



# AI Accelerator & GPU

Flexible Utilization for Superlative AI Acceleration

# Flexible Utilization for Superlative AI Acceleration

Boost your edge AI with drop-in accelerator modules designed for seamless performance gains. Our PCIe, MXM, and AI Accelerator products ensure easy integration into existing infrastructures without changes to your current infrastructure. Whether accelerating deep learning, image processing, or other demanding AI tasks, get the best of efficiency and adaptability. When space is at a premium, small form factor solutions like Aetina's MXM leverages parallel processing in ultra-compact COTS packages, bringing top-tier computing to embedded systems. Accelerate the AI performance you need, even in low-power consumption and limited spaces.

## PCIe

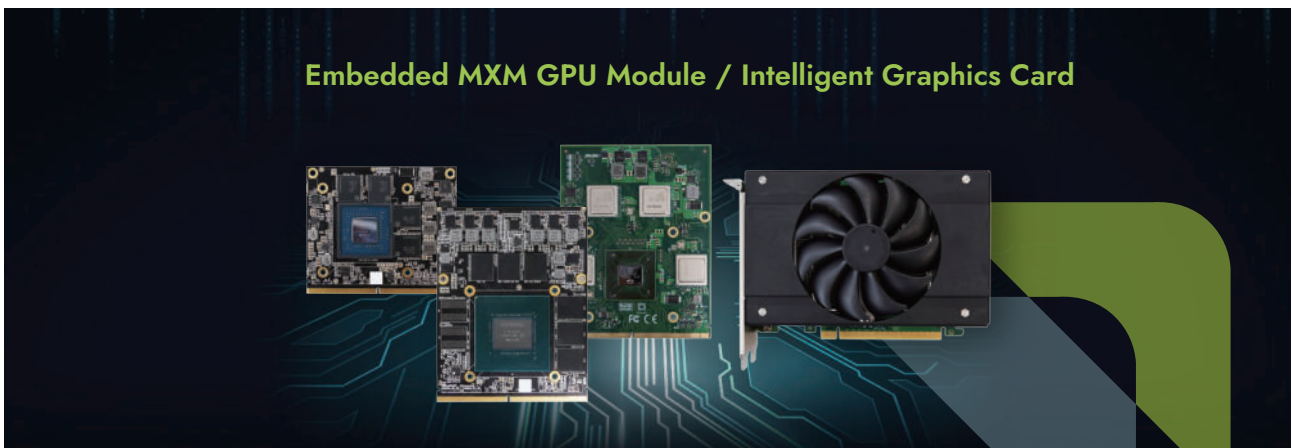
Create stunning visual experiences and AI-accelerated applications without any delay using Aetina's line of plug-and-play PCIe products powered by NVIDIA and Intel.

## MXM

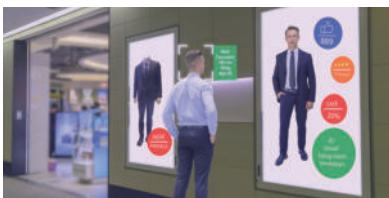
Aetina's MXM utilizes cutting-edge NVIDIA GPUs and AI ASIC processors, delivering unparalleled power efficiency and high-level computing capabilities. It is ideal for embedded systems that require high performance along with size, weight, and power (SWaP) optimization for edge AI applications.

## M.2

With low-power deep neural network inference, Aetina's M.2 is suitable for a broad range of market segments.



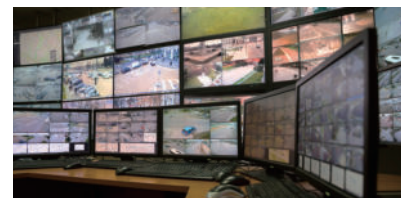
## Industry Applications



Retail



Logistics



Security



Gaming

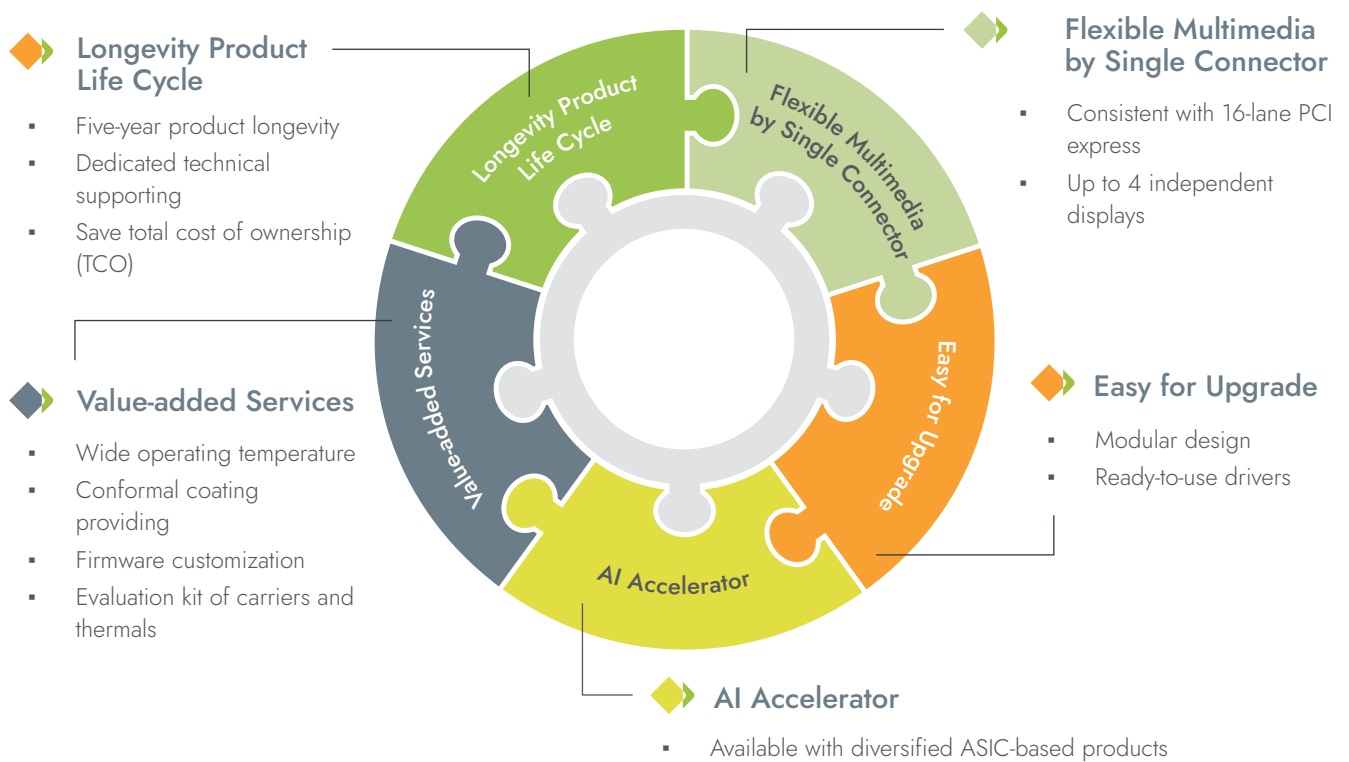


Medical



Transportation

# Purpose-Built for Embedded Vision AI



## Key Features



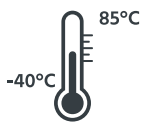
### Slim and Space-saving

- Small footprint and light weight
- Mounted flat to save mechanism space



### Golden Finger 30μ"

- Extra protection from scratch and damage
- Ensure stable and quality signals



### Extended Temperature

- Option temperature support : -20~70°C, -40~85°C
- Individual validation before shipping



### Conformal Coating

- Protection against dust, moisture and corrosion
- Improve MTBF



### CUDA Computing

- Up to thousands of CUDA cores
- Optimized parallel computing



### Visual Computing

- Dedicated for AI acceleration
- Real-time image processing



### Multi Displays

- 3840x2160 resolution
- DP++, HDMI outputs



### Configurable TDP

- Power cap customization
- Meet specific usage scenario

# PCIe Series

## Based on Intel Arc Graphics



### IA380-QUFL \*Preliminary

- Powered by Intel Arc A380E graphics
- Intel XMN 128 cores and 8 Xe-core, 6GB GDDR6 memory
- PCIe Gen 4 x8 interface
- 4.096 TFLOPS peak FP32 performance

## Based on NVIDIA RTX Graphics



### N4060-VSFX-A1

- Powered by NVIDIA RTX 4060 Ada Lovelace architecture
- 3072 CUDA cores, 24 RT cores and 96 Tensor cores, 8GB GDDR6 memory
- PCIe Gen 4 x16 interface
- 15.11 TFLOPS peak FP32 performance



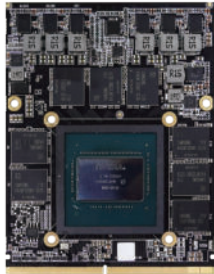
\*Preliminary



Model Number	IA380-QUFL	N4060-VSFX-A1
Engine	Intel Arc A380E Architecture: Intel Arc Intel XMN Cores: 128 Xe engine: 128 Xe-core: 8 Core Clock (MHz): 2000(Base) / 2250(Boost)(TBD) Floating Point Performance: 4.096 TFLOPS	NVIDIA GeForce RTX 4060 Architecture: NVIDIA Ada Lovelace CUDA Cores: 3072 Tensor Cores: 96 RT Cores: 24 Core Clock (MHz): 1830(Base) / 2460(Boost) Floating Point Performance: 15.11 TFLOPS
Memory	Size: 6GB GDDR6 Clock: 15.5 Gbps Interface Width: 96-bit Bandwidth (GB / sec): 186	Size: 8GB GDDR6 Clock: 17 Gbps Interface Width: 128-bit Bandwidth (GB / sec): 272
Support	PCI Express 4.0 x8 DirectX: 12 Ultimate OpenVINO OpenCL 3.0 OpenGL 4.6 Vulkan 1.3 Intel Deep Link	PCI Express 4.0 x8 DirectX: 12 Ultimate Open GL 4.6 Vulkan 1.2
Display	Resolution: 7680x4320 Max: 4x Display Display connector: Mini DisplayPort 2.0	Resolution: 7680x4320 Max: 4x Display Display connector: DisplayPort1.4a / HDMI 2.1
Power Consumption	Total Graphics Power (TGP): 50-75 W (TBD) Min. System Power Requirement2: 350W or better power supply (TBD)	Total Graphics Power (TGP): 115 W Min. System Power Requirement2: 550 W Supplementary Power Connectors): 8-PIN
Form Factor	Low-profile / Single slot	ATX
Dimension (W x D)	169.69 x 68.85 mm (6.68" x 2.71")	152.05 x 114.15 mm (5.98" x 4.49")
Net Weight	0.248 kg (0.5467 lb) (TBD)	0.401 kg (0.8841 lb)
Temperature	Standard Operating Temp. : 0°C ~ +55°C Storage Temp. : -40°C ~ +85°C	Standard Operating Temp. : 0°C ~ +55°C Storage Temp. : -40 °C ~ +85°C
Humidity	10~90%, Non-condensing	10~90%, Non-condensing
OS Support	Windows 10 / 11 64-bit	Windows 10 / 11 64-bit / Linux
Certification	CE / FCC / UKCA	CE / FCC / UKCA

# MXM Series

## ◆ Based on NVIDIA Ada



### MX5000A-WP

- NVIDIA RTX 5000 Ada Generation Embedded GPU based on NVIDIA Ada Lovelace architecture
- 9728 CUDA cores, 76 RT cores and 304 Tensor cores
- 41.15 TFLOPS peak FP32 performance
- 16GB GDDR6 memory
- PCIe Gen 4 x16 interface
- Support Error Correction Code(ECC)



### MX3500A-SP

- NVIDIA RTX 3500 Ada Generation Embedded GPU based on NVIDIA Ada Lovelace architecture
- 5120 CUDA cores, 40 RT cores and 160 Tensor cores
- 23.04 TFLOPS peak FP32 performance
- 12GB GDDR6 memory
- PCIe Gen 4 x16 interface
- Support Error Correction Code(ECC)

## ◆ Based on AI ASIC



### MX2000A-VP

- NVIDIA RTX 2000 Ada Generation Embedded GPU based on NVIDIA Ada Lovelace architecture
- 3072 CUDA cores, 24 RT cores and 96 Tensor cores
- 12.99 TFLOPS peak FP32 performance
- 8GB GDDR6 memory
- PCIe Gen 4 x8 interface
- Support Error Correction Code(ECC)



### AI-MXM-H84A

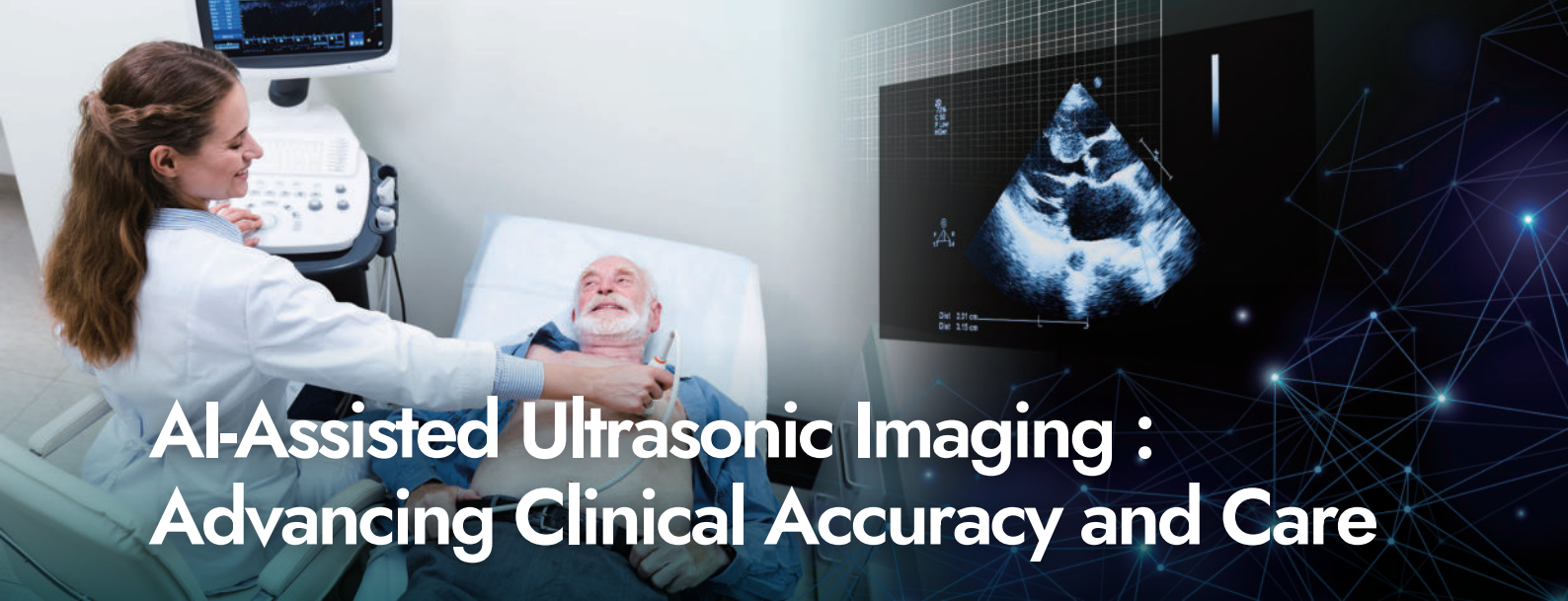
- Powered by 4 x Hailo-8 AI Processors
- MXM Type B small form factor
- Delivers up to 104 TOPS of AI performance at a typical power consumption of 25W
- Dedicated enablement S/W package and AI developer tools, with out-of-the-box support for state-of-the-art NN models





Model Number	MX5000A-WP	MX3500A-SP	MX2000A-VP	AI-MXM-H84A
<b>Engine</b>	NVIDIA RTX 5000 Ada Architecture: NVIDIA Ada Lovelace CUDA Cores: 9728 Tensor Cores: 304 RT Cores: 76 Floating Point Performance: 41.15 TFLOPS	NVIDIA RTX 3500 Ada Architecture: NVIDIA Ada Lovelace CUDA Cores: 5120 Tensor Cores: 160 RT Cores: 40 Floating Point Performance: 23.04 TFLOPS	NVIDIA RTX 2000 Ada Architecture: NVIDIA Ada Lovelace CUDA Cores: 3072 Tensor Cores: 96 RT Cores: 24 Floating Point Performance: 12.99 TFLOPS	4 x Hailo-8 AI processor with up to 26 TOPS and best-in-class power efficiency AI performance: 104 TOPS
<b>Memory</b>	Size: 16GB GDDR6 Speed: 18 Gbps Interface Width: 256-bit Bandwidth (GB/sec): 576	Size: 12GB GDDR6 Speed: 18 Gbps Interface Width: 192-bit Bandwidth (GB/sec): 432	Size: 8GB GDDR6 Speed: 16 Gbps Interface Width: 128-bit Bandwidth (GB/sec): 256	N/A
<b>Support</b>	PCI Express 4.0 x16 DirectX: 12 Ultimate Open GL 4.6 Vulkan 1.2	PCI Express 4.0 x16 DirectX: 12 Ultimate Open GL 4.6 Vulkan 1.2	PCI Express 4.0 x8 DirectX: 12 Ultimate Open GL 4.6 Vulkan 1.2	PCI Express 3.0 x16 Supported TensorFlow and ONNX
<b>Display</b>	Resolution: 7680x4320 Max: 4x DisplayPort	Resolution: 7680x4320 Max: 4x DisplayPort	Resolution: 7680x4320 Max: 3x DisplayPort	N/A
<b>Power Consumption</b>	Total Graphics Power (TGP): 115 W	Total Graphics Power (TGP): 115 W	Total Graphics Power (TGP): 60 W	25W (Typical power consumption)
<b>Form Factor</b>	MXM Graphics Module Version 3.1, Type B	MXM Graphics Module Version 3.1, Type B	MXM Graphics Module Version 3.1, Type A	MXM Graphics Module Version 3.1, Type B
<b>Dimension (W x D x H)</b>	82.0 x 105.0 mm (3.22" x 4.13")	82.0 x 105.0 mm (3.22" x 4.13")	82.0 x 70.0 mm (3.22" x 2.75")	82 x 105 mm
<b>Net Weight</b>	0.06 kg (0.1323 lb)	0.0592 kg (0.1305 lb)	0.037 kg (0.082 lb)	0.05 kg
<b>Temperature</b>	Standard Operating Temp. : 0°C ~ +55°C Storage Temp. : -40°C ~ +85°C	Standard Operating Temp. : 0°C ~ +55°C Storage Temp. : -40°C ~ +85°C	Standard Operating Temp. : 0°C ~ +55°C Extended Operating Temp. : -40°C ~ +85°C Storage Temp. : -40°C ~ +85°C	Standard: Operating Temp. : 0°C ~ +70°C Storage Temp. : -40°C ~ +85°C
<b>Humidity</b>	10~90%, Non-condensing	10~90%, Non-condensing	10~90%, Non-condensing	10~90%, Non-condensing
<b>OS Support</b>	Windows 10 / 11 64-bit	Windows 10 / 11 64-bit	Windows 10 / 11 64-bit	Windows 10 / 11 64-bit Linux 64-bit
<b>Certification</b>	CE / FCC / UKCA	CE / FCC / UKCA	CE / FCC / UKCA	CE / FCC

\*Product Specifications Are Subject To Change Without Prior Notice



# AI-Assisted Ultrasonic Imaging : Advancing Clinical Accuracy and Care

Investment in healthcare is sharply increasing, particularly AI in ultrasound. An AI-powered ultrasound solution can be implemented to improve patient outcomes by increasing the accuracy of diagnoses. One of Aetina partners, global manufacturers of diagnostic and prenatal ultrasound equipment, adopted Aetina ASIC-based MXM AI-MXM-H84A to speed up the auto-segmentation model to reduce image adjustment time and task operated manually.

AI-MXM-H84A drives unprecedented AI performance and provides the advanced image analytics deep learning model at a high-frame rate while in low latency. As a result, assessments become faster and more accurate.

## ► Benefits

- High flexibility and scalability to compliant with existing X86 or ARM systems
- Handles heavy inference workloads with low latency
- Comprehensive software package such as AI development tools and customization services

## ► Results

- Improves patient diagnosis outcomes through optimized accuracy
- Saves ultrasound AI system maintenance costs

