

Nebula Accelerator X3

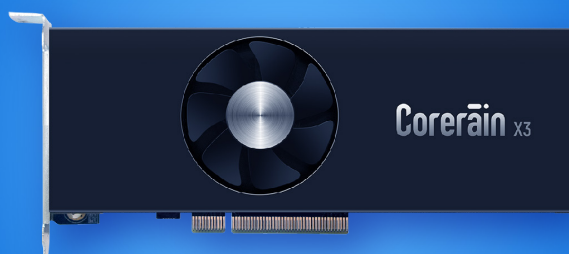
星空 X3 加速卡

Corerain
鲲云科技

搭载 CAISA 芯片

高达 95.4% 芯片利用率

* 基于 benchmark 测试结果



ResNet-50	ResNet-152	VGG19	Inception-v4	YOLOv3	SSD-ResNet50	SSD-FPN	*KY-SSD	*U-Net	DeepLabv3+	U-Net Industrial
3.06 ms	8.68 ms	18.33 ms	14.6 ms	31.06 ms	21.96 ms	95.98 ms	2.31 ms	343.01 ms	119.37 ms	74.07 ms
1306.93 FPS	460.27 FPS	218.01 FPS	273.75 FPS	125.75 FPS	182.16 FPS	40.71 FPS	1676.19 FPS	11.39 FPS	32.73 FPS	54.01 FPS
92.32%	95.43%	78.53%	69.21%	82.37%	77.06%	70.64%	80.80%	85.37%	67.08%	64.97%

注: Batch=4, INT8, 百分比为芯片利用率, 以上为基于 TensorFlow 的神经网络模型, *KY-SSD 和 *U-Net 为自定义网络



高性能低延时



高通用性



高算力性价比



简单部署

一款面向边缘端、服务器端进行深度学习推断的AI专用计算加速卡

基于全球首款数据流AI芯片, 提供10.9 TOPS的算力, 最高可实现95.4%芯片利用率。

X3加速卡采用PCIe 3.0 x8接口, 可与X86架构、Arm架构服务器兼容。用户可通过RainBuilder编译工具链将TensorFlow、Caffe、PyTorch及ONNX等框架开发的算法模型无缝部署到星空加速卡上, 为支持深度学习的云端及边缘服务器提供高性能、低延时、高算力性价比的计算加速方案。

主芯片 **CAISA**
峰值算力 **10.9 TOPS**
芯片利用率 **高达 95.4%**

接口 **PCIe 3.0 x8**
内存 **板载 8GB DDR4**

功耗 **峰值功耗 56 W, 平均动态功耗 23.8 W**
电源 **PCIe 供电**
散热方式 **主动散热 (单风扇)**

工作温度 **-20 ~ 70°C (环境温度)**
外形尺寸 **167.50mm x 68.85mm x 20.00mm**
(PCIe 全高全长标准, 单槽位)

应用
设备



小型主机



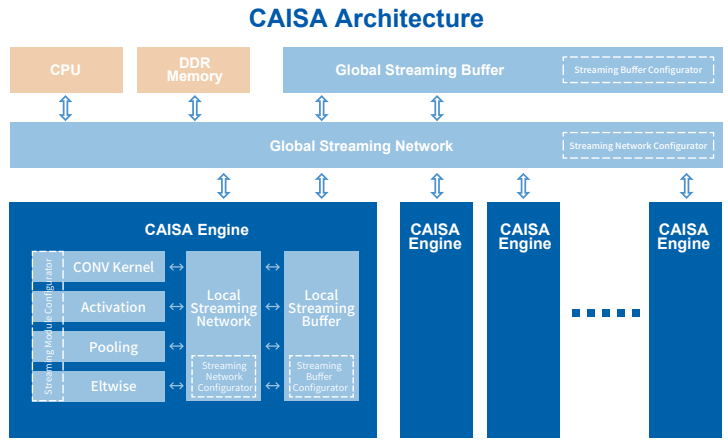
服务器



工控机

1 定制数据流 CAISA® 架构

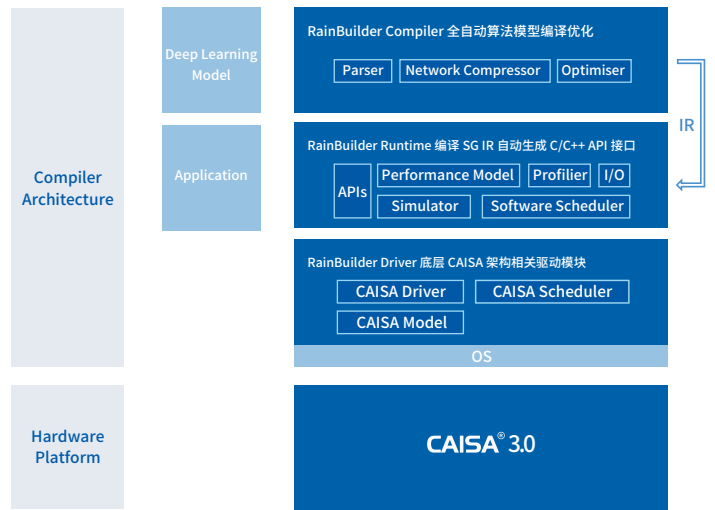
- 高算力性价比：时钟级精确的计算
- 高通用性：数据流动态重组
- 高软件易用性：算法端到端自动化部署



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

