

科研进展 | 新型光学神经网络芯片，使快速训练成为可能

人工智能（AI）正以惊人的速度发展，其所消耗的计算资源每3~4个月翻一番，计算芯片的性能提升速度已难以跟上算力需求的增长速度。在此背景下，多种框架结构的张量核心应运而生，其中光学神经网络芯片可以填补传统微电子计算芯片在速度、延迟和能效等关键指标上的短板，被认为是提升AI算力的重要发展方向。

光学神经网络芯片主要用于执行占据人工神经网络90%算力的张量运算。光学神经网络芯片以光波作为载体进行计算，传输过程就是计算过程，其在计算速度和能效方面有着无可比拟的优势。光学神经网络芯片在计算过程中不会产生焦耳热，因此没有散热问题的困扰。光波的振幅、波长、偏振态和相位都可以携带信息进行计算，并且光路在交叉传输时互不干扰，相比于电子，光子更擅长于张量运算。近年，虽然光学神经网络芯片领域不断有新的技术被突破，但是由于光损耗、单元器件密度、波长通道数等因素的限制，人们对光学神经网络芯片能否高效地处理大型数据、以及能否在模型训练方面具有优势一直持怀疑态度。

针对以上问题，中山大学研究团队与不列颠哥伦比亚大学、女王大学、拉瓦尔大学、加拿大国家研究院、清华大学团队携手，基于铌酸锂薄膜光电集成技术，创新性地把光的波动性和电子的粒子性结合起来，在光学神经网络芯片领域取得了新的进展。他们提出了一种新的简洁的光学神经网络芯片框架结构——实现了一整层的神经网络，计算速度达到120 GOPS，同时输入和输出的数量能够被灵活地调整以应对多种人工智能任务。该光学神经网络芯片还支持快速原位训练，权重的刷新速度达到60 GHz。研究人员首次在光上实现了负数与负数的乘法，为聚类AI任务的训练提供了新的方案。另外，基于上述光学神经网络芯片，该工作还提出了首个端到端、纳秒级延迟、无需数字处理器辅助的超大型光学神经网络架构（见图1）。

相关结果以“120 GOPS Photonic tensor core in thin-film lithium niobate for inference and in situ training”为题在*Nature*

Communications 上发表。论文的第一作者为中山大学林忠劲副教授，通讯作者为中山大学蔡鑫伦教授和不列颠哥伦比亚大学 Lukas Chrostowski 教授。

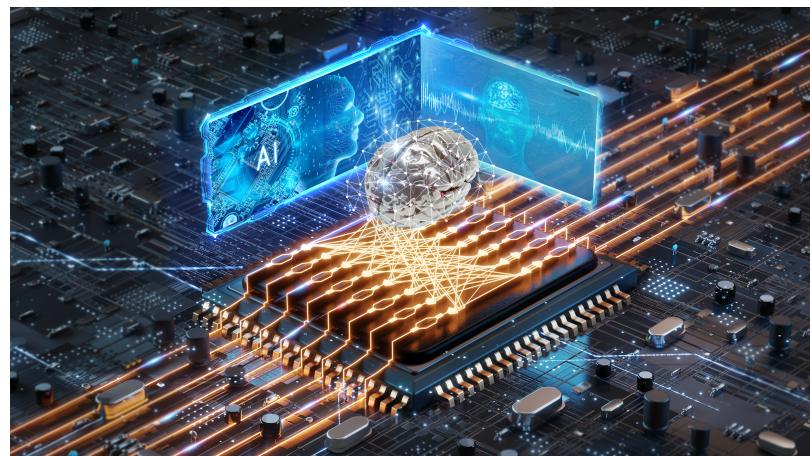


图 1 超大型光学神经网络架构概念图

工作原理与概念

如图 2 所示，与其他光学神经网络芯片复杂的结构不同，该工作提出的光学神经网络芯片只由 1 个激光器、2 个铌酸锂薄膜电光调制器、1 个平衡光探测器和 1 个光生载流子积分器构成。虽然结构如此简单，但却可以单独执行人工神经网络中的层与层之间的全连接计算。两个铌酸锂薄膜电光调制器分别负责把输入和权重转换为光信号，光依次穿过两个调制器后，输入和权重就完成了乘法操作并映射在光强上，然后利用积分器对不同时间的光强进行积分，就完成了加权总和操作。通过改变积分时间，可以随意改变输入的维度，因此所提的光学神经网络芯片可以执行超大维度输入的加权总和操作（见图 2c）。

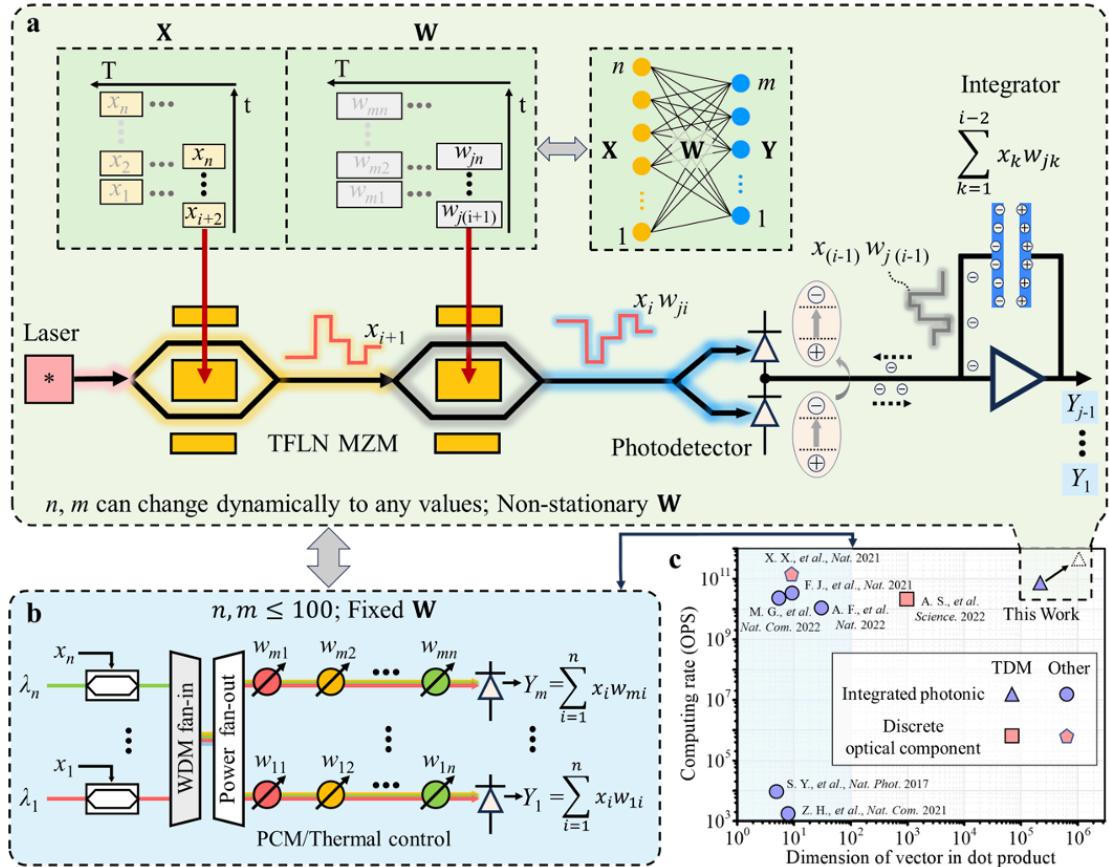


图 2 所提的光学神经网络芯片工作原理与概念

a 所提的光学神经网络芯片的结构示意图；b 基于波分复用的光学神经网络芯片的结构示意图；c 不同类型光学神经网络芯片在工作原理、尺寸、计算速度和加权总和的输入维度方面的性能对比；TDM:时分复用。

光学神经网络芯片全集成

该工作突破了铌酸锂薄膜光芯片的光子引线键合封装工艺，实现了铌酸锂薄膜光芯片和激光器的混合集成（见图 3）。激光器和铌酸锂薄膜光芯片之间通过聚合物三维光波导连接。由于光探测器不能直接在铌酸锂薄膜材料平台上制备，因此通过倒装焊的方式实现铌酸锂薄膜光芯片和光探测器之间的连接。

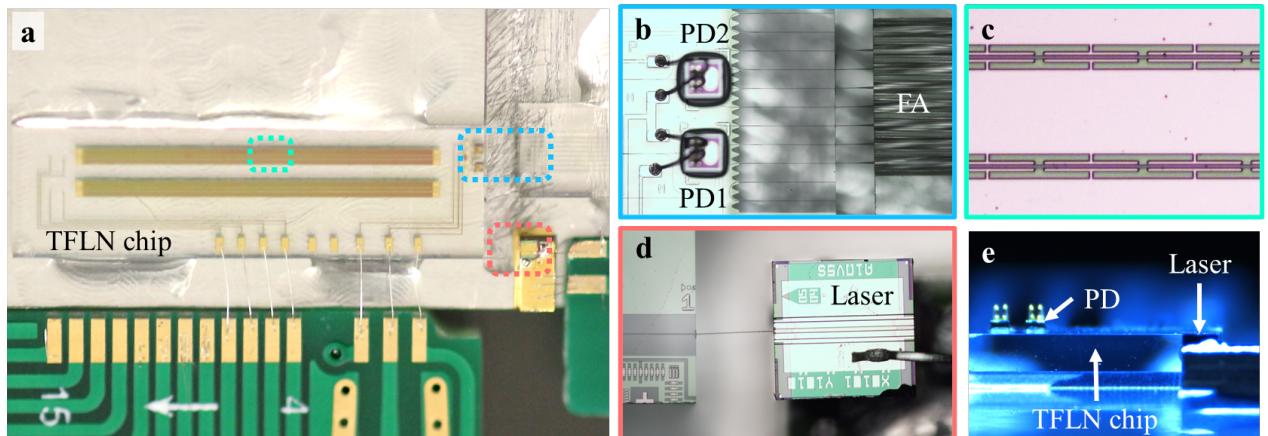


图 3 所提的光学神经网络芯片全集成

a 光学神经网络芯片整体照片；b-d 分别为光学神经网络芯片的光探测器、调制区域、激光器部位的显微镜图；e 光学神经网络芯片的侧视图以供理解光探测器、激光器和铌酸锂薄膜光芯片之间位置关系。

识别 AI 任务执行

为了评估所提的光学神经网络芯片在执行监督型机器学习任务方面的性能，如图 4a 所示，该工作用光学神经网络芯片构建了多层感知神经网络，并对 112×112 像素的数字图像进行识别。图 4b 展示的是基于所提的光学神经网络芯片进行原位模型训练的工作流程：输入和权重从中心处理器传送到光学神经网络芯片，光学神经网络芯片完成加权总和运算并把结果传回中心处理器。图 4c 表明所提的光学神经网络芯片的精度足以完成原位模型训练任务。另外，在图像识别方面，光学神经网络芯片的性能可以媲美全中心处理器（见图 4d 和 4e）。

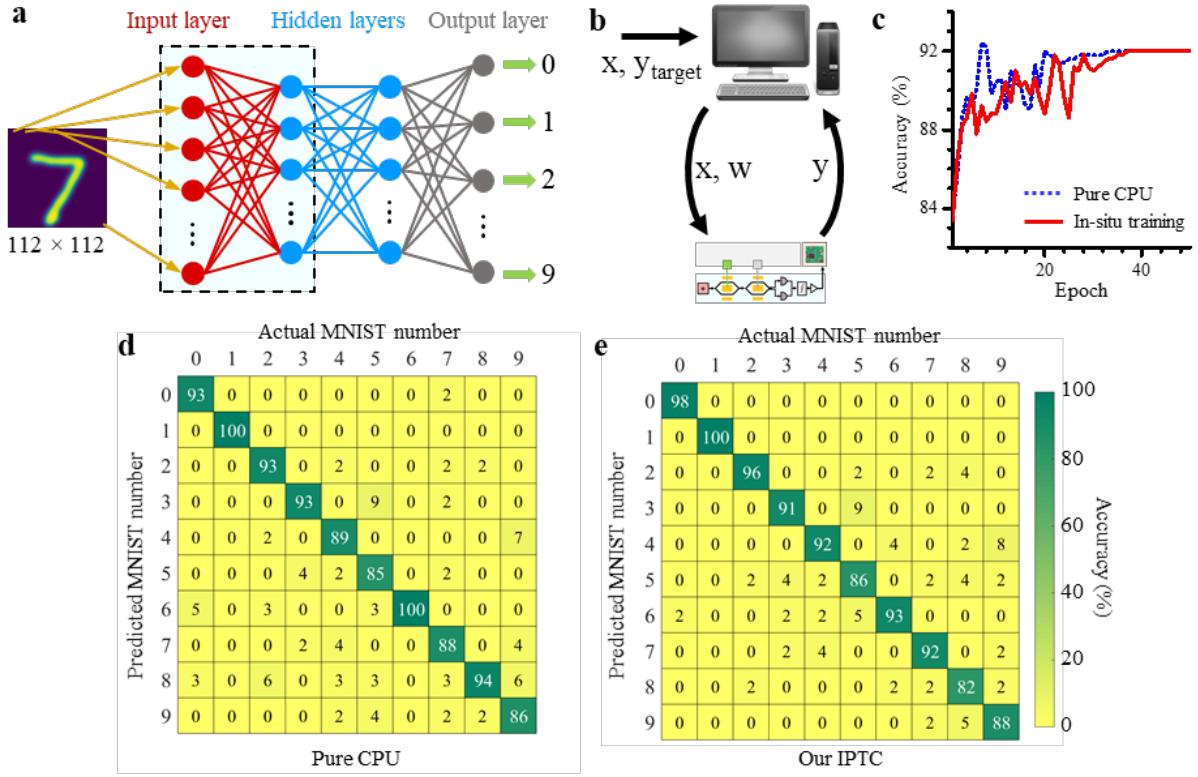


图 4 利用所提光学神经网络芯片执行图像识别任务

a 多层感知神经网络结构示意图；b 原位模型训练的工作流程 ‘c 光学神经网络芯片进行训练的迭代曲线’ d 和 e 分别为全中心处理器和所提光学神经网络芯片进行图像识别的结果。

聚类 AI 任务执行

与图像识别任务不同，图像聚类任务是无监督机器学习任务，其训练的过程涉及到负数与负数之间的乘法运算，然而在之前的光计算方案中，这一操作无法完成。针对这一问题，该工作提出了首个能够执行负数与负数相乘的方案（见图 5a）。图 5b 表明所提的光学神经网络芯片可以实现与全中心处理器的结果。图 5c 和 5d 展示的是利用所提的光学神经网络芯片实现的图像聚类结果。对于这个任务，光学神经网络芯片的处理速度是正常的中心处理器（Intel i9-9900 @ 3.10 GHz）的 5 倍。

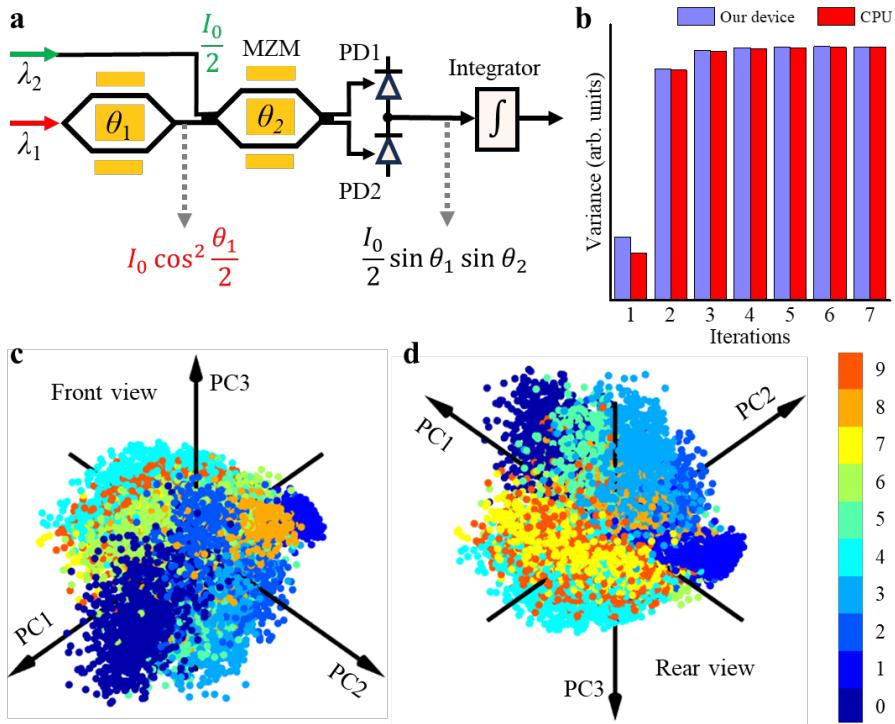


图 5 利用所提光学神经网络芯片执行图像聚类任务

a 负数与负数乘法运算的原理; b 光学神经网络芯片进行训练的迭代曲线; c 和 d 为所提光学神经网络芯片进行图像聚类任务的结果。

超大型光学神经网络结构

在上述光学神经网络芯片的工作基础上, 综合分析时分复用和波分复用的光学神经网络的优点, 该工作提出了新的超大型光学神经网络(见图 6)。在理论上, 该光学神经网络可以实现 500 TOPS 的计算速度, 并且能够在纳秒级别完成高维度数据的处理。

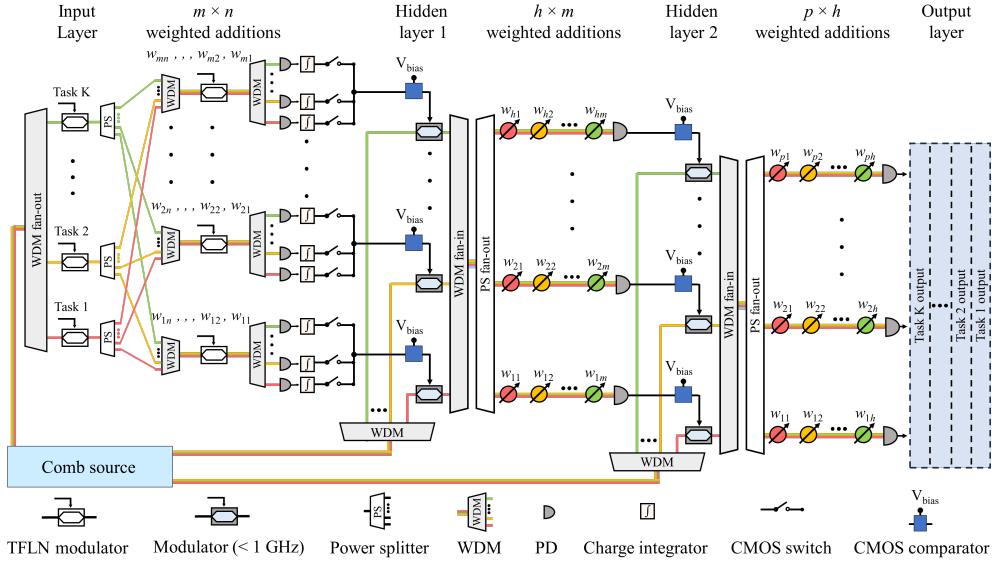


图 6 超大型光学神经网络结构示意图

总结与展望

利用该光学神经网络芯片，研究人员实现了 AI 系统性能的提升，以原位模型训练的方式，实现了 112×112 像素的数字图像识别（监督学习）和聚类（无监督学习）AI 任务。该研究工作不仅推动了光学神经网络芯片的发展，还为未来构建超大型光学神经网络奠定了基础。

（转载自“爱光学”）

原文链接：<https://doi.org/10.1038/s41467-024-53261-x>



林忠劲副教授

林忠劲，中山大学副教授，围绕光电子集成技术，致力于光学测量仪器小型化和智能化，包括片上偏振/光谱测量与调控、光学神经网络芯片、感算一体等。发表论文 40 多篇，包括 *Light: Science & Applications*、*Nature*

Communications、Optica、Laser & Photonics Reviews 等。片上偏振/光谱测量方面的研究成果被技术评估公司 SOVAR 评为 20 佳技术之一。