Firefly

CSA1-N8S1684X

8-node 256TOPS cluster server

Product features

256TOPS powerful computing power

The SOPHON AI processor is BM1684X and supports up to 256TOPS (INT8) peak computing power or 128TFLOPS (FP16/BF16) or 16TFLOPS (FP32) high-precision computing power, which can meet the application requirements of deep learning model development.

Powerful multi-channel video full-process processing

It supports up to 256 channels of H.265/H.264 1080p@25fps video decoding, 256 channels of H.265/H.264 1080P@25fps HD video processing (decoding + AI analysis), and 96 channels of H.265/H.264 1080p@25fps video encoding, meeting the needs of various AI application scenarios such as face detection, license plate recognition, and smart cities in video streaming.

Standard 1U rack server

Highly dense and tightly deployed. The server can be configured with up to 8 BM1684X computing modules, and users can customize the number of computing modules and storage configuration at will. The standard 1U rack server chassis design is designed to fit most types of racks in the data center.

Product features

Rich algorithms and strong practicality

It supports the transplantation of various algorithms such as human/vehicle/object recognition, video structuring, trajectory behavior analysis, etc., with high security and high reliability, and can be flexibly applied to various product research and development.

Open SDK, one-stop AI development kit

SOPHON SDK (BMNNSDK2) is a one-stop deep learning development toolkit that provides a series of software tools such as underlying driver environment, compiler, inference deployment tool, etc. Support mainstream frameworks such as Caffe/TF/PyTorch/Mxnet/Addle, support mainstream network models and custom operator development, support Docker containerization, and enable rapid deployment of algorithm applications.

A wide range of applications

It is widely used in application scenarios such as edge computing, cloud storage, blockchain, multi-channel video encoding and decoding, intelligent security, and multi-application opening.
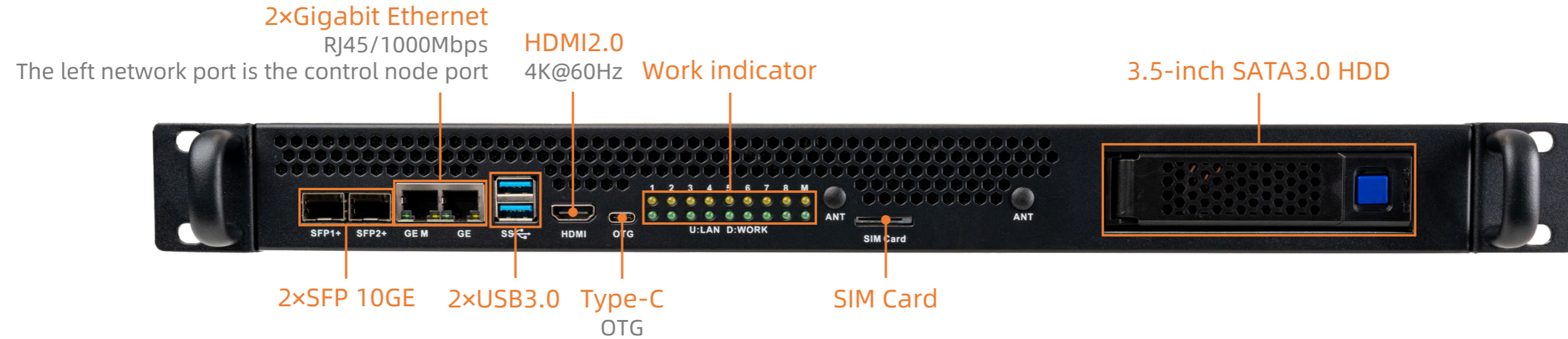
Specifications

		Specifications
Basic Specifications	Product name	Computing power cluster server
	Product model	CSA1-N8S1684X
	AI computing power	256TOPS (INT8) peak computing power, 128TFLOPS (FP16/BF16) computing power, 16TFLOPS (FP32) high-precision computing power
	Encoding/Decoding	H.265/H.264 1080p@6400fps video decoding, 1080P@6400fps HD video processing (decoding + AI analysis), H.265/H.264 1080p@2400fps video encoding
	Number of nodes	8 compute nodes (up to 64 ARM cores) + 1 control node
	CPU	Compute node: BM1684X octa-core (A53×8) 64-bit processor, with a maximum frequency of 2.3 GHz Control node: RK3588S octa-core (4× Cortex-A76+4×Cortex-A55) 64-bit processor, with a maximum frequency of 2.4GHz
	RAM	16GB LPDDR4 × 8 (Number of Compute Nodes)
	Storage	32GB eMMC × 8 (Number of Compute Nodes), 3.5-inch SATA 3.0 hard drive × 1
	Power	300W AC power supply (input: 100V AC~240V AC)
	Power consumption	Normal: 160W~180W, Max: 220W~280W
	OS	Linux
	BMC	Integrated BMC management system, providing a web-based management interface, BMC management system can be redeveloped
	Deep learning framework	TensorFlow / PyTorch / Paddle / Caffe / ONNX / MXNet / DarkNet
	Size	Standard 1U rack servers: 490mm × 390mm × 44.4mm
	Environment	Operating Temperature: 0ºC ~ 50ºC, Operating Humidity: 10% ~ 90%RH(non-condensing)
Interface Specifications	Internet	SFP 10GE × 2, Gigabit Ethernet (RJ45) × 2 (1 control node port, 1 common network port), 4G LTE/5G network (optional)
	Display	HDMI2.0 × 1 (4K@60Hz, main processor core board display)
	USB	USB3.0 HOST × 2, Type-C × 1 (For processor core board debugging)
	Fan module	6 high-speed cooling fans

Interface description

Front view

2×Gigabit Ethernet
RJ45/1000Mbps
The left network port is the control node port

HDMI2.0
4K@60Hz

Work indicator

3.5-inch SATA3.0 HDD

2×SFP 10GE

2×USB3.0

Type-C
OTG

SIM Card

Rear view

Power module

300W AC power supply (input: 100V AC~240V AC)

Fan module

Dimension

390mm

44.4mm

490mm

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.