все со всеми» в биологических нейронных цепях. Однако есть обширные свидетельства частичных горизонтальных связей между нейронами в пределах данного слоя коры головного мозга, и эти связи могут принести множественные преимущества, описанные здесь: эффективность в смысле занимаемого пространства и потребления энергии, гибкие вычисления и возможность завершения визуальных форм. Другое соображение состоит в том, что взаимные связи, когда блок  $i$  соединяется с блоком  $j$ , а блок  $j$  соединяется с блоком  $i$ , являются скорее исключением, чем правилом в биологии, особенно если силы этих соединений должны быть симметричными.

## 7.7. НЕЙРОННЫЕ СЕТИ МОГУТ РЕШАТЬ ЗАДАЧИ РАСПОЗНАВАНИЯ

Как нейронные сети могут решать задачи распознавания? Рассмотрим простую задачу зрительного распознавания. Представьте, что у нас есть набор изображений, состоящий из рукописных версий числа 3 и рукописных версий числа 7 (рис. 7.8А). Люди могут взглянуть на каждую картинку и быстро сказать, что слева – 3, а справа – 7. Теперь рассмотрим нейронную сеть. Точная архитектура нейронной сети на данный момент не актуальна; пока мы можем думать о любой из сетевых архитектур, показанных на рис. 7.5, а подробнее поговорим о различных архитектурах в разделах 8.3–5. На входе в сеть – интенсивность каждого пикселя изображения. Размер примеров на рис. 7.8А составляет  $16 \times 16$  пикселей, поэтому у нас будет 256 входных блоков (массив из 256 чисел, объединяющий все строки в матрице изображения). Активация каждого входа – это значение интенсивности от 0 (черный) до 255 (белый). Каждое изображение будет иметь различную комбинацию этих 256 значений. Как мы обсуждали в разделе 2.11, мы можем рассматривать эти числа как грубую визуализацию частоты срабатывания ганглиозных клеток сетчатки в ответ на изображение.

Мы можем попытаться классифицировать изображения на основе этих 256 значений напрямую. В качестве альтернативы можно построить нейронную сеть, подобную той, что показана на рис. 7.4, с 256 входными блоками (вместо четырех входов, показанных на этом рисунке). Вооружившись множеством образцов изображений «3» и «7», нейронную сеть можно обучить регулировать силу соединения, чтобы найти подходящие признаки, которые могут облегчить различение двух групп изображений. В разделе 8.6 мы обсудим, как можно отрегулировать силу этих соединений. А пока предположим, что мы уже обучили сеть. После обучения нейронная сеть извлекает набор признаков из каждого изображения. Мы можем представить эти особенности и построить все изображения в графике многомерной функции, подобном тому, который показан на рис. 7.8В, где каждая точка соответствует разному изображению. Количество измерений соответствует количеству признаков, то есть количеству блоков в нейронной сети до слоя классификации. Слой классификации вывода должен иметь столько блоков, сколько существует классов для задачи различения – в данном случае два блока вывода: один указывает на наличие «3», а другой – на наличие «7».

Мы можем представить себе каскад вычислений, которые происходят при перемещении от одного слоя к другому, как набор вычислений, которые производятся при перемещении от сетчатки до зрительной коры (гл. 5 и 6). Используя это отображение, мы можем рассматривать активацию каждого из выходных блоков как грубую визуализацию частоты срабатывания нейронов в зрительной коре головного мозга. Точная область зрительной коры не имеет отношения к текущему обсуждению здесь; мы вернемся к сравнению нейронных сетей и откликов нейронов в разных частях зрительной коры в разделе 8.14.



**Рис. 7.8** ♦ Схематический пример задачи распознавания, решаемой нейронной сетью. (А) Рассмотрим набор из множества изображений, представляющих рукописные цифры 3 и 7, только два из которых показаны здесь. (В) Интенсивность пикселей может быть передана в нейронную сеть, которая извлечет набор признаков. Затем каждое изображение будет представлено точкой в многомерном пространстве, состоящем из нескольких признаков. Здесь все цифры 3 представлены белыми квадратами, а все 7 – серыми кружками. Набор данных используется для обучения классификатора (схематично представлен здесь пунктирной линией), для разделения двух типов изображений. (С) Учитывая новое изображение, которое не использовалось во время обучения, классификатор пометит его как 3 или 7, в зависимости от того, с какой стороны линии на графике оно оказывается

В разделе 6.7 мы проиллюстрировали, как классификатор может научиться различать разные типы изображений на основе спайковой частоты популяции нейронов (рис. 6.4). Теперь мы можем использовать ту же процедуру для классификации изображений на основе признаков, извлеченных нейронной сетью. Пунктирная линия на рис. 7.8В представляет собой классификатор: изображение классифицируется как цифра 3, если точка опускается ниже

пунктирной линии на этом графике, и как цифра 7 в противном случае. Конечно, пунктирная линия соответствует гиперплоскости в многомерном пространстве, потому что, как правило, выбор делается из более чем двух объектов. Кроме того, процедура может быть легко расширена до задачи классификации с несколькими классами, а не только с двумя (например, классификация всех десяти рукописных цифр, раздел 8.10).

Если теперь нам будет представлено новое изображение, то есть изображение, которое не использовалось для обучения нейронной сети, мы можем снова вычислить значения активации и нанести точку, соответствующую новому тестовому изображению, на тот же график (рис. 7.8С). Таким образом, классификатор может присвоить новому изображению название «3» или «7». Расширяя сопоставление между блоками в сети и нейронами в головном мозге, мы можем заявить, что активность популяции нейронов в зрительной коре считывается нейронами в другой области мозга, которая в конечном итоге отвечает за нашу способность говорить «это изображение является тройкой» и «это изображение – семеркой». В гл. 8 мы более подробно рассмотрим архитектуру нейронных сетей (раздел 8.5), насколько хорошо их работа соответствует каскаду вычислений в центральной зрительной коре (раздел 8.14) и насколько хорошо они могут объяснить зрительно-моторную координацию (разделы 8.12 и 8.13).

## 7.8. ЭКСТРЕМАЛЬНЫЙ БИОЛОГИЧЕСКИЙ РЕАЛИЗМ: ПРОЕКТ «BLUE BRAIN»

Прежде чем закончить эту главу, мы вернемся к понятию вычислительных моделей и абстракции. Многие биологи твердо убеждены в том, что упрощенные сети, подобные описанным здесь, не способны уловить сложность и богатство нейробиологических схем. Это наблюдение, конечно, совершенено верно.

На другом конце спектра архитектур сетей можно встретить такие проекты, как проект «Blue Brain» («Синий мозг»)<sup>1</sup>. Этот проект направлен на привнесение значительного количества биологического реализма с использованием сложных сетевых симуляций. Амбициозная цель – в один прекрасный день создать *in silico* точную копию мозга грызуна, а может быть, и человеческого мозга. В отличие от абстракций, используемых в нейронных сетях, проект направлен на создание биофизически реалистичных симуляций отдельных нейронов и включение данных о формах нейронов, а также взаимосвязях между ними.

Современные нейронные сети, даже в упрощенном и абстрактном формате, показанном на рис. 7.4, имеют огромное количество настраиваемых

<sup>1</sup> Проект назван так, поскольку осуществлен на суперкомпьютере IBM Blue Gene. – Прим. ред.

параметров (раздел 8.9). Построение биологически подробных моделей нейронных схем добавляет много порядков сложности с точки зрения количества параметров, поддающихся настройке. Например, должны ли модели учитывать детальную геометрию каждого дендрита, расстояние между нейронами, количество миелина, окружающего каждый аксон, различные биофизические свойства мириад различных типов интернейронов? Список биологических свойств нейронов можно продолжать очень долго. Для многих из этих дополнительных параметров у нас все еще нет достаточных данных, чтобы ограничить степень реалистичности моделей. Даже если бы у нас было достаточно экспериментальных данных, чтобы описать огромное пространство параметров, совсем не очевидно, что мы хотели бы включить все мелочи биологического механизма. Предыдущее краткое обсуждение адекватного уровня абстракции и реализма при моделировании одиночных нейронов в равной степени применимо здесь, в контексте сетевых моделей.

## 7.9. Обзор главы

- Чтобы понять, как работает зрение, важно построить количественные вычислительные модели.
- Мы используем модели, в разной степени абстрагированные; биологическая основа зрения в них упрощена, чтобы выделить из нее основные принципы вычислительной работы нейронов.
- Нейрон «интегрировать и сработать» состоит из интегратора с утечкой и определяет важные признаки того, как входной сигнал, полученный нейроном, преобразуется в выходную активность.
- Операция свертки позволяет извлекать одни и те же визуальные признаки по всему визуальному полю.
- Основные элементарные вычисления включают фильтрацию, нормализацию, объединение и нелинейности.
- Объединение нескольких блоков приводит к архитектурам нейронных сетей с новыми вычислительными свойствами, которые в конечном итоге сводятся к комбинации простых элементарных шагов.
- Нейронные сети обычно включают в себя соединения следующих типов: соединения с прямой связью, горизонтальные соединения и нисходящие соединения.
- Комбинируя эти разные типы соединений, можно создавать самые разные архитектуры нейронных сетей.
- Рекуррентные нейронные сети на основе атTRACTоров, такие как сеть Хопфилда, могут демонстрировать интересные динамические свойства, которые позволяют экономить энергию, обеспечивают гибкую вычислительную мощность и демонстрируют устойчивость к возмущениям.
- Нейронные сети могут решать визуальные задачи.

## ЛИТЕРАТУРА

См. дополнительные ссылки на <http://bit.ly/2HpAqRm>.

- Dayan, P., and Abbott, L. (2001). *Theoretical neuroscience*. Cambridge: MIT Press.
- Gabbiani, F., and Cox, S. (2010). *Mathematics for neuroscientists*. London: Academic Press.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *PNAS* 79: 2554–2558.
- Koch, C. (1999). *Biophysics of computation*. New York: Oxford University Press.
- Markram, H. (2006). The blue brain project. *Nat Rev Neurosci* 7: 153–160.

Дополнительный контент вы можете скачать по ссылке <http://bit.ly/2HpAqRm>.

# Глава 8

## Учим компьютеры видеть

Мы прошли долгий путь с момента наших первых шагов по определению основных свойств зрения в гл. 1. Мы начали с характеристики пространственной и временной статистики естественных изображений (гл. 2). Мы описали зрительно-моторную координацию – как наблюдатели воспринимают и реагируют на изображения вокруг себя (гл. 3). Исследования повреждений мозга помогли определить определенные области коры головного мозга, отвечающие за обработку различных типов зрительной информации (гл. 4). Мы исследовали, как нейроны сетчатки, таламуса иentralной зрительной коры головного мозга реагируют на множество различных стимулов (гл. 2, 5 и 6).

В этой главе мы объединим все эти фрагменты феноменологических наблюдений в согласованную теоретическую основу, чтобы понять, как нейронные цепи организуют обработку зрительной информации. Мы представим вычислительные модели, которые воплощают эту теоретическую основу, давая машинам возможность начать видеть и интерпретировать визуальный мир вокруг нас.

### 8.1. КРАТКОЕ ОПИСАНИЕ И ОПРЕДЕЛЕНИЯ

Мы начнем с обобщения важнейших наблюдений из предыдущих глав, чтобы определить ограничения для решения проблемы зрения. Теория зрения, реализованная с помощью вычислительной модели, должна удовлетворять следующим восьми требованиям.

1. *Избирательность.* Визуальная система демонстрирует замечательную степень избирательности, или способность различать образы, которые кажутся похожими на уровне пикселей (например, произвольные трехмерные формы, созданные из скрепок, символов, букв и разных лиц). Модель должна уметь различать изображения, похожие в пиксельном пространстве, но представляющие разные объекты (рис. 1.4).
2. *Толерантность к трансформациям.* Тривиальным решением для достижения высокой селективности было бы запоминание всех пикселей изображения, т. е. алгоритм прямого сопоставления форм объектов (рис. 1.4). Этот тип алгоритма не допускает никаких изменений изобра-

жения. Однако объект может давать бесконечное количество вариантов проекций на сетчатку (рис. 1.3). Эти преобразования изображения возникают из-за изменений положения объекта относительно точки фиксации внимания, видимого размера объекта, поворота или перемещения по глубине, изменения контраста, освещения, цвета или его окклюзии (рис. 3.6). Важность адекватного сочетания избирательности и толерантности составляет критическую особенность систем машинного зрения и серьезную задачу для вычислительных моделей.

3. *Скорость.* Зрение – очень быстрый процесс, что подчеркивается многими психофизическими экспериментами, а также нейрофизиологическими записями регистраций сигналов активности мозга у людей и обезьян (раздел 3.6). Примерно за 150 мс мы можем определить присутствие объектов на изображении и получить хорошее первое впечатление о том, что происходит. Эта скорость накладывает ограничение на количество вычислительных шагов, которые визуальная система может использовать для решения задач зрительного распознавания.
4. *Универсальность.* Мы можем распознать самые разные предметы. Оценки точного числа категорий объектов, которые могут распознать приматы, сильно различаются в зависимости от нескольких предположений и экстраполяций. Некоторые типы объектов, с которыми человек особенно хорошо знаком, могут быть особенно для него интересными. С этими объектами может быть связана большая изощренность структур коры; они могут обрабатываться быстрее. Однако независимо от точного количества образов, которые могут быть идентифицированы приматами, и независимо от неравномерного распределения по классам объектов, существует общая система, способная различать множество самых разных образов. Мы даже можем различать образы, видимые впервые (это называется обучением с первой попытки).
5. *Реализуемость в алгоритмах компьютерной обработки изображений.* Успешная теория зрения должна быть описана достаточно подробно, чтобы ее можно было реализовать с помощью вычислительных алгоритмов обработки изображений. Алгоритм, работающий с изображением, принимает изображение или последовательность изображений в качестве входных данных и, произведя вычисления, посылает решение на выход. Требование такого количественного алгоритма является важным, потому что вычислительная реализация позволяет нам запускать симуляции и количественно сравнивать эффективность модели с поведенческими реакциями людей или животных. Моделирование также позволяет напрямую сравнивать вычислительную работу данной модели с нейрофизиологической активностью на различных этапах обработки изображений. Вычислительная модель может быть протестирована с теми же изображениями, которые используются в поведенческих или нейрофизиологических экспериментах.

В отличие от моделей, работающих непосредственно с изображениями, существуют различные увлекательные идеи и теоретические по-

строения о зрении, которые не реализуются с помощью вычислительных алгоритмов. Мы можем назвать эти идеи основанными на языке схематизациями или вербальными моделями. В качестве конкретного примера – вербальная модель может утверждать, что в зрительной системе есть фильтры, которые извлекают информацию о цвете, краях изображения, текстуре и наличии лиц. Вербальные модели могут быть полезны в области моделирования зрения и могут вдохновить на разработку вычислительных моделей. Однако вербальные модели недостаточно конкретизированы и поэтому склонны к неверной интерпретации (что такая текстура или лицо? Как извлекается информация о цвете?). Вербальные модели не дают количественных прогнозов (насколько быстро подопытный будет различать изображения с разной ориентацией краев? Какова будет частота срабатывания популяции нейронов, которые отличают одно лицо от другого?). Поскольку вербальные модели не определены должным образом, они не являются фальсифицируемыми. Более того, мы не можем легко сравнивать различные вербальные модели между собой, или вербальные модели с количественными моделями, или вербальные модели с поведенческими либо нейронными реакциями.

Алгоритмическая реализация заставляет нас строго формулировать предположения и формализовывать вычислительные шаги. Таким образом, вычислительные модели можно сравнивать не только с другими вычислительными моделями, но и с поведенческими и нейронными реакциями людей и животных. Реализация также может помочь нам отладить теорию, обнаруживая ошибочные скрытые предположения и случаи, когда эффективность расходится с поведенческими или нейронными реакциями.

6. *Ограничение зрением приматов.* Для простоты здесь мы следуем примеру предыдущих глав и ограничиваем обсуждение вычислительных моделей зрением приматов. Есть сильное сходство в зрении на поведенческом и нейрофизиологическом уровнях между макаками (один из основных видов для нейрофизиологических исследований) и людьми. Некоторые из этих моделей могут быть применимы к другим видам (например, кошкам и грызунам). Иные особенности модели могут потребовать доработки и модификации для других видов. Но если озабочиться построением модели зрения беспозвоночных (например, мух), скорее всего, потребуются более радикальные общие изменения в моделях.
7. *Биофизическое подобие.* Мы стремимся напрямую связать теоретические основы зрения с реальными нейронными цепями мозга, таким образом делая мост между тремя уровнями анализа, предложенными Дэвидом Марром и Томазо Поджио (раздел 1.9). В итоге, вычислительная реализация должна быть основана на нейронных сетях, а это означает, что модель должна быть способной объяснить, как происходят вычисления, базируясь на работе основных элементов вычислений в нейронных цепях, то есть нейронов. Мы ограничиваемся моделями, которые являются биофизически подобными, и при этом пропускаем

обширную литературу по компьютерному зрению, в которой исследователи пытаются решить аналогичные задачи без прямого обращения к схемам зрительной коры головного мозга.

Чисто инженерные подходы к зрению полезны с практической точки зрения независимо от того, имеют ли они какое-либо отношение к нейронным цепям реального мозга. В конечном счете компьютеры могут быть успешными в шахматах без какой-либо прямой связи с тем, как в эту игру играют люди, или полет самолета имеет весьма опосредованное отношение к полету птицы, подходы компьютерного зрения могут достичь высокой эффективности в визуальных задачах, не имитируя нейронные цепи. Такие алгоритмы могут быть полезны в различных повседневных задачах, но они не представляют собой биофизически подобную модель зрения приматов.

Преимущество внимания к нейронным цепям и их поведению состоит в том, что для решения задач, которые могут быть легкими для людей и трудными для машин, мы можем черпать вдохновение из биологии. То, что мы настаиваем на биологическом подобии, не должно означать, что подобные биологическим операции или архитектура в существующих нейронных сетях необходимы, не говоря уже о достаточности, для решения визуальных задач. Вполне вероятно, что в этот естественный порядок нам придется внести существенные изменения, но окончательный вид реализации должен иметь биологическое подобие. Поиск правильного уровня детализации при определении биофизического подобия остается интересной задачей (разделы 7.2 и 7.8).

8. *Ограничение только зрительной системой.* Зрительная система не изолирована от остального мозга. Между зрительной корой и другими сенсорными областями коры, системами памяти в медиальной височной доле, лобными долями и иными областями мозга существует множество связей. Несмотря на то что мы часто операционализируем<sup>1</sup> несколько задач, пытаясь максимально отделить зрение от других процессов, в реальном мозге границы между зрением и другими видами деятельности часто стираются. Связи вне зрительной системы играют важную роль в обработке зрительного сигнала, преимущественно через сигналы обратной связи, которые включают ожидания (например, расчет вероятности того, что присутствие льва в офисе микроскопично), через предварительно накопленные знания (например, объект похож на другой объект, с которым мы знакомы) и через кросс-модальную интеграцию (например, объект – возможно, музыкальный инструмент из-за того, что он издает музыкальный звук). В качестве первоначального упрощения и в качестве стратегии решения сложной задачи мы ограничиваем обсуждение исключительно зрительной системой.

---

<sup>1</sup> Операционализация – это процесс превращения теории или концепции в практический инструмент, пригодный для использования в ходе эмпирического конфликтологического исследования. – Прим. ред.

## 8.2. Общие темы в моделировании центрального потока обработки зрительного сигнала

Несколько исследователей предложили вычислительные модели, которые стремятся уловить некоторые важные принципы, лежащие в основе преобразований сигнала в центральном зрительном потоке приматов. Прежде чем обсуждать некоторые из этих моделей более подробно, мы начнем с определения тем, общих для многих моделей.

Входными данными для моделей зрительной коры обычно является изображение, определяемое матрицей, которая содержит кодированный цвет каждого пикселя. Чаще это трехмерная матрица с интенсивностями красного, зеленого и синего (RGB) для каждого пикселя. Модели могут работать с одномерными входными значениями в градациях серого, и также можно расширить модели до более чем трех входных измерений, например, для рассмотрения зрения животных, которые могут видеть за пределами видимой людьми части спектра. Динамические входы могут подразумевать последовательность кадров изображения. Из-за типа доступных изображений и видео, а также из-за требуемых вычислительных ресурсов большинство моделей имеют дело с пространственно-ограниченной версией полного визуального поля; например, в известном исследовании компьютерного зрения, проведенном группой Джейфри Хинтона, в качестве входных данных использовались изображения размером  $224 \times 224$  пикселя.

Поскольку основное внимание часто уделяется вычислительным свойствам центральной зрительной коры, многие исследователи игнорируют сложности моделирования вычислений в сетчатке и LGN. Пиксели предназначены для грубого представления продукции ганглиозных клеток сетчатки или клеток LGN. Карта соответствия между пикселями и угловой координатой поля зрения не всегда явно выражена в моделях. Большинство моделей в качестве входных данных используют изображения с одинаковым разрешением по всей площади, без учета дискретизации, зависящей от эксцентричности, которая проявляется как в сетчатке, так и по всей зрительной коре (рис. 2.7 и 2.8). Эти предположения, конечно, относятся к числу многих чрезмерных упрощений в типичных вычислительных моделях; изображения проходят несколько преобразований, прежде чем ганглиозные клетки сетчатки передают информацию в LGN и в кору (гл. 2). Включение более точного учета сетчатки и схем LGN, скорее всего, улучшит эффективность и надежность существующих моделей зрения.

Большинство моделей имеют иерархическую и глубокую структуру, которая имитирует приблизительно иерархическую архитектуру центральной зрительной коры (рис. 1.5, раздел 5.12). Свойства глубоких нейронных сетей привлекли большое внимание в вычислительном мире, хотя математика обучения в глубоких нейронных сетях с нелинейными откликами гораздо менее понятна, чем в их поверхностных аналогах. Неокортекс имеет стра-

тегию «разделяй и властвуй», и разработчики компьютерных моделей приняли ее тоже. Согласно этой стратегии, сложная задача делится на множество более простых задач (раздел 5.10). Поднимаясь вверх по иерархической структуре модели, блоки на более высоких уровнях обычно имеют большие рецептивные поля, реагируют на более сложные визуальные особенности и демонстрируют повышенную степень толерантности к трансформации предпочтаемых ими объектов.

Большинство вычислительных моделей явно или неявно предполагают, что «кора есть кора»; то есть существуют канонические микросхемы и вычисления, которые повторяются снова и снова по всей зрительной области. Таким образом, визуальная обработка может быть аппроксимирована иерархией последовательных вычислительных шагов, каждый из которых довольно прост и включает в себя основные биофизически подобные операции, такие как вычисление скалярных произведений, применение нелинейного преобразования к интегрированной активности в нейронной соме и нормализация выходного сигнала (раздел 7.2).

## 8.3. АРСЕНАЛ МОДЕЛЕЙ

Самой старой идеей для распознавания визуальных объектов является сопоставление образов, когда модель хранит в памяти определенное количество образов и любое новое изображение сравнивается на уровне пикселей, пиксель за пикселеем, с этими шаблонами. Прямое сопоставление образов на уровне пикселей для распознавания образов не работает. Даже сдвиг на один пиксель может создать серьезные проблемы для алгоритма, который просто сравнивает входные данные с сохраненным шаблоном попиксельно.

Как отмечалось в начале этой главы, ключевой проблемой при визуальном распознавании является то, что один и тот же объект может вызывать бесконечное количество разных проекций на сетчатке глаза. Если бы все объекты всегда были представлены в стандартизированном положении, масштабе, повороте и освещении, распознавание было бы значительно проще. На этом понятии основаны несколько подходов по попытке преобразовать входное изображение в стандартизованный (каноничный) формат путем сдвига, масштабирования и поворота объектов. Необходимые преобразования обычно достаточно сложны. Хотя искусственные вычислительные стратегии могут преодолеть некоторые из этих проблем, не совсем ясно, как мозг может реализовывать такие сложные операции, как вращение и прочие преобразования, а также нет какой-либо очевидной связи этого семейства моделей с нейрофизиологическими откликами нейронов, наблюдаемыми вдоль тракта центральной зрительной коры.

Множество моделей основаны на описании объекта как его частей и соотношений между ними. Идея этого подхода заключается в том, что может существовать небольшой словарь частей объекта и небольшой набор возможных связей, которые действуют как строительные блоки всех объектов. Это предположение можно проследить в истории до известной работы Дэвида Марра (1945–1980).

Разработчики и исследователи искусственного интеллекта ввели понятие структурных описаний. Точно так же, как математическая функция с «хорошим» поведением<sup>1</sup> может быть разложена на сумму ее составляющих (например, в виде параметрической функции или Фурье-представления), идея представлять объекты как простую сумму их частей привлекательна, потому что это, возможно, поможет легче распознать их без учета возможных преобразований. В простейших реализациях эти модели основаны на обнаружении определенной группы частей, характеризующих объект, подход, который допускает то, что перестановка частей не ухудшит распознавание объекта, но на самом деле это так (например, дом с гаражом на крыше и дымовая труба на полу). Более сложные версии включают взаимодействие между частями объекта и относительное расположение частей объекта. Этот подход превращает задачу распознавания объекта в задачу распознавания частей объекта плюс задачу распознавания характерных соотношений между его частями. Не совсем очевидно, что распознавание частей объекта должно быть проще, чем распознавание объектов, и не очевидно, что любой объект может быть однозначно и лаконично описан универсальным и небольшим словарем составляющих его более простых частей. Различие между предметами и частями также не определено должным образом, хотя имели место несколько вычислительных реализаций, основанных на структурных описаниях деталей. Что еще более важно, не совсем очевидно, как эти структурные описания связаны с нейрофизиологией центральной зрительной коры. Несмотря на эти предостережения, идея разложения объекта на части и вычислительные преимущества композиционного представления привлекательны и заслуживают дальнейшего изучения.

Серия вычислительных алгоритмов, обычно основанная на литературе по нейронным сетям, пытается построить глубокие структуры, целью которых является реконструкция входных данных. Одна из версий этого типа моделей называется *автокодировщиком*. В экстремальной версии данного типа сети нет потери информации по всей глубокой иерархии, а обратные сигналы несут информацию, способную воссоздать произвольные входные данные в нижних визуальных областях. Для таких глубоких сетей с автокодировщиком есть интересные приложения – в частности, возможность выполнения уменьшения размерности. Однако цель коры прямо противоположна идеальной реконструкции ввода; ее цель – это потеря информации биологически полезными способами (разделы 5.10 и 5.13). Неравенство обработки данных предусматривает, что объем информации<sup>2</sup>, содержащийся в сигнале, не может быть увеличен с помощью какой-либо обработки без добавления дополнительной внешней информации. Рассмотрим входное изображение, которое обрабатывается с помощью последовательности шагов А – В – С.

<sup>1</sup> Имеется в виду так называемая «гладкая» функция. – Прим. ред.

<sup>2</sup> Здесь и далее под «информацией» имеется в виду «полезная информация», поскольку добавление шума однозначно добавит объем информации, но ничего хорошего от такой операции не получится. В качестве предельного случая мы можем получить т. н. «белый шум», который содержит максимально возможный объем информации для данных параметров сигнала, однако «полезной» информации он не содержит вообще. – Прим. ред.

Массив информации на уровне С может содержать меньше информации об исходном изображении, чем на уровне А (в качестве тривиального примера мы можем сигнал на шаге В умножить на 0). Массив информации на уровне С может содержать тот же объем информации, что и на уровне А (в качестве тривиального примера мы можем скопировать сигнал из А в В и далее в С). Однако представление на уровне С не может содержать больше информации об исходном изображении, чем само исходное изображение; этапы обработки не могут создать новую информацию. По мере продвижения по зрительной системе, без добавления внешней информации информационное содержание должно либо уменьшаться, либо оставаться прежним. Не понятно, зачем строить целую сеть для копирования входных данных (даже если для копии требуется меньше данных). Другими словами, ключевая цель вентральной зрительной коры состоит в извлечении *важной* информации, такой как идентичность объекта, несмотря на изменения входных данных на уровне пикселей.

Не вся информация теряется из-за обработки вентральной корой, даже если эта информация безразлична для данной задачи. Например, можно считывать местоположение объекта с нейронов в нижней височной коре, несмотря на их относительную устойчивость к изменениям положения (разделы 6.5 и 6.7). Однако в то время как сетчатка имеет нейроны с размерами рецептивного поля, охватывающими несколько минут дуги ( $1^\circ$  угла зрения равен 60 угловым минутам), и они могут отслеживать временные изменения с частотой 100 Гц, нейроны в нижней височной коре более грубы в пространственном и временном измерениях. Несмотря на то что у нас нет полного понимания того, какая информация теряется при преобразованиях вдоль вентральной зрительной коры, а какая сохраняется, ясно, что некоторая информация теряется; целью обработки коры головного мозга не является точная реконструкция.

В сообществе нейрофизиологов существует несколько подходов к «метрическому» моделированию, при которых исследователи пытаются параметрически задать пространство образов, а затем записать активность нейронов вдоль вентрального зрительного потока в ответ на заданные параметры. Например, в некоторых случаях исследователи начинают с представления подопытным различных образов в поисках стимула, вызывающего наиболее сильные реакции. Впоследствии они манипулируют «предпочтительным» стимулом (объектом), удаляя различные его части и оценивая, как эти преобразования изменяют отклики нейронов. Хотя эти подходы интересны, они страдают от трудностей, присущих рассмотрению произвольных форм, которые могут или не могут составлять действительно «предпочтительные» стимулы<sup>1</sup>. Кроме того, в некоторых случаях изученные трансформации выявляют только антропоморфные предубеждения относительно того, какие особенности могут иметь значение. Другой подход – параметрическое определение форм. Например, в нескольких интересных исследованиях рассматривалось семейство простых геометрических фигур, параметризованных

---

<sup>1</sup> То есть оперируют изначально ограниченным набором входных стимулов. – Прим. ред.

различными типами кривизны, и моделировались нейронные реакции в шестимерном пространстве, определяемом суммой гауссиан с параметрами, заданными кривизной, ориентацией, относительным положением и абсолютной величиной и положением элементов контура на дисплее. Такой подход привлекателен, потому что он имеет замечательное свойство, позволяющее исследователям строить «настраиваемые кривые», подобные тем, которые использовались для представления активности блоков в более ранних визуальных областях. Однако этот подход также делает серьезные предположения о типах форм, предпочитаемых блоками. Эти параметризованные дескрипторы формы стимула с трудом можно отнести к вычислительным моделям обработки изображения, применимым к любому возможному естественному изображению.

Визуальная система делает гораздо больше, чем просто распознавание, и развитие визуальной системы в онтогенезе включает частично неконтролируемое обучение из окружающей среды во время выполнения естественных задач ориентации, социальных взаимодействий и игр. Многие вычислительные модели зрения были протестираны в задачах распознавания визуальных образов. Привлекательной особенностью задач распознавания является то, что их можно легко оценить на поведенческом, нейрофизиологическом и вычислительном уровнях. Кроме того, многие задачи распознавания могут быть непосредственно протестираны на животных. Прежде чем мы углубимся в современные модели обработки изображений, давайте определим различные подходы к решению задачи распознавания объектов.

## 8.4. ОБЩАЯ СХЕМА РЕШЕНИЯ ЗАДАЧИ

### РАСПОЗНАВАНИЯ ОБРАЗОВ

Рисунок 8.1 иллюстрирует типичный подход компьютерного зрения к решению задач зрительного распознавания. Рассмотрим серию  $M$  помеченных изображений  $(\mathbf{x}_i, y_i)$ , где  $i = 1, \dots, M$ ,  $\mathbf{x}$  – это матрица, представляющая изображение, а  $y$  – название категории (например, присутствует лицо или нет). Набор признаков  $\mathbf{f}$  извлекается из изображений:  $\mathbf{f}_i = g(\mathbf{x}_i)$ , где  $g$  – вычислительная модель. Эти признаки могут включать такие свойства, как края форм, основные составляющие части и цвета, среди многих других признаков. Выбор этих признаков является одним из ключевых аспектов, разделяющих разные алгоритмы компьютерного зрения. Функция  $g$ , извлекающая признаки, может быть жестко запрограммирована или иметь несколько параметров – мы назовем их  $w$ , – которые требуют настройки для конкретной задачи распознавания.

Мы рассмотрим настройку этих параметров в разделе 8.6. А пока предположим, что параметры  $w$  известны и фиксированы. После извлечения адекватного набора признаков  $\mathbf{f}$  используется схема контролируемого обучения для изучения карты соответствия между этими параметрами и названиями  $y$ . Например, может использоваться классификатор машины опорных вектор-

ров (SVM) с линейным ядром для изучения соответствия между признаками и названиями. Этот процесс аналогичен считыванию из популяции нейронов, описанному в разделе 6.7, за исключением того, что здесь мы используем вычисленные, а не биологически извлеченные признаки, закодированные в спайковой частоте.



**Рис. 8.1** ♦ Общая схема задач визуальной классификации. Признаки извлекаются из изображения (или видео). Эти признаки используются для тренировки классификатора посредством контролируемого обучения. Полученный результат применяется для присвоения названий новым изображениям (т. е. изображениям, отличным от тех, которые использовались во время обучения)

Далее следует процедура перекрестной проверки (раздел 8.8), разделяющая данные на обучающий набор и тестовый набор, чтобы гарантировать, что алгоритм оценивается на новых данных, — то есть чтобы гарантировать, что просто запоминание каждой точки данных, представленных при обучении, не приведет к положительному результату классификации. Некоторым исследователям нравится разделять данные на обучающий набор, основной набор и набор для тестирования. В этом случае исследователи точно настраивают гиперпараметры модели, учитывая набор для обучения и набор для проверки, а затем используют набор тестов для окончательной оценки результата. После обучения алгоритм оценивается с изображениями тестового набора. Используя разные алгоритмы, применяемые к одним и тем же данным, можно количественно сравнить достоинства альтернативных вычислительных моделей.

## 8.5. ВОСХОДЯЩИЕ ИЕРАРХИЧЕСКИЕ МОДЕЛИ ВЕНТРАЛЬНОГО ЗРИТЕЛЬНОГО ПОТОКА

Давайте вернемся к типу иерархических моделей нейронных сетей, который представлен в разделах 7.3 и 7.5. Иерархическая сетевая модель может быть описана серией уровней  $l = 0, 1, \dots, N$  (рис. 7.4). Каждый слой содержит  $n_l \times n_l$  блоков, помещенных в матрицу (матрица не обязательно должна быть квадратной), каждый блок имеет ограниченное поле восприятия и, следовательно, активируется определенным местом изображения. Кроме того, в каждом месте может быть несколько различных фильтров  $K_l$  (иногда называемых ядрами), создающих набор таких матриц; например, входное изображение ( $l = 0$ ) может иметь  $K_0 = 3$  цвета. Такой набор матриц в каждом

слое называется тензором. Активность всех блоков на каждом уровне может быть представлена значением  $x_l$  ( $x_l \in \mathbb{R}^{n_1 \times n_l \times K_l}$ ). Каждое входное значение  $x$  обычно представлено скаляром, называемым его активацией, что можно грубо интерпретировать как спайковую частоту блока. Изображение подается на вход в начальный слой;  $x_0$  представляет значения яркости пикселей в изображении. Другое изображение приведет к другому набору активаций  $x_i$  (в предыдущем разделе мы использовали нижний индекс  $i$  для обозначения активаций, или признаков, для каждого отдельного изображения, которые здесь опущены для простоты и которые не нужно путать с субиндексом  $l$ , используемым здесь для обозначения данного слоя).

От одного уровня к другому матрицы преобразуются с помощью различных операций свертки (раздел 7.7), нелинейных операций, таких как операция ReLU (рис. 7.2), и операций субдискретизации, таких как операция на рис. 7.6. В наиболее типичном сценарии эти операции выполняются между уровнем  $l$  и уровнем  $l + 1$ . Все эти операции вместе называются *блоком свертки*, включая саму операцию свертки, ReLU и операцию субдискретизации. Фактически один биологический нейрон может выполнять все эти вычисления, что схематично показано на рис. 7.1.

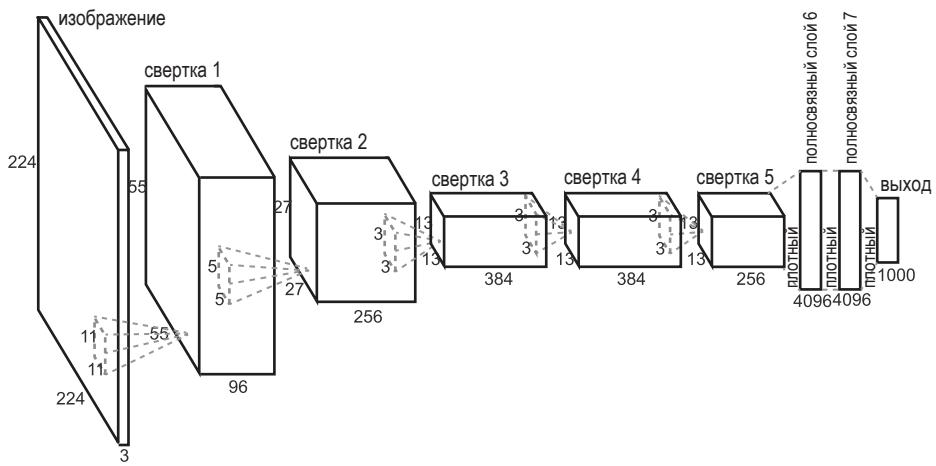
Эта формулировка, основанная на последовательной обработке, предполагает, что активность на данном уровне зависит исключительно от модели активности на предыдущем уровне. Это упрощение подразумевает, что игнорируются как минимум три типа соединений (рис. 7.4): (i) соединения, которые «пропускают» уровень в иерархии (например, синапсы из V1 непосредственно в область V4, пропускающие область V2); (ii) нисходящие связи (например, синапсы от V2 к V1) и (iii) связи внутри слоя (например, горизонтальные связи между нейронами с аналогичными предпочтениями в V1). Некоторые варианты, такие как так называемая архитектура ResNet, также включают соединения, которые обходят некоторые уровни (представленные в разделе 7.4).

В большинстве моделей негласно предполагается, что существуют общие правила, суммарно определяемые выражением «кортекс – это кортекс», так что только несколько типов преобразований допускаются при вычислениях от одного уровня к другому. Одной из первых моделей, нацеленных на описание распознавания объектов, вдохновленных нейрофизиологическими открытиями Хьюбела и Визеля (раздел 5.4), был неокогнитрон, разработанный японским ученым – специалистом в области информатики Кунихико Фукусима. В этой модели были две возможные операции: функция линейной настройки (выполняемая «простыми» клетками) и нелинейная операция «ИЛИ» (выполняемая «сложными» нейронами, раздел 5.5). Эти две операции чередовались и повторялись на нескольких уровнях вычислительной иерархии. Эта модель продемонстрировала, что такие линейные и нелинейные каскады могут обеспечить анализ масштаба и положения в задаче распознавания букв. Архитектура неокогнитрона вдохновила несколько последующих исследований.

Одной из таких попыток расширить вычислительные возможности неокогнитрона является модель HMAX, разработанная Максом Ризенхубером, Томасом Серром и Томазо Поджио из Массачусетского технологического института. Эта модель характеризуется чисто прямой и иерархической ар-

хитектурой. Изображение, представленное значениями оттенков серого, сворачивается с помощью фильтров Габора (раздел 5.7) в нескольких масштабах и положениях, чтобы имитировать ответы простых ячеек в V1. Подобно другим вычислительным алгоритмам, модель состоит из каскада линейных и нелинейных операций. У этих операций есть только две разновидности в данной модели: операция настройки и операция максимальной субдискретизации. HMAX и подобные архитектуры были подвергнуты нескольким тестам, включая сравнение с психофизическими измерениями и нейрофизиологическими реакциями, которые мы обсудим позже в этой главе.

Это семейство моделей называется глубокими нейронными сетями, в отличие от *неглубоких* сетей, целью которых является выполнение всех вычислений на одном уровне. Поскольку эти модели могут работать с изображением, их также можно напрямую использовать для решения задач компьютерного зрения. Действительно, в последнее десятилетие эти глубокие иерархические модели получили признание и распространение в компьютерном зрении. Одна такая модель, часто называемая *AlexNet*, представленная Алексом Крижевски и Джоном Хинтоном в 2012 году (рис. 8.2), вызвала переполох в мире компьютерного распознавания, поскольку она привела к значительному улучшению способности идентифицировать объекты в соревнованиях по компьютерному зрению. Впоследствии AlexNet стимулировал прогресс во многих других задачах распознавания образов (разделы 9.2 и 9.4). Важной особенностью AlexNet и многих других нейронных сетей является возможность настройки параметров  $w$  для повышения эффективности.



**Рис. 8.2** ♦ Глубокая сверточная нейронная сеть. Схема архитектуры глубокой сверточной нейронной сети AlexNet, состоящей из входного слоя, пяти сверточных слоев, двух полносвязанных слоев и выходного слоя. Числа под каждым слоем обозначают размер в пикселях. Числа внутри каждого слоя обозначают размер сверточных фильтров. Измененное по Крижевский и др., 2012

Исторически сложилось так, что многие усилия в разработке компьютерного зрения заключались в попытках делать все более совершенные функции

для извлечения уникальных признаков из изображения. Затем они передавались подходящему классификатору. Все обучение, зависящее от задачи, происходило на уровне классификатора. Было обнаружено, что различные типы признаков обычно полезны для задач классификации объектов, включая извлечение краев, цветов, главных компонентов, инвариантные к сдвигу преобразования признаков (SIFT), углы, пространственно-частотную декомпозицию и многие другие. Классификатор, такой как SVM, отвечал за изучение карты соответствия между этими функциями и соответствующими названиями изображений, как показано на рис. 8.1. В неокогнитроне и в первоначальной реализации архитектуры HMAX веса между уровнями устанавливались вручную и фиксировались.

В отличие от этих подходов, основная работа в настоящее время заключается в создании полностью обучаемых (end-to-end) систем, обычно с фиксированной архитектурой со случайно инициализированными весами, но где все веса являются переменными и могут быть изменены для достижения наилучшей точности классификации результата.

## 8.6. Изучение весовых параметров

Мы еще не описали, как устанавливаются параметры извлечения признаков  $w$ . В общем мы будем рассматривать модели нейронных сетей (раздел 7.3), где основными параметрами являются веса, которые диктуют, как активность в данном блоке влияет на активность в постсинаптических целевых блоках, как правило, на следующем уровне. Давайте теперь рассмотрим пример способа изучения этих весов, иллюстрирующий замечательные вычисления, которые могут выполняться нейронными сетями. Веса можно определить контролируемым образом (когда у нас есть названия для каждого изображения) или неконтролируемым образом (например, автоматически извлекая статистические закономерности из изображений окружающей среды). Здесь мы сосредоточимся на стратегиях контролируемого обучения и вернемся к неконтролируемой настройке весов в разделе 8.17.

Один из самых ранних примеров биологически вдохновленного вычислительного алгоритма, двухслойная нейронная сеть, называемая *перцептроном*, может быть обучена для выполнения некоторых нетривиальных задач классификации. Представьте, что у нас есть данные, которые мы хотим разделить на две возможные группы. Например, может быть набор изображений собак или кошек (каждое изображение содержит только одно животное), и мы хотим научить алгоритм различать, содержит ли изображение собаку или кошку. Каждое изображение, индексируемое  $i$ , представляет собой матрицу значений оттенков серого, которые могут быть векторизованы и представлены  $x_i$ . С каждым изображением у нас есть ассоциированное название  $y_i = +1$  (собака) или  $-1$  (кошка). У нас есть обучающий набор, состоящий из нескольких таких изображений-примеров. В этом типе задач всегда важно разделить данные на обучающий набор (используемый для подбора параметров) и тестовый набор (используемый для оценки эффективности модели), процесс называется *перекрестной проверкой* (раздел 8.8; см. дальнейшее

обсуждение в разделе 9.12). В двухуровневой сети перцепtronов мы будем считать, что входом для выходного блока является  $w \cdot x$ , где  $\cdot$  представляет собой скалярное произведение. Выходной сигнал у принимает значение  $+1$ , если  $w \cdot x - \gamma > 0$ , и  $-1$  в противном случае, где  $\gamma$  – пороговое значение. Правило обучения перцептрона говорит нам, как выбрать веса  $w$ , чтобы минимизировать ошибку в данной задаче классификации.

Вместо задачи бинарной классификации нас может заинтересовать аппроксимация заданной выходной функции  $h(s)$  (например,  $h(s)$  может представлять собой частоту срабатывания нейрона в коре головного мозга в ответ на стимул  $s$ ). Для заданного стимула  $s$   $h(s)$  является целевым выходом для нейронной сети, и мы определяем  $\hat{h}(s)$  как фактический выход сети. Ошибка – это квадрат разницы между этими двумя переменными:  $E = (h(s) - \hat{h}(s))^2$ ; это евклидово расстояние является типичным способом оценки ошибки, которая обладает прекрасным свойством дифференциации; это скоро станет полезным в обсуждении. Градиентный спуск относится к изменению  $w$  для минимизации ошибки в этой задаче путем внесения корректировок в  $w$  в направлении наибольшего изменения ошибки,  $w \rightarrow w + \epsilon \nabla_w E$ , где  $\epsilon$  – скорость обучения, а  $\nabla_w E$  – градиент ошибки в направлении  $w$ .

Задачи классификации не обязательно ограничивать двумя классами, например кошки и собаки. В общем, цель состоит в том, чтобы взять изображение и присвоить ему название, ярлык. Например, цель может заключаться в том, чтобы определить, содержит ли изображение рукописные цифры 0, 1, 2, … 9, как в базе данных MNIST (раздел 8.10), для различия кошек от собак, или идентифицировать лицо.

Ситуация усложняется, когда у нас несколько слоев. Теперь изменение весов в одном слое влияет на следующий слой, который, в свою очередь, влияет на очередной следующий слой, и эти изменения распространяются на весь выход. Нам необходимо учитывать все эти взаимозависимости при настройке весов в глубокой нейронной сети. Одним из наиболее успешных способов корректировки весов с помощью контролируемого обучения в глубокой нейронной сети является *обратное распространение*, при котором разница между целевыми выходами и текущими выходами (то есть ошибка) распространяется обратно через градиентный спуск по всей сети. Обратное распространение – это элегантный пример применения цепного правила в дифференциальном исчислении<sup>1</sup>.

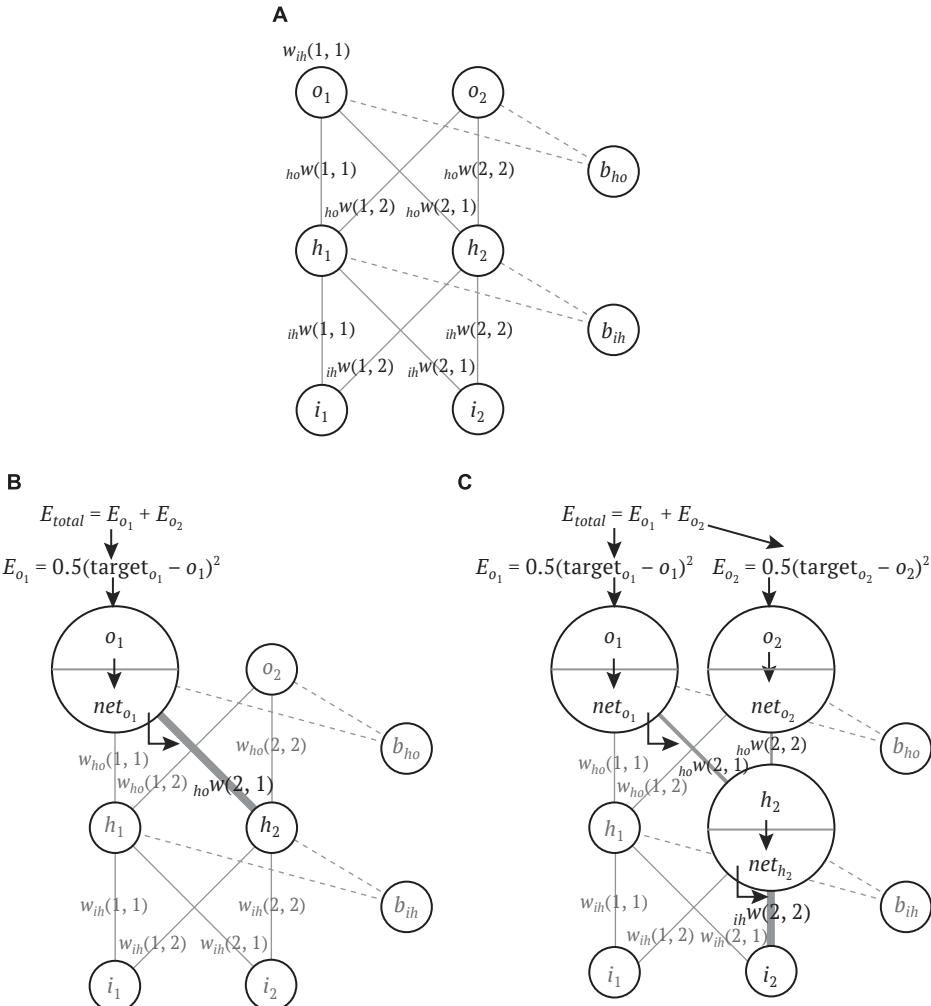
Давайте рассмотрим один простой пример, шаг за шагом, чтобы описать концепцию обратного распространения. Рассмотрим трехуровневую сеть, показанную на рис. 8.3А. Сеть состоит из входного слоя с двумя блоками, значения активации которых представлены  $i_1, i_2$ ; скрытого слоя с двумя скрытыми блоками, значения активации которых представлены как  $h_1, h_2$ ; и выходного слоя с двумя блоками вывода с активациями  $o_1, o_2$ . Термин «скрытый блок» в нейронной сети – это несколько странное название, которое относится ко всем блокам, не являющимся ни входом, ни выходом. Возможно, термин «промежуточный блок» был бы более разумным, но этот сленг получил широкое распространение. Вес от входного блока  $i$  к скрытому

<sup>1</sup> Правило дифференцирования сложной функции. – Прим. ред.

## 214 ♦ Глава 8

блоку  $h$  равен  $w_{ih}(i, h)$ , а вес от скрытого блока  $h$  к выходному блоку  $o$  равен  $w_{ho}(h, o)$ . Смещение для скрытого слоя –  $b_{ih}$ , смещение для выходного слоя –  $b_{ho}$ . Чистый вклад в каждый скрытый блок:

$$\begin{aligned} net_{h_1} &= i_1 * w_{ih}(1, 1) + i_2 * w_{ih}(2, 1) + b_{ih}; \\ net_{h_2} &= i_1 * w_{ih}(1, 2) + i_2 * w_{ih}(2, 2) + b_{ih}. \end{aligned} \quad (8.1)$$



**Рис. 8.3** ♦ Весовые параметры в глубокой нейронной сети можно определить с помощью обратного распространения ошибки. Глубокие сверточные нейронные сети используют обратное распространение – эффективную процедуру для обучения весов в режиме контролируемого обучения. (A) Пример трехуровневой нейронной сети. (B) Чтобы изменить вес  $w_{ho}(2,1)$ , мы вычисляем его влияние на общую ошибку, используя цепное правило (подробности см. в тексте). (C) Аналогично, чтобы изменить вес  $w_{ih}(2,2)$ , мы распространяем ошибку по всей сети. (Подробности см. в тексте; адаптировано из Мэтта Мазура)

Вместо операции ReLU (рис. 7.2) здесь мы вычисляем выход каждого скрытого блока, передавая входные данные через нелинейную логистическую функцию<sup>1</sup>:

$$\begin{aligned} h_1 &= \frac{1}{1 + e^{-net_{h_1}}}; \\ h_2 &= \frac{1}{1 + e^{-net_{h_2}}}. \end{aligned} \quad (8.2)$$

Выходы скрытых блоков – это входы блоков вывода. Чистый ввод в каждый блок вывода равен:

$$\begin{aligned} net_{o_1} &= h_1 * w_{ho}(1, 1) + h_2 * w_{ho}(2, 1) + b_{ho}; \\ net_{o_2} &= h_1 * w_{ho}(1, 2) + h_2 * w_{ho}(2, 2) + b_{ho}, \end{aligned} \quad (8.3)$$

и они также проходят через логистическую функцию:

$$\begin{aligned} o_1 &= \frac{1}{1 + e^{-net_{o_1}}}; \\ o_2 &= \frac{1}{1 + e^{-net_{o_2}}}. \end{aligned} \quad (8.2)$$

Теперь, учитывая входы  $i_1, i_2$  (мы можем рассматривать их как изображения, которые пытаемся классифицировать), мы получаем выходы  $o_1, o_2$ . Наши целевые значения – это выходы  $target_{o_1}, target_{o_2}$ ; мы можем рассматривать эти целевые значения как желаемые вероятности для названий класса 1 и класса 2. Например, если мы классифицируем изображение как кошку или собаку, желаемые вероятности могут быть 0 и 1 для кошек и 1 и 0 для собак. Общая ошибка составляет

$$E_{total} = E_{o_1} + E_{o_2} = 0.5[(target_{o_1} - o_1)^2 + (target_{o_2} - o_2)^2]. \quad (8.5)$$

Теперь представьте, что мы меняем один из весов, скажем  $w_{ho}(2, 1)$ , – что мы ожидаем от общей ошибки (рис. 8.3В)? Если мы изменим  $w_{ho}(2, 1)$ , это вызовет изменение в  $net_{o_1}$  (но не в  $net_{o_2}$ ; см. уравнение (8.3)). Изменение  $net_{o_1}$ , в свою очередь, вызовет изменение  $o_1$  (но не  $o_2$ ). Изменение  $o_1$  повлияет на  $E_{total}$ . Мы можем вычислить, какое изменение общей ошибки мы получим, используя цепное правило и разложив градиент  $E_{total}$  относительно  $w_{ho}(2, 1)$  на следующие части:

$$\frac{\partial E_{total}}{\partial w_{ho}(2, 1)} = \frac{\partial E_{total}}{\partial o_1} \frac{\partial o_1}{\partial net_{o_1}} \frac{\partial net_{o_1}}{\partial w_{ho}(2, 1)}. \quad (8.6)$$

Здесь определение квадратичной ошибки в уравнении (8.5) пригодится, потому что легко вычислить производные. Каждый из этих трех факторов

<sup>1</sup> Логистическая функция – решение простого нелинейного дифференциального уравнения первого порядка, на графике выглядит как сигмоида. – Прим. ред.

## 216 ♦ Глава 8

---

можно легко вычислить, взяв производные в уравнениях (8.3), (8.4) и (8.5) соответственно:

$$\frac{\partial E_{total}}{\partial o_1} = 2 * 0.5 * (o_1 - target_{o_1}); \quad (8.7)$$

$$\frac{\partial o_1}{\partial net_{o_1}} = o_1(1 - o_1); \quad (8.8)$$

$$\frac{\partial net_{o_1}}{\partial w_{ho}(2,1)} = h_2. \quad (8.9)$$

Чтобы уменьшить общую ошибку, мы изменим веса в соответствии с

$$w_{ho}(2,1) \rightarrow w_{ho}(2,1) - \varepsilon \frac{\partial E_{total}}{\partial w_{ho}(2,1)}, \quad (8.10)$$

где  $\varepsilon$  – скорость обучения, которая контролирует, насколько велики изменения весов на каждом шаге.

Мы можем выполнить ту же процедуру, чтобы изменить  $w_{ho}(2,2)$ ,  $w_{ho}(1,2)$  и  $w_{ho}(1,1)$ . В общем, в нейронной сети мы хотим изменить все веса, чтобы результат был как можно ближе к цели. Как изменить веса переходов с входных блоков на скрытые, например  $w_{ih}(2, 2)$ ? Мы следуем той же процедуре (рис. 8.3С), распространяя ошибку в обратном направлении от  $E_{total}$  до веса, который мы хотим изменить. Прохождение нескольких уровней требует еще нескольких операций, но мы следуем тем же идеям, что и раньше. Зависимость полной ошибки от  $w_{ih}(2, 2)$  проходит через скрытый блок  $h_2$  (а не  $h_1$ ). Следовательно, мы можем написать

$$\frac{\partial E_{total}}{\partial w_{ih}(2,2)} = \frac{\partial E_{total}}{\partial h_2} \frac{\partial h_2}{\partial net_{h_2}} \frac{\partial net_{h_2}}{\partial w_{ih}(2,2)}. \quad (8.11)$$

Последние два фактора очевидны (по аналогии с уравнениями (8.8) и (8.9)):

$$\frac{\partial h_2}{\partial net_{h_2}} = h_2 * (1 - h_2);$$

$$\frac{\partial net_{h_2}}{\partial w_{ih}(2,2)} = i_2.$$

Зависимость полной погрешности от  $h_2$  проходит через оба выходных блока. Следовательно, первый множитель в уравнении (8.11) становится

$$\frac{\partial E_{total}}{\partial h_2} = \frac{\partial E_{o_1}}{\partial h_2} + \frac{\partial E_{o_2}}{\partial h_2} = \frac{\partial E_{o_1}}{\partial o_1} \frac{\partial o_1}{\partial net_{o_1}} \frac{\partial net_{o_1}}{\partial h_2} + \frac{\partial E_{o_2}}{\partial o_2} \frac{\partial o_2}{\partial net_{o_2}} \frac{\partial net_{o_2}}{\partial h_2};$$

$$\frac{\partial net_{o_1}}{h_2} = w_{ho}(2,1);$$

$$\frac{\partial net_{o_2}}{h_2} = w_{ho}(2,2).$$

Согласно уравнению 8.8, имеем  $\frac{\partial o_1}{\partial net_{o_1}} = o_1 * (1 - o_1)$  и, аналогично,  $\frac{\partial o_2}{\partial net_{o_2}} = o_2 * (1 - o_2)$ . Согласно уравнению 8.7, имеем  $\frac{\partial E_{o_1}}{\partial o_1} = (o_1 - target_{o_1})$  и, подобным образом,  $\frac{\partial E_{o_2}}{\partial o_2} = (o_2 - target_{o_2})$ . Нам нужно изменить вес  $w_{ih}(2,2)$  на следующую величину:

$$w_{ih}(2,2) \rightarrow w_{ih}(2,2) - \epsilon \frac{\partial E_{total}}{\partial w_{ih}(2,2)}.$$

Прелесть этих шагов в том, что мы можем применять цепное правило независимо от того, насколько глубока сеть. Обычно ученые и инженеры создают нейронные сети, состоящие из более чем сотни слоев, и обучаются ими, используя, по сути, ту же процедуру обратного распространения, описанную здесь. Кроме того, итерация назад от последнего уровня позволяет избежать избыточных вычислений промежуточных членов при использовании цепного правила, так что все предыдущие члены из поздних уровней могут быть повторно использованы для более ранних уровней.

Учитывая пример с входными значениями  $i_1$ ,  $i_2$  и целевыми выходными значениями  $target_{o_1}$  и  $target_{o_2}$ , мы выполняем шаги, описанные ранее, чтобы изменить все веса в сети. В общем, у нас есть много примеров, состоящих из пар входных и целевых выходных значений (раздел 8.4). При *стохастическом градиентном спуске*<sup>1</sup> мы просматриваем эти примеры один за другим, изменения веса после каждой итерации. Также может быть создана группа, в которой расчеты выполняются для нескольких примеров перед фактическим изменением весов. Слово *стохастический* относится к выбору примеров и случайному упорядочиванию.

После того как мы пройдемся по всем примерам в обучающем наборе данной итерации, мы начнем новую итерацию, перенастроив веса. Эта процедура продолжается до схождения результатов. Скорость обучения  $\epsilon$  играет важную роль. Если скорость обучения слишком велика, результаты процедуры могут разойтись. Если она слишком мала, схождение может оказаться очень медленным, и алгоритм может застрять в локальных оптимумах. Было разработано несколько эвристических процедур для корректировки скорости обучения, иногда даже ее изменения на протяжении обучения (с более быстрым обучением в начале и более медленным обучением к концу). Также возможно применять разные скорости обучения для разных уровней. Есть множество других вариантов, которые выходят за рамки этой главы, и их можно найти в учебниках по компьютерному зрению и машинному обучению.

Как правило, начальные значения веса присваиваются случайным образом, но были проведены эмпирические исследования преимуществ различных начальных условий. Биологический мозг, скорее всего, не начинается с полностью случайных весов связности. Существует внутренняя структура,

---

<sup>1</sup> «Стохастический» – случайный. – Прим. ред.

которая действует на правила пластиности во время последующего обучения. Эта начальная структура может быть продуктом эволюции, а также продуктом развития индивида, зависящим от опыта его деятельности.

Особенно интересное начальное условие возникает, когда сеть «предобучена» (предварительно обучена) решению несколько иной задачи, отличной от той, которую мы стремимся решить. Применение весов, полученных при обучении в процессе решения одной задачи, к другой задаче называется «переносом обучения». Представьте, например, что мы хотим построить сеть, распознающую рукописные буквы. Можно сначала обучить сеть распознавать рукописные цифры и использовать веса из этой предварительно обученной сети в качестве начальных условий для задачи распознавания букв. Многие вычисления, необходимые для создания распознавателя рукописных цифр, могут использоваться совместно с вычислениями, необходимыми для создания распознавателя рукописных букв. Таким образом, запуск такой предварительно обученной сети может ускорить обучение, а также привести к такой же точности с использованием меньшего количества обучающих примеров. Не сразу очевидно, какие задачи подходят для такого подхода к переносному обучению. Интуитивно понятно, что если две задачи очень похожи, тогда этот подход может быть выгодным, но если две задачи слишком разные, то предварительное обучение может не дать никаких преимуществ.

Как отмечалось ранее, цепное правило позволяет распространять ошибку через глубокие сети со множеством уровней, а большее количество уровней обычно означает большее количество весов, которые необходимо настроить. Чтобы настроить большое количество весов, полезно иметь много примеров пар входов и целевых выходов. В случае алгоритмов классификации изображений эти примеры представлены в виде изображений и названий.

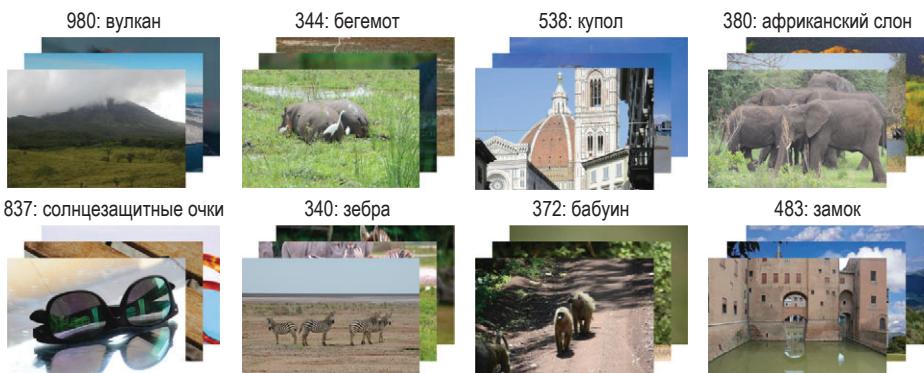
## 8.7. БАЗЫ ДАННЫХ НАЗВАНИЙ

В сообществе разработки компьютерного зрения достигнут значительный прогресс в большом количестве задач категоризации изображений. Этот прогресс был вызван сочетанием увеличенных вычислительных ресурсов, доступа к большому количеству цифровых изображений (а также видео) и захватывающих дискуссий на академических конференциях.

В последнее десятилетие количество цифровых изображений, доступных в сети, резко возросло. В 2019 году пользователи загружали порядка нескольких миллиардов цифровых изображений в день (например, Facebook: ~300 млн изображений в день; Instagram: ~100 млн изображений в день). Кроме того, многие пользователи непреднамеренно оказались чрезвычайно полезны сообществу компьютерного зрения, предоставляя все большее и большее количество контента в сочетании с «тегами», краткими подписями, «лайками» и другими комментариями. Каждую минуту в 2019 году люди делали больше цифровых фотографий, чем общее количество фотографий, доступных во всем мире столетие назад. Также резко увеличилось количество

загружаемых видеоматериалов (например, YouTube: 500 тысяч часов видео в день). Параллельно с доступностью изображений теперь существуют также доступные платформы, такие как Amazon Mechanical Turk, где пользователи могут получать подборки изображений за небольшую плату. Исследователи загружают свои изображения и платят пользователям Amazon Mechanical Turk, чтобы они называли их, что привело к созданию замечательных баз данных с аннотациями и названиями объектов на изображениях. Изображения, контент и сопутствующий экспоненциальный рост вычислительной мощности компьютеров открыли двери для использования сетей для задач распознавания с миллионами настраиваемых параметров.

Типичным примером является крупномасштабный проект зрительного распознавания ImageNet. Эта база данных состоит из цветных изображений, загруженных из интернета, каждое из которых имеет сопровождающую справку, что на нем изображено. В типичном случае эти справки включают в себя такие категории, как «вулкан», «бегемот», «купол» или «африканский слон» (рис. 8.4). Версия ImageNet 2014 года широко использовалась для сравнения различных алгоритмов компьютерного зрения при классификации объектов и включала 1000 классов объектов, 1 281 413 изображений для обучения (732–1300 изображений на класс) и 100 000 изображений для тестирования (100 изображений на класс).



**Рис. 8.4** ♦ Примеры изображений из базы данных ImageNet. Доступность баз данных, состоящих из миллионов изображений с названиями, значительно повысила эффективность алгоритмов контролируемого обучения при категоризации объектов

Тот факт, что изображения загружаются из интернета, является благословением и проклятием: благословением, потому что эти изображения обладают широким разнообразием признаков, когда объект, ответственный за название изображения, может появляться в разных положениях, в разных масштабах, поворотах, цветах, освещенности, различной степени окклюзии и других вариантах. В некотором грубом приближении это отражает естественное представление объектов в мире. Такое приближение не совсем точное, потому что эти изображения фильтруются через объективы и на них накладываются предубеждения фотографов. Например, изображений

бегемота в дождливой ночи, скорее всего, очень мало. Изображения, взятые из интернета, также являются проклятием из-за их неконтролируемого характера и большого количества других несколько разных контекстуальных факторов, которые способствуют классификации. Например, на трех изображениях «куполов» на рис. 8.4 (верхний ряд, третий столбец) пиксели в верхнем левом углу в основном синие. Кажется вероятным, что когда люди фотографируют купола, изображения оказываются на фоне неба, и наверху наблюдается большая площадь синего цвета. Напротив, ни один из примеров «бабуинов» (нижний ряд, третий столбец) не содержит синего цвета вверху. Тем не менее синий цвет наверху не является уникальным отличительным признаком куполов. Многие другие изображения также обычно содержат синий цвет вверху (например, вулканы, слоны, зебры и замки). Также, скорее всего, есть изображения куполов без синего цвета вверху и изображения бабуинов с синим цветом вверху. Дело в том, что в изображениях есть много сложных корреляций, которые лишь минимально связаны с самими названиями объектов. В зависимости от конкретной задачи и цели эти контекстные корреляции могут представлять помеху или, наоборот, являться полезными.

Еще одно любопытное свойство базы данных ImageNet заключается в том, что некоторые категории изображений весьма интригуют. Есть много названий категорий, которые мне пришлось бы искать в словаре (например, линь или юнко). В добавок многие из этих 1000 классов соответствуют специализированным и уточненным группам животных (сколько людей может отличить хлыстохвостую ящерицу от аллигаторовой или зеленой ящерицы, от ящерицы комода и плащеносной ящерицы?). Тем не менее компьютеры обучены распознавать эти категории с самого начала, и различие между хлыстохвостой и плащеносной ящерицей может быть столь же кристально ясным для компьютера, как для человека различие между солнцезащитными очками и головой. Цель соревнований по классификации заключается в количественной оценке и сравнении разных алгоритмов компьютерного зрения, что легко сделать с помощью различий между хлыстохвостой и плащеносной ящерицами. Мы вернемся к обсуждению дополнительных аспектов обучающих баз данных в разделе 9.12.

Алгоритмы компьютерного зрения, основанные на подходах с контролируемым обучением со случайно инициализированными весами, жадно требуют данных. ImageNet содержит порядка 1000 категорий, 1000 примеров на категорию, что дает  $10^6$  изображений. База данных такого размера не была доступна примерно до 2012 года, и поэтому эти изображения стали хорошей экспериментальной площадкой для разработки, уточнения и построения более сложных глубоких сверточных нейронных сетей. Учитывая огромное количество цифрового контента, следует ожидать мощнейшего роста размеров баз данных изображений.

Еще один заслуживающий внимания аспект ImageNet и аналогичных баз данных заключается в том, что он дает возможность прямого сравнения и тестирования различных алгоритмов. Сравнение того, как алгоритм X обрабатывает базу данных  $I_x$ , с тем, как алгоритм Y обрабатывает другую базу  $I_y$ , является сложной задачей, чем-то вроде сравнения яблок с бананами. Хотя

это простая концепция, тесты, основанные на стандартных базах данных, не распространены в других областях. Например, в области нейробиологии почти каждая лаборатория создает свои собственные задачи, используя свои изображения, созданные на заказ, визуализируя результаты, которые сложно интегрировать и сравнивать, а обмен данными все еще находится в зачаточном состоянии.

## 8.8. ПЕРЕКРЕСТНАЯ ПРОВЕРКА НЕОБХОДИМА

Вооружившись такой большой базой данных изображений с названиями, мы готовимся к обучению вычислительного алгоритма для изучения карты соответствия между отличительными признаками изображения и названиями изображений. Крайне важно, что нам нужно не просто запомнить каждую пару «изображение – название». Вместо этого мы хотим получать правильные названия для *новых* изображений, которые алгоритм никогда раньше не видел. Чтобы избежать явного механического запоминания, замаскированного под высокую эффективность, критически важно использовать *перекрестную проверку*, разделяя изображения внутри каждой категории названий на обучающий набор и тестовый набор. Все параметры модели могут быть изменены по усмотрению только во время исследования обучающего набора изображений, но нам не разрешается изменять какие-либо другие параметры при оценке модели на тестовом наборе.

В большинстве случаев мы используем случайные разбиения одной и той же базы данных на неперекрывающиеся наборы для обучения и наборы для тестирования. Доля изображений, которые входят в обучающую выборку, может не иметь большого значения; в обучающий набор мы можем взять хоть 50 % изображений из общего набора, или 70 %, или даже все, кроме одного, которое будет использоваться для тестирования в каждой итерации. Эта процедура повторяется несколько раз, и результаты обобщаются как средняя эффективность по всем случайным разбиениям плюс мера вариации внутри набора. В качестве контроля полезно случайным образом перемешивать названия изображений, повторять ту же процедуру, а также сообщать среднее значение и изменение эффективности в случае перемешивания названий.

В моделях глубокой сверточной сети этап обучения обычно сводится к контролируемому изменению весов посредством обратного распространения ошибки (раздел 8.6). Однако также возможно исследовать другие особенности модели, включая ее архитектуру, количество слоев, размер каждого слоя и вычислительные предустановки, если мы ограничимся обучающей выборкой. После обучения алгоритм тестируется с новыми изображениями, и определяется доля изображений, которые названы правильно.

Как правило, разделение данных на обучающий и тестовый наборы выполняется случайным образом. Поскольку в базе данных нет дубликатов, нет обрезанных картинок одних и тех же изображений и нет других потенциальных затруднений, случайного разделения должно быть достаточно, чтобы избежать самообмана с помощью механического запоминания, хотя мы еще

обсудим перекрестную проверку в разделе 9.10. Важно отметить, что базы данных следует тщательно контролировать, чтобы избежать таких проблем, как дублирование или слегка измененные версии одного и того же изображения, для того чтобы правильно оценить эффективность модели. Например, рассмотрим алгоритм распознавания лиц знаменитостей. Людям нравится обрезать одни и те же изображения знаменитостей снова и снова и загружать их в свои любимые социальные сети. Если случайное разбиение приводит к тому, что обучающий набор содержит ту же или практически такую же картинку, что и тестовый набор, то мы не выполняем перекрестную проверку должным образом и можем ввести себя в заблуждение, полагая, что алгоритм более результативный, по сравнению с тем, чем он на самом деле является.

## 8.9. ПРЕДУПРЕЖДЕНИЕ: МНОЖЕСТВО ПАРАМЕТРОВ!

Для современных нейронных сетей со множеством уровней интригующий аспект процедуры обратного распространения, описанный в разделе 8.6, который следует изучить на примерах, заключается в том, что корректировка весов включает в себя огромное количество свободных параметров. Рассмотрим изображение размером  $256 \times 256$  пикселей с тремя цветами: это 196 608 возможных вводных. Если существует 1000 возможных названий категорий выходных данных, то в простейшем сценарии картирования входных данных непосредственно на выходы у нас будет около  $200 \times 10^6$  параметров. База данных ImageNet содержит порядка  $10^6$  изображений. Другими словами, количество весов в нейронной сети (свободные параметры) может быть на порядки больше, чем количество обучающих примеров.

Иметь больше параметров, чем ограничений, может быть проблематично. В качестве простого примера давайте вернемся к основам линейной алгебры и рассмотрим систему линейных уравнений с четырьмя неизвестными. В общем, если у нас есть четыре независимых уравнения, мы гарантированно получим единственное решение. Однако если у нас есть только два независимых уравнения, система недоопределенна; без каких-либо дополнительных соотношений существует бесконечное количество возможных решений. Та же проблема возникает и при построении кривой. Если кривая имеет 10 свободных параметров, а у нас есть только пять наборов координат точек, ей принадлежащих, существует бесконечное количество решений, и можно получить сколько угодно кривых, которые будут точно, с нулевой ошибкой, описывать данные, но не будут экстраполироваться на возможные новые данные. В качестве примера рассмотрим график зависимости числа женщин от числа мужчин в данном штате, показывающий данные из пяти штатов. Вам, скорее всего, не захотелось бы использовать многочлен 10-й степени для описания данных!

Эти классические примеры недоопределенных систем и избытка параметров изучаются в математических классах средней школы. Что удивительно, так это то, что самые успешные модели компьютерного зрения, доступные сегодня, работают именно в этом режиме. В качестве иллюстрации проблемы гиперпараметризации – мы можем случайным образом переставить назва-

ния в обучающем наборе в ImageNet: слона обозначить как «стул», другого слона назвать «дерево», автомобиль обозначить как «солнцезащитные очки» и так далее. Систему компьютерного зрения можно обучить для достижения высокой эффективности даже на *обучающем наборе с этими, случайным образом перемешанными изображениями*. Другими словами, сеть с большим количеством параметров, чем обучающих примеров, обладает огромной дополнительной силой и даже может запомнить всю базу данных. Конечно, такая сеть может ошибиться, что еще раз подчеркивает важность перекрестной проверки. В настоящее время существует значительный интерес к пониманию того, как нейронные сети могут по-прежнему хорошо работать на тестовом наборе и избегать перегрузки, несмотря на огромное количество свободных параметров.

Один из способов смягчить потенциальную проблему перегрузки параметрами – использовать больше ограничений для их числа. Во многих задачах компьютерного зрения один из способов увеличить количество ограничений – получить больше данных. Получение данных с названиями может быть узким местом во многих практических приложениях, а также в более биологически подобных механизмах обучения. Следовательно, есть интерес к увеличению объема данных без прироста количества названий для похожих изображений, идея, обычно называемая «расширение базы данных». Для примера, зачем это нужно, рассмотрим такую базу данных, как ImageNet. Можно взять каждое изображение и обрезать его, развернуть по горизонтали, размыть, повернуть, добавить шум, а затем использовать его в качестве отдельного обучающего примера с тем же названием.

## **8.10. ИЗВЕСТНЫЙ ПРИМЕР: РАСПОЗНАВАНИЕ ЦИФР В СЕТИ ПРЯМОГО РАСПРОСТРАНЕНИЯ, ОБУЧЕННОЙ МЕТОДОМ ГРАДИЕНТНОГО СПУСКА**

В качестве примера применения этих идей рассмотрим задачу научиться распознавать рукописные цифры от 0 до 9. В знак уважения к введению Кернигана и Ричи в кодирование на языке программирования С большинство курсов программирования начинаются с изучения с того, как вывести текст «привет, мир» на экран. В машинном обучении эквивалент «привет, мир» – это научиться писать код для определения рукописных версий цифр от 0 до 9. База данных MNIST (Модифицированный национальный институт стандартов и технологий) состоит из 60 000 обучающих изображений и 10 000 тестовых изображений.

В 1998 году Янн ЛеКун и его коллеги разработали сеть прямого распространения, обученную методом градиентного спуска, которая могла достаточно хорошо выполнять эту задачу, достигнув коэффициента ошибок в 7,6 %, что было очень неплохо для того времени (при случайному угадывании доля ошибок составила бы 90 %, потому что есть 10 возможных классов). В более поздней вычислительной модели 2019 года коэффициент ошибок со-

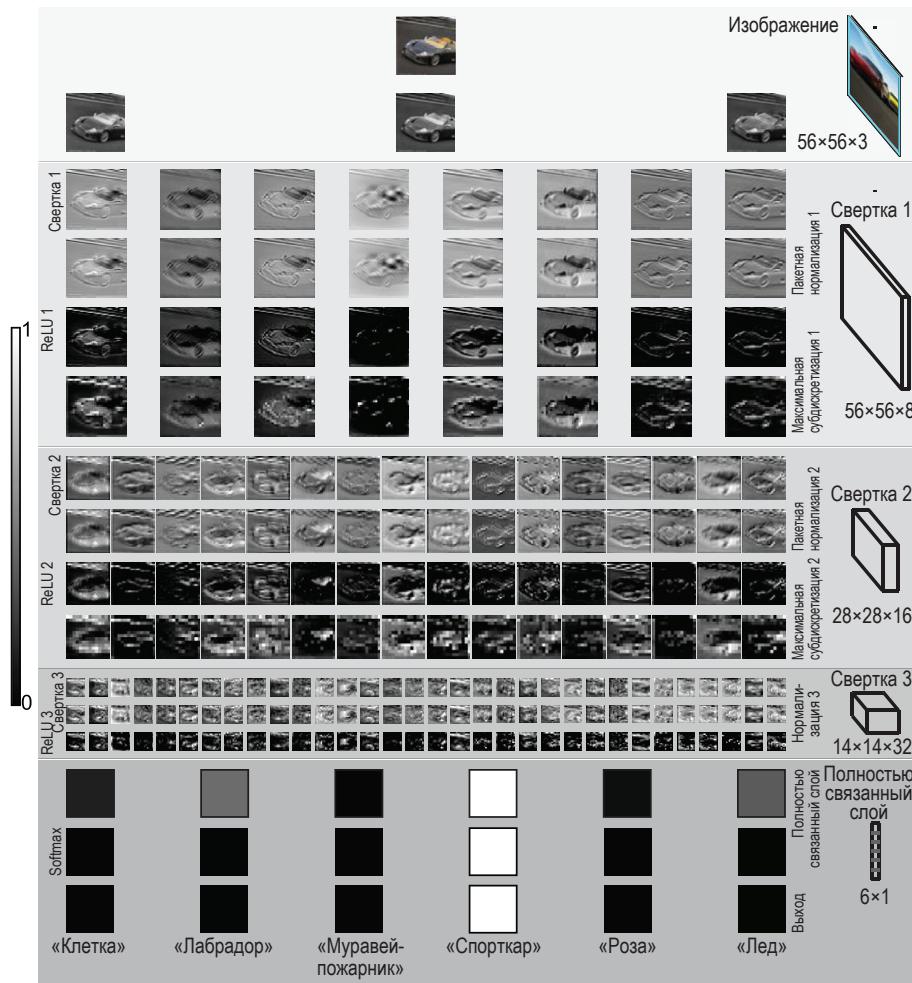
ставляет всего 0,21 % (то есть около одной ошибки в 500 изображений). Эта вычислительная модель включает в себя комбинацию нескольких глубоких сверточных нейронных сетей, стратегию, которая является общей для систем компьютерного зрения: создание множества экспертных систем и суммирование их прогнозов. Распознавание рукописных цифр – это пример задачи, в которой компьютеры достигли точности, сравнимой с той, что показывает человек, если не лучше.

## 8.11. ГЛУБОКАЯ СВЕРТОЧНАЯ НЕЙРОННАЯ СЕТЬ В ДЕЙСТВИИ

Далее мы шаг за шагом проиллюстрируем, как генерируются все выходные данные глубокой сверточной нейронной сети. Мы хотим показать активацию всех блоков. Поскольку современные сети обычно имеют огромное количество блоков, в целях иллюстрации мы рассмотрим упрощенную сеть (рис. 8.5). Эта сеть принимает в качестве входных данных цветное изображение размером  $56 \times 56$  пикселей в 3 цветах. Сеть имеет три сверточных слоя и обучена классифицировать шесть категорий изображений из ImageNet (рис. 8.4): изображения биологических клеток, собак-лабрадоров, муравьев-пожарников, спорткаров, роз и льда. Эти шесть категорий были выбраны случайным образом в иллюстративных целях. Выходной слой содержит шесть значений, по одному для каждой категории. Сеть обучается методом обратного распространения ошибки (раздел 8.6) с использованием случайно выбранной обучающей выборки. Примерно после 1200 итераций сеть достигает точности 84 % на обучающем наборе и 76 % на тестовом наборе (где вероятность случайно правильного выбора – одна шестая, или 16,7 %). Как отмечалось при обсуждении перекрестной проверки (раздел 8.8), разница в результате между точностью обучения и точностью тестирования обычно отражает чрезмерную подгонку сети к конкретным примерам изображений, наблюдавшихся во время обучения.

На рис. 8.5 показана активация каждого блока в сети для одного конкретного изображения спорткара. Для визуализации каждого блока значения активации были нормализованы, чтобы оставаться в пределах от 0 до 1, и они показаны как значения в градациях серого (см. цветовую шкалу слева от рисунка). Изображение (вверху) состоит из трех каналов: R, G и B (которые в этом примере выглядят очень похожими, потому что изображение в основном серое). На первом этапе (свертка 1) есть восемь фильтров, каждый размером  $3 \times 3$  пикселя (как фильтры, показанные на рис. 7.6). Количество фильтров в каждом слое – одно из многих архитектурных решений, которые мы должны принять при построении модели; здесь число было произвольно установлено равным восьми для первого слоя в иллюстративных целях. Мы все еще можем видеть подобие входного изображения при активации первого слоя, где каждый из различных фильтров подчеркивает определенные особенности. Свернутое изображение проходит этап пакетной нормализации, за которым следуют выпрямляющие линейные блоки (рис. 7.2). Пакет-

ная нормализация – это метод, который улучшает скорость, эффективность и стабильность нейронных сетей за счет нормализации входных данных до заданного уровня. Наконец, выходные данные проходят этап максимальной субдискретизации, который уменьшает размер с  $56 \times 56$  до  $28 \times 28$ . Слой Свертка 1 состоит из  $56 \times 56 \times 8 = 25\,088$  блоков, а максимальный уровень пула 1 состоит из  $28 \times 28 \times 8 = 6272$  блоков.



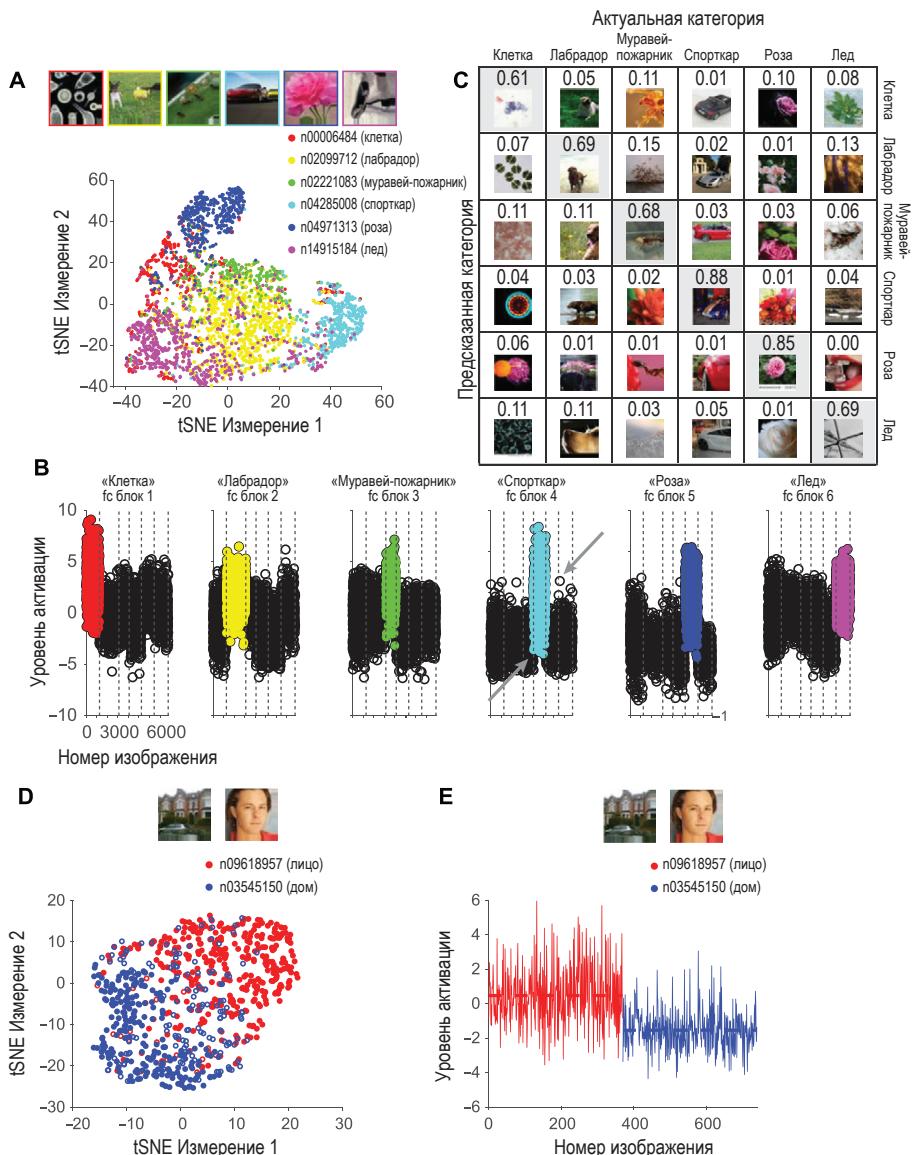
**Рис. 8.5 ♦** Пример выходных данных глубокой сверточной нейронной сети. Нейронная сеть, в которой входное изображение проходит через три слоя с операцией свертки и выдает шесть выходных значений вероятности (подробности см. в тексте). Размер каждого сверточного фильтра представлен справа. Сеть обучена классифицировать изображения из шести групп, показанных внизу. Этот рисунок демонстрирует активацию каждого блока в сети в ответ на изображение спортивного автомобиля, показанное вверху (см. линейку масштаба слева). Из шести блоков вывода (внизу) блок «спорткар» показывает максимальную активацию; следовательно, сеть правильно определяет название в этой задаче

Второй и третий слои проходят те же шаги с 16 и 32 фильтрами соответственно, все размером  $3 \times 3$  пикселя. Активации показаны в виде «картинки» в каждом квадрате на рис. 8.5 путем объединения всех блоков на данном шаге и для данного фильтра. По мере того как мы проводим вычисления от слоя Свертка 1 к ReLU 3, результирующие картинки все меньше и меньше похожи на исходные. Вся цель сети состоит не в том, чтобы создать изображение, похожее на исходное, а в том, чтобы извлечь адекватные признаки, которые могут решить задачу классификации.

Значения ReLU 3 передаются на полносвязный (fc) слой, состоящий из шести выходов, что отражает вероятность того, что название изображения соответствует каждой из шести возможных категорий. Для изображения в этом примере блок 4 в данном слое показывает максимальную активацию, которая соответствует спортивному автомобилю. Однако другие блоки в этом слое по-прежнему показывают ненулевые вероятности. Результирующие значения  $z_1, \dots, z_6$  передаются через функцию *softmax* (максимальная субдискретизация), ####, операцию, которая преобразует значения в вероятности, которые увеличивают значение до 1, а затем через пороговое значение, чтобы получить окончательное значение «победитель получает все». Эти окончательные значения активации указывают наиболее вероятное название. В этом случае сеть правильно определяет, что под изображением стоит название «спорткар».

Таким образом, активацию блоков полносвязных слоев (fc) после функции максимальной субдискретизации (*softmax*) можно интерпретировать как вероятность для каждого из шести названий. Трудно визуализировать паттерны активации в ответ на каждое из  $\sim 8000$  изображений в этом шестимерном пространстве. Для представления матрицы активаций  $8000 \times 6$  существует множество методов уменьшения размерности, включая анализ главных компонентов, метод главных компонент, многомерное масштабирование и другие. На рис. 8.6А мы использовали метод уменьшения размерности с причудливым названием *t-распределенное стохастическое сетевое вложение*, или сокращенно *tSNE*<sup>1</sup>. Описание математики этого метода выходит за рамки данной главы (см. работы Ван дер Маатена и Хинтона, и, конечно же, для этого есть страница в Википедии). В *tSNE*, как и в других методах уменьшения размерности, похожие изображения (в смысле их расстояния в шестимерном fc-пространстве) представлены ближайшими точками, а различные изображения представлены точками, которые находятся дальше. Каждая точка представляет одно изображение, и точки окрашены в соответствии с фактическими названиями (их цвет не является частью *tSNE*, которая является полностью неконтролируемой процедурой и не использует никаких названий). Изображения с одинаковым названием (одного цвета) обычно группируются. В конечном итоге точность классификации зависит от сравнения между результатом выходного слоя (рис. 8.5) и истинными названиями, но разделение в этом двумерном пространстве *tSNE* дает интуитивно понятный намек на то, что можно адекватно разделить изображения с разными названиями.

<sup>1</sup> Или *стохастическое вложение соседей с t-распределением*. – Прим. ред.



**Рис. 8.6 ♦ Изучение активности выходного слоя в глубокой сверточной нейронной сети.** (A) Визуализация активности шести блоков fc из модели, показанной на рис. 8.5, в ответ на каждое изображение после уменьшения размерности до двух (двумерное представление) с использованием техники, называемой tSNE (см. текст). Каждый цвет обозначает отдельную категорию объекта. (B) Активация каждого из шести полносвязных (fc) блоков в ответ на каждое изображение. Вертикальные пунктирные линии разделяют категории объектов; цвет подчеркивает изображения из той категории, которую должен классифицировать блок fc. (C) Матрица ошибок показывает долю изображений, идентифицированных соответственно заданной категорией названия. (D)–(E) Несмотря на то что эта сеть никогда не была обучена распознавать человеческие лица или дома, мы можем обнаружить блоки, которые по-разному реагируют на лица и дома, и можем использовать эту же сеть для обнаружения лиц или домов. Измененное из работы Креймана, 2018

Примечательно, что нейронная сеть, представленная здесь, относительно проста, чтобы более наглядно показать активацию всех блоков. Современная сеть обеспечит более высокую точность, а кластеры будут намного лучше разделены.

Можно также использовать ту же технику уменьшения размерности для визуализации активации блоков в каждом из других слоев. Что особенно интересно в слое  $fc$ , так это то, что его значения активации напрямую коррелируют с выходными данными сети. Другими словами, с точки зрения поведения этой сети мы могли бы назвать блок  $fc$  4 «спорткаром». Мы также можем построить график активации каждого из шести блоков  $fc$  для каждого изображения (рис. 8.6В). Как и ожидалось, в среднем блок «спорткар» (блок  $fc$  4) показывает более высокую активацию для изображений с названием «спорткар» (выделено голубым цветом на рис. 8.6В). Однако есть некоторые изображения, содержащие спортивные автомобили, которые вызывают низкую активацию в этом блоке (например, серая стрелка внизу), и есть изображения, которые не содержат спорткаров, но все же вызывают высокую активацию в этом блоке (например, серая стрелка вверху). Другими словами, блок  $fc$  4 «спорткар» может не активироваться многими изображениями спортивных автомобилей, но может показывать высокую активацию другими изображениями, которые не содержат спортивный автомобиль. В разделе 9.9 мы познакомимся с методами, которые можно использовать для описания того, какие типы изображений вызывают высокую активацию блоков в сети, а также для нейронов в головном мозге.

Выходная мощность сети зависит от максимального значения  $fc$  по всем шести блокам. Если истинное название для данного изображения – «спорткар», но активация в блоке  $fc$  4 «спорткара» ниже, чем активация в другом блоке  $fc$ , тогда сеть совершила ошибку, выбрав другое название. И наоборот, если истинное название для данного изображения не является «спорткаром», но блок  $fc$  4 показывает максимальную активацию, тогда изображение будет ошибочно опознано как спорткар. На рис. 8.6С показано, как часто возникают эти ошибки, в виде матрицы ошибок. Столбцы указывают фактические названия категорий, а строки указывают предсказанные названия категорий. Все записи по диагонали – правильные ответы. Например, в 88 % случаев, когда изображение содержало спортивный автомобиль, сеть правильно назвала его спорткаром (строка 4, столбец 4 матрицы). Иногда на изображении был спорткар, но в сети он был помечен как «лед»; это произошло для 5 % изображений спортивных автомобилей в тестовой выборке (строка 6, столбец 4 матрицы). В других случаях на изображении был изображен лед, но сеть решила, что это «спорткар»; это произошло для 4 % изображений льда (строка 4, столбец 6 матрицы).

Шесть категорий объектов, представленных на рис. 8.5 и 8.6, были выбраны случайным образом. Блоки  $fc$  также активируются изображениями, которых не показывали сети раньше. Одним из преимуществ работы с моделями, обрабатывающими изображения, является то, что мы можем напрямую количественно оценить реакцию сети на любое изображение. Например, мы можем спросить, будут ли те же блоки  $fc$  той же сети активироваться изображениями домов или человеческих лиц. Важно отметить, что мы не переобучаем сеть

с этими новыми изображениями. Веса фиксированы, и мы просто отслеживаем активацию блоков fc. Сеть никогда раньше не видела дома или лица; однако блоки в сети по-прежнему активируются этими изображениями (точно так же, как нейроны в нашей зрительной коре были бы активированы, если бы нам показали единорога, даже если мы никогда его раньше не видели).

Опять же, мы используем tSNE для вычисления активации всех шести блоков fc для всех изображений домов и лиц (рис. 8.6D). Несмотря на то что сеть никогда не обучалась с этими изображениями, она все же может довольно хорошо различать их: сеть достигла точности в 86 % в различении лиц и домов (где вероятность составляет 50 %). Блок fc, который показал наиболее четкое разделение между лицами и домами, был блоком «спорткар» (блок fc4, рис. 8.6E). Этот отряд показал более сильную активацию по лицам ( $0,47 \pm 1,72$ ) по сравнению с домами ( $-1,54 \pm 1,18$ ). Если бы исследователь проводил исследование с помощью этой сети и показывал бы только изображения лиц и домов, исследователь, возможно, назвал бы данный блок «лицевым блоком». Тем не менее активация этого блока fc для спорткаров была  $4,59 \pm 2,27$ . Таким образом, простого показа набора случайных изображений недостаточно для интерпретации активации блоков в сети (см. также обсуждение в разделах 6.4 и 9.9).

Упражнение по оценке активности сети для изображений, которые она никогда раньше не видела, позволяет нам сделать еще один вывод. Без переобучения сеть может решать задачи визуальной классификации, которым она не была обучена. Словарь признаков и вычислений, полученный сетью при попытке разделить шесть произвольных классов случайных объектов, достаточно богат, чтобы можно было различить другие категории изображений. Можно даже продолжить и, начав с сети, предварительно обученной различать эти шесть категорий, переобучить сеть для новой задачи. Такая переподготовка – еще один пример трансферного обучения, представленного в разделе 8.6: сначала обучение одной задаче, а затем использование предварительно обученной сети в качестве начального условия для освоения новой задачи.

## 8.12. ОШИБАТЬСЯ СВОЙСТВЕННО И ЛЮДЯМ, И АЛГОРИТМАМ

В том типе задач классификации изображений, который мы обсуждали, истинное название устанавливается людьми. Вернемся к базе рукописных цифр MNIST; эти цифры имеют названия, присвоенные людьми. Только сравнивая с человеческим поведением, мы можем определить, допускает ли компьютерная модель зрения ошибки.

На рис. 8.6В и 8.6С мы показали, как вычислительные модели допускают ошибки, снова ссылаясь на названия, предоставленные людьми. Точно так же на рис. 8.7 показана матрица ошибок, сравнивающая выходные данные двухуровневой нейронной сети и людей для данных базы MNIST.



**Рис. 8.7** ♦ Матрица ошибок для базы данных MNIST. Значения, выделенные серым цветом в записи  $(i, j)$ , указывают процент тестовых изображений MNIST, которые принадлежат названию категории  $j$  и были классифицированы как название категории  $i$  двухуровневой нейронной сетью. Диагональные записи соответствуют правильной классификации, а недиагональные записи являются ошибками. Примеры ошибок показаны для каждой возможной комбинации.

В квадратах выделен наиболее вероятный тип ошибки для каждого числа

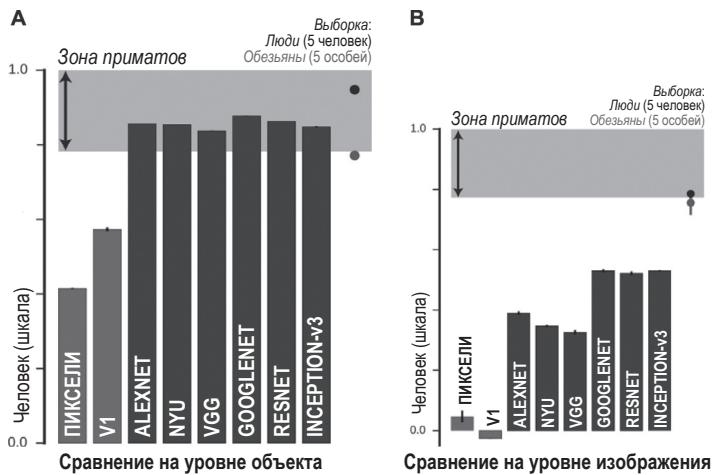
Средняя эффективность этой сети составила 95 %. Конечно, есть множество лучших алгоритмов, но здесь мы намеренно используем эту двухуровневую сеть, чтобы показать ошибки, допущенные алгоритмом. Все проценты по диагонали соответствуют случаям, когда нейронная сеть правильно классифицировала изображения. Не все цифры были классифицированы одинаково хорошо. Точность для числа 7 составила 89,3 %, то есть число семь было легче спутать с другими числами, возможно, отражая неоднородность того, как люди рисуют семерки. Номер 1 имел самую высокую точность – 98,9 %. В тестовом наборе из 10 000 изображений, показанном на рис. 8.7, некоторых ошибок не произошло; например, сеть никогда не принимала 1 за 0 или 0 за 1. Самые сложные для распознавания цифры в каждом столбце выделены рамкой. В худшем случае число 7 было перепутано с числом 9, что произошло в 5,1 % случаев, когда была представлена семерка.

Глядя на некоторые из примеров ошибок на рис. 8.7, интуитивно понятно, как модель могла неверно интерпретировать некоторые из этих цифр. Например, число 6 в нижнем ряду действительно похоже на 0. Ошибаться могут не только алгоритмы, но и люди. Что в этом контексте значило бы для людей совершать ошибки? Мы могли бы рассмотреть поведенческий эксперимент, в котором группу испытуемых просят классифицировать эти изображения. Конечно, этот набор испытуемых не должен зависеть от первоначального набора людей, которые изначально установили истинные названия. Таким образом, мы можем оценить степень вариабельности зри-

тельного распознавания между людьми тестовой группы. Таким же образом мы можем сравнивать эффективность решения задачи распознавания людей и животных. Например, обезьяну можно обучить поведенческому различению набора изображений, а затем мы можем напрямую сравнивать названия, данные человеком и обезьяной одному и тому же изображению в серии изображений.

Раджмалингем и его коллеги пошли по этому пути, чтобы получить сравнительные характеристики эффективности зрительного распознавания между обезьянами, людьми и вычислительными моделями (рис. 8.8). Авторы рассмотрели 24 объекта – в том числе «слона», «шорты», «гаечный ключ» и другие. Повернутые и масштабированные фотографии этих объектов включались в набор естественных изображений. Людям или обезьянам показывали тестовое изображение в течение 100 мс, после чего шел экран выбора, содержащий канонический образ тестового объекта и канонический образ одного из других 23 объектов. Подопытные должны были указать, какое из двух изображений соответствовало показанному только что. На основе поведенческих измерений авторы вычислили различимость каждого объекта по сравнению с любым другим объектом (сравнения на уровне объекта, рис. 8.8A), усреднив все изображения (все повороты и фоны). Они также вычислили различимость каждого изображения по сравнению с любым другим изображением, где изображение представляет собой конкретную комбинацию объекта, поворота и фона (сравнение на уровне изображения, рис. 8.8B). Ответы людей вполне коррелировали друг с другом, что было оценено путем исключения некоторых из них и сравнения их результатов с остальными (черная точка на рис. 8.8A, согласованность  $> 0,9$ ). На уровне отдельных изображений больше вариабельности, что демонстрирует степень согласованности между людьми экспериментальной группы  $\sim 0,8$  на рис. 8.8B. Обезьяны также преуспели в этой задаче и показали согласованность с людьми  $\sim 0,8$  на рис. 8.8A и 8.8B (серые точки).

Далее авторы рассмотрели несколько вычислительных моделей, протестированных на точно таких же изображениях. Они рассмотрели шесть глубоких сверточных нейронных сетей: AlexNet, NYU, VGG, GoogleNet, ResNet и Inception-v3. Ученые, занимающиеся компьютерным зрением, распространяли свой биологический пыл на названия своих моделей. Названия на рис. 8.8 соответствуют некоторым популярным моделям компьютерного зрения, и мы не будем вдаваться в подробности различия их архитектур (все эти архитектуры являются расширениями и вариациями той, что показана на рис. 8.2). Достаточно сказать, что все эти модели содержат от семи слоев (AlexNet) до более чем 150 уровней (ResNet), что они были обучены на базе данных ImageNet (рис. 8.4) и успешно справились с задачами распознавания. Авторы также рассмотрели модель, в которой использовалась только информация на уровне пикселей, и модель, имитирующую вычисления, выполняемые в первичной зрительной коре (V1). Отображение показателей эффективности в модели попиксельного сравнения – это нижняя из всех возможных отметок, которую следует включить в эти сравнения: в конце концов, если можно решить данную задачу, используя только значения яркости пикселей, зачем возиться с более сложными моделями?



На уровне объекта (рис. 8.8А) все модели глубокой сверточной нейронной сети (но не попиксельная модель и V1) показали замечательную степень согласованности с человеческим поведением. Эти модели были немного больше похожи на людей, чем на обезьян, и разница была несколько больше разброса результатов разных людей. Напротив, для сравнений на уровне изображений (рис. 8.8В), хотя все глубинные модели лучше соответствуют человеческой реакции, чем модели на основе сравнения пикселей или модели V1, все они терпят неудачу по сравнению с человеком. Хотя и люди, и модели могут неплохо, но не идеально, справиться с этой задачей, характер ошибок людей и моделей различен при рассмотрении каждого изображения в отдельности. При рассмотрении всего разнообразия положений, размеров и поворотов объектов обнаружилась большая согласованность между разными людьми или между обезьянами и людьми, чем между моделями и людьми.

## 8.13. ПРОГНОЗИРОВАНИЕ ДВИЖЕНИЙ ГЛАЗ

Мы можем дополнительно ограничить вычислительные модели, выйдя за рамки оценки того, можем ли мы сопоставить характер ошибок в задачах классификации изображений. Одной из важных особенностей зрительно-моторной координации человека<sup>1</sup> является быстрая последовательность

<sup>1</sup> Есть еще термин «координация глаз–рука», но здесь он некорректен – во многих экспериментах рука не задействована вообще. – Прим. ред.

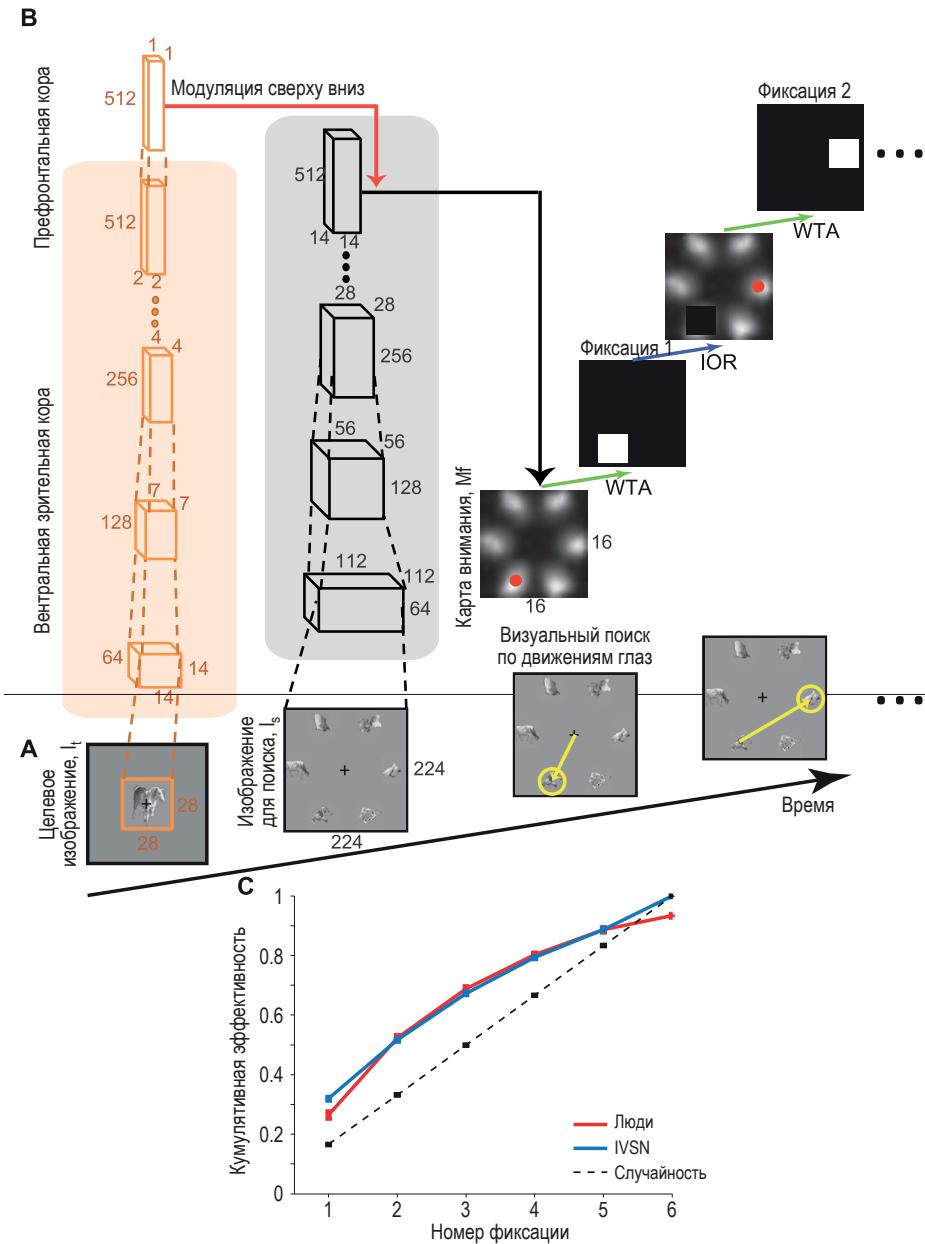
движений глаз, которые происходят примерно три раза в секунду при обычных условиях просмотра (раздел 2.4). Типы моделей глубоких нейронных сетей, которые мы описали до сих пор, не имеют ничего общего с движениями глаз, но мы можем модифицировать модели, чтобы предсказать, какие особенности изображения являются заметными и могут управлять движениями глаз.

Одно из самых важных особенностей изображения – резкие изменения движений. Если человек в поле нашего зрения вдруг начинает бежать, наши глаза немедленно обратятся на этого человека. В то время как такие временные изменения дают сильные сигналы, притягивающие наше внимание, пространственные изменения в изображении, включая контраст, изменения цвета и изменения текстуры, также вызывают смещение пространственного внимания через движение глаз. Например, если мы смотрим на изображение, на котором все серое, кроме желтой машины, эта машина обязательно привлечет наше внимание. Эти признаки «заметности» широко изучались в психофизической литературе. Во многих случаях эти признаки были заново открыты людьми, снимающими фильмы, а также в рекламной индустрии.

Цели задания влияют и на движения глаз. Например, если мы ищем свою машину на стоянке, наш взгляд может быть привлечен к местам, где есть машины, а не к зданиям или небу, особенно к другим машинам того же цвета и формы. Мы можем даже не обращать внимания на другие важные признаки, такие как движение (по всей вероятности, наша машина не движется). Целевые функции берут верх и играют важную роль в определении нашей пространственно-временной последовательности движений глаз.

Рассмотрим пример задачи зрительного поиска (рис. 8.9А). Подопытным предоставляется целевое изображение, содержащее искомый объект (в данном примере лошадь). Затем подопытным предоставляется изображение для поиска, содержащее шесть объектов, находящихся вокруг точки фиксации. Один из этих шести объектов – лошадь, а другие пять – отвлекающие стимулы. В принципе, можно решить задачу поиска путем простого перебора, сопоставления образов, исчерпывающего перемещения шаблона того же размера, что и целевой объект, по всему изображению до тех пор, пока не будет найдено идеальное совпадение. Чтобы избежать такого решения, объекты на изображении для поиска отображаются в другом масштабе, а также случайным образом поворачиваются. Более того, целевая лошадь фактически является другим экземпляром из той же категории объектов (то есть в данном случае другой лошадью).

Нас интересует алгоритм зрительного поиска, который может эффективно находить целевой объект (то есть без полного сканирования всего изображения), делать это выборочно (чтобы отличить целевой объект от отвлекающих) и инвариантно по отношению к изменениям в масштабе, повороте и даже к другому экземпляру целевого объекта. Кроме того, мы также хотим проверить, может ли алгоритм улавливать фундаментальные особенности того, как люди двигают глазами, чтобы решить задачу поиска.



**Рис. 8.9** ♦ Нейронная сеть, прогнозирующая движения глаз во время зрительного поиска. (A) Задача зрительного поиска, при которой подопытным или модели нужно двигать глазами, чтобы найти целевой объект (слева) на изображении. (B) Схематическое изображение модели инвариантной визуальной поисковой сети (IVSN) (подробности см. в тексте). (C) Совокупная эффективность людей (красный) и IVSN (синий) в задаче (пунктирная линия указывает на вероятность). По материалам Чжан и др., 2018

На рис. 8.9В показана схема такой вычислительной модели. В основе модели лежит глубокая сверточная нейронная сеть, иносказательно называемая здесь «центральной зрительной корой», которая извлекает визуальные признаки из целевых и поисковых изображений. В данном случае авторы использовали нейронную сеть VGG, которая является одной из сетей, также протестированных на рис. 8.8. Эта сверточная нейронная сеть была предварительно обучена с помощью ImageNet (рис. 8.4), чтобы иметь обширный словарь визуальных характеристик естественных изображений.

Модель должна хранить в памяти информацию о целевом объекте, чтобы иметь возможность искать его в поисковом изображении. Нам также необходимо решить, какие особенности целевого изображения модель должна хранить в памяти. Должна ли модель сохранять все объекты на каждом уровне сети? Сохранение всех признаков будет соответствовать полному многоуровневому представлению целевого объекта с точки зрения ответов уровня 1, ответов уровня 2 и так далее. В качестве альтернативы мы могли бы сохранить только ответы уровня 1. Проблема с сохранением объектов исключительно на уровне 1 заключается в том, что эти низкоуровневые функции слишком чувствительны к метрическим свойствам изображения и не идеальны для поиска объектов, которые были масштабированы и повернуты. Напротив, мы можем сохранять только ответы верхнего уровня; этот подход проиллюстрирован на рис. 8.9В. Для упрощения модель идеально и бесконечно хранит все особенности изображения в верхнем слое модели вентральной зрительной коры. На самом деле этот тип памяти, часто называемый рабочей памятью, быстро распадается в течение нескольких секунд. Эта часть модели называется «префронтальной корой», потому что именно в ней исследователи обнаружили нейроны, которые играют важную роль в хранении информации при решении задач, связанных с рабочей памятью.

Информация о целевом объекте используется для модуляции активации модели в ответ на поисковое изображение. Эта модуляция сверху вниз выполняется параллельно по всему изображению. Результатом является карта активации, которая, по существу, описывает, насколько каждая часть изображения похожа на цель, где сходство определяется высокоуровневыми функциями, хранящимися в префронтальной коре, и где пространственное разрешение зависит от того, какой уровень иерархии модулируется. Этот паттерн активации называется картой внимания. В данном примере эта карта имеет разрешение  $16 \times 16$  областей. Механизм «победитель получает все» (WTA) выбирает максимум на карте внимания. Положение, соответствующее ему, становится положением первой «фиксации» модели.

Если в этом месте цель обнаружена, поиск заканчивается. Если цель не найдена, модель возвращается к карте внимания. Поскольку модель детерминирована, если мы снова выберем максимум, она опять и опять будет фиксироваться в одном и том же месте. Чтобы избежать этой проблемы, модель использует механизм подавления бесконечного возврата (IOR), означающий, что модель никогда не перемещает точку своей фиксации в место,

которое она уже выбирала ранее. Механизм «победитель получает все» выбирает следующий максимум на карте внимания для фиксации. Таким образом, добавив несколько вычислительных шагов, мы можем использовать глубокую сверточную нейронную сеть для выполнения последовательности движений глаз и определения местоположения целевых объектов на поисковом изображении.

Работает ли это? Во-первых, давайте рассмотрим поведение человека в этой задаче (рис. 8.9С). Поскольку имеется шесть объектов (одна цель и пять отвлекающих стимулов), вероятность того, что цель будет найдена при первой фиксации, равна 1/6, при второй фиксации вероятность нахождения цели составляет 2/6 и так далее. Люди добиваются лучшего результата, превышающего случайность. Они не могут найти цель за одну фиксацию, но могут сделать это с вероятностью примерно 1/3. Эти цифры не слишком критичны; точные вероятности, скорее всего, зависят от множества факторов, таких как размер объектов, их расстояние от точки фиксации, степень сходства отвлекающих факторов с целью, количество отвлекающих факторов и насколько отлична цель в поисковом изображении от того, что было показано в качестве целевого. Независимо от количественных показателей, люди могут эффективно находить целевые объекты. Любопытно, что за шесть фиксаций у людей результат чуть ниже 100 %, что меньше простой вероятности, потому что у людей нет запрета на бесконечное возвращение, как реализовано в модели. Люди – упрямые существа, и иногда они переводят взгляд обратно в то же место, даже если в этом месте нет цели.

Затем давайте рассмотрим модель, изображенную на рис. 8.9, протестированную на тех же изображениях и задачах, которые использовались в психофизическом эксперименте над людьми. Модель работает на удивление хорошо: она может локализовать целевые объекты более эффективно, чем последовательное или случайное сканирование всего изображения. И люди, и модель могут находить цель, демонстрируя инвариантность к некоторым ее характеристикам, учитывая особые условия постановки эксперимента, заключающиеся в использовании различных образцов из одних и тех же категорий, а также масштабировании и повороте объектов. Следует отметить, что модель никогда раньше не видела эти конкретные объекты, и тем не менее она могла находить объекты даже без какого-либо предметно-ориентированного обучения. Сильное сходство между человеческим поведением и поведением модели частично случайно. Эксперименты с другими изображениями, включая поиск объектов на естественных изображениях или знаменитый пример поиска Уолдо, показывают, что модель не полностью соответствует движениям человеческих глаз. Следует отметить, что модель не была обучена соответствовать человеческому поведению; т. е. в этой процедуре нет согласования данных. Модель, обученная для классификации объектов, может быть адаптирована для задачи зрительного поиска и объяснения движений человеческих глаз без настройки каких-либо параметров.

## 8.14. ПРОГНОЗИРОВАНИЕ ЧАСТОТЫ ВОЗБУЖДЕНИЯ НЕЙРОНОВ

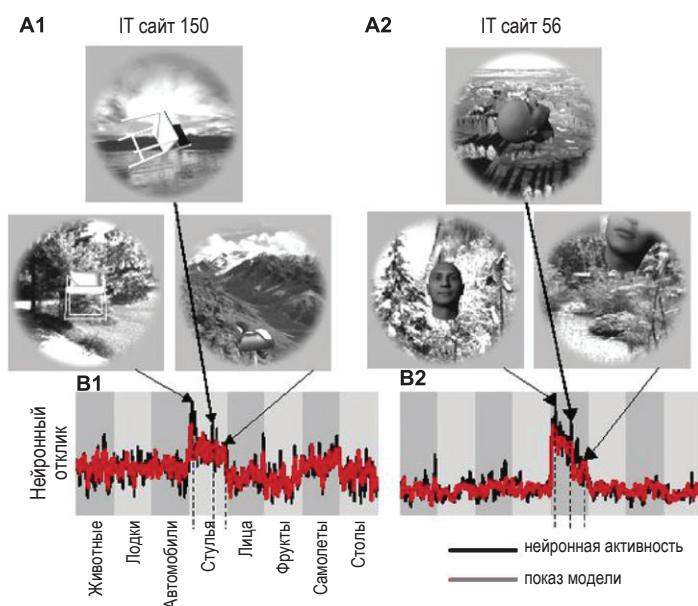
Предыдущие два раздела демонстрируют, что модели глубокой сверточной нейронной сети, обученные для задач классификации объектов, могут обеспечить разумное приближение первого порядка к зрительно-моторной координации человека и обезьяны как с точки зрения таких задач, как категоризация объектов, так и с точки зрения других задач, таких как совершение определенных движений глазами при визуальном поиске. Тем не менее современные модели не обеспечивают идеального соответствия человеческому поведению. Ранее мы указали несколько случаев, когда модели могут качественно воспроизвести определенные особенности зрительно-моторной координации (например, выполнение классификации на уровне объекта с помощью движений глаз во время зрительного поиска объектов в представленном массиве изображений). Однако в других аспектах есть достаточно возможностей для улучшения (например, эффективности классификации объектов на уровне изображения, движений глазами во время зрительного поиска в естественных изображениях). Мы еще вернемся к некоторым положительным провалам современных моделей в разделах 9.11 и 9.12.

Теперь обратим внимание на то, что происходит внутри мозга, и зададим вопрос, могут ли существующие модели уловить внутренние механизмы функции зрения человека. Даже если бы у нас была модель, которая исключительно хорошо объясняла зрительно-моторную координацию, это вовсе не означало бы, что внутренняя работа мозга и модели одинаковы. Мозг и модели могут решать одну и ту же задачу совершенно разными способами. Понимание различий между работой мозга и вычислительными механизмами моделей может вдохновить исследователей на разработку более совершенных моделей.

Вопрос о том, схожи ли внутреннее устройство мозга и модели, требует дальнейшего уточнения. Если мы опустимся до уровня отдельных молекул, «аппаратные средства» у мозга и модели будут совсем разными. Чтобы оценить, отражает ли модель особенности функции нейронной цепи – и, следовательно, может ли модель помочь нам лучше расшифровать нейронные механизмы обработки зрительного сигнала, – нам необходимо определить, какие особенности нейронной функции мы хотим понять и объяснить. Естественный вопрос – попытаться объяснить частоту срабатывания нейронов (спайковую частоту). Как обсуждалось в главах 2, 5 и 6, нейронные спайки<sup>1</sup> представляют собой золотой стандарт для изучения нейронных функций и основной механизм, с помощью которого нейроны могут посыпать сигналы на большие расстояния. Поэтому мы рассматриваем, может ли модель предсказать количество импульсов в единицу времени, испускаемых нейроном в ответ на представленное изображение.

<sup>1</sup> Резкие максимумы регистрируемого потенциала. – Прим. ред.

Пример такого анализа показан на рис. 8.10. Исследователи показали обезьяне обширную коллекцию изображений, ведя запись активности нейронов в нижней височной коре (ITC). Подобно исследованию, описанному на рис. 8.8, обезьянам были представлены изображения животных, лодок, лиц и пяти других категорий объектов. Объекты были помещены на естественный фон (примеры изображений показаны на рис. 8.10A). Как описано в разделе 6.2, нейроны ITC демонстрировали избирательную реакцию на различные типы изображений. Например, запись нейронной активности на рис. 8.10A1/B1 (черный сигнал на регистрации) демонстрирует более высокие отклики на изображения, содержащие стулья, и в меньшей степени – самолеты (где «отклики» определены здесь как общее количество спайков в фиксированном окне от 70 до 170 мс после появления стимула). Как обсуждалось в гл. 2 и 5, нейронные реакции могут изменяться при повторном предъявлении одного и того же стимула; ответы нейронов, показанные на рис. 8.10B, представляют собой средние значения по десяткам повторений одного и того же стимула.



**Рис. 8.10 ♦** Вычислительные модели могут демонстрировать приблизительное соответствие частоты по сравнению с частотой срабатывания нейронов. **(A)** Примеры изображений, показанных обезьяне, при записи активности двух разных участков (участок 150 и участок 56) в нижней височной коре (ITC). **(B)** Нейронные ответы (черный) по сравнению с предсказанными ответами от модели глубокой сверточной нейронной сети (красный). Каждая координата по оси x соответствует одному из 1600 различных изображений, разделенных на восемь категорий объектов (животные, лодки и другие).

Измененное из Яминс и др., 2014

Те же изображения можно передать в глубокую сверточную нейронную сеть, извлекая значения активации в каждом слое, как показано на рис. 8.5. Затем мы можем сравнить активации модели с откликами нейронов. Один из способов сделать это сравнение – взять значения активации в данном слое и построить линейную карту на ответах данного нейрона. Эта процедура линейной подгонки имеет один свободный параметр для блоков нейронной сети, а количество уравнений равно количеству изображений. Некоторые изображения используются для подгонки к линейной карте, а остальные используются для проверки того, насколько хорошо модель может аппроксимировать отклики нейронов на новые изображения (красный сигнал на рис. 8.10B1/B2). Корреляция между прогнозируемыми и фактическими ответами нейронов лучше при использовании модельных активаций из более высоких уровней, чем из ранних уровней глубоких сверточных сетей, это предполагает, что для объяснения активности нейронов в ИС могут потребоваться более сложные функции. Эти модели обычно дают более 50 % дисперсии нейронных ответов. Таким образом, несмотря на то что модели нейронных сетей далеки от сложности биологических тканей, активация блоков обеспечивает хорошее начальное приближение для предсказания нейронных ответов.

## 8.15. ВСЕ МОДЕЛИ НЕТОЧНЫ, НО НЕКОТОРЫЕ ПОЛЕЗНЫ

Современные глубокие сверточные нейронные сети привлекательны, потому что они выполняют многие из требований, сформулированных в разделе 8.1. Кроме того, в разделе 9.5 мы покажем, что эти модели успешно справились с широким спектром задач обработки изображений реального мира. Как обсуждалось в предыдущих трех разделах, эти модели также обеспечивают неточное, но вполне разумное приближение первого порядка к визуальной нейрофизиологии и зрительно-моторной координации людей и приматов.

Любопытно, что глубокие сверточные нейронные сети могут улавливать особенности поведения и физиологии, несмотря на то что они абстрагированы от реальных нейронных цепей. Как обсуждалось в разделах 7.3 и 7.4, обоснованное предположение о правильном уровне абстрагирования при моделировании нейронных цепей имеет важное значение для прогресса в исследованиях. Биологи, изучающие глубокую сверточную нейронную сеть, потрясены отсутствием множества реальных элементов, присутствующих в нервной ткани. Если упомянуть лишь некоторые из них, от более крупных до более мелких, зрительная система характеризуется завораживающе сложным набором взаимосвязей (рис. 1.5), большинство из которых отсутствуют в современных математических моделях. Мы также знаем, что существует много разных типов нейронов, включая, по крайней мере, десятки, если не сотни различных типов интернейронов в мозге, тогда как современные модели имеют, по сути, только один или несколько различных типов вычислительных блоков, в зависимости от того, как посчитать. Нейроны характери-

зуются сложным строением, и пространственно-временное распределение входных сигналов к разным дендритам может иметь значительное влияние на биофизику вычислений одиночных нейронов. Биохимики могут даже задаться вопросом о сложных вариантах экспрессии примерно 20 000 генов человеческого генома в различных типах нейронов.

В то время как биологи беспокоятся о многочисленных отличиях моделей от этих и других характеристик реальных нейронных цепей, на другом конце научного спектра психологи обеспокоены тем, что в современных вычислительных моделях такого абстрагирования слишком мало. Психологов шокирует отсутствие множества различных вариантов структур. Если представить лишь некоторые из них на все более и более высоких уровнях абстракции, эти модели не имеют четкого семантического смысла (см. раздел 6.8), помимо того что навязывается названиями, используемыми во время обучения. Кроме того, здравые предположения о визуальном мире, включая понятие «объектности» или понятие агентов, которые взаимодействуют друг с другом, явно не включены в эти модели. Психологи на основе самоанализа утверждают, что эти концепции имеют решающее значение для интерпретации визуального мира. Некоторые психологи также утверждают, что мы не можем понять визуальную обработку отдельно от того, как мы взаимодействуем с миром, и что зрение не может быть отделено от языка.

Обсуждение биологических и психологических компонентов, которые отсутствуют в существующих моделях, можно приблизительно сопоставить с тремя уровнями анализа Томазо Поджио и Дэвида Марра (раздел 1.9). Психологи склонны думать о высокоуровневых вычислениях, которые система может захотеть реализовать, а биологи склонны думать об оборудовании, необходимом для выполнения всех этих вычислений. Важной целью моделей является преодоление этих уровней анализа путем создания алгоритмов, которые могут реализовать эти вычисления, и путем связывания этих алгоритмов с реальным биологическим оборудованием. Вклад как биологов, так и психологов будет неоценимым для дальнейшего улучшения существующих вычислительных моделей.

## 8.16. ГОРИЗОНТАЛЬНЫЕ И НИСХОДЯЩИЕ СИГНАЛЫ ПРИ РАСПОЗНАВАНИИ ОБРАЗОВ

Одно из нескольких упрощений в глубоких сверточных нейронных сетях, заслуживающее дальнейшего изучения, – это отсутствие горизонтальных и нисходящих сигналов. Мы знаем, что в неокортике много горизонтальных и обратных проекций. Функции нисходящих связей менее изучены на нейрофизиологическом уровне, но нет недостатка в вычислительных моделях, иллюстрирующих богатый набор вычислений, которые могут возникнуть с включением такой связи.

Практически все вычислительные модели подразумевают, что нисходящие сигналы играют решающую роль во время обучения. Процедура обратного

распространения ошибок, описанная в разделе 8.6, требует передачи ошибок по сети сверху вниз. Однако чисто восходящие модели не используют нисходящие сигналы после обучения и во время визуальной обработки.

Некоторые модели использовали нисходящие соединения, чтобы направлять внимание к конкретным местоположениям или функциям в изображении (разделы 5.17, 8.13). Направление внимания к определенным частям изображения может значительно улучшить качество распознавания, облегчая проблемы, связанные с сегментацией изображения и его беспорядком. Модель, представленная на рис. 8.9, использует нисходящие сигналы для направления движений глаз, то есть прямого внимания, к конкретным местам, которые могут содержать искомый объект.

Горизонтальные и нисходящие сигналы также могут играть важную роль в распознавании частично закрытых объектов. Когда доступна только частичная информация об объекте, визуальная система должна иметь возможность выполнять завершение формы и интерпретировать изображение на основе полученного ранее опыта (раздел 3.5). Рекуррентные сети на основе аттракторов могут извлекать идентичность хранимых воспоминаний из частичной информации (раздел 7.6). Точно так же вычислительные модели объединили восходящие архитектуры с сетями аттракторов на вершине иерархии. Рекуррентная динамика, основанная на аттракторах, может помочь сделать выводы на основе частичной информации и, таким образом, распознать сильно закрытые объекты. В дополнение к горизонтальным связям нисходящие сигналы также могут играть важную роль во время завершения форм, предоставляя предварительно сохраненную информацию, которая влияет на сенсорные реакции снизу вверх.

Идея о том, что нисходящие сигналы могут нести релевантную для задачи распознавания априорную информацию, была поддержана несколькими предложениями, формулирующими визуальное распознавание как задачу байесовского вывода. Учитывая три уровня зрительного каскада (например, LGN, V1 и более высокие области, такие как V2) и обозначая активность на этих трех уровнях как  $x_0$ ,  $x_1$  и  $x_h$  соответственно, вероятность получения заданного образца ответа в V1 ( $x_1$ ) зависит как от сенсорных входов, так и от обратной связи из более высоких областей:

$$P(x_1|x_0) = \frac{P(x_0|x_1)P(x_1|x_h)}{P(x_0|x_h)}, \quad (8.12)$$

где  $P(x_1|x_h)$  представляет смещения обратной связи, передающие априорную информацию.

## 8.17. ПРЕДИКТИВНОЕ КОДИРОВАНИЕ

Интересная версия того, как нисходящие сигналы могут использоваться во время зрительного распознавания, была предложена Раджешем Рао и Даной Баллард, которые утверждали, что обратная связь обеспечивает *предиктивные* (предсказывающие) сигналы, тогда как восходящие сигналы передают

разницу между сенсорными входами и нисходящими предсказаниями. Например, рассмотрим явление ингибиования окружения (раздел 5.5): если мы представим оптимально ориентированную синусоидальную решетку в пределах рецептивного поля нейрона V1, спайковая частота увеличивается с размером стимула до определенного уровня; когда размер стимула превышает размер воспринимающего поля, возбуждение начинает уменьшаться с увеличением размера стимула. Согласно модели предиктивного кодирования, нейроны в более высоких областях с большими размерами рецептивного поля (например, V2) отправляют сигнал обратной связи, который может предсказать ответы, а нейрон V1 вычитает эти прогнозы из сенсорных входов, тем самым приводя к меньшему отклику для больших размеров стимула. В самом деле, заглушающая активность нейронов V2 приводит к уменьшению эффекта ингибиования окружения, то есть более сильным ответам в V1 на более сильные стимулы.

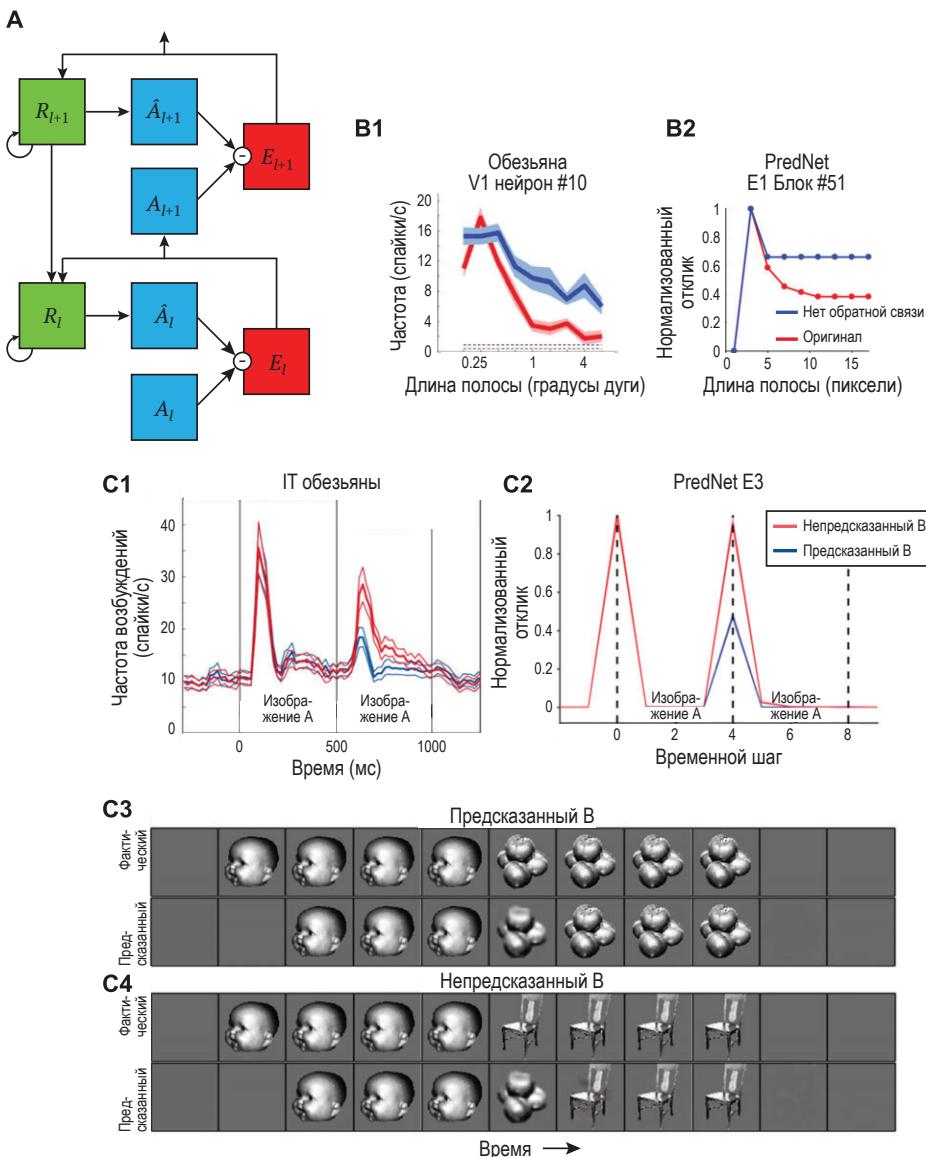
Прогнозы могут делаться не только в пространственной, но и во временной области. Постоянный визуальный стимул (при отсутствии каких-либо внешних или внутренних изменений, таких как движения головы или глаз) может быть предсказан. Согласно модели предиктивного кодирования, сигналы обратной связи приводят к уменьшению уровня первичного отклика. Действительно, затухающие реакции на постоянные раздражители являются нормой для всей зрительной системы (раздел 2.9).

Такие идеи предиктивного кодирования могут быть распространены на многослойную сеть. Архитектура модели, показанная на рис. 8.11A, состоит из нескольких уровней (только два из которых показаны на схеме); каждый уровень состоит из четырех типов блоков: входных блоков (синий,  $A_l$ ), рекуррентных блоков (зеленый,  $R_l$ ), блоков распространения ошибок (красный,  $E_l$ ) и блоков прогнозирования (синий,  $\hat{A}_l$ ). Если мы удалим или заглушим рекуррентные блоки, то путь от  $A_l$  к  $E_l$  до  $A_{l+1}$  и  $E_{l+1}$  будет стандартной глубокой сверточной нейронной сетью. Рекуррентные блоки выдают нисходящие сигналы. Если мы перейдем от более высоких уровней к более низким уровням, то создадим прогрессивно расширяющееся распространение, которое можно рассматривать как генеративную деконволюционную сеть, аналогичную другим алгоритмам для генерации изображений, которые обсуждаются в следующей главе (разделы 9.8 и 9.9). Исследователи называют эту сеть *PredNet*.

В этой сети блоки ошибок передают разницу между предсказаниями и входными данными на следующий уровень. Рекуррентные блоки принимают в качестве входных данных как ошибку в текущем слое, так и нисходящий сигнал следующего слоя. В отличие от стандартных глубоких сверточных нейронных сетей, подобных показанной на рис. 8.2, эта сеть демонстрирует богатую динамику: активация каждого блока изменяется с течением времени.

Пусть  $x_t$  представляет входной кадр в момент времени  $t$ . Для первого слоя  $A_l^t = x_t$ . Для следующих уровней входные блоки вычисляют свертку (плюс исправление и субдискретизация) по активации блоков ошибок в предыдущем слое:

$$A_l^t = \text{MAXPOOL}(\text{ReLU}(\text{CONV}(E_{l-1}^t))). \quad (8.13)$$



**Рис. 8.11 ♦ PredNet, архитектура глубокого предиктивного кодирования.** (A) Схематическое изображение двух уровней архитектуры PredNet. Блоки  $R$  отправляют предсказательные сигналы обратной связи на предыдущий уровень. Восходящие сигналы передают разницу между предсказанными сигналами и сигналами с предыдущего шага (см. в тексте). (B) Ингибиование окружения нейрона V1 обезьяны (B1) и блока уровня 1 PredNet (B2). Отклики на оптимальную полосу увеличивающейся длины в исходных условиях (красный) и при отсутствии обратной связи (синий). (C) Прогнозирование последовательности IT-нейронами обезьяны (C1) или блоком уровня 3 PredNet (C2), когда второй стимул является предсказуемым (синий) или непредсказуемым (красный) (C3).

Измененное из Лоттер и др., 2018

Рекуррентные блоки объединяют исходящие сигналы от рекуррентных блоков верхнего уровня и ошибки, распространяемые в текущем уровне. Входные данные с верхнего уровня должны быть подвержены операции увеличения разрешения (передискретизации) из-за операции субдискретизации при передаче от одного уровня к другому. Рекуррентные вычисления внутри слоя подразумевают наличие горизонтальных соединений, которые связывают блоки  $R_1$ . Мы обсуждали модель с такими горизонтальными связями, сеть Хопфилда, в разделе 7.6. Многие современные модели используют другую реализацию рекуррентной сети, известную как модуль *долгой краткосрочной памяти* (LSTM), особый тип рекуррентного модуля, который хорошо подходит для изучения долгосрочных зависимостей в данных. Мы можем схематично описать активацию рекуррентных блоков как

$$R_i^t = \text{CONVLSTM}(E_i^{t-1}, R_i^{t-1}, \text{UPSAMPLE}(R_{i+1}^t)). \quad (8.14)$$

Прогнозы вычисляются непосредственно из повторяющихся активаций блока:

$$\hat{A}_i^t = \text{ReLU}(\text{CONV}(R_i^t)). \quad (8.15)$$

Блоки ошибок сигнализируют о разнице между входными данными и прогнозами с обоими возможными знаками:

$$E_i^t = [\text{RELU}(A_i^t - \hat{A}_i^t), \text{RELU}(\hat{A}_i^t - A_i^t)]. \quad (8.16)$$

Эту сеть можно обучить сквозным способом. Архитектура глубокой сверточной нейронной сети, описанная на рис. 8.5, была обучена распознаванию объектов. Напротив, модель PredNet на рис. 8.11 была обучена предсказанию следующего кадра в видеопоследовательностях. Исследователи обучили сеть, используя видео, снятые с камеры, установленной на автомобиле, и настроили сеть, чтобы предсказывать следующий кадр. Функция потерь основана на минимизации разницы между прогнозируемым кадром видео и фактическим кадром. Она может быть основана на последовательных кадрах или коротком интервале между двумя кадрами. Следует отметить, что процедура обучения аналогична формализму обратного распространения ошибки, описанному в разделе 8.6, хотя тип функции здесь отличается.

В то время как в предыдущих примерах функция потерь была задана разницей между фактическими и предсказанными названиями изображения, здесь сеть не обучается с использованием каких-либо явных названий или даже какого-либо явного представления об объектах. Тип процедуры обучения, при котором функция потерь непосредственно встроена во входную последовательность без какой-либо внешней аннотации, называется *обучением с самоконтролем*. Некоторые люди называют этот сценарий *неконтролируемым обучением*, но предпочтительнее использовать термин «неконтролируемое» в ситуации, когда нет функции потерь для каждого изображения, видео или пробной версии, например при кластеризации данных, при применении

tSNE, обучении Хебба<sup>1</sup>, пластиности, зависящей от спайковой частоты, или аналогичных механизмов.

Что может сделать такая сеть, как PredNet? Для начала модель может предсказать следующий кадр в видеопоследовательностях. В конце концов, это то, для чего модель обучалась. Модель может достичь этих прогнозов видео даже в других видео, которые отличаются от тех, на которых она была обучена. Кроме того, активации блоков могут использоваться для классификации объектов. Несмотря на то что сеть явно не обучена для классификации объектов, она обучается различию достаточно богатого набора признаков, которые относятся к естественным изображениям, и линейный классификатор может использоваться для присвоения названий объектам с использованием этого набора признаков.

Поэтому мы можем оценить сеть, используя те же тесты, описанные ранее для восходящих нейронных сетей, включая оценку эффективности классификации объектов, а также сравнения их выходных данных с поведенческими и нейронными данными экспериментов на людях и приматах. Например, блоки в сети показывают ингибицию окружения (рис. 8.11В, красная кривая). Подобно нейронам V1 обезьяны (рис. 8.11В), блоки в первом слое показывают большую активацию для более длинных полосок вплоть до определенной величины, а затем при увеличении длины активация уменьшается. В значительной степени ингибиция окружения происходит из-за исходящих сигналов, что демонстрируется подавлением блоков  $R$  в сети (рис. 8.11В, синяя кривая).

Ингибиция окружения можно рассматривать как форму пространственного прогнозирования. Модель также может делать временные прогнозы, такие как предсказание следующего кадра в видеопоследовательности. Одна из методик, которая широко используется в нейробиологии, – это задача обучения на последовательности, в которой животные узнают, что данный стимул В обычно следует за стимулом А (рис. 8.11С). Обезьян можно обучить работе с такого типа временными непредвиденными обстоятельствами, и нейроны в ИТС будут демонстрировать более низкий отклик на предсказанный второй стимул по сравнению с новым, неожиданным, вторым стимулом (рис. 8.11С, ср. ответ на предсказанный В (синий) по сравнению с непредсказанным В (красный)). Блоки ошибок на уровне 3 в PredNet также показывают этот тип реакции на новый, неожиданный стимул, и демонстрируют более низкую активацию, когда второй стимул ожидаем (рис. 8.11С).

Несмотря на то что модель PredNet никогда не была обучена присваивать названия объектам, или проявлять ингибицию окружения, или выполнять обнаружение нового стимула, эти и другие биологические свойства в данном типе сети проявляются, когда она обучается делать предсказания следующего кадра в видеопоследовательностях. Это появление несвязанных свойств особенно захватывающе, потому что оно предполагает, что фунда-

---

<sup>1</sup> Принцип Хебба заключается в том, что если один нейрон многократно стимулирует какой-либо другой нейрон, то сила связи между двумя нейронами увеличивается. – Прим. ред.

ментальные особенности архитектуры визуальной системы могут развиваться благодаря опыту с естественной статистикой мира, без необходимости тренировать модель с миллионами маркированных названиями примеров в контролируемом режиме. В общем, можно создавать биологически подобные нейронные архитектуры, которые учатся извлекать фундаментальные структуры мира в режиме самоконтроля. Некоторые биологические свойства проявляются в этих сетях естественным образом, через обучение их базовым принципам, таким как предиктивное кодирование.

## 8.18. Обзор главы

- Биологически подобные вычислительные модели визуальной обработки должны работать с изображениями, должны быть основаны на архитектуре нейронных цепей мозга и должны отображать фундаментальные свойства избирательности, инвариантности, скорости и обобщения.
- Современные развитые модели зрения основаны на иерархической архитектуре «разделяй и властвуй», состоящей из уровней, которые последовательно обрабатывают информацию.
- При подъеме по иерархии сети блоки охватывают большие размеры воспринимающего поля, отображают предпочтения для более сложных признаков и проявляют все большую толерантность к метрическим преобразованиям этих признаков.
- Глубокие сверточные нейронные сети обучаются от начала до конца так, что все веса в сети модифицируются в соответствии с заданной функцией потерь без необходимости ручной настройки параметров модели.
- Одним из основных способов изучения весов в глубоких сверточных нейронных сетях является использование градиентного спуска, реализуемого алгоритмом обратного распространения ошибки.
- Большие базы данных, такие как ImageNet, позволяют всесторонне обучать нейронные сети глубокой свертки посредством контролируемого обучения.
- Современные сети имеют огромное количество настраиваемых параметров, поэтому возникает вопрос, как их обобщить и избежать чрезмерной сложности.
- Перекрестная проверка – важный шаг для избежания получения завышенных значений эффективности, которые не экстраполируются на новые данные.
- После обучения отклики блоков в сети на любое произвольное изображение могут быть быстро вычислены. Эти ответы можно напрямую сравнить с поведенческими и нейрофизиологическими измерениями в экспериментах над людьми и приматами.
- Глубокие сверточные нейронные сети обеспечивают аппроксимацию первого порядка для первичных поведенческих характеристик, и сети могут также аппроксимировать структуру ошибок в задачах зрительного распознавания и характер движений глаз во время зрительного поиска.

- Активация блоков в сети также может быть использована для приблизительного прогнозирования ответов биологических нейронов во всей центральной зрительной коре, при предъявлении зрительных стимулов.
- В современных моделях отсутствуют многие биологические механизмы низкого уровня, а также многие психологические качества высокого уровня.
- Нисходящие сигналы необходимы для коммуникации сенсорных входов с памятью и предыдущими знаниями о мире.
- Нисходящие сигналы важны во время обучения (например, обратное распространение).
- Нисходящие сигналы также играют важную роль во время визуальной обработки, объединяя восходящие сенсорные сигналы с предиктивными сигналами, основанными на данных более высоких уровней.

## ЛИТЕРАТУРА

См. дополнительные ссылки на <http://bit.ly/36RxOGX>.

- Krizhevsky, A.; Sutskever, I.; and Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Presented at Neural Information Processing Systems, Montreal.
- Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience* 2: 79–87.
- Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience* 2: 1019–1025.
- Serre, T. (2019). Deep learning: the good, the bad and the ugly. *Annual Review of Vision* 5: 399–426.
- Yamins, D. L.; Hong, H.; Cadieu, C.F.; Solomon, E. A.; Seibert, D., and DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America* 111: 8619–8624.

Дополнительный контент вы можете скачать по ссылке <http://bit.ly/36RxOGX>.

# Глава 9

---

## К миру с разумными машинами, которые смогут видеть и интерпретировать мир вокруг нас

В предыдущей главе мы представили идею прямого сравнения вычислительных моделей с поведением человека в визуальных задачах. Например, мы оцениваем, как модели классифицируют изображение по сравнению с тем, как классифицируют то же самое изображение люди. В некоторых задачах типы ошибок, допускаемых вычислительными моделями, могут быть аналогичны человеческим ошибкам. Здесь мы углубимся в то, что современные алгоритмы компьютерного зрения могут и что не могут делать. Мы подчеркнем огромную мощность современных вычислительных моделей, в то же время отметив некоторые из их ограничений и увлекательную работу, которую нам предстоит проделать для создания более совершенных моделей.

Есть много задач по распознаванию изображений, с которыми компьютеры справляются уже значительно лучше людей. Простым примером является возможность считывать штрих-коды, такие как те, которые используются в супермаркете для маркировки продуктов. Даже если бы люди в принципе могли пройти длительную подготовку по считыванию штрих-кодов, в этой конкретной задаче им было бы чрезвычайно трудно достичь эффективности машины. В большинстве супермаркетов человеку все еще необходимо перевернуть продукт, найти штрих-код и разместить штрих-код таким образом, чтобы сканер мог его обработать. Этот уровень человеческого вмешательства, скорее всего, скоро исчезнет, но в некотором смысле интересно отметить, что локализовать штрих-код и правильно расположить его по-прежнему лучше получается у людей, чем у машин.

Здесь наблюдается двойное расхождение в отношении того, какие задачи люди могут решить легко (определение местонахождения штрих-кода и правильное позиционирование продукта для его считывания), а какие задачи легки для машин (расшифровка штрих-кода). Задача может показаться несколько ограниченной: все сводится к измерению ширины полос и расстояний. Человек решает сложную задачу инвариантности (распознавание изображения в разных масштабах, положениях и углах, как на рис. 3.6), размещая объект в нужном месте. Аналогичный случай можно сделать и для считывания кодов быстрого ответа (QR-кодов). Как мы обсудим вскоре, существует множество других визуальных задач, в которых компьютеры уже сравнялись с людьми или даже превосходят их. Есть также много визуальных задач, в которых машинам еще предстоит пройти долгий путь, чтобы достичь уровня эффективности человека. Ханс Моравек, Родни Брукс и Марвин Мински сформулировали это разделение между машиной и человеческими способностями в парадоксе Моравека. Парадокс заключается в том, что относительно легко наделить компьютеры эффективностью уровня взрослого человека на традиционных тестах интеллекта и невероятно сложно дать машинам навыки годовалого ребенка с точки зрения восприятия и мобильности.

Что будет означать, если вычислительные алгоритмы будут соответствовать или превосходить людей во всех возможных визуальных задачах? Представьте себе мир, в котором машины смогут по-настоящему видеть и интерпретировать визуальный мир вокруг нас, – это мир, в котором машины могут пройти визуальный тест Тьюринга.

## 9.1. ВИЗУАЛЬНЫЙ ТЕСТ ТЬЮРИНГА

Алан Тьюринг (1912–1954) был одним из великих умов XX века и пионером в информатике (и теории искусственного интеллекта. – Прим. ред.). В своей основополагающей статье 1950 года он предложил «игру в имитацию», в которой ряд вопросов задается как человеку, так и компьютеру. Тьюринг предположил, что если мы не можем различить, какие ответы были получены от человека, а какие – от компьютера, то мы должны считать, что этот компьютер обладает интеллектом.

Термин «интеллект» плохо определен и используется в разных смыслах. Кроме того, понятие машинного интеллекта является движущейся мишенью: как только компьютеры могут решить данную задачу (например, победить чемпионов мира в игре в шахматы или го), критики неизменно утверждают, что такой подвиг *не является* фактической демонстрацией интеллекта (хотя те же самые эксперты до того, как компьютеры победили людей, утверждали обратное). Эти люди часто имеют в виду совершенно бесполезное определение интеллекта: интеллект – это то, на что компьютеры не способны! Чтобы избежать подобных несуразиц, стандартным методом оценки интеллекта стал тест Тьюринга.

Мы можем определить специализированную версию теста Тьюринга для визуального интеллекта (рис. 9.1). Предположим, мы показываем человеку

или компьютеру изображение (или видео без звука). Важно, чтобы на содержимое изображения не накладывалось никаких ограничений: это может быть кадр, извлеченный из фильма Диснея, картина Кандинского или фотография, подобная изображенной на рис. 9.1. Нам разрешено задать любой вопрос, касающийся изображения. Например, мы можем спросить, есть ли на нем дерево, сколько на нем автомобилей, есть ли у кого-нибудь шляпа, находится ли человек в шляпе ближе к зрителю, чем дерево, есть ли наш друг Джон на картинке, выглядит ли Джон счастливым, смешная ли это картинка или грустная, сколько людей едут на велосипеде и так далее. Если мы не можем различить, исходят ответы от человека или от компьютера, то можем претендовать на победу. Мы сможем утверждать, что с поведенческой точки зрения люди интерпретируют изображения так же, как наш алгоритм компьютерного зрения.



**Рис. 9.1** ♦ Визуальный тест Тьюринга. Изображение может быть произвольным, и мы можем задать любой вопрос, касающийся изображения. Если мы не можем различить, исходят ответы от человека или от вычислительного алгоритма, то говорим, что алгоритм прошел визуальный тест Тьюринга

Здесь уместны несколько пояснений и уточнений. Если бы кто-нибудь задал мне вопросы об изображении на китайском языке, я бы не смог на них ответить. Это не отказ моей зрительной системы; это просто показывает, что я не говорю по-китайски. Я могу пройти визуальный тест Тьюринга, но я бы не прошел тест Тьюринга на китайском языке! Следовательно, определение теста Тьюринга на зрение предполагает, что у нас есть способ кодировать вопросы и ответы в формате, понятном компьютеру. Например, если мы спросим, выглядит Джон счастливым или нет, компьютер должен уметь определять, что означает «счастливый». Мы стремимся ограничить тест Тьюринга строгой визуальной обработкой и отделить ее от понимания языка.

Язык, конечно, еще один увлекательный аспект познания, и мы хотим, чтобы компьютеры тоже могли использовать язык. Можно даже расширить тест Тьюринга, включив в него и зрение, и язык. Например, позже в этой главе мы кратко обсудим задачу создания подписей к изображениям, то есть как придумать краткое описание изображения. Однако главная задача в этой главе – пройти визуальный тест. Поэтому мы определяем тест Тьюринга строго в области зрения. Мы по-прежнему хотим, чтобы компьютер мог ответить на любой вопрос, но нас не волнует, понимает ли компьютер слова вопроса, знает он грамматику или нет.

Чтобы компьютер мог ответить, выглядит Джон счастливым или нет, нужно обучить компьютер на изображениях счастливых людей и изображениях людей, которые не выглядят счастливыми. В качестве альтернативы мы могли бы придумать другие способы научить компьютер тому, как выглядят счастливые люди. Это обучение интерпретации задания применимо и ко всем другим вопросам. Если мы хотим знать, едет ли женщина на синем велосипеде, компьютеру необходимо понимать, что означает женщина, ездить, синий и велосипед. Конечно, то же самое верно и для человеческого зрения, даже если мы склонны принимать это как должное и недооценивать этот очевидный момент. Точно так же, как я не смог бы ответить на вопросы по-китайски, если мы спросим человека, есть ли на картинке *карга*, человек не сможет нам ничего ответить, если не знает, что означает слово «*карга*» (устаревшее слово, имеется в виду старуха).

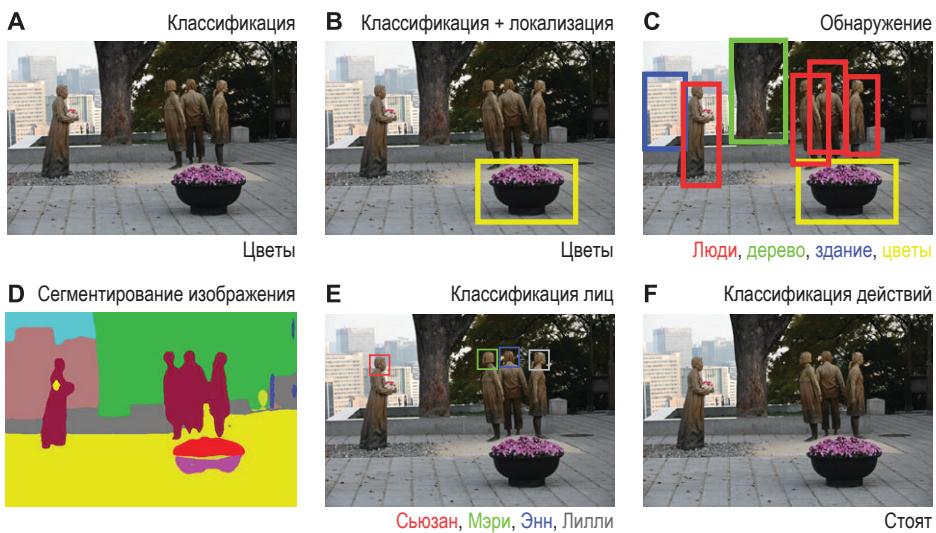
В этой формулировке теста Тьюринга важно, чтобы количество вопросов оставалось бесконечным. Например, можно построить вычислительную модель, которая отлично распознает, присутствует наш друг Джон на картинке или нет; то есть идеальный детектор Джона, который может распознать Джона даже лучше, чем мы. Такая вычислительная модель была бы неплохой, но она не прошла бы визуальный тест Тьюринга. Точно так же можно построить модель, которая может маркировать каждый пиксель изображения (этот пиксель является частью дерева; этот пиксель является частью красной машины; этот пиксель является частью Джона). Такая модель была бы еще более впечатляющей, но она не могла бы ответить ни на один произвольный вопрос об изображении, например счастлив Джон или нет, и, следовательно, модель также не прошла бы визуальный тест Тьюринга.

В то время как тест Тьюринга в том виде, в каком он определен до сих пор, фокусируется на человеческом восприятии, мы также можем сделать тест Тьюринга для зрения крысы, что означает алгоритм, который неотличим от поведения крысы при выполнении визуальных задач. Мы тоже можем определить тест Тьюринга для визуального тестирования годовалого младенца, имея в виду алгоритм, который неотличим от поведения годовалого младенца. Точно так же некоторые люди могут обладать довольно специализированными знаниями, например орнитолог, который может классифицировать разные типы птиц, или врач, который может диагностировать определенные состояния, просматривая рентгеновские снимки. Для этих случаев можно было бы определить ограниченные версии теста Тьюринга, такие как машина, которую невозможно отличить от мирового эксперта-орнитолога с точки зрения классификации птиц по их фото.

## 9.2. Компьютерное зрение повсюду

Несмотря на огромный прогресс в компьютерном моделировании обработки изображений, мы все еще далеки от того, чтобы создавать алгоритмы, которые могли бы пройти визуальный тест Тьюринга. Большинство разработок компьютерного зрения сосредоточены на конкретных наборах вопросов или задач на пути к созданию систем, которые могут пройти общий тест Тьюринга. Было разработано множество интересных алгоритмов для решения нескольких взаимосвязанных задач компьютерного зрения (рис. 9.2).

Одной из наиболее распространенных задач является *классификация объектов* (рис. 9.2А): компьютеру предоставляется изображение, и он должен присвоить объекту название из фиксированного числа возможных. Например, есть ли в изображении дерево [да|нет]? Какой из следующих объектов присутствует на изображении [люди|дерево|здание|цветок]? Другой пример классификации объектов – это задача клинической диагностики на основе изображений; например, содержит ли изображение маммограммы опухоль [да|нет]? Еще один пример классификации объектов – это задача распознавания лиц (рис. 9.2Е); например, кто на этом изображении [Сьюзан|Мэри|Энн|Лилли]?



**Рис. 9.2** ♦ Типичные задачи для тестирования моделей компьютерного зрения. (А) Классификация объектов. (Б) Классификация и локализация объектов. (С) Обнаружение объекта. (Д) Сегментирование изображения. (Е) Классификация лиц. (Ф) Классификация действий

При присвоении изображению названия эти названия могут быть вложены в структуры и иерархии. Например, некоторые психологи ссылаются на *категоризацию* объекта (содержит изображение автомобиль или лицо?), в отличие от идентификации объекта (какая марка и модель машины, чье конкретно

лицо?). С вычислительной точки зрения это, по сути, одна и та же задача, и можно разработать иерархические алгоритмы, которые будут отвечать на эти вопросы последовательно или параллельно.

Интригующий и вездесущий аспект человеческого языка – это определение категориальных различий, выходящих за рамки точных визуальных характеристик изображения; понятие семантических категорий обсуждалось в разделе 6.8 (рис. 6.5). Например, мы можем объединить изображения муравьев, змей, львов, птиц и дельфинов и отнести их к категории животных. Если мы тренируем компьютерное зрение, которое отлично распознает муравьев и змей, исключительно муравьев и змей, алгоритм может не понять, что птица – это другой вид животных. Эта неспособность экстраполировать на другое животное может показаться серьезной проблемой для компьютерного зрения: конечно, практически любой человек может сказать, что птица – это животное. Однако не ясно, смогут ли люди преуспеть в этой же задаче с тем же типом обучения, которому подвергались компьютеры. Представьте себе человека, который знает муравьев и змей, но никогда не видел других животных. Учитывая изображение птицы (без движения, без контекстной информации или любой другой подсказки; помните, что мы желаем как можно точнее сопоставить человеческую задачу с компьютерной задачей; в противном случае люди имеют несправедливое преимущество), сможет ли человек понять, что птица – это другой вид животных? Можно подумать, что да. Однако трудно представить, каким было бы его или ее понимание «животности», если бы весь их визуальный опыт ограничивался исключительно статическими изображениями муравьев и змей. Мы часто склонны недооценивать количество имеющегося у нас визуального опыта.

Другой вариант классификации объектов – это задача *верификации объекта*: по двум (или более) изображениям определить то, соответствуют ли изображения одному и тому же объекту. Например, сотрудник службы безопасности аэропорта сравнивает фото на паспорте и человека перед ним, чтобы оценить, соответствует ли человек изображению на фото или нет. Еще одна задача связана с поиском изображений; учитывая изображение, получить все экземпляры похожих изображений из базы данных. Например, нужно получить изображения из интернета, которые похожи на данное изображение.

Расширяя задачу классификации объектов, были разработаны алгоритмы *обнаружения объектов* или *локализации объектов* (рис. 9.2В и С). В этих задачах цель состоит в том, чтобы очертить рамку вокруг интересующего объекта на изображении. Например, «найдите всех пешеходов на изображении». Прогресс в локализации объектов быстро ускорился с разработкой базы данных MSCOCO, которая содержит детализированные контуры объектов из 80 общих категорий. Одним из примеров обнаружения объекта является возможность поместить лицо на изображении в рамку (обнаружение лица), которая в наши дни используется в цифровых камерах для фокусировки на лицах. Современные алгоритмы могут обнаруживать и размещать рамки вокруг нескольких объектов на изображении. Эти усилия дали колossalный толчок к возможности разработки беспилотных автомобилей, которые оснащены

датчиками для обнаружения других автомобилей, пешеходов, автомобильных полос и многих других объектов, представляющих интерес.

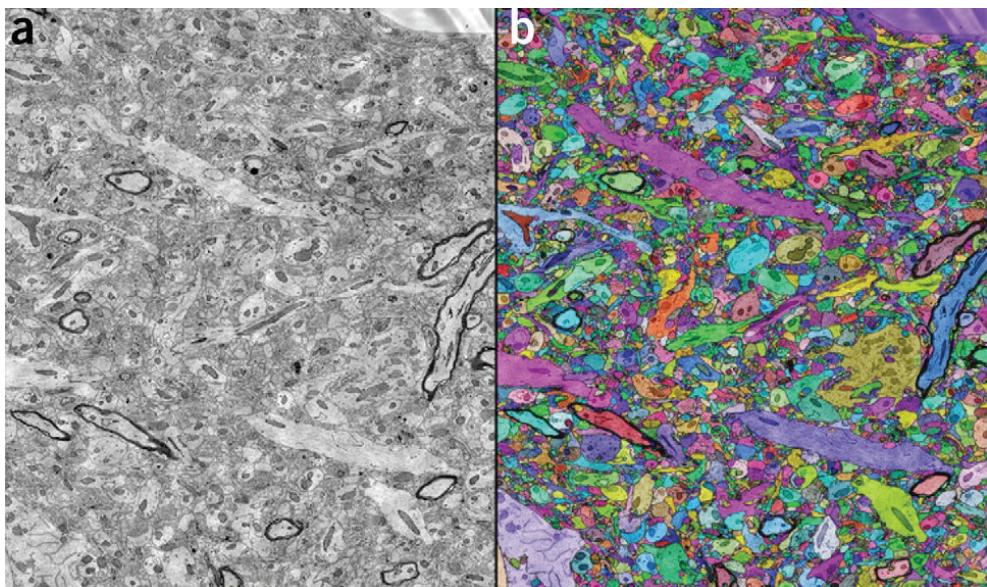
С задачей обнаружения объекта связан вопрос *сегментации* изображения, *выделения объекта* на изображении, цель которого – очертить контур данного объекта (рис. 9.2D). Первоначальная карта выделенных объектов на изображении может быть сделана путем адекватного определения их краев. Однако более сложные задачи часто требуют более глубокого понимания взаимосвязей между различными частями объекта. Примером сложной задачи для выделения объекта является случай зебры: алгоритм должен выделить зебру в целом, а не маркировать каждую полосу как отдельный объект. Другая типичная задача сегментации возникает при окклюзии. Например, рассмотрим повернутые буквы В на рис. 3.8: алгоритм сегментации изображения должен выделить каждую букву целиком, а не просто отмечать каждый фрагмент буквы как отдельный объект. Исследователей могут заинтересовать алгоритмы для выделения всех объектов на изображении, а не для локализации каждого отдельного объекта определенного класса. *Семантическое выделение границ* относится к определению контуров объектов на изображении без учета краев, которые не разделяют объекты.

В литературе широко обсуждается проблема курицы и яйца: идет ли сегментация до распознавания или распознавание идет первым. Когда может быть определена глубина, на основе бинокулярности зрения, окклюзии и движения, сегментация может произойти раньше, до распознавания. Однако когда единственные сигналы основаны на яркости, нет четких биологических свидетельств того, что сегментация происходит до распознавания или наоборот. Возможно, оба вычисления происходят параллельно. Во многих практических приложениях классификация, обнаружение и выделение объектов часто сочетаются.

Пример приложения, объединяющего все три задачи, включает анализ изображений микроскопии в клеточной биологии. Биологов интересует алгоритм, который может автоматически обнаруживать клетки заданной формы, отмечать их заданным цветом и подсчитывать их. Особенно сложная и захватывающая задача в этом направлении была выдвинута сообществом исследователей, работающих над картированием взаимосвязей в нервной системе на основе изображений, полученных с помощью электронной микроскопии (рис. 9.3). Эти изображения являются визуализацией последовательности тонких слоев внутренней структуры нервной ткани с высоким разрешением; цель состоит в том, чтобы автоматически проследить связи каждого нейрона с другими по этим изображениям. Под *сегментацией образца* понимается разделение и маркировка каждого пикселя изображения. Например, на снимках электронной микроскопии мы хотим пометить каждый дендрит, сому, аксон нейрона, каждую глиальную и другие типы клеток. Мы особенно хотим проследить за дендритами и аксонами в нескольких слоях, чтобы сопоставить, где они проходят и где находятся синаптические соединения с другим нейроном.

*Распознавание действий* означает способность идентифицировать действия на изображении или видео (рис. 9.2F, рис. 9.4). Человек играет в футбол [да|нет]? Какое из этих действий выполняет человек [играет на

виолончели|чистит зубы|играет в боулинг|жонглирует мячом]? Распознавание действий может быть основано на отдельных изображениях, но оно также привело к разработке соответствующих баз данных видео. В спорте люди заинтересованы в создании систем компьютерного зрения, которые могут автоматически анализировать игру с мельчайшими подробностями, включая обнаружение отдельных игроков, их отслеживание и определение того, что они делают (например, пробежка с мячом, пас, дриблинг, удар по воротам).

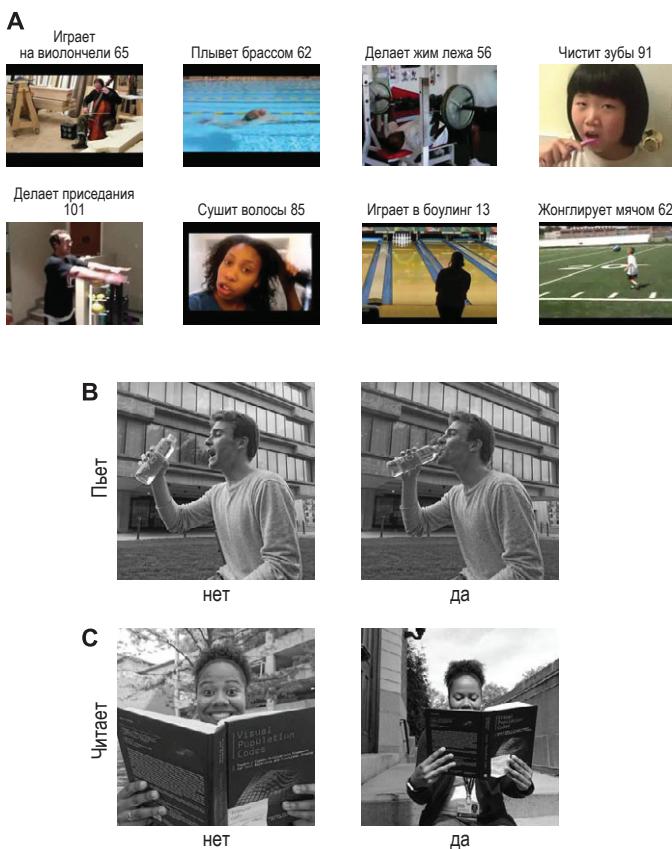


**Рис. 9.3 ♦** Алгоритмы сегментации изображения могут помочь картировать нейронные связи. (А) Электронная микрофотография среза коры головного мозга мышей размером 40×20 мкм. (В) Автоматическая компьютерная сегментация, где каждый клеточный объект отображается отдельным цветом, наложенным на исходное изображение. Воспроизведено по Лихтман и др., 2014

Распознавание и отслеживание действий – это примеры, когда многие задачи компьютерного зрения, определенные ранее, переплетаются и должны быть объединены. Приложения для распознавания действий также получили широкое распространение среди биологов, изучающих поведение животных. Традиционно количественная оценка поведения животных была утомительной и трудоемкой задачей: аспирант, интересующийся поведением мыши, может легко установить камеру для записи часов и часов поведенческих данных. Анализ этих данных обычно требовал долгих часов изучения этих видео и субъективного описания поведения животного. В настоящее время некоторые системы могут объективно и надежно выполнять эти типы аннотаций: подходы компьютерного зрения могут автоматически анализировать видео, количественно определять объем времени, затрачиваемого на различные формы поведения, и описывать последовательность различных типов движений. Еще одно широко распространенное применение систем

распознавания действий – охранные системы наблюдения. Кого-то может заинтересовать обнаружение «аномального» поведения возле дома, в аэропорту или на многолюдном концерте. Ученые, занимающиеся компьютерным зрением, называют эту задачу *обнаружением аномалий*.

Распознавание действий – хороший пример того, как постановка эксперимента и базы данных могут сделать задачи легкими или сложными. Распознать, играет ли кто-то на виолончели или жонглирует футбольным мячом, по типам изображений, показанных на рис. 9.4А, несложно. Однако определить, читает человек или нет, на основе типов изображений, показанных на рис. 9.4С, может быть значительно труднее. Мы обсудим это еще раз в разделе 9.10.



**Рис. 9.4** ♦ Структура базы исходных данных может упростить или усложнить задачу распознавания. (A) База данных UCF-101 видео с названными действиями (Сoomро и др., 2012). Здесь показан первый кадр в восьми примерах из 101 категории действий. В заголовках указывается номер категории и описание. (B)–(C) Сложный набор данных для распознавания действий, где подопытные должны в бинарной форме указать, пьет он или нет (B), читает или нет (C)

Список приложений компьютерного зрения настолько обширен и растет так быстро, что вполне возможно, что к тому времени, когда читатель про-

читает эти строки, уже будет множество впечатляющих новых достижений в этой области.

## 9.3. ДОБАВЛЕНИЕ ВРЕМЕННОЙ ИНФОРМАЦИИ С ПОМОЩЬЮ ВИДЕО

Исторически сложилось так, что многие исследования компьютерного зрения ограничивались анализом статических изображений. Частично работа была сосредоточена на статических изображениях, потому что и люди, и машины могут довольно хорошо распознавать объекты на таких изображениях. Акцент на статические изображения также отчасти является исторической случайностью: было проще создавать базы данных со статическими изображениями, они занимают меньше места на жестком диске и требуют меньше вычислительных ресурсов для обработки. Эти технические ограничения сегодня менее актуальны.

В естественных условиях просмотра есть несколько подсказок, которые зависят от интеграции информации с течением времени. Эти динамические подсказки могут значительно улучшить классификацию объектов. Типичный случай, когда важна временная интеграция, – это распознавание действия. Хотя можно распознать действия исключительно по статическим изображениям (например, рис. 9.4), обычно как людям, так и компьютерам распознать действие значительно проще на видео. Например, может оказаться сложно определить, говорит человек или нет, используя только статичное изображение.

Современные модели распознавания действий из пространственно-временных входных данных, основанные на архитектурах глубокой сверточной нейронной сети, можно разделить на три группы: (i) сети с трехмерными сверточными фильтрами, в которых пространственные и временные характеристики обрабатываются вместе с помощью трехмерных сверток; (ii) двухпотоковые сети, в которых один поток обрабатывает пространственную информацию, а другой поток получает оптический информационный поток из последовательности кадров, и два потока объединяются на поздней стадии для классификации; (iii) сети, которые подпитываются повторяющейся архитектурой, такой как долговременная краткосрочная память (LSTM) (раздел 8.17), которая интегрирует пространственные характеристики с временными данными.

Временная информация актуальна для многих других задач, помимо распознавания действий. Выделение объектов, как правило, значительно упрощается с помощью данных видео. Важность временных изменений для сегментации объясняет распространенное использование камуфляжа в животном мире. При отсутствии движения подходящие цвета, контраст и текстуры помогают животным ускользнуть от внимания хищников или, по крайней мере, выиграть достаточно времени, чтобы убежать. Особенно сложно сегментировать объекты на зрительной периферии, однако нейроны

с рецептивными полями, расположенными на больших эксцентризитетах, остаются очень чувствительными к зрительному движению. Кроме того, движение – один из самых сильных индикаторов заметности снизу вверх сети.

Временная информация также может играть важную роль в зрительном обучении. В элегантном эксперименте кошек выращивали в условиях стробоскопического освещения, то есть с короткими включениями света, как на дискотеке; такое освещение не позволяет увидеть непрерывное, плавное движение. Развитие первичной зрительной коры у этих кошек было ненормальным с точки зрения избирательности ориентации, бинокулярной интеграции, обнаружения движения и размеров рецептивных полей. Эти результаты дополнительно подтверждают обсуждение в разделе 2.2 статистики естественных стимулов, управляющих свойствами настроек нейронов в зрительной системе.

Кроме того, поскольку объекты не появляются и не исчезают мгновенно, использование видеоданных может естественным образом помочь людям и компьютерным моделям научиться распознавать объекты с разных ракурсов. Видеопоследовательности автоматически предоставляют биологически правдоподобный способ выполнения «увеличения данных» путем получения множества похожих изображений объекта от изображения с одним названием (раздел 8.9). Другой пример того, как временная информация может использоваться для визуального обучения, – это случай обучения с самоконтролем для предсказания будущих событий, обсуждаемый в алгоритме PredNet в разделе 8.17 (рис. 8.11).

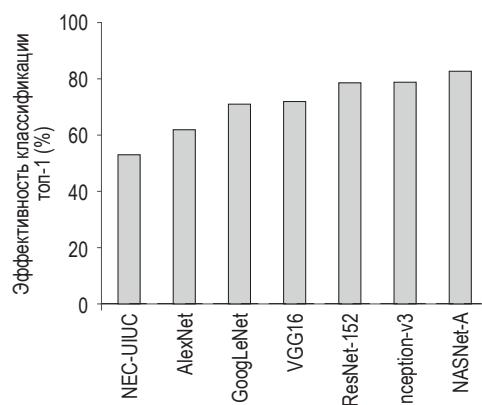
## 9.4. ОСНОВНЫЕ ЭТАПЫ КЛАССИФИКАЦИИ ОБЪЕКТОВ

В разделе 8.7 мы представили несколько баз данных изображений, таких как ImageNet, которые сыграли важную роль в разработке вычислительных моделей зрительного распознавания (рис. 8.4). Эти базы данных были созданы для крупномасштабных задач визуального распознавания, в которых исследователи соревнуются, чтобы получить меньшие классификационные ошибки.

Хороший способ сказать об эффективности в этих соревнованиях – указать на «топ-1» точности классификации, когда модель присваивает одно название каждому изображению, и результат будет либо правильным, либо неправильным. Многие приложения компьютерного зрения сообщают о более мягкой и менее точной характеристике: «топ-5» точности классификации, когда модели разрешается создавать пять разных названий для каждого изображения, и результат считается правильным, если хотя бы одно из этих названий верно. Одним из оправданий выбора показателя топ-5 является то, что некоторые естественные изображения, извлеченные из интернета, содержат несколько объектов. Изображение может содержать собаку, но также и дерево; поэтому связь между этим изображением и названием дерева является произвольной. На том же изображении легко можно было обозначить и одну собаку. Хотя в этих показателях есть смысл, отчеты с применением топ-5 преувеличивают точность алгоритмов и затрудняют прямое сравнение

с эффективностью человека. Например, рассмотрим изображение из базы данных ImageNet (где есть 1000 возможных названий), показывающее исключительно дерево на улице. Название изображения – «дерево». Вычислительный алгоритм может предоставить следующие пять названий, отсортированных в порядке убывания вероятности, указанном числами в скобках: слон (вероятность = 0,62), холодильник (0,31), машина (0,02), дерево (0,02), лед (0,01). В сумме эти вероятности составляют 0,98, а не 1, потому что оставшиеся  $1000 - 5 = 995$  категорий в сумме дают 0,02. Эти пять названий можно было бы считать правильным ответом в соответствии с пятеркой лучших показателей точности, но они несколько странные. Люди не сказали бы, что изображение с вероятностью 0,62 содержит слона и с вероятностью 0,31 содержит холодильник! Другие базы данных, такие как MSCOCO, маркируют несколько объектов на изображении, поэтому можно проверить точность нескольких названий.

На рис. 9.5 показана эффективность топ-1 в ImageNet для нескольких вычислительных моделей, многие из которых выиграли соревнования по классификации объектов за последнее десятилетие, и некоторые из них уже упоминались в гл. 8. Эффективность в топ-1 на данный момент несколько превышает 80 %, а эффективность в топ-5 составляет почти 95 %. Эти показатели впечатляют, учитывая, что существует 1000 классов, и, следовательно, простая вероятность составляет 0,1 %. С учетом аргументов, приведенных в предыдущем абзаце, непросто напрямую сравнить эти показатели эффективности с человеческими, особенно с показателями топ-5. Люди не очень хороши в классификации из 1000 категорий: трудно запомнить эту 1000 названий, и у людей может быть много предубеждений в отношении запоминания и использования одних названий в ущерб другим. Кроме того, как мы обсуждали в разделе 8.7, некоторые категории изображений в ImageNet являются несколько эзотерическими (сколько раз вы видели изоподу, эркер



**Рис. 9.5** ♦ Эволюция эффективности базы данных ImageNet. Эффективность классификации топ-1 объектов, основанной на базе данных ImageNet. Каждый столбец относится к разному вычислительному алгоритму. Вероятность случайного совпадения – 0,1 %

или кирасу?). Люди потенциально могут быть обучены так же, как алгоритмы на рис. 9.5, чтобы стать экспертами в распознавании изопод, эркеров, кирас или любых других 997 названий. Независимо от этих соображений, неофициальные измерения эффективности человека в этой базе данных дают уровень точности от 90 до 95 %. Следовательно, даже со всеми их ограничениями современные алгоритмы могут выполнять классификацию объектов на изображениях ImageNet так же или даже лучше, чем люди.

Следует отметить, что эффективность классификации топ-1 не всегда является отличным показателем. Например, в следующем разделе мы рассмотрим задачу анализа клинических изображений. Рассмотрим конкретное заболевание, которое встречается у одного из 10 000 человек. Предположим, что мы обучаем алгоритм и он достигает эффективности 99,99 %. На первый взгляд такая эффективность кажется весьма впечатляющей. Тем не менее легко достичь эффективности 99,99 %, просто указав, что ни одно изображение не содержит признаков болезни! Такой банальный алгоритм вообще непригоден. Алгоритм будет иметь 9999 истинно отрицательных, 0 истинно положительных, 1 ложноотрицательный и 0 ложных срабатываний. В частности, в ситуациях, когда существует разница между количеством изображений с каждым названием (проблема несбалансированной классификации), полезно определить два показателя, точность и отклик:

$$\text{отклик} = \frac{\text{истинно положительные}}{(\text{истинно положительные} + \text{ложноотрицательные})};$$

$$\text{точность} = \frac{\text{истинно положительные}}{(\text{истинно положительные} + \text{ложноположительные})}.$$

Алгоритм, утверждающий, что ни одно из изображений не показывает наличие болезни, имеет нулевой отклик и нулевую точность, даже несмотря на то, что он достигает точности 99,99 %. И наоборот, рассмотрите другой алгоритм, который *также* бесполезен, который маркирует все изображения как показывающие доказательства болезни. Этот алгоритм будет иметь 0 истинно отрицательных, 1 истинно положительный, 0 ложноотрицательных и 9999 ложноположительных результатов. Отклик будет равен 1, что может показаться довольно приятным, за исключением того, что точность будет очень низкой, несмотря на высокий уровень отклика. Те же идеи часто обсуждаются в классах статистики как ошибка типа I (ложные срабатывания) и ошибка типа II (ложные отрицательные результаты). Для особо увлеченных некоторые исследователи также используют другую метрику, называемую оценкой  $F_1$ , которая представляет собой гармоническое среднее значение точности и отклика:

$$F_1 = 2 \frac{\text{точность} \cdot \text{отклик}}{\text{точность} + \text{отклик}} = \frac{\text{истинно положительные}}{\frac{\text{истинно положительные} + \text{ложноотрицательные}}{0,5(\text{ложноположительные} + \text{ложноотрицательные})}}.$$

В зависимости от характера задачи и последствий ошибок ложноположительные срабатывания могут быть намного хуже, чем ложноотрицательные, и наоборот. Можно назначать веса в функциях потерь, чтобы по-разному

наказывать за различные типы ошибок. Например, если принять, что отклик в  $\beta$  раз важнее точности, можно определить

$$F_\beta = (1 + \beta^2) \frac{\text{точность} \cdot \text{отклик}}{(\beta^2 \cdot \text{точность} + \text{отклик})}$$

(что эквивалентно  $F_1$ , когда  $\beta = 1$ ).

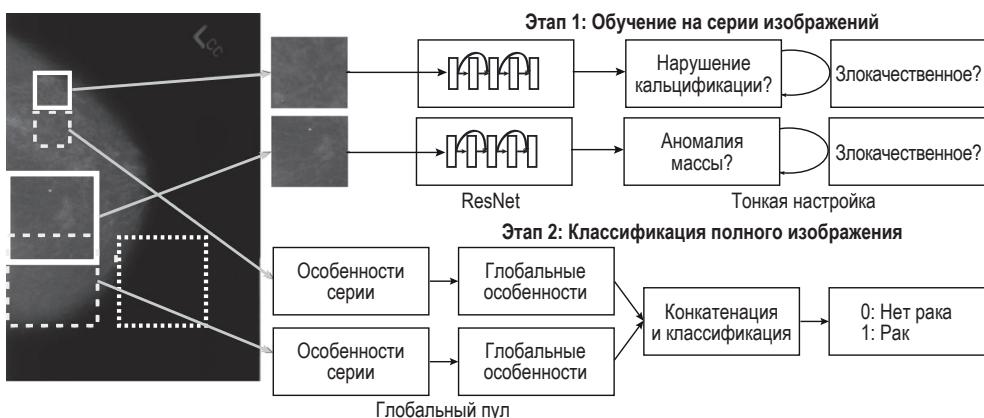
Независимо от конкретных показателей, очевидно, что был достигнут заметный прогресс в задачах классификации объектов (рис. 9.5). Сам AlexNet продемонстрировал существенный рост по сравнению со всеми своими предшественниками, что привело к быстрому исследованию более глубоких и сложных архитектур, которые повысили эффективность более чем на 20 % за период менее чем в десять лет. Это заметное улучшение показателей академических соревнований привлекло внимание многих людей, желающих заняться разработкой приложений для распознавания образов.

## 9.5. ПРИЛОЖЕНИЯ АЛГОРИТМОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ КЛАССИФИКАЦИИ РЕАЛЬНЫХ ОБЪЕКТОВ ФИЗИЧЕСКОГО МИРА

Успех соревнований по классификации изображений вдохновил разработчиков во многих областях компьютерного зрения. Одним из первых реальных приложений было оптическое распознавание символов (OCR), которое быстро стало популярным в сортировке почты на основе почтовых индексов, написанных от руки. Теперь есть удобные приложения, которые могут переводить в символы даже рукописные строки математических формул. С одной стороны, некоторые математические символы относительно просты; с другой – математические символы, возможно, менее стереотипны, и обучающих данных меньше, чем в других приложениях OCR. Алгоритмы компьютерного зрения уже добились быстрого прогресса во множестве интересных приложений; далее мы обсудим лишь несколько примеров.

Область, которая быстро изменяется с развитием компьютерного зрения, – это анализ клинических изображений. Клинический диагноз, основанный на изображениях, иногда можно упростить до зрительного распознавания образов. Клиницисты могут комбинировать информацию, полученную при диагностике на основе изображений, с большим количеством другой информации, включая анамнез, генетическую информацию, симптомы и прочее. Как объединить эти различные источники информации в методы автоматической диагностики – эта задача интересна сама по себе, но она выходит за рамки нашего текущего обсуждения. Здесь мы ограничиваем задачу диагностики строгим анализом изображений. Например, радиолог может исследовать маммограмму, чтобы определить, есть ли опухоль в грудной железе (рис. 9.6). База данных, состоящая из множества изображений маммограмм, аннотированных экспертами, может быть легко использована для обучения алгоритмов компьютерного зрения. Американское онкологическое общество

рекомендует делать маммограмму, обычно состоящую из двух рентгеновских снимков каждой груди, всем женщинам один или два раза в год, в зависимости от возраста. Такое количество маммограмм приводит к большому количеству изображений (около 40 млн изображений в год только в Соединенных Штатах). Задача важна, потому что ранняя диагностика может иметь решающее влияние на выбор курса действий. Подсчитано, что радиологи анализируют порядка 10 000 случаев в год; рентгенолог с тридцатилетним опытом, возможно, изучал 300 000 случаев. В настоящее время алгоритм компьютерного зрения можно обучить на гораздо большем количестве примеров, чем человек-клиницист может увидеть за всю свою жизнь.



**Рис. 9.6** ♦ Компьютерное зрение может помочь в постановке диагноза с помощью анализа изображений. Пример алгоритма обнаружения рака на маммограммах молочных желез. Измененное по Лоттер, 2018

Алгоритмы компьютерного зрения преуспели в самых разных разработках диагностики по изображениям. Для обучения и тестирования этих алгоритмов необходимы достоверные аннотации, предоставляемые клиницистами. Следует отметить, что люди – существа своеобразные. Клиницисты не всегда соглашаются друг с другом в диагностике данного изображения (вариативность между экспертами). Более того, врачи иногда даже не соглашаются с самими собой при многократном тестировании на одних и тех же изображениях (вариативность эксперта)! В случае обнаружения опухоли молочной железы вычислительные алгоритмы выступают сейчас на уровне или даже лучше, чем врачи-клиницисты. Другими словами, различия между современным алгоритмом компьютерного зрения и экспертом-человеком такие же, как и вариативность внутри эксперта и между экспертами. Будущие поколения могут рассматривать людей, пытающихся диагностировать изображения, так же, как мы сейчас рассматриваем человека, пытающегося прочитать штрих-код в супермаркете или вычислить квадратный корень из 17 в уме<sup>1</sup>.

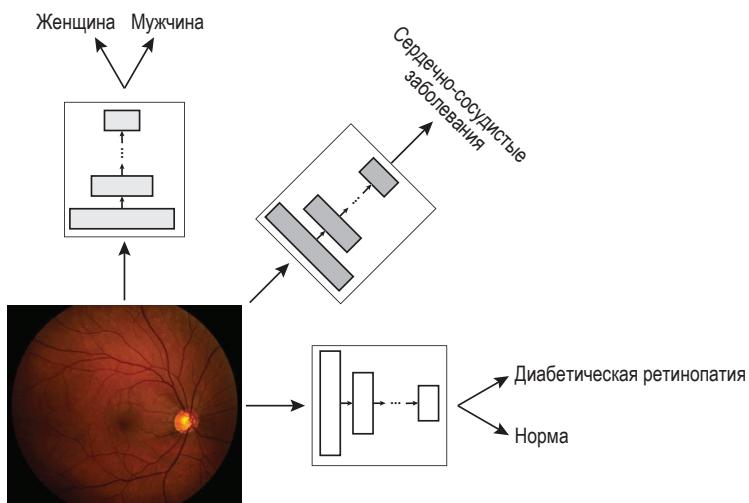
<sup>1</sup> Что совсем несложно, другое дело – вычислить корень 17-й степени из числа 2. – Прим. ред.

Хотя наличие или отсутствие опухоли является центральным вопросом, представляющим интерес для подавляющего большинства обследований груди, иногда могут возникать другие важные вопросы, которые врачи могут задать об изображении. Например, иногда случаются неожиданные находки, когда человека сканируют для диагностики данного состояния X (например, рака груди), сканирование не выявляет никаких результатов относительно X, но радиолог обнаруживает другие аномалии, которые приводят к другому диагнозу – Y. Такие случайные находки проблематичны для современных алгоритмов компьютерного зрения, потому что они могут быть крайне редкими. Алгоритмы являются сверхспециализированными и превосходят рентгенологов в обнаружении состояния X, но никогда не были обучены обнаружению редкого состояния Y. Одним из возможных компромиссов в качестве начального решения этой задачи было бы, чтобы системы компьютерного зрения помечали такие изображения как аномальные и отправляли их обратно человеку для дальнейшего изучения.

Случайные открытия представляют собой одну из областей, где люди все еще могут превосходить машины в диагностике клинических изображений – они могут найти то, что компьютеры упустят. Верно и обратное: машины могут на клинических изображениях обнаружить новые закономерности, которые ранее не обнаруживались людьми. Интригующий пример этого явления возник, когда исследователи разрабатывали методы компьютерного зрения, исследуя фотографии глазного дна сетчатки для диагностики состояния, известного как диабетическая ретинопатия (рис. 9.7). Диабетическая ретинопатия – это состояние, которое может возникнуть у пациентов с диабетом, когда высокий уровень сахара в крови вызывает набухание и кровотечение сосудов сетчатки. Эти кровеносные сосуды можно исследовать на фотографиях глазного дна, которые представляют собой изображения задней части глаза, используемые офтальмологами для диагностики заболевания. После сбора сотен тысяч маркированных изображений алгоритм компьютерного зрения с глубоким обучением быстро научился ставить диагнозы соответственно тому, как это делают врачи, что на данном этапе неудивительно.

Указать диагноз – это только один из вопросов, которые можно задать об этих изображениях. Исследователи решили применить свои алгоритмы машинного обучения к другим вопросам на тех же изображениях. Удивительно, но ученые-информатики спросили, могут ли они извлечь другие типы информации из фотографий глазного дна. Например, вместо того чтобы решать «да или нет» для диабетической ретинопатии, они обучили одни и те же алгоритмы прогнозированию возраста пациента. Алгоритмы смогли довольно точно определить возраст с абсолютной погрешностью менее 3,5 лет. Затем исследователи оценили, могут ли они предсказать пол субъекта. Удивительно, но они смогли сделать это исключительно хорошо, с площадью под кривой рабочей характеристики приемника (ROC) 0,97. Кривая ROC представляет собой график зависимости вероятности правильного определения от вероятности ложной тревоги. Достичь высоких показателей обнаружения за счет высокого уровня ложных тревог (заявляя, что каждое изображение показывает болезнь; см. предыдущий раздел) или низкого

уровня ложных срабатываний без какого-либо правильного обнаружения (утверждая, что ни одно изображение не показывает болезнь) тривиально. Хороший алгоритм будет иметь низкий уровень ложных тревог и высокую вероятность обнаружения. Лучшее, что может достичь алгоритм, – это уровень 1.0; простая вероятность дает 0,5. Квалифицированные офтальмологи никогда не могли оценить чей-либо пол по фотографиям глазного дна вообще. Возможно, они никогда не задумывались об этом вопросе; в конце концов, перед клиницистами находятся сами пациенты и записи о них. Однако даже после того, как врачам сообщили, что на этих изображениях присутствует информация о поле и возрасте, и попросили врачей указать пол или возраст, они не смогли этого сделать. Не совсем ясно, какие именно особенности изображения использует алгоритм для определения пола или возраста. Можно предположить, что, возможно, врачи, как мужчины, так и женщины, могут располагать аппарат так, чтобы делать снимки глазного дна в среднем немного ближе к пациентам-женщинам, чем к пациентам-мужчинам, при получении этих изображений. Алгоритмы вполне могли уловить такую небольшую подсознательную установку. В качестве альтернативы, возможно, существуют настоящие тонкие различия между женскими и мужскими кровеносными сосудами в сетчатке. Независимо от того, справедливо ли это объяснение, этот пример показывает, что компьютерное зрение может обнаруживать особенности изображения, которые не очевидны даже для специалистов в этой области.



**Рис. 9.7** ♦ Вычислительные алгоритмы могут делать неожиданные находки в анализе наблюдений. Пример клинического применения компьютерного зрения – фотографирование задней части глаза (фотография глазного дна) и использование глубокой сверточной сети для диагностики диабетической ретинопатии (Поплин и др., 2018). Кроме того, алгоритмы компьютерного зрения можно обучить задавать другие вопросы по тому же изображению, включая прогнозирование пола объекта или даже риска сердечно-сосудистых заболеваний

Оценка возраста и пола объекта по фотографиям глазного дна, возможно, не особенно увлекательна с практической точки зрения. Самая загадочная находка возникла, когда исследователи решили задать еще более смелый вопрос: можно ли предсказать риск сердечно-сосудистых заболеваний по фотографиям глазного дна? Исследователи в области цифровых технологий обнаружили, что они могут предсказать сердечно-сосудистые заболевания по фотографиям глазного дна с площадью под кривой ROC 0,7. Этот результат весьма примечателен, потому что этот вопрос, о котором офтальмологи не задумывались; это вопрос, который чрезвычайно актуален с клинической точки зрения. Компьютерный анализ представляет собой дополнительную информацию, которая бесплатно поступает из фотографии глазного дна без каких-либо дополнительных действий или клинических испытаний. Что, возможно, еще более примечательно, так это то, что алгоритм компьютерного зрения смог предсказать сердечно-сосудистые заболевания лучше, чем оценка риска Фрамингема, которая считается одним из лучших индикаторов риска сердечно-сосудистых заболеваний на основе десятилетий клинической работы. Алгоритмы компьютерного зрения могут не только научиться диагностировать изображения, как врачи, но они также могут научить нас новому об этих изображениях.

Есть несколько ситуаций, когда необходимо классифицировать огромное количество изображений (или видео). Автоматическая классификация изображений нашла применение не только в клинической диагностике. Например, компьютерное зрение пролило свет на грандиозную задачу классификации галактик и экзопланет по изображениям телескопов. Существует огромное количество изображений, которые помогут нам понять форму галактик и охарактеризовать планеты за пределами Солнечной системы, но у нас недостаточно астрофизиков, чтобы классифицировать все эти изображения. Астрофизики обратились к краудсорсингу, вовлекая общественность в просмотр изображений и научившись классифицировать галактики. Это идеальная среда для применения методов распознавания образов компьютерного зрения: за последние несколько лет алгоритмы машинного обучения сделали много захватывающих открытий. Концептуально похожий пример – категоризация растений и животных. Компьютерное зрение использовалось для классификации флоры и фауны, быстро превзойдя любого наивного наблюдателя и став предметом зависти для опытных биологов.

Еще одна задача классификации изображений, которая была радикально преобразована компьютерным зрением, – это идентификация лиц. Существует множество приложений для алгоритмов автоматического распознавания лиц. Многие смартфоны имеют алгоритмы, которые используют лица для входа в систему, что не так давно было только в научно-фантастических фильмах. Facebook теперь может находить фотографии, на которых изображен конкретный человек, даже когда этот человек не отмечен тегом. Количественные исследования идентификации лиц показали, что системы компьютерного зрения лучше, чем судебно-медицинские эксперты, и даже лучше, чем так называемые суперраспознаватели, люди с необычайной способностью распознавать и запоминать лица. Также растет индустрия прило-

жений безопасности, основанных на функциях распознавания лиц. Приложения безопасности в ближайшем будущем могут полагаться и на алгоритмы классификации распознавания действий. Наряду с достижениями в области распознавания лиц ведутся энергичные и своевременные обсуждения вопросов конфиденциальности. Вполне вероятно, что очень скоро будет довольно сложно пройти по улице, не будучи узнанным. Сценарий Большого брата Джорджа Оруэлла с камерами, способными распознавать людей, теперь технически осуществим.

Захватывающий прогресс в области беспилотных автомобилей также был вызван прогрессом в области компьютерного зрения – с такими задачами, как определение местоположения пешеходов, автомобилей, знаков «STOP», светофоров, других знаков, полос движения, тротуаров и даже животных, велосипедов или аномальных объектов на дороге. Хотя большинство приложений компьютерного зрения полагаются на видео или изображения с обычных камер, изображения не должны ограничиваться такими источниками. Например, беспилотные автомобили могут одновременно использовать информацию с нескольких камер и многих других датчиков. В области компьютерного зрения был достигнут настолько большой прогресс, что большинство инженеров, пытающихся создать автомобили с автоматическим управлением, думают, что главные проблемы, стоящие впереди, выходят за рамки компьютерного зрения и включают вопрос принятия решений, юридические вопросы и надежность.

Другие приложения алгоритмов компьютерного зрения все еще находятся в стадии разработки, но будут готовы довольно скоро. Например, существует большой интерес к интеллектуальному поиску изображений или видео в контенте (в литературе по компьютерному зрению это называется поиском изображений). Поиск в интернете по содержанию (в отличие от поиска по слову «собака» и использования названия для поиска текста или изображений с названием собаки) открывает двери для целого набора приложений. Первоначальные прототипы таких поисковиков уже созданы.

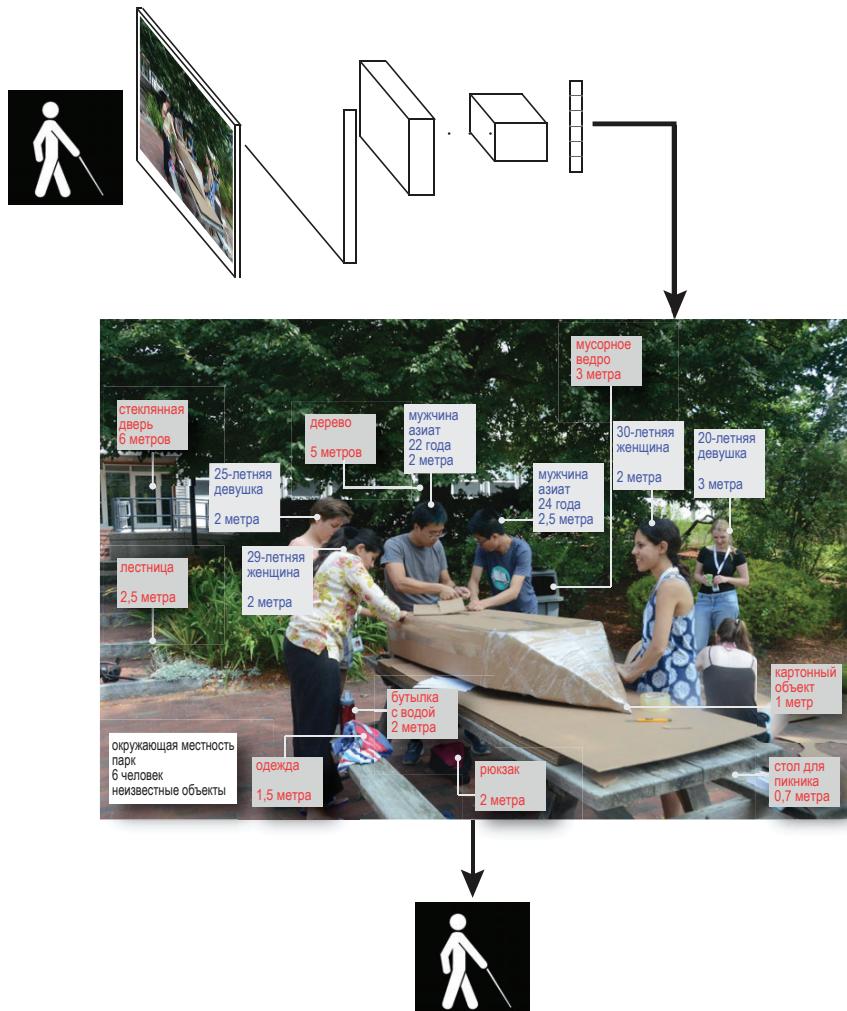
В предыдущем разделе были представлены достижения в распознавании лиц. Эти алгоритмы позволяют искать людей по фотографиям и могут иметь множество увлекательных приложений, таких как поиск пропавших без вести людей или поиск давнего друга. Прогресс в распознавании лиц может вскоре привести к появлению банкоматов, способных распознавать клиентов. Автомобили и дома также могут в ближайшем будущем узнавать своих владельцев по лицам. Прогресс в распознавании людей и распознавании действий может радикально изменить систему проверки безопасности в многолюдной среде, включая аэропорты, стадионы и, возможно, каждую улицу в больших городах. Работа над приложениями компьютерного зрения для проверки безопасности и других целей уже ведется в нескольких крупных городах<sup>1</sup>.

---

<sup>1</sup> Например, чтобы попасть в некоторые офисы или спортивные центры, сейчас уже недостаточно электронного пропуска – нужно, чтобы вас узнала система распознавания лиц. – Прим. ред.

## 9.6. Компьютерное зрение может помочь людям со зрительной дисфункцией

Особенно интересным применением систем компьютерного зрения является помочь людям с нарушениями зрения, особенно слепым (рис. 9.8). В одних только Соединенных Штатах примерно один миллион человек являются сле-



**Рис. 9.8** ♦ Компьютерное зрение может помочь людям с ослабленным зрением. Пример потенциального подхода к использованию компьютерного зрения для помощи людям с нарушениями зрения. Слепой человек может носить камеру, которая подключается к алгоритму компьютерного зрения и может интерпретировать окружающую сцену. Алгоритм компьютерного зрения может предоставлять информацию о людях, объектах, расстояниях и относительном местоположении объектов в режиме реального времени

пыми и около 3,25 млн человек страдают нарушениями зрения. В сочетании с высококачественными и относительно недорогими камерами алгоритмы компьютерного зрения могут помочь переваривать выходные данные цифровых камер для передачи информации слепым. Большинство телефонов в наши дни могут определять местоположение человека по координатам GPS, но вскоре можно будет получать еще более точную информацию, указав на телефон и заставив его определять направление к определенным магазинам, автобусным остановкам или достопримечательностям. Телефоны также могут помочь в чтении указателей и меню ресторана. Однако слепые люди нуждаются и заслуживают гораздо большего.

Интересным применением компьютерного зрения было бы восстановление зрительной функции людям с серьезными нарушениями зрения. Под восстановлением «зрительной функции» мы не обязательно подразумеваем, что слепой человек будет *видеть* так же, как зрячий. Напротив, это относится к способности быстро и точно передавать информацию, которую могут использовать слепые люди. Слепой легко мог бы носить камеру на лбу или в подвеске. Представьте себе алгоритм, который может называть каждый объект в поле зрения камеры (выделение и классификация объекта). Как мы можем передать такую обширную информацию слепому? Изображение стоит тысячи слов. Коротко говоря, мы получаем богатое представление о нашем окружении, которое сильно отличается от простого названия каждого объекта.

Это представление подчеркивает определенные особенности изображения, игнорируя при этом другую, менее важную информацию. Например, нас может не интересовать форма каждой ветви на ближайшем дереве, хотя мы могли бы получить доступ к этой информации, если бы захотели. Вместо этого нас больше интересует, не приближается ли к нам на полной скорости велосипед. С первого взгляда мы можем различать расстояния, отношения между объектами и даже действия и намерения. Даже если бы мы могли точно обозначить все объекты на изображении, визуальное ориентирование – это гораздо больше; к этой теме мы еще вернемся в конце данной главы. Основная задача помочь слепым – предоставить им *актуальную* информацию в режиме реального времени.

В качестве примечания: мы могли бы легко распространить эти идеи и на улучшение зрительных способностей людей с нормальным зрением. Легко представить камеру, которая дала бы нам немедленный доступ к 360-градусному обзору мира или предоставила бы нам доступ к другим частям светового спектра, к которым наши глаза нечувствительны, например к инфракрасному свету. Мы все «слепы» в инфракрасном и ультрафиолетовом диапазонах частот или со стороны затылка, но у нас есть инструменты, которые могут обнаруживать такого рода сигналы. Системы компьютерного зрения могут помочь нам анализировать и интерпретировать эти изображения. Следует отметить, что основные операции свертки, нормализации, субдискретизации и исправления (раздел 8.5) не зависят от того, исходят ли сигналы из видимой части спектра или из инфракрасных, ультрафиолетовых либо других источников. В общем, компьютерное зрение может помочь восстановить и, возможно, даже улучшить зрение человека.

## 9.7. ГЛУБОКИЕ СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ РАБОТАЮТ И ЗА ПРЕДЕЛАМИ ЗРЕНИЯ

Те же математические операции, которые используются для анализа изображений, полученных с фотографий, можно распространить и за пределы видимой части спектра. Кроме того, нет причин ограничиваться только светом. Хотя в центре нашего внимания находится обсуждение систем компьютерного зрения, интересно отметить, что та же математика, те же типы архитектур и одни и те же типы алгоритмов обучения вышли далеко за рамки зрения.

Распознавание образов привело к успеху в решении множества других задач. Например, систем распознавания речи; систем, предлагающих автоматические ответы на электронные письма; систем для прогнозирования погоды, фондового рынка или поведения потребителей. Многие другие задачи также были революционизированы глубокими сверточными нейронными сетями, изначально разработанными для распознания изображений<sup>1</sup>. Каждая из этих областей требует обучения с различными типами данных, изменения входных данных и, в некоторых случаях, внесения корректировок в сами архитектуры. Однако в основе этих областей за пределами зрения лежит аналогичная математическая задача: обучение нейронной сети научиться извлекать адекватные признаки из данных и затем классифицировать полученные особенности. Изменяется ввод: вместо использования пикселей в пространстве RGB в случае распознавания речи можно использовать спектrogramму частот звука как функцию времени для обработки звуков. Однако последующие этапы обработки и процедура обучения этих алгоритмов очень похожи, если не совсем идентичны, во многих приложениях.

В нейробиологии идея о том, что одинаковые вычислительные принципы могут использоваться для различных задач, иногда формулируется как «кора есть кора» (раздел 8.2), ссылаясь на гипотезу о том, что одни и те же основные архитектурные принципы соблюдаются в визуальной, слуховой и тактильной системах. Несомненно, существуют важные различия между модальностями, и инженеры также точно настраивают свои алгоритмы для каждого приложения. Однако в первом приближении некоторые из основных ингредиентов, похоже, подходят для множества, казалось бы, различных задач.

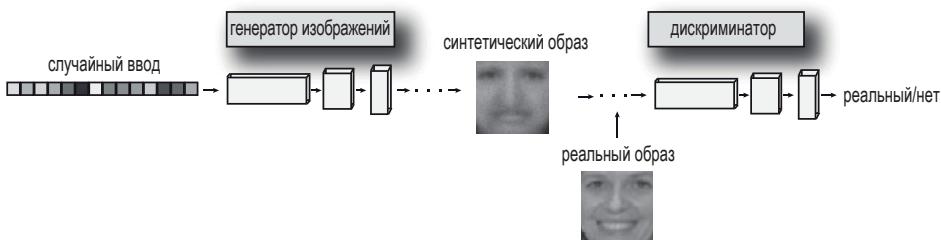
## 9.8. ГЕНЕРАТОРЫ ИЗОБРАЖЕНИЙ И GAN

Основная схема в большинстве приложений компьютерного зрения, которые мы обсуждали до сих пор, следует структуре, показанной на рис. 8.2. Изображение обрабатывается через нейронную сеть, которая учится извлекать его

<sup>1</sup> Одно из последних достижений – разложение ранее сведенной фонограммы на отдельные дорожки инструментов и голосов. Например, это позволило восстановить звук старых записей «Битлз» в фильме «Get Back» Питера Джексона. – Прим. ред.

особенности (признаки) для решения поставленной задачи. Еще одним замечательным достижением глубоких сверточных нейронных сетей была идея повернуть этот процесс в обратном направлении и использовать эти особенности для *генерации* изображений. Вычислительные модели, обсуждаемые до сих пор, представляют собой дискриминационные алгоритмы, которые присваивают описательные названия изображениям или их частям. Напротив, целью генеративных алгоритмов является не присвоение названия, а создание новой выборки из данного распределения. В контексте зрения это обычно сводится к созданию новых изображений или видео. Особенно успешным подходом к генерации изображений является использование генеративных состязательных сетей GAN, рис. 9.9.

Сети GAN состоят из двух основных компонентов: генератора изображений и дискриминатора изображений. Генератор изображений можно рассматривать как инвертированную глубокую сверточную нейронную сеть. В типичной глубокой сверточной нейронной сети входом является изображение, а выходом – набор признаков. В генераторе изображений входными данными является серия признаков, а выходными данными – изображение. Например, используя случайные исходные данные, целью может быть создание изображений реалистичных лиц. Дискриминатор изображений принимает на вход как реальные изображения, так и изображения, созданные генератором; его задача – определить, похоже изображение на настоящее или нет. Два компонента обучаются совместно: генератор пытается обмануть дискриминатор, а дискриминатор пытается поймать генератора – изготовителя фальшивок.



**Рис. 9.9 ♦** Генеративные состязательные сети (GAN) играют в игру «полиция против воров». Генеративная состязательная сеть – это алгоритм, который создает новые образцы из заданного распределения, например генерирует новые изображения. Алгоритм состоит из двух основных компонентов: генератора изображений и дискриминатора изображений. Генератор можно рассматривать как инвертированную глубокую сверточную нейронную сеть, использующую признаки изображений в качестве входных данных и создающую изображения в качестве выходных. Дискриминатор берет образцы от генератора и реальных изображений и определяет, соответствуют сгенерированные изображения настоящим или нет

Такие генераторы изображений нашли забавное применение в нескольких областях. Одна из этих областей – перенос стиля. Можно взять произвольный снимок и пересказать его в стиле известной картины. Можно использовать

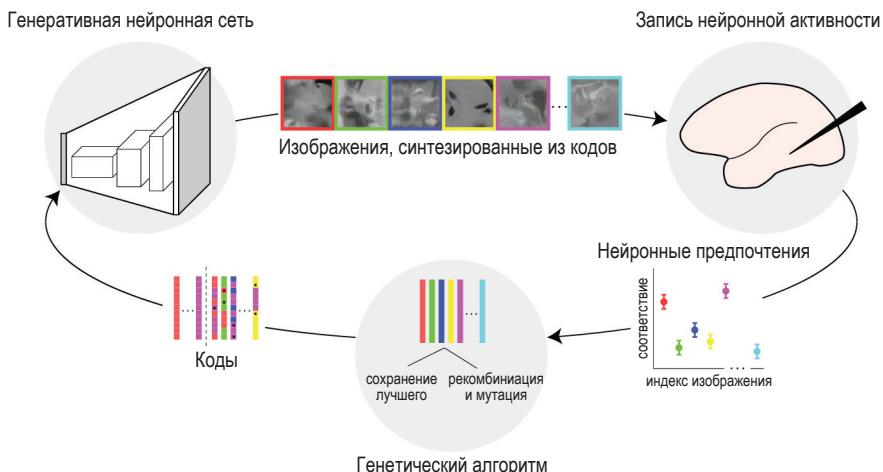
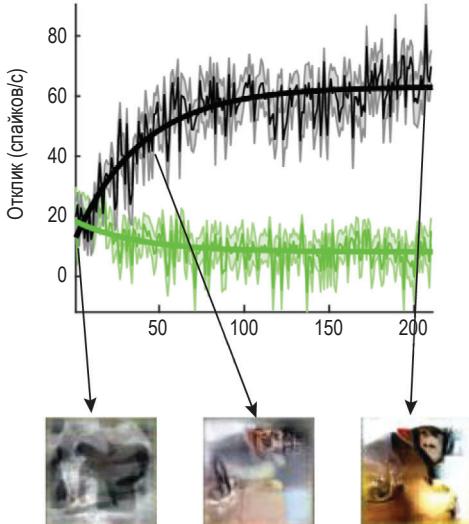
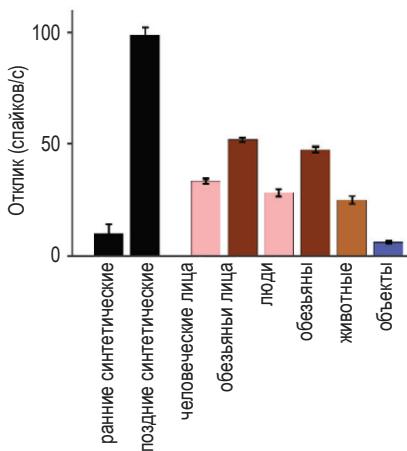
GAN, чтобы объединить разные лица, сделать лицо похожим на знаменитость или визуализировать, как данный человек может выглядеть, когда он или она станет старше. Другое приложение – графическое искусство. Недавно изображение «Портрет Эдмона Белами», созданное GAN, было продано Christie's за серьезную сумму в размере 432 500 долларов.

Другие GAN сосредоточились на попытках создавать реалистичные фотографии. На самом деле наивному глазу бывает сложно отличить подделку от настоящей фотографии. Эти алгоритмы вызывают большую заинтересованность не только у Голливуда. Поговорка «верь глазам своим» может потребовать серьезного пересмотра в эпоху изощренных цифровых подделок.

## 9.9. DEEP DREAM И XDREAM: НАСТРОЙКИ ВЫЧИСЛИТЕЛЬНЫХ БЛОКОВ И БИОЛОГИЧЕСКИХ НЕЙРОНОВ

Особенно интересным является использование генераторов изображений, чтобы помочь справиться с проклятием размерности при изучении свойств настроек нейронов в зрительной коре головного мозга (рис. 5.10). Семейство методов, первоначально называвшееся поэтическим названием *DeepDream*, было введено разработчиками для визуализации типов изображений, предпочтаемых блоками в глубоких сверточных нейронных сетях. При рассмотрении этих нейронных сетей мы знаем архитектуру и все веса; другими словами, мы можем математически точно определить активацию каждого блока. В этих условиях мы можем инверсировать процесс, чтобы спросить, какие типы изображений приведут к высокой активации для данного блока. Здесь «функция потерь» – это единичная активация (которая должна быть максимизирована), и мы все еще можем применять алгоритм градиентного спуска, представленный в разделе 8.6, за исключением того, что мы вычисляем производные по отношению к самому изображению вместо изменения весов сети.

Теперь представьте, что мы хотим сгенерировать изображения, которые максимально активируют нейрон в мозге, а не блок в нейронной сети. Ситуация намного сложнее, когда дело доходит до нейронных сетей в биологическом мозге, где мы не знаем архитектуры, не говоря уже о весах. Чтобы обойти эти проблемы, Уилл Сяо и его коллеги разработали алгоритм *XDream* (расширение DeepDream с эволюцией в реальном времени для максимизации активации, рис. 9.10), который был кратко представлен в разделе 6.4. Алгоритм состоит из трех компонентов: (i) генератора изображений, (ii) механизма оценки соответствия каждого изображения и (iii) метода поиска для создания следующего набора изображений (рис. 9.10A). Генератор изображений представляет собой инвертированную глубокую сверточную нейронную сеть, аналогичную алгоритмам, представленным в предыдущем разделе. Генератор изображений принимает набор признаков в качестве входных дан-

**A****B****C**

**Рис. 9.10** ♦ Генераторы изображений могут помочь беспристрастно исследовать поведение нейронов. (A) Многообещающим недавним применением генераторов изображений является разработка алгоритмов с обратной связью для исследования настройки нейронов. Схема алгоритма XDream, состоящего из генератора изображений, нейронных записей и генетического алгоритма. (B) Частота срабатывания нейрона нижней височной коры увеличивается с каждой итерацией алгоритма XDream (синтетические изображения, черный цвет), создавая изображения, которые лучше, чем эталонные естественные изображения. (C) В то время как средний отклик этого нейрона на естественные изображения может привести некоторых исследователей к выводу о настройке на лица, синтетические изображения вызывают еще более высокую частоту срабатывания

ных и создает цветное изображение. Начальные условия – случайные изображения. Затем алгоритм оценивает изображения, созданные генератором, и упорядочивает их в соответствии с функцией соответствия, определяемой тем, что мы хотим максимизировать. Например, алгоритм может максимизировать активацию конкретного блока в сети, среднюю активность всех блоков в данном слое или стандартное отклонение активности блоков внутри слоя. В неврологии функцией соответствия может быть частота срабатывания данного нейрона в ответ на изображения (как показано в нескольких примерах в гл. 5 и 6). После ранжирования изображений на основе функции соответствия XDream использует алгоритм генетического поиска для выбора, удаления и рекомбинации начального набора признаков для создания нового раунда изображений. Важно отметить, что XDream не делает никаких априорных предположений о настройке нейронов и не требует каких-либо знаний об архитектуре или весах нейронной сети либо мозга; алгоритм требует только способа оценки степени соответствия для каждого изображения.

XDream может визуализировать признаки, представляемые блоками в нейронных сетях. Он может обнаруживать изображения, которые вызывают высокую активацию – экстраполируя на разные уровни, разные архитектуры и даже разные режимы обучения. Примечательно, что XDream также очень эффективен в обнаружении изображений, которые вызывают высокую активацию реальных биологических нейронов (раздел 6.4). Без каких-либо предположений о корковых связях или установок о предпочтениях нейронов и в рамках ограничений, вводимых биологическими записями сигналов, алгоритм генерирует изображения, которые запускают высокую частоту срабатывания (рис. 9.10В). Эти синтетические изображения оказываются столь же эффективными, а в некоторых случаях более эффективными, чем случайные естественные изображения, которые десятилетиями использовались в нейробиологии (рис. 9.10С).

## 9.10. РАЗМЫШЛЕНИЯ О ПЕРЕКРЕСТНОЙ ПРОВЕРКЕ И ЭКСТРАПОЛЯЦИИ

В этой главе мы выделили некоторые замечательные достижения алгоритмов компьютерного зрения. Теперь сменим тему и рассмотрим некоторые критические для современных алгоритмов проблемы и некоторые захватывающие возможности, которые ждут нас впереди. Начнем с критического вопроса генерализации (обобщения). В разделе 8.8 мы представили концепцию перекрестной проверки. Чтобы снизить риск перебора и самообмана, полагая, что наши алгоритмы лучше, чем они есть на самом деле, очень важно разделить данные на обучающий и независимый наборы тестов.

Что не совсем понятно в большинстве приложений компьютерного зрения, так это то, насколько набор тестов должен отличаться от обучающего набора. В большинстве типичных сценариев у нас есть большая база данных, и мы случайным образом выбираем некоторые изображения для обучения,

а остальные для тестирования. То, насколько нас удовлетворят результаты, во многом зависит от того, насколько набор тестов будет отличным от обучающего набора. В тривиальном примере ранее мы упоминали о потенциальной проблеме дублирования изображений в базах данных (раздел 8.8). Предположим, что изображение 5000 и изображение 8000 на самом деле идентичны, и предположим, что случайный выбор вносит изображение 5000 в обучающий набор, а изображение 8000 – в контрольный. Конечно, это не настоящая перекрестная проверка, и правильная классификация изображения 8000 не должна считаться достижением алгоритма. В чуть более сложном примере предположим, что теперь изображение 8000 идентично изображению 5000, за исключением одного пикселя, или что изображение 8000 является слегка обрезанной версией изображения 5000. Хотя мы можем следовать всем правилам перекрестной проверки и адекватно разделить изображения в независимый набор тестов, адекватная оценка эффективности проблематична, если тестовые изображения очень похожи на изображения в обучающем наборе.

Есть более тонкие и пагубные варианты этой проблемы. Многие базы данных основаны на изображениях из интернета. В изображениях, которые люди загружают в интернет, могут быть сильные предубеждения и ложные корреляции. Например, представьте, что мы хотим построить алгоритм распознавания Пизанской башни в Италии. Туристы, посещающие Пизу, часто фотографируют знаменитую башню и загружают эти фотографии в интернет. Есть множество позиций, с которых можно сфотографировать Пизанскую башню, и количество туристов очень велико (порядка  $10^6$  туристов ежегодно). Люди делают эти снимки с самых разных ракурсов, но не со всех возможных. Например, люди могут снимать башню с определенных улиц, могут быть определенные места, где люди обычно сидят, но мало кто использует дроны для съемки с воздуха. Может быть разница и в отношении того, что именно содержат изображения (например, большинство людей фотографируют всю башню, а не ее части; многие фото содержат участки газона вокруг башни). Может быть разница в цвете неба, окружающего башню (например, в солнечный день намного больше снимков, а во время грозы – очень мало). Собрать все изображения Пизанской башни и выполнить соответствующую перекрестную проверку, чтобы убедиться, что тестовые изображения не слишком похожи на изображения в обучающей выборке, является сложной задачей. Если перекрестная проверка не будет проведена чрезвычайно тщательно, алгоритм может достичь высокой точности в распознавании Пизанской башни, но при этом потерпит неудачу с необычным снимком, сделанным с дрона в дождливый день. Другими словами, алгоритм легко подгоняется под обучающие данные, несмотря на наши лучшие намерения и все усилия по разделению обучающих и тестовых наборов данных.

Эта проблема не ограничивается известными достопримечательностями. Например, многие люди любят хвастаться приготовленной едой, загружая фотографии в социальные сети. Рассмотрим все фото омлетов в сети. Они в основном сняты с одного и того же ракурса? Омлеты обычно на тарелке? На многих фотографиях тарелка белая? Большинство снимков сделаны при более или менее равномерном освещении кухни? В некоторых из них есть

вилки и ножи? Но скажите, сколько в сети есть фотографий омлета, свисающего с ветки дерева в парке в дождливый день?

Еще один пример из этого семейства проблем можно почерпнуть из задачи распознавания действий, показанной на рис. 9.4. Кадры на рис. 9.4A взяты из хорошо известной базы данных видео для распознавания действий, UCF101. Без какой-либо сложной обработки, используя только отдельные кадры и информацию на уровне пикселей, можно сделать вывод, что если изображение содержит много синих пикселей, оно, вероятно, соответствует «брассу», тогда как если изображение содержит много зеленых пикселей, скорее всего, это соответствует «жонглированию мячом». Другие действия также содержат много синего или зеленого цвета, но тем не менее в этой задаче можно добиться значительно более высокой эффективности без какого-либо острого понимания изображений, не говоря уже о понимании того, что означают названия действий. Напротив, контролируемые наборы данных, показанные на рис. 9.4B, значительно сложнее: здесь задача состоит в том, чтобы определить, пьет человек или нет. Есть много разных способов питья (из чашки, из бутылки, из соломинки, из рук в качестве сосуда, из питьевого фонтанчика). Настоящий классификатор действий, способный различать картинки, изображающие, как кто-нибудь пьет, должен уметь обобщать все эти условия. Мы не можем добиться значительно более высокой эффективности в задаче на рис. 9.4B, просто рассматривая количество синих пикселей. Превышение вероятности при классификации на уровне пикселей является хорошим показателем того, что задача слишком проста, что есть сильное сходство между обучающими и тестовыми изображениями и что база изображений может иметь значительную степень избыточности.

Из-за этих типов корреляций в изображениях баз данных контекстная информация в алгоритмах компьютерного зрения имеет тенденцию играть заметную роль. Алгоритмы могут адекватно вывести правильное название, даже если сам объект полностью закрыт, исключительно на основе статистики контекстной информации. Например, светофоры, как правило, сняты на улице и обычно располагаются в верхней части изображения. Хотя это можно рассматривать как выгодное использование статистики изображений, верно и обратное: нейронные сети могут неправильно классифицировать объект, помещенный вне контекста. Контекстная информация может помочь и людям (раздел 3.7); однако люди, как правило, более невосприимчивы к манипуляциям с изображениями, таким как размещение объектов вне контекста.

Не все реальные приложения зависят от генерализации. Например, если Facebook желает автоматически присвоить тег изображениям Пизанской башни, загруженным его пользователями, Facebook может достичь 99%-ной точности, упустив только аэрофотоснимки, сделанные во время грозы. Другие приложения могут критически нуждаться в подготовке к неожиданностям. Мы хотим, чтобы беспилотные автомобили могли распознать корову, пересекающую шоссе, даже если это случается крайне редко.

Проблема перекрестной проверки связана с вопросом систематической ошибки в базах данных для обучения (в сообществе специалистов по компьютерному зрению это называется систематической ошибкой базы данных).

Например, предположим, что мы создаем алгоритм для обнаружения опухолей молочной железы с использованием маммограмм от белых женщин в возрасте от 50 до 60 лет, которые живут в Калифорнии. Будет ли алгоритм работать с белыми женщинами того же возраста из Массачусетса? А из Европы? Будет ли алгоритм работать с афроамериканками или азиатками? Будет ли алгоритм работать с женщинами от 30 до 80? Проблема предвзятости в обучающих данных недавно была освещена в новостях в связи с задачами систем распознавания лиц, которые лучше справлялись с определенными этническими группами, чем с другими.

Следует отметить, что проблема предубеждений характерна не только для компьютерного зрения. Предубеждения зрительного распознавания распространены и в человеческом зрении. Радиологи, обученные распознавать рак груди на маммограммах белых женщин в возрасте 50 лет, также могут потерпеть неудачу при тестировании на маммограммах других групп женщин. В случае идентификации лиц существуют хорошо известные человеческие предубеждения, основанные на том, где люди растут и в какой степени им приходилось сталкиваться с лицами разных этнических групп.

Генерализация – важное и желаемое свойство вычислительных алгоритмов. Способность делать выводы на основе перекрестно проверенных данных недостаточно хорошо определена и зависит от того, насколько четким, неизбыточным является набор тестов. Один из способов попытаться количественно оценить эту проблему – провести различие между интерполяцией (обобщение внутри распределения) и экстраполяцией (обобщение вне распределения). Опять же, что именно подразумевается под распределением, не совсем точно определено, но, по крайней мере, это дает возможность начать количественную оценку способности алгоритмов экстраполировать за пределы своего обучающего набора.

## 9.11. СОСТАЗАТЕЛЬНЫЕ ИЗОБРАЖЕНИЯ

Мы выделили некоторые из захватывающих достижений в том, как вычислительные алгоритмы обрабатывают изображения и как машинное зрение может соответствовать или даже превосходить эффективность человека во многих задачах. Однако следует проявлять осторожность, прежде чем думать, что машины вот-вот пройдут общий визуальный тест Тьюринга. Есть еще много визуальных задач, которые машины не могут решить. Кроме того, в визуальных задачах машины относительно легко обмануть (например, рис. 9.4).

Одним из примеров сбивающего с толку поведения глубоких сверточных нейронных сетей является случай *состязательных изображений* (*adversarial images*), когда минимальные изменения изображения резко меняют прогнозируемый класс<sup>1</sup> (рис. 9.11). Состязательные изображения кажутся людям

<sup>1</sup> К «состязательным изображениям» относятся такие изображения, которые содержат умышленно измененные пиксели, с целью вызвать некорректную работу нейронной сети, но при этом не имеющие существенных для человека изменений. Иногда их называют «вредоносными изображениями». – Прим. ред.

похожими, почти идентичными, но они определяются системой компьютерного зрения как разные. Например, два изображения на рис. 9.11 практически неотличимы для наблюдателей, однако глубокая сверточная сеть правильно классифицировала левое изображение как «кукурузу» и неверно пометила изображение справа как «шноркель». Учитывая алгоритм, который вынужден назначать название изображению из двух вариантов, А или В, неизбежно будет граница, на которой, с совсем незначительными изменениями изображения, мы перейдем от А к В. Разделение между двумя названиями в пространстве изображений сродни тому, как если бы мы стояли на весьма неопределенной границе между двумя состояниями, например пытаемся точно определить, где начинается дождь, когда переходим от точки А к точке В. Эти противоречавшие образы обычно создаются с использованием знаний о границах категорий, и, умышленно изменяя несколько пикселей, мы можем подтолкнуть изображение к альтернативному названию. Как и в алгоритме DeepDream, представленном в разделе 9.10, процесс создания состязательных изображений включает градиентный спуск на пикселях самого изображения.

988: кукуруза



801: шноркель



**Рис. 9.11** ♦ Состязательные изображения ошибочно классифицируются вычислительными алгоритмами, однако человеческому мозгу они кажутся неотличимыми. Два этих изображения кажутся неотличимыми для людей. Однако современные компьютерные алгоритмы классифицируют левое как «кукурузу», а правое – как «шноркель». Изображение справа было создано путем добавления небольшого количества шума к изображению слева, вдоль определенных направлений

Что интригует в таких примерах, так это глубокое различие между машинным и человеческим восприятием. Во многих реальных приложениях видение мира, как это делают люди, может быть весьма актуальным. Фактически была целая индустрия исследователей, разрабатывающих «состязательные атаки», чтобы сбить с толку системы компьютерного зрения, вместе со столь же энергичным сообществом защитников от таких состязательных атак. Например, можно спросить, вернется ли изображение справа на рис. 9.11 обрат-

но к кукурузе после его масштабирования, изменения цвета, использования разных версий одной и той же сети (начиная с разных случайных начальных условий) или с использованием разных архитектур. Эти примеры ясно показывают, что даже когда современные алгоритмы могут правильно называть множество изображений, современные глубокие сверточные нейронные сети не обязательно видят мир так, как это делают люди.

Состязательные примеры не уникальны для компьютерного зрения. Люди также страдают от таких враждебных примеров, просто создать их для людей намного сложнее, потому что мы не можем вычислять градиенты в своих биологических сетях, как это делают искусственные нейронные сети. Даже без таких градиентов психологи обнаружили множество изображений, которые сбивают людей с толку. Люди склонны ошибаться во многих визуальных иллюзиях, которые заставляют нас видеть несуществующие вещи (гл. 3).

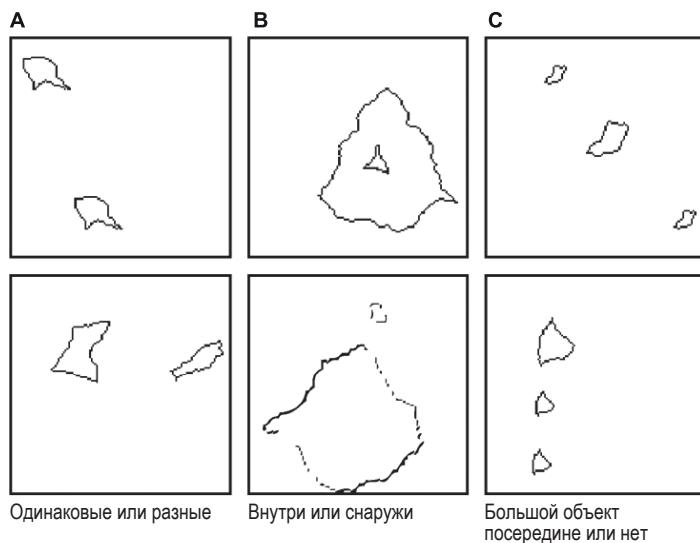
В общем, люди и современные системы компьютерного зрения допускают схожие ошибки в задачах классификации объектов (раздел 8.12). Однако многие изображения могут обмануть системы компьютерного зрения, а не людей, и наоборот. Эти результаты показывают, что даже наши лучшие системы компьютерного зрения все еще не полностью учитывают возможности зрительного распознавания человека. Поскольку можно обнаружить такое двойное расхождение между машинным и человеческим зрением, эти результаты показывают, что современные глубокие сверточные нейронные сети все еще не могут пройти визуальный тест Тьюринга. Мы можем легко отличить машину от человека, показав изображение справа на рис. 9.11.

## 9.12. ОБМАНЧИВО ПРОСТЫЕ ЗАДАЧИ, БРОСАЮЩИЕ ВЫЗОВ АЛГОРИТМАМ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Состязательные изображения специально созданы для того, чтобы обмануть вычислительные алгоритмы. Также возможно бросить вызов вычислительным алгоритмам в основных визуальных задачах, которые не предназначены для конкретной цели перемещения изображений через границы категорий. Хотя есть много визуальных задач, по которым компьютеры превосходят людей, например при чтении штрих-кодов, есть также много простых визуальных задач, где компьютеры легко обмануть (рис. 9.11).

Многие визуальные задачи, легкие для человека, представляют собой серьезную проблему для современных архитектур. Рассмотрим примеры на рис. 9.12, взятые из набора 23 задач на визуальное мышление, предложенных группой Дона Гемана. Имея набор положительных (верхняя строка) и отрицательных (нижняя строка) примеров, нам нужно выяснить, какое правило дает возможность классифицировать новые изображения. Люди быстро понимают, что это правило «одинаковы или отличаются», с учетом перемещения, две фигуры на рис. 9.12A, «внутри или снаружи» на рис. 9.12B

и то, находится ли самая большая из трех фигур между двумя другими или нет, на рис. 9.12С. Даже если люди никогда раньше не видели этих конкретных примеров и задач, они могут быстро вывести правила. Затем люди могут использовать эти правила, чтобы рассуждать о новых примерах. Группа Томаса Серра показала, что современные модели компьютерного зрения не справляются с этими задачами, несмотря на обширную подготовку с использованием до миллиона примеров.



**Рис. 9.12** ♦ Некоторые, казалось бы, простые задачи бросают вызов существующим алгоритмам. Задача включает в себя обучение классификации изображений на две группы в соответствии с определенными фиксированными, но неизвестными правилами. Здесь показаны три типа правил: (A) одинаковые или разные, (B) внутри или снаружи, (C) большой объект посередине или нет. Положительные примеры показаны в верхней строке, а отрицательные – в нижней. Воспроизведено по Флере и др., 2011

Связанным примером является база данных CLEVR, состоящая из изображений, содержащих несколько геометрических форм, таких как сферы, кубы и цилиндры различных размеров, цветов и свойств поверхности. Задача включает в себя ответы на такие вопросы, как, например, больше ли красный цилиндр слева от синего куба, чем красный цилиндр справа от синего куба, или совпадает ли количество больших объектов с количеством металлических объектов. Современные сети, похоже, адекватно учатся отвечать на эти вопросы, когда их обучают и тестируют на одних и тех же комбинациях форм и цветовых свойств. Однако при тестировании с новыми комбинациями форм и цветов (например, когда сеть никогда не сталкивалась с синим цилиндром во время обучения, хотя она видела много синих кубов и много красных цилиндров) сети не смогли сделать обобщения.

## 9.13. ВЫЗОВЫ, КОТОРЫЕ СТОЯТ ВПЕРЕДИ

Был достигнут значительный прогресс в обучении компьютеров зрению. Мы уже окружены машинами, которые могут успешно использовать алгоритмы автоматического зрения в реальных приложениях. Поразительный прогресс в области компьютерного зрения позволяет нам думать, что эту задачу мы уже почти решили. Действительно, известные газеты предлагали заголовки с заявлениями, намекающими на то, что компьютерное зрение – дело решенное. Однако я бы сказал, что мы все еще очень далеки от прохождения общего визуального теста Тьюринга и что лучшие достижения еще впереди.

В дополнение к некоторым задачам, обсуждавшимся в предыдущих разделах (состязательные изображения, генерализация, визуальное обоснование в простых задачах), быстро развиваются и подчеркивают прогресс задачи, связанные с комментариями к изображениям (также связанные с системами ответов на вопросы об изображениях). Цель этих задач – дать краткое и «релевантное» описание<sup>1</sup>. В отличие от задач категоризации, количественно оценить результаты сложнее. Кроме того, эти задачи могут смешивать зрение и язык, как это указано в начале главы. Тем не менее алгоритмы комментариев к изображениям представлены в обзоре, завершающем данную главу, и при этом подчеркивают интересные задачи, стоящие перед нами в этой области.

На рис. 9.13 показан современный пример создания комментариев к изображениям, который основан на результатах, полученных с помощью бота для создания подписей, комментариев к изображению (примерно ноябрь 2018 г.). Важно указать дату, потому что я подозреваю, что в ближайшие годы мы увидим серьезные улучшения. Комментарии, предоставляемые этим алгоритмом, впечатляют. Система хороша для обнаружения людей, даже для количественной оценки того, содержит изображение одного человека (рис. 9.13A) или несколько людей (рис. 9.13D). Система также может определять пол на рис. 9.13A и делает разумное предположение о том, счастливы ли люди на рис. 9.13D (я сам запечатлен на этой картинке и могу свидетельствовать, что был очень счастлив, подозреваю, как и большинство людей, посещающих Пизанскую башню). Система также правильно делает вывод, что человек сидит на рис. 9.13A и стоит на рис. 9.13D. Кроме того, система обнаруживает другие особенности сцены, включая наличие стола на рис. 9.13A, воды на рис. 9.13B и здания на рис. 9.13D. Многие другие объекты не описаны, что, возможно, разумно, учитывая, что цель состоит в том, чтобы подписать изображение, а не упомянуть каждый отдельный объект. Еще одно предостережение при использовании подписи к изображениям в качестве испытательной задачи для компьютерного зрения состоит в том, что мы не знаем, не упоминаются ли конкретные объекты только потому, что они не были обнаружены, или потому, что алгоритм посчитал эти объекты не слишком важными.

<sup>1</sup> То есть максимально соответствующее изображению. – Прим. ред.



**Рис. 9.13** ❖ Успехи и проблемы при создании подписей к изображениям.  
Четыре примера результатов из системы создания подписей к изображениям

Немного удивительно, что система не описывает Пизанскую башню на рис. 9.13D, учитывая, что такие памятники содержатся в непомерном объеме обучающих данных. Еще более удивительно, на рис. 9.13A присутствует довольно заметная ложка, которая не была описана. Также кажется вероятным, что многие люди описали бы невесту на рис. 9.13B. Система не может работать с линейными рисунками (рис. 9.13C), но приятно, что алгоритм смог осознать свои ограничения и признать, что он не может описывать линейные рисунки. Отличить линейные рисунки от фотографий, возможно, не так уж сложно, особенно если изображение имеет значительное количество белых пикселей, несколько черных пикселей и практически не имеет текстур. Людям относительно легко понять, что на рис. 9.13C изображены три человека, хотя не совсем ясно, как именно происходит это умозаключение. Современные алгоритмы для создания подписей к изображениям, подобные проиллюстрированному здесь, возможно, имеют малый опыт обучения с по-

мощью рисунков, если он вообще есть. Напротив, большинство людей очень часто сталкивались с символикой, лежащей в основе контурных рисунков.

Один из простых способов обмануть эти системы подписей – зашифровать изображение. Например, мы можем разделить изображение на четыре квадранта и случайным образом их переставить. Изображение по большей части теряет смысл, однако подписи практически не меняются. Если мы представим фотографию глазного дна с рис. 9.7 (только фотография глазного дна без остальной части рисунка), система ответит: «Я не могу описать изображение, но я вижу свет, сидящего, лампу». Похвально, что система понимает, что она не может полностью описать изображение, что изображение отличается от ее обучающего набора. На изображении действительно есть свет. Система, вероятно, видела много примеров, когда слово «свет» соотносилось со словом «лампа», добавляя его в описание.

В нашем примере немного сложнее понять, откуда взялось слово «сидеть». Проблема объяснения того, откуда берутся названия, характерна для глубоких нейронных сетей, ее многие люди критиковали. Учитывая большое количество параметров в системе, не всегда легко выразить словами, почему система выдает тот или иной результат. Люди могут придумать объяснения постфактум, но не всегда легко понять, как они к ним пришли. Радиологи не склонны подробно объяснять, как они ставят диагноз, и от них, конечно же, не требуется придумывать объяснение того, что делают нейроны в их мозге. Людям было бы сложно дать механистическое объяснение того, почему они думают, что видят дерево, как на рис. 8.1.

Следует отметить, что тот же тип архитектуры, который используется в комментариях к изображениям, можно обучить, чтобы превзойти врачей в интерпретации одних и тех же фотографий глазного дна. Те же самые архитектуры можно обучить обнаружению Пизанской башни. Каждый из этих вопросов требует отдельных шагов обучения. Напротив, врач может оценить фотографии глазного дна, а также понять, что происходит на рис. 9.13, в то время как многие современные глубокие сверточные сети являются сверхспециализированными для конкретных задач, и обучить нейронные сети выполнять несколько задач непросто.

Для прохождения визуального теста Тьюринга необходимо уметь ответить на любой вопрос об изображении, а не просто быть обученным отвечать на вопросы одного типа. Понятно, что можно задать много вопросов об изображениях на рис. 9.13. Какими бы впечатляющими ни были эти подписи, они даже близко не подходят к решению теста Тьюринга на зрение. В комментариях совершенно не отражены фундаментальные особенности сцены, что происходит, кто, что с кем и почему делает. Люди могут смотреть на эти изображения и понимать отношения между различными объектами, их относительное положение и почему они находятся там, где находятся, и даже делать выводы о том, что произошло раньше или что может произойти дальше.

Что еще более интригующе, все эти изображения должны быть несколько любопытными или забавными. В заключение я хотел бы выделить пример проблемы, которую считаю чрезвычайно сложной: понимание человеческого чувства юмора на основе изображений. Конечно же, несмотря на то

что понятие смешного является субъективным и зависит от возраста, пола и культурного происхождения, между разными людьми существуют сильные корреляции в том, что смешно, а что нет.

Давайте рассмотрим рис. 9.13С в качестве примера. Что смешного в этом изображении? Чтобы понять, что происходит на изображении, нам нужно включить не просто информацию на уровне пикселей, не только названия конкретных объектов, но также их символизм и относительные взаимодействия. Квадратная шапочка вместе с другими признаками, представляющими одежду человека в центре, плюс его относительное положение по отношению к другим людям, заставляет нас думать, что он судья. Обратите внимание, что именно сочетание многих из этих признаков, в их совокупности, приводит нас к такому пониманию. Каждого отдельного признака было бы недостаточно. Человек, сидящий ниже судьи, скорее всего, является обвиняемым (или, что менее вероятно, свидетелем). Этот вывод частично основан на рубашке человека с горизонтальными полосами, но в основном на его относительном положении и понимании расположения судьи и обвиняемого в суде. Мы можем сделать вывод, что третье лицо – полицейский, что согласуется с его одеждой, но также с тем фактом, что он стоит и что находится позади обвиняемого.

Расшифровав, что человек в центре – судья, мы понимаем, что он держит молоток, он кричит и бьет молотком по столу. Обвиняемый тоже злится, глядя судье в глаза. Что любопытно – то, что обвиняемый, похоже, тоже держит молоток. Это наблюдение кажется нам необычным: обвиняемый не должен держать молоток, не говоря уже о том, чтобы пользоваться им. Отклонение от нормы – вот суть того, почему изображение смешное: оно изображает неожиданный сценарий. Если мы уберем несколько пикселей, изображающих молоток обвиняемого, изображение сразу станет менее интересным. Конечно, юмор субъективен и может варьироваться от человека к человеку.

Даже если люди не сочтут рис. 9.13С забавным, они все равно могут понять всю символику, действия, кто эти люди и как они связаны друг с другом. Независимо от того, забавно то или иное изображение, люди могут интерпретировать то, что происходит на рис. 9.13С, в первый же раз, когда они видят это изображение. Людям не требуется обширная подготовка с черно-белыми рисунками людей в суде, чтобы понять это изображение. Нам необходим значительный объем мировых знаний, чтобы понять и интерпретировать рис. 9.13С. Предсказать, является ли изображение забавным или нет, еще больше усложняется тем фактом, что даже если бы мы обучили алгоритм понимать всю символику на рис. 9.13С, это не помогло бы ни понять, почему рис. 9.13А интригует, ни сделать вывод, что, возможно, произошло на рис. 9.13В.

Существуют тривиальные, грубые и в конечном итоге неинтересные решения, которые могут обеспечить превосходную эффективность в задаче различения смешного и несмешного. Добавление большого количества изображений, подобных тем, что на рис. 9.13, в глубокую сверточную сеть, обученную с помощью контролируемого обучения, может привести к некоторой способности расшифровывать, смешно это или нет, более чем в 50 %

случаев. Например, среди забавных изображений много мультфильмов или рисунков. Система может быстро научиться отличать рисунки от настоящих фотографий. Если рисунки соотносятся с более «забавными» надписями, тогда система может работать достаточно хорошо. Однако на самом деле модель ничего не знала бы о юморе. Удаление молотка у обвиняемого на рис. 9.13С не изменит название для этого типа модели, даже несмотря на то, что эта простая манипуляция радикально меняет забавность изображения. Эта манипуляция с изображениями – очередной пример проблем с подгонкой и предвзятостью, подробно описанных в разделе 9.11. Хорошо контролируемая визуальная задача должна гарантировать, что названия не коррелируют с какими-либо другими свойствами, кроме тех, которые изучаются.

Определение того, является ли изображение забавным или нет, иллюстрирует современные проблемы, связанные с включением дополнительных знаний в визуальную обработку. Однако стоит отметить, что нет физических ограничений на то, что компьютеры могут делать. Если мы можем это сделать, то сможет и компьютер. За последнее десятилетие был достигнут значительный прогресс в обучении компьютеров выполнению множества задач, которые традиционно считались прерогативой исключительно людей. Любой настольный компьютер может играть в шахматы на соревнованиях, а лучшие компьютеры могут побеждать чемпионов мира по шахматам. IBM Watson преуспела в похожей на мелочи игре Jeopardy. Более того, будучи несовершенной, Siri и связанные с ней системы делают огромные успехи, становясь лучшими помощниками в мире. В области зрения вычислительные алгоритмы уже способны выполнять определенные задачи, такие как распознавание цифр в полностью автоматическом режиме на уровне человеческих возможностей, разделение изображений из интернета на 1000 различных категорий, обнаружение лиц в кадре для их съемки, распознавание лиц для входа в систему на смартфоне или анализ клинических изображений, задачи в области астрономии, и многое другое. Хотя люди по-прежнему превосходят самые сложные современные алгоритмы в большинстве визуальных задач, разрыв между способностями машин и человеческим зрением быстро сокращается.

Значительный прогресс был достигнут в описании распознавания зрительных образов в рамках строгой теории. Однако существуют заметные пробелы в нашем понимании функциональной и вычислительной архитектуры центральной зрительной коры. Предварительные шаги позволили выявить важные принципы вычислений в неокортексе, включая глубокие нейронные цепи, которые могут разделять и решать сложные задачи, и восходящие цепи, которые делают быстрые вычисления за счет постепенного повышения избирательности и толерантности к трансформации объектов. В отличие от пути от сетчатки до первичной зрительной коры, у нас нет количественного описания предпочтений нейронов вдоль центрального зрительного пути. Более того, некоторые вычислительные модели не дают четких и проверяемых прогнозов относительно системных характеристик центральной зрительной коры на физиологическом уровне. Вычислительные модели могут выполнять некоторые сложные задачи распознавания. Однако для подавляющего

большинства задач распознавания машинное зрение по-прежнему значительно уступает человеческим возможностям. Следующие несколько лет, скорее всего, принесут много новых сюрпризов в этой области. Мы сможем описать схемы зрительной коры с беспрецедентным разрешением на экспериментальном уровне, сможем оценивать сложные теории, требующие больших вычислительных ресурсов, в течение вполне реалистичного времени. Точно так же, как молодое поколение не удивляется машинам, которые могут играть в шахматы на соревнованиях не хуже людей, следующее поколение вряд ли удивится интеллектуальным устройствам, которые видят мир так же, как мы.

## 9.14. ОБЗОР ГЛАВЫ

- Машина пройдет визуальный тест Тьюринга, если мы не сможем отличить ее ответы от человеческих ответов на любой произвольный вопрос о любом изображении.
- Компьютерное зрение показало замечательные успехи в различных задачах, включая классификацию объектов, обнаружение объектов, выделение объектов на изображении и классификацию действий.
- Успех в визуальных задачах привел к появлению множества реальных приложений, включая алгоритмы распознавания лиц, визуальную интерпретацию окружающей обстановки для беспилотных автомобилей, анализ клинических изображений, классификацию галактик по астрономическим снимкам и многое другое.
- Инвертирование сверточных сетей открыло двери для алгоритмов, генерирующих синтетические изображения. Одно из применений генераторов изображений – систематическое изучение свойств настройки нейронов вдоль центральной зрительной коры.
- Несмотря на быстрый прогресс, приложения компьютерного зрения остаются весьма уязвимыми. Алгоритмы относительно легко обмануть, и есть много задач, которые просты для человека, но очень сложны для машин, например определение того, находится ли некая форма внутри или снаружи другой.
- Из-за большого количества параметров часто не ясно, насколько хорошо современные алгоритмы компьютерного зрения могут экстраполировать свои возможности обработки изображения на новые сценарии, в отличие от простой интерполяции между обучающими выборками. Генерализация является важным требованием для будущих алгоритмов машинного зрения.
- Остается еще много интересных задач, для решения которых нужно научить компьютеры видеть и интерпретировать мир так, как это делают люди. В качестве примера сложной задачи – обучение систем компьютерного зрения тому, как определять, является изображение забавным или нет; это выходит далеко за рамки возможностей существующих систем.

## ЛИТЕРАТУРА

См. дополнительные ссылки на <http://bit.ly/2t53QRd>.

- Lotter, W.; Kreiman, G.; and Cox, D. (2020). A neural network trained for prediction mimics diverse features of biological neurons and perception. *Nature Machine Learning*. 2: 210–219.
- Poplin, R.; Varadarajan, A.; Blumer, K.; et al. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering* 2: 158–164.
- Russakovsky, O.; Deng, J.; Su, H., et al. (2014). ImageNet Large Scale Visual Recognition Challenge. In: CVPR: 1409.0575.
- Szegedy, C.; Zaremba, W.; Sutskever, I.; et al. (2014). Intriguing properties of neural networks. In: International Conference on Learning Representations.
- Turing, A. (1950). Computing machinery and intelligence. *Mind* LIX: 433–460.

Дополнительный контент вы можете скачать по ссылке <http://bit.ly/2t53QRd>.

# Глава 10

---

## Зрительное сознание

Как обсуждалось в последних двух главах, в компьютерном зрении был достигнут значительный прогресс. Машины становятся довольно эффективными для решения широкого круга визуальных задач. Подростков не удивляет телефон, который распознает их лица. Беспилотные автомобили – предмет повседневных дискуссий в современном мире. Наличие в доме камер, которые могут определять настроение человека, скорее всего, не так уж и далеко. Теперь представьте себе мир, в котором есть машины, которые могут визуально интерпретировать мир так же, как мы. Чтобы быть более точным, представьте себе мир, в котором есть машины, которые могут гибко отвечать на, казалось бы, бесконечное количество вопросов по данному изображению. Предположим, что мы не можем отличить ответы, данные машиной, от ответов, которые дал бы человек; то есть предположим, что машины могут пройти визуальный тест Тьюринга, как он определен в разделе 9.1. Можно ли утверждать, что такая машина может *видеть*? Обладает ли такая машина *зрительным сознанием*?

Большинство людей все равно на этот вопрос ответят «нет». Они будут утверждать, что такая машина – не что иное, как очень сложный алгоритм, способный извлекать соответствующий ответ из набора пикселей. Они утверждают, что машины могут победить чемпиона мира по шахматам или го, но при этом не «понимают» игру. Они могут указать на то, что все люди разные. Люди могут воспринимать изображение, испытывать чувства по поводу изображения, смеяться над изображением или бояться его содержания; изображение вызывает ощущение особого качества (*qualia*).

*Qualia* – интригующий термин, введенный философами; словарь определяет *qualia* как «внутренний, субъективный компонент чувственного восприятия, возникающий в результате стимуляции органов чувств различными явлениями». Это определение не кажется особенно полезным для определения того, обладает ли наша необычная машина, которая может пройти визуальный тест Тьюринга, сознанием или нет. Тем не менее этого расплывчатого определения должно хватить, пока мы не получим лучший, который напрямую основан на строгом понимании того, как *qualia* может

отображаться в функции нейронных цепей. Тест Тьюринга определяется исключительно как оценка результата серии вопросов и ответов, касающихся интерпретации изображений. Такое наблюдаемое поведение (ответы на вопросы) не обязательно отражает то, что люди или машины испытывают при воздействии данного изображения. Было бы полезно иметь операционное определение<sup>1</sup> зрительного сознания с помощью теста Тьюринга, аналогичного тому, который был представлен в предыдущей главе. Наличие такого теста Тьюринга даст нам определить, может ли машина обладать сознанием, а также может помочь определить, какие виды животных обладают сознанием.

Чтобы продвинуться к определению сознания и qualia, пора вернуться к мозгу. Мы сопровождали информацию и были свидетелями ее приключений при ее перемещении и обработке вдоль центрального зрительного потока, начиная с фотонов, падающих на сетчатку, и кончая удивительными реакциями нейронов в нижней височной коре. На протяжении всего этого каскада процессов мы обнаружили нейроны, которые реагируют на изменение освещения в определенных местах поля зрения; мы восхищались нейронами, которые избирательно активируются разными типами форм; мы обсуждали, насколько нейроны толерантны к изменениям свойств стимула; нас заинтриговали нейроны, которые могут реагировать на воображаемые вещи, которые отражают то, чего нет во внешнем мире, например иллюзорные контуры; мы обнаружили нейроны, которые возбуждаются в отсутствие визуального стимула, как следствие таинственного процесса возникновения образов изображения. По мере продвижения по визуальной иерархии возрастает степень сходства между свойствами ответа нейрона и поведенческими возможностями распознавания. Попутно мы, возможно, забыли о глубоком свойстве нашего визуального опыта, а именно о субъективном ощущении зрения и переживании зримого, визуального мира. Как нейронная активность вызывает эти субъективные чувства? Какие биологические механизмы ответственны за qualia?

Приходить к конкретным определениям сознания может быть несколько преждевременным. Некоторые исследователи пытались провести различие между сознанием, осознанием, qualia и субъективным восприятием. Например, философ Дэвид Чалмерс предложил зарезервировать термин «осознанность» для обозначения отчетного и доступного содержания осознаваемой картины мира, в то время как другие термины связаны с непосредственным опытом независимо от отчетов. Здесь я буду использовать все эти термины как синонимы. Скорее всего, смешивание этих терминов – не лучшая идея, и будущая работа поможет нам прояснить наше понимание нюансов сознательного восприятия. На данный момент, вместо того чтобы пытаться дать точное определение, мы рассмотрим конкретные эксперименты, направленные на выяснение биологических механизмов, которые коррелируют с сознательным восприятием определенного образа. В контексте этих экс-

<sup>1</sup> Описание явления в терминах операций, которые необходимо произвести для подтверждения наличия явления, измерения его продолжительности и величины. – Прим. ред.

периментов вопросы хорошо определяются путем сопоставления осознанных образов с поведенческими отчетами экспериментов. Также существуют параллели «без отчета» с теми экспериментами, в которых мы воображаем, что эти образы идентичны, независимо от поведенческих моторных регистраций.

Вопрос о субъективном осознании в контексте зрительного восприятия является частью более грандиозной темы сознания. Зрительное сознание – лишь один из примеров ощущений, которые демонстрирует наш мозг. Оно может быть особенно доминирующим по отношению к другим ощущениям для приматов, но есть и иные особенности сознательного опыта, которые не зависят от зрения. Другие ощущения включают слуховое сознание, чувство боли, любви, желания и голода. Извечный вопрос о том, как физическая система может вызвать сознание, обсуждался философами, клиницистами и учеными на протяжении тысячелетий. За последние два десятилетия вопрос интерес к использованию современных методов нейробиологии для дальнейшего понимания цепей и механизмов, с помощью которых нейроны извлекают и различают сознательное содержимое. Здесь мы сосредоточимся на этих экспериментах и теориях в рамках темы обработки зрительного сигнала.

## 10.1. НЕПОЛНЫЙ СПИСОК ВОЗМОЖНЫХ ОТВЕТОВ

Механистическое объяснение зрительного сознания в конечном итоге должно быть выражено в терминах фундаментальных физических структур, поддерживающих qualia, то есть нейронов и их взаимодействий. Однако сложно представить, как физические системы могут обладать субъективным восприятием. Имеет смысл предположить, что отдельные атомы не обладают qualia и не вызывают их. Связь физического реализма с миром субъективных ощущений – это, пожалуй, один из самых сложных вопросов всех времен. Кажется, что в наших учебниках физики нет ни одной главы для чего-либо, хотя бы отдаленно похожего на сознание. Но в учебниках физики также нет главы о генетике, а мы можем проследить путь от атомов к молекулам, к богатой химии молекул углерода, к структуре ДНК и к генетике. В случае с сознанием нам не хватает даже бледного наброска такого пути.

На протяжении многих лет предлагалось множество ответов на вопрос, как в физической системе может возникнуть феномен сознания. Мы не сможем отдать должное всем этим гипотезам или подробно обсудить их все. Вместо этого мы грубо классифицируем эти идеи и перечислим некоторые из основных ответов, которые ученые предлагали на протяжении веков.

1. «Религиозный», «дуалистический» и «нефизический» ответы. Это не-научные объяснения, которые заставляют вспомнить о душе, разумной сущности, «переводчике» или какой-либо форме связи между физическими системами и нефизическими сущностями (между физикой и метафизикой. – Прим. ред.). Часто проводится различие между мозгом, физическим субстратом и «разумом», эфирной сущностью,

которая, возможно, связана или не связана с мозгом, в зависимости от того, к кому вы с этим вопросом обратитесь. Существует множество вариантов этих объяснений, включая упоминания в Библии и сочинениях Платона, Аристотеля, Фомы Аквинского, Рене Декарта, Карла Поппера, Зигмунда Фрейда и даже таких выдающихся нейробиологов, как Джон Экклс. Для простоты я беру на себя смелость объединять все формы дуализма в одну группу, которую я называю «религиозно-дуалистическими» ответами. Однако следует отметить, что между этими мыслителями есть существенные различия; конечно же, не все из них разделяют философию дуализма по религиозным причинам. Ситуация усложняется тем, что некоторые религиозные мыслители *не признают* дуализма. Я просто указываю, что любое объяснение, не основанное на физике – и, соответственно, на науке о мозге, – требует некоторого дополнительного «волшебного эликсира». Этот волшебный эликсир называют душой, умом или разумом.

Дуализм между мозгом и «разумом» пронизывает наш словарный запас. Мы говорим о том, чтобы «обрести разум», «сохранять разум свободным» или «границы разума». Более того, даже первоклассные нейробиологи, которые не признают дуализм, тоже используют странные дуалистические описания, например «Мозг знает о наших решениях раньше нас» или «Наш мозг не сообщает нам всего, что он знает. А иногда он идет дальше и активно вводит нас в заблуждение». Трудно искоренить длинную и темную тень картезианской дихотомии между разумом и мозгом.

2. «Мистический» ответ. Сторонники этой идеи, включая гигантов калибра Томаса Нагеля, Фрэнка Джексона и Дэвида Чалмерса, утверждают, что наука просто не может исчерпывающе объяснить феномен сознания. Есть несколько вариантов этой идеи, в том числе такие утверждения, как «система не может понять саму себя», или «ответ слишком сложен для нашего простого мозга», или «наука полагается на объективные измерения, а сознание требует субъективного понимания». Этот пораженный подход не кажется особенно продуктивным. В отсутствие каких-либо убедительных доказательств того, что наука не может решить проблему, кажется, лучше попытаться и потерпеть неудачу, чем не пытаться вообще. Еще более неприятным является тот факт, что этот ответ нелегко опровергнуть, не решив сначала проблему сознания, тем самым сделав его замкнутым предположением.
3. Сознание как иллюзия. Некоторые философы, такие как Дэниел Деннетт, утверждали, что сознания не существует. Следовательно, нет ничего, что требует объяснения в терминах мозговых цепей. Согласно этой точке зрения, сознание не является реальным феноменом; сознание, его ощущение – это просто иллюзия. Но какая это необычная иллюзия! Мы добились невероятного прогресса в понимании нейронной основы множественных зрительных иллюзий. Например, когда мы воспринимаем иллюзорные контуры, мы знаем, что магии нет; есть настоящие нейроны, которые энергично реагируют на эти контуры и явно представляют линии, которые мы видим (раздел 5.15).

У нас даже есть вычислительные модели, которые предполагают, как нейронные реакции на иллюзорные контуры могут возникать посредством интеграции через горизонтальные соединения сигналов от других нейронов, отвечающих на реальные контуры. Было бы особенно интересно иметь возможность предоставить аналогичное механистическое объяснение нейронной основы сознательных ощущений независимо от того, называются эти ощущения иллюзиями восприятия или нет.

4. Сознание как эпифеномен. Родственная версия сознания как иллюзии – это представление о том, что сознание – это эпифеномен. Это предложение утверждает, что сознание не имеет причинной силы – то есть что сознание не может вызывать никаких изменений в физическом состоянии системы. Как только соединяются несколько нейронов и сложных сетей, возникает чувство сознания. Согласно этой точке зрения, это чувство бесполезно. Для иллюстрации этого предложения часто используется следующая аналогия: компьютер может нагреваться, когда он выполняет свою работу, но это тепло само по себе не служит какой-либо цели; это просто побочное следствие оборудования, используемого для выполнения фактических вычислений. Однако в этом случае мы также хорошо понимаем, откуда берется это тепло с точки зрения физических законов. Было бы не менее увлекательно дать механистическое объяснение нейронной основы «сознательного тепла» независимо от того, служит ли оно какой-либо цели или нет.
5. Сознание и новые законы физики. Другие, такие как блестящий математик и физик Роджер Пенроуз, утверждают, что нам нужны новые, еще не открытые законы физики для объяснения сознания. Аргумент состоит в том, что действующие законы в чем-то недостаточны. Это предположение вполне может оказаться верным. Однако, по крайней мере исторически, новые законы были открыты при попытке описать экспериментальные результаты с помощью существующих законов, которая оказалась безуспешной. Еще лучше – показать, что существующие законы приводят к ошибочным предсказаниям, несовместимым с эмпирическими выводами. Утверждение априори о необходимости новых законов, кажется, пропускает важный шаг в научном исследовании. Возникают интересные философские и практические вопросы о том, когда накапливается достаточно доказательств, чтобы мы имели право предположить, что нынешняя парадигма неверна. Около двух десятилетий ученые думают о том, как объяснить сознание на основе активности нейронных цепей; кажется, сейчас еще не настало время, чтобы заявить, что современные законы физики не могут объяснить явление сознания. Пенроуз и другие могут быть правы, но мы с уважением просим их дать нам больше времени, чтобы попытаться решить проблему, используя только мощную артиллерию современной физики.

В отличие от упомянутых подходов, несколько нейробиологов заинтересовались, возможно, более прямым представлением о том, что сознание

возникает в результате определенных взаимодействий внутри нейронных цепей, которые определены известными нейробиологическими принципами. Сознание – это реальный феномен, присущий организму; как и любой другой феномен, сознание заслуживает механистического объяснения. Нет необходимости прибегать к волшебному эликсиру или требовать новых законов физики. Сознание вполне можно рассматривать как иллюзию в том смысле, что все наши воспринимаемые образы являются конструкциями, созданными мозгом. Более того, кажется преждевременным сомневаться в том, обладает сознание причинной силой или нет, учитывая, что мы все еще делаем первые, предварительные шаги к определению сознания в терминах науки о мозге. Согласно этой схеме, у нас уже есть ключевые ингредиенты для объяснения сознания. Какие цепи, когда и как нейронная активность управляет сознанием, еще предстоит определить с помощью научных исследований без привлечения новых законов или метафизики. Мы предполагаем, что сознание можно и нужно объяснять нейробиологическими терминами и что нет предела нашей способности найти ответ. Мы до сих пор не понимаем многих аспектов работы мозга. На самом деле я бы сказал, что мы до сих пор не понимаем большинство аспектов работы мозга. Если бы мне пришлось угадывать и сравнивать историю нейробиологии с историей исследований в области физики, я бы сказал, что нейробиология все еще находится в доньютоновском состоянии. Однако этот восхитительный уровень неведения не означает, что мы должны сдаться и обратиться к ранее приведенным объяснениям для всех наблюдений, связанных с функцией мозга, которые мы до сих пор не можем понять.

Нейробиологический подход к изучению сознания включает несколько рабочих предположений:

1. **Мы обладаем сознанием.** Сознание – это не эпифеномен. Визуальные данные вызывают ощущения, которое надежны, воспроизведимы и в основном универсальны для всех людей. Следовательно, сознание заслуживает объяснения, как и любое другое эмпирическое наблюдение, например приливы, положение луны, паттерны возбуждения ганглиозных клеток сетчатки или восприятие иллюзорных контуров.
2. Другие животные тоже имеют сознание. Это предположение позволяет нам исследовать сознание животных, не только человека. Кажется, слишком рано подводить черту и однозначно диктовать, какие животные имеют сознание, а какие нет. Представляется разумным предположить, что бактерии не обладают какой-либо формой зрительного сознания, даже если они могут улавливать свет для фотосинтеза. Трудно сказать о других видах, кроме бактерий. Растения также могут улавливать свет и осуществляют фотосинтез, наряду со многими другими захватывающими процессами; однако рабочее предположение объяснения феномена сознания, основанного на нейронных цепях, также исключило бы их из обсуждения на наличие сознания. Как только мы поймем нейронные механизмы, составляющие сознание, мы сможем понять, что некоторые виды – скажем, фруктовых мушек дрозофил – могут показать все признаки, свидетельствующие о наличии зритель-

ного сознания. С другой стороны, мы можем прийти к пониманию, что визуально запускаемое поведение дрозофилы – это чисто автоматические рефлексы, которые не обладают сознанием вовсё. Сейчас еще слишком рано говорить об этом, и мы должны держать наш мозг открытым к восприятию (а не разум, потому что это было бы дуалистично!), и мы должны быть готовы удивляться новым научным данным.

3. В этой главе мы сосредоточимся на зрительном сознании. В изучении зрительного сознания есть несколько преимуществ: мы знаем о нейроанатомии и нейрофизиологии зрительной системы больше, чем о других областях (гл. 2, 5 и 6), у нас есть модели, основанные на обработке изображений (гл. 7–9), и мы можем строго контролировать длительность и содержание стимула при регистрации поведения (гл. 3). Другие исследователи начали изучать сознание и в других областях, помимо обработки изображений. Мы ожидаем, что сможем обобщить то, что узнаем из зрения, на другие ощущения (например, боль, запах, самосознание). Изучение визуальных вычислений в мозге стимулировало прогресс во многих областях нейробиологии, включая другие сенсорные модальности, а также исследования в области обучения, памяти, принятия решений и иных процессов. Поэтому мы надеемся, что, как только мы добьемся прогресса в выяснении нейронных механизмов, которые объясняют зрительное сознание, результаты могут быть перенесены и на другие его проявления<sup>1</sup>.

Сосредоточение внимания на зрительном сознании упускает из виду многие увлекательные особенности сознания. Некоторые из этих тем включают сны, осознанные сновидения, внетьесные переживания, галлюцинации, медитацию, лунатизм, гипноз, понятие qualia и чувства. Мы не имеем в виду, что это неинтересные или неактуальные темы. Многие смелые ученые исследуют эти проявления сознания.

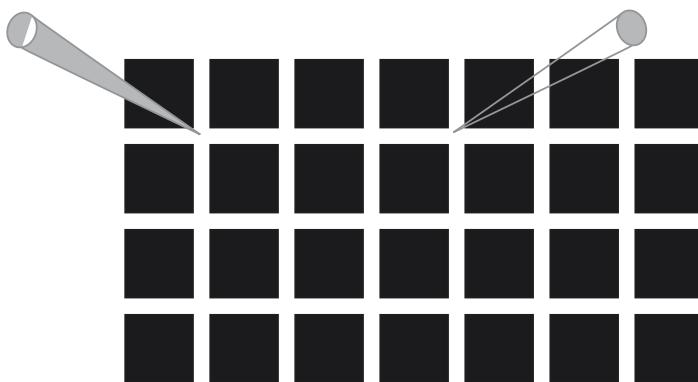
4. Нам нужно явное и механистическое представление. Только ограниченный набор частей мозга должен быть ответственным за содержимое сознания. Недостаточно утверждать, что сознание находится в мозге. Мы хотели бы иметь количественные модели зрительного сознания, похожие по духу и, возможно, даже по формату и архитектуре на типы моделей, обсуждаемые в главах 7–9. Мы надеемся, что эти модели позволят нам предсказать, как сознательные ощущения влияют на нейронную активность, и считывать сознательное восприятие по нейронной активности.

---

<sup>1</sup> Возможно, то, что автор называет «сознанием», в отечественной науке называется «психика» – «субъективный образ объективного мира». А. Н. Леонтьев определил следующий критерий наличия психики у живого существа: психика есть у тех существ, кто обладает способностью отражать abiотические (не связанные непосредственно с жизнедеятельностью) раздражители. Например, мы ощущаем приятный запах кофе, хотя этот запах никак не влияет на нашу жизнь, или антилопа в саванне пугается и убегает, услышав рычание льва, хотя непосредственно звук не представляет для нее никакой опасности. Леонтьев назвал эту способность чувствительностью. Для того чтобы связать запах с пищей или звук с опасностью, и нужна психика. – Прим. ред.

## 10.2. Поиск НСС: нейронные корреляты сознания

НСС (нейронные корреляты сознания, рис. 10.1) определяются как минимальный набор нейронных событий и механизмов, которые в совокупности достаточны для *специфических сознательных ощущений*. НСС – это минимальный набор. Такое решение, как «весь здоровый человеческий мозг обладает сознанием», не очень продуктивно. Нейронные механизмы должны быть *достаточными*, а не просто необходимыми, чтобы представлять сознательные ощущения.



**Рис. 10.1** ♦ Нейронные корреляты сознания (НСС). Любое ощущение должно быть связано с минимальным и явным представлением. Например, если бы мы записали активность нейронов, у которых есть рецептивное поле, расположенное на пересечении квадратов в этой знаменитой иллюзии, мы бы ожидали, что НСС будет активным тогда и только тогда, когда субъект воспринимает черное пятно на этом пересечении, в любой момент времени

Мы не учтываем так называемые стимулирующие факторы, такие как сердце или холинергические системы, возникающие в стволе мозга. Мы ищем корреляты для конкретного содержания сознательных ощущений, например возникающих при виде лица, в отличие от общих аспектов, таких как сознательное/бессознательное.

Совершенно очевидно, что не вся активность мозга напрямую связана с сознательным восприятием окружающего мира в любой данный момент времени. Это не означает, что эти мозговые процессы не являются необходимыми или интересными. Например, значительные ресурсы мозга и конкретные группы нейронов служат контролю дыхания, осанки и ходьбы. Но, за некоторыми исключениями, в большинстве случаев мы просто не осознаем этих процессов.

Особенно впечатляющий материал по сложной обработке мозга, которая не достигает осознания, дает пациент, изученный Мелвином Гудейлом и Дэ-

видом Милнером, что описано в разделе 4.6. У этого пациента было серьезное повреждение центрального зрительного потока, в то время как дорсальный поток был относительно сохранен. Пациент не мог распознавать образы и не осознавал их, но все же мог оперировать с этими образами с относительно высокой точностью. Например, пациент не мог сообщить об ориентации щели, но мог довольно точно поместить конверт в щель. Поиски NCC связаны с выяснением того, какие нейронные процессы коррелируют с сознательным содержанием, а какие – нет.

## 10.3. ОСОЗНАННЫЙ ОБРАЗ ДОЛЖЕН БЫТЬ «ЯВНЫМ»

Увидев объект, нейроны сетчатки активируются. Фактически стимуляция каждого из фоторецепторов сетчатки соответствующим ему образом и силой должна вызывать восприятие этого объекта. Означает ли это, что фоторецепторы сетчатки составляют желаемый NCC? Не совсем. Нейроны сетчатки активируют нейроны в LGN, которые, в свою очередь, активируют нейроны в первичной зрительной коре, которые, в свою очередь, передают информацию в более высокие области центральной зрительной коры.

Несколько линий доказательств предполагают, что активность в ранних зрительных областях от сетчатки до первичной зрительной коры вряд ли является локусом NCC. Яркий пример – то, что происходит, когда мы смотрим телевизор. У ТВ-монитора есть определенная частота обновления; то есть он показывает много кадров в секунду – скажем, 50. Ганглиозные клетки сетчатки и нейроны в первичной зрительной коре головного мозга энергично возбуждаются из-за этих быстрых изменений уровня сигнала, следующих с частотой обновления экрана, временно увеличивая спайковую частоту в ответ на каждую новую вспышку кадра. Однако наше восприятие практически игнорирует то, что информация подается покадрово, прерывисто; мы воспринимаем непрерывное движение без каких-либо мерцаний, если только частота обновления экрана не очень мала. Другими словами, есть ответы RGC, которые не достигают сознательного восприятия. И наоборот, содержание восприятия может включать в себя сигналы, которые напрямую не отражаются RGC. Яркий пример – слепая зона (раздел 2.5). Если прикрыть один глаз, образуется область поля зрения, для которой в глазу просто нет фоторецепторов. Однако мы не видим пустой или черной скотомы в этом регионе. Мозг заполняет сцену, несмотря на отсутствие информации, поступающей от сетчатки глаза в слепой зоне. Мы также редко замечаем мигания наших век, несмотря на то что для RGC весь мир на мгновение становится темным.

Важным аспектом NCC является то, что представление визуальной информации должно быть «явным». Если есть нейроны, представляющие информацию, о которой мы не знаем в данный момент времени, то эти нейроны не могут быть частью NCC в этот момент времени. Как отмечалось ранее, некоторые нейроны контролируют наше дыхание и то, как мы ходим, но обычно мы не осознаем их активность. Точно так же воспринимаемые нами образы напрямую не коррелируют с активностью нейронов сетчатки.

Что такое явное представление, и как мы когда-либо узнаем, что найдем его? В конце концов, очевидно, что для зрения необходима информация от RGC. Что делает это представление явным или неявным? Один из способов определить явное представление – это возможность декодировать информацию через однослойную нейронную сеть (разделы 6.7 и 7.7). В простейшем случае перцептрон должен уметь декодировать информацию: если у нас есть популяция нейронов с активностями  $x_1, x_2, \dots, x_n$ , то классификатор перцептронов можно выразить как  $g(w_1x_1 + w_2x_2 + \dots + w_nx_n)$ , где  $g$  – нелинейная функция, подобная пороговой. Явное представление может по-прежнему зависеть от совместной активности в популяции нейронов, причем акцент делается на том, легко ли она декодируется, в отличие от типа неявной информации, присутствующей в сетчатке.

Если мы видим стул, то этот стул представлен активностью RGC, но мы не можем считывать наличие или отсутствие стула на сетчатке, используя однослойную сеть. Аналогичным образом компьютер может содержать представление информации о стуле на цифровой фотографии. Однако, как мы обсуждали в предыдущих главах, для декодирования такой информации требуется каскад из множества вычислений. Информация об объектах в явном виде не представлена в пикселях цифровой фотографии. Точно так же сетчатка не содержит явного представления об образах, возникающих в нашем восприятии.

Явное представление содержимого зрительного восприятия в любой момент времени не должно соответствовать частоте обновления монитора, должно быть способно заполнить недостающую информацию в слепом пятне и должно быть подвержено визуальным иллюзиям таким же образом, как это диктует восприятие. Например, рассмотрим треугольник Каниджи (раздел 3.1): воспринимаемый образ края, когда его нет, предполагает, что должны быть нейроны, которые представляют этот субъективный край. Нейроны сетчатки не реагируют на такие иллюзорные контуры, в отличие от нейронов в области коры V2 (раздел 5.15).

## 10.4. ЭКСПЕРИМЕНТАЛЬНЫЕ ПОДХОДЫ К ИЗУЧЕНИЮ ЗРИТЕЛЬНОГО ВОСПРИЯТИЯ

Пример треугольника Каниджи и другие зрительные иллюзии предлагают многообещающий путь исследования нейронных коррелятов зрительного сознания путем определения того, какие нейронные процессы совпадают с субъективным восприятием образов. Особенно плодотворный экспериментальный подход заключается в том, чтобы сосредоточить внимание на ситуациях, когда один и тот же визуальный стимул может привести к зрительному осознанию только иногда, но не в каждом случае (рис. 10.2).

Одним из примеров является рассмотрение восприятия вблизи порогов различия. Например, стимул может быть трудно обнаружить, если уменьшить его контраст. Если контраст достаточно высок, подопытные могут обнаруживать зрительный стимул в большей части случаев его предъяв-

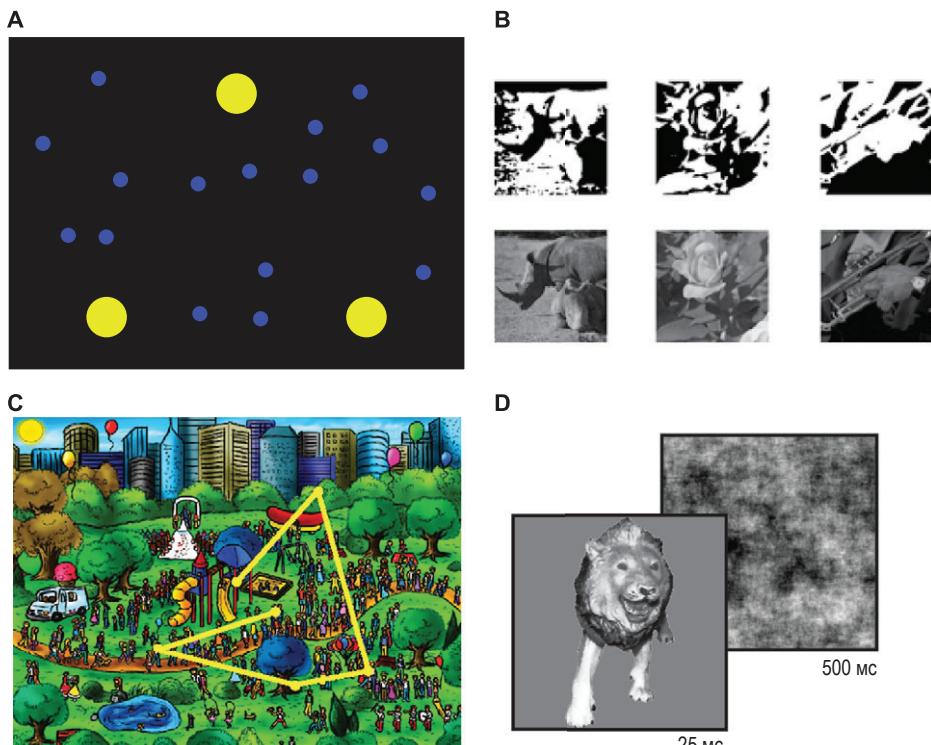
ления. Если контраст слишком низкий, подопытные по большей части не могут обнаружить стимул. Вблизи порога существует промежуточная область, в которой наблюдатели иногда могут видеть стимул, а в иногда – нет, что оценивается с помощью измерения поведенческих реакций. Один и тот же физический стимул иногда приводит к его восприятию, но иногда нет. Предположим, что мы можем гарантировать, что предъявляем точно такой же стимул, а также что глаза фиксируются в одном и том же месте и что нет никаких других изменений. В этих условиях кажется разумным предположить, что нейронные реакции сетчатки во всех испытаниях одинаковые, независимо от того, воспринимается стимул или нет. Однако что-то должно измениться где-то в мозге в некоторых испытаниях, чтобы подопытные сообщали, что они видят стимул. В этом случае мы можем исследовать, где, когда и как возникают нейронные реакции вдоль зрительной коры, которые коррелируют с субъективным восприятием.

Похожая ситуация может быть достигнута в экспериментах по обратной маскировке, когда стимул вспыхивает в течение короткого промежутка времени, после чего быстро сменяется шумом (рис. 10.2D, раздел 3.6). Если продолжительность стимула достаточно велика, подопытные могут легко его увидеть. Если продолжительность достаточно мала, подопытные никогда не увидят его. Существует промежуточный режим длительностью порядка 25 мс, при котором подопытные сообщают о том, что видели стимул только в некоторых, но не во всех испытаниях.

Другой пример – интерпретация изображений, которые трудно распознать, например изображений Муни. Эти изображения представляют собой ухудшенные черно-белые изображения, которые сложно интерпретировать с первого взгляда. Известный пример – иллюзия далматинской собаки. Еще несколько примеров показаны на рис. 10.2В. Рассмотрим пример в верхнем левом углу на рис. 10.2В. На первый взгляд кажется, что изображение содержит множество черных пятен произвольной формы, случайно разбросанных по всей поверхности. Тем не менее изображение содержит носорога в естественной обстановке. Если кто-то внимательно изучит контуры или посмотрит на полутоновый аналог этого изображения (внизу слева на рис. 10.2В), то он сможет легко узнать носорога, а также интерпретировать остальную часть сцены. То же изображение и предположение о том же месте фиксации может привести к интерпретации его или как шума, или как носорога. Мы предполагаем, что нейронное представление изображения на уровне сетчатки неразличимо между интерпретациями шума и носорога. Однако должно быть представление о носороге, которое появляется, возможно, в нейронах нижней височной коры (раздел 6.2), и это представление активируется тогда и только тогда, когда наблюдатель правильно интерпретирует изображение.

Мы можем наблюдать это ежедневно во время визуального поиска. Представьте, что мы ищем ключи от машины на загроможденном столе или ищем Уолдо на рис. 10.2С. Глаза сканируют стол в течение нескольких секунд с помощью множества саккад. Иногда мы фиксируемся непосредственно на ключах от машины, но не осознаем, что наш взгляд упал на ключи, и продолжаем поиск. В конце концов, наш взгляд останавливается на ключах, и мы начинаем осознавать, что нашли их. Это случай двух фиксаций, предположим для

простоты, в одном и том же месте с одним и тем же визуальным стимулом, одна без осознания, а другая с осознанием.



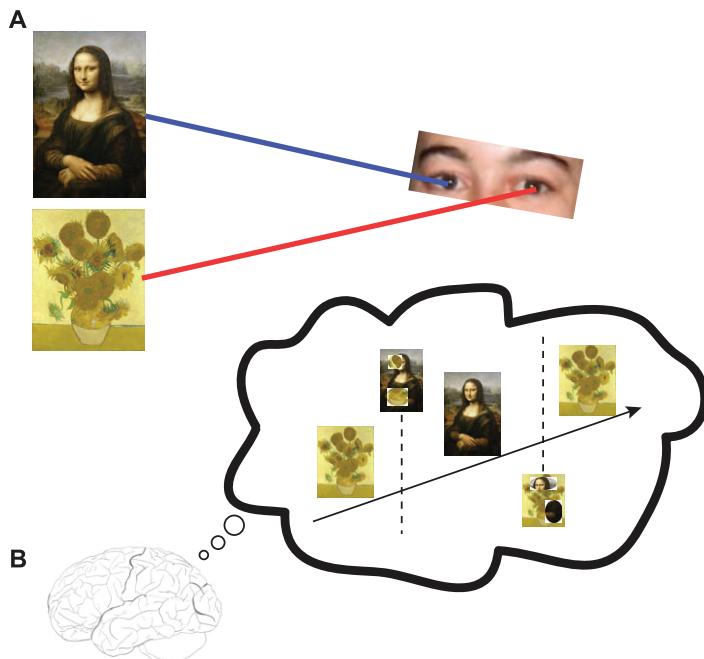
**Рис. 10.2** ♦ Примеры задач, используемых для изучения NCC. (А) Слепота, вызванная движением. Когда синие точки перемещаются, желтые круги периодически исчезают из поля зрения восприятия. (В) Изображения Муни. Обычно трудно интерпретировать изображения в верхнем ряду. Обращение к их аналогам в градациях серого (нижний ряд) немедленно делает эти изображения интерпретируемыми. (С) Во время визуального поиска наблюдатели часто фиксируются на целевом объекте и продолжают поиск, не осознавая этого. (Д) Обратная маскировка может сделать стимул невидимым

Аналогичная ситуация возникает при явлениях *слепоты невнимания* (перцептивная слепота) и *слепоты к изменению*. Во время слепоты невнимания наблюдатели не замечают полностью видимый объект, предположительно потому, что внимание сосредоточено где-то в другом месте. Хорошей демонстрацией этого явления является хорошо известное видео, где две команды, черная и белая, делают передачи друг другу двух баскетбольных мячей. Испытуемых просят подсчитать количество передач между членами одной команды. Без ведома испытуемых (и заранее прошу прощения, если я испортил эффект для читателя) человек, замаскированный под гориллу, медленно проходит через середину сцены. Примечательно, что около половины испытуемых совершенно не замечают гориллу. Без сомнения, инфор-

мация о горилле достигает ганглиозных клеток сетчатки и, возможно, также первичной зрительной коры, а может быть, даже более высоких областей в пределах зрительной коры. Однако многие испытуемые совершенно не обращают внимания на присутствие гориллы. В связанном случае слепоты к изменениям наблюдатели не замечают, что что-то было изменено на дисплее. Один экземпляр включает в себя многократное мигание изображения с коротким пустым интервалом между ними. В альтернативных вспышках наблюдается существенное изменение изображения; например, у одного человека может измениться цвет брюк. Несмотря на то что наблюдатели могут свободно перемещать глаза, чтобы внимательно изучить изображение, часто бывает довольно сложно заметить изменение, на обнаружение которого могут потребоваться десятки секунд.

Особым типом визуальной иллюзии, оказавшим влияние на изучение зрительного сознания, являются бистабильные восприятия. Известный пример бистабильного восприятия – куб Неккера. Одну и ту же картинку можно увидеть в двух разных конфигурациях. В случае куба Неккера можно добровольно переключаться между двумя возможными интерпретациями одного и того же стимула.

Такой волевой контроль невозможен в случае явления, известного как *бинокулярное соперничество* (рис. 10.3). В нормальных условиях информация,



**Рис. 10.3 ♦ Бинокулярное соперничество.** (A) Стимул (Джоконда) показан одному глазу, а другой стимул (подсолнухи) показан другому глазу. (B) Воспринимаемый образ обычно чередуется между двумя возможными образами с переходными периодами частичного соперничества, когда два образа сливаются

передаваемая правым и левым глазами, сильно коррелирует. То, что видят правый и левый глаза, не идентично: небольшие различия между данными правого и левого глаз дают сильные сигналы для получения трехмерной информации. Что произойдет, если мы подадим два совершенно разных стимула на правый и левый глаза? В этих условиях наблюдатели воспринимают либо один стимул, либо другой, чередуя их, казалось бы, случайным образом, – соперничество между входами от двух глаз.

Обширные психофизические исследования предоставили обширную информацию об условиях, которые приводят к перцептивному преобладанию того или иного визуального стимула, о том, что можно или нельзя делать с подавляемой информацией, и о динамике, лежащей в основе перцептивных изменений. Что особенно интересно в этом явлении, так это то, что в достаточно хорошем первом приближении общий зрительный стимул (оба изображения) постоянен, но субъективное восприятие чередуется между двумя возможными интерпретациями визуального мира.

Простую демонстрацию бинокулярного соперничества можно вызвать, свернув лист бумаги и посмотрев в него одним глазом. Когда оба глаза открыты и одна рука держит лист бумаги, один глаз может фокусироваться на удаленных объектах, а другой – на руке перед вами. Воспринимаемый образ таинственным образом чередуется между рукой и объектами, видимыми вдалеке через видимое отверстие в вашей руке.

Продолжительность доминирования каждого из двух стимулов следует гамма-распределению, и восприятие непроизвольно переключается с одного стимула на другой, иногда проходя через смешанное восприятие, известное как частичное соперничество. Это как если бы мозг был запрограммирован на понимание того, что не может быть двух разных объектов в одном месте одновременно. Эти два объекта соревнуются за образ, возникающий в сознании; один из них на мгновение побеждает, но ожесточенная конкуренция продолжается, и в конце концов другой объект берет верх. Хотя название и формат представления, казалось бы, предполагают конкуренцию между монокулярными каналами, некоторые свидетельства предполагают, что конкуренция также имеет место на более высоком уровне, между восприятием двух образов: (i) можно вызвать монокулярное *соперничество*, более слабое явление, при котором конкуренция между двумя возможными интерпретациями входных данных имеет место, даже если входные данные представлены только одному глазу посредством наложения; (ii) стимулы могут быть расположены так, что половина информации об объекте представляется одному глазу, а половина – другому глазу; вместо того чтобы испытывать чередование двух половинных образов, возникает соперничество между двумя целостными образами, что требует объединения информации от двух глаз; (iii) замечательные эксперименты, в которых стимулы быстро перемещаются от одного глаза к другому, дополнительно показывают, что конкуренция может происходить на уровне самого восприятия объекта, а не между двумя глазами.

Существует несколько разновидностей бинокулярного соперничества. *Подавление вспышки* относится к ситуации, когда стимул – скажем, «Джоконда» – отображается монокулярно – скажем, в правый глаз. Сразу после этого

«Джоконда» остается в правом глазу, но новый раздражитель – скажем, «Подсолнухи» – попадает в левый глаз. В этих условиях новый стимул, «Подсолнухи», доминирует, а старый стимул, «Джоконда», полностью подавляется. Если два стимула остаются на экране, по одному для каждого глаза, в конечном итоге возникает бинокулярное соперничество, и восприятие начинает чередоваться между двумя образами. Интересным вариантом является феномен *непрерывного подавления вспышкой*, когда «Джоконда» остается на правом глазу, в то время как серия стимулов непрерывно направляется на левый глаз. В этих условиях наблюдатели воспринимают непрерывный поток вспыхивающих раздражителей, и «Джоконда» может оставаться невидимой в течение нескольких минут.

Как и в других примерах, мы ожидаем, что активность RGC не будет коррелировать с изменением воспринимаемого образа при переключении между одной интерпретацией изображения и другой во время бинокулярного соперничества. С другой стороны, NCC должен напрямую коррелировать с изменениями воспринимаемого образа.

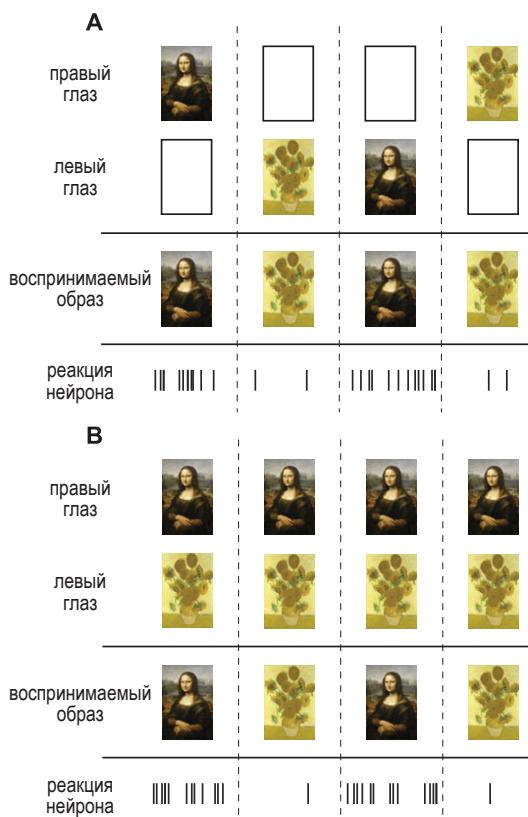
## 10.5. НЕЙРОННЫЕ КОРРЕЛЯТЫ ЗРИТЕЛЬНОГО ВОСПРИЯТИЯ ВО ВРЕМЯ БИНОКУЛЯРНОГО СОПЕРНИЧЕСТВА

Феномен бинокулярного соперничества подвергается широкому изучению на нейрофизиологическом уровне. Исследователи ищут нейронные изменения, которые коррелируют с субъективными переключениями между сигналами, приходящими в тот или иной глаз. Интересным свойством бинокулярного соперничества является то, что это явление может быть вызвано стимулами практически любой формы. Бинокулярное соперничество может происходить путем показа горизонтальной решетки для правого глаза и вертикальной решетки для левого глаза или изображения лица для правого глаза и изображения решетки для левого глаза. Вооружившись способностью исследовать реакции нейронов вдоль вентрального пути (гл. 5 и 6), мы можем спросить, коррелирует ли возбуждение нейронов, активируемых этими стимулами, с субъективным восприятием образов в отчете эксперимента.

Никос Логотетис и его сотрудники тщательно изучили этот вопрос на всей зрительной коре. Они использовали множество хитрых стратегий, чтобы научить обезьян сообщать о воспринимаемом образе во время перцептивных изменений. Например, периоды бинокулярного представления случайным образом смешивались с периодами монокулярного представления, которые можно использовать в качестве контроля, чтобы гарантировать, что обезьяна правильно сообщает о том, что видит.

Исследователи записали активность зрительно избирательных нейронов, которые сильнее реагируют на предъявляемый стимул А по сравнению с другим стимулом В (аналогично примерам, показанным в разделе 6.2). Затем

исследователи представили А одному глазу и В другому глазу (рис. 10.4). В частности, они регистрировали активность нейрона в нижней височной коре головного мозга, который сильнее реагировал на изображение орангутанга, чем на изображение абстрактного узора, во время монокулярной презентации – или во время бинокулярной презентации, когда один и тот же стимул подавался в оба глаза. Примечательно, что когда орангутанг и абстрактный узор были представлены во время эксперимента по бинокулярному соперничеству, динамические изменения частоты срабатывания нейронов сильно коррелировали с отчетами о зрительном восприятии обезьяны: когда обезьяна указывала, что видит орангутанга, нейрон показывал высокую спайковую частоту, тогда как всякий раз, когда обезьяна показывала, что видит абстрактный узор, нейрон показывал низкую частоту срабатывания. Изменения в частоте срабатывания предшествовали отчетам о восприятии образа на несколько сотен миллисекунд, что согласуется с идеей о том, что нейронные реакции отражают смену воспринимаемого образа и что требует-



**Рис. 10.4** ♦ Схематическое представление реакции нейрона на тот или иной образ во время бинокулярного соперничества. **(A)** Во время монокулярного представления нейрон сильнее реагирует на «Джоконду», чем на «Подсолнухи». **(B)** Во время бинокулярного соперничества нейрон проявляет более сильную реакцию всякий раз, когда субъект сообщает, что видит «Джоконду»

ся время, чтобы вызвать соответствующий моторный выход, который можно зарегистрировать. Подавляющее большинство нейронов нижней височной коры продемонстрировали такое поведение – их активность коррелировала с субъективными отчетами о воспринимаемом образе.

Активность нейронов в средней височной доле человека обнаруживает такие же корреляции с восприятием того или иного образа. Во всех этих экспериментах как на обезьянах, так и на людях нейронные реакции могут предшествовать поведенческому отчету о перцептивных переходах на несколько сотен миллисекунд. По крайней мере частично, это может указывать на то, что у нас нет очень точных способов измерения точного времени перцептивного перехода, а поведенческие регистрации могут иметь временную задержку. Тем не менее, что любопытно, нейроны медиальной височной доли человека могут активироваться задолго до перцептивных переходов, даже за 1000 мс, а во фронтальных областях некоторые нейроны активировались даже раньше. Кажется маловероятным, что такое может быть приписано исключительно задержке поведенческих регистраций. Следовательно, эти нейроны могут быть вовлечены в еще плохо изученные предсознательные механизмы, которые в конечном итоге приводят к перцептивным переходам.

В отличие от корреляций, наблюдавшихся для нейронов в ИТС и медиальной височной доле, активность нейронов в V1 обычно не соответствовала данным регистрации. Первичные нейроны зрительной коры указывали на физическое присутствие предпочитаемых ими стимулов, и в большинстве случаев их активность не коррелировала с поведенческими реакциями обезьяны на воспринимаемые образы. Промежуточные визуальные области, такие как V4 и область МТ, показали результаты, которые находились посередине между результатами в V1 и ИТС. Другими словами, по мере того как мы поднимаемся по визуальной иерархии, увеличивается доля нейронов, которые коррелируют с отчетом о субъективном восприятии.

Точная пропорция нейронов, которые коррелируют с перцептивными переходами в данной области, может зависеть от экспериментальных условий. Например, элегантное исследование показало, что в области МТ изменение стимула и контекста может привести к тому, что разные нейроны будут показывать изменения спайковой частоты, сопровождающиеся изменениями в восприятии. Другими словами, NCC может не быть статичным, а скорее динамически зависеть от задачи и условий эксперимента.

Одна из проблем, связанных с этими экспериментами, заключается в том, что нам необходимо получить поведенческую реакцию от подопытных, чтобы выяснить, что они восприняли. Указывают ли нейронные реакции на сознательно воспринятые образы, или они просто отражают решения и моторные сигналы? Для разграничения этого было разработано несколько экспериментальных методов, основанных на оригинальных способах считывания восприятия без регистрации поведенческой реакции. В этих так называемых методах «без отчета» используются размер зрачка или другие независимые сигнатуры того или иного стимула для определения перцептивных переходов без явной поведенческой реакции. Результаты метода «без отчета», по-видимому, подтверждают результаты более ранних исследова-

ний, показывая нейронные корреляты субъективных восприятий, особенно в высших эшелонах зрительной коры.

Другой вопрос, который был поднят в связи с интерпретацией исследований, направленных на отслеживание коррелятов сознательного восприятия, заключается в том, отражают ли реакции нейронов изменения в сознательном восприятии или отражают только изменение внимания. В большинстве случаев внимание и сознание тесно взаимосвязаны, и мы осознаем то, на что обращаем внимание. Однако можно провести эксперименты, в которых внимание и сознание диссоциированы. Эти эксперименты показывают, что наблюдатели могут сознательно воспринимать объект или сцену в отсутствие механизмов внимания сверху вниз нейронных структур. Кроме того, наблюдатели также могут обращать внимание на объекты, ускользающие от восприятия.

## 10.6. ТРЕБОВАНИЯ К NCC

Эксперименты с бистабильным восприятием, такие как бинокулярное соперничество, проложили путь к первоначальному пониманию того, как изменения активности в конкретных структурах нейронной активности коррелируют с переходами в субъективном восприятии. В то же время в мозге есть много других нейронов, которые продолжают выполнять свои обязанности независимо от текущего контента сознания.

Что могло бы стать доказательством обнаружения NCC? Параллельно с обсуждением вычислительных моделей в гл. 7–9 мы ищем количественное описание субъективного восприятия. В гл. 8 мы утверждали, что полное вычислительное описание зрения должно быть способно предсказывать нейронные реакции на любое произвольное изображение (раздел 8.14), а также предсказывать поведенческие реакции в любой визуальной задаче в ответ на любое изображение (разделы 8.12 и 8.13). Распространяя это определение на область зрительного сознания, для полного описания нейронного коррелята зрительного сознания должны быть выполнены четыре условия:

1. Мы должны уметь количественно прогнозировать нейронные реакции с учетом состояния восприятия<sup>1</sup>. Например, во время бинокулярного соперничества мы должны быть в состоянии предсказать нейронную активность нейронов в различных областях мозга в зависимости от состояния восприятия подопытного.
2. И наоборот, мы должны уметь предсказывать состояния восприятия по ответам нейронов. Регистрируя активность популяций нейронов (определенные типы нейронов, цепи и области для NCC), мы должны уметь определять, что подопытный сознательно воспринимает в данный момент времени.

<sup>1</sup> Здесь состояние восприятия подразумевает, какой из предложенных ему зрительных стимулов в данный момент сознательно воспринимает подопытный. – Прим. ред.

3. Мы должны быть в состоянии вызвать определенный образ, активируя соответствующие нейронные структуры (например, с помощью электрической стимуляции, раздел 4.9). Эти нейронные структуры могут находиться в одной или в нескольких областях мозга. Результирующий, воспринимаемый образ должен быть конкретным (например, образ женщины, сидящей в парке на открытом воздухе рядом с деревом), а не просто световыми фосфенами, вызванными путем одновременной активации кластеров нейронов в первичной зрительной коре. Кроме того, в эксперименте с бинокулярным соперничеством стимуляция NCC должна давать возможность изменить состояние восприятия подопытного. Такое расширенное понятие NCC постулирует, что возбуждение этих специфических нейронных цепей напрямую и причинно связано с состоянием восприятия. Следовательно, даже если субъект спит, активация NCC должна вызвать сновидение или галлюцинацию со специфическим зрительным образом.
4. Мы должны уметь отключать или подавлять определенное состояние восприятия, изменяя схемы нейронной активности. В эксперименте с бинокулярным соперничеством мы могли гарантировать, что подопытные не воспринимают один из стимулов, отключив соответствующий NCC. Опять же, поскольку NCC несет прямую и причинную ответственность за восприятие, в принципе, мы могли бы показать изображение женщины, сидящей в парке рядом с деревом, и подопытный не воспримет ничего из этого, если соответствующий NCC будет инактивирован. Эта манипуляция должна быть специфичной для конкретного содержания, о котором сигнализирует NCC (например, закрытие глаз для снижения активности всех нейронов в зрительной системе не будет являться проверкой этого требования).

Излишне говорить, что мы все еще далеки от понимания нейронных коррелятов зрительного сознания при соблюдении этих четырех условий. Тем не менее эти вопросы стали важной областью исследований, и мы можем оказаться приятно удивлены, увидев значительный прогресс в данной области в ближайшие годы.

## 10.7. ИНТЕГРИРОВАННАЯ ТЕОРИЯ ИНФОРМАЦИИ

Предыдущие разделы были сосредоточены на результатах экспериментов, с помощью которых исследователи пытались выяснить, какие конкретные паттерны распределения активности нейронов коррелируют с субъективным восприятием или не коррелируют с ним. Эти эмпирические данные привели к появлению понятия об относительном порядке, в котором различные области могут активироваться во время сознательного восприятия образов. Относительный порядок активации различных нейронных цепей во время перцептивных переходов приводит к идее глобального рабочего пространства, которое принимает сенсорную информацию и распространяет эту информацию «глобально» или, по крайней мере, на множество других об-

ластей мозга. Некоторые исследователи предположили, что распространение на другие области мозга вызывает изменения в субъективном восприятии.

Параллельно с эмпирическими наблюдениями паттернов распределения активности нейронов, сопровождающих восприятие зрительных образов, за последнее десятилетие была разработана элегантная, амбициозная и противоречивая теоретическая основа, заслуживающая обсуждения: теория интегрированной информации (ИТ) Джоулио Тонони. В упрощенной форме основная идея, лежащая в основе ИТ, заключается в том, что сознательный опыт представляет собой информацию и что это представление уникально. Эта теория содержит набор из пяти аксиом (рис. 10.5) и дает количественное определение информации и интеграции. Эти пять аксиом утверждают, что (i) сознание существует как уникальный внутренний опыт (*внутреннее существование, экзистенция*), (ii) сознательный опыт состоит из множества феноменологических элементов (*состав*)<sup>1</sup>, (iii) сознание специфично (*информация*), (iv) сознательный опыт един и нередуцируем (*интеграция*) и (v) контент сознания ограничен пространством и временем (*исключение*). Затем теория выводит из этих аксиом постулаты, устанавливающие необходимые условия для того, чтобы система проявляла эти свойства сознательного опыта.

Согласно ИТ, динамическая система взаимосвязанных частей характеризуется метрикой, обозначаемой  $\Phi$  («фи»), которая имеет ненулевое положительное значение, когда система не может быть описана меньшими, относительно независимыми подсистемами. Чем больше  $\Phi$ , тем больше интегрированной информации имеет система. Теория постулирует, что сознательный опыт пропорционален  $\Phi$ . Определение  $\Phi$  состоит из двух шагов: (i) выполнить мнимое разбиение системы и вычислить  $\Phi$ , меру того, насколько две части влияют друг на друга (т. е. насколько хорошо мы можем предсказать эволюцию системы на основе условной вероятности перехода), и (ii) определить  $\Phi$  как наиболее «грубое» такое разбиение, которое минимизирует  $\Phi$ . Элегантным образом теория дает конкретные математические определения для вычисления этих величин с учетом динамических переходов в системе взаимосвязанных частей, таких как нейронная цепь.

Основная проблема при тестировании структуры ИТ заключалась в том, что для реальных систем эти уравнения представляют из себя практически невыполнимую вычислительную задачу. Время вычисления с увеличением размера системы растет экспоненциально. Макс Тегмарк и другие недавно разработали приближение для вычисления  $\Phi$  с использованием теории графов, доведя вычисления до полиномиальной зависимости от размера системы и сделав этот алгоритм легко применимым к большим объемам записей регистрации физиологических сигналов.

Эта теория особенно элегантна, поскольку начинается с аксиом и предлагает конкретные количественные определения, что отличает ее от других дискуссий о сознании, которые носят чисто качественный характер. В то же время теория делает много противоречивых предсказаний. Любой объект – сотовый телефон или даже стул, на котором мы сидим, – имеет определенное значение  $\Phi$ . Можно ожидать, что неодушевленные предметы или бактерии должны

<sup>1</sup> То есть сознание структурировано. – Прим. ред.

иметь  $\Phi = 0$ , но это не то, что утверждает теория. Эти объекты могут иметь низкие значения  $\Phi$ , возможно, даже пренебрежимо малые, но не нулевые. Интуитивно хотелось бы, чтобы любая теория указывала на полное отсутствие сознания у стула, а не на то, что у него есть микроскопическое сознание.



**Рис. 10.5 ♦** Аксиомы интегрированной теории информации (IIT). IIT предлагает пять фундаментальных аксиом о природе сознательного опыта: (1) внутренний опыт; (2) состав; (3) информация; (4) интеграция и (5) исключение. По материалам Тонони и Кох, 2015

Возможно, это чисто семантическая задача, которую можно решить, установив пороговое значение для  $\Phi$ .

Другой сбивающий с толку вывод ИИТ состоит в том, что в принципе возможно создавать относительно простые искусственные системы с высокими значениями  $\Phi$  (для знатоков примером являются так называемые матрицы Вандермонда). Однако кажется несуральным, что такие искусственные системы проявляют сознание. Конечно, вывод интроспективного самоанализа, что эти предсказания противоречат здравому смыслу, не делает их неверными. В науке есть множество примеров, когда противоречащие интуиции предсказания приводили к новым захватывающим открытиям. Наука должна руководствоваться экспериментально проверяемыми предсказаниями и эмпирическими результатами, а не нашими пристрастиями или интуицией.

В конце концов, будет интересно проверить интегрированную теорию информации эмпирически. Независимо от того, является ли ее теоретическая основа полностью верной, или потребует пересмотра и уточнения, как почти все другие теории в науке, или даже если она полностью неверна, – это количественная теория, которая была предложена для объяснения одной из самых неуловимых тайн человеческого существования – сознания.

## 10.8. Обзор главы

- Феномен сознания обсуждается на протяжении тысячелетий мыслителями из самых разных областей знания, но только недавно оно стало важной темой исследований для скрупулезных теоретиков и экспериментаторов нейробиологии.
- Экспериментальные усилия были сосредоточены на поиске необходимых и достаточных нейронных коррелятов сознания, НСС.
- Для изучения зрительного сознания было разработано несколько экспериментальных методов, в которых входной сигнал постоянен, но восприятие его может меняться во времени. Эти эксперименты включают обратную маскировку, манипуляции с вниманием, визуальный поиск и бистабильное восприятие, такое как бинокулярное соперничество.
- Во время бинокулярного соперничества нейронные реакции в высших отделах зрительной коры коррелируют с динамическими изменениями содержимого сознания (воспринимаемыми образами).
- Полное описание НСС потребует количественной вычислительной модели, которая может предсказать нейронные реакции с учетом состояния восприятия, а также может предсказать состояние восприятия с учетом ответов нейронов. Активация или подавление НСС должны вызывать или заглушать определенные состояния восприятия.
- Интегрированная теория информации (ИИТ) – это первая количественная теоретическая основа, цель которой – объяснить возникновение сознания как функции динамической системы со взаимосвязанными частями.

## ЛИТЕРАТУРА

См. дополнительные ссылки на <http://bit.ly/2FHXycS>.

- Chalmers, D. (1996). *The conscious mind: in search of a fundamental theory*. New York: Oxford University Press.
- Crick, F. (1994). *The astonishing hypothesis*. New York: Simon & Schuster.
- Koch, C. (2005). *The quest for consciousness*, 1st ed. Los Angeles: Roberts & Company Publishers.
- Leopold, D. A., and Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends in Cognitive Sciences* 3: 254–264.
- Tononi, G. (2005). Consciousness, information integration, and the brain. *Prog Brain Res* 150: 109–126.

Дополнительный контент вы можете скачать по ссылке <http://bit.ly/2FHXycS<sup>1</sup>>.

---

<sup>1</sup> В этой книге автор использует понятие «зрительное сознание», мало распространенное в российской научной литературе. См. примечание к разделу 1.8. – Прим. ред.

# Предметный указатель

## A

Alexnet, 19

## C

CRISPR, 97

## D

DeepDream, 19, 20, 271, 277

## F

fovea, 55, 66, 156

## G

GAN, 19, 269, 270, 271

GoogleNet, 231

## I

ImageNet, 19, 219, 220, 222, 223, 224, 231, 235, 246, 247, 258, 259, 286

Inception-v3, 231, 259

ITC, 19, 107, 108, 112, 119, 120, 140, 141, 144, 152, 153, 154, 155, 156, 157, 158, 160, 161, 162, 163, 164, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 238, 239, 245, 303

## L

LGN, 19, 31, 35, 63, 64, 95, 104, 111, 117, 123, 126, 130, 134, 136, 137, 144, 145, 149, 169, 173, 179, 204, 241, 295

LSTM, 19, 244, 257

## N

NCC, 294, 295, 298, 301, 303, 304, 305, 308  
NYU, 231

## P

PredNet, 19, 242, 243, 244, 245, 258

## Q

qualia, 287, 288, 289, 293

## R

ResNet, 19, 210, 231, 259, 262  
RGC, 20, 57, 58, 59, 60, 61, 62, 63, 64, 65, 147, 148, 169, 295, 296, 301

## S

Softmax, 20

## T

t-распределенное стохастическое сетевое вложение, 226

## U

UCF101, 20, 275

## V

V1, 20, 102, 103, 104, 105, 112, 114, 117, 123, 125, 127, 128, 130, 131, 133, 134, 135, 136, 137, 138, 139, 140, 141, 144, 145, 146, 148, 149, 150, 151, 158, 169, 173, 179, 181, 185, 190, 210, 211, 231, 232, 241, 242, 243, 245, 303  
VGG, 19, 231, 235

## X

XDream, 20, 271, 272, 273

## A

Аблация, 108  
Автокодировщик, 206  
Агнозия, 13, 109  
Адаптация, 30, 65, 168, 169, 182  
Акинетопсия, 106  
Аксоны, 57, 99, 181, 182, 186  
Активность нейронов, 59, 61, 94, 104, 117, 129, 149, 207, 242, 294, 303, 304  
Амакриновые клетки, 53, 54  
Амблиопия, 89  
Анортоскопическое восприятие, 80  
Аперцептивная зрительная агнозия, 111  
Ассоциированная зрительная агнозия, 111  
АтTRACTор, 193  
Ахроматопсия, 55, 108, 149

**Б**

Бинокулярное поле, 89  
 Бинокулярное соперничество, 299, 301, 304, 308  
 Биофизически подобные модели, 202  
 Биполярные клетки, 53, 54  
 Бистабильное восприятие, 38  
 Блок, 19, 180, 181, 182, 183, 185, 186, 187, 188, 189, 194, 209, 213, 215, 216, 225, 226, 227, 228, 229, 271

**В**

Центральная зрительная кора, 141, 153, 158  
 Центральная кора, 105  
 Вентральный поток, 107, 153  
 Верхний бугорок, 30, 63  
 Верхний холмик, 30, 63, 95  
 Визуальная система, 26, 27, 69, 73, 141, 178, 201, 241  
 Визуальное различение, 80, 91  
 Визуальное распознавание, 29, 76, 92  
 Визуальное сознание, 287  
 Визуальное сопоставление форм, 110  
 Визуальный поток, 152  
 Визуальный тест Тьюринга, 41, 249, 250, 251, 252, 276, 278, 285, 287  
 Вместимость системы, 29  
 Восприятие, 33, 36, 37, 38, 42, 52, 58, 63, 66, 68, 69, 72, 75, 80, 82, 83, 84, 85, 86, 88, 91, 115, 117, 118, 148, 149, 163, 168, 179, 287, 292, 293, 294, 295, 296, 300, 301, 302, 305, 308  
 Восходящие связи, 126  
 Выходные нейроны сетчатки, 20  
 Вычислительные модели зрения, 147, 153, 208

**Г**

Ганглиозные клетки, 20, 53, 54, 57, 58, 66, 88, 95, 204  
 Ганглиозные клетки сетчатки, 20, 57, 58, 66, 88, 95, 204  
 Генерализация, 273, 280  
 Генеративные состязательные сети, 19, 159, 270  
 Генератор изображений, 159, 270  
 Гиппокамп, 39, 113, 124, 152  
 Гладкие сверхмалые биполярные клетки, 54  
 Глубокая сверточная нейронная сеть, 19, 211, 235  
 Глубокая структура, 204

Глубокие нейронные сети, 204, 211

Глубокие сверточные нейронные сети, 239  
 Горизонтальные клетки, 53, 54  
 Горизонтальные связи, 128, 210  
 Градиентный спуск, 213, 217, 277

**Д**

Дендритные ветви, 63, 126  
 Дендритные шипы, 180  
 Дендриты, 126, 181, 182, 186  
 Дискриминатор изображений, 270  
 Диффузные биполярные клетки, 54  
 ДНК, 21, 97, 99, 289  
 Дорсальная кора, 105  
 Дорсальный поток, 153, 295

**Ж**

Желтое пятно, 56

**З**

Зрительная адаптация, 85  
 Зрительная кора головного мозга, 102  
 Зрительное воображение, 170  
 Зрительное восприятие, 39, 68, 289  
 Зрительное распознавание, 27, 195

**И**

Идентификация, 82, 86, 111, 252, 265  
 Идентификация форм, 27  
 Иерархическая сетевая модель, 209  
 Избирательность, 27, 28, 29, 36, 42, 120, 153, 157, 160, 161, 162, 163, 185  
 Иллюзии, 56, 57, 69, 72, 75, 77, 83, 85, 86, 91, 115, 147, 149, 291, 294, 296, 299  
 Иллюзия  
     маски, 86  
     Мюллера-Лайера, 83  
     составного лица, 72  
 Инвагинирующие сверхмалые биполярные клетки, 54  
 Ингибиование  
     возвращения, 51  
     окружения, 60, 133, 242, 245  
 Ингибирующий канал, 99  
 Интегрированная теория информации, 306  
 Интеллект, 249  
 Интуиция, 37, 178, 185  
 Ионный канал, 98  
 Ипсолатеральный сигнал, 64  
 Искусственный интеллект, 19

**К**

- Калькариновая борозда, 104  
 Категоризация, 91, 237, 252, 265  
 Клетки  
     M-типа, 63  
     P-типа, 63  
     с off-центром, 60  
     с on-центром, 60  
 Когнитивные функции, 21, 114, 125  
 Колбочки, 53, 54, 55, 66, 88, 149  
 Компьютерная томография, 101  
 Компьютерное зрение, 253, 264, 265, 268, 273, 280  
 Коннектом, 13, 58  
 Контралатеральный сигнал, 64  
 Кора головного мозга, 31, 102, 124, 140, 152, 175  
 Кортекс, 103, 124, 210  
 Кортикальные ткани, 125  
 Кортикальный гомункулус, 114

**Л**

- Латеральная область, 125  
 Латеральное коленчатое ядро, 19, 66, 123  
 Лобная кора, 39  
 Локализация объектов, 252, 253  
 Локальные цепи, 97

**М**

- Магнитно-резонансная томография, 101  
 Магноцеллюлярные клетки, 63  
 Математические модели, 40  
 Матрица ошибок, 227, 229, 230  
 Машина опорных векторов, 20  
 Медиальная височная доля, 153  
 Медиальная область, 125  
 Межполушарные связи, 153  
 Микроэлектрод, 34, 36  
 Миндалевидное тело, 153  
 Модель Ходжкина–Хаксли, 185  
 Модуль долгой краткосрочной памяти, 244  
 Модуляция внимания, 172  
 Монокулярное соперничество, 300  
 Моторная область, 114  
 Мультистабильные восприятия, 38

**Н**

- Неинвазивные методы, 34, 156  
 Нейронная схема, 31  
 Нейронные корреляты зрительного восприятия, 38, 301

- Нейронные сети, 192, 195, 197, 198, 214, 217, 223, 239, 246, 269, 271, 275, 278, 282  
 Неокогнитрон, 19, 210  
 Неокортекс, 124, 125  
 Неокортикальные цепи, 129  
 Нижняя височная кора, 19, 36, 107, 112, 121, 152  
 Нисходящие связи, 126, 210

**О**

- Обнаружение  
     аномалий, 256  
     объектов, 253, 285  
 Обработка изображений, 36, 42  
 Обратная маскировка, 82, 85  
 Обратная проекция, 63  
 Обратное распространение, 213, 214, 247  
 Обучающий набор, 165, 209, 212, 221, 274  
 Обучение с самоконтролем, 258  
 Одностороннее пространственное игнорирование, 106  
 Окклюзия, 77, 161, 162  
 Оператор разности гауссиан, 61  
 Операция свертки, 190  
 Оптогенетика, 98  
 Отключение генов, 97

**П**

- Палочки, 53, 54, 66, 88  
 Парагиппокампальная извилина, 153  
 Парвоцеллюлярные клетки, 63  
 Первичная зрительная кора, 20, 31, 102, 121, 125, 127, 158  
 Перекрестная проверка, 212, 274  
 Периринальная кора, 153  
 Перцептивная слепота, 298  
 Перцептрон, 186, 296  
 Пирамидные нейроны, 126  
 Плексiformные клетки, 54  
 Повреждения, 33, 39, 94, 95, 99, 101, 102, 105, 106, 107, 108, 111, 121, 152  
 Подавление вспышки, 300  
 Поле зрения, 49, 56, 57, 62, 65, 66, 89, 102, 103, 115, 116, 125, 134, 172, 268  
 Популяция нейронов, 166, 296  
 Постсинаптические нейроны, 126  
 Предиктивные сигналы, 241  
 Предпочтения  
     настройки ориентации, 134  
     нейронов, 134, 141  
 Пресинаптическая активность, 182  
 Префронтальная кора, 153

Принцип эффективного кодирования, 46  
 Проекция на сетчатку, 73  
 Прозопагнозия, 33, 112  
 Промотор, 99  
 Пространственное внимание, 149  
 Простые нейроны, 131  
 Психофизика, 38, 68, 91

**Р**

Распознавание, 23, 26, 27, 29, 72, 73, 74, 75, 76, 78, 82, 83, 92, 102, 105, 106, 111, 124, 205, 206, 208, 241, 249, 254, 257, 261, 284, 285  
 действий, 254  
 лиц, 27, 29, 75, 266, 284, 285  
 образов, 28  
 Рекуррентная нейронная сеть, 19  
 Рекуррентные нейронные сети, 193, 198  
 Рекуррентные сети Хопфилда, 192  
 Рекуррентные соединения, 128  
 Ретрансляционные клетки, 64  
 Рецептивное поле, 36, 59, 66, 130, 132, 145, 148, 150, 154, 294  
 РНК, 97

**С**

Саккады, 49, 50, 51, 55, 69, 80  
 Сверхмалые ганглиозные клетки, 54  
 Сегментация объектов, 254  
 Сети прямого распространения, 192, 194, 223  
 Сети Хопфилда, 192, 193, 194  
 Сетчатка, 26, 30, 43, 51, 53, 59, 65, 148, 207, 296  
 Симултанагнозия, 106  
 Скотома, 101  
 Слепое пятно, 57, 69  
 Слепота к изменению, 298  
 Слепота невнимания, 298  
 Сложные нейроны, 131, 132  
 Сновидения, 171, 293  
 Сознание, 39, 102, 289, 290, 291, 292, 293, 294, 304, 306, 308  
 Соматосенсорная область, 114  
 Состояние восприятия, 304, 305, 308  
 Состязательные изображения, 277, 280  
 Спайковая частота, 60, 162, 202  
 Стереопсис, 89  
 Стержневые биполярные клетки, 54  
 Стимул, 59, 62, 80, 82, 115, 117, 118, 119, 120, 132, 139, 142, 148, 150, 168, 169, 170, 171, 172, 213, 242, 243, 245, 296, 297, 298, 299, 300, 301

Столбчатая структура, 134  
 Субдискретизация, 190, 191, 242  
 Субдуральные электроды, 113  
 Субъективное восприятие, 37, 38, 39, 50, 91, 300  
 Супрахиазматическое ядро, 30, 63

**Т**

Таламус, 19, 30, 60, 64, 143  
 Теория зрительного распознавания, 29  
 Тестовое изображение, 197, 231  
 Тестовый набор, 165, 209, 212, 221  
 Тolerантность, 27, 28, 29, 42, 73, 78, 92, 131, 151, 160, 162, 163, 246  
 Треугольник Канижа, 69

**Ф**

Фальсифицируемые модели, 177  
 Фиксация, 81  
 Фовеальная область, 55, 103, 134, 145  
 Фосфены, 114, 115  
 Фоторецепторы, 53, 54, 295  
 Функция автокорреляции, 45  
 Функция Габора, 135

**Х**

Хрусталик, 52

**Ц**

Цветовая слепота, 54  
 Центральная ямка, 55  
 Центр фиксации, 55, 103  
 Цитоархитектонические карты, 102

**Э**

ЭКГ, 29, 156  
 Экспрессия генов, 97  
 Экстраполяция, 91  
 Эксцентризитет, 55, 60, 103, 115, 145, 204, 258  
 Электростимуляция, 33, 115  
 Энторинальная кора, 153  
 Эффект инверсии, 72

**Ю**

Юмор, 283

**Я**

Явное представление, 148, 296

Книги издательства «ДМК ПРЕСС»  
можно купить оптом и в розницу  
в книготорговой компании «Галактика»  
(представляет интересы издательств  
«ДМК ПРЕСС», «СОЛООН ПРЕСС», «КТК Галактика»).

Адрес: г. Москва, пр. Андропова, 38;  
тел.: (499) 782-38-89, электронная почта: [books@aliants-kniga.ru](mailto:books@aliants-kniga.ru).

При оформлении заказа следует указать адрес (полностью),  
по которому должны быть высланы книги;  
фамилию, имя и отчество получателя.

Желательно также указать свой телефон и электронный адрес.  
Эти книги вы можете заказать и в интернет-магазине: [www.a-planeta.ru](http://www.a-planeta.ru).

Гэбриел Крейман

## **Биологическое и компьютерное зрение**

Главный редактор *Мовчан Д. А.*  
*dmkpress@gmail.com*

Зам. главного редактора *Сенченкова Е. А.*  
*Перевод Люско И. Л.*

Научные редакторы *Киселева Т. Б., Люско Т. И.*

*Корректор Синяева Г. И.*

*Верстка Чаннова А. А.*

*Дизайн обложки Бурмистрова Е. А.*

Гарнитура PT Serif. Печать цифровая.

Усл. печ. л. 25,51. Тираж 200 экз.

Веб-сайт издательства: [www.dmkpress.com](http://www.dmkpress.com)

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 рассматриваются биологические основы зрения. Обсуждаются различные аспекты зрительного восприятия, показано, какие структуры нейронов его обеспечивают, какие процессы происходят в мозге при распознавании образов. Затрагивается тема иллюзий и расстройств зрительного восприятия.

Во второй части представлены различные структуры математических моделей – нейронных сетей, а также принципы их работы. Рассказывается, в частности, о глубоких сверточных нейронных сетях, машинном обучении и генеративных состязательных сетях.

Последняя глава посвящена исследованию самого феномена сознания.

Среди рассматриваемых тем:

- структура нейронных цепей в мозге человека;
- результаты новейших исследований зрительного восприятия животных и человека;
- современные разработки в области искусственного интеллекта: распознавание визуальных образов математическими моделями, основанными на нейронных сетях, построенных по подобию биологических;
- эксперименты в биологических и вычислительных исследованиях зрения, проверка и критическое осмысление гипотез в этой сфере.

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



Гэбриел Крейман – профессор Центра мозга, разума и машин, а также детской больницы Школы медицины Гарвардского университета. Лауреат премии New Innovator Национального института здравоохранения (NIH), премии за достижения Национального научного фонда (NSF), премии Писарта за исследования зрения, а также премии Фонда исследований Макнайта.

Интернет-магазин:  
[www.dmkpress.com](http://www.dmkpress.com)

Оптовая продажа:  
КТК «Галактика»  
[books@aliants-kniga.ru](mailto:books@aliants-kniga.ru)



ISBN 978-5-93700-100-9

9 785937 001009 >