Horizon
Robotics

OpenExplorer™
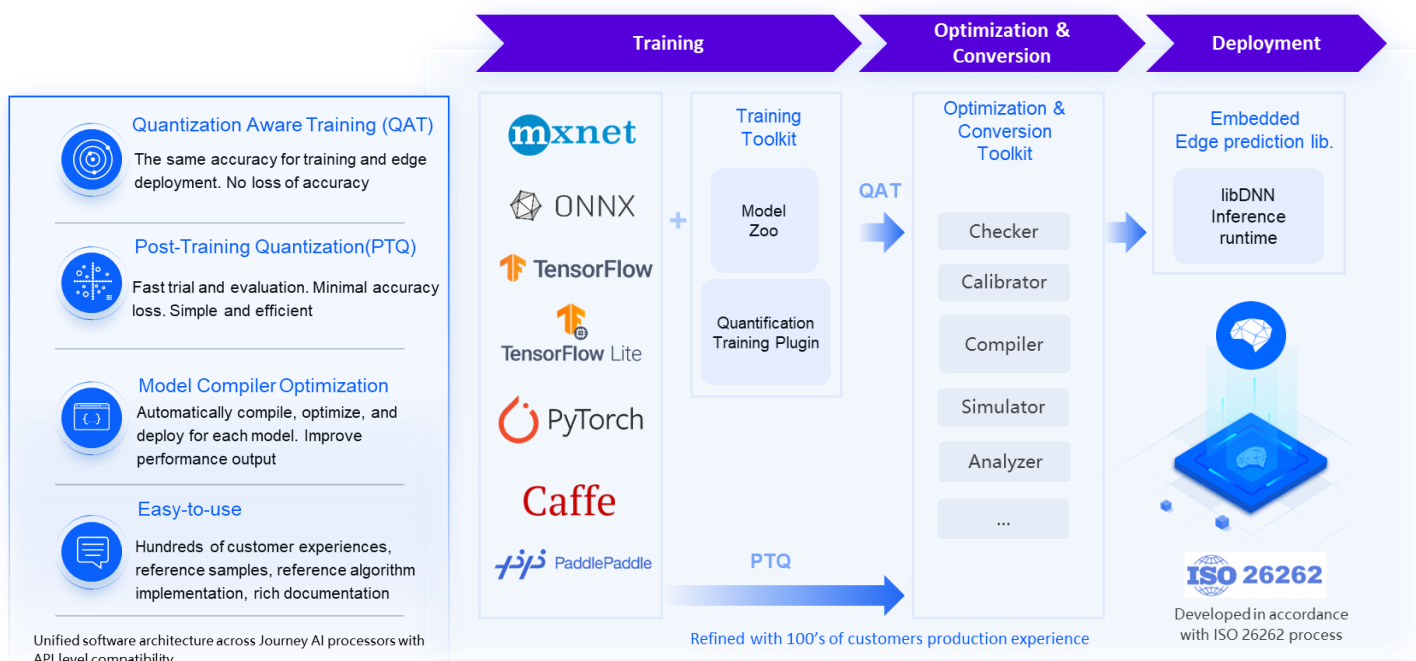## A Developer's Toolkit for Horizon Robotics Journey™ Automotive Solutions

Train, optimize and deploy your Deep Learning models for Horizon BPU™ engine

Horizon OpenExplorer™ toolkit empowers deep learning software developers to fully leverage the performance and energy efficiency of Horizon's Brain Processing Unit™ (BPU™). Built on top of Linux, this comprehensive toolkit is easy to use. With OpenExplorer, developers can train their proprietary neural networks on popular frameworks and quantize for performance without loss of accuracy. The OpenExplorer model optimizing compiler automatically delivers the best BPU performance for your deep learning models, ready to be deployed to Journey automotive grade solutions for your smart automotive applications.

Flexibility

Performance

Ease of use
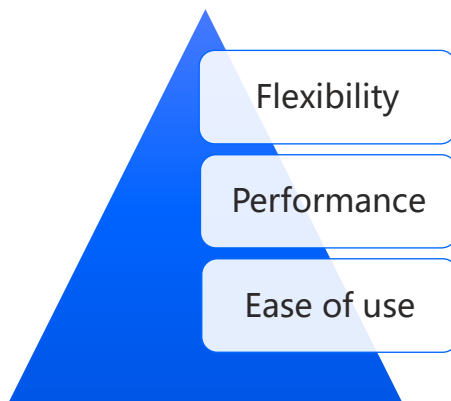
OpenExplorer is a modern, flexible toolkit, enabling neural networks developers to achieve an optimal performance and energy efficiency

OpenExplorer toolkit workflow:

Training | **Optimization & Conversion** | **Deployment**

Quantization Aware Training (QAT)
The same accuracy for training and edge deployment. No loss of accuracy

Post-Training Quantization(PTQ)
Fast trial and evaluation. Minimal accuracy loss. Simple and efficient

Model Compiler Optimization
Automatically compile, optimize, and deploy for each model. Improve performance output

Easy-to-use
Hundreds of customer experiences, reference samples, reference algorithm implementation, rich documentation

Unified software architecture across Journey AI processors with API level compatibility

mxnet
ONNX
TensorFlow
TensorFlow Lite
PyTorch
Caffe
PaddlePaddle

Training Toolkit
Model Zoo
Quantification Training Plugin

QAT

PTQ

Optimization & Conversion Toolkit
Checker
Calibrator
Compiler
Simulator
Analyzer
...

Embedded Edge prediction lib.
libDNN Inference runtime

ISO 26262
Developed in accordance with ISO 26262 process

Refined with 100's of customers production experience
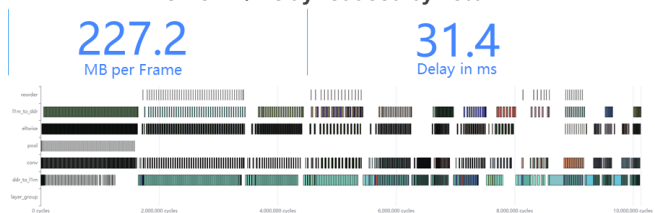
Train and quantize:

- Framework support: ONNX, Tensorflow, PyTorch, Caffe.
- Two methods of quantization supported: Quantization-aware training for near zero accuracy loss
- Post-training quantization for ease-of-use and rapid porting to Journey™ solution BPU™.
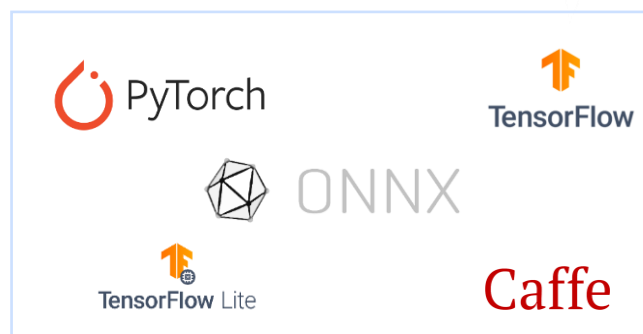
Before Compiler Optimizations: Limited by memory access
Perf 1x / Delay 1x

730.9
MB per Frame

124.76
Delay in ms

Analyze, compile and optimize:

- Model structure is checked to run efficiently on the BPU™
- The Analyzer provides a snapshot of resource utilization
- The Model compiler automatically performs optimizations, such as layer split and fuse, to increase data and instruction level parallelism. It also schedules store and load processes as efficiently as possible for the workload
- Result is a higher inference performance at low latency
- Performance improves with more recent version of popular neural networks models

After Compiler Optimizations: No memory access bottleneck
Perf 3.2x / Delay reduced by 75%

227.2
MB per Frame

31.4
Delay in ms

MobileNet-V2 @ 224x224x3 classification from ImageNet

Deploy models:

- Compiled models are deployed easily using the embedded prediction library (LibDNN)
- Only simple processes such as loading the compiled models, preparing the input and output space, and calling the inference interfaces are required in sequence
- Horizon provides abundant models and application samples with documentation for automotive use cases to accelerate your project

Horizon support

Customer success is our number one objective. Horizon provides documentation, design examples, training services and support, for a compelling developer's experience. Please visit our website at en.horizon.cc or contact your Horizon representative for more details.