

原创 DPU 重构计算模式

■本报记者 袁一雪

在过去的十年,电子驱动的计算处理器对人工智能(AI)发展产生了巨大影响。但随着技术的发展,电子硬件的性能即将触碰到“天花板”,而只有突破“天花板”才能让人工智能发展更进一步。

能够替代电子驱动计算处理器的光学计算处理器,一直为研究人员所倚重。它不仅能够克服电子学的固有限制,还可以将能效、处理速度和计算吞吐量提高几个数量级。

目前,已经有研究人员利用光学计算的特性,构建专用光学处理器,其在数据和信号处理方面所表现出的性能远远超过现在的电子处理器。然而,现有的光学 AI 加速器只能为特定神经网络体系结构或特定任务定制单一功能。换言之,对于不同任务的不同 AI 算法,现在的“设备”显得有些“力不从心”。不仅如此,现有的神经网络模型复杂度和实验性能水平较低,导致网络性能(如分类精度)与最先进的电子神经网络相比存在较大差距。

近日,中国工程院院士戴琼海团队提出了一种光电可重构计算模式,针对光学网络设计空间的灵活性有限、难以集成理想的非线性运算,以及难以灵活控制复杂的数据流等问题提出了新的解决思路。

将 AI 加速处理器推向新高度

5月5日,这篇第一作者是清华大学博士研究生周天祝、通讯作者为戴琼海的论文登上《自然—光子学》期刊封面。该论文作者之一、清华脑与认知科学研究院及未来芯片技术高精尖创新中心特聘研究员林星在接受《中国科学报》采访时表示,针对现有神经网络模型复杂度低、可重构难以实现以及实验性能不佳等难题的关键原因,该团队通过以下三方面进行了突破:

首先,该团队在空频域开辟了衍射光学计算的新型光电计算架构。“衍射智能光计算架构,缩写为DPU,是基于光学衍射的物理现象,通过构建大规模的光学相控阵列以实现高复杂度的光学互联和神经网络架构。其原理具备高度的可扩展性和集成潜力。”林星解释道,“我们这次的研究工作通过采用高通量可编程的光电器件,融合光计算和电计算的互补性优势,将其在构建人工智能加速处理器的应用推向了一个新的高度。”

其次,这次提出的 DPU 充分利用了光的波粒二象性,神经网络权重的调整通过控制光波传播的波前分布来实现,采用光电效应实现人工神经元的功能,包括对衍射光场的线性加权求和以



“衍射光电智能计算体系架构和理论方法在很大程度上解决了领域内困扰已久的难题,为研发高性能光电智能处理器和芯片打开了更为广阔的空间,也为促进该领域的发展及其产业落地做出了一定贡献。”

该研究论文登上《自然—光子学》封面

及非线性激活响应。林星解释称,系统运行过程衍射光计算部分几乎承担所有计算操作,而采用高通量可编程的光电器件并结合电子计算的灵活特性,能够实现高速数据调控及大规模网络参数的编程和网络结构的重构。此外,由于光计算部分采用了三维光场传播,避免了对图像和视频的向量化处理和寻址操作,从而实现了视觉信息的高效处理。

最后,此次研究提出了一种自适应在线训练方法来校正系统误差累积,而这种自适应的训练方法通过测量神经网络内部的状态实现对计算参数的实时调整。因此,在不依赖网络结构且不增加系统复杂度的情况下,该 DPU 能够普适于现有的光电智能计算系统。

“可以说我们这次研究工作所提出的衍射光电智能计算体系架构和理论方法在很大程度上解决了上述领域内困扰已久的难题,通过构建原型系统及可扩展性分析验证了光电智能计算的优越性和巨大潜力,为研发高性能光电智能处理器和芯片打开了更为广阔的空间,也为促进该领域的发展及其产业落地做出了一定贡献。”林星补充道。

分类识别精度有所提升

取得这样的成绩并不容易。事实上,

在研究之初,该团队研究人员就已经对前路崎岖有着充足的心理准备。“我们的研究目标是要通过解决光电智能计算领域中的难点问题,提高其在执行分类和识别等任务上的模型与实验精度。”林星回忆说。

起初,研究人员构建的全连接的网络模型性能有限,也曾经受到系统不稳定的困扰。不过,解决问题的过程就是团队不断创新的过程。团队成员经过反复讨论并调整光电神经网络构架和超参数,最终创建了网络中网络模型。“我们的光电神经网络在 MNIST 手写数据集和 Fashion-MNIST 时尚产品数据集上的模型精度都超越了卷积神经网络 LeNet-4,并逼近 LeNet-5。”林星表示。

而且,通过不断进行的硬件系统优化,包括更换所采用的光电器件及其相应机械固定件设计等,以及开发自适应训练算法,研究人员也提升了系统实验的稳定性。

“此外,我们也花了大量的时间优化和升级光电器件的数据接口,以提升系统带宽和数据调控速度,最终实现了高速的图像和视频的分类识别。”林星告诉记者,“运行同样的神经网络,光电计算系统与特斯拉 V100 图形处理器(GPU)相比,计算速度提高了8倍,系统能效提升超过1个数量级,核心模块计算能效则能够提升4个数量级。”

有望让我国在新一轮科技变革中占得先机

“长期以来,我国在电子计算以及芯片领域受到高精度光刻机等制约,一直处于落后位置。光电计算为我国摆脱这一制约提供了可能。”戴琼海告诉《中国科学报》,“光电智能计算与芯片是一个非常广阔的前沿研究领域,这需要多学科交叉融合。希望我们的研究工作能够吸引国内更多不同背景的科研人员和研究机构从事这个领域的研究,从而推动我国在该领域学术圈乃至产业生态圈的建设。”

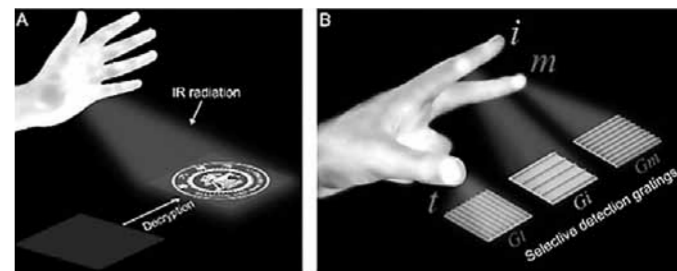
戴琼海团队抓住光学计算与芯片发展的机遇,在其原理验证和原型系统的构建等方面率先展开了原创性的研究工作,所提出的计算范式将促进领域光电材料、光电器件、加工工艺以及集成方法的研发。

谈及未来,戴琼海与林星皆表示,将进一步挖掘光电智能计算技术的潜力,致力于解决系统小型化和芯片集成的难点问题。“我们将使用超材料构建片上相控阵列实现原型系统的芯片集成,从而更为全面地体现光电智能计算的优越性。”戴琼海说。

巴黎中央理工—高等电力学院教授 Daniel Brunner 与瑞士联邦理工学院洛桑分校教授 Demetri Psaltis 认为,这项研究的准确性与能源效率,都可以和竞争性数字 NN 基准测试模型以及现代 GPU 相媲美,并证明了光学解决方案与未来高性能计算的相关性。“这项研究的非线性运算放大了数据表示的维数,最终使 NN 能够挖掘出隐藏的特征,进而利用它们来进行具有挑战性的计算。尽管其互连受到严格限制,但是它们通过级联多层实现了竞争性性能。”

“这一工作属于我们研究路径规划中的阶段性成果。我们预期最终集成化的光电智能芯片能够实现光载海量信息的实时处理,将极大促进其在终端计算、边缘端计算、数据中心的应用,推动自动驾驶、机器人技术领域的发展。其中,小型化的空间光智能光电计算系统可应用于空间光通信,与视觉成像系统融合实现光载视觉信息(包括高维光谱视觉信息及光场视觉信息)的实时处理。同时,集成化的光电智能计算系统还能够和光纤通信融合,实现大规模光信息的高效处理,应用于无线通信、云计算和数据中心等领域,从而引领新一代智能产业和信息技术的变革。”林星展望道。

相关论文信息: <https://doi.org/10.1038/s41566-021-00796-w>



手可作为一个天然且智能化的红外光源。(A)整只手作为光源和(B)每根手指作为独立光源的示意图。

本报道 近日,上海交通大学材料科学与工程学院教授冯涛团队在以人手作为智能红外光源进行信息识别研究中获得重要进展。相关研究在线发表于美国《国家科学院院刊》。

红外光源作为光源的一种,在传感、能源、通信、军事等领域得到了广泛应用。相对于人造红外光源,人体是一个天然的红外光源,发射的红外光峰值在 9 微米左右,正好处于大气透射窗口(7.5 微米~14 微米)且发射的为非相干光。基于这一原理,研究人员在这项工作中研究了手作为一个天然智能红外光源在不同系统里面的应用:整只手作为光源集成到防伪系统中进行信息识别;每根手指作为一个独立的红外光源集成到信号生成系统中进行手语识别。

研究人员发现,在将手集成到基于反射的防伪系统中进行信息识别的工作中,防伪标签由铝基底上喷涂透明的聚二甲基硅氧烷(PDMS)图案构成。金属铝对红外光具有很强的反射率,而 PDMS 聚合物由于官能团的红外吸收导致反射率降低。当手作为红外光源放在防伪标签一侧时,防伪图案的不同部分反射的红外光强度不同,被红外相机捕捉并可视化,从而使肉眼不可见的防伪图案被识别出来。

通过调控 PDMS 薄膜的红外反射率,研究人员进一步实现了多级彩色信息加密与识别,与单色的防伪图案相比,进一步提高了信息加密的等级与安全性。论文第一作者、上海交通大学材料科学与工程学院博士生安顺告诉《中国科学报》,不仅整只手可以作为一个光源,每根手指也可以作为独立的红外光源且发射的红外光可以被金属反射型光栅衍射。通过调控光栅的周期及取向,每根手指发射的红外光可以选择性地被特定的光栅衍射,从而将该光栅“照亮”。光源的数量以及相对位置可以通过改变手势进行调节。当改变手势时,红外相机可以从同一个光栅阵列中识

别出不同的红外衍射图案,且这些衍射图案与手势之间一一对应。如将手作为密码集成到防伪系统中,将会进一步提高防伪的等级,以及信息加密的安全性。

这种手势与红外衍射图案之间的对应关系进一步被集成到信号生成系统中,以实现手语识别功能。研究人员设计了一种仅需 3 根手指即可产生 26 种手语以及相应的 26 种红外衍射图案的手语识别系统,并将这 26 种红外衍射图案与 26 个英文字母一一对应。与传统的直接识别手语轮廓的手语识别系统相比,本研究将复杂的手势转换成简单的红外衍射图案进行识别,不仅避免了环境光线以及复杂背景对于识别的干扰,在手势分割、特征提取等方面也具有较高的鲁棒性,采用简单的算法即可以实现实时、快速的手语识别。

同时,这项研究证明了手作为一个天然且智能化的红外光源在不同系统中的应用。将手集成到这些系统里不仅提高了系统的可持续性,并且由于大脑直接参与光源的调控,所以也进一步提高了系统的智能性和可控性。

“这项工作为提高人工系统的可持续性和智能性提供了一种新思路,可以促进以人体红外光源取代人造红外光源,从而在更多领域的应用。”安顺说。

(黄辛) 相关论文信息: <https://www.pnas.org/content/118/15/e2021077118>

速速

“AI 文案大师”让朋友圈更有“范”

本报道 日前,北京智源人工智能研究院(简称智源研究院)悟道文澜团队发布“AI 文案大师”小程序,可以自动生成相应的文案。

据介绍,“AI 文案大师”小程序可以实现“输入一张图片,输出一句文案”的功能。在这个过程中,“图片”和“文本”两种数据分别对应于“视觉”和“语言”两种模态。

机器学习领域将这种输入输出数据分别属于不同模态的任务称为“跨模态”任务。例如,可以训练模型根据输入图片中提取出的视觉信息,自由地生成相应的描述性自然语言文本,这样的任务称为“图像描述”。此外,还可以把图片作为输入,提取其视觉特征,并在现有的文本库中检索最符合图片特征的文字描述,这样的任务称为“图文检索”或“图文匹配”。

智源研究院悟道文澜团队相关负责人表示,“视觉—语言”多模态预训练有助于提取到两种模态数据中的关键信息,从而实现更高质量的图文匹配。为了实“AI 文案大师”系统要求的图文匹配质量,需要提升神经网络的表征能力。



“AI 文案大师”演示的“图文匹配”实例

得益于近年来自监督学习在提取数据内在关键知识方面的进展,智源研究院悟道文澜团队采用先进的跨模态对比学习(自监督学习算法中的一种)算法 CMCL,利用某一个“图文对”中的图像模态或文本模态来构建该“图文对”的负样本,并设计了一种基于队列的词典来扩大负样本数目,从而得到更好的数据表征。

未来,智源研究院悟道文澜团队还将继续发布各种基于大规模中文多模态预训练模型的研究成果和有趣应用。(郑金武)

新一代全栈信创云推出“一云多芯”

本报道 近日,我国企业级云计算产品及服务提供商易捷行云(EasyStack)在北京举行发布会,推出其“一云多芯、全栈可进化”的新一代全栈信创云。会上,易捷行云联合创始人兼首席运营官王瑞琳透露,基于他们对 200 多家金融行业用户需求调研发现,50%以上的金融行业会考虑把国产化信息技术应用创新建设和信息化重构两步并作一步走。此外,70%的金融用户在相应的数据中心建设上都会采用信创云,而 90%的用户都有一云多芯的需求。

据介绍,在易捷行云已服务的 1000 多家大型用户中,金融行业用户约占 10%。目前,他们已经开始为金融用户建设信创云。会上,易捷行云还发布了新一代金融全栈信创云及“新一代金融全栈信创云解决方案及成功案例白皮书”。

在发布会上,易捷行云创始人兼 CEO 陈喜伦判断,2021 年将是“信创云元年”。但他也表示,中国有巨大的市场,超大规模的用户和复杂的应用场景,也有一批高层次人才,中国的信息技术应用创新云是“世界级水准的产品”。(赵广立)

开源软件供应链重大基础设施建设势在必行

■武延军

全球范围内,开源软件已经成为了基础设施的核心要素,是构成操作系统、数据库等重要“卡脖子”基础软件的核心“元器件”。能否为设备、系统、产业和行业提供高质量、高可靠的、可持续发展的开源软件供应,关系到国内当前和未来 IT 科研、产品与生态的核心竞争力,甚至是“生命线”。然而,近年来,国内开源软件供应链“卡脖子”事件频频发生,开源软件产业仍面临着不少根本问题。

为了贯彻落实发展开源软件的国家战略,实现开源软件的可靠供应,需要尽快打造核心基础设施支撑,解决基础软件核心技术受制于人的问题,推动我国软件产业走上自主创新、自立自强之路。建设开源软件供应链重大基础设施势在必行。

开源席卷全球软件产业

自上世纪 70 年代软件商业化以来,它为世界创造了巨大的财富。近年来全球市值排名前七的公司都有一个共同特点,即核心竞争力都来自于软件。

工业和信息化部公布的数据显示,2020 年中国软件产业总收入 81616 亿元,占 GDP 超过 8%,已经成为国民经济的支柱产业。

当前,开源是软件产业的主流模式。Synopsys 2020 年抽样审计报告显示,99%的商业软件含有开源组件,75%则直接由源代码组成。

同时,国内的开源产业也已经日趋成熟。面向开源及私有软件项目的托管平台 GitHub 在 2020 年 4 月的统计数据显示,来自中国的源代码贡献量已经仅次于美国。就在今年年初, Linux 内核 5.10 版本中,华为的内核

代码贡献位列第一。

不仅如此,开源软件对关键产业具有重要的战略支撑作用。“缺芯少魂”是产业面临的巨大挑战。比如在被“卡脖子”的芯片领域,开源软件就是芯片的核心竞争力和生态的重要支撑,软件优化可充分发挥芯片的性能;同时,一款芯片能否被主流开源软件(如 Linux 内核)所支持,决定了该芯片的应用范围和市场份额。

“卡脖子”风险

尽管我国已经开始积极推动开源生态的建设,但国内开源软件产业仍面临着根本问题。

首先是产业价值不高。以美国红帽(Red Hat)公司和国内主要操作系统厂商对比为例,前者在 2019 年的收入约为 30 亿美元,而后者年收入则在亿元人民币规模。

其次,创新创业支撑不足。近年来,美国纷纷诞生了一些基于开源的独角兽企业,如著名的开源协作软件 Slack 和开源云计算软件 Snowflake,市值分别已经达到了 200 亿和 700 亿美元。在国内,极度缺乏这样基于开源的高价值的创新企业公司。

其三,开源生态受制于人。华为在 2019 年发布会上指出,谷歌依托安卓操作系统的 GMS(谷歌移动服务)对华断供,至少影响了 100 亿美元的海外销售收入。

事实上,开源软件供应链“卡脖子”事件频频发生,已经给国内软件产业敲响了警钟。

例如,Docker 是云计算领域最重要的开源应用容器引擎。2020 年 8 月 13 日起,它的企业版 DockerEE 和 DockerHub 禁止被美国政府列入贸易管制

“实体清单”的企业使用,一批中国企业、科研院所和高校受到直接影响。

CentOS 是国内服务器领域使用最多的开源操作系统,2020 年 12 月,红帽公司宣布将于 2021 年年底停止维护 CentOS 8,给中国企业造成了大量的应对成本。

再如 Openwall 的“隐形断供”。Openwall 是开源基础软件安全预警平台。漏洞共享、安全预警是操作系统等基础软件产业的重要环节,可国内在这个领域仍然处于空白状态。由于获取受限,国内基础软件存在 2 周以上的“安全预警空白区”。

除此之外,国内的开源软件供应链还面临新型 OpenChain 的“准入”风险。OpenChain 是开源软件合规性标准,目标是在交换开源软件解决方案的组织之间建立信任基准,确保程序被设计成为合规工作。Linux 基金会采用快速过会的方式将 OpenChain 转变成国际标准,意味着国内软件企业必须符合 OpenChain 标准才能进入国际市场。但国内本就缺乏开源软件使用的合规性审核,这一标准的实施平添了壁垒,将限制国内软件产品进入国际“大循环”。

为国产软件“定魂筑根”

国家“十四五”规划中提及,补齐产业链、供应链短板,实施产业基础再造工程,加大重要产品和关键核心技术攻关力度,发展先进适用技术,推动产业链供应链多元化。

因此,为了贯彻落实发展开源软件的国家战略,实现开源软件的可靠供应,今年在多方支持下,中国科学院软件研究所(以下简称软件所)与软件所南京软件技术研究院(以下简称软件所南京院)启动了开源软件供应链重大基