



莱迪思 sensAI 4.1 工具和 IP 将低功耗 FPGA 变为 网络边缘智能 AI/ML 计算引擎

摘要

网络边缘设备的爆发式增长推动着新应用的开发, 这些应用可以将大量原始数据转化为有用的、可操作的信息, 便于实时决策。莱迪思 sensAI 4.1 解决方案集合提供即用的 AI/ML 工具、IP 核、硬件平台、参考设计和演示以及定制设计服务, 将网络边缘设备 and 应用快速推向市场。

本文由莱迪思赞助

引言

毫无疑问，你已经读过或听说过，由于网络边缘设备数量激增，产生了不断增长的巨量数据流，这些设备包括自动驾驶汽车、物联网设备、消费电子产品，甚至是笔记本电脑和个人电脑。根据多项估算，截至 2025 年，运行的物联网设备将达到数百亿个。这些设备以连续数据流的形式向云端发送各种形式的数据，数据速率也千差万别。总体来说，这些设备将生成大量原始数据，且数据量随着时间的推移不断增加。

安全摄像头、自动驾驶汽车和 PC 中的视频录像机会生成高码率、高分辨率的视频流。物联网设备则生成中等码率的数据，汇聚到大数据流中。多种其他类型的物联网传感器（测量温度、压力、位置、光照水平等）会生成低码率数据流，但很快此类传感器的数量将会达到数十亿。因此，即使是这些低码率数据流也可以在进入云端之前汇聚成更大的、高码率的数据流。

5G 无线网络和其他高速网络技术的兴起，包括微微基站 (Picocell)、长距离物联网网络（如 LoRaWAN）以及全球联网卫星网络（如 SpaceX 不断扩展的星链宽带网络和 Swarm Technologies 基于卫星的物联网网络），提供了广泛而快速的云端访问（注：星链于 2021 年 8 月收购了 Swarm Technologies）。这些通信和网络技术加速了新兴的网络边缘计算设备和应用的发展。

新兴的网络边缘设备和应用包括自动驾驶汽车、机器人、自动化生产、远程监控、供应链和物流系统，以及保障公共和私人安全的视频监控。市场对这些网络边缘系统的需求极速增长，因为它们可以提高效率、降低运营成本并改善用户体验。但无论我们建设多少无线和有线通信基础设施，超量数据的骇浪都可能压垮或者堵塞这些通向云端的数据管道。

网络边缘本地化处理有助于疏通数据管道

这些趋势表明如今需要尽量在网络边缘数据产生的地方进行更多的处理，减少传输到云端的数据量。物联网和其他网络连接设备的爆发式增长是推动新的网络边缘设备开发的主要动力，这也进一步刺激了新应用的开发，从而将原始数据转换为有用的、可操作的信息，支持快速决策，实时应对不断变化的情形。

在网络边缘计算发展的早期阶段，公司主要关注将数据远距离传输到数据中心的成本问题。最初，网络边缘应用的一大特征是需要访问存储在云端和连接到云的其他计算机中的数据。这些早期应用通常不是实时应用；数百毫秒甚至数秒的响应时间都是可以接受的。然而，物联网设备的发展以及对网络边缘实时处理、分析和响应的需求不断增长，推动了网络边缘技术强有力的发展，同时也伴随着更大的设计挑战。

网络边缘处理使得计算和数据存储越来越靠近收集数据的设备端，而不是在数千里之外的数据中心进行分析和决策。网络边缘的实时应用通常不容许高延迟，因此处理、分析和决策必须转移到设备本身。这些网络边缘设备包括自动驾驶汽车、物联网传感器、安全摄像头、智能手机、笔记本电脑和个人电脑等。因此网络边缘计算的潜力巨大。

数据重压之下，云端无法包揽一切

智能手机和物联网设备的指数级增长推动了网络边缘计算的发展，这些设备无处不在，必须连接到互联网才能向云端发送信息或从云端接收信息。一些物联网设备（例如摄像机）在运行过程中会生成大量数据。

其他物联网设备，如温度传感器，会生成少量数据，但由于这样的传感器数量可达数十亿，为云端处理带来了极大的负担。因此，基于网络边缘的处理十分必要，不仅可以降低云端的网络通信成本和云存储成本，还能避免云端数据通道过载。

网络边缘产品和应用的开发人员越来越多地采用人工智能和机器学习 (AI/ML) 算法来匹配和识别复杂的模式，以帮助分析数据并据此做出决策。事实上，AI/ML 技术的使用增长极其迅猛。

如今 AI/ML 算法被视为高效处理原始数据的必要手段，因为它们可以识别出传统的算法程序难以解析和识别的复杂、多维度的数据模式。一些特定的 AI/ML 应用包括检测、识别、辨认和计数人员或物体；资产和存货追踪、环境感知、声音和语音检测和识别、系统健康监测以及系统维护调度等。

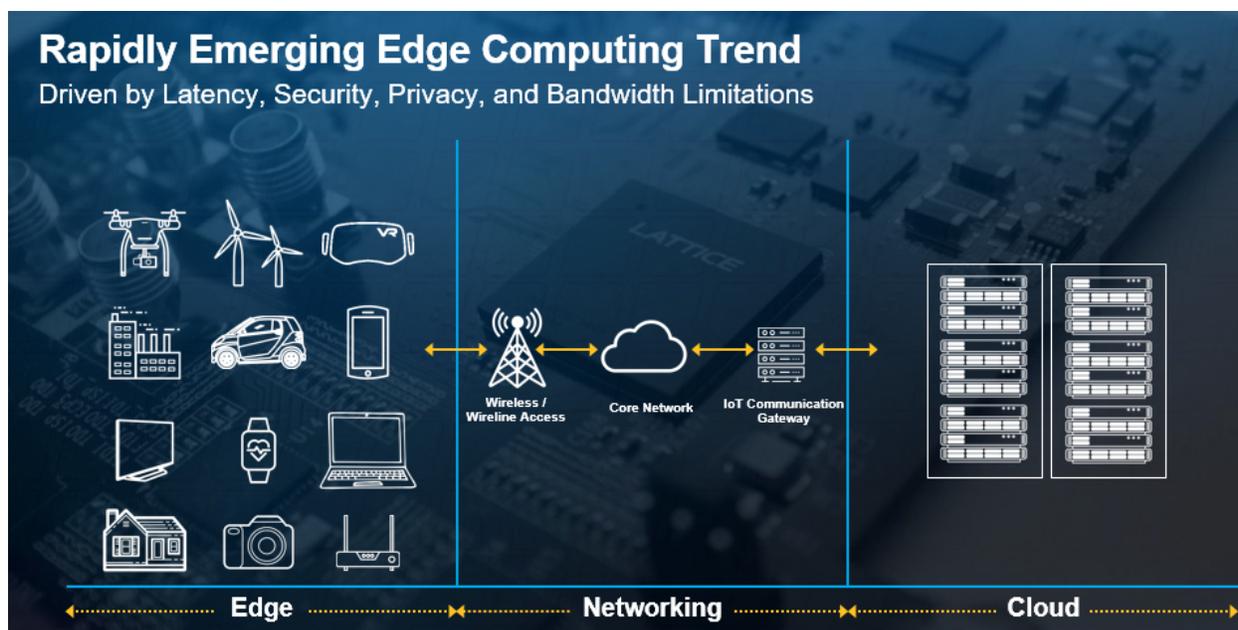


图1. 网络边缘计算的趋势(图片来源:莱迪思)

许多可以利用 AI/ML 功能的网络边缘应用需要在极具严苛的功耗限制下运行。这些广泛分布的设备通常依靠电池供电。此类应用在各种网络边缘环境中比比皆是，包括工厂、农场、办公楼、零售店、医院、仓库、街道和住宅。随着它们数量的增加，这些设备需要在仅充一次电或者仅依靠收集和存储能量的情况下运行较长时间，甚至可能是几个月或几年。

因此，许多设备需要在大部分时间里处于睡眠或休眠状态，在设备处于非活动状态时大部分电路应处于低功耗待机模式。然后激活事件会在需要时启动设备。在此类应用中，以超低功耗运行的基础电路系统必须保持待命，等待激活事件，然后根据需要为设备的其余部分供电。

FPGA 以低功耗实现 AI/ML

对低运行功耗和 AI/ML 算法实现的需求似乎与低功耗网络边缘设备设计的要求相互冲突。然而，这两种复杂的设计要求其实并不矛盾。莱迪思最新的 FPGA——低功耗、小尺寸、高性能的 CertusPro-NX 系列器件——专为满足低功耗网络边缘设备的诸多设计要求而定制。这些 FPGA 可以支持多个传感器、显示器，支持高分辨率视频、网络连接和网络边缘 AI/ML 处理。

与此同时，莱迪思最新发布的 sensAI 解决方案集合 4.1 版本提供了即用的 AI/ML 工具、IP 核、硬件平台、参考设计和演示以及定制化设计服务，有助于设计团队开发新的网络边缘设备，并将其快速推向市场。最新版本的 sensAI 支持 CertusPro-NX FPGA。

莱迪思 sensAI 解决方案集合可加速端到端的 AI/ML 模型训练、验证和编译。莱迪思在 2021 年初发布的 sensAI 4.0 中新增了 sensAI Studio 设计环境，这是一种基于图形用户界面 (GUI) 的工具，可帮助开发人员快速构建机器学习应用。在使用莱迪思 sensAI 4.1 中的工具设置网络边缘计算设计，并且采用莱迪思 iCE40 UltraPlus、CrossLink-NX、ECP5 和 CertusPro-NX FPGA 时，可以在超低功耗下实现实时的 AI/ML 功能——功耗低至 1mW 到 1W。

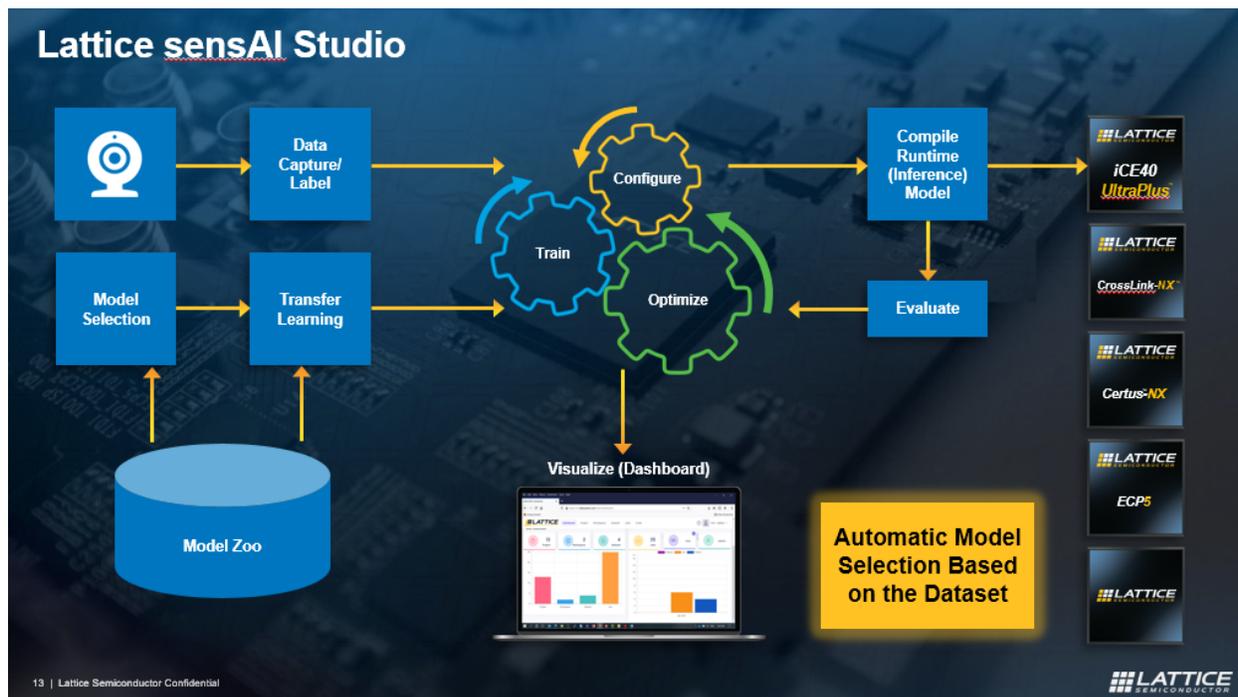


图 2. 莱迪思 sensAI Studio 设计环境加速端到端的 AI/ML 模型训练、验证和编译。(图片来源:莱迪思)

随着 sensAI 4.1 支持莱迪思 CertusPro-NX FPGA 系列产品, sensAI 的性能也有了较大提升,除了已有的对象检测和追踪应用之外,还新增了对多个对象实时分类等应用。sensAI 4.1 解决方案集合包括更新的神经网络编译器,还兼容其他广泛使用的机器学习平台,包括最新版本的 Caffe、Keras、TensorFlow 和 TensorFlow Lite。

莱迪思 sensAI 4.1 解决方案集合中的 IP 核包括三种类型的卷积神经网络(CNN)加速器——CNN、CNN Plus 和 CNN Compact——以及一个 CNN 协处理器引擎。CNN IP 核能让开发人员使用其他人发布的广泛使用的各类 CNN,例如 Mobilenet v1/v2、Resnet、SSD 和 VGG,或者根据需要自定义 CNN 模型。sensAI 4.1 CNN 加速器利用莱迪思 FPGA 的并行处理能力、分布式存储器和 DSP 资源,极大简化了超低功耗 AI 设计的实现。加速器核利用 FPGA 的可编程逻辑来实现低功耗神经网络,包括极其高效的二值神经网络(BNN),能够以毫瓦级超低功耗实现 CNN。

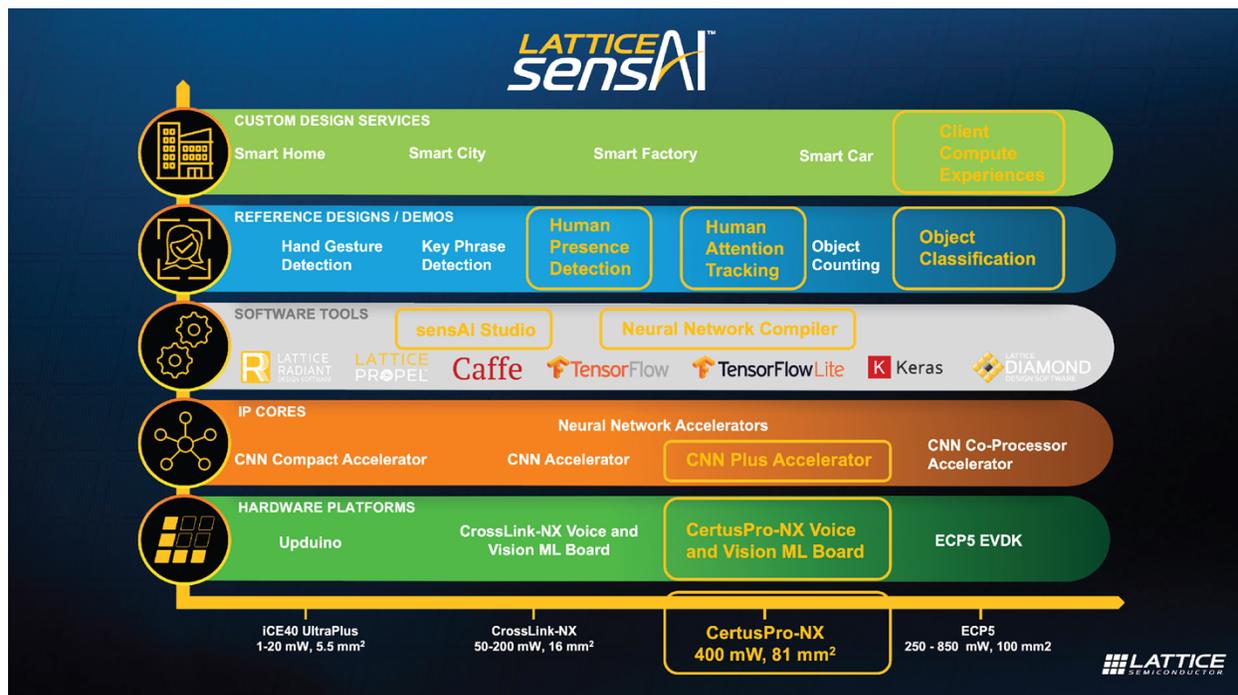


图 3. 莱迪思 sensAI 解决方案集合可开发基于莱迪思 FPGA 的 AI/ML 设备。(图片来源:莱迪思)

莱迪思 sensAI 4.1 参考设计

莱迪思 FPGA 提供可编程 I/O, 经配置可支持传感器接口常用的多种电气接口标准。公司还提供许多硬核和软核 IP 模块以支持不同的传感器通信协议。由于 FPGA 长期以来在传感器融合方面具有显著优势, 因此莱迪思 sensAI 4.1 的设计旨在简化网络边缘设备中基于多个传感器的 AI/ML 推理功能的开发, 实现智能的传感器融合。sensAI 4.1 解决方案集合包括许多参考设计示例, 演示了多种智能传感器融合的应用案例, 它们可以同时运行, 实现深入的情景感知。这些参考设计包括:

- **手势检测**

该参考设计使用 IR 图像传感器, 实现了一个基于 AI 的低功耗手势检测系统。该参考设计提供了一个训练数据集、可使用常用神经网络训练工具训练的脚本以及一个神经网络模型, 方便用户进行修改。

- **关键词检测**

该参考设计使用数字 MEMS 麦克风持续检测关键词话语。设计人员可以使用深度学习框架(例如 Caffe、Tensorflow 或 Keras)更新提供的训练数据集, 为系统添加唤醒词功能。参

考设计包括一个训练数据集、可使用常用神经网络训练工具训练的脚本以及一个神经网络模型，方便用户进行修改。

- *人脸检测*

该参考设计使用图像传感器实现基于 CNN 的人脸识别，并且可以通过修改训练数据库来识别其他类型的目标。

- *人员侦测*

该参考设计使用 CMOS 图像传感器持续检测人员的存在。基于此设计的 AI 系统可以使用深度学习框架(例如 Caffe 或 Tensorflow)更新所提供的训练模型来检测和定位任何感兴趣的目标。该参考设计包括一个神经网络模型、一个训练数据集和可使用常用训练工具训练的脚本。

- *目标检测、分类、追踪和计数*

该参考设计提供了目标检测、分类、追踪和计数的示例，拥有完整的设计，包括用于莱迪思开发板的 FPGA RTL、神经网络模型、示例训练数据集以及用于重新创建和更新设计的脚本。

可以使用 AI 的常见和潜在的网络边缘应用

使用 AI/ML 算法提高众多网络边缘设备(例如自主机器人、环境控制和视频安全摄像头)的性能具有明显优势，而其他类型的网络边缘设备也可以从中受益，例如 PC 和笔记本电脑。莱迪思正与合作伙伴和客户合作，利用多模式、智能传感器融合和 AI/ML 技术，不断提升 PC/笔记本电脑用户的体验，并显著降低笔记本电脑的运行功耗，在某些应用中，电池使用时间最多提高了 28%。

哪些设备特性能发挥出潜在的价值呢？

PC 和笔记本电脑在 24 小时内的使用情况差异很大，一般在白天工作时间集中使用。然而，即使是在工作时间，它们也会有休息状态。人们会偶尔休息，中午也会用餐，这些时间他们通常会让计算机保持运行状态，确保他们打开的各种应用不被关闭。

将 AI/ML 分析和决策与计算机现有的传感器(摄像头和麦克风)相结合，实现智能传感器融合，能让 PC 或笔记本电脑感知周围环境，从而决定何时关闭显示器和 CPU，以及何时应该给它们重新供电。

存在检测最简单的用途是在周围无人时关闭计算机。当用户长时间远离屏幕时，注意力追踪功能可以调暗计算机屏幕并激活低功耗模式。充当智能传感器中心的低功耗、小尺寸 FPGA 可以接收来自计算机传感器的输入，然后根据情况决定为哪些组件供电。

解决隐私和安全问题

同样，这些功能也可以增强计算机的隐私和安全性。计算机的内置会议摄像头可用于监控用户身后的背景，检测是否有人从用户的肩膀后面窥视。如果计算机被配置为保护隐私，当授权用户背后有人疑似在偷窥计算机屏幕时，它可以弹出警告提醒用户甚至自动调暗屏幕。需要注意的是，使用这些解决方案，所有推理数据都保存在 FPGA 本地。仅将元数据传递给 SoC，这进一步增强了隐私并提高了安全性。

优化用户体验

AI/ML 功能还可以增强计算机用户的总体体验。例如，基于 AI/ML 的面部取景功能可以利用内置视频会议摄像头的较高分辨率来裁剪和居中用户的头像，为视频会议提供较好的画面。与会者还可以在会议期间移动的同时，他们的图像依然保持居中。同样，手势识别可以为笔记本电脑或 PC 或任何其他支持视频的物联网设备添加非接触式操作功能。

健康方面的益处

许多公司现在明确表示要保障员工的健康，基于 AI/ML 的感知功能可以通过弹出提醒和其他措施帮助避免重复性压力损伤，并利用计算机的视频传感器确保员工实际上采纳了给出的休息建议。

AI/ML 应用还可用于检测用户的姿势，这可能是造成重复性压力损伤的另一个因素。这些利用传感器主动反馈的特性可以用来开发健康应用，这明显优于目前企业中所使用的简单的定时提醒，能够有效应对压力相关的工伤。

所有这些通过 AI/ML 实现的功能都可以帮助供应商打造出对企业买家更具吸引力的 PC 和笔记本电脑，并且所有这些功能都可以通过 sensAI 4.1 解决方案集合和莱迪思低功耗 FPGA 的特性来实现。

这种 FPGA 的使用方式超越了长期以来 FPGA 开发的标志性功能——传感器连接和融合，并且基于成熟的 AI/ML 算法，新增了传感器信号分析和决策制定功能。AI/ML 的加入使得 FPGA 成为低功耗系统控制器，可管理系统功能、增强用户体验并通过降低整体系统运行功耗大幅延长电池寿命。

结论:数十亿网络边缘设备的庞大市场有待开发

凭借其多个低功耗 FPGA 系列产品和支持这些产品系列的 sensAI 4.1 解决方案集合,莱迪思致力于为数十亿计的网络边缘设备带来 AI/ML 技术。因此,网络边缘应用是一个极具潜力的目标市场。

根据多方面的估算,全球广泛的地域需要数百亿个网络边缘设备来满足大量网络边缘市场的需求,这对于 FPGA 业务来说非常具有吸引力——当然这种规模对于任何行业来说都是如此。莱迪思发布 sensAI 4.1 解决方案集合及其低功耗、小尺寸 FPGA 系列,则是直接瞄准了网络边缘应用和市场。莱迪思的 sensAI 4.1 解决方案集合是一种网络边缘应用的创新开发工具,系统开发人员可以借此为各种市场开发灵活、针对特定应用、基于 FPGA 的 AI/ML 推理解决方案。

版权所有 © 2021 TIRIAS Research。TIRIAS Research 保留本文的所有权利。

未经 TIRIAS Research 事先书面明确许可,禁止转载全文或部分内容。

本文包含的信息在撰写时被视为真实可靠的,但不保证其准确性或完整性。

产品和公司名称可能是其各自所有者的商标(TM)或注册商标(®)。

本文内容系对统计数据 and 信息的解释和分析,这些数据和信息或已全面公开,或由可靠的机构或个人发布。