



INNOVATE VISION SOLUTIONS FROM EDGE TO CLOUD

INTEL® DISTRIBUTION OF OPENVINO™ TOOLKIT

Open Visual Inference & Neural Network Optimization
Includes Intel® Deep Learning Deployment Toolkit

OpenVINO™



Awarded
by the
Embedded
Vision
Alliance

30-3-30 Presentation
IAGS Compute Performance Developer Products (CPDP)
October 2019

Intel Computer Vision/AI Portfolio

EXPERIENCES



TOOLS

Intel® Parallel Studio XE
Intel® System Studio
Intel® Media SDK

Intel® Distribution of OpenVINO™ toolkit
Intel® SDK for OpenCL™ Applications
Nauta

FRAMEWORKS



LIBRARIES

Intel® Data Analytics Acceleration Library

Intel® Distribution for Python*

Intel® Math Kernel Library

Intel® nGraph™ Compiler

Movidius Stack

HARDWARE



Compute



Memory & Storage



Networking



Visual Intelligence

UNLEASH FULL POTENTIAL

Optimization Notice

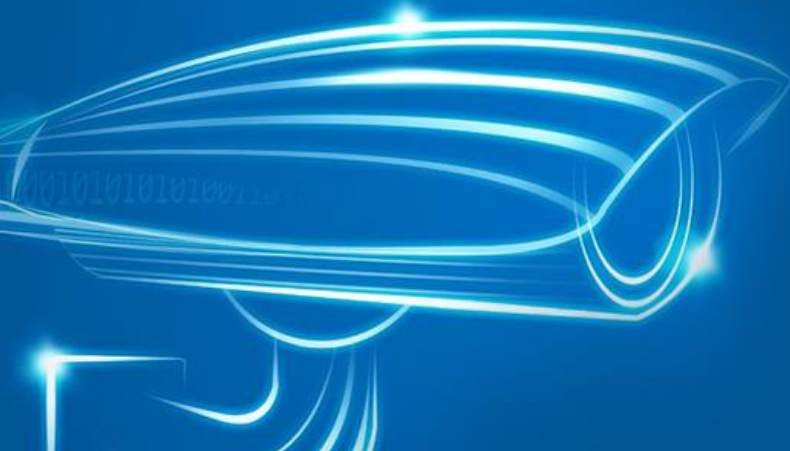
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

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



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 development of **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, NLP, Audio, 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.

Latest version is 2019 R3



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 target¹
- **14x** inferencing speed over the baseline solution¹
- 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](#)

Optimization Notice

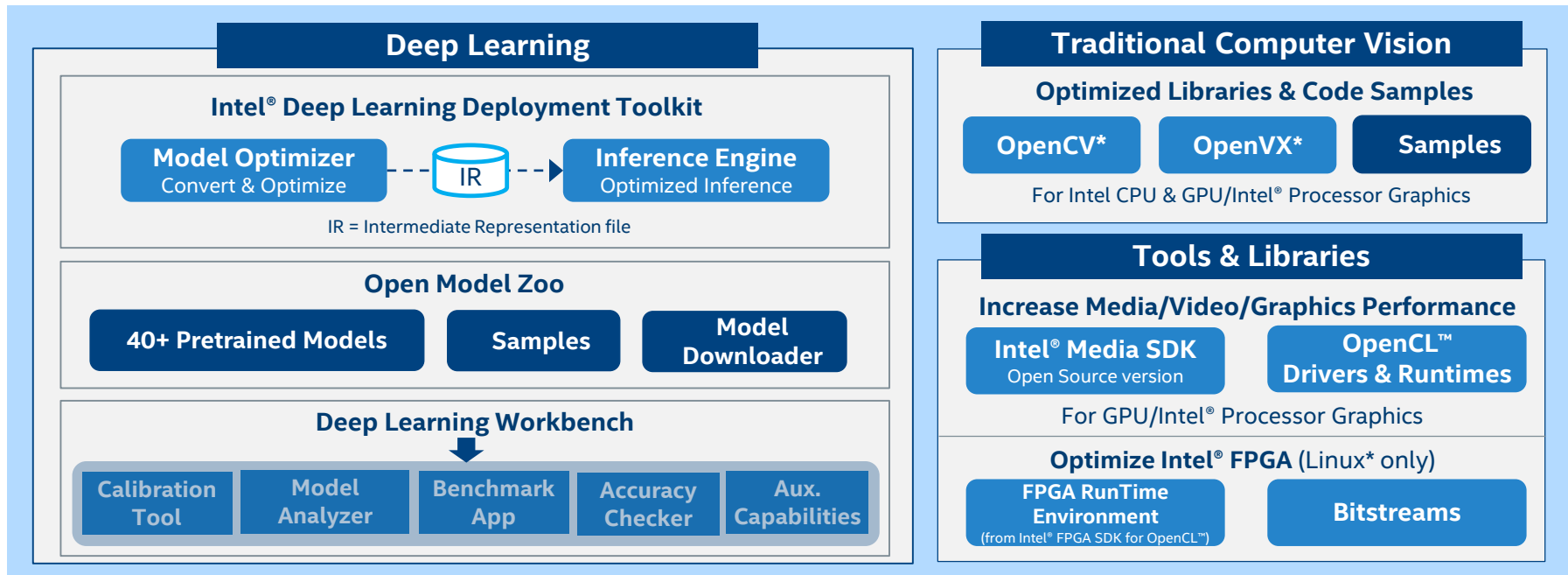
Copyright © 2019, Intel Corporation. All rights reserved.

*Other names and brands may be claimed as the property of others.

¹See white paper for performance details.



What's Inside Intel® Distribution of OpenVINO™ toolkit



OS Support: CentOS* 7.4 (64 bit), Ubuntu* 16.04.3 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 (deep learning functions support for Intel CPU/GPU/NCS/GNA).

Optimization Notice

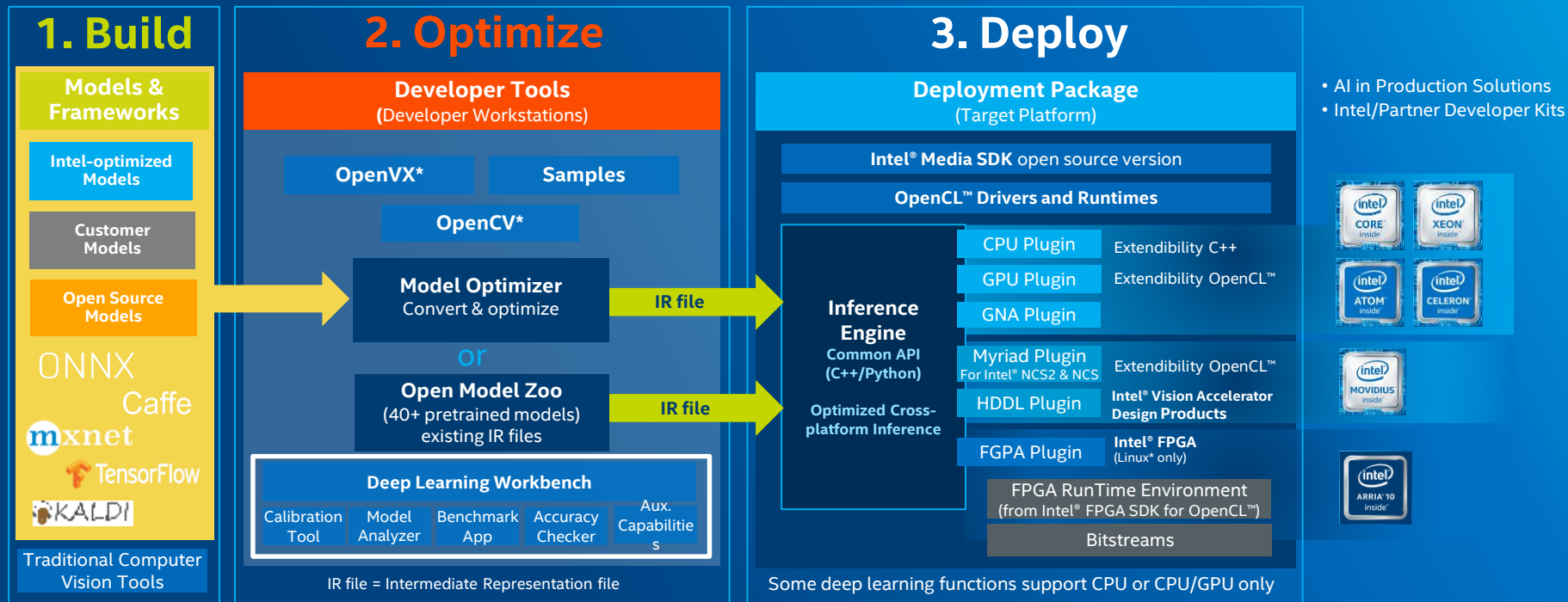
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

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



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



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



Quick Guide: What's Inside the Intel Distribution vs Open Source version

Tool/Component	Intel® Distribution of OpenVINO™ toolkit	OpenVINO™ toolkit (open source)	Open Source Directory
Installer (including necessary drivers)	✓		
Intel® Deep Learning Deployment toolkit			
Model Optimizer	✓	✓	/opencv/dldt/tree/2018/model-optimizer
Inference Engine	✓	✓	/opencv/dldt/tree/2018/inference-engine
Intel CPU plug-in	✓ Intel® Math Kernel Library (Intel® MKL) only ¹	✓ BLAS, Intel® MKL ¹ , jit (Intel MKL)	/opencv/dldt/tree/2019/inference-engine
Intel GPU (Intel® Processor Graphics) plug-in	✓	✓	/opencv/dldt/tree/2019/inference-engine
Heterogeneous plug-in	✓	✓	/opencv/dldt/tree/2019/inference-engine
Intel GNA plug-in	✓	✓	/opencv/dldt/tree/2019/inference-engine
Intel® FPGA plug-in	✓		
Intel® Neural Compute Stick (1 & 2) VPU plug-in	✓	✓	/opencv/dldt/tree/2019/inference-engine
Intel® Vision Accelerator based on Movidius plug-in	✓		
40+ Pretrained Models - incl. Model Zoo (IR models that run in IE + open sources models)	✓	✓	https://github.com/opencv/open_model_zoo
Samples (APIs)	✓	✓	/opencv/dldt/tree/2018/inference-engine
Demos	✓	✓	https://github.com/opencv/open_model_zoo
Traditional Computer Vision			
OpenCV*	✓	✓	https://github.com/opencv/opencv
OpenVX (with samples)	✓		
Intel® Media SDK	✓	✓ ²	https://github.com/Intel-Media-SDK/MediaSDK
OpenCL™ Drivers & Runtimes	✓	✓ ²	https://github.com/intel/compute-runtime
FPGA RunTime Environment, Deep Learning Acceleration & Bitstreams (Linux* only)	✓		

Optimization Notice

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

¹Intel MKL is not open source but does provide the best performance

²Refer to readme file for validated versions



ACCELERATE PERFORMANCE FOR DEEP LEARNING INFERENCE



Intel® Deep Learning Deployment Toolkit

For Deep Learning Inference

Model Optimizer

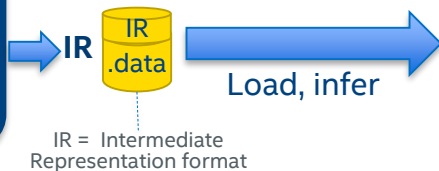
- **What it is:** A Python*-based tool to import trained models and convert them to Intermediate representation.
- **Why important:** Optimizes for performance/space with conservative topology transformations; biggest boost is from conversion to data types matching hardware.



Trained Models

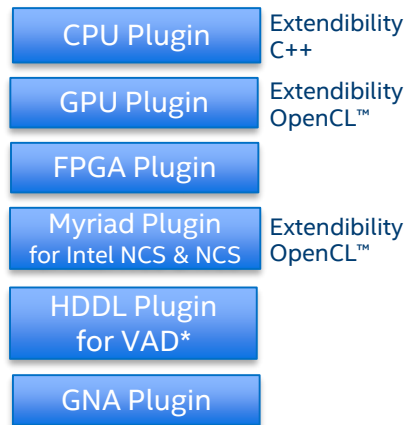
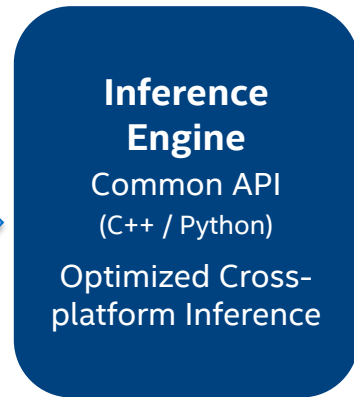


Model Optimizer
Convert & Optimize



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 GPU/Intel® Processor Graphics, Intel® NCS = Intel® Neural Compute Stick (VPU)

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

Optimization Notice

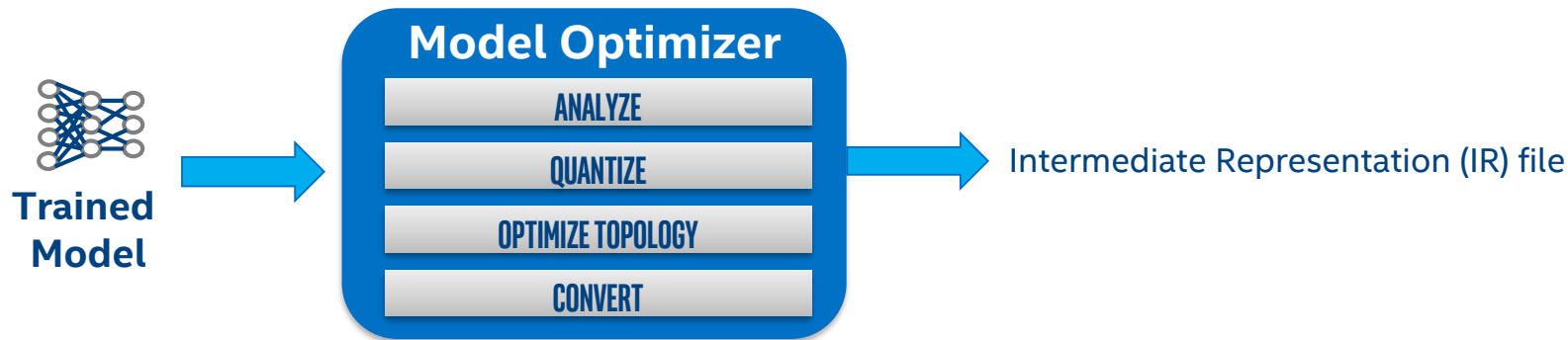
Copyright © 2019, Intel Corporation. All rights reserved.

*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



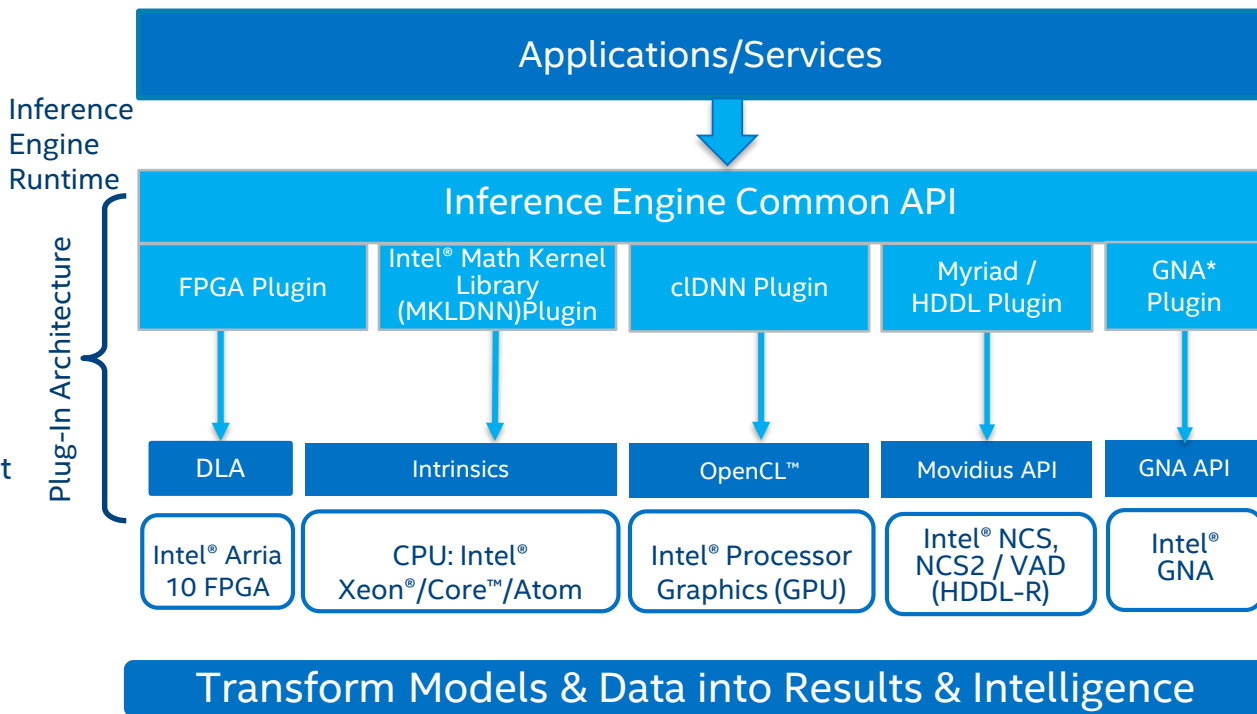
Improve Performance with Model Optimizer



- 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.

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 (CPU inference remains at FP32)



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



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

Binary Models

- | | | |
|-------------------------------|----------------------------|-------------------|
| ▪ Face Detection Binary | ▪ Vehicle Detection Binary | ▪ ResNet50 Binary |
| ▪ Pedestrian Detection Binary | | |

Save Time with Deep Learning Samples

Use Model Optimizer & Inference Engine for Public Models & Intel Pretrained Models

- Object Detection
- Standard & Pipelined Image Classification
- Security Barrier
- Object Detection SSD
- Neural Style Transfer
- Object Detection for Single Shot Multibox Detector using Asynch API+
- Hello Infer Classification
- Interactive Face Detection
- Image Segmentation
- Validation Application
- Multi-channel Face Detection

WHAT'S NEW IN OPENVINO 2019 R3

NETWORK LOADING OPTIMIZATIONS

Reduced model load times for faster performance

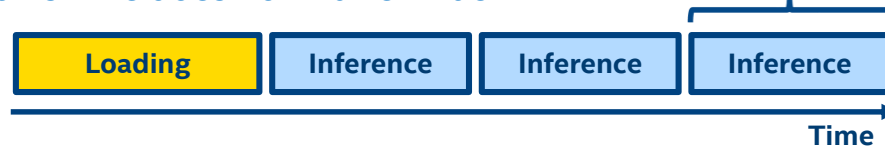
Audience: Helpful when shape size changes from inference to inference, and resizing is undesirable (e.g., leads to accuracy degradation)

Problem: Shape change requires reloading of the model which can be slow

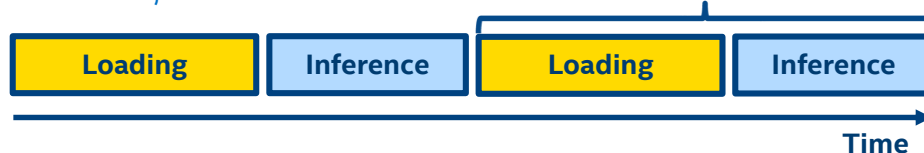
UseCase: Input shape is defined by a previous network in the pipeline (i.e., in the case of object detection/classification), or ROI is defined by operator (common case in medical applications)

Performance

Usual case: loading happens only once, so its time does not matter much



Changing shape case: loading happens before each inference, so its time matters



Loading time optimization decreases total inference time in this case



CPU

iGPU

VPU

FPGA

GNA

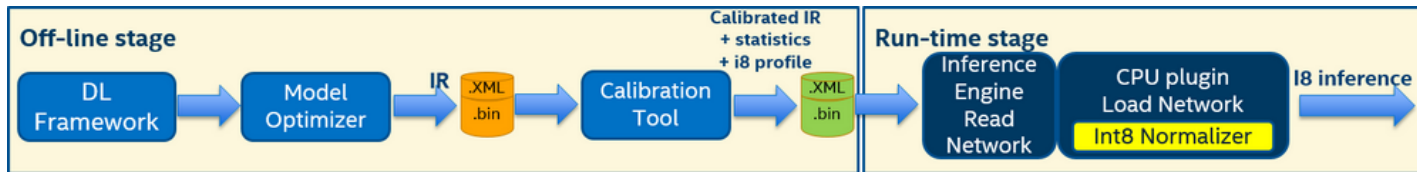
Win 10

Win Serv

Linux

Mac

FP32->INT8 MODELS CALIBRATION FLOW



Calibration tool

- Selects of quantization ranges & base layers (Convolution, ScaleShift, FullyConnected)

Inference Engine Normalizer

- Defines the quantization rules: float->[0..255] or float ->[-127..127]
- Converts neighbor layers to int8, build chains executed in int8
- Interpret ranges depending on next layers

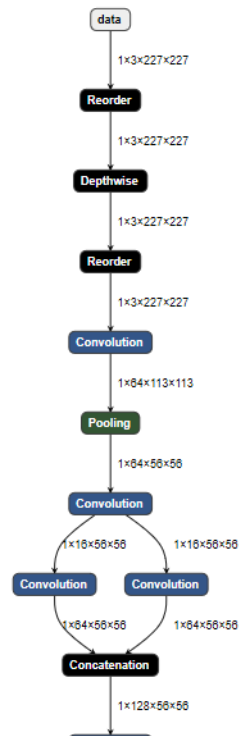
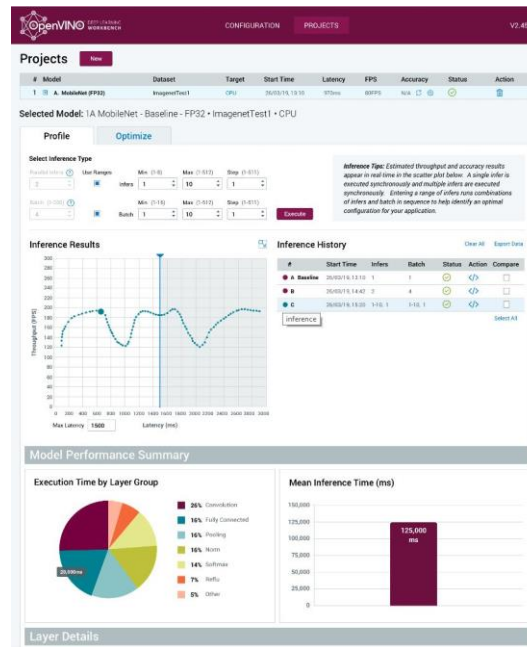
Calibrated IR

- Stores unmodified topology and original weights
- Stores ranges for quantization in Normalizer
- Can be consumed by any plugin supporting FP32 IR and continue to work

Deep Learning Workbench (Preview)

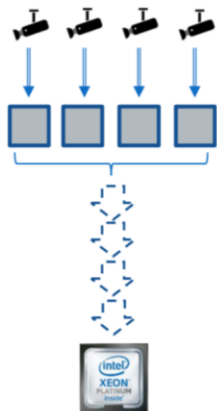
Deep Learning Workbench capabilities

- 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

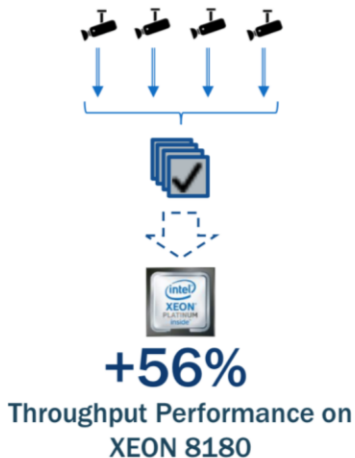


BATCH VS ASYNC EXECUTION

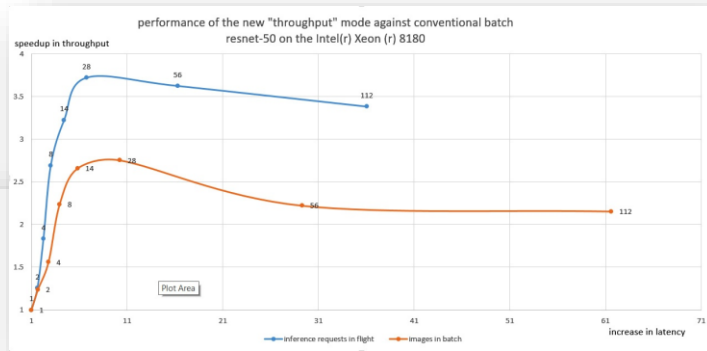
SERIALIZED EXECUTION



BATCHED EXECUTION



PARALLEL EXECUTION

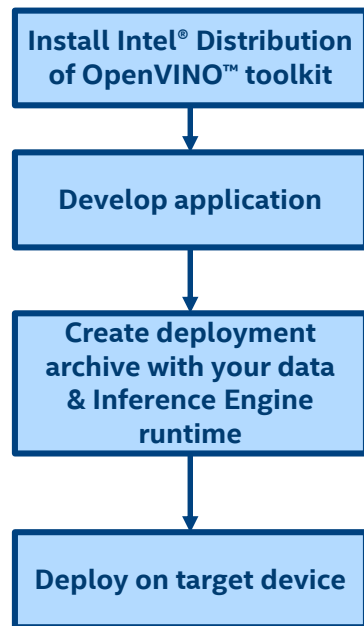


<https://www.intel.ai/cpu-inference-performance-boost-openvino/>

Command Line Deployment Manager

- Generate an optimal, minimized runtime package for selected target device.
- Deploy Inference Engine with pre-compiled application-specific data such as models, config, and a subset of required hardware plugins.
- Achieve deployment footprint to be several times smaller than the development footprint.

For more details, see [Introduction to CLI Deployment Manager](#)



Target	Size, MB
CPU only	65
GPU only	26
Myriad only	22
HDDL only	27
GNA only	15

Measurements for deployment archives based on 2019 R3

Inference Engine Centric API

- New IE **Core** class which provides the same functionality as plugins:
 - QueryNetwork(device), LoadNetwork(device), ImportNetwork(device), AddExtension(device)
 - SetConfig(device), SetLogCallback, GetVersions(device)
- Registers and manages plugins (no needs to create plugins explicitly):
 - Via configuration file for default IE plugins
- Does not allow to get reference to particular plugin
- API operates with devices, not with plugins
- Lazy initialization: plugins are initialized only after device is used in Core

MULTI-DEVICE SUPPORT

Automatic load-balancing between devices (inference requests level)

Any combinations of inference devices

- CPU + iGPU (very ubiquitous!)
- Multiple NCS2, etc

As easy as “-d **MULTI**:HDDL,GPU” cmd-line option of your favorite OpenVINO sample

C++ example (Python is similar):

```
// New IE-centric API
Core ie;

ExecutableNetwork exec = ie.LoadNetwork(network, {{"DEVICE_PRIORITIES", "HDDL,GPU"}}, "MULTI");

// Old plugin-centric API
auto plugin = PluginDispatcher().getPluginByDevice("MULTI:CPU,GPU");

ExecutableNetwork executable_network = plugin.LoadNetwork(network, config);
```



Instance Segmentation (MaskRCNN)

UseCase: Crowded scenes, close-up objects and many other scenarios require precise boundary of the object (vs just bounding box) to be solved. MaskRCNN implementations help provide such segmentations.



Image Retrieval

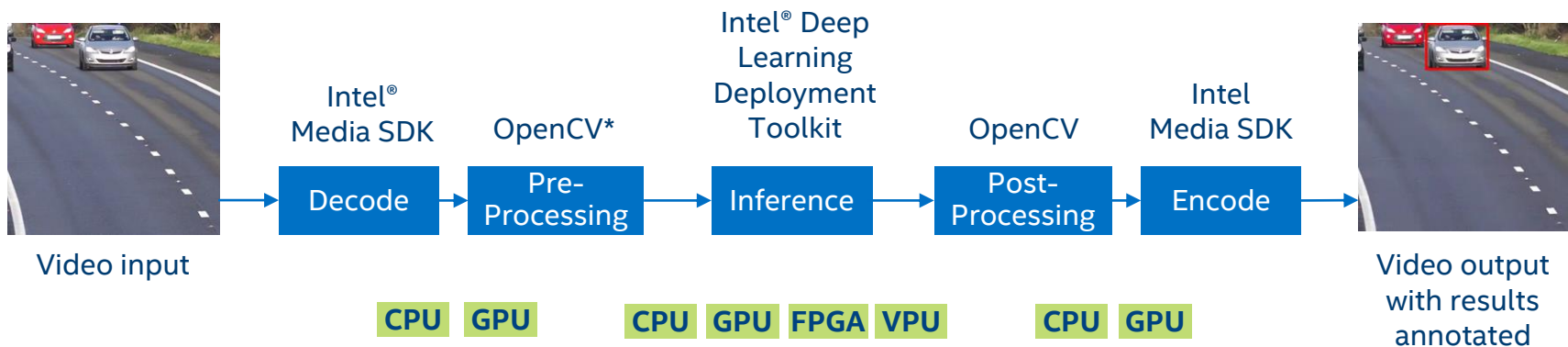
UseCase: Matches an artificial patterned image to its real-world instance, such as a textile. The model achieves top1/top5/top10 accuracy of 82%/92%/93% on 100 validation videos (with ~2.5K patterns in the gallery)



Super Resolution

UseCase: A common operation on MFPs is to upscale an image from 150 ppi (received via network) to 600 ppi (good for printing). DL-based super resolution produces significantly better output image in comparison to the bi-cubic upscale (26.42 vs 20.16 PSNR)

End-to-End Vision Workflow



Call to Action, Resources

Download Free ▶

[Intel® Distribution of OpenVINO™ toolkit](#)

Get started quickly with:

- [Developer resources](#)
- [Online webinars, tool how-tos & quick tips](#)
- [Hands-on developer workshops](#)

Support

- Connect with Intel engineers & computer vision experts at the public [Community Forum](#)

<https://software.intel.com/en-us/openvino-toolkit>



Select Intel customers may contact their Intel representative for issues beyond forum support.

[Optimization Notice](#)

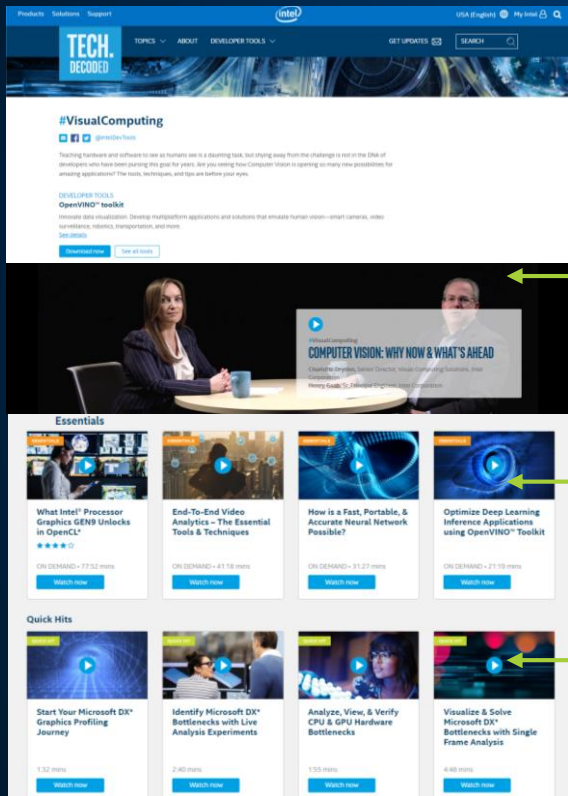
Copyright © 2019, Intel Corporation. All rights reserved.

*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.



Technical Specifications - Intel® Distribution of OpenVINO™ toolkit

	Intel® Platforms	Compatible Operating Systems
Target Solution Platforms	CPU <ul style="list-style-type: none"> 6th-10th generation Intel® Core™ and Xeon® processors 1st and 2nd generation Intel® Xeon® Scalable processors 	<ul style="list-style-type: none"> Ubuntu* 18.04.3 LTS (64 bit) Microsoft Windows* 10 (64 bit) CentOS* 7.4 (64 bit) macOS* 10.13 & 10.14 (64 bit)
	<ul style="list-style-type: none"> Intel® Pentium® processor N4200/5, N3350/5, N3450/5 with Intel® HD Graphics 	<ul style="list-style-type: none"> Yocto Project* Poky Jethro v2.0.3 (64 bit)
	Iris® Pro & Intel® HD Graphics <ul style="list-style-type: none"> 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 graphics¹) 	<ul style="list-style-type: none"> Ubuntu 18.04.3 LTS (64 bit) Windows 10 (64 bit) CentOS 7.4 (64 bit)
	FPGA <ul style="list-style-type: none"> 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) 	<ul style="list-style-type: none"> 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)
Development Platforms	Intel® Vision Accelerator Design Products <ul style="list-style-type: none"> Intel® Vision Accelerator Design with Intel® Arria10 FPGA 	<ul style="list-style-type: none"> Ubuntu 18.04.2 LTS (64 bit)
	<ul style="list-style-type: none"> Intel® Vision Accelerator Design with Intel® Movidius™ VPUs 	<ul style="list-style-type: none"> Ubuntu 8.04.3 LTS (64 bit) Windows 10 (64 bit)
Additional Software Requirements	<ul style="list-style-type: none"> 6th-10th generation Intel® Core™ and Intel® Xeon® processors 1st and 2nd generation Intel® Xeon® Scalable processors 	<ul style="list-style-type: none"> Ubuntu* 18.04.3 LTS (64 bit) Windows* 10 (64 bit) CentOS* 7.4 (64 bit) macOS* 10.13 & 10.14 (64 bit)
	Linux* build environment required components <ul style="list-style-type: none"> OpenCV 3.4 or higher CMake* 2.8 or higher 	<ul style="list-style-type: none"> GNU Compiler Collection (GCC) 3.4 or higher Python* 3.4 or higher
External Dependencies/Additional Software	Microsoft Windows* build environment required components <ul style="list-style-type: none"> Intel® HD Graphics Driver (latest version)[†] Intel® C++ Compiler 2017 Update 4 Python 3.4 or higher 	<ul style="list-style-type: none"> OpenCV 3.4 or higher CMake 2.8 or higher Microsoft Visual Studio* 2015
		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.

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



IOT DEV CLOUD BY INTEL®

INTEL IOT DEVELOPER CLOUD

Develop Using Jupyter Notebook

Terminal Access to Device

Choose Hardware(s) for Workload
Compute/ Inference
(CPU/GPU, VPU, Accelerator)

Choose CSD Edge Compute
(Amazon Green Grass, Azure Edge)

Scalable XEON
based dev server



Dynamic
Storage

Login to
IoT DevCloud



Developer



Dedicated
Access



Queue
(H/W, S/W)



Visualize
Output

IoT Devkits, RRKs, OxM Hardware



- Pre-installed s/w stacks (OS, SDK, Tools)
- 50GB Dynamic storage allocated to all users
- Metering, Profiling tools available to easily evaluate hardware
- Examples to easily get started

INTEL IOT DEVELOPER CLOUD

Cloud-based Heterogeneous Edge Development Environment



INTEL IOT DEVELOPER CLOUD

Cloud-based Heterogeneous Edge Development Environment

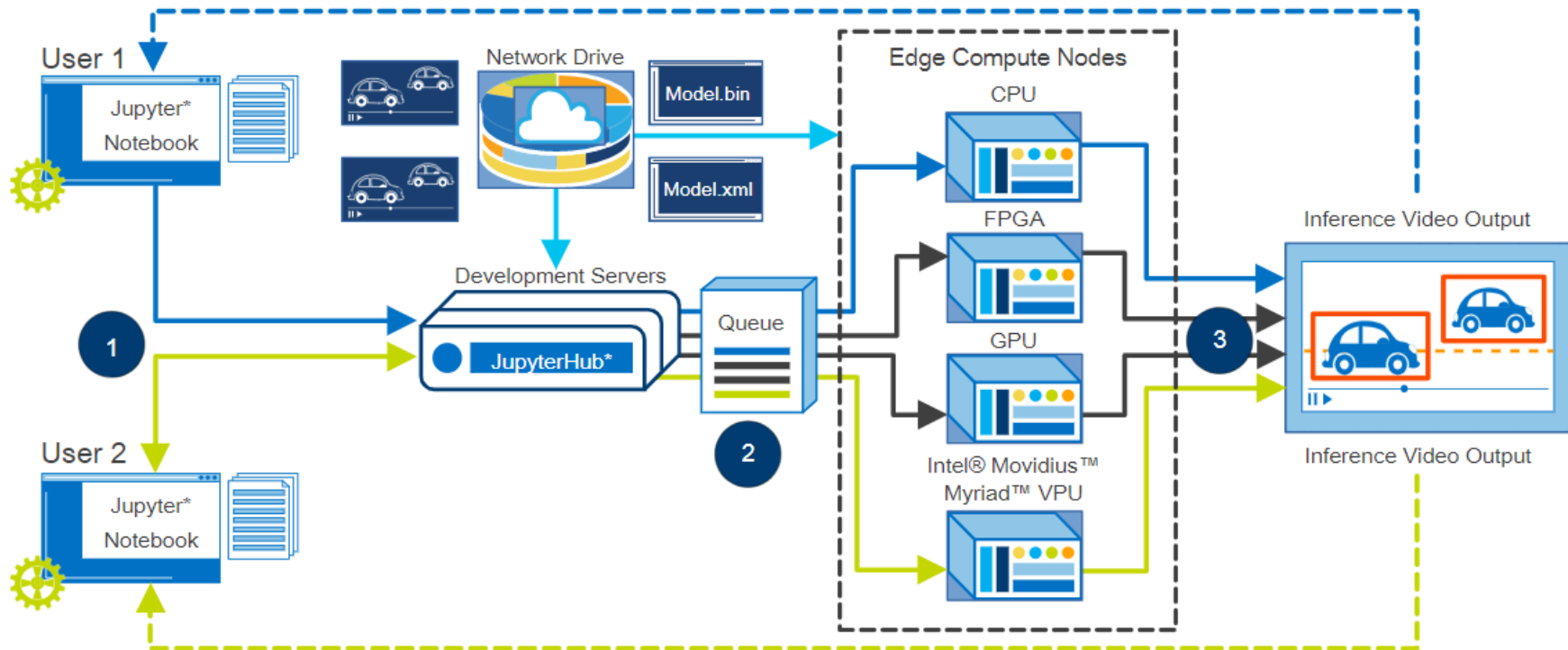
Key Benefits

- Easily evaluate the power of Intel's heterogeneous edge compute processors: CPU, CPU with Integrated Graphics, FPGA, VPU
- Learn to develop vision inference solutions with OpenVINO™ toolkit
- Evaluate and compare various IoT Developer Kit offerings, software stacks and edge stacks
- Getting started immediately without waiting to buy edge hardware

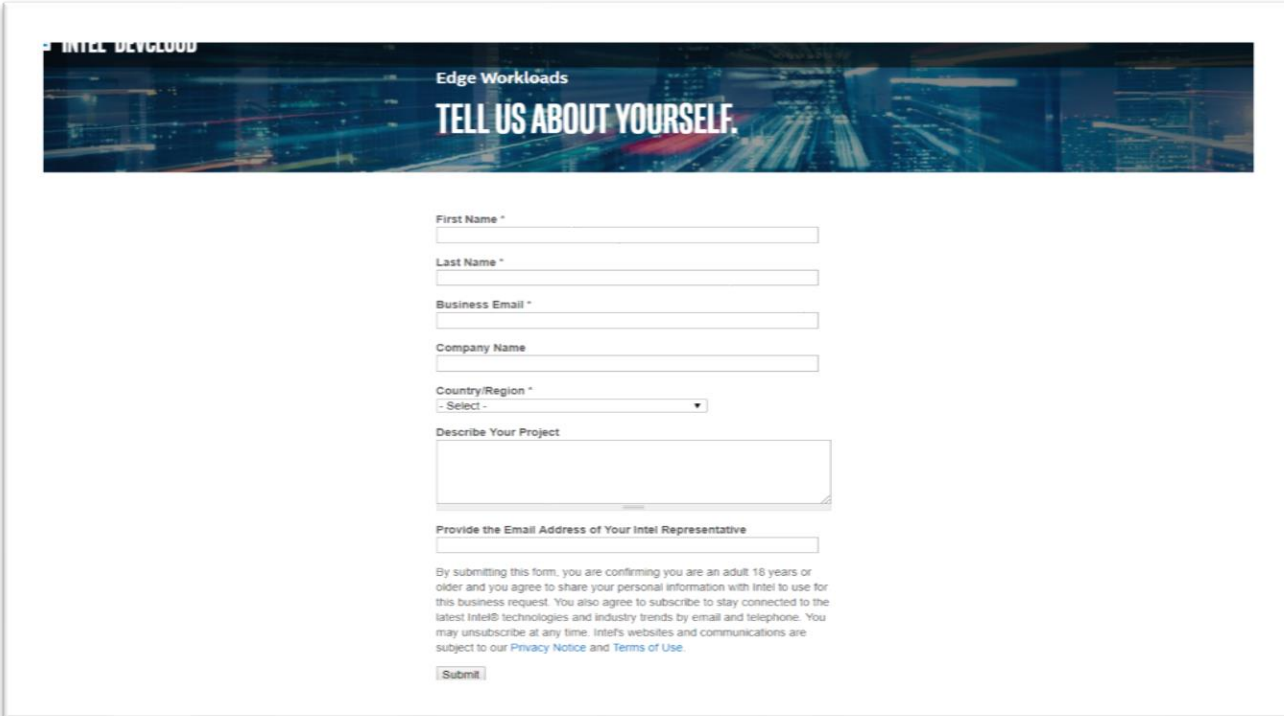
Business Opportunity

- Scale openVINO™ Toolkit enablement across OEM/ODM, SIs, ISVs and Developers
- Scale workshops and trainings WW without the need to ship hardware
- Understand how developers are using the various Edge offerings from Intel and partners
- Showcase customer use cases that run best on IA (sales enablement tool, architecture conversion)

INTEL IOT DEVELOPER CLOUD



PERSONAL ACCOUNT APPLICATION



The screenshot shows a web form for signing up for Intel DevCloud Edge Workloads. At the top, there is a banner with the text "INTEL DEVCLOUD" on the left, "Edge Workloads" in the center, and "TELL US ABOUT YOURSELF." on the right. Below the banner, the form contains several input fields: "First Name *", "Last Name *", "Business Email *", "Company Name", and "Country/Region *" (with a dropdown menu). There is also a large text area labeled "Describe Your Project". Below this, there is a field for "Provide the Email Address of Your Intel Representative". At the bottom, there is a "Submit" button. A disclaimer paragraph is located above the "Submit" button, stating that by submitting the form, the user confirms they are an adult 16 years or older and agrees to share their personal information with Intel for business purposes, including staying connected to Intel's technologies and industry trends via email and telephone. The disclaimer also mentions that the user may unsubscribe at any time and that Intel's websites and communications are subject to its Privacy Notice and Terms of Use.

INTEL DEVCLOUD

Edge Workloads

TELL US ABOUT YOURSELF.

First Name *

Last Name *

Business Email *

Company Name

Country/Region *

Select -

Describe Your Project

Provide the Email Address of Your Intel Representative

By submitting this form, you are confirming you are an adult 16 years or older and you agree to share your personal information with Intel to use for this business request. You also agree to subscribe to stay connected to the latest Intel® technologies and industry trends by email and telephone. You may unsubscribe at any time. Intel's websites and communications are subject to our [Privacy Notice](#) and [Terms of Use](#).

Submit

<https://software.intel.com/en-us/devcloud/edge/sign-up>

Welcome to the IoT DevCloud by Intel!

Powered by Colfax Connect

INTRODUCTION

to the IoT DevCloud by Intel



EXPERIENCE

the IoT DevCloud by Intel in action



CONNECT

from your home computer to the cloud



- Introductory content
- Getting started with Dev cloud
- Tutorials
- Samples that showcase technology
- Develop/ test own code/ models
- Manage storage

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

Optimization Notice

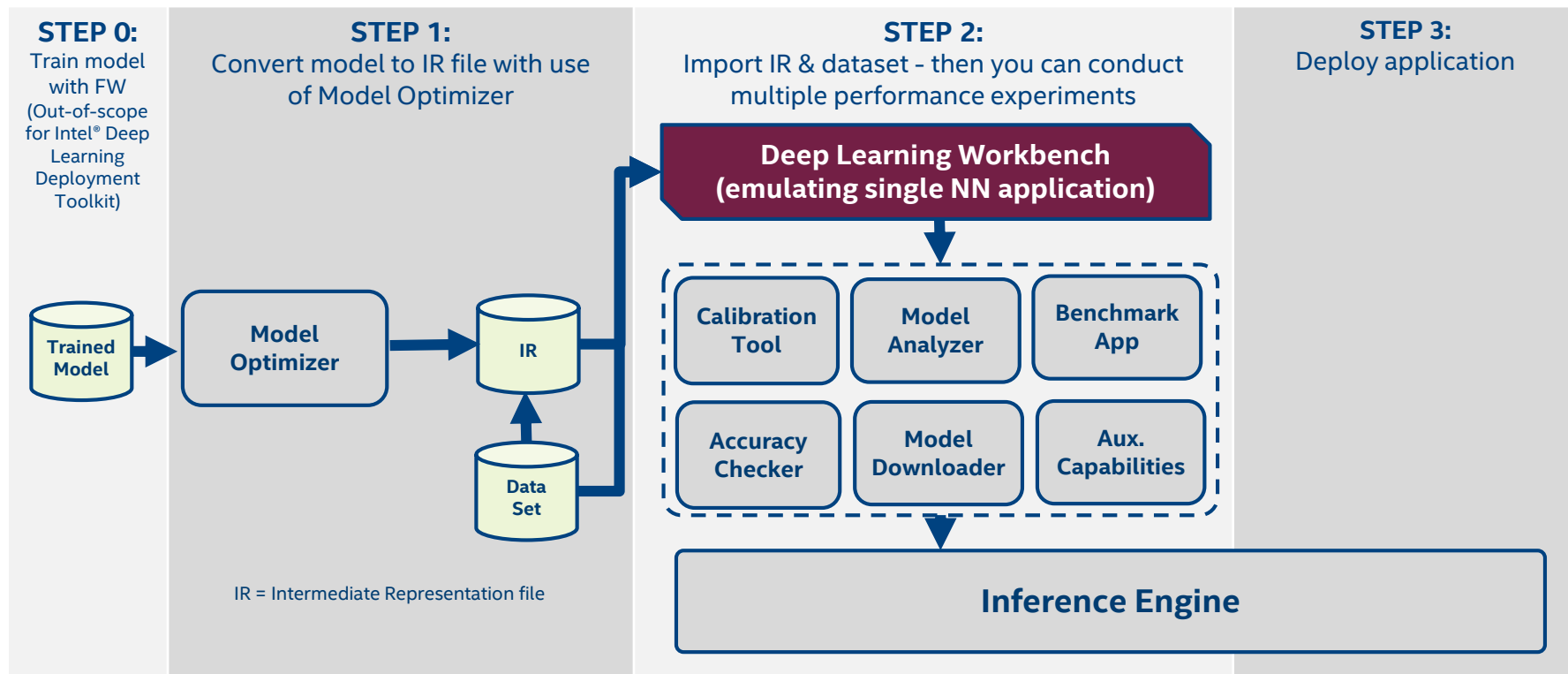
Copyright © 2019, Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.

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






Deep Learning Workbench Data Flow



DL Workbench Look and Feel (INT8, profiling, accuracy, graph)

 **OpenVINO™** Open Visual Inference for Neural Networks

CONFIGURATION

Version: 1.0.2151.226e4275

Projects

#	Model	Dataset	Target	Precision	Start Time	Latency (ms)	Throughput (FPS)	Accuracy (%)	Status	Actions
1	squeezeNet1.1	ImageNet_1000_224x224	CPU	FP32	08/07/19, 04:40	5.87	433.19	54.6		
	squeezeNet1.1 - Int 8			MIXED	08/07/19, 04:42	60.81	1,022.19	54.5		

Selected Model: squeezeNet1.1 - Int 8 • ImageNet_1000_224x224 • CPU

Profile **Optimize**

Select Inference Type

Parallel streams: 2

Use ranges: ☒

Batch(1-256): 2

Min(1-24): 2

Max(1-24): 8

Step(1-23): 2

Min(1-256): 2

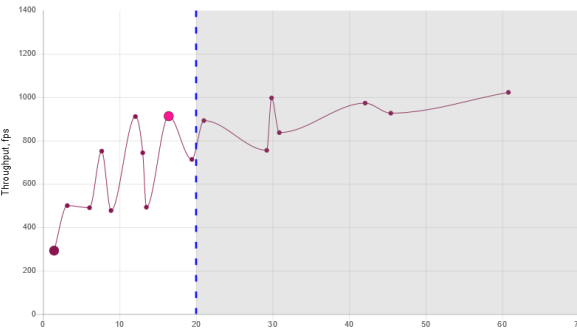
Max(1-256): 8

Step(1-255): 2

Execute

Inference Tips: Estimated throughput and accuracy results appear in real-time in the scatter plot below. A single infer is executed synchronously and multiple inferences are executed asynchronously. Entering a range of inferences runs combinations of inferences and batch in sequence to help identify an optimal configuration for your application.

Inference Results



In range 0-1000:
Max. Latency 20

Inference History

#	Start Time	Parallel Streams	Batch	Status	Filter	
A	Baseline	08/07/19, 04:42	Sync	1		<input type="checkbox"/>
B	Group Inference	08/07/19, 04:49	Group: 2-8, 2	2-8, 2		<input type="checkbox"/>

