

Решения Xilinx для умных городов: от Cloud до Edge



Алексей Попов

Alexey.Popov@avnet.eu



We Are Xilinx

Founded in 1984



\$3.16B

Revenue
(FY20)



~4,500

Employees Worldwide



60K+

Customers



60+

Industry Firsts



4,400+

Patents

Our Customers are Innovators in...



Aerospace &
Defense



ProAV &
Broadcast



Automotive



Consumer



Data Center



Industrial &
Vision



Healthcare &
Sciences



Test & Measurement,
and Emulation



Wired
Communications



Wireless
Communications

➤ Smart City Applications



Transportation

Public Health

Infrastructure Security

Access Control

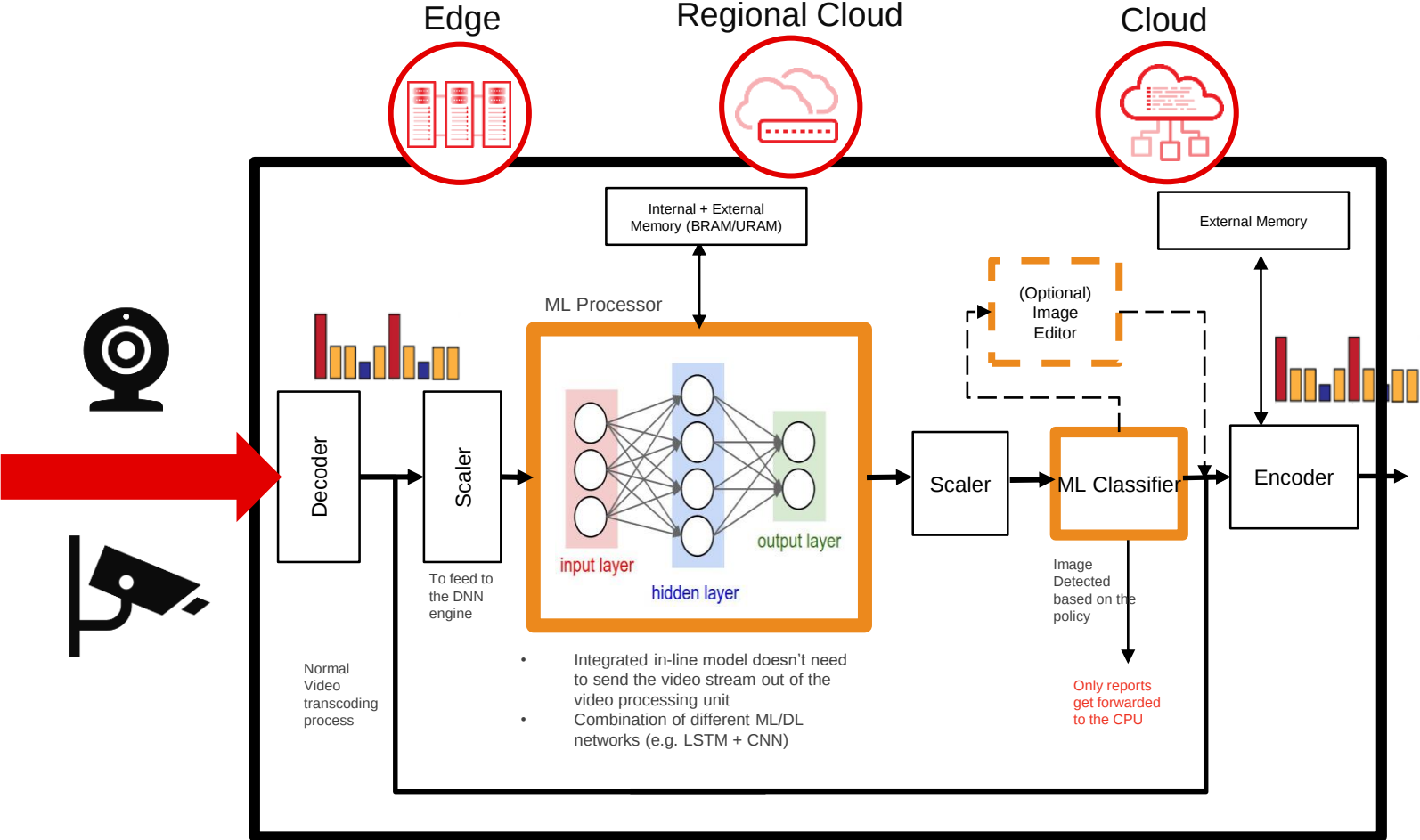
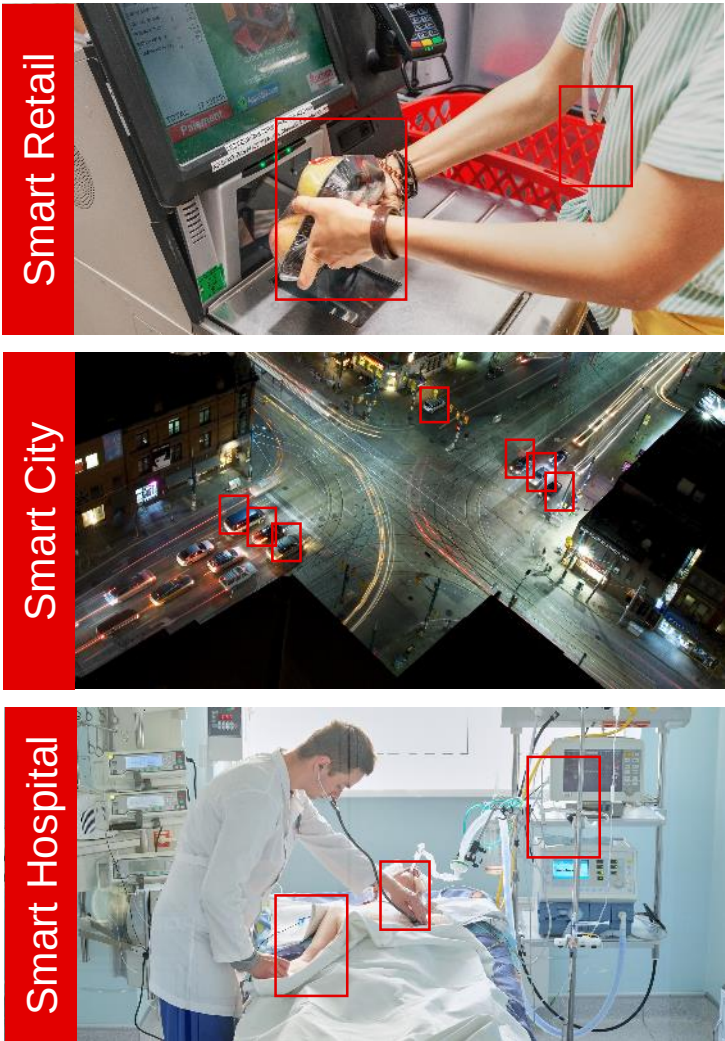
Smart Traffic

Crowd/Event
Detection

Public Safety

Smart Building

Xilinx Smart World Target Markets



Video Machine Learning Streaming Server (VMSS)

ALVEO Enables Real-time Computing



FAST

Built for high throughput, ultra-low latency
Real-time computing



ADAPTABLE

Deploy optimized domain-specific architectures
Adapt to changing algorithms



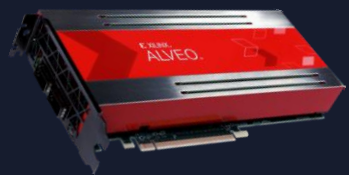
ACCESSIBLE

Deploy in the cloud or on-premises
Rich set of accelerated Applications



▶ Xilinx Alveo Product Lineup

ALVEO™ U200



UltraScale+ Architecture

1,182k LUTs

Dual slot, FH3/4L

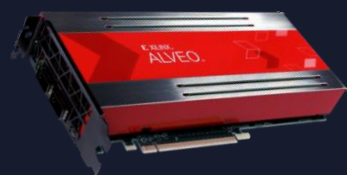
64GB DDR, 77GB/sec

PCIe Gen3

2x QSFP 28 (100GbE)

< 225W

ALVEO™ U250



UltraScale+ Architecture

1,728k LUTs

Dual slot, FH3/4L

64GB DDR, 77GB/sec

PCIe Gen3

2x QSFP 28 (100GbE)

< 225W

ALVEO™ U280



UltraScale+ Architecture

1,304k LUTs

Dual slot, FH3/4L

8GB HBM2, 460GB/sec

PCIe Gen3, Gen4, CCIX

2x QSFP 28 (100GbE)

< 225W

ALVEO™ U50



UltraScale+ Architecture

872k LUTs

Single slot, HHHL

8GB HBM2, 460GB/sec

PCIe Gen3, Gen4, CCIX

1x QSFP 28 (100GbE)

< 75W

ALVEO™ U25



UltraScale+ Architecture

520k LUTs

Single slot, HHHL

6GB DDR4

2x PCIe Gen3x8

2x SFP28 (2x10/25GbE)

< 75W

Vitis AI: Unified AI Inference Solution Stack

User Applications
and Demo Zoo



Frameworks



Caffe

Vitis AI Models

Model Zoo

Custom Models

Vitis AI
Development Kit

AI Optimizer

AI Quantizer

AI Compiler

AI Profiler

AI Library

- > Support both edge and cloud
- > Support AI model zoo
- > Several releases before this Vitis AI release

Vitis Runtime

Deep Learning
Processing Unit (DPU)

DPU

CNN-
Zynq

CNN-
Alveo

LSTM-
Alveo

CNN-AIE

LSTM-AIE

...

Xilinx Cloud Data Center Customers



Alibaba F3 series

High-performance acceleration card using
VU9P



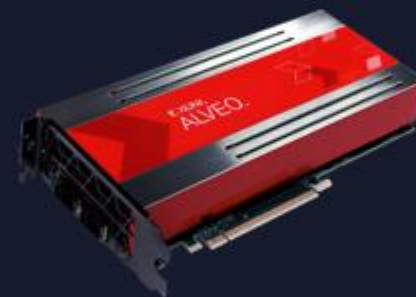
Fast AI, Deep learning, & IO for cloud



Amazon Sagemaker



Microsoft Azure FaaS



Powering Next Wave of HPC Services



Xilinx App Store

Pre-Built Applications for
Alveo Data Center Accelerator Cards
and Kria SOMs



Select Application by Product



Applications for Alveo
[Shop Alveo Apps >](#)



Applications for Kria
[Shop Kria Apps >](#)



Supported Workload

- ☐ Machine Learning 6
- ☐ Video and Image Processing 6
- ☐ Data Analytics 5
- ☐ High Performance Computing 3
- ☐ Networking and Security 2

On-Premises Solution

- ☐ Alveo U50 11
- ☐ Alveo U200 9
- ☐ Alveo U280 4
- ☐ Alveo U250 4
- ☐ Alveo U30 2

Cloud Solution

- ☐ AWS 5

Results per page 30 60 120 150

Results 1-22 of 22

Product ▾ Vendor ▾ **Card** Table

High Performance Computing

2D Reverse Time Migration (RTM)

Reverse Time Migration (RTM) is an important seismic imaging

[Get Started](#)



Data Analytics

Anti-money laundering application

The Anti-money laundering application is used to detect

[Get Started](#)



Machine Learning

Aupera Facial Recognition Solution

Aupera Facial Recognition Solution is end-to-end

[Get Started](#)



Data Analytics

DataBase Aggregation Demo

Database Aggregation demo is an application for hash

[Get Started](#)



Data Analytics

DataBase Join Demo

Database Join demo is an application for hash join

[Get Started](#)



Video and Image Processing

DeepField-SR

DeepField-SR is a fixed functional hardware accelerator

[Get Started](#)



Video and Image Processing

DeePoly ThunderImage

ThunderImage is DeePoly's leadership product for image

[Get Started](#)



Machine Learning

Edgecortex Dynamic Neural Accelerator DNA-

Dynamic Neural Accelerator® DNA-F100 is Edgecortex's first-

[Get Started](#)



Machine Learning

Edgecortex Dynamic Neural Accelerator DNA-

Dynamic Neural Accelerator® DNA-F200 is Edgecortex's

Data Analytics

Ethernet KVS by Axonerve

Key Value Store (KVS) is a valid technique in systems such as

Machine Learning

FPGA Accelerated Visual Search Engine

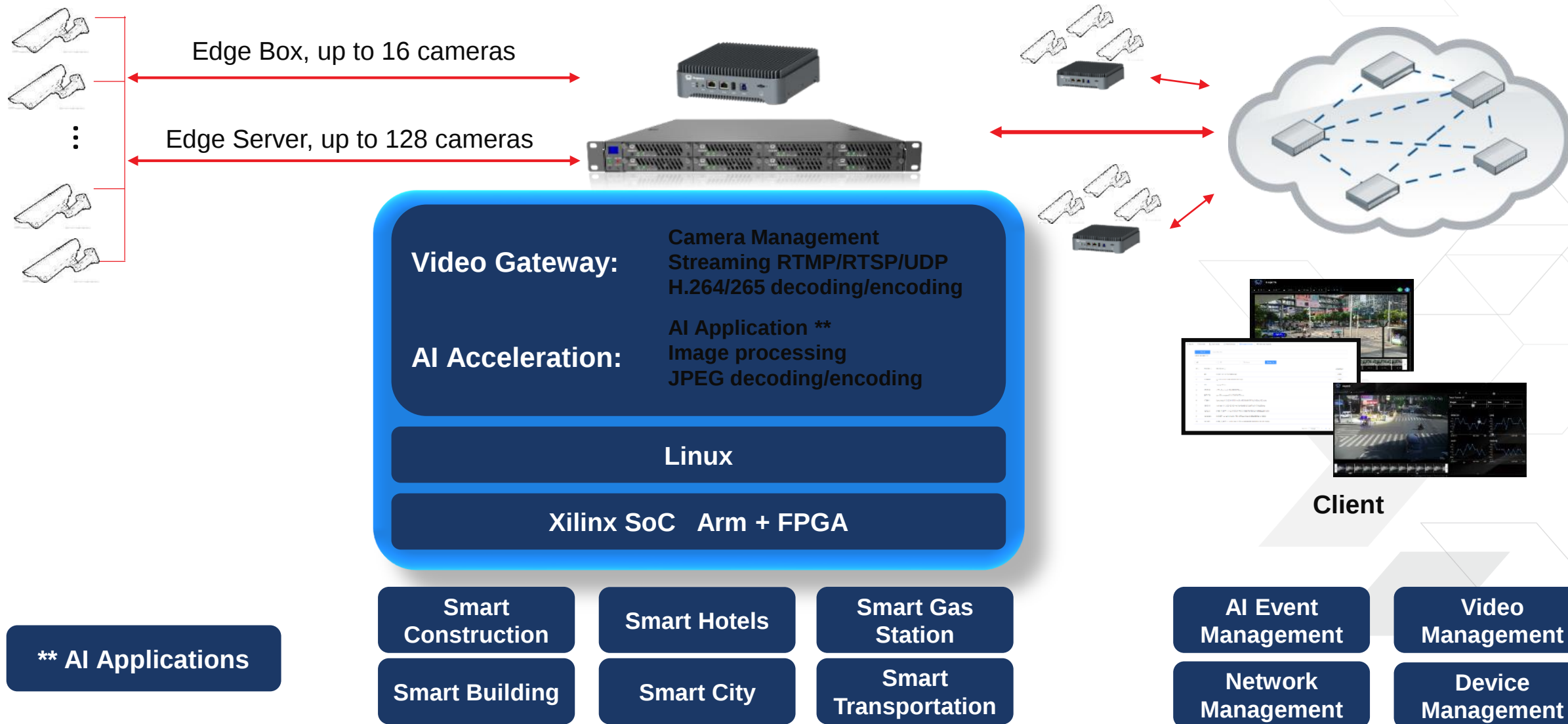
Moffett AI offers a breakthrough Visual Search as a Service

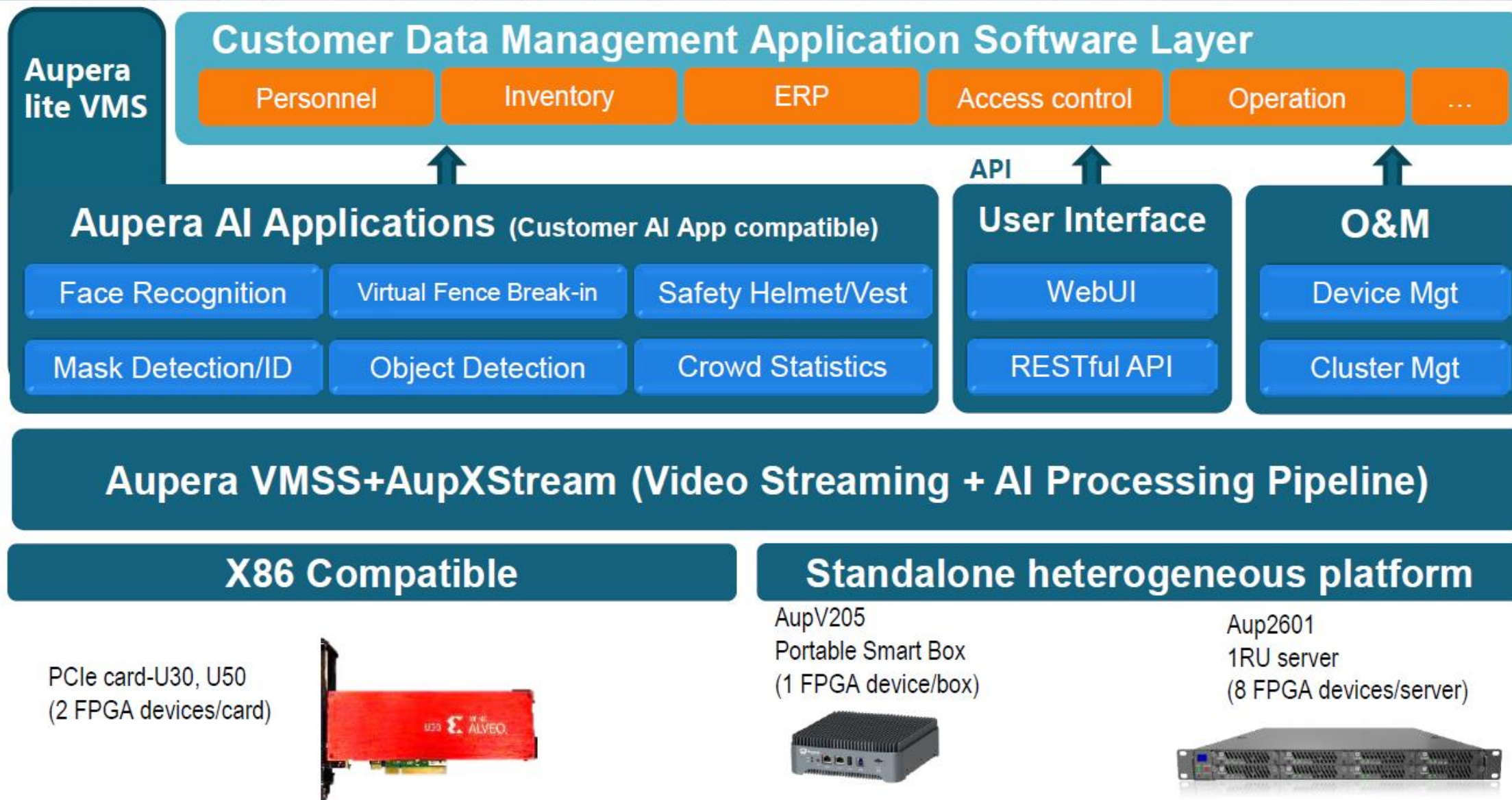
Video and Image Processing

HEVC/H.265 Encoder – Bitrate Optimized

NGCodec's RealityCodec is a distribution broadcast-quality

Aupera AI Platform

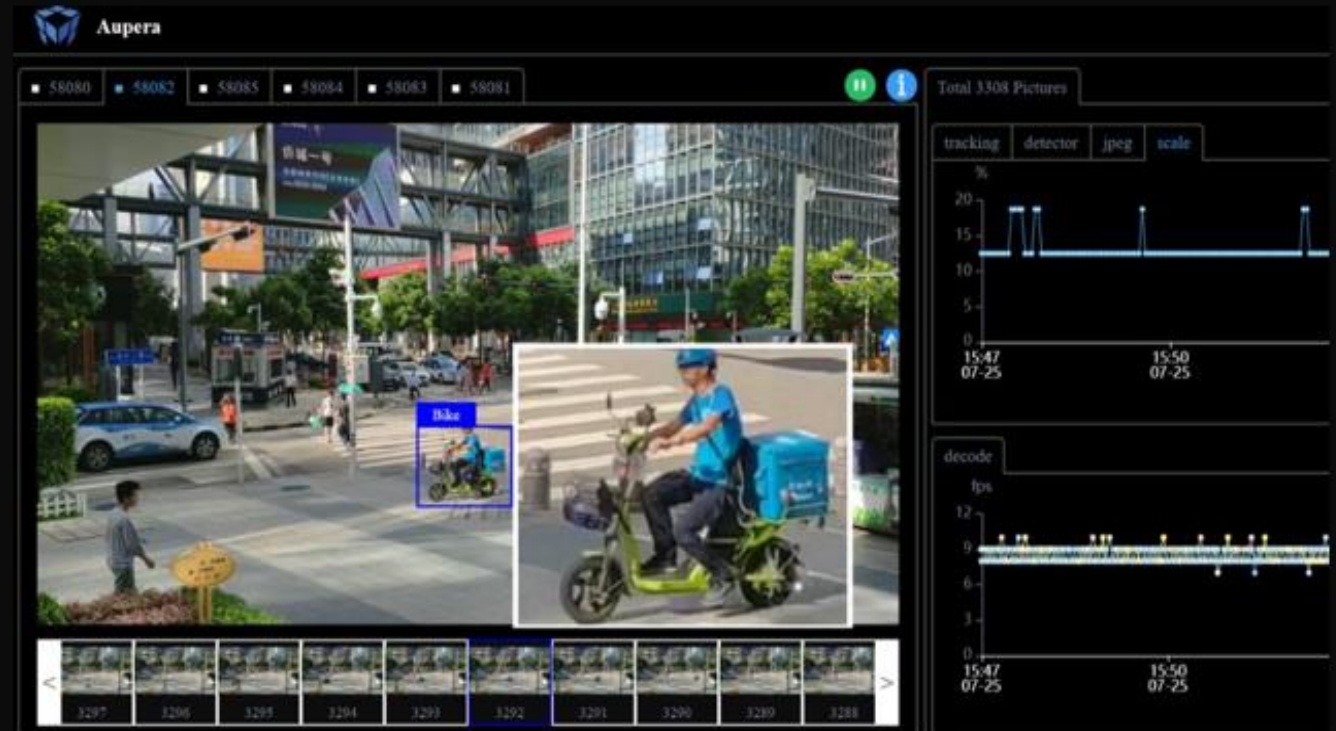




Xilinx MPSoC Based End to End Solution



- **Face Identification application**
- **Virtual fence break-in detection**
- **Dynamic gallery updates**
- **Algorithms field training**



- Pedestrian detection/tracking
- Car/Bike detection/tracking



- **VIP customers database management**
- **Car plate, car type recognition**
- **Gas pump occupancy**



- Safety helmet/vest detection
- Virtual fence break-in detection



- **Railway track obstacles detection**
- **Crowd statistics**
- **Train carriage human detection**

Mipsology Zebra

Neural Networks

- ▶ Supports CNN without modification
- ▶ Delivered with tested networks: AlexNet, CaffeNet, GoogLeNet V1, Inception V3, Inception V4, VGG16, VGG19, ResNet50, ResNet152, Nin, YoloV1, YoloV2, YoloV3, VDSR, SSD, MobileNet...
- ▶ Accelerated layers: convolution, fully connected, max/average pooling, concat, softmax, batch norm, scale, add eltwise, reorg, up sampling, depth to space, reduce mean, dilated convolution, squeeze, separable depth wise, clip to value, relu, leaky relu, relu6, sigmoid...
- ▶ Automatic split of graph
- ▶ Up to 3.2 billion weights
- ▶ Up to 1 million layers
- ▶ Illimited number of convolutions
- ▶ Single or multiple outputs
- ▶ Up to 1360x1360x3 input images
- ▶ Up to 24 simultaneous independent users

Zebra for Accelerating Neural Network Inference

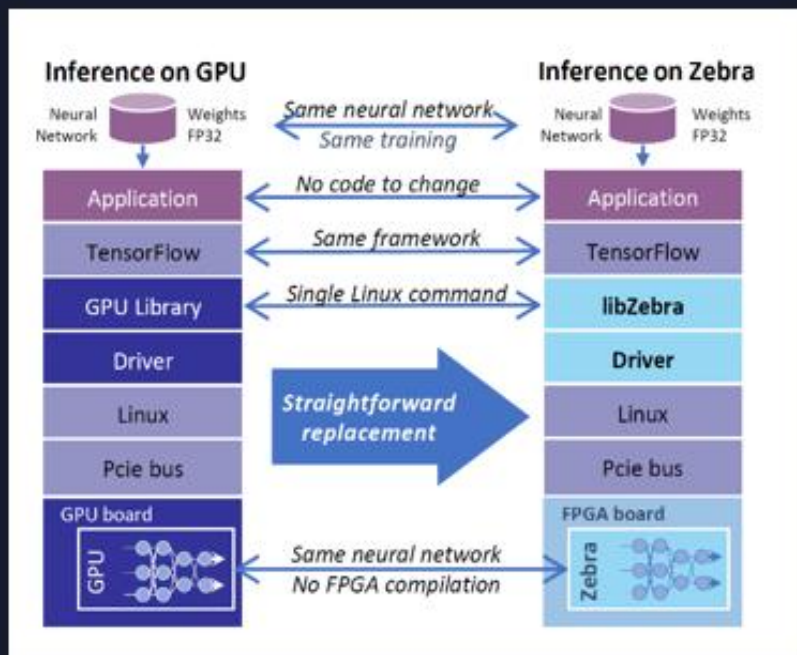
High Performance | Fast to Deploy | Easy to Use

- ✓ Supports many frameworks to compute any CNN
- ✓ Type a single Linux command to replace CPUs or GPUs for inference
- ✓ Delivers high performance in data centers, at the edge, under your desk

by Mipsology



ZEBRA BY MIPSOLGY



Migration from GPU or CPU

- ▶ Trained parameters from GPU/CPU training without changes
- ▶ No proprietary training or re-training needed and no pruning required
- ▶ Usable immediately
- ▶ Similar accuracy as FP32

Adaptable.
Intelligent.





Решения Xilinx для умных городов: от Cloud до Edge

Обзор ускоренных приложений для умных городов на платформе системы на модуле KRIA SoM K26

Alexander Vlasov

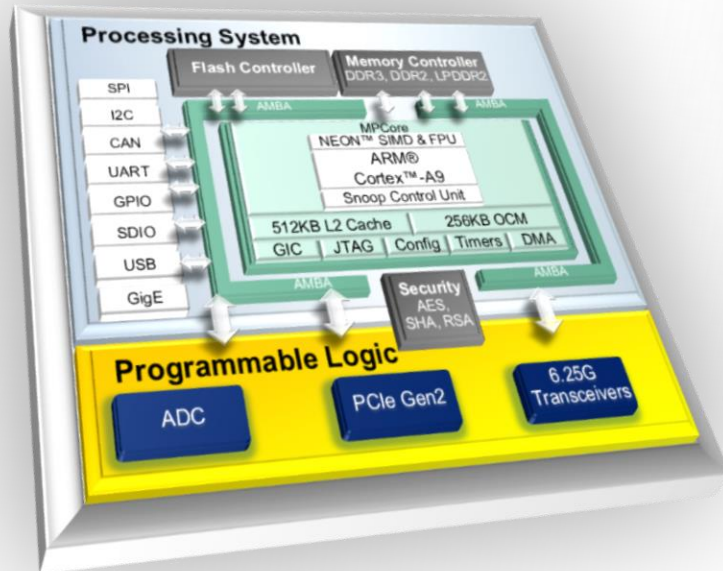
Alexander.Vlasov@avnet.eu

SoC / MP SoC Xilinx - микропроцессорные гетерогенные системы на кристалле

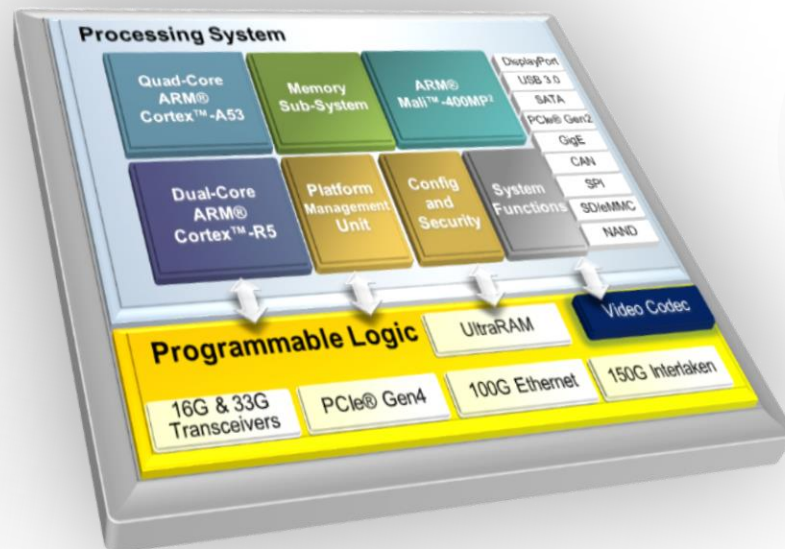
ZYNQ®

ZYNQ®
UltraSCALE+

XILINX
VERSAL™
AI Edge



2012

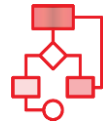


2016



2021

Versal AI Edge - обзор архитектуры



Scalar Engines

- Embedded Compute
- Safety-Critical Operation



PCIe Gen4 & CCIX

- 2X PCIe & DMA bandwidth
- Cache-coherency



DDR4 Memory

- 3200-DDR4, 4266-LPDDR4
- 2X bandwidth/pin



Transceiver Leadership

- Broad range, 1G → 32G
- Power-optimized



Adaptable Engines

- Re-architected for faster timing closure
- Tune for power or performance
- Adaptable for sensor fusion, any workload



Intelligent Engines

- AI Compute
- Diverse DSP Workloads



Accelerator RAM

- Reduces need for DDR
- Enhances AI Compute



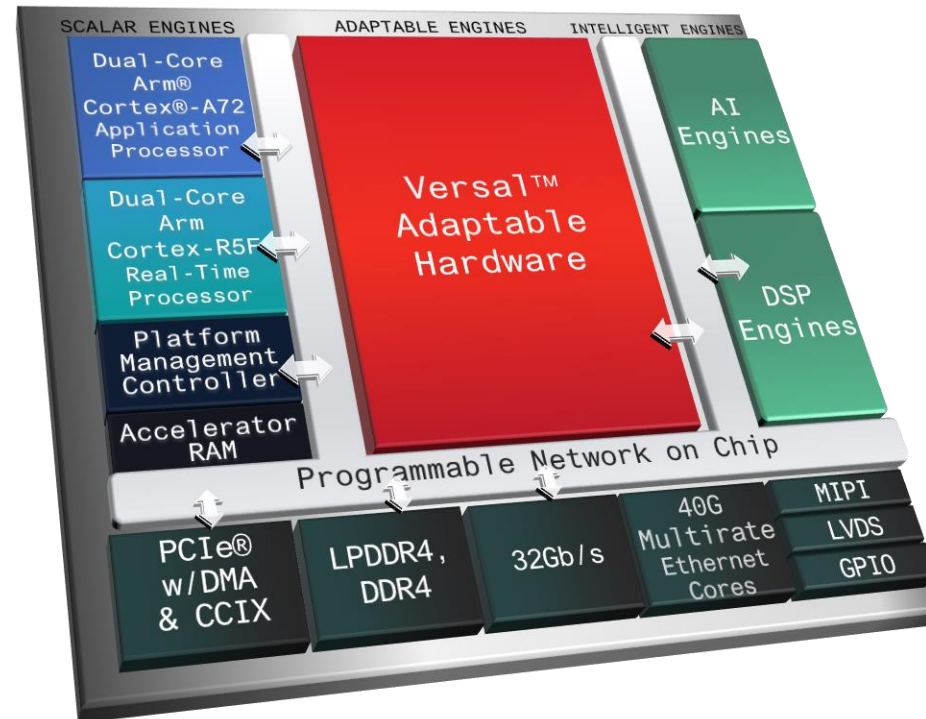
Programmable NoC

- Guaranteed Bandwidth
- Enables SW Programmability



Multi-rate Ethernet

- 10G/25G/40G
- Flexible, multi-standard



Programmable I/O

- Any interface or sensor
- Includes 3.2Gb/s MIPI

1: 4X logic resources per configurable logic block

Представление нового решения от Xilinx – системы на модуле Kria K26 SOM



Pre-Built HW Acceleration

- Application designs included for common AI functions
- ‘Out-of-the-box’ ready for SW developers



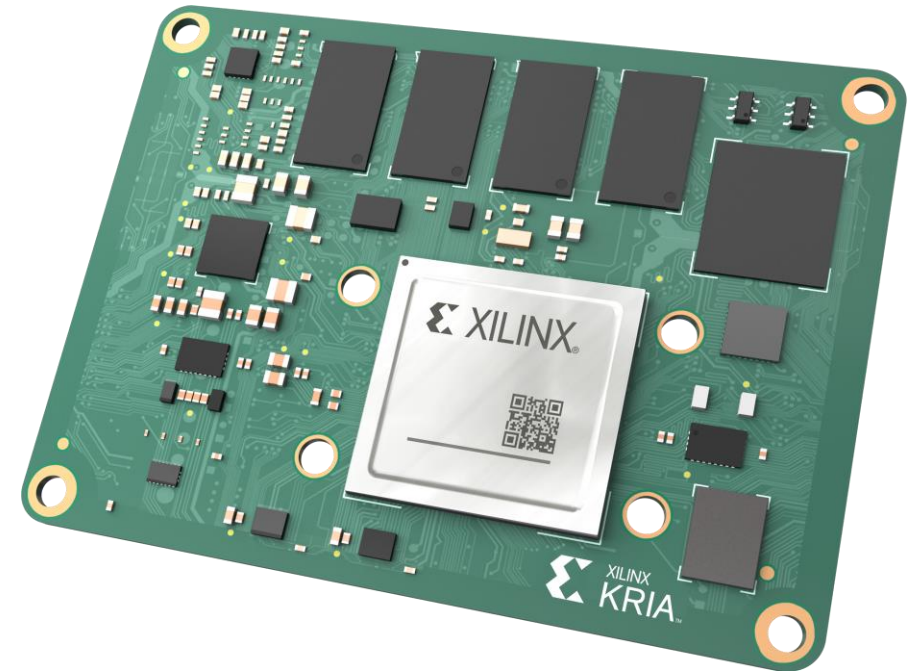
3X Performance¹

- AI at half the power with low latency
- Accelerates the whole application from AI to control



Future Proof

- Adaptable to changing AI and sensor requirements
- Ruggedized for industrial life cycles
- Path from concept to production to design revisions



Embedded Design Simplicity with System Flexibility

1: Xilinx Benchmarks vs. Jetson Nano published data (<https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-nano/>)

Xilinx SOM - Портфолио

Kria KV260 Vision AI Starter Kit

For Evaluation and Development Use



Разработка



Kria K26

Production Module
Fully Qualified and Certified



C-Grade

For Commercial Environments
Operating Temp 0°C to 85°C
2 Year Warranty

I-Grade

For Rugged Environments
Operating Temp -40°C to 100°C
3 Year Warranty

Использование

К26 SOM Обзор характеристик

Модуль построен на базе Zynq® UltraScale+™ MPSoC архитектуры



COMPUTE

Application Processor	64-bit Quad-Core Arm® Cortex®-A53
Real-Time Processor	32-bit Dual-Core Arm Cortex-R5F
Graphics Processor	Arm Mali™-400MP2
Programmable Logic	256K System Logic Cells
Deep Learning Processor	4K INT8 (upgradable to INT4)
Video Codec (H.264/H.265)	Up to 32 Streams (total resolution ≤ 4Kp60)
Memory	26.6Mb On-Chip SRAM
Security	IEC62443 Security w/HW Root-of-Trust

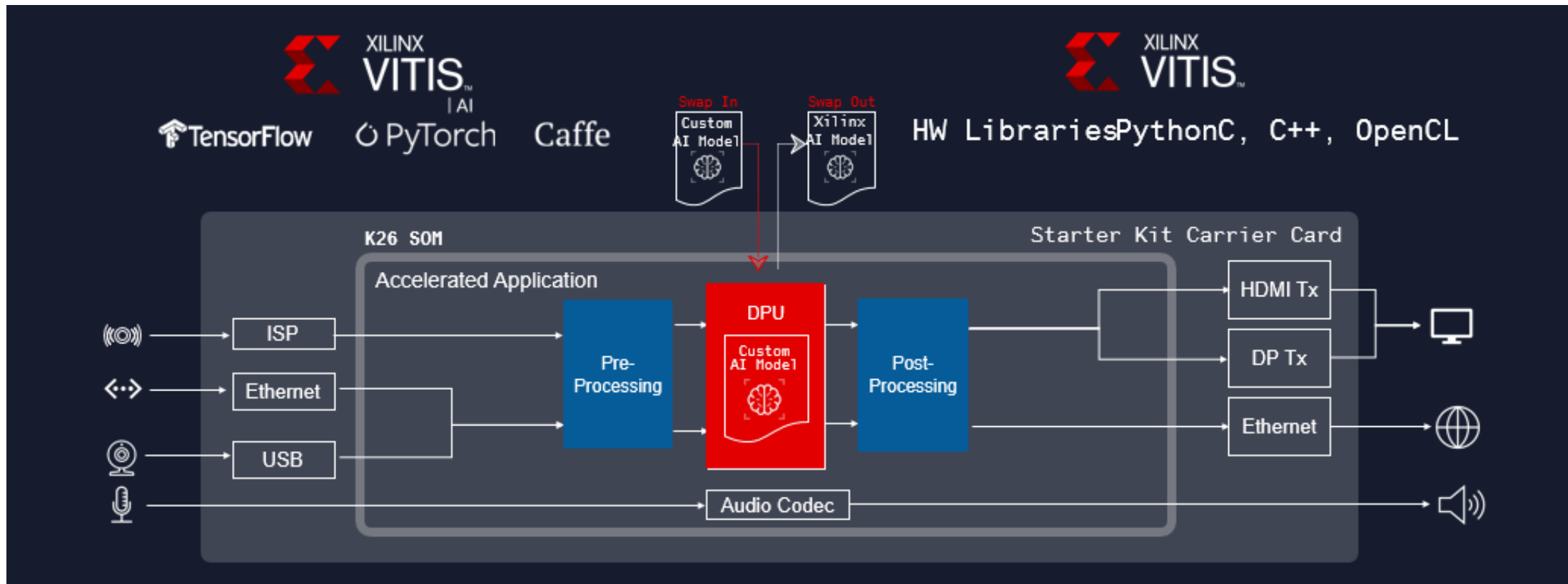
INTERFACES

Camera	11 x4 Full MIPI or sub-LVDS Interfaces 1 x4 SLVS-EC Interfaces
USB	4x USB 2.0 / 3.0
Multi-Media	DisplayPort, HDMI
Network	1Gb up to 40Gb Ethernet (w/GigE Vision)
Memory Interface	4GB 64-bit DDR4
Transceivers	4x 12.5Gb/s, 4x 6Gb/s
Mechanical	77 x 60 x 11mm w/ dual 240-pin connectors

1.4 TOPs

Старт с более высокого уровня абстракции с помощью ускоренных приложений (**Accelerated Applications**)

- ▶ Vitis™ AI development environment to swap AI models running on Deep Learning Processing Unit (DPU)
- ▶ Vitis core development kit to customize vision pipeline via HW-accelerated libraries, C, C++, OpenCL, Python

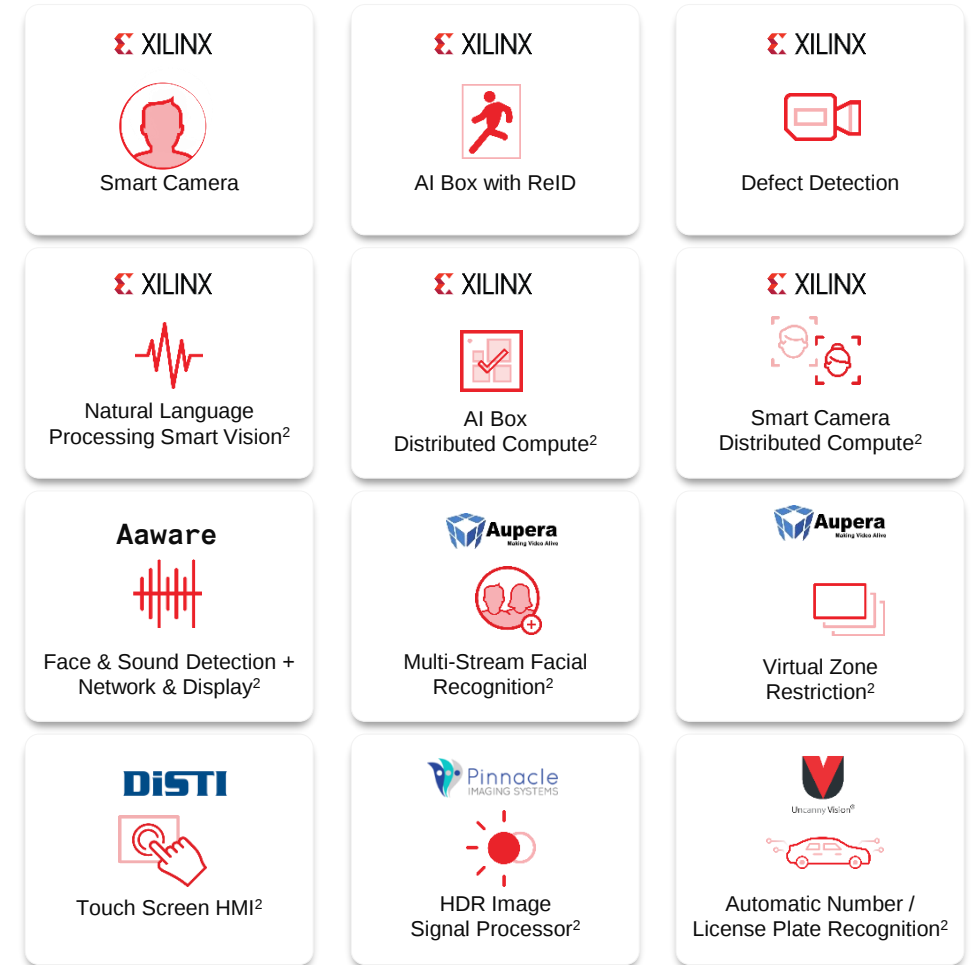


Широкий выбор готовых приложений для KRIA SOM, доступных на App Store

Vision accelerated applications
“do the work for you”

Production-ready applications now
available from both Xilinx and partners

App Store model encourages partners to
deliver high quality applications



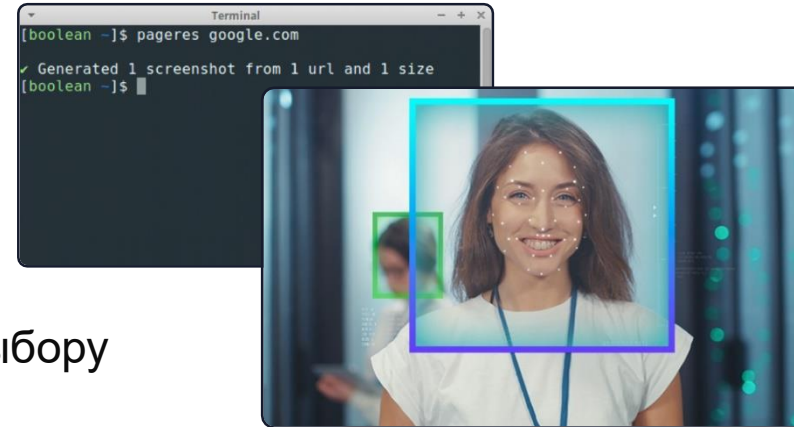
First Embedded App Store for Edge Applications

1: Supports face detection and other models in Xilinx Model Zoo

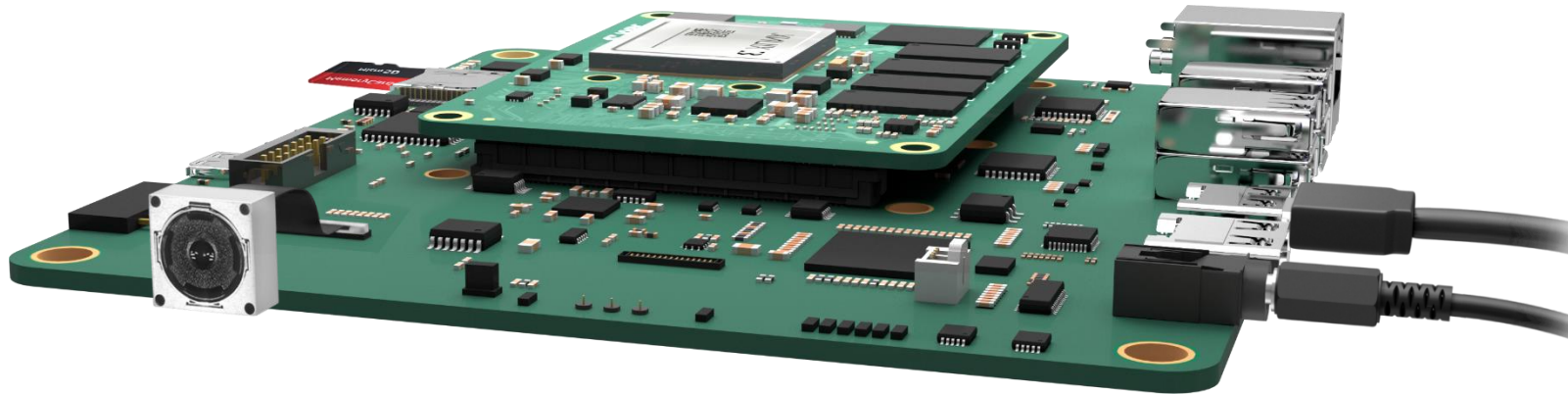
2: App coming soon

Готовый к разработке Стартовый Комплект + Периферийные устройства + Ускоренное Приложение

1. Подключите камеру, кабели и монитор
2. Вставьте запрограммированную карту microSD
3. Подайте питание на плату
4. Загрузите Ускоренная Приложение по вашему выбору
5. Запустите Ускоренное Приложение



GETTING STARTED WEB PAGE



Запуск менее чем за 1 час, опыт работы с ПЛИС не требуется

Примеры готовых решений “из коробки”, доступных на Xilinx App Store для конкретных применений



Smart Camera – Smart Security

Требования к приложению	Вывод (inference) ИИ по входному потоку и вывод результата через дисплей или Ethernet
Ценность предложения	Входное разрешение до 4K Адаптивные источники ввода и вывода. Развертывание пользовательской модели ИИ на DPU



Uncanny Vision – ANPR – Smart City

Требования к приложению	Автоматическое распознавание номерных знаков и символов
Ценность предложения	Облачная панель мониторинга и серверная часть Переобучение для конкретной страны с использованием соответствующего набора данных (dataset)

AI Box – Retail Analytics

Требования к приложению	AI Box + Возможность отслеживать и повторно идентифицировать людей или объекты
Ценность предложения	Поддерживается до 4 потоков. Обнаружение пешеходов + повторная идентификация на нескольких потоках



AI Box – Smart City

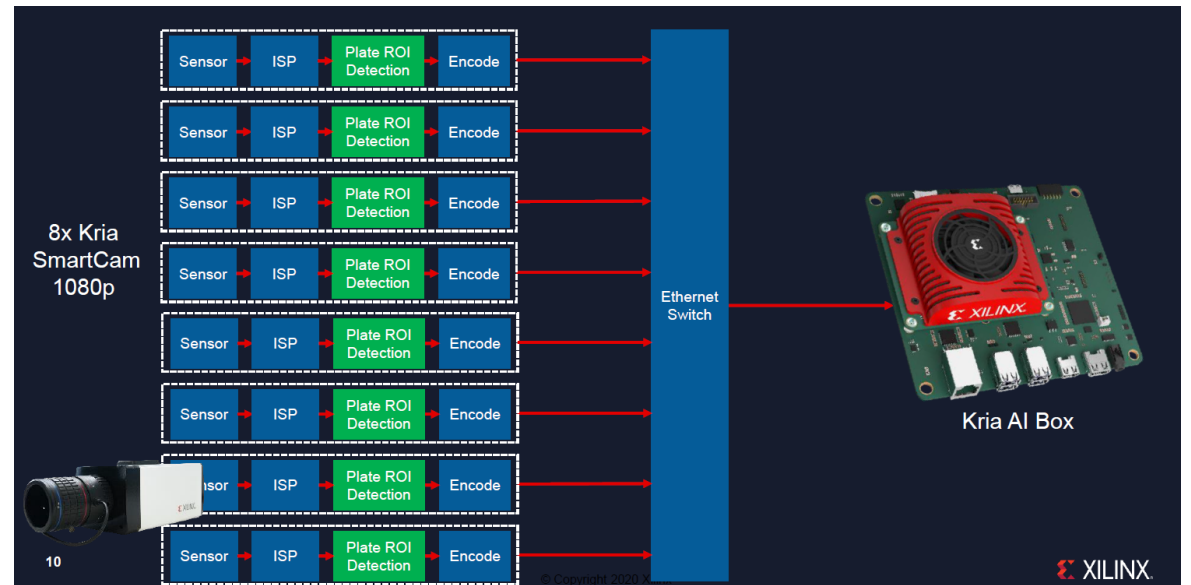
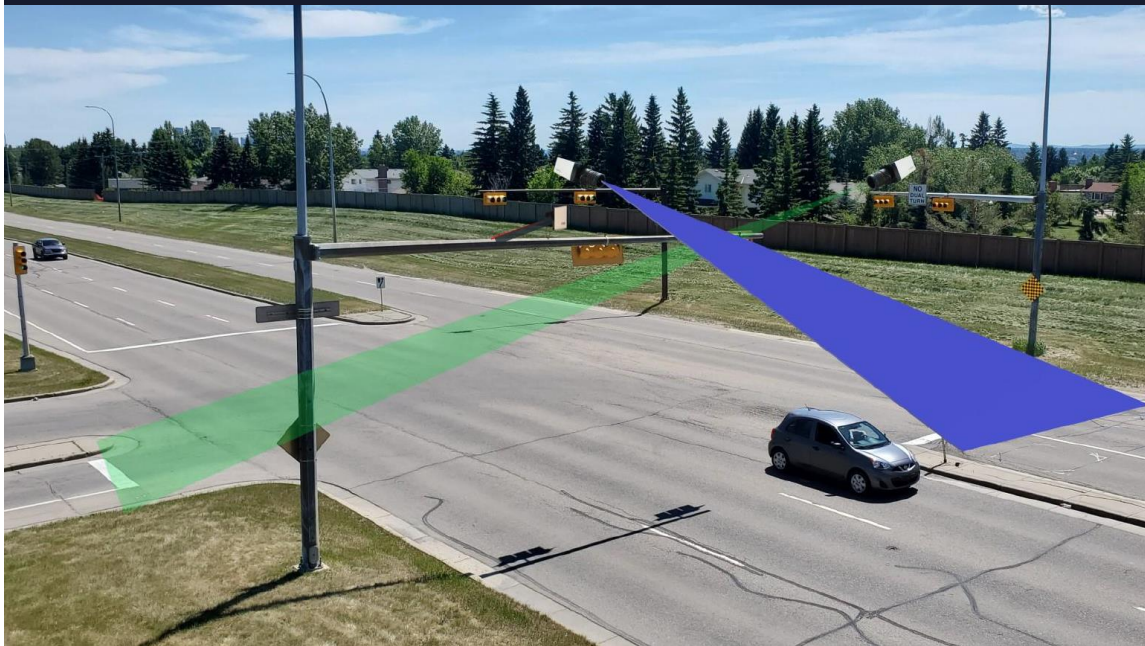
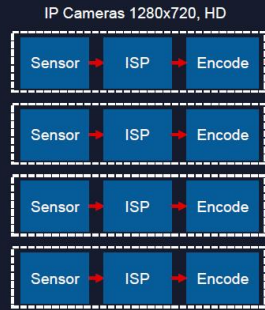
Требования к приложению	Вывод (inference) ИИ по входному потоку и вывод результата через дисплей или Ethernet
Ценность предложения	Поддерживается до 4 потоков. Обнаружение лиц или пешеходов на нескольких потоках

AI Box для автоматического распознавания номерных знаков (ALPR/ANPR). Ускоренное приложение от Uncanny Vision



- Видео конвейер с несколькими моделями искусственного интеллекта для обнаружения транспортных средств, номерных знаков, классификации транспортных средств и распознавания номерных знаков
- До 2 потоков декодирования H. 264/H. 265 с разрешением 1080p
- Комплексное решение LPR, включая информационные панели, удаленный мониторинг, черный и белый списки транспортных средств
- Открытый REST APIs для пользовательской интеграции
- Законченное приложение с включением аппаратного дизайна

AI Box ANPR



Распознавание номерных знаков (ANPR/LPR)

Высокая точность сквозного AI конвейера (точность 98+%)

- Обработывает широкий спектр сложных типов номерных знаков
- Сложенные символы, символы, временные таблички
- Поддерживает Stop-n-Go и Свободный поток, односторонний и двусторонний, несколько видов полос движения
- Открытый API для интеграции с любой сторонней системой
- Гарантия точности уровня обслуживания (SLA)



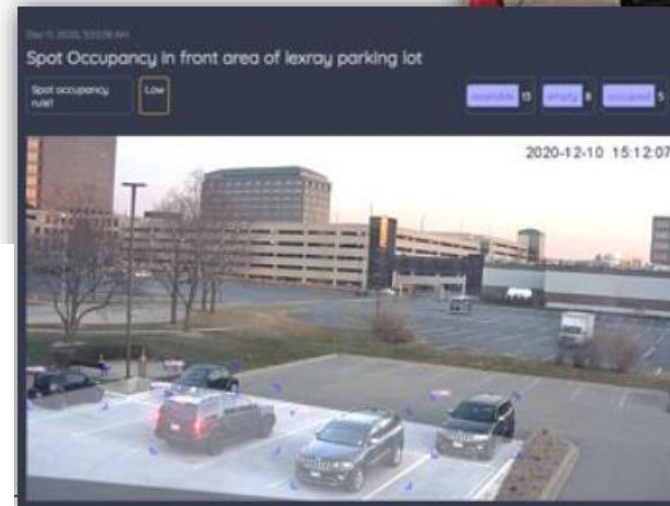
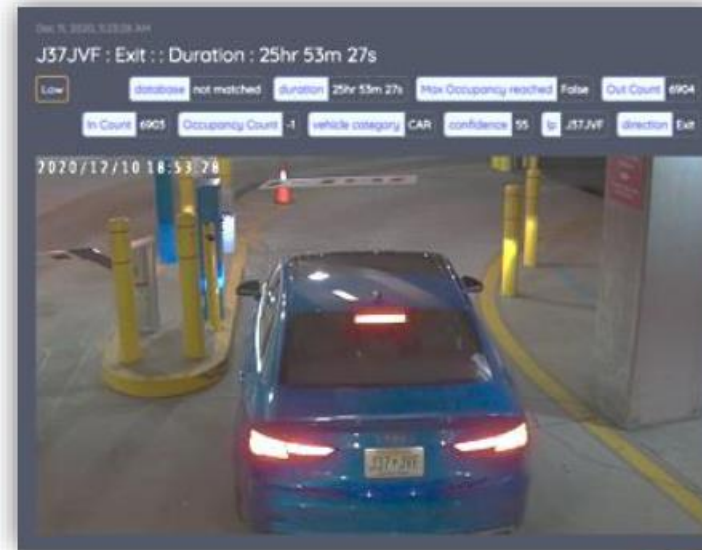
Умная Парковка & Умные ворота/шлагбаум, управление доступом

Распознавание номерных знаков у ворот

- Автоматизированный контроль доступа
- Контроль посетителей
- Контроль оплаты парковки

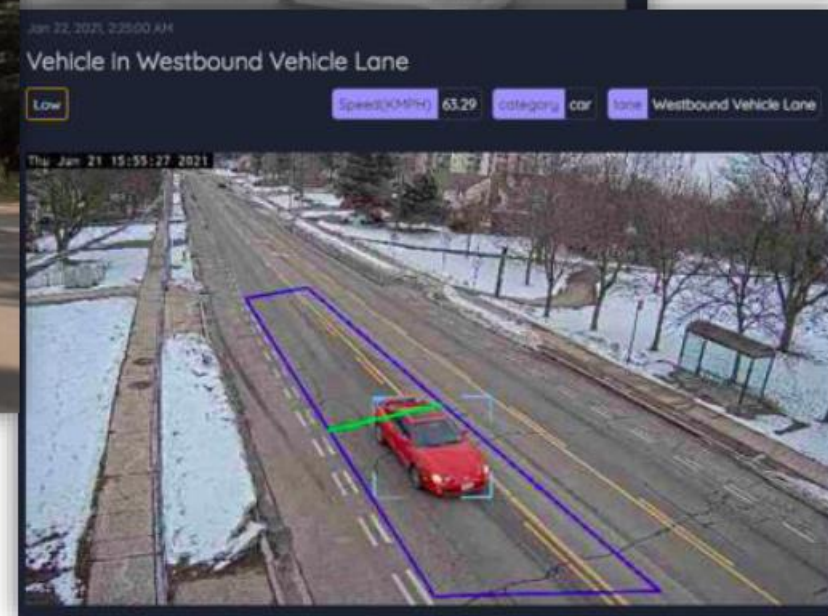
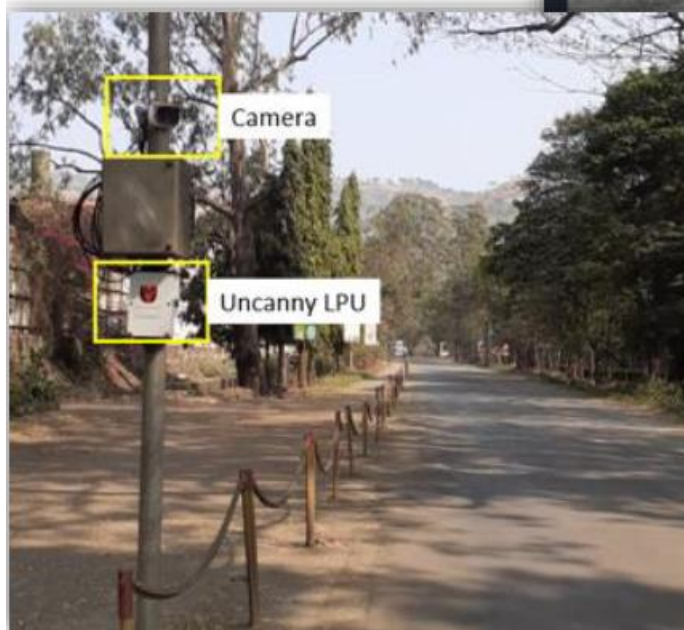
Заполняемость мест

- Автоматизация парковки автомобилей / грузовиков
- Оцифровка заполняемости
- Edge AI и облачная панель мониторинга

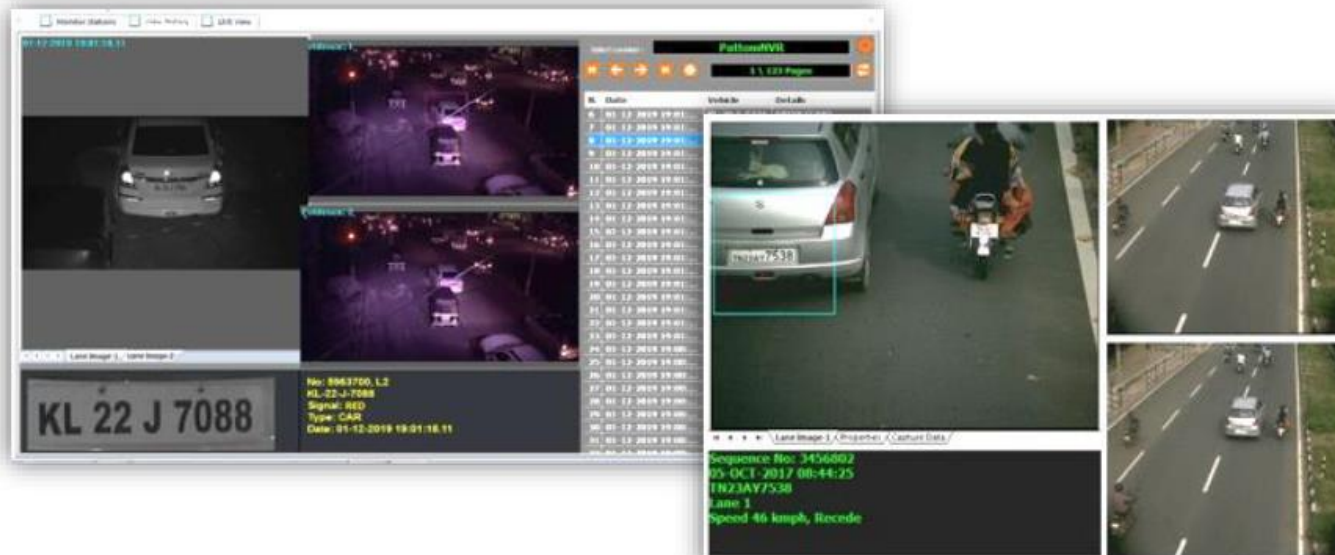
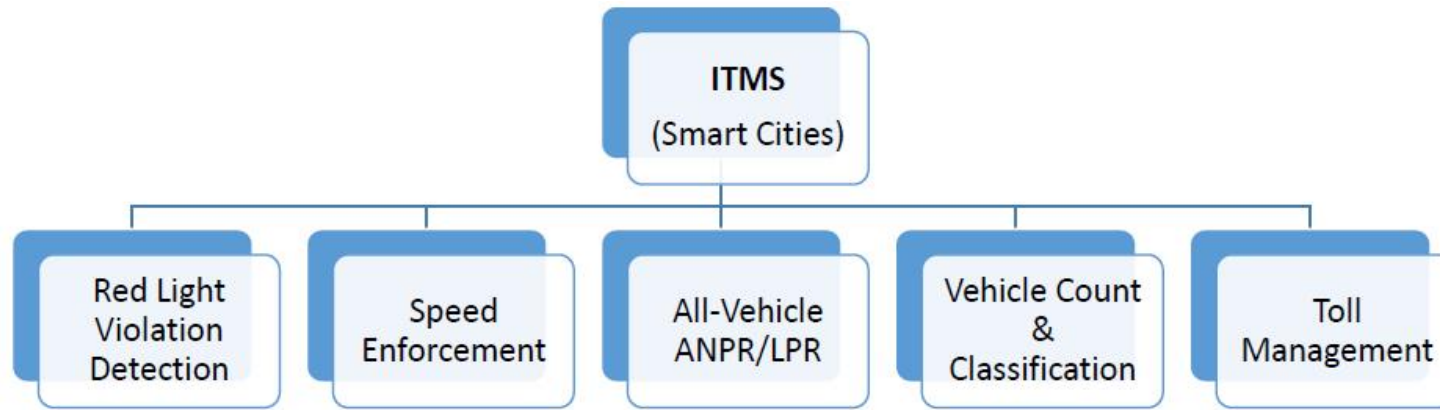


Контроль скорости с автоматическим распознаванием номерных знаков

- Контроль скорости без радара
- Недорогое краевое решение
- Классификация транспортных средств
- Интегрированная функция ANPR



Интеллектуальное решение для управления трафиком (ITMS)



Документ с таблицами сравнения производительности и KRIA SOM и конкурирующими решениями – WP529

Table 1: K26 SOM vs. Nvidia Jetson Features Comparison

No.	Feature	Xilinx's K26 SOM	Nvidia Jetson Nano ⁽¹⁾	Nvidia Jetson TX2 ⁽¹⁾
1	Application Processor	Quad-core Arm® Cortex®-A53 MPCore™ up to 1.5GHz	Quad-Core Arm Cortex-A57 MPCore processor	Dual-Core Nvidia Denver 2 64-Bit CPU and Quad-Core Arm Cortex-A57 MPCore processor
	Real-time Processors	Dual-core Arm Cortex-R5F MPCore up to 600MHz		
2	GPU	Mali™-400 MP2 up to 667MHz (primarily used for graphics rendering)	128-core Nvidia Maxwell GPU	256-core Nvidia Pascal GPU
3	Machine Learning Throughput	1.36TOPS ⁽²⁾	472GFLOPs (FP16)	1.33TFLOPs (FP16) 1.26TFLOPs for TX2i
4	Camera Interfaces	MIPI: Up to 44 DPHY2.0 lanes, Up to 11 Cameras, Max BW 10Gb/s BW per interface, Up to 16 virtual channels per interface	12 DPHY1.1 x 4 lanes Up to 4 cameras Max BW 6Gb/s per interface	12 DPHY 1.1 lanes Up to 6 cameras Max BW 6Gb/s per camera
		SLVS, LVDS: 11 x4 SLVS or LVDS cameras interfaces	NA	NA
		SLVS-EC: 4 lane, 5Gb/s /lane	NA	NA
5	Display Interface	DP1.2 x 2 lane	2x HDMI 2.0, DP 1.2, eDP 1.2, 2x MIPI DSI x2 lanes (1.5Gb/s /lane)	2x HDMI 2.0, DP 1.2, eDP 1.2, 2x MIPI DSI
		Additional HDMI 2.0 (GTs), DisplayPort 1.4 (GTs) with Soft IPs, MIPI DSI x4 lane (2.5Gb/s /lane) w/ Soft IPs	NA	NA
6	Video Encode H.264/H.265	Up to 32 simultaneous streams, Max resolution 4K @60FPS Color format: 422 8/10 bpc and 420 8/10bpc	Up to 9 streams; Max resolution 4K @30 FPS 420 8bpc	Up to 8 streams H.265, 14 streams of H.264, Max resolution 4K @60FPS 420 8bpc
7	Video Decode H.264/H.265	Up to 32 simultaneous streams, Max resolution 4K @60FPS Color format: 422 8/10bpc and 420 8/10bpc	Up to 9 streams; Max resolution 4K @60 FPS 420 8bpc	Up to 32 streams H.265, 16 streams of H.264, Max resolution 4K @60FPS 420 8bpc

Table 1: K26 SOM vs. Nvidia Jetson Features Comparison (Cont'd)

No.	Feature	Xilinx's K26 SOM	Nvidia Jetson Nano ⁽¹⁾	Nvidia Jetson TX2 ⁽¹⁾
8	Wireless	GTR M.2/SATA	M.2 Key-E site on carrier	802.11a/b/g/n/ac 2x2 867Mb/s Bluetooth 4.1 BCM4354 on module
9	Ethernet	4x 10/100/1000 Base-T Ethernet Additional Ethernet ports with Soft IP in HP I/Os	10/100/1000 Base-T Ethernet	10/100/1000 Base-T Ethernet
10	USB	2xUSB3.0, 2x USB2.0	4x USB 3.0 + Micro-USB 2.0	USB 3.0 + USB 2.0
11	PCIe®	PCIe Gen2 x 4 PCIe Gen3 x4 in with Soft IP on GTs	PCIe Gen2 x4 lanes	PCIe Gen2 x5 lanes
12	High-speed I/O (GTs) Provides Additional Interface Support	4x GTH transceivers in programmable logic can be configured to support a plethora of high-speed protocols such as SLVS-EC, PCIe Gen 3, HDMI, 10GE and many more	NA	NA
13	I/O Flexibility	69 3.3V I/Os, 116 1.8V I/Os allows users to create highly flexible and configurable I/O interfaces in programmable logic	NA	NA
14	Programmable Logic	256K System Logic Cells, 1248 DSPs, 26.6Mb on-chip memory allows users to implement custom accelerators for vision and ML functions	NA	NA
15	DRAM	4GB 64-bit DDR4	4GB 64-bit DDR4	8GB 128-bit LPDDR4?
16	eMMC	16GB	16GB	32GB
17	Flash	512MB QSPI	NA	NA
18	Socket Carrier Card Interface	Two 240-pin connectors	260-pin edge connector	400-pin board-to-board connector

Table 2: Deep Learning Models Performance Comparison

No.	Model	Image Size	Xilinx K26 B3136 DPU		Xilinx K26 B4096 DPU		Nvidia Jetson Nano		Nvidia Jetson TX2	
			FPS (Latency Optimized) ⁽¹⁾	FPS (Throughput Optimized) ⁽²⁾	FPS (Latency Optimized)	FPS (Throughput Optimized)	FPS (Latency Optimized)	FPS (Throughput Optimized)	FPS (Latency Optimized)	FPS (Throughput Optimized)
1	Inception V4	299x299	19	19.1	30.3	30.4	11	13	24	32
2	VGG-19	224x224	17.9	17.9	17.4	17.4	10	12	23	29
3	Tiny Yolo V3	416x416	88.2	92.6	148.0	161.3	48	49	107	112
4	ResNet-50	224x224	49	49.1	75.6	75.9	37	47	84	112
5	SSD Mobilenet-V1	300x300	129.6	133.4	192.1	200.4	43	48	92	109
6	SSD ResNet34	1200x1200	1.6	1.6	2.5	2.5	1	1	3	2

https://www.xilinx.com/support/documentation/white_papers/wp529-som-benchmarks.pdf

Основные ссылки на материалы для Kria SOM

Основная страница на сайте Xilinx с основной информацией по Kria K26 SoM модулю и соответствующему стартовому комплекту – <https://www.xilinx.com/products/som/kria.html>

Ссылка на страницу, посвящённую модулю Kria K26 SoM на сайте Avnet: <https://www.avnet.com/wps/portal/silica/products/product-highlights/2021/xilinx-systemonmodules/>

Общая презентация System-on-Modules Xilinx: <https://www.xilinx.com/publications/ebooks/introducing-adaptive-system-on-modules.pdf>

Брошюры с общим описанием Kria K26 SoM: <https://www.xilinx.com/publications/product-briefs/xilinx-k26-product-brief.pdf>
https://www.xilinx.com/support/documentation/white_papers/wp528-simplified-som.pdf

Datasheet на модуль Kria K26: https://www.xilinx.com/support/documentation/data_sheets/ds987-k26-som.pdf

Основные показатели и таблицы сравнения Kria K26 SoM с модулями Nvidia приведены в этом документе: https://www.xilinx.com/support/documentation/white_papers/wp529-som-benchmarks.pdf

Брошюра с общим описанием Kria KV260 Vision AI Starter Kit: <https://www.xilinx.com/publications/product-briefs/xilinx-kv260-product-brief.pdf>

Datasheet на Kria KV260 Vision AI Starter Kit: https://www.xilinx.com/support/documentation/data_sheets/ds986-kv260-starter-kit.pdf

User Guide Kria KV260 Vision AI Starter Kit: https://www.xilinx.com/support/documentation/user_guides/som/1_0/ug1089-kv260-starter-kit.pdf

WP528 - Achieving Embedded Design Simplicity with Kria SOMs https://www.xilinx.com/support/documentation/white_papers/wp528-simplified-som.pdf

Ссылки на соответствующие видеоматериалы на сайте Xilinx:
[New Video: Getting Started with the KV260 Starter Kit](#)

Evaluate Uncanny ANPR: <https://www.uncannyvision.com/partner/xilinx> <https://www.uncannyvision.com/products/uncanny-anpr>
<https://www.xilinx.com/products/app-store/kria/auto-license-number-recognition.html>



Thank You

