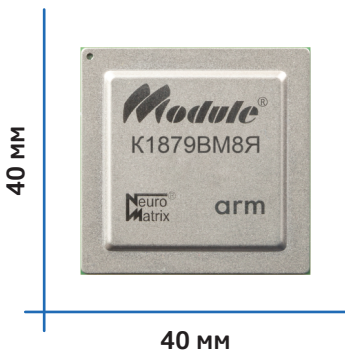




Mezzo Mini



на базе
СБИС K1879BM8Я



Спецвычислитель, подключаемый по шине PCIe к несущей плате.

Решает задачи реализации нейронных сетей, цифровой обработки сигналов и изображений.
Работает с 32- и 64-разрядными данными с плавающей точкой одинарной и двойной точности.

Технические характеристики

Процессор и топология:	Интерфейсы и память:	ПО и NN:
<p>DSP: 16 тензорных ядер NMC4 – 1 ГГц.</p> <p>RISC: 5 RISC ядер Arm Cortex A5 – до 800 МГц.</p> <p>Производительность: FP32 – 512 GFLOP/s. FP64 – 128 GFLOP/s.</p> <p>Техпроцесс: 28 нм КМОП.</p> <p>Мощность (типичая/максимальная): 12 Вт / 25 Вт.</p> <p>Температурный диапазон: -60С +85С.</p> <p>Корпус: 1444 HFCBGA.</p>	<p>5 Гб памяти DDR3L (до 32 Гб/с). PCIe 2.0x4 (Root и Endpoint). Ethernet 100 Мб/с. SPI. UART/CAN. RTC. GPIOx7. JTAG.</p>	<p>Драйверы: Windows (7,10), Linux (включая "AstraLinux" и "Эльбрус ОС").</p> <p>NMC SDK: IDE (open VS-code), компиляторы, отладчики ARM и NMC. Bare Metal SDK. БЗИО, БУПВ (MPI, GAS), OpenCL. Набор специализированных библиотек математических функций (BLAS, ЦОС).</p> <p>Neuro Matrix Deep Learning: Фирменный компилятор ИНС DarkNet и ONNX. Библиотека для применения ИНС. Библиотека трансляции моделей. Библиотека для подготовки изображений. Набор утилит для применения.</p>

Решает задачи



Производительность в реальных тестах

Ниже приведены значения производительности модуля при реализации (Inference) некоторых современных глубоких нейронных сетей.

FPS – число обработанных кадров изображения в секунду.

Latency – задержка до получения первого обработанного кадра.

Batch-mode – режим одновременной обработки нескольких кадров.

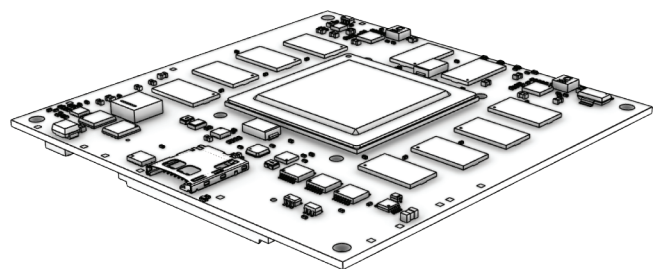
NMDL – производительность реализации нейронных сетей.

Нейронная сеть (размер изображения в пикселях)	FPS NM Card/ Card Mini, NM Mezzo/ Mezzo Mini (multi unit mode)	Latency NM Card/ Card Mini, NM Mezzo/ Mezzo Mini (multi unit mode)	NM Card/ Card Mini, NM Mezzo/ Mezzo Mini (batch mode)	Latency NM Card/ Card Mini, NM Mezzo/ Mezzo Mini (batch mode)
alexnet (227x227)	12,6	79	13	308
inception v3 (299x299)	12,8	78	20,3	197
inception v3 (512x512)	3,93	254	5,44	735
resnet 18 (224x224)	25	40	47	85
resnet 50 (224x224)	12,2	82	20,6	194
squeezenet (224x224)	74,4	13	100	40
u-net (512x512)*	2	500	2	2000
yolo v2 tiny (416x416)	21	47	30,4	132
yolo v3 (416x416)	3,7	270	4,5	889
yolo v3 tiny (416x416)	27,3	36	35,3	113
yolo v5s (640x640)	3,3	303	3,4	1176

* В модели u-net произведена замена слоев transposed convolution на upsampling

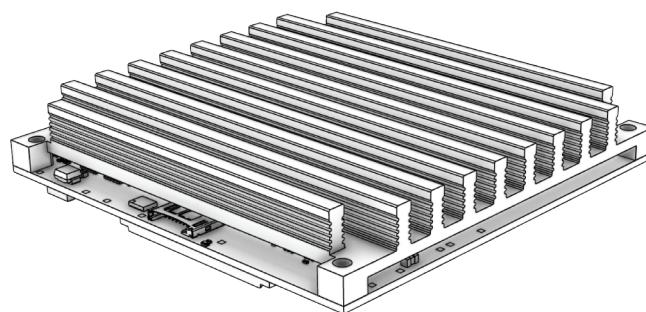
Габариты модуля

Исполнение без радиатора



100x100x13 мм

Исполнение с радиатором



100x100x25 мм

Масса модуля 300 гр.



Руководство по эксплуатации
доступно по ссылке

