

编者按

当前人工智能技术迅猛发展,在算法理论、大数据和强计算能力的支持下,以深度学习技术为代表的人工智能技术在近期迅速应用于众多领域. AI 芯片、定制加速器、智能计算平台等也成为学术和产业热点. 从技术发展趋势来看,进入人工智能大发展的时代,异构并行进一步深化,面向智能应用定制计算加速器成为当前热点. 定制领域不断扩展,从卷积神经网络加速器、深度学习加速器、通用机器学习加速器到图计算加速器等;器件基础也不断深化,从可重构 FPGA, ASIC 芯片到新型忆阻器,全面呈现百花齐放态势.

《中国科学: 信息科学》特组织“面向智能应用的定制计算加速器技术”专题,邀请各位学者撰文呈现最新的技术成果. 本专题共收录了 6 篇文章,内容涉及众核处理器、通用向量 DSP 深度学习加速器、基于 FPGA 加速器、图计算加速器和定制化深度学习加速器.

江南计算技术研究所李宏亮等撰写的文章“面向智能计算的国产众核处理器架构研究”综述了众核处理器体系结构主要技术路线,介绍了国产众核处理器采用多种技术为人工智能应用提供技术支撑情况.

国防科技大学王慧丽等撰写的文章“基于通用向量 DSP 的深度学习硬件加速技术”综述了深度学习硬件加速的多种技术,介绍了兼顾向量 DSP 通用性和深度学习硬件加速器专用型的可重构加速阵列体系结构方案.

北京大学卢丽强等撰写的文章“面向卷积神经网络的 FPGA 设计”重点综述了深度学习中卷积神经网络的常用 4 种算法,介绍了在 FPGA 上进行硬件优化设计的方法.

图数据结构能够表达复杂的关联关系,其应用非常广泛,对此类应用进行加速的研究工作正在日益受到关注. 上海交通大学王靖等撰写的文章“面向图计算的内存系统优化综述”重点综述了异构架构图计算中内存系统的管理及优化方法,分析了 GPU, FPGA, ASIC, PIM 等的架构特点与内存方面的优化工作,同时总结了图计算在内存方面的机遇与挑战.

引用格式: 陈云霁, 窦勇, 梁云, 等. 面向智能应用的定制计算加速器技术专题编者按. 中国科学: 信息科学, 2019, 49: 245-246, doi: 10.1360/N112019-00064

清华大学严佳乐等撰写的论文“低功耗神经网络计算技术研究”提出了可重构神经网络加速器硬件架构,从硬件资源高效调度、访存优化设计和数据复用 3 个方面介绍了相关技术成果.

最后一篇文章“深度学习驱动的区域专用架构”来自北京比特大陆科技有限公司的马立伟研究员,该文首先分析人工智能发展阶段,同时从熵的角度讨论深度学习领域专用结构,观点很有新意.另一方面,讨论了深度学习数字计算的数制基础问题,为专用深度学习加速器的稀疏与量化研究提供了新颖的技术思路.

智能应用仍在持续发展,面向领域进行定制化设计是当前处理器体系结构研究的一个重要方向,期盼相关研究人员能够把握这个机遇,将应用与结构设计紧密耦合,创新发展,取得丰硕成果.

特约编辑: 陈云霄 中国科学院计算技术研究所
 窦 勇 国防科技大学
 梁 云 北京大学
 汪东升 清华大学