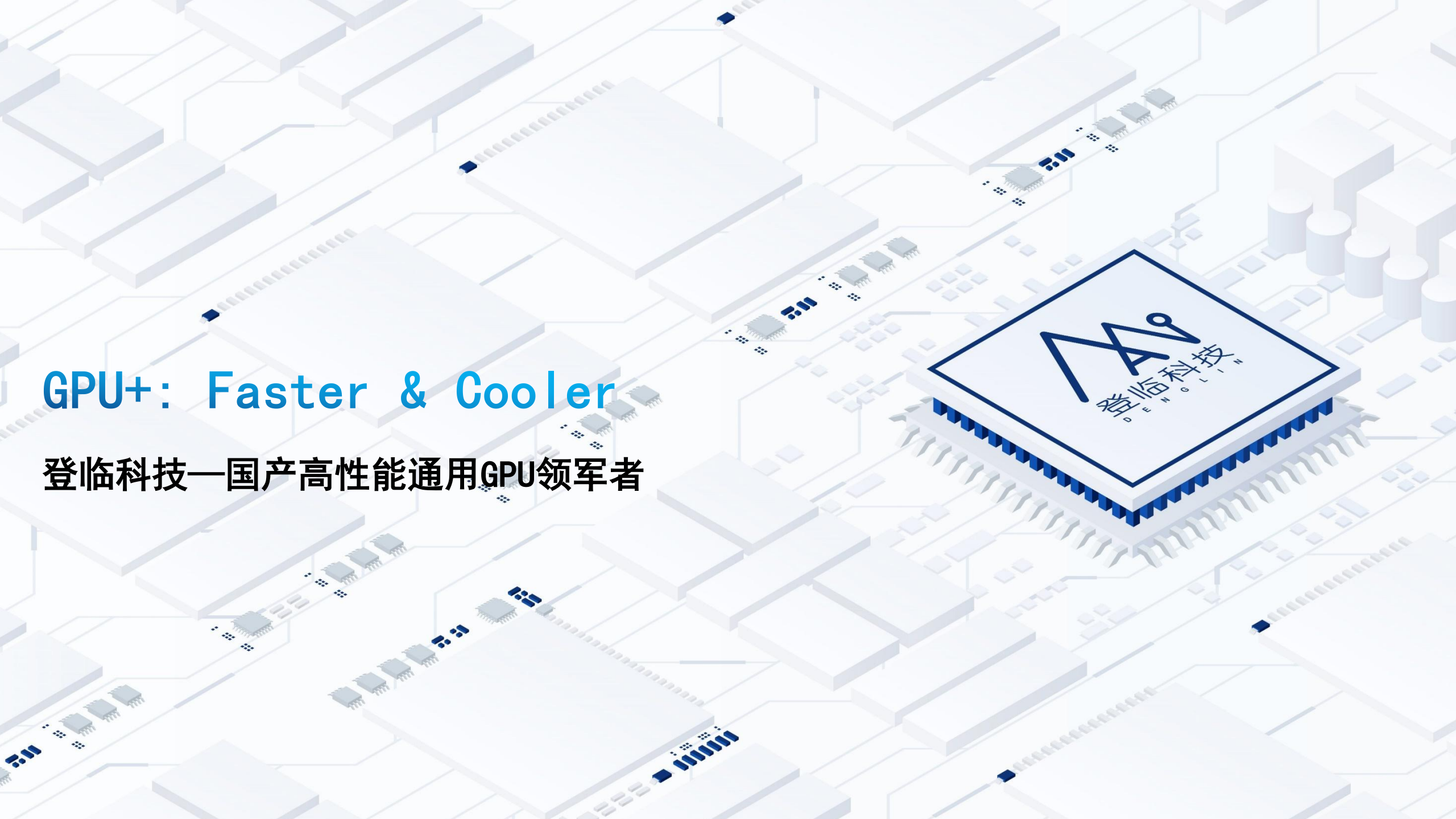
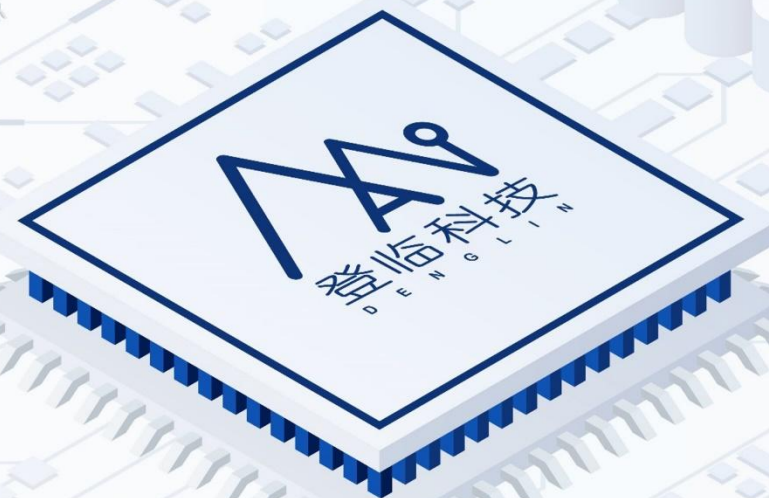


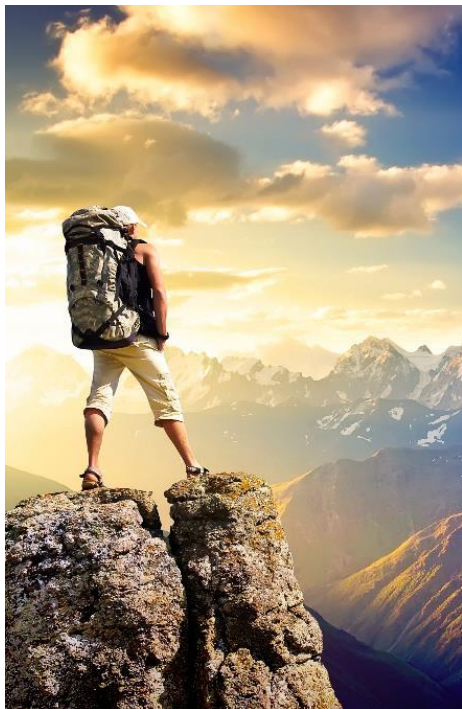


GPU+: Faster & Cooler

登临科技—国产高性能通用GPU领军者



公司介绍



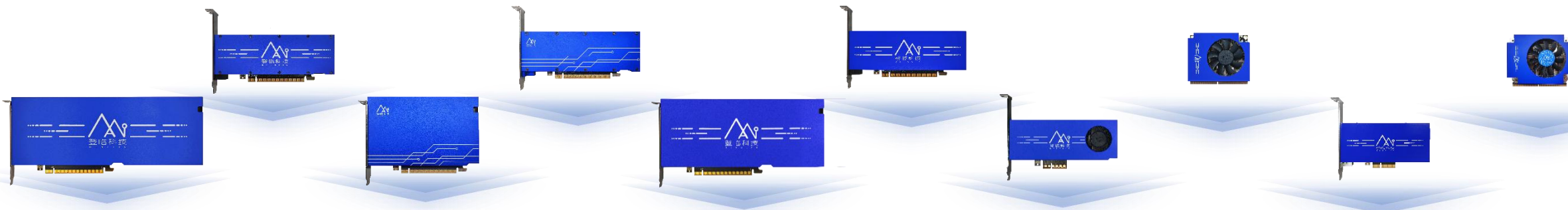
登临成立于2017年底，专注于**高性能、通用计算平台的芯片研发与技术创新**，致力于打造云边端一体、软硬件协同、训练推理融合的前沿芯片产品和平台化基础系统软件。

核心团队成员来自图芯、S3、英伟达、AMD、海思、阿里等全球知名的芯片设计公司，人均拥有**超过20年**的GPU研发和产品商业化的经验，产品运营团队来自苹果、华为、思科、飞利浦等国际性企业，精通高质量、高可靠性产品运营与供应链管理。

总部位于**苏州**，在**上海、北京、成都、杭州**和**深圳**各地均设有研发中心。

登临产品线：

- 推理、训练
- 云边端
- 从人工智能向自动驾驶，高性能计算，图形加速等领域扩展



公司优势 | 核心技术

核心技术

可扩展的、软件定义的、片内异构的体系结构



硬件兼容现有的软件生态 (CUDA/OpenCL等)



算法模型平滑迁移, 极大节约移植成本

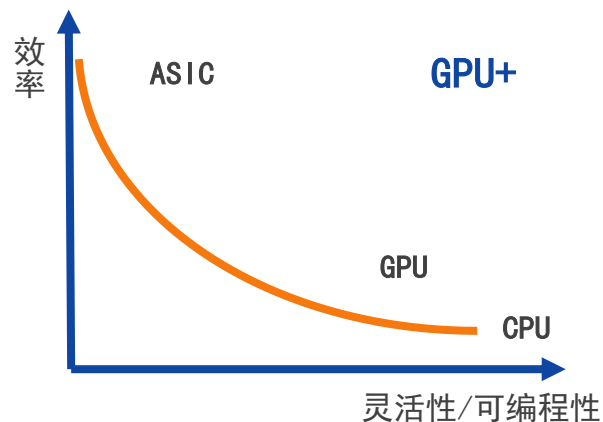


支持推理、训练、通用计算



GPU+

兼具GPU的通用性及ASIC的高效率



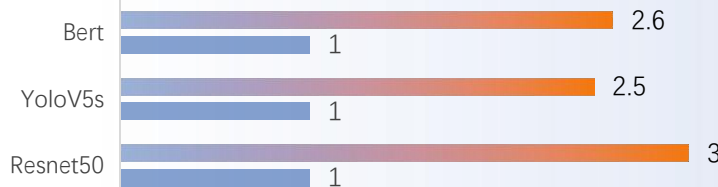
鸣天™ 芯片架构:

软件定义的片内异构通用AI域处理器



登临纳适系列产品 vs 国际主流旗舰产品

■ KS38 ■ 国际主流旗舰产品



2~3x
性能提升

3~4x
能效提升

高能效比
高通用性

大模型

DeepSeek	通义千问	MiniCPM
Baichuan	LLAMA	CHATIMG
STABLE DIFFUSION	VICUNA	OPT
GPT	LLaVA	Ultravox
Pixtral	GLM	天鹰
BLOOM		

传统模型

分类

Resnet
Inception
SqueezeNet
MobileNet
ShuffleNet
VGG
DenseNet
FCN
.....

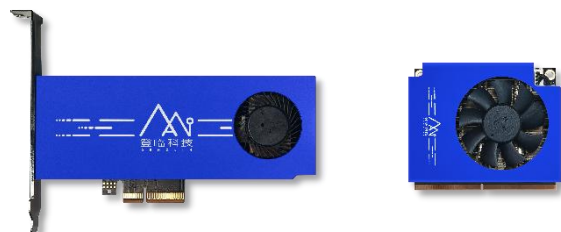
检测

Yolo
SSD
SwinTransformer
dbNet
Openpose
Blazenet
Pointpillar
.....

NLP

bert
跟踪
pplcnct
分割
Deeplab
语音
hifiNet
.....

纳适™ II 系列 KS20



KS20

关键特性		规格描述
		KS20
有效算力	INT4	140 TOPS
	INT8	70 TOPS
	FP16	35 TFLOPS
	TF32	18 TFLOPS
数据类型		INT4/INT8/BF16/FP16/TF32/FP32
内存		16GB/32GB
PCIe		PCIe 4.0 x 4
形态		MXM、PCIe、M2
典型功耗 (Typical)		25W
视频解码		最大60路 1080P@30FPS
视频编码		最大30路 1080P@30FPS
峰值内存带宽		102.4 GBytes/s

优势



超高能效

- 可适配NVR、AIBOX等边缘设备
- 可支持无风扇设计



高算力GPU

- 单卡支持**30B MoE**模型



生态

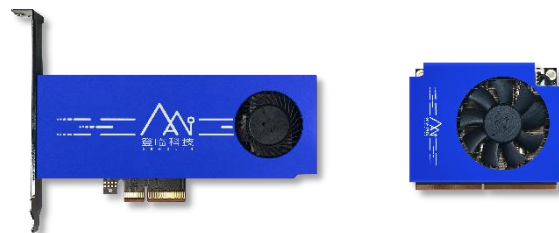
- 边缘侧提供单精度算力
- 支持CUDA/OPENCL
- 支持WINDOWS/LINUX



先进架构

- 针对DeepSeek等大模型特殊优化
- 训推一体
- 支持多框架大模型推理

纳适™ II



KS20

应用场景



智慧城市

视频结构化
智能金融
智慧园区
智慧能源

.....



智慧工业

外观检测
缺陷分类
装配确认
检测
分选定级

.....



大模型

信创PC智能助手
公文写作
本地知识库
编程辅助
备课出卷

.....

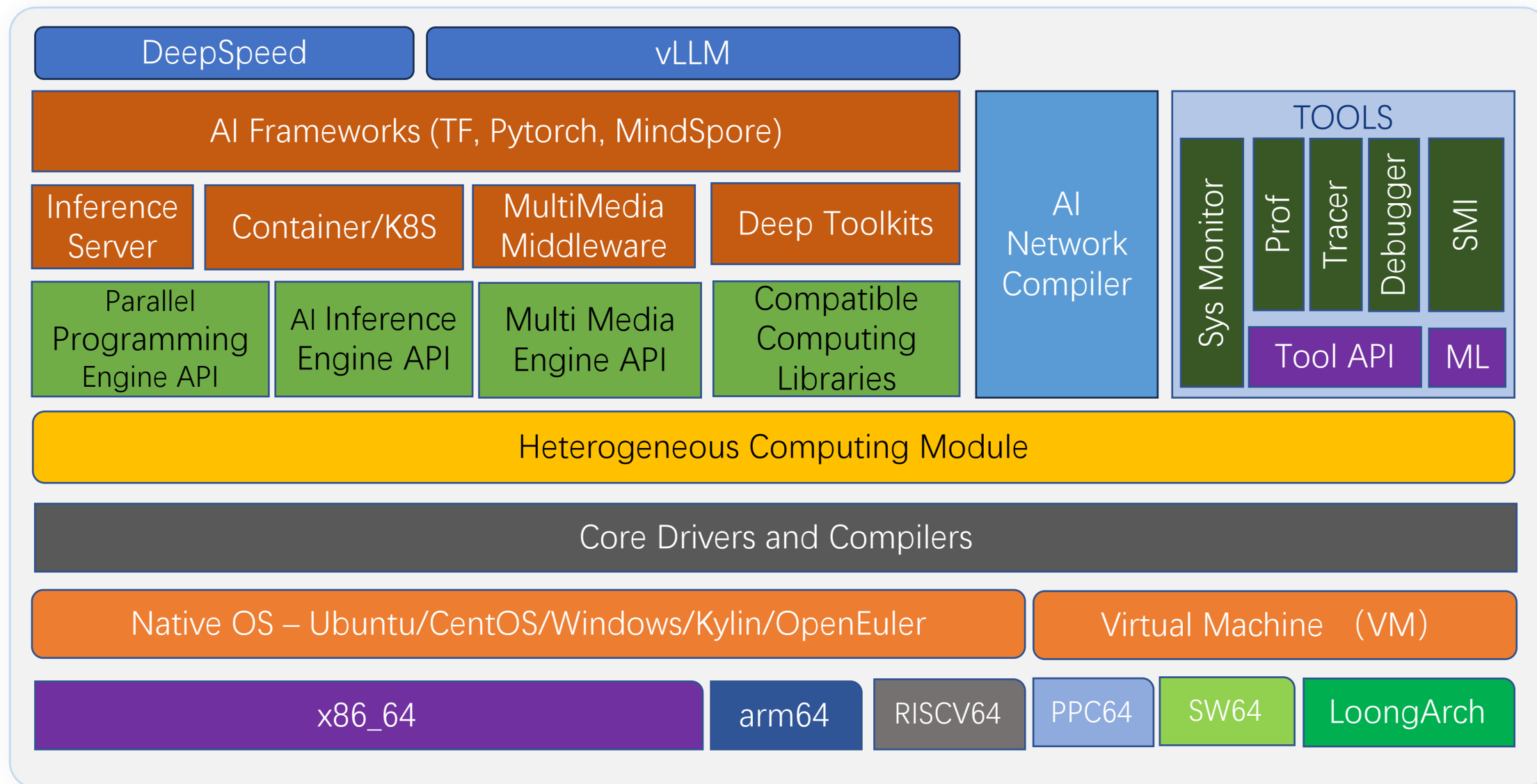


无人车/无人机

雷达
V2x
低空经济
等

.....

登临软件栈概览



登临基于dICU的cuda编程文档

从官方获取开发文档，从dICU programming guide获取详细的编程指南。

登临客户文档中心

探索我们的知识库，查找您需要的技术文档、使用指南和参考资料

SDK和应用软件文档

登临SDK, SDK驱动和CUDA适配、算子库, 常用工具和K8S, 虚拟化等文档

- ✓ SDK驱动和CUDA适配
- ✓ 算子库文档与API参考
- ✓ 常用工具与K8S集成指南
- ✓ 虚拟化技术支持文档

查看文档

登临PyTorch使用文档

登临PyTorch是基于PyTorch使用登临AI加速卡的深度学习框架

- ✓ 产品镜像使用说明
- ✓ 模型训练与优化教程
- ✓ 性能调优与基准测试
- ✓ 示例代码与API参考

查看文档

CUDA

dICU

- Denglin Compute Compiler V2 Command Options
- Denglin Hamming V2 Texture Feature Guide
- Denglin Hamming V2 dICU Extended API
- Denglin Hamming V2 dICU Extended Math API
- Denglin Hamming V2 dICU Programming Guide**
 - 1 Introduction
 - 2 Programming Model
 - 3 Programming Interface
 - 4 dICU Driver API Reference

查看文档

- 3.2.5.2 CUDA Arrays
 - 4 dICU Driver API Reference
 - Appendix A. C++ Language Extensions
 - A.1 Surface Functions
 - A.1.1 Surface Object API
 - A.2 Printf Functions
 - Format Specifiers
 - Limitations
 - Examples

Denglin Hamming™ V2 dICU Programming Guide

1 Introduction

KS20安装部署和编程示例 -1

1. 实体卡安装：KS20自带主动散热风扇，在PC上选择PICE gen4x4的插槽安装固定即可，LED灯为绿灯常亮即安装完毕。开机进入系统后，打开终端，输入`lspci | grep 1e27`；能够看到设备即说明安装成功。



```
kaitian@kaitian-90x9z0s0kx:~/桌面$ lspci | grep 1e27
02:00.0 Co-processor: Device 1e27:0006 (rev 01)
kaitian@kaitian-90x9z0s0kx:~/桌面$
```


KS20安装部署和编程示例 -3

3. SDK安装：从官方获取SDK安装文件和CUDA11头文件，解压在同一个目录下后sudo bash执行即可，注意需要使用root用户执行。看到如下界面说明已经安装成功：

目录结构

```
root@kaitian-90x9z0s0kx:/home/kaitian/下载/denglinSDK# ll
总计 801724
drwxrwxr-x 3 kaitian kaitian    4096 10月20日 16:24 ./
drwxr-xr-x 6 kaitian kaitian    4096 10月20日 16:16 ../
drwxrwxrwx 3 root   root      4096 2025年  3月10日 /
-rw-rw-r-- 1 1010 1010    527806  8月20日 16:14 cuda_11.7_cudnn_8.5.0.patch
-rw-rw-r-- 1 kaitian kaitian 31526912 10月14日 15:43 cudall.tar
-rwxrwxr-x 1 1010 1010 394160067  8月20日 16:14 denglin_hamming_v2-4.020101.71-202508201444-manylinux_2_28-x86_64.run*
-rw-rw-r-- 1 kaitian kaitian 394711040 10月14日 15:40 denglin_hamming_v2-4.020101.71-202508201444-manylinux_2_28-x86_64.tar
-rw-rw-r-- 1 1010 1010    1696  8月20日 16:14 requirements_x86_python3.10.txt
-rw-rw-r-- 1 1010 1010    1553  8月20日 16:14 requirements_x86_python3.11.txt
-rw-rw-r-- 1 1010 1010    1383  8月20日 16:14 requirements_x86_python3.12.txt
-rw-rw-r-- 1 1010 1010    1497  8月20日 16:14 requirements_x86_python3.7.txt
-rw-rw-r-- 1 1010 1010    1428  8月20日 16:14 requirements_x86_python3.8.txt
-rw-rw-r-- 1 1010 1010    1539  8月20日 16:14 requirements_x86_python3.9.txt
root@kaitian-90x9z0s0kx:/home/kaitian/下载/denglinSDK#
```

成功页面

```
patching file thrust/system/cuda/detail/cub/device/dispatch/dispatch_spawn_orig.cuh
patching file thrust/system/cuda/detail/cub/iterator/tex_obj_input_iterator.cuh
patching file thrust/system/cuda/detail/cub/thread/thread_load.cuh
patching file thrust/system/cuda/detail/cub/thread/thread_store.cuh
patching file thrust/system/cuda/detail/cub/util_arch.cuh
patching file thrust/system/cuda/detail/cub/util_device.cuh
patching file thrust/system/cuda/detail/cub/util_ptx.cuh
patching file thrust/system/cuda/detail/cub/warp/specializations/warp_reduce_shfl.cuh
patching file thrust/system/cuda/detail/cub/warp/specializations/warp_scan_shfl.cuh
patching file vector_types.h
[LOG] CUDA headers installed
[LOG] Previous installation cleaned
Activate Release Package @ /usr/local/denglin/sdk
[LOG] Environment setup completed
[LOG] Installation completed
[LOG] Installation successful
root@kaitian-90x9z0s0kx:/home/kaitian/下载/denglinSDK# bash denglin_hamming_v2-4.020101.71-202508201444-manylinux_2_28-x86_64.run
```

KS20安装部署和编程示例 -4

4. 开始cuda编程：登临提供的dlCU编译器为dlcc，可以按照文档编译第一个简单的cu程序。
注意：需要先source /usr/local/denglin/sdk/env.sh激活SDK，通过which dlcc看到如下界面。

```
(sdk) root@kaitian-90x9z0s0kx:/home/kaitian/下载/denglinSDK# source /usr/local/denglin/sdk/env.sh
Activate Release Package @ /usr/local/denglin/sdk
(sdk) root@kaitian-90x9z0s0kx:/home/kaitian/下载/denglinSDK# which dlcc
/usr/local/denglin/sdk/bin/dlcc
(sdk) root@kaitian-90x9z0s0kx:/home/kaitian/下载/denglinSDK# a
```

Cu程序

```
// Kernel definition
__global__ void MatAdd(float A[N][N], float B[N][N],
                      float C[N][N])
{
    int i = threadIdx.x;
    int j = threadIdx.y;
    C[i][j] = A[i][j] + B[i][j];
}

int main()
{
    ...
    // Kernel invocation with one block of N * N * 1 threads
    int numBlocks = 1;
    dim3 threadsPerBlock(N, N);
    MatAdd<<<numBlocks, threadsPerBlock>>>>(A, B, C);
    ...
}
```

dlcc编译命令

3.1 Compilation with dlcc

```
dlcc -o target_file_name.bc source_file_name.cu
```

For details on how to use dlcc, see document *Denglin Compute Compiler V2 Command Options.xlsx*.

KS20安装部署和编程示例 -5

5. 参考程序：登临在SDK中提供了cuda编程的应用实例，可以进入/usr/local/denglin/sdk/samples文件夹，进一步参考了解更多基于KS20做cuda编程和应用的相关示例：

```
(sdk) root@kaitian-90x9z0s0kx:/usr/local/denglin/sdk/samples# ll
总计 72
drwxr-xr-x 16 root root 4096 10月20日 16:25 ./
drwxr-xr-x 11 root root 4096 10月20日 16:25 ../
drwxr-xr-x  7 root root 4096 10月20日 16:25 cuda/
drwxr-xr-x  4 root root 4096 10月20日 16:25 dlblas/
drwxr-xr-x  4 root root 4096 10月20日 16:25 dldnn/
drwxr-xr-x  4 root root 4096 10月20日 16:25 dlfft/
drwxr-xr-x  4 root root 4096 10月20日 16:25 dljpeg/
drwxr-xr-x  5 root root 4096 10月20日 16:25 dlrnd/
drwxr-xr-x  4 root root 4096 10月20日 16:25 dlsolver/
drwxr-xr-x  4 root root 4096 10月20日 16:25 dlsparse/
drwxr-xr-x  5 root root 4096 10月20日 16:25 dlvid/
drwxr-xr-x 12 root root 4096 10月20日 16:25 network/
drwxr-xr-x  4 root root 4096 10月20日 16:25 plugin/
drwxr-xr-x  2 root root 4096 10月20日 16:25 pycuda/
drwxr-xr-x  3 root root 4096 10月20日 16:25 quantize/
-rwxr-xr-x  1 1010 1010 5518  8月20日 16:12 READMEv2.md*
drwxr-xr-x  2 root root 4096 10月20日 16:25 requirements/
(sdk) root@kaitian-90x9z0s0kx:/usr/local/denglin/sdk/samples# pwd
/usr/local/denglin/sdk/samples
(sdk) root@kaitian-90x9z0s0kx:/usr/local/denglin/sdk/samples#
```

登临广泛应用场景



AI创作



智能云



在线教育



欺诈检测



语音识别



智能客服



内容推荐



自然语言处理



娱乐



社交互动



采集



编辑



播报



智慧农业



机器人



智慧仓储



无人配送



智慧物流



交易场景



AR/VR



医疗



砥砺前行 登临未来