

# Nebula Accelerator X3

## 星空 X3 加速卡

Corerain  
鲲云科技

搭载 CAISA 芯片

芯片利用率高达 95.4%

\* 基于 benchmark 测试结果



| ResNet-50              | ResNet-152            | YOLOv3                 | SSD-ResNet50           | U-Net Industrial      |
|------------------------|-----------------------|------------------------|------------------------|-----------------------|
| 3.06 ms<br>1306.93 FPS | 8.68 ms<br>460.27 FPS | 31.06 ms<br>125.75 FPS | 21.96 ms<br>182.16 FPS | 74.07 ms<br>54.01 FPS |

\* Batch=4, INT8, 以上为基于 TensorFlow 的神经网络模型

\* 测试结果来源于中国人工智能产业发展联盟 (AIIA) 委托中国信息通信研究院完成的 DNN Benchmark 测评项目 <http://www.aiaa.org.cn/benchmark/zh/top1.html?now=5>



四核 CAISA 引擎



高算力性价比



毫秒级延时

星空 X3 加速卡是一款深度学习推断专用计算加速卡, 提供 10.9 TOPS 算力, 最高可实现 95.4% 芯片利用率。

星空 X3 加速卡采用 PCIe 3.0 x 8 接口, 兼容 x86、Arm 架构服务器。可通过 RainBuilder 编译工具链将 TensorFlow、Caffe、PyTorch、ONNX (MXNet、PaddlePaddle) 等框架开发的算法模型部署到星空 X3 加速卡上, 为支持深度学习的服务器提供高性能、低延时、高通用性、高算力性价比的计算加速方案

主芯片 CAISA  
峰值性能 10.9 TOPS  
芯片利用率 最高 95.4%

接口 PCIe 3.0 x8  
内存 板载 8GB DDR4

功耗 TDP 56 W, 典型功耗 23.8 W  
电源 PCIe 供电  
散热方式 主动散热

工作温度 -20 ~ 70°C  
外形尺寸 PCIe 半高半长标准, 单槽位

应用  
设备



小型主机



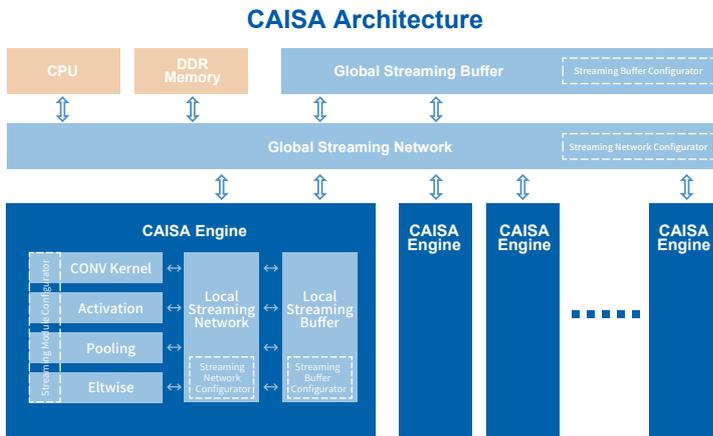
服务器



工控机

## 1 定制数据流 CAISA® 架构

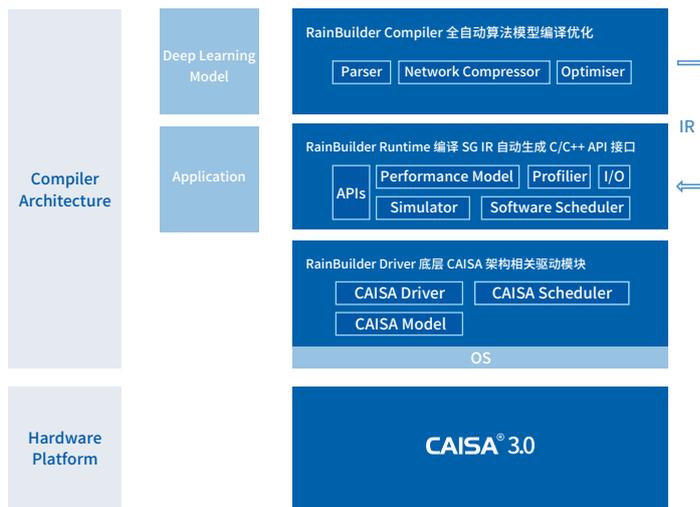
CAISA 3.0 是鲲云科技为深度学习神经网络定制的高性能 AI 计算架构，通过数据流流动次序来控制计算顺序，消除指令操作导致的额外开销，让深度学习网络的算子级数据流图可以实现高效流水线运算。



## 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(MXNet、PaddlePaddle) 等主流开发框架，可实现神经网络模型在鲲云星空系列产品上的高效部署及运行。

- 无缝链接 支持 TensorFlow、Caffe、PyTorch、ONNX(MXNet、PaddlePaddle) 等深度学习开发框架
- 简单易用 支持 C++/Python 的 API，简单两步，无需了解复杂的底层技术知识
- 通用性高 支持 ResNet、YOLO、DeepLab 等主流算法推断模型



## 应用步骤

