

AADS:让仿真技术告别“五毛特效”

■本报记者 王之康 通讯员 吕文倩 刘天佑 任博文

近年来,自动驾驶越来越多地为人们熟知和重视,在北京之外,上海、深圳等城市也在积极为自动驾驶车辆开辟测试道路。就在最近,刚刚浮出水面的京雄高速设计方案显示,将设自动和半自动驾驶专用车道。

一方面,自动驾驶的热潮愈演愈烈;另一方面,其安全性仍然为人们所担忧——尽管不久前,美国检方认定优步(Uber)在一年前因自动驾驶测试而发生的交通事故中并不承担刑事责任,但减少和避免自动驾驶汽车发生交通事故一直都是该领域的刚性需求。

相较于花费十几年甚至更久时间让自动驾驶车辆接受一定数量的道路测试,通过仿真系统测试来评估、提高其安全性极具可操作性和现实意义。近日,李伟等7名来自南京航空航天大学、北京大学、浙江大学、香港大学等高校的在读博士生研发的增强自动驾驶仿真系统(AADS)就突破了现有仿真技术存在的许多障碍,将这一手段变得更加实用。目前,相关论文《基于数据驱动算法的增强自动驾驶仿真》(AADS: Augmented autonomous driving simulation using data-driven algorithms)已发表在《科学》杂志子刊《科学·机器人学》上。

一种有效方式

所谓仿真系统,就是使用项目模型将特定于某一具体层次的不确定性转化为其对目标影响的一种系统。目前,此类系统在执法、防御和医疗训练等领域较为常见。

回到自动驾驶领域,“仿真系统不仅可以测试和验证自动驾驶汽车在环境感知、导航和控制方面的性能,还能生成大量有标签的训练数据以训练深度神经网络等机器学习方法。”在接受《中国科学报》采访时,南京航空航天大学自动化学院博士生李伟表示。

当前,不少公司开发了仿真系统,比如英特尔的 CARLA、微软的 AirSim、英伟达的 Drive Constellation 等,它们主要是利用游戏引擎或高保真计算机图形(CG)模型来创建驾驶场景,即结合计算机图形、物理建模和机器人运动规划技术来构建一个合成环境,然后在环境中对运动的车辆进行动画制作和渲染。

不过,这些仿真系统虽然取得了不错的合成渲染结果,却存在很大局限。一方面,由于对高保真环境模型的需求,创建逼真 CG 模型的成本非常高;另一方面,汽车、行人等运动障碍的动画制作较为生硬,缺乏真实场景的灵活性,而且系统无法生成城市环境中的诸多不同场



李伟

“所谓仿真系统,就是使用项目模型将特定于某一具体层次的不确定性转化为其对目标影响的一种系统。”

景,丰富性不够。

“简单来说,利用现有技术开发的仿真系统,最大的问题就是不够真实。”李伟解释道,“因为它是虚拟出来的,所以效果跟真实路测有很大的差距。”

《基于数据驱动算法的增强自动驾驶仿真》就致力于解决这些问题。

最大亮点是真实

“该研究成果介绍了一种增强现实的自动驾驶仿真系统,通过模拟交通流来增强现实世界图像,进而创建逼真的、媲美现实世界渲染的仿真场景,为自动驾驶车辆提供更可靠且廉价的实验室模拟方法,可大规模用于训练和测试评估自动驾驶系统的规划决策算法。”对李伟的这项研究,他的导师、南京航空航天大学自动化学院教授龚华军予以肯定。

李伟向《中国科学报》介绍说,逼真的道路场景来自于实采,即使用激光雷

达(LiDAR)和相机扫描街景,然后根据获得的轨迹数据,为汽车、行人等生成交通流,并将其合成到背景中;同时,合成图像还可以基于不同视角和传感器模型(LiDAR 或相机)进行再合成,最后应用于自动驾驶测试。

“自动驾驶汽车依赖于感知模块和导航模块,感知模块接收并解释现实世界的信息,导航模块根据感知的信息作出决定,比如转向哪里、是否刹车或加速等。”美国肯塔基大学教授、李伟在美国联合培养期间的导师杨睿刚说,不同于现有的仿真系统,AADS 能更准确、更真实地模拟自动驾驶汽车在路上接收到的信息,在感知层面更接近于真实世界。

“打个比方,如果说一些厂商做的仿真系统看上去像动画,那么 AADS 就是真人 COSPLAY。”李伟举例说道,“该研究成果之所以受学术界认可,就是因为其真实性,同时这也是它的最大亮点。”

他表示,仿真与实际场景之间必然存

在一定的差异,而他所做的,就是尽量缩小这个差异。“如果在一个不够真实的环境中通过了模拟测试,实际应用时仍然可能遇到各种问题;如果在一个非常接近真实的环境中通过了模拟测试,实际应用效果就会好很多。”李伟说,越接近真实,测试就越有效。

弥补与现实环境的差异

在杨睿刚看来,这项研究真是一种突破性的仿真技术,将更精确、更真实地测试自动驾驶系统,极大地加快自动驾驶技术的成熟与落地。

实际上,李伟及研究团队也曾面临现有仿真系统在呈现真实交通流、驾驶员与行人的互动等场景中存在诸多不足的问题。

“最大的难题,同样是如何将道路的真实情况进行仿真合成。”他说,现有的技术一般都是做室内场景图像合成,直接拓展到做室外场景图像合成并不能取得预期效果。

于是,他们就专门做了一种针对室外仿真系统图像合成的技术,仅用少量图像就可以实现视图插值和 extrapolation。“与之前的方法相比,它可以生成伪像素更少的高质量图像。”

“我们这个系统其实就是在弥补与现实环境的差异,现在我们的研究应该是最接近真实的。”李伟说,“目前学术界和工业界尚未有特别针对自动驾驶的街景合成技术,而我们的研究则在这一点上完成了突破。”

据了解,该研究目前已经部分应用在自动驾驶系统的仿真、验证环节。

“当然,提升更多传感器的仿真效果(如 LiDAR),增加仿真场景的多样性、仿真更多现有自动驾驶系统容易失效的场景,能够进一步提升该系统的仿真能力。”杨睿刚说,同时,加入并行测试的功能,才能让该仿真技术适应企业级的大规模测试场景。

对此,李伟有更加具体的打算。

“我们不仅希望能够做一篇比较好的论文,更希望把研究成果应用到更多地方。在此之前,还有很多研究工作要做。”他说,仿真系统本身是比较庞大的,包括图像仿真、激光雷达仿真以及车流移动仿真等,还需要一些更深的技术将整个系统加以完善。“比如,现在的合成基于采集的图像,但采集的图像是固定的,所以我们需要通过一些方法,实现白天合成晚上、晴天合成阴天,只有这样,驾驶系统才能仿真各种各样的场景,提高测试覆盖面,并减轻前期采集的工作量。”

速递

英特尔宣布退出 5G 智能手机调制解调器业务

本报讯 美国当地时间 4 月 16 日,英特尔公司宣布计划退出 5G 智能手机调制解调器业务,并完成对其他调制解调器业务机会的评估,包括 PC、物联网设备及其他以数据为中心的设备。不过,英特尔表示“还将继续投资其 5G 网络基础设施业务”。

英特尔表示,他们将继续履行对现有 4G 智能手机调制解调器产品线的客户承诺,但不会在智能手机领域推出 5G 调制解调器产品,包括最初计划于 2020 年推出的产品。

“我们对于 5G 和网络‘云化’的机遇感到兴奋,但对于智能手机调制解调器业务而言,显然已经没有明确

的盈利和获取回报的路径。”英特尔公司首席执行官司睿博(Bob Swan)明确表示,“我们正在评估我们的选择,让创造的价值得以实现,包括在一个 5G 世界中广泛的、以数据为中心的平台和设备的机遇。”

这一消息是在苹果与高通宣布达成和解协议后几小时发布的。两大科技巨头已经就专利授权等问题达成了全面和解,此举也意味着苹果未来 5G 手机可能采用高通生产的基带芯片,而非英特尔。

英特尔预计在即将于 4 月 25 日公布的 2019 年第一季度财报及电话会议上披露更多信息。(赵广立)

“1+1+2+N”模式开启海淀“城市大脑”

本报讯 近日,海淀(中关村科学城)“城市大脑”发布会在京召开。会议发布了海淀“城市大脑”顶层设计规划纲要,并成立海淀区城市大脑科技产业联盟。

海淀“城市大脑”建设总体架构为“1+1+2+N”模式,即:一张感知网、一个智能云平台、两个中心(大数据中心、AI 计算处理中心)、N 个创新应用(前期聚焦 5 个示范应用)。以遍布海淀全域的城市感知网络为硬件基础,以城市大数据为核心资源,以物联网、云计算、大数据、

人工智能为关键技术,以政府主导、多元参与、共建共享为机制保障,对海淀全域进行全感知、全互联、全分析、全响应、全应用。

第四范式创始人、CEO 戴文渊表示:“作为首批联盟企业,我们希望把 AI 从感知阶段升级到认知决策的阶段。当然这还不是终点,有了认知,AI 所做的决策也不一定都是对的,需要形成闭环,让机器知道错误、实现自我学习,这样就可以进一步迭代,不断进化认知引擎。”(田瑞颖)

全球首个人脸断点导航智能停车场在广州落地

本报讯 近日,人工智能整体解决方案企业 Aibee 发布 AI-BeeParking 智能停车场解决方案。基于 AI-Bee Parking 解决方案打造的全球首个人脸断点导航智能停车场已在广州 K11 正式启用。

通过“人脸断点导航”智慧屏幕,顾客在停车场寻车时,只需看一眼经过的屏幕,就能立即获得基于当前位置的实景导航服务,轻松寻车。除“寻车”服务,它还能解决顾客到达停车场后,找不到目的地的问题。未来还将上线车位引导、刷脸缴费、增值服务等功能,提升顾客从到场到离场全流程的智慧体验。(田瑞颖)

相比传统技术方案,AI-BeeParking 解决方案大大减少了摄像头相关硬件投入,平均一个摄像头可管理车位是过去的 5~10 倍,降低了停车场建设与运营成本。

Aibee 由百度研究院前院长林元庆创立,成立一年多,就已在中美两地拥有全职技术人员 200 余人。在成立之初,Aibee 就明确了“精准零售”的战略路线,并迅速与 K11、红星美凯龙等线下零售头部企业达成紧密合作。人脸断点导航智能停车场的成功落地应用,正是 Aibee 与 K11 携手树立线下零售数字化“新标杆”的力作。(田瑞颖)

学术经纬

量化神经处理芯片是边缘计算的答案吗

■程健

边缘计算随着人工智能和物联网的高速发展,已经成为科技界、学术界、商界竞相追捧的热门领域。特别是随着智能驾驶、智能安防、智慧商业等行业对终端 AI 技术能力的要求越来越高,就更需要背后的深度神经网络在算法层面、应用层面不断创新。

边缘计算是大趋势

深度神经网络的训练与应用只能在高性能的计算机集群上进行,为此各大科技巨头都搭建了高性能深度学习平台来处理海量数据。所以最初人们对人工智能和物联网的规划是,大量的数据在云端的数据中心被分析并决策后,再将结果发还给终端。也就是说,云端负责人工智能的部分,边缘端负责数据采集及执行。

但在实际应用中,在云端做智能处理的设想遇到很多问题。

首先,数据传输成本高。随着数据量激增传输带宽压力也不断增加,边缘端的无线传输模块也必须支持高速无线传输,这就需要更大的功耗,这与边缘端低功耗的需求是冲突的。

其次,无法接受网络延时。许多终端应用场景对延时非常敏感,比如自动驾驶、智能安防等场景下的网络传输带来的延时(几十毫秒、几秒甚至掉线)是无法接受的。

还有,工业物联网中对于安全和隐私的需求,也使得云计算在物联网时代的发展受限。边缘计算可以保护隐私,并且做到自主操作,无需依赖网络连接情况。

边缘计算可以很好地解决这些问题,根据实际应用需求,边缘端既可以独立作为智能处理模块,也可以和云端配合(边缘端做一些分析处理和过滤,再交给云端)。这样的解决方案在延时、带宽和功耗等方面都有明显优势。因此,边缘计算是智能化发展的必经之路。

边缘计算的挑战和应对措施

边缘计算在实际应用终端需要深度神经网络满足存储占用小、网络精度高、计算速度快、电量消耗低、低延时等诸多条件才能真正发挥作用,因此,边缘计算的技术核心是高算力、低成本、体积小、低功耗、低延时。

高算力容易理解,随着深度神经网络模型变得越来越复杂,其参数和运行时间也相应增加(以 AlexNet 为例,该神经网络包含超过六千万个参数,对于单幅图像进行分类的过程就涉及多达 7.3 亿次浮点运算)。出于成本和体积的考虑,边缘端不能加 DRAM(动态随机存储);要实现低功耗、低延时,则需要考虑内存内计算——这样既不需要 DRAM,也可以节省内存存取的外开开销,但同样首先需要减少网络计算量。

由此可知,如何解决深度神经网络的加速与压缩,是边缘计算实现落地的核心问题之一。

同时,如何在网络精度保持不变或略有下降的前提下减少网络的计算量和存储,从而实现更加有效的加速压缩,也是存在着许多困难和挑战。

有些压缩方法,虽然能降低存储,但由于其特殊存储方式,需要解码操作后才能用于网络计算,该类压缩方法只能降低网络对磁盘的占用,不能减少网络对内存的消耗,因此不能降低功耗,也没有真正做到加速。

对于目前基于 CPU 的计算平台,许多加速方法虽然能达到很高理论加速比,但实际加速效果远低于理论。以网络稀疏的加速方法为例,由于大部分元素为零,因此理论上有效计算量小,但由于无规则稀疏,非零元素在内存不连续,因此仍然需要频繁访问内存。

除此之外,内存内计算面临的挑战是,目前内存内计算的实现方案本质上是做模拟计算,因此计算精度有限,浮点

32 位的高精度(神经网络普遍使用全精度)是无法实现的。

定点量化计算方式的明显优势

面对以上挑战,定点量化的方式表现出明显优势,体现在三个方面。一、网络在训练好后,实际应用中并不需要全精度,而量化恰是用更低的数值精度实现神经网络的加速压缩;二、量化的方式有多种,但由于定点量化在硬件如 FPGA 上能够取得很高的运行效率,因此更为有效;三、使用比特位数越低,加速效果越明显。

然而,目前常用的定点量化方法,尤其是当量化比特数非常低的时候,如 1 比特或者 2 比特量化,很难达到与全精度(FP32)相近的性能,这也是量化技术发展过程中所面临的挑战。

那么,中国科学院自动化研究所南京人工智能芯片创新研究院(AiRiA 研究院)的量化神经处理芯片是如何解决以上问题的?

中科院自动化研究所基于十几年对量化处理技术的积累,形成了独具优势的量化模型压缩处理技术,可以极大简化整个计算过程,且与全精度性能结果对比,几乎没有精度损失。

基于中科院自动化所的量化处理技术,AiRiA 研究院开发出量化神经处理器 QNPU 架构,并实现了支持 4 比特、2 比特甚至任意 1 比特的量化处理技术,达到国际领先水平。基于 QNPU 架构,AiRiA 研究院的芯片团队自主开发了适合多种应用场景的量化神经处理芯片 Watt-A1,将 QNPU 的优势发挥到极致。

在 QNPU 架构下,大规模神经网络可实现在片上计算,从而减少甚至无需访问外部存储器,解决了需要极大功耗、带宽和体积成本的“内存墙”难题,使得带宽和体积成本的“内存墙”难题,使得在满足低成本、低功耗、低延时的条



程健

如何结合各企业实际及产品的特性,与客户协同创新,探索具体的解决方案,最终为企业、为社会创造价值,才是边缘计算的答案。

件下,仍然可以保持高算力,且精度几乎没有损失。

边缘计算是未来的重要趋势,能很好地应对这些技术挑战,是关键的第一步。而如何结合各企业实际及产品的特性,与客户协同创新,探索具体的解决方案,最终为企业、为社会创造价值,才是边缘计算的答案。

(作者系 AiRiA 研究院常务副院长,本报记者赵广立整理)

360 车联网发布汽车安全大脑

本报讯 4 月 15 日,360 车联网战略新品发布会在北京召开。360 车联网总裁邓邱伟提出汽车安全大脑概念,并推出 360 汽车安全大脑,旨在将技术运用到车载终端,使车载终端具备链接智能,形成安全闭环从而减少风险产生。

会上,360 还发布了两款终端新品——360 流媒体后视镜记录仪 M320 和 360 儿童安全座椅。其中,流媒体后视镜记录仪 M320 是对抗“视觉盲区”的重要产品。它搭载 9.35 英寸高清液晶大屏,同时拥有前置影像 1440P,可以更大视角消除盲区,完成多

种天气和光照条件下后路多车道的清晰展示。

360 儿童安全座椅配置了 ISOFIX 接口、五点式安全带、F1 赛车供应商的钛合金扣,高强度钢骨架、记忆棉材料等丰厚的用料。此外,该座椅可连接手机蓝牙,其通风系统可缓解孩子久坐不适感,父母通过灯光颜色便可感知孩子体温。

“智能汽车是大势所趋,无论是虚拟世界的通信互联安全,还是物理世界的驾驶行为安全,其带来的安全问题都在与日俱增。”邓邱伟表示,这也是 360 致力于此的初衷。(赵广立)

混合云向边缘延伸: ZStack Mini 超融合和 ZStack CMP 发布

本报讯 4 月 16 日,以“互联融合,混合云向边缘延伸”为主题的 2019ZStack 新品和合作伙伴战略发布会在北京举行。会上,国产自主创新开源云计算服务商 ZStack 发布了 ZStack Mini 超融合一体机(以下简称 ZStack Mini)和 ZStack 多云管理平台(以下简称 ZStack CMP)两款新品。

据 ZStack 产品研发总监宁波介绍,ZStack Mini 结合 ZStack 核心引擎的灵活强大和双引擎高度集成化优势,只需两节点即可起步,降低用户 TCO 总成本,提供超小型云计算一体机,具备功能完善、轻量易用、弹性扩展、高可

用部署、助力边缘计算等特性。ZStack 高级产品经理丁阔表示,ZStack CMP 深度结合用户需求,管理多种公有云、私有云及多种异构云基础设施,具备多云平台接入、全栈资源管理、智能监控运维、云费用管理、IT 自服务、聚合监控大屏等特性,尤其适合大型集团客户。

“云计算实质上是一种普惠技术,而 IaaS 层作为云计算的基础,只有通过产品化、标准化才能真正落地实现普惠。ZStack Mini 和 ZStack CMP 两大新品,都是 ZStack 围绕 IaaS 领域的延伸和拓展。”ZStack 创始人兼 CEO 张鑫表示。(田瑞颖)