

OPENVINO – NEW DIRECTIONS OF 2019

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, NLP, Audio, text) & more.



HIGH PERFORMANCE, PERFORM AI AT THE EDGE



STREAMLINED & OPTIMIZED DEEP LEARNING INFERENCE

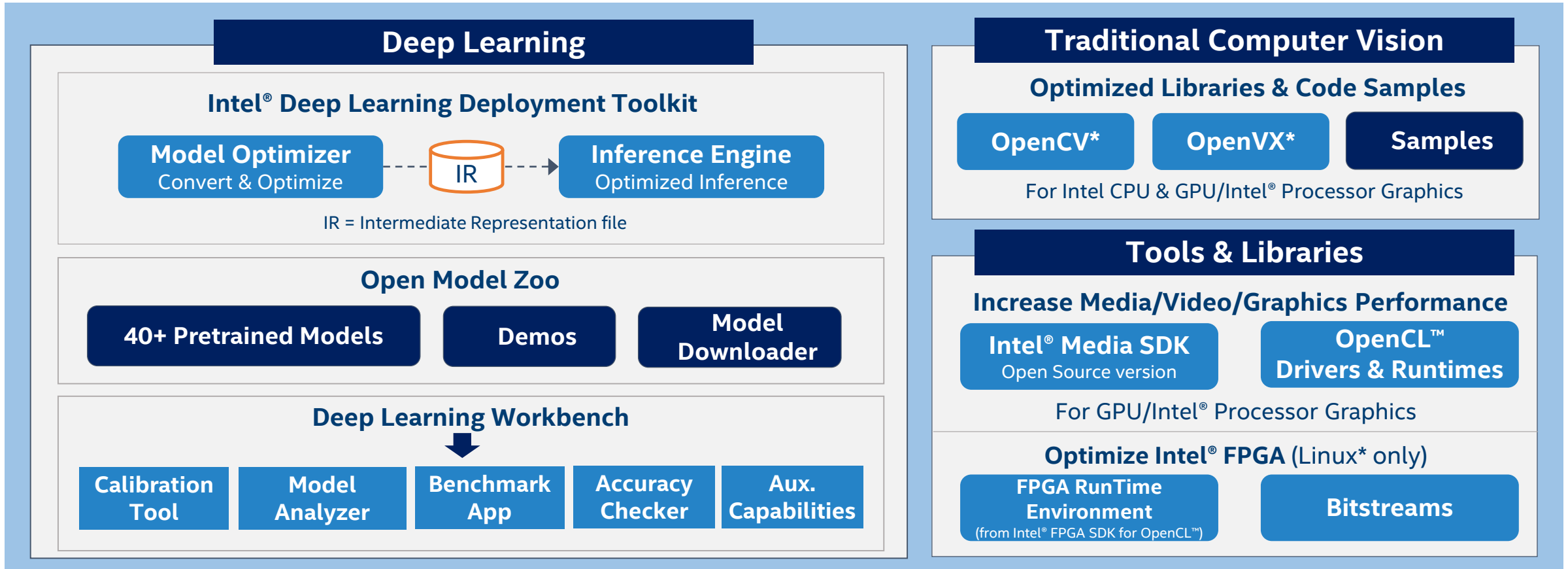


HETEROGENEOUS, CROSS-PLATFORM FLEXIBILITY

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

Open Source version ▶ 01.org/openvinotoolkit

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)



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

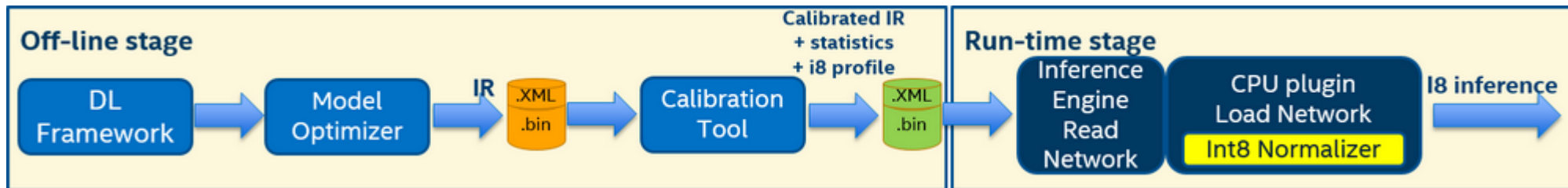
Int8 support on CPU

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

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

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

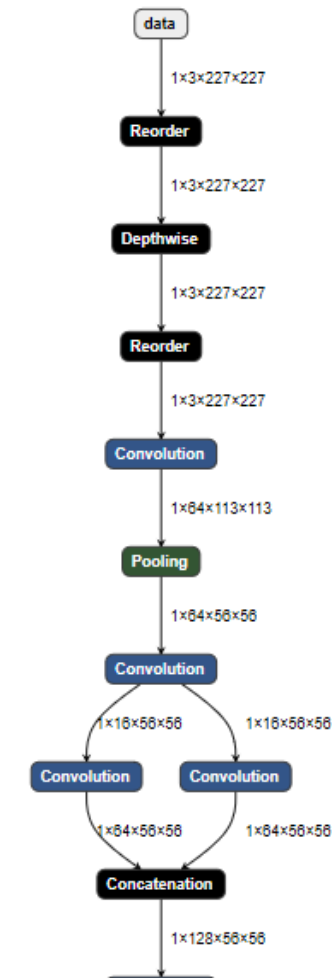
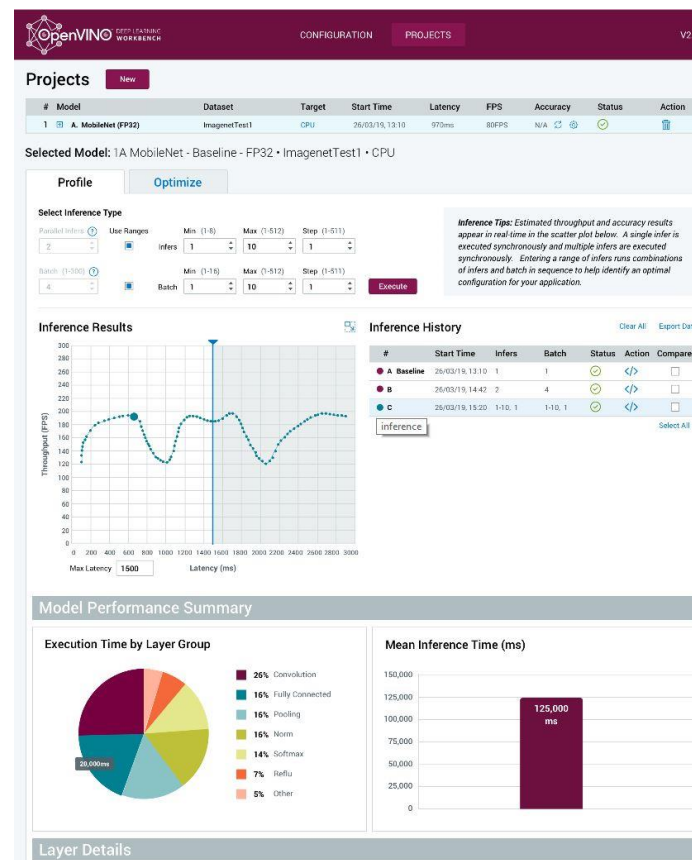


- Calibration in **“simplified” mode** – see the maximum of potential performance gain from Int8 without accuracy calculation
- Pass full calibration process to get working Int8 model with accuracy statistics

Deep Learning Workbench (Preview)

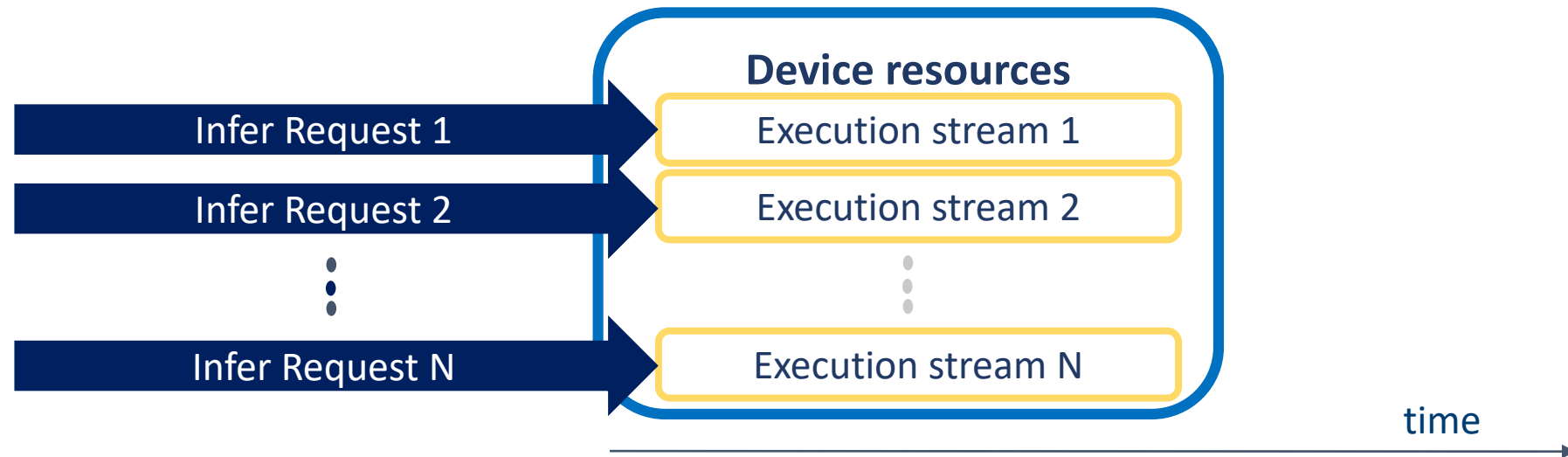
Deep Learning Workbench capabilities

- 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



Inference Engine “Throughput” mode for CPU and iGPU

- **Latency** – inference time of 1 frame (ms).
- **Throughput** – overall amount of frames inferred per 1 second (FPS)
- **“Throughput” mode** allows the Inference Engine to efficiently run multiple infer requests simultaneously, greatly improving the overall throughput.
- Device resources are divided into execution **“streams”** – parts which runs infer requests in parallel



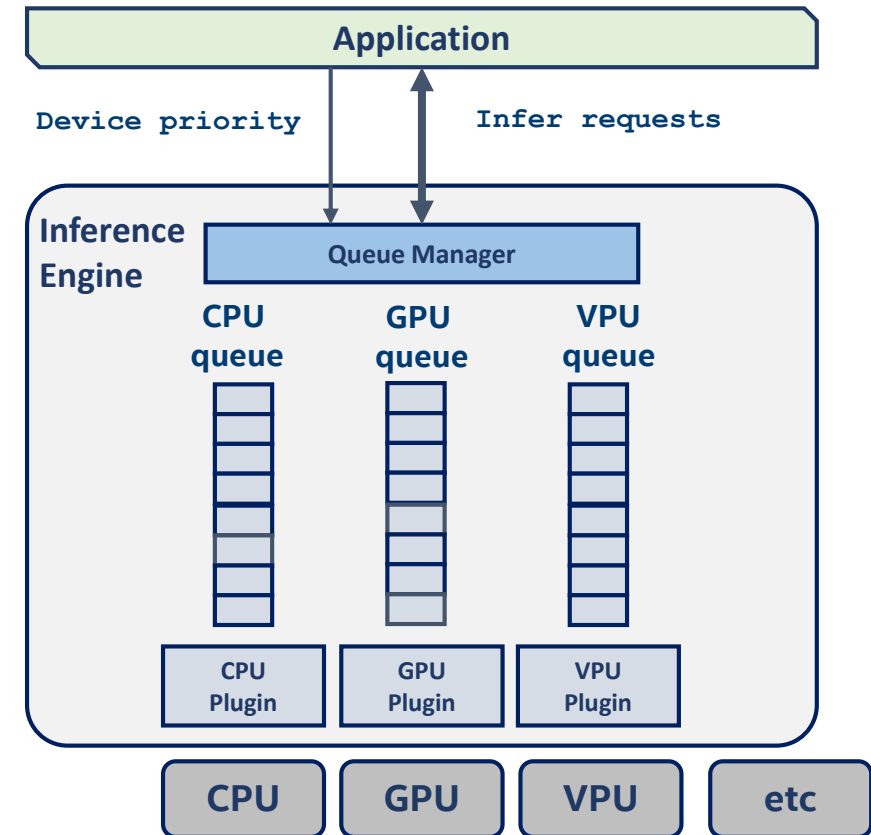
Inference Engine Multi-Device Support

Automatic load-balancing between devices (inference requests level)

- Any combinations of the following devices are supported (CPU, iGPU, VPU, HDDL)
- As easy as “-d **MULTI**:CPU,GPU” for cmd-line option of your favorite sample/demo
- C++ example (Python is similar)

```
// New IE-centric API
Core ie;
ExecutableNetwork exec = ie.LoadNetwork(network, {{"DEVICE_PRIORITIES",
"CPU,GPU"}}, "MULTI");

// Old plugin-centric API
auto plugin = PluginDispatcher().getPluginByDevice("MULTI:CPU,GPU");
ExecutableNetwork executable_network = plugin.LoadNetwork(network,
config);
```



Speed Deployment with Pretrained Models & Demos

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

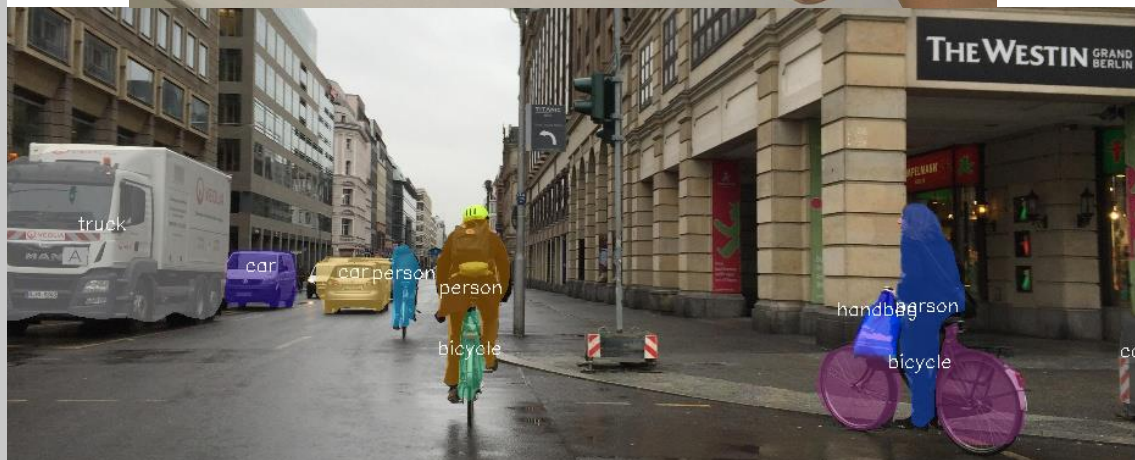
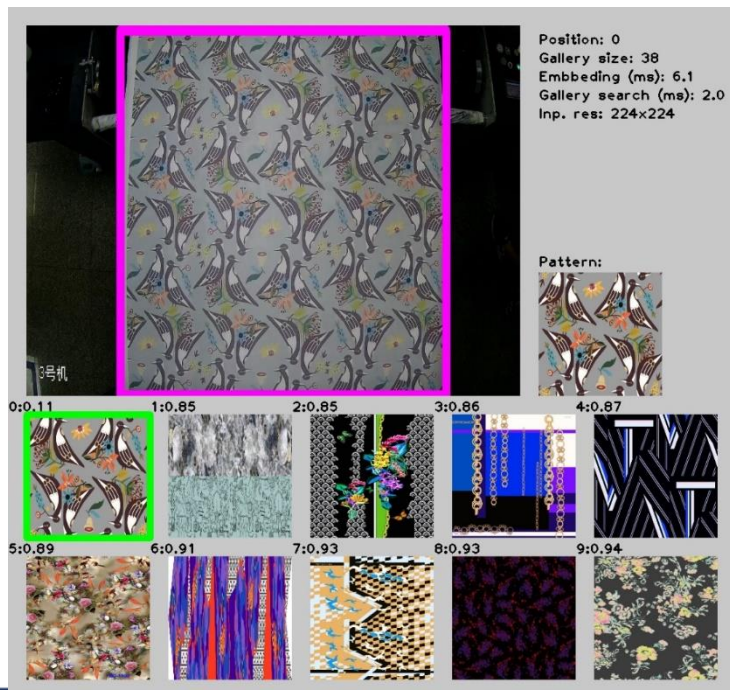
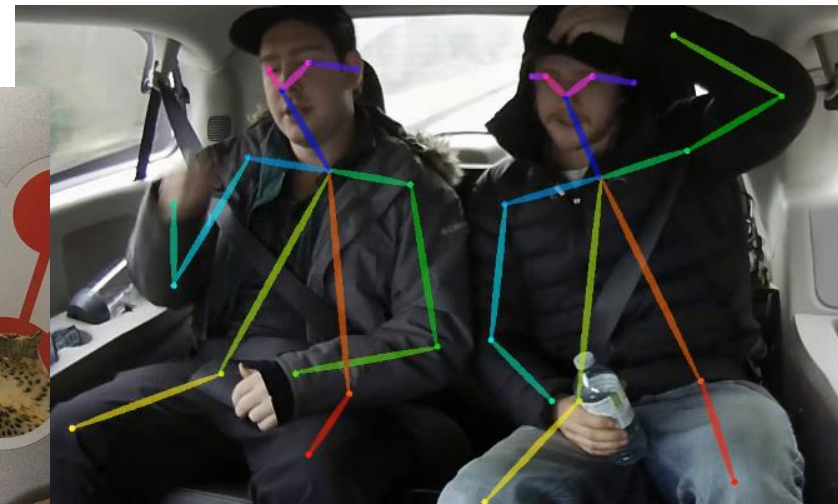
Pretrained Models in Intel® Distribution of OpenVINO™ toolkit

- 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

Open Model Zoo – Use Case Examples



Other OpenVINO-related Open Source projects

OpenVINO Training Extensions

Convenient environment to train Deep Learning models and convert them using OpenVINO™ Toolkit for optimized inference

PyTorch

- Action recognition
- Face recognition
- Human pose estimation
- Instance segmentation
- Object Detection
- Face Detection
- Person Vehicle Bike Detector
- Segmentation of thoracic organs
- Super resolution

.....

TensorFlow

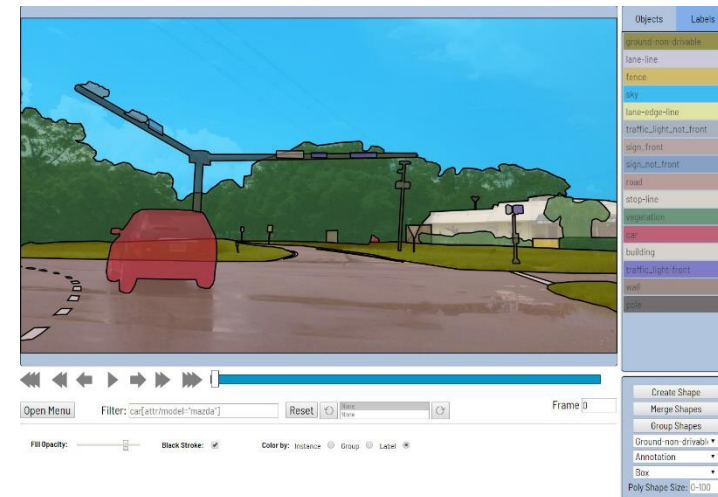
- Action Detection
- License Plate Recognition
- Person Vehicle Bike Detector
- SSD Object Detection
- Text recognition

.....

github.com/opencv/opencvino_training_extensions

Computer Vision Annotation Tool (CVAT)

Free, online, interactive video and image annotation tool for computer vision.



github.com/opencv/cvat

Other improvements of 2019

- Model Loading Optimization
- CLI Deployment Manager Tool
- New Inference Engine centric APIs
- Supports serialized FP16 Intermediate Representation
- Support of use-cases for machine translation, natural language processing, and speech processing and recognition
- Binary distribution methods (yum, apt, docker)
- Open Sourced VPU (NCS and NCS2) plugins
- Preview of VPU custom layers (NCS and NCS2) support
- Read more at the release Notes:

<https://software.intel.com/en-us/articles/OpenVINO-RelNotes>

To be continued...

Q&A

DISCLOSURES

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

The products and services described may contain defects or errors known as errata which may cause deviations from published specifications. Current characterized errata are available on request. No product or component can be absolutely secure.

Copies of documents which have an order number and are referenced in this document may be obtained by calling 1-800-548-4725 or by visiting www.intel.com/design/literature.htm.

Intel, the Intel logo, [List the Intel trademarks in your document] are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries.

