光子神经形态自主导航的硬件实现

Yonghang Chen[1], Shuiying Xiang[1*]、Xintao Zeng[1], Mengting Yu[1], Tao Zou[1],Shangxuan Shi[1], Xingxing Guo[1], Yanan Han[1], Yahui Zhang[1*]和 Yue Hao[1]

[1]西安电子科技大学综合业务网国家重点实验室，中国西安 *710071*。

关键词：光子神经形态计算、脉冲强化学习、光子脉冲神经元、自主导航

摘要：强化学习（RL）是推动人工智能（AI）从感知迈向决策的核心技术，但在传统电子硬件上的部署会受到冯·诺依曼架构带来的高延迟和高能耗限制。本文提出一种用于神经形态自主导航的光子脉冲双延迟深度确定性策略梯度（TD3）强化学习架构，并在带可饱和吸收体（DFB-SA）的分布式反馈激光器阵列上进行了实验验证。该混合架构将光子脉冲 Actor 网络与双连续值 Critic 网络相结合，其中 Actor 的末端非线性脉冲激活层部署在 DFB-SA 激光器

阵列上。在自主导航任务中，系统获得了 58.22±17.29 的平均奖励和 80%±8.3 % 的成功率。软硬件协同推理的估算能耗为 0.78 nJ/inf，超低延迟为 191.20 ps/inf；在有障碍物干扰和无障碍物干扰的任务场景中，协同推理错误率分别为 0.051% 和 0.059%。针对误差激活通道的仿真结果与预期响应完全一致，验证了 DFB-SA 激光器的动态特性。该架构展现出与大规模光子线性计算芯片集成的巨大潜力，可实现功能完备的光子计算以及低功耗、低延迟的神经形态自主导航。

1. 引言

强化学习（RL）是一类机器学习方法，智能体通过与环境持续交互来学习最优策略。其理论基础将行为心理学中的试错学习机制与最优控制理论中的奖励优化原则相结合 [1-4]。作为序列决策的核心范式，强化学习推动了多个领域的重大进展，包括自动驾驶 [5-7]，策略博弈 [8-11]，机器人控制 [12-14]，以及许多其他应用 [15-20]。近期的算法发展，包括双延迟深度确定性策略梯度（TD3） [21], 近端策略优化

（PPO）[22], 深度确定性策略梯度（DDPG）[23], 和深度Q网络（DQN）[15], 显著提升了强化学习处理高维状态空间、提高样本效率以及稳定策略学习的能力。然而，将这些算法部署在传统冯·诺依曼架构上，仍受制于存储器与计算彼此分离这一固有特性，由此造成显著的延迟和能耗开销，严重限制了强化学习在自主导航等应用中的实时性能 [24]。作为一种新兴的高性能计算范式，光子计算凭借超快的信号传播速度和固有的并行性，展现出巨大潜力 [25-28]。通过以光子而非电子作为信息载体，光子计算利用光的固有速度和并行特性，实现了高能效、低延迟的线性运算 [29-43]。现有的光子线性计算实现方案主要包括基于微环谐振器（MRR）的权重库 [31-33]，级联马赫-曾德尔调制器（MZM）[34, 35], 相变材料（PCM）交叉阵列 [36, 37], 以及马赫-曾德尔干涉仪（MZI）网格 [38-43]。然而，尽管近期已有进展 [44-46], ，实现功能完备的光子计算系统仍然面临挑战，尤其是在实现非线性运算方面。在此背景下，光子脉冲神经网络（PSNN）应运而生，成为一种前景广阔的计算框架，将光子计算的高速、高能效处理能力与脉冲神经网络的事件驱动和稀疏动态特性相结合 [47-55]。2019年，Feldmann等人通过实验展示了首个完全集成的光学脉冲神经形态

基于相变材料的网络，实现了对光学编码字母的模式识别，每次突触权重更新的能耗约为 60 pJ [47]。2020年，Robertson等人报告了一种基于垂直腔面发射激光器（VCSEL）的光子脉冲神经元，实现了对数字图像中定向边缘和全向边缘的超快速检测 [48]。2021年，Xiang等人提出了一种完全基于VCSEL、用于监督学习的全光脉冲神经网络；理论上展示了光学数字分类，同时避免了在光学系统中难以实现的横向抑制机制 [49]。2022年，Lee等人提出并实现了一种基于Izhikevich模型的事件驱动光电脉冲神经元，在MNIST手写数字识别任务中达到97%的准确率，能效为每焦耳每秒50万亿次运算（TOPS/J）[50]。2023年，Xiang等人制造并通过实验验证了一种基于带可饱和吸收体的法布里-珀罗激光器（FP-SA）的光子脉冲神经元芯片，通过软硬件协同计算实现了单神经元模式分类 [51]。同年，该团队进一步开发并展示了一种基于DFB-SA 激光器的光子突触核心芯片，该芯片首次在单个器件中同时实现线性加权和非线性脉冲激活，在MNIST数据集上达到87%的准确率 [52]。最近，Talukder等人构建了一个包含40,000个神经元的大规模光子脉冲神经网络，利用具有慢速抑制反馈的改进Ikeda映射实现仿生兴奋性。该网络仅使用8.5%的神经元，就在MNIST上达到77.5%的准确率，并首次展示了与硬件兼容的同时扰动随机逼近（SPSA）训练

[53]。这些进展共同展现了光子脉冲神经网络在能效、处理速度和稀疏性方面的优势。尽管如此，其在自主导航领域的应用仍未得到探索。Gao等人提出了一种采用法布里-珀罗激光器的、基于光子脉冲神经元的避障系统，该系统利用频率编码模拟生物单目视觉，并通过实验验证了超快速、低功耗机器人避障的可行性 [54]。然而，该系统的决策过程依赖预设阈值，缺乏强化学习智能体与环境之间的自主交互和在线学习，因此无法实现全流程神经形态自主导航。因此，将光子脉冲神经网络的超快速感知决策能力与强化学习的序贯决策优化能力相结合，有望为构建智能、自适应、低功耗、低延迟的神经形态自主导航系统提供一条可行途径。

本文提出了一种用于神经形态自主导航的光子脉冲 TD3 强化学习架构，通过软硬件协同推理实现高能效的智能决策。本文的主要贡献如下：

1. 我们提出了一种光子脉冲 TD3 强化学习框架，将基于脉冲的 Actor 网络与双连续值 Critic 网络相结合。该算法针对光子神经元芯片的硬件约束进行了优化，其中 Actor 网络采用正权重无偏置结构，以确保硬件兼容性。在自主导航任务中，所提出的架构取得了 58.22±17.29 的平均奖励和 80%±8.3% 的导航成功率

在软件仿真中，证明了采用光子脉冲神经网络进行自主导航的可行性。

2. Actor网络的最终非线性脉冲激活层部署在DFB-SA激光器阵列上。该系统的估算推理能耗为 0.78 nJ/inf，延迟低至 191.20 ps/inf，凸显了光子计算在强化学习中的固有低功耗和低延迟优势。

3. 在有障碍物干扰和无障碍物干扰两种场景下进行了软硬件协同推理实验。协同推理误差率分别为 0.051% 和 0.059%，验证了基于 PSNN 的强化学习在神经形态自主导航中的高精度和可靠性。此外，使用算法生成的输入对误差激活通道进行山田模型仿真的结果与预期的脉冲响应完全一致，从理论上验证了 DFB-SA 激光器的动态特性。

2. 方法

2.1. 光子脉冲 TD3 强化学习架构

所提出的光子脉冲强化学习架构基于 TD3 算法，这是 DDPG方法的一种增强型变体。TD3 旨在缓解 Actor-Critic框架中函数近似误差导致的 Q值过估计偏差，从而提升算法稳定性和整体性能 [21]。这一目标通过三项关键创新实现：采用双 Critic

网络（裁剪双重Q学习）、引入目标策略平滑正则化，以及采用延迟策略更新。这些机制共同稳定训练、提高策略优化的准确性，并提升复杂连续控制任务中的性能。凭借出色的稳定性和可扩展性，TD3 已广泛应用于需要在高维状态—动作空间中进行精确决策的领域，包括机器人控制 [56, 57], 自动驾驶[58, 59], 和无人机通信 [60, 61]。

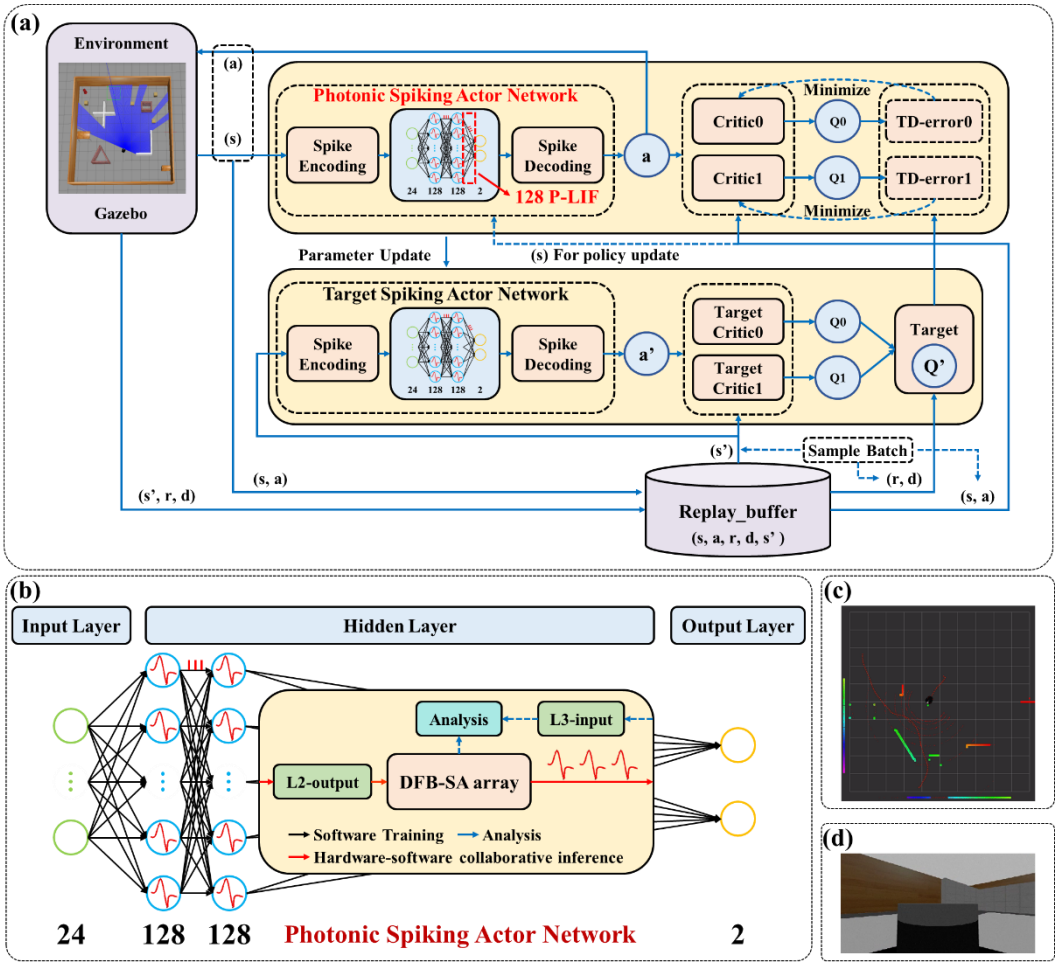


图1。光子脉冲TD3强化学习系统的总体架构与实现细节。(a) 光子脉冲TD3强化学习系统的总体架构。(b) 软硬件协同推理框架。(c) 实时导航路径图。(d) 机器人的机载摄像头视图。

图 1(a) 展示了所提出的光子脉冲 TD3 强化学习架构。强化学习智能体由基于 PSNN 的 Actor 网络和两个基于人工神经网络（ANN）的 Critic 网络组成。光子脉冲 Actor 网络根据当前状态生成两个连续动作——机器人的线速度和角速度，而 Critic 网络则评估相应的状态-动作 Q 值。采用基于 ANN 的 Critic 网络能够实现更稳定的梯度估计并加快收敛，从而与光子脉冲 Actor 网络的事件驱动和高能效特性相辅相成，形成均衡的学习框架。我们在 Gazebo 仿真环境中考虑了一项连续控制任务 [62]，，机器人通过激光雷达（LiDAR）传感器感知障碍物，并通过强化学习实现无地图动态避障和目标导向导航。整个系统由三个核心模块组成：智能体、Gazebo 仿真环境和差速驱动控制系统，它们通过机器人操作系统（ROS）框架相互连接，从而实现基于激光雷达的无地图动态导航 [63]。仿真环境、状态空间和动作空间的详细配置见 SI-1。

在自主导航任务中，光子脉冲 Actor 网络处理来自 Gazebo 仿真环境的实时状态数据。静态状态输入首先通过脉冲编码进行转换：沿时间维度扩展，并在 T 个时间步上重复，以匹配脉冲神经网络的时间序列处理范式。随后，编码序列被传入光子脉冲 Actor

网络，其中每个隐藏层均采用漏积分发放（LIF）神经元来产生时序脉冲响应。随后，通过时间聚合（即对时间取平均）对这些脉冲输出进行解码，以还原静态表征，再使用双曲正切函数对其归一化，从而生成连续的速度指令。训练过程中，在动作中注入高斯噪声以鼓励探索。此外，还采用强制探索策略： 当激光雷达检测到附近有障碍物时，智能体会以一定概率执行随机动作，以摆脱局部最优。每次交互后，环境都会返回奖励、下一状态和终止信号，并将这些信息存储在经验回放缓冲区中。进行策略评估时，从经验回放缓冲区中随机采样的小批次数据由采用双Q架构的双人工神经网络评论家网络（800×600×1）处理。目标网络通过最小化时序差分误差来稳定 Q 值估计。目标动作由一个独立的目标光子脉冲演员网络生成，并加入截断噪声以减轻 Q 值高估；该目标 Actor 遵循与主演员网络相同的脉冲编码–LIF激活–脉冲解码流程。Actor 网络每训练两步更新一次，利用评论家网络提供的梯度最大化期望状态–动作值。这种延迟策略更新策略与软目标网络同步相结合，可确保收敛过程稳定且高效。完整算法流程的伪代码见 SI-2。

2.2. 硬件感知算法优化与光子实现 n

为确保硬件兼容性和通用适用性，光子脉冲

Actor网络在硬件感知约束下进行协同设计，以便部署于通用光子线性计算平台。尽管当前光子线性处理器已通过实验展示了最高达 128×128 [43] 的矩阵运算规模，我们仍将 Actor 网络配置为 24×128×128×2 ，作为具有代表性且兼容硬件的示例。此配置展示了我们的方法在现有光子线性计算架构能力范围内的可扩展性和可实现性，同时保持较高的控制精度。在软件预训练期间，隐藏层权重被限制为正值，并移除所有偏置项，从而确保训练参数可直接映射至光子线性计算硬件。最终的非线性脉冲激活层通过 DFB-SA激光器阵列部署于硬件中，并利用激光脉冲动力学执行光学非线性计算。该设计可与大尺度光子矩阵-向量乘法芯片无缝集成，实现全功能光子计算架构，其中线性和非线性运算均完全在光学域中执行。仿真结果验证了该架构的性能和稳定性，为后续硬件部署及协同推理实验提供了理论指导。

2.3. 软硬件协同推理架构

为验证所提出的光子脉冲 TD3 强化学习架构的有效性，采用软硬件协同推理方案来实现光子脉冲 Actor 网络。图1(b)展示了整体协同推理

架构。该过程从软件预训练开始：使用单个时间步（T = 1）和替代梯度法从头训练光子脉冲 Actor 网络，获得已纳入第2.2节所述硬件约束的预训练模型。预训练完成后，使用 DFB-SA激光器阵列在硬件中部署最终的非线性脉冲激活层，通过激光脉冲动力学实现光学非线性计算。随后，将预训练软件模型对应的输入-输出数据馈入光子硬件，以评估其响应。将实验测得的光学输出与软件生成的对应结果进行比较，以量化协同推理误差率，从而验证光学推理与数字推理之间的一致性。图1(c)和图1(d)分别展示了机器人的实时导航轨迹和机载摄像头视图，直观呈现仿真环境以及机器人的自主导航行为。

2.4 实验装置

光子脉冲 Actor 网络的最终非线性脉冲激活层采用 DFB-SA激光器阵列进行实验部署。DFB-SA 激光器的制备细节与参考文献 [64] 中所报告的相似。图 2(a) 展示了该层的实验装置，包括实验中所使用的 DFB-SA激光芯片的显微图像。在示意图中，红线表示光路，蓝色虚线表示电路。

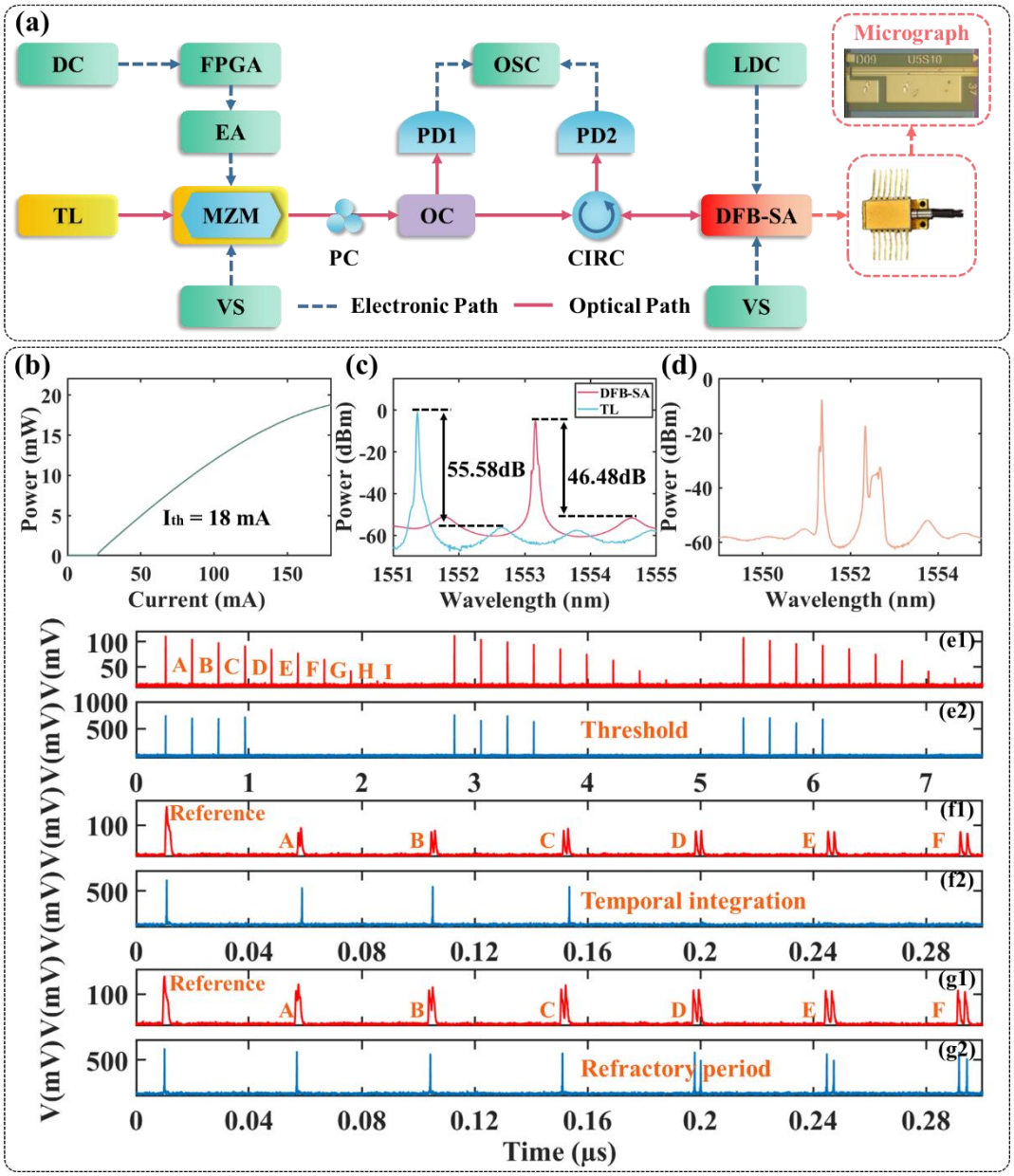


图 2。基于 DFB-SA 激光器的光子脉冲神经元的实验装置及特性。(a) 光子脉冲 Actor 网络中最终激活层的实验装置。DC：数字计算机；FPGA：现场可编程门阵列；EA：电子放大器；TL：可调谐激光器；MZM：马赫-曾德尔调制器；VS：电压源；PC：偏振控制器；OC：光耦合器；PD：光电探测器；OSC：示波器；CIRC：环形器；LDC：激光二极管控制器。(b) DFB-SA 激光器的 PI 曲线。(c) TL 和 DFB-SA 激光器的光谱。(d) DFB-SA 激光器的光谱。(e1)–(g2) DFB-SA 激光器的类神经元响应。

可调谐激光器（TL）提供光载波，并将其注入马赫-曾德尔调制器（MZM，Fujitsu FTM7928FB），而电压源（VS）提供驱动电压。由现场可编程门阵列生成的电子信号

（现场可编程门阵列）在数字计算机（DC）的控制下，经电子放大器（EA）放大后注入马赫-曾德尔调制器进行调制。随后，调制后的光信号经由三端口光环行器（CIRC）注入 DFB-SA 激光器，并由偏振控制器（PC）调节偏振态。激光二极管控制器（LDC，ILX Lightwave LDC3724B）为 DFB-SA 激光器的增益区提供高精度温度控制和低噪声偏置电流，同时电压源向可饱和吸收体施加反向偏置电压。DFB-SA 激光器的输出由光电探测器（PD，安捷伦 HP11982A）探测，并用示波器（OSC，是德科技 DSOZ592A）记录，随后与从光耦合器（OC）分出的调制光信号一并进行分析。图 2(b) 给出了 DFB-SA 激光器的 PI 曲线，其阈值电流为 18 mA。图 2(c) 给出了自由运行可调谐激光器和 DFB-SA 激光器的光谱，测得二者的边模抑制比（SMSRs）分别约为 55.58 dB 和 46.48 dB，表明其具有优异的单模性能和光束质量，这对实验至关重要。图 2(d) 给出了在来自可调谐激光器（TL）的注入下 DFB-SA 激光器的输出光谱。图 2(e1)-(g2) 展示了 DFB-SA 激光器的关键神经拟态动力学——阈值、时间积分和不应期——从而验证了其模拟复杂光子脉冲神经元行为的能力。实验过程中，DFB-SA 激光器温度保持在 22 °C，可饱和吸收体的反向偏置电压设为 0.76 V，增益区的偏置电流设为 27.9 mA，注入功率 P_{inj} 为 97.96 μW。

2.5 基于山田模型的 DFB-SA 激光器理论模型

2002年，山田稔提出了对自持脉动现象的理论分析，并研究了速率方程在窄带结构半导体激光器中的应用。该模型可用于分析 DFB-SA 激光器的动态特性 [65]。在此，我们修改山田模型，以模拟基于 DFB-SA 激光器的光子脉冲神经元。山田模型的数学表达式如下：

$$\frac{dN_1}{dt} = \frac{I}{eV_1} - \frac{N_1}{\tau_s} - \frac{\Gamma_1 g_1}{V_1}(N_1 - N_{g_1})S \tag{1}$$

$$\frac{dN_2}{dt} = -\frac{N_2}{\tau_s} - \frac{\Gamma_2 g_2}{V_2}(N_2 - N_{g_2})S \tag{2}$$

$$\frac{dS}{dt} = [\Gamma_1 g_1(N_1 - N_{g_1}) + \Gamma_2 g_2(N_2 - N_{g_2})]S - \frac{S - S_{\text{ext}}}{\tau_{\text{ph}}} + \beta_{\text{sp}}\frac{N_1 V_1}{\tau_{\text{s}}} \tag{3}$$

此处，下标 1 和 2 分别表示增益区和可饱和吸收体。N表示载流子密度，g 表示增益系数，Γ 表示限制因子，V表示腔体积，N_g 表示透明载流子密度。S_{ext} 表示外部注入光子的数量，β_{sp} 表示自发发射系数，τ_{ph} 表示光子寿命，τ_{s} 表示载流子寿命，I表示施加于增益区的偏置电流。为简化起见，模型未考虑增益区与可饱和吸收体之间的载流子扩散。相应的参数值列于表 1。

表 1. DFB-SA 激光器的对应参数值

参数	增益区	可饱和吸收体
增益系数 g	1.8794 × 10^{11} 米³/秒	0.594 × 10^{-11} m^3/s
限制因子 Γ	0.08	0.08
腔体积 V	172.8 × 10^{-18} m^3	7.2 × 10^{-18} m^3
透明载流子密度 N_g	1.4 × 10^{24} m^{-3}	1.6 × 10^{24} m^{-3}
光子寿命 τ_{ph}	2.56 × 10^{-3} ns	2.56 × 10^{-3} ns
载流子寿命 τ_{S}	1 ns	1 ns

3. 结果

本节介绍采用光子脉冲 TD3 强化学习架构并结合软硬件协同推理进行神经形态自主导航的实验结果。图 3(a1) 和 3(a2) 分别展示了基线软件算法的平均奖励曲线和导航成功率曲线。由于训练曲线在收敛后仍出现振荡，因此对最后 20% 的训练数据计算均值和标准差，得到平均奖励 58.22±17.29，成功率为 80%±8.3%。算法所用的环境配置和训练参数见补充信息 SI-3。收敛后，选取有障碍物干扰和无障碍物干扰两种任务场景，对光子脉冲 Actor 网络中最终的非线性脉冲激活层进行实验验证。由于机器人的初始位置和目标位置均为随机设定，不同任务中的状态-动作样本对数量有所不同。两种任务场景（有障碍物干扰和无障碍物干扰）的演示视频分别见补充视频 1 和补充视频 2。

图 3(b1) 和 3(b2) 展示了起点和目标位置、完整的导航轨迹，以及预期脉冲激活输出与实验测得的脉冲激活输出之间的差异分布，其中 1 表示假阳性脉冲（过度激活），-1 表示假阴性脉冲（漏激活）。实验结果证实，DFB-SA 阵列的激活输出与预期激活高度吻合。为便于表述，将无障碍物干扰的场景称为任务 1，将存在障碍物干扰的场景称为任务 2。

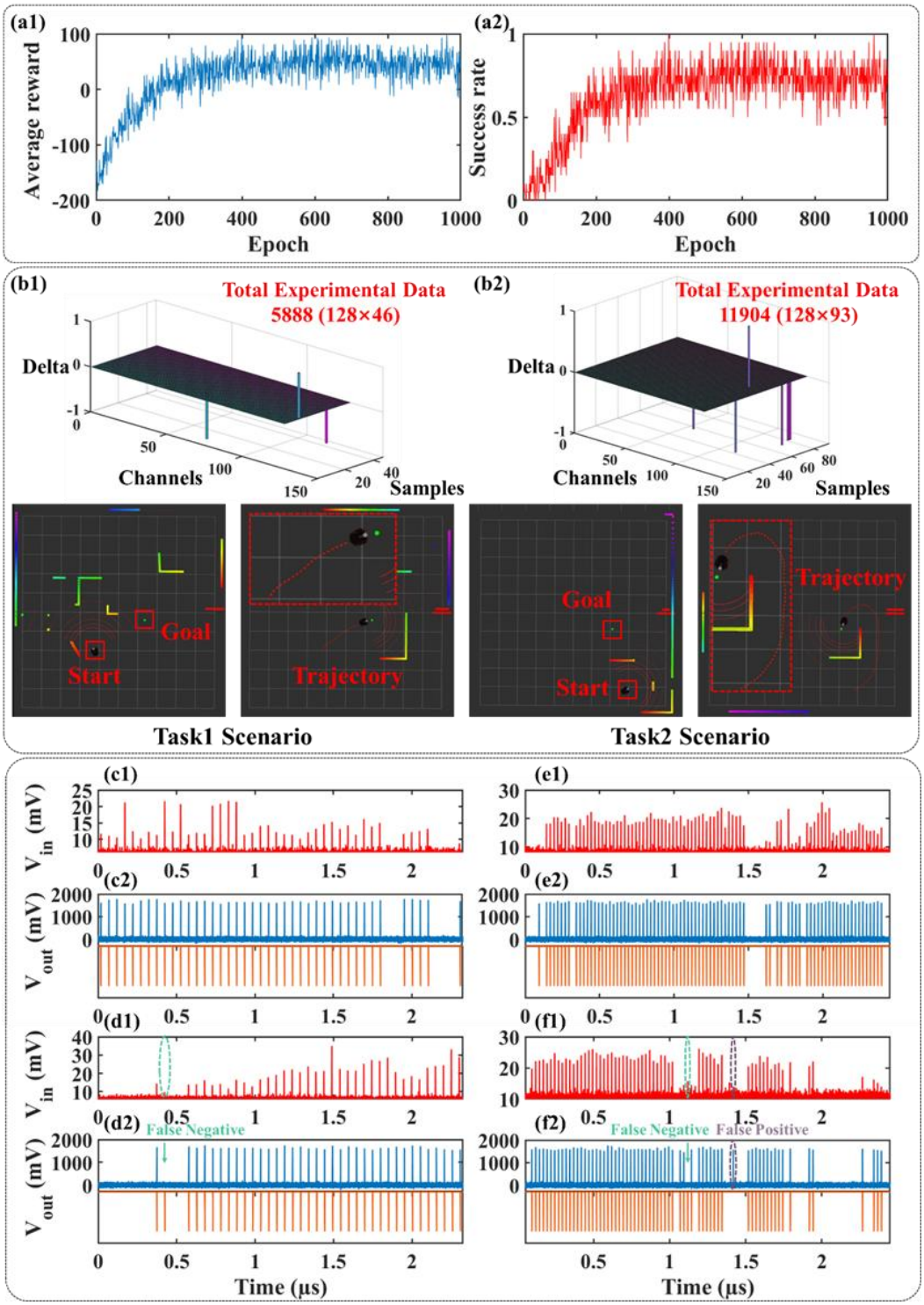


图3。自主导航任务的软件训练结果和实验非线性输出。(a1) - (a2)：软件训练的平均奖励和成功率。(b1) - (b2)：两个任务场景中预期脉冲激活输出与实验测得的脉冲激活输出之间的差异分布。(c1) - (f1)：两个任务场景下实验输入和输出的对比示例。红色曲线表示实验输入，蓝色曲线表示实验测得的脉冲激活输出，橙色曲线表示预期的脉冲激活输出。

在任务 1 中，5,888 个样本（128×46）中仅观察到 3 处差异，软硬件协同推理误差率为 0.051%。在任务 2 中，11,904 个样本（128×93）中出现了 7 处差异，对应的误差率为 0.059%。图 3(c1)–(f2) 展示了两个任务场景中具有代表性的输入输出对比。对于任务 1，图 3(c1)–(c2) 显示通道 1 中的脉冲激活匹配正确，而图 3(d1)–(d2) 显示通道 69 中存在不匹配：DFB-SA 激光器未能在 0.43 μs 时产生脉冲。对于任务 2，图 3(e1)–(e2) 展示了通道 1 中匹配正确的脉冲激活，而图 3(f1)–(f2) 则显示通道 114 中在 1.12 μs 时出现了过度激活的脉冲，并在 1.42 μs 时缺少一个脉冲。所提出的架构具有良好的可扩展性，可扩展至两个已测试场景之外的多种导航任务。SI-4 中提供了所有激活不匹配情况的完整输入输出对比。

此外，我们评估了所提出的光子脉冲 TD3 强化学习架构的推理延迟和能耗。单个DFB-SA激光芯片的功耗估计为$1.097 \times 27.9 + 0.76 \times 1.93 = 32.07$ mW，对应于128个器件的总功耗$P_{\text{DFB-SA}} = 4.10$ W，这证明了其低功耗优势[52]。单个DFB-SA激光芯片的自脉动频率为5.23 GHz，对应于电光延迟$\tau_{\text{DFB-SA}} = 1 / 5.23\ \text{GHz} = 191.20$ ps [66]。因此，光子脉冲 TD3 强化学习架构实现了每次推理191.20 ps的推理延迟，能耗为$E_{\text{inf}} = P_{\text{DFB-SA}} \times \tau_{\text{DFB-SA}} = 0.78$ nJ/inf。与Eyeriss架构[67] 的2083 ps/inf推理延迟以及

PopSAN 架构[68], 的11.98 μJ/inf能耗相比，所提出的方法显著降低了延迟，并显著提高了能效。

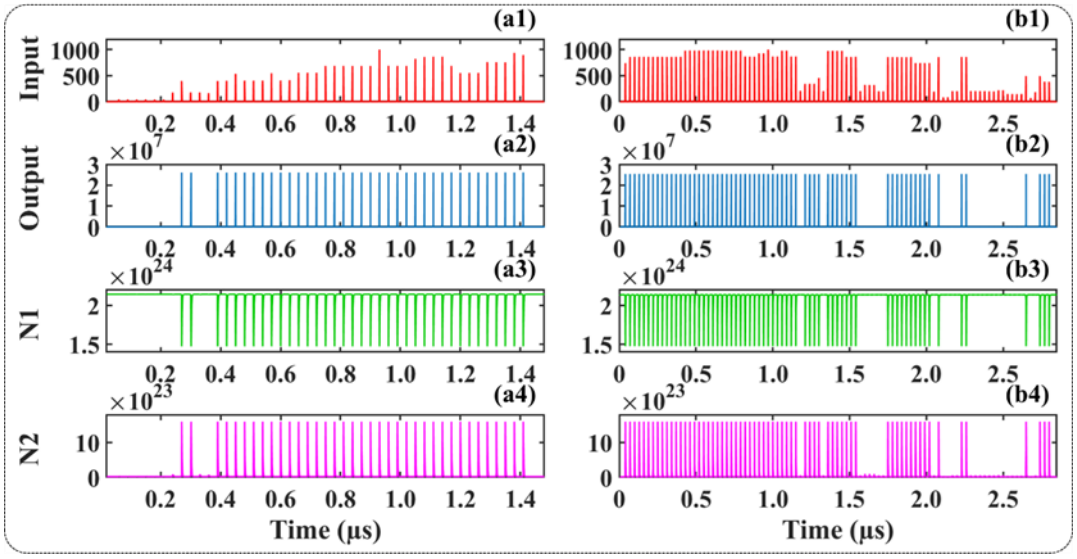


图 4。自主导航任务中误差激活通道输入数据对应的 DFB-SA 激光器仿真结果。（a1）-（a4）：任务 1 中通道 69 的仿真结果；（b1）-（b4）：任务 2 中通道 114 的仿真结果。（a1）、（b1）：外部注入光信号；（a2）、（b2）：DFB-SA 激光器的输出；（a3）、（b3）：增益区中的载流子密度；（a4）、（b4）：可饱和吸收体中的载流子密度。

最后，基于实验结果，我们进一步使用山田模型仿真误差激活通道对应的算法输入数据，以从理论上验证 DFB-SA 激光器的动态特性。代表性仿真结果如图 4 所示，其中图 4(a1)–(a4) 对应任务 1 中的通道 69，图 4(b1)–(b4) 对应任务 2 中的通道 114。可以看出，仿真的脉冲激活与预期输出完全一致，证实了实现 DFB-SA 激光器精确激活的理论可行性。两项任务中所有误差激活通道的完整仿真结果见 SI-5。

4. 讨论

本节进一步研究更大网络规模和更长时间步下光子脉冲神经网络的性能。为确保公平比较，仅改变光子脉冲 Actor 网络的隐藏层维度和时间步长，其他所有参数保持不变，并取消光子硬件兼容训练中施加的正权重约束。在网络规模分析中，较小的网络表现出更强的振荡和较差的收敛性（100×100）；因此，在固定时间步 T = 1的条件下，分别采用隐藏层维度 128×128、 256×256 和 512×512 进行训练比较。在时间步分析中，将隐藏层维度固定为 128×128，并将时间步设为 T = 1、2 和 4。基线软件算法的平均奖励和导航成功率曲线如图 5 所示。

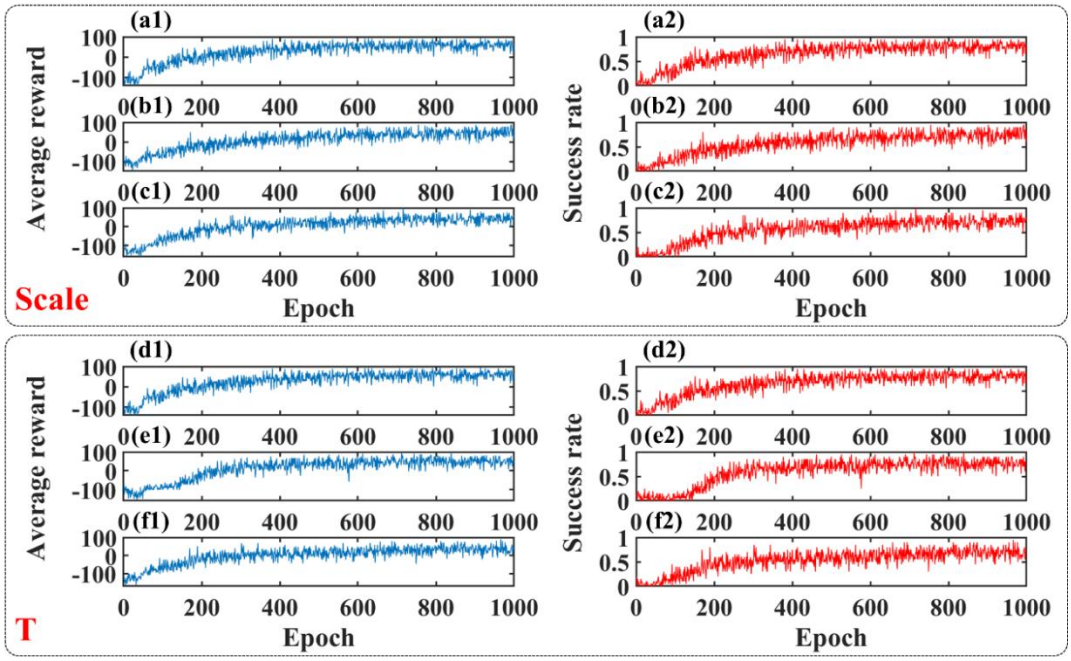


图5. 不同网络尺度和时间步下软件训练的平均奖励和导航成功率。隐藏层维度：(a1) - (a2)，128×128；(b1) - (b2)，256×256；(c1) - (c2)，512×512。时间步：(d1) - (d2)，T = 1；(e1) - (e2)，T = 2；(f1) - (f2)，T = 4。

随后，我们分析了训练结果。由于曲线在收敛后出现振荡，因此对训练数据最后20%的部分计算均值和标准差，并以“均值 ± 标准差”的形式汇总结果，如表2所示。结果表明，无论是增大隐藏层维度，还是延长时间步，都无法有效提升 PSNN 的性能。相反，这两个因素都容易使训练过程不稳定，导致奖励曲线出现更明显的振荡。可能的解释包括反向传播期间梯度消失或梯度爆炸、网络参数数量增加导致过拟合，以及硬件资源限制造成性能下降。尽管较大的网络和较长的时间步偶尔能达到更高的峰值性能——某些训练周期中的导航准确率最高可达100%——但它们在训练早期也会产生较低的奖励值。相比之下，隐藏层维度为 128×128 、时间步为 T = 1 时，能够在最大限度降低硬件资源需求的同时，为自主导航任务实现稳定高效的训练。

表2. 不同隐藏层维度和时间步的训练结果

消融研究	**参数**	**平均奖励**	**成功率**
尺度	100 × 100	37.73 ± 20.58	69.7% ± 10.1%
	128 × 128	58.22 ± 17.29	80% ± 8.3%
	256 × 256	44.60 ± 20.97	74% ± 10.3%
	512 × 512	40.25 ± 19.26	71.6% ± 9.5%
T	1	58.22 ± 17.29	80% ± 8.3%
	2	48.82 ± 18.71	75.8% ± 9.1%
	4	35.36 ± 20.92	69.4% ± 10%

5. 结论

本研究提出并通过实验验证了一种用于神经形态自主导航的光子脉冲 TD3 强化学习架构，通过硬件与软件协同推理实现了高能效的智能决策。该架构结合了基于脉冲的 Actor 网络与双连续值 Critic 网络，并在 Gazebo 仿真环境中成功完成导航任务，平均奖励为58.22±17.29，导航成功率为80%±8.3%。光子 Actor 网络的末端非线性脉冲激活层由 DFB-SA激光器阵列实现。该系统的推理能耗估算为0.78 nJ/inf，超低延迟为191.20 ps/inf，凸显了光子计算在强化学习中固有的低功耗和低延迟优势。硬件–软件协同推理进一步展现出很高的准确率，在无障碍场景和障碍物干扰场景中的误差率分别低至0.051%和0.059%。此外，我们针对误差激活通道，基于算法生成的结果开展了山田模型仿真，结果与预期的脉冲响应完全一致，从而从理论上验证了 DFB-SA 激光器的动态特性。这些结果为将所提架构与大规模光子线性计算芯片集成、实现全功能光子计算奠定了坚实基础。针对隐藏层维度和时间步的消融研究进一步证实了所采用网络配置的合理性。尽管

当前工作聚焦于相对简单的导航任务，未来将致力于大规模系统集成、多任务泛化以及光电协封装。利用光子计算固有的并行性和超快响应，所提出的方法为实现下一代实时神经形态自主导航系统提供了一条颇具前景的途径。

数据可用性

支持本研究结论的数据可根据合理请求向通讯作者获取。

通讯作者

*电子邮箱： syxiang@xidian.edu.cn

*电子邮箱：zhangyahui@xidian.edu.cn

支持信息

支持信息可从Wiley在线图书馆或作者处获取。

致谢

本研究得到国家自然科学基金委员会

（编号：62535015、62575231）；中央高校基本科研业务费（QTZX23041）；西安电子科技大学交叉学科探索专项资助项目（TZJH2024009）的资助。

经济利益冲突

作者声明不存在任何经济利益冲突。

个人简介



Yonghang Chen 于2002年出生于中国陕西省。他目前在中国西安的西安电子科技大学攻读硕士学位。他的研究兴趣包括光子神经形态计算和光子脉冲神经网络。



Shuiying Xiang于1986年出生于中国江西省。她获得了博士学位 .

2013年获中国成都西南交通大学学位。

她现任中国西安西安电子科技大学综合业务网国家重点实验室教授。她是140多篇研究论文的作者或合著者。她的研究兴趣包括垂直腔面发射激光器、神经形态光子系统、类脑信息处理、混沌光通信和半导体激光器动力学。

Xintao Zeng 于1998年出生于中国福建。目前，他正在西安中国西安电子科技大学攻读学士学位。他的研究兴趣包括光子神经形态计算、光子脉冲神经网络、光电芯片与集成技术，以及硅光电芯片与集成技术。

Mengting Yu于2002年出生于中国江西省。她正在中国西安的西安电子科技大学攻读硕士学位。她的研究兴趣包括硅光子MZI芯片，以及基于机器人学的光子脉冲强化学习。

Tao Zou 于2000年出生于中国四川省。他正在中国西安的西安电子科技大学攻读硕士学位。他的研究兴趣包括现场可编程门阵列、类脑信息处理和脉冲神经网络。

Shangxuan Shi 于2000年出生于中国山西省，目前正在中国西安的西安电子科技大学攻读理学硕士学位，研究方向为光子神经形态计算和光子脉冲神经网络。

Xingxing Guo于1993年出生于中国江西省吉安市。她在中国西安的西安电子科技大学获得博士学位。她的研究兴趣包括垂直腔面发射激光器、储备池计算和混沌通信。

Yanan Han于1996年出生于中国宁夏回族自治区。她在中国西安电子科技大学获得博士学位。她的研究兴趣包括半导体激光器的动力学及其应用、随机数发生器和类脑信息处理。

Yahui zhang于1993年出生于中国河北省。她在中国西安电子科技大学获得博士学位。她的研究兴趣包括神经形态光子系统、类脑信息处理和随机数发生器。



Yue Hao于1958年出生于中国重庆。他于1991年在中国西安的西安交通大学获得博士学位。他是IEEE高级会员、中国电子学会常务理事，以及陕西省电子学会、集成电路行业协会和陕西省半导体照明协会执行委员会主席。他是

中长期规划中“核心电子器件、高端通用芯片和基础软件产品”重大科技专项实施专家组组长。他是中国人民解放军总装备部微电子技术专家组组长。他是电子信息科学与工程专业全国教学指导委员会副主任委员。曾任第九届、第十届中国人民政治协商会议委员和第十一届中华人民共和国全国人民代表大会代表。他现任西安电子科技大学副校长，任职于宽带隙半导体技术国家重点学科实验室和西安电子科技大学微电子学院，工作地点为中国西安。他目前的研究方向包括宽禁带半导体材料与器件。

参考文献

[1]. R. S. Sutton, A. G. Barto, 《强化学习》, J. Cogn. Neurosci., 1999, 11, 126。[2]. R. S. Sutton, A. G. Barto, 《强化学习：入门》, MIT出版社, 1998。[3]. X. Wang等, 《深度强化学习：综述》, IEEE神经网络与学习系统汇刊, 2022, 35, 5064。[4]. V. François-Lavet等, 《深度强化学习导论》, Found. Trends Mach. Learn., 2018, 11, 219。[5]. S. Aradi, 《自动驾驶车辆运动规划中的深度强化学习综述》, IEEE Trans. Intell. Transp. Syst., 2020, 23, 740。[6]. D. Isele等, 《自动驾驶车辆中的安全强化学习》, Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., 2018, 1。[7]. J. Dinneweth等, 《自动驾驶车辆的多智能体强化学习：综述》, Auton. Intell. Syst., 2022, 2, 27。[8]. D. Silver等, 《借助深度神经网络和树搜索掌握围棋》, Nature, 2016, 529, 484。[9]. D. Silver等, 《在没有人类知识的情况下掌握围棋》, Nature, 2017, 550, 354。[10]. D. Silver等, 《通过自我对弈掌握国际象棋、将棋和围棋的通用强化学习算法》, Science, 2018, 362, 1140。[11]. V. Mnih等, 《通过深度强化学习实现人类水平控制》, Nature, 2015, 518,

529。[12]. P. Kormushev等，《机器人学中的强化学习：应用与现实挑战》，Robotics，2013，2，122。[13]. L. Brunke等，《机器人学中的安全学习：从基于学习的控制到安全强化学习》，Annual Review of Control, Robotics, and Autonomous Systems，2022，5，411。[14]. L. C. Garaffa等，《移动机器人探索中的强化学习：综述》，IEEE神经网络与学习系统汇刊，2021，34，3796。[15]. V. Mnih等，《使用深度强化学习玩Atari游戏》，arXiv预印本 arXiv:1312.5602，2013年。[16]. T. P. Lillicrap等，《使用深度强化学习进行连续控制》，arXiv预印本 arXiv:1509.02971，2015年。[17]. N. Frémaux等，《使用基于脉冲神经元的连续时间演员-评论家框架进行强化学习》，PLoS计算生物学，2013，9，e1003024。[18]. A. Charpentier等，《经济学与金融中的强化学习》，计算经济学，2023年，62，425。[19]. N. Shinn等，《Reflexion：采用语言强化学习的语言智能体》，神经信息处理系统进展，2023，36，8634。[20]. E. Lin等，《用于不平衡分类的深度强化学习》，应用智能，2020年，50，2488。[21]. S. Fujimoto等，《解决演员-评论家方法中的函数逼近误差》，国际机器学习会议论文集，2018，1587。[22]. J. Schulman等，《近端策略优化算法》，arXiv预印本 arXiv:1707.06347，2017年。[23]. D. Silver等，《确定性策略梯度算法》，国际机器学习会议论文集，2014，387。[24]. S. Xiang等，《综述：用于光学神经计算的光子器件、架构与算法》，半导体学报，2021，42，023105。[25]. Y. Shen等，《利用相干纳米光子电路进行深度学习》，Nature Photonics，2017，11，441。[26]. G. Wetzstein等，《利用深度光学和光子学进行人工智能推理》，Nature，2020，588，39。[27]. D. A. B. Miller，《用于低能耗信息处理与通信的阿焦耳光电子学》，光波技术杂志，2017，35，346。[28]. B. J. Shastri等，《面向人工智能与神经形态计算的光子学》，Nature Photonics，2021，15，102。[29]. J. Feldmann等，《使用集成光子张量核心进行并行卷积处理》，Nature，2021，589，52。[30]. X. Xu等，《面向光学神经网络的11 TOPS光子卷积加速器》，Nature，2021，589，44。[31]. A. N. Tait等，《利用硅光子权重库构建神经形态光子网络》，Scientific Reports，2017，7，7430。[32]. S. Xu等，《使用集成光子电路进行高阶张量流处理》，Nature Communications，2022，13，7970。

[33]. B. Bai 等，基于微梳的集成光子处理单元，Nature Communications，2023年，14，66。[34]. S. Xu 等，用于复杂深度学习回归的光学相干点积芯片，Light Sci. Appl.，2021年，10，221。[35]. Z. Lin 等，用于推理和原位训练的薄膜铌酸锂 120 GOPS 光子张量核心，Nature Communications，2024年，15，9081。[36]. W. Zhou 等，采用电可编程权重组的存内光子点积引擎，Nature Communications，2023年，14，2887。[37]. M. Wei 等，将相变材料单片集成至代工制造的硅光子学后端工艺中，Nature Communications，2024年，15，2786。[38]. Y. Tian 等，具有低学习能力损失的可扩展紧凑型光子神经芯片，Nanophotonics，2022年，11，329。[39]. Y. Wan 等，面向片上光学处理器的高效随机并行梯度下降训练，Opt.-Electron. Adv.，2025年，7，230182。[40]. Z. Xu 等，大规模光子芯粒 Taichi 实现 160-TOPS/W 通用人工智能，Science，2024年，384，202。[41]. J. Ouyang 等，用于硅芯片上优化问题的 16 通道光子求解器，Chip，2025年，4，100117。[42]. S. Hua 等，具有超低延迟的集成式大规模光子加速器，Nature，2025年，640，361。[43]. S. R. Ahmed 等，通用光子人工智能加速，Nature，2025年，640，368。[44]. J. Wu 等，面向人工智能的模拟光学计算，Engineering，2022年，10，133。[45]. C. Huang 等，光子神经网络的前景与应用，Adv. Phys. X，2022年，7，1981155。[46]. Y. Shi 等，基于带可饱和吸收体的自脉动 DFB 激光器的光子集成脉冲神经元芯片，Photonics Res.，2023年，11，1382。[47]. J. Feldmann 等，具有自学习能力的全光脉冲神经突触网络，Nature，2019年，569，208。[48]. J. Robertson 等，利用光子脉冲 VCSEL 神经元进行图像边缘检测，Opt. Express，2020年，28，37526。[49]. S. Xiang 等，基于全 VCSEL 全光脉冲神经网络的计算原语：监督学习与模式分类，IEEE神经网络与学习系统汇刊，2020年，32，2494。[50]. Y. J. Lee 等，基于受 Izhikevich 启发的模型、采用事件驱动飞焦耳光电神经元的光子脉冲神经网络，Opt. Express，2022年，30，19360。[51]. S. Xiang 等，基于集成式法布里-珀罗激光器与可饱和吸收体的光子脉冲神经元芯片实现软硬件协同计算，Optica，2023年，10，162。[52]. S. Xiang 等，用于卷积脉冲神经网络的光子集成神经突触核心，Opt.-Electron. Adv.，2023年，6，230140。[53]. R. Talukder 等，采用延迟与秩序编码进行训练、利用稀疏性的 40,000 神经元脉冲光子神经网络，Neuromorph. Comput. Eng.，2025年。[54]. S. Gao 等，基于单个

光子脉冲神经元实现超快速障碍物规避的硬件实现，Laser Photonics Rev.，2023年，17，2300424。[55]. Q. Zhang 等，基于单个微环谐振器中多波长辅助与竞争的全光数字逻辑和神经形态计算，Opt.-Electron. Sci.，2025年。[56]. B. Huang 等，基于改进 TD3算法的巡检机器人导航，Sensors，2024年，24，2525。[57]. J. Wu 等，A-TD3：一种适用于连续动作空间的自适应异步双延迟深度确定性算法，IEEE Access，2022年，10，128077。[58]. T. Xu 等，复杂场景下基于改进 TD3算法的端到端自动驾驶决策方法，Sensors，2024年，24，4962。[59]. K. Liu 等，利用 TD3驱动的双边控制提升自动驾驶车辆队列的混合交通稳定性，Sustainability，2025年，17，4790。[60]. X. Gao 等，基于改进 TD3算法的无人作战飞行器空战机动决策最优策略，Drones，2024年，8，501。[61]. T. H. Nguyen 等，面向偏远地区 RSMA 增强型无人机辅助下行通信的 TD3 优化框架，Remote Sens.，2023年，15，5284。[62]. N. Koenig、A. Howard，Gazebo 开源多机器人仿真器的设计与使用范式，Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.，2004年，2149。[63]. K. Takaya 等，使用 ROS 和 Gazebo 测试移动机器人的仿真环境，Proc. Int. Conf. Syst. Theory Control Comput.，2016年，96。[64]. X. Zeng 等，基于 DML 和 DFB-SA 激光芯片的模式分类光子脉冲神经网络，Opt. Express，2025年，33，12045。[65]. M. Yamada，窄条半导体激光器自持脉动现象的理论分析，IEEE J. Quantum Electron.，2002年，29，1330。[66]. S. Xiang 等，用于脉冲强化学习的非线性光子神经形态芯片，arXiv 预印本 arXiv:2508.06962，2025年。[67]. R. Machupalli、M. Hossain、M. Mandal，深度神经网络 ASIC加速器综述，Microprocessors Microsyst.，2022年，89，104441。[68]. G. Tang、N. Kumar、R. Yoo 等，利用群体编码脉冲神经网络进行连续控制的深度强化学习，Conf. Robot Learn.，2021年，2016。