

Выбор платформы для **встраиваемых систем,** **оснащенных машинным зрением**

Максим Сорока
ноябрь 2017
адаптация для PDF



Содержание

- Повышение информативности
- Платформы
- Примеры применения

Зачем нужна информация?

- Повышение надежности работы алгоритмов
- Статистический анализ
- Обучение нейронных сетей

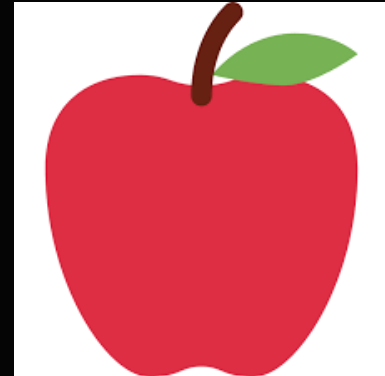
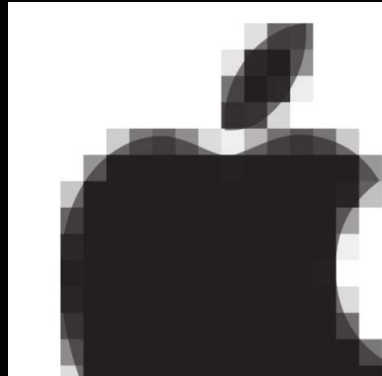
"Кто владеет информацией, тот владеет миром"
(Син-Цзы, Н. Ротшильд, У.Черчилль и др.).

Эффективное использование информации?

- Нужная **информация**
- Нужным **людям** (машинам, алгоритмам)
- В нужный **момент времени**

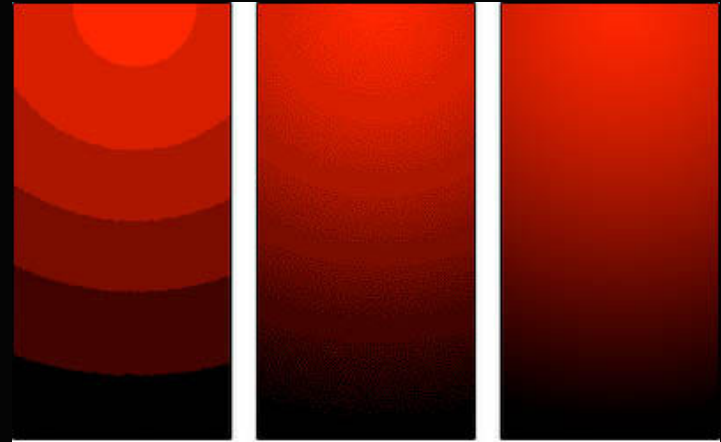
Изображение без сжатия

- Сохранение деталей
- Большой объем



Глубина цвета

Разрядность	Оттенки
8	256
10	1024
12	4096
14	16384



Объем данных

Размер кадра HD 1920 x 1080
Частота – 25 кадров в секунду
8 разрядов

H.264 Кадр* - 50 Кбайт
Поток - 10Мбит
10 камер на линию Ethernet 100Мбит
Запись на диск** до 200 камер

MJPEG Кадр* - 250 Кбайт
Поток - 50Мбит
2 камеры на линию Ethernet 100Мбит
Запись на диск** до 40 камер

* Усредненный размер сжатого кадра
** Скорость записи 250Мбайт в секунду

Кадр - 2 Мбайт (без сжатия)
Поток - **500Мбит**
2 камеры на линию Ethernet **1000Мбит**
7 камер на линию USB3 5000 Мбит (400МБ)
Запись на диск** до 5 камер



Повышение информативности

Технические параметры камер

Детализация:

- Исходные данные **без сжатия**
- **Высокое разрешение**
- Глубина цвета 8-10-12 разрядов
- **Частота** кадров



Повышение информативности

Технические параметры камер

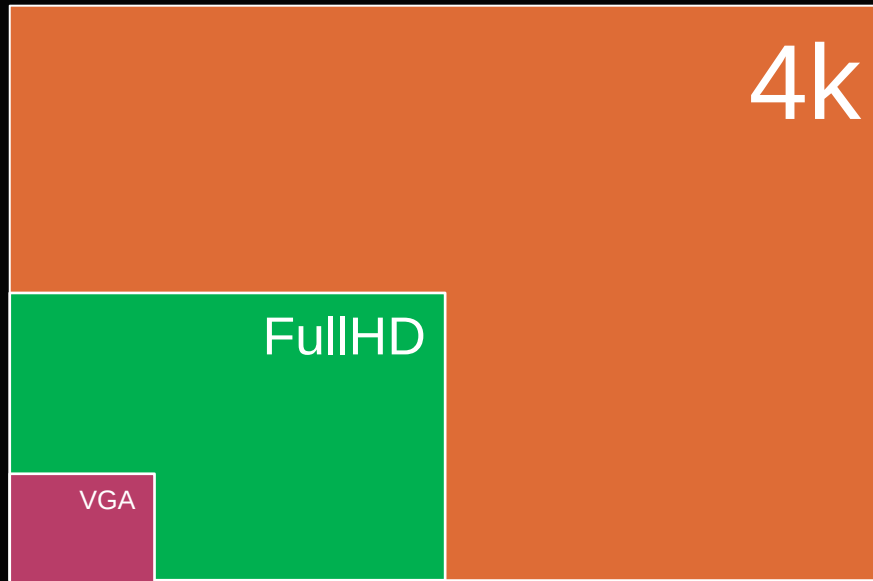
Управление процессом захвата изображений

- **Параметры съемки:** Экспозиция, ROI, подсветка, синхронизация и пр.
- **Совместная работа** с другими камерами (VIS, IR, ToF и пр.) и датчиками



Размер имеет значение

Увеличение размера кадра и скорости ввода изображений



300 fps!

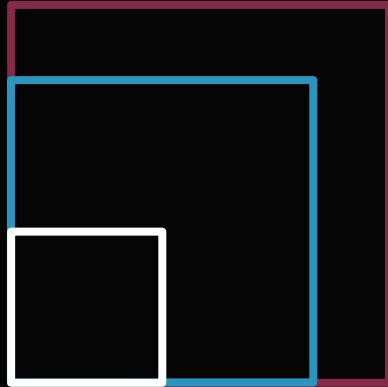
Размер имеет значение

Большой пиксель:

Чувствительность

Глубина резкости

Требования к оптике



$$2 \times 2 = 4$$

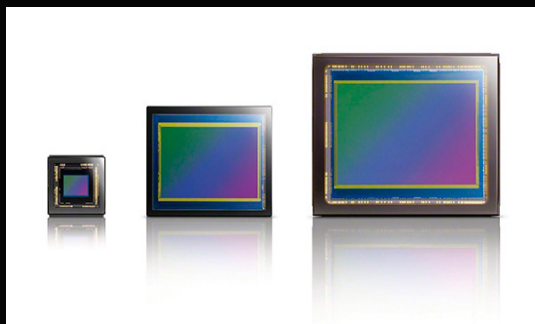
$$4 \times 4 = 16$$

$$5 \times 5 = 25$$

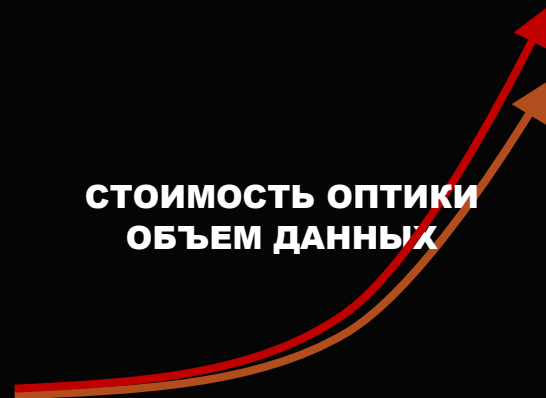
Размер имеет значение

Увеличение размера кадра влечет за собой требования к оптике и каналам связи

- o VGA (640x480) 1/3"-2/3";
- o HD (1920x1080) 2/3"-1";
- o HD 4k (4096x2160) 1.1" ;
- o 71 Мп 1+";



**СТОИМОСТЬ ОПТИКИ
ОБЪЕМ ДАННЫХ**



От мала до велика

Basler ace U серия: acA1920-40um
FullHD, 40fps SONY IMX249, 5.86мкм

€399*

Basler ace L серия: acA4112-30um
12Mp, 30fps SONY IMX253, 3.45мкм

€2499*

* Указана стоимость без НДС



Интерфейсы

Интерфейсы

- GigE – 1Gb;
- USB3 – 5 Gb;
- CameraLink – 10 Gb;
- CoaXPress – 25 Gb;
- CSI-2, 3xC-PHY - 24Gb;
- PCIe, Gen3 – 64 Gb.

2k FullHD (1920x1080, 8 bit) без сжатия

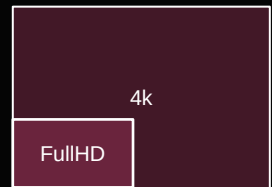
50 fps

165 fps

340 fps (4k 60fps)

600 fps (4k 100fps)

1500 fps (4k 300fps).



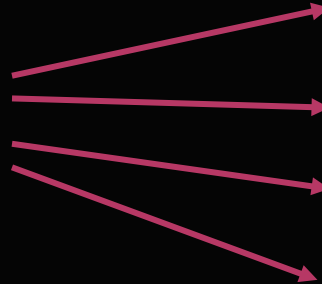
Разница в подходах



Готовый функционал



Видеонаблюдение
и запись.
Стандартные
алгоритмы.



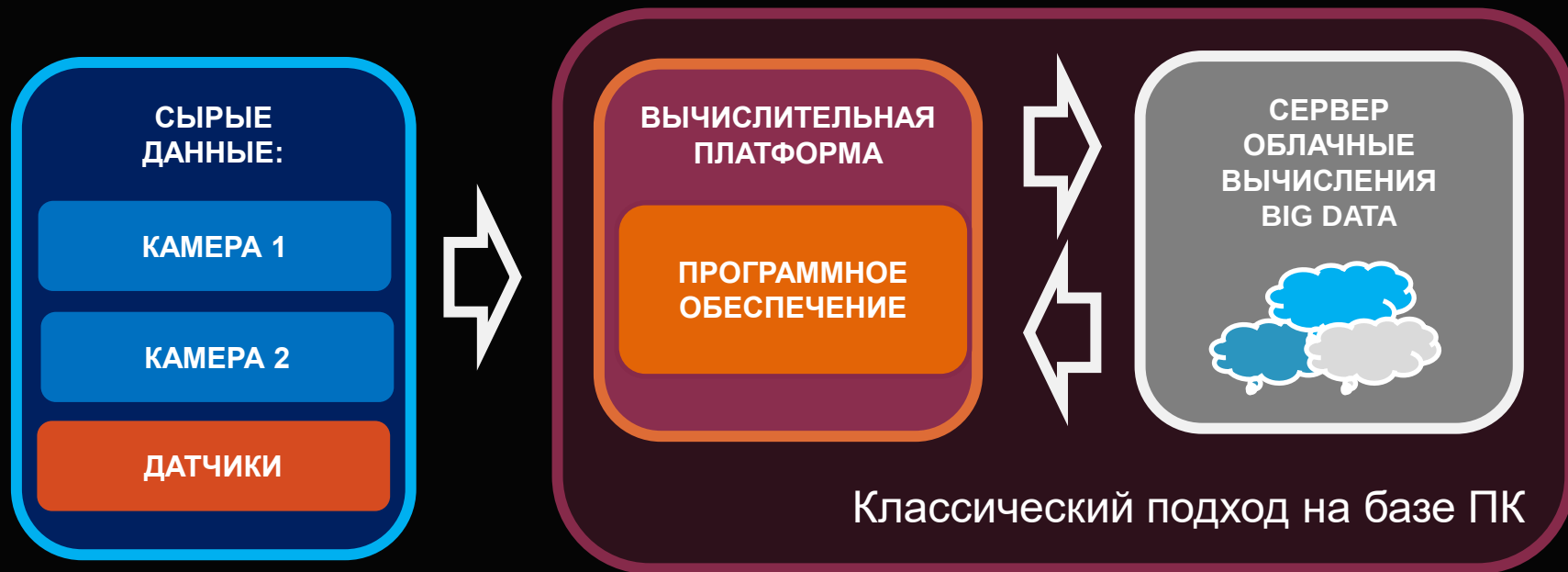
Гибкость и
возможность выбора

- Распознавание
- Идентификация
- Положение
- Отслеживание
- ...

Архитектура



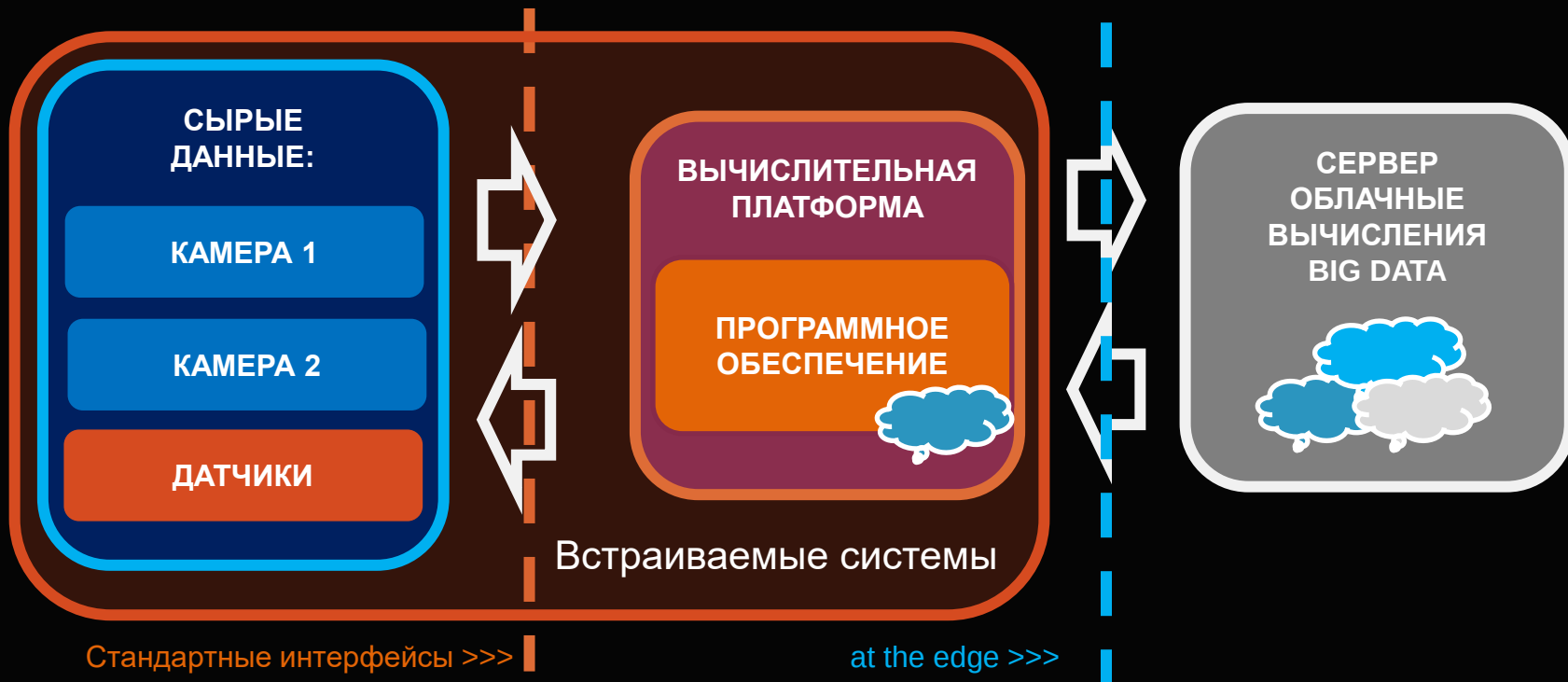
Классический подход



Интеллектуальные камеры



Встраиваемые системы



Повышение информативности

Встраиваемые системы: оборудование и средства разработки

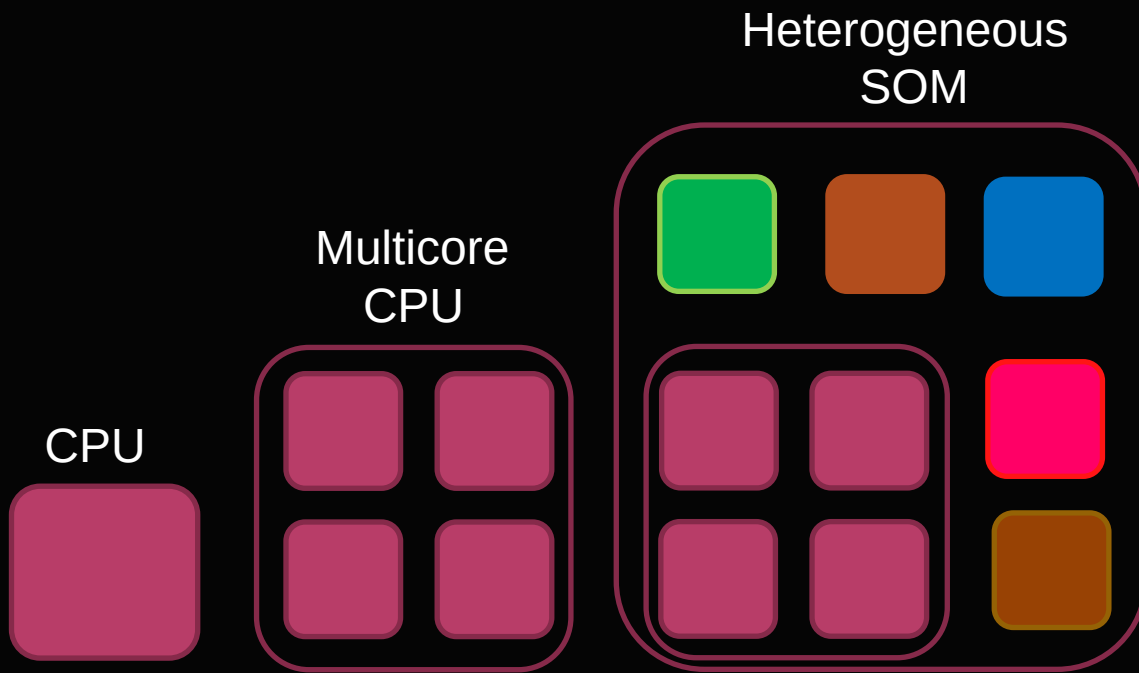
- Выбор вычислительной платформы
- Средства разработки программного обеспечения



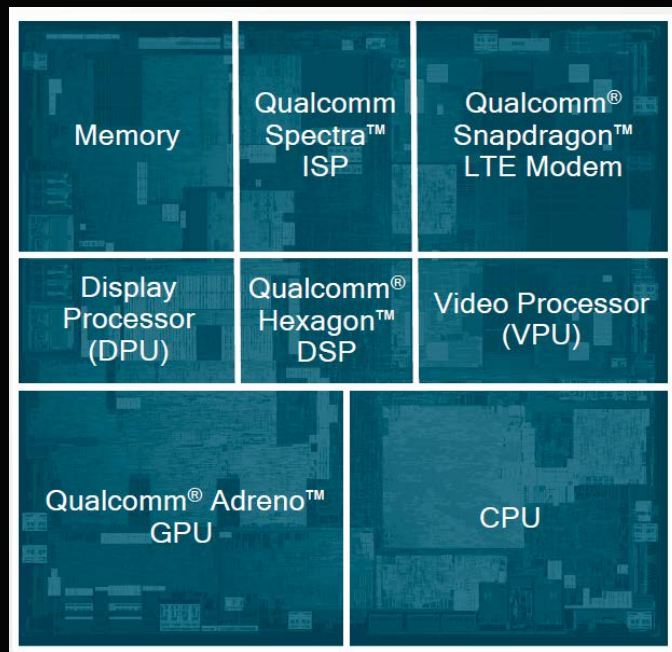
Выбор платформы

- ARM: Qualcomm
- x86: intel
- FPGA: XILINX ZinQ
- GPU: NVIDIA

Эволюция процессоров



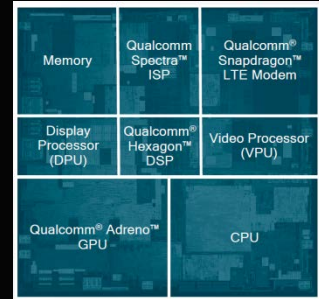
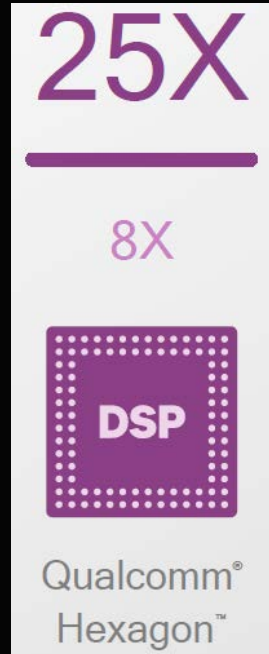
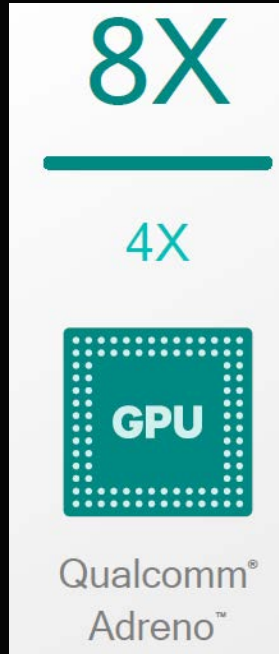
Qualcomm Snapdragon 625



Qualcomm Snapdragon DLN

Потребление

Производительность



Qualcomm® Snapdragon™ Neural Processing Engine (NPE)

Июль 2017

- Программная среда для реализации нейронных сетей на базе процессоров Qualcomm, Android и Linux(Ubuntu)
- Используется гетерогенная архитектура Qualcomm® Kryo™ CPU, Qualcomm® Adreno™ GPU или Qualcomm® Hexagon™ DSP
- Поддерживаются популярные архитектуры сетей Caffe, Caffe2 и Tensorflow, а также дополнительные слои

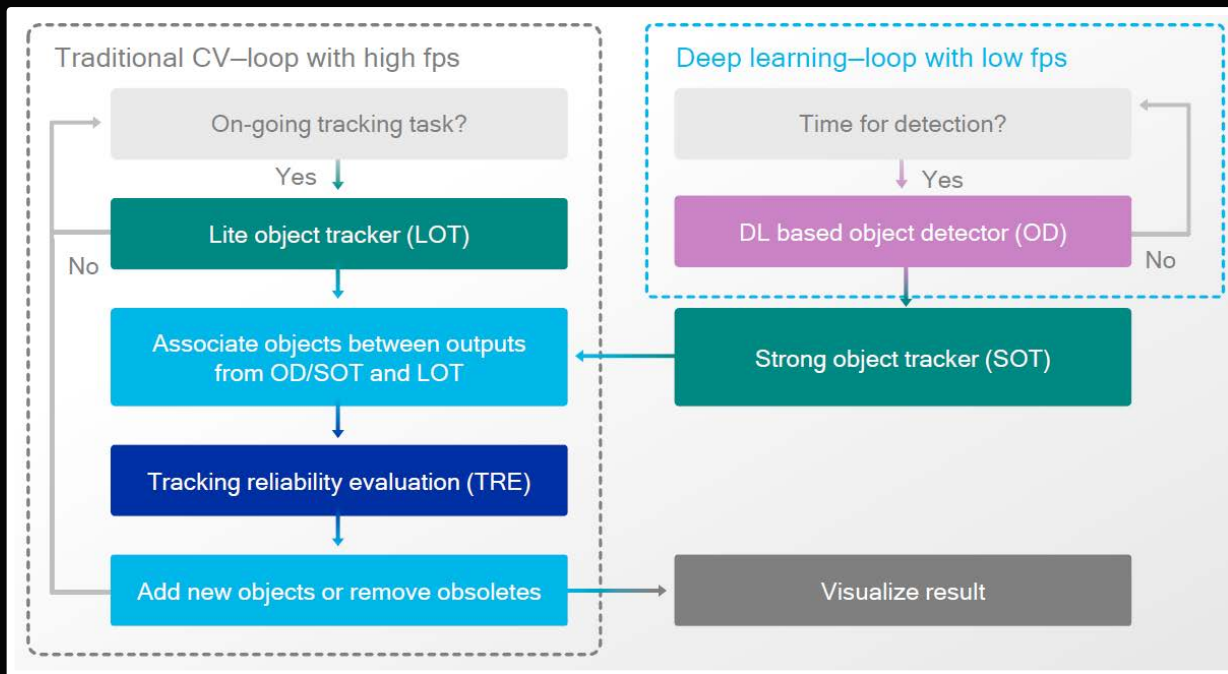
Гибридная архитектура

- **Машинное зрение:**
обработка изображений,
выделение областей
интереса, трекинг
объектов, идентификация
- **Нейронные сети:**
распознавание лиц,
классификация

KPIs	Pure DL solution (Detector + Classifier)	CV + Light-weight DL Classifier
MACs	~20G	~20M in average and ~150M in the worst case
Processing frame rate	2-3 fps	25fps
Memory BW (MBps)	1000	450
CPU Utilization	8 cores @ 90% util.	2 cores @ 60% util.

По материалам Qualcomm Technologies, Inc.

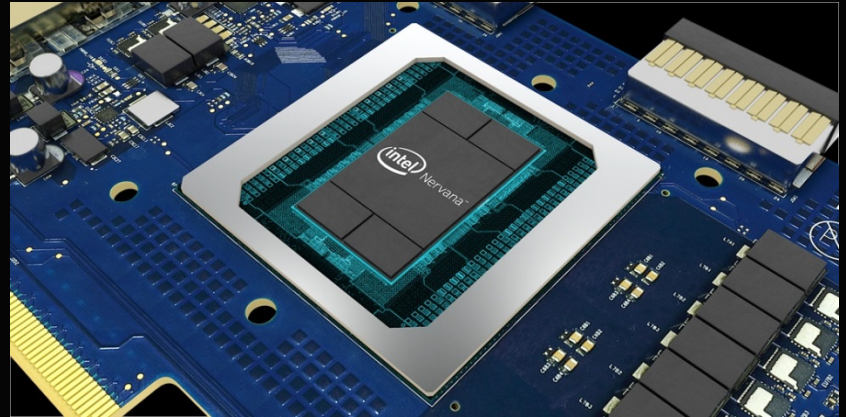
Два цикла



Intel

“...the goal we set last year of achieving 100 times greater AI performance by 2020”

Brian Krzanich is the chief executive officer of Intel Corporation.



Intel. Стратегия развития

- Слияния и поглощения
- Собственные разработки
- Программное обеспечение



Intel. Инвестиции

\$30+ миллиардов в течение последних трех лет

- Altera (2015)
- Itseez (2016)
- Nervana (2016)
- Movidius (2016)
- Mobileye (2017)

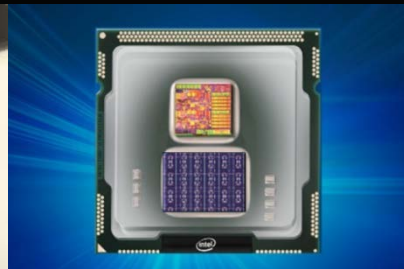


Intel. Компоненты ИИ

- Intel® Xeon® мощные процессоры. Новое поколение Lake Crest, Knights Crest с поддержкой нейронных сетей
- 8th Gen Intel® Core™ процессоры
- Intel® Nervana™ технологии нейронных сетей и обработки данных
- Intel Movidius™ - нейронные сети в кристалле.
- Intel FPGA – программируемые логические устройства

Intel. Новые разработки

- Loihi - first-of-its-kind self-learning neuromorphic chip
- Movidius Neural Compute Stick, Google Clips
- Intel® Nervana™ Neural Network Processor
- Intel® Xeon Phi™ с поддержкой нейронных сетей



Intel + Google. Google Clips. Стратегический союз.

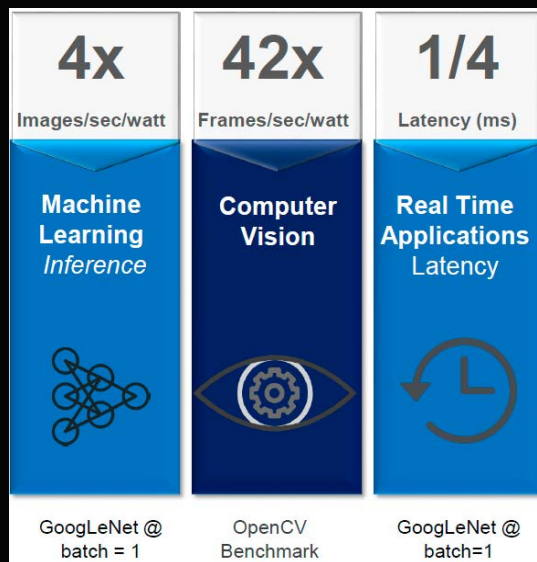




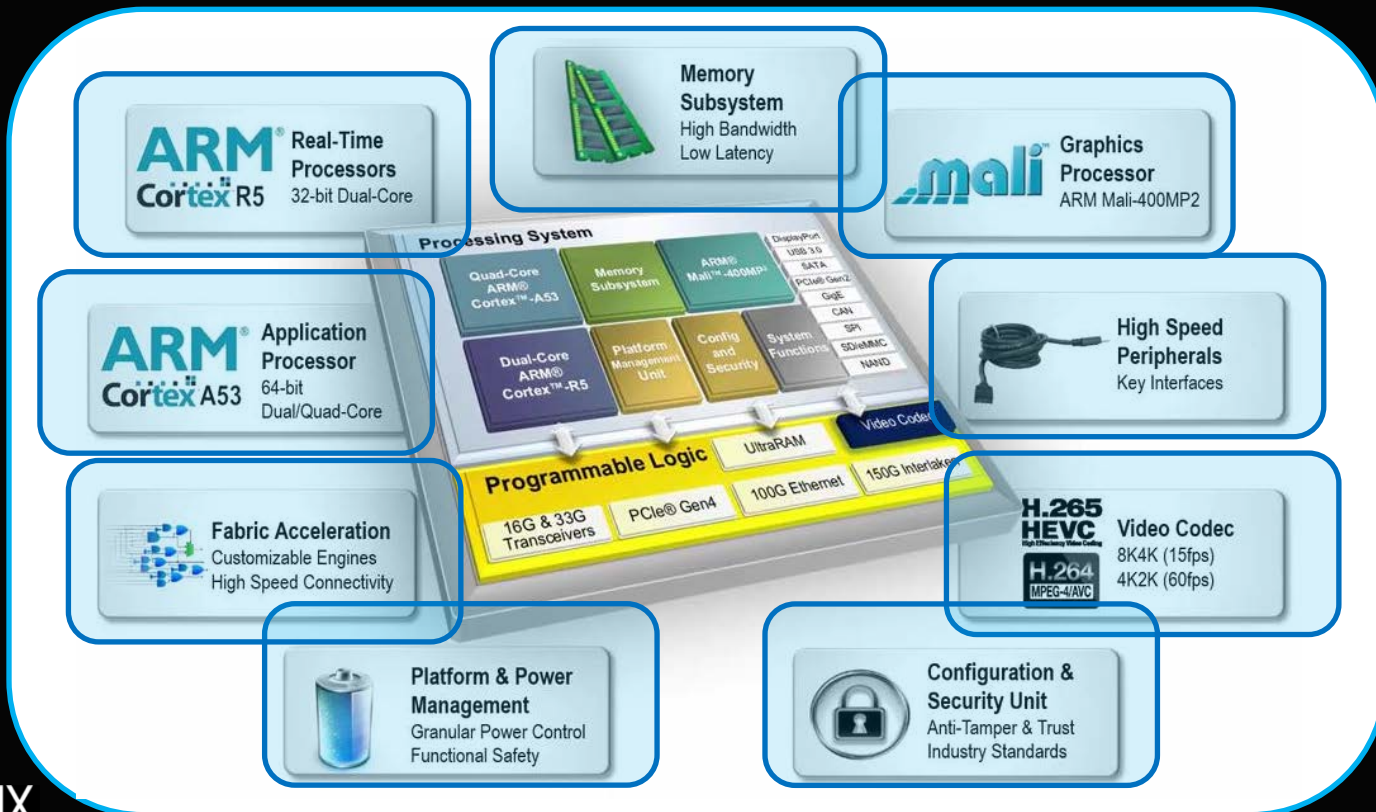
Преимущества FPGA

- Аппаратный специализированный процессор всегда быстрее программной эмуляции.
- Процессор может быть оптимизирован под различные нагрузки
- Конфигурацию программно определяемого процессора всегда можно легко обновить/изменить

Преимущества FPGA



Xilinx ZynQ основные свойства



Проблема внедрения

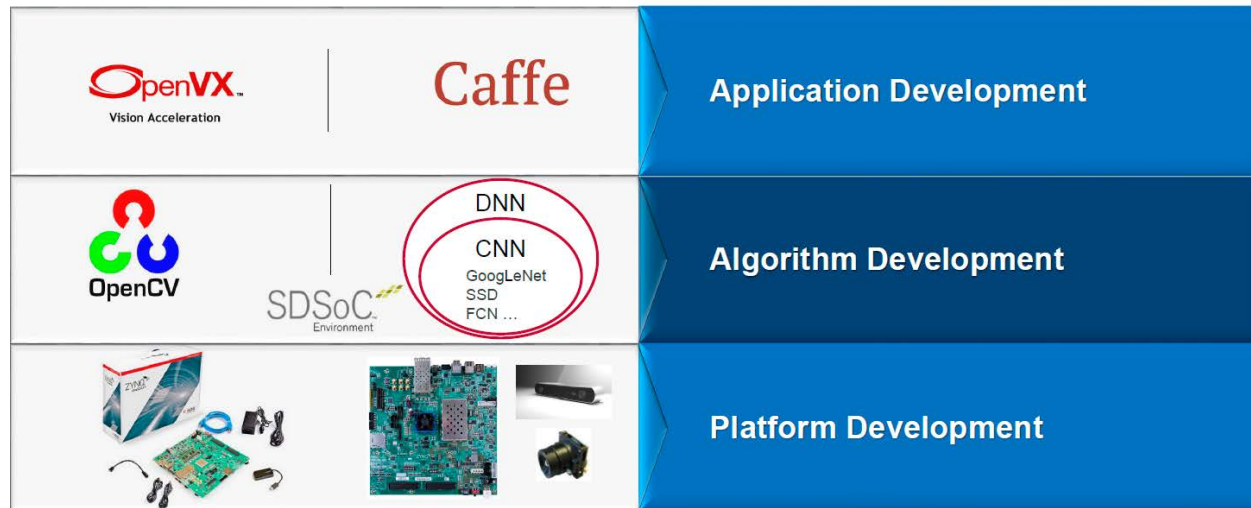
Сложность разработки программного обеспечения долгое время препятствовала массовому использованию FPGA

Решение?



reVision Stack

reVISION[™] Stack



Аппаратная оптимизация

КОМПИЛЯЦИЯ



```
main() {
    cv::imread(A);
    cv::stereoRectify(A,B,C,D);
    cv::stereoLBM(C,D,out);
    cv::imshow(out);
}
```

оценка

Time [%]	Time
	4504 ms (100 %)
	4500 ms (100 %)
	4370 ms (97,1 %)
	4370 ms (97,1 %)
	3474 ms (77,2 %)
	3473 ms (77,2 %)
	3471 ms (77,2 %)
	3412 ms (75,8 %)
	2597 ms (57,7 %)
	2274 ms (50,5 %)
	1933 ms (43 %)
	1895 ms (42,1 %)
	1782 ms (39,6 %)
	1593 ms (35,4 %)
	1567 ms (34,8 %)
	1444 ms (32,1 %)

ОПТИМИЗАЦИЯ

HW functions

Name	Clock Frequency (MHz)
stereoRectify	300
stereoLBM	300

```
main() {
    cv::imread(A);
    xF:stereoRectify<line>(A,B,C,D);
    xF:stereoLBM<win,n_disp>(C,D,out);
    cv::imshow(out);
}
```



результат



Результат 1



	Xilinx ZU9	Nvidia TX1
Frames/s	60	0.1
Power (W)	4.8	7.9
Latency (ms)	16.7	10000
Utilization	15%	100%

NVIDIA использует CUDA OpenCV

В расчеты не включены затраты на ввод и вывод изображений
Окно 53x53

4K60 Dense Optical Flow

ZCU102 Evaluation Platform (\$2,495)

MIPI камеры



По материалам компании Xilinx

Результат 2



	Xilinx ZU9	Nvidia TX1
Frames/s	140	28
Power (W)	4.8	7.9
Latency (ms)	7.1	35.7
Utilization	14%	100%

NVIDIA использует CUDA OpenCV

В расчеты не включены затраты на ввод и вывод изображений

Stereo Disparity map

ZCU102 Evaluation Platform (\$2,495)

USB3 камеры



По материалам компании Xilinx

NVIDIA



nVIDIA®

Технологии

- GPU сервера для облачных вычислений
- GPU акселераторы для обработки изображений
- GPU платформы для мобильных приложений

NVIDIA DGX-1

1 NVIDIA TESLA V100

The first GPU architecture to incorporate Tensor Core technology designed for AI.

2 NEXT GENERATION NVIDIA NVLINK

High-speed interconnect 300 GB/s per GPU, 10X faster than current PCIe Gen3 x16 interconnections.

3 TWO INTEL XEON CPUS

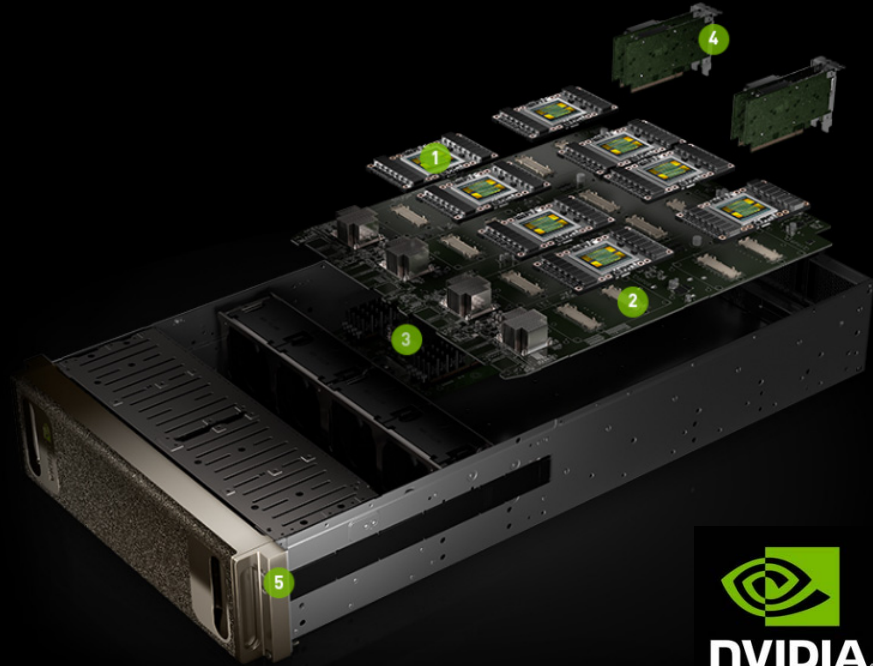
For boot, storage management, and deep learning framework coordination.

4 QUAD EDR IB

High-bandwidth and low-latency, with a total of 800 Gb/s of communication.

5 THREE RACK-UNIT ENCLOSURE

Fits in a compact rack space, using 3,200 Watts of maximum system power.



Встраиваемые ускорители



Беспилотные автомобили

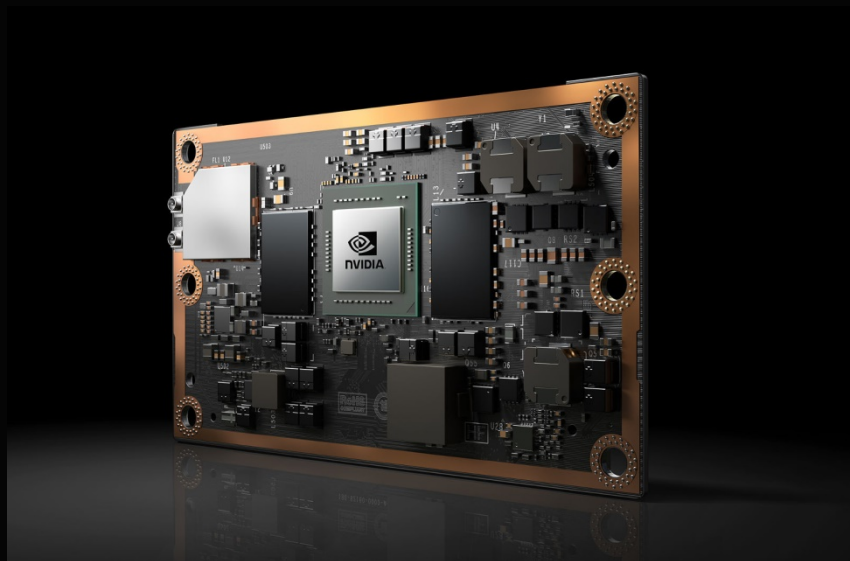


Встраиваемые вычислители с поддержкой NVIDIA GeForce 1050/1080



NVIDIA Jetson TX1/TX2

Встраиваемый компьютер для искусственного интеллекта



Jetson TX2

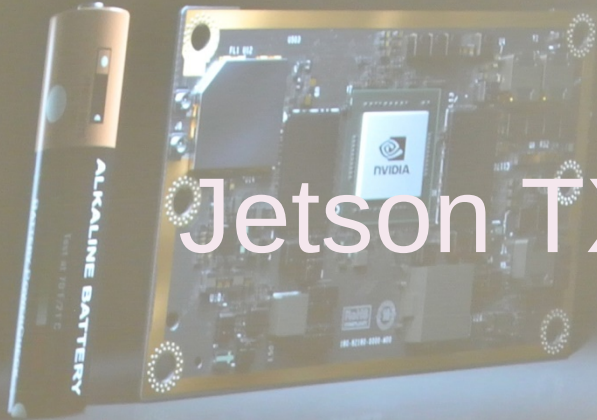
JETSON TX2 EMBEDDED AI SUPERCOMPUTER

Advanced AI at the edge

JetPack SDK

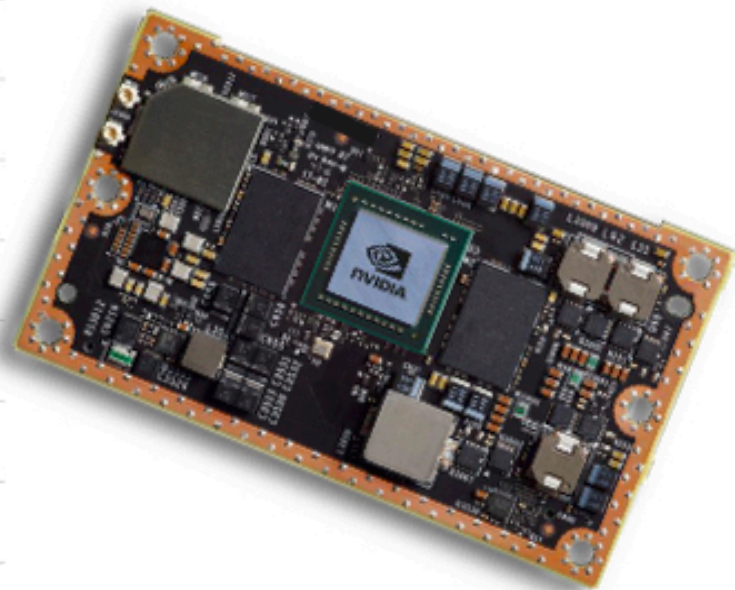
< 7.5 watts full module

Up to 2X performance or 2X energy efficiency



7.5 watts) delivers up to 2x energy efficiency vs. Jetson TX1 maximum per
module (< 15 watts) delivers up to 2x performance vs. Jetson TX1 maximum per

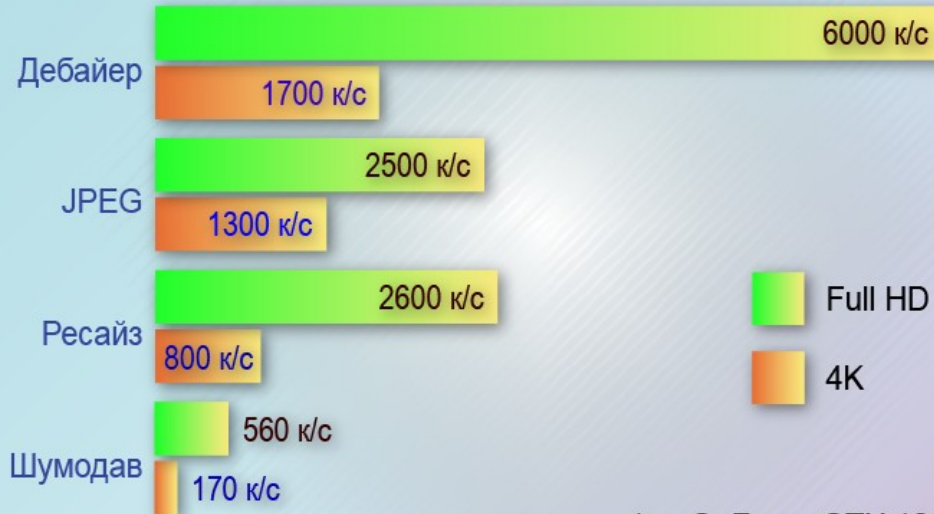
	JETSON TX1	JETSON TX2
GPU	Maxwell	Pascal
CPU	64-bit A57 CPUs	64-bit Denver 2 and A57 CPUs
Memory	4 GB 64 bit LPDDR4 25.6 GB/s	8 GB 128 bit LPDDR4 58.4 GB/s
Storage	16 GB eMMC	32 GB eMMC
Wi-Fi/BT	802.11 2x2 ac/BT Ready	802.11 2x2 ac/BT Ready
Video Encode	2160p @ 30	2160p @ 60
Video Decode	2160p @ 60	2160p @ 60 12 bit support for H.265, VP9
Camera	1.4Gpix/s Up to 1.5Gbps per lane	1.4Gpix/s Up to 2.5Gbps per lane
Mechanical	50mm x 87mm 400-pin Compatible Board to Board Connector	



GPU эффективность

Производительность Fastvideo SDK*

© Фаствидео 2016



*на GeForce GTX 1080 GPU

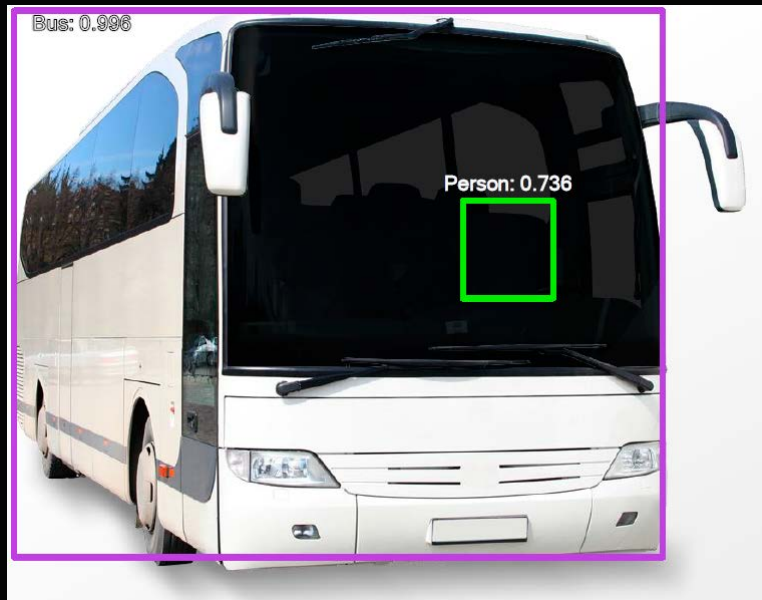
GPU: GeForce и Tegra X2

<i>Fastvideo SDK / NVIDIA GPU</i>	<i>GeForce 1080</i>	<i>Quadro P6000</i>	<i>Tegra X1</i>	<i>Tegra X2</i>
JPEG Encoder				
2K gray (8-bit, q=90%)	0.216 x5	0.11	1.5	1.2
2K (8-bit, q=90%, 4:2:0)	0.36	0.17	2.4	2.0
2K (8-bit, q=90%, 4:4:4)	0.40	0.21	3.7	3.1
4K gray (8-bit, q=90%)	0.55	0.35	6.3	5.0
4K (8-bit, q=90%, 4:2:0)	0.78	0.51	9.4	7.9
4K (8-bit, q=90%, 4:4:4)	1.12	0.74	14.9	12.5
4K gray (12-bit, q=90%)	0.83	0.54	11.2	8.5
4K (12-bit, q=90%, 4:2:0)	1.22	0.82	17.9	13.4
4K (12-bit, q=90%, 4:4:4)	1.90	1.32	28.6	22.4

Указано время в мс без операций ввода изображений
 Меньшее значение означает более высокую производительность

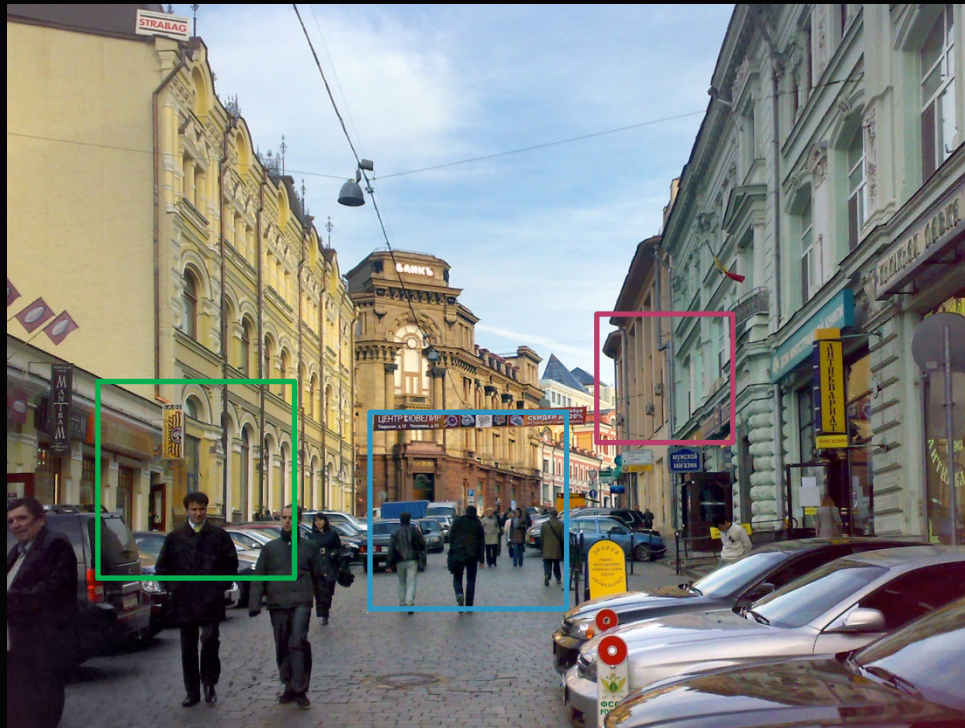
Информативность изображения

Нейронные сети распознают
объекты лучше человека



Замена механики

Захват большой площади без механизмов перемещения



Выводы 1

- Технические характеристики и возможности интеграции камер машинного зрения позволяют **повысить содержание информации** в исходном изображении
- Применение современных вычислительных средств — открывает **широкие возможности применения** этой информации как в системах обеспечения безопасности, так и в других областях: на транспорте, в спорте, в бизнесе

Выводы 2

- Мобильные приложения, Android – Qualcomm
- Обучение нейронных сетей - NVIDIA
- Реализация алгоритмов машинного зрения и нейронных сетей во встраиваемых системах: FPGA (Intel-Altera, Xilinx), Jetson TX2 или специализированные процессоры: Movidus и пр..

Ключевое значение имеет среда разработки ПО !



Основана в **1995** году в Петербурге

Поставка компонентов систем машинного зрения:

- Камеры
- Оптика
- Вычислители

Интеграция **промышленных систем**:

- Измерения и испытания
- Машинное зрение и визуальный контроль
- Робототехнические комплексы

WWW.VITEC.RU