XILINX

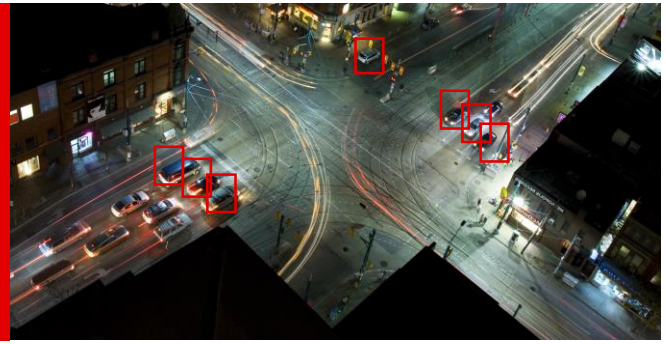
The Future of Smart World

Seong H Kim, Ph.D.
Sr. Director Datacenter Systems Architect

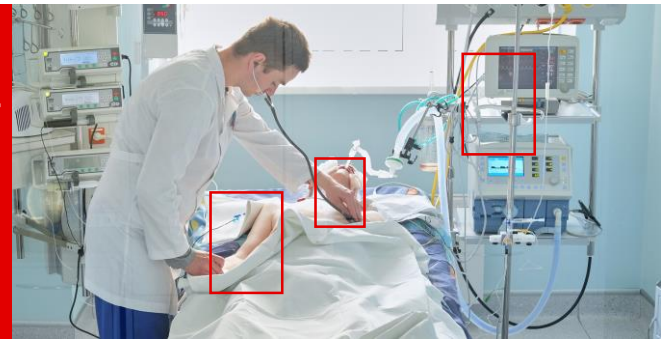
The Smart World Trend

Smart Retail

Smart City



Smart Hospital



Retail loss reached more than $60B in 2019 in the US alone, according to the National Retail Federation

By 2040, an additional 1.5 billion people will be living in the world's cities

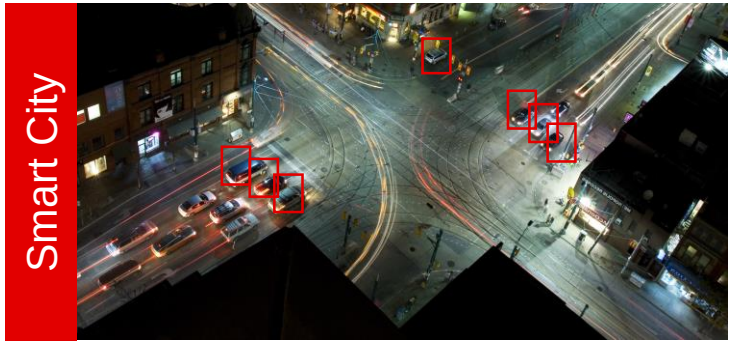
Quickly evolving machine learning technology, the industry is equipped to enable intelligence

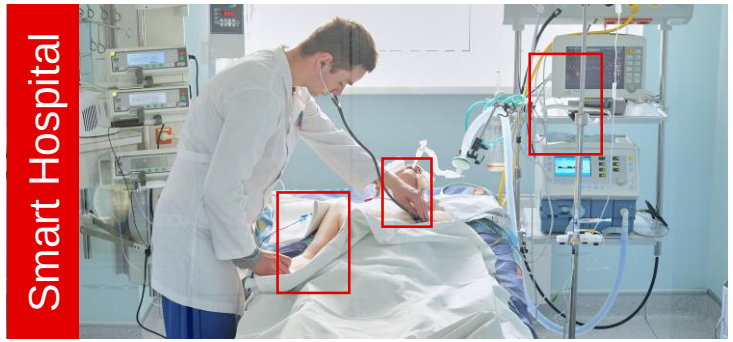
New applications are emerging, such as Smart City, Smart Retail, Smart Hospital, Smart Factory and Building, and others

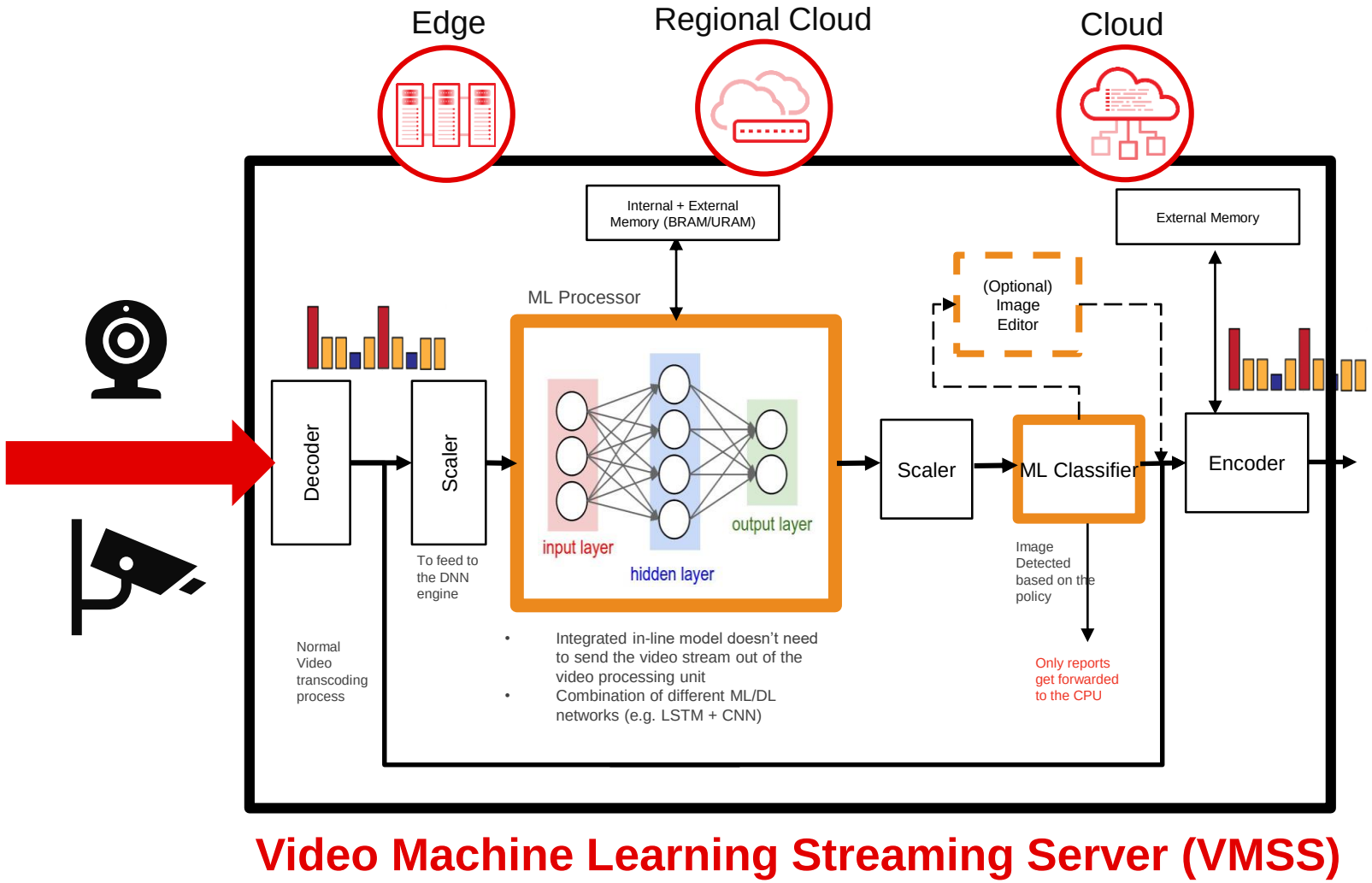
The foundation of new technology enabling Smart World applications is Video processing and ML inferencing

XILINX

Xilinx Smart World Target Markets



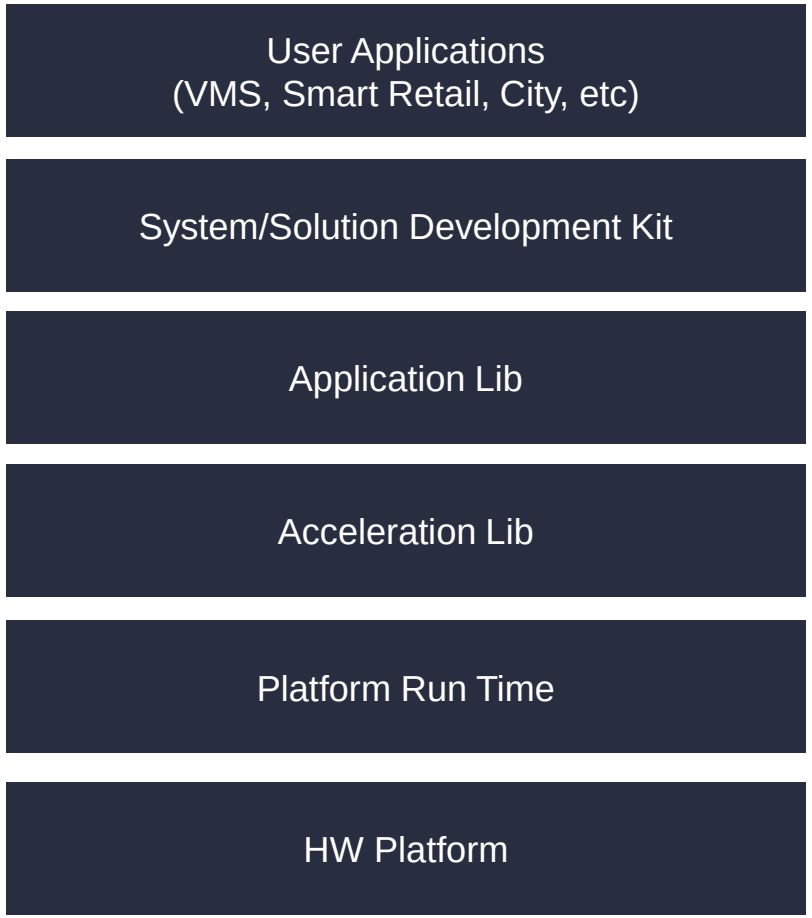


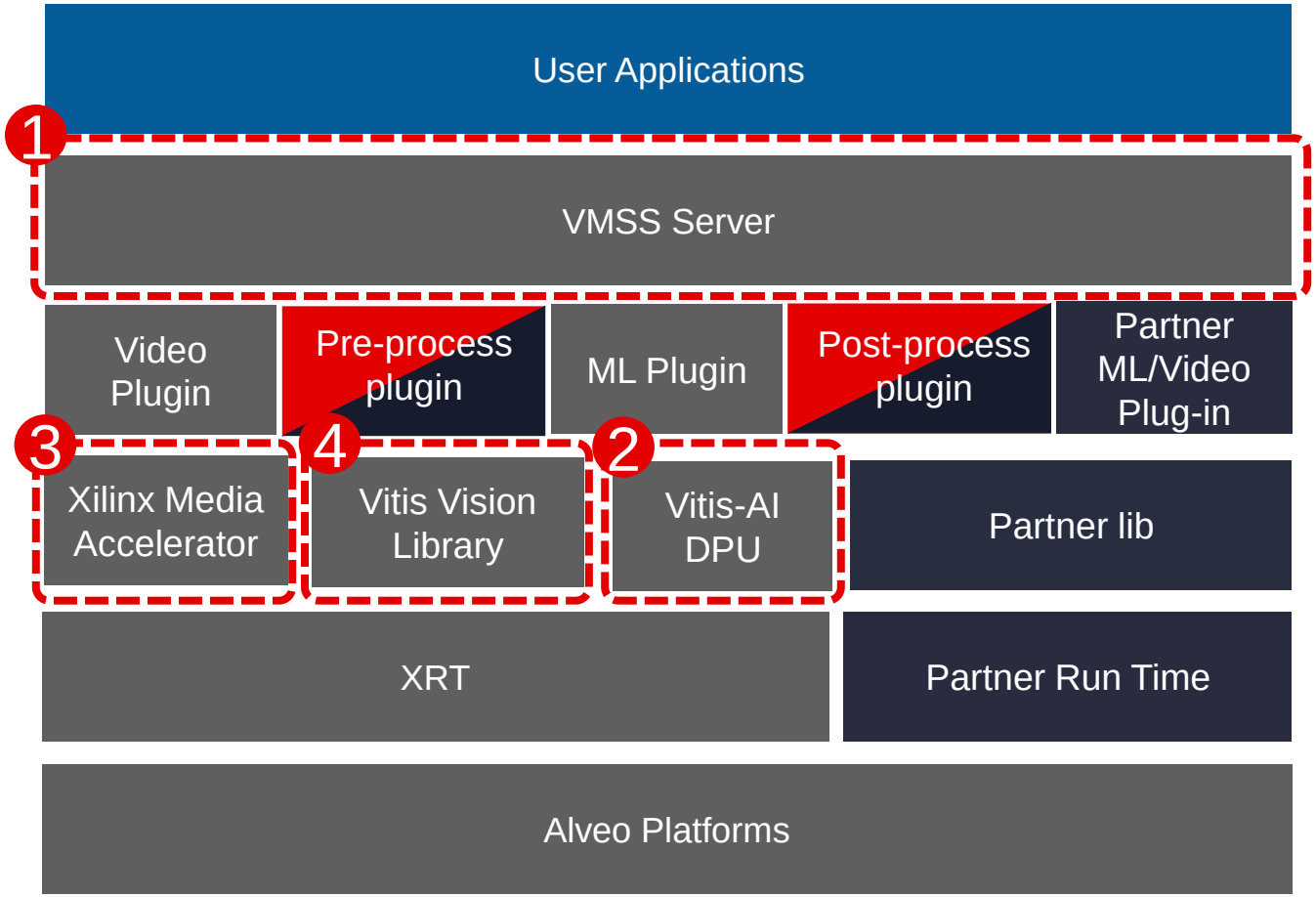


Video Machine Learning Streaming Server (VMSS)

XILINX

Xilinx Video Analytics Development Environment





XILINX

“ML + Video” = VMSS

1 VMSS Server

VMSS (Video Machine Learning Streaming Server) is a **development software framework** designed to ingest and process image data to produce ML inference results utilizing FPGA hardware acceleration

Pre-Processing

- **Video Preprocessing**
  - Image decode and preprocessing
  - Video decode and scaling
  - Crop and buffering

Data Processing

- **Workloads**
  - Machine learning
  - Other compute acceleration
  - Additional ML inferencing

Post-Processing

- **Conversion of Results for Consumption**

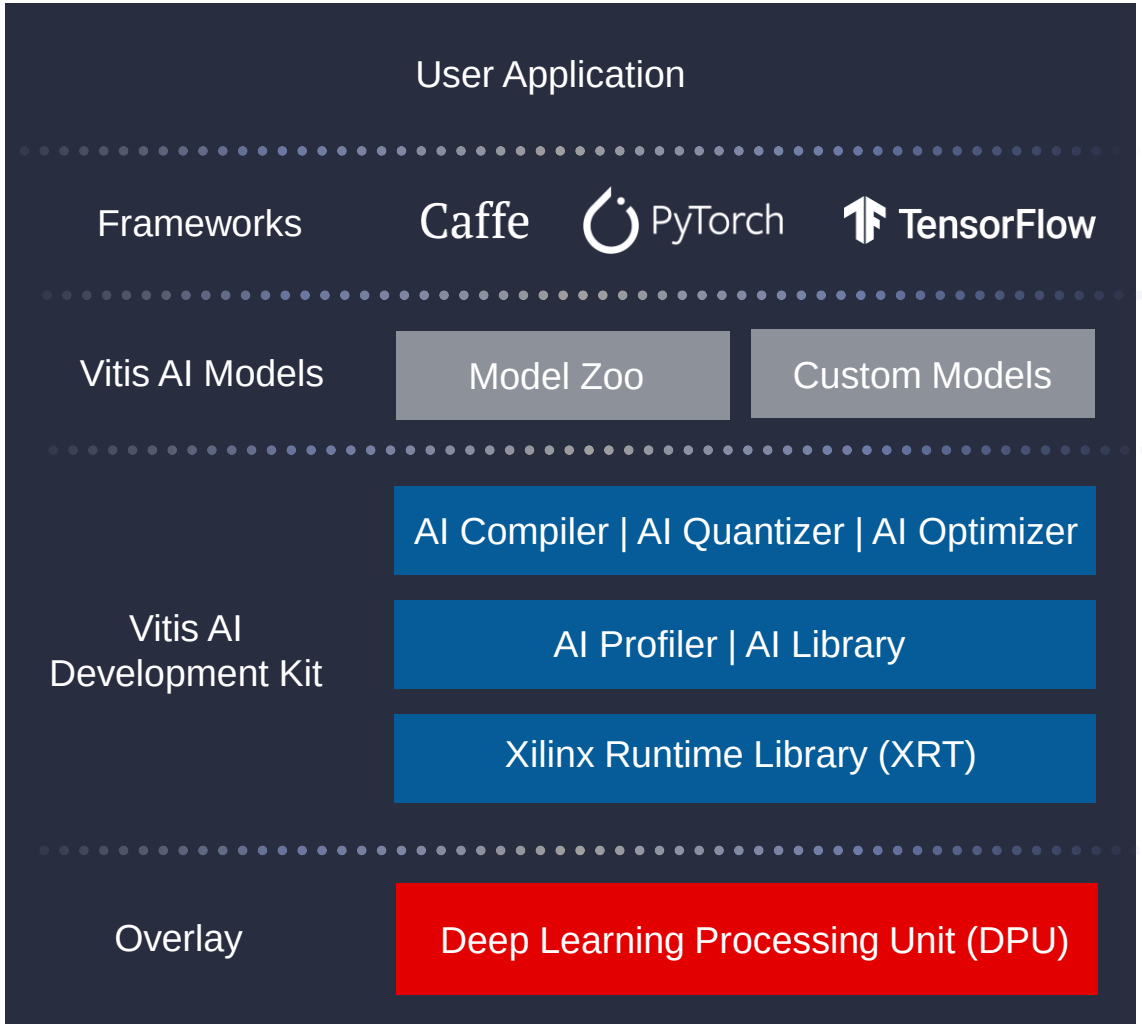
Custom App

- **Database record**
- **Data Storage:**
  - Acceleration on Data Access Functionalities

Xilinx Video Machine Learning Streaming Server (VMSS)

Xilinx Vitis AI

2 Vitis AI DPU

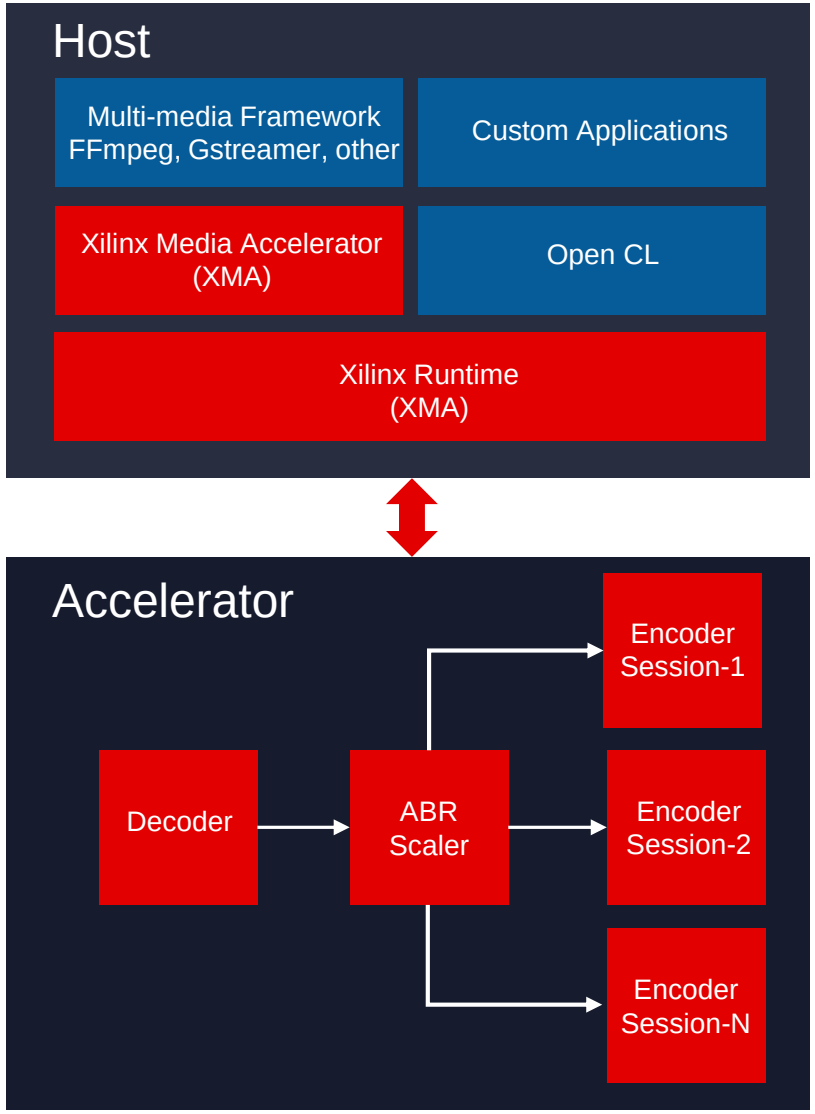


- Supports mainstream frameworks and the latest models capable of diverse deep learning tasks
- Provides a comprehensive set of **pre-optimized models** that are ready to deploy on Xilinx devices. Customers can find the closest model and start re-training for your applications
- Provides a powerful open source **quantizer** that supports pruned and unpruned model quantization, calibration, and fine tuning.
- The **AI profiler** provides layer by layer analysis to help with bottlenecks
- The AI library offers open source high-level C++ and Python APIs for maximum portability from edge to cloud.

XILINX

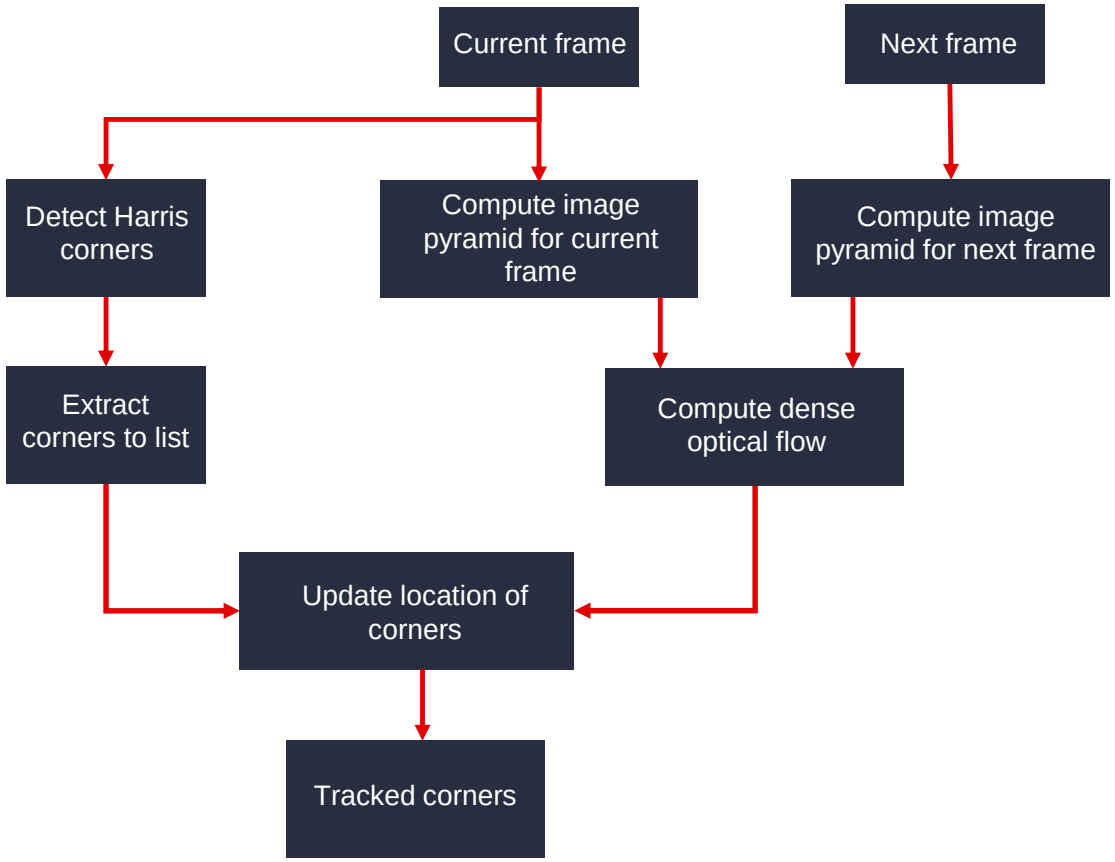
Xilinx XMA (Xilinx Media Accelerator)

3 Xilinx Media Accelerator



- Xilinx Media Accelerator (XMA) library (libxmaapi) is a host interface meant to simplify the development of applications managing and controlling **video accelerators such as decoders, scalers, filters, and encoders.**
- The libxmaapi is comprised of two API interfaces:
  - an application interface and a plugin interface.
- From a high-level perspective, the XMA sits between a media framework (i.e. FFmpeg) and the Xilinx runtime (XRT)

Xilinx Vision Library

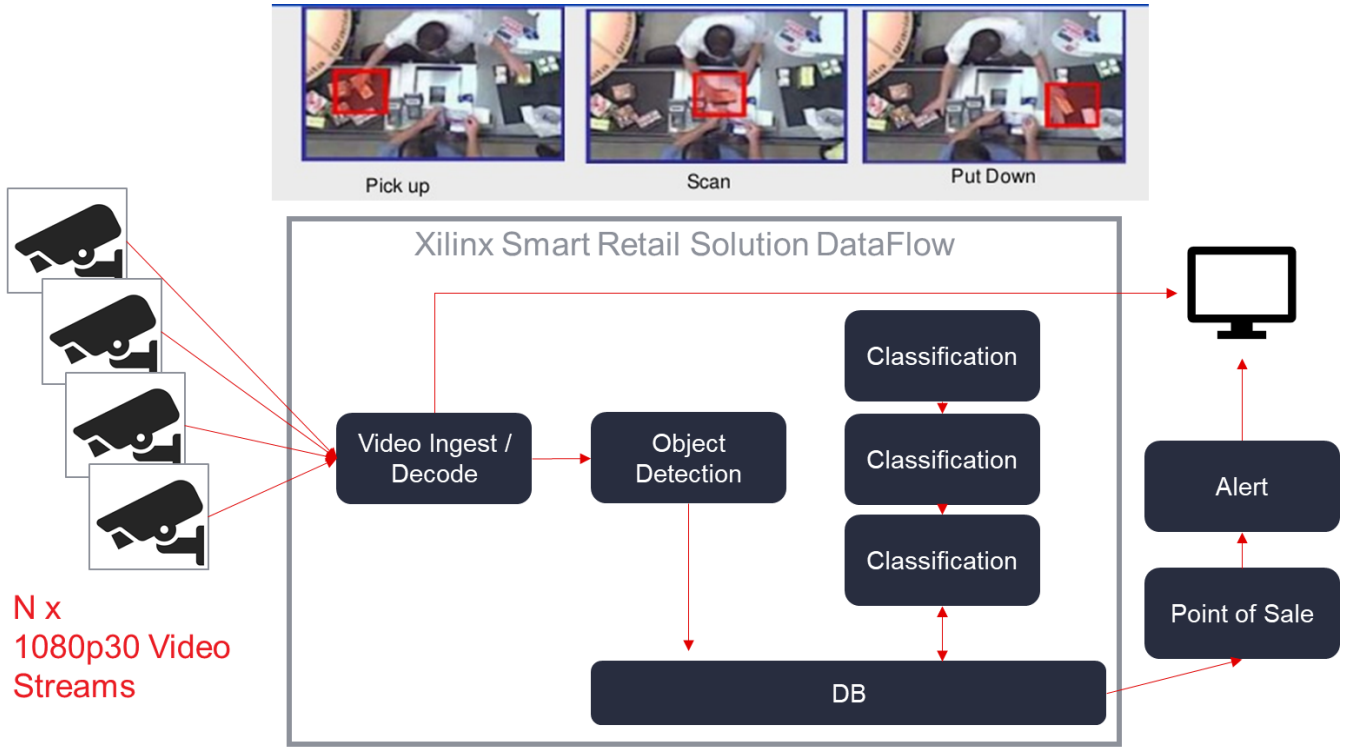


Corner tracking example

- All Vitis Vision functions are tested against OpenCV version 3.4.2.
- Performance-optimized kernel and primitive functions for:
  - Color and bit-depth conversion, Channel Extractions, Pixel-wise Arithmetic Operations
  - Geometric Transforms, Image Statistics, Filters
  - Feature Detection and Classifiers
  - 3D Reconstruction
  - Motion Analysis and Tracking
- Support for color image processing and multi-channel support

Use Case 1: Smart Retail Solution

- Retail loss from theft reaching $61.7B in 2019
  - “Retail shrink totaled $61.7 billion in 2019 amid rising employee theft and shoplifting/ORC” To fight losses, retailers started using point-of-sale (PoS) analytics, security cameras, wired alarms
- Smart Retail Features:
  - Recognize the specific product (Millions of products)
  - Low latency object classification and detection. Typical e2e latency is 100msec. Multiple classification results required to raise any alarm
- Smart Retail Use Cases:
  - Mis-scan: no item scanned, but objects detected by the ML
  - Ticket swapping: Item scanned doesn’t match with ML object classification results

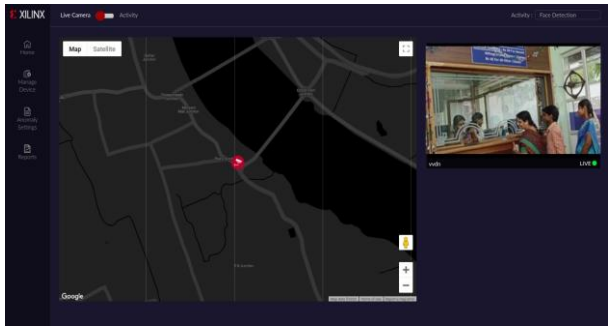


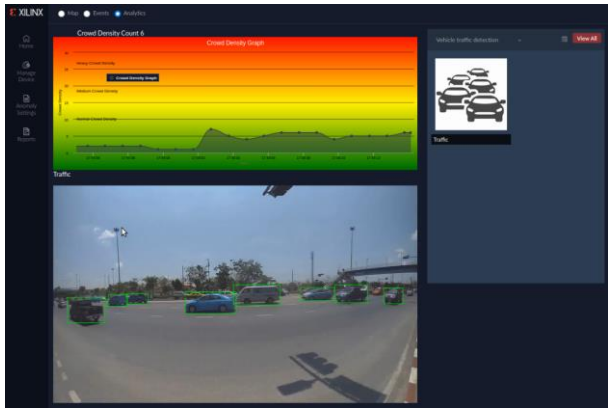
Source: https://nrf.com/media-center/press-releases/retail-shrink-totaled-617-billion-2019-amid-rising-employee-theft-and

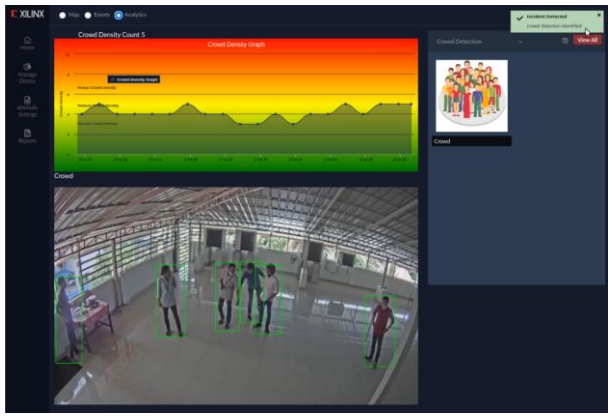
Use Case 2: Smart City Solution

- Features
  - Support multiple cameras
  - Event alert based on machine learning classification
  - Supports video recording and database recording
  - Utilizing Xilinx Video Machine Learning Streaming Server (VMSS) framework
  - Low latency event detection based on non batch ML inferencing
  - Custom plug-in supported
  - PoC reference source code can be provided for a quick TTM

- Details
  - Software Design Document (SDD)
  - GStreamer application development
  - Model creation, Model integration and support
  - Socket based application
  - Web GUI design with HTTP server support
- Modules
  - Activity Manager
  - Face recognition - Surveillance Security
  - Face mask detection
  - Crowd detection
  - Vehicle license plate detection
  - Vehicle traffic monitoring

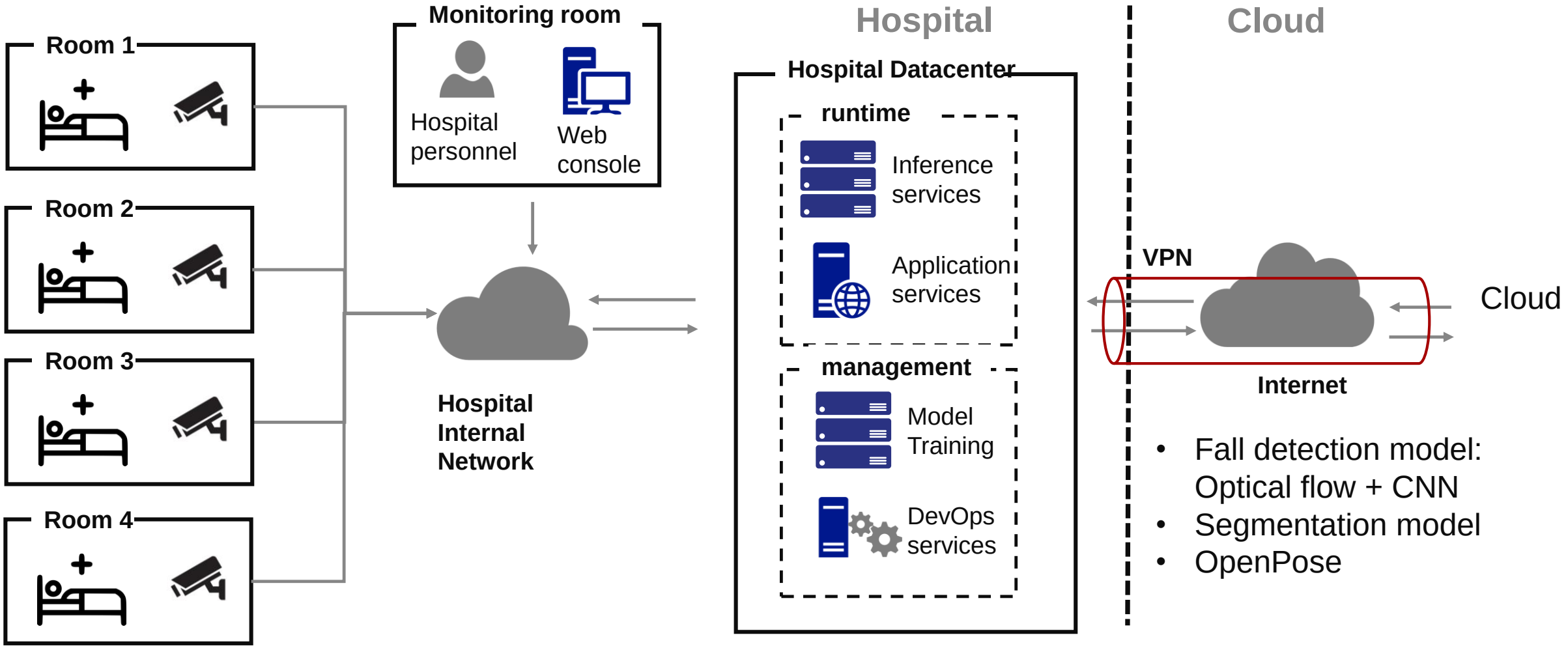






Use Case 3: Smart Hospital Solution

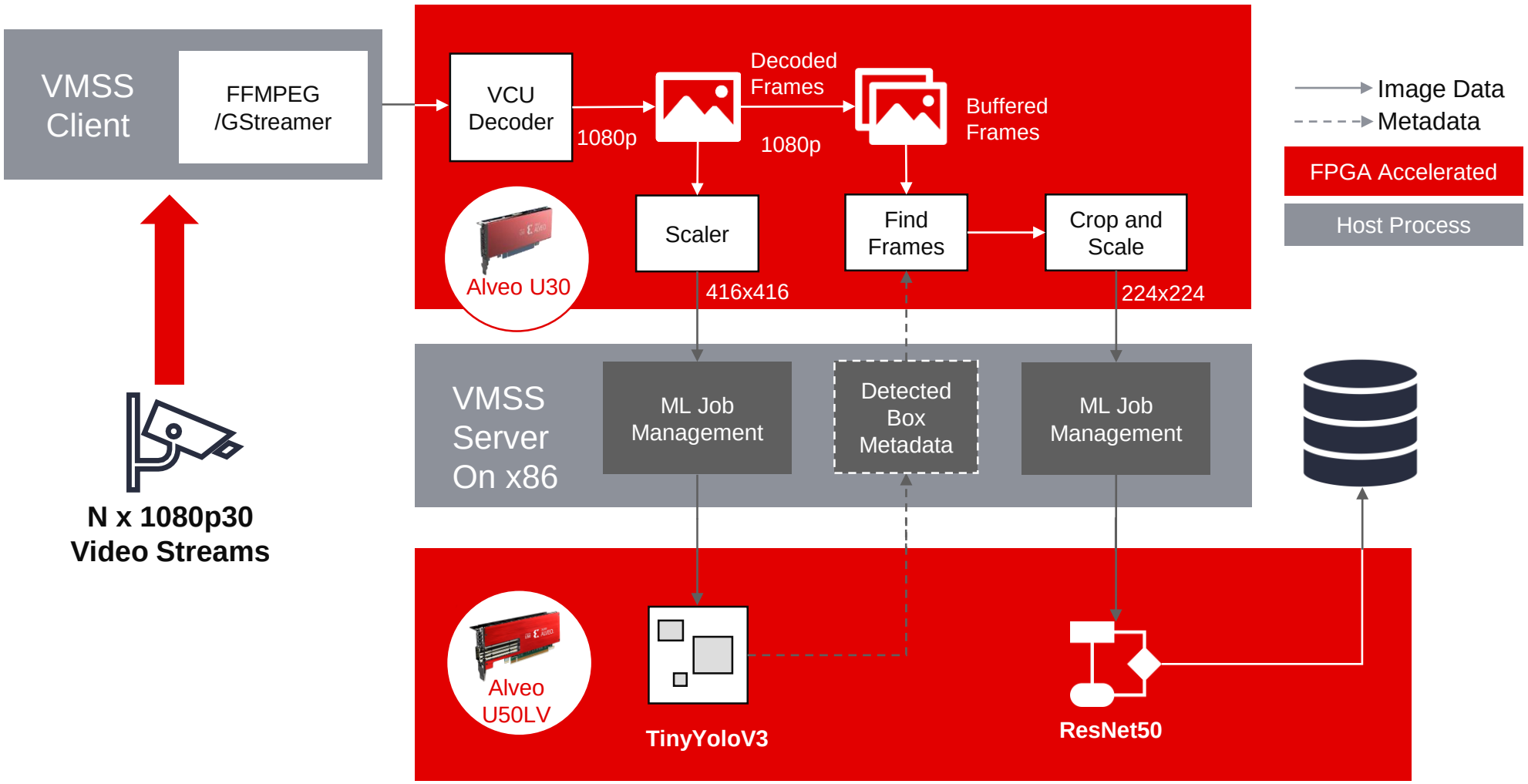
- Virtual Patient Observation (VPO)



- Fall detection model: Optical flow + CNN
- Segmentation model
- OpenPose

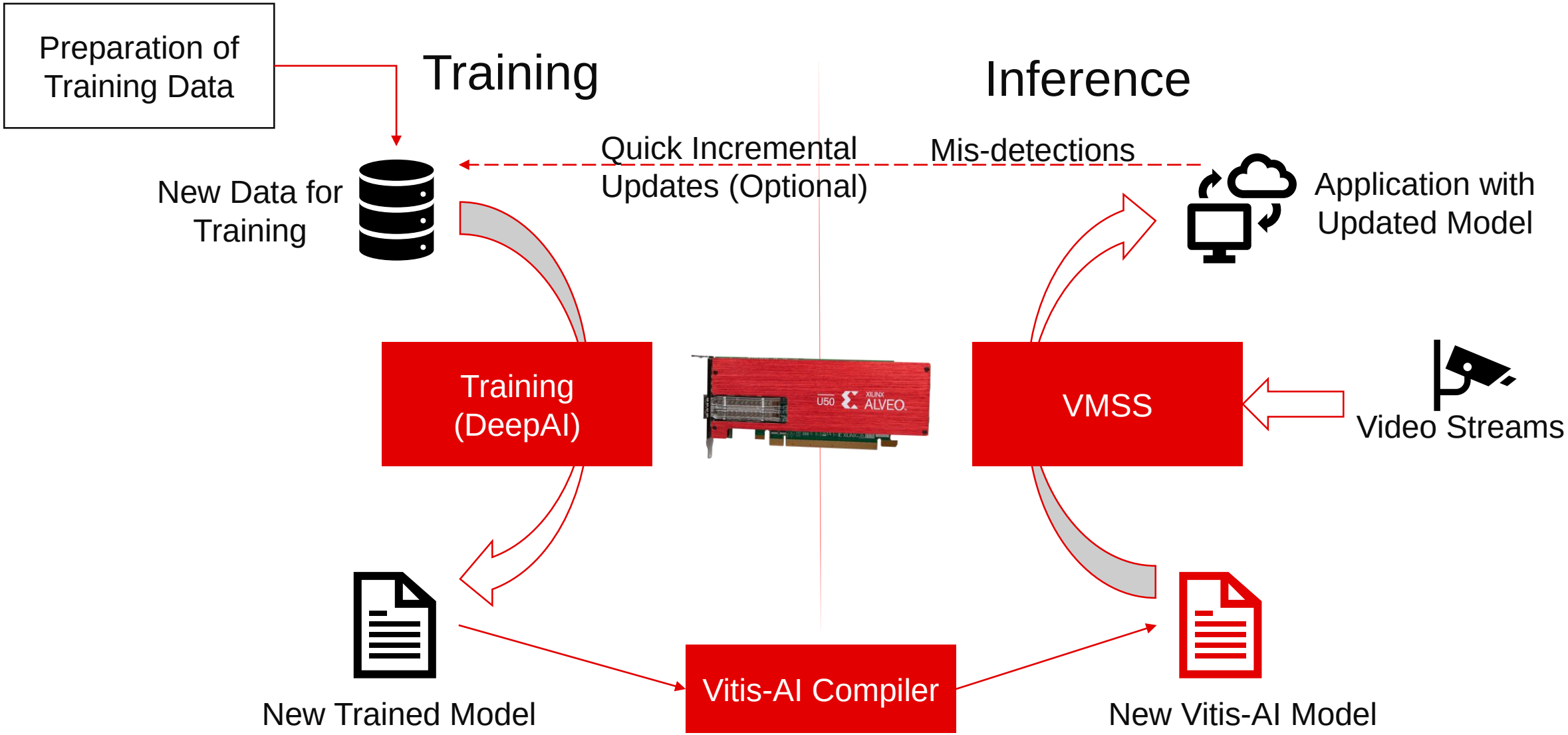
XILINX

U30 + U50LV Dataflow: How to Enable Best Performance



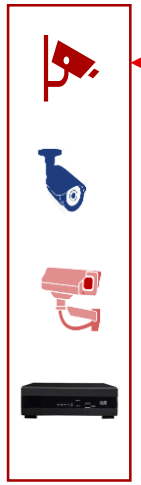
XILINX.

Training at the Edge



XILINX

Appliances – Edge AI Solution



Edge deployment for
video streaming & analytics

Video Gateway

Camera Management ONVIF
Streaming
RTMP/RTSP/UDP/TCP
H.264/265 decoding/encoding
Cloud platform interfacing
Cluster management

Aupera
Making Video Alive

AI Acceleration

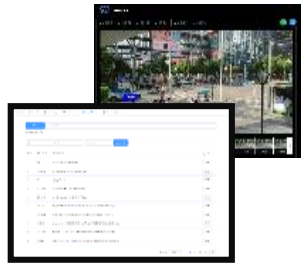
Full pipeline AI Solution for:
Smart
Building/City/Retail/Construction

Linux

Xilinx SoC Arm + FPGA

Analytics Output

VMS
(easy integrate with 3rd party)



Display dashboard

Restful API/Push API

IoT Cloud

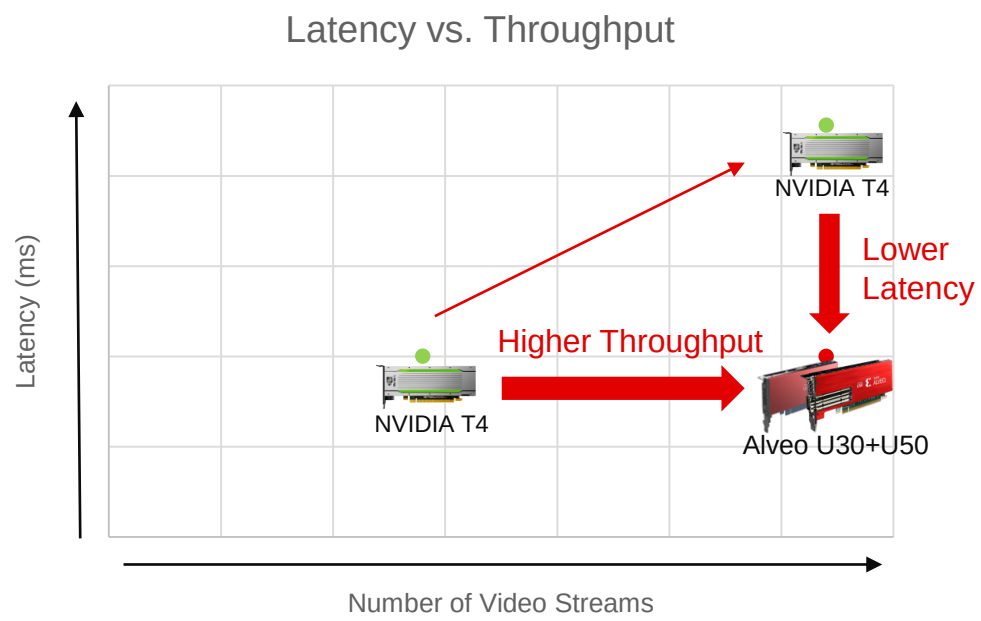
Tencent Cloud aws Azure ...

Industry Solutions:

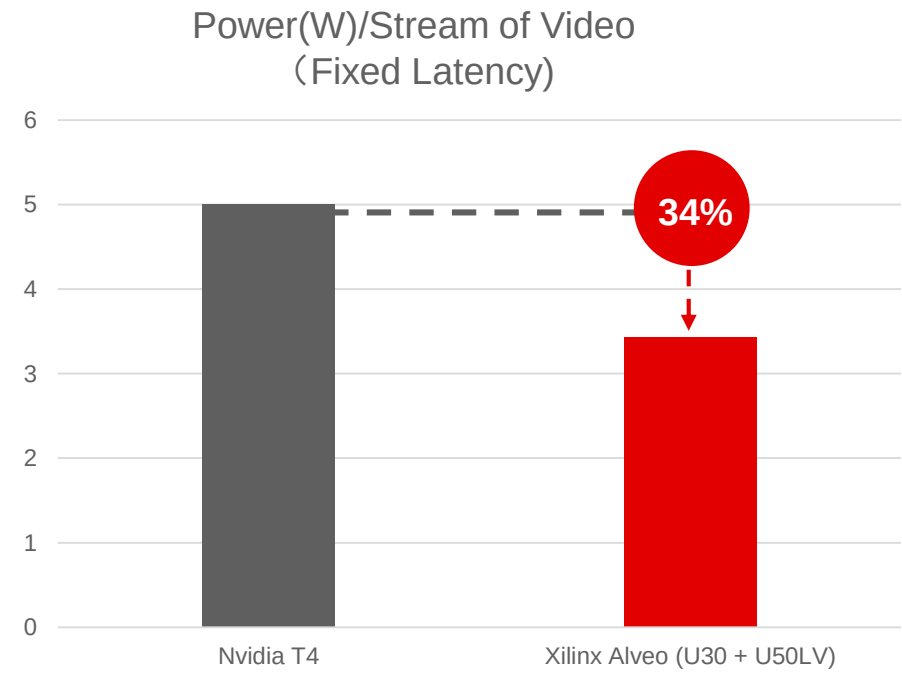
Smart Building | Smart City | Smart Retail | Smart Construction

High Throughput Real-Time Processing

Higher Performance



Lower Power Requirements



XILINX

What are the Differentiators?

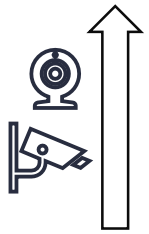
1) Optimal use of resources (independently scalable workloads)

2) Customizable, and low latency Higher performance Smart Video processing and Data Pre-processing

4) Custom plug-ins for HW Acceleration and SW Processing

Alveo U30

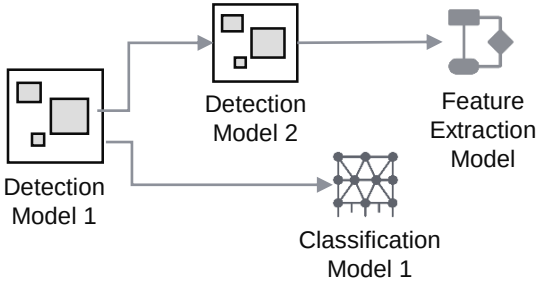
Video Decoder → Image/Data Pre-Processing



3) Multiple NN running Dynamically, asynchronously, simultaneously without performance penalty

DPU Inference Engine

Day



Alveo U50LV

Night

Post-Processing /Database Plugins

DB

Streaming

User Applications

5) Design for Exchangeability (DFX)
Day-time: inferencing
Night-time: (Re)Training, Video Encoding, DB

XILINX

Key Xilinx Advantages – Future Smart World Solution

Adaptability

The Xilinx platform allows to take any video and ML technology, including new codecs, new camera streaming protocols, and new Neural Networks including specialized NNs with custom layers.

Agility

The Xilinx platform makes customer-performed in-field updates quickly and efficiently.

Scalability

The Xilinx platform is easy to scale. Growing needs for processing many more video streams requires hassle-free scalability to allow adding new cards and platforms *without imposing complexity and without* adding too much power.

Visit Our Partners In The ISV Showcase

Migration and Acceleration

Mipsology

- Toolset delivering easy migration of existing AI applications
- High-performance plug-and-play AI inference accelerator

Edge AI Training

DEEPAI

- AI training at the edge on FPGA with a 10x performance/cost advantage vs GPUs
- Support for Tensorflow, PyTorch and Keras.

Smart Retail and Smart City

Aupera
Making Video Alive

- Full video AI solutions at the edge
- Smart city, smart building, and smart retail
- High efficiency, low latency, and scalability

Xilinx Mission

Building the Adaptable, Intelligent World