

# 奥创智能网

首页 人工智能 VR/AR 智能硬件 机器人 无人驾驶 互联网+ 科学技术

当前位置： 首页 > 智能硬件 > 揭秘寒武纪神经网络处理器 深度学习处理器指令集DianNaoYu

## 揭秘寒武纪神经网络处理器 深度学习处理器指令集DianNaoYu

li, yong 智能硬件 03-16 1580 0

就在全世界媒体的焦点锁定于谷歌AlphaGo连续3盘战胜李世石的同时，中国科学院计算技术研究所陈云霁、陈天石课题组提出的深度学习处理器指令集DianNaoYu被计算机体系结构领域顶级国际会议ISCA2016（International Symposium on Computer Architecture）所接收，其评分排名所有近300篇投稿的第一名。模拟实验表明，采用DianNaoYu指令集的寒武纪深度学习处理器相对于x86指令集的CPU有两个数量级的性能提升。

 揭秘寒武纪神经网络处理器 深度学习处理器指令集DianNaoYu

（寒武纪处理器）

### 寒武纪处理器是长期技术积累的成果

陈天石和陈云霁研究员在2014年5月刊发的《中国计算机学会通讯》（第10卷第5期）写道：

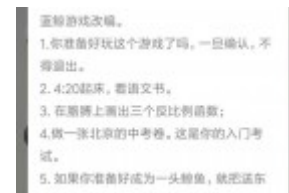
## 编辑推荐



从翻译到日常出行 无所不在的人工智能！



谷歌DeepMind很厉害，微软成立新AI实验室挑战



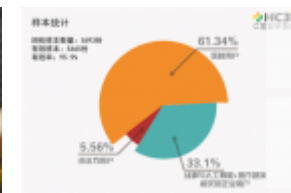
黑客玩改良版“蓝鲸”游戏 结局万没想到



摩拜单车最早是蓝色：名字又土又难听



人工智能有多智能？



2016-2017年度人工智能+医疗市场分析及趋势报

You reached the end.

“从2008年到现在，我们已经在体系结构和人工智能的交叉研究方向上工作了6年。作为国际上为数不多的几个长期开展此方向研究的团队之一，我们在不被认可中坚持了下来，并尝试通过自己的努力来改善这个领域的环境（当然近年来环境的改善也得益于深度学习的兴起），最终得到了学术界一定程度的肯定。

回想起来，如果我们紧紧跟随国际学术圈的热点走，我们是很难拿到ASPLOS最佳论文奖的。原因有两个方面：第一，当我们看到别人的“热点”论文发表后再去跟着做，可能已经比别人晚了若干年。尤其是体系结构领域，论文的工作周期非常长（ASPLOS2014上发表的论文，我们在2012年就启动相关工作了），要跟上热点很困难。第二，当跟随热点时，我们的工作不可避免会被视为对某个过往论文的改进。这种改进效果必须非常显著，机理必须非常清晰，才能打动顶级会议挑剔的审稿人。这对于论文写作提出了很高的要求，而中国大陆研究者往往在英文论文写作上不占优势。但这里存在一个矛盾：紧跟多变的国际学术圈热点，论文不容易在顶级会议上发表；而探讨的问题若不是国际学术圈热点，论文同样很难在顶级会议上发表。

面对这个矛盾，我们的看法是：研究者应该坚持自己的学术理想，重视论文但不为论文发表所左右；同时尽力宣传自己的研究方向，推动这个方向被国际学



99%的人都不去用 人工智能电视让你变智障



中国人工智能产业创新联盟在京成立 近200家



阿里影业新战略：建新基础设施 做服务者而非



天猫国际618同比成交超500% 引爆全球新零售狂

### 最新文章

### 热评文章

### 随机文章

猎豹移动旗下直播平台Live.me获6000万美元融资

优步伦敦恐袭后仍按“峰时定价”被批低级到家

天天快递网点老板跑路 快递员为讨薪暂扣千份快..

“中国最美导游”孙洁：搭建中欧旅游更大“桥梁”

携程CEO孙洁参加中欧工商峰会推广中国旅游

沃尔玛为新员工培训用上VR 模拟“黑五”大促销

要逆天！人工智能还能识别动物的情绪了

旧金山或将封杀快递机器人 因为马路服务于人类


苹果推出智能音箱HomePod 挑战亚马逊和谷歌

百度之后 知乎也要“灵活”实名制了

术圈的主流认可。经过长期的等待和努力，也许有一天，自己的研究方向就会成为热点。到那时，过去的一切坎坷都会被证明是值得的。”

截至目前，陈天石博士和陈云霄研究员就光寒武纪系列的技术成果，已经斩获两篇AS PLOS，两篇ISCA，一篇MICRO，一篇HPCA，这些是计算机体系结构方面国际四大顶级会议，只不过只有科研圈子里关注，普通人还不明白其中的意义。

因此，寒武纪处理器并非借着借着阿法狗对决李世石的热点横空出世，而是长期技术积累的结果，并在数年前就于高端国际会议中荣获殊荣。事实上，中国在智能芯片上是引领世界的——以寒武纪芯片为例，美国的哈佛、斯坦福、MIT、哥伦比亚等高校都跟在我们后面做。铁流认为，寒武纪之所以过去一直不见于主流媒体，仅仅是国内很多媒体人缺乏专业知识和辨识能力，而导致了更愿意鼓吹国外所谓“高科技”，却对真正非常具有含金量的自主技术选择性失明。

 揭秘寒武纪神经网络处理器 深度学习处理器指令集DianNaoYu

陈云霄（左）和陈天石（右）兄弟

## 寒武纪神经网络处理器庐山真面目

目前，寒武纪系列已包含三种原型处理器结构：

寒武纪1号（英文名DianNao，面向神经网络的原型处理器结构）；

寒武纪2号（英文名DaDianNao，面向大规模神经网络）；

寒武纪3号（英文名PuDianNao，面向多种机器学习算法）。

DianNao是寒武纪系列的第一个原型处理器结构，包含一个处理器核，主频为0.98GHz，峰值性能达每秒4520亿次神经网络基本运算，65nm工艺下功耗为0.485W，面积3.02mm<sup>2</sup>。在若干代表性神经网络上的实验结果表明，DianNao的平均性能超过主流CPU核的100倍，但是面积和功耗仅为1/10，效能提升可达三个数量级；DianNao的平均性能与主流GPGPU相当，但面积和功耗仅为主流GPGPU百分之一量级。

**DianNao的核心问题是如何让有限的内存带宽喂饱运算功能部件，使得运算和访存平衡，从而达到高效能比。难点在于选取运算功能部件的数量、组织策略以及片上RAM的结构参数。**由于整个结构参数空间有上千万种选择，模拟器运行速度不及真实芯片的十万分之一，不可能蛮力尝试各种可能的设计参数。为解决此问题，使用了一套基于机器学习的处理器性能建模方法，并基于该性能模型最终为DianNao选定了各项设计参数，在运算和访存间取得了平衡，显著提升了执行神经网络算法时的效能。

**即便数据已经从内存取到了片上，搬运的能耗依然非常高。NVidia首席科学家Steve Keckler曾经指出，在40nm工艺下，将64位数据搬运20毫米所花的能耗是做64位浮点乘法的数倍。**

因此，要降低处理器功耗，仅仅降低运算功耗是不够的，必须优化片上数据搬运。中科院计算所提出对神经网络进行分块处理，将不同类型的数据块存放在不同的片上R

AM中，并建立理论模型来刻画RAM与RAM、RAM与运算部件、RAM与内存之间搬运次数，进而优化神经网络运算所需的数据搬运次数。相对于CPU/GPU上基于cache层次的数据搬运，DianNao可将数据搬运减少10~30倍。



揭秘寒武纪神经网络处理器 深度学习处理器指令集DianNaoYu

(DianNao结构)

DaDianNao在DianNao的基础上进一步扩大了处理器的规模，包含16个处理器核和更大的片上存储，并支持多处理器芯片间直接高速互连，避免了高昂的内存访问开销。在28nm 工艺下，DaDianNao的主频为606MHz，面积67.7 mm<sup>2</sup>，功耗约16W。单芯片性能超过了主流GPU的21倍，而能耗仅为主流GPU的1/330。64芯片组成的高效能计算系统较主流GPU的性能提升甚至可达450倍，但总能耗仅为1/150。

**虽然神经网络已成为模式识别等领域的主流算法，但用户很多时候可能倾向于使用其他一些经典的机器学习算法。**

例如程序化交易中经常使用线性回归这类可解释性好、复杂度低的算法。在此背景下，寒武纪3号多用途机器学习处理器PuDianNao应运而生，当前已可支持k-最近邻、k-均值、朴素贝叶斯、线性回归、支持向量机、决策树、神经网络等近十种代表性机器学习算法。PuDianNao的主频为1GHz，峰值性能达每秒10560亿次基本操作，面积3.51mm<sup>2</sup>，功耗为0.596W（65nm工艺下）。PuDianNao运行上述机器学习算法时的平均性能与主流GPGPU相当，但面积和功耗仅为主流GPGPU百分之一量级。



( PuDianNao版图 )

## 神经网络处理器的市场前景

在上世纪80年代，因人工智能无法达到公众/投资人的预期，导致整个行业陷入低谷。

近年来，随着算法、应用和工艺三个方面都发生了剧烈的变化，神经网络处理器涅槃重生。



2006年，Hinton、LeCun和Bengio等人提出了深度学习方法，在深层人工神经网络的训练上取得了巨大的突破。

简单地说，深度学习方法在传统的人工神经网络训练中增加了一个预训练阶段，即用无监督学习对每一层网络进行一次专门的训练，然后才用有监督学习对整个网络进行总体训练。通过深度学习方法，人工神经网络的效果一举赶上甚至显著超过了支持向量机等其他机器学习方法，在IBM、谷歌、微软、科大讯飞、百度等公司很多工业级图像和语音处理应用上取得了非常好的效果。

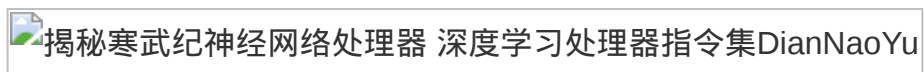
为什么深度学习会有效，暂时还没有传统的统计学习理论方面的完美证明。目前一种比较直观的解释是：**分层预训练相当于对输入数据进行逐级抽象，这暗合生物大脑**



**的认知过程比较**（例如人脑就是一种深层的神经网络，在认知过程中会逐层将看到的原始图像或听到的原始声波最终抽象成语义符号）。

既然人工神经网络已经重新成为最有效的认知任务处理算法（至少是之一），只要人工智能健康发展，专门的神经网络处理器自然能随着产业发展而茁壮成长。

另外，随着日常生活显然需要进行大量的认知活动，自然而然地，计算机体系结构研究者的目光必须要从传统的科学计算转到认知任务上。事实上，很多大公司已经认识到这一点。Intel和IBM等传统的硬件厂商都已经成立了专门的部门进行认知任务处理的研究。而现在体系结构研究中最常见的测试集Parsec中近半数都是认知类的应用（如bodytrack、facesim、freqmine、streamcluster、vips等）。在认知任务已经成为了当前计算机最主要的任务之一的情况下，用户和市场自然会有加速人工神经网络的需求。



因此，笔者认为只要不发生社会和媒体过分炒作人工智能，最终导致整个行业陷入低谷的情况正如80年代已经发生过的情况，寒武纪的市场前景是非常值得期待的——寒武纪处理器失败的风险就是社会和媒体过分炒作人工智能，如果最后人工智能的发展速度达不到公众（投资人）预期（这必然会发生，例如现在很多媒体，甚至谷歌自己都发话天网就要造出来），那么整个领域都会陷入大低谷，覆巢之下焉有完卵。

至于将来寒武纪产业化成果几何，还请由时间去检验。

---

标签： DianNaoYu   处理器   寒武纪   指令集   深度学习   神经网络


---

《 上一篇  
中科院提出研发深度学习处理器架构“寒武纪”


下一篇 >>  
Nvidia宣布基于Tegra X1打造的Jetson TX1登陆中国市场 首款针对深度学习神经网络而设计的嵌入式开发板




Uber收购神秘创业公司  
涉足人工智能

 继谷歌、Facebook  
公开其深度学习平台  
后 亚马逊宣布开源  
其机器学习工具  
DSSTNE

继谷歌、Facebook公开  
其深度学习平台后 亚马  
逊宣布开源其机器学习  
工具DSSTNE

 用深度学习检查胸部  
X光照片预测疾病

用深度学习检查胸部X光  
照片预测疾病

 什么是深度学习与机  
器视觉？看这一篇就  
够了

什么是深度学习与机器  
视觉？看这一篇就够了

---

揭秘寒武纪神经网络处理器 深度学习处理器指令集DianNaoYu：等您坐沙发呢！

---

发表评论



昵称

邮箱

网址



<div></div>		
表情	还能输入 <b>240</b> 个字	发布评论

---

[关于我们](#) [联系我们](#) [免责声明](#) [投稿](#)

Copyright © 2014-2017 奥创智能网

备案号：鲁ICP备15012884号-1 . 网站统计