

边缘计算不再“边缘”

■本报记者 赵广立

“边缘计算”火了。近日,主营为电网提供边缘计算服务的友讯达公司因在互动易平台回复了投资者关于边缘计算的提问,公司股价连续4个交易日大涨;而以网宿科技为代表的相关概念股,短短数日内股价几乎翻番,火爆程度可见一斑。

作为5G衍生概念,边缘计算究竟有何魔力?

“5G商用时代来临,数据量将更加巨大、复杂,对计算提出更高要求,同时也为发展人工智能、边缘计算带来了新机遇。”浪潮服务器产品部副总经理陈彦灵告诉《中国科学报》,如果说3G、4G是云计算的时代,那么5G就是边缘计算时代,“因为5G通信网络更加去中心化,需要在网络边缘部署小规模或者便携式数据中心,以满足超高可靠、超低时延通信的需求”。

分久必合 合久必分

据市场研究机构IDC预测,全球物联网终端设备安装数量有望在2019年达到256亿台,2020年将超过500亿的终端与设备联网。届时边缘计算市场规模将超万亿元,成为与云计算平分秋色的新兴市场。

关于云计算与边缘计算,有个形象的例子:云计算就像章鱼的大脑,边缘计算就类似于章鱼的那些小爪子,每个爪子就是一个小型的机房,并靠近具体的实物。简单来说,边缘计算的主要价值在于减少超大流量对网络的消耗,同时为各种应用提供低延时和开放的网络能力。

人工智能创业公司地平线创始人兼CEO余凯,一直是边缘计算的拥趸。“地平线从创立的第一天开始,就把边缘计算作为自己的核心使命。”余凯在接受《中国科学报》采访时说,“边缘计算会覆盖生活的每一天。”

余凯之所以把边缘计算看得这么重要,是看到它确实会让城市生活变得更加安全、更加美好。“城市是人工智能重要的场景,在这个场景中有交通、有驾驶、有大量的人机交互,意味着城市是数据大量产生和爆发的地方,而计算一定是发生在数据爆发的场景里。”

余凯说,无论是在家、在路上,还是在工作场景中,大量的场景下会产生大量的数据。信息技术研究与顾问公司Gartner预计,到2025年,全球数据量将以10倍的速度剧增至163ZB,这将远超过任何互联网公司现在所储备的数据量。

“分久必合,合久必分,这在计算上也在发生。”余凯说,在PC时代,所有的计算都是在电脑上,是本地边缘计算,随着互联网的发展,数据越来越往云端,数据中心迁移,计算也慢慢转移到了云端,即云计算;而从移动互联网到物联网,则是另外一个趋势——更多计算从中央向边缘迁移,这要求计算的实时性和可靠性,



未来边缘计算和中心计算在市场份额方面将平分秋色

5G商用时代来临,数据量将更加巨大、复杂,对计算提出更高要求,同时也为发展人工智能、边缘计算带来了新机遇。因为5G通信网络更加去中心化,需要在网络边缘部署小规模或者便携式数据中心,以满足超高可靠、超低时延通信的需求。

以及设备在离线的环境下还可以正常工作。“比如自动驾驶汽车过隧道的时候,即便没有信号也能够正常行驶,这一定要靠不依赖网络和云端的边缘计算。”

“我认为汽车未来就是‘四个轮子上的数据中心’。”余凯说,“自动驾驶对延迟要求很高,需要边缘计算,而且自动驾驶每升高一个级别,算力的需求几乎呈指数级升高。”

此外,余凯说,很多安防类企业,现在靠WiFi联网处理的计算方式容易受信号和带宽的影响,因此也需要边缘计算。“还有很多因素如安全性、性价比等,最核心的是,大量的计算要在边缘,而不是在数据中心。”

与原有计算业态互补

在陈彦灵看来,随着5G商用的日益临近,实时、智能、安全、隐私这四大趋势催生了边缘计算与端智能的崛起,而5G通信的超低时延与超高可靠要求,使得边缘计算成为必然选择。不过,在5G商用场景驱动下,边缘计算平台建设是一个“扩展项”,它将云计算平台、大数据平台等共同组成城市发展的数字基础设施。业界的共识是,现代IT之所以具有更高效率,云计算的集中处理思维居功至伟。尤其是数据量暴增的当下,集中处理

数据、全局分析、智能调度是挖掘数据价值的一般准则。也就是说,即便是边缘计算崛起,云计算作为计算的中心位置,也不会消亡。

对此,中国信息通信研究院云计算与大数据研究所所长何宝宏认为,边缘计算与中心化的计算,甚至终端的计算是要相互协作的,目的是为了配合通信(如移动边缘计算)、存储(如内容分发网络)、安全(如防火墙)等应用,因此很可能纵向发展成为原有产业的补充。

“再直白点,边缘计算很可能会形成一些新产品,但(暂时)不可能形成一个新行业。”何宝宏说。

数字通信技术专家、中国工程院院士邬贺铨也是倡议推进边缘计算技术与产业繁荣发展的专家之一,但他更多倡导对边缘计算概念及其内涵有更为理性、清醒的认识。

“边缘计算绝不是单一的计算能力。”邬贺铨举例说,通信基站的边缘计算首先要进行大数据的采集,同时实现加密,在对大数据进行消化后,再回送网络的数据链应该是减少的,也就意味着对数据进行过滤,而为了过滤,它可能需要缓存。因此,边缘计算会跟缓存数据采集、数据分析、过滤、计算能力集中在一起。一位服务器行业从业者告诉记者,一

般而言,边缘计算处理简单的数据,云计算处理相对复杂的信息,两者配合才能够达到更高的效率和可靠性,而边缘计算更准确地说,应该是对云计算的一种补充和优化。

挑战犹存 忌炒作过甚

任何一项技术的发展都应该是渐进式而不是“火箭式”,更何况边缘计算所面向的场景——5G和物联网的爆发还没有真正到来。因此,对于近期边缘计算概念的火热,何宝宏认为,边缘计算只是一种解决方案,还没有创造出新的核心技术或做重大的技术组合,也没有诞生新的应用,更没有创造出新的商业模式。“近期对边缘计算的炒作,太过分了”。

此外,何宝宏还指出,概念热炒之下,有许多“碰瓷”的、蹭热点的,这对技术的发展不利,他在其个人微博上不客气地写道:“边缘计算就是个筐,要想化装成为边缘计算,把自己‘边缘化’还是很容易的。只要不是明显位于核心位置的技术,比如内容分发网络、物联网、无人驾驶、车联网、基站和VR/AR,甚至卖PC、智能手机和机顶盒的,都可以自称为边缘计算;如果不巧你拥有的技术明明白白不在边缘位置,也一点不用担心,去编个新故事还是很容易的,比如简单地加上边缘两个字,像边缘服务器、边缘接入、边缘路由器、边缘网络、边缘数据中心、边缘云、边缘智能、边缘区块链节点等。”

在邬贺铨看来,边缘计算技术应该是一个体系,但目前包括他本人对这个体系的了解都“只是边缘,根本没有进入到核心,很多东西不太清楚”。因此,他提出了“边缘计算十问”,希望找到答案。

比如,邬贺铨写道:5G移动边缘计算(MEC)应该下沉到什么位置(越靠下反应越快,数量越多)?一个云计算节点连接的移动边缘计算(或雾计算)节点的合理数量是多少?计算能力如何在云计算与边缘计算间优化配置? MEC间通过云计算节点互通还是直接通信?同样是边缘计算节点,它们之间要不要互通,或者说必须间接互通?

在此背景下,余凯认为,软硬件结合在边缘计算领域机会更大。他说,未来的终端对于效率更敏感,这需要软硬件更协同。此外,余凯告诉记者,边缘计算实时性、低延迟的要求其实很高,它对算力的要求一点都不低。因此,基于边缘的人工智能处理器是未来科技竞争的主战场,是一个制高点。

“现在边缘计算占比还很小,互联网巨头都是以云计算作为出发点和根据地。从市场的重要性和空白点两个角度来看,我们选择边缘计算这个市场定位。”余凯认为,未来边缘计算和中心计算在市场份额方面将平分秋色,“大数据训练在中心计算,而数据计算执行在边缘计算,两者是相互配合的关系”。

基于单帧图像实现的。由于对输入图像的要求低,3D重建技术可以利用普通光学摄像头作为感知设备,进而将推动更多智能应用成为可能。比如,该技术将使美颜类App无需结构光摄像头也能具备高精度的瘦身与动画成功功能等。

不过,吴凡也提出,基于单帧图像的3D重建技术对原始图像要求降低的同时,却对背后的技术提出了更高挑战,比如技术需要从单帧图像中推理出人体或人脸的3D形态,并通过光学透视、阴影叠加等基本光学原理准确预测出各个关键点在3D空间的位置和朝向,才能得到人体的姿态或表情信息。

合等优势。

在视觉零售领域,英特尔的视觉解决方案能够使用图像识别技术帮助客户轻松实现自助结算。例如京东边缘知识系统所采用的英特尔边缘计算服务器,能够支持多个摄像头以每秒150帧的速度处理图像,相比重力感应技术,这套系统的识别准确度更高,且能够支持更多的库存,在提供更好购物体验的同时降低运维成本。

此外,英特尔还发布了专注于加速深度学习并将视觉数据转换为业务洞察的OpenVINO工具包,以充分帮助企业边缘侧快速实现高性能计算机视觉与深度学习的开发,进而有效帮助零售企业削减部署成本,简化运营、管理库存、优化供应链和增强推销能力。

前沿扫描



黄钊金提出的“打分”新算法

图像分割出新方法

近日,华中科技大学硕士生黄钊金作为第一作者完成研究成果 Mask Scoring R-CNN,在COCO图像实例分割任务上超越了何恺明(首获CVPR“最佳论文奖”的中国学者)的Mask R-CNN,拿下了计算机视觉顶会CVPR 2019的口头报告。该成果从5000多篇投稿中脱颖而出,成为最顶尖的5.6%。

黄钊金师从华中科技大学电信学院副教授王兴刚,王兴刚也是这篇论文的作者之一。目前,相关算法已经开源。

什么是Mask R-CNN?它其实是一种简洁、灵活的实例分割框架。2017年,计算机研究者何恺明凭借Mask R-CNN的研究成果拿下ICCV 2017最佳论文奖。

与何恺明的研究成果相比,黄钊金的Mask Scoring R-CNN在性能上有什么优越性呢?在论文中,研究人员提出了一种给算法的“实例分割假设”打分的新方法。这个分数打得是否准确,会影响实例分割模型的性能。

这些模型在实例分割任务里,虽然输出结果是一个蒙版,但打分却是和边界框目标检测共享的,都是针对目标区域分类置信度算出来的分数。这个分数和图像分割蒙版的质量未必一致,用来评价蒙版的质量,可能就会出偏差。

于是,这篇CVPR 2019论

文就提出了一种新的打分方法——给蒙版打分,被称之为蒙版得分(Mask Score)。

Mask Scoring R-CNN中提出的计分方式很简单:不仅仅直接依靠检测得到的分类算分,而且还让模型单独学一个针对蒙版的得分规则,MaskIoU head。

MaskIoU head是在经典评估指标AP(平均正确率)启发下得到的,会拿预测蒙版与物体特征进行对比。MaskIoU head同时接收蒙版head的输出与ROI的特征(Region of Interest)作为输入,用一种简单的回归损失进行训练。

最后,同时考虑分类得分与蒙版的质量得分,就可以评估出质量了。评测方法公平公正,实例分割模型性能自然也就上去了。

实验证明,在挑战COCO benchmark时,在用MS R-CNN的蒙版得分评估时,在不同基干网络上,AP始终提升近15%。

在COCO 2017测试集(Test-Dev set)显示的成绩对比中,无论基干网络是纯粹的ResNet-101,还是用了DCN、FPN、MS R-CNN的AP成绩都比Mask R-CNN高出一点几个百分点。在COCO 2017验证集上,MS R-CNN的得分也优于Mask R-CNN。(赵利利)

相关论文信息: <https://arxiv.org/abs/1803.01534>

速递

北京工业互联网技术创新与产业发展联盟成立

本报讯 为实施《北京工业互联网发展行动计划(2018-2020年)》,加快推动北京市工业互联网创新发展,3月17日,由清华大学牵头推动的北京工业互联网技术创新与产业发展联盟在京成立。

该联盟成员涵盖企业、高校、科研机构等多家单位。中国工程院院士、清华大学副校长尤政表示,清华大学后续将充分发挥在人才、教育和科研方面的优势,从工业互联网共性技术开源创新和工业互联网技术应用等多方面发力,为联盟的建设和北京市工业互联网产业的发展提供相关的支持。

2018年,工业互联网创新发展工程支持项目中近一半由北京市承担。北京还是我国统一建设的5个工业互联网标识解析国家顶

级节点之一。工业和信息化部信息和软件服务业司巡视员李颖指出,成立北京工业互联网产业联盟有利于催生北京经济增长新引擎,打造京津冀工业互联网协同发展新局面,成为引领全国产业升级的新典范。

“国家工业互联网产业联盟会员单位有三成来自北京,北京工业互联网发展资源好,优势强。北京工业互联网产业联盟的成立不仅为北京,也将为我国工业互联网产业发展提供重要支撑。”中国信息通信研究院总工程师余晓晖在致辞中表示,国家工业互联网产业联盟和中国信息通信研究院后续将大力支持北京工业互联网产业联盟的建设,共同推进北京市和全国制造业的转型升级。(李惠钰)

科大讯飞在韩国建首个合资公司

本报讯 3月18日,科大讯飞与韩国知名企业HANCOM集团在合肥举行签约仪式,就双方成立合资子公司Accuflly.AI进行签约。科大讯飞智能服务BU总理解飞在致辞中表示,这是科大讯飞在韩国成立的首个合资公司,此次签约合作“是科大讯飞进行海外合作一个重要的里程碑”。

根据协议,此次双方共建的合资子公司Accuflly.AI,将在双方集团平等互信基础上,致力于开展构建海外全方位的人工智能(AI)合

作系统,辅助科大讯飞AI生态系统的全球化战略。该公司将在智能办公、翻译机、金融呼叫中心等领域进行拓展,同时会根据韩国市场具体需求,发挥合作伙伴优势,逐步开拓教育、医疗、汽车以及快速实现本地化销售的硬件设备等业务。

据介绍,Accuflly.AI公司将设置产品部门、研发部门和业务管理部门,并实行共同代表制,以“共同打造更多面向韩国甚至走向世界的通用性产品”。(赵广立)

云从科技3D人体重建技术获突破

本报讯 (记者赵广立)3月19日,记者从云从科技获悉,这家中科院旗下人工智能(AI)创企的3D人体重建技术同时在该领域三大权威数据集上取得新突破:将原有最低“误差”大幅降低30%,刷新了世界纪录。

3D重建领域通常以“误差”作为衡量算法能力的主要指标,这里的误差特指生成模型与实际图像的差别。显然,误差越低,精度越高,意味着技术性能越好。

据介绍,在该领域三大权威数据集Human3.6M、Surreal和UP-3D上,云从科技3D人体重建技术“全身精度误差”在Surreal上从75.4毫米降低到52.7毫

米,“关节精度误差”从55.8毫米降低到40.1毫米,在Human3.6M上的关节精度误差从59.9毫米降低到46.7毫米。此外,技术的执行速度从之前的上百毫秒降低到只需5毫秒。

上述突破是如何实现的?云从科技技术人员团队、资深算法研究员吴凡介绍,针对人体具有丰富多样的姿态和穿着的特点,他们提出了一套全新的基于人体3D纵深预测的3D信息表征方式。这种表征方式通过对三原色图像(RGB,不含深度信息)的分析来预测人体的3D形态和姿势,并用6万多个点完整描绘人体,从而在人体重建技术上取得速度与精度

的双突破。

“(新算法下的)结果呈现出来的模型更精细,帧率更是高达200帧每秒,原本受实时显示限制而无法实现的应用可以——实现。”吴凡说,这将极大地推动相关智能图像应用的落地。

记者了解到,传统的3D重建技术大多需要连续的或是多视角的图像,在硬件上一般需采用双目摄像机或结构光摄像机等设备,因此在手机等便携设备上尚难以实现;另一方面,专用设备还会增加部署成本,增加大规模普及3D重建技术的难度。

云从科技这一3D人体重建技术是

英特尔工作负载整合策略助推智能零售落地

本报讯 (记者李惠钰)近日,2019中国智慧零售数字化高峰论坛在苏州召开。英特尔同期举办了2019英特尔智能POS产品技术研讨会,集中展示了包括芯片、机器视觉、人工智能等在内的一系列产品和解决方案,并分享了英特尔通过工作负载整合策略携手合作伙伴打造数字化智慧门店的构想。

当前,零售领域迎来了设备爆发式的增长,线下零售店中的智能设备纷繁复杂,如果每一个设备都部署一个服务器,无形中提升了成本且不利于管理。因此,英特尔提供整体解决方案进行工作负载整合,把小型边缘计算集成到中央服务器,从而降低服务器成本,提升计算效率。如此一来,最大程度地惠及了零售商和消费者,让产业链中的每一个环节都能够提

升效率并且优化体验,进而满足用户的个性化需求。

所谓工作负载整合,就是把原来在不同设备上分立的负载,通过虚拟化等技术,整合到一个单一的高性能的计算平台上,从而实现综合、复杂的功能。各个功能子系统既能分享设备提供的计算、存储、网络等资源,同时还能具有一定的独立性,以避免彼此的相互影响。

“零售行业技术应用方面有非常众多的子系统,甚至包括不同的单片机的方案,我们为什么不把它们整合到POS机里面去呢?”在英特尔物联网销售市场部中国区总监谢青山看来,机器视觉、人工智能等方面的应用,POS以其计算力能够成为一个很好的负载整合的落脚点。

“负载整合的概念就是把工作负载放在边缘计算的一个平台上,于零售商来说,POS就是个很好的载体,就是一个边缘计算产品的形态。当然,可能还需要针对现在POS机形态、功能、性能等需要实现综合、复杂的功能。各个功能子系统既能分享设备提供的计算、存储、网络等资源,同时还能具有一定的独立性,以避免彼此的相互影响。”

为了应对当前零售应用场景和需求的升级,英特尔还推出了一系列针对零售场景的产品和技术,为零售企业搭建拥有丰富场景的“云—网—端”整体解决方案。针对支付系统,英特尔SPOS能够兼容英特尔低功耗酷睿处理器U和凌动系列,低功耗的紧凑设计可提供适合各种应用和细分市场的极具竞争力的高效解决方案,有效降低设计和开发成本,以满足众多尖端外形的要求,还具备工作负载整

觉、深度学习和TensorFlow框架等研究工作。

贾扬清本科和硕士毕业于清华大学,后在美国加州大学伯克利分校获得计算机科学博士学位。

贾扬清现在刚满35岁,是全世界最受关注的AI科学家之一。攻读博士期间,贾扬清创立并开源了深度学习框架Caffe,同时他也是TensorFlow核心作者之一。(赵鲁)