Software

INTEL® DISTRIBUTION OF OPENVINO™ TOOLKIT.

Jamal EL YOUSSEFI
Senior Software Application Engineer
Intel Architecture, Graphics and Software Group

INTEL® DISTRIBUTION OF OPENVINO™ TOOLKIT

Take your computer vision solutions to a new level with deep learning inference intelligence.

What it is

A toolkit to accelerate **high performance computer vision** & **deep learning inference into vision/AI applications** used from edge to cloud. It enables deep learning on hardware accelerators and easy deployment across multiple types of Intel® platforms.

Who needs this product?

- Computer vision, deep learning software developers
- Data scientists
- OEMs, ISVs, System Integrators

Usages

Security surveillance, robotics, retail, healthcare, AI, office automation, transportation, non-vision use cases (speech, text) & more.

HIGH PERFORMANCE, PERFORM AI AT THE EDGE

STREAMLINED & OPTIMIZED DEEP LEARNING INFERENCE

HETEROGENEOUS, CROSS-PLATFORM FLEXIBILITY

Free Download ▸ software.intel.com/openvino-toolkit

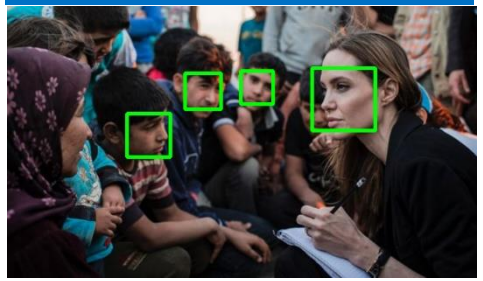
Open Source version ▸ 01.org/openvinotoolkit

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

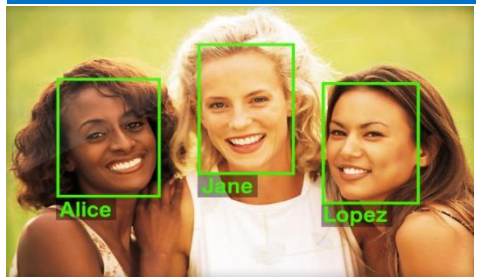
OpenVINO™ Toolkit Uses Cases

PEOPLE DETECTION

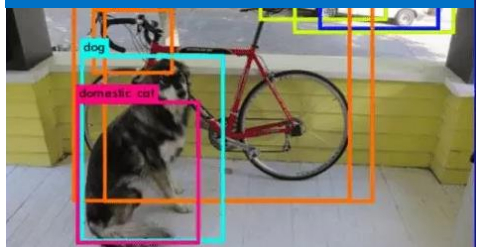
FACE DETECTION



FACE RECOGNITION



OBJECT DETECTION & CLASSIFICATION



RETAIL INTELLIGENCE

VEHICLE DETECTION

VEHICLE CLASSIFICATION

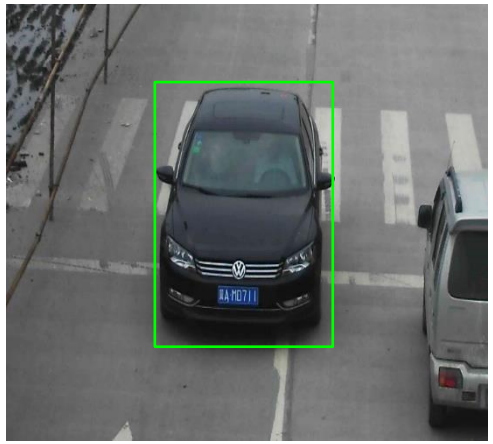
AUTOMOTIVE

INDUSTRIAL INSPECTION

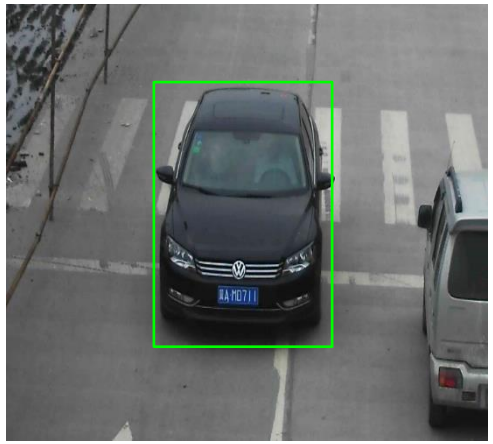
Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

OpenVINO™ Toolkit Uses Cases

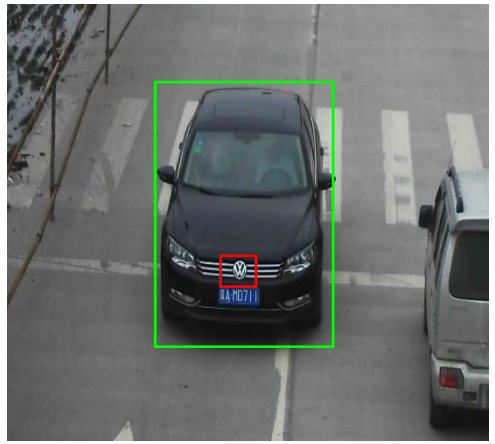
A Car



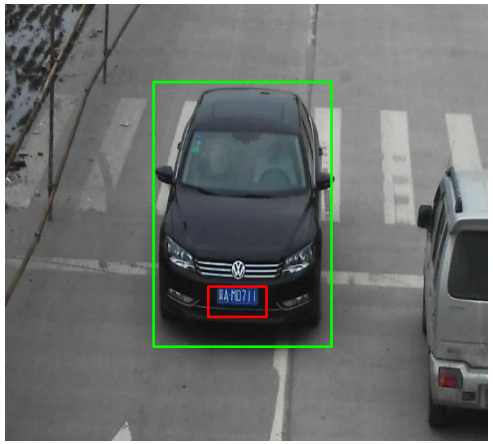
A Black Car



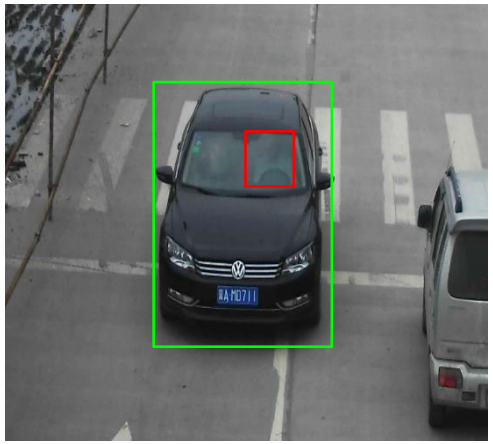
Volkswagen Passat



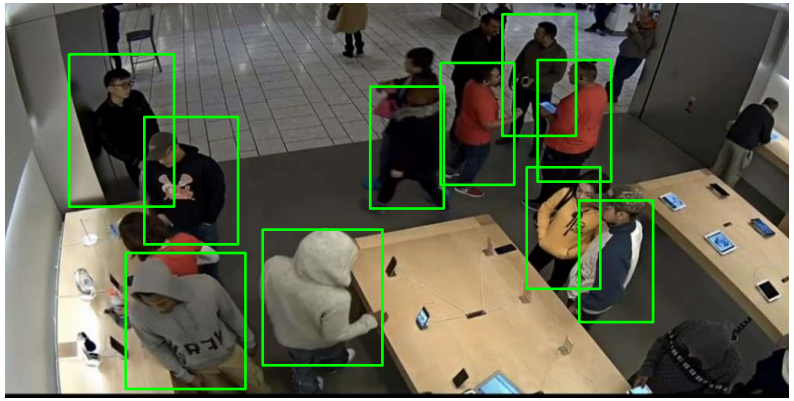
license plate number



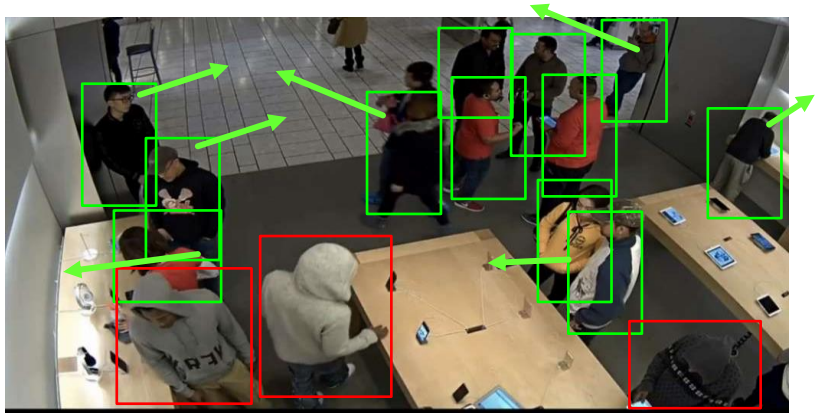
Not the owner !!



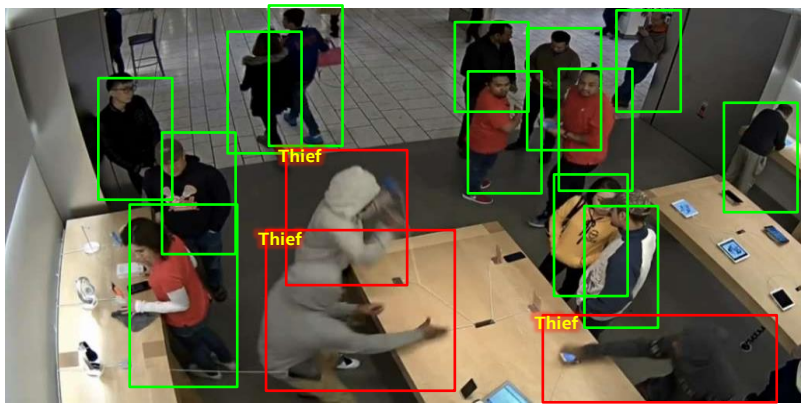
People Detection



People Tracking



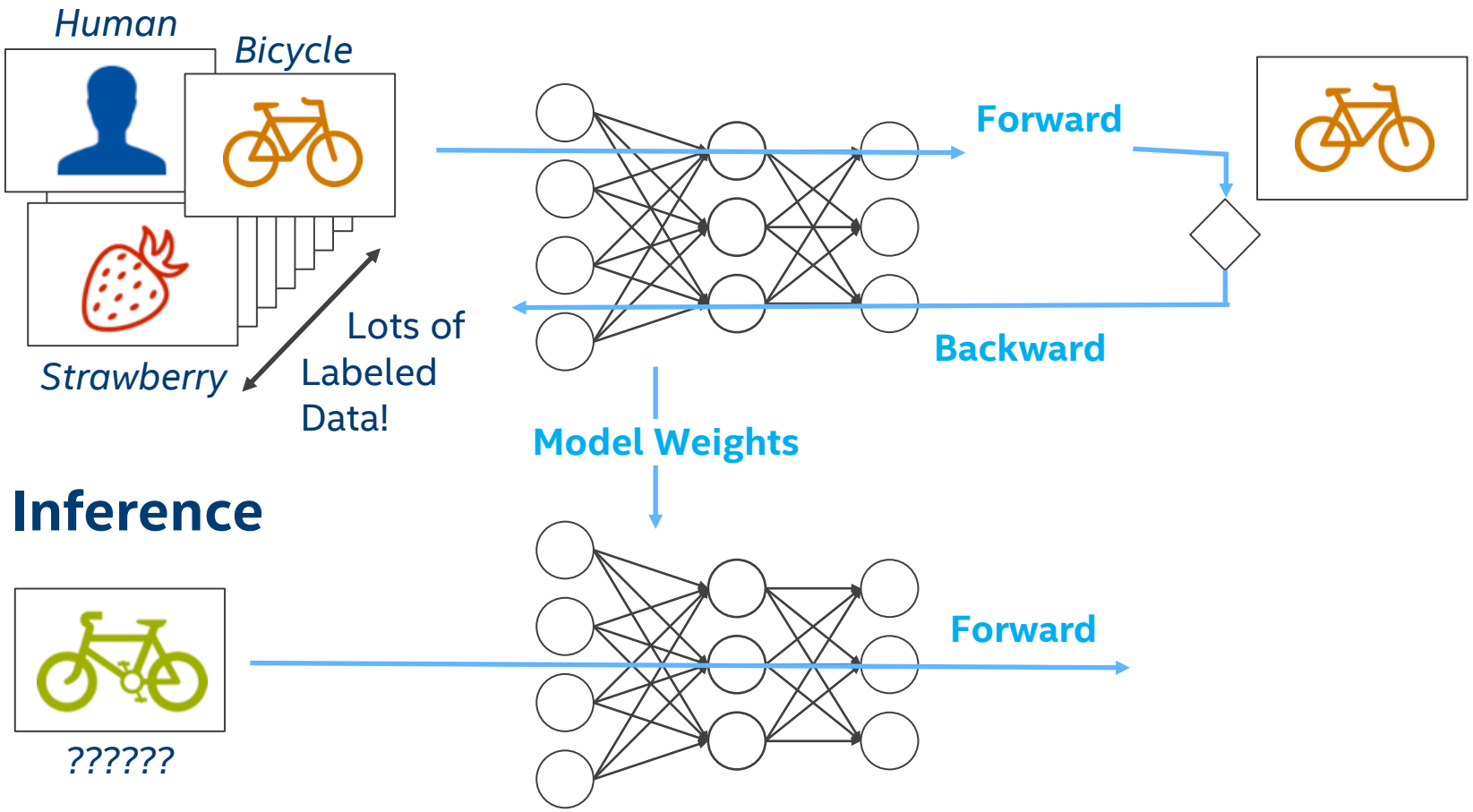
Analyze behavior/ intentions



Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Deep Learning: Training vs. Inference

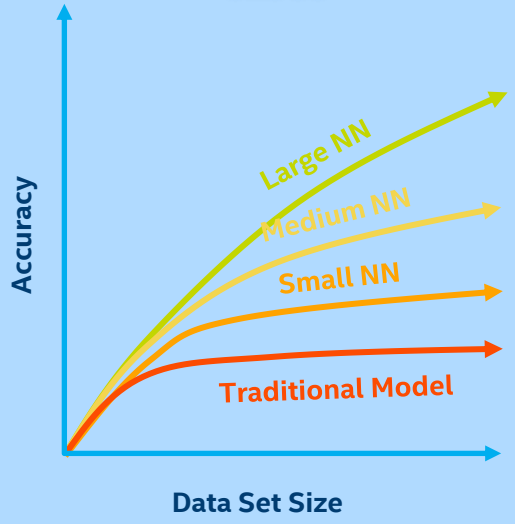
Training



Inference

Did You Know?

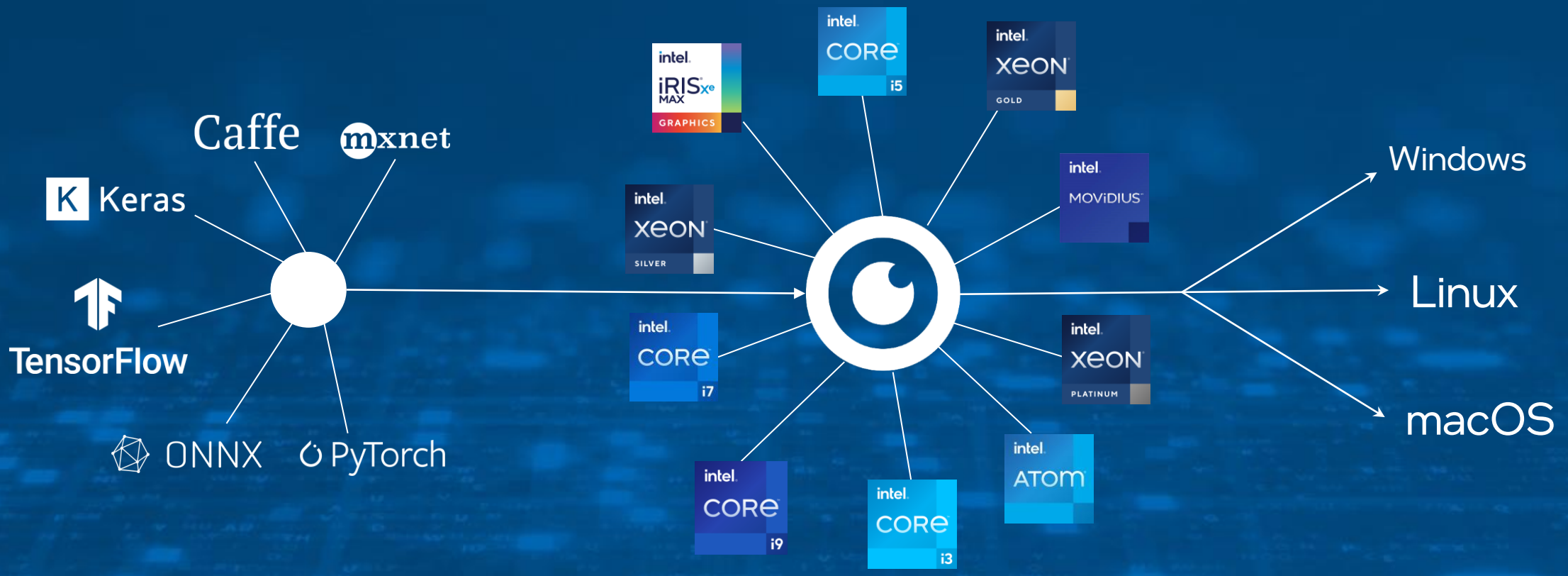
Training requires a very large data set and deep neural network (many layers) to achieve the highest accuracy in most cases



Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

OpenVINO™ toolkit in a nutshell

Advanced capabilities to streamline deep learning deployments



Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

What's Inside Intel® Distribution of OpenVINO™ toolkit

Deep Learning

Intel® Deep Learning Deployment Toolkit

Model Optimizer
Convert & Optimize

IR

Inference Engine
Optimized Inference

IR = Intermediate Representation file

Open Model Zoo

- 280+ Pretrained Models
- Demos
- Model Downloader

Deep Learning Workbench

- Calibration Tool
- Model Analyzer
- Benchmark App
- Accuracy Checker
- Aux. Capabilities

Traditional Computer Vision

Optimized Libraries & Code Samples

- OpenCV*
- OpenVX*
- Samples

For Intel® CPU & GPU/Intel® Processor Graphics

Tools & Libraries

Increase Media/Video/Graphics Performance

- **Intel® Media SDK** Open Source version
- **OpenCL™ Drivers & Runtimes**

For GPU/Intel® Processor Graphics

Optimize Intel® FPGA (Linux* only)

- **FPGA RunTime Environment** (from Intel® FPGA SDK for OpenCL™)
- **Bitstreams**

OS Support: Ubuntu* 20.04.3 LTS (64 bit), Microsoft Windows* 10 (64 bit), Yocto Project* version Poky Jethro v2.0.3 (64 bit), macOS* 10.13 & 10.14 (64 bit)

Intel® Architecture-Based Platforms Support

Intel® Vision Accelerator Design Products & AI in Production/ Developer Kits

An open source version is available at 01.org/openvinotoolkit (some deep learning functions support Intel CPU/GPU only).

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

OpenVX and the OpenVX logo are trademarks of the Khronos Group Inc.
OpenCL and the OpenCL logo are trademarks of Apple Inc. used by permission by Khronos

Open Model Zoo

Pre-Trained Model

TensorFlow ONNX KALDI mxnet Caffe

Supported Frameworks and Formats ▸ https://docs.openvinotoolkit.org/latest/openvino_docs_MO_DG_prepare_model_Supported_Frameworks_Layers.html
Configure the Model Optimizer for your Framework ▸ https://docs.openvinotoolkit.org/latest/_docs_MO_DG_prepare_model_Config_Model_Optimizer.html

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

OpenVINO™ DEVELOPER JOURNEY

1 BUILD 2 OPTIMIZE 3 DEPLOY

280+

Pre-Trained + Optimized Models

- Object Detection
- Object Recognition
- Gesture Recognition
- Machine Translation
- Text Recognition
- Question Answering
- Compressed Models
- Semantic Segmentation
- Human Pose Estimation
- Action Recognition
- Image Retrieval
- Image Processing
- Instance Segmentation
- Text Detection
- Text Spotting

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Intel® Deep Learning Deployment Toolkit

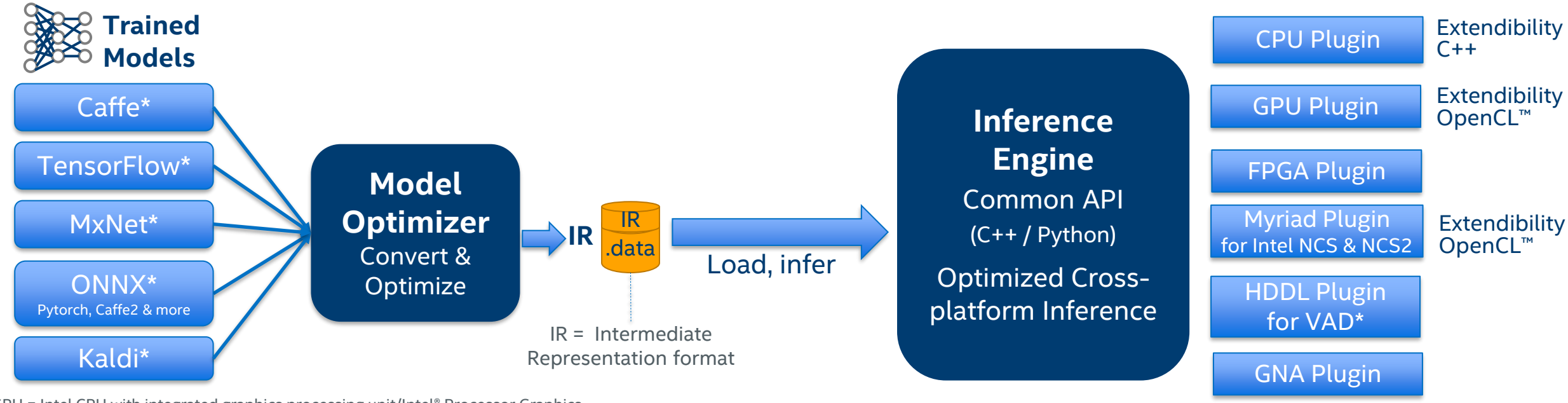
For Deep Learning Inference

Model Optimizer

- **What it is**: A Python*-based tool to import trained models and convert them to Intermediate representation.
- **Why important**: Optimizes for performance/space with conservative topology transformations; biggest boost is from conversion to data types matching hardware.

Inference Engine

- **What it is**: High-level inference API
- **Why important**: Interface is implemented as dynamically loaded plugins for each hardware type. Delivers best performance for each type without requiring users to implement and maintain multiple code pathways.



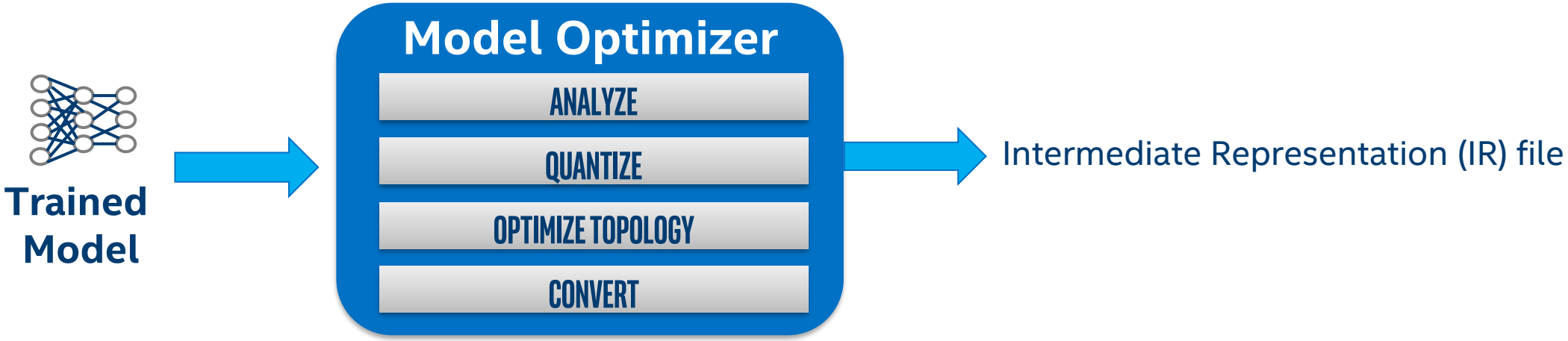
GPU = Intel CPU with integrated graphics processing unit/Intel® Processor Graphics

*VAD = Vision Accelerator Design Products (HDDL-R)

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

OpenCL and the OpenCL logo are trademarks of Apple Inc. used by permission by Khronos

Improve Performance with Model Optimizer

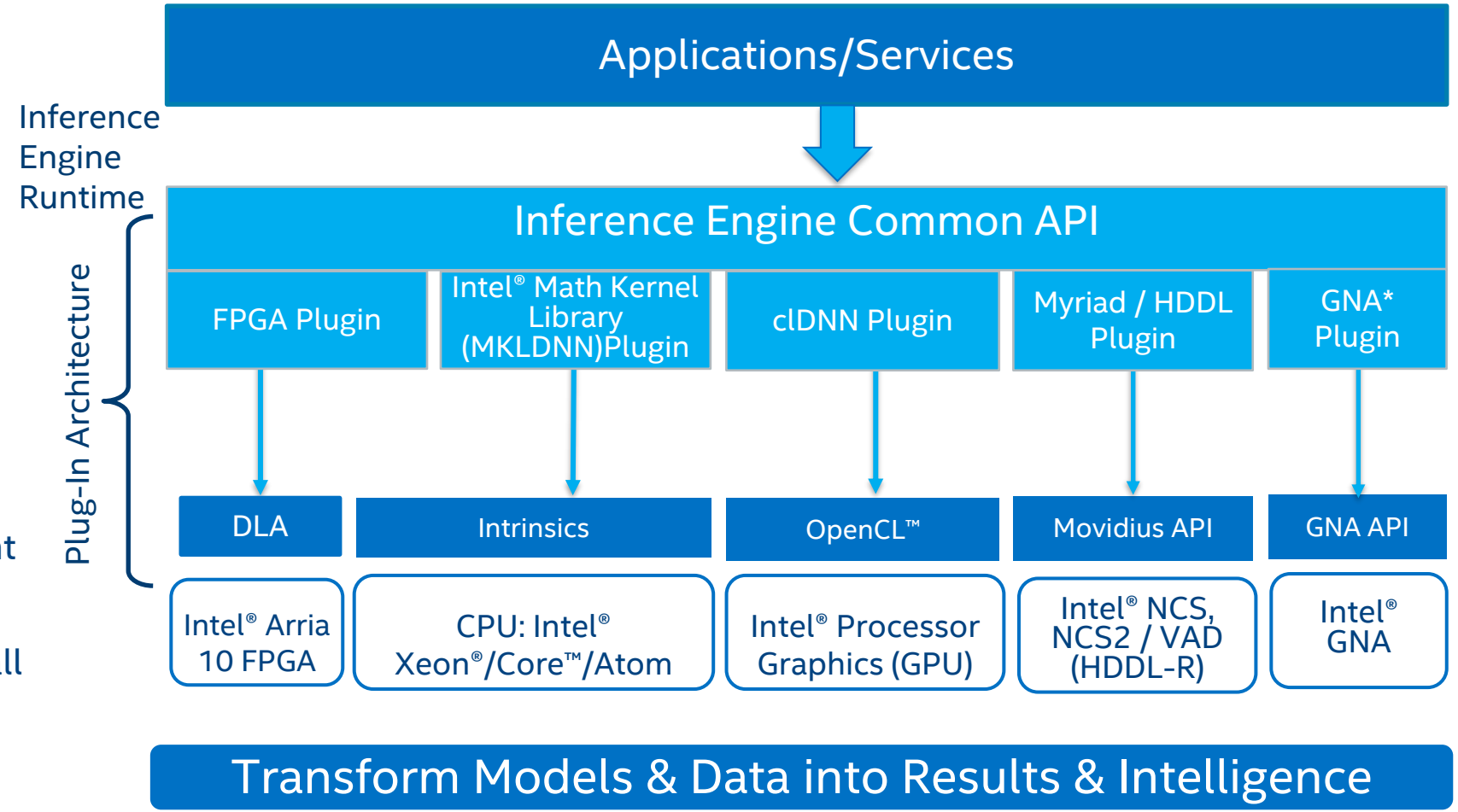


- Easy to use, Python*-based workflow does not require rebuilding frameworks.
- Import Models from many supported frameworks: Caffe*, TensorFlow*, MXNet*, Kaldi*, exchange formats like ONNX* (Pytorch*, Caffe2* and others through ONNX).
- 100+ models for Caffe, MXNet, TensorFlow validated. Supports all ONNX* model zoo public models.
- Extends inferencing for non-vision networks with support of LSTM, Bert, GNMT, TDNN-LSTM, ESPNet and more.
- IR files for models using standard layers or user-provided custom layers do not require Caffe.
- Fallback to original framework is possible in cases of unsupported layers, but requires original framework.

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Optimal Model Performance Using the Inference Engine

- Simple & unified API for inference across all Intel® architecture
- Optimized inference on large IA hardware targets (CPU/GEN/FPGA)
- Heterogeneity support allows execution of layers across hardware types
- Asynchronous execution improves performance
- Futureproof/scale your development for future Intel® processors
- Supports serialized FP16 IR across all plugins / platforms



GPU = Intel CPU with integrated graphics/Intel® Processor Graphics/GEN
GNA = Gaussian mixture model and Neural Network Accelerator

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.
OpenVX and the OpenVX logo are trademarks of the Khronos Group Inc.
OpenCL and the OpenCL logo are trademarks of Apple Inc. used by permission by Khronos

Using the Intel® Distribution of OpenVINO™ toolkit

Advanced Capabilities to Streamline Deep Learning Deployment

1. Build

Models & Frameworks

- Intel-optimized Models
- Customer Models
- Open Source Models

ONNX
Caffe
mxnet
TensorFlow
KALDI

Traditional Computer Vision Tools

2. Optimize

Developer Tools
(Developer Workstations)

OpenVX* | Samples

OpenCV*

Model Optimizer
Convert & optimize

IR file

or

Open Model Zoo
(40+ pretrained models)
existing IR files

IR file

Deep Learning Workbench

Calibration Tool | Model Analyzer | Benchmark App | Accuracy Checker | Aux. Capabilities

IR file = Intermediate Representation file

3. Deploy

Deployment Package
(Target Platform)

Intel® Media SDK open source version

OpenCL™ Drivers and Runtimes

Inference Engine
Common API (C++/Python)

Optimized Cross-platform Inference

- CPU Plugin — Extendibility C++
- GPU Plugin — Extendibility OpenCL™
- GNA Plugin
- Myriad Plugin (For Intel® NCS2 & NCS) — Extendibility OpenCL™
- HDDL Plugin — **Intel® Vision Accelerator Design Products**
- FGPA Plugin — **Intel® FPGA** (Linux* only)

FPGA RunTime Environment
(from Intel® FPGA SDK for OpenCL™)

Bitstreams

Some deep learning functions support CPU or CPU/GPU only

- AI in Production Solutions
- Intel/Partner Developer Kits

Intel® NCS = Intel® Neural Compute Stick (VPU)

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

OpenVX and the OpenVX logo are trademarks of the Khronos Group Inc.
OpenCL and the OpenCL logo are trademarks of Apple Inc. used by permission by Khronos

End-to-End Vision Workflow

Video input

Intel® Media SDK → Decode → OpenCV* → Pre-Processing → Intel® Deep Learning Deployment Toolkit → Inference → OpenCV → Post-Processing → Intel Media SDK → Encode

CPU GPU | CPU GPU FPGA VPU | CPU GPU

Video output with results annotated

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Inference Engine Code Examples in C++ and Python

Samples

Simple applications demonstrating API features and simple functionality.

Part of DLDT

- classification_sample_async
- hello_reshape_ssd
- speech_sample
- style_transfer_sample

https://github.com/opencv/dldt/tree/2019/inference-engine/samples

Demos

More complex applications demonstrating complicated scenarios, public and Intel models inference.

Part of Open Model Zoo

- security_barrier_camera_demo
- object_detection_demo
- smart_classroom_demo
- human_pose_estimation_demo

https://github.com/opencv/open_model_zoo/tree/master/demos

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

How to measure performance ?

Benchmark app – command line application which calculates the latency (ms) and throughput (fps). Recommended way to estimate Inference Engine performance.

- Distributed in binary tool and as part of IE samples
- Usage:

```
$ benchmark_app -m <model> -i <input> -d <device>
```

- Automatically queries and sets optimal values of the following options to saturate the device:
  - **"-nireq"** – number of Infer requests
  - **"-nstreams"** (CPU, GPU, VPU) – number of execution "streams"

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Speed Deployment with Pretrained Models & Samples

Expedite development, accelerate deep learning inference performance, speed production deployment

Pretrained Models in Intel® Distribution of OpenVINO™ toolkit

- Age & Gender
- Face Detection–standard & enhanced
- Head Position
- Human Detection–eye-level & high-angle detection
- Detect People, Vehicles & Bikes
- License Plate Detection: small & front facing
- Vehicle Metadata
- Human Pose Estimation
- Action recognition – encoder & decoder
- Text Detection & Recognition
- Vehicle Detection
- Retail Environment
- Pedestrian Detection
- Pedestrian & Vehicle Detection
- Person Attributes Recognition Crossroad
- Emotion Recognition
- Identify Someone from Different Videos–standard & enhanced
- Facial Landmarks
- Gaze estimation
- Identify Roadside objects
- Advanced Roadside Identification
- Person Detection & Action Recognition
- Person Re-identification–ultra small/ultra fast
- Face Re-identification
- Landmarks Regression
- Smart Classroom Use Cases
- Single image Super Resolution (3 models)
- Instance segmentation
- and more…

Binary Models

- Face Detection Binary
- Pedestrian Detection Binary
- Vehicle Detection Binary
- ResNet50 Binary

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Python Object Detection code sample

```
from openvino.inference_engine import IENetwork, IECore

ie = IECore()
net = IENetwork(model=model_xml, weights=model_bin)

input_blob = next(iter(net.inputs))
out_blob = next(iter(net.outputs))
n, c, h, w = net.inputs[input_blob].shape

exec_net = ie.load_network(network=net, num_requests=2, device_name="CPU")

cap = cv2.VideoCapture(0) # camera 0
frame = cap.read()

in_frame = cv2.resize(frame, (w, h))
in_frame = in_frame.transpose((2, 0, 1))  # Change data layout from HWC to CHW
in_frame = in_frame.reshape((n, c, h, w))

exec_net.start_async(request_id=cur_request_id, inputs=[in_frame])

if exec_net.requests[cur_request_id].wait(-1) == 0:
          res = exec_net.requests[cur_request_id].outputs[out_blob]
          for obj in res[0][0]:
                # Draw only objects when probability more than specified threshold
                if obj[2] > 0.5:
                    xmin = obj[3]
                    ymin = obj[4]
                    xmax = obj[5]
                    ymax = obj[6]
                    class_id = int(obj[1])
```

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

SECURITY BARRIER RECOGNITION MODEL USING INTEL® DEEP LEARNING DEPLOYMENT TOOLKIT

Load Input Image(s)

Run Inference 1:
Model
vehicle-license-plate-detection-barrier-0007

Detects Vehicles

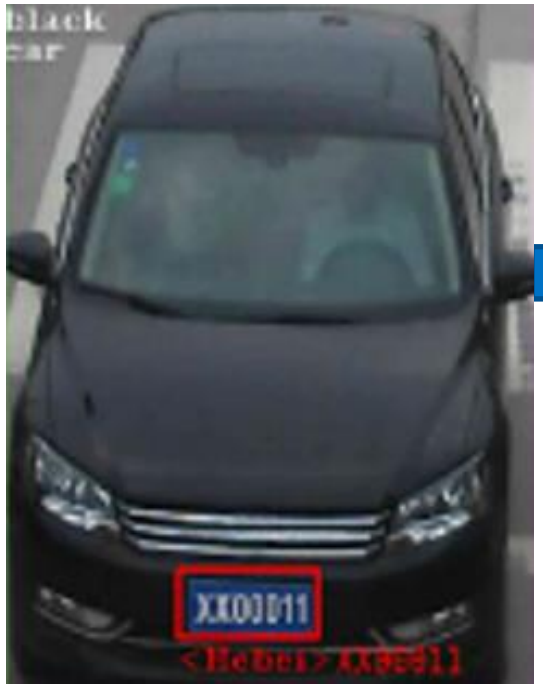
Run Inference 2:
Model
vehicle-attributes-recognition-barrier-0010

Classifies vehicle attributes

Run Inference 3:
Model
license-plate-recognition-barrier-0001

Detects License Plates

Display Results



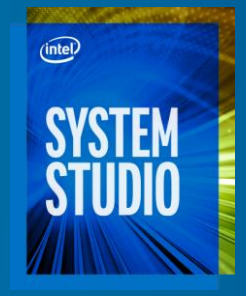
Vehicle Detection Time : 30.10 ms (33.23 fps)
Vehicle Attribs Time (averaged over 2 detections) :6.26 ms (159.71 fps)
LPR Time (averaged over 1 detection) :5.04 ms (198.43 fps)

Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Common & Seamless Developer Resources

Find it all at Intel® Developer Zone: software.intel.com

SOFTWARE DEVELOPMENT TOOLS



- Intel® Media SDK
- Intel® Distribution for Python*
- Intel® Math Kernel Library for Deep Neural Networks

HARDWARE

AI & IoT Developer Kits & Accelerators

ECOSYSTEM SUPPORT

- Services & programs
- Events, training, engagements
- Expert forums & Intel Priority Support

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

OpenVINO™ Toolkit

Open Source Version

- Provides flexibility and availability to the developer community to extend OpenVINO™ toolkit for custom needs
- Components that are open sourced
  - **Deep Learning Deployment Toolkit** with **CPU, GPU, Heterogeneous, Myriad** (for Intel® Neural Compute Stick (Intel® NCS) & Intel® NCS2), and **GNA** plugins github.com/opencv/dldt
  - **Open Model Zoo** - Includes pretrained models, model downloader, demos and samples: github.com/opencv/open_model_zoo
- See FAQ and next slides for key differences between the open source and Intel distribution

Learn More ▶ 01.org/openvinotoolkit

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

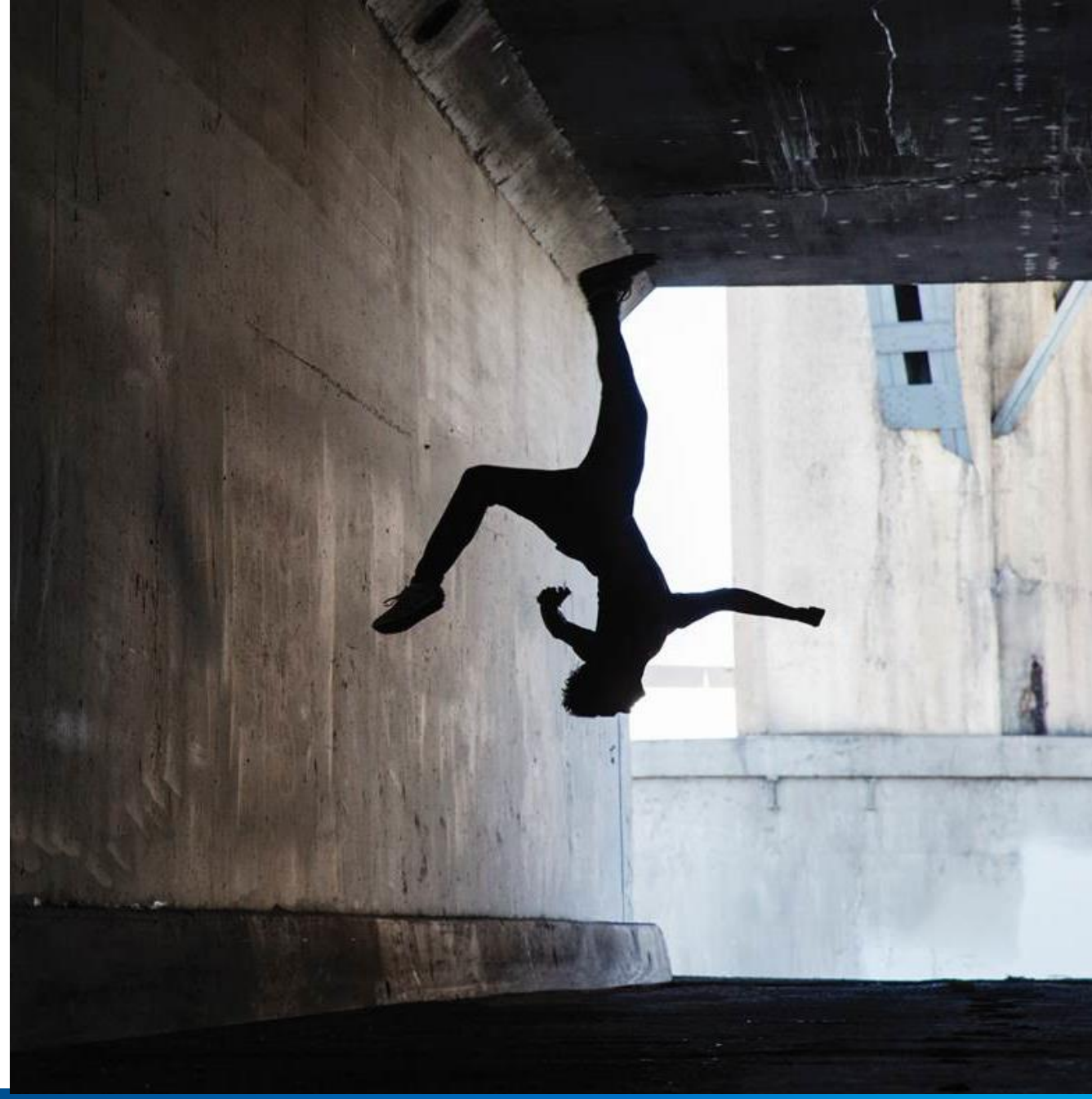
Resources

Download Intel® Distribution for OpenVINO™ Toolkit: software.intel.com/en-us/openvino-toolkit

Open Sourced version: 01.org/openvinotoolkit

Documentation: docs.openvinotoolkit.org/

Forum: software.intel.com/en-us/forums/computer-vision



Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

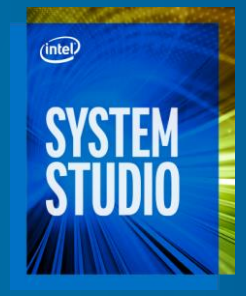
Q&A

BACKUP

Common & Seamless Developer Resources

Find it all at Intel® Developer Zone: software.intel.com

SOFTWARE DEVELOPMENT TOOLS



- Intel® Media SDK
- Intel® Distribution for Python*
- Intel® Math Kernel Library for Deep Neural Networks

HARDWARE

AI & IoT Developer Kits & Accelerators

intel MOVIDIUS inside

intel ATOM x7 inside | intel MOVIDIUS inside

intel CORE i5 inside | intel CORE i7 inside | intel XEON inside | intel MOVIDIUS inside | intel ARRIA 10 inside

ECOSYSTEM SUPPORT

- Services & programs
- Events, training, engagements
- Expert forums & Intel Priority Support

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Additional Services & Programs

Intel® Solutions Select & Intel® IoT Solutions Alliance

Find partner solutions based on Intel® architecture

Learn more >

AI: In Production Community

Accelerate the development of AI-centric applications at the edge

Learn more >

Intel® IoT RFP Ready Kits

Get use-case-specific technology offerings that can be easily customized to quickly respond to RFPs

Learn more >

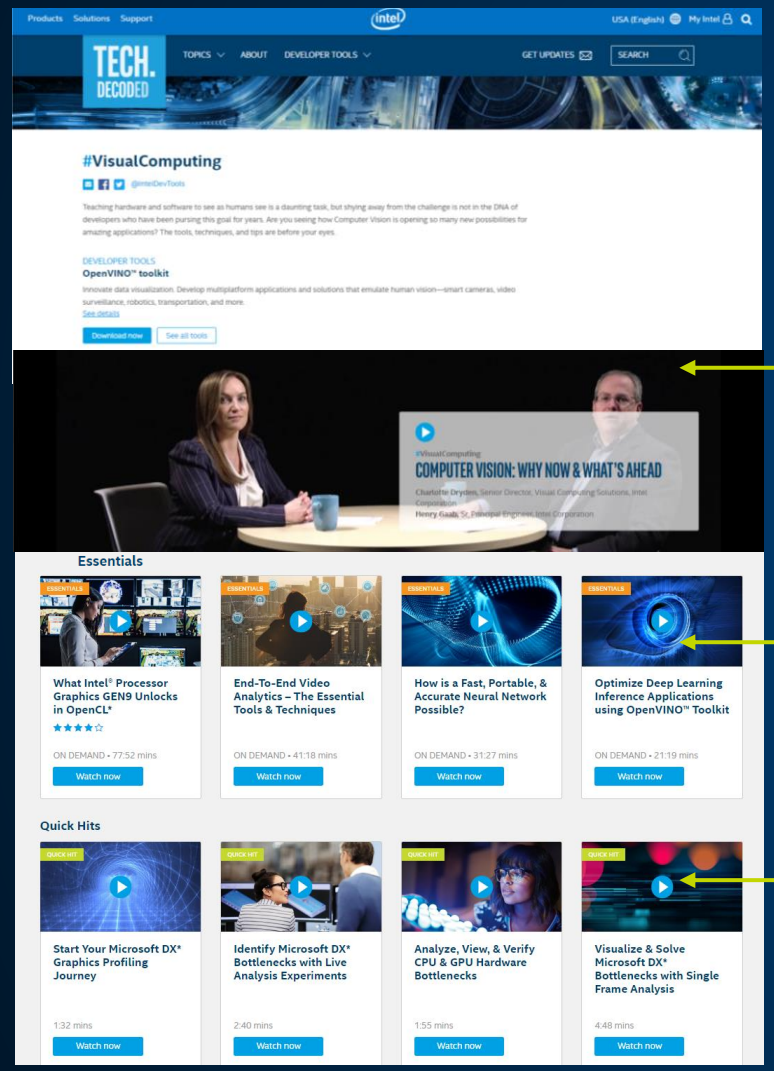
Intel® IoT Market Ready Solutions

Submit use-case-specific solutions, get unique support and scaling opportunities

Learn more >

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Get the Most from Your Code: Software Tools Training



Put key optimization strategies into practice with Intel development tools - **TechDecoded.intel.io**

Big Picture Videos

Discover Intel's vision for key development areas.

Essential Webinars

Gain strategies, practices and tools to optimize application and solution performance.

How-to Videos & Articles

Learn how to do specific programming tasks using Intel software tools.

Intel® Distribution of OpenVINO™ toolkit

Multiple Visionary videos

9 Webinars

27 How-to videos, articles

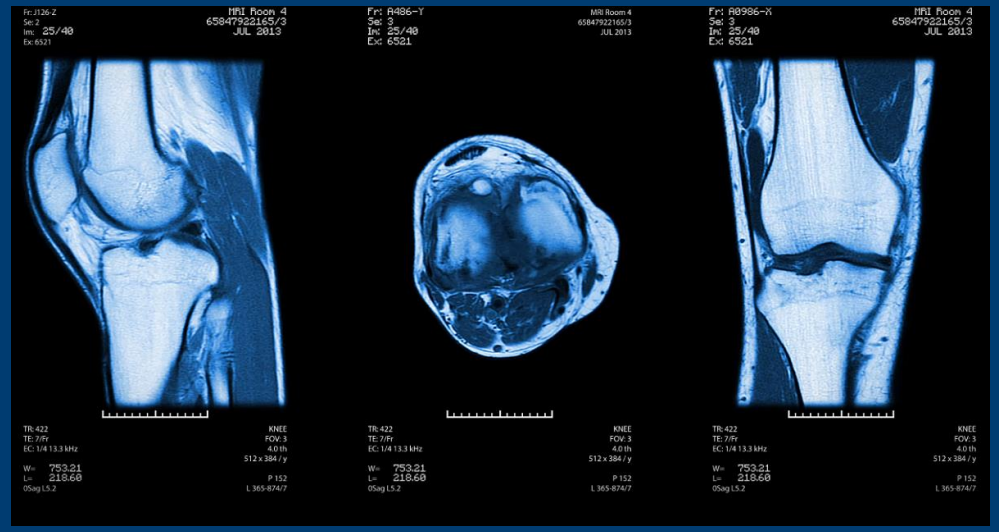
OVERALL DEVELOPMENT TOPICS

- Visual Computing
- Code Modernization
- Systems & IoT
- Data Science
- Data Center & Cloud Computing

Easily Find Training for all the Essential Software Tools You Need

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Key Vision Solutions Optimized by Intel® Distribution of OpenVINO™ toolkit



Philips

Intel teamed with Philips to show that servers powered by Intel® Xeon® Scalable processors & Intel® Distribution of OpenVINO™ toolkit can efficiently perform deep learning inference on patients' X-rays & computed tomography (CT) scans, without the need for accelerators. Achieved breakthrough performance for AI inferencing:

- **188x** increase in throughput (images/sec) on Bone-age prediction model.[1]
- **38x** increase in throughput (images/sec) on Lung segmentation model. [1]

"Intel® Xeon® Scalable processors and OpenVINO toolkit appears to be the right solution for medical imaging AI workloads. Our customers can use their existing hardware to its maximum potential, without having to complicate their infrastructure, while still aiming to achieve quality output resolution at exceptional speeds."
— Vijayananda J., chief architect and fellow, Data Science and AI, Philips HealthSuite Insights, India

White Paper

[1]See white paper for performance details.

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Key Vision Solutions Optimized by Intel® Distribution of OpenVINO™ toolkit

GE Healthcare*

The Intel® Distribution of OpenVINO™ toolkit helped GE deliver optimized inferencing to its deep learning image-classification solution. By bringing AI to its clinical diagnostic scanning, GE no longer needed an expensive 3rd party accelerator board, achieving:

- **5.9x** inferencing performance above the target[1]
- **14x** inferencing speed over the baseline solution[1]
- Improved image quality, diagnostic capabilities, and clinical workflows

With the OpenVINO™ toolkit , we are now able to optimize inferencing across Intel® silicon, exceeding our throughput goals by almost 6x," said David Chevalier, Principal Engineer for GE Healthcare.
"We want to not only keep deployment costs down for our customers, but also offer a flexible, high-performance solution for a new era of smarter medical imaging. Our partnership with Intel allows us to bring the power of AI to clinical diagnostic scanning and other healthcare workflows in a cost-effective manner."

Intel-GE Healthcare, Intel® Distribution of OpenVINO™ Optimizes Deep Learning Performance for Healthcare Imaging

[1]See white paper for performance details.

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Security for Stadiums at the World Cup 2018

RESULT axxonsoft

9000+ SURVEILLANCE CAMERAS

used to protect

2MILLION+ FANS

The integrated surveillance system connected to cameras at stadiums, which transmitted video data to operational HQ in each city.

Intel® Distribution of OpenVINO™ toolkit allowed Axxonsoft to distribute the neural network video analytics of the video across all available Intel hardware, for zone entry detection, abandoned objects detection, and facial recognition.

Optimization Notice

See case study for details.

Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Train Safety

RESULT

KLAS TELECOM

IN-TRAIN VISION PLATFORM

Enables pedestrian & vehicle identification at crossroads + on-train empty seat detection

The platform provides multiple functions from onboard Wi-Fi to computer vision applications such as human/vehicle detection at crossroads, onboard empty seat detection and intruder detection.

OpenVINO provides a scalable, high performance common platform across a variety of hardware for greater efficiencies.

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Demonstrated Industry Success

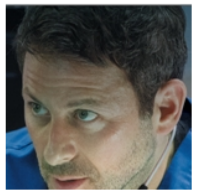
Access Developer Success Stories for details & more examples

NexCOBOT Delivers Robotics with AI for Industry 4.0L

NexCOBOT, a NEXCOM company, offers a flexible, modular robotics solution integrating artificial intelligence (AI) with machine vision, powered by Intel® Vision Accelerator Design products and optimized by Intel® Distribution of OpenVINO™ toolkit.

Philips Performs AI-Driven Medical Imaging Efficiently and Cost-Effectively on Intel® CPU–Based Systems

Philips demonstrates breakthrough performance for AI inferencing of healthcare workloads run on servers powered by Intel® Xeon® Scalable processors and optimized with the OpenVINO™ toolkit.



QNAP/IEI Develop AI Solutions for Healthcare

Together, Intel and QNAP/IEI have come up with a solution that offers developers, data scientists, medical researchers, and students a quick-to-deploy computer vision system combining a workstation, deep learning software development kit (Intel® Distribution of OpenVINO™ toolkit), and powerful NAS.

For Somatic, Deep Learning Brings Touch to Robots

Deep learning algorithms that enable touch as well as vision can create tremendous opportunities for robotics applications. Intel® Software Development Tools can help devs take advantage of enabling technologies that bring touch to the forefront and are fueled by the latest artificial intelligence (AI) advances.

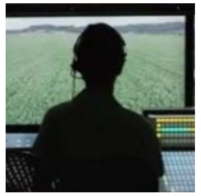
ADLINK and Touch Cloud Deliver AI Solutions Powered by Intel® Vision Products

A

DLINK, Touch Cloud, and Intel provide a turnkey AI engine to assist in data analytics, detection, classification, and prediction for a wide range of use cases, optimized by Intel Distribution of OpenVINO toolkit.

Advantech Addresses Major Retail Dilemmas with Intel® Vision Products

Advantech deploys a combination of hardware and software including cameras, AI deep learning, and video analysis technology optimized by Intel® Distribution of OpenVINO™ toolkit and Intel® Vision Accelerator Design products.



Agent VI* Delivers the Next Generation of Digital Security and Surveillance Solutions

Intel® Distribution of OpenVINO™ toolkit is the centerpiece of computer Agent VI's next-gen vision solutions.

GeoVision Gets a 24x Deep Learning Algorithm Performance Boost

GeoVision turbo-charges its deep learning facial recognition solution using Intel® System Studio and the OpenVINO™ toolkit.

GE Healthcare and Intel Optimize Deep Learning Performance for Healthcare Imaging

Intel® Math Kernel Library and OpenVINO™ toolkit help bring the power of AI to clinical diagnostic scanning and other healthcare workflows.

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Technical Specifications - Intel® Distribution of OpenVINO™ toolkit

	Intel® Platforms	Compatible Operating Systems
Target Solution Platforms	**CPU** ▪ 6th-10th generation Intel® Core™ and Xeon® processors ▪ 1st and 2nd generation Intel® Xeon® Scalable processors	▪ Ubuntu* 18.04.3 LTS (64 bit) ▪ Microsoft Windows* 10 (64 bit) ▪ CentOS* 7.4 (64 bit) ▪ macOS* 10.13 & 10.14 (64 bit)
	▪ Intel® Pentium® processor N4200/5, N3350/5, N3450/5 with Intel® HD Graphics	▪ Yocto Project* Poky Jethro v2.0.3 (64 bit)
	Iris® Pro & Intel® HD Graphics ▪ 6th-10th generation Intel® Core™ processor with Intel® Iris™ Pro graphics & Intel® HD Graphics ▪ Intel® Xeon® processor with Intel® Iris™ Pro Graphics & Intel® HD Graphics (excluding E5 product family, which does not have graphics[1])	▪ Ubuntu 18.04.3 LTS (64 bit) ▪ Windows 10 (64 bit) ▪ CentOS 7.4 (64 bit)
	FPGA ▪ Intel® Arria® FPGA 10 GX development kit ▪ Intel® Programmable Acceleration Card with Intel® Arria® 10 GX FPGA operating systems ▪ OpenCV* & OpenVX* functions must be run against the CPU or Intel® Processor Graphics (GPU)	▪ Ubuntu 18.04.2 LTS (64 bit) ▪ CentOS 7.4 (64 bit)
	VPU: Intel Movidius™ Neural Compute Stick:, Intel® Neural Compute Stick2	Ubuntu 18.04.3 LTS (64 bit) CentOS 7.4 (64 bit) Windows 10 (64 bit) macOS* (64 bit) Raspbian (target only)
	Intel® Vision Accelerator Design Products ▪ Intel® Vision Accelerator Design with Intel® Arria10 FPGA ▪ Intel® Vision Accelerator Design with Intel® Movidius™ VPUs	▪ Ubuntu 18.04.2 LTS (64 bit) ▪ Ubuntu 8.04.3 LTS (64 bit) ▪ Windows 10 (64 bit)
Development Platforms	▪ 6th-10th generation Intel® Core™ and Intel® Xeon® processors ▪ 1st and 2nd generation Intel® Xeon® Scalable processors	▪ Ubuntu* 18.04.3 LTS (64 bit) ▪ Windows® 10 (64 bit) ▪ CentOS* 7.4 (64 bit) ▪ macOS* 10.13 & 10.14 (64 bit)
Additional Software Requirements	Linux* build environment required components ▪ OpenCV 3.4 or higher ▪ GNU Compiler Collection (GCC) 3.4 or higher ▪ CMake* 2.8 or higher ▪ Python* 3.4 or higher	
	Microsoft Windows* build environment required components ▪ Intel® HD Graphics Driver (latest version)† ▪ OpenCV 3.4 or higher ▪ Intel® C++ Compiler 2017 Update 4 ▪ CMake 2.8 or higher ▪ Python 3.4 or higher ▪ Microsoft Visual Studio* 2015	
External Dependencies/Additional Software		View Product Site, detailed System Requirements

Optimization Notice

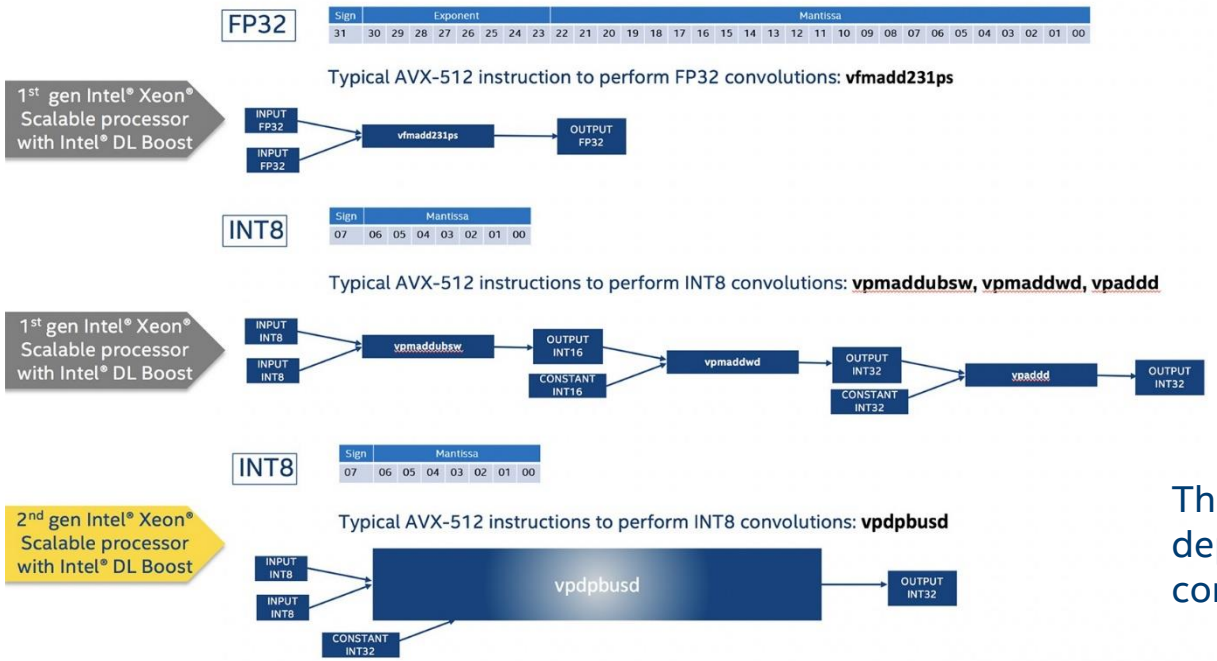
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

[1]Graphics drivers are required only if you use Intel® Processor Graphics (GPU).

Int8 support on CPU – why important?

Significant performance boost and little loss of accuracy because

- Benefit from less data size on Intel® platforms with Intel® AVX-512, Intel® AVX2, Intel® SSE4.2
- Take advantage from VNNI (Vector Neural Network Instructions) on 2nd Generation Intel® Xeon® Scalable

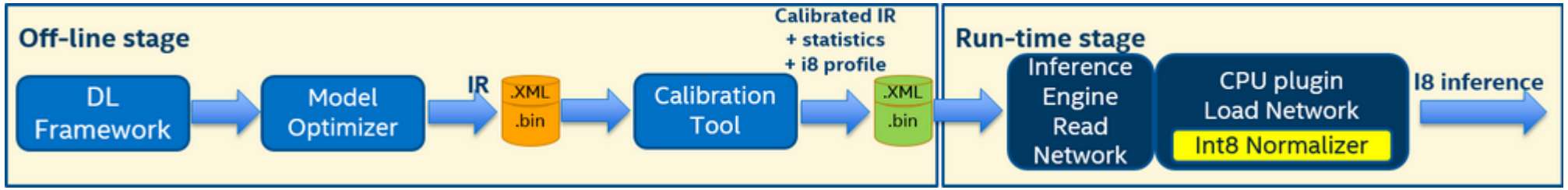


The size of performance gain is dependent on the topology (how big is convolutional part) and system

Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Int8 Calibration – Calibration Tool

Calibration tool - command line app which collects statistics from FP32 or FP16 IR (intermediate representations)



Calibration in **"simplified" mode** – see the maximum of potential performance gain from Int8 without accuracy calculation

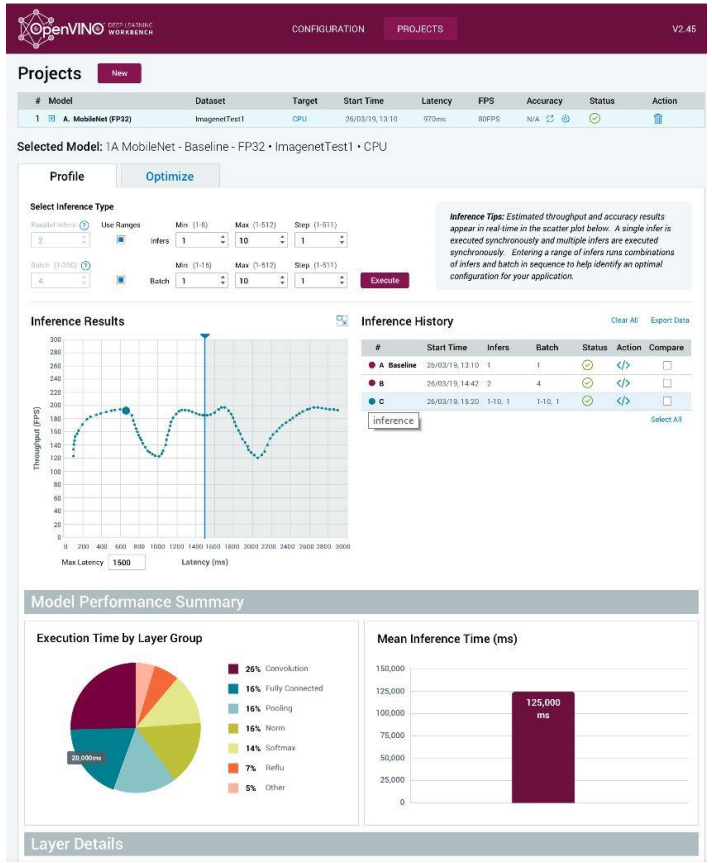
Pass full calibration process to get working Int8 model with accuracy statistics

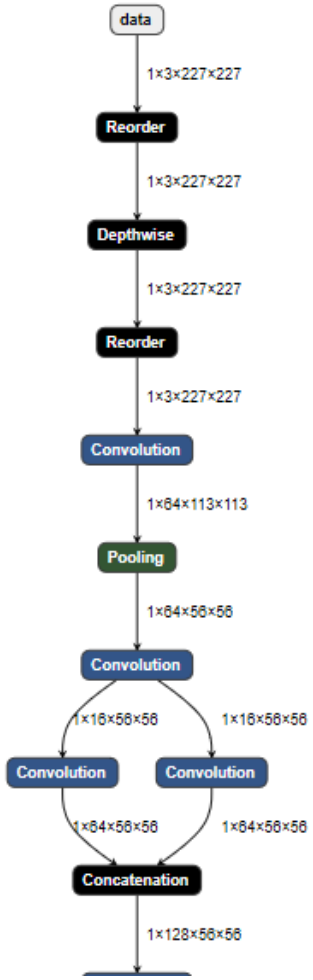
Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Deep Learning Workbench (Preview)

Deep Learning Workbench capabilities

- Web-based tool - UI extension of Intel® Distribution of OpenVINO™ toolkit functionality
- Visualizes performance data for topologies/ layers to aid in model analysis
- Automate analysis for optimal performance configuration (streams, batches, latency)
- Experiment with int8 calibration for optimal tuning
- Provide accuracy info through accuracy checker
- Direct access to Models from public set of Open Model Zoo





Optimization Notice
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

Notices & Disclaimers

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit www.intel.com/benchmarks.

INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS". NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO THIS INFORMATION INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

Copyright © 2019, Intel Corporation. All rights reserved. Intel, Pentium, Xeon, Core, VTune, OpenVINO, and the Intel logo are trademarks of Intel Corporation in the U.S. and other countries.

Optimization Notice

Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.
Notice revision #20110804

intel®

Software