MPU+: A Transformative Solution for Next-Gen AI at the Edge

Dr. Petronel Bigioi

Chief Executive Officer

FotoNation

FotoNation®

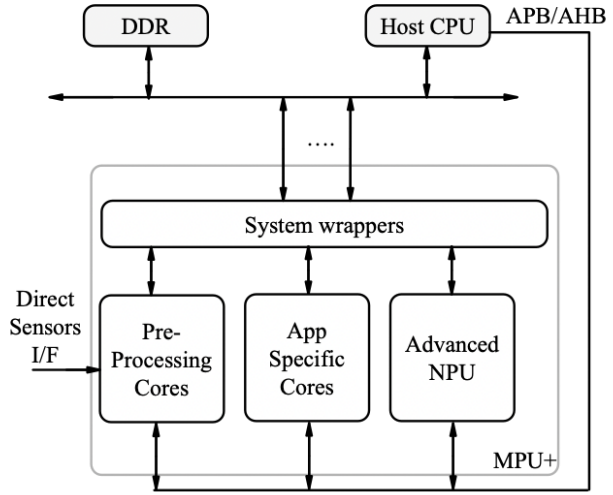
Cloud AI Applied for Edge Solutions

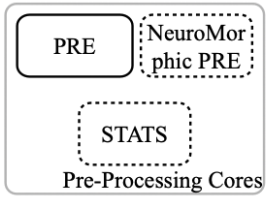
- **Cloud AI**: Large, resource-intensive, and requires constant connectivity
- **Edge AI**: Optimized for efficiency, running on limited hardware without cloud dependency
- **Optimization Limits**: Techniques like quantization and pruning help but aren't universal solutions
- **Flexibility Challenge**: General-purpose edge AI engines often struggle with diverse workloads
- **Best Approach**: Application-specific NPUs should be integrated into a heterogeneous AI platform for optimal performance

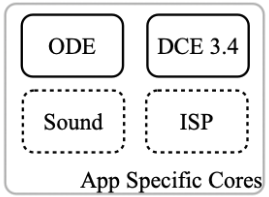
FotoNation®

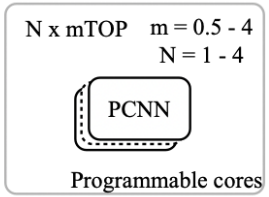
Media Processing Next-Generation Platform for Edge AI

- High-performance, low-power edge AI platform that enables performance improvements for real-world applications
- AI inference heterogeneous platform that contains accelerators adapted for classes of signal processing
- Gains: speed and latency, energy efficiency, adaptability and scalability



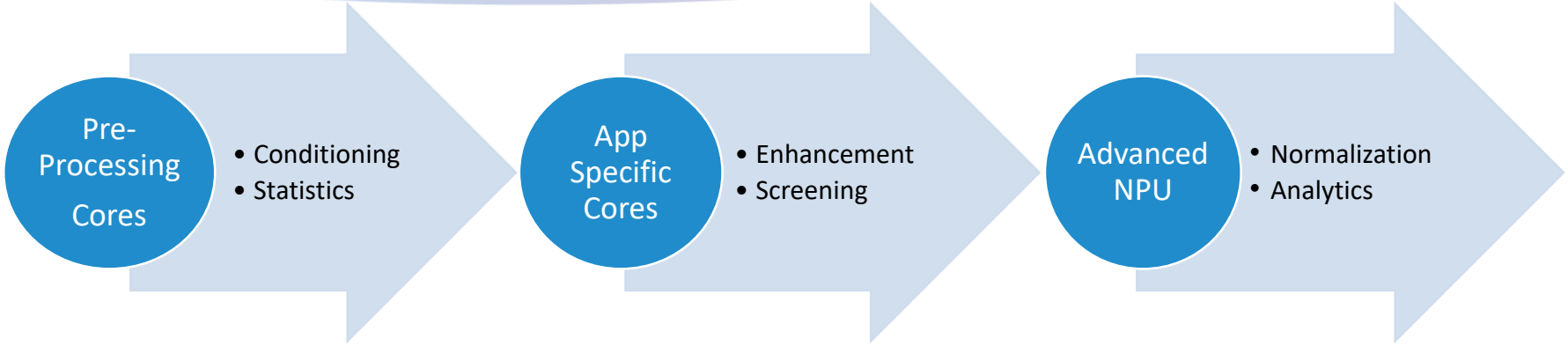






FotoNation®

MPU+ Processing & Analytics Flow



- Signal conditioning and stats done with traditional approaches
- Enhancement done in a hybrid way (traditional & AI), using a software defined 4k@60 pipeline
- Screening done with dedicated optimized neural processors (object, keyword, etc..)
- Advanced analytics done with on-the-fly normalization (e.g., for video – ROI scale, rotation, illumination) via a special signal-processing-capable load data module (a.k.a. layer0), part of our NPU (a.k.a. PCNN)

FotoNation®

MPU+ Hardware Cores

Pre-Processing Cores

PRE (-processing)

On-the-fly signal conditioning and statistics

Neuromorphic PRE

On-the-fly event-to-texture, making neuromorphic sensors compatible with classic NN processing/fabric

App Specific Cores

Distortion Correction Engine

HQ resampling engine supporting static and dynamic correction grids

Audio Source Separation Engine

Sound source separation optimized for voice isolation

Image Signal Processing

Traditional ISP blocks interconnected via AI fabric

Object Detection Engine

Multiple class ultra long-range object detector

Advanced NPU

PCNN (Programmable CNN)

Configurable 512-4096 ops/cycle/core multi-core neural inference engine optimized for signal analytics and processing.

Built in “layer 0” for ROI normalization, FP16 equivalent native pipeline, infers encrypted models, optional quantization and compression support, etc.

FotoNation®

MPU+ Edge AI Enabled Solutions

Always-on intelligence while the device is “off”

(e.g., facial analytics as an enabler, voice denoising, keyword hunting, etc..)

Surveillance & monitoring

(e.g., always-on devices for perimeter protection)

Smart IoT: TVs to drones …

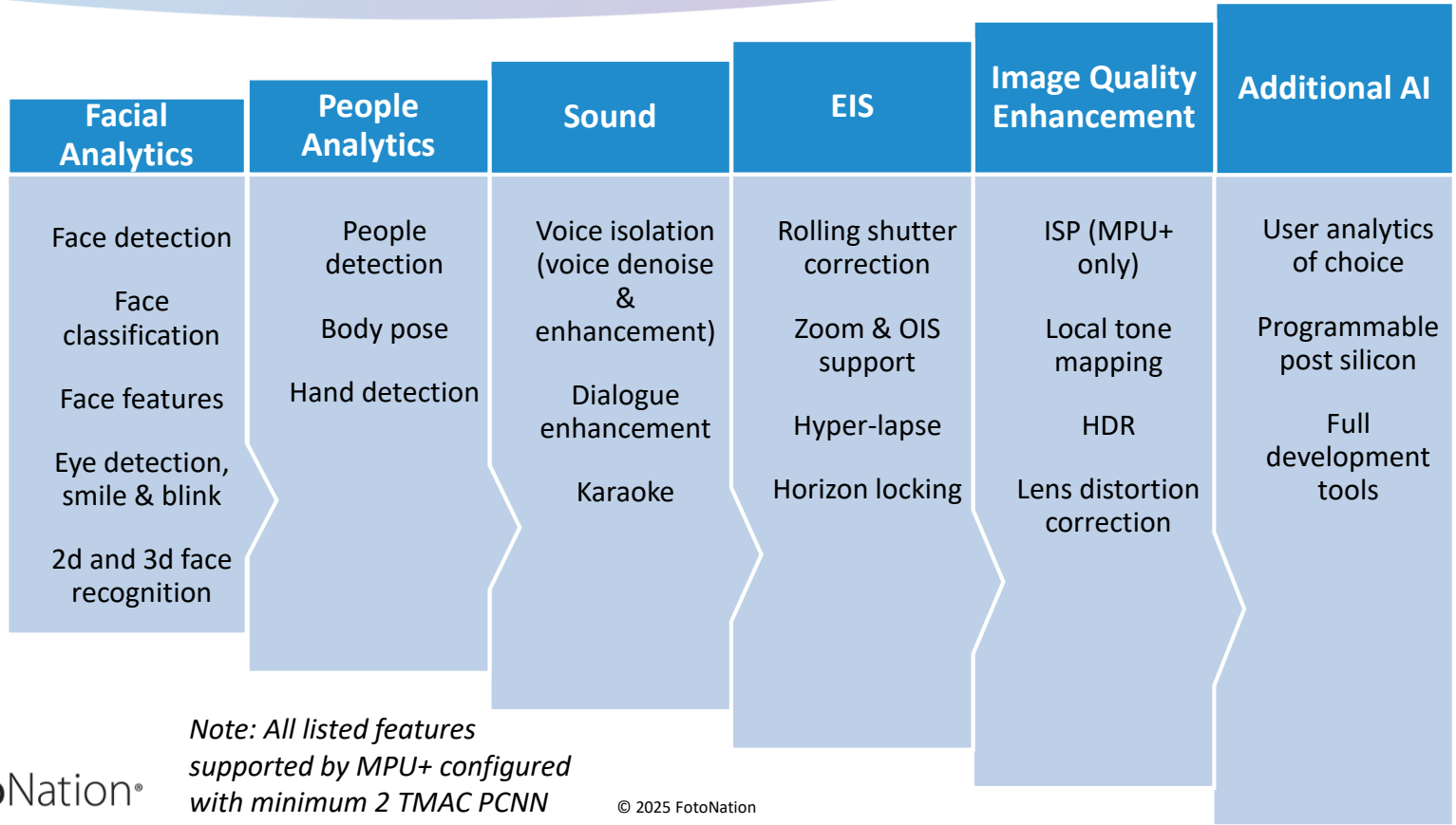
(e.g., ultra-low-power presence, engagement, demographics and personalization)

Wearables

(e.g., voice isolation, multi-modal sensor AI tokenizer, etc.)

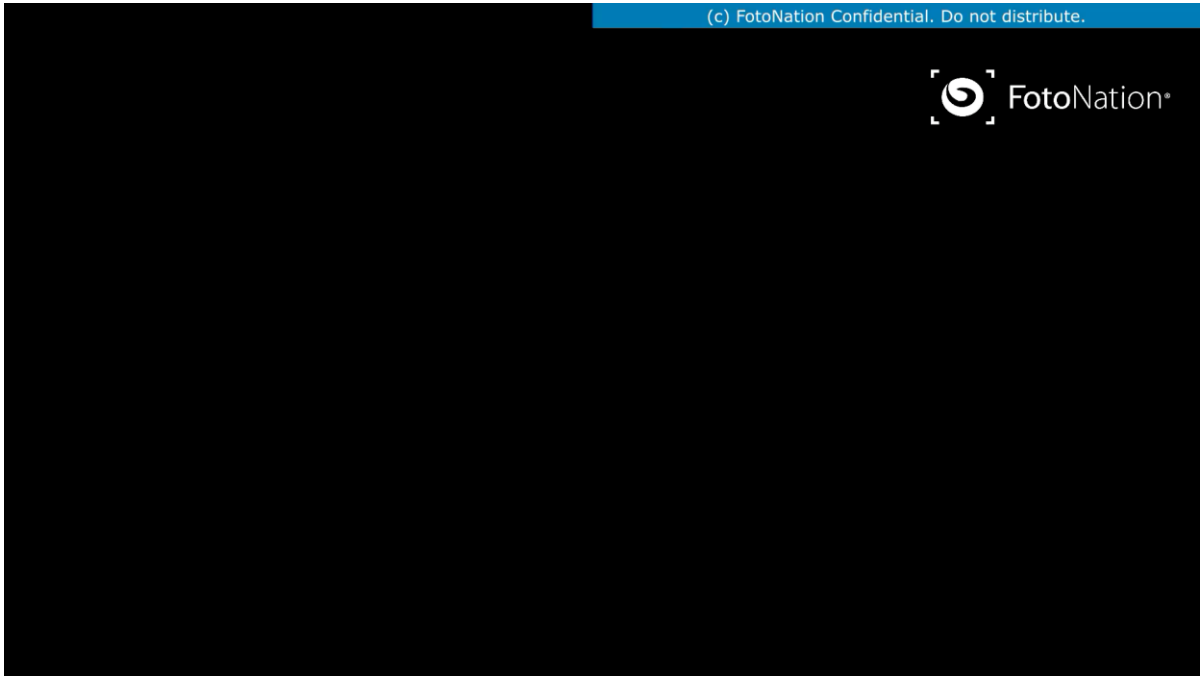
FotoNation®

MPU+ Supported Features

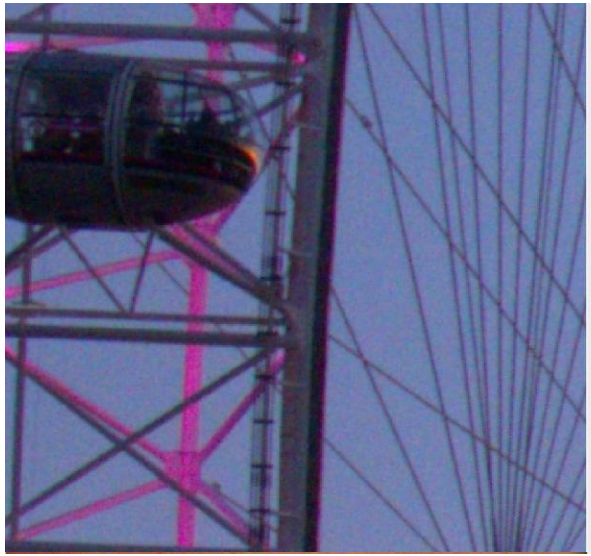


Note: All listed features supported by MPU+ configured with minimum 2 TMAC PCNN

MPU+: Seeing is Believing



MPU+ Image Quality Preview



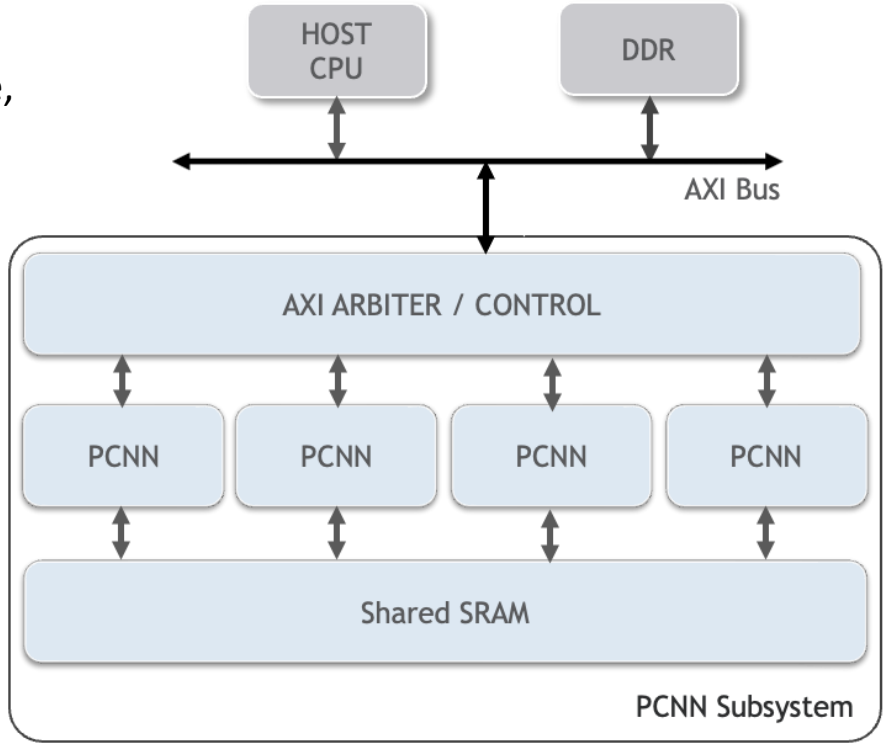
OpenCV reference

Leading OEM production

FotoNation MPU+

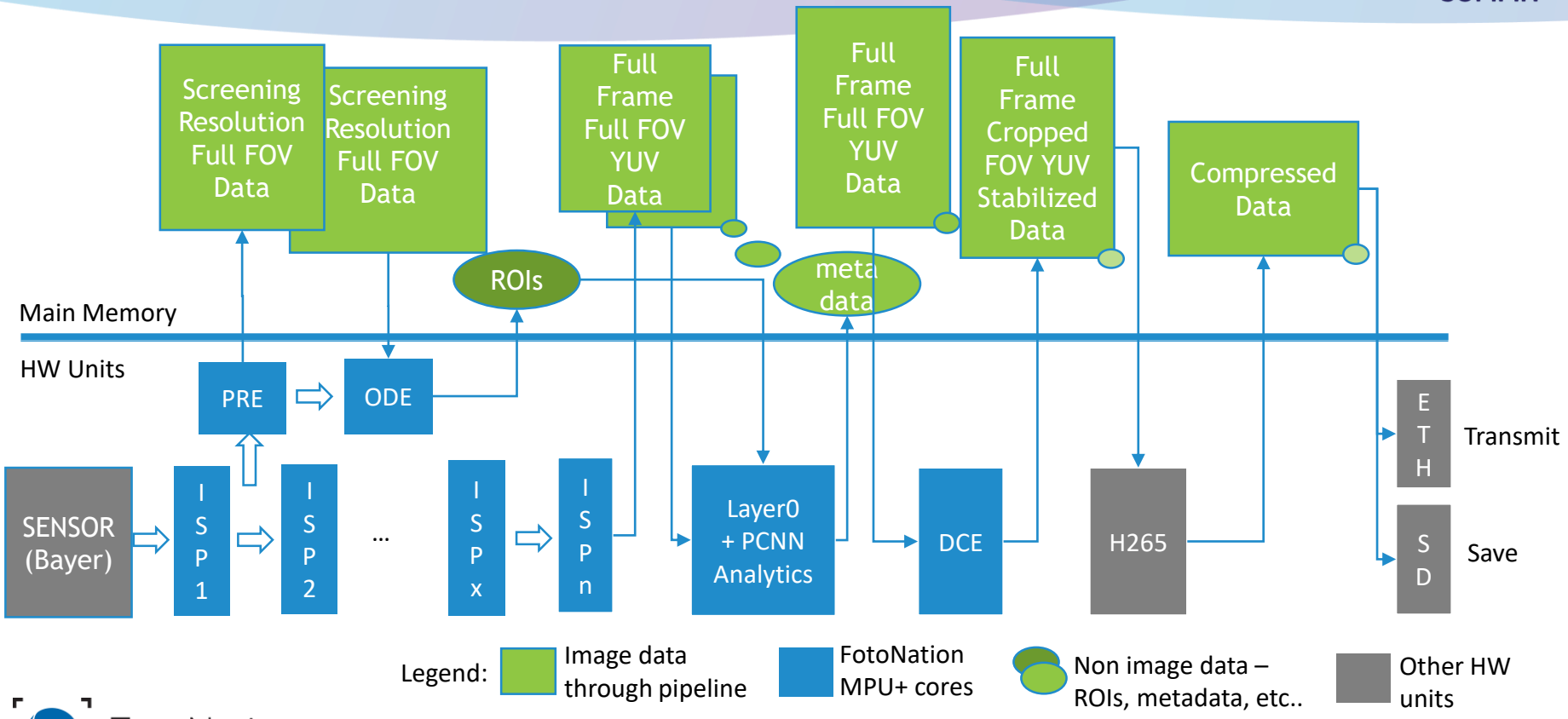
MPU+ Under the Hood — Programmable CNN Subsystem

- Up to 4 TFLOPS per core, 80%+ MAC efficiency
- Built-in “layer-0” engine for enhancement, scale, rotation and data format conversion
- Native support for 8, 10, 12, 14, 16-bit I/O data
- Half-float equivalent internal math pipeline
- 2-bit up to 16-bit weights encoding (either integer, fixed point of half-float)
- Support for compression, quantization, and on-the-fly model decryption
- Designed for optimized memory bandwidth
- Designed for either 2D shared SRAM or 3D (hybrid bonding) with SRAM or alternative



FotoNation

MPU+ Under the Hood – Video & CV Flow



FotoNation®

The Cost of Generality vs Application Specific AI

State-of-the-art

- 48 TOPS (2x24 TOPS cores) Edge-AI platform
- Silicon cost: estimated 10 sq mm per core
- Operating frequency: 768 MHz
- Tech node: unknown
- Reference model*: 30 fps (only one core used); MAC efficiency is around 0.7%
- Power consumption: unknown

FotoNation MPU+

- 0.25 TOPS (single core) ODE
- Silicon cost: less than 0.4 sq mm (350 Kgates + 380 KB SRAM)
- Operation frequency: 1,000 MHz
- Tech node: 12 nm
- Reference model*: 30 fps; MAC efficiency around 70%
- Power consumption: 25 mW (logic), 28 mW (SRAM), 55.25 mW (DDR)

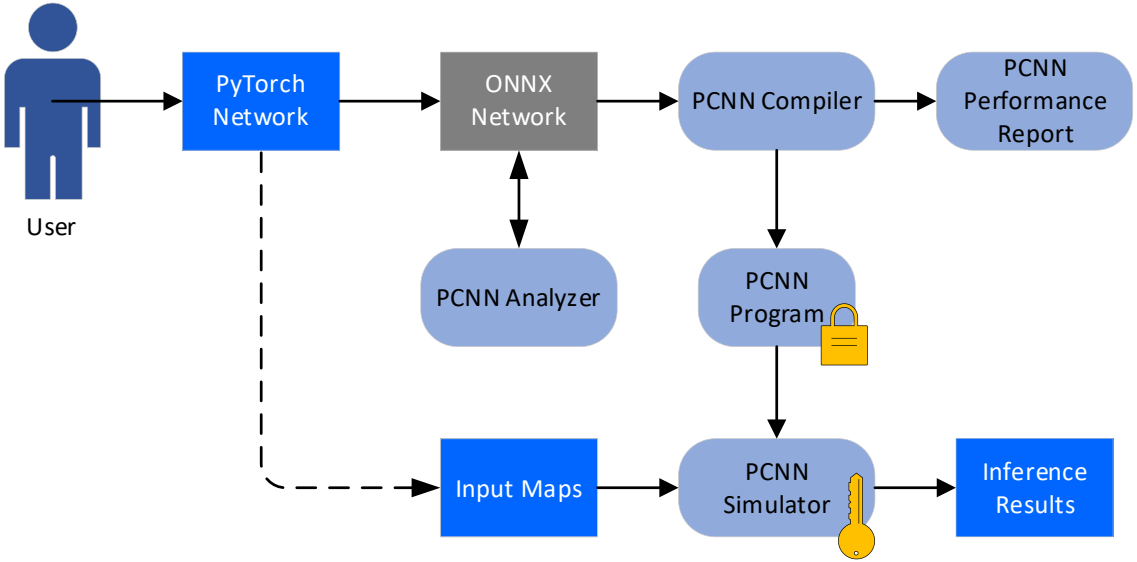
COST OF GENERALITY – 10 sq mm vs 0.4 sq mm

** Reference model used: multi-class object detector, 2.8 GMAC/frame, working Full-HD resolution, min object 5x5 pixels*

FotoNation®

MPU+ Toolchain

- **Easy-to-use** tools for model deployment on MPU hardware, with optimal allocation of hardware resources



- Transform ONNX to PCNN/ODE binary program
- Bit-exact simulation of the hardware
- Check compatibility & automatic modifications

Performance optimization

- Advanced network graph optimization algorithms
- FP32 to FP8 quantization for weights and feature maps
- Platform details (latencies) used in compilation

FotoNation®

Conclusions

- **The Shift to Edge AI**: A strong industry trend is driving intelligence to the device for **real-time processing, lower latency, enhanced autonomy, improved privacy, and better scalability**.
- **Evolving Methods & Architectures**: Simply combining various processors (CPU/GPU/DSP/etc.) for edge AI **sacrifices efficiency, increasing cost and complexity** instead of optimizing performance.
- **MPU+ as the Future of Edge AI**: A **versatile signal processing platform** with **heterogeneous neural engines** tailored for specific signal processing tasks—delivering **high efficiency, ultra-low latency, and top-tier performance** for next-gen edge AI applications.
- **Seamless Integration**: **MPU+ can serve as the video/audio processor of choice**, available as an **IP core or (soon) chiplet**, making it an attractive solution for **multiple industries**.

FotoNation®

Resources

FotoNation website:

https://www.fotonation.com

2025 Embedded Vision Summit

FotoNation MPU+ demo available at booth 621 in the exhibit hall

Thank you!

Dr. Petronel Bigioi

petronel.bigioi@fotonation.com

FotoNation®