

新智元 (/author/index?authorId=511)

简介：▼

(/aut

hor/

inde

x?au

thor

Id=5

11)

第一个手机端分布式深度学习系统，设计自动化顶会 DATE 最佳论文

2017-03-28 21:51 李静怡、张易

新智元报道

作者：李静怡、张易

【新智元导读】欧洲最大的设计自动化会议 DATE 昨天在瑞士洛桑召开。杜克大学陈怡然教授组关于移动平台深度学习计算的研究荣获嵌入式软件类最佳论文。陈怡然等人提出了本地分布式移动计算系统 MoDNN，让 DNN 能在移动端做并行计算。实验表明，MoDNN 加速 DNN 计算表现良好，展现了 DNN 应用中移动平台的巨大潜力。

【进入新智元公众号，在对话框输入“170328”下载论文】

最新 (/)

团队号 (/channel/index?type=263&name=team)

移动开发 (/channel/index?type=272&name=mobile)

Java (/channel/index?type=257&name=java)

Python (/channel/index?type=268&name=python)

个人号 (/channel/index?type=271&name=person)

前端 (/channel/index?type=264&name=web)

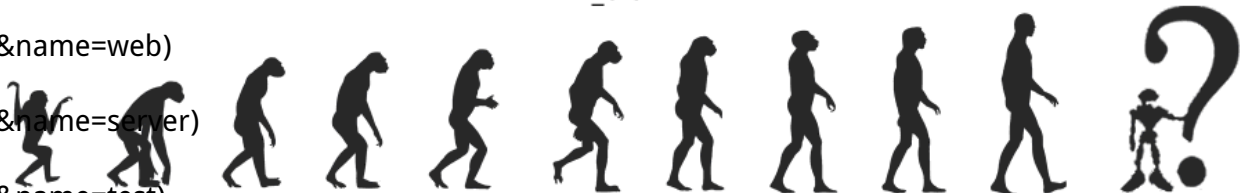
后端 (/channel/index?type=269&name=server)

测试 (/channel/index?type=266&name=test)

更多 (/channel/index?type=270&name=other)



点击右上角
分享文章到朋友圈
欢迎关注公众号
AI_era



DATE——Design, Automation and Test in Europe，是欧洲最大的设计自动化会议。DATE 汇集的人群从研究者、开放商到终端用户，几乎覆盖了整个生态。具体说，设计师和设计自动化的用户、研究人员和供应商，以及电子电路和系统的软硬设计、测试和制造的专家。此外，DATE 非常强调 IC/SoC，可重配置硬件，还有嵌入式系统。

2017 年的 DATE 本月 27 号在瑞士洛桑揭幕。在为期 5 天的会议中，除了常规的主旨演讲、论文发表、讲座和研讨会，还有一个商业展出，展示的内容包括最先进的设计和测试工具、方法、IP 以及设计服务，还有可重构和其他的硬件平台，比如汽车、无线、电信和多媒体应用。可以说，相比其他顶会，DATE 的内容可能与我们的生活关系更为密切。

每届顶会的论文都代表了领域热门研究方向、最新研究方法，其中优秀最佳论文值得关注。2017 年 DATE 评选出了 4 篇最佳论文，点击下载《开发者大会》[下载 \(/download/dev.apk\)](#)

最新 (/) 篇最佳论文，设计 (D Track)、应用 (A Track)、测试 (T Track) 和嵌入式软件 (E Track) 四个分类各一篇。其中，2017 年 DATE 嵌入式软件 E Track 的最佳论文奖被授予了杜克大学陈怡然教授组关于移动平台深度学习计算的文章——《团队号 (/channel/index?type=263&name=team) MoDNN：用于深度神经网络的本地分布式移动计算系统》。

移动开发 (/channel/index?type=272&name=mobile)

Java (/channel/index?type=257&name=java)

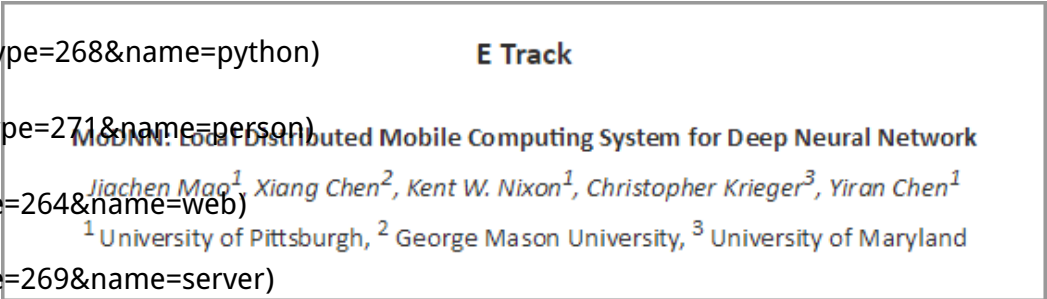
Python (/channel/index?type=268&name=python)

个人号 (/channel/index?type=271&name=person)

前端 (/channel/index?type=264&name=web)

后端 (/channel/index?type=269&name=server)

测试 (/channel/index?type=266&name=test)



*投稿时陈怡然教授尚在匹兹堡大学，今年初已转到杜克大学工作

更多 (/channel/index?type=270&name=other)

在这篇论文中，研究人员提出了一个可以通过无线局域网 (WLAN) 工作的深度神经网络 (DNN) 的本地分布式移动计算系统，叫做 MoDNN。MoDNN 可以通过在多个移动设备之间引入执行的并行性，显著加速 DNN 的计算。

DNN 的应用越来越广，不仅是在大型的数据中心，现在，有些智能手机的功能也强大到能运行某些深度学习，举几个常见的例子，过滤掉麦克风里的杂音，或者删除加速计采集的数据里不必要的信号。

深度学习可以大幅提升手机 App 的性能，让智能手机变得更加“智能”。这一切的基础，就是在移动端也能运行 DNN。

下载《开发者大全》

下载 (/download/dev.apk)

×

最新 (/) 研究人员表示，“据们所知，本文是第一篇利用WLAN中的各种移动设备作为DNN计算资源的论文，在执行并行性增强和
数据传输方面有多项创新”。
团队号 (/channel/index?type=263&name=team)

移动开发 (/channel/index?type=272&name=mobile)
下面我们就来具体看一下陈怡然教授组的这项研究成果。

Java (/channel/index?type=257&name=java)

Python (/channel/index?type=268&name=python)

个人号 (/channel/index?type=271&name=person)

前端 (/channel/index?type=264&name=web)

后端 (/channel/index?type=269&name=server)

测试 (/channel/index?type=266&name=test)

更多 (/channel/index?type=270&name=other)



论文 MoDNN：用于深度神经网络的本地分布式移动计算系统

下载《开发者大全》

下载 (/download/dev.apk)



最新 (/)

团队号 (/channel/index?type=263&name=team)

移动开发 (/channel/index?type=272&name=mobile)

Java (/channel/index?type=257&name=java)

Python (/channel/index?type=268&name=python)

个人号 (/channel/index?type=271&name=person)

前端 (/channel/index?type=264&name=web)

后端 (/channel/index?type=269&name=server)

测试 (/channel/index?type=266&name=test)

MoDNN: Local Distributed Mobile Computing System for Deep Neural Network

Jiachen Mao[†], Xiang Chen[‡], Kent W. Nixon[†], Christopher Krieger[¶], and Yiran Chen[†]

[†]University of Pittsburgh, USA; [‡]George Mason University, USA; [¶]University of Maryland, USA

[†]{jim35, kwn2, yic52}@pitt.edu; [‡]xchen26@gmu.edu; [¶]Krieger@lps.umd.edu

作者：毛驾臣、陈翔、Kent W. Nixon、Christopher Krieger，陈怡然

摘要

虽然深度学习神经网络(DNN)在许多应用中被广泛使用，但是通常难以在资源受限的设备（例如移动平台）上部署DNN。一些现有的尝试主要集中在客户端 - 服务器计算模式或DNN压缩模型中，这需要基础设施的支撑或专门的训练。本文提出了MoDNN——一个用于DNN应用的本地分布式移动计算系统。MoDNN 可以将已训练的DNN模型分割到多个移动设备上，减轻设备级计算成本和内存使用，从而加速DNN计算。我们还设计了两种模型分区方案以最小化非并行数据的传递时间，包括唤醒时间和传输时间。实验结果表明，当工作节点数从2增加到4时，MoDNN可以加速DNN计算2.17-4.28×。除了并行执行之外，性能加速也部分来自于数据传送时间的减少，例如对于传统2D网格分区，减少了30.02%。

背景介绍

移动网络日益增长的带宽激发了移动设备上多媒体应用的大规模增长，这涉及密集的对象识别和分类任务。深度神

最新 (/channel/index?type=263&name=team) 代表性的例子是VGG，它代表了2014年 ImageNet 大规模视觉识别挑战 (ILSVRC14) 的最先进性能。VGG具有15M的神经元、144M的参数和3.4B的连接。当部署在移动设备时，VGG花费大约16秒来完成一个图像的识别过程，这在实际中是不可容忍的。

移动开发 (/channel/index?type=272&name=mobile)

Java (/channel/index?type=257&name=java) 由于巨大的计算量和移动设备的有限计算资源之间的鸿沟对用户体验产生了不利影响，并催生了一些研究来填补这一鸿沟。例如，客户端 - 服务器模式是一种直接的解决方案，可以有效地将高计算成本卸载到外部基础设施上：Hauswald et al.(2014) 提出了一种流水线机器学习结构中的数据卸载方案；Li et al.(2014) 建立了DNN训练的高效分布式参数服务器框架。此外，也有许多研究致力于减少DNN的计算工作量，例如模型压缩：Han et al.(2015)使用三级流水线深度压缩DNN模型：剪枝，受训量化和霍夫曼编码；Chen et al (2015) 引入了低成本哈希函数将权重分组到哈希桶中以用于参数共享。

个人号 (/channel/index?type=271&name=person)

前端 (/channel/index?type=264&name=web)

后端 (/channel/index?type=269&name=server)

测试 (/channel/index?type=266&name=test) 我们发现，有一个重要的场景，在以前的研究中还没有充分考察。那就是，在本地分布式移动计算系统上运行DNN。与由外部基础设施支持单个移动设备的客户端 - 服务器模式相比，本地分布式移动计算系统有几个重要的优势，包括更多的本地计算资源、更高的隐私、对网络带宽更少的依赖等等。

更多 (/channel/index?type=270&name=other)

这篇论文的主要贡献是：

- 1. 调查了使用多个授权的、支持WiFi的、用于DNN计算的移动设备在WLAN中构建计算集群的方法。携带了测试数据（例如图像）的移动设备充当 Group Owner (GO)，其他设备充当工作节点；
- 2. 基于两类DNN层的特性（卷积层和完全连接层）和不同移动设备的计算能力，提出两种分区方案以最小化移动设备之间的数据传送时间；



最新 (/) 3. 在计算集群中的每个移动设备上使用中间件来调度整个执行过程。

团队号 (/channel/index?type=263&name=team)

移动开发 (/channel/index?type=272&name=mobile) 据我们所知，本文是第一篇利用 WLAN 中的各种移动设备作为DNN计算资源的论文，在执行并行性增强和数据传输方面有多项创新”。实验结果表明，当工作节点数从2增加到4时，由于实现了高执行并行性，以及数据传输时间显著减少，M

Java (/channel/index?type=257&name=java) 计算2.17-4.28X。

Python (/channel/index?type=268&name=python)

个人号 (/channel/index?type=260&name=personal) MoDNN的系统框架

前端 (/channel/index?type=264&name=web)

后端 (/channel/index?type=269&name=server) 图1是MoDNN的系统框架概述，包括三个主要组件：

测试 (/channel/index?type=266&name=test) 1) 由GO和多个工作节点形成的本地分布式网络集群；

更多 (/channel/index?type=270&name=other) 2) 将DNN模型分割到工作节点上的模型处理器；

3) 执行DNN的数据传递和识别服务的中间件。

图1

我们注意到卷积层（CL）的计算成本主要取决于其输入大小。因此，我们引入了Biased One-Dimensional Partition (BODP)的方案来划分CL。相反，完全连接层（FL）的内存使用主要由层中的权重数量决定。鉴于此，专门针对稀疏FL引入了由Modified Spectral Co-Clustering（MSCC）和Fine Grain Cross-Partition（FGCP）组成的权重分割方案。值得注意的是，

最新 (/) , 一旦DNN被训练，DNN模型分区只需要在应用程序中执行一次。因此，只要训练的DNN保持相同，分区成本可以由系统的执行来分摊。

团队号 (/channel/index?type=263&name=team)

移动开发 (/channel/index?type=272&name=mobile)

Java (/channel/index?type=257&name=java) 更多关于系统实现和实验设置的技术细节，请查阅论文。

Python (/channel/index?type=268&name=python)

个人号 (/channel/index?type=262&name=person)

前端 (/channel/index?type=264&name=web)

后端 (/channel/index?type=269&name=server) 在这篇论文中，作者提出了 MoDNN 本地分布式移动计算系统，以实现 DNN 在移动平台上的并行计算。由于卷积层和完全连接层被认为是影响总体执行时间的主要DNN组件，因此作者提出了几种高级分区方案，即 BODP、MSCC 和 FGCP，

测试 (/channel/index?type=266&name=test) 以平衡每个工作节点的工作负载，最小化数据传送时间。实验表明，在DNN计算上，MoDNN 比线性性能加速表现更好，展现了DNN应用中移动平台的巨大潜力。

更多 (/channel/index?type=270&name=other)

【进入新智元公众号，在对话框输入“170328”下载论文】

下载《开发者大全》

下载 (/download/dev.apk)



最新 (/)

团队号 (/channel/index?type=263&name=team)

移动开发 (/channel/index?type=272&name=mobile)
3月27日，新智元开源·生态AI技术峰会暨新智元2017创业大赛颁奖盛典隆重召开，包括BAT在内的中国主流AI公司、600多名行业精英齐聚，共同为2017中国人工智能的发展画上了浓墨重彩的一笔。

Java (/channel/index?type=257&name=java)

Python (/channel/index?type=268&name=python)
点击阅读原文，查阅文字版大会实录

个人号 (/channel/index?type=271&name=person)

前端 (/channel/index?type=264&name=web)
2017年4月14日，新智元·生态AI技术峰会暨新智元2017创业大赛颁奖盛典隆重召开，包括BAT在内的中国主流AI公司、600多名行业精英齐聚，共同为2017中国人工智能的发展画上了浓墨重彩的一笔。

后端 (/channel/index?type=269&name=server)
● 腾讯科技：<http://v.qq.com/live/p/topic/26417/preview.html>
测试 (/channel/index?type=266&name=test)
● 阿里云栖社区：<http://yq.aliyun.com/webinar/play/199>
更多 (/channel/index?type=270&name=other)
● 爱奇艺：http://www.iqiyi.com/l_19rrfgal1z.html

分享：

下载《开发者大全》
登录

下载 (/download/dev.apk)



最新 (/)

来说两句吧...

团队号 (/channel/index?type=263&name=team)

移动开发 (/channel/index?type=272&name=mobile)

Java (/channel/index?type=257&name=java)

Python (/channel/index?type=268&name=python)

个人号 (/channel/index?type=271&name=person)

十条网正在使用畅言 (http://changyan.kuaizhan.com/)

前端 (/channel/index?type=264&name=web)

后端 (/channel/index?type=269&name=server)

测试 (/channel/index?type=266&name=test)

更多 (/channel/index?type=270&name=other)



【配色】超美的30例春日色彩搭配，治愈你的小

热评话题

- 离开北京之后，你真能活得更好？ - 腾讯科技 | 十...
- 重磅| 英国研究出仿生皮肤 让机器人触觉超过人类...
- MySQL 数据库增量数据恢复案例 - 数据库开发 | ...
- 全栈工程师都关注了这些 - 数据库开发 | 十条 (htt...

下载《开发者大全》

下载 (/download/devapk)

热点：苹果获得iPhone 6重大胜诉；腾讯成全球...



最新 (/)

小黄车ofo出现大面积网络故障：无法开锁或无法...

团队号 (/channel/index?type=263&name=team)

程序员喝酒解释专业术语 - 数据库开发 | 十条 (htt...

移动开发 (/channel/index?type=272&name=mobile)

Java (/channel/index?type=257&name=java)	新智元 更多文章
Python (/channel/index?type=268&name=python)	【327峰会直击】新智元“春节”引爆AI原力，群星纵论中国AI大棋局 (/html/511/201703/2651995346/1.html)
个人号 (/channel/index?type=271&name=person)	ImageNet 2017启幕，海康威视浦世亮谈2016夺冠绝技及深度学习+安防？ 新智元AI (/html/511/201703/2651995305/1.html)
前端 (/channel/index?type=264&name=web)	【最详尽的GAN介绍】王飞跃等：生成式对抗网络 GAN 的研究进展与展望 (/html/511/201703/2651995256/1.html)
后端 (/channel/index?type=269&name=server)	徐立：1200层神经网络夺冠ImageNet，深度学习越深越好？ 新智元 AI 领军人物专访 (/html/511/201703/2651995230/1.html)
测试 (/channel/index?type=266&name=test)	【重磅】腾讯任命张潼为AI Lab主任，“无”恩达时代BAT人工智能三国杀（投票） (/html/511/201703/2651995179/1.html)
更多 (/channel/index?type=270&name=other)	

猜您喜欢
词云看两会，教你四步知道2016年政府重心在哪里！ (/html/386/201604/2650813008/1.html)
《K8s与云原生应用》之云原生应用的设计理念与挑战 (/html/324/201608/2664334760/1.html)
Open-Falcon发布新版本了 (/html/490/201508/207718773/1.html)
一个万亿级市场的悄悄到来 (/html/637/201608/2649619131/1.html)
Perl[Tk] 实时读入数据到表格 (/html/401/201610/2650169266/1.html)

下载《开发者大全》

下载 (/download/dev.apk)



- 最新10条
- Copyright © 十条网 (http://www.10tiao.com/) | 京ICP备13010217号 (http://www.miibeian.gov.cn/) | 关于十条 (/html/aboutus/aboutus.html) | 开发者大全 (/download/index.html)
- 团队号 (/channel/index?type=263&name=team)
- 移动开发 (/channel/index?type=272&name=mobile)
- Java (/channel/index?type=257&name=java)
- Python (/channel/index?type=268&name=python)
- 个人号 (/channel/index?type=271&name=person)
- 前端 (/channel/index?type=264&name=web)
- 后端 (/channel/index?type=269&name=server)
- 测试 (/channel/index?type=266&name=test)
- | 更多 (/channel/index?type=270&name=other)

下载《开发者大全》

下载 (/download/dev.apk)

