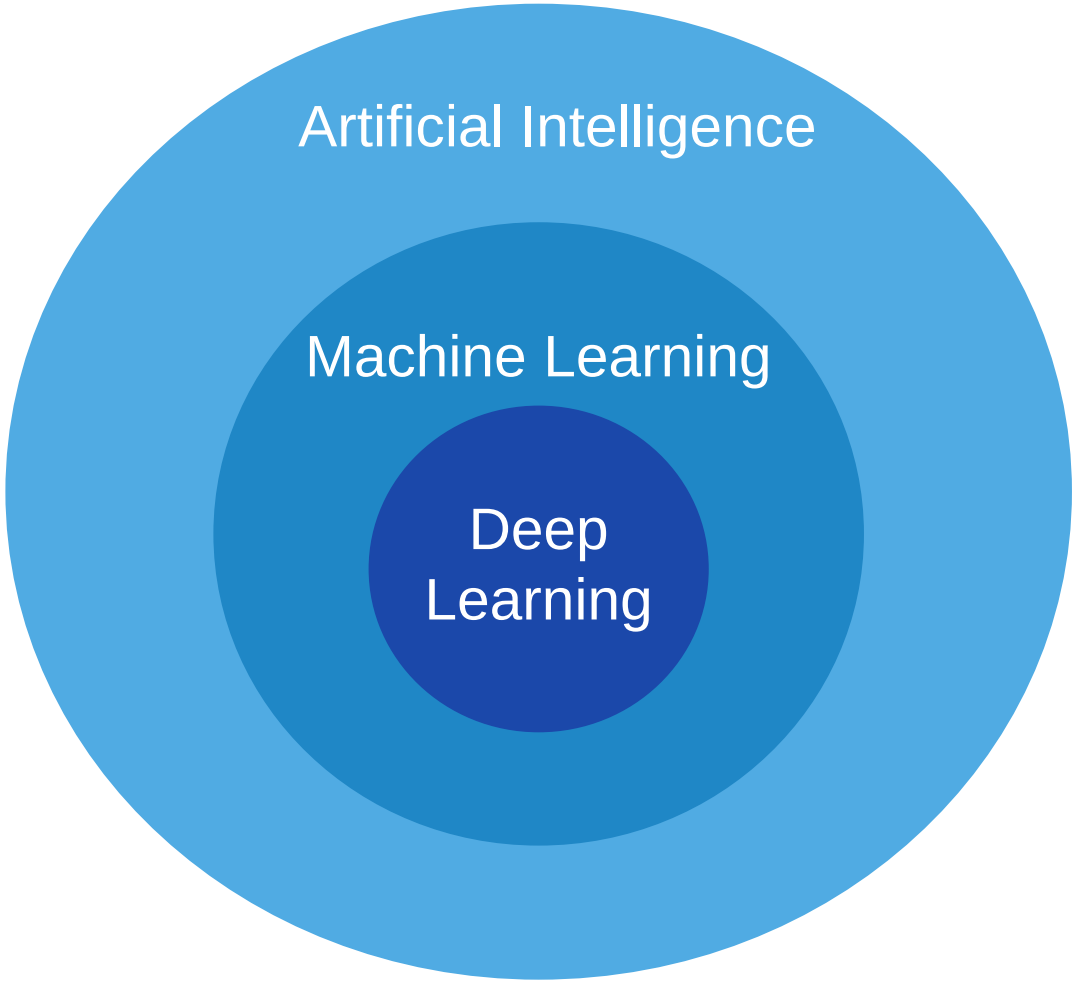
Machine Learning at the Edge with i.MX 8M Plus

Meng Ju Lin

SECURE CONNECTIONS FOR A SMARTER WORLD

Artificial Intelligence, Machine Learning and Deep Learning



Artificial Intelligence

- The very broad concept of using machines to do “smart” things and act intelligently **like a human**

Machine Learning

- One of many ways to implement AI
- The concept that if you give machines a lot of data, they can learn how to do smart things on their own, without having to be explicitly programmed to do that action.
- **Self learning and self improving**

Deep Learning

- One of many ways to implement machine learning (ML)
- Uses “**Neural Networks**” that can learn and make intelligent decisions on its own
- Needs a **lot** of data

Machine Learning @ Edge Encompasses Domains

Vision

- ADAS and Driver Monitoring
- Surveillance Systems for Security or Factory Monitoring
- Package Detection
- Assembly line visual defect recognition

Voice/Sound

- Keyword actions
- Voice commands
- Audio Alarm Analytics (Breaking Glass/Baby Crying)

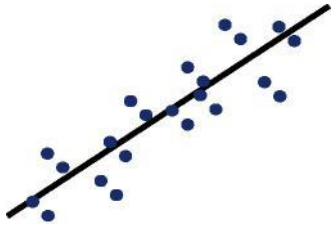
Anomaly Detection

- Agriculture and Industry Quality Control/Analytics
- Motor performance and analysis
- Smartwatch health monitoring

What Can Machine Learning Do

Regression (Calculation)

- Predict continuous values



X=a, y=?

Classification (Choice)

- Recognition, object detection

- It is a ( )
- A: Dog B: Cat C: Cow *D: Neither*

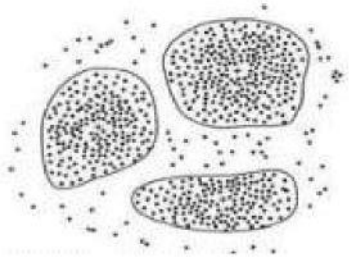
Anomaly detection (Judgement)

- Detect abnormal conditions

- Heart is going to malfunction? Y/N

Clustering

- Discover patterns / partitions



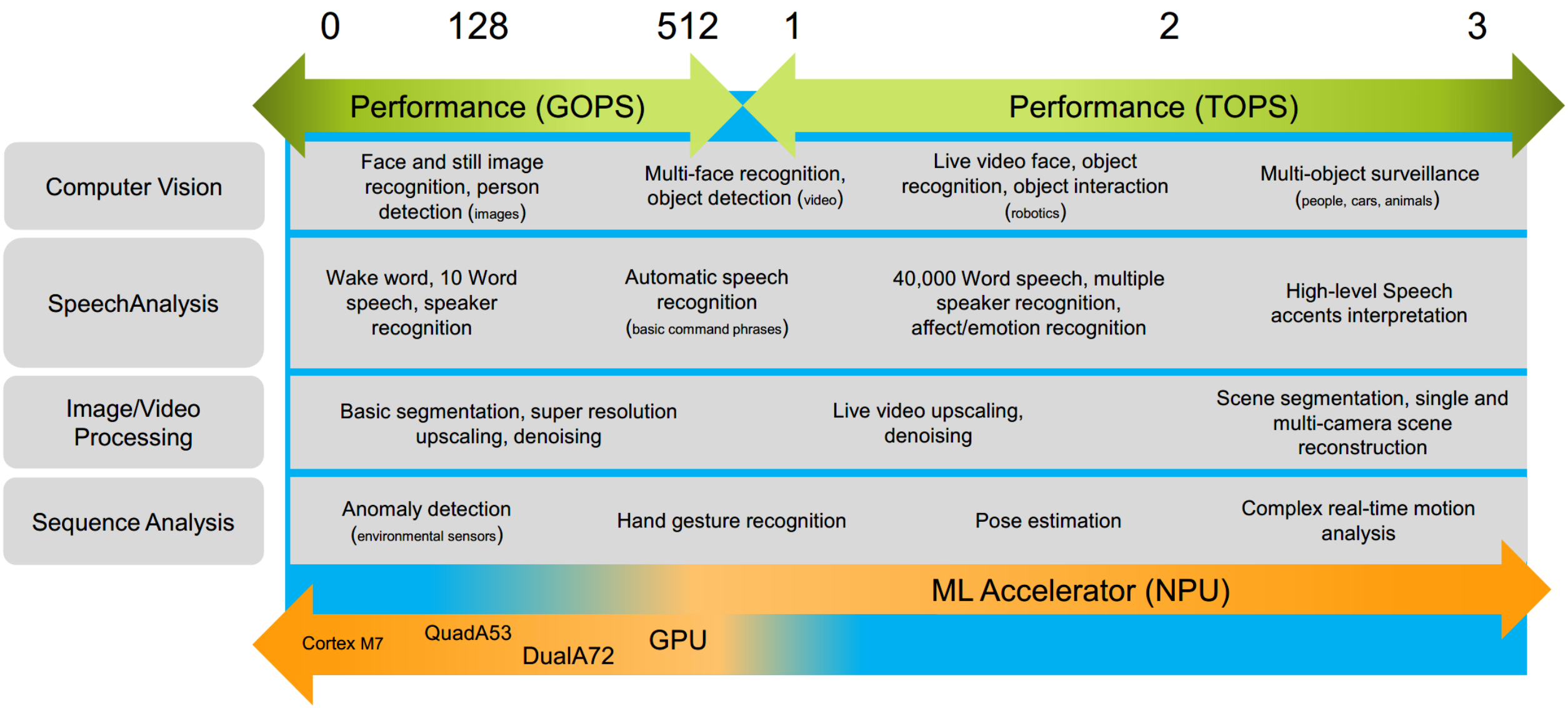
- Find crowds
- No need labels

Learn strategies

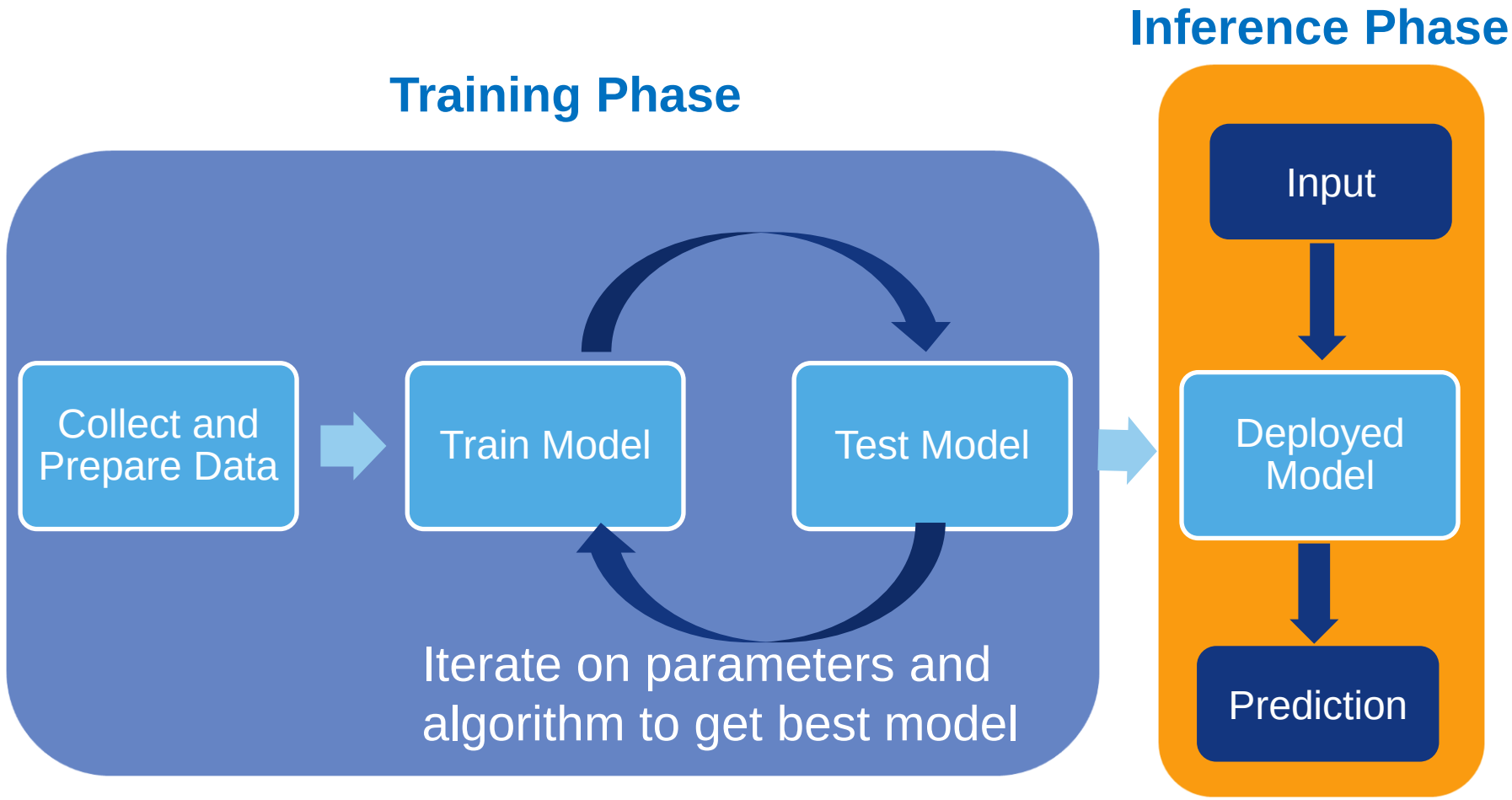
- Reinforcement Learning

- How to play the game?

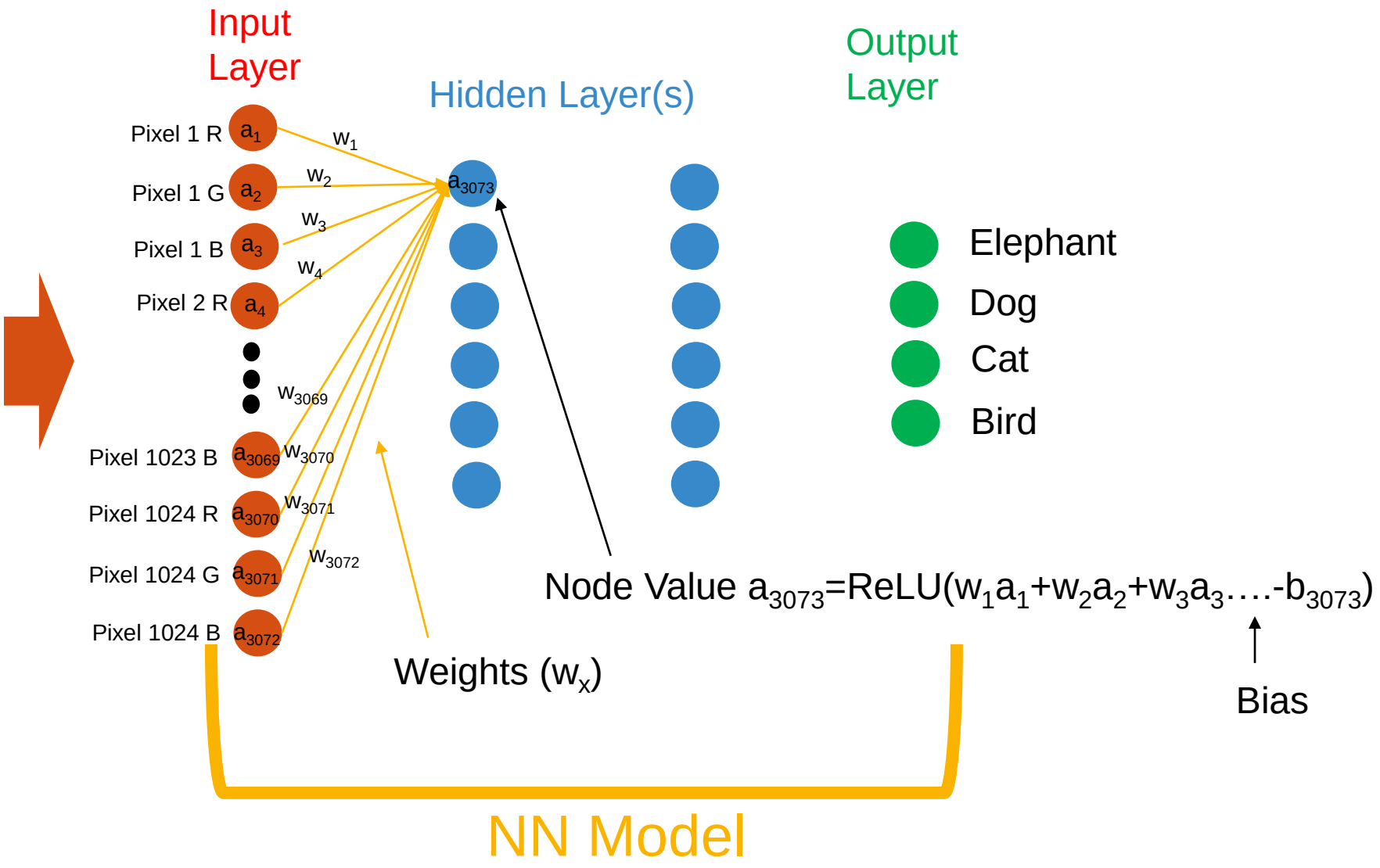
Machine Learning Use Cases



Machine Learning Process

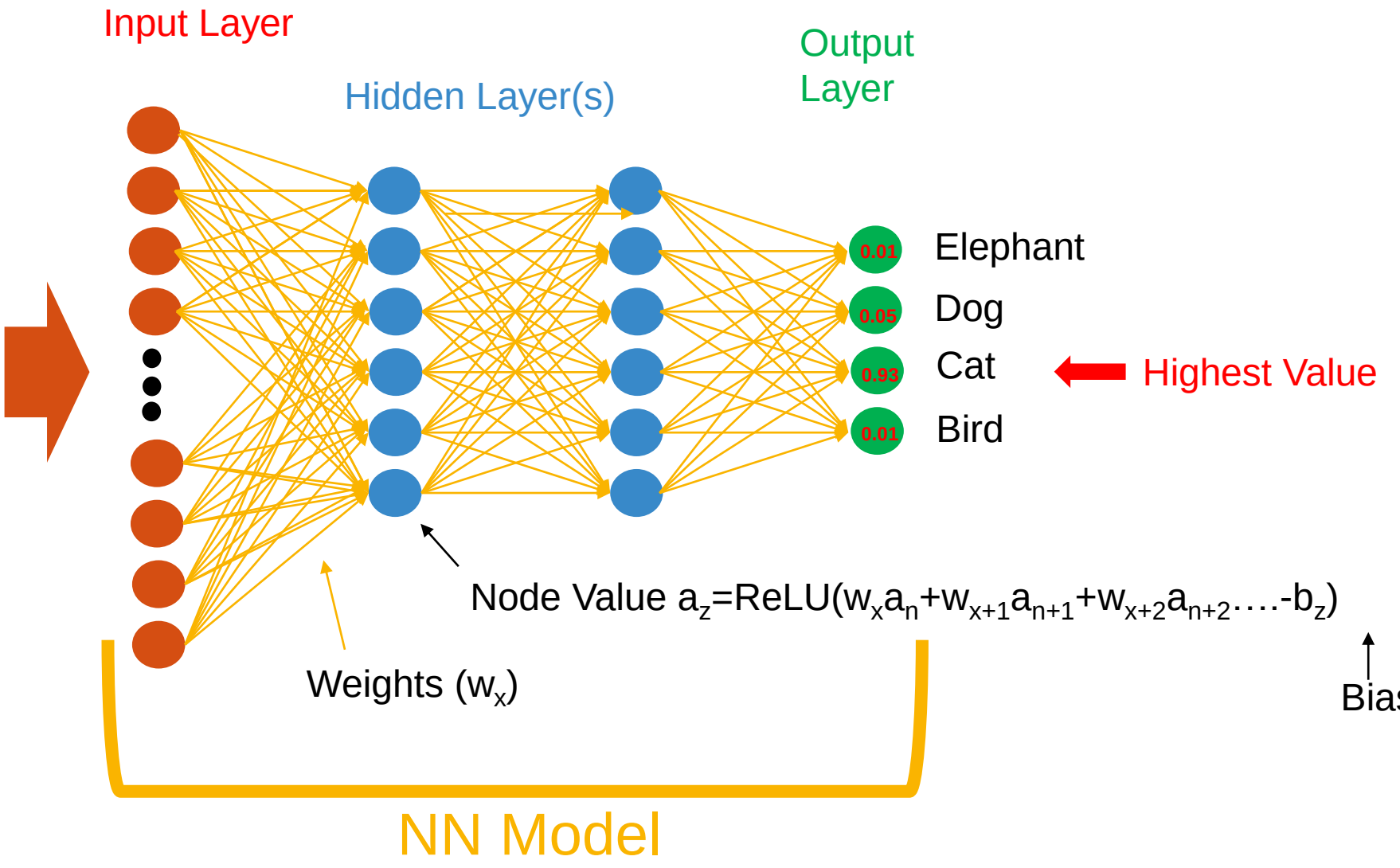


Very Simplified Neural Network Model

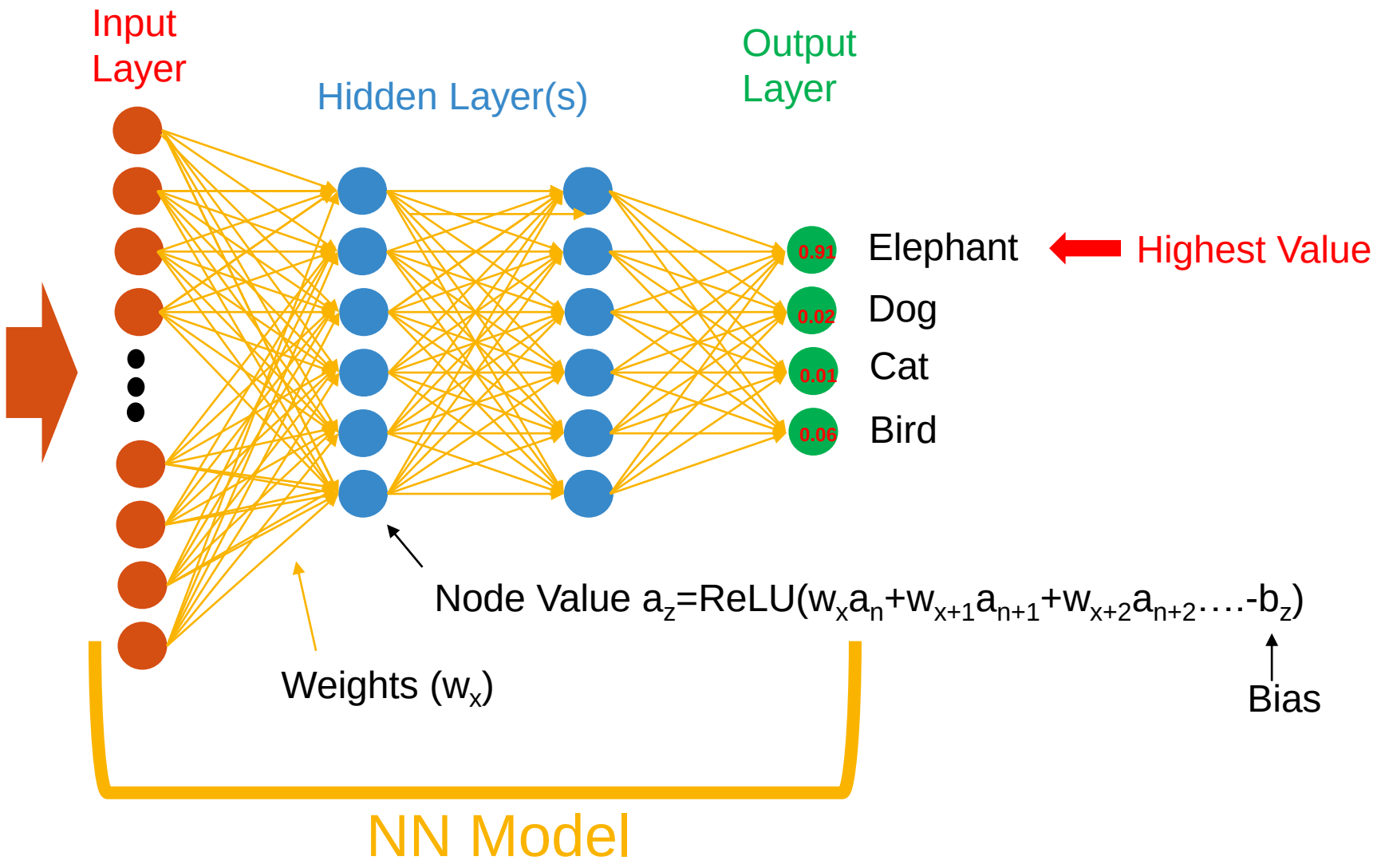


For 32x32 pixel 3 color image, there are 32*32*3=3072 input nodes.

Very Simplified Neural Network Model



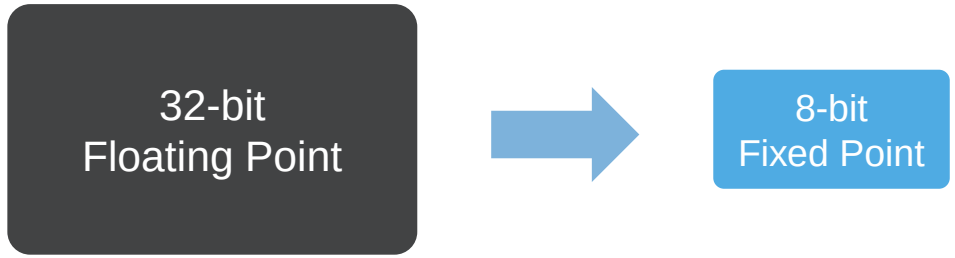
Very Simplified Neural Network Model



Quantization and Pruning

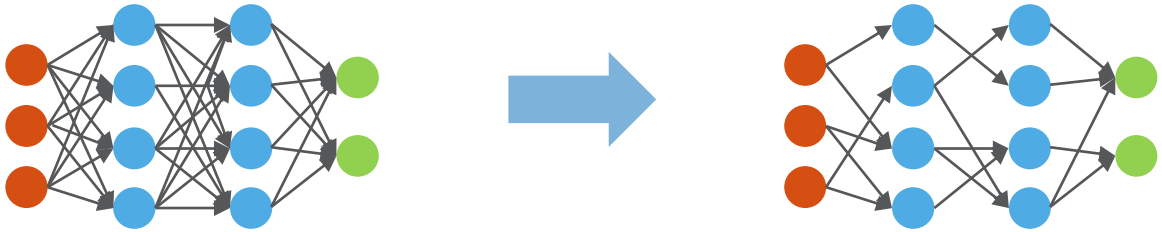
Quantization

- Transform 32-bit floating point weights → 8-bit fixed point weights
  - Reduces model size by 4x
  - Fixed point math quicker than floating point
  - Usually little loss of accuracy



Pruning

- Remove low importance weights and biases from a neural network
  - Recommended to retrain model after pruning



NXP Broad-based Machine Learning Solutions & Support

EIQ MACHINE LEARNING

eIQ™ ML Enablement

eIQ (edge intelligence) for edge AI/ML inference enablement

Open source technologies (TensorFlow Lite, Arm NN, Glow, ONNX)

Support for i.MX 8 family, i.MX RT

Integrated into NXP development environments (MCUXpresso, Yocto/Linux)

DIY

EIQ AUTO AI

eIQ™ Auto AI Enablement

Deep Learning toolkit for S32V23x processors

Optimization: Prunes, quantizes, compresses the Neural Network

Automated neural net layer deployment to optimum available compute resource

Auto Quality Inference Engine: A-SPICE qualified inference engine

Automotive Grade

CORAL

Third Party SW and HW

Google Coral Dev Board

i.MX 8M Mini Development Kit for Amazon® Alexa Voice Service

Au-Zone Network Development Tools

Arcturus video applications

SensiML tools for sensor analysis

…. And more

SLN-ALEXA-IOT

Turnkey Solutions

Alexa Voice Services (AVS) solution

- i.MX RT106A (kit – SLN-ALEXA-IOT)

Local voice control solution

- i.MX RT106L (kit – SLN-LOCAL-IOT)

Face & emotion recognition solution

- i.MX RT106F (kit – SLN-VIZN-IOT)

Fully Tested

i.MX 8M Plus machine learning compute engines

Machine Learning Accelerator (1GHz)

- Primary Use: Multi-camera classification/detection

Quad Arm® Cortex-A53 (1.8GHz)

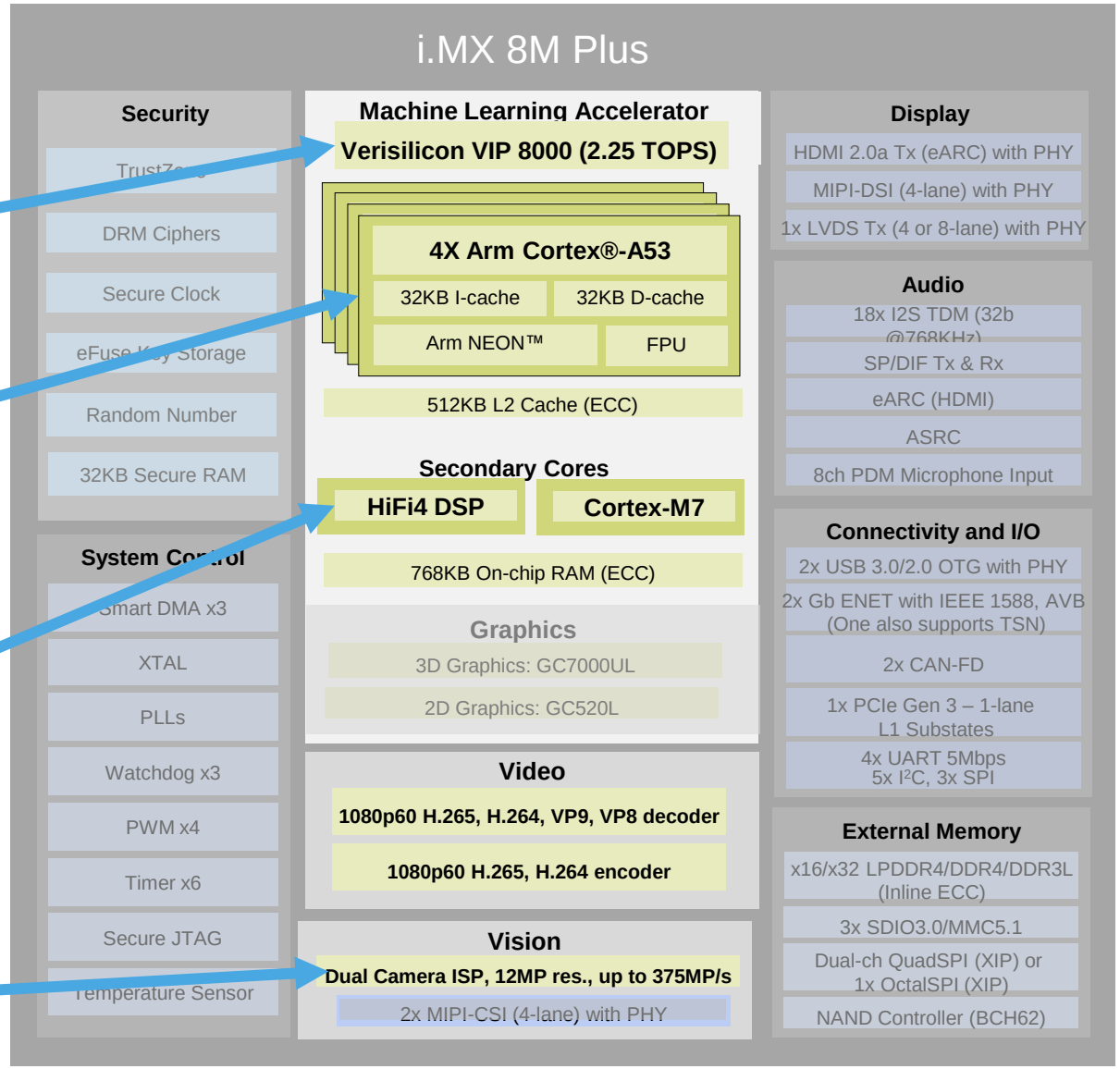
- Primary Use: Speech command recognition, object detect/classification

Cortex-M7+HiFi4 DSP (800MHz)

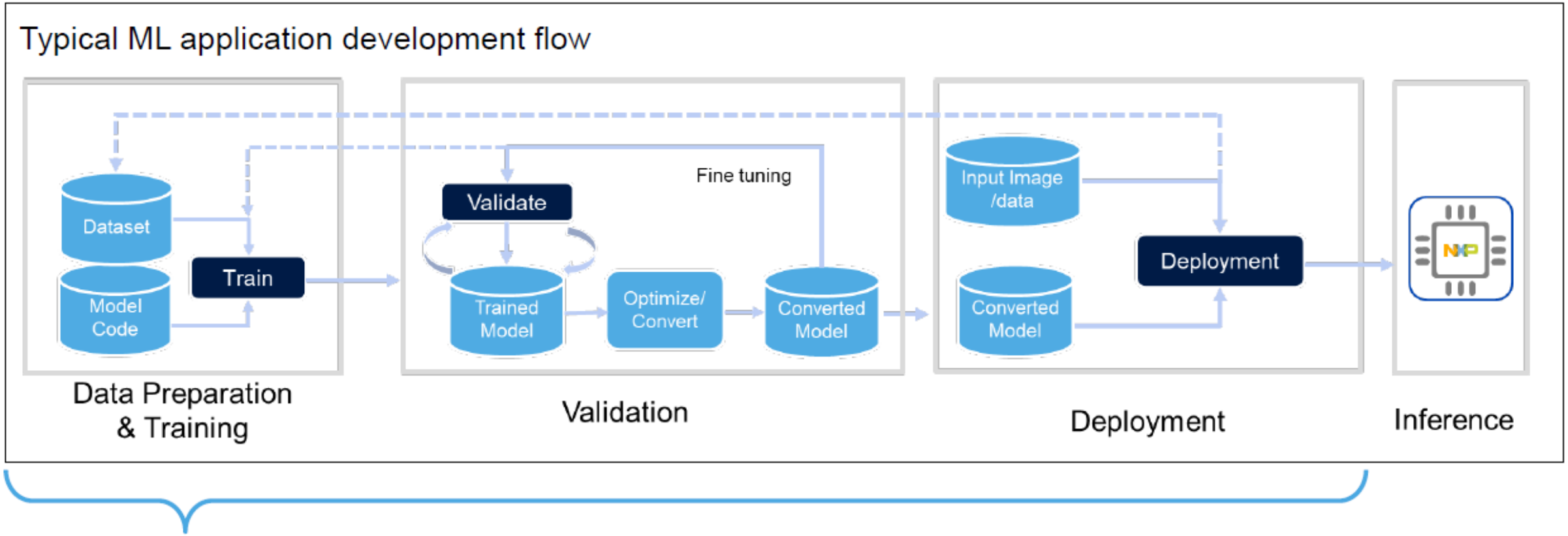
- Primary Use: Keyword detection, sensor fusion

2 channel Image Signal Processor (ISP)

- Primary Use: Scaling, de-warping, image enhancement



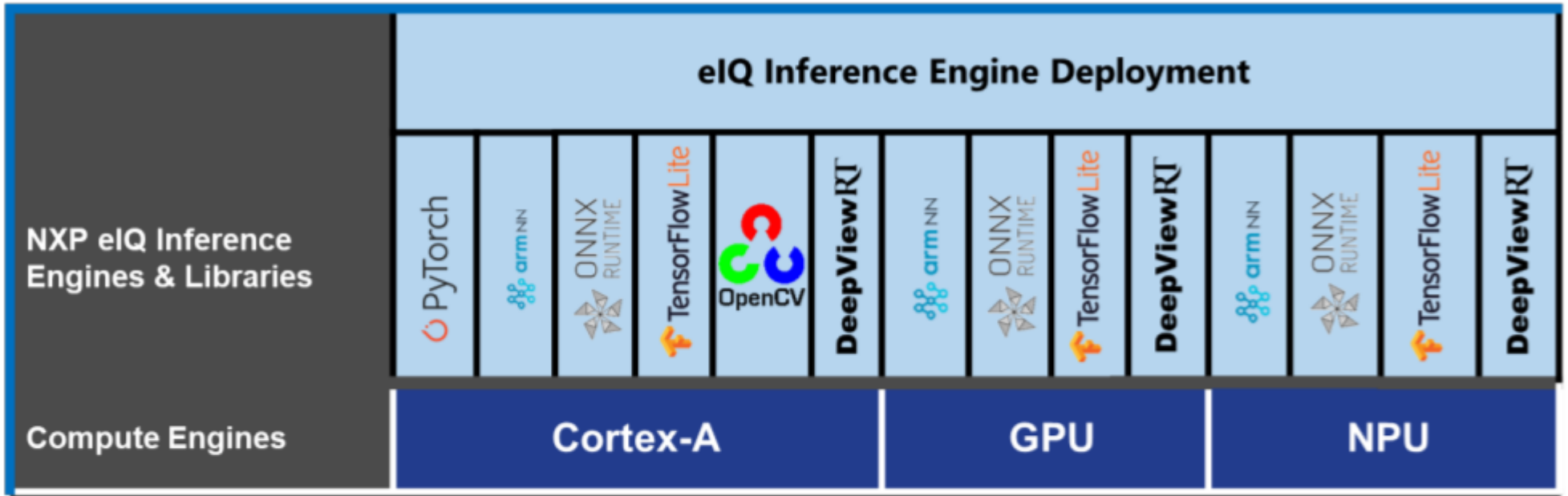
eIQ™ MACHINE LEARNING SOFTWARE DEVELOPMENT ENVIRONMENT



NXP's **eIQ ML Software** provides a collection of development tools, utilities and libraries for building ML applications using NXP MCUs and applications processors (MPUs).

eIQ ML software can be leveraged as part of a user's existing flow or can be used for the complete flow depending on the ML application targeted.

NXP eIQ Supported Compute Engines vs. Inference Engines



NXP eIQ Inference Engines & Libraries	eIQ Inference Engine Deployment													
	PyTorch	arm NN	ONNX RUNTIME	TensorFlow Lite	OpenCV	DeepViewRT	arm NN	ONNX RUNTIME	TensorFlow Lite	DeepViewRT	arm NN	ONNX RUNTIME	TensorFlow Lite	DeepViewRT
Compute Engines	Cortex-A						GPU				NPU			
i.MX 8M Plus	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
i.MX 8QuadMax	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	NA	NA	NA	NA
i.MX 8QuadXPlus	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	NA	NA	NA	NA
i.MX 8M Quad, Nano	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	NA	NA	NA	NA
i.MX 8M Mini, 8ULP	✓	✓	✓	✓	✓	✓	NA	NA	NA	NA	NA	NA	NA	NA

✓ Supported NA (Not Applicable)

i.MX 8M Plus PyeIQ Demo

Pre-requisite

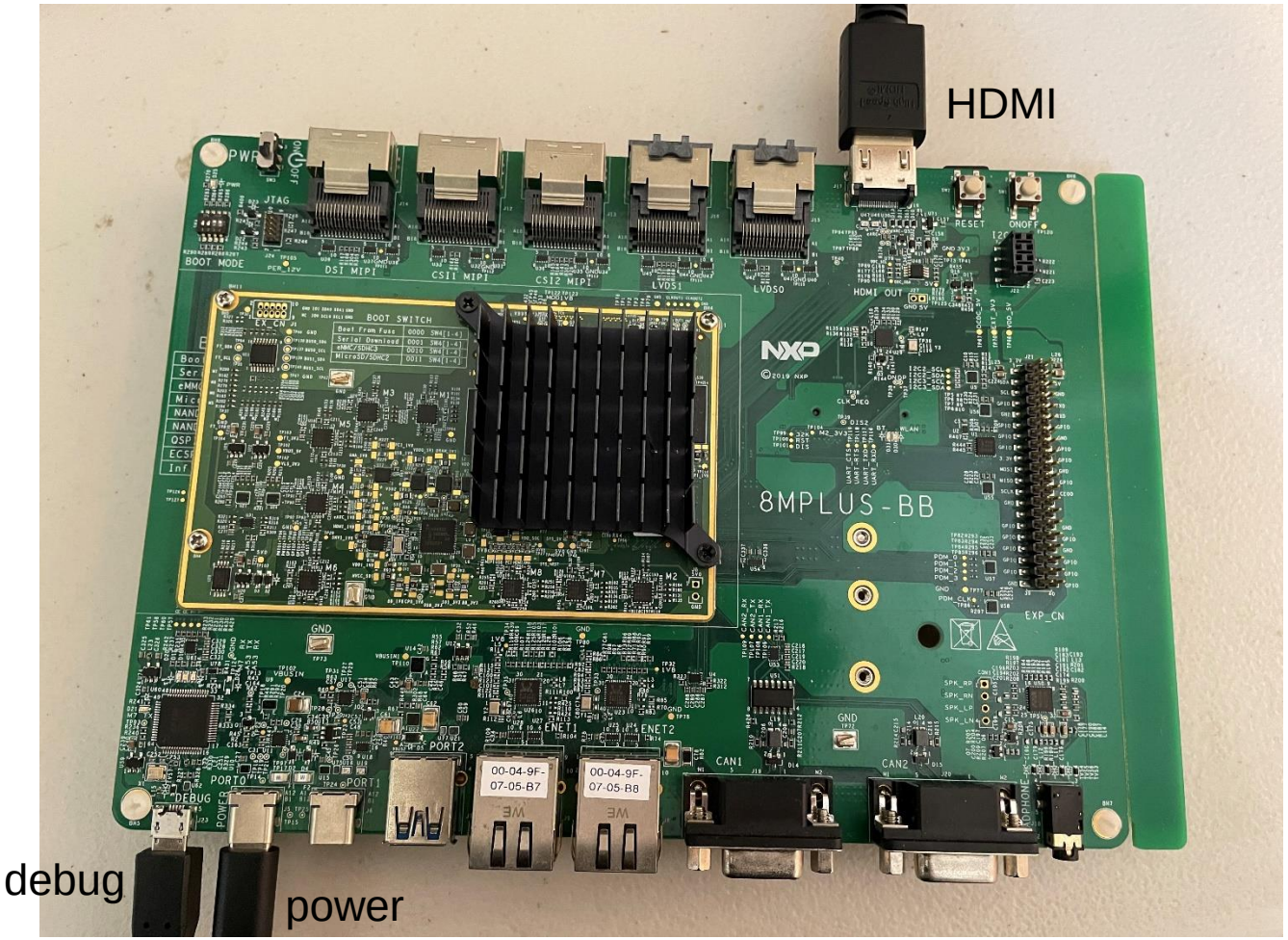
Hardware:

- i.MX8M Plus EVK
- HDMI port connects to monitor
- Debug port connects to PC
- And power port connects to power

Software:

- Install PyeIQ under debug terminal

```
#pip3 install pyeiq
```



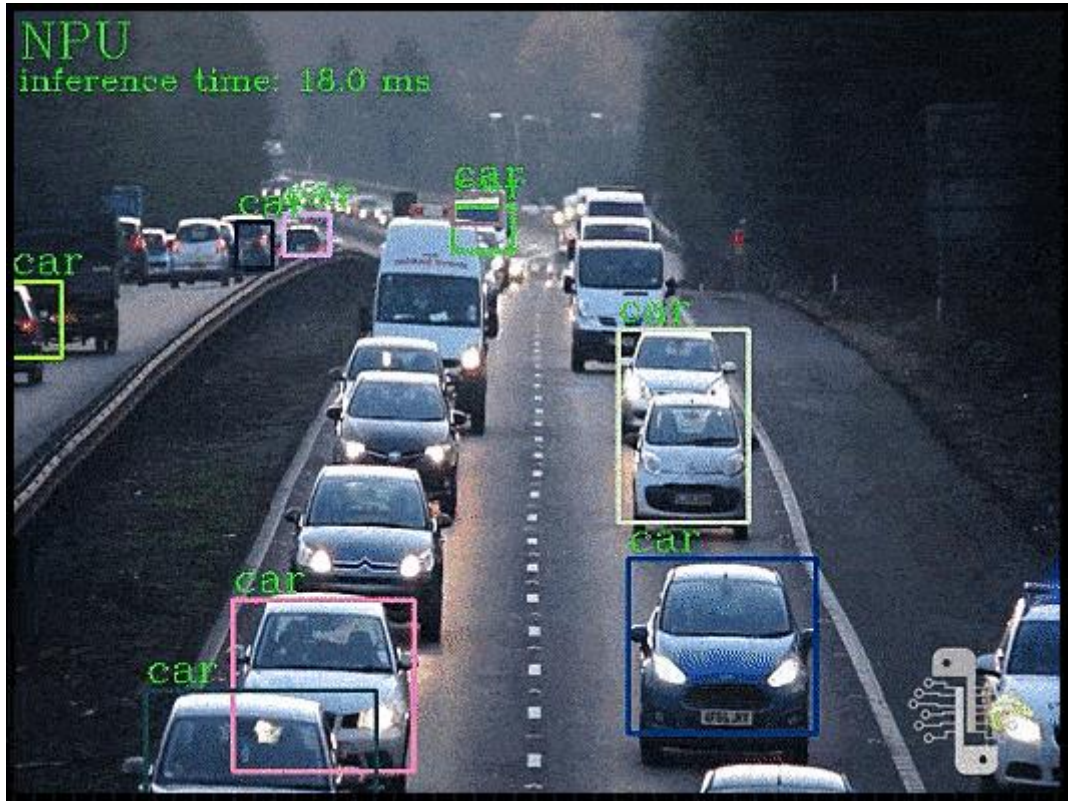
PyeIQ demo – Switch Detection Video

- This demo uses
  - Tensorflow Lite as an inference engine
  - Single Shot Detection as default algorithm

- Run Switch Detection Video Demo

  #pyeiq --run switch_video

  - Type CPU or NPU/GPU in the terminal to switch the compute engines.



PyeIQ demo – Object Detection (1/2)

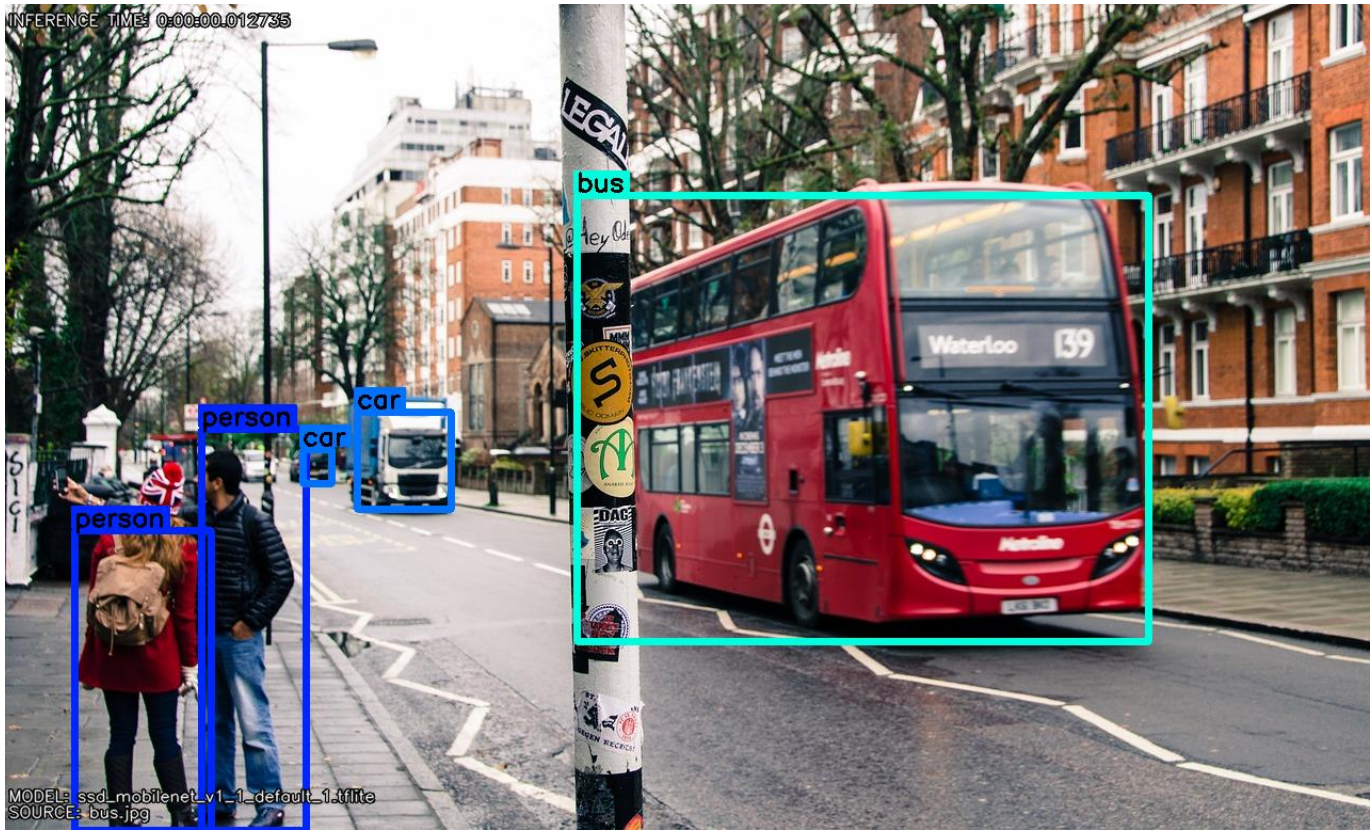
- This demo uses
  - Tensorflow Lite as an inference engine
  - Single Shot Detection as default algorithm

- Using image for inference

```
#pyeiq --run object_detection_tflite
```

or

```
#pyeiq --run object_detection_tflite --image=/path_to_the_image
```



PyeIQ demo – Object Detection (2/2)

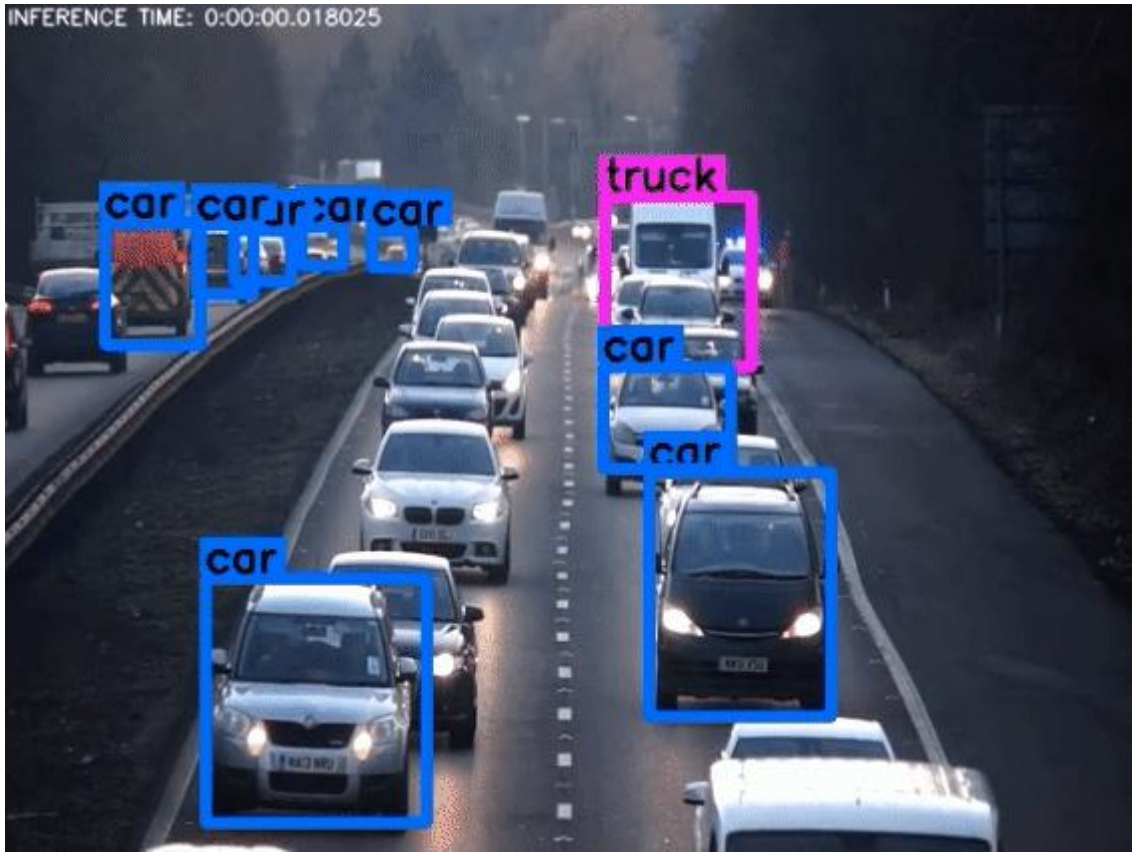
- This demo uses
  - Tensorflow Lite as an inference engine
  - Single Shot Detection as default algorithm

- Using video source for inference

```
#pyeiq --run object_detection_tflite --video_src=/path_to_the_video
```

- Using video camera or webcam for inference

```
#pyeiq --run object_detection_tflite --video_src=/dev/video<index>
```



PyeIQ demo – Object Classification(1/2)

- This demo uses
  - Tensorflow Lite as an inference engine
  - MobileNet as default algorithm

- Using image for inference

```
#pyeiq --run object_classification_tflite
```

or

```
#pyeiq --run object_classification_tflite --image=/path_to_the_image
```

PyeIQ demo – Object Classification (2/2)

- This demo uses
  - Tensorflow Lite as an inference engine
  - MobileNet as default algorithm

- Using video source for inference

```
#pyeiq --run object_classification_tflite --video_src=/path_to_the_video
```

- Using video camera or webcam for inference

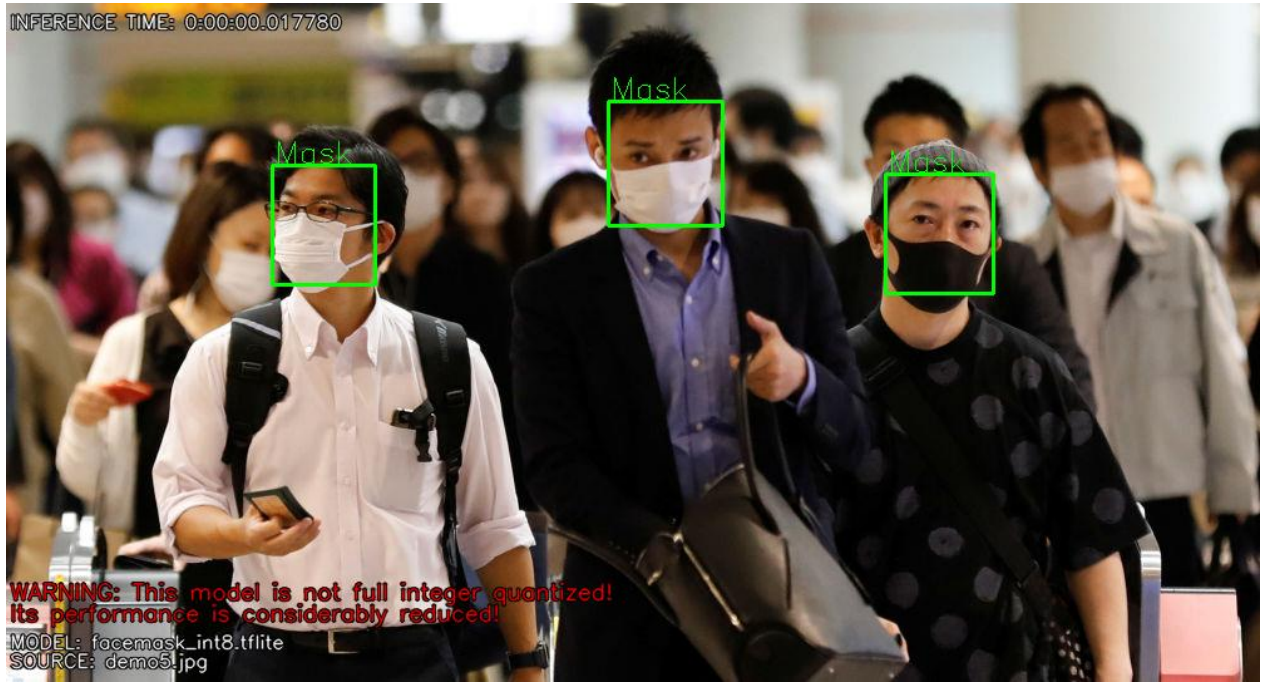
```
#pyeiq --run object_classification_tflite --video_src=/dev/video<index>
```

PyeIQ demo – Covid19 Detection

- This demo uses
  - Tensorflow Lite as an inference engine
  - Single Shot Detection as default algorithm
- Using image for inference

```
#pyeiq --run covid19_detection
```

```
#pyeiq --run covid19_detection --image=/path_to_the_image
```



- Using video source for inference

```
#pyeiq --run covid19_detection --video_src=/path_to_the_video
```

- Using video camera or webcam for inference

```
#pyeiq --run covid19_detection --video_src=/dev/video<index>
```

Reference

- AI and Machine Learning Training Academy

https://www.nxp.com/design/training/ai-and-machine-learning-training-academy:TS-MACHINE-LEARNING-AND-AI

- i.MX 8M Plus

https://www.nxp.com/products/processors-and-microcontrollers/arm-processors/i-mx-applications-processors/i-mx-8-processors/i-mx-8m-plus-arm-cortex-a53-machine-learning-vision-multimedia-and-industrial-iot:IMX8MPLUS

- PyeIQ 3.x Release User Guide

https://community.nxp.com/t5/Blogs/PyeIQ-3-x-Release-User-Guide/ba-p/1305998

- eIQ toolkit

https://www.nxp.com/design/software/development-software/eiq-ml-development-environment/eiq-toolkit-for-end-to-end-model-development-and-deployment:EIQ-TOOLKIT

- AI and Machine Learning

https://www.nxp.com/applications/enabling-technologies/ai-and-machine-learning:MACHINE-LEARNING

Q&A

SECURE CONNECTIONS
FOR A SMARTER WORLD

SECURE CONNECTIONS
FOR A SMARTER WORLD