全爱科技后羿智盒 GPU AI 大模型开发套件
HOUYI-1000B
技术白皮书

全爱科技（上海）有限公司

商标声明

全爱科技® 后羿 JTDS 二郎神

和其他全爱商标均为全爱科技（上海）有限公司的商标。

本文档提及的其他商标或注册商标，由各自的所有人拥有。

注意事项：

您购买的产品、服务等应受全爱科技商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，全爱公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

全爱科技（上海）有限公司

地址：上海市闵行区剑川路 920 号 2 栋 3 层 邮编：200240

网址：www.quanaichina.com

文档更新记录

版本	日期	更新记录
0.1.1	2025-5-29	初版发布

操作系统支持版本如下表：

操作系统版本	Ubuntu 22.04
全爱科技 硬件产品：	全爱科技后羿智盒 GPU AI 大模型开发套件 HOUYI-1000B

目录

1 产品说明

1.1 概述

全爱科技 后羿智盒 GPU AI 大模型开发套件 HOUYI-1000B 提供完整的 GPU 并行计算和 NPU 人工智能开发环境，支持高性能计算、深度学习训练与推理等应用。

1.2 产品特点

1. CPU 12 核/8 核 2.65 GHz
2. 最大可提供 50 TOPS INT8 算力，适用于 AI 训练、推理及高性能计算场景。
3. 支持多路 H.264/H.265 硬件编解码：
   解码：2*4K 60fps。
   编码：2*8K 30fps。

1.3 外观结构

全爱科技 后羿智盒 GPU AI 大模型开发套件 HOUYI-1000B 采用 ITX 工业主板形态紧凑的结构设计，外观如图 1-1 所示。

图 1.1 外观图

按键与接口说明

图 1.2 正面接口图

表 1-1 按键与接口说明表

前面板接口			
1	Console 调试接口	2	GbE1 千兆以太网口
3	GbE0 千兆以太网口	4	DP 口
5	USB_1 USB 口	6	USB_2 USB 口
7	Type-c 下载口	8	12V 电源口
内部接口			
9	USB 口	10	SIM 卡槽
11	预留按键	12	软件升级按钮
13	风扇电源插座	14	DP 口
15	串口 U24	16	串口 U25
17	PCIE5.0 X8	18	M.2 NVME

1.4 系统框图

全爱科技 后羿智盒 GPU AI 大模型开发套件 HOUYI-1000B 硬件系统，系统框图如 1-3 所示。

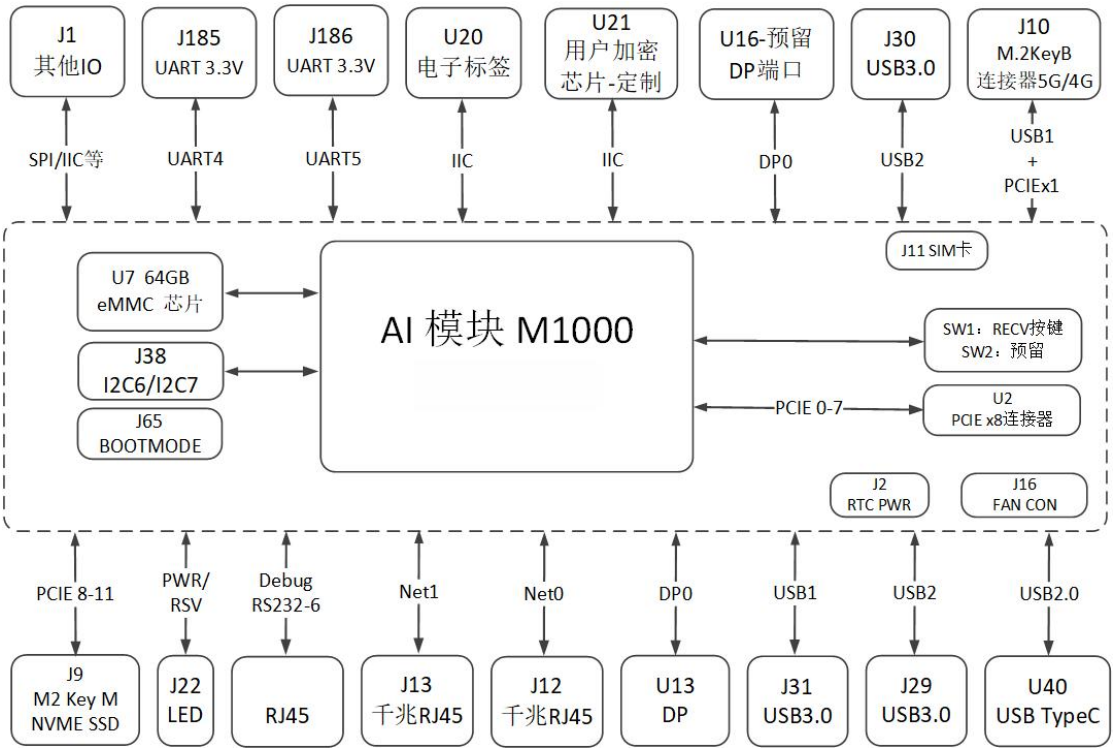

图 1.3 系统框图

2 功能说明

2.1 基本规格

表 2-1 硬件基本规格

全爱型号	HOUYI-1000B-16	HOUYI-1000B-32
规格类目	SoM 16G	SoM 32G
SoM 模组尺寸	60mmX82mm,MXM314PIN	
CPU 性能	8*ARM A78,2.65GHz	12*ARM A78,2.65GHz
AI 算力	50TOPS INT8（稠密算力）	
内存容量	16GB(适合 10B 以内 LLM)	32GB（适合 14B 以内 LLM）
内存带宽	102.4GB/s(LPDDR5)	
编码能力	2*4K 60fps	
解码能力	2*8K 30fps	
ISP 能力	Y	
PCIE 接口	PCIe5.0	
千兆网口	2 组 Ethernet 千兆	
显示能力	DP/eDP 1.4b(2 with MST)	

2.2 大模型能力

测试模型：DeepSeek-R1-Distill-Qwen-7B-GPTQ-Int4-MTT

模型来源：

https://www.modelscope.cn/models/hiyangdong/DeepSeek-R1-Distill-Qwen-7B-GPTQ-Int4-MTT

部署工具：musa-vllm

重要测试参数：

- 输入 token：2048

表 2-2 gpu-memory-utilization=0.60下测试结果

并发数	首token延时（s）	Token延时（s）	吞吐率（Tokens/s）
1	3.1549328730034176	0.10320266452325053	10.88199535879809
2	3.974085571753676	0.11247585913685329	9.995954396734625
5	7.012584995599172	0.13836560577729282	8.129758216841617
10	11.064425944249525	0.1832553381996903	6.155944294998657
20	并发数设太高会爆显存，crash	并发数设太高会爆显存，crash	并发数设太高会爆显存，crash

3 接口介绍

3.1 调试 接口

连接 RJ45 的调试线缆，进行打印系统日志。

3.2 PCIe 接口

有 1 组 PCIe X8接口，最高支持 PCIe Gen5 规格。

3.3 USB 接口

共有 3 组 USB 接口，每组 USB 接口都可以支持 USB 2.0 和 USB3.2 Gen1 的规格。每组 USB 3.2 Gen1 接口只有一对 TX/RX 信号。

3.4 DP 接口

对外提供 2 个 DP 接口可以兼容支持 eDP，不兼容支持 HDMI。

支持同屏同显和同屏异显。

3.5 以太网 接口

支持 2 组千兆以太网接口。

3.6 串口 接口

表3-1 提供2个串口连接位置，串口引脚图如下表所示。

管脚	名称	管脚	名称
1	3.3V	2	TX
3	RX	4	GND

3.7 电源 接口

供电接口使用普通的 DC 插头，电源输入电压为 12V，供电功率不低于 36W，若低于 36W 可能会出现瞬时供电不足的现象，导致系统异常。

3.8 Type C 接口

对外提供一个 Type-C 接口类型 主要用来对接调试主机进行下载文件。

3.9 M.2 Key M 连接器

M.2 Key M 连接器支持用户配置 NVME SSD 盘。默认选择 NVME 模式，支持 2280 规格形态。

表 3-2 M.2 Key M 连接器 引脚定义

管脚	名称	管脚	名称
1	GND	2	3V3
3	GND	4	3V3
5	PERn3	6	NC
7	PERp3	8	NC
9	GND	10	NC
11	PETn3	12	3V3
13	PETp3	14	3V3
15	GND	16	3V3
17	PERn2	18	3V3
19	PERp2	20	NC
21	GND	22	NC
23	PETn2	24	NC
25	PETp2	26	NC
27	GND	28	NC
29	PERn1	30	NC
31	PERp1	32	NC
33	GND	34	NC
35	PETn1	36	NC
37	PETp1	38	DEVSLP (O)
39	GND	40	NC
41	PERn0	42	NC
43	PERp0	44	NC
45	GND	46	NC
47	PETn0	48	NC
49	PETp0	50	PERST# (O) (0/1V8/3V3)
51	GND	52	CLKREQ# (I/O) (0/1V8/3V3)
53	REFCLKn	54	PEWAKE# (I/O) (0/1V8/3V3)