Using the Intel Distribution of the OpenVINO Toolkit for Deploying Accelerated Deep Learning Applications [2021.1]

intel.

Notices and disclaimer

- INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

- Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at www.intel.com.

- 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.

- © Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

intel

Optimization notice

AI CHANGING AND ENABLING EVERY INDUSTRY

AI software market is projected to reach USD 126.0 billion in annual worldwide revenue by 2025[1]

Deep learning software revenue is estimated to grow to USD 67.2 billion by 2025[2]

Global deep learning chip market is expected to reach USD 29.4 billion by 2025[3]

AGRICULTURE
Achieve higher yields and increase efficiency

ENERGY
Maximize production and uptime

EDUCATION
Transform the learning experience

GOVERNMENT
Enhance safety, research, and more

FINANCE
Turn data into valuable intelligence

HEALTH
Revolutionize patient outcomes

INDUSTRIAL
Empower truly intelligent Industry 4.0

MEDIA
Create thrilling experiences

RETAIL
Transform stores and inventory

SMART HOME
Enable homes that see, hear, and respond

TELECOM
Drive network and operational efficiency

TRANSPORTATION
Automated driving

1. Tractica, Artificial Intelligence Software Market, 2020

2. Tractica, deep learning research, 2018

3. AlliedMarketResearch, Deep Learning Chip Market, 2018

Deep learning breakthroughs and opportunities

Machines able to meet or exceed human image and speech recognition

Image Recognition
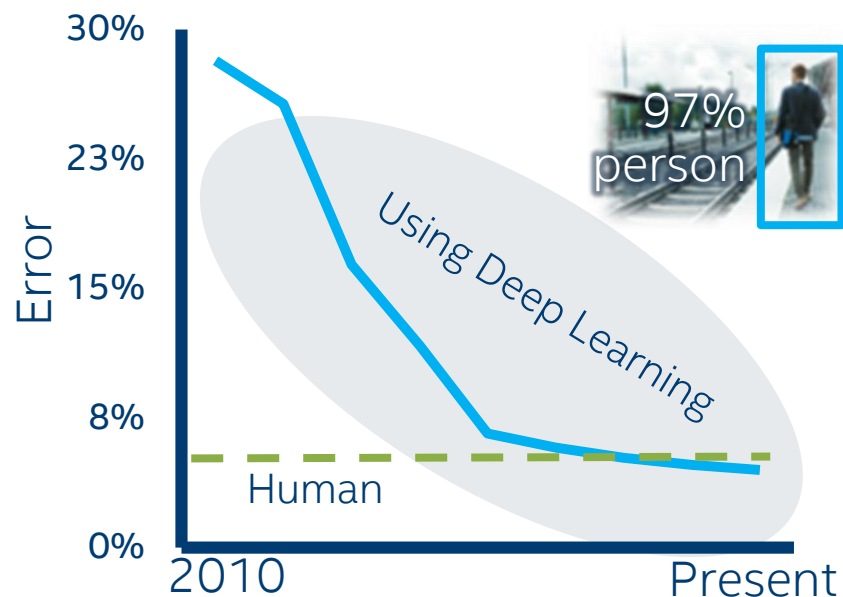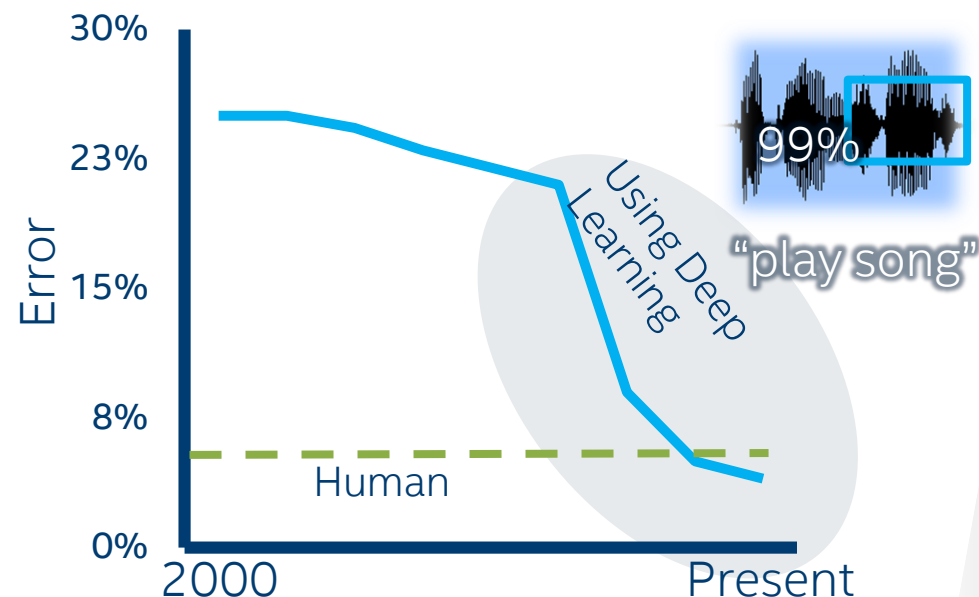
Error

30%
23%
15%
8%
0%

Using Deep Learning

97% person

Human

2010 — Present

Speech Recognition

Error

30%
23%
15%
8%
0%

Using Deep Learning

99% "play song"

Human

2000 — Present

ADDITIONAL ECONOMIC IMPACT DRIVEN BY AI

$13 TRILLION IN 2030

Deep Learning Development Cycle

Data acquisition and organization

Create models

Integrate trained models with application code

Adjust models to meet performance and accuracy objectives

Import

Build
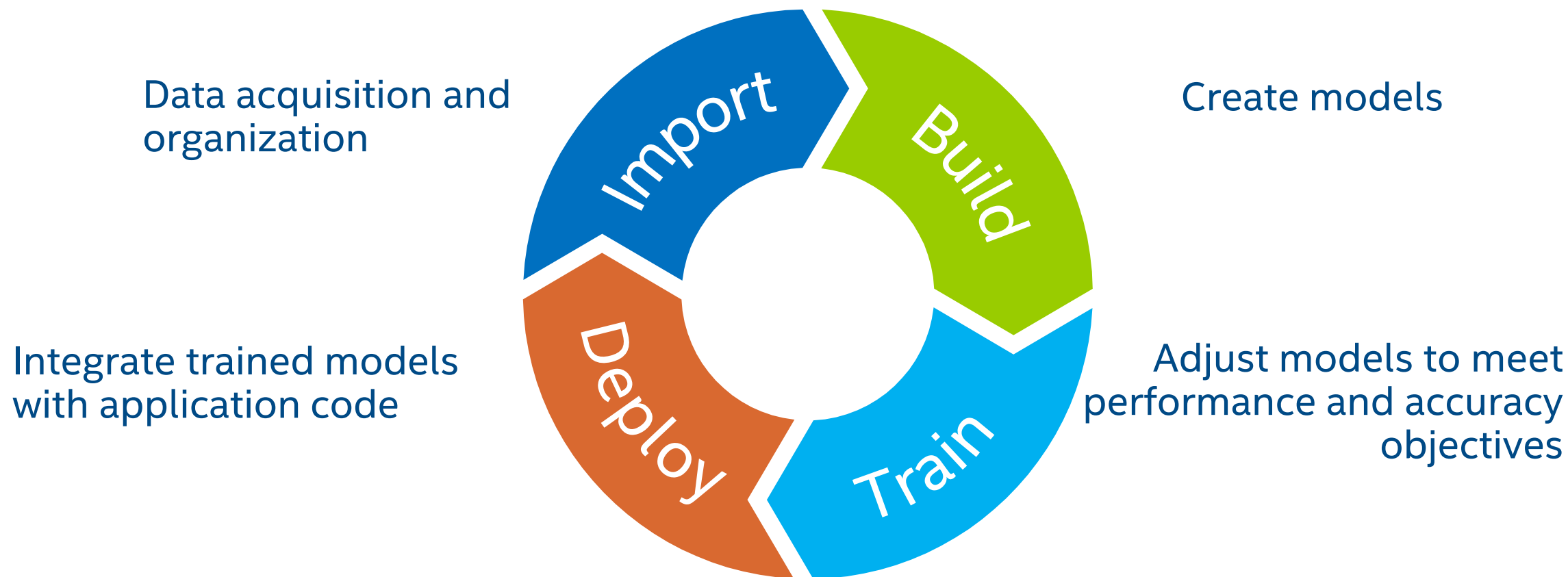
Deploy

Train

Intel® Distribution OpenVINO™ Toolkit Provides Deployment from Intel® Edge to Cloud

intel.

Deep Learning: Training vs. Inference

Training

Human
Bicycle

Strawberry

Lots of Labeled Data!

Forward

Backward

Model Weights

Inference

??????

Forward
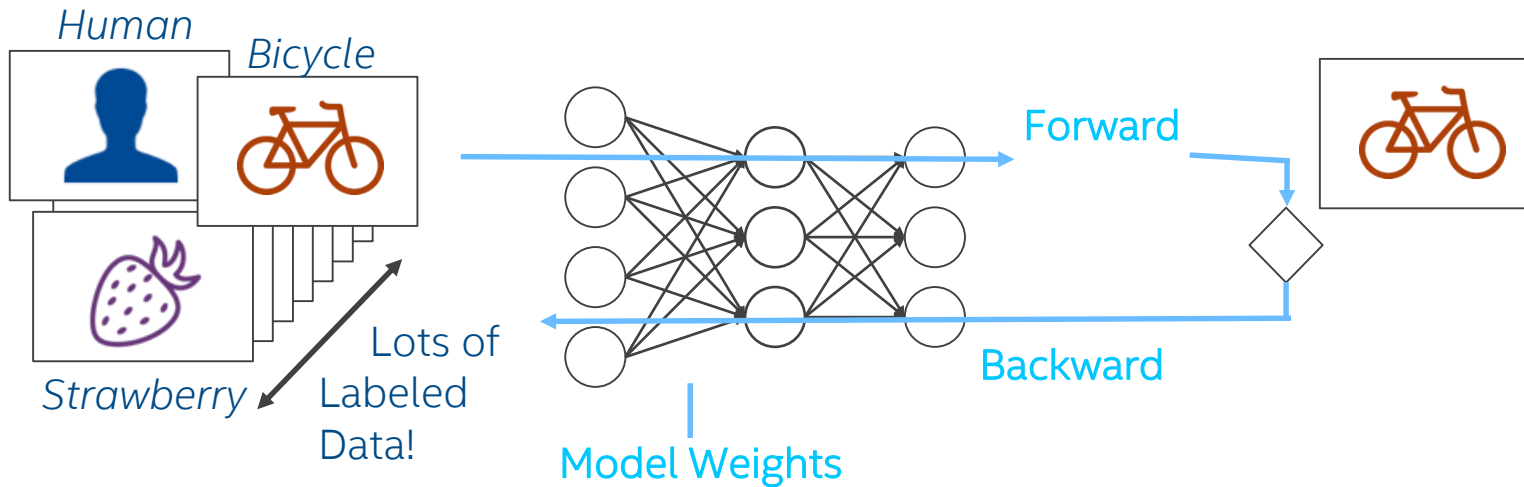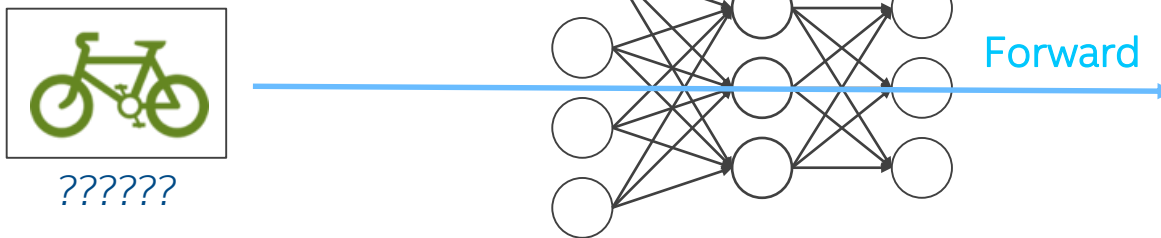
Did You Know?

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

Accuracy

Large NN
Medium NN
Small NN
Traditional Model

Data Set Size
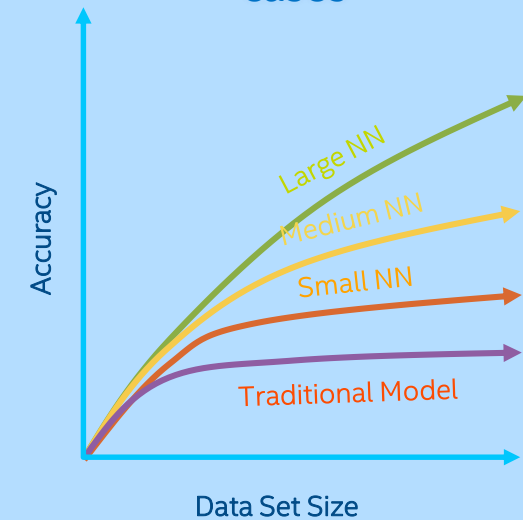
AI COMPUTE CONSIDERATIONS

How do you determine the right computing for your AI needs?

WORKLOADS
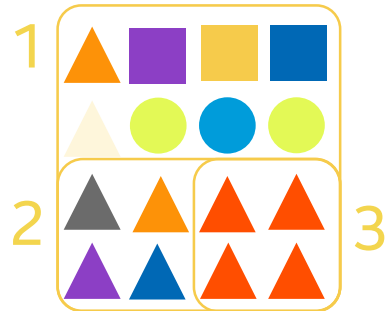
What is my workload profile?

REQUIREMENTS

What are my use case requirements?

DEMAND
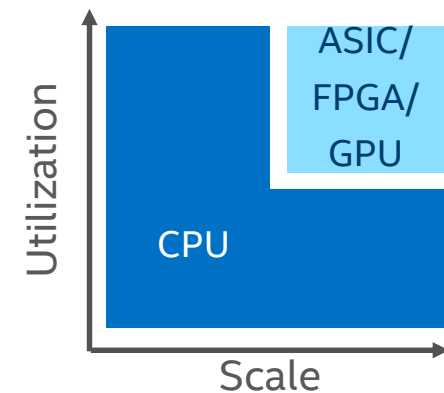
How prevalent is AI in my environment?

WHY INTEL AI COMPUTE?
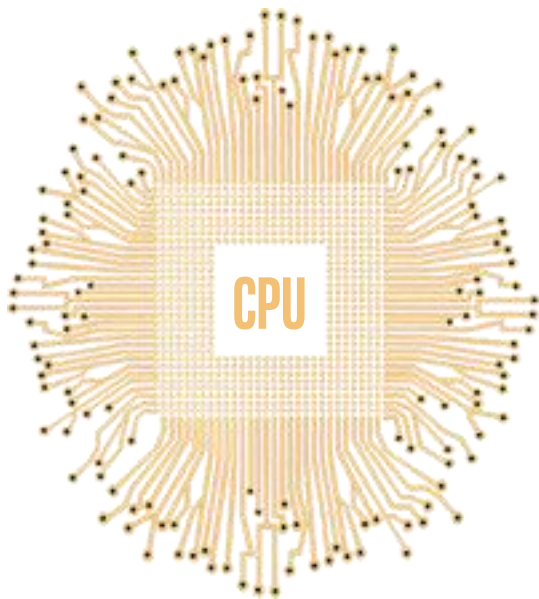
MAXIMIZE	OPTIMIZE	SIMPLIFY	LEAD
CPU	CPU FPGA — GPU ASIC	OPTIMIZED SW DATA PIPELINE ANALYTICS & AI SUPPORT MOVE/STORE	
Get the most out of the foundation for AI from the CPU leader	Choose the right compute for you from the one with all the options	Reduce "moving parts" by building on an optimized AI platform	Lead your industry by aligning with the builder of next-gen AI solutions

Intel® distribution of OpenVINO™ toolkit

- Tool Suite for High-Performance, Deep Learning Inference
- Fast, accurate real-world results using high-performance, AI and computer vision inference deployed into production across Intel® architecture from edge to cloud

High-Performance,
Deep Learning Inference
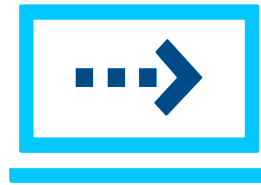
Streamlined Development,
Ease of Use

Write Once,
Deploy Anywhere

New Features from OpenVINO Toolkit 2021.1

- Support for Tiger Lake (11th generation Intel® Core™ processors)
- New capabilities in OpenVINO™ Model Server

- Support for TensorFlow 2.x
- Support for non-computer vision workloads
- (Coming in Q4) Beta Release: Integration of OpenVINO™ toolkit DL Workbench and Intel® DevCloud for the Edge

- Support for GNA 2.0

Three steps for the Intel® Distribution of OpenVINO™ toolkit

1 Build   ## 2 Optimize   ## 3 Deploy

Trained Model

TensorFlow   Caffe   mxnet
KALDI   ONNX
PyTorch   PaddlePaddle

Open Model Zoo
100+ open sourced & optimized pre-trained models available

Model Optimizer
Converts and optimizes trained model using a supported framework

Read, Load, Infer

IR Data
Intermediate **R**epresentation (.xml, .bin)

Inference Engine
Common API that abstracts low-level programming for each hardware

Post-Training Optimization Tool
Reduces model size into low-precision without re-training

Deep Learning Workbench
Visually analyze and fine-tune

Deep Learning Streamer

OpenCV   **OpenCL™**

Code Samples & Demos
(e.g. Benchmark app, Accuracy Checker, Model Downloader)

Model Server
gRPC/ REST Server with C++ backend

Deployment Manager

intel ATOM   intel CORE i7   intel XEON

intel IRIS Pro GRAPHICS

intel MOVIDIUS   Intel GNA (IP)   intel ARRiA 10

Additional Tools and Add-ons from the OpenVINO GitHub Repo

Computer Vision Annotation Tool
This web-based tool helps annotate videos and images before training a model

Dataset Management Framework
Use this add-on to build, transform and analyze datasets

Neural Network Compression Framework
Training framework based on PyTorch* for quantization-aware training

[*NEW*] **OpenVINO™ Model Server**
Scalable inference server for serving optimized models and applications

Training Extensions
Trainable deep learning models for training with custom data

Speed up development with open source resources

Open source resources with pre-trained models, samples and demos

Computer Vision

Object detection
Object recognition
Reidentification
Volumetric segmentation
Semantic segmentation
Instance segmentation
3D reconstruction
Human pose estimation
Image processing
Action recognition
Image super resolution

Audio, Speech, Language

Language processing
Speech to text
Text detection
Text recognition
Natural Language Processing

Other
(Data Generation, Reinforcement Learning)

Compressed models
Image retrieval

And more..

Model Downloader
▪ Provides an easy way of accessing a number of public models as well as a set of pre-trained Intel models

Accuracy Checker
▪ Check for accuracy of the model (original and after conversion) to IR file using a known data set

PRE-TRAINED MODELS
https://github.com/opencv/open_model_zoo

Speed up development with open source resources

Open source resources with pre-trained models, demos, and tools

The Open Model Zoo demo applications are console applications that demonstrate how you can use your applications to solve specific use-cases.
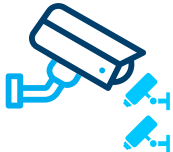
Smart Classroom
Recognition and action detection demo for classroom settings

Multi-Camera, Multi-Person
Tracking multiple people on multiple cameras for public safety use cases

Gaze Estimation
Face detection followed by gaze estimation, head pose estimation and facial landmarks regression.

Weld Porosity Detection
Demonstrates how to find defects in welding

Person Inpainting
Removes unwanted people in images or videos

And more..

DEMO APPLICATIONS

https://github.com/opencv/open_model_zoo

Choose between Release Types

Standard Releases vs Long-Term Support Releases

Standard Release (3-4 releases a year): Users looking to take advantage of new features, tools and support in order to keep current with the advancements in deep learning technologies

Long-Term Support Release: Users looking for a stable and reliable version that is maintained for a longer period of time, and are looking for little to no new feature changes

intel

Supported OS and Install Options [2021.1]

https://software.intel.com/content/www/us/en/develop/tools/openvino-toolkit.html

- **Operating Systems**
  - Ubuntu 18.04.x long-term support (LTS), 64-bit
  - CentOS 7.6, 64-bit (for target only)
  - Yocto Project v3.0, 64-bit (for target only and requires modifications)
  - Microsoft Windows* 10 64-bit
  - macOS* 10.15
  - Raspbian* Buster, Stretch

- **Install From Images and Repositories**
  - GitHub
    - https://github.com/openvinotoolkit/openvino.git
  - Anaconda Cloud
    - https://anaconda.org/intel/openvino-ie4py

- Python* Package Installer (PIP)
  - https://pypi.org/project/openvino-python/
- Docker
  - Install from Image file
  - Download from DockerHub
- APT
  - sudo apt-cache search install-openvino-runtime-ubuntu18
- YUM
  - sudo install intel-openvino-runtime-centos7

- **Intel® Edge Software Hub**
  - Edge Insights for Vision