北京智源人工智能研究院
BEIJING ACADEMY OF ARTIFICIAL INTELLIGENCE

白皮书 | 人工智能的认知神经基础
Brain and Machine Intelligence

智源人工智能的认知神经基础重大研究方向 编著

北京智源人工智能研究院
2022 年 1 月

版权声明

指导专家

刘　嘉 智源首席科学家，清华大学脑与智能实验室研究员
宋　森 智源研究员，清华大学生物医学工程系研究员
吴　思 智源研究员，北京大学心理与认知科学学院教授
方　方 智源研究员，北京大学心理与认知科学学院教授
余　山 智源研究员，中国科学院自动化研究所研究员
陈良怡 智源研究员，北京大学未来技术学院教授

编写组成员

张　博 智源博士后，人工智能的认知神经基础重大研究方向
苏　杰 智源博士后，人工智能的认知神经基础重大研究方向
蒋龙生 智源博士后，人工智能的认知神经基础重大研究方向
陈智强 智源博士后，人工智能的认知神经基础重大研究方向
陈路瑶 智源博士后，人工智能的认知神经基础重大研究方向
邹晓龙 智源博士后，人工智能的认知神经基础重大研究方向
刘　祥 智源博士后，人工智能的认知神经基础重大研究方向
徐琳璐 智源博士后，人工智能的认知神经基础重大研究方向
秦方博 中国科学院自动化研究所助理研究员
韩　程 中国科学院自动化研究所博士研究生

搭建脑科学与人工智能的桥梁

智源研究院院长 黄铁军

智源研究院 2021 年度《人工智能的认知神经基础白皮书》如期和大家见面了！延续去年的传统，今年的白皮书盘点了神经科学、认知科学、智能技术等相关领域的重要进展；同时，与去年不同的是，除了从认知科学和神经科学两大领域系统梳理重要进展及对人工智能的启示外，今年还集中介绍了类脑视觉、脑机接口和交叉学科技术这三个方向的热点和趋势，以飨读者！

脑科学对人工智能的重要性不言而喻。把人工智能这个概念送上历史舞台的 1956 年达特茅斯夏季研讨会共讨论了七大问题，问题 3 就是“神经网络：一群神经元是如何形成概念的？”，我认为这是人工智能需要回答的最重要的问题，也是脑科学需要回答的最重要的问题。

“一群神经元”，这是神经科学的研究对象，“形成概念”，这是认知科学的研究对象，这个最重要的问题，正是认知科学和神经科学的连接点。认知科学研究智能现象，主要采用自顶向下的方法，神经科学研究脑的结构，主要采用自底向上方法。

认知科学和神经科学都属于脑科学，它的研究对象是脑及其智能现象，被称为“自然科学的最后疆域”，进展速度不如人工智能那么让人眼花缭乱。这是因为，人工智能是一门技术，目的是构造越来越智能，因而越来越复杂的系统，它的进步比较容易看得到。相比之下，生物神经系统是个盘根错节的黑暗丛林，生物智能是复杂的动力学现象，还缺乏有效的数学工具，因此任何一点儿进步都十分艰难。

人工智能并不能因为进步快而沾沾自喜。当前人工智能系统和生物神经系统相比，还是小巫见大巫。例如智源研究院去年发布的人工智能大模型“悟道 2.0”，参数规模达到 1.75 万亿，但还不到人类大脑连接数量的 2%，而且其基本单元和连接方式都比生物系统简单得多。视觉是研究人员最多、应用最广的方向，但是已有视觉模型都难望生物视觉之项背，今年热点是视觉大模型，如果要在像素级进行视觉空间关系训练，集合全球算力都不够，更遑论时空关系联合训练。

说到算力，人们往往会说强大的人脑是个低功耗系统，这是认识错位。用人工智能的术语来说，人脑的低功耗是“推理”过程低功耗，而不是“训练”过程低功耗。人脑是亿万年进化的产物，进化就是一种训练过程，大自然训练出人脑这个复杂网络，消耗了巨量太阳能，相比之下，全球算力功耗算得了什么呢？

这就是人工智能离不开脑科学的原因。以“机器学习+大数据/复杂环境+大算力”模式训练大规模智能模型，确实可以解决不少问题，但天下没有免费的午餐，强大智能是以巨大训练成本为前提的，训练人脑花费的“天价”，人类付得起吗？因此，借鉴生物大脑这个已经训练成功的“蓝本”，模拟生物大脑的精细神经结构和信息加工机理，却可能是实现更强大、更通用人工智能的最短路径。

借鉴脑科学研究成果，并不是说默默等待脑科学最新进展，事实上，脑科学大量已有进展尚未在人工智能领域得到有效利用。例如，目前人工神经网络所用的神经元模型，还是 1943 年的麦卡洛克-皮茨（M-P）模型，训练的理论依据，还是 1949 年提出的赫布学习规则（Hebb Learning Rule）。在脑科学领域，有许多与智能行为密切相关的认知范式、神经活动机理等“宝藏”等待人工智能领域研究者开发和利用，并以此推动生物智能启发的人工智能模型算法研究新范式。

因此，智源研究院于 2020 年 8 月，设立“人工智能的认知神经基础重大研究方向”，就是要促进脑科学和人工智能的交叉，促进两个领域学者的交流和合作。作为认知神经基础重大研究方向的重要成果，智源生物智能开源开放平台已经在去年正式上线。同时智源研究院还在去年设立了生命模型研究中心，从模拟高精度生命系统的角度开展交叉领域前沿探索。

为了进一步加强脑科学和人工智能的合作，架起连接脑科学与人工智能的实际桥梁，我专门造了一个新词：“智元（Wiston）”，意思是具有独立智能功能的基本神经回路。事实上，脑科学已经发现了很多“智元”，例如这份报告第 2 章提到的位置细胞和网格细胞、第 3 章提到的吸引子网络、赢者通吃网络，众所周知的视皮层简单细胞和复杂细胞，以及近期热门的记忆痕迹细胞等，已经遍及感知、定位、学习、决策、记忆等多种智能。可惜的是，这些进展都没跳出“细胞/神经元”这个神经科学术语，因此我提出“智元”概念，就是要把相对独立的智能和实现这种智能的一群神经元（及其网络连接）作为一个整体单元。以“智元”作为基本单元构造的人工智能系统，将是可解释、可预期和可信任的。

当然更重要是，从“智元”开始，我们就已经开始回答“一群神经元是如何形成概念的？”这个最重要的问题了。

前 言

近年来人工智能技术得到了快速的发展，引起了各界的广泛关注。随着计算机算力和大数据可及性的快速提升，以深度人工神经网络为核心的人工智能系统在物体识别、自然语言处理等领域取得了令人瞩目的成绩，在围棋、星际争霸等竞技游戏中一骑绝尘，甚至在蛋白质结构解析、提出和解决数学难题等方面展现出超越人类专家的潜力。但目前的人工智能与通用智能之间，还存在巨大的能力鸿沟。而大脑作为通用智能的唯一样本，为人工智能的发展提供了重要参照。智源 “人工智能的认知神经基础”重大方向（Brain and Machine Intelligence）旨在从生物脑如何实现智能的角度，对于人工智能的发展提出有启发的问题，提供可资借鉴的原理、模型、算法和系统实现方案，从而促进类脑智能的发展，推动人工智能向人类水平，甚至超越人类的水平逐渐逼近。每年发表的白皮书就是我们的尝试之一，希望通过它向大家梳理脑科学、认知科学和类脑智能方向上最值得关注的动态和进展，并分享我们对于这些方向未来发展趋势的思考。

计算神经科学的先驱，英国科学家 David Marr 曾经提出，可以从三个层面理解脑的工作原理，首先是计算的层面 (Level of Computation)，即脑在做什么计算，以及为什么要做这个计算；其次是表征/算法的层面(Level of Representation/Algorithm)，即脑在计算过程中的信息如何表征，选择什么算法来实现计算目标；最后是物理实现的层面(Level of Implementation)，即脑选择什么样的硬件实现形式来执行这些计算。今年的白皮书中，上述三个层面的研究进展都会有所涉及。

在计算层面，我们重点介绍了具身认知（Embodied Cognition）理论和全局工作空间（Global Workspace Theory, GWT）理论。与当前主流人工智能主要基于被动观察与识别，往往不具有具体物理形态的范式不同，具身认知认为，认知过程无法脱离身体而进行，推广开来，整个环境和个体的行为同样是认知的重要组成部分。个体通过感知外部环境，进行决策，生成相应动作与环境交互，以此改变环境，这个过程周而复始，促成了智能的形成和发展。全局工作空间理论则是

由美国心理学家 Bernard Baars 在上世纪 80 年代作为一种意识模型而提出的认知架构，后来发展为“全局神经元工作空间”（Global Neuronal Workspace, GNW）。GNW 如同一个分布式路由器，同各个脑区的众多神经元存在关联，从而可以放大、维持信息，并提供给各个处理模块使用，从而实现全局的信息共享和处理。

在表征/算法层面，我们今年聚焦于脑中认知地图的表征以及神经流形这两个重要的研究领域。位于脑中海马体及其邻近脑区中存在表征空间特征的位置细胞（Place cell）和网格细胞（Grid cell），近年来的研究揭示这一系统可能不仅涉及空间记忆与导航，而且可能参与了物理空间认知以外的信息处理，比如图片空间、嗅觉空间，甚至关系空间的表征，提示脑中可能用一套通用的机制在处理一系列表面上截然不同，但是具有深刻共性的信息维度。神经流形(Neural manifold)则是利用动力学的理论和观点来理解众多神经元构成的群体如何开展高效计算的有力工具。通过流形向量场这一精确的数学语言对神经电生理信号进行分析已经开始回答很多有关神经群体编码的关键问题。

在物理实现层面，我们重点介绍了受生物视网膜启发的动态视觉传感器（Dynamical vision sensor，简称 DVS）和脉冲摄像头（Spiking camera）。与传统的视觉传感器不同，这两类模拟视网膜的感知设备能够将图像信息转化为脉冲事件流进行表征，具备高动态范围、高时间分辨率、低能量消耗以及高像素带宽等特性。相应的，我们也系统地梳理了适宜于处理脉冲事件流信号，并可以开展运动目标快速探测、有效跟踪和精确识别的类脑视觉计算模型和算法。

在上述三个方面的内容之外，我们还针对脑科学与类脑智能研究中近年来涌现的新技术，特别是脑机接口技术、新型脑成像、连接组学与数据处理方法等进行了梳理和介绍。脑机接口通过对于脑活动信息的检测和调控，在脑与外部世界间建立直接的信息通讯接口。这一技术的发展，有望对于人与环境、人与人的交互方式带来根本变化，从而引起社会、经济、教育、军事、医疗等众多领域的颠覆性变革。新型脑成像、连接组学与数据处理方法，展现了以往观察不到的神经活动细节，解析了神经网络中各部分的相互作用机制，从而促进人们进一步理解神经系统的设计原则。

编写白皮书的过程是我们一年一度盘点神经科学、认知科学、智能技术等相关领域重要进展的过程，也是我们不断思考什么是智能，以及如何发展类脑智能的过程。希望这些努力能让对于这些领域的进展感兴趣，也对回答这些问题感兴趣的读者有所收获。与此同时，经过人工智能的认知神经基础方向各位同仁一年多的努力，智源生物智能开源开放平台（Bio-Intelligence Opensource Platform, BIOSP）已经在 2021 年正式上线，该平台旨在通过开源开放数据、模型、算法、软件工具等一站式科研资源的方式，为认知科学、神经科学和计算科学及相关交叉领域的研究人员、学生和相关从业者搭建一个服务智能科学研究的平台型基础设施，进而推动和支撑国内脑启发的通用智能研究工作。希望每年一版的白皮书和不断完善的开源开放平台能够助力中国脑-智研究的交叉融合，促进类脑通用智能的早日实现。

目 录

第1章 认知科学对人工智能的启示

近年来，人工智能领域在第三次浪潮爆发后经历了快速的发展，许多特定领域的专用人工智能算法已经大幅度超越了人类的水平，并在工业生产和社会生活中得到了广泛的应用。尽管如此，主流的观点仍然认为，目前深度学习算法的本质依然是海量数据驱动的统计学习，距离人类更加复杂的高级认知功能仍然存在本质上的差别。如何弥补这种差异，从而推动人工智能从弱人工智能到强人工智能的转变，已经成为许多从业者开始思考并着手解决的重大难题。

认知科学（Cognitive Science）是一门研究认知如何工作的交叉学科，自诞生之初便与人工智能有着密不可分的关系。认知科学的相关理论数次推动了人工智能的发展，而人工智能作为人类模拟大脑功能的尝试，其本身也可以看作是认知科学理论的一种实践和验证。在本章中，我们将简要介绍认知科学的具身主义流派以及可能对实现通用人工智能具有一定指导意义的全局工作空间理论，并对它们与人工智能的关系做一些简单梳理。

1.1 具身主义认知科学的兴起

1.1.1 符号主义与联结主义认知科学

在探索智能的道路上，现代意义的认知科学主要经历了两个时代：符号主义时代（Symbolism）和联结主义时代（Connectionism）[1]。符号主义尝试通过操作具有特定含义的符号来实现“智能”，这一思想被后人概括为物理符号系统，典型的例子是 Alan Turing 在 1936 年提出的图灵机概念（图 1.1 左），通过读写头在纸带上标记二进制信息（有孔和无孔）来实现相应的计算功能。图灵机概念的成功让以 Allen Newell 和 Herbert A. Simon 为首的研究者们相信，通过对符号进行操纵，有限的符号最终可生成无限的信息，最终实现智能。符号主义浪潮推动了电子计算机的发展，使其在 20 世纪的战争、工业、甚至我们的生活中被广泛使用，而基于符号主义的人工智能也取得了专家系统、计算机推理等诸多辉煌的成就……尽管当时许多研究者认为真正意义上的人工智能近在眼前，但符号主义在那些不适定问题（ill-posed problems）上却屡屡受挫，止步不前。

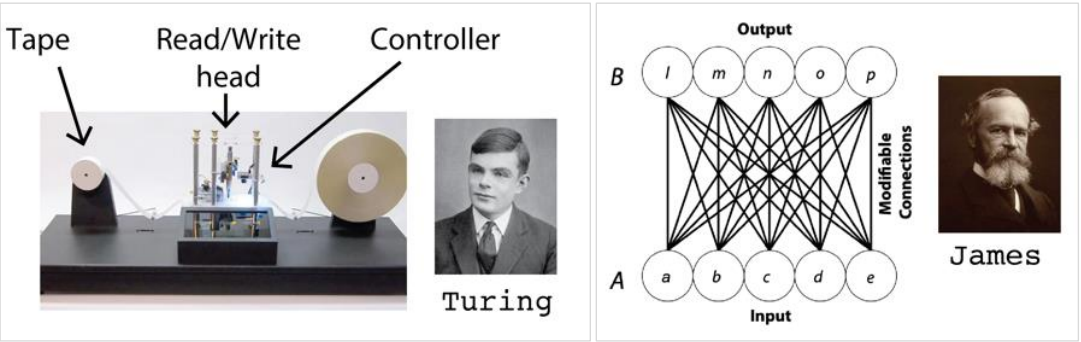


图 1.1 左：符号主义的代表——图灵机；
右：哲学家 William James 在 1890 年提出的最早的连接主义模型[1]

人们开始意识到，古典认知科学所倡导的符号主义衍生出的产品与人脑的智能相差甚远。要实现智能，应该让机器的运作机制向人脑的神经元机制靠拢，由此，受神经科学的发展推动，联结主义时代到来，虽然人工神经网络的雏形早在 1890 年已经由哲学家 William James 提出（图 1.1 右）。相比于物理符号系统直接读取特定的符号信息，人工神经网络尝试读取输入源的统计形态信息，并以表征的形式在输入和输出信号之间建立统计关系，以达到学习和预测的目的。在经

历了几次起起落落之后，当前，由联结主义思想衍生出的深度神经网络（Deep Neural Network，DNN）已取得了巨大成功，尤其在人脸识别、图像重建等领域，深度神经网络为人们的生产生活提供了许多便利。

值得注意的是，符号主义与联结主义虽然源自不同的哲学思想，但并不意味着两者水火不相容。基于联结主义的神经网络虽然能够很好的解决图像分类、识别、语音识别、语义理解等任务，但其背后的原理和可解释性问题一直困扰着人们，而符号主义有着更深刻的哲学和数学基础，在处理串行等问题上更加简洁有效。因此，近年也有一些研究者尝试构造混合模型，以综合这两者的特长。

历史上，人工智能的几次繁荣和低谷都与符号主义和联结主义认知科学的发展密切相关（图 1.2）。虽然基于联结主义思想的深度神经网络目前还处在发展的高峰，但受限于样本量小、泛化能力差、能耗大、语义理解欠缺等瓶颈，当前的深度神经网络所达到的“智能”与人们所向往的类脑通用智能还相差甚远。那么，我们如何做才能实现这样的智能？结合神经科学近年来的重要发现，我们认为，以 Lawrence Shapiro 为代表的学者提出的具身主义浪潮会在不远的将来到来。

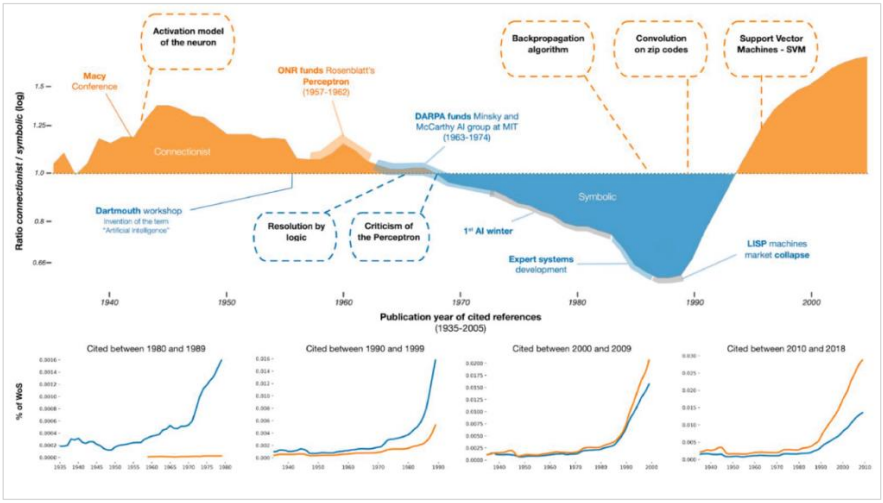


图 1.2 符号主义与联结主义相关文献数量随时间的变化，修改自[2]

1.1.2 具身认知与强化学习

古典认知科学中的三明治模型（sandwich theory）认为，由智能驱动的认知过程可以视作一个由感知、思考、和动作（sense-think-act）这 3 个独立的

元素所构成的回路[1]（图 1.3），通常人们主要关注的是其中的 Think，却有意无意的将另外两部分弱化。而具身认知（Embodied cognition）认为，人的认知过程无法脱离身体而进行，推广开来，整个环境和个体的行为同样是认知的重要组成部分，个体（agent）通过感知外部环境，产生思想并通过计算后，生成相应动作与环境交互，以此改变和影响环境，这个过程周而复始，这就是智能。

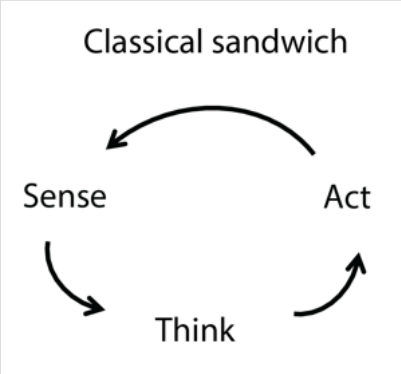


图 1.3 古典认知科学中的三明治模型[1]

地球上的动物经历了几亿年的演化（Evolution）而表现出了显著的具身智能，这使得它们能够在复杂的环境中生存、学习，并与其他个体、其他物种和环境进行交互。在行动中，动物为了趋利避害往往会更加频繁的采取对自己有利的行为策略。经过一段时间的学习之后，这些行为被强化（reinforce），甚至变成习惯而固定下来，这种学习方式称为强化学习（Reinforcement Learning）。在强化学习中，智能体不断与环境进行交互并得到反馈（Feedback），通过试错（trial-and-error）的方式去总结哪些行动可能会带来更好的收益（Reward），以便于更好的适应环境。如果我们把时间尺度放大，在个体的强化学习之外，自然或环境本身还会提供一种优化算法，即通过自然选择筛选种群，并通过基因突变来避免陷入局部极值点。

基于具身认知，李飞飞团队提出了一个同时包含这两者的计算框架，称为深度进化强化学习（Deep Evolutionary Reinforcement Learning, DERL）[3]。在该框架下，智能体可以在多个复杂环境中执行不同的任务。在这项研究中创建的具身智能体可以在平地、多变地形等不同环境中执行巡视、导航、避障、探索、逃脱、爬坡、推箱子和控球等多种不同的任务（图 1.4）。DERL 为计算机模拟实验中大规模具身智能体的创建打开了一扇门，这有助于获得有关学习和进化如何

协作以在环境复杂性，形态智能以及控制的可学习性之间建立复杂关系的科学见解。此外，DERL 还减少了强化学习的样本低效性的情况。智能体的创建不仅具有所需使用的数据更少的优势，而且还可以泛化解决其他多种形式的新任务。

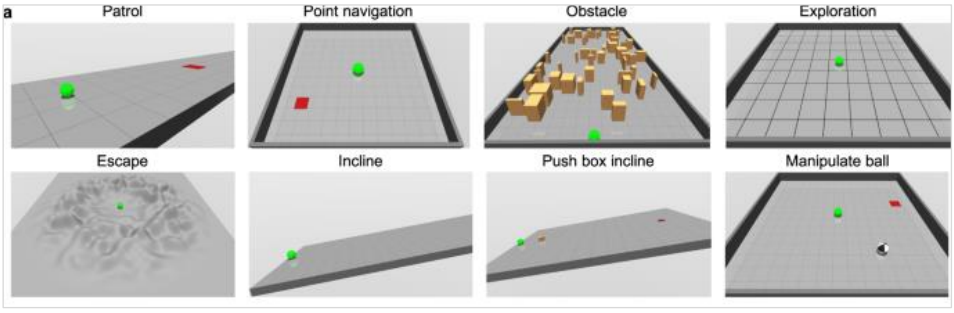


图 1.4 具身智能体能够在不同环境中执行多种任务[3].

无独有偶，DeepMind 团队也进行了相似的研究[4]，通过自动生成大量不同的环境和游戏目标，智能体可以接受各种各样任务的训练（图 1.5），在大规模的开放（Open-Ended）环境中，智能体甚至学会了举一反三，做到了现有深度神经网络难以做到的零样本学习（Zero-Shot Learning）。强化学习和进化对于具身智能体和通用智能的重要性可见一斑。

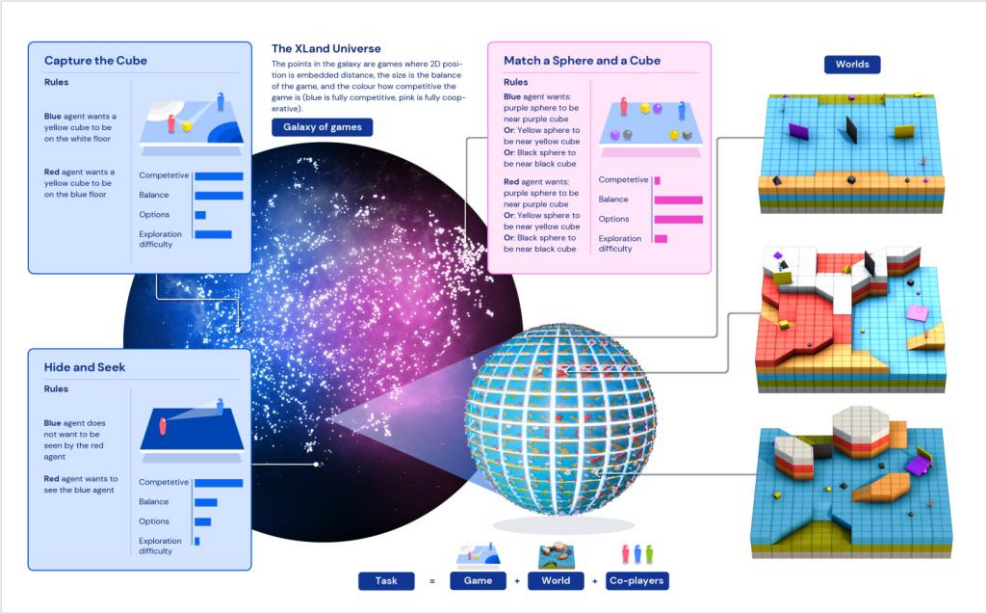


图 1.5 Open-Ended Learning 中的开放环境 XLand[4]

具身智能体的一个显著特征是可以利用不同的感觉器官获取环境的信息进行整合，并执行多种不同的任务。例如，动物们可以通过视觉、听觉、触觉等不

同的感官获取环境信息，并进行觅食、逃跑、迁徙等等。而当前的人工智能大多只能执行非常单一的或者少数任务，即便是 DeepMind 宣称的可以执行几百万种不同任务的智能体，仍然有具体任务相似性太高的缺陷。与之形成鲜明对比的是，生物智能体能够执行的任务种类要多得多，并且通常涵盖多种不同的认知功能。

在认知科学中，我们经常通过不同的任务范式去研究人类智能的一些特征甚至缺陷。这些任务范式通常是为了特定实验目的定制的，然而其中的一些单一任务对于目前的人工智能而言仍然有一定的难度，更不用说让智能体同时完成多种任务。值得一提的是，在智源的生物智能开源开放平台中，我们开放了 30 多种不同的人类认知行为范式，实验主题包括客体识别、注意、记忆、语言、数量感、音乐、空间认知等，每种任务都包含大量人类被试的行为数据。我们认为，这些任务可以供新的具身智能体在开放环境中学习用，也可以作为测试任务评估训练后的智能体的认知能力，并与人类智能进行对比。我们希望这批数据能够为人工智能发展多任务能力提供一些帮助。

1.1.3 多智能体交互与共识主动性

在具身认知中，与其他个体的交互也是智能体与环境交互的重要组成部分，不同智能体之间可能存在合作、竞争等不同的交互模式。社会认知（social cognition）通常主要关注多个个体之间，或者个体与群体之间的交互行为。例如，两个或多个个体间可重复进行的社会决策往往在博弈论（Game Theory）的框架下进行研究。这些理论对于多智能体交互固然具有重要的意义，但在大量智能体同时活动的环境中，智能体之间进行直接对话的方式往往并不能达到好的效果，甚至难以完成。

在一些低等动物中，尽管每个个体的智能非常有限，但众多个体组成的群体却能涌现出一定的智能（群体智能）。例如，鱼群能够结队行进，防御捕食者，提高觅食成功率；蚂蚁搬运食物时往往走的是最短的路径等等。每只蚂蚁在它走过的路径上都会留下信息素，并尽可能沿着信息素浓度高的路径前进，而信息素会随时间挥发，于是最短路径上信息素的浓度更高。人们借鉴这种现象创造了蚁群算法和粒子群优化等算法，并且这种现象在无人机编队等多智能体互动中也得

到了充分的关注。

在宏观层面，共识主动性不仅仅出现在低等动物中，根据其定义，人类在社会活动和文明的进程中也会通过共识主动性机制与其他人进行间接的交互，尤其在互联网时代，任何人对于互联网环境都可以造成直接或间接的干预，从而可能对其他人造成或多或少的影响。科研社区、开源社区以及基于区块链技术的金融社区等等都体现出了人类社会中的共识主动性，而在可以预见的将来，当元宇宙普及之后，这种作用可能会更加明显。

在微观层面，大脑的智能也可以看作功能相对单一的大量神经元涌现出的群体智能。同鸟群和鱼群类似，通常只有临近的神经元之间存在直接交流，信息通过这种局部的交互也能够传遍大脑并进行计算加工。事实上，神经生物学的研究表明，神经元的生长发育、突触的建立可能也体现了一种共识主动性：神经元通过发放神经递质、代谢产物等改变其附近的微环境，并利用组织液中的化学物质决定自己的行为，从而与环境中的其他神经元进行间接交互。甚至已经有研究者开始考虑在人工神经网络中加入共识主动性机制。

1.2 全局工作空间理论

1.2.1 人类的认知架构

伴随着具身主义思想的发展，以及多智能体交互需求的不断上升，促使了对环境中个体的认知架构研究。科学家们一直试图将人类的心智（Mind）理论化，并通过形式化建模的方式来构建认知架构，以实现人工智能。认知科学和神经科学近几十年的研究已经表明，大脑是模块化的，不同的区域具有特异的不同功能，例如人脑的梭状回面孔区（fusiform face area, FFA）负责面孔的识别，韦尼克区（Wernicke's area）负责语言语义理解，额叶眼动区（frontal eye fields, FEF）负责扫视运动等等。那么，这些区域如何相互配合，完成“在嘈杂的人群中看到熟人，听到他说话时盯住嘴巴，同时利用嘴型和不甚清楚的声音听懂他在跟你打招呼并走过去聊天”这样的日常行为呢？这就涉及到了我们将要介绍的全局工作空间理论（Global Workspace Theory, GWT）[6,7]。

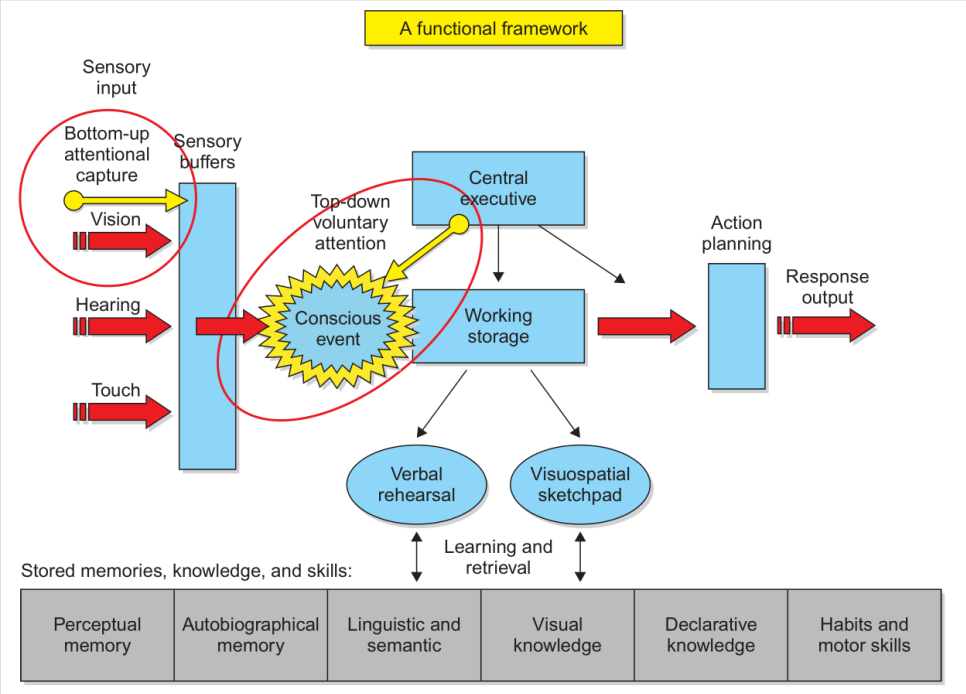


图 1.6 GWT 认知架构的功能框架草图[5]

全局工作空间理论是由美国心理学家 Bernard Baars 在上世纪 80 年代提出的一种认知架构，它最初作为一种意识模型而被提出，是现代认知科学的一个重要理论。该理论认为，大脑可以分成一些具有特定功能的模块，当感知觉输入或任务需求激发了某些模块的响应之后，这些响应会相互竞争，通过选择性注意机制，某些信息会进入全局工作空间，并在不同模块之间进行广播（broadcast），以此完成不同模块之间的信息交流，并合作完成不同的任务。而当信息进入全局工作空间并分发到其他模块时，意识就此产生（图 1.6）。GWT 理论通常可以用“剧场隐喻”（theater metaphor）来理解[8]（图 1.7）。在“意识剧场”中，选择性注意像聚光灯一样照亮了舞台上的一个区域。这个亮点揭示了意识的内容：演员们进行表演、演讲或者相互交流。导演、编剧、场景设计师等工作人员藏在幕后的黑暗中，他们塑造了舞台上的可见活动，但它们本身是不可见的。舞台中正在上演的内容也被播送给同样处在黑暗中的观众（即大脑的其他部分）。

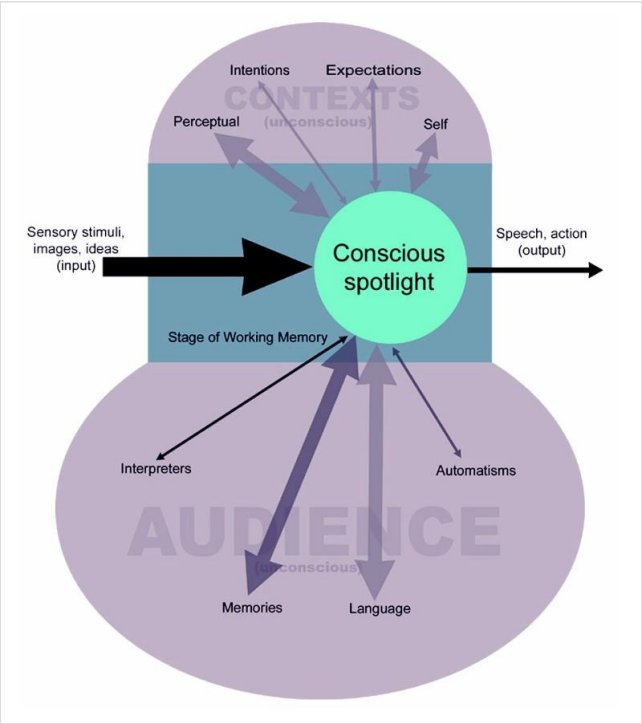


图 1.7 GWT 的剧场隐喻[8]

Dehaene 和 Changeux 等人提出了一个全局工作空间架构的神经元版本，即所谓“全局神经元工作空间”（Global Neuronal Workspace, GNW）[9,10]。在他们的模型中，一些局部的、专用的、模块化的皮层区域构成了一个单独的计算空间，各个模块可能具有各自的层级结构，但不同部分可以并行、分布式处理特定的信息，如感知觉、运动、记忆等等。第二个计算空间是由一些广泛分布的兴奋性神经元（称为 GNW 神经元）和具有长程连接的轴突组成，能够通过下行连接选择性地调动或抑制特定模块传入的信息。在他们的模型中，这种分布式的神经元群体具有自下而上接收信息并将自上而下的信息传输给任何一个处理器的能力，从而选择和广播信息（图 1.8）。这种大范围广播允许不同的认知模块都能够接收到信息，被认为有助于未知问题的解决，例如通过调动不同的信息处理模块进行竞争或合作，从而更容易找到解决问题的路径。

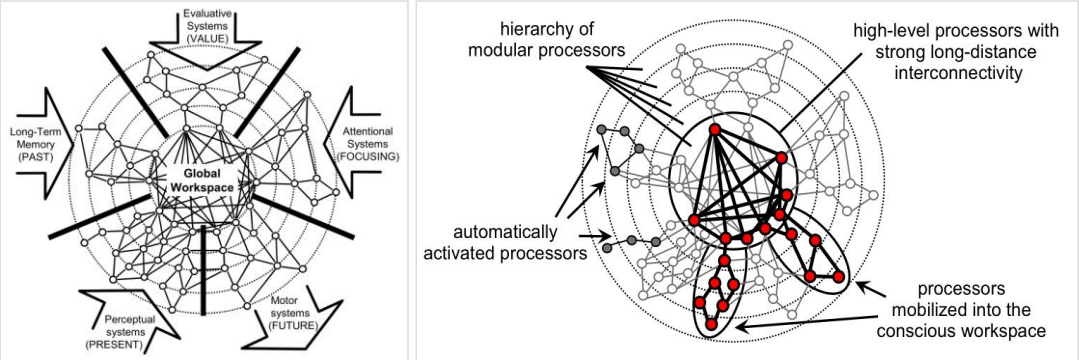


图 1.8 Global Neuronal Workspace [11, 12]

GNW 的激活是非线性的，具有“全或无”（all-or-none）的特性，即一旦有信息进入，便会迅速诱发全局工作空间的广播，这称作“引燃”（ignition），这种现象已经在人和动物的实验中得到了证实（图 1.9）。引燃可能由外部刺激所触发，例如黑暗里的一盏灯、周围车辆的鸣笛；或者受正在执行的任务相关脑区触发，例如在回忆时发生“知晓感”（feeling of knowing），话到嘴边却无法提取记忆内容；甚至可能在休息时自发随机产生。GNW 还具有独占性（exclusive），某群神经元的激活能够抑制其余的神经元，如果某个模块的信息激活了全局的活动模式，其他模块的信息将无法进入全局工作空间，因此全局工作空间只能够串行处理信息，并且不同子系统之间会存在竞争。这种机制符合意识的一些特征，例如状态单一，容量有限、顺次发生，也能够解释诸如非注意盲视（Inattentional Blindness）、注意瞬脱（Attentional Blink）等认知现象。

GNW 如同一个分布式路由器，同各个脑区的无数神经元存在关联，从而可以放大、维持信息，并提供给各个信息处理模块和丘脑皮层环路使用。大脑的前额叶皮层（prefrontal cortex, PFC）、背外侧前额叶皮层（dorsolateral prefrontal cortex, DLPFC）、下顶叶皮层（inferior parietal cortex）、前颞叶皮层（anterior temporal cortex）、前后扣带回皮层（anterior/posterior cingulate cortex, ACC/PCC）、楔前叶（precuneus）等脑区，各自有其独特的功能和连接模式，但相互之间存在广泛的连接，任何一个区域获取的信息都可以迅速提供给其他脑区。这些脑区之间密切的双向连接为引燃（ignition）创造了条件，从而能够触发突然的、集体的协同活动在全脑广播。

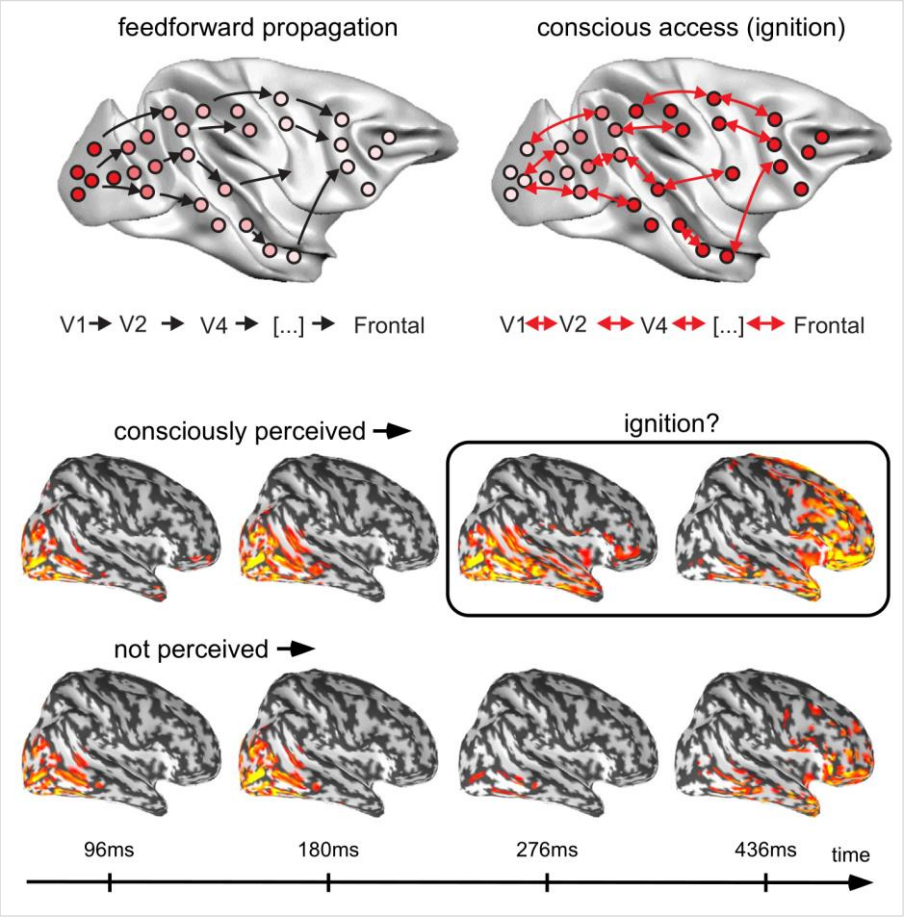


图 1.9 猕猴和人脑中的引燃[13]

2017 年，Christof Koch 团队在小鼠的屏状核（claustrum）发现了三个巨大神经元（Giant Neuron）[14]，这些神经元跨越大脑的两个半球，缠绕在整个大脑周围，与大脑负责感觉信息、负责行为反应的许多区域都有连接，在神经元层面符合全局工作空间的特征，被认为可能是意识的开关。

GWT 不仅仅是一个概念模型，Dehaene，Changeux 等人提出的神经元动力模型（Dehaene-Changeux Model，DCM）即为 GNW 的一种计算机模拟[15]。通过分别建模单个神经元、丘脑皮层柱网络和具有长程连接的由网络组成的网络（图 1.10），DCM 模拟了生物脑中观测到的丘脑-皮层震荡，以及网络自发或刺激诱发的引燃（ignition）等现象。

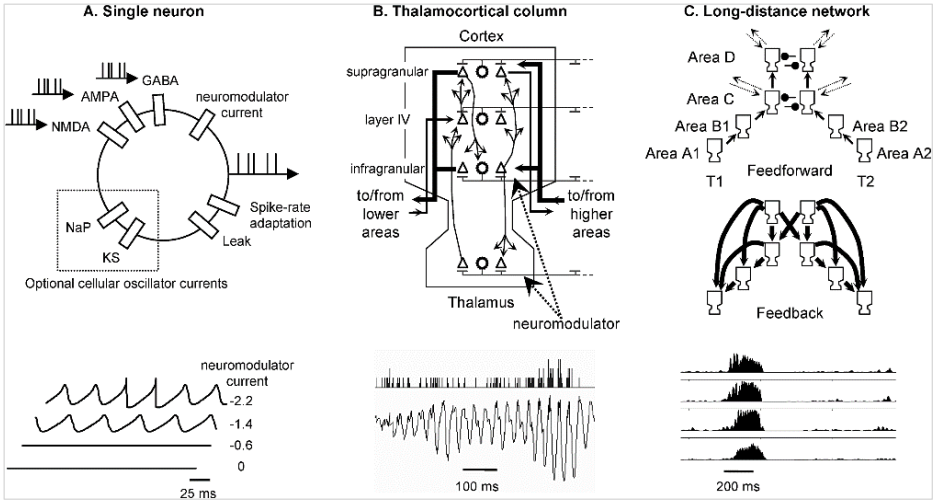


图 1.10 Dehaene-Changeux 模型[15]

Franklin 等人在全局工作空间理论的基础上构建了一个更加通用的认知架构，称为学习型智能分配代理（Learning Intelligent Distribution Agent, LIDA），是一种生物学启发的综合性、可以计算实现的概念模型[16]。LIDA 模型以 LIDA 认知循环（cognitive cycle）为基础（图 1.11）。LIDA 把认知循环看作是一个认知原子，其中包含了更高层次的认知过程、思考、推理、问题解决、计划、想象等。每个认知循环分为三个阶段：感知理解阶段、注意阶段以及动作选择和学习阶段，各个阶段分别由若干相互作用的模块构成，如图 1.11 所示。在每个认知周期中，LIDA 智能体首先通过更新其对环境外部和内部特征的表示，尽可能好地理解其当前的状况（current situational model）。通过一种竞争过程，它决定哪些信息最需要注意，并将这些信息广播，使其成为当前意识的内容，于是智能体能够选择适当的行动去执行。需要指出的是，LIDA 认知循环中的各个模块并不与大脑中的功能模块直接对应，它们更多的是一种思维或心智意义上的功能模块。虽然模块在图中用明显的边界表示，但它们有非常丰富的交互，可能很难清晰的拆分开。另外，在 LIDA 模型中，除了意识和行为选择部分以外，其他过程都可以异步、并行的处理。

LIDA 模型实现并充实了全局工作空间理论，并且涵盖了人类认知的很大一部分，为许多认知过程提供了合理的解释，被认为有可能作为理解心智如何运作的工具。同时，LIDA 框架被认为可能对通用人工智能（AGI）的实现具有重要的帮助[17,18]。除此之外，Blum 等人还基于 GWT 构建了意识图灵机（Conscious

Turing Machine，CTM，图 1.12），认为可以用于构建具有意识的人工智能系统。

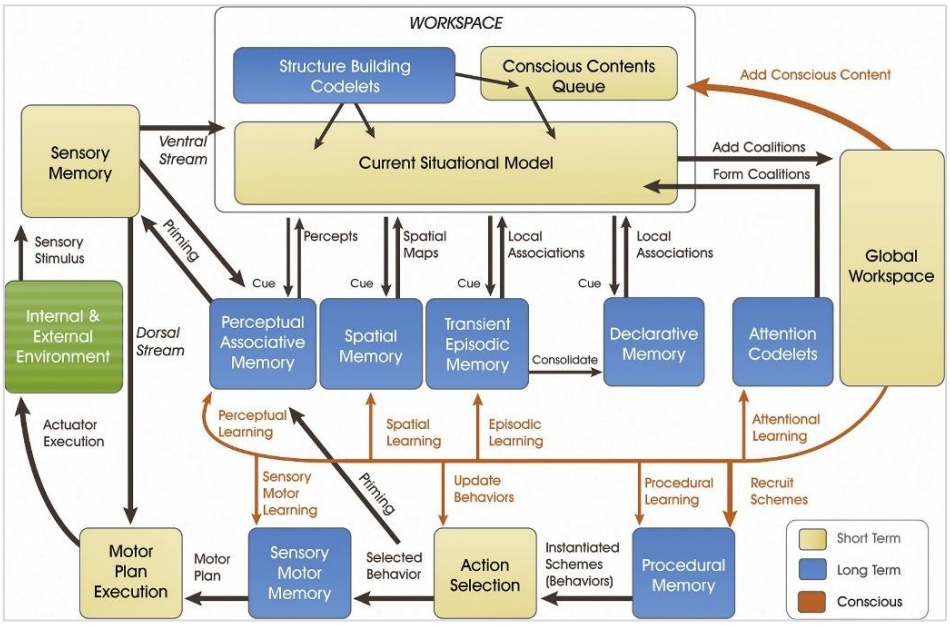


图 1.11 LIDA 模型中的认知循环[19]

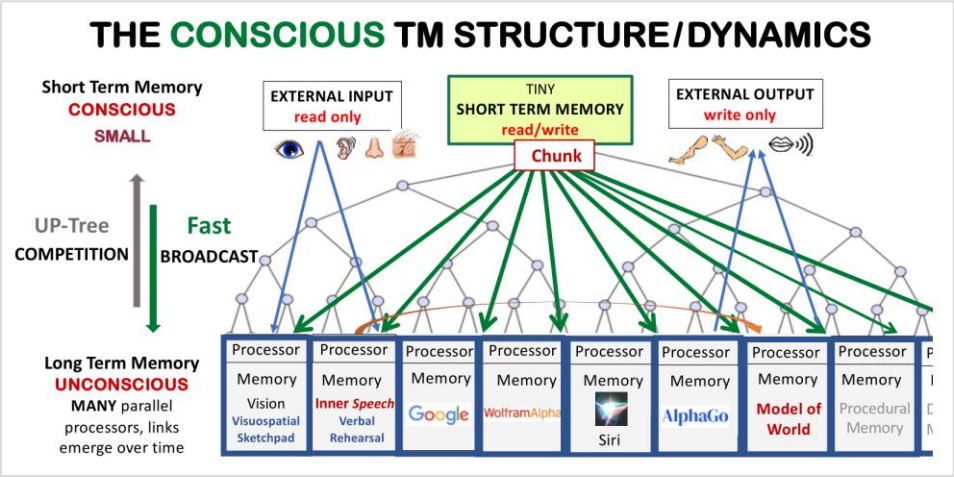


图 1.12 意识图灵机（CTM）[20]

1.2.2 元认知与元学习

现代计算科学的创始人艾伦·图灵（Alan Turing）和约翰·冯·诺依曼（John von Neumann）认为，机器最终能够模仿包括意识在内的大脑的所有能力，而当前的深度学习和人工智能所解决的计算问题主要与人脑中的无意识认知加工相对应。意识似乎是实现通用人工智能（AGI）的过程中无法避开的一个话题，针

对机器能否拥有意识的问题，Dehaene 等人提议将人类的意识相关计算分成三个水平[21]。

无意识加工（unconscious processing, C0）包括了大部分人类的智能，例如知觉恒常性、语义提取、决策、学习等，大多在潜意识或无意识状态即可完成。图 1.13（上）展示了面孔加工中潜意识下的视觉不变性（subliminal view-invariant），如果首先呈现同一个人的面孔进行阈下刺激，即便是完全不同视角的照片也能促进面孔信息的加工，并降低 FFA 区域的激活强度，这种现象称为潜意识启动（subliminal priming）。图 1.13（下）的双眼抑制实验中，阈下刺激也能够进行有效的证据积累，从而影响正确率和反应时间。此外，在强化学习中，即使线索、奖励等信号低于意识的阈值，人类的学习过程也能继续进行。

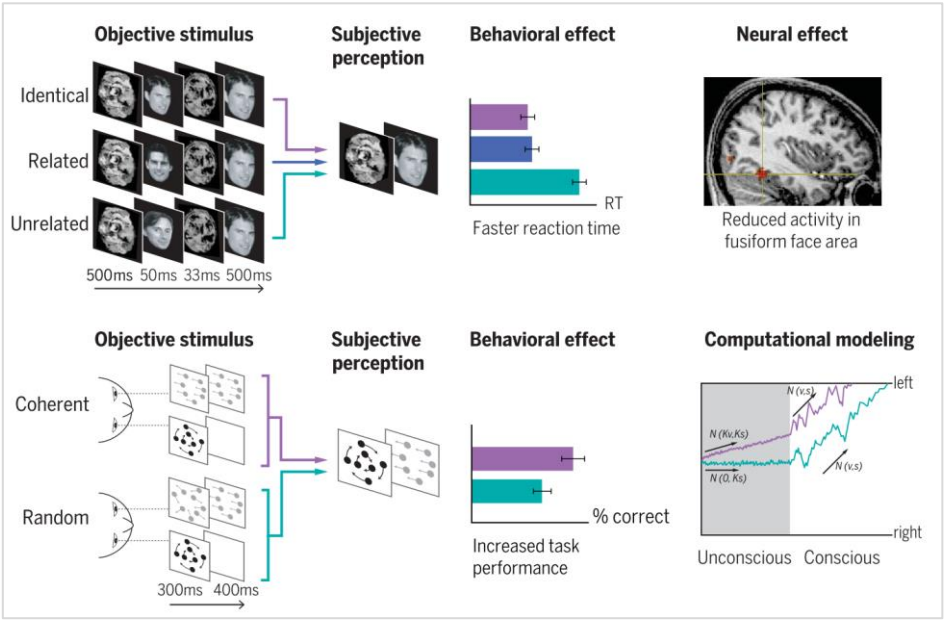


图 1.13 无意识加工的研究示例[21]

事实上，由于人类意识的独占性和顺序性，必然有大量复杂的计算和推断过程需要在无意识层面进行，而这些无意识计算可以在大脑的不同区域异步、并行发生。当前的人工智能已经能够完成大部分 C0 级别的计算过程，例如面孔和客体识别、语言理解等等，甚至在许多方面已经超越了人类的水平。

在 Dehaene 等人看来，意识包含了两种不同类型的信息加工。第一种意义的意识（consciousness in the first sense, C1）称为总体可用性（global availability），主要对应意识的传递意义，即有意识的信息需要进一步的处理

时，不同的大脑功能模块都可以获取。例如，对于“司机意识到燃油指示灯亮起来”这一场景，“燃油指示灯亮起来”这个信息可以被记忆、回想，可以被谈论，还可以用于规划接下来的行动等等。C1 可以看作是一种解决信息共享问题的信息处理架构，无意识计算模块（C0）的信息被整合、筛选，进入意识的全局工作空间，从而可以在不同模块之间进行分享。

除此之外，他们认为还存在第二种意义的意识（consciousness in the second sense, C2），即所谓自我监控（self-monitoring）。如果说 C1 意识反映了其具有访问外部信息的能力，那么 C2 意识则通过其表征自己的能力来体现。具体来说，这是一种能够监控自己的信息加工过程，并获得其状态和信息的能力。这种意识与通常所说的内省（introspection）相对应，即认知神经科学和心理学中的“元认知”（metacognition）。

元认知，即“对认知的认知”（cognition about cognition, knowing about knowing），最初由美国教育心理学家 Flavell 提出，指的是个人对自己认知活动的认识（监控）与调节（控制）过程，而 C2 意识主要指的是元认知监控过程。人类大脑在做出任何决定的时候，都会同时评估该决策的可信度，于是人类对自己的选择或多或少会感到一定的自信。自信程度（confidence）是元认知相关研究中经常采用的所谓第二类任务（type 2 tasks）中需要被试主动汇报的一个行为指标，它可以被定义为对一个决定或计算是否正确的主观概率。类似的概念在学习、记忆等任务中都存在，例如对所学知识的信任程度（judgement of learning, JOL）、记忆是否可靠（feelings of knowing, FOK）等等。对于婴幼儿实验和动物实验，还经常采用决策后赌注（post decision wagering, PDW）等方式测量。

认知神经科学的研究认为，元水平（meta-level）与客体水平（object-level）的认知加工过程是在物理上分离开的，元认知主要与前额叶脑区关系密切，并且具有一定的通用性，不同范畴的客体水平认知加工都可以引发元认知系统的响应，并且元认知系统能够利用统一的方式表征元认知信号，同时还能够进行区分，进行特定的元认知控制。

Dehaene 等人认为，C1 和 C2 这两种意识虽然可能存在交集，但很大程度上

是正交、互补的两个部分。无意识加工的内容（C0）同样可以存在自我监控（C2），同时有意识的内容（C1）也可能没有经过监控和评估，于是被试丢失了信息加工中的概率信息，错误的估计了感知的准确性，甚至感到过于自信（over confident）。Baars 等人认为，在 GWT 框架下，元认知可以分为有意识的元认知和无意识元认知两种，上述情况被称为“吸收”（absorbed experiences），在这种状态下，有意识的元认知（C2）能力被最小化，而无意识元认知可能会继续工作，并有机会在某些情况下上升到意识层面，引燃全局工作空间。LIDA 框架把元认知也作为一个基本模块，从而任何全局广播都有机会触发元认知代理，并启动元认知过程。但也有学者认为，全局工作空间中的表征总是与元认知（自信程度）伴生[22]。

元认知对智能体具有重要的意义。在学习阶段，它能够帮助智能体了解自己的优势和缺点，并灵活安排如何更有针对性的学习或训练；在决策中，元认知能够帮助智能体在信心不足时增加计算资源甚至求助，以免错误的决策造成大的损失；在与其他个体交流时，元认知可以心理理论（Theory of Mind）相互配合，对自己和他人的状态做出评估并决定自己的行为。但如今的大多数机器学习系统都缺乏自我监控，在人工神经网络中，尽管通常都能够给出决策和对应的置信度（conference），但除了贝叶斯网络（Bayesian networks）以外，神经网络的置信度与人类元认知监控中的自信程度并不相同，而且从根本上讲，目前的神经网络尚不具备自省能力。因此，如何在人工智能中引入元认知，是一个值得思考的方向。

除此之外，元学习（meta-learning）也是人工智能中的一个重要的概念，并且近几年也得到了较多的关注[23]。在认知神经科学中，元学习是元认知的一个分支，是指学习如何学习的能力，例如学习调整现有学习算法和策略，以及如何利用现有模型和知识有效地解决新任务。这种元学习能力对于使现有的人工智能系统更具适应性和灵活性以有效地解决新任务非常重要。针对 Dehaene 的意识分级，元学习依赖于 C2 意识，或者本身也是 C2 的组成部分，只有这样，智能体才能监控自己的学习，并对学习中的行动进行控制。

1.2.3 深度学习与全局隐空间理论

前面我们提到过，如今的深度学习已经能够很大程度上解决 C0 级别的认知加工过程，但对于 C1 和 C2 级别的认知，即意识的模拟似乎还有一定的距离。我们有各种不同的神经网络用于处理一些具体的问题，但如何在一个系统中完成多个不同的工作流并灵活的协调它们，仍然是一个难题。DeepMind 曾提出一种称为 PathNet 的网络架构，展现出了强大的、灵活的性能和跨任务模块的泛化能力。Jeff Dean 总结的下一代人工智能框架 Pathways，其基本思路也是需要将单独解决视觉、听觉、语言等不同功能的模块整合起来，构建足够灵活，可以完成多种不同任务的新一代人工智能。

在 GWT 框架下，这个问题看起来已经变得不是那么遥不可及，VanRullen 等人认为采用深度学习的思想实现 GWT 的时机已经成熟，并给出了实现深度学习的全局隐空间（Global Latent Workspace，GLW）的路线图（图 1.14）。首先，GWT 需要若干独立的专用模块，这些模块在深度学习中不难实现，而各个模块都有各自的高阶隐空间（high-level latent space）。在深度学习中，隐空间指的是经过训练，神经网络中形成的用于编码输入空间关键特征的层。神经网络通过它进行高级的概念表征，例如物体的特征、词义、动作序列等等。这些模块可以是经过预训练的网络，用于进行感知、自然语言处理、长期记忆、强化学习、运动控制等等，模块的能力决定了整个全局工作空间系统的能力和能够执行的任务范围。

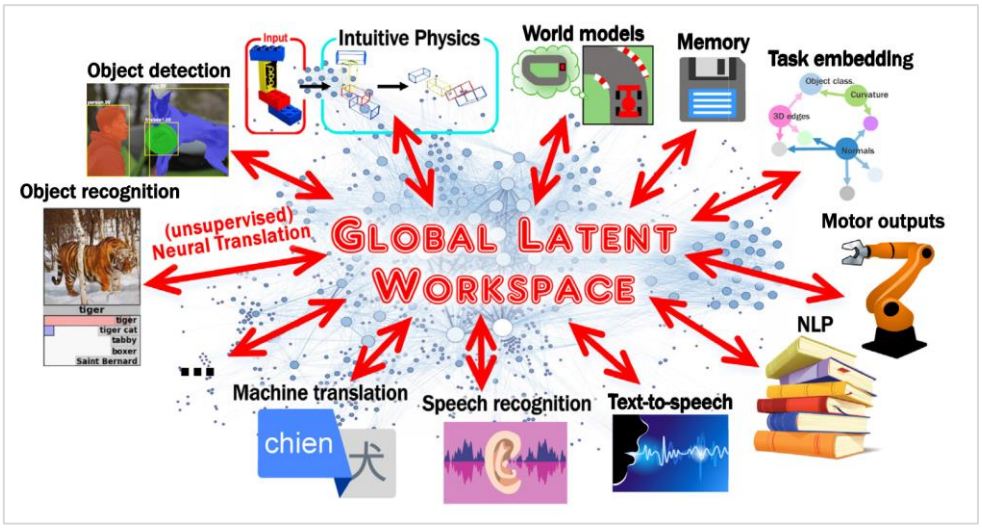


图 1.14 深度学习中的 GLW[12]

其次，还需要一个全局的隐空间（GLW）。GLW 本质上是一个非模态的、独立、共享的隐空间，负责在不同模块对应的隐空间之间进行变换。这个空间的维度应该与其输入的各个隐空间的内在维度（intrinsic dimension）相同或更高，但远低于所有空间维度的和。这个约束可以确保在每个时刻只有相关的信息可以被编码，从而迫使这个系统通过注意力机制在相互竞争的输入中进行选择。空间之间的变换可能需要满足一些近似拓扑变换的约束，以保持其流形（manifold）的关键性质（参考第 2 章内容）。

在大脑中，注意决定了哪些信息被有意识地感知，哪些被忽略；在原始的 GWT 框架中，选择性注意也是信息进入全局工作空间的主要途径。注意同样是深度学习中大家所关注的焦点，例如在自然语言处理和机器视觉中广泛使用的 transformer 架构便基于注意机制，虽然该“注意”与神经科学中的注意并不相同。GLW 也可以借鉴 transformer 的注意机制，通过密钥查询匹配过程（key-query matching process）选择哪些信息作为 GLW 的输入。

当某个模块被注意选择后连接到工作空间，其隐空间中的激活向量便被复制到 GLW 中。这个副本便成为 GLW 与该模块之间双向连接的接口。之后，GLW 中的信息便立即被广播，或曰变换到其他模块的隐空间，于是该信息变得全局可用，尽管接收到信息的模块不见得总会使用这些信息。

GLW 架构符合全局工作空间理论，如果我们将各个模块看作不同的智能体，而 GLW 看作环境，这同时也符合共识主动性的定义。GLW 的整体功能远大于各个模块的功能之和，但通过共用的全局隐空间，它避免了模块之间两两互连造成的复杂结构。VanRullen 等人认为，GLW 将能够显著提高人工智能的性能和鲁棒性，并且也更加容易进行迁移学习。

1.3 总结与展望

在本章中，我们首先介绍了认知科学的具身主义流派，指出基于具身认知思想，在开放环境中进行强化学习的智能体在多任务解决中具有非常高的潜力。接下来具体介绍了全局工作空间理论和它的一些变种、实现以及在深度学习中的作

用。限于篇幅和笔者的水平，以上内容难免挂一漏万，或许还有非常重要的观点并没有被笔者注意到而错过，或许有更好的观点值得探讨。我们希望，我们整理的这些内容能够拓宽大家的思路，甚至对自己的研究有所启发。

我们注意到，人工智能领域也正在经历一次较大的范式转移。其一，各个特定领域的专用算法逐渐逼近性能的极限，许多新算法的性能提升已经不够明显，而预训练大模型如 GPT-3 等开始在应用中占据主导地位，人们倾向于对预训练模型进行微调（Fine-tuning）以适应多种不同的任务；其二，人们开始越来越多的关注如何让人工智能更加通用，能够完成更多不同的任务，适应更多不同的状况；另外，随着具身主义的兴起，人们开始不再局限于大脑本身，而开始考虑环境的反馈和智能体与环境的交互，DeepMind，OpenAI 等业界巨头都基于强化学习取得了前所未有的突破。

随着大模型的能力越来越大，能够执行的任务也越来越多，不少科学家认为，通用人工智能（Artificial General Intelligence, AGI）已经呼之欲出，并认为类脑智能是走向通用人工智能的一个可能的突破口，而强化学习则是通往通用人工智能的道路。

刚刚过去的 2021 年被许多人称为“元宇宙元年”，包括 Meta（原 Facebook）、微软等在内的各大科技巨头纷纷押注，认为元宇宙可能是下一个技术热点，也引得国内大量企业机构效仿。

或许在不久的将来，智能体和人类代理同时在元宇宙内互动，共建元宇宙内容。智能体可以通过网格细胞机制在元宇宙内徜徉，并在无限开放的环境中不断学习、进化；智能体可以拥有视觉、听觉、触觉等不同的感知功能，也拥有视觉、语言、记忆等不同的功能模块，模块之间通过全局工作空间机制自由组合不同模块的功能，从而能够完成各种不同的任务、与人类交流、合作，以至于人类无法分辨其中的个体是否是真人；智能体之间可以像人类社会中一样进行直接的互动，或者通过共识主动性机制在元宇宙中留下自己的印记，同时受到其他智能体和人类的印记所影响，甚至共同完成一些宏大的任务，解决一些仅靠个体的智能无法解决的问题……

真实宇宙中，在无数的巧合中诞生了地球，地球上环境事宜，恰好能够孕育

生命，并最终进化出了迄今为止智能水平最高的人类。在元宇宙中，智能体有机会在更加精心设计的环境中进行更加大胆甚至天马行空的行为和演化。从这个角度来理解，元宇宙的意义便是通过无数人类的建设和互动，为通用人工智能的诞生和进化提供一个足够开放和巨大的环境和生态，为超越人类智能水平的 AI 提供土壤和生存空间。或许，这才是各大科技巨头押宝元宇宙的根本原因。

参考文献

[1] Dawson, M. R. (2013). *Mind, body, world: foundations of cognitive science*. Athabasca University Press.

[2] Cardon, D., Cointet, J. P., & Mazières, A. (2018). Neurons spike back: The invention of inductive machines and the Artificial Intelligence controversy. Réseaux (Vol. 211).

[3] Gupta, A., Savarese, S., Ganguli, S., & Fei-Fei, L. (2021). Embodied Intelligence via Learning and Evolution. *arXiv preprint arXiv:2102.02202*.

[4] Team, O. E. L., Stooke, A., Mahajan, A., Barros, C., Deck, C., Bauer, J., ... & Czarnecki, W. M. (2021). Open-ended learning leads to generally capable agents. *arXiv preprint arXiv:2107.12808*.

[5] Baars, B. J., & Gage, N. M. (2010). *Cognition, brain, and consciousness: Introduction to cognitive neuroscience*. Academic Press.

[6] Baars, B. J. (1993). *A cognitive theory of consciousness*. Cambridge University Press.

[7] Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Progress in brain research*, *150*, 45-53.

[8] Baars, B. J. (1997). In the theatre of consciousness. Global workspace theory, a rigorous scientific theory of consciousness. *Journal of consciousness Studies*, *4*(4), 292-309.

[9] Dehaene, S., Kerszberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the national Academy of Sciences*, *95*(24), 14529-14534.

[10] Dehaene, S., Sergent, C., & Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences*, *100*(14), 8520-8525.

[11] Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, *105*(5), 776-798.

[12] VanRullen, R., & Kanai, R. (2021). Deep learning and the Global Workspace Theory. *Trends in Neurosciences*.

[13] Van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., & Roelfsema, P. R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science*, *360*(6388), 537-542.

[14] Reardon, S. (2017). A giant neuron found wrapped around entire mouse brain. *Nature News*, *543*(7643), 14.

[15] Dehaene, S., & Changeux, J. P. (2005). Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLoS biology*, *3*(5), e141.

[16] Franklin, S., Strain, S., Snaider, J., McCall, R., & Faghihi, U. (2012). Global workspace theory, its LIDA model and the underlying neuroscience. *Biologically*

Inspired Cognitive Architectures, *1*, 32-43.

[17] Snaider, J., McCall, R., & Franklin, S. (2011, August). The LIDA framework as a general tool for AGI. In *International Conference on Artificial General Intelligence* (pp. 133-142). Springer, Berlin, Heidelberg.

[18] Faghihi, U., & Franklin, S. (2012). The LIDA model as a foundational architecture for AGI. In *Theoretical Foundations of Artificial General Intelligence* (pp. 103-121). Atlantis Press, Paris.

[19] Franklin, S., Madl, T., Strain, S., Faghihi, U., Dong, D., Kugele, S., ... & Chen, S. (2016). A LIDA cognitive model tutorial. *Biologically Inspired Cognitive Architectures*, *16*, 105-130.

[20] Blum, M., & Blum, L. (2021). A theoretical computer science perspective on consciousness. *Journal of Artificial Intelligence and Consciousness*, *8*(01), 1-42.

[21] Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it?. *Science*, *358*(6362), 486-492.

[22] Shea, N., & Frith, C. D. (2019). The global workspace needs metacognition. *Trends in cognitive sciences*, *23*(7), 560-571.

[23] Langdon, A., Botvinick, M., Nakahara, H., Tanaka, K., Matsumoto, M., & Kanai, R. (2022). Meta-learning, social cognition and consciousness in brains and machines. *Neural Networks*, *145*, 80-89.

第2章 神经科学进展

解码和理解大脑神经元的功能和机制，是神经科学领域科学家永恒的目标。一直以来，脑科学研究被局限在了不同维度：例如，从认知维度可以将脑科学划分为视觉、记忆、推理、决策等方向，从研究方法维度又有有创（基于动物模型）和无创（基于影像扫描技术）研究的区分。因此，受限于研究方法、专业壁垒和技术瓶颈等因素，脑科学研究往往由独立的科学团队在各自的研究领域向前推进。

然而，近年来多个脑科学领域的发现同时指出了神经元的编码和计算机制具有普适性，例如在记忆研究中发现的神经计算机制，也适用于对视觉、决策等认知机制的解释。这一发现正在促使脑认知研究从神经元的单一信息表征朝通用信息表征过渡、以及从个体神经元表征向群体神经元表征过渡。

本章首先以空间认知领域的发现为例介绍单神经元编码。通过介绍位置细胞（Place cell）和网格细胞（Grid cell）的实验发现，以及人们对其机制理解的演化过程，本章阐释当前最新的神经科学进展。在此基础上，本章从单神经元编码过渡到神经元群体编码，着重介绍神经流形这个当前非常流行的神经元群体编码描述方式。

2.1 单神经元编码与抽象表征

1948 年，Tolman 通过一个行为导航实验提出了认知地图假说[1]。他认为，我们所生活的环境就像是一张张地图，而这些地图以某种形式保存在了我们的大脑中。得益于此，我们可以轻而易举的判断目标方位，即使闭上眼睛，也可以感受到身边物品的位置。30 年后，随着编码自身位置的位置细胞被发现，托尔曼的假设得到了验证，也自此打开了探索神经元编码外部空间机制的大门。

2.1.1 从位置细胞，网格细胞到物理世界的神经编码

空间认知研究的鼻祖可以追溯于 Tolman 在 1948 年的一项空间行为实验[1]。在他的实验中，大鼠被训练在迷宫中觅食（图 2.1，左），当大鼠熟悉了从起点 A 到终点 G 的路径后，Tolman 更换了迷宫，新的迷宫除了保留起点与终点位置不变以外，包含了多个可以抵达终点的路径（图 2.1，中）。实验结果显示，即使终点不可见，大鼠在新迷宫中主动选择距离终点最短的路径（“6”号路径）抵达终点（图 2.1，右）。说明在起点和终点位置明确的情况下，大脑可以利用这些信息构建出一个认知地图为大鼠的运动进行导航。

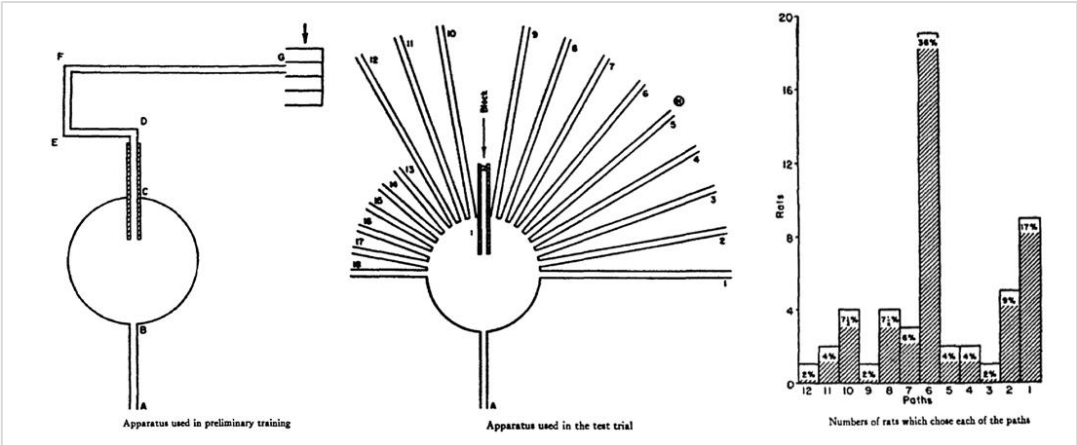


图 2.1 左：实行导航任务用的迷宫环境，大鼠被训练从 A 点到 G 点获取食物奖励；中：变化后的迷宫环境，但起点和终点位置不发生变化；右）行为结果显示，即使目标位置不可见，绝大多数大鼠选择了通往奖励位置的正确路径[1]。

直到1971年,O’Keefe和Dostrovsky发现了认知地图存在的证据(图2.2),即存在于大脑海马体内的位置细胞（Place cell）[2]。每一个位置细胞编码空间环境中的特定位置，换言之，大鼠在导航过程中，经过空间中的每个位置时，都会有特定的位置细胞放电表征该位置。Bostock 等人[3]进一步测试了同一只大鼠依次在两个颜色不同但形状相同的环境中执行导航任务,发现位置细胞可以对两个环境的相同位置表现选择性放电（remapping），这个发现进一步确定位置细胞的放电表征的是位置信息而非纹理信息。除了自身位置信息以外，后续的研究又分别在 1984 年和 2008 年在内嗅皮层发现了编码自身方向的方向细胞（direction cell）和编码环境边界位置的边界细胞（border cell）等一系列细胞类型。基于这些发现，人们意识到大脑对物理世界的编码是依托于大脑内这些表达特定信息的细胞而实现的。然而,相比于这些编码明确的细胞类型以外(例如，位置、方向、边界），网格细胞（Grid cell）的发现又重新给空间认知的神经机制带来了迷雾。网格细胞于2005年由Mosor夫妇在大鼠脑内的内嗅皮层被发现[4]，和位置细胞不同的是，网格细胞拥有多个放电场（firing field）（图2.2），这些放电场布满空间且呈现规则的网格状分布。有研究指出，这些网格的形态并不受大鼠的运动速度和方向影响，而仅由环境本身决定。这表示网格细胞的网格结构与环境中的距离和方向有关,因而网格细胞对空间的表征具有重要作用。

但是，这种规则结构是如何形成的、以及它的神经机制和功能具体是什么当前均还没有定论。主流的假设是网格细胞的规则放电形态是神经震荡的干涉结果，而反对的声音认为，如果网格细胞放电场由神经震荡形成，那么观察到的空间中位置的活动表征应该是周期性的，而不应该是单一放电场的形态。而后的几年，空间编码的神经机制研究进展放缓，直到Constantinescu等人在2016年，才揭示了网格细胞的普适性特点[5]，作为里程碑式的发现，神经元的普适性引发了人们对神经元工作机制的重新思考与讨论。

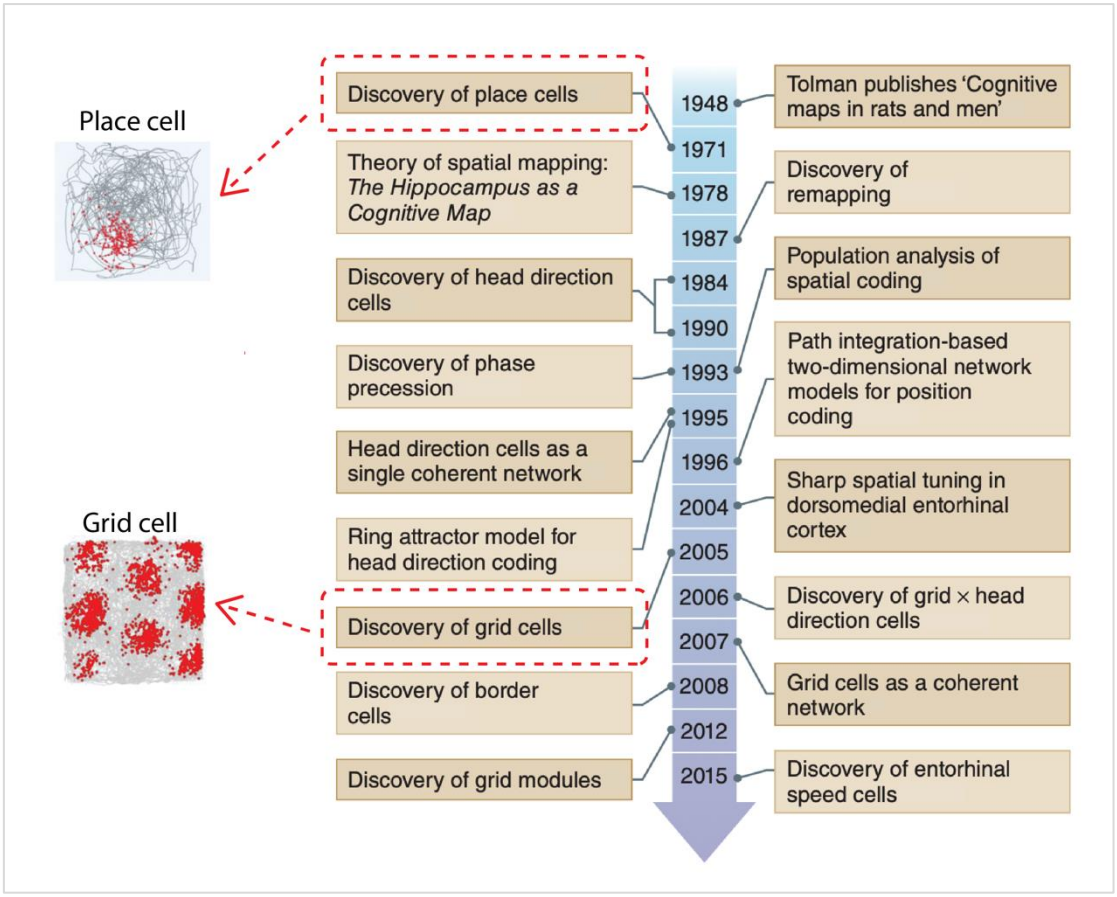


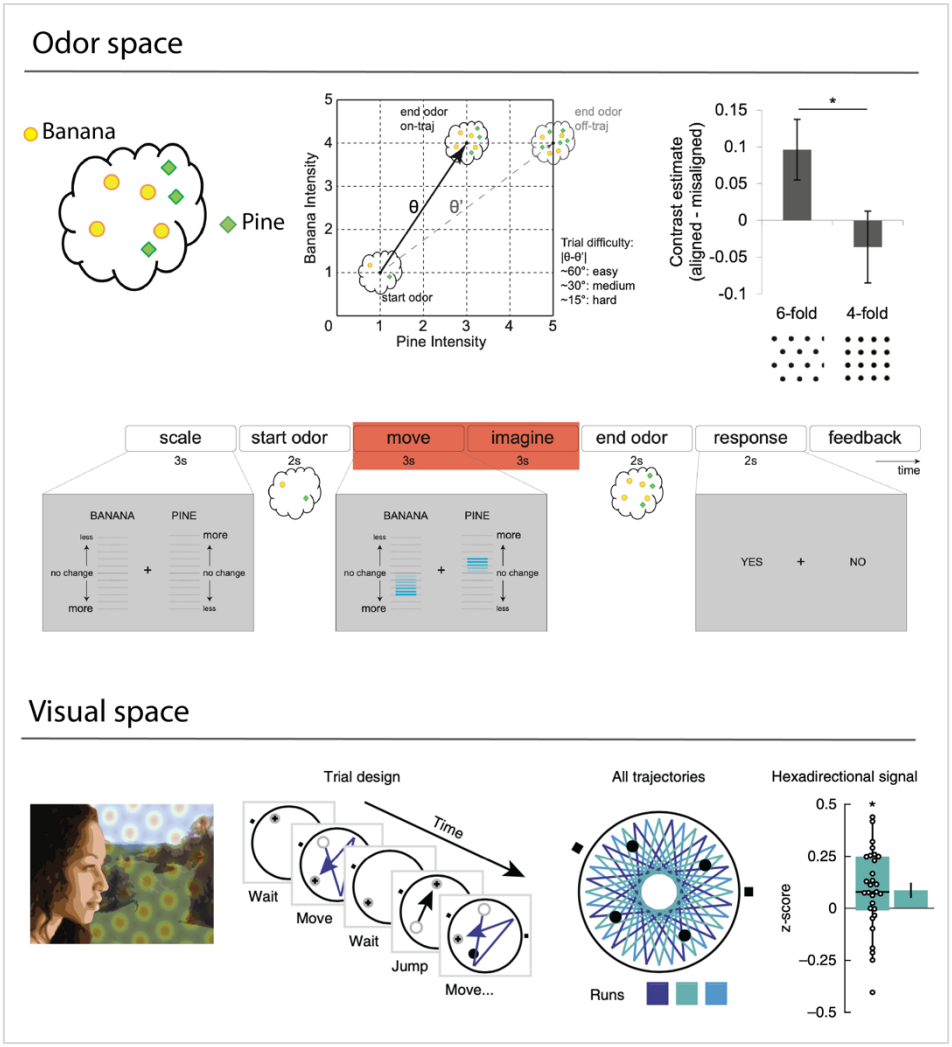
图 2.2 以空间认知领域为例阐述神经元信息编码的历史历程[4]。自 Tolman 于 1948 年建立认知地图假说以来，过去 50 年间人们在大脑的内侧颞叶中发现了编码物理世界不同信息的神经元，海马的位置细胞（Place cell） 与内嗅皮层的网格细胞（Grid cell） 是典型。左侧是大鼠的位置与网格细胞在方形空间下的放电示例）。

2.1.2 从物理空间到抽象空间的神经编码

正当人们针对网格细胞的形成机制争论的喋喋不休的时候，新的实验证据显示，在负责空间信息编码的内侧颞叶以外的脑区，也发现了网格细胞存在的证据。2010 年，Doeller 等人借助 MRI 方法成功在健康人类被试的脑中检测到了网格细胞的活动信号[6]。不过出乎意料的是，网格细胞活动不仅出现在了内侧颞叶的内嗅皮层，而且是有选择的分布于全脑，这包括内侧前额叶、后扣带回皮质、楔前叶、外侧颞叶皮层等脑区，即传统上广泛被认知神经科学家所熟知的脑默认网络（default mode network）。默认网络是一个脑功能网络，其激活常见于各类人类认知功能研究的实验范式，包括长短期记忆、想象、场景重建、价值评估与决策等。既然网格细胞存在于默认网络，且默认网络又参与编码如此广泛的认知

功能，那么是否有可能网格细胞也参与编码空间以外的信息？带着这个问题，2016 年 Constantinescu 等人率先调查了网格细胞是否对编码概念信息具有选择性。实验中，通过要求人类被试调节卡通鸟的脖子和腿长度这一简单的任务，Constantinescu 等人观察到当人类被试在由鸟的脖子和腿的长度构成的二维概念空间中“运动”时，大脑的内侧颞叶以及默认网络被激活，意味着网格细胞参与了这一与空间任务无关的认知行为。同时证明了网格细胞不仅编码空间信息（例如位置、方向、边界等），也编码非空间信息的假设。 Constantinescu 的实验似乎从侧面暗示着我们，我们从未真正理解神经元的工作机制。

图 2.3 网格细胞编码非物理空间（视觉空间[7]和嗅觉空间[8]）的新证据。



自 Constantinescu 等人的 2016 年发现后，截止到目前，人们正在对网格细胞的抽象信息编码机制进行广度这个层面上的进一步探索，既然由卡通动物的脖

子长度和腿长度构成的抽象结构可以诱发网格细胞活动，那么任何信息之间关系所构成的我们所理解的“空间”都应该诱发相应的神经活动。例如，Bao 等人构建了基于嗅觉的“二维空间”（图 2.3），通过调控每个维度（一种味道）的浓度成功诱发了网格细胞活动[8]。特别需要强调的是，相比较于上文所提到的记忆和嗅觉维度，近期的一项研究指出了网格细胞参与视觉信号编码的证据，Nau 等人在基于视觉的任务中[7]，要求被试用眼睛注视屏幕上移动的圆点，随着屏幕上圆点在屏幕上划过呈特定角度的直线并重复多次后，结果显示内嗅皮层被眼球运动这个行为成功激活（图 2.3）。相比于需要大脑腹侧脑区编码的客体与嗅觉空间，Nau 等人的实验中，人类被试除了眼球运动以外并不需要编码额外信息，该结果说明，仅仅是由大脑背侧编码的眼球运动也可以诱发网格细胞的活动，说明其实际参与了我们用眼睛观察外部环境的认知过程。以上这些实验结果表明，虽然网格细胞的信息编码机制目前仍还并不完全清楚，但已有的这些证据显示，网格细胞对我们的视觉、嗅觉、记忆等多维度的认知过程都有所涉及，展现了其具有超越单一维度的通用属性。

网格细胞参与编码多种认知能力的现象，从生物进化的角度来讲，符合“Neuronal Recycling”假设[9]，即网格细胞在生物进化的初期可能就已经存在了，例如仅编码简单的认知功能，随着生物的认知功能进化，才渐渐参与对更全面的认知功能的编码。一方面，支撑这个假设的证据是网格细胞已经在多个生物物种的大脑中被发现，包括大鼠、蝙蝠、猴子、以及人类；另一方面，近期的人工智能研究指出，在人工递归神经网络中，也观察到了网格细胞的自涌现现象。Banino 等人通过长短期记忆网络（LSTM）训练智能体完成空间导航任务，网络的隐藏层中，有 25.2%的人工神经元表现出了类似网格细胞的放电形态[10]。 这些现象进一步印证了网格细胞对于认知与智能来说，具有普适性特点。

综上所述，本小节概述了自网格细胞于 2005 年被发现以来，人们对神经元研究的最新进展与理解，即我们大脑神经元的机制不仅局限于单一维度，且展现了较强的可泛化性与通用性。涉及到其通用性的特征可概括如下，第一，网格细胞被发现参与编码了和物理空间无关的抽象信息，支持了大脑创建并利用多客体空间协调物体间关系的假说。第二，即使排除物理空间和抽象空间等信息的影响

后，网格细胞仍可通过眼球运动被激活，证明其参与了大脑对视觉信息的表征。第三，在排除了基于生物体这一媒介后，网格细胞仍在深度神经网络的隐藏层中被发现，说明神经元对信息的编码在自然界中存在相似性，即基于生物的神经元和基于机器的神经元之间存在基于自然的通用性。因此，我们期待未来的研究继续对神经元的普适性表征机制进行探索，它将对解码人脑智能的本质，以及对通用智能的实现具有重要意义。

2.2 神经元群体编码：神经流形

神经科学家发现单个神经细胞所编码的信息比之前认为的更为复杂。比如，Nieh 等人发现海马体中位置细胞在参与位置编码的同时还参与了对决策证据的编码[11]；Elsayed 等人发现相同的运动神经元在运动准备阶段和运动执行阶段都有发放[12]。这些新发现难以被单神经元编码理论解释，越来越依赖对神经元群体编码的理论研究。人们询问着一个简单的问题，如何开发出有效的方法来研究神经元群体编码。

近年来，随着电极阵列等大规模神经元记录技术的发展，我们可以同时记录上万神经元的活动，从神经元群编码的角度理解神经元群的活动愈显重要和迫切[13]。因为某一种神经表征有来自成千上万的神经元活动的共同贡献，所以我们需要使用更简洁的方式来分析这些复杂信号。在本节，我们介绍一个当前非常流行的神经元群体编码分析视角：神经流形(neural manifold)。本节内容将介绍什么是流形，有关流形的实验发现，流形如何解释神经元编码，以及如何在研究分析中使用流形。

2.2.1 什么是神经流形

流形作为一个几何概念在 19 世纪 40 年代被黎曼等人提出[14]。现代数学中对流形的研究主要集中在两个分支：拓扑学和微分几何。在拓扑学中，流形被定义为一个局部近似于欧氏空间的拓扑空间，其结构的连接属性是主要的被研究对象。在微分几何中，拓扑空间的基础上流形的定义增加了平滑属性从而要求其可微。拥有平滑属性的流形被称为黎曼流形。虽然流形作为一个整体可能有复杂的

形状，但是其中的每一个点附近都可以近似看成一个相同维度的欧式空间，而这些局部欧氏空间不一定共享一个全局坐标系。简而言之，流形可以被想象成若干个局部欧式空间覆盖在相同维度的全局空间上。

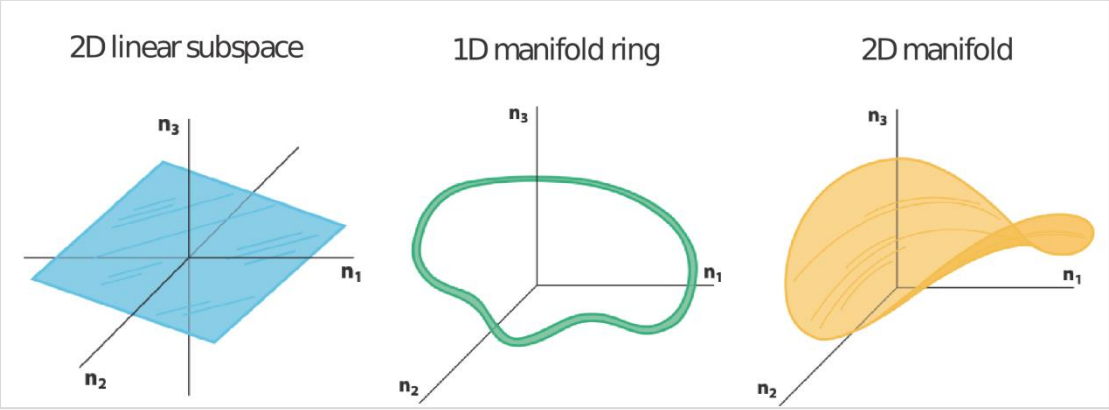
将流形的概念引入到神经科学的文献最早可以追述到 1947 年[15]。这里流形作为一个名词仅仅被用来指代褶皱的大脑皮层。系统地将流形作为数学工具来分析神经网络的是 S.Amari。在 1990—2000 Amari 提出了信息几何,用流形的概念描述神经网络的训练[16]。信息几何中的参数空间是一个高维空间，其每一个轴代表一个连接权重，表征神经网络的流形处在这个参数空间内。近二十年来流形越来越多地被用来描述大脑中神经元的群体活动[17]，因此我们可以称之为神经流形。

为了能够理解什么是神经流形，我们需要从几何视角描述一群神经元的活动状态。 Seung 和 Lee 在 2000 年提出可以把神经元的群体活动状态描述成一个向量，该向量中的每一个分量代表一个神经元的活动[18]，我们可以想象一个高维坐标系，其中一个向量代表一个点。这个高维的状态空间中每一个轴代表一个神经元的放电活动。如果有N个神经元参与该刺激的编码，那对应的坐标系代表了一个N维的神经活动空间(neural response space)。近年来，神经科学家提出了一个通用的流形假说：大脑将外部信息编码为高维神经活动空间内的低维流形。

神经元群体编码受到两方面的约束：一方面是来自外部刺激；另一方面来自神经元连接结构。对于给定的外部刺激，神经元群体编码的维度低于神经活动空间的维度。假设外部刺激处于某一状态，并且被观测的所有神经元的放电活动达到稳定状态，那么该刺激状态近似地对应于某一个神经编码向量，即神经活动空间内的某一个点。当外部刺激从一种状态变化到另一种状态的时候，神经活动空间内的点也从一个位置变化到另一个位置。同一任务中刺激状态的变化远比高维空间中的点的位置变化受到更多的约束，比如在图像识别任务中，如果同一图像绕中心旋转，该刺激改变的状态只是一个旋转角，与之对应的应该是高维神经活动空间内点的位置沿一条 1 维曲线的变化[18]。该类刺激在神经活动空间中的编码是一条 1 维曲线。以此类推，如果同类外部刺激有两个种不同的状态改变，那

么该类刺激的编码是一个 2 维曲面，如图 2.4 中示例，如果外部刺激有更多种状态改变，比如m种，那么该刺激的编码是一个m维“曲面”，或更准确地称为m维流形。假设外界任务刺激的变量数m远少于大脑神经元的数量N，那么同一类外部刺激的神经编码是嵌在高维神经活动空间的一个低维流形。

图 2.4 三维空间中的 1 维和 2 维流形示例[19]



神经元群体编码的维度同时也受限于大脑内部的连接结构。例如，当神经流形编码运动[20]和抽象知识[11]时，虽然这些表征不直接依赖外部刺激，但其神经计算依赖于特定神经元的相互连接来实现，因此产生的群体活动受到神经网络中功能与结构连接的限制[21]。这样的约束也可能导致神经元群体活动被限制在一个低维流形上。

2.2.2 有关神经流形的实验发现

既然外部刺激和神经元连接结构是群体神经元放电的两个约束条件，那么通过对外部刺激以及神经元之间的通讯进行干预和调控，利用这种方式有利于我们进一步理解神经流形。近期的神经科学发现支持了这个观点。

关于外部刺激的调控对流形的影响，以我们上一章节介绍的网格细胞为例，我们知道网格细胞可以表征空间环境，且网格细胞相对应的流形概念也早在 2007 年就被提出[22]，如图 2.5a-c 所示，一个基于物理空间的神经群体编码在水平和垂直两端进行边缘连接后，即形成了表征这个简单的二维矩形环境的流形。但是，现实环境往往比实验室条件下构建的环境更复杂，例如环境往往需要有边界的清晰定义，那么通过调控环境让其发生变化，网格细胞的流形会受到什么样

的影响呢？Derdikman 等人提出了空间是以类似于马赛克的形式被神经元编码的假说[23],即每一个子空间(马赛克)作为一个环境单位被相应的流形所表征。为了验证这个假说，Derdikman 先训练大鼠熟悉一个方形的环境，神经元的网格状放电形态随即出现（图 2.5d），随后在环境中摆入障碍物，使原先的环境被分割成若干区域后，结果显示神经元的放电形态并没有维持原状（图 2.5e）,而是随环境变化形成了特定区域独特的编码（图 2.5f）。该实验结果表明 Derdikman 的假说是正确的，即神经元对复杂环境的每个子空间进行独立编码(图 2.5g)，并由相应的流形所表征。同时，流形具有可塑性，如果环境发生变化，相应的神经流形也随即变化。

关于对神经元连接结构的调控，根据上文介绍的神经流形定义，一个神经流形可以实现对我们一个具体认知能力的编码（例如移动眼球并观察特定区域），那么通过对构成流形的神经元信号进行人为调控，我们的认知能力会发生什么变化呢？带着这个问题，Sadtler 等人发表于 2014 年的工作指出，使用脑机接口技术，猕猴被训练移动屏幕中央圆点到周围的 8 个指定位置（图 2.5）[24]。当编码 8 个不同方向运动的神经元以及对应的神经流形被找到后，Sadtler 和同事通过对神经信号进行干扰，构造了流形内(within manifold)和流形外(outside manifold)干扰两个实验条件以检测神经放电与流形的几何位置关系对运动学习能力的影响。结果显示，“流形内”条件下，猕猴可以快速地学会并完成任务，而“流形外”条件下猕猴却无法完成任务。Sadtler 等人的研究结果显示，由神经元连接结构决定的神经流形形状对我们的认知能力具有决定性影响，并确实起到了束缚和限制的作用。这也为我们生活中常遇到的个体差异现象提供了一个可能的解释，例如有的人专长于音乐创作但不善于数学推理，可能就是受流形的形状影响。因此，如果流形是智能实现所不可或缺的元素，那么对流形形状和对其进一步的神经机制研究，将成为未来一个极具价值的待解决问题。

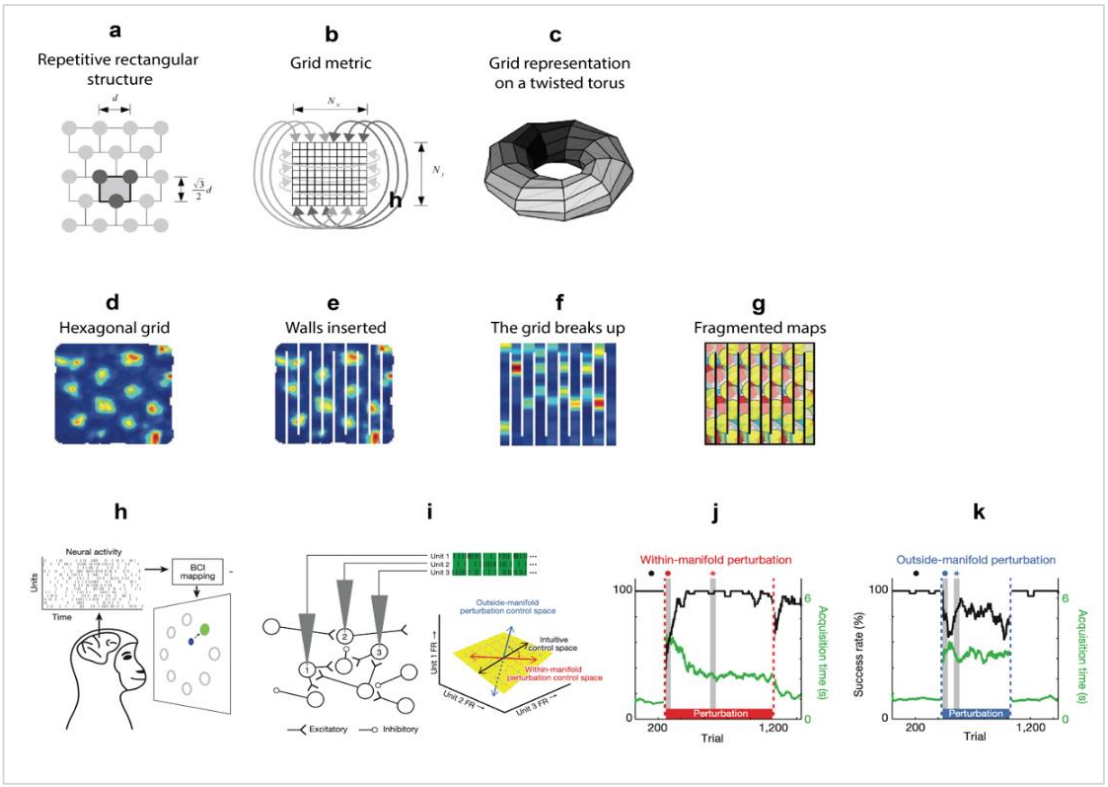


图 2.5 a-c,网格细胞流形表征的概念模型[22];d-g，网格细胞的规则放电形态随空间形态变化而变化[23]；h-k，流形在大脑内存在的神经科学证据[24]。

2.2.3 流形的维度

流形作为一个几何概念有一些固有的几何特征，比如维度、拓扑结构、大小、曲率等等。其中维度作为流形的最重要和最直接的特征吸引很多研究人员的兴趣。Sereno 和 Lehky 研究了物理空间感知任务中的神经流形的维度[25]。因为物理空间是 3 维，他们假设神经活动空间内对应的神经流形也应该是 3 维的。为了验证这个假设，他们让猕猴观察屏幕上不同位置的刺激，并采集猴子 80 个前下颞叶皮层(AIT)神经元和 73 个侧内顶叶(LIP)皮层神经元，构成了 80 维 AIT 空间和 73 维 LIP 空间。每一个位置刺激对应于神经活动空间内一个点。然后他们使用多维标度(MDS)的方法将高维空间降维成三维空间。他们发现在校正后刺激点在三维空间内的相对位置类似于这些刺激在原本物理空间内的相对位置。因此，他们认为在视觉皮层对物理空间的表征只需要 3 个维度。为了进一步验证在更抽象的客体识别任务中神经元群体编码用较低的维度表示，Lehky 和同事让猕猴观看 806 幅不同类别的照片并记录了 AIT 脑区 674 个神经元的电信号。他们用主成分

分析(PCA)对 674 维的神经活动空间进行降维，发现所有图像刺激在神经活动空间对应的点可以被降维到约 63 维的空间中[26]。

在以上的工作中，研究人员使用了线性降维的方法，因此降维后的空间仍为欧氏空间。然而，线性降维的方法无法明确降维后的维度与神经元编码所需要的变量数之间的关系。为了便于进行系统分析，Jazayeri 和 Ostojic 给出了一套完整的层级关系描述[21]。他们将包含了所有可能状态的神经活动空间的维度，也就是可观测神经元的个数，称为周遭维度(ambient dimension)。通过主成分分析等线性降维方法，神经元活动可以被降维到某个低维欧氏空间。这个欧氏空间的维度被称为嵌入维度(embedding dimension)。实际操作中，嵌入空间内的神经活动通常需要能解释 80%的神经活动。然而，嵌入维度也不代表编码所需要的最少独立变量数目。比如如果对应于外部刺激的流形是一个弯曲的 1 维曲线，那么这条曲线需要嵌在一个 3 维欧氏空间内。这里独立变量的数目是 1 而不是 3。编码所需要的最少独立变量数被定义为嵌入空间内神经流形的维度，也称内在维度(intrinsic dimension)。

神经流形往往是弯曲的，内在维度的计算因而需要更复杂的方法。确实，Chaudhuri 和同事发现小鼠对头部朝向角度的编码就是几千维的神经活动空间中的一维环[27]。他们分别在小鼠清醒和睡眠的时候记录了前背侧丘脑核(ADn)的神经信号。Chaudhuri 和同事发现很多线性降维的方法并不能找到复杂流形的真正结构和维度。为此，Chaudhuri 和同事们开发了一个复杂的降维方法。其中最重要的一步是用持续同调(persistent homology)的方法确定流形的拓扑结构。简单而言，拓扑结构可以根据空间内的实体的个数和实体上“洞”的个数来区分，如图 2.6a，一条直线有一个实体没有洞。任意一条曲线也只有一个实体没有洞，因此任意一条曲线跟一条直线是拓扑同构的；然而，一个环有一个实体和一个洞，因此环和一条曲线拓扑不同构。持续同调方法如图 2.6b 所示，以神经活动空间内的每一点为中心以相同半径确定一个球体。随着半径的增大，球体会相互连接形成一个新的实体，构成新的拓扑结构，比如图 2.6b 中的红色，青色，蓝色，和黄色的环代表实体上的洞。黄色的环持续的时间最长，该拓扑结构也最持续，

因此该拓扑结构被确定为流形的拓扑结构。基于这个拓扑结构，我们可以使用相关维度等方法来估计流形的维度。

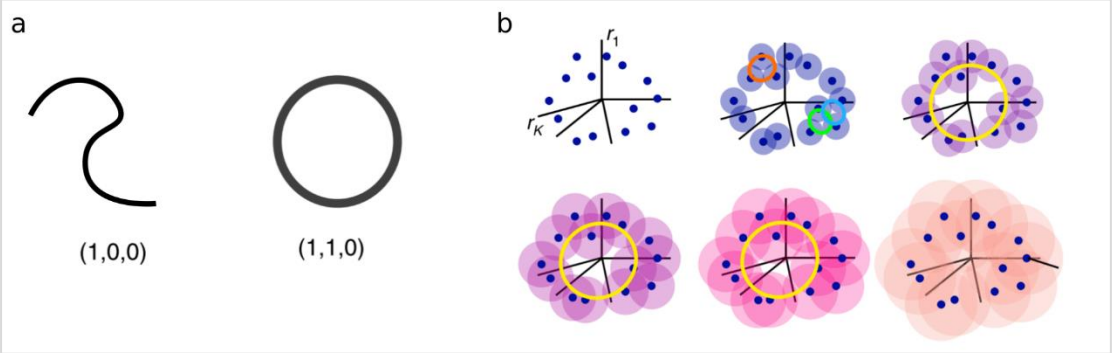


图 2.6 a)曲线和环是不同的拓扑结构；b）用持续同调方法确定拓扑结构[27]

流形的内在维度反映了刺激、运动等表征的最少独立变量数，而这些变量的信息通常在流形的嵌入空间中被处理和传递[21]。关于大脑信息处理的一个经典假设是，不同脑区之间，以及大脑和周围神经系统之间，经常通过线性解码进行交流，而嵌入空间的组织形式可能有利于这种线性解码。

2.2.4 流形与线性解码的关系

嵌入空间形成了神经活动空间的子空间，而子空间就提供了一种线性的方式来实现复杂计算。Elsayed 和同事们在研究猕猴动作任务中发现当两个子空间相互垂直的时候，同一群神经元所进行的不同计算可以被有效隔离[12]。他们让猕猴观察屏幕上不同位置的一个目标点，但是必须在收到开始命令后猴子才能去触摸该目标点。因此这个任务分为两个阶段：开始命令前的准备阶段和开始命令后的运动阶段。他们记录了初级运动皮层（M1）和背侧前运动皮层（PMd）中的 127 个神经元的信号。因为准备阶段的神经活动和运动阶段的神经活动是相关的，所以他们关心的问题是为什么准备阶段的神经活动不会立刻触发猴子的肌肉响应。他们用主成分分析法将准备阶段和运动阶段的神经元信号分别从 127 维空间降为两个对应的 10 维子空间。他们发现这两个子空间非常接近正交（图 2.7）。所

以，通过用线性变换的方式让执行不同计算的子空间相互正交，同一群神经元可以互不干扰地表征多种信息，从而同一个神经元可以同时参与不同的表征。

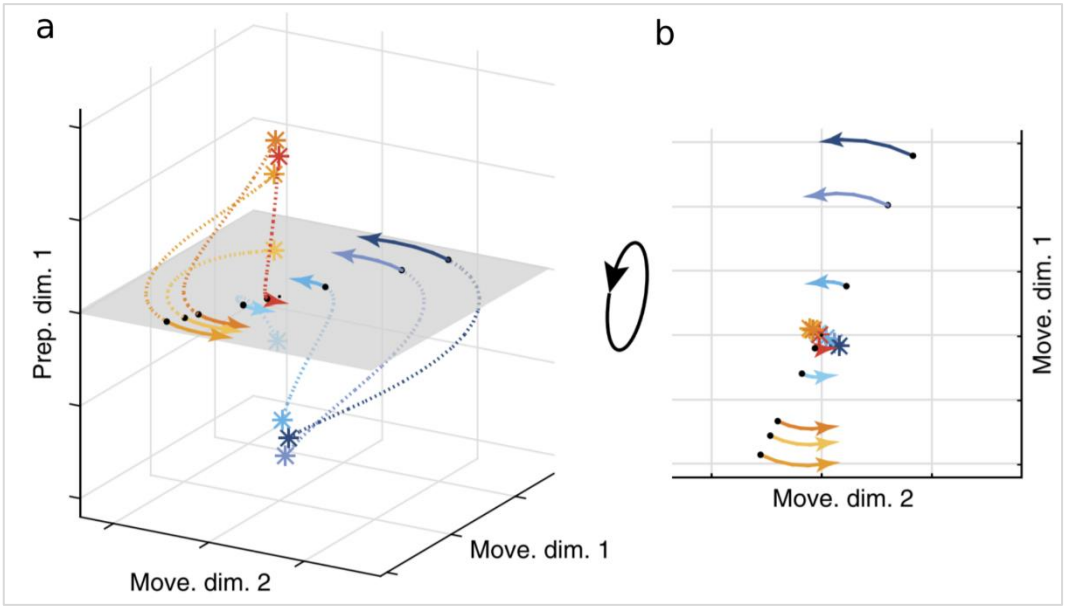


图 2.7 a）相互垂直子空间的 3 维投影：准备子空间投影为竖直轴，运动子空间投影为水平面；b）该 3 维投影的俯视图[12]

基于大脑线性解码假说和流形假说，研究人员能够进一步分析神经网络是如何在嵌入空间内处理信息。比如在物体识别任务中，外部世界的同一类刺激对应于神经活动空间内的一个流形，大脑的识别功能可以被认为是通过对不同的流形在嵌入空间内的分割来实现（图 2.8）。可以想象如果某个流形跟其他流形重叠在一起，那么它跟其他流形就无法被有效区分。不同流形被区分的能力跟流形的几何形状息息相关。为了量化不同流形被区分的能力，Cohen 和同事基于统计力学发明了分类容量(classification capacity)这一量化指标，并探索了人工深度神经网络中分类容量跟神经流形的维度、半径、和中心位置之间的关联[28]。这里人工神经网络所有计算单元的活动强度构成一个神经活动空间。其中，每一层的计算单元的活动强度构成一个嵌入空间。嵌入空间内的对应于不同类刺激的神经流形拥有如图 2.8 所示的中心位置和边界锚点。在图 2.8a 和 b 中，区分不同的流形会依赖于不同的锚点。同一个流形中锚点的分布确定了该流形的半径。Cohen 和同事发现在完成训练的神经网络中，越接近输出的层中的神经流形维度和半径越低，不同流形的中心相关性越低，而分类容量越高，因此神经流形的几何特征决定了分类容量。在这些几何特征中，流形维度对分类容量的贡献最大。

基于此结果，他们推测大脑中的信息处理很可能也是通过更改流形的几何特征使其更便于线性分割。

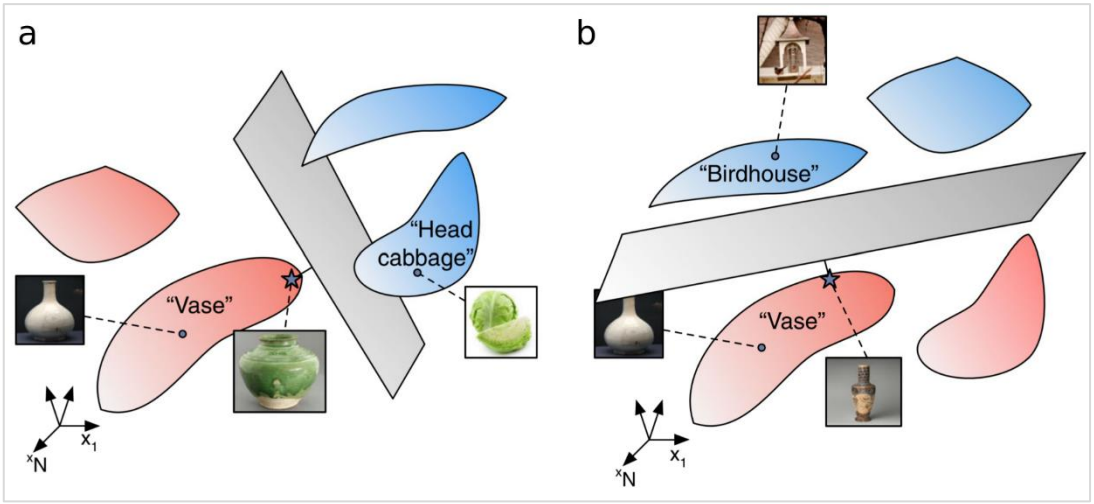


图 2.8 线性分类花瓶流形和白菜流形(a)与花瓶和鸟屋(b)使用了不同的锚点（星号）[28]

类似这样的线性解码不光可以被用来实现分类功能，它也很可能是大脑预测功能的实现方式。人和动物的很多行为依赖于大脑依据当前的视觉感知信息预测下一刻的信息。然而外部视觉刺激的变化往往是非线性的[29]。如图 2.9a,视觉场景随时间的展开可以描述成像素空间内一条轨迹。这条轨迹可以看成一个曲率非常高的 1 维流形。Henaff 和同事认为经过视觉通路的处理，如果对应于外部刺激的 1 维流形的曲率变得更加接近直线，那么预测下一个点的位置就会变得更加简单。基于这样的假设，Henaff 和同事们让猕猴观测不同的视频片段，并记录初级视觉皮层(V1)几十个神经元的信号[30]。他们发现在神经活动空间内的 1 维流形的曲率相对于像素空间变小(图 2.9b)。作为对照，他们人为编辑了一些在像素空间曲率很小但是不自然的视频片段。这些非自然的片段对应的神经流形曲率很大（图 2.9c）。 基于这样的结果，他们推测大脑视觉通路的作用是将神经活动空间内的时间轨迹矫直从而进行线性预测。

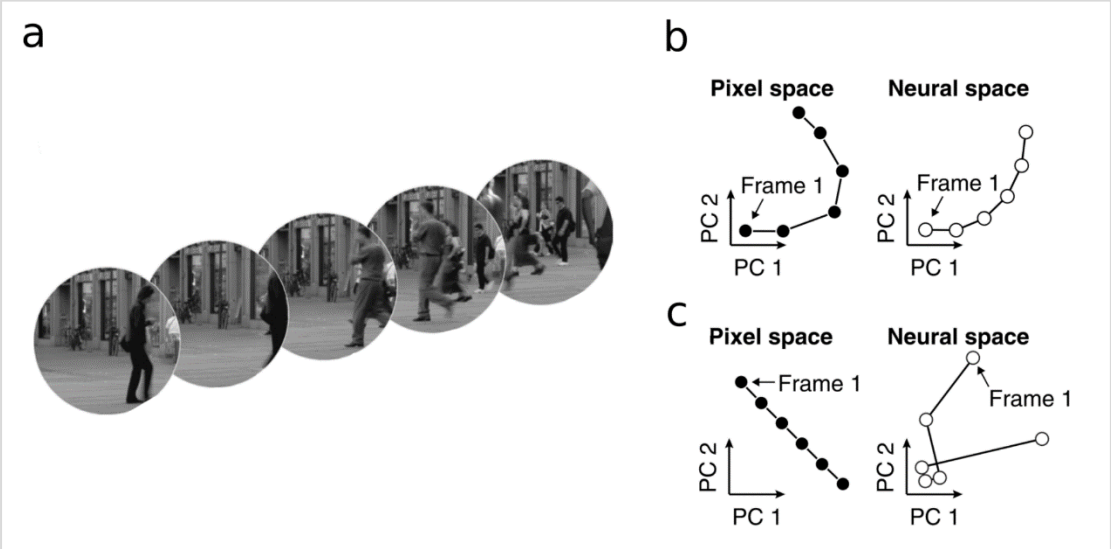


图 2.9 a）自然视觉场景的展开[29]；b）随机视频对应像素空间和神经活动空间内的 1 维流形主成分分析；c）非自然视频在像素空间和神经活动空间内的 1 维流形主成分分析[30]

神经活动空间内的轨迹是一个动态过程。这个过程可以看成是点在流形上运动。流形的形状限制了这个点的独特动力学属性。在上文提到的研究中，Chaudhuri 和同事也发现表征小鼠头部转动的 1 维流形是稳定的连续吸引子：流形上的点处于稳定态从而不会在流形上运动，但是流形外的其他点会向流形移动[27]。所以，流形的形状对神经活动空间内的点的运动有束缚作用。

2.2.5 流形上的动力学

为了理解流形的几何形状如何限制流形上点的运动，我们首先需要介绍一下空间内点的动态过程是什么。点的动态过程是指一个点处在空间内某一位置时，它向下一位置运动的速度大小和方向是确定的，可以表示为 $\dot{v} = f(v)$，这里，向量$\dot{v}$代表点的速度，向量v代表点的位置。函数$f(\cdot)$为一个位置确定一个对应的速度。所以，当一个点处在空间内某一个初始位置时，它会沿着该位置的速度方向运动到下一个位置，再沿着下一个位置的速度运动到下下个位置，依此类推，这个点的运动轨迹就被确定下来。因此$f(\cdot)$提供了点的动力学的完整形式化表达。

我们可以从几何视角来看$f(\cdot)$如何将动力学和流形联系在一起。函数$f(\cdot)$为空间内的每一个位置确定了一个速度向量，形成一个速度向量场，如图 2.10a

上图所示。这个速度向量场以几何方式保存了$f(\cdot)$的所有信息，因此向量场的几何表达和$f(\cdot)$的形式表达是等价的。又因为神经活动空间内的点的运动被约束在神经流形上，所以我们可以用流形上的速度向量场来等价描述点的动力学（图 2.10a 下图）。值得注意的是通常流形是一个弯曲空间，因此速度向量不再存在于流形之内，而是在流形不同的位置与流形相切。这些切向向量表达在流形所处的嵌入空间内，因此他们与嵌入空间有着相同的维度。当然，这些切向向量同时也可以表达在更高维度的神经活动空间内，但是嵌入空间的使用使我们避免处理不必要的高维信息，从而使我们对速度向量场的分析变得简单。

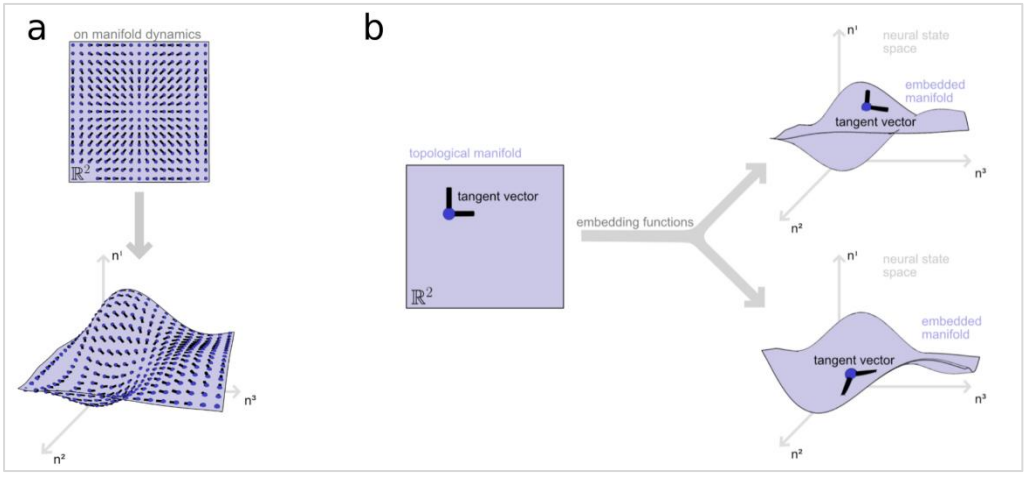


图 2.10 a）2 维向量场和 2 维流形向量场；
b）相同的向量场可以通过不同的嵌入方式形成不同流形向量场[31]

Claudi 和 Branco 发现流形的拓扑特性可以被用来简化对速度向量场的分析[31]。他们对流形向量场的描述分成两步：1）在一个简单的空间内——比如欧氏空间——描述向量场；2）将该简单空间通过拓扑不变的方式置于到嵌入空间中的流形内（图 2.10a）形成复杂流形向量场。这样做的好处在于同一个简单向量场可以通过不同的嵌入方式构成不同的流形向量场（图 2.10b），因而不需要对各种各样不同几何形状的流形单独定义向量场。因为嵌入空间是由神经活动空间线性降维而来，因此它们之间存在着线性变换关系。通过这个关系，嵌入空间内定义的低维向量场可以线性变换为神经活动空间内的高维向量场。这样，我们可以从简单的低维向量场，通过嵌入变换和线性变换得到复杂的神经活动空间内的高维向量场，表示为$\{(v_i, \dot{v}_i)\}$，　$i = 1, 2, \ldots$。

在知道神经活动空间内点的动态过程如何被约束在流形上之后，我们仍然关心这样的动态过程是如何被神经网络所实现，从而保证该动态过程的生物可实现性。很多的研究都指出循环神经网络（RNN）可以作为验证神经动力学的重要测试平台[17, 12, 21]。

2.2.6 流形向量场和循环神经网络

用一个循环神经网络模拟大脑某脑区神经网络的理想情况是循环网络内各个计算单元之间的连接权重接近大脑神经网络内各个神经元之间的突触权重。通常各个计算单元的连接权重可以通过神经元活动数据和优化算法训练得到，从而间接确定循环网络的连接权重[32, 33]。最近，Pollock 和 Jazayeri 提出了一个方法用流形向量场直接合成循环网络内的连接权重[34]。合成的循环网络与流形向量场拥有相同的动力学特性。他们在滞后颜色识别任务中成功验证了这个方法。如图 2.11a 所示，滞后颜色任务需要被试先观看某一颜色，然后在一段时间之后选择颜色环上的相同颜色。因为外部刺激的颜色取自一个闭合的颜色环，所以对应于该任务的神经流形应该是一个 1 维环（图 2.11b）。环上每一点的速度的方向与环相切，其大小假设由一个沿着环的波形曲线确定。循环神经网络所表达的动力学特性需要近似于这个环上的向量场（图 2.11c）。循环神经网络中计算单元的动态过程可以简化地表达为$\dot{x} = W\sigma(x)$，这里向量$\dot{x}$代表所有计算单元的变化量，x代表所有计算单元的当前活动量，$\sigma(\cdot)$是计算单元本身已知的非线性函数，W是未知的循环网络的连接矩阵。我们知道x等价于神经活动空间中向量v。通过给定的流形向量场$\{(v_i, \dot{v}_i)\}$，我们可以得到一个线性方程组$\dot{v}_i = W\sigma(v_i)$，$i = 1, 2, \ldots$。我们可以解出W从而确定循环神经网络的连接权重。这个根据流形合成而非根据数据优化的方法具有很高的可解释性。

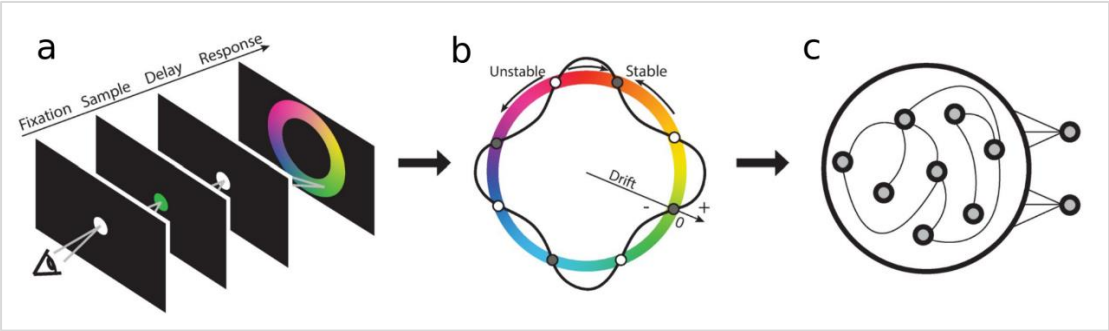


图 2.11 a）滞后颜色识别任务；b）环状流形上的速度向量的方向与环相切，大小由环上的波形函数确定；c）根据流形向量场计算循环网络连接权重[34]

2.2.7 总结和展望

本章从神经科学的角度，以内侧颞叶海马的位置细胞和内嗅皮层的网格细胞为例，介绍了单个神经元的普适性特征。自从 1971 年 O’Keefe 和同事 Dostrovsky 发现单神经元对特定空间位置的选择性放电现象以来，随着过去 50 年神经元信号观测技术的进步，以及对神经元编码机制的研究积累，当今的神经科学研究正在从单神经元编码主导的时代朝神经元群体编码主导的时代过渡，在这个具有历史意义的特殊时期，为了应对解码群体神经元放电机理的新挑战，神经流形将提供了一个新的视角，为神经表征的研究引入基于几何分析的新概念和新工具。我们认为其中最重要的概念是 Jazayeri 和 Ostojic 的三层维度框架[21]。

利用这个三层维度框架，我们可以在 Marr 的计算-算法-实现的三个层次上[35]系统地分析神经流形如何实现神经元群体编码。首先，在实现层，我们需要进行生物学观测，采集神经元信号，这些被观测的神经元的活动构成了神经活动空间的坐标轴。接着，我们需要将神经活动空间降维成嵌入空间。低维的嵌入空间简化了我们在计算层对神经元动力学信号的分析，使我们可以通过形式化或者流形向量场的方式对其进行描述。最后，在算法层，我们可以通过循环神经网络对所观测脑区进行模拟。这样的模拟可以发生在嵌入空间或者神经活动空间。在嵌入空间内，我们可以验证神经网络的模拟是否正确反映计算层的动力学描述；在神经活动空间内，我们可以对比模拟的计算单元信号是否跟观测到的神经元信号一致。

神经内部世界才刚刚给我们展现了它的一角。当前的神经科学仅初步的挖掘了流形潜能。流形的更多特性和它背后成体系的数学分析工具将持续不断地为我们理解神经元群体编码机制提供动力。

参考文献

[1] Tolman, E. C. (1948). Cognitive maps in rats and men. Psychological review, 55(4), 189.

[2] O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. Brain research.

[3] Bostock, E., Muller, R. U., & Kubie, J. L. (1991). Experience-dependent modifications of hippocampal place cell firing. Hippocampus, 1(2), 193-205.

[4] Moser, E. I., Moser, M. B., & McNaughton, B. L. (2017). Spatial representation in the hippocampal formation: a history. Nature neuroscience, 20(11), 1448-1464.

[5] Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. (2016). Organizing conceptual knowledge in humans with a gridlike code. Science, 352(6292), 1464-1468.

[6] Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. Nature, 463(7281), 657-661.

[7] Nau, M., Schröder, T. N., Bellmund, J. L., & Doeller, C. F. (2018). Hexadirectional coding of visual space in human entorhinal cortex. Nature neuroscience, 21(2), 188-190.

[8] Bao, X., Gjorgieva, E., Shanahan, L. K., Howard, J. D., Kahnt, T., & Gottfried, J. A. (2019). Grid-like neural representations support olfactory navigation of a two-dimensional odor space. Neuron, 102(5), 1066-1075.

[9] Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. Neuron, 56(2), 384-398.

[10] Banino, A., Barry, C., Uria, B., Blundell, C., Lillicrap, T., Mirowski, P., ... & Kumaran, D. (2018). Vector-based navigation using grid-like representations in artificial agents. Nature, 557(7705), 429-433.

[11] Nieh, E. H., Schottdorf, M., Freeman, N. W., Low, R. J., Lewallen, S., Koay, S. A., Pinto, L., Gauthier, J. L., Brody, C. D., & Tank, D. W. (2021). Geometry of abstract learned knowledge in the hippocampus. Nature, 595(7865), 80–84. https://doi.org/10.1038/s41586-021-03652-7

[12] Elsayed, G. F., Lara, A. H., Kaufman, M. T., Churchland, M. M., & Cunningham, J. P. (2016). Reorganization between preparatory and movement population responses in motor cortex. Nature Communications, 7(1), 13239. https://doi.org/10.1038/ncomms13239

[13] Steinmetz, N. A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., ... & Harris, T. D. (2021). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. Science, 372(6539).

[14] Scholz, E. (1999). The Concept of Manifold, 1850–1950. In History of Topology (pp. 25–64). Elsevier. https://doi.org/10.1016/B978-044482375-5/50003-1

[15] Pitts, W., & McCulloch, W. S. (1947). How we know universals the perception of auditory and visual forms. The Bulletin of Mathematical Biophysics, 9(3), 127–147. https://doi.org/10.1007/BF02478291

[16] Amari, S. I. (1998). Natural gradient works efficiently in learning. Neural computation, 10(2), 251-276.
[17] Chung, S., & Abbott, L. F. (2021). Neural population geometry: An approach for understanding biological and artificial neural networks. Current Opinion in Neurobiology, 70, 137–144. https://doi.org/10.1016/j.conb.2021.10.010
[18] Seung, H. S., & Lee, D. D. (2000). The Manifold Ways of Perception. Science, 290(5500), 2268–2269. https://doi.org/10.1126/science.290.5500.2268
[19] Vyas, S., Golub, M. D., Sussillo, D., & Shenoy, K. V. (2020). Computation through neural population dynamics. Annual Review of Neuroscience, 43, 249-275.
[20] Gallego, J. A., Perich, M. G., Miller, L. E., & Solla, S. A. (2017). Neural Manifolds for the Control of Movement. Neuron, 94(5), 978–984. https://doi.org/10.1016/j.neuron.2017.05.025
[21] Jazayeri, M., & Ostojic, S. (2021). Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. Current Opinion in Neurobiology, 70, 113–120. https://doi.org/10.1016/j.conb.2021.08.002
[22] Guanella, A., Kiper, D., & Verschure, P. (2007). A model of grid cells based on a twisted torus topology. International journal of neural systems, 17(04), 231-240.
[23] Derdikman, D., & Moser, E. I. (2011). A manifold of spatial maps in the brain. Space, Time and Number in the Brain, 41-57.
[24] Sadtler, P. T., Quick, K. M., Golub, M. D., Chase, S. M., Ryu, S. I., Tyler-Kabara, E. C., ... & Batista, A. P. (2014). Neural constraints on learning. Nature, 512(7515), 423-426.
[25] Sereno, A., & Lehky, S. (2011). Population Coding of Visual Space: Comparison of Spatial Representations in Dorsal and Ventral Pathways. Frontiers in Computational Neuroscience, 4, 159. https://doi.org/10.3389/fncom.2010.00159
[26] Lehky, S. R., Kiani, R., Esteky, H., & Tanaka, K. (2014). Dimensionality of Object Representations in Monkey Inferotemporal Cortex. Neural Computation, 26(10), 2135–2162. https://doi.org/10.1162/NECO_a_00648
[27] Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A., & Fiete, I. (2019). The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. Nature Neuroscience, 22(9), 1512–1520. https://doi.org/10.1038/s41593-019-0460-x
[28] Cohen, U., Chung, S., Lee, D. D., & Sompolinsky, H. (2020). Separability and geometry of object manifolds in deep neural networks. Nature Communications, 11(1), 746. https://doi.org/10.1038/s41467-020-14578-5
[29] Hénaff, O. J., Goris, R. L. T., & Simoncelli, E. P. (2019). Perceptual straightening of natural videos. Nature Neuroscience, 22(6), 984–991. https://doi.org/10.1038/s41593-019-0377-4
[30] Hénaff, O. J., Bai, Y., Charlton, J. A., Nauhaus, I., Simoncelli, E. P., & Goris, R. L. T. (2021). Primary visual cortex straightens natural video trajectories. Nature Communications, 12(1), 5982. https://doi.org/10.1038/s41467-021-25939-z
[31] Claudi, F., & Branco, T. (2021). Differential geometry methods for constructing

manifold-targeted recurrent neural networks [Preprint]. Neuroscience. https://doi.org/10.1101/2021.10.07.463479

[32] Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature, 503(7474), 78–84. https://doi.org/10.1038/nature12742

[33] Mohan, K., Zhu, O., & Freedman, D. J. (2021). Interaction between neuronal encoding and population dynamics during categorization task switching in parietal cortex. Neuron, 109(4), 700-712.e4. https://doi.org/10.1016/j.neuron.2020.11.022

[34] Pollock, E., & Jazayeri, M. (2020). Engineering recurrent neural networks from task-relevant manifolds and dynamics. PLOS Computational Biology, 16(8), e1008128. https://doi.org/10.1371/journal.pcbi.1008128

[35] Peebles, D., & Cooper, R. P. (2015). Thirty years after Marr's vision: Levels of analysis in cognitive science. In Topics in cognitive science (Vol. 7, Issue 2, pp. 187–190). Wiley Online Library.

第3章 类脑视觉

当前，以深度卷积神经网络为代表的机器视觉模型在图像识别等任务中取得了重大成功，但依然远逊于人类的视觉系统。深度卷积神经网络主要近似地模拟了大脑视觉系统中的腹侧前馈通路，为了发展更好的视觉系统，从图像识别走向图像理解，我们需要更多地从大脑借鉴，发展类脑视觉。

类脑视觉的目标是模拟生物视觉系统的架构和计算原理，发展高效的视觉计算模型与算法，支撑类脑智能的应用。尽管我们每个人都能切身地感受到人类的视觉功能远远超过了当前的机器视觉，但到目前为止，我们还没有发展出高效的类脑视觉技术，更没能在一个杀手级应用（killer application）中展现突破性进展（如 AlphaGo 在围棋比赛、AlphaFold 在蛋白质结构预测等），从而令人信服地展现类脑视觉的优点。那么问题出在哪里呢？简单地说，这是因为生物视觉与当前机器视觉在计算任务和信号表达上都存在根本的不同。比如，深度学习网络在做图像识别时，处理的是静止图片信息；而生物视觉系统在实现认知功能时，处理的是时空运动模式信息。而当前脉冲神经网络模型并没有充分考虑这些根本性的不同，因而没有体现出生物视觉的诸多优点。因此，为了真正借鉴生物大脑的视觉智能，我们有必要发展全新的类脑视觉范式，界定新的适合类脑视觉的认知计算任务，并研发出新的类脑视觉计算模型。

本节将围绕构建新的类脑视觉范式来展开，首先介绍当前一些拟视网膜的类脑相机，其原理是从信号的采集模式上就开始类脑；其次介绍模拟生物神经网络的三个类脑计算模型，包括运动目标快速检测，运动目标预测追踪，以及运动目标识别，它们构成了类脑视觉的基本功能模块；最后展望类脑视觉的未来发展之路。

3.1 类脑视觉从采集信号开始

与当前机器视觉主要处理静态图像相比，生物视觉的一大特征是加工时空动态信息（图 3.1）。从视网膜开始，生物神经系统接受的输入就是动态光流。这些光流信号在视网膜内被转化为脉冲序列信号，然后通过层级加工被传输到大脑视觉皮层；大脑各功能区域的信息加工以及区域之间的信息传递也是通过神经元群所产生的脉冲序列来完成。总的说来，生物视觉系统加工的信息都是以神经脉冲序列为表达形式的时空动态信息。静止图像其实只是我们的“幻觉”。因此，为了仿真生物视觉，我们有必要在信号源上就抓住生物视觉时空动态输入的特点。

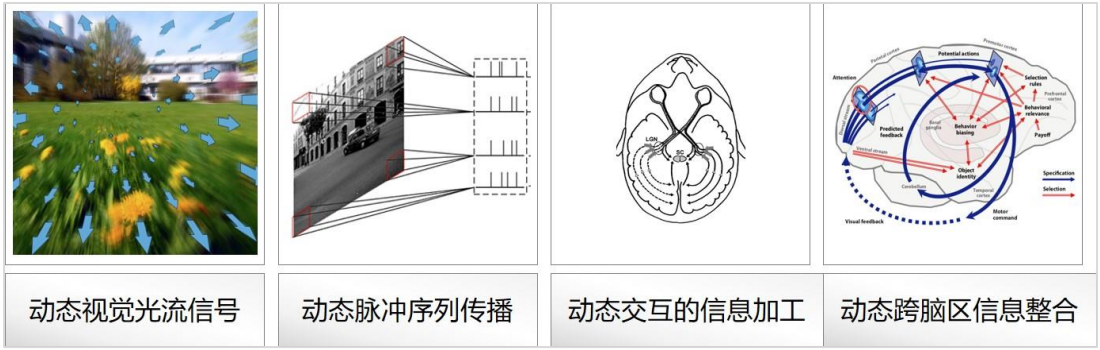


图 3.1 生物视觉系统加工基于脉冲序列的时空动态模式

近年来，类脑感知器件技术快速发展，已经可以实现对外部视觉场景高时间分辨率的、持续的感知，并将光信号转换为脉冲序列进行表征。这里主要介绍两种类脑感知器件，一种是动态视觉传感器（Dynamical vision sensor，简称 DVS），另一种是脉冲摄像头（spiking camera），如 vidar 脉冲摄像头。两者均受启发于生物大脑的视网膜系统。

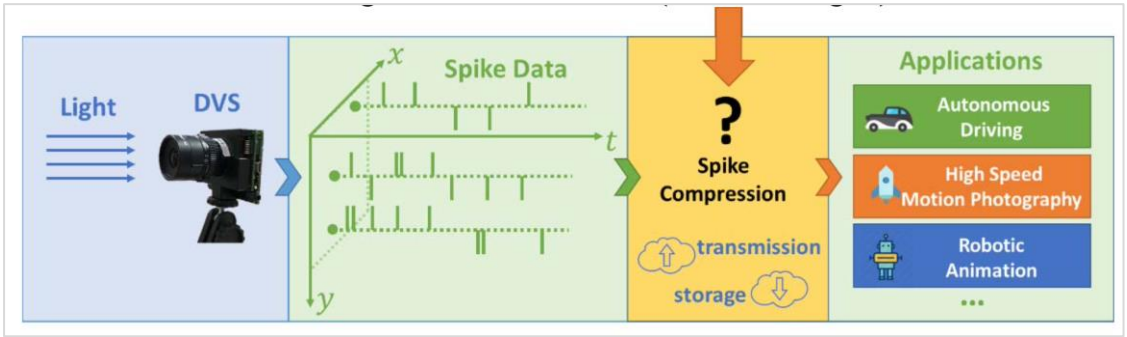


图 3.2 DVS 示意图以及其脉冲信息处理所面临的挑战[1]

DVS 是一种拟视网膜的感知设备。传统的摄像头是按固定的时间频率（比如 30fs）获取一帧帧图像信息，每一帧图像信息包含了每一个像素点的视觉信息，而忽略像素亮度变化的信息。DVS 以非同步化的方式感知外部图像中每一个像素点亮度的变化，并输出一个脉冲事件流，如图 3.2 所说。DVS 会以一种地址-事件表征的方式输出这些脉冲信号。每一个脉冲信号包含四部分信息，水平位置x，垂直位置y，脉冲发放时刻t，以及表征正负性的值p，即$< x, y, t, p >$。前三者表征了脉冲信号的时空位置，后者表征了亮度变化的方向。在 DVS 中，一个像素位置的亮度变化只有超过一定的阈值时，才会发放一个脉冲信号，故 DVS 产生的脉冲信号在空间上是稀疏的，时间上是离散的。作为一种生物启发的方法，近年来，DVS 已经被用于模拟一个三层的视网膜系统，实现一个简化的感光细胞-双极细胞-神经节细胞通路[2]。DVS 相比于传统的摄像头，其具备高动态范围，高时间分辨率，低能量消耗以及高像素带宽等特性 [3]。DVS 只记录了相对的光强变化，更多的是对视觉的运动信息进行了感知，而忽略视觉场景的纹理细节等。

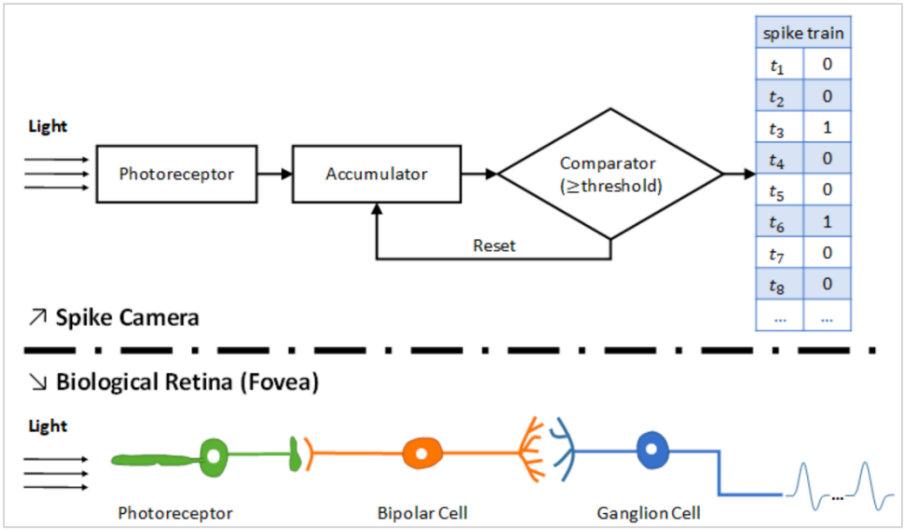


图 3.3 脉冲摄像头[4]

脉冲摄像头受启发于大脑的中央凹视觉系统[4]。脉冲摄像头构建了一个像素处理阵列，每一个像素位置，含有一个模拟到数字的转换器（analog-to-digital converter，简称 ADC）和累加器。转换器模块可以将每一个像素位置的光强转换

为电压，随后输出被送入累加器，如图 3.3 所示。累加器对输入光强进行累加，当累积的输入超过一定的阈值时，一个脉冲放电产生，同时累加器被重置。在一个采样时刻，如果一个脉冲发放产生，输出为 1 ，否则为 0。转换器类似一个感光细胞，而累加器类似一个整合发放神经元，整个过程可以视为对光感受器-双极细胞-神经节细胞的简化模拟。不同于 DVS 记录光强的相对变化，脉冲摄像头可以对绝对的光强进行表征。一个像素位置的光强越大，对应的转换器的输出越大，累加器会更快地达到阈值，从而脉冲摄像头中对应的像素位置脉冲发放越频繁。每一个累加器的输出和重置均为非同步化的。基于脉冲摄像头的脉冲序列，可以重构出场景的纹理特征。脉冲摄像头对数据的采集可以高达毫秒级时间分辨率，理论上，我们可以从脉冲数据中，获取任意时刻场景的光强信息。

3.2 类脑视觉的基本计算模型

基于类脑感知器件采集的脉冲序列，我们可以进一步模拟生物视觉系统的信息加工方法，发展类脑的计算模型。视觉功能应用有很多，但基本模块主要包括三个：目标检测、目标追踪、和目标识别。人工神经网络领域已经发展了大量的算法，但这些算法主要都是基于静止图像加工，不能处理好时空动态模式。简单的把前馈人工神经网络脉冲化并不能完成这样的计算任务。我们需要借鉴生物神经系统的结构特点来发展高效的类脑视觉计算模型，尤其是要引入生物神经网络的动力学计算过程，因为生物神经网络本身就是自然长期进化来优化加工时空动态信息的结果。

接下来，我们简要介绍基于生物神经网络发展而来的类脑视觉计算的几个初步模型，包括运动目标快速探测、运动目标预测追踪、以及运动目标识别。

3.2.1 运动目标快速探测的类脑模型

生物大脑对外部刺激的快速反应对动物的生存至关重要，无论是在捕食者突然出现时迅速逃跑，还是在复杂的地形中迅速而平稳地通过，都需要对视觉信息快速处理与响应。

为了在自然环境中生存，生命体进化出了迅速对视觉刺激做出响应的能力。例如，人类被试可以在 150 ms 内完成复杂的视觉信息处理，猕猴视觉皮层的神经元的响应延时只有几十毫秒。同时，实时快速处理信号变化的能力在工程上也有很高的需求。对于脉冲视觉信号而言，其和传统视觉信号有着很大不同，最近新研发出来的基于脉冲放电的 Vidar 相机就是一个例子[4]，它具有高达 40,000 帧/秒的采样速度，远超普通相机的 60 帧/秒，因此可以捕捉到极高速运动的物体及物体上的细节。为了用高速脉冲视觉信号进行实时探测，我们需要专门设计可以实时快速处理这样信号的算法。

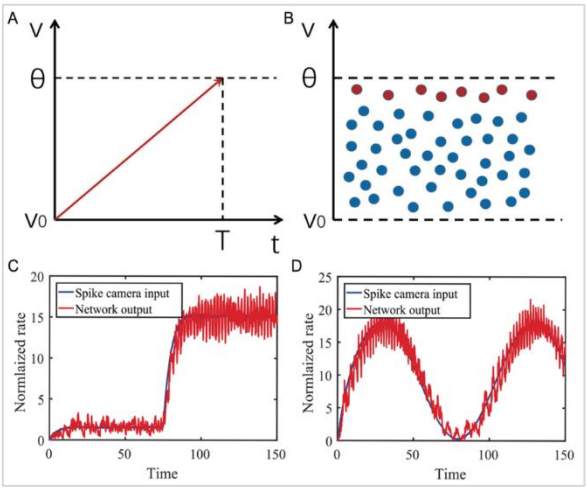


图3.4 兴奋-抑制平衡网络对输入变化的快速响应。（A）单神经元接受“干净”输入的整合发放过程，神经元的整合时间由神经元的膜电位时间常数决定。（B）神经元群的膜电位分布使得神经元群可以快速的响应外部刺激，红色的点表示接近发放阈值的神经元，其可以快速地响应外部刺激变化。（C,D）兴奋-抑制平衡的神经网络对不同的，随时间变化的刺激可以快速响应，其中蓝色的曲线为脉冲摄像头的输入，红色的曲线为神经元群的输出[6]。

一般的神经网络如果没有特别设计，其对快速变化的刺激会有响应延迟。比如，神经元整合外部的信息需要一定的时间，这个时间取决于神经元的膜电位时间常数。那么神经系统是通过什么方式实现了对外部刺激的快速响应？已有理论研究指出，兴奋-抑制平衡网络（excitation-inhibition balanced neural network，下面简称为“平衡网络”）可以帮助网络实现对外部变化刺激的快速响应[5]。兴奋-抑制平衡是生物神经网络的基本特性，已被大量的实验所证实。在平衡网络中，每个神经元接收到的兴奋性与抑制性输入大致互相抵消，从而使得神经元以较低频率发放，并使得每个突触对神经元的状态有较大影响，使总输入得以产生大的

波动，令发放的脉冲序列十分不规则。平衡网络可以快速地追踪变化的刺激。对于单个神经元，其整合外部的信息需要一定的时间，这个时间取决于神经元的膜电位时间常数，如图 3.4A 所示。当一个神经元群共同接收一个信号的时候 ，如果此时系统内的某种噪音促使不同神经元膜电位处于不同的水平，这样会维持一个膜电位的分布，那么总会有神经元的膜电位会离发放阈值较近，此时可以迅速地通过其发放行为的变化来对输入变化做出响应。此时，这个集群就作为一个整体，其响应时间不被单个神经元的积分时间所限制，如图 3.4B 所示。这种机制的关键就是防止神经元的同步放电以及维持神经元群中膜电位的稳定分布，而异步放电则恰好是 E-I 平衡网络的特征之一。Tian 等人将该模型推广到类脑计算中，在脉冲摄像头数据上证实了兴奋-抑制平衡网络可以实现对高速运动目标的快速探测[6]（图 3.4C-D）。

3.2.2 运动目标预测跟踪的类脑模型

视觉信息在大脑中传递时存在不可避免的延迟，比如视觉信号从视网膜到初级视觉皮层需大约 50~80ms，单个神经元的响应延迟大概为 10~20ms。如果这些延迟不能被合适的补偿，我们对运动物体的感知将会滞后物体在外部世界的真实位置，这将为大脑处理运动信息，如目标追踪等任务带来巨大的不便。Mi 等人研究了生物神经网络实现预测跟踪的计算机制，提出基于连续吸引子网络，并引入负反馈机制可以实现网络对外部运动输入信号的预测追踪[7]。吸引子指的是一个动力学系统在不接受外界输入情况下靠自身动力学就能维持的非静息的稳定状态，其被广泛认为是神经系统表达信息的方式[8]。经典的 Hopfield 模型没有考虑神经元之间连接的对称结构，因而其具有的吸引子在空间上是相互孤立的,如图 3.5A 所示。而连续吸引子网络考虑神经元之间的连接具有空间平移不变性的对称结构，这时网络具有一簇连续的、而不是孤立的吸引子。这些吸引子在参数空间上紧密排列，构成一个连续的状态子空间， 如图 3.5B-C 所示。连续吸引子网络被成功地用于解释运动方向编码、视觉朝向编码、头朝向编码、空间位置编码等, 并被直接的实验数据证实其存在于生物大脑中[8]，如图 3.5D 所示。

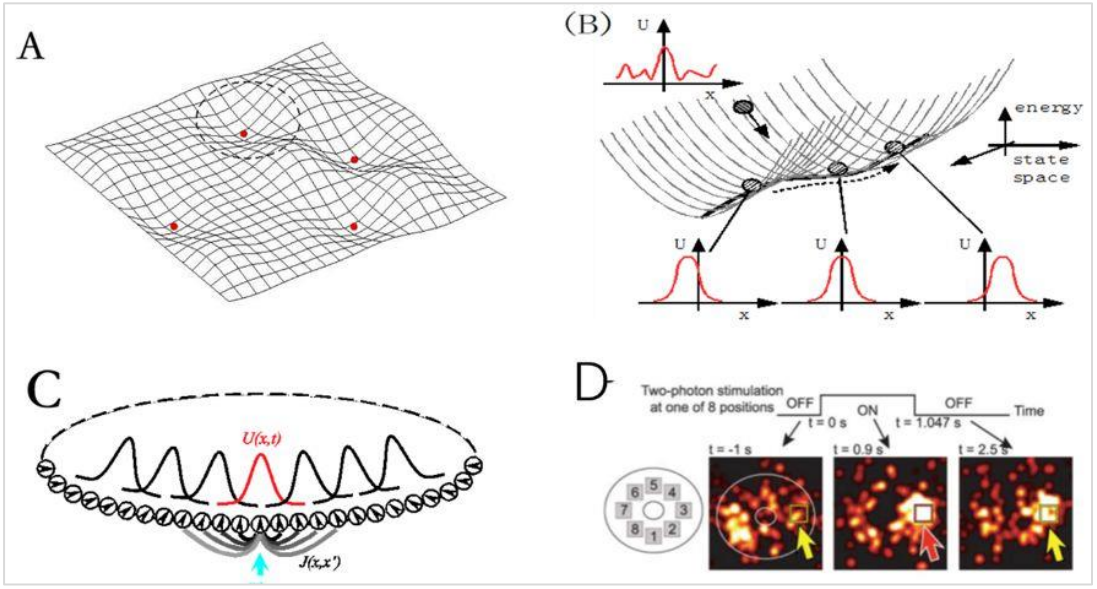


图 3.5 (A) 离散的吸引子神经网络示意图。每个吸引子是能量空间中的局部极小点。(B) 一维 CANN 示意图。一簇连续的吸引子在能量空间中构成一个一维的平滑子空间，在其上系统随遇平衡。每个吸引子对应于一个高斯波包活动状态。(C) 编码朝向的一维 CANN 示意图。(D) 果蝇大脑中头朝向神经元的分布图[9]

在连续吸引子构成的状态子空间中，由于能量函数是平的，系统状态处于随遇平衡；这意味着在外部微小输入的驱动下，系统可以轻松改变状态，得网络状态能够平滑跟踪外部运动输入，但实现的跟踪是滞后于运动目标的实时位置的，因为网络中神经元的反应和信号传递都需要时间。这一点解释不了头朝向跟踪的实验发现，即神经系统实现的运动跟踪是提前的，且领先一个恒定的时间值，几乎不依赖目标的运动速度。为了实现预测跟踪，Mi 等人研究发现，如果在神经元动力学中引入神经系统广泛存在的负反馈机制（其可以是单个神经元发放频率的自适应（spike frequency adaptation），如图 3.6A 所示）[7]、神经元之间突触的短时程衰减(short-term depression)、或者不同层间的反馈抑制等），那么连续吸引子网络就可以实现时间上恒定领先的运动预测跟踪，如图 3.6B 所示。Mi 等首先发现在引入互反馈机制后，连续吸引子网络能维持一个行波解(travelling wave)，且行波的速度（网络的内在速度）由负反馈强度调制。在接受运动输入的情况下，网络中波包移动速度被锁定到输入的运动速度，但其空间位置是领先还是落后于目标位置则是由网络内在速度与目标运动速度的相对大小来决定的，如图 3.6C 所示：当内在速度大于目标运动速度时，预测就发生了。理论研究表明，在很大

一段速度范围内，网络波包领先的距离和目标速度成正比，即预测的时间是固定的，这一点和实验数据吻，如图 3.6D 所示。

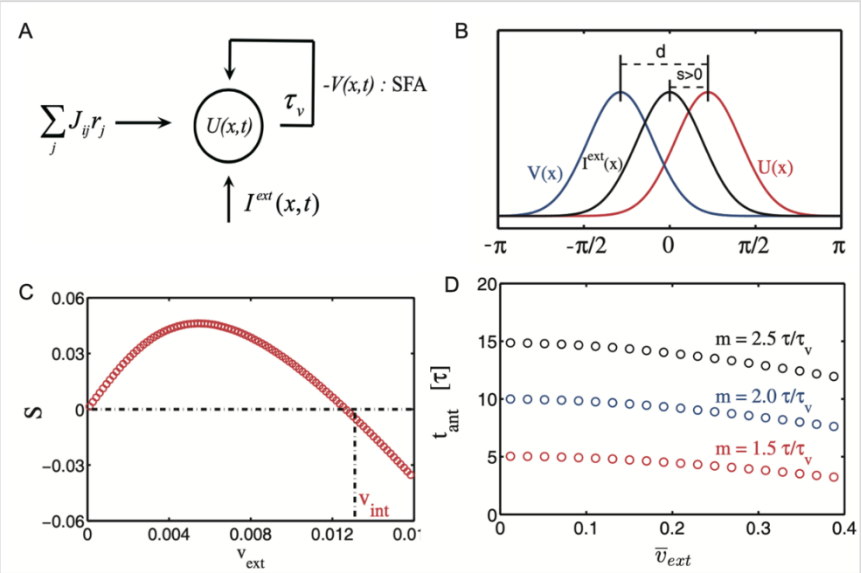


图 3.6 引入了适当负反馈作用的连续吸引子网络可以实现运动目标的预测跟踪。A，负反馈机制示意图；B，网络波包领先外部输入示意图；C，领先的条件是内在速度大于目标速度；D，在很大一段速度范围内，领先时间近似为常数[7]。

生物连续吸引子网络的这种强大预测跟踪能力为我们提供了一种类脑的运动目标预测跟踪算法。该算法的优点包括：1）预测跟踪的时间是恒定的，几乎不依赖于目标运动速度；2）模型的参数是可以根据任务，理论上预先设定的，不需要大数据训练；3）连续吸引子网络也可以被类脑芯片实现[10]。

3.2.3 运动目标识别的类脑模型

生物系统的视觉通路分为皮层上通路和皮层下通路。目前大多数神经科学的视觉研究的焦点集中在皮层上通路，如深度卷积神经网络被认为是模仿了大脑的皮层上通路中的腹侧通路。皮层上的腹侧和背侧通路对于视觉感知是非常重要的，但生物的皮层下通路同样非常重要。对于高等动物而言，皮层下视觉通路在本能地快速检测危险信号时非常重要 [11]。对于低等动物，因为缺乏新皮层，它们的视觉不可避免地只能主要依靠皮层下视觉通路来实现。一些关于先天盲人的实验表明，视觉皮层受损的盲人在某些情况下依然能绕过障碍物或者在无意识的情况下检测到警觉性的视觉刺激[12]，这背后的机理归功于皮层下视觉通路。

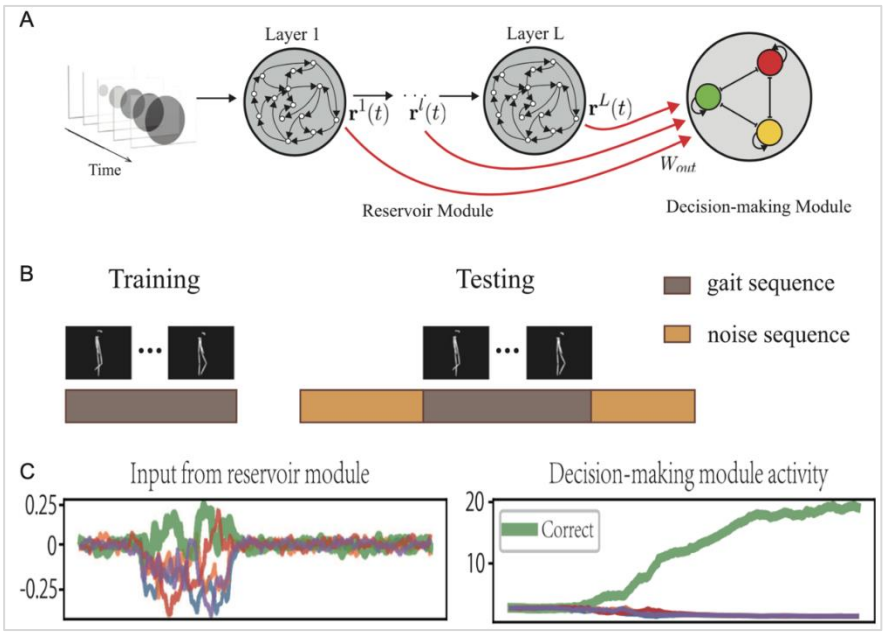


图 3.7（A）模拟皮层下视觉通路的模型框架，其包含两部分，库网络和决策网络。(B)基于事件驱动的步态识别任务。(C) 在基于事件驱动的步态序列样例下，库网络和决策网络的响应神经元活动[14]。

皮层下视觉通路是一条进化上非常保守的通路。在这条通路中，视网膜将视觉信号传递到上丘核团（superior colliculus），然后传递到丘脑枕（pulvinar），最后传递到视觉皮层进行整合。研究认为，这条通路主要负责运动物体的快速识别，特别在介导生物的逃跑反应过程中起到了关键作用。最近的实验也发现[13]，上丘核团与整合时空信息进行决策之间存在重要的因果关系。Lin 等人最近对从视网膜到上丘的通路进行了建模[14]，提出了新的类脑运动模式识别算法。该算法由两个部分组成，一个是深度库网络，它负责将外部输入的时空序列投影到高维的网络活动空间中；另一个是决策网络，决策模型中存在相互抑制和自兴奋，可以以一种生物合理的方式不断累积接受到的不同类别，不同强度的证据，据此来作出判断，如图 3.7A 所示。Lin 等将传统的二分类决策模型进行了简化，并推广到了多分类决策任务,并对决策模型进行了细致的理论分析和参数分析。和库网络进行结合，该类脑算法能像生物一样高效地利用少量样本对时空序列进行识别,并在步态识别的应用中证明了模型的实用性。更有意思的是，该模型可以以一种基于事件驱动的方式进行运动目标识别，即当外部输入为背景噪声时，决策网络神经活动处于静息状态，但当外部输入为用信号时，决策网络开始累积信息并

做出决策，如图 3.7B,C 所示。该模型可以应用于一般性的基于脉冲信号的时空模式识别。

3.3 总结与展望

更加贴近生物大脑的脉冲神经网络模型，早在上世纪 90 年代就被誉为“第三代神经网络模型”[15]。但是由于多种因素制约，目前基于脉冲信号的类脑视觉算法还远远落后于人工神经网络模型，我们急需新的类脑视觉范式来促进该领域的发展。近年来，随着基于脉冲信号的摄像头的大量应用，我们已经可以采集大规模的复杂场景的脉冲时空数据；随着更深入理解生物大脑智能，我们可以发展更优的适合加工脉冲时空数据的类脑计算模型；同时随着神经拟态芯片的快速发展，我们可以有更高效的运算平台来支撑类脑智能技术的应用。这些相关领域的协同发展可以大大促进类脑视觉的进步。下面，我们展望未来类脑视觉新范式发展的一些核心要素：

（1）更生物的类脑感知器件。未来我们需要发展模拟视网膜更多计算特性的类脑感知器件。比如，DVS 记录了运动信息，近似模拟了视网膜的外周信息处理，而脉冲摄像头记录了细节纹理信息，近似模拟了视网膜的中央凹信息处理。将两者信息进行融合，可以更好地模拟生物大脑的视网膜感知功能[16]。

（2）更智能的类脑视觉计算模型。前面只是介绍了模拟生物视觉系统的三个简单模型，即目标检测、跟踪、与识别。真实生物视觉系统具有更丰富的结构和计算功能，比如视觉信息可以通过多通路并行加工，视觉认知由全局到局部，先验知识通过反馈连接来影响我们对外部输入的感知等。我们可以借鉴这些生物视觉系统的结构特点，发展更优、更丰富的类脑视觉计算模型。此外，我们还可以借鉴机器学习的训练方法，从数据和功能出发来优化类脑计算模型。

（3）更合适的类脑视觉应用场景。由于生物视觉系统的存在是为了生命体更好地适应自然环境，其计算优势也体现在与周围环境的高效动态交互。因此，为了体现类脑视觉的优点，我们需要界定合适的类脑视觉应用场景，而不是简单应用于深度学习所擅长的计算任务。

（4）类脑视觉的编程环境。深度网络的快速发展除了依赖大数据和超算外，Pytorch，TensorFlow 等便捷软件的出现也做出了关键贡献，这些编程工具大大地降低了领域的入门门槛，推动了领域的高速发展。同样，针对脉冲序列加工的包含了神经网络动力学过程的类脑计算模型也需要高效、方便的编程平台来助力。目前已有了一些类脑计算软件平台，如 BrainPy（灵机）[17]等。

参考文献

[1] Bi, Z., Dong, S., Tian, Y., & Huang, T. (2018, March). Spike coding for dynamic vision sensors. In 2018 Data Compression Conference (pp. 117-126). IEEE.

[2] Posch, C., Serrano-Gotarredona, T., Linares-Barranco, B., & Delbruck, T. (2014). Retinomorphic event-based vision sensors: bioinspired cameras with spiking output. Proceedings of the IEEE, 102(10), 1470-1484.

[3] Gallego, G., Delbrück, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., ... & Scaramuzza, D. (2020). Event-based vision: A survey. IEEE transactions on pattern analysis and machine intelligence, 44(1), 154-180.

[4] Zhu, L., Dong, S., Huang, T., & Tian, Y. (2019, July). A retina-inspired sampling method for visual texture reconstruction. In 2019 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1432-1437). IEEE.

[5] S. Denève, C. K. Machens. Efficient codes and balanced networks[J]. Nature Neuroscience, 2016,19:375–382.

[6] Tian, G., Li, S., Huang, T., & Wu, S. (2020). Excitation-Inhibition Balanced Neural Networks for Fast Signal Detection. Frontiers in Computational Neuroscience, 14, 79.

[7] Mi, Y., Fung, C. C., Wong, M. K. Y., & Wu, S. (2014). Spike frequency adaptation implements anticipative tracking in continuous attractor neural networks. Advances in neural information processing systems, 1(January), 505.

[8] Wu, S., Wong, KYM., Fung, CCA., Mi, Y., and Zhang, W. (2016). Continuous attractor neural networks: candidate of a canonical model for neural information representation. F1000 Research, 66(16), 209-226

[9] Sung Soo Kim, Hervé Rouault, Shaul Druckmann, Vivek Jayaraman. Ring attractor dynamics in the Drosophila central brain. Science V.356, 849-853, 2017.

[10] Jing Pei et al. (2019). Towards artificial general intelligence with hybrid Tianjic chip architecture. Nature, v.572, p.106–111.

[11] De Franceschi G, Vivattanasarn T, Saleem A B, et al. Vision guides selection of freeze or flight defense strategies in mice[J]. Current biology, 2016, 26(16): 2150-2154.

[12] Zeki S. REVIEW: Parallel Processing, Asynchronous Perception, and a Distributed System of Consciousness in Vision[J]. The Neuroscientist, 1998, 4(5): 365-372.

[13] Latimer, K. W., & Huk, A. C. (2021). Superior colliculus activates new perspectives on decision-making. Nature Neuroscience, 24(8), 1048-1050.

[14] Lin, X., Zou, X., Ji, Z., Huang, T., Wu, S., & Mi, Y. (2021). A brain-inspired computational model for spatio-temporal information processing. Neural Networks, 143, 74-87.

[15] Maass, W. (1997). Networks of spiking neurons: the third generation of neural network models. Neural networks, 10(9), 1659-1671.

[16] Zhu, L., Li, J., Wang, X., Huang, T., & Tian, Y. (2021). NeuSpike-Net: High Speed Video Reconstruction via Bio-Inspired Neuromorphic Cameras. In Proceedings of

the IEEE/CVF International Conference on Computer Vision (pp. 2400-2409).
[17] Wang, C., Jiang, Y., Liu, X., Lin, X., Zou, X., Ji, Z., & Wu, S. (2021, December). A Just-In-Time Compilation Approach for Neural Dynamics Simulation. In International Conference on Neural Information Processing (pp. 15-26). Springer, Cham.

第4章 脑机接口技术与应用

脑中神经网络的活动是人感觉、认知、决策和执行的物质基础。自然进化塑造了视、听、触等感官以及运动、语言等作为脑与环境进行信息交互的自然接口，新兴的脑机接口则是通过对于脑活动信息的检测和调控，在脑与外部世界间建立直接的信息通讯接口。这一技术的发展，有望对于人与环境、人与人的交互方式带来根本变化，从而引起社会、经济、教育、军事、医疗等众多领域的颠覆性变革。为了获得高质量的脑电信号或精准地刺激大脑神经元，最有效的方式就是通过电极阵列与目标脑区的神经元细胞直接接触。然而，由于这种方式侵入性较高具有损伤大脑组织的风险，且需要线缆进行信息传输，在安全性和便捷性方面面临诸多挑战。随着柔性电极技术、植入式芯片以及植入技术的发展，植入式脑机接口技术逐渐向着无线化、便携化方向发展，其安全性也大大提高，有望开启脑机接口研究的一个新纪元。在应用上随着技术的快速发展，脑机接口逐渐被应用到人类受试者身上，其功能也覆盖到运动、触觉、语言、视觉等多个方面，并向着‘脑控’与‘控脑’的双向信息闭环发展。本章节将从技术进展和应用研究两个大的方面来介绍脑机接口的发展。

4.1 脑机接口技术及其发展趋势

大脑中枢神经元膜电位的变化会产生峰电位或者动作电位，并且神经细胞突触间传递的离子移动会产生场电位，而这些神经电生理信号代表了大脑的活动状态或意图。通过在大脑特定的位置和深度插入电极，再利用电子设备采集并放大这些神经生理信号，就可以实现神经信号的记录。脑机接口技术通过直接在大脑与外部环境之间建立一种不依赖于外周神经和肌肉的通道，从而实现大脑与外部设备的直接交互。简单来说，我们目前常用的交互方式是大脑-肢体-外界设备，而脑机接口技术的目的是直接建立大脑-外界设备的通道。该技术能够在人（或其他动物）脑与外部环境之间直接建立沟通通道，以达到控制设备或起到监测、改善/恢复、增强大脑认知功能等作用。

从信息在脑与机之间的流向分析，脑机接口包括下行脑机接口即“脑控”和上行脑机接口即“控脑”两种，分别指用脑活动信号控制外部设备和从外部调控脑活动和脑功能两种应用模式，当前的发展趋势更多的是结合上述两者的双向脑机接口，从而实现“脑控“和“控脑”的信息闭环。从技术实现的路径来看，脑机接口可以分为非侵入式和侵入式两类。非侵入式接口利用放置于头皮外的传感器探测脑活动，安全易用，但是由于所能获得的脑活动信号较为粗略，只能支持较低速率的脑机通讯，更适用于脑状态的检测和调控。侵入式接口通过将电极植入颅骨内的大脑皮层，使电极阵列与目标脑区的神经元细胞直接接触，从而实现高带宽、高质量脑电信号的传递，长远看是实现高速脑机接口甚至人脑与 AI 的混合智能更可能的技术途径。

作为多学科的交叉研究领域，脑机接口技术涉及了神经科学、电子工程、材料科学、人工智能等多个学科，具有重大的科研价值和广泛的应用前景，并在最近取得了重要的技术突破，特别是马斯克的 NeuraLink 公司在侵入式脑机接口方面取得了引人注目的成果。这吸引了包括各国政府、商业机构、学术机构的广泛关注和大量资源投入，有望在未来数年进入技术驱动和需求牵引的正反馈循环，从而实现整个脑机接口领域的快速发展。可以预见未来相关的技术、应用和市场生态也将成为重要的科技竞争要地。

目前脑机接口在技术上的热点和趋势是不断减少侵入式接口的创伤程度、提高安全性和性价比，最终利用微创形式实现高速脑机接口。传统的多通道神经接口技术，通过在颅内插入电极进行神经信号采集或神经电刺激，但是需要通过线缆将采集到的信息传送给体外的神经信号处理系统。近期的植入式脑机接口技术通过无线通信手段完成脑电向外部控制设备的信号传输，无需使用冗杂的线缆，甚至集成片上信号处理功能，为未来真实的脑机接口应用提供了更加安全便捷和更加可行的技术手段。为了进一步减小脑机接口对大脑的损伤，柔性电极具有更好的生物兼容性，但是柔性电极柔软和纤细的特点也导致其难以操作和难以植入。植入机器人技术可以实现对柔性电极的精准植入操作，从而解决人的手眼精度和稳定性无法满足植入要求的问题。此外编解码技术、混合智能技术等都是目前研究的热点，本章在技术方面主要介绍植入式脑机接口芯片和柔性电极植入机器人技术近期在国内外的突破或进展。

4.2 植入式脑机接口芯片

侵入式脑机接口粗略可分为有线和无线两种（如图 4.1）。有线方式通过线缆将采集到的神经信号传输到外部处理平台，通过外部处理平台进行信号的预处理、放大和其他处理，这种方式容易引发感染并且也非常不方便。无线方式通过植入式脑机接口芯片集成了神经信号的记录、预处理等功能，并通过芯片上无线模块进行数据传输，进一步地在双向脑机接口中甚至能够输出可调节的神经电刺激信号。目前在技术上脑机接口芯片主要面临两大需求：一是高通量低功耗，即扩大采集通道数量记录更多神经元的活动，这样才能更精确地获取用户意图，进一步实现神经活动的调控以及外部控制，同时为了实现长期稳定的安全植入，芯片的体积需要足够小，功耗尽量低。另一个是无线化，这个是针对通信和供电而言的，相对于笨重的有线电缆和电池，无线链路可以降低感染风险，改善外观并且方便用户的活动。

图 4.1 有线脑机接口（左）和无线脑机接口（右） [1]

4.2.1 高通量低功耗技术

为了实现高通量低功耗的性能需求，最新研究方案主要集中在两个方面：

1. **组网扩展通道数量。**由于神经信号具有一致性，可以复制实现多通道的扩展。Neuralink 公司提出的 1024 通道的全植入脑机接口 SOC（System on Chip，片上系统）就是 4 个相同的 256 通道子单元的组网，每个子单元主要包括神经记录前端、刺激器、数字信号处理器以及局部能源管理模块[2]。这样的框架设计直接对子单元性能以及各单元网络互连功能提出挑战，而目前关于如何协调控制各单元的研究较少。国内，复旦大学研制了全无线侵入式 64 通道脑机接口芯片模组，未来可以实现多芯片组网，中科院自动化研究所完成了 16 和 64 通道脑机接口芯片的研发和流片，下一步计划进行支持 512 以及更高通道的芯片研制，从而满足未来脑机接口更高通量的性能需求。

2. **片上实现神经信号的处理。**无线链路的带宽难以满足高通量原始神经信号的实时传输，因此需要在片上实现海量数据的处理，而且这对实时调控神经系统是很有必要的。针对这一需求，目前相关研究主要从两个方面进行突破。一是在信号传输前进行信号压缩，只传输感兴趣的原始信号或者经过提取后的神经特征。其中常用的神经特征包括尖峰信号、特定频带内的神经信号能量、熵等。除了检测阈值交叉事件（threshold-crossing events），此外最新有实验证明提取多个神经元尖峰带能量（spiking-band power）特征可以进一步降低功耗[3]。二是在面向闭环控制时，采用信号采集和神经电刺激协同控制的设计思路，结合刺激触发条件处理神经信号。而目前片上数字处理的关键技术瓶颈是如何在满足

应用需求的前提下，降低算法复杂度和芯片功耗来提高脑机接口性能。也有研究从基础神经科学编解码的角度出发，分析对神经信号质量以及特征的要求，从而放宽脑机接口芯片设计指标来实现低功耗小体积[4]。

4.2.2 无线化技术

无线化是指芯片支持无线供电、无线通信的功能，这是当前研究热点，也是未来脑机接口芯片走向实用化的必然要求。就供电模块而言，小型化电池已经难以满足高通量的芯片功耗和长期稳定植入的需求。而应用于脑机接口领域的无线供电方案主要包括，近场电感耦合、远场电磁耦合、光伏供电、超声波功率传输以及通过收集人体自身能量进行供电[5]等。除了提高传输功率满足芯片的功耗需求（目前基于片上阈值处理的芯片每通道消耗功率最少至μW量级[4]），还要考虑传输距离，传输系统的灵活性和鲁棒性等多种因素。目前最常用的无线供电方案是近场电感耦合技术，通过一对耦合线圈产生谐振实现能量传输，从而向植入式脑机接口芯片供电。该方案在短距离内传输效率很高，但是需要线圈对准，大大限制了使用者的活动范围。最近，也有一些团队对新的供电方式展开探索。比如国内中科院自动化所团队尝试人体媒介能量传输技术[6]，即利用人体的高导电性的特点，以人体作为媒介，将发送端置于体表，接收端植入体内，通过小型化电极实现能量在植入物内的传输，具有体积小且灵活性高的独特优势。该方案是人体信道通信技术向无线供电应用的延伸，尽管面向体表供电的技术相对成熟[7]，但针对植入式脑机接口的应用研究尚少。

而关于无线通信方案，其带宽、传输速率、传输距离以及功耗是主要性能指标。除了蓝牙，WiFi 和超宽带等传统通信方式，目前大部分研究集中在使用一对线圈实现数据和功率的同时传输来减小系统体积和重量，但是依旧面临着通信数据速率低，传输距离不够远的问题。而人体信道通信技术是近几年在可穿戴设备间通信领域的研究热点，利用人体作为通信信道的通信方式，具有低功耗，小尺寸，且安全性高的优势，有望成为植入式脑机接口中最佳的无线数据传输方式。国内中科院自动化所团队突破人体信道建模、信道自适应补偿等关键技术，并成

功研制人体信道通信芯片，传输速率最高可达 60Mbps，达到国际先进水平，为实现脑机接口无线通信链路提供新的研究思路。

4.2.3 未来展望

脑机接口芯片的未来发展趋势必然是高通量、低功耗、无线化且全植入的，这对各模块的设计性能和芯片的集成度提出了巨大挑战。对于精确获取并调控神经元的活动，目前最高 1024 采集通道数是远远不够的，有观点认为通道数至少需要扩大至 10 万级以上才有可能满足未来的应用需求。这就需要植入式芯片具备更高通量、更低功耗的无线通信能力和更高效能的无线供电技术。目前脑机接口芯片由于缺乏应用需求，为了控制成本，制造工艺多采用 65nm 制程，但是最新的芯片工艺已经发展至 8nm 甚至更低，这说明脑机接口芯片在技术上还有巨大的发展潜力和空间。随着近年来更加广泛的社会关注和更多的资源投入，未来脑机接口芯片技术有望在技术进步和需求牵引双重驱动下快速迭代发展。总体而言，研发高性能植入式脑机接口芯片，实现长期稳定且灵活度高的植入系统是非常有希望的。

4.3 柔性电极植入机器人

为使植入式脑机接口芯片对大脑的损伤最小化，应最小化电极与组织间的应力与阻抗差异，最小化植入物整体尺寸，最小化对管脉系统的损伤，并在此基础上最大化电极的数量与分布范围。柔性微电极是实现该目标的有效手段，其宽度在几十微米量级，柔性和纤细的特点带来了生物兼容性好、对脑组织损伤小的优点，但同时也带来了杨氏模量小（杨氏模量是描述固体材料抵抗形变能力的物理量）、难以操作、难以植入的技术难题，人的手眼精度与稳定性无法满足柔性微电极的植入。植入机器人可以实现对柔性微电极的精准植入操作，主要基于显微成像、立体视觉等精密感知技术，以及末端微执行器的规划与控制技术，此外植入机器人还要具有术前规划、术中导航、颅骨配准、颅窗成像、激光切除等功能[8-12]。近几年，随着技术的发展，植入机器人技术有重大进展，下面我们将介绍近期国内外在这项技术上的进展。

4.3.1 国际研发进展

2019 年，美国加利福尼亚大学旧金山分校的 Hanson 等人首次研制出了一种面向微创神经接口的柔性电极植入机器人，采用柔性薄膜聚合物微电极、刚性植入针和植入机器人，实现了一种灵活、稳定、安全的“缝纫机”式新型电极植入范式[13]。如图 4.1 所示，柔性电极丝被整齐排列于可替换盒中，电极丝末端从盒中露出并具有环状开孔，植入机器人在计算机视觉引导下将直径 25 微米的植入针尖穿过一个电极丝的末端孔，使该电极丝与相邻电极剥离并可自由移动，然后带动电极植入大脑皮层的目标位置。基于对皮层表面的成像，在规划植入点时可以避开脑部血管。实验中在大鼠的体感皮质区植入了 42 根柔性电极丝，可实现信号的稳定采集。

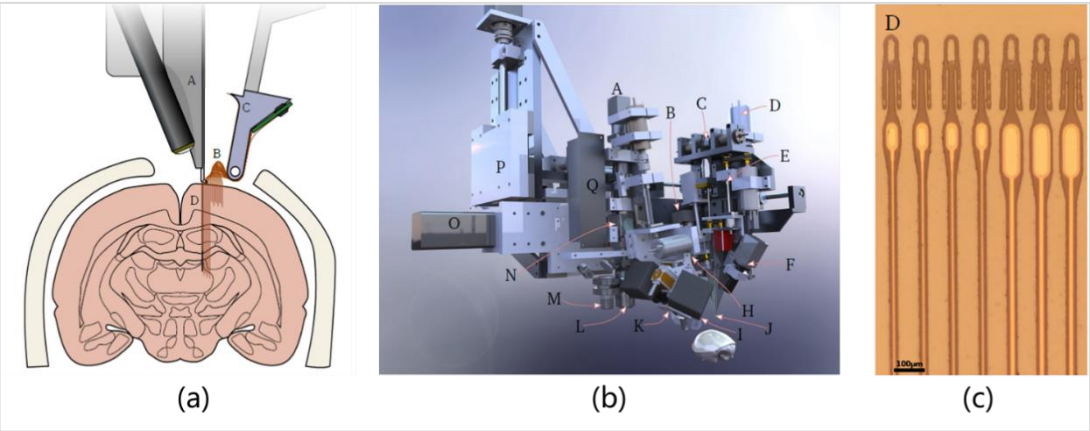


图 4.1 “缝纫机式”植入机器人：(a) 概念图；(b)植入机器人；(c)柔性微电极[12]

同年 2019 年，美国 NeuraLink 公司的创始人 Elon Musk 发布了一种集成化的柔性微电极植入机器人，首次实现了高带宽脑机接口，在实验中成功将 96 根电极丝植入鼠脑，可同时采集多达 3072 个通道的电信号，该原型机代表着脑机接口领域的当前尖端水平[14]。如图 4.2(a)所示，植入机器人的末端安装有植入针、夹持钳、接触力传感器、多波长光源、体视显微镜、多路相机等精密设备，在视觉引导下，机器人每分钟可以植入 6 根电极丝，含有 192 个通道。在 19 次手术实验中，在有医生手动控制辅助微调的情况下，机器人电极植入的平均成功率可达 87.1%。如图 4.2(b)所示，植入在 $4\times7mm^2$ 的颅骨切除区域内进行，电极丝间距不小于 300 微米，覆盖了较大的皮层区域。植入机器人是 NeuraLink 公司

的关键技术，并逐渐向具有临床应用能力的层次发展，如图 4.2(c,d)所示，工业设计公司 Woke Studio 为 Neuralink 的植入机器人设计了优美的曲面外壳，可分散人对侵入式手术的注意力。2020 年，NeuraLink 通过机器人将具有无线传输能力的脑机接口器件植入猪脑中，可捕捉 1024 通道的动作电位信号并进行无线传输，实现对猪的体感与触觉的无线实时监控。2021 年，NeuraLink 进一步将脑机接口器件植入了猕猴大脑的手臂与手部运动区，实现了猕猴通过脑机接口玩电子游戏。

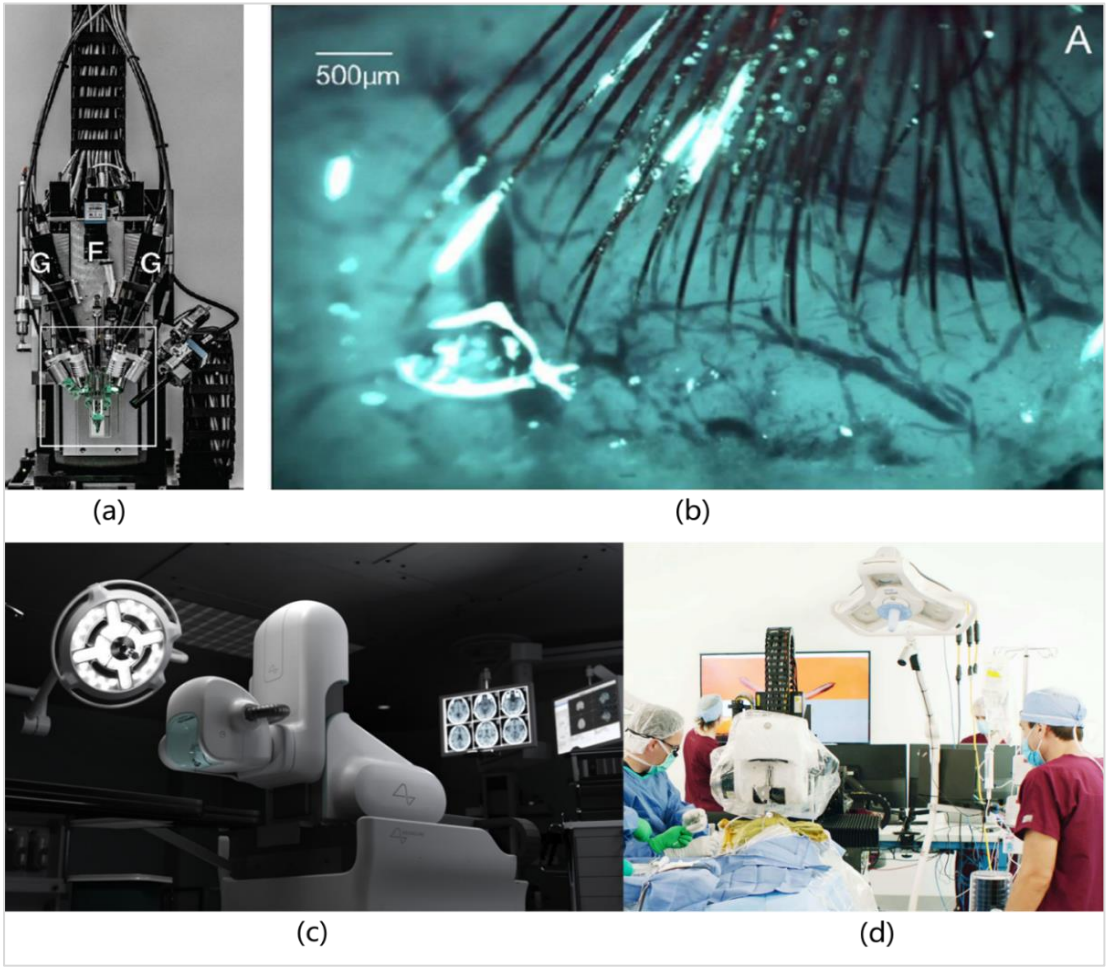


图 4.2 Neuralink 植入机器人与应用[14,15,16]
(a)实物图；(b)电极植入；(c)产品图；(d)产品图

4.3.2 国内研发进展

2021 年，中国科学院自动化研究所的研究人员成功研制了柔性电极精密植入机器人，该机器人具有跨尺度协同运动能力，可实现大范围空间区域内的自

动、精准植入，末端植入装置具有微米量级的运动分辨率，为植入手术提供了更大的灵活性和准确性。机器人通过智能感知实现对电极、植入靶点等目标在三维空间位姿的实时精密测量，并为植入模块提供精确的反馈控制信息。目前该机器人已成功实现小鼠大脑皮层的柔性电极植入实验，有望在不久的将来为柔性电极在大动物的精准植入甚至人类临床实验提供有力的技术支撑。另外，中国科学院半导体研究所团队也研发了半自动柔性电极植入装置，将为柔性电极在脑科学实验中的灵活使用提供便利。

4.3.3 面临的挑战

机器人柔性电极植入技术是一个新兴领域，目前主要在鼠、猪和猴等动物上进行实验，该技术的最终目标是实现服务于人类的临床医学应用，且需要进行大量的探索与验证。2021 年 12 月，NeuraLink 创始人 Elon Musk 在采访中表示，他们有望在 2022 年获得 FDA 许可后首次实现在脊椎神经受损的患者脑部植入柔性电极。对于体积较大的人脑，为了实现更充分的脑机交互，需要增加电极通道数量，甚至在不同部位植入多个集成器件，所以不仅对植入机器人的手术时间效率提出更高要求，而且对于植入手术规划与跨尺度定位提出了更大的挑战。柔性电极的植入成功率是植入手术结果评估的重要指标，代表着目前先进水平的 NeuraLink，其植入成功率也只能达到 87%，未来应进一步提高单根电极的植入成功率，减少无效和不可靠电极导致的植入资源损失。

4.4 脑机接口技术的应用

技术和应用的发展相互促进，相辅相成。技术的发展使得更多更加复杂的应用不再遥不可及，应用的发展也驱使着技术的进步与更新迭代。一般而言，根据信号流的方向，我们可以将脑机接口应用分为下行脑机接口（“脑控”）和上行脑机接口(“控脑”)：下行脑机接口如运动控制假肢/机械臂脑机接口，主要实现从大脑读取神经信号并解码控制外部设备的功能；上行脑机接口则通过将外部传感系统感知的信息编码成可激活大脑产生响应的光/电/声/磁等信号，对大脑进行调控，如用于虚拟触觉重建的脑机接口及用于恢复视/听觉功能的视/听皮层

假体。接下来我们将分别介绍上/下行脑机接口在人类受试者的最新进展，并总结当前脑机接口在进一步向临床应用发展时主要面临的技术瓶颈。

4.4.1 下行脑机接口

下行脑机接口的研究和应用相对较多，下面我们主要介绍运动控制脑机接口、语言脑机接口（语言假体）以及颅内神经记录等方面的研究及最新进展。

（1）运动控制脑机接口

植入式运动皮层脑机接口通过直接采集人类大脑运动皮层神经元发放信号，解码与运动相关的神经发放模式，进而生成指令控制外部机械臂或假肢产生相应的运动，可帮助瘫痪患者实现运动控制。

近年来，有关运动控制的脑机接口已逐渐由猕猴实验进展到瘫痪病人实验。2015 年，加州理工学院 Anderson 团队在瘫痪病人的后顶叶皮层植入电极，提取神经信号并解析获得运动控制指令，使受试患者成功控制外部机械手臂给自己喂饮料[17]；2017 年，Shenoy 团队在三名瘫痪病人的运动脑皮层植入了神经信号采集电极，通过解码信号控制电脑屏幕光标，在打字任务中速度较之前研究提高了三倍[18]。进一步地，随着感觉神经接口的发展， 2020 年美国巴泰尔科研中心(Battelle Institute)和俄亥俄州立大学韦克斯纳医学中心(the Ohio State University Wexner Medical Center)的研究团队提出的双向脑机接口，能帮助一位脊髓严重损伤的受试者同时恢复运动与触觉功能，该系统不仅能让受试者仅靠触觉就能感知到物体，还能够感知握持或捡拾物体时所需的压力[19]，如图 4.4 所示。

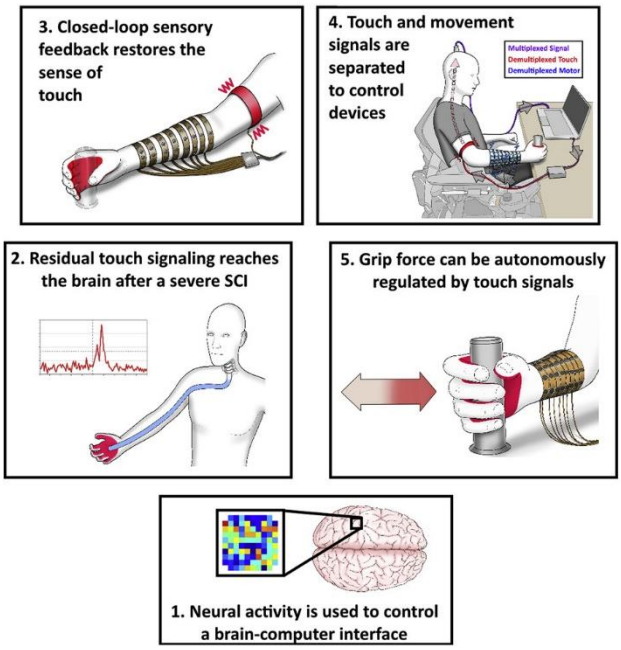


图 4.3 重建触觉感知的脑机接口[19]

（2）语言脑机接口

对于单纯无法说话的残障人士而言，文字书写和手语可以作为有效可靠的交流替代手段。但是，四肢瘫痪的失语残障人士可能因严重的行动障碍而无法操作辅助器具，需另辟蹊径打通他们与家人、朋友和护理人员的交流渠道，以提升患者的自主权和生活质量。

2021 年 7 月，美国加州大学旧金山分校 Edward Chang 等通过在受试患者的感觉运动皮层植入 128 通道的 ECoG 电极，直接从大脑皮层活动中解码单词和句子，成功帮助一位瘫痪超过 15 年的失语男子 BRAVO1 恢复了 “说话” 能力[20]。在该项研究中，受试患者被以文本形式呈现目标单词，并尝试大声“说”出该目标。研究者收集受试患者在练习任务期间的神经活动数据以训练、微调和评估解码系统，使用深度学习技术从神经活动中进行预测受试者的语言内容，并实时反馈给受试患者，如图 4.5 所示。具体而言，解码系统的语言检测模型处理任务期间神经活动的每个时间点，并实时检测单词生成尝试的开始和偏移；单词分类模型结合练习任务期间收集的神经数据，通过处理从检测到的单词生成尝试开始前 1 秒到后 3 秒的神经活动来预测和生成单词的概率模型。自然语言模型和 Viterbi 解码器在句子任务期间实时解码神经活动中的完整句子：结合英

语语言结构特征，自然语言模型在给出前一个词后计算下一个词的概率情况；Viterbi 解码器根据来自单词分类器的预测单词概率和来自自然语言模型的单词序列概率，确定最可能的单词序列。这项研究一方面证明，这名失语超过 15 年的患者的感觉运动皮层仍然保留了对语音的功能表征；另一方面表明位于大脑软膜表面的高密度 ECoG 植入物可能适用于长期而直接的语言神经假体应用。

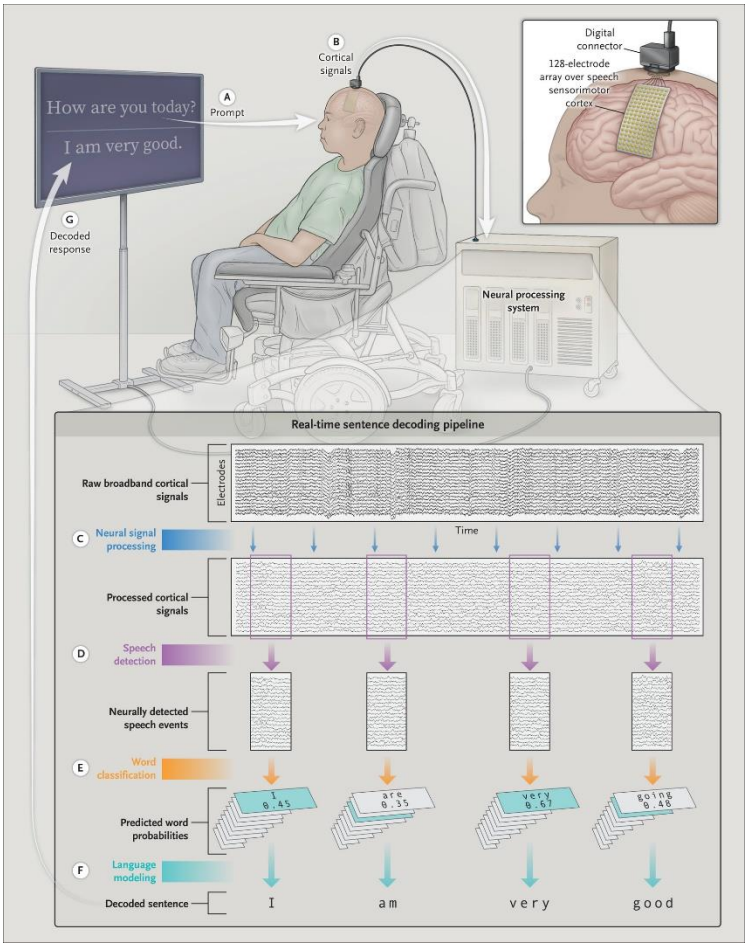


图 4.4 语言脑机接口系统框图[20]

（3）颅内神经记录

脑机接口的发展带动颅内植入术、神经信号采集及解码技术突飞猛进，使得一些原本属于认知科学领域难以在人体上得以验证的科学问题寻得了合适的技术手段而得以突破和解决。其中，颅内立体脑电（stereo-electroencephalography，sEEG）采集技术可以在人类活体大脑中获取最为直接的神经电信号证据，用于探索和揭秘人类认知行为对应的神经机制。

通过视觉线索回忆学习到的时间序列是基于经验的神经可塑性的一种重要形式。2021 年，北京大学方方教授团队及合作者利用 sEEG 技术在清醒的人类视觉皮层中观察到了这种重新激活模式[21]。如图 4.6 示，通过视网膜拓扑定位技术，研究者首先定位了早期视皮层 V1-V3 内各电极位点的视觉空间感受野。实验分为前测、学习和后测三个阶段：在学习阶段，被试反复观看屏幕上快速移动的光点；在前后测试阶段，在运动轨迹的不同位置短暂呈现光点，记录感受野位于运动轨迹的电极信号，考察光点能否诱发运动序列的重放。对神经信号分析发现，基于 Gamma 信号的互相关分析及对视觉诱发电位的分析均显示，斑点的闪光能够触发沿着运动路径的下游感受野的神经活动回放，并且这种影响是只有当线索出现在感受野附近时才会观察到。对这种学习效应的时间特征的进一步分析发现，回放的速度比光点在物理移动时的速度要快，这表明回放过程中存在时间压缩。此外，该回放效应会随着时间的推移快速衰减。在此之前，科学家仅在啮齿动物等非人类动物的感知皮层观察到这一过程，而颅内脑电采集技术为人类早期视皮层在线索诱发下的回放效应提供了直接的电生理证据。

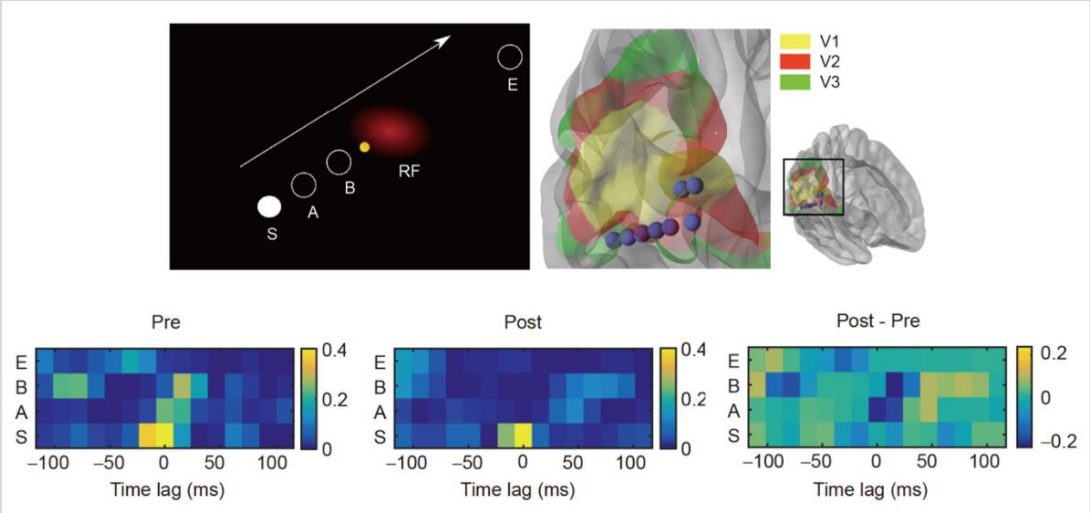


图 4.5 实验刺激与立体颅内脑电记录位点（上）
基于 Gamma 信号的互相关图显示回放效应（下）[21]

4.4.2 上行脑机接口

下行脑机接口主要将大脑神经信号解码来控制外部设备即“脑控”，而从外部对大脑施加刺激实现“控脑”对脑机接口的实际应用也非常重要。打通信息双向交互的链路，有利于实现带有反馈的闭环控制系统，同时上行脑机接口技术可

应用于恢复某些患者的大脑功能或增强正常人的大脑效能等，下面我们以触觉反馈和视皮层假体为例简述其发展现状。

（1）触觉反馈

运动控制脑机接口需要解码大脑的运动指令，而触觉反馈脑机接口则需要将外部传感器的触觉传感信息编码成大脑可以感知和理解电刺激信号。相比于运动脑机接口，感觉脑机接口的研究相对较少，早期针对猕猴的研究主要集中探索能诱发产生空间位置、力的大小等感觉的有效电刺激范式。2011 年，杜克大学的研究团队在 Nature 上发表文章表明在虚拟手臂接触不同物体的时候施加不同的微电刺激到感觉皮层可以使猕猴产生不同的虚拟触觉[22]；2013 年芝加哥大学的研究团队在 PNAS 上发文表明，通过改变皮层微电刺激的位置及频率幅值能够使猕猴产生不同手部位置不同力的大小的虚拟触觉[23]；2015 年美国加州大学 Sabes 团队在 Nature Neuroscience 上发文表明，通过不同的皮层微电流刺激，可以使猕猴感知虚拟物体的空间位置以及与其手部的距离[24]。

随着研究人员在猕猴实验上的理论和技术积累，应用于人体触觉感觉重建的脑机接口也取得了亮眼的突破。2016 年，研究人员在瘫痪病人的大脑中负责处理运动功能和手指、手掌感觉的区域植入了四个微电极阵列，通过电刺激使病人通过机械手获得了非常接近自然的触觉感知[25]。到 2020 年，美国巴泰尔科研中心 Patrick Ganzer 团队报告的双向脑机接口技术可采集和识别残存的、低于知觉反应范围的触觉信号，并结合颅内微电流刺激技术将这些触觉信号加强到可被感知到的强度，实现对受试者触觉的恢复[19]。

（2）视皮层脑机接口

近年来，科学家们通过开发视皮层假体（visual cortical prothesis，VCP），在利用电刺激视觉皮层帮助视觉障碍者恢复基础的视觉能力方面，取得了系列进展。微电极是视皮层假体的重要组成部件，随着微纳加工制造领域的技术革新，具有更高通道的微电极阵列或具有更高柔韧性的薄膜电极阵列逐渐被应用在 VCP 的设计中。

2020 年，Pieter R. Roelfsema 等借助了在生物兼容性和电学特性上表现优异的犹他电极阵列，实现了可调控通道数上千的视皮层假体，通过电刺激成功使

猴子能“看到”移动的点、线条以及字母[26]。在此基础上，2021 年由 Roelfsema 教授与西班牙米格尔·埃尔南德斯·德埃尔切大学（Miguel Hernández University）细胞生物学教授 Eduardo Fernández Jover 合作的一项研究，进一步将 96 通道犹他电极阵列植入到失明患者右侧枕叶皮层靠近 V1 和 V2 的边界处，尝试以优化后的皮层内微弱电刺激恢复失明受试者的视觉功能[27]。先前开发的皮层视觉假体一般将电极植入涉及所有视觉信息的初始处理的纹状皮层(V1)。然而，V1 区域被埋在很难到达的距状沟中，且包含许多需要避开的血管。考虑到先前对猴子和人类的研究结论：V1 和 V2 区域的光幻视阈值相似，并且引发的感知质量具有可比性。Eduardo 教授研究组最终确定在受试者的纹外区域植入电极。在确保受试者能区分电刺激诱发的和自发的光幻视后，研究人员确定了最佳电刺激参数，并研究了单点电刺激与多点电刺激的光幻视阈值差异。此外，该研究还通过同步电刺激和神经活动采集实验探索了电刺激与神经元活动的关系，通过设置不同的刺激位点研究了电极间距、刺激时间间隔等因素对幻视尺寸、幻视感知灵敏度等的影响，最后通过最高 16 个通道的多点电刺激成功让受试者识别一些字母和物体边缘，如图 4.7 所示。这项研究是高密度犹他电极构成的脑机接口在人类视觉皮层的第一例应用，证明了慢性皮层微刺激的安全性和有效性，显示了其在恢复盲人功能性视力方面的巨大潜力。

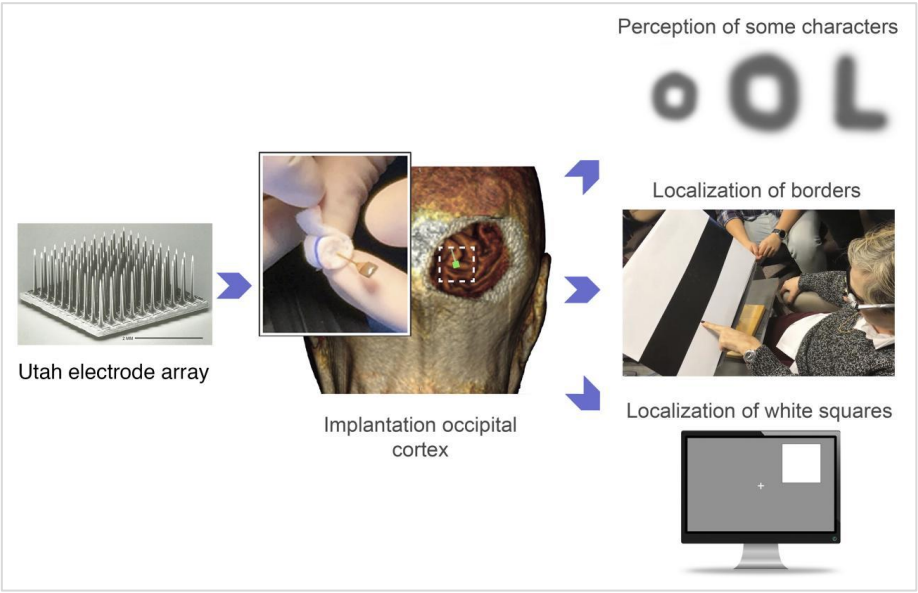


图 4.6 视皮层假体示意图[27]

4.4.3 未来展望

伴随着实际应用提出的便携易操作需求，临床大脑皮层假体将使用具有稳健功能、无线驱动的电极阵列。随着微纳加工及封装技术、集成芯片技术和无线传输技术的进一步发展，使用者将不必时刻与体积庞大、连线冗杂的大型外部计算直接互连，从而实现实用性更高的便携式脑机接口。

在硬件系统的支撑下，研究安全有效的电刺激范式是用于恢复感知觉的脑机接口亟需解决的问题。回顾 2020 年 Daniel Yoshor 等开发的动态电刺激技术，该技术通过控制各通道电流的时间顺序，匹配目标形状，并传递到电极阵列，从而产生连贯的视觉感知，使参与者 “看到”电刺激所示踪的形状，并在触摸屏上准确地“打印”出这些形状[28]。这种创新的电刺激范式值得进一步深入探讨其迁移到颅内微电极电刺激层级的可能性，以及对听觉皮层假体等针对其他脑区的脑机接口应用的启发。

4.5 总结与展望

2020 年，慕尼黑工业大学的 Gordon Cheng 等人指出，神经科学、机器人学与人工智能的融合提出了一种新的挑战——神经工程学（Neuroengineering），精细神经解码、机器人到脑的反馈、双向脑-机适应和软体/混合结构机器人等技术是实现脑-机协同的关键[29]。高带宽的脑机接口技术和更加精细的神经解码技术有望让瘫痪或残障患者获得更多自由，例如文本交流、玩电子游戏、艺术创作，甚至可以通过读取大脑信号并相应的刺激肌肉神经，使患者在一些程度上恢复运动能力。进一步来讲，构建触觉或视听觉的神经假体将触觉或视听觉信息直接反馈给人类的大脑，将控制和反馈形成一个闭环，可以有效提高人类对假肢的反应式控制能力或外界的感知能力。基于柔性微电极的植入式脑机接口技术在安全性和成本上达到可以推广的程度后，植入式脑机接口有望成为普通人所必备的设备，增强人类的信息获取与处理能力，并与虚拟世界、机器人、智能手机等外界平台有着无缝化的交互与协同，实现人脑与 AI 的混合智能。

参考文献

[1] Zhang M, Tang Z, Liu X, et al. Electronic neural interfaces[J]. Nature Electronics, 2020, 3(4): 191-200.

[2] Yoon D Y, Pinto S, Chung S W, et al. A 1024-channel simultaneous recording neural SoC with stimulation and real-time spike detection[C]//2021 Symposium on VLSI Circuits. IEEE, 2021: 1-2.

[3] Nason, S.R., Vaskov, A.K., Willsey, M.S. *et al.* A low-power band of neuronal spiking activity dominated by local single units improves the performance of brain–machine interfaces. *Nat Biomed Eng* **4,** 973–983 (2020).

[4] Even-Chen, N., Muratore, D.G., Stavisky, S.D. *et al.* Power-saving design opportunities for wireless intracortical brain–computer interfaces. *Nat Biomed Eng* **4,** 984–996 (2020).

[5] Won, S.M., Cai, L., Gutruf, P. et al. Wireless and battery-free technologies for neuroengineering. Nat Biomed Eng (2021). https://doi.org/10.1038/s41551-021-00683-3.

[6] C. Han, J. Mao, X. Wang, S. Yu and Z. Zhang, "Body Channel Based Wireless Power Transfer Method for Implantable Bioelectronics," 2021 IEEE International Symposium on Circuits and Systems (ISCAS), 2021.

[7] Li, J., Dong, Y., Park, J.H. et al. Body-coupled power transmission and energy harvesting. Nat Electron 4, 530–538 (2021). https://doi.org/10.1038/s41928-021-00592-y.

[8] Y. Wang, G. Liang, F. Liu, Q. Chen and L. Xi, "A Long-Term Cranial Window for High-Resolution Photoacoustic Imaging," IEEE Transactions on Biomedical Engineering, vol. 68, no. 2, pp. 706-711, 2021.

[9] P. Aebischer, S. Meyer, M. Caversaccio and W. Wimmer, "Intraoperative Impedance-Based Estimation of Cochlear Implant Electrode Array Insertion Depth," IEEE Transactions on Biomedical Engineering, vol. 68, no. 2, pp. 545-555, 2021.

[10] C. Li, X. Fan, J. Hong, D. W. Roberts, J. P. Aronson and K. D. Paulsen, "Model-Based Image Updating for Brain Shift in Deep Brain Stimulation Electrode Placement Surgery," IEEE Transactions on Biomedical Engineering, vol. 67, no. 12, pp. 3542-3552, 2020.

[11] K. M Boergens, A. Tadić, M. S Hopper, et al., Laser ablation of the pia mater for insertion of high-density microelectrode arrays in a translational sheep model, Journal of Neural Engineering, vol. 18, no.4, 2021.

[12] Higueras-Esteban, et al., "Projection-Based Collision Detection Algorithm for Stereoelectroencephalography Electrode Risk Assessment and Re-Planning," in IEEE Access, vol. 9, pp. 105180-105191, 2021.

[13] L. H. Timothy, C. A. Diaz-Botia, V. Kharazia, et al., The “sewing machine” for minimally invasive neural recording, BioRxiv, 2019.

[14] E. Musk and Neuralink. “An integrated brain-machine interface platform with

thousands of channels.” Journal of Medical Internet Research, vol. 21, no. 10, 2019.
[15] https://neuralink.com/
[16] https://techcrunch.com/2020/08/28/take-a-closer-look-at-elon-musks-neuralink-surgical-robot/
[17] Aflalo, Tyson, et al. "Decoding motor imagery from the posterior parietal cortex of a tetraplegic human." Science 348.6237 (2015): 906-910.
[18] Pandarinath, Chethan, et al. "High performance communication by people with paralysis using an intracortical brain-computer interface." Elife 6 (2017): e18554.
[19] Ganzer, Patrick D., et al. "Restoring the sense of touch using a sensorimotor demultiplexing neural interface." Cell 181.4 (2020): 763-773.
[20] Moses, David A., et al. "Neuroprosthesis for decoding speech in a paralyzed person with anarthria." New England Journal of Medicine 385.3 (2021): 217-227.
[21] Lu, Junshi, et al. "Cue-triggered activity replay in human early visual cortex." Science China Life Sciences 64.1 (2021): 144-151.
[22] O’Doherty, Joseph E., et al. "Active tactile exploration using a brain–machine–brain interface." Nature 479.7372 (2011): 228-231.
[23] Tabot, Gregg A., et al. "Restoring the sense of touch with a prosthetic hand through a brain interface." Proceedings of the National Academy of Sciences 110.45 (2013): 18279-18284.
[24] Dadarlat, Maria C., Joseph E. O'doherty, and Philip N. Sabes. "A learning-based approach to artificial sensory feedback leads to optimal integration." Nature neuroscience 18.1 (2015): 138-144.
[25] Flesher, Sharlene N., et al. "Intracortical microstimulation of human somatosensory cortex." Science translational medicine 8.361 (2016): 361ra141-361ra141.
[26] Chen, Xing, et al. "Shape perception via a high-channel-count neuroprosthesis in monkey visual cortex." Science 370.6521 (2020): 1191-1196.
[27] Fernández, Eduardo, et al. "Visual percepts evoked with an intracortical 96-channel microelectrode array inserted in human occipital cortex." The Journal of Clinical Investigation 131.23 (2021).
[28] Beauchamp, Michael S., et al. "Dynamic stimulation of visual cortex produces form vision in sighted and blind humans." Cell 181.4 (2020): 774-783.
[29] G. Cheng, S. K. Ehrlich, M. A. L. Nicolelis, Neuroengineering challenges of fusing robotics and neuroscience. Sci. Robot. 5, eabd1911(2020)

第5章　交叉学科技术进展

技术进步推动科研发展，在神经科学领域亦是如此。我们想要了解神经系统，不仅需要研究神经系统的组织结构、设计原理，还需要探索神经系统实现各种功能的工作机制。这些信息的获取均离不开各项技术的支持。每一项新技术的产生或突破，带来的不仅是研究方法上的进步，还有分析视角上的变革，以及全新的数据带来的新发现。比如，超分辨显微成像技术的发展，让研究人员看到了以往观察不到生命细节；多组分标记技术的发展，推动人们理解复杂系统中各组分的相互作用机制；电镜成像技术的发展，推动人们对神经系统设计原则的理解，探索功能行使的组织结构基础；随着人工智能的发展，新算法的开发与应用，让以往无法分析的大数据被高速化处理，研究人员可从海量的数据中提取更丰富的信息。2021 年是一个技术进展的丰收年，本章节将从成像、连接组学、大数据处理等方面，来简析这些进展为神经科学发展带来的新见解和可能性。

5.1 高精度高信息量的数据获取方法

探索神经系统的工作原理需要数据作为基础，因此获取高精度（如，高时间分辨率、高空间分辨率）、高信息量（如，跨脑区、三维空间、多组分、多水平、复杂网络等）的数据是近年来神经科学相关技术的发展方向。高精度数据的获取，能够帮助研究人员尽可能真实的还原神经系统的设计原理、动力学变化过程，引导研究者从更精细的角度探索神经系统的工作机制。高信息量数据则能够推动研究者更好地理解复杂系统的相互作用机制，从更全面的角度提取复杂网络信息。

5.1.1 稀疏解卷积通过计算提高成像分辨率

2014 年诺贝尔化学奖被授予了荧光超分辨显微技术，该技术利用荧光分子的化学开关特性（PALM/FPALM/STORM）或者物理的直接受激辐射现象（STED），实现超越衍射极限的超分辨成像。尽管如此，活细胞中的超分辨率成像仍然存在两个主要瓶颈：（1）超分辨率的光毒性限制了观察活细胞中精细生理过程；（2）受限于荧光分子单位时间内发出的光子数，时间和空间分辨率不可兼得。

受限于这个瓶颈，为了在活细胞上达到 60 nm 空间分辨率极限，现有超分辨率成像手段需要强照明功率（kW~MW/mm^2）、特殊荧光探针和长曝光时间（> 2 s）。强照明功率引起的强漂白会破坏真实荧光结构的完整性，长曝光时间在图像重构时导致运动伪影，降低有效分辨率。因此迄今为止，基于光学硬件或者荧光探针的改进均无法进一步提升活细胞超分辨率的时空分辨率。

2021 年 11 月，北京大学陈良怡教授团队与哈尔滨工业大学李浩宇教授团队合作发表 Nature Biotechnology 封面文章，提出了稀疏解卷积（Sparse deconvolution）方法[1]。稀疏解卷积是基于“荧光图像的分辨率提高等价于图像的相对稀疏性增加”这个通用先验知识，结合陈良怡教授团队在 2018 年提出的信号空时连续性先验知识[2]而发明的两步迭代解卷积算法。该方法突破硬件限制，首次实现通用计算荧光超分辨率成像，可提供目前活细胞光学成像中超高空间分辨率（60nm）下，速度更快（564Hz）、成像时间更长（>1h）的超分辨成像。

通过稀疏解卷积算法来实现荧光超分辨率成像，从算法层面打破了基于光学硬件或者荧光探针的改进无法进一步提升活细胞时空分辨率的僵局，与目前基于特定物理原理或者特殊荧光探针的超分辨率方法都不相同。且该方法可以与现有的多数商业荧光显微镜结合，有效提升它们的空间分辨率（见图 5.2

图5.3）。该算法在多色成像（图 5.1）和活体成像（图 5.4）中也表现出非常好的效果。

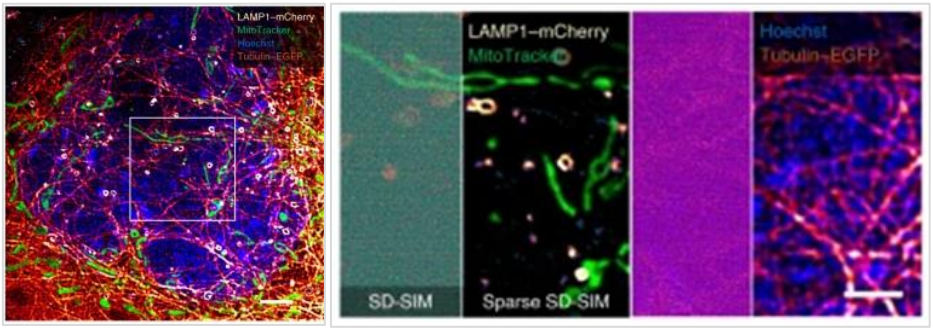


图 5.1 稀疏 SD-SIM 进行四色超分辨成像[1]左：对溶酶体、线粒体、微管和细胞核进行四色(LAMP1-mCherry，黄色；MitoTracker，绿色；Hoechst，蓝色；Tubulin-EGFP，棕色)活细胞 SR 成像。中、右：左图白色框区域的放大视图。比例尺：左：5μm；中、右 3μm。

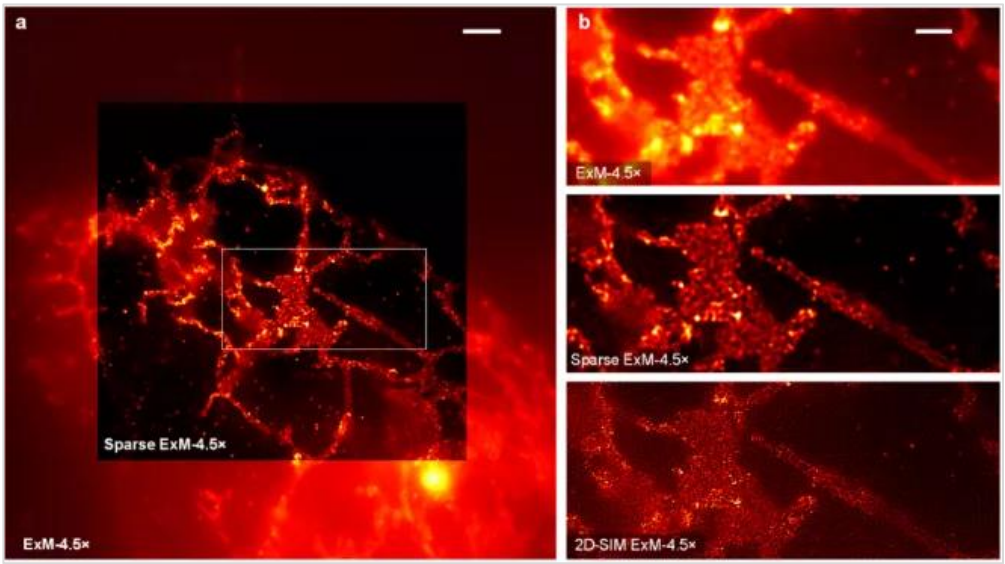


图 5.2 Sparse ExM 解析膨胀细胞内质网[1]（a）Sec61β-GFP 在 COS-7 细胞中的 ExM 图像。背景为膨胀 4.5 倍的细胞（ExM-4.5×），中心为稀疏解卷积重建后的 Sparse ExM-4.5×图像。（b）ExM-4.5×、Sparse ExM-4.5×和 2D-SIM ExM-4.5×下白色框包围的放大区域。比例尺：（a）1μm；（b）500nm。

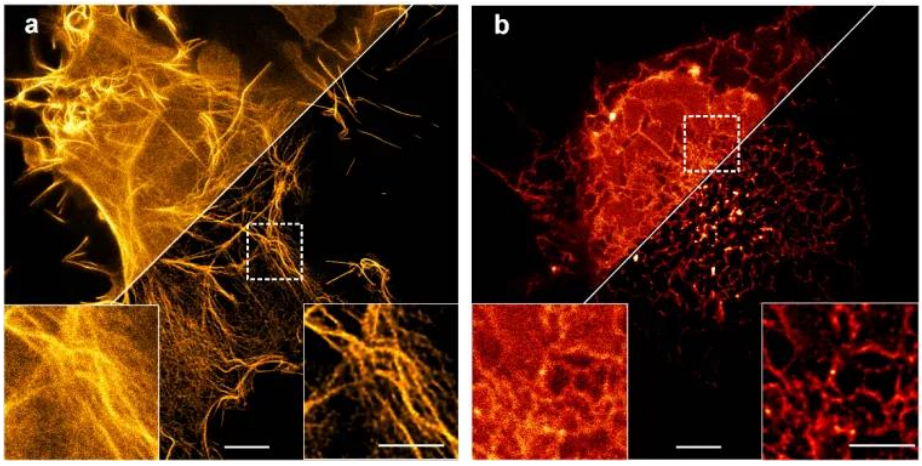


图 5.3 利用稀疏解卷积提升 STED 显微镜的空间分辨率[1]（a，b）在 STED 显微镜下直接（左上）或经过稀疏解卷积(右下)观察到表达 Lifeact-GFP（a）或 Sec61β-GFP（b）的 COS-7 活细胞。图在下方显示了在 STED(左下插图)下和稀疏反褶积(右下插图)后白色框区域的放大视图。比例尺：（a，b）5 μm；放大视图区域 2 μm。

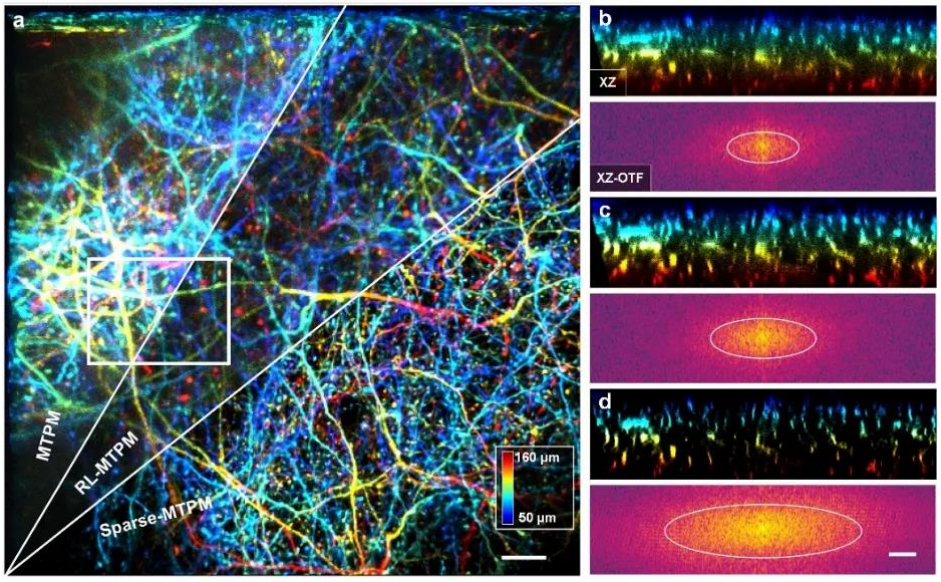


图 5.4 应用微型化双光子显微成像（Sparse-MTPM）活体三维成像转基因小鼠视皮层 Thy1-GFP 标记的神经元树突棘，深度为 50~160 μm。稀疏解卷积处理后可观察到显著的频谱延展，强于传统的 RL（Richardson-Lucy）解卷积（RL-MTPM），更清楚观察到树突棘三维结构[1]。

稀疏解卷积方法取得了如下突破和创新：1）解决了 Richardson–Lucy 反卷积应用到生物成像中的噪声和先验知识问题，拓展了它在生物成像中的实际应用；2）利用稀疏结构光超分辨成像在活细胞中实现了同时高时空分辨率长时程成像；

3）方法具有普适性，可以广泛用于宽场成像和其它超分辨成像技术，提高这些成像方法的分辨率。

5.1.2 多色成像揭示系统全景组分

多色成像是指使用多种荧光染料来检查一个样本中的不同元素，以实现同时观察多种组分的分布与相互作用过程。生物系统是一个复杂的系统，神经系统又最为复杂。在中枢神经系统中，多种高度分化的细胞类型混合并存，仅人类的前额叶中便包含至少 21 种细胞亚型[3]。并且单个神经元的轴突向多个方向延伸并长距离投射构成复杂的网络[4]。因此，除了获得高时空分辨率的神经活动信息外，在神经系统的原有三维结构中同时标记、分析多种组分对于理解复杂的神经系统具有同样重要的意义。

在现有技术方法中，这一目标可通过荧光成像、质谱成像和循环免疫荧光等方法来实现。所谓荧光成像（Fluorescence imaging），是指通过荧光标记试剂或荧光抗体对无法直接观察的无色透明的细胞器或特定蛋白质进行荧光标记从而实现可视化的手法。通过荧光成像能够对细胞或蛋白质的形态或结构，以及生命活动进行观察与分析。而质谱成像是通过质谱直接扫描生物样品成像，可以在同一张组织切片或组织芯片上同时分析数百种分子的空间分布特征。该方法对不同组分的分析主要是通过测定生物分子的质荷比来实现的。但是以上这些技术都受限于成像颜色与样本深度，难以在较大尺度上对生物样本进行三维多色成像。比如受限于荧光光谱太宽，传统荧光成像方法通常不能同时分析超过 5 种组分[5, 6]；而质谱成像与循环免疫荧光只适用于薄样本，需要应用连续切片才能成像较大尺度样本。

2021 年 10 月，哥伦比亚大学的闵玮团队公布了他们最新开发的三维多色成像方法[7]。在先前的研究中，该团队利用组织透明化技术助力拉曼散射显微镜实现厚样本成像。而在最新的研究中，他们进一步设计并拓展了能够在拉曼静默区（1800-2800cm^{-1}）成像的 MARS 染料，并优化了适用于 MARS 染料的组织透明化技术 Raman DISCO （rDISCO）。将二者结合，融合拉曼显微成像技术，首次实现了在毫米厚度的脑组织切片中一次性成像 11 种分子组分。这一技术的开发

不仅把多色成像的深度提升了 10 到 100 倍，更是打破了荧光成像的颜色壁垒，使得更多的组分信息能够被同时获得，为研究人员理解神经系统中各组份在三维空间中的分布与相互作用，以及描绘组分间的网络连接起到了技术上的推动。利用这一技术可以同时标记多种组分的特点，研究人员发现了波形蛋白（vimentin）与胶质纤维酸性蛋白（GFAP）在星形胶质细胞发育过程中的转换过程，即波形蛋白在快速髓鞘化阶段逐渐被 GFAP 替代。因为该技术在进行多色成像的同时完好的保留了样本原本的三维结构，因此可测量处于不同分化阶段的星形胶质细胞到其最近的血管的距离，这也是反应中枢神经系统发育与组织情况的关键参数。最后，研究者还借助成像获得的信息，在小鼠小脑样本的 3D 空间中绘制了蛋白质相关网络图，这将有助于人们了解复杂组织结构的组织原则，对探索复杂生物系统的底层架构和设计原则具有一定的帮助。

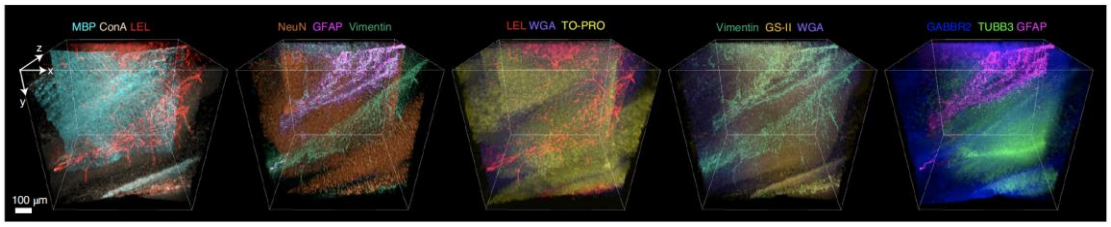


图 5.5 在厚度为 1mm 的小鼠小脑切片中进行 11 种组分的体成像。摘自[7]

5.1.3 脑连接组反应组织设计原则

了解神经系统完整且精细的组织结构，挖掘神经网络的物理架构，是科研人员长久以来的追求。一些研究者认为，大脑的终极奥秘就藏在这份精细的“大脑地图”中。目前为止，通过电镜成像技术绘制脑图谱是获取最为全面连接组信息的方法。随着电镜技术的日益发展，人们已经获得包括线虫、果蝇、斑马鱼、小鼠以及人在内的或全脑或局部脑组织的连接组。人工智能技术也从图像分割、自动重建以及系统性分析等层面对大规模脑连接组的获得与信息提取提供了帮助。

电镜成像的超高分辨率可以解析脑组织中每一个神经元的空间分布、形态，以及轴突、树突的数量、分布与走向；更进一步，神经元的每一个突触也能被精确的定位、计量，并且统计出这些突触分别与哪些神经元相连接；一些神经元中的细胞器数量等信息也能够被清楚观察到。这就好比将一台性能精良的计算机的

设计图纸详细且清晰的呈现在我们面前。计算机行使功能离不开硬件的支持，同样神经系统对内部、外部信息进行加工，并促使生物最终作出相应的行为，这些过程依赖于其物质结构基础。因此研究人员试图获得脑连接组的初衷在于破解生物智能产生的物质结构基础，人们想知道是怎样的组织设计原则涌现了生物智能。

自从 1986 年研究人员首次获得线虫全脑接组以来[8]，脑连接组的获取仍然是一项耗时耗力的工作。并且科学家们对连接组学的争议也持续已久，其中最著名的便是“Top ten arguments against connectomics”（连接组学十大争议）[9]。

“连接组学十大争议”

Top Ten Arguments against Connectomics

十、神经回路的结构不等同于功能

Number ten: circuit structure is different from circuit function

九、存在没有突触的信号和没有信号的突触，即信号实际连接和连接结构并不完全等同

Number nine: signals without synapses and synapses without signals

八、存在“垃圾”突触

Number eight: ‘junk’ synapses

七、相同结构可能对应很多功能

Number seven: same structure, many functions

六、相同功能可以基于很多结构

Number six: same function, many structures

五、对现象的统计可能更本质

Number five: statistical synapses should suffice

四、脑结构复杂度远超意志本身

Number four: the mind is no match for the complexity of the brain

三、连接组只是更加昂贵的描述性的神经解剖学

Number three: merely descriptive neuroanatomy, just more expensive

二、连接组并不能帮助我们得到很多

Number two: not much was learned from the connectomes we have

一、静态信息并不意味着工作机制

Number one: it’s a static picture, this is connected to that, but you wouldn’t know how it worked

在这 10 项争议中，人们主要质疑“结构”并不等于“功能”，在二者之间仍然有许多不确定的因素，影响研究者直接从结构信息中提取功能产生的机制。而支持连接组学的科学家们也认可这些争议的存在，但是他们同样认为脑连接组的获

得，尤其是多样本量的脑图谱的绘制，对于人们破解大脑的奥秘是十分重要且不可规避的。正如资深的连接组学研究者 Jeff Lichtman 曾表示过：“我们之所以对连接组孜孜以求，是因为我们认为它是获取神经系统部分基础信息的唯一途径”[10]。

近年来，随着电镜成像技术与人工智能的发展，大规模、快速获取脑连接组的方法日趋完善。大量项目的开展与数据积累，为神经科学的探索提供了许多新视角与新发现。2020 年 1 月，来自霍华德·休斯医学研究所珍妮莉亚研究园区（Janelia Research Campus, Howard Hughes Medical Institute）的研究团队发布了达到突触级别的果蝇一半脑的连接组“Hemibrain”[11]。在这份数据中，神经科学家发现了数十种可能与飞行导航相关的新神经元与环路。这项工作被誉为揭示果蝇如何整合感觉信息并将其转化为行为的重要里程碑。2021 年 5 月，哈佛大学和谷歌团队发布了首份人脑连接组数据[12]，他们通过高速扫描电子显微镜成像完成了 1 立方毫米人类颞叶组织的连接组图谱。从这 1.4PB 的数据中，研究人员分割、重构了 57216 个细胞、数亿个神经突和 1.337 亿个突触连接。在这海量的数据中，他们发现了诸如轴突卷曲并相互盘旋的神经元，以及具有两个轴突的神经元等在其他动物中从未见过的新型细胞。正是如此精细的电镜数据支持研究者发现了以往无法观察到的新现象，并向人们展示了大脑的复杂度以及对脑功能探索的很长的路。但在这些新发现背后人们也存在一些疑虑——“n of 1”问题。样本源自单一的独立个体，这个人特定的基因构成，独特的人生经历，使得样本具有非常高的特异性，人们很难判断在该样本中发现的现象与规律是专属于这一人的独特现象，还是足以代表千千万万人类的普遍现象。因此能够大规模获得多个样本的连接组便显得重要起来。

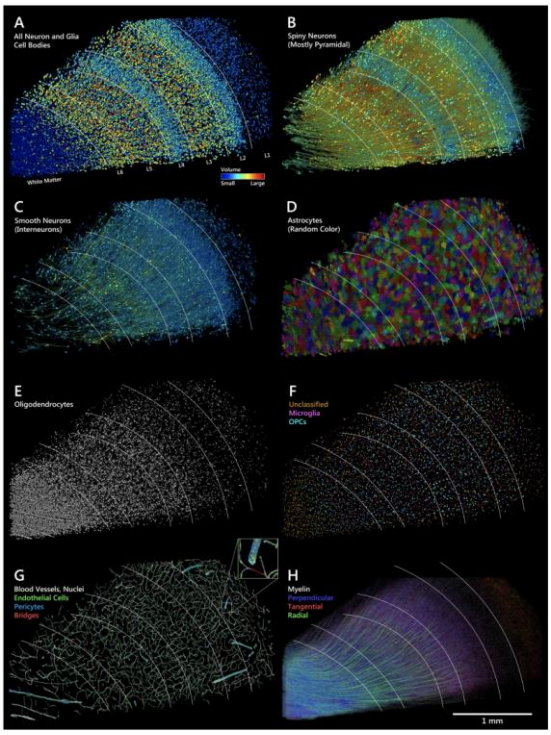


图 5. 6 人脑样本中细胞、血管、髓鞘的分布[12]

目前为止能够进行大规模连接组获取的物种只有线虫，现有的技术已经发展到在大约一个月内能完成一只秀丽隐杆线虫的脑图谱绘制，利用这一模式生物开展的研究也证明了大规模连接组的力量。2021 年 9 月，来自哈佛大学的 Aravinthan D.T. Samuel 团队发表在 Cell 杂志上的工作向我们展示了如何利用连接组来预测动物行为[13]。研究人员同时记录了线虫在交配过程中的行为学与全脑范围内的神经影像学（钙成像）数据，并绘制了多条线虫的脑连接图谱。通过光电联合技术将神经系统的功能活动映射到连接组中可获得线虫的脑活动图谱，并从中分析确定线虫在交配过程中加工环境信息的神经机制。研究人员发现，神经元间的相关性不是固定不变的，而是与线虫当下的行为状态有关，并且随着行为的展开，特定功能神经元的动力学变化也随之呈现。通过比较 8 条线虫的脑活动图谱，人们发现它们的功能特性是如此的明显且一致，甚至可以用来预测第九条线虫的行为。事实上，当研究者精确的损毁了一个参与“转动”的神经元后，该线虫（新的样本）确实失去了转动的能力。-

另外，来自多伦多西奈山医院和哈佛大学的研究者发表在 2021 年 8 月 Nature 上的工作比较了 8 条具有相同基因的秀丽隐杆线虫的连接组[14]，这些线

虫分别处于幼虫到成虫的不同阶段。研究人员发现尽管这些线虫具有相同的遗传基因，但它们的大脑中仍有多达 40%的神经元连接是不相同的，并且这些具有个体差异的连接与那些在每个个体中都存在的相似连接相比，它们在连接度上更弱，而那些更强的连接（拥有 100 个或更多的突触连接）在个体间呈现了稳定的一致性。针对这一现象，Lichtman 认为脑连接可能包含两种类型，一类是可变连接，一类是保守连接。如果事后能够证明生物通过这些保守连接来支持生存所必需的神经活动，那么可变连接的变化水平则可成为评估连接组的一个重要指标。

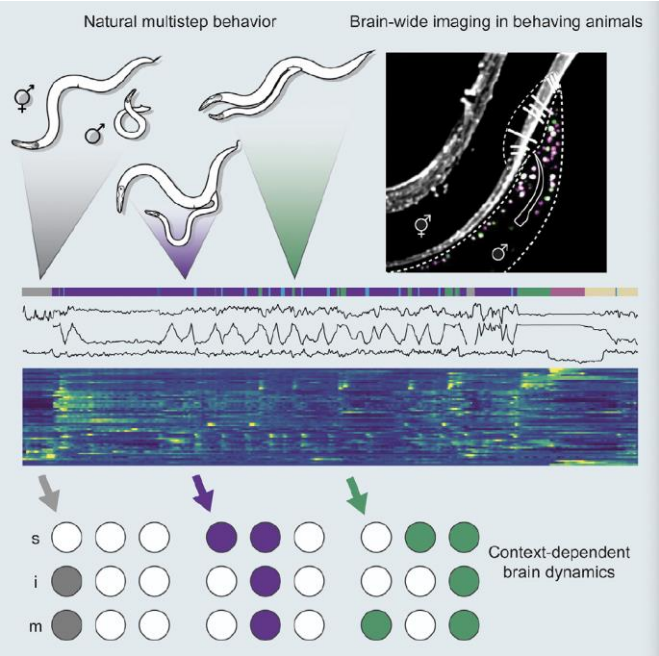


图 5.7 线虫的全脑成像显示神经元具有功能特异性，且神经元间的神经活动相关性不是固定不变的，而是依据行为而改变[13]。

可见，当连接组信息积累到一定水平时，一些与功能形成相关的重要机制就可能会从随机的海洋中涌现出来。如今脑连接组正在成为一项资源，人们希望将连接组学作为一种工具和数据库来研究神经环路的工作机制。也许有一天，当人们将大量的连接组数据汇聚在一起，结合更为丰富的功能信息，便能够实现脑连接组学最初的目标——破解生物智能的奥秘。

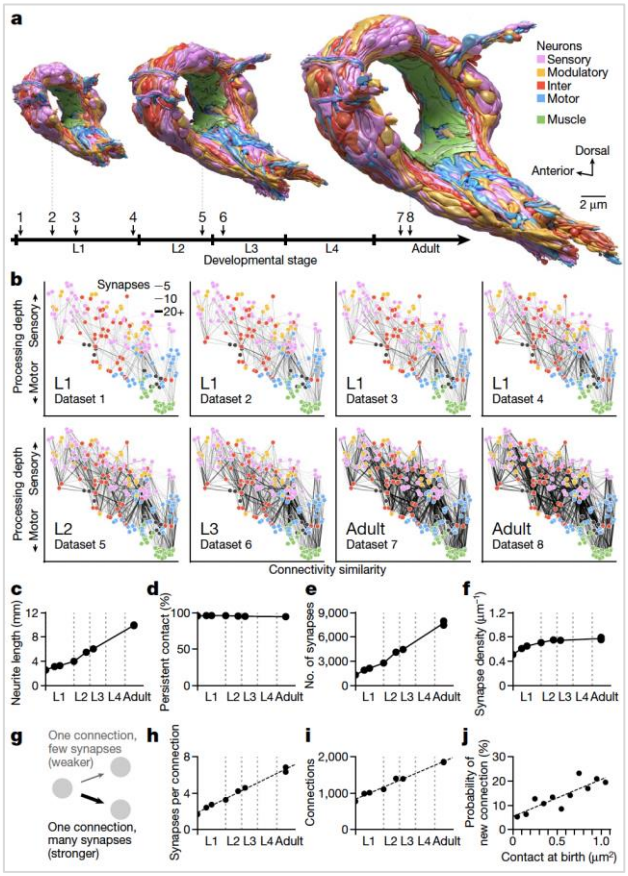


图 5.8 发育中的大脑突触与连接数量会不断增多但整体的几何结构保持不变。摘自[13]

5.2 智能化数据处理手段

生命科学已经进入大数据时代，且在技术的不断推动下，生命科学领域尤其脑科学领域的高质量数据正在不断积累。图像处理和以组学为代表的生物大数据分析领域尤甚。这两个领域另一个重要的特点是，尽管数据中蕴藏着巨大的信息量，数据本身需要精细准确的提取与正确的统计才能保证结论的可靠性。随着信息量大，以人工处理为主传统分析方法已逐渐力不从心。而以深度神经网络为代表的机器学习领域正不断涌现出新的数据处理与分析方法。下面我们将介绍过去一年中出现的具有代表性的新方法。

5.2.1 更智能的图像数据处理

了解生物组织的空间结构对于基础研究和转化研究都至关重要。虽然组织成像技术进展突飞猛进，但解释它们产生的图像等数据是一项重大的计算挑战。目

前尚不存在促进对这些数据集进行大规模分析和解释的工具。最明显的例子是缺乏高通量、高准确性的通用图像细胞分割算法——组织成像使用的是完整的样本，为了从图像中提取单细胞信息，必须在图像采集后正确分配每个像素所属的细胞，这个过程是下游分析（如细胞类型识别和组织邻域分析）的基础，因此对整个图像数据分析具有决定性意义。

2021 年 11 月，美国加州帕萨迪纳加州理工学院生物与生物工程系的 David Van Vale 与斯坦福大学病理学系的 Michael Angelo 通过大规模的数据注释和深度学习解决了组织成像数据中的细胞分割问题[15]。为解决大规模数据注释问题，他们搭建了 TissueNet 图像数据集，这是一个包含了来自九个器官和六个成像平台的 100 万对成对标记的全细胞和细胞核组织图像。基于 TissueNet 上的数据，他们设计了完整的人工标记与模型训练流程（详见图 5.9），并开发了一种使用深度学习的分割算法 Mesmer，实现了组织成像数据中细胞核分割和全细胞分割。Mesmer 比以前的方法具有更好的速度和准确性，更为重要的是 Mesmer 能适应 TissueNet 中所有组织类型和成像平台的多种类型图像数据，并在全细胞分割方面达到人类水平的性能。同时 Mesmer 实现了关键细胞特征的自动提取，例如蛋白质信号的亚细胞定位。当前，所有底层代码和模型都已许可作为社区资源发布。

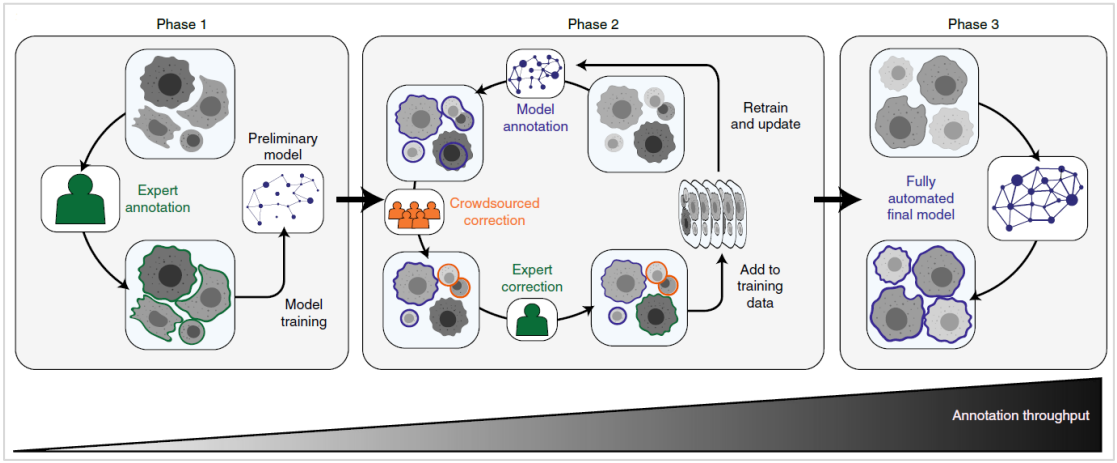


图 5.9　模型训练分为三个阶段。第 1 阶段，人工创建注释来训练初级模型。第 2 阶段，用初级模型对新数据做预测，随后由人工校正。模型在此过程中同步进化。这样在减少了错误率的同时提高了新数据的注释速度。第 3 阶段，无需人工校正即可运行准确的模型[15]。

深度神经网络同样提升了神经科学图像处理能力。在神经科学研究中，钙成像是一种以单细胞分辨率监测神经回路活动的方法。然而，钙成像本质上容易受

到检测噪声的影响，在高帧率或低激发光剂量成像时尤甚。2021 年 8 月，清华大学脑与认知科学研究院、自动化系戴琼海课题组在 Nature Methods 发表研究论文介绍了他们开发的 DeepCAD 算法[16]。此算法的独特之处在于这是一种自监督的深度学习算法。该算法不需要任何高信噪比 (SNR) 观测，仅使用单个低信噪比的钙成像视频序列即可实现网络训练（见图 5.10）。DeepCAD 可以抑制检测噪声，将信噪比提高十倍以上，增强了神经元提取和神经脉冲推断的准确性，促进了神经回路的功能分析。利用这种方法，研究者可以通过降低对图像质量的要求获得其他方向的（如时空分辨率）更多信息。

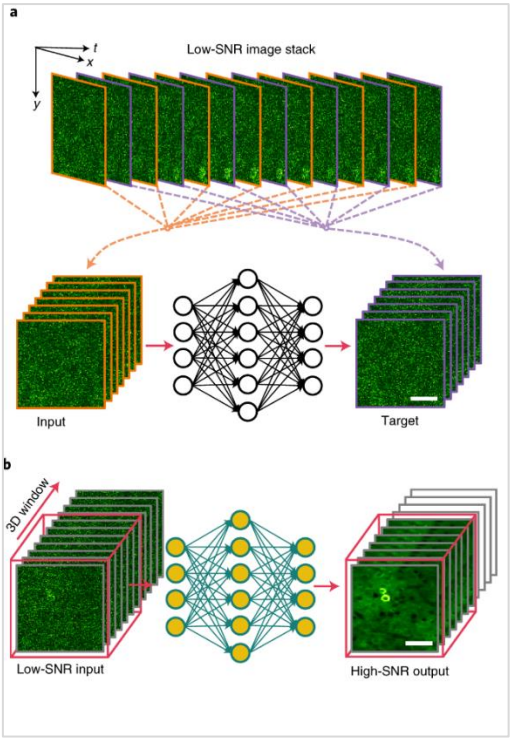


图 5.10 a) DeepCAD 自监督训练示意图。模型训练过程中不需要高信噪比信号。只需要将原始的低信噪比延时成像中取出固定时间窗口长度的图栈。将该图栈相互间隔的划分的两个子栈。分别当作输入数据和目标数据来训练一个 3D U-Net 神经网络。训练后的该网络即具有滤噪功能。b) 训练好的 DeepCAD 实际运行效果[16]。

5.2.2 智能化的生物大数据分析

随着以组学和超分辨成像等为代表的技术不断进步，生物实验室中产生的数据也在飞快地积累。如何解释海量的高精度、高质量数据也是一个不可回避的问

题。本节我们将介绍本年度在细胞的蛋白质组结构描述和神经科学领域的全脑匹配两个领域产生的典型的智能化生物数据处理技术。

细胞蛋白质组结构可视化

从结构上来看，真核细胞由细胞器等大尺度结构组成，这些成分又会逐级分解成更小的成分，例如凝聚物和蛋白质复合物，最终形成一个具有至少四个数量级多尺度、模块化结构。有关如何描述这些复杂的亚细胞结构，当前存在两种方法——蛋白质荧光成像和蛋白质-蛋白质关联分析。尽管数据特征大不相同，两种方法对蛋白质在结构中的定位具有很大的互补性：蛋白质成像通过蛋白质相对细胞核等标志物的位置定位蛋白，而蛋白质-蛋白质关联则通过蛋白质附近的蛋白质组成来定位该蛋白。目前，两种方法都已高度自动化，产生了大量数据。但由于他们具有不同的数据质量和分辨率，所以通常只能被单独分析而不能结合在一起进行更深入的研究。如果能够将这两个不同领域的数据结合起来进行统一分析，将极大提升人类对细胞内部结构和蛋白质性质的认识。

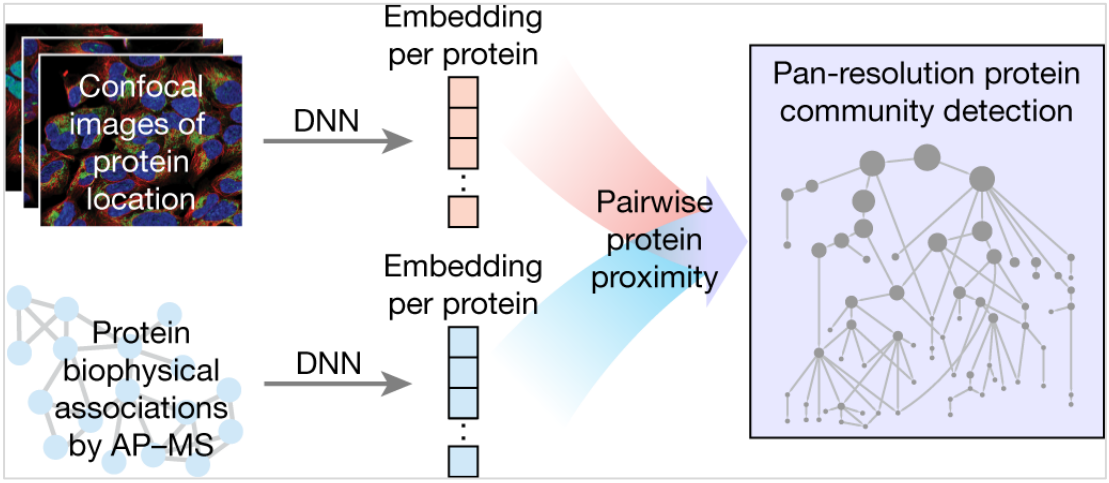


图 5.11 利用神经网络方法将蛋白质的成像和关联数据生成每种蛋白质结构图流程示意图。利用这种方法可以生成细胞中蛋白的泛尺度可视化的结构[17]。

2021 年 11 月，美国加州大学圣地亚哥分校 Trey Ideker 研究组与瑞典皇家理工学院以及斯坦福大学 Emma Lundberg 研究组合作，提出了融合两类数据的一种可靠方法[17]。利用深层神经网络方法，他们整合了人类蛋白质图谱（Human Protein Atlas）[18]的免疫荧光图像与 BioPlex 蛋白相互作用组学数据库（BioPlex InterRactiome）[19]中的亲和纯化质谱（AP-MS）结果。深度神经网络被用来将两

种数据对每个蛋白的测量投影到低维度空间，将这两个空间中每个蛋白的坐标整合、调校以形成两两蛋白间距离后，最终构建了多尺度集成细胞图谱 MuSIC1.0（Multi-scale integrated cell）。MuSIC 能够涵盖了从非常小（小于 50 nm）到非常大尺度（大于 1μm）的信息。

利用 MuSIC，文章作者解析出 69 个亚细胞系统（并且其中可能有一半是新发现的）。MuSIC 提高了成像分辨率，同时为蛋白质相互作用提供了一种空间可视化的描述，为在蛋白质组范围内整合不同类型的数据并构建细胞图谱铺平了道路。

跨模态小鼠全脑配准

在神经科学领域，近年来出现了一系列全脑尺度神经绘制项目，使用了包括双光子断层扫描、荧光断层扫描、光片显微镜等方式进行大规模三维体成像，或使用磁共振技术实现全脑成像。将这些多维全脑图像配准到标准脑图谱上对于表征神经元类型和构建脑神经回路图至关重要。然而目前来说，跨模态图像配准依然具有挑战性。

“所谓工欲善其事，必先利其器。我的团队现阶段的主要目标，仍然是继续完善大规模脑数据的处理和分析平台，为神经生物学家提供高效、强大的计算分析工具，助力脑科学研究以及中国脑计划。”——屈磊

2021 年 11 月，东南大学脑科学与智能技术研究院（以下简称东大脑智院）与美国艾伦脑科学研究所彭汉川教授与安徽大学电子信息工程学院副院长屈磊教授合作，开发了跨模态配准流程 mBrainAligner[20]。该方法基于相干标记点映射（coherent landmark mapping，CLM）图像配准框架,集成了深度学习技术以实现可靠的跨模态全脑配准。与传统的“检测-匹配-筛选”策略不同，CLM 使用“检测-映射”策略来系统地解决跨模态配准的问题，他们还使用深度神经网络在配准准确性和鲁棒性方面进一步提升了 CLM。结果表明，与现有方法相比，mBrainAligner 可以解决荧光显微光学切片断层成像（fMOST）数据到 Allen 标准脑图谱[21]的配准这一在当前具有挑战性的问题。且 mBrainAligner 具有更高

的准确性，这就为大规模单神经元分析工作提供了基础。他们还使用其他方式（包括 LSFM、VISoR 和磁共振成像）验证了 mBrainAligner 的泛化能力和性能。

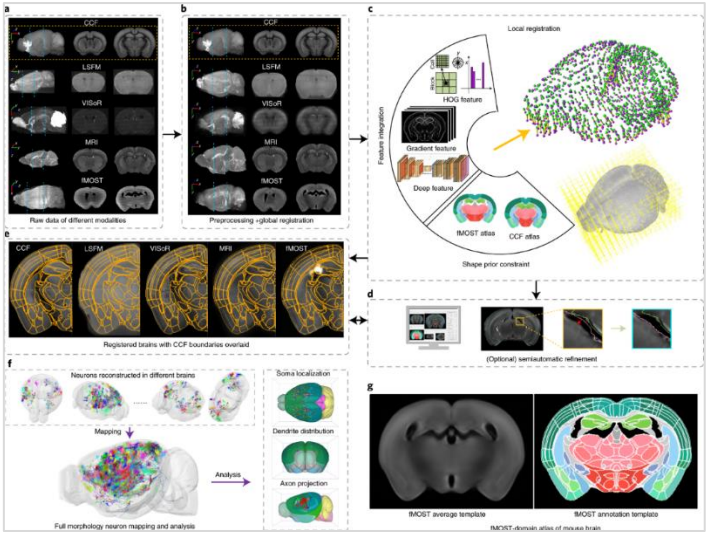


图 5.12 mBrainAligner 配准流程概述。a)左：不同模态下的原始脑图像的最大强度投影，中和右：冠状切片。b)经过预处理后和全局配准的大脑。c)基于 CLM 的局部配准算法示意图。d）可选的半自动配准模块。e)不同模态的脑数据经过配准后，叠合在标准脑模板 CCFv3 中。f)将树突、轴突和细胞体映射到规范坐标空间以进行可视化、比较和分析。g)使用 mBrainAligner 构建的小鼠大脑的 fMOST 图谱[20]。

mBrainAligner 将整个小鼠大脑图像与 Allen 公共坐标框架图谱（Common Coordinate Framework atlas，CCFv3）[21]对齐。从而实现了 31 个高分辨率 fMOST 脑图像和 1741 个小鼠全脑神经元形态重建数据到 Allen 公共坐标框架图谱（Common Coordinate Framework atlas，CCFv3）的映射，这些来自不同小鼠的细胞在统一的坐标框架下能够被自动地进行脑区划分并进行分析，方便了数据的横向对比，最终实现多模态数据的统一整合。

5.3 总结与展望

对生命科学来说，知识的疆界很大程度上取决于技术的发展程度。从更大的范围、更高的分辨率、更多的信息量等方面获取生物系统的信息，并从中提取生物系统的工作机制，一直是神经科学乃至整个生物学的研究方向之一。同时，高质量高通量数据的涌入又促进了生命科学领域数据处理方法不断更新变革。以往以人工为主的处理手段越来越乏力，而自于机器学习与人工智能领域的新方法能

够精确处理海量的数据，并对数据做出更加合理的统计、分析与预测。可以这么说，来自于数据观测采集与数据处理两个方面的技术进步，很大程度上推动了神经科学乃至生命科学研究的进步。

在当前，机器学习与人工智能领域的成果在生命科学领域的应用仍然不够深入，应用范围也局限在（多维）图像分析处理，大规模组学分析等少数领域。值得欣喜的是，2021 年出现了一些跨模态，能够处理分析多种类型数据的新技术和新算法，进一步融合了生命科学与神经科学领域不同类型的数据，更加深入地理解实验结果和数据，促进了新的知识和理论的产生。另一方面，数据处理方法的通用性也是当前生命科学和神经科学面临的问题。算法往往局限于某些特定情况采集的数据中，妨碍了算法进一步推广应用。如何提升算法的稳定性与泛用性，依然是对机器学习与人工智能领域的挑战。高通量的，高泛用性的数据采集，数据质量控制与质量提升以及数据分析算法将极大的提升生命科学，尤其是是神经科学、认知科学对结果的分析速度与程度，加速人们对本领域知识的开拓与理解。

神经科学、认知科学不但是人类认知自我的途径，也担负着为人工智能提供原理性依据的任务。因此，机器学习与人工智能新技术对上述领域的推动终将“反哺”自身，两者相辅相成的发展必将产生更加富有创造性、颠覆性的科研成就。

参考文献

[1] Zhao, W., et al., Sparse deconvolution improves the resolution of live-cell super-resolution fluorescence microscopy. Nat Biotechnol, 2021.

[2] Huang, X., et al., Fast, long-term, super-resolution imaging with Hessian structured illumination microscopy. Nat Biotechnol, 2018. 36(5): p. 451-459.

[3] Luo, C., et al., Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. Science, 2017. 357(6351): p. 600-604.

[4] Lichtman, J.W. and W. Denk, The big and the small: challenges of imaging the brain's circuits. Science, 2011. 334(6056): p. 618-23.

[5] Tsurui, H., et al., Seven-color fluorescence imaging of tissue samples based on Fourier spectroscopy and singular value decomposition. J Histochem Cytochem, 2000. 48(5): p. 653-62.

[6] Cutrale, F., et al., Hyperspectral phasor analysis enables multiplexed 5D in vivo imaging. Nat Methods, 2017. 14(2): p. 149-152.

[7] Shi, L., et al., Highly-multiplexed volumetric mapping with Raman dye imaging and tissue clearing. Nat Biotechnol, 2021.

[8] White, J.G., et al., The structure of the nervous system of the nematode Caenorhabditis elegans. Philos Trans R Soc Lond B Biol Sci, 1986. 314(1165): p. 1-340.

[9] Morgan, J.L. and J.W. Lichtman, Why not connectomics? Nature Methods, 2013. 10(6): p. 494-500.

[10] Brouillette, M. New Brain Maps Can Predict Behaviors. 2021; Available from: https://www.quantamagazine.org/new-brain-maps-can-predict-behaviors-20211206/.

[11] Scheffer, L.K., et al., A connectome and analysis of the adult Drosophila central brain. Elife, 2020. 9.

[12] Shapson-Coe, A., et al., A connectomic study of a petascale fragment of human cerebral cortex. bioRxiv, 2021.

[13] Susoy, V., et al., Natural sensory context drives diverse brain-wide activity during C. elegans mating. Cell, 2021. 184(20): p. 5122-5137 e17.

[14] Witvliet, D., et al., Connectomes across development reveal principles of brain maturation. Nature, 2021. 596(7871): p. 257-261.

[15] Greenwald, N.F., et al., Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nature Biotechnology, 2021.

[16] Li, X.Y., et al., Reinforcing neuron extraction and spike inference in calcium imaging using deep self-supervised denoising. Nature Methods, 2021. 18(11): p. 1395-+.

[17] Qin, Y., et al., A multi-scale map of cell structure fusing protein images and interactions. Nature, 2021. 600(7889): p. 536-542.

[18] Thul, P.J., et al., A subcellular map of the human proteome. Science, 2017.

356(6340).
[19] Huttlin, E.L., et al., Architecture of the human interactome defines protein communities and disease networks. Nature, 2017. 545(7655): p. 505-509.
[20] Qu, L., et al., Cross-modal coherent registration of whole mouse brains. Nature Methods, 2021.
[21] Keller, P.J. and M.B. Ahrens, Visualizing Whole-Brain Activity and Development at the Single-Cell Level Using Light-Sheet Microscopy. Neuron, 2015. 85(3): p. 462-483.

结 语

早在上个世纪六十年代，麦卡锡、明斯基、香农等众多计算科学家在美国达特茅斯学院开会研讨“如何用机器模拟人的智能”，并首次提出了“人工智能（AI, Artificial Intelligence）”的概念。这一概念的提出，标志着人工智能学科的诞生。

在随后的几十年中，人工智能的发展就像坐过山车一样遭遇了几轮起起伏伏，其中就包含了两次艰难的“寒冬期”。人工智能发展的第一次低谷发生在1974 年到 1980 年，由于人们提出了一些不切实际的目标，随之而来的是外行对人工智能的巨大期望值。然而，受当时计算机性能及数据库缺失等技术瓶颈的影响，实际问题的复杂性让科学家和政府部门对人工智能的前景产生质疑，数学科学家莱特希尔甚至尖锐地指出人工智能的研究已经完全失败。此后，专家系统的成功研发带领人工智能经历了一段短暂的崛起，但很快苹果和 IBM 公司开发的台式计算机性能便超过了搭载专家系统的通用型计算机，人工智能领域再次步入低谷。90 年代后期和 21 世纪初期，随着互联网技术的发展，人工智能的研究在新平台、新机遇下强势崛起，以 IBM 的“深蓝”计算机系统及深度学习为代表的新兴产物及其应用不断给人们带来惊喜。纵观人工智能的发展简史，似乎正印证了罗伊·阿玛拉所说的“我们总是高估一项科技所带来的短期效益，却又低估它的长期影响”。

如今，计算机算力的巨幅提高让人工智能有了更多的应用场景，人工智能在各类视频游戏、围棋和扑克等竞技类比赛中胜过人类专家的例子比比皆是。随着科技巨头的深度参与，人工智能在教育、交通、通信和医疗等方面的应用取得了令人瞩目的突破。在这里，我们认为，除了计算机硬件算力的贡献，经常被忽视的、却十分重要的贡献是来自于认知神经科学的研究成果及学科思想。

首先，认知神经科学可以树立智能的基准并检验当前的人工智能算法。考虑到人工智能算法旨在模拟、拓展并最终超越自然智能所能达成的功能，描述并探索生物智能的认知神经科学便为智能提供了行之有效的量化基准。此外，新一代人工智能的最新进展大多源自于数理模型的更迭与硬件算力的飞跃。在此基础上，

进一步探寻现有智能算法在认知神经科学范畴的映射研究，能从更多角度增强检验人工智能算法的可靠性。

其次，认知神经科学为人工智能算法提供了可行有效的潜在优化方案。尽管当前人工智能算法在各个领域都取得了长足进步，然而如何在能耗限制的情况下实现同等优秀的性能是实际应用中人工智能算法面临的重大挑战。在平衡算力与能耗的问题上，未来科学家们或许可以采用认知神经科学的研究手段，识别出人脑核心的生物计算算法，以启发优化人工智能算法。

更为重要的是，认知神经科学的基础研究可以为构建人工智能的新型算法和架构提供丰富的灵感来源。当前的人工智能算法，比如卷积神经网络，Transformer 网络等，核心架构皆可以追溯到认知神经框架下的视觉层级加工、注意与工作记忆等概念。随着本白皮书介绍的神经元解码、脑机接口及高时空分辨率成像等技术的进一步发展，我们可以期待认知神经科学携手人工智能科学开创新的纪元。

北京智源人工智能研究院“人工智能的认知神经基础重大研究方向”旨在将神经科学、认知科学和信息科学进行交叉融合，加强人工智能和脑科学的双向互动和螺旋发展，揭示生物智能系统的精细结构和工作机理，构建功能类脑、性能超脑的智能系统，以视觉等功能和典型模式动物作为参照物构建智能水平测试平台，为人工智能未来发展探索可行道路。在此背景下，此次 2021 版的白皮书延续了去年的传统，梳理了系统神经科学、认知神经科学、计算神经科学、脑机接口以及神经科学新技术领域的重要突破，从多个角度介绍了脑科学及相应的智能技术，以及二者交叉领域的重要进展，并结合本方向六位智源学者的最新探索成果，提出了一些对领域发展的观察与思考。

本年度《人工智能的认知神经基础白皮书》的发布，以及智源生物智能开源开放平台的正式上线，都是我们为进一步深入开展学科交叉，实现人工智能与认知神经科学相互启发和促进做出的尝试。期望我们的这些尝试与努力能为读者带来些许助益，对于本书的疏漏与不足之处，敬请各位读者批评指正。

关于我们

北京智源人工智能研究院

北京智源人工智能研究院（Beijing Academy of Artificial Intelligence，BAAI）成立于 2018 年 11 月，是在科技部和北京市委市政府的指导和支持下，由北京市科委和海淀区政府推动成立的新型研发机构。

智源研究院的愿景是，聚焦原始创新和核心技术，建立自由探索与目标导向相结合的科研体制。支持科学家勇闯人工智能科技前沿“无人区”，挑战最基础的问题和最关键的难题，推动人工智能理论、方法、工具、系统和应用取得变革性、颠覆性突破。营造全球最佳的学术和技术创新生态，推动北京成为全球人工智能学术思想、基础理论、顶尖人才、企业创新和发展政策的源头，率先成为国际领先的人工智能创新中心。推动人工智能产业发展和深度应用，改变人类社会生活，促进人类、环境和智能的可持续发展。

智源「人工智能的认识神经基础」重大研究方向

智源研究院「人工智能的认知神经基础（Brain and Machine Intelligence）」重大研究方向旨在将神经科学、认知科学和信息科学进行交叉融合，加强人工智能和脑科学的双向互动和螺旋发展，揭示生物智能系统的精细结构和工作机理，构建功能类脑、性能超脑的智能系统，以视觉等功能和典型模式动物作为参照构建智能水平测试平台，为人工智能未来发展探索可行道路。

智源生物智能开源开放平台

智源研究院「生物智能开源开放平台（Bio-Intelligence Opensource Platform, BIOSP）」是由智源「人工智能的认知神经基础」重大研究方向科学家团队共同发起建成的国内首个智能科学研究基础设施。该平台集成并开源了从人类认知行为范式数据库（CogNet）、到生物脑神经活动多模态数据库（BrainDB）、到类脑模型算法、到国产开源软件工具及领域前沿动态等内容，旨在通过开放数据、模型、算法、软件工具等一站式科研资源的方式，为认知科学、神经科学和计算科学及相关交叉领域的研究人员、学生和相关从业者搭建一个服务智能科学研究的平台型基础设施，进而推动和支撑国内脑启发的通用智能研究工作。

编者介绍

张博，智源博士后，合作导师为智源首席科学家、清华大学脑与智能实验室刘嘉教授。2020 年北京大学心理与认知科学学院获博士学位，主要研究方向为：人脑视觉、记忆神经机制、神经计算与模拟、影像图像的处理与分析等。

苏杰，智源博士后，合作导师为智源首席科学家、清华大学脑与智能实验室刘嘉教授。2020 年北京师范大学心理学部脑与认知科学研究院获博士学位，主要研究方向为：学习与决策认知神经科学、元认知、视觉认知等。

蒋龙生，智源博士后，合作导师为智源研究员、清华大学脑与智能实验室宋森教授。2021 年克莱姆森大学机械工程系获博士学位，主要研究方向:抓握动作的视觉认知机理、类人决策行为模型等。

邹晓龙，智源博士后，合作导师为智源研究员、北京大学心理与认知科学学院吴思教授。2018 年北京师范大学系统科学学院获博士学位，主要研究方向为：计算神经科学，类脑计算等。

陈路瑶，智源博士后，合作导师为智源研究员、北京大学心理与认知科学学院方方教授，华中科技大学微电子学与固体电子学博士，海外加州大学洛杉矶分校联合培养博士，主要研究方向为脑机接口，神经调控等。

陈智强，智源博士后，合作导师为智源研究员、中国科学院自动化所余山研究员。2021 年于中国科学院自动化研究所获得博士学位，主要研究方向为：脑启发的深度本征网络、类脑智能等。

刘祥，智源博士后，合作导师为智源研究员、北京大学未来技术学院陈良怡教授。2020 年北京大学前沿交叉学科研究院获博士学位，主要研究方向为：斑马鱼全脑神经信号数据分析；全脑双光子成像信号处理；结合行为学的斑马鱼全脑模块

划分与神经回路分析；细胞生物学大数据统计分析。

徐琳璐，智源博士后，合作导师为智源研究员、北京大学未来技术学院陈良怡教授。2020 年于中国科学院生物物理研究所获博士学位，主要研究方向为：斑马鱼视觉与斑马鱼学习记忆的神经机制。

秦方博，中国科学院自动化研究所助理研究员，中国科学院大学博士，主要研究方向：柔性电极植入机器人、智能视觉感知、精密操作控制。

韩程，中国科学院大学自动化研究所博士研究生，主要研究方向：面向脑机接口的无线供电技术、人体媒介能量传输技术。

联系方式

闫亚琼 智源科研项目经理

邮箱：yqyan@baai.ac.cn

地址：北京市海淀区成府路 150 号智源大厦

北京智源人工智能研究院

官网：https://www.baai.ac.cn

邮箱：press@baai.ac.cn

地址：北京市海淀区成府路 150 号智源大厦