32T High Computing Power AI Box

AIBOX-1684X

AIBOX-1684

V1.0 2024-9-12

T-CHIP INTELLIGENCE TECHNOLOGY

Product features

32TOPS AI processor with ultra-high computing power

SOPHON AI Processor BM1684X/BM1684, features an octa-core ARM Cortex-A53 with up to 2.3GHz of frequency and a 12nm lithography process. AIBOX-1684X has a computing power of up to 32Tops (INT8). AIBOX-1684 has a computing power of up to 17.6Tops (INT8).

Powerful multi-channel video AI performance

The AI box supports up to 32 channels of 1080P H.264/H.265 video decoding and 12 channels of H.264/H.265 video encoding, making it ideal for various AI applications such as face detection and license plate recognition on video streaming.

The private deployment of large language models

Support the private deployment of ultra-large-scale parameter models under the Transformer architecture, including large language models such as LLaMa2, ChatGLM, and Qwen, as well as large vision models like ViT, Grounding DINO, and SAM.

Private deployment of AIGC image generation models

Support the private deployment of the Stable Diffusion V1.5 image generation model in the AIGC field and Docker container management technology.

Product features

Multiple deep learning frameworks

Support traditional network architectures such as CNN, RNN, and LSTM; a variety of deep learning frameworks, including TensorFlow, PyTorch, MXNet, PaddlePaddle, and ONNX, as well as custom operator development.

Strong network communication capability

With dual Gigabit Ethernet (1000Mbps/RJ45), the AI box ensures high-speed and stable network communication, meeting the needs of various application scenarios.

All-aluminum alloy enclosure for heat dissipation

The industrial-grade all-metal enclosure with aluminum alloy structure for thermal conduction. The side of the top cover features a grille design for efficient heat dissipation. Its top cover is a porous hexagonal design, combining elegance with high efficiency. The compact, exquisite device operates stably and meets the needs of various industrial-grade applications.
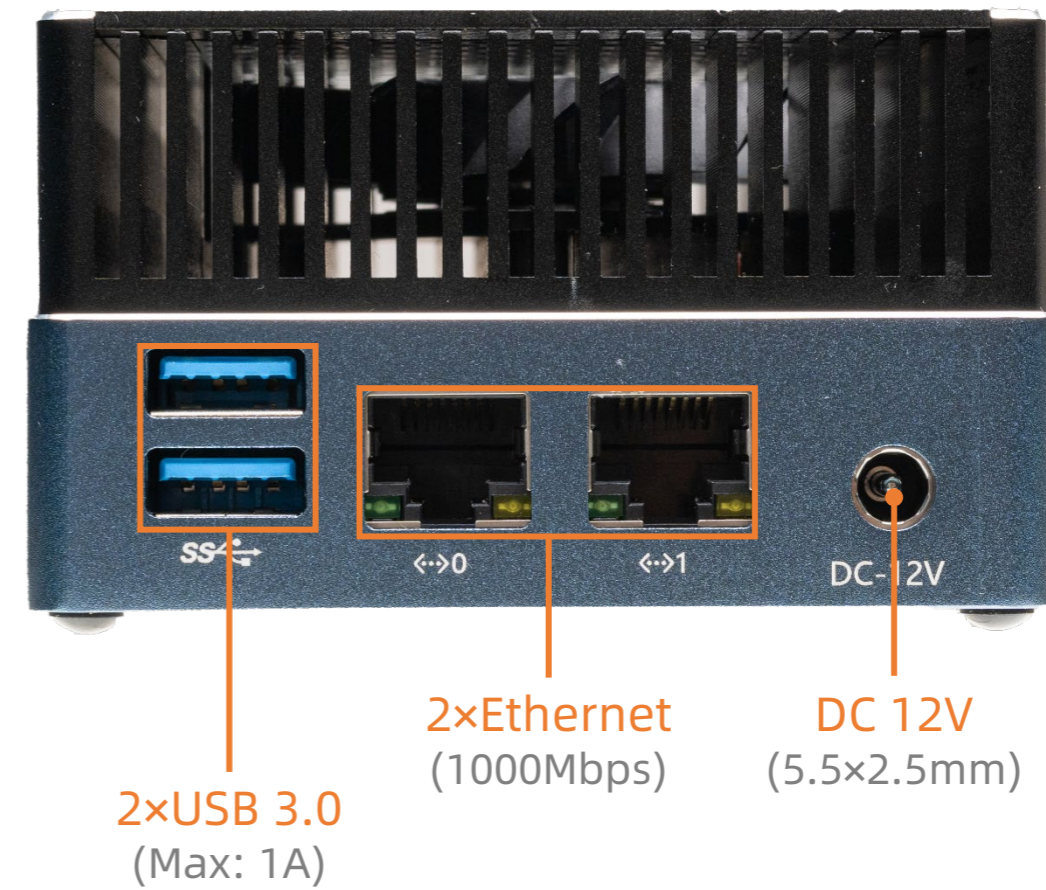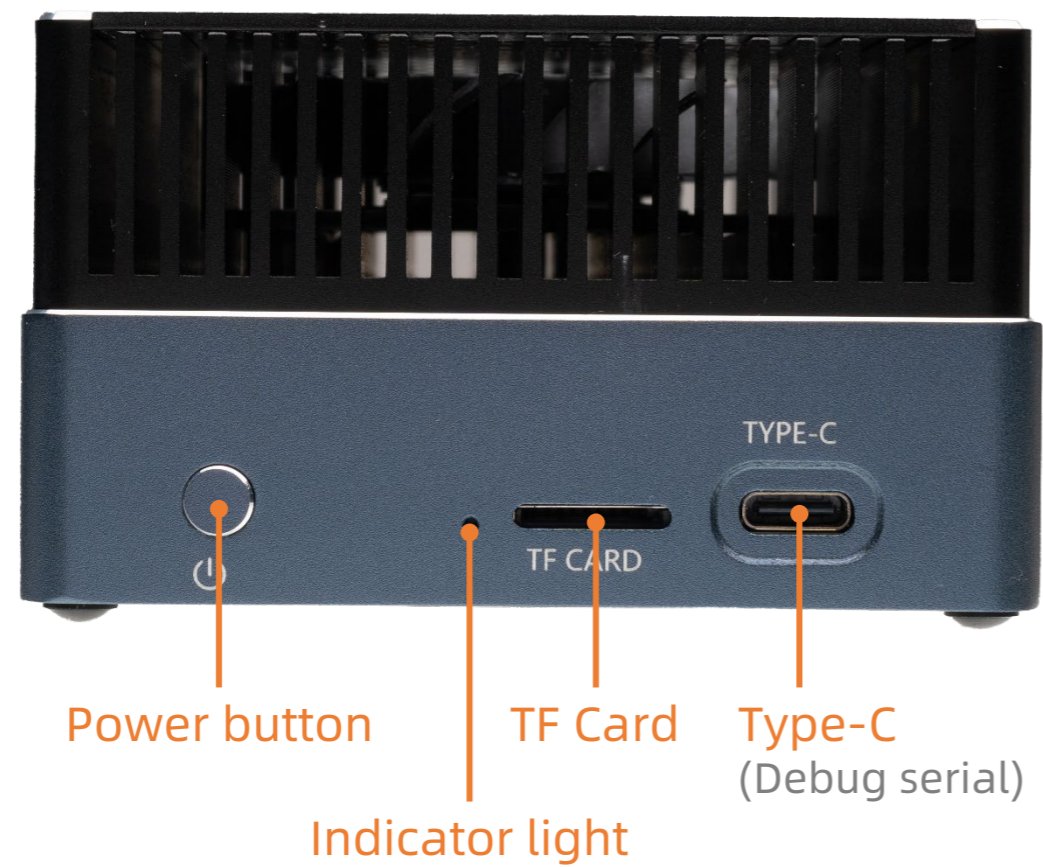
A wide range of applications

The device is widely used in intelligent surveillance, AI education, services based on computing power, edge computing, private deployment of large models, and data security and privacy protection.
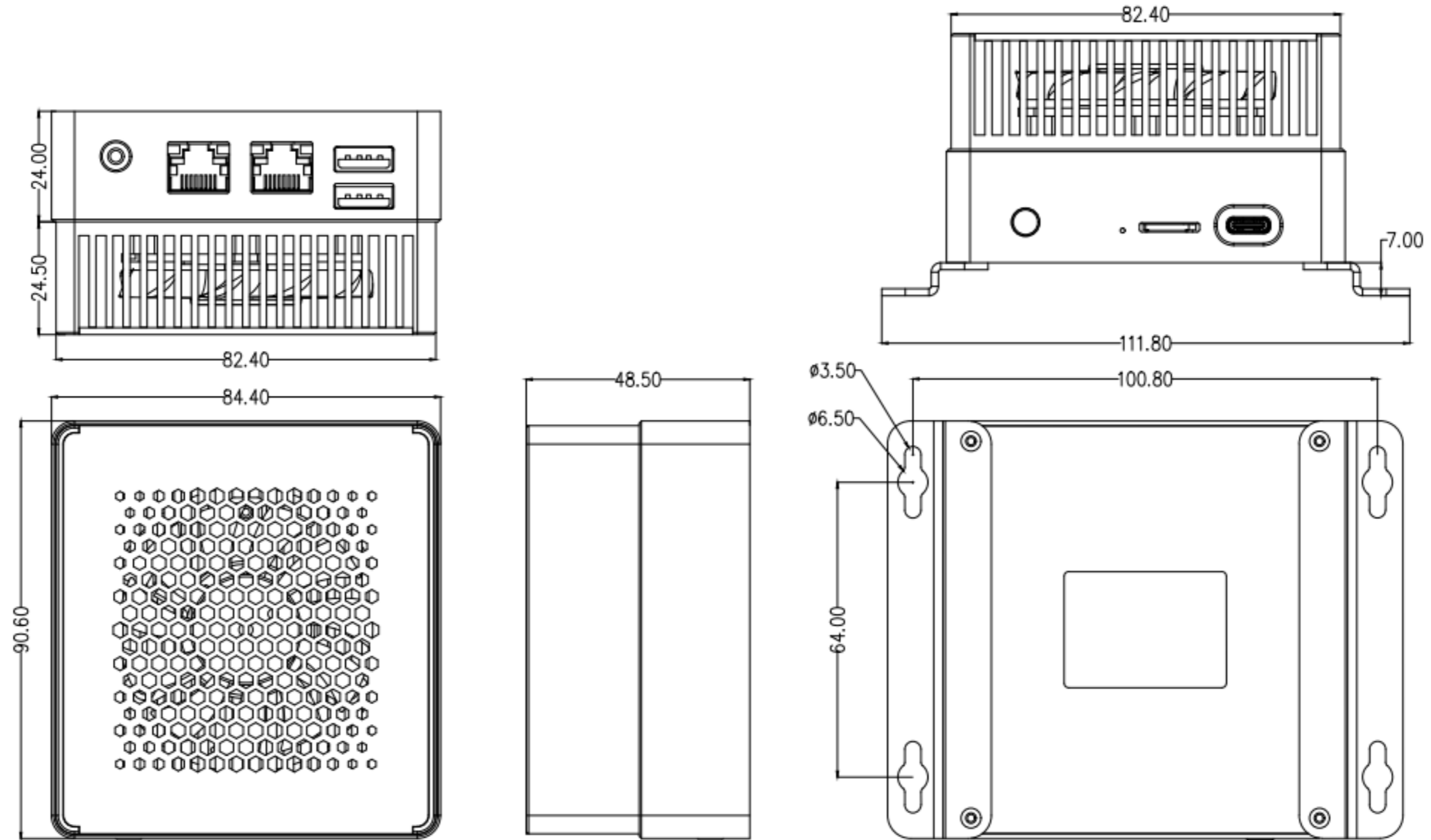
Specifications

		AIBOX-1684X	AIBOX-1684
Basic Specifications	SOC	SOPHON BM1684X	SOPHON BM1684
	CPU	High-performance octa-core ARM A53, 12nm lithography process, frequency up to 2.3 GHz	
	TPU	32TOPS (INT8), 16TFLOPS (FP16/BF16), 2TFLOPS (FP32)	17.6TOPS (INT8), 2.2TOPS (FP32), 35.2TOPS (INT8, enable winograd)
	VPU	32-channel H.265/H.264 1080P@25fps video decoding, 1-channel H.265 8K@25fps video decoding 32-channel 1080P@25fps HD video processing (decoding + AI analysis), 12-channel H.265/H.264 1080P@25fps video encoding JPEG image encoding and decoding can reach 1080P@600fps	32-channel H.265/H.264 1080P@30fps video decoding 2-channel H.265/H.264 1080P@25fps video encoding MJPEG image encoding and decoding can reach 1080P@480fps
	RAM	8GB/12GB/16GB LPDDR4/LPDDR4X	6GB/12GB/16GB LPDDR4/LPDDR4X
	Storage	32GB/64GB/128GB eMMC, 1 × TF Card	
	Power	DC 12V/4A (5.5 × 2.5mm)	DC 12V/3A (5.5 × 2.5mm)
	Power consumption	Normal: 20.4W(12V/1700mA), Max: 43.2W(12V/3600mA)	Normal: 9.6W(12V/800mA), Max: 26.4W(12V/2200mA)
	OS	Linux	
	Software Support	• The private deployment of ultra-large-scale parameter models under the Transformer architecture, including large language models such as LLaMa2, ChatGLM, and Qwen, as well as major visual models like ViT, Grounding DINO, and SAM. • The private deployment of the Stable Diffusion V1.5 image generation model in the AIGC field. • Traditional network architectures such as CNN, RNN, and LSTM; a variety of deep learning frameworks, including TensorFlow, PyTorch, MXNet, PaddlePaddle, Caffe and ONNX, as well as custom operator development • Docker container management technology	• Traditional network architectures such as CNN, RNN, and LSTM; a variety of deep learning frameworks, including TensorFlow, PyTorch, MXNet, PaddlePaddle, Caffe and ONNX, as well as custom operator development • Docker container management technology
	Dimension	90.6mm × 84.4mm × 48.5mm	
	Weight	≈ 420g	
	Environment	Operating temperature: -20℃ ~ 60℃, Storage temperature: -20℃ ~ 70℃, Storage humidity: 10% ~ 90%RH (non-condensing)	
Interface Specifications	Ethernet	2 × Gigabit Ethernet (1000Mbps/RJ45)	
	USB	2 × USB3.0 (Max: 1A), 1 × Type-C (Debug serial)	
	Button	1 × Power button	

Interface description

Power button

Indicator light

TF Card

Type-C
(Debug serial)

TYPE-C

TF CARD

2×USB 3.0
(Max: 1A)

2×Ethernet
(1000Mbps)

DC 12V
(5.5×2.5mm)

SS

0

1

DC-12V

Dimension

T-CHIP INTELLIGENCE TECHNOLOGY

Contact Us
(+86)18688117175

E-mail
global@t-firefly.com

Website
https://en.t-firefly.com/

Address
Room 2101, Hongyu Building, #57 Zhongshan 4Rd, East District, Zhongshan, Guangdong, China.

Firefly