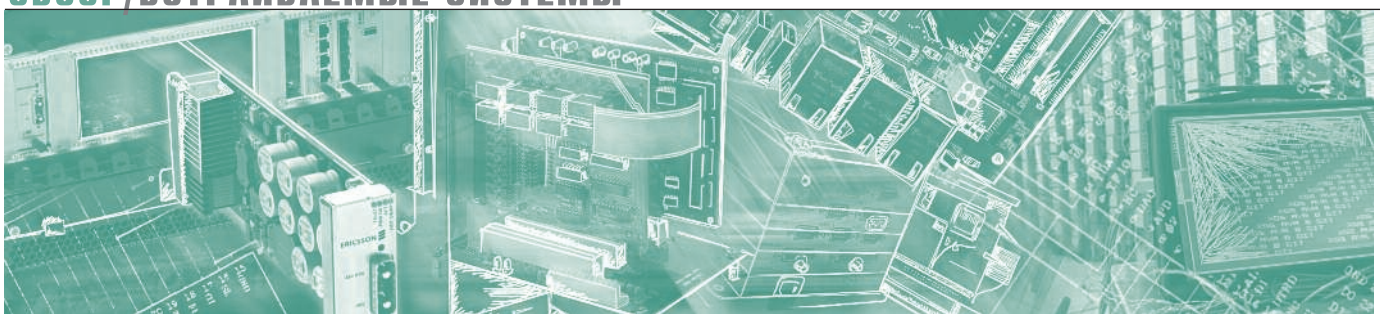


Юрий Широков

Лошадиные силы для искусственного интеллекта

Компания IEI Integration Corp. – один из ведущих поставщиков промышленных компьютеров. В последнее время IEI предлагает гибкие и высокоинтегрированные решения для умных городов в области транспортных систем, домашней автоматизации, интеллектуального управления розничной торговлей, телемедицины. В статье приведены примеры программно-аппаратных решений IEI.

В нашу эпоху информационного взрыва появляются различные цифровые сервисы, такие как OTT (Over-the-Top – сетевая доставка видеоконтента на устройство пользователя), MSO (Multiple-System Operator – многосистемный кабельный оператор), CDN (Content Delivery/Distribution Network – сеть дистрибуции контента), SaaS (Software as a Service – программное обеспечение как услуга), сталкивающиеся с нехваткой вычислительных ресурсов. Поставщикам цифровых услуг требуются вычислительные мощности, способные справиться с огромным количеством информации, данных, аудио и видео. В прошлом, в отсутствие особых лимитов по объёмам, для повышения производительности мы просто увеличивали число серверов, но теперь требования компактности становятся всё существеннее. Для экономной модернизации сервера можно использовать только карты расширения FPGA или GPU, часто имеющие ограниченную функциональность. Но, к счастью, создано решение – интеллектуальные универсальные компактные вычислительные ускорители семейства Mustang.

Mustang-200

Это продвинутый, доступный и масштабируемый акселератор для ускорения компьютерных вычислений и работы приложений (рис. 1). Карта формата PCIe оснащена двумя процессорами Intel® Core™ i5 или i7, 32 Гбайт (4×8 Гбайт) оперативной памяти и твердотельным накопителем Intel NVMe ёмкостью 1 Тбайт (2×512 Гбайт). Она может быть встроена в существующую систему, без больших финансовых затрат обеспечивая ей высокую производительность вычислений. Вы можете установить даже несколько Mustang-200 и распределять задачи между ними по мере необходимости, что ещё более увеличит вычислительные возможности (табл. 1). Вот вкратце преимущества описываемой платы.

Рис. 1. Общий вид Mustang-200

Таблица 1

Сравнение карт-акселераторов

Характеристика	Ускорители общего назначения	Специализированные ускорители на базе ASIC, FPGA	Гибкие решения на базе архитектуры x86
Многозадачность	Нет	Нет	Да
Гибкость	Низкая	Низкая	Высокая
Энергопотребление	Высокое	Низкое	Низкое
Стоимость разработки	Высокая	Высокая	Низкая

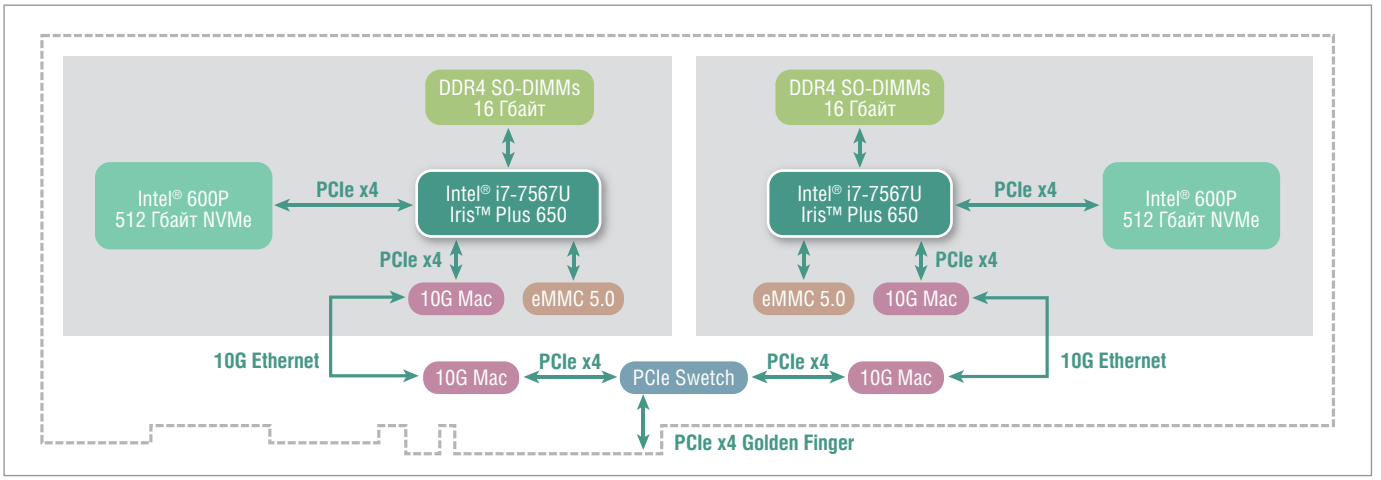


Рис. 2. Структура Mustang-200

- Поддержка узлов децентрализованной вычислительной архитектуры на базе 10-гигабитного Ethernet и вычислительной архитектуры x86.
- Прекрасно интегрированная ОС QNAP QTS Lite обеспечивает гибкую и безопасную среду разработки.
- Технология виртуализации: виртуальная машина (VM) и технология контейнеров.
- Плата встраивается в стандартный сервер, совместимый с PCI Express x4, x8, x16.
- Даёт возможность увеличить вычислительную мощность без изменения или добавления серверов.
- Позволяет достичь более высоких плотностей вычислений при снижении общей стоимости.

Рассмотрим более подробно структуру Mustang-200 (рис. 2). Каждый процессор карты Mustang-200 оснащён 16 Гбайт оперативной памяти (2×8 Гбайт) и 512 Гбайт твердотельным накопителем Intel® серии 600P NVMe. После установки в слот PCIe x4 хост-компьютер подключается к обоим вычислительным узлам Mustang-200 с помощью сети 10 GbE. Преимущество сетевой структуры заключается в том, что не требуется никакого проприетарного оборудования, за счёт чего достигается более низкая стоимость и универсальность. Вычислительные узлы работают под управлением QTS Lite – облегчённой версии удостоенной наград операционной системы QTS QNAP, предустановленной на eMMC (SSD-накопитель, интегрированный на плату).

Итак, компания IEI, уже 20 лет специализирующаяся на разработке промышленных компьютеров и платформ для систем автоматизации, создала аппаратную платформу, а QTS, поддерживающая технологию виртуализации, подходит для создания приложений по обработке и защите данных. Вместе они образуют оптимальное сочетание для разработчиков различных приложений.

В ускорителе Mustang-200 каждый процессор работает независимо, так что можно «привязать» задачи к любым узлам по вашему выбору и осуществлять контроль работы каждого узла в режиме реального времени. Mustang-200 не требует никакого дополнительного оборудования: он просто устанавливается в существующую систему. Если не хватает текущей вычислительной мощности, вы всегда можете добавить дополнительные Mustang-200, так как они работают независимо друг от друга. Максимальное количество плат ограничено только числом доступных слотов PCIe x4 в системе (рис. 3).

Рис. 3. Параллельная работа Mustang-200

Рис. 4. Запуск контейнерных приложений на акселераторе Mustang-200

Благодаря высокой производительности и масштабируемости эта плата подходит и для облачных вычислений: при взаимодействии с облачными сервисами есть возможность предварительно обработать данные в вашей организации локально, а затем загрузить в облачную платформу только подготовленную информацию. Таким образом достигается серьёзная экономия расходов на трафик при использовании облачной платформы.

Интегрированная операционная система QTS Lite поддерживает такие технологии виртуализации, как контейнеры и виртуальные машины. Можно преобразовать свою физическую систему в виртуальную (P2V) и назначить её одному из узлов на базе Mustang-200. Неважно, какое программное обеспечение используется, вы можете таким образом портировать его в Mustang-200, будь то ПО искусственного интеллекта, программы для научных исследований или различные симуляторы. При этом Mustang-200 обеспечивает вычисления в режиме реального времени, пакетные и параллельные вычисления, автоматическую балансировку загрузки вычислительных узлов, объединение нескольких карт в кластер под управлением ОС QTS Lite, а также имеет API для управления операционной системой.

Возможности Mustang-200

Контейнерные приложения

Контейнерная станция QNAP объединяет технологии LXC (Linux Container) и Docker® Lightweight Virtualization, позволяющие управлять несколькими независимыми системами Linux® на сетевом накопителе QNAP, а также загружать тысячи приложений (рис. 4). Контейнерная станция расширила концепцию JeOS (Just enough OS) и использует лёг-

Рис. 5. Обеспечение информационной безопасности

кую технологию виртуализации, позволяющую разработчикам и IT-администраторам свободно переключаться между Mustang-200 и облаком. Это даёт возможность быстро инсталлировать микросервисы, легко разворачивать приложения IoT, исполнять огромное количество популярных приложений.

Виртуальная машина

Станция виртуализации QNAP является полноценным решением для виртуализации на основе x86 IEIxQNAP Mustang-200. Она позволяет управлять несколькими виртуальными машинами (VM), запущенными на Mustang-200. Станция виртуализации добавляет Mustang-200 исключительную многосторонность: благодаря ей можно создавать вычислительную среду высокой плотности на базе виртуальных машин, запуская в них программы или службы, которые до этого работали на нескольких отдельных ПК.

Видео и большие данные в реальном времени

Транскодирование (кодирование и декодирование) видеопотоков – также достойная сфера применения Mustang-200. Мощные процессоры Mustang-200 легко обрабатывают изображения высокой чёткости 360° объёмного видео. Сетевая структура Mustang-200 также отлично подходит для рендер-ферм, где требуется много параллельных вычислительных ресурсов.

- Для ускорения транскодирования мультимедиа совместно с центральным процессором в Mustang-200 задействован графический процессор (GPU). Это позволяет повысить производительность в приложениях искусственного интеллекта (виртуальная реальность и глубокое обучение), при обработке больших данных, аналитики и в других инженерных приложениях.
- Как карта расширения Mustang-200 предлагает быструю и масштабируемую интеграцию с существующими стандартными архитектурами серверов, создавая платформу для потоковой передачи видео.
- Mustang-200 обеспечивает полное аппаратное транскодирование.

Mustang для разработчиков

Платформа разработки QPKG (система управления программными пакетами) предназначена для использования профессиональными разработчиками программного обеспечения, сетевыми и системными интеграторами, а также независимыми разработчиками программного обеспечения для интеграции оборудования и ПО и разработки приложений. Для разработки приложений можно использовать следующие возможности:

1) платформа разработки QTS Lite App (QPKG) позволяет разработчикам проектировать программные приложения, работающие на Mustang-200;
2) Development Toolkit (API & SDK): разработчики могут создавать программные приложения для смартфонов или ПК, реализующие удалённый доступ и управление файлами на Mustang-200.

Для обеспечения информационной безопасности и защиты в системе QTS Lite присутствуют 7 встроенных механизмов защиты (рис. 5), которые сэкономят усилия и время на разработку приложений, что позволит сосредоточиться на сути задачи.

В заключение данного раздела приведён свод технических характеристик платы и описание расположенных на ней коннекторов (табл. 2, рис. 6, 7). Далее мы поговорим об одном из самых

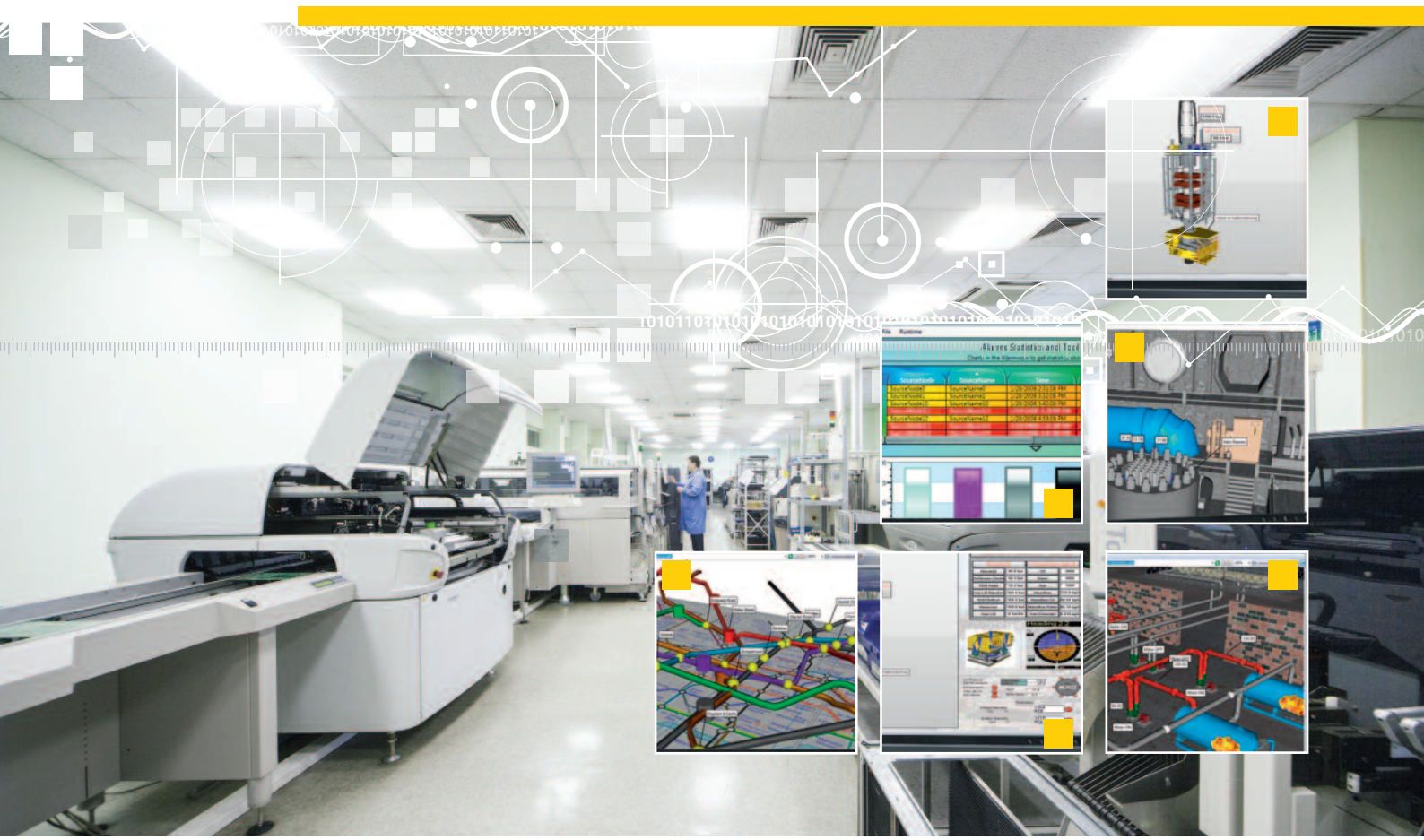
Таблица 2

Спецификация Mustang-200

Базовый чипсет	Два Intel® Kabylake ULT CPU
	Intel® Core™ i7-7567U (28 W), кэш-память 4 Мбайт, до 4,00 ГГц
	Intel® Core™ i5-7267U (28 W), кэш-память 4 Мбайт, до 3,50 ГГц
Графическая подсистема	Intel® Iris™ Plus Graphics 650 (GT3e) • базовая частота 300 МГц • максимальная динамическая частота 1,05 ГГц • встроенная графическая DRAM 64 Мбайт на ЦП
Аппаратное декодирование видео	H.264, H.265/HEVC
	MPEG2, M/JPEG
	VC-1
	VP8 (8 бит)/VP9 (10 бит)
Аппаратное кодирование видео	H.264, H.265/HEVC
	MPEG2, M/JPEG
	VC-1
	VP8 (8 бит)
Память	4×DDR4 8 Гбайт SODIMM (2×DDR4 8 Гбайт SODIMM на каждый ЦП)
Накопитель	2×Intel® SSD 600P (512 Гбайт, M.2 80 мм PCIe 3.0 x4, 3D1, TLC)
Физический интерфейс PCIe	PCI Express x8
Интерфейс маршрутизации трафика	PCI Express Gen2 x4
Внешние интерфейсы	Кнопка рестарта
Индикатор	7-сегментный
Диапазон рабочих температур	0…+40°C
Охлаждение	Сдвоенный вентилятор
Габаритные размеры (Д×Ш×В)	40×210×111 мм
Диапазон рабочей влажности	10…90%

Анализ технологических показателей в реальном времени

Решения на базе программных продуктов ICONICS



ЧТО?

- Управление эксплуатацией оборудования
- Снижение затрат
- Энергоменеджмент

КАК?

- Диагностика состояния оборудования с возможностью прогнозирования сбоев. Учёт наработки, экспертные карты, вероятностный анализ
- Анализ нештатных режимов. Частота возникновения, поиск взаимосвязи, анализ времени реакции персонала
- Анализ потребления энергоресурсов. Данные о потреблении в реальном времени, сравнение с идеальной моделью и плановыми показателями, сравнение с историческими данными, индикаторы энергоэффективности. Поддержка анализа в рамках энергоменеджмента по ГОСТ 50001:2011

ICONICS

PortalWorX Productivity Analytics Facility AnalytiX Energy AnalytiX Alarm Analytics

НОРВИКС ТЕХНОЛОДЖИ

Тел.: +7 (495) 232-1817
Факс: +7 (495) 232-1649
Эл. почта: inform@norvix.ru

Официальный партнёр компании ПРОСОФТ
www.norvix.ru

Реклама

Рис. 6. Расположение коннекторов Mustang-200 с лицевой стороны

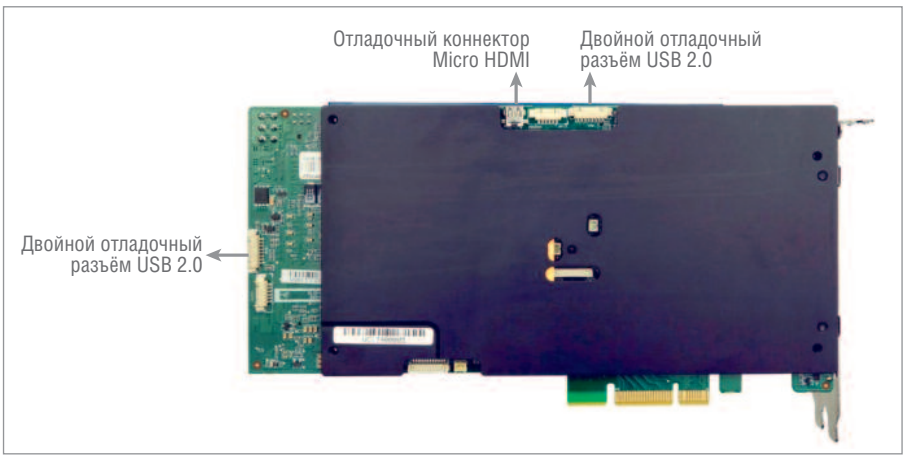
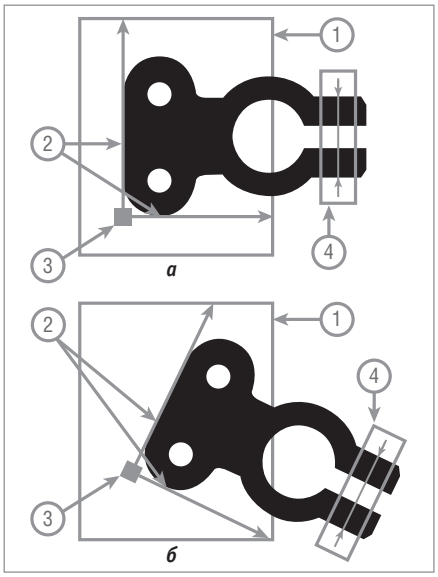


Рис. 7. Расположение коннекторов Mustang-200 с тыльной стороны



Условные обозначения: 1…4 – точки, по которым производится идентификация образа.

Рис. 8. Контроль изделий на конвейере (*а* и *б* – деталь на конвейере в различных положениях)

Рис. 9. Распознавание лиц в потоке

Рис. 10. Эволюция искусственного интеллекта

интересных и перспективных применений устройств акселерации вычислений – о системах искусственного интеллекта (ИИ), этого мостика в будущее, с которым многие связывают очередную технологическую революцию.

Нейронные интеллектуальные системы

Все перечисленные свойства и возможности ускорителей Mustang позволяют с успехом использовать их в приложениях на основе искусственного интеллекта. ИИ, в свою очередь, применим практически в любых приложениях, где востребованы компьютеры. Развитием первых систем ИИ являются системы машинного обучения, действующие на основе наборов правил и инструкций. Дальнейшая эволюция породила системы ИИ глубокого обучения на базе нейронных алгоритмов. В настоящее время именно они особенно эффективны при решении задач классификации и распознавания образов.

В двух словах разницу между «простым» машинным (Machine Learning – ML) и глубоким (Deep Learning – DL) обучением можно охарактеризовать так: при традиционном подходе за обучение машины отвечает человек, задающий компьютеру правила игры и проводящий в соответствии с ними обучение машины, при глубоком обучении машина способна сама генерировать правила и создавать по ним модели оптимального поведения. Примеры классического (ML) ИИ: компьютерное зрение (распознавание лиц, рукописного текста, шаблонов). Примеры ИИ с глубоким обучением (DL): распознавание речи, машинный перевод, классификация изображений.

Системы ML хорошо подходят для применений с анализом типовых систем высокой плотности, например, механических изделий на конвейере (рис. 8). DL имеет преимущества при большой вариативности исследуемых образов, например, при анализе качества пищевых продуктов, распознавании лиц в потоке, при медицинской диагностике (рис. 9). Несмотря на значительно более высокую сложность реализации и ресурсоёмкость, по мере роста производительности вычислительных устройств именно DL-системы становятся фаворитами.

Одной из перспективнейших алгоритмических технологий для распознавания и классификации изображений считается технология свёрточной нейронной сети (convolutional neural network, CNN), рис. 10. Поэтому CNN сегодня рассматривается как неотъемлемая часть технологий глубокого обучения. Её принципы были сформулированы еще в 1988 году французским учёным Яном Лекуном. Не вдаваясь в сложности матричных вычислений, суть работы свёрточной нейронной сети

InduKey®

УСТРОЙСТВА ВВОДА INDUKEY — НЕМЕЦКАЯ НАДЕЖНОСТЬ

- Различные типы клавиатур, в том числе силиконовые, вандалостойкие и стеклянные емкостные
- Степень защиты до IP68
- Различные варианты интерфейсов
- Модели для применения в промышленности, медицине, лабораториях и информационных системах
- Инновационные указательные устройства
- Разработка заказных изделий

ProSoft®

ОФИЦИАЛЬНЫЙ ДИСТРИБЬЮТОР

(495) 234-0636
INFO@PROSOFT.RU

WWW.PROSOFT.RU

Реклама

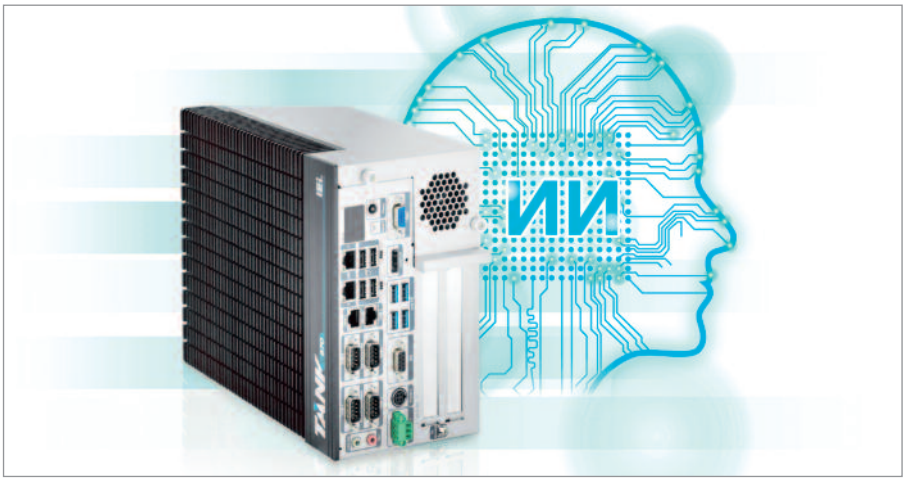


Рис. 11. Компактная AIoT-платформа TANK

можно сформулировать так: она абстрагируется от деталей в пользу поиска и выделения характерных существенных признаков изображения, которые затем используются в качестве ключевых при распознавании образов. В силу некоторых особенностей математического аппарата вычислительные алгоритмы свёрточной нейронной сети хорошо поддаются распараллеливанию, поэтому для их работы ускорители семейства IEI Mustang хорошо подходят.

Интеллектуальный TANK

TANK производства компании IEI (рис. 11) – это встраиваемая компактная платформа AIoT (от Artificial Intelligence и Internet of Things – «ИИ + Интернет вещей»), обладающая следующими достоинствами:

- поддержка ЦП Intel® Core™/Xeon® 6-го или 7-го поколения, чипсетов Intel® Q170/C236 и памяти DDR4;
- поддержка двух независимых дисплеев высокого разрешения;
- высокоскоростные интерфейсы ввода/вывода, сгруппированные на одной стороне корпуса;
- встроенный коннектор питания для обеспечения питанием плат расширения;
- отличные возможности расширения аппаратной части;
- предустановленная ОС Ubuntu 16.04 LTS;
- предустановлены пакет Intel® Distribution of Open Visual Inference & Neural Network Optimization (OpenVINO™) toolkit, Intel® Media SDK, Intel® System Studio и Arduino® Create.

Программный пакет OpenVINO™ базируется на свёрточных нейронных сетевых алгоритмах. Среда помогает оптимизировать предобученные модели глубокого обучения, такие как Caffe, MXNET, и Tensorflow. В общей сложности пакет содержит более двадцати предобученных и поддерживает более ста пользовательских моделей (включая Caffe, MXNet, TensorFlow, ONNX, Kaldi). Он позволяет пользователям легко разворачивать приложения глубокого обучения на базе архитектуры Intel, реализующие концепцию «один пакет

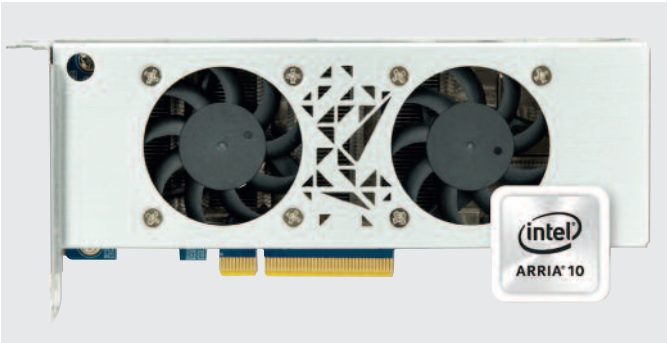


Рис. 12. Mustang-F100-A10

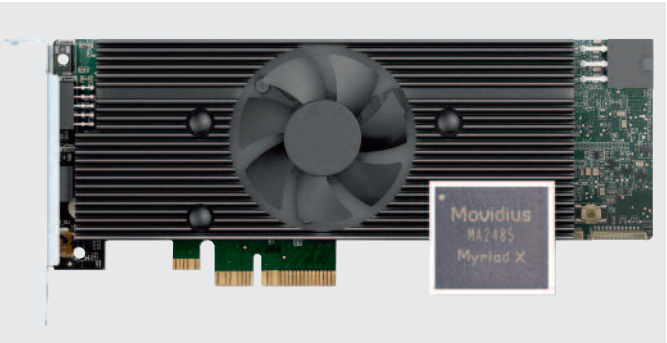


Рис. 13. Mustang-V100-MX8

SDK – множество платформ-ускорителей» (CPU, GPU, FPGA, VPU).

Как показано на рис. 11, TANK обладает двумя слотами для плат расширения. Это слоты PCIe (x16), которые, в частности, могут быть использованы для повышения вычислительных возможностей платформы путём добавления карт-акселераторов Mustang-200, Mustang-F100-A10 (ЦП – FPGA-матрица Intel® Arria® 10 GX1150, рис. 12) и Mustang-V100-MX8 (ЦП – восемь видеопроцессоров Intel® Movidius™ Myriad™ X MA2485, рис. 13).

Таким образом, TANK – это хороший выбор для построения систем ИИ, машинного зрения, контроля дорожного трафика, идентификации людей. Благодаря повышенной устойчивости к вибрации (с установленным SSD-накопителем TANK соответствует требованиям стандарта MIL-STD-810G 514.6 C-1) и расширенному диапазону рабочих температур (–20...+60°C) платформа может использоваться в мобильных приложениях на транспорте и в промышленности.

Заключение

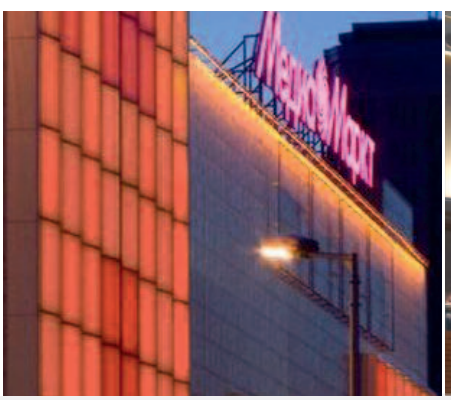
На базе передовых разработок компании IEI возможна реализация проектов практически в любых отраслях. Особенно выгодно будет применение устройств IEI там, где требуется компактность, повышенная надёжность и вычислительная мощность, а именно такие задачи становятся всё более распространёнными с развитием технологий Индустрии 4.0. Программно-аппаратные платформы IEI будут интересны не только системным интеграторам, но и конечным заказчикам, которым компания ПРОСОФТ готова предложить комплексные решения на их основе.

Статья подготовлена по материалам компании IEI

E-mail: textoed@gmail.com

XLight

Серия светодиодных прожекторов
для систем архитектурно-художественного освещения



Преимущества

– Компактные размеры
– Широкий диапазон рабочих температур –40…+60°C
– Степень защиты IP65
– Высокая вандалоустойчивость
– Широкая номенклатура вариантов исполнения
– Высокие экономичность и эффективность
– Гарантия 3 года

(495) 232-1652 / info@xlight.ru / www.xlight.ru

Реклама