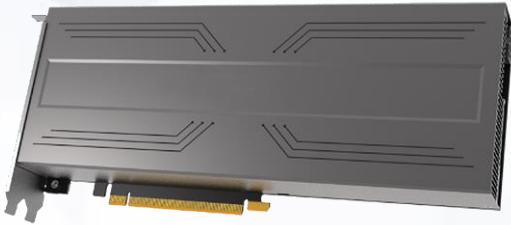


Smart Computing Card SC7 224T



Smart Computing Card SC7 224T is a computing power card for cloud AI applications. It can be used for large language model (LLM) inference, text-to-image generation, etc. Also, it can efficiently adapt to mainstream AI algorithms in the market to realize audio and video auditing, audio and video transcoding, structuring, behavior analysis, etc. These features empower various fields such as smart cities, smart transportation, smart energy, smart finance, smart telecommunications and smart industry.

Great Computing Power

- Peak performance 224 TOPS INT8
- Peak performance 112 TFLOPS FP16/BF16
- Support INT8/FP16/BF16/FP32 mixed precision calculations

Excellent Encoding and Decoding Capabilities

- 256 Channels 25fps HD video hardware decoding
- 96 Channels 25fps HD video hardware encoding
- Up to 8K resolution of video and image decoding

Wide Adaptation

- Adapt to various servers
- Adapt to various operating systems and mainstream algorithm frameworks

Application Scenarios



AIGC



Smart
Transportation



Smart Power



Smart Oil
and Gas



Smart Cities



Smart
Business Hall



Smart
Manufacturing



Smart
Warehousing

Product Specifications

Product Number		SC7 224T
CPU		64-core A53 338560 DMIPS 2.3GHz
Deep learning computing power	INT8	224 TOPS peak performance
	FP16/BF16	112 TFLOPS peak performance
	FP32	14 TFLOPS peak performance
Video/image coding and decoding performance	Video decoding capacity	H.264 & H.265: 1080P @6400fps
	Video decoding resolution	8K / 4K / 1080P / 720P / D1 / CIF
	Video coding capacity	H.264 & H.265: 1080P @2400fps
	Video coding resolution	4K / 1080P / 720P / D1 / CIF
	Image coding and decoding capacity	JPEG: 3200 images/second @1080P (interleave off)
	Image coding/decoding maximum resolution	32768 * 32768
Memory	Standard	128GB LPDDR4x 1024bit 470GB/s
PCIe	Data channel	PCIe Gen4 X16
Power	Maximum power	300W
Working environment	Temperature range	0°C ~ 45°C
	Method of heat dissipation	Passive
Structure size	Length * height * width	full height, full length, and double width
Deep learning frameworks		TensorFlow / PyTorch / Paddle / Caffe / ONNX / MXNet / DarkNet
Operating system support		Ubuntu / CentOS / Debian / UOS / Kylin
Main control CPU support		Compatible with Intel、AMD、Feiteng、Haiguang、Shenwei、Loongson, etc

Application-optimized Software SDK



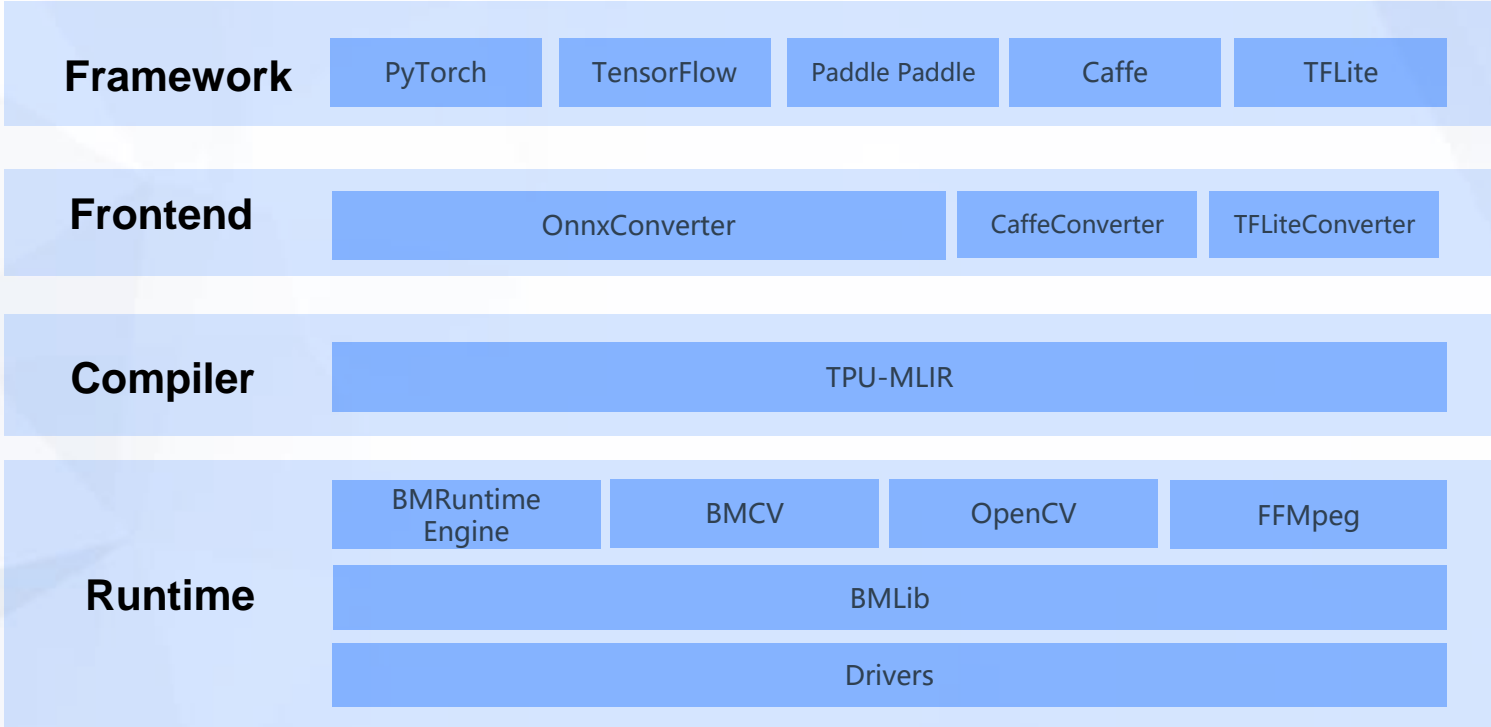
- Support various deep learning frameworks: TensorFlow, PyTorch, ONNX, PaddlePaddle, Caffe, etc
- Built-in 150 operator implementations and user-defined operators
- Support dynamic tensor size
- Support control flow networking
- Support user quantitative model import
- Support automatic quantitative strategy search
- Supports immediate data verification during compilation



- Efficiently adapt to various deep learning algorithms Support large model reasoning, such as ChatGLM, LLaMA, Falcon
- Support AI drawing, SD V1.5, SDXL
- Support various image classification and detection networks, such as Resnet50, Yolo series
- Support various natural language processing networks, such as BERT series
- Support various voice networks, such as Tacotron2



- Provides powerful development libraries (BMCV, FFMPEG coding and decoding interface, OPENCV coding and decoding interface, OPENCV extension)
- Support Docker containerization, Kubernetes extension management



Unified cloud-edge-end tool chain