Deep Learning for Embedded Vision System

Hai Tao, Dr.

Credits to all my colleagures who make this presntation possible

Jan. 11th, 2017

Vion Technologies Co., Ltd.

Vion Technologies: A Leader in the Field of Computer Vision

Founded in 2005

2016 NEEQ/ Series B Funding

2014.12 Series A Funding

- Vion Technologies Co. Ltd., founded in 2005, currently employs 200+ talented staffs. The company is developing CV HW/SW total solutions for intelligent transportation systems (ITS), smart video surveillance systems and business intelligence systems.
- Huge potential for CV products in ToB markets
  - Every year more than 40 million surveillance cameras are sold globally (IDC data analysis)
  - High resolution (720p, 1080p, even 4K resolution) IP cameras are replacing the D1 resolution analog cameras
  - Better algorithms enable more applications in ToB applications
  - High performance, low power consumption, low cost processors are available

- **IOT+Computer Vision, Where Are the Applications ?**
- Embedded CV Hardware
- GPU, VPU, and FPGA

Industry Applications

Smart Traffic

大梁路与西环路东_2015年01月09日09时34分17秒597毫秒_东向西_车道1_左拐闯红灯

Intersection violation capture & smart plate number recognition & light control

Transit & emergency vehicle lane use capture

Smart parking management

Parking Violation Capture

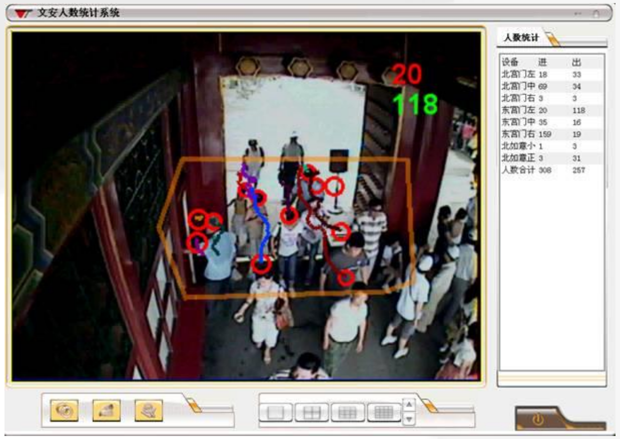
Industry Applications

VionVision

Smart People Counting

Malls

Retail Stores



设备	进	出
北宫门左	18	33
北宫门中	69	34
北宫门右	3	3
东宫门左	20	118
东宫门中	35	16
东宫门右	159	19
北加意小	1	3
北加意正	3	31
人数合计	308	257

Cultural Attraction Guest Traffic

Transit People Counting

Subway People Counting

Theatre People Counting

Industry Applications

Public Security, City Management, Banking, Rail, Border Control and Many More ...

Security & Counterterrorism: Fighting

Security & Counterterrorism: Chasing

Banking: ATM Protection

Rail: Driver Fatigue

Mining: Production Safety

Intruder Alerts

- IOT+Computer Vision, Where Are the Applications ?
- **Embedded CV Hardware**
- GPU, VPU, and FPGA

Smart Cameras

Sensor Rich ITS Camera

- Sensor rich (multi-axis/temp)
- 3/6/8MP 25fps
- High performance platform
- 3G/4G/WIFI
- Smart traffic industry

Smart Traffic Camera

- Integrated image sensing and analysis
- Wifi probe & iBeacon
- POE powered
- Patented exterior design, screw free installation
- H.264 real-time video output
- 2-year data storage

Bus People Counting

- Integrated image sensing and analysis
- RS485, GPIO
- Patented exterior design specially for transit
- H.264real-time video output
- 2-year data storage
- IP65, sealed against dust & water

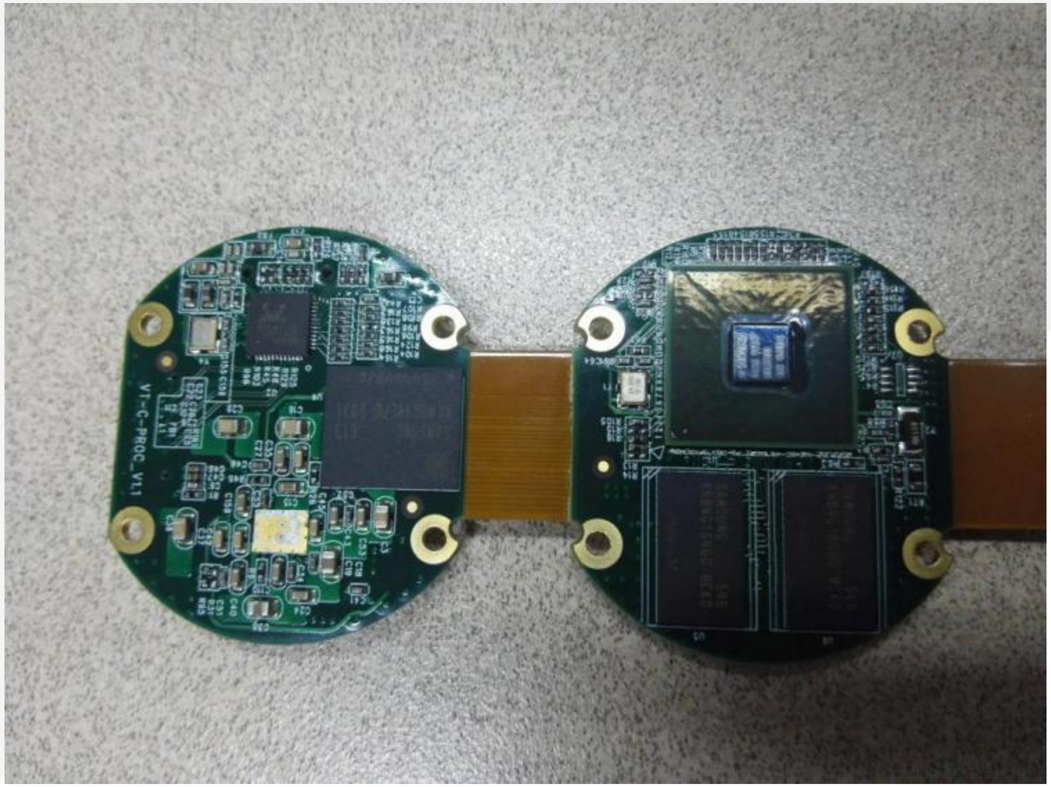
Smart Cameras - 4K ePolice Camera

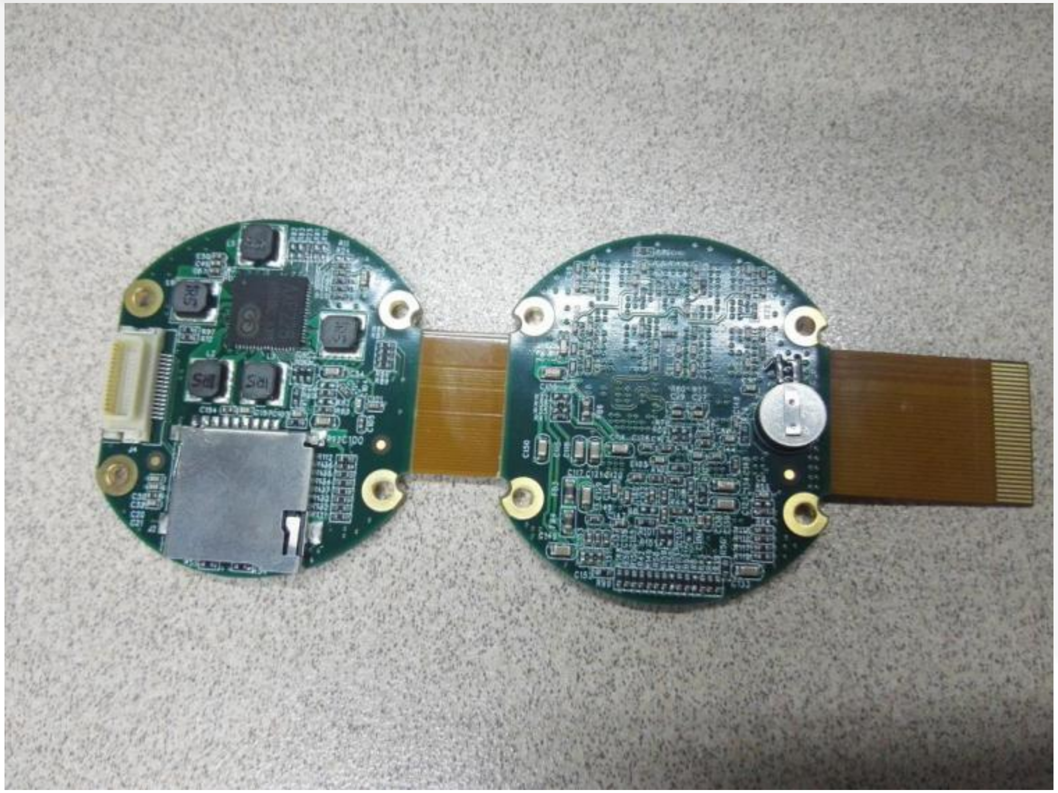
Spec: 4K resolution 4/3' CCD, Ambarella processor, Xilinx FPGA module

Applications: ePolice at road intersections, covering 4 lanes. The first 4K@25fps ePolice in the world

Release data: 2016 Q3

Smart Cameras - People Counting





Spec: ARM processor, compact form format

Applications: People counting for Shopping malls and retail stores.

Release Date: 2016 Q3

Front-End AI Module - Tarsier I

VionVision

Tarsier I Module - A Step to Smart Edge Device

Low Power

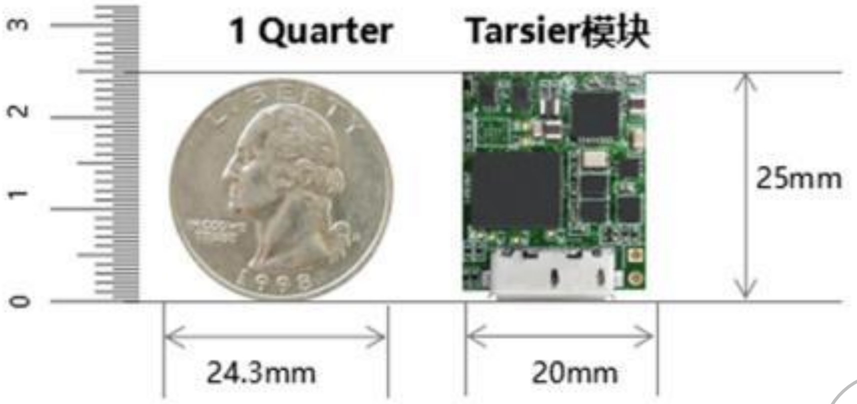
<1.5W

Multi-Modal

Video & Audio

Interface

camera, other processors



High Performance

Deep Learning (CNN , Recurrent DNN) > 40GFLOPS

Low Cost

<$15

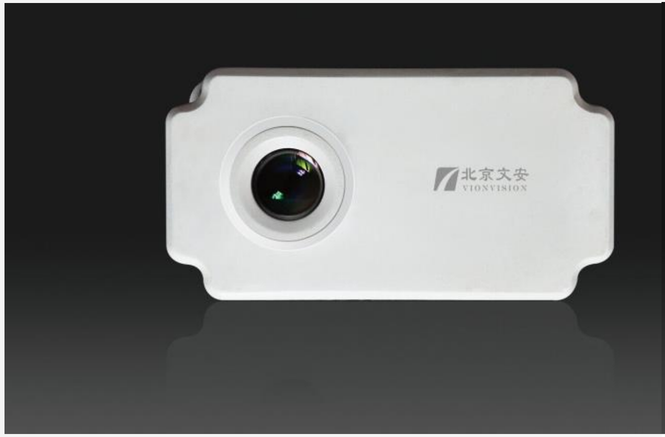
IC Technology

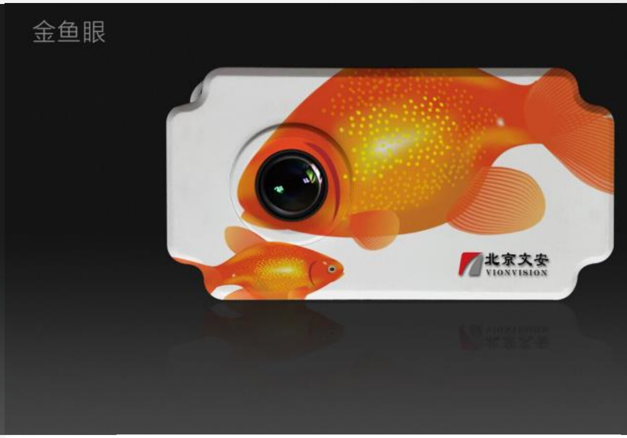
28nm low power

Quick Time-to-Market

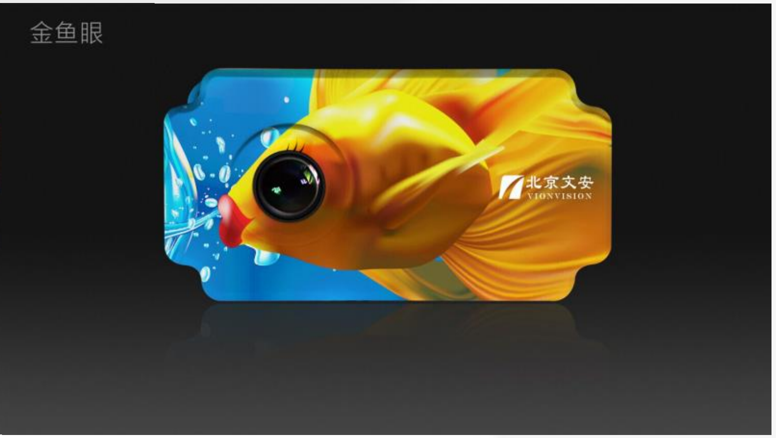
16'Q3

Smart Cameras Design- Bus Traffic Counting

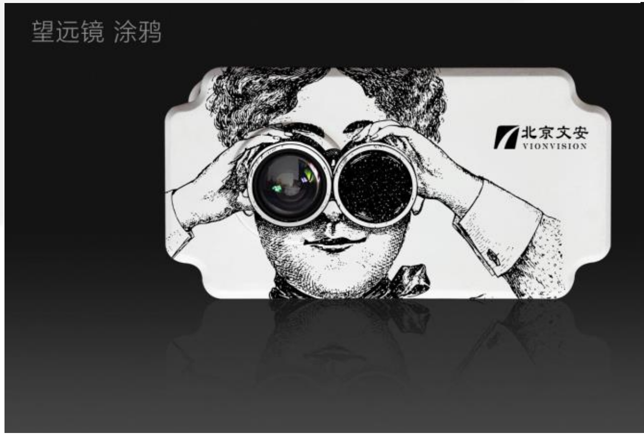




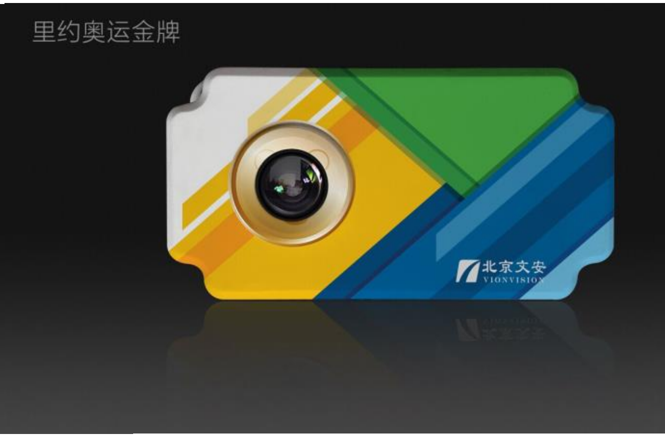
金鱼眼



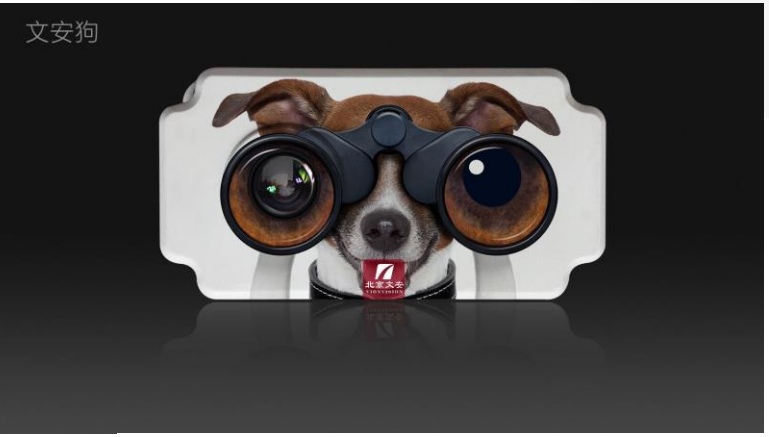
金鱼眼



望远镜 涂鸦



里约奥运金牌



文安狗

Back-End GPU Processing Units

CBox - Single GPU Unit

- GPU platform, 300 GFLOPS
- Analog/IP Video Input
- 2.5 inch hardrive & EMMC
- USB3.0, dual gigabyte LAN

Front End Control Terminal

- High performance , 300 GFLOPS
- 4 3.5” hard drives
- USB 3.0
- Dual gigabyte network ports

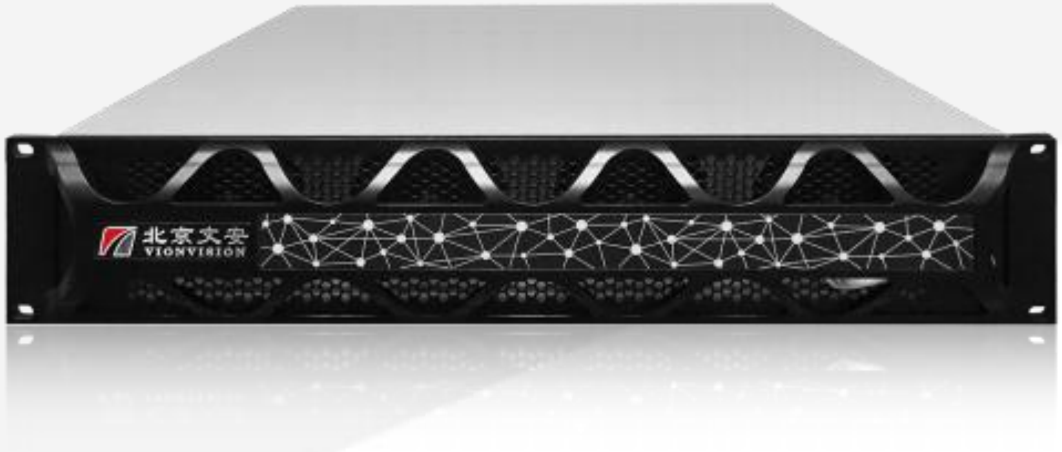
Smart Video & Audio Analysis Terminal

- Dual GPU, 600 GFLOPS
- 8 analog video & audio input
- Hard drive & EMMC storage
- 4 alarm in, 2 out

High Density GPU Cluster Server

- 40 nVidia GPUs
- 80ch 1080P H.264 decoding
- Processing up to 160ch@D1 or 80ch@1080p

Back-End GPU Processing Units - StarNet I



Spec: 40 nVidia GPUs, <600W, analyze up to 160ch@D1 or 80ch@1080p in real time

Applications: ITS, crowd management, IVS in various industries

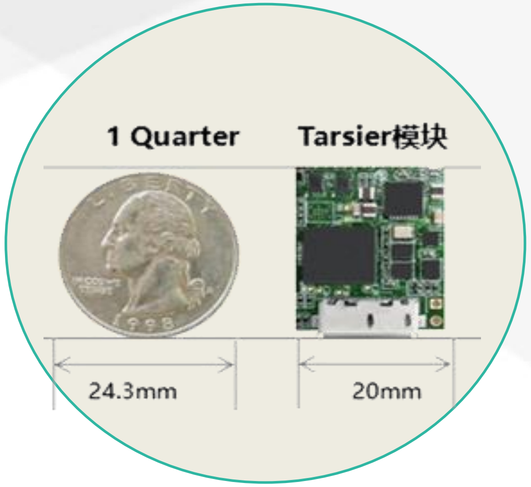
Release date: Q3,2016

- IOT+Computer Vision, Where Are the Applications ?
- Embedded CV Hardware
- GPU, VPU, and FPGA

DNN Speed on TK1, MA2450

<12W

<1.5W



- Nvidia TK1: **120ms**/frame
- Movidius MA2450: **140ms**/frame

Nvidia Tegra K1: CNN Implementation

- GPU for detection (relatively low frequency) and CPU for tracking
- Memory footprint is optimized via buffer sharing and TK1's unified mem mechanism
- Maximize CPU & GPU utilization via nvidia asynchronous ops and streams.
- cuDnn library for general layers
- Non-standard layers are implemented based on fine-tuned kernels
- 1x1 convolution, Balance between MACs & accuracy
- Balance between depth & width, depth for more representative power

Movidius Implementation

- fp16 is used with no accuracy loss
- Net architecture is tuned based on depth, width, kernel size
- Convolution/bias/relu/pooling -> combined layer
- All combined layer operations run in the on-chip CMX memory
- DDR and CMX exchange data when a combined layer is completed
- Implement 2D convolution in assembly kernel
- Bias, relu and pooling are done via processor intrinsics
- Make full use of the underlying "SIMD" shave architecture

Movidius Implementation: Inter-shave Task Parallelism

- Output feature map oriented strategy

  Put each shave in charge of several output feature maps, with load balanced among all shaves

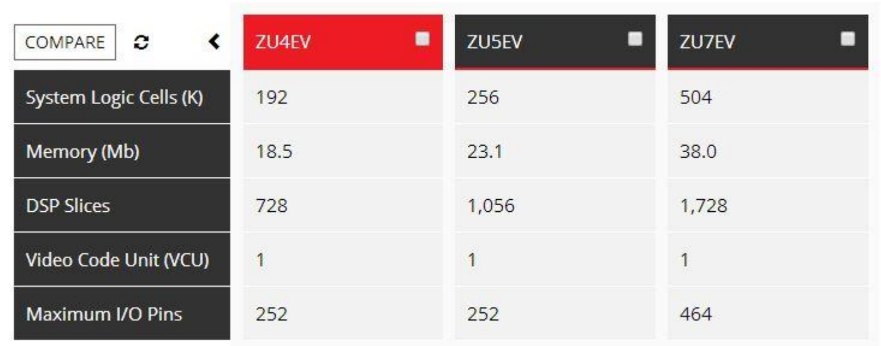
  Input feature map oriented strategy

each shave processor could take charge of "a band" of input feature maps, and compute all output channels of that spatial "band"

The above strategies are employed according to each layer's specific configurations, to minimize the amount of data transferred.

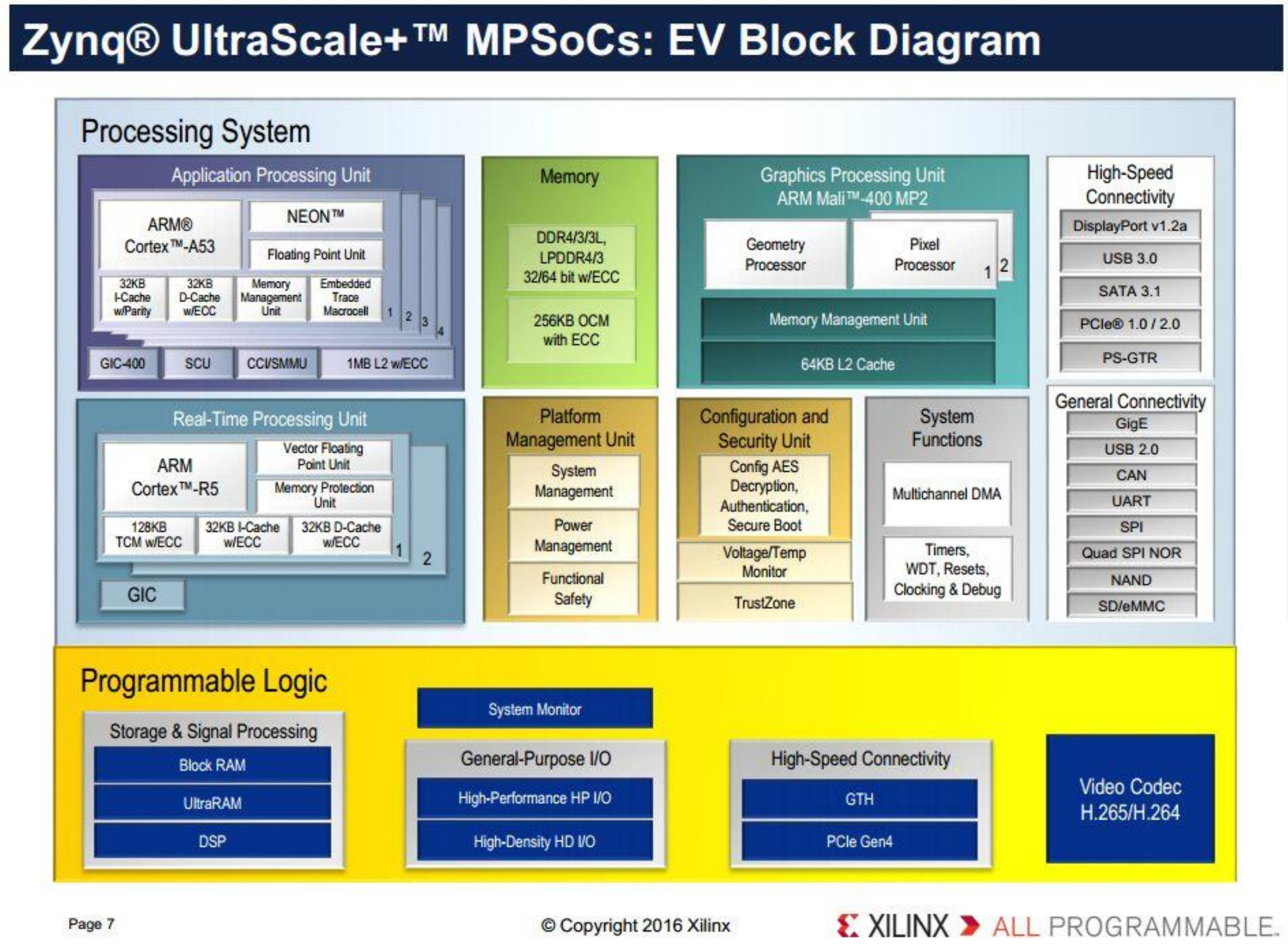
FPGA and DNN - Pottwal Project

VionVision



COMPARE

	ZU4EV	ZU5EV	ZU7EV
System Logic Cells (K)	192	256	504
Memory (Mb)	18.5	23.1	38.0
DSP Slices	728	1,056	1,728
Video Code Unit (VCU)	1	1	1
Maximum I/O Pins	252	252	464



Zynq® UltraScale+™ MPSoCs: EV Block Diagram

Processing System

Application Processing Unit: ARM® Cortex™-A53; NEON™; Floating Point Unit; 32KB I-Cache w/Parity; 32KB D-Cache w/ECC; Memory Management Unit; Embedded Trace Macrocell; 1 2 3 4; GIC-400; SCU; CCI/SMMU; 1MB L2 w/ECC

Memory: DDR4/3/3L, LPDDR4/3 32/64 bit w/ECC; 256KB OCM with ECC

Graphics Processing Unit ARM Mali™-400 MP2: Geometry Processor; Pixel Processor; 1 2; Memory Management Unit; 64KB L2 Cache

High-Speed Connectivity: DisplayPort v1.2a; USB 3.0; SATA 3.1; PCIe® 1.0 / 2.0; PS-GTR

Real-Time Processing Unit: ARM Cortex™-R5; Vector Floating Point Unit; Memory Protection Unit; 128KB TCM w/ECC; 32KB I-Cache w/ECC; 32KB D-Cache w/ECC; 1 2; GIC

Platform Management Unit: System Management; Power Management; Functional Safety

Configuration and Security Unit: Config AES Decryption, Authentication, Secure Boot; Voltage/Temp Monitor; TrustZone

System Functions: Multichannel DMA; Timers, WDT, Resets, Clocking & Debug

General Connectivity: GigE; USB 2.0; CAN; UART; SPI; Quad SPI NOR; NAND; SD/eMMC

Programmable Logic

Storage & Signal Processing: Block RAM; UltraRAM; DSP

System Monitor

General-Purpose I/O: High-Performance HP I/O; High-Density HD I/O

High-Speed Connectivity: GTH; PCIe Gen4

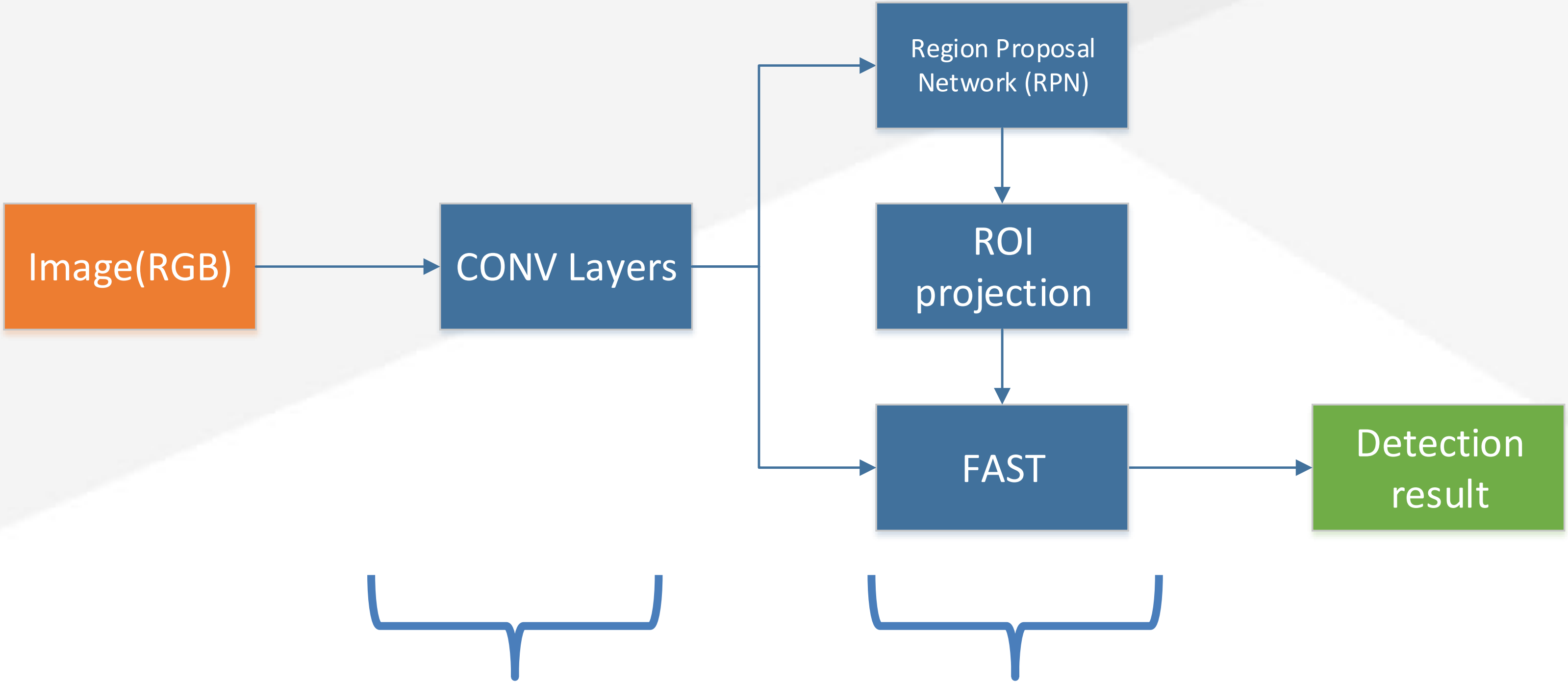
Video Codec H.265/H.264

Page 7

XILINX ALL PROGRAMMABLE.

The detection of neural network (Faster_RCNN)

The detection of neural network (Faster_RCNN)



Most of the computation :

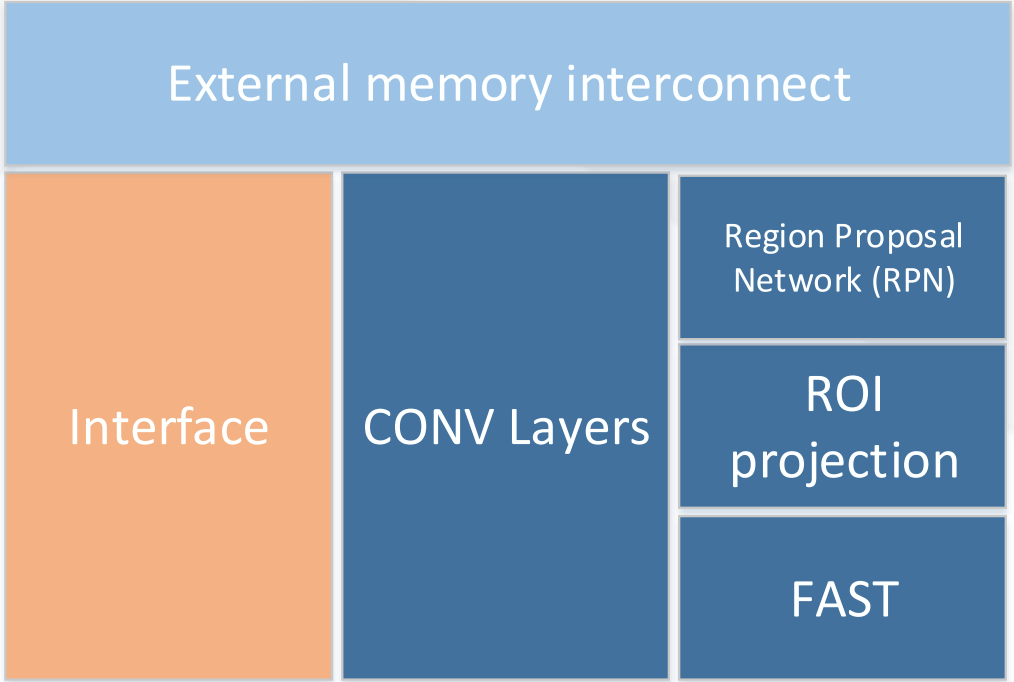
$$X = \sum_{i=0}^{n} x_i w_i$$

$$Y = \begin{cases} 0, & X < 0 \\ X, & X \geq 0 \end{cases}$$

Softmax, NMS, Coordinate inversion and so on

Design Features

- Global pipeline
- Ping-Pong
- Reduced data interaction
- SIMD
- Int8



FPGA and DNN - Pottwal Project

Performance

- Up to 8 channels of 1080p@30 detection
- Effective performance ：1.2T ops
- PE computational efficiency ：87.2%
- Latency：11.5ms

Platform	Performance(Effective)	Power	Performance per Watt
Our FPGA Platform	1.2T ops	7W	171.4G ops/W
NVIDIA TX1	220G ops	10W	22G ops/W
NVIDIA TK1	55G ops	10W	5.5G ops/W
Movidius MA2450	40G ops	1.5W	27G ops/W

VionVision

Vision without Limits!

VionVision

Vion Technologies Co., Ltd.

Vion Core Team

VionVision

Hai Tao, Dr., Founder&CEO

Tsinghua Univ. BS'91, MS'93; UIUC PdD'99; Sarnoff 99-01; UCSC Assoc.&Tenured Prof. 01-10. US NSF 2004 Young Career Award. Pulished 150+ papers in CV, 10+ US patents.

Tianshu Wang, Product Director

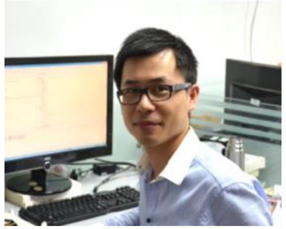
Xian Jiaotong Univ. BS'93,PhD'03; Microsoft Research 97-03, IBM Research 03-10; Lenovo Research 10-16, joined Vion in 2016.

Jun Song, CTO

Tsinghua Univ. Math, BS'01, MS'04; Responsible for all R&D work. Leads the smart traffic product core development & hardware system design.

Fan Yang, Director, Smart Counting

Tsinghua Univ. EE, BS'03, MS'06; Manager: business intelligence group; Manager: smart counting product line.



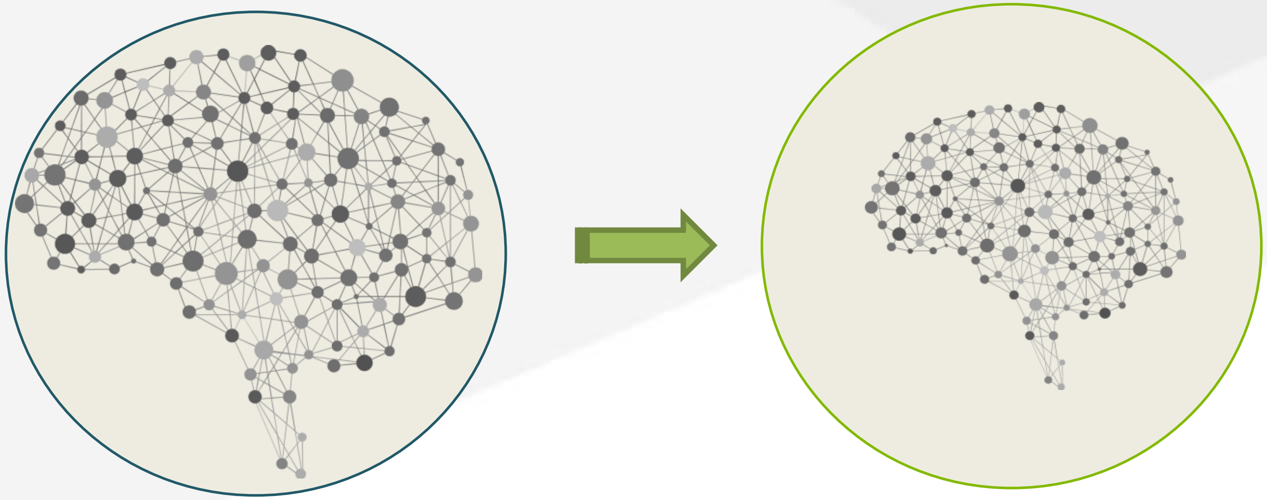
Yu Lin, Director, Vision System

Tsinghua Univ. AE, BS'03, MS'06; Manager: smart city product line; Manager: face recognition and intelligent video analysis group.

Xiang Zheng, Director, ITS

Tsinghua, CS, BS'01, MS'04; CV algorithm expert; data department manager; Rich vision product experience.

Embedded CNN Structure



- decrease the model size, less than **1 million** params
- limit the complextity to **1.5GMAC**, < 2% of VGG

Embedded CNN Performance

VionVision

- Detection Rate >**89%** (FDDB)
- **5%** lower than VGG (0.2FP/frame)
- Face detection scale from **20 pixels** to **400 pixels**
- Detection Rate >**83%** for real unconstrained local scenarios (illumination, expression, occlusion, pose)