CN 51-1346/O4 ISSN 1003-501X (印刷版) ISSN 2094-4019 (网络版)

光电工程

Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

光子神经网络研究进展

项水英，宋紫薇，张雅慧，郭星星，韩亚楠，郝跃

引用本文:

项水英，宋紫薇，张雅慧，等. 光子神经网络研究进展[J]. 光电工程，2024， **51**(7): 240101.

Xiang S Y, Song Z W, Zhang Y H, et al. Progress in the research of optical neural networks[J]. *Opto-Electron Eng*, 2024, **51**(7): 240101.

https://doi.org/10.12086/oee.2024.240101

收稿日期: 2024-05-04; 修改日期: 2024-06-27; 录用日期: 2024-06-28

相关论文

Pattern recognition in multi-synaptic photonic spiking neural networks based on a DFB-SA chip

Yanan Han, Shuiying Xiang, Ziwei Song, Shuang Gao, Xingxing Guo, Yahui Zhang, Yuechun Shi, Xiangfei Chen, Yue Hao

Opto-Electronic Science 2023, **2**(9): 230021 doi: 10.29026/oes.2023.230021

Photonic integrated neuro-synaptic core for convolutional spiking neural network

Shuiying Xiang, Yuechun Shi, Yahui Zhang, Xingxing Guo, Ling Zheng, Yanan Han, Yuna Zhang, Ziwei Song, Dianzhuang Zheng, Tao Zhang, Hailing Wang, Xiaojun Zhu, Xiangfei Chen, Min Qiu, Yichen Shen, Wanhua Zheng, Yue Hao

Opto-Electronic Advances 2023, **6**(11): 230140 doi: 10.29026/oea.2023.230140

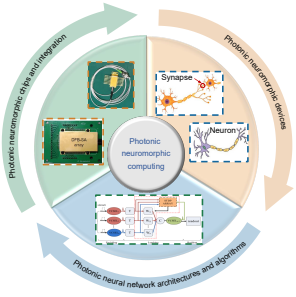
Opto-Electronic Engineering
光 电 工 程

Review

2024 年, 第 51 卷, 第 7 期

DOI: 10.12086/oee.2024.240101

光子神经网络研究进展



项水英[1*]，宋紫薇[2]，张雅慧[1]，郭星星[1]，韩亚楠[1]，郝　跃[3]

[1]西安电子科技大学空天地一体化综合业务网全国重点实验室，陕西 西安 710071；

[2]空军工程大学基础部，陕西 西安 710051；

[3]西安电子科技大学微电子学院宽禁带半导体国家工程研究中心，陕西 西安 710071

摘要：在数据海量化、信息化的时代，电子计算机处理系统所面临的算力和能耗等性能要求愈发严苛，传统冯·诺依曼架构存在“内存墙”和“功耗墙”瓶颈，加之摩尔定律放缓甚至失效，使得电子芯片在计算速度和功耗方面遇到极大挑战，利用光计算替代传统电子计算将是解决当前算力与功耗问题的极具潜力的途径之一。本文系统地梳理了片上集成和自由空间的光子神经网络架构与算法方面的研究进展，详细介绍了典型的研究工作，然后讨论并对比了这两种光子神经网络的优劣势，以及光子神经网络的训练策略等。最后探讨了光子神经网络面临的挑战，并对其未来发展进行了前瞻性的展望。

关键词：光子神经网络；片上集成；自由空间；类脑计算

中图分类号：TN25　　**文献标志码：**A

项水英，宋紫薇，张雅慧，等. 光子神经网络研究进展 [J]. 光电工程，2024，**51**(7): 240101

Xiang S Y, Song Z W, Zhang Y H, et al. Progress in the research of optical neural networks[J]. *Opto-Electron Eng*, 2024, **51**(7): 240101

Progress in the research of optical neural networks

Xiang Shuiying[1*], Song Ziwei[2], Zhang Yahui[1], Guo Xingxing[1], Han Yanan[1], Hao Yue[3]

[1]State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an, Shaanxi 710071, China;

[2]Fundamentals Department, Air Force Engineering University, Xi'an, Shaanxi 710051, China;

[3]State Key Discipline Laboratory of Wide Bandgap Semiconductor Technology, School of Microelectronics, Xidian University, Xi'an, Shaanxi 710071, China

Abstract: In the era of massive data and information, electronic computer processing systems face increasingly greater demands on computing power and energy consumption. Bottlenecks such as the "memory wall" and "power wall" inherent in the traditional von Neumann architecture, coupled with the slowing down or even invalidation of Moore's Law, have posed significant challenges to electronic chips in terms of computing speed and power consumption. Utilizing optical computing as an alternative to traditional electronic computing represents one of the

收稿日期：2024-05-04；修回日期：2024-06-27；录用日期：2024-06-28

基金项目：国家重点研发计划项目 (2021YFB2801900，2021YFB2801901，2021YFB2801902，2021YFB2801903，2021YFB2801904，2018YFE0201200)；国家优秀青年科学基金项目 (62022062)；国家自然科学基金项目 (61974177)；中央高校基本科研业务费专项 (QTZX23041)

*通信作者：项水英，syxiang@xidian.edu.cn。

most promising avenues to address current challenges in computing power and power consumption. This review systematically summarized the research progress of optical neural network architectures and algorithms in both on-chip integration and free space, and described typical research efforts in detail. Then, the advantages and disadvantages of these two types of optical neural networks and the training strategies of optical neural networks were discussed and compared. Finally, the potential challenges that optical neural networks may encounter were discussed in depth, and a forward-looking perspective on their future development was offered.

Keywords: optical neural network; on-chip integration; free space; brain-like computing

1 引 言

人类科技进步过程中，大自然提供了源源不断的灵感源泉和创新策动力，并从多个维度为人类的科技发展贡献了卓越智慧。比如，通过深入研究鸟类飞行的生物学原理，人们得以解决复杂的航空技术挑战；再如，借鉴蚁群精巧的搜索策略，科学家们开发出了高效的优化算法；同样地，植物光合作用过程中所展现出来的高效能量转换机制，则为新型能源生产和利用提供了极具价值的参考模型。诸如此类源自生物界的启示不胜枚举，它们持续不断地为人类现实生活及生产领域提供了一系列精密而巧妙的解决方案。然而，大自然最令人叹为观止的造物之一无疑是人类大脑，它生动地诠释了“智能”的概念，通过模仿人脑的处理机制建立接近乃至超越人脑智能的机器，一直是人们的一个朴素理念，因此，人工智能 (artificial intelligence，AI)[1] 应运而生，人脑也成为衡量 AI 系统能力的标准参照物。

人工智能通过模拟人脑中信息存储和处理机制等智能行为，使机器具有一定程度的智能水平。它作为一门新兴学科诞生于 20 世纪 50 年代中期，随后便在充满未知的道路上不断探索，其发展历程如图 1 所示，曾于 20 世纪 50 年代末和 80 年代初先后两次步入发展高峰，但因为技术瓶颈、应用成本等局限性均在短暂繁荣后跌入低谷。自 2006 年深度学习算法[2] 的提出，以及在信息技术的引领下，随着数据信息的快速积累和运算能力的大幅提升，人工智能技术应用取得了突破性发展，并第三次站在了科技发展的浪潮之巅，这次的研究狂潮至今仍在持续。2016 年，由深度神经网络驱动的计算机程序 AlphaGo 战胜了人类顶尖围棋选手，这不仅刷新了普通大众对 AI 技术的看法，而且促使 AI 上升至国家战略层面。2023 年，世人见证了 ChatGPT 在全球范围的大火，以生成式人工智能为代表的新一代 AI 的问世，改变了 AI 技术与应用的发展轨迹，加速了人与 AI 的互动进程，是 AI 发展史上的新里程碑。近年来，我国《新一代人工智能发展规划》《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》等文件相继出台，从国家层面对人工智能进行系统布局。特别地，在 2024 年的两会工作报告中，“人工智能+”这一概念首次被纳入国家发展战略的核心内容，这无疑揭示了我国对人工智能技术及其融合应用的高度重视和前瞻布局。

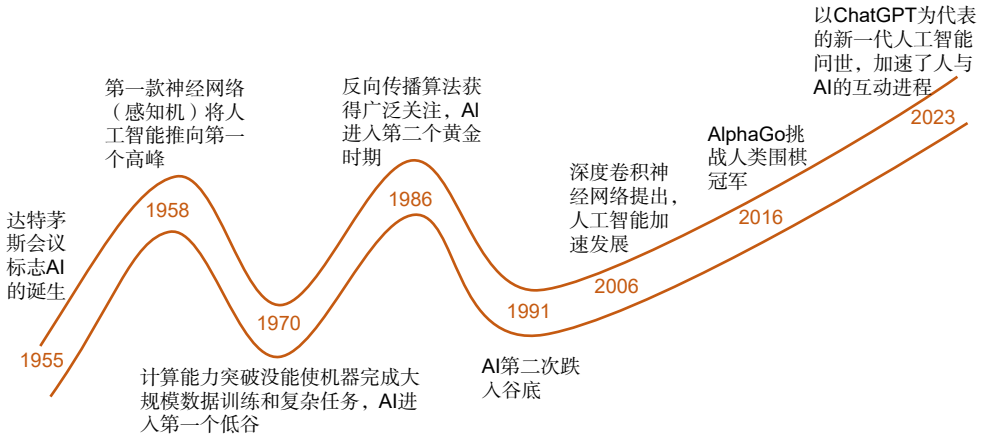


图 1 人工智能发展历程

Fig. 1 The development history of artificial intelligence

现如今，几乎最先进的人工智能算法都使用神经网络实现，并在基于冯·诺依曼架构的数字计算机上运行，然而，神经网络计算模型与冯·诺依曼体系架构在关键特征上有着本质的区别。一方面，神经网络是一种受大脑神经突触结构启发的计算模型，网络中并没有明显区分计算单元和存储单元，而是将网络信息蕴含在神经元和突触之中，在外部刺激进入网络传输的同时便完成了计算，相反，基于冯·诺依曼架构的数字计算机中计算单元和存储单元则是物理上分离的芯片，计算前需要先从存储单元读取出相应信息再转移至计算单元中进行计算；另一方面，神经网络具有高度并行和分布式的特性，而冯·诺依曼架构本质上是顺序执行，最佳情况下是使用多处理器实现顺序并行。因此，这两种架构之间的显著差异导致计算速度减慢、能耗增加，例如，AlphaGo 在击败世界冠军之前一共进行了 16 万场对局，使用 50 个 GPU 进行了 4 周的训练，消耗大约 4×10^{10} 焦耳的能量，相当于一个成年人维持 10 年新陈代谢所需的能量；据斯坦福大学人工智能研究所发布的《2023 年人工智能指数报告》称，GPT-3 单次训练耗电量高达 1287 兆瓦时，成本约为 140 万美元，对于多模态的 GPT-4，耗电量和成本更加巨大；而人脑仅需 20 瓦左右的能量就可以完成各种复杂的任务。因此，与人脑智能相比，能源消耗问题已成为限制人工智能发展的一大障碍。此外，在 2006 年深度学习出现之前，用于 AI 训练的算力需求基本符合摩尔定律[3]；之后，算力需求不断增长，2012 年后，AI 模型训练算力需求远超摩尔定律所预测的算力供给量；随着大数据与大模型蓬勃发展，训练算力需求呈指数级增长，目前的电子芯片能提供的算力和 AI 的算力需求高度不匹配，无法满足现代技术日益增长的计算需求。因此，需要探索各种新兴方法、创新理论、新型器件以及新颖的计算架构来开发性能更出色、能耗更低的新一代计算机。

上世纪 80 年代，科学家 Mead 受生物神经系统的启发，首次提出了类脑智能[4] 的概念。自此以后，实现类脑智能成为人类一直追求的梦想，并逐步上升至国家战略层面。进入 21 世纪后，各国纷纷布局脑科学领域的重大项目，全球主要国家的“脑计划”概况如图 2 所示。2005 年，瑞士洛桑联邦理工学院脑与心理研究所发起蓝脑计划 (blue brain project)，利用逆向工程模拟脑，“复制”人脑所有的活动以及内部发生的各种反应；在此计划基础上，欧盟于 2013 年启动了人类大脑计划 (human brain project)[5]；同年，美国政府启动脑科学计划 (BRAIN initiative)[6-7]，旨在支持创新技术的开发和应用，以促进对大脑功能的动态理解，该计划于 2022 年迭代至 2.0 版本[8]，目标是期望改变神经科学研究的方法和人脑疾病治疗的方式；紧接着在 2014 年，日本启动综合神经技术用于疾病研究的脑图谱计划 (brain mapping by integrated neurotechnologies for disease studies, Brain/MINDS)[9]，该计划包括绘制非人灵长类大脑的结构和功能图谱、发展创新神经技术、人类大脑的结构功能重建和临床研究三个主题，在 Brain/MINDS 的框架基础上，日本在 2018 年进一步启动了 Brain/MINDS Beyond 计划；

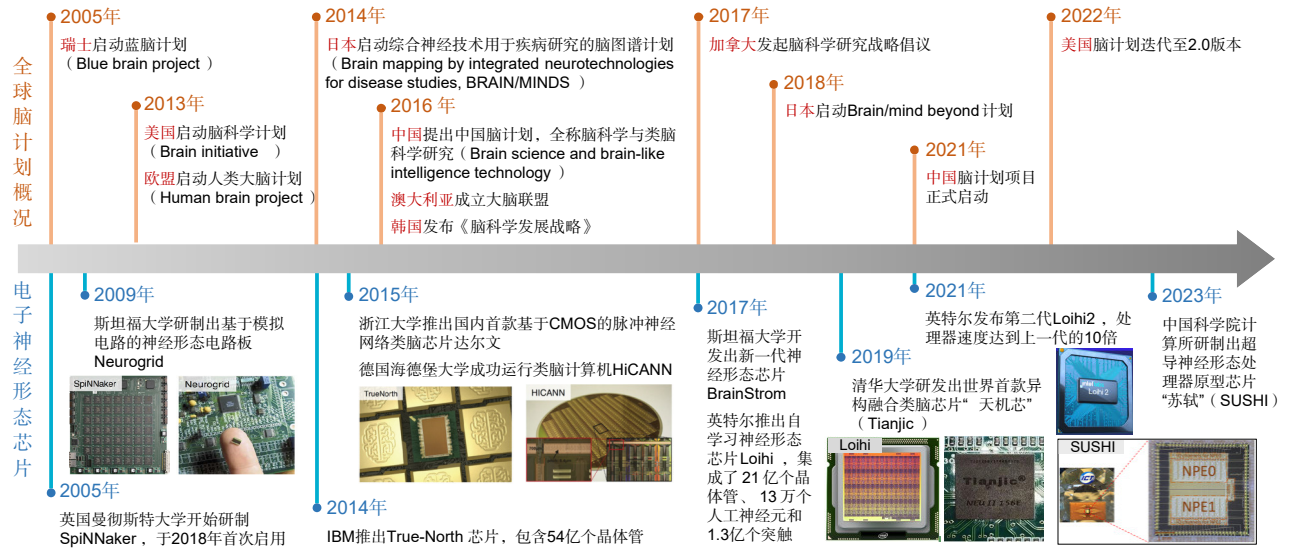


图 2 类脑智能的国家战略概况及硬件实现[5-12,15-23]

Fig. 2 National strategy overview and hardware implementation of brain-like intelligence[5-12,15-23]

2016 年，我国提出脑科学与类脑科学研究 (brain science and brain-like intelligence technology)[10-12]，采用“一体两翼”的战略架构布局，即以研究脑认知的神经原理为“主体”，以研发脑重大疾病诊治新手段和脑机智能新技术为“两翼”开展研究；2021 年，我国科技部发布科技创新 2030——“脑科学与类脑研究”重大项目，中国脑计划项目正式启动。因此，受脑信息处理机制启发的类脑计算引起了国内外广泛关注并成为全球科研热点。类脑计算[13] (即神经形态计算) 致力于开发模拟大脑的人工智能技术，通过借鉴大脑的神经系统结构及其处理信息的基本规律及机制，力求在硬件实现与软件算法等多个维度对现有的计算体系与系统实施根本性的革新，以构建低能耗、高性能的计算系统[14]。它基于非冯·诺依曼架构，颠覆了传统的计算模式，更具生物可解释性。

类脑智能的基石是类脑神经形态硬件芯片，通过构建模拟大脑底层结构的全新硬件平台，打破冯·诺依曼体系结构固有的性能限制，并在解决智能任务时获得速度和效率方面的优势。近些年来，基于微电子技术与平台构建的神经形态硬件取得了显著进展，如图 2 所示，自 1989 年类脑工程概念[4] 提出，先后有英国曼彻斯特大学研制的 SpiNNaker[15]、斯坦福大学研制的 Neurogrid[16]、IBM 研制的 True-North[17]、德国海德堡大学的 HiCANN[18]、浙江大学研制的达尔文类脑芯片 (Darwin)[19]、英特尔研制的 Loihi[20] 和 Loihi2[21]、清华大学研发的天机芯 (Tianjic)[22]、中国科学院研制的“苏轼”(SUSHI)[23] 等神经形态芯片问世。然而，电子连接面临严重的带宽-连接密度权衡的问题[24-25]，以及电子晶体管的特征尺寸和集成度逐渐逼近摩尔定律所揭示的物理极限，基于电子硬件的神经形态芯片已经遇到了速度、能耗和面积等方面的电子瓶颈，其能效和计算速度难以进一步提高。

采用光子计算方法替代传统电子计算方法，是克服摩尔定律所带来的物理极限以及冯•诺依曼架构局限性的有效途径之一，有望为解决当前算力与功耗的瓶颈问题提供强大动力。光子的高速、大带宽、高并行性和低串扰等特性特别适应于高互连密度的网络，在硬件神经形态计算领域受到越来越多的关注。因此，光子神经形态计算[26-27] 是光子学和神经形态计算交叉融合的前沿研究领域，充分利用了光子学的优势，通过克服电子瓶颈来显著加快计算速度和减少资源能耗，从而推动智能计算系统的高效运行，以满足日益增长的计算需求和能源效率要求。光子神经形态计算通过光学硬件模拟生物大脑基本构建块 (即神经元和突触)，并将其组合成适当规模的神经网络，以光子作为信息传输载体，利用光和光学器件的特性，来完成计算过程。光子神经网络功耗低、并行性高、速度快，在满足大量数据处理需求的同时降低了能耗，其发展历程可以追溯到 20 世纪 80 年代[28]，当时主要是基于光电效应和光学器件实现神经网络的模拟，随着光学技术的不断进步，如今的光子神经网络已经发展成为一种结合光学计算和神经网络算法的全新领域。

光子神经网络由大量复杂互连的线性层构成，矩阵乘法以光速进行，有效加速人工神经网络中密集的矩阵乘法计算过程，并减少能量和时间的消耗，通过非线性光学元件来实现人工神经网络中的非线性计算，一旦完成光神经网络的训练，整个结构就能以光速进行光信号计算，而无需额外的能量输入[29]。其中，光互连具有高度的并行性，光束可以在空间交叉而不会产生串扰，而且光的传播速度快，时延和色散都可以忽略不计[30-31]。光互连技术主要分为波导光互连与自由空间光互连两大类别，基于光互连机制，光子神经网络主要包括基于集成光学平台和基于自由空间光学两种实现形式。

本综述旨在回顾光子神经网络的研究进展，重点介绍该领域已经取得的显著性成果，并前瞻性地展望其未来的发展趋势及所面临的关键挑战和难题。首先综述了片上集成的光子神经网络的研究进展，包括基于半导体激光器、基于硅光微环谐振器 (micro-ring resonator，MRR)、基于马赫-增德尔干涉仪 (Mach-Zehnder interferometer，MZI) 和基于相变材料 (phase change material，PCM) 的光子神经网络；然后介绍了基于自由空间光学的光子神经网络的研究进展，主要介绍了基于衍射光学元件的光子神经网络；进一步介绍了光子神经网络的训练算法；最后讨论了光子神经网络面临的挑战，并展望其未来的研究方向和应用前景。

2 片上集成的光子神经网络

基于集成光学平台的光子神经网络使用光波导引导光波传播来实现互连，光波导通常分为两类：一种是圆柱形光波导，即光纤；另一种是集成光波导，包括平面光波导和条形介质光波导。基于半导体激光器的光子神经网络通常使用光纤互连实现，因此，本节分别介绍了基于半导体激光器的光子神经网络和基于

集成光波导平台的光子神经网络。

2.1 基于半导体激光器的光子神经网络

2016 年开始，西安电子科技大学研究团队围绕基于半导体激光器的光子神经网络展开了系统深入的研究，如图 3 所示。研究初期重点探索了光子神经形态计算的新原理器件，针对光子脉冲神经元，首次建立了基于偏振转换的垂直腔面半导体发射激光器 (vertical-cavity surface-emitting laser，VCSEL) 光子神经元理论模型，成功实现了类神经元的兴奋性响应[32]，并在 VCSEL 耦合级联系统中实现了兴奋性响应的稳定级联传输[33]，在此基础上，提出了双偏振光注入 VCSEL 产生兴奋性和抑制性动力学响应的新方案[34]，并分析了脉冲编码方案和存储特性[35-36]；2019 年首次发现偏振模竞争效应引起的类神经元抑制响应[37]，随后基于该抑制响应单步实现了非线性的异或 (exclusive OR，XOR) 运算[38]。此外，还提出了带有饱和吸收区的 VCSEL (VCSEL with a saturable absorber，VCSEL-SA) 实现多路脉冲编码的方案[36]。针对光学突触，2018 年建立了基于垂直腔面半导体光放大器 (vertical-cavity semiconductor optical amplifier，VCSOA) 的光学突触可塑性理论模型[39] 并进行了实验验证[40]，2020 年提出了基于双偏振光注入 VCSEL 的实时光突触可塑性方案[41]。

基于新原理器件的研究成果，进一步通过仿真和实验研究了基于这些器件的光子脉冲神经网络的理论模型与算法，从 2019 年开始建立了基于 VCSEL-SA 的“光子神经元-突触-学习算法”一体化光子脉冲神经网络物理模型，先后提出无监督学习算法[42]、有监督学习算法[43-44]、赢者通吃竞争学习机制[45] 和“时延-权重协同可塑性”监督学习算法[46]，完成脉冲模式识别[43]、声源定位[47]、脉冲序列学习[48]、联想记忆[49]、数字模式识别[50]、数独求解[51]、运动检测和方向识别[52] 等任务；进一步建立了基于 VCSEL-SA 的多层光子脉冲神经网络，提出监督学习算法实现 XOR 运算以及其他的逻辑运算任务[53]，提出结合 STDP 规则和梯度下降机制的改进学习算法实现对鸢尾花数据集和乳腺癌数据集的分类[44]，提出改进的反向传播算法通过理论结合实验完成非线性 XOR 运算、鸢尾花分

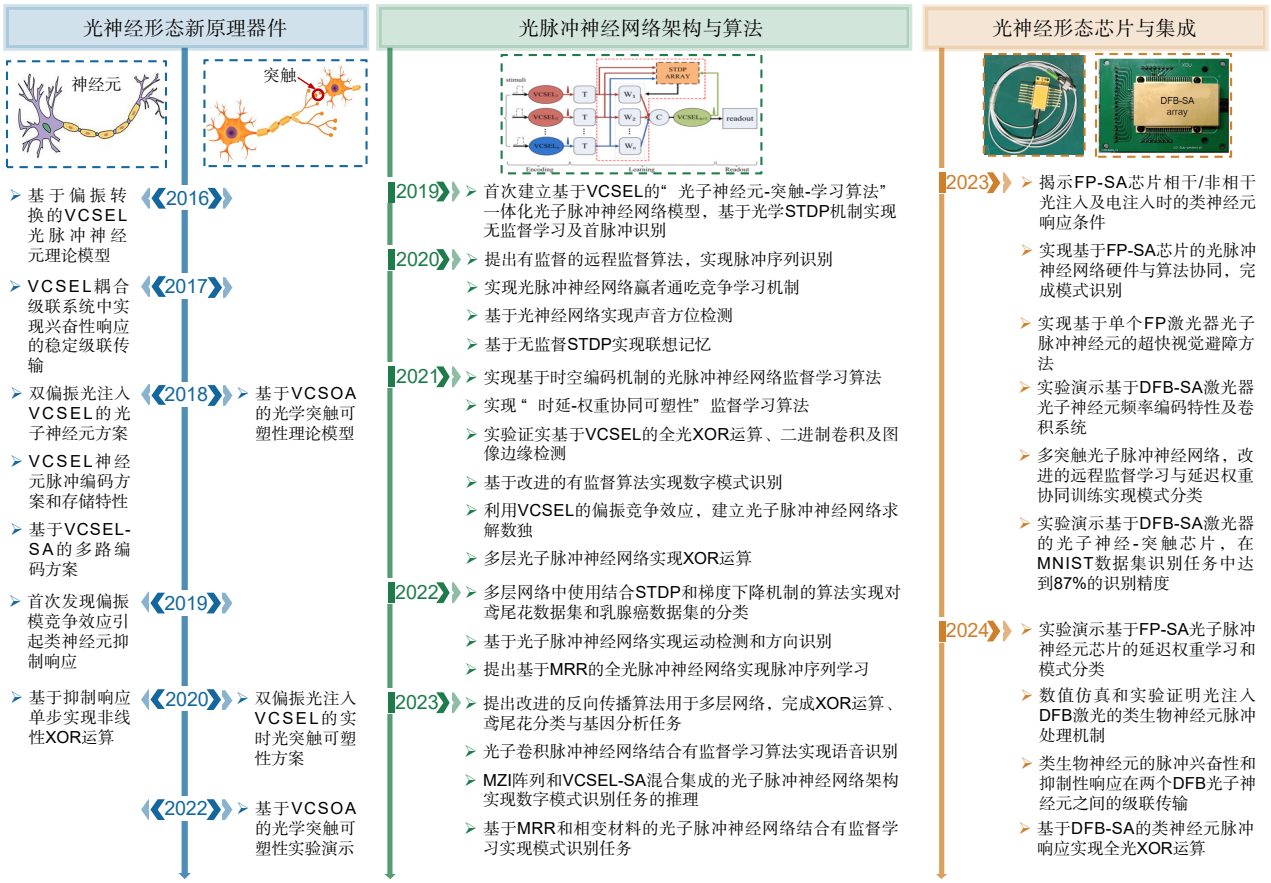


图 3 西安电子科技大学研究进展[32-72]

Fig. 3 Research progress of Xidian University[32-72]

类与基因分析任务[54-55]。同时，又提出一种光子卷积脉冲神经网络结构，结合有监督学习算法实现语音识别任务[56]。此外，研究了硅基光器件构建的光子神经网络，提出一种基于硅光 MRR 的全光脉冲神经网络实现脉冲序列学习[57]，进一步构建了基于 MRR 和相变材料的光子脉冲神经网络，并结合有监督学习实现了模式识别任务[58]，随后，提出一种硅基光 MZI 阵列和 VCSEL-SA 混合集成的光子脉冲神经网络架构实现数字模式识别任务的推理[59]。

近几年，在光子脉冲神经网络的理论模型与算法、硬件平台和集成技术持续进步的推动下，光子神经形态芯片及其集成的研究取得了显著进展，先后研制了法布里-珀罗 (Fabry–Pérot，FP) 激光器、含可饱和吸收区的法布里-珀罗 (FP with a saturable absorber，FP-SA) 激光器、分布式反馈 (distributed feedback，DFB) 激光器和含可饱和吸收区的分布式反馈 (DFB with a saturable absorber，DFB-SA) 激光器，在硬件平台研究了它们的类神经元特性，并构建了光子脉冲神经网络完成类神经计算任务。对于 FP-SA 激光器，首先研究了其在相干光注入和非相干光注入下的类神经元响应[60-61]，进一步实现了硬件-算法协同计算的光脉冲神经网络[62]，随后研究了由三个级联 FP-SA 构成的多层光子脉冲神经网络的硬件实现[63]，实验演示了基于 FP-SA 光子脉冲神经元芯片的延迟权重学习和模式分类[64]，以及单个双集成电极光子脉冲神经元的脉冲信息处理[65]；基于单个 FP 激光器光子脉冲神经元模拟了视觉神经系统的超快视觉避障方法并进行了实验验证[66]。对于 DFB 激光器，数值仿真和实验证明了光注入 DFB 激光器可以实现类生物神经元的脉冲处理机制[67]，并探索了类生物神经元的脉冲兴奋性和抑制性响应在两个 DFB 光子神经元之间的级联传输[68]。对于 DFB-SA 激光器，首先研究了其类神经元响应，基于此实现了全光 XOR 运算[69]；然后基于 DFB-SA 激光器频率编码的特性，演示了一种基于 DFB-SA 激光器光子神经元的卷积系统[70]；随后提出了一个多突触光子脉冲神经网络，将改进的远程监督学习算法与延迟权重协同训练方法相结合来实现模式分类[71]；最后，实验演示了基于 DFB-SA 激光器的光子神经-突触芯片，单个 DFB-SA 激光器可以同时实现线性加权和非线性脉冲激活，进一步基于制作的四通道 DFB-SA 激光器阵列实现了脉冲卷积神经网络的矩阵卷积运算，在 MNIST 数据集识别任务中达到 87% 的识别精度[72]。

除了西安电子科技大学研究团队的贡献，英国思克莱德大学 Hurtado 研究团队自 2010 年以来，也致力于基于半导体激光器的光子神经网络的研究，他们通过理论与实验研究了基于 VCSEL 的光子神经形态计算系统，首先探索了基于光注入 VCSEL 的偏振转换效应[73] 模拟生物神经元的兴奋性响应[74] 和抑制性响应[75]，研究了级联系统中脉冲的可控产生[76] 与传播[77]；然后展示了由单个 VCSEL 光子神经元构建的硬件友好的光子脉冲神经网络硬件系统，利用 VCSEL 光子神经元的输入积分、阈值与脉冲激发特性，以超快的纳秒速率演示了重合检测和模式识别任务[78-79]，结合时分复用成功地演示了全光图像边缘特征检测[80]；2024 年，基于多个 VCSEL 神经元构建的光子脉冲神经网络，成功实现并实验证明了超快速率的检测以及图像和视频输入中目标模式的跟踪[81]。除此之外，美国麻省理工学院 Englund D 研究团队在 2023 年实验证明了一个基于 VCSEL 阵列的时空复用光子神经网络系统[82]，采用微米级 VCSEL 阵列作为神经元编码，该阵列具有高效的电光转换和紧凑的排布，并通过零差光电倍增技术，实现了在量子噪声极限下的矩阵运算和具有瞬时响应的基于检测的光学非线性，相比于最先进的数字处理器，该系统的能效和计算密度分别提高了 100 倍和 20 倍。当前，基于半导体激光器的光子神经网络正快速发展，致力于实现低能耗和更高集成度的非线性计算，同时朝着大规模、超高速和高集成度的光学神经网络的构建不断推进。

2.2 基于集成光波导平台的光子神经网络

集成光子平台和器件的逐步成熟[83] 极大推进了基于集成光波导互连技术的光子神经网络的研究，主要包括基于 MRR 阵列的光子神经网络、基于 MZI 网络的光子神经网络和基于 PCM 的光子神经网络，表 1 总结了这三种网络的规模集成度与采用的技术方法。

2.2.1 基于 MRR 阵列的光子神经网络

单个 MRR 通常由一个环形波导和一或两个直波导构成，具有特定的光学谐振特性，可以在特定波长或频率处具有较大的透过率，它尺寸非常小，半径通常为几微米，因此该方法易于扩展，可以做成小规模的阵列。多个 MRR 级联形成的阵列，可以实现非相干的矩阵计算，其中每个微环被调谐到特定的波长或频率，并独立控制这个波长通道上的透射系数，其阵列的规模可以根据矩阵乘运算的规模进行设计。如

表 1　三种光子神经网络的性能对比

Table 1　Performance comparison of three photonic neural networks

网络类型	计算规模/单元	集成度/(单元/mm^2)	技术方法
基于MRR阵列的光子神经网络	~10^3	~10^3	光学谐振
基于MZI网络的光子神经网络	~10^4	~10^3	光学干涉
基于PCM的光子神经网络	~10^3	~10^2	晶态切换

图 4 所示为一个简单的 4×4 MRR 阵列，其完成矩阵乘运算的原理为：输入矢量 $\boldsymbol{X}$ 编码为不同功率的光信号加载在不同的波长上，不同波长的光束经过多个微环并调节透过系数，系数矩阵即为权重矩阵 $\boldsymbol{W}$，最后输出的总功率矢量 $\boldsymbol{Y}=\boldsymbol{WX}$ 即为矩阵乘运算的结果。由于 MRR 在光子集成电路中的紧凑封装、波分复用兼容性和可调谐性等优势，MRR 阵列适合作为片上光子突触并具备可重构特性。

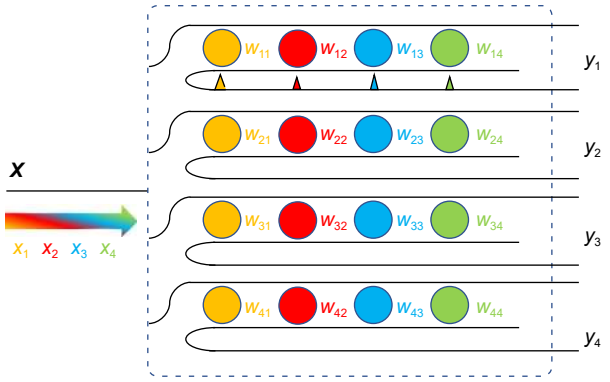


图 4　4×4 MRR 阵列示意图

Fig. 4　Schematic diagram of a 4×4 MRR array

近年来，国内外众多研究团队对基于 MRR 阵列的光子神经网络展开了研究，其进展如图 5 所示。2017 年，普林斯顿大学 Tait 等人开发了世界上首个递归光子神经网络架构[84]，并展示了其超快的计算速度，该光子递归神经网络芯片设计方案，采用 24 个基于电光调制器 (electro-optic modulator，EOM) 的光子神经元节点，16 个 MRR 作为光子权重器实现 4×4 矩阵乘法，执行一个微分系统模拟功能测试时运算速度是传统计算方法的 294 倍。2018 年，乔治华盛顿大学 Mehrabian 等人提出了基于硅基 MRR 和波分复用技术实现了卷积神经网络加速器的方案[85]，运行速度理论上有望比电子卷积神经网络的运行速度快上千倍。2020 年，普林斯顿大学研究团队实现了基于片上 MRR 权重库对盲源信号的光子独立成分分析[86]，同时提出了基于数字电子和模拟光子的卷积神经网络硬件架构，速度有望达到同期最先进的图形处理单元的 2.8 到 14 倍，并节省近 25% 的能量[87]。2021 年，美国 Sunny 等人提出了一种基于 MRR 权重库的硅光神经网络加速器 CrossLight[88]，网络中的卷积层和全连接层分开处理，实现了更高的分辨率、更好的能源效率和更高的吞吐量。2022 年，日本东京大学 Ohno 等人提出了一种基于 4×4 MRR 交叉阵列实现硅基可编程光神经网络[89]，具有 30 TOPS/W 的计算效率，

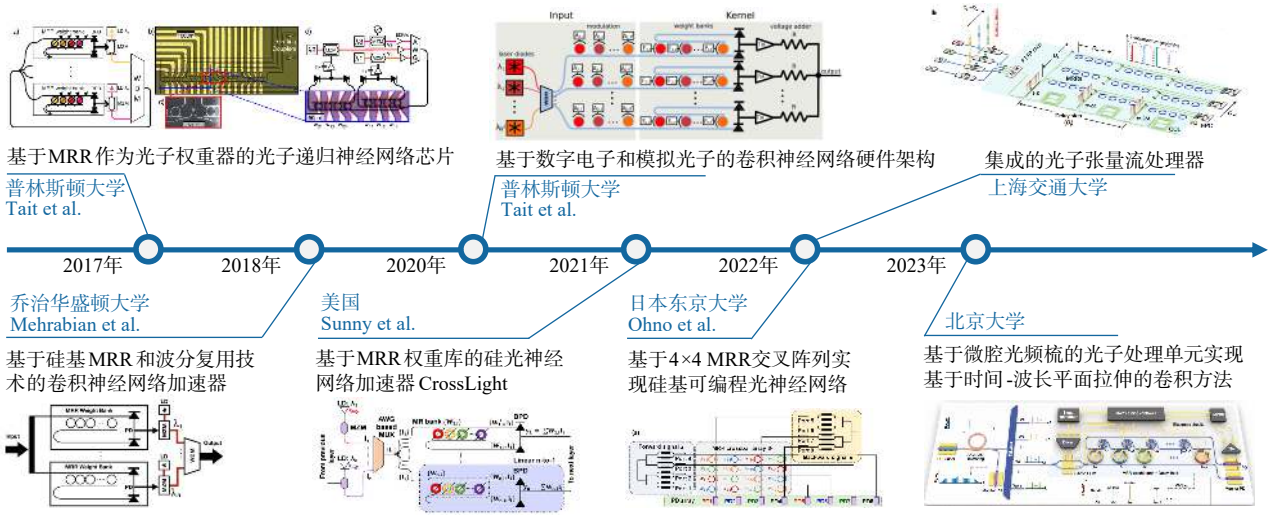


图 5　基于 MRR 阵列的光子神经网络研究进展[84-85,87-91]

Fig. 5　Research progress in photonic neural networks based on MRR arrays[84-85,87-91]

在 Iris 数据集的分类任务中预测准确率为 93%，并利用 MRR 交叉阵列的转置矩阵运算的片上反向传播实现了片上训练。2022 年，上海交通大学邹卫文团队提出了一种集成的光子张量流处理器[90]，通过混合操作光学波长、空间维度和时间延迟，在光域直接表示和处理高阶张量，实现高速张量卷积运算，处理器的内核权重在 MRR 内实现，数据寄存由嵌入式光延迟结构完成，通过重新配置 MRR 的权重和重复使用处理器芯片，实验演示了张量处理；芯片以 20 Gbaud 的调制速率工作，核心部件算力密度为 588 $GOP\cdot s^{-1}\cdot mm^{-2}$，并演示了一种用于视频动作识别的卷积神经网络，其准确率达到 97.9%。2023 年，北京大学王兴军团队提出并演示了一种基于微腔光频梳的光子处理单元[91]，实现基于时间-波长平面拉伸的卷积方法，其中，微腔光频梳提供了一个相干多通道光源，片上硅 EOM 在微腔光频梳的每个梳齿上编码高速数据流 MRR 阵列实现内核权重，嵌入式光学延迟线实现数据缓存，并开发了一个专用的程序精确控制所有单独的组件，实现了 9 bit 的权重精度和 1.04 $TOP\cdot s^{-1}\cdot mm^{-2}$ 的光子核心算力密度；实验验证了该卷积神经网络能够完成图像边缘检测和数字识别分类任务，MNIST 数据集的测试准确率达到 96.6%。

上述研究成果展示了光子神经网络中 MRR 作为光子权重器的强大潜力，通过结合机器学习算法与光子计算的特性，研究者们开发了一系列创新的网络架构，例如基于 MRR 的递归神经网络和卷积神经网络结构，并展示了其在图像识别、数据分类等特定任务中与电子系统竞争甚至超越的能力。在未来基于 MRR 阵列的光子神经网络研究中，首先要增强 MRR 阵列的多功能性与可编程性，使光子神经网络能够动态适应不同的计算任务，实现更高效的资源利用和更广泛的应用范围；其次，应持续提升 MRR 的性能参数，如降低插入损耗和提高调制带宽，并探索更高效的光互连技术，以支持更大规模光子神经网络的构建，满足更复杂算法的需求；最后，利用异质集成技术，高效地将 MRR 与其他光电器件进行整合，实现全光计算系统。

2.2.2 基于 MZI 网络的光子神经网络

MZI 的结构允许调整光程差来实现不同相位的干涉，输入的光信号通过 MZI 时会被分成两束，然后在空间上重新组合，形成一个新的光信号，这个新的光信号可以被认为是输入光信号的线性组合，因此它可以实现光学矩阵运算。MZI 是天然的最小矩阵运算单元，利用 MZI 搭建的光学网络可以扩展至任意维度的任意矩阵计算，如图 6 展示了经典的三角形和矩形结构 4×4 MZI 网络。与 MRR 阵列类似，这些 MZI 网络都可在片上集成，且调制速率都可大于 10 G；不同之处在于，MRR 阵列是显式计算，传输矩阵可以直接生成，而 MZI 网络属于隐式计算，其传输矩阵需要迭代或者解析换算产生。

1994 年，Reck 等人首次证明了使用光学设备 MZI 构造任意有限维酉矩阵进行片上矩阵运算，提出了基于三角形网格结构的酉矩阵分解[92]。2016 年，牛津大学 Clements 等人提出了矩形网格结构的酉矩阵分解法[93]，通过将 MZI 的排布形状从三角形转化为矩形，能够减少一半的光学深度，同时也增加了计算网络的误差容忍度。在此基础上，利用 MZI 完成线性矩阵运算的光子神经网络得到了广泛的研究，近几年的研究进展如图 7 所示。2017 年，麻省理工学院 Shen 等人基于 56 个 MZI 实现了 4×4 矩阵运算[94]，通过在电域引入非线性激活函数来模拟神经元，实现的前馈全连接光神经网络识别四种基本元音的准确率达到 76.7%，比当时最新的电子芯片快两个数量级以上，但使用的能量不到电子芯片的千分之一。同年，乔治华盛顿大学 George 等实现了全光的卷积神经网络[95]，其中利用 MRR 做延时，使用 MZI 网络做矩阵计算。2019 年，加利福尼亚大学伯克利分校 Fang 等人仿真研究了 GridNet (网格网络) 和 FFTNet (快速傅里叶变换网络) 两种光神经网络[96]，都采用了 8×4 的 MZI 线性矩阵运算器，在大规模矩阵运算中具有较高

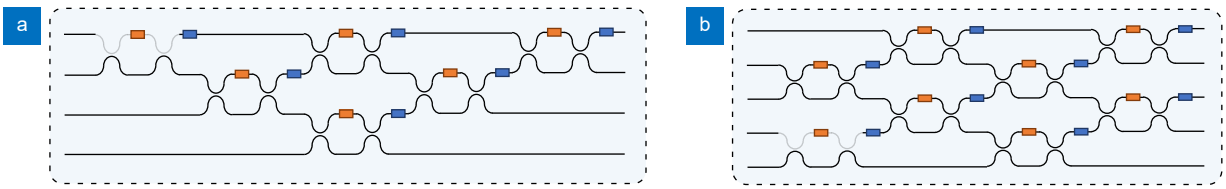


图 6 4×4 MZI 网络。(a) 三角形结构；(b) 矩形结构

Fig. 6 4×4 MZI network. (a) Triangle structure; (b) Rectangular structure

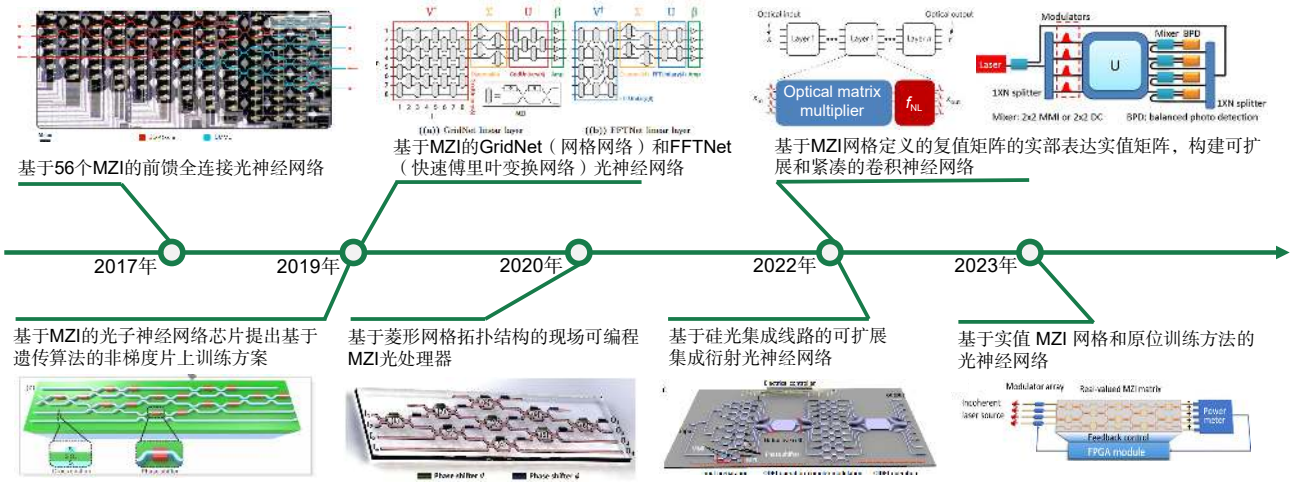


图 7 基于 MZI 网络的光子神经网络研究进展[94,96-97,99-101,103]

Fig. 7 Research progress in photonic neural networks based on MZI mesh[94,96-97,99-101,103]

的计算精度，在手写数据集的分类任务中的分类准确率分别为 98% 和 95%。2019 年，北京邮电大学研究团队研究了基于 MZI 的光子神经网络芯片在线训练算法，提出基于遗传算法/粒子群算法的非梯度片上训练方案[97]，通过仿真分别实现了光神经网络在 Iris 数据集、Wine 数据集上的在线训练。2020 年，麦吉尔大学 Shokraneh 等人提出了一种基于菱形网格拓扑结构的现场可编程 MZI 光处理器[98]，实现了各种尺寸的光神经网络，与三角形网格相比，菱形网格对相位误差和损耗容限有更高的鲁棒性，并通过理论分析发现矩形结构比三角形结构实现的单层 4×4 光神经网络在高损耗和相位误差条件时具有更高的分类精度[99]。2022 年，重庆联合微电子中心田野等人基于 MZI 网格定义的复值矩阵的实部表达实值矩阵，构建了可扩展和紧凑的卷积神经网络，并通过实验测试了 MNIST 识别任务[100]。2022 年，新加坡南洋理工大学研究团队提出了基于硅光集成线路的可扩展集成衍射光神经网络[101]，实现并行傅里叶变换、卷积运算和特定应用的光学计算，N 个输入的网络仅需使用 N 个 MZI 和两个超紧凑衍射单元 (平板波导)，其中 MZI 阵列实现幅度和相位调制，平板波导作为衍射单元实现傅里叶变换和反变换，由于使用了片上紧凑型衍射单元，所提出架构的占地面积和功耗与输入数据维度成线性比例，而不是传统光神经网络框架中的二次比例，占地面积和能耗方面减少了约 10 倍，该网络在 Iris、MNIST 和 Fashion MNIST 数据集上分别获得 98.3%、89.3% 和 81.3% 的测试准确率，与以前基于 MZI 的光神经网络具有同等的高精度。2022 年，华中科技大学张新亮团队提出了一种自监测全光神经网络实现全光激活和非侵入性电监测功能[102]，单芯片包含 16 个 MZI 和 4 个非线性锗硅光电二极管 (PD)，其中 PD 用于构建片上光学神经元，其非线性光学吸收效应实现了全光激活函数，实验中的三层光网络能够在对象分类和语义分割任务中进行原位训练和学习，在 MNIST 数据集上准确率为 97.3%。随后在 2023 年，该团队提出一种用于执行实值非相干光矩阵向量乘的简化 MZI 网格[103]，其性能优于传统的 MZI 网格，并利用该 MZI 网格构建了一个光子神经网络，使用原位粒子群优化训练成功完成了 Iris 分类任务，通过引入匹配的片上非线性激活函数，该 MZI 网格可以级联到单个芯片上，所提出的实值 MZI 网格和原位训练方法具有空间效率高、能效高、可扩展以及对制造误差的鲁棒性等特点，适用于大规模光子神经网络。

综上所述，国内外学者们相继提出了基于 MZI 的光学前馈神经网络、卷积神经网络、衍射神经网络和脉冲神经网络等多种网络架构，在网络中通过电子控制实现动态配置，增强了系统的多功能性和可编程性，使得光子神经网络能够灵活适配各类算法需求，在图像识别、数据分类和语义分割等应用中展示出广阔的应用潜力。此外，为降低 MZI 网络的规模和能耗，积极探索了网络新架构和优化方法；并通过改进 MZI 的设计和制造工艺进行性能优化，提高光子神经网络的计算容量和稳定性，并降低能耗。然而，基于 MZI 的光子神经网络当前面临的主要挑战为尺寸大和实际能耗高，这在一定程度上限制了其在大规

模集成和能耗优化方面的应用。因此，未来的研究需要进一步减小器件尺寸、降低能耗和提高性能，同时解决大规模光子电路中的信号传输与互连问题，实现高效的数据传输；然后，结合非线性光学效应或与电子元件混合集成，以实现更复杂的非线性计算，提升光子神经网络的处理能力；最后，强化算法与硬件的协同优化，设计针对特定任务的高效算法，充分发挥光子神经网络在高速、低耗和高并行度方面的独特优势。

2.2.3 基于 PCM 的光子神经网络

PCM 通过电或光激发在其晶态和非晶态之间实现可逆的快速切换，进而改变光路的传导特性，实现光信号的调制与存储，可以用来编码信息，执行类似神经元激发和突触权重调整的操作，为光子脉冲网络提供了基础构件。由于其快速响应、非易失性以及良好的光电器件兼容性，在硅基光子学集成平台上展现出巨大潜力，通过结合硅光子技术和 PCM，能够构建出集存储和计算于一体的光子神经元和突触结构，实现高效的全光信号处理和光子计算，这对于构建大规模、高速、低功耗的光子神经网络至关重要。

2011 年，Wright 等提出将 PCM 用于算术和生物启发计算[104]，并首次提供了基于 PCM 的处理器的原理实验证明，演示了加、乘、除、减四种基本运算，并同时存储运算结果。同年，Kuzum 等报道了基于 PCM 的新型纳米级电子突触[105]，用于光学数据存储和非易失性存储，PCM 中的连续电阻转换被用来模拟生物突触的特性，从而实现突触学习规则。2017 年，英国牛津大学 Cheng 等人制备了基于 PCM 和集成波导的光子突触芯片[106]，揭示了光子突触权重的调控机制，可使用通过波导的脉冲直接改变突触权重，模拟了突触可塑性。2019 年，美国普渡大学 Chakraborty 等基于相变材料 $Ge_2Sb_2Te_5$ (GST) 研究了非易失性光子突触[107]，突触圆环的最小半径为 1.5 μm，可实现高密度突触阵列。2019 年，德国明斯特大学 Feldmann 等人研发了基于 PCM 和 MRR 的全光学脉冲神经突触网络芯片[108]，避免了传统计算架构中存储单元和处理单元的物理分离，如图 8(a) 所示，其具体工作原理是使用 PCM 单元对输入脉冲进行加权，然后通过 MRR 将相应波长的光耦合到单模波导中进行功率求和，当单模波导中累积的光功率超过某个阈值时，最后一个 MRR 上的 PCM 单元切换晶体状态并产生输出脉冲，从而完成光域的计算，这个芯片上构建了包含 4 个光子脉冲神经元和 60 个突触的网络，可以进行有监督和无监督学习，并成功演示了四个 15 pixel 的字母模式识别任务。随后该团队于 2021 年提出了基于光学张量核的专用集成光子硬件处理器[109]，如图 8(b) 所示，采用基于 PCM 的存内计算架构，利用光学频率梳和波长复用技术实现多路并行处理，使用 16×16 的 PCM 集成阵列实现矩阵-向量乘法运算，运算速度可达每秒 10^{12} 次乘加运算。2023 年，牛津

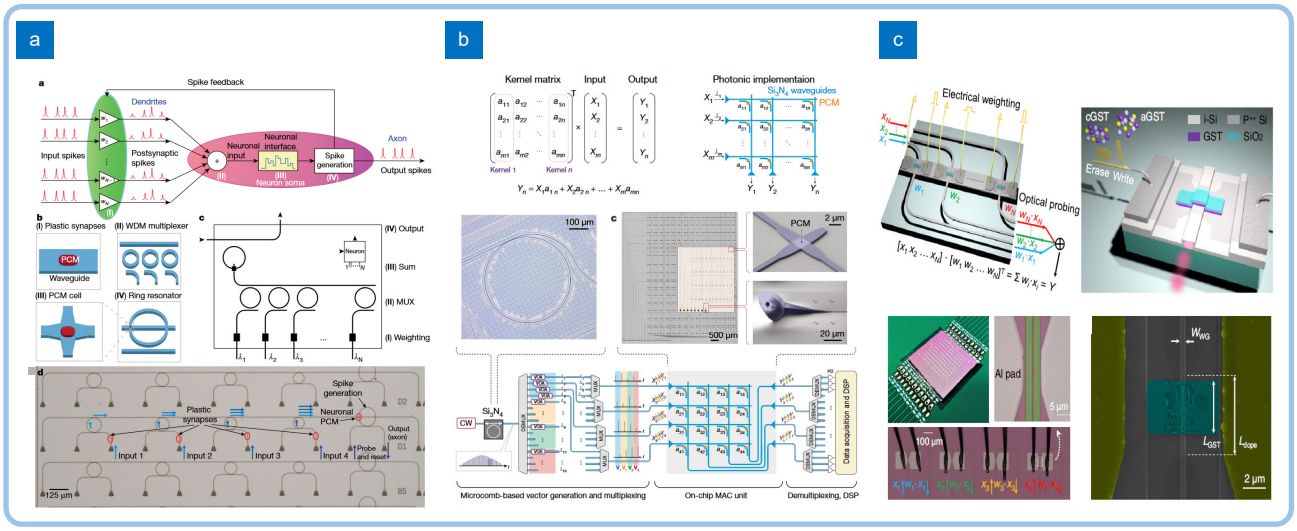


图 8 基于 PCM 的光子神经网络。(a) 全光脉冲神经突触网络的原理和实验[108]；(b) 基于光学张量核的专用集成光子硬件处理器架构[109]；(c) 基于非易失性 PCM 存储单元的存内光电混合计算系统[110]

Fig. 8 Photonic neural network based on PCM. (a) Principle and experiment of all-optical spiking neurosynaptic network[108]; (b) Integrated photonic hardware accelerator architecture based on photonic tensor cores[109]; (c) In-memory photonic–electronic computing platform based on non-volatile electronically reprogrammable PCM memory cells[110]

大学 Zhou 等研究人员展示了一种基于非易失性 GST 存储单元的存内光电混合计算系统[110]，如图 8(c) 所示，其支持灵活的电编程和高通量光学计算，其中电子电路完成权重库的设置和存储，光学电路执行标量乘法和乘加运算，实验揭示了在该非易失性电子可编程 PCM 存储单元作为卷积层的卷积神经网络中，对 MNIST 数据集中手写数字和时尚产品识别的推理准确率分别为 87% 和 86%。

这些研究工作聚焦于脉冲处理与学习机制的探索，通过精确控制 PCM 的相变特性，成功模拟了光子神经网络中的神经元激活与突触连接功能，PCM 的非易失性和可塑性支持了类似于生物神经网络的学习过程，即通过调整相变状态来改变光信号的传递效率，从而实现权重调制；同时致力于将 PCM 器件集成到复杂的光子神经网络架构中，如多层网络和卷积神经网络等，并在算法层面不断创新，以满足光子计算技术的广泛应用需求。接下来，针对基于 PCM 的光子神经网络的研究需要聚焦以下几个方向：一是优化 PCM 材料和器件设计，降低相变过程中的能量消耗，同时提高相变的稳定性和重复性，实现长期稳定运行和大规模应用；二是进一步提升 PCM 光子器件的集成度，开发更复杂的网络架构，并探索更加灵活的编程方法，使网络应对多样化的任务需求；三是结合光器件特性与机器学习算法，开发新的计算范式、网络架构和调控算法。

2.2.4 其他集成光波导平台的光子神经网络

2014 年，比利时根特大学 Vandoorne 等人在绝缘硅平台上实现了包含 16 节点的集成无源硅光子储备池芯片[111]，成功实现了 2 位 XOR 操作、12.5 Gbit/s 的 5 位标头识别和语音数字分类任务。2021 年，澳大利亚斯威本科技大学徐兴元等人提出了一个通用光学向量卷积加速器[112]，使用了 10 个 3×3 卷积核，基于光学频率梳产生 90 个波长光信号，采用波分复用、时分复用、空分复用技术实现光卷积运算加速器，单个处理内核的计算速度超过 10 TOPS，手写数字图像识别准确率达 88%。2022 年，美国宾夕法尼亚大学 Ashtiani 等人报道了一种集成的硅光端到端光学深度神经网络[113]，它使用传播计算来执行亚纳秒图像分类，对手写字母二分类的准确率高于 93.8%，进一步四分类的准确率高于 89.8%。2023 年，清华大学陈宏伟团队提出了一种基于绝缘衬底上的硅 (silicon-on-insulator，SOI) 的片上集成的衍射光学神经网络[114]，可以执行高集成度和低功耗特性的机器学习任务，并提出了一种补偿由芯片制造和实验实现阶段引起的系统误差的算法，从而增加系统的抗噪声能力，实验验证了 1 个隐藏层和 3 个隐藏层的片上 DONN 在 Iris 分类任务中的准确率分别为 86.7% 和 90%，3 个隐藏层的片上 DONN 在 MNIST 数据集上准确率达到 96.3%，该网络克服了空间衍射光子神经网络的体积限制，提高了集成度，与其他集成光子神经网络相比，该芯片摆脱了波导数目的制约，更容易实现计算单元的片上大规模拓展。同年，中国科学院半导体研究所李明团队研制出一款超高集成度光学卷积处理器[115]，通过两个 4×4 多模干涉耦合器和四个移相器构造了三个 2×2 相关的实值卷积核，创新性地将波分复用技术结合光的多模干涉，以波长表征核元素，输入到输出的映射实现了卷积中的乘法运算过程，波分复用和光电转换实现了卷积中的加法运算，通过调节四个热调移相器实现了相关卷积核重构，实验验证了该处理器的手写数字图像特征提取和分类能力，图像特征提取精度达到 5 bit，对 MNIST 手写数字数据集的手写数字进行十分类，准确率达 92.17%。

2.3 总　结

片上集成的光子神经网络具有集成度高、能量效率高和稳定性好等优点，通过光子集成技术，可以在很小的芯片上集成大量的光学元件，实现高密度的信息处理，有利于小型化和便携化设备的开发。同时，由于光子在芯片内直接处理信息，减少了电子与光子之间的转换损耗，整体能效较高，加之标准化的制造工艺保证了器件性能的一致性和可靠性，有利于大规模生产和应用。然而，目前基于片上集成的光子神经网络仍面临制造难度高、成本较大和可调谐性相对不足等挑战，高精度的光子集成芯片制造工艺复杂，初期研发和生产成本相对较高，且一旦制造完成，调整光学参数的灵活性有限，可能影响到算法和网络架构的适应性。因此，这种网络架构适合于追求高集成度、低功耗和稳定性的应用场景。

3 自由空间的光子神经网络

自由空间的光子神经网络利用光在开放空间中的传播和交互来模拟生物神经元和突触的功能，实现神

经形态计算和信息处理，它不依赖于固定的光波导结构，而是通过透镜、反射镜、光栅、空间光调制器等光学元件在三维空间内操纵光束处理信息。

3.1 基于衍射光学元件的光子神经网络

衍射光学元件通常采用微/纳米蚀刻工艺来设计二维分布的衍射单元，每个衍射单元都具有特定的形态和折射率来调节激光波前的相位分布，通过每个衍射单元后，激光在一定距离处衍射和干涉，产生特定的强度分布。图 9 为近年来衍射光子神经网络的研究进展。

2018 年，加州大学洛杉矶分校的 Lin 等人首次提出一种能够进行图像分类的全光深度学习框架——衍射深度神经网络 (diffractive deep neural network，D^2NN)[116]，该架构利用多层衍射表面，并经过计算机训练以光学方式执行任务，采用 3D 打印制造出了这种光学架构，可以完成手写数字和时尚产品的图像识别以及太赫兹光谱成像镜头的功能，该网络对 10000 幅手写数字图像的分类准确率为 93.39%，实现了高准确率和低功耗。然而，该网络的训练仍然是由电子计算机完成的，通过参数化和 3D 打印虚拟衍射层无法实现快速实时编程，实验环境也受到使用太赫兹光源的限制。同年，美国斯坦福大学 Chang 等人提出了一种基于优化的衍射光学元件的光学卷积层[117]，在光电混合卷积神经网络中，光卷积层取代了电神经网络中卷积层所需处理量最大的部分，其余部分仍由计

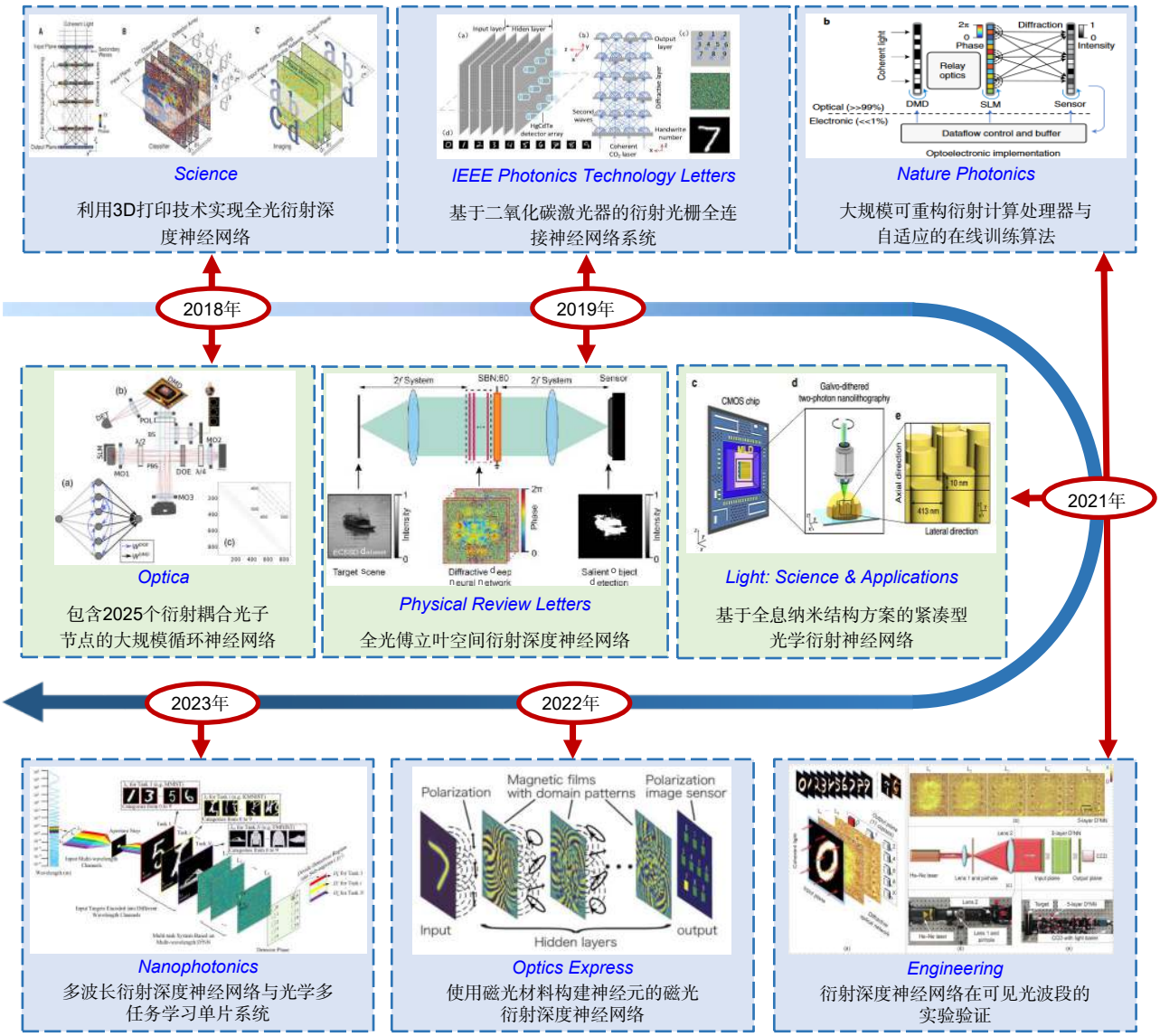


图 9 衍射光子神经网络的研究进展[116-125]

Fig. 9 Research progress in diffractive optical neural network[116-125]

算机处理，采用这样的网络布局，图像处理性能得以提高，卷积层的功耗大致可视为零，从而大大降低了神经网络的总体功耗。此外，西班牙 Bueno 等人实现了一个包含 2025 个衍射耦合光子节点的大规模循环神经网络[118]，其中，空间光调制器将网络节点状态编码到光幅度中，通过偏振分束器和衍射光学元件成像到相机上，从而实现了网络节点之间的耦合，并通过强化学习实现了性能卓越的时间序列预测。

2019 年，天津大学研究团队提出了一种基于二氧化碳激光器的衍射光栅全连接神经网络系统[119]，与 Lin 等人的工作[116] 相比，该系统所采用的光源是波长为 10.6 μm 的长波红外光，比太赫兹光源更容易获得，而且，使用矩阵光栅来代替 3D 打印的衍射层，衍射光栅的像素尺寸从 400 μm 减少到 5 μm (减少了 80 倍)，每层之间的距离减少了 3 倍，D^2NN 的衍射光栅成本更低。同年，清华大学 Yan 等人提出了用于全光图像处理的傅立叶空间衍射深度神经网络 (F-D^2NN)[120]，之前的衍射神经网络的运算都是在实数域完成，该系统将图像进行傅里叶变换后再经过神经网络处理，系统中加入了非线性元件——特殊材料的光折变晶体 (SBN:60)，这个晶体会随着介质中光强度的变化改变折射率，从而对神经元施加复激活函数；对于 MNIST 数据集，5 层 F-D^2NN 的准确率为 96.4%，10 层 F-D^2NN 的准确率为 98.1%，与 D^2NN 相比，F-D^2NN 更复杂，能够以光速执行高级计算机视觉任务 (例如分割)，而且在一些任务上有着更高的正确率和稳定性。

2021 年，哈尔滨工业大学 Chen 等人通过实验验证了 D^2NN 在可见光波段的功能[121]，同时提出了将神经元大小与可见光谱中的波长相关联的公式，在实验中使用石英制作了一个相位掩模，并使用 CCD 相机记录了最终的输出强度，实验结果证明了 D^2NN 应用领域的拓展。同年，清华大学戴琼海团队提出了大规模可重构衍射计算处理器[122]，并提出自适应的在线训练算法，在手写数字图像和人类动作视频分类问题中，实验分类精度与电子计算方法相当。2021 年，上海理工大学研究团队提出了基于全息纳米结构方案的紧凑型光学衍射神经网络完成全光推理[123]，该网络比其他衍射设备和集成光子硬件的算力提高了 3 至 5 个数量级，在单层纳米尺度每平方厘米部署超过 5 亿个神经元，密度达到人类大脑神经元的 1/400。

2022 年，日本长冈技术科学大学 Fujita 等人提出了一种磁光衍射深度神经网络 (MO-D^2NN)[124]，其创新之处在于使用磁光 (magneto-optical, MO) 材料构建神经元，当线性偏振光穿过 MO 材料或从 MO 材料反射时，由于 MO 效应，偏振平面发生旋转和椭圆形，该系统中输入光通过隐藏层中的 MO 效应进行偏振调制和衍射，并使用偏振图像传感器捕获输出信号；MO-D^2NN 对在 MNIST 数据集的分类准确率大于 90%，与其他光学衍射层相反，MO 衍射层是可重写的。

2023 年，Duan 等人采用联合优化方法设计了多波长 D^2NN，提出了一种新的光学多任务学习单片系统[125]，通过将多任务编码到多波长通道中并利用衍射光场的波长维度，所提出的光学多任务学习方法可以以光速并行实现不同的任务；光学多任务功能在单片系统中实现，不需要衍射调制层的机械移动，从而显著降低了系统的复杂性，分析表明，所提出的方法可以显著缓解多任务之间的竞争，并保持每个任务的性能，随着任务数量的增加，多波长 D^2NN 在实现光学多任务学习方面表现出更大的优势；通过使用波分复用技术执行光学多任务学习，该方法可以扩展到其他光子神经网络架构，同时实现高并行、高精度和高通用性的能力。同年，清华大学戴琼海院士团队制造出了一款全模拟光电计算芯片 (all-analog chip combining electronic and light computing，ACCEL)[126]，通过融合光域计算和模拟域电计算来实现神经网络的计算，在光域中，该芯片通过一个多层光学衍射神经网络，针对所输入的高分辨率图像，以光速来进行特征提取和数据降维，然后由一个光电二极管阵列接收衍射网络的输出，并通过光电效应转换成模拟电流信号，通过这种光域处理，极大地减小了数据维度，从而降低光电转换的规模，实现超高算力和超低功耗，其算力达到目前高性能商用芯片的三千余倍，能效达到四百万余倍。

基于衍射光学的深度神经网络架构显著消除了对光电转换元件的需求，可以更有效地实现全光神经网络；此外，衍射神经网络的神经元数量显著增加，网络结构更加复杂，其应用领域更加广泛。然而，目前的衍射神经网络仍需要数字微镜器件等设备来加载图像和需要计算机用于训练网络，而且需要专注于光学非线性元件的研究和应用。

3.2 其他自由空间的光子神经网络

2019 年，香港科技大学研究团队实现了一个具有可调线性操作和非线性光学激活函数的全光神经网络[127]，其中线性运算由空间光调制器和傅里叶透镜

编程，而非线性光学激活函数在具有电磁感应透明度的激光冷却原子模块实现，该硬件系统可以针对不同的应用进行重新配置，而无需修改物理结构；该团队设计了两层全连接的全光神经网络，成功实现对原型 Ising 模型不同相位的分类，证实了其在机器学习应用中的能力和可行性。同年，麻省理工学院 Bernstein 研究团队提出了一种基于相干零差检测的新型光子加速器[128]，该加速器既能实现全连接网络，也能实现卷积网络，它快速、低功耗、结构紧凑、易于扩展到大规模，可利用标准自由空间光学元件实现大规模空间多路复用，其中突触连接的乘加操作具有千兆赫兹的高速度和亚焦耳的极低能量；随后，又提出了一种数字的非相干零差检测的大规模可重构光子神经网络[129]，使用大量的光扇出和基于干涉的光电倍增的组合，更加节能和吞吐量更高。2020 年，法国玛丽居里大学 Rafayelyan 等人提出一种能实现 5 万个光学节点的储备池计算系统[130]，其关键光学部件包括了空间光调制器、散射介质和相机，其中非线性由散射介质提供，该系统成功预测了大型时空混沌数据集；与传统计算机相比，实验研究证实提出的光学方案计算时间几乎与光子节点数量无关。2024 年，清华大学方璐团队首创分布式广度光计算架构，研制出全球首款大规模干涉-衍射异构集成芯片——太极 (Taichi)[131]，该芯片融合衍射光计算大规模并行优势与干涉光计算灵活重构特性，将衍射编解码与干涉特征计算进行部分/整体重构复用，以时序复用突破通量瓶颈，实现了 160 TOPS/W 的能量效率，赋能光计算实现自然场景千类对象识别、跨模态内容生成等人工智能复杂任务。

2020 年，浙江大学 Qian 等人提出了一种基于复合惠更斯超表面的衍射神经网络[132]，展示了所有光学逻辑运算的通用框架，如图 10(a) 所示，该网络中入射平面波首先通过输入层的特定逻辑运算进行空间编码，然后通过隐藏层 (即复合惠更斯元表面) 进行进一步解码，精心设计的超表面将编码光散射到输出层的两个小指定区域之一，提供输出逻辑状态的信息，在一个概念性的微波实验中，在两层电介质超表面上成功地实现了与、或、非三个基本的逻辑运算。2021 年，美国华盛顿大学 Wu 等研究人员展示了一种多模光子核[133]，由基于相变材料制成的相位梯度超表面的可编程波导模式转换器组成，这种相变超表面模式转换器利用 PCM 相变过程中的大折射率变化来以高达 64 级的高模态对比度精度控制波导两个空间模式的转换，在矩阵向量乘计算中以 6 位精度表示权重参数，如图 10(b) 所示，基于该转换器作为可编程内核构建的多模光学卷积神经网络，可以高精度地执行边缘检测和模式识别等图像处理任务，展示了该网络在大规模光学计算中的可行性和潜力。2022 年，东南大学 Liu 等人设计了一种基于多层数字编码超表面阵列的可编程衍射深度神经网络[134]，如图 10(c) 所示，超表面上的每个超构原子与两个放大器芯片集成充当一个活跃的人工神经元，提供 35 dB (从−22 dB 到 13 dB) 的动态调制范围，可编程信息超表面的引入增强了网络的可编程性，通过开发用于现场学习的强化学习算法和用于数字编码的离散优化算法，证明该系统可以处理各种深度学习任务，包括图像分类、移动通信编解码和实时多波束聚焦。

3.3 总　结

自由空间光子神经网络凭借其灵活性高、可扩展性强的特点，能够在不拘泥于固定集成结构的情况下，快速适应算法变化和网络设计创新，通过增加光学组件可以轻松扩大网络规模，不受集成度的限制。然而，这种网络也面临着尺寸较大不利于小型化和集成，对环境因素敏感导致稳定性和精度问题，以及光线在自由空间传播可能存在的损耗导致能效相对较低等挑战，限制了其在小型化设备和低能耗场景的应用。因此，自由空间光子神经网络适用于快速原型设计和算法探索以及处理大规模数据集。

4　光子神经网络的训练

训练是神经网络构建过程中的一个重要步骤，它决定着整个系统的性能，光子神经网络的训练问题是扩展光子神经网络应用的制约因素。目前的光子神经网络的训练方法可分为两类：硬件感知的非原位训练和片上原位训练，其研究进展如图 11 所示。

非原位训练是指在数字计算机的帮助下进行的训练，硬件感知的非原位训练将光神经网络的训练过程转移到数字计算机上，并利用各种线路感知训练技术，在训练过程中尽可能精确地捕捉线路行为并对其进行建模，同时考虑各种非理想效果，然后在训练过程中注入光子线路模型，以减少训练和真实推理之间的差距。2020 年，Gu 等人提出一种噪声感知量化方案[135]，使光子神经网络适应低精度控制和具有移相器噪声的非理想环境，该方案通过粗梯度近似和酉投影，实现光子神经网络的低精度电压控制，并缓解相应的精度

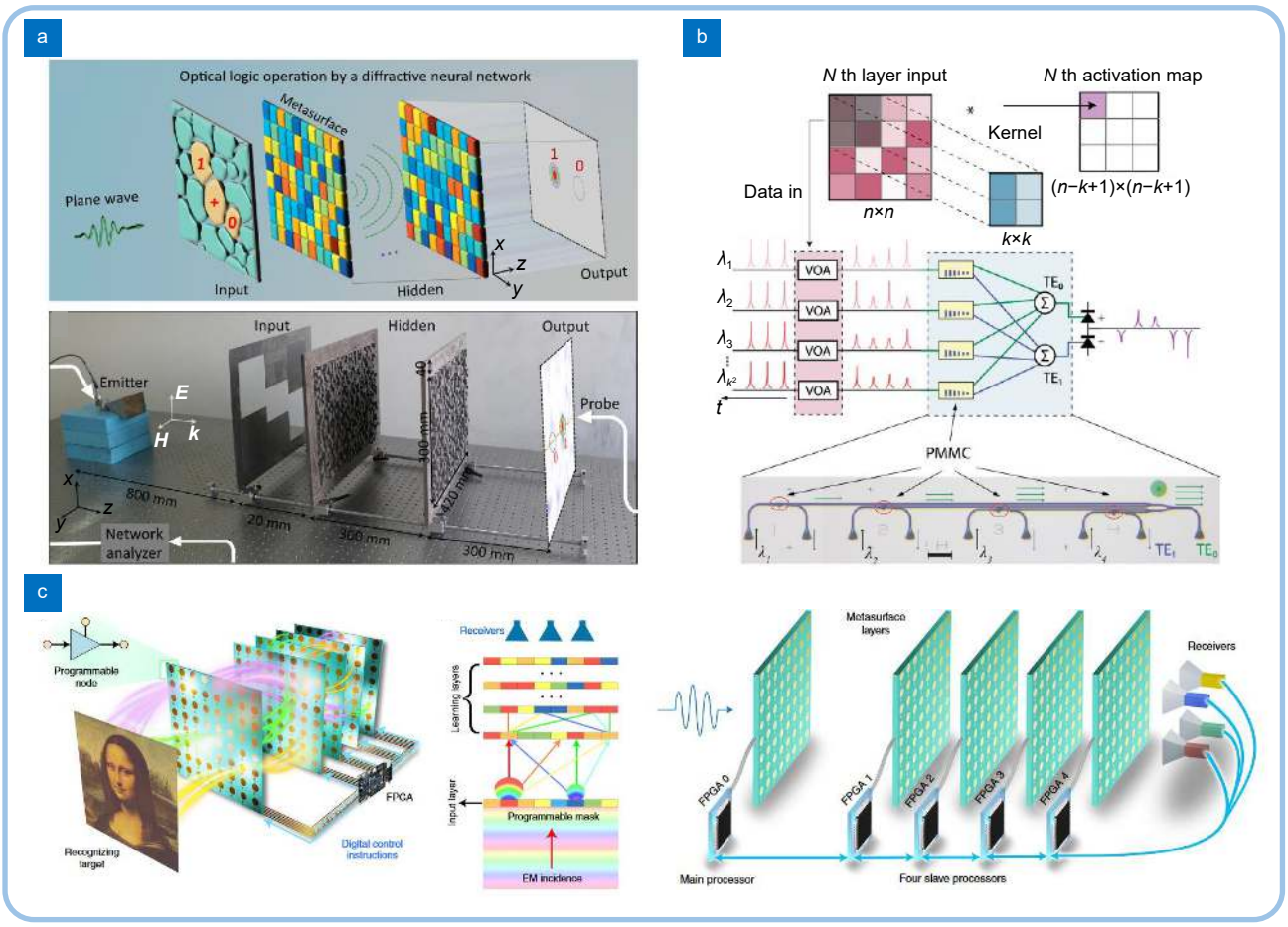


图 10 基于超表面的光子神经网络。(a) 基于复合惠更斯超表面的衍射神经网络实现光学逻辑运算的原理和实验[132]; (b) 基于相变超表面可编程模式转换器阵列作为光子计算核心的光子卷积神经网络[133]; (c) 基于多层数字编码超表面阵列的可编程衍射深度神经网络[134]

Fig. 10 Photonic neural network based on metasurfaces. (a) The principle and experiment of optical logic operations performed by a diffractive neural network based on a compound Huygens' metasurface[132]; (b) Optical convolutional neural network based on the phase-change metasurface mode converter as a photonic computing core[133]; (c) A programmable diffractive deep neural network based on a multi-layer digital-coding metasurface array[134]

下降，同时该团队还提出一种保护群 Lasso 正则化技术，通过动态抑制鲁棒性较差的权重矩阵块来保护网络免受移相器噪声的影响；实验结果表明，在各种控制精度和设备噪声下，基于噪声感知训练的四层光子神经网络比基线方法具有更高的推理精度和更低的方差。2021 年，Mourgias-Alexandris 等基于相干硅集成电路将噪声容忍线性神经元架构方案与噪声感知训练方法相结合[136]，实现了高性能的光子深度学习模型，该模型与最先进的相干网络相比，每个轴突的片上计算速率和分类精度高出 6 个数量级和 7%，与 Gu 等人的工作[135] 相比，本工作使用基于双 IQ 调制器的相干线性神经元架构，其中单个片上权值简单地由相移和幅度调制元件定义，与具有级联 MZI 的相应相干布局相比，该方式显著提高了噪声容忍特性。2022 年，Kirtas 等人提出了一种量化感知训练框架来训练具有有限精度的光子深度学习模型[137]，有效降低光子深度学习模型的精度要求，从而减少对昂贵的高速和高精度模数转换/数模转换的需求，并降低与这些组件相关的硬件成本，在较低的比特分辨率下显著提高了模型性能，在各种任务和不同的光子结构中，包括全连接、卷积和递归网络，实验证明了所提出的方法的有效性，该方法应用于上述 Mourgias-Alexandris 研究团队的工作[136] 中可以降低硬件的复杂性和成本，因此，该方法提供了一种更全面的方法来训练光神经网络，在提高模型的鲁棒性和性能的同时降低硬件的复杂性和成本。2022 年，Feng 等人开发了一个多级硬件感知训练框架[138]，利用测量数据和人工智能算法开发了一种基于神经网络的可微光子集成线路估计

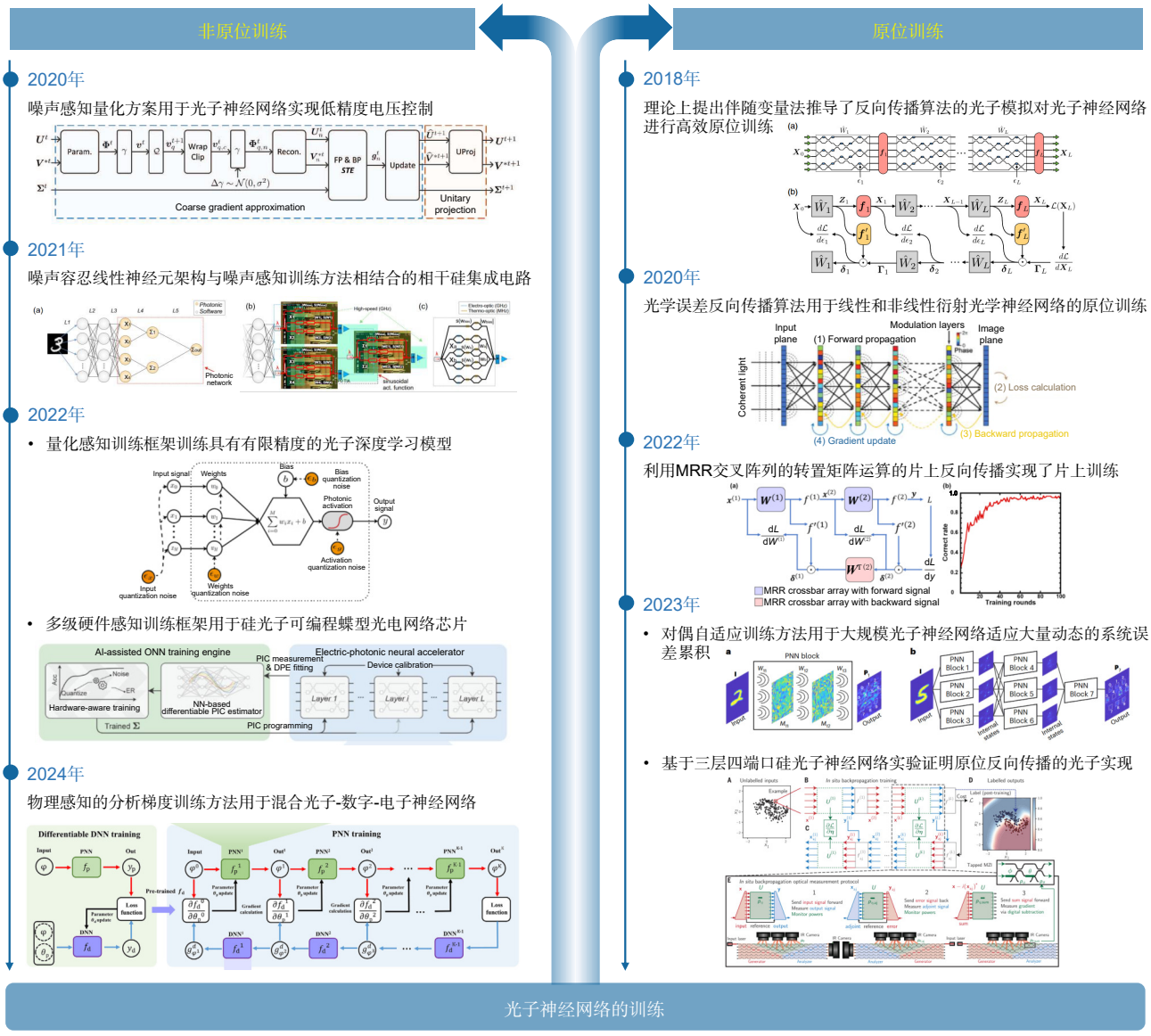


图 11 光子神经网络的训练的研究进展[89,135-144]

Fig. 11 Research progress in training optical neural networks[89,135-144]

器，在正向和梯度反向传播过程中对实际物理芯片的行为进行了明确的建模，实现基于梯度的物理变化感知优化，对网络中非理想行为进行建模并预测真实光学神经芯片的响应，结合该训练框架，在硅光子可编程蝶型光电网络芯片上对提出的硬件高效的光学子空间神经网络结构进行了实验验证，最大限度地降低了所需的设备编程精度、减少了芯片面积和提高了噪声鲁棒性，与前面的工作[135-137]不同，所提出的训练框架部署在光学子空间神经网络，通过使用更小的光学组分以及更快、更高效的光电转换技术，进一步提高网络的性能，突破了光神经网络的可扩展性和完整性的极限，并在提高硬件效率的情况下为下一代高性能 AI 加速器创造了新的设计范式。2024 年，Zhan 等人提出一种物理感知的分析梯度训练方法用于混合光子-数字-电子神经网络[139]，该方法以分治策略计算分析梯度，使用光子神经网络执行前向传递，通过学习光子神经网络的独特物理变换建立可微分的数字神经网络模型，利用该数学神经网络模型来计算分析梯度用于反向传播，补充物理系统无法完成的训练循环部分，该方法利用可微数学神经网络来表示不可微光子神经网络，克服了光神经网络训练中芯片不可微性带来的困难，简化了混合光子数字网络的训练过程并降低了成本，使用 8×8 光子神经网络芯片上对训练算法进行了验证，在各种机器学习任务中展现出巨大的

潜力。

原位训练旨在直接在芯片上进行训练，从而能够考虑各种芯片上的非理想性，这种方法通过将光子硬件的真实行为直接纳入训练过程，最大限度地提高精度。2018 年，斯坦福大学 Hughes 等人在理论上提出伴随变量法推导了反向传播算法的光子模拟[140]，能够对光子神经网络进行高效原位训练，该方法通过干扰伴随场和正向场来实现片上反向传播，直接测量出梯度信息作为原位强度测量，可以以高效和可扩展的方式直接在设备内部进行训练，该方法广泛适用于集成和自适应光学系统，使自配置和自动优化成为可能。2020 年，清华大学 Zhou 等人提出光学误差反向传播算法用于线性和非线性衍射光学神经网络的原位训练[141]，基于光互易性和相位结合原理测量正向和反向传播的光场，准确地计算损耗函数相对于衍射层权重的梯度，从而加快训练速度，提高核心计算模块的能量效率，基于可重构衍射光神经网络系统数值模拟验证了所提出的原位光学训练方法在物体分类和矩阵矢量乘法任务中实现了与使用电子计算机进行训练相当的精度，所提出的原位光学训练系统存在顺序读入模式和现有空间光调制器成本较高的局限性，随着集成光子学中可编程片上光电子设备的出现可以缓解这些问题。2022 年，Ohno 等人利用 MRR 交叉阵列的转置矩阵运算的片上反向传播实现了片上训练[89]，但它仅在数值模拟中实现，没有芯片上的演示。在 Hughes 等人和 Zhou 等人提出的原位训练方法[140-141]中，用于反向光场传播的额外硬件配置可能会导致反向传播过程中的梯度计算不准确，使得原位训练方法在训练具有重大系统误差的大规模光神经网络方面仍然面临巨大挑战，阻碍了先进架构的构建，并限制了模型在执行复杂人工智能任务时的性能。因此，2023 年，清华大学 Lin 等人提出了大规模光子神经网络的对偶自适应训练方法[142]，通过网络的精准建模和对偶反向传播，使网络能够适应大量动态的系统误差累积，并在部署过程中保持其性能，实现任务推理性能的巨大提升，在动态系统误差环境下，成功训练了包含 28 万神经元的光子神经网络，在分类任务上的训练性能大幅优于当前训练方法。同年，Hughes 研究团队在 2018 年提出的原位训练算法[140]基础上，构建了一个具有可编程移相器和光功率监测的三层四端口硅光子神经网络芯片，基于该芯片实验证明了原位反向传播的光子实现[144]，通过干扰正向和反向传播的光来测量移相器电压的反向传播梯度，使用原位反向传播解决分类任务，在给定误差的情况下模拟了在 64 端口光子神经网络使用原位反向传播训练 MNIST 图像的识别，测试准确率大于 94%，与数字模拟计算相当，然而，这种实现方式需要额外的功率/相位监测器、更快的微控制器和更精确的检测器，增加了硬件控制的复杂性并带来了可扩展性问题。

硬件感知的非原位训练不直接在硬件上进行训练计算，仅涉及到模型在硬件上的部署和调整，通过在训练过程中动态调整硬件级和模型级性能设置，以平衡模型质量和训练速度；由于它并不直接在硬件上进行计算，可能无法充分利用硬件的并行性和效率，因此，这种训练方法更适用于需要在多种硬件平台上进行训练，并需要在训练过程中平衡模型质量和训练速度的场景。片上原位训练则直接在硬件上进行神经网络的训练，通过使用特定的算法，可以在硬件上直接计算每个神经网络层的梯度向量，充分利用硬件的并行性和效率，从而实现高效的并行化和快速的训练速度，还考虑到了硬件的非理想性，使训练对噪声具有鲁棒性。然而，这种训练方法可能受限于特定硬件的特性和容量，对于大型模型或数据集可能不太适用，它更适用于需要超快训练速度且对硬件的并行性和效率有较高要求的场景。

5 总结与展望

光子神经网络作为人工智能技术与光子技术的交叉学科产物，通过融合两个领域的优势，打破传统电子神经网络存在的性能瓶颈，构建出高速低功耗的新型计算架构。本文系统介绍了片上集成与自由空间两种光子神经网络的研究进展，详细介绍了该领域内典型的研究工作。其中，自由空间的光子神经网络具备潜在的大带宽和并行处理能力，将部分计算转化为光学处理以减轻软件和硬件的压力，同时降低成本和节省时间，适用于大规模计算任务。然而，自由空间中的光学神经网络当前仍面临多重挑战，首要问题是其训练和构建仍需借助计算机和光电转换设备，其中，非线性实现的难题尤为突出。此外，光束在自由空间中传播时，能量利用效率低下，且易受环境因素影响，导致其稳定性较差以及控制复杂度较高。随着光纤通信互连、纳米光学等技术的日趋成熟，用于神经计算的硅光集成电路和相变材料得到了快速发展，相较于

自由空间的光子神经网络，片上集成的光子神经网络具有高能效、紧凑的面积布局以及灵活的可编程性等诸多优势，然而，在大规模集成的过程中，如何确保准确性和稳健性，以更好地应对日益复杂和大规模的计算任务，仍是当前亟待解决的问题。

当前，人工智能技术正蓬勃发展，其在机器视觉、智能感知及自动驾驶等多个领域都展现出强大的应用潜力，光子神经网络作为一种新型人工智能计算架构，其应用领域正逐渐从简单的模式识别应用拓展至更广泛复杂的应用。本综述归纳总结了目前已报道的光学神经网络架构在人工智能多个特定应用场景的表现，包括 MNIST 手写数字识别、图像边缘检测、鸢尾花识别、元音识别及非线性逻辑运算等典型任务，展现了其在人工智能领域的应用潜力。相较于电神经网络，光神经网络展现出了更低的功率消耗、更少的参数设置以及更快的计算速度。在应对如 MNIST 数据集等简单应用场景时，光神经网络所表现出的准确性足以与电神经网络相媲美。然而，当面对复杂数据集时，光神经网络的准确性仍需经过更为严格的验证。因此，在光神经网络真正能够广泛投入实际应用之前，尚需经历一段漫长的探索与验证过程。

光子神经网络是光子学、人工智能和光电集成等多个领域的交叉融合的产物，相较于已经相对成熟的电子神经网络，其在可扩展性、集成度、规模化封装、小型化、实用化和可训练性等方面仍有很大的进步空间。特别是光电器件固有的非理想性和低稳定性特性，对光子神经网络的可训练性、集成度和规模化构成了严峻的挑战，导致构建功能复杂的神经网络模型面临更严苛的要求，反过来，这些性能问题也限制了光子神经网络在特定领域充分发挥其独特优势。因此，光子神经网络的发展迫切需要克服以下挑战，从根本上突破其发展瓶颈。具体而言，在高性能大规模光子集成领域，面临着大规模异质异构集成、光电融合集成等先进工艺、高效调制与高灵敏探测等难题；在高密度光电封装方面，则需应对百量级光阵列耦合、百量级高速电封装、高密度互连等挑战；在光电速率匹配方面，需要解决高速率、低功耗模数转换、光/电时分复用与解复用等问题；此外，在驱动软件与应用开发方面，亦需要开发先进的系统与驱动软件，以构建完善的软件生态。

展望未来，光子神经网络正呈现出一种多层次、跨领域、全方位的发展格局，从材料器件、系统架构到算法实现，形成了一条自下而上、层层递进、互为支撑的技术发展脉络。在器件层面，致力于进一步降低光脉冲神经元和光突触器件的单元功耗、提高速率，并努力开发大规模光脉冲神经元和光突触阵列，以增强网络规模的可扩展性；在系统架构层面，积极探索光脉冲神经元与光突触芯片的异质异构集成技术，并深入研究光计算芯片与电控制芯片的光电融合集成，力求在硬件层面实现多层光子神经网络的构建，并实现光神经网络芯片的小型化；在算法层面，专注于深度脉冲神经网络算法的研究及其应用的拓展，以期实现从专用计算到通用计算的跨越。这样，不仅能够深度挖掘光子特性的潜能，更可以实现了与人工智能算法的深度融合，展现出光子神经网络在构建新型智能计算体系中的广阔前景。

综上，光子神经网络将光子的高并行、高速传输、低能耗等特性与神经网络的强大计算能力完美结合，在图像分类、目标检测和动作识别等计算机视觉任务，以及文本分类、语音识别和情感分析等自然语言处理任务方面，展现出前所未有的潜力和优势，进而更广泛地应用于高性能计算与数据中心、图像处理、自动驾驶和生物医学等复杂场景，为这些领域带来颠覆性变革，并将推动光子神经网络这一全新计算范式的发展，引领下一代人工智能计算科技迈向更加高效、智能的未来。

参考文献

[1] McCarthy J, Minsky M L, Rochester N, et al. A proposal for the Dartmouth summer research project on artificial intelligence: August 31, 1955[J]. *AI Mag*, 2006, **27**(4): 12−14.

[2] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, **521**(7553): 436−444.

[3] Moore G E. Cramming more components onto integrated circuits[J]. *Electronics*, 1965, **38**(8): 114–117.

[4] Mead C. Neuromorphic electronic systems[J]. *Proc IEEE*, 1990, **78**(10): 1629−1636.

[5] Amunts K, Ebell C, Muller J, et al. The human brain project: creating a European research infrastructure to decode the human brain[J]. *Neuron*, 2016, **92**(3): 574−581.

[6] Insel T R, Landis S C, Collins F S. The NIH BRAIN initiative[J]. *Science*, 2013, **340**(6133): 687−688.

[7] Martin C L, Chun M. The BRAIN initiative: building, strengthening, and sustaining[J]. *Neuron*, 2016, **92**(3): 570−573.

[8] Ngai J. BRAIN 2.0: transforming neuroscience[J]. *Cell*, 2022, **185**(1): 4−8.

[9] Okano H, Sasaki E, Yamamori T, et al. Brain/MINDS: a Japanese national brain project for marmoset neuroscience[J].

Neuron, 2016, **92**(3): 582−590.

[10] Poo M M. Whereto the mega brain projects?[J]. *Natl Sci Rev*, 2014, **1**(1): 12−14.

[11] Poo M M, Du J L, Ip N Y, et al. China brain project: basic neuroscience, brain diseases, and brain-Inspired computing[J]. *Neuron*, 2016, **92**(3): 591−596.

[12] Poo M M, Xu B, Tan T N. Brain science and brain-inspired intelligence technolog—an overview[J]. *Bull Chin Acad Sci*, 2016, **31**(7): 725−736
蒲慕明, 徐波, 谭铁牛. 脑科学与类脑研究概述[J]. 中国科学院院刊, 2016, **31**(7): 725−736

[13] Huang T J, Shi L P, Tang H J, et al. Research on multimedia technology 2015——advances and trend of brain-like computing[J]. *J Image Graphics*, 2016, **21**(11): 1411−1424.
黄铁军, 施路平, 唐华锦, 等. 多媒体技术研究: 2015——类脑计算的研究进展与发展趋势[J]. 中国图象图形学报, 2016, **21**(11): 1411−1424.

[14] Xiang S Y, Song Z W, Gao S, et al. Progress and prospects of photonic neuromorphic computing (Invited)[J]. *Acta Photonica Sin*, 2021, **50**(10): 1020001.
项水英, 宋紫薇, 高爽, 等. 光神经形态计算研究进展与展望 (特邀)[J]. 光子学报, 2021, **50**(10): 1020001.

[15] Painkras E, Plana L A, Garside J, et al. SpiNNaker: a 1-W 18-core system-on-chip for massively-parallel neural network simulation[J]. *IEEE J Solid-State Circuits*, 2013, **48**(8): 1943−1953.

[16] Benjamin B V, Gao P R, McQuinn E, et al. Neurogrid: a mixed-analog-digital multichip system for large-scale neural simulations[J]. *Proc IEEE*, 2014, **102**(5): 699−716.

[17] Merolla P A, Arthur J V, Alvarez-Icaza R, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface[J]. *Science*, 2014, **345**(6197): 668−673.

[18] Schemmel J, Brüderle D, Grübl A, et al. A wafer-scale neuromorphic hardware system for large-scale neural modeling[C]//*2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2010: 1947–1950.
https://doi.org/10.1109/ISCAS.2010.5536970.

[19] Ma D, Shen J C, Gu Z H, et al. Darwin: a neuromorphic hardware co-processor based on spiking neural networks[J]. *J Syst Archit*, 2017, **77**: 43−51.

[20] Davies M, Srinivasa N, Lin T H, et al. Loihi: a neuromorphic manycore processor with on-chip learning[J]. *IEEE Micro*, 2018, **38**(1): 82−99.

[21] Orchard G, Frady E P, Rubin D B D, et al. Efficient neuromorphic signal processing with loihi 2[C]//*2021 IEEE Workshop on Signal Processing Systems (SiPS)*, 2021: 254–259.
https://doi.org/10.1109/SiPS52927.2021.00053.

[22] Shi L P, Pei J, Deng N, et al. Development of a neuromorphic computing system[C]//*2015 IEEE International Electron Devices Meeting (IEDM)*, 2015: 4.3.1–4.3.4.
https://doi.org/10.1109/IEDM.2015.7409624.

[23] Liu Z S, Chen S, Qu P Y, et al. SUSHI: ultra-high-speed and ultra-low-power neuromorphic chip using superconducting single-flux-quantum circuits[C]//*Proceedings of the 56th Annual IEEE/ACM International Symposium on Microarchitecture*, 2023: 614–627.

[24] Miller D. Device requirements for optical interconnects to silicon chips[J]. *Proc IEEE*, 2009, **97**(7): 1166−1185.

[25] Nahmias M A, De Lima T F, Tait A N, et al. Photonic multiply-accumulate operations for neural networks[J]. *IEEE J Sel Top Quantum Electron*, 2020, **26**(1): 7701518.

[26] Tait A N, Nahmias M A, Tian Y, et al. Photonic neuromorphic signal processing and computing[M]//Naruse M. *Nanophotonic Information Physics: Nanointelligence and Nanophotonic Computing*. Berlin: Springer, 2014: 183–222.
https://doi.org/10.1007/978-3-642-40224-1_8.

[27] Shastri B J, Chang J, Tait A N, et al. Ultrafast optical techniques for communication networks and signal processing[M]//Wabnitz S, Eggleton B J. *All-Optical Signal Processing: Data Communication and Storage Applications*. Cham: Springer, 2015: 469–503.
https://doi.org/10.1007/978-3-319-14992-9_15.

[28] Hopfield J J. Neural networks and physical systems with emergent collective computational abilities[J]. *Proc Natl Acad Sci*, 1982, **79**(8): 2554−2558.

[29] Liu J, Wu Q H, Sui X, et al. Research progress in optical neural networks: theory, applications and developments[J]. *PhotoniX*, 2021, **2**(1): 5.

[30] Tsai F C F, O'Brien C J, Petrović N S, et al. Analysis of optical channel cross talk for free-space optical interconnects in the presence of higher-order transverse modes[J]. *Appl Opt*, 2005, **44**(30): 6380−6387.

[31] Hu W H, Li X J, Yang J K, et al. Crosstalk analysis of aligned and misaligned free-space optical interconnect systems[J]. *J Opt Soc Am A*, 2010, **27**(2): 200−205.

[32] Xiang S Y, Wen A J, Pan W. Emulation of spiking response and spiking frequency property in VCSEL-based photonic neuron[J]. *IEEE Photonics J*, 2016, **8**(5): 1−9.

[33] Xiang S Y, Zhang H, Guo X X, et al. Cascadable neuron-like spiking dynamics in coupled VCSELs subject to orthogonally polarized optical pulse injection[J]. *IEEE J Sel Top Quantum Electron*, 2017, **23**(6): 1−7.

[34] Xiang S Y, Zhang Y H, Guo X X, et al. Photonic generation of neuron-like dynamics using VCSELs subject to double polarized optical injection[J]. *J Lightwave Technol*, 2018, **36**(19): 4227−4234.

[35] Zhang Y H, Xiang S Y, Gong J K, et al. Spike encoding and storage properties in mutually coupled vertical-cavity surface-emitting lasers subject to optical pulse injection[J]. *Appl Opt*, 2018, **57**(7): 1731.

[36] Zhang Y H, Xiang S Y, Guo X X, et al. Polarization-resolved and polarization- multiplexed spike encoding properties in photonic neuron based on VCSEL-SA[J]. *Sci Rep*, 2018, **8**(1): 16095.

[37] Zhang Y, Xiang S, Guo X, et al. All-optical inhibitory dynamics in photonic neuron based on polarization mode competition in a VCSEL with an embedded saturable absorber[J]. *Opt Lett*, 2019, **44**(7): 1548−1551.

[38] Xiang S Y, Ren Z X, Zhang Y H, et al. All-optical neuromorphic XOR operation with inhibitory dynamics of a

single photonic spiking neuron based on a VCSEL-SA[J]. *Opt Lett*, 2020, **45**(5): 1104−1107.

[39] Xiang S Y, Gong J K, Zhang Y H, et al. Numerical implementation of wavelength-dependent photonic spike timing dependent plasticity based on VCSOA[J]. *IEEE J Quantum Electron*, 2018, **54**(6): 8100107.

[40] Song Z W, Xiang S Y, Cao X Y, et al. Experimental demonstration of photonic spike-timing-dependent plasticity based on a VCSOA[J]. *Sci China Inf Sci*, 2022, **65**(8): 182401.

[41] Xiang S Y, Han Y N, Guo X X, et al. Real-time optical spike-timing dependent plasticity in a single VCSEL with dual-polarized pulsed optical injection[J]. *Sci China Inf Sci*, 2020, **63**(6): 160405.

[42] Xiang S Y, Zhang Y H, Gong J K, et al. STDP-based unsupervised spike pattern learning in a photonic spiking neural network With VCSELs and VCSOAs[J]. *IEEE J Sel Top Quantum Electron*, 2019, **25**(6): 1700109.

[43] Xiang S Y, Ren Z X, Song Z W, et al. Computing primitive of fully VCSEL-based all-optical spiking neural network for supervised learning and pattern classification[J]. *IEEE Trans Neural Networks Learn Syst*, 2021, **32**(6): 2494−2505.

[44] Fu C T, Xiang S Y, Han Y N, et al. Multilayer photonic spiking neural networks: generalized supervised learning algorithm and network optimization[J]. *Photonics*, 2022, **9**(4): 217.

[45] Zhang Y H, Xiang S Y, Guo X X, et al. The winner-take-all mechanism for all-optical systems of pattern recognition and max-pooling operation[J]. *J Lightwave Technol*, 2020, **38**(18): 5071−5077.

[46] Han Y N, Xiang S Y, Ren Z X, et al. Delay-weight plasticity-based supervised learning in optical spiking neural networks[J]. *Photonics Res*, 2021, **9**(4): B119−B127.

[47] Song Z W, Xiang S Y, Ren Z X, et al. Photonic spiking neural network based on excitable VCSELs-SA for sound azimuth detection[J]. *Opt Express*, 2020, **28**(2): 1561−1573.

[48] Song Z W, Xiang S Y, Ren Z X, et al. Spike sequence learning in a photonic spiking neural network consisting of VCSELs-SA with supervised training[J]. *IEEE J Sel Top Quantum Electron*, 2020, **26**(5): 1700209.

[49] Wang S H, Xiang S Y, Han G Q, et al. Photonic associative learning neural network based on VCSELs and STDP[J]. *J Lightwave Technol*, 2020, **38**(17): 4691−4698.

[50] Zhang Y H, Xiang S Y, Guo X X, et al. A modified supervised learning rule for training a photonic spiking neural network to recognize digital patterns[J]. *Sci China Inf Sci*, 2021, **64**(2): 122403.

[51] Gao S, Xiang S Y, Song Z W, et al. All-optical Sudoku solver with photonic spiking neural network[J]. *Opt Commun*, 2021, **495**: 127068.

[52] Gao S, Xiang S Y, Song Z W, et al. Motion detection and direction recognition in a photonic spiking neural network consisting of VCSELs-SA[J]. *Opt Express*, 2022, **30**(18): 31701−31713.

[53] Xiang S Y, Ren Z X, Zhang Y H, et al. Training a multi-layer photonic spiking neural network with modified supervised learning algorithm based on photonic STDP[J]. *IEEE J Sel Top Quantum Electron*, 2021, **27**(2): 7500109.

[54] Zhang Y H, Xiang S Y, Han Y N, et al. BP-based supervised learning algorithm for multilayer photonic spiking neural network and hardware implementation[J]. *Opt Express*, 2023, **31**(10): 16549−16559.

[55] Song Z W, Xiang S Y, Zhao S H, et al. A multi-layer photonic spiking neural network with a modified backpropagation algorithm for nonlinear classification[J]. *Opt Commun*, 2023, **546**: 129806.

[56] Xiang S Y, Zhang T R, Han Y N, et al. Neuromorphic speech recognition with photonic convolutional spiking neural networks[J]. *IEEE J Sel Top Quantum Electron*, 2023, **29**(6): 7600507.

[57] Han Y N, Xiang S Y, Zhang Y N, et al. An all-MRR-based photonic spiking neural network for spike sequence learning[J]. *Photonics*, 2022, **9**(2): 120.

[58] Zhang Y N, Xiang S Y, Han Y N, et al. Supervised learning and pattern recognition in photonic spiking neural networks based on MRR and phase-change materials[J]. *Opt Commun*, 2023, **549**: 129870.

[59] Song Z W, Xiang S Y, Zhao S T, et al. A hybrid-integrated photonic spiking neural network framework based on an MZI array and VCSELs-SA[J]. *IEEE J Sel Top Quantum Electron*, 2023, **29**(2): 8300211.

[60] Zheng D Z, Xiang S Y, Guo X X, et al. Experimental demonstration of coherent photonic neural computing based on a Fabry–Perot laser with a saturable absorber[J]. *Photonics Res*, 2023, **11**(1): 65−71.

[61] Song Z W, Xiang S Y, Guo X X, et al. Nonlinear neural computation in an integrated FP-SA spiking neuron subject to incoherent dual-wavelength optical pulse injections[J]. *Sci China Inf Sci*, 2023, **66**(12): 229405.

[62] Xiang S Y, Shi Y C, Guo X X, et al. Hardware-algorithm collaborative computing with photonic spiking neuron chip based on an integrated Fabry–Perot laser with a saturable absorber[J]. *Optica*, 2023, **10**(2): 162−171.

[63] Guo X X, Xiang S Y, Zhang Y H, et al. Hardware implementation of multi-layer photonic spiking neural network with three cascaded photonic spiking neurons[J]. *J Lightwave Technol*, 2023, **41**(20): 6533−6541.

[64] Han Y N, Xiang S Y, Gao S, et al. Experimental demonstration of delay-weight learning and pattern classification with a FP-SA-based photonic spiking neuron chip[J]. *J Lightwave Technol*, 2024, **42**(5): 1497−1503.

[65] Zhang Y H, Xiang S Y, Guo X X, et al. Spiking information processing in a single photonic spiking neuron chip with double integrated electronic dendrites[J]. *Photonics Res*, 2023, **11**(12): 2033−2041.

[66] Gao S, Xiang S Y, Song Z W, et al. Hardware implementation of ultra-fast obstacle avoidance based on a single photonic spiking neuron[J]. *Laser Photonics Rev*, 2023, **17**(12): 2300424.

[67] Xiang S Y, Gao S, Shi Y C, et al. Experimental demonstration of a photonic spiking neuron based on a DFB laser subject to side-mode optical pulse injection[J]. *Sci China Inf Sci*, 2024, **67**(3): 132402.

[68] Gao S, Xiang S Y, Zheng D Z, et al. Cascadable excitability

and inhibition in DFB laser-based photonic spiking neurons[J]. *Opt Commun*, 2024, **554**: 130207.

[69] Zhang Y N, Xiang S Y, Song Z W, et al. Evolution of neuron-like spiking response and spike-based all-optical XOR operation in a DFB with saturable absorber[J]. *J Lightwave Technol*, 2024, **42**(6): 2026−2035.

[70] Yu C Y, Xiang S Y, Zhang Y N, et al. Neuromorphic convolution with a spiking DFB-SA laser neuron based on rate coding[J]. *Opt Express*, 2023, **31**(26): 43698−43711.

[71] Han Y N, Xiang S Y, Song Z W, et al. Pattern recognition in multi-synaptic photonic spiking neural networks based on a DFB-SA chip[J]. *Opto-Electron Sci*, 2023, **2**(9): 230021−230021.

[72] Xiang S Y, Shi Y C, Zhang Y H, et al. Photonic integrated neuro-synaptic core for convolutional spiking neural network[J]. *Opto-Electron Adv*, 2023, **6**(11): 230140.

[73] Hurtado A, Henning I D, Adams M J. Optical neuron using polarisation switching in a 1550nm-VCSEL[J]. *Opt Express*, 2010, **18**(24): 25170−25176.

[74] Hurtado A, Schires K, Henning I D, et al. Investigation of vertical cavity surface emitting laser dynamics for neuromorphic photonic systems[J]. *Appl Phys Lett*, 2012, **100**(10): 103703.

[75] Robertson J, Deng T, Javaloyes J, et al. Controlled inhibition of spiking dynamics in VCSELs for neuromorphic photonics: theory and experiments[J]. *Opt Lett*, 2017, **42**(8): 1560−1563.

[76] Hurtado A, Javaloyes J. Controllable spiking patterns in long-wavelength vertical cavity surface emitting lasers for neuromorphic photonics systems[J]. *Appl Phys Lett*, 2015, **107**(24): 241103.

[77] Deng T, Robertson J, Hurtado A. Controlled propagation of spiking dynamics in vertical-cavity surface-emitting lasers: towards neuromorphic photonic networks[J]. *IEEE J Sel Top Quantum Electron*, 2017, **23**(6): 1800408.

[78] Robertson J, Hejda M, Bueno J, et al. Ultrafast optical integration and pattern classification for neuromorphic photonics based on spiking VCSEL neurons[J]. *Sci Rep*, 2020, **10**(1): 6098.

[79] Robertson J, Wade E, Kopp Y, et al. Toward neuromorphic photonic networks of ultrafast spiking laser neurons[J]. *IEEE J Sel Top Quantum Electron*, 2020, **26**(1): 7700715.

[80] Robertson J, Kirkland P, Alanis J A, et al. Ultrafast neuromorphic photonic image processing with a VCSEL neuron[J]. *Sci Rep*, 2022, **12**(1): 4874.

[81] Robertson J, Kirkland P, Di Caterina G, et al. VCSEL-based photonic spiking neural networks for ultrafast detection and tracking[J]. *Neuromorph Comput Eng*, 2024, **4**(1): 014010.

[82] Chen Z J, Sludds A, Davis R, et al. Deep learning with coherent VCSEL neural networks[J]. *Nat Photonics*, 2023, **17**(8): 723−730.

[83] Wang J W, Sciarrino F, Laing A, et al. Integrated photonic quantum technologies[J]. *Nat Photonics*, 2020, **14**(5): 273−284.

[84] Tait A N, De Lima T F, Zhou E, et al. Neuromorphic photonic networks using silicon photonic weight banks[J]. *Sci Rep*, 2017, **7**(1): 7430.

[85] Mehrabian A, Al-Kabani Y, Sorger V J, et al. PCNNA: a photonic convolutional neural network accelerator[C]//*2018 31st IEEE International System-on-Chip Conference (SOCC)*, 2018: 169–173. https://doi.org/10.1109/SOCC.2018.8618542.

[86] Ma P Y, Tait A N, De Lima T F, et al. Photonic independent component analysis using an on-chip microring weight bank[J]. *Opt Express*, 2020, **28**(2): 1827−1844.

[87] Bangari V, Marquez B A, Miller H, et al. Digital electronics and analog photonics for convolutional neural networks (DEAP-CNNs)[J]. *IEEE J Sel Top Quantum Electron*, 2020, **26**(1): 7701213.

[88] Sunny F, Mirza A, Nikdast M, et al. CrossLight: a cross-layer optimized silicon photonic neural network accelerator[C]//*2021 58th ACM/IEEE Design Automation Conference (DAC)*, 2021: 1069–1074. https://doi.org/10.1109/DAC18074.2021.9586161.

[89] Ohno S, Tang R, Toprasertpong K, et al. Si microring resonator crossbar array for on-chip inference and training of the optical neural network[J]. *ACS Photonics*, 2022, **9**(8): 2614−2622.

[90] Xu S F, Wang J, Yi S C, et al. High-order tensor flow processing using integrated photonic circuits[J]. *Nat Commun*, 2022, **13**(1): 7970.

[91] Bai B W, Yang Q P, Shu H W, et al. Microcomb-based integrated photonic processing unit[J]. *Nat Commun*, 2023, **14**(1): 66.

[92] Reck M, Zeilinger A, Bernstein H J, et al. Experimental realization of any discrete unitary operator[J]. *Phys Rev Lett*, 1994, **73**(1): 58−61.

[93] Clements W R, Humphreys P C, Metcalf B J, et al. Optimal design for universal multiport interferometers[J]. *Optica*, 2016, **3**(12): 1460−1465.

[94] Shen Y C, Harris N C, Skirlo S, et al. Deep learning with coherent nanophotonic circuits[J]. *Nat Photonics*, 2017, **11**(7): 441−446.

[95] George J K, Nejadriahi H, Sorger V J. Towards on-chip optical FFTs for convolutional neural networks[C]//*2017 IEEE International Conference on Rebooting Computing (ICRC)*, 2017: 1–4. https://doi.org/10.1109/ICRC.2017.8123675.

[96] Fang M Y S, Manipatruni S, Wierzynski C, et al. Design of optical neural networks with component imprecisions[J]. *Opt Express*, 2019, **27**(10): 14009−14029.

[97] Zhang T, Wang J, Dan Y H, et al. Efficient training and design of photonic neural network through neuroevolution[J]. *Opt Express*, 2019, **27**(26): 37150−37163.

[98] Shokraneh F, Geoffroy-gagnon S, Liboiron-Ladouceur O. The diamond mesh, a phase-error- and loss-tolerant field-programmable MZI-based optical processor for optical neural networks[J]. *Opt Express*, 2020, **28**(16): 23495−23508.

[99] Shokraneh F, Geoffroy-Gagnon S, Liboiron-Ladouceur O. Towards phase-error- and loss-tolerant programmable MZI-based optical processors for optical neural networks[C]//*2020 IEEE Photonics Conference (IPC)*, 2020: 1–2. https://doi.org/10.1109/IPC47351.2020.9252466.

[100] Tian Y, Zhao Y, Liu S P, et al. Scalable and compact photonic neural chip with low learning-capability-loss[J]. *Nanophotonics*, 2022, **11**(2): 329−344.

[101] Zhu H H, Zou J, Zhang H, et al. Space-efficient optical computing with an integrated chip diffractive neural network[J]. *Nat Commun*, 2022, **13**(1): 1044.

[102] Shi Y, Ren J Y, Chen G Y, et al. Nonlinear germanium-silicon photodiode for activation and monitoring in photonic neuromorphic networks[J]. *Nat Commun*, 2022, **13**(1): 6048.

[103] Wu B, Liu S J, Cheng J W, et al. Real-valued optical matrix computing with simplified MZI mesh[J]. *Intell Comput*, 2023, **2**: 0047.

[104] Wright C D, Liu Y W, Kohary K I, et al. Arithmetic and biologically-inspired computing using phase-change materials[J]. *Adv Mater*, 2011, **23**(30): 3408−3413.

[105] Kuzum D, Jeyasingh R G D, Lee B, et al. Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing[J]. *Nano Lett*, 2012, **12**(5): 2179−2186.

[106] Cheng Z G, Ríos C, Pernice W H P, et al. On-chip photonic synapse[J]. *Sci Adv*, 2017, **3**(9): e1700160.

[107] Chakraborty I, Saha G, Roy K. Photonic in-memory computing primitive for spiking neural networks using phase-change materials[J]. *Phys Rev Appl*, 2019, **11**(1): 014063.

[108] Feldmann J, Youngblood N, Wright C D, et al. All-optical spiking neurosynaptic networks with self-learning capabilities[J]. *Nature*, 2019, **569**(7755): 208−214.

[109] Feldmann J, Youngblood N, Karpov M, et al. Parallel convolutional processing using an integrated photonic tensor core[J]. *Nature*, 2021, **589**(7840): 52−58.

[110] Zhou W, Dong B W, Farmakidis N, et al. In-memory photonic dot-product engine with electrically programmable weight banks[J]. *Nat Commun*, 2023, **14**(1): 2887.

[111] Vandoorne K, Mechet P, Van Vaerenbergh T, et al. Experimental demonstration of reservoir computing on a silicon photonics chip[J]. *Nat Commun*, 2014, **5**(1): 3541.

[112] Xu X Y, Tan M X, Corcoran B, et al. 11 TOPS photonic convolutional accelerator for optical neural networks[J]. *Nature*, 2021, **589**(7840): 44−51.

[113] Ashtiani F, Geers A J, Aflatouni F. An on-chip photonic deep neural network for image classification[J]. *Nature*, 2022, **606**(7914): 501−506.

[114] Fu T Z, Zang Y B, Huang Y Y, et al. Photonic machine learning with on-chip diffractive optics[J]. *Nat Commun*, 2023, **14**(1): 70.

[115] Meng X Y, Zhang G J, Shi N N, et al. Compact optical convolution processing unit based on multimode interference[J]. *Nat Commun*, 2023, **14**(1): 3000.

[116] Lin X, Rivenson Y, Yardimci N T, et al. All-optical machine learning using diffractive deep neural networks[J]. *Science*, 2018, **361**(6406): 1004−1008.

[117] Chang J L, Sitzmann V, Dun X, et al. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification[J]. *Sci Rep*, 2018, **8**(1): 12324.

[118] Bueno J, Maktoobi S, Froehly L, et al. Reinforcement learning in a large-scale photonic recurrent neural network[J]. *Optica*, 2018, **5**(6): 756−760.

[119] Lu L D, Zhu L Q, Zhang Q K, et al. Miniaturized diffraction grating design and processing for deep neural network[J]. *IEEE Photonics Technol Lett*, 2019, **31**(24): 1952−1955.

[120] Yan T, Wu J M, Zhou T K, et al. Fourier-space diffractive deep neural network[J]. *Phys Rev Lett*, 2019, **123**(2): 023901.

[121] Chen H, Feng J N, Jiang M W, et al. Diffractive deep neural networks at visible wavelengths[J]. *Engineering*, 2021, **7**(10): 1483−1491.

[122] Zhou T K, Lin X, Wu J M, et al. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit[J]. *Nat Photonics*, 2021, **15**(5): 367−373.

[123] Goi E, Chen X, Zhang Q M, et al. Nanoprinted high-neuron-density optical linear perceptrons performing near-infrared inference on a CMOS chip[J]. *Light Sci Appl*, 2021, **10**(1): 40.

[124] Fujita T, Sakaguchi H, Zhang J, et al. Magneto-optical diffractive deep neural network[J]. *Opt Express*, 2022, **30**(20): 36889−36899.

[125] Duan Z Y, Chen H, Lin X. Optical multi-task learning using multi-wavelength diffractive deep neural networks[J]. *Nanophotonics*, 2023, **12**(5): 893−903.

[126] Chen Y T, Nazhamaiti M, Xu H, et al. All-analog photoelectronic chip for high-speed vision tasks[J]. *Nature*, 2023, **623**(7985): 48−57.

[127] Zuo Y, Li B H, Zhao Y J, et al. All-optical neural network with nonlinear activation functions[J]. *Optica*, 2019, **6**(9): 1132−1137.

[128] Hamerly R, Bernstein L, Sludds A, et al. Large-scale optical neural networks based on photoelectric multiplication[J]. *Phys Rev X*, 2019, **9**(2): 021032.

[129] Sludds A, Bernstein L, Hamerly R, et al. A scalable optical neural network architecture using coherent detection[J]. *Proc SPIE*, 2020, **11299**: 112990H.

[130] Rafayelyan M, Dong J, Tan Y Q, et al. Large-scale optical reservoir computing for spatiotemporal chaotic systems prediction[J]. *Phys Rev X*, 2020, **10**(4): 041037.

[131] Xu Z H, Zhou T K, Ma M Z, et al. Large-scale photonic chiplet Taichi empowers 160-TOPS/W artificial general intelligence[J]. *Science*, 2024, **384**(6692): 202−209.

[132] Qian C, Lin X, Lin X B, et al. Performing optical logic operations by a diffractive neural network[J]. *Light Sci Appl*, 2020, **9**(1): 59.

[133] Wu C M, Yu H S, Lee S, et al. Programmable phase-change metasurfaces on waveguides for multimode photonic convolutional neural network[J]. *Nat Commun*, 2021, **12**(1): 96.

[134] Liu C, Ma Q, Luo Z J, et al. A programmable diffractive deep neural network based on a digital-coding metasurface array[J]. *Nat Electron*, 2022, **5**(2): 113−122.

[135] Gu J Q, Zhao Z, Feng C H, et al. ROQ: a noise-aware quantization scheme towards robust optical neural networks with low-bit controls[C]//*2020 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2020: 1586–1589. https://doi.org/10.23919/DATE48585.2020.9116521.

[136] Mourgias-Alexandris G, Moralis-Pegios M, Tsakyridis A, et al. Noise-resilient and high-speed deep learning with coherent silicon photonics[J]. *Nat Commun*, 2022, **13**(1): 5572.

[137] Kirtas M, Oikonomou A, Passalis N, et al. Quantization-aware training for low precision photonic neural networks[J]. *Neural*

Networks, 2022, **155**: 561−573.

[138] Feng C H, Gu J Q, Zhu H Q, et al. A compact butterfly-style silicon photonic–electronic neural chip for hardware-efficient deep learning[J]. *ACS Photonics*, 2022, **9**(12): 3906−3916.

[139] Zhan Y C, Zhang H, Lin H X, et al. Physics-aware analytic-gradient training of photonic neural networks[J]. *Laser Photonics Rev*, 2024, **18**(4): 2300445.

[140] Hughes T W, Minkov M, Shi Y, et al. Training of photonic neural networks through *in situ* backpropagation and gradient measurement[J]. *Optica*, 2018, **5**(7): 864−871.

[141] Zhou T K, Fang L, Yan T, et al. *In situ* optical backpropagation training of diffractive optical neural networks[J]. *Photonics Res*, 2020, **8**(6): 940−953.

[142] Zheng Z Y, Duan Z Y, Chen H, et al. Dual adaptive training of photonic neural networks[J]. *Nat Mach Intell*, 2023, **5**(10): 1119−1129.

[143] Wu T W, Menarini M, Gao Z H, et al. Lithography-free reconfigurable integrated photonic processor[J]. *Nat Photonics*, 2023, **17**(8): 710−716.

[144] Pai S, Sun Z H, Hughes T W, et al. Experimentally realized in situ backpropagation for deep learning in photonic neural networks[J]. *Science*, 2023, **380**(6643): 398−404.

作者简介

【通信作者】项水英 (1986-)，女，博士，西安电子科技大学通信工程学院教授，国家级青年人才，主要研究方向包括光神经形态计算、类脑信息处理、光脉冲神经网络、光电芯片与集成技术等，主持国家重点研发计划项目 1 项、国家自然科学基金项目 4 项，发表 SCI 检索论文 100 余篇。

E-mail: syxiang@xidian.edu.cn

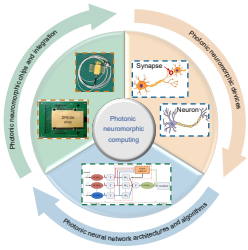
宋紫薇 (1996-)，女，博士，空军工程大学基础部讲师，主要研究方向包括光神经形态计算、光脉冲神经网络、光脉冲神经元和硅基光突触。

E-mail：1064971297@qq.com

扫描二维码，获取PDF全文

Progress in the research of optical neural networks

Xiang Shuiying[1*], Song Ziwei[2], Zhang Yahui[1], Guo Xingxing[1], Han Yanan[1], Hao Yue[3]



Research progress of photonic neuromorphic computing

Overview: The era of big data has placed greater demands on the computing power and speed of electronic computer processing systems. Issues such as the "memory wall" and "power wall" inherent in the traditional von Neumann architecture, coupled with the slowing down or even invalidation of Moore's Law, have posed significant challenges to electronic chips in terms of computing speed and power consumption. Utilizing optical computing as an alternative to traditional electronic computing represents one of the most promising avenues to address current challenges in computing power and power consumption.

This review systematically summarized the research progress of optical neural network (ONN) architectures and algorithms in both on-chip integration and in free space, and described typical research efforts in detail. In terms of on-chip integrated ONNs, the research progress of ONNs based on semiconductor lasers, silicon micro-ring resonators, Mach-Zehnder interferometers, and phase change materials was presented. Meanwhile, progress in research on free-space-based ONNs, including diffractive deep neural networks and metasurface-based ONNs, was summarized. Then, the advantages and disadvantages of these two types of ONNs were discussed and compared. The free-space-based ONNs have excellent parallel computing capabilities and are suitable for large-scale computing tasks. But they suffer from large volume and high complexity. In contrast, on-chip integrated ONNs have the advantages of scalability, high power efficiency, compact footprint, and high programmability. However, how to ensure accuracy and robustness in the process of large-scale integration to better cope with increasingly complex and large-scale computing tasks is still an urgent problem to be solved. In addition, training is an important step in the construction of neural networks and determines the performance of the entire system. Therefore, the research progress of the in-situ training method and the hardware-aware offline training method used in ONNs was introduced.

At last, the potential challenges that ONNs may encounter were discussed in depth, and a forward-looking perspective on their future development was offered. From the material and devices, to the system architecture, and ONNs are presenting a multi-level, cross-domain, and comprehensive development pattern for the algorithm implementation. By thoroughly exploring the potential of photon properties and deeply integrating them with artificial intelligence algorithms, the broad prospects and infinite possibilities of ONNs in building new intelligent computing systems can be demonstrated. Advances in ONNs can promote the development of the new computing paradigm of photonic brain-like computing, leading computing technology toward a more efficient and intelligent future.

Xiang S Y, Song Z W, Zhang Y H, et al. Progress in the research of optical neural networks[J]. *Opto-Electron Eng*, 2024, **51**(7): 240101; DOI: 10.12086/oee.2024.240101

Foundation item: Project supported by National Key Research and Development Program of China (2021YFB2801900, 2021YFB2801901, 2021YFB2801902, 2021YFB2801903, 2021YFB2801904, 2018YFE0201200), National Outstanding Youth Science Fund Project of National Natural Science Foundation of China (62022062), National Natural Science Foundation of China (61974177), and the Fundamental Research Funds for the Central Universities (QTZX23041)

[1]State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an, Shaanxi 710071, China; [2]Fundamentals Department, Air Force Engineering University, Xi'an, Shaanxi 710051, China; [3]State Key Discipline Laboratory of Wide Bandgap Semiconductor Technology, School of Microelectronics, Xidian University, Xi'an, Shaanxi 710071, China

* E-mail: syxiang@xidian.edu.cn