

DA600系列

DA600 大模型一体机系列

DA600 是万物纵横基于**RK3588/RK3588J**芯片打造的高性能人工智能深度学习边缘计算终端。产品搭载八核 64 位处理器（4×A76+4×A55），主频高达**2.4GHz**，内置**6TOPS@INT8**独立 NPU，支持 8K 超高清视频编解码，兼容**Ubuntu、Debian、麒麟**等多类操作系统，兼具工业级稳定可靠性与丰富扩展接口。



DA600 支持灵活扩展 **20TOPS/60TOPS/160TOPS@INT8** 系列 M.2 算力模组，采用「主控（CPU + 基础 NPU）+ 专用算力模组」架构，可高效运行 **1.5B ~ 35B** 参数大语言模型与多模态大模型。

DA600 针对大模型场景做了专属散热与供电强化，配备 **2路M.2 PCIe** 高速接口，支持**双算力模组并行**，最高可提供 **326TOPS@INT8**异构算力，为边缘侧大模型部署提供强劲算力支撑。

产品特性

强悍算力内核，端侧 AI 高效处理

- ✓ **AI 算力强劲：**集成 6 TOPS@INT8 算力 NPU 神经网络处理器，支持 INT4/INT8/FP16 多精度运算，轻松应对图像识别、实时视频分析、智能检测等端侧 AI 密集任务。
- ✓ **多核算力支撑：**搭载八核 CPU（4×A76+4×A55），为系统运行、业务逻辑与多任务并发处理提供稳定、流畅的算力保障。

灵活算力扩展，按需弹性升级

- ✓ **专业扩展架构：**独立散热与供电系统，保障高性能持续稳定输出；标配 2 路 M.2 PCIe 高速接口，带宽充足、扩展灵活。
- ✓ **多档算力可选：**支持 20TOPS/60TOPS/160TOPS@INT8 算力模组灵活选配，可按需升级，异构算力最高可达 326TOPS@INT8，满足从小规模到超高性能场景需求。

专业级视频处理，超高清流畅体验

- ✓ **超高清视频能力：**支持 8K@60fps H.265、8K@30fps H.264 视频解码，8K@30fps H.265/H.264 视频编码，支持多路高清摄像头接入，实现超高清可视化呈现。
- ✓ **强悍图像性能：**搭载 32MP 专业 ISP，支持 HDR、3D 降噪，解码能力强劲，低照度、复杂光影环境下画面依旧清晰。

工业级可靠设计，严苛环境稳定运行

- ✓ **宽温无风扇设计：**全铝合金被动散热机箱，可选商业级（0℃~60℃）/ 工业级宽温（-20℃~70℃），IP51 防护等级，无惧工业现场、户外等恶劣环境。
- ✓ **丰富工业接口：**双千兆网口、RS-485/232、DI/DO 等接口齐全，可直连传感器、PLC、工业相机，快速完成系统集成部署。

全场景扩展能力，适配多元部署需求

- ✓ **预留 4 路 M.2（M.2 NGFF*1、M.2 4G/5G*1、M.2 PCIe*2）扩展接口，**支持 4G/5G、Wi-Fi、算力模组、SATA SSD 及高速 NVMe SSD 存储扩展，兼顾无线联网、算力扩展、高速读写与大容量存储，满足智慧工厂、安防监控、智能交通、智慧城市等多场景落地。

产品规格

规格参数		DA600	DA600J
核心配置	芯片平台	RK3588	RK3588J
	主控处理器	八核64位大小核架构，4*Cortex-A76 + 4*Cortex-A55@2.4GHz	
	主控算力	6 TOPS@INT8	
	扩展算力(选配)	20TOPS / 60TOPS / 160TOPS @INT8	
	GPU	ARM Mali-G610 MC4 GPU, 专用2D图形加速模块	
	ISP	32MP ISP, 支持HDR 和3DNR	
	内存	默认8G LPDDR5, 可选 16G	
	存储	默认128G eMMC, 可选 32G / 64G	
	扩展存储	M.2 SSD: 支持NGFF SSD 2242,2260,2280/NVMe SSD 2280 (机箱内部) TF卡槽: 支持插入TF存储卡	
	视频编解码	解码: 支持 H.265/H.264/AV1/VP9/AVS2 , 最高 8K@60fps (7680×4320 分辨率) 编码: 支持 H.264/H.265 , 最高 8K@30fps (7680×4320 分辨率)	
基本参数	电源	DC 12VDC/5A	
	散热方式	被动散热	
	工作温度	0℃~60℃	-20℃ ~ 70℃
	工作湿度	10%~90%RH, 无凝结	
	结构尺寸	243mm × 180mm × 82mm	
	重量	不带协处理器: 2967.7 g 带 2 张 20TOPS 算力模组: 3031.2 g	
	防护等级	IP51	
接口参数	网络接口	RJ45 ×2: 支持接入10/100/1000Mbps网络 WiFi: 支持板载扩展 WiFi/蓝牙模组 4G/5G: 支持通过M.2 B-KEY接口扩展4G/5G模组	
	视频输出	HDMI IN ×1 (支持4K@60fps视频输入) HDMI OUT ×1 (支持8K@60fps视频输出)	
	音频接口	Line_IN ×1: 音频输入, Line_OUT ×1: 音频输出 (标准3.5mm音频接口)	
	其他接口	USB3.0 ×4 / SIM卡槽 ×1 / RS-485 ×2 / RS-232 ×1 / DI ×1 / DO ×2 / DEBUG ×1 / 对外供电 (5V) ×1	
	指示灯	PWR (电源) ×1 / SYS (系统) ×1 / HDD (硬盘状态) ×1 / WWAN (4G/5G模组状态) ×1	
	可扩展接口	M.2 (4G/5G) ×1 / M.2 (NGFF SSD) ×1 / M.2 (NVMe SSD/算力模组) ×2 / WIFI+BT×1 / GPS&北斗双模定位	
软件配置	系统版本	Debian12 / Ubuntu 22.04 / 麒麟 可选	
	软件支持	支持TensorFlow / ONNX / Caffe / PyTorch / MxNet / DarkNet等多种深度学习框架;	

大模型算力扩展-M.2 算力模组可选配置



M.2模组

DA600系列支持2路M.2 PCIe 标准接口扩展，可选配三款差异化算力模组，按需升级大模型算力；采用“通用主控+专用AI算力”灵活组合，覆盖从轻量化推理到工业级复杂任务的全场景需求。

系列型号	DA600(J)-R182X	DA600(J)-HM50	DA600(J)-DL20
算力模组 算力指标	20 TOPS@INT8	160 TOPS@INT8 100 FLOPS @bFP16	60 TOPS@INT8 120 TOPS@INT4 30TFLOPS@BF16/FP16
算力模组 核心架构	3D 堆叠 DRAM+8 核 NPU (近存计算)	SRAM-CIM 存算一体 (第二代“天璇”IPU)	异构计算架构 (Minsky™平台)
核心定位	高性价比大模型协处理器	高能效中端大模型主力引擎	工业级高性能多模态算力标杆
显存容量	5GB DRAM	12GB/24GB/48GB LPDDR5 (192-bit 位宽)	8/16GB LPDDR5 (128-bit 位宽)
显存带宽	1TB/s+	153.6GB/s	102.4GB/s
适配模型参数	3B-7B LLM/VLM (Qwen2.5-7B、Qwen3-8B、Qwen3-VL-4B等)	7B-35B LLM/VLM (Qwen3.5-9B、Qwen3.6-35B_A3B、Qwen3-VL 等)	7B-13B LLM/VLM (DeepSeek、InternVL 等)
核心优势	同源适配，驱动零开发；成本可控，延迟低至 0.1s	能效比 16 T/W；支持多芯扩容，多系统兼容	工业级宽温（-20℃~70℃）；32 路视频解码；7x24 小时稳定运行
算力模组 典型功耗	≤15W (同声传译场景仅 6W)	≤15W	≤25W

核心价值

本地部署 · 低延迟 · 零API 成本

- 覆盖全量级模型：

从 1.5B 轻量化模型到 35B 大参数量模型，可按需搭配算力模组，适配不同复杂度场景需求。
- 主流模型全面适配：

支持 Gemma-2B、LlaMa2-7B、Qwen2.5、Qwen3、Qwen3.5 等主流大模型。
- 极致多模态性能：

算力模组搭载专用 NPU+RK3588 ISP 视觉处理单元，支持文本、图像、视频一体化多模态推理，端到端延迟低至 0.1s。

模型类型	支持代表模型	推荐算力模组
轻量化 LLM（3B-7B）	Qwen 2.5 0.5B/1.5B/3B/7B Qwen 3 0.6B/1.7B/4B/8B GLM Edge	20TOPS
主流 LLM（7B-13B）	Qwen3-8B Qwen2.5-7B DeepSeek-7B	160TOPS 60TOPS
百亿参数 LLM（35B+）	Qwen3-30B Qwen3.5-9B Qwen3.6-35B-A3B	160TOPS
多模态 VLM	Qwen2.5-VL-1.5B~14B Qwen3-VL-35B-A3B InternVL3 YOLO	20 TOPS 160TOPS

大模型官方性能指标——R182X协处理器（截止2026年6月）

R182X——LLM

模型名称	输入长度	输出长度	首字时延 TTFT (ms)	输出速度 TPS
Qwen 2.5-3B	128	128	83.09	100.75
Qwen 2.5-7B	128	128	158.28	69.37
Qwen 3-4B	128	128	106.58	86.87

R182X——VLM

模型名称	Vision 分辨率	ViT(ms)	LLM TTFT(ms)	输出速度 TPS
FastVLM_1.5B	512*512	144	48.19	148.47
InternVL3-2B	448*448	190.80	47.93	148.26
Qwen2.5-VL-7B	392*392	279.34	159.40	70.02
Qwen3-VL-4B	384*384	158.89	108.29	89.69

R182X——ViT/CNN

模型	图像分辨率	单核帧率	多核帧率
YOLOv5s	640x640	35.4	212.6
ResNet50V2	224x224	113.7	851.3
ViT_b_p16	224x224	48.9	/
DINOv3	224x224	54.8	331.0
DepthAnythingV2	448x252	/	27.3

模型精度保持

模型类别	模型名称	数据集	考察重点	原始精度	RKNN3 精度	数据类型
LLM	Qwen3-4B	GSM8k	数学能力	90.6	89.84	W4A16
LLM	Qwen3-4B-Instruct	IFEval(strict)	指令遵循	88.73	90.42	
LLM	Qwen3-4B-Instruct	BFCL(avg)	函数调用	84.94	85.34	
VLM	InternVL3.5-4B	MMBench(cn)	图片理解	78.69	77.41	
CNN	YOLOv8s	Coco2017	目标检测	0.39(AP@.5:.95)	0.380	W8A8
CNN	ResNet50V2	Imagenet	图片分类	0.729(Top1)	0.721	

大模型官方性能指标——HM50协处理器（截止2026 年6月）

HM50——VLM性能测试汇总

Model	Size	PrefillLen(k)	Ctx(k)	Batch	NChip	Input(k)	Output(k)	Prefill(tps)	Decode(tps)	Vision(fps)
Qwen2.5-VL	7B	0.25	8	1	1	0.25	0.05	1169.48	22.35	5.23
Qwen2.5-VL	7B	0.25	8	1	1	0.5	0.05	1166.63	22.25	5.24
Qwen2.5-VL	7B	0.25	8	1	1	1	0.05	1166.53	21.92	5.24
Qwen2.5-VL	7B	0.25	8	1	1	2	0.05	1115.22	21.34	5.21
Qwen2.5-VL	7B	0.25	8	1	1	4	0.05	1023.32	19.19	5.24
Qwen2.5-VL	7B	0.25	8	1	1	7.95	0.05	868.83	17.48	5.24
Qwen3-VL	30b_a3b	0.25	8	1	1	0.25	0.05	535.53	28.54	8.23
Qwen3-VL	30b_a3b	0.25	8	1	1	0.5	0.05	534.36	28.44	8.25
Qwen3-VL	30b_a3b	0.25	8	1	1	1	0.05	537.87	27.94	8.21
Qwen3-VL	30b_a3b	0.25	8	1	1	2	0.05	524.82	25.76	8.21
Qwen3-VL	30b_a3b	0.25	8	1	1	4	0.05	488.06	22.14	8.23
Qwen3-VL	30b_a3b	0.25	8	1	1	7.95	0.05	434.96	19.59	8.24
Qwen3-VL	4B	0.25	32	1	1	0.25	0.05	3015.81	30.24	16.67
Qwen3-VL	4B	0.25	32	1	1	0.5	0.05	2986.93	29.67	16.7
Qwen3-VL	4B	0.25	32	1	1	1	0.05	2942.91	28.76	16.79
Qwen3-VL	4B	0.25	32	1	1	2	0.05	2694.4	27.75	16.7
Qwen3-VL	4B	0.25	32	1	1	4	0.05	2333.26	26.17	16.75
Qwen3-VL	4B	0.25	32	1	1	8	0.05	1853.63	23.41	16.75
Qwen3-VL	4B	0.25	32	1	1	16	0.05	1318.71	19.3	16.8
Qwen3-VL	4B	0.25	32	1	1	31.95	0.05	812.86	14.9	16.77
Qwen3-VL	8B	0.25	32	1	1	0.25	0.05	2181.26	21.44	10.99
Qwen3-VL	8B	0.25	32	1	1	0.5	0.05	2167.41	21.11	11.01
Qwen3-VL	8B	0.25	32	1	1	1	0.05	2131.02	20.63	11
Qwen3-VL	8B	0.25	32	1	1	2	0.05	1999.02	20.16	11.01
Qwen3-VL	8B	0.25	32	1	1	4	0.05	1794.95	19.31	11
Qwen3-VL	8B	0.25	32	1	1	8	0.05	1497.96	17.75	10.98
Qwen3-VL	8B	0.25	32	1	1	16	0.05	1128.48	15.3	11.01
Qwen3-VL	8B	0.25	32	1	1	31.95	0.05	754.68	12.38	10.96

Model	Size	PrefillLen(k)	Ctx(k)	Batch	NChip	Input(k)	Output(k)	Prefill(tps)	Decode(t ps)	Vision(fps)
Qwen3.5	0.8B	0.25	256	1	1	0.25	0.05	6690.44	110.5	2.01
Qwen3.5	0.8B	0.25	256	1	1	0.5	0.05	6677.75	109.73	2.01
Qwen3.5	0.8B	0.25	256	1	1	1	0.05	6664.77	105.83	2.01
Qwen3.5	0.8B	0.25	256	1	1	2	0.05	6620.46	103.7	2.01
Qwen3.5	0.8B	0.25	256	1	1	4	0.05	6480.96	101.26	2.01
Qwen3.5	0.8B	0.25	256	1	1	8	0.05	6251.86	95.87	2.01
Qwen3.5	0.8B	0.25	256	1	1	16	0.05	5863.87	86.99	2.01
Qwen3.5	0.8B	0.25	256	1	1	32	0.05	5212.83	74.22	2.01
Qwen3.5	0.8B	0.25	256	1	1	64	0.05	4270.56	56.36	2.01
Qwen3.5	0.8B	0.25	256	1	1	128	0.05	3139.62	38.44	2.01
Qwen3.5	0.8B	0.25	256	1	1	255.95	0.05	2054.41	23.84	2.01
Qwen3.5	2B	0.25	256	1	1	0.25	0.05	5187.01	65.85	1.4
Qwen3.5	2B	0.25	256	1	1	0.5	0.05	5156.94	65.76	1.4
Qwen3.5	2B	0.25	256	1	1	1	0.05	5150.75	64.52	1.4
Qwen3.5	2B	0.25	256	1	1	2	0.05	5122.19	63.69	1.4
Qwen3.5	2B	0.25	256	1	1	4	0.05	5049.64	62.39	1.4
Qwen3.5	2B	0.25	256	1	1	8	0.05	4905.31	60.82	1.4
Qwen3.5	2B	0.25	256	1	1	16	0.05	4662.94	56.89	1.4
Qwen3.5	2B	0.25	256	1	1	32	0.05	4238.52	50.98	1.4
Qwen3.5	2B	0.25	256	1	1	64	0.05	3595.36	41.98	1.4
Qwen3.5	2B	0.25	256	1	1	128	0.05	2758.95	31.26	1.4
Qwen3.5	2B	0.25	256	1	1	255.95	0.05	1884.24	20.8	1.4
Qwen3.5	4B	0.25	256	1	1	0.25	0.05	2303	31.81	1.4
Qwen3.5	4B	0.25	256	1	1	0.5	0.05	2299.06	31.7	1.4
Qwen3.5	4B	0.25	256	1	1	1	0.05	2295.52	31.31	1.4
Qwen3.5	4B	0.25	256	1	1	2	0.05	2273.18	31	1.4
Qwen3.5	4B	0.25	256	1	1	4	0.05	2234.48	30.52	1.4
Qwen3.5	4B	0.25	256	1	1	8	0.05	2164.45	29.58	1.4
Qwen3.5	4B	0.25	256	1	1	16	0.05	2034.1	27.99	1.4
Qwen3.5	4B	0.25	256	1	1	32	0.05	1819.19	25.32	1.4
Qwen3.5	4B	0.25	256	1	1	64	0.05	1500.94	21.26	1.4
Qwen3.5	4B	0.25	256	1	1	128	0.05	1113.01	16.07	1.4
Qwen3.5	4B	0.25	256	1	1	255.95	0.05	734.65	10.84	1.4
Qwen3.5	9B	0.25	256	1	1	0.25	0.05	1791.45	21.31	0.87
Qwen3.5	9B	0.25	256	1	1	0.5	0.05	1788.27	21.25	0.87
Qwen3.5	9B	0.25	256	1	1	1	0.05	1786.97	21.11	0.87
Qwen3.5	9B	0.25	256	1	1	2	0.05	1772.91	20.95	0.87
Qwen3.5	9B	0.25	256	1	1	4	0.05	1749.84	20.75	0.87
Qwen3.5	9B	0.25	256	1	1	8	0.05	1706.11	20.3	0.87
Qwen3.5	9B	0.25	256	1	1	16	0.05	1624.98	19.56	0.87
Qwen3.5	9B	0.25	256	1	1	32	0.05	1484.45	18.26	0.87
Qwen3.5	9B	0.25	256	1	1	64	0.05	1265.55	16.06	0.87
Qwen3.5	9B	0.25	256	1	1	128	0.05	977.97	12.96	0.87
Qwen3.5	9B	0.25	256	1	1	255.95	0.05	673.06	9.41	0.87

Model	Size	PrefillLen(k)	Ctx(k)	Batch	NChip	Input(k)	Output(k)	Prefill(tps)	Decode(tps)	Vision(fps)
Qwen3.6	35B-A3B	0.25	256	1	2	0.25	0.05	859.97	34.7	0.87
Qwen3.6	35B-A3B	0.25	256	1	2	0.5	0.05	861	34.61	0.87
Qwen3.6	35B-A3B	0.25	256	1	2	1	0.05	856.94	34.22	0.87
Qwen3.6	35B-A3B	0.25	256	1	2	2	0.05	857.46	33.66	0.87
Qwen3.6	35B-A3B	0.25	256	1	2	4	0.05	856.54	33.43	0.87
Qwen3.6	35B-A3B	0.25	256	1	2	8	0.05	844.56	32.57	0.87
Qwen3.6	35B-A3B	0.25	256	1	2	16	0.05	823	31.15	0.87
Qwen3.6	35B-A3B	0.25	256	1	2	32	0.05	783.06	28.66	0.87
Qwen3.6	35B-A3B	0.25	256	1	2	64	0.05	711.41	24.79	0.87
Qwen3.6	35B-A3B	0.25	256	1	2	128	0.05	601.81	19.65	0.87
Qwen3.6	35B-A3B	0.25	256	1	2	255.95	0.05	460.45	13.89	0.87
Qwen3.6	35B-A3B	0.25	256	1	1	0.25	0.05	768.17	30.08	0.87
Qwen3.6	35B-A3B	0.25	256	1	1	0.5	0.05	769.69	30.03	0.87
Qwen3.6	35B-A3B	0.25	256	1	1	1	0.05	769.08	29.56	0.87
Qwen3.6	35B-A3B	0.25	256	1	1	2	0.05	765.98	29.3	0.87
Qwen3.6	35B-A3B	0.25	256	1	1	4	0.05	763.84	28.91	0.87
Qwen3.6	35B-A3B	0.25	256	1	1	8	0.05	759.04	28.25	0.87
Qwen3.6	35B-A3B	0.25	256	1	1	16	0.05	737.67	27	0.87
Qwen3.6	35B-A3B	0.25	256	1	1	32	0.05	705.17	24.84	0.87
Qwen3.6	35B-A3B	0.25	256	1	1	64	0.05	645.62	21.47	0.87
Qwen3.6	35B-A3B	0.25	256	1	1	128	0.05	551.62	16.89	0.87
Qwen3.6	35B-A3B	0.25	256	1	1	255.95	0.05	426.96	11.91	0.87
Qwen3.6	27B	0.25	128	1	1	0.25	0.05	596.38	7.09	0.87
Qwen3.6	27B	0.25	128	1	1	0.5	0.05	596.28	7.08	0.87
Qwen3.6	27B	0.25	128	1	1	1	0.05	595.12	7.03	0.87
Qwen3.6	27B	0.25	128	1	1	2	0.05	590.65	7	0.87
Qwen3.6	27B	0.25	128	1	1	4	0.05	582.03	6.95	0.87
Qwen3.6	27B	0.25	128	1	1	8	0.05	566.09	6.84	0.87
Qwen3.6	27B	0.25	128	1	1	16	0.05	536.86	6.65	0.87
Qwen3.6	27B	0.25	128	1	1	32	0.05	486.73	6.28	0.87
Qwen3.6	27B	0.25	128	1	1	64	0.05	410.21	5.69	0.87
Qwen3.6	27B	0.25	128	1	1	127.95	0.05	311.97	4.82	0.87
Qwen3.6	27B	0.25	256	1	2	0.25	0.05	726.61	10.33	0.87
Qwen3.6	27B	0.25	256	1	2	0.5	0.05	727.84	10.34	0.87
Qwen3.6	27B	0.25	256	1	2	1	0.05	726.85	10.26	0.87
Qwen3.6	27B	0.25	256	1	2	2	0.05	722.92	10.2	0.87
Qwen3.6	27B	0.25	256	1	2	4	0.05	716.4	10.13	0.87
Qwen3.6	27B	0.25	256	1	2	8	0.05	704.13	9.99	0.87
Qwen3.6	27B	0.25	256	1	2	16	0.05	681.06	9.73	0.87
Qwen3.6	27B	0.25	256	1	2	32	0.05	638.78	9.25	0.87
Qwen3.6	27B	0.25	256	1	2	64	0.05	567.97	8.44	0.87
Qwen3.6	27B	0.25	256	1	2	128	0.05	465.23	7.16	0.87
Qwen3.6	27B	0.25	256	1	2	255.95	0.05	341.56	5.53	0.87

大模型官方性能指标——HM50协处理器

HM50——LLM性能测试汇总

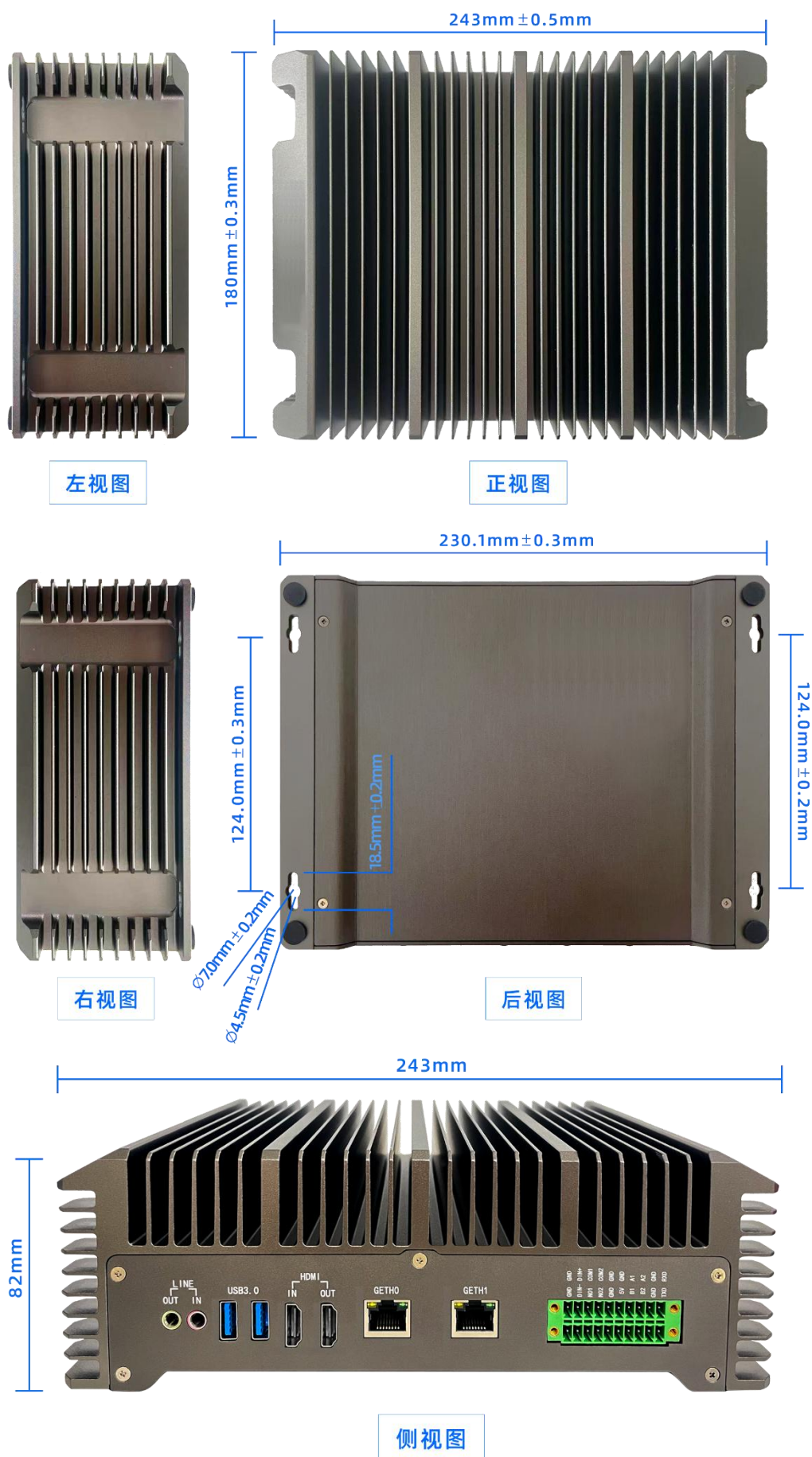
Model	Size	PrefillLen(k)	Ctx(k)	Batch	NChip	Input(k)	Output(k)	Prefill(tps)	Decode(tps)	Vision(fps)
CoPaw-Flash	9B	0.25	8	1	1	0.25	0.05	1650.66	20.64	0
CoPaw-Flash	9B	0.25	8	1	1	0.5	0.05	1649.58	20.61	0
CoPaw-Flash	9B	0.25	8	1	1	1	0.05	1639.31	15.92	0
CoPaw-Flash	9B	0.25	8	1	1	2	0.05	1631.89	20.18	0
CoPaw-Flash	9B	0.25	8	1	1	4	0.05	1616.38	20.24	0
CoPaw-Flash	9B	0.25	8	1	1	7.95	0.05	1569.39	20.09	0
DeepSeek-R1-Qwen3	8B	0.25	32	1	1	0.25	0.05	2453.06	22.36	0
DeepSeek-R1-Qwen3	8B	0.25	32	1	1	0.5	0.05	2437.41	22.02	0
DeepSeek-R1-Qwen3	8B	0.25	32	1	1	1	0.05	2407.77	21.48	0
DeepSeek-R1-Qwen3	8B	0.25	32	1	1	2	0.05	2240.06	21.02	0
DeepSeek-R1-Qwen3	8B	0.25	32	1	1	4	0.05	1986.44	20.14	0
DeepSeek-R1-Qwen3	8B	0.25	32	1	1	8	0.05	1630.19	18.6	0
DeepSeek-R1-Qwen3	8B	0.25	32	1	1	16	0.05	1202.33	16.1	0
DeepSeek-R1-Qwen3	8B	0.25	32	1	1	31.95	0.05	787.71	13.12	0
DeepSeek-R1-Qwen3	8B	0.25	32	2	1	0.25	0.05	2451.89	38.01	0
DeepSeek-R1-Qwen3	8B	0.25	32	2	1	0.5	0.05	2437.3	37.12	0
DeepSeek-R1-Qwen3	8B	0.25	32	2	1	1	0.05	2403.58	35.87	0
DeepSeek-R1-Qwen3	8B	0.25	32	2	1	2	0.05	2235.98	34.6	0
DeepSeek-R1-Qwen3	8B	0.25	32	2	1	4	0.05	1985.43	32.28	0
DeepSeek-R1-Qwen3	8B	0.25	32	2	1	8	0.05	1629.2	28.51	0
DeepSeek-R1-Qwen3	8B	0.25	32	2	1	16	0.05	1202.36	23.09	0
DeepSeek-R1-Qwen3	8B	0.25	32	2	1	31.95	0.05	787.55	17.48	0
GPT-OSS	20B	0.25	256	1	1	0.25	0.05	1432.48	33.34	0
GPT-OSS	20B	0.25	256	1	1	0.5	0.05	1441.67	33.15	0
GPT-OSS	20B	0.25	256	1	1	1	0.05	1452.86	31.92	0
GPT-OSS	20B	0.25	256	1	1	2	0.05	1419.66	30.68	0
GPT-OSS	20B	0.25	256	1	1	4	0.05	1351.18	28.6	0
GPT-OSS	20B	0.25	256	1	1	8	0.05	1219.18	25.2	0
GPT-OSS	20B	0.25	256	1	1	16	0.05	1025.09	20.39	0
GPT-OSS	20B	0.25	256	1	1	32	0.05	776.42	14.76	0
GPT-OSS	20B	0.25	256	1	1	64	0.05	523.59	9.52	0
GPT-OSS	20B	0.25	256	1	1	128	0.05	317.12	5.57	0
GPT-OSS	20B	0.25	256	1	1	255.95	0.05	176.67	3.05	0
GPT-OSS	20B	0.25	32	1	1	0.25	0.05	1439.58	33.42	0
GPT-OSS	20B	0.25	32	1	1	0.5	0.05	1461.51	33.07	0
GPT-OSS	20B	0.25	32	1	1	1	0.05	1463.01	31.93	0
GPT-OSS	20B	0.25	32	1	1	2	0.05	1427.05	30.64	0
GPT-OSS	20B	0.25	32	1	1	4	0.05	1357.86	28.59	0
GPT-OSS	20B	0.25	32	1	1	8	0.05	1221.69	25.23	0
GPT-OSS	20B	0.25	32	1	1	16	0.05	1027.97	20.39	0
GPT-OSS	20B	0.25	32	1	1	31.95	0.05	770.61	14.94	0

Model	Size	PrefillLen(k)	Ctx(k)	Batch	NChip	Input(k)	Output(k)	Prefill(tps)	Decode(tps)	Vision(fps)
Qwen2.5	7B	0.25	8	1	1	0.25	0.05	1300.12	25.11	0
Qwen2.5	7B	0.25	8	1	1	0.5	0.05	1301	25.02	0
Qwen2.5	7B	0.25	8	1	1	1	0.05	1301.43	24.82	0
Qwen2.5	7B	0.25	8	1	1	2	0.05	1260.61	24.52	0
Qwen2.5	7B	0.25	8	1	1	4	0.05	1180.77	22.24	0
Qwen2.5	7B	0.25	8	1	1	7.95	0.05	1034.31	20.64	0
Qwen2.5	7B	0.25	8	2	1	0.25	0.05	1299.78	37.63	0
Qwen2.5	7B	0.25	8	2	1	0.5	0.05	1300.75	37.53	0
Qwen2.5	7B	0.25	8	2	1	1	0.05	1299.67	37.01	0
Qwen2.5	7B	0.25	8	2	1	2	0.05	1260.31	36.33	0
Qwen2.5	7B	0.25	8	2	1	4	0.05	1179.95	34.88	0
Qwen2.5	7B	0.25	8	2	1	7.95	0.05	1033.79	32.29	0
Qwen3	0.6B	0.25	32	1	1	0.25	0.05	10432.43	86.75	0
Qwen3	0.6B	0.25	32	1	1	0.5	0.05	10172.32	82.74	0
Qwen3	0.6B	0.25	32	1	1	1	0.05	9773.33	77.59	0
Qwen3	0.6B	0.25	32	1	1	2	0.05	8743.59	72.4	0
Qwen3	0.6B	0.25	32	1	1	4	0.05	7171.56	64.55	0
Qwen3	0.6B	0.25	32	1	1	8	0.05	5345.57	53.55	0
Qwen3	0.6B	0.25	32	1	1	16	0.05	3544.36	39.89	0
Qwen3	0.6B	0.25	32	1	1	31.95	0.05	2120.47	27.95	0
Qwen3	1.7B	0.25	32	1	1	0.25	0.05	7304.76	58.3	0
Qwen3	1.7B	0.25	32	1	1	0.5	0.05	7177.15	56.46	0
Qwen3	1.7B	0.25	32	1	1	1	0.05	7055.86	53.66	0
Qwen3	1.7B	0.25	32	1	1	2	0.05	6461.93	51.73	0
Qwen3	1.7B	0.25	32	1	1	4	0.05	5565.3	47.73	0
Qwen3	1.7B	0.25	32	1	1	8	0.05	4400.91	41.47	0
Qwen3	1.7B	0.25	32	1	1	16	0.05	3099.16	32.67	0
Qwen3	1.7B	0.25	32	1	1	31.95	0.05	1951.17	24.08	0
Qwen3	14B	0.25	16	1	1	0.25	0.05	1328.27	12.79	0
Qwen3	14B	0.25	16	1	1	0.5	0.05	1320.47	12.65	0
Qwen3	14B	0.25	16	1	1	1	0.05	1302.3	12.46	0
Qwen3	14B	0.25	16	1	1	2	0.05	1213.72	12.3	0
Qwen3	14B	0.25	16	1	1	4	0.05	1075.44	11.96	0
Qwen3	14B	0.25	16	1	1	8	0.05	881.55	11.34	0
Qwen3	14B	0.25	16	1	1	15.95	0.05	648.62	10.58	0
Qwen3	14B	0.25	32	1	2	0.25	0.05	1725.17	21.89	0
Qwen3	14B	0.25	32	1	2	0.5	0.05	1719.87	21.68	0
Qwen3	14B	0.25	32	1	2	1	0.05	1699.69	20.99	0
Qwen3	14B	0.25	32	1	2	2	0.05	1622.71	20.66	0
Qwen3	14B	0.25	32	1	2	4	0.05	1494.97	19.99	0
Qwen3	14B	0.25	32	1	2	8	0.05	1295.09	18.82	0
Qwen3	14B	0.25	32	1	2	16	0.05	1026.16	16.81	0
Qwen3	14B	0.25	32	1	2	31.95	0.05	724.67	14.32	0

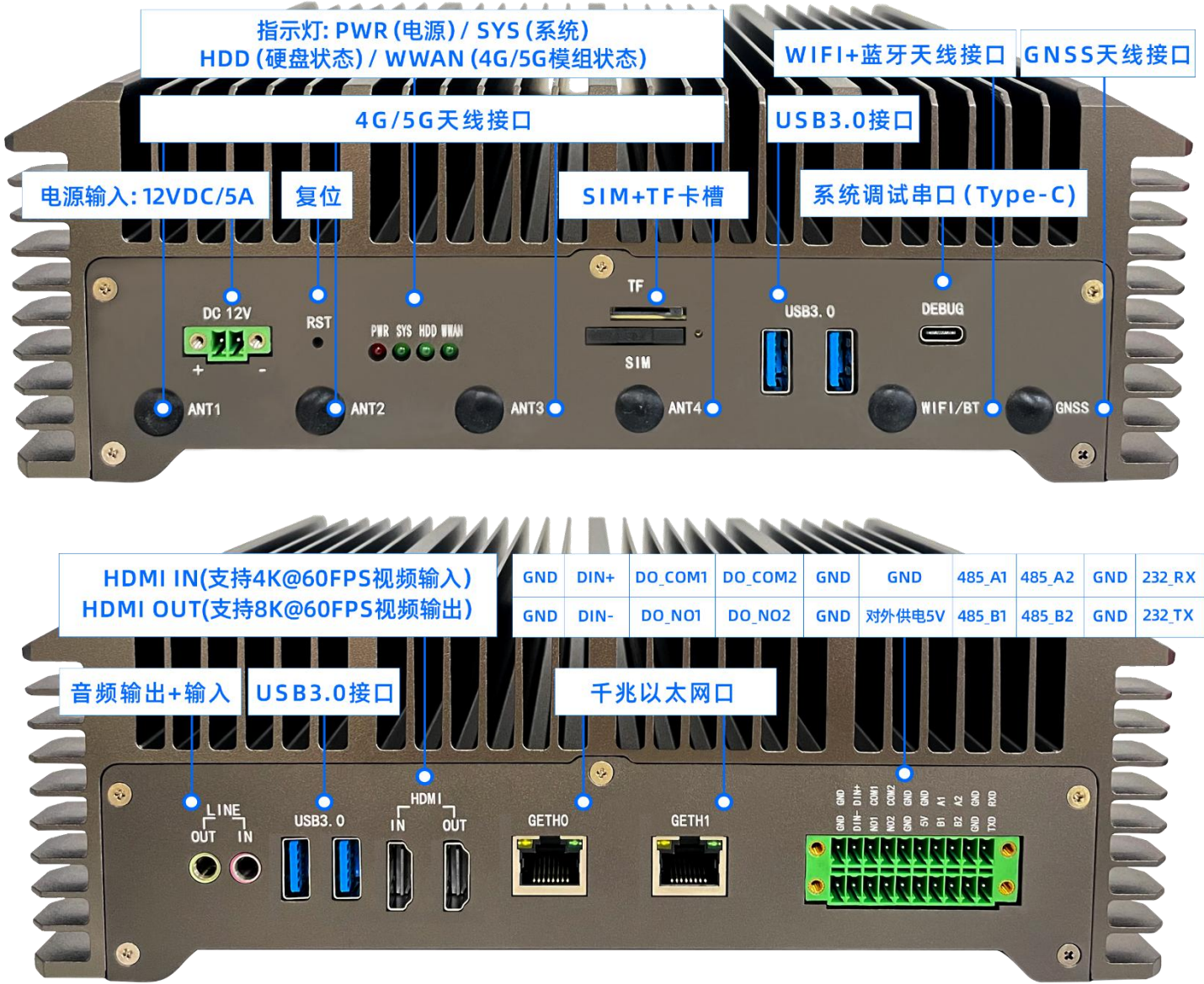
Model	Size	PrefillLen(k)	Ctx(k)	Batch	NChip	Input(k)	Output(k)	Prefill(tps)	Decode(tps)	Vision(fps)
Qwen3	4B	0.25	32	1	1	0.25	0.05	3195.34	23.03	0
Qwen3	4B	0.25	32	1	1	0.5	0.05	3161.35	22.57	0
Qwen3	4B	0.25	32	1	1	1	0.05	3098.86	22.02	0
Qwen3	4B	0.25	32	1	1	2	0.05	2829.65	21.54	0
Qwen3	4B	0.25	32	1	1	4	0.05	2433.82	20.6	0
Qwen3	4B	0.25	32	1	1	8	0.05	1919.4	19.01	0
Qwen3	4B	0.25	32	1	1	16	0.05	1352.64	16.4	0
Qwen3	4B	0.25	32	1	1	31.95	0.05	849.55	13.32	0
Qwen3	8B	0.25	16	4	1	0.25	0.05	2312.91	57.27	0
Qwen3	8B	0.25	16	4	1	0.5	0.05	2296.38	55.19	0
Qwen3	8B	0.25	16	4	1	1	0.05	2258.9	53.32	0
Qwen3	8B	0.25	16	4	1	2	0.05	2112.91	50.46	0
Qwen3	8B	0.25	16	4	1	4	0.05	1886.47	45.44	0
Qwen3	8B	0.25	16	4	1	8	0.05	1562.4	38.05	0
Qwen3	8B	0.25	16	4	1	15.95	0.05	1161.1	31	0
Qwen3	8B	0.25	32	1	1	0.25	0.05	2322.72	22.45	0
Qwen3	8B	0.25	32	1	1	0.5	0.05	2306.02	22.07	0
Qwen3	8B	0.25	32	1	1	1	0.05	2274.67	21.56	0
Qwen3	8B	0.25	32	1	1	2	0.05	2122.8	21.13	0
Qwen3	8B	0.25	32	1	1	4	0.05	1892.83	20.2	0
Qwen3	8B	0.25	32	1	1	8	0.05	1566.29	18.67	0
Qwen3	8B	0.25	32	1	1	16	0.05	1154.62	16.21	0
Qwen3	8B	0.25	32	1	1	31.95	0.05	745.13	11.5	0
Qwen3	30b_a3b	0.25	128	1	1	0.25	0.05	1179.75	31.31	0
Qwen3	30b_a3b	0.25	128	1	1	0.5	0.05	1222.4	30.78	0
Qwen3	30b_a3b	0.25	128	1	1	1	0.05	1244	29.04	0
Qwen3	30b_a3b	0.25	128	1	1	2	0.05	1204.55	27.89	0
Qwen3	30b_a3b	0.25	128	1	1	4	0.05	1116.98	26.25	0
Qwen3	30b_a3b	0.25	128	1	1	8	0.05	976.55	23.48	0
Qwen3	30b_a3b	0.25	128	1	1	16	0.05	773.49	19.46	0
Qwen3	30b_a3b	0.25	128	1	1	32	0.05	542.57	14.52	0
Qwen3	30b_a3b	0.25	128	1	1	64	0.05	339.75	9.67	0
Qwen3	30b_a3b	0.25	128	1	1	127.95	0.05	193.92	5.91	0
Qwen3	30b_a3b	0.25	32	1	1	0.25	0.05	1197.97	32.47	0
Qwen3	30b_a3b	0.25	32	1	1	0.5	0.05	1265.11	32	0
Qwen3	30b_a3b	0.25	32	1	1	1	0.05	1277.01	30.15	0
Qwen3	30b_a3b	0.25	32	1	1	2	0.05	1241.68	29.16	0
Qwen3	30b_a3b	0.25	32	1	1	4	0.05	1143.92	27.52	0
Qwen3	30b_a3b	0.25	32	1	1	8	0.05	997.34	24.78	0
Qwen3	30b_a3b	0.25	32	1	1	16	0.05	786.09	20.61	0
Qwen3	30b_a3b	0.25	32	1	1	31.95	0.05	549.09	16.22	0
Gemma 4	26b_a4b	0.25	128	1	1	0.25	0.05	1317.17	27.05	0
Gemma 4	26b_a4b	0.25	128	1	1	1	0.05	1401.93	26.02	0
Gemma 4	26b_a4b	0.25	128	1	1	2	0.05	1330.96	24.98	0
Gemma 4	26b_a4b	0.25	128	1	1	4	0.05	1342.04	23.16	0
Gemma 4	26b_a4b	0.25	128	1	1	8	0.05	1340.28	19.94	0
Gemma 4	26b_a4b	0.25	128	1	1	16	0.05	1241.37	15.34	0
Gemma 4	26b_a4b	0.25	128	1	1	48	0.05	999.17	8.38	0
Gemma 4	26b_a4b	0.25	128	1	1	100	0.05	548.92	1.58	0

Model	Size	PrefillLen(k)	Ctx(k)	Batch	NChip	Input(k)	Output(k)	Prefill(tps)	Decode(tps)	Vision(fps)
Qwen3-Coder	30b_a3b	0.25	128	1	1	0.25	0.05	1186.91	31.49	0
Qwen3-Coder	30b_a3b	0.25	128	1	1	0.5	0.05	1201.65	30.92	0
Qwen3-Coder	30b_a3b	0.25	128	1	1	1	0.05	1207.09	29.17	0
Qwen3-Coder	30b_a3b	0.25	128	1	1	2	0.05	1175.24	28.02	0
Qwen3-Coder	30b_a3b	0.25	128	1	1	4	0.05	1101.61	26.34	0
Qwen3-Coder	30b_a3b	0.25	128	1	1	8	0.05	969.98	23.55	0
Qwen3-Coder	30b_a3b	0.25	128	1	1	16	0.05	767.18	19.57	0
Qwen3-Coder	30b_a3b	0.25	128	1	1	32	0.05	543.1	14.53	0
Qwen3-Coder	30b_a3b	0.25	128	1	1	64	0.05	341.49	9.67	0
Qwen3-Coder	30b_a3b	0.25	128	1	1	127.95	0.05	188.02	5.91	0
Qwen3-Coder	30b_a3b	0.25	128	1	2	0.25	0.05	1800.19	38.64	0
Qwen3-Coder	30b_a3b	0.25	128	1	2	0.5	0.05	1849.71	38.29	0
Qwen3-Coder	30b_a3b	0.25	128	1	2	1	0.05	1807.37	35.26	0
Qwen3-Coder	30b_a3b	0.25	128	1	2	2	0.05	1780.94	34.19	0
Qwen3-Coder	30b_a3b	0.25	128	1	2	4	0.05	1660.24	32.41	0
Qwen3-Coder	30b_a3b	0.25	128	1	2	8	0.05	1497.1	29.12	0
Qwen3-Coder	30b_a3b	0.25	128	1	2	16	0.05	1253.95	24.29	0
Qwen3-Coder	30b_a3b	0.25	128	1	2	32	0.05	937.23	18.22	0
Qwen3-Coder	30b_a3b	0.25	128	1	2	64	0.05	622.84	12.19	0
Qwen3-Coder	30b_a3b	0.25	128	1	2	127.95	0.05	359.7	7.46	0
Qwen3-Coder	30b_a3b	0.25	256	1	2	0.25	0.05	1764.65	38.36	0
Qwen3-Coder	30b_a3b	0.25	256	1	2	0.5	0.05	1779.39	37.99	0
Qwen3-Coder	30b_a3b	0.25	256	1	2	1	0.05	1817.58	35.41	0
Qwen3-Coder	30b_a3b	0.25	256	1	2	2	0.05	1783.77	34.11	0
Qwen3-Coder	30b_a3b	0.25	256	1	2	4	0.05	1662.99	32.34	0
Qwen3-Coder	30b_a3b	0.25	256	1	2	8	0.05	1513.65	29.05	0
Qwen3-Coder	30b_a3b	0.25	256	1	2	16	0.05	1249.73	24.21	0
Qwen3-Coder	30b_a3b	0.25	256	1	2	32	0.05	937.94	18.23	0
Qwen3-Coder	30b_a3b	0.25	256	1	2	64	0.05	623.4	12.21	0
Qwen3-Coder	30b_a3b	0.25	256	1	2	128	0.05	359.65	7.31	0
Qwen3-Coder	30b_a3b	0.25	256	1	2	255.95	0.05	199.77	4.11	0

产品尺寸



接口描述



应用场景



智能会议终端

核心需求：低成本实现会议离线转录、多语言翻译，避免云端数据泄露，无需专业 IT 维护。

DA600+算力模组方案：

- DA600 提供 Line-IN 音频接口（接入会议麦克风）、HDMI 显示接口（输出转录内容），支持 4G/5G 扩展。
- 算力模组负责 ASR 语音转文字、LLM 文本总结本地存储转录文件。



轻量安防分析

核心需求：实时识别异常事件，生成视频摘要，减少人工回看时间，带宽成本低。

DA600+算力模组方案：

- DA600 接入1080P 摄像头，支持本地 SD 卡存储（最大 2TB）。
- 算力模组运行模型，实现异常事件预警，预警信息可通过 DA600 串口联动声光报警器。



工业视觉质检

核心需求：检测电子元件表面划痕、缺角，配合产线，兼顾检测效率与功耗。

DA600J+算力模组方案：

- DA600J 接入工业相机提供 DI/DO 接口，支持 - 20℃~70℃宽温（车间环境适配）。
- 算力模组运行定制化瑕疵检测模型，10W 低功耗无需额外散热装置。

保修描述

保修范围

- 保修期限：主机保修1年（自购买日期起，具体以销售凭证为准）
- 免费服务：非人为故障，凭有效保修卡可免费维修
- 不保修范围：随机附件、人为损坏（如碰撞、电压异常、误操作等）及超出保修期的故障

不保修的情况

以下情形需支付维修费：

- 无法提供有效保修凭证，或凭证与产品不符
- 人为损坏（如带电插拔、运输磕碰、未按说明书操作等）
- 自行拆卸、改装或未经授权维修
- 自然灾害等不可抗力导致的故障

用户须知

- 配合义务：维修时需提供产品SN码及故障描述
- 数据安全：维修可能导致数据丢失，请提前备份，本公司不承担责任
- 更换部件：维修更换的故障部件归本公司所有

其他声明

- 本公司保留条款解释权及维修操作权限
- 超保或非保修故障可提供有偿服务，费用详询售后

注：购买后请核对保修信息，妥善保存凭证



官方公众号



官方抖音



官方淘宝

四川万物纵横科技股份有限公司



公司地址

成都市高新区天府五街花漾锦江B座7层



官方网站

<http://www.iotdt.com>



联系电话

191-1390-7060