



独家专栏 | 如何利用深度学习搭建一个最简单的无人驾驶系统



雷锋网

百家号

10-29 19:46

雷锋网按：本文为雷锋网独家专栏，作者系佐思产研研究总监周彦武，雷锋网经授权发布。

国内最牛的无人驾驶厂家的运算平台是这样的：

一个英特尔至强 E5 的 CPU，拥有 12 内核，30MB 的二级缓存，运算能力为 400GOPS/s，功率消耗为 400 瓦（应该是第一代至强 E5，目前 E5 最高功率消耗都不超过 180 瓦），八个英伟达的 K80 GPU，每一个 K80 的运算能力为 8TOPS/s，通过 PCI-E 连接，每一个 K80 的功率消耗为 300 瓦，合计是 2800 瓦的功率，取得 64.4 TOPS/s 的运算能力。

和 Waymo 的车一样，需要一个备份系统，当然这个备份系统处于待机状态，功耗可能不到 500 瓦，但是为了有足够的冗余，同时还要考虑到传感器的功率消耗（传感器功率一般都很低，即便是 Velodyne 的 HDL-64E，典型功耗仅 60 瓦），整体系统的功率设计为 5000 瓦，原车的电源系统当然不行，需要一个发电机，或者双电瓶设计。这一套系统价格大约 3 万美元。

这样的设计只能用于 Demo，量产自然不能是这样。

FPGA、CPU 与 GPU



雷锋网

百家号

最近更新：10-29 19:46

简介：雷锋网——关注智能与未来！

作者最新文章

CNCC 2017最后一天，两场论坛引爆互联网和人工智能

2013年首次亮相的RoboBee，如今却能“上天入海”

小心吸尘器偷录家中视频：LG Hom-bot系列智能机器人有安全漏洞

相关文章



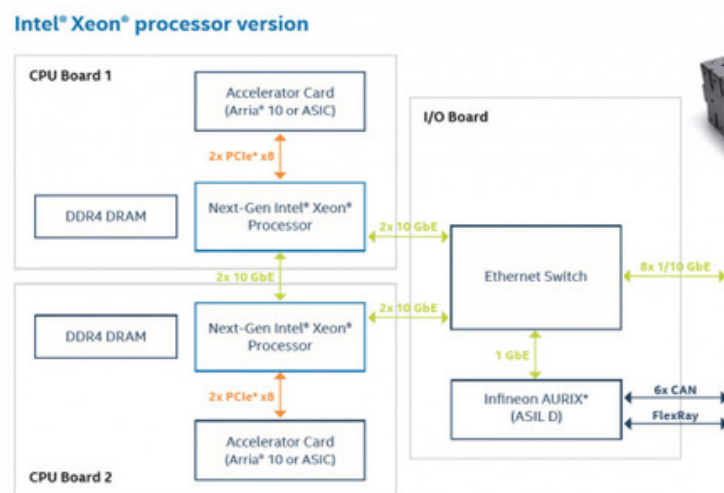
阿尔法狗退役，最强“新狗”AlphaGo Zero横空...
网易财经 10-19

NVIDIA 程亚冰：深度解析英伟达完全自动驾驶...



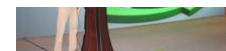
可能是这样的：未来汽车运算平台架构是一片 CPU 或者说 SoC, 对应大部分控制和运算量不大但逻辑关联比较多的运算，一片 FPGA 或 GPU 做加速，对应运算量很大，但内部几乎没有逻辑关联的运算，例如车辆识别算法，行人识别算法，车道线识别算法。

再加一片安全控制 MCU，通常是英飞凌的 Aurix 系列 MCU，使整体系统达到 ASIL D 级安全等级。即便是这样的系统，其功耗最少也在 500 瓦以上。



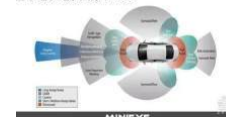
以英特尔的 Go 最为典型（如上图），Waymo 用 FCA 改造的无人车很有可能使用了类似的运算平台。

[百度首页](#) [AllinOneE](#)



深鉴科技获蚂蚁金服投资，对标英伟达水平的...
网易财经 10-25

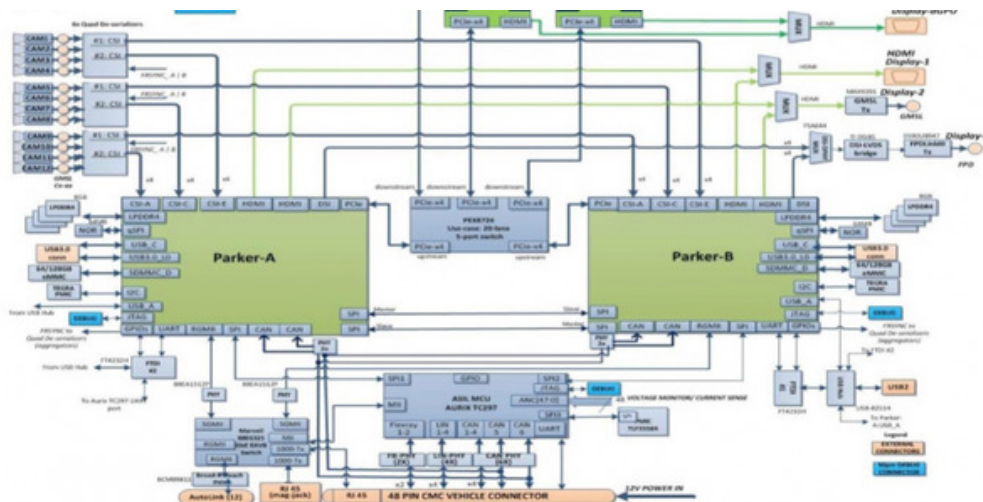
基于感知技术的ADAS



ADAS的前世今生和一个世界难题！
智东西 10-28



极客生涯 | 他用了42300个手工焊接的晶体管制...
科技e哥 10-23



上图为英伟达 Drive PX2

PX2 与英特尔 Go 基本上如出一辙，只不过用 GPU 取代了 FPGA。FPGA 作为一种高性能、低功耗的可编程芯片，可以根据客户定制来做针对性的算法设计。

所以在处理海量数据的时候，FPGA 相比于 CPU 和 GPU，优势在于：FPGA 计算效率更高，FPGA 更接近 I/O。FPGA 不采用指令和软件，是软硬件合一的器件。

对 FPGA 进行编程要使用硬件描述语言，硬件描述语言描述的逻辑可以直接被编译为晶体管电路的组合。所以 FPGA 实际上直接用晶体管电路实现用户的算法，没有通过指令系统的翻译。



中要实现一个 2 输入的与门，那么该与门的输入输出对应关系为：

输入	输出
0, 0	0
0, 1	0
1, 0	0
1, 1	1

既然有了此对应逻辑关系，那么在 SRAM 对应的地址中只需存储相应的输出数值，而输入数值作为地址来查表即可得到正确的输出值即可。

当输入地址为 0,0 时，输出 0 地址里存储的值 0，当输入地址为 0,1 时，输出 1 地址里存储的值 0，当输入地址为 1,0 时，输出 2 地址里存储的值 0，当输入地址为 1,1 时，输出 3 地址里存储的值 1。

某种意义上讲，FPGA 是不计算的，它只是个存储器，对逻辑组合电路，先预先存储不同输入对应的输出组合，然后根据输入数据，选择对应的输出。这就决定了 FPGA 拥有无论伦比的速度和低功耗。但是 FPGA 没有 CPU 和 GPU 的取指令和指令译码能力，这就注定无法单独使用，通常会加一个 ARM 内核的 CPU 来处理比较简单的指令，这样的 FPGA 叫 SoC FPGA。

这样一来，FPGA 的适用面广了，但是性能肯定要下降。尽管如此，还是比 CPU 和 GPU 的效率要高不少。

GPU 是 GPU 把晶体管更多用于计算单元，而不像 CPU 用于数据 Cache 和流程控制器。这样的设计是因为并行计算时每个数据单元执行相同程序，不



GPU 中一个逻辑控制单元对应多个计算单元，同时要想计算单元充分并行起来，逻辑控制必然不会太复杂，太复杂的逻辑控制无法发挥计算单元的并行度，例如过多的 if...else if...else if... 分支计算就无法提高计算单元的并行度，所以在 GPU 中逻辑控制单元也就不需要能够快速处理复杂控制。

所以 GPU 适合空间域的数据处理，数据之间缺乏时间顺序上的逻辑关联，数据的独立性强，图像是最适合 GPU 处理的，FPGA 并不占优势，而视频之间的语义关联则不适合 GPU，比如行人过马路这个视频，要根据前后视频关联来分析行人是要停下来还是加速通过，这种深度学习应用，FPGA 比 GPU 更合适。

而行人识别，车道线识别，GPU 更适合。同理，语音识别也是如此，要知道上下文语义关联，FPGA 就比 GPU 更合适。

也就是说，FPGA 更适合 RNN 和 LSTM，GPU 更适合 CNN。LSTM 可以看成 RNN 的变种，比 RNN 多了 memory cell 和 forget gate，语音信号又是一种上下文相关的长时相关性时序信号，所以 LSTM 做语音识别精读最高。微软 IBM 阿里采用的都是 LSTM。

LTSM 与无人驾驶

LSTM 和无人车有什么联系？这就是目前最火的深度学习与无人车的结合。传统的无人车使用大量的传感器和栅格法来做环境感知和决策，这种方法成本太高了。

于是就有人想出了利用深度学习来学习人类驾驶习惯的方法，只用一个不到 10 美元的消费级摄像头做传感器，用深度学习代替栅格法，这就是百度所说的 Road Hackers。



Road Hackers 是百度的自动驾驶开放平台，是可在真实道路上实现端到端模式的高级自动驾驶模型。

Road Hackers 平台首期将开放 1 万公里自动驾驶训练数据，并已具备提供海量中国路情驾驶数据开放，基于深度学习的自动驾驶算法演示，以及自动驾驶算法 Benchmark 评比等能力。

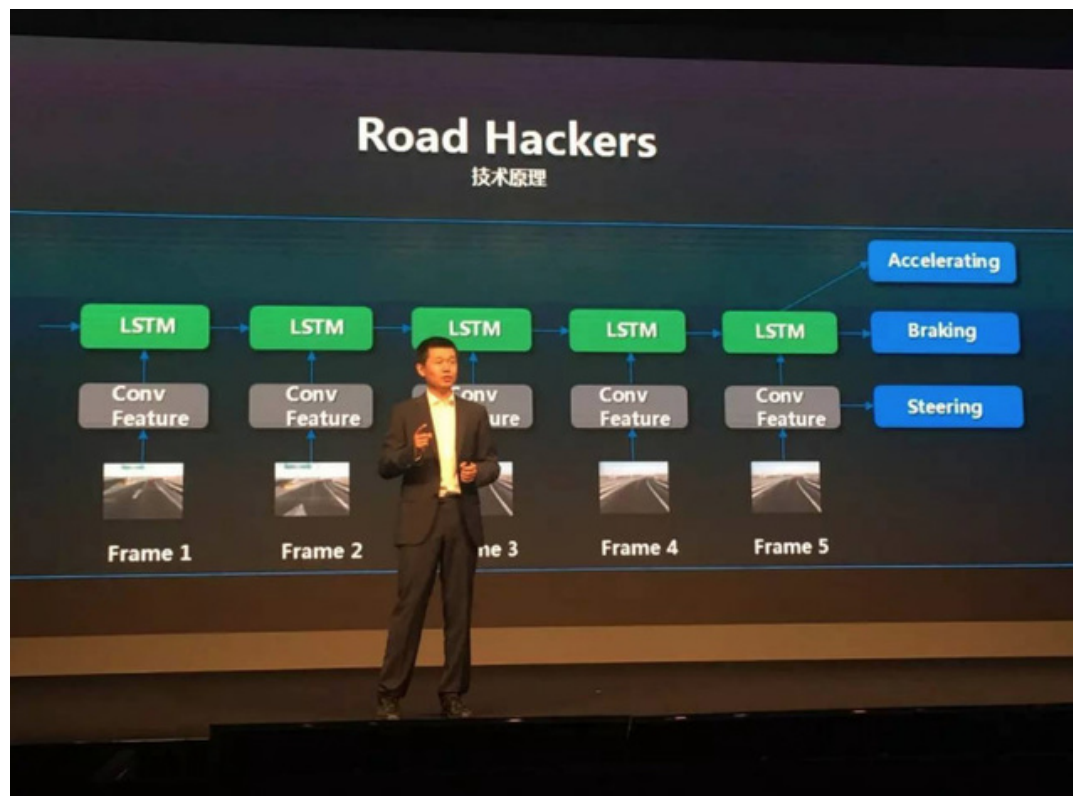
CES Asia 上的路测车辆采用 Road Hackers 平台技术，集成有全国首例基于单摄像头的端到端深度学习自动驾驶解决方案。

红色的线代表方向盘转向的角度。绿色的线代表用算法预测出的司机对方向盘的转动。从红绿两条线的趋势可以看出，收敛的趋势越来越明显。以前只



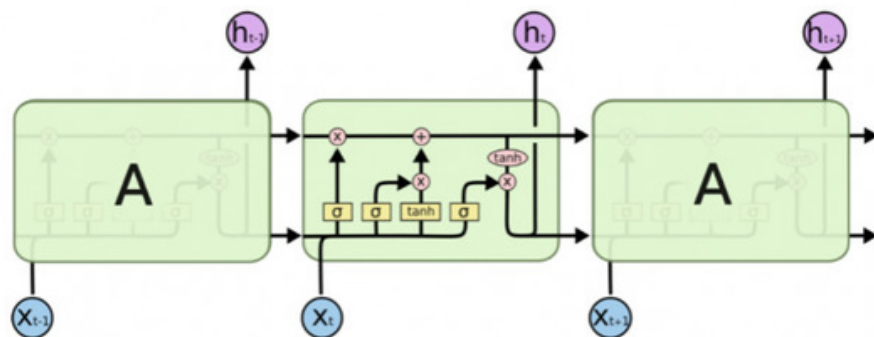
场。

[百度首页](#) [AllinOneE](#)



与百度类似，George Hotz 创办的 Comma.ai 也使用了类似的方法，并且在 2016 年 8 月公布了源代码，人人都可以下载研究。

简单解释一下 LSTM 的工作原理。LSTM 区别于 RNN 的地方，主要就在于它在算法中加入了一个判断信息有用与否的「处理器」，这个处理器作用的结构被称为 cell。



(LSTM 的内部处理器和三重门)

一个 cell 当中被放置了三扇门，分别叫做输入门、遗忘门和输出门。一个信息进入 LSTM 的网络当中，可以根据规则来判断是否有用。只有符合算法认证的信息才会留下，不符的信息则通过遗忘门被遗忘。说起来无非就是一进二出的工作原理，却可以在反复运算下解决神经网络中长期存在的大问题。

目前已经证明，LSTM 是解决长序依赖问题的有效技术，并且这种技术的普适性非常高，导致带来的可能性变化非常多。各研究者根据 LSTM 纷纷提出了自己的变量版本，这就让 LSTM 可以处理千变万化的垂直问题。这就像人类的学习过程，不断地重复正确的选择，直到形成牢固的记忆。

而 FPGA 实质上就是一种特殊的记忆体，所以 LSTM 更适合用 FPGA 来实现。

当然，这种只靠一个摄像头的无人驾驶显然极不靠谱。



其次，目前大多是基于单目的深度学习，单目在目标距离判断上是靠像素大小来估测的，无法得到深度信息，距离估测的准确性很低，尤其在高低起伏的路面，例如重庆和贵阳的道路或立交桥，单目所估测的距离可能 50% 都是错误的。

再次，摄像头的 FOV 与探测距离之间是矛盾的，FOV 越窄，距离就越远，但是车辆周围就会出现盲区，这是很容易出事故的，FOV 越宽，距离就越近，同时超过 70 度的广角会出现失真。即使三目摄像头也无法完全解决问题。

最后，低照度情况下肯定不能用。不过这是一种低成本的尝试，仍然值得赞扬。这种方法可以做 L2 或 L3 级的无人驾驶。

本文仅代表作者观点，不代表百度立场。系作者授权百家号发表，未经许可不得转载。