LLM-8850 Card

SKU:AI-001

AI
LLM-8850 CARD
M.2 M-KEY AI ACCELERATION CARD

M5STACK

MICRO TURBINE FAN

ALUMINUM ALLOY CNC HEAT SINK

M.2 M-KEY 2242

42.6x24x9.7mm

LLM 8850 ACCELERATOR

7W@3.3V
8GB LPDDR4
24 TOPS@INT8
A55 8-Core 1.7GHz
8K@30FPS H264/H265

Axera AX8850 SoC
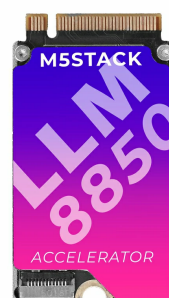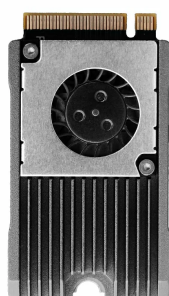
Description

LLM-8850 Card is an **M.2 M-KEY 2242** AI acceleration card designed for edge devices, combining a compact 42 mm form factor with the Axera AX8850 SoC delivering 24 TOPS @ INT8 performance. It enables host devices such as Raspberry Pi 5, RK3588 SBCs, and x86 PCs to easily expand multimodal large-model and video analysis capabilities with a simple plug-and-play design. The card is equipped with an active cooling system featuring a micro turbine fan and aluminum alloy CNC fins, with fan speed intelligently controlled by the onboard EC based on temperature–current curves. Even under long-term full load, it can maintain stable low temperatures to prevent thermal degradation inside enclosed cases.

The onboard **DCDC + PMIC** power chain is managed in real time by the EC to achieve "power on demand, cooling on demand," greatly improving overall system stability. Supports **AXCL Runtime**, with C / Python APIs enabling one-click deployment of mainstream CNN, Transformer, LLM, and multimodal models such as YOLO-v8/11, CLIP, Whisper, Llama3.2, InternVL3, Qwen3; and utilizes the AX8850's VPU hardware pipeline for H.264/H.265 8K encoding/decoding with simultaneous encode-decode transcoding acceleration and scaling/cropping, bridging AI and video stream processing. Host devices can directly invoke the hardware video codec via ffmpeg.

Tutorial

LLM-8850 Card

This tutorial introduces the quick start process, model list, and advanced usage of the LLM-8850 Card.

Features

- Ultra-compact form factor: NGFF M.2 M-KEY 2242 size, supports PCIe 2.0 ×2 lanes plug-and-play
- High-performance NPU: 24 TOPS @ INT8, octa-core Cortex-A55 1.7 GHz CPU
- Intelligent cooling / power supply: onboard turbo fan + CNC aluminum alloy integrated heatsink, EC-monitored temperature-current-speed closed loop
- High-bandwidth memory: 64-bit LPDDR4x, 4266 Mbps speed, 8GB capacity
- Hardware video engine: 8 K @ 30 fps H.264/H.265 encoding, 8 K @ 60 fps decoding, supports 16-channel 1080 p parallel decoding
- Secure boot & encryption: AES / DES / 3DES / SHA-256 hardware security module
- Native AXCL: one-click operation for CNN, Transformer, CLIP, Whisper, Llama3.2, Qwen3, InternVL3 full-stack models, supports H.264/H.265 simultaneous encode-decode transcoding

Includes

- 1 x LLM-8850 Card

Applications

- Industrial / commercial SBC compute upgrade: local execution of object detection and defect recognition on Raspberry Pi 5, RK3588, TI AM62x boards
- Embodied intelligent robots: AMR / AGV / service robots with plug-and-play local "perception-decision-control" chains
- AIPC & edge intelligent terminals: provide offline Copilot, customer service Q&A, meeting subtitles & instant translation within mini PCs
- Intelligent NVR / NAS upgrades: adding multi-channel AI license plate capture, event summary search + hardware transcoding to existing storage boxes
- Smart interactive devices: enable low-latency local LLM + TTS dialog for voice assistants, smart doorbells, advertising displays
- AI vision gateways: real-time passenger flow statistics and dangerous behavior warnings at traffic intersections and park gates

Specifications

Specification	Parameter
SoC	Axera AX8850
CPU	Octa-core Cortex-A55 1.7 GHz
NPU	24 TOPS @ INT8
Video Encoder	8 K @ 30 fps H.264/H.265 encoding, supports scaling / cropping
Video Decoder	8 K @ 60 fps H.264/H.265 decoding, supports 16 channels 1080 p parallel decoding, supports scaling / cropping
Memory	64-bit LPDDR4x, 4266 Mbps, 8GB capacity
Storage	32Mbits QSPI NOR Flash (used for Bootloader only)
Form Factor	M.2 M-KEY 2242, PCIe 2.0 ×2
Cooling	Micro turbo fan + integrated aluminum alloy CNC heatsink, EC intelligent temperature control
Operating Ambient Temperature	0 ~ 60 °C
Full Load Temperature at Room Temp	70 °C
Power Supply	7W @ 3.3V
Product Size	42.6 x 24.0 x 9.7mm
Product Weight	14.7g
Package Size	66.0 x 44.0 x 13.5mm
Gross Weight	19.8g

Learn

Reminder

The device may become hot during operation. Do not touch it to avoid burns.

Power Supply

When connecting to a Raspberry Pi or other PCs, use a switching power supply adapter (non-PD protocol) with DC 5V@3A power supply capability. If a PD power supply adapter is used, it may fail to output the maximum power normally due to protocol compatibility issues, which could further cause abnormal device operation.

> **Device Requirements**
>
> Due to insufficient internal interrupt resources of the Raspberry Pi, when using PCIe to convert to 4-channel M.2, M.2 SSDs and LLM-8850 cannot be used simultaneously.
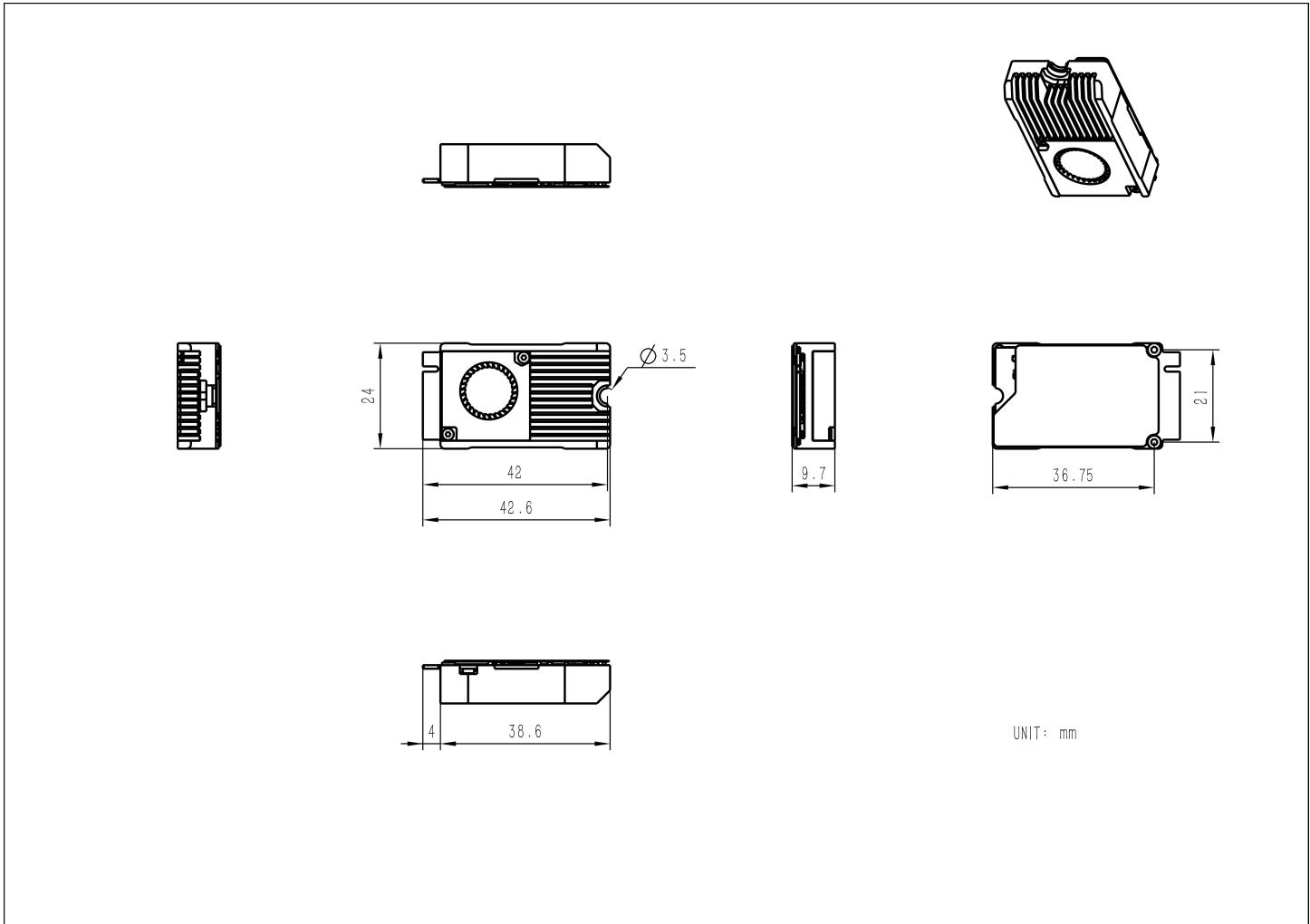
Hardware Compatibility

Feature	Supported
M.2 Direct Plug-in	✅
PCIe Direct Plug-in	❌
PCIe to M.2	✅
USB3 to M.2	❌
USB4 to M.2	✅

System Compatibility

Operating System	Supported
Ubuntu20.04	✅
Ubuntu22.04	✅
Ubuntu24.04	✅
Debian12	✅
Debian13	✅
Windows10	✅
Windows11	✅
macOS	❌
WSL	❌
VMware	❌
VBox	❌

Model Size

- LLM-8850 Card Model Size PDF

UNIT: mm

Softwares

Quick Start

- LLM-8850 Card User Guide

Video

- LLM-8850 Card Product Introduction and Function Demonstration

AI-001-LLM8850-video_EN.mp4