



RISC-V融合AI技术的芯片 实践和应用

进迭时空 K1 AI-CPU的产品化

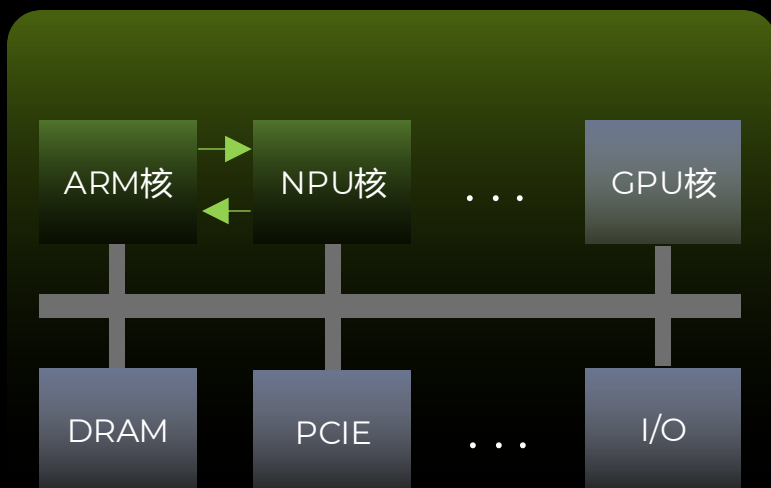


- 一、RISC-V融合AI技术
- 二、进迭时空K1 AI-CPU
- 三、支持模型及实践
- 四、落地应用

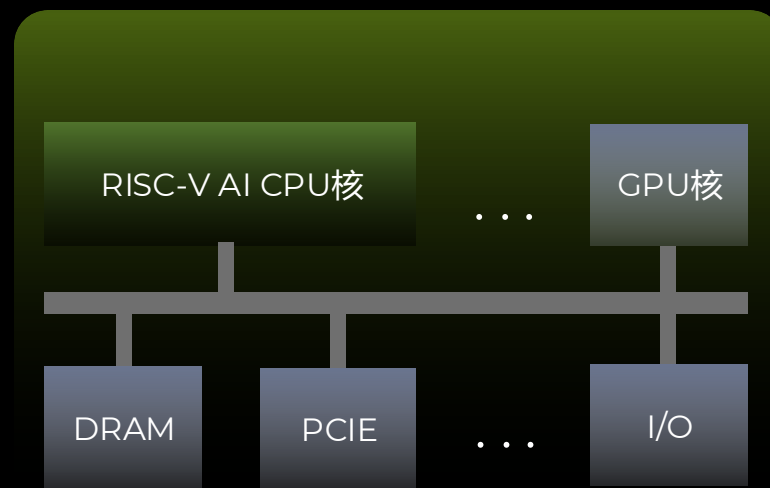
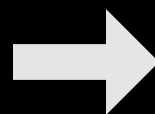


RISC-V融合AI技术

ARM架构中心控制模式



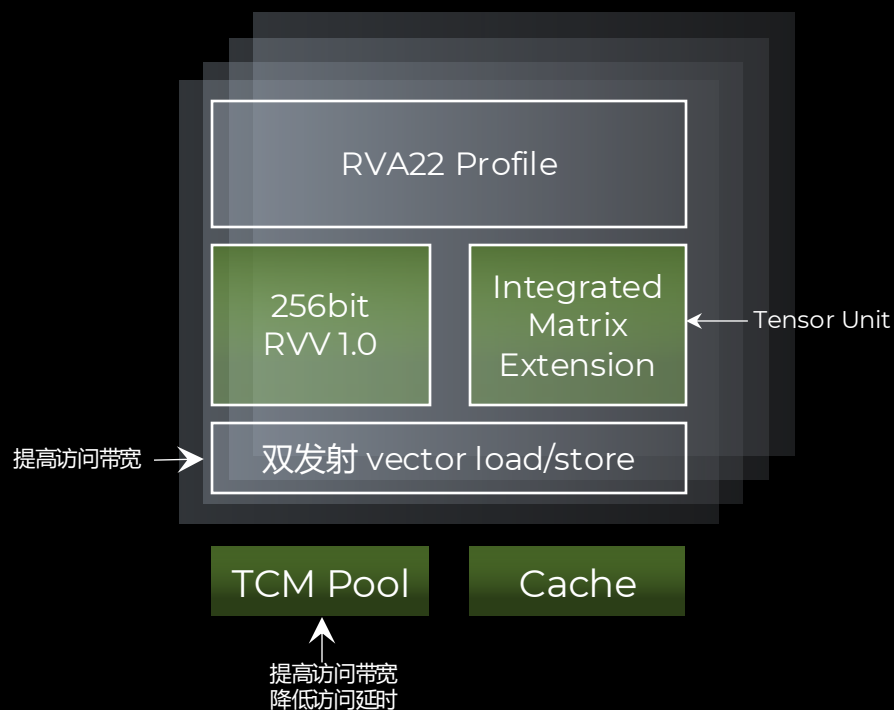
RISC-V架构创新模式



- 以CPU形式提供AI算力
- 融入主流的CPU推理生态



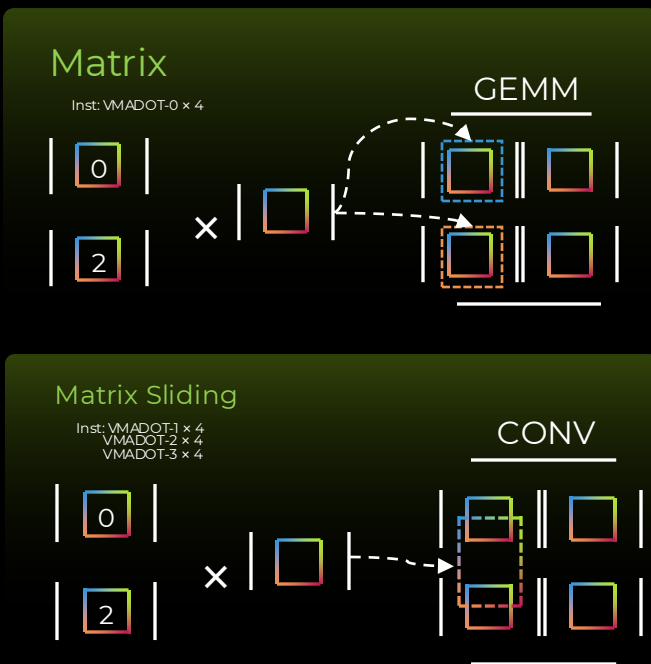
X60™：原生AI算力的智算核



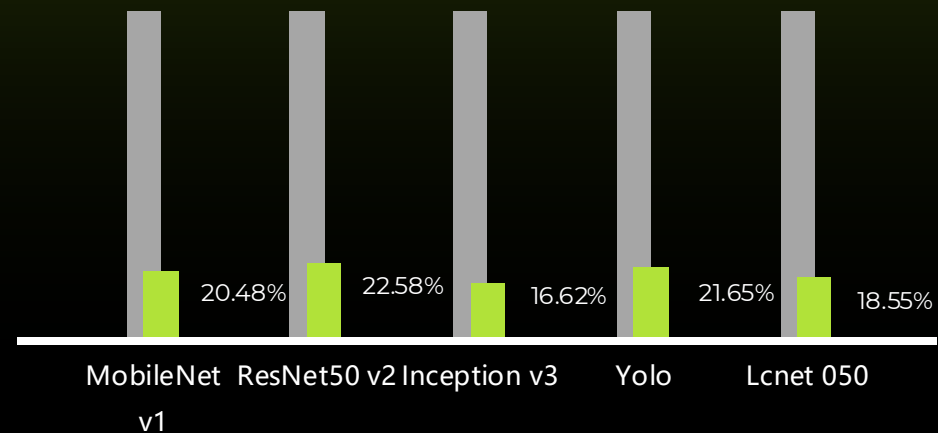
- 全球首款支持256bit RVV1.0的RISC-V处理器
- 双发射8级按序
- 双发射的Vector load/store
- RVA22 Profile + Vector 1.0 + IME (Integrated Matrix Extension)
- Cache+TCM的双缓存架构
- 复用RVV寄存器，扩展16条IME指令



X60™ 扩展 16 条 AI 指令



相同AI算法
X60™ 指令仅为Arm Neon的20%



使用RWV1.0加速的软件库和SpacemitT IME指令





SpacemiT Key Stone™ K1

- 自研RISC-V智算核X60™，RVA22 Profile
- 8核同构X60™，频率最高2.0GHz
- CPU算力>50 KDMIPS，AI算力 2.0 TOPS@INT8
- 单核算力是ARM Cortex-A55的1.3倍
- 接入全球主流的CPU推理生态，支持所有AI模型部署
- 最大16GB LPDDR4X
- 典型功耗3.5W



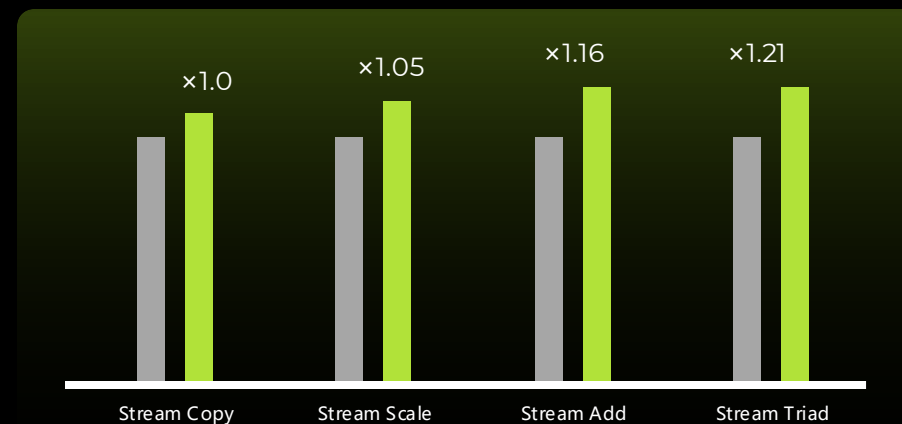
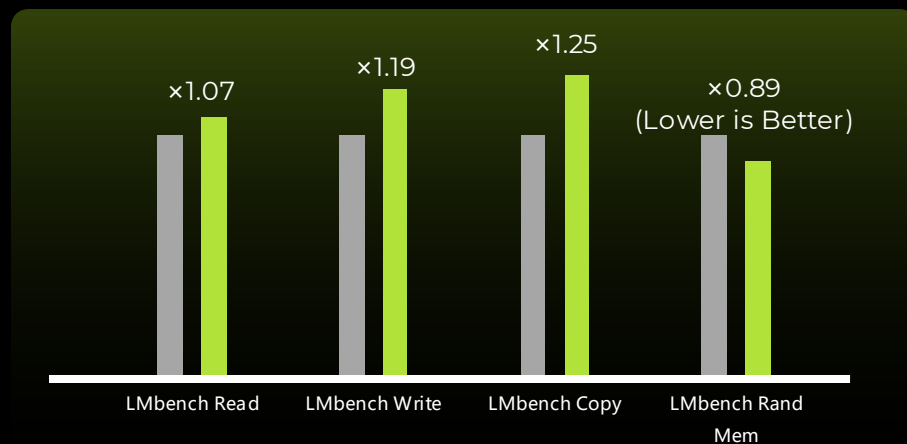
X60™智算核性能指标

存储性能 | 运算性能 | 浮点性能 | 能耗指标



存储性能对比

Relative Performance (Higher is Better)

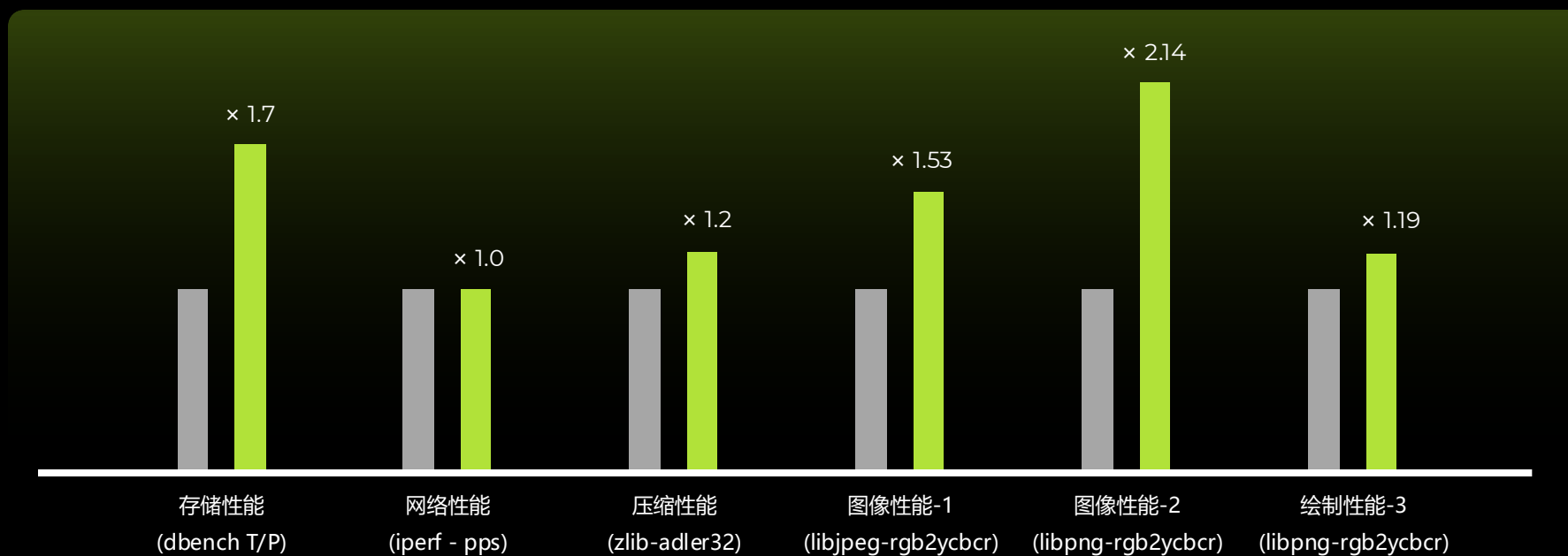


使用SpacemiT LLVM Compiler



运算性能对比

Relative Performance (Higher is Better)



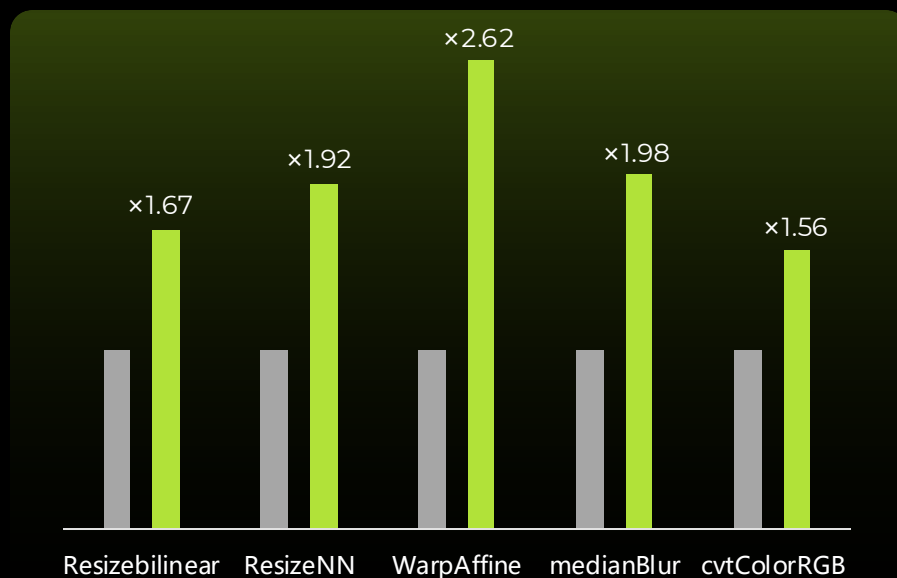
使用SpacemiT LLVM Compiler和RVV 1.0加速的软件库



向量计算性能对比

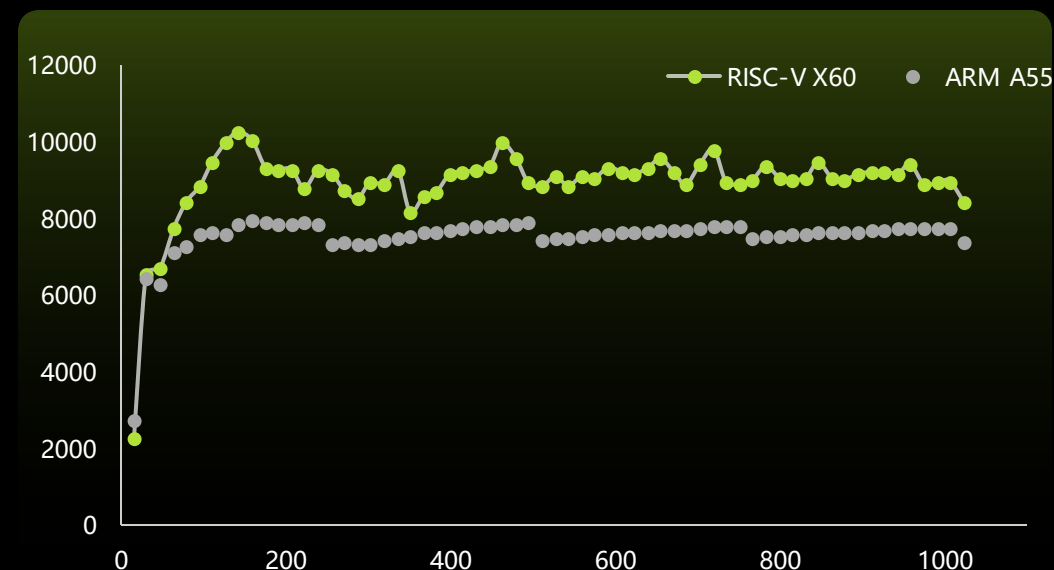
Relative Performance (Higher is Better)

OpenCV性能



OpenCV (K1 X60 vs 某款主流ARM A55芯片): 基于RVV 1.0加速的OpenCV, 如预处理、颜色空间转换、图形学等优化图像处理场景的性能

Eigen-SGEMM性能

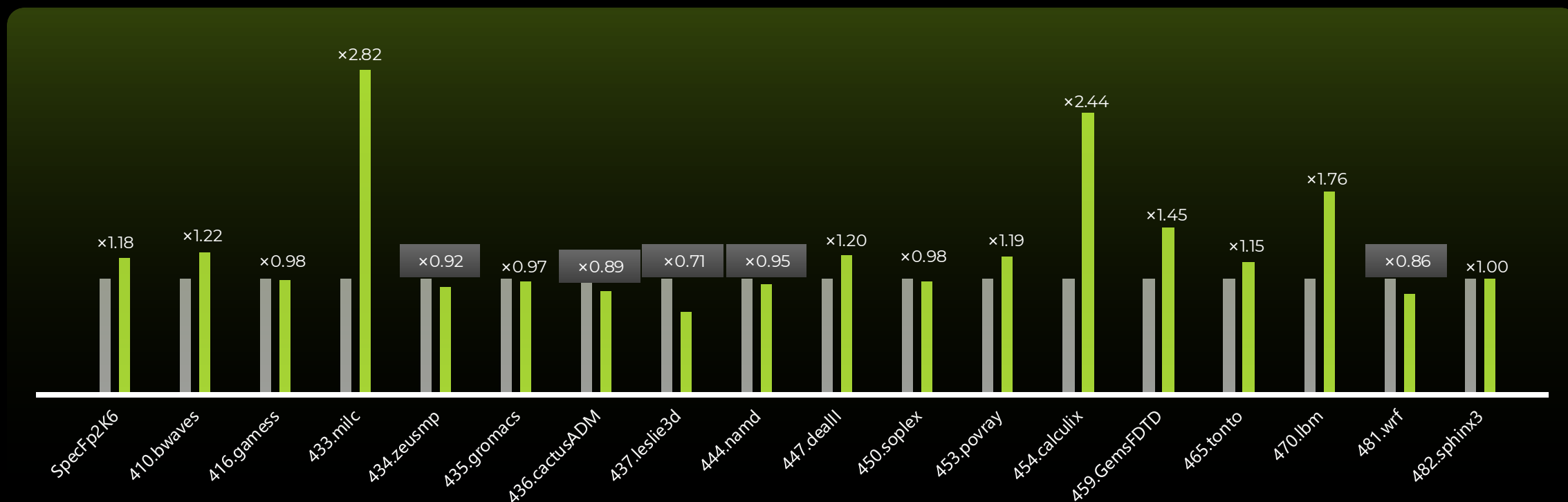


使用SpacemiT LLVM Compiler和RVV 1.0加速的软件库



浮点性能对比

Relative Performance (Higher is Better)

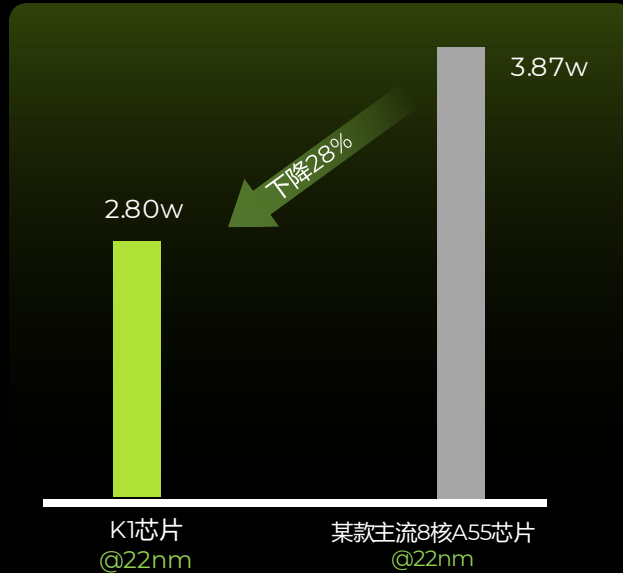


使用SpacemiT LLVM Compiler , opts : -ofast ...



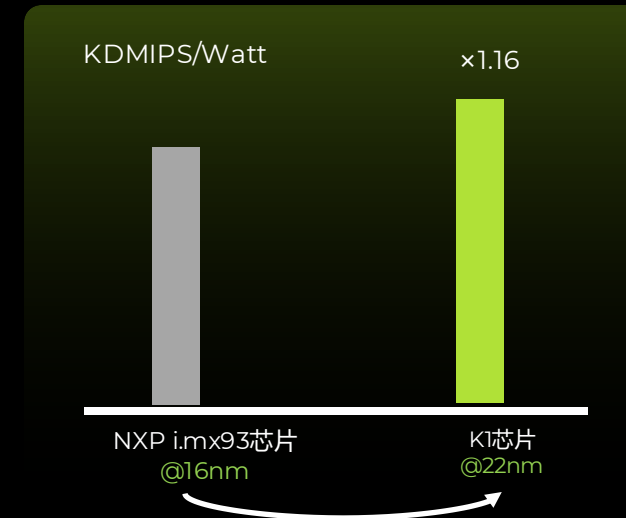
功耗对比

同等工艺
与某款主流8核A55芯片对比



测试场景：Linux下8核运行stress-ng

代差工艺
与NXP A55芯片能效对比



超越代际的能效表现

A55@16nm功耗

The screenshot shows two tables from an NXP document. The top table, 'POWERPERFORMANCE BENCHMARK (IN LINUX 5.10.71)', compares the A55 chip with other processors. The bottom table, 'Power Measurements', shows power consumption in different states.

芯片	频率	核心数	功耗 (W)	性能 (KDMIPS/Watt)
A55	1.7GHz	8	4.45	0.81 (0.0075)
A55	1.7GHz	8	4.45	0.81 (0.0075)
A55	1.7GHz	8	4.45	0.81 (0.0075)

芯片	频率	核心数	功耗 (W)	性能 (KDMIPS/Watt)
A55	1.7GHz	8	4.45	0.81 (0.0075)
A55	1.7GHz	8	4.45	0.81 (0.0075)
A55	1.7GHz	8	4.45	0.81 (0.0075)

Source: <https://www.nxp.com/design/training/nxps-latest-ultra-low-power-and-edge-ai-application-processors-i-mx-8ulp-i-mx-93:TIP-230606-NXP-Documents:INTRODUCTION-TO-i.MX-93-APPLICATIONS-PROCESSOR-FAMILY>, Page 18.

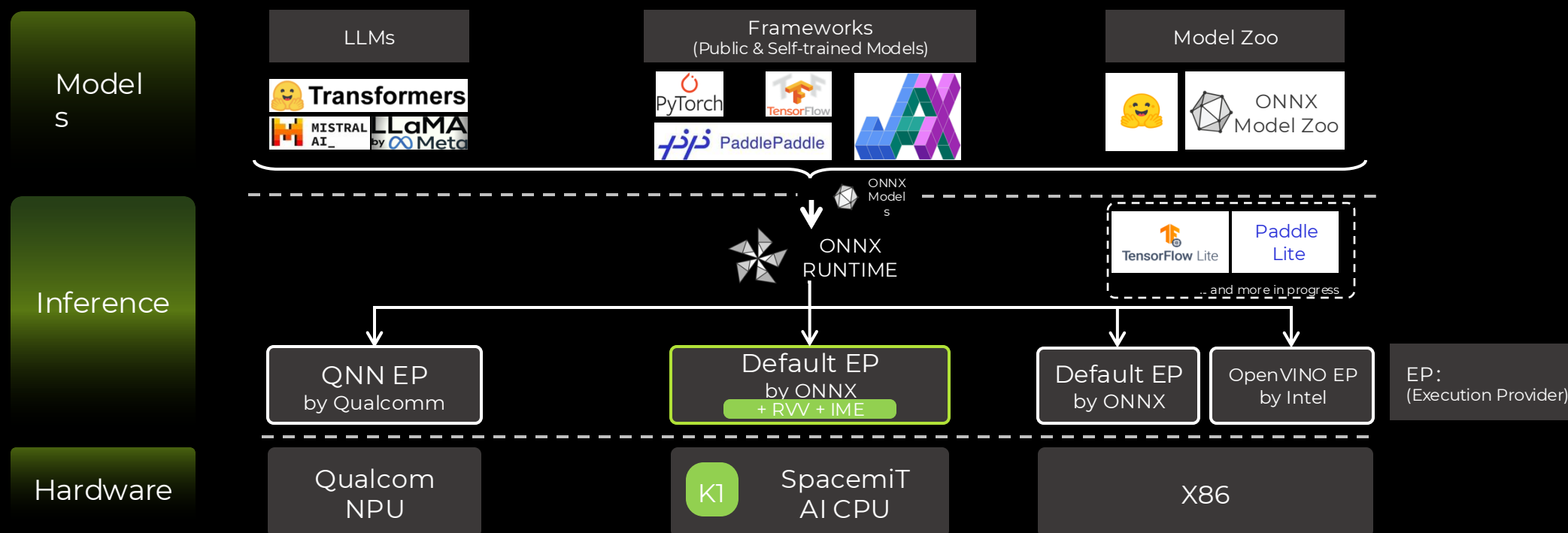


K1 支持模型及实践

部署框架 | 模型适配 | 系统生态

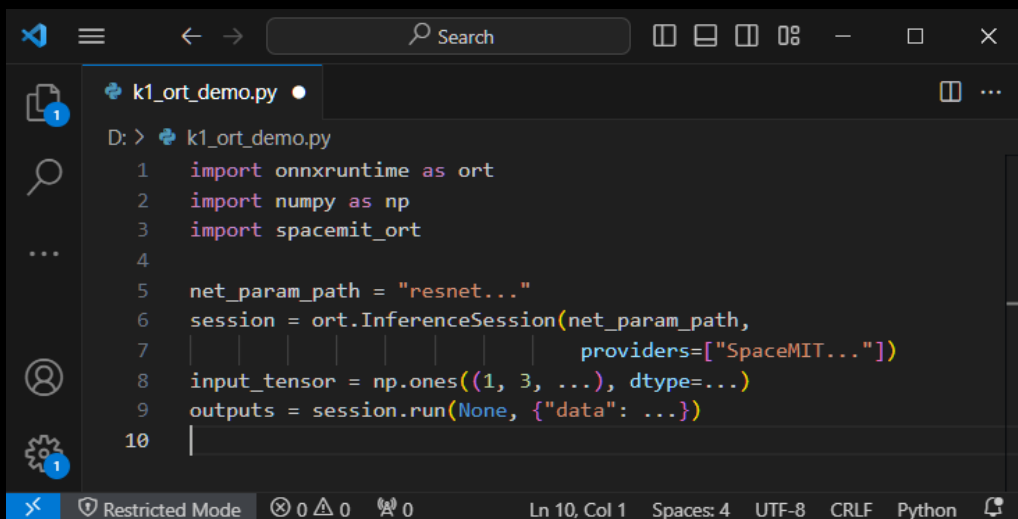


AI开发接入全球主流AI推理生态



对AI模型的适配能力

轻松部署



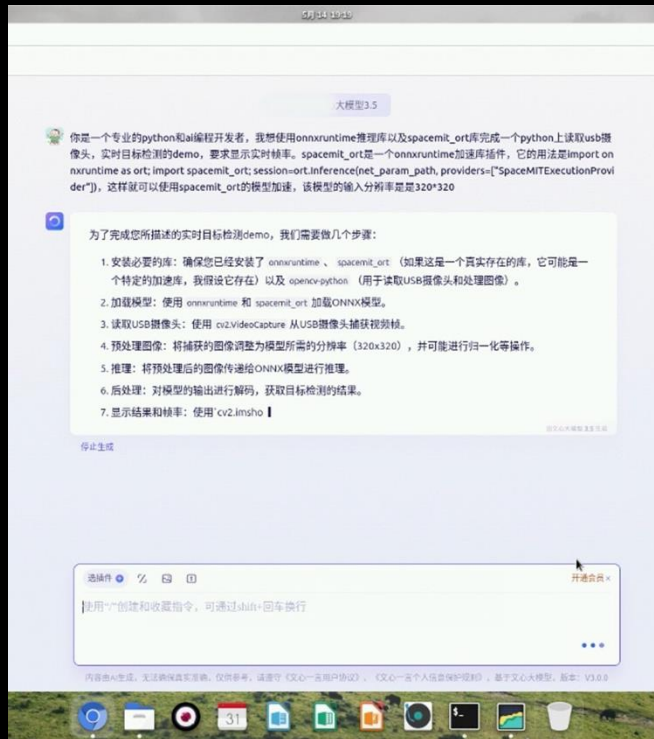
```
k1_ort_demo.py
D: > k1_ort_demo.py
1 import onnxruntime as ort
2 import numpy as np
3 import spacemit_ort
4
5 net_param_path = "resnet..."
6 session = ort.InferenceSession(net_param_path,
7                               providers=["SpaceMIT..."])
8 input_tensor = np.ones((1, 3, ...), dtype=...)
9 outputs = session.run(None, {"data": ...})
10
```

可运行所有AI模型

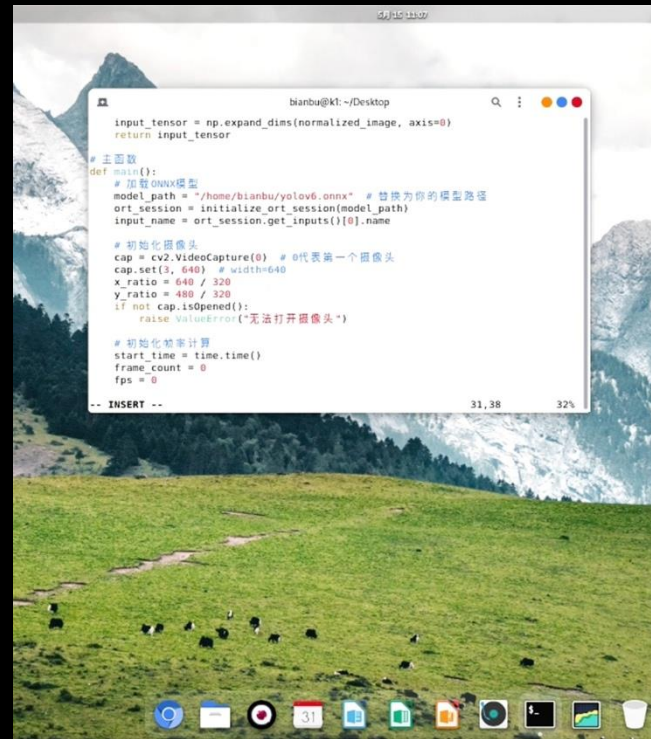
模型类型	常见模型	K1-X60	MLPerf Edge	某ARM芯片的NPU	Google EdgeTPU
图像分类/Image Classification	MobileNet	✓	✓	✓	✓
对象检测/Object Detection	SSD-MobileNetV1, SSD-ResNet34	✓	✓	✓	✓
图像分割/Image Segmentation	3DU-Net	✓	✓	✓	×
自然语言处理 / NLP	BERT	✓	✓	×	×
语义分割/Semantic Segmentation	DeepLabV3-MobileNetV2	✓	✓	×	✓
语音识别/Speech Recognition	RNNT	✓	✓	×	✓
大模型/LLM	Llama-2-7	✓	×	×	×
姿态估计/Pose Estimation	PoseNetMobileNetV1	✓	×	×	✓



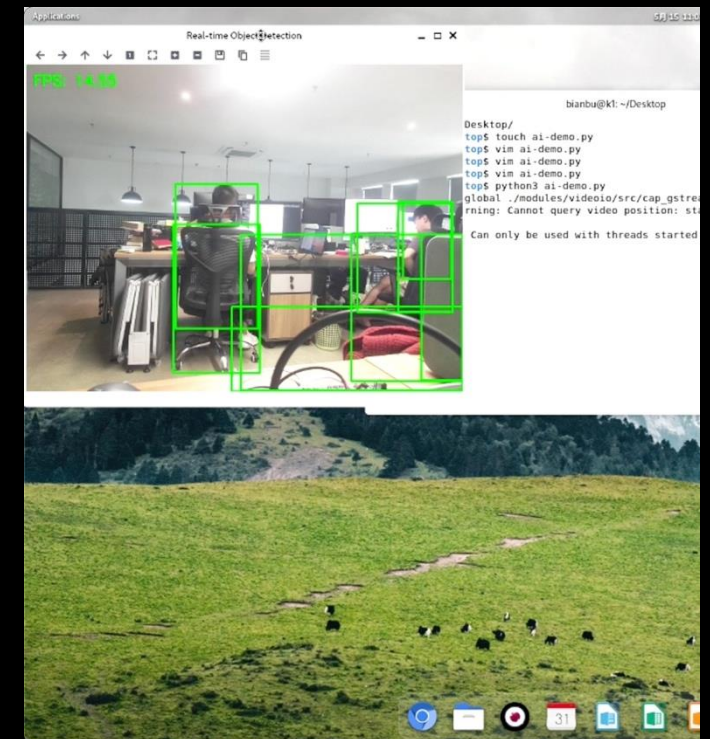
轻松部署



给大模型喂入spacemit_ort
库API以及功能要求



复制代码到python, 填入模
型路径、依赖设备号



执行脚本, 完成部署





多模型适配

SpacemiT在K1上已经适配：

类别	模型	参数量
NLP	qwen1.5	4B
	qwen2	1.5B
	qwen2	0.5B
	phi3	3.8B
	phi2	2.7B
	minicpm-llama-1b	1B
	minicpm-llama-2b	2B
	tinyllama	1.1B
	gemma	2B
	chatglm2.0	6B
	llama3-cn-8b	8B
MultiModel	MobileVLM-V2	1.7B
	MiniCPM-V2	2.8B



RISC-V系统生态已具备量产条件

Linux发行版



进迭自研OS



RTOS



嵌入式Linux



AI推理框架



编程语言



应用开发框架



应用



MPV

编译器



Ready

Developing



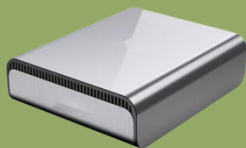
RISC-V 融合AI落地应用

K1应用领域 | AI应用案例



K1芯片应用领域

NAS



智能机器人



边缘计算



SBC



笔记本电脑



工业控制



...



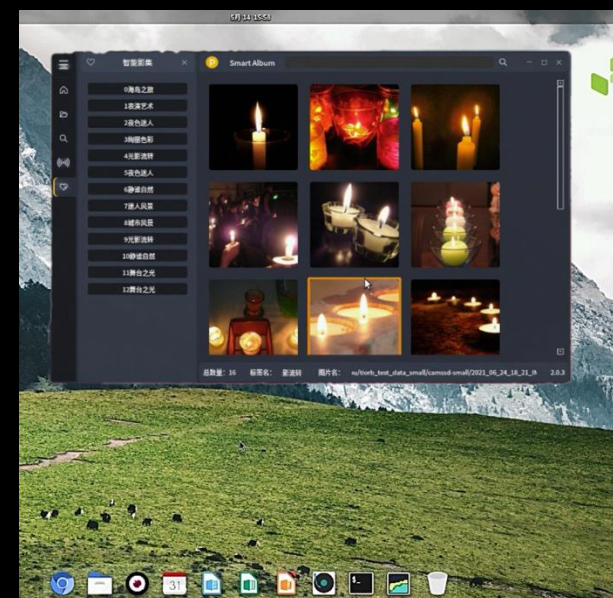
智能图库

- 基于传统CNN+Transformer的组合应用--【AI通用算力】
- AI多模态带来的新体验：文搜图、文生图、图搜图



智能图库

Nii图文搜索 | 智能影集生成 | 场景分类 | 人脸聚类算法 | 照片检测



✓ 在K1中完成适配并商用



机器视觉

- 双目SLAM (Vector) : K1的向量能力可构建半稠密世界观
- NeRF (CNN) : 基于图像生成3D场景新视图



机器视觉

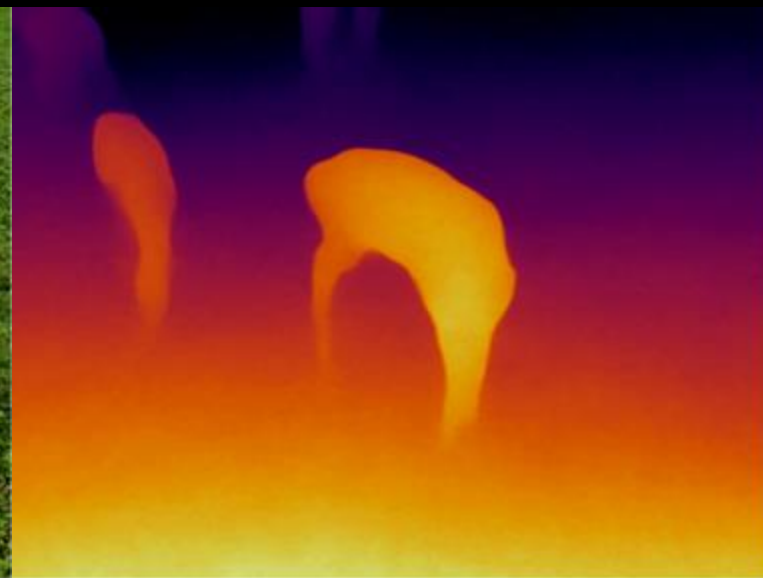
半稠密点云建图



Raw image



Depth Anything



✓ 在K1中完成移植调优



端侧大模型

- 基于LLM行业应用：交通、校园等行业的新AI升级
- Robot Computer：自主意识机器人

✓ K1正在开发中



RISC-V + AI, 驱动数智未来





THANK YOU!



进迭时空 - RISC-V高性能芯片



spacemit.com