Imagination

ACCELERATE THE AIOT WITH AI

© Imagination Technologies 2020

Yin Huang
Senior Business Development Manager
Yin.huang@imgtec.com

Who we are

>11 Billion
Chips shipped

~38%
Mobile GPU IP market share

Thousands
Of Patents

~43%
Automotive GPU IP market share

#1
NNA top Quant 8
AI-Benchmark 2019

© Imagination Technologies 2020

The Imagination IP family

Imagination

Imagination

The best solution for embedded graphics, AI, compute

Graphics

Broad suite of products covering embedded graphics needs across all markets

Compute

Dedicated Compute & AI hardware IP

PowerVR GPU

Scalable cores with best PPA
+ Safety Critical Automotive Cores

PowerVR Ray Tracing

Architecture for advanced modelling of light

NNA

PowerVR Neural Network Accelerators
AI Compute Software, Tools & Libraries

EPP

Ethernet Packet Processor

Embedded intelligence at low power enables the AIoT

Imagination

IoT

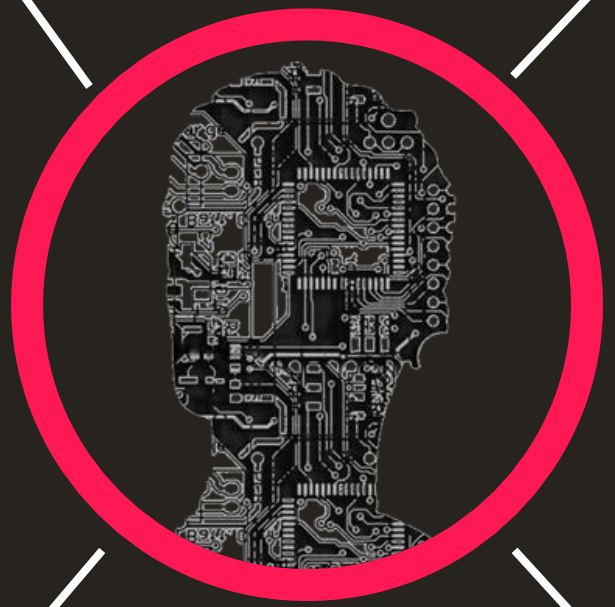
- Retail PoS automation
- Security
- Smart Cities
- Smart Buildings
- Asset Tracking
- Foot-fall analytics

Consumer/Media/AR/VR

- Image enhancement
- Gesture Control
- Viewer analytics
- Voice UI
- Environment awareness
- User interaction

Automotive

- Emergency braking
- Driver monitoring
- Lane departure warning
- Navigation& safety
- Collision warning
- Traffic sign recognition



Mobile

- Facial identification
- Picture annotations
- Application monitoring
- Speech recognition
- Advanced camera filters
- Smart Assistance

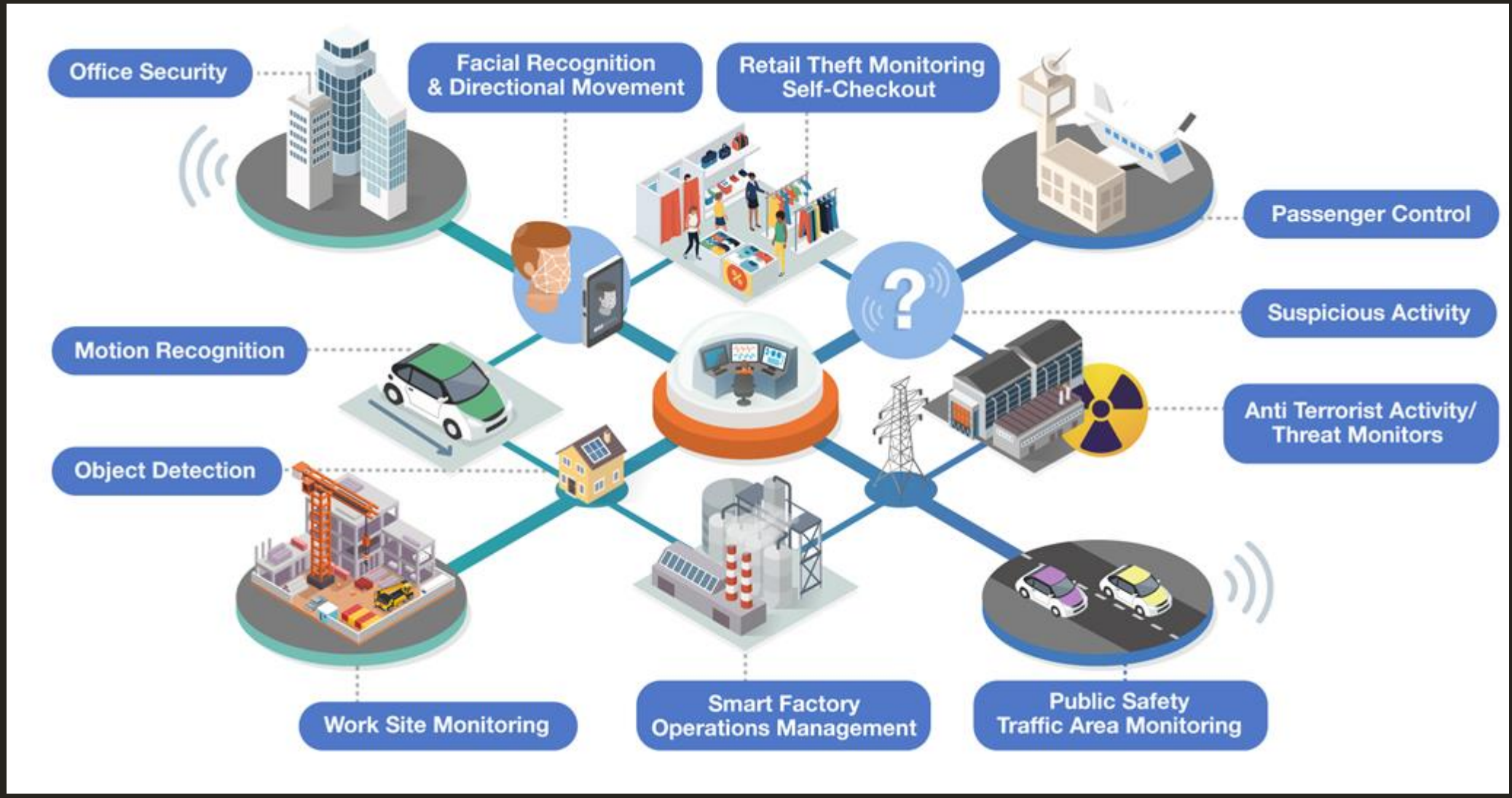
Industrial

- Defect detection
- Object detection
- Context awareness

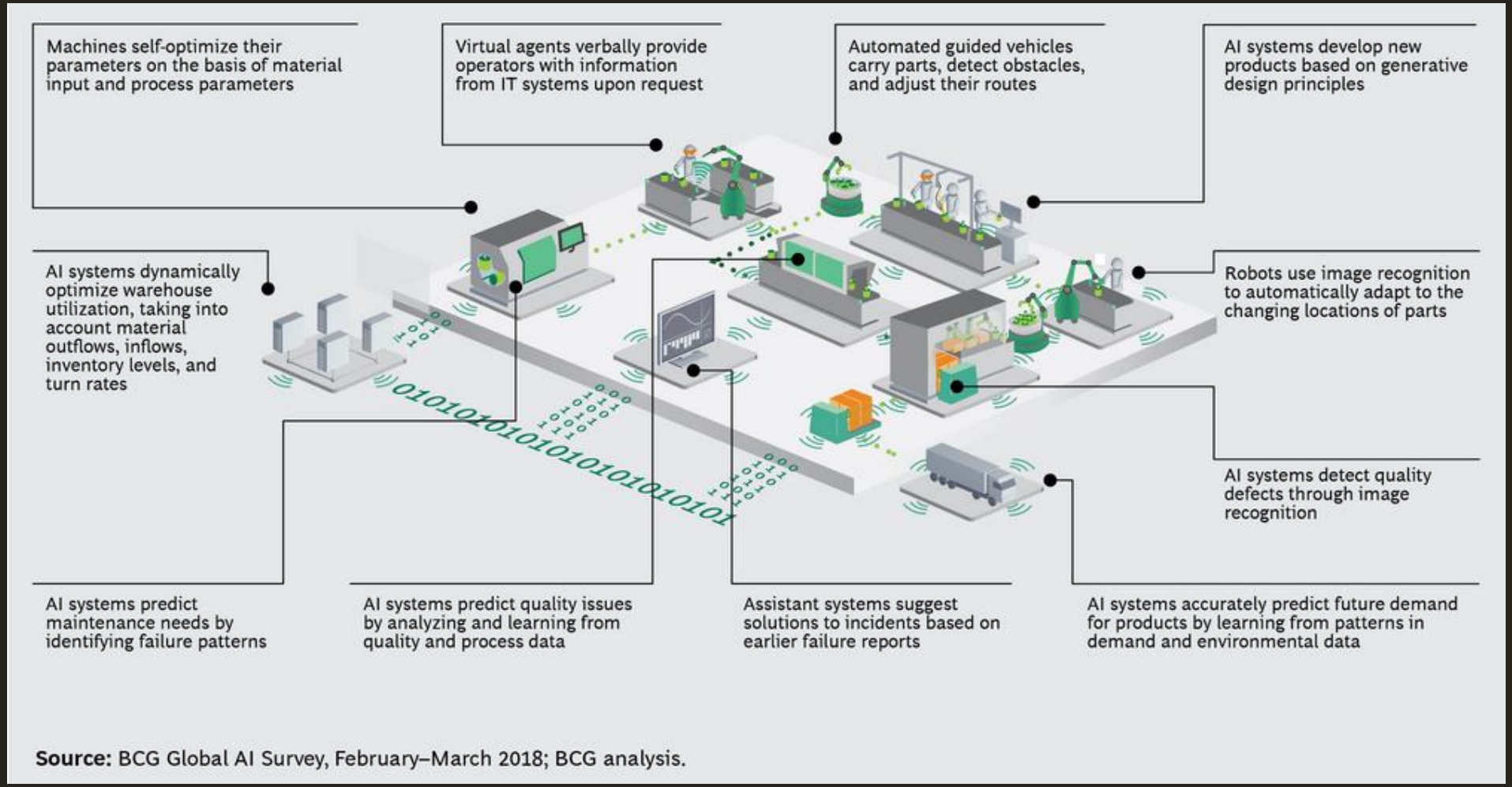
Smart Surveillance

- People tracking
- Abnormal behaviour alerts
- Queue Monitoring
- Event detection
- Engagement recognition
- Vehicle detection

Enabling the Smart City with AI & Connectivity

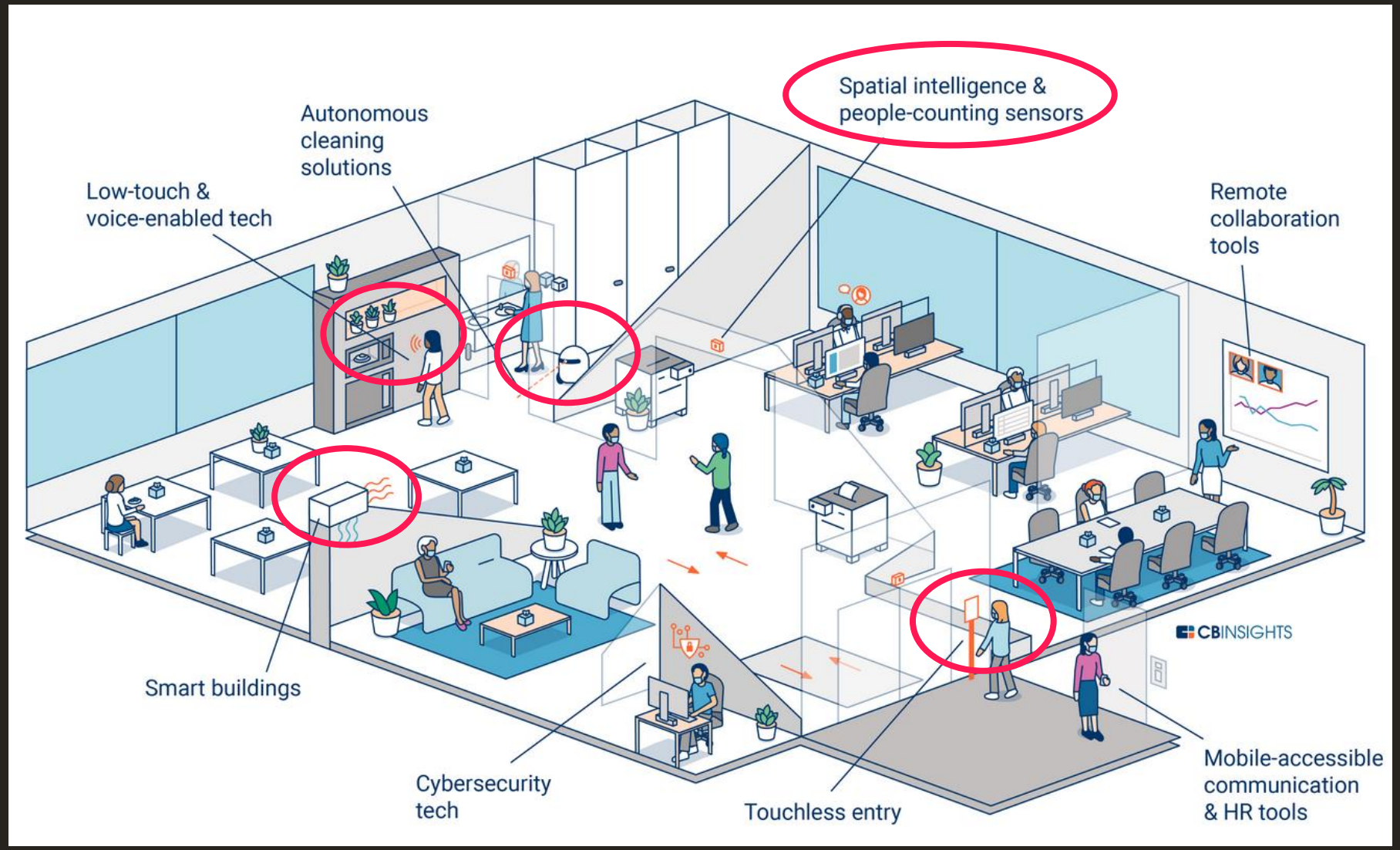


AIoT enables Smart Factory + Robotics



Source: BCG Global AI Survey, February–March 2018; BCG analysis.

AIoT enables Smart Workplace



Source: CBInsights, Imagination Technologies

AIoT enables Smart camera urban infrastructure

Application

People Tracking

AI Technology

Face Identification

Application

Queue Monitoring

AI Technology

Scene Recognition

Application

Suspicious Behaviour

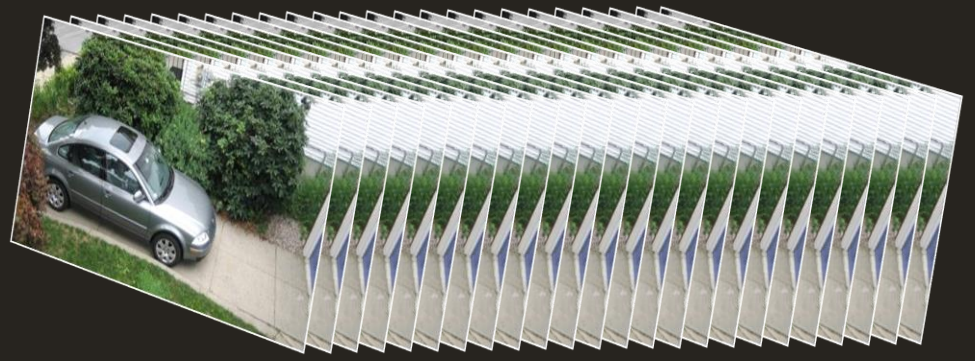
AI Technology

Object Detection

Edge Devices + Cloud Brain = Efficiency

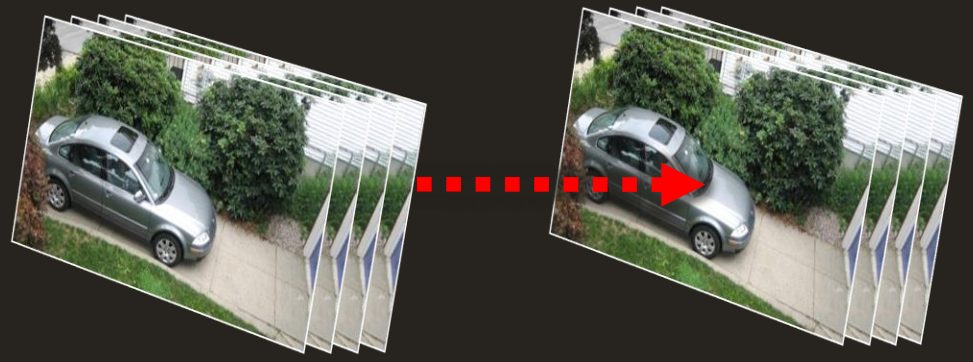
Only send vital data

Standard camera



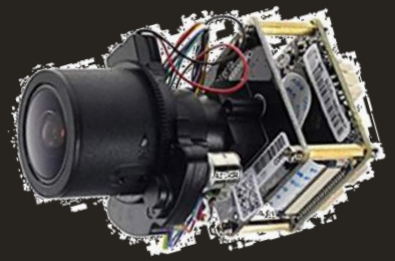
Send **all** video frames

Camera with motion sensor



Only **interesting** video frames

Smart camera with GPU + NNA

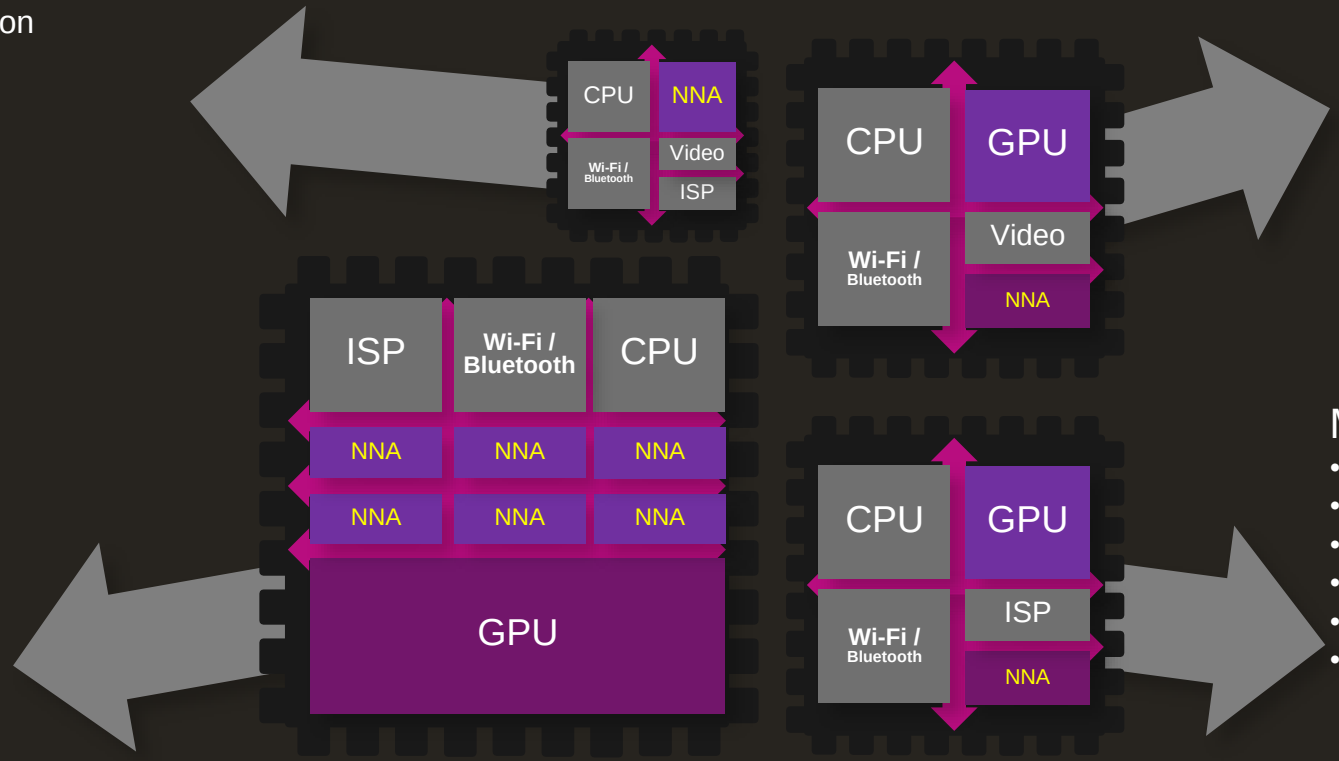


```
0  0  0  1111  0  0  0  0  111
```

Convert video to **metadata**

Imagination flexibility: Addressing the market

Security Camera
- People detection & recognition
- Emergency detection
- Emergency identification
  - Breaks
  - Fire/Smoke
  - Leak Identification



STB/DTV Platform
- User Recognition/Detection
- Video scene recognition
- Commercial detection

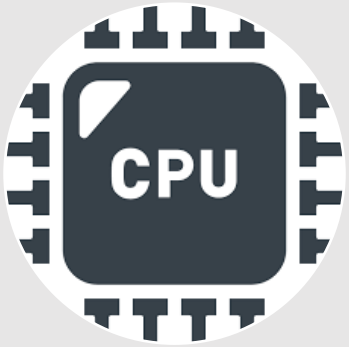
Automotive
Non-ADAS
- Lane detection warning
- Driver distraction warning
- Street Sign Detection
- Blind spot warning

ADAS
- Environment Recognition
- Obstacle detection
- Feed into ASIL Brain

Mobile
- Scene Detection/Recognition
- Image Depth Estimation
- Super Resolution
- Speech Recognition
- Noise Reduction
- GPU post processing
  - MSAA
  - Depth of field

Smart AI

Imagination

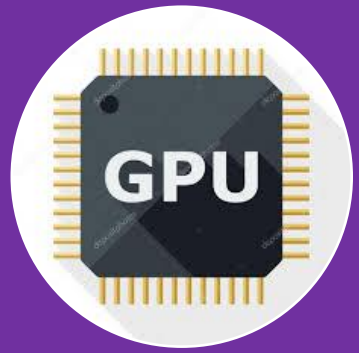


CPU

- Fully Flexible
- BUT inefficient and slow for high compute workloads

DSP

- Fully Flexible
- BUT hard to program – no standardisation, INT focussed



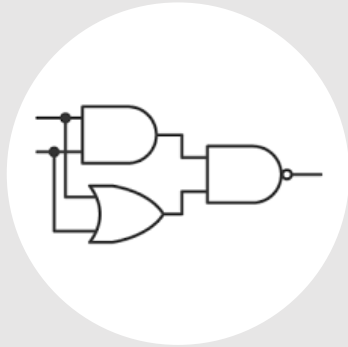
GPU

- Fully Flexible
- Standardised APIs for Compute, Float and INT support

Neural Network Accelerator

- Configurable
- Lowest power with domain specific flexibility

Best partners



Fixed Function

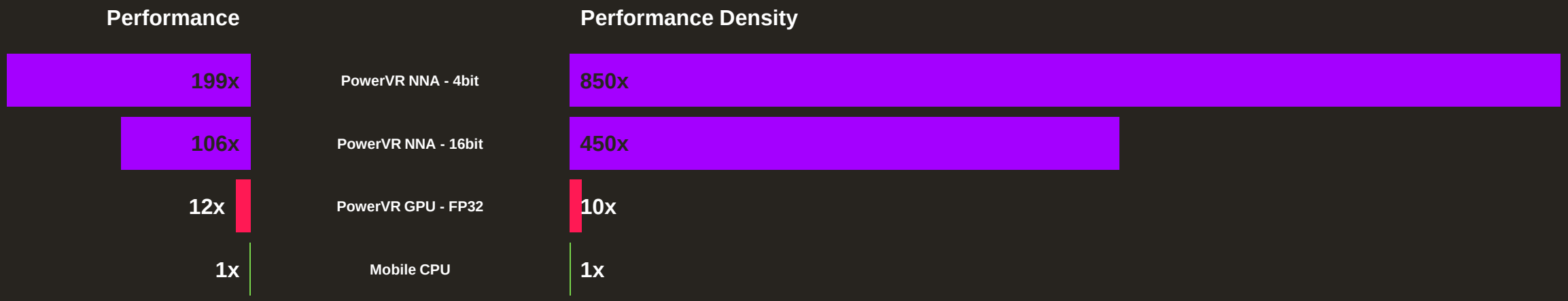
- Single usage case
- Lowest power BUT zero flexibility

Why the dedicated accelerator approach?

Imagination

Extracting the best performance for the area budget

PowerVR NNA Solution Comparison



Neural Networks have high bandwidth and computation requirements

A dedicated architecture addresses both of these issues

Cost efficient because of mobile experience

PowerVR NNA is designed to deliver the best performance per mm^2

INTRODUCING IMG SERIES3 NNA

Optimised for

Performance

Cost

Ease of use

© Imagination Technologies 2020

Series3NX- range

For Low Power IoT, battery powered and energy harvesting devices

PERFORMANCE

0.6 TOPS	1.2 TOPS	2.4 TOPS	5 TOPS	10 TOPS
AX3125	AX3145	AX3365	AX3385	AX3595

TOPS

- IoT
- Smart camera
- Wearables
- Speech rec
- Speech NLP
- Smartphone
- Home Assistant
- AR/VR
- ADAS/AV

© Imagination Technologies 2020

INTRODUCING IMG SERIES4 NNA

Imagination's

new architecture

for autonomy

© Imagination Technologies 2020

New Series4NX-MC Range

Imagination

The IP Business Model

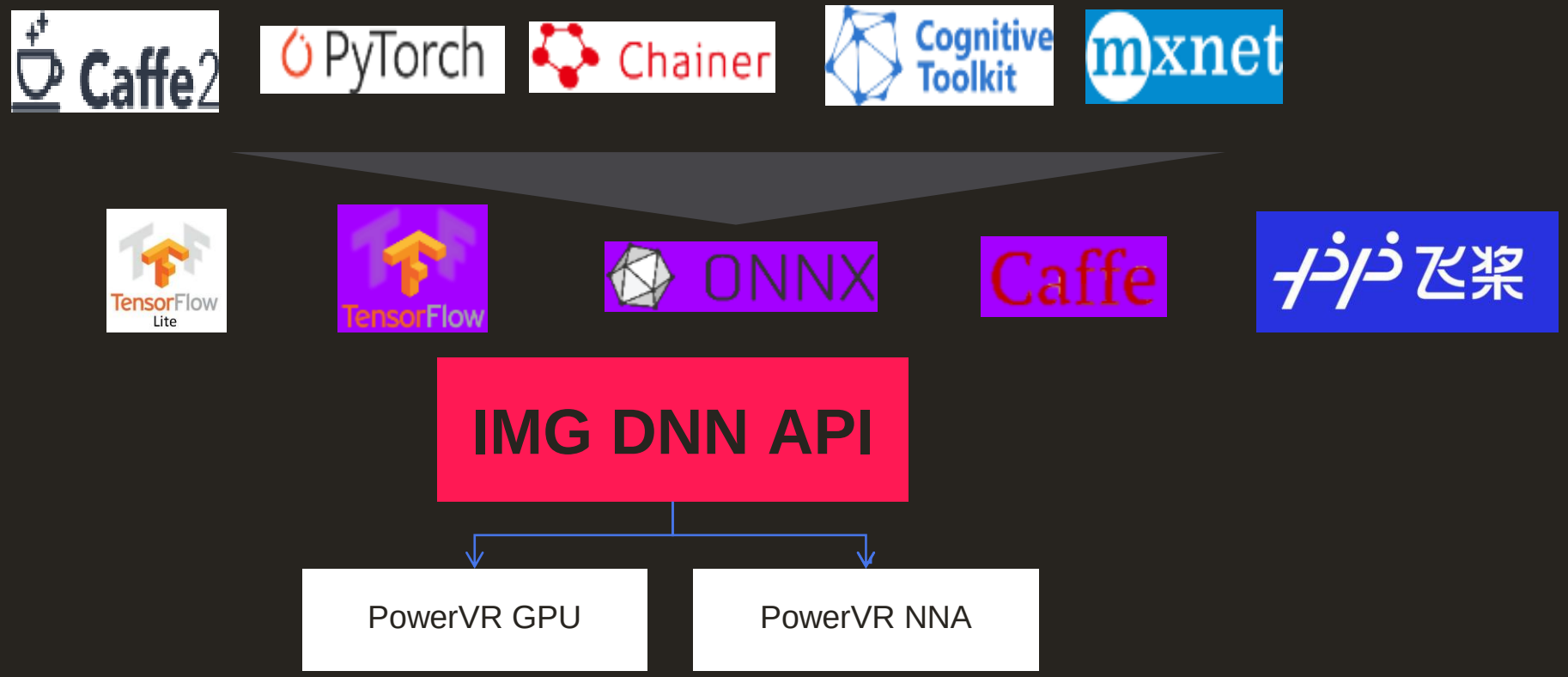
Imagination

BUILDING THE AIOT ECOSYSTEM TOGETHER

Tools Eco System

Imagination

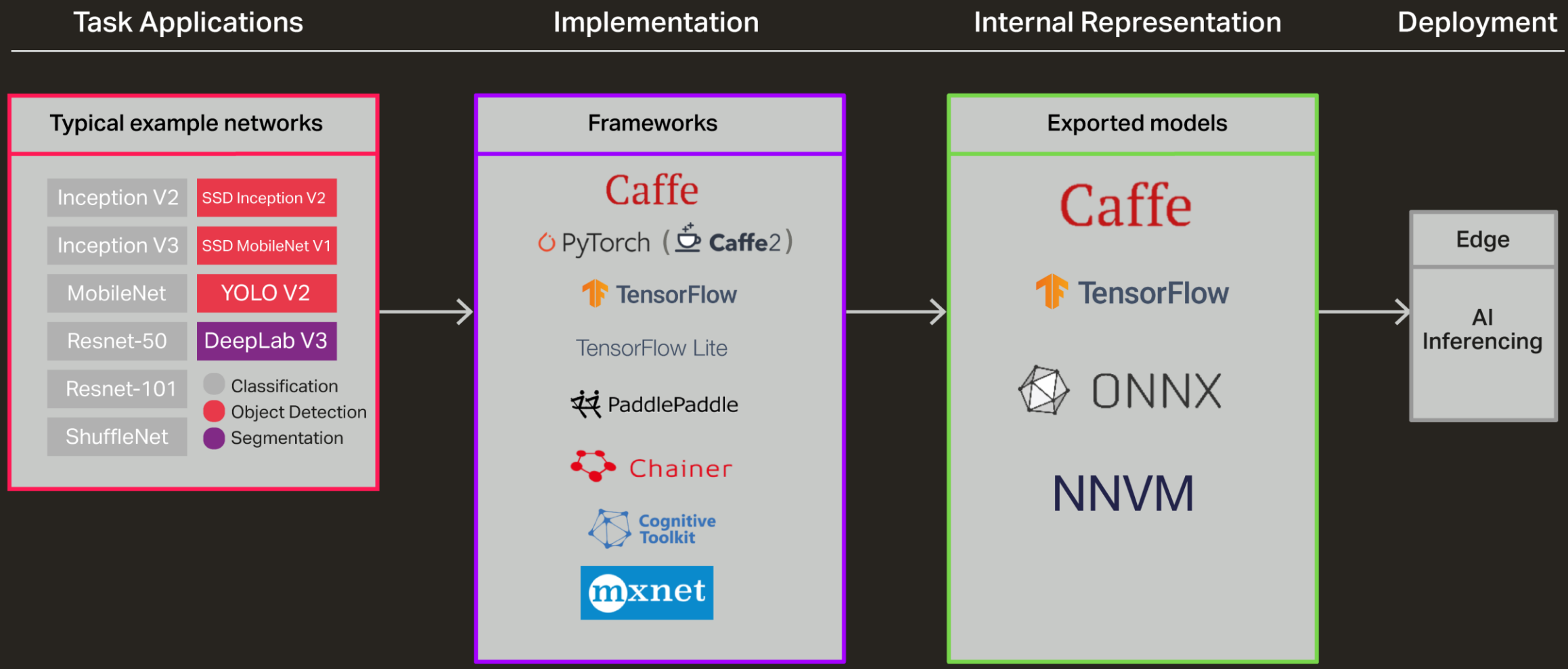
Your network, one API, optimised for GPU and NNA



Cost	Lowering effort to support a wide range of networks
Performance	Zero copy memory sharing between IP, driver-level synchronisation
Ease of use	Common tools across GPU and NNA

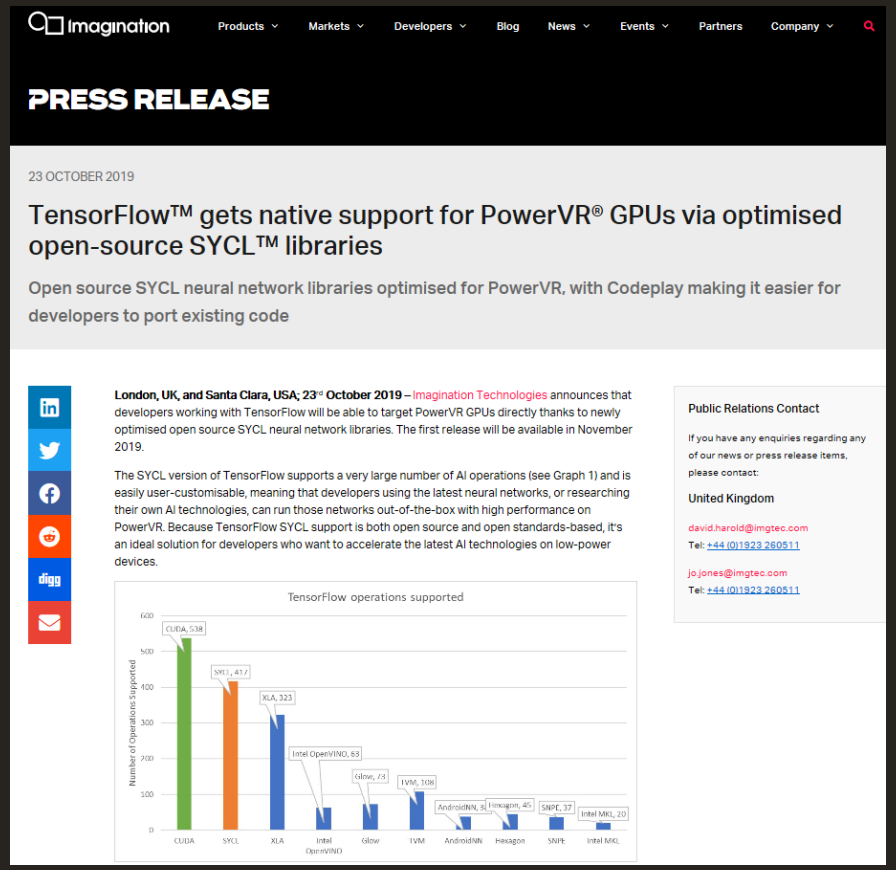
Tools Eco System

Your network, one API, optimised for GPU and NNA



© Imagination Technologies 2020

Tools Eco System



Imagination

RISC-V partnerships

RIOS

RISC-V

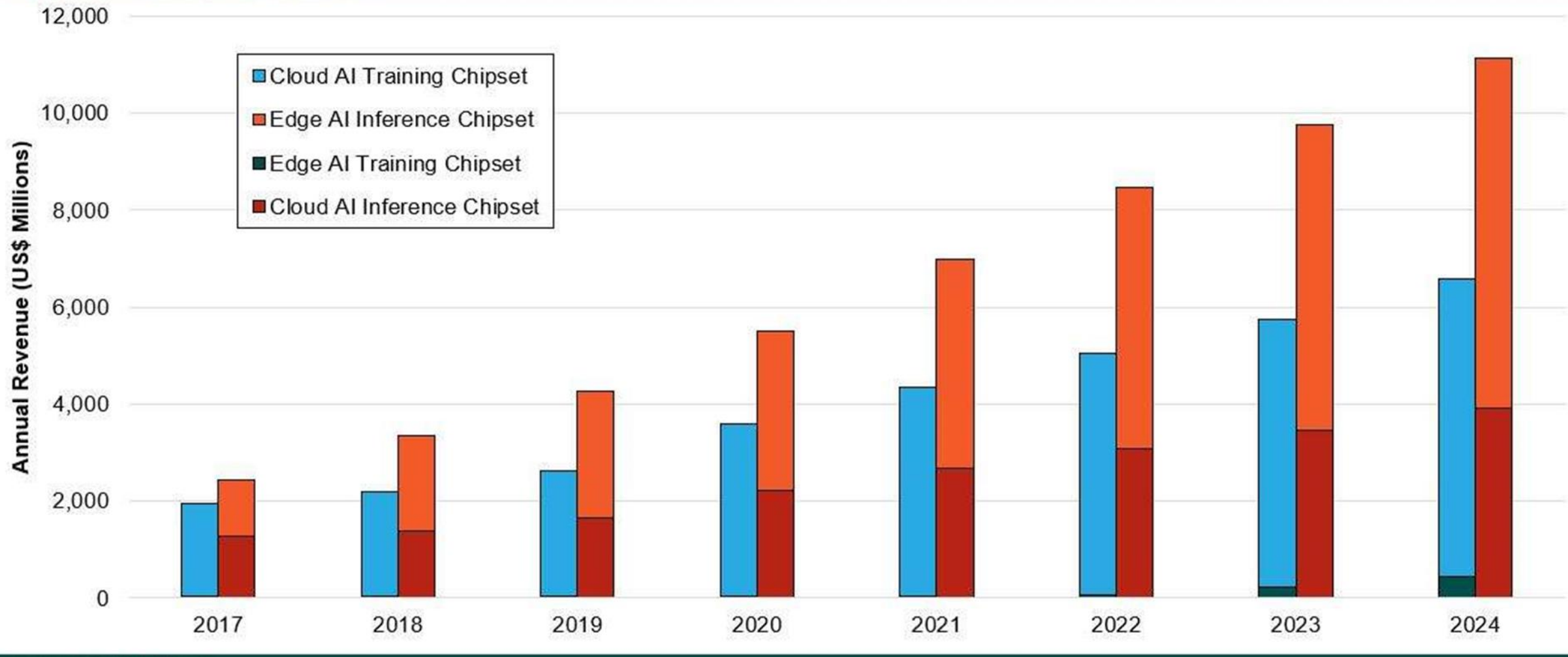
SiFive

ANDES TECHNOLOGY

Tools Eco System

Imagination

Total Annual Revenue from AI Chipset Sales by AI Inference and Training
World Markets, Forecast: 2017 to 2024



Source: ABI Research

Application Demo

NNA licensed in multiple markets

Demos available

Evaluation available

Detection of people

NNA running on FPGA acceleration

GPU on chrome book
https://www.youtube.com/watch?v=Bm9g9ZxFoHQ&feature=youtu.be

Silicon proven

Driver monitoring

Identification of multiple people on NNA

Rendering on GPU

https://www.youtube.com/watch?v=rLKfbj39E0Q

GPU + NNA

Smart surround view

People detection using GoogleNet SSD on NNA

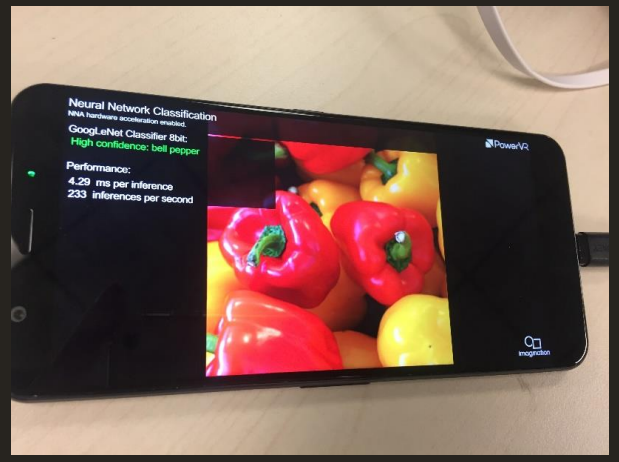
Surround view on GPU

https://www.youtube.com/watch?v=mZ5rvuXEu40

Application Demo

T710 Platform

IMG 2NX NNA inside phone， Android system



Max Freq : NNA 800M, CPU 2.0G, DDR 1.536G

IMG 2NX NNA inside development board, Linux system

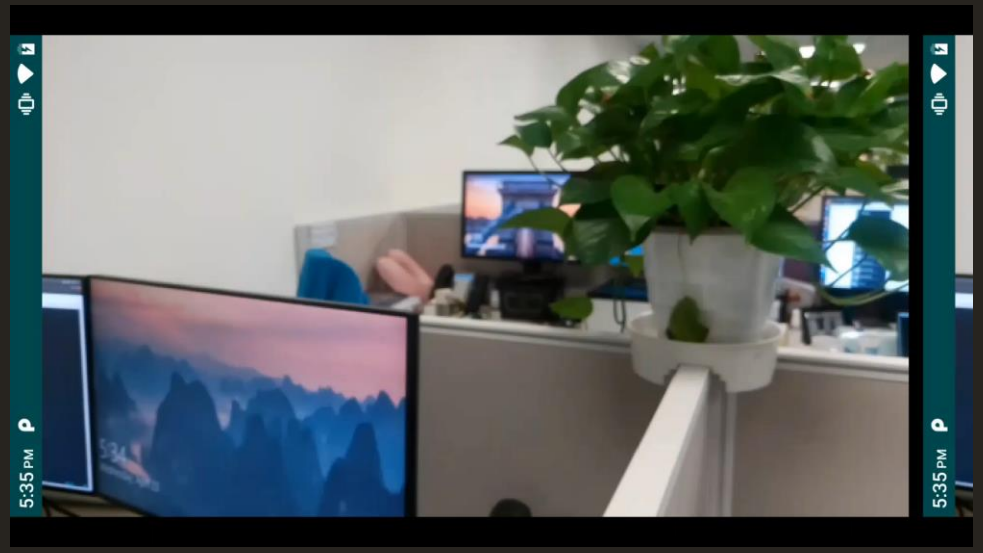
Imagination

Face detection & recognition

NNA practice

- This is one demo based on Android running one T710 based phone
- Java / C++ solution
- Application has two parts.
  - The frontend is face detection
  - The backend is face recognition
- Those networks are converted to MBS files by offline SDK mapping tools
- Also can run with Tensorflow lite

- For face recognition part
- When 160x160 resolution and 2NX running at 800M and 16 bit, the total bandwidth is 80M , and inference 195FPS per sec. It is useful when you recognize several people together.



Imagination

Human Pose detection

NNA practice

- Another demo based on Android running one T710 based phone
- It is based on Pose network
- Network are converted to MBS files by offline SDK mapping tools
- It is useful for suspicious behavior monitor

Input size 16 bit : 368x432
Total bandwidth : 450.00 MB
inference per sec : 31.80

Application Demo

Eye-mouth openness

NNA Practice

- It is one eye-mouth openness demo based on T710 Ubuntu development board
- Python solution
- We re-trained the object detection network to meet this scenario requirement.
- It can quickly detect the eye /mouth open or close. And it also can detect the yawn.
- It is one useful case to monitor car driver behavior

THNAK YOU