白皮书

FPGA AI 套件

英特尔® FPGA 和 SoC 配合英特尔® FPGA AI 套件和 OpenVINO™ 工具包，推动嵌入式产品、边缘 AI 和机器学习应用的发展

作者

Jahanzeb Ahmad
英特尔公司
高级解决方案架构师

Mark Jervis
英特尔公司
高级解决方案架构师

Rama Venkata
英特尔公司
高级 AI 技术营销经理

随着企业运营的节奏日益加快，人们对快速响应的期望日益提升，决策逐渐从数据中心转向网络边缘。无论是要尽量保障繁忙作业，避免闲置的车间生产线、在手术室等待分析结果的医生，还是正在待命前往扑灭熊熊山火的消防队、寻找各种洞察从而帮助修复珊瑚礁的科学家，又或是在顾客焦急等待购物服务的零售环境中，企业都必须对系统进行相关配置，以收集信息，获得可行洞察，并实时帮助做出决策或提供分析结果。在越来越多的情况下，似乎只有完全自动化决策的及时性才足以满足要求。

如今，边缘采集的数据量十分庞大。据 Gartner 预测，到 2025 年，将有多达 75% 的企业数据会在传统数据中心以外生成[1]。Gartner 高级研究总监 Santhosh Rao 表示：“踏上业务数字化之旅的企业和机构已经意识到，必须采用去中心化的方法来满足数字业务基础设施的要求。随着数据体量与增长速度的提升，将信息流式传输至云端或数据中心进行处理的效率也愈发低下[1]。”

将采用人工智能（AI）和机器学习（ML）算法的计算能力转移到更靠近数据产生的位置，甚至在许多情况下在数据产生的边缘直接提供上述能力，能够实现全新的实时用例，拓展潜在的新收入来源，同时防止敏感数据在网络中流转后再进入数据中心。实现对边缘数据的实时响应需要至少四种技术的有效组合：

- 边缘计算
- 人工智能（AI）
- 高速网络
- 云

企业必须在整个基础设施中整合这些技术，才能获得云边协同智能的全部优势。将更多具有 AI 功能的设备和算力安排在边缘，可以在提升数据处理量的同时也生成更多数据，从而实现更复杂的 AI 用例，进而获得更多可行洞察。

企业的核心可能位于数据中心或云端，但远离该核心位置的边缘却能涵盖除此以外的所有数据收集、处理、存储和通信功能。边缘包含以下部分：

- 边缘设备，即生成、收集、处理和/或使用数据的资产，包括智能摄像头、工业传感器、机器人、自动驾驶车辆、可穿戴设备、智能手机、智能扬声器和无人机等设备。
- 边缘基础设施，即能够从不同来源聚合众多数据流的设备，如本地服务器、网关和网络视频录像机等。

[1] 这些边缘设备都可以从 AI 功能中受益。

目录

边缘的定义 2
边缘 AI 用例 2
实施 AI 需要满足哪些要求？ 2
为什么 FPGA 是 AI 实施的理想选择？ 3
FPGA 十分契合网络边缘及核心众多终端市场的需求 4
医疗应用中的 AI 4
工业和制造业应用中的 AI 4
面向英特尔® FPGA 和 SoC 的边缘就绪型 AI 工具套件 6
结论 8
参考资料 8

边缘的定义

边缘设备通常是一些小型设备（例如，智能手表或智能摄像头），边缘环境中几乎没有空间容纳又大又重的组件。同时，边缘设备的供电也往往十分有限。这意味着边缘硬件必须要高效利用空间和电能。这些设备还必须提供高性能，甚至要能足以添加 AI 工作负载来处理本地收集的数据。

虽然在执行 AI 推理工作负载时，边缘设备可以且经常是独立运行，但对于 AI 训练而言，连接多个边缘设备以实现联邦学习能带来众多益处。联邦学习使边缘设备可协作学习并共享预测模型，同时所有训练数据都位于边缘设备，不必存储在云端，从而提高了数据安全性。

能够以边缘集群或网络服务器等形式支持更全面或更复杂的边缘计算的硬件性能往往高于独立的边缘设备。这类硬件可能也会根据需要使用安全或连接功能，从而支持指定用例。

以下是两个在边缘使用 AI 的真实案例：

- 在制造业和工业领域，边缘计算蕴含巨大潜力。例如，奥迪的内卡苏尔姆工厂每天要组装多达 1,000 辆汽车，而每辆汽车大约有 5,000 个焊点。也就是说，仅一家工厂每天就需要检查 500 万个焊点。如果每天都要人工检查上百万个焊点，不仅成本高昂，耗时费力，而且也并不现实，更别说奥迪的目标是希望能够以出色的精度实现焊点全检。
- 在一些濒危物种的栖息地，环保人员进入可能会造成问题，而智能摄像头和视频分析则有望帮助监控与保护这些地方。例如，珊瑚礁修复通常需要潜水员下水进行监控。他们需要潜入水中直接收集数据，或手动拍摄珊瑚礁的视频或图像，供日后分析。这种数据收集方法可能会干扰野生动物行为，无意中影响研究结果。此外，数据采集也很有限，因为潜水员一次只能在水下安全停留大约 30 分钟。菲律宾的 CoRaiL 项目通过利用智能摄像头和 AI 增强型视频分析来研究珊瑚礁的韧性，成功解决了这些问题。

边缘 AI 用例

以下是一些边缘 AI 用例：

1. 在医疗领域，AI 有许多潜在用途，医疗影像就是其中非常主流的一种。每天会产生成千上万的医疗影像，如 CT 扫描、X 光和 MRI 等，每张影像都需要经过仔细分析来发现其中异常，实现准确诊断。
2. 在零售层面，机器视觉能够可靠地读取条码、文本和数字，以帮助管理、跟踪和分析库存水平，确保重要材料由需要的人员掌握。互联的响应式智能数字标牌可以根据顾客行为与喜好为顾客推荐产品或优惠。这又能促进零售商了解他们向消费者发出的讯息何时真正起到了效果。自助服务设施和无人商店可为顾客提供一系列服务，打造个性化购物体验。与此同时，机器学习又可以分析遍布整个商店的摄像头采集的多个视频流，帮助实时识别潜在犯罪行为。
3. 在美国，机器人被用来对医院表面进行紫外线（UV）消毒，这样既能够有效杀死病毒，也避免了紫外线对人类造成伤害。机器人能够利用 AI 在医院内导航，先确认所在空间没人，再用紫外线对该区域进行消毒。在该用例中，AI 的采用有助于确保整个医院的安全，同时尽量保持繁忙区域开放正常运营，以供使用。
4. 具备 AI 功能的智能摄像头能够带来巨大价值，实现重复性日常任务的自动化，从而解放员工，使他们能专注于应对更复杂的挑战。例如，基于 AI 的车牌识别被广泛应用于各种场景，包括防止未经授权车辆进入的安保应用，以及自动放行，让注册用户能够直接驱车进入洗车区等等。

实施 AI 需要满足哪些要求？

在整个企业内广泛实施 AI 时，务必要确保以下三大基础设施要素均具备处理 AI 工作负载的足够性能。这三大要素分别是：边缘设备、边缘基础设施和云。实施 AI 的具体要求包括：

- 高性能：AI 工作负载往往计算密集度高，因此在进行 AI 训练或推理的地方，必须具备强大的计算性能。
- 低时延：AI 的一大优势在于能够支持实时决策。将 AI 工作负载转移到边缘位置（即使只是将部分 AI 工作负载转移到边缘），有助于降低决策时延。
- 高容量：AI 依赖大量数据，因此，运行 AI 的基础设施必须确保计算、存储和内存容量能够胜任任务，从而避免瓶颈。
- 可靠的安全性：AI 工作负载需要大量越来越敏感的数据（例如，在医疗或公共安全领域）。无论是何种 AI 工作负载，运行它们的设备和软件都必须安全可靠。

英特尔提供众多技术和解决方案，可在满足上述要求的同时，支持企业和机构实现从边缘到云的 AI 工作负载。图 1 展示了探索英特尔® AI 解决方案的典型初始框架，但最终的解决方案也将取决于低时延或定制板外形等特定要求。面向 AI 的英特尔® 边缘技术解决方案能够在各类设备上实现高性能推理，这些设备包括本地服务器、PC、摄像头、机器人和无人机等。由于在 AI 领域并不存在“一体适用”的解决方案，因此英特尔推出了包括 CPU、GPU、VPU 和 FPGA 在内的产品组合，旨在提供低时延推理，帮助消除数据瓶颈。英特尔® oneAPI AI 分析工具套件（AI 套件）和英特尔® 分发版 OpenVINO™ 工具包以一套统一的 AI 开发工具支持广泛的英特尔计算设备。

图 1. 英特尔的 AI 产品组合框架

为什么 FPGA 是 AI 实施的理想选择？

人的大脑中有近 1,000 亿个神经元。尽管这已经是个天文数字，但将这些神经元组织成为网络的神经连接数量更是达到了百万亿级，而神经连接的数量显著影响着大脑的能力。FPGA 内的互连性就类似于人脑中的神经连接。FPGA 内的可编程逻辑结构也以类似的方式相互连接，这就是为什么英特尔® FPGA 是神经网络和其他 AI 工作负载实施的理想选择。在逻辑和线路互连的层面上，FPGA 的比特级动态可编程性就好比灵活的大脑，可以调整注意力，专注于当前的特定任务。此外，一直以来，FPGA 的外部 I/O 也具备其他硬件架构所不具备的出色灵活性，可以连接到雷达、音频、振动和视觉等各种来源的传感器。这些特性能够让信号实时进出 FPGA，达到可媲美人脑的高级智能水平。

英特尔® FPGA 家族包括英特尔® Cyclone® 10 GX FPGA、英特尔® Arria® 10 GX FPGA 和英特尔® Stratix® 10 GX FPGA 等。这些产品具备 I/O 灵活性、低功耗（或每次推理的能耗）和低时延，本就可在 AI 推理上带来优势。这些优势在三个全新的英特尔® FPGA 和片上系统（SoC）家族的产品中又得到了补充，使得 AI 推理性能进一步获得了显著提升。这三个家族分别是英特尔® Stratix® 10 NX FPGA 以及英特尔® Agilex™ FPGA 家族的新成员：英特尔® Agilex™ D 系列 FPGA，和代号为“Sundance Mesa”的全新英特尔® Agilex™ 设备家族。这些英特尔® FPGA 和 SoC 家族包含专门面向张量数学运算优化的专用 DSP 模块，为加速 AI 计算奠定了基础。

第一款采用张量模块的英特尔® FPGA 是英特尔 2020 年 6 月 18 日推出的英特尔® Stratix® 10 NX FPGA。英特尔® Stratix® 10 NX FPGA 的张量模块架构针对 AI 计算中常用的矩阵-矩阵或矢量-矩阵乘法和加法运算进行了优化，旨在高效地用于各种不同规模的矩阵。该张量模块支持 INT8 和 INT4 数据计算，并通过共享指数支持 FP16 和 FP12 块浮点的数字格式。

此前的英特尔® Agilex™ 设备家族就已经配备可变精度数字信号处理（DSP）模块，能够提供多种 AI 功能，而集成在全新英特尔® Agilex™ FPGA 和 SoC FPGA 结构中的 DSP 模块在此前模块设计的基础上，还引入了英特尔® Stratix® 10 NX FPGA 中所用的张量模块的多种功能。采用 AI 张量模块的增强型 DSP 引入了两种全新的重要运算：面向 AI 的张量处理能力和面向信号处理应用的复数支持。此类应用包括快速傅里叶变换（FFT）和复杂有限脉冲响应（FIR）滤波器等。

第一种模式可通过 INT8 张量模式增强 AI。该模式可在一个采用 AI 张量模块的增强型 DSP 中提供 20 次 INT8 乘法。与之前的英特尔® Agilex™ 设备家族相比，INT8 计算密度提升高达 5 倍。张量模式使用两列的张量结构，同时具备 INT32 和 FP32 的级联和累加功能，还支持块浮点指数，以改善推理精度和低精度训练。此外，可变精度 DSP 的 AI 功能也有所增强。矢量模式也已经从四个 INT9 乘法器（Multiplier）升级到了六个 INT9 乘法器。这些模式对以 AI 为中心的张量数学运算和各类 DSP 应用格外有用。

应用	乘法器	每个 DSP 模块的能力		提升*
		之前的英特尔® Agilex™ 设备	采用 AI 张量模块的增强型 DSP*	
AI，信号处理	INT8	4 OPS	20 OPS	5 倍
	INT9	4 个乘法器	6 个乘法器	50%
信号处理	16 位复数乘法器	需要 2 个 DSP 模块	1 个 DSP 模块	2 倍

图 2. AI 和 DSP 计算密度的数量级提升

*限英特尔® Agilex™ D 系列 FPGA 和代号为 Sundance Mesa 的全新英特尔® Agilex™ 设备家族提供。

第二种新模式是复数运算，可在运行复数乘法时使张量模块的性能翻倍。过去，复数乘法需要两个 DSP 模块，但这一全新英特尔® Agilex™ FPGA 和 SoC FPGA 家族产品在一个采用 AI 张量模块的增强型 DSP 中就可进行 16 位定点复数乘法运算。

FPGA 十分契合网络边缘及核心众多终端市场的需求

许多数据中心以外的终端市场都很适合采用 FPGA 来实现应用和 AI 计算功能所需的逻辑，从而支持在本地处理数据。这些终端市场包括：

- 医疗和生命科学，包括医疗监护仪、具有图像识别和物体检测功能的 2D 诊断设备（例如，X 射线设备和内窥镜），以及其他类型的病理学检测、基因组测序、手术机器人等设备。
- 军事和航空航天，包括无人飞行载具 (UAV)、目标检测、雷达侦测和分类等。
- 工业应用，可用于在边缘增加基于 AI 的检测和实时控制。
- ProAV（专业影音）系统，包括具备人脸识别功能，从而可以实现镜头自动平移/缩放和背景消除的视频会议摄像头、具备自动人脸检测，从而可以实现精准对焦的摄影棚用摄像头。
- 广播视频，包括从标准动态范围 (SDR) 到高动态范围 (HDR) 的转换、不同视频分辨率的智能转换，以及可变帧率视频的采集和显示。
- 消费级应用，包括具备人眼检测和跟踪功能，从而可以实现立体成像的 3D 显示器。

下面介绍了几个 AI 在医疗行业和工业/制造业的深入应用示例：

医疗应用中的 AI

患者和医护人员的人口结构正在变化，同时人们愈发希望在降低医疗成本的同时改善医疗成果。这些因素推动着 AI 在医疗行业的广泛应用。AI 有助于提高基于 MRI 和 CT 成像的癌症诊断准确性；基于 AI 的信息系统和机器人手术设备可协助外科医生手术；AI 还能通过基于全球数据集的模型改善罕见疾病的治疗。

AI 增强型内窥镜摄像头就是 AI 在医疗领域应用的一个具体示例，它已经广泛用于各个医疗科室，如神经内科、骨科、泌尿科和妇科。这种内窥镜摄像头系统越来越多地能够支持高级成像功能，包括边缘增强和色彩校正，为医生提供更清晰、更容易解读的图像。英特尔® FPGA 的高性能、小尺寸和低能耗特性，能够为内窥镜摄像平台额外提供其所需的实时功能。

英特尔® FPGA 还赋能内窥镜摄像头制造商支持包括以下各种 AI 增强型检测在内的多种 AI 用例：

- 结直肠筛查中的息肉检测
- 内窥镜食道筛查中与巴雷特食管相关的异常增生检测

美国每年要进行 1,600 多万次结直肠镜检查，仅美国每年诊断出的巴雷特食管病例就多达 20 多万例。因此，AI 增强型医学成像功能将显著提升这些内窥镜检查的效率和准确性。这些基于 AI 的全新功能和优化将推动 AI 在内窥镜中应用的持续增长，帮助医生满足日益增长的微创检查和手术需求。

工业和制造业应用中的 AI

现代制造业是一个复杂的体系，包含众多系统。各类传感器、摄像头和执行器组成了一个相互连接、联网控制的分层架构。英特尔® FPGA 广泛用于整个分层架构，可确保满足硬实时和安全要求。此外，制造业正在经历第四次工业革命，运营技术 (OT) 系统与信息技术 (IT) 系统愈发融合，构成了更智能、更灵活的工厂，提供更高效、更自动化的生产，同时需要的人为干预也更少。

各种工业应用和制造工厂所广泛使用的通信技术，包括 5G、工业网关和智能网卡 (NIC) 等均采用英特尔® FPGA。英特尔® FPGA 被用于工作负载需要 I/O 灵活性、直接数据提取能力、确定性计算能力、更低运行功耗，以及需要在严苛的工业环境中运行的场景。在这些场景中，AI 技术皆可提供支持，并且也越来越多地被用于制造业应用的视觉和非视觉类任务。

表 1 展现了工业和制造业中 AI 的广泛应用。

	压铸	纺织	电器和电子产品	手工装配
任务	▪ 包装、部件或表面缺陷检测 ▪ 线上质量控制/保证 ▪ 产品缺陷检测 ▪ 原材料外观检测 ▪ 资产管理	根据历史数据预测未来结果 ▪ 预测产品质量 评估设备健康状况并预测维护需求 ▪ 预测产量波动	▪ 识别机会/需要改进的流程 ▪ 安全的工人协作 ▪ 取放和分拣 ▪ 托盘装货/卸货 ▪ 焊接 ▪ 机器维护 ▪ 视觉辅助机器人 ▪ 训练/编程 ▪ 能耗优化型运动控制 ▪ 能够导航和避障的 AMR 感知	▪ 具备软实时功能的控制流程 ▪ 优化流程效率
价值	▪ 降低生产成本 ▪ 减少客户退货	▪ 缩短工厂故障时间 ▪ 防止昂贵的维护费用	▪ 提高工厂生产和效率	▪ 整合多供应商解决方案，并实现互操作性 ▪ 管理生命周期管理成本

表 1. AI 在现代工厂中的用例示例

边缘设备采用的英特尔® FPGA 可以安装在其他传感和控制功能旁边，或与之集成，协同实现高能效的 AI 工作负载。例如，英特尔® FPGA 可直接提取摄像头或传感器数据，以确定性计算、低时延和高吞吐量运行工作负载。英特尔® FPGA 可直接从一个或多个摄像头提取视频数据，对视频进行预处理（包括优化对比度和曝光、增强边缘、校正色彩），在视频流中抓取相关帧，并进行特征或缺陷检测，这一切都能实时完成。可编程逻辑控制器（PLC）制造商已在其新款 PLC 中采用小型 AI 引擎，为下一代控制器添加智能功能。在生产线上，AI 增强型视觉检测能够比人工更快、更准确地进行缺陷检测和质量评估。

将 AI 增强型功能与视频处理管道或其他功能紧密结合，是建立优化的实时工业系统的重要方式，而低功耗和低时延正是其中的关键因素。表 2 展示了 AI 增强型图像和视频处理在工业和制造业中的部分用途：

AI 让机器人设置和训练工作变得更加容易，可以教会机器人感知周围环境并作出相应反应。此类场景下，机器人可执行的任务包括根据示范进行学习、取放物体、使用直接反馈来控制末端执行工具（如焊接工具），以及与其他机器人或工人实现安全的协作。这些任务同样适用于自主移动机器人（AMR），因为它们也必须实时使用传感器数据来构建并持续更新周边环境地图，从而进行相应的导航。AMR 使用的 AI 功能必须具备低能耗的特点，以免对机器人电池造成太多负担，但同时还必须具备稳定性和低时延，能够与其他工作负载紧密整合。

工业边缘还涉及许多非视觉类的 AI 应用。工厂所有者迫切希望提升应用效率，从而降低总体拥有成本（TCO）并提高产量。影响 TCO 的因素包括尽可能减少故障时间。基于 AI 的机器健康状况评估和预测性维护就可以做到这一点。借助电压、温度、振动和声音等非侵入式传感器，机器学习工作负载可准确地检测新出现的问题，并在问题恶化、影响生产或导致生产中断之前，就对维护或修理需求进行预测。

压铸	纺织	电器和电子产品	手工装配
▪ 破损/缺陷铸件 ▪ 压铸溢流口缺失 ▪ 孔洞堵塞 ▪ 翘曲变形缺陷	▪ 经/纬向缺陷 ▪ 编织缺陷 ▪ 脏点 ▪ 破损性瑕疵	▪ 组件实体不匹配 ▪ 组件类型不匹配 ▪ 组件部件编号不匹配 ▪ 组件位置不匹配 ▪ 电路短缺	▪ 人数检测 ▪ 产量检测 ▪ 工人行为检测

表 2. AI 增强型功能发挥重要作用的工业视觉检查和检测示例

英特尔® FPGA 常用在边缘通过 AI 和机器学习来处理这些传感器信号，以降低工厂的网络带宽需求，并快速识别问题，避免了将未经处理的数据发送至云端，再等待云端发回决策而产生的时延。主流原始设备制造商（OEM）正在使用 AI 来动态计算能效更高的运动路径。例如，控制多轴机器人以尽可能高的效率和尽可能低的能耗在规定时间内将物体从 A 点移动到 B 点，AI 可用于增强闭环控制算法，而低时延 AI 算法和确定性计算能力对此类任务至关重要。英特尔® FPGA 非常适合用于实现 AI 增强型闭环算法。

在全球范围内，工业和制造业客户都非常关心以下这些问题：

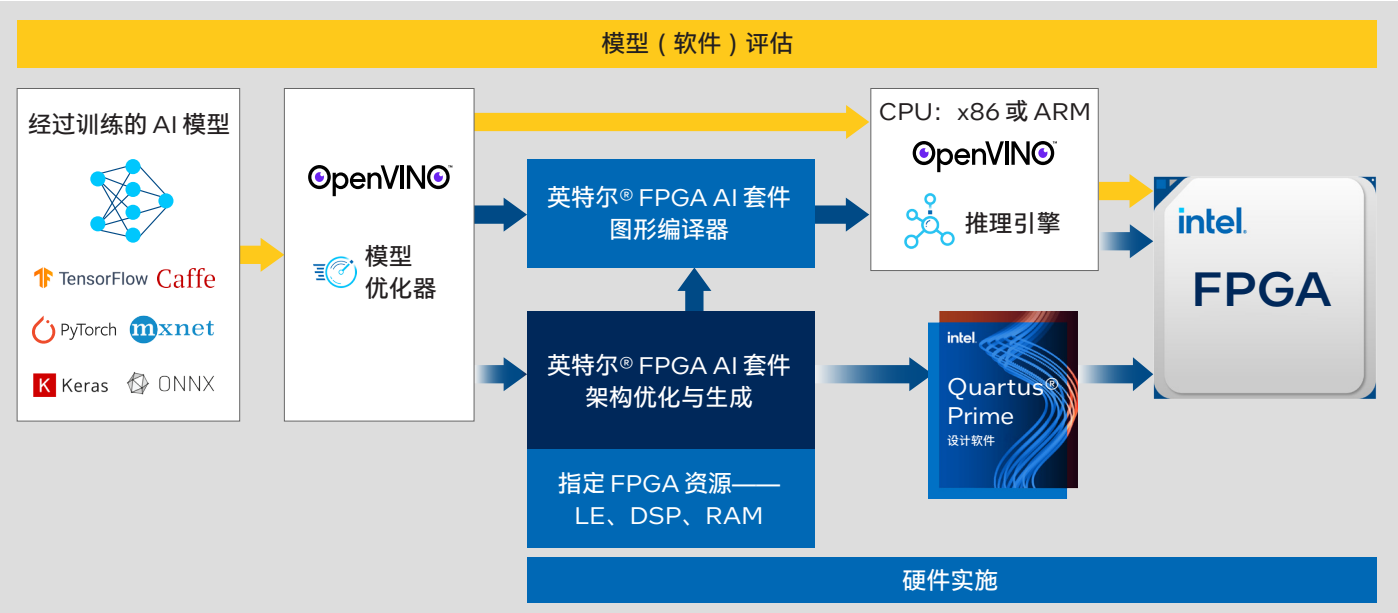
- 如何满足越来越高的产品质量要求？
- 如何优化工厂运营，提高产量和效率？
- 如何更好地预测并缩短设备和系统的故障时间？
- 如何更快地应对和适应市场需求变化？

AI 有助于解决这些问题。正因如此，英特尔® FPGA 正积极在其原有的传感和控制功能基础上新增 AI 功能。这些功能是建立高效工业系统、满足工厂边缘硬实时要求所必备的。

面向英特尔® FPGA 和 SoC 的边缘就绪型 AI 工具套件

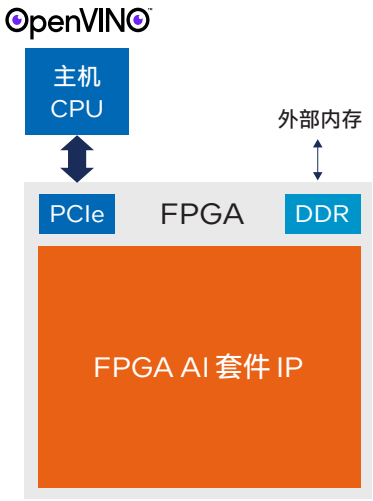
分布式 AI/机器学习边缘解决方案往往十分复杂，开发难度非常高。英特尔提供的开发工具和软件致力于推动开放标准并支持容器化和云原生开发，从而帮助开发人员简化他们的工作流程并加速分布式边缘解决方案的部署。对于使用英特尔® FPGA 和 SoC 进行 AI/机器学习应用开发的人员，英特尔提供以下开发工具：

1. 英特尔® oneAPI AI 分析工具套件（AI 套件）为数据科学家、AI 开发人员和研究人员提供了一套他们熟悉的 Python 工具和框架，可加速端到端的数据科学与分析管道。英特尔通过该工具套件提供支持底层计算优化的 oneAPI 库，从而尽可能地提高从预处理到机器学习的各种工作负载的性能，并提供互操作性以实现高效的模型开发。
2. 英特尔® 分发版 Python 支持 TensorFlow、Keras、PyTorch、oneDNN 和 BigDL 等常用库和框架，可用于在多个英特尔® 计算平台上创建高速的机器学习应用。这些工具可支持面向一系列广泛的 AI/机器学习工作负载快速开发应用。
3. 英特尔® 分发版 OpenVINO™ 工具包支持对边缘计算机视觉用例至关重要的深度学习应用开发。
4. 英特尔® FPGA AI 套件使 FPGA 设计人员、机器学习工程师和软件开发人员能够高效地优化基于英特尔® FPGA 的 AI 设计。利用常见和主流的行业框架（如 TensorFlow、PyTorch）以及 OpenVINO™ 工具包，英特尔® FPGA AI 套件中的实用程序可加速基于 FPGA 的 AI 推理开发，同时还充分利用了英特尔® Quartus® Prime 软件强大且成熟的 FPGA 开发流程。
5. 英特尔® FPGA AI 套件非常灵活，可以针对各种系统级用例进行配置。用户可以通过一键式流程，生成优化的 AI 推理 IP 模块，并集成到英特尔® Quartus® Prime 软件中。用户还可针对英特尔® FPGA AI 套件中的架构优化器指定设备资源（DSP、内存、逻辑单元）和吞吐量。这种独特的定制能力对于探索设计以及嵌入式 AI 应用的尺寸、重量和能耗等维度的优化都至关重要。

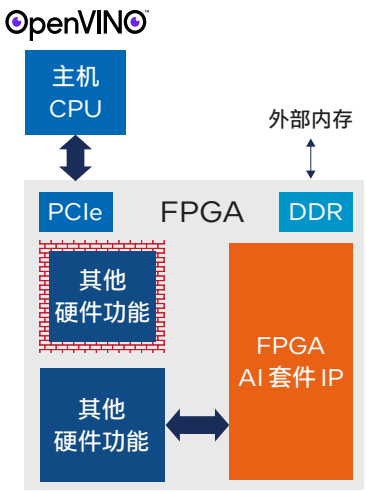


您可通过以下四种方式利用英特尔提供的开发工具，将英特尔® FPGA 和 AI/机器学习集成至您的系统：

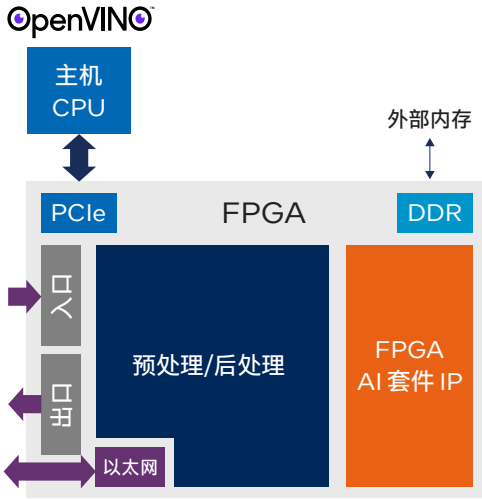
1: 采用基于 FPGA 的 AI/机器学习加速器，实现 CPU 分流。主机 CPU 通过 PCIe 接口与 AI/机器学习加速器进行通信。英特尔® FPGA 直接支持与主机的英特尔® CPU 进行 PCIe 连接。



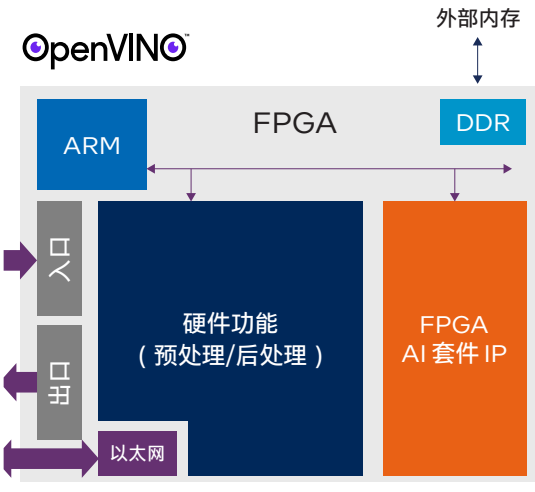
2: 采用英特尔® FPGA 实施 AI 加速器和额外逻辑，实现多功能 CPU 分流。英特尔® FPGA 为主机的英特尔® CPU 提供 AI/机器学习加速，并实施应用所需的任何额外逻辑。与示例 1 一样，主机 CPU 通过 PCIe 接口与 AI/机器学习加速器进行通信。



3: 提取/内联处理 + AI。基于 FPGA 的 AI 加速器直接提取数据，并使用 AI 和算法工作负载处理数据，然后再通过 PCIe 连接将处理过的数据和推理传输至主机的英特尔® CPU。



4: 英特尔® SoC FPGA 利用集成的 CPU（ARM 或 Nios® 处理器内核）充当 AI/机器学习加速器，直接提取并处理数据，实施 AI/机器学习推理，然后通过以太网网络将处理过的数据和推理传输至云端。FPGA 还负责实施应用所需的任何额外逻辑电路。



结论

从构建商、集成商、云和网络提供商到开发人员，在整个边缘价值链中，英特尔深耕数十年，积累了丰富经验。英特尔根据与各类用例客户的合作，开发出了一系列专门用于解决常见集成问题的解决方案，并依托不断优化和创新的成熟开发人员生态系统提供数百个预配置包。您可通过以下方式利用该生态系统缩短开发时间，更快地获得成果：

- **使用可立即部署的企业 AI 解决方案。**英特尔® AI Builders 涵盖全球 300 多个主流 AI 软件、硬件和服务提供商，提供 150 多款解决方案，涉及各种用例和各个市场，使所有企业都能快速采用 AI。
- **确保高水平的 AI 部署。**面向 AI 的英特尔® 精选解决方案利用已在英特尔® 至强® 可扩展处理器和其他英特尔® 平台上进行优化并通过了基准测试和验证的解决方案，帮助您简化和加快基础设施部署。
- **减少开发和协作挑战。**英特尔® AI: In Production 利用英特尔® 技术、软件工具、开发套件、代码样本和广泛的英特尔合作伙伴和开发人员生态系统的解决方案，帮助加快 AI 走向生产之路。
- **英特尔® FPGA AI 套件和 OpenVINO™ 工具包**关注深度学习推理 FPGA IP 创建和集成的简易性，解决了英特尔® FPGA 和 SoC 部署中的“最后一英里”问题。

如果您正在开发需要 AI 功能的边缘或核心设备，欢迎联系当地的英特尔现场销售代表，了解英特尔能够如何为您的团队提供帮助。

参考资料

- Rob van der Meulen，“What Edge Computing Means for Infrastructure and Operations”（边缘计算对基础设施和运营意味着什么），2018 年 10 月 3 日。
- Carl Zimmer，“100 Trillion Connections: New Efforts Probe and Map the Brain’s Detailed Architecture”（100 万亿个连接：全新研究探索并绘制大脑详细结构），《科学美国人》，2011 年 1 月 1 日。
- “Intel® FPGA AI Suite melds with OpenVINO™ toolkit to generate heterogeneous inferencing systems”（英特尔® FPGA AI 套件与 OpenVINO™ 工具包共同发力生成异构推理系统）

WP- 01321-1.0