

# Nebula Accelerator X9

## 星空加速卡 X9

Corerain  
鲲云科技

搭载 CAISA 芯片

高达 95.7% 芯片利用率

\* 基于 benchmark 测试结果



ResNet50	ResNet101	ResNet152	VGG16	VGG19	Inception V4	Yolo V3	SSD-FPN	*KY-SSD	*U-Net	*DeepLab V3+	SSD-ResNet50	U-Net Industrial
3.06 ms	5.86 ms	8.67 ms	15.18 ms	18.4 ms	14.51 ms	31.3 ms	96.81 ms	2.33 ms	335.85 ms	112.73 ms	21.83 ms	73.18 ms
5235.5 FPS	2730.6 FPS	1846.0 FPS	1054.5 FPS	869.9 FPS	1103.0 FPS	511.2 FPS	165.2 FPS	6868.8 FPS	47.6 FPS	141.9 FPS	733.1 FPS	218.6 FPS
92.5 %	94.8 %	95.7 %	74.8 %	78.3 %	69.7 %	83.7 %	71.7 %	82.8 %	89.3 %	72.7 %	77.5 %	65.7 %

注: Batch=16, INT8, 百分比为芯片利用率, 以上为基于 TensorFlow 的神经网络模型



高性能低延时



高通用性



高算力性价比



简单易部署

### 面向智算中心及高性能服务器端的 AI 专用计算加速卡

搭载 4 颗全球首款数据流 AI 芯片, 提供 43.6TOPS 峰值算力。

最高可实现 95.7% 芯片利用率, 3 毫秒内的延时。

采用 PCIe3.0x16 接口, 全高全长, 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。

主芯片

CAISA x 4

架构

定制数据流架构 CAISA 3.0

芯片利用率

高达 95.7%

峰值算力

43.6 TOPS

PCIe

PCI Express3.0 x16

内存

板载 32GB DDR4

电源

PSU 供电

散热方式

被动散热 (无风扇)

形态尺寸

313.9mmx 106.65 mm (PCIe 全高全长标准, 双槽位)

应用  
设备



智算中心



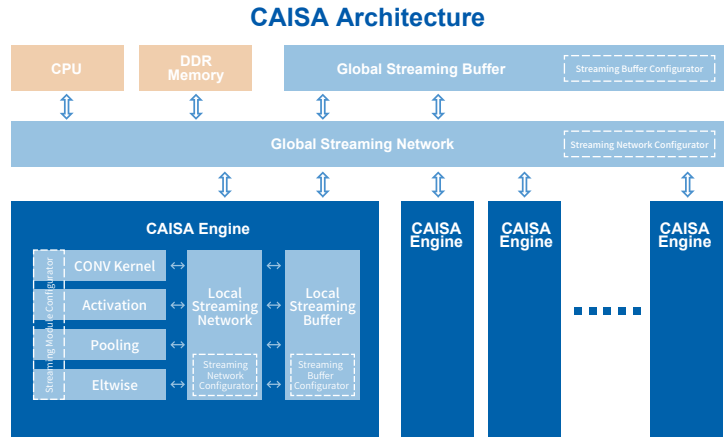
云计算中心



高性能计算

## 1 定制数据流 CAISA® 架构

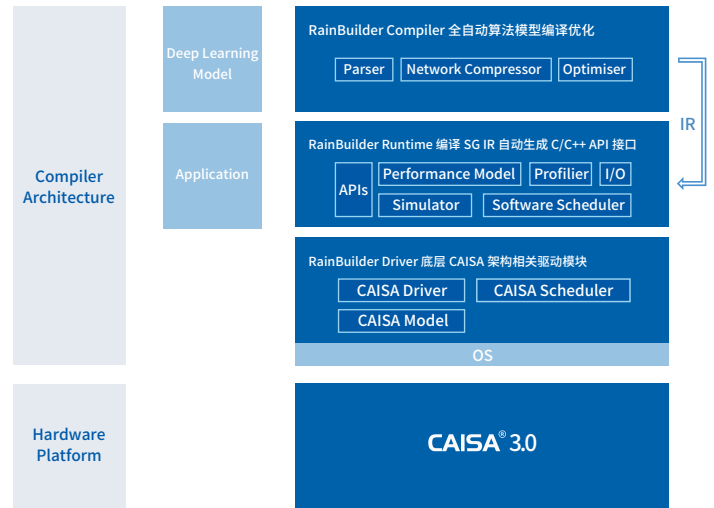
- 高算力性价比：时钟级精确的计算
- 高通用性：数据流动态重组
- 高软件易用性：算法端到端自动化部署



## 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

