1. **Patterson, D. et al. Carbon emissions and large neural network training. Preprint at https:// arxiv.org/abs/2104.10350 (2021).**

**碳排放与大型神经网络训练**

最近，机器学习 （ML） 的计算需求迅速增长，这带来了许多成本。估算能源成本有助于衡量其对环境的影响并找到更环保的策略，但如果没有详细信息，这是具有挑战性的。我们计算了最近几个大型模型（T5、Meena、GShard、Switch Transformer 和 GPT-3）的能源使用和碳足迹，并完善了发现 Evolved Transformer 的神经架构搜索的早期估计。我们重点介绍了以下提高能源效率和 CO2 当量排放 （CO2e） 的机会： 大型但稀疏激活的 DNN 可以消耗大型密集 DNN 的 <1/10 的能量，而不会牺牲准确性，尽管使用了同样多甚至更多的参数。地理位置对 ML 工作负载调度很重要，因为即使在同一个国家和同一个组织内，无碳能源和由此产生的 CO2e 的比例也会相差 ~5X-10X。我们现在正在优化大型模型的训练地点和时间。特定的数据中心基础设施很重要，因为云数据中心的能源效率可能比典型数据中心高 ~1.4-2 倍，并且其中面向 ML 的加速器可能比现成系统有效 ~2-5 倍。值得注意的是，DNN、数据中心和处理器的选择可以减少高达 ~100-1000 倍的碳足迹。这些重要因素也使得对能源成本的追溯估算变得困难。为了避免计算错误，我们认为需要大量计算资源的ML论文应该在可行的情况下明确能源消耗和CO2e。我们正在努力在未来的研究中提高能源使用和二氧化碳当量的透明度。为了帮助减少 ML 的碳足迹，我们认为能源使用和 CO2e 应该是评估模型的关键指标，我们正在与 MLPerf 开发人员合作，将训练和推理期间的能源使用纳入此行业标准基准。

1. **Reuther, A. et al. Survey of machine learning accelerators. In 2020 IEEE High Performance Extreme Computing Conference (HPEC) 1–12 (IEEE, 2020).**

**机器学习加速器的研究**

多核处理器和加速器的进步为机器学习技术在各种应用中的更大探索和应用打开了闸门。这些进步，以及包括摩尔定律在内的几个趋势的细分，促使处理器和加速器爆炸式增长，这些处理器和加速器有望提供更强大的计算和机器学习能力。这些处理器和加速器有多种形式，从 CPU 和 GPU 到 ASIC、FPGA 和数据流加速器。本文调查了这些处理器和加速器的当前状态，这些处理器和加速器已经公开发布了性能和功耗数据。将性能和功耗值绘制在散点图上，并讨论和分析了该图上趋势中的许多维度和观察结果。例如，图中有一些关于功耗、数值精度以及推理与训练的有趣趋势。然后，我们选择并基准测试了两个商用的低尺寸、重量和功耗 （SWaP） 加速器，因为这些处理器对于最适用于 DoD 和其他 SWaP 受限用户的嵌入式和移动机器学习推理应用最感兴趣。我们确定它们在真实世界图像和神经网络模型中的实际性能，将这些结果与报告的性能和功耗值进行比较，并根据某些嵌入式应用中使用的英特尔 CPU 对其进行评估。

**[3]Xia, Q., & Yang, J. J. Memristive crossbar arrays for brain-inspired computing. Nat. Mater. 18, 309–323 (2019).**

**用于类脑计算的忆阻横杆阵列**

忆阻器件的工作机制基于离子迁移，其开关动力学和电行为类似于突触和神经元的开关动力学和电行为，使这些器件成为类脑计算的候选者。它们内置于大规模横杆阵列中以形成神经网络，通过直接使用物理定律以大规模并行性执行高效的内存计算。人工突触和神经元之间的动态相互作用使网络具有监督和无监督的学习能力。此外，它们能够与来自传感器的模拟信号连接，而无需模拟/数字转换，从而减少了处理时间和能源开销。尽管许多模拟表明这些网络在类脑计算方面的潜力，但大规模忆阻阵列的实验实现仍处于起步阶段。本文着眼于通过忆阻实现实现的高效类脑计算的进展、挑战和可能的解决方案，既可以作为深度学习的加速器，也可以作为脉冲神经网络的构建块。

[4]**Burr, G. W. et al. Neuromorphic computing using non-volatile memory. Adv. Phys. X 2, 89–124 (2017).**

**使用非易失性存储器的神经形态计算**

非易失性存储器 （NVM） 器件的密集交叉阵列代表了实现大规模并行和高能效神经形态计算系统的一种可能途径。我们首先回顾了NVM设备在三种计算范式中的应用的最新进展：脉冲神经网络（SNN）、深度神经网络（DNN）和“记忆计算”。在 SNN 中，NVM 突触连接由本地学习规则更新，例如尖峰-时间-依赖性-可塑性，这是一种直接受生物学启发的计算方法。对于 DNN，NVM 数组可以表示突触权重矩阵，以模拟但大规模并行的方式实现反向传播等算法所需的矩阵-向量乘法。与基于 GPU 的 DNN 训练相比，这种方法可以在功耗和速度方面提供显着的改进，适用于具有商业意义的应用程序。然后，我们调查了最近的研究，其中不同类型的NVM器件（包括相变存储器，导电桥接RAM，丝状和非丝状RRAM以及其他NVM）已被提出作为突触或神经元，用于神经形态计算应用。根据电导动态范围、（非）线性和（a）电导响应的对称性、保持性、耐久性、所需开关功率和器件可变性等特性，评估了这些器件的相关优点和局限性。

# **Khaddam-Aljameh, R. et al. HERMES core—a 14nm CMOS and PCM-based in-memory compute core using an array of 300ps/LSB linearized CCO-based ADCs and local digital processing. In 2021 Symposium on VLSI Circuits (IEEE, 2021).**

# **HERMES Core – 基于 14nm CMOS 和 PCM 的内存计算内核，使用基于 CCO 的 300ps/LSB 线性化 ADC 阵列和本地数字处理**

我们展示了采用14nm CMOS设计和制造的256×256内存计算（IMC）内核，具有后端集成的多级相变存储器（PCM）。它包括 256 个基于线性化电流控制振荡器 （CCO） 的 ADC，间距紧凑，间距为 4μm，以及一个执行仿射缩放和 ReLU 操作的本地数字处理单元。介绍了一种新的CCO频率线性化技术，在1 GHz以上工作时，可实现精确的片上矩阵矢量乘法（MVM）。 当使用两个内核进行深度学习（DL）推理时，给出了MNIST和CIFAR-10数据集上的测量分类精度。测得的能效为10.5 TOPS/W，性能密度为1.59 TOPS/mm2.

# **Narayanan, P. et al. Fully on-chip MAC at 14nm enabled by accurate row-wise programming of PCM-based weights and parallel vector-transport in duration-format. In 2021 Symposium on VLSI Technology (IEEE, 2021).**

# **14 nm 的全片上 MAC，通过对基于 PCM 的权重进行精确的逐行编程和持续时间格式的并行矢量传输来实现**

使用模拟非易失性存储器 （NVM） 进行深度学习的硬件加速需要具有高器件良率、高精度乘法 ACcumulate （MAC） 操作的大型阵列，以及用于实现任意深度神经网络 （DNN） 拓扑的路由框架。在本文中，我们介绍了一款用于模拟 AI 推理的 14 纳米测试芯片，它包含多个相变存储器 （PCM） 器件阵列，每个阵列能够存储 512 个相变存储器×512 个唯一的 DNN 权重，并在数据位置执行大规模并行 MAC 操作。DNN 激励使用并行且可重构的 2-D 网格上的持续时间表示在芯片上传输。为了准确地将推理模型传输到芯片，我们描述了一种闭环调谐 （CLT） 算法，该算法对每个权重中的四个 PCM 电导进行编程，实现 <3% 的平均权重误差。逐行编程方案和相关电路允许我们同时在多达 512 个权重上执行 CLT。我们表明，测试芯片可以在两个不同的 DNN 上实现接近软件等效的精度。我们演示了使用MNIST的全片上两层网络（精度下降~0.6%）的瓦片到瓦片传输，并展示了使用循环长短期记忆（LSTM）网络在长序列（最多10 000个字符）上对错误传播的弹性，实现片外激活和矢量向量运算，以生成用于下一个片上MAC的循环输入。

# [7]Kohda, Y. et al. Unassisted true analog neural network training chip. In 2020 IEEE International Electron Devices Meeting (IEDM) (IEEE, 2020).

# **无辅助真模拟神经网络训练芯片**

使用电阻处理单元 （RPU） 的模拟内存计算已被提议用于神经网络 （NN） 训练。然而，硬件演示仅限于使用一些数字仿真来辅助模拟芯片功能。使用电容器作为模拟权重，我们报道了第一个模拟神经网络训练芯片，其中所有倍数和累加（MAC）功能在模拟交叉点阵列中执行，并且所有权重并行更新。该芯片测量出 92.7% 的全 MNIST 训练准确率，实时运行时间比数字系统快。

# [8]Marković, D., Mizrahi, A., Querlioz, D. & Grollier, J. Physics for neuromorphic computing. Nat. Rev. Phys. 2, 499–510 (2020).

# **神经形态计算的物理学**

生物大脑神经神态与神经网络；将深度神经网络结构引入物理系统

神经形态计算从大脑中汲取灵感，为信息处理创建节能硬件，能够完成高度复杂的任务。使用标准电子设备构建的系统通过模仿大脑的分布式拓扑结构来实现速度和能量的提升。扩大此类系统的规模并使其能源使用、速度和性能提高几个数量级，需要对硬件进行革命。我们讨论了在用于数据处理的算法和纳米级材料中加入更多的物理学如何对神经形态计算领域产生重大影响。我们回顾了利用物理学增强人工神经网络计算能力的惊人成果，这些成果使用电阻开关材料、光子学、自旋电子学和其他技术。我们讨论了可能导致这些方法走向成熟的路径，走向可以实时推断和学习的低功耗、小型化芯片。

**[9]Wetzstein, G. et al. Inference in artificial intelligence with deep optics and photonics. Nature 588, 39–47 (2020).**

# **使用深度光学和光子学的人工智能推理**

光学运算系统应用于视觉计算系统的可能性

众多应用中的人工智能任务需要加速器来实现快速和低功耗的执行。光学计算系统可能能够满足这些特定领域的需求，但尽管进行了半个世纪的研究，通用光学计算系统尚未成熟为实用技术。然而，人工智能推理，特别是对于视觉计算应用，可能会为基于光学和光子系统的推理提供机会。在本期文章中，我们回顾了最近在人工智能应用中的光学计算方面的工作，并讨论了其前景和挑战。

到目前为止[10-22] 的方法还无法应用反向传播算法来原位训练非常规的新型硬件。

1. **Romera, M. et al. Vowel recognition with four coupled spin-torque nano-oscillators. Nature 563, 230–234 (2018).**

# **使用四个耦合自旋扭矩纳米振荡器进行元音识别**

将元音编码到SHG输入光波的方法

近年来，人工神经网络已成为人工智能的旗舰算法[1](https://www.nature.com/articles/s41586-018-0632-y" \l "ref-CR1" \o "Silver, D. et al. Mastering the game of Go without human knowledge. Nature 550, 354–359 (2017).).在这些系统中，神经元激活函数是静态的，计算是通过标准算术运算实现的。相比之下，神经启发计算的一个突出分支包含大脑的动力学性质，并建议赋予神经网络的每个组件动态功能，例如振荡，并依赖于新兴的物理现象，例如同步2,3,4,5,6，用于解决小型网络的复杂问题7,8,9,10,11.这种方法对于硬件实现特别有趣，因为新兴的纳米电子器件可以提供紧凑且节能的非线性自振荡器，模拟生物神经元的周期性脉冲活动12,13,14,15,16.振荡器之间的动态耦合可用于介导人工神经元之间的突触通信。以这种方式使用纳米器件的一个挑战是实现学习，这需要对其耦合振荡进行精细控制和调整[17](https://www.nature.com/articles/s41586-018-0632-y" \l "ref-CR17" \o "Vassilieva, E., Pinto, G., de Barros, J. A. & Suppes, P. Learning pattern recognition through quasi-synchronization of phase oscillators. IEEE Trans. Neural Netw. 22, 84–95 (2011).);纳米器件的动力学特性可能难以控制，并且容易产生噪声和可变性[18](https://www.nature.com/articles/s41586-018-0632-y" \l "ref-CR18" \o "Vodenicarevic, D., Locatelli, N., Araujo, F. A., Grollier, J. & Querlioz, D. A nanotechnology-ready computing scheme based on a weakly coupled oscillator network. Sci. Rep. 7, 44772 (2017).).在这里，我们展示了自旋电子纳米振荡器出色的可调谐性，即通过电流和磁场在很宽的范围内精确控制其频率的可能性，可用于应对这一挑战。我们成功地训练了一个由四个自旋扭矩纳米振荡器组成的硬件网络，通过根据自动实时学习规则调整它们的频率来识别口语元音。我们表明，高实验识别率源于这些振荡器的同步能力。我们的结果表明，通过赋予小型硬件神经网络振荡和同步等非线性动力学特征，可以实现非平凡的模式分类任务。

1. **Shen, Y. et al. Deep learning with coherent nanophotonic circuits. Nat. Photon. 11, 441–446 (2017).**

# **使用相干纳米光子电路进行深度学习**

人工神经网络是受大脑信号处理启发的计算网络模型。这些模型极大地提高了许多机器学习任务的性能，包括语音和图像识别。然而，今天的计算硬件在实现神经网络方面效率低下，这在很大程度上是因为其中大部分是为冯·诺依曼计算方案设计的。在开发电子架构方面已经付出了巨大的努力，这些架构经过调整，以实现具有更高计算速度和准确性的人工神经网络。在这里，我们提出了一种新的全光学神经网络架构，原则上，它可以为传统的推理任务提供比最先进的电子设备更高的计算速度和能效。我们使用可编程纳米光子处理器通过实验演示了该概念的基本部分，该处理器在硅光子集成电路中具有 56 个可编程 Mach-Zehnder 干涉仪的级联阵列，并展示了其在元音识别中的实用性。

[12]**Prezioso, M. et al. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. Nature 521, 61–64 (2015).**

# **基于金属氧化物忆阻器的集成神经形态网络的训练与操作**

尽管半导体集成电路技术取得了所有进步，但人类大脑皮层的极端复杂性使得具有相当数量的设备的神经形态网络的硬件实现极具挑战性。基于混合CMOS/忆阻器电路的CrossNets是提供相当复杂度、运行速度更快且功耗可控的最有前景的候选产品之一。在这些电路中，通常的CMOS堆栈增加了一个或多个交叉层，每个交叉点都有可调节的双端忆阻器。最近，在改进这种忆阻横棒的制造技术及其与CMOS电路的集成方面取得了重大进展，包括首次演示了其垂直集成。另外，已经有几个离散忆阻器作为神经形态网络的人工突触的演示。最近，这种实验被扩展到相变忆阻器件的横杆阵列。然而，这种器件的调整需要在每个交叉点上增加一个晶体管，因此它们的缩放前景不如金属氧化物忆阻器那么令人印象深刻，金属氧化物忆阻器的非线性 I-V 曲线可实现无晶体管操作。在这里，我们报告了无晶体管金属氧化物忆阻器横杆的首次实验实现，其器件可变性降低到足以证明简单集成神经网络（单层感知器）的成功运行。可以使用 delta 规则算法的粗粒度变体在原位教授该网络，以将 3x3 像素的黑/白图像完美分类为 3 类。我们相信，这次演示是朝着实现更大、更复杂的忆阻神经形态网络迈出的重要一步。

# [13]Euler, H.-C. R. et al. A deep-learning approach to realizing functionality in nanoelectronic devices. Nat. Nanotechnol. 15, 992–998 (2020).

# **在纳米电子器件中实现功能的深度学习方法**

使用了用DNN对器件的输入输出特性进行数学建模的方法

许多纳米级器件需要精确优化才能运行。当端子和联轴器的数量增加时，将它们调整到所需的操作状态变得越来越困难和耗时。缺陷和器件间差异阻碍了使用基于物理场的模型的优化。深度神经网络 （DNN） 可以对各种复杂的物理现象进行建模，但到目前为止，它主要用作预测工具。在这里，我们提出了一种通用的深度学习方法，以有效地优化复杂的多端纳米电子器件，以实现所需的功能。我们展示了在硅中掺杂原子的无序网络中实现功能的方法。我们用DNN对器件的输入输出特性进行建模，随后通过梯度下降对DNN模型中的控制参数进行优化，以实现各种分类任务。当相应的控制设置应用于物理设备时，生成的功能与 DNN 模型预测的一样。我们希望我们的方法能够为复杂（量子）纳米电子器件的快速原位优化做出贡献。

1. **Hughes, T. W., Williamson, I. A., Minkov, M. & Fan, S. Wave physics as an analog recurrent neural network. Sci. Adv. 5, eaay6946 (2019).**

# **作为模拟循环神经网络的波物理**

声波信号在逆向设计非均匀介质的传播

模拟机器学习硬件平台有望比数字机器学习硬件平台更快、更节能。在声学和光学中发现的波动物理学是构建时变信号模拟处理器的自然候选者。在这里，我们确定了波物理学动力学与循环神经网络计算之间的映射。这种映射表明，可以使用神经网络的标准训练技术来训练物理波系统来学习时态数据中的复杂特征。作为演示，我们展示了逆向设计的非均匀介质可以对原始音频信号进行元音分类，因为它们的波形在原始音频信号中散射和传播，从而实现了与递归神经网络的标准数字实现相似的性能。这些发现为新型模拟机器学习平台铺平了道路，该平台能够在其原生领域快速有效地处理信息。

1. **Wu, Z., Zhou, M., Khoram, E., Liu, B. & Yu, Z. Neuromorphic metasurface. Photon. Res. 8, 46–50 (2020).**

**神经形态超表面**

超表面已被用于实现光学功能，如聚焦和光束控制。他们使用亚波长纳米结构来控制光的局部振幅和相位。在这里，我们表明这种控制也可以实现人工神经推理的新功能。我们证明了超表面可以通过将来自物体的光聚焦到与物体类别相对应的不同空间位置来直接识别物体。

1. **Furuhata, G., Niiyama, T. & Sunada, S. Physical deep learning based on optimal control of dynamical systems. Phys. Rev. Appl. 15, 034092 (2021)**

**基于动力系统最优控制的物理深度学习**

深度学习是人工智能技术的骨干，可以看作是一种多层前馈神经网络。深度学习的本质是信息层传播。这表明深度神经网络和动态系统之间存在联系，因为信息传播是由动态系统的时间演化明确建模的。在这项研究中，我们基于连续时间动力系统的最优控制进行模式识别，适用于物理硬件实现。该学习基于伴随方法对动态系统进行最优控制，并利用基于系统时间演化的深度（虚拟）网络结构对输入信息进行处理。作为一个关键的例子，我们将基于动力学的识别方法应用于光电延迟系统，并证明延迟系统的使用允许仅使用几个控制信号进行图像识别和非线性分类。这与传统的多层神经网络形成鲜明对比，后者需要训练大量的权重参数。该方法为最优控制问题框架下的深度网络处理机制提供了见解，并提出了实现物理计算硬件的途径。

1. **Lin, X. et al. All-optical machine learning using diffractive deep neural networks. Science 361, 1004–1008 (2018).**

# **使用衍射深度神经网络的全光机器学习**

我们引入了一种全光衍射深度神经网络（D2NN）架构，该架构可以在基于深度学习的无源衍射层设计后学习实现各种功能。我们通过实验证明了该框架的成功，创建了3D打印的D2NN，这些D2NN学会了在太赫兹光谱下实现手写数字分类和成像镜头的功能。借助现有的大量 3D 打印和其他光刻制造方法以及空间光调制器，这种全光学深度学习框架可以以光速执行基于计算机的神经网络可以实现的各种复杂功能，并将在全光学图像分析、特征检测和物体分类中找到应用，还支持新的相机设计和光学组件，可以学习使用 D2NN 执行独特的任务。

1. **Miller, J. F., Harding, S. L. & Tufte, G. Evolution-in-materio: evolving computation in materials. Evol. Intell. 7, 49–67 (2014).**

**材料计算的演化**

物质进化 （EIM） 是通过计算机控制进化 （CCE） 对物理系统的操纵。它的立场是，要从物理系统获得有用的功能，需要应用高度特定的物理信号并将系统置于特定的物理状态。它认为CCE是实现这一目标的有效方法。这样做的一个潜在优势是，人工进化可以潜在地利用那些过于复杂而无法理解或迄今为止未知的物理效应。EIM最常用作在物理系统中实现计算的方法。该方法是模拟计算和经典计算的混合体，因为它使用经典计算机对物理系统或模拟设备进行编程。到目前为止，EIM只在一组相当有限的物理和化学系统中进行了尝试。这篇综述论文研究了过去与EIM相关的工作，并讨论了这些工作背后的历史基础。它描述了最新的发展，分析了这项工作的利弊以及仍然存在的挑战。

1. **Chen, T. et al. Classification with a disordered dopant-atom network in silicon. Nature 577, 341–345 (2020).**

# **硅中无序掺杂原子网络的分类**

分类是生物和人工神经网络都擅长的一项重要任务[1,2](https://www.nature.com/articles/s41586-019-1901-0" \l "ref-CR1" \o "Hubel, D. H. & Wiesel, T. N. Receptive fields of single neurones in the cat’s striate cortex. J. Physiol. 148, 574–591 (1959).).在机器学习中，非线性投影到高维特征空间可以使数据线性分离[3,4](https://www.nature.com/articles/s41586-019-1901-0" \l "ref-CR3" \o "Haykin, S. Neural Networks and Learning Machines (Pearson Prentice Hall, 2008).)，简化了复杂特征的分类。这种非线性投影在传统计算机中的计算成本很高。一种很有前途的方法是利用物理材料系统，因为它们具有很高的计算密度，因此本质上执行这种非线性投影[5](https://www.nature.com/articles/s41586-019-1901-0" \l "ref-CR5" \o "Torrejon, J. et al. Neuromorphic computing with nanoscale spintronic oscillators. Nature 547, 428–431 (2017).)、固有的并行性和能源效率[6,7](https://www.nature.com/articles/s41586-019-1901-0" \l "ref-CR6" \o "Tanaka, G. et al. Recent advances in physical reservoir computing: a review. Neural Netw. 115, 100–123 (2019).).然而，现有的方法要么依赖于系统的时间动力学，这需要顺序数据处理，因此阻碍了并行计算[5,6,8](https://www.nature.com/articles/s41586-019-1901-0" \l "ref-CR6" \o "Tanaka, G. et al. Recent advances in physical reservoir computing: a review. Neural Netw. 115, 100–123 (2019).)，或采用难以扩大规模的大型材料系统[7](https://www.nature.com/articles/s41586-019-1901-0" \l "ref-CR7" \o "Lin, X. et al. All-optical machine learning using diffractive deep neural networks. Science 361, 1004–1008 (2018).).在这里，我们使用一种并行的纳米级方法，其灵感来自大脑中的过滤器[1](https://www.nature.com/articles/s41586-019-1901-0" \l "ref-CR1" \o "Hubel, D. H. & Wiesel, T. N. Receptive fields of single neurones in the cat’s striate cortex. J. Physiol. 148, 574–591 (1959).)和人工神经网络[2](https://www.nature.com/articles/s41586-019-1901-0" \l "ref-CR2" \o "LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436–444 (2015).)执行非线性分类和特征提取。我们利用了跳频传导的非线性[9,10,11](https://www.nature.com/articles/s41586-019-1901-0" \l "ref-CR10" \o "Mott, N. F. Conduction in glasses containing transition metal ions. J. Non Cryst. Solids 1, 1–17 (1968).)通过硅中硼掺杂原子的电可调网络，通过人工演化重新配置网络，实现不同的计算功能。我们首先求解规范的双输入二进制分类问题，实现所有布尔逻辑门[12](https://www.nature.com/articles/s41586-019-1901-0" \l "ref-CR12" \o "Bose, S. K. et al. Evolution of a designless nanoparticle network into reconfigurable Boolean logic. Nat. Nanotechnol. 10, 1048–1052 (2015).)在室温下，使用纳米材料系统进行非线性分类。然后，我们发展我们的掺杂剂网络以实现特征滤波器[2](https://www.nature.com/articles/s41586-019-1901-0" \l "ref-CR2" \o "LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436–444 (2015).)可以对修改后的国家标准与技术研究院手写数字数据库执行四输入二进制分类。与直接应用于原始数据的线性分类器相比，我们基于材料的过滤器的实施大大提高了分类精度[13](https://www.nature.com/articles/s41586-019-1901-0" \l "ref-CR13" \o "Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324 (1998).).我们的研究结果为小尺寸和节能计算建立了硅基电子学的范式[14](https://www.nature.com/articles/s41586-019-1901-0" \l "ref-CR14" \o "Xu, X. et al. Scaling for edge inference of deep neural networks. Nat. Electron. 1, 216–222 (2018).).

1. Bueno, J. et al. Reinforcement learning in a large-scale photonic recurrent neural network. Optica 5, 756–760 (2018).

# **大规模光子递归神经网络中的强化学习**

光子神经网络的实现作为一种潜在的颠覆性未来技术，已经获得了相当大的关注。在大规模神经网络中展示学习对于将光子机器学习基板建立为可行的信息处理系统至关重要。到目前为止，在完全并行和高效的学习硬件中实现具有大量非线性节点的光子神经网络是缺乏的。我们展示了一个由多达2025个衍射耦合光子节点组成的网络，形成了一个大规模的递归神经网络。使用数字微镜设备，我们实现了强化学习。我们的方案是完全并行的，无源权重最大限度地提高了能效和带宽。计算输出有效地收敛，我们实现了非常好的性能。

1. **Tanaka, G. et al. Recent advances in physical reservoir computing: a review. Neural Netw. 115, 100–123 (2019).**

# **物理储层计算研究进展**

物理储层计算：只在输出层进行训练

储层计算是一种适用于时间/顺序数据处理的计算框架。它源自几个循环神经网络模型，包括回声状态网络和液态机。储层计算系统由一个储层组成，用于将输入映射到高维空间，以及一个用于从储层中的高维状态进行模式分析的读数。储液器是固定的，只有读数是用简单的方法（如线性回归和分类）训练的。因此，与其他递归神经网络相比，储层计算的主要优势是快速学习，从而降低训练成本。另一个优点是，没有自适应更新的储液器适合使用各种物理系统、基板和设备进行硬件实现。事实上，这种物理储层计算在各个研究领域都引起了越来越多的关注。本文旨在通过根据储层类型对物理储层计算进行分类，概述物理储层计算的最新进展。本文讨论了物理储层计算的当前问题和观点，以进一步拓展其实际应用，开发下一代机器学习系统。

1. **Appeltant, L. et al. Information processing using a single dynamical node as complex system. Nat. Commun. 2, 468 (2011).**

# **使用单个动态节点作为复杂系统的信息处理**

在信息驱动的社会中，人们非常需要新的信息处理方法。受大脑处理信息能力的启发，最近引入的被称为“储层计算”的范式表明，复杂的网络可以有效地执行计算。在这里，我们介绍了一种新颖的架构，该架构将通常需要的大量元素减少到具有延迟反馈的单个非线性节点。通过电子实现，我们通过实验和数字展示了语音识别基准测试中的出色性能。互补的数值研究也表明，在时间序列预测基准方面具有出色的性能。这些结果证明，即使在最简单的表现形式中，延迟动力学系统也可以进行有效的信息处理。这一发现为储层计算的可行和资源节约型技术实施铺平了道路。

# [23]Mouret, J.-B. & Chatzilygeroudis, K. 20 years of reality gap: a few thoughts about simulators in evolutionary robotics. In Proc. Genetic and Evolutionary Computation Conference Companion 1121–1124 (2017).

# **20 年的现实差距：关于进化机器人模拟器的一些想法**

进化机器人（ER）中的模拟器通常被认为是“暂时的邪恶”，直到可以在真正的机器人上进行实验。然而，经过 20 多年的 ER 研究，大多数实验仍然在模拟中进行，没有任何迹象表明这种情况在未来几年会发生变化。在这篇简短的论文中，我们描述了模拟器对ER的要求，我们尝试了什么，以及我们如何在许多实验中成功跨越“现实鸿沟”。我们认为，未来的模拟器在预测适应度值时需要能够估计其置信度，以便避免未准确模拟的行为。

1. Howison, T., Hauser, S., Hughes, J. & Iida, F. Reality-assisted evolution of soft robots through large-scale physical experimentation: a review. Artif. Life 26, 484–506 (2021).

# **基于大规模物理实验的软体机器人现实辅助演化研究进展**

在这篇综述中，我们介绍了现实辅助进化的框架，总结了将基于模型和无模型的方法相结合以改进物理实体软机器人设计的日益增长的趋势。在计算机模拟中，数据驱动的模型使用真实世界的实验数据构建、调整和改进目标系统的表示。通过使用这些数据驱动的模型模拟大量虚拟机器人，优化算法可以照亮多个设计候选者，以便转移到现实世界。实际上，大规模的物理实验有助于多个候选设计的制造、测试和分析。自动化装配和可重新配置的模块化系统使实际设计评估的数量比以前多得多。通过物理实验收集的大量地面实况数据可以返回到虚拟环境中，以改进数据驱动的模型并指导优化。将设计过程建立在物理实验的基础上，可确保虚拟机器人设计的复杂性不会超过模型限制或可用的制造技术。我们概述了在现实辅助进化框架下物理实体软机器人设计的关键发展。

# de Avila Belbute-Peres, F., Smith, K., Allen, K., Tenenbaum, J. & Kolter, J. Z. End-to-end differentiable physics for learning and control. Adv. Neural Inf. Process. Syst. 31, 7178–7189 (2018).

# **结合可微分方程求解器和图神经网络进行流体流动预测**

求解大型复偏微分方程 （PDE），例如计算流体动力学 （CFD） 中出现的偏微分方程，是一个计算成本高昂的过程。这促使人们使用深度学习方法来近似偏微分方程解，但这些方法预测的仿真结果通常不能很好地推广到真正新颖的场景。在这项工作中，我们开发了一种混合（图）神经网络，它将传统的图卷积网络与网络本身内部的嵌入式可微流体动力学模拟器相结合。通过将实际的 CFD 模拟器（在问题的更粗略的分辨率表示上运行）与图网络相结合，我们表明我们既可以很好地推广到新情况，又可以从神经网络 CFD 预测的大幅加速中受益，同时也大大优于单独的粗略 CFD 模拟。

# [26]Degrave, J., Hermans, M., Dambre, J. & Wyffels, F. A differentiable physics engine for deep learning in robotics. Front. Neurorobot. 13, 6 (2019).

# **用于机器人深度学习的可微分物理引擎**

机器人技术的一个重要领域是控制器的优化。目前，机器人在这个优化过程中往往被当作一个黑匣子，这也是进化算法、强化学习等无导数优化方法无处不在的原因。当使用基于梯度的方法时，模型保持较小或依赖于雅可比的有限差分近似。随着参数数量的增加，这种方法会迅速变得昂贵，例如在深度学习中发现的参数。我们建议实现一个现代物理引擎，它可以区分控制参数。此引擎针对 CPU 和 GPU 均实现。首先，本文展示了这种引擎如何加快优化过程，即使是对于小问题也是如此。此外，它还解释了为什么这是深度Q学习的替代方法，用于在机器人技术中使用深度学习。最后，我们认为这是机器人深度学习的一大步，因为它为优化机器人开辟了新的可能性，包括硬件和软件。

1. Molesky, S. et al. Inverse design in nanophotonics. Nat. Photon. 12, 659–670 (2018).

# **纳米光子学中的逆向设计**

计算逆向设计方法（基于所需功能特征发现光学结构的算法技术）的最新进展已经开始重塑纳米光子学可用的结构景观。在这里，我们概述了光子优化这一新兴领域的关键发展的横截面：从基础结果的回顾到非线性、拓扑、近场和片上光学应用的动机。

1. **Peurifoy, J. et al. Nanophotonic particle simulation and inverse design using artificial neural networks. Sci. Adv. 4, eaar4206 (2018)**

# **使用人工神经网络的纳米光子粒子模拟和逆向设计**

我们提出了一种使用人工神经网络来近似多层纳米粒子光散射的方法。我们发现，网络只需要对数据进行少量采样的训练，就可以将模拟逼近到高精度。一旦神经网络经过训练，它就可以比传统模拟快几个数量级来模拟这种光学过程。此外，经过训练的神经网络可用于通过使用反向传播来解决纳米光子逆设计问题，其中梯度是解析的，而不是数值的

1. **Stern, M., Arinze, C., Perez, L., Palmer, S. E. & Murugan, A. Supervised learning through physical changes in a mechanical system. Proc. Natl Acad. Sci. USA 117, 14843–14850 (2020)**

# **通过机械系统的物理变化进行监督学习**

监督学习：使用已知正确输出的训练集和测试集的学习

机械超材料通常设计为显示对规定力的所需响应。在某些应用中，很难准确指定所需的力-响应关系，但很容易获得力和所需响应的示例。在这里，我们提出了一个在薄而折痕的薄片中进行监督学习的框架，该框架通过实际体验训练示例来学习所需的力-响应行为，然后至关重要的是，对以前看不见的测试力做出正确的反应（概括）。在训练过程中，我们使用训练力折叠床单，促使局部折痕刚度与他们所经历的应变成比例变化。我们发现，这种学习过程重塑了折叠板材中固有的非线性，以便显示以前看不见的测试力的正确响应。我们展示了学习表中的训练误差、测试误差和表大小（模型复杂度）之间的关系，并将它们与机器学习算法中的对应项进行了比较。我们的框架展示了如何通过局部物理学习过程雕刻无序机械材料的崎岖能量景观，以显示所需的力-响应行为。

1. **Zhou, F. & Chai, Y. Near-sensor and in-sensor computing. Nat. Electron. 3, 664–671 (2020).**

# **近传感器和传感器内计算**

通常用于感官网络的节点数量正在迅速增长，导致大量冗余数据在感官终端和计算单元之间交换。为了有效地处理如此大量的数据并降低功耗，有必要开发在感官网络附近或内部运行的计算方法，并可以减少传感和处理单元之间的冗余数据移动。在这里，我们研究了近传感器和传感器内计算的概念，其中计算任务部分转移到感觉终端。我们将函数分为低级和高级处理，并讨论了不同物理传感系统近传感器和传感器内计算的实现。我们还分析了该领域的现有挑战，并为使用先进制造技术的集成传感和处理单元的硬件实现提供了可能的解决方案。

1. **Martel, J. N., Mueller, L. K., Carey, S. J., Dudek, P. & Wetzstein, G. Neural sensors: learning pixel exposures for HDR imaging and video compressive sensing with programmable sensors. IEEE Trans. Pattern Anal. Mach. Intell. 42, 1642–1653 (2020).**

**神经传感器：使用可编程传感器学习用于 HDR 成像和视频压缩传感的像素曝光**

相机传感器依靠全局或卷帘快门功能来曝光图像。这种固定功能方法严重限制了传感器捕捉高动态范围 （HDR） 场景和解析高速动态的能力。空间变化像素曝光已被引入作为一种强大的计算摄影方法，用于对传感器上的辐照度进行光学编码，并通过计算恢复场景的附加信息，但现有方法依赖于启发式编码方案和笨重的空间光调制器来光学实现这些曝光功能。在这里，我们介绍了神经传感器作为一种方法，以端到端的方式与可微分图像处理方法（如神经网络）一起优化每像素快门功能。此外，我们还演示了如何利用新兴的可编程和可重构传感器处理器直接在传感器上实现优化的曝光功能。我们的系统考虑了传感器的特定局限性，以优化物理上可行的光学代码，并在模拟和真实场景的实验中评估其快照 HDR 和高速压缩成像的性能。

1. **Mennel, L. et al. Ultrafast machine vision with 2D material neural network image sensors. Nature 579, 62–66 (2020).**

# **采用 2D 材料神经网络图像传感器的超快机器视觉**

近年来，机器视觉技术取得了巨大的飞跃，现在正在成为各种智能系统不可或缺的一部分，包括自动驾驶汽车和机器人技术。通常，视觉信息由基于帧的相机捕获，转换为数字格式，然后使用机器学习算法（如人工神经网络 （ANN））进行处理[1](https://www.nature.com/articles/s41586-020-2038-x" \l "ref-CR1" \o "LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436–444 (2015).).然而，通过整个信号链传递的大量（大部分是冗余的）数据会导致低帧速率和高功耗。因此，已经开发了各种视觉数据预处理技术2,3,4,5,6,7提高 ANN 中后续信号处理的效率。在这里，我们证明了图像传感器本身可以构成一个ANN，它可以同时感知和处理光学图像而不会延迟。我们的器件基于可重构的二维（2D）半导体[8,9](https://www.nature.com/articles/s41586-020-2038-x" \l "ref-CR8" \o "Manzeli, S., Ovchinnikov, D., Pasquier, D., Yazyev, O. V. & Kis, A. 2D transition metal dichalcogenides. Nat. Rev. Mater. 2, 17033 (2017).)光电二极管10,11,12阵列，网络的突触权重存储在连续可调的光响应度矩阵中。我们演示了监督和无监督学习，并训练传感器对图像进行分类和编码，这些图像以每秒 2000 万个箱的吞吐量光学投射到芯片上。

**[33]Brooks, R. A. Intelligence without reason. In Proc. 12th International Joint Conference on Artificial Intelligence Vol. 1, 569–595 (Morgan Kaufmann, 1991).**

计算机和思想是共同将人工智能定义为一门学科的两个类别。人们普遍认为，过去三十年来，人工智能方面的工作对计算机体系结构的各个方面产生了重大影响。在本文中，我们也提出了相反的主张;计算机体系结构的状态对我们的思维模式产生了重大影响。冯·诺依曼（Von Neumann）的计算模型在特定方向上引领了人工智能。生物系统中的智能是完全不同的。最近在基于行为的人工智能方面的工作产生了新的智能模型，这些模型在精神上更接近生物系统。他们使用的非冯·诺依曼计算模型与生物计算有许多共同的特征。

**[34]Hooker, S. The hardware lottery. Preprint at https://arxiv.org/abs/2009.06489 (2020).**

硬件、系统和算法研究社区历来有不同的激励结构和波动的动机来明确地相互接触。这种历史处理方式很奇怪，因为硬件和软件经常决定哪些研究想法成功（和失败）。本文引入了“硬件彩票”一词来描述一个研究想法何时获胜，因为它适合可用的软件和硬件，而不是因为这个想法优于其他研究方向。早期计算机科学史上的例子说明了硬件彩票如何通过将成功的想法视为失败来延迟研究进度。鉴于领域专用硬件的出现，这些经验教训尤为突出，这使得偏离研究思想的人迹罕至的道路的成本越来越高。本文认为，计算进步带来的收益可能会变得更加不平衡，某些研究方向进入快车道，而其他研究方向的进展则进一步受阻。

1. **Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25, 1097–1105 (2012).**

**使用深度卷积神经网络进行 ImageNet 分类**

我们训练了一个大型深度卷积神经网络，将 ImageNet LSVRC-2010 竞赛中的 120 万张高分辨率图像分类为 1000 个不同的类别。在测试数据上，我们分别实现了 37.5% 和 17.0% 的前 1 名和前 5 名错误率，这比以前最先进的要好得多。神经网络有 6000 万个参数和 650,000 个神经元，由五个卷积层组成，其中一些卷积层后面是最大池化层，以及三个完全连接的层，最终是 1000 路 softmax。为了加快训练速度，我们使用了非饱和神经元和非常高效的 GPU 卷积运算实现。为了减少全连接层中的过拟合，我们采用了一种最近开发的称为“dropout”的正则化方法，该方法被证明非常有效。我们还在 ILSVRC-2012 竞赛中输入了该模型的变体，并取得了 15.3% 的前 5 名测试错误率，而第二名的参赛作品则为 26.2%。

# Lin, H. W., Tegmark, M. & Rolnick, D. Why does deep and cheap learning work so well? J. Stat. Phys. 168, 1223–1247 (2017).

# **为什么深度和廉价的学习效果如此之好？**

我们展示了深度学习的成功不仅取决于数学，还取决于物理学：尽管众所周知的数学定理保证了神经网络可以很好地逼近任意函数，但具有实际意义的函数类别通常可以通过“廉价学习”来近似，其参数比通用函数少得多。我们探讨了物理学中经常遇到的属性，如对称性、局域性、组合性和多项式对数概率，如何转化为非常简单的神经网络。我们进一步认为，当生成数据的统计过程具有物理学和机器学习中普遍存在的某种分层形式时，深度神经网络可能比浅层神经网络更有效。我们使用信息论对这些主张进行形式化，并讨论了与重整化群的关系。我们证明了各种“无扁平定理”，表明当有效的线性深度网络不能被不损失效率的浅层网络精确近似时，例如，我们证明了 *[数学处理误差]* 变量不能在单个隐藏层中使用少于 2^n 个神经元相乘。

1. **Grollier, J. et al. Neuromorphic spintronics. Nat. Electron. 3, 360–370 (2020).**

# **神经形态自旋电子学**

神经形态计算使用受大脑启发的原理来设计电路，这些电路可以执行计算任务，其能效优于传统计算机。然而，使用传统电子设备创建人工神经元和突触的方法目前受到这些组件的能量和面积要求的限制。利用电子的磁性和电学特性的自旋电子纳米器件可以提高能源效率并减少这些电路的面积，磁隧道结作为神经形态计算元件特别令人感兴趣，因为它们与标准集成电路兼容并且可以支持多种功能。在这里，我们回顾了用于神经形态计算的自旋电子学器件的发展。我们研究了磁隧道连接如何充当突触和神经元，以及磁性纹理（如域壁和skyrmions）如何充当神经元。我们还探索了基于自旋电子学的神经形态计算任务实现，例如联想记忆中的模式识别，并讨论了在扩展这些系统时存在的挑战。

1. **Mitarai, K., Negoro, M., Kitagawa, M. & Fujii, K. Quantum circuit learning. Phys. Rev. A 98, 032309 (2018).**

### 量子电路学习

我们提出了一种用于近期量子处理器机器学习的经典量子混合算法，我们称之为量子电路学习。由我们的框架驱动的量子电路通过调整在其上实现的参数来学习给定的任务。参数的迭代优化使我们能够绕过高深度电路。理论研究表明，量子电路可以近似非线性函数，数值模拟进一步证实了这一点。该框架将低深度量子电路和用于机器学习的经典计算机混合在一起，为近期量子器件在量子机器学习中的应用铺平了道路。

1. **Poggio, T., Banburski, A. & Liao, Q. Theoretical issues in deep networks. Proc. Natl Acad. Sci. USA 117, 30039–30045 (2020).**

# **深度网络中的理论问题：逼近、优化和泛化**

梯度下降与参数优化

虽然深度学习在许多应用中都很成功，但在理论上还没有得到很好的理解。然而，深度学习的一个令人满意的理论特征开始出现。它涵盖了以下问题：1）深度网络的表示能力 2）经验风险的优化 3）梯度下降技术的泛化特性 ---为什么当网络过度参数化时，尽管没有显式正则化，但预期误差不会受到影响？在这篇综述中，我们讨论了这三个领域的最新进展。在近似理论中，浅层网络和深层网络都被证明是以指数级参数数（函数维数为指数）为代价来逼近有界域上的任何连续函数。然而，对于组合函数的子集，卷积类型的深度网络可以对维度具有线性依赖性，这与浅层网络不同。在优化中，我们讨论了指数损失函数的损失情况，并表明随机梯度下降将高概率找到全局最小值。为了解决分类任务的泛化问题，我们使用经典的均匀收敛结果来证明在每一层权重矩阵的单位范数约束下最小化代理指数型损失函数的合理性——因为分类的有趣变量是权重方向而不是权重。我们的方法得到了几个独立的新结果的支持，为深度过度参数化ReLU网络的泛化性能的难题提供了解决方案，揭示了潜在的隐藏复杂性控制的起源。

1. **Scellier, B. & Bengio, Y. Equilibrium propagation: bridging the gap between energy-based models and backpropagation. Front. Comput. Neurosci. 11 (2017).**

**平衡传播：弥合基于能量的模型与反向传播之间的差距**

弥合当前反向传播原理与真实生物过程的差别，存在自由相和扰动相，达到平衡状态意味着能量函数的最小化

我们介绍了平衡传播，这是一个基于能量的模型的学习框架。它只涉及一种神经计算，在第一阶段（预测时）和训练的第二阶段（在目标或预测误差显示后）进行。尽管该算法像反向传播一样计算目标函数的梯度，但它不需要为第二阶段进行特殊的计算或电路，其中误差是隐式传播的。均衡传播与对比赫比学习和对比散度有相似之处，同时解决了两种算法的理论问题：我们的算法计算定义明确的目标函数的梯度。由于目标函数是根据局部扰动定义的，因此平衡传播的第二阶段仅对应于将预测（不动点或稳态分布）推向减少预测误差的配置。在循环多层监督网络的情况下，输出单元在第二阶段略微向目标靠拢，输出层引入的扰动在隐藏层中向后传播。我们表明，在第二阶段，“反向传播”的信号对应于误差导数的传播，并编码目标函数的梯度，当突触更新对应于尖峰时间依赖性可塑性的标准形式时。这项工作使大脑可以实现类似于反向传播的机制变得更加合理，因为泄漏积分器神经计算在我们的模型中同时执行推理和错误反向传播。这两个阶段之间唯一的局部区别是是否允许突触变化。我们还通过实验表明，具有 1、2 和 3 个隐藏层的多层循环连接网络可以通过平衡传播在排列不变的 MNIST 任务上进行训练。

1. **Ernoult, M., Grollier, J., Querlioz, D., Bengio, Y. & Scellier, B. Equilibrium propagation with continual weight updates Preprint at https://arxiv.org/abs/2005.04168 (2020).**

# **平衡传播与持续权重更新**

平衡传播 （EP） 是一种将机器学习和神经科学联系起来的学习算法，它通过计算与时间反向传播 （BPTT） 的梯度非常匹配的梯度，但学习规则在空间中局部。给定输入x和相关目标y，EP分两个阶段进行：在第一阶段，神经元自由地向第一稳态进化;在第二阶段，输出神经元被推向y直到它们达到第二个稳定状态。然而，在现有的 EP 实现中，学习规则在时间上不是本地的：权重更新是在第二阶段的动力学收敛后执行的，并且需要第一阶段的信息，而这些信息在物理上不再可用。在这项工作中，我们提出了一种名为持续平衡传播（C-EP）的EP版本，其中神经元和突触动力学在整个第二阶段同时发生，因此权重更新及时变得局部。这种在空间和时间上都局部的学习规则为极节能的 EP 硬件实现提供了可能性。我们从理论上证明，如果学习率足够小，在第二阶段的每个时间步长，神经元和突触的动力学都遵循BPTT给出的损失梯度（定理1）。我们演示了在 MNIST 上使用 C-EP 进行训练，并将 C-EP 推广到神经元通过不对称连接连接的神经网络。我们通过实验表明，网络更新越遵循BPTT的梯度，它在训练方面的表现就越好。这些结果使EP更接近生物学，更好地遵守硬件限制，同时保持其与反向传播的密切联系。

1. **Laborieux, A. et al. Scaling equilibrium propagation to deep convnets by drastically reducing its gradient estimator bias. Front. Neurosci. 15 (2021).**

# **通过大幅减少梯度估计器偏差，将均衡传播扩展到深度卷积网络**

平衡传播 （EP） 是受生物启发的 Backpropagation Through Time （BPTT）的对应物，由于其强大的理论保证和其学习规则在空间上的局部性，促进了专用于学习的节能硬件的设计。然而，在实践中，EP并不能比MNIST更难扩展到视觉任务。在这项工作中，我们表明，使用有限微移所固有的 EP 梯度估计中的偏差是造成这种现象的原因，并且取消它允许通过 EP 训练深度 ConvNet，包括具有不同前向和后向连接的架构。这些结果突出了EP是一种可扩展的方法，用于计算深度神经网络中的误差梯度，从而激励其硬件实现。

1. **Martin, E. et al. Eqspike: spike-driven equilibrium propagation for neuromorphic implementations. iScience 24, 102222 (2021).**

# **EqSpike：用于神经形态实现的尖峰驱动平衡传播**

寻找可以在[神经形态系统的](https://www.sciencedirect.com/topics/engineering/neuromorphic-system" \o "Learn more about neuromorphic systems from ScienceDirect's AI-generated Topic Pages)局部约束下实现的基于尖峰的学习算法，同时实现高精度，仍然是一个艰巨的挑战。平衡传播是[反向传播](https://www.sciencedirect.com/topics/engineering/backpropagation" \o "Learn more about backpropagation from ScienceDirect's AI-generated Topic Pages)的一种很有前途的替代方案，因为它只涉及局部计算，但到目前为止，面向硬件的研究主要集中在基于速率的网络上。在这项工作中，我们开发了一种名为EqSpike的脉冲[神经网络算法](https://www.sciencedirect.com/topics/engineering/artificial-neural-network" \o "Learn more about neural network algorithm from ScienceDirect's AI-generated Topic Pages)，该算法与神经形态系统兼容，通过平衡传播进行学习。通过仿真，我们在MNIST手写数字数据集（美国国家标准与技术研究院混合）上获得了97.6%的测试识别准确率，类似于基于速率的均衡传播，并且与用于脉冲神经网络的替代学习技术相比具有优势。我们展示了 EqSpike 在 [硅](https://www.sciencedirect.com/topics/materials-science/silicon" \o "Learn more about silicon from ScienceDirect's AI-generated Topic Pages)与[图形处理单元](https://www.sciencedirect.com/topics/engineering/graphics-processing-unit" \o "Learn more about graphics processing units from ScienceDirect's AI-generated Topic Pages)相比，神经形态技术可以分别将推理和训练的能耗降低三个数量级和两个数量级。最后，我们还表明，在学习过程中，EqSpike 权重更新表现出一种与尖峰时间相关的可塑性，突出了与生物学的可能联系。

1. **Dillavou, S., Stern, M., Liu, A. J., & Durian, D. J. Demonstration of decentralized, physics-driven learning. Preprint at https://arxiv.org/abs/2108.00275 (2021).**

# **去中心化、物理驱动学习的演示**

在典型的人工神经网络中，神经元根据中央处理器的全局计算进行调整，但在大脑中，神经元和突触会根据局部信息进行自我调整。最近有人提出对比学习算法来训练物理系统，如流体、机械或电气网络，以根据局部演化规则执行机器学习任务。然而，迄今为止，由于创建基于自身对两组全球边界条件的响应自主演化的元素的工程挑战，此类系统仅在计算机模拟中实施。在这里，我们介绍并实现了一个物理驱动的可变电阻网络对比学习方案，使用电路局部比较两个相同网络在两组不同边界条件下的响应。利用这一创新，我们的系统可以有效地进行自我训练，在不使用中央处理器或外部信息存储的情况下优化其电阻值。一旦系统针对指定的变构、回归或分类任务进行了训练，该任务随后就会通过物理命令快速自动地执行，以最大限度地减少响应给定电压输入的功耗。我们证明，与典型的计算机不同，这种学习系统由于其分散式学习而对极端损坏（以及制造缺陷）具有鲁棒性。因此，我们的双网络方法很容易扩展到超大型或非线性网络，其中其分布式特性将是一个巨大的优势;只有 500 个边缘的实验室网络已经超过了其计算机模拟网络。

1. **Hermans, M., Burm, M., Van Vaerenbergh, T., Dambre, J. & Bienstman, P. Trainable hardware for dynamical computing using error backpropagation through physical media. Nat. Commun. 6, 6729 (2015).**

# **用于动态计算的可训练硬件，使用物理介质的误差反向传播**

神经网络目前在数字冯·诺依曼机器上实现，这些机器没有充分利用其固有的并行性。我们演示了如何使用一类新型可重构动态系统进行模拟信息处理，从而缓解了这个问题。我们用于动态模拟计算的通用硬件平台由具有非线性反馈的倒数线性动力系统组成。由于互易性（光和声音的传播等许多物理现象的普遍属性），误差反向传播（将此类系统调整为特定任务的关键步骤）可能发生在硬件中。这可能会显著加快优化过程，为受神经启发的硬件的可扩展性提供重要优势。在本文中，我们使用一个经过实验验证的示例和一个概念示例表明，此类系统可以为构建高度可扩展、完全动态的模拟计算机提供一种简单的机制。

1. **Hughes, T. W., Minkov, M., Shi, Y. & Fan, S. Training of photonic neural networks through in situ backpropagation and gradient measurement. Optica 5, 864–871 (2018).**

# **基于原位反向传播的光子神经网络训练**

我们提供了一种基于伴随方法的光子神经网络训练协议。网络相对于其可调自由度的梯度是通过物理反向传播光学误差信号来计算的。

1. **Lopez-Pastor, V. & Marquardt, F. Self-learning machines based on Hamiltonian echo backpropagation. Preprint at https://arxiv.org/abs/2103.04992 (2021).**

# **基于哈密顿回波反向传播的自学习机**

物理自学习机可以定义为一个非线性动态系统，可以在数据上进行训练（类似于人工神经网络），但作为可学习参数的内部自由度的更新是自主发生的。这样，既不需要外部处理和反馈，也不需要了解（和控制）这些内部自由度。我们介绍了在任何时间可逆哈密顿系统中自学习的一般方案。我们用数值方式说明了这种自学习机在耦合非线性波场情况下的训练。

1. **Hubara, I., Courbariaux, M., Soudry, D., El-Yaniv, R. & Bengio, Y. Quantized neural networks: training neural networks with low precision weights and activations. J. Mach. Learn. Res. 18, 6869–6898 (2017).**

# **量化神经网络：使用低精度权重和激活训练神经网络**

我们介绍了一种在运行时训练量化神经网络 （QNN） 的方法，---具有极低精度（例如 1 位）权重和激活的神经网络。在训练时，量化的权重和激活用于计算参数梯度。在前向传递期间，QNN 大大减少了内存大小和访问，并用按位运算取代了大多数算术运算。因此，功耗有望大幅降低。我们在 MNIST、CIFAR-10、SVHN 和 ImageNet 数据集上训练了 QNN。由此产生的 QNN 实现了与 32 位对应项相当的预测精度。例如，我们具有 1 位权重和 2 位激活的 AlexNet 量化版本实现了51%top-1 精度。此外，我们还将参数梯度量化为 6 位，这使得仅使用逐位运算即可计算梯度。在 Penn Treebank 数据集上测试了量化的递归神经网络，并实现了与仅使用 4 位的 32 位对应网络相当的精度。最后但并非最不重要的一点是，我们编写了一个二进制矩阵乘法 GPU 内核，使用它可以使我们的 MNIST QNN 运行速度比使用未优化的 GPU 内核快 7 倍，而不会损失任何分类精度。QNN 代码可在线获取。

1. **Frye, R. C., Rietman, E. A. & Wong, C. C. Back-propagation learning and nonidealities in analog neural network hardware. IEEE Trans. Neural Netw. 2, 110–117 (1991).**

# **模拟神经网络硬件中的反向传播学习和非理想性**

介绍了使用光控神经网络进行自适应学习的实验结果。作者使用非线性系统识别和信号预测这两个潜在神经网络应用领域中的示例问题来研究模拟神经硬件的能力。这些实验研究了模拟硬件系统典型的各种非理想性的影响。他们表明，使用大量非均匀组件阵列的网络可以执行模拟通信，其精度比考虑到网络元素变化程度的预期要高得多。还研究了其他常见的非理想性的影响，如噪声、权重量化和动态范围限制

1. **Cramer, B. et al. Surrogate gradients for analog neuromorphic computing. Preprint at https://arxiv.org/abs/2006.07239 (2020).**

# **模拟神经形态计算的代理梯度**

为了以较低的代谢成本快速处理时间信息，生物神经元将输入整合为模拟总和，但与峰值、二进制事件进行及时通信。模拟神经形态硬件使用相同的原理来模拟具有出色能效的脉冲神经网络。然而，由于设备不匹配和缺乏有效的训练算法，在此类硬件上实例化高性能尖峰网络仍然是一个重大挑战。在这里，我们介绍了一个基于代理梯度的通用在环学习框架，该框架可以解决这些问题。使用 BrainScaleS-2 神经形态系统，我们表明学习可以自我纠正设备不匹配，从而在视觉和语音基准测试中实现具有竞争力的尖峰网络性能。我们的网络显示稀疏的尖峰活动，平均而言，每个隐藏的神经元和输入远少于一个尖峰，以高达 85 k 帧/秒的速率执行推理，功耗低于 200 mW。总之，我们的工作为模拟神经形态硬件上的低能耗脉冲网络处理设定了几个新的基准，并为未来的片上学习算法铺平了道路。

# **[51]**Adhikari, S. P. et al. Memristor bridge synapse-based neural network and its learning. *IEEE Trans Neural Netw. Learn. Syst.* **23**,1426–1435 (2012).

# **忆阻器桥接突触神经网络及其学习**

提出了一种基于忆阻器桥突触的多层神经网络的模拟硬件架构及其学习方案。在所提出的架构中使用忆阻器桥突触解决了模拟神经网络实现中关于非易失性权重存储的主要问题之一。为了补偿忆阻器桥突触的空间不均匀性和非理想响应，该文还提出了一种适用于所提出的神经网络架构的改进芯片在环学习方案。在所提出的方法中，初始学习在软件中进行，硬件网络通过独立学习网络的每个单层神经元来学习软件训练网络的行为。单层神经元学习的前向计算是在电路硬件上实现的，然后是由主机辅助的权重更新阶段。与传统的芯片在环学习不同，通过忆阻器桥突触和提出的学习方案，消除了为计算每个时期的权重更新而读出突触权重的需要。还介绍了硬件架构以及在三位奇偶校验网络和汽车检测网络上成功实现所提出的学习。

**[52]Lillicrap, T. P., Cownden, D., Tweed, D. B. & Akerman, C. J. Random synaptic feedback weights support error backpropagation for deep learning. Nat. Commun. 7, 13276 (2016).**

# **随机突触反馈权重支持深度学习的误差反向传播**

大脑通过多层神经元处理信息。这种深层架构在表征上很强大，但使学习变得复杂，因为当犯错时很难识别负责的神经元。在机器学习中，反向传播算法通过将错误信号与每个神经元轴突上的所有突触权重相乘来分配责任。然而，这涉及到一种精确、对称的向后连接模式，这在大脑中被认为是不可能的。在这里，我们证明了这种强大的架构约束对于有效的错误传播并不需要。我们提出了一个令人惊讶的简单机制，通过将错误乘以随机突触权重来分配责任。这种机制可以在多层神经元之间传输教学信号，并且在各种任务中与反向传播一样有效。我们的研究结果有助于重新提出关于大脑如何使用错误信号的问题，并消除长期以来关于算法对学习的限制的假设。

1. **Launay, J., Poli, I., Boniface, F., & Krzakala, F. Direct feedback alignment scales to modern deep learning tasks and architectures. Preprint at https://arxiv.org/abs/2006.12878 (2020)**

# **直接反馈对齐扩展到现代深度学习任务和架构**

尽管是深度学习的主力军，但反向传播算法并不是万能的。它强制执行顺序层更新，从而阻止训练过程的高效并行化。此外，其生物学合理性也受到挑战。已经设计了替代方案;然而，在突触不对称的约束下，没有一个可以扩展到现代深度学习任务和架构。在这里，我们挑战了这一观点，并研究了直接反馈对齐对神经视图合成、推荐系统、几何学习和自然语言处理的适用性。与之前仅限于计算机视觉任务的研究相比，我们的研究结果表明，它成功地训练了大量最先进的深度学习架构，其性能接近微调的反向传播。与普遍的看法不同，我们的工作支持在没有重量运输的情况下可以解决具有挑战性的任务。

# [54]Paszke, A. et al. PyTorch: an imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32, 8024–8035 (2019).

# **PyTorch：命令式风格的高性能深度学习库**

GPU深度学习；批处理；反向自动微分

深度学习框架通常关注可用性或速度，但不是两者兼而有之。PyTorch 是一个机器学习库，它表明这两个目标实际上是兼容的：它提供了一种命令式和 Python 编程风格，支持代码作为模型，使调试变得容易，并与其他流行的科学计算库保持一致，同时保持高效并支持硬件加速器，如 GPU。  
在本文中，我们详细介绍了驱动 PyTorch 实现的原则，以及这些原则如何反映在其架构中。我们强调 PyTorch 的每个方面都是一个常规的 Python 程序，由其用户完全控制。我们还解释了其运行时关键组件的谨慎和务实的实现如何使它们能够协同工作以实现令人信服的性能。  
我们演示了各个子系统的效率，以及 PyTorch 在几个常见基准测试中的整体速度。

[55]**LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324 (1998).**

**基于梯度的学习应用于文档识别**

使用反向传播算法训练的多层神经网络构成了成功的基于梯度的学习技术的最佳示例。给定适当的网络架构，基于梯度的学习算法可用于合成复杂的决策面，该决策面可以对高维模式（例如手写字符）进行分类，只需最少的预处理。本文回顾了应用于手写字符识别的各种方法，并将它们与标准手写数字识别任务进行了比较。 卷积神经网络是专门为处理 2D 形状的可变性而设计的，其性能优于所有其他技术。现实生活中的文档识别系统由多个模块组成，包括字段提取、分段识别和语言建模。一种新的学习范式，称为图转换器网络（GTN），允许使用基于梯度的方法对这种多模块系统进行全局训练，以最小化整体性能测量。描述了两种在线手写识别系统。实验证明了全局训练的优势，以及图转换器网络的灵活性。还描述了用于读取银行支票的图形转换器网络。 它使用卷积神经网络字符识别器与全局训练技术相结合，为商业和个人支票提供记录准确性。它被商业部署，每天读取数百万张支票。

1. **Xiao, H., Rasul, K., & Vollgraf, R. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms.**

# **Fashion-MNIST：用于机器学习算法基准测试的新型图像数据集**

我们展示了 Fashion-MNIST，这是一个新的数据集，包含来自 10 个类别的 70,000 种时尚产品的 28x28 灰度图像，每个类别有 7,000 张图像。训练集有 60,000 张图像，测试集有 10,000 张图像。Fashion-MNIST 旨在作为原始 MNIST 数据集的直接替代品，用于对机器学习算法进行基准测试，因为它共享相同的图像大小、数据格式以及训练和测试拆分的结构。该数据集可在此 [https URL](https://github.com/zalandoresearch/fashion-mnist) 免费获得

1. **Spoon, K. et al. Toward software-equivalent accuracy on transformer-based deep neural networks with analog memory devices. *Front. Comput. Neurosci*. 53, (2021).**

**在基于Transformer的深度神经网络上实现与模拟存储设备相当的软件精度**

深度学习的最新进展是由不断增加的模型大小推动的，网络增长到数百万甚至数十亿个参数。如此庞大的模型需要快速且节能的硬件加速器。我们研究了基于非易失性存储器的模拟AI加速器，特别是相变存储器（PCM）的潜力，用于自然语言处理应用的软件等效精确推理。我们通过结合噪声感知训练来对抗固有的 PCM 漂移和噪声源，以及低至 INT6 的精度降低的数字注意力块计算，展示了一条通往 BERT（来自 Transformer 的双向编码器表示）的软件等效精度的途径

1. **Kariyappa, S. et al. Noise-resilient DNN: tolerating noise in PCM-based AI accelerators via noise-aware training. *IEEE Trans. Electron Devices* 68, 4356–4362 (2021).**

**抗噪 DNN：通过噪声感知训练在基于 PCM 的 AI 加速器中容忍噪声**

DNN噪声处理

基于相变存储器 （PCM） 的“模拟 AI”加速器由于内存计算提供的能效，在边缘应用中的推理中变得越来越重要。然而，PCM 设备固有的噪声源会导致深度神经网络 （DNN） 权重值不准确。这种不准确性会导致模型精度严重下降。为了解决这个问题，我们提出了两种技术来提高DNN的噪声弹性：1）漂移正则化（DR）和2）乘法噪声训练（MNT）。我们评估了在图像分类上训练的卷积网络和在语言建模上训练的递归神经网络，并表明我们的技术在一个月内将模型准确性提高了 12%。

1. **Gokmen, T., Rasch, M. J. & Haensch. W. The marriage of training and inference for scaled deep learning analog hardware. In *2019 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2019).**

# **规模化深度学习模拟硬件的训练和推理的结合**

电阻式横杆阵列是高效执行深度神经网络 （DNN） 推理工作负载的有前途的候选者。神经网络的权重矩阵映射到横杆阵列上的电导值，然后用作向量矩阵乘法引擎。尽管这种映射看起来很简单，但我们表明，对于大规模 DNN，权重必须来自考虑硬件诱导约束（如 ADC、DAC、噪声和器件故障）的训练过程，才能使推理任务在由横杆阵列组成的模拟硬件上成功运行。

1. **Rasch, M. J. et al. A flexible and fast PyTorch toolkit for simulating training and inference on analog crossbar arrays. In *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS)* (IEEE, 2021).**

# **灵活快速的 PyTorch 工具包，用于模拟横杆阵列上的训练和推理**

我们介绍了 IBM Analog Hardware Acceleration Kit，这是一个全新的开源工具包，用于在 PyTorch 中以方便的方式仿真模拟横杆阵列（[可在此 https URL](https://github.com/IBM/aihwkit) 上免费获得）。该工具包正在积极开发中，以“模拟瓦片”的概念为中心，该概念捕获在横杆阵列上执行的计算。模拟瓦片是构建块，可用于使用模拟组件扩展现有网络模块，并使用 PyTorch 框架的灵活性组成任意人工神经网络 （ANN）。模拟磁贴可以方便地配置，以模拟大量不同的模拟硬件特性及其非理想特性，例如器件间和周期间的变化、电阻器件响应曲线以及重量和输出噪声。此外，该工具包还可以设计自定义晶胞配置，并使用先进的模拟优化算法，如Tiki-Taka。此外，可以将向后和更新行为设置为“理想”，以便为仅针对推理加速的芯片启用硬件感知训练功能。为了评估此类芯片随时间推移的推理精度，我们提供了在相变存储器硬件上校准的统计编程噪声和漂移模型。我们的新工具包完全由 GPU 加速，可用于方便地估计材料特性和未来模拟技术的非理想性对任意人工神经网络精度的影响。

1. **Falcon, W. et al. PyTorch Lightning (2019); <https://github.com/PyTorchLightning/pytorch-lightning>**
2. **Biewald, L. Experiment Tracking with Weights and Biases (2020); <https://www.wandb.com/>**
3. **Kasim, M. F. et al. Building high accuracy emulators for scientific simulations with deep neural architecture search. Preprint at <https://arxiv.org/abs/2001.08055> (2020).**

# **通过深度神经架构搜索为科学模拟构建高精度仿真器**

使用神经网络训练数学模型

计算机模拟是科学发现的宝贵工具。然而，精确的仿真往往执行缓慢，这限制了它们在广泛的参数探索、大规模数据分析和不确定性量化的适用性。通过使用机器学习构建快速仿真器来加速仿真的一种有前途的途径需要大型训练数据集，而通过慢速仿真获得这些数据集的成本可能高得令人望而却步。在这里，我们提出了一种基于神经架构搜索的方法，即使在训练数据数量有限的情况下，也可以构建准确的仿真器。该方法使用相同的超架构、算法和超参数，在天体物理学、气候科学、生物地球化学、高能量密度物理学、聚变能和地震学等 10 个科学案例中成功地将模拟速度提高了 20 亿倍。我们的方法本身还提供了仿真器的不确定性估计，进一步增加了使用它们的信心。我们预计这项工作将加速涉及昂贵模拟的研究，允许更广泛的参数探索，并实现新的、以前不可行的计算发现。

1. **Rahmani, B. et al. Actor neural networks for the robust control of partially measured nonlinear systems showcased for image propagation through diffuse media. *Nat. Mach. Intell.* 2, 403–410 (2020).**

# **用于稳健控制部分测量的非线性系统的演员神经网络，展示了通过漫反射介质传播图像**

物理系统的输出，例如通过雾照射激光笔的光斑形成的加扰图案，通常很容易通过直接测量获得。然而，选择这种系统的输入以获得所需的输出是困难的，因为这是一个不恰当的问题;也就是说，有多个输入产生相同的输出。通过散射介质的信息传输就是这个问题的一个例子。用于成像的机器学习方法已在光子学中非常成功地实现，以从扭曲的强度衍射图案输出中恢复散射系统的原始输入相位和振幅对象。然而，控制这种系统的输出，而没有输入的例子，这些输入可以产生用户想要产生的输出对象类别的输出，是一个具有挑战性的问题。在这里，我们提出了一种在线学习方法，用于在输出端采集仅强度测量样本时，通过多模光纤投影任意形状。这个投影系统是非线性的，因为检测到的是强度，而不是复幅。我们展示了高达~90%的图像投影保真度，这与通过相位和振幅测量完全表征系统的黄金标准方法相当。所提出的方法的通用性和简单性可能会在只有部分测量的情况下为实际应用中的目标导向控制提供一种新的方法。

1. **Karniadakis, G. E. et al. Physics-informed machine learning. *Nat. Rev. Phys.* 3, 422–440 (2021).**

# **基于物理的机器学习**

尽管在利用偏微分方程（PDE）的数值离散化来模拟多物理场问题方面取得了巨大进展，但仍然无法将噪声数据无缝地整合到现有算法中，网格生成仍然很复杂，并且无法解决由参数化偏微分方程控制的高维问题。此外，解决隐藏物理的逆问题通常成本高昂，并且需要不同的公式和复杂的计算机代码。机器学习已经成为一种很有前途的替代方案，但训练深度神经网络需要大数据，而这些数据并不总是可用于解决科学问题。取而代之的是，这种网络可以从通过执行物理定律获得的额外信息（例如，在连续时空域中的随机点）进行训练。这种基于物理的学习集成了（噪声）数据和数学模型，并通过神经网络或其他基于核的回归网络实现它们。此外，还可以设计专门的网络架构，自动满足一些物理不变量，以获得更好的准确性、更快的训练和改进的泛化。在这里，我们回顾了将物理学嵌入机器学习的一些流行趋势，介绍了一些当前的能力和局限性，并讨论了物理知情学习在正向和逆向问题中的各种应用，包括发现隐藏的物理和解决高维问题。

[66]Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: A next-generation hyperparameter optimization

framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*,

2623–2631 (2019).

# **Optuna：下一代超参数优化框架**

本研究旨在为下一代超参数优化软件引入新的设计准则。我们提出的标准包括 （1） 允许用户动态构建参数搜索空间的运行定义 API，（2） 搜索和修剪策略的高效实现，以及 （3） 易于设置的多功能架构，可以部署用于各种目的，从可扩展的分布式计算到通过交互式界面进行的轻量级实验。为了证明我们的观点，我们将介绍Optuna，这是一款优化软件，是我们在开发下一代优化软件方面努力的结晶。作为一款采用运行定义原则设计的优化软件，Optuna 是同类产品中的第一款。我们将介绍在开发符合上述标准的软件时所需的设计技术，并通过实验结果和实际应用来展示我们新设计的强大功能。我们的软件在 MIT 许可（[此 https URL）](https://github.com/pfnet/optuna/)下可用。

[67]**Liu, W. et al. Programmable controlled mode-locked fiber laser using a digital micromirror device. *Opt. Lett.* 42, 1923–1926 (2017).**

# **使用数字微镜器件的可编程受控锁模光纤激光器**

基于数字微镜器件（DMD）的任意光谱振幅整形器被集成到大模面积光子晶体光纤激光器腔中。整形器充当腔内可编程滤波器，提供从正常到异常的大可调色散。因此，在具有瓦特级高输出功率的不同色散状态下实现了锁模。通过在DMD上编程不同的滤光片配置文件，激光器产生具有可调中心波长和可控带宽的飞秒脉冲。在合适的腔体色散和泵浦功率条件下，通过改变滤波器的振幅传递函数直接获得设计形状的光谱。结果表明，基于DMD的腔内滤波器在锁模光纤激光器中灵活控制脉冲动力学的多功能性。

1. **Matthès, M. W., del Hougne, P., de Rosny, J., Lerosey, G. & Popoff, S. M. Optical complex media as universal reconfigurable linear operators. *Optica* 6, 465–472 (2019).**

# **光学复杂介质作为通用可重构线性算子**

（Control DMD的编码方法）

使用光学器件执行线性运算是电信、光学模拟计算和机器学习等许多领域的关键组成部分。对于其中许多应用，关键要求是对制造不准确的鲁棒性、可重构性和可扩展性。我们提出了一种使用复杂光学介质（如多模光纤或散射介质）作为由波前整形驱动的计算平台进行线性运算的方法。给定一个大的随机传输矩阵，表示光在这种介质中的传播，我们可以通过找到合适的输入和输出投影仪来提取任何所需的较小线性算子。我们通过使用空间光调制器查找输入波前来演示这一概念，该调制器使复杂介质在光场上充当所需的复值线性算子。我们实验性地构建了几个16×1616×16运算符，并讨论我们方法可扩展性的基本限制。它提供了可重新配置、稳健且易于制造的线性光学模拟计算单元的前景。

1. **Popoff, S. M. & Matthès, M. W. ALP4lib: q Python wrapper for the Vialux ALP-4 controller suite to control DMDs. *Zenodo* <https://doi.org/10.5281/zenodo.4076193> (2020).**

# ALP4lib：用于 Vialux ALP-4 控制器套件的 Python 包装器，用于控制 DMD

相关论文：

# **Turning Optical Complex Media into Universal Reconfigurable Linear Operators by Wavefront Shaping**

# **通过波前整形将光学复杂介质转换为通用的可重构线性算子**

使用光学器件执行线性运算是电信、光学模拟计算和机器学习等许多领域的关键组成部分。对于其中许多应用，关键要求是对制造不准确性和可重构性的鲁棒性。目前定制的光子器件或相干光子电路的设计只能部分满足这些需求。在这里，我们提出了一种通过使用复杂的光学介质（如多模光纤或薄散射层）作为由波前整形驱动的计算平台来执行线性运算的方法。给定一个代表光在这种介质中传播的大型随机传输矩阵 （TM），我们可以通过找到合适的输入和输出投影仪来提取所需的更小的线性算子。我们讨论了我们的方法可以实现的线性变换大小的基本上限，并提供了一个实验演示。对于后者，我们首先使用非干涉相位检索方法检索复杂介质的TM。然后，我们利用大量的自由度，使用空间光调制器（SLM）找到输入波前，使由SLM和复杂介质组成的系统在光场上充当所需的复值线性算子。我们实验性地构建了几个16×16复值算子，并且能够随意从一个运算符切换到另一个运算符。我们的技术为可重构、稳健且易于制造的线性光学模拟计算单元提供了前景。

1. **Hillenbrand, J., Getty, L. A., Wheeler, K. & Clark, M. J. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am*. 97, 3099–3111 (1995).**

**美式英语元音的声学特征**

本研究的目的是复制和扩展 Peterson 和 Barney （PB） [J. Acoust. Soc. Am. 24， 175-184 （1952）] 对元音声学的经典研究。45 名男性、48 名女性和 46 名儿童在 h-V-d 音节中产生了元音 /i、I、e、epsilon、ae、a、[符号：见文本]、O、U、u、lambda、3 iota/。F1-F4的共振峰轮廓是使用自定义交互式编辑工具从LPC光谱测量的。为了与PB数据进行比较，共振峰模式在目视检查判断为最大稳定的时间进行采样。对共振峰数据的分析表明，在F1和F2的平均频率以及相邻元音之间的重叠程度方面，现有数据与PB数据之间存在许多差异。与最初的研究一样，听力测试表明，这些信号几乎总是被识别为说话者想要的元音。判别分析表明，元音的分离度比基于共振峰模式静态样本的PB数据更差。但是，如果包括持续时间和频谱变化信息，则可以高度准确地分离元音。

1. **Veit, A.,Wilber, M. & Belongie, S. Residual networks behave like ensembles of relatively shallow networks Preprint at <https://arxiv.org/abs/1605.06431> (2016).**

# **残差网络的行为类似于相对浅层网络的集合**

在这项工作中，我们提出了一种对残差网络的新解释，表明它们可以被看作是许多不同长度的路径的集合。此外，残差网络似乎通过在训练期间仅利用短路径来实现非常深的网络。为了支持这一观察结果，我们将残差网络重写为路径的显式集合。与传统模型不同，通过残差网络的路径长度各不相同。此外，一项病变研究表明，这些路径表现出类似集成的行为，因为它们不强烈地相互依赖。最后，也是最令人惊讶的是，大多数路径都比人们预期的要短，并且在训练过程中只需要短路径，因为较长的路径不会产生任何梯度。例如，具有 110 层的残差网络中的大部分梯度来自深度仅为 10-34 层的路径。我们的研究结果揭示了似乎能够训练非常深的网络的关键特征之一：残差网络通过引入可以在非常深的网络范围内携带梯度的短路径来避免梯度消失问题。