

FPGA 计算卡

73%~94% 芯片利用率

*基于benchmark测试结果



注: Batch=1, INT 8 以上为基于TensorFlow的神经网络模型,
*KY-SSD和*U-Net为自定义网络

ResNet-50	ResNet-101	VGG 16	Inception-v4	YOLOv3	SSD-FPN	*KY-SSD	*U-Net
4.87 ms	8.76 ms	21.49 ms	17.87 ms	38.48 ms	113.33 ms	2.97 ms	445.18 ms
205.7 FPS	114.2 FPS	46.5 FPS	55.9 FPS	25.9 FPS	8.8 FPS	337.5 FPS	2.2 FPS

一款面向边缘端、服务器端进行深度学习推断的FPGA计算加速卡

基于鲲云自主研发的定制数据流架构CAISA 3.0, 最高可实现94%的芯片利用率。

NA-100c采用PCIe 3.0 x8接口, 功耗低于60 W, 无需额外供电。用户可通过RainBuilder编译工具链将TensorFlow、Caffe、PyTorch及ONNX(MXNet)等框架开发的算法模型无缝部署到星空加速卡上, 为支持深度学习的边缘端及数据中心提供高性能、低延时、高算力性价比的计算加速方案。

FPGA	Intel® Arria® 10 GX 1150
峰值性能	1.64 TOPS
芯片利用率	73%~94%
接口	PCIe 3.0 x8
内存	板载8GB DDR4
电源	预留 PCIe 6-pin 12V 外部电源接口*
功耗	< 60 W (12V, @5A)
散热方式	主动散热 (单风扇)
工作温度	5°C~60°C (环境温度)
形态尺寸	169.5 mm x 68.7 mm (PCIe半高半长标准, 单槽位)

应用
设备



小型主机



服务器

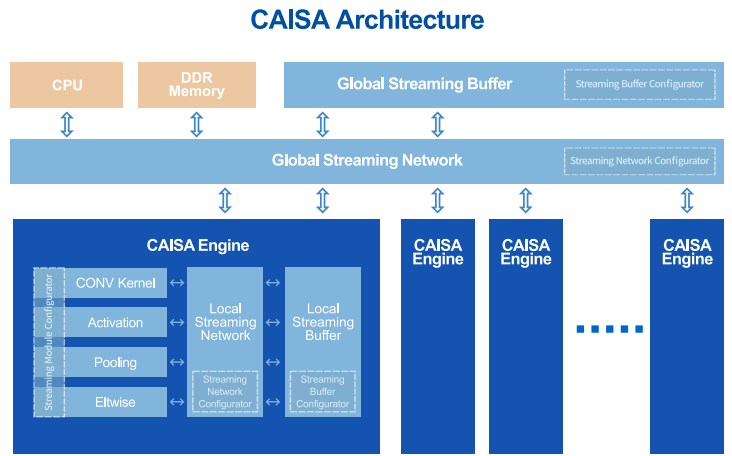


工控机

*通常标准的PCIe插槽可提供75W电源; 此处预留功能可为不同系统设定提供电源需求

1 定制数据流 CAISA[®] 架构

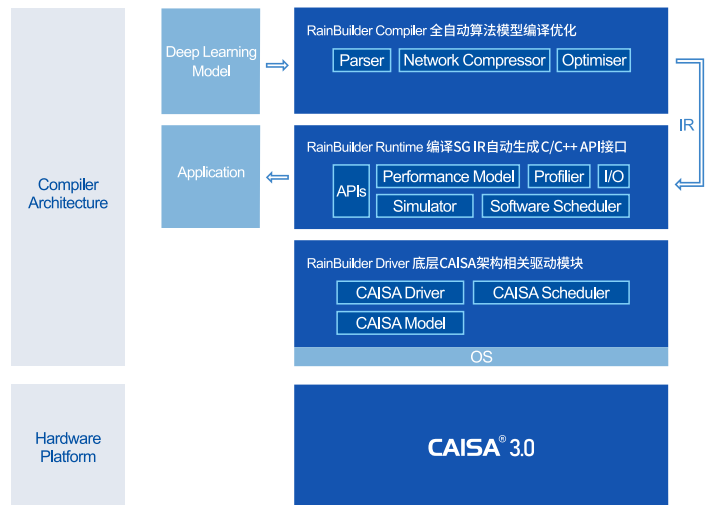
- 高算力性价比：时钟级精确的计算
- 高通用性：数据流动态重组
- 高软件易用性：算法端到端自动化部署



2 端到端编译工具链 RainBuilder[®]

RainBuilder是一款针对CAISA架构的深度学习算法编译工具链，兼容TensorFlow、Caffe、PyTorch、ONNX等主流开发框架，开发者只需两步，即可实现神经网络模型推演在鲲鹏AI加速卡上的高效部署及运行。

- 支持TensorFlow、Caffe、PyTorch、ONNX等主流框架下深度学习算法无缝链接
- 通用性高，支持VGG、YOLO、ResNet等多种CNN类算法模型
- 简单易用，支持基于C++/Python的标准开发流程，无需了解芯片架构



应用步骤

