

光学学报

人工智能纳米光子学:光学神经网络与纳米光子学

栾海涛^{1,2}, 陈希^{1,2}, 张启明^{1,2}, 蔚浩义^{1,2}, 顾敏^{1,2*}

¹上海理工大学光子芯片研究院, 上海 200093;

²上海理工大学光电信息与计算机工程学院人工智能纳米光子学中心, 上海 200093

摘要 人工智能技术,特别是人工神经网络的创新引领了许多领域的应用革命,如网络搜索、计算机识别和语言、图像的识别技术。近年来纳米光子学的发展为传统的人工神经网络技术,特别是光学神经网络的发展带来了全新的物理视角以及截然不同的实现方法。一方面,纳米光子学是一门研究光与材料在纳米尺度相互作用的科学,可以带来全新的技术,如超分辨光学加工技术和超分辨光学成像技术,进而推动微纳尺度上多种功能的光学神经网络的实现。另一方面,纳米光子学中光子传播的多频段、高速度、低功耗的特点,促使了光学神经网络向着小体积、高密度、低功耗的方向发展。人工神经网络自身的发展也促使神经网络算法(如逆向设计、深度学习)在纳米光子学器件的设计中发挥前所未有的作用,以满足纳米光子学器件对自身功能、体积、集成度、计算功能的日益增长的要求。以神经网络的发展为起点,阐述人工神经网络特别是光学神经网络的发展趋势,以及人工神经网络与纳米光子学相互促进的发展历程。

关键词 光学器件; 人工智能; 人工神经网络; 光学神经网络; 纳米光子学; 光学人工智能

中图分类号 O439

文献标志码 A

doi: 10.3788/AOS202141.0823005

Artificial Intelligence Nanophotonics: Optical Neural Networks and Nanophotonics

Luan Haitao^{1,2}, Chen Xi^{1,2}, Zhang Qiming^{1,2}, Yu Haoyi^{1,2}, Gu Min^{1,2*}

¹*Institute of Photonic Chips, University of Shanghai for Science and Technology, Shanghai 200093, China;*

²*Centre for Artificial-Intelligence Nanophotonics, School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China*

Abstract Innovations in artificial intelligence, particularly artificial neural networks, have revolutionized applications in many areas, such as big-data search, computer recognition, and language and image recognition. The development of nanophotonics in the past decades has brought physical perspectives and different approaches to the implementation and the development of traditional artificial neural network technologies, especially optical neural networks. On the one hand, nanophotonics is a field studying the interaction of light and matter at the nanoscale, which can lead to new techniques, such as super-resolution optical lithography and super-resolution optical imaging technology, therefore in turn promoting the implementation of optical neural networks with multiple functions at the micro/nano scale. On the other hand, due to the characteristics of multi-bands, high speed, and low power consumption of light propagation, nanophotonics is accelerating the development of optical neural networks with compact size, high density, and low power consumption. Meanwhile, the development of artificial neural networks has also promoted neural network algorithms (such as reverse design and deep learning) as a new toolbox for the design of novel nanophotonics devices to meet the growing requirements of the function, volume, integration, and computing function of nano-photonics devices. In this paper, starting with the development of neural networks, we

收稿日期: 2021-03-18; 修回日期: 2021-04-01; 录用日期: 2021-04-06

基金项目: 国家自然科学基金(61975123, 11974247)、上海张江国家自主创新示范区专项发展基金重大项目(ZJ2019-ZD-005)

* E-mail: gumin@usst.edu.cn

review the development of artificial neural networks, especially the development of optical neural networks. The reciprocal development between artificial neural networks and nanophotonics is reviewed.

Key words optical devices; artificial intelligence; artificial neural networks; optical neural networks; nanophotonics; optical artificial intelligence

OCIS codes 230.3120; 230.4320

1 引言

人工智能技术的进步正以前所未有的广度和深度改变社会。从技术上来讲,这些进步极大地依赖于人工神经网络(ANNs)的发展。人工神经网络作为人工智能的核心技术,由 McCulloch 等^[1]在 1943 年提出,并首次基于硅基电子芯片的冯诺依曼计算机体系上通过数学和算法的形式实现,并模拟了生物神经网络的信息处理方式^[2-3],这标志着人工神经网络研究时代的序幕。经过 50 年的发展,人工神经网络已经有了实质的进步,特别是其自我学习能力、并行信息处理能力^[4],得到了众多应用领域的重视。然而,基于硅基电子芯片的冯诺依曼计算机越来越不能满足人工智能技术日益增长的大规模数据的处理需求,如多线程、多任务、快速图像处理的要求,这是由其芯片中的二进制逻辑的数据处理方法所限制的。据估计,如果用最先进的由冯诺依曼计算机构造的超级计算机来实时模拟 100% 的人脑功能,能耗将会高达 500 MW^[5-6],更不用说其巨大的物理体积。而人脑作为一个生物神经网络,其能耗只有数瓦,而且体积只有数十立方厘米。此外,硅基电子芯片还存在电子信号互相干扰的问题^[7],大大阻碍了从技术上实现具有高密度连接的人工神经网络。

硅基电子芯片的这些内在缺陷,极大地刺激着科学家们对于光计算和光学神经网络(ONNs)的研究^[8-11]。这是因为,光是一种优异的信息传播和处理媒介,具有高度并行的能力,而且光波可以在三维空间中进行无干扰的传播,特别是光传播的速度高达 10^8 m/s,在信息处理时几乎没有时间延迟^[12],这些优点使得光学神经网络极有可能以其大数据的传输与处理能力取代冯诺依曼电子器件^[13]。1978 年,斯坦福大学的 Goodman 等^[14]首次提出光的矢量矩阵耦合器的概念。这一概念的提出成为了光计算的起点,也极大地促使了光学神经网络的发展^[15]。随着集成光学、光电子学和纳米光子学研究的发展和深入,光学神经网络与电子器件的结合可实现光电混合神经网络。与此同时,低能耗、高并行度的集成

技术也可以用来构造光学处理器^[16]。随着高级人工神经网络特别是深度学习算法的发展,人工神经网络的发展也为纳米光子学器件的设计提供了一种高效的方法^[17],并进一步促进了光学神经网络以及纳米光子学的发展。

本文简要讨论了人工神经网络的基本理论及其分类,介绍了基于传统光学器件的光学神经网络及其发展所遇到的瓶颈,如全息光器件、空间光调制器等;介绍了纳米光子学在突破传统光学神经网络瓶颈中所做出的贡献,尤其是微纳光波导技术与三维激光加工技术在构造纳米光子学神经网络中的作用;阐述了人工神经网络与纳米光子学如何通过“深度学习与逆向设计”实现相互辅助发展;总结了光学神经网络发展中的一个重要的部件——光学突触的最新发展。最后对全文进行总结,展望人工智能与纳米光子学的未来发展趋势。

2 人工神经网络与纳米光子学

2.1 人工神经网络

人工神经网络可用于模拟生物神经网络^[18]处理信息的能力(图 1)。生物神经元包含基本的 4 个部分:轴突、树突、细胞体,以及突触^[19-20]。一个神经元细胞通常包含很多轴突的分支,用于接收神经网络中传来的信号。信号沿着轴突从一个神经元传达到另一个神经元。这样的信息传递过程通常是一对多或者多对多的传递。除此之外,生物神经元传递的信号是一种特别的阈值信号,这种信号与传统的冯诺依曼计算机中的二进制数据完全不同,图 1(b)左侧展示了一个生物神经细胞的 spiking 信号特征。因此,人工神经网络试图从这一角度模拟信号在不同神经元之间的传递方式,构造神经网络算法,实现类似生物神经式的信息处理方式。这种方式可以简单地用图 1(b)中的结构表示^[21]。同样地,人工神经网络中的神经元可以接收多个并行的输入信号,通过数学模型中的加权对这些信号进行处理,可以实时改变权重,通过神经元处理之后得到一个阈值输出信号,完成类似生物神经元计算的功能。

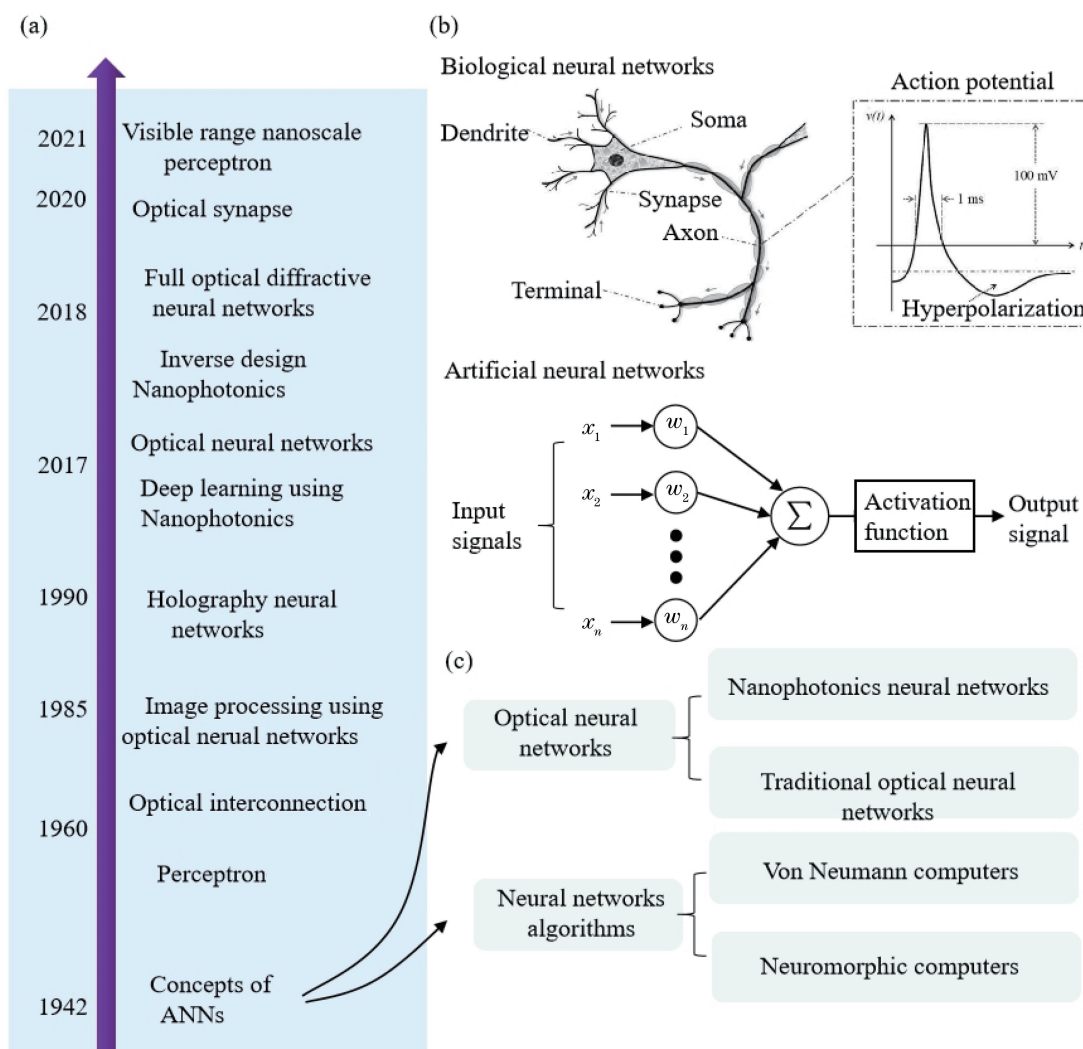


图 1 人工神经网络的基本概念与光学神经网络发展的时间线。(a)光学神经网络发展的时间轴;(b)生物神经网络的结构、信号特征^[18],以及人工神经网络的基本原理(其中 x_1, x_2, x_n 分别代表输入信号, w_1, w_2, w_n 代表计算权重);(c)人工神经网络的分类

Fig. 1 Concepts of artificial neural networks, and development timeline of optical neural networks. (a) Development timeline of optical neural networks; (b) features of structures and signals of biological neural networks^[18] and basic principle of artificial neural networks (x_1, x_2 , and x_n represent input signals, and w_1, w_2 , and w_n represent calculation weights); (c) general categories of artificial neural networks

人工神经网络大致可以分为两种。一种是基于硅基电子器件的神经网络,包括基于冯诺依曼计算机和神经形态电子计算机构造的功能性网络;另外一种是基于光学器件实现的光学神经网络,其中包含基于传统光学器件的光学神经网络,以及基于纳米光子学发展的纳米光子学神经网络。因为篇幅限制,本文将只讨论那些被应用于光学神经网络的神经网络算法,不对其他神经网络算法进行过于详细的讨论。

基于对生物神经网络信息处理特点的模拟,1957年,Rosenblatt^[22]提出了感知器的概念,它是

由两层神经元组成的神经网络(图1)。感知器是首个具有学习功能的人工神经网络,但这种神经网络只能完成简单的线性分类任务。Minsky和Papert^[23]合作撰写了《Perceptrons》一书,从数学上论证了感知器的缺点,尤其是感知器无法解决像“异或”(XOR)这样的简单分类任务,这对人类进行人工神经网络的研究又提出了更深层次的挑战。

神经网络形态相关领域的研究在20世纪80年代开始日益活跃,在诸多研究的基础上,1986年Rumelhart等^[24]在*Nature*上发表了题为《Learning Representations by Back-Propagating

Errors》的论文,讨论了基于反向传播(BP)算法的神经网络,这成为人工智能研究的全新方向。1987年,Lecun^[25]提出了神经网络的反向传播学习算法,利用变分法原理对反向传播算法进行了改进,缩短了机器学习的时间,加快了反向传播算法的计算速度。

此后,Lecun等^[26-27]提出了卷积神经网络的概念,并对这个卷积神经网络进行了训练。卷积神经网络通常由输入层、隐藏层和输出层组成,作为一种特殊的深度神经网络,其卷积神经元之间的连接是局部的,同一层中某些神经元之间可能共享连接权重。卷积神经网络是计算机视觉、语音识别、语音合成、图像合成和自然语言处理领域的标准算法之一^[28-29],这些研究工作奠定了实现光学神经网络并进行信息处理和推算的理论基础。

2.2 光学神经网络

2.2.1 传统光学神经网络

在神经网络算法发展的基础上,光学神经网络也逐渐得到特别的关注。尤其是以光学器件构造的光学神经网络在图像处理等应用中有着得天独厚的

优势。1985年,Psaltis等^[30-31]第一次演示了光学神经网络的实验并进行了光学神经网络对图像识别的探索,特别是1990年,Psaltis等^[32]又首次提出了将光学全息技术应用到神经网络中,在这个工作中,共轭镜面被用来提高全息信号中的阈值、反馈以及增益,最终实现神经网络的功能[图2(a)]。2004年Yeh等^[33]提出了利用矩阵光栅实现光学神经网络的方法。空间光调制器也被应用于构造光学神经网络[图2(b)]^[34]。这些光学神经网络虽然提高了工作效率以及系统的可靠性,但因为受硬件响应速度的限制,计算时的收敛速度非常慢。此外,Lu等^[35]也提出了一种利用高分辨率的显示器构造光学神经网络的方法,实现了一种动态的、可编程的信息存储。但是受器件的成像畸变以及光学探测能力的限制,这项工作实现的神经元数目非常少。Yang也提出了一种基于液晶显示器构造透镜阵列的方法,实现光学神经网络。然而受液晶显示的二维限制,这种方法不能稳定地控制神经元输入信号的动态阈值。

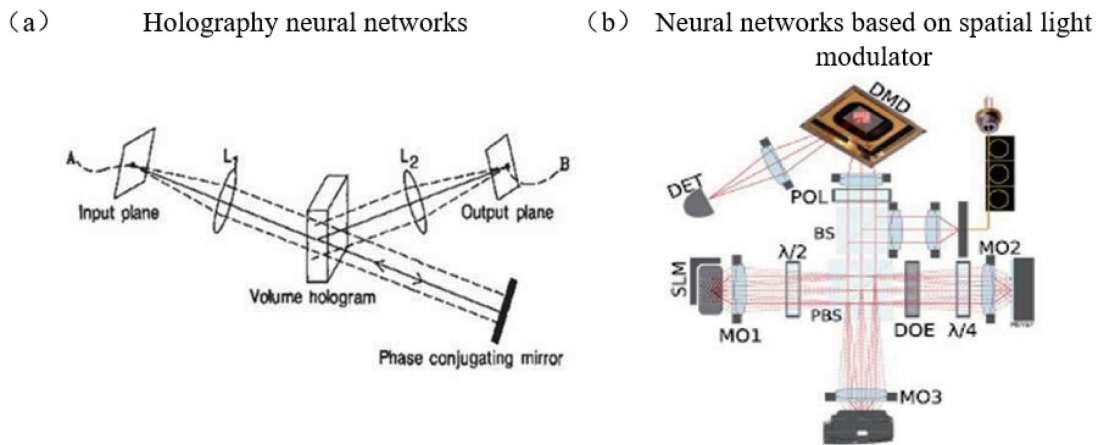


图 2 基于传统光学器件的光学神经网络。(a)全息神经网络^[32];(b)基于空间光调制器的光学神经网络^[34]

Fig. 2 Optical neural networks based on traditional optical devices. (a) Holography neural network^[32]; (b) optical neural network based on spatial light modulator^[34]

总体而言,传统光学器件体积过大,极大地限制了光学神经网络中神经元的数目和密度,也制约了其在应用领域的发展。受到20世纪八九十年代微纳光学器件制造水平、光电技术水平和大规模网络尚未普及应用的制约,光学神经网络研究在当时还未引起科研工作者足够的重视和热情。

2.2.2 纳米光子学神经网络

直到21世纪初,随着纳米光子学领域研究的进步,以及互联网技术(特别是光纤通信技术)在人类生活中的广泛应用,如何快速与节能地处理和传输

海量数据成为了科学研究的重点,此时人们重新将目光聚焦在处理速度超快、能耗极低的光学人工神经网络的构建及实现上。许多计算机领域的科学家Woods等^[9,36]在多篇科学评论中重新对光学特别是纳米光子学在构造人工神经网络以及光学计算方面的应用进行了大胆的猜测。与此同时,纳米光子学的发展也极大地促进了光学神经网络朝着高速度信号处理、微型集成化、低功耗、高神经元密度的方向快速发展。

第一类纳米光子学神经网络是基于光波导器件

的神经网络,如图 3 所示,图中 MRR 为微波共振器,BPD 为平衡光电二极管,LD 为激光二极管,MZM 为马赫-曾德尔调制器,WDM 为波长分离耦合器。2017 年,Tait 等^[37]研发了世界上首个光子神经网络,并展示了其超快的运算速度。物理上,这个光子神经网络使用微型圆形波导实现,神经元节点的相应特征通过纳米刻蚀的方法嵌入硅基底里。当输入入射光信号时,该系统会对在阈值处工作的

激光器的输出进行调制。在这个阈值附近,入射信号的微小参数变化都会对输出信号产生显著影响。波分复用技术通过在系统中的节点处使用一种特定波长的光来实现。系统的非线性特征通过一个反馈电路实现。同年,MIT 大学的 Shen 等^[38]开发了一种新型的利用光信号代替电信号的芯片架构,这种架构的特征是使用了多条光束并通过光波干涉产生干涉图样,这些干涉图样承载着预设结果。

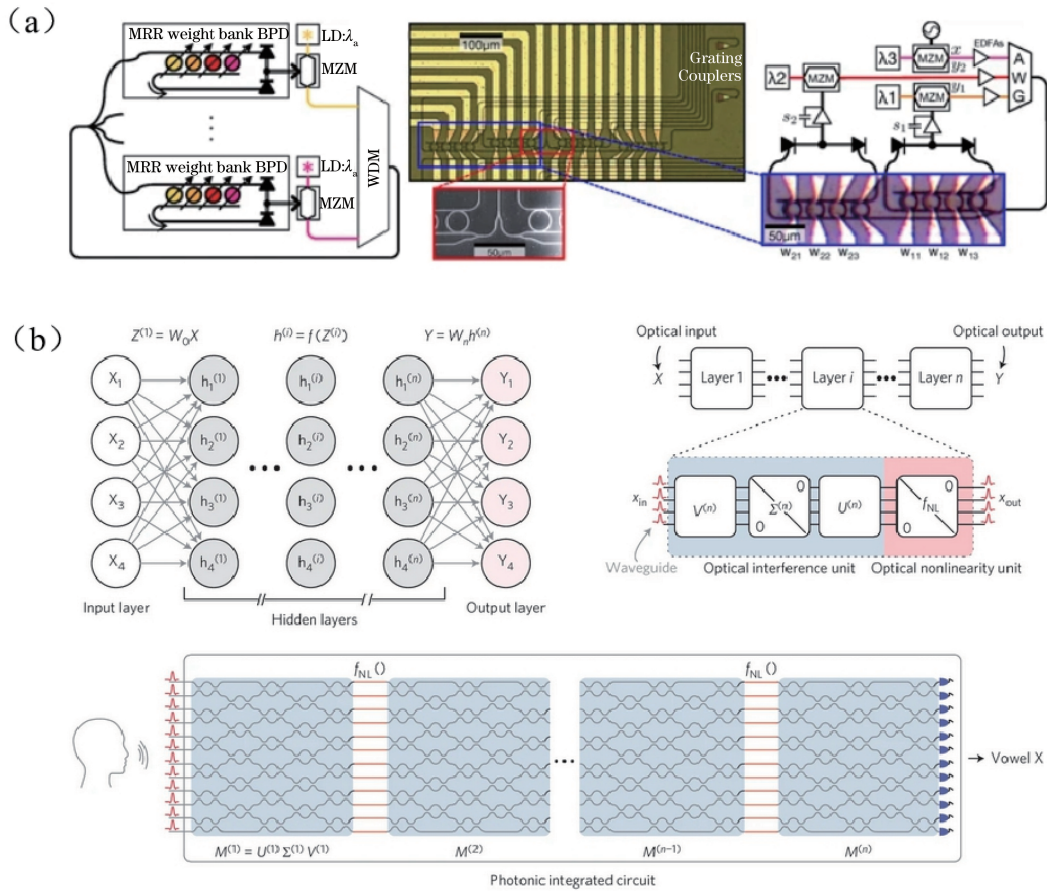


图 3 以微纳光波导为基础的纳米光子学神经网络实例。(a)光学神经网络原型机以及实验图^[37](左:原理图;中:实验图;右:信号传输图及其扫描电子显微镜下的实验装置);(b)以光信号为载体的新型芯片架构^[38](左上:人工神经网络的结构示意图,包含信号输入层、隐藏层和信号输出层;右上:分层显示的人工神经网络的结构;下:全光可集成的神经网络)

Fig. 3 Examples of nanophotonics neural networks based on micro/nano optical waveguides. (a) Prototype of optical neural network and experimental diagram^[37] (left: principle diagram; middle: experimental configuration; right: signal transmission image and scanning electron microscopic image of experimental setup); (b) structural design of a novel computing chip using optical signal as carrier^[38] (top-left: schematic diagram of artificial neural network, which includes input layers, hidden layers, and output layers; top-right: layered demonstration of artificial neural network; bottom: full optical integratable neural networks)

另外一类光学神经网络是基于衍射光学的构架,如图 3 所示。2018 年,Lin 课题组提出了一种全光学的深度学习结构,即衍射深度神经网络(D2NN)[图 4(a)],这是基于一种深度学习算法的无源衍射层,该结构利用误差反向传播法对衍射神

神经网络进行训练,理论上能以接近光速的速度处理输入信息,最终实现复杂的机器学习^[39]。物理上,此工作使用 3D 打印技术得到 D2NN 光学结构,从实验上实现了手写数字和时尚产品的图形分类功能,其工作波长位于太赫兹波段。2018 年,Chang

小组提出在 CNN 前端替换一个光学卷积层 (opt-convo) 的方案, 以提高神经网络计算性能的同时大尺度地降低能耗^[40]。光学卷积层是利用光学器件实现的卷积层, 其特点是大带宽、互联、并行处理、处理速度接近光速、功耗几乎接近零。

上海理工大学顾敏团队基于利用 3D 激光直写

技术进行超精细加工光敏材料的经验^[41-44], 成功地将全光学机器学习芯片的工作波段从太赫兹波段扩展至可见光波段^[45], 如图 4 (b) 所示。2020 年, Miscuglio 等^[46]提出了一种具有加速功能的光学卷积神经网络, 实现了每秒能够处理 PB(1 PB=2⁵⁰ B) 级数据量的计算能力。

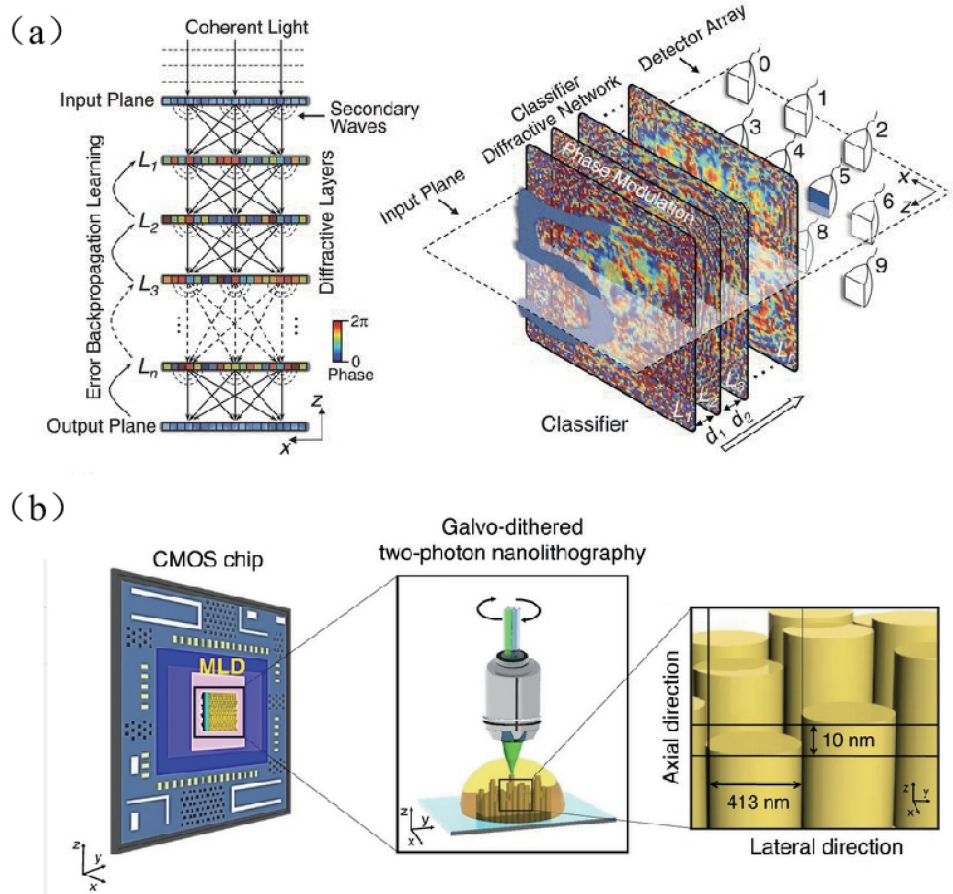


图 4 以衍射元件为基础的纳米光子学神经网络实例。(a) 衍射神经网络^[39] (左: 由多层透射/反射结构组成的衍射神经网络; 右: 实验中用于训练多层衍射神经网络的过程); (b) 通过双光子三维激光刻蚀技术加工的与 CMOS 集成的多层光子芯片示意图^[45] (中: 振镜扫描的双光子激光加工示意图; 右: 多层衍射神经网络芯片结构的内部示意图, 最小的间距为 10 nm)

Fig. 4 Examples of nanophotonics neural networks based on diffractive elements. (a) Diffractive neural networks^[39] (left: diffractive neural networks based on multi-layer transitive/reflective structures; right: experimental demonstration of training process using multi-layer diffractive neural networks); (b) schematic diagram of multi-layer photonic chip intergrated with CMOS fabricated by two-photon 3D laser etching technique^[45] (middle: diagram of two photon laser processing using galvanometer scanning; right: inner structure of multi-layer nanoscale diffractive neural networks. The smallest resolution is 10 nm)

2021 年, 哈尔滨工业大学 Chen 等^[47]将 D2NN 扩展到可见光波长, 提出了包括修正公式的一般理论, 以解决波长、神经元大小和制造限制之间的矛盾。这种新型可见光 D2NN 分类器用于识别可见光波长为 632.8 nm 的不变目标和已更改的目标 (即已被覆盖或更改的目标), 此项工作获得的实验

分类精度 (84%) 和数值分类精度 (91.57%) 量化了理论设计与制造系统性能之间的匹配。2021 年, 上海理工大学顾敏教授领导的研究团队提出的一种集成全息感知器, 将上述假设变成了现实, 他们研究出的一种基于机器学习的纳米级 AI 光学芯片可以实现以光速进行无源全光推理、神经密度比人脑高 10

倍、计算能力比电子高 10 多个数量级的处理能力。在这项工作中,科研人员利用一流的激光 3D 纳米打印技术加工光学传感器,其神经元密度可以达到每平方厘米超过 5 亿个神经元。这些智能纳米元件的纳米特征尺寸可将解密器的计算速度提升至上限 400 ExaFLOPS(ExaFLOPS 指每秒 10^{18} 次浮点运算),与集成光子硬件相比,其每秒操作数量增加 5 个数量级^[48]。

2.3 逆向设计与纳米光子学

纳米光子学的发展一方面给光学神经网络的实现提供了一个全新的物理基础,实现了光学神经网络的高速度、高密度的信息处理能力。另一方面,人工神经网络的发展,特别是近年来基于深度学习算法的逆向设计的发展给予了物理学家全新的手段和方法来设计种类繁多的纳米光子器件,如光学非线性、手性超材料、微型激光器、超透镜、拓扑光子晶体等。本节首先介绍了纳米光子学的应用领域,基于此展示了逆向设计方法如何帮助物理学家构建全新的纳米光子学器件。

纳米光子学是一门研究光与物质在纳米尺度相互作用的学科。纳米光子学包含光学成像^[49-50]、纳米光学加工^[42]、光子晶体^[51]、超材料^[52-53]、超表面^[54-55]、超结构^[56]和等离子体纳米结构^[57],近 20 年在精确控制光与物质相互作用和开发新型应用等方面取得了重大的进展。在这些应用中,结构与材料的设计都起着核心作用,目前主要有两种设计方法。第一种方法是基于物理理论的方法,该方法基于从先前或相关实践中获得的知识以及科学知识,例如通过反射透射理论可得到多层介质膜的透射谱精确解、用米散射理论可得到简单的球形纳米颗粒的精确解。但当结构的几何和空间排列变得复杂时,找到正确的结构来实现期望的光子特性并不容易。对于复杂的结构,需要利用第二种方法对电磁场进行数值模拟,如有限时域差分法、有限元法等。一般来说,这些电磁场数值模拟从一定的初始和边界条件出发,通过空间和时间上的离散麦克斯韦方程组来解决设计问题。通过建立足够的网格和迭代步骤,可以准确地计算出给定结构的光学性质。然而,人们经常需要对几何结构进行微调,并迭代地执行模拟,以逐步接近目标响应。这个过程很大程度上依赖于以往设计模板的经验,受数值模拟的效率和时间的限制,在寻找最优结构时只能调整有限的设计参数。

逆向设计与传统设计方法不同,由目标出发以

确定表征问题特征的参数。逆向设计通过将先进的人工智能算法与数值模拟相结合,实现复杂的参数优化,并可实现更大自由度的设计空间。在工程应用中,逆向设计已经广泛用于地球物理、生物、医疗及建筑等领域。最近,逆向设计广泛应用在纳米光子学的设计,通过结合智能优化算法、深度学习算法,实现了超越传统设计方法的纳米光子学器件。

2.3.1 基于优化算法的纳米光子学逆向设计

优化算法是指为了实现一定目标,通过控制参数变量寻找最优方案或函数值,以满足某种功能性指标。优化算法擅长于求解各种优化问题的应用技术,比如图像处理、生产调度、自动控制和机械设计等。为了有效地搜索庞大的设计空间,最早的光学逆向设计的过程通常采用人工智能的优化算法。优化算法是一种根据概率、按照固定步骤寻求问题的最优解的过程,算法不断地迭代、执行同一步骤,直到结束。常见的优化算法可以分为梯度下降算法与进化算法两大类。这种逆向设计算法使人们能够发现非直观的、不规则形状的光子结构,这些结构在许多应用中要比以经验设计的结构表现得更好,如硅光子元件、光子晶体和超材料。

梯度下降算法是逆向设计中使用最广泛的优化算法。当使用梯度下降算法时,目标函数的极小值成为计算的目标。通过沿着梯度的反方向(也就是下降最快的方向)将变量移动一定距离,目标函数便会逐渐减小,最终达到最小。拓扑优化算法与最速下降法是最常见的两种梯度下降算法。

拓扑优化算法在纳米光子学中得到了大量的应用,其基本思想是把器件的设计网格化成像素,将每一个像素的材料属性都当作一个设计变量。在材料设计中,设计变量的总数可以很大,因此器件的设计不局限于常见的几何结构。根据每次迭代过程的数值模拟结果计算梯度,更新器件设计中每个像素的材料属性,实现最终器件设计的优化。

如图 5(a)所示,拓扑优化算法用于设计 Z 形弯曲光子晶体波导^[58]。光波在通过较小的曲率半径的波导结构时,产生较大的能量损耗,基于经验的光子晶体波导的设计可以通过有限的几何结构优化波导的弯曲半径。通过拓扑优化算法调整 Z 形弯曲处 5 个小孔的形状,实现对光子晶体波导在小曲率半径处的带宽与透过率的优化,并通过实验在硅基二维光子晶体上验证了 200 nm 的传输带宽。

图 5(b)展示了利用逆向设计实现的波长分束器,其将入射的多波长光束分成短波长与长波长两

路^[59]。波长分束器的设计是通过多个步骤实现。第一阶段,利用“目标优先”的方法实现了分束器结构的初始分布,该阶段允许电磁场在分束器的模场分布符合目标,但可以允许电磁场分布不符合麦克斯韦方程组。第二阶段,在第一阶段的基础上通过优化算法减少不符合麦克斯韦方程组的地方,并设置限制条件让这一过程符合麦克斯韦方程组,通过最速下降法计算梯度、找到最优解,从而得到用来微

调分束器的连续渐变的折射率分布。最后,对连续渐变的折射率分布进行二值化处理,将其转变为空气与硅的折射率,并利用最速下降法对 1300 nm 与 1550 nm 附近的波段进行优化。整个设计过程需要在 GPU 加速下计算 36 h,才能完成时域有限差分法(FDTD)的模拟计算,最终的结果如图 5(b)所示,功能区域包含不规则形状的空穴,不同的输入波可被以低插入损耗方式引导到不同的输出口。

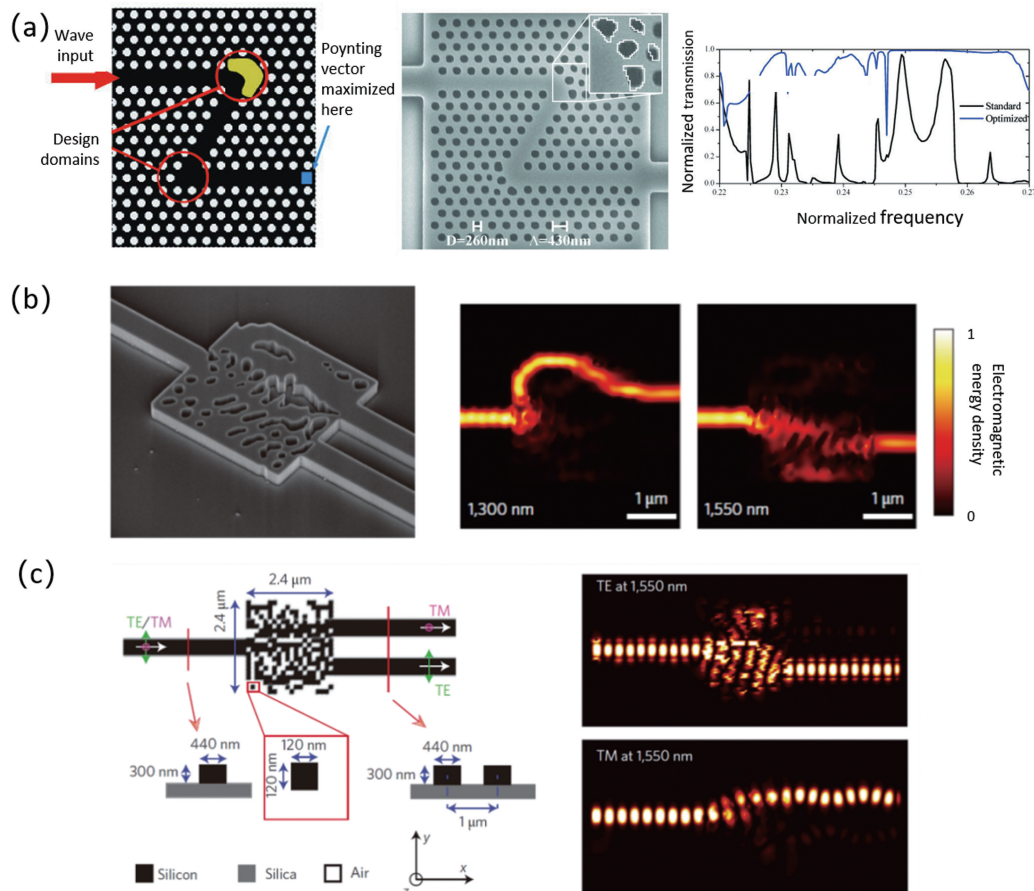


图 5 拓扑优化方法的逆向设计实例。(a)通过拓扑优化方法逆向设计 Z 形光子晶体波导结构^[58]; (b)通过梯度下降方法逆向设计波导波长分束器^[59]; (c)通过直接搜索方法逆向设计波导偏振分束器^[60]

Fig. 5 Examples of inverse designs of topology optimization methods. (a) Inverse design of Z-shape photonic crystal waveguide by topology optimization method^[58]; (b) inverse design of wavelength splitter by gradient descent method^[59]; (c) inverse design of polarization beam splitter by direct search method^[60]

图 5(c)为通过逆向设计实现的不同偏振光束[横电(TE)波和横磁(TM)波]的偏振分束器,其设计区域大小为 $2.4 \mu\text{m} \times 2.4 \mu\text{m}$ ^[60]。首先把设计区域的平面离散化成 20×20 的像素,每个像素的尺寸为 $120 \text{ nm} \times 120 \text{ nm}$,其折射率为硅或者空气的折射率,分别对应 1 或 0 的状态。将整个器件的高度方向也离散化成 10 nm 的台阶。优化过程中根据 TE 和 TM 波的传输效率定义了品质因数,然后随机改变每个像素的 0 和 1 的状态,如果像素在改变

状态后品质因数降低,则保持像素状态不变。这样的像素优化过程作为完整迭代过程的一步,通过重复多次的迭代过程,当品质因数低于设定的阈值则结束优化的流程,整个流程约需要 140 h,相比于基于梯度的算法需要更长的时间而且当设计的变量更多时,效率更低。

遗传算法(GA)是一种速度快、优化参数少的神经算法,在图像处理领域有重要的应用。这是因为其算法结构简单,可用于模拟自然中生物的遗传。

在遗传算法中,每一个待求解问题的候选解被抽象成为种群中一个个体的基因。种群中个体基因的好坏由待求解问题中表示个体基因的候选解的所得值来评判。种群中的两个个体通过交叉,以一定的概率交换一个或者多个基因并产生新的后代。每一个交叉都有一定的概率发生变异。发生变异的个体的某一位或某几位基因会变异成其他值,最终将以个体的适应度值为概率选取个体,并将其保留至下一代。遗传算法也被大量用于逆向设计光学器件,如超表面、光波导器件。通过遗传算法可以实现优化圆孔形状的超透镜焦点的特性^[61],实现超越传统设计的光学天线^[62],根据多种需求设计同时偏振、颜色敏感的纳米结构^[63],以及设计等离子体共振纳米结构色^[64]。

2.3.2 基于深度学习的纳米光子学逆向设计

从根本上说,基于优化算法的逆向设计是基于规则的方法,包含逐个案例的迭代搜索步骤,通常依赖于每一步的数值模拟来产生中间结果,以帮助修改搜索策略。这种随机算法受其随机搜索特性的限制,无法进行多约束问题的复杂设计。因为神经网络是一种模仿生物神经系统的新的信息处理模型,具有独特的结构。计算机领域中已经出现了多种类型的神经网络模型及相应的算法。其中,神经网络误差后传算法是一种最为典型的算法,在 1986 年被 Rumelhart 等^[24]提出。

最基础的神经网络的结构是多层感知器(MLP)。如图 6(a)所示,该结构尽管相对简单,在理论上已被证明是一种通用的逼近器,能够拟合任何可用有限数目神经元表达的连续函数。在具有复杂和特定任务架构的现代深度学习模型中,多层感知器经常作为瓶颈层来提取有意义的特征,例如作为高维数据(如图像)的紧凑表示。其中,一种基于双向多层感知器的深度学习模型可用于设计等离子体纳米结构^[65]。在早期研究的基础上,研究人员借助更深层次的网络结构,成功地模拟了光子结构与其光学特性之间复杂的物理关系。如图 6(b)所示,待设计的结构是一个由 8 个参数表示的 h 型金属结构:3 个连续参数(臂的长度和旋转角度)和 5 个二元参数(表示某些臂的存在性)。设计的目标是在水平方向和垂直方向的偏振光照射下的两个反射光谱。每个光谱离散为 43 个数据点,将材料性质(如金属结构与金属结构底层、覆盖层的折射率)表示为 25 个参数的向量。通过在 18000 个样本上训练了 8 层的几何预测网络和 6 层的光谱预测网络,使模型

能够同时作为快速模拟器和逆向设计工具。数据集的规模、输入复杂度和模型深度表明,拥有更大数据集的更深层模型可望解决逆向光子设计任务。

随着模型架构、培养策略和应用等方面的改进,基于多层感知器的模型的适应性也迅速提高。双向多层感知器神经网络通过集成学习策略实现手性超材料的逆向设计^[66],当入射光是左旋圆偏振光或右旋圆偏振光时,这些材料会表现出不同的反应。两个双向神经网络分别称为主网络和辅助网络,可用于对反射光谱、圆二色性光谱和设计参数三个物理量的互连进行建模。通过引入辅助网络,超材料的散射谱前向预测以及设计参数逆检索的准确性得到了大幅提高。该模型的功能也得到了扩展,能够根据期望的手性响应的几个参数检索可能的超材料设计,或根据设计参数直接预测圆二向色谱。

在大数据集上进行训练时,基于多层感知器的逆向设计模型往往不能收敛,这是因为存在多个满足相似要求但设计差异很大的候选模型。为了解决这个问题,一种用于逆向设计的串联训练方法被提出^[67]。这个方法首先训练一个正演建模网络,将设计映射到光学响应。将预训练好的前向网络连接到逆向设计网络的输出,前向预测误差作为监控信号。通过这种串联配置的间接训练,在正演模型的指导下,逆向检索输出被强制收敛到一种数据,这有效地解决了逆向设计问题中由于非唯一性这一基本属性而引起的数据不一致性问题。

除了设计问题之外,深度学习还可以用于解决光子学中的其他问题,比如模态分类和波导中有效的折射率检索^[68-69],通过降维来揭示等离子体纳米天线的基础物理^[70]和近场操作。作为深度学习模型的普通形式,多层感知器需要向量化的输入和输出。通过参数化模型的输入和输出,计算几个离散和无量纲元素组成的向量,多层感知器模型可以适应广泛的光子应用,而不考虑底层物理特性。

2.4 人工光学突触

在摩尔定律逐渐失效的当今时代,传统冯诺依曼架构的计算机的特点是:运算单元和存储单元相互独立,无法进一步降低能耗和提高集成度。因此,发展运算和存储合一的类脑结构的新型计算机系统成为了当务之急,其中对于负责信息传递的人工突触的研究更是重中之重。这是因为,生物的神经系统主要是由一种被称为神经元的细胞组成,神经元可以感知环境变化并存储相应的信息,然后向临近

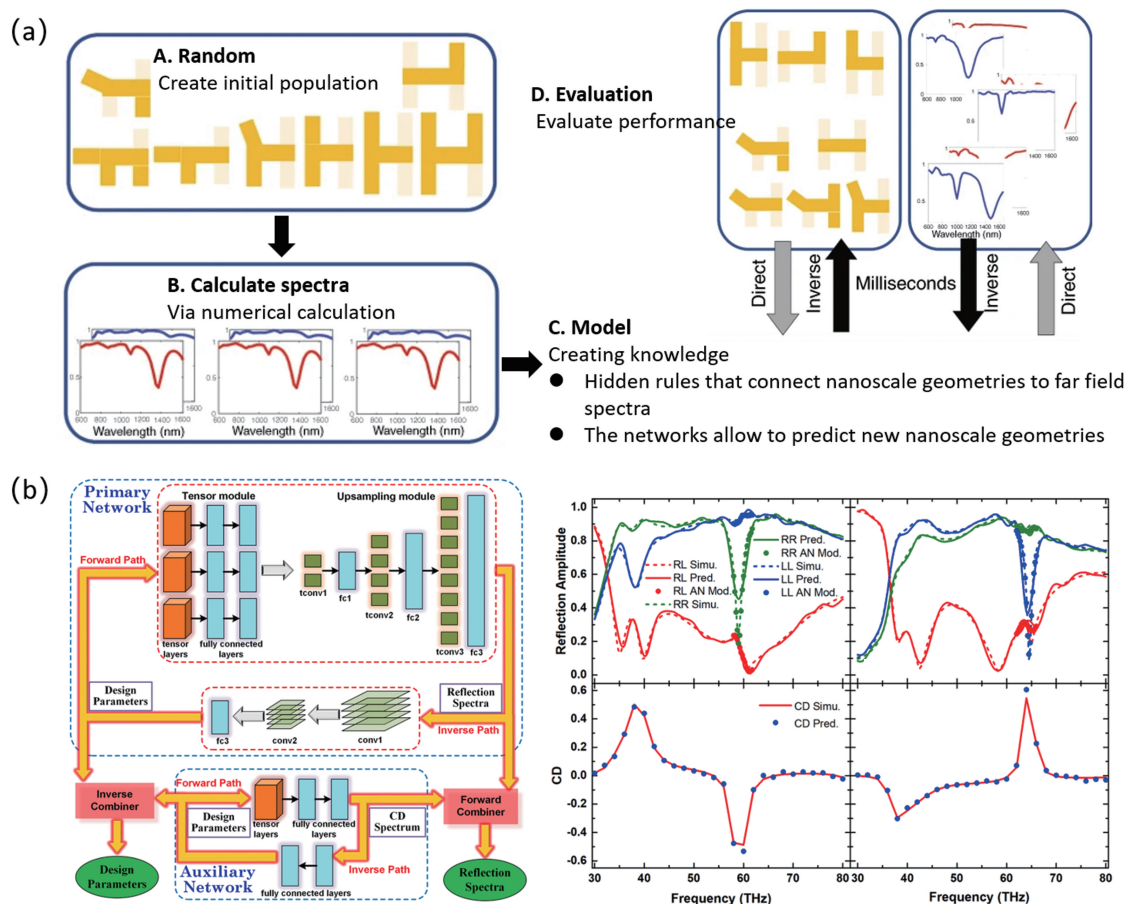


图 6 基于多层感知器的等离子体结构设计实例。(a)利用基于双向多层感知器的深度学习模型设计等离子体纳米结构^[65]；(b)利用双向多层感知器神经网络实现手性超材料的逆向设计^[66]

Fig. 6 Examples of design of plasmonic devices based on multi-layer perceptron. (a) Plasmonic nanostructures designed using deep learning model based on bidirectional multi-layer perceptron^[65]; (b) inverse design of chiral metamaterials designed using bidirectional multi-layer perceptron neural network^[66]

的神经元传递信息,这种信息传递的关键功能是由神经元结构中的突触来完成。前后神经元的质膜在突触的位置紧密贴合,突触前细胞发生动作电位时释放神经递质,神经递质传导到突触前膜。这时突触打开前膜上由电位控制的离子通道,神经递质通过突触间隙扩散到突触后膜上,实现神经元信息的传递。人类大脑之所以具有飞焦级的信息传递能耗、多任务并行处理的能力和卓越的感知功能,与突触的结构特点和工作机理密不可分。

注意到,目前的人工突触仍是主要基于电子器件实现,通过施加电压信号实现电流的类神经响应特性。相比之下,若电流的响应信号是由光脉冲激发,这种光学人工突触可以达到抛开运算存储间传输瓶颈的超低能耗、大幅突破的器件集成度、不受带宽限制的超高运算速度以及不被电磁干扰的超强稳定性。此外,光学人工突触的激发特性同时决定了其可以实现非接触式的信息处理,这对于无线通信

技术的发展更有帮助。

2.4.1 光学人工突触的基本特征

作为人工智能的基础器件,光学人工突触应具有宽带宽和激发光脉冲能耗小的特点。更重要的是,光学人工突触需要展示出人脑神经系统的一些基本特征,如短期/长期记忆模式切换、配对脉冲易化(PPF)、学习行为和光可塑性。下面针对每个特征进行介绍。

短期记忆(STM)和长期记忆(LTM)是人类记忆行为的两种类型,它们都被认为是由突触可塑性引起的^[71-72]。STM存储在海马体中表现为神经连接的暂时性弱增强,持续几分钟甚至几秒钟后几乎完全消失。与之相比,LTM则是神经连接的暂时性强化的表现,其作用时间明显长于STM,能够保持几天到几年。特别是在进行了具有巩固效应的重复排练之后,STM将可以转化成LTM的形式。如图7所示,使用不同数量或频率的脉冲光刺激可以

实现 STM、LTM, 以及 STM 到 LTM 的跃迁。响应光信号的衰减时间和衰减后的光信号强度随着数

量或频率的增加而增加, 可用于模拟人脑中 STM 到 LTM 的过渡。

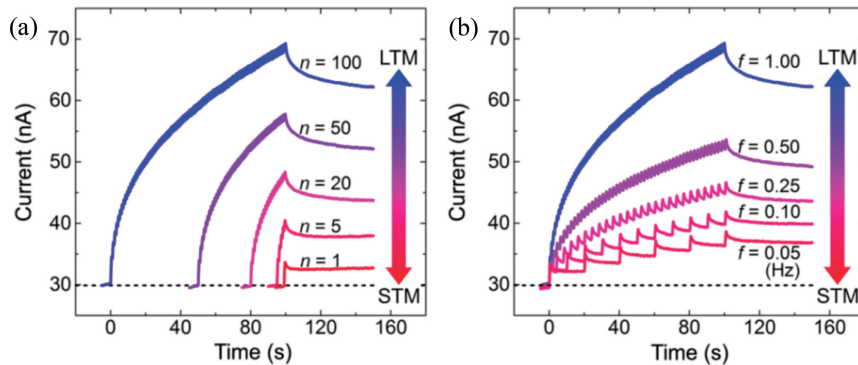


图 7 光学人工突触的长/短记忆模式及在光脉冲数量和频率变化下两种模式的切换^[73]。(a)光脉冲数量变化;(b)频率变化
Fig. 7 Switch between long-term and short-term memory modes of photonic synapses under adjustment of number of pulses and frequency^[73]. (a) Change of number of pulses; (b) change of frequency

配对脉冲易化效应是在一对具有可变时间间隔 (Δt) 的光脉冲作用下产生的特殊形态的电流响应信号, 其强度与这两个脉冲之间的时间间隔密切相关^[73]。当光脉冲刺激作用于人工突触时, 突触需要以典型的神经形态方式处理获得的光响应结果。电

流响应结果表明第二个光脉冲引起更大的光响应。使用 A_1 和 A_2 分别表示由第一、二个光脉冲引起的电流响应强度, A_2/A_1 随 Δt 的增加呈现双指数函数衰减 (图 8), 这是配对脉冲易化的特征, 它对识别和解码神经系统中的视觉和听觉信号等时间信息至关重要。

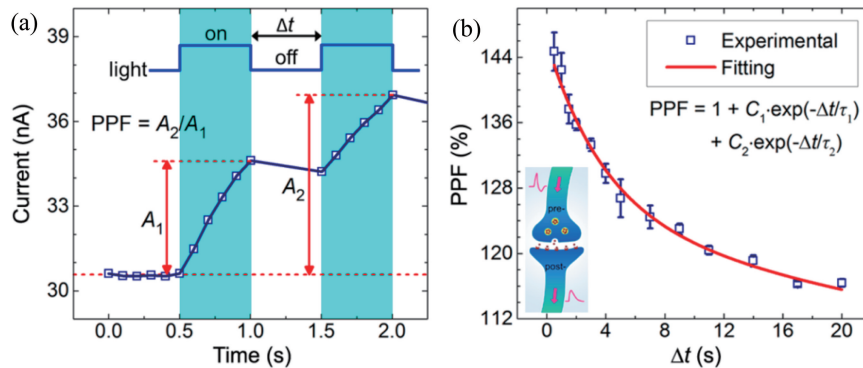


图 8 光学人工突触的配对脉冲易化^[73]。(a)电流变化曲线;(b)脉冲间隔对 PPF 效率(A_2/A_1)的影响
Fig. 8 Paired-pulse facilitation of photonic synapses^[73]. (a) Current change curves; (b) relationship between PPF efficiency (A_2/A_1) and pulse interval

光学人工突触的学习行为如图 9 所示。如果对突触相继施加两个时间间隔可控的光脉冲序列, 响应电流信号强度首先由脉冲作用增强, 然后在序列间隔内自发衰减至中间水平, 这类似于一个人在学习一段时间后会忘记部分所学信息的表现。为了弥补信号的衰减, 第二次刺激过程中需要的脉冲数量会显著少于第一次的数量。在相同时间内, 第二次刺激过程后信号强度的衰减要比第一次刺激过程后小。这种行为类似于以下事实: 一个人通常需要更少的时间来重新学习以前记忆的丢失信息, 并且重新学习的过程可以显著增强记忆的稳定性。需要强调的是, 光学人工突触最好可以具备令所有记忆效

应都能通过删除记忆来完全消除的途径, 令突触能够轻松恢复活性, 这可通过调节突触光可塑性效应中的抑制性后突触电流 (IPSC) 来实现。

如图 10 所示, 光学人工突触的光可塑性的实现基础是兴奋性后突触电流 (EPSC) 和 IPSC 的产生^[74-75]。在神经系统中, 神经递质会在当电脉冲应用于前突触膜的时候释放, 导致后突触电流的变化。EPSC 产生时兴奋神经递质被释放, 导致后突触电流的增大。与之相反, IPSC 发生在抑制神经递质释放时, 导致后突触电流的减小。对于相同的输入图像, EPSC 和 IPSC 自然会分别导致较大和较小的电流, 从而实现突触输出信号的调节。目前科学

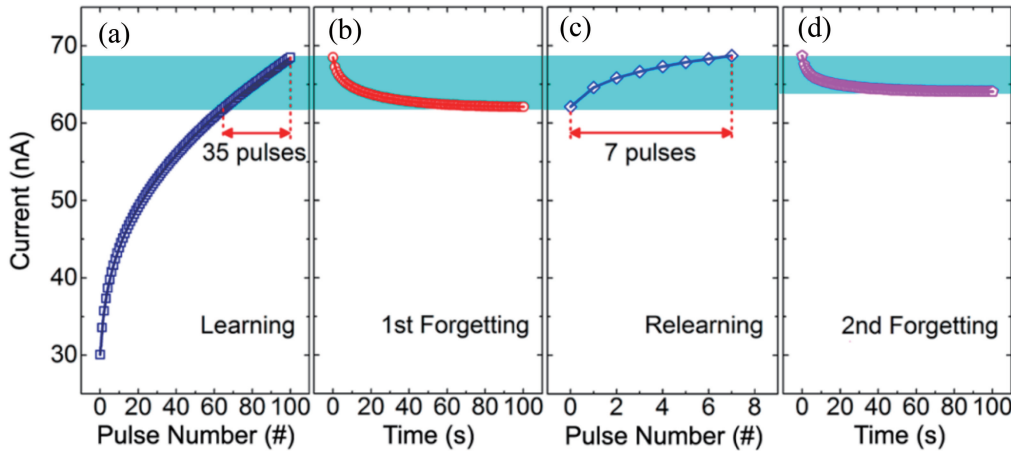


图 9 光学人工突触的学习功能^[73]。(a)第一次学习;(b)第一次遗忘;(c)第二次学习;(d)第二次遗忘

Fig. 9 Learning-experience behavior of photonic synapses^[73]. (a) 1st learning; (b) 1st forgetting; (c) 2nd learning; (d) 2nd forgetting

家已经发现了多种实现 EPSC 和 IPSC 的机理,如在不同光脉冲时间、不同脉冲频率、不同光波长和施加电压信号条件下实现 EPSC 和 IPSC 等。根据这些机理,还可以将不同曝光条件下电流响应结果调制为相同的值。此特性非常类似于人眼中的虹膜功能,它可以通过调节瞳孔大小来稳定各种光线环境下的视觉效果。

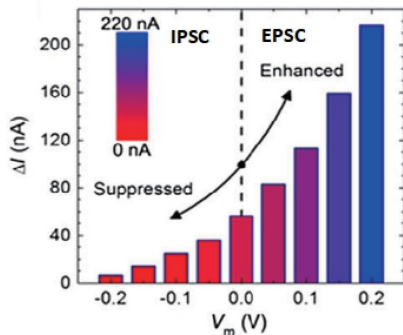


图 10 光学人工突触基于兴奋性后突触电流和抑制性后突触电流(调节外加电压)实现的光可塑性^[73]

Fig. 10 Plasticity of photonic synapses achieved based on excitation post synaptic currents and inhibition post synaptic currents (tuning applied electric voltage)^[73]

2.4.2 光学人工突触的工作原理以及实例

光学人工突触功能产生的必要条件是器件具有两个电阻态,即低电阻态和高电阻态。电阻态之间的转变可以在光脉冲作用下完成,脉冲消失后向最初的电阻态返回。在特定功率的光脉冲照射下,器件的电流响应信号的演化过程存在以下趋势:光脉冲会导致明显的电流信号变化,脉冲断开时电流逐渐衰减。如果器件的电学特性具有良好的非易失

性,电流信号衰减的拟合结果一般与 Kohlrausch 拉伸指数函数吻合,其表达式为

$$I_t = \Delta I \times \exp[-(t/\tau)^\beta] + I_c$$

式中: τ 和 β 分别表示弛豫时间常数和拉伸指数; I_c 为背景电流; ΔI 为电流变化。将器件暴露于具有不同周期和强度的光刺激下,较长的照明时间或较高的光强度可能导致较大的 ΔT 。

目前科学家已研发出数种途径来实现这样的功能,如氧化物半导体的持续性光电导效应^[76-77]、复合型半导体界面的肖特基接触^[78-79]、金属/半导体界面的光伏驱动的电导控制^[80-81]和低维材料的光诱导栅控效应^[82-83]。铟镓锌氧化物(IGZO)是一种典型的半导体材料,具有独特的持续性光电导效应^[84-85]。使用光强为 0.6 mW/cm^2 的紫外光照射时,电子由禁带被激发至导带,从而产生电流。撤掉光源后,因为持续性光电导效应的作用,光电流的衰减很慢,仍能保持一段时间的电流响应。基于氧化物半导体的光学人工突触结构如图 11 所示,紫外脉冲光作为外界刺激在 IGZO 中产生电流信号,脉冲断开时信号衰减,由此模拟突触的工作特征。

肖特基势垒是在导体和半导体材料相接触时产生的,同时半导体能带在界面处出现弯曲。选择能带结构匹配、具有出色的透明导电特性和丰富的用于电子捕获的表面缺陷材料,如 ITO/Nb: SrTiO₃ 异质结,通过界面形成的肖特基接触模拟光学人工突触。Nb: SrTiO₃ 的电子亲和能为 3.9 eV ,由于 Nb 的重掺杂,其费米能级位于导带底部,加上 ITO 的大逸出功(4.5 eV),ITO/Nb: SrTiO₃ 界面处会存在肖特基势垒^[86-87]。由于具有较高的势垒高度,这

个材料体系可以显著降低隧道漏电流,使得异质结保持在高阻态。光照下界面缺陷(主要是氧空位)中捕获的电子被释放,留下一些带正电荷的空陷阱提供额外的电势,因此势垒的高度将变低,电子隧穿的可能性增加,从而将异质结转换为低电阻状态。实验表明采用光强为 30 mW/cm^2 的蓝光照射 ITO/Nb:SrTiO₃ 异质结 100 s 会产生 55 nA 的电流变化,照射结束后电流逐渐衰减。

CH₃NH₃PbI₃ 是一种已被广泛研究的光伏材料,它具有的碘空位迁移的低势垒使其具有应用于光学突触的前景^[88-89]。将银薄膜沉积在 CH₃NH₃PbI₃ 的前表面和后表面,形成三明治结构,光照将抑制碘空位迁移的形成,加速碘空位的湮灭,从而导致 CH₃NH₃PbI₃ 电导的下降。通过改变光照参数可以实现对电导下降速度的控制,形成类似于后突触电流的电导衰减信号。

光诱导栅控效应是在低维结构中显示的光电导增益现象,这种情况下入射的光子不但可以增加体系中的光生载流子数目,而且使得结构表面俘获光生载流子,从而形成类浮栅的结构,以进一步调控结构的电导,因此通过较少的光子辐照可获得超高的电流增益。这种效应已被报道用于设计制作基于单壁碳纳米管的光学人工突触。如图 11 所示,这种光学人工突触使用 HfO₂ 作为栅介质, n 型掺杂硅薄膜作为栅电极。520 nm 的激光脉冲照射时,在硅薄膜栅控效应的作用下,光电流迅速衰减,其衰减程度和激光脉冲的频率呈正比^[89]。

2.4.3 光学人工突触的应用

光学人工突触可以模拟神经系统,特别是认知视觉信号的视觉系统的基本特征,它具有模仿人类感官记忆的前景。在人工智能技术飞速发展的当代,研发可以在各种环境条件下工作的先进机器人系统迫切需要智能的光电感知器件,以实现在科学、工业和军事环境中的人工智能视觉。这样的器件应是有源并且可控的,它需要具有一定的反馈和控制机制,以在变化的照明强度下实现自适应响应。

2018 年, Sun 等^[90] 使用 IGZO-烷基化石墨烯人工突触器件实现了对光脉冲信号的感知记忆功能。他们应用 200 个包括光刺激和电动刺激的混合脉冲,其中光脉冲的波长为 405 nm,功率为 $100 \mu\text{W}$,延迟时间为 50 ms;电脉冲的电压为 -1 V,延迟时间为 50 ms。研究发现混合脉冲具有相对于电脉冲的电流增强效应,可以产生强度大得多的响应信号。然后将由 785 个模拟神经元构成的美国国家标准与

技术研究院(MNIST)的手写数字图像作为光学信息输入人工突触,再使用 10 个输出神经元表示 0~9 的感知输出。在输入和输出神经元之间可以实现 7850 个人工突触的信息传递,再通过 EPSC 效应调节输出的电流权重,最终可在 20 个电脉冲态下实现准确的图像识别,而在只使用电脉冲的情况下,则需要 100 个电脉冲才能达到与之相当的识别准确度。

与 Sun 等使用黑白 MNIST 图像不同, Seo 等^[91] 报道了 h-BN/WSe₂ 异质结结构对彩色光脉冲信号的感知记忆功能,使用了两个异质结来分别实现人工突触和感光的功能。实验发现只存在突触时的图像识别率只有 40%,但在感光异质结的辅助下,识别率提高至 90%。相比于红光和绿光,器件在蓝光脉冲下会得到最明显的电信号。对于 MNIST 图像的学习功能, Lü 等^[92] 沿用了 7850 个人工突触的器件设计,使用碳量子点作为人工突触材料,发现经过 15000 次训练,该结构对于图形“A”和“I”的识别率分别为 73%和 65%。Ham 等^[93] 使用 CH₃NH₃PbI₃ 钙钛矿器件制作的人工突触,在 0.1 V 的外加电压下,通过施加功率为 4.82 nW 的光脉冲即可使得经过 2000 次训练后的识别率高达 82.7%,如果只用电压脉冲达到同样的识别率,能耗则要高 2600 倍。

2019 年, Gao 等^[73] 使用 ITO/Nb:SrTiO₃ 异质结制作人工光电突触,不仅证明了其可模仿人类的视觉记忆,而且展示了该器件的多维可控性。实验中从异质结突触阵列上随机选择 9 个阵列点,以识别和记忆 3×3 像素组成的视觉输入图像,每个阵列点受到 1000 s 固定时间、脉冲宽度为 0.5 s、频率为 1 Hz 的光刺激,然后进行突触后电流的衰减测量。实验表明通过突触后电流对比可以很好地识别输入图像,并展示出对输入图像的存储特征。同时,突触电流的强度可以通过改变光脉冲频率来调节。此外,通过向 ITO 电极施加外部电压 V_m 灵活地调节异质结的光响应效率。 V_m 对视觉记忆效果的调制类似于真实的视觉记忆中受到许多因素(如个人兴趣、情绪、注意力等)影响的现象。Gao 等^[73] 设定字母“L”、“I”和“H”分别表示一个人的低、中和高强度的视觉记忆。对阵列点施加产生 IPSC 和 EPSC 的电压(产生 IPSC 的外部电压是 -0.15 V,产生 EPSC 的外部电压是 0.15 V),在相同的观看时间后,该结构对于输入图像显示出明显的由弱到强的记忆。这些结果有望在未来的电子眼、多功能机器人、智能传感器等方面得到广泛的应用。

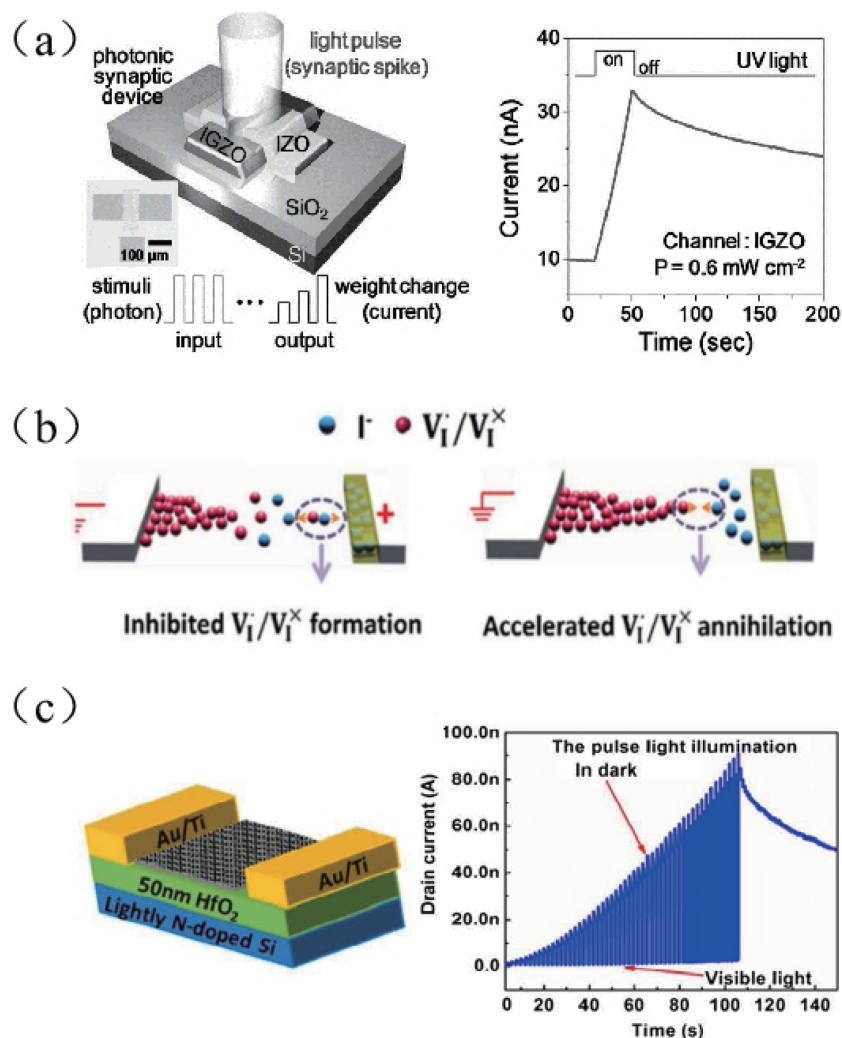


图 11 光学人工突触的结构和功能。(a) 铜镓锌氧化物光学人工突触的结构(左)和光脉冲激发下的电流变化曲线(右)^[83]；(b) $\text{CH}_3\text{NH}_3\text{PbI}_3$ 光学人工突触中光伏效应驱动的电导控制机理^[89]；(c) 利用单壁碳纳米管并通过光诱导栅控效应制作的光学人工突触的结构(左)和电流变化曲线(右)^[93]

Fig. 11 Structures and performances of photonic synapses. (a) Photonic synapses based on indium gallium zinc oxides (left), and curve of electric current change under optical pulse excitation (right)^[83] ; (b) principle of conductivity control of photonic synapses based on $\text{CH}_3\text{NH}_3\text{PbI}_3$ driven by photovoltaic effect^[89] ; (c) structure of photonic synapses based on light induced grating control effect using carbon nanotubes (left) and electric current change curve (right)^[93]

3 总结与展望

总结了传统光学、纳米光子学、神经网络算法在近半个世纪以来交互发展的时间线,其中包含了多个前沿课题,即人工神经网络、传统光学神经网络、纳米光子学神经网络、逆向设计与纳米光子学器件,以及最前沿的光学神经网络中的光学突触。一方面,光学神经网络,特别是纳米光子学神经网络为神经网络算法的实现提供了一个完全不同于传统电子芯片构架的物理平台,实现了多种多样的光学神经

网络的极其广泛的应用场景。纳米光子学神经网络的进步极有可能全面地突破当今人工神经网络计算的瓶颈,并带来全新的物理计算模型。另一方面,人工神经网络又为纳米光子学器件的设计与开发提供了一个强大的工具,特别是逆向设计技术的发展为科学家提供了一个探索高效的、目标驱动的设计新型纳米光子学器件的方法。

纳米光子学与人工神经网络是两个看似不同的学科,但事实上却可以紧密联系、协同发展。这种协同发展必须将传统光学、纳米光子学、计算科学、材

料学相结合,并且充分利用光学这一媒介在信息传导和处理中的优势,设计出一种完全不同于传统的计算机体系。这种发展最终也会促进高效光子神经网络的出现,提升现有人工神经网络计算的效率和尺度。光学神经网络的进一步发展也会使得其本身被应用于纳米光子学的逆向设计。在这样一条发展道路上,纳米光子学与神经网络协同发展的最终目标是构造新型纳米光子学结构和材料,实现小体积、高神经元密度、低功耗的光学神经网络。

正如当代电子计算机(冯诺依曼式计算机)的设计并没有遵循古典的计算器械(算盘)的设计思路一样,未来的人工神经网络甚至纳米光子学芯片的设计也应该突破当今冯诺依曼电子计算机的枷锁。只有不局限于现有的体系,系统性地整合光学、纳米光子学、材料学、计算科学的发展,才有可能促使全光的光学神经网络的诞生,实现以光的速度为基准的人工智能计算,并最终促使一个全新的人工智能时代早日到来。

参 考 文 献

- [1] McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity[J]. The bulletin of Mathematical Biophysics, 1943, 5(4): 115-133.
- [2] Krenker A, Bester J, Kos A. Introduction to the artificial neural networks [M] // Artificial Neural Networks - Methodological Advances and Biomedical Applications. London: InTech, 2011: 25693.
- [3] Thompson R. The neurobiology of learning and memory[J]. Science, 1986, 233(4767): 941-947.
- [4] Dayhoff J E, DeLeo J M. Artificial neural networks [J]. Cancer, 2001, 91(S8): 1615-1635.
- [5] Dzierma Y, Schuermann M, Melchior P, et al. Optimizing adjuvant stereotactic radiotherapy of motor-eloquent brain metastases: sparing the nTMS-defined motor cortex and the hippocampus [J]. Frontiers in Oncology, 2021, 11: 628007.
- [6] Sengupta B, Stemmler M B. Power consumption during neuronal computation[J]. Proceedings of the IEEE, 2014, 102(5): 738-750.
- [7] Schwabe R J, Zelinger S, Key T S, et al. Electronic lighting interference[J]. IEEE Industry Applications Magazine, 1998, 4(4): 43-48.
- [8] Chaisakul P, Marris-Morini D, Frigerio J, et al. Integrated germanium optical interconnects on silicon substrates[J]. Nature Photonics, 2014, 8(6): 482-488.
- [9] Woods D, Naughton T J. Photonic neural networks [J]. Nature Physics, 2012, 8(4): 257-259.
- [10] Feitelson D G. Optical computing: a survey for computer scientists [J]. Applied Optics, 1989, 28: 2182-2183.
- [11] Zhang Q M, Yu H Y, Barbiero M, et al. Artificial neural networks enabled by nanophotonics[J]. Light: Science & Applications, 2019, 8: 42.
- [12] Athale R A, Psaltis D, Wagner K. Optical computing: introduction by the guest editors to the feature in the 1 May 1988 issue[J]. Applied Optics, 1988, 27(9): 1641-1642.
- [13] Engheta N. Circuits with light at nanoscales: optical nanocircuits inspired by metamaterials[J]. Science, 2007, 317(5845): 1698-1702.
- [14] Goodman J W, Dias A R, Woody L M. Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms [J]. Optics Letters, 1978, 2(1): 1-3.
- [15] Caulfield H J, Dolev S. Why future supercomputing requires optics [J]. Nature Photonics, 2010, 4(5): 261-263.
- [16] Xiang L, Wang Q, Nie D, et al. Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image [J]. Medical Image Analysis, 2018, 47: 31-44.
- [17] Molesky S, Lin Z, Piggott A Y, et al. Inverse design in nanophotonics [J]. Nature Photonics, 2018, 12(11): 659-670.
- [18] Haddad W, Hui Q, Bailey J. Human brain networks: spiking neuron models, multistability, synchronization, thermodynamics, maximum entropy production, and anesthetic cascade mechanisms [J]. Entropy, 2014, 16(7): 3939-4003.
- [19] Harris K M, Jensen F E, Tsao B. Three-dimensional structure of dendritic spines and synapses in rat hippocampus (CA1) at postnatal day 15 and adult ages: implications for the maturation of synaptic physiology and long-term potentiation [J]. The Journal of Neuroscience, 1992, 12(7): 2685-2705.
- [20] Koch C, Segev I. The role of single neurons in information processing [J]. Nature Neuroscience, 2000, 3(Suppl): 1171-1177.
- [21] Nahmias M A, Shastri B J, Tait A N, et al. Neuromorphic photonics [J]. Optics and Photonics News, 2018, 29(1): 34-41.
- [22] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain [J]. Psychological Review, 1958, 65(6): 386.
- [23] Minsky M, Papert S A. Perceptrons [M]. Cambridge: The MIT Press, 2017.
- [24] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors [J].

- Nature, 1986, 323(6088): 533-536.
- [25] Lecun Y. PhD thesis: modeles connexionnistes de l'apprentissage (connectionist learning models) [D]. Paris: Université de Paris, 1987.
- [26] Lecun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4): 541-551.
- [27] Cun Y L, Boser B, Denker J S, et al. Handwritten digit recognition with a back-propagation network[C]// Advances in Neural Information Processing Systems, June 1, 1990, San Francisco, CA, United States. San Francisco: Morgan Kaufmann, 1990: 396-404.
- [28] Yamashita R, Nishio M, Do R K G, et al. Convolutional neural networks: an overview and application in radiology [J]. Insights into Imaging, 2018, 9(4): 611-629.
- [29] Ning Y S, He S, Wu Z Y, et al. A review of deep learning based speech synthesis [J]. Applied Sciences, 2019, 9(19): 4050.
- [30] Psaltis D, Farhat N. Optical information processing based on an associative-memory model of neural nets with thresholding and feedback [J]. Optics Letters, 1985, 10(2): 98-100.
- [31] Farhat N H, Psaltis D, Prata A, et al. Optical implementation of the Hopfield model [J]. Applied Optics, 1985, 24(10): 1469-1475.
- [32] Psaltis D, Brady D, Gu X G, et al. Holography in artificial neural networks [J]. Nature, 1990, 343(6256): 325-330.
- [33] Yeh S L, Lo R C, Shi C Y. Optical implementation of the Hopfield neural network with matrix gratings [J]. Applied Optics, 2004, 43(4): 858-865.
- [34] 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.
- [35] Lu T, Wu S, Xu X, et al. Two-dimensional programmable optical neural network [J]. Applied Optics, 1989, 28(22): 4908-4913.
- [36] Kirchain R, Kimerling L. A roadmap for nanophotonics [J]. Nature Photonics, 2007, 1(6): 303-305.
- [37] Tait A N, de Lima T F, Zhou E, et al. Neuromorphic photonic networks using silicon photonic weight banks [J]. Scientific Reports, 2017, 7(1): 7430.
- [38] Shen Y C, Harris N C, Skirlo S, et al. Deep learning with coherent nanophotonic circuits [C]//2017 IEEE Photonics Society Summer Topical Meeting Series (SUM), July 10-12, 2017, San Juan, PR, USA. New York: IEEE Press, 2017: 189-190.
- [39] 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.
- [40] Chang J, Sitzmann V, Dun X, et al. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification [J]. Scientific Reports, 2018, 8(1): 12324.
- [41] Li X, Lan T H, Tien C H, et al. Three-dimensional orientation-unlimited polarization encryption by a single optically configured vectorial beam [J]. Nature Communications, 2012, 3: 998.
- [42] Gan Z, Cao Y, Evans R A, et al. Three-dimensional deep sub-diffraction optical beam lithography with 9 nm feature size [J]. Nature Communications, 2013, 4: 2061.
- [43] Yue Z, Xue G, Liu J, et al. Nanometric holograms based on a topological insulator material [J]. Nature Communications, 2017, 8: 15354.
- [44] Yue Z J, Ren H R, Wei S B, et al. Angular-momentum nanometrology in an ultrathin plasmonic topological insulator film [J]. Nature Communications, 2018, 9: 4413.
- [45] Gu M, Fang X Y, Ren H R, et al. Optically digitalized holography: a perspective for all-optical machine learning [J]. Engineering, 2019, 5(3): 363-365.
- [46] Miscuglio M, Hu Z B, Li S R, et al. Massively parallel amplitude-only Fourier neural network [J]. Optica, 2020, 7(12): 1812-1819.
- [47] Chen H, Feng J N, Jiang M W, et al. Diffractive deep neural networks at visible wavelengths [EB/OL]. [2021-03-15]. [https://www.sciencedirect.com/science/article/pii/S2095809921000448? via%3Dihub](https://www.sciencedirect.com/science/article/pii/S2095809921000448?via%3Dihub).
- [48] 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: Science & Applications, 2021, 10: 40.
- [49] Hell S W, Wichmann J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy [J]. Optics Letters, 1994, 19(11): 780-782.
- [50] Betzig E, Patterson G H, Sougrat R, et al. Imaging intracellular fluorescent proteins at nanometer resolution [J]. Science, 2006, 313(5793): 1642-1645.
- [51] Johnson S G, Mekis A, Fan S, et al. Molding the flow of light [J]. Computing in Science & Engineering, 2001, 3(6): 38-47.

- [52] Jahani S, Jacob Z. All-dielectric metamaterials [J]. *Nature Nanotechnology*, 2016, 11(1): 23-36.
- [53] Mokkaapati S, Beck F J, de Waele R, et al. Resonant nano-antennas for light trapping in plasmonic solar cells [J]. *Journal of Physics D: Applied Physics*, 2011, 44(18): 185101.
- [54] Hsiao H H, Chu C H, Tsai D P. Metasurfaces: fundamentals and applications of metasurfaces [J]. *Small Methods*, 2017, 1(4): 1770041.
- [55] Minovich A E, Miroschnichenko A E, Bykov A Y, et al. Functional and nonlinear optical metasurfaces [J]. *Laser & Photonics Reviews*, 2015, 9(2): 195-213.
- [56] Baranov D G, Wersäll M, Cuadra J, et al. Novel nanostructures and materials for strong light-matter interactions [J]. *ACS Photonics*, 2018, 5(1): 24-42.
- [57] Pelton M, Aizpurua J, Bryant G. Metal-nanoparticle plasmonics [J]. *Laser & Photonics Review*, 2008, 2(3): 136-159.
- [58] Borel P I, Harpøth A, Frandsen L H, et al. Topology optimization and fabrication of photonic crystal structures [J]. *Optics Express*, 2004, 12(9): 1996-2001.
- [59] Piggott A Y, Lu J, Lagoudakis K G, et al. Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer [J]. *Nature Photonics*, 2015, 9(6): 374-377.
- [60] Shen B, Wang P, Polson R, et al. An integrated-nanophotonics polarization beamsplitter with $2.4 \times 2.4 \mu\text{m}^2$ footprint [J]. *Nature Photonics*, 2015, 9(6): 378-382.
- [61] Huntington M D, Lauhon L J, Odom T W. Subwavelength lattice optics by evolutionary design [J]. *Nano Letters*, 2014, 14(12): 7195-7200.
- [62] Feichtner T, Selig O, Kiunke M, et al. Evolutionary optimization of optical antennas [J]. *Physical Review Letters*, 2012, 109(12): 127701.
- [63] Wiecha P R, Arbouet A, Girard C, et al. Evolutionary multi-objective optimization of colour pixels based on dielectric nanoantennas [J]. *Nature Nanotechnology*, 2017, 12(2): 163-169.
- [64] Kristensen A, Yang J K W, Bozhevolnyi S I, et al. Plasmonic colour generation [J]. *Nature Reviews Materials*, 2017, 2: 16088.
- [65] Malkiel I, Mrejen M, Nagler A, et al. Plasmonic nanostructure design and characterization via deep learning [J]. *Light, Science & Applications*, 2018, 7: 60.
- [66] Ma W, Cheng F, Liu Y M. Deep-learning-enabled on-demand design of chiral metamaterials [J]. *ACS Nano*, 2018, 12(6): 6326-6334.
- [67] Liu D J, Tan Y X, Khoram E, et al. Training deep neural networks for the inverse design of nanophotonic structures [C] // 2019 Conference on Lasers and Electro-Optics (CLEO), May 5-10, 2019, San Jose, CA, USA. New York: IEEE Press, 2019: 1-2.
- [68] Alagappan G, Png C E. Modal classification in optical waveguides using deep learning [J]. *Journal of Modern Optics*, 2019, 66(5): 557-561.
- [69] Alagappan G, Png C E. Deep learning models for effective refractive indices in silicon nitride waveguides [J]. *Journal of Optics*, 2019, 21(3): 035801.
- [70] Kiarashinejad Y, Abdollahramezani S, Zandehshahvar M, et al. Deep learning reveals underlying physics of light-matter interactions in nanophotonic devices [J]. *Advanced Theory and Simulations*, 2019, 2(9): 1900088.
- [71] Bessonov A A, Kirikova M N, Petukhov D I, et al. Layered memristive and memcapacitive switches for printable electronics [J]. *Nature Materials*, 2015, 14(2): 199-204.
- [72] Ohno T, Hasegawa T, Tsuruoka T, et al. Short-term plasticity and long-term potentiation mimicked in single inorganic synapses [J]. *Nature Materials*, 2011, 10(8): 591-595.
- [73] Gao S, Liu G, Yang H L, et al. An oxide Schottky junction artificial optoelectronic synapse [J]. *ACS Nano*, 2019, 13(2): 2634-2642.
- [74] Li Y, Zhong Y, Zhang J, et al. Activity-dependent synaptic plasticity of a chalcogenide electronic synapse for neuromorphic systems [J]. *Scientific Reports*, 2014, 4: 4906.
- [75] Zhang J Y, Dai S L, Zhao Y W, et al. Recent progress in photonic synapses for neuromorphic systems [J]. *Advanced Intelligent Systems*, 2020, 2(3): 1900136.
- [76] Lee M, Lee W, Choi S, et al. Brain-inspired photonic neuromorphic devices using photodynamic amorphous oxide semiconductors and their persistent photoconductivity [J]. *Advanced Materials*, 2017, 29(28): 1700951.
- [77] Iwasaki T, Itagaki N, Den T, et al. Combinatorial approach to thin-film transistors using multicomponent semiconductor channels: an application to amorphous oxide semiconductors in In-Ga-Zn-O system [J]. *Applied Physics Letters*, 2007, 90(24): 242114.
- [78] Wang Z, Yin M, Zhang T, et al. Engineering incremental resistive switching in TaO_x based memristors for brain-inspired computing [J].

- Nanoscale, 2016, 8(29): 14015-14022.
- [79] Hu D C, Yang R, Jiang L, et al. Memristive synapses with photoelectric plasticity realized in $\text{ZnO}_{1-x}/\text{AlO}_y$ heterojunction [J]. ACS Applied Materials & Interfaces, 2018, 10(7): 6463-6470.
- [80] Zhu X J, Lu W D. Optogenetics-inspired tunable synaptic functions in memristors [J]. ACS Nano, 2018, 12(2): 1242-1249.
- [81] Wang Y, Lü Z, Chen J R, et al. Photonic flash memory: photonic synapses based on inorganic perovskite quantum dots for neuromorphic computing [J]. Advanced Materials, 2018, 30(38): 1870287.
- [82] Shao L, Wang H L, Yang Y, et al. Optoelectronic properties of printed photogating carbon nanotube thin film transistors and their application for light-stimulated neuromorphic devices [J]. ACS Applied Materials & Interfaces, 2019, 11(12): 12161-12169.
- [83] Wang H, Zhao Q, Ni Z, et al. A ferroelectric/electrochemical modulated organic synapse for ultraflexible, artificial visual-perception system [J]. Advanced Materials, 2018, 30(46): e1803961.
- [84] Jeon S, Ahn S E, Song I, et al. Gated three-terminal device architecture to eliminate persistent photoconductivity in oxide semiconductor photosensor arrays [J]. Nature Materials, 2012, 11(4): 301-305.
- [85] Bera A, Peng H Y, Lourembam J, et al. A versatile light-switchable nanorod memory: wurtzite ZnO on perovskite SrTiO_3 [J]. Advanced Functional Materials, 2013, 23(39): 4977-4984.
- [86] Robertson J. Band offsets of wide-band-gap oxides and implications for future electronic devices [J]. Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures, 2000, 18(3): 1785-1791.
- [87] Park Y, Choong V, Gao Y, et al. Work function of indium tin oxide transparent conductor measured by photoelectron spectroscopy [J]. Applied Physics Letters, 1996, 68(19): 2699-2701.
- [88] Azpiroz J M, Mosconi E, Bisquert J, et al. Defect migration in methylammonium lead iodide and its role in perovskite solar cell operation [J]. Energy & Environmental Science, 2015, 8(7): 2118-2127.
- [89] Löper P, Stuckelberger M, Niesen B, et al. Complex refractive index spectra of $\text{CH}_3\text{NH}_3\text{PbI}_3$ perovskite thin films determined by spectroscopic ellipsometry and spectrophotometry [J]. The Journal of Physical Chemistry Letters, 2015, 6(1): 66-71.
- [90] Sun J, Oh S, Choi Y, et al. Optoelectronic synapse based on IGZO-alkylated graphene oxide hybrid structure [J]. Advanced Functional Materials, 2018, 28(47): 1804397.
- [91] Seo S, Jo S H, Kim S, et al. Artificial optic-neural synapse for colored and color-mixed pattern recognition [J]. Nature Communications, 2018, 9: 5106.
- [92] Lü Z, Chen M, Qian F S, et al. Photonic synapse: mimicking neuroplasticity in a hybrid biopolymer transistor by dual modes modulation [J]. Advanced Functional Materials, 2019, 29(31): 1970212.
- [93] Ham S, Choi S, Cho H, et al. Photonic artificial synapses: photonic organolead halide perovskite artificial synapse capable of accelerated learning at low power inspired by dopamine-facilitated synaptic activity [J]. Advanced Functional Materials, 2019, 29(5): 1970031.